Correction of *ERCC1* deficiency in mice

JIM SELFRIDGE

Thesis presented for the Degree of Doctor of Philosophy
Institute of Cell and Molecular Biology
University of Edinburgh
1999

DECLARATION

The composition of this thesis, and the work presented in it are my own, unless otherwise stated. The experiments were designed in collaboration with my supervisor Prof. David W. Melton.

James Selfridge

To Carol-Ann
and
In Memory of My Mother

Table of Contents

Abstract

The omnipresence of DNA-damaging agents and the chemical instability of certain chemical bonds in DNA have made it essential for living organisms to develop systems of DNA repair. Repair of damaged DNA is necessary to prevent lesions from causing disruption of essential cellular functions or from converting into permanent mutations that lead to malignancy or cell death. Nucleotide excision repair (NER) is just one of the repair systems that has evolved to protect cells from the consequences of DNA damage. The NER system protects DNA from the widest variety of lesions induced by agents such as UV irradiation, cross-linking agents and free radicals. Defects in NER are associated with the human inherited disorder xeroderma pigmentosum, which predisposes to skin cancer. We have previously reported the generation of DNA repair deficient mice by the targeted inactivation of the NER gene *ERCC1*. The mice were born severely runted and died, prior to weaning, with liver failure. Investigations into the consequences of *ERCC1* deficiency in other tissues have been severely constrained by the limited lifespan of the knockout mice. One of the initial reasons for generating the *ERCC1*-deficient mice was to study UV-induced tumourogenesis. Although *ERCC1* has not been associated with any known human disorder these mice serve as a model for NER deficiency.

The aims of this work were to study the pattern of *ERCC1* expression in mouse tissues to determine if a tissue specific expression was evident. Secondly, we aimed to bring about the phenotypic rescue of the lethal liver phenotype of the *ERCC1* nulls in order to study the consequences of *ERCC1* deficiency in other tissues.

In all mammals the skin is the first line of defence against the harmful effects of UV irradiation. A novel *ERCC1* mRNA has been identified in mouse skin. Subsequent characterisation of the transcript demonstrated that the difference between the normal and skin-specific *ERCC1* mRNA was at the 5' end and is due to differential initiation of transcription. As with the normal *ERCC1* transcript the novel skin-specific transcript did not appear to be induced by UV irradiation.

A functional *ERCC1* minigene was constructed to facilitate subsequent analysis of the upstream promoter region and identification of the sequences involved in the regulation of the observed skin-specific pattern of expression. The minigene

corrected the UV sensitivity of *ERCC1*-deficient cultured cells but did not exhibit the characteristic skin-specific expression pattern at the level of mRNA analysis.

A pilot study for a potential gene therapy approach to increasing the lifespan of the *ERCC1* knockout mice was completed. Recombinant ecotropic retroviruses containing *ERCC1* coding sequences were produced using a murine leukaemia virus derived viral vector and viral packaging cell line. The resultant *ERCC1* retrovirus was shown to partially correct the UV sensitivity associated with pools of *ERCC1*-deficient mouse embryonic fibroblasts. Clones, subsequently isolated from the transduced pools, were shown to be phenotypically corrected to wild type levels. The observed phenotypic correction correlated with expression of a retroviral *ERCC1* mRNA.

A transthyretin regulated *ERCC1* transgene was used to bring about the targeted expression of *ERCC1* in the liver. Pronuclear injection of the transgene was used to produce a transgenic mouse line containing ~5 copies of the transgene integrated at a single site. Expression of this transgene on an *ERCC1*-deficient background resulted in the correction of the lethal liver phenotype. Transgene positive nulls were not as severely runted and survived for between 9 and 12 weeks compared to the three-week survival of the transgene negative *ERCC1*-deficient mice. The consequences of *ERCC1* deficiency in other tissues were studied. Abnormalities were identified in the skin and kidneys of adult transgene positive nulls.

An investigation into the consequences of DNA repair deficiency on UV-B induced immunosuppression revealed that antigen presenting epidermal Langerhans cells, in the transgene positive nulls, did not show the normal pattern of accumulation in the lymph nodes following UV irradiation.

Acknowledgements

So many people have helped me, in so many ways, throughout this time that it is difficult to know where to start with my thanks. If it had not been for the continued support and encouragement of my supervisor David Melton, things may have been very different. I joined David to work as a technician in his lab over ten years ago and throughout this time he has continually given me every opportunity to progress. It has been a privilege for me to work with him. I am very grateful for all his advice and all he has taught me, not to mention the continued employment.

I will always be grateful to the PhD students from 'the good old days', Nik Somia and Patrick Costello, who helped me find my feet, amongst other things, when I first arrived in the lab. In particular I would like to thank Simon Thompson for his continued friendship and guidance during this time.

Thanks go to all the members of the Melton lab, past and present: Thomas Magin, Jim McWhir, Sabine Leitgeb, Tracey White, Richard Moore, Darren Bentley, Fatima Nuñez, Kan-Tai Hsia, Caroline Holmes, Liz Jamieson, Niki Redhead, Michael Chipchase and special thanks to Carolanne McEwan for her help with genotyping more mice than I care to remember. Despite having worked in the lab for sometime, it was not until I began this PhD that I was forced into the strange world of 'tissue culture' and I would like to thank Ann-Marie Ketchen for holding my hand throughout what could have been a traumatic experience. Any success I have had in tissue culture, she assures me, is down to her expert tuition. Jean Ramsay and Joan Smail have provided a steady supply of clean glassware and sexual harassment throughout my time here and I am grateful to both, for both.

I am extremely grateful to Douglas Scott and the staff (Lesley, Stuart, Les, Eileen, Moira, Hilda and Avril) of the Ann Walker building for all of their help in all things mouse-related. Thanks also to the ICMB photographers, Graham Brown, Frank Johnston and Dave Haswell for their advice and help with preparing the figures presented here. I must also thank our collaborators Ali El-Ghorr and Mary Norval for their patience and understanding that *ERCC1*-deficient mice don't grow on trees. Donald MacLeod and Margaret Robertson both acted as excellent teachers of pronuclear injections, it really isn't as easy as it looks on the video. My glamorous

assistant, Jill Douglas was a great help when it came to performing the necessary sequence alignments.

It would be difficult, not to mention dangerous, for me to ignore the contributions of the other friends I have made in my time at Kings Buildings. Sandra Bruce, Jeanette Goman and Fiona Gray have all tried their best to ensure that my metaphorical keel remained even. Thanks for all their support and guidance through the rough seas. Jill Douglas, Ann-Marie Ketchen, Ali Alloueche, Mike Dyson and Laurie Cooper have all helped ensure that my alcohol consumption was always in a social context. I have known Laurie for almost 15 years and in that time I cannot recall a single conversation about science and for that I also thank him.

I would also like to thank my family for all their encouragement, particularly after my mother died in 1994 and a lot of this didn't seem worth it. They say that behind every great man there is a great woman. Whilst I make no claims about being a great man, I will always be indebted to my 'great woman', my wife Carol-Ann. Without her unwavering support over the last few years absolutely none of this would have been possible. Thanks Nan!

Abbreviations

μ	micro
ψ	psi
A	Adenosine
AT	ataxia telangiectasia
BER	base excision repair
bp	base pair(s)
BS	Bloom's syndrome
C	cytosine
C-	carboxy-terminal
cDNA	DNA complimentary to RNA
CHO cells	Chinese hamster ovary cells
cpm	counts per minute
CS	Cockayne's syndrome
DNA	deoxyribonucleic acid
DMSO	dimethylsulphoxide
EDTA	ethanediaminetetraacetic acid
ERCC	excision repair cross-complementing
ES	embryonic stem
EtBr	ethidium bromide
FA	Fanconi anaemia
FACScan	fluorescence activated cell scanning
FCS	foetal calf serum
g	gram(s)
G	guanosine
GS	glycerol shock
HPRT	hypoxanthine phosphoribosyltransferase
J	joule(s)
kb	kilobase
l	litres
LB	Luria broth

LC	Langerhans cells
LTR	long terminal repeats
m	messenger
m	metre
M	molar
MLV	murine leukaemia virus
MMR	mismatch repair
MOPS	3-[*N*-morpholino]propane sulphonic acid
N-	amino terminal
NCS	new-born calf serum
NER	nucleotide excision repair
nt	nucleotide(s)
OD	optical density
p	plasmid
PCR	polymerase chain reaction
PGK	phosphoglycerate kinase
PrP	prion protein
r	resistance (superscript)
r	ribosomal
RACE	rapid amplification of cDNA ends
RNA	ribonucleic acid
RNase	ribonuclease
RT	reverse-transcriptase
SV40	Simian virus 40
T	thymidine
TB	terrific broth
TG	transgene
TK	thymidine kinase
TTD	Trichothiodystrophy
TTR	transthyretin
UTR	untranslated region

UV	ultraviolet
V	volts
WT	wild type
XP	xeroderma pigmentosum

Chapter 1
Introduction

1.1 Foreword

Perhaps more than any other property, genomic stability is essential in all organisms. As the carrier of the genetic code it is essential, for normal cell function and ultimately for survival, that the integrity of the DNA molecule is maintained. The reliable maintenance of any form of data is generally dependent on the stability of the medium on which it is stored and in this respect cellular DNA is no different. The data that is, in this case, DNA is continually under threat of molecular changes and rearrangements, each of which can potentially result in changes to the informational accuracy of the genetic code. Alterations to the DNA molecule can arise, either spontaneously as a consequence of normal cellular activities, or as a result of environmental agents. Irrespective of the origin of the DNA lesion, in order to prevent the manifestation of possible deleterious consequences, all living organisms have evolved a number of mechanisms for repairing damaged DNA.

As was the case for most areas of molecular biology, initial awareness and subsequent understanding of the mechanisms involved in the repair or tolerance of damaged DNA came from studies of prokaryotic systems in particular the bacterium *Escherichia coli.* The old adage of, 'what is true for *E. coli* is true for the elephant' is certainly relevant in most biological processes and to a certain extent DNA repair is no exception. Although many of the DNA repair mechanisms have been conserved throughout evolution, the eukaryotic organism is clearly more complex. In simple terms, the eukaryotic genome is very much larger ($3.0x10^9$ compared to $4.2x10^6$ bp) and composed of multiple chromosomes. Additionally, the multicellular nature of higher eukaryotes demands strict regulation of cell division and differentiation. DNA damage is known to impair both of these processes. Accordingly, some of the mechanisms of DNA repair present in eukaryotes are much more complex than those characterised in *E. coli.* Today, the focus of many DNA repair studies is towards developing a greater understanding of how these DNA repair mechanisms function in mammalian systems.

1.2 DNA damage

The simplest classification of DNA damage is either as spontaneous or environmental in origin. Whilst spontaneous damage refers to lesions that arise as a consequence of normal cellular processes such as DNA replication, environmental damage refers to lesions induced by exogenous factors such as UV irradiation. However, even this simplest of classification systems is not unambiguous as the chemical consequences of some spontaneous damage can result in identical changes to those brought about by environmental agents. For example, DNA strand breaks introduced by clastogenic agents such as ionising radiation can also result from endogenous factors activated by normal cellular metabolism. The variety of spontaneous alterations that can occur and the range of environmental agents that lead to DNA damage are reviewed in Friedberg, Walker and Siede (1995).

Base mismatches arising from replicative infidelity are believed to occur once in every 10^9 nucleotides inserted during DNA synthesis. However, inappropriate base pairing can also result from the spontaneous deamination of cytosine, adenine and guanine leading to the presence of uracil, hypoxanthine and xanthine in the DNA molecule, respectively. In addition to the incorporation of uracil as a consequence of deamination of cytosine, it can also occur if the synthesis of TMP from dUMP is inhibited. This results in an increase in the abundance of dUTP relative to TTP and thus promotes the incorporation of uracil relative to thymine into DNA. It is possible for each of the bases to experience a transient change in its bonding capacity, a tautomeric shift. These tautomers include the rare enol forms of thymine and guanine as well as the rare imino forms of adenine and cytosine. As such, if any of the bases in the template strand is present in the form of its tautomer, during DNA synthesis, misincorporation of bases in the daughter strand can occur. The loss of bases, depurination and depyrimidination can result from spontaneous hydrolysis of the glycosyl bonds. Finally, spontaneous damage of DNA can result from attack by reactive oxygen species or free radicals. These reactive species, such as H_2O_2, can be the products of normal metabolic processes and are capable of inducing helical distorting base damage. Lesions arising from the attack of free radicals on DNA include 8-hydroxyguanine and thymine glycol. The guanine lesion is capable of

mispairing during replication, whilst the thymine glycol appears to act as a replicative block.

The vast range of environmental DNA damaging agents and the lesions they induce makes it impracticable to present a comprehensive list within the context of this thesis. Instead, an overview of the principle classes of agents and their lesions is presented. Ionising radiation itself induces a variety of DNA lesions including single and double strand breaks, base and sugar damage and the indirect damage caused by free radicals formed as a consequence of the radiation.

Alkylating agents are chemicals which damage DNA by introducing bulky adducts. Alkylation being the introduction of methyl or ethyl groups to reactive sites of the bases and phosphodiester backbone. Guanine is particularly susceptible to alkylation, becoming methylated at the O^6 position, forming O^6-methylguanine, which is capable of mispairing with thymine.

Cross-linking agents, including mitomycin C and certain psoralens, can damage DNA by the introduction of inter- and intrastrand cross-links. Interstrand cross-links are effective in the prevention of strand separation and as such are capable of preventing normal transcriptional and replicative processes.

A number of chemicals present as environmental pollutants, including benzo[a]pyrene present in cigarette smoke, can be metabolised by cells to form electrophilic reactants. The resultant DNA lesion varies according to the actual agent. Generally, they result in the addition of helix distorting bulky chemical adducts.

Finally, UV radiation is probably the most extensively studied environmental DNA damaging agent. The UV spectrum is divided into three wavelength bands of UVA (400-320 nm), UVB (320-290 nm) and UVC (290-100 nm). Solar radiation consists principally of UVA and UVB, with UVC being absorbed by the Ozone layer. Despite this, laboratory studies of UV induced damage often use UVC (260nm) emitting sources which introduce DNA lesions identical to those caused by the longer, and more biologically significant, UVB wavelengths. UV irradiation can result in a whole host of lesions including cross-links and strand breaks. When DNA is exposed to UV radiation with a wavelength of 260nm adjacent pyrimidines become fused as covalent links are formed between the respective carbon 6 and 5 atoms. The resultant

cyclobutane pyrimidine dimer leads to distortion of the DNA helix. Another UV irradiation product is the pyrimidine-pyrimidine [6-4] photoproduct, which is a noncyclic bond between adjacent pyrimidines. The bond forming between the carbon 6 of the 5' pyrimidine and the carbon 4 of the 3' pyrimidine. Again this photoproduct formation brings about distortion of the DNA helix.

It is evident that there are a vast number of different DNA lesions caused by a equally varied number of agents or events. However, rather than maintaining several pathways specific for each of the DNA lesions, eukaryotes have evolved a limited number of repair pathways capable of dealing with several types of lesion.

1.3 Mechanisms of DNA repair

The survival of any given cell is dependent on the appropriate expression and functioning of the considerable number of proteins encoded by the cellular DNA. Cell viability is thus heavily dependent on the efficient removal of lesions from its DNA. A comprehensive discussion of each of the cellular responses to DNA damage would be outwith the scope of this thesis, instead a concise description of the principle repair pathways will be presented. Included in the many repair pathways evident in eukaryotic cells are direct reversal of damage, base excision repair, mismatch repair, double strand-break repair and finally, the focus of this thesis, nucleotide excision repair

1.4 Direct reversal of DNA damage

In the direct repair pathways damaged DNA is returned to its normal state by means of an enzyme-catalysed activity (reviewed by Sancar, 1996). The classic example of direct reversal of damage is photoreactivation. In *E. coli* the enzyme DNA photolyase repairs UV induced cyclobutane pyrimidine dimers by using a light stimulated electron transfer reaction to monomerise the dimer. A similar correcting activity has not been confirmed in higher eukaryotes. The simple rejoining of single strand breaks by DNA ligase has been observed in *E. coli*, but the significance of this process in mammalian cells remains ambiguous. For ligation to occur the free ends of the DNA require 3'OH and 5'P termini. In most cases the strand breaks commonly introduced

by radiation result in damage to the DNA strand ends, and as such require processing by additional enzymes prior to ligation.

An example of a direct repair pathway that is evident in all species is the repair of alkylation damage by the enzyme O^6-methylguanine transferase (O^6-MGT). The DNA lesion O^6-methylguanine is introduced by alkylating agents such as methylnitro-nitrosoguanidine (MNNG) and methylnitrosourea (MNU). The repair of the lesion involves the transfer of the methyl group from the DNA to a cysteine residue on the enzyme resulting in the permanent inactivation of the enzyme and the formation of S-methyl cysteine.

1.5 Double-strand break repair by end joining

Double-strand breaks (DSBs) are introduced into DNA by environmental agents such as ionising radiation or during failed recombination reactions. In higher eukaryotes DSBs are repaired by non-homologous or illegitimate end joining (reviewed by Wood, 1996). The DNA-dependent protein kinase (DNA-PK) is pivotal in V(D)J recombination and non-homologous end joining. The actual nature of the mechanism by which DNA-PK effects DNA repair has yet to be characterised, although possible models have been proposed (summarised by Jackson and Jeggo, 1995). The DNA-PK consists of a catalytic subunit and an associated DNA binding element (Ku). The binding of Ku to the DNA ends activates the DNA-PK, which is subsequently thought to recruit accessory repair proteins to the lesion where it might be responsible for their activation by phosphorylation. The actual *in vivo* targets for this phosphorylating activity remain to be determined. It is possible that the DNA-PK molecule serves to align the two ends of the double strand break, acting as a bridging factor, to facilitate ligation.

The double-strand break repair mechanism has characteristics in common with the V(D)J rearrangement process, the lymphoid site-specific recombination mechanism that is responsible for generating a diverse repertoire of immunoglobin and T-cell receptor genes by recombining various gene segments. Both processes involve double strand breaks and the joining of non-homologous ends. Severe combined immune deficient (*scid*) mutant mice are deficient in B- and T-cells due to a defect in

V(D)J recombination. Cells from these mice are sensitive to ionising radiation suggesting that the scid protein plays a role in double-strand break repair. A number of X-ray sensitive rodent cell lines have been identified and the cloning of the complementing (XRCC) genes will enable further elucidation of the double strand break repair mechanism.

1.6 Double strand break repair by homologous recombination

In addition to the non-homologous end joining mechanism, double-strand breaks can be repaired by the more biologically complex mechanism of homologous recombination (reviewed by Kanaar *et al.,* 1998). In this mechanism the second intact copy of the chromosome acts as the repair template, to ensure faithful restoration of any sequence loss at the break point. The *RAD52* epistasis group genes, from *S. cerevisiae*, have been identified as those required for DSB repair by homologous recombination. A number of mammalian homologues to these yeast genes have also been isolated, pointing to evolutionary conservation of this pathway. The ends of the double strand break are processed, by a nuclease activity, to produce single stranded 3' overhangs. The single-stranded DNA ends from the damaged chromosome invade the homologous intact duplex DNA, initiating strand exchange. Following the formation of a joint molecule between the damaged and undamaged duplex DNAs, DNA is synthesised from the damaged ends using the intact strands as templates (branch migration). Finally the crossed DNA strands (Holliday junctions) are resolved, by a resolvase which cleaves to the two separate duplexes. It is often suggested that this more accurate repair pathway is favoured in the unicellular lower eukaryotes and mammalian germ and stem cells where the non-homologous end-joining is more likely to result in errors if nucleotides are added or lost at the break and the less accurate end-joining mechanism is tolerated in differentiated somatic cells where much of the genome is no longer functional.

1.7 Base excision repair

Base excision repair (BER) refers to the pathway in which DNA bases, damaged by spontaneous hydrolysis or attack by free radicals, are generally removed and DNA is

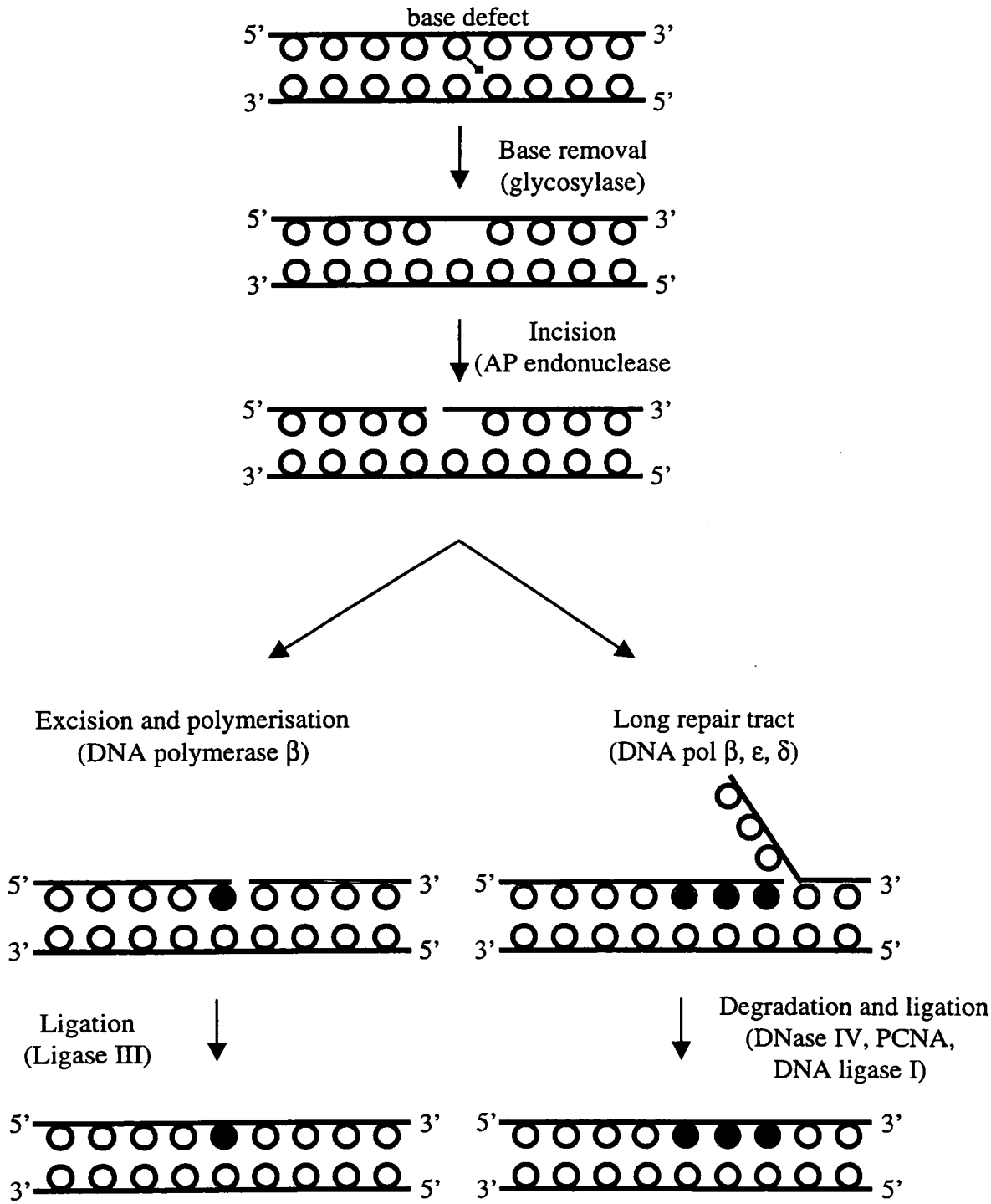
repaired (reviewed in Friedberg, Walker and Siede, 1995; Wood, 1996 and Lindahl, Karran and Wood, 1997). In BER the damaged or inappropriate base is recognised by one of several DNA glycosylases, which hydrolyses the gylcosylic bond linking the damaged base to the deoxyribose sugar and thus releases the base, generating an abasic site (Figure 1.1). Each of the glycosylases, including uracil-DNA glycosylase and 3-methyladenine-DNA glycosylase, is specific for a small subset of lesions. The repair of the resultant apurinic or apyrimidinic (AP) site is initiated by the action of an AP endonuclease, which produce incisions in the phosphodiester backbone by hydrolysing the bond to the 5' of the AP site. The completion of BER involves the removal of the 5' phosphate residue, generated by the action of the AP endonuclease, repair synthesis and ligation. The majority of AP sites are repaired by DNA polymerase β (pol β) which adds a single nucleotide to the 3' end and removes the 5' terminal phosphate residue. Subsequent ligation is performed by DNA ligase III. Alternatively, longer repair tracts can result when the polymerase (either polβ, polε or polδ) adds a number of nucleotides, in a nick translation reaction displacing the parental strand in the process. This generates an overhanging 5' terminal single stranded tract, which is removed by the action of structure specific nuclease DNase IV. This pathway shows a requirement for PCNA. Ligation of the termini is carried out by DNA ligase I. Recently, it was shown that the base excision repair of oxidised pyrimidine lesions, such as thymine glycol, is activated by the nucleotide excision repair enzyme XPG. The function of XPG in the base excision repair of oxidative damage is distinct from its NER endonuclease activity indicating overlapping functions for the components of the various repair pathways (Klungland *et al.*, 1999).

1.8 Mismatch repair

The DNA mismatch repair (MMR) pathway serves to correct errors resulting from inappropriate incorporation of nucleotide bases during replication. The discovery that mutations in mismatch repair genes segregate with hereditary non polyposis colon cancer has lead to a greater interest in this postreplicative repair process, in recent years (reviewed by Lindahl, Karran and Wood, 1997 and Jiricny, 1998). In mammals, mismatch repair is carried out using protein homologues of the *E. coli* MMR proteins

Figure 1.1 Schematic representation of BER pathways

A model for the mechanism of base excision repair, adapted from Wood 1996. The sugar phosphate backbone is indicated by lines, whilst bases are depicted by circles. Open circles represent bases present in the DNA molecule before repair and closed circles are the bases introduced by the repair mechanism. See text for details.


base defect
5'
3'
3'
5'
Base removal
(glycosylase)
5'
3'
3'
5'
Incision
(AP endonuclease
5'
3'
3'
5'
Excision and polymerisation
(DNA polymerase β)
Long repair tract
(DNA pol β, ε, δ)
5'
3'
3'
5'
5'
3'
3'
5'
Ligation
(Ligase III)
Degradation and ligation
(DNase IV, PCNA,
DNA ligase I)
5'
3'
3'
5'
5'
3'
3'
5'

MutS and MutL, indicating that this repair pathway has been conserved during evolution. At this point in time, whilst there is some understanding of how mismatches are recognised, little is known about the mechanisms which identify the incorrect strand or the processes involved in the excision and repair of the mismatched sequences.

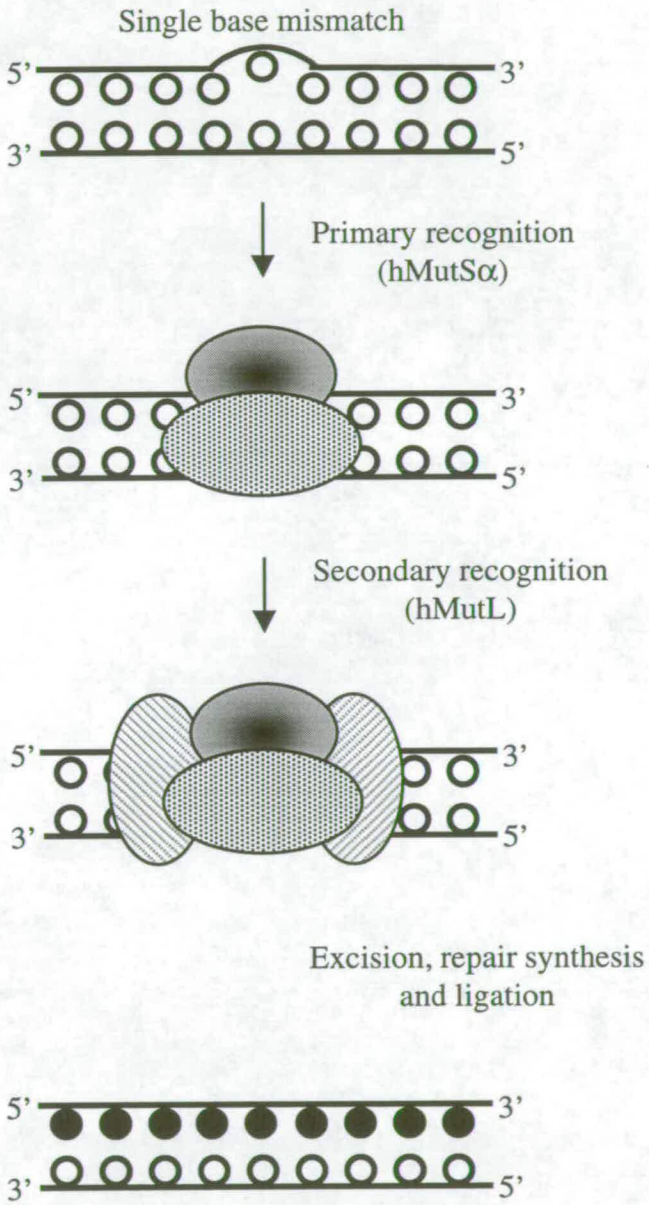
Mismatch recognition is performed by heterodimeric protein complexes hMutSα (hMSH2 and hMSH6, human homologues of MutS) or hMutSβ (hMSH2 and hMSH3, human homologues of MutS). The hMutSα complex binds preferentially to single base mismatches, single base and two base loops, whilst hMUTSβ recognises mismatches resulting in loops of more than two bases. The hMutL complex (hMLH1 and hPMS2, the human homologues of MutL) is recruited to the DNA-protein complexes whereby it acts as a secondary recognition factor which is required to bind before excision and repair can take place. Following mismatch recognition the incorrect base and surrounding sequences up to 1kb away are excised, before repair synthesis using the normal template strand takes place. The precise nature of this excision, polymerisation and ligation process remains to be determined for higher eukaryotes. Recently, the DNA replication protein, proliferating cell nuclear antigen (PCNA), has been implicated in MMR. It is possible that the interaction between MMR proteins and PCNA may be responsible for the strand discrimination. The subsequent excision and repair processes may employ proteins responsible for analogous functions in the nucleotide excision repair pathway (NER). Evidence supporting the interaction between MMR and NER was gained from yeast two-hybrid analysis which showed that the MMR MSH2 interacted with six different NER proteins (RAD1, RAD2, RAD3, RAD10, RAD14 and RAD25) (Bertrand *et al.*, 1998). The precise nature and biological significance of these interactions remains to be determined. A model for MMR is presented in Figure 1.2.

1.9 Nucleotide excision repair

Nucleotide excision repair (NER), the principle focus of the work to be presented in this thesis, is perhaps the most extensively studied of the repair pathways. This

Figure 1.2 Schematic representation of the mechanism of human mismatch repair

A model for the mechanism of mismatch repair, adapted from Lindahl 1997. The sugar-phosphate backbone is indicated by lines, whilst bases are depicted by circles. Open circles represent bases present in the DNA molecule before repair and closed circles are the bases introduced by the repair mechanism. Repair proteins are represented by shaded ellipses. See text for details.


Single base mismatch
5'
3'
3'
5'
Primary recognition
(hMutSα)
5'
3'
3'
5'
Secondary recognition
(hMutL)
5'
3'
3'
5'
Excision, repair synthesis
and ligation
5'
3'
3'
5'

pathway is capable of removing the broadest range of DNA lesions including the UV-induced photoproducts (cyclobutane pyrimidine dimers and 6-4 photoproducts), bulky chemical adducts and intrastrand cross-links. In the simplest of terms NER consists of five steps: damage recognition, dual incision, excision, repair synthesis, and ligation (reviewed by Sancar, 1996; Wood, 1996; and Lindahl, Karran and Wood, 1997).

As with the other repair pathways described already, the study of NER in higher eukaryotes was greatly aided by the characterisation of the less complicated NER pathway in *E. coli*. In *E. coli* DNA lesions are first recognised by the $UvrA_2B$ helicase complex, the UvrA subunits then dissociate leaving the UvrB bound at the site of damage where it recruits the UvrC subunit which in turn mediates the incisions to either side of the lesion. The lesion is then excised, by the combined action of UvrD and DNA polymerase I, as a 12-13 base oligonucleotide with the bound Uvr subunits. The resultant gap is filled in by DNA polymerase I and sealed by DNA ligase.

In mammals, NER consists of two subpathways, the first pathway global genome repair (GGR) deals with DNA lesions throughout the genome, whilst the second pathway, transcription-coupled repair (TCR), is involved in the repair of lesions in actively transcribed genes. The gene products involved in mammalian NER were largely identified by means of studies involving NER mutant cell lines from the seven xeroderma pigmentosum complementation groups (A through G), and UV-sensitive rodent cell lines defective in their excision repair cross complementing (*ERCC*) genes. It has been shown that some of the *ERCC* genes are identical to the XP genes, where this is true, the XP nomenclature has been adopted. The current model for mammalian NER indicates the involvement of 25 polypeptides.

The initial step in mammalian NER is the identification of the DNA lesion. This is believed to be performed by the proteins XPA and RPA which both bind preferentially to damaged DNA and have been shown to interact with each other (Robins *et al.*, 1991 and He *et al.*, 1995). In addition to the XPA-RPA complex, both XPC and XPE have been shown to bind preferentially to damaged DNA, although the precise role of XPE in NER is as yet undetermined. *In vitro* reconstitution studies

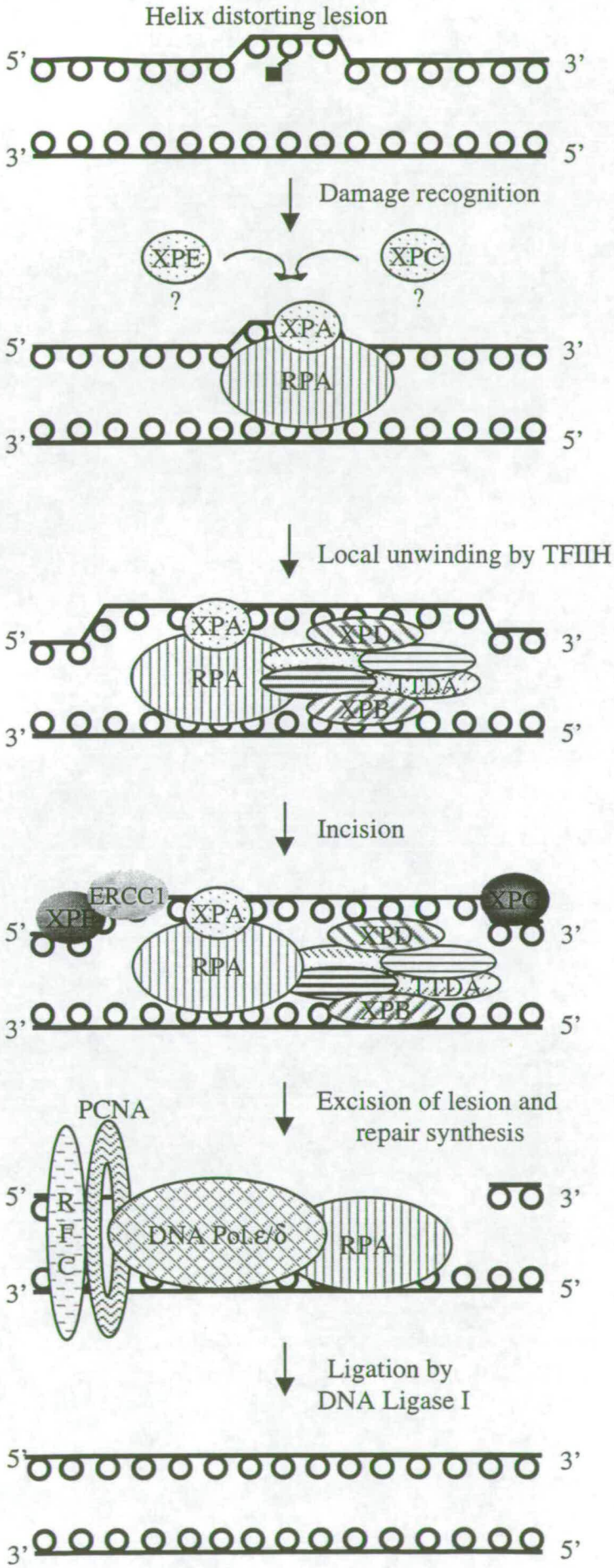
have shown that XPE is not essential for repair up to the point of incision *in vitro.* XPC is not required for transcription-coupled repair (Venema *et al.,* 1991). Recently published *in vitro* studies have shown that XPC, as a complex with HHR23B, may be responsible for damage detection in the global repair pathway prior to damage verification by XPA (Sugasawa *et al.,* 1998). The XPA-RPA complex recruits the basal transcription factor TFIIH, which has as two of its subunits the DNA helicases XPB and XPD (van Vuuren *et al.,* 1994), to the damaged site. The action of the helicases serves to unwind the DNA in the damaged region. This localised unwinding facilitates the access of the endonucleases to the region of damage. Whilst the DNA is unwound by the action of the helicases, RPA binds to the undamaged DNA strand where it is believed to prevent inappropriate incision of the undamaged strand by endonucleases.

The first incision at the damaged site is performed by the XPG nuclease, which binds to RPA on the 3' side of the lesion (O'Donovan *et al.,* 1994). The lesion-carrying strand is then cut at about 6 to 9 bases 3' to the lesion (Huang *et al.,* 1992 and Moggs *et al.,* 1996). The 5' incision is made by a second structure specific endonuclease, the ERCC1-XPF complex (Sijbers *et al.,* 1996), 16 to 25 bases 5' to the lesion. Both *in vivo* and *in vitro* studies have indicated that the ERCC1-XPF endonuclease is positioned at the damaged site by virtue of an interaction between ERCC1 and the XPA component of the bound XPA/RPA complex (Li *et al.,* 1994 and Saijo *et al.,* 1996). The damaged 22-30 base oligonucleotide is then removed with some of the repair proteins still bound (Mu *et al.,* 1996). The size of the excised fragment and the positions of the incisions relative to the lesion vary according to the nature of the lesion. The remaining single stranded gap is filled in by the concerted action of the still bound RPA, DNA polymerase δ or ε, PCNA, replication factor C (RFC) and finally DNA ligase seals the repair patch (Figure 1.3).

Bohr *et al.* (1985) first reported that UV-induced pyrimidine dimers were more efficiently repaired in active genes than in the genome overall. This analysis was later extended to show that the dimers were more efficiently removed from the transcribed strand and that the rate repair of the non-transcribed strand did not differ from overall genome repair. The precise nature of the mechanism behind this transcription-

Figure 1.3 Schematic representation of the mechanism of nucleotide excision repair in mammals

A model for the mechanism of mammalian nucleotide excision repair, adapted from Wood, 1996. The sugar-phosphate backbone is indicated by lines, whilst bases are depicted by circles. Open circles represent bases present in the DNA molecule before repair and closed circles are the bases introduced by the repair mechanism. Repair proteins are as identified and are represented by shaded ellipses. See text for details.


Helix distorting lesion
5'
3'
3'
5'
Damage recognition
XPE
?
XPC
?
XPA
RPA
5'
3'
3'
5'
Local unwinding by TFIIH
XPA
XPD
5'
3'
RPA
TTDA
XPB
3'
5'
Incision
ERCC1
XPF
XPA
XPC
XPD
5'
3'
RPA
TTDA
XPB
3'
5'
PCNA
Excision of lesion and
repair synthesis
5'
3'
R
F
C
DNA Pol ε/δ
RPA
3'
5'
Ligation by
DNA Ligase I
5'
3'
3'
5'

coupled repair remains to be determined. It is believed that the repair mechanism is initiated when the RNA polymerase II complex becomes stalled at a DNA lesion in the transcribed strand of the duplex. Additional polypeptides, including CSA and CSB, are believed to be involved in altering the conformation of the stalled RNA polymerase to facilitate the initial binding of the repair factors XPA, RPA (TFIIH is already bound in its capacity as a transcription factor), after which the repair process proceeds as described for GGR.

1.10 Human inherited disorders of nucleotide excision repair (NER)

Genomic instability is associated with a number of human inherited disorders including Fanconi anaemia (FA), ataxia telangiectasia (AT) and Blooms syndrome (BS). Each of these diseases is characterised by hypersensitivity to particular mutagenic agents and a predisposition to cancer. At a cellular level, these disorders are characterised by defective cellular responses to DNA damage as opposed to defects in excision repair or its auxiliary processes. Although cells from FA patients show defective repair of cross-links, additional evidence points to a primary defect of cell cycle control, failure to prevent the replication of cross-linked DNA resulting in chromosome damage. Cells from AT patients show hypersensitivity to ionising radiation thought to be as a consequence of the failure to arrest the cell cycle in order to permit the repair of strand breaks. The hypermutability associated with Bloom's syndrome is thought to be as a consequence of a reduced ability to resolve certain DNA structures generated during replication, leading to error prone recombination.

Defective NER has been associated with three diverse human inherited disorders xeroderma pigmentosum (XP), Cockayne's syndrome (CS) and the photosensitive form of trichothiodysrophy (PIBIDS). The clinical features of the classical forms of these disorders will be described in this section (reviewed in Friedberg, Walker and Siede, 1995 and Cleaver, 1995).

Xeroderma pigmentosum (XP) is a rare autosomal recessive disease, occurring in the population at between 1 in 250,000 and 1 in 40,000 according to geographic location. Patients are clinically characterised by severe photosensitivity of exposed regions of the skin, pigmentation abnormalities, a predisposition to skin cancers (2000-fold

higher than normal), elevated incidence of tumours on other sun exposed tissues including tongue and eyes, increased rate of internal cancers (20-fold higher than normal), and neurological abnormalities including deafness and progressive neurodegeneration, are evident in 20% of XP patients. The average age for the onset of skin neoplasias is 8 years, 50 years younger than in the general population.

Cockayne's syndrome (CS) is an extremely rare disorder in which the patients are clinically characterised by virtue of early presentation of UV sensitivity, with significant erythema following minimal sun exposure, dwarfed appearance following arrest of growth and development, impaired sexual development, progressive neurological abnormalities resulting from neurodemyelination and mental retardation. Additional features of CS include deafness, dental caries and retinal pigmentary degeneration. CS patients do not show pigmentation abnormalities or present a predisposition to cancers.

Trichothiodystrophy (TTD) is another autosomal recessive disorder. The clinical features of TTD include brittle hair and nails resulting from deficiency of cysteine-rich matrix proteins, ichthyosis (fish-like scaled skin), short stature, mental retardation, and distinctive facial features (protruding ears and receding chin). A form of TTD, with additional symptoms is represented by the acronym PIBIDS, indicating the presence of photosensitivity, intellectual impairment, brittle hair, ichthyosis, decreased fertility and short stature. Photosensitive TTD patients do not present a predisposition for skin cancers.

The principle clinical features of these disorders are summarised in the table overleaf (Table 1.1)

Table 1.1 Summary of the clinical features of XP, CS and photosensitive TTD

Clinical feature	XP	CS	TTD
UV-sensitivity	+ +	+	+
Abnormal skin pigmentation	++	-	-
Skin cancer	+ +	-	-
Brittle hair and nails	-	-	+
Short stature	-	+	+
Mental retardation	-	+	+
Neurodemyelination	-	+	?
Neurological degeneration	+	-	?

1.11 Genetic basis and complexity of the human inherited disorders XP, CS and TTD

Clinically, XP, CS and TTD present as quite distinct diseases yet there is evidence that these disorders do, infact, overlap with each other. In some cases patients have presented combined features of XP/CS and XP/TTD.

The cause of the XP phenotype was first identified as defective NER when cells from XP patients were seen to display increased sensitivity to DNA damaging agents and a reduced level of unscheduled DNA synthesis (UDS). It was noted that the levels of sensitivity and UDS showed marked variation between patients. The highest degree of sensitivity was found in cells from XPA patients whilst cells from XPE patients showed a near normal sensitivity to UV irradiation. Subsequent cell fusion studies revealed that this variation was a function of heterogeneity at the molecular level. XP patients fall into seven complementation groups (XPA through XPG) and an additional variant group XPV (clinical symptoms of XPV match the classical form without neurological abnormalities, cells do not exhibit NER deficiency). The genes for the A, B, C, D, F and G complementation groups have now been cloned. In many cases the XP genes have been found to be identical to genes isolated from earlier studies involving 11 UV-sensitive rodent complementation groups, mainly Chinese hamster ovary cell derived. The genes complementing these rodent cells

Table 1.2 Summary of NER disease complementation groups and genes.

Complementation Group (gene)	Cloned ERCC Gene	Clinical presentation	Gene function
XPA (Tanaka,1989)		XP	Damage recognition
XPB (Weeda *et al.*, 1990)	3	XP/ XP-CS / TTD	DNA helicase, subunit of TFIIH
XPC (Legerski and Peterson , 1992)	-	XP	Global genome repair
XPD (Weber, 1988)	2	XP/ XP-CS / TTD	DNA helicase, subunit of TFIIH
XPE	-	XP	Damage recognition ? (Gene not cloned)
XPF (Thompson,1994)	4	XP	Endonuclease, 5' of lesion
XPG (Mudgett and MacInnes, 1990)	5	XP / XP-CS	Endonuclease, 3' of lesion
XPV	-	XP	Unknown (Gene not cloned)
CSA	-	CS	Transcription-coupled repair (Gene not cloned)
CSB (Troelstra *et al.*, 1990)	6	CS	Transcription-coupled repair
TTDA	-	TTD	TFIIH subunit (Gene not cloned)
-	1 (Westerveld 1984)	-	Endonuclease, 5'of lesion

were denoted excision repair cross-complementing (*ERCC*). The relationship between the previously cloned *ERCC* genes and the XP complementation groups is summarised in Table 1.2.

Cell fusion experiments revealed that there were two complementation groups in the classical form of CS; CSA and CSB. However it has been shown that in addition to these groups patients with XPB, XPD or XPG can present a combined XP/CS phenotype. Similarly, it was found that the XPB or XPD genes complemented the underlying molecular defects in several TTD patients. These XP genes did not complement one additional TTD complementation group, TTDA. Together these findings suggest that in addition to the eight XP complementation groups, there are five 5 CS complementation groups (including the three XP/CS groups) and three TTD complementation groups (including the two XP/TTD groups).

The identification of the genetic complexity underlying these disorders raised a number of questions concerning the mechanisms by which mutations in certain genes could give rise to disorders with phenotypic consequences as divergent as a 2000-fold higher incidence to skin cancer (XP) and no predisposition to cancers (CS and TTD). This enigma will be considered further in the following section.

1.12 Correlation between NER deficiency and predisposition to cancer

If XP is first considered in isolation, it has been shown that most of the XP genes have a clear role in NER. Despite retaining the transcription-coupled repair pathway, XPC patients still exhibit a predisposition to skin cancers, suggesting that the cancer proneness is a function of defective GGR. The excepted model being that failure to repair DNA damage leads to mutations which then result in carcinogenesis when mutations occur in genes involved with the control of cell growth and division. The cancer predisposition of the NER proficient XPV group is apparently caused by a defective post-replication repair process.

At the molecular level the cells from XP, TTD and CS patients are all defective in NER, CS cells being deficient in only the transcription-coupled element of NER, yet at the clinical level the symptoms are so divergent. Despite the shared cellular defect in NER, it has been suggested that the clinical symptoms presented by CS and TTD

patients represent defects in basal transcription, reflecting the roles played by the relevant gene products in both transcription and DNA repair. Whilst it would be predicted that the complete loss of transcription would be lethal, it is not difficult to imagine how impaired transcription could lead to symptoms such as growth retardation. Consequently, both CS and TTD have been described as 'transcription syndromes'. This hypothesis will be considered by focusing on the XPD gene (reviewed by Chu, 1996). It has been shown that mutations in XPD are associated with XP, XP/CS and TTD. XPD is a subunit of the TFIIH basal transcription factor, which in addition to its role in NER is involved in loading RNA polymerase II onto promoters to initiate transcription. An XPD mutation affecting the repair capacity of TFIIH would result in XP. Whilst a mutation that impaired the transcriptional capacity of TFIIH could compromise the synthesis of certain proteins, for example the sulphur-containing proteins absent from the hair of TTD patients. This model is referred to as the 'transcription syndrome' hypothesis. Intrinsic, to this hypothesis is the suggestion that the phenotypic consequences of mutations in the XPD gene would correlate with particular regions (functional domains) of the XPD gene product. It appears that the mutations, characterised thus far, are clustered in the carboxy-terminal region of the gene and the mutations do not reveal either XP or TTD domains. It is clear that no single mutation is shared by an XP and TTD patient and, as such, the mutations appear to be XP or TTD specific. This hypothesis provides a plausible explanation for the 'non-repair' symptoms associated with TTD and CS. It appears that the overlapping phenotypic complexity of these disorders stems from the additional functions played by NER proteins in other cellular processes.

1.13 Cloning and characterisation of *ERCC1*

Of the many genes involved in the complex mechanism of NER, the work presented in this thesis focuses on *ERCC1*. A review of the *ERCC1* studies performed to date will be presented. Genetic complementation of UV sensitive rodent cell lines (1-11) led to the molecular cloning of the excision repair cross complementing genes 1 to 6 (*ERCC1-6*). As described previously, subsequent studies revealed that the majority of these genes were complementing genes for several of the XP groups (*ERCC2, 3, 4, 5*)

in addition to the CS group B (*ERCC6*). The remaining *ERCC* gene *ERCC1* failed to complement any of the known NER disease complementation groups (van Duin *et al.*, 1989), and has yet to be associated with any other known human disorder. All of the other cloned NER genes have been associated with inherited disorders and, as such, have been shown not to be essential for viability. It remains possible that a human *ERCC1* inherited disorder remains to be identified or that *ERCC1* deficiency is not compatible with viability in humans. The survival of xeroderma pigmentosum patients, completely lacking in NER, suggests that if *ERCC1* deficiency in humans is not compatible with survival it may have a further, as yet, unconfirmed function, the inactivation of which is lethal.

The human *ERCC1* gene, the first mammalian DNA repair gene to be cloned, was cloned by virtue of its ability to complement the UV sensitive CHO cell line 43-3B (Westerveld *et al.*, 1984). The *ERCC1* mutation in this cell line was only recently characterised as a missense mutation at the 98th amino acid residue (Hayashi *et al.*, 1998). The gene consists of ten exons spanning 15kb of chromosome 19 and gives rise to a major mRNA transcript of 1.1kb, with alternative splicing and differential polyadenylation resulting in additional transcripts of 1kb, 3.4kb and 3.8kb respectively (van Duin *et al.*, 1987). The role of the alternatively spliced 1kb transcript, lacking exon 8, is unclear as it has been demonstrated that it does not encode a repair proficient product (van Duin *et al.*, 1986). The cloned human *ERCC1* cDNA was used to isolate the mouse *ERCC1* homologue, which also consists of ten exons spanning about 16kb of mouse chromosome 7 (van Duin *et al.*, 1988). At the nucleotide level the sequence similarity between the two coding regions is 82%. The 3' end of the *ERCC1* gene, in mouse and man, overlaps with the 3' end of another gene *ASE1* (antisense *ERCC1*), the function of which remains to be elucidated (van Duin *et al.*, 1989).

In both cases, the expression of *ERCC1* is driven by a promoter devoid of all classical promoter elements such as GC-rich regions, TATA and CAAT boxes. *ERCC1* transcription is not induced by UV irradiation and occurs at low levels in all tissues and developmental stages examined to date. This finding is in agreement with the promoters of other NER genes including XPA (van Oostrom *et al.*, 1994).

1.14 The ERCC1 protein

The ERCC1 protein is known to complex with the XPF protein to form a structure specific endonuclease responsible for making the 5' incision in mammalian NER (Sijbers *et al.*, 1996). Additional functions for ERCC1 remain to be determined, although the hypersensitivity of *ERCC1* mutants to cross-linking agents and homology with yeast recombination proteins may indicate a role in a recombination pathway.

The *ERCC1* cDNA has an open reading frame that encodes a polypeptide of 297 amino acids in humans, and 298 amino acids in mouse, with an estimated molecular weight of 33 kDa. The overall identity between the encoded proteins is 85% with most of the amino acid variation being at the N-terminal region of the proteins. There is 70% identity in the first 100 amino acids, 97% identity in the next 100 amino acids and 89% identity in amino acids 200 to 298. The encoded proteins harbour a putative DNA binding domain and a nuclear location signal (NLS). The NLS is located at the N-terminal region of the protein between residues 17-23, whilst the DNA binding domain spans residues 134-156 (Hoeijmakers *et al.*, 1986).

The translated mouse and human *ERCC1* sequences show extensive homology with the *Saccharomyces cerevisiae RAD10* and *Saccharomyces pombe swi10*$^{+}$ genes. The homology between the mouse and human ERCC1 proteins and the yeast RAD10 and swi10^{+} is most significant in the middle region of the proteins (corresponding to the C-terminal regions of the yeast proteins). There is no significant homology between the N-termini of the proteins. In yeast, RAD10 complexes with RAD1 to form an endonuclease which makes the 5' NER incision, analogous to the role of the *ERCC1/XPF* complex in mammals. The yeast proteins are known to play a role in mitotic recombination, such a function remains to be shown for the mammalian homologues (Tomkinson *et al.*, 1993 and Bardwell *et al.*, 1994). The mammalian proteins also have an additional 83 C-terminal residues that are absent from the yeast homologues. This C-terminal 'extension' of the mammalian proteins shares homology with regions of the *E. coli* NER excision protein UvrC (van Duin *et al.*, 1986 and Hoeijmakers *et al.*, 1986). The observed homology points to the conservation of basic DNA repair mechanisms through evolution.

Mutational analysis has indicated that the repair capacity of the ERCC1 protein was not compromised by deletion of the first 91 N-terminal residues, whereas deletion of only 5 residues at the C-terminal resulted in complete inactivation of the protein (Sijbers *et al.*,1996). The C-terminal deletions effect the region of shared homology with the UvrC of *E. coli*, the C-terminal region of which was recently shown to be involved in DNA binding and incision (Moolenaar *et al.*, 1998). Similarly, it has been shown that the overexpression of an ERCC1 protein with only two additional residues inserted in the C-terminal region resulted the induction of sensitivity to cross-linking agents in repair proficient cells, suggesting that the mutant protein competed against the endogenous wild type ERCC1 to impair the repair process (Belt *et al.*, 1991). The low level of ERCC1 protein is a common feature of cells from XPF patients, suggesting that where ERCC1 is unable to form a complex the free protein is not stable (Yagi *et al.*, 1997).

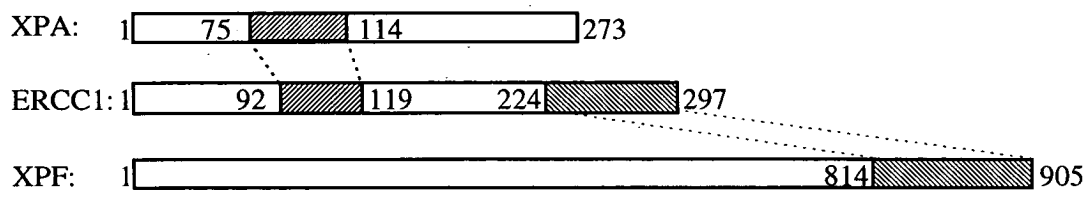
1.15 The interactions of ERCC1 with XPF and XPA

The first evidence of an interaction between ERCC1 and the XPF proteins came from the observation that cell free extracts from rodent *ERCC1* and *ERCC4* mutant cells as well as human XPF cells did not complement each other *in vitro*. This indicated that the complementing activity that was absent from the *ERCC1* mutant cells was in some way associated with the complementing activity of *ERCC4* and *XPF* cells. It was already known that the *ERCC1* cDNA did not complement either *ERCC4* or *XPF* cells, but at this point although the complementing *ERCC4* and *XPF* genes had yet to be cloned they were believed to be equivalent. Furthermore it was shown that when ERCC1 was depleted from wild type cell extracts (using antibodies), the depleted extract was no longer able to complement the *ERCC4* or *XPF* cells (van Vuuren *et al.*, 1993). It was subsequently demonstrated that the correcting activities for the *ERCC1, ERCC4* and *XPF* co-purified as a high molecular weight species of approximately 100-120 kDa. This complex was used to correct the repair defects of each of the groups *in vitro* (Biggerstaff *et al.*, 1993). Further purification and characterisation of the complex showed that it contained ERCC1 and a 112kDa species, presumably ERCC4 or XPF, which was capable of endonuclease activity

(Park *et al.*, 1995). It was later confirmed that *ERCC4* and *XPF* were equivalent (Sijbers *et al.*, 1996). Thus it was known that ERCC1 formed a heterodimeric complex with XPF to constitute an endonuclease.

Only recently has the precise nature of the interaction between ERCC1 and XPF been determined. Using a series of C- and N-terminal truncated recombinant ERCC1 and XPF proteins it has been demonstrated that the minimal domain for stable XPF binding spans residues 224 to 297 of the human ERCC1, whereas residues 814-905 in XPF were deemed sufficient for the stable binding to ERCC1 (de Laat *et al.*, 1998). Somewhat surprisingly, the interaction domains in ERCC1 and XPF do not coincide with the binding domains previously identified for the *S. cerivisiae* homologues RAD10 and RAD1. Indeed, the binding domain in ERCC1 maps to the C-terminal extension that shares homology with the *E. coli* NER protein UvrC. The relative positions of the protein interaction domains of ERCC1, XPF and XPA are illustrated in the Figure below (Figure 1.4).

Figure 1.4 Summary of ERCC1 interaction domains



A schematic representation of the known interaction domains of ERCC1, XPF and XPA. Amino acid residues (human) are numbered and the interaction domains are indicated by diagonally striped boxes.

In addition to its endonuclease forming association with XPF, ERCC1 has also been shown to interact with the XPA protein. The currently accepted model for NER suggests that DNA lesions are first recognised and bound by XPA, the primary recognition protein. It is believed that this interaction is responsible for the appropriate positioning of the endonuclease at the damaged site. XPA itself interacts

with numerous other factors including RPA and the basal transcription factor TFIIH. The ERCC1/XPF complementing complex was isolated from cell free extracts by binding to an XPA affinity column, demonstrating the formation of a ternary complex between XPA and the ERCC1/XPF heterodimer (Park and Sancar, 1994). Two hybrid analysis and *in vitro* studies with truncated recombinant proteins have shown that there is a direct interaction between residues 92-119 of ERCC1 and residues 75-114 of XPA (Lei *et al.*, 1994) (Figure 1.4). Analysis of the UV sensitive rodent cell line CHO43-3B has shown that the mutant form of ERCC1 expressed in this cell line is capable of binding with XPF *in vitro* but fails to interact with XPA (Hayashi *et al.*, 1998).

1.16 Gene targeting

The simplest and most direct method of determining the function of any gene is to assess the consequences of its loss of function. Whilst a number of animal models of human disorders have been identified following spontaneous and randomly introduced mutations, simple factors of animal husbandry mean that to rely on this approach to identify and study the consequences of specific gene inactivating mutations would be too time consuming and extremely costly. The development of gene-targeting technology has meant that we are no longer reliant on random mutagenesis events to provide animal models of human disorders. Gene targeting is defined as "homologous recombination between DNA sequences residing in the chromosome and newly introduced DNA sequences, enabling the transfer of any desired alteration of the cloned sequences into the genome of a living cell" (Cappechi, 1989). Whereas it was initially only possible to 'knockout' genes, today's advanced technology means a variety of other alterations are possible (reviewed by Melton, 1994). In many cases more subtle gene alterations are required to faithfully mimic actual human disorders, as with the point mutations in the prion protein (*PrP)* gene associated with familial Creutzfeldt-Jakob disease (a neurodegenerative disorder) (Moore *et al*, 1995). In cases where the inactivation of a gene results in embryonic lethality (Bentley *et al*, 1996) conditional deletion of the gene in specific

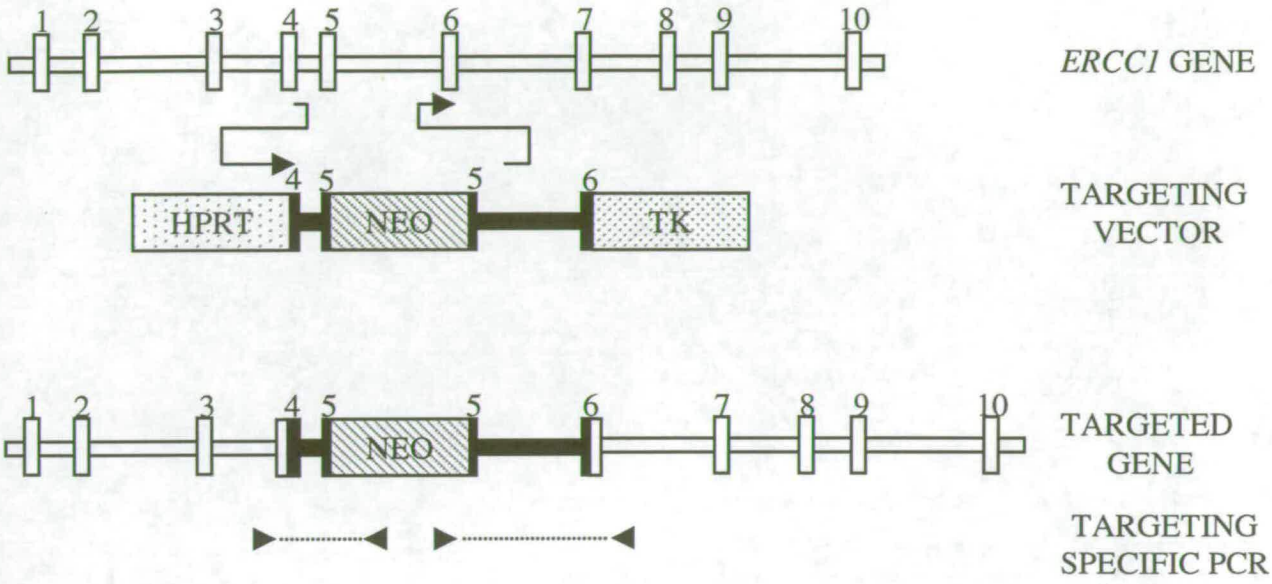
tissues of the adult can be used as a means of studying the consequences of loss of gene expression in later stages of development (Selbert *et al.*, 1998).

The success of a gene-targeting strategy relies on the ability of the introduced cloned sequences to recombine with homologous endogenous sequences and, most importantly, the ability to enrich for, and identify these events. When DNA is introduced to a cell it can either randomly integrate into the genome, or integrate by homologous recombination. To enrich for the latter event a positive-negative selection strategy is employed (Mansour *et al.*, 1988). Positive selection simply identifies those cells that have integrated the exogenous targeting vector DNA. The most frequently used positive selection system involves selecting for the bacterial neomycin phosphotransferase gene (*neo*^{r}) by resistance to culturing in G418. Selection against random integration events is achieved by flanking the homologous sequences with negative selectable markers, such as the Herpes simplex virus *thymidine kinase* gene (*TK*) or, alternatively, a mouse hypoxanthine phosphoribosyltransferase (HPRT) minigene. In the event of a random integration event, the linearised targeting vector integrates by its ends and, consequently, the negative selectable markers will be integrated into the genome. Selection against these markers is achieved by the culturing of transfected cells in medium with the toxic thymidine analogue gancyclovir (GC) and 6-thioguanine (6-TG), respectively. After selective enrichment, homologous recombination events are then confirmed using a PCR assay specific for the targeted allele before further comprehensive analysis by Southern blot. The targeted inactivation of the mouse *ERCC1* gene by Selfridge *et al.* provides an illustration of the successful application of such a positive-negative selection strategy. In this instance positive selection and insertional inactivation was achieved by inserting a *neo*^{r} cassette into exon 5 of the *ERCC1* gene. Negative selection was enabled by including both, a *TK* and an *HPRT* minigene at either end of the region of homology (Figure 1.5).

