# The Patchy Human Genome:
## DNA Double-strand Breaks in Human Cells can be Repaired by the Capture of Other DNA Fragments

Kevin C.E. Little
Department of Medicine, Division of Experimental Medicine
McGill University, Montreal

August 2004

Doctor of Philosophy (Ph.D.)

A thesis submitted to McGill University in partial fulfilment of the requirements of the degree of doctor of philosophy (Ph.D.)

*Your file Votre référence*
*ISBN: 0-494-12886-0*
*Our file Notre référence*
*ISBN: 0-494-12886-0*

# TABLE OF CONTENTS

# ACKNOWLEDGEMENTS

"To finish the moment, to find the journey's end in every step of the road, to live the greatest number of good hours, is wisdom."

— Ralph Waldo Emerson

This thesis process has taught me a great deal about research, and about science, but mostly it has afforded me the opportunity to learn about myself. Like the day to day work in the laboratory, the outcome of this self–discovery was not always what one might have expected, sometimes disappointing and at other times illuminating. But, like any experiment, the results are all you have to work from – and if you went through the process properly, results don't lie. It is up to the observer to see universal truths in these few grains of sand, and this entire wondrous process has taught me how to see them, within and without.

I have received tremendous support and encouragement during the course of this work, for which I am deeply grateful. Members of our lab, past and present, have been invaluable resources for technical and moral support. For particular technical assistance, I thank Nancy Ringuette for her help with FISH slides. I wish to especially thank our technician Roxane Lussier for her enthusiasm, and eternal willingness to answer my many questions. I also thank Hugo Wurtele for inspiring discussions, and showing me the art in research. I cannot express enough the gratitude and admiration I have for my supervisor, Pierre Chartrand, who created an environment of exceptional opportunity. He has taught me by example that philosophy and vision make a scientist truly great.

Just as I could never have accomplished the research aspect of this work without the contributions of the aforementioned people, the support I have received from my entire family has been boundless. Nature first inspired me to learn the way of the scientist, and has now humbled and awed me with her generosity, bestowed upon us as a baby girl, Clara. For my wife Claude, who has stood with me through the best and the worst, and whose love has allowed me to appreciate each glorious moment of it all, this has been our journey together.

# ABSTRACT


## English

The human genome is composed of very few genes (the DNA which encodes proteins), and the remaining 98% of all our DNA is made up of repetitive sequences or "junk" DNA with little or no known function. Our complexity, and the cause of many genetic diseases such as cancer, is brought about by the differential expression of that genetic material, in part due to shuffling and movement of chromosomes within the cell nucleus. A major threat to the integrity of the genome is the occurrence of a DNA double–strand break (DSB). These DSBs occur frequently in every cell, and if repaired improperly or left unattended can lead to genomic rearrangements and cell death. With the publication of the human genome sequence, we were able to establish two systems to investigate the possibility that DNA fragments can insert into breaks sites during DSB repair in human cells.

We find that both foreign and human genomic DNA can insert into extrachromosomal and chromosomal DSBs. Genomic instability syndromes, like those which result from deficiencies in repair proteins, still permit this DSB insertional repair process, however the spectrum of source material provided can differ. The deregulation of replication origins, such as the amplification of sequences flanking viral integration sites, can lead to the spread and further gene amplification of DNA by this insertion mechanism.

Our demonstration that DSB insertional repair takes place in human cells provides a mechanism for reshuffling genomic DNA, and acquiring new sequences. We propose that a selective advantage is conferred upon a cell able to insert DNA at a DSB, providing a complexity of gene content and interaction which could explain the origins and evolutionary role of the vast majority of the human genome.

## Français

Le génome humain est composé de très peu de séquence géniques (l'ADN qui code pour les protéines), et d'environ 98% d'ADN répétitif sans fonction connue. La complexité de l'humain, et la cause des maladies génétiques comme le cancer, sont le résultat de l'expression différentielle du matériel génétique causé par le mélange et le mouvement des chromosomes dans le noyaux de la cellule. L'intégrité du génome est menacé par les nombreuses cassures double–brins de l'ADN (CDB). Ces CDBs se produisent fréquemment dans chaque cellule, et lorsqu'elles sont mal réparées ou non surveillées, celles–ci mènent aux réarrangements chromosomiques et à la mort cellulaire. Avec la publication de la séquence du génome humain, nous avons été capable d'établir deux systèmes dans des cellules humaines, pour étudier la possibilité d'insertion de fragments d'ADN dans des sites de brisures pendant la réparation des CDBs .

Nous avons observé que l'ADN, provenant de sources exogènes et endogènes, peut s'insérer dans des CDBs extrachromosomiques ou chromosomiques. Les cellules souffrant d'instabilité génomique, comme ceux qui résultent des insuffisances en protéines de réparation, sont toujours capables de réparer les CDBs avec l'insertion de fragment d'ADN. Par contre, en comparaison avec les cellules normales, la source du matériel peut différer. La dérégulation des origines de réplication, ainsi que l'amplification des séquences génomiques flanquant les sites d'intégration des virus, peuvent mener à la diffusion et à l'amplification des gènes par ce mécanisme d'insertion.

Nous avons mis en évidence que la réparation des CDBs par insertion a lieu chez l'humain. Ce mécanisme permet une réorganisation de l'ADN génomique et l'aquisition de nouvelles séquences qui confère un avantage sélectif aux cellules. Ce mécanisme augmente la complexité du génome et les interactions entres les gènes, et pourrait expliquer les origines et le rôle évolutif de la grande majorité du génome humain.

# DECLARATION OF ORIGINAL CONTRIBUTION

The work described in this thesis constitutes novel and original contributions to the field of DNA repair, and to our understanding of the human genome, as follows:

- The specific identification of insertions of foreign and human genomic DNA fragments at sites of deliberate double–strand breaks (DSBs), both extrachromosomally and within chromosomes, in human cells (termed "DSB insertional repair").
- Establishment of two separate assays which are capable of identifying, and characterizing, events of DSB misrepair that result specifically in large insertions at the break site.
- Development of a dual–expression plasmid for the I–Sce I endonuclease and GFP, which permits identification and separation of transfected cells.
- Demonstration that DSBs can act as recipient sites for fragment insertions, as well as initiating events of fragment production at the donor loci. This indicates multiple DSBs are co-localized to the same site during DNA repair in human cells.
- Discovery that endogenous sites in the human genome are cleaved by the I–Sce I endonuclease, despite some degeneracy of the 18 bp recognition sequence.
- Demonstration that following insertion at a DSB, inserted fragments:
    - Can span exons and/or introns
    - Can be expressed from the new locus (as with cases of T-ag)
    - Can be inherited by daughter cells with no apparent detriment to cell viability
- Demonstration that integration of sequences spanning the SV40 origin of replication, in the presence of T-ag expression, can cause amplification of human genomic DNA which can be used during DSB insertional repair.

# 1. INTRODUCTION

This is an extraordinarily exciting time for biologists. A wave of new technologies are peeling back the veil of mystery which has long obscured our understanding of the organization and functioning of the human genome. The visualization and tracking of nucleic acid and protein movements in real–time and in three dimensions is possible with new labelling techniques and protein fusions, in conjunction with computer–aided microscopy. In concert with genomic sequencing data, and whole–genome expression arrays technologies, these new tools put us at a unique frontier in our understanding of our origins and the fundamental processes of life.

In this Introduction, we provide a brief overview of our genomic make–up, and how it is organized in the cell nucleus. While systems exist to reduce the threat of genetic alterations, and maintain homeostasis, therein lie also the seeds of change – the means to rearrange and revolutionize genomes on an evolutionary scale. We discuss this evolutionary process in the context of how a human cell responds to damage incurred in its DNA, particularly in the form of a double–strand break (DSB).

Our focus is a mechanism for repairing DSBs which has only recently been characterized in other organisms but not in humans: the insertion of DNA fragments into DSB sites. We introduce this topic in the context of work that was ongoing in our lab during development of this project. The objective of this thesis work was to explore the possible occurrence of this DSB insertional repair in a human system, and avail ourselves of the recently published human genome sequence to identify and characterize the source of any insertions.

## *1.1 The Contents of the Human Genome*

The human "genome" is a term applied to the collective assembly of deoxyribonucleic acid (DNA) in the cell nucleus, to the particular set of DNA sequences that distinguish one individual from another, and to the entire genetic complement that defines us as a

species. At the cellular level, a genome consists of approximately $3.2x10^9$ nucleotide base pairs (3.2 Gb) distributed between 23 long molecules of DNA, the chromosomes. Each somatic cell contains twice this amount, 6.2 Gb in 46 chromosomes (22 autosomal pairs and 2 sex chromosomes).

Drafts of the precise order of these bases, the DNA sequence, were published in 2001 by two groups (each one a composite of source DNA from a few individuals) [1,2]. Both groups estimate that when analysis of the raw data is complete, and we better understand how to properly identify and define a gene, the human genome will be found to contain about 32,000 genes. These genes are usually composed of multiple exons (segments which code for proteins) interspersed by introns (non–coding segments which are excised before the messenger molecule, RNA, is translated into protein). As it stands, 1.1% – 1.5 % of the human genome sequence spans exons, with another 24% – 36% in introns [1,2].

One might wonder then, with the exception these exons, what is the remaining 98% of the genome composed of? Analyses of the draft human genome confirmed earlier predictions, and established that one–half of our genome is most certainly made up of ancient retroviruses and transposable elements which fall into four major classes: LINEs, SINEs, LTRs, and DNA transposons (see Figure 1). The discovery of these repetitive DNAs, with no corresponding explanation for any adaptive purpose, led to this fraction of the genome (in numerous organisms) being called "junk", or "selfish" DNA [3,4]. Repetitive DNA does help to explain observations that the overall size of a genome is not proportional to an organism's complexity: what is called the *C*–value paradox.

Long interspered nuclear elements (LINEs) make up about 21% of our genomic DNA. An intact LINE element is 6.1 kb long, and encodes its own proteins for making cDNA copies of itself and integrating those elsewhere in the genome. However, the vast majority of LINEs in the genome are not the complete 6.1 kb actively retrotransposing type, but rather are truncated and thought to be incompetent for retrotransposition. There are nearly 900,000 copies of LINEs averaging only 900 bp long scattered thoughout the genome, and more than half of those belonging to the only active family of LINE, L1 [1,2].

Short interspersed nuclear elements (SINEs) are another class of repeats, accounting for over 13% of all the DNA in our genome. SINEs are only 100 bp – 400 bp long when full–length, can be transcribed into RNA but encode no proteins, and share the same sequence as LINEs at their 3' end. This has long suggested that SINEs spread by attracting the LINE machinery to their transcipts for retrotransposition. There are about 1,600,000 copies of SINEs throughout the genome, and over two–thirds of those belong to the only active family of SINE, the Alu [1,2].

Long terminal repeat retroposons (LTRs) are 6 kb – 11 kb long when full–length, make up over 8% of the genome, and also encode their own reverse transcriptase and other machinery for making and integrating copies. Like LINEs, the vast majority of LTRs are not active, but in this case due to the loss of sequences internal to the flanking repeats. Other mRNAs which may co–opt the reverse transciptase activity of LTRs or LINEs may be responsible for the partial "pseudogene" sequences making up nearly 1% of the genome [1,2].

The DNA transposons are a varied group of elements which do not reverse transcribe, but rather encode a transposase enyzme that cuts out terminal repeat sequences and pastes them elsewhere in the genome. These are 2 kb – 3 kb full–length, and make up nearly 3% of the genome, but are believed to no longer be active [1,2].

In all, these transposable elements can be recognized as contributing to about 45% of the bulk of our genome. There are also the psuedogenes, simple sequence repeats (SSRs), and several large segmental duplications between chromosomes. The remaining one–half of the genome is "unique" non–coding DNA, which is probably derived largely from repetitive DNA which has degraded and genetically drifted beyond our ability to recognize it [1,2].

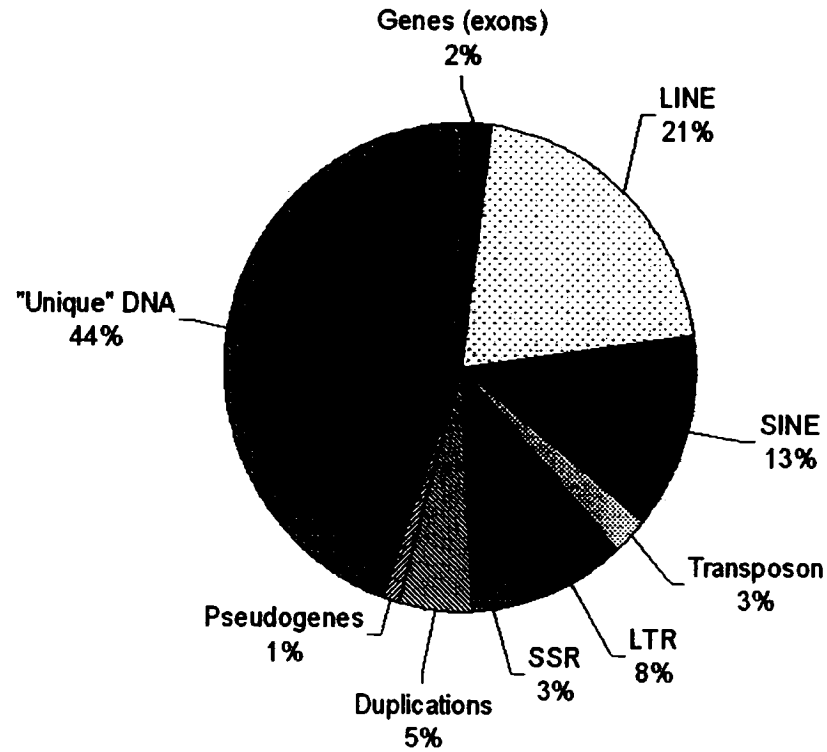



Figure 1. DNA content of the human genome. These estimates are taken from the analysis of the draft human genome sequence published by the International Human Genome Sequencing Consortium [1]. The estimated size of the human genome is $3.2 \times 10^9$ base pairs per haploid genome, broken down by percentage of nucleotides into the following categories. LINE: long interspersed nuclear element. SINE: short interspersed nuclear element. LTR: long terminal repeat retroposons. SSR: simple sequence repeat. Duplications: large intrachromosomal and interchromosomal segmental blocks which are present two or more times. Pseudogenes: potentially coding or non-coding sequences derived by reverse transcription from other RNAs. "Unique" DNA: unidentifiable as any known sequence, believed to be largely the result of genetic drift of repetitive elements.

## 1.2 Genome Complexity

At 3,200 Mb our genome is certainly larger than those of some simpler organisms frequently used as model systems, such as: the yeast *Saccharomyces cerevisiae* (12 Mb); the nematode worm *Caenorhabditis elegans* (100 Mb); or the fruitfly *Drosophila melanogaster* (180 Mb) [1,2]. However, other creatures we may shy from labelling as more complex than ourselves have in fact phenomenally larger genomes, such as: the pine tree *Pinus resinosa* (68,000 Mb); or the unicellular protozoan *Amoeba dubia* (670,000 Mb) [5]. What was perhaps most surprisng from the comparisons of the sequenced genomes is that our 32,000 genes is a figure not far from the 13,600 genes of the fruitfly, or the 19,100 genes of the worm *C. elegans* [6,7] . These last two examples also demonstrate that genome size does not necessarily correlate with gene number, either.

One consolation for the humbling paucity of genes we have relative to other organisms is that we mix up what we do have in more complex ways. Once a gene is transcribed by the cell into a messenger RNA, a process of alternative splicing can remove not only the introns but some of the exons as well, producing different protein products from a single gene. Compared to other organisms, in humans there is a greater proportion of genes subject to splicing, and with a greater number of splice variants per gene. One estimate puts 60% of all human genes as candidates for alternative splicing, with an average of 3.2 splice variants per gene (compared to 22% of *C. elegans* genes at 1.3 variant forms per gene) [1]. While exon sizes are the same between species, our introns are significantly longer, which increases the chance an intron will contain a splice site. Alternative splicing may generate up to 5 times as many primary protein products in the human than are in the fruitfly or worm [1,2]. What this translates to, literally, is a greater complexity of proteins encoded by novel combinations of the same shared building blocks.

Most proteins contain multiple structural or functional units, called domains, and any one particular arrangement of domains constitutes a protein family [8]. We have virtually the same repertoire to draw from as most other organisms, as evidenced by the fact nearly 90% of the protein domains in humans are shared by fruitfly and *C. elegans* proteins [9].

But of the estimated 1278 families of proteins in humans, only 94 of these are specific to vertebrates (responsible for traits such as acquired immunity, increased cell signalling, and complexification of programmed cell death). Being able to pick and choose which exons to include in a transcript allows a mix–and–match of functional motifs per protein. So while we may not necessarily have that many different protein domains to work with compared to other species, part of our complexity stems from the fact we use more domains per protein, and in new combinations [10].

Now that we have an understanding of what our genome contains, and the complexity of its expression, we move on to look at how all this DNA is packaged and used within the cell nucleus.

## *1.3 Nuclear Organization: Packing and Moving*

### 1.3.1 The Nucleosome

On a local scale, DNA in the nucleus is complexed with a variety of proteins into chromatin. The primary structural organizers of chromatin are the histone proteins, which form an octamer around which approximately 147 bp of every 200 bp of DNA is coiled, to generate a nucleosome. This octamer can be composed of variant forms of histones, and these are subject to a number of post–translational modifications such as acetylation, methylation, and phosphorylation [11]. These modifications alter the affinity for histone binding to DNA, contributing to relatively open or closed conformations of chromatin called euchromatin and heterochromatin, respectively [12,13]. The degrees of condensation permit or prevent exposure of the associated DNA to proteins such as transcription factors, thus influencing the levels and timing of transcription, replication, and repair [14,15].

Experiments which integrate foreign genes into various genomic locations demonstrate that the chromatin context influences expression patterns, as well as the priority assigned to replication of a chromosomal region [16–18]. The pattern of post–translational modifications constitutes a histone code, which is heritable by daughter cells and transmissible even through the germ line [19,20]. This creates an epigenetic level of gene

regulation beyond just the sequence of DNA bases. In this way the differential expression of genes can lead to cellular differentiation, and possibly even speciation, while the disruption of appropriate chromatin patterns can lead to tumorigenesis [21–24].

### 1.3.2 Chromosomes and Territories

On a larger organizational scale, the nucleosomal strand is condensed by linker histones and other chromatin proteins for packaging into chromosomes. The challenge lies in compacting nearly 2 m of DNA into a nucleus 6 μm in diameter, while maintaining a degree of order that permits its use [25]. Evidence of DNA zig–zags, coils, twists, and folds has been used to invoke different models of higher–order packaging into loops, rosettes, solenoids, and an array of contortions [26–28]. Compared to a naked linear strand of DNA, chromosomes in the interphase nucleus are compacted on the order of 1000–fold, and 10,000–fold at metaphase, which must be undone and then re–established again with each cycle of DNA replication and cell division [25]. Our knowledge of exactly how our DNA is packaged, such that the necessary regions are accessible and functional at the required time, leaves much to the imagination and begs further investigation.

A great deal of study in the past several years, in conjunction with the development of imaging techniques, has given us a much better idea of how chromosomes are positioned and accessed by proteins in the nucleus. Interphase is the stage in the cell's cycle when it is typically fulfilling its role as a given cell type, expressing general house–keeping as well as cell type–specific genes from chromosomes which are at their least condensed state, and chromatin structures the least repressive. Interphase chromosomes occupy distinct volumes in the nucleus, referred to as "territories". The pattern of chromosome positioning and juxtaposition of territories is believed to coincide with gene expression differences during the cell cycle, development, and onset of tumorigensis [29,30].

How these arrangements are established and maintained from generation to generation also remains somewhat of a mystery. The possibility that similar patterns may persist along the metaphase plate could provide one way to assign addresses to chromosome

neighbourhoods during cell division [31,32]. Though there is still some controversy as to the extent to which structural proteins in the nucleus constitute a nuclear scaffold or matrix (or are merely artefacts of laboratory extraction techniques), chromosome attachments to proteins such as lamins may provide some means of maintaining and re–establishing patterns [33–35].

### 1.3.3 Concentrating Nuclear Activities into Specific Regions

Proteins aggregate into clusters as distinct foci at sites of DNA replication, transcription, and DNA damage [36–38]. There can be more than one chromosomal site included in such a cluster, which has led to proposed models of multiple DNA strands being pulled through a "factory" complex, which is de–commissioned and re–commissioned to perform different functions in the same general nuclear location [38–41]. This begins to integrate the machinery for, and reshape our understanding of, what have traditionally been studied as separate processes: DNA replication, repair, recombination, and transcription [42].

Generally, chromosomes with higher gene densities are located closer to the centre of the nucleus [43]. Localization of some chromosomes relative to the nuclear periphery can change as cells exit the cell cycle and become quiescent or senescent [44]. There is also evidence that larger chromosomes assume positions closer to the periphery of the nucleus [45]. However, only some cases present evidence for differences in relative chromosome positioning between normal and tumour cells, and a degree of variation in chromosome position may simply reflect normal variations in levels of gene expression from one cell to the next, and one cell generation to the next [46–48].

Most studies support the idea that active genes are localized to the periphery of a territory, or can decondense locally to extend out from the usual territorial confines (and infringe on the borders of neighbouring territories), making them more accessible to the high concentrations of transcriptional proteins found between territories [49–51]. The positioning of genes within territories, and of syntenic chromosome regions within the

nucleus, can be conserved across evolution from mouse to non–human primates to human in a gene–density dependent manner [52,53].

A non–random folding and positioning of chromosomes in the nucleus means loci can find themselves in the same 3–dimensional neighbourhood from one cell to the next. This can be advantageous for maintaining patterns of gene expression which have proven successful, but also provides a basis for recurrent translocations between different chromosomes, or between sites widely spaced along the same chromosome [54–57]. Perhaps the best example of several chromosomes converging into a single site for expression is the nucleolus, where high levels of ribosomal RNA are transcribed from multiple gene copies on several co–localized chromosomes [58]. This spatial arrangement in the nucleolus has been described to breakdown entirely upon inhibition of rRNA synthesis, suggesting gene activity itself may (in part) dictate recurrent chromosome movements and positioning within the nucleus [59].

We see then how the arrangement of DNA into nucleosomes, and then into chromosomes, exerts influence over the way in which it is used by the cell. The position a given segment of DNA occupies in the nucleus helps dictate its role, and priority. We now turn our attention to the primary lesion of DNA which risks altering this positional, functional, and genomic integrity in the nucleus: the DNA double–strand break.

## *1.4 DNA Double–strand Breaks (DSBs)*

Damage to DNA occurs frequently, and double–strand breaks (DSBs) are regarded as the most dangerous form of DNA lesion. DSB repair involves more complex co–ordination than the repair of other types of DNA damage, and the potential impact of their misrepair is more severe [60]. DSBs can be caused by external sources (chemicals or ionizing radiation), or by endogenous agents such as oxygen radicals from cellular respiration. Some of the cell's enzymes purposely induce DSBs to initiate events, such as crossing over of meiotic chromosomes. The RAG proteins "cut–and–paste" immunoglobulin genes to generate a variety of antigen receptors during V(D)J recombination, and

topoisomerases allow one or two DNA strands to pass through another to undo the knotting and tension of chromosome condensation [61,62]. A recent estimate puts the number of endogenous DSBs occurring in normal human fibroblasts at 50 breaks per cell, every cell cycle [63].

Breaks can trigger a number of signalling pathways in the cell. The presence of even a single DSB can cause cell cycle arrest, or programmed cell death (apoptosis) [64,65]. DSBs also affect the initiation and progress of DNA replication, as they can inhibit not only the advancement of replication forks but the activation or "firing" of replication origins [66]. It is vital to a cell's survival therefore that DSBs be recognized and dealt with, and human cells have evolved a number of mechanisms for repairing them. However, as with all systems, DNA repair can be imperfect and result in genomic rearrangements, such as large deletions or translocations of chromosomal segments. We thus turn our attention to the cellular responses to DNA double–strand breaks.

## *1.5 DSB Repair*

In order to respond to DNA damage, the cell must have the capacity to first recognize it, and then carry out its repair. This can be envisioned as occurring in three phases (though they may not be entirely distinct): sensing, marking, and repairing the DSB.

### 1.5.1 Sensing the DSB

The group of proteins principally responsible for sensing, and subsequently responding to DNA damage, is a family of kinases which includes: ATM (ataxia telangiectasia mutated), DNA–PKcs (the DNA–dependent protein kinase, catalytic subunit), and ATR (ataxia telangiectasia Rad3 related). Each one plays multiple non–redundant roles in the cell, even if they sometimes act upon the same substrates. They phosphorylate a variety of targets to activate DNA repair proteins, or trigger cascades which signal cells to halt (or re–initiate) the cell cycle at different stages or "checkpoints" [67]. While DNA–PKcs is involved more specifically with DNA repair and activation or suppression of apoptosis,

ATM and ATR play broader roles in the cell, both in repair and checkpoint control [68]. One substrate upon which all three kinases act is p53, a major mediator of checkpoints which can slow replication to allow the cell time to repair any DNA damage, and activate apoptosis [69,70].

These three kinases are exceptionally large proteins, and can physically associate with broken DNA ends [71]. This may mean they provide a physical protein foundation upon which various other protein components can assemble (and disassemble), like the functional "factories" observed to form in the nucleus [65]. For instance, ATM phosphorylates the checkpoint kinase Chk2 at sites of DSBs, and Chk2 can then diffuse throughout the nucleus to phosphorylate its own targets and control progression of the cell cycle [72,73]. By concentrating DNA repair into foci, with multiple chromosomal sites and kinases per "factory", the cell can better monitor both its cycling and genomic integrity. Having multiple copies of the kinase on site and functioning in concert enables the cell to transmit its signal more quickly.

Perhaps the least understood aspect of the DNA damage response is precisely how these kinases detect damage in the first place [67]. Interestingly, physical changes in the stresses and topology of chromatin that can result from a break have recently been suggested to be the "damage sensor" event which triggers autophosphorylation and activation of the cellular pool of ATM [74]. Again, the packaging of DNA in the nucleus and its alteration may have functional consequences, this time at the level of cell cycle control and DNA repair.

