Demonstration of Tissue Factor expression using a Tissue Factor – Green Fluorescent Protein Reporter Model

Charles Keith Rizleigh Gomez

Haemostasis and Thrombosis
MRC Clinical Sciences Centre
Imperial College Faculty of Medicine
London

July 2007

Submitted in part fulfilment of the requirements of the University of London
for the degree of Doctor of Philosophy

Acknowledgements

A number of individuals provided invaluable support and assistance during this project. In particular this work would not have been possible without the guidance of Dr. John McVey who supervised the entire project.

The work was carried out within the facilities of the Clinical Sciences Centre of the Medical Research Council, London. Invaluable advice and technical support was provided by my colleagues in the Haemostasis and Thrombosis group: Adam Carpenter, Dr. Geoff Kemball-Cook and Dr. Yan Li. I would also like to thank Dr. Kate Michaelides for her assistance in developing tissue culture protocols, Prof. Ted Tuddenham for valuable discussions and Jonathan Godwin and Zoe Webster for their technical support in the generation of transgenic mice.

Declaration

The experimental work described in this thesis was carried out solely by the author with the following exceptions. In these experiments the work was carried out in conjunction with colleagues in the Clinical Sciences Centre, MRC, London:

1. Blastocyst injections – Jonathan Godwin, Transgenics and ES cell Laboratory
2. BAC targeting and immunohistochemistry – Adam Carpenter, Haemostasis and Thrombosis
3. Pronuclear injections – Zoe Webster, Transgenics and ES cell Laboratory

C.K.R. Gomez J.H. McVey

Abstract

Blood coagulation is initiated when circulating factor (F) VII comes into contact with membrane bound tissue factor (TF). Subsequently a series of enzymatic cleavages results in the formation of insoluble fibrin. The coagulation network is of fundamental importance in maintaining the integrity of the circulation and is closely integrated with other processes such as inflammation. Both under- and over-activity of the network have significant pathological consequences and thus the network is tightly regulated. The first point of regulation is in the formation of the TF•FVIIa initiating complex. Circulating FVII is kept separate from membrane-bound TF by an intact vascular wall. The initiation of coagulation is thus dependant on the site of TF expression. In addition abnormal TF expression has been shown to be important in various pathological processes such as atherosclerosis and cancer metastasis. The accurate demonstration of TF activity is of critical importance in the understanding of these processes.

This thesis describes the development of a TF-Enhanced Green Fluorescent Protein (EGFP) reporter model for analysis of TF expression. A targeting vector containing the cDNA for EGFP in-frame with and flanked by regions homologous to the mouse TF gene was generated. The vector was designed such that EGFP would be knocked in to the TF locus by homologous recombination to produce a fusion protein with EGFP tagged onto the intact cytoplasmic domain of TF. Two strategies were used for generating transgenic mice. A targeting vector was electroporated into mouse embryonic stem (ES) cells and correctly targeted ES cells were then used in blastocyst injections. Alternatively a bacterial artificial chromosome (BAC) containing the mouse TF gene and flanking regions was targeted in a bacterial host using a similar vector. The targeted BAC was used in pro-nuclear injections of fertilized mouse ova resulting in random integration into the mouse genome. The function of the TF-EGFP reporter model was investigated by analysis of transgene expression in whole tissues and specific cell types taken from transgenic mice.

Contents

Abbreviations used in the text

(c)DNA	(complementary) deoxyribonucleic acid
(E)GFP	(enhanced) green fluorescent protein (if italicised refers to the gene)
(m)F3	(mouse) tissue factor gene
(m)RNA	(messenger) ribonucleic acid
APC	activated protein C
BAC	bacterial artificial chromosome
CHO	Chinese hamster ovary
CmR	chloramphenicol resistance
CPK	Corey, Pauling and Kulton atomic modelling scheme
C_T	cycle threshold in quantitative real-time PCR
DAB	3,3' diaminobenzidine
DAPI	4',6-diamidino-2-phenylindole
DSB	double strand break
EC	endothelial cell
EDTA	ethylenediaminetetraacetic acid
EGF	epidermal growth factor
ELISA	enzyme-linked immunosorbent assay
ES	embryonic stem
FACS	fluorescence-activated cell sorting
FCS	foetal calf serum
GAPDH	glyceraldehyde-3-phosphate dehydrogenase
HA	homologous arm
HEK	human embryonic kidney
HR	homologous recombination
HRP	horseradish peroxidase
HSVTK	herpes simplex virus thymidine kinase
LB	Luria-Bertani
LPS	lipopolysaccharide
MAP	mitogen-activated protein
MEM	minimum essential medium
MGB	minor groove binding
MNP	mouse normal plasma
MRC	Medical Research Council

NeoR	neomycin resistance
NFQ	non-fluorescent quencher
NHR	non-homologous recombination
PAR	protease-activated receptor
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PDB	Protein Data Bank
PDI	protein disulphide isomerase
PE	phycoerythrin
PL	phospholipid
PMA	phorbol 12-myristate 13-acetate
PT	prothrombin time
RE	reaction efficiency in quantitative real-time PCR
RIPA	radioimmunoprecipitation
RT	reverse transcriptase
SpecR	spectinomycin resistance
TBE	tris borate EDTA buffer
TBS-A	tris base saline albumin
TF	tissue factor
TFPI	tissue factor pathway inhibitor
TG	transgenic
UTR	untranslated region
VEGF	vascular endothelial growth factor
WT	wild-type
YAC	yeast artificial chromosome

Chapter 1 - Introduction

1.2 Overview

It is now universally recognised that the principal physiological initiator of blood coagulation is the complex formed by the interaction between tissue factor (TF) and factor (F) VIIa. Regulation of TF expression controls the site of this interaction and therefore controls the initiation of the coagulation network. Aberrant TF expression has significant pathological consequences. In mouse models with reduced TF levels there is extensive morbidity caused by haemorrhage whereas in humans thrombosis and cancer metastasis are directly associated with increased TF expression. Thus accurate demonstration of TF expression is of fundamental importance in understanding these and many other processes.

This thesis describes the development of a TF-enhanced green fluorescent protein (EGFP) reporter model for analysis of TF expression. A targeting vector containing the cDNA for EGFP in-frame with and flanked by regions homologous to the mouse TF gene (*mF3*) was generated. EGFP was 'knocked-in' to the TF locus in embryonic stem (ES) cells, or in bacteria carrying a bacterial artificial chromosome (BAC) containing the mouse TF gene and flanking regions, by homologous recombination. The predicted fusion protein product contains the fluorescent marker tagged on at the C-terminal end of the intact cytoplasmic domain of TF. Two strategies were used for generating transgenic mice. The targeting vector was electroporated into mouse ES cells and correctly targeted ES cells (resulting from homologous recombination at the endogenous locus) were then used in blastocyst injections. Alternatively a targeted BAC was used in pro-nuclear injections of fertilized mouse ova resulting in random integration into the mouse genome. The TF-EGFP reporter model was characterised by analysis of mRNA and protein expression in transgenic mice resulting from pro-nuclear injection of targeted BAC.

1.3 A brief history of tissue factor

In the early part of the nineteenth century it was generally believed that coagulation was an intrinsic property of blood. This was consistent with the finding that blood clotted when removed from the circulation and placed in a test tube. It was presumed that blood vessels and extravascular tissues had some anticoagulant property that prevented clot formation *in vivo*. In 1834 De Blainville reported that

animals receiving intravenous injections of suspensions of brain tissue died from intravascular coagulation. At the time this was a revolutionary discovery as it provided direct evidence that there were procoagulant extravascular tissue components in contradiction of prevailing theory. Subsequently other researchers were able to induce similar effects by injection of a variety of tissue extracts confirming De Blainville's findings and leading to the concept of an extravascular trigger for blood coagulation (reviewed in Bachli, 2000; and Owen, 2001).

The exact nature of the procoagulant tissue component remained the matter of much debate. In the early part of the twentieth century many researchers thought that it was directly responsible for converting fibrinogen to fibrin and that the tissue component was therefore the same as the previously discovered thrombin. Gradually it became clear that other enzymes were involved in intermediate steps in blood coagulation and while the tissue component could initiate the network, it was not by itself directly responsible for the majority of fibrin formed. These disagreements regarding the function of the procoagulant tissue component spawned a variety of terms in the literature. Thromboplastin and thrombokinase were perhaps the most widely used but terms such as tissue globulin, koagulin and cell-fibrinogen were also used to refer to essentially the same component. There were diverse names for other coagulation factors as well and to produce some uniformity the International Committee for the Nomenclature of Blood Clotting Factors was formed. In 1958 agreement was reached on a system in which the coagulation factors were identified by Roman numerals (reviewed in Giangrande, 2003). Thromboplastin was designated as FIII. More recently for a few coagulation factors the Roman numeral nomenclature has been replaced by descriptive names and now FIII and thromboplastin (subsequently CD142) are more commonly referred to as tissue factor (TF).

1.4 Tissue factor structure

The gene for mouse (m) TF, *mF3*, spans some 11 kb on chromosome 3 and is split into six exons (Mackman *et al.*, 1992). Exon 1 encodes the signal peptide, exons 2-5 the extracellular domains which form the major part of the molecule and exon 6 the transmembrane-spanning domain and cytoplasmic tail (Table 1.1). The primary translation product of the mRNA is a 294 residue polypeptide (Hartzell *et al.*, 1989). Removal of the signal peptide leaves a mature protein in which the N-terminal residues (1-223) contain the extracellular domains, followed by a 23 residue transmembrane domain and a 20 residue cytoplasmic tail.

Exon	Length (bp)	Peptide region	Residues (aa)	
			Start	End
1	194	Signal	-28	-1
2	106	F1	+1	109
3	218			
4	179	F2	110	223
5	160			
6	937	Transmembrane	224	246
		Cytoplasmic	247	226

Table 1.1 Exons in the mouse TF gene with corresponding protein domains. F1 and F2 refer to the fibronectin type III-like domains.

The extracellular region consists of two fibronectin type III-like domains (F1 and F2) and has sequence homology with members of the cytokine receptor class II family. There are four potential N-linked glycosylation sites although the significance of carbohydrate moieties in the function of mTF remains unclear. Human (h) TF similarly has three potential glycosylation sites, of which two are glycosylated *in vivo* but the recombinant protein produced in bacteria (and therefore without carbohydrate side chains) has full procoagulant function (Paborsky & Harris, 1990; Paborsky *et al.*, 1991).

Although the crystal structure of mTF is not yet available, hTF and the human TF·FVIIa (inhibited) complex have been crystallised (Harlos *et al.*, 1994; Muller *et al.*, 1994; Banner *et al.*, 1996a). An alignment of the predicted TF polypeptide sequences from different species demonstrates that there is a greater conservation of residues in the extracellular domains in comparison with the transmembrane and cytoplasmic domains (Figure 1.1). This similarity is presumably maintained in their tertiary structures as in many cases TF and FVIIa from different species can successfully propagate coagulation *in vitro* (Kadish *et al.*, 1983; Janson *et al.*, 1984). However, there are subtle differences between species and in some cases these are well known to affect function. For example in the measurement of activity levels of some FVII mutants causing FVII deficiency in humans, lower results are obtained with rabbit or bovine thromboplastin compared with recombinant human thromboplastin

(Takamiya *et al.*, 1995). This suggests that at the interface between TF and FVII there are slight differences in the critical residues that can affect binding. The interaction with FVIIa occurs extensively over both extracellular domains and various critical residues have been identified by site-directed mutagenesis studies (Ruf *et al.*, 1994; Gibbs *et al.*, 1994; Muller *et al.*, 1994; Kelley *et al.*, 1995). An important requirement for this interaction is that TF forms a firm surface for FVII to attach to. This rigidity is contributed to by the presence of a single intra-domain disulphide bridge in each of the fibronectin type III-like domains (Figure 1.2).

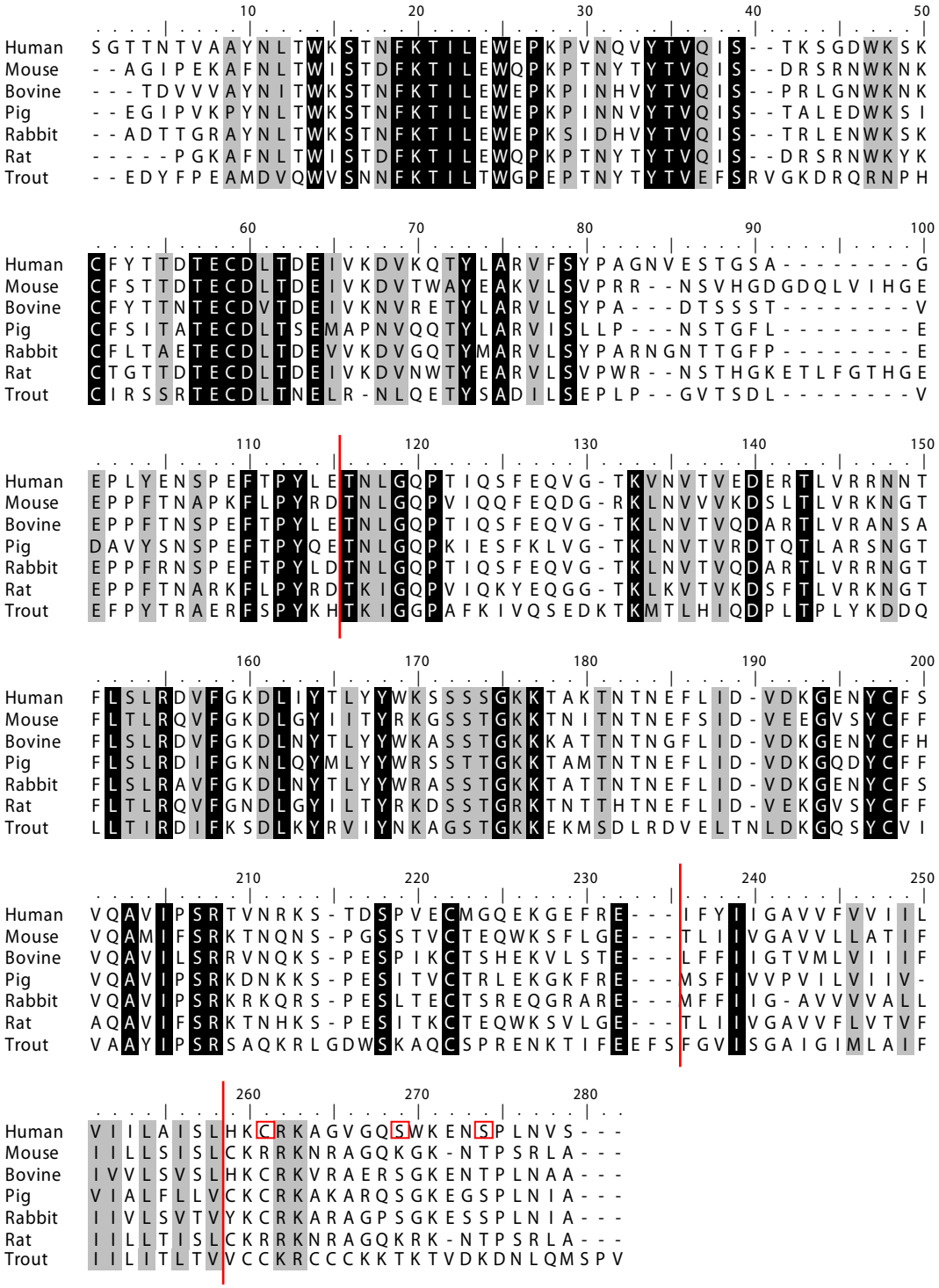


Figure 1.1 Alignment of the mature TF polypeptide sequence from selected species created using Clustal W with PAM120 matrix. White on black and black on grey represent positions with complete and high similarity respectively. Red vertical lines demarcate the domain boundaries. The red boxes indicate the palmitoylated and phosphorylated residues in human.

The F1 and F2 domains form a platform for binding and activation of FVII and localization of the resulting TF·FVIIa complex to a membrane surface. FVIIa on its own has very limited proteolytic activity but this is increased by more than a millionfold by the co-factor function of TF (Morrissey, 2001). The association of the TF·FVIIa complex with appropriate phospholipids is clearly important. TF truncated after the extracellular domains provides a much reduced enhancement of FVIIa activity towards macromolecular substrates, than the full-length, membrane-bound protein. At the C-terminal end of FVII is a Gla domain that contains γ-carboxylated glutamate residues that bind Ca^{2+} ions (Figure 1.3). This domain is usually important in interacting with the membrane. In the TF·FVIIa complex the transmembrane domain of TF serves to provide this association and despite its proximity to the TF molecule, the FVII Gla domain is dispensable for binding to TF (Neuenschwander & Morrissey, 1994). The exact nature of the membrane anchor in the TF·FVIIa complex seems to be unimportant. Replacement of the transmembrane domain with a phosphatidylinositol membrane anchor restores full function to the truncated protein, indicating that it is the location on phospholipid rather than the transmembrane domain itself which is critical (Paborsky *et al.*, 1991).

In hTF the cytoplasmic tail contains protein kinase C-dependent phosphorylation sites at Ser253 and Ser258 (Mody & Carson, 1997) and a palmitoylation site at Cys245 (Bach *et al.*, 1988). In mTF there is similarly a palmitoylation site at Cys247 and a single phosphorylation site at Thr261. Although there is relatively poor conservation between species in the cytoplasmic domain there is maintenance of potential palmitoylation and phosphorylation sites across species (Figure 1.1). These post-translational modifications are seen *in vivo* in mice and humans which is normally indicative of a role in intra-cellular signalling pathways. The role of TF·FVIIa in signalling pathways will be discussed later but it appears that the cytoplasmic domain is not required for the normal physiological functions of TF since mice expressing only a truncated from of the protein lacking the cytoplasmic tail are viable and have no obvious abnormalities (Melis *et al.*, 2001). However it does seem that the cytoplasmic domain may be important in some pathological processes.

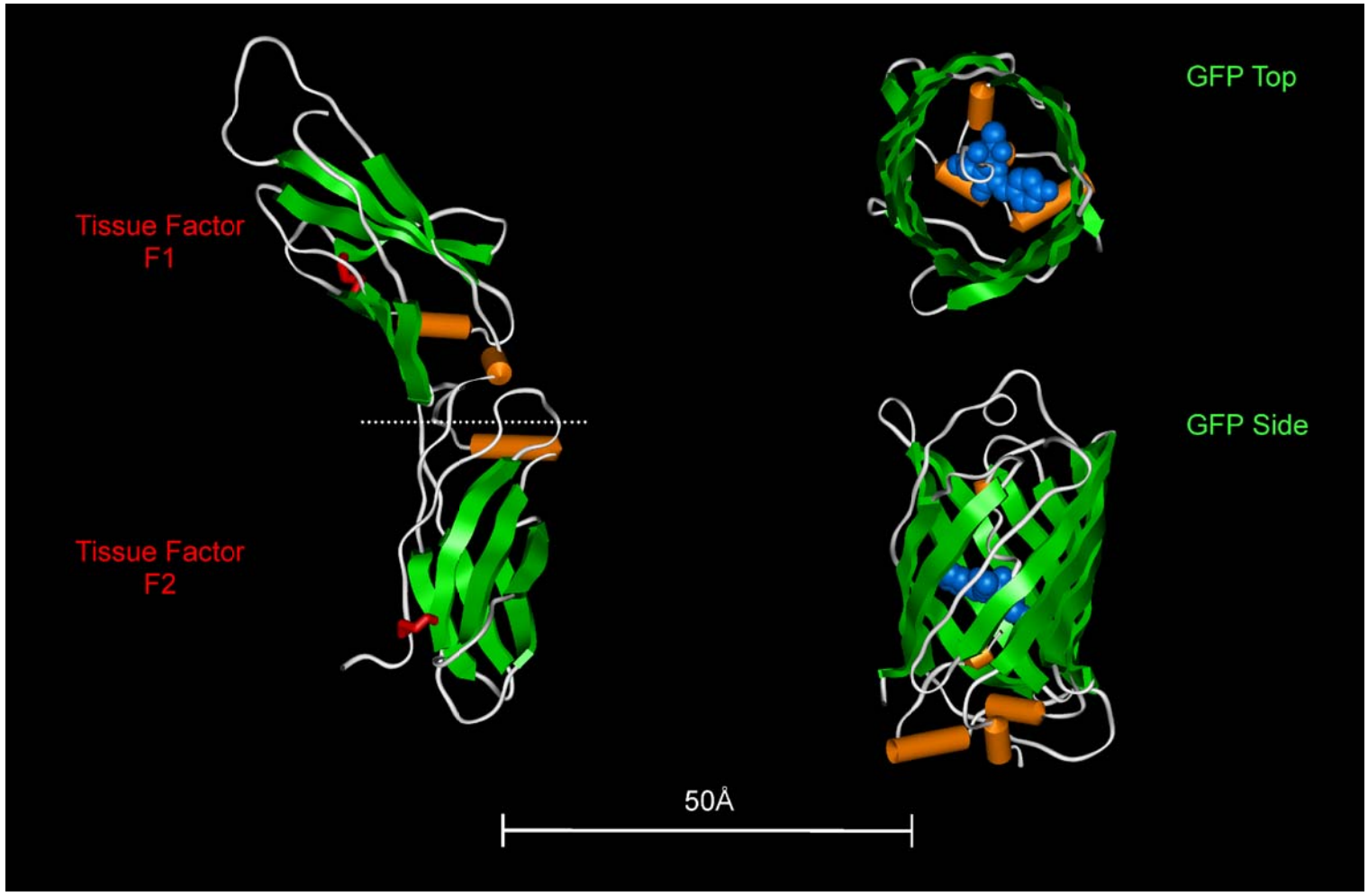


Figure 1.2 Secondary structure of TF and GFP S65T drawn to scale. β-sheets are in green and α-helices in orange. The intra-domain disulphide bonds in TF are shown by red lines. The dashed white line shows the approximate boundary between the two fibronectin III-like domains which are orientated at an angle of 125°. GFP is shown in side and top-down views. The fluorophore (blue CPK spheres) lies at the centre of a cylinder formed by several β-sheets. PDB IDs: 1BOY (Harlos *et al.*, 1994) and 1C4F (Elsliger *et al.*, 1999).

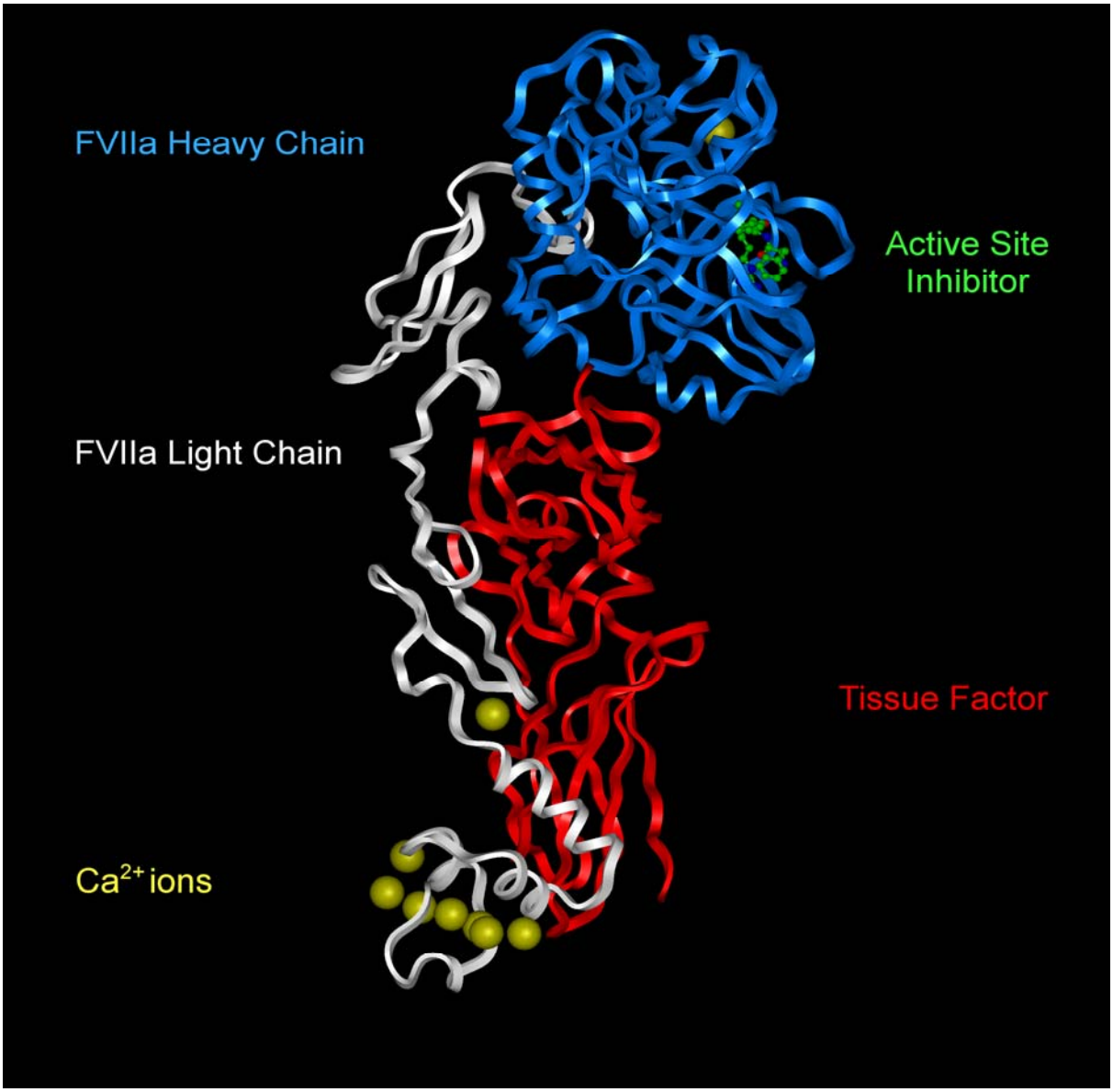


Figure 1.3 Molecular graphic demonstrating the interaction between the extracellular domains of TF (red ribbons) and FVIIa (heavy chain in blue ribbons, light chain in white ribbons). The active site inhibitor is shown as a ball and stick model. In this representation the membrane platform would be at the bottom of the figure. The activation of FX is dependent on a number of Ca^{2+} ions (yellow CPK spheres) which are bound in the Gla, EGF1 and catalytic domains of FVIIa. PDB ID: 1 DAN (Banner *et al.*, 1996b).

1.5 Tissue factor in normal physiology

1.5.1 Initiation of coagulation

The purpose of the coagulation network is to produce a focused burst of fibrin clot formation at the site of vascular injury. Not only does the process need to be rapid, it also needs to be tightly controlled to prevent the potentially disastrous consequences of clot dissemination through the intact vasculature. TF expression patterns in human and mouse tissues have been demonstrated by *in situ* hybridization using nucleic acid probes against TF mRNA (Wilcox *et al.*, 1989; Luther *et al.*, 1996). TF protein has similarly been demonstrated in primates and rodents by antibodies using classical immunohistochemistry (Drake *et al.*, 1989; del Zoppo *et al.*, 1992; Bogdanov *et al.*, 2006). These studies showed that TF is expressed in the outer adventitia of blood vessels and various extravascular supporting tissues such as fibrous or connective tissue. It is not normally present in the inner media of the blood vessel or the luminal endothelial lining. This has led to the concept of TF forming a 'haemostatic envelope' around blood vessels. In the intact vasculature the physical separation of TF from circulating FVII prevents coagulation. Disruption of the vessel lining allows the TF·FVIIa complex to form thereby initiating the network (reviewed in McVey, 1999; Morrissey, 2001; and Gomez & McVey, 2006). Thus control of TF expression patterns provides the first point of regulation for the coagulation network.

The original 'waterfall' hypothesis of coagulation described two independent pathways by which coagulation could be initiated. The TF·FVIIa complex initiated the extrinsic pathway while contact factors initiated the intrinsic pathway. While this hypothesis fitted with the results of *in vitro* tests of coagulation, it could not adequately explain the phenotypes of classical haemophilia patients who bled despite having an intact extrinsic pathway. Revision of the waterfall hypothesis has produced the currently accepted paradigm in which coagulation *in vivo* is initiated entirely through the TF pathway (Figure 1.4). The first step occurs when TF binds to FVII or FVIIa to form the TF·FVIIa complex, also referred to herein as the initiator complex.

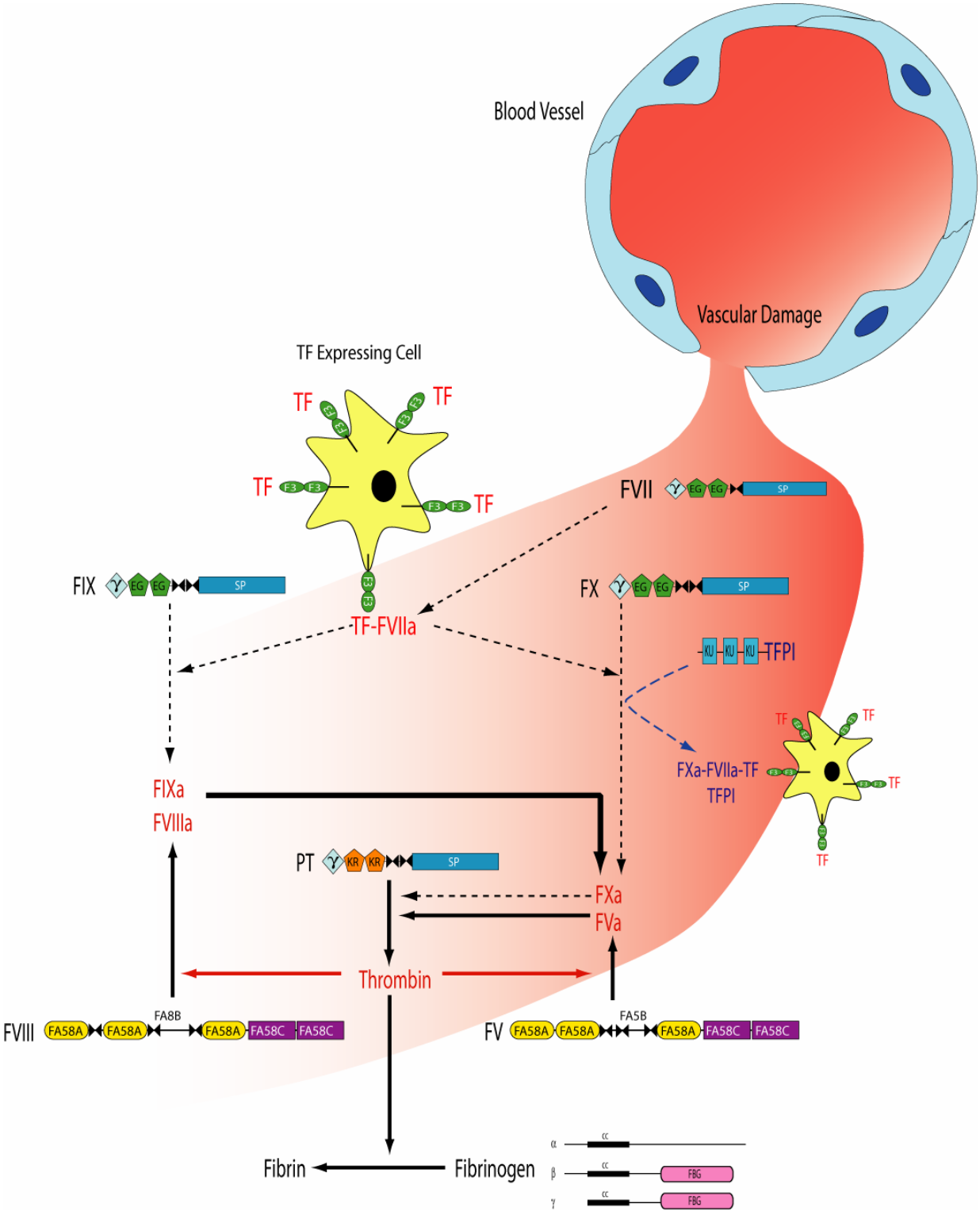


Figure 1.4 TF-dependent initiation of coagulation following disruption of the vessel wall. Extravasation allows FVII to form the initiator complex with TF generating a small amount of thrombin (dashed lines). The thrombin thus generated participates in various positive feedback reactions resulting in amplification of coagulation and culminating in the formation of a fibrin clot (solid lines). The domain organisation of proteins is represented. Zymogens and pro-cofactors are shown in black, active forms in red and inhibitory reactions in blue. Reproduced courtesy of Dr. J. McVey.

About 1% of total FVII circulates in the activated form in normal individuals, although exactly how this FVIIa fraction is maintained is unclear. Zymogen FVII bound to TF is extremely sensitive to proteolytic activation which results in the formation of a neo N-terminus that is inserted into the catalytic domain adjacent to the active site. A number of enzymes may activate FVII *in vitro,* such as FXa, FIXa and FVIIa itself. There is also some evidence for a specific FVII activating protease (Romisch *et al.*, 1999). Which of these are physiologically important remains uncertain. In any case TF·FVIIa may then form further macromolecular complexes with its substrates FIX and FX. Following these initial steps propagation and amplification of coagulation to generate the large burst of localized thrombin essential for haemostasis requires the various factors and co-factors depicted in Figure 1.4.

Generally in mammalian biology serine proteases are inhibited by serine protease inhibitors (serpins) but *in vitro* data indicates that plasma serpins such as antithrombin inactivate the initiator complex too slowly to be relevant and that the main inhibitor of the initiator complex is tissue factor pathway inhibitor (TFPI) (van't Veer & Mann, 1997). This Kunitz-type inhibitor binds to form a quaternary complex that inhibits both FVIIa and FXa. It has been shown that inactivation of TF·FVIIa by TFPI is dependent on FXa indicating that TFPI only inhibits the complex after the pathway to thrombin generation has been commenced (Broze, Jr., 1995). It therefore serves to shut down the initiator complex once its purpose has been served thereby limiting the spread of coagulation. On account of the dependence on FXa, it is generally assumed that TFPI binds free FXa first and TF·FVIIa subsequently. However, there is kinetic data suggesting that TFPI can bind the enzyme-product complex TF·FVIIa·FXa and thus it remains to be resolved which of these pathways occurs *in vivo* (Baugh *et al.*, 1998; Panteleev *et al.*, 2002) (Figure 1.5). The mechanism of inhibition by TFPI is quite different from the typical action of serpins on serine proteases. Following cleavage of the scissile bond the serpin remains covalently linked to the serine protease acting as a 'suicide substrate' and inhibiting the enzyme. The Kunitz domains of TFPI interact specifically with their target protease: K1 with FVIIa and K2 with FXa. Each contains a reactive centre loop which is inserted into the active site but is resistant to cleavage, thereby acting as a pseudo-substrate (Bajaj *et al.*, 2001). This leads to irreversible inhibition and the macromolecular complex TF·FVIIa·FXa·TFPI which is internalized and degraded intra-cellularly (Sevinsky *et al.*, 1996).

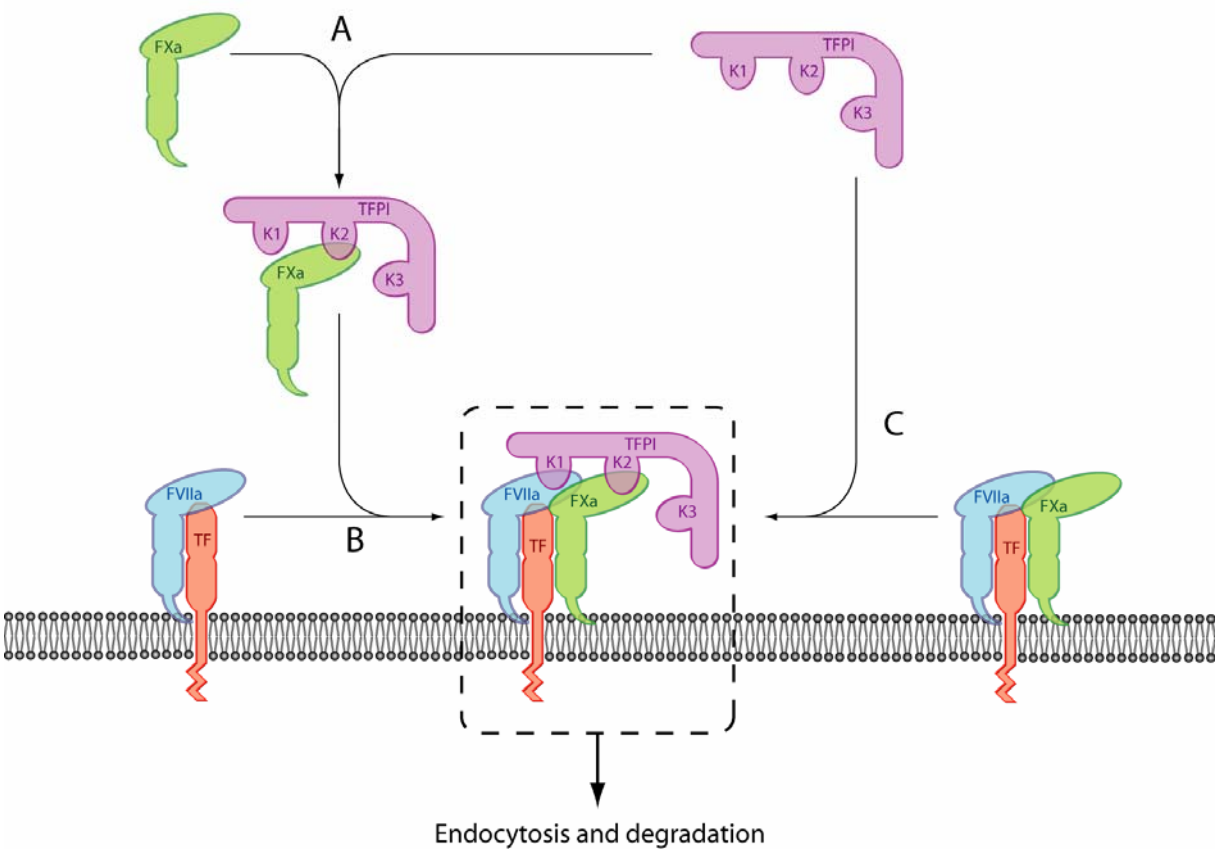


Figure 1.5 Possible mechanisms for the inhibition of the initiator complex by TFPI. TFPI inhibits FVIIa and FXa through its K1 and K2 domains respectively. TFPI either binds free FXa first (A) before binding to TF·FVIIa (B) or binds directly to TF·FVIIa·FXa (C). The role of the K3 domain and C-terminus is unclear although the latter may be involved in stabilising the complex on the membrane. The serine proteases are irreversibly bound in the final TF·FVIIa·FXa·TFPI macromolecular complex which is internalised for degradation.

1.5.2 Tissue factor in development

In addition to its role in coagulation it is now clear that TF has additional roles in other physiological systems. TF knockout mice (*F3-/-*) generally die *in utero* between days 8.5–10.5 (Bugge *et al.*, 1996; Carmeliet *et al.*, 1996; Toomey *et al.*, 1996). This is a critical time in mouse embryogenesis and knockout models of a number of proteins are lost at the same gestational age with a variety of defects. An obvious possible explanation for the foetal loss is haemorrhage due to lack of TF and a failure to maintain haemostasis. However knockout models of other coagulation factors, such as FX and fibrinogen, can develop normally *in utero* before suffering fatal

haemorrhages in the neonatal period (Suh *et al.*, 1995; Aasrum & Prydz, 2002). These models indicate that deficiencies of proteins that solely affect haemostasis may be tolerated up to the peri-natal period and that TF must have *in utero* functions that are independent of its role in fibrin clot formation. Exactly what these functions are remains a matter of debate. It may be that TF activates various signalling pathways and it is interesting that knockouts of some protease activated receptors (PARs) have the same phenotype as *F3-/-* mice. What is clear is that the functions of TF *in utero* are dependent on the extracellular domains as lack of the cytoplasmic domain has no detrimental effect on development.

The phenotypes of surviving knockout mice are similar to the clinical phenotypes of homologous coagulation factor deficiencies in humans. Individuals with congenital deficiencies of FX and fibrinogen experience haemorrhagic symptoms from the peri-natal period. There have been no reports of individuals deficient in TF, presumably because these defects would be lethal *in utero*. TF null embryos are characterised by a lack of vessel formation in the yolk sac. This suggests that TF is required for formation or maintenance of the foetal vasculature. Knockout models of other proteins thought to be required for vascular integrity, such as thrombin and vascular endothelial growth factor (VEGF), die at the same stage *in utero* and have a similar appearance. Additional evidence that TF is required for vascular integrity comes from *in vitro* studies showing that inhibition of TF activity results in impairment of vascular network formation (James *et al.*, 2002).

An important question that the TF knockout data raises is whether TF acts alone in embryogenesis or requires FVII to be present. This is partly addressed by the FVII knockout model which develops normally *in utero* and then, as with humans with severe FVII deficiency, suffers fatal peri-natal haemorrhage (Rosen *et al.*, 1997). Although this suggests that FVII production from the embryo itself is not required for foetal development, it does not exclude the possibility that maternal FVII may enter the foetus by transplacental transfer. In the knockout study FVII protein was undetectable by ELISA in *F7-/-* embryos. However, in humans with FVII deficiency haemostasis may be adequate despite undetectable FVII protein, indicating that the lower detection limit of currently available assays can miss physiologically relevant levels of FVII (McVey *et al.*, 2001). Recent work has shown that if the mouse FVII knockout is produced on a background where the maternal FVII level is very low foetal loss occurs *in utero* rather than at term (personal communication from Dr E. Rosen, Indiana, USA). In this situation maternal FVII levels are so low that transplacental transfer is reduced to negligible levels and this data suggests that it is

the TF·FVIIa complex, rather than TF acting alone, that is of critical importance for embryogenesis.

1.6 Tissue factor in pathological processes

1.6.1 Cardiovascular disease

In normal tissues TF expression is tightly regulated and limited to a few specific cell types. Cells which either do not express TF constitutively or express it at very low levels may be induced to produce TF under specific conditions. As previously mentioned the luminal endothelium surface has to be free of active TF to prevent thrombosis. Atherosclerosis is a disease which clearly demonstrates the pathological consequences of aberrant TF expression. This disorder is characterized by endothelial dysfunction ultimately resulting in atherosclerotic plaque formation in a thickened vessel wall. In these lesions the media becomes infiltrated by macrophages derived from circulating monocytes and thickened by proliferation of smooth muscle cells. This results in the formation of pathological foam cells which contain TF (Kaikita *et al.*, 1999; Corti & Badimon, 2002). The rupture of vulnerable plaques leads to exposure of active TF to circulating FVII and subsequent thrombosis (Toschi *et al.*, 1997; Moons *et al.*, 2002; Steffel *et al.*, 2006). Monocytes and endothelial cells (EC) may be induced to express TF by a variety of stimuli, including inflammatory mediators such as lipopolysaccharide, cytokines, lipoproteins and granulocytes (Pearson, 1999; Lwaleed *et al.*, 2001). Much of the data on stimulation of TF expression in these two cell types is derived from *in vitro* studies using primary cells or cell lines. While there is good evidence that cells derived from the monocytic lineage may express TF under specific pathological conditions *in vivo*, evidence for EC expression of TF *in vivo* is scant. Identification of the critical mechanisms involved in the induction of TF in atherosclerotic plaques *in vivo* remains an important focus for future research.

In animal models injury to the vessel wall may be used to study the role of the coagulation network in the healing process. Disruption of the endothelium in normal vessels leads to a local inflammatory response and upregulation of TF expression in smooth muscle cells (D'Andrea *et al.*, 2003). In this case the accumulation of TF occurs without significant involvement of monocytes or neutrophils. The interaction between the coagulation network and inflammatory pathways is important in controlling repair of the vessel wall. However increased TF expression can lead to recurrent thrombosis following injury. The expression of anticoagulant molecules in

the injured vessel walls reduces thrombus formation and TF expression (Chen *et al.*, 2006a).

1.6.2 Neoplasia

Cancer patients have a much higher incidence of thrombosis than the normal population. In many cases thrombosis is the presenting feature of occult malignancy. Many factors contribute to the thrombophilia associated with malignancy and aberrant TF expression has been implicated as a major causative factor. Cancerous organs express high levels of TF in a dysregulated pattern (Callander *et al.*, 1992) and cancer patients have higher levels of circulating TF than normals (Santucci *et al.*, 2000; Fernandez & Rickles, 2002). Activation of blood coagulation is an obvious consequence of this but TF has additional and more fundamental roles in carcinogenesis. Tumours are often rapidly growing in comparison with normal tissues and as such require an expanding vascular supply network. Neovascularisation is therefore seen in many tumours and is dependent on various factors including TF, VEGF, various cytokines and thrombin (Fernandez & Rickles, 2002). In a study using a human melanoma cell line, tumours expressing high levels of TF were more vascular than those with low levels of TF with tumour VEGF production being regulated by TF (Abe *et al.*, 1999). Similarly in a mouse tumour model, tumour angiogenesis and size was found to be highly dependent on TF expression (Belting *et al.*, 2004). Interestingly both of these studies showed that the cytoplasmic domain of TF was important for regulation of TF-dependent angiogenesis.

The dependence of tumours on TF makes it an excellent therapeutic target for novel anti-cancer treatments. The protein is more highly expressed in cancers than normal tissues suggesting that TF-targeted treatments would have a high therapeutic index. A few studies have shown that immunotherapy against TF-expressing cells can induce tumour regression and reduce relapse (Fernandez & Rickles, 2002). While this is encouraging, it is unlikely that TF-directed therapies can be curative in cancer as it is not a causative factor, although it may well be a useful target for adjunctive treatments in the future. Neovascularisation is a characteristic of some non-neoplastic disorders also, such as diabetic eye disease and arthritis. The role of TF-dependent angiogenesis in these disorders has not been thoroughly investigated but they may well prove to be suitable targets for TF directed therapies also.

1.7 Tissue factor signalling

The fact that TF expression may influence various apparently separate biological processes such as coagulation and inflammation suggests an involvement in crosstalk between these systems. The TF cytoplasmic domain contains post-translational modifications which are often indicative of a role in triggering intra-cellular signalling pathways. Furthermore the homology between TF and members of the cytokine receptor class II family suggests potential as a signalling receptor.

Early studies of potential TF·FVIIa signalling demonstrated that binding of FVIIa to TF in tissue culture cells induced intra-cellular Ca^{2+} fluxes and activation of the MAP kinase pathway (Rottingen *et al.*, 1995; Poulsen *et al.*, 1998; Camerer *et al.*, 1999). These reports showed that signalling did not involve the TF cytoplasmic domain but did require the binding of FVIIa. Replacement of FVIIa with active site inhibited FVIIa (FVIIai) abolished the cellular responses even though FVIIai binds effectively to TF. Thus in normal cells signalling is via the TF·FVIIa complex, rather than TF alone, and is dependent on the proteolytic function remaining intact. One would therefore expect that the main physiological inhibitor of the complex, TFPI, would regulate its signalling activity and this does indeed seem to be the case (Ahamed *et al.*, 2005). Signalling by other coagulation factors has also been shown to be dependent on proteolytic activity but until the discovery of the PARs, the mechanism remained obscure. PARs are unusual receptors in that the signalling ligand is uncovered following proteolytic cleavage in the extracellular N terminus (reviewed in Coughlin, 2000). This generates a new N terminus which is inserted into the transmembrane part of the PAR resulting in the cellular response. The requirement for enzymatic activity is thus explained.

A family of PARs is found in both humans and mice (Table 1.2). Thrombin signals via PARs 1, 3 and 4 while the TF·FVIIa complex appears to signal via PAR2. Following recruitment of FX and formation of the TF·FVIIa·FXa complex signalling may occur via PAR1 or PAR2 (Camerer *et al.*, 2000; Riewald *et al.*, 2001). FXa may also signal in the absence of TF·FVIIa. Therefore, it is unclear whether physiologically relevant PAR activation by the initiating complex is dependent on FXa. Because of the overlap in agonist activity between TF·FVIIa and FXa it remains difficult to differentiate which serine protease is predominantly responsible for PAR activation. In addition there is different platelet PAR expression between humans and mice. In humans platelets express PAR1 and 4, whereas in mice it is PAR3 and 4. This makes extrapolation from mouse PAR knockout models less reliable when determining

which PAR activations are important in signalling in man during activation of coagulation. Nevertheless, it does seem that TF·FVIIa does have physiologically relevant signalling roles that are independent of FXa as suggested by the fact that *F10-/-* mice can develop normally *in utero* while *F3-/-* mice do not (Aasrum & Prydz, 2002).

		PAR1	PAR2	PAR3	PAR4
Expression on platelets by species		Human	Not expressed	Mouse	Human and Mouse
Agonist	Thrombin	+	-	+	+
	FXa	+	+	-	-
	TF·FVIIa	?	+	-	-
	APC	+	-	-	-

Table 1.2 Activation of PARs by serine proteases. PARs are expressed in a variety of cells but important species differences are seen in platelets. The effect of selected serine proteases on the known PARs is shown, + = activated, - = not activated, ? = unclear.

As TF signalling seems to be protease-dependent in normal physiology the cytoplasmic domain appears to be dispensable for this function, despite the implications of its post-translational modifications. This hypothesis is supported by studies involving targeted disruption of the cytoplasmic domain. As noted above the *F3-/-* mouse fails to develop *in utero*. Disruption of the extracellular domains of TF results in the same phenotype as the full knockout (Parry & Mackman, 2000). In contrast, mice with TF lacking only the cytoplasmic domain develop normally *in utero*, have a phenotype indistinguishable from normal and the truncated TF molecule demonstrates normal procoagulant function (Melis *et al.*, 2001). The data seem conclusive that in normal physiology neither procoagulant nor coagulation-independent functions of TF require the cytoplasmic domain. The situation seems to be different in the pathological processes already described. Metastasis and tumour growth seem to be dependent on the cytoplasmic domain with the conclusion that

TF signalling occurs via different pathways depending on the specific cell type and its milieu.

One of the main areas of interest in signalling by coagulation factors is in the cross-talk between coagulation and inflammation. The induction of TF in various cell types by inflammatory mediators has been well documented. Inflammation has effects on other coagulation factors also with the net result that it is generally associated with a procoagulant state. Less is known about the reverse situation where coagulation factors modulate inflammatory pathways but this is an area of intensive study. It seems likely that the protein C pathway is the most important part of the coagulation network for regulating, and being influenced by, inflammation. Activated protein C (APC) can signal via PAR1 and the influence of various PAR family members on inflammation has been well recognised (Ruf, 2004; Pawlinski & Mackman, 2004). The influence of TFPI on inflammation has also been well documented. For example administration of TFPI ameliorates the effects of *E. coli*-induced sepsis in various animal models (Creasey *et al.*, 1993; Lwaleed & Bass, 2006). The inflammatory reaction to wire-induced vessel injury is also reduced by expression of TFPI (Chen *et al.*, 2006b). In these experiments a similar effect was seen when hirudin or TFPI was expressed in the vessel wall. This indicates that in this situation it is thrombin generation dependent on TF·FVIIa during the inflammatory reaction that is relevant.

1.8 Circulating tissue factor

It is well known that TF protein may be detected circulating freely in the blood, urine and cerebrospinal fluid (Fareed *et al.*, 1995). Levels of circulating TF are often increased in disease states (Santucci *et al.*, 2000). However, it is worth noting that cell lysis and apoptosis are common in disease and cause circulatory release of many cellular proteins. The presence of TF in the circulation suggests that there may be a soluble form of the protein or that the protein may circulate inserted in membranes in blood cells or cell-derived microparticles. This somewhat contradicts the traditional view of TF as an entirely extravascular initiator of coagulation.

A secondary transcript of *F3* has recently been described in which the gene is alternatively spliced to remove exon 5 (Bogdanov *et al.*, 2003). This produces a TF protein with a novel C terminus lacking a transmembrane domain, but with most of the extracellular region intact. The alternatively-spliced TF was initially described in humans but has recently been detected in mice and is present in blood, tissues and

freshly formed thrombi, representing 10-30% of total TF antigen (Bogdanov *et al.*, 2006; Szotowski *et al.*, 2006). A hotly debated question is whether this soluble form of the protein can initiate or contribute to thrombus formation. This issue remains unresolved with contradictory data from studies using different methods of assaying TF activity (Bogdanov *et al.*, 2003; Butenas *et al.*, 2005; Szotowski *et al.*, 2006). Alternatively-spliced TF can demonstrate procoagulant activity in assays with added or induced phospholipid and supra-physiological levels of FVIIa. However the procoagulant activity of the shortened form is significantly reduced in comparison with full length TF and its relevance in physiological conditions *in vivo* remains doubtful.

Circulating TF may also be found in the form of full-length membrane-bound TF associated with blood cells or cellular microparticles. Giesen and colleagues (1999), demonstrated that perfusion of collagen-coated slides with normal human blood resulted in the formation of TF-containing thrombi. As there was no expected source of TF in the slides this indicated that its source must have been the blood. Further analysis of the immunostaining in this study suggested that the TF originated from monocytes and neutrophils. The induction of TF expression in monocytes has been well documented *in vitro* and there is good evidence that this may occur in physiologically relevant conditions also. Regulation of monocyte TF expression is dependent on interactions with cell surface markers such as CD14 and other leucocytes and platelets (Eilertsen & Osterud, 2005). Some, but not all, studies have demonstrated that neutrophils, platelets and possibly lymphocytes may also express TF (Osterud *et al.*, 2000; Mechiche *et al.*, 2005; Maugeri *et al.*, 2006).

Microparticles are cellular fragments of membrane sometimes combined with membrane-associated proteins such as TF. They provide a procoagulant surface as they contain negatively charged phosphatidylserine. It seems likely that microparticles that are relevant for coagulation are derived from monocytes, although other cell types such as ECs and vascular smooth muscle cells have also been implicated (Morel *et al.*, 2004). The proposed role for microparticles in coagulation is that they can provide blood-borne TF in a procoagulant environment to contribute to thrombus formation on the platform of a platelet plug. The association of microparticles with platelets is dependent on P-selectin and its principal receptor P-selectin glycoprotein ligand 1 (Falati *et al.*, 2003). P-selectin has also been implicated in the generation of TF-containing microparticles from monocytes (Celi *et al.*, 1994). The hypothesis explaining the relationship between P-selectin and TF has been recently reviewed by Wagner and colleagues (2005) and

may be summarised as follows. P-selectin is released from α-granules onto the platelet surface following activation. It may then stimulate the formation of procoagulant TF-containing microparticles from monocytes. These microparticles contribute to the propagation of coagulation on activated platelets in a P-selectin dependent manner.

1.9 Tissue factor encryption

It is clear that if blood-borne TF can contribute to thrombus formation then additional regulatory systems must exist, as this bypasses the physical barrier that controls initiation with extravascular TF. This has led to the hypothesis of 'TF encryption' defined by Bach as "the post-translational suppression of TF procoagulant activity on the cell surface" (Bach, 2006).

There are various mechanisms by which encryption might occur. A key feature of the TF·FVIIa complex is that full activity is only seen on an appropriate phospholipid surface. The platelet is the likely source of this surface *in vivo* as it provides a circulating membrane that is localised to the site of clot formation. The phospholipid composition of the platelet membrane plays a critical part in regulating the activity of the macromolecular complexes that form the coagulation network, including TF·FVIIa. In the resting state negatively charged phosphatidylserine is sequestered on the inner surface of the phospholipid bilayer and while in this state, the formation of coagulation complexes is not supported. Following platelet activation the membrane is effectively 'flipped' so that phosphatidylserine is now on the outer surface which provides the necessary platform for coagulation (reviewed in Bach, 1998; and Gomez & McVey, 2006). Membrane flipping occurs in association with Ca^{2+} influx and is an energy-dependent process. The importance of this process is demonstrated by the rare platelet disorder Scott syndrome, which is characterised by a failure of membrane flipping and associated with a mild bleeding diathesis (Albrecht *et al.*, 2005).

The composition of the plasma membrane may affect TF activity in other ways. Normally the fatty acid chains of phospholipids are arranged in an irregular fashion in the membrane. In some situations they are lined up parallel to each other in a more orderly fashion, forming microdomains referred to as lipid rafts (Simons & Ikonen, 1997). Further specializations of these microdomains lead to the formation of caveolae, which are flask-shaped invaginations of the membrane with lipid rafts lining the bottom of the pocket (Harder & Simons, 1997). Recent data indicate that TF

may be localised within these microdomains and that the organisation of the phospholipid membrane prevents its procoagulant activity (Dietzen *et al.*, 2003; Fortin *et al.*, 2005). Following appropriate stimuli, TF and the associated membrane are packaged together and bud off from the cell to form microvesicles or microparticles which may circulate freely and promote coagulation (del Conde *et al.*, 2005).

It has recently been proposed that regulation of the formation of the intra-domain disulphide bonds may be used to mask or unmask TF procoagulant activity. The presence of disulphide bonds that control function rather than act purely as stabilisers has been suggested in a number of proteins (Schmidt *et al.*, 2006). It has long been known that the enzyme complex protein disulphide isomerase (PDI) can catalyse formation and cleavage of these disulphide bonds (Ellgaard & Ruddock, 2005). PDI activity has been detected in association with intra-cellular organelles and attached to the outer cell membrane. In 1991 experiments were reported in which the importance of each disulphide bond in TF was demonstrated by replacing the cysteines with serines (Rehemtulla *et al.*, 1991). Without the disulphide bonds the protein still folded such that it could be detected by conventional antibodies. Loss of the disulphide bond in the F1 domain (Cys49-Cys57) had no significant effect on function while loss of the disulphide in the membrane-proximal F2 domain (Cys186-Cys209) resulted in a marked reduction in procoagulant function. More recent data has been reported showing that TF lacking the Cys186-Cys209 bond is found *in vivo* and that disulphide bonds in TF are regulated by PDI (Ahamed *et al.*, 2006). The reduced TF is able to bind and activate FVII forming a reduced TF·FVIIa complex, but does not bind FX. Surprisingly, the reduced TF·FVIIa is still able to signal via cleavage of PAR2. Thus it appears that reduction of the disulphide bond leads to loss of procoagulant activity but not direct signalling potential. As the reduced TF·FVIIa does not form a complex with FXa it is unlikely that its direct signalling will be subject to inhibition by TFPI.

In summary there is increasing data on blood-borne TF and its regulation, but the physiological role of this form of the protein remains to be conclusively defined. It remains to be seen whether one particular method of TF encryption predominates in normal physiology or if various TF encryption mechanisms co-exist.

1.10 Demonstration of tissue factor expression

It is clear from the above discussion that many physiological and pathological processes are critically dependent on the pattern of TF expression and exposure to blood. Localization of TF protein has thus been an area of extensive investigation. Much of the data on sites of TF expression have been provided by immunohistochemical techniques as previously described. This has greatly enhanced our understanding of the role of TF in various processes, but immunohistochemistry has well documented limitations. Clearly the specificity and affinity of the primary antibody are of critical importance and in these regards the commercially available anti-TF antibodies have often been unsatisfactory. As our understanding of the coagulation network increases it is becoming apparent that this is a rapidly changing, dynamic process. This has led to the development of new tools for investigation of coagulation such as real-time monitoring of thrombus formation (Falati *et al.*, 2002). Since immunological detection methods have been used in these systems there is a concern that antibodies used for detection may also interfere with protein function. This is of particular concern with TF detection, as its procoagulant activity is dependent on the extracellular domains which generally provide the epitopes for antibody recognition. Furthermore, optimal resolution in immunological methods requires fixation of tissues, which is not feasible in a live detection system. The Haemostasis and Thrombosis group at the MRC has been seeking improvements in TF detection for many years. Various strategies have been used to develop new antibodies, such as the use of DNA vaccines, to optimise immunological detection (Mumford *et al.*, 2005). However, the limitations of these strategies have led the group to investigate TF detection using a reporter model.

1.10.1 Green fluorescent protein

Reporter genes have long been used in molecular biology as methods of demonstrating gene expression. In principle the reporter is inserted into the host by genetic engineering techniques so that its expression reflects expression of the gene of interest. In the past commonly used reporter genes included those coding for the bacterial enzyme β-galactosidase (*lacZ*) and members of the luciferase family. The actions of these enzymes produce either coloured products or generate light. These reporter proteins have been widely used in gene expression studies and continue to have a valuable role. *LacZ* in particular provides a strong signal that is readily visualized in histological sections without the need for immunological detection

methods. However, the gene is expressed under prokaryotic control and the most commonly used substrate, X-gal, is toxic to eukaryotic cells. For these reasons *LacZ* is unsuitable for analysis of mammalian gene expression under physiological conditions or in live cells. For this purpose the naturally occurring fluorescent proteins have become the preferred choice.

Green fluorescent protein (GFP) was first described in the 1960s (Shimomura *et al.*, 1962). This protein was isolated from the jellyfish *Aequorea victoria* and its key property is the emission of a bright light with peak wavelength of 509 nm following excitation by blue light. The protein and fluorophore are stable in mammalian cells making it an ideal candidate for mammalian gene expression analysis. Since the 1970s there has been intensive study of the characteristics of this protein and knowledge of the tertiary structure has enabled the production of mutant variants with increased usability. The main part of the protein consists of a number of β-pleated sheets arranged parallel to each other to form a barrel shape. The fluorophore is comprised of residues Ser65, Tyr66 and Gly67 which are directed to the core of the molecule as shown in Figure 1.2 (Elsliger *et al.*, 1999). A variant of the protein in common use as a reporter gene at the present time is enhanced green fluorescent protein (EGFP), which contains the mutations Phe64Leu and Ser65Thr at the site of the fluorophore. This leads to improved folding at 37 °C and increased ionization of the fluorophore (Tsien, 1998). The successful use of EGFP in reporter assays has promoted the production of colour variants with emission in the cyan and yellow wavelengths. With the introduction of the red fluorescent protein from *Discosoma sp.* (DsRed) we now have a variety of reporters available for studying expression of several genes simultaneously. These reporter genes offer a solution to the limitations imposed by immunological detection methods in the coagulation network.

1.10.2 Previous studies with Tissue Factor–GFP reporter models

In order to investigate the feasibility of using GFP as a reporter gene for mTF, a TF-EGFP reporter model under control of a modified mammalian promoter has been previously created and studied in this group (Mumford, 2003). A similar model has been described in which EGFP was used as a reporter for rabbit TF (Fortin *et al.*, 2005). The model previously used in the Haemostasis and Thrombosis group will be briefly described.

Expression of a mTF-EGFP fusion protein was first analysed in Chinese hamster ovary (CHO) cells. The mammalian expression vector pcDNA3.1 (Invitrogen)

was modified to express an mTF-EGFP fusion protein. First, the mTF cDNA without the C-terminal stop codon was inserted into the multiple-cloning site of pcDNA3.1. The EGFP cDNA was then cloned into the vector following and in-frame with the mTF coding sequence. Control vectors contained either EGFP cDNA or mTF cDNA inserted separately into pcDNA3.1.

CHO cells transfected with the vector expressing EGFP alone had a pan-cytoplasmic pattern of fluorescence, while those expressing the mTF-EGFP fusion protein showed fluorescence localized to the plasma membrane and intra-cytoplasmic vesicles (Figure 1.6). TF procoagulant activity was found to be the same in CHO cells expressing mTF-EGFP and those expressing mTF alone. No TF procoagulant activity was demonstrable in CHO cells transfected with the control empty vector. Growth of CHO cells was analysed by a ^{3}H-thymidine incorporation assay and found to be equivalent in CHO cells expressing mTF, mTF-EGFP, EGFP, and in CHO cells transfected with unmodified pcDNA3.1 and in untransfected CHO cells.

These results indicate that an mTF-EGFP fusion protein shows a similar expression pattern to that expected of TF from previous immunohistochemical studies. Tagging of the mTF molecule with EGFP impairs neither procoagulant function nor cell viability.

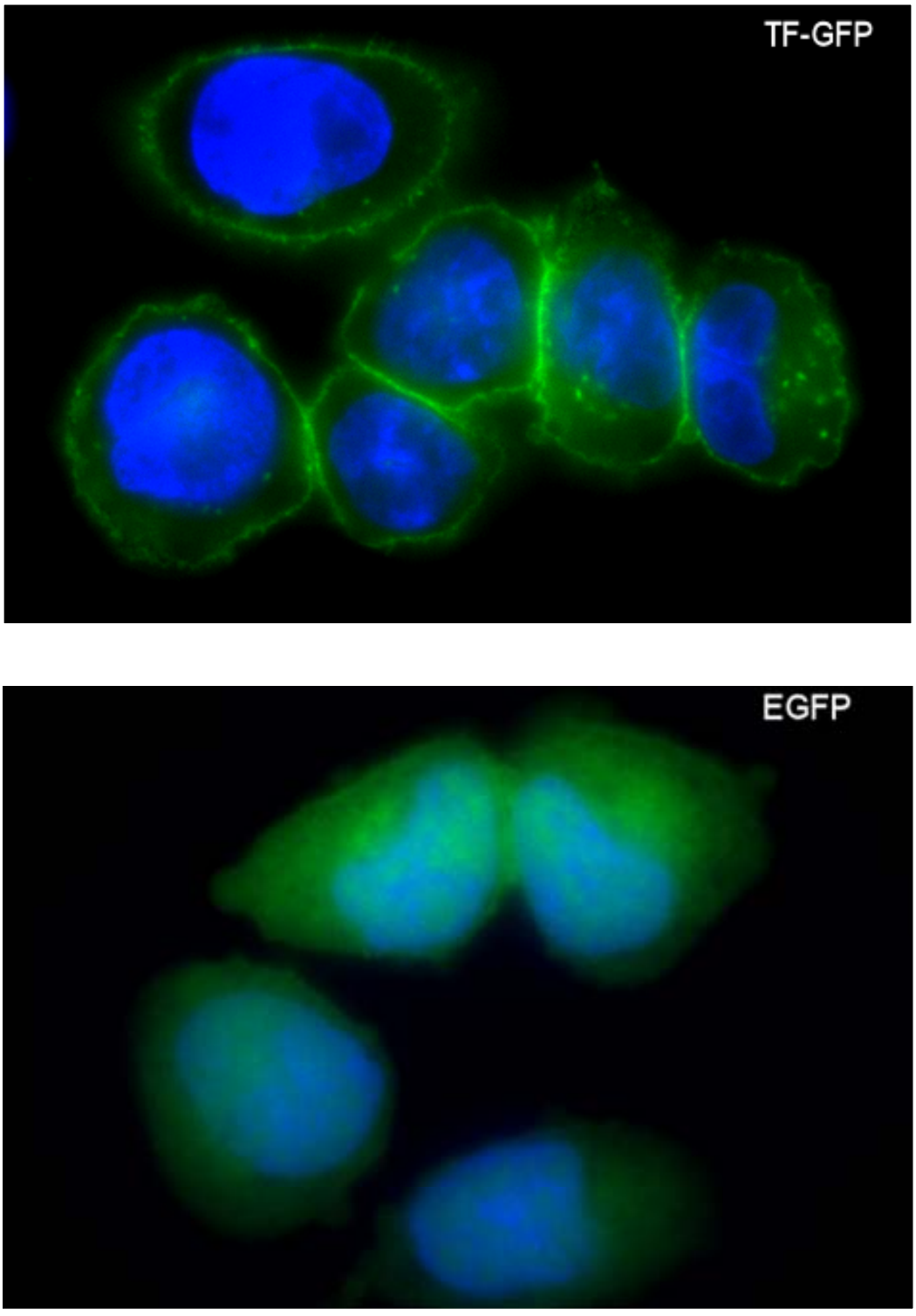


Figure 1.6 Fluorescence micrographs of CHO cells transfected with vectors expressing a TF-GFP fusion protein (top) or EGFP alone (bottom). Slides are counterstained with DAPI. With the TF-GFP vector fluorescence is seen localized to the plasma membrane and in discrete regions in the cytoplasm compared with the pan-cytoplasmic distribution seen with the control EGFP vector. Images courtesy of Dr. A. Mumford.