Figure 1.5 Strategy for the targeted inactivation of *ERCC1* in mouse embryonic stem cells (HM-1)

A representation of the positive negative selection procedure adopted for targeting strategies in *HPRT*-deficient embryonic stem cells. The 15kb mouse *ERCC1* gene is shown schematically, exons are numbered and shown by vertical open boxes and introns are depicted by horizontal open boxes. The linear targeting vector is also shown schematically, homologous sequences of the targeting vector are represented by filled boxes and exons are numbered. *HPRT* minigene is represented by a heavily stippled box. Herpes simplex virus thymidine kinase gene (*TK*) is represented by a lightly stippled box. Neomycin transferase cassette (NEO) is represented by a diagonally striped box.

The targeting vector consisted of a region of homology spanning exons four to six of the *ERCC1* gene, with two flanking markers for negative selection as shown. In the event of random integration of the vector sequences the negative selectable markers are also integrated into the genome. Only after homologous recombination events are the cells are neomycin positive and G418 resistant, HPRT negative and 6TG resistant, *TK* negative and gancyclovir resistant. The arrows indicate the predicted homologous recombination event. The structure of the targeted *ERCC1* allele is shown schematically and shows the integrated vector sequences as filled boxes and the endogenous gene sequences as open boxes. The arrowheads indicate the positions of primers for targeting specific PCRs (adapted from Selfridge *et al.*, 1992)


1 2 3 4 5 6 7 8 9 10
ERCC1 GENE
4 5 5 6
HPRT
NEO
TK
TARGETING VECTOR
1 2 3 4 5 5 6 7 8 9 10
NEO
TARGETED GENE
TARGETING SPECIFIC PCR

1.17 Production of knockout mice using an *HPRT*-deficient embryonic stem cell line (HM-1)

The targeting of an allele in cultured embryonic stem cells is generally the first step towards generating mice heterozygous and homozygous for the targeted alteration (Mcwhir *et al*, 1993 and Redhead *et al*, 1996). Mouse embryonic stem cells are derived from the inner cell mass of blastocysts, *in vivo* these cells differentiate to form all cell types in the developing mouse. When cultured *in vitro* the cells maintain their pluripotency (Evans *et al*, 1981). Thus, following successful gene targeting, ES cells can be microinjected into a host blastocyst, which is then surgically re-implanted into the uterus of a pseudopregnant recipient mouse. The pluripotent stem cells are capable of contributing to all tissues of the developing animal, including the germline. The chimaeric offspring can be identified by means of their heterogeneous coat colouring. For instance, if the stem cell line in question is derived from a chinchilla coloured 129/Ola mouse, when these cells are injected into a BalbC (white coat) host blastocyst any chimaeras will have a mixed chinchilla and white coat. The extent of chimaerism will be reflected by how much chinchilla colouring is evident in the coat. If the pluripotent targeted stem cells have indeed contributed to the germline, any alteration to the genome can be transmitted to the offspring of the chimaeras following test matings. Germline transmission can also be monitored by means of observing the coat and eye colours of the offspring from the chimaeras, where ES cell derived progeny result in different colours from host embryo derived progeny, due to the presence of different markers in the two strains. In the event of germline transmission of a targeted autosomal locus 50% of the offspring will inherit the targeted allele. Thus by means of gene targeting in mouse embryonic stem cells and generation of chimaeras by blastocyst injection it is possible to generate mice carrying an engineered genomic alteration (Figure1.6).

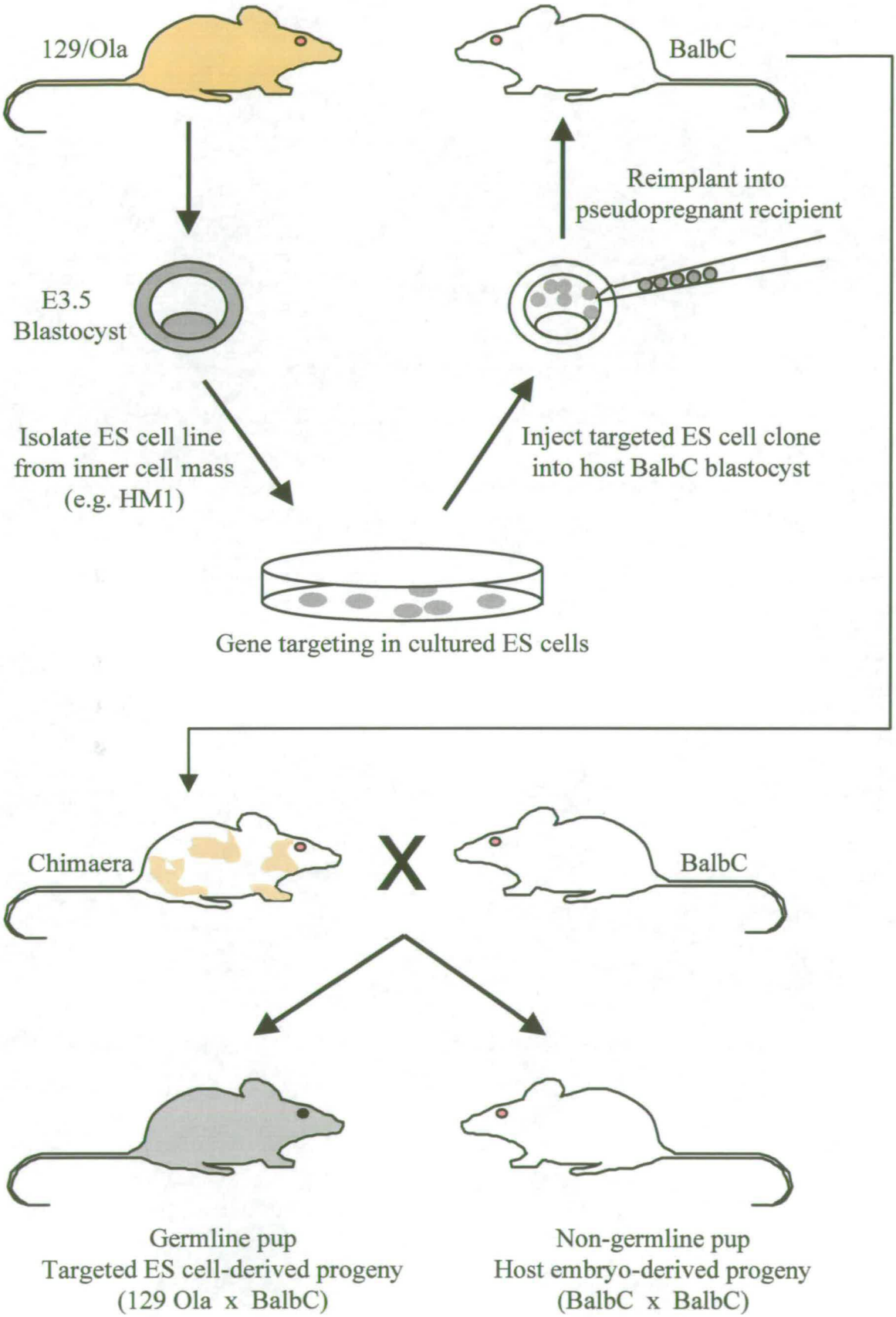
The *HPRT*-deficient stem cell line HM-1 was derived from blastocysts from an *HPRT*-deficient 129/Ola mouse line. This mouse strain had been generated by blastocyst injection of E14TG2A a spontaneous *HPRT*-deficient derivative of the wild type E14 stem cell line (Hooper, 1987). The HM-1 cell line was shown to contribute to the germline more frequently than the original E14TG2A line (Magin *et*

Figure 1.6 Generation of mice from gene targeted embryonic stem cells

Schematic representation of the procedure involved in generating mice heterozygous for a targeted gene alteration (adapted from Melton, 1994).

ES cells are isolated from the inner cell mass of blastocysts from strain 129/Ola mice. Gene targeting is performed whilst in culture and targeted ES cells are injected into the blastocyst cavity of a host embryo from a different strain (e.g. BalbC). Injected embryos are implanted into pseudopregnant foster mothers for development to term. Chimaeric offspring are identified by means of altered coat colour.

To assess whether the ES cells have contributed to the germline of these animals, the chimaeras are test mated (e.g. with BalbC). Germline transmission of the 129/Ola derived gametes is determined by transmission of the 129/Ola markers for coat and eye colour. The 129 derived HM-1 cells are homozygous for the white bellied agouti (A^W) allele, and homozygous for the chinchilla (c^{ch}) allele at the albino locus, which confers a tan (chinchilla) coat colour. The HM-1 cells also carry a homozygous recessive allele (*p*) at a third locus (*P*) which confers pink eyes. The BalbC recipient embryos and studs are homozygous for the wild type alleles at the agouti (*A*) and pink eyed (*P*) loci as well as being albino (*c/c*). The albino locus is epistatic to the pink eyed locus so BalbC have white coats and pink eyes. Following test matings between HM-1 chimaeras and BalbC studs germline transmission of the ES cell markers is recognisable at birth; germline pups are grey with black eyes and will have the genetic make-up c^{ch}/c, *p/P*, A^W/A. The combination of a single chinchilla allele with the alleles of the agouti locus results in a grey coat colour, and eye pigmentation is also restored. 50% of the germline offspring will carry the targeted allele. Non germline pups have the genetic make-up of the BalbC strain and are thus white with pink eyes.


129/Ola
BalbC
Reimplant into
pseudopregnant recipient
E3.5
Blastocyst
Isolate ES cell line
from inner cell mass
(e.g. HM1)
Inject targeted ES cell clone
into host BalbC blastocyst
Gene targeting in cultured ES cells
Chimaera
X
BalbC
Germline pup
Targeted ES cell-derived progeny
(129 Ola x BalbC)
Non-germline pup
Host embryo-derived progeny
(BalbC x BalbC)

al., 1992). The male HM-1 cells are *HPRT*-deficient by virtue of lacking the first two exons of the X-linked *HPRT* allele (Thompson, 1989). HPRT catalyses an early step in the purine salvage pathway in mammals. It is possible to select against *HPRT* activity *in vitro* by culturing cells in medium with the toxic purine analogue 6-TG. Thus *HPRT* minigenes can be used as negative selectable markers when targeting is performed in HM-1 cells (see Figure 1.5). When targeting is performed in an *HPRT*-deficient ES cell line such as HM-1 it is also possible to select for *HPRT* activity. *HPRT* is selected for by culturing cells in medium with hypoxanthine, aminopterin and thymidine (HAT). Aminopterin inhibits *de novo* synthesis of purines and pyrimidines. The *HPRT* activity enables cells to bypass the aminopterin induced metabolic block by utilising hypoxanthine and thymidine via the purine and pyrimidine salvage pathways. Thus *HPRT* minigenes can also be used as positive selectable markers if necessary (see section 1.17).

1.18 Introduction of subtle gene alterations by 'double replacement targeting'

In many instances the complete inactivation or 'knockout' of a gene is not the desired result of a gene targeting experiment. When attempting to model a human disorder it is often the case that the human mutation being modelled does not actually result in a completely inactive gene, thus it would not be possible to make direct comparisons between the phenotypic consequences of a partially active human allele and a null allele in a mouse model. To be able to mimic the more subtle mutations of a human disorder a method of 'double replacement' gene targeting has been developed (see Figure 1.7)(Stacey *et al.*, 1994). This targeting strategy requires the use of an *HPRT*-deficient ES cell line such as HM-1 in order to exploit the ability to select for and against *HPRT* activity. A first round of gene targeting is performed to introduce an *HPRT* selectable marker into the target locus. The *HPRT* minigene is selected for by culturing cells in medium supplemented with HAT. In this round of targeting, negative selection is enabled by incorporating a *TK* gene into the targeting vector. The first round of targeting generates a knockout allele.

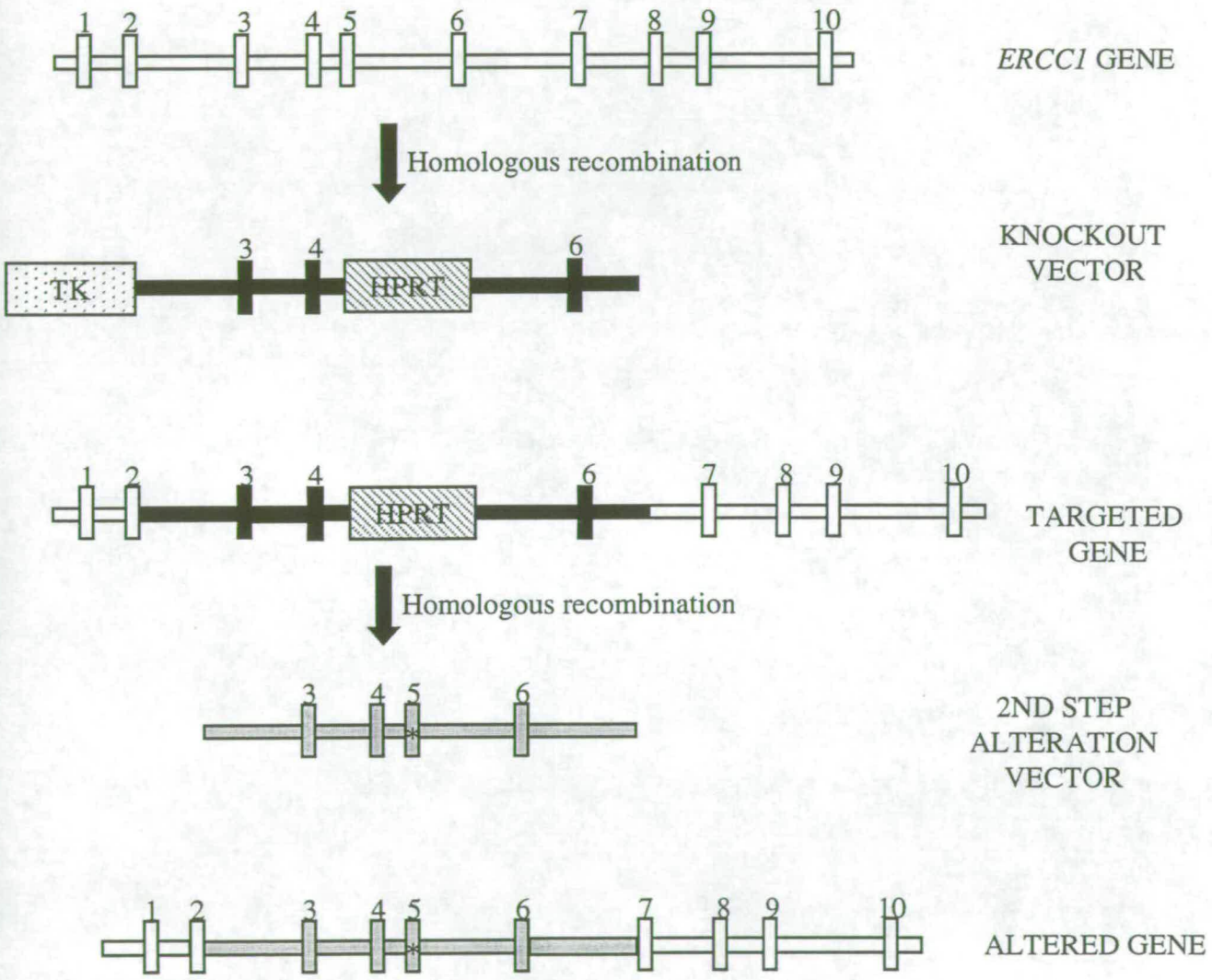
The second targeting step is designed to replace the inactivating *HPRT* marker gene with sequences homologous to the target gene bearing an engineered subtle

Figure 1.7 Double-replacement gene targeting strategy

A hypothetical two step 'double replacement' strategy for the introduction of a point mutation in exon five of the mouse *ERCC1* gene is shown schematically. The endogenous gene is represented by vertical open boxes for exons and horizontal boxes for introns. The homologous sequences of the first round 'knockout' vector are shown in black, the *HPRT* minigene is shown as a striped box and the *TK* gene is shown as a stippled box. Homologous sequences of the second round targeting vector are shown in grey and the subtle alteration is indicated by an asterisk.

The first targeting step is designed to knockout the gene by replacing one of the exons and its flanking sequence with an *HPRT* minigene. The knockout vector also contains a TK gene to enable positive-negative selection. Selection at this stage is for HPRT (HAT) and against TK (gancyclovir). Homologous recombination events are confirmed by PCRs specific for the targeted allele.

In the second targeting step the *HPRT* marker itself is replaced by the gene segment carrying the subtle alteration. Selection at this stage is for the loss of *HPRT* activity using 6TG. Random integration of the vector will not result in loss of the minigene. The presence of the subtle alteration can be confirmed using PCR analysis, often exploiting restriction fragment length polymorphisms resulting from the alteration.


1 2 3 4 5 6 7 8 9 10
ERCC1 GENE
Homologous recombination
TK
3 4
HPRT
6
KNOCKOUT VECTOR
1 2 3 4
HPRT
6 7 8 9 10
TARGETED GENE
Homologous recombination
3 4 5 6
2ND STEP ALTERATION VECTOR
1 2 3 4 5 6 7 8 9 10
ALTERED GENE

alteration. Selecting for the loss of the *HPRT* marker gene activity enriches for homologous recombination events. Selection against *HPRT* activity is achieved by culturing cells in the toxic purine analogue 6-TG. At both targeting steps the homologous recombination events are identified using PCR. As the second step introduces a subtle change at the target locus it is often necessary to use restriction fragment length polymorphism and, or sequence analysis to confirm the presence of point mutations(Moore *et al.*, 1995).

1.19 Conditional gene targeting

The deletion of a particular gene may result in an embryonic lethal (Bentley *et al.*, 1996) or early developmental lethal (McWhir *et al.*, 1993) phenotype and as such further analysis of gene function *in vivo* is somewhat limited. A further advance in gene targeting methodology, utilising site specific recombinases and homologous recombination, has provided a means of bypassing such phenotypic constraints. This strategy enables genes to be deleted in specific tissues or at desired stages of development (reviewed by Kilby *et al.*, 1993 and Cohen-Tannoudji, 1998).

The most frequently used method exploits the *loxP/Cre* recombinase system from the bacteriophage P1. Cre (causes recombination) has been identified as a site specific recombinase that is capable of catalysing recombination between 34bp recognition sequences called *loxP* (locus of cross over) sites. As part of its life cycle the bacteriophage P1 replicates as a plasmid in *E. coli.* The *loxP/Cre* recombinase system serves to maintain the replicating genome as a monomer, dimers are resolved by Cre mediated recombination. When the *loxP* sites are arranged in the same (head to tail) orientation the recombination event results in the deletion of the intervening sequences as a circular molecule carrying one of the *loxP* sites, the second *loxP* site remains in its original position. To exploit this mechanism for the purposes of gene targeting, conventional targeting strategies are employed to flank the gene or gene segment of interest with two *loxP* sites (described as floxing) in a head to tail orientation. With expression of *Cre* the sequences lying between the *loxP* sites are excised and the gene is inactivated. Inversion of the sequences between two *loxP* sites occurs if the recognition sequences are arranged in a head to head orientation.

Furthermore, it is possible for intermolecular recombination events to occur between *loxP* sites on different chromosomes, resulting in reciprocal translocations.

Following the floxing of a locus in ES cells, the key to the success of this system lies with the chosen method of *Cre* expression. It is possible to observe Cre-mediated recombination *in vitro* following the transient expression of *Cre* in the targeted ES clone (Xiao and Weaver, 1997). In the context of conditional or stage specific gene deletion, it is first necessary to produce germline mice with the floxed allele. The recombinase is then introduced by mating the floxed mice with mice carrying a *Cre* transgene under the control of an appropriate tissue specific, stage specific or inducible promoter (Schwenk *et al.*, 1998). The floxed locus will only be deleted in those cells where or when *Cre* is expressed. Thus conditional gene targeting can be used to monitor the consequences of gene inactivation that might otherwise have been lethal. It would be difficult to underestimate the potential in this area of gene manipulation. The limiting factor at this stage is the production and characterisation of the appropriate *Cre* expressing transgenics. Advances in this technology will be facilitated by a recent agreement to establish a 'library' of *Cre* mice and the setting up of a database of all the *Cre* mice characterised so far (meeting review by Rossant and McMahon, 1999).

1.20 Animal models

As the number of cloned human genes continues to grow so, it seems, does the number of knockout mice. The ability to modify the mouse genome by means of homologous recombination in embryonic stem cells facilitates the use of mouse models to study most biological systems and to model human diseases related to loss of gene function. One of the systems that has benefited from this form of analysis is the mechanism of DNA repair. A recent review of the field listed 52 independent mouse knockout models that are associated with various cellular responses to DNA damage (Friedberg *et al.*, 1998). Several of these mutant models involve genes associated with human diseases arising from defective DNA repair, such as xeroderma pigmentosum (e.g. Sands *et al.*, 1995). However, it is often the case that the phenotypic consequences of gene inactivation in the mouse do not accurately

model the consequences in humans. For example, *HPRT* deficiency in man results in the severe neurological disorder Lesch-Nyhan syndrome (Lesch and Nyhan, 1964) whilst *HPRT*-deficient mice are phenotypically normal, reflecting species differences in the affected purine metabolism pathway (Hooper *et al.*, 1987). Whilst this in itself can provide useful insights into differences between analogous biological systems in man and mouse, it can restrict any development of therapies for the diseases in question. With this in mind, examples of mouse models of defective NER are considered below.

1.21 *XPA*-deficient mice: a model for XP?

Of the seven human XP complementation groups XPA patients generally present the most severe symptoms including neurological abnormalities in 20% of cases (Cleaver, 1995). There have been two independent reports describing the generation and characterisation of *XPA*-deficient mice (de Vries *et al.*, 1995 and Nakane *et al.*, 1995). Whilst the structure of the targeted alleles differed, a *neo* cassette inserted into exon four of one mouse and exons three and four deleted in the other, both mice were confirmed as complete nulls by northern or western blot analysis. In both cases the mice were perfectly viable, apparently normal and fertile. NER deficiency was demonstrated in terms of the hypersensitivity of fibroblast cell lines, derived from these animals, to UV and genotoxic compounds. Residual repair of <5% was measured in the mouse fibroblasts, compared to <2% seen in human XPA cells. A marked predisposition to UVB induced skin cancer, relative to wild type littermates, was also reported for both 'knockout' strains. A subsequent publication reported that 21% of XPA mice aged over 1 year developed spontaneous tumours of the liver (de Vries *et al.*, 1997). In this context the animals accurately modelled the human patients and were suitable models for studying the high incidence of skin cancers in group A XP patients.

At the ages of 13 and 18 months neither of the mouse models exhibited any of the neurological abnormalities, including progressive neurodegeneration, associated with patients from the XPA complementation group.

1.22 *CSB*-mutated mice: a model for Cockayne's syndrome?

The *CSB* gene was disrupted by the insertion of a premature stop codon in the fifth exon, to closely mimic a mutation previously characterised in a human patient. Whilst the mutated allele gave rise to a mRNA fusion transcript, no wild type protein was detected. The mutant CSB mice were viable and showed a small but significant growth deficit compared to wild type animals. Behavioural analysis was interpreted to suggest evidence of mild neurological dysfunction. In this context the mice were said to display mild Cockayne's syndrome symptoms. Unlike human CSB patients the fertile mice did not exhibit impaired sexual development nor did they show any signs of dysmyelination.

As with the human cells, the fibroblast cell lines derived from the CSB mice were defective in the transcription-coupled repair, whilst global genome repair was unaffected. Although human CSB patients are sensitive to UV they do not show any increased incidence of skin cancer. However, the CSB mice were not only sensitive to UV but revealed a clear predisposition to skin cancers. It was suggested that the difference in cancer predisposition observed between humans and mice was a result of global genome repair in the mouse being less efficient than in humans. Further it was felt that the short lifespan and repeated hospitalisation of CS patients may conceal any predisposition to UV induced skin cancers in humans.

1.23 XPD mutated mice: a model for trichothiodystrophy?

The complete inactivation of *XPD* in the mouse resulted in a pre-implantation embryonic lethal phenotype (de Boer *et al.*, 1998). The severity of the phenotype was attributed to the abolition of basal transcription following the inactivation of the TFIIH transcription factor. Subsequently, mice were generated carrying a point mutation in the *XPD* gene which mimicked a mutation characterised in a patient with the photosensitive form of TTD. The arginine 722 to tryptophan mutation in exon 22 was engineered using a single-step 'gene-cDNA fusion targeting' strategy. Basically, the 3' end of the mouse gene was replaced with the equivalent 3' end sequences from the human cDNA, carrying the point mutation. It was argued that this approach of

'humanising' the C-terminal region of the mouse protein would result in a more accurate reflection of the human patient.

Mice homozygous for this mutation reflected several of the symptoms of TTD patients. The observed UV sensitivity, brittle hair due to deficiency of the cysteine rich matrix proteins, retarded growth, skin abnormalities, reduced fertility, and reduced lifespan all faithfully mirror the human condition. The *XPD* mutated mice suffered occasional tremors which may have been indicative of neurological abnormalities, although no defects in myelination were seen. It has been suggested that the manifestation of neurological abnormalities in humans is as a consequence of greater accumulation of DNA lesions in cells of the nervous system.

It was noteworthy that the mouse embryonic fibroblasts with this mutation exhibited the same increased UV sensitivity (1.5 fold higher than wild type) as the equivalent human cells. However, as with the CSB mice the TTD mice revealed a clear predisposition to UVB and chemical-induced skin cancers, representing a major divergence from the TTD patients. Once more it was suggested that the increased incidence of cancers in the mouse might reflect a less efficient global repair mechanism in these animals relative to humans. As the components of NER are conserved between the two species this would suggest that the differences in cancer predisposition, seen between animal models and the human patients, may lie in a repair regulating mechanism which differs between the two species. It is possible that cell cycle arrest in rodent cells might not be as efficient as in humans and as such rodent cells may be more inclined to proceed with cell division without completion of DNA repair, resulting in an accumulation of mutations.

1.24 *ERCC1*- deficient mice: the original NER deficient mouse model

ERCC1 deficiency has yet to be associated with any human disorder and it had been suggested, on this basis, that *ERCC1* deficiency in the mouse would result in an embryonic lethal phenotype. The *ERCC1*-deficient mice produced by this laboratory constituted the first mouse model of defective NER (McWhir *et al.*, 1993). Since the production of the first *ERCC1*-deficient mice, we have produced a second independent *ERCC1* knockout line (Melton, unpublished data) and two additional

ERCC1 targeted lines have been reported by a group in Holland (Weeda *et al.*, 1997). In our original knockout, inactivation was brought about by the insertion of a *neo*^{r} cassette into exon five, whilst the first of the Dutch knockouts involved the insertion of a *neo*^{r} cassette into exon seven. The second line reported by the Dutch involved the introduction of a subtle mutation; a seven amino acid carboxy-terminal truncation was engineered by introducing a premature stop codon at residue 292. Our second *ERCC1* knockout line (KT#209), was produced as an intermediate step towards the production of mice carrying a floxed *ERCC1* allele, lacks exons 3, 4 and 5 and has an *HPRT* minigene inserted in between exons 2 and 6 (Figure 1.8a). Although the structures of the targeted alleles vary between each of the mice, if all of the targeted alleles are complete knockouts less variation in the phenotypic consequences would be anticipated.

We first reported that *ERCC1*-deficient mice were viable but died prematurely before weaning at the age of three weeks. The *ERCC1* nulls were severely runted compared to wild type littermates, weighing only 20% of wild type (Figure 1.8b). Detailed analysis suggested that the cause of death in these animals was likely to be hepatic failure. At birth, some hepatocyte nuclei were already enlarged and polyploid; by the age of three weeks, the number of enlarged nuclei had increased and polyploid and aneuploid nuclei were detected. Polyploidy is normally only observed in adult livers. Further, elevated levels of p53 were detected in the liver, brain and kidneys of the *ERCC1* nulls, indicating persistence of DNA damage in these tissues. Despite the detection of p53 there was no evidence of increased apoptosis in the corresponding tissues. The general cachexic appearance of the mutant mice was characteristic of significantly aged animals. Northern blot analysis carried out on tissues from the *ERCC1* nulls showed that an *ERCC1/neo*^{r} fusion transcript, believed to result from polyadenylation within the *neo*^{r} cassette, was detectable.

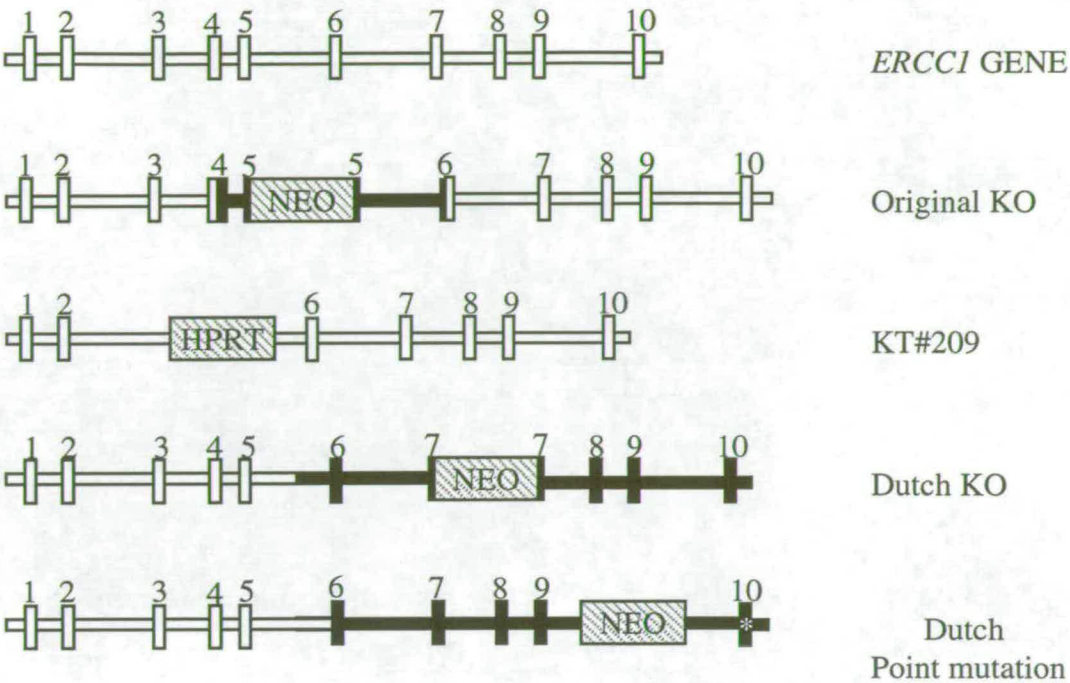
The Dutch *ERCC1* knockout mice survived for up to 78 days whilst those carrying the subtle mutation lived for up to 6 months, but were infertile. The improved survival of the Dutch mice relative to our nulls may be a consequence of differing genetic backgrounds or could possibly reflect a residual ERCC1 activity in these lines. In addition to the phenotypic consequences previously reported by ourselves,

Figure 1.8 Comparison of *ERCC1* targeted alleles and the consequences of *ERCC1* inactivation in the mouse

a, Schematic representation of the known *ERCC1* targeted allele structures. Exons are numbered and indicated by vertical boxes , introns are indicated by horizontal boxes. Selectable markers are represented by shaded boxes, and the point mutation is indicated by an asterisk. The regions of the targeted alleles that correspond to the integrated sequences of the receptive targeting vectors are shown in black. The original knockout allele involved inactivation by insertion of a *neo*^{r} cassette in exon 5, the KT#209 allele is inactivated by replacement of exons 3 to 5 with an *HPRT* minigene. The Dutch knockout allele involved inactivation by insertion of a *neo*^{r} cassette in exon 7, the subtle alteration involved the introduction of point mutation (stop codon) in exon 10 and a selectable marker into intron 9.

b, Illustration of the severe runting caused by *ERCC1* deficiency. Comparison of a three week old *ERCC1* null (right) and a wild type littermate (left)

a.


1 2 3 4 5 6 7 8 9 10
ERCC1 GENE
1 2 3 4 5 NEO 5 6 7 8 9 10
Original KO
1 2 HPRT 6 7 8 9 10
KT#209
1 2 3 4 5 6 7 NEO 7 8 9 10
Dutch KO
1 2 3 4 5 6 7 8 9 NEO 10 *
Dutch
Point mutation

b.

the Dutch mice also exhibited nuclear abnormalities of the kidney and additional histological evidence indicative of renal dysfunction. They also reported an absence of subcutaneous fat and early onset deposition of ferritin in the spleen, which can result from an increased turnover of erythrocytes. The development of these additional features is likely to be as a consequence of the extended lifespan of these animals relative to our nulls. All in all, it is suggested that the phenotype associated with *ERCC1* deficiency is more characteristic of the consequences of cell cycle arrest and premature ageing than simple DNA repair deficiency. This idea was further supported by the phenotypic similarity between the *ERCC1* mutants and that reported for mice overexpressing the cyclin-dependent kinase inhibitor p21 in hepatocytes. As one of the principle effectors of p53, overexpression of p21 leads to cell cycle arrest in G2 (Wu *et al.*, 1996).

It is worth noting that our original knockout and the two Dutch mutants could potentially give rise to transcript with the ability to encode more than the amino-terminal half of ERCC1. It remains possible that whilst these targeting strategies have resulted in the inactivation of the NER capacity of *ERCC1* a further as yet undetermined function remains intact. Analysis of different, laboratory generated, ERCC1-deficient rodent cell lines has shown some phenotypic heterogeneity supportive of this idea (Busch *et al.*, 1997). Whilst most *ERCC1* mutants are sensitive to both UV and mitomycin C (cross-linking agent), some were only sensitive to one of these mutagenic agents. Our second knockout line carries a deletion of coding sequences from the N-terminal region, which may reveal additional phenotypic consequences resulting from inactivation of any secondary *ERCC1* functions. Initial analysis of these nulls has confirmed the essential phenotypic features of the other *ERCC1* null strains. More detailed analysis, of what may be the only true *ERCC1* null, may be required to reveal any subtle phenotypic variations.

Embryonic fibroblasts derived from our original null and the two Dutch mutant *ERCC1* lines were shown to be hypersensitive to UV, with virtually no incision following irradiation, confirming that the null animals were indeed NER deficient. More detailed analysis carried out on fibroblasts from our original null line has

shown that the cells were only mildly sensitive to the cross-linking agent mitomycin C, suggesting that ERCC1 is not essential for the recombination mediated repair of interstrand cross-links in mouse cells or that this function is not effected by the targeting (Melton *et al.*, 1997). The fibroblasts were not deficient in the homologous recombination pathway. In mouse cells the G1 arrest induced by UV irradiation is less extensive than in human cells, consequently replication proceeds along an unrepaired template. To prevent the accumulation of strand breaks at these unrepaired lesions, a mechanism of illegitimate recombination is employed which leads to increased incidence of chromatid exchanges. The frequency of these exchanges was reduced in *ERCC1* null fibroblasts relative to wild type cells, suggesting that ERCC1 may normally play a role in the promotion of these recombination events.

It has been shown that *ERCC1* -deficiency is indeed compatible with development to term in the mouse, however the gestation period in man is significantly longer and as such the consequences of the analogous deficiency may be developmentally lethal in humans. Whilst the *ERCC1*-deficient animals have provided a useful system in which to investigate ERCC1 function *in vivo,* the short lifespan of the animals has precluded a number of further studies such as exploring the possible role of *ERCC1* in carcinogenesis.

1.25 Project Aims

In the limited number of cells and tissues studied to date it has been reported that *ERCC1* is expressed constitutively, but at low abundance. It would seem logical for the expression of *ERCC1* to be altered where there was a greater biological demand for the protein. We speculated that biological demand on ERCC1 would perhaps be greater in the skin, as it receives the greatest exposure to UV. We originally produced *ERCC1*-deficient mice, by targeted inactivation of the *ERCC1* gene in mouse embryonic stem cells, with the intention of studying the mechanisms involved in the UV-induction of skin cancer. Such studies were precluded by the severe phenotype associated with *ERCC1* deficiency in the mouse. The mice are born severely runted and die, before weaning, with liver failure. The principal aims of the work presented in this thesis, were:

1.To study the pattern of *ERCC1* expression in mouse tissues to determine if a tissue specific pattern of expression was evident.

2. To extend the lifespan of the *ERCC1* nulls so as to be able to study the consequences of *ERCC1* deficiency in other tissues, such as the skin, in adult mice.

Chapter 2
Methods and Materials

Materials

2.1 Laboratory reagents and suppliers:

BDH Ltd.: dimethyldichlorosilane, hydrochloric acid, Sarkosyl NL30, xylene, xylene cyanol.

Boehringer Mannheim UK Ltd.: deoxyribonucleotides, proteinase K, 3'/5' RACE Kit.

The Boots Company plc: paraffin oil.

Difco Laboratories: agar, bacto-tryptone, bacto-yeast extract.

Fisher Scientific UK Ltd./Fisons Scientific Equipment: 3M blotting paper, boric acid, butanol, calcium chloride, citric acid, disodium hydrogen orthophosphate, disodium hydrogen orthophosphate, EDTA, filter paper, glacial acetic acid, hydroxyquinolin, isoamyl alcohol, isopropanol, sodium acetate, sodium chloride, sodium citrate, sodium hydroxide, sodium dihydrogen orthophosphate, potassium acetate, trichloroacetic acid.

Gibco BRL Life Technologies Ltd.: 100bp DNA marker ladder, agarose, formamide, guanidine hydrochloride, *Hind* III digested λ DNA marker ladder, phenol, Tris, low melting point agarose, NACS columns, hybridisation bags, urea.

ICN Biomedicals Ltd.: Dulbecco's PBS tablets.

Millipore (UK) Ltd.: disposable sterile filters, Immobilon P.

National Diagnostics: ecoscint A.

New England Nuclear Life Science Products: Genescreen *Plus*.

Promega (UK) Ltd.: oligo dT, random hexamers, RNasin.

Qiagen Ltd.: tip 500 columns.

Sigma-Aldrich Company Ltd.: bovine serum albumin, bromophenol blue, dextran sulphate, DTT, ethidium bromide, ficoll, formaldehyde, glycerol, herring sperm DNA, IPTG, mineral oil, MOPS, new methylene blue, NP-40, parafilm M, PMSF, sodium bicarbonate, thymidine, Triton X-100, Tween 20, X-Gal.

2.2 DNA /RNA modifying enzymes

Boehringer Mannheim UK Ltd.: Klenow polymerase, M-MuLV reverse transcriptase, restriction endonucleases.

Gibco BRL Life Technologies Ltd.: bacterial alkaline phosphatase, restriction endonucleases, T4 DNA kinase, *Taq* DNA polymerase.

New England Biolabs (UK) Ltd.: restriction endonucleases, T4 DNA ligase.

2.3 Radioactive reagents

Amersham International plc.: Redivue [α-^{32}P]-dCTP (~3000 Ci/mmole, 10mCi/ml).

2.4 Antibiotics

Sigma-Aldrich Company Ltd.: ampicillin, penicillin G, streptomycin.

2.5 Mammalian cell culture reagents

Becton Dickinson Labware: plasticware.

Difco Laboratories: trypsin.

Gibco BRL Life Technologies Ltd.: foetal calf serum, Glasgow's Modified Eagle Medium (BHK21), horse serum, L-glutamine, new-born calf serum, non-essential amino acids, sodium bicarbonate, sodium pyruvate.

Sigma-Aldrich Company Ltd.: b-mercaptoethanol.

2.6 Bacterial culture reagents

Difco Laboratories: agar, bacto-tryptone, bacto-yeast extract

2.7 Oligonucleotides

All oligonucleotides were synthesised by the Oswel DNA service. The sequences of all primers used are given in Table 2.1.

Table 2.1 Oligonucleotides

CODE	SEQUENCE (5'→3')	DESCRIPTION
033M	CCAGTGTTGAAGTTTG TGCG	*ERCC1* exon 4 (429-448)
035M	CGAAGGGCGAAGTTC TTCCC	*ERCC1* exon 5 (598-579)
159M	TAGCCAGCTCCTTGA GAGCC	*ERCC1* exon 6 (657-638)
529N	GGTGCAGTCAGCCAA GATGC	*ERCC1* exon 6 (683-664)
N1138	CTTCCTCCTGGTGGGT GGTCC	*ERCC1* exon 1 (161-142)
Exon 1	TGAGGGCTGTGGCCG ACTTTCCTCGTCCTT	*ERCC1* exon 1 (140-110)
H4395	CATGCCATGGACCCT GGGAAGGACGAG	*ERCC1* cDNA 5'end (99-119)
H4396	GAAGATCTTCTCGAG TCATCGAGGCACTTTG AGG	*ERCC1* cDNA 3'end (996-977)
Oligo dT anchor	GACCACGCGTATCGA TGTCGACTTTTTTTTT TTTTTTTV	5'/ 3' RACE (Boehringer Mannheim)
PCR anchor	GACCACGCGTATCGA TGTCGAC	5'/ 3' RACE (Boehringer Mannheim)
M4956	GGTTCGAAATGACCG ACCAAGCG	Neomycin cassette

Media

2.8 Bacterial culture media

Luria Broth (LB): 1% (w/v) bacto-tryptone, 0.5% (w/v) bacto-yeast extract, and 0.5% (w/v) NaCl, adjusted to pH 7.2 with NaOH.

Luria Agar: 1.5% agar, 1% (w/v) bacto-tryptone, 0.5% (w/v) bacto-yeast extract, and 0.5% (w/v) NaCl, adjusted to pH 7.2 with NaOH.

Terrific Broth (TB): 1.2% (w/v) bacto-tryptone, 2.4% (w/v) bacto-yeast extract, 0.4% (v/v) glycerol, 170mM potassium dihydrogen orthophosphate, and 72mM dipotassium hydrogen orthophosphate.

When required, ampicillin was added immediately prior to use, to a final concentration of 100μg/ml.

2.9 Mammalian tissue culture media

Glasgow's Modified Eagle's Medium (GMEM) (McPherson, 1962, with modifications by W. House, Medical Research Council Institute of Virology, University of Glasgow, 1964) was used as the base culture medium for mouse and human fibroblast cell lines.

Freezing medium for long term storage of cells consisted of the appropriate supplemented culture medium with additional DMSO and FCS to final concentrations of 10% (v/v) and 20% (v/v) respectively.

Fibroblast culture medium: 10% (v/v) FCS, 1X non-essential amino acids, 1mM sodium pyruvate, 25U/ml penicillin, 25μg/ml streptomycin, in GMEM.

ψCRE culture medium: 10%(v/v) NCS, 1X non-essential amino acids, 1mM sodium pyruvate, 25U/ml penicillin, 25μg/ml streptomycin, in GMEM.

2.10 Solutions and buffers

bacterial alkaline phosphatase buffer: 530mM Tris-HCl pH 7.5, 95mM $MgCl_2$, 50mM DTT

Carnoy fixative: 3 parts methanol, 1 part glacial acetic acid

CA: 24 parts chloroform, 1 part isoamyl alcohol

Citrate buffer: 8.55% sucrose, 5% DMSO, 40mM tri-sodium citrate, pH 7.6

denaturation buffer: 0.5M NaOH, 1.5M NaCl

dialysis buffer: 10mM Tris-HCl (pH7.5) 0.1mM EDTA (pH7.5)

electroporation buffer: 21mM HEPES, 5mM D-glucose, 8mM Na_2HPO_4, 5mM KCl, 140mM NaCl

FACS stock solution: 3.4M tri-sodium citrate, 0.5M Tris –HCl, 1% NP40, 0.52mg/ml spermine tetrahydrochloride, pH 7.6

FACS Solution A: 30µg/ml trypsin in stock solution

FACS Solution B: 0.5mg/ml trypsin inhibitor, 0.1mg/ml ribonuclease inhibitor in stock solution

FACS Solution C: 0.42mg/ml propidium iodide, 1mg/ml spermine tetrahydrochloride in stock solution

formalin: 4% (v/v) formaldehyde in phosphate buffered saline

formamide sample buffer: 2.3X MOPS, 50% de-ionised formamide, 11% formaldehyde

injection buffer: 10mM Tris pH7.5, 0.1mM EDTA pH 7.5

10X kinase buffer: 50mM Tris-HCl pH 7.5, 100mM $MgCl_2$, 100mM DTT, 10mM ATP, 0.5µg/ml BSA

10X MOPS: 200mM 3-(*N*-morpholino) propane-sulphonic acid (MOPS), 50mM sodium acetate, 10mM EDTA, pH 7.0

neutralising buffer: 3M NaCl, 0.5M Tris-HCl, pH 7.0

northern blot stripping solution: 0.01% (w/v) SDS, 0.01X SSC

oligo labelling buffer: 50µl solution A, 125µl solution B, 75µl solution C

Solution A: 1.25M Tris-HCl pH 8.0, 0.125M $MgCl_2$, 25mM ß-mercaptoethanol, 0.5mM each of dGTP, dATP and dTTP.

Solution B: 2M HEPES buffer adjusted to pH 6.6 with NaOH.

Solution C: Random hexadeoxyribonucleotides OD_{260nm}=90 units/ml in TE

PI: 50mM Tris-HCl pH 8.0, 10mM EDTA, 100µg/ml RNase A

PII: 20mM NaOH, 1% (w/v) SDS

PIII: 2.55M potassium acetate pH 4.8

PCA: 25 parts redistilled phenol, 24 parts chloroform, 1 part isoamyl alcohol

10X PCR: 50mM KCl, 1.5mM $MgCl_2$, 10mM Tris-HCl pH 8.3, 0.01% (w/v) gelatin, 0.45% (v/v) Triton X-100, 0.45% (v/v) Tween 20

QBT: 750mM NaCl pH 7.0, 50mM MOPS, 15% (v/v) ethanol, 0.15% (v/v) Triton X-100

QC: 1M NaCl pH 7.0, 50mM MOPS, 15% (v/v) ethanol

QF: 1.25M NaCl pH 8.2, 50mM MOPS, 15% (v/v) ethanol

10X RT buffer: 500mM Tris-HCl pH 8.3, 60mM $MgCl_2$, 400mM KCl, 10mM DTT

5X sample buffer: 20% (v/v) glycerol, 100mM EDTA, 0.1% (w/v) bromophenol blue

20X SSC: 3M NaCl, 0.3M tri-sodium citrate, pH 7.0

Southern blot stripping solution: 1% (w/v) SDS, 0. 1X SSC

5X T4 DNA Ligase: 250mM Tris-HCl pH7.6, 50mM $MgCl_2$, 5mM ATP, 5mM DTT, 25% (w/v) polyethylene glycol-8000

10X TBE: 0.9M Tris-HCl, 0.9M boric acid, 20mM EDTA

tail buffer: 400mM NaCl, 10mM Tris-HCl, 3mM EDTA, 1% (w/v) SDS

TE buffer: 10mM Tris-HCl, 1mM EDTA, pH 8.0

transformation buffer: 50mM $CaCl_2$,10mM Tris-HCl

2.11 Bacterial strains

The *E. coli* strain DH5α was used for the propagation of all plasmids. The genotype of DH5α is: *supE44, • lacU169 (ϕ 80lacZ• M15), hsdR17, RecA1, endA1, gyrA96, thi-1*, and was first described by Hanahan (1983).

2.12 Plasmids

The plasmids used are presented in Table 2.2

Table 2.2 Plasmids

PLASMID	DESCRIPTION	REFERENCE
pBluescript II	Cloning vector	Thummel *et al.*, 1988
pTZME	Mouse ERCC1 cDNA	van Duin *et al.* 1988
pMFG	Retroviral vector	Dranoff *et al.*, 1993
pTTR EX1V3	Transgene vector	Yan *et al.* 1990
p91	Mouse actin cDNA	Minty *et al.*, 1991
pGAPDH	Mouse GAPDH cDNA	Fort *et al.*, 1985
pHPT5	Mouse HPRT cDNA	Konecki *et al.*, 1982

2.13 Mammalian cell culture lines

The cell lines used are presented in Table 2.3.

Table 2.3 Mammalian cell lines

CELL LINE	DESCRIPTION	REFERENCE
PF 20	Immortalised wild type mouse embryonic fibroblast cell line.	Melton *et al.*, 1998
PF24	Immortalised *ERCC1*-deficient mouse embryonic fibroblast cell line.	Melton *et al*, 1998
CHO9	Wild-type Chinese hamster ovary cell line.	Wood and Burki, 1982
CHO43-3B	*ERCC1*-deficient Chinese hamster ovary cell line.	Wood and Burki, 1982
Keratinocyte (WT)	Mouse wild type keratinocytes	Melton, unpublished data
Keratinocyte (WT)	*ERCC1*-deficient mouse keratinocytes	Melton, unpublished data
ψCRE	Retroviral packaging cell line.	Dranos and Mulligan, 1988

Methods

Bacterial culture

2.14 Growth of *E. Coli*

Cells were grown at 37°C either in suspension in LB or TB with shaking, or on the surface of Luria agar plates. Cells grown on Luria agar plates supplemented with ampicillin were incubated at 30°C to prevent growth of satellite colonies.

2.15 Storage of *E. Coli*

For long term storage, 900µl of a fresh culture, grown overnight in LB supplemented with antibiotic where necessary, was mixed with 100µl DMSO, and frozen at -70°C. Cells were recovered by scraping the surface of the frozen culture with a sterile inoculating loop and streaking onto a Luria agar plate.

For short term storage (4-6 weeks) bacteria were streaked onto Luria agar plates, and stored at 4°C.

2.16 Transformation of *E. coli*

Bacteria were transformed with plasmid DNA by the method of Mandel and Higa (1970) with the modifications by Dagert and Ehrlich (1974). An inoculating loop was used to scrape bacterial growth from a fresh Luria agar plate of *E. Coli* and was added to 50ml of LB supplemented with $MgCl_2$ to a final concentration of 20mM. Cells were grown with vigorous shaking at 37°C until the OD_{600nm} of the suspension equalled 0.2, and the cells were in log phase growth. The cells were chilled on ice for 5 minutes, and then spun down at 1,500rpm for 15 minutes at 4°C. The pellet was gently resuspended in 20ml ice-cold transformation buffer, incubated for 30 minutes on ice, subsequently re-pelleted at 1,500rpm for 15 minutes at 4°C, and finally resuspended in 2ml ice-cold transformation buffer. The suspension of competent cells was then incubated on ice for a minimum of 2 hours prior to use.

Approximately 10ng of plasmid DNA in a volume of 5-10µl was added to 100µl of competent cells, which were then incubated on ice for 30 minutes. Following heat shock at 37°C for 5 minutes, 400µl of warm LB was added, and the mixture

incubated at 37°C for 1 hour to allow expression of antibiotic resistance. Thereafter, the cells were spread onto Luria agar plates with ampicillin selection, and incubated overnight at 30°C.

Where appropriate, 30µl of 2% (w/v in dimethylformamide) X-Gal and 20µl of 100mM IPTG were also spread over the surface of the plate to enable blue-white colour selection of colonies containing recombinant plasmids.

Mammalian cell culture

2.17 Growth of mammalian fibroblasts

Mammalian fibroblasts were maintained in supplemented GMEM at 37°C in a 5% CO_2 atmosphere in a humidified tissue culture incubator (Forma Scientific).

To passage cells, the medium was aspirated, the bottom of the flask or dish rinsed with 1-2ml of trypsin (0.05% w/v), then a further 1-2ml of trypsin added and incubated at 37°C until the cells had become detached. Thereupon 5 volumes of medium were added, the cell suspension transferred to a sterile 50ml tube, and cells pelleted by 5 minutes centrifugation at 1,300rpm. The supernatant was discarded, and the cells resuspended in an appropriate volume of medium, and an aliquot transferred into a tissue culture flask.

For long term storage, cells were kept in liquid nitrogen. Cultures to be frozen were trypsinised, spun down, the supernatant removed, and cells resuspended in ice cold freezing medium. 1ml aliquots were placed at -20°C for 2 hours, then transferred to -70°C overnight, before removal into liquid nitrogen. Frozen aliquots were thawed rapidly at 37°C and diluted in 9ml of medium, before being pelleted by centrifugation at 1,300rpm for 5 minutes. The supernatant was discarded, the cells resuspended in an appropriate volume of medium, and transferred into a flask.

2.18 Transfection of ψCRE cells

Cloned DNA sequences were introduced into the retrovirus producer cell line using a calcium phosphate precipitation protocol with a glycerol shock, as follows. A confluent 90mm dish of ψCRE cells were split 1:10 the day before the transfection. Added to an Eppendorf tube containing 0.5ml HBS was 10µg of retroviral vector

DNA and 1.0μg of a selectable marker construct and then 31μl of 2M $CaCl_2$. The solution was left to precipitate at room temperature for 40 minutes. The medium was aspirated from the ψCRE dish and the calcium phosphate precipitate added to the cells and left at room temperature for 20 minutes, with frequent rocking of the dish. Following this, 5ml of medium was added back to the dish and the cells incubated at 37°C for 4 hours. The medium was once again aspirated from the dish and this time replaced with 2.5ml of 15% glycerol in HBS (v/v) and incubated at 37°C for 3.5 minutes. The glycerol was then aspirated and the dishes rinsed with 5ml of medium and then cultured overnight in non-selective medium. The following day the medium was aspirated and replaced with medium containing appropriate selection.

2.19 Retroviral infection of mouse fibroblasts

In an attempt to optimise the viral titre from the retroviral producer cell line ψCRE, the harvesting of the supernatant was timed to coincide with the cultures reaching confluence. To ensure that they were actively dividing, the target cells were spit 1:20 on the day prior to the infection. The viral supernatant was removed from the cells using a syringe and filtered through a 0.45 micron filter. Polybrene was added to the supernatant to a final concentration of 8μg/ml. The medium was aspirated from the target cells and replaced with the viral supernatant. The infection was allowed to proceed for 6 hours at 37°C. The viral supernatant was then replaced with normal medium.

2.20 Electroporation

In order to achieve transient expression of cloned DNA sequences in cultured mammalian cells, a linearised ERCC1 minigene vector was introduced by co-electroporation with a selectable HPRT minigene construct under the control of the mouse phosphoglycerate kinase (*PGK-1*) gene promoter. Cells to be electroporated were trypsinised, pelleted by centrifugation, and $5.0X10^6$ cells were resuspended in 0.8ml of electroporation buffer. The suspension was mixed with 120μg of plasmid DNA in 100μl sterile distilled water, and transferred into an electroporation cuvette (0.4cm electrode, Bio-Rad). The cuvette was slotted into a Gene Pulser (Bio-Rad),

and the cells pulsed at 300V, 500μFd capacitance. After 10 minutes incubation at room temperature, the cells were added to 20mls supplemented GMEM and transferred to a flask. After culturing overnight, the medium was changed to selective medium containing HAT.

2.21 UV survival assays

To measure UV survival of cultured cells, 10^5 cells were plated per 30mm dish. The following day the medium was aspirated from the dishes and replaced with 0.2 ml of sterile PBS and the dishes were UV irradiated (254nm) at a fluence of 0.15 J $m^{-2}s^{-1}$. The medium was then replaced and the dishes were incubated for a further five days when the cells were fixed in Carnoy fixative and stained with Crystal Violet. The degree of cell growth was then determined by extracting the dye from the stained cells using 70% ethanol and measuring the optical density at 575nm. Each UV dosage was done in duplicate and survival was determined relative to unirradiated control dishes as the ratio of the mean OD / mean OD of cells in unirradiated dishes.

Nucleic acid isolation

2.22 Small-scale preparation of plasmid DNA

A modification of the method described by Ish-Horowicz and Burke (1981) was used to prepare amounts of plasmid DNA up to 10μg. Bacteria were grown overnight at 37°C in 5ml LB or TB supplemented with ampicillin to a final concentration of 50μg/ml. 1.5ml of the culture was transferred to a microfuge tube, and the cells pelleted by centrifugation for 5 minutes. The pellet was resuspended in 100μl of solution P1, and the cells lysed by addition of 200μl of solution P2 with 5 minutes incubation on ice. 150μl of cold solution P3 was added, with gentle mixing, and the tube incubated on ice for a further 15 minutes. Precipitated chromosomal DNA, SDS and protein were sedimented by centrifugation for 10 minutes in a microcentrifuge, and the supernatant, containing the plasmid DNA, removed to a fresh tube. The supernatant was centrifuged again for a further 10 minutes to remove residual debris, and then plasmid DNA was precipitated by addition of 1ml cold ethanol. Following 10 minutes centrifugation, the nucleic acid pellet was washed twice with 70% (v/v)

ethanol, vacuum dried, resuspended in 50µl sterile distilled water and stored at -20°C.

2.23 Large scale preparation of plasmid DNA (Quiagen prep.)

For preparation of larger amounts of plasmid DNA, up to 500µg, a column purification method was employed. Bacteria were grown for 24 hours at 37°C in 50ml of TB supplemented with ampicillin to a final concentration of 100µg/ml. Cells were pelleted by spinning at 4,000rpm for 20 minutes at 4°C. The pellet was resuspended in 10ml of solution P1 supplemented with RNase to a final concentration of 0.5µg/ml, and the cells lysed by addition of 10ml of solution P2 with 5 minutes incubation at room temperature. 10ml of cold solution P3 was added, with gentle mixing, and the tube incubated on ice for a further 20 minutes. Precipitated chromosomal DNA, SDS and protein were sedimented by centrifugation at 16,000rpm for 10 minutes at 4°C. The supernatant was removed promptly, filtered and applied to a Qiagen-tip 500 column, previously equilibrated with 10ml of QBT buffer. After washing twice with 30ml QC buffer, the plasmid DNA was eluted from the column with 15ml QF buffer. 12ml of isopropanol was added to precipitate the DNA, which was then spun down at 10,000rpm for 10 minutes at 4°C. The DNA pellet was resuspended in 400µl of sterile distilled water, prior to transfer into an microfuge tube, whereupon the DNA was reprecipitated with 40µl of P3 and 1ml of cold absolute ethanol. Following 10 minutes centrifugation, the nucleic acid pellet was washed twice with 70% (v/v) ethanol, vacuum dried, resuspended in 200µl sterile distilled water and stored at -20°C.

2.24 Preparation of mammalian genomic DNA

DNA was prepared from mouse tail samples, tissue samples and fibroblasts by the same method. Cultured cells were harvested by scraping in ice cold PBS, and pelleted by centrifugation at 1,300rpm for 5 minutes. Cell pellets and tissue samples were digested overnight at 37°C in 750µl tail buffer supplemented with proteinase K to a final concentration of 280µg/ml. The supernatant was extracted twice with 750µl PCA and vigorous shaking, and subsequently once with 750µl CA to remove traces

of phenol. The DNA was precipitated for 10 minutes at room temperature by addition of 750µl of isopropanol. Following 10 minutes centrifugation, the nucleic acid pellet was washed twice with 70% (v/v) ethanol, vacuum dried, resuspended in 200µl sterile distilled water and stored either at 4°C in the short term or at -20°C for longer periods.

2.25 Preparation of mammalian RNA (RNAzol™ B method)

A modification of the method described by Chomczynski and Sacchi (1987) was used for the isolation of total RNA from all cultured cells and tissues. The method used the commercially available reagent RNAzol™ B (Biogenesis Ltd.) and followed the protocol supplied by the manufacturers. Briefly, approximately 300mg of tissue was homogenised with 6ml RNAzol™ B using an Ultra-Turrax T25 homogeniser (IKA-Laboratechnik). Cultured cells were lysed directly in the culture dish by the addition of RNAzol™ B (1ml per 3.5ml petri dish) and the RNA was solubilised by pipetting the lysate a few times. The RNA was extracted by the addition of 0.1ml chloroform per 1ml of homogenate, followed by vigorous shaking of the mixture for 15 seconds prior to placing on ice for 5 minutes. The suspension was then centrifuged at 12000g for 15 minutes at 4°C. The aqueous phase was transferred to a fresh tube and the RNA precipitated by the addition of an equal volume of isopropanol. The RNA was left to precipitate at 4°C for at least 15 minutes and then pelleted by centrifugation at 12000g. The RNA pellet was then washed in 8mls 70% ethanol by vortexing followed by centrifugation at 7500g for 8 minutes at 4°C. The pellet was allowed to air dry before being resuspended in 200µl of sterile distilled water for storage at -70°C.

2.26 Estimation of DNA concentrations

DNA samples were diluted in 1ml of distilled water and the absorbency at 260 and 280nm measured in a spectrophotometer (Perkin-Elmer, Lambda 15, UV/VIS Spectrophotometer). Double stranded DNA of concentration 50µg/ml has an OD_{260nm} = 1.0. The ratio OD_{260nm}/OD_{280nm} gives an estimate of nucleic acid

purity. A value of around 1.8 indicates that the preparation is not significantly contaminated with protein or phenol.

2.27 Estimation of RNA concentrations

RNA samples were diluted in 1ml of distilled water and the absorbency at 260 and 280nm was measured in a spectrophotometer (Perkin-Elmer, Lambda 15, UV/VIS spectrophotometer). An OD_{260nm} = 1.0 represents an RNA concentration of 40 µg/ml. The ratio OD_{260nm}/OD_{280nm} gives an indication of nucleic acid purity. A value of OD_{260nm}/OD_{280nm} = 2.0 indicates a pure preparation of RNA.

DNA manipulation

2.28 Restriction of DNA with endonucleases

DNA was digested with 10 units of the desired endonuclease per µg of DNA using buffer and temperature conditions recommended by the manufacturer. For double digests involving enzymes with different recommended buffers, reactions were either carried out using an intermediate buffer, or DNA was digested in two sequential digests separated by a DNA purification procedure. Digestion reactions were terminated by heating at 65°C for 10 minutes, phenol extraction, or by the addition of 5X sample buffer for agarose gel electrophoresis.

2.29 Dephosphorylation

Bacterial alkaline phosphatase (BAP) is a phosphomonoesterase that hydrolyses 3' and 5' phosphates from DNA and RNA. It is suitable for removing 5' phosphates prior to end labelling and for dephosphorylating linearised plasmid vectors prior to insert ligation. Approximately 1ng of DNA, in 90µl of sterile distilled water, was mixed with 10µl of 10X BAP buffer and 66 units of BAP. The reaction mix was incubated for one hour at 65°C, following which the DNA was purified by phenol extraction to remove all traces of the enzyme.

2.30 Phosphorylation and ligation of PCR products

In the process of cloning PCR products it was necessary to phosphorylate their 5' termini by means of a kinase reaction. Reactions were performed on gel purified PCR products in 10µl volumes containing 1µl 10X kinase buffer, and 1µl of T4 DNA kinase. Following incubation at 37°C for $1^1/_2$ hours, a further 1µl of 10X kinase buffer was added, along with linearised plasmid vector DNA and 2 units of T4 DNA ligase. The total volume was made up to 20µl with sterile distilled water, and the ligation reaction incubated overnight at 15°C, after which it was used to transform competent *E. coli.*

2.31 Blunt-ending of DNA fragments

Overhanging ends, generated by restriction endonucleases were converted to blunt ends by 'filling-in' using the Klenow fragment of *E. coli* DNA polymerase I. Approximately 1µg of DNA , in 39µl sterile distilled water, was mixed with 5µl 10X nick translation buffer, 5µl of 5mM dNTP's and 6 units of Klenow. The reaction mix was incubated at room temperature for 1 hour and the reaction stopped by phenol extraction.