### 1.5.2 Marking the DSB

One of the earliest indications of a DSB is the rapid appearance of a phosphorylated histone H2AX (called γ–H2AX), and its spread outwards in both directions from the break, marking up to several megabases of DNA [75]. Though it is perhaps not essential for recruitment of repair factors to the DSB site, H2AX phosphorylation assists in the retention of repair proteins within foci, and seems an important identifying step in the initiation of DNA repair [76]. ATM is the principal kinase responsible for phosphorylating

H2AX at DSBs, but both DNA–PKcs and ATR can also target H2AX, with ATR operating at stalled replication forks [77–79]. Immunofluorescence studies with γ–H2AX show that some marked regions (presumed DSBs) appear to remain fairly motionless in the nucleus, while others are able to move and even co–localize [77,80,81]. This implies that not all DSBs, once identified and marked by the cell, remain in one spot but may be recruited to new neighbourhoods or functional "factories".

### 1.5.3 Repairing the DSB

Following the identification of a break, the cell initiates one of two distinct repair pathways: homologous recombination (HR) or non–homologous end joining (NHEJ). In essence, HR makes use of a template from which to copy sequences and replace what may have been lost at the break. NHEJ simply reseals a break by uniting broken ends, requiring little or no homology between them. In yeast (which has a much smaller genome, with very little non–coding DNA) HR is the preferred repair pathway, however mammalian cells use NHEJ most often [82].

It is unclear exactly what determines the choice to follow the NHEJ pathway versus HR. NHEJ is efficient throughout the cell cycle, while HR increases after G1 when its protein components become more highly expressed, and only then do HR proteins co–localize with γ–H2AX foci [81,83–85]. The majority of repair events, even in G2, are normally still carried out by HR–independent mechanisms in mammalian cells [86]. If a rapid end–joining event seems possible, and the NHEJ pathway is not compromised, this is generally carried out first and HR is suppressed [87,88]. The extent to which NHEJ is preferred by human cells is suggested (amongst others) by a previous study in our lab, which found 95% of the detectable repair events of a DSB in a mouse LINE sequence were carried out by NHEJ [89].

The Rad52 protein executes the repair of breaks by HR, binding to the ends to protect them from exonucleases, and recruiting Rad51 to carry out the search and strand invasion of a homologous donor sequence [90]. Rad52 also appears to be involved in the end–to–end

joining reaction that completes HR [91]. In a similar respect, the Ku70/80 heterodimer mediates NHEJ, also by binding and protecting broken ends, and then with DNA–PKcs it recruits the XRCC4/Ligase IV complex to carry out end–joining [82]. Perhaps part of the reason for the preferential use of NHEJ is the ubiquitous presence of Ku, which also binds at the ends of chromosomes (telomeres), and is present at mammalian origins during the initiation of DNA replication [92,93]. A truncated form of Ku86 is even active in end–joining repair of DNA damage in mitochondria, the cellular organelles of respiration which contain their own DNA [94].

Perhaps one way of guiding the choice of repair pathway is through the Mre11/Rad50/Nbs1 (MRN) complex of proteins, which directly binds to the broken ends of DSBs. Tracing the movements of the MRN complex (by labelling Mre11) shows it co–localizes with sites of DNA replication in S–phase, as well as with tracks of DSBs caused by irradiation throughout the cell cycle [81,95]. ATM controls this interaction between MRN and DNA, by activating the Nbs1 component so that it can associate with γ–H2AX [96].

The MRN complex plays a role in both HR and NHEJ repair pathways, though at different stages. Nuclease activities of MRN are active at the initiation steps of HR, resecting DNA ends to permit binding by Rad52 [60]. In contrast, in NHEJ this nucleolytic activity of MRN is only sometimes recruited, and then at a later step of the repair process. Ku binds first to the broken DNA ends, and interacts with DNA–PKcs, and only afterwards MRN may be recruited alongside the XRCC4/Ligase IV complex (which carries out the end–joining reaction) [60,82].

The decision to follow HR or NHEJ could be decided if there were direct competition between Rad52 and Ku for binding to the ends of a DSB. If so, then Rad52's requirement for MRN means it may be the speed with which MRN can arrive at a break, and be activated by ATM, that determines which DSB repair pathway is followed (i.e. whether Rad52 can bind and compete with Ku). The involvement of MRN (and its nuclease activity) in only some NHEJ events could also explain why some NHEJ junctions are

accompanied by no loss of sequence, whereas others may have a few bases or significantly larger degradations from the ends prior to rejoining [97,98].

The features of DSB repair (and NHEJ in particular) which became very interesting to us, and were the basis for this thesis project, were suggestions that fragments of DNA were sometimes inserting at DSB sites. We now look at this evidence in more detail, and how it and other work in our lab directed us to begin this project.

## *1.6 On the Origins of the Thesis*

The impetus for this thesis project began during preparation for my predoctoral exam. It was then that a number of concepts such as genome organization, repetitive DNA, DSB repair, and DNA integration began to converge. The work I had been involved in prior to this, during my first 2 years in the lab, included studies on recombination using plasmid DNA transfected into mammalian (mouse) cells. Our lab had developed a technique to identify regions of the genome which physically and functionally interacted with one another, by following where in the genome a plasmid would integrate after it had undergone HR with a genomic sequence [99]. We extended these studies to show that when a plasmid undergoes HR, it is somehow potentiated to integrate into the genome [100]. These integrations seemed to occur at DSBs, as evidenced by the apparently random distribution of integrations in the absence of induced DSBs, and by the elevated detection of integrations at sites where we could induce a DSB [100,101]. Laboratory protocol refinement had long ago demonstrated that to get a foreign piece of DNA like a plasmid to integrate into a cell's genome, the process was more efficient if the plasmid were cut and made linear, rather than left as a closed circular DNA molecule [102]. This was, in essence, creating broken ends on the plasmid, like a DSB.

At the same time, ongoing work in our lab was using L1 sequences (the LINE elements we discussed above) as substrates in HR studies, sometimes induced by DSBs [89,103–105]. The abundance of LINEs throughout mammalian genomes had previously been explained solely by retrotransposition followed by genetic drift (to account for the truncations and

mutations observed), and our group was demonstrating HR between LINEs could also spread copies [106]. However, the abundance of LINEs in the genome, coupled with the integration studies we were doing, led us to think that perhaps another mechanism could be responsible for some of the spread of LINE fragments: direct insertion into DSBs.

The results of these projects from our lab suggested that the process which repaired DSBs in chromosomes might also operate on the linear ends of foreign (plasmid) DNA and bring the two "breaks" together. We began to ask ourselves why the proteins which bind broken ends of a chromosome, like Ku, couldn't also bind to the free ends of transfected DNA? Or, for that matter, couldn't the broken ends of other kinds of DNA (from mitochondria, viruses, or bacterial DNA) also be recognized by the cell's DSB repair machinery in a similar fashion? The evidence for such events of DSB inserts was just emerging in the literature, from a few sources, which we now turn our attention to.

## *1.7 Evidence of Insertions at DSBs*

A major difficulty in studying DSB repair is being able to identify a break at a specific site, which can then be followed and analyzed. One way of overcoming this was revealed with the discovery of the rare–cutting endonuclease I–Sce I, which has a long recognition sequence (making it unlikely to occur naturally in a mammalian genome) [107,108]. The 18 bp I–Sce I site could be artificially introduced into a plasmid or genome, and upon expression of the enzyme I–Sce I gave experimentalists a way to induce a cut (DSB) in a precise location.

Two studies in tobacco plants found cases of I–Sce I sites being repaired along with "filler" DNA, from plasmids or the tobacco genome, up to 1.3 kb long [109,110]. In yeast, three groups reported insertions into I–Sce I (or the comparable HO endonuclease) induced DSBs [111-113]. These all gave similar results, showing retrotransposon (Ty 1) DNA could insert in fragments, suggesting NHEJ rather than active retrotransposition was responsible. Two of these also identified fragments from the mitochondrial genome inserting at the DSB, some of them compound non–contiguous fragments, totalling

approximately 200 bp per insert [111,112]. The systems used by these groups, however, generally enriched for particular repair events, and suggested inserts were occurring fairly infrequently (less than 1% of all DSB repair events).

Less direct evidence of insertions at induced DSBs existed from mammalian cell studies. In hamster cells one group found 3 insertion events (45 bp – 205 bp long) at an I–Sce I site, and a different group found another 3 insertions (0.9 kb – 2.1 kb long) [114,115]. These involved transfected DNA (the I–Sce I expression plasmid) and sequences presumed to be hamster genomic DNA, based on their similarity to human repetitive DNA [114,115].

(It was only during the course of this thesis work that a system was specifically applied, using I–Sce I, to study insertions at DSBs in mouse cells [116]. Like the previous groups, they found evidence of transfected and repetitive DNA being "captured" during DSB repair, generally 200 bp – 500 bp in size, but with one exceptional insert of nearly 3.4 kb [116,117]. And though some of the mouse genome had been sequenced by this point, it was still far from complete, preventing detailed identification of the non–transfected inserts.)

Hence a few groups had employed I–Sce I to study DSB repair, but the evidence for insertions at the DSB was usually not well characterized. Often the goals of their assays were focused on other categories of repair, and large alterations (like inserts, deletions, or translocations) were dismissed as an entire category of "other" events. Also, these assays employed primer extension/polymerase chain reaction (PCR) to analyze repair events (which requires intact annealing sites), and used polymerases with the capabilities to synthesize only short stretches of DNA (up to about 3.5 kb). This placed limitations on how much DNA could be degraded from the ends or inserted at the DSB, and still be detected. And perhaps most importantly, the evidence from mammalian systems had little reference sequence data from which to identify the source of any of these inserts.

Alltogether this evidence suggested that insertions of DNA fragments occurred at DSBs in a wide variety of organisms – plants, yeast, and rodents. What we set out to do was to see if this phenomenon also occurred in humans.

## *1.8 Objectives of the Thesis*

The goals of my thesis work have been:

1. To identify if DNA can insert at sites of DSBs in human cells
2. To characterize those inserts

To do this, we established two systems, to study the repair of DSBs from either: 1) linear extrachromosomal plasmids; or 2) chromosomal sites at which we could induce a DSB. We chose methods other than PCR to detect repair events, so as to encompass a range of insertion sizes which may have been ignored by previous studies. The availability of the human genome sequence meant we could likely identify the source of any insertions we detected. As such, we chose to conduct both our assays primarily in normal human fibroblasts and equivalent cells from a patient with ataxia telangiectasia (AT), deficient for the ATM kinase.

### 1.8.1 ATM as a Model for Comparison

As we have seen, ATM is a key regulator of a number of repair and signalling processes in the cell. Levels of ATM gene expression are not affected by cell cycle or DNA damage, and the protein itself can activate its targets in response to DSBs at all stages of the cell cycle [118]. Patients with AT suffer from, amongst other things, increased incidences of lymphomas and other cancers, immune deficiencies, and neurological deterioration [119]. Cells which lack functional ATM have an increased sensitivity to ionizing radiation, demonstrate a greater degree of genomic instability and chromosomal breakage, and show increased rates of somatic mutation [90,120,121]. ATM may even affect chromosome positioning in the nucleus, as AT cells show increased associations between chromosome ends (telomeres) and the proteins which make up the proposed nuclear matrix [122].

In order to determine if the process of DNA fragment insertion actually does take place in human cells, we considered there may be difficulty in detecting large–scale events of this nature in a normal human cell line. For this reason, we chose to compare our results from normal cells with those from a patient with AT. Due to its role in numerous aspects of DNA damage sensing, signalling, and repair, ATM–deficient cells have demonstrated (in our lab and others) that they permit more DSB misrepair than normal cells. In addition, the increased chromosome breakage associated with AT cells suggested a possibility of accelerating the process of "shuffling" by DSB insertional repair.

# 2. MATERIALS AND METHODS:

## *2.1 General Methodology*

All DNA modifying and restriction enzymes were purchased from New England Biolabs (NEB – Pickering, ON).

### 2.1.1 Oligonucleotides and Primers

All oligonucleotides (oligos) were designed and/or analyzed using the GeneTool Lite v1.0 freeware (previously available from http://www.doubletwist.com) to ensure proper annealing temperatures, appropriate GC content, and to avoid any hairpin structure formations or multiple annealings to a given template sequence. Oligos were ordered from Alpha DNA (Montreal, QC), were delivered lyophilized, and were resuspended in sterile, deionized distilled water ($ddH_2O$) to a stock concentration of 200 μM. For a complete list of oligos used, refer to Appendix A.

### 2.1.2 Gel Purification and Ethanol Precipitation of DNA

Hereafter "gel purification" refers to the isolation of individual restriction bands from agarose gel following electrophoresis, phenol cleaning, ethanol precipitation, and resuspension as follows: 0.7% – 1.5% agarose gels were stained with ethidium bromide, the desired DNA fragment visualized under UV and excised with a clean scalpel. The gel slice was transferred to a 1.5 mL snap–cap tube (Eppendorf/Brinkmann – Mississauga, ON), mechanically broken up with a sterile pipette tip, and mixed with an equal volume pure phenol. Following 30 sec of vigorous vortexing, the tube was placed at -80C for 20 min, thawed, vortexed again, and centrifuged 4 min at 11000Xg. The supernatant was transferred to a clean tube, and an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1) was added. The mixing, centrifugation, and recuperation was repeated, and an equal volume of chloroform:isoamyl alcohol (24:1) was added and the cycle repeated again to remove all traces of phenol.

This last supernatant was transferred to a new tube for "ethanol precipitation", which hereafter refers to the addition of 1/10 volume of 3M sodium acetate (NaOAC), mixed gently by inversion, and addition of 2 volumes of 95% ethanol to precipitate the DNA. The sample was centrifuged at 11000Xg for 8 min, the supernatant discarded, and the DNA pellet washed twice with 500 µL chilled 70% ethanol to remove salt traces, air dried, and resuspended in 100 µL – 1000 µL $ddH_2O$ + RNAse A (Invitrogen – Burlington, ON).

### 2.1.3 Plasmid DNA Preparation and Analysis

All plasmid screening was done by a modified alkaline lysis miniprep protocol [123]. Bacterial colonies were picked individually with sterile toothpicks, and left to soak 2 minutes in a 1.5 mL snap–cap tube containing 1mL Terrific Broth (TB – Invitrogen) plus appropriate antibiotic. The toothpick was discarded, the tube closed, and grown overnight in a shaker at 37C. The next day, bacteria was pelleted by 12 sec spin at 11000Xg, the supernatant discarded by inverting the tube, and the pellet resuspended in the ~50 µL remaining in the tube by vortexing 5 sec. Lysis of bacterial cells was done by addition of 300 µL TENS (Tris–EDTA buffer plus 0.1N NaOH and 5% SDS), and vortexing for 5 sec, followed by addition of 150 µL 3M NaOAC, and a further 5 sec vortex. Bacterial debris was pelleted by centrifugation for 2 min at 11000Xg, and the supernatant containing the plasmid DNA transferred to a clean 1.5 mL tube by decanting. DNA was precipitated (as above) and resuspended in 100 µL $ddH_2O$ + RNAse A.

For larger quantities of plasmid DNA, pure enough for use in transfections, a "maxiprep" was performed. Individual bacterial colonies were picked and the toothpick left to inoculate 300 mL of TB (plus appropriate antibiotic) overnight in a shaker at 37C. Plasmid DNA was then extracted, cleaned on a column, precipitated and resuspended using the Plasmid Maxi Kit (Qiagen – Mississauga, ON).

### 2.1.4 Bacterial Transformation

All bacterial transformations were carried out as follows: 1 μL – 2 μL of DNA (resuspended in $ddH_2O$) were mixed with 50 μL *Escherichia coli* ElectroMAX DH10B (Invitrogen) per sample, and electroporated using the GeneZapper 450/2500 (IBI, Kodak/Mandel – Saint–Laurent, QC) at 21 μF capacitance, 2500 Volts and 200 Ω resistance, in 0.2 cm gap Electroporation Cuvettes (Bio–Rad Laboratories – Mississauga, ON). Following electroporation, samples were immediately transferred to 5 mL tubes containing 1 mL of SOC medium (Invitrogen), incubated in a shaker at 37C for 1 hr, then plated on Luria Broth (LB)/agar (Invitrogen) and grown overnight at 37C. For selection, LB/agar plates were prepared with 50 μg/mL ampicillin or 50 μg/mL kanamycin (and additionally with 40 μg/mL X–gal [5–bromo–4–chloro–3–indolyl–ß–D–galactoside] (all Invitrogen) for blue/white selection when cloning into the lacZ of pBluescript SK+ constructs).

### 2.1.5 DNA Sequencing

Manual DNA sequencing was performed on plasmid DNA with the Thermo Sequenase Radiolabelled Terminator Cycle Sequencing Kit (Amersham Biosciences – Piscataway, NJ), run at 65W – 75W on 6% Acrylamide gel, and exposed on autoradiography film (Kodak) for 12 hrs – 96 hrs. Automated DNA sequencing was performed on ABI Prism 3100, 3700, or 3730 XLGenetic analyzers (Applied Biosystems – Foster City, CA) through the in–house services of the CHUM (Hopital Notre Dame, Montreal, QC) and McGill University Genomics facilities (Montreal, QC), with the BigDye Terminator Cycle Sequencing Kits v2.0 or 3.0 (Amersham Biosciences).

### 2.1.6 DNA Sequence Analysis

All searches of sequence databases were performed as follows: searches of a given sequence (to identify source and degree of homology) were done by BLAST on the NCBI site (http://ncbi.nih.gov) and/or using the Ensembl Genome Browser (http://www.ensembl.org), and/or by BLAT at the UCSC Genome Bioinformatics site

(http://genome.cse.ucsc.edu). These shall hereafter be referred to as NCBI, Ensembl, and UCSC, respectively. NCBI and Ensembl both link sequence similarities to disease databases which were used in analyses, such as Online Inheritance In Man (http://www.ncbi.nlm.nih.gov/omim) and the Mitelman Database of Chromosomes Aberrations in Cancer (http://cgap.nci.nih.gov/Chromosomes/Mitelman), hereafter referred to as OMIM and Mitelman, respectively. UCSC provides a great deal of detail on a given genomic sequence and its surroundings, including the nature of repetitive sequences, which was also used in our analyses.

## *2.2 Cell Culture*

### 2.2.1 Cell Lines

Six adherent human cell lines were used in our studies: GM00637, GM05849, NTERA–2 cl.D1, M059J and M059K, and HeLa. One simian fibroblast line expressing SV40 T–ag was used in the extrachromosomal DSB assay: COS–7. Initial tests with the chromosomal DSB vectors were done in mouse LTA fibroblasts. Both of the GM cell lines were obtained from the Coriell Institute for Medical Research (Camden, NJ); all other cell lines from American Type Culture Collection (ATCC – Manassas, VA).

GM00637 are SV40–transformed fibroblasts (human female) whose phenotype is normal, or unaffected. GM05849 (a.k.a. AT5BIVA) are SV40–transformed fibroblasts from a male patient with ataxia telangiectasia (AT), a disease characterized by chromosomal breakage and genomic instability. The GM05849 cells are deficient for the DNA damage response and cell cycle checkpoint control kinase ATM (ataxia telangiectasia mutated). The ATM gene is at 11q22.3, and the mutation in this cell line is a homozygous deletion of 2 amino acids at positions 2427 and 2428 [124]. The use of this pair of cell lines provided us with a comparison of human fibroblasts, both transformed with SV40, with normal (GM00637) and inactivated ATM (GM05849) [125].

NTERA–2 cl.D1 (NT2) is a pluripotent testicular embryonal carcinoma cell line. They were chosen for two reasons: 1) they can be induced to differentiate along

neuroectodermal lineages, into postmitotic neurons, upon exposure to retinoic acid [126]; and 2) they have previously been reported to express high levels of LINE–1 transcripts [127]. This enabled us to have a cell line with which we could establish a clone with a single integration site of our vector (for the chromosomal DSB assay), which could later be differentiated if so desired. Epigenetic modifications that accompany differentiation, such as methylation, are believed to influence the level of LINE–1 expression in human cells [128], and we wished to have a means of comparing potential DSB insertions in the context of a constant DSB site with varying LINE–1 activity levels.

M059J and M059K were both derived from the same tumour specimen taken from a male patient with malignant glioblastoma. M059K expresses normal levels of DNA–dependent protein kinase, catalytic subunit (DNA–PKcs), and are proficient for DSB repair. In contrast, M059J is the only human cell line that exhibits a deficiency in a component of the non–homologous end joining (NHEJ) DSB repair machinery, with reduced DNA–PKcs activity and subsequent DSB repair deficiency [129]. The use of this pair of cell lines provided us a comparison of DSB repair outcomes with and without proficient NHEJ.

HeLa are cervical epithelial cells from an adenocarcinoma which were used in our study as a human cell line which lacks expression of SV40 T–ag. COS–7 are African green monkey (*Cercopithecus aethiops*) kidney cells transformed with an origin–defective mutant of SV40, but express T–ag. These two cell lines were used in the extrachromosomal DSB assay to test for functional SV40 origins of replication and T–ag expression from some of the repair products.

LTA (TK-) are mouse fibroblasts which lack endogenous Thymidine Kinase (TK) activity. We used these cells in previous recombination and DNA integration assays in the lab [100], and here to test for functional HSV–TK activity of our vector constructs.

### 2.2.2 Cell Culturing

All cells were incubated in Falcon treated cell culture dishes or flasks (BD Biosciences – Mississauga, ON) at 37C with humidity and 5% $CO_2$. All cells were cultured in Complete Medium (CM), consisting of DMEM/F12 with L–glutamine, 10% Fetal Bovine Serum (FBS), and 100 I.U. penicillin:100 ng/mL streptomycin (all from Wisent – St. Bruno, QC). In addition, for the M059 cell lines the CM was supplemented with 1.0 mM Sodium Pyruvate (Wisent) and 0.05 mM Non–Essential Amino Acids (Mediatech – Herndon, VA). Cells were maintained by passage (split 1/2 – 1/3) with fresh medium every 2 days – 3 days. NT2 and M059 cells were passaged by draining all media from the dish followed by incubation for 2 min – 3 min in phosphate buffered saline (PBS – Wisent) supplemeted with 2.5 mM EDTA + 2% FBS, and taken up by gentle pipetting in CM. All other cells were passaged by rinsing with PBS alone, adding 1/10 – 1/25 volume of Trypsin–EDTA (Wisent) for 2 min – 3 min, and taken up by vigorous pipetting in CM.

### 2.2.3 Cell Storage

For freezing, cells in culture were harvested by rinsing in PBS, trypsinization (as above), and were resuspended to $1x10^6 – 2x10^6$ cells/mL in DMEM/F12 + 20% FBS and 10% dimethyl sulfoxide (DMSO)(Sigma–Aldrich – Oakville, ON). Samples were aliquoted into 1 mL freezer tubes and slowly cooled to -80C before storage in liquid nitrogen. For thawing, tubes were taken directly from liquid nitrogen, thawed at 37C, and the cells plated into dishes containing prewarmed and pH balanced medium (kept 1 hr in the incubator prior to plating).

### 2.2.4 Cell Transfection by Electroporation

All cells were transfected by electroporation, an efficient method of DNA entry for both the extrachromosomal and chromosomal DSB assays. In addition, for the generation of clones in the chromosomal DSB assay, electroporation promotes more events of single plasmid integration, rather than concatamers at a given locus, which simplified analysis of repair events [130].

As with bacterial transformation, we used the GeneZapper 450/2500 apparatus, but here at 950 μF capacitance, 270 Volts and 0 Ω resistance, in 0.4 cm gap cuvettes. Cells were passaged every 24 hrs for 3 days prior to any transfections to be maintained at approximately 70% confluence and ensure active cell cycling (which improves transfection efficiency and cell survival) [131]. Cells were harvested (as above), rinsed twice in PBS (pelleting by centrifugation for 8 min at 1000Xg, draining, and resuspending), counted with a haemocytometer, and resuspended in DMEM/F12 to $6.25x10^6$ cells/mL. DNA samples were added to 800 μL ($5x10^6$ cells) in the cuvette, left 5 min, electroporated, left at room temperature 2 min, then plated into dishes of preconditioned media and incubated.

### 2.2.5 Selection of Clones

Initial tests of pOdin and pZaphod were performed by transfection into LTA cells (along with the original pNTKV–1901 for comparison). As LTA lack endogenous TK activity, both Neomycin phosphotransferase (Neo) and HSV–TK (TK) selectable markers were tested to ensure their functions remained intact. Neo selection was done by adding 400 μg Geneticin (G418 Sulfate – Invitrogen)/mL CM. TK selection was done by adding HAT (100 μM hypoxanthine, 2 μM aminopterin, 15 μM thymidine), such that clones from cells lacking endogenous TK but expressing HSV–TK survive in HAT medium.

Clones from human cells which were first selected for Neo were then treated with Ganciclovir, which kills cells expressing HSV–TK. The endogenous TK activity of the human cells, while not permitting positive selection in HAT, does act differently on Ganciclovir than HSV–TK and thus permits negative selection. Only the viral kinase effectively acts upon Ganciclovir to convert it to a mono–phosphate, and the cell's endogenous kinases then activate it to a di– and tri–phosphate, DHPGTP, which competes with dGTP for incorporation during chain elongation [132].

### 2.2.6 Isolation of Clones

Following transfection, marker gene selection resulted in formation of distinct colonies of clonal cells. Typically 2 weeks – 3 weeks of selection in G418 (for the presence of Neomycin resistance, as in the case of the initial pOdin and pZaphod mother clone generation), or in Ganciclovir (following transfection with a I–Sce I expression plasmid, to induce DSBs and inactive HSV–TK) was needed for the appearance of distinct clone colonies. To isolate these colonies into individual culture wells of a 96–well plate, the dish was rinsed in PBS, drained, and the individual colonies picked by gentle pasteur pipetting with several drops of trypsin to detach the cells from the adherent dish.

### 2.2.7 Cell Lysis PCR

During the transfer of clones to 96–well dishes, some residual cells remain in the pasteur pipette, and could be subjected to a lysis solution and PCR amplification of the I–Sce I sites. The lysis solution consists of 6M Guanidinium HCl, 7.5M Ammonium Acetate, 20% N–Lauroyl Sarcosine Sodium Salt, and 10 mg/mL Proteinase K. The pasteur pipette which was used to transfer the clone is rinsed into a PCR tube filled with 500 μL PBS, centrifuged for 10 min at 8000 rpm, and the pellet resuspended in 10 μL of lysis solution. This is incubated at 62C for 1 hr, before using one half of the solution for PCR. Conditions for PCR with the "CFP/CRP" primer pair (which amplify the region surrounding the pOdin I–Sce I site to give 154 bp fragment in pOdin, and a 126 bp fragment in pZaphod) are: 94C 1 min, 57C 1 min, 72C 1 min 15 s (10 cycles); 94C 1 min, 57C 1 min, 72C 1 min 15 s + 5 s per cycle (20 cycles). These conditions gave very faint signals, and further tests involved doing PCR directly on the pipette PBS wash, which was boiled at 95C for 10 min and/or frozen at -80C to rupture cells.