1.11 Aims of project

The above data establish the suitability of EGFP as a reporter gene for TF. Although the model previously used in this lab and the rabbit model described by Fortin and colleagues are useful tools, they are not designed to study TF expression under endogenous regulation, as they are randomly integrated and utilise a viral promoter to drive transcription. To demonstrate TF expression under endogenous regulation, the EGFP coding sequence must be inserted into one of the endogenous TF alleles.

The purpose of this project was to investigate whether a mTF-EGFP gene generated by homologous recombination could act as a reporter of mTF expression and thus further our understanding of the role of TF in normal physiology and pathological states. Therefore the specific aims of the project are:

1. To target the mF3 locus by homologous recombination such that the cDNA for EGFP is 'knocked-in' at the 3' end. The targeting will be carried out in mouse ES cells and in bacteria carrying a BAC containing the mF3 gene and genomic flanking regions.
2. To inject targeted ES cells into blastocysts in order to generate chimaeric mice expressing the mTF-EGFP gene. These mice will be bred to generate offspring homozygous for the reporter model.
3. To purify targeted BAC from bacterial DNA and use it in pro-nuclear injection of fertilized mouse embryos. This will generate transgenic mice expressing the reporter model alongside the endogenous TF alleles.
4. To demonstrate appropriate expression of the reporter gene in mice and compare the expression pattern of the reporter model with immunohistochemical localization of TF.
5. To use these transgenic mice to study mTF expression in embryonic development and in murine models of human disease.

Chapter 2 - Materials and Methods

Procedures referred to herein as 'standard methods' were carried out according to protocols detailed in Current Protocols in Molecular Biology (edited by Ausubel *et al.*, 2006), available online at http://www.mrw.interscience.wiley.com/cp (DOI:10.1002/0471142727). Unless otherwise stated, all laboratory reagents were analytical grade and manufactured 'in house' (listed in appendix 1) or obtained from commercial suppliers (listed in appendix 2). Commercial kits and equipment were used according to the manufacturers' instructions unless otherwise indicated.

Genomic and cDNA nucleotide sequences are numbered with +1 referring to the A of the ATG translation start codon. In peptide sequences +1 indicates the first residue at the N-terminus of the mature peptide. Reference sequences for TF are those derived from the C57BL/6J strain as listed at the Mouse Genome Informatics (MGI) website at http://www.informatics.jax.org/, MGI ID: 88381. Reference sequence accession numbers for *mF3*: AC129311 (genomic) and AK137642 (cDNA). The reference TF peptide sequence is the predicted translation product of AK137642. The reference sequence used for EGFP cDNA is accession number U76561.

Details of oligonucleotide primer sequences and polymerase chain reaction (PCR) conditions are given in appendix 3. All animal work was carried out under licences granted by the Home Office in accordance with the Animals (Scientific Procedures) Act 1986 and associated codes of practice as detailed on the Home Office website at http://scienceandresearch.homeoffice.gov.uk/animal-research/.

2.1 Preparation and analysis of vectors

2.1.1 Preparation of plasmid DNA

Bacteria transformed with the required plasmid were grown in LB broth containing the appropriate antibiotic at 30-37 °C overnight in a shaker incubator set at 300 rpm. Colonies were initially inoculated into 5 mL culture overnight and then the 5 mL culture was used to inoculate 500 mL LB broth. To make aliquots of bacterial cultures for freezing, 900 µL of culture was mixed with 100 µL 10X Hogness freezing medium and allowed to equilibrate at room temperature for a few hours before storage at -70 °C. For selection antibiotics were used at the following concentrations: ampicillin 50 mg/L, carbenicillin 50 mg/L, spectinomycin 25 mg/L

and kanamycin 25 mg/L. Small scale plasmid DNA preparations (mini-prep) from 1 mL of culture were carried out by phenol chloroform extraction and ethanol purification according to standard methods or kits supplied by Qiagen. Large scale plasmid DNA preparations (maxi-prep) from 500 mL of culture were carried out using a CsCl gradient according to standard methods or kits supplied by Qiagen. Quantification of DNA was carried out by labelling DNA with Picogreen (Invitrogen) and measuring emission at 520 nm in a F-2500 fluorescence spectrophotometer (Hitachi) or a Spectramax Gemini EM fluorescence plate reader (Molecular Devices).

2.1.2 Enzymatic manipulation and cloning of DNA

Purified plasmid DNA and PCR amplicons were digested with restriction endonucleases (Promega or NEB) in buffers supplied by the manufacturers. The restriction fragment required for insertion was purified after separation by agarose gel electrophoresis. The plasmid DNA required to receive the insert was digested to produce compatible ends and dephosphorylated with calf intestinal alkaline phosphatase (Promega). After dephosphorylation plasmids were incubated at 70 °C for 10 minutes to inactivate the phosphatase prior to ligation.

Dephosphorylated plasmid and restriction digested insert were incubated in a 1:3 molar ratio with 1 U of T4 DNA ligase (Promega) in 20 µL ligase buffer containing 1 mmol/L ATP. The mixture was incubated at room temperature for 4 hours or at 16 °C overnight. 1 µL of the ligation mixture was transformed into 100 µL MAX Efficiency DH5αF′ IQ or GeneHogs DH10B-derived competent cells (Invitrogen) using a standard heat-shock method. 250 µL SOB medium was added and the bacteria allowed to recover in a shaking incubator at 37 °C for 1 hour before 50-150 µL of culture were spread on LB agar plates containing antibiotics for selection. Plates were incubated at 37 °C overnight or for 48-72 hours at room temperature. Antibiotic resistant colonies were picked with sterile pipette tips and either streaked onto LB agar plates for PCR screening or placed directly into 5 mL of LB broth containing the appropriate antibiotic. For screening to assess correct ligation of the insert by PCR, the pipette tip used to streak the colony was agitated in 50 µL PCR reaction mix.

Undigested PCR amplicons were cloned by DNA topoisomerase catalysed TA ligation into pCR2.1-TOPO vector (Invitrogen) using the manufacturer's TOPO TA Cloning Kit.

2.1.3 Analysis of DNA fragments

DNA was mixed with 10X loading buffer for electrophoresis on agarose gel containing 0.7-2% agarose (FMC Bioproducts) and 0.2 mg/L ethidium bromide (Sigma-Aldrich) in TBE. The gel was visualized under ultra-violet light in a Gel Doc trans-illuminator and images were captured using Quantity One software (Bio-Rad). In this document, images demonstrating electrophoretic separation of nucleic acids are shown with molecular weight markers given in kb, unless otherwise indicated.

2.1.4 Purification of DNA fragments

For isolation of restriction fragments or PCR amplicons, products were separated by agarose gel electrophoresis. The gel was viewed under long wavelength (to reduce DNA nicking) ultra-violet light using a trans-illuminator (BDH), and the required band excised with a scalpel. The product was extracted from the excised gel block using the QIAquick gel extraction kit (Qiagen).

2.1.5 Identification of transformed bacterial clones by PCR

In experiments that required screening of less than a few hundred clones, a PCR based strategy was used. Individual clones were picked by sterile pipette tip and streaked on agar plates labelled with a grid reference. The same pipette tips were then immediately placed in pre-prepared PCR reaction tubes. Positive clones were then cultured from the streaked agar plates.

2.1.6 Identification of transformed bacterial clones by colony lifts

In experiments that required screening of more than a few hundred clones, an *in situ* colony lift procedure followed by hybridisation with radio-labelled probes was used. Hybond-N+ nucleic acid transfer membranes (GE Healthcare) were placed on the agar plates containing the colonies and marked to ensure correct orientation. The majority of the bacteria in each colony immediately stick to the membrane but sufficient organisms are left behind to reconstitute the colony after overnight incubation. The nylon membranes with attached colonies were soaked in 10% sodium dodecyl sulphate (SDS) for 5 minutes followed by 10 minutes in Southern denaturation solution, 15 minutes in Southern neutralisation solution and finally 5 minutes in 20X SSC. The released DNA was cross-linked to the membrane by exposure to 1200 µJ ultra-violet light before hybridisation as described below. Clones containing the sequence of interest appear as dots on photographic film. The film is

then aligned with the original agar plate using the orientation marks to allow identification and expansion of the positive clones.

2.1.7 Polymerase chain reaction

Oligonucleotide primers for PCR were designed using Oligo (Molecular Biology Insights Inc.) and were obtained from Sigma-Genosys. Details of primer sequences and PCR conditions are given in appendix 3. The binding sites of primers are shown in appendices 4 and 5.

PCR was performed using Red Hot DNA polymerase (Abgene) for amplicons of 1000 base pairs or less. For generation of longer amplicons or where increased fidelity was required, Platinum *Taq* DNA Polymerase High Fidelity (Invitrogen) or Easy-A High Fidelity PCR Cloning Enzyme (Stratagene) was used. Primers were used at a final concentration of 1 µmol/L. The reactions were performed on a DNA Engine Tetrad 2 Peltier thermal cycler (Bio-Rad). All PCR products were verified by agarose gel electrophoresis.

2.1.8 DNA sequencing

PCR products (approximately 50 ng) or purified plasmid DNA (approximately 500 ng) was used as the template in DNA sequencing reactions. Amplicons were filtered using YM-100 Microcon columns (Millipore) to remove residual PCR primers. Oligonucleotide primers were diluted to 30 nmol/L. Sequencing was carried out by the Genetics Core Facility, MRC Clinical Sciences Centre using a 3730 xl capillary DNA analyser and accompanying software (Applied Biosystems). Chromatograms were interpreted and sequences assembled into databases using the Staden package available as open source from http://staden.sourceforge.net/ (Bonfield *et al.*, 1995).

2.2 Culture and targeting of embryonic stem cells

2.2.1 Culture of embryonic stem cells

Three murine ES cell lines were used in this project: 129/1 (kindly provided by Dr. Graham Kay, Northwick Park Hospital, London), E14 (kindly provided by Prof. Martin Hooper, Univ. Edinburgh) and EL M3 (Open Biosystems). All three lines were derived from the 129/Sv mouse strain and had been previously used in successful gene targeting studies (Hooper *et al.*, 1987; Kay *et al.*, 1993; Brambrink *et al.*, 2006).

Two murine fibroblast cell lines were used: STO (ECACC #86032003) and SNL, which is a STO line stably transfected with a gene conferring resistance to G418,

kindly provided by Prof. Elizabeth Robertson, Harvard, USA. For use as a feeder layer for ES cells, fibroblasts were prevented from further mitotic division (inactivation) by incubation in media containing 10 mg/L mitomycin C (Invitrogen) for 3 hours at 37 °C.

Details of the culture conditions for different ES cell lines are given in Table 2.1. Tissue culture flasks and plates (Nunc, VWR) were coated with 0.1% gelatin in PBS (Invitrogen) prior to use. If required, inactivated SNLs were plated at a density of 2 x 10^5 per cm^2. ES cells were plated at a density of 1.5-4 x 10^5 per cm^2. Cells were cultured in Dulbecco's modified Eagle's medium (D-MEM, Invitrogen) containing 15% foetal calf serum (FCS, Autogen Bioclear) 2 mmol/L L-glutamine, 1 mmol/L sodium pyruvate, 1% non-essential amino acids, 0.1 mmol/L β-mercaptoethanol (all from Invitrogen) and 1000 U/mL leukaemia inhibitory factor (ESGRO, Chemicon International). In order to prevent differentiation media was changed at least every other day and cells were split before reaching 80% confluency. For storage in liquid nitrogen ES cells were suspended in freezing medium containing 65% culture medium, 25% FCS and 10% dimethyl sulphoxide (Sigma) at a concentration of 10^6 cells per mL.

ES cell line	Fibroblast layer required	Passage at transfection	% Normal karyotype at transfection	G418 concentration
129/1	STO or SNL	16	75-85%	200 mg/L
E14	None	23	75-85%	100 mg/L
EL M3	STO or SNL	13	80-90%	200 mg/L

Table 2.1 Details of ES cell lines used for targeting**.**

For detaching cells in culture, flasks were first washed twice with PBS and then incubated for 2 minutes with 0.25% trypsin, 0.04% EDTA in Hanks BSS (Invitrogen). Following detachment the trypsin was neutralised with culture medium.

ES cells were frequently checked for mycoplasma contamination using MycoAlert (Cambrex).

2.2.2 Karyotyping

Cell lines with karyotype instability have reduced germline potential and therefore the karyotypes of ES cells were established prior to targeting and use in blastocyst injections. ES cells were harvested for karyotyping the day after plating.

Three hours prior to harvesting ethidium bromide 1.5 mg/L and Colcemid (Invitrogen) 50 µg/L were added to the culture medium to arrest the cells in metaphase and extend the chromosomes. Cells were harvested with trypsin and after gentle resuspension in 75 mmol/L KCl, incubated for 10 minutes at 37 °C. The cells were pelleted and the supernatant removed. Fresh ES fixative was slowly added to the cell pellet and cells were gently resuspended by flicking the tube. Cells were stored in ES fixative at 4 °C for at least 24 hours prior to further analysis.

For the preparation of metaphase spreads, stored cells were spun down and resuspended in fresh ES fixative. The cell suspension was picked up with a Pasteur pipette and single drops were applied to slides from a height of 1-2 m. Slides were allowed to air-dry and stained with DAPI. Metaphase spreads were examined in an inverted laser scanning confocal microscope system (Leica TCS SP1 DM IRB). Photographs were taken of 20-30 clearly defined single cell chromosome groups for each analysis (Figure 2.1). Chromosomes were counted and the karyotype expressed as the percentage of cells with a normal chromosome complement (Table 2.1).

Figure 2.1 Representative metaphase spreads from ES clones taken prior to blastocyst injection. The spreads show the normal chromosome complement of 20 pairs with no gross structural abnormalities.

2.2.3 Electroporation of embryonic stem cells

ES cells were cultured through two passages prior to transfection. The targeting vector was maxi-prepped, linearized with *Dra*III (short construct, pES-TC-S) or *Sfi*I (long construct, pES-TC-L), extracted with phenol chloroform and ethanol precipitation and resuspended at 500 mg/L. ES cells were detached with trypsin and resuspended in PBS at 10^7 cells per mL. 8 x 10^6 ES cells and 25 µg linearized vector were mixed in an electroporation cuvette and incubated at room temperature for 5 minutes. The cuvette was placed in an electroporator (GenePulser II, BioRad) set to capacitance 500 µF and voltage 240 V and an electrical pulse applied. This produced a time constant of 7 ± 0.4 ms. Following electroporation the ES cells were allowed to recover on ice for 10 minutes. The contents of a single cuvette were then plated onto two 60 mm tissue culture petri dishes.

Selection agents were applied the day after electroporation. The effective concentration of the positive selection agent G418 (Invitrogen) was empirically determined for each cell line and is given in Table 2.1. This was the minimum concentration that resulted in complete cell loss at six days in stock ES cells. The negative selection agent ganciclovir (Cymevene, Roche) was used at 2 µmol/L. Media containing selection agents was changed daily. Ganciclovir was omitted from the culture medium after 8 days. Positive selection with G418 was maintained until clones were ready for picking.

2.2.4 Selection and expansion of transfected clones

Colonies representing clonal growth from single ES cells were picked 10-11 days after electroporation. The culture media in the petri dish was replaced with PBS. Each colony was aspirated in a small amount of PBS with a 20 µL pipette by applying the pipette tip to the centre of the colony. Colonies were disaggregated in 96-well plates by applying trypsin for 1-2 minutes, followed by pipetting, before transfer to fresh 96-well plates pre-plated with gelatin ± SNL fibroblasts. Clones were grown in media free of selection agents for 2-3 days at which point they were 70-80% confluent. At this point they were split into two preprepared 96-well plates. For 129/1 and EL M3 clones one plate was preplated as usual with gelatin and fibroblasts while the other was preplated with gelatin only. E14 plates always contained gelatin only. After 2-3 days gelatin and fibroblast pre-treated plate was frozen at -70 °C while the other plate was cultured for 5-6 days to allow the colonies to become over confluent.

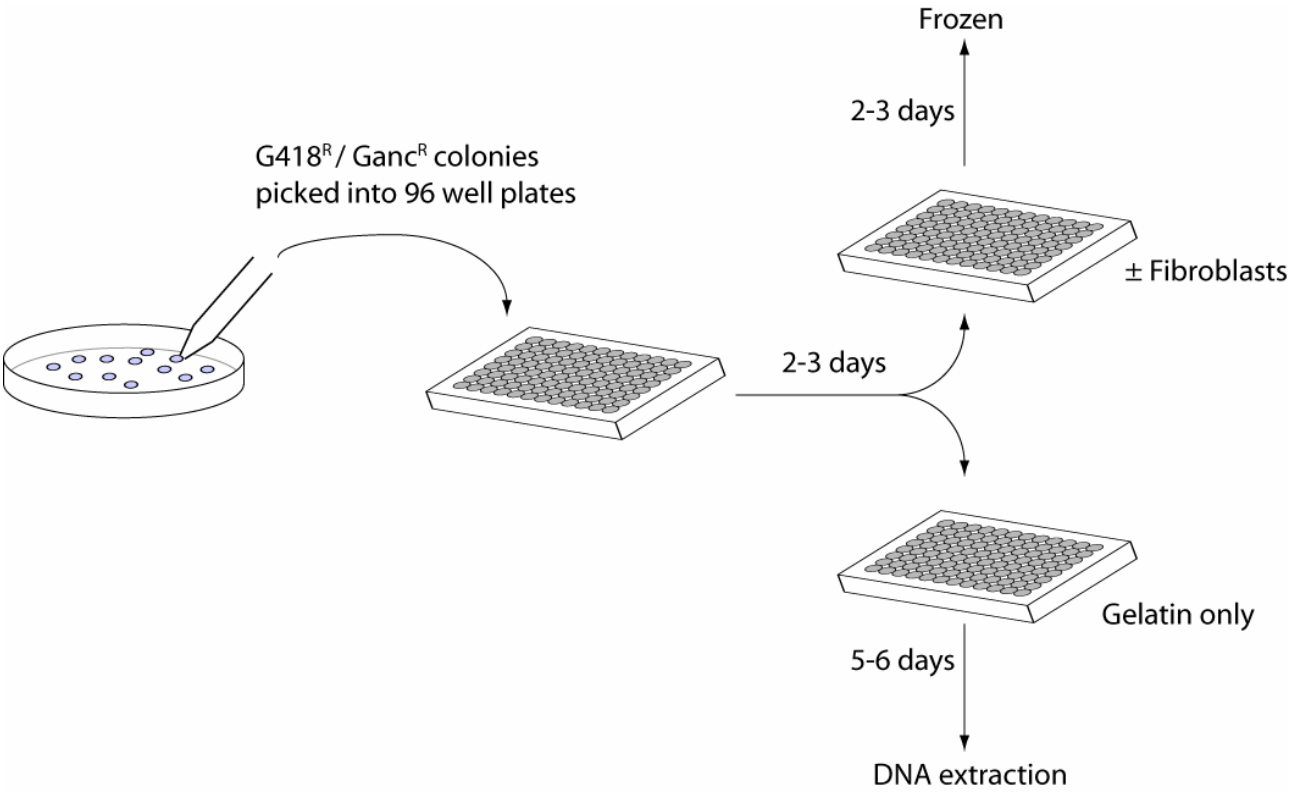


Figure 2.2 Procedure for the expansion of G418 and ganciclovir resistant colonies. Duplicate plates are generated of which one is frozen and the other used for DNA extraction and screening. Each transfer is counted as one passage.

This plate was subsequently used to for DNA extraction and screening. The management of picked colonies is shown in Figure 2.2.

2.2.5 DNA extraction and PCR screening of ES clones

Clones were allowed to become fully confluent in 96-well plates, resulting in on average 10^5 ES cells per well. After washing with PBS, 50 µL of ES lysis buffer with proteinase K (1 g/L) added on the day of use was applied to each well. Plates were incubated overnight in a humid atmosphere at 55 °C.

The following day DNA was precipitated *in situ* by the addition of absolute ethanol and left at room temperature for 2 hours. Once strands of precipitated DNA were clearly visible the solvent was removed by inverting the plate. DNA strands were washed three times with 70% ethanol and left to dry at room temperature for 15-30 minutes. DNA was redissolved in 33 µL water and 3 µL were removed for screening by PCR. This was added to 50 µL of PCR mix containing primers NP2 and MTFP2 with Platinum *Taq* DNA polymerase.

The remaining 30 µL were prepared for screening by Southern blot analysis with labelled probes. The DNA was mixed with 10 µL restriction digest mix containing 1X restriction enzyme buffer, 1 mmol/L spermidine, 100 mg/L bovine serum albumin, 50 mg/L RNase A and 50 U *Bam*HI. The plate was then incubated overnight at 37 °C. 30 µL of digested DNA was mixed with 10X loading buffer and electrophoresed for 6-18 hours in 0.8% agarose gel. Restriction fragments were transferred onto nylon membrane by Southern blotting.

2.2.6 Southern blot

Southern blot was carried out using a standard method. DNA fragments were separated by agarose gel electrophoresis and nicks induced by exposure to 1200 µJ ultra-violet light. The gel was then washed in Southern denaturation solution (20 minutes X2) followed by Southern neutralisation solution (20 minutes X2) before overnight transfer onto Hybond-N+ nucleic acid transfer membranes (GE Healthcare). DNA fragments were subsequently detected by radio-labelled or digoxigenin-labelled probes.

2.2.7 Generation of probes for DNA hybridisation

A 777 bp probe was prepared by PCR amplification from 129/1 genomic DNA using primers PbES2F and PbES2R. The amplicon was cloned into pCR2.1-TOPO vector. For labelling, probes were removed by digestion with *Eco*RI and gel purified.

2.2.8 Screening with radio-labelled probes

Radio-labelled probes were prepared by PCR from gel purified restriction fragments using Megaprime DNA labelling system (GE Healthcare) with 1.85 MBq of [α-^{32}P] dCTP, specific activity 3000 Ci/mmol, per 50 µL labelling reaction. Following removal of unincorporated [α-^{32}P] dCTP by filtration through a Sephadex NICK column (GE Healthcare), labelling efficiency was measured by counting dpm from 1 µL of labelled probe in an Easicount 4000 (Scotlab).

Following Southern transfer, nylon membranes were placed in ULTRAhyb hybridization buffer (Ambion) containing radio-labelled probe at a minimum final concentration of 10^6 dpm per mL. Hybridisation tubes were placed in a rotor in an incubator at 42 °C overnight. The following day membranes were washed twice for 5 minutes in 2X SSC, 0.1% SDS at 42 °C followed by two 15-minute washes in 0.1X SSC, 0.1% SDS at 42 °C. Radioactivity was detected by exposure of membranes to Kodak

BioMax MR film (Sigma-Aldrich) or detection in a Storm Phosphorimager system (GE Healthcare).

2.2.9 Screening with digoxigenin-labelled probes

Digoxigenin-dUTP was incorporated into probes using the DIG High Prime DNA Labelling and Detection Starter Kit II (Roche). The hybridization procedure was the same as for radio-labelled probes. Probes were bound by anti-digoxigenin antibody conjugated with alkaline phosphatase (Roche). The chemiluminescent enzymatic dephosphorylation of CSPD by alkaline phosphatase was detected by exposure to Kodak BioMax MR film (Sigma-Aldrich).

2.3 Targeting of the bacterial artificial chromosome

The protocols in section 2.3 were developed in the laboratory of Dr. N. Copeland, National Cancer Institute, Maryland, USA and are detailed on the faculty website at http://recombineering.ncifcrf.gov/. This work was carried out in collaboration with A. Carpenter, Haemostasis and Thrombosis, MRC Clinical Sciences Centre.

A major concern with procedures involving BACs is that very large fragments of DNA are fragile and easily sheared. Therefore pipetting and mixing of DNA was kept to minimum. For transfer of BAC in small volumes 1 mL pipette tips were cut with sterile blades to reduce shear forces caused by pipetting across a small bore.

2.3.1 Preparation of DNA from BAC hosts

DH10B bacteria transformed with the BAC of interest were obtained from the BACPAC Resources Centre, Children's Hospital, Oakland, USA (http://bacpac.chori.org/home.htm). Bacteria were cultured at 32 °C on LB agar or broth containing chloramphenicol 12.5 mg/L. BAC DNA was prepared using the reagents present in the Qiagen Large Construct kit according to a previously published method (Liu *et al.*, 2003).

2.3.2 Pulse field gel electrophoresis

Conventional agarose gel electrophoresis is unsuitable for separation of large fragments of DNA. Size analysis of BAC fragments was carried out using the Chef-DR II Pulsed Field Electrophoresis system (Bio-Rad). 1% agarose gels were prepared using ultra pure DNA grade agarose (Bio-Rad). Gels were run in 0.5X TBE and

maintained at a constant 16 °C. The voltage was set to 6 V/cm and pulse lengths were 5-15 s. The run time was 18 hours. Chef DNA λ ladder (Bio-Rad) was used as a molecular marker.

2.3.3 Homologous recombination in bacteria

For manipulation by homologous recombination BACs were first transformed into electrocompetent EL350 cells (a kind gift of Dr. N. Copeland). These are modified DH10B that contain the *red* genes of bacteriophage λ under temperature control and the *cre* gene under control of an arabinose-inducible promoter (Lee *et al.*, 2001). The *red* genes are responsible for controlling homologous recombination while *cre* codes for the recombinase that catalyses recombination at loxP sites.

BAC-containing EL350 were grown overnight in 5 mL LB broth. The following day when the growth curve had reached a plateau, bacteria were placed in a shaking water bath at 42 °C for 15 minutes to induce expression of the *red* genes. The container was then immediately placed in ice for 15 minutes. Bacteria were then centrifuged at 2,500 g at 0 °C for 6 minutes and the pellet resuspended in ice-cold water. This step was repeated twice and renders the bacteria electrocompetent. The pellet was resuspended in 50 µL ice-cold water and placed in an electroporation cuvette with 100 ng of a 7.5 kb *Bgl*II – *Nhe*I fragment from pBAC-TC. Electroporation was carried out using an *E. coli* pulser (Bio-Rad) set to 1.75 kV, 25 µF with the pulse controller set to 200 Ω. Typically this produced a time constant of 4 ms. 0.5 mL of LB broth was added and cuvettes were incubated at 32 °C for 1 hour. 50-200 µL of the electroporated bacteria were then plated onto LB agar containing spectinomycin 25 mg/L and placed in an incubator overnight at 32 °C.

BAC-containing host colonies were picked into 5 mL LB broth and the pipette tip used to inoculate a PCR mix containing the primers MTFI5F and EGFP2R. Correctly targeted clones were grown overnight and either stored briefly at 4 °C or entered directly into the next step. 1 mL of culture was inoculated into 10 mL fresh LB broth and incubated for 2 hours. Then L (+) arabinose (Sigma-Aldrich) was added to a final concentration of 0.1% and the mixture cultured for a further 1 hour to induce *cre* expression. The culture was then plated onto LB agar containing chloramphenicol 12.5 mg/L and incubated overnight. Colonies positively selected by chloramphenicol resistance were streaked on LB plates containing spectinomycin 25 mg/L. Reacquisition of sensitivity to spectinomycin indicated successful 'floxing' of the neomycin resistance cassette. This was confirmed by PCR screening using the

primers GFP3F and 3HAR, which generates a 547 bp amplicon in the floxed gene compared with a 1747 bp amplicon before floxing.

2.4 Generation of transgenic mice

2.4.1 Blastocyst injections with ES cells

This procedure was carried out by J. Godwin, Transgenic and ES cell laboratory, MRC Clinical Sciences Centre and will be briefly described here. Day 2 blastocysts were removed from C57BL/6J females mated with C57BL/6J males. 10-20 ES cells were introduced using a micro-injector and blastocysts were re-implanted into pseudopregnant CBA mice. 8-9 blastocysts were returned per surrogate female.

2.4.2 Pro-nuclear injections with BAC DNA

This procedure was carried out by Z. Webster and J. Sealby, Transgenic and ES cell laboratory, MRC Clinical Sciences Centre and will be briefly described here. Mice were a hybrid strain produced by crossing CBA with C57BL/6J. F1 superovulated hybrid females were mated with F1 hybrid males. F2 fertilized embryos were harvested from the oviducts of newly plugged females. BAC DNA at 0.5-2 mg/L was injected using a micro-injector into one or both pro-nuclei. The male pronucleus was preferentially injected as it is usually the larger of the two. Following penetration of the pronucleus sufficient volume was injected to produce visible swelling without disruption of the membrane. This equated to a volume of 1-2 pL per injection containing 0.5-4 fg or 2-20 molecules of BAC DNA. At surgery 15 embryos were implanted into one uterine horn in each pseudopregnant CBA foster mothers.

2.5 Analysis of transgenic mice

2.5.1 Genotyping of transgenic mice

Tail clips were performed on 3-week old mice. DNA was extracted from 5 mm of tail using the Wizard Genomic DNA purification kit (Promega). The presence of the transgene was detected by PCR using primer pair GFP3F and 3HAR. The control PCR to verify adequacy of the genomic DNA template used primer pair MTFex3F and MTFex3R.

2.5.2 Preparation of RNA and cDNA from tissues and monocytes

RNA was prepared from 10^6 cells in culture, 40 - 60 mg of tissue or 10^5 monocytes using the RNAqueous-4PCR Kit (Ambion). Tissues were mostly processed fresh but on occasion were snap frozen in liquid nitrogen and equilibrated in RNAlater-ICE (Ambion) for processing at a later date. For RNA extraction from tissues, samples were first homogenized in lysis buffer with an Ultra-Turrax T8 mechanical disruptor (IKA-Werke). Quantification of RNA was carried out with Ribogreen (Invitrogen). Synthesis of cDNA was carried out with 0.1-0.5 µg total RNA as template. Reverse transcriptase (RT) was primed with random hexamers or oligo dT using the First Strand cDNA Synthesis Kit for RT-PCR (Roche Applied Science). Specific primers were then used to generate the required amplicon from the cDNA by PCR.

2.5.3 Mouse blood sampling

Mice were given a terminal dose of 40 mg pentobarbitone by intra-peritoneal injection and the thoracic cavity opened. Blood was drawn by cardiac puncture with a 2 ml syringe attached to a 23G needle. Syringes were pre-filled with 10 µL 15% K_3EDTA. By this method the volume of blood obtained was 20-30 µL per g body weight (generally about 800 µL from a 30 g adult male). At least 2 mL whole blood was used for each experiment, achieved by pooling the blood from 3-4 animals. Blood samples were kept on ice prior to use.

2.5.4 Induction of tissue factor activity in whole blood

Lipopolysaccharide (LPS from *E.coli* strain 0111:B4) and phorbol 12-myristate 13-acetate (PMA) were obtained from Sigma. The optimal dose and sampling times were empirically determined using previous reports as guidance (Crossman *et al.*, 1990; Osterud *et al.*, 2000; Molor-Erdene *et al.*, 2005). 1 mL EDTA anticoagulated whole blood was placed in a 1.5 mL polypropylene microtube with 100 µg/L LPS and 30 µg/L PMA. Samples were incubated at 37 °C in a Thermomixer comfort (Eppendorf) set to shake at 100 rpm for 1 in every 4 minutes. Control samples were kept at 4 °C for the same time period. Samples were taken after 2 hours of stimulation for RNA quantitation and after 6 hours of stimulation for measurement of TF procoagulant activity and flow cytometry.

2.5.5 Purification of a monocyte-rich fraction from mouse blood

For analysis of induction of TF expression and protein by LPS and PMA, a monocyte-rich fraction was purified from whole blood by filtration through a density gradient. The appropriate density gradient was derived empirically. Hypertonic D-MEM, with a saline concentration of 189 mmol/L, was prepared by adding 170 µL of 2 mol/L NaCl to 10 mL D-MEM (Invitrogen). This mild increase in osmolarity causes swelling of mouse monocytes which aids their separation from lymphocytes (personal communication from Dr. J. Graham, JG Research Consultancy, Liverpool, UK). A working solution (ρ = 1.22 g/mL) of 40% iodixanol was prepared by adding 2 volumes iodixanol 60% stock (Optiprep, Axis-Shield) to 1 volume hypertonic D-MEM. The density gradient (14.6% iodixanol, ρ = 1.08 g/mL) was prepared by adding 1 volume iodixanol 60% stock to 1.82 volumes hypertonic D-MEM.

1 mL whole blood was gently mixed with 750 µL iodixanol working solution. 5 mL iodixanol density gradient was layered on top, taking care not to disturb the blood. 0.5 mL hypertonic D-MEM was layered on top of the density gradient as shown in Figure 2.3A. The sample was then centrifuged at 700 g for 30 minutes at 4 °C.

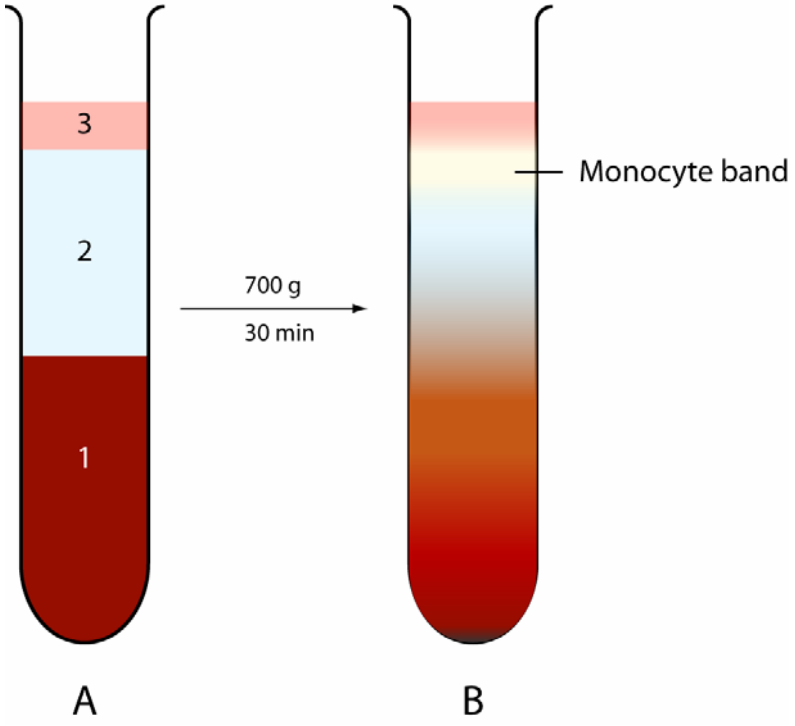


Figure 2.3 Separation of a monocyte-rich fraction from whole mouse blood using a density gradient. A) tube before centrifugation containing 1 – whole blood, 2 – density gradient and 3 – D-MEM. B) after spinning the monocyte-rich band is visible between the D-MEM and density gradient.

After spinning, the centrifuge rotor was allowed to decelerate without braking. A thin band containing the monocytes was seen at the top of the sample (Figure 2.3B). The top 1 mL of filtrate containing the band was aspirated and mixed with 0.5 mL hypertonic D-MEM in a fresh microtube. The microtube was centrifuged at 400 g for 5 minutes. The supernatant was discarded and the remaining mononuclear cell pellet was resuspended in the appropriate solvent for further analysis. The composition of the monocyte-rich fraction was confirmed by microscopy and flow cytometry. Typically more than 95% of cells were monocytes and approximately 10^5 monocytes were extracted from 1 mL whole blood.

2.5.6 Semi-quantitative real-time RT-PCR

Expression of the wild-type TF allele and transgenic TF-GFP allele was analysed by measuring mRNA transcripts for each allele. A commercially available assay containing primers and MGB (minor groove binding) probe for TF was obtained from Applied Biosystems (assay ID Mm00438853_m1). The primers in this assay produce an amplicon spanning the exon 1-2 boundary and thus amplify both the wild-type and transgene mRNAs. This assay is designated TFc in this project. In order to differentially amplify the two alleles custom MGB probes and PCR primers were designed using Primer Express v 1.5 (Applied Biosystems) and obtained from the same company. These assays are herein designated WT for the wild-type allele and EGFP for the transgene, with primer and probe sequences given in appendix 3. These assays are designed to span the exon 5-6 boundary. Both assays use the same forward primer but the reverse primer and MGB probe are specific for either the wild-type or transgenic allele.

Assays and cDNA template were mixed in TaqMan Universal PCR Master Mix, No AmpErase UNG (Applied Biosystems). The final volume was 20 µL containing 3 µL of cDNA generated by a 20 µL first strand synthesis reaction (see above), 900 nmol/L of each primer and 250 nmol/L MGB probe. Assays were performed in duplicate in 96-well plates in a DNA Engine Opticon 2 continuous fluorescence detection system (Bio-Rad). Amplification conditions were 95 °C for 5 minutes followed by 40 cycles of 95 °C for 20 s, 60 °C for 60 s and a plate reading. Data was obtained using Opticon Monitor 3.1 (Bio-Rad) and exported for further analysis in Excel (Microsoft). Cycle threshold was defined as the cycle number at which fluorescence became greater than 4 standard deviations above the mean baseline value. The mean baseline value was the mean fluorescence in cycles 3-10 over the whole plate.

2.5.7 Tissue factor procoagulant activity assay

Monocyte-rich fractions from 1 mL whole blood were resuspended in 100 µL TBS-A and entered into a coagulation assay to estimate TF activity. Assays were performed in duplicate in a water bath at 37 °C. Diagen B & A platelet substitute (Diagnostic Reagents) was used as a source of phospholipid (PL) and pooled mouse normal plasma (MNP) was obtained from Harlan-Sera laboratories. Mouse brain extracts suspended in acetone (Sigma-Aldrich) were diluted in TBS-A to provide TF standards. 50 µL of PL, 50 µL MNP and 50 µL of sample or standard were mixed in a round-bottomed plastic tube and incubated at 37 °C for 30 s. 50 µL of freshly prepared 25 mmol/L $CaCl_2$ was added to initiate coagulation and the time taken for clot formation measured with a stopwatch.

2.5.8 Flow cytometry

Monocyte-rich fractions from 1 mL whole blood were resuspended in 100 µL PBS and equally divided between two microtubes. Samples were incubated with rat monoclonal anti-mouse CD11b (ab25501, Abcam) 2 mg/L or rat IgG2a isotype control 2 mg/L (eBioscience) for 1 hour at room temperature. Both the antibody and isotype control were conjugated with PE/Cy5. After incubation with antibody or control samples were washed three times in PBS by centrifugation at 300 g for 5 minutes. Finally the cells were resuspended in 500 µL PBS.

Samples were analysed in a FACS Calibur system (BD Biosciences). Data was acquired using the CellQuest software provided with the system. Amplifiers were set to linear for scatter channels and log for fluorescence channels. Cell preparations were sorted at a flow rate of 30 cells per second.

2.5.9 Protein extraction from tissues

Mouse tissues were either analysed fresh or snap frozen in liquid nitrogen for later processing. During protein extraction and immunoprecipitation samples and solutions were kept at 4 °C at all times. Tissues were immersed in RIPA or NP-40 lysis buffer with Complete protease inhibitor, 1 tablet per 25 mL (Roche). Typically 1 mL of lysis buffer was used for 100-200 mg tissue. Samples were homogenized with an Ultra-Turrax T8 mechanical disruptor (IKA-Werke). Microtubes were then placed in an end-over-end rotator for 2 hours and subsequently centrifuged at 15,000 g for 20 minutes. The supernatant containing the soluble protein was transferred to a fresh tube and the precipitate containing nuclear material and insoluble cell debris was

discarded. Total protein recovered was measured using the Bio-Rad Protein Assay (Bio-Rad). Typically the final protein concentration was 10-20 g/L indicating that recovered protein represented approximately 10% of the starting tissue weight.

2.5.10 Immunoprecipitation

TF has been estimated to represent just 0.001% total protein in measurements taken from bovine brain (Bach *et al.*, 1981). Therefore TF or GFP were immunoprecipitated from the total protein tissue extracts prior to Western blot analysis. The immunoprecipitation was carried out with approximately 10 mg of protein extract as starting material.

Protein A Sepharose CL-4B beads (GE Healthcare) were used to pre-clear the tissue extracts of immunoglobulin. Beads were hydrated and suspended in immunoprecipitation buffer at 200 g/L. 4 mg (20 µL) of the protein A bead slurry was added per 1 mL tissue protein extract and placed in an end-over-end rotator for 6 hours at 4 °C. Beads were pelleted by centrifugation at 10,000 g for 10 minutes and discarded. The supernatant was transferred to a fresh microtube and the volume made up to 500 µL by the addition of immunoprecipitation buffer. Sheep anti-rabbit IgG magnetic Dynabeads M-280 (Invitrogen) were washed and resuspended in immunoprecipitation buffer at 1-2 x 10^7/mL. Approximately 10^6 magnetic beads were added to each sample.

For immunoprecipitation of TF rabbit polyclonal anti-mouse TF (#4515, American Diagnostica) was used at 1:33 dilution (30 mg/L). For immunoprecipitation of GFP rabbit polyclonal anti-GFP (ab290, Abcam) was used at 1:250 dilution. Samples were placed in an end-over-end rotator at 4 °C overnight.

2.5.11 Western blot

Protein electrophoresis was carried out in 10% polyacrylamide gels (NuPAGE Novex bis-tris gel, Invitrogen) using the XCell SureLock mini-cell system (Invitrogen) with NuPAGE MES SDS running buffer (Invitrogen). Following immunoprecipitation samples were placed in magnetic holders and beads were washed 3 times with immunoprecipitation buffer. Retained beads were resuspended in 30 µL Western reducing buffer and denatured by heating at 95 °C for 5 minutes. 15 µL per lane were loaded onto the gel.

After electrophoresis proteins were transferred on to Hybond ECL nitrocellulose membrane using the Trans-Blot semi-dry system (Bio-Rad) with Novex

TrisGly transfer buffer (Invitrogen). Membranes were blocked in 5% (w/v) non-fat milk in Western blocking buffer for 1 hour. All subsequent washing and incubation steps were carried out in 0.1% milk in Western blocking buffer. Membranes were incubated overnight at 4 °C with primary antibody. For detection of TF, rabbit polyclonal anti-mouse TF (#4515, American Diagnostica) was used at 1:200 dilution (5 mg/L). For detection of GFP, mouse monoclonal anti-GFP (mAb 11E5, Invitrogen) was used at 1:400 dilution (0.5 mg/L).

The binding of primary antibody was visualized in an Odyssey infrared imaging system (Li-Cor). Membranes were incubated for 1 hour at room temperature with secondary antibody, either Alexa Fluor 680 conjugated goat anti-mouse IgG (A21058, Invitrogen) used at 1:1000 dilution (1 mg/L) or IRDye 800CW conjugated goat anti-rabbit IgG (Li-Cor) used at 1:5000 dilution (0.2 mg/L). Finally membranes were washed in 0.1% Tween-20 in PBS before scanning.

A soluble form of mTF that had previously been generated in this laboratory was used as a control. This consisted of the extracellular domains (residues 1-223) with an N-terminal tag used for purification. In this case the tag consisted of part of the protein C sequence. The recombinant soluble TF had been expressed in *E. coli*. A commercially available GFP control was obtained from Clontech (Cat. # 632373).

In this document, images demonstrating electrophoretic separation of proteins are shown with molecular weight markers given in kDa, unless otherwise indicated. The protein standard marker used was the SeeBlue Plus2 ladder (Invitrogen).

2.6 Microscopy

2.6.1 Brightfield microscopy

Smears from whole blood or the monocyte-rich fraction were made on Superfrost 100 slides (VWR) and stained with May-Grunwald-Giemsa. Immunohistochemistry slides from mouse tissues were prepared as detailed below. Slides were visualised using a Nikon Eclipse E400 upright, compound microscope connected to a Nikon DN100 Digital Net camera for image capture.

2.6.2 Fluorescence microscopy

Tissue sections were placed in PBS containing 20% sucrose at 4 °C overnight. Samples were then mounted in Tissue-Tek O.C.T. Compound (Sakura Finetek) and

immersed in liquid nitrogen. Frozen tissues were stored at -20 °C until cutting. 7 µm sections were cut using a Leica CM1800 microtome set at -20 °C and placed onto polylysine coated slides (BDH). Sections were stained with Vectashield mounting medium with DAPI (Vector Laboratories). Slides were viewed using Leica TCS SP1 or Leica TCS SP5 inverted laser scanning confocal microscope systems. DAPI nuclear fluorescence was stimulated with an ultraviolet laser, excitation λ = 365 nm, and emission was detected in the range 410-540 nm. EGFP fluorescence was stimulated with an argon laser, excitation λ = 488 nm. For EGFP detection a variety of bandwidths were tried in the range 500-600 nm. The most suitable bandwidth was found to be 500-550 nm as this reduced the signal from autofluorescence. Images were captured using the Leica Confocal Software provided with the microscopes.

2.6.3 Immunohistochemistry

Freshly-dissected tissues were fixed in 4% formaldehyde overnight. The preparation of paraffin-embedded tissue sections was carried out by the Pathology Department, Hammersmith Hospital, London and A. Carpenter, Haemostasis and Thrombosis, MRC Clinical Sciences Centre. 4-7 µm sections were transferred to slides and dewaxed using a standard method of washing with xylene followed by ethanol. Antigen was unmasked by incubating slides in 0.01M sodium citrate pH 6.9 at 90 °C for 10 minutes. Non-specific binding was blocked by incubating with 10% goat serum (DakoCytomation) in PBS at room temperature for 1 hour. Primary and secondary antibodies were diluted in 10% goat serum in PBS and applied for 1 hour at room temperature. TF was detected using a rabbit anti-mouse TF antibody at 1:100 dilution, kindly provided by Dr. J. Morrissey, University of Illinois. This antibody was prepared by immunising rabbits using a soluble, recombinant form of mouse TF containing only the extra-cellular domains (residues 1-223). Crude serum was prepared from immunised rabbits and used without further purification. Pre-immune serum was used as a control. GFP was detected using rabbit anti-GFP at 1:200 dilution (ab290, Abcam). Pre-immune serum from the same manufacturer was used as a control.

The secondary antibody used was goat anti-rabbit immunoglobulin conjugated with alkaline phosphatase at 1:75 dilution (D0487, DakoCytomation). Between antibody applications slides were washed three times in 0.05% Tween-80 (Calbiochem) in PBS. The activity of alkaline phosphatase was detected using the Fast Red Substrate Chromogen system (DakoCytomation). Slides were incubated with the

substrate and levamisole 1 mmol/L, to inhibit the activity of endogenous alkaline phosphatases, for 30 minutes. Pre-immune goat serum was used as a control.

An alternative method was also tested for anti-GFP immunohistochemistry. In this method the primary antibody was rabbit anti-GFP at 1:1000 dilution (A11122, Invitrogen). A biotin-streptavidin detection system was used (Vector Laboratories). In this system the secondary antibody is a biotinylated goat anti-rabbit immunoglobulin used at 1:300 dilution. Streptavidin conjugated horseradish peroxidase (HRP) is then applied to bind the captured biotin. HRP activity is visualised following application of the chromogenic substrate 3,3' diaminobenzidine (DAB) which results in formation of a brown dye.

After antibody detection slides were either viewed without further processing or counterstained with haematoxylin using standard methods. Slides were viewed with an upright conventional light microscope as detailed above.

2.7 Software

In addition to standard office programs and the software already mentioned, various specialized applications were used during this project. Molecular graphics were generated using Insight II (Accelrys) on a Silicon Graphics Inc (SGI) Octane2 workstation running IRIX 6.5. Bioedit Sequence Alignment Editor © Tom Hall, Ibis Therapeutics, Carlsbad, California was used for sequence analysis and generation of alignments. Plasmid diagrams were drawn with Simvector 4.0 (Premier Biosoft). Image stacks generated by microscopy were analysed with ImageJ (open source from National Institute of Health, USA available at http://rsb.info.nih.gov/ij/).

Chapter 3 - Targeting of Embryonic Stem Cells

3.1 Introduction

In the last two decades there has been a great increase in the number of studies involving manipulation of the mammalian genome. This was made possible by a number of breakthroughs in the early 1980s, including the isolation of ES cells and the demonstration of their ability to contribute to the germline of host embryos (reviewed in Ledermann, 2000). In 1985 Smithies and colleagues reported the first successful targeting in mammalian cells when they used a plasmid to introduce mutations into the β-globin gene. Four years later mice with targeted disruption of the hypoxanthine phosphoribosyltransferase gene were described (Koller *et al.*, 1989). This represented the first mouse knockout model of human disease.

Targeting of ES cells has become the most widely-used method of generating transgenic mice carrying a modification at an endogenous locus. Although various mammalian cell types have been successfully used in gene targeting experiments, ES cells remain the preferred choice for generation of transgenic animals on account of their totipotency, i.e. their ability to differentiate into any of the cell types found in a multicellular organism. The first step in the process requires the generation of a targeting vector, or construct, containing the sequence to be inserted flanked by regions carrying identical sequence to that at the target locus: the homologous arms (HAs). The targeting construct is then introduced into ES cells and homologous recombination (HR) allowed to occur during subsequent cell divisions. HR is the process by which DNA is exchanged between sister chromosomes and commonly occurs in two normal processes: repair of double strand breaks (DSB) and during meiosis. In DSB repair the alignment of homologous regions allows the damaged DNA to be regenerated using the intact sequence as a template. During meiosis the alignment of homologous regions is followed by transfer of genetic material between sister chromatids, a key feature of the evolutionary process. HR is thus a natural phenomenon that may occur in any cell. Because of its potential for genetic alteration HR is tightly controlled and achieved by complex mechanisms that are not yet fully understood.

Gene targeting utilises the inherent HR potential in mammalian cells by providing a homologous region for alignment to the gene of interest. As HR occurs the crossover process leads to integration of the sequence to be inserted specifically

at the locus of interest. In contrast non-homologous recombination (NHR) occurs when DNA sequences are inserted randomly at any position in the genome, often at regions containing a DSB or single strand break.

Targeting at the *mF3* locus has been successfully carried out by a number of groups (Bugge *et al.*, 1996; Carmeliet *et al.*, 1996; Toomey *et al.*, 1996; Melis *et al.*, 2001). All of these studies involved disruption of the TF locus such that the targeted allele produced no TF protein or only a truncated form. There are no reports in the literature of knocking-in a sequence into the intact locus such that the full length protein is produced fused to a reporter protein. The purpose of this part of the project was to target ES cells such that one of the two *mF3* alleles expresses EGFP fused to the intact TF protein. These heterozygous ES cells may then be used in blastocyst injections to generate chimaeric mice with the TF-EGFP fusion gene incorporated in the germline.

3.2 Design of the targeting construct

The targeting construct initially used in this chapter was generated during a earlier project in this research group (Mumford, 2003). The construct design follows well-established strategies for targeting in mammalian cells. The sequence to be inserted is flanked by two regions, the HAs, containing sequence identical to regions of *mF3*. The insert contains the reporter gene, in this case EGFP, and a gene encoding resistance to a specific drug. Unless recombination occurs within the insert, the incorporation of the reporter gene is then linked to drug resistance and can be used to positively select for targeting events.

The drug resistance gene used encodes for the bacterial enzyme aminoglycoside-3-phosphotransferase and is carried in the NeoR cassette. Aminoglycosides have their bactericidal effect by interfering with the function of ribosomal subunits thereby causing a generalised disruption of protein synthesis, resulting in cell death (Mehta & Champney, 2002). Certain aminoglycosides such as paromycin, hygromycin and gentamicin (G418) are sufficiently toxic to eukaryotic cells to be used as selection agents in tissue culture (Eustice & Wilhelm, 1984). Aminoglycoside-3-phosphotransferase confers resistance to the aminoglycoside group of antibiotics by phosphorylating the drug. The NeoR cassette includes a promoter that drives expression in eukaryotic cells and ensures that the gene is expressed irrespective of the site of integration in the genome. G418 resistance is

now well-established as one of the main methods of positive selection for mammalian targeting events.

The main drawback of HR as a mechanism for gene targeting is its extremely low frequency in comparison with random genomic integration by NHR. The addition of G418 to the tissue culture medium positively selects for insertion of the targeting construct. However the majority of G418 resistant clones are generated by NHR. In order to select against NHR clones a second marker encoding sensitivity to a specific drug is placed in the targeting construct outside the HAs. The drug most commonly used in these targeting projects is ganciclovir (9-(1,3-dihydroxy-2-propoxy)methyl guanine). For full activity ganciclovir requires phosphorylation to the active triphosphate (ganciclovir-5'-triphosphate) by the enzyme thymidine kinase (Matthews & Boehme, 1988). This metabolite is a potent and selective inhibitor of viral DNA polymerase, but at high intra-cellular levels inhibits mammalian DNA polymerases also. The targeting construct contains the gene for herpes simplex virus thymidine kinase (HSVTK) which confers sensitivity to ganciclovir. As HR only results in insertion of the sequence between the HAs, it would not be expected to lead to integration of the HSVTK cassette. In NHR the expectation is that the whole targeting construct is integrated thereby leading to sensitivity to ganciclovir. Unfortunately, phosphorylation of ganciclovir can be catalysed by a few mammalian kinases, such as phosphoglycerate kinase. Consequently this drug suffers from a significant bystander effect that limits its usefulness.

It is now well-established that the expression of some mammalian genes is adversely affected by the presence of bacterial sequences within the gene. HR between *mF3* and the targeting construct results in the presence of the bacterial NeoR cassette between the gene and its 3' untranslated region (UTR). Although the exact function of the 3' UTR is unknown in *mF3*, it is widely believed that this region has a post-transcriptional regulatory role (Mackman *et al.*, 1992; Mackman, 1997). It is therefore desirable to remove the NeoR cassette after it has served its purpose during the selection of ES cell clones. Commonly used methods for removal of genetic material in gene targeting procedures involve site-specific recombination systems such as the Cre-Lox or FLP-FRT systems (Metzger & Feil, 1999; Thomson & Ow, 2006). The enzymes mediating recombination do so in a highly efficient and specific manner without the requirement of additional co-factors. Although both systems are bacterial in origin they have activity in eukaryotic cells and have been successfully used in the removal of specific sequences in transgenic mice (Nagy, 2000;

Branda & Dymecki, 2004). The Cre-Lox system has been selected for use in this project because transgenic mice expressing cre, both ubiquitously and in an organ-specific manner, are readily available. In addition loxP sites are frequently present in vectors, which will be advantageous if plasmid modification is required. For these reasons the NeoR cassette has a loxP site at either end orientated in the same direction. The mechanism of cre-mediated recombination is further discussed in chapter 4. Transgenic mice carrying the targeted *mF3* allele will be crossed with mice expressing cre recombinase, resulting in excision of the intervening sequence between the loxP sites.

3.3 Assembly of pES-TC-S

The components of the targeting construct were inserted by sequential sub-cloning into pBluescript KS (-) vector (Stratagene) by Dr. A. Mumford, Haemostasis and Thrombosis, MRC Clinical Sciences Centre. The resulting targeting construct, pES-TC-S, with selected sequence details, is shown in Figure 3.1. The HAs were generated by high fidelity PCR from a 129/Sv cosmid genomic library (Poustka *et al.*, 1984) using primers with the 3′ ends modified to contain the desired restriction sites. The EGFP-loxP-NeoR-loxP fragment was excised from pSP72 (Godwin *et al.*, 1998) and inserted in-frame in-between the last codon at the 3′ end of *mF3* and the natural stop codon. The NeoR cassette is flanked by loxP sites to allow its subsequent removal by cre recombinase. The HSVTK fragment was excised from pIC 19R-HSVTK (Mansour *et al.*, 1988). The entire insert was sequenced to verify that no errors had been introduced during the cloning process. It was particularly important to demonstrate that the EGFP coding sequence had been inserted appropriately in relation to *mF3* (Figure 3.1B) and that the loxP sites were of the same sequence and correctly orientated (Figure 3.1C).

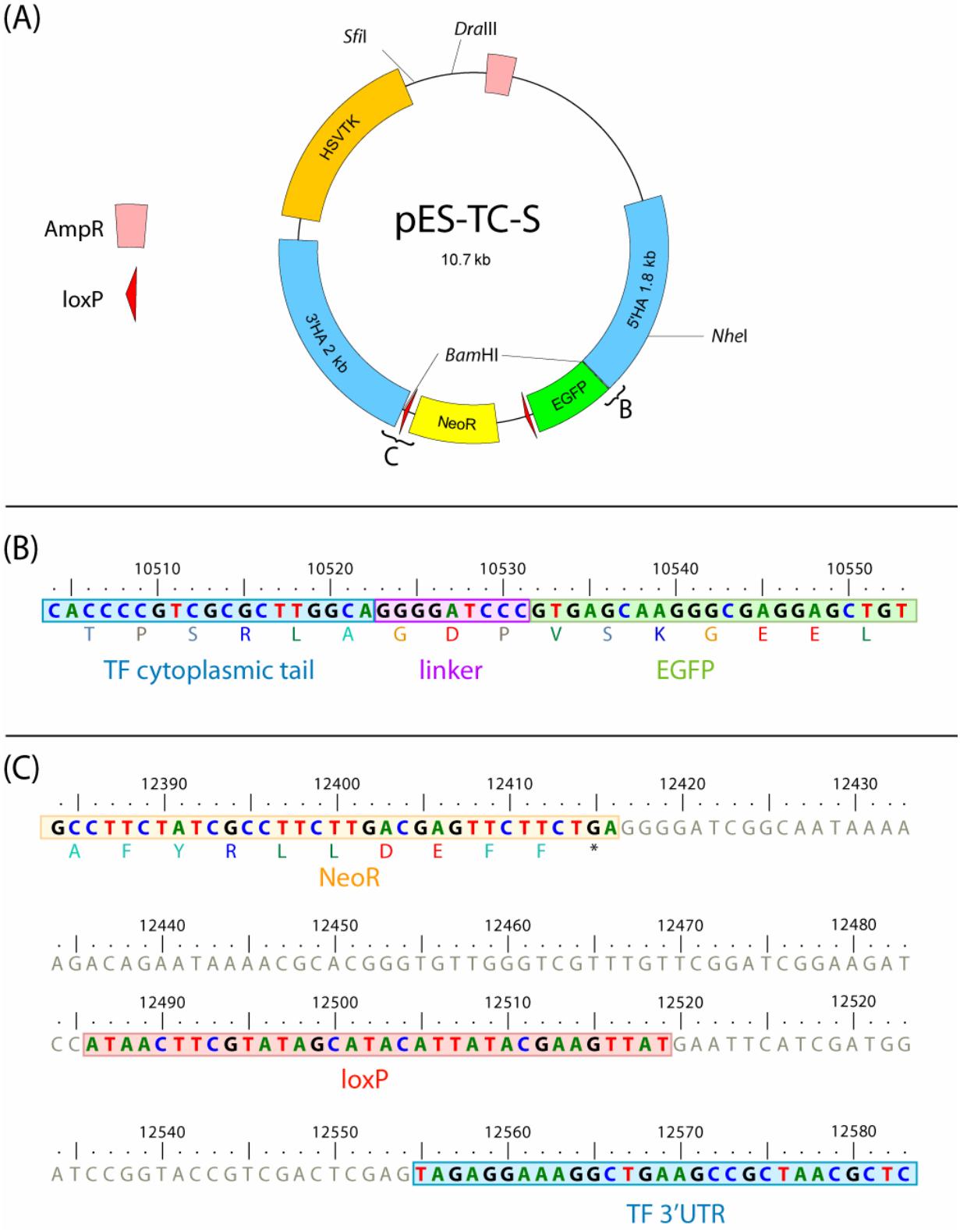


Figure 3.1 (A) Map of pES-TC-S. The position of restriction sites used in the project are shown. Sequence details of selected regions are shown below. (B) The region between the 5' HA and EGFP showing the position of the linker tripeptide that maintains the correct reading frame. (C) The region between NeoR and the 3' HA showing the position of the loxP site. Sequence numbering is that of the predicted EGFP-targeted *mF3* gene. The single letter translations of coding regions are given below the sequence.

The key features of the targeting construct are:

i) The HAs totalling 3.8 kb are made from DNA isogenic with that in the selected ES cell lines. The 5′ HA consists of exon 5, intron 5 and exon 6, up to the last codon of the mF3 coding sequence. The 3′ HA consists of the natural stop codon and the 3′ UTR.

ii) The sequence to be inserted (non-homologous region) consisting of EGFP-loxP-NeoR-loxP and totalling 2 kb is placed in between the HAs. NeoR includes a self-contained promoter and will be expressed irrespective of the integration site. EGFP expression requires correct insertion in-frame with a pre-existing coding sequence and downstream of a suitable promoter.

iii) The coding sequence for a linker tripeptide (Gly-Asp-Pro) is placed in between the TF and EGFP coding sequences to ensure that the latter is in-frame with the former.

iv) The cDNA for herpes simplex virus thymidine kinase (HSVTK) is placed after the HAs, so that it would not be incorporated during a HR event but would be inserted during NHR. This should render clones that become resistant to G418 by NHR also sensitive to ganciclovir. ES cells that undergo HR should not become sensitive to ganciclovir.

The mechanism of HR between the targeting construct and *mF3* leading to incorporation of the desired sequences at the 3′ end of the gene is shown in Figure 3.2.

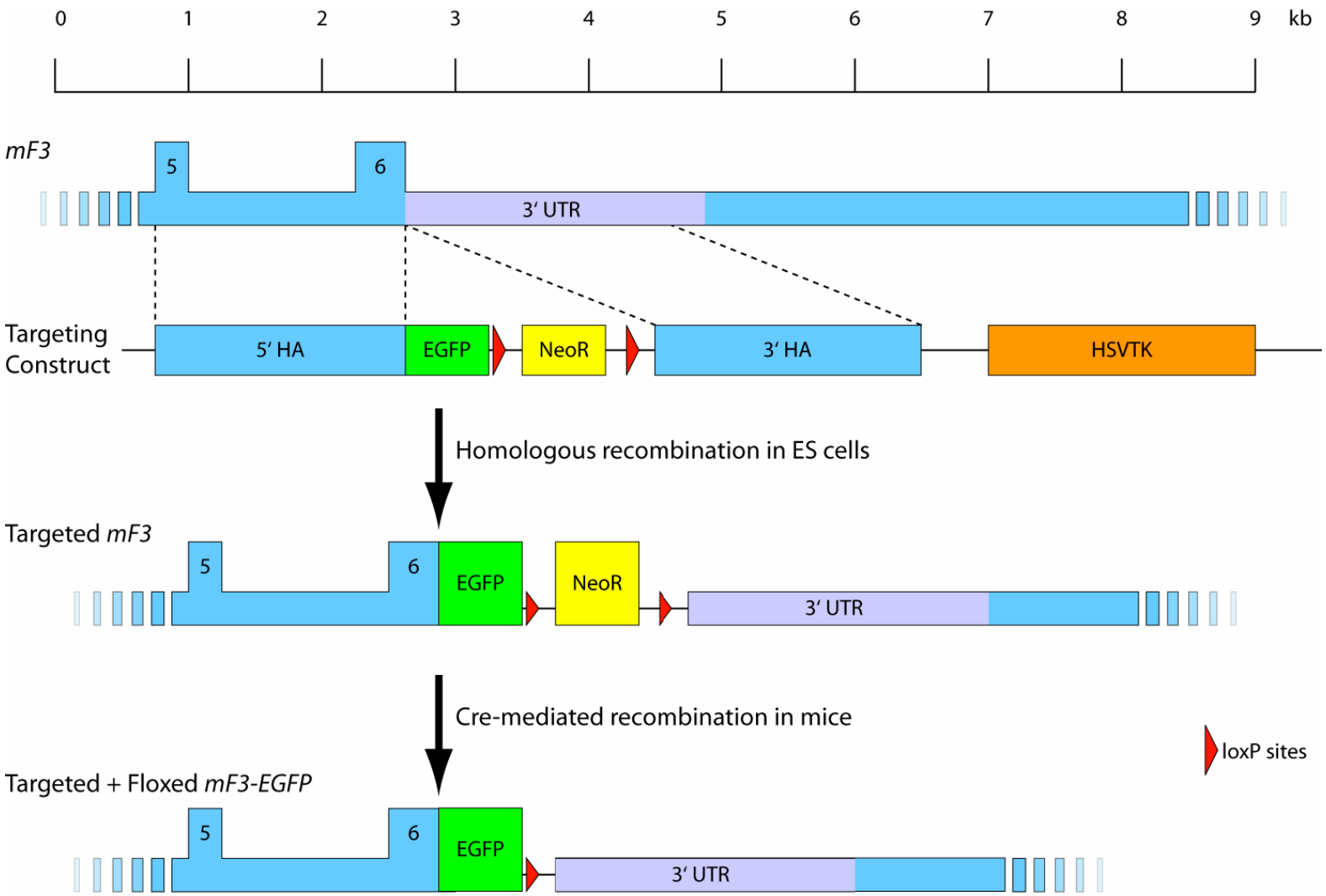


Figure 3.2 Targeting strategy for insertion of EGFP at the *mF3* locus in ES cells using pES-TC-S. Exons in *mF3* are shown by numbered blocks. TC represents the linearized targeting construct. The positions where HAs map to *mF3* are shown by dashed lines. HR leads to the insertion of the EGFP-loxP-NeoR-loxP cassette between exon 6 and the intact 3′ UTR. The HSVTK cassette is not incorporated during HR but would be included in NHR. Subsequent floxing in cre-expressing mice leads to removal of the NeoR cassette.