2.32 Ligation

Linearised vector and insert DNA fragments were gel purified as described (in section 2.34) prior to ligation. Approximately 200ng of insert DNA was ligated in a reaction with vector to insert DNA concentrations at a 3:1 molar ratio. Ligations were carried out in a 20µl reaction volume containing 4µl of 5X T4 ligase buffer and 2 units of T4 DNA ligase and incubated overnight at 15°C. Products of the ligation reaction were then used to transform *E. coli.*

Electrophoresis of nucleic acids

2.33 Electrophoresis of DNA in agarose gels

DNA fragments were separated on 0.8-2% (w/v) electrophoresis grade agarose gels containing 0.5µg/ml ethidium bromide and 1X TBE. DNA samples were mixed with $^1/_5$ volume of 5X sample buffer prior to loading. Electrophoresis was carried out

horizontally at 30-100V in a 1X TBE buffer system. *Hind* III digested bacteriophage λ DNA marker ladder and a 100bp DNA ladder were used as size markers. After electrophoresis, DNA was visualised by UV illumination.

2.34 Recovery of DNA from agarose gels

If DNA was to be recovered from agarose gels, low melting point agarose was used, and the desired section of gel excised following electrophoresis. An equal volume of 1X TBE, 0.2M NaCl was added, and the gel melted at 65°C. 1µl of glycogen was added to act as a carrier, and increase the proportion of DNA ultimately recovered. Next, an equal volume of redistilled phenol saturated with 1X TBE , 0.1M NaCl was added, the mixture vortexed until an emulsion was formed, then spun down in a microfuge for 5 minutes. The aqueous phase was removed into a fresh tube, and 100µl of butanol added to remove traces of phenol. The mixture was vortexed, then spun down for 1 minute and subsequently the butanol layer discarded. 1/10th volume 3M sodium acetate pH 5.0 and $2^1/_2$ volumes of cold absolute ethanol were added, and the DNA precipitated at -70°C for 20 minutes. Following 10 minutes centrifugation, the nucleic acid pellet was washed twice with 70% (v/v) ethanol, vacuum dried, and resuspended in 10µl sterile distilled water.

2.35 Purification of DNA for pronuclear injections

Approximately 100µg of plasmid was digested with a suitable restriction enzyme under appropriate buffer conditions. The restriction digest was electrophoresed slowly at 50V overnight on a large agarose gel without the addition of ethidium bromide. Following electrophoresis, a full length slice of the gel was cut and stained with ethidium bromide to enable the position of the linear DNA fragment to be determined and then isolated. The gel slice was then placed into dialysis tubing with a small volume of the electrophoresis buffer for electro-elution at 100V for 1 hour. The buffer was then collected from the tubing and the DNA precipitated with 1/10th volume 3M sodium acetate and $2^1/_2$ volumes absolute ethanol. The pellet was washed in 70% ethanol, dried, resuspended in 100µl of TE + 0.5M NaCl. The DNA was then purified by chromatography on a NACS PREPAC™ Cartridge (GIBCO

BRL). The column was equilibrated with 0.5M NaCl in TE, prior to the loading of the DNA. After washing, the DNA was eluted in 1ml of 2M NACl in TE.
The purified DNA was then dialysed against 2 litres dialysis buffer for 36 hours with one change of buffer. The DNA was recovered using a syringe and transferred to Eppendorf tubes for storage at -20°C.

2.36 Electrophoresis of RNA in agarose gels

RNA samples were electrophoresed on denaturing 1% (w/v) agarose gels containing 0.5µg/ml ethidium bromide, 1X MOPS and 0.66M formaldehyde. 20µg of total RNA in 20µl of sterile distilled water was added to an equal volume of formamide sample buffer and $^1/_4$ volume of 5X sample buffer. Samples were heated for 5 minutes at 65°C and snap chilled on ice immediately prior to loading. Electrophoresis was carried out in a 1X MOPS buffer system at 100V for 3-4 hours.

Transfer of nucleic acids from agarose gels to membranes

2.37 Transfer of DNA from agarose gels to membranes (Southern blot)

This procedure was originally developed by Southern (1975) and modified by Smith and Summers (1980). For Southern analysis of genomic DNA, 10µg of DNA was cut with an appropriate restriction enzyme, and the resulting fragments were separated by agarose gel electrophoresis. The gel was soaked in denaturation buffer for 30 minutes preceding DNA transfer onto Genescreen *Plus* nylon membrane (New England Nuclear) by capillary action using denaturation buffer as the transfer medium (Reed and Mann 1985). A wick made of wet blotting paper was placed on a platform, with both ends of the blotting paper immersed in denaturing buffer in a reservoir underneath the platform. The gel was laid on top of the wick, bordered with cling film. A sheet of nylon membrane, previously equilibrated in denaturing buffer) was placed on top, overlaid by three sheets of moistened blotting paper and a stack of dry paper towels. A weighted glass plate was put on top of the paper towels. Transfer was allowed to proceed for 12-24 hours, after which time the membrane was neutralised for 30 minutes in neutralising buffer, and air-dried.

2.38 Transfer of RNA from agarose gels to membranes (northern blot)

Following electrophoresis the gel was soaked for two 20 minute periods in 10X SSC, with gentle agitation. Transfer was as described for DNA, except that 10X SSC was used as the transfer medium. When transfer was complete, the membrane was rinsed in 2X SSC and baked for 2 hours at 80°C under a vacuum.

Nucleic acid hybridisation

2.39 Labelling DNA by random priming with hexadeoxyribonucleotide primers

Radioactively labelled DNA was obtained using the randomly primed DNA labelling method (Feinberg and Vogelstein 1983). This method is based on the hybridisation of a mixture of hexanucleotides to the DNA, and allows small amounts of DNA to be labelled to high specific activities. Many sequence combinations are represented in the hexanucleotide primer mixture, which leads to binding of primer to template in a statistical manner. The complementary strand is synthesised from the 3'OH termini of the hexanucleotide primers using Klenow polymerase, simultaneously incorporating radiolabelled dCTP into the newly synthesised DNA strand. DNA fragments to be labelled were purified from agarose gels. Approximately 10ng of DNA dissolved in 32µl of sterile distilled water was denatured by boiling for 5 minutes, followed by incubation at 37°C for 10 minutes. 10µl of oligo labelling buffer, 20µg of BSA, 50µCi of α-^{32}P-dCTP and 2 units of Klenow polymerase were added, and the mixture incubated overnight at room temperature. 4.5mg of sonicated herring sperm DNA was added to the probe, which was then denatured prior to hybridisation by boiling for 5min.

2.40 Separation of unincorporated radionucleotides

Unincorporated nucleotides were separated from the labelled DNA by chromatography on a NICK® Column (Pharmacia). The column was equilibrated with 400µl TE before the labelling reaction was added to the column. The labelling reaction was run into the column by the addition of 400µl TE. The labelled DNA was then eluted into a fresh Eppendorf tube by the addition of a further 400µl TE.

2.41 Hybridisation

Membranes onto which DNA had been transferred were blocked by prehybridising in 30ml of 6X SSC, 1% (w/v) SDS, and 10% (w/v) dextran sulphate with 3.5mg of denatured sonicated herring sperm DNA for two hours placed in an oven at 65°C in a hybridisation bag (Gibco BRL). Hybridisation was performed by addition of denatured radiolabelled probe to the prehybridisation mixture, and incubation for a further 12-24 hours at 65°C. Membranes to which RNA was bound were treated identically, except that incubations were performed at 60°C. Following hybridisation, non-specifically bound DNA molecules were removed by washing. The membrane was first immersed in 2X SSC at room temperature for 5 minutes with agitation, and then twice in 2X SSC, 1% (w/v) SDS for 30 minutes at either 65°C or 60°C. Finally the membrane was rinsed in 0.1X SSC for 10 minutes at room temperature, and sealed in a plastic bag to prevent drying out. Radioactive DNA molecules bound to the membrane were then visualised by autoradiography or phosphorimagery.

2.42 Removal of probes and re-use of blots

Radiolabelled DNA probes hybridised to nucleic acids immobilised on nylon were removed by the method described in the manual supplied with Genescreen *Plus* membranes. Southern filters were boiled three times for 10 minutes, with gentle shaking, in 0.1X SSC, 1% (w/v) SDS. Northern filters were washed five times for 3 minutes in hot 0.1X SSC, 0.01% (w/v) SDS. The blots were then autoradiographed to confirm that dehybridisation was complete. Filters could then be dried and hybridisation repeated as above.

2.43 Autoradiography

Autoradiography was used to visualise, on film, radioactive molecules hybridised to the membrane. Autoradiography was performed using Cronex (Du Pont) X-ray film in a cassette with intensifying screens (Cronex Lightning Plus, Du Pont). Cassettes were stored at -70°C during exposure to slow the reversal of activated bromide crystals to their stable form and give an enhanced signal.

2.44 Phosphorimagery

Phosphorimagery was performed using a storage phosphor screen (Molecular Dynamics). After exposure of the screen, signals were visualised using a Molecular Dynamics phosphorimager, and analysed with ImageQuant software (Version 3.3).

2.45 DNA sequencing

DNA sequencing was performed using an ABI PRISM dye terminator cycle sequencing reaction ready kit (Perkin-Elmer Corporation), on an Omnigene thermal cycler (Hybaid Ltd.). Extension products were then separated on an ABI PRISM 377 DNA sequencer (Perkin-Elmer).

Approximately 0.4μg of double stranded template DNA was mixed with 3.2pmoles of the required oligonucleotide primer and 8μl of terminator ready reaction mix in a final volume of 20μl. The reaction mixture was overlaid with a drop of mineral oil before thermal cycling as follows: 96°C for 30 seconds, 50°C for 15 seconds, 60°C for 4 minutes, repeated for 25 cycles.

Extension products were purified by the addition of 2.0μl 3M sodium acetate pH 5.0, 50μl absolute ethanol, and incubation on ice for 10 minutes followed by 15 minutes centrifugation. The pellet was washed with 70% (v/v) ethanol, vacuum dried and stored dry at 4°C before electrophoresis. Electrophoresis was performed by the DNA sequencing facility within the Institute of Cell and Molecular Biology.

2.46 Amplification of DNA by the polymerase chain reaction

All PCRs were carried out in 50μl volumes in 1X PCR buffer supplemented with EDTA, to a final concentration of 0.4mM, and dNTPs, to a final concentration of 0.1mM, and 2.5 units of *Taq* DNA polymerase. Approximately 100ng of DNA was used per reaction, along with 300ng of each primer. Cycle conditions varied for different primers pairs, generally, denaturation was carried out at 94°C for 1 minute, followed by 1 minute at an appropriate annealing temperature, and synthesis at 72°C for 30 seconds to 2 minutes. 35-40 cycles were carried out for all genotyping PCRs. All experiments were carried out using a Techne PHC-2 thermal cycler.

2.47 Rapid screening using PCR

To facilitate the rapid screening of colonies of cultured mammalian cells, a PCR-based strategy that bypassed the need for DNA purification was employed (McMahon and Bradley 1990). When colonies were picked from dishes to be transferred into multiwell plates, half of each colony (circa 1,000 cells) was transferred into a 1.5ml microfuge tube. The cells were pelleted for by centrifugation for 30s, and the medium removed using a pipette. The pellet was resuspended in 40µl 1X PCR buffer to which proteinase K had been added to a final concentration of 0.1mg/ml, and incubated at 65°C for two hours. The proteinase K was inactivated by heating at 90°C for 15 minutes, and 10µl of sample used as a template for PCR.

2.48 RT-PCR

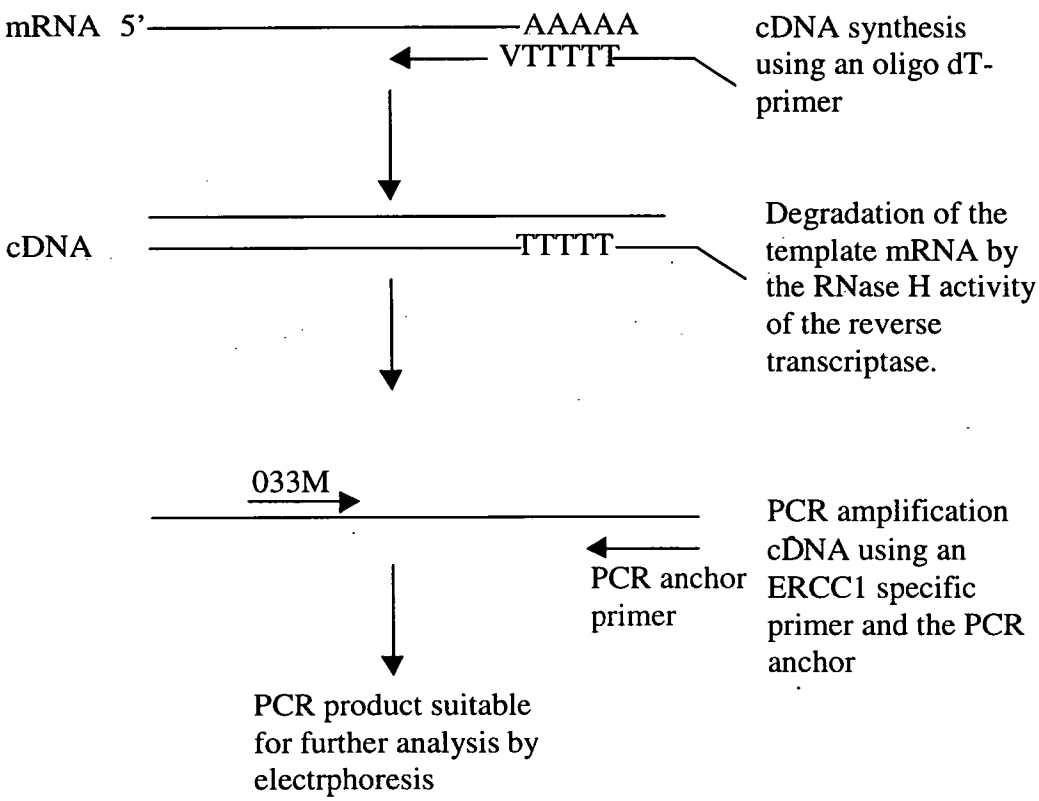
Reverse transcription was carried out on 20µg of total RNA. To denature the RNA, samples were heated at 70°C for 2 minutes and then snap cooled on ice. Reverse transcription was carried out at 43°C for 1 hour in 1X RT buffer with 0.5µl RNasin, dNTPs to a final concentration of 1.25mM, 0.1µg of random hexamers, 25µg of oligo-dT, and 30 units of reverse transcriptase. The reaction was then heated at 75°C for 1 minute snap cooled on ice before a second round of reverse transcription was carried out by the addition of a further 0.5µl RNasin and 30 units of reverse transcriptase. After incubation at 43°C for 1 hour the reaction was diluted in 200µl of TE and stored at 4°C. For subsequent procedures 10µl of the total cDNA pool was used.

2.49 3' RACE

Characterisation of the 3' end of the ERCC1 cDNAs was performed using 5'/ 3' RACE Kit (Boehringer, cat. no. 1734792). All reagents used in the procedure were supplied by the manufacturer, excluding the allele specific primers. A detailed protocol supplied by the manufacturer was followed throughout the procedure. An overview of the principles involved in 3' RACE and an abridged protocol are presented in this section. The method exploits the naturally occurring poly(A)-tail of messenger RNAs as a target site for PCR primers. First strand cDNA synthesis is

started from the poly(A)-tail using an oligo dT-anchor primer. The oligo dT-anchor primer is in actual fact a mixture of dT primers with a single non T at the 3' end, the primer then binds to the inner end of the poly(A)-tail. The same primer codes for three restriction sites at the 5' end to facilitate subsequent cloning of any PCR products. After this synthesis, the cDNA is amplified by PCR using a PCR anchor primer, which is specific for the last twenty-two nucleotides at the 5'end of the oligo dT-anchor primer, and a forward primer specific for ERCC-1 (033M). The resultant PCR product was used for subsequent characterisation by cloning and restriction analysis.

Figure 2.1 An overview of 3'RACE.



2.50 5' RACE

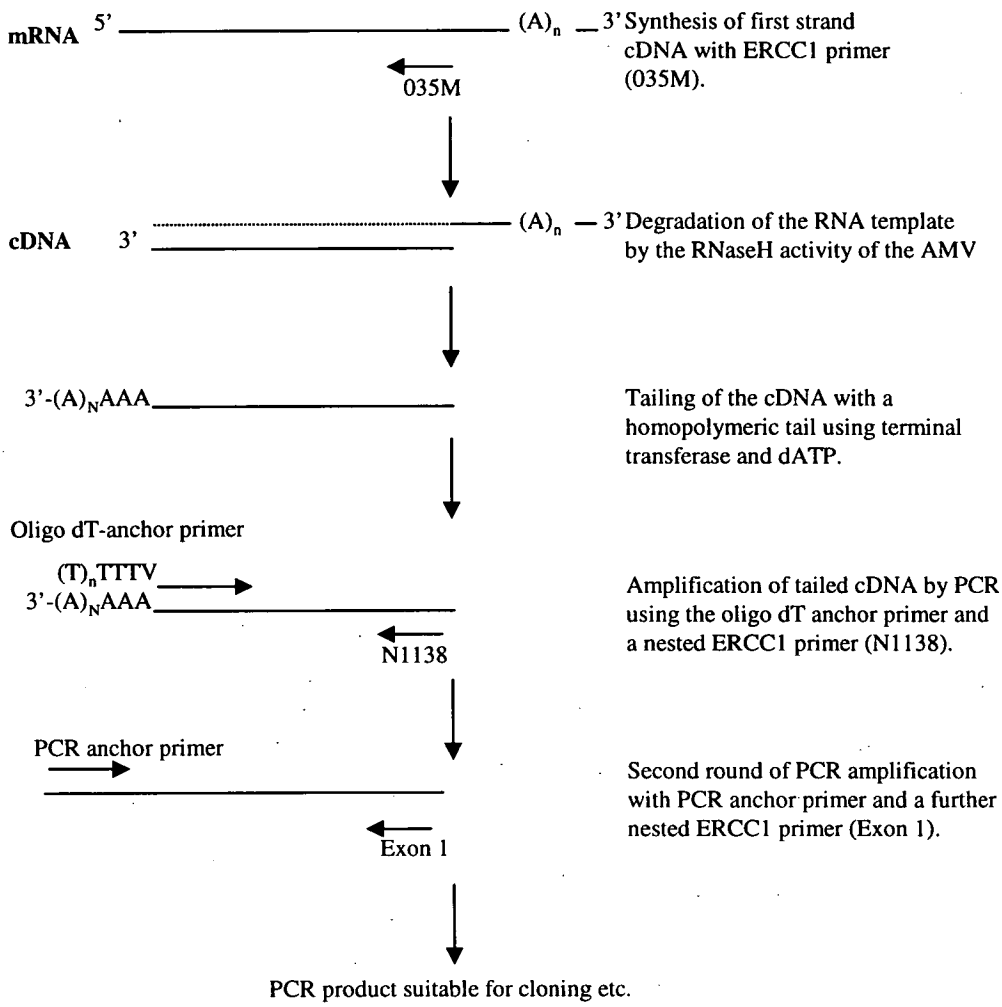
Characterisation of the 5' end of the ERCC1 cDNAs was performed using a 5'/ 3' RACE kit (Boehringer, cat. no.1734792). All reagents used in the procedure were

supplied by the manufacturer, excluding the relevant ERCC1 specific primers. The detailed protocol supplied by the manufacturer was followed throughout the procedure. For the purposes of illustration, an overview of the principles behind the technique along with an abridged version of the protocol are presented in this section. (See Figure 2.2)

The procedure requires the amplification of unknown DNA sequences from a messenger RNA (mRNA) template between known internal sequences and the unknown sequences of the 5' end of the mRNA. This method is referred to as RACE, the rapid amplification of cDNA ends and was first described by Frohmann et al. (1994).

The first strand of cDNA was synthesised from total RNA using a primer specific for ERCC1 (035M), AMV reverse transcriptase and deoxynucleotides. The resultant cDNA was then purified from unincorporated primers and deoxynucleotides using the High Pure PCR Purification Kit component of the RACE Kit. A homopolymeric A-tail was added to the 3' end of the cDNA using terminal transferase. The tailed cDNA was then amplified by PCR using a second ERCC1 specific primer (N1138) and an oligo dT-anchor primer. This cDNA was then amplified with a further round of PCR using a nested ERCC1 (Exon 1) primer in combination with the anchor primer. The cDNA resultant at this stage was then cloned into a suitable vector to enable further characterisation by restriction analysis and sequencing.

Figure 2.2. Overview of 5' RACE analysis.



Animal procedures

2.51 Animal husbandry

Mice were maintained in accordance with established animal care guidelines (Poole 1989), and all procedures were licensed by the Home Office under the animals (Scientific Procedures) Act 1986. The mice were kept under a 12 hour light and 12 hour dark cycle, with standard mouse diet and tap water. Genotyping was by PCR analysis carried out on the biopsied distal 1cm of tail, removed after the pups were weaned into single sex cages at 21 days. Individual cage members were identified by means of coded ear marking.

2.52 Conventional transgenesis by pronuclear injection

The protocol adopted for the generation of transgenic mice by injection of DNA directly into the pronuclei of fertilised mouse eggs is described in detail in Manipulating the Mouse Embryo by Hogan, Costantini and Lacy (1986).

2.53 Superovulation of donor females

In order to increase the yield of fertilised eggs from each of the female donors it was necessary to induce super-ovulation by the administration of gonadotrophins. Up to eight female F1 hybrid CB6 mice aged between 4 and 6 weeks were injected intraperitoneally with 5 units of pregnant mare's serum (PMS) which mimics the endogenous follicle stimulating hormone (FSH). After a period of approximately 42 hours the mice were then injected intraperitoneally with human chorionic gonadotrophin (hCG) which mimics luteinizing hormone (LH). Each female was then placed in a cage with a CB6 stud male and checked for a copulation plug the following morning.

2.54 Recovery of fertilised eggs

Plugged females were humanely culled by cervical dislocation before bilateral dissection of the oviducts. The oviducts were then transferred to a 35mm petri dish containing M2 medium. The eggs, and attached cumulus cells, were released using watchmakers forceps to tear open the oviduct. The eggs were then transferred to a

second petri dish containing M2 with hyaluronidase at about 300μg/ml and incubated at room temperature to remove cumulus cells. After rinsing in a further dish of M2 the fertilised eggs were transferred to a dish containing M16 and cultured at 37°C until required.

2.55 Injection of mouse egg pronuclei

All pronuclear injections were carried out using a Narishige micromanipulator system (Narishige Co Ltd.). Holding pipettes and injection needles were pulled from borosilicate glass capillaries (Clark Electromedical Instruments) using a Sutter Instruments Co. (model P-87) needle puller and a Research Instruments Ltd forge. Embryos were transferred to a drop of M2 on a coverslip, overlayed with mineral oil. Following injection the surviving eggs were transferred to an M16 drop and cultured at 37°C until required.

2.56 Embryo transfers

Eggs that survived injection were either transferred on the day of injection or after overnight culturing into the oviduct of a 0.5 p.c pseudopregnant CB6 female. To compensate for the number of embryos not developing to term and thus avoiding very small litters, between 20 and 30 injected eggs were transferred into each pseudopregnant recipient.

2.57 Tissue histology

All tissue samples were fixed in 4% formalin for at least 48 hours and then stored in 70% ethanol until sectioning. All sectioning and staining of fixed tissues was performed by the Department of Pathology, University of Edinburgh. The tissues were embedded in paraffin wax and sectioned at 5μm onto microscope slides. The mounted sections were stained using haematoxylin and eosin for subsequent histological analysis.

2.58 Fluorescence activated cell scanning

Fluorescence activated cell scanning (FACScan) refers to the application of flow cytometry in the determination of the DNA content of a cell. A laser light (488nm) is used to excite fluorochrome molecules, causing them to emit light of a greater wavelength, which is detected by the FACScan. Propidium iodide is a commonly used DNA dye. The propidium iodide intercalates in the DNA double helix and fluoresces strongly red following exposure to laser light. The extent of the fluorescence varies according to the DNA content and is measured by the FACScan.

In all cases the analysis was carried out on fresh tissue samples. A $0.5cm^3$ tissue biopsy was homogenised using a sterile scalpel and suspended in 100µl of citrate buffer. After adding 450µl of FACS Solution A, the suspension was incubated at room temperature for 10 minutes with gentle mixing. The action of the trypsin was stopped and cellular RNA degraded by the addition of 325µl of FACS Solution B and incubation at room temperature for 10 minutes. The nuclear DNA was then dyed with the fluorescent stain propidium iodide, by the addition of 250µl of FACS Solution C and incubation on ice for 10 minutes. The suspension was filtered through sterile gauze before analysis on a flow cytometer.

The analysis was carried out using a Becton Dickson FACScan, and the resultant data analysed using Lysis II software.

2.59 ATPase staining of Langerhans cells

To identify phosphatase-positive cells, epidermal sheets were first fixed in cacodylate formaldehyde fixative (4% formaldehyde, 0.8M sodium cacodylate and 0.2M sucrose) and then washed three times for ten minutes in 10ml of Trismal buffer (20mM Trismal maleate, 0.2M sucrose, pH 7.3). They were then incubated at 37°C for 30 minutes in Trismal buffer containing 20mM $MgCl_2$. 4mM $Pb(NO_3)$ and 0.33mM ADP. After three washes in Trismal buffer, the samples were immersed in 0.5% ammonium sulfide for 5 minutes and then washed 4 times in 10ml of distilled water before mounting in a 90% glycerol, 10% PBS medium. Stained cells were then counted microscopically.

2.60 Mixed skin lymphocyte reaction

Epidermal cells were prepared from mouse ears by incubation in a 1% trypsin solution. After counting, 0.5 x 10^5 epidermal cells were incubated, in 96 well plates at 37°C, with 10^5 spleen cells from strain C3H female mice in complete RPMI medium containing 10% foetal bovine serum. After 5 days in culture 0.7 μCi [H^3]thymidine was added to the well. [H^3]thymidine incorporation was measured when cells were harvested 24 hours later. The specific proliferative response was calculated as the mean of 5 replicate wells after subtracting the mean of the negative control consisting of EC cultured alone.

2.61 Accumulation of Langerhans cells in the draining lymph nodes

The draining, axillary and auricular lymph nodes were removed and pooled for each of the experimental groups. A single suspension of lymph node cells was prepared by mechanical disaggregation through a 200-mesh stainless steel gauze. The DC were purified by density gradient centrifugation. Briefly, 8ml of the cell suspension was added to a 10ml conical tube and was underlayed with 2ml of 14.5% metrizamide in RPMI-FCS medium. The tubes were then centrifuged for 15 minutes at 600*g*. The DC enriched population at the interface was collected, washed and resuspended in RPMI-FCS medium. The number of DC within this fraction was assessed morphologically by conventional light microscopy. The mean number of DC present within a lymph node was calculated.

Chapter 3
Identification and molecular characterisation of a novel *ERCC1* transcript in mouse skin

The mammalian genome contains an estimated 80,000 (Antequera and Bird, 1993) genes and by various mechanisms, including differential splicing, can encode an even greater number of polypeptides. Fortunately, in order for a cell to survive it is not required to express every single gene, all of the time. In many cases gene expression is determined to a great extent by biological requirement. This control of gene expression can vary from simple, expression or no expression, to various levels of expression depending on the individual cell type or tissue and the biological requirement for expression.

It would not be unreasonable to imagine that *ERCC1* expression could be altered in tissues where it would have a greater biological significance. The NER pathway, in which *ERCC1* is involved, is known to be the principle pathway responsible for the removal of UV-induced DNA lesions and the tissue which receives the greatest exposure to UV is the skin. It might therefore be reasonable to expect that the expression of *ERCC1* may display different properties in the skin in an attempt to deal with a more demanding role in that tissue. It was with this in mind that we decide to take a detailed look at the expression pattern of *ERCC1* in mouse tissues. In the limited number cells and tissues studied to date it has been reported that *ERCC1* is expressed constitutively, but at low abundance.

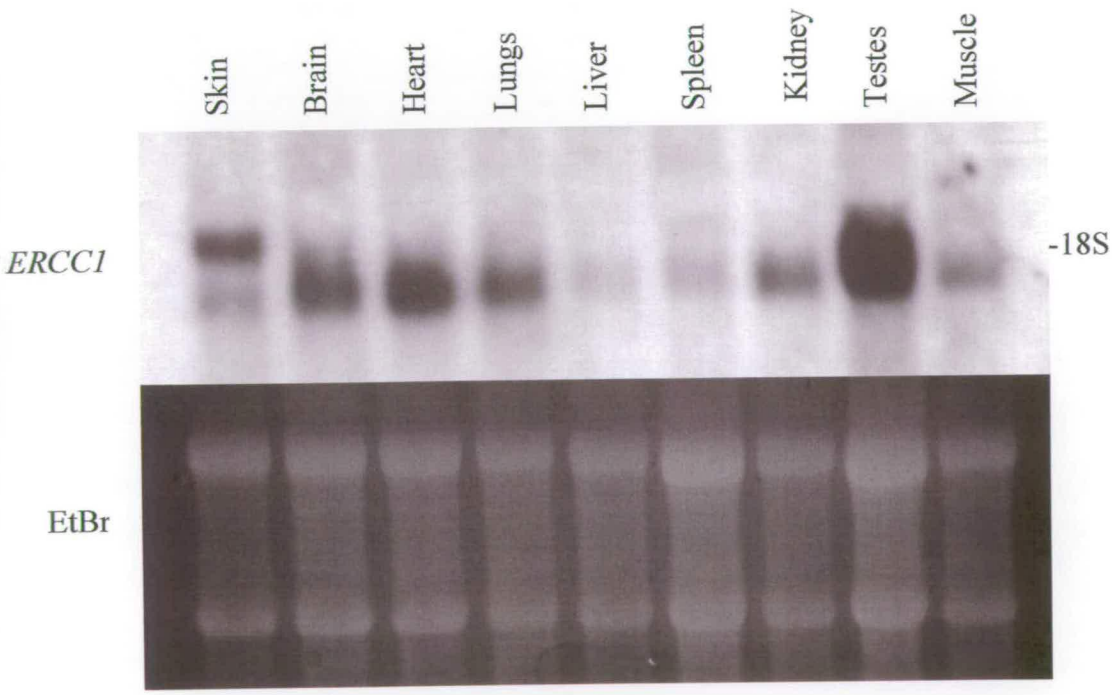
3.1 Identification of a novel *ERCC1* transcript in mouse skin

A panel of tissues from a wild type male mouse was analysed by northern blot analysis to investigate further the endogenous pattern of *ERCC1* expression (Figure 3.1). Total RNA (30ug) from each of the tissues was separated by electrophoresis, transferred onto a nylon membrane and the *ERCC1* transcript detected by hybridisation to an *ERCC1* cDNA probe.

The ethidium bromide staining of the gel is presented in the lower panel and suggests that the amount of RNA loaded in each of the tracks was equivalent. The detection of *ERCC1* mRNA by probing with *ERCC1* cDNA is shown in the top panel. The normal 1.1kb *ERCC1* transcript was evident in all tissues examined. By eye, the level of expression was lowest in the liver and generally low in all other tissues

Figure 3.1 Identification of a novel *ERCC1* in mouse skin

Northern blot analysis of tissue RNA from a wild type male mouse. Upper panel, total RNA was prepared from a selection of mouse tissues, electrophoresed on a 1.4% agarose-formaldehyde gel, transferred onto a nylon membrane and probed with an 800bp BamHI fragment of the *ERCC1* cDNA, corresponding to exons 1 to 8. The position of the 18S rRNA is indicated. Lower panel, ethidium bromide staining to illustrate RNA loads.


Skin
Brain
Heart
Lungs
Liver
Spleen
Kidney
Testes
Muscle
ERCC1
-18S
EtBr

except for the testes, where the level of expression of this transcript was clearly much higher than in any of the other tissues. There was a suggestion of some heterogeneity in the size of the *ERCC1* transcript detected within both the brain and heart, as seen by hybridisation to a higher molecular weight smear up from the normally discrete transcript signal. The pattern of expression seen in the skin was clearly different from any other tissue. In addition to the normal *ERCC1* transcript, a second larger transcript was evident. In terms of abundance the novel *ERCC1* transcript was the major transcript in skin. The size of the transcript extends beyond the upper limits of the heterogeneous smear identified in the brain and heart and there was no evidence to suggest expression of this transcript in any of the other tissues examined. The elevated level of expression observed in the testes may be as a consequence of the proposed role played by *ERCC1* in recombination.

3.2 The novel *ERCC1* transcript is evident in some cultured cell lines

Having identified a novel *ERCC1* transcript in mouse skin it was important to us to see if the transcript was also evident in any cultured cells. Initial attempts to isolate good quality RNA from skin biopsies were often technically problematic, thought to be due to a greater abundance of nucleases associated with the tissue. Identifying an *in vitro* source of the transcript would greatly facilitate any subsequent efforts towards the molecular characterisation of the novel *ERCC1* transcript and the elements affecting its expression.

A panel of cultured cell lines, from those readily available in our laboratory was screened by northern blot analysis for the presence of the novel *ERCC1* transcript (Figure 3.2). Total RNA was isolated from the SV40 transformed wild-type human fibroblast line MRC5V1, the *HPRT*-deficient mouse embryonic stem cell line HM1, the mouse L-cell line RJK691, the wild-type Chinese hamster ovary cell line CHO9, the *ERCC1* deficient Chinese hamster ovary cell line CHO43 3B, the transformed wild-type mouse embryonic fibroblast line PF20, the transformed *ERCC1* deficient embryonic fibroblast line PF24, a wild-type mouse keratinocyte line Ker.(WT), an *ERCC1* deficient mouse keratinocyte line Ker.(null) and a wild-type mouse primary neuronal culture HMB4.4. The RNA was electrophoresed on a 1.4% agarose-

Figure 3.2 The novel *ERCC1* transcript is present in some cultured cell lines

Northern blot analysis of RNA samples from a range of cultured cell lines. Total RNA was prepared from a transformed human fibroblast cell line (MRC5V1), a mouse embryonic stem cell line (HM1), a mouse L cell line (RJK691), a wild-type mouse fibroblast line (PF20), an *ERCC1* null fibroblast line (PF24), a wild-type mouse keratinocyte line (Ker.WT), an *ERCC1* null mouse keratinocyte line (Ker. Null) and a wild type mouse neuronal culture (HMB4.4). Upper panel, RNA was electrophoresed through a 1.4% agarose-formaldehyde gel, transferred onto a nylon membrane and probed with an 800bp BamHI fragment of the *ERCC1* cDNA, corresponding to exons 1 to 8. Positions of the 18S and 28S rRNAs are indicated. Lower panel, ethidium bromide staining to illustrate RNA loads.

MRC 5V1
HM1
RJK691
CHO9
CHO43 3B
PF20T (WT)
PF24T(null)
Keratinocyte.(WT)
Keratinocyte.(null)
HMB 4.4
ERCC1
28S
18S
EtBr

formaldehyde gel, transferred onto a nylon membrane and *ERCC1* transcript detected by hybridisation to a mouse *ERCC1* cDNA probe.

The ethidium bromide staining of the gel is presented in the lower panel and indicates that the amounts of RNA in each lane are essentially equivalent. The normal 1.1kb transcript was present in all lanes except for the *ERCC1*-deficient fibroblasts and keratinocytes. The high molecular weight species evident in the null keratinocyte lanes was believed to be the transcript arising from the *ERCC1* targeted allele (McWhir *et al.*, 1993), however, the targeted allele transcript was not evident in the PF24 lane. This could be a consequence of partial degradation of the RNA sample preventing detection of a higher molecular weight transcript normally present at very low abundance. The nature of the *ERCC1* mutation in the CHO43 3B cell line is such that a transcript is still evident although no functional protein is made (Hayashi *et al.*, 1998).

In the MRC5V1 lane there appeared to be a number of additional larger transcripts, these did not correspond with the size of the novel *ERCC1* transcript and were most likely due to contaminating DNA in the RNA sample. The transcripts in HMB 4.4, the neuronal culture, exhibited a degree of heterogeneity similar to that observed in mouse brain. The *ERCC1* pattern of expression first identified in mouse skin was also mirrored in the wild-type mouse fibroblast and keratinocyte lanes. In these lanes the major transcript was again visibly larger than the normal 1.1kb transcript observed in other cell types. It was also evident from this blot, assuming equal loading of RNA, that the level of *ERCC1* expression in these two cell lines was higher than in any of the other cell lines tested, with expression apparently highest in the keratinocytes.

3.3 Estimated size of the novel, skin specific *ERCC1* transcript

Further northern blot analysis was performed to provide an estimation of size for the novel skin specific *ERCC1* transcript. Duplicate panels of total RNA from ES cells and mouse skin were electrophoresed on a 1.4% agarose-formaldehyde gel, transferred onto a nylon membrane and then the two panels were probed separately using an *HPRT* cDNA fragment and an *ERCC1* cDNA fragment (data not shown). A

comparison between the size of the skin specific *ERCC1* transcript and the size of the *HPRT* transcript detected in ES cells, showed that they were very similar. The *HPRT* mRNA is known to be 1.5kb (Costello, 93). Thus the novel *ERCC1* transcript is approximately 400bp larger than the normal 1.1kb transcript.

3.4 The size of the coding region of the *ERCC1* skin specific transcript is unaltered

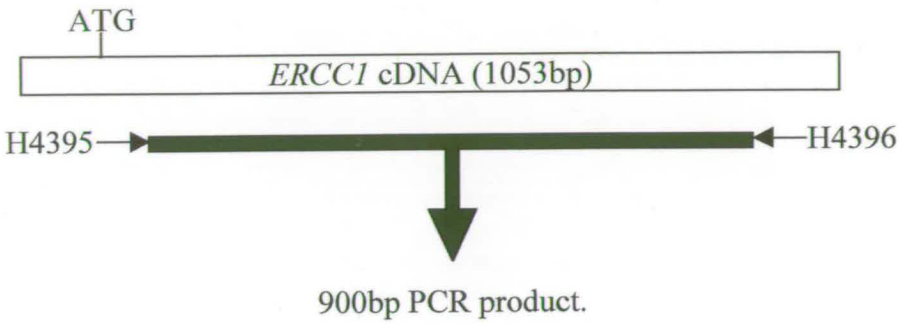
Having identified a novel *ERCC1* transcript of approximately 1.5kb it was necessary to discover the molecular basis for this difference in size. The first possibility we considered was that the skin specific transcript contained additional coding sequences and ultimately encoded a different skin-specific NER protein. RT-PCR was used to compare the size of the *ERCC1* coding region of the normal transcript from ES cells with that of the larger skin specific transcript.

Total RNA from ES cells and two separate wild-type mouse skin samples was reverse transcribed using an oligo-dT primer to generate a pool of cDNAs. The *ERCC1* cDNA was then PCR amplified from these pools using primers specific for exons 1 and 10, electrophoresed on a 1% agarose gel and visualised by ethidium bromide staining. At this stage PCR products, of the predicted 900bp size, were clearly visible in the ES cell lanes and were faintly evident in only one of the skin sample lanes (data not shown). To confirm that the PCR products were *ERCC1* cDNA and not simply artefacts of PCR, Southern blot analysis was performed. The electrophoresed PCR products were denatured and transferred onto a nylon membrane, then probed using an *ERCC1* cDNA probe (Figure 3.3). A schematic representation of the *ERCC1* cDNA and the positions of the PCR primers is shown in the top panel. The detection of *ERCC1* sequences by probing with *ERCC1* cDNA is shown in the lower panel. The Southern blot confirms, by hybridisation, that the PCR products resultant from RT-PCR analysis of total RNA from ES cells and mouse skin were *ERCC1* cDNA. The absence of a positive signal in the second skin lane was probably due to technical difficulties experienced with initial attempts at the isolation of RNA from skin biopsies. It is possible that this RNA sample was slightly degraded and as a result the initial round of reverse transcription was not succesful

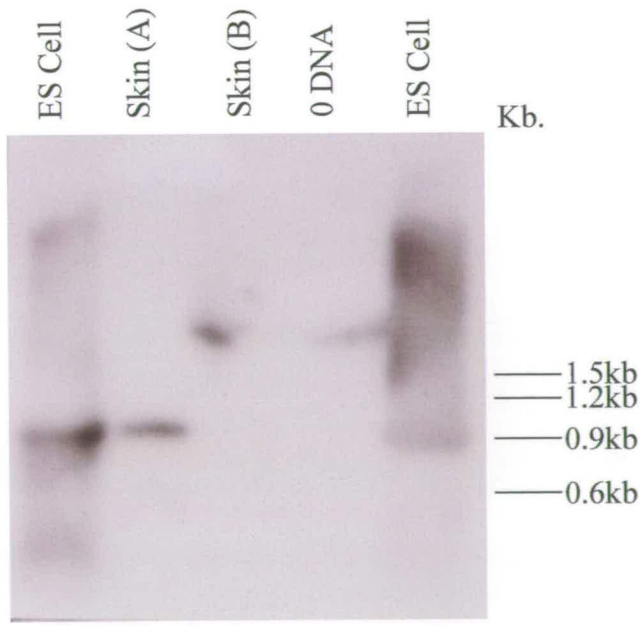
Figure 3.3 RT-PCR analysis of *ERCC1* transcripts in ES cells and mouse skin

a, the *ERCC1* cDNA is represented schematically, with the relative positions of the *ERCC1* primers shown. Scale, 1kb to 10 cm. b, Southern blot analysis of *ERCC1* RT-PCRs. Total RNA from ES cells and two separate wild-type mouse skin samples was reverse transcribed using an oligo-dT primer. The *ERCC1* cDNA was then PCR amplified from these pools using primers specific for exons 1 (H4395) and 10 (H4396), electrophoresed on a 1% agarose gel, transferred to a nylon membrane and probed with an *ERCC1* cDNA fragment spanning exons 1 to 8.

a.


ATG
ERCC1 cDNA (1053bp)
H4395
H4396
900bp PCR product.

b.


ES Cell
Skin (A)
Skin (B)
0 DNA
ES Cell
Kb.
1.5kb
1.2kb
0.9kb
0.6kb

and consequently the second stage PCR failed. Where *ERCC1* sequences were detected, in the ES cell lanes and in the first skin lane, it was evident that the products were the same size, the predicted 900bp. Whilst this result may indicate that the coding region of the skin specific transcript is equal in size to that of the normal transcript, it remains a formal possibility that the product amplified from the skin RNA was derived from the normal sized transcript which is also evident in this tissue. Furthermore, the analysis presented here fails to exclude the possibility of an exon being included in the skin specific transcript which lies either upstream or downstream of the known coding sequences

3.5 Characterisation of the 3' end of the *ERCC1* cDNA in mouse skin and ES cells

The 3' UTR and poly(A) tail have been implicated in the mRNA stability and translational efficacy of a number of genes. It is generally accepted that the longer the 3'UTR the more stable the mRNA. It was possible that the increased size of the skin specific transcript was the result of differential polyadenylation giving rise to a more stable *ERCC1* transcript. Increasing the half-life of the transcript could be seen as one way of meeting the increased demands on the protein in skin, by enabling more protein to be translated. Furthermore the differential polyadenylation of *ERCC1* transcripts in fibroblasts has previously been reported (van Duin *et al.*, 1987).

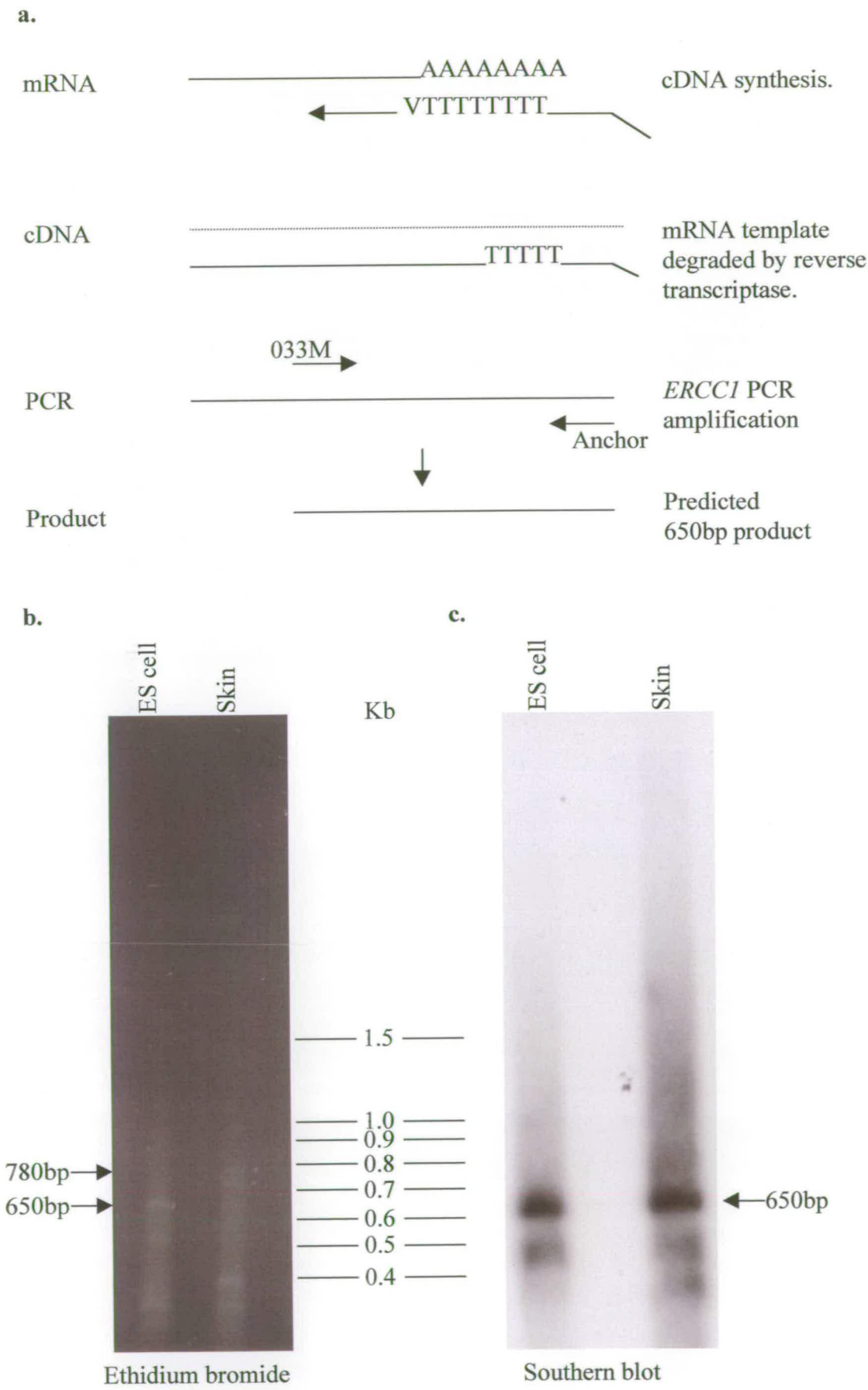
The 3' ends of the *ERCC1* cDNAs from ES cells and mouse skin were characterised using 3' RACE (rapid amplification of the cDNA ends) analysis. The method is essentially a modified RT-PCR protocol enabling the amplification of the 3' end of a cDNA where upstream sequence data are known. The method exploits the naturally occurring poly(A)-tail of mRNAs as a site for PCR priming. First-strand cDNA synthesis is performed using an oligo dT- anchor primer, which is actually a pool of oligo dT primers with a single non T nucleotide at the 3' end, forcing the primer to bind to the inner limit of the poly (A) tail. The anchor element is a stretch of 22 nucleotides at the 5' end of the primer which contain restriction sites (SalI, ClaI, and MluI) to facilitate subsequent cloning steps where required. Following the cDNA synthesis the 3' end of the cDNA of interest is amplified directly using a PCR anchor

primer in combination with a gene specific primer. The resultant product is suitable for cloning and subsequent restriction or sequencing.

Total RNA samples (2μg) from ES cells and mouse skin were reverse transcribed using the oligo dT- anchor primer. Without purification of the resultant cDNA pool, PCR amplification of the 3' end of the *ERCC1* cDNA was performed. Using 1μl of the cDNA reaction, PCR was performed with the anchor primer (a 22mer homologous to the 5' end of the oligo dT-anchor primer) and a primer specific for exon 4 of the *ERCC1* cDNA (033M). The predicted size for the resultant PCR product was 664bp, from the published *ERCC1* cDNA sequence, assuming the normal polyadenylation site was used (van Duin *et al.*, 1988). Half of the resultant PCR product was then electrophoresed on a 1.2% agarose gel and visualised by ethidium bromide staining. To be able to distinguish between PCR products that were *ERCC1* cDNA products and not simply artefacts of PCR, Southern blot analysis was performed. The electrophoresed PCR products were denatured and transferred onto a nylon membrane, then probed using an *ERCC1* cDNA probe (Figure 3.4). Panel a, shows a schematic representation of the 3' RACE analysis performed on the *ERCC1* cDNA. Panel b, the ethidium bromide stained gel showed an apparent difference between the PCR profiles resultant from the two cDNA sources. The ES cell lane contained a heterogeneous DNA smear, the largest of the clear bands was approximately 650bp. The PCR profile resultant from the mouse skin PCR showed an equally heterogeneous DNA smear, however, in this case the largest distinct band was visible at approximately 780bp. At this stage the PCR results suggested that the 3' ends of the *ERCC1* cDNAs in ES cells and mouse skin may be different. Panel c, the Southern blot analysis of the PCR products showed a distinctly different result from that seen after ethidium bromide staining. The *ERCC1* probe hybridised to distinct bands in both the ES cell and mouse skin lanes. In ES cell lane the probe hybridised to the 650bp band and a smaller band of approximately 500bp. In the mouse skin lane the probe also hybridised to bands of 650bp and 500bp as well as a third band of approximately 400bp. The 780bp band visible after ethidium bromide staining did not hybridise with the *ERCC1* probe and was therefore most likely an artefact of the PCR reaction. The Southern blot analysis of the PCRs showed that the

Figure 3.4 3' RACE analysis of *ERCC1* cDNAs in skin and ES cells

a, A schematic representation of the principles of 3' RACE analysis is shown, cDNA was synthesised by reverse transcription of total RNA, using an oligo dT-anchor primer. The position of the variable nucleotide (A, C or G) is indicated as V. The *ERCC1* cDNA was amplified using a primer specific for *ERCC1* exon 4 in combination with an anchor primer. The PCR product was then visualised by gel electrophoresis. b, ethidium bromide stained gel of electrophoresed PCR products from 3' RACE analysis of ES cell and skin cDNAs. The 650bp product from ES cells, and an 780bp from skin are indicated. c, Southern blot analysis of the 3' RACE PCR products. Following electrophoresis, the PCR products were transferred to a nylon membrane and probed with an *ERCC1* cDNA fragment corresponding to exons 1 to 8. The hybridising band at 650bp is indicated.


a.
mRNA
AAAAAAAA
VTTTTTTTT
cDNA synthesis.
cDNA
mRNA template degraded by reverse transcriptase.
TTTTT
033M
PCR
ERCC1 PCR amplification
Anchor
Product
Predicted 650bp product
b.
c.
ES cell
Skin
Kb
1.5
1.0
0.9
0.8
0.7
0.6
0.5
0.4
780bp
650bp
650bp
Ethidium bromide
Southern blot

3' end of the *ERCC1* cDNA in mouse skin was not larger than the 3' end in ES cells. The smaller hybridising bands in skin and ES cells most likely represent artefacts of PCR rather than actual shorter *ERCC1* transcripts. Once again, the products amplified from the skin RNA sample could have been derived from the normal 1.1kb *ERCC1* transcript which is also present in skin, at a much lower level than the 1.5kb transcript. However, the failure to detect a larger product in skin suggests that it is unlikely that the difference between the *ERCC1* mRNAs is due to differential polyadenylation.

3.6 Characterisation of the 5' end of the *ERCC1* cDNA in mouse skin and ES cells

So far it has been suggested that the increased size of the novel skin specific transcript was not due to additional sequences within the coding region or within the 3' UTR. A simple process of elimination would then suggest that the difference between the two transcripts must lie at the 5' end. To investigate this hypothesis the 5' ends of the *ERCC1* cDNAs from ES cells and mouse skin were characterised using 5' RACE analysis.

Once again, 5' RACE is an RT-PCR based method which enables the characterisation of unknown 5' cDNA ends where internal cDNA sequence data are available. The first strand cDNA synthesis is carried out using a gene specific 3' to 5' primer, a homopolymeric A-tail is added to the 3' end of the cDNA by the action of terminal transferase. The tailed cDNA is then PCR amplified using a second gene specific primer, in combination with the oligo dT-anchor primer. A further round of PCR is carried out using the anchor primer in combination with a third nested gene specific primer. The resultant product is suitable for cloning and subsequent restriction analysis or sequencing.

Total RNA samples (2μg) from ES cells, keratinocytes and mouse skin were reverse transcribed using a primer specific for exon 5 of the mouse *ERCC1* cDNA (035M). The resultant cDNA pool was purified using a PCR product purification kit. The PCR product binds to a glass fibre platform within a microfuge tube in the presence of guanidine-thiocyanate, the unincorporated nucleotides and any protein or salt

contaminants are removed by washing, before the PCR product is eluted in re-distilled water. A poly(A)-tail was added to the 3' end of the first strand by the action of terminal transferase. A first round of PCR was carried out on 5µl of the tailed cDNA using an *ERCC1* primer specific for exon 1 (N1138) in combination with the oligo dT-anchor primer. The resultant PCR products were visualised by electrophoresis on a 1.5% agarose gel and ethidium bromide staining. At this stage there was an apparent difference of only ~200bp between the strongest product visible from ES cells and the largest product from mouse skin (Figure 3.5b). A further round of PCR was performed, on the ES cell and skin PCR products, using a nested *ERCC1* primer from exon 1 in combination with the anchor primer. The resultant PCR products were electrophoresed on a 1.5% agarose gel and visualised by ethidium bromide staining (Figure 3.5b). Panel a, shows a schematic representation of the *ERCC1* cDNA and the relative positions of the primers used in the 5' RACE analysis are indicated. The positions of the transcriptional and translational start sites are also indicated for reference. The transcriptional start site indicated is that mapped for the human gene (van Duin *et al.*, 1987) Panel b, the first panel shows the ethidium bromide stained gel of the first round PCRs, all lanes showed a heterogeneous smear of DNA. The strongest band in the ES cell lane was approximately 280bp, compared with a band of almost 450bp in skin. There was no evidence of this 450bp product in the keratinocyte lane. Following this observation the keratinocyte sample was not used for subsequent rounds of PCR. The ethidium bromide stained gel of the nested PCR showed an apparent difference between the PCR profiles from the ES cell and skin cDNA sources. The ES cell lane contained a single diffuse band of approximately 150 to 200bp. The PCR profile resultant from the mouse skin PCR again showed a heterogeneous DNA smear, however, in this case a distinct band was evident at approximately 550bp. An additional diffuse band, believed to be an artefact of PCR, was clearly evident at approximately 150bp.

This apparent size difference of almost 400bp closely matches the size difference as estimated from the earlier northern blot analysis. The size of the 5' RACE product from ES cells maps it closely to the predicted transcriptional start site for the normal transcript, as previously determined for the human gene. The 150-200bp RACE

product was amplified from nt position +42, this would then map the end point to a nt position between -108 and -158. The transcriptional start site in humans was mapped to nt position -112. In contrast to the RT-PCR and 3' RACE analysis, the 5' RACE did reveal differences between the products evident in stem cells and skin. At the level of PCR fragment analysis it appears that the difference between the *ERCC1* skin specific transcript and the normal transcript may be linked to differences at the 5' end of the mRNA.

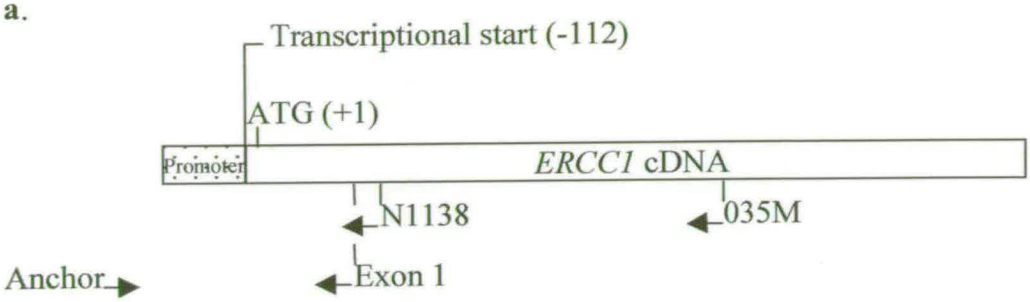
3.7 Cloning of the 5' RACE products from ES cells and mouse skin

Multiple PCRs were performed on a dilution of the first round PCRs from both sources, in order to generate sufficient product for cloning. The PCRs for each source were pooled and separated by low melting point gel electrophoresis. The 150-200bp product was isolated and purified from the ES cell PCRs and the 550bp product was isolated and purified from the skin PCRs. The fragments were then cloned blunt ended into the SmaI site of pBluescriptII SK+. Recombinants were identified by blue/white colour selection and characterised by restriction enzyme analysis (not shown). The 5' RACE clone from ES cells was labelled ES/ 5' RACE #1, and restriction analysis confirmed that the insert was approximately 200bp. In order to map the end point of the cDNA more precisely, the insert was sequenced (Figure 3.6) It was previously assumed that the transcriptional start site in mouse would be the same as that already determined for humans. As such the site for initiation of mouse *ERCC1* transcription mapped to a position upstream of the start of the published mouse cDNA sequence (accession number: X07414) and within the published promoter sequence (accession number: X07413). With this in mind and for the purposes of illustration the sequence of the HM1 5' RACE clone, ES/ 5' RACE #1, was aligned with a sequence compiled from the promoter region and the cDNA. The 120 bases at the 5' end of the published cDNA sequence being identical to the 3' limit of the published promoter region (sequence alignments were performed using Wisconsin Package Version 9.1, Genetics Computer Group (GCG), Madison, Wisc., and were carried out by Jill Douglas). Sequencing of the HM1 5' RACE clone revealed the size of the insert to be 189bp. The sequence was numbered from the

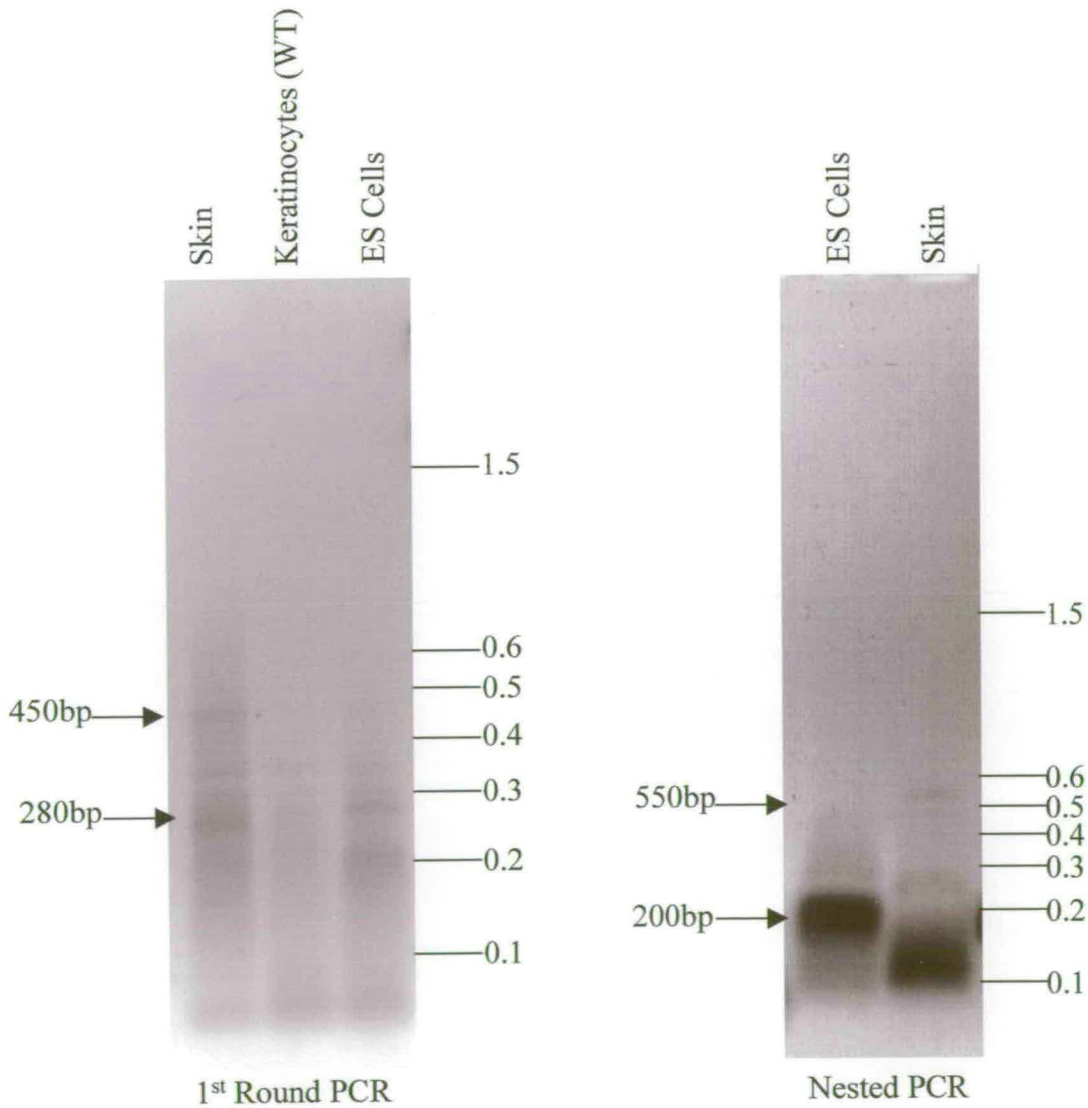
Figure 3.5 5' RACE analysis of *ERCC1* cDNAs in skin and ES cells

a, A schematic representation of the *ERCC1* cDNA and promoter region. The promoter region is indicated as a dotted box, the cDNA region as an open box. The relative positions of primers used for reverse transcription (035M) and subsequent rounds of PCR amplification (N1138, Exon 1 and Anchor) are shown. The positions of the transcriptional (human gene) and translational start sites are indicated. Shown to scale: 1cm = 100bp. b, After reverse transcription using primer 035M and addition of a homopolymeric poly(A)-tail, the tailed *ERCC1* cDNA was amplified using primers N1138 and the anchor primer. The PCR product was visualised by gel electrophoresis and ethidium bromide staining (first panel), the positions of bands of 450bp in the skin and 280bp in the ES cell track are indicated. A further round of nested PCR was performed on these products using the Exon 1 and anchor primers. The second round PCR products were electrophoresed and visualised by ethidium bromide staining (second panel), the positions of bands of 200bp in ES cells and 550bp in the skin track are indicated.

a.


Transcriptional start (-112)
ATG (+1)
Promoter
ERCC1 cDNA
N1138
035M
Anchor
Exon 1

b.


Skin
Keratinocytes (WT)
ES Cells
1.5
0.6
0.5
0.4
0.3
0.2
0.1
450bp
280bp
1st Round PCR
ES Cells
Skin
1.5
0.6
0.5
0.4
0.3
0.2
0.1
550bp
200bp
Nested PCR

translational start site, with the A of the ATG being +1. The Sequence analysis showed that the homopolymeric A-tail, of 14 nucleotides (nt), added by terminal transferase to the end of the cDNA, began at nt position -113. This mapped the 5' end of the *ERCC1* cDNA in mouse ES cells to nt position -112. This coincides exactly with the position of the transcriptional start site previously determined for the human *ERCC1* gene. The sequence alignment showed that the amplified *ERCC1* sequence matched the published mouse sequences except for the last four 5' nucleotides.