## *2.3 Genomic DNA Preparation*

Genomic DNA (gDNA) was extracted from cultured cells as follows: $1x10^6$ – $1x10^7$ cells were rinsed, trypsinized, pelleted by centrifugation and rinsed in PBS again, pelleted and then resuspended in 500 μL – 5 mL proteinase lysis solution (50 mM Tris–HCl pH 7.6,

100 mM EDTA, 100 mM NaCl, 1% SDS, 1.6 mg/mL final concentration Proteinase K [Invitrogen] in dd$H_2O$). Samples were incubated at 60C overnight in a gently shaking waterbath. The following day, RNAse A was added to 0.85 mg/mL final concentration and the sample incubated at 37C for 1 hr (minimum). The sample was then phenol cleaned (with two rounds of pure phenol at the outset), precipitated with 0.7 volumes of isopropanol (instead of ethanol) and NaOAC, spooled onto the tip of a sealed sterile glass pipette, air dried, and resuspended in 100 µL – 500 µL 1X TE.

### 2.3.1 Genomic DNA PCR Screening

PCR was used to screen gDNA preps for the presence of pOdin and pZaphod plasmids, to ensure they were not lost from cell populations cultured for long periods of time, or cultured without G418. Attempts were made to maximize this protocol by performing duplex PCR reactions, using 2 sets of primers in the same reaction tube (one for the plasmid, another for a known gene such as Rb as a positive internal control to establish presence of gDNA in the sample). This duplex PCR would ensure that, given a positive signal for the control band (Rb), we could interpret a lack of signal from primer sets flanking the I–Sce I sites to mean cutting had occurred *in vivo* and the repair event resulted in either loss of the annealing sites or an intervening sequence addition too large to detect.

## *2.4 I–Sce I Resistant Band PCR*

Assessment of I–Sce I cutting *in vivo* is traditionally done by amplifying the I–Sce I site by PCR, and digesting the PCR product with I–Sce I to determine if any of it remains undigested. We used primer set "O.DSB" which anneal to generate products of 1113 bp for pOdin and 1088 bp for pZaphod. Digestion of the PCR product with I–Sce I produces fragments of 508 bp and 605 bp from pOdin, and 177 bp and 911 bp from pZaphod.

## *2.5 Recovery of Small Cellular DNA (by Hirt)*

To recover extrachromosomal DNA from cells in culture, we modified the method used by Hirt [133]. To enrich the Hirt extract for molecules which had entered the cell nucleus (required for DNA repair, replication, or transcription), we first performed an isolation of cell nuclei by centrifugation on a sucrose gradient prior to the Hirt extraction.

### 2.5.1 Nuclear Isolation

Adherent cells in culture (approximately $1x10^6$ cells per 100 mm x 25 mm circular culture dish) were harvested by trypsinization, drained and resuspended in 500 µL PBS, then combined on ice with 1.0 mL chilled $ddH_2O$ + 500 µL of Nuclear Pellet Solution (NPS: 1.28 M sucrose, 40 mM Tris–HCl pH 7.5, 20 mM $MgCl_2$, 4% Triton X–100) and left 10 min. The mixture was then spun 15 min at 1300Xg at 4C to pellet nuclei, drained, and resuspended in 250 µL NPS + 750 µL $ddH_2O$. Following another spin of 15 min at 1300Xg at 4C, the nuclear pellet was drained and resuspended in 400 µL PBS, and either used immediately for Hirt extraction or stored at -20C.

### 2.5.2 Hirt Extraction

An equal volume of Hirt Extraction Solution (0.6% SDS, 10 mM Tris–HCl pH 8.0, 1 mM EDTA) was added to harvested cells in PBS, or to isolated nuclei (as above). Following mixture by inverting the tube several times, and incubation at room temperature for 20 min, NaCl was added to 1 M final concentration, mixed gently, and left at 4C overnight. The following day, the sample was centrifuged for 15 min at 11000Xg at 4C to pellet larger DNA fragments (and any cellular debris), and the supernatant (Hirt extract) transferred to a new tube. The sample was then phenol cleaned, ethanol precipitated, and resuspended in 50 µL – 100 µL $ddH_2O$.

In the case of the extrachromosomal DSB assay, where a "time = 0" control was established, the 800 µL of cells + DNA from the cuvette were electroporated, and immediately divided in half. 400 µL was plated and incubated as usual for the experimental sample, while the other 400 µL (the t = 0 control) was immediately added

to 400 µL Hirt Extraction Solution on ice (to limit any enzymatic activity) and subjected to Hirt extraction as above.

## *2.6 Southern Blot and Hybridization Analysis*

Genomic DNA (gDNA) was digested overnight with the appropriate restriction enzyme, or plasmid DNA digested 2 hrs – 3 hrs, before gel electrophoresis and transfer of the DNA onto nylon membranes for hybridization. 10 µg – 15 µg gDNA were loaded and run on 0.7% – 1.0% agarose gel overnight at low voltage (15 V– 20 V) for maximum resolution. The gel was denatured for 30 min in 1.5 M NaCl/0.5 M NaOH at room temperature before being transfered via capillary action onto a filter paper wick blot assembly, consisting of (bottom to top): reservoir of denaturation/transfer solution, solid support surface above reservoir, 3 sheets Whatman 3MM paper (as wick, ends soaking in reservoir), the denatured gel, then a charged nylon membrane (Hybond N+ – Amersham), 3 sheets Whatman 3MM paper, and a stack of absorbant paper all cut to the exact size of the gel, topped by a small weight. The level of solution in the reservoir, the amount of absorbant paper that remained dry, and the lack of short–circuits were all checked to ensure continuous and uninterrupted flow of solution through the gel and into the membrane. Transfer times were overnight for gDNA blots, 1 hr for plasmid DNA blots. The membrane was rinsed briefly in 2X SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate), UV crosslinked, and air dried before being used for hybridization.

### 2.6.1 Dot–Blot

For gross assessment of particular DNA (ex. SV40) copy numbers in cell lines, genomic DNA was transferred to Hybond N membrane with the use of a 96–well dot–blot vacuum manifold (BioRad). Genomic DNA was vortexed, resuspended in serial dilutions in 0.4M NaOH/10 mM EDTA, and boiled at 95C for 10 min (then placed on ice) prior to blotting. The membrane was pre–soaked in $ddH_2O$, placed in the manifold, and with the vacuum active each well was pre–loaded with 400 µL $ddH_2O$. Once drained, 400 µL of each sample were loaded per well, and allowed to drain briefly under vacuum. Each well was

then loaded with 400 μL 0.4M NaOH, drained, the manifold dismantled, the membrane rinsed briefly in 2X SSC, UV crosslinked, and air dried prior to hybridization.

### 2.6.2 Membrane Hybridization

Hybridization was done with a $^{32}$P–labelled dCTP random–primed probe prepared using the NEBlot Kit (NEB). Briefly, plasmid DNA probes were labelled by denaturation (boiling at 95C for 5 min) followed by polymerase synthesis using Klenow + radioactive cytosine and a mix of random primers, incubated at 37C for 1 hr. Unincorporated nucleotides were removed on a G–50 Sephadex spin column (Amersham). The membrane was rehydrated in 2X SSC, rolled between sheets of nylon mesh, and unrolled into glass bottles for incubation (prehybridiation, hybridization, and washes all done in a rotating Hybaid oven at 65C). Prehybridization was done in hybridization solution (4X SSC, 8% formamide, 0.75% SDS, 1.5 mM EDTA, 0.75% skim milk, 0.07g/mL dextran sulfate) plus 1.5 mg/mL denatured unlabelled salmon sperm DNA, for 3 hrs prior to addition of probe to eliminate non–specific binding. The labelled probe was then denatured and added to the prehybridized bottle and incubated at 65C for 18 h. The membrane was successively washed in the hybridization oven with 3 washes each (20 min each wash) of 1X SSC + 0.1% SDS, then 0.3X SSC + 0.1% SDS, sealed in plastic and exposed to X–ray film at -80C for 1 day – 14 days.

## *2.7 Fluorescence In Situ Hybridization (FISH)*

Visualization of the locations of specific DNA sequences in cell nuclei was done by first preparing interphase and metaphase cells on glass microscope slides, then hybridization with fluorescently labelled DNA probe.

### 2.7.1 Cell Slide Preparation

Actively cycling cell cultures (in 100 mm dishes) were arrested in metaphase by the addition of 40 ng/mL Colcemid (Invitrogen) and incubation for 2 hrs at 37C. Cells were harvested both from the culture dish surface (by trypsinization) and from the medium,

pelleted by gentle centrifugation (1000Xg max), and rinsed with PBS. Hypotonic shock (to rupture the cells) was done by resuspension of the cell pellet in 10 mL 0.75% sodium citrate, incubation for 20 min at 37C, then centrifugation. Fixation was done by gently resuspending the pellet in 10 mL fixative (3:1 methanol:glacial acetic acid), incubating 10 min at room temperature, pelleting by centrifugation and draining (leaving 1 mL above pellet). The pellet was gently resuspended by pipetting air through it, and another 10 mL total fixative added and incubated again. This fixation wash was repeated a total of 3 times, leaving the resuspended pellet in 1 mL total. Slides were then prepared by dropping the fixed suspension (3 drops with a pasteur pipette) onto chilled glass microscope slides, and air dried at room temperature.

The fixed slides were washed in 2X SSC (45 min, 37C), then treated with pepsin (20 µL pepsin 10%, 40 mL HCl 0.01 N for 15 min, 37C). At room temperature slides were then washed in PBS (for 5 min), treated with formaldehyde (1 ml formaldehyde 37%, 0.18 g $MgCl_2$, 39 ml PBS for 5 min), and washed again (PBS for 5 min). The slides were then placed in 3 successive washes of 1 min each in 70%, 85%, and 100% ethanol and dried at 37C for 1 hr.

### 2.7.2 Hybridization with Fluorescent Probe

For hybridization, both slides and probe DNA were denatured. Slides were incubated at high temperature in solution to denature chromosomal DNA (70% formamide in 2X SSC, 70C for 2.5 min), rinsed in successive ethanol washes (as above) chilled to -20C, and dried at 37C. Probes were made from plasmid or SV40 genomic DNA using the Bionick Labelling System Kit (Invitrogen). Approximately 400 ng of labelled probe DNA was combined with 10 µg unlabelled human Cot–1 DNA in hybridization buffer (50% formamide, 10% dextran sulfate, 2X SSC, 0.1% SDS, 1X Denhardt's solution, 1 mg salmon sperm DNA) to a total of 20 µl per slide. This probe mixture was denatured by boiling (95C for 10 min), then placed on ice. The probe was added to the slides, sealed with a cover slip, and incubated in a humid container at 37C overnight. The following day, the coverslip was removed and the slides washed (0.3% NP–40 for 2 min at 72C,

then 0.1% NP–40 for 1 min at room temperature). The slides were prevented from drying with new coverslips, were stored in the dark, and visualized with TSA Fluorescence Systems Kits (PerkinElmer Life Sciences – Boston, MA) on a Nikon Eclipse TE300 inverted microscope.

## 2.8 Fluorescence Activated Cell Sorting (FACS)

The efficiency of transfection and subsequent transgene expression was determined by assessing the proportion of cells expressing GFP, and at what intensities. 48 hrs post–transfection cells were harvested and resuspended in PBS + 2% FBS to $7.5x10^6$ cells/mL, kept on ice, and evaluated against a background of mock–transfected cells of the same line using the Coulter EPICS XL flow cytometer (Coulter Electronics – Ville St. Laurent, QC) and analyzed using Win MDI software (http://facs.scripps.edu).

For physical separation of the highest–expressing proportion of cells, FACS was performed 36 hrs post–transfection using the FACStar cytometer (BD Biosciences). Cells were harvested, resuspended in PBS + 1% FBS, and filtered on a 70 μM Falcon cell strainer, and kept at room temperature for 3 hrs – 4 hrs during the sorting. Following sorting, cells were recuperated into 10 mL CM, pelleted by gentle centrifugation, resuspension, and plated into 6–well culture dishes of preconditioned medium.

## 2.9 Plasmid Rescue

Plasmid rescue from mammalian cells is best performed using bacterial strains which are deficient for the recombination protein RecA, such as the DH10B strain used in our study [134]. To recover the integrated vector from the cellular genome for analysis, genomic DNA (gDNA) was harvested from the cells, digested with appropriate restriction enzymes, and ligated under dilute conditions to produce circular DNAs. This mixture was then used to transform bacteria, and colonies screened by miniprep (as above).

Restriction enzymes used for plasmid rescue were chosen which cleaved outside of the sequences necessary for plasmid replication in bacteria (leaving the ampicillin resistance and Col E1 origin intact), as well as the DSB site(s), and cutting within the cellular genome in order to recuperate a junction which enabled us to identify the integration site. Hind III was initially used, but produced relatively larger plasmids when cutting within the genome, hence the majority of later work was done using Bgl II, which cleaves more frequently in the adjacent genomic DNA producing smaller more manageable plasmids.

The detailed plasmid rescue procedure was done as follows: 10 μg of clean gDNA was digested overnight in the appropriate enzyme, phenol cleaned, ethanol precipitated and resuspended in 20 μL $ddH_2O$. The concentration of the digested DNA was re–evaluated on agarose gel to be certain exactly 5 μg was used with 4 μL T4 Ligase in a total volume of 200 μL (ligase buffer + $ddH_2O$), and ligated overnight at 16C. The ligation mix was then phenol cleaned, precipitated using n–butanol instead of ethanol (same procedure but using 10 volumes n–butanol with no NaOAC, followed by two washes with 70% ethanol) and resuspended in 5 μL $ddH_2O$. 2.5 μL was then used to transform bacteria.

# 3. RESULTS

## *3.1 Extrachromosomal DSB Repair Assay*

### 3.1.1 Experimental Design

The most direct method for demonstrating that DNA double–strand break repair can be accompanied by the insertion of other DNA fragments, is to introduce a linear plasmid into cells and then recover any plasmids which have been recircularized with an increase in size. We chose the plasmid pBluescript SK+ as our template, which generates blue bacterial colonies (when grown on medium containing X–gal) by expressing lacZ. We digested the plasmid at one site (linearized) within the lacZ gene prior to transfection into human cells, and recovered low molecular weight DNA from the cells by Hirt extraction after 48 hrs – 72 hrs. Alkaline lysis miniprep and restriction digests were done on bacterial colonies which lost lacZ activity (producing white colonies), indicating imprecise DSB repair during recircularization of the linear plasmid. Those candidates with a possible insertion at the DSB, as evidenced by an increase in plasmid size, were then sequenced (see Figure R1). The results of this work done in two human cell lines, GM00637 (wild type) and GM05849 (ATM–deficient, prone to chromosome breakage and genomic instability) are published in Oncogene [135]. Some of the data not shown in the article is included below, including FISH and Southern blot studies, as well as less detailed results from other cell lines.

**Extrachromosomal DSB Insertional Repair Assay**

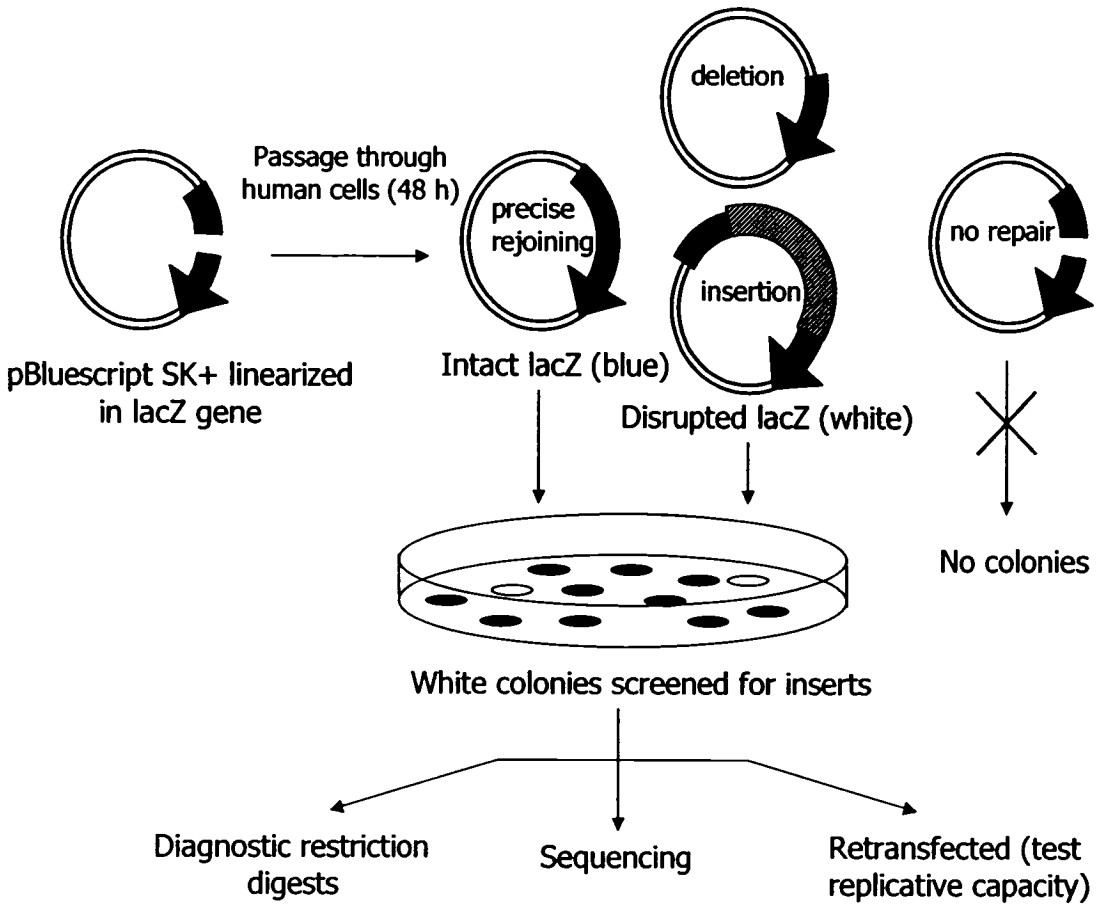



Figure R1. Assay for insertions during extrachromosomal DNA double-strand break (DSB) repair. Human cell lines were transfected by electroporation with linearized plasmid pBluescript SK+. Linearization was primarily done by blunt digestion with Sma I (although a Hind III/Bam HI double digest was also studied) in the lacZ gene. Plasmid was then dephosphorylated with CIP, extracted on agarose gel, phenol cleaned, ethanol precipitated, and resuspended in water. Electroporation was performed and after 48 hrs, nuclei were harvested and low molecular weight DNA isolated by Hirt extraction. DH10B bacteria were transformed with the Hirt extract, and plated on LB agar + ampicillin/X-gal. Plasmids which remain linear transform bacteria poorly, whereas molecules recircularized by the human cells (representing DSB repair events) readily generate bacterial colonies, provided the ampicillin-resistance gene is still intact. Disruption of lacZ results in white colonies, representing imprecise DSB repair (deletion or insertion). All white colonies were screened by alkaline lysis miniprep, restriction enzyme digestion, and run on 1% agarose gel to identify events of insertion greater than ~200 bp. Insertions were then sequenced in part or in full, beginning with primers annealing to pBluescript.

### 3.1.2 Consecutive Refinement of Experimental Procedure

The extrachromosomal DSB assay was performed at three different times (Experiments 1 – 3), slightly modifying the experimental conditions each time to more accurately assess the repair efficiency and fidelity of the cells used. Each of the three experiments did generate events of insertion at the DSB site, and are reported below.

Experiment 1 (Exp 1) was done without isolating only linearized plasmid, prior to transfection into human cells, and as such the frequencies of precise versus imprecise DSB repair could not accurately reflect the cells' capacities (i.e. there was an over–representation of blue colonies with intact lacZ from incompletely digested input DNA).

Experiment 2 (Exp 2) was then performed with gel–purified linear plasmid, and this time by isolating the cell nuclei from the transfected cells after 48 hrs incubation (to enrich for plasmids which entered the nucleus, and thus would have a greater chance of undergoing repair). However, nuclei were not harvested for the "time = 0" control samples, which could complicate direct comparison with the "time = 48 hrs" samples.

To refine and standardize the experimental conditions, Experiment 3 (Exp 3) was done such that nuclei were isolated at both "time = 0" and "time = 48 hrs", which reduced the number of colonies generated but did not appear to have any significant impact on the frequencies (relative to Exp 2). To generate a greater number of events, Exp 3 transfections were performed in triplicate for each cell line, GM00637 and GM05849. Also for Exp 3, immediately following the electroporation we removed 1/3 of the cuvette contents for use as the "time = 0" control, whereas the remaining 2/3 were plated and cultured for 48 hrs prior to extraction of nuclei and Hirt (the "time = 48 hrs" samples).

#### *3.1.2.1 Extrachromosomal DSB Structure: DSB Inserts at Staggered Ends*

The primary goal of this project was to identify if indeed DNA fragments were available for insertion into DSBs, in several human cell lines of interest. Of lesser interest to us was

the influence of the type of DSB on insertion – that is, the nature of the broken ends involved in repair. Extrachromosomal DSB repair efficiency in rodent and human cells has been reported to be the same regardless of the DSB end structure (from various restriction digests), and only the fidelity of repair differs [97,121,136–138]. In an attempt to avoid excluding any particular class of insertion events, we chose to focus our study on a blunt–ended DSB. We used a Sma I digest, which would present no cohesive ends to any potential inserts (prior to whatever processing, such as Mre11–mediated degradation, might occur in the cells prior to ligation) [139]. Blunt DSBs have been shown to repair with one of the highest degrees of fidelity of any break structure generated by restriction enzymes – that is, they recircularize a linear plasmid with fewer deletion or insertion events [136]. This suggests that, if anything, our assay underestimates the frequency at which insertions occur during DSB repair in human cells, since all naturally–occurring DSBs are not blunt ended.

However, to ensure that insertion events could also occur with a different DSB structure, the intial experiment (Exp 1) also analyzed some white colonies generated from a staggered–ended DSB. A double digest with Hind III/Bam HI was used, which excises the intervening 24 bp and generates non–cohesive staggered ends as follows:

5′ ...A^ ^G A T C C... 3′
3′ ...T T C G A^ ^G... 5′

These staggered DSB events proved more difficult to screen, since any repair of the DSB leads to inactivation of lacZ, due to the excised sequences. However, of 288 white colonies studied (harvested by Hirt from GM05849 after 48 hrs), one event ("HBC–2") was characterized as an insertion of 1060 bp of the *E. coli* genome. This was likely the result of contamination of the input plasmid DNA with genomic DNA from the host bacteria used to propagate large quantities of the plasmid. Eliminating this possible contamination was another reason we began gel purifying the linearized plasmid prior to transfection in subsequent experiments (Exp 2 and 3). This event did however demonstrate that the phenomenon of DSB insertions was not limited to only blunt DSBs.

We also compared a Sma I digest alone versus Sma I + CIP. Slightly fewer white colonies were observed for both cell lines in the CIP–treated samples in Exp 1, though these were not all fully characterized to identify the type of DSB misrepair (see Table R–A). To reduce the chances of the linearized plasmid religating upon itself, we chose to apply CIP treatment in all subsequent experiments (Exp 2 and 3).

### *3.1.2.2 DSB Insert Screening Method 1: Restriction Digests Detect Many Events*

When white colonies were generated from a Hirt transformation, several methods for rapid isolation of plasmid DNA were attempted, including direct visualization of the colony DNA on gel without an overnight growth propagation, direct hypotonic lysis and gel loading of complete bacterial samples, or "cracking" lysis of crude plasmid fractions. The alkaline lysis miniprep was finally decided upon as the most efficient means of obtaining plasmid DNA in sufficient quantity and purity to both screen for insert events (by restriction digests, see Figure R2), then characterize them by further digestions and sequencing.