3.4 Screening of ES cell clones

A single targeted ES clone had been previously generated in 28B cells in this laboratory using the targeting construct pES-TC-S by Dr. A. Mumford (2003). Repeated blastocyst injections with this clone had failed to generate chimaeric mice indicating that the clone was not able to contribute to the germline to any significant extent. However, the targeted clone was used as source of DNA to verify the screening strategy to be used for subsequent ES cell clones.

3.4.1 Efficiency of the DNA extraction process

After targeting, clones that were resistant to G418 and ganciclovir were picked and expanded in 96-well plates. Duplicate plates were made of which one was frozen and the other used for DNA extraction. The DNA was screened for the presence of the targeted allele by two methods. From the total DNA extracted, 10% was used in a PCR based screen and the remaining 90% was screened by Southern Blot.

In order to maximise the amount of DNA available for screening, wells were incubated for a further 1-2 days after they had become fully confluent. ES colonies were grown on gelatin, rather than a fibroblast feeder layer, to eliminate the possibility of contamination from fibroblast DNA. The mean number of ES cells present in a fully-confluent well was 5 x 10^5 (range 10^5-10^6, n = 24). The efficiency of the DNA extraction process was assessed by comparing the amount of DNA extracted against the total amount of genomic DNA that would be expected to be present in a well. The weight of a single copy of the mouse genome may be calculated by the equation shown below. This assumes that the average molar weight (*MW*) of a nucleotide is 308 g and that a single copy of the mouse genome has 3.4 x10^9 base pairs (*G)*. Avogadro's constant is denoted by the symbol N_A.

$$\text{Molecular weight of mouse genome} = \frac{MW \times G \times 2}{N_A}$$

$$= \frac{308 \times (3.4 \times 10^9) \times 2}{6.022 \times 10^{23}}$$

$$= 3.5 \text{ pg}$$

From this it can be calculated that 5 x 10^5 ES cells would contain 3.5 µg of genomic DNA. The actual mean amount of DNA obtained was 2.5 µg (range 1-5 µg, n=24), or about 2/3 of the amount expected.

3.4.2 Verification of the PCR screen

The PCR screen utilised a 5′ primer, NP2, in the NeoR cassette and a 3′ primer, MTFP2, in the region downstream of the *mF3* 3′ UTR and outside the sequence contained in the 3′ HA. This generated a 2381 bp amplicon that was specific for targeted clones generated by HR (amplicon A in Figure 3.3). A second control PCR reaction was used to confirm the adequacy of the DNA template. For the control PCR both primers, N5 and N3, were positioned within the NeoR cassette. This generates a 443 bp amplicon in any clone with integration of the intact targeting construct and is positive in both HR and NHR clones (amplicon B in Figure 3.3).

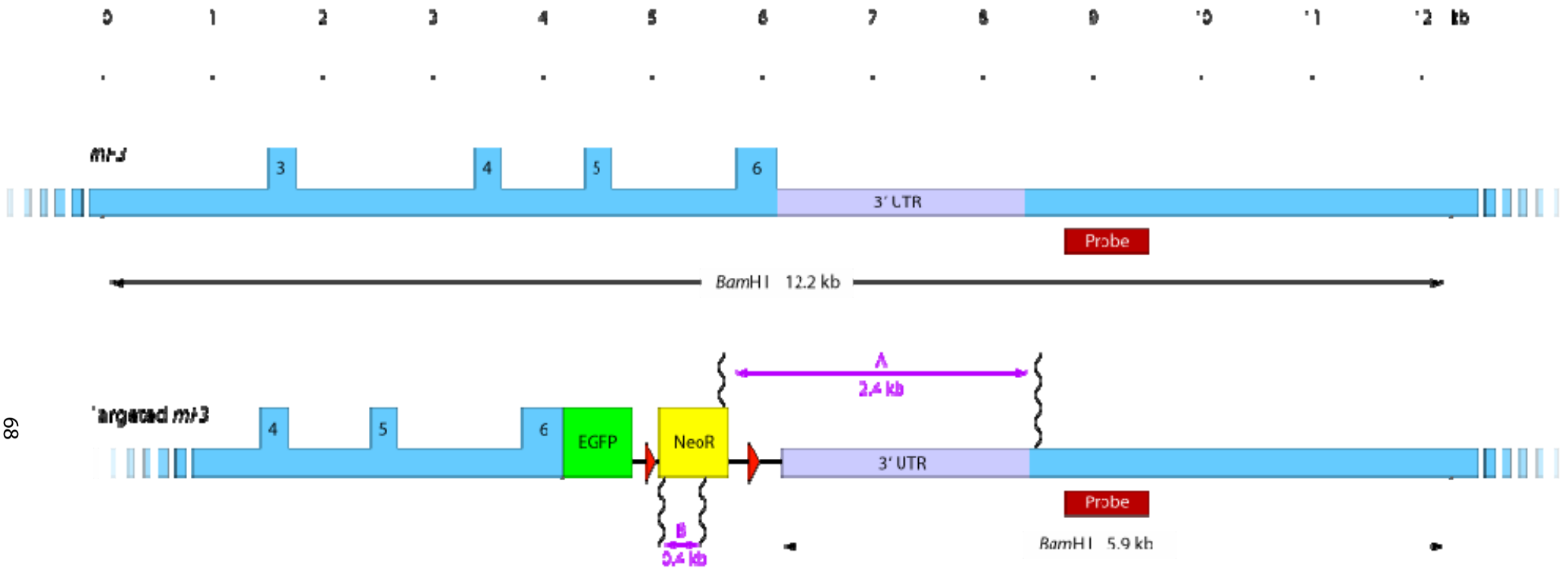


Figure 3.3 Strategies for screening for HR in targeted ES cell clones. Dashed lines indicate *Bam*HI restriction sites. Vertical wavy lines indicate primer binding sites. PCR amplicons are shown by purple, arrowed lines (A and B) with the amplicon size shown below. The location of the probe used in Southern blot analysis is shown by a red rectangle. HR is detected by PCR with NP2 and MTFP2 which produces amplicon A. PCR with primer pair N5 – N3 produces amplicon B in both HR and NHR clones.

Standard *Taq* DNA polymerase (Red Hot, Abgene) was unable to consistently amplify the 2381 bp HR-specific template. Therefore, two high fidelity enzyme systems were assessed for use in the screening assay: Easy-A PCR cloning enzyme (Stratagene) and Platinum *Taq* DNA polymerase (Invitrogen). The product literature recommended that the amount of genomic DNA required for efficient amplification of large templates is 50-100 ng for Easy-A and 100 ng for Platinum *Taq*.

The average amount of DNA available for the PCR screen (in 10% of the total extracted from a confluent well) is 250 ng, although from wells with the lowest number of cells this amount may be about 100 ng. The PCR screen therefore needs to be able to amplify the template from this lower amount of DNA. Figure 3.4 demonstrates that Easy-A was unable to satisfactorily amplify the template from less than 500 ng of genomic DNA. Platinum *Taq* was able to amplify the template from 50 ng of DNA and a fainter band was seen with 25 ng of DNA. As Platinum *Taq* had the greater sensitivity over the required range, subsequent screening for targeted colonies was carried out using this enzyme.

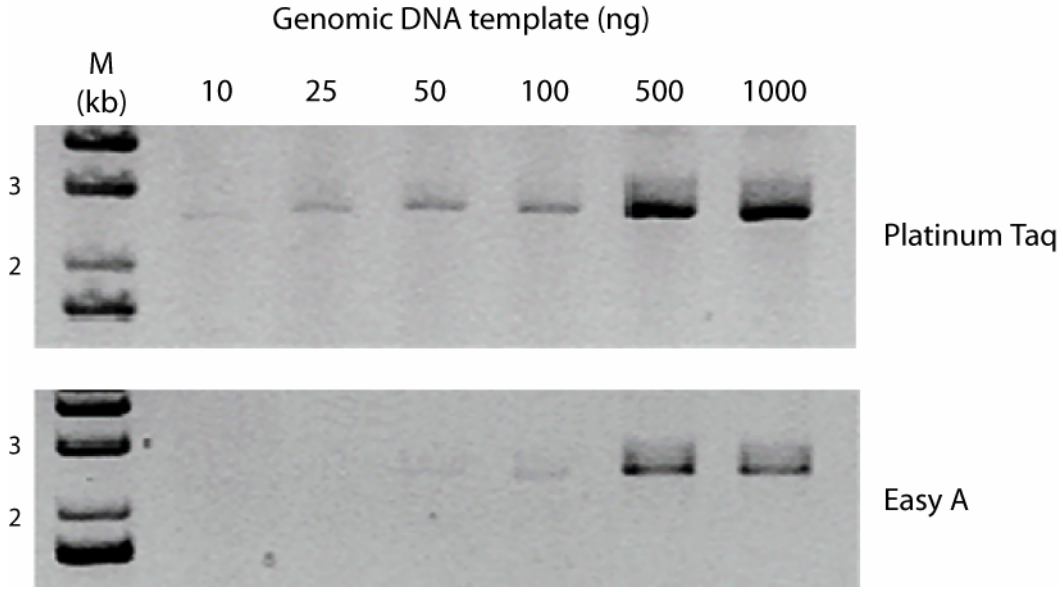


Figure 3.4 Comparison of the sensitivity of two high fidelity DNA polymerases in the reaction amplified by primer pair NP2 and MTFP2. The same DNA template was entered into each reaction and quantities are given above the corresponding lane. For use in the PCR screen the enzyme has to be capable of amplifying a product from a minimum of 100 ng of DNA.

The amount of DNA available from wells with the lowest amount of cells is likely to be towards the lower end of the range at which Platinum *Taq* was shown to have sensitivity. It is consequently undesirable to split the sample into two aliquots for separate PCR reactions and preferable to use a duplex mixture containing both the HR-specific and control PCR primer sets. Running PCRs in multiplex increases the chances of primer-primer interactions and false priming. The 5′ primers used in the two reactions, NP2 and N5, are 816 bp apart raising the possibility that both could act as forward primers with the reverse primer used in the HR-specific reaction, MTFP2. This would be undesirable as it would result in the production of two higher molecular weight bands in correctly targeted clones. One would expect that these bands would be of lower intensity than a single band from the same template, reducing the sensitivity of the screening test and making interpretation more difficult. To eliminate the potential adverse effects of a multiplex reaction, experiments were carried out using the control DNA to verify that performing the screen in duplex format produced only the same bands as seen with the reactions in isolation.

3.4.3 Comparison of digoxigenin and α-32P labelled probes in Southern blot screening

Restriction maps of *mF3* and the predicted TF-EGFP transgene sequence were examined to identify differences that would distinguish the wild-type and targeted alleles. In the wild-type allele digestion with *Bam*HI generates a 12.2 kb fragment at the 3′ end of the gene. *Bam*HI sites are present in the targeting construct between the 5′ HA and EGFP and also between the last loxP site and the 3′ HA. Thus correct integration of the targeting construct produces two bands of 6.3 and 5.9 kb in place of the 12.2 kb wild-type fragment (Figure 3.3).

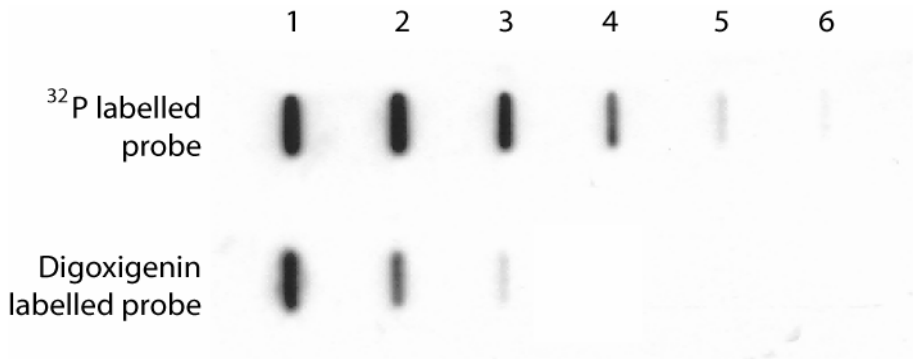
A probe for use in Southern blot analysis was generated by PCR using the primer pair PbES2F and PbES2R. This produces a 777 bp amplicon that hybridizes downstream of the 3′ HA.

The standard method for labelling of probes involves the incorporation of radioactive nucleotides. A potentially less hazardous method involves the incorporation of digoxigenin-conjugated nucleotides which may then be detected using an enzyme-conjugated anti-digoxigenin antibody. This technique has been used successfully to detect transgenes with multiple copies in the genome. However, targeting at an endogenous locus requires detection of a single copy of the

transgene in the genome. It is unclear whether digoxigenin-labelled probes have sufficient sensitivity for this purpose.

To investigate the possibility of using digoxigenin labelling for Southern analysis the 777 bp probe described above was labelled with either digoxigenin or α-^{32}P. Genomic DNA was prepared from EL M3 ES cells and quantified. Serial dilutions of the digested DNA were made to reflect the range of DNA amounts that would be available for screening by Southern blot. These were applied to a membrane in duplicate using a slot blot according to the table in Figure 3.5. Adjacent rows of the slot blot containing duplicates of digested DNA were separated and probed with the manufacturer's recommended amounts of digoxigenin- or radio-labelled probe.

(A)



(B)

Column	Amount of DNA (µg)	No. of template copies
1	10	2.9×10^6
2	5	1.4×10^6
3	2	5.7×10^5
4	1	2.8×10^4
5	0.5	1.4×10^5
6	0.25	7.1×10^4

Figure 3.5 (A) Varying amounts of genomic DNA were applied to a slot blot and probed with either radio- or digoxigenin-labelled probe. (B) The amount of DNA applied to each position and the number of template copies that this represents is shown in the table below.

In the previous section it was determined that, on average, about 2.5 µg of genomic DNA would be available for screening by Southern blot. However, as there is variation in the number of ES cells present in a well and in the efficiency of the DNA extraction process, the amount of DNA available may be as low as 1 µg. The Southern screen therefore needs to be able to detect the template in this amount of DNA. Figure 3.5 demonstrates that the digoxigenin-labelled probe was unable to detect the template in less than 2 µg of genomic DNA, whereas the same probe sequence labelled with radioactivity was sensitive down to 0.5 µg, and possibly lower. The radioactive labelling method was therefore used to generate probes for screening ES cells.

3.4.4 Using EGFP fluorescence to screen for targeted clones

The requirement for a screening strategy stems from the fact that the selection markers used are relatively inefficient at enriching the mixture of clones in favour of HR as opposed to NHR. An obvious selection marker that would differentiate between HR and NHR is the expression of EGFP. Unlike the NeoR cassette which has its own promoter and is expressed independently of the site of integration, EGFP expression will only take place if the gene is inserted correctly in relation to an endogenous promoter and in-frame with the endogenous coding sequence. This is designed to occur with correct targeting of *mF3*, but is very unlikely with random integration in the genome. Screening clones by EGFP fluorescence may then be a way of greatly enriching the pool for those that have undergone HR.

In order to establish the feasibility of this strategy it is first necessary to demonstrate that *mF3* is expressed in ES cells, as the endogenous promoter may not be active in this cell type. *mF3* is known to be expressed by cells at embryo day 6.5 (Luther *et al.*, 1996), however it is not known if expression is seen in ES cells in culture. To investigate this, total RNA was isolated from the three ES cell lines used and from murine fibroblasts and analysed for *mF3* expression by RT-PCR. Fibroblasts are known to express functional TF (Mackman *et al.*, 1992) and are therefore a suitable positive control. Following the RT step, cDNA generated from TF mRNA was detected using the primer pair MTFCF and MTFCR producing a 624 bp amplicon. GAPDH was amplified with primer pair MGAPDHF and MGAPDHR producing a 531 bp band, to act as a transcription control. Dependence of the reaction on reverse transcriptase (RT) was demonstrated by a no RT control.

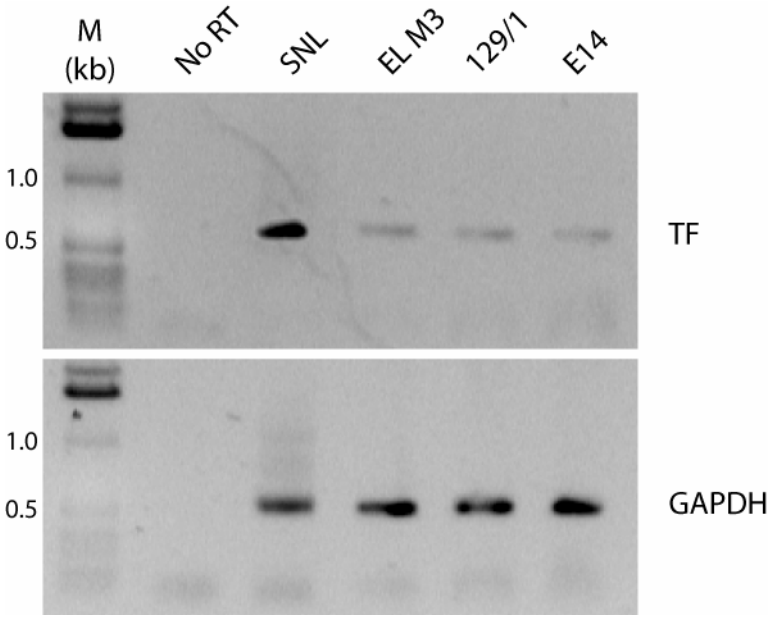


Figure 3.6 RT-PCR showing TF expression (top panel) with GAPDH shown as a transcription control. No RT and fibroblast (SNL) controls were analysed alongside samples from the three ES cell lines. Relatively faint levels of expression can be seen in all ES cell lines.

TF expression was detected in all three ES cell lines (Figure 3.6). Although this was not a quantitative assay, the same amount of total RNA was entered into each reaction and therefore some information about levels of expression may be gained. The intensity of the GAPDH band was similar with all samples. The TF band was of a lower intensity in ES cells compared with SNL cells. There are two possible explanations for this finding. Firstly, that TF expression does occur in ES cells but at a lower level than in fibroblasts. The second explanation is that ES cells do not express TF at the pluripotent stage but may differentiate into TF expressing cells. Although ES cells were cultured in the presence of LIF, this does not completely prevent differentiation occurring in a few cells. Therefore it is possible that the low level of TF expression seen might originate from a small population of other cells derived from ES cells, such as fibroblasts.

In order to establish whether this level of *mF3* expression would be sufficient to produce detectable EGFP fluorescence, electroporated ES cells were assessed by flow cytometry. EL M3 cells were electroporated with pES-TC-L and selected as normal. A control reaction was performed in which cells were electroporated with a vector containing the EGFP coding sequence driven by a CMV promoter with a NeoR

cassette under separate control (pcDNA 3.1-EGFP, Invitrogen). The control reaction was subject to positive selection with G418 only. 12 days after electroporation, when more than 100 colonies were visible macroscopically on each plate, the colonies were examined under a fluorescence microscope. About 1 in 5 colonies showed fluorescence in the control reaction, while no fluorescent colonies were seen in the cells electroporated with pES-TC-L. The colonies were then trypsinized and analysed for fluorescence by flow cytometry (Figure 3.7). This demonstrated that an EGFP positive population was detectable in the control reaction but that no fluorescent population was detectable in the cells electroporated with pES-TC-L. This result indicated that although *mF3* expression was detectable in ES cells the level was insufficient to produce detectable fluorescence from the attached EGFP reporter gene. Alternatively, ES cells may not have the necessary post-translational mechanisms for production of TF protein and the mRNA detected may in fact represent ectopic transcription.

In a population where the targeted cells (positives) are at a relatively low level the critical feature of the screening test is to have a high negative predictive value, so as not to mistakenly classify targeted cells as negative. Because the level of *mF3* expression is low in ES cells, using EGFP expression as a screening tool does not satisfy this criteria. Therefore ES clones were not routinely assessed for fluorescence after targeting.

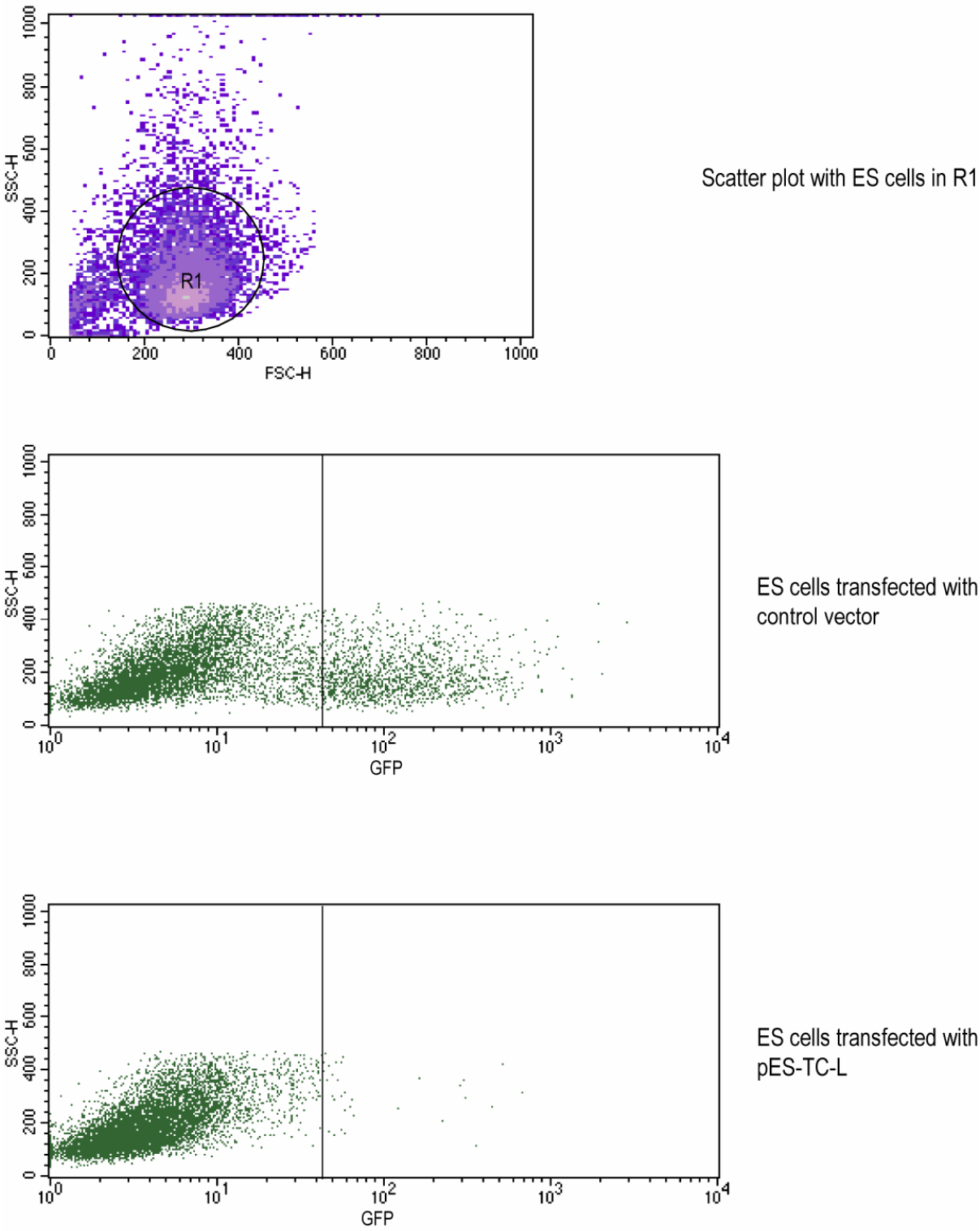


Figure 3.7 EL M3 cells were electroporated with pES-TC-L or an EGFP expressing control vector and assessed by flow cytometry. The top panel shows the gating of ES cells by forward and side scatter. In the middle panel an EGFP positive population is detectable in the control reaction. In the lower panel the ES cells transfected with pES-TC-L are negative for EGFP fluorescence.

3.5 Targeting with pES-TC-S

The targeted 28B cell line that had been generated during previous work in this laboratory had failed to contribute to the germline following blastocyst injections. There are a number of possible causes for this, such as the presence of cryptic chromosomal abnormalities. Further testing of the 28B cell line had shown that it was infected with mycoplasma. This may have prevented germline contribution and required that the infected stocks of this cell line were destroyed. As no uninfected 28B stocks were available, two different cell lines were initially selected for targeting. Both 129/1 and E14 cell lines had previously been shown to have germline potential in published murine gene targeting projects (Hooper *et al.*, 1987; Kay *et al.*, 1993). However, stocks of the two cells lines were already at a higher passage than would be ideal for a targeting project and consequently a third ES cell line was obtained: EL M3. These were of a lower passage (Table 2.1) and have also been successfully used in previously reported studies (Brambrink *et al.*, 2006). Prior to targeting, stock EL M3 cells were used in control blastocyst injections to generate mice that were >90% chimaeric by coat colour. Subsequent breeding of these mice and the presence of agouti pups in the F1 generation confirmed the germline potential of EL M3 cells in this laboratory.

pES-TC-S was linearized by digestion with *Dra*III and 25 µg were used to electroporate 8 x 10^6 ES cells. The results of 14 separate targeting experiments involving 22 electroporations are shown in Table 3.1.

ES cell type	Total ES cells electroporated	$G418^R$ & $Ganc^R$ Colonies	Colonies screened	NHR clones	HR Clones
129/1	11.2 x 10^7	607	586	585	0
E14	3.2 x 10^7	285	267	265	0
EL M3	3.2 x 10^7	465	457	457	0
Total	17.6 x 10^7	1357	1310	1307	0

Table 3.1 Results of 22 electroporations using pES-TC-S. $G418^R$ & $Ganc^R$ refers to G418 and ganciclovir resistance. All colonies were screened by a combination of PCR and Southern blot. NHR = non-homologous recombinant, HR = homologous recombinant clones.

Electroporation with 129/1 and E14 cells produced fewer G418 and ganciclovir resistant colonies than were obtained with EL M3 cells. In comparison with the other two cell lines, EL M3 cells grew more rapidly after splitting and thawing into well-demarcated ES cell colonies. These results may be explained by the relatively lower passage of the EL M3 cell line. Higher passage ES cells have a tendency to differentiate, which is evidenced by a spreading of colonies with flattened cells at the colony edge. Furthermore, senescence is likely to render cells less tolerant of the electroporation process.

Relatively few experiments were possible with E14 cells. This is because stocks at passage 23 were of limited number. Expansion of stocks was attempted, which inevitably resulted in an increase in passage number. In addition to the comments above regarding the effects of senescence, this increase was associated with a reduction of metaphases demonstrating a normal karyotype to less than 70%. This is below the acceptable limit for use in blastocyst injections.

In summary, more than 1300 colonies were screened, by both Southern blot and PCR analysis, after targeting with pES-TC-S and none were found to have undergone HR. It was therefore apparent that a modification of the targeting strategy was required.

3.6 Assembly of pES-TC-L

Experiments with the first targeting construct, pES-TC-S, failed to produce any targeted clones. The low efficiency of HR is the most likely explanation for this. In order to improve the HR rate, modifications were made to the targeting construct to increase the length of the 5′ HA.

Due to the lack of suitable restriction sites, it was not possible to simply linearize pES-TC-S at the proximal end of the 5′ HA and insert part of the *mF3* sequence that lies upstream of this region. Therefore the components of the targeting construct were subcloned into intermediate vectors and then the extension to the 5′ HA was inserted to generate a new vector.

3.6.1 Generation of p1083

pES-TC-S contains a single *Nhe*I site in the 5′ HA which corresponds to g.9842 in *mF3*. Digestion of pES-TC-S with *Sfi*I and *Nhe*I produces a 6.7 kb fragment consisting of part of the original 5′ HA, EGFP, NeoR, the 3′ HA and HSVTK. This combination of restriction enzymes is not seen in the multiple cloning site of most

plasmids. pBST(+) is a version of pBS(+) (Stratagene) in which the polylinker has been replaced by a custom made sequence containing *Sfi*I and *Nhe*I sites. T7 and T3 primers are positioned at either end of the polylinker to allow verification of the insert by PCR. This cloning vector was kindly provided by Dr. I Garner, Delta Biotechnology. The cloning of the 6.7 kb *Sfi*I – *Nhe*I fragment from pES-TC-S into pBST(+) creates the intermediate plasmid p1083 (Figure 3.8A). Correct formation of this vector was confirmed by restriction digest and PCR across the inserted ends.

3.6.2 Generation of p1103

Digestion of *mF3* with *Nhe*I produces a 4.3 kb fragment between the cut sites at g.5573 and g.9842. Initially, attempts were made to generate this fragment by PCR using high fidelity DNA polymerases. A number of PCR amplicons were subcloned into T-vector and sequenced. No amplicons were generated without errors and on average there were 5-10 incorrect nucleotides per 4.3 kb amplicon. The desired fragment was therefore produced by digestion of genomic EL M3 DNA.

Attempts to directly shotgun clone the g.5573 – g.9842 fragment into p1083 from a mixture of genomic *Nhe*I fragments failed. However this fragment was successfully shotgun cloned into pBST(+) (Figure 3.8B). In shotgun cloning the clones containing the desired fragment represent a very small percentage of the total as the entire population of genomic restriction digest fragments is available for insertion. Therefore a high throughput screening technique is required and the clones were screened by Southern blot analysis of colony lifts. This second intermediate plasmid was designated p1103.

3.6.3 Generation of pES-TC-L

The p1103 vector backbone contains 14 *Msp*I sites while there are none in the inserted *mF3* fragment. p1103 was digested with *Nhe*I and *Msp*I and the g.5573 – g.9842 fragment was then cloned into the single *Nhe*I site in p1083 corresponding to g.9842 (Figure 3.8C). Simultaneous digestion with *Msp*I prevents reformation of p1103 in this step. The resulting new targeting vector was designated pES-TC-L. This has similar features to pES-TC-S with the significant difference being an increase in the size of the 5′ HA from 1.8 to 4.9 kb (Figure 3.9). The final construct was verified by a combination of digestion with selected restriction endonucleases and direct sequencing of PCR amplified products at regular intervals across the insert.

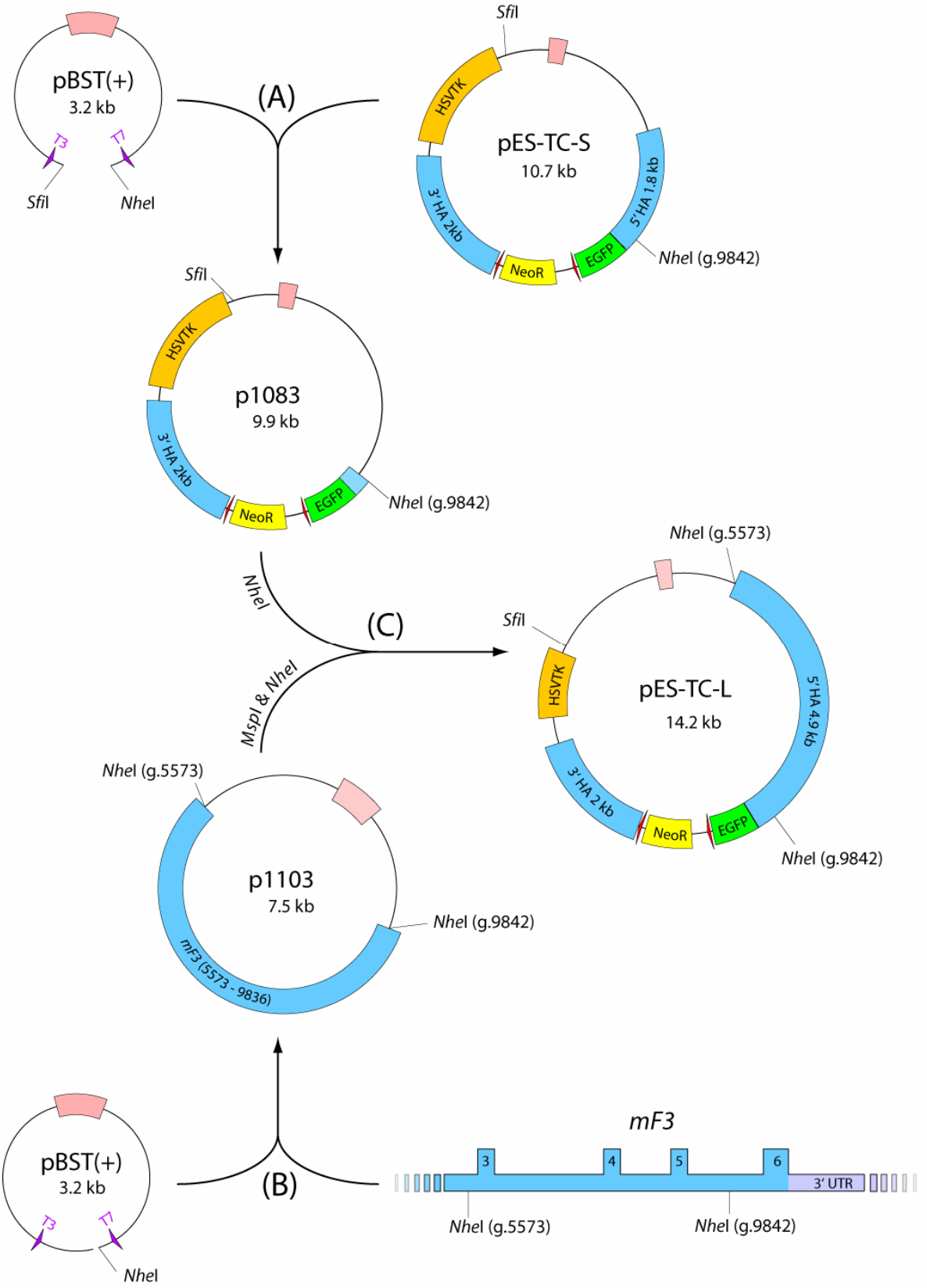
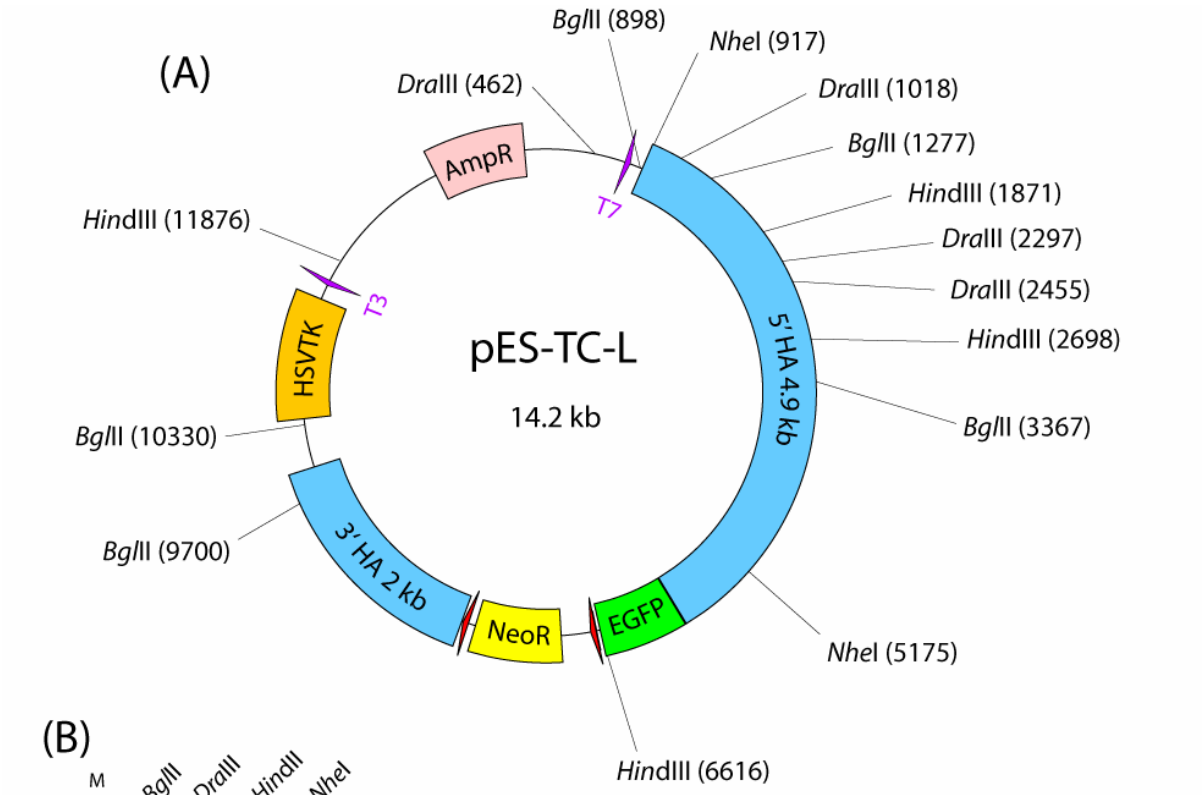
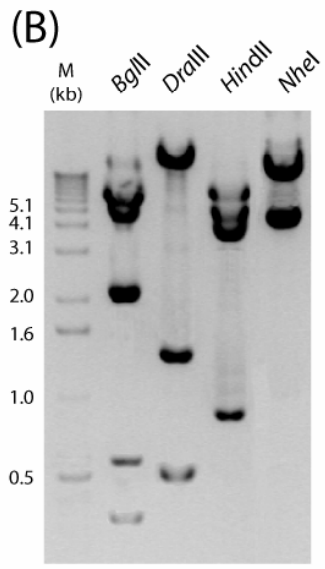
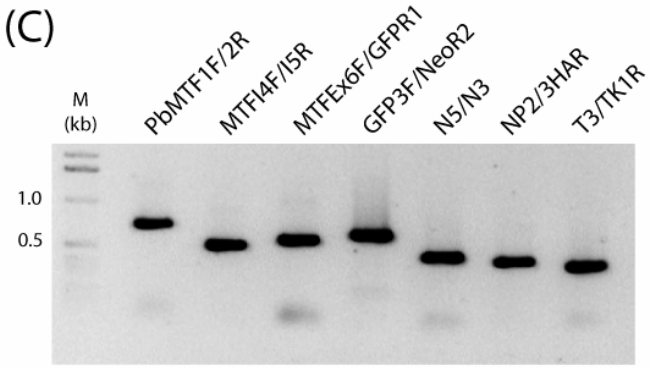


Figure 3.8 The generation of pES-TC-L, with a longer 5' HA, from pES-TC-S. (A) The *Sfi*I – *Nhe*I fragment from pES-TC-S is subcloned into pBST(+) to make p1083. (B) The appropriate *Nhe*I fragment from *mF3* is cloned in to pBST to make p1103. (C) p1083 is digested with *Nhe*I and p1103 with *Nhe*I and *Msp*I. The fragments are ligated to create pES-TC-L.







Enzyme	Fragment Sizes (kb)
*Bgl*II	0.4, 0.6, 2.1, 4.7, 6.3
*Dra*III	0.2, 0.5, 1.3, 12.2
*Hin*dIII	0.8, 3.9, 4.2, 5.3
*Nhe*I	4.3, 9.9

Figure 3.9 Map of pES-TC-L targeting vector with extended 5' HA. (A) Position of features and restriction sites. (B) Restriction fragments produced following digestion with selected endonucleases. The sizes expected from the predicted sequence are shown in the table. (C) PCRs using primer pairs positioned at regular intervals in the construct was used to confirm that the construct had been correctly generated.

3.6.4 Verification of the extended 5'HA sequence

The importance of using isogenic DNA in the HAs has already been stated and even single nucleotide mismatches reduce the efficiency of HR. The 3' HA had been previously sequenced and shown to have complete sequence identity with the reference sequence . The 5' HA was sequenced in two pES-TC-L clones, which showed the correct patterns on the diagnostic restriction digests, using the primer pairs detailed in Appendix 3. The same amplicons (with gene specific primers to replace the primers in the vector sequence) were generated for sequencing from genomic EL M3 DNA. The sequences were assembled and there were no differences in the 5'HA sequence between the clone DNA and ES cell genomic DNA. The consensus sequence from EL M3 cells showed the following variations when compared with the C57BL/6J reference sequence:

1. g.5975_5982 del TCTCTCTC ins ACAC
2. g.9268_9269 ins TG
3. g.9291 T>C
4. g.10208 G>A

An alignment of the EL M3 and reference sequences is shown in Appendix 4. As EL M3 cells are derived from the 129/Sv strain these changes may represent polymorphic variation between strains. There is no published 129/Sv reference sequence for comparison. The first two polymorphisms are caused by deletions or insertions in dinucleotide repeat elements. The third polymorphism is a nucleotide substitution. All of the variations are intronic and distant from splice junctions. Therefore no changes would be expected in the coding sequence as a result of these polymorphisms.

During the final step in the creation of pES-TC-L (Figure 3.8C) 95% of clones generated showed abnormalities in either the diagnostic restriction digests or PCR products. An explanation for this may be seen on examination of the 5' HA genomic sequence. There are several regions of dinucleotide repeats in the 5' HA: TC repeats at g.5935_5982 and g.6420_6449, AC repeats at g.5983_6024 and g.6925_6942, TG repeats at g.6450_6469 and g.9225_9268. These can be seen in Appendix 4. HR is a much more efficient process in bacteria than in mammalian cells. This is discussed in greater detail in the next chapter. As little as approximately 20 bp of identical sequence can be sufficient for HR in bacteria. Sequencing of some of these clones

showed a high frequency of recombination events within the extended 5′ HA. In the majority of cases recombination occurred between the intronic dinucleotide repeat elements listed.

3.7 Targeting with pES-TC-L

Table 3.1 demonstrates that more G418 and ganciclovir resistant colonies were obtained with EL M3 cells than with the other cell lines. Therefore targeting experiments with the longer targeting construct pES-TC-L were carried out using EL M3 cells only. pES-TC-L was linearized by digestion with *Sfi*I and 25 µg were used to electroporate 8 x 10^6 ES cells. Two separate targeting experiments were carried out with this combination.

ES cell type	Total ES cells electroporated	$G418^R$ & $Ganc^R$ Colonies	Colonies screened	NHR clones	HR Clones
EL M3	1.6 x 10^7	268	255	251	4

Table 3.2 Results of two electroporations using pES-TC-L. $G418^R$ & $Ganc^R$ refers to G418 and ganciclovir resistance. All colonies were screened by a combination of PCR and Southern blot. NHR = non-homologous recombinant, HR = homologous recombinant clones

Four HR clones were identified after targeting with pES-TC-L (Table 3.2). Conventionally the *absolute targeting frequency* is defined as the ratio of the number of independent HR targeting events to the number of cells electroporated, while the *relative targeting frequency* is the ratio of HR events to NHR events (Capecchi, 1989c). Using these criteria gives an absolute targeting frequency of 1:4 x 10^6 and a relative targeting frequency of 1:63. Representative Southern blot and PCR screening images demonstrating the presence of targeted clones are shown in Figure 3.10.

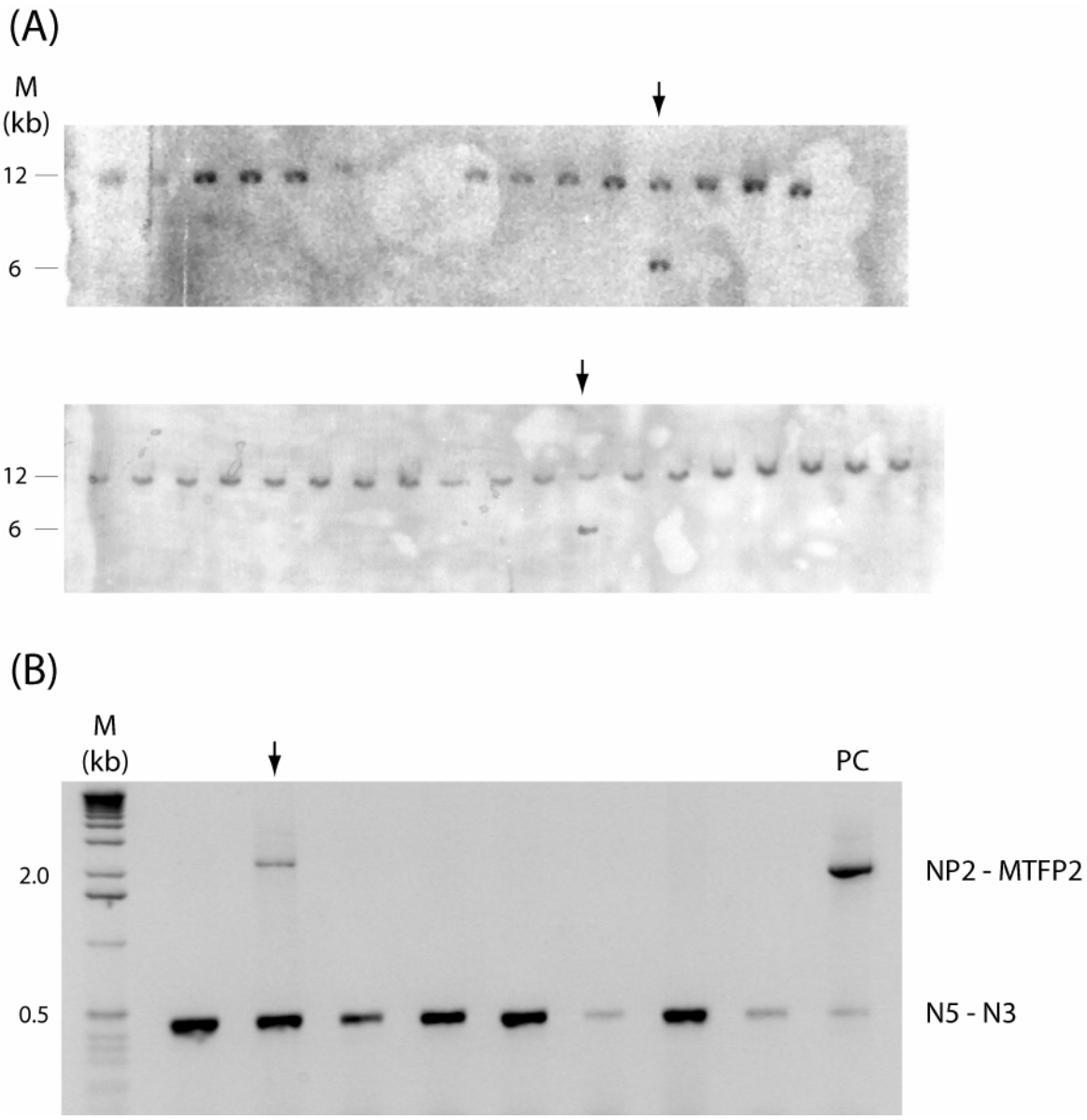


Figure 3.10 Representative images showing the results of screening of EL M3 clones targeted with pES-TC-L. Correctly targeted clones are arrowed. (A) Images of gels generated during Southern blot screening taken using a phosphorimager. The wild-type alleles produce a band of 12.2 kb. A correctly targeted clone has an additional band of 5.9 kb. (B) Images of gels generated during PCR screening taken using a UV trans-illuminator. PC represents the positive control. The control reaction, amplified by the primer pair N5 – N3, produces a 0.4 kb band in all clones. The HR-specific reaction, amplified by primer pair NP2 – MTFP2, produces a 2.4 kb band in correctly targeted clones only. The primer pairs are shown on the right of the gel, level with the corresponding amplicon.

3.8 Blastocyst injections with targeted EL M3 clones

The four correctly-targeted EL M3 clones were expanded in preparation for use in injections into C57BL/6J blastocysts. One clone failed to grow after thawing and was discarded. All cells were injected at passage 19. Metaphase spreads from the three remaining clones, numbered 17, 33 and 109, were analysed prior to injections. In all cases this showed a greater than 80% normal chromosome number. A total of seven separate blastocyst injection and re-implantation experiments were carried out as detailed in Table 3.3.

Injection	Clone	ES cells per blastocyst	Blastocysts per female	Blastocysts returned	Pregnant females	Pups born	Chimaeric pups	% agouti coat colour in chimaeras
1	33	10-20	9	27	2	0	0	-
2	33	15-20	8	24	1	0	0	-
3	109	10-18	8	32	3	10	0	-
4	17	7-12	8	48	4	9	2	5% (M) 5% (M)
5	33	8-15	10	60	4	8	2	20% (M) 20% (F)
6	109	12-20	11	22	1	0	0	-
7	109	15-20	8	16	2	4	0	-
Totals				**229**	**17**	**31**	**4**	

Table 3.3 Results of blastocyst injections with targeted EL M3 clones. Between 4-6 pseudopregnant females were used in each experiment and the number becoming pregnant is shown. Pups were counted 1 week after birth. Chimaerism was assessed after 3 weeks and is expressed as the percentage of agouti colouring in the coat. The sex of the chimaeric pup is given in brackets after the percentage

Between 7 and 20 El M3 cells were injected into each blastocyst. 8-11 blastocysts were re-implanted unilaterally into one uterine horn of the pseudopregnant foster mothers. From a total of 229 re-implanted blastocysts, 31 pups were born that survived the first week. Of these 27 had an entirely black coat and 4 showed some agouti colouration. 2 of the chimaeric pups had a very low level of chimaerism by coat colour of 5%, while the other 2 had slightly better levels of 20%. The EL M3 cells are derived from male embryos and therefore a female chimaeric pup is unlikely to have a significant contribution to her germline from the injected cells. From the three male chimaeras the 20% agouti pup was selected for breeding experiments. At the time of writing no pups had been sired.

3.9 Discussion

Targeting the mammalian genome in ES cells remains the preferred method for generating transgenic mice in most settings. However, the results from this chapter demonstrate some of the difficulties inherent in this process.

3.9.1 Design of the targeting constructs

In this project, experiments with the pre-existing targeting construct, pES-TC-S, failed to produce any HR clones after screening 1310 colonies. This construct had been previously used in a targeting project and a single HR clone was identified from 331 screened colonies (Mumford, 2003). Because only one correctly targeted clone was produced it is difficult to draw any firm conclusions from this data regarding the absolute targeting frequency associated with this construct.

The detection of large numbers of NHR clones in the targeting experiments using pES-TC-S indicated two things. Firstly, that the linearized targeting construct was successfully introduced into cells by the electroporation procedure and secondly, that sufficient numbers of ES cells had survived the procedure. This suggests that the main reason why correctly targeted clones were not obtained was because of inefficiency of the HR mechanism. The rate of HR between a targeting construct and the homologous genomic DNA sequence is dependent on a number factors in any given targeting procedure. Some of these are fixed but among the variable factors, the most easily altered is the length of the HAs.

The key requirements for HAs are now well-established (Ayares *et al.*, 1986; Capecchi, 1989b; Hasty *et al.*, 1991). It is essential that the DNA is isogenic with the target as even single nucleotide mismatches reduce the chance of successful

recombination. This may be ensured by using DNA from the ES cells as the template for the HAs. After this, the size of the HAs is perhaps the most critical feature. The length of homologous sequence between the targeting vector and gene of interest is directly correlated with the efficiency of HR up to at least 10 kb (Hasty *et al.*, 1991; Deng & Capecchi, 1992). In this regard it is the combined length of both arms that is directly related to the rate of HR. Therefore improvements may be seen by extending one side alone. Although the length of the HAs is undoubtedly important the negative impact of having a long non-homologous region has also been well demonstrated. This is of particular importance where the aim of the targeting is to knock-in a gene rather than simply disrupt the targeted locus. Gene knock-outs are generally achieved by replacing a critical part of the gene by a selectable marker. The non-homologous region need only consist of this marker. However, to knock-in a reporter gene requires the presence of additional coding sequence in the non-homologous region. In these constructs it is more accurate to say that it is the ratio of the homologous to the non-homologous region that determines the HR rate.

In order to assess the requirements for targeting at the TF locus it is useful to examine the constructs used in previously reported studies. These are listed in Table 3.4.

Reference / Vector	Size of HAs (kb)	Corresponding regions in *mF3*	Size of NHS (kb)	HAs:NHR	Effect on *mF3*
Bugge *et al.*, 1996	3.7	5′ UTR – IVS 5	3.2	1:1	Exons 1-3 replaced by HPRT
Carmeliet *et al.*, 1996	10.5	5′ UTR – exon 6	1.8	5.8:1	Exon 3 deleted
Toomey *et al.*, 1996	5.9	IVS 2 – 5	1.5	3.9:1	Exon 3 replaced by NeoR
Melis *et al.*, 2001	9.5	IVS 2 – 3′UTR	1.8	5.3:1	Cytoplasmic tail not transcribed
pES-TF-S	3.8	Exon 5 – 3′ UTR	2.0	1.9:1	EGFP knocked in at 3′ end
pES-TF-L	6.9	IVS2 – 3′ UTR	2.0	3.5:1	

Table 3.4 Comparison of the targeting constructs used in this project with those in published studies of targeting at the *mF3* locus. The combined size of both HAs is given together with the region that they correspond to within *mF3*. The size of the non-homologous region sequence (NHS) is given and the ratio of the sizes of the HAs to the NHS. The effect on the targeted locus is shown in the final column.

It can be seen from this table that there are no published studies in which a gene has been knocked-in to the intact *mF3* locus. In three of the previous studies the aim was to knock-out the gene and thus the exact site of integration at the *mF3* locus was not critical. In the study by Melis and colleagues two stop codons were inserted before the cytoplasmic tail. In three of the four reports the homologous to non-homologous ratio was greater than that seen with pES-TC-S and pES-TC-L. It was not possible to obtain any data regarding the absolute or relative targeting frequencies seen in these studies. However, from the size of these constructs it does seem that a relatively high homologous to non-homologous ratio is required to target the *mF3* locus in comparison with successful targeting at most other loci.

3.9.2 Homologous recombination in ES cells

Various strategies have been used to try and directly increase the intrinsic HR potential within a targeted cell. For example there is good evidence that members of the *RAD* family of proteins play a direct role in DSB repair by regulating HR and NHR

(Vasquez *et al.*, 2001). ES cells derived from knockout models of some of these genes have been shown to have enhanced HR potential. Unfortunately, there are additional consequences of interfering with the complex genetic control of chromosomal repair and as yet this remains an area for additional investigation.

As targeting by HR is essentially a harnessing of the cell's DNA repair mechanism another approach is to stimulate this process by controlled DNA damage in the target locus or HAs. This has been achieved by creating HAs with stripping of one strand of the DNA at the ends of the arms. Alternatively breaks may be introduced specifically within the target locus using endonucleases, although this requires initial targeting of the locus with a recognition site motif (Carroll, 2004). A third method of simulating DNA damage involves the use of triplex-forming oligonucleotides which bind to specific complementary regions and stimulate the repair mechanism (Knauert & Glazer, 2001; Kalish *et al.*, 2005). However, the techniques for using these strategies remain to be fully elucidated and as yet they do not offer a clearly-defined improvement over standard targeting procedures.

All of the strategies discussed above are difficult to use in practice. A relatively simple way of altering the targeting efficiency is by modification of the method of DNA delivery. Early attempts at targeting utilised micro-injection. Although this had the benefit of very high targeting efficiency, in that integration occurred in almost every cell injected, it is laborious and requires specialist skills. This is discussed further in the next chapter. Micro-injection has now been superseded by mass DNA delivery methods, predominantly electroporation. The much-reduced targeting efficiency associated with mass delivery methods, is compensated for by the far-greater number of cells exposed to the construct. Efforts to improve the efficiency of transfection, such as the use of Lipofectamine, have generally led to an increase in NHR and are not of proven benefit (reviewed in Vasquez *et al.*, 2001)

The absolute frequency for HR is generally in the region of one targeting event per 10^5-10^7 electroporated cells (Capecchi, 1989a; Vasquez *et al.*, 2001). The HR rate of 1 in 4 x 10^6 electroporated cells seen with pES-TC-L is consistent with this. The main problem in most targeting projects is that NHR occurs at a much higher rate of 1 in 10^2 to 10^4 cells. This would mean that without selection thousands of clones would need to be screened to find a few correctly targeted ones. Positive selection is generally used to reduce the number of clones that need to be screened. However, this has its drawbacks. The gene encoding for the positive selection marker has to be part of the sequence to be inserted. This means that the non-homologous region is

increased in size, thereby negatively impacting upon the rate of HR. In most cases the gene to be targeted is not expressed in ES cells and so the positive selection marker is introduced under the control of its own strong promoter. Expression of the selection marker is therefore independent of the site of integration and positive selection does not help to differentiate between HR and NHR.

In a few cases the sequence inserted allows for expression of the selectable marker to be directed by the targeted loci. The benefit of this is that selection would then distinguish between HR and NHR as it is unlikely that random integration would place the selection marker such that it was directed by an endogenous promoter. However, when this possibility was explored in this project it was found that the level of *mF3-EGFP* expression in ES cells was insufficient to produce detectable fluorescence.

A conventional strategy using positive selection for any integration event coupled with negative selection of NHR clones was used. Although positive selection remains an effective way of limiting the number of clones surviving the electroporation process, negative selection remains of relatively little benefit. Ganciclovir, for example, suffers from the 'bystander effect' which results in toxicity in clones that don't contain the HSVTK gene. Thus HR clones are also sensitive to it, albeit to a lesser extent than NHR clones (personal communication from Dr. A. Porter, Imperial College, London). This appears to be because of some intrinsic kinase activity within ES cells. To allow for this and prevent inhibition of HR clones, Ganciclovir was removed from the culture medium before colonies were picked.

3.9.3 Screening of ES cell clones

The traditional method for screening the large numbers of clones generated in a targeting project is Southern blot analysis. This is a laborious process and is one of the major rate-limiting steps. PCR offers a much more rapid detection method for specific sequences. Therefore a PCR screening method was investigated to assess if it was sufficiently sensitive to be used as the sole method of screening. The major constraint on the PCR screen is that one primer must bind exclusively in the targeting construct and the other exclusively in genomic DNA. Effectively this dictates that the PCR primers are placed as close as possible to, but without overlapping, one of the HAs, usually the shorter of the two. However, the amplicons remain large and in this project measured 2.4 kb when using the primer pair NP2 and MTFP2.

Increasing amplicon lengths above 1 kb raises specific problems with regard to the efficiency of the PCR. Conventional DNA polymerases used in PCR, such as *Taq* polymerase, have a terminal transferase activity which adds nucleotides to extend the amplicon, but lack a 3' to 5' exonuclease activity. Thus if errors occur during nucleotide incorporation, they are not corrected. As the amplicon is extended the number of mismatched nucleotides increases, reducing the stability of the double stranded product and ultimately leading to strand separation and premature detachment of the polymerase. The result is a mixture of variable length amplicons with nucleotide mismatches that may prevent amplicon extension in subsequent rounds. The error rate for *Taq* polymerase is about 8 x 10^{-6} per base pair per cycle (Cline *et al.*, 1996). Put another way this means that in the standard reaction conventionally used to assess fidelity (the *lacI*OZα-Based Fidelity Assay), 16% of 1 kb products will contain errors after 20 cycles.

This problem may be overcome by using a DNA polymerase with additional 3' to 5' exonuclease activity, also described as the 'proofreading' function. With these high fidelity DNA polymerases, nucleotide mis-incorporation is recognised and the proofreading function excises the incorrect nucleotide. This allows the correct nucleotide to be inserted and extension to occur over a longer distance. However, with early high fidelity enzymes, such as *Pfu* polymerase, the proofreading activity resulted in degradation of primers and template before the PCR reaction reached the optimal temperature for amplicon extension. This necessitated the use of 'hot start' methods which are impractical for high throughput, 96-well based, procedures. Advances in high fidelity enzyme technology have negated the need for hot start by blockade of the proofreading activity at low temperatures using thermolabile inhibitors such as antibodies. This ensures that the proofreading function is only released after the first denaturation step in the reaction, as with the Easy-A and Platinum *Taq* enzymes used for long range PCRs in this project.

Another issue with long-range PCR is that the template requirements are more stringent than for short (< 500 bp) amplicons. Not only does the template have to be of good quality with few impurities, but larger amounts of template are required. Both the HR-specific PCR screen and Southern blot analysis with radio-labelled probe were shown to have the necessary sensitivity when good quality genomic DNA was used as a template. However, the performance of both assays was reduced with genomic DNA from ES cell clones. The probable explanation for this is that the high volume method used for DNA extraction in 96-well format is relatively

crude. The DNA produced is likely to have a significant amount of impurities, such as contaminating protein, that might well interfere with the performance of the screening assays.

The effects of having reduced quality ES cell DNA were clearly seen in this project. Although the HR-specific amplicon is much longer than the control amplicon, it was generally of a lower intensity after gel electrophoresis, despite the fact that ethidium bromide staining is directly related to amplicon length. This indicates that the efficiency of the HR-specific reaction was very much less. As the key requirement of the screening test was that it had a very high sensitivity (minimal chance of false negatives) this meant that PCR could not be relied upon as the sole screening strategy and had to be used alongside Southern analysis.

3.9.4 Choice of ES cell lines

ES cells are used in whole-animal targeting projects because of their pluripotency. The derivation of high-quality ES cells, that retain the ability to contribute to a host embryo's germline, remains a procedure that is technically very demanding. The majority of murine ES cell lines used in targeting experiments are derived from the 129/Sv strain. In comparison with other strains these cells retain their pluripotency when introduced into embryos from heterologous strains (reviewed in Ledermann, 2000).

Factors which are important in maintaining ES cell pluripotency and preventing unwanted differentiation include low passage, growth in media that inhibits differentiation and splitting cultures before confluence. Evidence of this was seen in this project, where a tried and tested cell line (E14) had to be abandoned because with increasing passage the cells became both less able to survive electroporation and less able to contribute to the germline. This is because the other main factor is maintenance of cells with a high percentage of normal karyotype. Unfortunately, chromosomal instability increases as ES cell lines are passaged, effectively putting a limit on expansion. Therefore new lines, which are untested in terms of germline potential, have to be freshly derived.

Assessment of karyotype in this project was carried out using the traditional method of arresting cells in metaphase and then counting the number of chromosomes in cells spread on slides. Unfortunately this only identifies gross numerical abnormalities. More minor abnormalities, which can nevertheless significantly reduce a cell line's germline potential, can only be detected by more

detailed chromosome analyses such as g-banding cytogenetics and fluorescent *in situ* hybridisation (FISH) (Russo, 2000). Although these techniques are well-established in humans and are commonplace in clinical practice, mouse cytogenetics has previously lagged behind. However, newer technologies such as spectral karyotyping have been applied to a variety of mammalian cells including mouse (Bayani & Squire, 2001). As these techniques become more widely used it is likely that mouse ES cell karyotyping will become much more accurate in the future.

3.9.5 Blastocyst injections

Although targeted ES cell clones were generated the overall results from a large number of blastocyst injections was poor. 31 out of 229 re-implanted blastocysts produced viable pups, giving a birth rate of 14%. This does not compare well with most targeting projects where the birth rate is generally in the region of 25-50% using 129/Sv derived cells. There are a number of potential explanations for this.

The three cell lines used had different culture and selection conditions despite being derived from the same strain. The most obvious of these was the need to grow 129/1 and EL M3 cells on a layer of fibroblasts while E14 cells grew on gelatin alone. Similarly experience in the ES cell and Transgenic laboratory in this institute has shown that blastocyst injection procedures need to be tailored to each cell line (personal communication from J. Godwin). In particular the optimal number of ES cells that should be injected is quite variable and not simply dependent on the size of the cultured cells. The EL M3 cell line that was successfully targeted had not previously been used for targeting projects in this institute. The most favourable injection conditions were therefore determined empirically by injection with stock cells. However, as stock ES cells have not been subjected to the same experimental conditions as targeted clones, the conditions determined can only be considered a guide. For this reason blastocyst injections were performed with a range of ES cell numbers.

A recurring problem was variability in the number and quality of blastocysts obtained. Even in humans, where extensive research has made *in vitro* fertilisation a widely-used clinical procedure, the results from ovarian hyperstimulation and embryo re-implantation are very variable. In mice the number of blastocysts harvested is highly dependent on the environment. For example small variations in temperature in the animal house can have dramatic effects.

Finally, the technical skill required for these procedures in a small mammal should not be underestimated. Even the best reported results indicate that the majority of re-implanted blastocysts fail to result in viable pups.

3.10 Summary

Attempts to knock-in EGFP at the *mF3* locus using a vector that had been generated during a previous project proved unsuccessful. This was due to the very low rate of HR achievable with the original targeting construct. The vector was reconstructed and a significant improvement in the efficiency of HR was achieved. A number of targeted ES cell clones were generated using the modified vector.

These clones were used in blastocyst injections with the result that a small number of low-level chimaeric mice were produced. Breeding experiments with these chimaeric mice are still in their infancy but have not as yet resulted in generation of mice heterozygous for the *mF3-EGFP* targeted allele.

Chapter 4 - Targeting a bacterial artificial chromosome

4.1 Introduction

In the early 1990s a number of genome sequencing projects were initiated with the intention of identifying all the genes from a particular organism. These projects required cloning vectors capable of accepting and maintaining large fragments of genomic DNA (> 100 kb). Conventional cloning plasmids have a maximum insert capacity of about 10-20 kb and are often maintained in multiple copies in the bacterial host increasing the chance of unwanted recombination events. Although cosmids have an increased insert capacity of 20 – 40 kb, this is smaller than many genes and remains an impractical size for genome sequencing. The problem of insert size was solved following modification of a plasmid to include the *E. coli* F factor. A plasmid containing the F factor can accept DNA fragments of 300 kb or more and is maintained at a low-copy number of 1 – 2 per cell (Shizuya *et al.*, 1992). With further modifications these vectors have now been optimised and used to construct genomic libraries of mouse and human DNA. On account of the size of the insert they are referred to as "artificial chromosomes" and have been cloned in bacterial (BAC) and yeast (YAC) hosts (Osoegawa *et al.*, 2000). YACs are prone to insertion of non-contiguous DNA fragments and are typically less stable than BACs. In addition separating the plasmid DNA from the host genomic DNA is also more difficult in yeasts, making BACs the preferred vectors for sequencing projects (She, 2003).

Although BACs were originally designed with large scale sequencing in mind they are effectively mini-chromosomes, making them also suitable tools for gene targeting projects. In contrast with smaller vectors, BACs can contain the entire target gene with associated cis-acting regulatory elements, thus enabling expression of the target-reporter fusion product to remain under physiological control. The previous chapter discussed the difficulties associated with targeting in mammalian cells and in particular the very low rate of HR in ES cells. In contrast the mechanisms involved in HR in bacteria are much more efficient, so targeting a mammalian region within a bacterial host is relatively straightforward. In fact sequence identity over regions as short as 60 bases is sufficient for HR to occur (Zhang *et al.*, 1998).

Culturing bacteria in the laboratory is much less time-consuming and technically less demanding than culturing ES cells. Coupled with the much greater

targeting efficiency this means that far fewer clones need to be screened and correctly targeted clones can be expanded in a few days as opposed to several weeks. In projects where multiple genes are targeted simultaneously or the same locus is targeted by different constructs, targeting in bacteria can have very significant advantages.