The 5' RACE clone from skin was labelled Skin/ 5' RACE #17, and restriction analysis confirmed the insert was approximately 580bp. The insert was then sequenced in order to map the end point of the cDNA in skin more precisely. (Figure 3.7) It was known that the transcriptional start site for the human *ERCC1* mapped to a position upstream of the published cDNA sequence (accession number: X07414) and within the published promoter sequence (accession number: X07413). Once again, for the purposes of illustration the sequence of the skin 5' RACE clone, Skin/ 5' RACE #17, was aligned with a sequence compiled from the promoter region and the cDNA. (sequence alignments were performed using Wisconsin Package Version 9.1, Genetics Computer Group (GCG), Madison, Wisc., and were carried out by Jill Douglas). Sequencing of the skin 5' RACE clone revealed the size of the insert to be 579bp. The 5' end sequence of the Skin/ 5' RACE #17 insert was dominated by polypyrimidine/polypurine tracts (CT repeats). As previously, the sequence was numbered with the A of the translational start (ATG) being numbered +1. The Sequence analysis showed that a homopolymeric A-tail, of 16 nucleotides (nt), added by terminal transferase to the end of the cDNA, began at nt position -502. This mapped the 5' end of the *ERCC1* cDNA in mouse skin to nt position -501, 389 nt upstream of the transcriptional start site mapped in mouse ES cells and human cells. This apparent difference fitted well with the predicted size, 1.5kb of the *ERCC1* skin specific transcript, compared to the 1.1kb normal transcript. The sequence alignment showed that the amplified sequences matched the published sequences, except for a single base mismatch at nt position -31. In the 5' RACE clone the nt at this position was a T, compared with a C in the published sequence. This nt difference was not observed in the analogous region of the HM1 5' RACE clone and as such it most

Figure 3.6 Sequence of the ES/ 5'RACE #1 insert aligned with the published mouse *ERCC1* promoter and cDNA sequences

The sequence of the ES/ 5'RACE #1 insert was used to precisely map the end point of the 5' RACE product in ES cells. The sequence (HM1race) was aligned with a composite of the published mouse *ERCC1* cDNA and promoter sequences (promocDNA) (Accession numbers: X07414 and X07413). The sequence corresponding to the anchor primer is shown in red, the sequence corresponding to primer Exon 1 is shown in blue. The translational start site is shown in pink and is numbered +1. The position of the transcriptional start site (human) is indicated by an asterisk at nt -112. The sequence corresponding to the homopolymeric tail added by terminal transferase is underlined (single). The position of the four mismatched nucleotides are underlined (double).


-222 -173
promocDNA AAGCTTCCAT CCGCCCCAAA CCACAGCGGT CCTCCAGGAC CATAGAGAGC
HM1race ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
-172 -123
promocDNA AGCGCGAATA GAGTTCCCCG CTCTAACTCC TCCGGGGAGC AGCGAGACGA
HM1race ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~ACCAC GCGTATCGAT GTCGACTTTT
-122 * -73
promocDNA GCGAAGGGCC AGAGGGCCGG AAGTGAGTCT AGCAGGAGTT GTGCTGGCTG
HM1race TTTTTTTTTT GAGCGGCCGG AAGTGAGTCT AGCAGGAGTT GTGCTGGCTG
-72 -23
promocDNA TGCTGGCGTT GTGTCGCCTC TGTTTCCCCC CGTGGTATTT CCTTCTAGGC
HM1race TGCTGGCGTT GTGTCGCCTC TGTTTCCCCC CGTGGTATTT CCTTCTAGGC
-22 +1 28
promocDNA ATCGGGAAAG ACCAGGCCCC AGATGGACCC TGGGAAGGAC GAGGAAAGTC
HM1race ATCGGGAAAG ACCAGGCCCC AGATGGACCC TGGGAAGGAC GAGGAAAGTC
29 42
promocDNA GGCCACAGCC CTCA
HM1race GGCCACAGCC CTCA

likely reflects a PCR error. The 5' end of the cDNA in mouse skin extends beyond the limit of the published mouse promoter sequence, as such it was not possible to verify the integrity of the sequence using database searches. Subsequently a genomic clone encompassing this region was isolated within our laboratory (Kan-Tai Hsia, unpublished data.).The 1053bp genomic fragment encompassed all but 9nts, from nt -231 to nt -223, of the skin 5' RACE clone. This has enabled sequence alignment and comparisons to be made between the genomic clone and the sequence of the Skin/ 5' RACE #17 insert. The genomic fragment contained stretches of CT repeats; a $(CT)_{34}$ beginning at nt -428 and a $(CT)_4$ at nt -358. The sequence of the Skin/ 5' RACE #17 insert matched the sequence at the 3' end of the genomic fragment between nts -351 and nt -232. Upstream of nt -351 the two sequences become quite divergent. The 5' RACE clone contains the $(CT)_4$ repeats but, although it does have a polypyrimidine/polypurine tract at the 5' end, it does not align perfectly with the adjacent $(CT)_{34}$ repeats of the genomic fragment. In the Skin/5' RACE #17 insert, the absolute number of CT repeats was greater although the tracts were disrupted. In addition to the $(CT)_4$ repeat shared with the genomic fragment, the Skin/5' RACE #17 insert also had $(CT)_6$, $(CT)_7$, $(CT)_5$, $(CT)_8$ and a $(CT)_{30}$ repeat only 7 nt downstream of the 5' end point of the skin cDNA.

In summary, the 5' limits of the HM1 and skin 5' RACE clones have been mapped by sequence analysis of the cloned RACE fragments. The 5' end of the normal *ERCC1* cDNA from ES cells was positioned at nt -122, which matches the transcriptional start site reported for the human *ERCC1* gene. The 5' end of the skin 5' RACE clone was mapped to nt position -501, 389nts upstream of the transcriptional start site used for the normal *ERCC1* transcript. This suggests that the transcription of *ERCC1* in skin may be driven by an alternative upstream promoter element not utilised in other tissues. Sequence analysis of the skin 5' RACE clone revealed an extensive tract of CT repeats at the 5' limit of the clone, the functional significance of such sequences will be discussed later.

Figure 3.7 Sequence of the Skin/ 5'RACE #17 insert aligned with the published mouse *ERCC1* promoter and cDNA sequences and unpublished 5' flanking sequences

The sequence of the Skin/ 5'RACE #17 insert was used to precisely map the end point of the 5' RACE product in skin. The sequence (skin) is shown aligned with a composite of the published mouse *ERCC1* cDNA, promoter sequences (5'promcDNA) (Accession numbers: X07414 and X07413) and the 3' end of the unpublished 5' flanking sequence. The 3' end of the unpublished 5' flanking sequence is shown in italics. The sequence corresponding to the anchor primer is shown in red, the sequence corresponding to primer Exon 1 is shown in blue. The translational start site is shown in pink and is numbered +1. The position of the transcriptional start site (human) is indicated by an asterisk at nt -112. The CT repeat tracts in both sequences are underlined, genomic sequence: broken underline, insert sequence : double underline. A single mismatch between the 5'RACE sequence and the published elements of the composite sequence is shown in bold type. Mismatches with the unpublished 5' flanking sequence are not indicated.

```
            -537                                                    -488
5'promcDNA  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~GTGA GTCAATGTCT
      skin  CCACGCGTAT CGATGTCGAC TTTTTTTTTT TTTTTTCTTT TTTCTCTCTC

            -487                                                    -438
5'promcDNA  GTGGGCTGAG TCCACCCACG CTTACAGGGA AGGGCGTCCT CTGAAATCTC
      skin  TCTCTCTCTC TCTCTCTCTC TCTCTCTCTC TCTCTCTCTC TCTCTCTCTC

            -437                                                    -388
5'promcDNA  ACCTCCCTCC TCTCTCTCTC TCTCTCTCTC TCTCTCTCTC TCTCTCTCTC
      skin  TCTGTGTGTG TCTCTCTCTC TCTCTCTGTC TCTCTCTCTG TCTCTCTCTC

            -387                                                    -338
5'promcDNA  TCTCTCTCTC TCTCTCTCTC TCTCTCTGTC TCTCTCTCCT TTCCTCCTTT
      skin  TCTCTGTCTC TCTCTCTCTT TGTCTCTGTC TCTCTCTCCT TTCCTCCTTT

            -337                                                    -288
5'promcDNA  CCCTCCCCCA CCTCCTTTGA TCATCACTGC GCCGGATCTG GAGTCTGGGA
      skin  CCCTCCCCCA CCTCCTTTGA TCATCACTGA GCCGGATCTG GAGTCTGGGA

            -287                                                    -238
5'promcDNA  AGCGCTTAAG AGGGCCTTGG AACACAACTA CTTGAAGTCA AAGTCTCCCA
      skin  AGCGCTTAAG AGGGCCTTGG AACACAACTA CTTGAAGTCA AAGTCTCCCA

            -237                                                    -188
5'promcDNA  GGTACC~~~~ ~~~~~AAGCT TCCATCCGCC CCAAACCACA GCGGTCCTCC
      skin  GGTACCAAGA TGCACAAGCT TCCATCCGCC CCAAACCACA GCGGTCCTCC

            -187                                                    -138
5'promcDNA  AGGACCATAG AGAGCAGCGC GAATAGAGTT CCCCGCTCTA ACTCCTCCGG
      skin  AGGACCATAG AGAGCAGCGC GAATAGAGTT CCCCGCTCTA ACTCCTCCGG

            -137                          *                         -88
5'promcDNA  GGAGCAGCGA GACGAGCGAA GGGCCAGAGG GCCGGAAGTG AGTCTAGCAG
      skin  GGAGCAGCGA GACGAGCGAA GGGCCAGAGG GCCGGAAGTG AGTCTAGCAG

            -87                                                     -38
5'promcDNA  GAGTTGTGCT GGCTGTGCTG GCGTTGTGTC GCCTCTGTTT CCCCCCGTGG
      skin  GAGTTGTGCT GGCTGTGCTG GCGTTGTGTC GCCTCTGTTT CCCCCCGTGG

            -37                                          +1          13
5'promcDNA  TATTTCCTTC TAGGCATCGG GAAAGACCAG GCCCCAGATG GACCCTGGGA
      skin  TATTTCTTTC TAGGCATCGG GAAAGACCAG GCCCCAGATG GACCCTGGGA

            14                              42
promocDNA   AGGACGAGGA AAGTCGGCCA CAGCCCTCA
     skin   AGGACGAGGA AAGTCGGGCA CAGCCCTCA
```

3.8 Detection of the skin specific *ERCC1* transcript using a probe isolated from the mouse skin 5' RACE clone (skin/5' RACE #17)

Having characterised the 5' end of the *ERCC1* cDNA in mouse skin using 5' RACE analysis it was important to confirm the integrity of the result. Evidence that the 5' RACE clone was actually cDNA derived would be gained by demonstrating that the 5' end of the RACE clone was suitable for use as a transcript specific probe for the 1.5kb skin transcript.

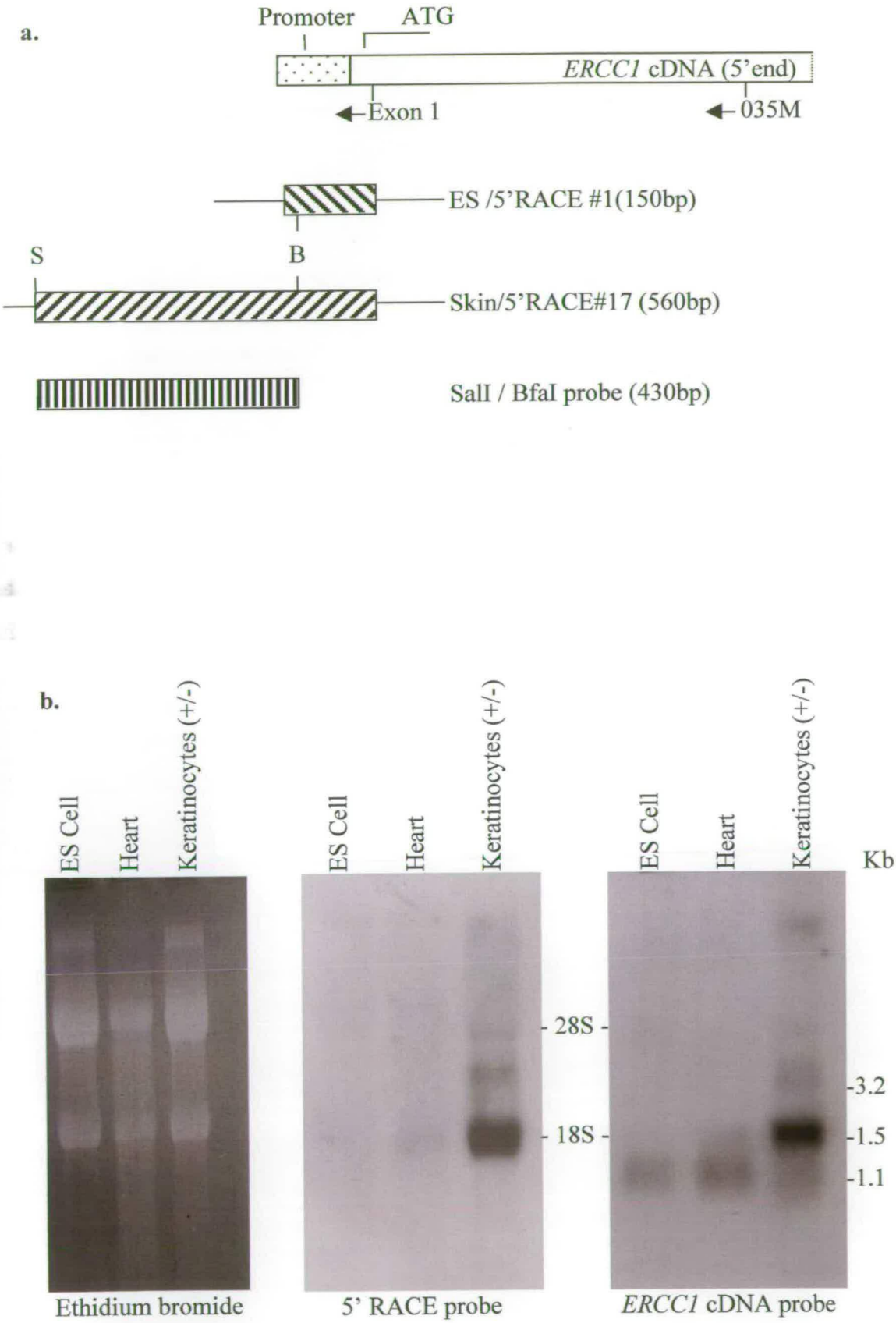
The previously described sequencing of both 5' RACE clones enabled restriction sites to be mapped for each of the cloned fragments. With this information it was possible to identify a region of cDNA unique to the skin clone (skin/5' RACE #17). A 430bp SalI/BfaI fragment, which shared only 14bp of homology with the ES cell cDNA 5' end, was identified as a suitable probe (Figure3.8). Panel a, shows a schematic representation of the 5' end region of the mouse *ERCC1* cDNA, the positions of the translational start site and the location of the *ERCC1* primers are indicated for reference. The relative positions and sizes of the 5' RACE clones from both sources are shown along with a representation of the fragment identified as a suitable probe.

Total RNA (30ug) from ES cells, wild type mouse heart and *ERCC1* heterozygous keratinocytes was separated by electrophoresis, transferred onto a nylon membrane and the *ERCC1* transcript detected by hybridisation to the SalI/BfaI 5' RACE probe (Figure 3.8). The membrane was then stripped and reprobed using an *ERCC1* cDNA probe spanning exons 1 to 8. An *ERCC1* heterozygous keratinocyte RNA was used for this blot to further test the specificity of the transcript specific probe. As detailed in the introduction, there is known to be a 2.8kb transcript associated with the *ERCC1* targeted allele in ES cells. It was predicted that this transcript should be detected in *ERCC1* heterozygous keratinocytes at 3.2kb

Panel b, the ethidium bromide stained gel is shown in the first panel. The RNA loading in each lane appeared reasonably well matched. The second panel shows the filter probed with the SalI/BfaI 5' RACE fragment. It was seen that, as predicted, the fragment did not hybridise with anything in the ES cell lane. In the heart lane there was a suggestion of a faint signal, equal in size to the skin specific transcript. This

Figure 3.8 Northern blot analysis using a probe specific for the novel *ERCC1* skin specific transcript

a, Shows a schematic representation of the 5' end of the *ERCC1* cDNA and promoter region. The promoter region is indicated as a dotted box, the cDNA region as an open box. The relative positions of the RT-PCR primers are indicated for reference, and the translational start site (ATG) is also indicated. The relative positions and sizes of the cloned 5'RACE products from ES cells and skin are shown below. The region of the skin/5'RACE #17 insert used for the probe is also indicated. B, BfaI; S, SalI. Shown to scale 1cm =100bp. b, northern blot analysis. Total RNA (30μg) from ES cells, heart and *ERCC1* heterozygous keratinocytes was electrophoresed through a 1.4% agarose-formaldehyde gel (first panel), transferred onto a nylon membrane and probed with a 430bp SalI/BfaI fragment from Skin/5'RACE #17 (second panel) before being stripped and reprobed with an 800bp BamHI fragment of the *ERCC1* cDNA, corresponding to exons 1 to 8 (third panel). Positions of the 1.45kb and 1.1kb transcripts, as well as 18S and 28S rRNAs are indicated.


a.
Promoter
ATG
ERCC1 cDNA (5'end)
Exon 1
035M
ES /5'RACE #1(150bp)
S
B
Skin/5'RACE#17 (560bp)
SalI / BfaI probe (430bp)
b.
ES Cell
Heart
Keratinocytes (+/-)
Ethidium bromide
5' RACE probe
ERCC1 cDNA probe
28S
18S
Kb
3.2
1.5
1.1

signal could be explained as a consequence of contamination from the neighbouring lane of keratinocyte RNA. A more likely explanation is, however, that the signal arises from the heterogeneous smear associated with the *ERCC1* transcript in heart. This smear may represent a degree of heterogeneity with *ERCC1* transcriptional initiation in the heart. Similar results were seen when this fragment was used to probe brain RNA, which also results in an heterogeneous smear when probed with a cDNA fragment (data not shown). In the third, *ERCC1* heterozygous keratinocyte lane, the probe hybridised strongly with a transcript of 1.5kb. There was no evidence of hybridisation to any smaller transcripts. A larger transcript in the region of 3.0kb, was also evident in this lane as predicted for the targeted allele. The final panel shows the same membrane, this time probed using an *ERCC1* cDNA probe spanning exons 1 to 8. The normal 1.1kb transcript was detected in each of the lanes. In the heart there was also the typical heterogeneous smear above the normal transcript. The keratinocyte sample exhibits the distinctive skin specific pattern of expression, where the 1.5kb transcript was clearly the major species and was detected at greater levels of abundance than the 1.1kb transcript. The transcript from the targeted allele was also evident.

These results add significant weight to the evidence already presented, from the 5' RACE analysis, that the size difference observed between the normal and skin specific *ERCC1* transcripts is due to differences at the 5' end of the transcript. Furthermore, the northern blot analysis has suggested that the although the pattern of expression in skin and keratinocytes is clearly distinct from the pattern seen in brain and heart, the 1.5kb transcript was also present in these tissues at very low abundance.

3.9 Lack of UV inducibility of the novel *ERCC1* skin transcript

Identifying a novel transcript for *ERCC1* was in its own right an interesting result. The tissue specificity of the novel *ERCC1* transcript added a much greater biological significance to the discovery. As previously mentioned, the skin is the body's first line of defence against the harmful effects of UV irradiation. The NER pathway, in which *ERCC1* is involved, is the principal repair mechanism for removing UV

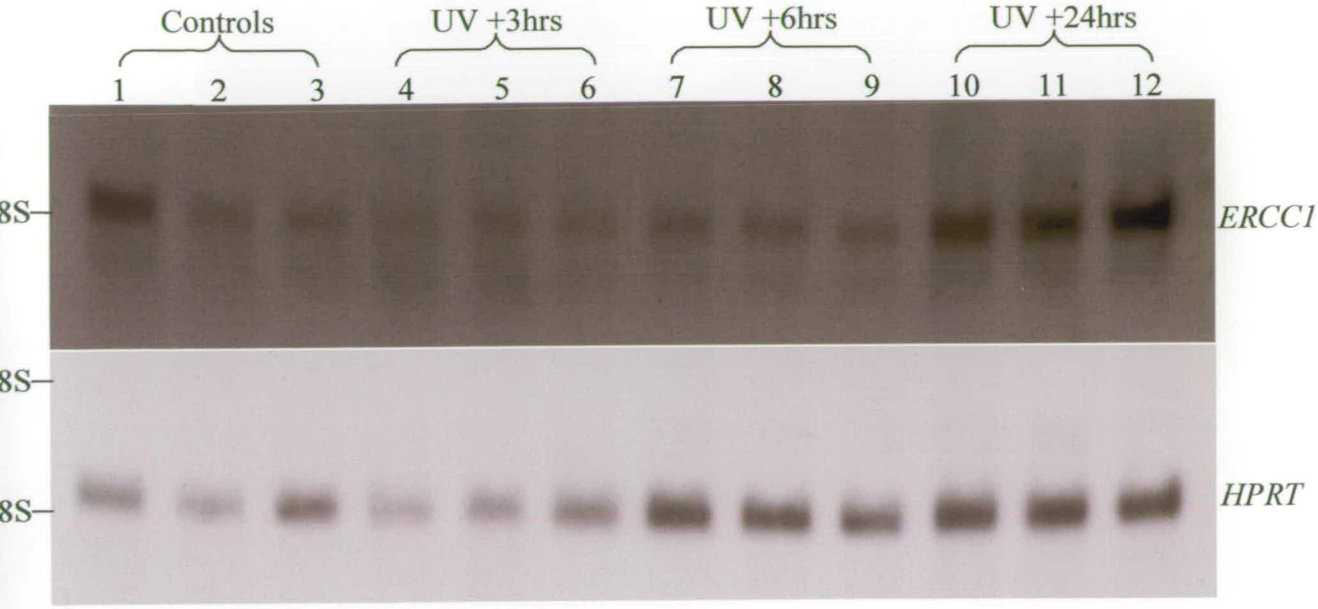
induced DNA lesions. In this context the skin specific expression pattern of the novel *ERCC1* transcript seems perfectly logical. Therefore it would also seem reasonable to imagine that a gene involved in the repair of UV induced DNA damage might itself be UV inducible. It has been reported that the expression of the normal transcript is not induced by UV in Hela cells (van Duin *et al.,* 1987). To investigate the biological significance of the novel *ERCC1* transcript in skin, we decided to see if this 1.5kb transcript was UV inducible.

Twelve adult male BalbC mice were anaesthetised by intraperitoneal injection of Hypnorm and Hypnovel. Whilst under anaesthesia a $6cm^2$ area of the back was shaved, exposing the skin directly to the UV source. Anaesthesia was used to prevent mice from being shaded from the UV source as a consequence of huddling behaviour. The shaved backs of the mice were then exposed to a single 2000 Jm^{-2} dose of UVB light. The average minimal erythemal dose for mice being 1500 Jm^{-2}. To investigate induction of the *ERCC1* gene expression, animals were culled and total RNA was extracted from the areas of shaved dorsal skin at various time points after the UV irradiation. Three animals were culled immediately after anaesthesia and shaving to serve as non irradiated controls. Total RNA was extracted from three irradiated animals at: 3, 6 and 24 hours after UV irradiation. Northern blot analysis was performed to determine whether *ERCC1* mRNA levels were altered following UV irradiation. Total RNA (30ug) from each of the skin samples was separated by electrophoresis, transferred onto a nylon membrane and the *ERCC1* transcript detected by hybridisation to an *ERCC1* cDNA probe. The membrane was later stripped and reprobed using a mouse hypoxanthine phosphoribosyl transferase *(HPRT)* cDNA probe to standardise for any variations in RNA loading in each lane (Figure 3.9). Panel a, the detection of *ERCC1* mRNA by probing with *ERCC1* cDNA is shown in the top panel. The normal pattern of expression was evident in each sample; the 1.5kb transcript remained the major transcript relative to the 1.1kb. The lower panel shows the *HPRT* reprobe of the filter to standardise for loading. At first glance it appeared that the level of *ERCC1* was unchanged, relative to the controls, at 3 and 6 hours after irradiation. It then appeared to be elevated twenty-four hours after irradiation. It can be seen that this apparent elevated level of *ERCC1* correlated with

Figure 3.9 Lack of UV induction of *ERCC1* in skin

a, northern blot analysis of skin RNA from UV irradiated mice. Upper panel, total RNA was prepared from the dorsal skin of mice at various time points after receiving a UVB dose of 2000 Jm^{-2} , electrophoresed on a 1.4% agarose-formaldehyde gel, transferred onto a nylon membrane and probed with an 800bp BamHI fragment of the *ERCC1* cDNA, corresponding to exons 1 to 8. Lower panel, the filter was then stripped and reprobed with a mouse *HPRT* cDNA probe. The positions of the 18S and 28S rRNAs are indicated. b, phosphorimager analysis of *ERCC1* expression following UV irradiation. The values from the phosphorimage were obtained using the ImageQuant V3.3 image analysis package. The signal values represent the area, in arbitrary units, under the peak corresponding to the *ERCC1* and *HPRT* mRNA bands. The ratio represents the value of the *ERCC1* signal divided by the value of the *HPRT* signal, i.e. the standardised *ERCC1* value. The % control is the mean mRNA level in irradiated samples relative to the non irradiated controls, expressed as a percentage.

a.



b.

Sample	Time	*ERCC1* Signal	*HPRT* Signal	Ratio (*ERCC1/HPRT*)	Mean of ratios (±SE)	Average of ratios (3) as % Control
1	control	202.1	315.5	0.64	0.49 (±0.096)	
2	control	117.8	221.3	0.53		100
3	control	101.5	324.1	0.31		
4	+3hrs	96.3	221.3	0.43	0.36 (±0.058)	
5	+3hrs	103.0	259.7	0.40		73.4
6	+3hrs	79.5	326.0	0.24		
7	+6hrs	89.2	487.6	0.18	0.21 (±0.02)	
8	+6hrs	91.7	428.1	0.21		43.5
9	+6hrs	83.0	328.8	0.25		
10	+24hrs	184.3	457.9	0.40	0.48 (±0.062)	
11	+24hrs	207.7	480.0	0.43		97.2
12	+24hrs	260.1	427.8	0.60		

an elevated *HPRT* signal in the same lanes which suggests disproportionate loading of RNA in these lanes. Phosphorimagery was used to quantify the level of *ERCC1* relative to the *HPRT* signal (Figure 3.9). The ratio of the *ERCC1* signal to the *HPRT* signal was calculated for each of the twelve samples and the mean of the three ratios from each time point was used for comparisons between the groups. The control group, which received no UV irradiation, represents the normal level of *ERCC1* expression in skin and the mean ratio of this group (0.49) was regarded as the normal level (100%). The mean ratio of the samples taken 3hrs after irradiation was 0.36 or 73% of normal, for the samples taken 6hrs after irradiation the mean of the ratios was 0.21 or 43% of normal, and finally for the samples taken 24hrs after irradiation the mean of the ratios was 0.48 or 97% of normal. It was clear that, at the time points at which samples were taken, there was no evidence of increased expression of *ERCC1* in skin. The pattern of expression that had emerged did show some changes in the abundance of *ERCC1* in the skin. After irradiation the level of *ERCC1* appeared to fall; to 73% of the normal level after 3hrs and then further to 43% of the normal level after 6hrs, before returning to 97% of the normal level 24hrs after irradiation. This apparent reduction in the level of *ERCC1* may reflect an increase in the rate of turnover of the *ERCC1* mRNA as the cells attempt to deal with lesions induced by the UV irradiation. The expression returns to the normal level after 24hrs, presumably once the insult has been dealt with. In summary, although the levels of *ERCC1* appeared to be affected by UV, the expression of the *ERCC1* skin specific transcript was not induced by UV irradiation as determined by northern blot analysis.

3.11 Discussion

The expression pattern of a gene is, by and large, determined by the biological function and demand on the protein that the gene encodes. Patterns of gene expression take many forms including: tissue specific patterns of expression; where genes are only switched on in certain tissues or cell types; inducible gene expression, where the expression of a gene is switched on or elevated in response to biological demand or environmental stresses; constitutive expression, where genes are expressed at all times in all tissues. Mouse *ERCC1* has been reported as being

expressed in all tissues and cells at low abundance. The gene gives rise to a major transcript of 1.1kb, with differential splicing resulting in a smaller transcript of 1.0kb and differential polyadenylation giving rise to transcripts of 3.4 and 3.8kb (van Duin *et al.*, 1988).

We have identified and characterised a novel mouse *ERCC1* transcript. Northern blot analysis of mouse tissues has revealed that the pattern of *ERCC1* expression in skin differed from all other tissues examined, although the normal 1.1kb transcript is evident in skin, the major transcript was an estimated 1.5kb. This novel transcript was also identified in a number of cultured cell lines including: embryonic fibroblasts and as one might expect, mouse keratinocytes. The presence of a skin specific transcript in embryonic fibroblasts may reflect the manner in which the embryonic fibroblast cell cultures are established; essentially whole embryos at day 12.5 gestation are disaggregated and cultured *in vitro*. As such the resultant fibroblasts may be derived from a range of tissues including the dermis, and thus some cells may express genes in a 'tissue specific' manner. The apparently higher level of expression of the novel *ERCC1* transcript observed in keratinocytes compared to the PF20 cells may be due to the heterogeneous origins of the cells that make up the fibroblast cell line. If the transcript is 'skin specific' and if the cells are not all of a dermal derivation they may not all be capable of expressing the novel transcript, whereas the keratinocytes are by their nature all of a dermal derivation and thus all capable of expressing the skin specific transcript. As mentioned above, there are a number of transcripts associated with *ERCC1*, however none of these other transcripts has been associate with a tissue specific pattern of expression. This skin specific pattern of expression may have arisen in order to meet a greater biological demand for NER in the skin. As the skin is the body's first line of defence against environmental UV irradiation, it is reasonable to assume that the incidence of UV induced lesions is higher in skin than, for instance, in the kidneys. It would therefore seem reasonable that the expression of a gene involved in the removal of such lesions is altered to meet the possible increased biological demand. Although skin specific transcripts have not been reported for any of the other NER genes characterised so far, this may

simply be because this tissue is seldom chosen for such analysis due to difficulty with the RNA extraction.

An *ERCC1* cDNA with altered coding sequences, the result of alternative splicing of exon 8, has previously been described (van Duin *et al*, 1986). We have used RT-PCR to demonstrate that the size of the coding region of the skin specific transcript matched that of the normal 1.1kb transcript of ES cells. Similarly, although differentially polyadenylated *ERCC1* transcripts have been reported (van Duin *et al.*, 1987), we have shown that the skin specific transcript is not the product of differential polyadenylation. Using 3' RACE analysis it was shown that the 3' RACE products from the normal transcript was the same size as the 3' RACE product from skin.

Using 5' RACE we have demonstrated that the increased size of the skin specific *ERCC1* transcript is most likely due to differential initiation of transcription, probably utilising an, as yet, uncharacterised upstream promoter. The transcriptional start site of the normal *ERCC1* transcript in mouse was shown to be at the same nucleotide position as that reported for humans, nt -112 (van Duin *et al.*, 1987). It was noted that the last 4 nts at the 5' end of the ES cell 5' RACE clone (GAGC) did not match the published sequence for that region (AGAG). Whilst this may reflect genuine differences between the mouse strains used in each case, it is more likely that the sequence divergence was an artefact of the PCR amplification. The sequence obtained from the corresponding region of the skin 5' end did match the published sequence. The 5' RACE analysis carried out using skin RNA mapped the 5' end of the tissue specific cDNA to nt position -501, 389 nts upstream from the transcriptional start site of the normal transcript. This is in agreement with the estimated size of the skin specific transcript as being almost 1.5kb. Sequencing of the skin 5' RACE clone revealed a striking number of CT repeats at the 5' end, but the absolute number of repeats was apparently less extensive in a cloned genomic DNA fragment spanning the region. Whilst it is possible that the sequence heterogeneity and the increase in the absolute number of CT repeats could be an artefact of the RT-PCR, an alternative explanation is that the number of repeats are, in this instance, a reflection of the different mouse strains from which the genomic fragment (129/Ola)

and skin 5' RACE clone (BalbC) were derived. Differences in the sizes of the *ERCC1* introns as well as additional restriction fragment length polymorphisms have previously been reported between these strains (Selfridge *et al.*, 1992).

CT repeats are abundant in eukaryotic genomes and have been implicated in a number of biological mechanisms. In *Drosophila*, CT repeats have been shown to have a role in the regulation of the transcription of genes for two heat shock proteins, *hsp26* and *hsp70*. The double-stranded $(CT)_n/(AG)_n$ has been shown to be a binding site for the GAGA transcription factor (GAF)(Wilkins and Lis, 1998). It is believed that the binding of GAF to the CT element plays an important role in the maintenance of DNase I hypersensitive sites in the promoters, which are themselves important for heat-induced expression (Lu *et al*, 1993). Although CT repeats are abundant in rodents, the GAF has yet to be detected in mouse. In promoters that lack a TATA box, CT repeats can act as initiator elements and in many cases are sufficient to drive gene expression (Weis and Reinberg, 1992). The promoter of the chicken malic enzyme gene lacks a TATA box and uses multiple transcription initiation sites (Hodnett *et al.*, 1996). It has been suggested that the $(CT)_7$ tract within the promoter region may function as an alternative promoter in certain tissues or at specific stages of development. It is believed that the ability of the CT repeats to form non-B-DNA structures is of greater functional significance than the binding of any specific protein to the sequence. Increasing the extent of the CT repeats, which increased the likelihood of non-B-DNA structures forming, was shown to increase the transcription level of a reporter gene (Xu and Goodridge, 1998). In this context it is very probable that the CT repeats associated with the *ERCC1* gene play an important role in the regulation of transcription although the sequences that control the skin specific expression are likely to be located upstream of CT tracts.

The sequence divergence between the genomic fragment and the 5' end of the skin RACE clone does cast some doubt over the exact site of transcriptional initiation in skin. Further analysis of the 5' end of the skin specific transcript, using either primer extension or RNase protection analysis would be required to confirm the exact site of transcriptional initiation in skin. However the result of the 5' RACE analysis in skin was supported by the demonstration that 430bp from the 5' end of the Skin/ 5' RACE

#17 insert acted as a transcript specific probe for the 1.5kb transcript, confirming that this region shares homology with the skin specific transcript.

Finally, we have shown that the novel *ERCC1* transcript was not induced by UV irradiation. Although the UVC induction of the *Ligase I* gene, which is involved in NER, has previously been reported (Montecucco *et al*, 1995), this result was in accordance with findings that other NER genes including, XPA and XPB, or the normal *ERCC1* transcript are not induced following UV irradiation (reviewed in Hoeijmakers, 1993). We did observe an apparent increase in the turnover of the *ERCC1* mRNA following irradiation, perhaps as more protein is made to meet the increased biological demand.

This work constitutes the first report of a skin specific transcription pattern associated with a gene involved in NER. The exact mechanism by which this pattern of expression is controlled remains to be elucidated. We speculate the tissue specificity is determined by sequences within a recently isolated genomic fragment, which may harbour an upstream promoter.

Chapter 4

Use of an *ERCC1* minigene to complement DNA repair deficiency in cultured cells

Eukaryotic gene expression is driven by promoter sequences that are generally located upstream of the transcriptional initiation point. It is with these promoter sequences that the basal transcription factors interact in order that the RNA polymerase can then bind. The polymerase responsible for synthesising heterogeneous nuclear RNA, the precursor to mRNA, is RNA polymerase II. Within any promoter there are a number of elements that can determine the efficiency with which the promoter will drive gene expression. The 'classical promoter' elements include: the TATA box, the CAAT box and the GC box. The TATA box (consensus sequence: TATA(T/A)A(T/A)), generally located about 30 nucleotides upstream of the transcriptional initiation site, is implicated in the determination of the start site. The CAAT box (consensus sequence: GGCCAATCT), located about 80 nucleotides upstream from the transcriptional start site, is believed to be involved in controlling the frequency of transcription initiation. The GC box (consensus sequence: GGGCGG) is often present in multiple copies within promoters and binds the transcription factor Sp1.

The characterisation of the mouse *ERCC1* promoter has been reported (van Duin *et al.*, 1988). It was shown that the *ERCC1* promoter, located within 170bp upstream of the transcriptional start site, lacked any clearly identifiable classical promoter elements such as CAAT or TATA boxes. It was shown that 170bp of 5' genomic sequence was able to drive expression of a complete human *ERCC1* cDNA fragment, as measured by the correction of the DNA repair defect associated with CHO43-3B cells. This finding is in keeping with the promoters of other NER genes including XPA (van Oostrom *et al.*, 1994). We have just described (Chapter 3) the identification of a novel skin specific *ERCC1* transcript which arises as the result of transcription initiating at a point upstream of the currently published promoter sequence. This suggests that expression of the *ERCC1* skin specific transcript is under the control of an upstream promoter not previously characterised.

The aim of the work described here was, to construct a functional *ERCC1* minigene which could be used subsequently to identify the control elements responsible for the skin specific pattern of *ERCC1* transcription, by conventional deletional analysis.

4.1 Construction of an *ERCC1* minigene

As previously shown, the normal 1.1kb *ERCC1* transcript is also expressed in skin but is much less abundant than the 1.5kb skin specific transcript. As such, using the 5' flanking sequence from the mouse *ERCC1* gene with a reporter gene, such as chloramphenicol acetyl transferase (CAT), would not have been suitable for the identification of the control elements required for the skin specific pattern of expression. It has been shown that the promoter associated with the normal transcript lies within 170bp of the initiation site, so, depending on the mechanism for the skin specific expression and the cell type in which the assay was performed, CAT expression might still be predicted from a deleted promoter incapable of driving the skin specific pattern of expression if the 170bp promoter remains intact. Using gene fragments readily available in our laboratory an *ERCC1* minigene was constucted in order to define the elements responsible for the skin specific pattern of expression. A minigene was constructed in which a 1.05kb stretch of 5' flanking sequence was linked on to an *ERCC1* gene fragment and the 3' end of the *ERCC1* cDNA which carries the polyadenylation signal. Due to the uncertaintity that remains, at this stage, about the precise transcriptional start site in skin it remains a formal posibility that some of the necessary elements for the skin specific pattern of expression may lie upstream of the 5' flanking sequence utilised in this minigene. The 1.05kb 5' flanking sequence used in the minigene construct was cloned and sequenced by Kan-Tai Hsia (Figure 4.1). The nucleotide sequence of this region showed that the fragment contained a number of interesting sequence motifs, including a stretch of 25 CA repeats at the 5' end in addition to the CT repeats, described earlier (Chapter 3), at the 3' end. As with CT repeats, CA repeats are widespread in the genome and are believed to play a role in chromatin structure. More significantly, it was found that this fragment contained a clearly identifiable TATA box at nt position 456 (TATATAA) and the consensus sequence for a GC box, which binds the transcription factor Sp1, at nt position 305 (GGGCGG). This is the first time that classical promoter elements have been identified in association with an *ERCC1* promoter (Hsia, unpublished data). The TATA box in this fragment is located at approximately 322bp upstream of the putative transcriptional initiation site in skin, as

Figure 4.1 Nucleotide sequence of the mouse *ERCC1* 5' flanking region

The nucleotide sequence of the mouse *ERCC1* 5' flanking region (pRK1.05). The stretch of CA repeats is shown in red. Consensus sequences for GC and TATA boxes are shown in yellow and green respectively. The nt position for initiation of transcription in skin is shown in pink. The CT repeat tract is shown in blue. Data kindly provided by Kan-Tai Hsia.

1 60

GAATTCACAG CGTAGGCATG TGCTTGTACA CACACACACA CACACACACA CACACACACA

61 120

CACACACACA CACACACACC TCACAACTAG GAAAACATTA AAAAAGATAT AGAGGGATTA

121 180

CATATGATGA CCTTTGGGCG TGCCCTCTCT CAATACACGA AGACCAAAAG AAAATGGAGA

181 240

ATACGTTGGT AAATGCATGC AAAATTTTAA ATGGAGAATT TTTAAAAGTA TGAACTCGAA

241 300

GGAAAGATTT GCAGCCTAAT GAAGTGACTA ACAGCATCAG CCCTTAAGCA TCACTCCAAC

301 360

AAATGGGCGG TAATCAAAAG ATGCTCATCT CCTTTTCCTC CCCCTTTTTA ATTTAAAAAA

361 420

ACCGTTGAAT TGCTCATTTC GAAAAAATGG GCAAAGGATA TGAGCCAATC TTTGTTAAAA

421 480

GGAATAGGCA AAGGAAATGA GCAAGAACAA AAAAATATAT AAGACCATTA AACTATTTCC

481 540

TTAGCAATCA CAGAACTACA AGTTGGAACG GGGTGAAACT GTCCTTCCAG GAAATGAGTC

541 600

ATACTAAGTC ATTCATGGGG CCCAGGTCTC GGAACACAGG TGTCCCTTCC TTCAGGAAGC

601 660

TCGTCCTCAC ACCCCAGCTG GGTCAGGTGC TCCCCCTAAC ACGGGCTCAA TGTCCTCTGG

661 720

GCTCTCCCTT CCCGACCCTG CCCGCTCTGG GTTGTCCCTG TCTGACTGTC CCACTGGACT

721 780

TTAAGCCACA GCTGTCGGTA TCAGTCCCTG CTTGGGATGC AAGTGGGGAA TGGCTTAGAA

781 840

TTAGTGAGTC AATGTCTGTG GGCTGAGTCC ACCCACGCTT ACAGGGAAGG GCGTCCTCTG

841 900

AAATCTCACC TCCCTCCTCT CTCTCTCTCT CTCTCTCTCT CTCTCTCTCT CTCTCTCTCT

901 960

CTCTCTCTCT CTCTCTCTCT CTCTGTCTCT CTCTCCTTTC CTCCTTTCCC TCCCCCACCT

961 1020

CCTTTGATCA TCACTGAGCC GGATCTGGAG TCTGGGAAGC GCTCAAGAGG GCCTTGGAAC

1021 1053

ACAACTACTT GAAGTCAAAG TCTCCCAGGT ACC

determined by the 5' RACE analysis. Typically TATA boxes are situated about 25bp upstream of the startpoint, this may suggest that the sequence motif identified in the 5' flanking fragment is simply coincidental and does not function as a TATA box, or, alternatively the 5' end point determined by the 5' RACE analysis is not exact.

The cloning strategy used to construct the *ERCC1* minigene is shown schematically in Figure 4.2. The minigene was derived from three parental plasmids available within our laboratory. The 5' flanking region believed to harbour the upstream promoter was isolated as a 1.05kb SstI/KpnI fragment from a genomic subclone, pRK1.05 (provided by Kan-Tai Hsia). This was hooked up to a 9.3kb KpnI fragment of the gene carrying exons 1 to 7, isolated from the genomic subclone pKpnI 9.3 (provided by Kan-Tai Hsia). The KpnI site within exon 7 was utilised to add on the 3' end of the cDNA carrying exons 7 to 10, the polyadenylation signal and a short 3' UTR, as a 300bp KpnI/EcoRI fragment, in BluescriptII SK+. Recombinants were identified by blue/white colour selection and subsequently characterised by restriction analysis. A single positive clone was identified and the integrity of each of the cloning sites was confirmed by sequencing (data not shown). The resultant minigene, *ERCC1* MG #13, was 10.4kb and could be linearised by restriction at the NotI site in the BluescriptII SK+ polylinker.

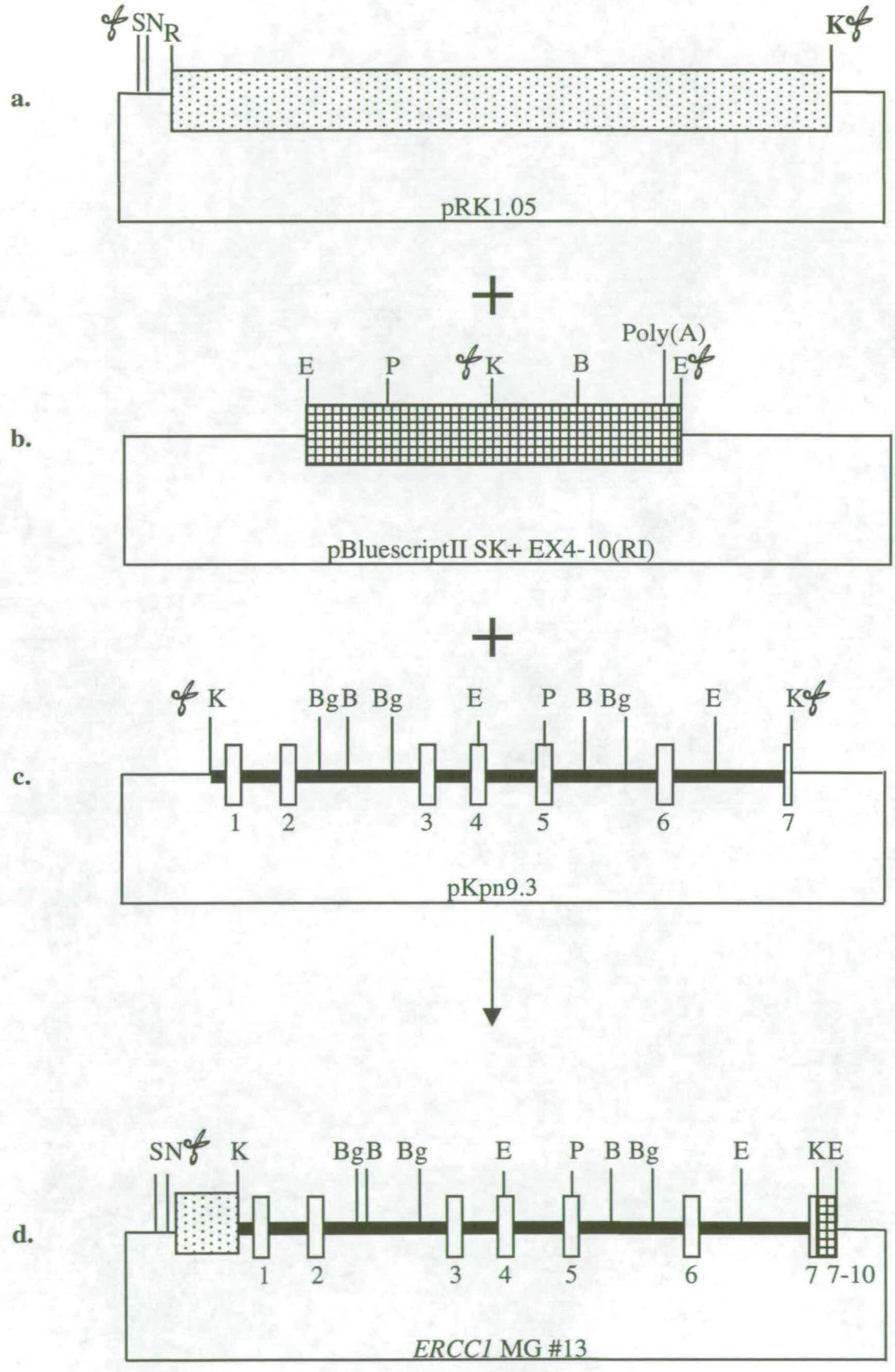
4.2 Expression of the *ERCC1* minigene (*ERCC1 MG #13*) in DNA repair deficient mouse embryonic fibroblasts (PF24)

As described earlier, the spontaneously transformed embryonic fibroblast cell line (PF24) was derived from an *ERCC1* deficient embryo and has been shown to be hypersensitive to UV as a consequence of DNA repair deficiency. It has been shown that the pattern of *ERCC1* expression observed in wild type embryonic fibroblasts (PF20) matches the pattern observed in skin and keratinocytes. As such the PF24 cell line was regarded as suitable for the *in vitro* study of the expression pattern of the minigene. Northern blot analysis after transfection of the PF24 cell line with the minigene construct would enable us to determine if all the sequence elements necessary for the skin specific pattern of expression were present within the

Figure 4.2 Construction of the *ERCC1* minigene (*ERCC1* MG #13)

a, The 5' flanking sequence clone pRK 1.05 was cut with SstI and KpnI to liberate the insert. Drawn to scale: 1cm=100bp. b, The exon 7 to 10 region of the cDNA subclone pBluescriptII SK+ EX4-10 (RI), was isolated as a 300bp KpnI/EcoRI fragment. Drawn to scale: 1cm=100bp. c, The 9.3kb KpnI genomic fragment carrying exons 1 to 7 was isolated from pKpn 9.3. Drawn to scale: 1cm=1kb. d, Each of the isolated fragments were cloned into SstI/ EcoRI cut pBluescriptII SK+. Drawn to scale 1cm=1kb.

Plasmid backbone shown as thin line. 5'Flanking sequence shown as stippled box. cDNA sequence shown as hatched box. Exons shown as numbered open boxes. Introns shown as thick line. Restriction sites used for cloning and subsequent analysis shown as follows: EcoRI; E, BamHI; B, BglII; Bg, KpnI; K, PstI; P, SstI; S, NotI; N. Restriction sites cut for cloning or later linearising are indicated by ✂.


a.
SNR
K
pRK1.05
+
b.
Poly(A)
E
P
K
B
E
pBluescriptII SK+ EX4-10(RI)
+
c.
K
BgB
Bg
E
P
B
Bg
E
K
1
2
3
4
5
6
7
pKpn9.3
d.
SN
K
BgB
Bg
E
P
B
Bg
E
KE
1
2
3
4
5
6
7
7-10
ERCC1 MG #13

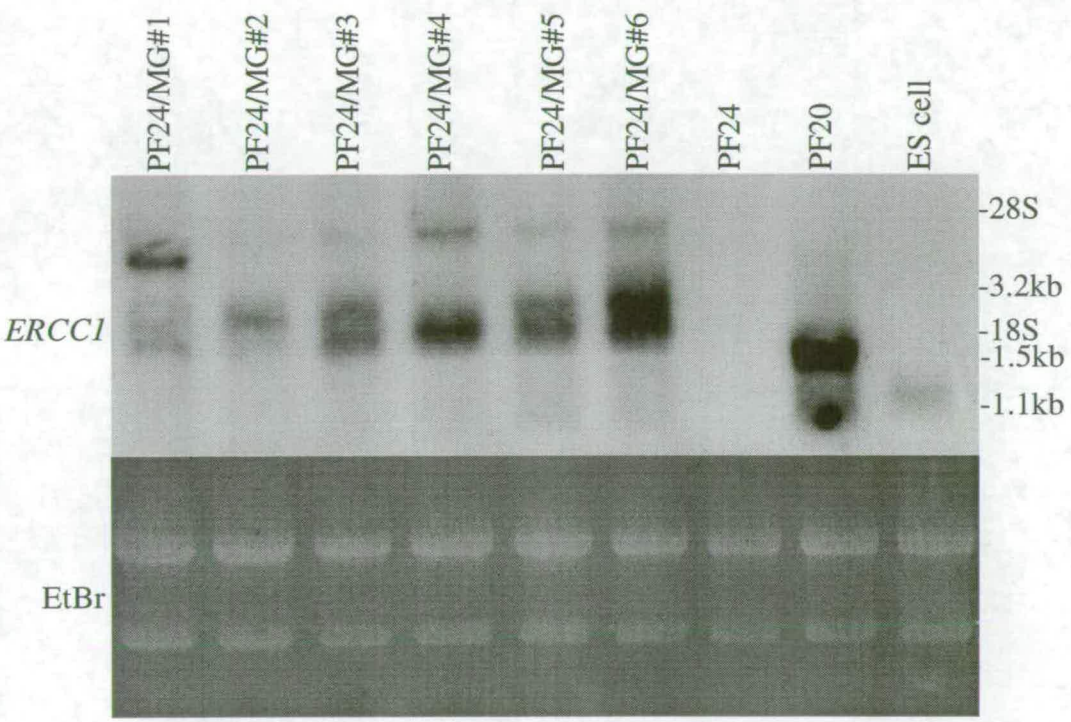
minigene. This was essential before any deletional analysis of the promoter could be carried out.

The original *ERCC1* gene targeting took place in the *HPRT*-deficient ES cell line, HM1 (Selfridge *et al.,* 1992). Consequently, a proportion of the *ERCC1*-deficient animals generated from the targeted ES cell line also carried the X-linked, non-functioning *HPRT* allele from the HM1 cell line. It has been established that the ERCC1-deficient PF24 cell line was derived from an *HPRT*-deficient male embryo (Melton unpublished data) a factor which was exploited in the gene transfer experiments described here. *HPRT* catalyses an early step in the purine salvage pathway in mammals. It is possible to select for HPRT activity *in vitro,* using medium supplemented with hypoxanthine, aminopterin and thymidine (HAT). As aminopterin inhibits *de novo* synthesis of purines and pyrimidines, *HPRT* activity is required for cells to bypass the aminopterin-induced metabolic block by utilising the hypoxanthine and thymidine in the purine and pyrimidine salvage pathways. This enables *HPRT* minigenes to be used as selectable markers.

PF24 cells were co-transfected with the *ERCC1* MG #13 and the *PGK* driven *HPRT* minigene, *PGKHPRT,* in a 3:1 ratio, respectively. Both vectors were linearised with NotI prior to transfection. Stably transfected cells were identified by culturing in HAT selective medium. Individual colonies were then picked and screened by PCR for the presence of the *ERCC1* minigene. Details of the PCR assay used to distinguish between wild type and targeted alleles are presented in Figure 6.2 (Chapter 6). The PCR primers 033M and 035M give rise to a 600bp product from exons 4 to 5 of the wild type allele and fail to generate a product from the targeted allele. Using this primer pair when screening transfected PF24 cells can only give rise to a PCR product if the *ERCC1* minigene is present; exons 4 and 5 are present within the minigene in their wild type configuration. Out of twelve colonies screened, all were found to be positive for the presence of the *ERCC1* minigene (data not shown). Six of the twelve positive clones were expanded in culture for northern blot analysis (Figure 4.3).

Figure 4.3 *ERCC1* minigene (ERCC1 MG #13) expression in repair-deficient embryonic fibroblasts (PF24)

Northern blot analysis of RNA from minigene-positive PF24 cell lines. Upper panel, total RNA was prepared from the minigene transfected cell lines PF24/MG#2, PF24/MG#4 and PF24/MG#6, a wild type mouse embryonic fibroblast cell line (PF20), an *ERCC1*-deficient mouse embryonic fibroblast cell line (PF24) and a mouse ES cell line (HM1). The RNA (30μg) was electrophoresed on a 1.4% formaldehyde-agarose gel, transferred onto a nylon membrane and probed with an 800bp BamHI fragment of the *ERCC1* cDNA corresponding to exons 1 to 8. Positions of the 18S and 28S rRNAs and estimated transcript sizes are indicated. Lower panel, ethidium bromide staining to illustrate RNA loads.


PF24/MG#1
PF24/MG#2
PF24/MG#3
PF24/MG#4
PF24/MG#5
PF24/MG#6
PF24
PF20
ES cell
ERCC1
-28S
-3.2kb
-18S
-1.5kb
-1.1kb
EtBr

Total RNA was isolated from six minigene positive PF24 cell lines, PF24/MG#1-#6, the wild type PF20 cell line, the parental PF24 cell line and the HM1 ES cell line. The RNA (30μg) was separated by gel electrophoresis, transferred onto a nylon membrane and *ERCC1* transcripts detected by hybridisation to an *ERCC1* cDNA probe.

The ethidium bromide staining of the gel is presented in the lower panel and indicates that the amounts of RNA in each lane are, by eye, equivalent. Only the normal 1.1kb transcript was present in HM1 lane, whereas the skin specific pattern of expression was evident in the PF20 lane. Once again, as described in chapter 3.2, there was no evidence of a transcript from the targeted allele in the PF24 lane, perhaps due to partial degradation of the RNA sample preventing the detection of higher molecular weight species present in low abundance. Multiple transcripts were evident in each of the minigene positive lanes, however, the predicted skin specific pattern of expression, the presence of an abundant 1.5kb transcript, was not readily detected. In each of the lanes there was a suggestion of weak signals from hybridising mRNA at the predicted sizes of 1.1kb and 1.5kb, but not in the proportions normally detected in skin or the PF20 cell line. In addition to these weakly hybridising bands there were much stronger signals of higher molecular weight evident in each of the lanes. In lanes 1, 3, 5, and 6 there were two additional strongly hybridising bands, estimated to be in the region of 3.2kb, present in equal proportions to each other. In lane 4 only the smaller of the two additional transcripts was evident. In lane 2 the larger of the two additional bands was clearly detected and there was evidence of hybridisation to a band equal in size to the smaller additional band. Although these transcripts do not fit the size predictions, the pattern of expression evident in lane two, a higher molecular weight species present in greater abundance than the lower molecular weight species, is reminiscent of the skin specific pattern of expression.

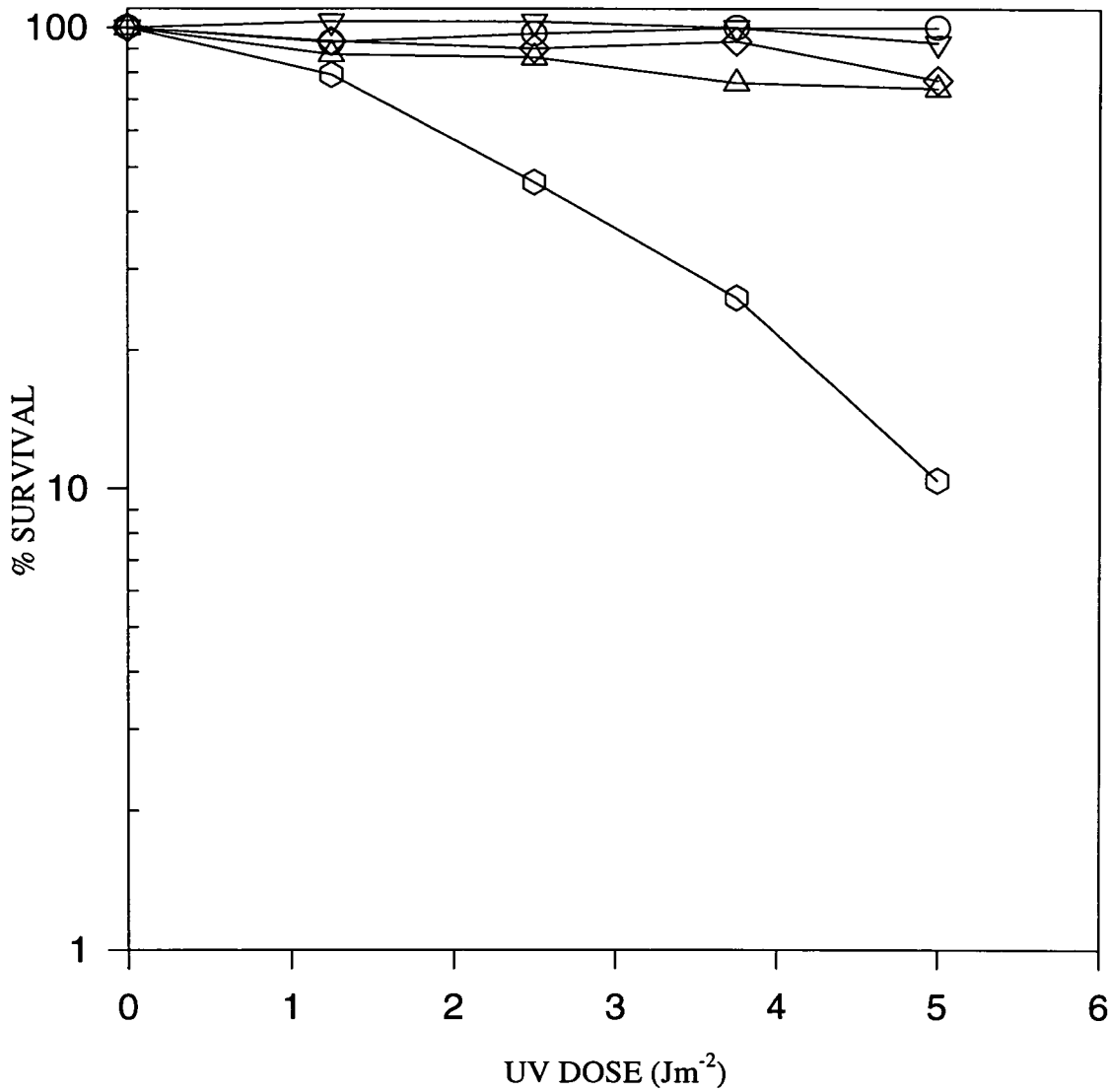
4.3 Expression of the *ERCC1* minigene corrects the UV sensitivity of *ERCC1*-deficient fibroblasts (PF24)

The northern blot analysis described above confirmed that the *ERCC1* minigene was being transcribed, but the transcripts detected did not fit the predicted sizes and

aberrant processing of the mRNA could not be ruled out at this stage. It was therefore necessary to establish that the minigene transcripts gave rise to fully functional ERCC1 protein. Rescue of the PF24 UV sensitivity following transfection of the cells with the *ERCC1* minigene would confirm the functionality of the minigene construct. Three of the minigene positive cell lines were expanded in culture for use in UV sensitivity experiments. The selected lines, PF24/MG#2, PF24/MG#4 and PF24/MG#6 all displayed different patterns of minigene expression at the level of northern blot analysis (see Figure 4.3). They were selected to investigate whether there was any correlation with any of the patterns and gene function. For the UV experiments, $2x10^4$ cells were plated in small dishes and cultured overnight before being irradiated with various doses of UV irradiation. The dishes were maintained in culture until control dishes, not exposed to UV, reached confluence. At this point the cells were fixed and stained to quantify the UV sensitivity of the minigene positive cell lines. The UV survival curves obtained are presented in Figure 4.4. The graph illustrates the survival in terms of the cells surviving each dose of UV irradiation as a percentage of the untreated controls. The graph shows that the minigene positive PF24 cell lines were not sensitive to UV. The survival profile of the UV sensitive, non-transfected PF24 cells showed that 79% of the cells survived the UV dose of 1.25 Jm^{-2}, 46% survived the 2.5 Jm^{-2} dose, 26 % survived the 3.75 Jm^{-2}, and 10 % survived the highest dose of 5 Jm^{-2}. In the case of the wild type PF20 cell line, the survival profile indicates that 100% of the cells survived the highest dose of 10 Jm^{-2}. The survival profile of the cell line PF24/MG #4 only showed a slight degree of UV sensitivity at the highest UV dose, with 93% survival at the 5 Jm^{-2} dose. The PF24/MG#6 cell line showed a similar resistance to the UV irradiation with 94 and 90% survival at the lower UV doses of 1.25 and 2.5 Jm^{-2}. At the higher doses of 3.75 and 5 Jm^{-2} survival of 94 and 77% was observed. The PF24/MG#2 cells survival profile showed that at the lower doses of 1.25 and 2.5 Jm^{-2} the survival was 88 and 86%, whilst at the higher doses of 3.75 and 5 Jm^{-2} the survival was measured at 76 and 73%.

Figure 4.4 Correction of the UV sensitivity of *ERCC1*-deficient fibroblasts (PF24) by following transfection with an *ERCC1* minigene (*ERCC1* MG #13)

The UV survival curves for wild type (PF20), *ERCC1*-deficient (PF24) mouse fibroblasts and three minigene-positive PF24 cell lines are shown. Each point represents the mean of two separate experiments. Circle, PF20; hexagon, PF24; triangle, PF24/MG#2; inverted triangle, PF24/MG#4; diamond, PF24/MG#6.


100
10
1
% SURVIVAL
0
1
2
3
4
5
6
UV DOSE (Jm-2)

Thus the expression of the minigene in each of the cell lines analysed was seen to correct the UV sensitivity of the repair deficient PF24 cells to near wild type levels. There were no significant differences between the survival profiles of the minigene positive cell lines and as such it was not possible to consider any correlation between any of the observed patterns of minigene expression and gene function.