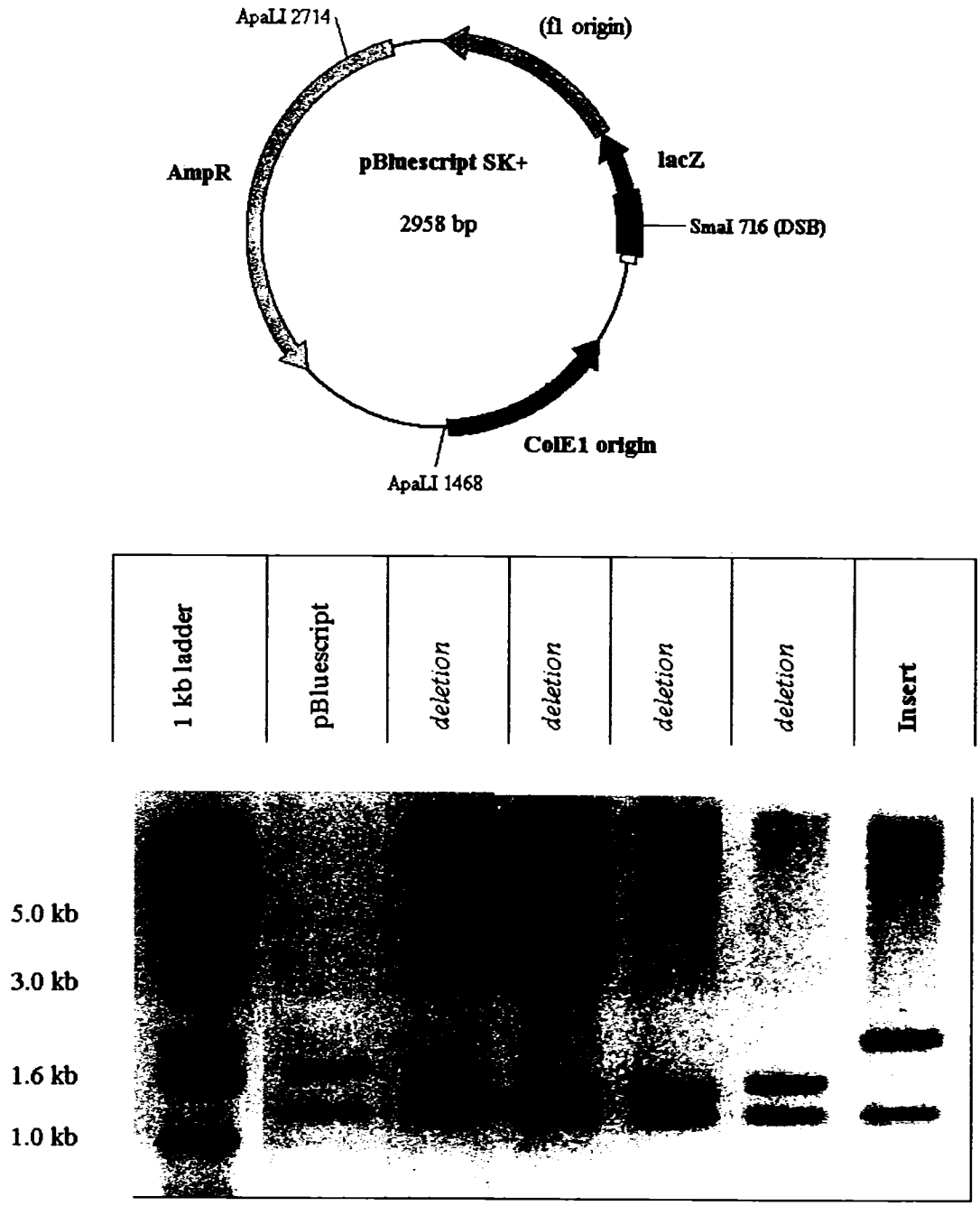



Figure R2. Example of extrachromosomal DSB assay screening method by restriction digest. Top image: DSB template plasmid pBluescript. Bottom image: Plasmid DNA from white colonies (candidates for DSB misrepair) is digested Apa LI and run on agarose gel to identify the 1.2 kb band, and any increases in size of the 1.8 kb (which spans the Sma I DSB site) or additional bands. Image taken from a single ethidium bromide stained gel, with some lanes edited out for clarity (inverted colours). All samples shown are white colonies generated from GM05849 Experiment 3 (Hirt A3).

### *3.1.2.3 DSB Insert Screening Method 2: PCR Misses Some Insert Events*

A method for screening white colonies for insertions at the DSB using PCR was employed in Experiment 1 (see Figure R3). PCR screening did detect some insertions, up to 1 kb in size, which were then sequenced. However, this method failed to generate specific PCR products of the DSB site when larger insertions occurred, and gave only a less-specific PCR product shared with other plasmids.

Verified later by sequencing, this PCR was as successful as the restriction analysis at identifying a deletion of 108 bp ("ASCB-11") and an insertion of 1060 bp of *E. coli* DNA ("HBC-2"). However, larger insertions such as the nearly 4kb insert from chr 4q32 ("ASCB-10"), or 5.2 kb from Xq22 ("ASCB-7") failed to give a specific signal by PCR, despite an elongation time of 4 min and the fact that the primer annealing sites in both cases remained intact. Modifying the reaction conditions (by varying temperatures, polymerases, and cycle times) proved equally unable to detect the larger inserts. As such, all transfections done after Exp 1 were screened for inserts by enzymatic digests alone, without the pBS Insert PCR which would overlook a significant number and class of DSB insert events. This is a direct illustration of the advantage of plamid screening over PCR for detection of large inserts, and could explain how previous groups' studies would have overlooked large insertions as an outcome of DSB repair.

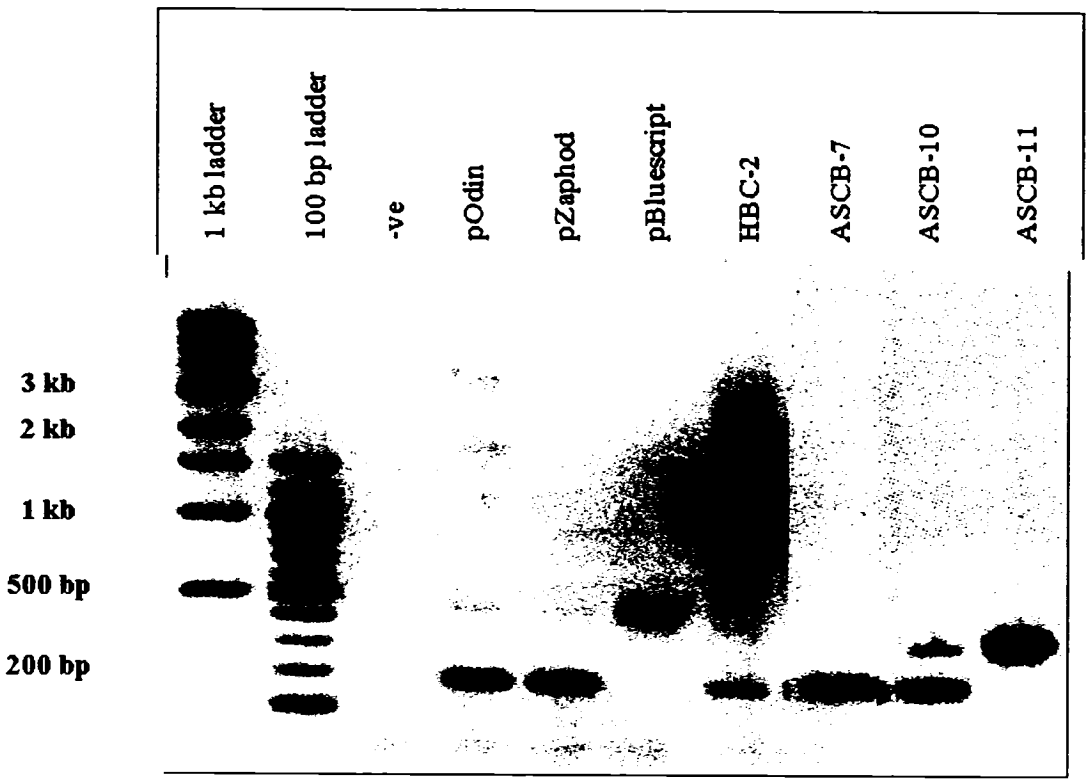



Figure R3. "pBS Insert" PCR to identify inserts at extrachromosomal DSBs in GM05849 cells from Experiment 1. Ethidium bromide stained agarose gel (inverted colours). Plamids (pBluescript) were linearized with Sma I ("S") or Hind III/Bam HI ("HB"), followed by dephosphorylation with CIP ("C"). At 48 hrs post-transfection nuclei were isolated and used for Hirt extraction. White colonies (indicating inactivation of lacZ during DSB repair) generated following bacterial transformation were screened for increases in size by enzymatic digestion, and some by this PCR reaction, prior to sequencing. The primer set "pBS Insert" generates a specific fragment of 458 bp from pBluescript, and a non-specific band at 200 bp in the presence of other plasmid DNA (such as pOdin, pZaphod, or events of insertion in pBluescript too large to amplify by PCR). This image is taken from a single gel photograph, with some intervening lanes edited out for clarity.

### 3.1.3 Repair Frequency and Fidelity (GM00637 and GM05849 Cells)

#### *3.1.3.1 Comparing the 3 Experiments*

The number and frequency of blue and white colonies recovered from all three experiments is summarized in Table R–A. The most significant difference between Exp 1 and the other two was the failure to linearize all input plasmid, which is reflected in the difference of white colonies recovered at 48 hrs. Despite the reduction of this background by gel purifying the linearized plasmid in later steps, there remained some blue colonies in all of the "time = 0" samples tested.

The improved conditions increased the proportion of white colonies (from Exp 1 to Exp 2) at 48 hrs from 1.75% to 3.59% in GM00637, and from 4.07% to 22.67% in GM05849. Frequencies remained relatively similar between Exp 2 and the triplicate Exp 3, although there was some variation within replicate experiments, such as G1 and G2 (2.7% vs 5.6% white colonies). The variation became clear once screening for inserts began, and the source of some of the insert DNA was identified (see below).

Comparing the blue/white frequencies, especially from Exp 3, we note that there is a 5– to 10–fold increase in white colonies from the GM05849 (AT) cells, suggesting a greater degree of DSB misrepair than in the GM00637 (normal) cells. The presence of undigested plasmid in the samples was confirmed with the "1 ng Control", whereby transformation of 1 ng of each DNA sample (prior to being used for transfection) was done directly into bacteria, yielding blue colonies. Less colonies were recovered from the Hind III/Bam HI controls, as the double digestion would have left less plasmid completely uncut than the single Sma I digest. Overall, this suggests that the frequency of DSB misrepair, including DSB inserts, is under–represented in our study.

**A) GM00637 (normal) cells**

| Experiment # (Digest) | | Time | # blue | # white | Total # | *% blue* | *% white* |
|---|---|---|---|---|---|---|---|
| Exp 1 | *Sma I* | **0** | 9000+ | 7 | 9000+ | *99.92%* | *0.08%* |
| | | **48 hrs** | 687 | 13 | 700 | *98.14%* | ***1.86%*** |
| | *Sma I + CIP* | **0** | 11392 | 0 | 11392 | *100.00%* | *0.00%* |
| | | **48 hrs** | 1011 | 18 | 1029 | *98.25%* | ***1.75%*** |
| Exp 2 *(Sma I + CIP)* | | **0** | 2864 | 0 | 2864 | *100.00%* | *0.00%* |
| | | **48 hrs** | 1640 | 61 | 1701 | *96.41%* | ***3.59%*** |
| Exp 3 *(Sma I + CIP)* | Hirt G1 | **0** | 2574 | 1 | 2575 | *99.96%* | *0.04%* |
| | | **48 hrs** | 359 | 10 | 369 | *97.29%* | ***2.71%*** |
| | Hirt G2 | **0** | 559 | 0 | 559 | *100.00%* | *0.00%* |
| | | **48 hrs** | 910 | 54 | 964 | *94.40%* | ***5.60%*** |
| | Hirt G3 | **0** | – | – | – | – | – |
| | | **48 hrs** | 326 | 10 | 336 | *97.02%* | ***2.98%*** |

**B) GM05849 (AT) cells**

| Experiment # (Digest) | | Time | # blue | # white | Total # | *% blue* | *% white* |
|---|---|---|---|---|---|---|---|
| Exp 1 | *Sma I* | **0** | 10000+ | 0 | 10000+ | *100.00%* | *0.00%* |
| | | **48 hrs** | 4776 | 252 | 5028 | *94.99%* | ***5.01%*** |
| | *Sma I + CIP* | **0** | 10000+ | 0 | 10000+ | *100.00%* | *0.00%* |
| | | **48 hrs** | 4973 | 211 | 5184 | *95.93%* | ***4.07%*** |
| Exp 2 *(Sma I + CIP)* | | **0** | 3280 | 0 | 3280 | *100.00%* | *0.00%* |
| | | **48 hrs** | 174 | 51 | 225 | *77.33%* | ***22.67%*** |
| Exp 3 *(Sma I + CIP)* | Hirt A1 | **0** | 1858 | 2 | 1860 | *99.89%* | *0.11%* |
| | | **48 hrs** | 129 | 46 | 175 | *73.71%* | ***26.29%*** |
| | Hirt A2 | **0** | 190 | 0 | 190 | *100.00%* | *0.00%* |
| | | **48 hrs** | 1495 | 593 | 2088 | *71.60%* | ***28.40%*** |
| | Hirt A3 | **0** | 159 | 0 | 159 | *100.00%* | *0.00%* |
| | | **48 hrs** | 827 | 261 | 1088 | *76.01%* | ***23.99%*** |
| **1 ng Control** | | | | | | | |
| Sma I | | | 360 | 0 | 360 | *100.00%* | *0.00%* |
| Sma I + CIP | | | 334 | 0 | 334 | *100.00%* | *0.00%* |
| Hind III + Bam HI | | | 6 | 0 | 6 | *100.00%* | *0.00%* |
| Hind III + Bam HI + CIP | | | 3 | 0 | 3 | *100.00%* | *0.00%* |

Table R–A. Bacterial colonies generated from linearized pBluescript SK+ following passage in A) GM00637 and B) GM05849 cells. Experiment 3 was performed in triplicate, whereby three parallel transfections generated three Hirt extracts for each cell line (G1–G3 from GM00637 and A1–A3 from GM05849). Though used to transform bacteria several times each, the frequencies listed here for the Exp 3 Hirts are calculated from a single transformation per Hirt only. "1 ng Control" was done by transforming bacteria directly with plasmid DNA from Exp 1 (which was not gel purified), confirming the presence of undigested plasmid in the input DNA.

### 3.1.4 Sources of DSB Inserts and Repair Efficiency Determination

A variety of DNA sources contributed to extrachromosomal DSB inserts (for details, see below). These include transfected (plasmid, or contaminant bacterial), or human chromosomal DNA. An unexpected source of DSB inserts was the discovery of plasmids which were repaired using fragments of simian virus 40 (SV40) DNA. While both GM00637 and GM05849 were cell lines established by immortalization with SV40, we had not forseen the possibility this would act as a source of DNA fragment inserts. We also isolated events of SV40 and human chromosomal DNA which shared recurring junctions, suggesting the source of these events was from integrated viral genomes.

We detected one event of SV40 DNA insertion from each cell line in Exp 2: "GM–GB19" from GM00637, which had a total of 5 kb inserted (not all of it characterized); and "ASCB–9" from GM05849, which had 3 kb inserted, consisting of 335 bp of SV40 followed by a region of chromosome 18q21. Perhaps even more surprising was the discovery of the same plasmid from multiple colonies. This proved to be events of DSB insertional repair which permitted the repaired locus (the new plasmid) to replicate in human cells, prior to Hirt extraction (for more detail, see sections below).

As described above, the best characterized set of DSB insert results, and those which can most accurately be compared (due to the finalized experimental conditions), come from Exp 3. By eliminating all white colonies which harbour a plasmid present in more than one copy from our counts, we are left with the number of unique insert events. Once we were able to determine the extent to which DSB insert events had resulted in autonomous replication of plasmids, we used the frequencies from Exp 3 to establish the impact DSB insertional repair has on overall DSB repair capacity in these cells (see Table R–B).

We observed 4% of the DSBs repaired in normal cells (GM00637) suffered a loss of lacZ, whereas 20% of colonies showed misrepair events in AT cells (GM05849). Of these white colonies, or DSB misrepair events, 23% from normal cells involved a DSB insert, a 10–fold difference from the 2.4% in AT cells. Overall, however, the proportion

of DSB inserts per total repair events was roughly the same in both cells: 0.86% in normal cells versus 0.46% in AT cells.

| | GM00637 cells (normal) n = 9 | | GM05849 cells (AT) n = 13 | |
|---|---|---|---|---|
| | **number** | **percentage** | **number** | **percentage** |
| **Total colonies (DSB repaired)** | 4498 | | 13416 | |
| Blue (lacZ intact) | 4318 | 96.00% | 10635 | 79.27% |
| White (lacZ disrupted) | 180 | 4.00% | 2781 | 20.73% |
| | | | | |
| **White colonies (mutants) analyzed** | 143 | | 1499 | |
| Insertions (total) | 42 | 29.37% (of whites); **1.17% (of total)** | 135 | 9.01% (of whites); **1.87% (of total)** |
| Insertions (replicates) | 11 | | 102 | |
| | | | | |
| **Unique mutants (whites)** | 132 | 92.31% (of insertions); **3.69% (of total)** | 1397 | 93.20% (of insertions); **19.32% (of total)** |
| Inserts of SV40 alone | 13 | | 0 | |
| Inserts of SV40 + human DNA | 4 | | 4 | |
| Inserts of human DNA | 0 | | 13 | |
| Inserts of pBluescript | 5 | | 13 | |
| Inserts of unidentified DNA | 9 | | 3 | |
| | | | | |
| **Unique Inserts** | 31 | 23.48% of unique whites; **0.86% (of total)** | 33 | 2.36% of unique whites; **0.46% of total** |

Table R–B. Efficiency, fidelity, and characterization of repair of double–strand breaks in normal (GM00637) and AT (GM05849) human cells from Experiment 3. Each cell line was transfected in triplicate, subjected to Hirt extraction, and each extract used to transform DH10B bacteria at least 3 times (for a total of "n" transformations). Some repair products produced multiple colonies, later identified as autonomously replicating, which were counted only once to give "unique" insertion events. These figures are calculated from the Hirts generated in Experiment 3, whereby Hirt G1, G2, and G3 were each transformed three times for a total of 4498 colonies from GM00637. Hirt A1 was transformed 4 times, whereas A2 and A3 were transformed 5 times each, for a total of 13416 colonies from GM05849.

### 3.1.5 Characterization of Extrachromosomal DSB Inserts

Both GM00637 and GM05849 cell lines were capable of using other pBluescript molecules, or fragments thereof, as inserts at DSBs (though these were difficult to analyze in any detail). While both cell lines enabled human and SV40 DNA inserts, there was a distinct difference in the ways in which this DNA was used. Both cell lines had repair events which combined human and SV40 DNA: in GM00637 it was always chr 12q21.33 in combination with SV40, whereas in GM05849 it was always chr 18q21.31 with SV40 (see Table R–C). However, while many (13) insert events in GM00637 consisted solely of SV40, no events from GM05849 used SV40 DNA alone. Conversely, while many (13) insert events in GM05849 were from human DNA without any SV40, no such events were found in GM00637. When SV40 was used as an insert source in GM05849 it was always accompanied by the same region of 18q21.31. When human DNA was used as an insert in GM00637, it was only ever 12q21.33, and always accompanied by SV40.

#### *3.1.5.1 Miscellaneous DNA Sources*

Some events of insertion, as evidenced by increases in plasmid size by restriction digests, could not be characterized by sequencing, and remained "unidentified". Other insertions produced plasmids with doublet/overlapping sequences (in which the sequencing primer annealed in more than one place), along with restriction patterns or partial sequences showing insertion of pBluescript DNA.

As mentioned in the previous section on Screening Methods, the intial experiments identified an event of *E. coli* DNA being inserted at the DSB, as well as larger inserts from chr 4q32 and Xq22. Event "A3E–101" showed an insertion of 460 bp from a predicted stress–response gene of *Corynebacterium glutamicum* (NCBI identifier "gi:21323710", positions 144179 – 144638, from website at http://ncbi.nih.gov). There were 0 bp lost from either side of the Sma I site, and no overlap at either junction between the plasmid and insert sequences. This is a widely–used bacterium in the commercial production of yogurt.

Upon our request the technicians at Coriell (from where the GM05849 cells were originally obtained) screened their cell stock and verified no *C. glutamicum* infection was present, though this does not exclude the possibility our cell stocks were later contaminated. This event occurred using the gel–purified input plasmid, in contrast with the event from Exp 1 ("HBC–2") where 1060 bp of *E. coli* DNA had inserted. There was no evidence this bacterium, nor any plasmids containing its DNA, had ever been used in ours or adjacent labs. ATCC distributes the bacterium, so it may have been cross–contamination from one of our other ATCC cell lines, or perhaps from the cell culture media. Whether the *E. coli* stock we used was somehow contaminated, or that this fragment was derived from some other source which remains unidentified in the DNA databases, remains unknown. Regardless of the source, "A3E–101" represents an interesting event of DSB repair insertion, and another example of capture of foreign DNA.

All together, the events described above showed that: 1) insertions do occur at extrachromosomal DSB sites in human cells; and 2) those inserts could be derived from (at least) either transfected DNA, or human genomic DNA. What follows is a description of the three major categories of DSB insert events we characterized from the normal and AT cells: human chromosomal fragments (in GM05849 only), simian virus 40 (SV40, in GM00637 only), and human DNA + SV40 (in both cell types).

## Table R–C. Insertions of human and SV40 DNA during DSB repair in normal and AT human cells

**A) DSB repaired plasmids with inserts recovered from normal (GM00637) cells**

| Event | Insert size | 5' loss | 5' overlap | 5' seq (from pBluescript into insert...) | (...from insert into pBluescript) 3' seq | 3' overlap | 3' loss |
|---|---|---|---|---|---|---|---|
| | | | | SV40 (Rh911) and **12q21.33 (NT_019546.15)** | | | |
| G1–9 | 5156 bp | 12 bp | ?: ? | (54 bp unidentified): 2630 – 4984... | ...**19110382 – 19111892:** (15 bp unidentified) | ? | 91 bp |
| G2–6 | 2615 bp | 510 bp | 0 bp | (1 bp "T" inserted): **19111752 – 19111342...** | ...123 – 1, 5180 – 4977 | 1 bp | 81 bp |
| G3–10 | 4851 bp | 680 bp | 2:0:1: 5 bp | (duplicated pBluescript 12 bp): **19110595 – 19110619:** 3877 – 3799: 2570 – 2637... | ...**19111189 – 19111467:** (1 bp "C" inserted) | 0 bp | 67 bp |
| G3C–12 | 3887 bp | 623 bp | 0 bp | (1 bp "T" inserted): 2460 – 3129... | ...512 – 902: **19110242 – 19110507** | 1: 3 bp | 27 bp |
| | | | | | | | |
| | | | | SV40 (strain Rh911) | | | |
| G3C–5 | 2114 bp | 840 bp | 2 bp | 3662 – 4598... | ...5098 – 5180, 1 – 593: (7 bp unidentified) | ? | 59 bp |
| | | | | SV40 (strain SV40–GM00637H Variant 1) | | | |
| GM–GB19 | ~ 5 kb | 0 bp | 0 bp | 188 – 1:5249 – 5017... | ? | ? | ? |
| G2–7 | ~ 4.5 kb | 228 bp | 0 bp | (2 bp "CG" inserted): 291 – 1, 5249 – 5047.... | ..1183 – 786 | ? | ? |
| G1Dp–4 | ~ 3.7 kb | ? | ?: 0 bp | (375 bp unidentified): 292 – 1, 5249 – 4976: 4649 – 4495... | ...5057 – 4972: 4649 – 4465: 2796 – 2440: (22 bp unidentified) | 0: 3: ? bp | 48 bp |
| G2B–11 | ~ 10 kb | 616 bp | ? | (44 bp unidentified): 924 – 295... | ...417 – 1, 5249 – 4972: 4649 – 4641: (12 bp unidentified) | 0 : ? bp | 120 bp |
| G2–2 | 2602 bp | 53 bp | 4:6 bp | (inverted pBluescript pos 715–649): 1867 – 1416... | ( inverted pBluescript pos 750–728) ...5048 – 4972:4649 – 4191: (60 bp unidentified): | 0: ?: 2 bp | ? |
| G1Dp–14 | ~ 6 kb | 572 bp | ?: ? | (12 bp unidentified): 1869 – 1375... | ...4077 – 3466 | 2 bp | 248 bp |
| G1B–6 | ~ 8 kb | 128 bp | 0 bp | (4 bp "AAAG" inserted): 3143 – 3992... | ...1780 – 2565 | 2 bp | 416 bp |
| G2C–2 | ~ 7 kb | 10 bp | 0 bp | 3513 – 3784... | ? | ? | ? |
| G2C–11 | ~ 5 kb | 657 bp | 2 bp | 3816 – 3231... | ...3885 – 3146 | 2 bp | 134 bp |
| G1Dp–12 | ~ 2.2 kb | 393 bp | ?: 0 bp | (8 bp unidentified): 4032 – 4649: 4972 – 5242... | ...1197 – 1840 | 0 bp | 47 bp |
| G3B–3 | 854 bp | 15 bp | 0: 0 bp | 4647–4649: 4972–5249, 1 – 345... | ...3 – 575 | 2 bp | 9 bp |
| G3C–18 | ~ 5.5 kb | ? | ? | ? | ...396 – 1128: (2 bp "TT" inserted) | 0 bp | 117 bp |

## Table R–C (continued). Insertions of human and SV40 DNA during DSB repair in normal and AT human cells

**B) DSB repaired plasmids with inserts recovered from AT (GM05849) cells**

| Event | Insert size | 5' loss | 5' overlap | 5' seq (from pBluescript into insert...) | (...from insert into pBluescript) 3' seq | 3' overlap | 3' loss |
|---|---|---|---|---|---|---|---|
| | | | | SV40 (Rh911) and **18q21.31 (NT_025028.13)** | | | |
| A2–16 | ~ 6 kb | 192 bp | 3: 1: 5 bp | 201 – 1, 5180 – 5065: **2474670 – 2474998**: 537 – 1, 5180 – 5065... | ...2136 – 1271: **2479779 – 2480327** | 3: 1 bp | 33 bp |
| A2B–16 | ~ 8 kb | 94 bp | 0: 0 bp | **2481416 – 2481846: 2474350 – 2474601...** | ...523 – 1, 5180 – 4467: (inverted pBluescript pos 858 – 832) | 7: 7/8 bp | 133 bp |
| A3E–154 | ~ 5 kb | 539 bp | 1: 5 bp | 4331 – 5180, 1 – 537: **2474998 – 2474161**... | **...2474925 – 2474182** | 0 bp | 130 bp |
| ASCB–9 | ~ 3 kb | 5 bp | 2 bp | **2473046 – 2473680...** | ...**2474704 – 2474998**: 537 – 202 | 5: 0 bp | 17 bp |
| | | | | | | | |
| | | | | **Human Chromosomal DNA** | | | |
| A3B–61 | 459 bp | 0 bp | 0 bp | **1q23.3** (NT_004668.15) 10454503 – 10454043 | 10454503 – 10454043 (NT_004668.15) **1q23.3** | 1 bp | 0 bp |
| A2–52 | 499 bp | 0 bp | 0 bp | **17p12** (NT_010718.13) 13648740 – 13648884... | 13648740 – 13649239 (NT_010718.13) **17p12** | 0 bp | 0 bp |
| A2C–177 | 1605 bp | 154 bp | 0 bp | **12q24.32** (NT_009755.16) 4075873 – 4076650... | ...4076730 – 4077478 (NT_009755.16) **12q24.32** | 1 bp | 156 bp |
| ASCB–10 | 3195 bp | 0 bp | 2 bp | **4q32.2** (NT_016606.16) 24791147 – 24791806... | ...24794149 – 24794342 (NT_016606.16) **4q32.2** | 0 bp | 176 bp |
| ASCB–7 | 5146 bp | 0 bp | 0 bp | **Xq22.2** (NT_011651.13) 26108489 – 26108322... | ...26103475 – 26103343 (NT_011651.13) **Xq22.2** | 0 bp | 2 bp |
| AS2.38 | 6081 bp | 0 bp | 0 bp | **3q26.31** (NT_005612.14) 79668202 – 79668692... | ...79673701 – 79674283 (NT_005612.14) **3q26.31** | 0 bp | 0 bp |
| A2C–69 | 6356 bp | 0 bp | 0 bp | **6p24.3** (NT_034880.3) 8784620 – 8784092... | ...8778839 – 8778264 (NT_034880.3) **6p24.3** | 0 bp | 0 bp |
| A3B–94 | 17214 bp | 157 bp | 2 bp | **9q31.1** (NT_008470.16) 4708036 – 4708775... | ...4724972 – 4725249 (NT_008470.16) **9q31.1** | 4 bp | 9 bp |
| A2–62 | 17590 bp | 0 bp | 0 bp | **12q15** (NT_029419.10) 32799044 – 32799214... | ...32816182 – 32816635 (NT_029419.10) **12q15** | 1 bp | 0 bp |
| | | | | | | | |
| A3B–55 | 1373 bp | 53 bp | 4: 0 bp | (inverted pBluescript pos 715–684): **18q21.32** (NT_025028.13) 5826222 – 5825588... | (13 bp unidentified) ...5825679 – 5824862 (NT_025028.13) **18q21.32**: | ? | 148 bp |
| A3E–297 | 132: 6027 bp | 0 bp | 0: 5 bp | (132 bp pBluescript pos 2456–2587): **3p25.2** (NT_022517.16) 12453073 – 12453556... | ...12458476 – 12459100 (NT_022517.16) **3p25.2** | 2 bp | 20 bp |
| A3E–99 | 9742: 171 bp | 466 bp | 3 bp | **3p12.2** (NT_022459.13) 13902222 – 13902758... | (NT_022459.13) **3p12.2: 3p12.3** ...13911408 – 13911964: 11701869 – 11702040 | 0: 0 bp | 209 bp |
| A3C–72 | 16049 bp | 126 bp | ? | (10 bp unidentified): **3p14.2** (NT_022517.16) 59740917 – 59741678... | ...59756330 – 59756957 (NT_022517.16) **3p14.2** | 0 bp | 44 bp |

Table R–C (previous pages). Insertions of human and SV40 DNA during DSB repair in normal (GM00637) and AT (GM05849) human cells. Each plasmid repair event is listed, along with known or approximate ("~") insert size, determined by restriction digest pattern and sequencing. "5' loss" and "3' loss" indicates number of bases degraded from either side of the Sma I linearization site in the template pBluescript SK+, before encountering the insert sequences. "5' overlap" and "3' overlap" indicates homologies shared between the 2 fragments at their junction. Multiple junctions are separated by a colon (":"). For inserts containing both SV40 and human DNA, positions in regular font refer to SV40 genome (Rh911), and positions in bold font refer to NCBI contigs: for GM00637 cells, chromosome 12q21.33 contig NT_019546.15; for GM05849 cells, chromosome 18q21.31 contig NT_025028.13. All other human DNA inserts are listed following the contig number (in brackets), with positions along the contig in regular font.