Increased understanding of the proteins involved in bacterial HR and control of their expression has significantly improved targeting of mammalian genes in bacterial hosts (reviewed in Muyrers *et al.*, 2001). In comparisons between wild-type micro-organisms HR in *S. cerevisiae* seems to be more efficient than in *E. coli*. The use of the *red* gene group from bacteriophage λ in place of the endogenous *recBCD* gene cluster in *E. coli* derived from the DH10B strain has increased the rate of recombination in bacteria to make it comparable with that in yeast (Poteete *et al.*, 2002). The two *red* genes *exo* and *bet* control recombination while an additional λ gene, *gam*, abrogates the exonuclease activity of RecBCD. The λ genes are under control of the temperature sensitive repressor λ*cI857*, thus allowing the conditions for efficient HR to be temporarily provided when the culture temperature is raised to 42°C (Lee *et al.*, 2001).

The EL350 bacteria used during this part of the project have been specifically modified to optimise the hosting of BACs. They are derived from the DH10B strain and in addition to containing the *red* phage λ genes they also contain the *cre* gene under control of the P_{BAD} promoter which is only functional in the presence of arabinose. Therefore in this strain HR may be induced by increasing the temperature and cre-mediated recombination is induced by adding arabinose to the culture medium (Lee *et al.*, 2001).

Targeting mammalian genes within a BAC in a bacterial host is increasingly viewed as an alternative to targeting directly in mammalian cells. However, this option does have some drawbacks. The most obvious of these is that having carried out the targeting procedure in bacteria, the BAC containing the correctly targeted gene then has to be introduced into the mammalian genome. The method typically used requires injection of carefully prepared BAC DNA into the male pro-nucleus of fertilized ova. Not only is this a technically demanding procedure, but also the BAC DNA has to be handled carefully, as it is easily sheared on account of its large size. Following injection the BAC DNA is randomly integrated into the host genome. Although BACs are chosen to contain large regions of flanking sequence around the gene of interest, this nevertheless raises the possibility that elements adjacent to the

integration sites may influence expression of the targeted gene. In addition shearing of the BAC may lead to loss of critical regulatory elements and abnormalities of gene expression.

Furthermore, the endogenous alleles remain unaltered within the host genome. In tightly regulated genes with feedback pathways controlling the total amount of translated product, this may lead to competition with the newly inserted targeted gene. Although this effect would also occur following targeting in ES cells, it is more of an issue when a single transgene competes with two endogenous genes. It is therefore particularly important to ensure that the targeting does not put the modified gene at a disadvantage to the wild-type gene.

4.2 Identification of appropriate BACs

The National Centre for Biotechnology Information (NCBI) Clone Registry at http://www.ncbi.nlm.nih.gov/genome/clone/index.html was interrogated in July 2004 to find BAC clones spanning the *mF3* locus on chromosome 3. A number of clones were identified of which one had been fully sequenced (RP24-404P10, accession no. AC129311.6). A further 11 clones had been mapped to the same region by alignment of their end sequences as demonstrated in Figure 4.1. The generation of BAC clones and features of the vectors used has been previously described (Osoegawa *et al.*, 2000; Zeng *et al.*, 2001).

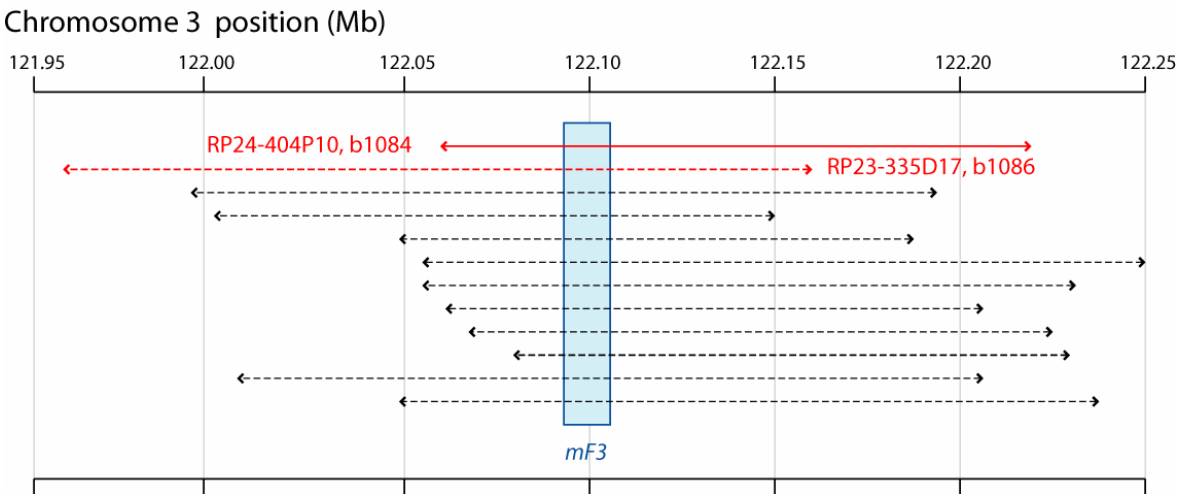


Figure 4.1 Map view of part of chromosome 3 showing the position of the *mF3* locus (blue rectangle) and BAC clones listed in the NCBI clone registry. The solid line represents the single fully sequenced clone. Dashed lines indicate clones sequenced at the ends only. The two clones selected for this project are shown in red.

RP24-404P10 was selected for targeting as the availability of the full sequence allows accurate restriction mapping and PCR amplification. An additional clone, RP23-335D17, was selected as this contained the longest 5′ flanking region upstream of *mF3* of the available BACs. At the time of selection the full sequence for clone RP23-335D17 had not been published but the clone's position in the mouse genome had been verified by sequencing across the ends of the genomic fragment.

Although the promoter region of *mF3* has been characterised (Mackman *et al.*, 1992), the mechanisms controlling expression of the gene remain to be fully elucidated. With some genes cis-acting regulatory elements may be present many kilobases upstream of the promoter, or less commonly, in the 3′ flanking region. The binding sites for transcription factors and elements mediating responsiveness to stimuli such as lipopolysaccharide (LPS) are contained in the -227 to -1 bp sequence immediately before the start of transcription in *mF3*. However, the possibility of unknown regulatory elements further upstream of the gene remains. For this reason the selected BAC clones contained a large amount of sequence upstream of *mF3*. The 5′ flanking regions of the clones are 42 kb for RP24-404P10 and 141 kb for RP23-335D17.

Mapping of the clones in chromosome 3 showed that RP23-335D17 contains no known genes other than *mF3*. RP24-404P10 contains one additional gene: *Abcd3*. The protein encoded by this gene is a member of the superfamily of ATP-binding cassette (ABC) transporters. *Abcd3* is transcribed in the reverse direction to *mF3* and there is no evidence that transcription or regulation of either gene has any effect on the other. *Abcd3* is 60 kb in length and the 3′ ends of *mF3* and *Abcd3* are 23 kb apart.

The selected clones were obtained from the BACPAC Resource Centre, Children's Hospital Oakland Research Institute, California, USA (http://bacpac.chori.org/). The clones were designated as b1084 and b1086 in the laboratory. Vector maps and specific features of the selected BACs are shown in Figure 4.2.

In order to demonstrate that the clones contained the *mF3* sequence BACs were analysed by PCR using four sets of primers designed to generate amplicons at various positions spanning the *mF3* locus. This produced amplicons of the same size in the BACs as in mouse genomic DNA and sequencing confirmed their identity (Figure 4.3). The presence of the expected flanking regions was confirmed by restriction digest with *Xho*I, *Not*I and *Sfi*I which demonstrated bands of the same size as that predicted from the genomic reference sequence (Figure 4.3).

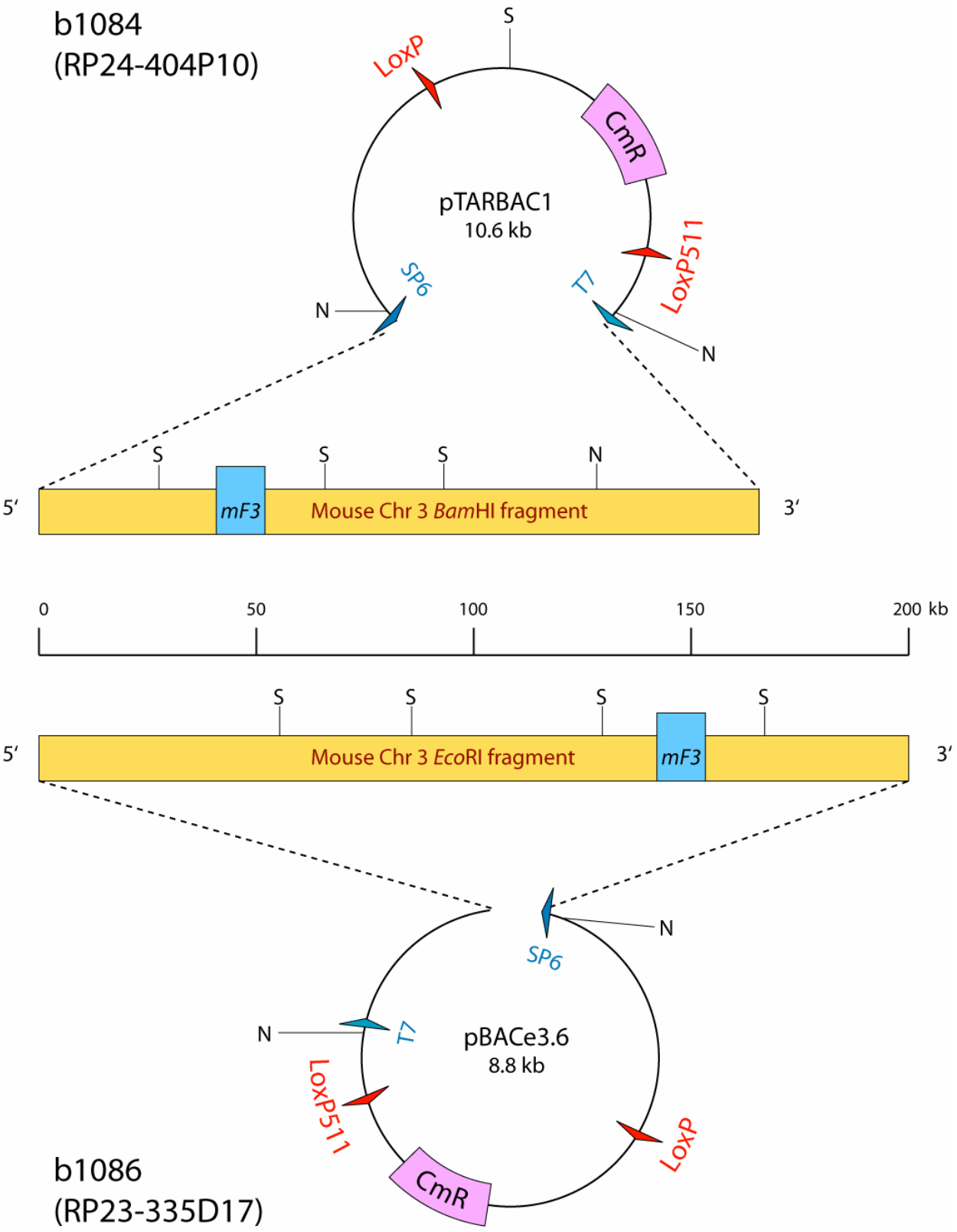
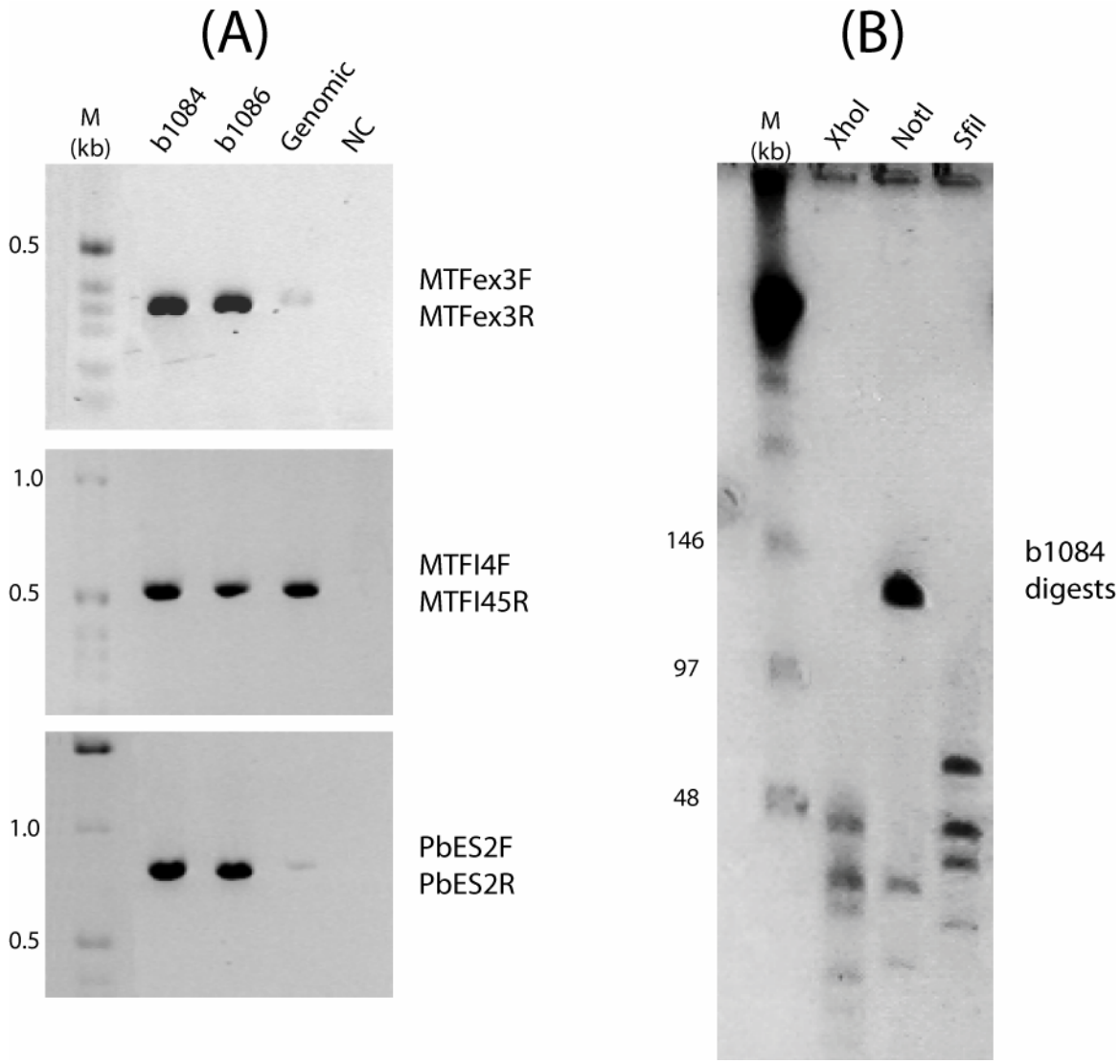


Figure 4.2 Vector maps of the BAC clones b1084 and b1086. Partial digests of mouse genomic DNA (yellow rectangles) were ligated into vectors modified to accept large inserts. The inserts are drawn according to the scale shown. The position of *mF3* in the inserts is indicated by a blue box. The position of *Not*I (N) and *Sfi*I (S) sites are shown. CmR = chloramphenicol resistance cassette. LoxP / LoxP511 = wild-type and mutant LoxP sites.



Enzyme	Predicted Bands (kb)
*Xho*I	10, 12, 16, 18, 25, 42
*Not*I	11, 27, 134
*Sfi*I	21, 34, 49, 67

Figure 4.3 Verification of BAC clones by PCR and restriction digest. A) PCR reactions using the primer pairs indicated to the right of each gel. Genomic – mouse genomic DNA, NC – no template control. B) Representative pulsed field electrophoresis gel showing restriction fragments from b1084 with 3 enzymes. The table shows the sizes of predicted bands above 10 kb.

4.3 Modification of the targeting construct

The targeting construct described in the previous chapter utilised expression of aminoglycoside-3-phosphotransferase (NeoR) as a positive selection marker for recombination and insertion of the construct. Although this gene is of bacterial origin the cassette containing the gene has a eukaryotic promoter in place of the native promoter to allow expression in mammalian cells. The construct therefore had to be modified to include a prokaryotic selection marker that would be expressed in bacteria.

The short mammalian targeting construct, pES-TC-S, was modified for BAC targeting as the lengths of homologous sequences required for successful HR in bacteria are much shorter than for mammalian cells. Using the longer targeting construct, pES-TC-L, might reduce the rate of HR in this case as the homologous arms are not from isogenic DNA and the polymorphic variations noted between the El M3 and C57BL/6J sequences might reduce the efficiency of recombination.

pES-TC-S was transformed into EL350 cells. The *cre* gene was induced by the addition of arabinose, which results in removal of the NeoR cassette, leaving a single loxP site. Loss of this cassette was confirmed by PCR using primers GFP3F and 3HAR. The resulting intermediate plasmid was termed p1096 (Figure 4.4). pRetroNeo was kindly provided by Dr. C. Huxley, Imperial College Faculty of Medicine, London. This plasmid contains a NeoR cassette with a eukaryotic promoter and spectinomycin resistance cassette (SpecR) with the native prokaryotic promoter. The origin of replication comes from the R6K plasmid and is only functional after expression of the π protein. This protein is a product of the *pir* gene which is not present in most bacteria derived from *E. coli*, including the DH10B-derived EL350 strain (Magin-Lachmann *et al.*, 2003). pRetroNeo was grown and purified from π-expressing PIR1 cells (Invitrogen) and electroporated into the EL350 cells containing p1096. After *cre* induction bacteria were selected by spectinomycin resistance. In this situation spectinomycin resistance is indicative of successful insertion at the loxP site in p1096 (as pRetroNeo cannot replicate in EL350) and the insert is therefore lost without recombination. The new targeting construct is further referred to as pBAC-TC. The orientation and integrity of the newly inserted NeoR – SpecR cassette was confirmed by PCR and sequencing.

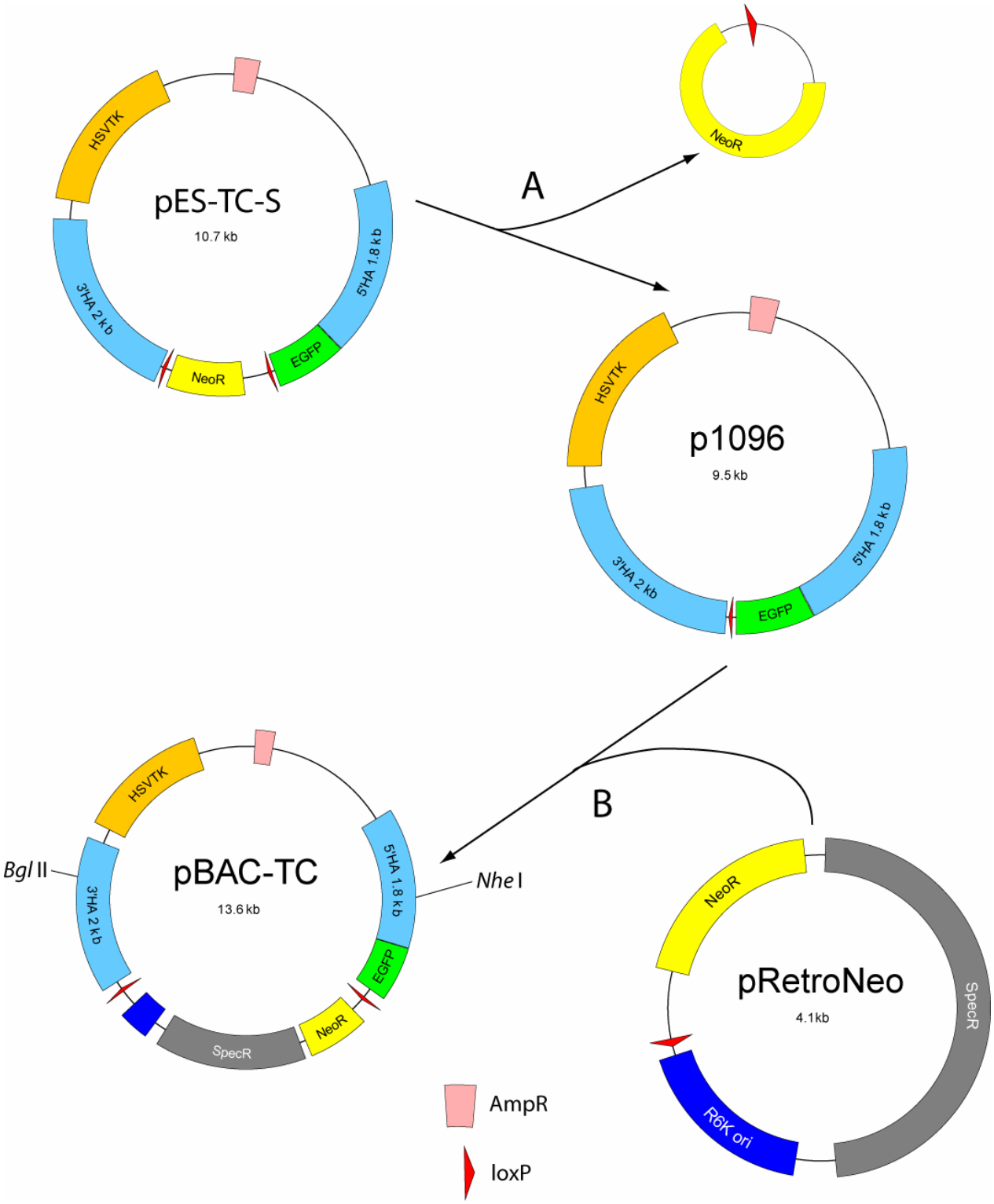


Figure 4.4 Modification of the mammalian targeting construct for use in bacteria in two steps mediated by cre recombinase in EL350 cells. A) Recombination between the loxP sites in pES-TC-S results in excision of the NeoR cassette leaving a single loxP site in the intermediate plasmid, p1096. B) Recombination between the loxP sites in p1096 and pRetroNeo results in union of the two plasmids. This generates pBAC-TC containing a spectinomycin resistance cassette (SpecR) with a prokaryotic promoter.

4.4 Generation of targeted BACs

4.4.1 Targeting strategy in bacteria

Targeting constructs for mammalian cells are designed to ensure maximum length in the homologous arms after linearization. For prokaryotic targeting it is important to reduce the amount of bacterial sequence present in the inserted DNA. This is because the efficiency of HR is so much greater in bacteria than in mammalian cells that any vector sequence present is likely to recombine with homologous sequences in the bacterial genome. For this reason the targeting construct is generally linearized by cutting within the homologous arms. HR at the required locus requires a minimum of about 20 bp, but the rate of HR increases exponentially up to about 75 bp (Watt *et al.*, 1985). Extension of the homologous sequence beyond this size produces less marked increases in the rate of HR. However, the importance of sequence identity remains paramount as it is in mammalian cells and even single nucleotide mismatches in longer HAs significantly reduce the efficiency of HR.

The strategy for targeting *mF3* in BACs is shown in Figure 4.5. pBAC-TC was digested with *Bgl*II and *Nhe*I at the sites shown in Figure 4.4 and the 7.5 kb region between the two cut sites purified by gel electrophoresis and extraction. As both digest sites are within the HAs this reduces their length to 680 bp for the 5'HA and 1818 bp for the 3'HA.

The original host for the BACs was the DH10B *E. coli* strain. BAC DNA was purified and electroporated into the EL350 strain for targeting. EL350 cells were cultured overnight and expression of *exo, bet* and *gam* was induced by raising the temperature to 42°C for 15 minutes. The *Bgl*II – *Nhe*I fragment from pBAC-TC was then introduced by electroporation. In a similar manner to that previously described in ES cells, HR at the *mF3* locus leads to insertion of EGFP at the 3' end of the gene. The NeoR – SpecR cassette with its own promoter is inserted downstream of EGFP and provides the positive selection marker. Bacteria were then spread on spectinomycin-containing agar plates. In bacteria the rate of NHR is not likely to be significantly greater than the rate of HR; therefore negative selection was not used.

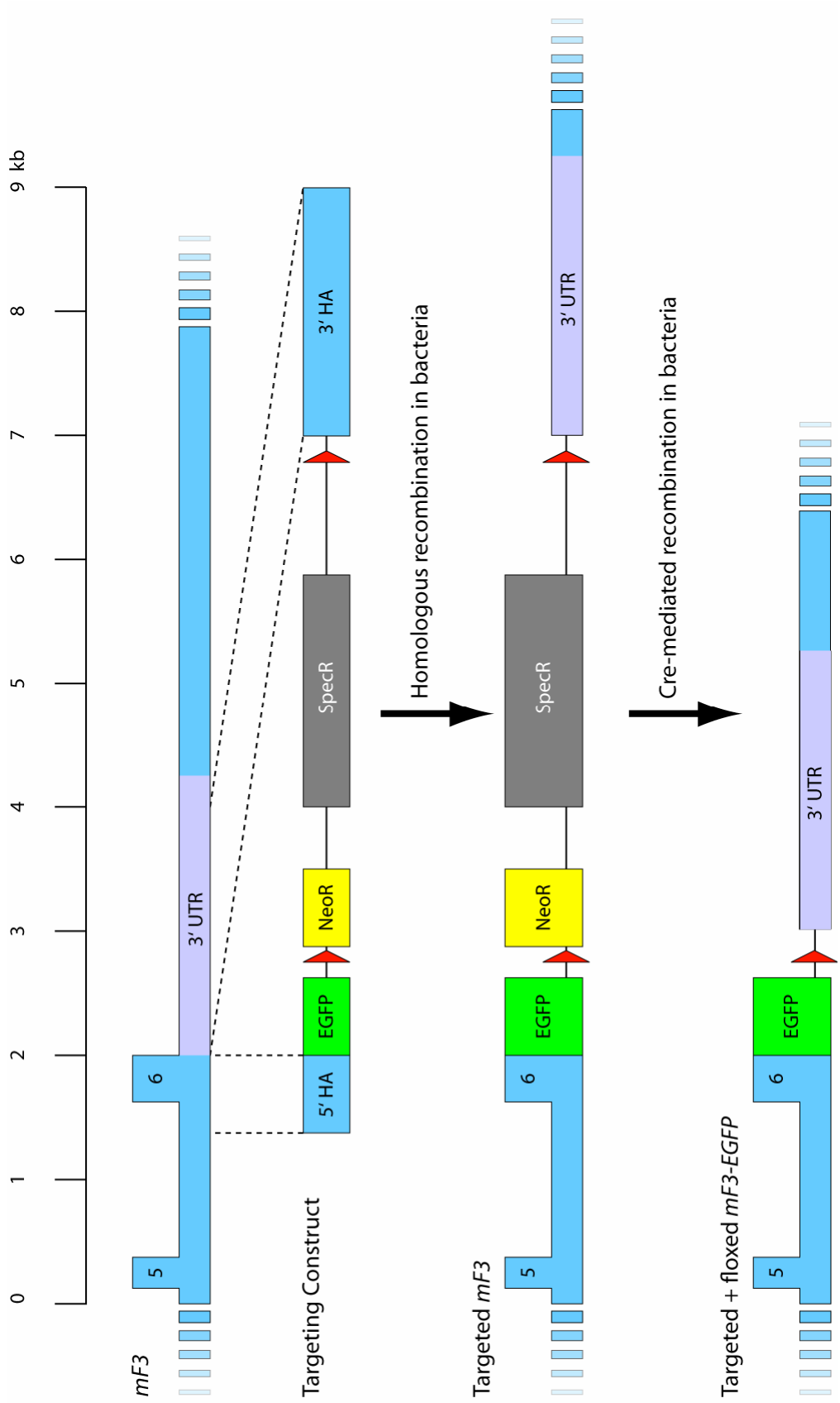


Figure 4.5 Targeting strategy for insertion of EGFP at the *mF3* locus in BACs. Exons in *mF3* are shown by blue numbered blocks. Targeting construct represents the 7.5 kb *BglII* – *NheI* fragment from pBAC-TC. The positions where HAs map to *mF3* are shown by dashed lines. After insertion by HR and floxing with cre recombinase, EGFP has been knocked in between the end of exon 6 and the intact 3' UTR.

4.4.2 Screening of BAC clones

PCR amplification of DNA mini-prepped from spectinomycin-resistant clones was used as the primary method for assessing correct integration of the targeting construct. As with screening of ES cells the amplicon sequence ran from outside the HAs to within the inserted sequence. The primers used for screening ES cell clones could not be used for screening in bacteria as replacement of the NeoR cassette with the NeoR – SpecR cassette in the targeting construct is associated with loss of the NP2 binding site. However shortening of the 5'HA allows amplification of an 887 bp HR-specific amplicon using MTFI5F, which binds upstream of the 5'HA, and EGFP2R which binds in the EGFP coding sequence (amplicon A in Figure 4.6). As described in the previous chapter the primers N5 and N3 were used to provide a control reaction for quality of the DNA template.

Purification of BAC DNA generates much larger amounts of template than extraction of DNA from ES cells. This enabled the use of a digoxigenin-labelled probe to confirm the results of PCR screening. The restriction digest strategy used for ES cells remains applicable in bacteria. The digestion of wild-type *mF3* with *Bam*HI yields a 12.2 kb fragment reaching from exon 3 to beyond the 3' UTR. After successful HR with the targeting construct two further *Bam*HI sites are introduced reducing the size of the fragment to 5.9 kb (Figure 4.6). Hybridisation to the same probe sequence used for ES cell screening was used to detect the change in the size of the *Bam*HI restriction fragment (Figure 4.6). In correctly targeted ES cells both the 12.2 and 5.9 kb fragments were detected (Figure 4.7B), as it is very unlikely that HR would occur in both alleles. As there is just a single allele in the bacterial host only the 5.9 kb fragment is seen after correct targeting.

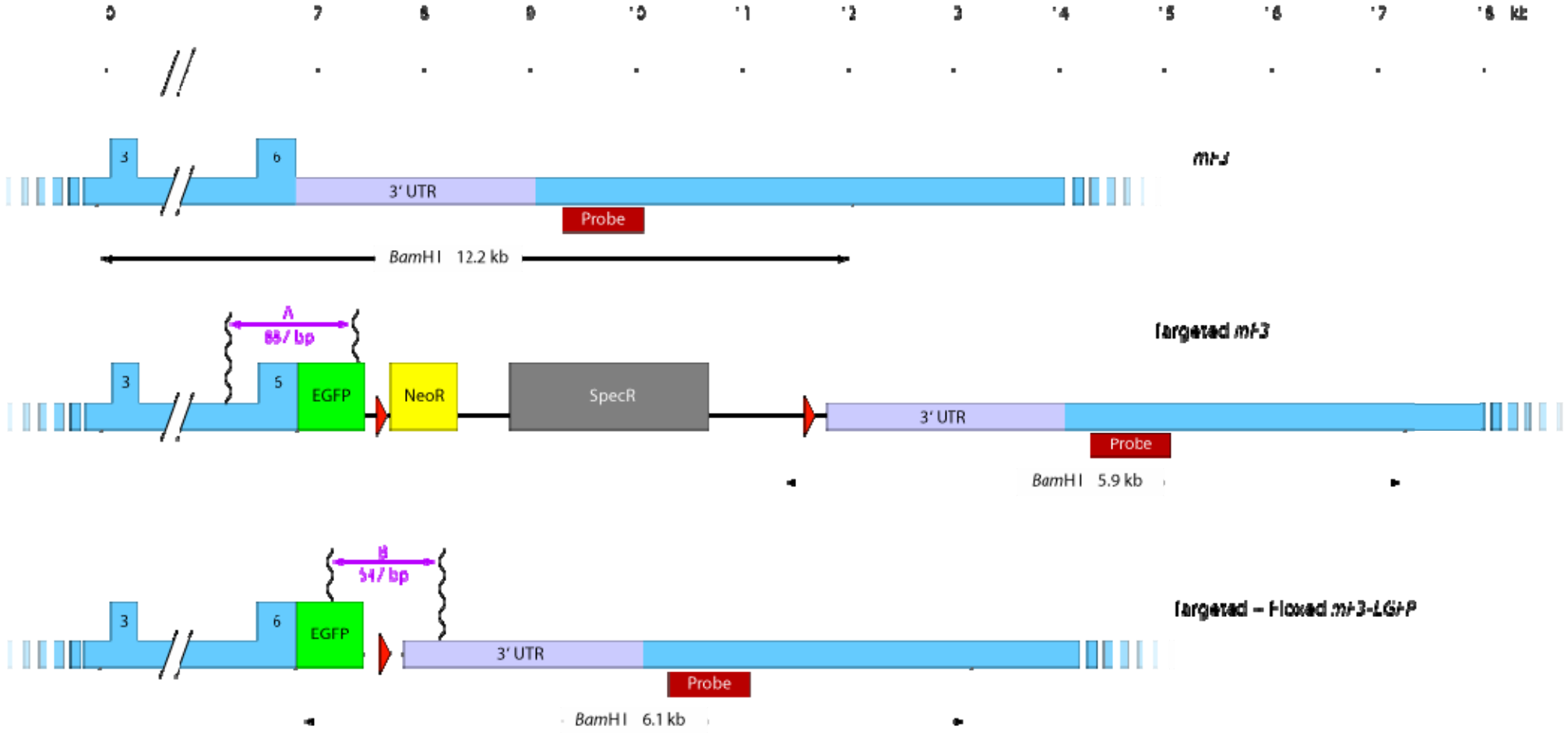


Figure 4.6 Strategies for screening for HR and cre-mediated recombination in BACs. Dashed lines indicate *Bam*HI restriction sites. Vertical wavy lines indicate binding sites for the primers used in the PCR screen. The location of the probe used in Southern blot is shown by a red rectangle. PCR amplicons are shown by purple lines with arrows at both ends. HR is detected by PCR with MTFI5F and EGFP2R which produces amplicon A. Floxing out of the NeoR – SpecR cassette is detected by PCR using GFP3F and 3HAR which produces amplicon B.

4.4.3 Results of BAC targeting

To assess the efficiency of the targeting strategy, BAC culture temperatures were either briefly elevated to 42°C to induce HR or kept at 32°C. The sample kept at the lower temperature gave an indication of the background rate of targeting construct insertion via non-homologous mechanisms.

Thus two cultures were grown on spectinomycin-containing plates for each of b1084 and b1086. Eight spectinomycin-resistant colonies from each plate were screened by PCR for correct insertion of the targeting construct, except in the control plate from b1086 where only 5 colonies were spectinomycin resistant (Table 4.1). The proportion of positive clones in the HR-induced cultures was 4/8 with b1084 and 7/8 with b1086 (Figure 4.7A). No correctly targeted colonies were detected in the control plates. Induction of HR increased the number of spectinomycin-resistant colonies above background by 10-fold in bacteria containing b1086 and almost 3-fold in those containing b1084 (Table 4.1).

BAC	HR induced at 42°C		Control kept at 32°C	
	SpecR colonies	PCR positive	SpecR colonies	PCR positive
b1084	200	4/8	70	0/8
b1086	50	7/8	5	0/5

Table 4.1 Results of targeting of BAC clones with pBAC-TC. The number of spectinomycin resistant (SpecR) colonies obtained from a single plate is shown. Bacterial cultures were either incubated at 42°C to induce HR or kept at 32°C in the control. Colonies were screened by PCR and the proportion of positives is shown.

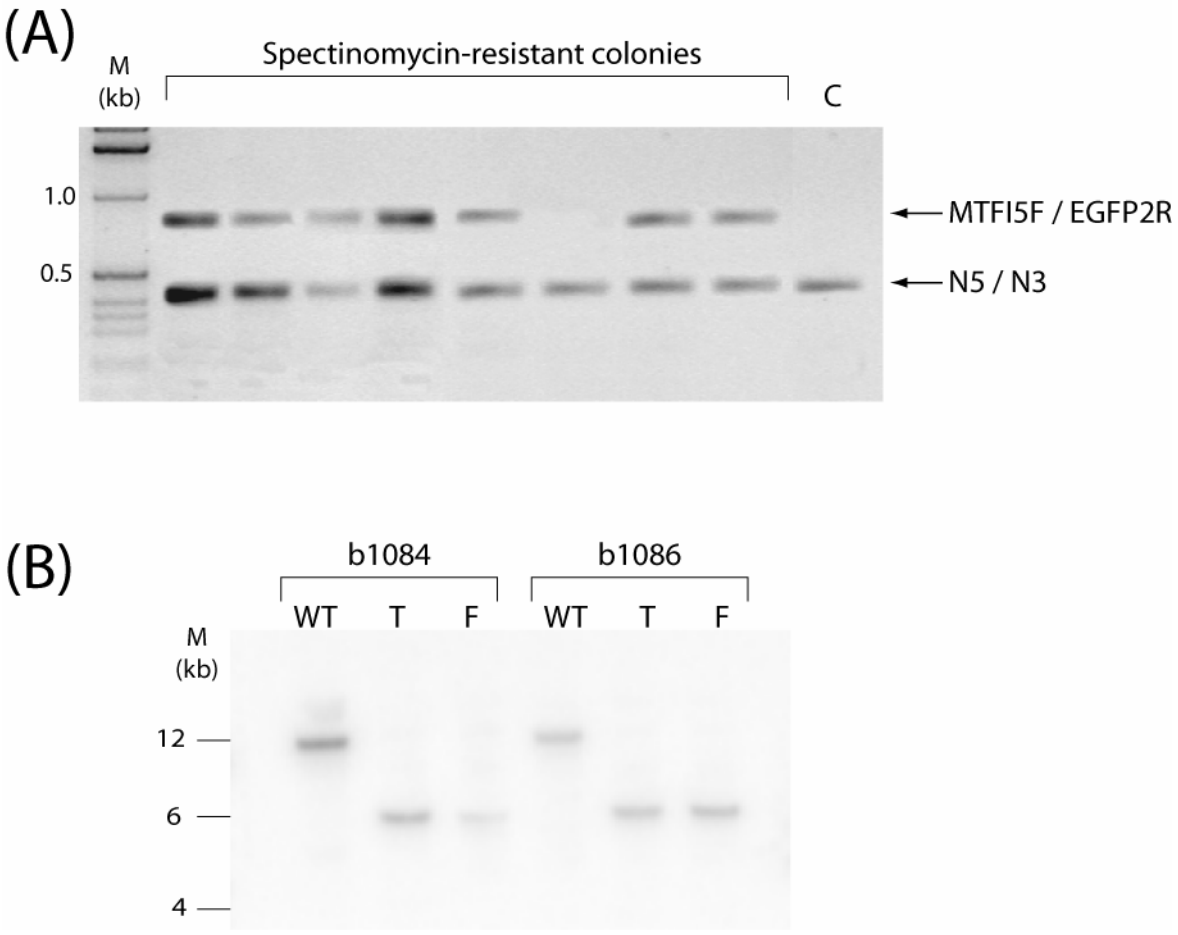


Figure 4.7 Representative images showing the results of screening of BAC colonies. A) PCR screen of spectinomycin resistant colonies using the primer pairs shown alongside the corresponding bands. The 887 bp amplicon generated by MTFI5F and EGFP2R is specific for correct targeting. The 443 bp amplicon generated by N5 and N3 confirms the adequacy of the DNA template. The control template (C) is the *Bgl*II – *Nhe*I fragment from pBAC-TC. In this gel 7/8 colonies are PCR positive. B) Hybridisation of digoxigenin-labelled probe to a *mF3* fragment produced by digesting BAC DNA with *Bam*HI. WT = wild-type, T = targeted, F = floxed. The 6 kb difference between wild-type and targeted BACs is easily seen. Targeted and floxed BACs are not distinguishable by size as the difference is only 0.2 kb.

4.4.4 Cre-mediated recombination

It has been well established that the presence of bacterial DNA sequences within a targeted gene can influence the expression of the gene after integration into the mammalian genome. The knocking-in of EGFP into the *mF3* locus results in the insertion of some 4 kb of prokaryotic sequence containing the NeoR – SpecR cassette. Once BAC clones have been selected by spectinomycin resistance, these cassettes are no longer required and it is therefore preferable to remove them via cre-mediated recombination.

The addition of L(+) arabinose to the culture medium induces expression of the *cre* gene and recombination between the two loxP sites in the targeted gene results in removal of the 4 kb region between the sites, leaving a single loxP sequence. After induction with arabinose, bacteria were grown first on chloramphenicol plates and then colonies were streaked on spectinomycin plates. Efficiency of the cre recombinase step was assessed by comparison with control cultures lacking arabinose. Reacquisition of sensitivity to spectinomycin in chloramphenicol-resistant colonies was suggestive of correct floxing of the targeted gene. This was confirmed by PCR using the primer pair GFP3F and 3HAR (amplicon B in Figure 4.6) which produces a 1.7 kb band in the targeted gene compared with a 0.5 kb band after floxing. The desired targeting and floxing events do not significantly alter the position of *Not*I restriction sites and the final BAC DNA was digested to check that the rest of the sequence had not been altered during these processes.

80 colonies from induced and control plates were streaked to assess spectinomycin sensitivity. Table 4.2 shows that with both BACs there was significant background recombination resulting in loss of spectinomycin resistance. This was more pronounced with b1084 such that the number of colonies losing spectinomycin resistance was only two-fold higher following the addition of arabinose. With b1086 induction of cre recombinase activity caused a four-fold increase in the number of colonies losing spectinomycin resistance.

Eight of the spectinomycin-sensitive colonies were screened by PCR and from the colonies induced with arabinose, three b1084 colonies and five b1086 colonies were positive (Table 4.2). Two b1086 colonies from the control culture were also positive, indicating a significant rate of background recombination in the absence of arabinose. PCR-positive clones were digested with *Not*I. Of the three b1084 clones none displayed the expected digestion pattern. With b1086 three out of five PCR

positive colonies showed the correct digestion pattern. The *Not*I digestion patterns generated by two PCR positive clones from b1084 (F_{1-2}) and two from b1086 (F_{3-4}) are shown in Figure 4.8B as examples. It can be seen that only F_4 has the same digestion pattern as the unmodified BAC which is what one would expect if the clone had been correctly floxed.

Examination of the digestion patterns after the temperature increase step to induce HR showed the same bands present in targeted as in WT BAC as expected. However after these clones were subjected to floxing the pattern of restriction fragments had changed in some, indicating that unexpected recombination events occurred after the addition of arabinose to the culture (Figure 4.8). These events were seen in all b1084 clones following repeated attempts to flox out the NeoR – SpecR cassette. Although such events were also seen in b1086 hosts after floxing, the majority of clones displayed the normal digestion pattern, indicating that only the desired loxP recombination had taken place.

In summary, both correctly targeted clones were produced from b1084 and b1086 with a high rate of efficiency using pBAC-TC. Subsequently cre-loxP-mediated recombination produced unexpected recombination events in all b1084 clones. Therefore only targeted and floxed b1086 clones were used in subsequent experiments.

BAC	With L(+) arabinose			Minus L(+) arabinose		
	SpecS colonies	PCR +ve	Correct *Not*I bands	SpecS colonies	PCR +ve	Correct *Not*I bands
b1084	40	3/8	0/3	20	0/8	-
b1086	65	5/8	3/5	15	2/8	0/2

Table 4.2 Results of induction of cre recombinase. BAC hosts were cultured with or without the addition of L (+) arabinose and the number of colonies that regained sensitivity to spectinomycin (SpecS) is shown. The proportion of PCR positive colonies and of those, the proportion with the correct *Not*I restriction pattern is shown. No correctly floxed clones were produced from using targeted b1084 clones, while 3/5 of targeted clones using b1086 were correctly floxed.

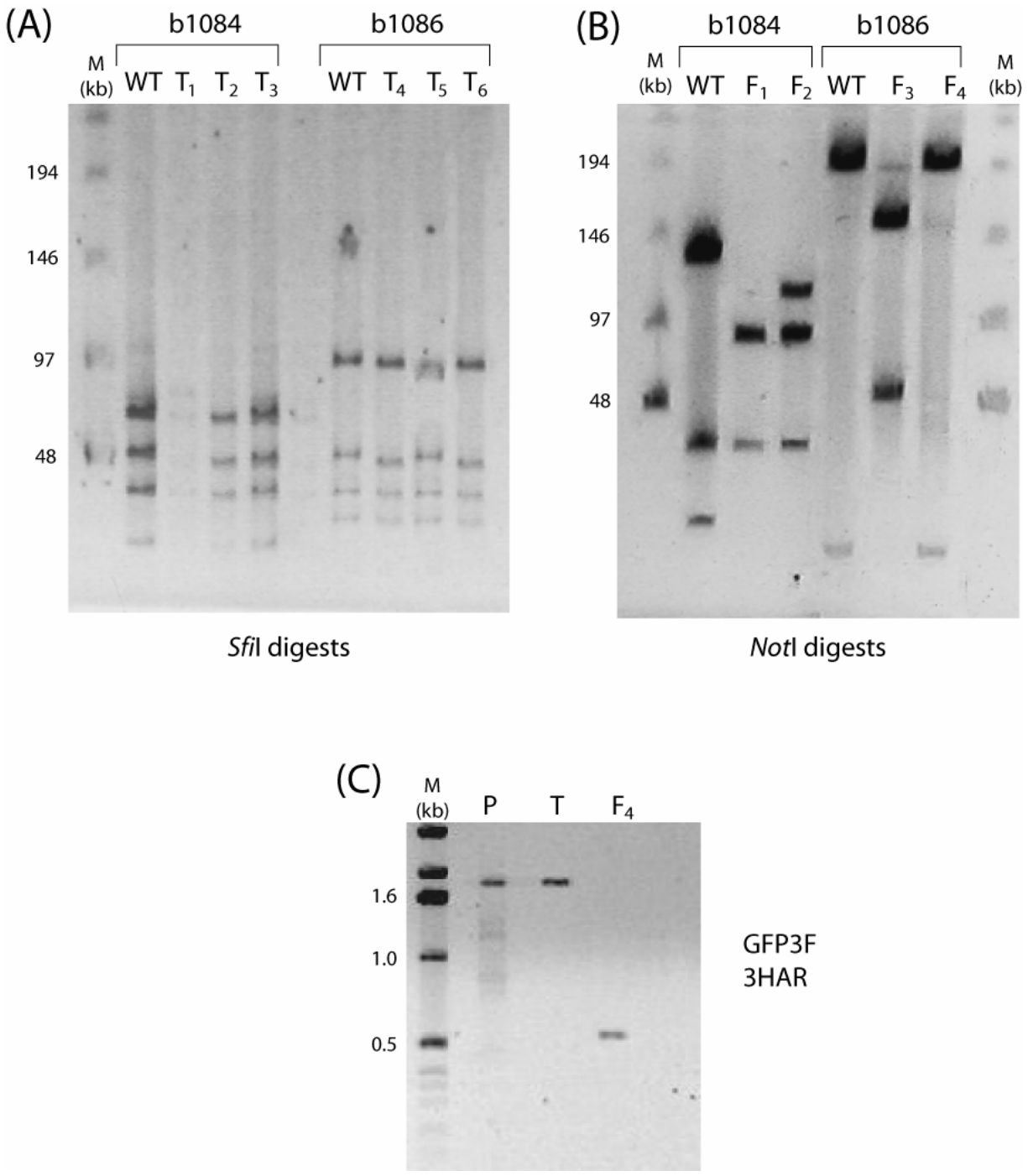


Figure 4.8 Representative images showing digestion patterns (A+B) and PCR bands (C) from BAC clones during targeting. A) Targeted (T_{1-6}) clones digested with *Sfi*I show the same digestion pattern as wild-type (WT). B) After floxing digestion with *Not*I shows a different pattern between WT and two floxed (F_{1-2}) clones from b1084. One of two floxed b1086 clones shows the correct digestion pattern (F_4), while the other does not (F_3). C) PCR using the primer pair GFP3F and 3HAR generates a 1.7 kb amplicon in pBAC-TC (P) and a targeted clone (T), and a 0.5 kb amplicon in the floxed clone (F_4).

4.5 Pronuclear injections with targeted BACs

BAC clones with EGFP targeted to the *mF3* locus and the NeoR – SpecR cassette correctly floxed out were used for pronuclear injections into fertilized embryos. As no b1084 clones were produced without aberrant recombination events only b1086 was used for injections (Table 4.3).

DNA was freshly prepared and the correct restriction digest pattern confirmed before dilution to a concentration of 0.5-2 mg/L. The first 6 injections were carried out using undigested, supercoiled BAC DNA. The ratio of live pups to embryos returned at each injection ranged from 1:4 to 1:15. Although the majority of foster mothers became pregnant following implantation, the number of live pups was significantly less than is normally seen following pronuclear injections with smaller vectors. DNA extracted from the tail clips of pups was assessed for the presence of the transgene using primer pair GFP3F and 3HAR (Figure 4.9). Pooling of the data from all 6 injections with undigested BAC DNA shows that a single transgenic mouse (designated TG17) was produced out of 79 pups.

Injection #	BAC prep	DNA conc. (mg/L)	Embryos returned	Females implanted	Pregnant females	Pups born	Transgenic
1	SC	2.0	60	4	3	9	-
2	SC	1.0	90	6	4	23	1
3	SC	0.5	105	7	6	23	-
4	SC	1.0	60	4	4	6	-
5	SC	1.0	30	2	2	2	-
6	SC	2.0	90	6	4	16	-
		Sub-total	435	29	23	79	1
7	Lin	1.0	105	7	6	26	1
		Total	540	36	29	105	2

Table 4.3 Results of pronuclear injections with b1084. The BAC preparation method is shown: SC = undigested, supercoiled; Lin = digested with *Not*I.

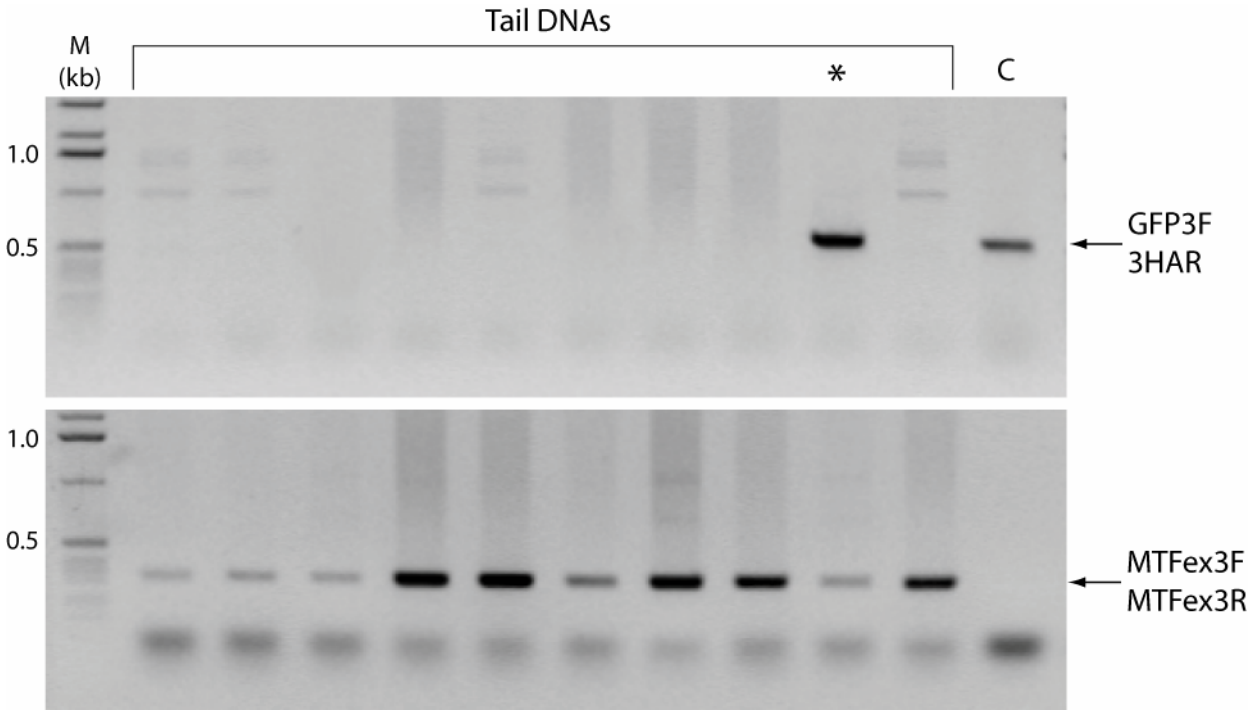


Figure 4.9 Representative gel showing screening of DNA taken from tails of potential transgenic pups. PCR primer pairs are shown to the right of corresponding amplicons. GFP3F – 3HAR is specific for the transgene while MTFex3F – MTFex3R amplifies a sequence in exon 3 of *mF3* and demonstrates the adequacy of the DNA sample. The single transgenic sample (*) shows amplicons with both primer pairs while the remaining samples are only positive with the control primer pair. The control templates (C) were targeted, floxed b1086 for the top panel and a no template primer mix for the bottom panel.

As the first 6 injections were of low efficiency in terms of the number of transgenic animals produced, the 7th injection was carried out using a linear fragment of DNA. b1086 was digested with *Not*I which cuts only in the vector backbone. The BAC was purified by extraction with phenol chloroform and precipitation with ethanol. From the single injection with a linearized BAC, 26 pups were produced of which 1 was transgenic (designated TG254).

The breeding of transgenic mice, determination of the transgene copy number and analysis of the TF-EGFP reporter model in these mice is described in chapter 5.

4.6 Discussion

4.6.1 Targeting selected BACs

Targeting by HR in bacteria was much more efficient than targeting in ES cells. Combining the results from both BAC clones showed that out of 16 clones that survived antibiotic selection, 11 (nearly 70%) had undergone HR. In comparison the best results with ES cells showed that with the larger targeting construct, pES-TC-L, 4 out of 255 selected clones had undergone HR (less than 2%). This meant that far fewer bacterial clones needed to be screened. In addition targeted BAC clones were generated in a matter of days as opposed to the months required for generation of targeted ES cell clones.

Although HR in bacteria proved to be an efficient and easily controllable process, recombination following the induction of cre recombinase produced unexpected results. In order to investigate this, it is first necessary to assess the potential loxP sites present in the targeted BAC (Table 4.4).

Vector	Site	Sequence 5′ – 3′
pBACe3.6	loxP	ATAACTTCGTATA-ATGTATGC-TATACGAAGTTAT
	loxP 511	ATAACTTCGTATA-GTATACAT-TATACGAAGTTAT
pTARBAC1	loxP (mod)	ATAATGGATGCTATACGAGGTTAT
	loxP 511	ATAACTTCGTATA-GTATACAT-TATACGAAGTTAT
pES-TC-S/L & pBAC-TC	loxP 5′	ATAACTTCGTATA-GCATACAT-TATACGAAGTTAT
	loxP 3′	ATAACTTCGTATA-GCATACAT-TATACGAAGTTAT
pRetroNeo	loxP 5′	ATAACTTCGTATA-ATGTATGC-TATACGAAGTTAT
	loxP 3′	ATAACTTCGTATA-ATGTATGC-TATACGAAGTTAT

Table 4.4 LoxP sites in the BAC vectors and plasmids used to generate the targeting constructs. Each loxP site contains two 13 bp complementary sequences separated by an asymmetrical 8 bp spacer (underlined).

LoxP sites consist of two 13 bp sequences that are the reverse compliment of each other. Joined together these sequences would form a palindrome. They are separated by an asymmetrical 8 bp spacer. The entire loxP site therefore contains 34 bp and the direction of the site is defined by the orientation of the spacer. For loxP sites to function as efficient recombination partners, the spacers must have identical sequence in either the same direction or in the reverse complement.

The BAC vector pBACe3.6 contains a wild-type lox P site and a mutant loxP511 site in the plasmid backbone (Figure 4.2). pTARBAC1 also contains a loxP511 site but the wild-type loxP site has been modified and is no longer functional. Figure 4.10 demonstrates the potential Cre-Lox recombinations that might occur in the targeted BAC.

At the start of recombination the loxP sites are arranged such that the sequences are aligned in the same direction. Recombination commences when the strands are broken at the start of the 8 bp spacer. The distal part of the loxP sites, with their attached downstream sequence, then swap position and the sites are reformed. If the loxP sites were in the same orientation at the start of the process, then this positional swap results in loss of the intervening sequence as shown in Figure 4.10B. Although cre recombinase can catalyse both the excision and re-insertion of the intervening sequence, the equilibrium is shifted towards excision because of dissociation of the intervening sequence (Nagy, 2000). In this situation joining of two distinct sequences is only detectable if supported by positive selection. An example of this is the insertion of pRetroNeo into p1096 to form pBAC-TC (Figure 4.4), which requires selection by spectinomycin.

The wild-type loxP site in the pBACe3.6 vector backbone is in the opposite orientation to those brought in by the targeting construct. Recombination between loxP sites in opposite orientation to each other results in inversion of the intervening sequence as shown in Figure 4.10C. As the distance between the loxP sites surrounding the NeoR – SpecR cassette and the vector backbone is approximately 55 kb, this inversion is less likely to occur than the desired excision. Clones with the inversion would produce no product with the primer pair GFP3F – 3HAR primers and are therefore negative in the PCR screen.

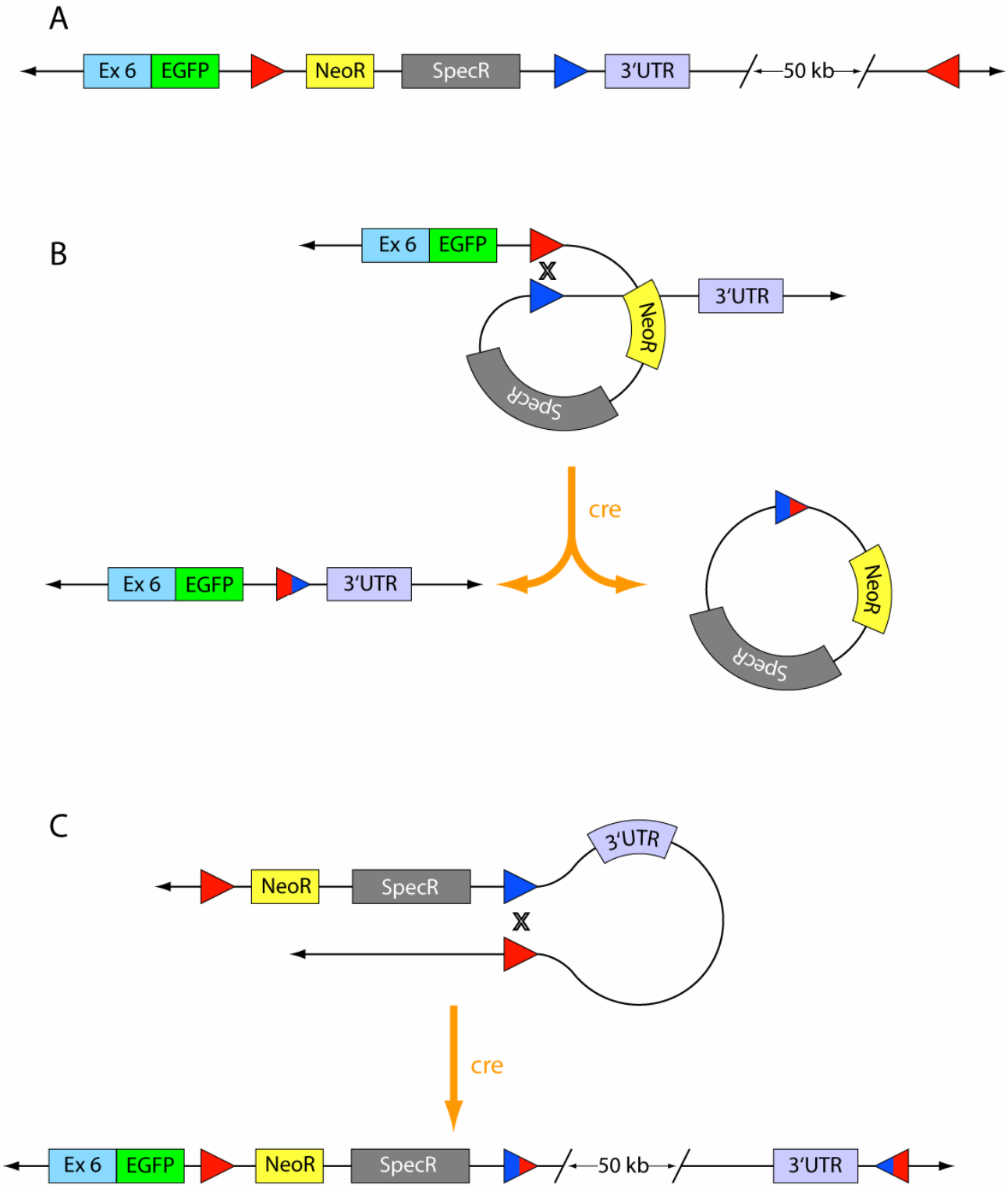


Figure 4.10 Potential Cre-Lox recombination events in the targeted BAC. A) Targeted BAC showing the position and direction of wild-type loxP sites (triangles). B) Recombination between the two same-orientation sites in the inserted sequence results in loss of the intervening sequence. C) Recombination between two sites orientated in the opposite direction results in inversion of the intervening sequence.

There are some reports that mutant sites might also represent potential recombination partners with the wild-type loxP sites in the targeting construct (Siegel *et al.*, 2001). However, a detailed analysis of cre recombinase activity indicates that in most cases recombination only occurs between homologous loxP sites (Lee & Saito, 1998; Branda & Dymecki, 2004; Shakes *et al.*, 2005). Recombination between wild-type and mutant loxP sites (such as loxP511) is much less efficient than that between homotypic sites. This type of recombination occurs at an estimated relative frequency of only 1 - 5% in comparison with that associated with homotypic sites (Branda & Dymecki, 2004). In practice it is only seen if cre recombinase activity occurs at very high levels. This is the probable explanation for the finding of recombination between non-homotypic sites in a few studies. In the procedure used in this project the levels of cre recombinase expression are designed to be moderate, therefore only recombination between the wild-type loxP sites is likely to be seen. Once again non-homotypic loxP recombination would not produce a band in the PCR screen and such clones would not be selected for further study.

Despite this, clones with the correct PCR amplicon subsequently showed an unexpected restriction digest pattern. It is clear that this cannot be explained by aberrant loxP recombination. The likely explanation is that recombination occurred between homologous sequences in the vector backbone and bacterial sequences in the targeting construct or mouse genomic DNA. If this is correct then the frequency of unexpected recombinations should vary between different BACs exposed to the same experimental conditions. This is consistent with the observation that unexpected recombination occurred much more frequently in b1084 than b1086.

It would be surprising if HR was catalysed by cre recombinase directly, as one of the advantages of this recombination system is its high specificity for loxP sites. An alternative possibility is that although the stimulus for expression of the *red* genes is transient, the proteins may persist for sufficient time to promote HR in subsequent steps. One possible way in which this could be addressed in future experiments would be to alter the length of the 42 °C incubation, thereby varying the levels of the expression of the *red* genes. In any case the results indicate that background levels of HR were increased by an unknown mechanism.

As expected, screening of BAC clones proved significantly easier than screening of ES cell clones. Detection of specific sequences by PCR or Southern blot analysis is much more straightforward if the template is present in a BAC background as opposed to the complexity of the entire mammalian genome. The reduced

complexity of the DNA background and the availability of DNA preparations containing higher concentrations of specific template enabled Southern analysis using digoxigenin-labelled rather than radio-labelled probes.

4.6.2 Efficiency of pronuclear injection in generating transgenic mice

Examination of the data from Table 4.3 shows that a total of 540 injected embryos were re-implanted to produce 105 live pups, or that approximately 20% of embryos were successfully carried through pregnancy. Although there are many reports of the generation of transgenic mice using BACs, few studies provide detailed analysis of the efficiency of the microinjection procedure. However, a number of studies have been carried out with embryos from a variety of species to assess comparative embryo survival rates following injection with plasmid DNA. The reported survival rates of 55-80% for mouse embryos *in vitro* is generally lower than that observed for other species such as rat, rabbit and pig (reviewed in Hirabayashi *et al.*, 2001). Although the mouse pronucleus is comparable in size with that of the rabbit and rat and smaller than that of the pig, it appears that mouse embryos are unable to tolerate the volume increases associated with injection as well as those from other species. Of surviving embryos reimplanted into foster mothers 10-25% produce live pups (Brinster *et al.*, 1985). Thus the 20% figure obtained in this project is comparable with other studies.

There are various reasons why only 1 in 5 embryos that appear to have survived the micro-injection and re-implantation process produce live pups. Undetected procedural damage is likely to account for the loss of the majority of embryos. In addition miscarriage might occur for a variety of reasons not necessarily related to the procedure, such as cryptic genetic abnormalities. Furthermore pups may be ingested by the mother before they can be counted. This is presumed to be because of phenotypic abnormalities in the newborns perceived by the mother. However, ingestion can be stimulated by a number of factors, some of which are environmental. The ingestion rate is widely believed to be greater amongst transgenic litters born to foster mothers than naturally produced litters. This seems to be the case even when the transgenic phenotype is indistinguishable from wild-type to researchers. Whether this is because of an increased incidence of undetectable abnormalities in transgenic pups or because of a lowered threshold for ingestion in foster mothers is unknown.

Overall efficiency for a targeting procedure is usually quoted as the percentage of transgenic mice amongst total live pups. In this project two transgenic mice were identified out of 105 total live pups, giving an overall efficiency of 2%. As one pup was produced from six injections with super-coiled DNA and one pup from a single injection with linearized DNA, this implies that linearized DNA is associated with a higher targeting efficiency. However, an accurate comparison between the two DNA preparation methods is not possible with only one targeting event on each side. In order to properly address this issue more targeting events are required from further injections. A review of the literature indicates that both preparation methods have similar success rates and that the choice of DNA preparation method is largely due to laboratory preference (Giraldo & Montoliu, 2001).

Commercial companies producing transgenic mice by pronuclear injection with plasmid DNA generally quote overall efficiency rates of 10-40%. In the first report describing the generation of transgenic mice by pronuclear microinjection with 9 kb plasmids by Gorden *et al*. (1980) the overall efficiency was 0-3%. Subsequent studies, which provided a detailed assessment of overall targeting efficiency with plasmids, found rates of 1-10% (Brinster *et al.*, 1985; Hirabayashi *et al.*, 2001). Injection techniques and equipment have improved in recent years with increasing use of this technology and most transgenic projects utilising small plasmids would now produce a higher overall efficiency than this. In the ES cell and transgenic laboratory in this institute the observed overall efficiency with small plasmids is 10-30% (personal communication Ms. Z. Webster).