4.4 Expression of the *ERCC1* minigene (*ERCC1* MG #13) in *ERCC1*-deficient Chinese hamster ovary cells (CHO43-3B)

In addition to using the *ERCC1* minigene in experiments designed to identify the sequences required for the skin specific pattern of expression it was also important to consider whether the minigene was expressed appropriately in other cell types. The cell types in which only the 1.1kb transcript is expressed include mouse ES cells and CHO cells. As *ERCC1*-deficient mouse ES cells were not available it would not have been possible to assess whether expression of the minigene in ES cells gave rise to a functional protein. An ERCC1-deficient CHO cell line was available for such analysis to be carried out. The aim of the work described in this section was to determine if, when expressed in CHO cells, the minigene would only give rise to the normal 1.1kb transcript.

The human *ERCC1* gene was originally cloned by virtue of its ability to complement the UV sensitive CHO cell line, CHO43-3B. The *ERCC1*-deficient status provided an ideal environment to test whether the mouse minigene was capable rescuing the UV sensitivity in another species. Recently it was shown that the molecular basis of the *ERCC1* defect in CHO43-3B cells was a point mutation resulting in an amino acid exchange at the 98th residue. Although the cells are ERCC1-deficient, expression of the gene gives rise to a transcript of normal size and at wild type levels (Hayashi *et al*, 1998). The fact that the mouse *ERCC1* cDNA probes hybridised with the hamster transcript meant that it would be less straightforward to detect expression of the minigene where the endogenous transcript was still present. It was felt that, as transgenes are generally present in multiple copies, an apparent increased abundance of the 1.1kb transcript might be sufficient to indicate minigene expression in the first instance. A secondary method for confirming transgene expression over the

endogenous gene would then be appropriate. The DNA mediated gene transfer was carried out using a spontaneously occurring *HPRT*-deficient clone, CHO43-3B(TG1) derived from the CHO43-3B cell line (Melton *et al*, 1998). This enabled an *HPRT* minigene to be used as a selectable marker in a co-transfection with the *ERCC1* minigene. A pilot study showed that the PCR assay used for the detection of the minigene, the 600bp exon 4 to 5 product, was specific for the mouse sequences of the minigene and did not give a positive result for CHO43-3B DNA (data not shown).

CHO43-3B(TG1) cells were co-transfected with the *ERCC1* MG #13 and the *PGK* driven *HPRT* minigene, *PGKHPRT*. Both vectors were linearised with *NotI* prior to transfection. Stably transfected cells were identified by culturing in selective HAT medium. Individual colonies were then picked and screened by PCR for the presence of the *ERCC1* minigene. Out of twelve colonies screened, all were found to be positive for the presence of the *ERCC1* minigene (data not shown). Six of the twelve positive clones were expanded in culture for northern blot analysis (Figure 4.5).

Total RNA was isolated from six minigene positive CHO43-3B(TG1) cell lines, CHO43-3B(TG1) /MG#2, 4, 6, 7, 10 and 11, the wild type CHO-9 cell line and the parental CHO43-3B cell line. The RNA (30µg) was separated by gel electrophoresis, transferred onto a nylon membrane and *ERCC1* transcripts detected by hybridisation to a mouse *ERCC1* cDNA probe.

The ethidium bromide staining is presented in the lower panel and suggests that, by eye, the two control tracks CHO-9 and CHO43-3B are overloaded relative to the other samples. The detection of *ERCC1* mRNA by probing with the mouse *ERCC1* cDNA is shown in the upper panel. The normal 1.1kb transcript was evident in every lane. Allowing for the apparent overloading of the two control tracks there was no evidence to suggest that the level of this transcript was elevated in any of the minigene positive cell lines. There were no additional distinct transcripts evident in lanes CHO43-3B(TG1)/MG#2, 6, 10 and 11. In the CHO43-3B(TG1)/MG#7 and 4 lanes an additional transcript, estimated at 1.5kb, was discernible. In summary, despite each of the cell lines being minigene positive, as determined by PCR, it appeared that the mouse *ERCC1* minigene was not expressed in all of the CHO cell lines.

Figure 4.5 *ERCC1* minigene (ERCC1 MG #13) expression in repair-deficient Chinese hamster ovary cells (CHO43-3B(TG1))

Northern blot analysis of RNA from minigene-positive CHO43-3B(TG1) cell lines. Upper panel, total RNA was prepared from the minigene transfected cell lines CHO43-3B(TG1)/MG#2, 4, 6, 7, 10, 11, a wild type Chinese hamster ovary cell line (CHO9), and the parental *ERCC1*-deficient Chinese hamster ovary cell line (CHO43-3B(TG1)) The RNA (30μg) was electrophoresed on a 1.4% formaldehyde-agarose gel, transferred onto a nylon membrane and probed with an 800bp BamHI fragment of the *ERCC1* cDNA corresponding to exons 1 to 8. Positions of the 18S and 28S rRNAs and estimated sizes of distinct transcripts are indicated. Lower panel, ethidium bromide staining to illustrate RNA loads.

CHO9
CHO43-3B(TG1)
CHO43-3B(TG1)/MG#2
CHO43-3B(TG1)/MG#4
CHO43-3B(TG1)/MG#6
CHO43-3B(TG1)/MG#7
CHO43-3B(TG1)/MG#10
CHO43-3B(TG1)/MG#11
ERCC1
-28S
-18S
-1.5kb
-1.1kb
EtBR

4.5 The UV sensitivity of ERCC1-deficient CHO cells (CHO43-3B) is corrected by expression of the mouse *ERCC1* minigene (*ERCC1* MG#13)

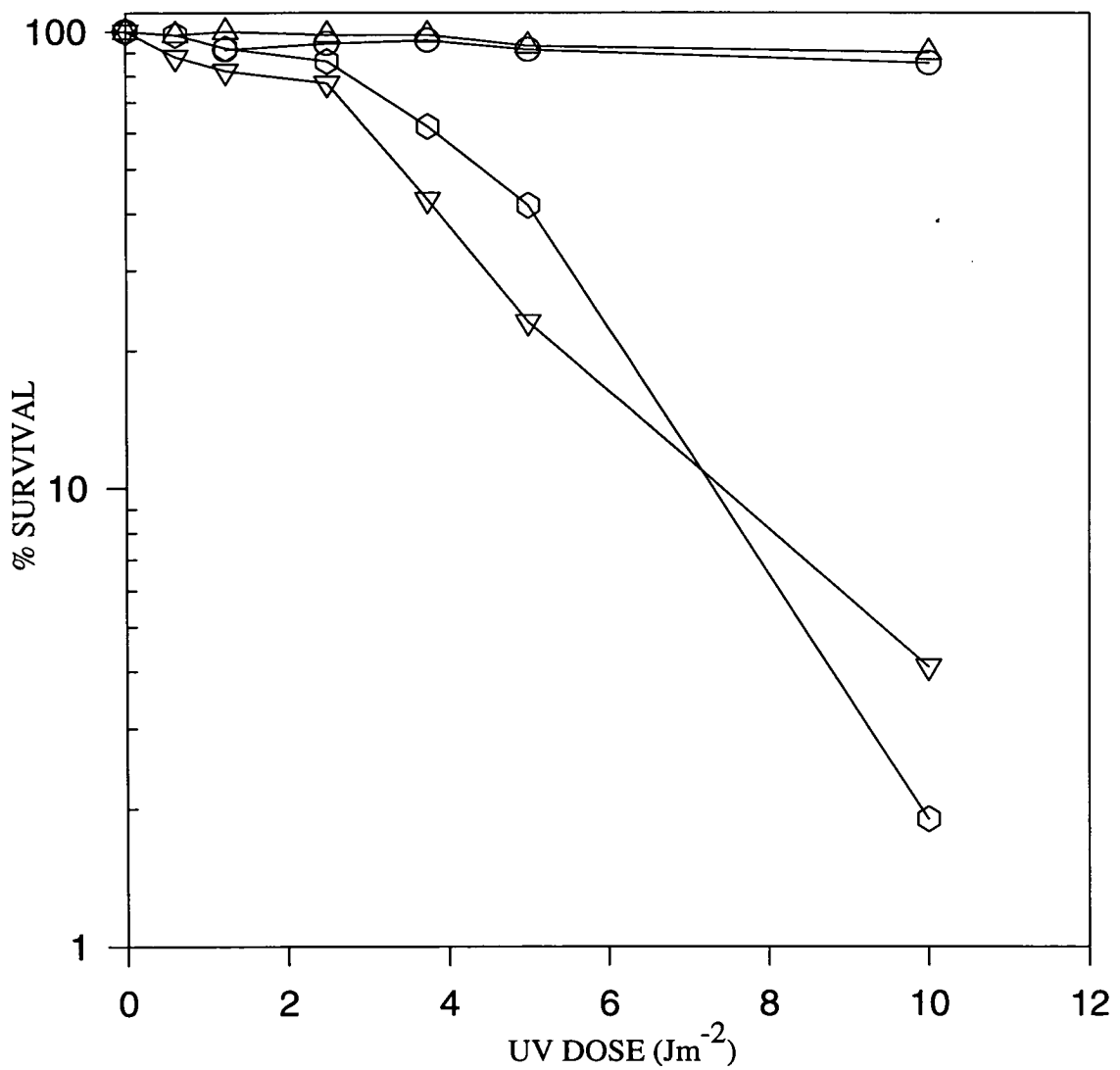
The northern blot analysis described above, suggested that the mouse *ERCC1* minigene was not expressed in the majority of the cell lines tested. Expression was detected in two cell lines by virtue of the transcript being different in size from the endogenous hamster transcript. It remained possible that the minigene was indeed being expressed appropriately, at low levels, in the other cell lines and that the presence of the endogenous hamster transcript prevented it from being detected. If this was the case, and assuming that the minigene was producing a functional protein, it was predicted that rescue of the CHO43-3B UV sensitivity would be evident in those cell lines where the minigene was being expressed in the manner suggested. To investigate further whether there was rescue of UV sensitivity, two of the cell lines were expanded in culture for use in UV sensitivity experiments. The cell lines selected were, CHO43-3B(TG1)/MG#2 which did not appear to be expressing the minigene at the level of northern blot analysis and CHO43-3B(TG1)/MG#7 in which an additional transcript was detected..

For the UV experiments, $2x10^4$ cells were plated in small dishes and cultured overnight before being irradiated with various doses of UV irradiation. The dishes were maintained in culture until control dishes, not exposed to UV, reached confluence. At this point the cells were fixed and stained to quantify the UV sensitivity of the minigene positive cell lines. The UV survival curves obtained are presented in Figure 4.6. The graph illustrates the survival in terms of the cells surviving each dose of UV irradiation as a percentage of the untreated controls

The UV survival curve of the CHO43-3B cell line illustrated its sensitivity to UV. The curve showed that although 92% of cells survived the 1.25 Jm^{-2} dose, survival was reduced to 61 and 41 % at the higher UV doses of 3.75 and 5 Jm^{-2} and only 2% survived a UV dose of 10 Jm^{-2}. The CHO43-3B(TG1)/MG#2 cell line exhibited a comparable degree of UV sensitivity. In this case 88% of cells survived the 1.25 Jm^{-2} dose, survival was reduced to 43 and 23% at the higher UV doses of 3.75 and 5 Jm^{-2} and 4% of the cells survived the 10 Jm^{-2} dose. The CHO43-3B(TG1)/MG#7 cell line did not appear to be sensitive to UV. In this case there was no cell death after the

Figure 4.6 Correction of the UV sensitivity of *ERCC1*-deficient Chinese hamster ovary cells (CHO43-3B(TG1)) following transfection with an *ERCC1* minigene (*ERCC1* MG #13)

The UV survival curves for wild type (CHO9), *ERCC1*-deficient (CHO43-3B(TG1)) Chinese hamster ovary cells and two minigene-positive CHO43-3B(TG1)cell lines are shown. Each point represents the mean of two separate experiments. Circle, CHO9; hexagon, CHO43-3B(TG1); triangle, CHO43-3B(TG1)/MG#7; inverted triangle, CHO43-3B(TG1)/MG#2.


100
10
1
% SURVIVAL
0
2
4
6
8
10
12
UV DOSE (Jm-2)

1.25 Jm^{-2} dose, survival at the higher UV doses of 3.75 and 5 Jm^{-2} was 98 and 93% and 90% of the cells survived the highest UV dose of 10 Jm^{-2}. This is in keeping with the survival curve of the wild type CHO-9 cells, which showed 91% survival at the 1.25 Jm^{-2} dose, 95 and 91 % survival at the 3.75 and 5 Jm^{-2} doses and 85% survival at the highest dose of 10 Jm^{-2}.

In summary, in the cell line where expression of the minigene was apparent, the UV sensitivity associated with the DNA repair deficient CHO cells, CHO43-3B, was rescued. The cell line in which only the endogenous transcript was evident remained sensitive to UV irradiation.

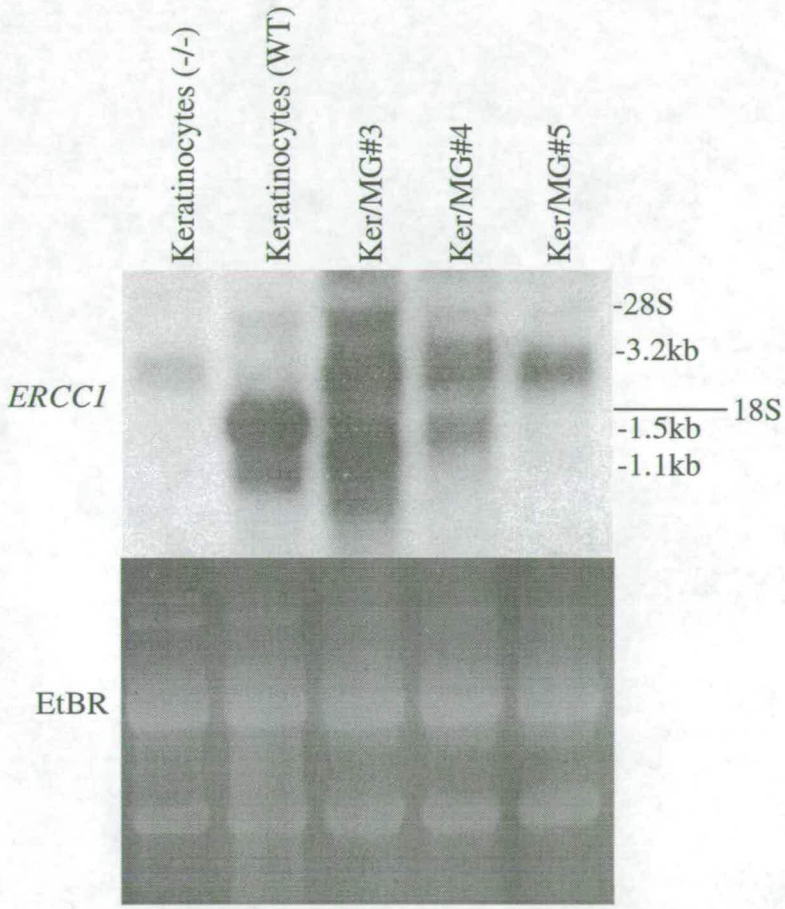
4.6 Expression of the *ERCC1* minigene (*ERCC1* MG #13) in cultured *ERCC1*-deficient mouse keratinocytes

We have already shown that the pattern of *ERCC1* expression observed in cultured mouse keratinocytes matched the pattern of expression seen in skin (Chapter 3). The major transcript in keratinocytes is the skin specific 1.5kb transcript. As such, keratinocytes were considered as an appropriate cell type to investigate further the elements involved in the control of this expression pattern using the *ERCC1* minigene. As with the PF24 cell line, the *ERCC1*-deficient keratinocytes (Ker.(-/-)) isolated in our laboratory were found to be *HPRT*-deficient (Melton, unpublished data). Once again this permitted the use of an *HPRT* minigene as a selectable marker in co-transfections with the *ERCC1* minigene (*ERCC1*/MG#13). Due to the technical difficulty associated with the successful culturing of mouse keratinocytes, all the culturing required for these experiments was carried out by Ann-Marie Ketchen.

ERCC1-deficient keratinocytes were co-transfected with the *ERCC1* MG #13 and the *PGK* driven *HPRT* minigene, *PGKHPRT*. Both vectors were linearised with NotI prior to transfection. Stably transfected cells were identified by culturing in selective HAT medium. Individual colonies were then picked and screened by PCR for the presence of the *ERCC1* minigene, as described in section 4.2. Out of six colonies screened, 3 were identified as *ERCC1* minigene positive and were expanded in culture for northern blot analysis (Figure 4.7)

Figure 4.7 *ERCC1* minigene (ERCC1 MG #13) expression in repair-deficient murine keratinocytes (Ker.(-/-))

Northern blot analysis of RNA from minigene-positive Ker(-/-) cell lines. Upper panel, total RNA was prepared from the minigene transfected cell lines Ker.(-/-) /MG#3, 4 and 5, a wild type murine keratinocyte cell line(Ker.(WT)), and the parental *ERCC1*-deficient murine keratinocyte cell line (Ker.(-/-)) The RNA (30μg) was electrophoresed on a 1.4% formaldehyde-agarose gel, transferred onto a nylon membrane and probed with an 800bp BamHI fragment of the *ERCC1* cDNA corresponding to exons 1 to 8. Positions of the 18S and 28S rRNAs and estimated sizes of distinct transcripts are indicated. Lower panel, ethidium bromide staining to illustrate RNA loads


Keratinocytes (-/-)
Keratinocytes (WT)
Ker/MG#3
Ker/MG#4
Ker/MG#5
ERCC1
-28S
-3.2kb
18S
-1.5kb
-1.1kb
EtBR

Total RNA was isolated from three minigene positive Ker.(-/-) cell lines, Ker.(-/-) /MG#3, #4 and #5, the wild type keratinocyte cell line and the parental *ERCC1*-deficient keratinocytes. The RNA (30μg) was separated by gel electrophoresis, transferred onto a nylon membrane and *ERCC1* transcripts detected by hybridisation to a mouse *ERCC1* cDNA probe.

The ethidium bromide staining of the gel is presented in the lower panel and indicates that the amounts of RNA in each lane are, by eye, essentially equivalent. Transcripts arising from the targeted *ERCC1* alleles were evident in the Ker.(-/-) lane, the size estimated for these transcripts was 3.2kb. In the wild type keratinocyte lane there were bands evident at 1.5kb and 1.1kb as expected. The expression patterns seen in the minigene positive lanes did not reflect the wild type pattern of expression. There was a band of approximately 3.2kb evident in each of the minigene positive lanes. Whilst the size of this band matches the size of the transcript from the *ERCC1* targeted alleles, the signals were noticeably stronger in the minigene positive lanes and as the ethidium staining of the gel did not suggest that these lanes were overloaded relative to the Ker. (-/-) lane, it is probable that the bands corresponded to minigene transcripts. A further transcript of approximately 1.5kb was evident in the Ker(-/-)/MG#4 lane, this fitted the predicted size for the skin specific transcript. There was no strong evidence of a hybridising species at 1.1kb in this lane and as such the normal pattern of *ERCC1* expression in keratinocytes was not obvious. In the Ker.(-/-)/MG#3 lane there were also additional transcripts evident, in this case none of the transcripts fitted the size predictions for correct expression of the minigene.

In summary, although *ERCC1* minigene expression was clearly detectable, the patterns of minigene expression seen in cultured keratinocytes did not reflect the normal pattern of *ERCC1* expression associated with this cell type.

4.7 The UV sensitivity of cultured ERCC1-deficient mouse keratinocytes is corrected by expression of the mouse *ERCC1* minigene (*ERCC1* MG#13)

The northern blot analysis described above, suggested that the minigene was being expressed in cultured mouse keratinocytes. However, as the transcripts did not fit the

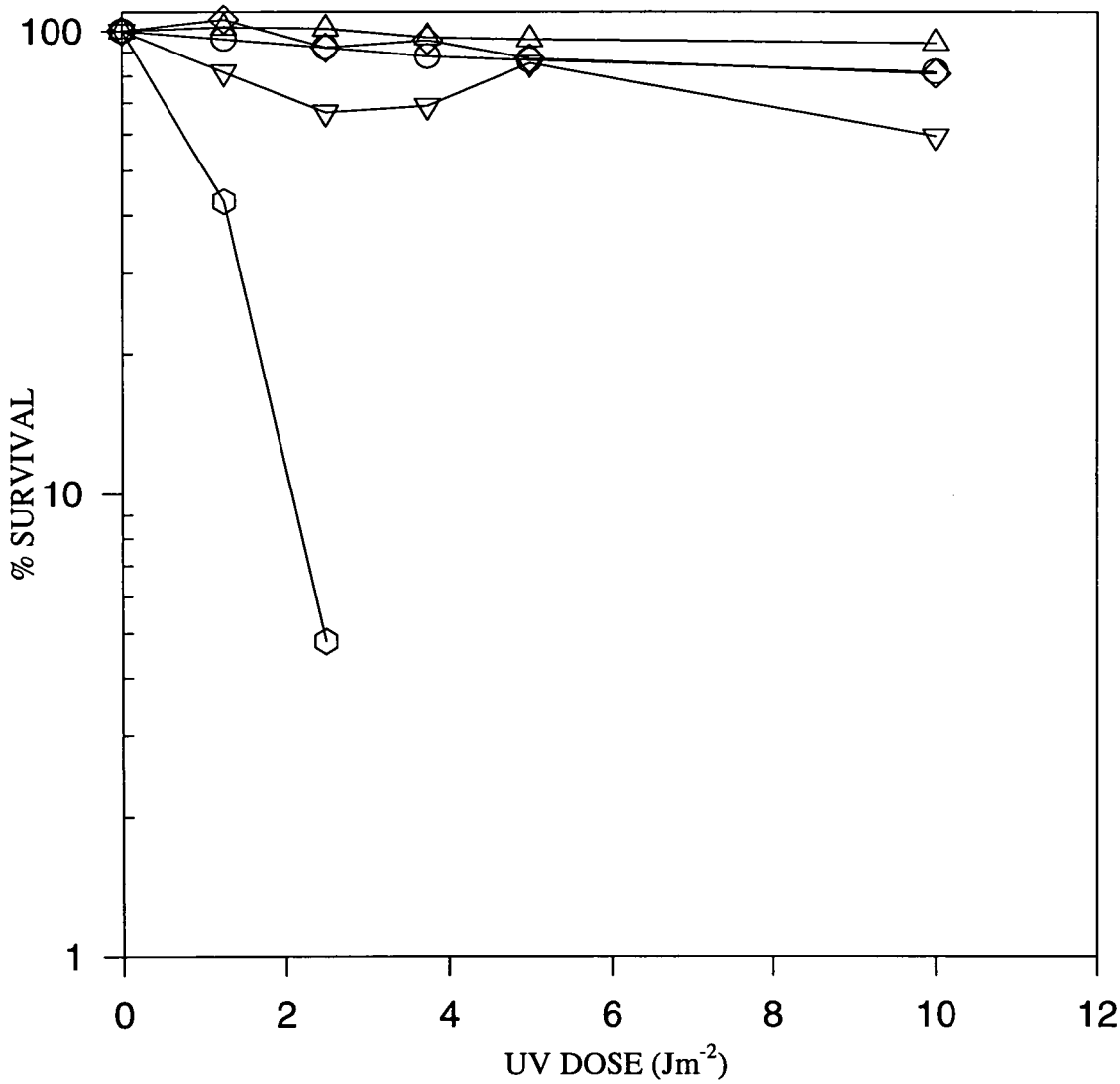
predicted sizes it was possible that the mRNAs detected, resulted from aberrant splicing of the minigene introns and as such would not be able to encode a fully functional protein. Rescue of the UV sensitivity associated with the Ker.(-/-) cell line would be indicative of the functionality of the minigene construct in keratinocytes. The keratinocyte UV survival assays were performed by Ann-Marie Ketchen.

The three minigene positive cell lines were expanded in culture for use in UV sensitivity experiments. For the UV survival experiments, $2x10^4$ cells were plated in small dishes and cultured overnight before being irradiated with various doses of UV irradiation. The dishes were maintained in culture until control dishes, not exposed to UV, reached confluence. At this point the cells were fixed and stained to quantify the UV sensitivity of the minigene positive cell lines. The UV survival curves obtained are presented in Figure 4.8

The graph illustrates the survival in terms of the cells surviving each dose of UV irradiation as a percentage of the untreated controls. The UV sensitivity of the *ERCC1*-deficient cell line(Ker.(-/-)) was illustrated by its survival curve. The curve showed 43 and 5% survival after UV doses of 1.25 and 2.5 Jm^{-2}. The wild type keratinocytes were not sensitive to these UV doses. The survival curve showed 96 and 92% survival following UV doses of 1.25 and 2.5 Jm^{-2} and 88, 86% survival after UV doses of 3.75 and 5 JM^{-2} and 82% survival following the highest UV dose of 10 Jm^{-2}. Each of the minigene positive cell lines had survival profiles similar to the wild type pattern. In the case of the Ker(-/-)/MG#3 cell line the survival profile showed 100% survival following UV doses of 1.25 and 2.5 Jm^{-2} and 97, 96% survival after UV doses of 3.75 and 5 JM^{-2} and 94% survival following the highest UV dose of 10 Jm^{-2}. The survival curve of the Ker.(-/-)/MG#4 cell line revealed 81 and 67% survival following UV doses of 1.25 and 2.5 Jm^{-2} and 69, 85% survival after UV doses of 3.75 and 5 JM^{-2} and 60% survival following the highest UV dose of 10 Jm^{-2}. Finally, the survival profile of the Ker.(-/-)/MG#5 cell line displayed 100 and 92% survival following UV doses of 1.25 and 2.5 Jm^{-2} and 95, 87% survival after UV doses of 3.75 and 5 JM^{-2} and 80% survival following the highest UV dose of 10 Jm^{-2}.

Figure 4.8 Correction of the UV sensitivity of *ERCC1*-deficient murine keratinocytes (Ker.(-/-)) following transfection with an *ERCC1* minigene (*ERCC1* MG #13)

The UV survival curves for wild type (Ker.(WT)), *ERCC1*-deficient (Ker.(-/-)) murine keratinocytes and three minigene-positive Ker. (-/-)cell lines are shown. Each point represents the mean of two separate experiments. Circle, Ker.(WT); hexagon, Ker.(-/-); triangle, Ker.(-/-)/MG#3; inverted triangle, Ker.(-/-)/MG#4; diamond, Ker.(-/-)/MG#5.


100
10
1
% SURVIVAL
0
2
4
6
8
10
12
UV DOSE (Jm-2)

In summary, it was apparent that the minigene positive cell lines, identified by PCR, were not sensitive to UV irradiation. The minigene transcripts detected by northern blot analysis were thus capable of encoding an ERCC1 protein capable of fulfilling its role in NER.

4.8 Transcription of the *ERCC1* minigene initiates at the upstream start site

The northern blot analysis carried out on each of the transfected cell lines has indicated that the *ERCC1* minigene is expressed in each of the different cell types. Although the UV survival data indicated that the transcripts were functional, in each case the resultant transcripts have not fitted with predicted sizes and the molecular basis of these size differences has not been determined. As a result it has not been possible to determine, from these blots, whether the upstream transcription initiation site, in the minigene construct, was being used. The use of the upstream transcriptional start site would be confirmed where any of the transcripts identified in the transfected cell lines hybridised to the 430bp SalI/BfaI 5' RACE probe (chapter 3.8). This probe is specific for the transcripts initiating at the upstream start site.

Total RNA (30μg) from selected minigene positive cell lines from each cell type was separated by electrophoresis, transferred onto a nylon membrane and *ERCC1* transcripts detected by hybridisation to the SalI/BfaI 5' RACE probe (Figure 4.9). The filter was then stripped and reprobed using a mouse *ERCC1* cDNA fragment spanning exons 1 to 8.

The aim of this analysis was to demonstrate that amongst the various minigene transcripts detected by the earlier northern blots, some were the result of transcription initiating at the upstream start site due to the inclusion of distal promoter elements in the minigene construct. By a comparison of which bands were detected using the 5' end probe and then the cDNA probe it was possible to identify which of these transcripts, if any, were the result of transcription from the upstream start site. The upper panel shows the filter probed with the exon 1 to 8 cDNA probe that hybridises to all *ERCC1* transcripts, irrespective of the site of transcriptional initiation. The lower panel shows the filter probed with the fragment specific for

Figure 4.9 Identification of minigene transcripts initiating at the upstream start site

Northern blot analysis of RNA from selected minigene-positive cell lines. Lower panel, total RNA (30μg) from wild type murine keratinocytes (Ker.(WT)), *ERCC1*-deficient keratinocytes (Ker.(-/-)), minigene positive keratinocyte cell lines (Ker.(-/-) /MG#3, 4 and 5), minigene positive embryonic fibroblast cell lines (PF24/MG#2, 4 and 6) and minigene positive Chinese hamster ovary cell lines (CHO43-3B(TG1)/MG#7 and 2) was electrophoresed on a 1.4% formaldehyde-agarose gel, transferred onto a nylon membrane and probed with a 430bp SalI/BfaI 5' RACE fragment. Upper panel, the filter was then stripped and reprobed using an 800bpBamHI fragment of the *ERCC1* cDNA corresponding to exons 1 to 8. Positions of the 18S and 28S rRNAs and estimated sizes of transcripts are indicated.

Keratinocytes(-/-)
Keratinocytes(WT)
Ker/MG#3
Ker/MG#4
Ker/MG#5
PF24/MG#2
PF24/MG#4
PF24/MG#6
CHO43-3B(TG1)/MG#7
CH043-3B(TG1)/MG#2
cDNA
-28S
-3.2kb
-1.5kb
-1.1kb
5' End
-28S
-3.2kb
-1.5kb
-1.1kb

transcripts initiating at the upstream start site. In the null keratinocyte lane transcripts arising from the targeted alleles were evident in both cases. In the case of the wild type keratinocytes only the 1.5kb transcript was detected when probed with the 5' end probe, whilst the 1.1kb transcript was also detected by the cDNA probe. In the minigene positive keratinocyte lanes a single distinct band of approximately 3.2kb was evident in each of the lanes when probed with the 5' end probe. When probed with the cDNA probe there were additional smaller transcripts evident in the Ker.(-/-)/MG#3 and #4 lanes. Using the exon 1 to 8 probe, the minigene positive PF24 cell lines showed a complex pattern of transcripts involving two bands at approximately 2.8 and 3.2kb. In the PF24/MG#6 cell line both of these bands are evident, in PF24/MG#4 only the 2.8kb band was detected and in PF24/MG#2 only the 3.2kb was detected. When probed with the 5' end fragment specific for transcripts from the upstream start site, only the 3.2kb bands in the PF24/MG#2 and #6 lanes were evident, suggesting that this transcript begins at the upstream start site. The 2.8kb transcript, presumably from the downstream start site, was not detected in the PF24/MG#4 and #6 lanes using this probe. The cDNA probe detected only the endogenous 1.1kb transcript in the CHO43-3B(TG1)/MG#2 lane whereas in the CHO43-3B(TG1)/MG#7 lane an additional band of approximately 1.5kb was also evident. When probed using the 5' end probe there was no band evident in the CHO43-3B(TG1)/MG#2 lane, however there was evidence of the band at 1.5kb. This would suggest that the mouse *ERCC1* minigene transcript identified in CHO cells initiated at the upstream start site.

In summary, this analysis has shown that for keratinocytes and mouse embryonic fibroblasts there is evidence to suggest that transcription of the mouse *ERCC1* minigene does initiate at the upstream start site in keeping with the pattern of transcription seen in wild type cells. However the apparent use of this transcriptional start site when the minigene was expressed in CHO43-3B cells does not concur with the normal pattern of *ERCC1* expression in wild type CHO-9 cells. Possible explanations for the aberrant sizes of the minigene transcripts will be discussed later.

4.9 Discussion

The transfection of cloned genes or their minigene derivatives into eukaryotic cells is a powerful and convenient tool for studying the control of gene expression. The use of minigenes is often necessary when the size or structure of a particular gene means that it is not amenable to simple cloning and transfection procedures. The *HPRT* minigene, used as a selectable marker in this chapter, is a good example of this problem. The endogenous mouse *HPRT* locus spans approximately 33kb and would not be suitable for use as a marker gene in that form, whereas, the *PGKHPRT* minigene was only 2.7kb and could be linearised without disruption of the coding sequences. It was deemed necessary to construct a mouse *ERCC1* minigene as the locus spans 15kb and was not available as a single cloned genomic fragment.

The primary goal of the work described in this chapter was to develop a functional *ERCC1* minigene for subsequent use in conventional deletional analysis of the promoter region. These studies are often carried out using the promoter region of the gene of interest coupled to a CAT reporter gene and the level of CAT enzyme activity detected is indicative of promoter activity (Li *et al.*, 1991). However, this approach could not be used to identify the sequences responsible for the skin specific pattern of expression that we have described, as the normal transcript driven by the 170bp promoter element is also present in skin. As such, it was predicted that CAT enzyme activity would continue to be detected from this promoter after the deletion of the upstream promoter. The CAT reporter gene assay could have been employed if the normal 1.1kb transcript was not evident in skin or cell lines displaying the skin specific pattern of *ERCC1* expression.

We have described the construction of a mouse *ERCC1* minigene. This involved combining the cloned 5' flanking sequence, believed to harbour the upstream promoter, with a cloned genomic fragment spanning exons 1 to 7, hooked onto the 3' end of the *ERCC1* cDNA including exons 7 to 10 and a short 3' UTR carrying the polyadenylation signal. The final size of the minigene was 10.6kb compared to 15kb of the endogenous locus. Given the limitations of using RACE analysis, alone, for the definition of the precise transcriptional start site in skin, it remains possible that the

5' flanking sequence included in the mingene did not harbour the complete promoter required for the skin specific expression pattern.

The minigene construct was used for the stable transfection of three different *ERCC1*-deficient cell types including: the *ERCC1*-deficient CHO cell line CHO43-3B as well as *ERCC1*-deficient embryonic fibroblast (PF24) and murine keratinocyte (Ker.(-/-)) cell lines derived from *ERCC1*-deficient mice generated in our laboratory (McWhir *et al.*, 1993). In each case where the minigene was being expressed, as determined by northern blot analysis, we observed complete rescue of the UV sensitive phenotype associated with each cell type. This confirmed that the minigene was capable of complementing the DNA repair defect in each of the cell types tested. The correction of the repair deficiency associated with CHO43-3B using the human and mouse *ERCC1* cDNAs driven by the SV40 early promoter has been reported (van Duin *et al.*, 1986 and 1988). The successful use of an XPA minigene to complement the repair defect in xeroderma group A cells has been demonstrated (Myrand *et al.*, 1996). The 4.2kb *XPA* minigene included 1.6kb of 5' flanking sequence and 0.3kb of 3' flanking sequence. RT-PCR analysis showed that the minigene expression was at wild type levels but northern blot analysis of the minigene transcript was not reported, so it was not clear if the minigene transcripts were of the predicted size.

Although the UV survival assays showed that the minigene was capable of encoding a functional ERCC1 protein, northern blot analysis revealed that, in each of the cell types analysed, the pattern of minigene expression did not reflect the pattern of expression seen with the endogenous gene in wild type cells of each type. The wild type embryonic fibroblasts and keratinocytes both display the skin specific pattern of expression; where the 1.5kb transcript is the major species but the 1.1kb transcript is still clearly detectable. In wild type CHO cells only the 1.1kb transcript is detected. In general, the minigene transcripts, estimated in the region of 3.2kb for the keratinocytes and fibroblasts and 1.5kb for the CHO cells, were larger than the expected 1.1 and 1.5kb. One possible explanation for the increased size of the minigene transcripts was aberrant splicing of introns from the pre-mRNA. This is however unlikely, as such RNAs would not be capable of encoding functional

proteins and it was demonstrated that the minigene did complement the DNA repair defect associated with each of the cell types. It remains possible that correctly spliced transcripts of the predicted sizes were present at such low levels that although they were capable of rescuing the UV sensitivity, they were not detectable by northern blot analysis.

Northern blot analysis using a probe specific for transcripts initiating at the upstream start site suggested that some initiation of minigene transcription was from this site, or at least upstream of the 1.1kb transcript start site. It is unlikely that the increased sizes of the transcripts were the result of transcription initiating from a site even further upstream as this would have extended beyond the 5' end of the minigene construct in order to produce the observed size differential of almost 1.7kb. There have been several reports of minigene expression studies that have used RNase protection assays or primer extension analysis to verify the initiation site of minigene transcription (Kiledjian and Kadesch, 1991 and Prochownik and Orkin, 1984). Such assays could not have been used to map the transcriptional initiation site of the mouse *ERCC1* minigene in the mouse embryonic fibroblasts or murine keratinocytes. As detailed in the introduction, the targeted alleles present in these cells are believed to give rise to non-functional fusion transcripts and as such the 5' ends of the transcripts from these alleles could not be distinguished from the 5' ends of minigene transcripts. Recently, a second *ERCC1*-deficient embryonic fibroblast cell line, which does not produce any transcripts from the targeted alleles, has become available in our laboratory (Hsia, unpublished data). Future *ERCC1* minigene expression studies will be performed using this cell line so that RNase protection and primer extension assays can be performed where necessary. It is possible that such analysis could be applied to the minigene transcripts evident in the CHO cells as the 5' end sequences would vary between the two species. The northern blot analysis of the minigene transcript detected in the transfected CHO cells suggested that the transcript had initiated upstream of the normal initiation site for transcription in CHO cells. RNase protection analysis could be used to verify if the upstream transcriptional initiation site was being used in this instance.

The process of 3' end formation in mRNA molecules involves cleavage of the pre-mRNA and then addition of the poly(A) tail. Studies of Chinese hamster *dihydrofolate reductase (DHFR)* minigenes showed that minigene constructs lacking the three endogenous polyadenylation sites were capable of transferring *DHFR* gene function (Venolia *et al.*, 1987). The discrete minigene mRNAs detected, did not match the sizes of the endogenous transcripts. It was suggested that polyadenylation of the *DHFR* minigene was occurring via recruitment of polyadenylation sites within the host genome or carrier plasmid DNA. It is most probable that the increased size of the minigene transcript observed here is due to incorrect polyadenylation. The 3' end of the minigene construct consists of exons 7 to 10 and a stretch of 3' UTR (48bp) including the polyadenylation signal derived from a cDNA clone. It is possible that there are genomic elements of the 3' UTR absent from this construct that are necessary for faithful and efficient polyadenylation of the *ERCC1* mRNA. The differential polyadenylation of *ERCC1* has been described (van Duin *et al.*, 1987) and suggests that there maybe additional sequences downstream that are involved in determining the site of polyadenylation. To achieve the appropriate polyadenylation it may be necessary to construct another minigene incorporating a more extensive 3' UTR, or alternatively substitute the polyadenylation site with a stronger poly(A) site such as the rabbit beta-globin poly(A) site. The human calcitonin/ CGRP-I pre-mRNA is processed in a tissue-specific manner into either calcitonin(CT) or calcitonin gene-related peptide (CGRPI). The respective mRNAs are the product of differential splicing and polyadenylation at either the fourth or sixth exons. It was reported that polyadenylation did not occur at the fourth exon of a human *calcitonin* minigene *in vitro.* Subsequent studies revealed that whilst polyadenylation at this site remained inefficient *in vitro*, there was a requirement for specific downstream sequences located within 39 nt of the cleavage site for optimal polyadenylation. When the exon 4 polyadenylation site was replaced with the rabbit beta-globin poly(a) site, efficient polyadenylation was observed (van Oers *et al.*, 1994). Similarly, improved expression of a human *CD2* minigene was reported following the addition of extended 5' and 3' flanking sequences (Zhumabekov *et al.*, 1995)

In wild type embryonic fibroblasts and keratinocytes the 1.5kb transcript is more abundant than the 1.1kb. We have found that where there were two minigene transcripts detected, this characteristic difference in the relative abundance of each transcript was not apparent. It is possible that sequences which are involved in establishing this ratio are absent from the minigene construct.

We have described the complementation of the DNA repair defect in three *ERCC1*-deficient cell types using a mouse *ERCC1* minigene. It was also shown that the pattern of minigene expression did not reflect the normal pattern of expression in each of these cell types with minigene transcripts being larger than predicted, probably due to aberrant polyadenylation. It would not be prudent to proceed with deletional analysis of the promoter region whilst the expression pattern of the minigene is so far removed from the endogenous pattern of *ERCC1* expression.

Chapter 5

Retroviral-mediated correction of UV sensitivity associated with *ERCC1* deficiency

ERCC1-deficient mice die severely runted before weaning, at three weeks of age, as the result of liver failure. The short lifespan of the *ERCC1* null animals has greatly reduced the extent of analysis we have been able to carry out on the consequences of *ERCC1* deficiency in other tissues. Increasing the lifespan of the *ERCC1* null animals would enable studies such as: investigations into the consequences of *ERCC1* deficiency at later stages of development as well as any predisposition to cancers.

As it stands, the *ERCC1* null mouse acts as an animal model for NER deficiency. The nature of this model, a single gene defect primarily affecting a single organ, also provides a suitable model for the development of a retroviral gene therapy system. The retroviral-mediated introduction of a functional *ERCC1* coding sequence into the liver of a null mouse was identified as a means of extending the lifespan of the null animals in order to facilitate the further investigations mentioned. This chapter describes the attempted retroviral-mediated gene transfer and rescue of the UV sensitive phenotype associated with *ERCC1* deficiency in cultured cells *in vitro* prior to the development of a therapeutic *in vivo* strategy.

In theory, gene therapy is quite simple: the introduction of corrective nucleic acid into a cell to alleviate a disease phenotype. In practice, numerous technical obstacles have had to be overcome before any clinical advances could be made. The major problems in this field of research have been the development of safe and efficient gene delivery systems. The advantage of retroviruses is that they can stably infect dividing cells by integrating into the host genome without expressing any immunogenic viral proteins. The integrated viral genome should, in theory, be maintained throughout the life of the host cell. As a result of extensive research, the retroviral gene delivery strategy is perhaps furthest along the road to reaching this goal. To create the safe retroviral vectors used in gene therapy, the life cycle of their naturally occurring counter parts are exploited. The normal life cycle begins when a retrovirus binds to and enters a cell where it injects its RNA genome and accessory proteins into the cytoplasm. The viral genome consists of three genes: *gag, pol* and *env,* which are processed into a number of polypeptides for replication, encapsidation, infection and reverse transcription. There are also three noncoding regions: the two long terminal repeats required for viral integration and gene

expression thereafter, and the psi (ψ) sequences necessary for packaging of the viral genome into virions. Once in the cytoplasm, reverse transcriptase from the virus, converts the viral RNA genome into DNA, which in turn enters the cell nucleus and integrates into the genome. Once integrated into the cell's genome the viral DNA directs the synthesis of viral proteins and RNA, utilising the transcription and translation machinery of the host cell. The resultant proteins once again encapsidate the viral RNA to form new virus particles which bud from the cell and go on to repeat the cycle.

The structure of a retroviral vector is essentially simple, comprising the 5' and 3' long terminal repeats (LTRs), a packaging sequence and the coding sequence of the gene of interest. The coding regions for the viral genes, *gag, pol* and *env* are deleted. For propagation of the recombinant virus it is necessary to provide the additional essential viral elements in *trans* using a cell line known as a packaging cell line. The packaging cell line is engineered to stably express the viral proteins encoded by the *gag, pol* and *env* but lacks the packaging sequence required for encapsidation of the viral genome. This packaging sequence is, however, present on the transgene vector and as a result it is only the RNA from the vector becomes encapsidated, and released from the packaging cell as an infectious virus. This safe viral vector is capable of infecting appropriate target cells, where it can integrate and lead to expression of the therapeutic gene. The viral vector lacks the viral genes required for further replication and encapsidation, thus it is unable to reproduce. The target specificity of a viral vector is determined by using a packaging cell line expressing the *env* gene of a retrovirus with a known tropism. In this case the murine ecotropism was determined by using a packaging cell line expressing the *env* gene from the Moloney murine leukaemia virus (MoMLV).

5.1 Production of recombinant *ERCC1* retrovirus

Recombinant ecotropic retroviruses containing *ERCC1* coding sequences were produced using the murine leukaemia virus (MLV) derived, MFG viral vector (Dranoff *et al.*, 1993) in combination with the viral packaging cell line ψCRE (Danos and Mulligan, 1988). In this system the ψCRE cell line, which stably expresses the retroviral proteins *gag, pol* and *env* is transfected with the retroviral vector encoding

the cDNA of interest. Expression of the cDNA is driven by the promoter activity of the LTR. As it is only the vector that includes the psi (ψ) packaging sequence, only the vector RNA is packaged into virions. The structure of the *ERCC1* viral vector (EMFGI) and the strategy for viral infection of cultured cells is outlined in Figure 5.1.

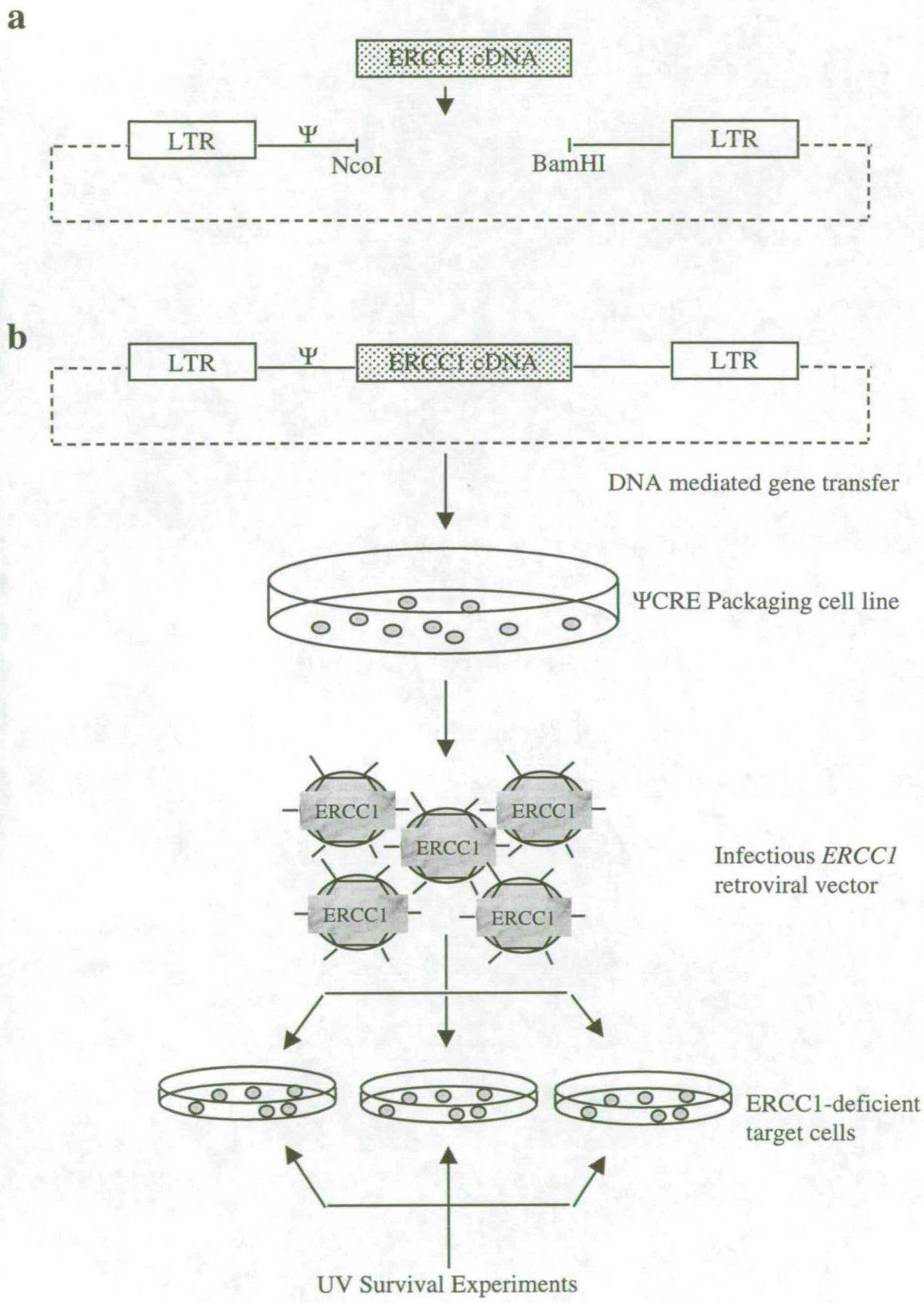
The *ERCC1* coding region was amplified by PCR from the plasmid pTZME using primers designed to introduce *NcoI* and *BglII* restriction sites, at the 5' and 3' ends respectively, required for later cloning into the pMFG plasmid. The sequence of the 5' end primer was : 5' CATGC'CATGGACC,CTGGGAAGGACGAG 3', the position of the NcoI site was underlined, use of this cloning site ensured that the cDNA is inserted into pMFG in the correct reading frame. The sequence of the 3' end primer was: 5' GAA'GATCTTCTCGAGTCATCGAGGC*ACT*TTGAGG 3', the BglII restriction sequence is underlined and the position of the translational stop codon is shown in italics. The PCR product was first cloned as a blunt ended fragment into the *EcoRV* site of Bluescript II SK+. The integrity of the coding sequence was confirmed by sequencing of the *ERCC1*/Bluescript clone (sequencing was performed by Carolanne McEwan). The *ERCC1* cDNA was isolated as a *NcoI / BglII* fragment and cloned into *NcoI/ BamHI* cut pMFG. Recombinants were identified by conventional restriction analysis of DNA from ampicillin resistant bacterial colonies. The ATG of the *NcoI site* is the translational start site of the viral envelope gene and now acts as the start site for the cloned cDNA. Sequencing carried out across the cloning sites confirmed the integrity of each of the sites. A single clone was used for each of the experiments described here.

The EMFGI vector was then used to stably transfect the ψCRE viral packaging cell line, resulting in the packaging of the EMFGI vector RNA into virions. Once transfected with the retroviral vector the cells should become retrovirus-producing cells. This approach was used to obtain four separate retrovirus producing cell lines; the first two lines EMFGI/ψCRE(GS) and EMFG/ψCRE were actually pools of ψCRE cells, transfected with and without a glycerol shock (GS) step, respectively.

Figure 5.1 Structure of the *ERCC1* retroviral vector and strategy for retroviral infections

a, cloning of the *ERCC1* retroviral vector EMFGI shown schematically. The NcoI/ BglII restricted *ERCC1* cDNA fragment (1015bp) was cloned into the NcoI/BamHI restricted pMFG vector (6.3kb). Dashed line, pUC19 plasmid backbone; solid line, MLV sequences including ψ packaging sequence; open boxes, MLV long terminal repeat (LTR) sequences; shaded box, *ERCC1* cDNA. The vector is not shown to scale.

b, strategy for retroviral infection of cultured *ERCC1*-deficient cultured cells. The ψCRE packaging cell line was stably transfected with the EMFGI vector which contains a ψ sequence necessary for packaging of the vector RNA. The resulting infectious virus contains the *ERCC1* retroviral-vector genome, which codes for *ERCC1* but does not code for the *gag, pol* and *env* proteins and is therefore unable to replicate autonomously. The virus was harvested within the media in which the producing cells were cultured and applied to the *ERCC1*-deficient target cells to permit infection to take place. After infection by the recombinant retrovirus phenotypic rescue was determined by reduced sensitivity to UV irradiation.


a
ERCC 1 cDNA
LTR
Ψ
NcoI
BamHI
LTR
b
LTR
Ψ
ERCC 1 cDNA
LTR
DNA mediated gene transfer
ΨCRE Packaging cell line
ERCC1
ERCC1
ERCC1
ERCC1
ERCC1
Infectious *ERCC1* retroviral vector
ERCC1-deficient target cells
UV Survival Experiments

At a later date the pool of cells arising from the transfection with glycerol shock was then plated at low density to enable individual colonies to be picked. Six out of the ten colonies picked contained EMFGI sequences, as determined by PCR detection of the *ERCC1* coding region from the viral vector. The assay developed to identify cells that contained the retroviral vector exploited the fact that the viral vector contained the *ERCC1* cDNA, which meant it was possible to use primers specific for exons 1 and 10 to amplify the 950bp cDNA insert. The endogenous genomic sequences could not be amplified as the product would be too large (~15kb). Two of these PCR positive clones, EMFGI/ψCRE #8 and EMFGI/ψCRE #9 were expanded for use as viral producing cell lines.

5.2 UV survival of *ERCC1*-deficient fibroblasts (PF24) following retroviral-mediated gene transfer

The spontaneously transformed embryonic fibroblast cell line derived from an *ERCC1*-deficient embryo (PF24) has been shown to be hypersensitive to UV and completely lacking in NER activity (Mcwhir *et al.*, 1993). Recombinant *ERCC1* retroviruses from four virus producing cell lines, described in section 5.1 of this chapter, were used in an attempt to alleviate this UV hypersensitivity.

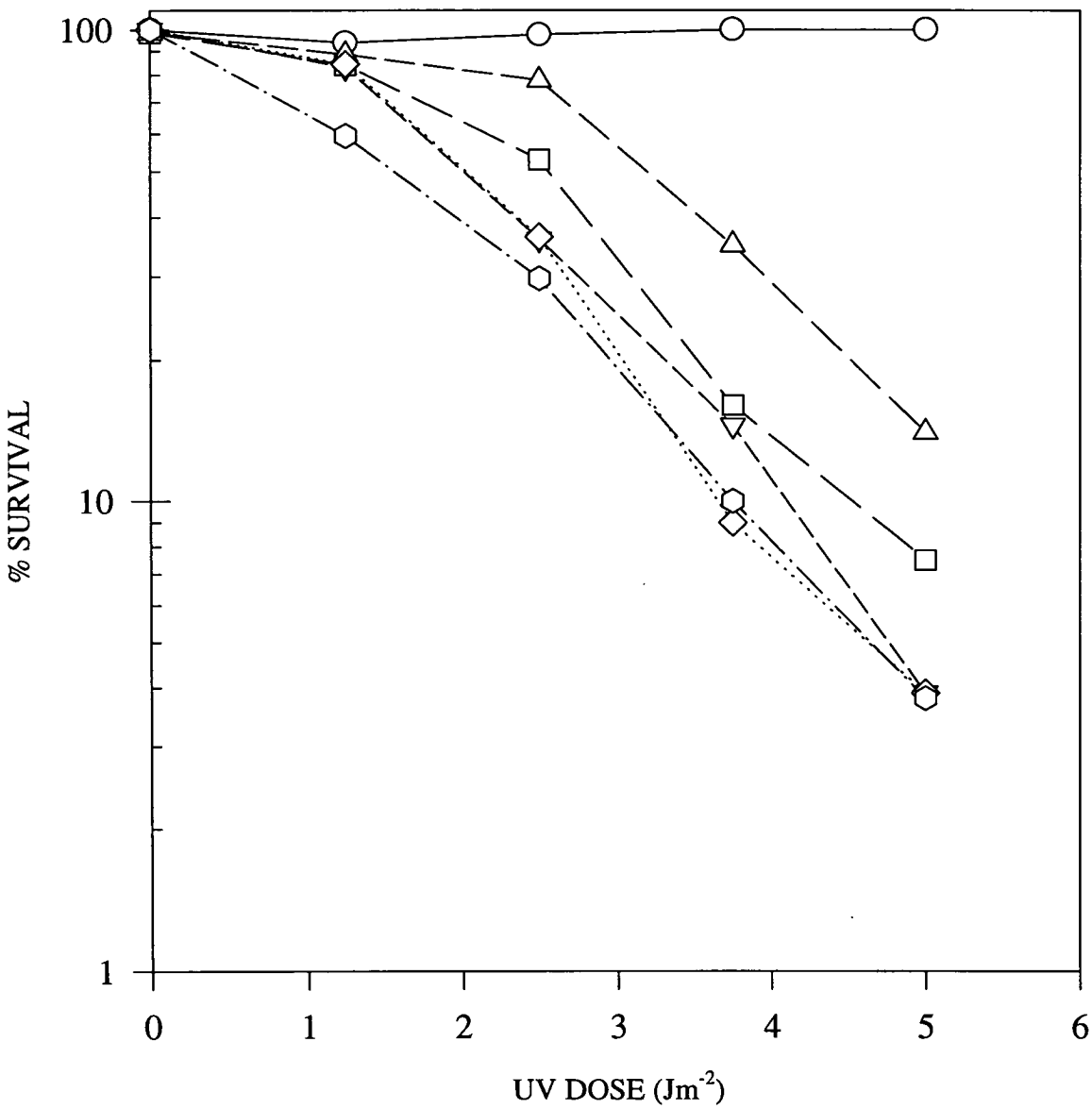
In each case the virus producing cells were cultured in a single dish, until almost confluent, to optimise the yield of virus. The culture medium was changed on the day before the virus was collected. After overnight culturing the virus-containing medium was harvested from the producing cells and added to a dish of target PF24 cells. The target cells had been split on the previous day to ensure that the cells were actively dividing. The infection was allowed to proceed for 6 hours before the virus containing medium was replaced with normal culture medium. The infected PF24 cells were expanded in culture for use in UV sensitivity experiments. For these experiments, $2x10^4$ infected cells were plated in small dishes and cultured overnight before being irradiated with various doses of UV irradiation. The dishes were maintained in culture until control dishes, not exposed to UV, reached confluence. At this point the cells were fixed and stained to quantify the UV sensitivity of the infected cells. The UV survival curves obtained are presented in Figure 5.2. The

graph illustrates sensitivity to UV in terms of the cells surviving each dose of UV irradiation as a % of untreated controls. The graph shows that the infections carried out using virus from the EMFGI/ ψUCRE producing line had only a minimal affect on the survival curve relative to the control PF24 survival curve. PF24 cells infected with virus from the EMFGI/ ψUCRE (GS) producing line still exhibited sensitivity to UV irradiation but in this case the survival profile was clearly distinguishable from the repair deficient profile of PF24 cells. At the lower UV doses of 1.25 and 2.5 Jm^{-2} the 98 and 78% survival of the infected cells was most reminiscent of the wild-type cells (PF20). At the higher UV doses of 3.75 and 5 Jm^{-2} the survival dropped to 33% and 7% respectively whereas survival of wild-type cells was unaffected. The survival curve for the PF20 cells showed 94 and 98% survival at the lower UV doses of 1.25 and 2.5 Jm^{-2} and 100% survival at the two higher UV doses. It was possible that the modest degree of phenotypic rescue observed was due to a low viral titre resulting from the use of a pool of virus producing cells, in which a proportion of the pool may not have been producing any virus, rather than a pure clone from this pool. The repair deficient PF24 cells were then infected in an identical manner by recombinant virus harvested from the two producing lines, EMFGI/ψCRE #8 and EMFGI/ψCRE #9 cloned from the EMFGI/ ψUCRE (GS) pool. However, the survival curves obtained following this approach showed little evidence of phenotypic correction of UV sensitivity. The survival profile of the cells infected by EMFGI/ψCRE #8 showed 83 and 36% survival the lower UV doses of 1.25 and 2.5 Jm^{-2} . At the higher doses of 3.75 and 5 Jm^{-2} survival fell to 14 and 4%. Similarly, the survival profile of the cells infected by EMFGI/ψCRE #9 showed 84 and 34% survival at the lower UV doses of 1.25 and 2.5 Jm-2. At the higher UV doses of 3.75 and 5 Jm^{-2} survival of 9 and 3% was observed.

In summary, the UV sensitivity of PF24 was seen to be partially corrected following retroviral-mediated gene transfer by viruses from the producing cell line EMFGI/ ψUCRE (GS).

Figure 5.2 UV sensitivity of *ERCC1* fibroblasts (PF24) following retroviral mediated gene transfer

The UV survival curves for wild type (PF20), *ERCC1*-deficient (PF24) mouse fibroblasts and PF24 cells following retroviral infection from four separate viral sources are shown. Each point represents the mean of two separate experiments. Circle, PF20; hexagon, PF24; triangle, PF24+EMFGI/ψCRE (GS); square, PF24+EMFGI/ψCRE; inverted triangle, PF24+EMFGI/ψCRE #8; diamond, PF24+EMFGI/ψCRE #9.


100
10
1
% SURVIVAL
0
1
2
3
4
5
6
UV DOSE (Jm-2)

5.3 UV resistant cells within pools of infected cells are corrected to wild-type levels

It was necessary to investigate further whether the limited degree of phenotypic correction observed thus far was a consequence of low viral titre or, alternatively, was due to some other factor which prevented the recombinant virus from bringing about complete phenotypic rescue. It was necessary to determine whether all the cells within pools of infected PF24 displayed the same modest degree of UV resistance or had some of the cells been corrected to the wild-type levels seen in PF20. Cells infected by recombinant virus from the EMFGI /ψCRE (GS) pool, which gave the best level of phenotypic correction thus far, were plated at low density and exposed to UV doses of between 3.5 and $5Jm^{-2}$. The irradiated cells were maintained in culture to enable UV resistant colonies to be identified and picked as clones. Each of the twelve UV resistant colonies picked were found to be positive, when screened by PCR, for the presence of the *ERCC1* cDNA element of the EMFG vector (not shown). Two of these UV resistant clones, PF24/UVR # 4 and #12, were expanded in culture for use in UV survival experiments. The UV survival curves obtained are presented in Figure 5.3. The graph illustrates sensitivity to UV in terms of % cells surviving at each UV dose. It is clear that the two PF24/ UVR clones exhibit a wild-type pattern of UV resistance as seen by the similarity to the survival curve of the PF20 cells.

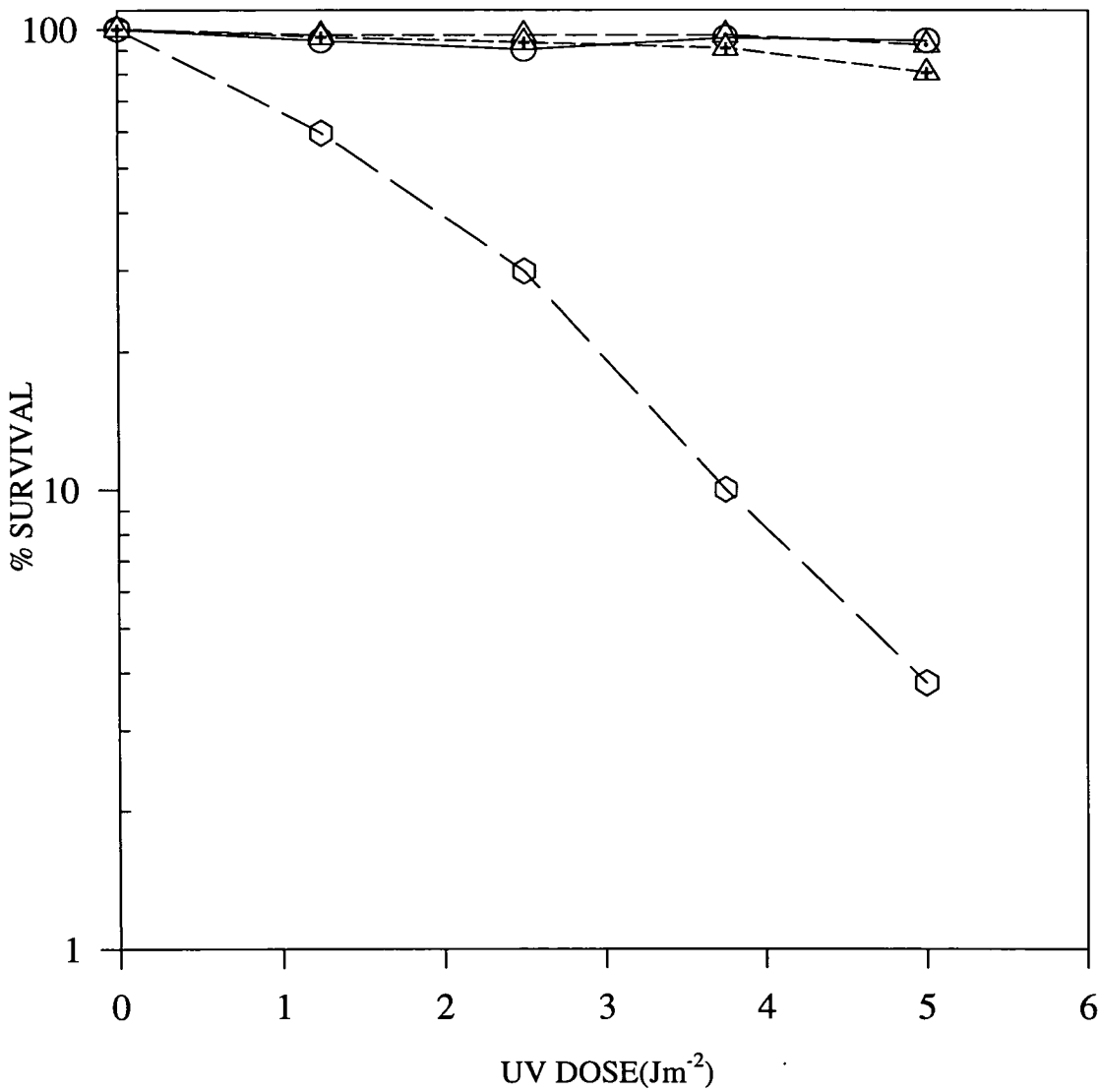
Thus the recombinant virus produced by the cell line EMFGI/ ψUCRE (GS) is capable of infecting repair deficient PF24 cells and bringing about phenotypic correction to wild-type levels.