### *3.1.5.2 Human DNA Inserts*

All characterized events (13 in total) of DSB insertion using human chromosomal DNA alone, without SV40, were observed in GM05849 AT cells (see Table R–C, B). We performed Southern blot analysis on all 13 of these insert events, which confirmed there was no detectable SV40 DNA present in any of the plamids (Figure R4). The inserted fragments ranged in size from 459 bp – 17590 bp. This likely reaches the uppermost limit of our ability to detect insertions, as in our experience any larger plasmids (particularly of 20 kb or more in size) become progressively more difficult for our bacteria to support. This implies that events of DSB insertion using fragments even larger than 18 kb may have occurred, but we were unable to propagate them in our extrachromosomal assay.

Of the human DNA insert events, all were characterized in part (or in full for the smaller inserts) by sequencing, and some by further restriction digests. This ensured they matched the profile of a single contiguous genomic stretch (from BLAST), as provided by the sequences we analyzed at both ends of the insert. Three events were not clearly composed of human DNA alone: 1) two events ("A3B–55" and "A3C–72") were accompanied by short (10 bp – 13 bp) unidentifiable sequences at the junction of pBluescript and the human DNA insert; and 2) a third event ("A3E–297") was accompanied by a 132 bp fragment of pBluescript, also at the junction. This latter case indicates that compound inserts (and from different sources) are also possible at extrachromosomal DSBs, though not as common as individual fragment inserts.

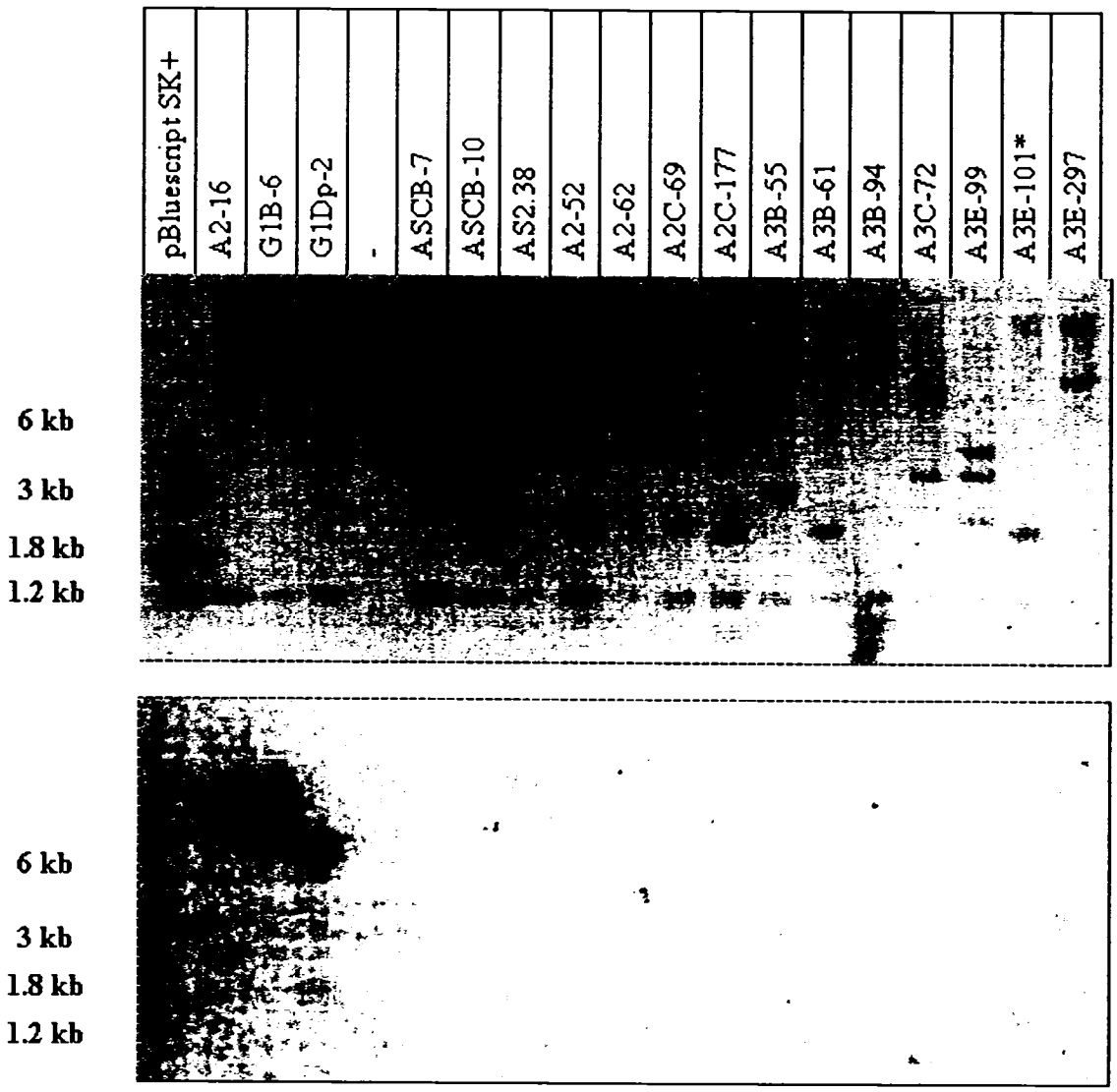



Figure R4. SV40 DNA content of large insertions characterized during extrachromosomal DSB repair. Plasmid samples were digested Apa LI, and run on agarose gel (Top image; ethidium bromide stained gel with inversed colours). The DNA was transferred to a membrane and used for Southern hybridization (Bottom image) with a SV40 probe (cut Hind III, exposure time = 5 min). Plasmid pBluescript was used as a negative control (only partially digested). Only the 3 samples used as positive controls hybridized positively for SV40. All 14 insertions from GM05849 (shown at right) contained no SV40 DNA associated with the insert. *: Event "A3E-101" had no human DNA, but a 460 bp insert from *Corynebacterium glutamicum*.

### *3.1.5.3 Characteristics of Human DNA Inserts*

There did not appear to be any obvious preference for chromosome type or number used as an insert source, aside from the use of chr 3 in four cases ("A3E–297", "A3E–99", "A3C–72", and "AS2.38"), and chr 12 in two cases ("A2C–177" and "A2–62"). Of the chr 3 inserts, 3/4 were derived from the same Hirt (A3), as were the two chr 12 inserts both derived from Hirt A2. It is possible that, in those particular cell sub–populations a given chromosome was more prone to generating fragments suitable for DSB insertion, and so becomes represented more frequently from a given Hirt extraction. However, the small size of the data set makes it difficult to explore this hypothesis.

Event "A3E–99" had nearly 10 kb of chr 3p12.2 inserted, joined to another 171 bp piece of LINE (L1ME2) from 3p12.3, 2.3 Mb away (telomeric to the large fragment). Given the instability inherent in this AT cell line, and the fact that all other genomic inserts were single contiguous fragments, it is possible that the rearrangement of the 2 portions of 3p12 occurred earlier in the genome, prior to our study.

As determined by analyses at the internet sites providing the UCSC Genome Browser (http://genome.cse.ucsc.edu) and MAR–Wiz (http://www.futuresoft.org/MAR–Wiz), the chromosomal DSB insertions from AT cells displayed no readily apparent bias for a number of structural features considered possible determinants of genomic organization and stability (see Table R–C).

#### 3.1.5.3.1 Recombination Rate

The local recombination rates of the genomic regions providing our inserts varied from very low to relatively high, from 0.1 – 2.8 cM/Mb (as per UCSC report). Variations in these rates have been suggested to reflect the likelihood of DSB formation which initiates meiosis [1]. Our results show the frequency of meiotic recombination of a given insert was unrelated to its likelihood of being used as an insert during DSB repair.

#### 3.1.5.3.2 Repetitive Elements

The types and quantities of repetitive DNA elements also failed to show any commonalities shared by all inserts. A given insert had anywhere from 0% – 75% of its DNA made up of repetitive elements (SINE, LINE, LTR, or other DNA repeats). Of the 13 events, 26 junctions exist between chromosomal DNA and pBluescript (if we include the 2 events with the short unidentified sequences at the junctions). Less than a third (8/26) of these junctions occurred within the group of repetitive elements reported to make up at least 45% of the human genome: 4/26 within an LTR; 3/26 within a LINE, 1/26 within a SINE [1]. That nearly 15% of the junctions lie within an LTR (Long Terminal Repeat retrotransposons) while only 12% are within a LINE is remarkable, as the inverse is the case for their genomic content: only about 8% of the human genome is made up of LTRs, while 20% is LINE [1]. One junction (4%) was in a SINE, which make up 13% of the genome. However, again our small sample size makes this difficult to ascertain as statistically significant for certain.

#### 3.1.5.3.3 MARs

The probability that a given sequence, relative to its surroundings, would associate with the nuclear matrix as a Matrix Attachment Region (MAR) was also calculated. The nuclear matrix is a (somewhat controversial) skeletal network of proteins upon which replication, transcription, and DNA repair activites might be carried out. Attachment of certain chromosomal regions to the matrix is widely thought to be responsible for higher–order chromatin folding, by anchoring chromosomes into loops at MARs, possibly creating distinct transcriptional domains [140]. Sites near MARs have been reported to be preferential sites for integration of DNA viruses in transformed cells, and prone to breakage, making this a possible feature responsible for producing fragments for DSB inserts (as well as localizing them to sites of DSB repair) [141].

We analyzed the genomic contig of our insert sequences, as well as the flanking 100 kb (50 kb on either side), for MAR character with the MAR–Wiz program, following a published technique [141]. This program assesses the potential a given region has for being a MAR based on similarity to known concensus sequences of topoisomerase cleavage sites, replication origins, DNA curvature, AT–richness, and other characteristics [142]. Only 1/13 events ("A2–62") bounded a

MAR with high probability, while the others ranged from 1 possible MAR (4/13 cases), 2 possible MARs (4/13 cases), 3 possible MARs (4/13 cases) all the way up to event "A3E–297" with 6 possible MARs (1/13 cases) within the flanking 100 kb. Again, no readily apparent bias or pattern was observed for MAR characteristics of the genomic regions within 50 kb either side of the inserts.

#### 3.1.5.3.4 Genes

In 4/13 cases, human inserts came from known genes: "AS2.38", "A3B–94", "A2–62" and "A3C–72". "AS2.38" inserted 6 kb of 3q26.31, from within a large intron between exons 2 – 3 of neuroligin 1 (NLGN1: Locus ID 22871 from NCBI at http://ncbi.nih.gov). "A3C–72" had 16 kb of 3p14.2 from the last intron of fragile histidine triad (FHIT: Locus ID 2272), in the area of the known fragile site FRA3B (see section 6.1, below). "A3B–94" had 17 kb of 9q31 encompassing exons 13 –17 and 3'–flanking sequences of inversin (INVS: Locus ID 27130). And "A2–62" had 17.5 kb of CCR4–NOT transcription complex, subunit 2 (CNOT2: Locus ID 4848) from within the first (alternate) intron and spanning exon 1.

#### 3.1.5.3.5 AT (or GC) Nucleotide Content

We assessed the AT (adenine and thymine) nucleotide content of the DSB inserts (not to be confused with "AT" for ataxia telangiectasia). The genome–wide content of AT is reported as 59%, with local variations related to gene density, cytogenetic banding, recombination rate, and repetitive element content [1,2]. Our inserts typically show AT content very close to this value (ranging from 50% – 67%), again suggesting no particular preference exists for a given genomic region generating an insert.

#### 3.1.5.3.6 Genomic Surroundings

We also looked at what genes were found in the areas surrounding our DSB inserts, and if any relationship to chromosomal breakage in the area of our inserts had been reported previously.

##### 3.1.5.3.6.1 Genomic Surroundings: Fragile Sites and Breakpoints

A "fragile site" (FS) is defined as a specific region of a chromosome which is prone to breakage when a cell is cultured under specific conditions, such as with inhibitors of DNA synthesis [143].

They can be common FS (widespread, in 50% – 100% of the population) or rare FS (found in less than 2.5% of the population), and are areas of instability frequently associated with genomic rearrangements, including amplification of oncogenes [144,145]. This propensity to break (i.e. to generate DSBs) suggested FS could be likely candidates for regions capable of generating fragments of chromosomal DNA (for use as inserts into other DSBs). However, since most FS are characterized by cytogenetics, and reported only by the chromosome band involved, we have a great deal of difficulty in localizing any known FS specifically to the sites which generated our inserts.

We compared the sites of our DSB insert sources with several surveys of known FS (see Table R–D). We found no particular correlation between any FS and the cytogenetic band, or even adjacent band, from which our DSB inserts were derived. The one exception was the source of DSB insert event "A3C–72", which came from within the FHIT gene at 3p14.2 in the region of the best characterized (and most frequently observed) FS, FRA3B. An extensive survey of tumour cell lines and somatic cell hybrids established a series of unique and recurring deletions, aphidicolin–induceable breakpoints, translocation breakpoints, and viral or plasmid integration sites within FRA3B [146]. Our 16 kb insert from "A3C–72" lies more than 200 kilobases telomeric to the region identified by this survey, although it is within the extreme boundaries others have proposed for FRA3B [147].

Chromosomal breakpoints are the junctions from translocations or gross rearrangements which are implicated in both cancers and inherited diseases [148]. We compared our inserts to a known survey of chromosomal breakpoints, the Gross Rearrangement Breakpoint Database (GRaBD), a searchable database available on the internet (at http://www.uwcm.ac.uk/uwcm/mg/grabd/), which lists 397 sequenced breakpoint junctions taken from published literature [148]. None of these breakpoints are within even several Mb of the source of our inserts.

The Cancer Chromosomes database at NCBI now combines the databases of SKY/FISH and CGH (Comparative Genomic Hybridization, a measure of copy–number alterations) from the Cancer Genome Anatomy Project with the Mitelman Breakpoint database [149]. Again, this

information is traditionally gathered at the broad cytogenetic level, and is fairly imprecise with respect to the exact locations and sequences we obtained from our inserts. This database is searchable but only at the level of resolution of large bands (ex. 18q21) and does not distinguish specific sub–bands (ex. 18q21.31 vs 18q21.32).

Dr. Felix Mitelman, the architect of the database and a foremost authority on human cytogenetics, acknowledged that any correlation between our exact insert source and a chromosomal aberration is difficult (personal communication). He states "there is a huge gap between the pure cytogenetic data in the database, expressed as band involvement in a karyotype, and sequence data, and there is no way to bridge the gap". However, we searched a subset of the database for only those cases of rearrangement (amplification, deletion, or translocation) which were recurrent (i.e. have been reported for more than one patient in the Mitelman Recurrent Aberrations database). This provides a (albeit rough) means of comparison of global genome stability and rearrangement propensity (with clinical manifestation) of the regions generating our inserts, and are listed in Table R–D.

#### 3.1.5.3.6.2 Genomic Surroundings: Gene Amplification

We compared the regions which generated our DSB inserts with surveys of oncogenes known to undergo amplification. The only case of note was event "A2–62", which had an insert from within the CNOT2 gene at 12q15. Several genes in the region of 12q13–15 are amplified in many sarcomas, including MDM2 (GeneID: 4193, which binds and inactivates p53), GLI (GeneID: 2735), and possibly others [150].

#### 3.1.5.3.7 Human DNA Inserts: Summary

We have observed that misrepair of DSBs in AT cells, which are known to exhibit genomic instability, enables remarkably large fragments of genomic DNA to be relocated to new loci. In this case, the recipient locus is an extrachromosomal break. Looking at a number of criteria, we find no particular characteristics common to these inserts, which suggests this phenomenon of DSB inserts is not limited to any specific region of the genome.

## Table R–D. Features of chromosomal regions from which human DNA inserts were derived.

| Event | Size (bp) | Locus | Recomb (cM/Mb) | % AT | MAR score of insert [# MARs: distance] | % content Repeats [junction] | Insert from Gene *Gene Neighbours* | Fragile Site/Breakpoint | Mitelman Recurrent Aberrations: # dif conditions (band) |
|---|---|---|---|---|---|---|---|---|---|
| A3B–61 | 459 | 1q23.3 | 1.8 | 66.59% | <0.15 [3: 28 and 30 kb 5'; 45 kb 3'] | 45% [5' in LTR] | *PBX1 (G)* | 1q: Most frequent donor in Jumping Translocations | **5 (1q23)**<br>– del (1)(q23): mesothelioma<br>– del and t with 19p (very frequent): ALL<br>– add (1)(q23): leukemia/lymphoma (HTLV+) |
| A2–52 | 499 | 17p12 | 2.2 | 51.00% | <0.05 [3: 7, 12 and 18 kb 5'] | 60% [3' in SINE] | – | 17p12: 0.17 % (SG) | **12 (17p12)**<br>– del (17)(p12): fibrosarcoma, TCL, osteosarcoma, AML, plasma cell leukemia, adenocarcinoma<br>– add (17)(p12): AML, osteosarcoma, RB, adenocarcinoma (frequent) |
| A2C–177 | 1605 | 12q24.32 | 2.8 | 59.58% | <0.30 [1: 48 kb 3'] | <5% [5' is in a 48 bp AT–rich repeat; *3' is 10 bp away from LINE L1P3*] | – | 12q24: 1.86% (SG) | **28 (12q24)**<br>– del (12)(q24): Wilms tumor, TCL, AML<br>– add (12)(q24): Hodgkin's, TCL, leukemia/lymphoma (HTLV+), Wilms tumor, ALL, myeloma<br>– t with (Xq11) or (18q21): FL |
| ASCB–10 | 3195 | 4q32.2 | 0.3 | 63.27% | <0.40 [3: 20 and 30 kb 5'; 27 kb 3'] | 20% [5' is in LINE] | – | – | **1 (4q32)**<br>– add (4)(q32): AML |
| ASCB–7 | 5146 | Xq22.2 | 0.4 | 63.00% | <0.10 [1: 49 kb 5'] | 75% [3' in LTR] | *MGC15737(G): is 685 bp 5'of insert* | Xq22: 13.89% (SG)<br>Xq22: 11 (H) | **7 (Xq22)**<br>– del (x)(q22): mesolthelioma, leiomyosarcoma, Wilms tumor<br>– add (X)(q22): squamous cell carcinoma |
| AS2.38 | 6081 | 3q26.31 | 1.1 | 63.02% | <0.50 [2: 38 and 49 kb 5'] | 55% [5' in LTR] | NLGN1 (in intron)<br>*EVI1 (G) = ectopic viral integration* | – | **20 (3q26)**<br>– inv and t (3;3)(q21;q26): frequent in AML, CML, others<br>– del (3)(q26): in Hodgkin's disease |
| A2C–69 | 6356 | 6p24.3 | 1.3 | 62.37% | <0.20 [2: 17 and 21 kb 5'] | 40% | – | *6p25*: 7.31% (SG) | **2 (6p24)**<br>– t(6;9;22)(p24;q34;q11) : CML<br>– add (6)(p24): B–cell neoplasm |
| A3B–94 | 17214 | 9q31.1 | 0.6 | 56.96% | <0.15 [1: 38 kb 3'] | 40% | INVS (spans 5 exons and 3' UTR) | *9q32*: 7.03% (SG) | **1 (9q31)**<br>– del (9)(q12q31) : AML |
| A2–62 | 17590 | 12q15 | 0.8 | 64.85% | 1 MAR within insert (in intron) [2: 33 and 47 kb 3'] | 60% [5' in LINE] | CNOT2 (mostly intron, spans 1 small exon)<br>*MDM1 and MDM2 within 2 MB 5', both amplified* | – | **8 (12q15)**<br>– inv, del, and t with many partners in rare disorders (germ cell tumours, etc.) not shared by other loci in this table |
| A3B–55 | 1373 | 18q21.32 | 2.0 | 62.82% | <0.20 [3: 4 kb 5'; 11 and 30 kb 3'] | <2% | *3 Mb away from other 18q21.32 (+SV40) inserts* | 18q21: 0.71% (SG) | **20 (18q21)**<br>– t (14;18)(q32;q21): very frequent with BCL2 gene in FL, also ALL<br>– del (18)(q21): Hairy cell leukemia, mesothelioma, AML, Wilms<br>– add (18)(q21): AML |

**Table R–D (continued). Features of chromosomal regions from which human DNA inserts were derived.**

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| A3E–297 | *132*: 6027 | *plasmid*: 3p25.2 | 0.6 | *50.38%*: 53.03% | <0.30 [6: 41 kb 5'; 4, 16, 25, 38 and 43 kb 3'] | 30% | *PPARG ends 37 kb 5'; SEN2L begins 7 kb 3'* | – | **9 (3p25)**<br>– del (3)(p21p25): mesothelioma<br>– add (3)(p25) : AML<br>– t(3;9;22)(p25;q34;q11) : CML |
| A3E–99 | 9742: 171 | 3p12.2: 3p12.3 | 0.1: 1.4 | 63.33%: 59.30% | <0.40 [1: 38 kb 5']: <0.50 [3: 1.5, 6 and 22 kb 3'] | 10% : 100% [all LINE (L1ME2)] | *ROBO1 begins 570 kb 5'; GBE1 ends 1.3 Mb 3': ROBO2 begins 310 kb 5'; ROBO1 ends 640 kb 3'* | – | **6 (3p12)**<br>– add (3)(p12): osteosarcoma<br>– del (3)(p12p21): AML |
| A3C–72 | 16049 | 3p14.2 | 1.9 | 61.41% | <0.70 [2: 7 kb and 27 kb 3'] | 60% – 5' in LTR | FHIT (intron 8–9) | FRA3B 3p14: 52.31% (SG) 3p14: 160 (H) | **11 (3p14)**<br>– del (3)(p14): AML, mesothelioma, melanoma, squamous cell carcinoma |

Table R–D. Features of chromosomal regions from which human DNA inserts were derived. Orientations: 5' and 3' refer to features upstream or downstream (respectively) of the insert along the chromosome, irrespective of the orientation of the fragment as it inserted into the pBluescript DSB. "Recomb" is the local recombination rate in cM/Mb, as per UCSC Genome Browser. "% AT" is the proportion of adenine + thymine nucelotides per total nucleotides inserted. "MAR score" (potential) and "nearest MAR" are relative calculations, based on a scan of 50 kb flanking either side of the insert. MAR score is the highest probability attained by any region within the insert of being a MAR, where 1.0 is the highest attainable probability of MAR. Distance of nearest MARs and their orientation (5' or 3' relative to insert source) is listed. Repetitive elements contained within the insert (SINE, LINE, LTR, and other DNA repeats) are listed, with [total content of repeats in bp/ total bp of insert] given as an approximate "% content". If either end of an insert is within a repetitive element, the junction (5' or 3') is listed along with the repeat class. Genes/Fragile Sites/Breakpoints: "G" is a feature reported by [143] as a fragile site, with types of rearrangements and nearby implicated genes; "SG" is mean frequency (%) of aphidicolin–inducible fragile sites in a mixed control population (reported by [144]); "H" refers to the number of times (out of 3023 lymphocytes studied) that a break or rearrangement was reported at a fragile site by [151]. "Mitelman Recurrent" lists the number of different conditions (and the chromosome band, in brackets) which report multiple cases associated with a rearrangement. Only some of the most common conditions associated with a rearrangement are listed. Type of rearrangement: "add" = addition, "del" = deletion, "inv" = inversion, "t" = translocation. Condition abbreviations (only the most common conditions are listed in the table): "CML" = chronic myeloid leukemia, "AML" = acute myeloid lukemia, "ALL" = acute lymphoblastic leukemia, "RB" = retinoblastoma, "TCL" = T–cell lymphoma, ("HTLV+" = cases of adult T–cell lymphoma/leukemia which test positive for human T–cell leukemia virus), "FL" = follicular lymphoma.