The overall efficiency with BACs in targeting projects is a little lower than for plasmids and generally in the range 5-20% (Giraldo & Montoliu, 2001; Abe *et al.*, 2004). A number of specific problems are presented by the size of the targeted BAC. Preparation of the BAC DNA has to be carried out with much greater diligence than that required for preparation of plasmid DNA. Any procedures which apply shear forces to the DNA, such as mixing or pipetting, must be kept to a minimum to reduce breakage of the DNA. Smaller fragments of DNA that might be generated after shearing are integrated much more readily into the mammalian genome increasing the potential for genetic disruption resulting in non-viable embryos. Shearing is less of a problem with BACs than YACs where it is estimated that less than half of genome-integrated YACs contain the whole YAC sequence (Schedl *et al.*, 1993).

In this institute's facility good results have been achieved in previous BAC targeting projects with supercoiled DNA. For these reasons pronuclear injections

were initially carried out using undigested DNA as this reduced the number of preparation steps required. However, injection of supercoiled DNA at the concentration normally used for plasmid injections (2 mg/L) proved difficult because of the high viscosity of the DNA suspension. The concern with reducing the DNA concentration is that as very low volumes are injected (1-2 pL), relatively few copies of targeted BAC are introduced into the pronucleus.

4.7 Summary

A BAC clone containing the *mF3* gene was successfully targeted and EGFP knocked-in at the 3′ end. Transgenic mice containing the *mF3-EGFP* fusion gene were produced following pronuclear injections of embryos with the targeted BAC clone.

Chapter 5 - Analysis of the TF-EGFP reporter in transgenic mice

5.1 Introduction

The previous chapter described how targeting of a BAC resulted in the insertion of the *EGFP* gene at the *mF3* locus, producing an *mF3-EGFP* fusion gene. Pro-nuclear injections of fertilized oocytes with the targeted BAC resulted in the production of two male transgenic mice. In the first transgenic mouse, TG17, both the point of linearization of the BAC and the position of integration in the genome are random. In the second transgenic mouse, TG254, the BAC was linearized in the vector backbone but the position of integration in the genome remains random. This chapter describes the methods used to assess whether the transgene was functioning as expected. The experiments were carried out entirely using mice derived from TG17 as TG254 was produced during the writing of this thesis.

Assessment of a transgene is normally carried out at a number of different stages along the pathway from genetic coding to mature protein function. Detailed analysis of the sequence of the targeted gene provides confirmation that the reporter gene is inserted as required and that the rest of the targeted locus remains intact. Expression of the targeted gene in comparison to the wild-type gene is used to ensure that the targeted gene is under normal regulatory control. Functional analysis may take a number of forms, the purpose of which is to establish that the fusion protein behaves in a similar manner to the endogenous protein in normal physiology and, where applicable, in pathological conditions. Generally this is done by a combination of quantitative and qualitative assays. Although the order of these assessments can reflect the biological pathways of transcription, translation and finally post-translational processing, in practice the functional assays are often performed first as a functioning transgene implies correct transcription and translation.

5.2 Analysis of the mF3-EGFP gene in transgenic mice

Genomic DNA was obtained from tail clips of mice. PCRs using the primer pairs MTFEx6F – MTFGFPR1 and GFP3F – 3HAR were used to generate amplicons for sequencing. This confirmed that the tripeptide spacer and the EGFP coding sequence were correctly inserted at the 3′ end of *mF3* before the terminal stop codon and that the proximal end of the 3′ UTR had identical sequence to the wild-type gene.

Southern blot analysis demonstrated the presence of bands of the expected size in transgenic mice indicating that there were no gross abnormalities at the 3′ end of the fusion gene.

5.3 Breeding of transgenic mice from TG17

The male transgenic mouse, TG17, was entered into a breeding program with wild-type C57BL/6J females to produce an F1 generation containing both transgenic and wild-type mice (Figure 5.1). The F1 transgenic mice have a single copy of the transgene and also carry both wild-type TF alleles, so their genotype may be written *mF3*$^{+/+}$*: mF3-EGFP*$^{+}$. Analysis of the TF-EGFP reporter model was carried out with experiments in these single copy transgenic F1 mice. Subsequently F1 *mF3*$^{+/+}$*: mF3-EGFP*$^{+}$ mice were crossed to produce mice with two copies of the transgene (*mF3*$^{+/+}$*: mF3-EGFP*$^{+/+}$) in the F2 generation.

The ratio of different genotypes in the F1-F3 generations may be predicted by the laws of Mendelian inheritance (Figure 5.1). However, the presence of a transgene may cause a deviation from predicted ratios because of a reduction in fertility or harmful effects that result in foetal or neonatal loss. The expectation is that the *mF3-EGFP* transgene should not have any such effects. Table 5.1 shows a summary of the progeny generated from breeding with the founder mouse TG17 and the observed frequency of pups with 1 or 2 copies of the transgene.

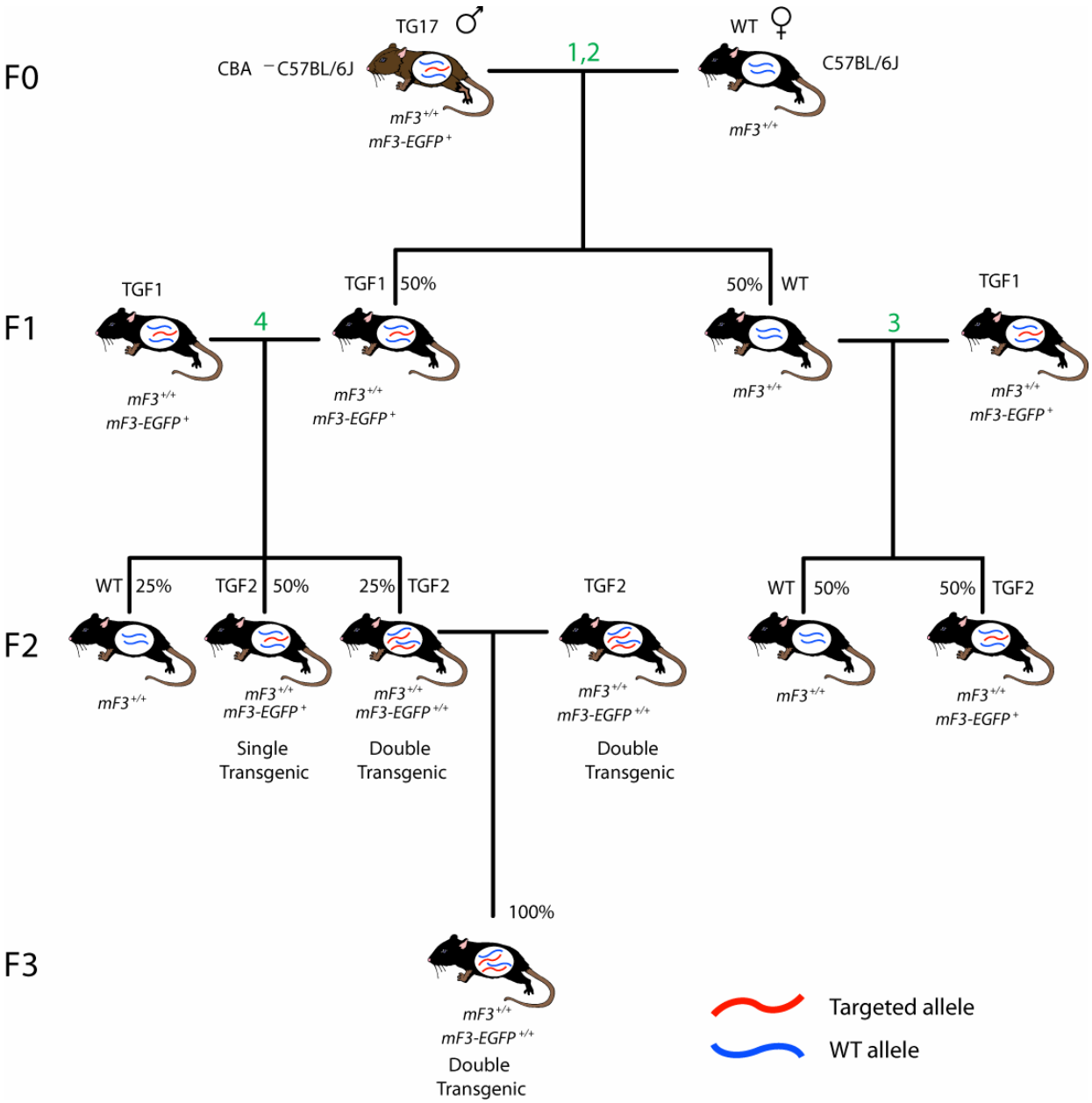


Figure 5.1 Breeding of transgenic mice from TG17. The mouse generation (F0-F3) is shown on the left. TG = transgenic, WT = wild-type. The genotype and percentage of progeny that would be expected from Mendelian inheritance is shown for each mating. Matings referred to in Table 5.1 are indicated by corresponding numbers in green (1-4). Crossing of the transgenic founder mouse carrying a single copy of *mF3-EGFP* with wild-type mice yields an F1 generation with 50% of progeny predicted to be transgenic. Crossing of transgenic mice in the F1 generation produces a predicted 25% of progeny carrying two copies of the transgene. A pure transgenic mouse line with two copies of the transgene is produced in the F3 and subsequent generations.

Mating No.	Mating pair	No. of matings	Pups born	Transgenic pups	% Transgenic	
					Exp	Obs
1	TG17 x WT1	4	37	6	50	18
2	TG17 x WT2-4	4	24	9	50	38
3	TGF1 x WT	6	47	18	50	38
4	TGF1 x TGF1	6	46	37	75	80

Table 5.1 Results of breeding with the TG17 founder male. The mating no. referred to in Figure 5.1 is in the first column. The first row shows the results of four matings between TG17 and the same wild-type female. The second row shows the results of matings between TG17 and three different wild-type females. Crosses in the F1 generation are shown in the bottom two rows. The final column shows the percentage of pups that were transgenic (either one or two copies) that were produced by all matings of that type.

From Figure 5.1 and Table 5.1 it can be seen that in the F1 generation 50% of mice are expected to be transgenic for *mF3-EGFP*. The first four matings involved TG17 and a single wild-type female. This is because experience in this transgenic laboratory has shown that chimaeric or transgenic founder males are more likely to breed successfully if kept with the same female initially (personal communication, J. Godwin). The percentage of transgenic pups from mating no. 1 was much lower than expected at 18%. Four subsequent matings with different wild-type females (mating no. 2) produced a higher transgenic percentage of 38%. The same percentage was obtained in crosses between F1 single-copy transgenic and wild-type mice (mating no. 3). As the percentages of transgenic pups produced from these crosses are all lower than predicted, the implication is that the transgene has a detrimental effect. However, in crosses between two F1 single-copy transgenic mice (mating no. 4) the percentage of transgenic (both single and double copy) offspring was 80%, slightly higher than the expected 75%. This indicates no adverse effect of the transgene.

There is no clear explanation for this apparent discrepancy in the results. However, the experience in the transgenic facility in this institute is that initial matings involving founder transgenic mice produce lower percentages of transgenic progeny than expected (personal communication, J. Godwin). It may be that in some situations foster mothers are able to distinguish transgenic offspring and remove them from the litter. If the distinguishing characteristic is not manifested in the

phenotype, it may not be apparent to researchers. In subsequent generations as the percentage of transgenic offspring in the litter increases, mothers may become more tolerant of transgenic pups. This would produce a rise in the observed percentage towards that predicted by the laws of Mendelian inheritance.

5.3.1 Allele quantification in transgenic mice

Distinguishing transgenic mice that carry either one or two copies of *mF3-EGFP* in the F2 generation generally requires mating of F2 transgenic mice with wild-type mice. The production of eight consecutive transgenic progeny from one or two such matings was taken as evidence that an F2 mouse had two copies of the transgene. This is based on a probability of 0.5 that any single progeny from a cross between a single-copy transgenic parent and a wild-type parent will be transgenic. There is a 99.6 % chance that if the F2 mouse has only a single copy of the transgene, at least one of its first eight offspring would be wild-type. It can therefore be safely assumed that a transgenic mouse producing eight consecutive transgenic progeny has two copies of the transgene.

Litter sizes from crossings of transgenic mice with pure-bred wild-type mice are often smaller than normal. Sometimes two matings are required to produce eight offspring. For males this can be achieved in a few weeks, but for females the process is longer on account of the requirement for the breeding cycle to be completed before the next mating. In addition to the time taken to establish the genotype, this strategy is wasteful as it increases the number of animals produced by the breeding program that are not suitable for study.

An alternative method for distinguishing mice with two copies of the transgene is by quantification of the transgene relative to the wild-type allele. This relies on real-time semi-quantitative PCR or RT-PCR. The design of this strategy is described later in this chapter. The forward primer used in the RT-PCR assay was replaced with a genomic DNA primer. DNA samples were extracted from ear punch biopsy, tail clips or blood samples following tail clip. Ultimately, allelic quantification using tail DNA did not have sufficient sensitivity to accurately distinguish single and double copy transgenic mice by copy number. Therefore the breeding method described above was used.

5.4 Breeding of transgenic mice from TG254

The second transgenic mouse, TG254, produced following pronuclear injection of targeted BAC was born after the experiments detailed in this chapter were undertaken. TG254 is currently entered into a breeding program to generate a second transgenic mouse line. It is expected that analysis of the transgene in this separate line will take place in the near future.

5.5 Microscopic analysis of the TF-EGFP reporter

Organs were removed from F1 mice and paraffin sections were made for immunohistochemical analysis. At the same time frozen sections were made for analysis by confocal fluorescence microscopy. Control sections were taken from wild-type littermates. TF expression patterns have previously been described in a variety of organs including heart, lung, kidney, brain and bowel (Drake *et al.*, 1989; Wilcox *et al.*, 1989; del Zoppo *et al.*, 1992; Flossel *et al.*, 1994; Luther *et al.*, 1996; Szotowski *et al.*, 2005; Bogdanov *et al.*, 2006). Of these the heart, lung and brain in particular have relatively high levels of TF in a well-characterised distribution. In this section references to previous studies indicate one or more of these reports.

Neural tissue lacks any significant fibrous or connective tissue and good quality sections are difficult to prepare even with the use of paraffin support. Therefore sections of heart and lung taken from transgenic and wild-type mice were assessed for distribution of TF and TF-EGFP by immunohistochemistry using antibodies against TF and GFP. The pattern of fluorescence was then assessed to see if this correlated with the immunological staining.

5.5.1 Immunohistochemistry of heart sections using anti-TF

The heart is an organ with a very high metabolic requirement and the myocardium tolerates interruption of blood supply very poorly. At the same time high pressures are created within the ventricular lumen increasing the potential for haemorrhage. It is therefore not surprising that TF levels are high in the heart. In mouse models of reduced TF expression, pathological effects are seen predominantly in the heart and central nervous system, confirming that these organs are particularly reliant on adequate TF expression (Pawlinski *et al.*, 2002; McVey, 2003).

The distribution of TF protein was demonstrated using a rabbit anti-TF antibody (Figures 5.2A-B and 5.3A-B). Binding of the primary antibody was visualised using a goat anti-rabbit immunoglobulin secondary antibody conjugated to alkaline

phosphatase. This produces a red signal after the addition of Fast Red substrate (DakoCytomation) at sites of antibody binding. As expected, similar patterns of TF expression were seen in heart sections from transgenic and wild-type mice. The signal intensity was not significantly different between transgenic and wild-type mice. TF was widely and diffusely expressed in the myocardium. Confirmation that the red signal was specific for the anti-TF antibody is shown by comparison with the pre-immune rabbit serum control (Figures 5.2C and 5.3C). In medium sized coronary arteries the intima, consisting of the endothelium and supporting connective tissue, was negative but localised foci of expression were seen in the media and outer adventitia. This pattern is the same as that seen in previous studies.

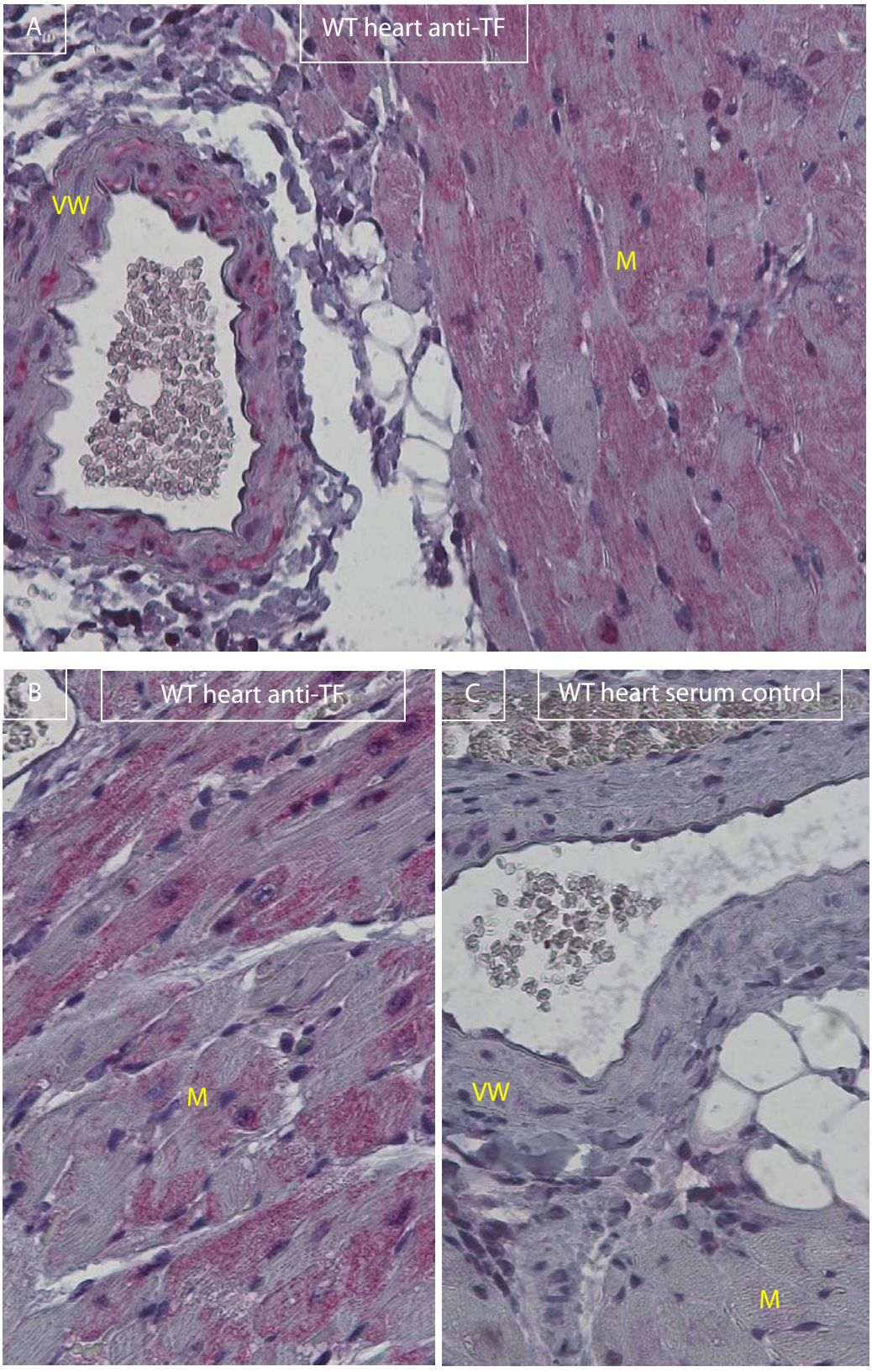


Figure 5.2 Sections of wild-type (WT) heart stained with anti-TF (A-B). Antibody binding is shown by red signal. (C) Control section with pre-immune serum replacing anti-TF. M = myocardium, VW = coronary artery wall. Sections are counter-stained with haematoxylin. x40 objective.

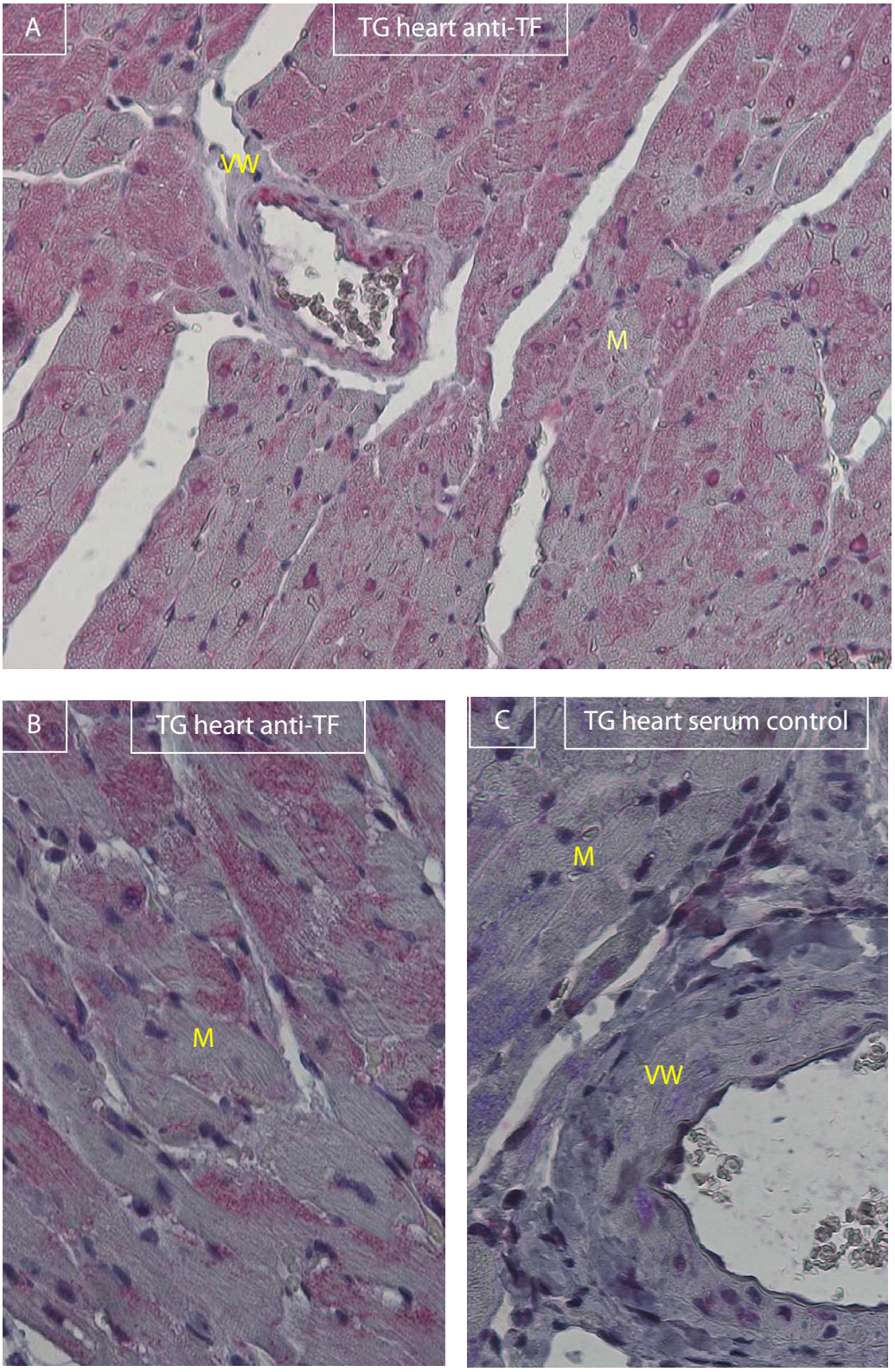


Figure 5.3 Sections of transgenic (TG) heart stained with anti-TF (A-B). Antibody binding is shown by red signal. (C) Control section with pre-immune serum replacing anti-TF. M = myocardium, VW = coronary vessel wall. Sections are counter-stained with haematoxylin. x40 objective.

5.5.2 Immunohistochemistry of heart sections using anti-GFP

Immunohistochemical staining using anti-GFP was of much lower intensity than with anti-TF. Haematoxylin counterstain obscured some of the pattern and therefore slides were prepared without counterstain.

If the reporter model was functioning correctly one would expect that signal would be detected in transgenic sections in a similar pattern to that seen with anti-TF. Wild-type sections should demonstrate no significant staining with anti-GFP. However, binding of anti-GFP was observed in sections from both wild-type littermates and transgenic mice (red stain in Figures 5.4A-B and 5.5A-B). A few fields in slides from transgenic mice showed intense anti-GFP staining in the adventitia of coronary arteries (arrows in Figure 5.5B). However, this pattern was not consistently repeated throughout a particular section and overall there was no significant difference between wild-type and transgenic mice.

Confirmation that the red staining was due to binding of anti-GFP is shown by comparison with the pre-immune serum controls which showed no staining in either wild-type (Figure 5.4C) or transgenic (Figure 5.5C) sections.

The detection of signal in the wild-type sections which do not contain GFP indicates that the primary anti-GFP antibody was binding in a non-specific fashion. A particular problem is that the distribution of binding with anti-GFP was similar to that observed with anti-TF, with staining particularly in the coronary vasculature. Therefore if weak signal was present from an appropriately expressed TF-EGFP reporter protein it would have been obscured by the non-specific binding of the anti-GFP antibody.

Ultimately the conclusion from this section must be that EGFP antigen was not consistently detectable in transgenic heart sections by immunohistochemistry.

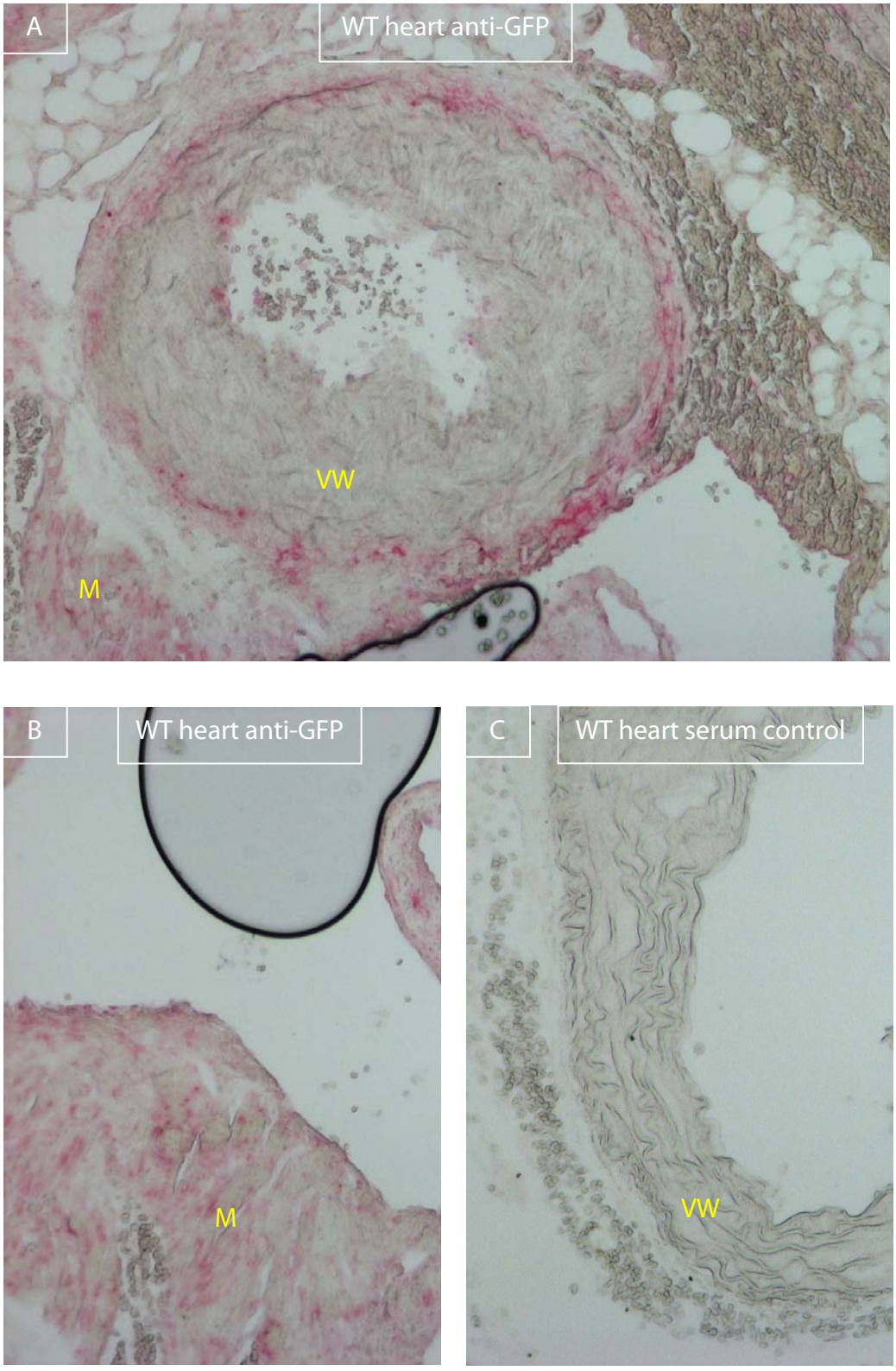


Figure 5.4 Sections of wild-type (WT) heart stained with anti-GFP (A-B). Antibody binding is shown by red signal. (C) Control section with pre-immune serum replacing anti-GFP. M = myocardium, VW = coronary artery wall. No counterstain has been applied. x40 objective.

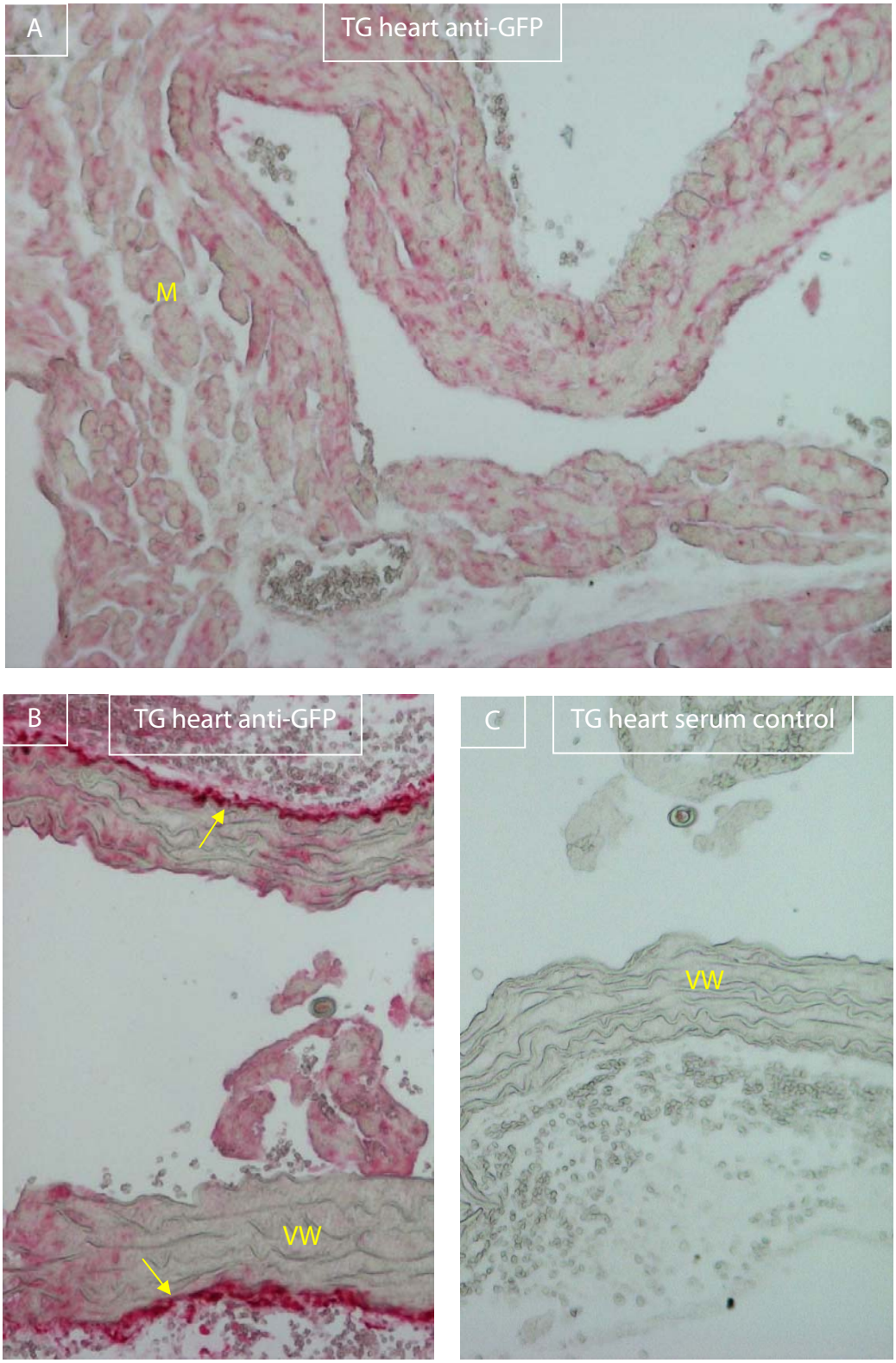


Figure 5.5 Sections of transgenic (TG) heart stained with anti-GFP (A-B). Antibody binding is shown by red signal. (C) Control section with pre-immune serum replacing anti-GFP. M = myocardium, VW = coronary artery wall. Arrows indicate high signal in the adventitia. No counterstain has been applied. x40 objective.

5.5.3 Fluorescent microscopy of heart sections

In sections from transgenic mice, fluorescence was detectable throughout the myocardium (Figure 5.6B). Particularly high contrast signal was obtained in the walls of the coronary vasculature (Figure 5.6A) and in some areas towards the luminal surface of the myocardium (Figure 5.6C). Fluorescence was detectable in a similar distribution in wild-type mice, indicating that much of the signal was due to autofluorescence (Figures 5.6D-F). Areas of high signal or high contrast green fluorescence were more readily seen in transgenic mice. However, the distribution of high signal was not consistent amongst identifiable structures in transgenic sections. Comparison of several sections from transgenic and wild-type littermates demonstrated that overall background levels of fluorescence were similar.

Very prominent filamentous structures were seen in the walls of coronary vessels at high power (arrowed in Figure 5.7A). The pattern of fluorescence, and proximity of these structures to the endothelium, suggested that the signal emanated from the internal elastic lamina. This membranous structure has been found to be negative for TF in previous histochemical studies. Therefore it is likely that this signal represents autofluorescence, particularly as it was often also fluorescent in wild-type sections.

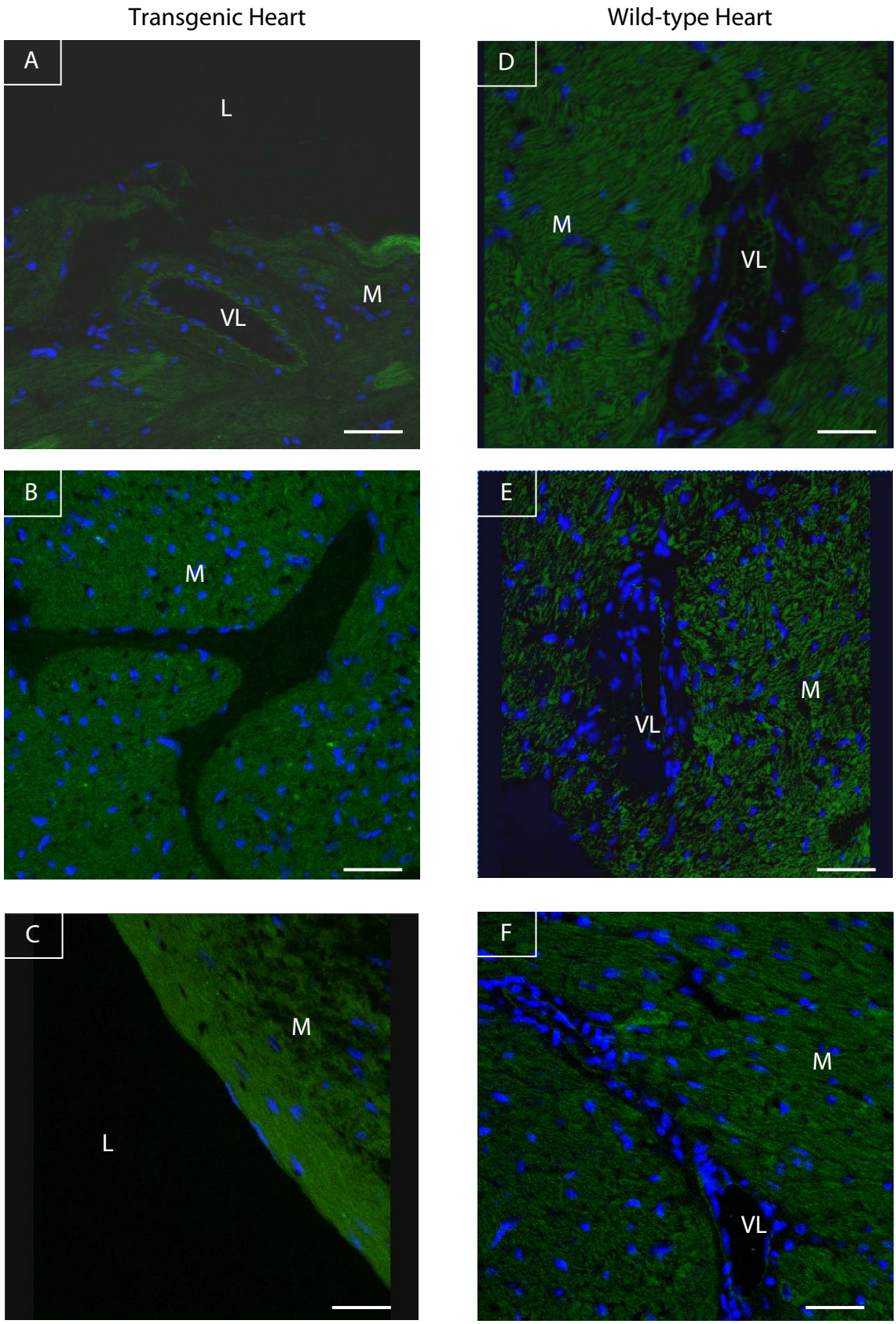


Figure 5.6 Fluorescence microscopy of heart sections. Magnification x80. Scale bars represent 50 µm. Images on the left (A-C) are taken from transgenic mice, those on the right (D-F) from wild-type littermates. L = ventricular or atrial lumen, VL = coronary vessel lumen, M = myocardium.

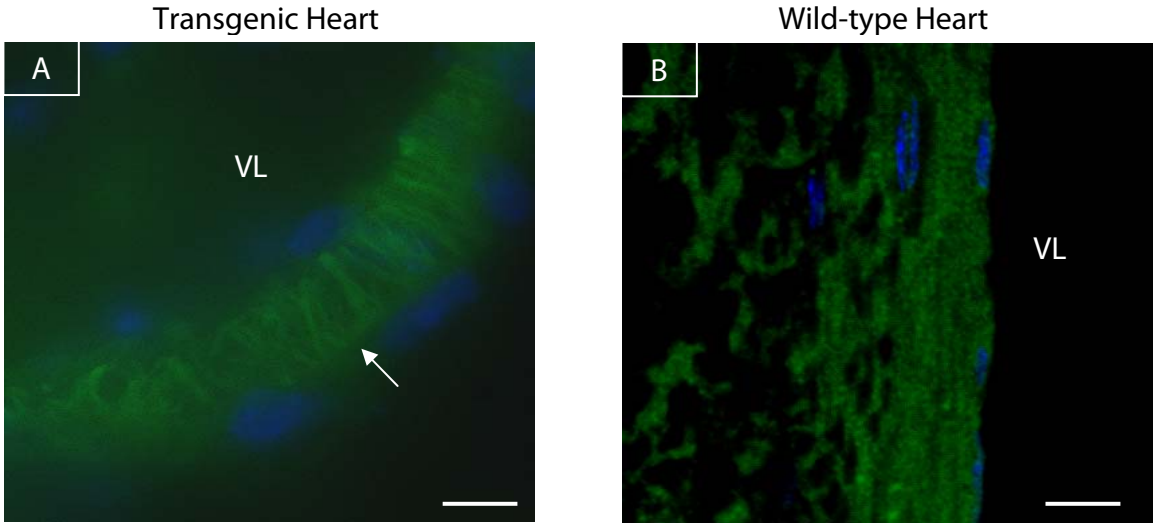


Figure 5.7 Fluorescence microscopy of heart sections showing detailed view of an intra-cardiac vessel wall. Magnification x200. Scale bars represent 20 µm (A) Transgenic. (B) Wild-type. VL = lumen of intra-cardiac vessel. The arrow indicates filamentous structures in the wall of a coronary vessel.

5.5.4 Immunohistochemistry of lung sections using anti-TF

The lungs are very vascular organs with thin-walled vessels to allow efficient gas exchange. Haemorrhage into the alveolar space is associated with severe morbidity; therefore procoagulant systems are particularly important. Pulmonary tissue is known to have a relatively high TF content and has been used in the past as a source from which TF may be readily purified (Nemerson, 1969). The epithelium is a highly specialized structure which provides the surface over which gas exchange occurs while simultaneously maintaining the integrity of the vasculature and airways. The basement membrane is critically important in these functions. The columnar epithelium lining the bronchial tree has a high secretory capacity that provides the surfactant necessary to lubricate and protect the airways. These histological structures are readily identified by conventional light microscopy and fluorescence microscopy.

Staining with anti-TF showed positive signal in the wall of pulmonary vessels and in the luminal surface of the columnar epithelium lining bronchioles (red staining in Figures 5.8 and 5.9). Pneumocytes forming the alveoli and lung parenchyma were generally negative. Pulmonary arterioles have a thinner media than similar sized coronary vessels, which reflects the difference in blood pressure in the two circulations. The focal anti-TF staining seen in the media of coronary vessels was not evident in the pulmonary vasculature but adventitial staining was seen. Identical distributions of anti-TF staining were seen in transgenic and wild-type sections.

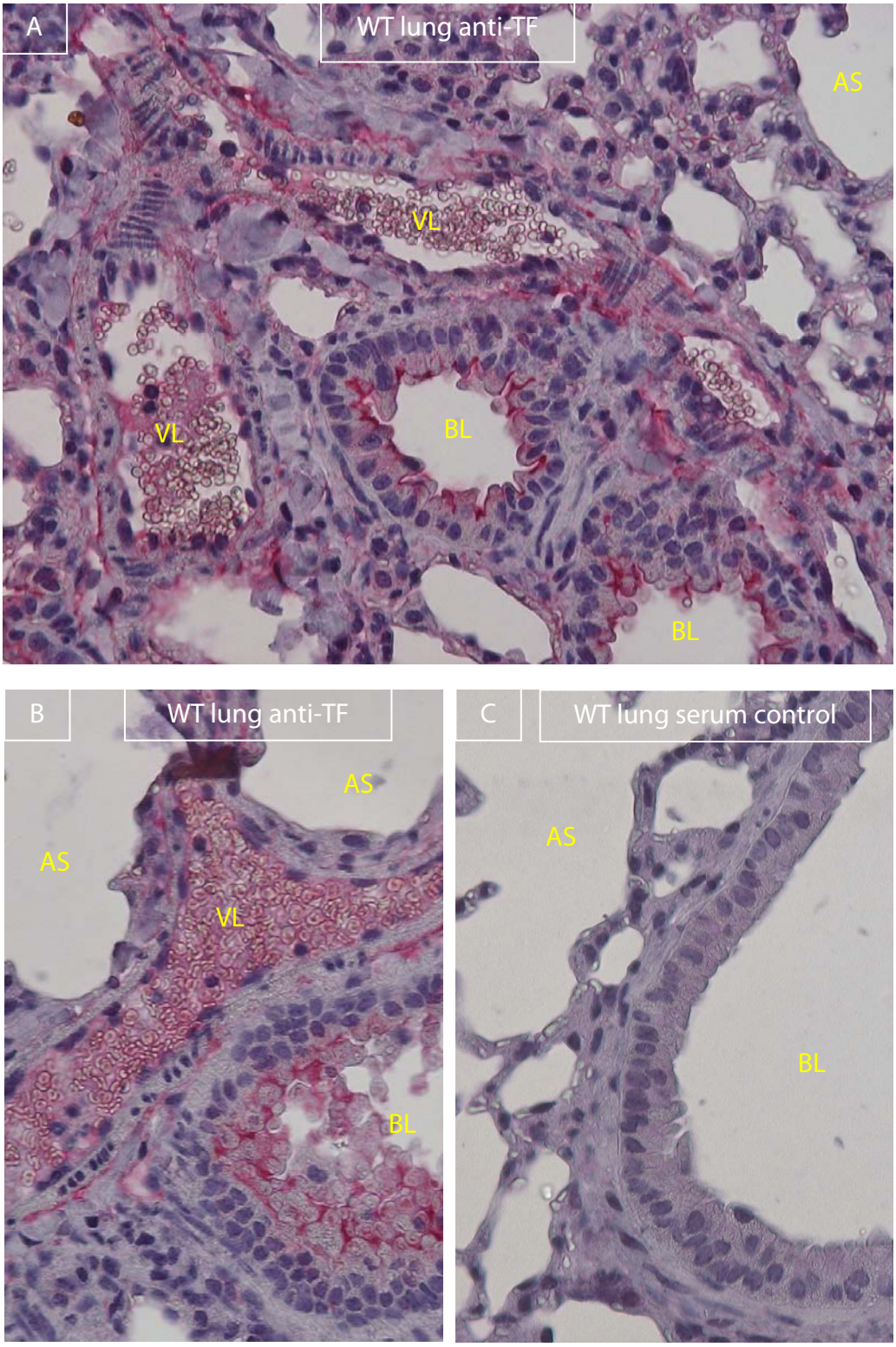


Figure 5.8 Sections of wild-type (WT) lung stained with anti-TF (A-B). Antibody binding is shown by red signal. (C) Control section with pre-immune serum replacing anti-TF. BL = bronchiolar lumen, AS = alveolar space, VL = vessel lumen. Sections are counter-stained with haematoxylin. x40 objective

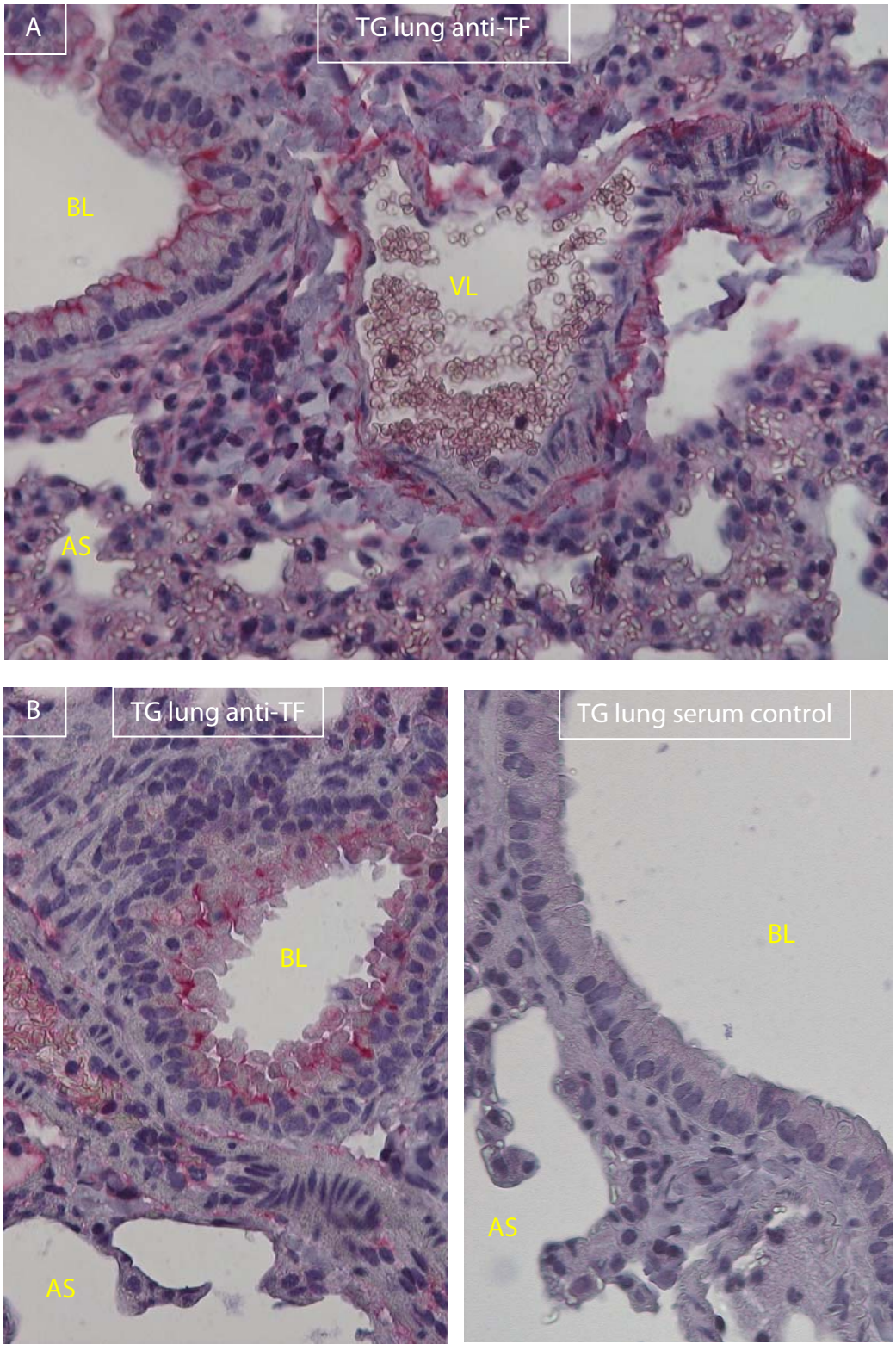


Figure 5.9 Sections of transgenic (TG) lung stained with anti-TF (A-B). Antibody binding is shown by red signal. (C) Control section with pre-immune serum replacing anti-TF. VL = vascular lumen, BL = bronchiolar lumen, AS = alveolar space. Sections are counter-stained with haematoxylin. x40 objective.

5.5.5 Immunohistochemistry of lung sections using anti-GFP

The results of staining with anti-GFP (AB290, Abcam) in lung sections were similar to that seen in heart sections. Positive signal was seen in both wild-type and transgenic sections indicating non-specific binding of the anti-GFP antibody (red signal in Figures 5.10 and 5.11). If the anti-GFP antibody was correctly detecting the EGFP tag in the fusion protein, one would expect that not only would there be absence of positive signal in the wild-type sections but also that the staining pattern would be similar to that observed with anti-TF in transgenic sections. In fact there were significant differences between the staining patterns observed with anti-TF and anti-GFP in lung sections from transgenic mice. The most striking positivity with anti-GFP was again in the columnar epithelium. But while the staining had been confined to the luminal surface with anti-TF (Figures 5.8A and 5.9A), with anti-GFP intense staining was seen within columnar cells, particularly in the basal regions (Figures 5.10A and 5.11A). This appeared to correlate with the intra-cytoplasmic position of the cells' secretory apparatus. Some patchy staining was seen in the lung parenchyma and the specialized alveolar pneumocytes which was not seen with anti-TF. A common characteristic of these structures is that there is heavy mucus secretion in the lining of the bronchial tree and also to a lesser extent in the alveoli. This may have been the target for non-specific anti-GFP binding. Staining in the pulmonary vessels was similar with anti-TF and anti-GFP.

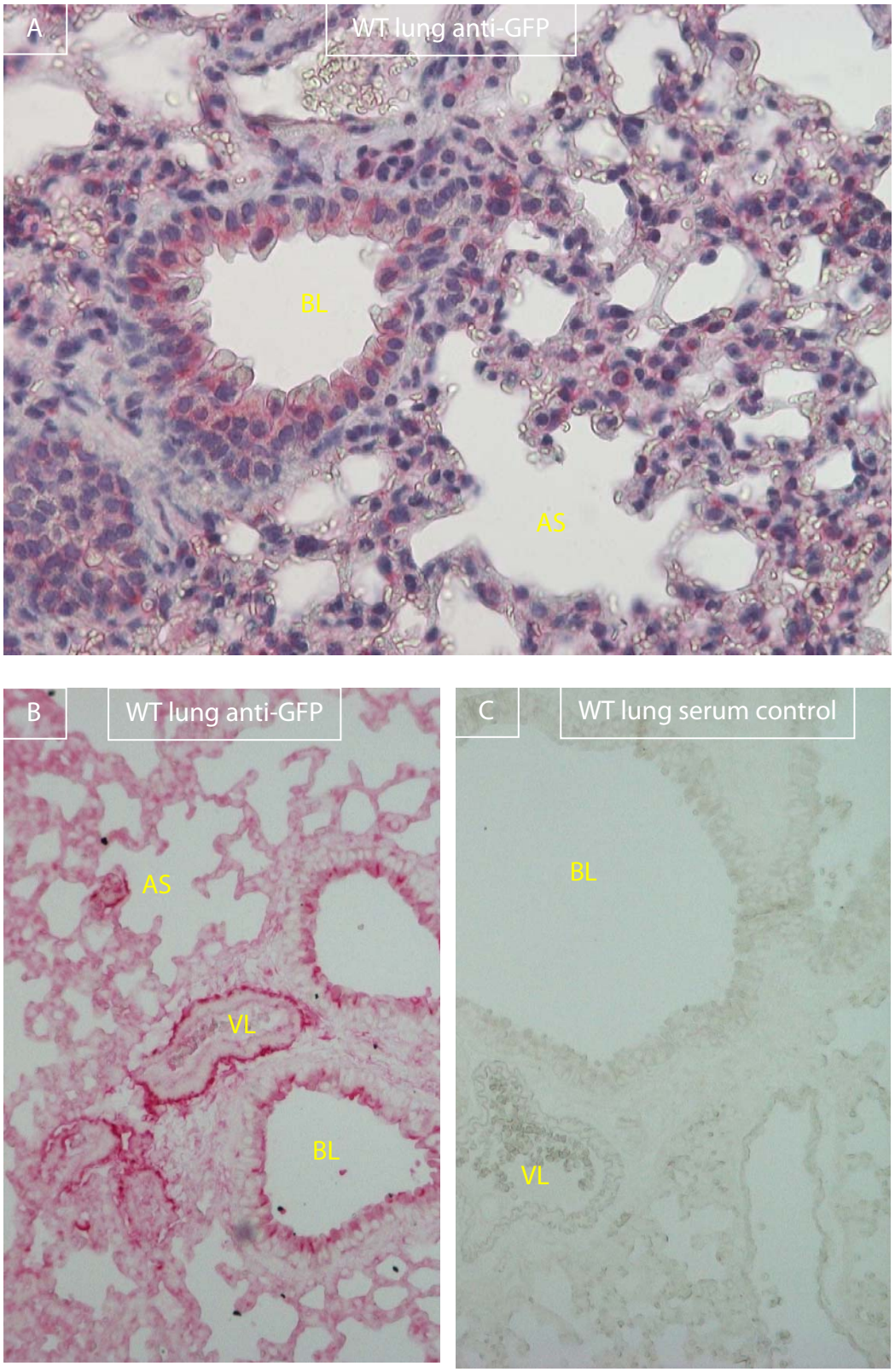


Figure 5.10 Sections of wild-type (WT) lung stained with anti-GFP shown by red signal. (A) Haematoxylin counterstain x40. (B) No counterstain x20. (C) Pre-immune serum control, no counterstain. BL = bronchiolar lumen, AS = alveolar space, VL = vascular lumen.

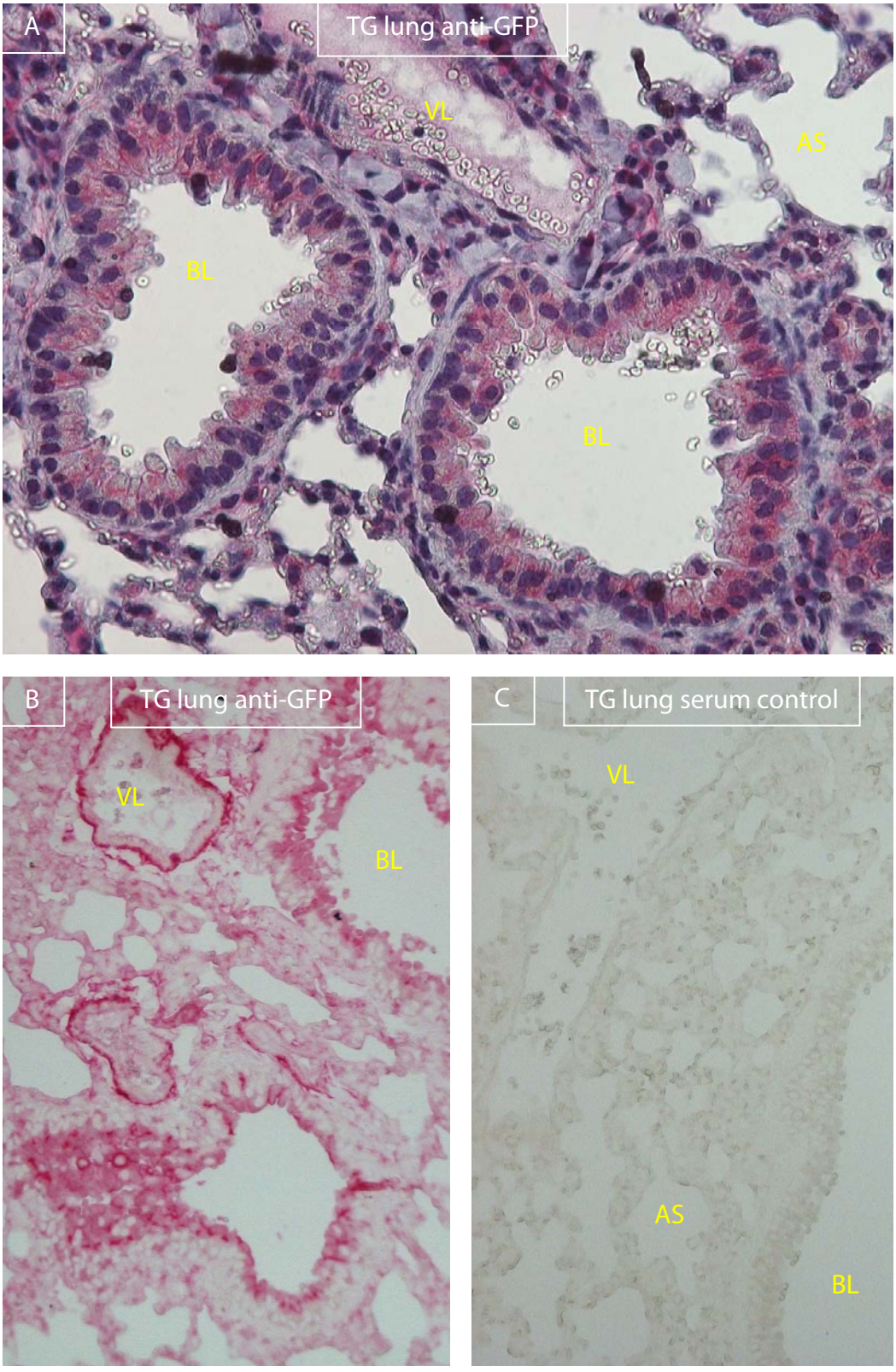


Figure 5.11 Sections of transgenic (TG) lung stained with anti-GFP shown by red signal. (A) Haematoxylin counterstain x40. (B) No counterstain x20. (C) Pre-immune serum control, no counterstain. BL = bronchiolar lumen, AS = alveolar space, VL = vascular lumen.

5.5.6 Anti-GFP immunohistochemistry using biotin-streptavidin amplification

It was clear from the results with anti-GFP in both heart and lung sections that non-specific binding was occurring that would mask any signal that might be detected from a functioning reporter protein. To try and overcome this problem immunohistochemistry was repeated using a different primary anti-GFP antibody (A11122, Invitrogen) and detection system. The alternative detection system used a biotinylated secondary antibody and a streptavidin-conjugated horseradish peroxidase (HRP). The presence of several biotin molecules on the secondary antibody allows the formation of multiple biotin•streptavidin•HRP complexes at the site of primary antibody binding. This amplifies the signal and allows the primary antibody to be used at greater dilution thereby reducing the potential for non-specific binding.

The results with the Invitrogen anti-GFP antibody were similar to that seen with the Abcam anti-GFP antibody. Staining was seen predominantly in the columnar epithelium but there was no significant difference in the staining pattern seen in transgenic and wild-type sections (brown signal in Figure 5.12).

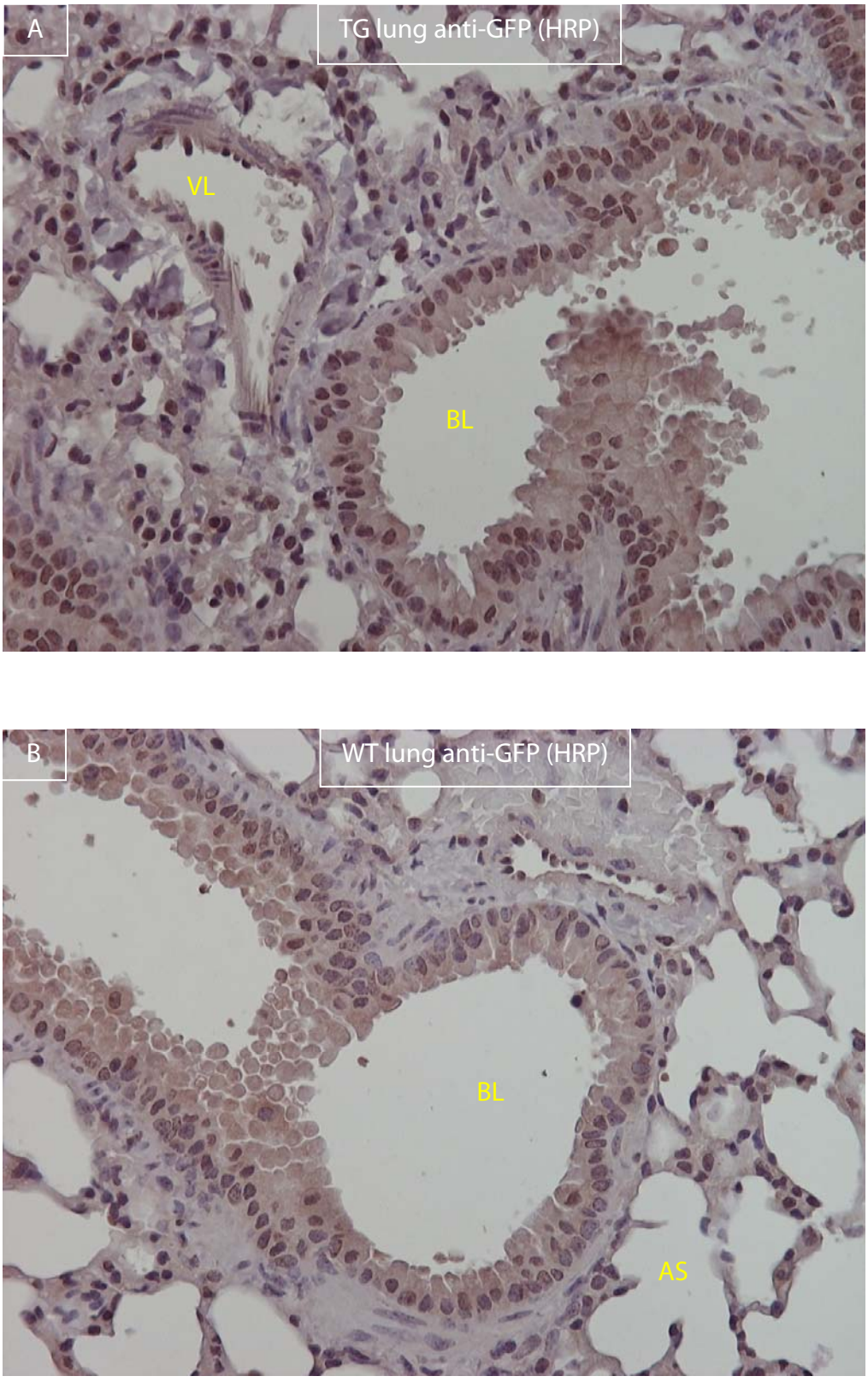


Figure 5.12 Sections from transgenic (A) and wild-type (B) lung stained with anti-GFP and counterstained with haematoxylin. The biotinylated secondary antibody was detected with streptavidin conjugated HRP shown by brown signal. BL = bronchiolar lumen, AS = alveolar space, VL = vascular lumen.

5.5.7 Fluorescent microscopy of lung sections

Fluorescence in the EGFP channel was readily detected in both transgenic and wild-type sections (Figures 5.13 and 5.14). The intensity and distribution of fluorescence was not significantly increased in transgenic mice.

The bronchiolar epithelium was positive with low intensity signal corresponding to columnar epithelium and high intensity signal corresponding to the basement membrane. With high power objectives (x63 and x100) fluorescence was detected from the lumen adjacent to columnar cells (Figures 5.14B, D and E). This suggests that the secretory products released from columnar cells were fluorescent. The pulmonary vasculature displayed fluorescence in a similar pattern to that seen previously in cardiac sections.

Although the pattern of fluorescence corresponded with that seen using the immunohistochemical methods for TF and EGFP detection, signal was detected in wild-type sections. This signal was due to autofluorescence within structures such as the bronchiolar epithelium.

5.5.8 Summary of microscopic analysis

The anti-TF antibody demonstrated similar patterns of TF expression to those reported in previous studies. The pattern was similar in transgenic and wild-type mice. Staining with anti-GFP proved less informative with non-specific binding masking any signal that might have been generated by a functioning TF-EGFP reporter protein. This problem was not resolved by changing the primary antibody or the secondary detection system. Similarly it was not possible to demonstrate significant differences between transgenic and wild-type mice using fluorescence microscopy.

These results indicate that although wild-type TF was expressed appropriately in transgenic mice, there was no additional detectable expression of the TF-EGFP reporter in mice derived from the TG17 founder mouse.