5.4 Increased resistance to UV irradiation correlates with *ERCC1* transcription

We have described the identification of UV resistant clones from within the pools of infected PF24 cells. It remained possible that the UV resistance of these clones was not a consequence of retroviral-mediated gene transfer. Such UV resistant clones could have arisen had the PF24 cultures been contaminated, at low levels, with the wild-type PF20 cells. PCR genotyping of the UV resistant clones confirmed that the

Figure 5.3 PF24 cells within infected pools are corrected to wild-type levels of UV sensitivity

The UV survival curves for wild type (PF20), *ERCC1*-deficient (PF24) mouse fibroblasts and UV resistant PF24 clones isolated from virally infected pools are shown. Each point represents the mean of two separate experiments. Circle, PF20; hexagon, PF24; triangle with dot, PF24 /UVR #4; triangle with cross, PF24 /UVR #12.


100
10
1
% SURVIVAL
0
1
2
3
4
5
6
UV DOSE(Jm-2)

cells were indeed *ERCC1* null (data not shown). It remained necessary to confirm that the retroviral-mediated gene transfer was responsible for the phenotypic correction. Evidence for this was gained by demonstrating that the observed correction of UV sensitivity correlated with the expression of a transgene derived *ERCC1* mRNA.

Northern blot analysis was performed on RNA extracted from each of the transduced PF24 cell lines, to demonstrate that the phenotypic correction observed in some of the cells was due to *ERCC1* expression, resultant from the retroviral-mediated gene transfer (Figure 5.4a). Total RNA (30μg) from the wild-type mouse embryonic fibroblast cell line PF20, the parental *ERCC1*-deficient mouse embryonic fibroblast cell line PF24 and the transduced cell lines PF24+EMFGI/ψCRE #8, PF24+EMFGI/ψCRE (GS), its UV resistant derivatives PF24/UVR #4 and PF24/UVR #12, was electrophoresed on a 1.4% agarose-formaldehyde gel, transferred onto a nylon membrane and *ERCC1* transcript detected by hybridisation to a mouse *ERCC1* cDNA probe. The blot was later stripped and reprobed using a *GAPDH* probe to standardise for any variations in RNA loading in each lane.

The lower panel shows the *GAPDH* reprobe of the filter to standardise for loading. The detection of *ERCC1* mRNA by probing with *ERCC1* cDNA is shown in the upper panel. The normal pattern of *ERCC1* expression was evident in the wild type PF20 lane, the 1.5kb major transcript and the normal 1.1kb transcript present at a lower abundance. In the PF24 lane there was no evidence of any mRNA species hybridising with the *ERCC1* probe. As expected, there was no suggestion of any hybridising transcripts in the PF24+EMFGI/ψCRE #8 lane. This was in keeping with the fact that these cells had not shown any degree of phenotypic correction as determined by UV survival. In the lanes of both of the UV resistant clones there was clear evidence of *ERCC1* hybridising bands at about 2.8kb, which matches the predicted size for an EMFGI transcript. Additional, larger hybridising species, believed to be retroviral pre-mRNA, were also evident in each of the lanes. In the PF24+EMFGI/ψCRE (GS) lane there was evidence of these transcripts at a much lower abundance, presumably reflecting the proportion of the cells within the pool that had been infected by the recombinant retrovirus. Phosphorimagery was used to

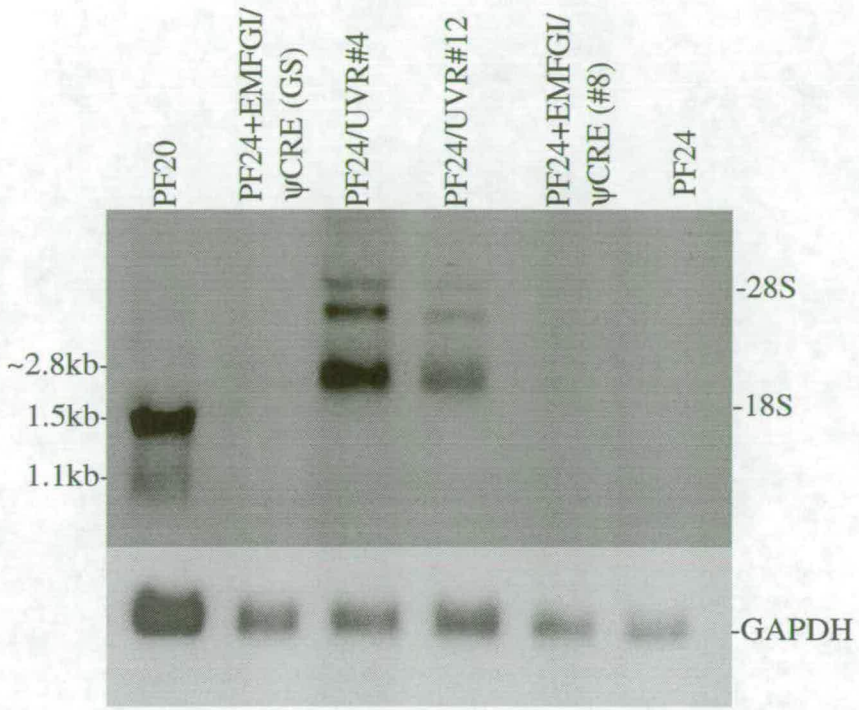
Figure 5.4 The UV resistance of transduced clones correlates with *ERCC1* transcription

a, northern blot analysis of RNA from transduced PF24 cell lines. Upper panel, total RNA was prepared from the transduced cell lines PF24+EMFGI/ψCRE (GS), PF24/UVR #4, PF24/UVR #12, PF24+EMFGI/ψCRE #8, wild type PF20 and the *ERCC1*-deficient PF24. The RNA was electrophoresed on a 1.4% formaldehyde-agarose gel, transferred onto a nylon membrane and probed with an 800bp BamHI fragment of the *ERCC1* cDNA corresponding to exons 1 to 8. Positions of the 18S and 28S rRNAs and estimated transcript sizes are indicated. Lower panel, the blot was then stripped and reprobed using a *GAPDH* cDNA fragment to standardise for RNA loading in each lane.

b, phosphorimager analysis of *ERCC1* transcript levels in the transduced pool and UVR clones. The values from the phosphorimage were obtained using the ImageQuant software V3.3 image analysis package. The signal values represent the area, in arbitrary units, under the peak corresponding to the ~2.8kb transcript. The ratio represents the value of the *ERCC1* signal divided by the value of the *GAPDH* signal, i.e. the standardised values. The level of *ERCC1* transcript in the pool relative to the two clones is indicated.

c, determination of the number of transduced cells in the infected pool. The number of UV resistant cells in the infected pool was determined by the number of colonies formed after exposure of 100000 cells to UV doses of 3.75 and 5 Jm^{-2}. The percentage survival was then corrected relative to the plating efficiency of unirradiated PF24+EMFGI/ψCRE (GS) cells which was 4.25%.

a.



b.

Sample	ERCC1 signal	GAPDH signal	Ratio ERCC1/GAPDH	Pool as % of each clone
Pool	782	897	0.9	-
UVR#4	24053	893	26.9	3.3
UVR#12	12814	1061	12.1	7.4
				5.3% Average

c.

# Cells plated	UV Dose (Jm^{-2})	# Surviving colonies	% Survival (cells plated)	% Survival corrected for the plating efficiency (4.25%)
100000	3.75	196	0.19	4.5
100000	3.75	185	0.18	4.2
100000	3.75	177	0.18	4.2
				4.3 Average
100000	5.0	83	0.083	1.9
100000	5.0	77	0.077	1.8
100000	5.0	68	0.068	1.6
				1.76 Average

quantify the amount of retroviral transcript present in the RNA from the infected pool relative to the UV resistant clones (Figure 5.4b). This analysis showed that the level of the *ERCC1* transcript in the infected pool lane was 5.3% of the average level of transcript evident in the UV resistant lanes. The phosphorimage analysis also showed a two-fold variation in the level of expression evident the two UV resistant clones

The level of *ERCC1* expression detected in the pool of infected cells could be regarded as an indirect indication of viral titre. Where the level of expression in the pool was measured at 5.3% of that in the clones it was estimated that somewhere in the region of only 5.3% of the cells in the pool had been infected by the retrovirus. Evidence to support this idea was gained from a cell plating experiment performed using the pool of infected cells. To determine the plating efficiency of the pF24+EMFGI/ψCRE (GS) cells, 1000, 500 and 200 cells were plated (in duplicates) and maintained in culture for one week. At this point the cells were fixed and stained to enable the number of colonies on each plate to be counted. The plating efficiency was determined by expressing the number of resultant colonies as a percentage of the number of cells plated. The plating efficiency for the cells of the infected pool was measured as 4.25% (data not shown). In parallel, 10^5 cells were plated and cultured overnight prior to UV irradiation, in order to determine the actual proportion of cells in the pool that had been infected by retrovirus, and were therefore UV resistant. The dishes (in triplicate) were exposed to UV doses of 3.75 Jm^{-2} and 5 Jm^{-2}, the UV doses from which the UV resistant clones were picked, then maintained in culture for one week. At this point the cells were fixed and stained and the number of colonies on each plate was counted (Figure 5.4c). The number of surviving colonies was used to determine the percentage of plated cells surviving. This figure was then adjusted relative to the plating efficiency of these cells by expressing the % cells surviving relative to the control plating efficiency (i.e.(% surviving cells / % plating efficiency). The result of this analysis showed that the average percentage of corrected cells in the pools surviving the UV dose of 3.75 Jm^{-2} was 4.3%. Suggesting that 4.3% of the cells in the pool were UV resistant as a result of retroviral mediated gene transfer. This figure is in accordance with the 5.3% correction estimated after the northern blot analysis The average percentage of cells in the pool surviving 5 Jm^{-2} was 1.76%. It

was first predicted that if cells were infected by retrovirus the observed correction would be to wild type levels of UV resistance as seen with the UV resistant clones. As such, the lower percentage survival of cells at the higher UV dose was not expected. This reduced survival at a UV dose of 5 Jm^{-2} may have been a reflection of heterogeneity in the levels of *ERCC1* being expressed in infected cells. It remained possible that the level of expression in some cells was not sufficient to confer resistance to the higher dose of UV. The earlier northern blot analysis performed on the UV resistant clones supported the idea of variation in the levels of *ERCC1* transcript between clones.

In summary, we have demonstrated that the observed phenotypic correction following retroviral-mediated gene transfer correlates with the expression of an *ERCC1* mRNA. Further, the abundance of this *ERCC1* mRNA appears to correlate with the degree of correction observed within the pool of infected cells and the subsequent UV clones derived from it. The level of *ERCC1* transcript detected in the infected pool was in accordance with the proportion of the cells within the pool that were resistant to a UV dose of 3.75 Jm^{-2}.

5.6 Discussion

The Achilles heel of gene therapy is currently gene delivery, the inability to deliver genes efficiently with stable expression. The aim of any gene therapy is to introduce successfully a nucleic acid into a defective cell such that the expression of those sequences will result in the correction of the cellular defect. Although the absolute nature of the defect can greatly influence the most appropriate strategy for gene delivery, most current approaches utilise viral vectors as the delivery system. The use of retroviral vectors typically results in the stable transduction of the majority of target cells (Scharfmann *et al.*, 1991 and reviewed in Smith 1995). The development of packaging cell lines that permit the production of high titres of replication-disabled recombinant virus (Danos and Mulligan, 1988) was crucial to the advancement of this technology.

Using the MFG vector for the stable transfection of the ψCRE packaging cell line we have engineered a retroviral based system to produce ecotropic recombinant

retroviruses carrying the murine *ERCC1* cDNA sequences. Using this system we have obtained four separate producing lines for the propagation of *ERCC1* retroviruses to be used in subsequent gene transfer studies.

The ability of the recombinant retrovirus to bring about phenotypic correction of *ERCC1*-deficient fibroblasts was demonstrated by the observed reduction of the UV sensitivity in transduced cells. The best phenotypic reversion was achieved using viruses harvested from pools of virus producing cells. The use of a glycerol shock step in the transfection of the packaging cell line resulted in a higher degree of correction when using the recombinant virus harvested from the pool of transfected cells. This is most likely due to the improved efficiency of transfection of the packaging cells resulting in a higher viral titre. The ψCRE packaging cell line was co-transfected with a selectable neomycin transferase marker gene and infections were carried out using the pools of cells that survived G418 selection. All of the cells in the pool should have contained the vector. When attempting to clone pure producing cell lines from the pools it was discovered that only 60% of the cells were positive for the retroviral vector. Furthermore when two of these cloned producing cell lines were used in retroviral-mediated gene transfer experiments they failed to result in any discernible phenotypic correction. Despite containing the retroviral vector sequences, these clones were apparently not producing therapeutic levels of recombinant retrovirus. It was not possible to linearise the retroviral vector prior to transfecting the ψCRE packaging line, so integration of the vector would have been random rather than being through the ends of a linear fragment. It is possible that when randomly integrating into the genome, the viral vector was disrupted in such a way as to prevent the vector genome from being packaged. Any disruption or loss of the packaging sequence would have this affect. Together these factors suggest that the resulting viral titre in these experiments was far from optimal. As an alternative to obtaining a producing line giving a higher titre of virus it would have been feasible to simply scale up the number of producing cells from which the virus was harvested. The successful transduction of target cells depends, to a certain degree, on the volume and concentration of virus that the target cells are exposed to and, as such,

this may have required the development of a technique to concentrate the virus. This would have been technically very difficult due to the fragile nature of retroviruses.

Retroviral-mediated gene transfer has been used to correct the NER defect associated with fibroblasts isolated from XPD patients (Quilliet *et al.*, 1996). In this case the MLV derived viral vector was propagated in the amphotropic ψCRIP packaging cell line to permit infection of the human fibroblasts. The vector also carried a selectable neomycin transferase marker gene to permit G418 selection of transduced cells. The correction of the repair defect in the XPD cells was only assessed in cells that had survived G418 selection. In the absence of a selectable marker in our system, the analogous results from the PF24 correction studies would be those carried out on the UV resistant clones isolated from transduced pools, where the UV could be regarded as a substitute selection for transduction. In this light the results obtained from these studies compare favourably with reports of retroviral-mediated correction of the DNA repair defects of xeroderma pigmentosum cells belonging to complementation groups A, B and C, where phenotypic correction was determined after primary selection for the presence of vector sequences (Zeng *et al.*, 1997). The ability to select for transduced cells prior to carrying out analysis of phenotypic correction is particularly useful in that it can permit *in vitro* studies to proceed where viral titre is low.

This work was embarked upon as a means of developing a gene therapy strategy aimed at increasing the lifespan of the *ERCC1*-deficient mouse. As such the goal of this study was to develop an efficient *in vivo* gene transfer technique using a retroviral vector. The use of the retroviral vector itself, whilst being particularly suited to the *in vitro* study described here, would have undoubtedly presented numerous technical hurdles when being used *in vivo*. It is a basic biological requirement that for retroviral infection to take place the target cell must be actively dividing. In the case of the *ERCC1*-deficient mouse, the target cells were those of the normally quiescent liver. Whilst it is possible to induce a degree of cell division within the liver by partial hepatectomy, it would not have been possible to perform such surgery on the *ERCC1*-deficient animals as they die before weaning. Without the active division of the target cells it is difficult to envisage how this system in its

current state could be successfully used for retroviral-mediated gene transfer aimed at rescuing the *ERCC1* mouse. It is now generally accepted that retroviral gene therapy is best suited to what are described as *ex vivo* gene therapy strategies, where target cells are removed from the patient and transduced *in vitro* before being transplanted back into the patient for repopulation. (Kay *et al.*, 92 and Grossman *et al.*, 94)

If this work was to be taken on any further, possible improvements to this system would have to include improved 'targeting' of the retrovirus by introducing tissue specific promoter elements to bring about expression in a hepatocyte specific manner. However, it is likely that this would still require *in utero* transduction of the *ERCC1* null hepatocytes in order to be most affective. It is difficult to imagine that the work required to overcome the technical hurdles posed by such a strategy would be justified by the potential outcomes in the event of any success. Realistically the phenotypic rescue of the DNA repair deficient *ERCC1* null mice should be attempted using a non-retroviral strategy.

Whilst it is clear that with time and further technological developments that retroviral-mediated gene transfer has a great deal to offer to the gene therapy clinician, it is less likely that the same technology will have much to offer the *ERCC1*-deficient mouse.

Chapter 6
Targeted *in vivo* expression of *ERCC1* in the liver

ERCC1-deficient CHO cells (CHO43-3B) and mouse embryonic fibroblasts (PF24) derived from *ERCC1*-deficient mice have been shown to have a 100-fold elevated spontaneous mutation rate at the hypoxanthine phosphoribosyltransferase (*HPRT*) locus in addition to a higher level of genome instability, as indicated by an increased frequency of micronuclei (Melton *et al.*, 1998). As such, we would speculate that *ERCC1*-deficient mice might be more susceptible to UV induced skin cancers than wild type mice. The premature death of our *ERCC1*-deficient mice has precluded any investigation relating to the incidence of skin cancers and *ERCC1* deficiency. When the *ERCC1*-deficient animals were first described, in addition to the striking liver abnormalities, it was also reported that elevated levels of the tumour suppressor protein p53 were detected in the liver, brain and kidney (McWhir *et al.*, 1993). The lethal liver phenotype of the *ERCC1*-deficient mice, which results in death before weaning, has greatly restricted studies of the consequences of *ERCC1* deficiency in these and other tissues. It is clear that extending the life of *ERCC1*-deficient mice would permit some, if not all, of these studies to be completed. One approach to this problem, currently being pursued in our laboratory, is the use of Cre/*loxP* mediated gene targeting, as described in the introduction chapter, to bring about the conditional inactivation of the *ERCC1* gene in the specific organ or system of interest, whilst maintaining the expression of *ERCC1* in the liver. However, this approach is time consuming, and reliant on the availability of additional transgenic mouse lines expressing Cre recombinase in tissue specific patterns.

We speculate, that by restoring *ERCC1* function in the liver, the lifespan of the *ERCC1*-deficient mice would be extended significantly and that this extended lifespan would enable further studies of the consequences of *ERCC1* deficiency in other tissues and systems. The work presented in this chapter describes the use of conventional transgenesis in the generation of a transgenic mouse line expressing *ERCC1* specifically in hepatocytes and the consequences of crossing this transgene onto an *ERCC1*-deficient background.

Conventional transgenesis involves the microinjection of cloned DNA sequences into the pronucleus of a fertilised mouse egg. The eggs are then surgically transplanted into a pseudopregnant recipient female and allowed to develop to term. The injected

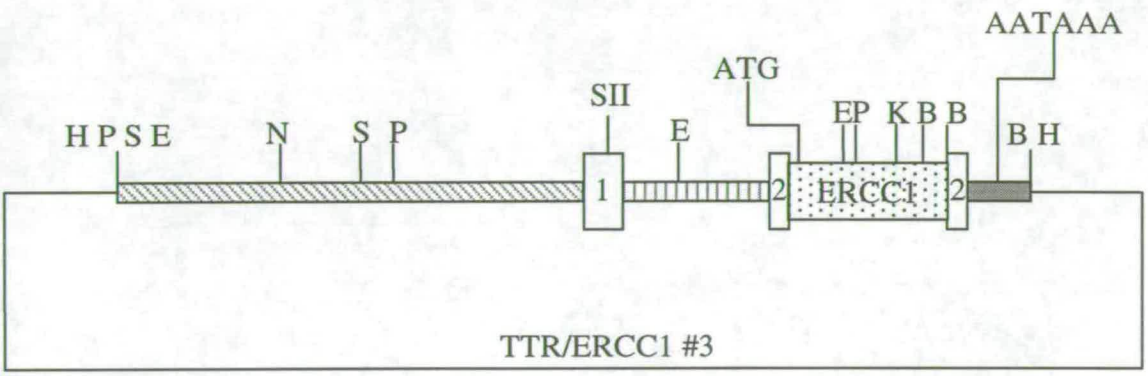
DNA becomes chromosomally integrated and is typically expressed in the resultant offspring. This approach has been used extensively to characterise the *cis* acting elements responsible for both tissue specific and stage specific expression of genes (Yan *et al.,* 1990) Many cloned genes that have been introduced into the mouse germline in this manner have displayed appropriate tissue or stage specific patterns of expression. Integration of the transgene DNA often occurs at the one-cell stage and is consequently present in every cell, including the germ cells, of the transgenic animal. However, if integration occurs at a later stage the resultant animals are described as 'mosaics' and some cells in the transgenic animal may not contain the transgene. The method by which the cloned sequences are introduced to the fertilised egg means that the number of copies of the transgene that become integrated into the genome can vary from 1 to several hundred (Robertson *et al.,* 1996). Generally, the transgene copies become integrated in a head to tail arrangement at a single site of integration (Brinster *et al.,* 1981 and Costantini and Lacy, 1981) although some integration events are more complex.

6.1 Construction of an *ERCC1* transgene regulated by the *transthyretin* (TTR) gene promoter

Transthyretin (TTR) is a thyroid hormone transport protein that is normally secreted by hepatocytes into serum and by the choroid plexus epithelium into the cerebral spinal fluid. The protein is not present in significant amounts in any other tissues. Studies carried out *in vitro* identified the upstream sequences that were required for *TTR* expression in hepatocytes (Costa *et al.,* 1996). Subsequently, conventional transgenesis was used to determine whether these sequences were sufficient to direct cell specific expression in the liver and choroid plexus *in vivo.* It was found that the minimal sequences (300bp) required for expression in hepatocytes were sufficient for normal expression in the liver but lead to aberrant expression in regions of the brain other than the choroid plexus. When 3kb of upstream sequences were included, transgene expression in the brain was restricted to the choroid plexus. When more

Figure 6.1 Structure of the transthyretin (*TTR*) regulated *ERCC1* transgene (TTR/*ERCCI #3)*

The structure of the *TTR/ERCC1 #3* transgene is shown schematically. The upstream (3kb) transthyretin regulatory region is shown by the diagonally striped box. Exons are represented by open boxes and are numbered. The first intron is shown as by the vertically striped box. The *ERCC1* cDNA sequence is represented by the light stippled box. The SV40 3' end is represented by the dark stippled box. The plasmid backbone is shown as a thin line. The relative positions of the translational start (ATG) and the polyadenylation signal (AATAAA) are indicated. Restriction sites used in cloning and subsequent analysis are shown as follows: H, HindIII; P, PstI; S, SstII; E, EcoRI; N, NcoI; SII, SstII; K, KpnI; B, BamHI. Shown to scale, 1cm = 500bp.


AATAAA
ATG
SII
EP K B B
H P S E
N
S P
E
B H
1
2
ERCC1
2
TTR/ERCC1 #3

than 6 copies of the transgene were present, expression was reported in the salivary gland and gut of some lines, but this was not apparent in all mice of the lineage (Yan *et al*, 1990). This minigene construct was later adapted for use as a vector suitable for the liver specific expression of any cloned sequences. The introduction of several point mutations in the first and second exons resulted in the abolition of translational start sites and introduction of unique cloning sites within the *TTR* minigene. The use of this vector for the successful liver targeted expression of p21 (a cyclin-dependent kinase inhibitor) and a mutant p53 (a tumour suppressor) has been reported. (Wu *et al.*, 1996 and Bowman *et al.*, 1996). A *TTRp21* transgene was constructed by cloning the p21 cDNA into the second exon of the *TTR* minigene backbone. Northern and western blot analysis of five *TTRp21* transgenic mouse lines revealed that expression of the transgene was restricted to the liver. Based on this data, the transthyretin minigene vector (pTTRI exV3) was considered as an ideal means of targeting expression of *ERCC1* to the liver using conventional transgenesis.

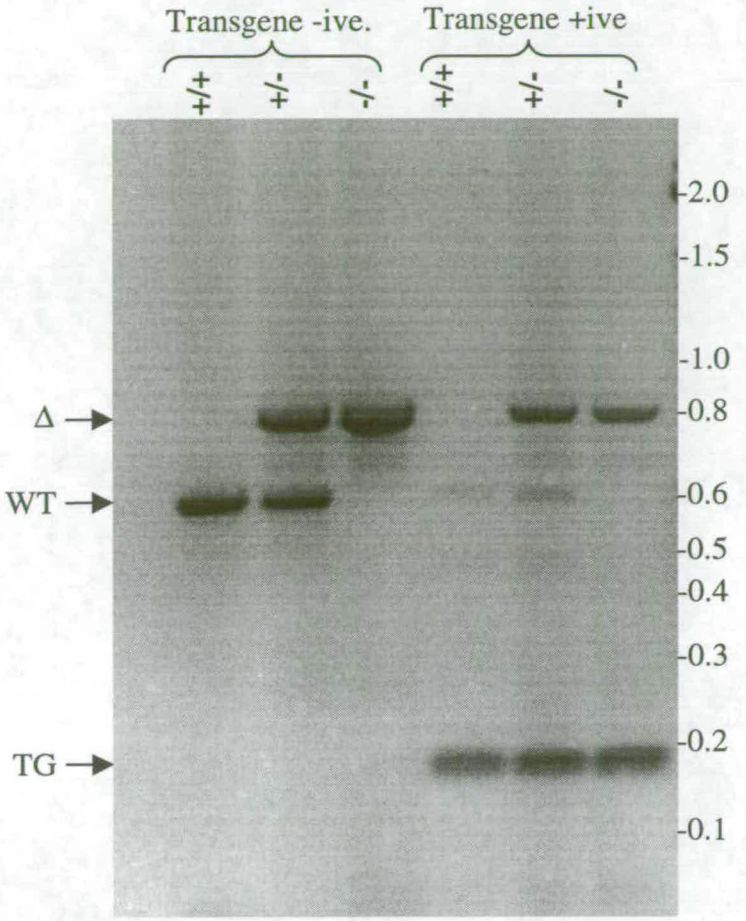
The structure of the *TTR/ERCC1* transgene is shown schematically in Figure 6.1. The exon 1 to 10 region of the *ERCC1* cDNA was isolated as a 900bp XbaI/HindIII fragment from *ERCC1*/Bluescript, this fragment includes only 5nt of *ERCC1* sequence upstream of the translational start signal and none of the *ERCC1* 3'UTR. The fragment was blunt-ended and cloned into the unique StuI site of p*TTR* exV3 (kindly provided by Dr. T. Van Dyke). Initiation of transgene transcription and polyadenylation is determined by the *TTR* regulatory sequences, whilst translation initiates from the ATG within the cloned cDNA fragment. Recombinants were characterised using restriction enzyme analysis. Two clones with the predicted structure were then sequenced to confirm the integrity of the translational start signal within the *ERCC1* cDNA insert. A single clone, *TTR/ERCC1* #3 was used for all the experiments described in this chapter.

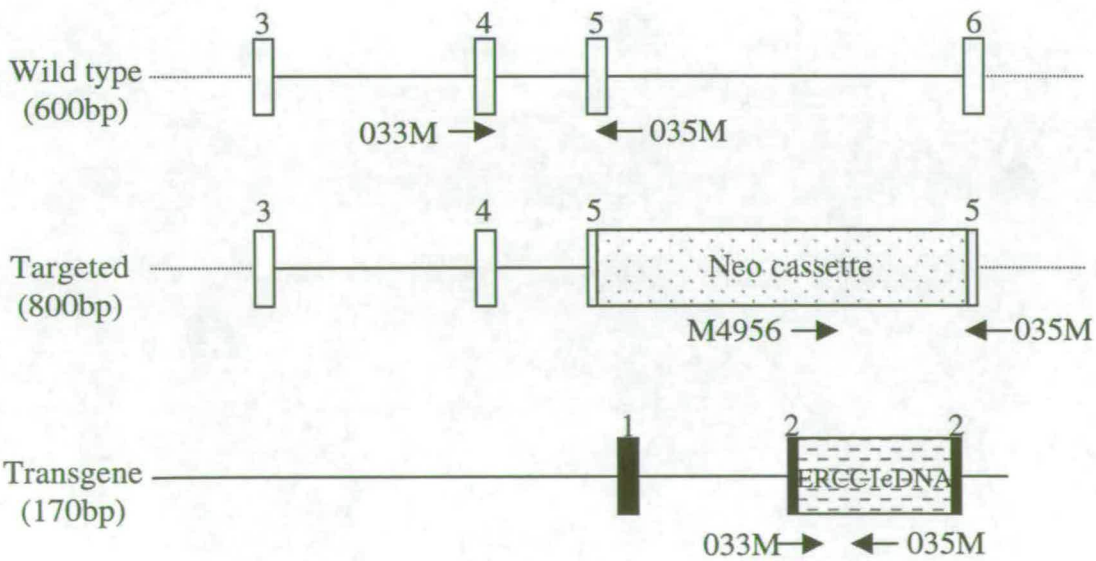
6.2 A single PCR assay capable of distinguishing between three *ERCC1* genes

The PCR assay used for the routine genotyping of our *ERCC1* breeding colony was also suitable for the genotyping of the *TTR/ERCC1* #3 transgenic line. The ability of this assay to identify both the wild type and targeted alleles in addition to the

Figure 6.2 PCR genotyping reaction that distinguishes between three *ERCC1* genes

The products of PCR genotyping reactions for the *ERCC1* wild type and targeted alleles and the transgene. Genomic DNA isolated from mouse tail biopsies was used as a template for PCRs carried out using primers 033M, 035M and M4956, and the products separated by electrophoresis. The 800bp product (from M4956 and 035M) is specific for the targeted allele(Δ), the 600bp product (from 033M and 035M) is specific for the wild type allele(WT), the 170bp product (from 033M and 035M) is specific for the transgene (TG). The strength of the endogenous wild type signal is reduced in the presence of transgene sequences. Transgene positive and negative samples of each *ERCC1* genotype were used. The relevant areas of the three templates are shown schematically below. The relative positions of the primers giving rise to the product from each template are indicated. Exons are numbered and shown by open boxes for *ERCC1* and closed boxes for *TTR* . Shown to scale, 1cm = 500bp.


Transgene -ive.
Transgene +ive
+/+
+/-
-/-
+/+
+/-
-/-
2.0
1.5
1.0
0.8
0.6
0.5
0.4
0.3
0.2
0.1
Δ
WT
TG


Wild type
(600bp)
3
4
5
6
033M
035M
Targeted
(800bp)
3
4
5
5
Neo cassette
M4956
035M
Transgene
(170bp)
1
2
2
ERCC1eDNA
033M
035M

transgene resulted in its use for all subsequent genotyping of *TTR/ERCC1* transgenic mice (Figure 6.2). The assay was performed as a single PCR reaction involving three primers (033M, 035M, and M4956). The 035M (*ERCC1* exon 5) and M4596 (*Neo* cassette) primers gave an 800bp product specific for the targeted allele. The 033M (*ERCC1* exon 4) and 035M (*ERCC1* exon 5) primers gave a 600bp product specific for the wild type allele, no product was seen from the targeted allele as the distance between the two primer binding sites on this allele (3.0kb) was too great for efficient amplification under the cycle conditions used. The same primer pair did give rise to a 170bp product specific for the transgene. It was observed that the 033M/035M product from the endogenous wild type allele became significantly weaker in the presence of the transgene, reflecting the propensity of the PCR reaction to favour the synthesis of smaller products where possible.

6.3 Production and identification of *TTR/ERCC1* transgenic mice

The transgene construct *TTR/ERCC1* #3 was restricted with HindIII and gel purified, as described in the methods chapter, to liberate the transgene fragment from the plasmid backbone. Transgenic mice were produced by the microinjection of the purified transgene fragment into the pronuclei of fertilised mouse eggs. All resultant offspring were screened by PCR for the presence of the transgene. In addition to the surviving animals, genotyping was also carried out *post-mortem* on three pups that were found dead soon after birth. It was found that two of these pups were transgene positive. A single female founder animal (TG#17) was identified from 23 mice that survived to weaning. All subsequent *TTR/ERCC1* transgenic lines were established from this animal.

6.4 Crossing of the *TTR/ERCC1* transgene onto an *ERCC1*-deficient background

The purpose of generating the *TTR/ERCC1* transgenic strain was to determine whether the targeted expression of *ERCC1* in the liver would be sufficient to rescue the lethal liver phenotype seen in *ERCC1*-deficient mice. Having successfully identified a founder animal it was necessary to cross the transgene onto an *ERCC1*-

deficient background. In the first instance the founder animal was crossed with a wild type MF1 male (MF1 mice typically have large litters) simply to generate more transgene positive offspring, for expression analysis and subsequent breeding. Transgene positive animals from this cross were then crossed with *ERCC1* heterozygotes from the *ERCC1* breeding colony. Transgene positive heterozygotes identified from these matings were then crossed to generate transgene positive *ERCC1* nulls. The animals on this background were given a TG# identification code. The *TTR/ERCC1* transgene was later crossed onto the second of our *ERCC1*-deficient strains (KT#209) by mating the female founder animal with a male heterozygote from that strain. Once again, transgene positive heterozygotes from this mating were then crossed to generate transgene positive *ERCC1* nulls. The animals on this background were given a KT/TG# identification code.

6.5 The transgene is present as a single array

The number of transgene arrays present in the founder animal was examined using Southern blot analysis. Genomic DNA isolated from a tail biopsy from TG#17 was cut with a range of restriction enzymes that were known not to cut within the transgene, including NotI, SalI, ClaI, XbaI and AatII. The DNA was separated by electrophoresis, transferred to a nylon membrane and the transgene array detected using an ERCC1 cDNA probe. As these restriction enzymes did not cut at sites within the transgene it meant that the enzymes would only cut at sites flanking the site of integration and each integration site would therefore produce a separate hybridising band on the Southern blot. The results of this analysis showed that only a single hybridising band was detected for each of the 5 restriction enzymes (data not shown). The interpretation of the Southern blots was not completely unambiguous; the use of rare cutting enzymes did not result in significant separation of the digested DNA by electrophoresis and as such it is possible that the single hybridising band, which was typically at the top of the DNA smear in each track, was actually present within uncut DNA. The use of pulse field gel electrophoresis (PFGE) would have enabled greater resolution of the high molecular weight DNA fragments.

However, further evidence that the transgene was present as a single array was gained from the genotyping of offspring from the first crosses between transgene positive *ERCC1* heterozygotes, set up to produce transgene positive nulls. The ratio of transgene positive to transgene negative offspring from these matings was 91:21 or 4.3:1 which does not appear to be strikingly different from the predicted ratio of 3:1, where a single transgene array is present. The predicted ratio for the presence of two segregating arrays would have been 15:1, assuming both parents in these initial breeding pairs each had two arrays. However, as a consequence of the breeding scheme used to generate the transgene positive *ERCC1* heterozygotes, it was possible that the parents in these crosses had either one or two arrays. Together, with the Southern blot observations, these results indicate that there is most likely only a single transgene integration site in the founder animal.

6.6 Transgene copy number determination by Southern blot analysis

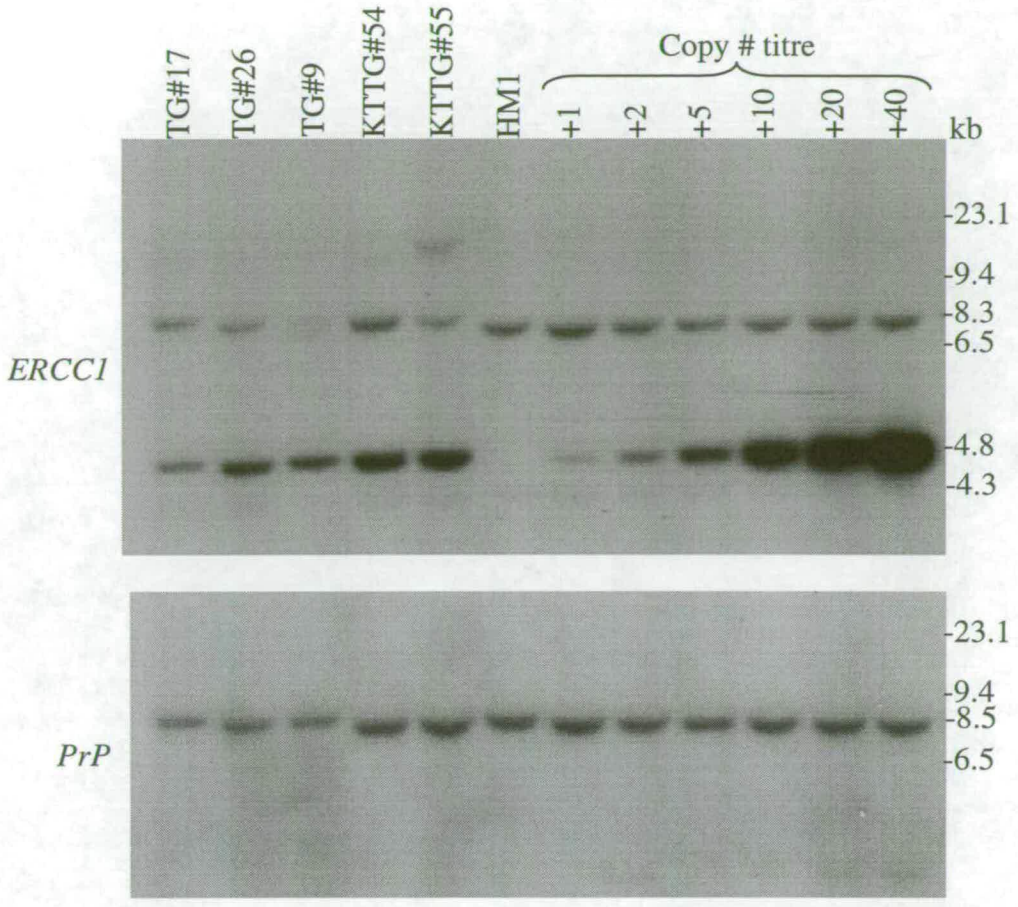
It has been reported that the expression of the *TTR* transgene is detected in tissues other than the liver if present in more than six copies. Southern blot analysis was used to determine the number of copies of the transgene integrated into the genome of the founder animal and F1 offspring from each of the crosses onto knockout backgrounds, each hemizygous for the transgene array.

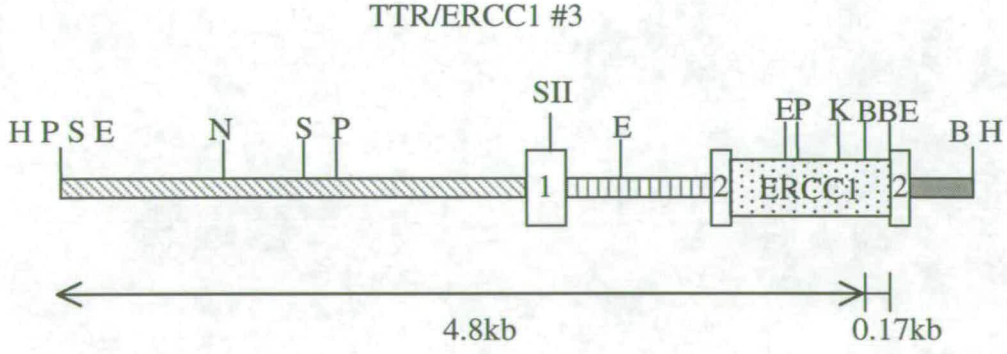
The mouse genome size is estimated to be $3x10^9$ bp and the transgene fragment was known to be 5.4kb. Using these figures it was possible to calculate the ratio of the two sizes, the transgene is roughly equivalent to 1/555,000 of the genome, and subsequently calculate the amount of transgene fragment equivalent to a single copy per genome for a 3µg amount of genomic DNA (1/555000 of 3µg is 5.4pg). In order to gauge the transgene copy number, 3µg mouse ES cell genomic DNA was spiked with various amounts of the transgene fragment, equivalent to 1, 2, 5, 10, 20 and 40 copies per mouse genome. The concentration of the ES cell DNA could be determined accurately by its absorbence at 260nm, however the tail biopsy DNA is heavily contaminated with RNA and cannot be accurately quantified in this manner. Approximate loads of 3µg were obtained by adjusting the amount of DNA digested so that the transgenic sample digests matched the appearance of the ES cell lanes, by

Figure 6.3 Transgene copy number determination by Southern blot analysis

Transgene copy number estimate by Southern blot analysis. Genomic DNA was prepared from tail biopsies of the founder animal (TG#17) and four F1 offspring (TG#26 and 9, KTTG#54 and 55). Genomic DNA isolated from mouse embryonic stem cells (HM1) was spiked with pre-determined amounts of transgene fragment DNA, equivalent to 1, 2, 5, 10, 20 and 40 copies of the transgene per genome. Upper panel, the DNA (3µg) was digested with *BamHI*, electrophoresed through an agarose gel, transferred onto a nylon membrane and probed with an *ERCC1* cDNA fragment. Lower panel, the blot was then stripped and reprobed using a mouse *PrP* probe to standardise for the amount of DNA in each track. The sizes of each of the hybridising bands are indicated (kb).

The transgene fragment is represented below, schematically. The region of the fragment corresponding to the 4.8kb hybridising transgene band is indicated. Shown to scale, 1cm = 500bp.


TG#17
TG#26
TG#9
KTTG#54
KTTG#55
HM1
Copy # titre
+1
+2
+5
+10
+20
+40
kb
23.1
9.4
8.3
6.5
ERCC1
4.8
4.3
23.1
9.4
8.5
6.5
PrP


TTR/ERCC1 #3
H P S E
N
S P
SII
E
EP K B B E
B H
1
2
ERCC1
2
4.8kb
0.17kb

eye. The copy number titration mixes and 3µg of each of the transgenic DNA samples were then digested using BamHI. Digestion of the transgene DNA with BamHI gave rise to two bands of 4.8 and 0.17kb, containing *ERCC1* homology, both of which could be used to assess the copy number using Southern blot analysis (Figure 6.3). The digested DNA was separated by electrophoresis, transferred to a nylon membrane and *ERCC1* sequences detected by hybridisation to an *ERCC1* cDNA probe. Due to the poor resolution of the internal 0.17kb band (data not shown), all copy number calculations were based on data obtained from the 5' end 4.8kb fragment. Although the 4.8kb band was not an internal fragment, the normal head to tail transgene integration pattern and the presence of a *BamHI* site at the 3' end of the fragment, meant the band was detected from the chromosomal array. However, the 5' end copy of each transgene array would result in a 'junction fragment' and would not give rise to a band of the predicted size and as such any copy number calculations based on this fragment will have an in-built error of 1 copy. The Southern blot analysis showed the presence of the endogenous 8.3 and weaker 4.3kb *ERCC1* band in each track. The transgenic sample TG#9 was an *ERCC1* heterozygote and consequently an additional band was evident at 6.7kb and a previously identified restriction fragment length polymorphism for the 8.3kb band was evident in TG#55 sample. Therefore, the endogenous *ERCC1* bands were not considered as suitable for internal standards. The transgene specific band of 4.8kb was strongly evident in each of the transgenic sample lanes and at appropriately varying intensities in the copy number titration lanes. To provide a standard for the loading of DNA in each track the blot was then stripped and reprobed with a *PrP* probe. Phosphorimagery was used on both blots to quantify the level of the transgene signal relative to the 8.5kb *PrP* band (Figure 6.4). It was then possible to calculate the ratio for each of the transgene signals to the *PrP* signals and compare it directly with the ratio of spiked transgene fragments to the *PrP* signals from the copy number titration. For instance, the *ERCC1* transgene/*PrP* ratio for 2 spiked copies was 0.76, the ratio for 5 spiked copies was 1.54, the ratio for the founder TG#17 was 0.65 whilst the ratios for the F1 offspring ranged from 1.01 to 1.4. The data from the phosphorimager analysis was used to plot a standard curve for the quantification of

Figure 6.4 Phosphorimage analysis of transgene copy number determination Southern blots

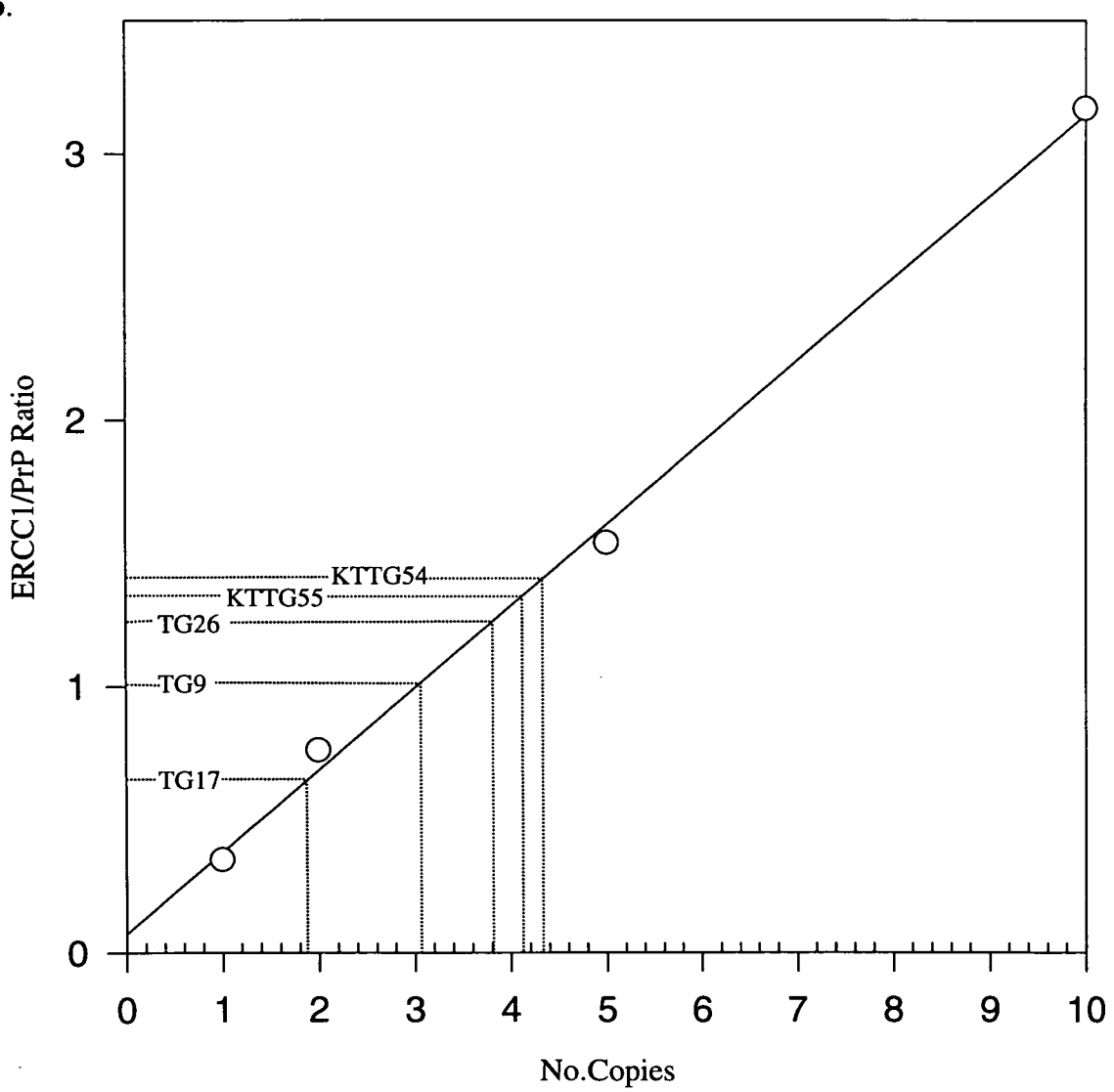
a, transgene copy number determination by phosphorimager analysis. The values from the phosphorimage were obtained using the ImageQuant software V3.3 image analysis package. The signal values represent the area, in arbitrary units, under the peak corresponding to the 4.8kb *TTR/ERCC1* bands and the 8.5kb *PrP* bands on the reprobed filter. The ratio represents the value of the transgene signal divided by the value of the *PrP* signal, i.e. the standardised transgene values. The copy number estimate was based on this standardised value relative to the values obtained for the spiked copy number titration.

b, ratio of *ERCC1/PrP* plotted against copy number for spiked copy titration. The *ERCC1/PrP* ratio for the spiked copy number samples was plotted against the number of spiked copies. Data points are represented by open circles, the best fit line is indicated. The relative positions of the equivalent *ERCC1/PrP* ratios are shown and extended down onto the copy number axis to determine actual copy numbers for each of the samples.

a.

Sample	TG Signal	PrP Signal	TG / PrP Ratio	Copy #
HM1	-	168.7	-	
+1	63.73	180.5	0.35	1
+2	111.5	146.1	0.76	2
+5	215.2	139.7	1.54	5
+10	531.8	167.5	3.17	10
+20	1025	150.1	6.83	20
+40	1798	169.5	10.6	40
17	75.1	115.8	0.65	2
26	146.6	117.8	1.24	4
9	111.8	110.8	1.01	3
54	221.2	165.6	1.34	4
55	231.1	165.3	1.40	4

b.



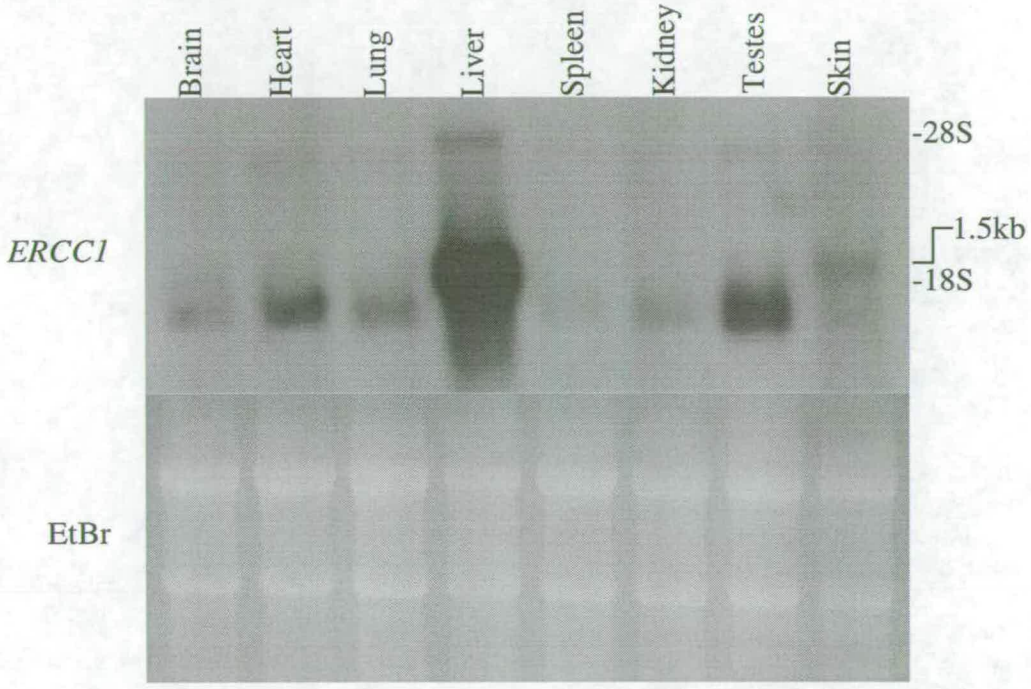
transgene copy number. The *ERCC1/PrP* ratio (i.e. the ratio of the spiked number of copies standardised for DNA load) plotted against copy number gave a linear result (Figure 6.4b). It was then possible to determine the copy number for each sample by using the previously determined *ERCC1/PrP* ratio for each of the transgenic samples to derive a copy number estimate from the standard curve of the copy number titration. This analysis gave the estimated copy numbers as: 2 copies for TG#17, 3 copies for TG#9 and about 4 copies for each of TG#26, KTTG#54 and KTTG#55.
Each of the copy number estimates suggested that the F1 offspring had a higher copy number (~4) than the founder TG#17 (~2). This may indicate that the founder animal was a transgenic mosaic. If chromosomal integration of the transgene array occurred at the two cell stage of development, only half of the cells in the founder animal would actually contain the transgene sequences and consequently it would appear to have only half of the copies present in its offspring.

6.7 Expression of *TTR/ERCC1* transgene mRNA is detected in the liver

To determine whether the integrated copies of the *TTR/ERCC1* transgene were being expressed, northern blot analysis was carried out on a panel of tissues from a transgene positive *ERCC1* wild type (F1) mouse (Figure 6.5). The estimated size for the *TTR/ERCC1* transgene transcript was 1.5kb. Analysis was carried out on RNA isolated from brain, heart, lung, liver, spleen, kidney, testes and skin. Total RNA (30μg) was electrophoresed on a 1.4% agarose-formaldehyde gel, transferred onto a nylon membrane and *ERCC1* transcripts detected by hybridisation to a mouse *ERCC1* cDNA probe.
The ethidium bromide staining of the gel is presented in the lower panel and suggests that the amount of RNA loaded in each of the tracks was equivalent. The detection of *ERCC1* hybridising mRNAs is shown in the upper panel. The normal 1.1kb band was evident at low levels in each of the tissues examined and the additional 1.5kb skin specific transcript was also detected. The endogenous level of the 1.1kb *ERCC1* mRNA is normally lowest in the liver. In this instance, there was an additional mRNA species of about 1.5kb, believed to be the transgene transcript, detected at extremely high levels in the liver. At this stage of analysis it appeared that the

Figure 6.5 Detection of *TTR/ERCC1* transgene mRNA expression in the liver using an *ERCC1* cDNA probe

Northern blot analysis of tissue RNA from a transgene positive *ERCC1* wild type male. Upper panel, total RNA was isolated from a range of mouse tissues, electrophoresed on a 1.4% formaldehyde-agarose gel, transferred onto a nylon membrane and probed with an 800bp BamHI fragment of the mouse *ERCC1* cDNA corresponding to exons 1 to 8. Positions of the 18S and 28S rRNAs and estimated transcript size are shown. Lower panel, the ethidium bromide staining to illustrate RNA loads.


Brain
Heart
Lung
Liver
Spleen
Kidney
Testes
Skin
-28S
1.5kb
-18S
ERCC1
EtBr

TTR/ERCC1 transgene was being expressed at high levels specifically in the liver. However, the heterogeneous nature of the endogenous *ERCC1* transcript in brain and heart along with the presence of the 1.5kb transcript in skin meant it was not possible to conclude that the transgene was not being expressed in these tissues.

To investigate the tissue specificity of the transgene expression further, it was necessary to isolate a probe specific for the transgene. A 300bp StuI/BamHI fragment was isolated from the parental transgene plasmid TTR ExV3. The fragment spanned the second *TTR* exon and contained the SV40 derived 3' end sequences and it was predicted that the probe would hybridise with endogenous *TTR* transcripts as well as transgene transcripts. To test the specificity of the fragment for *TTR* transcripts, northern blot analysis was performed on a panel of wild type mouse tissues. Total RNA (30μg) from a range of tissues was separated by electrophoresis, transferred onto a nylon membrane and probed using the 300bp StuI/BamHI TTR ExV3 fragment (Figure 6.6a). The northern blot showed that the probe hybridised to distinct transcripts of 0.7kb in the liver and brain. The hybridising mRNA species was much more abundant in liver than in brain. The sizes of the transcripts and pattern of expression both matched the expectations for the normal pattern of *TTR* expression. Thus it appeared that the fragment was suitable for use as a probe for the transgene transcripts which would also contain the *TTR* exon two sequences.

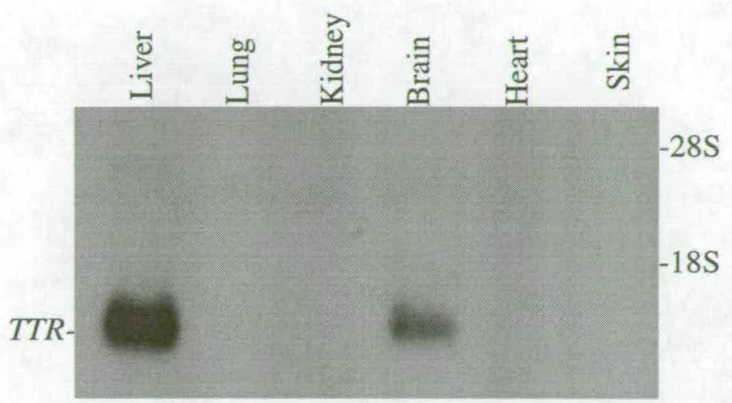
The StuI/BamHI fragment was then used for northern blot analysis of the *ERCC1* wild type transgene positive tissue RNAs. Total RNA (30μg) was separated by electrophoresis, transferred to a nylon membrane and transgene transcripts detected by hybridisation to the 300bp StuI/BamHI *TTR*/SV40 probe (Figure 6.6b). The endogenous *TTR* transcripts, of 0.7kb, were present in only liver and brain and in their normal abundance. An additional transcript, of approximately 1.5kb, was evident in the liver track. There was no evidence of hybridising mRNA species of this size in any of the other lanes.

In summary, this analysis has shown that the *TTR/ERCC1* transgene is expressed specifically in the liver. The expression at the level of mRNA greatly exceeds the normal low level of *ERCC1* in the liver but is less than the endogenous *TTR* expression.

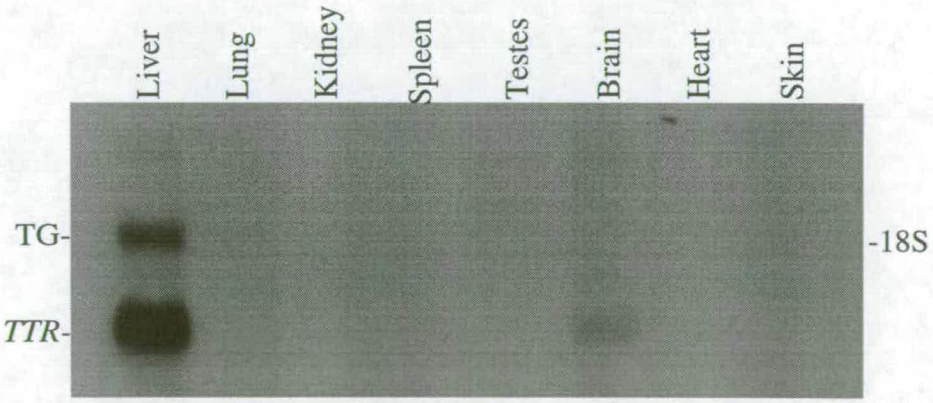
Figure 6.6 Identification of a probe specific for endogenous *TTR* and *TTR/ERCC1* transgene mRNA expression

a, total RNA was isolated from a range of wild type mouse tissues, electrophoresed on a 1.4% formaldehyde-agarose gel, transferred onto a nylon membrane and probed with a 300bp StuI/BamHI *TTR*/SV40 fragment from pTTR Ex V3. Positions of the endogenous *TTR* transcript, 18S and 28S rRNAs are shown. b, total RNA was isolated from a range of tissues from a transgene positive *ERCC1* wild type male mouse, electrophoresed on a 1.4% formaldehyde-agarose gel, transferred onto a nylon membrane and probed with a 300bp StuI/BamHI *TTR*/SV40 fragment from pTTR Ex V3. The positions of the endogenous *TTR* transcript, transgene transcript and 18S rRNA are indicated.

a.


Liver
Lung
Kidney
Brain
Heart
Skin
-28S
-18S
TTR-

b.


Liver
Lung
Kidney
Spleen
Testes
Brain
Heart
Skin
TG-
TTR-
-18S

6.8 Analysis of *TTR/ERCC1* expression on an *ERCC1*-deficient background

It was necessary to confirm the pattern of transgene expression on an *ERCC1*-deficient background. This analysis was carried out on tissues from a transgene positive, *ERCC1*-deficient mouse from each of the knockout backgrounds.

Initially, RNA was isolated from a panel of tissues from a transgene positive male on the original *ERCCI*-deficient background. Total RNA (30μg) from each of the tissues was separated by electrophoresis, transferred onto a nylon membrane and the transgene transcript detected by hybridisation to the 300bp StuI/BamHI TTR/SV40 probe (Figure 6.7a). The endogenous *TTR* transcript was detected in the liver and brain at expected levels. The transgene transcript was once more evident at high levels in the liver. Two additional and much larger species were also evident in this lane. These two additional transcripts may correspond to pre-mRNA species for the transgene and endogenous *TTR* transcripts. In the skin lane there was evidence of very weakly hybridising bands at the same sizes as the endogenous *TTR* and transgene transcripts. It was felt that this may have arisen as a consequence of contamination of the skin sample with liver RNA, at the time of isolation. In addition to the presence of these transcripts in skin there was also evidence of transcripts in the testes lane, again at much lower levels. In this case there was no detection of a transcript matching the size of the endogenous *TTR* mRNA and as such it is unlikely that the hybridising bands have arisen from contaminating liver RNA.

This analysis was then repeated using samples from a transgene positive *ERCC1*-deficient mouse on the KT#209 knockout background. A panel of tissues was analysed by northern blot analysis to investigate further the pattern of transgene expression (Figure 6.7b). Total RNA (30μg) from each of the tissues was separated by electrophoresis, transferred onto a nylon membrane and the transgene transcript detected by hybridisation to the 300bp StuI/BamHI *TTR*/SV40 probe. Lower panel, the endogenous *TTR* transcript was detected in the liver and brain at expected levels. The transgene transcript was once more evident at high levels in the liver and at a higher level than that evident in the upper panel. As with the transgene positive sample from the original background, two additional and much larger species were also evident in this lane. At this stage there was no evidence of transgene expression

Figure 6.7 Expression of the *TTR/ERCC1* transgene mRNA on two *ERCC1*-deficient backgrounds

a, total RNA from a range of tissues was isolated from a transgene positive *ERCC1*-deficient male mouse on the original *ERCC1*-deficient background, electrophoresed on a 1.4% formaldehyde-agarose gel, transferred onto a nylon membrane and probed with a 300bp StuI/BamHI *TTR*/SV40 fragment from pTTR Ex V3. Positions of the endogenous *TTR* transcript, transgene transcript, 18S and 28S rRNAs are shown. b, total RNA was isolated from a range of tissues from a transgene positive *ERCC1*-deficient male mouse on the KT#209 *ERCC1*-deficient background, electrophoresed on a 1.4% formaldehyde-agarose gel, transferred onto a nylon membrane and probed with a 300bp StuI/BamHI *TTR*/SV40 fragment from pTTR Ex V3. The positions of the endogenous *TTR* transcript (*TTR*), transgene transcript (TG) and 18S rRNA are indicated.