### *3.1.5.4 SV40 + Human DNA Inserts*

Both of the cell lines described so far, GM00637 (normal) and GM05849 (AT), were transformed with SV40. Dr. Lorraine Toji at Coriell (personal communication) confirmed that SV40 strain Rh911 (NCBI Locus ID: AF316140) had been used to transform GM05849. Rh911 was the most commonly used transforming strain during that period, and though they could not be totally certain Coriell stated it is most likely the same strain used to transform GM00637. From each cell line, we isolated 4 different events of DSB insertion which combined SV40 and human DNA: for GM00637, 12q21.33 and SV40; for GM05849, 18q21.31 and SV40 (see Table R–C). In both cases, the sequences we isolated from these events most closely matched strain Rh911.

#### 3.1.5.4.1 Chr 12q21 and SV40

In the case of GM00637, the 12q21.33 sequences span 1650 bp, from positions (pos) 19110242 to 19111892 (NCBI contig NT_019546.15). In "G1–9", the sequences are continuous from pos 19110382 – 19111892, whereas in "G3–10" this same stretch is interrupted. Chr 12q sequences (19110595 – 19110619) are followed by two discontiguous fragments of SV40 (Rh911 pos 3877 – 3799, then 2570 – 2637...). "G3C–12" indicates that chr 12q pos 19110242 is preceded by SV40 at pos 902 and before (the insert is: ...512 – 902: 19110242 – 19110507). *NB: three dots "..." indicates the sequence continues, but was not characterized past this point.* This suggests that either: 1) compound inserts occurred independently between SV40 and the same sequences of chr 12q, or more likely; 2) that duplication of some sequences in this area has occurred within the cell population, to generate multiple junctions.

This entire span (from pos 19110242 – 19111892) lies within the first (alternate) intron of Vezatin (NCBI Gene ID: 55591), a gene which encodes a transmembrane protein implicated in cell–cell adherens junctions [152]. Gene FGD6 (ID: 55785) begins 25 kb 5' of the span. The local recombination rate is relatively high, at 1.7 cM/Mb. The span begins 10 bp before a SINE (AluSp), and ends within a long SVA repeat. This highly repetitive nature of the span (approximately 95% repetitive elements) returns an exceptionally large number of BLAT

hits (46, with UCSC Genome Browser) which share over 90% homology with the span sequence over 90% (or more) of its length. The AT nucelotide content of the span is 39.31%, exceptionally low, which correlates with the high density of SINE elements concentrated in GC–rich regions of the genome [1,2].

There are 105 reported cases of chromosome aberrations in this gene's band (12q21.3 by the Mitelman database at CGAP, NCBI). While the span region begins with relatively high MAR potential (>0.50), this sharply tapers off for the second half of the span sequence to a remarkable score of 0.0. More interstingly, there is an exceptional density of high–scoring peaks with MAR potential in the flanking 100 kb, particularly 5' of the span. There are 8 probable MARs within 50 kb either side of the span: at 48 kb, 39 kb, 32 kb, 12 kb, 6 kb, and 2.5 kb 5'; and 7 kb and 14.5 kb 3' of the span.

#### 3.1.5.4.2 Chr 18q21 and SV40

In the case of GM05849, the 18q21.31 sequences span 8.8 kb, from pos 2473046 to 2481846 (NT_025028.13). In 3/4 events, the same junction exists between chr 18q pos 2474998 and SV40 pos 537. In "A3E–154", a contiguous sequence spans pos 2474998 – 2474161, whereas in "A2B–16" this region is truncated and joined to a discontiguous sequence (2481416 – 2481846: 2474350 – 2474601), and in "A2–16" this same stretch is truncated elsewhere and has a novel junction with SV40 (5180 – 5065: 2474670 – 2474998). The likelihood that these are all separate events of compound inserts, producing the same junctions, is very low.

This entire 8800 bp span (from pos 2473046 – 2481846) lies within the end of the WDR7 gene (NCBI Gene ID: 23335), a TGF–beta resistance associated gene. The span begins in the large intron 25 – 26, spans the small 105 bp exon 26, then continues into the final intron 26 – 27 (as per the Ensembl Exon Report at http://www.ensembl.org). According to the NCBI Gene report, WDR7 protein levels are exceptionally high in many transformed human cell lines, which correlates with increased metastatic potential. The local recombination rate of the span we have isolated is relatively low, at 0.8 cM/Mb.

There are 1498 reported cases of chromosomal aberrations involving the band of this gene (18q21.1 – q22 by the Cancer Genome Anatomy Project at NCBI: http://cgap.nci.nih.gov). The AT nucelotide content of this span is 60.40%, on par with the genome average. MAR character potential does not reach any greater than 0.35 across the entire span (relative to the flanking 50 kb on either side). There are 3 potential MARs within 50 kb either side of the span: at 29.5 kb, and 9.5 kb 5' of the span, and 10 kb 3' of the span – which represents slightly more MAR character than the average of our chromosomal inserts, but not as strongly as with the chr 12q span from GM00637.

In a genomic context, WDR7 is located 2 Mb centromeric of MALT1 (Gene ID: 10892, frequently translocated in lymphomas), and 7 Mb from both BCL–2 (596; recurrent translocation with chr 14q32 results in overexpression of an anti–apoptotic chimeric protein in lymphomas, leukemias, etc.) and FVT1 (2531; involved in a variant translocation with 2p11 in follicular lymphomas). POLI, or DNA polymerase iota (11201; an error–prone polymerase involved in translesion synthesis and rescue of stalled replication forks) is 2.5 Mb centromeric of WDR7, right next to MBD2 (8932; a deacetylase). DCC (1630; deleted in colorectal carcinoma and leukemia) is 3 Mb centromeric of WDR7, while the entire region from 18q21.3 to the telomere is frequently deleted in a number of conditions, including colorectal carcinomas and advanced adenomas [153].

### *3.1.5.5 SV40 Inserts*

The most probable explanation for these recurring regions in insert events is that SV40 had integrated into these loci (12q in GM00637, and 18q in GM05849), accompanied by (or resulting in) local instability which duplicated some genomic and viral sequences. To verify this hypothesis, we performed FISH analysis on both cell lines with a SV40 probe. We detected SV40 integrated into the region of 18q21 in GM05849, with some metaphases possibly showing an elongation of this chromome arm (see below). No clear signal, above considerable background hybridization, could be detected in GM00637.

Several groups have established different integration sites of SV40 in cell lines [141,154–160]. It is generally accepted that integration of small tumour viruses, including SV40, into the human genome is 1) random, in that no one particular site is preferred; and 2) required for cellular transformation to take place [141,161]. MARs, fragile sites, and areas of transcriptional activity (presumably due to the open chromatin conformation, and torsional stresses associated with gene expression) are "hotspots" for integration [162]. The exceptionally high MAR scores of the regions surrounding the chr 12 span we isolated support this idea of SV40 integrating preferentialy at MARs.

### *3.1.5.6 Characteristics of SV40 Inserts*

Previous groups have described variant SV40 genomes which have arisen in the GM00637 cell lines [163,164](also see J. Lednicky, direct GenBank submission at NCBI, gi:13517906, 2001). The GM05849 cell line has previously been characterized to display a diversity of integration sites of SV40 DNA, including amplifications, but no episomal SV40 [165](also J. Murnane, personal communication).

All events of SV40 inserts, when joined to human DNA, most closely matched the Rh911 strain of SV40 (see Figure R5). These viral genomes demonstrated an "archetypal" regulatory region, which contains only one copy of a 72 bp sequence which is repeated in other "non–archetypal" strains of SV40 [166]. The 13 different events of SV40 DNA insertion alone, without any detectable chr 12, displayed closest homology to a different SV40 genome than the Rh911 strain. These matched the GM00637H variant 1 (var1) genome (NCBI: AF345344), which has 2 copies of a 72 bp repeat in the enhancer region. The sole exception was event "G3C–5", which has an "archetypal" regulatory region lacking the enhancer duplication as is the case for strain Rh911. However, this event may be a case where not enough of the insert was sequenced to detect chr 12 sequences, if they were present.

Of the other 12/13 SV40–only repair products, 5 ("GM2–2", "G1Dp–4", "G2B–11", "G1Dp–12" and "G2–2") were sufficiently characterized to show they span the 323 bp A

gene deletion of the abundant and actively–replicating episomal viral genome gmSV40 [164]. This deletion spans the large T–antigen (T–ag) splice donor site, resulting in a loss of small t–antigen (t–ag) and a T–ag shortened by 14 amino acids. This gmSV40 has been assessed at up to 13,000 copies per cell, perhaps the result of the replicative advantage bestowed upon it by the truncated T–ag [164]. The DNA sequences we recovered which harbour the 323 bp deletion, and match var1, also exhibit a non–archetypal regulatory region, with duplicated enhancer sequences, known to provide a replicative advantage [167,167]. Our findings thus show that the SV40 sequences associated with chromosome 12 differ from those which are available for insertion from episomal viral DNA.

**SV40 Genome**

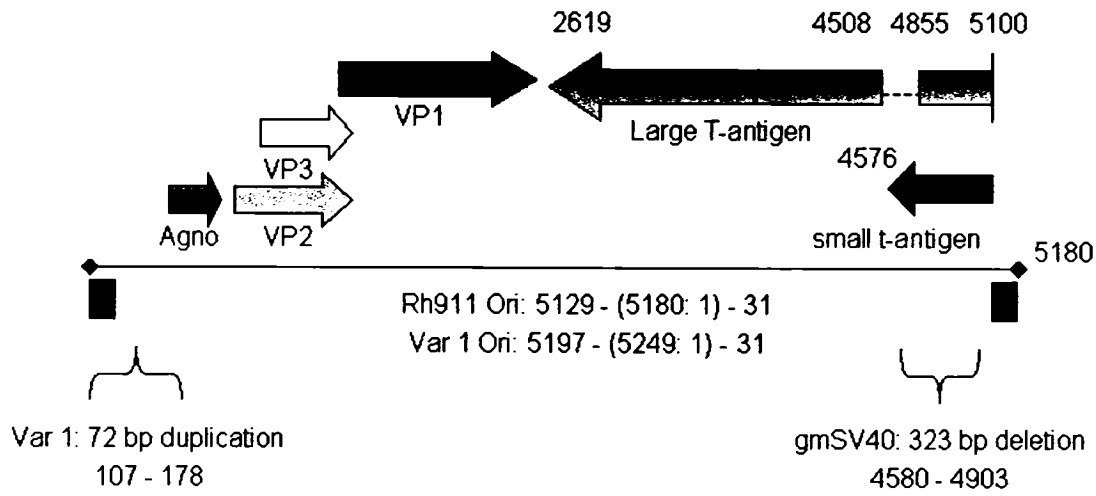



Figure R5. Genome of simian virus (SV40). The SV40 genome is a circular double–stranded DNA molecule, but is represented here as linear, for simplicity. All positions listed are in relation to SV40 strain Rh911 (total 5180 bp). Major feature differences (discussed in text) between Rh911 and GM00637H variant 1 ("Var 1") and gmSV40 strains are indicated. Black squares indicate the span of the viral origin of replication ("Ori"), with positions listed in between the squares. "Agno": agnoprotein; "VP1 – 3": viral coat proteins.

### 3.1.6 Analysis of Inserts for Involvement of Microhomology

Processing of broken ends prior to their ligation is often reported from studies done in mammalian cells, and with *in vitro* cell extract assays. There is evidence for some end–joining events making use of short 1 bp – 6 bp shared sequences on both ends, called "microhomologies", to determine where the processing of the break halts and the ligation takes place. However, there is equally ample evidence for NHEJ events which involve junctions with no shared homologies [97,168]. It is very difficult to evaluate if the presence of shared sequences at a junction is statistically significant, or whether they are present simply due to chance.

The number of bases shared by each junction, or "overlap", from our extrachromosomal DSB inserts is listed in Table R–B. A summary of the number of bases overlap per category of junction is listed below (see Table R–E). A high proportion of all the junctions (42%) share 0 bp overlap, which does not support a role for microhomologies in (at least) a considerable number of DSB insertional repair events. In particular, of the junctions between pBluescript and unique human DNA (inserts which do not include SV40 DNA), 15/24 (63%) have 0 bp overlap. Of the characterized plasmid junctions with human DNA, 13/26 (50%) show 0 bp bases were degraded (lost) from the Sma I DSB prior to the insertion of the human DNA. This strongly implicates a simple end–to–end ligation mechanism in the capture of chromosomal DSB inserts, which involved very little processing of the plasmid ends prior to DSB insertional repair.

| JUNCTION | Overlap (Number of bases) | | | | | | | TOTAL |
|---|---|---|---|---|---|---|---|---|
| | 0 bp | 1 bp | 2 bp | 3 bp | 4 bp | 5 bp | 6 bp | |
| plasmid: plasmid | | | 2 | | 1 | | | 3 |
| plasmid: SV40 | 7 | 2 | 6 | 1 | | | 1 | 17 |
| plasmid: human | 18 | 5 | 4 | 2 | 1 | 1 | | 31 |
| *(plasmid: unique human)* | *(15)* | *(3)* | *(3)* | *(1)* | *(1)* | *(1)* | | *(24)* |
| *(plasmid: Chr 12q21)* | *(1)* | *(1)* | | *(1)* | | | | *(3)* |
| *(plasmid: Chr 18q21)* | *(2)* | *(1)* | *(1)* | | | | | *(4)* |
| SV40: human | | 3 | | 1 | | 1 | | 5 |
| *(SV40: 12q21)* | | *(2)* | | | | | | *(2)* |
| *(SV40: 18q21)* | | *(1)* | | *(1)* | | *(1)* | | *(3)* |
| SV40: SV40 | 1 | | | 1 | | 1 | | 3 |
| **TOTAL** | **26** | **10** | **12** | **5** | **2** | **3** | **1** | **59** |
| **Percentage** | **42%** | **17%** | **20%** | **8%** | **3%** | **5%** | **2%** | **100%** |

Table R–E. Overlapping bases (microhomologies) found at fragment junctions from the extrachromosomal DSB repair assay. The number of junctions with a given number of shared bases (overlap) is listed by category, with subcategories in italics and parentheses. The percentage of junctions showing a given number of bases overlap is also listed.

### 3.1.7 Investigating the Autonomous Replication of Repaired Plasmids

Our initial screening of white colonies revealed that some plasmids from both the GM00637 (normal) cells and GM05849 (AT) cells which harboured inserts generated multiple colonies from a given bacterial transformation. Our method of transforming bacteria with the Hirt and allowing 1 hr recovery/growth prior to plating should permit, at most, 1 round of bacterial replication. In which case, we should see at most 2 bacterial colonies generated from a given plasmid present in the Hirt. The frequent occurrence of many colonies harbouring the same plasmid suggested some of the repair events had undergone replication. That is, following the DSB insertional repair event, but sometime during the 48 hrs prior to harvesting the Hirt supernatant, some of these newly repaired plasmids had replicated in the human cells.

#### *3.1.7.1 Dpn I Screening*

As an initial test of our hypothesis, that some plasmids had replicated in the human cells, I performed a Dpn I digestion on the Hirt extracts from Experiment 3 prior to transforming bacteria. Dpn I cleaves DNA which has been methylated by bacteria, but not DNA which has been synthesized in human cells. As such, plasmids replicated in bacteria (such as our input pBluescript), will be cut by Dpn I and fail to transform bacteria, hence no colonies. Meanwhile, any plasmids replicating in human cells will remain uncut, and able to generate colonies.

Following Dpn I treatment, restriction digests showed that 31/38 colonies from A2 were the same plasmid, which sequencing identified as repair event "A2–16". Of all 6 Hirts from Exp 3 (G1 – G3 and A1 – A3), Hirt A1 was the only one from which no multiple–copy plasmids were ever recovered (despite the fact it was used to transform bacteria five times). Consequently, when pre–treated with Dpn I, A1 was the only Hirt from which no white colonies were produced. Several undigested, background blue colonies were recovered from all Dpn I treated Hirts.

Further analysis of 2 events (isolated from the Dpn I treated Hirt G1), supplied evidence for the involvement of T–ag sequences. (Of the plasmids recovered, 9/15 were the same as event

"G1B–6"). "G1Dp–4" did not span an intact T–ag, and did not replicate in HeLa, although it did contain the intact origin sequences (with enhancer duplication) of var1 SV40. In contrast, "G1B–6" exhibited intact origin and T–ag sequences, and did replicate proficiently in HeLa cells. "G1Dp–4" also shows an inverted piece of SV40 DNA, suggesting the insertion is either from yet another variant episome, is the result of some rearrangement during or following the DSB repair, or is an insertion of multiple fragments of episomal SV40 into the same DSB.

### *3.1.7.2 Testing Repaired Plasmids for Replication*

Sequencing of the inserts revealed that all repair events which produced multiple colonies of this nature contained insertions of SV40 DNA which spanned the viral origin of replication (ori). This led us to investigate whether or not these plasmids were: 1) capable of replicating autonomously when reintroduced into human cells; and 2) whether or not they required their host cell to supply T–ag for replication.

Repaired plasmids which were candidates for autonomous replication were prepared individually by maxiprep and electroporated into 3 cell lines each: HeLa (human fibroblasts, T–ag negative), COS–7 (monkey kidney cells, T–ag positive), and the originating cell line, either GM00637 or GM05849 (both T–ag positive). Following 72 hrs incubation, low molecular weight DNA was harvested by Hirt extract, and fractions subjected to Dpn I digestion, followed by bacterial transformation, to determine if they were derived prokaryotically (from the original maxiprep) or via eukaryotic replication during passage in human cells [169]. All plasmids with an insert spanning the SV40 ori were replication competent in the presence of T–ag, and those spanning both ori and T–ag coding sequences were capable of autonomous replication even in HeLa cells, suggesting the DSB insertion event permitted ectopic expression of T–ag (see Table R–F).

As can be seen in Figure R6, outlining the structure of the SV40/human DNA inserts and their replicative capacity, several different junctions exist between the viral and human DNA, suggesting extensive rearrangement has taken place, especially in the GM05849 cells.

As seen from Figure R6, and Table R–F, of the GM00637 events, all those spanning the SV40 origin of replication were capable of replicating in the presence of T–ag expression, though none expressed their own T–ag. The only exception was event "G1B–6", which was not fully characterized to determine whether or not it contained any human DNA, but did contain a good deal of SV40 sequence (up to 8 kb total insert) and replicated in HeLa cells (which do not express T–ag).

In contrast, all events from the GM05849 cells which were tested replicated even in HeLa cells, indicating they were capable of expressing their own T–ag. Here the only exception was "ASCB–9", which did not appear to span the SV40 ori sequences, and was recovered only once from the Hirt transformation. Though not retransfected into the other cell lines to test its replicative capacity, Dpn I treatment and transformation of bacteria produced only 1 colony at the densest plating, indicating this plasmid was not able to replicate autonomously. (see Table R–F).

**Figure R6. Extrachromosomal DSB insertion events which captured both SV40 and human chromosomal DNA.**

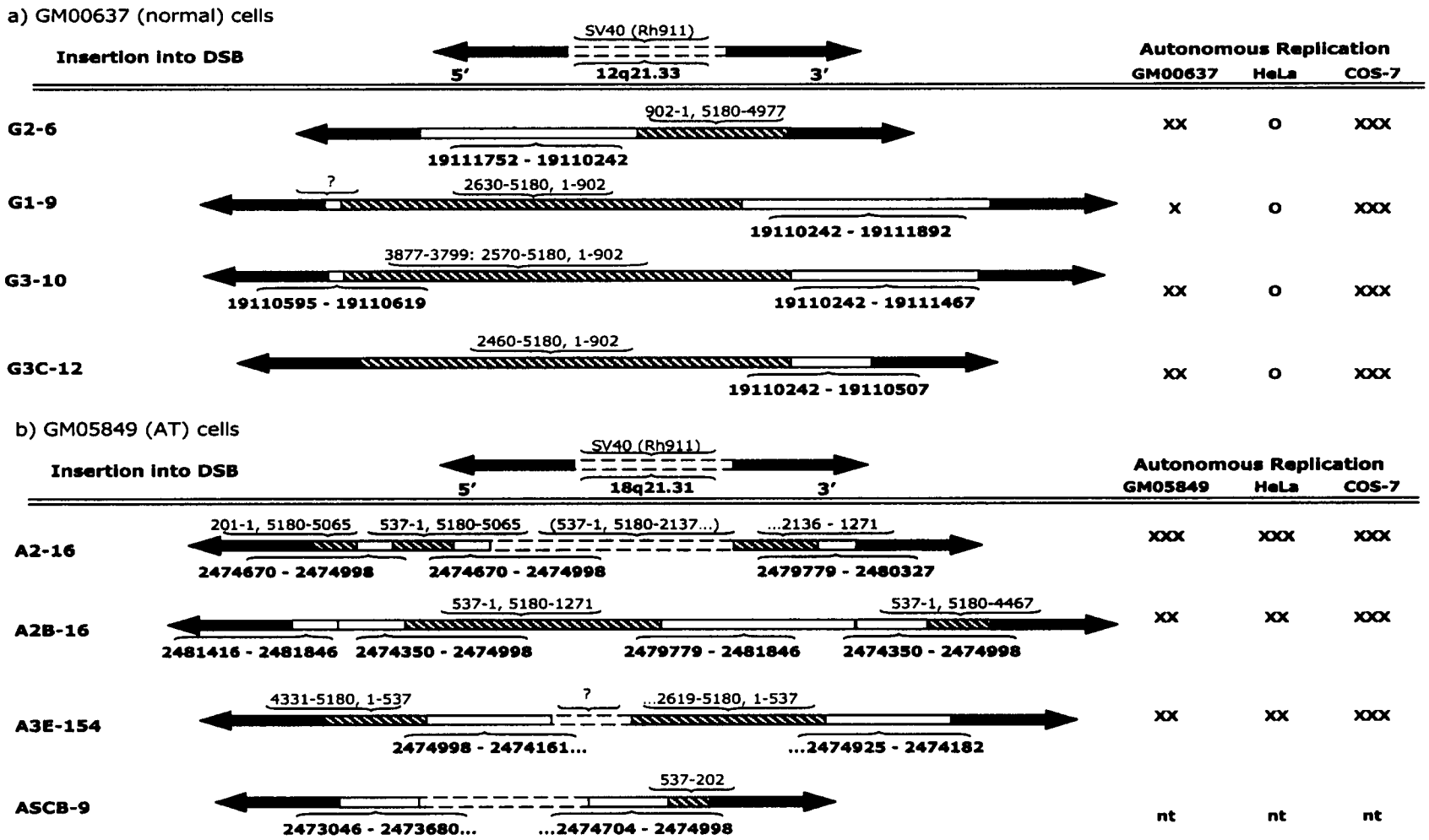


Figure R6 (previous page). Extrachromosomal DSB insertion events which captured both SV40 and human chromosomal DNA. Replication competence conferred to the repair locus was assessed by transfection of the plasmid into the cell line of origin (GM00637 or GM05849), HeLa (T–ag negative), and COS–7 (T–ag positive), and resistance to Dpn I digestion following Hirt extraction. Replicative capacity: X = low, XX = moderate, XXX = high level of replication, 0 = no replication, nt = not tested. Black arrows = pBluescript either side of DSB. Hatched boxes = SV40 DNA, white boxes = human DNA, dashed outline boxes = uncertain sequence, with likely sequence spans in brackets "( )". Nucleotide positions along the SV40 Rh911 genome (NCBI Accession : AF316140) are indicated above the insert, and positions along chromosome 12q21.33 (NCBI Accession :NT_019546.15) in a) or 18q21.31 (NT_025028.13) in b) are indicated below the insert.

| Plasmid | Cell Line | 1 X | | $10^{-1}$ X | | $10^{-2}$ X | | $10^{-3}$ X | | $10^{-4}$ X | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | uncut | cut Dpn I | uncut | cut Dpn I | uncut | cut Dpn I | uncut | cut Dpn I | uncut | cut Dpn I |
| pBluescript SK+ | ATM | c | 1 | c | 0 | 2240 | 0 | 179 | 0 | 125 | 0 |
| | HeLa | c | 39 | c | 4 | 633 | 0 | 168 | 0 | 36 | 0 |
| | COS-7 | c | 34 | c | 2 | 1501 | 1 | 189 | 0 | 36 | 0 |
| | GM | c | 42 | c | 6 | c | 2 | 1033 | 0 | 153 | 0 |
| A2-16 | ATM | c | c | 1403 | 2190 | 220 | 363 | 2 | 2 | 0 | 0 |
| | HeLa | c | c | 529 | 633 | 41 | 60 | 2 | 2 | 0 | 0 |
| | COS-7 | c | c | c | c | c | c | 718 | 853 | 95 | 111 |
| A2B-16 | ATM | np | 1223 | 248 | 165 | 34 | 17 | 2 | 2 | 0 | 0 |
| | HeLa | np | 1090 | 88 | 107 | 13 | 8 | 2 | 2 | 0 | 0 |
| | COS-7 | np | c | c | c | 1517 | 2393 | 248 | 752 | 110 | 84 |
| ASCB-9 | ATM | np | 0 | 546 | 0 | 103 | 0 | 13 | 0 | 1 | 0 |
| | HeLa | np | 1 | 108 | 0 | 17 | 0 | 2 | 0 | 0 | 0 |
| | COS-7 | np | 0 | 475 | 0 | 56 | 0 | 6 | 0 | 0 | 0 |
| A3E-154 | ATM | c | c | 341 | 575 | 46 | 75 | 2 | 6 | 0 | 1 |
| | HeLa | c | c | 265 | 448 | 36 | 54 | 2 | 13 | 0 | 1 |
| | COS-7 | c | c | c | c | c | c | c | c | 669 | 475 |
| G1-9 | GM | 146 | 5 | 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | HeLa | c | 0 | 318 | 0 | 35 | 0 | 4 | 0 | 1 | 0 |
| | COS-7 | c | c | c | c | c | c | 987 | 1312 | 189 | 265 |
| G2-6 | GM | 1662 | 612 | 170 | 70 | 30 | 4 | 3 | 2 | 0 | 0 |
| | HeLa | c | 0 | c | 0 | 218 | 0 | 24 | 0 | 6 | 0 |
| | COS-7 | c | c | c | c | c | c | c | c | 483 | 333 |
| G3-10 | GM | 1616 | 60 | 151 | 4 | 15 | 1 | 1 | 0 | 0 | 0 |
| | HeLa | c | 0 | 279 | 0 | 23 | 0 | 8 | 0 | 1 | 0 |
| | COS-7 | c | c | c | c | c | c | c | c | 341 | 385 |
| G3C-12 | GM | c | 97 | 509 | 13 | 50 | 2 | 4 | 0 | 1 | 0 |
| | HeLa | c | 0 | 542 | 0 | 50 | 0 | 3 | 0 | 0 | 0 |
| | COS-7 | c | c | c | c | c | c | 498 | 832 | 56 | 175 |
| G1B-6 | GM | 364 | 111 | 57 | 18 | 5 | 3 | 0 | 0 | 0 | 0 |
| | HeLa | c | c | 1412 | c | 179 | 312 | 25 | 37 | 2 | 3 |
| | COS-7 | c | c | c | c | c | c | 665 | 665 | 110 | 69 |
| G1Dp-4 | GM | c | 397 | 1072 | 31 | 131 | 4 | 18 | 0 | 1 | 0 |
| | HeLa | c | 0 | 567 | 0 | 49 | 0 | 9 | 0 | 1 | 0 |
| | COS-7 | c | c | c | c | c | c | 471 | 747 | 40 | 78 |

Table R-F. Assessment of autonomous replication capacity of some plasmids with DSB inserts. Candidate plasmids were transfected (4 μg plasmid DNA) into the cell line of origin (GM00637 or GM05849), HeLa (T-ag negative), and COS-7 (T-ag positive) ($2x10^6$ cells/transfection). Hirts extracted at 72 hrs, split, and ½ digested with Dpn I (which cleaves bacterially-methylated DNA, hence any colonies arise from plasmids which replicated post-transfection). The number of colonies per cell represent transformation of 1/45 of each Hirt. Plating was done 1 hr after transformation at 1X (= 1/10 of transformation volume), and four subsequent 1/10 dilutions ($10^{-1}$X – $10^{-4}$X). "c": covered with colonies; "np": not plated (were expected to generate too many colonies to count accurately).