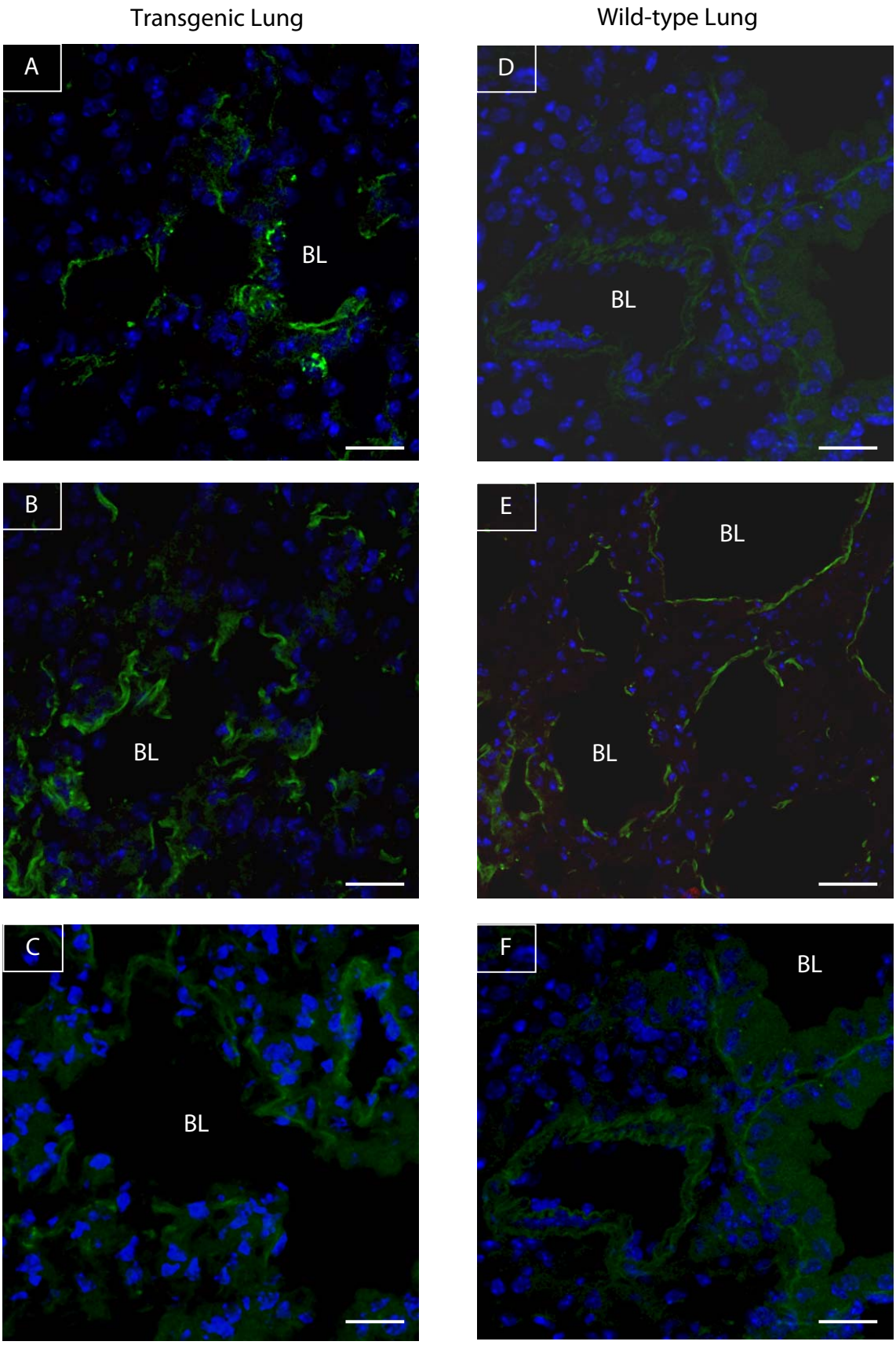


Figure 5.13 Fluorescence microscopy of lung sections. Magnification x80. Scale bars represent 50 µm. Images on the left (A-C) are taken from transgenic mice, those on the right (D-F) from wild-type littermates. BL = bronchiolar lumen.

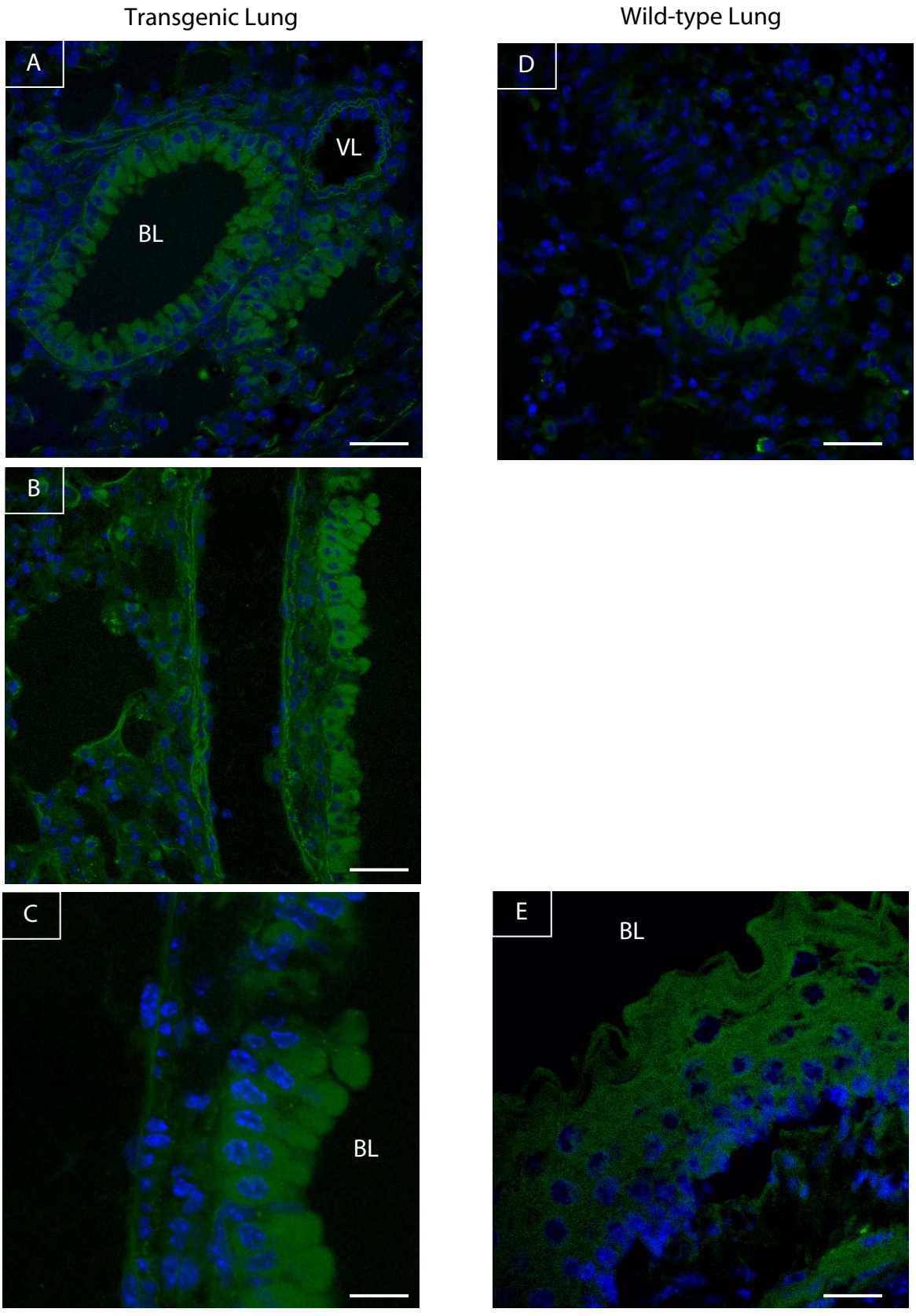


Figure 5.14 Fluorescence microscopy of lung sections. Images on the left (A-C) are taken from transgenic mice, those on the right (D&E) from wild-type littermates. Images A, B and D: magnification x126, scale bars represent 30 µm. Images C and E: magnification x200, scale bars represent 20 µm. BL = bronchiolar lumen, VL = vascular lumen.

5.6 Analysis of mF3-EGFP transcription in whole tissue samples

It was not possible to detect appropriate expression of TF-EGFP by microscopy. PCR analysis followed by direct sequencing and Southern blot analysis had previously indicated transmission of the transgene in the F1 and subsequent generations derived from the TG17 founder mouse. This suggested a defect in transcription, translation or post-translational modification of the transgene.

5.6.1 Design of real-time RT-PCR assays

Appropriate transcription of the transgene was assessed by real-time semi-quantitative reverse transcriptase (RT)-PCR. To provide the template for these assays, total RNA was first extracted from organs from transgenic and wild-type mice. Then cDNA was synthesized using RT primed with random hexamers. The cDNA provided the template for real-time PCR assays.

The principle of the real-time assay is that under specific conditions, the amount of product formed during PCR is dependent on the amount of template at the start of the reaction. The amount of product present in any given cycle is detected using a probe that hybridizes to a specific target sequence in the product (Figure 5.15A). The probe contains a fluorescent dye at the 5′ end and a quencher at the 3′ end. As long as the probe is intact, fluorescence is prevented by the close proximity of the dye and the quencher. During the extension part of the PCR cycle the DNA polymerase encounters and hydrolyses the bound probe (Figure 5.15D). This separates the dye from the quencher and the amount of fluorescence is therefore directly proportional to the amount of PCR product present. The PCR conditions are set such that the amount of product, and therefore fluorescence, is dependent only on the amount of starting template. If standards containing known amounts of template are included in the assay a true quantitative result may be obtained. Otherwise a semi-quantitative result is obtained, which allows comparison of the relative amounts of template present in various samples.

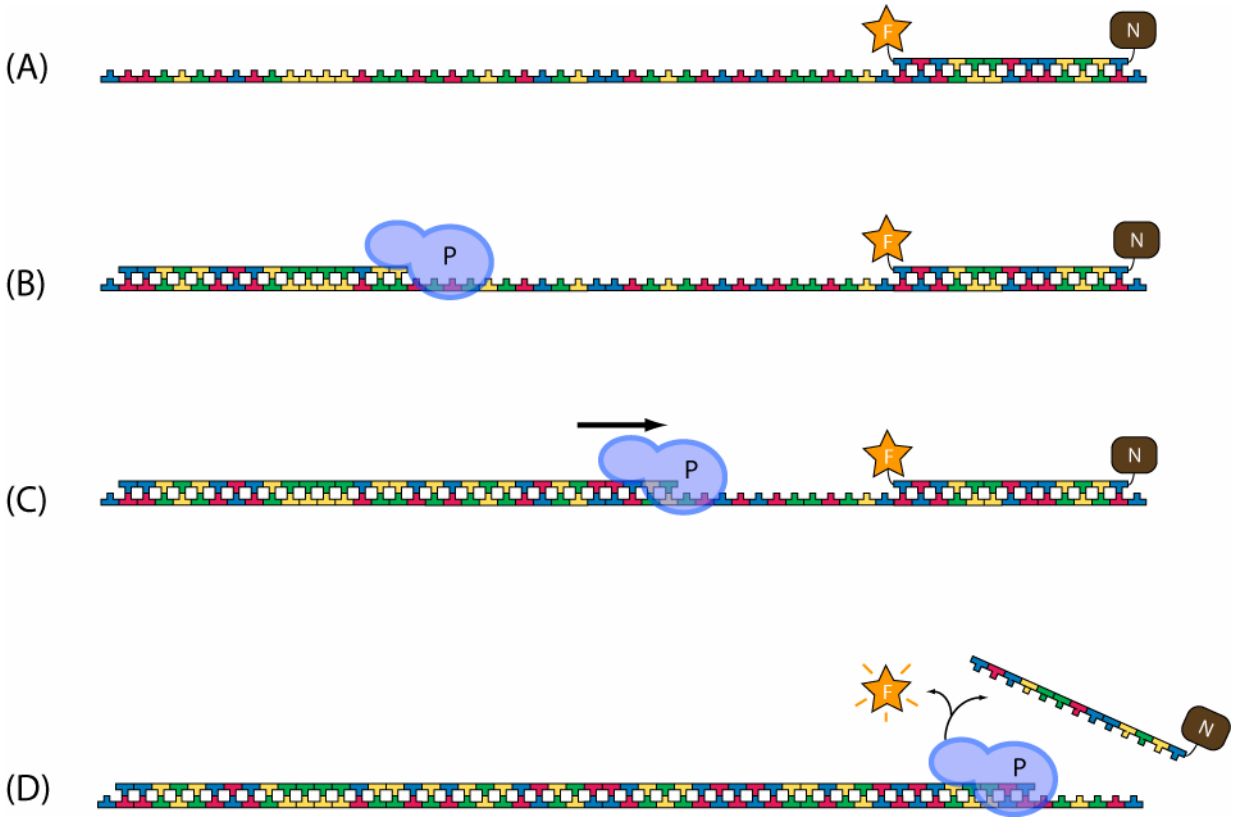


Figure 5.15 Quantitation of PCR products with fluorescence-labelled probes. (A) After denaturation of the double stranded DNA template, the MGB probe hybridizes to its specific target sequence. The fluorescent dye FAM (F) is at the 5′ end of the probe, with non-fluorescent quencher (N) at the 3′ end. (B) Subsequently, the PCR primer binds followed by DNA polymerase (P). (C) The polymerase extends the new DNA strand in the 5′ to 3′ direction. (D) When the polymerase encounters the probe the latter is displaced from the template releasing the fluorescent dye at its 5′ end.

The PCR reaction takes place in a system that monitors fluorescence in real-time. The critical measurement in real-time PCR quantitation is the cycle threshold (C_T). This is defined as the cycle number, or fraction thereof, at which the amount of fluorescence exceeds an arbitrary fixed point. Various equations may be used for obtaining the C_T. In this project the threshold fluorescence was set at 10 standard deviations above the mean fluorescence value over cycles 1-10. The C_T was then the cycle fraction when fluorescence first passed this point. The fluorescence detected in a no template control was taken as the baseline and subtracted from all values at the start of the threshold calculation. As the threshold is the same for all samples in an

assay, samples with greater amounts of template at the start of the reaction have relatively lower C_T values.

A critical factor in determining the relationship between the fluorescence level and amount of starting template is the PCR reaction efficiency (RE). This is given by the equation log RE = -1/*m*, where *m* is the slope of the curve measured during exponential increase. In an ideal reaction with an efficiency of 1.0, the amount of product doubles each cycle. In experiments where the results of separate PCR assays are analysed to indicate relative quantities of starting template, it is important that there is no significant difference between the REs of the assays. To satisfy this criterion the assays are designed to comply with very stringent parameters (set out in Primer Express v1.5, Applied Biosystems).

The latest probes used in real-time PCR are designed to be highly specific while maintaining a high PCR efficiency. The probes used in this project contain the fluorescent dye FAM at the 5′ end with a non-fluorescent quencher (NFQ) at the 3′ end. These probes also incorporate a minor groove binding protein (MGB) at the 3′ end. This serves to increase the specificity of hybridization while maintaining the optimal probe length. Without the MGB protein, increasing specificity is achieved by lengthening the probe which causes a detrimental rise in the probe's melting temperature (T_m).

Standardised assays for common genes are commercially available. A ready-made assay containing the PCR primers and MGB probe for *mF3* pre-mixed, was obtained from Applied Biosystems (Assay ID Mm00438853_m1). This assay is hereafter referred to as TFc. The probe sequence is given in Appendix 3 (TFc-MGB) and hybridization position shown in Appendix 5. The PCR primer sequences were not provided by the manufacturer.

The TFc assay amplifies a product spanning exons 1 and 2. It is therefore specific for the cDNA and will not amplify the gene. This product is found in both the wild-type and targeted alleles. Using the same design parameters as for TFc, two assays were produced to specifically amplify either the wild-type or targeted alleles. In order to prevent amplification of the gene the PCR products span the exon 5-6 boundary. To ensure that each MGB probe bound specifically to only one allele the hybridization position was limited to the end of exon 6 and the immediate 3′ sequence. As can be seen in Appendix 5 the probe binding region is specific to each allele. Assay WT measures expression of the wild-type allele, while assay EGFP measures the targeted allele. The probe (WT-MGB and EGFP-MGB) and primer

sequences (TFcDNA-F1, WTcDNA-R1 and EGFPcDNA-R2) are shown in Appendix 3 and their binding positions are shown on the cDNA map in Appendix 5.

Assays were performed in duplicate in 96-well plates in a DNA Engine Opticon continuous fluorescence detection system. Fluorescence was read at the end of extension in every PCR cycle. No-RT and no-template controls were included with every plate.

5.6.2 Results of real-time RT-PCR assays

Expression of the wild-type and targeted TF alleles was assessed in samples of heart, lung, brain and kidney taken from transgenic mice and wild-type littermate controls. The results are shown as graphs with the C_T and RE values tabulated below in Figures 5.16 – 5.19.

The results from heart, lung and brain were very similar. The TFc assay detects the total amount of TF expression resulting from the combined transcription of wild-type and targeted alleles. In all experiments the C_T values were lower in transgenic samples (dark blue squares) in comparison with wild-type controls (light blue squares), indicating a higher level of total TF expression in transgenic mice. This is consistent with the presence of three alleles expressed in transgenic mice compared with two in wild-type littermates.

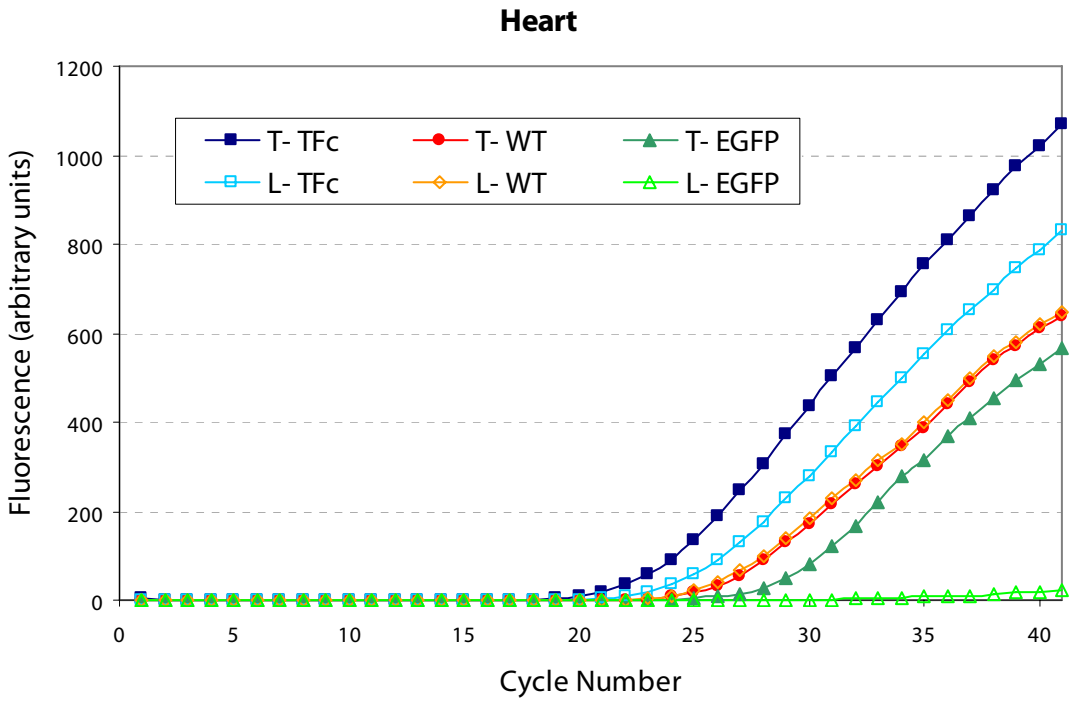
The WT assay detects only the wild-type allele and showed similar C_T values in both transgenic mice (red circles) and wild-type littermates (orange circles), indicating similar levels of expression of the endogenous allele. The EGFP assay detects only the targeted allele and as expected produced measurable C_T values in the transgenic samples only (dark green triangles).

The REs of the three assays were similar (0.91 – 0.96), but not identical. For this reason it was not possible to accurately quantitate relative expression levels of the two alleles. However, the data from the transgenic samples did show higher C_T values with the EGFP assay than with the WT assay. This indicates that the targeted allele was expressed at a lower level consistent with expression of a single, as compared with two wild-type, alleles.

The difference in the REs is most apparent when comparing the TFc and WT assays in wild-type littermates. These assays are indicated by the L-TFc (light blue squares) and L-WT (orange circles) traces in the graphs. Because there are only two alleles available for amplification in these samples and both assays will amplify both alleles, the C_T values would be similar if the REs for the two assays were identical. The

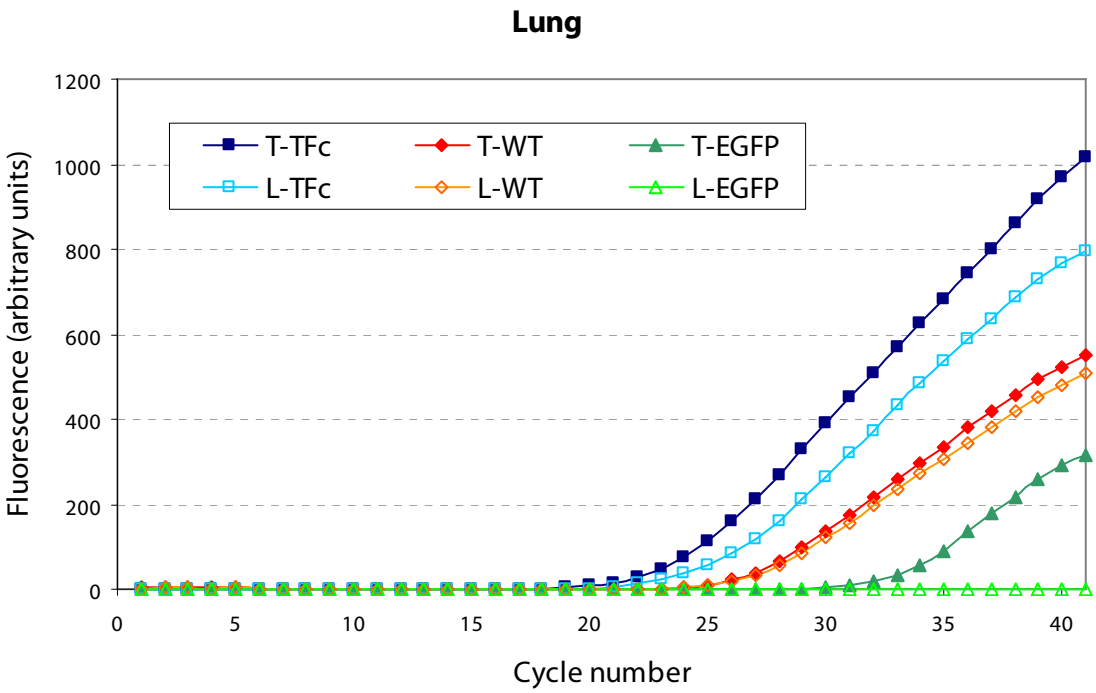
observed results showed that the C_T with the WT assay was 2-4 cycles higher than the C_T with the TFc assay in the same sample. As the same amount of cDNA was entered into both assays, this result indicates that the small difference in calculated RE was nevertheless significant enough to cause a difference in the C_T.

The results from kidney samples differed from those seen with other organs (Figure 5.19). No expression of the targeted allele was detectable in the kidney samples from transgenic mice. In addition the TFc control assay showed a slightly lower C_T in the sample from wild-type littermates. Whereas the samples from the other organs indicated expression from an additional allele in transgenic animals, this result indicates that total TF expression was similar in the kidneys of transgenic mice and wild-type littermates. The data is consistent with a lack of detectable expression of the transgene in the kidney. Since wild-type TF is expressed this is likely to be due to a site-specific integration event that prevents expression of the transgene in this particular tissue. It is clear that TF expression is carefully regulated by tissue specific mechanisms. It is possible that the specific regulatory processes active in the kidney resulted in different levels of expression of the two alleles. It should be noted that the total TF expression levels in the kidney samples were less than that seen in other organs. It may be that the combination of these factors reduced the transcription of the targeted allele to below detectable levels in the kidney. However, other transgenic lines would have to be examined to verify this.



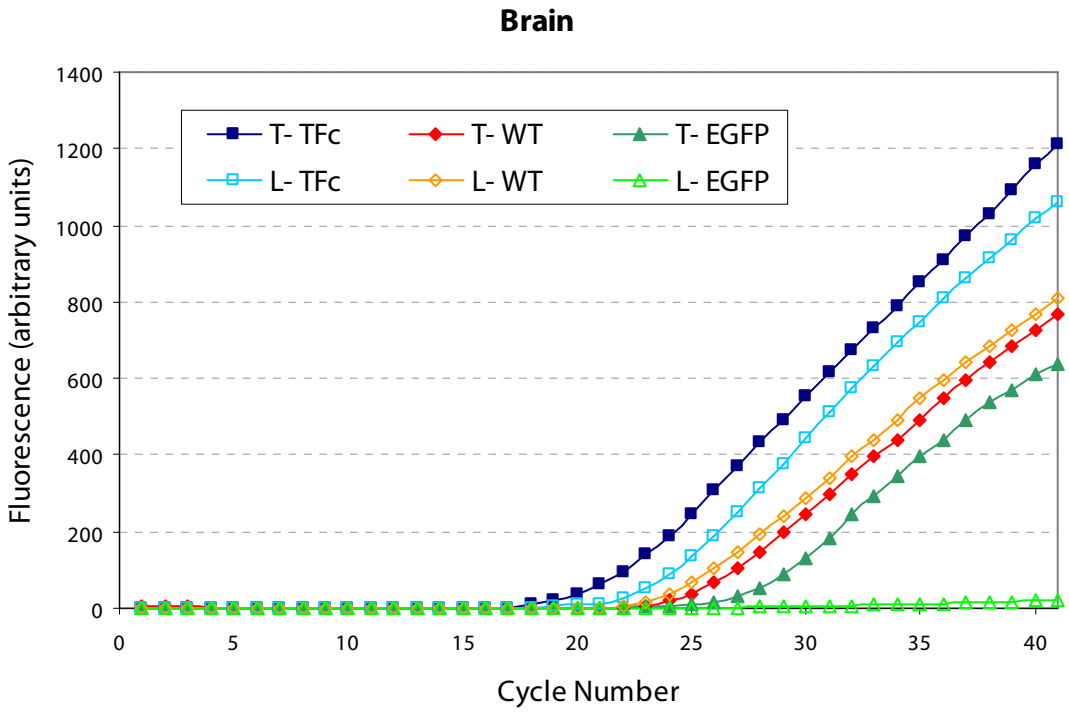
Transgenic RNA				Wild-type RNA			
Assay	Symbol	C_T	RE	Assay	Symbol	C_T	RE
T-TFc		20.0	0.96	L-TFc		21.7	0.96
T-WT		24.1	0.94	L-WT		24.5	0.94
T-EGFP		26.0	0.93	L-EGFP		-	-

Figure 5.16 Real-time RT-PCR quantitation of wild-type and targeted TF mRNAs from total heart RNA. The graph shows accumulation of fluorescence with increasing PCR cycles. For each curve the first letter denotes the sample source (T = transgenic, L = wild-type littermate). The PCR assay is given after the dash (TFc = TF control, WT = wild-type allele specific, EGFP = EGFP-targeted allele specific). For example T-TFc indicates that a transgenic sample was used in the TF control assay. The table shows the cycle threshold (C_T) and reaction efficiency (RE) for each assay.



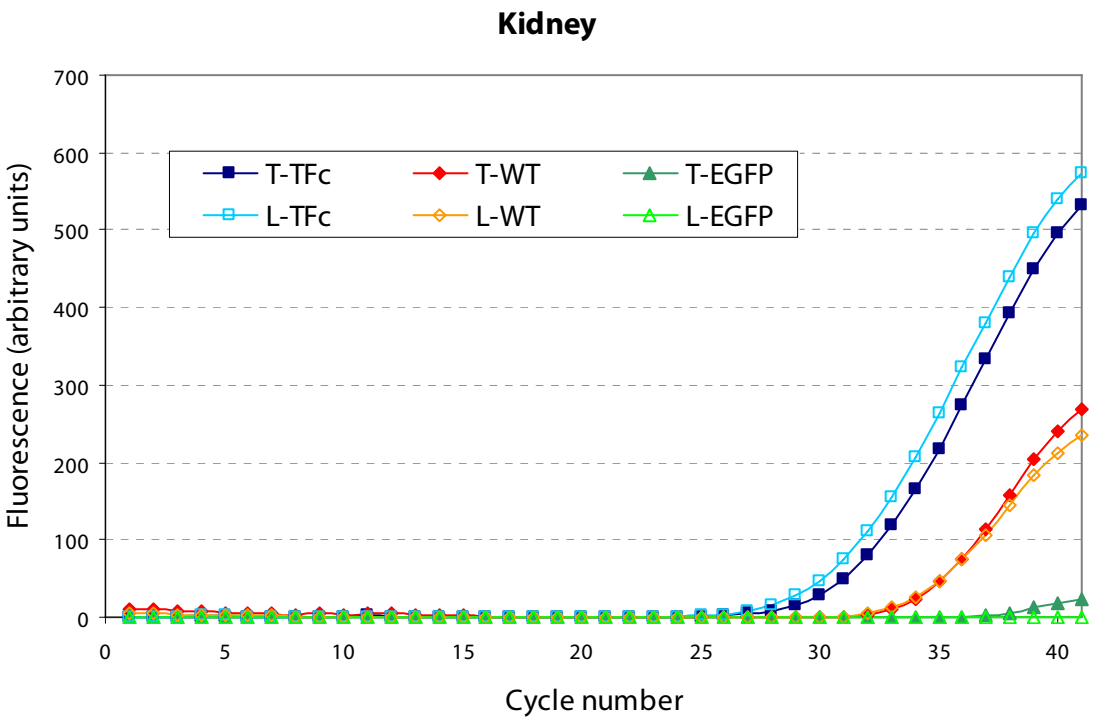
Transgenic RNA				Wild-type RNA			
Assay	Symbol	C_T	RE	Assay	Symbol	C_T	RE
T-TFc		20.2	0.96	L-TFc		21.6	0.95
T-WT		24.8	0.93	L-WT		25.1	0.92
T-EGFP		31.0	0.91	L-EGFP		-	-

Figure 5.17 Real-time RT-PCR quantitation of wild-type and targeted TF mRNAs from total lung RNA. The graph shows accumulation of fluorescence with increasing PCR cycles. For each curve the first letter denotes the sample source (T = transgenic, L = wild-type littermate). The PCR assay is given after the dash (TFc = TF control, WT = wild-type allele specific, EGFP = EGFP-targeted allele specific). For example L-TFc indicates that a sample from wild-type littermates was used in the TF control assay The table shows the cycle threshold (C_T) and reaction efficiency (RE) for each assay.



Transgenic RNA				Wild-type RNA			
Assay	Symbol	C_T	RE	Assay	Symbol	C_T	RE
T-TFc		18.0	0.96	L-TFc		20.6	0.96
T-WT		23.2	0.95	L-WT		22.4	0.95
T-EGFP		25.0	0.95	L-EGFP		-	-

Figure 5.18 Real-time RT-PCR quantitation of wild-type and targeted TF mRNAs from total brain RNA. The graph shows accumulation of fluorescence with increasing PCR cycles. For each curve the first letter denotes the sample source (T = transgenic, L = wild-type littermate). The PCR assay is given after the dash (TFc = TF control, WT = wild-type allele specific, EGFP = EGFP-targeted allele specific). For example T-WT indicates that a transgenic sample was used in the wild-type allele specific assay. The table shows the cycle threshold (C_T) and reaction efficiency (RE) for each assay.



Transgenic RNA				Wild-type RNA			
Assay	Symbol	C_T	RE	Assay	Symbol	C_T	RE
T-TFc		28.0	0.93	L-TFc		27.2	0.96
T-WT		32.7	0.90	L-WT		32.6	0.93
T-EGFP		-	-	L-EGFP		-	-

Figure 5.19 Real-time RT-PCR quantitation of wild-type and targeted TF mRNAs from total kidney RNA. The graph shows accumulation of fluorescence with increasing PCR cycles. For each curve the first letter denotes the sample source (T = transgenic, L = wild-type littermate). The PCR assay is given after the dash (TFc = TF control, WT = wild-type allele specific, EGFP = EGFP-targeted allele specific). For example L-EGFP indicates that a sample from wild-type littermates was used in the EGFP-targeted allele specific assay The table shows the cycle threshold (C_T) and reaction efficiency (RE) for each assay.

5.7 Analysis of TF-EGFP protein in tissues by Western blot

Transcription of the EGFP-targeted *mF3* gene was detected in whole organ RNA samples by real-time RT-PCR. In order to assess whether transcription of the gene correlated with the presence of protein in tissues, samples were analysed by Western blot.

Even in tissues with relatively high TF expression the protein forms a very small percentage of the total extractable protein. Brain is thought to have the highest amount of TF on a weight by weight basis, but even in this organ purification studies have shown that it only represents 0.001% of total protein (Bach *et al.*, 1981). It is therefore not surprising that detection of TF antigen in whole tissue extracts by direct Western blot analysis yields poor results. TF, in common with other transmembranous proteins, is poorly soluble in weak detergents. Solubilisation of the protein requires the use of strong detergents with vigorous mechanical lysis. The result is a low concentration TF-containing solution of high viscosity. Consequently to load sufficient amounts of TF on a gel requires relatively large volumes of the solution. This leads to poor separation of the constituents by gel electrophoresis.

To overcome these problems, TF was first purified and concentrated from whole tissue extracts by immunoprecipitation. The resulting material was much more amenable to electrophoresis followed by detection by standard Western blot methods.

Immunoprecipitation was carried out using both anti-TF and anti-GFP antibodies. When GFP was the target antigen a rabbit polyclonal antibody was used in the immunoprecipitation step and a mouse monoclonal antibody was used for Western detection. When TF was the target the same rabbit polyclonal antibody was used for both steps. Following immunoprecipitation the immune complexes containing the target protein were captured onto magnetic beads conjugated with an anti-rabbit immunoglobulin antibody. After washing the target protein is eluted from the magnetic beads by denaturation of the protein-immunoprecipitating antibody bond. The conditions of the denaturation step are designed such that the immunoprecipitating antibody is left bound to the magnetic beads.

If the denaturation is not correctly carried out the immunoprecipitating primary antibody may be released from the magnetic beads and may then represent a contaminant in the sample applied to the gel. This is of particular concern in the immunoprecipitation of TF because the immunoglobulin heavy chain migrates at about the same position as TF. Ordinarily this problem is overcome by using

antibodies from different species in the immunoprecipitation and Western blot. This was the strategy used when GFP was the target, but was not feasible in the case of TF because the only available antibodies were all raised in rabbits.

Therefore a control was required to indicate the presence of any contamination from the immunoprecipitating antibody in the Western blot. A recombinant, soluble form of TF (rTF) had previously been produced in this laboratory using a mammalian expression system in bacteria. This consists of the extracellular domains (residues 1-223) with a short protein C tag fused at the C-terminus. The protein C tag is used in purification of the protein through an affinity column. rTF migrates at a position distinct from the immunoglobulin heavy chain and would therefore demonstrate the presence of any contaminating immunoglobulin.

The total protein concentration in whole tissue extracts was 10-20 g/L. The volume of protein extract entered into the immunoprecipitation procedure was that containing 10 mg total protein. On the basis that TF represents 0.001% of total protein the amount of TF available for immunoprecipitation is then about 100 ng in brain samples. Whole tissue extracts were taken from brain and lung. The concentration of TF in lung is not known but because TF expression in the lung is restricted to very few cell types, it is likely to be less than in brain.

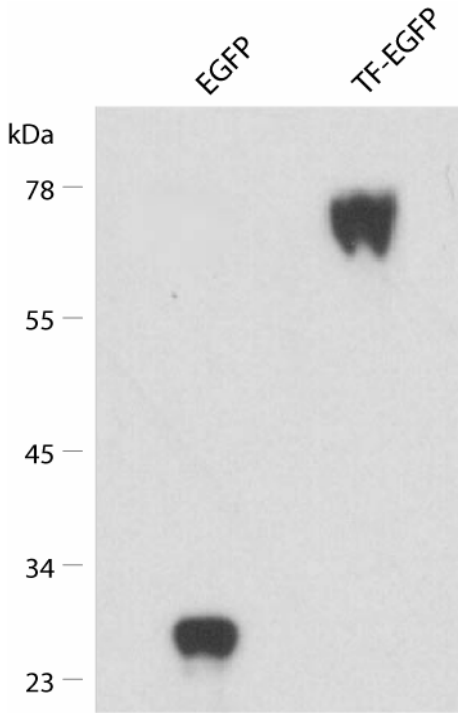


Figure 5.20 Western blot analysis of protein extracts from CHO cells expressing EGFP and TF-EGFP. The membrane has been stained with monoclonal anti-GFP. Reproduced from Mumford, 2003 with permission.

5.7.1 Results of Western blot analysis

TF subjected to SDS-PAGE (sodium dodecyl sulphate – polyacrylamide gel electrophoresis) under reducing conditions has a M_r of about 45 kDa. Although no source of TF-EGFP was available to act as a control in these experiments, previous data obtained in this laboratory confirmed that murine TF-EGFP expressed in CHO cells has an apparent M_r of about 75 kDa (Figure 5.20). Similarly a study of a rabbit TF-EGFP fusion protein expressed in HEK-293 cells found that the protein had a M_r of 73 kDa (Fortin *et al.*, 2005). A soluble GFP preparation with M_r of 27 kDa was used as a control to demonstrate appropriate specificity of the detection system.

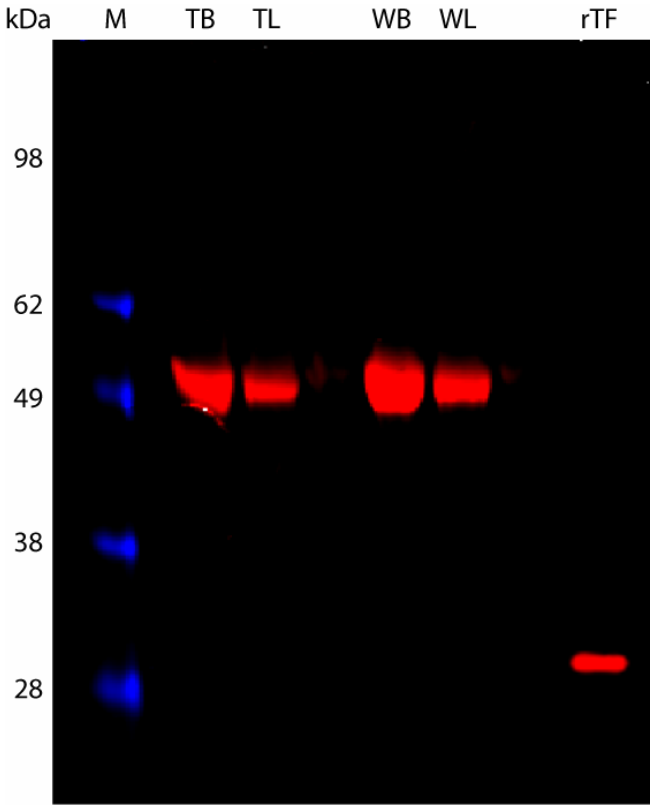


Figure 5.21 Western blot analysis of protein extracts from tissue samples from transgenic mice and wild-type littermates. Samples with 10 mg total protein, expected to contain about 100 ng TF, were immunoprecipitated with anti-TF prior to electrophoresis. The membrane has been stained with polyclonal anti-TF. M = marker, TB = transgenic brain, TL = transgenic lung, WB = wild-type brain, WL = wild-type lung, rTF = immunoprecipitated recombinant TF control. The size of protein standards is shown on the left. The marker at 98 kDa is not visualised in the infra-red detection system but is in the visible part of the spectrum.

Western blot analysis of immunoprecipitates from whole tissue extracts probed with rabbit anti-TF polyclonal antibody showed the same 45 kDa band in brain and lung samples from both transgenic and wild-type mice (Figure 5.21). The rTF control showed a single band at 28 kDa, consistent with the predicted mass. The absence of a higher molecular weight band confirms that there was no contaminating immunoprecipitating antibody. This indicates that the bands seen on the Western blot correspond to different forms of TF. The 45 kDa band in the tissue samples is at the correct position for wild-type TF. No additional bands were seen in the transgenic samples. If the transgenic samples contained a TF-EGFP fusion protein a further 75 kDa band would be expected in these samples but not in those from

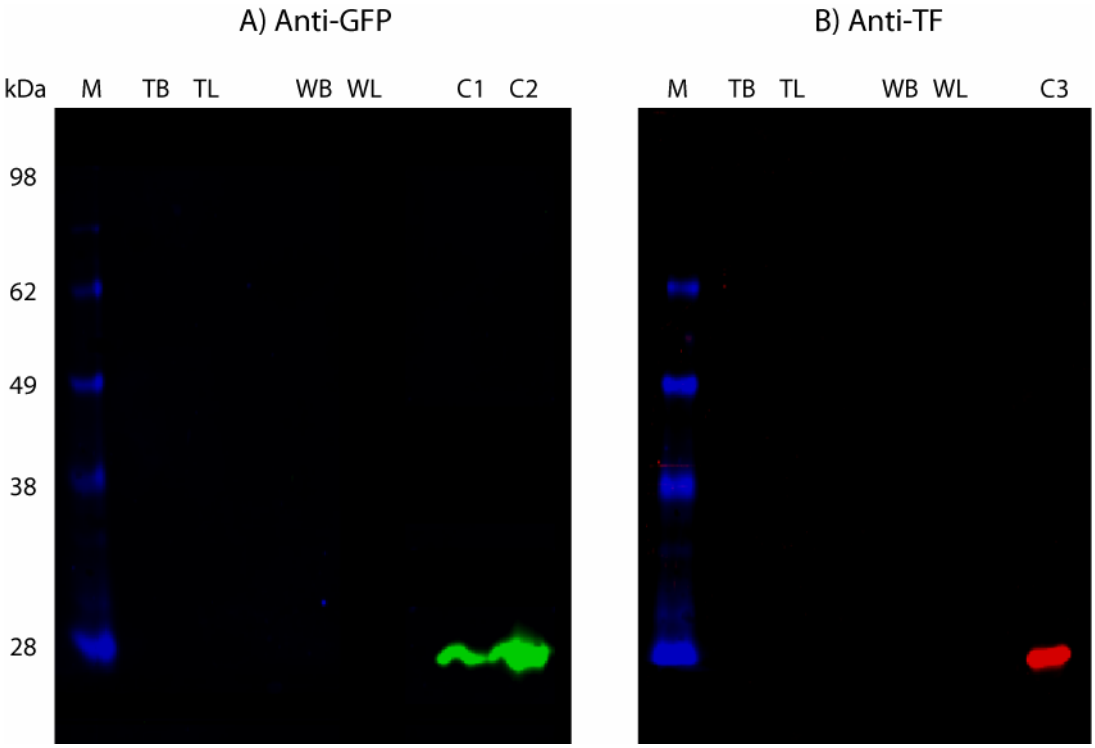


Figure 5.22 Western blot analysis of protein extracts from tissue samples from transgenic mice and wild-type littermates. Samples with 10 mg total protein were immunoprecipitated with anti-GFP prior to electrophoresis. A) Membrane stained with monoclonal anti-GFP. B) Membrane stained with anti-TF. M = marker, TB = transgenic brain, TL = transgenic lung, WB = wild-type brain, WL = wild-type lung, C1 = immunoprecipitated recombinant GFP control, C2 = recombinant GFP control applied directly to gel, C3 = recombinant TF control applied directly to gel. The size of protein standards is shown on the left. The marker at 98 kDa is not visualised in the infra-red detection system but is in the visible part of the spectrum.

wild-type mice. As no such band was detected this indicates that only wild-type TF protein was present in the transgenic organs.

Because the TF-EGFP fusion protein could not be detected with anti-TF antibody, the immunoprecipitation was repeated using polyclonal anti-GFP antibody. Western blot analysis was then carried using both anti-GFP and anti-TF antibodies. No bands were detected in any of the tissue samples indicating that there was no detectable GFP antigen in the samples (Figure 5.22). The control sample (Figure 5.22A, C1) demonstrated that GFP antigen could be immunoprecipitated and detected by this method.

Given that *mF3-EGFP* transcripts were readily detectable in samples from transgenic mice, the lack of detectable TF-EGFP fusion protein in extracts from the same tissues indicated that transcription of the allele did not result in production of mature protein. This suggests either an abnormality in post-transcription processing or that transcription occurred in cells that were unable to translate the signal into mature protein.

In order to assess whether the coding sequence had remained intact during the targeting process, the *mF3-EGFP* cDNA was amplified by PCR for direct sequencing. PCR primers MTFC-F1, MTFC-R2, MTFC-F3 and MTFC-R4 were designed to specifically amplify the *mF3-EGFP* cDNA and were tested to ensure that no product was amplified from the wild-type *mF3* cDNA. The PCR products were sequenced using the same primers and additional nested primers (MTFC-F6S, MTFC-R5S, MTFC-F8S and MTFC-R7S). Direct sequencing confirmed that there were no mutations in the coding sequence.

5.8 Analysis of mF3-EGFP transcription in monocytes

It can be assumed that an mRNA of the correct sequence would be correctly translated in cells that constitutively express the protein. However, the transcripts were detected in RNA samples of whole tissue. In a complex tissue there are very few cell types that would be expected to express TF constitutively. The possibility remains that a very sensitive detection technique such as PCR would be able to detect ectopic transcription from other cell types. Ectopic transcription would not be expected to result in appropriate protein production. This is a possible mechanism by which apparent gene transcription may be detected without subsequent protein detection.

To investigate this possibility analysis of transcription of the *mF3-EGFP* transgene was carried out in a single cell type: monocytes. As discussed in the introduction, the expression of TF in primary monocytes and monocyte cell lines has been extensively studied (reviewed in Lwaleed *et al.*, 2001). Monocytes do not express TF in their normal physiological condition but may be induced to do so when exposed to well-defined stimuli. Expression of functional TF may be induced in monocytes by stimulation with lipopolysaccharide (LPS) and phorbol myristate acetate (PMA) (van den Eijnden *et al.*, 1997; Osterud *et al.*, 2000; Rushworth *et al.*, 2005; Molor-Erdene *et al.*, 2005). A number of variables such as sex, age and the presence of leucocytes have been shown to influence the response of monocytes to LPS. In

particular CD14 on the surface of monocytes has been shown to influence the binding of LPS to its ligand (Gangloff *et al.*, 2005). Although CD14 is present on human monocytes it is not always detectable on mouse monocytes (Lai *et al.*, 1998; Khazen *et al.*, 2005). Nevertheless the presence of CD14 is a requirement for mouse monocytes to respond to LPS. This response is supported by the addition of serum to monocytes in culture medium and it is has been suggested that this is due to the adsorption of soluble CD14 in the serum on to the monocyte cell surface. There are no detectable stores of TF in resting monocytes. In addition the induction of functional TF in monocytes is preceded by an increase in TF mRNA levels, indicating that the monocytes synthesize TF *de novo* rather than releasing previously manufactured protein (Molor-Erdene *et al.*, 2005).

The expression of functional TF on monocytes was assessed by a coagulation test based on the prothrombin time (PT). Using this test the optimal concentrations of LPS and PMA were empirically determined. It was similarly found that stimulation in whole blood produced a more reproducible response than stimulation of purified monocytes in culture medium supplemented with foetal calf serum.

Following LPS and PMA stimulation of whole blood a monocyte-rich fraction was prepared by centrifugation through a density gradient. Examination of smears made from the fraction under a conventional light microscope showed that the majority of cells (~95% on differential count) were monocytes (Figure 5.23A-C). The remainder of the cells were small, mature lymphocytes. No contaminating granulocytes were seen. The microscopy findings were confirmed by flow cytometry using anti-CD11b, frequently used to identify mouse monocytes (Lai *et al.*, 1998) (Figure 5.23D).

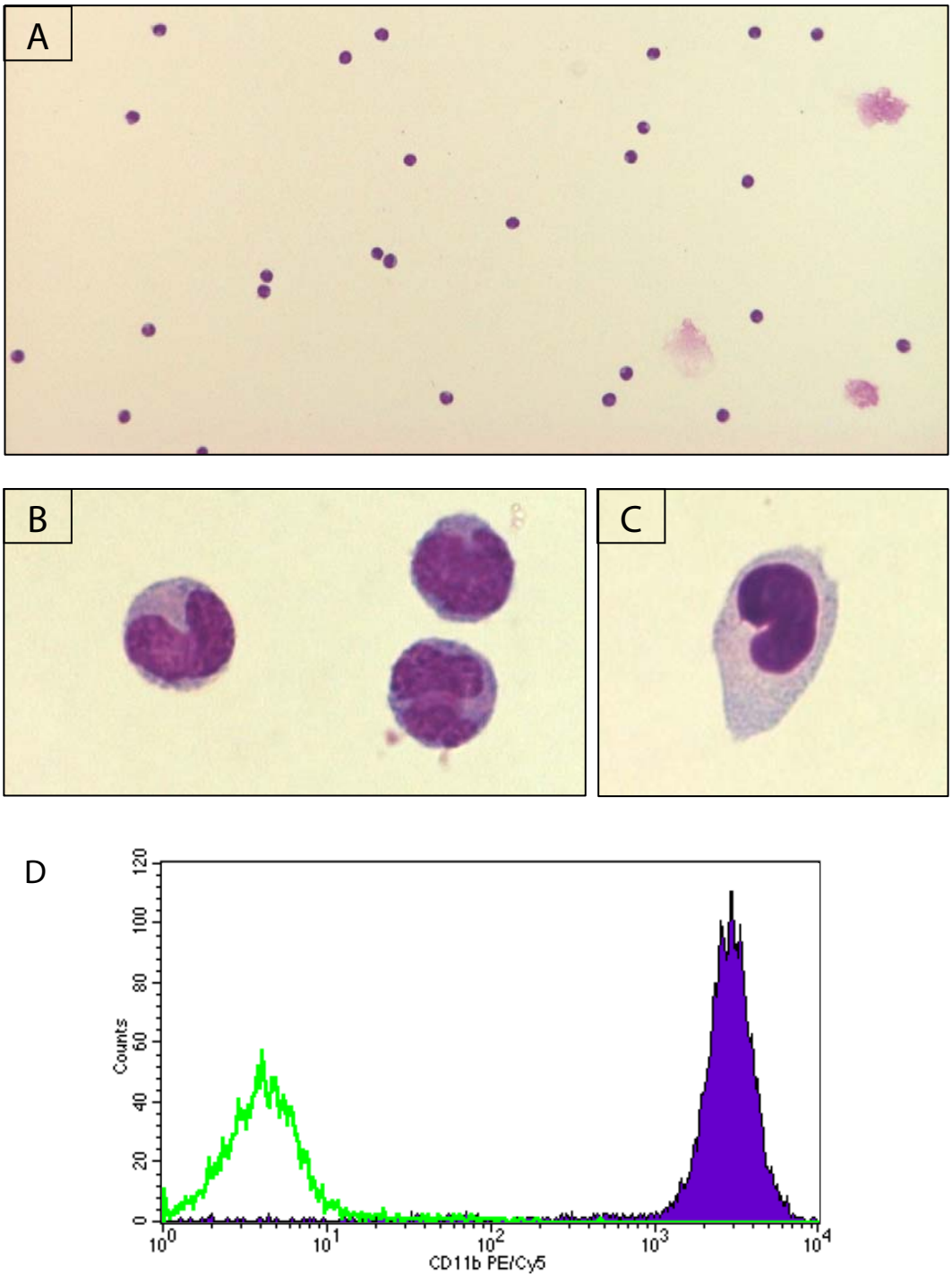


Figure 5.23 A mouse monocyte-rich fraction prepared by centrifugation across a density gradient. Light microscopy of smears stained with May-Grünwald-Giemsa.

(A) x10. (B-C) x100. All the cells show the morphological characteristics of monocytes.

(D) Flow cytometry of the monocyte-rich fraction using anti-CD11b monocyte-specific marker (blue-filled curve). The isotype control is shown as a green-outline curve.

5.8.1 Induction of TF procoagulant activity in transgenic monocytes

Purified monocytes taken from transgenic mice were stimulated for 6 hours with LPS and PMA. Procoagulant function was then assessed in a TF-sensitive coagulation test. Dilutions of an extract of mouse brain in acetone provided a TF standard. Monocytes purified from whole blood kept at 4 °C for the same time period provided an unstimulated control.

Unstimulated monocytes had a clotting time of 120s, the same as the buffer blank of the standard curve (Figure 5.24). This indicated that there was no functional TF protein available in resting monocytes. After 6 hours of stimulation with LPS and PMA the clotting time was reduced to 76s (mean value from 10 measurements, range 81 – 72 s), indicating that TF was now present on the cell surface. From the graph it can be estimated that this decrease in the clotting time represents at least a fifty-fold increase in the amount of functional TF following stimulation with LPS and PMA.

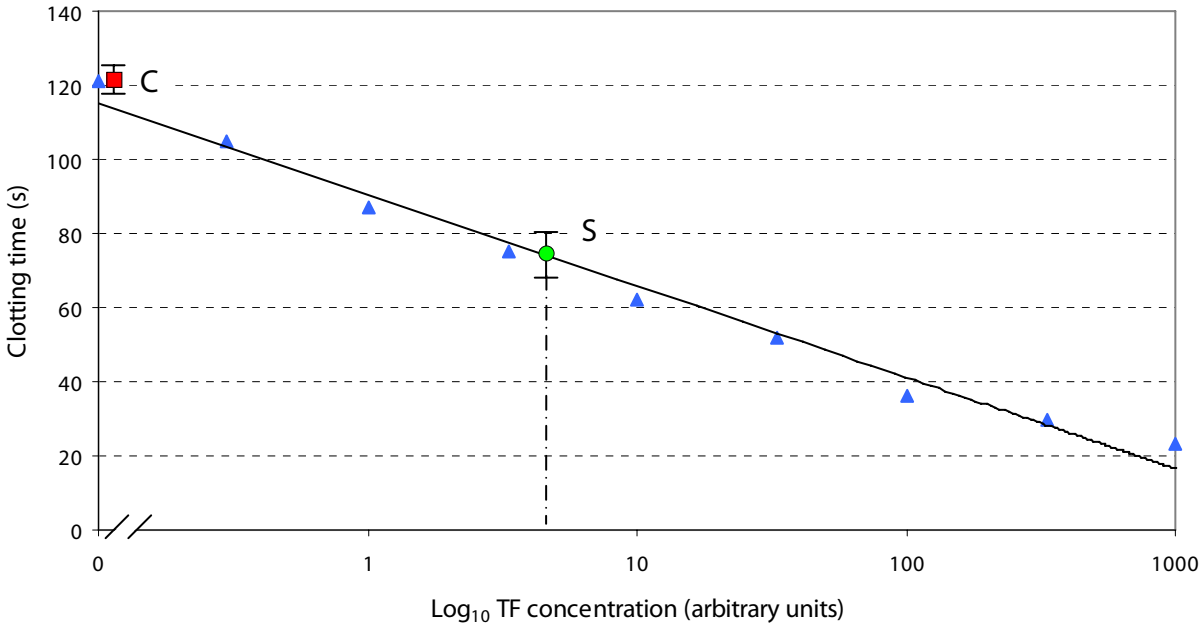


Figure 5.24 Graph of clotting time against TF concentration. Extract of mouse brain in acetone was diluted in TBS buffer to provide a TF standard curve (blue triangles). The clotting times obtained with control (C, red square) and LPS+PMA stimulated monocytes (S, green circle) are shown. Each data point represents the average of 10 measurements.

5.8.2 Analysis of TF-EGFP expression in transgenic monocytes

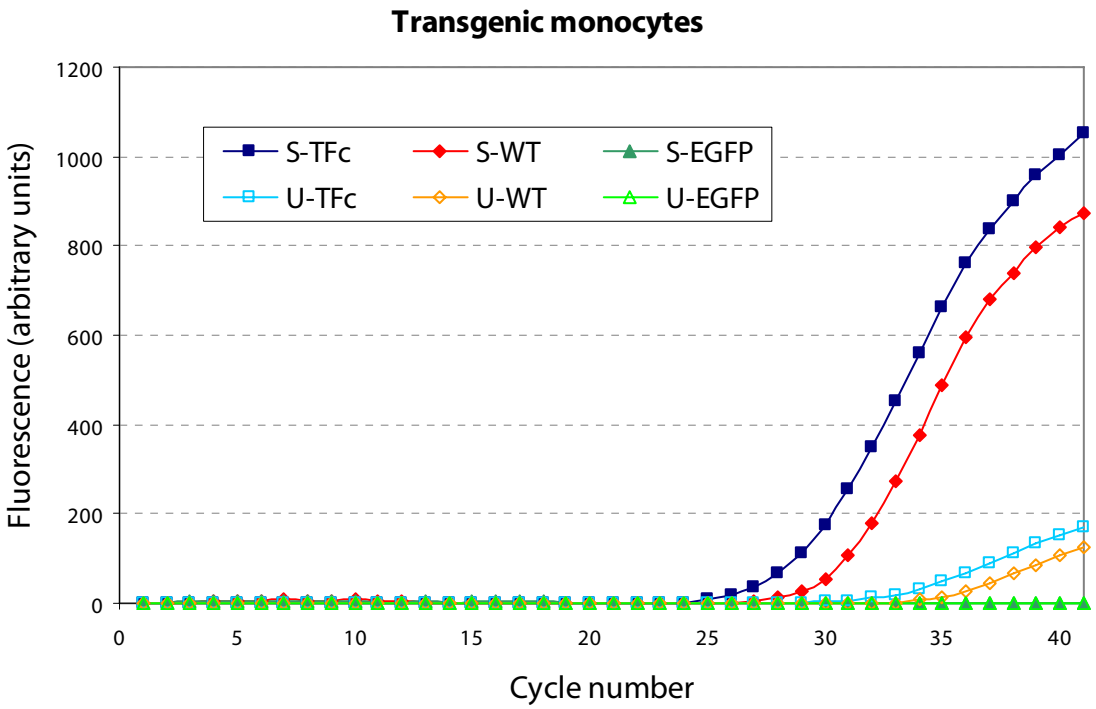
The results of the coagulation assay indicated that TF could be induced by LPS and PMA stimulation of transgenic monocytes. The real-time RT-PCR assays previously described were used to assess whether TF induction was preceded by an increase in transcription of both wild-type and targeted alleles. On this occasion, after 2 hours of stimulation half the sample was removed for purification of monocytes. RNA was extracted from these monocytes. The remainder of the sample was left under stimulation for a total of 6 hours. Monocytes were then purified and the clotting time tested as above to confirm that functional TF had been induced. As before the control was provided by whole blood kept at 4 °C.

A minimal amount of wild-type TF allele transcription was detectable in unstimulated monocytes (Figure 5.25, U-WT). The transcription level of the wild-type allele was significantly increased following LPS and PMA stimulation (Figure 5.25, S-WT). There was no detectable transcription of the targeted allele in either stimulated (Figure 5.25, S-EGFP) or unstimulated monocytes (Figure 5.25, S-EGFP).

5.8.3 Flow cytometry analysis of monocytes

To assess whether induction of TF procoagulant activity was accompanied by acquisition of EGFP fluorescence, monocytes were assessed by flow cytometry. Whole blood from transgenic mice was stimulated for 6 hours with LPS and PMA. Monocytes were purified and divided into two equal aliquots. One aliquot was used in a coagulation assay to confirm that TF procoagulant activity was present while the other was stained with CD11b and placed in the flow cytometer.

CD11b positivity confirmed that a monocyte rich fraction had been obtained but there was no detectable EGFP fluorescence in the stimulated monocytes (Figure 5.26). Although the coagulation test confirmed that TF procoagulant activity had been induced in comparison with the control monocytes the stimulated and unstimulated monocytes were indistinguishable in terms of EGFP fluorescence.



Transgenic RNA				Wild-type RNA			
Assay	Symbol	C_T	RE	Assay	Symbol	C_T	RE
S-TFc		26.7	0.97	U-TFc		33.2	0.87
S-WT		28.1	0.95	U-WT		34.9	0.83
S-EGFP		-	-	U-EGFP		-	-

Figure 5.25 Real-time RT-PCR quantitation of wild-type and targeted TF mRNAs from stimulated and unstimulated monocytes from transgenic mice. The graph shows accumulation of fluorescence with increasing PCR cycles. For each curve the first letter denotes the sample source (S = stimulated, U = unstimulated). The PCR assay is given after the dash (TFc = TF control, WT = wild-type allele specific, EGFP = EGFP-targeted allele specific). The table shows the cycle threshold (C_T) and reaction efficiency (RE) for each assay.

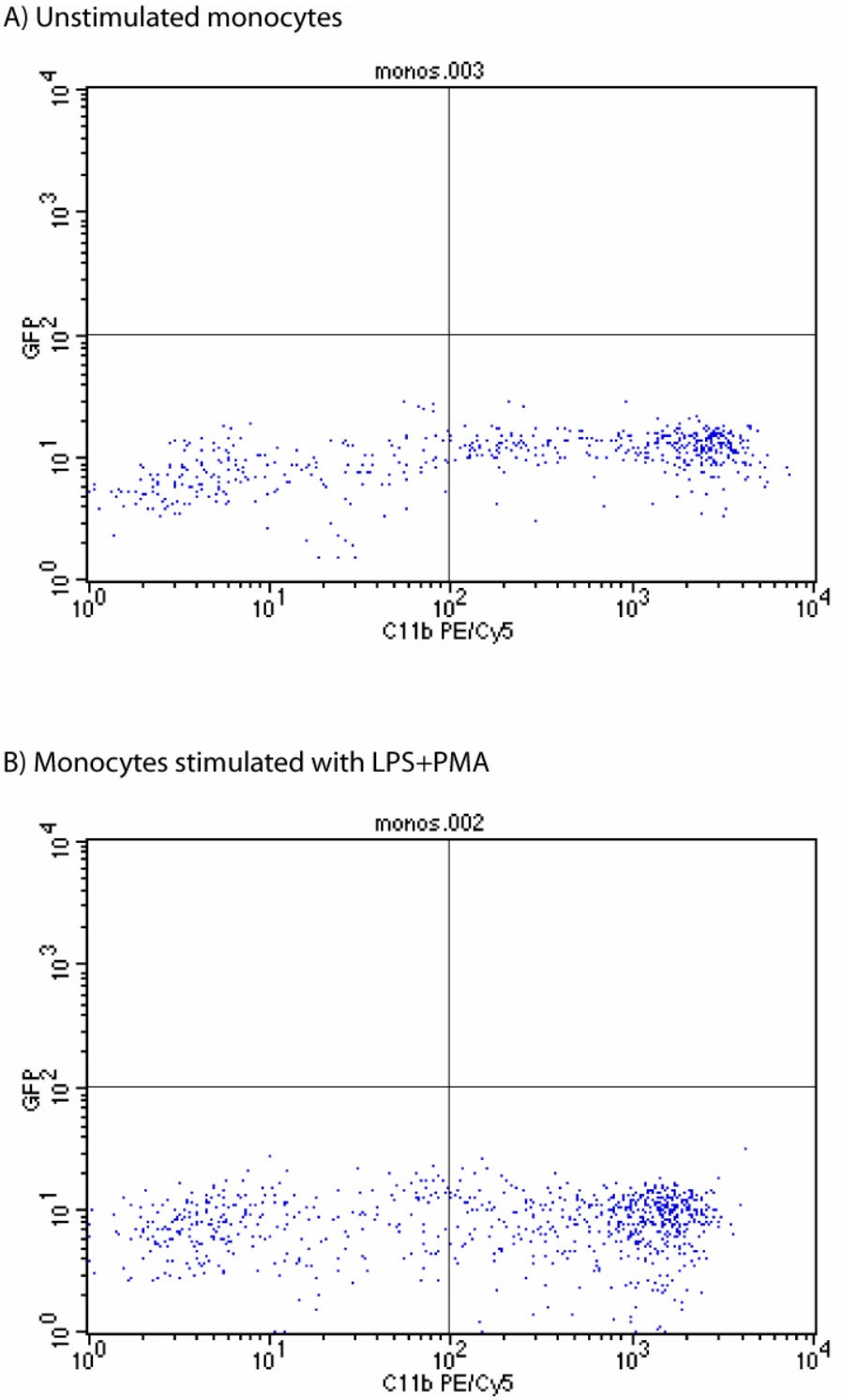


Figure 5.26 Dot plots from flow cytometry analysis of purified transgenic monocytes stained with the monocyte marker CD11b. Staining with CD11b is plotted against EGFP fluorescence. A) Control monocytes, B) Monocytes stimulated with LPS and PMA.

5.9 Discussion

5.9.1 Breeding of mice for analysis

One or the major limiting factors in analysis of the transgene was the time taken to breed sufficient numbers of transgenic mice. Mice have a short reproductive cycle of approximately 4 weeks in comparison to other commonly used transgenic animals such as rats or rabbits. However, their small size means that for experiments on blood components a number of animals are needed to provide a sufficient quantity of cells. In larger animals tissue samples may be removed by operative biopsy but this approach is not practical in mice because of the great technical difficulty and the small sample size that this would produce. Therefore experiments were carried out on organs from fully grown adult mice to maximise sample size.

This approach requires a large number of transgenic mice. In order to increase the sensitivity of the analysis it is preferable to perform the analysis on transgenic mice with 2 copies, rather than 1, of the transgene. However, the production of large numbers of double copy transgenic mice results in the generation of large numbers of unsuitable mice during a conventional breeding program. To try and minimize the time taken for production of these mice attempts were made to genotype progeny using the quantitative PCR assays developed for distinguishing the wild-type and targeted alleles. A limiting factor in these assays is the requirement for very high purity DNA samples. Contamination of the samples with protein or salts in particular markedly affects PCR efficiency. DNA samples taken from ear punch biopsy, tail biopsy and tail vein bleeds were all found to be unsuitable for producing reproducible quantitative PCR results. Although sufficiently high quality samples were obtained following cardiac puncture or whole tissue extraction, these require euthanasia of the subject.

The levels of coagulation factors may be affected by the variable genetic background found in different mouse strains. This may be due to polymorphisms within the genes themselves or caused by *trans*-acting factors. The analysis of expression levels of coagulation factors is therefore best carried out in a pure-bred genetic background. Gene targeting procedures are carried out in mixed strain mice because this generally yields better results in terms of the numbers of transgenic mice produced. However, there may then be variability between the expression levels of proteins from animals in the same litter. To reduce inter-subject variability

mice are back-crossed onto the desired background. This may take eight to ten generations before a sufficiently uniform background is obtained.

At the time of writing further breeding is ongoing to generate mice with two copies of the transgene on a pure C57BL6/J background.