Liver
Lung
Kidney
Spleen
Testes
Brain
Heart
Skin
-28S
-18S
TG-
TTR-
TTR

Liver
Lung
Kidney
Spleen
Brain
Heart
Skin
-28S
-18S
TG-
TTR-
TTR

in any tissues other than the liver. However, after prolonged exposure of the *TTR* probed filter, there was evidence of bands, of about 1.5kb, in the skin and brain lanes. The size of the band matched the size of the transgene transcript in liver. Phosphorimagery was used to determine the amount of the transcript in these tissues relative to the liver. The result of this analysis showed that the signal intensity of the transcript in skin and brain was 1.46 and 2.9%, respectively, of the transgene signal intensity in liver.

In summary, it has been shown that expression of the *TTR/ERCC1* transgene was detected at high levels in the livers of *ERCC1*-deficient mice from both knockout backgrounds, with some indication of variation in the levels of expression between the two lines. There was also evidence of very low levels of expression in other tissues including brain, testes and skin. It is possible that the non-specific expression of the transgene may reflect variation between individual animals.

6.9 Transgene expression is stable with age and persists through four generations

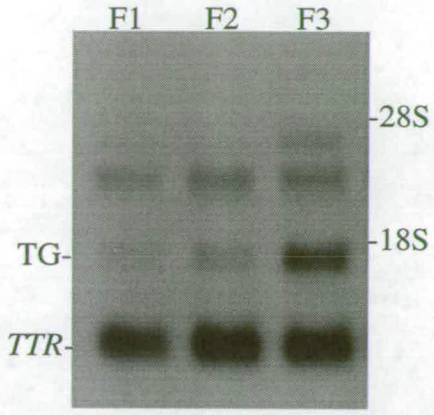
The aim of this work was to rescue the lethal liver phenotype associated with *ERCC1* deficiency, by the targeted expression of an *ERCC1* transgene. Simply crossing the transgene onto the original knockout background involved going through three generations of mice. It was important that the expression of the transgene was not silenced as it passed from generation to generation. The aim of targeting the expression of the transgene to the liver was primarily to extend the lifespan of the *ERCC1*-deficient mice to enable studies of the consequences of *ERCC1* deficiency in other tissues at later developmental stages. As such it was necessary to establish that the expression of the transgene was maintained in the liver throughout the life of an animal.

Northern blot analysis was used to confirm the stable expression of the transgene in the livers of mice from each of the first three generations (Figure 6.8a). Total RNA (30μg), from the livers of female mice from the first three generations after the founder, was separated by electrophoresis, transferred onto a nylon membrane and the transgene transcript detected by hybridisation to the 300bp StuI/BamHI

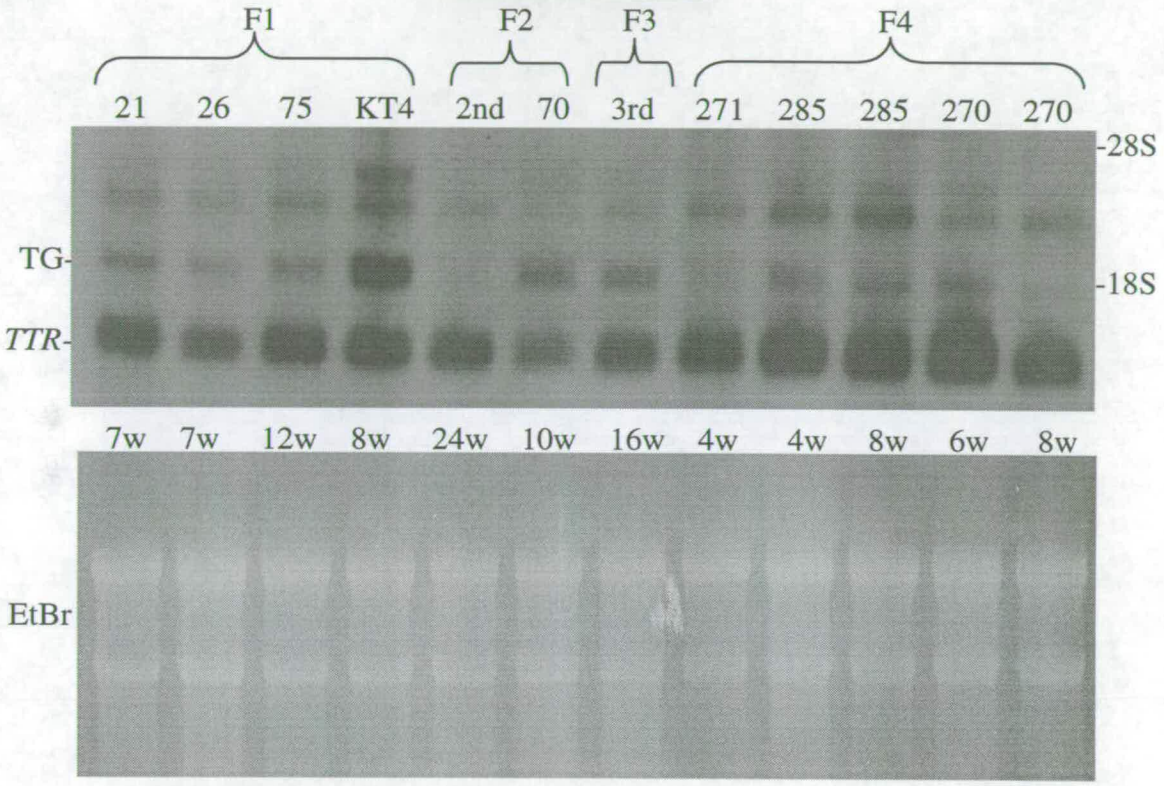
Figure 6.8 Expression of the *TTR/ERCC1* transgene mRNA is stable with age and through four generations

a, northern blot analysis of RNA from the livers of transgene positive animals from three different generations. Total RNA was isolated from liver tissue of animals representing three different generations: F1, F2 and F3. The RNA (30μg) was electrophoresed on a 1.4% formaldehyde-agarose gel, transferred onto a nylon membrane and probed with a 300bp StuI/BamHI *TTR*/SV40 fragment from pTTR Ex V3. Positions of the endogenous *TTR* transcript (*TTR*), transgene transcript (TG), 18S and 28S rRNAs are shown. b, northern blot analysis of RNA from the livers of transgene positive animals of various ages from four different generations. Total RNA was isolated from liver tissue of animals of various ages representing four different generations: F1, F2, F3 and F4. Upper panel, the RNA (30μg) was electrophoresed on a 1.4% formaldehyde-agarose gel, transferred onto a nylon membrane and probed with a 300bp StuI/BamHI *TTR*/SV40 fragment from pTTR Ex V3. The animal identification code is indicated above each lane and the age of the animal, in weeks, is shown below each lane. Positions of the endogenous *TTR* transcript (*TTR*), transgene transcript (TG), 18S and 28S rRNAs are shown. Lower panel, ethidium bromide staining to illustrate RNA loads. c, phosphorimager analysis of *TTR/ERCC1* transgene expression relative to endogenous *TTR* expression. The values from the phosphorimager were obtained using the ImageQuant software V3.3 image analysis package. The signal values represent the area, in arbitrary units, under the peaks corresponding to the transgene and endogenous *TTR* bands. The ratio represents the value of the transgene signal divided by the value of the endogenous *TTR* signal. The transgene mRNA expression is given as a percentage of the endogenous *TTR* expression.

a.



b.



c.

TG Sample	Genera-tion	Age (weeks)	*ERCC1* Status	*TTR/ERCC1* Signal	*TTR* Signal	*ERCC1/ TTR* Ratio	Expression (% endogenous)
#21	1	7	Wt	707.7	1595	0.44	44
#26	1	7	Wt	531.5	1220	0.44	44
#75	1	12	Wt	840	1888	0.44	44
KT #4	1	7	Wt	2375	2287	1.04	104
2ND	2	24	Wt	507.1	1648	0.31	31
#70	2	10	Δ	1114	1056	1.05	105
3RD	3	16	N/D	1081	1708	0.63	63
#271	4	4	Δ	829.5	2294	0.36	36
#285	4	4	Δ	1474	3521	0.42	42
#285	4	8	Δ	1374	3504	0.39	39
#270	4	6	Δ	1426	4894	0.29	29
#270	4	8	Δ	710.4	2446	0.29	29

TTR/SV40 probe. The endogenous *TTR* transcript was evident in its normal abundance in each of the tracks. The 1.5kb transgene transcript was also detected in each of the tracks, confirming that the transgene was expressed through generations. Where the endogenous *TTR* was regarded as a suitable indication of equal loading of RNA in each track, by eye, it appeared that the level of transgene expression had actually increased with each generation. At this point it was noted that the RNA was isolated from animals of different ages at each generation. The F1 sample was from a 32-week old female, the F2 sample was from a 24-week old female and the F3 sample was taken from a 16-week old female. It seemed possible that the level of transgene expression might have decreased as a function of age.

To investigate further whether the level of transgene expression was related to the age of the animal, the northern blot analysis was repeated with additional samples spanning four generations and incorporating animals of different ages from each generation (Figure 6.8b). Total RNA was isolated from the livers of three 7-week old and one 12-week old F1 animal, one 24-week old and one 10-week old F2 animal, a single 16-week old F3 animal and four F4 animals between 4 and 8-weeks old. In the case of two of the F4 animals, a partial hepatectomy was performed to enable analysis of the level of transgene expression within a single animal over a period of time. RNA was first isolated from the biopsied tissue and then again when the animal was culled after two or four weeks. The RNA was separated by electrophoresis, transferred onto a nylon membrane and the *TTR/ERCC1* transgene transcript detected by hybridisation to the 300bp StuI/BamHI *TTR*/SV40 probe. The ethidium staining of the gel is presented in the lower panel to illustrate the relative loading of RNA in each lane. The blot showed that the endogenous *TTR* transcript and the 1.5kb transgene transcript were evident in each of the lanes. By eye, the level of transgene expression generally appeared to be quite consistent across the samples. Phosphorimagery was used to quantify the level of transgene expression relative to the level of endogenous *TTR* expression for each sample (Figure 6.8c). Thus it was possible to express the level of transgene expression as a percentage of the endogenous *TTR* expression in each case.

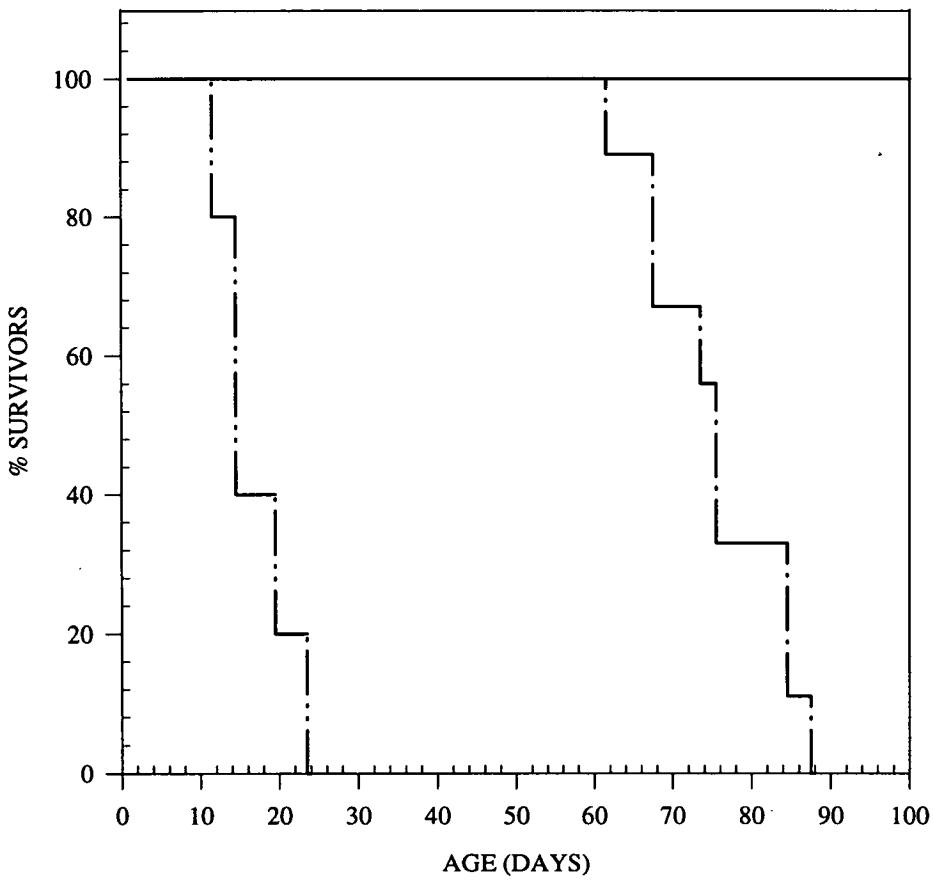
In the case of the F1 animals it was shown that the transgene expression was 44% of the endogenous *TTR* level in the RNA samples from the twelve-week and two of the 7-week old animals. The transgene level in the third 7-week old animal was seen to be 104% of the endogenous level. This sample was the only one taken from an animal on the KT#209 knockout background, it is possible that the elevated level of transgene expression seen in this sample reflects a difference in the genetic background. There was also some variation seen between the two F2 samples, the level of transgene expression in the 24-week old animal was 31% of the endogenous *TTR* compared to 105% of the endogenous level seen in the sample from the 10-week old animal. The level of transgene transcript in the 16-week old F3 animal was measured at 63% of the endogenous *TTR* level. The level of transgene expression in the F4 samples varied between 29 and 42% of the endogenous transcript. In the case of the samples taken from the same animals at different time points there was little evidence of any major changes in the level of transgene expression. In the samples taken only two weeks apart (at 4 and 6 weeks) the level of transgene expression was measured as 29% of the endogenous *TTR* level in both samples. In the samples taken four weeks apart (at 4 and 8 weeks) the level of transgene expression was measured as 42 and 39% of the endogenous *TTR* level, respectively. Thus whilst it was clear that there was some degree of variation in the absolute level of transgene expression across the samples (105 to 29% of the endogenous transcript), this variation did not appear to correlate with either the age of the animal or the generation from which the sample was taken. It is possible that the observed variation was a consequence of the number of transgene arrays present within an individual animal, it was not determined whether these animals were hemi- or homozygous for the transgene array.

6.10 Targeted *in vivo* expression of the *TTR/ERCC1* transgene in the liver extends the lifespan of *ERCC1*-deficient mice

ERCC1-deficient mice die severely runted before weaning as the result of liver failure. The premature death of these animals has significantly restricted the extent of any analysis we have been able to perform, relating to the consequences of

6.9 Improved survival profile of *TTR/ERCC1* transgene positive *ERCC1*-deficient mice

The percent of surviving mice is plotted as a function of time. Survival time was measured, in days, at the time when dying mice were sacrificed. Wild type, ———; *ERCC1*-deficient, — - - —; *TTR/ERCC1* transgene positive nulls, —— - ——.


100
80
60
40
20
0
% SURVIVORS
0
10
20
30
40
50
60
70
80
90
100
AGE (DAYS)

ERCC1 deficiency in other organs or at later stages of development. The principle aim of this work was to sufficiently extend the lifespan of these animals, by correction of the lethal liver phenotype, so as to permit these studies.

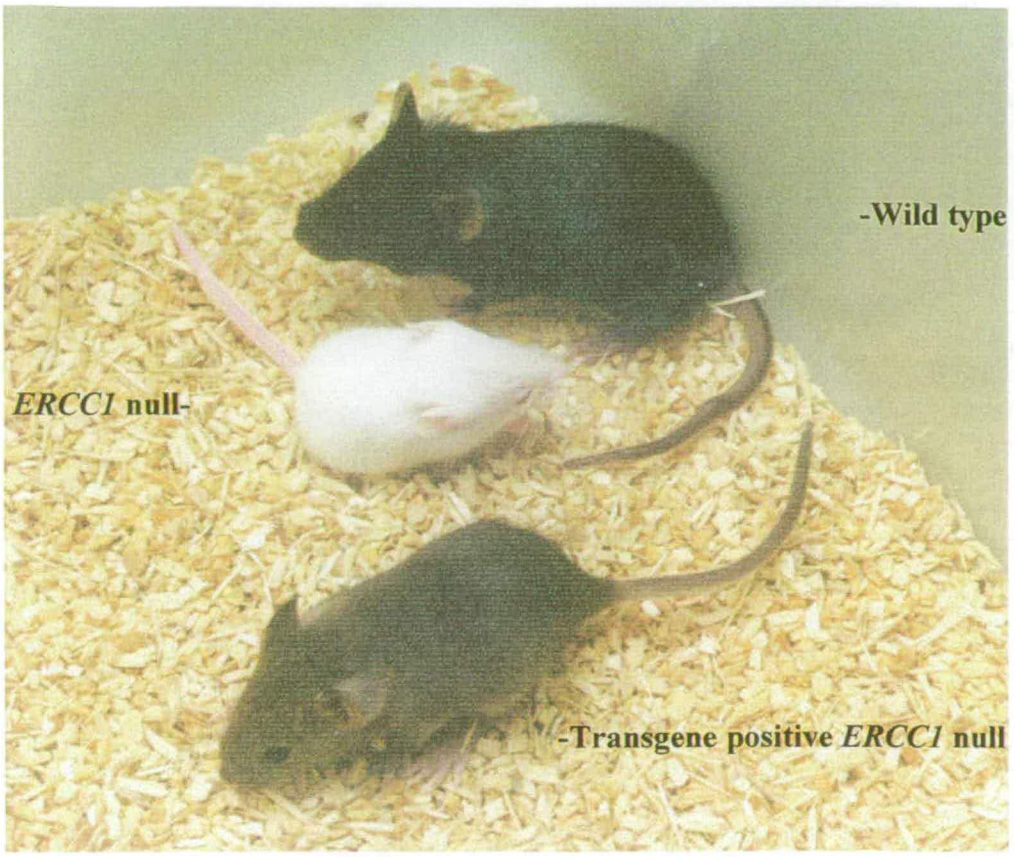
The improved survival profile of *TTR/ERCC1* transgene positive *ERCC1*-deficient mice relative to the transgene negative *ERCC1* nulls is shown in Figure 6.9. As previously reported (McWhir *et al.*, 1993), the *ERCC1* nulls (n=5) died before weaning at about 21 days of age. The survival profile of the *TTR/ERCC1* transgene positive *ERCC1* nulls (n=9) showed 100% survival up to 61 days of age, with all of the mice dying between day 61 and day 86. No wild type littermates (n=9) died within this period. The transgene positive nulls were seen to survive for three times as long their transgene negative null siblings, with some animals surviving as long as twelve weeks.

6.11 Targeted *in vivo* expression of the *TTR/ERCC1* transgene in the liver alleviates the severe runting phenotype observed in *ERCC1*-deficient mice

The severe runting associated with *ERCC1* deficiency in mice has been reported separately by McWhir *et al* (1993) and Weeda *et al* (1997). In both cases it was reported that *ERCC1*-deficient offspring were severely runted compared to their wild type siblings, with reported bodyweights ranging between 20 and 50% of the wild type weights. The *ERCC1*-deficient mice from our original knockout background weighed only 20% of their siblings. The bodyweights of *TTR/ERCC1* transgene positive *ERCC1*-deficient mice and their littermates, on this background, were measured at weaning (age 21-28 days). The average weight of a transgene positive *ERCC1*-deficient mouse was found to be 7.75g (±0.56) compared to 13.25g (±0.3) for wild type littermates. The bodyweight in terms of percentage of wild type was measured as 58% for the transgene positive *ERCC1*-deficient mice compared to the previously reported 20% for the knockout mice. Figure 6.10 shows the relative sizes of a *TTR/ERCC1* transgene positive *ERCC1* null, a wild type sibling and a transgene negative *ERCC1* null littermate. The clear size difference meant that transgene positive nulls were readily distinguished from transgene negative null littermates, by eye.

Figure 6.10 The severe runting phenotype associated with *ERCC1* deficiency is alleviated in *TTR/ERCC1* transgene positive nulls

A representative photograph of a *TTR/ERCC1* transgene positive *ERCC1*-deficient mouse, a wild type littermate and a transgene negative *ERCC1* null sibling at 20 days of age. The wild type mouse (black) is shown at the top of the photo, the transgene positive mouse(agouti) is shown at the bottom of the photo and the *ERCC1* null (white) is in the middle of the group The transgene positive mouse is readily distinguishable from its *ERCC1*-deficient littermate on the basis of the size difference.


-Wild type
ERCC1 null-
-Transgene positive *ERCC1* null

6.12 Targeted *in vivo* expression of the *TTR/ERCC1* transgene in the liver corrects the nuclear abnormalities associated with *ERCC1* deficiency

We have previously reported that *ERCC1*-deficient mice die as a consequence of liver failure. Histologically, perhaps the most striking phenotype reported for the *ERCC1* nulls was the nuclear abnormalities seen in the livers. Identical observations have been made for each of the known *ERCC1* knockout mice (McWhir *et al.*, 1993; Weeda *et al.*, 1997 and Melton, unpublished data). At three weeks of age the liver nuclei in wild type mice are predominantly diploid and uniform in size. *ERCC1* null mice at this age showed considerable variation in the sizes of hepatocyte nuclei, and flow cytometry analysis indicated the presence of aneuploid nuclei. We speculated that the incidence of abnormal nuclei in hepatocytes would be corrected by the targeted expression of the *TTR/ERCC1* transgene in the liver of an *ERCC1*-deficient animal.

In order to assess whether this phenotypic correction had indeed occurred, liver sections were taken from age matched transgene positive *ERCC1*-deficient animals and wild type littermates for comparison with sections taken from *ERCC1*-deficient animals on our original knockout background. The haematoxylin and eosin stained sections are presented in Figure 6.11. Similar results were seen for transgene positive nulls from both of our knockout backgrounds. The liver section representative of the *ERCC1* null animals confirmed the previously reported incidence of variable nuclear sizes, including the presence of some extremely large nuclei. In comparison, the nuclear sizes seen in the liver section representative of a wild type mouse showed greater uniformity in the sizes of the hepatocyte nuclei. Similarly, the liver section representative of the *TTR/ERCC1* transgene positive *ERCC1*-deficient animals showed much greater uniformity in the sizes of the hepatocyte nuclei compared with the *ERCC1* null sections. There was no evidence of the extremely large nuclei characteristically observed in the hepatocytes of *ERCC1*-deficient mice.

By way of confirming these histological observations, fluorescence activated cell scanning analysis (FACScan) was performed on liver biopsies from wild type, *ERCC1* null and transgene positive *ERCC1* null animals (Figure 6.11). The FACScan profiles represent the number of cells plotted against the DNA content of the nuclei

Figure 6.11 Targeted *in vivo* expression of the *TTR/ERCC1* transgene in the liver corrects the nuclear abnormalities associated with *ERCC1* deficiency

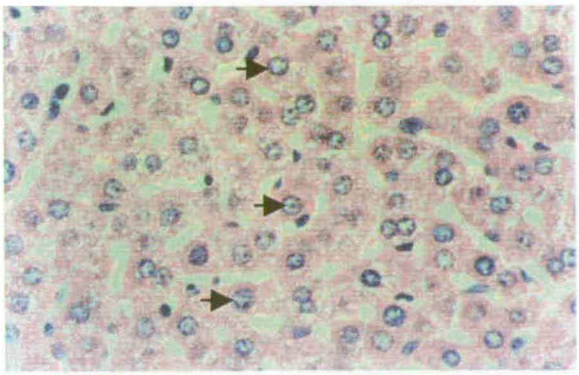
Comparison of the hepatocyte nuclear sizes and FACScan profiles between age matched wild type, *ERCC1* null and transgene positive *ERCC1* null mice.

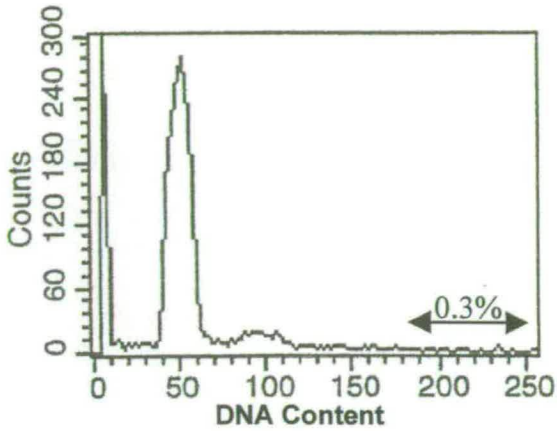
a, Left panel, haematoxylin and eosin stained liver section from a 3-week old wild type mouse. Representative uniformly sized nuclei indicated by arrows. Right panel, FACScan profile showing frequency distribution of DNA content of hepatocyte nuclei from a 3-week old wild type animal. The % of total nuclei with elevated DNA content (the region indicated by the double-headed arrow) is also shown.

b, Left panel, haematoxylin and eosin stained liver section from a 4-week old transgene positive *ERCC1* null mouse. Representative uniformly sized nuclei indicated by arrows. Right panel, FACScan profile showing frequency distribution of DNA content of hepatocyte nuclei from a 4-week old wild type animal. The % of total nuclei with elevated DNA content (the region indicated by the double-headed arrow) is also shown.

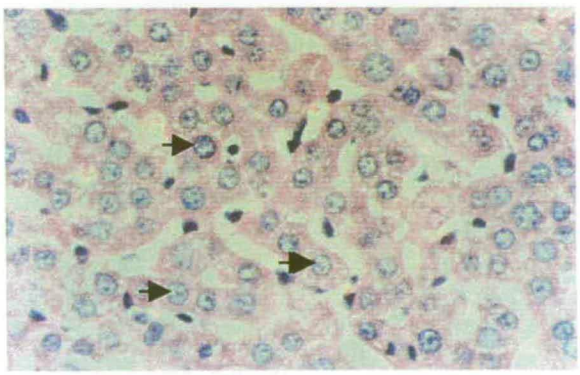
c, Left panel, haematoxylin and eosin stained liver section from 3-week old *ERCC1* null mouse. Characteristic abnormally large nuclei indicated by arrows. Right panel, FACScan profile showing frequency distribution of DNA content of hepatocyte nuclei from a 3-week old *ERCC1* null animal. The % of total nuclei with elevated DNA content (the region indicated by the double-headed arrow) is also shown.

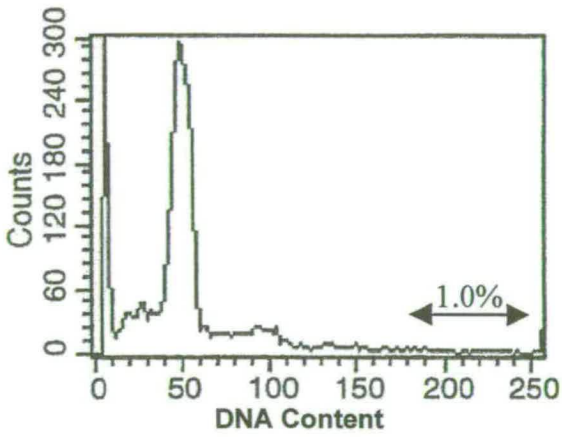
a.



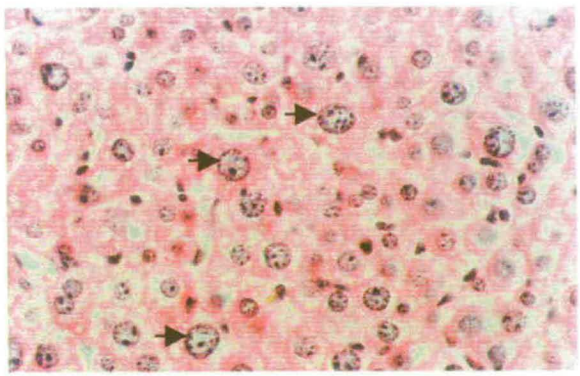

Counts
0 60 120 180 240 300
0.3%
0 50 100 150 200 250
DNA Content

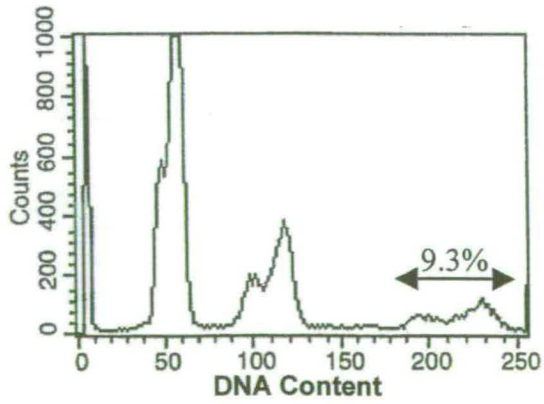
b.




Counts
0 60 120 180 240 300
1.0%
0 50 100 150 200 250
DNA Content

c.




Counts
0 200 400 600 800 1000
9.3%
0 50 100 150 200 250
DNA Content

as measured by the fluorescence of the propidium iodide stained nucleic acid. The FACScan profile representative of the wild type animals showed a major peak corresponding to normal diploid cells (cells at G1) a lesser peak representing cells that had undergone replication (cells at G2) was also evident, there was little evidence of any significant number of cells with DNA content incompatible with these cell cycle stages. This profile was all but duplicated by analysis carried out on liver tissue from the transgene positive *ERCC1*-deficient animals. The representative profile from this genotype also showed a major peak representing the hepatocytes with normal diploid DNA content, a lesser G2 peak was again evident. Furthermore, there was little evidence of cells with increased DNA content. These profiles contrast sharply with the profile representative of the *ERCC1* null animals. In addition to the major peak representing the normal diploid hepatocytes in G1, there were additional sizeable peaks corresponding to an increased number of tetraploid nuclei in G1 or diploid cells in G2 as well as peaks suggestive of both higher ploidy and aneuploid nuclei. The percentage of the total cells counted, falling into this region of greatly increased DNA content (8N plus) was determined to enable direct comparison between the respective FACScan profiles representative of each genotype. It was found that only 0.3 and 1.0% of the nuclei counted from wild type and transgene positive null samples had this elevated level of DNA content compared with 9.3% of nuclei from the *ERCC1* nulls. This result supports the histological observations of increased incidence of abnormally large nuclei in the hepatocytes of *ERCC1*-deficient mice and the correction of this phenotype, to wild type appearance, in the livers of *TTR/ERCC1* transgene positive *ERCC1* nulls.

In summary, the targeted expression of *ERCC1* in the hepatocytes of *ERCC1*-deficient mice results in the correction of the incidence of abnormally sized nuclei to apparent wild type levels.

6.13 Targeted *in vivo* expression of the *TTR/ERCC1* transgene in the liver does not correct the associated abnormalities in the skin

The aim of the work described in this chapter was to alleviate the phenotypic consequences associated with *ERCC1* deficiency in mice by the targeted expression

of an *ERCC1* transgene in the liver. We speculated that the restoration of expression of *ERCC1* to the liver would extend the limited lifespan of the *ERCC1* nulls sufficiently to permit further analysis of the consequences of *ERCC1* deficiency in tissues such as the skin. We have shown that the lifespan of the transgene positive *ERCC1* nulls was significantly extended relative to the *ERCC1*-deficient animals and, as such, the phenotypically rescued animals could potentially permit further analysis of any consequences of *ERCC1* deficiency in adult mouse skin. However, as described in chapter 6.7, we have detected evidence of very low levels of transgene mRNA expression in the skin of some transgene positive animals. If this expression was sufficient to bring about complete phenotypic correction in the skin it would preclude any further studies of the consequences of *ERCC1* deficiency in this tissue and systems relating to it.

In order to determine whether such phenotypic correction had occurred in the transgene positive *ERCC1* nulls, skin sections were taken from wild type mice and transgene positive nulls on both of our knockout backgrounds. The absence of subcutaneous fat from the dorsal skin of *ERCC1*-deficient mice has previously been reported (Weeda *et al.*, 1997 and Melton unpublished data). In this instance, for ease of dissection the skin biopsies were taken from the ears of the animals. The structure of the skin on the ears differs from body skin in that it lacks this subcutaneous layer of fat. The haematoxylin and eosin stained skin sections are presented in Figure 6.12. The wild type sections showed the normal thickness of the dermal layer and that the nuclei in the epidermal cells were of a uniform circular shape. The sections from the transgene positive *ERCC1*-deficient animals were both very similar. The nuclei within the epidermal layer appeared distorted as if they had been compressed in some way and the thickness of the dermal layer was reduced by approximately 50%.

In summary, despite the northern blot evidence, presented earlier, which suggested that there maybe expression of the transgene in skin, we have shown that the skin of the transgene positive *ERCC1*-deficient animals is not normal. This suggests that even if there is expression of the transgene in skin it is at such a low level so as not to be able to bring about phenotypic correction in this tissue. As such we consider that the skin of the transgene positive nulls can be regarded as *ERCC1*-deficient and

Figure 6.12 The phenotypic consequences of *ERCC1* deficiency in the skin are not corrected in *TTR/ERCC1* transgene positive *ERCC1* null mice

Comparison of skin sections between age matched wild type, *ERCC1* null and transgene positive *ERCC1* null mice.

a, haematoxylin and eosin stained skin section from a 7-week old wild type mouse. Large circular nuclei evident in the epidermal cells (E). Normal thickness of the dermal layer is indicated (D). The cartilaginous ear tissue is shown (C).

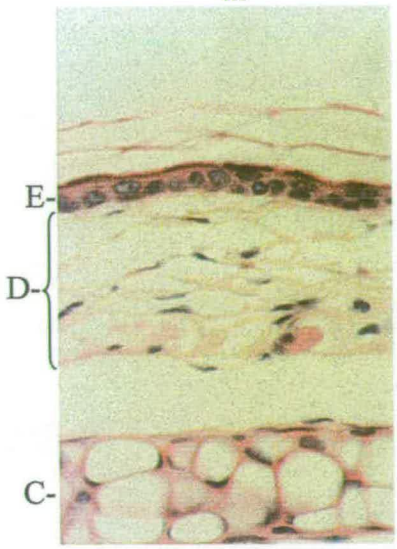
b, haematoxylin and eosin stained skin section from a 3-week old *ERCC1* null on the original knockout background. Large circular nuclei evident in the epidermal cells (E). Normal thickness of the dermal layer is indicated (D). The cartilaginous ear tissue is shown (C).

c, haematoxylin and eosin stained skin section from a 3-week old *ERCC1* null on the KT#209 knockout background. Large circular nuclei evident in the epidermal cells (E). Normal thickness of the dermal layer is indicated (D). The cartilaginous ear tissue is shown (C).

d, haematoxylin and eosin stained skin section from a 6-week old transgene positive *ERCC1* null on the original knockout background. Compressed nuclei evident in the epidermal cells (E). Thinning of the dermal layer is also evident (D). The cartilaginous ear tissue is shown (C).

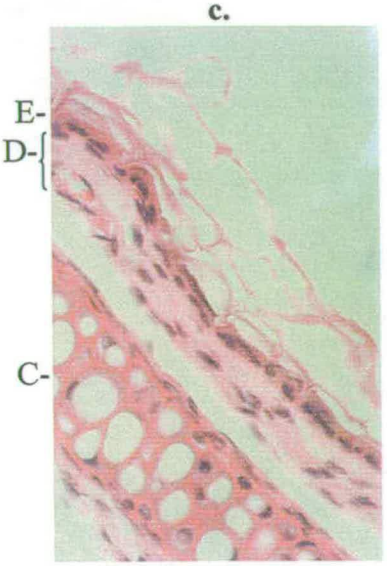
e, haematoxylin and eosin stained skin section from a 6-week old transgene positive *ERCC1* null on the KT#209 knockout background. Compressed nuclei evident in the epidermal cells (E). Thinning of the dermal layer is evident (D). The cartilaginous ear tissue is shown (C).

a.

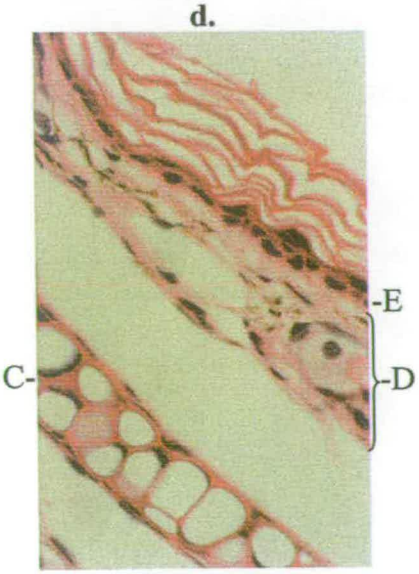


b.

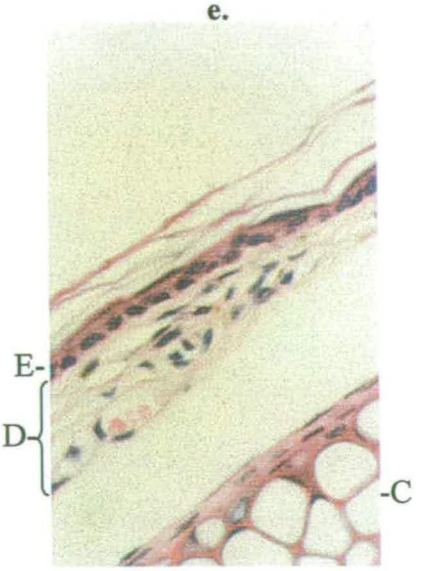
c.



d.



e.



is thus suitable for further studies of the consequences of *ERCC1* deficiency in this tissue and systems relating to it.

6.14 *TTR/ERCC1* transgene positive *ERCC1*-deficient mice exhibit kidney abnormalities

We have reported that the transgene positive *ERCC1* nulls live for up to 3 times as long as the transgene negative *ERCC1*-deficient animals. This extended lifespan has enabled us to carry out further analysis of the consequences of *ERCC1* deficiency in adult tissues. This study was carried out by sectioning the organs of transgene positive *ERCC1* nulls for comparison with sections from wild type littermates. The initial characterisation of our original *ERCC1* knockout line revealed detectable p53 protein in brain and medullary collecting tubules of the kidneys in these *ERCC1*-deficient animals (Mcwhir *et al.*, 1993). At this stage, despite the detectable p53, there was no evidence of any physical abnormalities in these 3-week old animals. We have been able to re-examine these organs in our phenotypically rescued transgene positive animals. For the first time we have been able to observe the consequences of DNA repair deficiency in the brains and kidneys of adult mice (aged 6 weeks and over). Comparison of the brain sections from our transgene positive nulls, from both knockout backgrounds, and age matched wild type littermates revealed no differences (data not shown).

Comparison of the kidney sections from our transgene positive nulls, from both knockout backgrounds, and age matched wild type littermates did reveal some changes in renal morphology (Figure 6.13). There was evidence of some abnormal enlarged nuclei, possibly polyploid, in cells of transgene positive mice on the original knockout background. Similar aberrant nuclei were identified in the animals on the KT#209 background to a lesser degree (not shown). Whilst there was some evidence of dilation of the renal tubules in the wild type sections, the tubule dilation appeared greater and more widespread in the transgene positive nulls from both knockout backgrounds. Furthermore there was evidence of tubules apparently blocked with protein casts, and a pyelonephritis like pattern. Pyelonephritis is a human condition in which the kidney becomes inflamed as the result of bacterial infection, treatment is

Figure 6.13 *TTR/ERCC1* transgene positive *ERCC1*-deficient mice exhibit kidney abnormalities

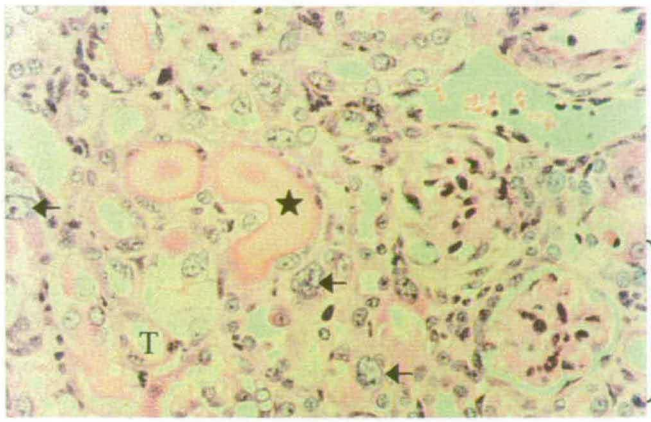
Comparison of kidney sections between age matched wild type, *ERCC1* null and transgene positive *ERCC1* null mice.

a, haematoxylin and eosin stained kidney section from a 6-week old transgene positive *ERCC1* null on the original background . A representative dilated renal tubule is indicated (T). Tubules apparently blocked with protein casts are shown (★). A representative glomerulus is indicated (G). Representative aberrant nuclei are indicated by arrows.

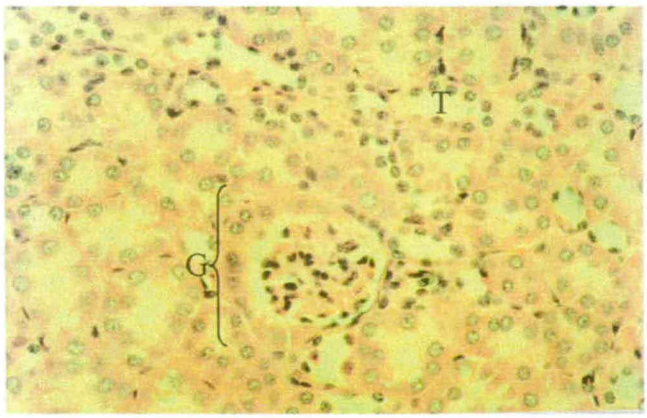
b, haematoxylin and eosin stained kidney section from a 7-week old wild type animal. A representative tubule is indicated (T). A representative glomerulus is indicated (G)

c, haematoxylin and eosin stained kidney section from a 6-week old transgene positive *ERCC1* null on the KT#209 knockout background. A representative dilated renal tubule is indicated (T). A representative glomerulus with abnormal structure is indicated (G)

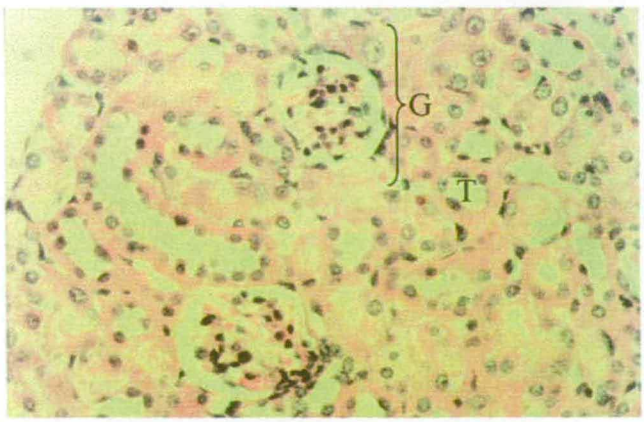
a.



b.



c.



generally by antibiotics. Rarely, when the infection is chronic, renal failure can follow. The appearance of the glomeruli in the transgene positive nulls was less ordered than in the sections from the wild type animals.

In addition to these histological observations it was also noted that, towards the latter stages of their lives, the abdomens of the transgene positive animals appeared distended. *Post mortem* examination of these animals revealed ascites, an accumulation of fluid in the peritoneal cavity. In humans ascites is commonly associated with failure of the heart or kidneys, as fluid leaks away from the circulating blood.

In summary, we have identified abnormalities in the kidneys of our adult transgene positive *ERCC1* nulls. Presumably, these abnormalities were not detected in our original knockout mice as they reflect the consequences of prolonged *ERCC1* deficiency through later stages of life not reached by the knockout mice. The *post mortem* observation of ascites in these animals may indicate renal failure, however, no assessment of renal function has been performed.

6.15 Discussion

Conventional transgenesis, by the pronuclear injection of cloned gene constructs, has enabled the study of the phenotypic effects of altered gene expression. Typically research has focussed on the effects of overexpression of genes (e.g. Wang *et al.*, 1998), the consequences of engineered changes within the control elements (e.g. Yan *et al.*, 1990) or the consequences of gene expression in an unusual cell type (e.g. Albers *et al.*, 1995). We have exploited the ability to target the expression of a transgene to a particular cell type by targeting the expression of *ERCC1* to hepatocytes.

The lethal liver phenotype of our *ERCC1* knockout mice has prevented us from studying the consequences of *ERCC1* deficiency in other tissues. We speculated that the targeted *in vivo* expression of *ERCC1* in the liver of our *ERCC1* knockout mice would extend the lifespan of these otherwise *ERCC1*-deficient animals. To bring about the targeted expression of *ERCC1* within the liver of mice, we constructed an *ERCC1* transgene by cloning the mouse *ERCC1* cDNA coding region into the second

exon of a transthyretin gene fragment, containing 3kb of transcriptional regulatory sequences within the promoter and the first two exons. This fragment was known to elicit high levels of hepatocyte specific transgene expression, with expression detectable in the choroid plexus of the brain only when 6 or more copies of the transgene were present (Yan *et al.,* 1990). Pronuclear injections carried out using this transgene resulted in the production of a single surviving transgenic founder animal from which the transgene was crossed onto both of our *ERCC1* knockout backgrounds. Two out of three pups that were found dead soon after birth were also genotyped as transgene-positive. The degraded condition of the DNA recovered from the cadavers prevented any analysis of transgene copy number in these animals. Whilst it was unlikely that the premature deaths of these animals was as a consequence of their transgene status, such a correlation could not be ruled out completely. It has been shown that overexpression of *ERCC1* in wild type Chinese hamster ovary cell lines resulted in an hypersensitivity to the DNA cross-linking agent melphalan, a nitrogen mustard derivative, without effecting sensitivity to UV irradiation (Bramson *et al,* 1993). It may be possible that these animals died as a consequence of overexpression of the transgene resulting in a dominant negative phenotype similar to that reported for the *in vitro* study. This would also suggest that the level of expression in the surviving transgenic founder was below the level for such an effect to be evident. An attempt to overexpress *ERCC1* in human cultured cells showed only a four-fold increase in protein levels despite evidence suggesting more than a thousand-fold increase in gene copy number and a dramatic increase in the level of *ERCC1* mRNA expression. Together these results may suggest that overexpression of the ERCC1 protein is not beneficial to the survival of a cell.

Further characterisation of the transgenic founder animal and F1 offspring indicated that, whilst the founder animal appeared to be mosaic with 1 to 2 copies of the transgene in a single array, between 3 and 4 copies of the transgene were detected in the offspring. Previous reports have indicated that the tissue specific expression of transgenes, where transcription is regulated by the *transthyretin* regulatory sequences, was most efficiently maintained where the copy number remains below 6 copies.

Animals with much higher copy numbers often had detectable levels of expression in other organs including kidney, salivary glands and intestines (Yan *et al.,* 1990).

Studies of LacZ reporter gene expression have shown that when transgenes are present in high copy number arrays gene expression can be silenced by mechanisms including hypermethylation of the transgene locus and formation of repressive chromatin confirmations (Garrick *et al.,* 1998). Using the Cre/*loxp* system of site specific recombination to generate transgenic mouse lines in which varying copy numbers of the LacZ transgene were present at the same chromosomal location, thereby eliminating positional effects, it was possible to study the effect of copy number alone on transgene silencing. It was found that in lines where in excess of 100 copies were present expression of the LacZ gene was very low. However when Cre-mediated reduction of the copy number was carried out, reducing the array to either 1 or 5 copies, expression was restored. Whilst it was unlikely that the array in our transgenic line would be subjected to repeat-induced silencing, it remained possible that other factors such as the chromosomal integration site may have had a repressive effect on transgene expression. We have shown, by northern blot analysis, that the *TTR/ERCC1* transgene was expressed at high levels, 20-100% of endogenous *TTR* levels, in the hepatocytes of all of the mice we analysed. Analysis of transgene expression on our *ERCC1* knockout backgrounds indicated that in some of the animals there was evidence of expression in other tissues including testes, skin and brain. The level of transgene expression in these other tissues was at very low levels and at less than 3% of the level of expression detected in the liver. It has been reported that expression of a *p21* transgene, using an identical *TTR* minigene vector backbone, was detected in the brain of a single transgenic line at the level of mRNA but p21 protein was not detected in the brain (Wu *et al.,* 1996). Analysis of expression of other *TTR* minigene based transgenes in skin has not been reported, presumably, as the analysis has not been performed. Expression of *TTR* minigene based transgenes in the testes has not previously been reported, where expression in this tissue has been analysed. The level and distribution of expression of the *TTR* minigene in non-hepatic organs was reported as being sporadic even within lines (Yan *et al.,* 1990).

Further, we have demonstrated that the expression of the *TTR/ERCC1* transgene was apparently stable with the age of the animals, although the absolute level of expression was shown to vary from animal to animal, expression was evident in an animal at 24 weeks of age, and persisted through all generations analysed thus far. This was particularly significant, in that, if expression of the transgene were to bring about phenotypic correction it was important that expression was maintained in order that correction continued. The variation in the absolute level of transgene expression we have reported did not correlate with age. Age dependant silencing of β-globin driven LacZ reporter genes has been described (Robertson *et al.*, 1996). Expression of the LacZ reporter gene in red blood cells was assessed in animals of various ages within a number of independent transgenic lines. It was found that expression fell by as much as 40-fold in some lines between the ages of 1 and 8 weeks. The extent by which expression declined was greatest in the lines with the highest number of transgene copies and was thought to be the result of progressive heterochromatinisation.

The *ERCC1* knockout mice reported by Weeda *et al.*, (1997) had a lifespan of 38 days whilst mice carrying a subtle mutation, resulting in a premature stop codon at position 292, lived for approximately 2 to 6 months depending on their genetic background. The considerable difference reported for the lifespan of the knockout mice (38 days) and mice with the subtle mutation (up to 6 months) may suggest that the protein arising from *ERCC1* allele carrying only the point mutation may retain some residual activity that contributes to this extended viability. The principle aim of the work reported here was simply to increase the lifespan of our *ERCC1*-deficient mice so as to enable further analysis of the consequences of *ERCC1* deficiency in other organs. We have reported that the *TTR/ERCC1* transgene positive *ERCC1* nulls survive for between 60 and 90 days compared to the transgene negative *ERCC1* nulls which generally die before weaning at 21 days of age. As such we can conclude that the targeted expression of *ERCC1* in the hepatocytes of our *ERCC1* nulls has resulted in an increased lifespan. We are now able to study the progressive consequences of *ERCC1* deficiency in other organs in our *ERCC1* knockout mice.

The observation of liver nuclear abnormalities has been confirmed for all of the *ERCC1* knockout and mutant mice reported to date (McWhir *et al.*, 1993; Weeda *et al.*, 1997 and Hsia unpublished data). We have shown that the targeted expression of an *ERCC1* transgene in the liver of our nulls results in the correction of this phenotype. The liver sections from our transgene positive nulls were seen to have a wild type appearance and this observation was further confirmed using FACScan analysis. The distribution of DNA content of hepatocyte nuclei from a four-week old transgene positive *ERCC1* null matched the distribution seen in a three-week old wild type animal. We speculate that the observed correction of aberrant nuclear sizes reflected a correction of the loss of hepatic function and as such the *ERCC1* null mice no longer died prematurely as a result of liver failure.

As with the liver abnormalities, the severe runting phenotype is common to all *ERCC1* targeted mice (McWhir *et al.*, 1993; Weeda *et al.*, 1997 and Hsia unpublished data). We have shown that this phenotype is partially corrected in our transgene positive nulls. We have reported that weight of the transgene positive *ERCC1* nulls was 60% of their wild type littermates, compared to the transgene negative nulls which weighed only 20% of the weight of the wild type animals (Mcwhir *et al.*, 1993). The severe runting associated with the nulls reflected the overall state of *ERCC1* deficiency in the animal and was not solely a consequence of *ERCC1* deficiency in the liver. As such it was not expected that the targeted expression of *ERCC1* in the liver would bring about complete correction of this phenotype. The degree of phenotypic correction that was observed served to illustrate the significance of corrected liver function in the context of growth and development.

Cutaneous abnormalities, including the lack of subcutaneous fat, in *ERCC1* null mice have been reported (Weeda *et al.*, 1997) and these abnormalities have been confirmed in our own knockout mice (unpublished data). Skin sections from the transgene positive *ERCC1*-deficient mice showed that the skin structure was noticeably altered from that of wild type animals. The skin sections showed an apparent distortion of nuclei in the epidermal cells and a thinning of the dermal layer. Although transgene expression (mRNA) had been detected in the skin of some mice it appeared that this expression was not sufficient to bring about phenotypic

correction of the associated cutaneous abnormalities. As such we regard the skin of the transgene positive nulls as being *ERCC1*-deficient and it is thus an appropriate organ in which to study the progressive consequences of *ERCC1* deficiency. There have been several reports concerning the predisposition of mice with defective NER to skin tumours (de Vries *et al.*, 1995; Nakane *et al.*, 1995 and Sands *et al.*,1995) The now extended lifespan of the transgene positive nulls would still preclude any long-term carcinogenicity studies but may permit studies of acute carcinogenesis to be performed. This would establish whether our *ERCC1* null mice were predisposed to developing skin cancer.

In addition to the liver, our original knockout mice showed detectable levels of the tumour suppressor protein p53 in both brain and kidneys. However there was no sign of any pathological abnormalities in these organs immediately prior to death at 21 days (Mcwhir *et al.*, 1993). Detailed histological examination carried out on the brain and kidneys from transgene positive animals aged between 6 and 8 weeks, showed there was still no sign of any pathology in the brain. However, in the kidneys of these animals there was evidence of nuclear abnormalities and dilated tubules, which in some cases contained leaked proteinaceous material, indicative of renal dysfunction. Similar abnormalities, protein casts and aberrant nuclei, were observed in the kidneys of the *ERCC1* targeted mice reported by Weeda *et al.*, (1997). In addition to the histological observations, we have also reported that, prior to dying, the transgene positive *ERCC1* nulls developed ascites, which can be an indication of renal failure. Although no assessment of renal function was carried out, it remains possible that the transgene positive *ERCC1* nulls died as a result of renal dysfunction associated with the *ERCC1*-deficient status of the kidneys in these animals. The abnormalities observed in the kidneys of these animals constitute the first additional consequence of *ERCC1* deficiency identified as a result of the extended lifespan of the nulls, resulting from the targeted expression of the *ERCC1* transgene in the liver.

As a whole, the result of this work will enable us to carry out further *in vivo* studies of the consequences of *ERCC1* deficiency that were previously constrained by the limited lifespan of our *ERCC1* null mice.

Chapter 7
Effects of UVB irradiation on epidermal Langerhans cells in *ERCC1*-deficient mice

The NER disorder xeroderma pigmentosum is principally characterised by hypersensitivity to UV light and the elevated incidence of skin cancers. The accepted theory of UV induced carcinogenesis suggests that skin neoplasias arise as a result of an accumulation of DNA damage in critical regions of the genome. When damage fails to be repaired before replication the damaged site is passed on as a cellular mutation. The accumulation of mutations in genes involved in the regulation of cell growth leads directly to carcinogenesis. This proposed, DNA repair focused model for carcinogenesis fails to consider the involvement of other cellular mechanisms in the observed predisposition to skin cancers. Pioneering work by Fischer and Kripke (1977) confirmed the carcinogenic capacity of UVB irradiation, long term exposure of mice to UV sources was shown to induce skin cancers. Further, they demonstrated that these tumours when transplanted into non-irradiated syngenic mice were destroyed by the host's immune response. When transplanted into UVB treated mice the tumours were seen to continue developing. These observations led to the proposal of a mechanism for tumour immunosurveillance which is impaired by UV irradiation. This immunosuppressive aspect of UV irradiation, combined with its mutagenic effects, may also contribute to the aetiology of the UV induced skin cancer associated with NER deficient XP patients.

The transgene positive *ERCC1* nulls do not survive long enough to perform any long-term carcinogenicity studies, and as such we are unable to determine whether these animals are predisposed to skin cancers. However, as they do survive into adulthood we are now in a position to assess the consequences of *ERCC1* deficiency on the immune functions implicated in preventing carcinogenesis. The majority of skin cancers initiate within the epidermis and as such cells with an immunological function in the epidermis are likely to be pivotal in establishing an anti-tumour response. The work presented in this chapter focuses on the effects of UVB exposure on the migratory properties and functional capacities of epidermal Langerhans cells in *ERCC1*-deficient mice. First, a brief overview of Langerhans cells in the context of their role within the skin immune system is presented (reviewed by Lappin, Kimber and Norval, 1996 and Muller , Halliday and Woods, 1997).

Epidermal Langerhans cells (LC) are the principal antigen presenting cells of the skin and are regarded as the outermost sentries of the immune system. The cells play an important role in the afferent arm of the immune response, migrating from the epidermis to the draining lymph node to present antigen to the T cells. LC belong to the family of dendritic cells and constitute a small, yet important, cell population of the skin immune system. They are derived from haematopoeitic stem cells of the bone marrow but little is known of what happens to them between leaving the bone marrow and reaching the skin. It is assumed that precursor cells migrate to the skin via the blood circulatory system. The LC are dendritic-shaped cells that are identified in epidermal sheets by staining for ATPase activity, LC being the only epidermal cells with clear cell surface expression of this enzyme. In keeping with their antigen presenting function Langerhans cells also express high levels of the major histocompatibility complex (MHC) class II molecules on their cell surface. Antigen is first trapped by the LC dendrites and then taken up by the cell whereby proteolysis within intracellular vesicles leads to presentation of the antigenic peptide. The MHC class II molecules possess a characteristic groove that can be loaded with antigen derived peptides, forming the antigen/ MHC complex which can interact with T cell receptors to activate T cells. In order to complete their T cell activation the LC express additional surface molecules, co-stimulatory elements, which interact with their respective ligands on the T cell, stabilising the binding and enhancing activation.

The potent antigen presentation capacity of the LC supports the idea that these cells may play a role in tumour immunosurveillance. The transformed tumour cells produce antigens which enable the immune system to distinguish between them and the normal cells of the body. Whilst LC are believed to be involved in the initiation of anti tumour immunity, the immunosuppressive effects of carcinogens such as UVB irradiation contribute to the failure of the tumour immunosurveillance system and enable tumour development. Following exposure to UVB irradiation the LC are depleted from the epidermis as they embark on an inappropriate migration to the lymph nodes. It is not clear what mechanism leads to the UVB irradiation causing the LC migration to the lymph nodes in the absence of any antigen to present. Further, if

antigen is present at the time of irradiation, the presentation of antigen under these conditions elicits a tolerising effect as opposed to the more appropriate immune response. It is believed that production of the immune response is dependent on the presence of the co-stimulatory signals on the cell surface of the LC and that UV irradiation somehow abrogates their effect. The depletion of the LC from the epidermis following UV irradiation renders the skin immune system less able to respond to any antigen, whether it be in the form of a chemical substance or a transformed cell, that may be encountered during this immunosuppressed period. Chronic exposure to UV irradiation is capable of exerting mutagenic effects in addition to prolonging the period of LC depletion.

The work presented in this chapter was carried out as part of a collaboration with Dr. Ali El-Ghorr and Dr. Mary Norval of the Department of Medical Microbiology, University of Edinburgh. Experimental procedures were performed by Dr. Ali El-Ghorr.

7.1 Langerhans cells are less abundant in the skin of *ERCC1*-deficient mice but show a normal pattern of depletion from the epidermis following UV irradiation

The number of LC in the epidermis of *ERCC1*-deficient mice, compared to wild type, was counted before and after UVB irradiation in order to determine whether the DNA repair deficient animals exhibited any enhancement of UV induced immunosuppression as measured indirectly by the depletion of LC from the epidermis. Previous attempts to study this in our original knockout mice were inconclusive. The lethal liver phenotype of these mice meant it was not possible to produce significant numbers of suitably aged null mice and the experimental procedures were not amenable to using young runted mice where the development of the ears was incomplete. Additionally it would not have been prudent to draw conclusions from analyses carried out using mice that were clearly not in good health prior to procedures. All the analyses presented in this chapter were performed using transgene positive *ERCC1* nulls on the KT #209 knockout background, unless otherwise stated, and wild type littermates were used as controls. All the mice used

were adults, aged between six and eight weeks and showed absolutely no signs of ill health at the time of the experiment. There were five mice of each genotype in the UV irradiated groups and six mice of each genotype in the non-irradiated control groups.

The mice were irradiated on their dorsal surfaces following shaving with electric clippers. To ensure equivalent doses were received, the mice were immobilised by administration of anaesthesia. A 1500 Jm^{-2} dose of broadband UVB was delivered using a bank of two TL20 W/12 lamps. This is equivalent to a minimal erythemal dose (MED) for wild type mice. Experimental analysis was performed 24 hours after the irradiation.

At 24 hours after the irradiation the dorsal skin of the *ERCC1* nulls was noticeably more erythemic than that of the irradiated wild type mice, indicating that the DNA repair deficient animals were more sensitive to UV irradiation. No detailed histological analysis was performed on the skin at this stage.

Twenty-four hours after the irradiation, the mice were humanely killed and one ear from each mouse was removed, split and epidermal sheets were prepared from the dorsal surfaces only. LC were stained using adenosine-5'-diphosphate (ADP) and the number of $ATPase^{+}$ cells were counted (Figure 7.1). The staining is a result of the reduction of the ADP to AMP by the ATPase on the LC surface. In the non-irradiated groups the average number of LC per mm^2 was 498 ±37 for the wild type mice and 366 ±24 for the ERCC1-deficient mice. Thus, there were significantly less LC present in the epidermis of the *ERCC1* nulls compared to wild type mice. Following UV irradiation the number of LC present in the epidermis of the wild type animals fell to 166 ±17, which constituted a depletion of 67% of the normal number. In the *ERCC1* nulls the number of LC present in the epidermis following UV irradiation was 137 ±20, which constituted a depletion of 63% of the normal number. It is therefore unlikely that DNA repair deficiency effects the UV induced depletion of these cells.

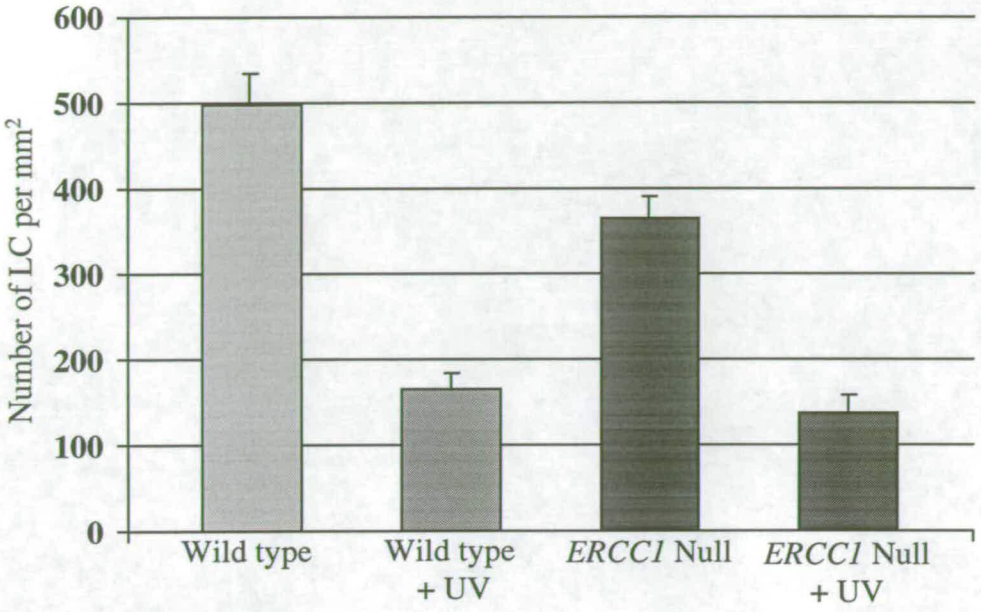
Morphologically the dendritic LC in non-irradiated skin of the *ERCC1* nulls were indistinguishable from those in wild type skin. Following UV irradiation a proportion of the LC that remained in the epidermis appeared to have lost their dendrites and

Figure 7.1 Langerhans cells present at a lower density in the epidermis of *ERCC1* nulls are depleted following UVB irradiation as normal

a. Epidermal density of LC following a single 1500 Jm^{-2} dose of UVB irradiation. LC were stained and counted in non-irradiated controls and in the irradiated groups twenty-four hours after the irradiation. Results represent the mean (±SEM) of LC densities in epidermal sheets from five mice per non-irradiated group and six mice per irradiated group. Results were compared using Students *t*-test; WT vs Null $p<0.005$, WT vs WT+UV $p<0.000$, Null vs Null+UV $p<0.000$.