#### *3.1.7.3 Evidence of DSB Insert Instability*

We do find evidence for rearrangements taking place after the DSB insert event. Plasmid "G1B–6" recurred in all transformations of the Hirt, again indicating replication took place. More interestingly, 2 forms of this plasmid arose, in which the 5' junction with pBluescript and subsequent insert sequences were identical, but in approximately half the colonies the size of the insert was doubled. This suggests there is instability inherent in the newly acquired SV40 sequences, which can undergo further rearrangement and amplification of DNA used as a DSB insert.

### 3.1.8 Extrachromosomal DSB Repair in NT2 Cells

The NT2 cell line, derived from a human teratocarcinoma, was originally chosen for study because it contained no SV40 DNA, and was reported to express LINE–1 transcripts at exceptionally high levels [127]. We had anticipated this might permit us to isolate events of insertion of retrotransposon material, although no such events were detected from our limited number of characterized plasmids. The conditions used for Exp 3 (above) were repeated for cell line NT2, such that transfections were performed in triplicate (see Table R–G).

| Transfection # | Time | # blue | # white | Total # | *% blue* | *% white* |
|---|---|---|---|---|---|---|
| **NT2** | | | | | | |
| 1 | 0 | 266 | 3 | 269 | *98.88%* | *1.12%* |
| | 48 hrs | 289 | 177 | 466 | *62.02%* | *37.98%* |
| 2 | 0 | 106 | 0 | 106 | *100.00%* | *0.00%* |
| | 48 hrs | 113 | 56 | 169 | *66.86%* | *33.14%* |
| 3 | 0 | 130 | 1 | 131 | *99.24%* | *0.76%* |
| | 48 hrs | 139 | 62 | 201 | *69.15%* | *30.85%* |
| Total 1 – 3 | **0** | **502** | **4** | **506** | ***99.21%*** | ***0.79%*** |
| | **48 hrs** | **541** | **295** | **836** | ***64.71%*** | ***35.29%*** |

Table R–G. Bacterial colonies generated from pBluescript SK+ (linearized Sma I + CIP) following passage in NT2 cells. Transfections were done in triplicate. DNA samples were gel purified after Sma I digest. Following transfection into cells, 1/3 of cuvette contents was used for "time = 0" control, while the remaining 2/3 were plated in preconditioned media and cultured for 48 hrs. Both fractions had their nuclei purified prior to Hirt extraction. 1/10 of each Hirt was used to transform DH10B bacteria, and plated on agar + ampicillin + X–gal dishes. Counts at "time = 0" are from a single transformation, whereas counts at 48 hrs are the sum of two transformations with each Hirt done at time = 48 hrs.

Nearly 96% of white colonies arose from deletions, the majority (approximately 70%) of those with 500 bp or more lost (as detected by restriction digests). This suggests that perhaps many more DSB repair events took place in the NT2 cells than were detected, which would have resulted in plasmids which lost a functional ampicillin resistance gene (and hence, would not generate a colony). Based upon restriction digests, 12/295 white colonies (4%) harboured insertions at the DSB, representing 1.5% of all DSB repair events detected (see Table R–H).

| NT2 | # colonies | *% of whites* | *% of total* |
|---|---|---|---|
| All colonies at 48 hrs | 836 | – | 100.00% |
| White colonies at 48 hrs | 295 | 100.00% | 35.29% |
| Inserts at DSB | 12 | 4.07% | 1.44% |

Table R–H. Frequency analysis of all colonies obtained from the extrachromosomal DSB assay in NT2 cells.

The degree of deletion, or degradation from the ends of the cut Sma I site, led to difficulty in accurately sequencing the entire DSB region, however even then sequencing of insertion events only revealed mixed fragments of pBluescript. For example, event "N1–67" had an insert of 25 bp of pBluescript, accompanied by a 34 bp loss from the 5' end of the cut Sma I and 443 bp lost from the 3' end. As there was no evidence of other kinds of insertion events, this suggests a relative lack of genomic or other DNA fragments made available to the DSB repair machinery in the NT2 cells, in contrast to the GM00637 and GM05849 cell lines. As they are neither transformed with SV40, nor subject to extensive genomic instability (like AT cells), NT2 demonstrate the phenomenon of DSB inserts is still a viable pathway to repair, using whatever DNA sources may be available.

### 3.1.9 Extrachromosomal DSB Repair in MO59 Cells

The MO59 cells are two lines derived from the same glioma, one which is relatively normal (MO59K) compared to the other (MO59J) which is deficient in the DNA–PK catalytic subunit of the non–homologous end joining pathway. Though performed only once, the results of the extrachromosomal DSB assay in the MO59J and K cell lines are presented here (see Table R–I). This was undertaken as an initial screening to see if a cell line other than the NT2 could provide evidence of insertions on the scale of the GM00637 or GM05849 cells.

| Cell line | Time | # blue | # white | Total # | *% blue* | *% white* |
|---|---|---|---|---|---|---|
| MO59J (DNA–PKcs$^-$) | **0** | 123 | 1 | 124 | *99.19%* | *0.81%* |
| | **48 hrs** | 103 | 3 | 106 | *97.17%* | *2.83%* |
| MO59K (wt) | **0** | 149 | 0 | 150 | *100.00%* | *0.00%* |
| | **48 hrs** | 64 | 6 | 70 | *91.43%* | *8.57%* |

Table R–I. Bacterial colonies generated from pBluescript SK+ (linearized Sma I + CIP) following passage in MO59 cells. Transfections were done only once (4 μg DNA, $2x10^6$ cells). DNA samples were gel purified after Sma I digest. Following transfection into cells, 1/3 of cuvette contents was used for "time = 0" control, while the remaining 2/3 were plated in preconditioned media and cultured for 48 hrs. Both fractions had their nuclei purified prior to Hirt extraction. 1/10 of each Hirt was used to transform DH10B bacteria, and plated on agar + ampicillin + X–gal dishes. Counts at "time = 0" are from a single transformation, whereas counts at 48 hrs are the sum of two transformations with each Hirt done at time = 48 hrs.

Of the 6/70 white colonies recovered from MO59K, and the 3/106 from MO59J at 48 hrs, all were deletions (as evaluated by restriction digests). It is difficult to draw any real conclusions from such a small data set, and as such it is not necessarily significant that the MO59J cells show a greater DSB repair fidelity than the MO59K cells. This could simply reflect a lack of DSB repair overall from MO59J, which would highlight any background (uncut) input plasmid and result in fewer white (and blue) repaired colonies. However, the relative number of total colonies isolated from both cell lines is the same, arguing against a lack of repair from MO59J. It would appear that overall there is a greater DSB repair fidelity from both of these cell lines (3% misrepaired whites from MO59J, 9% for MO59K) than from NT2 (35.29% white colonies) or GM05849 (19.32% unique white), and is more similar to the GM00637 DSB misrepair frequency (3.69% unique white).

## 3.1.10 SV40 Content of Cells and Extrachromosomal DSB Inserts

### *3.1.10.1 Dot-blot Evaluation of SV40 and Chr 18 in GM and NT2 Cells*

Dot-blot Southern hybridization of genomic DNA was used to perform a general assessment of the DNA content of the GM00637, GM05849, and NT2 cell lines (see Figure R7). The SV40 probe was made from the complete SV40 genome (digested Hind III). As a ready source of material from chromosome 18, some of the repaired plasmids from the GM05849 assay were analyzed for a unique sequence which would give a specific probe to assess any gross amplification of chromosomal DNA. The repetitive nature of the DNA from chromosome 12 prevented development of a specific probe.

A number of repair events captured the same region of chromosome 18q21.31, which shared junctions with regions of SV40 (see Table R–C). One of the repeating units of chr 18q21.31 and SV40 was isolated and used for making a probe to assess any amplification of this specific region of 18q21. Repair event plasmid "A2–16" was digested Bgl I to give a 977 bp fragment consisting of: SV40 (Rh 911) 5175 – 5065: chr 18 2474670 – 2474998: SV40 (Rh911) 537 – 1: 5180 – 5175. This sequence spans the SV40 origin of replication, as well as the agnoprotein coding sequences (pos 263 – 451). This chimeric probe proved to hybridize effectively to SV40 sequences, but did not generate any chr 18q21 signal.

Repair event "A3B–55", which contained no SV40 DNA, had inserted 1373 bp of chromosome 18q21.32 from a region 3 Mb distant from the recurring 18q21.31 sequences. A Pst I digestion of"A3B–55" gives a 1088 bp fragment which contained no repetitive elements, and which indicated no homology to other regions in the genome. As such, this fragment was chosen as a specific 18q21 probe (though distant from the SV40–associated site).

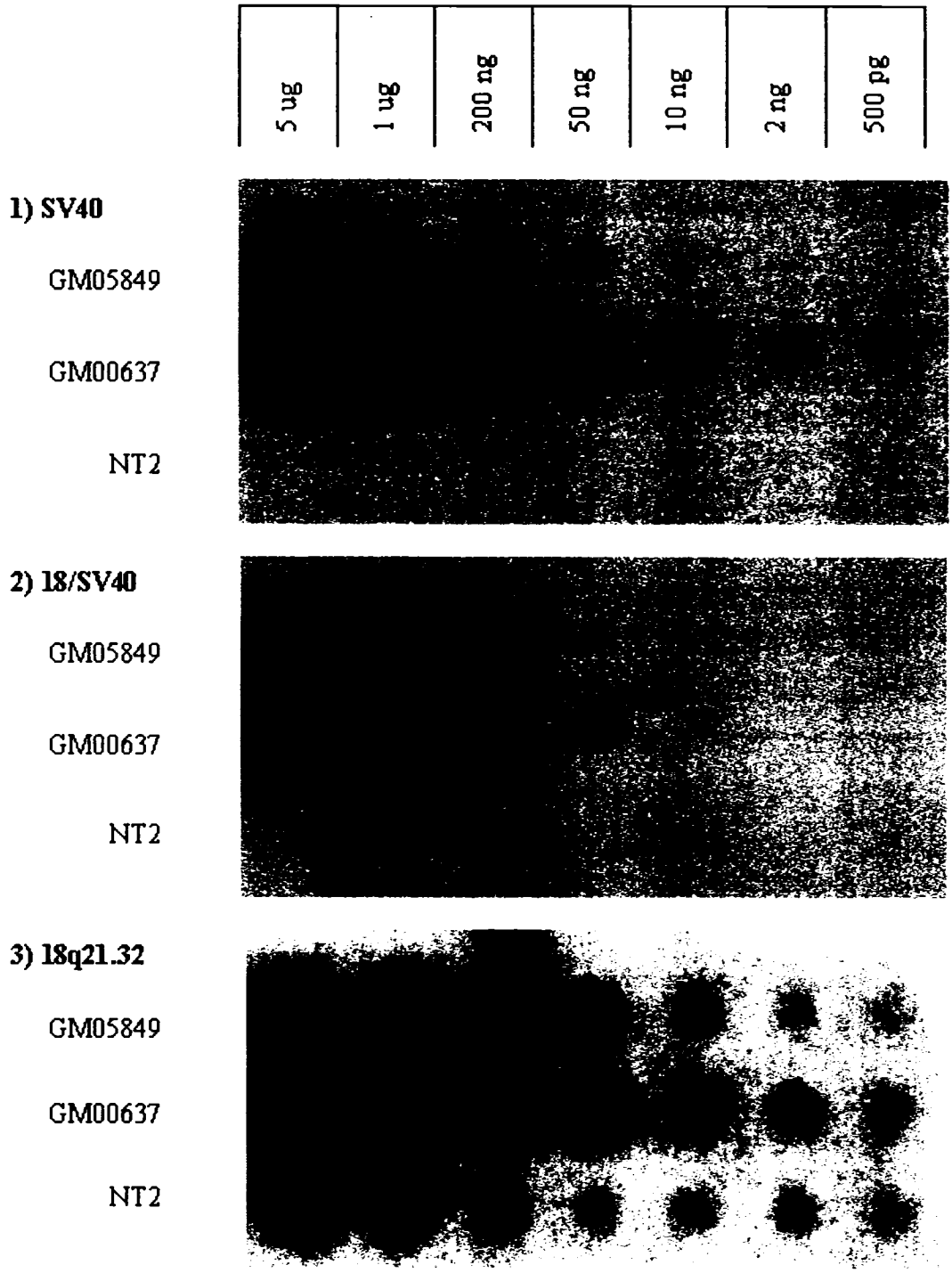



Figure R7. Southern dot-blot assessment of relative DNA content in different cell lines. 1) SV40, 2) chimeric 18q21/SV40, and 3) chr 18q21.32 probes. Equal amounts of genomic DNA were blotted in 5X serial dilutions with a vacuum manifold, and hybridized with the entire SV40 genome (digested Hind III, exposure time = 10 min), the chimeric 977 bp chr 18/SV40 probe (exposure time = 1 hr), or the 1.1 kb chr 18q21.32-specific probe (exposure time = 10 days).

The intensity of the signal from the pure SV40 probe versus the chimeric chr 18/SV40 probe is stronger in the GM00637 than in the GM05849, despite the fact the chimeric probe was derived from GM05849. This implies the SV40 content of GM00637 is greater than that of GM05849. The fact that only the chr 18 probe gives a signal for NT2 proves no SV40 is present in NT2, and that the chr 18q21.32 probe is specific.

Such a strong signal occurs from the SV40 probe, at as short a time as 10 minutes exposure, that the copy number of SV40 must be extremely high. This is a strong contrast to the 12 days required to obtain a signal from the chr 18q21.32 probe.

### 3.1.11 Southern Blot Analysis of GM Cells With SV40 and Chr 18

Both the GM00637 and GM05849 cell lines were subcloned (via transfection and selection for a transgene encoding Neomycin resistance). Genomic DNA from the mother cell line and subclones was digested Hind III, and blotted onto membranes. Successive rounds of hybridization and stripping were done on the same membrane with (in chronological order): 1) entire SV40 genome (digested Hind III); 2) chimeric chr 18/SV40 probe; and 3) chr 18q21.32 specific probe. The blots from the two cell lines were always stripped and reprobed with the same conditions and batch of probe, in parallel.

The SV40 probe gave a signal for GM00637 and only some of its subclones (see Figure S1), while GM05849 and all of its subclones produced a signal (see Figure S2). However, as with the dot–blot analyses, the signals were much stronger for GM00637 and its sublcones, indicating more SV40 in GM00637 than GM05849. The comparison of the entire SV40 genome versus the chimeric 18q21/SV40 probes, at equivalent exposure times, also confirms the dot–blot data, showing greater band intensity for the pure SV40 probe. For example, the pure SV40 blot at 5 days exposure (not shown), when compared to the chimeric 18q21/SV40 blot at 5 days, gives so much signal that none of the small bands can be distinguished. Also, the chimeric probe shows no hybridization to any human chromosomal DNA as demonstrated by the lack of additional signal in Figures S1(2) or S2(2), even after longer

exposures than 5 days, and as such confirms the results of the dot–blot analyses which showed a lack of hybridization signal for the NT2 cells with this probe.

Interestingly, while the GM05849 blot gave a signal for SV40 in every subclone, but with different intensities, only some of the GM00637 subclones gave any signal at all. It would appear that SV40 can be lost (at least to levels beyond our ability to detect) from subclones of GM00637, but those cells which do contain SV40 maintain it at relatively high levels overall. The presence of episomal variant SV40 genomes in GM00637, and the ability to eliminate them through subcloning, has been reported previously [164]. In contrast, GM05849 cells have previously been reported to contain no episomal SV40 [165].

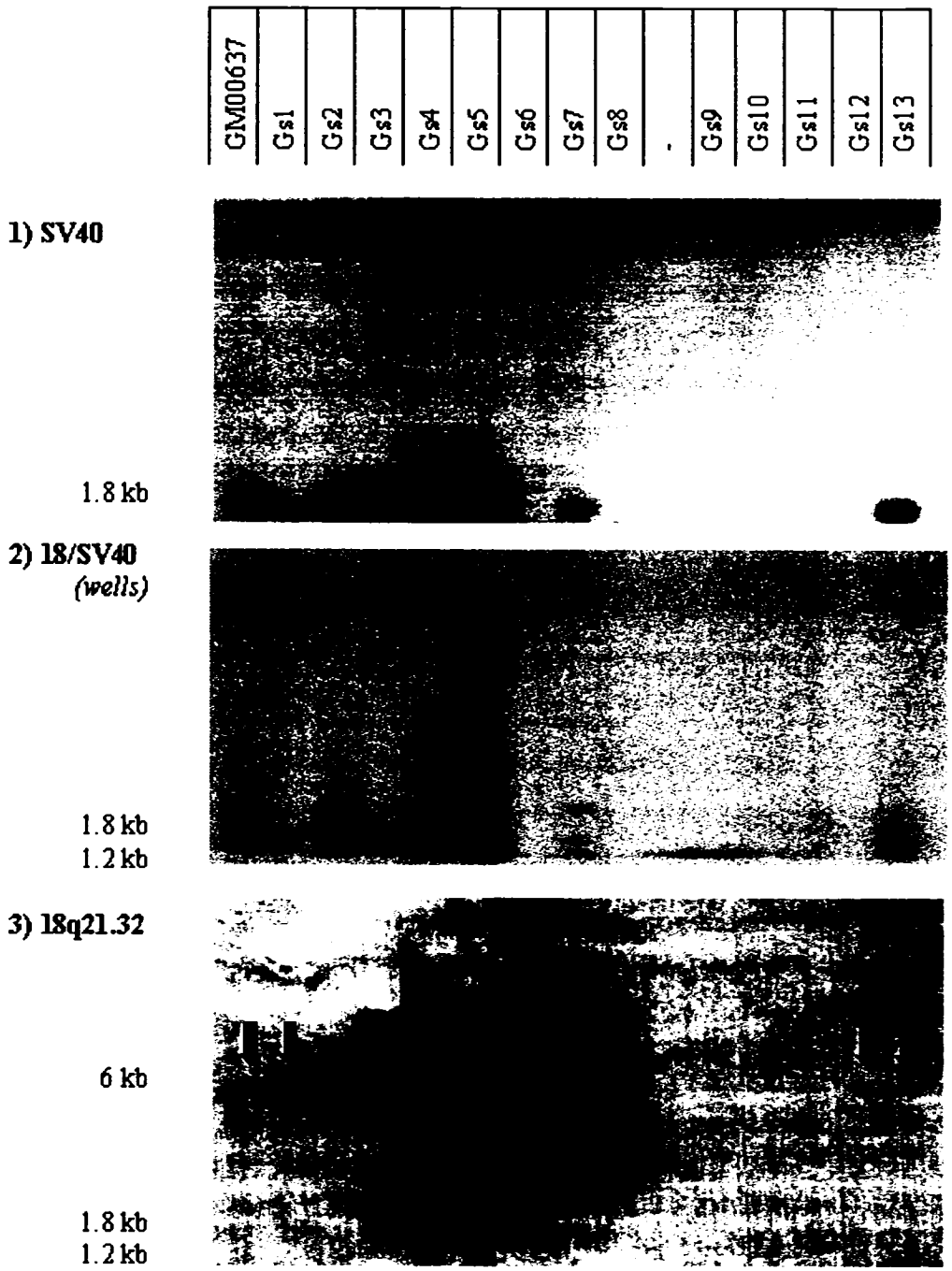



Figure S1. Southern blot of GM00637 cell line and subclones Gs1 – Gs13, probed for: 1) SV40 (1 hr exposure); 2) chimeric 18q21/SV40 (5 days exposure); and 3) 18q21.32 (21 days exposure). Genomic DNA is digested Hind III. In 1) only the 1.8 kb band appears (and not the 1.2 kb or any smaller band), because the membrane below these bands was obscured from the film. The membrane was cut at the bottom of the 1.2 kb range in such as way as to possibly eliminate other smaller SV40 bands from the image. In 3) black arrows point to a band at 6 kb (faintly present in most or all lanes); grey arrows point to bands remaining from hybridization with SV40 probe in lanes Gs4 and Gs5.

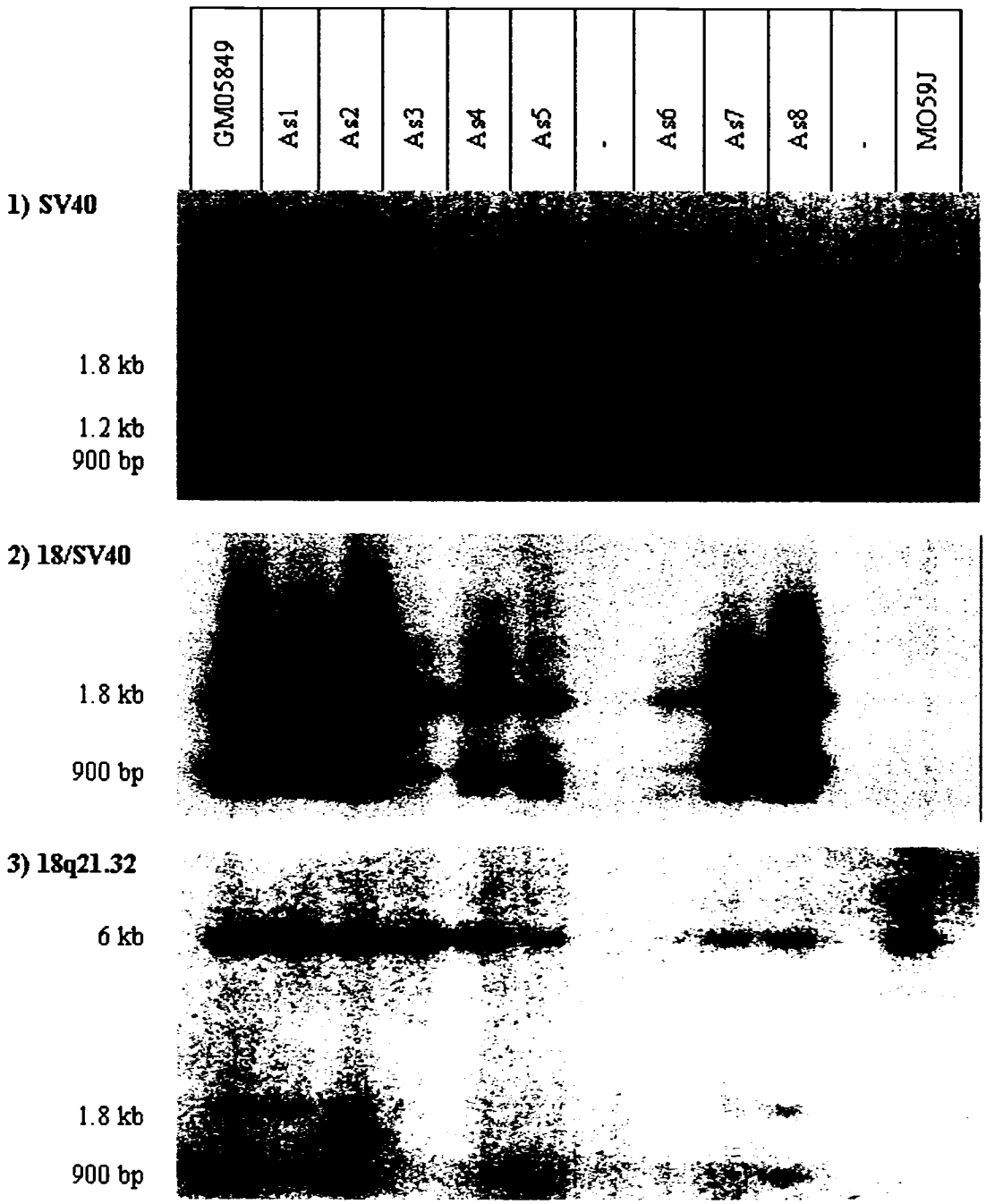



Figure S2. Southern blot of GM05849 cell line and subclones As1 – As8, probed for: 1) SV40 (1 hr exposure); 2) chimeric 18q21/SV40 (5 days exposure); and 3) 18q21.32 (21 days exposure). Genomic DNA is digested Hind III. The genome of SV40 was digested Hind III and labelled, then hybridized to the membrane. All lanes share bands at approximately 1.8 kb and 900 bp (though faint on this exposure, confirmed on longer exposure). MO59J is a control human cell line (not transformed with SV40).