5.9.2 Microscopic analysis of the TF-EGFP reporter model

It was clear from Southern blot analysis and PCR amplification that the targeted allele was correctly inherited during breeding in transgenic mice. It was not possible to fully sequence the entire targeted gene because of sequence identity with the wild-type alleles. Because of this it was not possible to differentially amplify the targeted gene away from the knocked-in sequence at the 3′ end. Specific amplification of the *mF3-EGFP* mRNA was possible and this demonstrated that the coding sequence was intact.

Ultimately the purpose of the TF-EGFP reporter model is to allow the direct visualisation of TF without the need for antibody-mediated detection systems. Attempts to visualise the reporter protein in a fluorescence microscope demonstrated some of the difficulties associated with this method.

GFP and its variants have been successfully used as reporter proteins in many studies of mammalian protein expression. In the vast majority of these studies the fluorescent protein is under the regulation of an exogenous promoter, often viral in origin. As these promoters are not under tight physiological control in mammalian cells, expression levels are relatively high and the fluorescent signal is easily detected. In this project EGFP fluorescence would be under tight regulation of an endogenous promoter and expression levels are expected to be low. It was clear from the images obtained from wild-type sections that there are various normal structures in mammalian tissues that autofluoresce within the bandwidth for EGFP emission. When the expression level of the target protein is low this causes difficulty in distinguishing the signal from the reporter from background fluorescence.

One method that is used to estimate background is to assess fluorescence at different wavelengths. When assessing background at the EGFP emission wavelength a comparison with emission in the channel used for TRITC detection is often used. However, this remains unsatisfactory because emission due to autofluorescence may well vary across the electromagnetic spectrum. Therefore, it is not certain that background fluorescence at a particular wavelength accurately predicts the amount of background that is present at a different wavelength. The most practical way of addressing this problem is by using a narrower bandwidth for

EGFP fluorescence detection. However, while this reduces the amount of background the signal intensity is also reduced so that weak signals may fall below the detection threshold. Attempts to use a narrower bandwidth for fluorescence microscopy in this project simply produced images with no detectable fluorescence. When visualising fluorescent proteins under endogenous control the sensitivity of the detection system may well be a limiting factor. Increasing the sensitivity leads to a reduction in the signal to noise ratio. In this particular case the separation of true positive signal from autofluorescence may only be possible in mice with two copies of the transgene.

5.9.3 Analysis of mF3-EGFP transcription and translation

The conclusion from the various functional analyses of the targeted allele is that transcription of the targeted allele did not lead to appropriate protein production in transgenic mice. RT-PCR assays demonstrated that the full-length *mF3-EGFP* mRNA was transcribed in whole tissues but not in kidney or following induction of TF in monocytes. Although the exact mechanisms regulating the expression of TF are not yet fully elucidated, it is clear that TF expression is restricted to a few specific cell types.

Regulation of TF expression may occur at transcription or post-transcription. After transcription the regulation of protein expression is generally achieved through post-translational pathways that modify the immature protein. TF distribution within cells occurs in a carefully controlled fashion. For example, it is detected at the cell membrane but there is little evidence for localisation in any cytoplasmic structures. This suggests that the protein is processed through quite specific post-translational pathways.

One possible explanation for the lack of TF-EGFP protein is that the presence of the EGFP tag interferes with the post-translational modification of TF preventing formation of the mature TF-EGFP protein. However, one would expect that if this were the case the incorrectly processed TF-EGFP protein would accumulate in some intracellular compartment in a degraded or partly degraded form. The Western blot analysis of whole tissue extracts analysed protein from all cellular components outside the nucleus. Therefore one would expect that a degraded protein would be detected as a band of unexpected size. It is unlikely that the degraded protein would be insoluble as both of the separate components (TF and EGFP) were solubilised in this project and the TF-EGFP protein had been solubilised in the same lysis buffer during previous work in this laboratory.

The most compelling evidence that the fusion of EGFP to TF does not have a detrimental effect on the target protein comes from the previous work detailed by Mumford (2003) and Fortin *et al.* (2005). The work by Mumford demonstrated that in CHO cells transfected with an *mF3-EGFP* fusion gene the TF-EGFP protein was expressed in similar fashion to the wild-type allele, while retaining procoagulant function and without any adverse effect on cell viability. TF-EGFP was distributed appropriately within the cell (see Figure 1.6). As CHO cells do not constitutively express TF there remains the possibility that EGFP might have a detrimental effect when the TF-EGFP allele is expressed in cells that have specific mechanisms for post-translation processing of TF. However, the study by Fortin *et al.*, (2005) described the expression of a TF-GFP fusion protein in cell types that have been shown to express TF *in vivo,* providing evidence against this. It is therefore unlikely that the lack of translation of *mF3-EGFP* transcripts into TF-EGFP was due to a negative effect of EGFP. In any case a defect in processing of the *mF3-EGFP* translation product can not explain the lack of transcription of the transgene in induced monocytes.

If the lack of production of the TF-EGFP protein was not due to an effect on post-translational pathways then the possibility that there was some fault in transcription of the transgene needs to be considered. The key result that needs to be explained here is that *mF3-EGFP* transcripts were detectable in whole tissue RNA preparations but not in those from induced monocytes.

The control of transcription is achieved through the action of various enhancers and repressors. The rate of TF transcription in a particular cell type is then determined by the balance of the effects of the specific enhancers and repressors acting in that cell type. Enhancers and repressors may have their effects by binding to regulatory elements that are adjacent to the coding sequence. Therefore the transgene will only be transcribed in a normal physiological manner if all the necessary regulatory elements are present.

It is evident that the major *cis*-acting regulatory elements must be intact in the transgene. Otherwise no transcripts from the transgene would have been detectable in whole tissue RNA preparations. The promoter and immediate 3′ regions of the *mF3* gene have been mapped indicating the binding sites of known transcription factors (Mackman *et al.*, 1992). However, the possibility of more distant regulatory sequences remains. The major drawback of the pro-nuclear injection strategy used for introducing the BAC into the fertilized oocytes is that the BAC is randomly sheared during integration into the mouse genome. This raises the

possibility that although the *mF3* coding sequence was intact there may have been disruption of some *cis*-acting regulatory elements.

Separation of the transgene from some of its regulatory elements might prevent transcription in some cell-types but not in others. This would be because of the differential effect of enhancers and repressors as described above. Ectopic transcription in particular, in cells which are not constitutively programmed to express TF, may well still be able to occur. The nature of ectopic transcription is that it can occur outside of the normal physiological mechanisms controlling expression of the gene concerned. This would account for the detection of *mF3-EGFP* transcripts in the real-time RT-PCR assays of whole tissue mRNA, where all cell types would contribute to the mRNA pool. However, separation of the gene from some of its regulatory elements is likely to prevent transcription in the cell types that would express TF *in vivo*. This could account for the presence of *mF3-EGFP* transcripts in RNA pools derived from a variety of cell types and the absence of the same transcripts in RNA preparations from induced monocytes.

If this explanation is correct then one would expect that the defect preventing production of the TF-EGFP is specific to the line of transgenic mice tested. Pro-nuclear injection with linearized BAC would be expected to result in integration of the BAC with a much lower chance of disruption of the targeted allele. This hypothesis will be tested with future experiments with a transgenic line derived from TG254.

5.10 Summary

Analysis of the TF-EGFP reporter model was carried out in transgenic mice generated from a founder mouse produced following pro-nuclear injection with a BAC in which the *EGFP* gene had been knocked-in to the *mF3* allele. Transmission of the targeted allele to the progeny of TG17 was demonstrated. Transcription of the targeted allele was demonstrated in whole tissue samples. However, it was not possible to conclusively demonstrate expression of TF-EGFP protein in this transgenic mouse line. It is proposed that this is due to disruption of the targeted allele preventing transcription of the targeted allele in appropriate cell types. It is likely that this defect is specific to this transgenic line and therefore unlikely to be repeated in separately derived transgenic mice.

Chapter 6 - Discussion

6.1 Project rationale

The purpose of this project was to generate a reporter model which would improve the detection of TF expression. This tool would then be used to further our understanding of the roles of TF in various physiological and pathological processes. The strategy chosen for generation of the reporter model was to genetically engineer the mouse genome to constitutively express a TF-EGFP reporter protein. The use of genetic engineering to create animal models remains a technically demanding process. Justification of this strategy comes from demonstration of the value of mouse models in this field and evidence that the proposed reporter model can be reasonably expected to realise the project's aim.

The basic structure of the haemostatic pathway is preserved across all jawed vertebrates (Davidson *et al.*, 2003). Consistent with this is the finding that animal models of haemostasis correlate well with observations on human haemostasis *in vivo*. The phenotypes produced by genetically engineered manipulations of coagulation factor levels in animals closely mirror clinical manifestations of human deficiencies. TF deficiency has not been described in humans, presumably because the phenotype would be lethal *in utero* as shown in mice. However, studies of TF in various pathological processes in humans, such as metastasis, correlate well with data from mouse models of diseases. These models have therefore improved our understanding of the functions of TF in humans.

The use of reporter models in demonstrating patterns of protein expression was discussed in chapter 1. In a dynamic process such as haemostasis, a critical feature of the reporter model is that it should accurately delineate protein expression without interfering with the process under investigation. This is an issue which is likely to limit the value of immunological detection strategies or any assay that interferes with function *in vivo*. A TF C-terminus fusion protein is unlikely to interfere with TF function because of the clear data demonstrating that the C-terminus is not required for TF function in normal physiology (Parry & Mackman, 2000; Melis *et al.*, 2001).

Thus a reporter model has advantages over conventional detection methods in demonstrating TF expression. The choice of GFP as the reporter is supported by the increasing data describing the use of this protein in a wide variety of different physiological pathways (reviewed in Tsien, 1998; and Hadjantonakis & Nagy, 2001).

Specific data on the suitability of EGFP to act as a reporter for TF, without interfering with TF function, comes from studies of a TF-EGFP reporter in cell lines *in vitro* (Mumford, 2003; Fortin *et al.*, 2005).

6.2 Strategies for production of transgenic mice

Although a reporter model carries significant advantages over other detection methods for single protein expression detection, its major drawback is the technical difficulty involved in generation of the model. The strategy initially employed in this project involved targeting in ES cells to create the *mF3-EGFP* fusion gene. This remains the most widely used method for manipulating the mammalian genome. The process is technically demanding, relatively time-consuming and very inefficient. In most successful targeting projects only a handful of correctly targeted ES cell clones are produced. It is clear that the success of a targeting project is heavily dependent on the efficiency of the targeting process. This is ultimately measured by the number of targeted clones generated (Godwin *et al.*, 1998).

6.2.1 Generation of transgenic mice using targeted ES cell clones

The first part of this project described the difficulties encountered in developing a sufficiently efficient ES cell targeting strategy. Many of the factors affecting targeting efficiency are unalterable. For example, targeting efficiency is highly locus-specific and knock-in projects in particular have very little flexibility in the selection of homologous regions. If the sequences in question are poor substrates for HR then there is currently little that can be done to improve this. Various strategies are currently under investigation to improve the efficiency of HR both in a genome-wide manner and at specific loci (Knauert & Glazer, 2001; Vasquez *et al.*, 2001; Carroll, 2004). However, on account of our limited understanding of HR regulation, these strategies are still some way from widespread use.

Of the factors influencing HR at a specific locus the only readily modifiable one is the length of the HAs. Lengthening of the 5' HA proved successful in this project at increasing the efficiency of HR and ultimately increasing the number of targeted clones. Although chimaeric mice were produced from different targeted ES cell clones, to date breeding experiments with these chimaeras have proved unsuccessful in producing transgenic pups. This is likely to be on account of a relatively poor representation of targeted cells in the chimaeras' germline. The normal method for assessing the contribution that targeted ES cells make to the germline is to estimate the percentage of agouti colouring in the chimaeras' coat.

During embryogenesis the reproductive organs are derived from the mesoderm, while the coat is derived from the ectoderm. Although it is possible that the representation of ES cells might differ between the two layers, the experience of most researchers is that the coat colour of a chimaeric animal is a good indicator of the potential for siring a transgenic pup. The experience in this institute is that transgenic pups are rarely produced from chimaeras with less than 75% agouti coat colouring (personal communication, J. Godwin, ES cell and Transgenics Laboratory). Unfortunately the chimaeric mice produced in this project had no more than 20% agouti coat colouring.

The difficulty encountered in this project in the generation of a transgenic mouse line from targeted ES cell clones is not uncommon in gene targeting projects. The process of electroporation followed by selection and expansion of targeted clones has a detrimental effect on the ES cells. This effect is greater than can be explained simply by senescence associated with repeated cell division. ES cells that are grown under optimal conditions without toxic additives can maintain their pluripotency at higher passages than encountered here. The adverse effects of positive and negative selection were most clearly seen in the E14 cell line. This line has been successfully used in targeting projects both in this institute and others for many years. In comparison with the other cell lines used the stocks were at a higher passage. Karyotypic instability was more evident in the E14 cells following clone selection than in the other ES cell lines.

Various factors may affect the germline potential of ES cells (reviewed in Koller & Smithies, 1992; and Prelle *et al.*, 2002). Much of our knowledge of these comes from studies of ES cell differentiation (reviewed in Desbaillets *et al.*, 2000). While the factors promoting differentiation are well characterised for some cell types, those that inhibit the contribution of pluripotent ES cells to the germline in an inner cell mass are less well characterised. For this reason assessment of ES cell germline potential remains relatively crude. By convention ES cells are assessed by microscopic examination of chromosomal number during metaphase, in the manner described in this project. This method can detect gross abnormalities that would certainly be detrimental, but would leave more subtle changes undetected. Of course, there are more stringent methods for detecting minor chromosomal aberrations. However, implementation of these would only be of value following studies of the impact of minor chromosomal variation on germline contribution. Without this differentiating polymorphic variation between different mouse strains from mutations is difficult. Even if assessment of the ES cell genotype was carried out

in fine detail on a genome wide scale, this would only for account for some of the factors influencing pluripotency. ES cell phenotype is heavily dependent on non-genetic factors such as the constituents of the culture medium. Subtle variations in phenotype which are undetectable to researchers can limit pluripotency. With very careful control of the culture milieu, ES cells may be kept in a pluripotent state through multiple passages (Prelle *et al.*, 2002). However, the effects of environmental factors on ES cell pluripotency have been largely investigated in experiments with untargeted wild-type cells. The targeting process and the need for selection agents in the culture medium may adversely affect pluripotency.

Cryptogenic abnormalities in the targeted ES cells may be a major impediment to the generation of high percentage agouti chimaeras. Additionally the blastocyst injection and re-implantation procedures are areas where a variety of factors are critically important. These procedures are technically demanding and success is heavily dependent on the experience within an institute with specific ES cell lines. For this reason laboratories tend to persist with ES cells from a particular line as the appropriate injection protocols vary between different lines. For example, the correct number of targeted ES cells that should be injected into the inner cell mass of blastocysts is dependent on the cell line. Generally the expectation is that if more cells can be introduced the relative contribution to the germline will be increased. However, increasing cell numbers in the inner cell mass increases the blastocyst volume leading to a reduction in viability. The correct number of cells for a particular line has to be determined empirically. For this reason the number of cells injected was varied in different experiments with the EL M3 cell line (see Table 3.3).

The preferred option is therefore to use ES cells that have been previously used successfully within an institute. Unfortunately, ES cells from a particular line are a limited resource and cannot be indefinitely expanded while retaining pluripotency. This is largely due to the problems of senescence associated with repeated cell division as described previously.

In addition to embryo viability a variety of factors can affect the outcome of pregnancy following blastocyst re-implantation. In particular environmental conditions, such as temperature in the animal house, can have a significant effect on the miscarriage rate. Different mouse strains vary in their suitability as foster mothers. The experience in this institute, in keeping with observations elsewhere, is that the C57BL/6J strain, from which the blastocysts are derived, produces relatively poor foster mothers as assessed by the number of live pups produced. For this reason the embryos were re-implanted into foster mothers bred from the CBA strain.

In summary, there are a number of potential explanations to account for the inability to produce a transgenic mouse line from correctly targeted ES cells. It is quite possible that a combination of factors amongst those described above were responsible. Determination of the factors adversely affecting the production of transgenic pups would be time-consuming and quite possibly unrewarding as some of the potential factors are not easily modifiable. Therefore, an alternative strategy was pursued in parallel to generate transgenic mice.

6.2.2 Generation of transgenic mice using a targeted BAC

Targeting of a BAC containing the *mF3* gene in a bacterial host is considerably easier than targeting in ES cells. A direct diagrammatical comparison of the two procedures can be seen in Figure 6.1 and the relative advantages and disadvantages of each strategy are listed below.

1. In both strategies, knock-in of *EGFP* at the *mF3* locus occurs via HR, which follows introduction of targeting construct DNA by electroporation. The efficiency of HR is very much greater in bacteria so that in this project the overall targeting efficiency was almost 70 % in bacteria and less than 2 % in ES cells.
2. The selection and expansion of HR clones can be performed in a few days in bacteria while this process takes several weeks in ES cells. This is because bacteria divide more rapidly and are easier to grow. Screening of clones is relatively simple because the high overall targeting efficiency means that sufficient numbers of clones are identified after a single experiment. Because very few correct targeting events occur in ES cells the screening strategy has to have a high negative predictive value. In this project ES cells were screened using a combination of Southern blot and PCR analysis which is much more time-consuming than PCR alone. Bacteria were screened by PCR alone, although selected clones were also analysed by Southern blot.
3. Both pro-nuclear injection and blastocyst injection are techniques requiring considerable technical skill. Re-implantation of fertilized embryos and blastocysts is similarly demanding. Therefore both strategies suffer from a poor return in terms of live pups relative to the number of re-implanted embryos

4. The major difference between the two strategies is that in ES cells one of the two endogenous loci is targeted, while in bacteria the targeted locus is exogenous and becomes randomly integrated in the genome. This is the major disadvantage of targeting in bacteria. Random integration in the genome carries with it the possibility of disruption of other genes and/or disruption of the targeted gene. Additionally in genes where the total amount of transcription product is carefully regulated there may be competition between all contributing alleles, so that a single transgene is disadvantaged when competing with two wild-type genes.

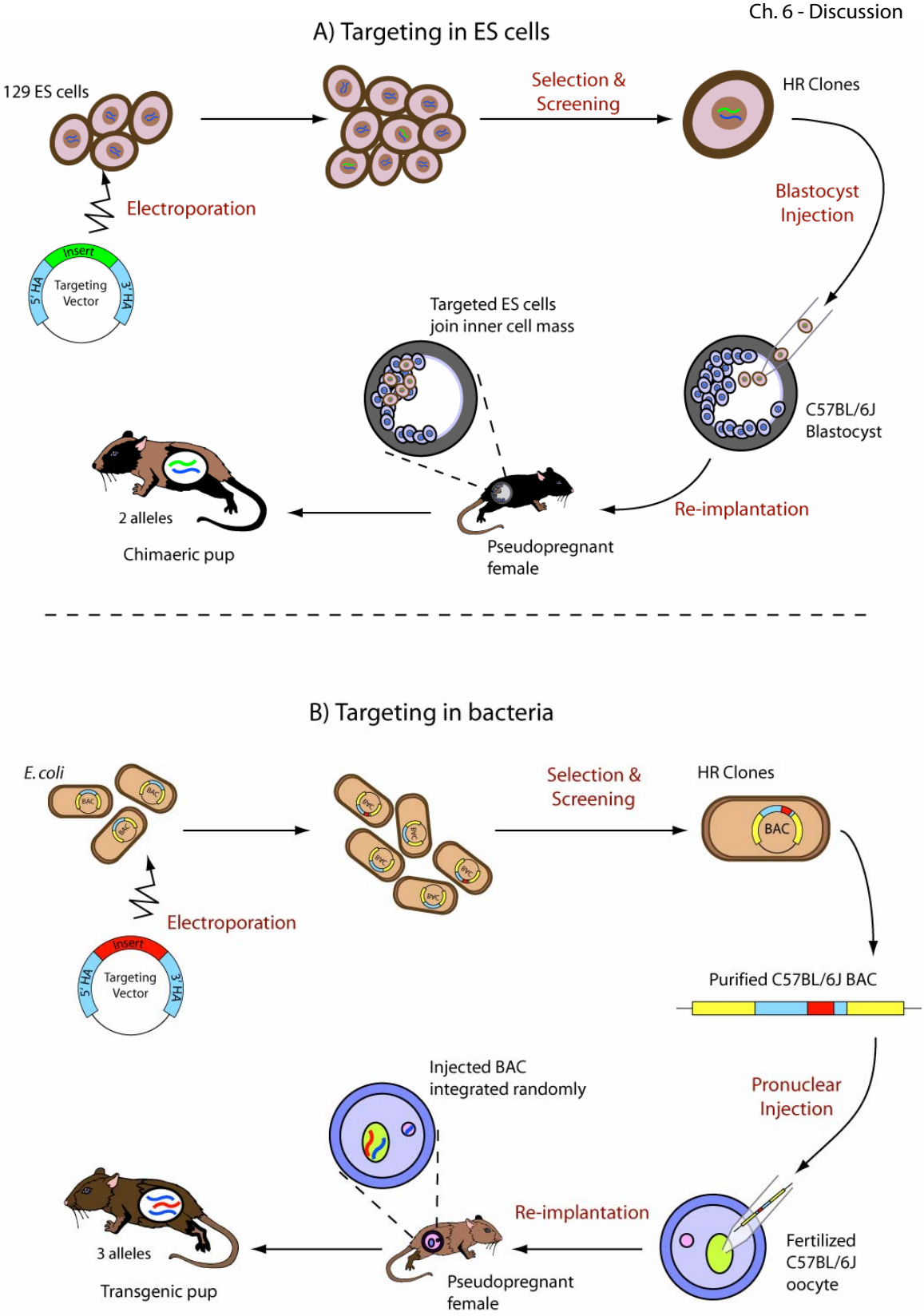


Figure 6.1 Comparison of the two strategies used for the generation of transgenic mice. Wavy lines in pups indicate different types of allele: blue = wild-type, green = targeted in BAC, red = targeted endogenous allele. A) Targeting in ES cells produces chimaeric pups with a single wild-type allele and a single targeted allele. B) Targeting in a BAC results in pups with two wild-type alleles and a single targeted allele that is randomly inserted in the genome.

6.3 Analysis of the reporter model in the TG17 transgenic line

Analysis of the reporter model was carried out in a line of transgenic mice produced from a single mouse (TG17) carrying the targeted allele. The results of this analysis showed that although mRNA produced from the transgene was detectable in whole tissue extracts, this did not translate into production of the fusion protein. When the analysis was restricted to a cell type with well characterised inducible TF expression, it was found that there was no production of mRNA from the transgene even though the wild-type allele was adequately expressed.

An explanation of these results needs to account for a difference of transcription of the transgene between cell types, which was not mirrored by a difference in transcription of the endogenous allele in the same samples. When transcription did occur there was no detectable protein produced from the transgene. The possibility that the fusion of EGFP onto the cytoplasmic tails of TF had a detrimental effect was considered. The results from studies of TF-EGFP expression under control of an exogenous promoter suggest that addition of EGFP does not have a detrimental effect. Expression of the fusion protein was seen in cell types that would not normally express TF and also those that may do so *in vivo*. Nevertheless the possibility remains that if the fusion protein were under control of the endogenous promoter rather than the exogenous promoter used in these studies a negative effect on expression might result.

An alternative explanation for the results described in chapter 5 is that during integration of the transgene into the genome the regulatory elements were disrupted such that transcription could occur in some cell types but not others. It was postulated that transcription might be occurring only in cell types that would not normally express TF and therefore lack the necessary post-translational pathways to produce a mature TF-EGFP protein. Although the TF promoter region has been thoroughly mapped it is still not clear exactly how TF expression is differentially regulated between different cell types. An example of this is seen in the work by Mackman and colleagues in which TF null mice are rescued by a minigene expressing human TF (Parry *et al.*, 1998). In this study it was found that inclusion of the region1.5 kb upstream of the gene did not result in adequate expression of the minigene despite the fact that this region contained all the transcription factor binding sites that were thought to be required for TF expression. The minigene was only expressed when more than 2 kb of the upstream sequence and part of intron 1 were included in the targeting vector. The additional regions do not contain any known

transcription factor binding sites and yet they clearly pay an important part in TF expression in some circumstances.

If the problems encountered in expression were due to disruption of the transgene regulatory elements during targeting, it would be expected that the abnormality is specific to the TG17 transgenic line. Transgenic mice generated using a linearized BAC in the pro-nuclear injection step would not be expected to have the same abnormality.

6.4 Design of reporter model

The use of reporter genes to indicate transcription of a gene of interest is now well established. A number of different reporter genes are available with distinct advantages and disadvantages. Selection of the reporter gene is dependent upon consideration of these in the context of the specific requirements of the proposed model. The key requirements of a reporter model capable of demonstrating TF expression dynamically *in vivo* are:

1. Accuracy over a period of time, such that signal does not lag behind expression or persist following removal of TF.
2. Lack of interference with the functions of TF under study.
3. Sensitivity to small changes in signal as TF is expressed at relatively low levels.

As analysis of the TF-GFP reporter model generated during this project has thus far been unable to demonstrate TF expression, it is important to consider whether alternative reporter genes might be more likely to satisfy these criteria.

There are two main categories of reporter gene that are suitable for insertion at endogenous loci. These are the enzymes that generate light or colour following proteolysis of an appropriate substrate and the fluorescent proteins that emit light following excitation at defined wavelengths. Amongst the fluorescent proteins variants of GFP remain the most widely used. The major advantage that this group has over other fluorescent proteins, such as variants of DsRed, is that the protein is relatively small and functions as a monomer. DsRed, in contrast, is an obligate tetramer which means it is more likely to interfere with TF function. The major drawback of fluorescent proteins is that, in comparison with enzymes, they provide a relatively low ratio of emitted photons per molecule. This is a particular issue with GFP because the peak emission wavelength overlaps with the bandwidth affected by

autofluorescence. This leads to reduced signal intensity because of increased background as demonstrated in chapter 5. In this respect DsRed has an advantage as its emission wavelength is at the red end of the spectrum which is less affected by autofluorescence (reviewed in Contag & Bachmann, 2002; Hadjantonakis *et al.*, 2003). In histological specimens for fluorescent detection fixation methods have to provide a compromise between preservation of tissue architecture and maintenance of the fluorescent signal. Paraffin embedding provides excellent preservation of architecture but significant loss of fluorescence. For this reason fluorescence was detected in frozen sections using methods that have been proven to preserve GFP signal.

The two main enzymes used as reporters are β-galactosidase and members of the luciferase family. Although β-galactosidase produces a strong colour change that amplifies native signal it is only suitable for use in fixed tissues. The bioluminescent luciferases have been widely used in live cell cultures and, to a lesser extent, living tissues. Sensitivity is greater than that seen with GFP because there is very little natural luminescence in mammalian tissues reducing the background signal. The major drawback is the requirement for a specific substrate to generate light. Although available substrates have good penetration into a variety of cell types in culture this is a concern in a live, dynamic system. If used as a TF reporter the substrate would have to have adequate and consistent penetration to all cell types *in vivo*.

In the vast majority of studies in the literature, expression of the reporter gene linked to the gene of interest is driven by strong viral promoters. Where the native promoter is under study, the strategy is often to place this in frame with the reporter gene in a construct which is then introduced into the required cells. Thus a combination of high copy number and/or strong promotion produces strong signal. There is much less experience in the use of reporter genes under endogenous control, but it is clear that sensitivity is a major issue. There is more data on the use of GFP in this situation than for other reporters, which is one reason for the choice of this protein as the reporter in this project. It may be that the bioluminescent enzymes may provide a viable alternative. Further work would be required to investigate whether the increase in sensitivity that these would provide would be accompanied by the necessary specificity.

6.5 Future work

In order to establish the cause of the lack of expression of the TF-EGFP reporter model in the TG17-derived mouse line, it is necessary to generate further transgenic mouse lines. Experimental data produced in this and other labs indicates that fusion of the EGFP moiety onto the intact TF cytoplasmic tail does not affect the normal expression or function of TF. Therefore, it is anticipated that production of further transgenic mice will result in a working TF-EGFP reporter model.

At the time of writing this work is proceeding along two lines. The results of ES cell targeting demonstrated that the extended targeting construct (pES-TC-L) was capable of producing targeted ES clones with an acceptable overall targeting efficiency. The subsequent failure in the production of mice with a high level of germline contribution from the targeted cells is likely to be caused by a combination of abnormalities in the targeted EL M3 clones and insufficient experience in the use of these specific cells in blastocyst injections.

There are now a number of commercial laboratories which specialise in the production of transgenic mice. The pES-TC-L construct will be used to generate new ES cell clones in a commercial laboratory. This should overcome any potential deficiencies in supplied ES cells or problems caused by lack of experience with particular ES cells. Targeting in ES cells remains the preferred method for the generation of transgenic mice because of the advantages of this method described above.

In case further attempts at ES cell targeting are unsuccessful, the second transgenic mouse, TG254, that was generated using linearized targeted BAC is currently entered into a breeding program. Linearization of the BAC at a site distant from the TF locus should significantly reduce the chance of disruption of the gene or *cis*-acting regulatory elements that might occur upon random integration of the BAC. Therefore transgenic mice derived from TG254 should not display the same abnormality of TF-EGFP protein production that was seen in the mice derived from TG17.

Ultimately, in transgenic mice produced by pronuclear injection of targeted BAC, successful expression of the TF-EGFP reporter model may occur at a relatively low level due to competition from the existing wild-type alleles. As discussed in chapter 5, fluorescence microscopy for the detection of EGFP may be hampered by the relatively low levels of expression seen when the EGFP fusion protein is under control of a tightly regulated endogenous promoter. Therefore it would be desirable

to reduce the possibility of a further reduction in levels of the reporter protein caused by competition with the wild-type alleles. To achieve this the transgenic mice will be crossed with heterozygous TF knockout mice (Parry *et al.*, 1998). This will allow the production of a transgenic line that is homozygous for the TF knockout but carries 2 copies of the *mF3-EGFP* transgene. Expression of the reporter model in this mouse line should be similar to that expected from the homozygous TF-EGFP mice that would be produced following ES cell targeting.

In the 4 years that have elapsed since the start of this project, there has been an increase in our knowledge of the various functions of TF. However, the techniques available for demonstration of TF expression remain essentially unchanged from the early 1990s and there remains a need for improvement in this area. The work outlined above is designed to utilise the advances made during this project to meet this need.

Reference List

Aasrum M & Prydz H (2002) Gene targeting of tissue factor, factor X, and factor VII in mice: their involvement in embryonic development. *Biochemistry (Mosc)* **67**, 25-32.

Abe K, Hazama M, Katoh H, Yamamura K & Suzuki M (2004) Establishment of an efficient BAC transgenesis protocol and its application to functional characterization of the mouse Brachyury locus. *Exp Anim* **53**, 311-320.

Abe K, Shoji M, Chen J, Bierhaus A, Danave I, Micko C, Casper K, Dillehay DL, Nawroth PP & Rickles FR (1999) Regulation of vascular endothelial growth factor production and angiogenesis by the cytoplasmic tail of tissue factor. *Proc Natl Acad Sci U S A* **96**, 8663-8668.

Ahamed J, Belting M & Ruf W (2005) Regulation of tissue factor-induced signaling by endogenous and recombinant tissue factor pathway inhibitor 1. *Blood* **105**, 2384-2391.

Ahamed J, Versteeg HH, Kerver M, Chen VM, Mueller BM, Hogg PJ & Ruf W (2006) Disulfide isomerization switches tissue factor from coagulation to cell signaling. *Proc Natl Acad Sci U S A* **103**, 13932-13937.

Albrecht C, McVey JH, Elliott JI, Sardini A, Kasza I, Mumford AD, Naoumova RP, Tuddenham EG, Szabo K & Higgins CF (2005) A novel missense mutation in ABCA1 results in altered protein trafficking and reduced phosphatidylserine translocation in a patient with Scott syndrome. *Blood* **106**, 542-549.

Ausubel F, Brent R, Kingston R, Moore D, Seidman J, Smith J & Struhl K (2006) *Current Protocols in Molecular Biology*: John Wiley & Sons, Inc.

Ayares D, Chekuri L, Song KY & Kucherlapati R (1986) Sequence homology requirements for intermolecular recombination in mammalian cells. *Proc Natl Acad Sci U S A* **83**, 5199-5203.

Bach R, Konigsberg WH & Nemerson Y (1988) Human tissue factor contains thioester-linked palmitate and stearate on the cytoplasmic half-cystine. *Biochemistry* **27**, 4227-4231.

Bach R, Nemerson Y & Konigsberg W (1981) Purification and characterization of bovine tissue factor. *J Biol Chem* **256**, 8324-8331.

Bach RR (1998) Mechanism of tissue factor activation on cells. *Blood Coagul Fibrinolysis* **9 Suppl 1**, S37-S43.

Bach RR (2006) Tissue Factor Encryption. *Arterioscler Thromb Vasc Biol*.

Bachli E (2000) History of tissue factor. *Br J Haematol* **110**, 248-255.

Bajaj MS, Birktoft JJ, Steer SA & Bajaj SP (2001) Structure and biology of tissue factor pathway inhibitor. *Thromb Haemost* **86**, 959-972.

Banner DW, D'Arcy A, Chene C, Winkler FK, Guha A, Konigsberg WH, Nemerson Y & Kirchhofer D (1996a) The crystal structure of the complex of blood coagulation factor VIIa with soluble tissue factor. *Nature* **380**, 41-46.

Banner DW, D'Arcy A, Chene C, Winkler FK, Guha A, Konigsberg WH, Nemerson Y & Kirchhofer D (1996b) The crystal structure of the complex of blood coagulation factor VIIa with soluble tissue factor. *Nature* **380**, 41-46.

Baugh RJ, Broze GJ, Jr. & Krishnaswamy S (1998) Regulation of extrinsic pathway factor Xa formation by tissue factor pathway inhibitor. *J Biol Chem* **273**, 4378-4386.

Bayani J & Squire JA (2001) Advances in the detection of chromosomal aberrations using spectral karyotyping. *Clin Genet* **59**, 65-73.

Belting M, Dorrell MI, Sandgren S, Aguilar E, Ahamed J, Dorfleutner A, Carmeliet P, Mueller BM, Friedlander M & Ruf W (2004) Regulation of angiogenesis by tissue factor cytoplasmic domain signaling. *Nat Med* **10**, 502-509.

Bogdanov VY, Balasubramanian V, Hathcock J, Vele O, Lieb M & Nemerson Y (2003) Alternatively spliced human tissue factor: a circulating, soluble, thrombogenic protein. *Nat Med* **9**, 458-462.

Bogdanov VY, Kirk RI, Miller C, Hathcock JJ, Vele S, Gazdoiu M, Nemerson Y & Taubman MB (2006) Identification and characterization of murine alternatively spliced tissue factor. *J Thromb Haemost* **4**, 158-167.

Bonfield JK, Smith K & Staden R (1995) A new DNA sequence assembly program. *Nucleic Acids Res* **23**, 4992-4999.

Brambrink T, Hochedlinger K, Bell G & Jaenisch R (2006) ES cells derived from cloned and fertilized blastocysts are transcriptionally and functionally indistinguishable. *Proc Natl Acad Sci U S A* **103**, 933-938.

Branda CS & Dymecki SM (2004) Talking about a revolution: The impact of site-specific recombinases on genetic analyses in mice. *Dev Cell* **6**, 7-28.

Brinster RL, Chen HY, Trumbauer ME, Yagle MK & Palmiter RD (1985) Factors affecting the efficiency of introducing foreign DNA into mice by microinjecting eggs. *Proc Natl Acad Sci U S A* **82**, 4438-4442.

Broze GJ, Jr. (1995) Tissue factor pathway inhibitor. *Thromb Haemost* **74**, 90-93.

Bugge TH, Xiao Q, Kombrinck KW*, et al.* (1996) Fatal embryonic bleeding events in mice lacking tissue factor, the cell-associated initiator of blood coagulation. *Proc Natl Acad Sci U S A* **93**, 6258-6263.

Butenas S, Bouchard BA, Brummel-Ziedins KE, Parhami-Seren B & Mann KG (2005) Tissue factor activity in whole blood. *Blood* **105**, 2764-2770.

Callander NS, Varki N & Rao LV (1992) Immunohistochemical identification of tissue factor in solid tumors. *Cancer* **70**, 1194-1201.

Camerer E, Huang W & Coughlin SR (2000) Tissue factor- and factor X-dependent activation of protease-activated receptor 2 by factor VIIa. *Proc Natl Acad Sci U S A* **97**, 5255-5260.

Camerer E, Rottingen JA, Gjernes E, Larsen K, Skartlien AH, Iversen JG & Prydz H (1999) Coagulation factors VIIa and Xa induce cell signaling leading to up-regulation of the egr-1 gene. *J Biol Chem* **274**, 32225-32233.

Capecchi MR (1989c) Altering the genome by homologous recombination. *Science* **244**, 1288-1292.

Capecchi MR (1989b) Altering the genome by homologous recombination. *Science* **244**, 1288-1292.

Capecchi MR (1989a) Altering the genome by homologous recombination. *Science* **244**, 1288-1292.

Carmeliet P, Mackman N, Moons L, *et al.* (1996) Role of tissue factor in embryonic blood vessel development. *Nature* **383**, 73-75.

Carroll D (2004) Using nucleases to stimulate homologous recombination. *Methods Mol Biol* **262**, 195-207.

Celi A, Pellegrini G, Lorenzet R, De BA, Ready N, Furie BC & Furie B (1994) P-selectin induces the expression of tissue factor on monocytes. *Proc Natl Acad Sci U S A* **91**, 8767-8771.

Chen D, Weber M, Shiels PG, Dong R, Webster Z, McVey JH, Kemball-Cook G, Tuddenham EG, Lechler RI & Dorling A (2006a) Post-injury vascular intimal hyperplasia in mice is completely inhibited by CD34+bone marrow-derived progenitor cells expressing membrane-tethered anticoagulant fusion proteins. *J Thromb Haemost*.

Chen D, Weber M, Shiels PG, Dong R, Webster Z, McVey JH, Kemball-Cook G, Tuddenham EG, Lechler RI & Dorling A (2006b) Post-injury vascular intimal hyperplasia in mice is completely inhibited by CD34+bone marrow-derived progenitor cells expressing membrane-tethered anticoagulant fusion proteins. *J Thromb Haemost*.

Cline J, Braman JC & Hogrefe HH (1996) PCR fidelity of pfu DNA polymerase and other thermostable DNA polymerases. *Nucleic Acids Res* **24**, 3546-3551.

Contag CH & Bachmann MH (2002) Advances in in vivo bioluminescence imaging of gene expression. *Annu Rev Biomed Eng* **4**, 235-260.

Corti R & Badimon JJ (2002) Biologic aspects of vulnerable plaque. *Curr Opin Cardiol* **17**, 616-625.

Coughlin S (2000) Thrombin signalling and Protease-Activated Receptors. *Nature* **407**, 258-264.

Creasey AA, Chang AC, Feigen L, Wun TC, Taylor FB, Jr. & Hinshaw LB (1993) Tissue factor pathway inhibitor reduces mortality from Escherichia coli septic shock. *J Clin Invest* **91**, 2850-2860.

Crossman DC, Carr DP, Tuddenham EG, Pearson JD & McVey JH (1990) The regulation of tissue factor mRNA in human endothelial cells in response to endotoxin or phorbol ester. *J Biol Chem* **265**, 9782-9787.

D'Andrea D, Ravera M, Golino P, *et al.* (2003) Induction of tissue factor in the arterial wall during recurrent thrombus formation. *Arterioscler Thromb Vasc Biol* **23**, 1684-1689.

Davidson CJ, Tuddenham EG & McVey JH (2003) 450 million years of hemostasis. *J Thromb Haemost* **1**, 1487-1494.

del Conde I, Shrimpton CN, Thiagarajan P & Lopez JA (2005) Tissue-factor-bearing microvesicles arise from lipid rafts and fuse with activated platelets to initiate coagulation. *Blood* **106**, 1604-1611.

del Zoppo GJ, Yu JQ, Copeland BR, Thomas WS, Schneiderman J & Morrissey JH (1992) Tissue factor localization in non-human primate cerebral tissue. *Thromb Haemost* **68**, 642-647.

Deng C & Capecchi MR (1992) Reexamination of gene targeting frequency as a function of the extent of homology between the targeting vector and the target locus. *Mol Cell Biol* **12**, 3365-3371.

Desbaillets I, Ziegler U, Groscurth P & Gassmann M (2000) Embryoid bodies: an in vitro model of mouse embryogenesis. *Exp Physiol* **85**, 645-651.

Dietzen DJ, Jack GG, Page KL, Tetzloff TA, Hall CL & Mast AE (2003) Localization of tissue factor pathway inhibitor to lipid rafts is not required for inhibition of factor VIIa/tissue factor activity. *Thromb Haemost* **89**, 65-73.

Drake TA, Morrissey JH & Edgington TS (1989) Selective cellular expression of tissue factor in human tissues. Implications for disorders of hemostasis and thrombosis. *Am J Pathol* **134**, 1087-1097.

Eilertsen KE & Osterud B (2005) The role of blood cells and their microparticles in blood coagulation. *Biochem Soc Trans* **33**, 418-422.

Ellgaard L & Ruddock LW (2005) The human protein disulphide isomerase family: substrate interactions and functional properties. *EMBO Rep* **6**, 28-32.

Elsliger MA, Wachter RM, Hanson GT, Kallio K & Remington SJ (1999) Structural and spectral response of green fluorescent protein variants to changes in pH. *Biochemistry* **38**, 5296-5301.

Eustice DC & Wilhelm JM (1984) Mechanisms of action of aminoglycoside antibiotics in eucaryotic protein synthesis. *Antimicrob Agents Chemother* **26**, 53-60.

Falati S, Gross P, Merrill-Skoloff G, Furie BC & Furie B (2002) Real-time in vivo imaging of platelets, tissue factor and fibrin during arterial thrombus formation in the mouse. *Nat Med* **8**, 1175-1181.

Falati S, Liu Q, Gross P, Merrill-Skoloff G, Chou J, Vandendries E, Celi A, Croce K, Furie BC & Furie B (2003) Accumulation of tissue factor into developing thrombi in vivo is dependent upon microparticle P-selectin glycoprotein ligand 1 and platelet P-selectin. *J Exp Med* **197**, 1585-1598.

Fareed J, Callas DD, Hoppensteadt D & Bermes EW, Jr. (1995) Tissue factor antigen levels in various biological fluids. *Blood Coagul Fibrinolysis* **6 Suppl 1**, S32-S36.

Fernandez PM & Rickles FR (2002) Tissue factor and angiogenesis in cancer. *Curr Opin Hematol* **9**, 401-406.

Flossel C, Luther T, Muller M, Albrecht S & Kasper M (1994) Immunohistochemical detection of tissue factor (TF) on paraffin sections of routinely fixed human tissue. *Histochemistry* **101**, 449-453.

Fortin JP, Rivard GE, Adam A & Marceau F (2005) Studies on rabbit natural and recombinant tissue factors: intracellular retention and regulation of surface expression in cultured cells. *Am J Physiol Heart Circ Physiol* **288**, H2192-H2202.

Gangloff SC, Zahringer U, Blondin C, Guenounou M, Silver J & Goyert SM (2005) Influence of CD14 on ligand interactions between lipopolysaccharide and its receptor complex. *J Immunol* **175**, 3940-3945.

Giangrande PL (2003) Six characters in search of an author: the history of the nomenclature of coagulation factors. *Br J Haematol* **121**, 703-712.

Gibbs CS, McCurdy SN, Leung LL & Paborsky LR (1994) Identification of the factor VIIa binding site on tissue factor by homologous loop swap and alanine scanning mutagenesis. *Biochemistry* **33**, 14003-14010.

Giesen PL, Rauch U, Bohrmann B, Kling D, Roque M, Fallon JT, Badimon JJ, Himber J, Riederer MA & Nemerson Y (1999) Blood-borne tissue factor: another view of thrombosis. *Proc Natl Acad Sci U S A* **96**, 2311-2315.

Giraldo P & Montoliu L (2001) Size matters: use of YACs, BACs and PACs in transgenic animals. *Transgenic Res* **10**, 83-103.

Godwin AR, Stadler HS, Nakamura K & Capecchi MR (1998) Detection of targeted GFP-Hox gene fusions during mouse embryogenesis. *Proc Natl Acad Sci U S A* **95**, 13042-13047.

Gomez K & McVey JH (2006) Tissue factor initiated blood coagulation. *Front Biosci* **11**, 1349-1359.

Gordon JW, Scangos GA, Plotkin DJ, Barbosa JA & Ruddle FH (1980) Genetic transformation of mouse embryos by microinjection of purified DNA. *Proc Natl Acad Sci U S A* **77**, 7380-7384.

Hadjantonakis AK, Dickinson ME, Fraser SE & Papaioannou VE (2003) Technicolour transgenics: imaging tools for functional genomics in the mouse. *Nat Rev Genet* **4**, 613-625.

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

Harder T & Simons K (1997) Caveolae, DIGs, and the dynamics of sphingolipid-cholesterol microdomains. *Curr Opin Cell Biol* **9**, 534-542.

Harlos K, Martin DM, O'Brien DP, Jones EY, Stuart DI, Polikarpov I, Miller A, Tuddenham EG & Boys CW (1994) Crystal structure of the extracellular region of human tissue factor. *Nature* **370**, 662-666.

Hartzell S, Ryder K, Lanahan A, Lau LF & Nathan D (1989) A growth factor-responsive gene of murine BALB/c 3T3 cells encodes a protein homologous to human tissue factor. *Mol Cell Biol* **9**, 2567-2573.

Hasty P, Rivera-Perez J & Bradley A (1991) The length of homology required for gene targeting in embryonic stem cells. *Mol Cell Biol* **11**, 5586-5591.

Hirabayashi M, Takahashi R, Ito K, Kashiwazaki N, Hirao M, Hirasawa K, Hochi S & Ueda M (2001) A comparative study on the integration of exogenous DNA into mouse, rat, rabbit, and pig genomes. *Exp Anim* **50**, 125-131.

Hooper M, Hardy K, Handyside A, Hunter S & Monk M (1987) HPRT-deficient (Lesch-Nyhan) mouse embryos derived from germline colonization by cultured cells. *Nature* **326**, 292-295.

James NJ, Ettelaie C & Bruckdorfer KR (2002) Inhibition of tissue factor activity reduces the density of cellular network formation in an in vitro model of angiogenesis. *Biochem Soc Trans* **30**, 217-221.

Janson TL, Stormorken H & Prydz H (1984) Species specificity of tissue thromboplastin. *Haemostasis* **14**, 440-444.

Kadish JL, Wenc KM & Dvorak HF (1983) Tissue factor activity of normal and neoplastic cells: quantitation and species specificity. *J Natl Cancer Inst* **70**, 551-557.

Kaikita K, Takeya M, Ogawa H, Suefuji H, Yasue H & Takahashi K (1999) Co-localization of tissue factor and tissue factor pathway inhibitor in coronary atherosclerosis. *J Pathol* **188**, 180-188.

Kalish JM, Seidman MM, Weeks DL & Glazer PM (2005) Triplex-induced recombination and repair in the pyrimidine motif. *Nucleic Acids Res* **33**, 3492-3502.

Kay GF, Penny GD, Patel D, Ashworth A, Brockdorff N & Rastan S (1993) Expression of Xist during mouse development suggests a role in the initiation of X chromosome inactivation. *Cell* **72**, 171-182.

Kelley RF, Costas KE, O'Connell MP & Lazarus RA (1995) Analysis of the factor VIIa binding site on human tissue factor: effects of tissue factor mutations on the kinetics and thermodynamics of binding. *Biochemistry* **34**, 10383-10392.

Khazen W, M'bika JP, Tomkiewicz C, Benelli C, Chany C, Achour A & Forest C (2005) Expression of macrophage-selective markers in human and rodent adipocytes. *FEBS Lett* **579**, 5631-5634.

Knauert MP & Glazer PM (2001) Triplex forming oligonucleotides: sequence-specific tools for gene targeting. *Hum Mol Genet* **10**, 2243-2251.

Koller BH, Hagemann LJ, Doetschman T, Hagaman JR, Huang S, Williams PJ, First NL, Maeda N & Smithies O (1989) Germ-line transmission of a planned alteration

made in a hypoxanthine phosphoribosyltransferase gene by homologous recombination in embryonic stem cells. *Proc Natl Acad Sci U S A* **86**, 8927-8931.

Koller BH & Smithies O (1992) Altering genes in animals by gene targeting. *Annu Rev Immunol* **10**, 705-730.

Lai L, Alaverdi N, Maltais L & Morse HC, III (1998) Mouse cell surface antigens: nomenclature and immunophenotyping. *J Immunol* **160**, 3861-3868.

Ledermann B (2000) Embryonic stem cells and gene targeting. *Exp Physiol* **85**, 603-613.

Lee EC, Yu D, Martinez d, V, Tessarollo L, Swing DA, Court DL, Jenkins NA & Copeland NG (2001) A highly efficient Escherichia coli-based chromosome engineering system adapted for recombinogenic targeting and subcloning of BAC DNA. *Genomics* **73**, 56-65.

Lee G & Saito I (1998) Role of nucleotide sequences of loxP spacer region in Cre-mediated recombination. *Gene* **216**, 55-65.

Liu P, Jenkins NA & Copeland NG (2003) A highly efficient recombineering-based method for generating conditional knockout mutations. *Genome Res* **13**, 476-484.

Luther T, Flossel C, Mackman N*, et al.* (1996) Tissue factor expression during human and mouse development. *Am J Pathol* **149**, 101-113.

Lwaleed BA & Bass PS (2006) Tissue factor pathway inhibitor: structure, biology and involvement in disease. *J Pathol* **208**, 327-339.

Lwaleed BA, Bass PS & Cooper AJ (2001) The biology and tumour-related properties of monocyte tissue factor. *J Pathol* **193**, 3-12.

Mackman N (1997) Regulation of the tissue factor gene. *Thromb Haemost* **78**, 747-754.

Mackman N, Imes S, Maske WH, Taylor B, Lusis AJ & Drake TA (1992) Structure of the murine tissue factor gene. Chromosome location and conservation of regulatory elements in the promoter. *Arterioscler Thromb* **12**, 474-483.

Magin-Lachmann C, Kotzamanis G, D'Aiuto L, Wagner E & Huxley C (2003) Retrofitting BACs with G418 resistance, luciferase, and oriP and EBNA-1 - new vectors for in vitro and in vivo delivery. *BMC Biotechnol* **3**, 2.

Mansour SL, Thomas KR & Capecchi MR (1988) Disruption of the proto-oncogene int-2 in mouse embryo-derived stem cells: a general strategy for targeting mutations to non-selectable genes. *Nature* **336**, 348-352.

Matthews T & Boehme R (1988) Antiviral activity and mechanism of action of ganciclovir. *Rev Infect Dis* **10 Suppl 3**, S490-S494.

Maugeri N, Brambilla M, Camera M, Carbone A, Tremoli E, Donati MB, de GG & Cerletti C (2006) Human polymorphonuclear leukocytes produce and express functional tissue factor upon stimulation. *J Thromb Haemost* **4**, 1323-1330.

McVey JH (1999) Tissue factor pathway. *Baillieres Best Pract Res Clin Haematol* **12**, 361-372.

McVey JH (2003) Your bleeding heart: lessons from low tissue factor expression in mice. *Trends Pharmacol Sci* **24**, 269-272.

McVey JH, Boswell E, Mumford AD, Kemball-Cook G & Tuddenham EG (2001) Factor VII deficiency and the FVII mutation database. *Hum Mutat* **17**, 3-17.

Mechiche H, Cornillet-Lefebvre P & Nguyen P (2005) A subpopulation of human B lymphocytes can express a functional Tissue Factor in response to phorbol myristate acetate. *Thromb Haemost* **94**, 146-154.

Mehta R & Champney WS (2002) 30S ribosomal subunit assembly is a target for inhibition by aminoglycosides in Escherichia coli. *Antimicrob Agents Chemother* **46**, 1546-1549.

Melis E, Moons L, De Mol M, Herbert JM, Mackman N, Collen D, Carmeliet P & Dewerchin M (2001) Targeted deletion of the cytosolic domain of tissue factor in mice does not affect development. *Biochem Biophys Res Commun* **286**, 580-586.

Metzger D & Feil R (1999) Engineering the mouse genome by site-specific recombination. *Curr Opin Biotechnol* **10**, 470-476.

Mody RS & Carson SD (1997) Tissue factor cytoplasmic domain peptide is multiply phosphorylated in vitro. *Biochemistry* **36**, 7869-7875.

Molor-Erdene P, Okajima K, Isobe H, Uchiba M, Harada N, Shimozawa N & Okabe H (2005) Inhibition of lipopolysaccharide-induced tissue factor expression in monocytes by urinary trypsin inhibitor in vitro and in vivo. *Thromb Haemost* **94**, 136-145.

Moons AH, Levi M & Peters RJ (2002) Tissue factor and coronary artery disease. *Cardiovasc Res* **53**, 313-325.

Morel O, Toti F, Hugel B & Freyssinet JM (2004) Cellular microparticles: a disseminated storage pool of bioactive vascular effectors. *Curr Opin Hematol* **11**, 156-164.

Morrissey JH (2001) Tissue factor: an enzyme cofactor and a true receptor. *Thromb Haemost* **86**, 66-74.

Muller YA, Ultsch MH, Kelley RF & de Vos AM (1994) Structure of the extracellular domain of human tissue factor: location of the factor VIIa binding site. *Biochemistry* **33**, 10864-10870.

Mumford AD (2003) The Generation of a Green Fluorescent Protein reporter model for Tissue Factor Expression. PhD , University of London.

Mumford AD, Chen D, Dorling A, Kemball-Cook G & McVey JH (2005) Generation of a polyclonal rabbit anti-mouse tissue factor antibody by nucleic acid immunisation. *Thromb Haemost* **93**, 160-164.

Muyrers JP, Zhang Y & Stewart AF (2001) Techniques: Recombinogenic engineering--new options for cloning and manipulating DNA. *Trends Biochem Sci* **26**, 325-331.

Nagy A (2000) Cre recombinase: the universal reagent for genome tailoring. *Genesis* **26**, 99-109.

Nemerson Y (1969) Characteristics and lipid requirements of coagulant proteins extracted from lung and brain: the specifity of protein component of tissue factor. *J Clin Invest* **48**, 322-331.

Neuenschwander PF & Morrissey JH (1994) Roles of the membrane-interactive regions of factor VIIa and tissue factor. The factor VIIa Gla domain is dispensable for binding to tissue factor but important for activation of factor X. *J Biol Chem* **269**, 8007-8013.

Osoegawa K, Tateno M, Woon PY, Frengen E, Mammoser AG, Catanese JJ, Hayashizaki Y & de Jong PJ (2000) Bacterial artificial chromosome libraries for mouse sequencing and functional analysis. *Genome Res* **10**, 116-128.

Osterud B, Rao LV & Olsen JO (2000) Induction of tissue factor expression in whole blood: lack of evidence for the presence of tissue factor expression in granulocytes. *Thromb Haemost* **83**, 861-867.

Owen CA (2001) *A History of Blood Coagulation*: Mayo Foundation.

Paborsky LR, Caras IW, Fisher KL & Gorman CM (1991) Lipid association, but not the transmembrane domain, is required for tissue factor activity. Substitution of the transmembrane domain with a phosphatidylinositol anchor. *J Biol Chem* **266**, 21911-21916.

Paborsky LR & Harris RJ (1990) Post-translational modifications of recombinant human tissue factor. *Thromb Res* **60**, 367-376.

Panteleev MA, Zarnitsina VI & Ataullakhanov FI (2002) Tissue factor pathway inhibitor: a possible mechanism of action. *Eur J Biochem* **269**, 2016-2031.

Parry GC, Erlich JH, Carmeliet P, Luther T & Mackman N (1998) Low levels of tissue factor are compatible with development and hemostasis in mice. *J Clin Invest* **101**, 560-569.

Parry GC & Mackman N (2000) Mouse embryogenesis requires the tissue factor extracellular domain but not the cytoplasmic domain. *J Clin Invest* **105**, 1547-1554.

Pawlinski R, Fernandes A, Kehrle B*, et al.* (2002) Tissue factor deficiency causes cardiac fibrosis and left ventricular dysfunction. *Proc Natl Acad Sci U S A* **99**, 15333-15338.

Pawlinski R & Mackman N (2004) Tissue factor, coagulation proteases, and protease-activated receptors in endotoxemia and sepsis. *Crit Care Med* **32**, S293-S297.

Pearson JD (1999) Endothelial cell function and thrombosis. *Baillieres Best Pract Res Clin Haematol* **12**, 329-341.

Polgar J, Matuskova J & Wagner DD (2005) The P-selectin, tissue factor, coagulation triad. *J Thromb Haemost* **3**, 1590-1596.

Poteete AR, Fenton AC & Wang HR (2002) Recombination-Promoting Activity of the Bacteriophage {lambda} Rap Protein in Escherichia coli K-12. *The Journal of Bacteriology* **184**, 4626-4629.

Poulsen LK, Jacobsen N, Sorensen BB, Bergenhem NC, Kelly JD, Foster DC, Thastrup O, Ezban M & Petersen LC (1998) Signal transduction via the mitogen-activated protein kinase pathway induced by binding of coagulation factor VIIa to tissue factor. *J Biol Chem* **273**, 6228-6232.

Poustka A, Rackwitz HR, Frischauf AM, Hohn B & Lehrach H (1984) Selective isolation of cosmid clones by homologous recombination in Escherichia coli. *Proc Natl Acad Sci U S A* **81**, 4129-4133.

Prelle K, Zink N & Wolf E (2002) Pluripotent stem cells--model of embryonic development, tool for gene targeting, and basis of cell therapy. *Anat Histol Embryol* **31**, 169-186.

Rehemtulla A, Ruf W & Edgington TS (1991) The integrity of the cysteine 186-cysteine 209 bond of the second disulfide loop of tissue factor is required for binding of factor VII. *J Biol Chem* **266**, 10294-10299.

Riewald M, Kravchenko VV, Petrovan RJ, O'Brien PJ, Brass LF, Ulevitch RJ & Ruf W (2001) Gene induction by coagulation factor Xa is mediated by activation of protease-activated receptor 1. *Blood* **97**, 3109-3116.

Romisch J, Feussner A, Vermohlen S & Stohr HA (1999) A protease isolated from human plasma activating factor VII independent of tissue factor. *Blood Coagul Fibrinolysis* **10**, 471-479.

Rosen ED, Chan JC, Idusogie E*, et al.* (1997) Mice lacking factor VII develop normally but suffer fatal perinatal bleeding. *Nature* **390**, 290-294.

Rottingen JA, Enden T, Camerer E, Iversen JG & Prydz H (1995) Binding of human factor VIIa to tissue factor induces cytosolic Ca2+ signals in J82 cells, transfected COS-1 cells, Madin-Darby canine kidney cells and in human endothelial cells induced to synthesize tissue factor. *J Biol Chem* **270**, 4650-4660.

Ruf W (2004) Protease-activated receptor signaling in the regulation of inflammation. *Crit Care Med* **32**, S287-S292.

Ruf W, Schullek JR, Stone MJ & Edgington TS (1994) Mutational mapping of functional residues in tissue factor: identification of factor VII recognition determinants in both structural modules of the predicted cytokine receptor homology domain. *Biochemistry* **33**, 1565-1572.

Rushworth SA, Chen XL, Mackman N, Ogborne RM & O'Connell MA (2005) Lipopolysaccharide-induced heme oxygenase-1 expression in human monocytic cells is mediated via Nrf2 and protein kinase C. *J Immunol* **175**, 4408-4415.

Russo A (2000) In vivo cytogenetics: mammalian germ cells. *Mutat Res* **455**, 167-189.

Santucci RA, Erlich J, Labriola J, Wilson M, Kao KJ, Kickler TS, Spillert C & Mackman N (2000) Measurement of tissue factor activity in whole blood. *Thromb Haemost* **83**, 445-454.

Schedl A, Larin Z, Montoliu L, Thies E, Kelsey G, Lehrach H & Schutz G (1993) A method for the generation of YAC transgenic mice by pronuclear microinjection. *Nucleic Acids Res* **21**, 4783-4787.

Schmidt B, Ho L & Hogg PJ (2006) Allosteric disulfide bonds. *Biochemistry* **45**, 7429-7433.

Sevinsky JR, Rao LV & Ruf W (1996) Ligand-induced protease receptor translocation into caveolae: a mechanism for regulating cell surface proteolysis of the tissue factor-dependent coagulation pathway. *J Cell Biol* **133**, 293-304.

Shakes LA, Garland DM, Srivastava DK, Harewood KR & Chatterjee PK (2005) Minimal cross-recombination between wild-type and loxP511 sites in vivo facilitates truncating both ends of large DNA inserts in pBACe3.6 and related vectors. *Nucleic Acids Res* **33**, e118.

She K (2003) So you want to Work with Giants: the BAC Vector. *Bioteach Journal* **1**, 69-74.

Shimomura O, Johnson F & Saiga Y (1962) Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, Aequorea. *J Cell Comp Physiol* **59**, 223-239.

Shizuya H, Birren B, Kim UJ, Mancino V, Slepak T, Tachiiri Y & Simon M (1992) Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in Escherichia coli using an F-factor-based vector. *Proc Natl Acad Sci U S A* **89**, 8794-8797.

Siegel RW, Jain R & Bradbury A (2001) Using an in vivo phagemid system to identify non-compatible loxP sequences. *FEBS Lett* **505**, 467-473.

Simons K & Ikonen E (1997) Functional rafts in cell membranes. *Nature* **387**, 569-572.

Steffel J, Luscher TF & Tanner FC (2006) Tissue factor in cardiovascular diseases: molecular mechanisms and clinical implications. *Circulation* **113**, 722-731.

Suh TT, Holmback K, Jensen NJ, Daugherty CC, Small K, Simon DI, Potter S & Degen JL (1995) Resolution of spontaneous bleeding events but failure of pregnancy in fibrinogen-deficient mice. *Genes Dev* **9**, 2020-2033.

Szotowski B, Antoniak S & Rauch U (2006) Alternatively spliced tissue factor: a previously unknown piece in the puzzle of hemostasis. *Trends Cardiovasc Med* **16**, 177-182.

Szotowski B, Goldin-Lang P, Antoniak S*, et al.* (2005) Alterations in myocardial tissue factor expression and cellular localization in dilated cardiomyopathy. *J Am Coll Cardiol* **45**, 1081-1089.

Takamiya O, Abe S, Yoshioka A, Nakajima K, McVey JH & Tuddenham EG (1995) Factor VIIShinjo: a dysfunctional factor VII variant homozygous for the substitution Gln for Arg at position 79. *Haemostasis* **25**, 89-97.

Thomson JG & Ow DW (2006) Site-specific recombination systems for the genetic manipulation of eukaryotic genomes. *Genesis* **44**, 465-476.

Toomey JR, Kratzer KE, Lasky NM, Stanton JJ & Broze GJ, Jr. (1996) Targeted disruption of the murine tissue factor gene results in embryonic lethality. *Blood* **88**, 1583-1587.

Toschi V, Gallo R, Lettino M*, et al.* (1997) Tissue factor modulates the thrombogenicity of human atherosclerotic plaques. *Circulation* **95**, 594-599.

Tsien RY (1998) The Green Fluorescent Protein. *Annu Rev Biochem* **67**, 509-544.

van den Eijnden MM, Steenhauer SI, Reitsma PH & Bertina RM (1997) Tissue factor expression during monocyte-macrophage differentiation. *Thromb Haemost* **77**, 1129-1136.

van't Veer C & Mann KG (1997) Regulation of tissue factor initiated thrombin generation by the stoichiometric inhibitors tissue factor pathway inhibitor, antithrombin-III, and heparin cofactor-II. *J Biol Chem* **272**, 4367-4377.

Vasquez KM, Marburger K, Intody Z & Wilson JH (2001) Manipulating the mammalian genome by homologous recombination. *PNAS* **98**, 8403-8410.

Watt VM, Ingles CJ, Urdea MS & Rutter WJ (1985) Homology requirements for recombination in Escherichia coli. *Proc Natl Acad Sci U S A* **82**, 4768-4772.

Wilcox JN, Smith KM, Schwartz SM & Gordon D (1989) Localization of tissue factor in the normal vessel wall and in the atherosclerotic plaque. *Proc Natl Acad Sci U S A* **86**, 2839-2843.

Zeng C, Kouprina N, Zhu B, Cairo A, Hoek M, Cross G, Osoegawa K, Larionov V & de JP (2001) Large-insert BAC/YAC libraries for selective re-isolation of genomic regions by homologous recombination in yeast. *Genomics* **77**, 27-34.

Zhang Y, Buchholz F, Muyrers JP & Stewart AF (1998) A new logic for DNA engineering using recombination in Escherichia coli. *Nat Genet* **20**, 123-128.

Appendix 1 – Reagents made in-house

Solution	Composition
10X Hogness freezing medium	36 mmol/L $K_2HPO_4.3H_2O$, 13 mmol/L KH_2PO_4, 20 mmol/L sodium citrate, 10 mmol/L $MgSO_4.7H_2O$, 44% glycerol (v/v)
20X SSC	3 mol/L NaCl, 0.3 mol/L sodium citrate
Alkaline SDS	200 mmol/L NaOH, 1% sodium dodecyl sulphate
CsCl gradient	295 g CsCl in 300 mL H_2O
ES Fixative (prepared fresh on day of use)	1 volume glacial acetic acid: 2.5 volumes methanol
Immunoprecipitation buffer	20 mmol/L tris base, 100 mmol/L NaCl, 0.05% Tween 80 (v/v), pH 8
Mammalian DNA lysis buffer	10 mmol/L tris base, 10 mmol/L EDTA, 10 mmol/L NaCl, 0.5% N-lauroylsarcosine sodium salt, 1 mg/mL proteinase K, pH 7.5
Nonidet-P40 buffer	20 mmol/L tris base, 137 mmol/L NaCl, 1% Nonidet-P40 (v/v), 10% glycerol (v/v), 2 mmol/L EDTA, pH 7.5
PBS (Phosphate buffered saline)	137 mmol/L NaCl, 2.7 mmol/L KCl, 8 mmol/L Na_2HPO_4, 2 mmol/L KH_2PO_4
Plasmid lysis buffer	50 mmol/L glucose, 25 mmol/L tris base, 10 mmol/L EDTA, pH 8
RIPA buffer	50 mmol/L tris base, 150 mmol/L NaCl, 1% Nonidet-P40 (v/v), 0.5% sodium deoxycholate, 0.1% sodium dodecyl sulphate, pH 8
Southern denaturation	1.5 mol/L NaCl, 0.5 mol/L NaOH

Southern neutralisation	1 mol/L tris base, 1.5 mol/L NaCl, pH 8
TBE	90 mmol/L tris base, 90 mmol/L boric acid, 2 mmol/L EDTA, pH 8
TBS-A	50 mmol/L tris base, 150 mmol/L NaCl, 1 g/L human albumin, pH 7.4
Western blocking buffer	50 mmol/L tris base, 150 mmol/L NaCl, 0.5% Triton X-100, pH 7.5
Western reducing buffer	25% NuPAGE® LDS sample buffer (Invitrogen), 20 mmol/L dithiothreitol

Reagents were obtained from Sigma-Aldrich, Fisher Scientific or Merck Biosciences. Unless otherwise indicated percentages refer to w/v.