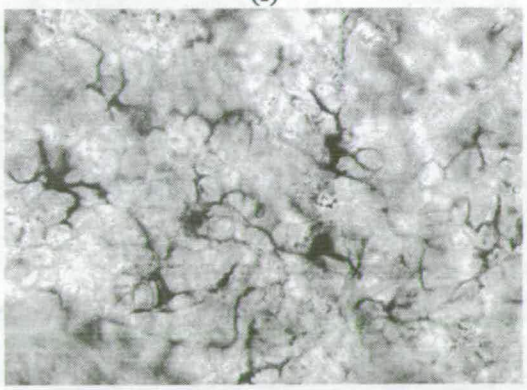
b. UV induced morphological changes in epidermal Langerhans cells. i. A non-irradiated wild type epidermis showing LC with clearly evident dendrites. ii. A non-irradiated *ERCC1* null epidermis showing LC present at a lower density but with evidence of dendrites. iii. Irradiated wild type epidermis showing depleted numbers of LC as well as morphological changes (less dendritic) in a proportion of the cells. iv. Irradiated *ERCC1* null epidermis showing depleted numbers of LC and evidence of morphologically altered (less dendritic) LC.

a.

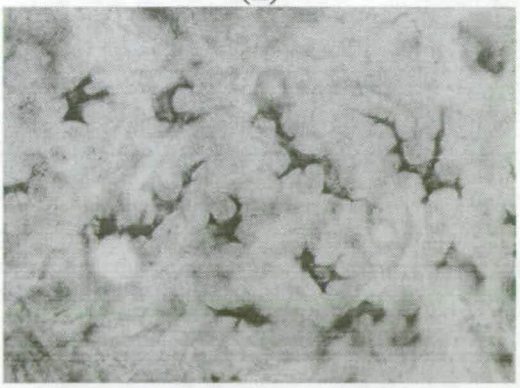

600
500
400
300
200
100
0
Number of LC per mm2
Wild type
Wild type + UV
ERCC1 Null
ERCC1 Null + UV

b.

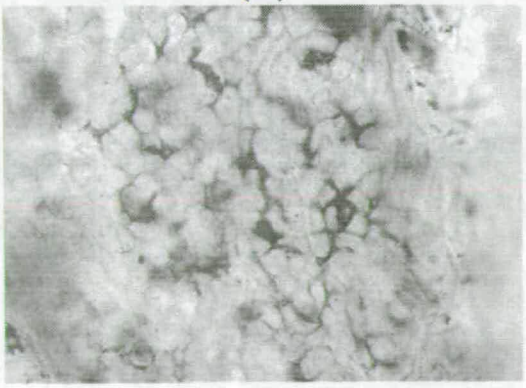
(i)



(ii)



(iii)



(iv)

looked rounded (Figure 7.1b). The extent of these differences was greater in the skin of the *ERCC1* nulls (estimated at 90% of the cells) than in wild type skin (estimated at 20% of the cells). This observation suggested that the LC in the epidermis of the *ERCC1* nulls were more sensitive to the UV irradiation.

7.2 The Langerhans cells in *ERCC1*-deficient mice are capable of antigen presentation as measured by the mixed skin lymphocyte reaction (MSLR)

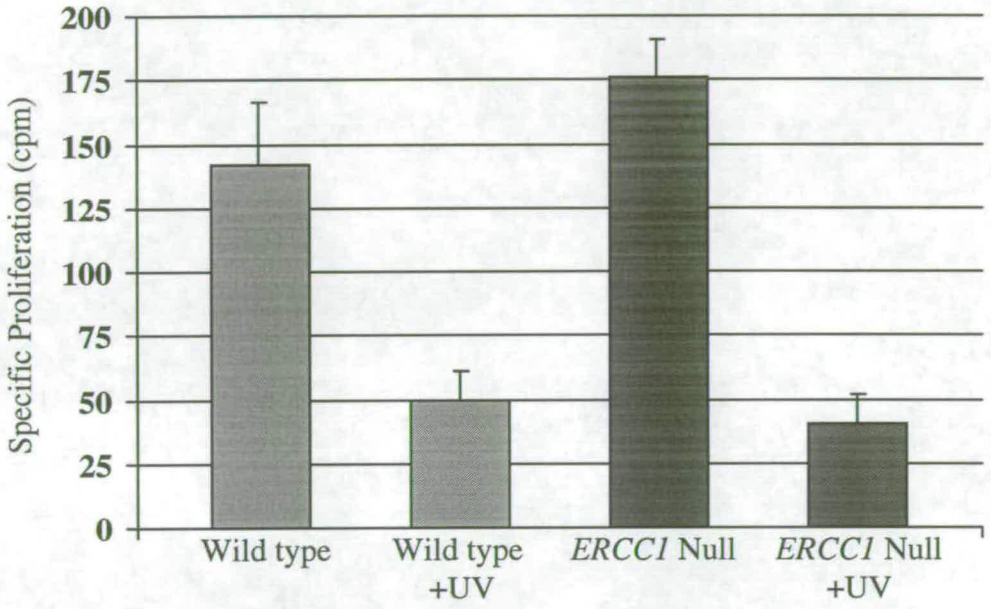
Having established that the Langerhans cells in the DNA repair deficient mice show the normal pattern of depletion from the epidermis following UV irradiation it was necessary to establish whether the LC in these animals were capable of presenting antigen. The antigen presenting capacity of the LC is determined by measuring the ability of the cells to stimulate proliferation of spleen cells *in vitro*.

The remaining ear from the above mentioned mice was removed for preparation of epidermal cells (EC). Spleen cells from strain C3H female mice were used as the responder cells in the MSLR. For each of the experimental groups, 0.5 x 10^5 EC were incubated together with 10^5 spleen cells. After culturing for five days at 37°C 0.7μCi [H^3]thymidine was added to each culture before harvesting the cells twenty-four hours later. The proliferative response of the spleen cells was determined by measuring the incorporation of the [H^3]thymidine. The net proliferative response was calculated as the mean of 5 experiments following the subtraction of the mean of control experiments, where EC were cultured alone (Figure 7.2).

This analysis showed that specific proliferative response (measured in counts per minute (cpm)) resulting from incubation with non-irradiated wild type epidermal cells was 143 cpm, compared to 50 cpm when incubated with EC cells from irradiated wild type animals. This result indicates that the antigen presenting capacity of the EC from wild type animals is reduced following UV irradiation. The reduction in the antigen presenting capacity is a direct reflection of the reduced number of LC in the epidermis following irradiation. The specific proliferative response resulting from incubation with non-irradiated *ERCC1* null EC was 176 cpm, compared to 40 cpm when incubated with EC cells from UV irradiated *ERCC1* null animals. This result indicates that, as with the wild type animals, the antigen presenting capacity of

Figure 7. 2 The epidermal cells of *ERCC1* nulls exhibit a normal antigen presenting capacity *in vitro*

A mixed skin lymphocyte reaction (MLSR) comparison of the antigen presenting capacity of epidermal cells from *ERCC1* null and wild type mice before and after irradiation. The results represent the mean (±SEM) specific proliferation of cells from five replicate wells per experimental group.


Specific Proliferation (cpm)
200
175
150
125
100
75
50
25
0
Wild type
Wild type +UV
ERCC1 Null
ERCC1 Null +UV

the EC from *ERCC1* nulls is reduced following UV irradiation. Thus the antigen presenting capacity of the EC from *ERCC1* nulls did not differ from those from wild type animals. Further, the observed morphological changes after UV irradiation did not appear to effect the antigen presenting function in this experiment.

7.3 Langerhans cells of the *ERCC1* nulls do not accumulate in the draining lymph nodes following UV irradiation

Having demonstrated that the LC were depleted from the epidermis of the *ERCC1* null mice it was necessary to investigate whether the cells had migrated to the draining lymph nodes to fulfil their antigen presentation function.

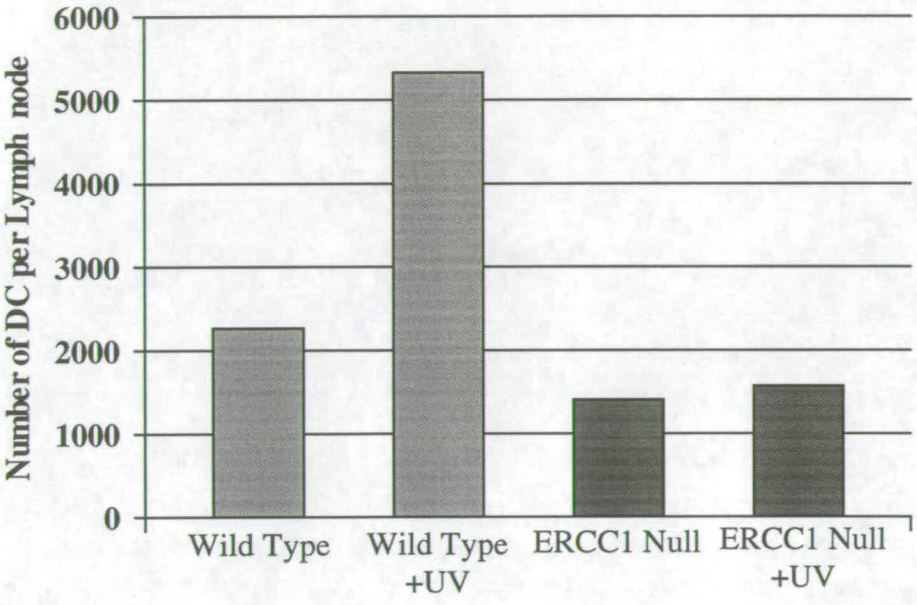
The accumulation of dendritic cells in the draining lymph nodes was measured. Twenty four hours after irradiation the mice were humanly killed and their auricular and axillary lymph nodes were removed and pooled. A single cell suspension was made from the pooled lymph nodes and the lymph node cells were counted microscopically, to establish the total number of cells per lymph node. The dendritic cells were then isolated by a process of centrifugation on a 14.5% metrizamide cushion. The dendritic cell enriched population at the cushion interface was counted microscopically to determine the number of dendritic cells per lymph node (Figure 7.3a). This analysis showed that, for the wild type animals, the number of dendritic cells per lymph node prior to UV irradiation was 2262 compared to 5533 following irradiation. This result illustrates the normal migratory response of epidermal LC following UV irradiation. The number of dendritic cells in the lymph nodes of the *ERCC1* nulls prior to irradiation was counted as 1412 compared with 1560 twenty-four hours after UV irradiation. There was no evidence to suggest that the LC from the *ERCC1* nulls accumulated in the draining lymph nodes following UV irradiation. This result was confirmed when the experiment was repeated (Figure 7.3b). The animals used in the second experiment included a mixture of transgene positive nulls from both of the *ERCC1* knockout backgrounds In the second experiment the number of dendritic cells per lymph node, in the wild type animals, was 1779 compared to 3083 following UV irradiation. The number of dendritic cells per lymph node, in the *ERCC1* null mice, was 1699 prior to irradiation and 1594 following irradiation.

Figure 7. 3 Following UVB irradiation LC from the epidermis of *ERCC1* null mice fail to accumulate in the draining lymph nodes

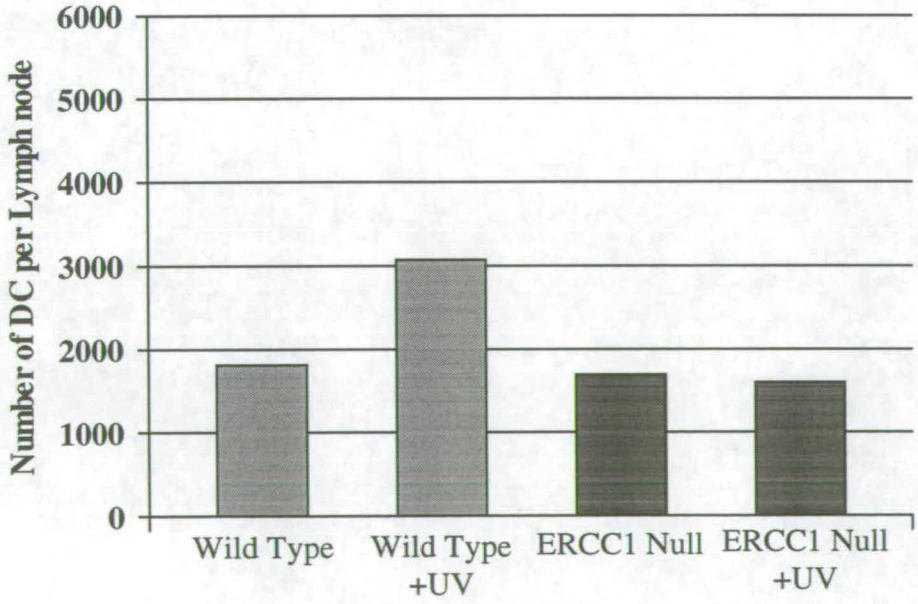
a. The number of dendritic cells per draining lymph node following a single 1500 Jm^{-2} dose of UVB irradiation. Results represent the mean number of dendritic cells per lymph node from 12 lymph nodes per non-irradiated group and 10 lymph nodes per irradiated group.

b. Repeated analysis of the number of dendritic cells per lymph node following UVB irradiation (1500 Jm^{-2}). Results represent the mean number of dendritic cells per lymph node from 24 lymph nodes per wild type group and 18 lymph nodes per *ERCC1* null group.

a.


Number of DC per Lymph node
6000
5000
4000
3000
2000
1000
0
Wild Type
Wild Type +UV
ERCC1 Null
ERCC1 Null +UV

b.


Number of DC per Lymph node
6000
5000
4000
3000
2000
1000
0
Wild Type
Wild Type +UV
ERCC1 Null
ERCC1 Null +UV

These results confirm that following UV irradiation LC cells do not accumulate in the draining lymph nodes of *ERCC1*-deficient mice.

7.4 Discussion

Patients with the DNA repair disorder xeroderma pigmentosum are hypersensitive to sunlight and predisposed to developing UV induced skin cancer. Whilst the nature of the DNA repair defect is well documented, little is understood about the mechanisms by which the defective NER results in the clinical symptoms such as hypersensitivity to sunlight. The carcinogenic properties of UVB irradiation have long been recognised and it is also recognised that UVB irradiation has an immunosuppressive effect (Fischer and Kripke, 1977). It has been suggested that, in patients with defective DNA repair, it is the combined immunosuppressive and mutagenic effects of UVB irradiation that lead to the observed increased incidence of skin cancer. The work presented in this chapter is a preliminary investigation into the consequences of *ERCC1* deficiency on the afferent limb of the skin immune system, namely the antigen presenting epidermal Langerhans cells.

We have shown that the density of LC in the epidermis of *ERCC1*-deficient mice was significantly lower than that seen in wild type littermates. It is not clear why the density of LC was lower in the *ERCC1* nulls prior to irradiation. The animals had been maintained in an animal facility with no windows and covered fluorescent lighting and as such it is unlikely that the lower density of LC was a consequence of an enhanced response to incidental exposure to UV irradiation prior to the experimental procedure. We have already reported that the structure of the ERCC1-deficient skin differs from that of the wild type skin (Chapter 6.13). It may be that the epidermis in our repair deficient animals is unable to provide the necessary physiological environment required to support the LC at the normal density. We have also observed that the *ERCC1*-deficient mice exhibited visible erythema following the UVB irradiation, unlike the wild type littermates. This suggested that the actual MED for the DNA repair deficient animals was less than the MED for wild type mice. In a previously reported study comparing the MED of XP patients with normal subjects it was shown that the MED of XPA, XPD and XPV patients was typically

lower than that of normal subjects. The XPA patient had the lowest MED at 0.1 of normal, the XPD patient had an MED of 0.33 of normal, whilst the XPV patients had an MED of about 0.5 of normal (Jimbo *et al.*, 1992).

Following UVB irradiation the LC depletion from the epidermis of the *ERCC1*-deficient animals was equivalent to that seen in the wild type littermates. In this respect the repair deficient animals exhibited a normal response to UVB irradiation. Of the cells that remained in the epidermis many were rounded and lacked dendrites and the morphological changes were more extensive in the *ERCC1* nulls (90%) compared to the wild type animals (20%). This observation suggested that the LC in the *ERCC1* nulls may have been more sensitive to the effects of the UVB irradiation. Alternatively, this may have been a reflection of the abnormal structure of the epidermis in these animals; the previously described compression of the epidermal layer may have afforded the LC less protection from the UV irradiation. The results we have reported here differ from results previously reported for similar studies carried out on *XPA*-deficient mice (Miyauchi-Hashimoto *et al.*, 1996). The structure of the skin in general and number of LC in the skin of the *XPA*-deficient animals did not differ from that of the wild type controls and following UV irradiation the depletion of LC from the *XPA*-deficient skin (59% depletion) was greater than that of the wild type controls (33% depletion). The depletion of LC from the epidermis is thought to be dose dependent and as such the differences observed between our mice and the *XPA*-deficient mice may reflect the different UV doses that the animals received. Whilst the *XPA*-deficient mice received a UVB dose of only 250 Jm^{-2} our animals received a dose of 1500 Jm^{-2}. It is possible that the 6-fold higher UV dose received by the mice in our study had the effect of masking differences in LC depletion that may have been evident at lower doses of UV. Although, when exposed to a single UVB dose of 1000 Jm^{-2} the LC depletion in *XPA*-deficient mice was reported to be almost 100%, compared to 62% in wild type animals. In a study carried out on XP patients, it was demonstrated that following exposure to a single MED dose of UV the depletion of LC from the skin of XP patients was greater than in normal subjects. However, when the same individuals were exposed to UV doses equivalent to three times the MED, the depletion of LC from the epidermis normal

subjects was similar to that seen in the XP patients (Jimbo *et al.*, 1992). The study carried out on the *XPA*-deficient mice also reported that the cells remaining in the epidermis had undergone morphological changes similar to those that we have reported. However, unlike our study, no changes were seen in wild type animals. Once again this is likely to be as a consequence of the lower UV dose administered in the *XPA* study.

We also reported that the LC in the epidermis of the *ERCC1*-deficient animals were capable of antigen presentation as shown by the MSLR. Although the capacity for antigen presentation was greatly reduced following UV irradiation, this reduction was mirrored in the wild type animals and as such does not reflect the consequences of DNA repair deficiency, rather it was a direct reflection of the reduced numbers of LC present in the skin following the normal migratory response after UV irradiation. Having already reported that the LC remaining in the epidermis of UV irradiated *ERCC1*-deficient animals exhibited considerable morphological changes the retention of their capacity to present antigen to any extent was quite surprising. It would not be unreasonable to assume that the loss of dendrites from these cells would have impaired their antigen presenting capacity to some extent. A study by Vink *et al.*(1996) demonstrated that the proportion of LC, containing DNA damage, in the draining lymph nodes peaked at ~10% twenty-four hours after irradiation. They went on to demonstrate that these migrated cells exhibited normal *in vivo* antigen presenting activity at this time. However, cells that were collected from lymph nodes following sensitising with an antigenic substance three days after the UV dose were less able to present antigen *in vivo*. This suggests that although DNA damage is present immediately after irradiation, the impairment of antigen presentation develops with time. It remains possible that we may have observed impaired antigen presentation in our *ERCC1* nulls relative to wild type if we had performed the MSLR four days after the initial irradiation.

Finally, we have reported that following UVB irradiation the LC of the *ERCC1*-deficient mice do not appear to accumulate in the draining lymph nodes, contrary to the situation observed in wild type mice. Whilst, this does not mean to say that the LC do not migrate to the lymph nodes it simply states that they do not accumulate in

the nodes. It could be that the LC of the repair deficient mice are more inclined to pass through the lymph node. It would not be unreasonable to suggest that DNA repair deficient LC might be more sensitive to UV, which could result in altered expression of the cell surface proteins required for the normal interactions with the T cells in the lymph nodes. The failure of such an interaction may mean that the LC do not accumulate. Further, we can not rule out the possibility that the LC in these animals are incapable of successful migration to the draining lymph node. It remains possible that the altered structure of the skin in these animals in some way perturbs the normal migratory pathway. If this were indeed to be the case, it would be reasonable to assume that the *ERCC1*-deficient animals would suffer from enhanced skin immunosuppression on the basis that the LC were incapable of antigen presentation *in vivo*. We do not have any evidence to suggest that these DNA repair deficient animals are immunocompromised relative to wild type animals.

In summary, we have shown that the density of LC in the skin of the *ERCC1*-deficient mice is lower than in wild type mice and following exposure to UVB irradiation these LC fail to accumulate in the draining lymph nodes. Although the antigen presenting capacity of the cells appears to be normal *in vitro*, the previous observations could indicate that the skin immune system may be impaired in these animals.

Chapter 8

Summary and concluding remarks

Transgenic animal technology and in particular the use of germline manipulation by gene targeting offers an ideal strategy for the *in vivo* analysis of the mechanisms of DNA repair. Knockout mice provide an opportunity to study the function of disrupted genes and to determine the consequences of the loss of function following exposure to genotoxic agents. Additionally, they have the potential to assist the development of gene therapies. The work presented in this thesis relates such analysis performed on our previously reported *ERCC1* knockout mice, an animal model for NER deficiency. *ERCC1* was previously reported as being expressed at low levels in all tissues and stages of development. Given that ERCC1 plays such an important role in the repair of UV induced DNA damage, we felt that it would not be unreasonable to expect that the expression of a DNA repair gene may be altered in a tissue where the gene product would be of greater biological relevance, namely the skin. Northern blot analysis of *ERCC1* expression in mouse tissues revealed that differential expression of the gene was indeed evident in the skin as well as cultured mouse keratinocytes and embryonic fibroblasts. In addition to the normal 1.1kb mRNA a novel *ERCC1* transcript was identified in these cell types and mouse skin. Characterisation of this transcript by 5' and 3' RACE analysis revealed that the 1.5kb skin specific transcript was produced by differential initiation of transcription, presumably from an upstream promoter. Preliminary mapping of the transcriptional initiation site, at 389 nt upstream of the transcriptional start site of the normal transcript, was based on the sequence analysis performed on the 5' RACE clone. It is intended that this initial mapping be confirmed using either primer extension or RNase protection analysis. This work constitutes the first report of a skin specific transcription pattern associated with an NER gene. It would be of interest to establish whether any of the other NER genes exhibit similar tissue specific expression patterns.

Subsequent analysis failed to provide evidence relating to the UV induction of this transcript. Although this ERCC1 mRNA was expressed at higher levels in the skin, relative to the normal 1.1kb transcript in other tissues, the biological significance of the novel transcript remains to be elucidated. Western blot analysis is currently being performed in our laboratory to determine whether the difference in expression seen at the level of mRNA results in any differences at the level of the protein. Preliminary

findings have not provided any evidence indicating significant differences at this level (Kan-Tai Hsia, unpublished data). We have also reported an elevated level of *ERCC1* expression evident in the testes, which may be indicative of some biological significance. Consequently, we are currently involved in studying the phenotypic consequences of *ERCC1* deficiency in this tissue.

Following on from the identification of the skin specific transcript we began work ultimately aimed at characterising the elements within the upstream promoter region responsible for this skin specific pattern of expression. The first step towards this analysis required the construction of a functional *ERCC1* minigene that could then be used, by means of conventional deletional analysis, to identify the control elements. A prototype minigene was constructed and shown to be functional by virtue of its ability to rescue the UV sensitivity of a variety of different *ERCC1*-deficient cell lines including mouse embryonic fibroblasts, keratinocytes and Chinese hamster ovary cells. Using northern blot analysis, we observed that the expression pattern of the minigene did not match the endogenous skin specific expression pattern, where the 1.5kb transcript is the more abundant species relative to the 1.1kb transcript, in any of the keratinocyte or embryonic fibroblast cells transfected with this construct. The failure to include any significant 3' flanking sequences in the minigene construct also meant that the minigene appeared to be subject to aberrant polyadenylation. It would not have been prudent to proceed with the proposed deletional analysis of the promoter region whilst the expression pattern of this *ERCC1* minigene was so far removed from the skin specific pattern. It is intended that the minigene should be redesigned to include more 3' flanking sequences in an attempt to overcome the problem of aberrant polyadenylation. Further, it may be beneficial to consider reducing the overall size of the minigene by replacing genomic fragments with cDNA sequences where appropriate.

The premature death of the *ERCC1* knockout mice, from liver failure, has prevented any significant studies of the consequences of *ERCC1* deficiency in adult tissues. The work presented in the second half of this thesis was aimed at rectifying this problem by extending the lifespan of these DNA repair deficient mice. This was first approached by attempting to develop an efficient retroviral gene therapy. We have

described the production of *ERCC1* retrovirus and its subsequent *in vitro* use in correcting the UV sensitivity of *ERCC1*-deficient mouse embryonic fibroblasts. Although we were able to demonstrate that a proportion of cells within a pool of infected fibroblasts were corrected to wild type levels, it was apparent that the viral titre from the viral producing pool was not sufficiently high enough to infect most target cells. Attempts to increase the viral titre using clones from the viral producing pool were unsuccessful. The limited degree of success achieved when using this system to correct the consequences of *ERCC1* deficiency *in vitro* meant that it was not plausible to consider attempting to rescue the lethal liver phenotype of the *ERCC1* knockout mice using this gene therapy strategy. It was concluded that the phenotypic rescue of the *ERCC1* null mice should be attempted using a non-retroviral strategy.

We speculated that, by restoring *ERCC1* function in the liver, the lifespan of the *ERCC1* nulls would be extended significantly to enable the further studies already mentioned. We have described how, using conventional transgenesis, we have generated mice expressing *ERCC1* specifically in hepatocytes and that crossing this line with our knockout line has increased the lifespan of our *ERCC1* null mice. A single transgenic line with hepatocyte specific expression was produced using an *ERCC1* transgene under the control of transthyretin regulatory sequences. Southern blot analysis indicated that the founder animal had 1 or 2 copies of the transgene present as a single array. Northern blot analysis of the expression pattern of this transgene revealed that it was expressed at high levels and essentially restricted to the liver. Expression was detected at very low levels in other tissues including skin and brain in some animals. The transgene expression was stable with age and through the four generations of mice generated thus far.

When crossed onto the *ERCC1* knockout backgrounds the transgene expression was shown to increase the lifespan of the *ERCC1* null mice by four-fold from the previous 21 days to up to 86 days. The severe runting associated with the *ERCC1* nulls was also alleviated; the transgene positive nulls weighed 58% of wild type compared to the 20% of wild type reported for the straight nulls. We have also shown that the expression of the *ERCC1* transgene corrects the liver nuclear abnormalities

previously associated with the *ERCC1*-deficient mice. Histological and FACScan analysis was used to demonstrate that the livers of the transgene positive nulls were indistinguishable from wild type livers.

We reported that histological analysis showed that the transgene positive nulls exhibited renal abnormalities, which had not previously been identified in our knockout mice. In addition to the presence of abnormal nuclei we also found dilation of the renal tubules to be a common feature in the kidneys of the adult transgene positive nulls. Frequently, these tubules appeared to be blocked with protein casts indicating that renal function may have been impaired. *Post mortem* examination of the transgene positive nulls revealed ascites and, bearing in mind the histological evidence, we suggested that the cause of death in these animals was likely to be renal failure.

Despite the reported observation of low levels of *ERCC1* expression evident in the skin of some of the transgenic animals we did not see any phenotypic consequences of this leaky expression. In all of the animals examined it was evident that the structure of the ERCC1-deficient skin was not the same as wild type. The nuclei of the epidermal cells appeared to be compressed and the dermal layer was consistently thinner than in wild type animals. We concluded that the skin of the transgene positive animals remained DNA repair deficient and as such served as a suitable tissue in which to study the consequences of NER deficiency in what were otherwise healthy animals. Essentially the aim of this work presented in this part of the thesis was achieved, the lifespan of the *ERCC1* nulls was extended to such an extent as to permit analysis of the consequences of *ERCC1* deficiency in other tissues.

Indeed, it was the success of the phenotypic rescue of the nulls by conventional transgenesis that made it possible to carry out the work presented in the final chapter of this thesis. We were able to investigate the consequences of *ERCC1* deficiency on the antigen presenting Langerhans cells of the skin immune system. In this respect, it was shown that the LC were present in the epidermis at a significantly lower density in the transgene positive nulls compared to wild type. Further, it was demonstrated that the *ERCC1* null LC were depleted from the epidermis by UVB irradiation as normal, but that they failed to accumulate in the draining lymph nodes as was

observed in wild type animals. We have speculated whether these differences stem from the basic histological differences between *ERCC1* null and wild type skin. Following the UVB irradiation we observed that the LC in the epidermis of the null mice exhibited more widespread morphological changes (dendritic loss) than seen in the wild type epidermis. It was believed that the apparent differences in the extent of dendritic loss would have been reflected in a reduced ability to present antigen. However, it was also demonstrated, using a mixed skin lymphocyte reaction, that whilst the antigen presenting ability of the *ERCC1* null LC was diminished following irradiation, the wild type LC exhibited an equivalent impairment. Although the antigen presenting capacity of the *ERCC1* null LC was apparently normal *in vitro*, we speculated that the initial observations, of lower LC density and failure to accumulate in the draining lymph nodes, may indicate that the skin immune system in these animals may well be effected as a consequence of the associated DNA repair deficiency. More extensive studies in this area have been restricted by the limited availability of transgene positive *ERCC1*-deficient mice.

The increased lifespan of the transgene positive *ERCC1* nulls enabled us to extend our previously reported studies of the *in vivo* consequences of *ERCC1* deficiency. In our laboratory, we have recently produce a mouse with a floxed *ERCC1* allele, which will enable us to further extend the studies presented here, by studying the consequences of *ERCC1* deficiency in specific tissues or at defined developmental stages.

Chapter 9

References

Albers, K.M., Davis, F.E., Perrone, T.N., Lee, E.Y., Liu, Y. and Vore, M. (1995) Expression of an epidermal keratin protein in liver of transgenic mice causes structural and functional abnormalities. *J Cell Biol.* **128:** 157-169

Antequera, F., Bird, A. (1993) Number of CpG islands and genes in human and mouse. *Proc. Natl. Acad. Sci. USA* **90:** 11995-11999

Bardwell, A.J., Bardwell, L., Tomkinson, A.E. and Friedberg, E.C. (1994) Specific cleavage of model recombination and repair intermediates by the yeast Rad1-Rad10 DNA endonuclease. *Science* **265:** 2082-2085

Belt, P.B.G., van Oosterwijk, M.F., Odijk, H., Hoeijmakers, J.H.J. and Backendorf, C. (1991) Induction of a mutant phenotype in human repair proficient cells after overexpression of a mutated human DNA repair gene. *Nucleic Acids Res.* **19:** 5633-5637

Bentley, D.J., Selfridge, J., Millar, J.K., Samuel, K., Hole, N., Ansell, J.D. and Melton, D.W. (1996) DNA ligase I is required for foetal liver erythropoiesis but is not essential for mammalian cell viability. *Nat. Genet.* **13:** 489-491

Bertrand, P., Tishkoff, D.X., Filosi, N., Dasgupta, R. and Kolodner, R.D. (1998) Physical interaction between components of DNA mismatch repair and nucleotide excision repair. *Proc. Natl. Acad. Sci. USA* **95:** 14278-14283

Biggerstaff, M., Szymkowski, D.E. and Wood, R.D. (1993) Co-correction of the ERCC1 , ERCC4 and xeroderma pigmentosum group F DNA repair defects *in vitro*. *EMBO J.* **12:** 3685-3692

Bohr, V.A., Smith, C.A., Okumoto, D.S., Hanawalt, P.C. (1985) DNA repair in an active gene: removal of pyrimidine dimers from the DHFR gene of CHO cells is much more efficient than in the genome overall. *Cell* **40:**359-369

Bowman, T., Symonds, H., Gu, L., Yin, C., Oren, M. and Van Dyke, T. (1996) Tissue specific inactivation of p53 tumor suppression in the mouse. *Genes. Dev.* **10:**826-835

Bramson, J. and Panasci, L. (1993) Effect of ERCC1 overexpression on sensitivity of Chinese hamster ovary cells to DNA damaging agents. *Cancer Res.* **53:** 3237-3240

Brinster, R.L., Chen, H.Y., Trumbauer, M., Senear, A.W., Warren, R. and Palmiter, R.D. (1981) Somatic expression of herpes thymidine kinase in mice following injection of a fusion gene into eggs. *Cell* **27:** 223-231

Busch, D.B., van Vuuren, H., de Wit, J., Collins, A., Zdzienicka, M.Z., Mitchell, D.L., Brookman, K.W., Stefanini, M., Riboni, R., Thompson, L.H., Albert, R.B., van Gool, A.J. and Hoeijmakers, J. (1997) Phenotypic heterogeneity in nucleotide excision repair mutants of rodent complementation groups 1 and 4. *Mutat. Res.* **383:** 91-106

Capecchi, M.R. (1989) Altering the genome by homologous recombination. *Science* **244:** 1288-1292

Chomczynski, P. and Sacchi, N. (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* **162(1):** 156-159

Chu, G. and Mayne, L. (1996) Xeroderma pigmentosum, Cockayne syndrome and trichothiodystrophy: do the genes explain the diseases? *Trends Genet.* **12(5):** 187-192

Cleaver, J. and Kraemer, K.H., Xeroderma pigmentosum and Cockayne syndrome. In *The metabollic and molecular bases of inherited disease.* 7th edition pp4393-4419, Eds. Scriver, C.R., Beaudet, A.R., Sly, W.S. and Valle, D.(McGraw-Hill Inc., 1995)

Cohen-Tannoudji, M. and Babinet, C. (1998) Beyond knockout mice: new perspectives for the programmed modification of the mammalian genome. *Mol. Hum. Repro.* **4(10):** 929-938

Costa, R.H., Lai, E. and Darnell, J.E. (1986) Transcriptional control of the mouse prealbumin (transthyretin) gene: Both promoter sequences and a distinct enhancer are cell specific. *Mol. Cell. Biol.* **6(12):** 4697-4708

Costantini, F. and Lacy, E. (1981) Introduction of a rabbit beta-globin gene into the mouse germ line. *Nature* **294:** 92-94

Dagert, M. and Ehrlich, S.D. (1974) Prolonged incubation in calcium phosphate improves competence of *Escherichia coli* cells. *Gene* **6:** 23-28

Danos, O. and Mulligan, R.C. (1988) Safe and efficient generation of recombinant retroviruses with amphotropic and ecotropic host ranges. *Proc. Natl. Acad. Sci. USA* **85:** 6460-6464

de Boer, J., de Wit, J., van Steeg, H., Berg, R.J.W., Morreau, H., Visser, P., Lehmann, A.R., Duran, M., Hoeijmakers, J.H.J. and Weeda, G. (1998b) A mouse model for the basal transcription/ repair syndrome trichothiodystrophy. *Mol. Cell* **1:** 981-990

de Boer, J., Donker, I., de Wit, J., Hoeijmakers, J.H.J. and Weeda, G. (1998a) Disruption of the mouse xeroderma pigmentosum group D gene DNA repair/basal transcription gene results in preimplantation lethality. *Cancer Res.* **58:** 89-94

de Laat, W.L., Sijbers, A.M., Odijk, H., Jaspers, N.G.J., Hoeijmakers, J.H.J. (1998) Mapping of interaction domains between human repair proteins ERCC1 and XPF. *Nucleic Acids Res.* **26:** 4146-4152

de Vries, A., van Oostrom, C.T.M., Dortant, P.M., Beems, R.B., van Kreijl, C.F., Capel, P.J.A., van Steeg, H. (1997) Spontaneous liver tumours and benzo(a)pyrene-induced lymphomas in XPA-deficient mice. *Mol. Carcinogenesis* **19(1):** 46-53

de Vries, A., van Oostrom, C.T.M., Hofhuis, F.M.A., Dortant, P.M., Berg, R.J.W., van Kreijl, C.F., Capel, P.J.A., van Steeg, H. and Verbeek, S.J (1995) Increased susceptibility to ultraviolet-B and carcinogens of mice lacking the DNA excision repair gene *XPA. Nature* **377:** 169-173

Dranoff, G., Jaffe, E., Lazenby, A., Golumbek, P., Levitsky, H., Brose, K., Jackson, V., Hamada, H., Pardoll, D. and Mulligan, R.C.(1993) Vaccination with irradiated tumour cells engineered to secrete murine granulocyte-macrophage colony stimulating factor stimulates potent, specific and long lasting anti-tumour immunity. *Proc. Natl. Acad. Sci. USA* **90:** 3539-3543

Evans, M. and Kaufman, M.H. (1981) Establishment in culture of pluripotent cells from mouse embryos. *Nature* **292:** 154-155

Feinberg, A.P. and Vogelstein, B. (1983) A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* **132:** 6-13

Fischer, M.S. and Kripke, M.L. (1977) Systemic alteration induced in mice by ultraviolet light irradiation and its relationship to ultraviolet carcinogenesis. *Proc. Natl. Acad. Sci. USA* **74 :** 1688-1692

Fort, P., Marty, L., Piechaczyk, M., el Sabrouty, S., Dani, C., Jeanteur, P. and Blanchard, J.M. (1985) Various rat adult tissues express only one major mRNA species from the glyceraldehyde-3-phosphate-dehydrogenase multigenic family. *Nucleic Acids Res.* **13:** 1431-1442

Friedberg, E.C., Meira, L.B. and Cheo, D.L. (1998) Database of mouse strains carrying targeted mutations in genes affecting cellular responses to DNA damage. Version 2. *Mutat. Res.* **407:** 217-226

Friedberg, E.C., Walker, G.C., Siede, W. *DNA repair and mutagenesis.* Washington, DC. ASM Press; 1995

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

Grossman, M., Raper, S.E., Kozarsky, K., Stein, E.A., Engelhardt, J.F., Muller, D., Lupien, P.J. and Wilson, J.M. (1994) Successful *ex vivo* gene therapy directed to liver in a patient with familial hypercholesterolaemia. *Nat. Genet.* **6:**335-341

Hanahan, D. (1983) Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Biol.* **166:** 557-589

Hayashi, T., Takao, M., Tanaka, K. and Yasui, A. (1998) ERCC1 mutations in UV-sensitive Chinese hamster ovary (CHO) cell lines. *Mutat Res.* **408:** 269-276

He, Z., Henricksen, L.A., Wold, M.S. and Ingles, C.J. (1995) RPA involvement in the damage-recognition and incision steps of nucleotide excision repair. *Nature* **374:** 566-569

Hodnett, D.W., Fantozzi, D.A., Thurmond, D.C., Klautky, S.A., MacPhee, K.G., Estrem, S.T., Xu, G. and Goodridge, A.G. (1996) The chicken malic enzyme gene: structural organisation and identification of triiodothyronine response elements in the 5' flanking DNA. *Arch. Biochem. Biophys.* **334:** 309-324

Hoeijmakers, J.H.J. (1993) Nucleotide excision repair II: from yeast to mammals. *Trends Genet.* **9:** 211-217

Hoeijmakers, J.H.J., van Duin, M., Westerveld, A., Yasui, A. and Bootsma, D.(1986) Identification of DNA repair genes in the human genome. Cold Spring Harbour Symposia on Quantitative Biology. **LI:** 91-101

Hogan, B., Costantini, F. and Lacy E. *Manipulating the mouse embryo* Cold Spring Harbour Laboratory, 1986

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

Huang, J-C., Svoboda, D.L., Reardon, J.T. and Sancar, A. (1992) Human excision nuclease removes thymine dimers from DNA by incising the 22nd phosphodiester bond 5' and the 6th phosphodiester bond 3' to the photodimer. *Proc. Natl. Acad. Sci. USA* **89:**3664-3668

Ish-Horowicz, D. and Burke, J.F. (1981) Rapid and efficient cosmid cloning. *Nucleic Acids Res.* **9:** 2989-2998

Jackson, S.P. and Jeggo, P.A. (1995) DNA double-strand break repair and V(D)J recombination: involvement of DNA-PK. *Trends Biochem. Sci.* **20:** 412-415

Jimbo, T., Ichihashi, M., Mishima, Y. and Fujiwara, Y. (1992) Role of excision repair in UVB-induced depletion and recovery of human epidermal Langerhans cells. *Arch. Dermatol.* **128:** 61-67

Jiricny, J. (1998) Eukaryotic mismatch repair: an update. *Mutat. Res.* **409:** 107-121

Kanaar, R., Hoeijmakers, J.H.J. and van Gent, D.C. (1998) Molecular mechanisms of DNA double-strand break repair. *Trends Cell Biol.* **8:** 483-489

Kay, M.A., Baley, P., Rothenburg, S., Leland, F., Fleming, L., Ponder, K.P., Liu, T., Finegold, G., Darlington, G. and Pokorny, W. (1992) Expression of human α1-antitrypsin in dogs after autologous transplantation of retroviral transduced hepatocytes. *Proc. Natl. Acad. Sci. USA* **89:**89-93

Kilby, N.J., Snaith, M.R. and Murray, J.A. (1993) Site-specific recombinase: tools for genome engineering. *Trends Genet.* **9:** 413-421

Kiledjian, M. and Kadesch, T. (1991) Post-transcriptional regulation of the human liver/bone/kidney *alkaline phosphatase* gene. *J. Biol. Chem.* **266:** 4207-4213

Klungland, A., Hoss, M., Gunz, D., Constantinou, A., Clarkson, S.G., Doetsch, P.W., Bolton, P.H., Wood, R.D. and Lindahl, T. (1999) Base excision repair of oxidative DNA damage activated by XPG protein. *Mol. Cell.* **3:** 33-42

Konecki, D.S., Brennand, J., Fuscoe, J.C., Caskey, C.T. and Chinault, A.C. (1982) Hypoxanthine-guanine phosphoribosyltransferase genes of mouse and Chinese hamster: construction and sequence analysis of cDNA recombinants. *Nucleic Acids Res.* **10:** 6763-6775

Lappin, M.B., Kimber, I., Norval, M. (1996) The role of dendritic cells in cutaneous immunity. *Arch. Dermatol. Res.* **288:** 109-121

Legerski, R. and Peterson, C. (1992) Expression cloning of a human DNA repair gene involved in xeroderma pigmentosum group C. *Nature* **359:** 70-73

Lesch, M. and Nyhan, W.L. (1964) A familial disorder of uric acid metabolism and central nervous system function. *Am. J. Med* **36:** 561-570

Li, L., Elledge, S.J., Peterson, C.A., Bales, E.S. and Legerski, R.J. (1994) Specific association between the human DNA repair proteins XPA and ERCC1. *Proc. Natl. Acad. Sci. USA* **91:** 5012-5016

Li, Y., Li, D., Osborn, K. and Johnson, L.F. (1991) The 5' flanking region of the mouse *thymidylate synthase* gene is necessary but not sufficient for normal regulation in growth-stimulated cells. *Mol. Cell. Biol.* **11:** 1023-1029

Lindahl, T., Karran, P. and Wood, R.D. (1997) DNA excision pathways. *Curr. Opin. Genetics Dev.* **7:** 158-169

Lu, Q., Wallrath, L.L., Granok, H. and Elgin, S.C.R. (1993) $(CT)_n$. $(GA)_n$ Repeats and heat shock elements have distinct roles in chromatin structure and transcriptional activation of the *Drosophila hsp26* gene. *J. Mol. Biol.* **13:** 2802-2814

Macmahon, A.P. and Bradley, A. (1990) The *Wnt-1* (*int-1*) proto-oncogene is required for development of a large region of mouse brain. *Cell* **62:** 1073-1085

Magin, T.M., McWhir, J. and Melton D.W. (1992) A new mouse embryonic stem cell line with good germline contribution and gene targeting frequency. *Nucleic Acids Res.* **20:** 3795-3796

Mandel, M. and Higa, A. (1970) Calcium dependent bacteriophage DNA infection. *J. Mol. Biol.* **53:** 159-162

Mansour, S.L., Thomas, K.R. and Capecchi, M.R. (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

McWhir, J., Selfridge, J., Harrison, D.J., Squires, S. and Melton, D.W. (1993) Mice with DNA repair gene (*ERCC1*) deficiency have elevated levels of p53, liver nuclear abnormalities and die before weaning. *Nat. Genet.* **5:** 217-224

Melton, D.W. (1994) Gene targeting in the mouse. *Bioessays* **16:** 633-638

Melton, D.W., Ketchen, AM., Nunez, F., Bonatti-Abbondandolo, S., Abbondandolo, A., Squires, S. and Johnson, R.T. (1998) Cells from *ERCC1*-deficient mice show increased genome instability and a reduced frequency of S-phase-dependent illegitimate chromosome exchange but a normal frequency of homologous recombination. *J Cell Sci.* **111:**394-404

Minty, A.J., Caravatti, M., Robert, B., Cohen, A., Daubas, P., Weydert, A., Gros, F. and Buckingham, M.E. (1981) Mouse actin messenger RNAs. Construction and characterisation of a recombinant plasmid molecule containing a complementary transcript of mouse α-actin mRNA. *J. Biol. Chem.* **256:** 1008-1014

Miyauchi-Hashimoto, H., Tanaka, K. and Horio, T. (1996) Enhanced inflammation and immunosuppression by ultraviolet radiation in xeroderma pigmentosum group A (XPA) model mice. *J. Invest. Dermatol.* **107:** 343-348

Moggs, J.G., Yarema, K.J., Essigmann, J.M. and Wood, R.D. (1996) Analysis of incision sites produced by human cell extracts and purified proteins during nucleotide excision repair of a 1,3-intrastrand d(GpTpG)-cisplatin adduct. *J. Biol. Chem.* **271:** 7177-7186

Montecucco, A., Savini, E., Biamonti, G., Stefanini, M., Focher, F. and Ciarrocchi, G. (1995) Late induction of human ligase I after UV-C irradiation. *Nucleic Acids Res.* **23:** 962-966

Moolenaar, G.F., Uiterkamp, R.S., Zwijnenburg, D.A. and Goosen, N. (1998) The C-terminal region of the *Escherichia coli* UvrC protein, which is homologous to the C-terminal region of the human ERCC1 protein, is involved in DNA binding and 5' incision. *Nucleic Acids Res.* **26:** 462-468

Moore, R.C., Redhead, N.J., Selfridge, J., Hope, J., Manson, J.C. and Melton, D.W. (1995) Double replacement gene targeting for the production of a series of mouse strains with different prion protein gene alterations. *Biotechnology* **13:** 999-1004

Mu, D., Hsu, D.S. and Sancar, A. (1996) Reaction mechanism of human DNA repair excision nuclease. *J. Biol. Chem.* **271:** 8285-8294

Mudgett, J.S. and MacInnes, M.A. (1990) Isolation of the functional human excision repair gene *ERCC5* by intercosmid recombination. *Genomics* **8:** 623-633

Muller, H.K., Halliday, G.M. and Woods, G.M. (1997) The skin immune system and tumor immunosurveillance. In *Skin immune system.* 2nd edition pp417-430, Edited by. Bos, J.D. (CRC Press, 1997)

Myrand, S.P., Topping, R.S. and States, J.C. (1996) Stable transformation of xeroderma pigmentosum group A cells with an *XPA* minigene restores normal DNA repair and mutagenesis of UV-treated plasmids. *Carcinogenesis* **17:** 1909-1917

Nakane, H., Takeuchi, S., Yuba, S., Saijo, M., Nakatsu, Y., Murai, H., Nakatsuru, Y., Ishikawa, T., Hirota, S., Kitamura, y., Kato, Y., Tsunoda, Y., Miyauchi, H., Horio, T., Tokunaga, T., Matsunaga, T, Nikaido, o., Nishimune, Y., Okada, Y. and Tanaka, K. (1995) High incidence of ultraviolet-B or chemical-carcinogen- induced skin tumours in mice lacking the xeroderma pigmentosum group A gene. *Nature* **377:** 165-168

O'Donovan, A. and Wood, R.C. (1993) Identical defects in DNA repair in xeroderma group G and rodent ERCC group 5. *Nature* **363:** 185-188

O'Donovan, A., Davies, A.A., Moggs, J.G., West, S.C. and Wood, R.C. (1994) XPG endonuclease makes the 3' incision in human DNA nucleotide excision repair. *Nature* **372:** 432-435

Park, C-H. and Sancar, A. (1994) Formation of a ternary complex by human XPA, ERCC1 and ERCC4 (XPF) excision repair proteins. *Proc. Natl. Acad. Sci. USA.* **91:** 5017-5021

Park, C-H., Bessho, T., Matsunaga, T. and Sancar, A. (1995) Purification and characterisation of the XPF-ERCC1 complex of human DNA repair excision nuclease. *J. Biol. Chem.* **270:** 22657-22660

Poole, T. (ed.), *The UFAW Handbook on the Care and Management of Laboratory Animals* 6th edition (Bath, Longmann Scientific and Technical, 1989)

Prochownik, E.V. and Orkin, S.H. (1984) *In vivo* transcription of a human *antithrombin III* minigene. *J. Biol. Chem.* **259:** 15386-15392

Quilliet, X., Chevier-Lagente, O., Eveno, E., Stojkovic, T., Destee, A., Sarasin, A. and Mezzina, M. (1996) Long-term complementation of DNA repair deficient human fibroblasts by retroviral transduction of the XPD gene. *Mutat. Res* **364:** 161-169

Redhead, N.J., Selfridge, J., Wu, C-L. and Melton, D.W. (1996) Mice with adenine phosphoribosyltransferase deficiency develop fatal 2,8-dihydroxyadenine lithiasis. *Hum. Gene Therapy.* **7:** 1491-1502

Reed, K.C. and Mann, D.A. (1985) Rapid transfer of DNA from agarose gels to nylon membranes. *Nucleic Acids Res.* **13:** 7207-7221

Robertson, G., Garrick, D., Wilson, M., Martin, D.I.K. and Whitelaw, E. (1996) Age-dependent silencing of globin transgenes in the mouse. *Nucleic Acids Res.* **24:** 1465-1471

Robins, P., Jones, C.J., Biggerstaff, M., Lindahl, T. and Wood, R. (1991) Complementation of DNA repair in xeroderma pigmentosum group A cell extracts by a protein with affinity for damaged DNA. *EMBO J.* **10:**3913-3921

Rossant, J. and McMahon, A. (1999) "Cre"-ating mouse mutants- a meeting review on conditional mouse genetics. *Genes. Dev.* **13:** 142-145

Saijo, M., Kuraoka, I., Masutani, C., Hanaoka, F. and Tanaka, K. (1999) Sequential binding of DNA repair proteins RPA and ERCC1 to XPA *in vitro. Nucleic Acids Res.* **24:**4729-4735

Sancar, A. (1996) DNA excision repair. *Ann. Rev. Biochem.* **65:** 43-81

Sands, A.T., Abuin, A., Sanchez, A., Conti, C.J. and Bradley, A. (1995) High susceptibility to ultraviolet-induced carcinogenesis in mice lacking *XPC. Nature* **377:** 162-165

Scharfmann, R., Axlerod, J.H. and Verma, I.M. (1991) Long-term *in vivo* expression of retrovirus mediated gene transfer in mouse fibroblast implants. *Proc. Natl. Acad. Sci. USA* **88:** 4626-4630

Schwenk, F., Kuhn, R., Angrand, P-O., Rajewsky, K. and Stewart, A.F. (1998) Temporally and spatially regulated somatic mutagenesis in mice. *Nucleic Acids Res.* **26:** 1427-1432

Selbert, S., Bentley, D.J., Melton, D.W., Rannie, D., Lourenco, P., Watson, C. and Clarke, A.R. (1998) Efficient BLG-Cre mediated gene deletion in the mammary gland. *Transgenic Res.* **7:** 387-396

Selfridge, J., Pow, A.M., McWhir, J., Magin, T.M. and Melton, D.W. (1992) Gene targeting using a mouse HPRT minigene/ HPRT-deficient embryonic stem cell system: Inactivation of the mouse *ERCC1* gene. *Somat. Cell Mol. Genet.* **18:** 325-336

Sijbers, A.M., De Laat, W.L., Ariza, R.R., Biggerstaff, M., Wei, Y-F., Moggs, J.G., Carter, K.C., Shell, B.K., Evans, E., De Jong, M.C., Rademakers, S., de Rooj, J., Jaspers, N.G.J., Hoeijmakers, J.H.J. and Wood, R.D. (1996) Xeroderma pigmentosum group F caused by a structure-specific DNA repair endonuclease. *Cell* **86:**811-822

Sijbers, A.M., van der Spek, P.J., Odijk, H., van den Berg, J., van Duin, M., Westerveld, A., Jaspers, N.G.J., Bootsma, D. and Hoeijmakers, J.H.J. (1996) Mutational analysis of the human nucleotide repair gene *ERCC1*. *Nucleic Acids Res.* **24:** 3370-3380

Smith, A. (1995) Viral vectors in gene therapy. *Ann. Rev. Microbiol.* **49:**807-838

Smith, G.E. and Summers, M.D. (1980) The bi-directional transfer of DNA and RNA to nitrocellulose or diazobenzyloxymethyl-paper. *Anal. Biochem.* **190:** 123-129

Southern, E.M. (1975) Detection of specific sequences among DNA fragments separated by agarose gel electrophoresis. *J. Mol. Biol.* **98:** 503-517

Stacey, A., Schnieke, A., McWhir, J., Cooper, J., Colman, A. and Melton, D.W. (1994) Use of double-replacement gene targeting to replace the murine α-*lactalbumin* gene with its human counterpart in embryonic stem cells and mice. *Mol. Cell. Biol.* **14:** 1009-10016

Sugasawa, K., Ng, J.M.Y., Masutani, C., Iwai, S., van der Spek, P.J., Eker, A.P.M., Hanaoka, F., Bootsma, D. and Hoeijmakers, J.H.J. (1998) Xeroderma pigmentosum group C protein complex is the initiator of global genome nucleotide excision repair. *Mol. Cell* **2:** 223-232

Tanaka, K., Satokata, I., Ogita, Z., Uchida, T. and Okada, Y. (1989) Molecular cloning of a mouse DNA repair gene that complements the defect of group-A xeroderma pigmentosum. *Proc. Natl. Acad. Sci. USA* **86:** 5512-5516

Thompson, L.H., Brookman, K.W., Weber, C.A., Salazar, E.P., Reardon, J.T., Sancar, A., Deng, Z. and Siciliano, M.J. (1994) Molecular cloning of the human nucleotide-excision-repair gene *ERCC4*. *Proc. Natl. Acad. Sci. USA* **91:** 6855-6859

Thompson, S., Clarke, A., Pow, A.M., Hooper, M.L. and Melton, D.W. (1989) Germ line transmission of a corrected *HPRT* gene produced by gene targeting in embryonic stem cells. *Cell* **56:** 313-321

Thummel, C.S., Boulet, A.M. and Lipshitz, H.D. (1988) Vectors for *Drosophila* P-element-mediated transformation and tissue culture transfection. *Gene* **74:** 445-456

Tomkinson, A.E., Bardwell, A.J., Bardwell, L., Tappe, N.J. and Friedberg, E.C. (1993) Yeast DNA repair and recombination proteins Rad1 and Rad10 constitute a single stranded-DNA endonuclease. *Nature* **362:** 860-862

Troelstra, C., Odijk, H., de Wit, J., Westerveld, A., Thompson, L.H., Bootsma, D. and Hoeijmakers, J.H.J. (1990) Molecular cloning of the Human DNA excision repair gene *ERCC6. Mol. Cell. Biol.* **10:** 5806-5813

van der Horst, G.T., van Steeg, H., Berg, R.J., van Gool, A.J. de Wit, J., Weeda, G., Morreau, H., Beems, R.B., van Kreijl, C.F., de Gruijl, F.R., Bootsma, D. and Hoeijmakers, J.H. (1997) Defective transcription-coupled repair in Cockayne syndrome B mice is associated with skin cancer predisposition. *Cell* **89:** 425-435

van Duin, M., de Wit, J., Odijk, H., Westerveld, A., Yasui, A., Koken, M.H.M., Hoeijmakers, J.H.J. and Bootsma, D (1986) Molecular characterisation of the human excision repair gene *ERCC1:* cDNA cloning and amino acid homology with the yeast DNA repair gene *RAD10. Cell* **44:**913-923

van Duin, M., Koken, M.H.M., van en Tol, J., ten Dijke, P., Odijk, H., Westerveld, A., Bootsma, D. and Hoeijmakers, J.H.J. (1987) Genomic characterisation of the human DNA excision repair gene *ERCC1. Nucleic Acids Res.* **15:** 9195-9213

van Duin, M., van den Tol, J., Hoeijmakers, J.H.J., Bootsma, D., Rupp, I.P., Reynolds, P., Prakash, L. and Prakash, S. (1989) Conserved pattern of antisense overlapping transcription in homologous human *ERCC1* and yeast *RAD10* DNA repair gene regions. *Mol. Cell. Biol* **9:** 1794-1798

van Duin, M., van den Tol, J., Warmendam, P., Odijk, H,. Meijer, D., Westerveld, A., Bootsma, D. and Hoeijmakers, J.H.J. (1988) Evolution and mutagenesis of the mammalian excision repair gene *ERCC1. Nucleic Acids Res. 16(12): 5305-5322*

van Duin, M., Vredeveldt, G., Mayne, L.V., Odijk, H., Vermeulen, W., Klein, B., Weeda, G., Hoeijmakers, J.H.J., Bootsma, D. and Westerveld, A. (1989) The cloned human DNA excision repair gene ERCC1 fails to correct xeroderma pigmentosum complementation group A through group I. *Mutat. Res.* **217:** 83-92

van Oers, C.C., Bakker, L. and Baas, P.D. (1994) The exon 4 poly(A) site of the human calcitonin/CGRP-I pre-mRNA is a weak site *in vitro. Biochim. Biophys. Acta* **1218:** 55-63

Van Oostrom, C.T.M., de Vries, A., Verbeek, S.J., van Kreijl, C.F. and van Steeg, H. (1994) Cloning and characterization of the mouse *XPAC* gene. *Nucleic Acids Res* **22:** 11-14

van Vuuren, A.J., Appeldoorn, E., Odijk, H., Yasui, A., Jaspers, N.G.J., Bootsma, D. and Hoeijmakers, J.H.J. (1993) Evidence for a repair enzyme complex involving ERCC1 and complementing activities of ERCC4, ERCC11 and xeroderma pigmentosum group F. *EMBO.J.* **12:** 3696-3701

van Vuuren, A.J., Vermeulen, W., Ma, L., Weeda, G., Appeldoorn, E., Jaspers, N.G.J., van der Eb, A.J., Bootsma, D. and Hoeijmakers, J.H.J and Humbert, S., Schaeffer, L. and Egly, J.-M. (1994) Correction of xeroderma pigmentosum repair defect by basal transcription factor BTF2 (TFIIH). *EMBO J.* **13:** 1645-1653

Venema, J., van Hoffen, A., Karcagi, V., Natarajan, A.T., van Zeeland, A.A. and Mullenders, L.H.F. (1991) Xeroderma pigmentosum complementation group C cells remove pyrimidine dimers selectively from the transcribed strand of active genes. *Mol. Cell. Biol.* **11:** 4128-4134

Venolia, L., Urlaub, G., and Chasin, L.A.(1987) Polyadenylation of Chinese hamster *dihydrofolate reductase* genomic genes and minigenes after gene transfer. *Somat. Cell Mol. Genet.* **13:** 491-504

Wang, P., Chen, H., Qin, H., Sankarapandi, S., Becher, M.W., Wong, P.C. and Zweier, J.L. (1998) Overexpression of human copper, zinc-superoxide dismutase (SOD1) prevents postischemic injury. *Proc. Natl. Acad. Sci. USA* **95:** 4556-4560

Weber, C.A., Salazar, E.P., Stewart, S.A. and Thompson, L.H. (1988) Molecular cloning and biological characterization of a human gene, *ERCC2,* that corrects the nucleotide repair defect in CHO UV5 cells. *Mol. Cell. Biol.* **8:** 1137-1146

Weber, C.A., Salazar, E.P., Stewart, S.A. and Thompson, L.H. (1990) *ERCC2:* cDNA cloning and characterization of a human nucleotide excision repair gene with high homology to yeast *RAD3. EMBO J.* **9:** 1437-1447

Weeda, G., Donker, I., de Wit, J., Morreau, H., Janssens, R., Vissers, C.J., Nigg, A., van Steeg, H., Bootsma, D. and Hoeijmakers, J.H.J.(1997) Disruption of mouse *ERCC1* results in a novel repair syndrome with growth failure, nuclear abnormalities and senescence. *Curr. Biol.* **7:** 427-439

Weeda, G., van Ham, R.C.A., Masurel, R., Westerveld, A., Odijk, H., de Wit, J., Bootsma, D., van der Eb, A.J. and Hoeijmakers, J.H.J. (1990) Molecular cloning and biological characterization of the human excision repair gene *ERCC3. Mol. Cell. Biol.* **10:** 2570-2581

Weis, L. and Reinberg, D. (1992) Transcription by RNA polymerase II: initiator-directed formation of transcription-competent complexes. *FASEB J* **6:** 3300-3309

Westerveld, A., Hoeijmakers, J.H.J., van Duin, M., de Wit, J., Odijk, O., Pastnik, A., Wood, R.D. and Bootsma, D. (1984) Molecular cloning of a human DNA repair gene. *Nature* **310:** 425-429

Wilkins, C.R. and Lis, J.T. (1998) GAGA factor binding to DNA via a single trinucleotide sequence element. *Nucleic Acids Res* **26:** 2672-2678

Wood, R.D. (1996) DNA repair in eukaryotes. *Ann. Rev. Biochem* **65:** 135-167

Wood, R.D. and Burki, H.J. (1982) Repair capability and the cellular age response for killing and mutation induction after UV. *Mutat. Res.* **95:** 505-514

Wu, H., Wade, M., Krall, L., Grisham, J., Xiong, Y. and Van Dyke, T. (1996) Targeted in vivo expression of the cyclin-dependent kinase inhibitor p21 halts hepatocyte cell-cycle progression, post-natal liver development and regeneration. *Genes Dev.* **10:** 245-260

Xiao, Y. and Weaver, D. (1997) Conditional gene targeted deletion by Cre recombinase demonstrates the requirement for the double-strand break repair Mre 11 protein in murine embryonic stem cells. *Nucleic Acids Res.* **25:** 2985-2991

Xu, G. and Goodridge, A.G.(1998) A CT repeat in the promoter of the chicken malic enzyme is essential for function at an alternative transcription start site. *Arch. Biochem. Biophys.* **358:** 83-91

Yagi, T., Wood, R.D. and Takebe, H. (1997) A low content of ERCC1 and a 120 kDa protein is a frequent feature of group F xeroderma pigmentosum fibroblast cells. *Mutagenesis* **12:** 41-44

Yan, C., Costa, R.H., Darnell, J.E., Chen, J. and Van Dyke, T. (1990) Distinct positive and negative elements control the limited hepatocyte and choroid plexus expression of transthyretin in transgenic mice. *EMBO J.* **9:** 869-878

Zeng, L., Quilliet, X., Chevier-Lagente, O., Eveno, E., Sarasin, A. and Mezzina, M. (1997) Retrovirus-mediated gene transfer corrects DNA repair defect of xeroderma pigmentosum cells of complementation groups A, B and C. *Gene Therapy* **4:** 1077-1084

Zhumabekov, T., Corbella, P., Tolaini, M. and Kioussis, D. (1995) Improved version of a human *CD2* minigene based vector for T cell-specific expression in transgenic mice. *J Immunol Methods* **185:** 133-140