Appendix 2 – Commercial suppliers

Abcam, Cambridge, UK

Abgene UK, Epsom, Surrey, UK

Accelrys, Cambridge, UK

Ambion (Europe) Ltd., Huntingdon, Cambridgeshire, UK

American Diagnostica, Stamford, Connecticut, USA

Autogen Bioclear, Calne, Wiltshire, UK

Axis-Shield, Kimbolton, Cambridgeshire, UK

BD Biosciences, Oxford, UK

BDH, Poole, Dorset, UK

Bio-Rad, Hemel Hempstead, Hertfordshire, UK

Cambrex BioScience, Wokingham, Berkshire, UK

Chemicon International, Temecula, California, USA

Clontech, Saint-Germain-en-Laye, France

DakoCytomation, Ely, Cambridgeshire, UK

Fisher Scientific UK Ltd, Loughborough, Leicestershire, UK

FMC Bioproducts, Rockland, Maine, USA

GE Healthcare, Amersham, UK

Hitachi, Wokingham, Berkshire, UK

IKA-Werke GMBH & Co., Staufen, Germany

Invitrogen, Paisley, UK

Leica Microsystems GmbH, Wetzlar, Germany

Li-Cor Biosciences UK Ltd., Cambridge, UK

Merck Biosciences Ltd. Beeston, Nottinghamshire, UK

Millipore, Watford, UK

Molecular Biology Insights Inc., Plymouth, USA

Molecular Devices, Wokingham, Berkshire, UK

NEB (New England Biolabs), Hitchin, Hertfordshire, UK

Nikon UK Ltd. Kingston-upon-Thames, Surrey, UK

Open Biosystems, Huntsville, Alabama, USA

Premier Biosoft International, Palo Alto, California, USA

Promega, Southampton, UK

Qiagen, Crawley, West Sussex, UK

Roche Applied Science, Indianapolis, USA

Sakura Finetek Europe B.V., Zoeterwoude, The Netherlands

Scotlab, Coatbridge, North Lanarkshire, Scotland

Sigma-Aldrich, Poole, Dorset, UK

Silicon Graphics Ltd., Reading, Berkshire, UK

Stratagene Europe, Amsterdam, The Netherlands

Vector Laboratories, Peterborough, UK

VWR International Ltd, Lutterworth, Leicestershire, UK

Diagnostic Reagents Ltd. Thame, Oxfordshire, UK

Harlan-Sera Lab, Loughborough, Leicestershire, UK

Appendix 3 – Oligonucleotide primers

Primers used in production of the targeting vectors

Primers	Sequence 5' to 3'	Annealing Temp °C	Amplicon size (bp)
PbES2F	TCCCTGCTCACATGACTTAT	60	777
PbES2R	GTGTGATTAAGTGAGGGCTC		
GFP3F	GGCACAAGCTGGAGTACAA	59	1747/547
3HAR	GGAGTAACAGGCCATGAAG		
MTFex3F	CCACAGACCCTAACCTCC	59	337
MTFex3R	GGTCATGAAGCACAAGCA		
NP2	TTGGCTACCCGTGATATTG	59	2381
MTFP2	GCTCTTTGTCCTGGATAAGTCA		
MTFI5F	ACTTTGTATGCTGGCCTAAA	59	887
EGFP2R	CTTGTGGCCGTTTACGTC		
N5	ATGTCGAGCAAACCCCG	59	443
N3	GGCACTTCGCCCAATAG		
T7	TAATACGACTCACTATAGGG	59	402
MTFex3R	GGTCATGAAGCACAAGCA		
TK1R	GGCACTCTGTCGATACCC	59	296
T3	AATTAACCCTCACTAAAGGG		
MTFex3U	CTCTTCAGGTTTTCCCAGTT	59	400
MTFex3R	GGTCATGAAGCACAAGCA		
3HAF	CTGGTCGGTTTTGTTTAC	57	424
TKR2	ATGCGAAGTGGACCTC		

Primers designed for sequencing the 5′ HA

Primers	Sequence 5' to 3'	Annealing Temp °C	Amplicon size (bp)
MTFex3F	CCACAGACCCTAACCTCC	59	337
MTFex3R	GGTCATGAAGCACAAGCA		
MTFSeq1FA	AATGCCTTCACTCTTGGTC	59	768
MTFSeq1RB	GGGGATTACCAAAACCAC		
MTFSeq2F	AGCCTCCATGTTGACTTTAC	59	673
MTFSeq2R	ACACAAACATACACCAGGCA		
MTFSeq3F	ATAGATTGGGTGTGGGATGT	59	688
MTFSeq3R	CATCGCTCATACAGAACACC		
PbMTF1F	TAATGTTGCCAAGGTGAAAA	59	743
PbMTF2R	TCTCTGAAAGGACCCTGC		
MTFSeq4F	CCCCCCCTTTTTAATAGA	59	503
MTFSeq4R	CAGTTCAGCCCCTTTTTAC		
MTFI4F	CTACCACTGAGTTGCATTCC	59	536
MTFI5R	CCCAAAGAGAATGAAGAATC		
MTFSeq5F	CAGAAATGAGCTTAGGGTGT	59	936
MTF4R	CCTAGGCCAGAGTGGGAAT		
MTFEx6F	AGCCTCCATGTTGACTTTAC	59	606
MTFGFPR1	CTCCTTGAAGTCGATGCC		

cDNA primers

Primers	Sequence 5' to 3'	Annealing Temp °C	Amplicon size (bp)
Primers for *mF3* cDNA			
MTFCF	TGAAGCCCCGAGACCTC	57	624
MTFCR	TTTCCCGTGCTTGAGCC		
Primers for *GAPDH* cDNA			
MGAPDHF	TCAACGACCCCTTCATTGAC	56	531
MGAPDHR	ATGCAGGGATGATGTTCTGG		
Primers for *mF3-EGFP* cDNA			
MTFC-F1	TGAAGCCCCGAGACCT	57	1067
MTFC-R2	AGATGAACTTCAGGGTCAGC		
MTFC-F3	GGACGGCGACGTAAAC	56	531
MTFC-R4	GCTTCAGCCTTTCCTCTACT		
MTFC-F6S	AAAACTGAACGTGGTTGTA	Primers used for sequencing only	
MTFC-R5S	GCAGGGTGAGGAATGTA		
MTFC-F8S	GACGGCAACTACAAGAC		
MTFC-R7S	GGGTGTTCTGCTGGTA		

Primers and probes for real-time RT-PCR

Primer/Probe	Sequence 5' to 3'	Annealing Temp °C	Amplicon size (bp)
WT assay			
TFcDNA-F1	TTCCTGGGAGAAACACTCATCA	59	184
WTcDNA-R1	CCGTGCAGGCAGTGTGA		
WT-MGB (Probe)	AGCCTTTCCTCTATGCC	69	
EGFP assay			
TFcDNA-F1	TTCCTGGGAGAAACACTCATCA	59	184
EGFPcDNA-R2	ACCACCCCGGTGAACA		
EGFP-MGB (Probe)	TCCTCGCCCTTGCT	69	
TFc control assay			
TFc-MGB (Probe)	GCGGGTGCAGGCATTCCAGAGAAAG	69	

The TFc control assay was obtained ready-made from Applied Biosystems. The MGB probe sequence was provided but details of the forward and reverse primers used in the TFc control assay were not provided by the manufacturer.

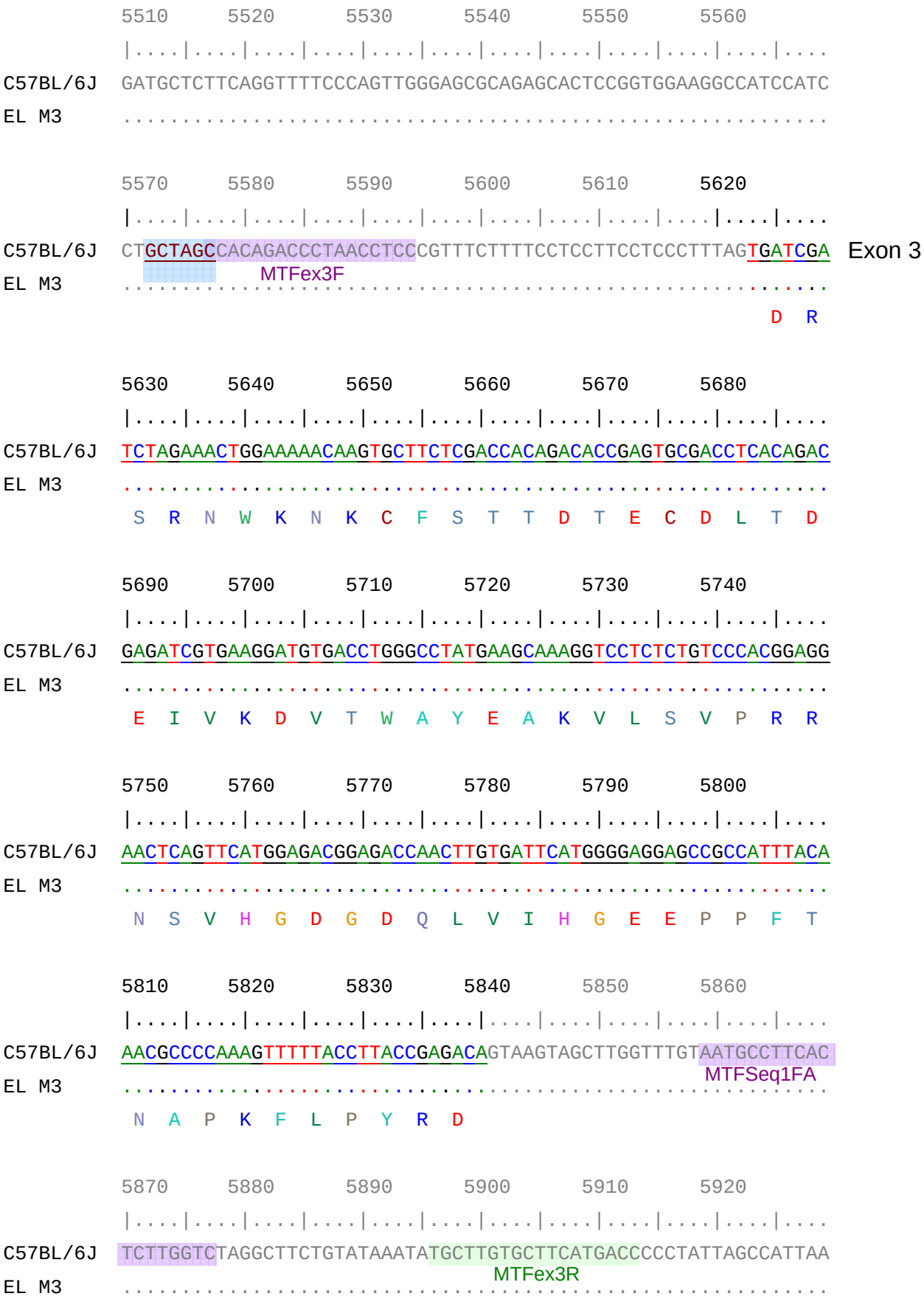
Appendix 4 – Sequence alignment in 5′ HA

Comparison of the 5′ HA sequence from EL M3 cells (derived from the 129/Sv strain) with the C57BL/6J reference sequence for the *mF3* gene, accession number AC129311. Numbering is according to the reference sequence with +1 representing the A of the ATG start codon. Exons are in coloured, underlined text with single letter translations below the sequence. The exon number is given in the right margin. Introns are in grey text with dinucleotide repeat regions of 15 bp or more underlined.

The positions of *Nhe*I sites used in the assembly of pES-TC-L are shown by dark red, underlined text on pale blue background. The binding sites of primers used in the sequencing of the 5′ HA are shown in lavender (forward primers) and light green (reverse primers). The primer name is given in the same colour below the binding site.

Sequence identity is shown by dots in the EL M3 sequence. Differences between the EL M3 and reference sequences are shown by black text on orange background and listed below:

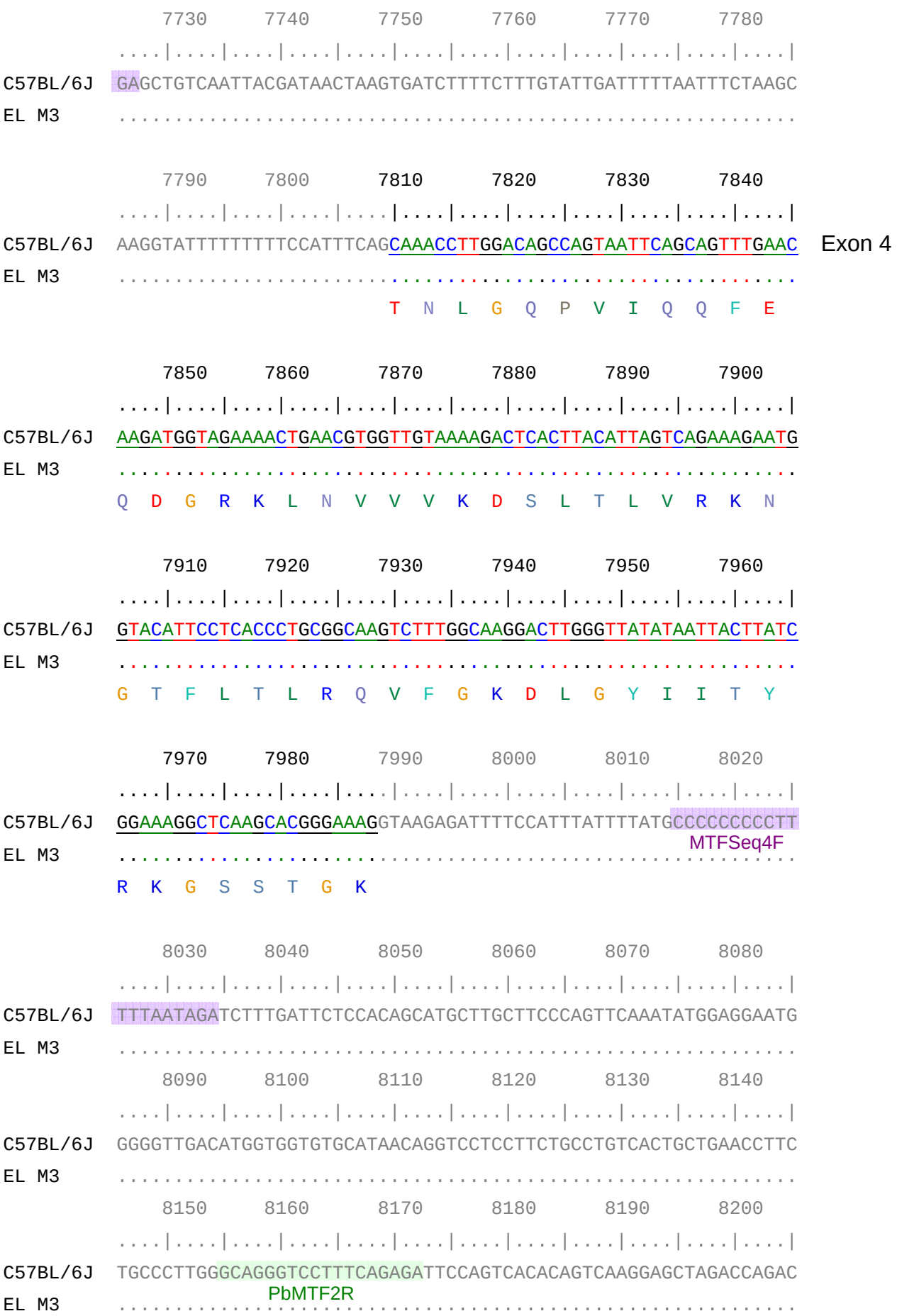
1. g.5975_5982 del TCTCTCTC ins ACAC
2. g.9268_9269 ins TG
3. g.9291 T>C
4. g.10208 G>A

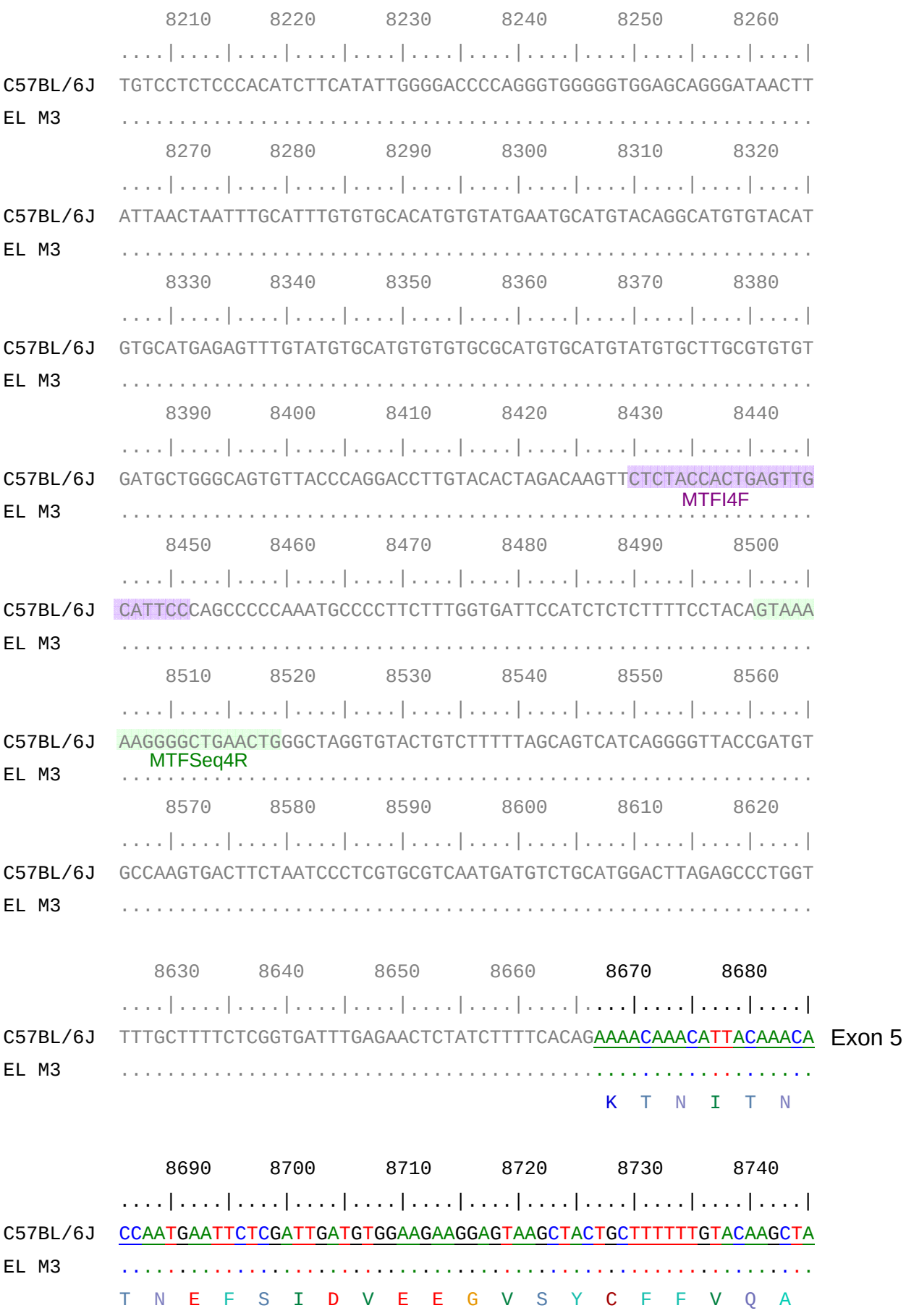

5510 5520 5530 5540 5550 5560
C57BL/6J GATGCTCTTCAGGTTTTCCCAGTTGGGAGCGCAGAGCACTCCGGTGGAAGGCCATCCATC
EL M3 .....
5570 5580 5590 5600 5610 5620
C57BL/6J CTGCTAGCCACAGACCCTAACCTCCCGTTTCTTTTCCTCCTTCCTCCCTTTAGTGATCGA
MTFex3F
Exon 3
EL M3 .....
D R
5630 5640 5650 5660 5670 5680
C57BL/6J TCTAGAAACTGGAAAAACAAGTGCTTCTCGACCACAGACACCGAGTGCGACCTCACAGAC
EL M3 .....
S R N W K N K C F S T T D T E C D L T D
5690 5700 5710 5720 5730 5740
C57BL/6J GAGATCGTGAAGGATGTGACCTGGGCCTATGAAGCAAAGGTCCTCTCTGTCCCACGGAGG
EL M3 .....
E I V K D V T W A Y E A K V L S V P R R
5750 5760 5770 5780 5790 5800
C57BL/6J AACTCAGTTCATGGAGACGGAGACCAACTTGTGATTCATGGGGAGGAGCCGCCATTTACA
EL M3 .....
N S V H G D G D Q L V I H G E E P P F T
5810 5820 5830 5840 5850 5860
C57BL/6J AACGCCCCAAAGTTTTTACCTTACCGAGACAGTAAGTAGCTTGGTTTGTAATGCCTTCAC
MTFSeq1FA
EL M3 .....
N A P K F L P Y R D
5870 5880 5890 5900 5910 5920
C57BL/6J TCTTGGTCTAGGCTTCTGTATAAATATGCTTGTGCTTCATGACCCCCTATTAGCCATTAA
MTFex3R
EL M3 .....

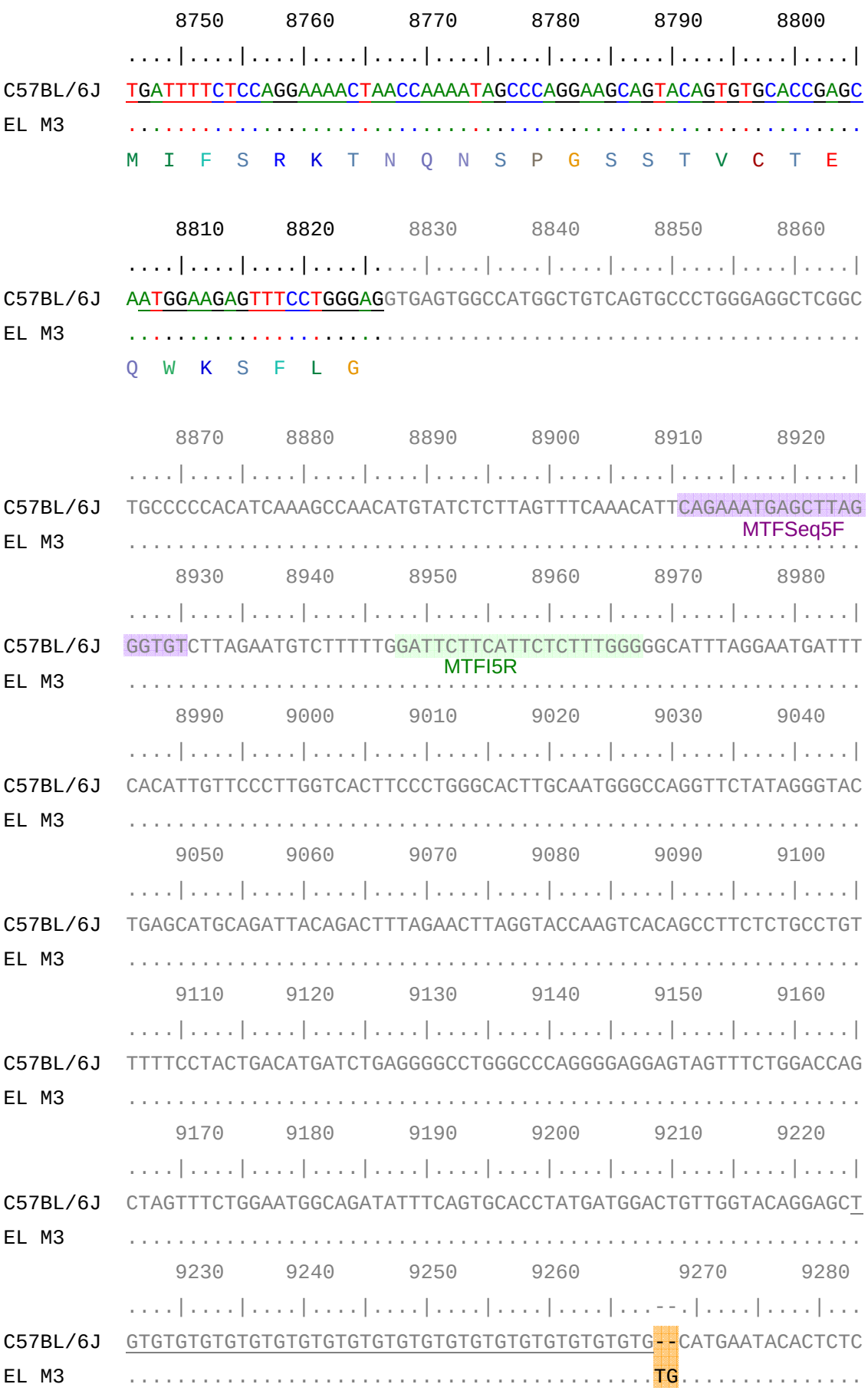
```
              5930      5940      5950      5960      5970      5980
              |....|....|....|....|....|....|....|....|....|....|..----..|
C57BL/6J      TTAGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC----ACA
EL M3         .............................................--------ACAC...
                 5990      6000      6010      6020      6030      6040
              ....|....|....|....|....|....|....|....|....|....|....|....|
C57BL/6J      CACACACACACACACACACACACACACACACACACACACACCAAGGAGATGCATCTTATAAG
EL M3         ............................................................
                  6050      6060      6070      6080      6090      6100
              ....|....|....|....|....|....|....|....|....|....|....|....|
C57BL/6J      TGTCTCTCTCTCTGCAGATGGGCAGGAGGTTCACAGAATGTGAACTCGTGTCTTGATCAC
EL M3         ............................................................
                  6110      6120      6130      6140      6150      6160
              ....|....|....|....|....|....|....|....|....|....|....|....|
C57BL/6J      TATGTCATCTACTCACTTTCTCCTCCAGTCTGCCGCCACTTCCTATCAGGAACCAGCTGG
EL M3         ............................................................
                  6170      6180      6190      6200      6210      6220
              ....|....|....|....|....|....|....|....|....|....|....|....|
C57BL/6J      AGGGTGCGCTAAAGCTGGGGTTGTGGCTTTTCCTCAGGCTACTGTGTCATATGCTGAATT
EL M3         ............................................................
                  6230      6240      6250      6260      6270      6280
              ....|....|....|....|....|....|....|....|....|....|....|....|
C57BL/6J      TTATATTCTTTCATCAGACCAGGGGTGCTAGTTTAGACATTTTCCTCCAATGAGCACAGA
EL M3         ............................................................
                  6290      6300      6310      6320      6330      6340
              ....|....|....|....|....|....|....|....|....|....|....|....|
C57BL/6J      TCCACAGGGAAGCCACTTGTGCCTGAAATTGGGCTTGTCAGCCTCCCTCGTGCTGGGTTT
EL M3         ............................................................
                  6350      6360      6370      6380      6390      6400
              ....|....|....|....|....|....|....|....|....|....|....|....|
C57BL/6J      GAATGTGTAGCATGCTTTCCCACCTTTGTCCAAGTGTGTGAGCCTGCCATCTAACAAGGA
EL M3         ............................................................
                  6410      6420      6430      6440      6450      6460
              ....|....|....|....|....|....|....|....|....|....|....|....|
C57BL/6J      AATCTTCTTTCCTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTGTGTGTGTGTGTG
EL M3         ............................................................
                  6470      6480      6490      6500      6510      6520
              ....|....|....|....|....|....|....|....|....|....|....|....|
C57BL/6J      TGTGTATGTATGTATATCTTCCTTCCTTTCTTTTTTCTTTCCTTCCATGTTTCAGAAGTA
                                               MTFSeq2F
EL M3         ............................................................
```

```
                6530      6540      6550      6560      6570      6580
                ....|....|....|....|....|....|....|....|....|....|....|
C57BL/6J        GACATTTCACAAGCTTATGCAGCACATTAGGAAGACTAAATCCCACTTCTTTTGCAAGGC
EL M3           ............................................................
                6590      6600      6610      6620      6630      6640
                ....|....|....|....|....|....|....|....|....|....|....|
C57BL/6J        CTGCTGGTTAAAACACAAGTGTTGTGGTTTTGGTAATCCCCATCTGTCCCCTTCTGATTG
                                          MTFSeq1RB
EL M3           ............................................................
                6650      6660      6670      6680      6690      6700
                ....|....|....|....|....|....|....|....|....|....|....|
C57BL/6J        CTCAGCGACGTGGATGGTGAGAAGGATGAAGGGCAATTGAGAAGATTTTACTTCCGATGT
EL M3           ............................................................
                6710      6720      6730      6740      6750      6760
                ....|....|....|....|....|....|....|....|....|....|....|
C57BL/6J        TTGCTCTAATGGTCTAACTACCCAAATGAATGTAGCTGCCTTCCTGGGGTCCCCACTGAC
EL M3           ............................................................
                6770      6780      6790      6800      6810      6820
                ....|....|....|....|....|....|....|....|....|....|....|
C57BL/6J        AAGCCTCTCTTCGGACTCCTTTGTATTTTTACCTTTGTGCTTCTGCACAACACGCGCATG
EL M3           ............................................................
                6830      6840      6850      6860      6870      6880
                ....|....|....|....|....|....|....|....|....|....|....|
C57BL/6J        CCAATCATGCTCCCCAAGCTGTGATTTCCCTTCATAGTTCTTGAGCTTTCAGCAGTCCAG
EL M3           ............................................................
                6890      6900      6910      6920      6930      6940
                ....|....|....|....|....|....|....|....|....|....|....|
C57BL/6J        ACCAAATCCTTTGAGATACACAACAGAACACACACAATCACACACACACACACACACTGT
EL M3           ............................................................
                6950      6960      6970      6980      6990      7000
                ....|....|....|....|....|....|....|....|....|....|....|
C57BL/6J        TCTGTGATTTGCACAGAGTGAGGACTTGTCTGACCATCACATGTGCTGCAGTTACCTCGT
EL M3           ............................................................
                7010      7020      7030      7040      7050      7060
                ....|....|....|....|....|....|....|....|....|....|....|
C57BL/6J        GTAACTGAATCAGGAACCAAGAACTTAAAATAGATTGGGTGTGGGATGTGTCAAGAGGTT
                                                  MTFSeq3F
EL M3           ............................................................
                7070      7080      7090      7100      7110      7120
                ....|....|....|....|....|....|....|....|....|....|....|
C57BL/6J        GTTGCAGGAAAACATTAAAAATTAACAAAATTAGAAGGTGACAAATGTGCACATAGTGGA
EL M3           ............................................................
```

```
                7130      7140      7150      7160      7170      7180
            ....|....|....|....|....|....|....|....|....|....|....|....|
C57BL/6J    CTTTCAAAAAAAAAAGTCTATGCCTGGTGTATGTTTGTGTATCATCTGTAACTCAGTTCC
                                    MTFSeq2R
EL M3       ............................................................
                7190      7200      7210      7220      7230      7240
            ....|....|....|....|....|....|....|....|....|....|....|....|
C57BL/6J    AATGATCAGATGCCATCTGACCTGTGGACCCTAGCCATGTATGTAGTACACACATCCATA
EL M3       ............................................................
                7250      7260      7270      7280      7290      7300
            ....|....|....|....|....|....|....|....|....|....|....|....|
C57BL/6J    GAACAGGCAAAATACATGCATAAAAATAAAAAAATATTTTTAAAAGTCTTAAAAAAAAAA
EL M3       ............................................................
                7310      7320      7330      7340      7350      7360
            ....|....|....|....|....|....|....|....|....|....|....|....|
C57BL/6J    AAAAACCCTTCCTGGGCTTCCCAGCCTCTAATGCATTGTAGCTAAAATACGGAAGCTTGG
EL M3       ............................................................
                7370      7380      7390      7400      7410      7420
            ....|....|....|....|....|....|....|....|....|....|....|....|
C57BL/6J    AAAACTTCTTGTCCAGCCTGGTAGCCATCACTCACACATGGCTACCAAACTTTTGAAATA
EL M3       ............................................................
                7430      7440      7450      7460      7470      7480
            ....|....|....|....|....|....|....|....|....|....|....|....|
C57BL/6J    TGGCTAATGTTGCCAAGGTGAAAAATATATGCCAGTTTTTGAAAAATATAAAATATCCAT
                  PbMTF1F
EL M3       ............................................................
                7490      7500      7510      7520      7530      7540
            ....|....|....|....|....|....|....|....|....|....|....|....|
C57BL/6J    AATATCCACTATATTGAGTGCATGTTAAAATGCCAATCTGTTAATGGACCTACTAAAGTA
EL M3       ............................................................
                7550      7560      7570      7580      7590      7600
            ....|....|....|....|....|....|....|....|....|....|....|....|
C57BL/6J    TATAATTAAAGTTAATTCCACAGGCTCCTTCTCACATTTTTCTAACATCTGCAAGGAAGG
EL M3       ............................................................
                7610      7620      7630      7640      7650      7660
            ....|....|....|....|....|....|....|....|....|....|....|....|
C57BL/6J    GTCTCCTTGTGCTTCCAGAGGACGGCATGGCTTGCTATGTGTTGAGACTCAGCTACAAGA
EL M3       ............................................................
                7670      7680      7690      7700      7710      7720
            ....|....|....|....|....|....|....|....|....|....|....|....|
C57BL/6J    CCGTCACCACTCCTGCTACTGGAGCTGATTCTGTACAGGTGTTCTGTATGAGCGATGGAG
                                              MTFSeq3R      MTFI3F
EL M3       ............................................................
```

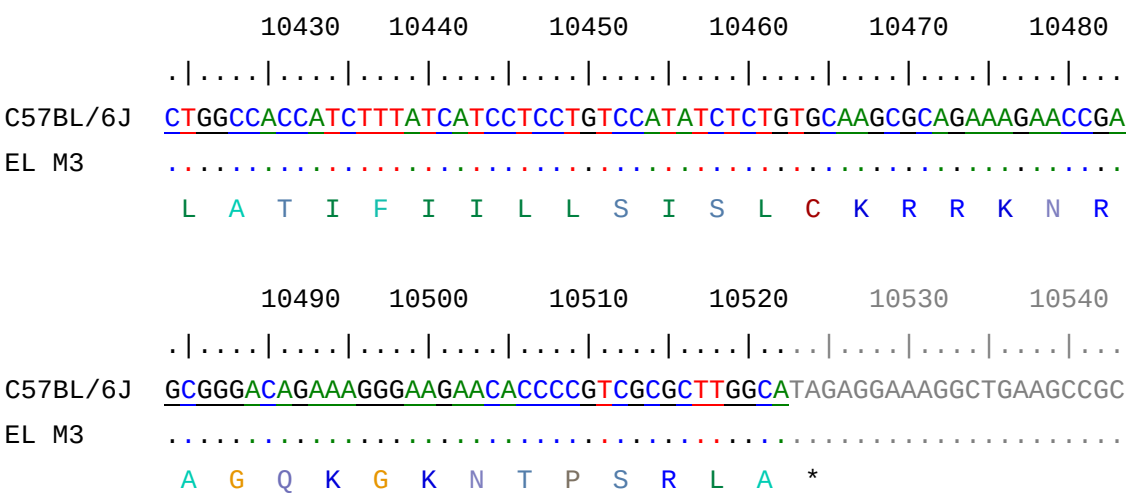

7730 7740 7750 7760 7770 7780
C57BL/6J GAGCTGTCAATTACGATAACTAAGTGATCTTTTCTTTGTATTGATTTTTAATTTCTAAGC
EL M3
7790 7800 7810 7820 7830 7840
C57BL/6J AAGGTATTTTTTTTTCCATTTCAGCAAACCTTGGACAGCCAGTAATTCAGCAGTTTGAAC
Exon 4
EL M3
T N L G Q P V I Q Q F E
7850 7860 7870 7880 7890 7900
C57BL/6J AAGATGGTAGAAAACTGAACGTGGTTGTAAAAGACTCACTTACATTAGTCAGAAAGAATG
EL M3
Q D G R K L N V V V K D S L T L V R K N
7910 7920 7930 7940 7950 7960
C57BL/6J GTACATTCCTCACCCTGCGGCAAGTCTTTGGCAAGGACTTGGGTTATATAATTACTTATC
EL M3
G T F L T L R Q V F G K D L G Y I I T Y
7970 7980 7990 8000 8010 8020
C57BL/6J GGAAAGGCTCAAGCACGGGAAAGGTAAGAGATTTTCCATTTATTTTATGCCCCCCCCCTT
MTFSeq4F
EL M3
R K G S S T G K
8030 8040 8050 8060 8070 8080
C57BL/6J TTTAATAGATCTTTGATTCTCCACAGCATGCTTGCTTCCCAGTTCAAATATGGAGGAATG
EL M3
8090 8100 8110 8120 8130 8140
C57BL/6J GGGGTTGACATGGTGGTGTGCATAACAGGTCCTCCTTCTGCCTGTCACTGCTGAACCTTC
EL M3
8150 8160 8170 8180 8190 8200
C57BL/6J TGCCCTTGGGCAGGGTCCTTTCAGAGATTCCAGTCACACAGTCAAGGAGCTAGACCAGAC
PbMTF2R
EL M3


8210 8220 8230 8240 8250 8260
C57BL/6J TGTCCTCTCCCACATCTTCATATTGGGGACCCCAGGGTGGGGGTGGAGCAGGGATAACTT
EL M3 .....
8270 8280 8290 8300 8310 8320
C57BL/6J ATTAACTAATTTGCATTTGTGTGCACATGTGTATGAATGCATGTACAGGCATGTGTACAT
EL M3 .....
8330 8340 8350 8360 8370 8380
C57BL/6J GTGCATGAGAGTTTGTATGTGCATGTGTGTGCGCATGTGCATGTATGTGCTTGCGTGTGT
EL M3 .....
8390 8400 8410 8420 8430 8440
C57BL/6J GATGCTGGGCAGTGTTACCCAGGACCTTGTACACTAGACAAGTTCTCTACCACTGAGTTG
MTFI4F
EL M3 .....
8450 8460 8470 8480 8490 8500
C57BL/6J CATTCCCAGCCCCCAAATGCCCCTTCTTTGGTGATTCCATCTCTCTTTTCCTACAGTAAA
EL M3 .....
8510 8520 8530 8540 8550 8560
C57BL/6J AAGGGGCTGAACTGGGCTAGGTGTACTGTCTTTTTAGCAGTCATCAGGGGTTACCGATGT
MTFSeq4R
EL M3 .....
8570 8580 8590 8600 8610 8620
C57BL/6J GCCAAGTGACTTCTAATCCCTCGTGCGTCAATGATGTCTGCATGGACTTAGAGCCCTGGT
EL M3 .....
8630 8640 8650 8660 8670 8680
C57BL/6J TTTGCTTTTCTCGGTGATTTGAGAACTCTATCTTTTCACAGAAAACAAACATTACAAACA
Exon 5
EL M3 .....
K T N I T N
8690 8700 8710 8720 8730 8740
C57BL/6J CCAATGAATTCTCGATTGATGTGGAAGAAGGAGTAAGCTACTGCTTTTTTGTACAAGCTA
EL M3 .....
T N E F S I D V E E G V S Y C F F V Q A


8750 8760 8770 8780 8790 8800
C57BL/6J TGATTTTCTCCAGGAAAACTAACCAAAATAGCCCAGGAAGCAGTACAGTGTGCACCGAGC
EL M3 ..... Exon 5 (cont.)
M I F S R K T N Q N S P G S S T V C T E
8810 8820 8830 8840 8850 8860
C57BL/6J AATGGAAGAGTTTCCTGGGAGGTGAGTGGCCATGGCTGTCAGTGCCCTGGGAGGCTCGGC
EL M3 .....
Q W K S F L G
8870 8880 8890 8900 8910 8920
C57BL/6J TGCCCCCACATCAAAGCCAACATGTATCTCTTAGTTTCAAACATTCAGAAATGAGCTTAG
MTFSeq5F
EL M3 .....
8930 8940 8950 8960 8970 8980
C57BL/6J GGTGTCTTAGAATGTCTTTTTGGATTCTTCATTCTCTTTGGGGGCATTTAGGAATGATTT
MTFI5R
EL M3 .....
8990 9000 9010 9020 9030 9040
C57BL/6J CACATTGTTCCCTTGGTCACTTCCCTGGGCACTTGCAATGGGCCAGGTTCTATAGGGTAC
EL M3 .....
9050 9060 9070 9080 9090 9100
C57BL/6J TGAGCATGCAGATTACAGACTTTAGAACTTAGGTACCAAGTCACAGCCTTCTCTGCCTGT
EL M3 .....
9110 9120 9130 9140 9150 9160
C57BL/6J TTTTCCTACTGACATGATCTGAGGGGCCTGGGCCCAGGGGAGGAGTAGTTTCTGGACCAG
EL M3 .....
9170 9180 9190 9200 9210 9220
C57BL/6J CTAGTTTCTGGAATGGCAGATATTTCAGTGCACCTATGATGGACTGTTGGTACAGGAGCT
EL M3 .....
9230 9240 9250 9260 9270 9280
C57BL/6J GTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG--CATGAATACACTCTC
EL M3 .....TG.....

```
                9290      9300      9310      9320      9330      9340
           .|...|....|....|....|....|....|....|....|....|....|....|...
C57BL/6J   TGGGTATTCAGTGTTGTGACTAAAAGCTCTGGGCCCCACAGAGAAGGGGCTCTGCTGTCA
EL M3      ......C.....................................................
                9350      9360      9370      9380      9390      9400
           .|....|....|....|....|....|....|....|....|....|....|....|...
C57BL/6J   GAGGAGACGACTAAGGGGGCAGTAGCTATGGCCTGGCTCTGTTAGAGCTTCAAGTCATTA
EL M3      .....................................................................
                9410      9420      9430      9440      9450      9460
           .|....|....|....|....|....|....|....|....|....|....|....|...
C57BL/6J   TGCCAACTAGGCCAGACTCGGGAAGACAAATGCCACATCTTCCTCTTATATGTGGAAAGC
EL M3      .....................................................................
                9470      9480      9490      9500      9510      9520
           .|....|....|....|....|....|....|....|....|....|....|....|...
C57BL/6J   TAGATTAAATGTATGTACATGTATGTATTTATATATGTATGTAGTGGGGTCATGAAACTA
EL M3      .....................................................................
                9530      9540      9550      9560      9570      9580
           .|....|....|....|....|....|....|....|....|....|....|....|...
C57BL/6J   GAAAGGAGACCATGGGAGAGAAGGAAGAAATGTAAAGGGGAAGGGTTAATAAAATAGAAG
EL M3      .....................................................................
                9590      9600      9610      9620      9630      9640
           .|....|....|....|....|....|....|....|....|....|....|....|...
C57BL/6J   TACTGAAAAGAGGCGGGGGATGGAGGAACTATCTAGGGAAAGAAACCAGAAAGAGACCAA
EL M3      .....................................................................
                9650      9660      9670      9680      9690      9700
           .|....|....|....|....|....|....|....|....|....|....|....|...
C57BL/6J   GGGGCATGGGGGCCAGGAATGAATGAGAACAAAGTCTAATTGCATGTTCTTCATAAAAAA
EL M3      .....................................................................
                9710      9720      9730      9740      9750      9760
           .|....|....|....|....|....|....|....|....|....|....|....|...
C57BL/6J   TACCATAATTAAACACATTACTTTGTATGCTGGCCTAAAAATAACTTTTTTAAATGCAAT
                                   MTFI5F
EL M3      .....................................................................
                9770      9780      9790      9800      9810      9820
           .|....|....|....|....|....|....|....|....|....|....|....|...
C57BL/6J   AATATGTTTAAGTATTCTGCCTTCTTGCCTCTGTTAACAGAAAGTTCTCTTTCCATTGGA
EL M3      .....................................................................
                9830      9840      9850      9860      9870      9880
           .|....|....|....|....|....|....|....|....|....|....|....|...
C57BL/6J   GCCCATTCCCACTCTGGGCTAGCCAGCTCCTTAGCCACAAGCCTCACTATTCCCTTCCCT
EL M3      .....................................................................
```

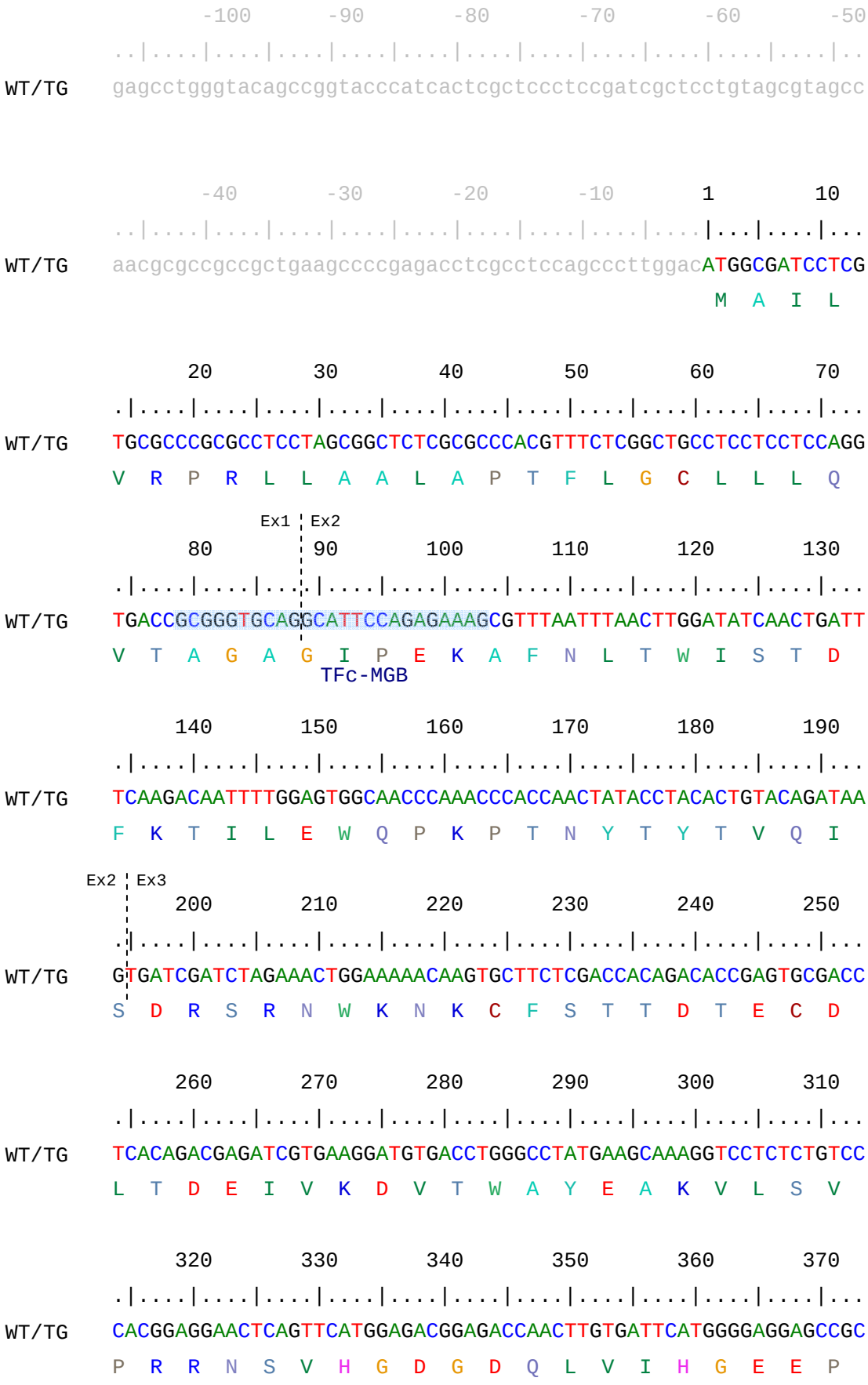

9890 9900 9910 9920 9930 9940
C57BL/6J AAGCCCTACACTATCTAGTTCTGATGCCCCAGGCTCTGCCTGTTTCCCTGCTTCATCTCT
EL M3 .....
9950 9960 9970 9980 9990 10000
C57BL/6J TAACCCAATAGATCCTCTGGGGTACTGACTACCCCAGCCAACTACTGTCTCTTCGTCTCC
EL M3 .....
10010 10020 10030 10040 10050 10060
C57BL/6J ACCATCAGCACTGTGCCCCCTCCTTCTCTGTCTTTCCCATTTATCTTCTGTTATAATCTC
EL M3 .....
10070 10080 10090 10100 10110 10120
C57BL/6J TGCCACCAAACTCAGGGTCGTTTTTTGGACTACATTTCCAACATGTGTTCTTCCATCCAT
MTFI5R1
EL M3 .....
10130 10140 10150 10160 10170 10180
C57BL/6J CTCCCGTGGCTCCTTTCTAGTCCAGGAACCTTGGGTAGCTCATTTAATGCTGTAGCCACT
EL M3 .....
10190 10200 10210 10220 10230 10240
C57BL/6J GTCAGTTTCTCTCAAGCGTATGAGGTCAGATAAATTCTTAATCAAAGAATGAGAGAAAGA
EL M3 .....A.....
10250 10260 10270 10280 10290 10300
C57BL/6J GTGGAAGAAATAAATGTCGGGTTTCAAAAAGGCGGGACGGTATGATTTGGGGTCACGCCC
EL M3 .....
10310 10320 10330 10340 10350 10360
C57BL/6J ACCGGCAAGTGGCGTTCGAGCCTCCATGTTGACTTTACATCTTTGATTTATTTTTTTGTT
MTFEx6F
EL M3 .....
10370 10380 10390 10400 10410 10420
C57BL/6J TGTTTGTTTGTCTGTTTTTGCTTTTCAGAAACACTCATCATTGTGGGAGCAGTGGTGCTC
Exon 6
EL M3 .....
E T L I I V G A V V L

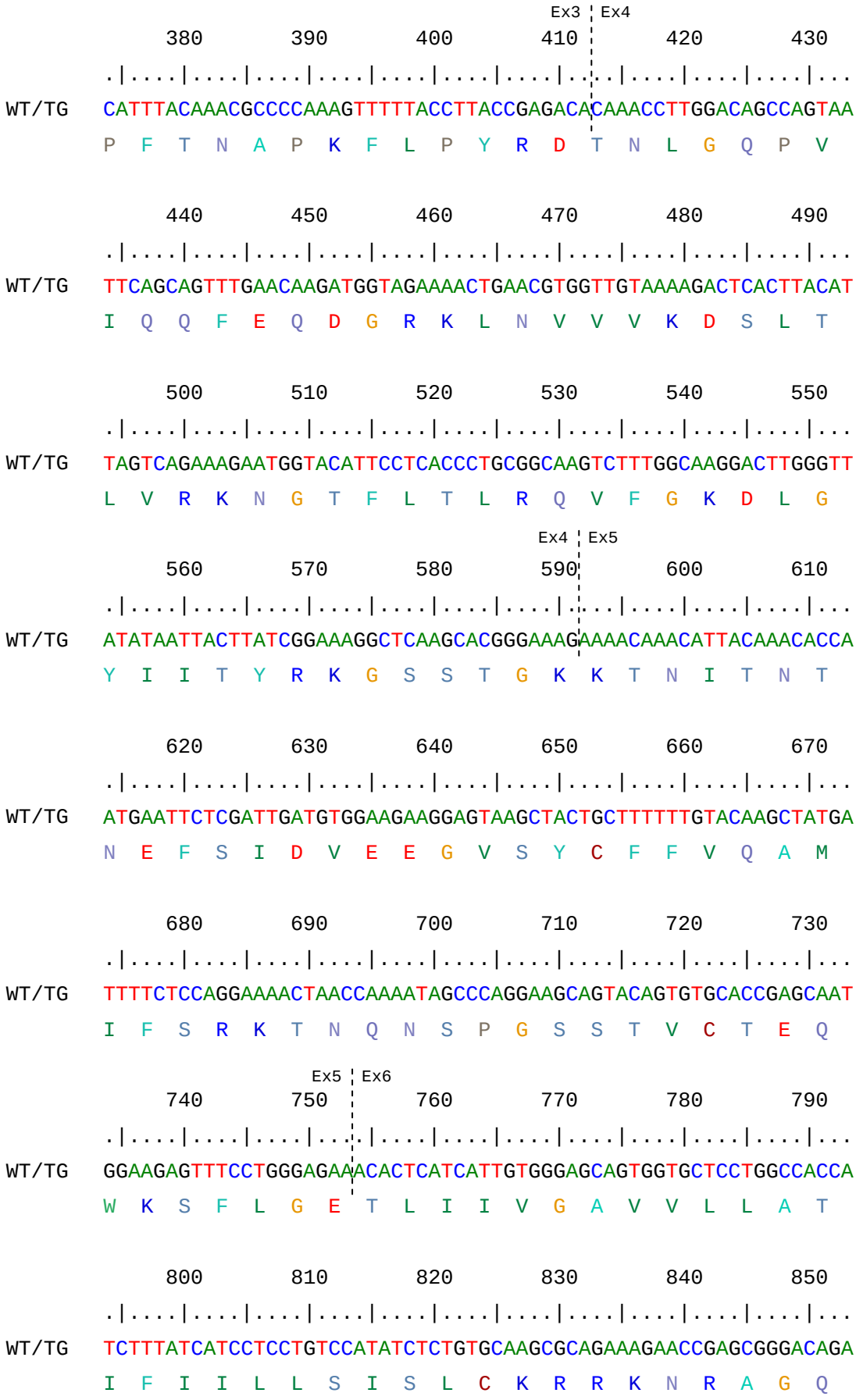

10430 10440 10450 10460 10470 10480
C57BL/6J CTGGCCACCATCTTTATCATCCTCCTGTCCATATCTCTGTGCAAGCGCAGAAAGAACCGA
EL M3 ....
L A T I F I I L L S I S L C K R R K N R
10490 10500 10510 10520 10530 10540
C57BL/6J GCGGGACAGAAAGGGAAGAACACCCCGTCGCGCTTGGCATAGAGGAAAGGCTGAAGCCGC
EL M3 ....
A G Q K G K N T P S R L A *

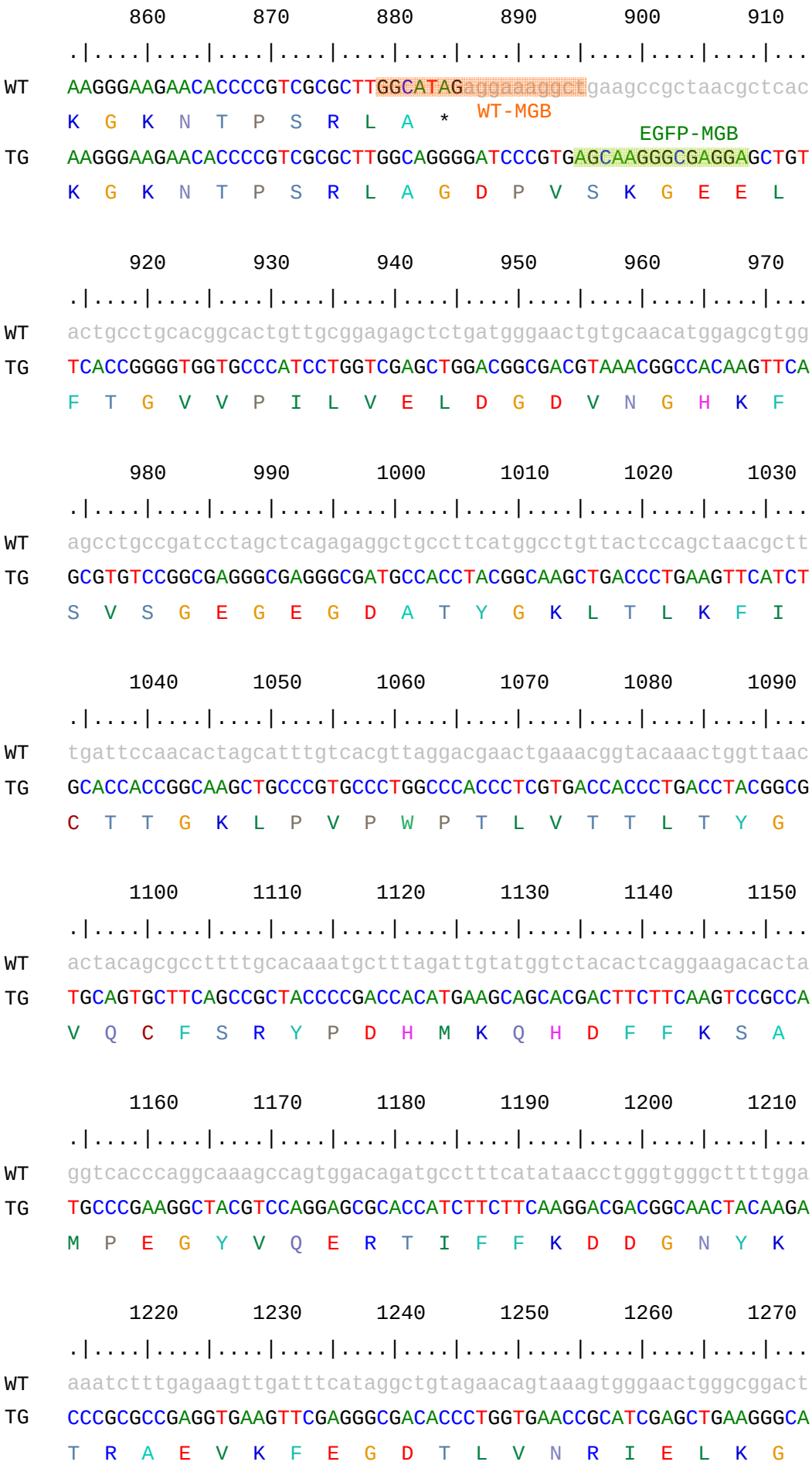
Exon 6 (cont.)

Appendix 5 – WT and TG cDNA alignments

Comparison of the cDNA sequences derived from the wild-type (WT) and transgenic (TG) alleles. WT/TG in the left margin indicates that the WT and TG alleles share the same sequence. From position 883 (the first base of the natural stop codon in the WT allele), the sequence varies, with the WT sequence printed above the TG sequence. Coding sequences are in coloured, capitalized text with the single letter amino acid translations given below. Untranslated regions are shown in grey, lowercase text. The position of exon boundaries in the gene sequence is indicated by vertical dashed lines. The position of MGB probes is indicated by coloured rectangles.


-100 -90 -80 -70 -60 -50
WT/TG gagcctgggtacagccggtacccatcactcgctccctccgatcgctcctgtagcgtagcc
-40 -30 -20 -10 1 10
WT/TG aacgcgccgccgctgaagccccgagacctcgcctccagcccttggacATGGCGATCCTCG
M A I L
20 30 40 50 60 70
WT/TG TGCGCCCGCGCCTCCTAGCGGCTCTCGCGCCCACGTTTCTCGGCTGCCTCCTCCTCCAGG
V R P R L L A A L A P T F L G C L L L Q
Ex1 Ex2
80 90 100 110 120 130
WT/TG TGACCGCGGGTGCAGGCATTCCAGAGAAAGCGTTTAATTTAACTTGGATATCAACTGATT
V T A G A G I P E K A F N L T W I S T D
TFc-MGB
140 150 160 170 180 190
WT/TG TCAAGACAATTTTGGAGTGGCAACCCAAACCCACCAACTATACCTACACTGTACAGATAA
F K T I L E W Q P K P T N Y T Y T V Q I
Ex2 Ex3
200 210 220 230 240 250
WT/TG GTGATCGATCTAGAAACTGGAAAAACAAGTGCTTCTCGACCACAGACACCGAGTGCGACC
S D R S R N W K N K C F S T T D T E C D
260 270 280 290 300 310
WT/TG TCACAGACGAGATCGTGAAGGATGTGACCTGGGCCTATGAAGCAAAGGTCCTCTCTGTCC
L T D E I V K D V T W A Y E A K V L S V
320 330 340 350 360 370
WT/TG CACGGAGGAACTCAGTTCATGGAGACGGAGACCAACTTGTGATTCATGGGGAGGAGCCGC
P R R N S V H G D G D Q L V I H G E E P


Ex3
Ex4
380 390 400 410 420 430
WT/TG CATTTACAAACGCCCCAAAGTTTTTACCTTACCGAGACACAAACCTTGGACAGCCAGTAA
P F T N A P K F L P Y R D T N L G Q P V
440 450 460 470 480 490
WT/TG TTCAGCAGTTTGAACAAGATGGTAGAAAACTGAACGTGGTTGTAAAAGACTCACTTACAT
I Q Q F E Q D G R K L N V V V K D S L T
500 510 520 530 540 550
WT/TG TAGTCAGAAAGAATGGTACATTCCTCACCCTGCGGCAAGTCTTTGGCAAGGACTTGGGTT
L V R K N G T F L T L R Q V F G K D L G
Ex4
Ex5
560 570 580 590 600 610
WT/TG ATATAATTACTTATCGGAAAGGCTCAAGCACGGGAAAGAAAACAAACATTACAAACACCA
Y I I T Y R K G S S T G K K T N I T N T
620 630 640 650 660 670
WT/TG ATGAATTCTCGATTGATGTGGAAGAAGGAGTAAGCTACTGCTTTTTTGTACAAGCTATGA
N E F S I D V E E G V S Y C F F V Q A M
680 690 700 710 720 730
WT/TG TTTTCTCCAGGAAAACTAACCAAAATAGCCCAGGAAGCAGTACAGTGTGCACCGAGCAAT
I F S R K T N Q N S P G S S T V C T E Q
Ex5
Ex6
740 750 760 770 780 790
WT/TG GGAAGAGTTTCCTGGGAGAAACACTCATCATTGTGGGAGCAGTGGTGCTCCTGGCCACCA
W K S F L G E T L I I V G A V V L L A T
800 810 820 830 840 850
WT/TG TCTTTATCATCCTCCTGTCCATATCTCTGTGCAAGCGCAGAAAGAACCGAGCGGGACAGA
I F I I L L S I S L C K R R K N R A G Q


860 870 880 890 900 910
WT AAGGGAAGAACACCCCGTCGCGCTTGGCATAGaggaaaggctgaagccgctaacgctcac
K G K N T P S R L A *
WT-MGB
EGFP-MGB
TG AAGGGAAGAACACCCCGTCGCGCTTGGCAGGGGATCCCGTGAGCAAGGGCGAGGAGCTGT
K G K N T P S R L A G D P V S K G E E L
920 930 940 950 960 970
WT actgcctgcacggcactgttgcggagagctctgatgggaactgtgcaacatggagcgtgg
TG TCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCA
F T G V V P I L V E L D G D V N G H K F
980 990 1000 1010 1020 1030
WT agcctgccgatcctagctcagaggctgccttcatggcctgttactccagctaacgctt
TG GCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCT
S V S G E G E G D A T Y G K L T L K F I
1040 1050 1060 1070 1080 1090
WT tgattccaacactagcatttgtcacgttaggacgaactgaaacggtacaaactggttaac
TG GCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCG
C T T G K L P V P W P T L V T T L T Y G
1100 1110 1120 1130 1140 1150
WT actacagcgccttttgcacaaatgctttagattgtatggtctacactcaggaagacacta
TG TGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCA
V Q C F S R Y P D H M K Q H D F F K S A
1160 1170 1180 1190 1200 1210
WT ggtcacccaggcaaagccagtggacagatgcctttcatataacctgggtgggcttttgga
TG TGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGA
M P E G Y V Q E R T I F F K D D G N Y K
1220 1230 1240 1250 1260 1270
WT aaatctttgagaagttgatttcataggctgtagaacagtaaagtgggaactgggcggact
TG CCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCA
T R A E V K F E G D T L V N R I E L K G

```
          1280      1290      1300      1310      1320      1330
     .|....|....|....|....|....|....|....|....|....|....|....|...
WT   tttctaacagtcgtacttttataaagcggtatttgggtgtttttccctcgaataggtac
TG   TCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCC
     I  D  F  K  E  D  G  N  I  L  G  H  K  L  E  Y  N  Y  N  S

          1340      1350      1360      1370      1380      1390
     .|....|....|....|....|....|....|....|....|....|....|....|...
WT   ttttggaagttcaaagcaagtggcaaactttcatataaacatgttaaatgcaggatattt
TG   ACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCC
     H  N  V  Y  I  M  A  D  K  Q  K  N  G  I  K  V  N  F  K  I

          1400      1410      1420      1430      1440      1450
     .|....|....|....|....|....|....|....|....|....|....|....|...
WT   ctgcttggggcatctttgtgatttgtactttcctacaatttagcactttaactgacaatg
TG   GCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCA
     R  H  N  I  E  D  G  S  V  Q  L  A  D  H  Y  Q  Q  N  T  P

          1460      1470      1480      1490      1500      1510
     .|....|....|....|....|....|....|....|....|....|....|....|...
WT   atggggtttaaacatttgacagccaactctatttttatacgactactatacaaagaaact
TG   TCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGA
     I  G  D  G  P  V  L  L  P  D  N  H  Y  L  S  T  Q  S  A  L

          1520      1530      1540      1550      1560      1570
     .|....|....|....|....|....|....|....|....|....|....|....|...
WT   acatatagttttatgatttaaggtacttagaaatgtttatggttaacattgtatatattt
TG   GCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCG
     S  K  D  P  N  E  K  R  D  H  M  V  L  L  E  F  V  T  A  A

          1580      1590      1600      1610      1620      1630
     .|....|....|....|....|....|....|....|....|....|....|....|...
WT   acataaaatttaaaggttttgt
TG   GGATCACTCTCGGCATGGACGAGCTGTACAAGTAAagcggccggccgcgactctagagga
     G  I  T  L  G  M  D  E  L  Y  K  *
```