The variable SV40 content of the GM05849 subclones is not due to a difference in quantity of gDNA loaded, since the intenisty of the SV40 bands in Figure S2(1) and (2) are not the same as the intensitites of the corresponding 18q21.32 specific bands in S2(3). In particular, subclone As.2 contains the greatest concentration of SV40 of any of the subclones, but in fact less DNA was loaded for that sample as compared to As.3, which has a darker band for the 18q21.32 (hence, more DNA loaded) but much less SV40 staining.

Taken together, this SV40 content data indicates that GM05849 does not entirely lose SV40 DNA during subcloning, but that the amount of viral DNA in a given cell can be amplified leading to more intense hybridization in some subclones. Also, the structure of the SV40 can change in a cell population over time, as evidenced by the appearance of a novel SV40 band in subclone As.2.

### 3.1.12 FISH Analysis of SV40 Integration

While no clear signal or integration point for SV40 could be detected in the GM00637 cells (see Figure F1), the GM05849 cells gave a very specific single integration signal (see Figure F2). GM00637 metaphases gave a high background signal. GM05849 cells show a signal with the SV40 probe towards the telomeric end of a small submetacentric chromosome. Two cytogeneticists in our institute independently identified this as chromosome 18, although in some metaphases there appeared to be elongation of this 18q arm (harbouring the SV40 signal).

The recurring junctions of SV40 and 18q21.31 sequenced from the extrachromosomal inserts of GM05849 gave strong indication this was an integration point for SV40 in this cell line. In addition, SV40 DNA was never captured alone as an insert, but always in conjunction with 18q sequences. In contrast, while GM00637 inserts sometimes showed recurring juntions between SV40 and chr 12q21.33, many other inserts were derived from SV40 DNA alone. The Southern blot analyses revealed all sublcones of GM05849 retain SV40, although it can become rearranged (producing novel Hind III fragments) and amplified (producing more intense bands).

## FISH Analyses of SV40 DNA in Human Cells

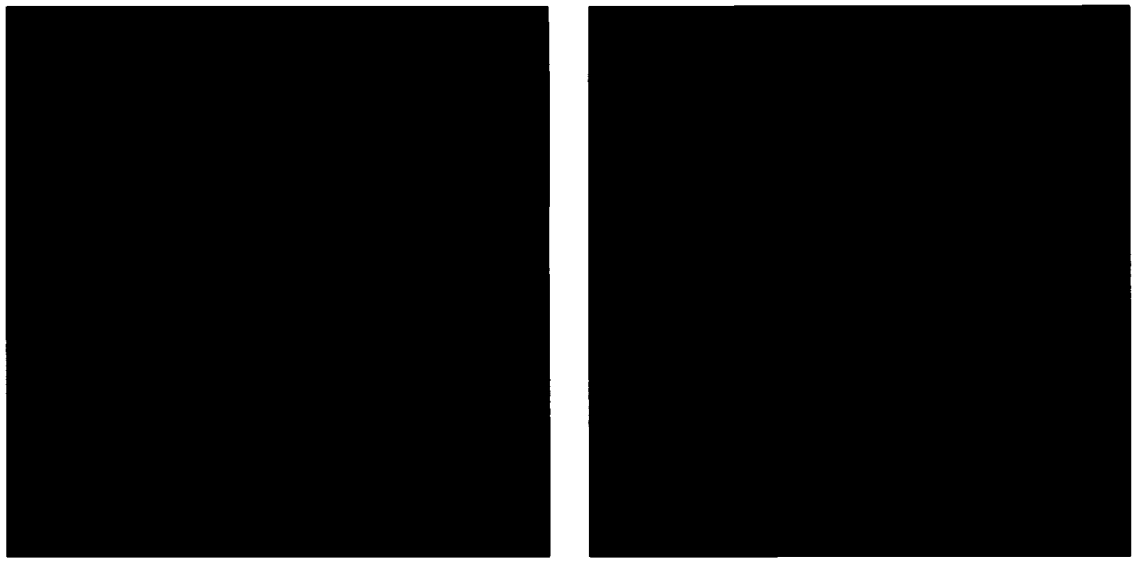


Figure F1. FISH analysis of cell line GM00637 probed with SV40. Two different metaphase spreads are shown, merged image (DAPI-stained and overlaid with image from fluorescence filter). No specific signal for SV40 is detected within a chromosome, only background hybridization.

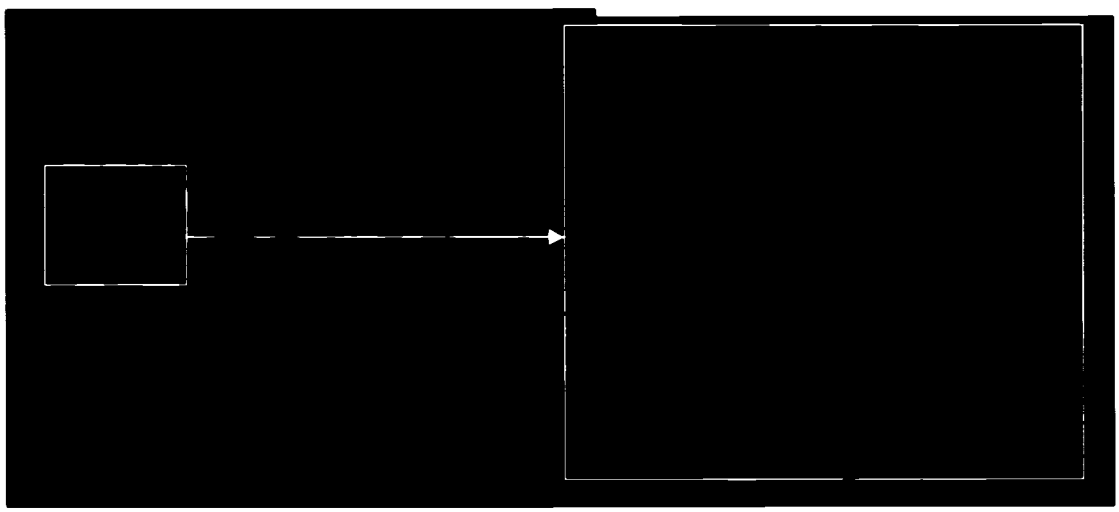


Figure F2. FISH analysis of cell line GM05849 probed with SV40. Merged image is shown (DAPI-stained metaphase spread overlaid with image from fluorescene filter for probe), full field (left) and magnified (right). Specific signal for SV40 is detected towards the end of the long arm of a smaller chromosome (independently identified by two cytogeneticists as chromosome 18, possibly with additional material added to the long arm in some cells).

In contrast to the AT cells, GM00637 subclones did not always retain detectable levels of SV40, but those cells which did maintain SV40 did so in large quantities relative to GM05849 cells. The FISH results show integration of SV40 in chr 18q of GM05849 and elongation of the arm in some cells of the population, and no background signal.

Taken together, these results suggest that SV40 is maintained in large quantities as episomal DNA in the GM00637 cells, which could account for the lack of specific integration signal from FISH and the high background signals. This also explains the predominance of SV40 DNA alone as a source of DSB inserts. On the other hand, the SV40 of GM05849 cells is not maintained extrachromosomally, but remains integrated at 18q21 and capable of causing amplification and rearrangement of the surrounding genomic region, leading to multiple fragment junctions as detected in the sequenced inserts.

### 3.1.13 Extrachromosomal DSB Inserts: Summary

We have demonstrated that insertions can occur at extrachromosomal DSBs (both blunt and staggered). These inserts derive from a variety of sources: transfected molecules (plasmid or bacterial DNA), episomal viral DNA, integrated viral DNA in conjunction with flanking human DNA, or human chromosomal DNA. Some inserts could not be fully characterized. While a few inserts involved the insertion of more than one fragment, the majority of inserts involved a lone DNA molecule ranging in size from 500 bp – 18 kb.

This DSB insert phenomenon is surprisingly frequent, as we detect it occurs in 0.5% – 1% of all DSB repair events (from normal and AT human fibroblasts). The limited number of DSB repair events characterized from other cell lines suggests this frequency may be equally high in other human cell types (ex. 1.5% of DSBs repaired with an insert in NT2). As described, our method has limitations to the size of insert it can detect, implying DSB insertional repair may account for an even greater proportion of DSB repair events, at least in an extrachromosomal context, than we report here.

It is conceivable that the differences observed in normal vs AT cells could result, not from a lack of these genomic fragments in the normal cells, but by their being out–competed by the presence of episomal SV40 DNA. If this were the case, and the episomal SV40 DNA fragments we detect in the normal cells were able to outcompete, then we would not expect to observe both cell lines sharing the same proportions of the other categories of capture events, which we did in fact detect. For instance, we observed 4/31(in normal) and 4/33 (in AT) distinct events of SV40 + human DNA being captured. Likewise, in the case of transfected linear molecules potentially outcompeting cotransfected plasmid, again in both cell lines we see approximately the same proportion of capture events which used our input linear plasmid: 5/31 in normal and 13/33 in AT. It is only with respect to the events of use of SV40 inserts alone in normal cells (13/31) and genomic DNA alone in AT cells (13/33) that a difference is observed between the cell lines.

## *3.2 Chromosomal DSB Repair Assay*

### 3.2.1 Introduction

In the preceeding section we have shown an alternative outcome of DNA repair which human cells are capable of carrying out, whereby fragments of DNA from a variety of sources (transfected, viral, and chromosomal) are able to insert into a break during DSB repair. There remains the possibility, however, that DSB repair on an extrachromosomal substrate, particularly a small foreign piece of plasmid DNA, may not be entirely representative of what goes on in the human genome. Perhaps plasmid DSBs become compartmentalized differentlly in the nucleus, not subject to the same constraints, conditions, or repair factors as chromosomal DNA.

DSB repair of a human chromosome must contend with compaction on the order of 1,000 times greater than naked DNA (during interphase), and spatial restriction to specific chromosmal territories. This limits the exposure of a given locus to a subsequent neighbourhood of certain other portions of its own, and a select few other, chromosomes [29]. As such the quantities, sizes, and sources of DNA fragments (if any) available to a chromosomal break for insertion could differ from the extrachromosomal results we obtained. We thus asked ourselves: could a system be designed that permitted investigation of DSBs at chromosomal sites of DNA damage, with the aim of identifying inserts?

The methodology needed to observe chromosomal DSB repair would require the ability to generate a specific double–strand break in chromosomal DNA, and then a means to identify events of insertion at that DSB. We therefore designed a system to generate unique and specific DSBs using the rare–cutting I–Sce I endonuclease. This site would be included in a plasmid, which would be integrated into the cellular genome, and thereafter could be induced to cut creating a unique DSB. This plasmid would also permit rescue of the repaired DSB from the genomic locus, to isolate and study repair events. The DSB sites would be placed in such a way as to activate a selectable marker when an insertion event took place, by inactivating the HSV–TK gene and thereby conferring resistance to Ganciclovir. (For details on plasmid contruction, refer to Appendix B).

## 3.2.2 Chromosomal DSB System: Design

### *3.2.2.1 Introduce an Inducible DSB Site*

In order to be able to create a specific double-strand break in the cellular genome, and follow the outcome of its repair, we introduced a unique sequence which could be cleaved upon expression of the I-Sce I restriction enzyme. This enzyme is a homing endonuclease isolated from a mobile mitochondrial intron of the yeast *Saccharomyces cerevisiae*, which recognizes and cleaves an 18 bp sequence [170]:

5' ... T A G G G A T A A ˇ C A G G G T A A T ... 3'
3' ... A T C C C ^ T A T T G T C C C A T T A ... 5'

With such a long recognition sequence, I-Sce I sites are expected to occur only once in every $7x10^{10}$ base pairs of random sequence, or once in 20 mammalian genomes, although the enzyme has been reported to tolerate some sequence degeneracy while still cutting [171]. A search of the genome databases at NCBI, Ensembl and UCSC reveals the exact 18 bp recognition sequence does not occur anywhere in the sequenced human genome. By first integrating the recognition site into the genome of the cell of interest, then expressing the enzyme in a clonal population of these cells, we induce a DSB at a specific location, select for events of DSB repair, and recuperate the original site from the genome as a plasmid to facilitate analysis.

Two plasmid substrates were constructed to act as an inducible DSB site. Both were modifications of pNTKV-1901, wherein oligonucleotide linkers were inserted to include particular restriction sites for both: 1) I-Sce I induction of a DSB; and 2) plasmid rescue (PR) by a variety of restriction enzyme combinations. PR requires excision of the integrated plasmid from the genome, and recircularization of the ends to permit efficient bacterial transformation with the plasmid. These sites needed to include the possibility of cutting once within the plasmid, then again in the adjacent genomic DNA, to permit identification of the genomic integration junctions (ideally on both sides) while maintaining the plasmid's

functional component sequences. The goal is to maximize the chances of successful PR, but minimize the risk of cutting within any DSB inserts.

The dual markers of pNTKV–1901 meant Neomycin resistance could be used as an initial selection to establish clones with integrated copies of the vector. The DSB sites were introduced in such a way that cutting, followed by a DSB insertion, should result in loss of HSV–TK expression. Since the human cells we wished to study all express endogenous TK, HSV–TK could not be used for positive selection (in HAT medium). Instead failure to express HSV–TK prevents the host cell from converting Ganciclovir to a toxic metabolite (thus killing only those which continue to express HSV–TK) regardless of the endogenous human TK. The first construct had an I–Sce I site introduced between the MCI Promoter and Thymidine Kinase (TK) sequences, and was named pOdin. The second construct had two I–Sce I sites introduced in opposite orientations (to avoid cohesive ends), one before the MCI Promoter and the other after the TK Poly A sequences, and was named pZaphod (see Figure C1, A).

## Chromosomal DSB Assay Plasmids

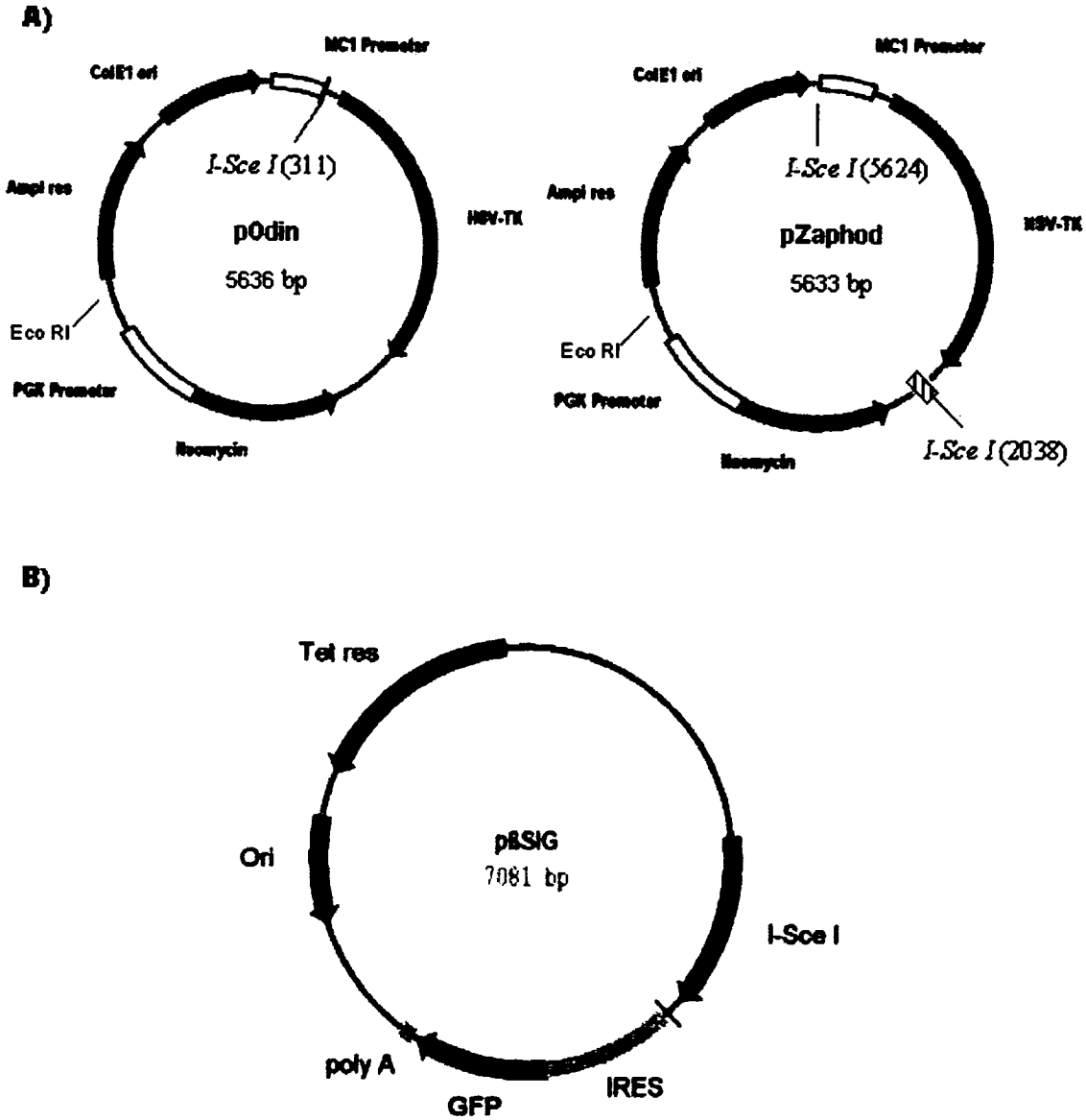



Figure C1. Plasmids used to study chromosomal double-strand breaks (DSBs) in human cells. A) DSB substrates: The pNTKV-1901 plasmid, encoding Neomycin resistance and HSV-TK, was modified to include recognition sites for the rare-cutting endonuclease I-Sce I. pOdin has an I-Sce I site placed between the MC1 promoter and HSV-TK, while pZaphod has two I-Sce I sites flanking the entire HSV-TK cassette. Linearization with Eco RI was performed before transfection. B) The pβactin-SceI plasmid (which expresses the I-Sce I endonuclease under control of the βactin promoter) was modified to express green fluorescent protein (GFP) from the same mRNA transcript by inclusion of an internal ribosomal entry site (IRES), creating plasmid pβSIG.

### *3.2.2.2 Identify Suitable Mother Clones*

To establish a mother clone which could be used in such an assay, and before moving on to the DSB–induction phase of the study, there are three principal criteria which must be met:

1. A single copy of the vector integrated
2. Ability to perform plasmid rescue (PR)
3. HSV–TK must be active (cells die in Ganciclovir selection)

Having more than one copy of the plasmid integrated in a clone would complicate the analysis of plasmid rescue, producing different sized plasmids and introducing regions of homology between multiple break sites, which could favour repair events other than the desired DSB inserts. We would want to apply such a system to different human cell lines with a variety of genetic backgrounds, to determine the influence of specific factors on DSB insertional repair. For each of these cell lines, we would want to have mother clones wherein the DSB substrate plasmid had integrated into various loci, so as to account for any site–specific variations of local chromatin context (early or late replication, active or repressed transcription, highly recombinogenic or not, etc.).

We showed with the extrachromosomal assay that techniques which rely on PCR (like those recently reported in plants, yeast and rodent cells) would have overlooked many potential DSB insertional repair events [109-114]. Hence a chromosomal DSB assay should not have to rely solely on PCR as a means to detect inserts. This requires that any mother clones harbouring a single integration of the DSB substrate plasmid must have retained the essential components of a functional plasmid, to be used in plasmid rescue. Since we would use electroporation of linearized plasmids (to favour single integrations), candidates would require minimal degradation of the linear ends upon integration to maintain the ampicillin resistance gene, origin of replication, and other plasmid sequences.

The extrachromosomal DSB assay suggests that 0.5% – 1% of all DSB repair events resulted in an insertion event. Since the process of inducing and characterizing a chromosomal DSB requires so much more work, and we could not be sure DSB inserts would be as frequent in

a chromosomal context, a means to enrich for insert events is highly desireable. Ensuring we apply the assay only to mother clones which will die when treated with Ganciclovir greatly increases the chances of observing insert events.

Once these three conditions have been met to begin the DSB analysis (a list which eliminates a large proportion of the clones generated), the clone must also display other characteristics. The cells require the ability to survive a second round of transfection (with the I–Sce I expression vector), and to express that plasmid at sufficently high levels to generate DSBs. They must also persist long enough in culture to permit harvesting of sufficient amounts of genomic DNA for analyses – a factor which becomes vital when we become aware of the toxicity of Ganciclovir to "bystander" cells (see below).

### *3.2.2.3 Induce the DSB*

To create DSBs we intially performed a number of transfections with pβactin–SceI (a gift of Maria Jasin), a plasmid which expresses the I–Sce I endonuclease under control of the βactin promoter. We had previously used this plasmid successfully in several studies in our lab, with sufficient cutting efficiency in rodent cells [89,100]. However, once we began applying pβactin–SceI to human cells, we found little evidence of I–Sce I cutting in our clones. Other attempts were made to create a dual expression plasmid for both I–Sce I and green fluorescent protein (GFP), which incorporated sequences that would later enable construction of an Adenovirus for high–level expression, though this construct proved to be of no use (see Appendix B). We succeeded in modifying pβactin–SceI by introducing an internal ribosomal entry site (IRES) sequence after the I–Sce I coding sequences, and before the GFP. This new plasmid (named pβSIG) encodes a single mRNA transcript which is then processed by the host cell to produce both I–Sce I and GFP (see Figure C1, B).

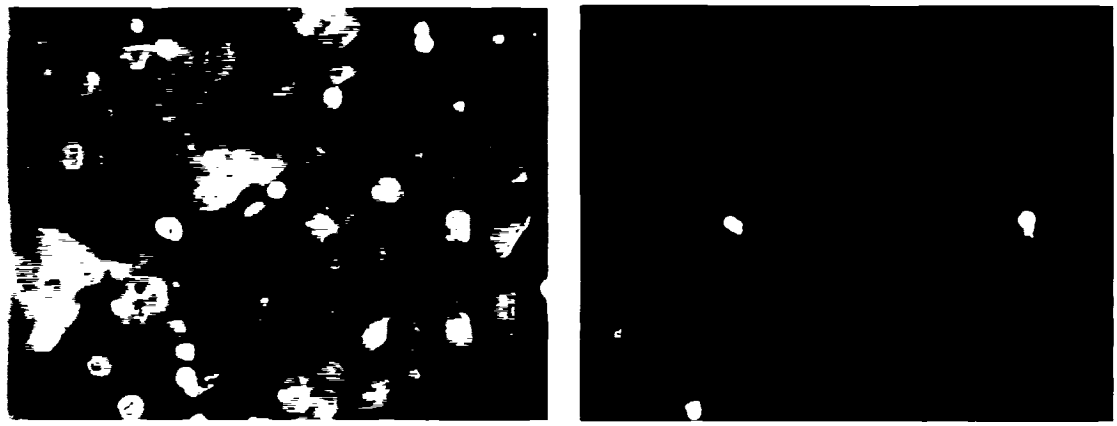
### *3.2.2.4 GFP Screening and Selection*

Not all mother clones survived equally well in culture, nor recovered well from the I–Sce I transfections. And even if they did survive well enough, the pβactin–SceI system gave us no way of determining which clones were permissive for I–Sce I expression when transfected.

The use of pβSIG thus carried 2 advantages: 1) it allowed us to identify which mother clones could be transfected efficiently, by evaluating GFP expression levels with fluorescence microscopy (see Figure C2) ; and 2) it provided us a means to isolate the proportion of transfected cells with the highest GFP (and hence I–Sce I) expression levels. Fluorescence activated cell sorting (FACS) can be used not only to count the number of cells expressing GFP (for further assessment of transfection efficiency), but to physically separate populations from the transfected pool based on their relative intensity of expression (see Figure C3). In this way, our efforts could be concentrated on studying clones which not only fulfilled the three principal criteria, but which we knew could survive the assay process and support a higher potential for DSB induction.

**Expression of GFP from Control and I–Sce I Plasmids**

**A) pEGFP–N1**



**B) pβSIG**

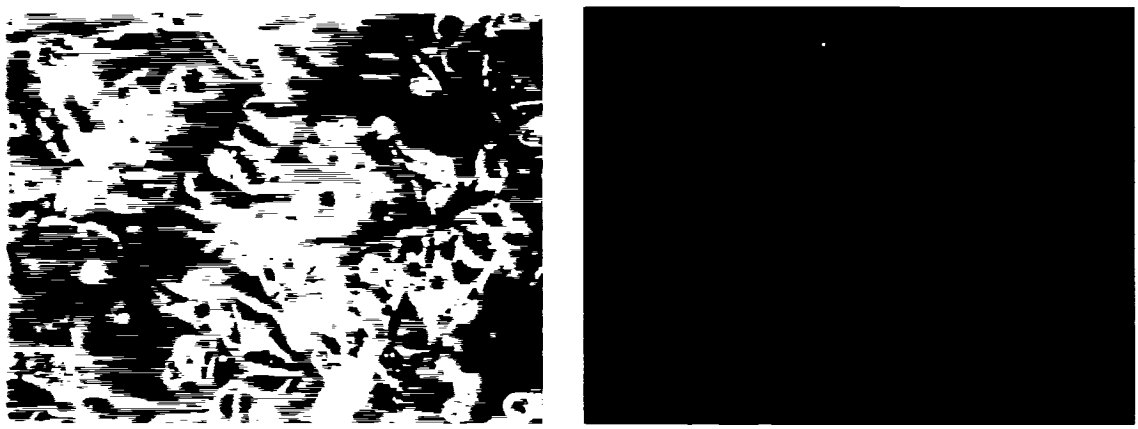


Figure C2. Light microscopy (left) and green fluorescence microscopy (right) of the same field of GM05849 cells transfected with a GFP control or pβSIG. $3.5x10^5$ cells were electroporated with 10 μg DNA and cultured in 6–well dish for 48 hrs prior to imaging (magnification: 200X). A) GM05849 cells transfected with pEGFP–N1 (a high–expression GFP control plasmid, containing the SV40 ori and hence replication competent in these cells). B) GM05849 cells transfected with pβSIG. Parallel transfections with 20 μg and 50 μg pβSIG resulted in greater cell death (with increased DNA concentration) at 48 hrs, with no apparent increase in GFP expression.

**Use of GFP to Sort I-Sce I Expressing Cells by FACS**

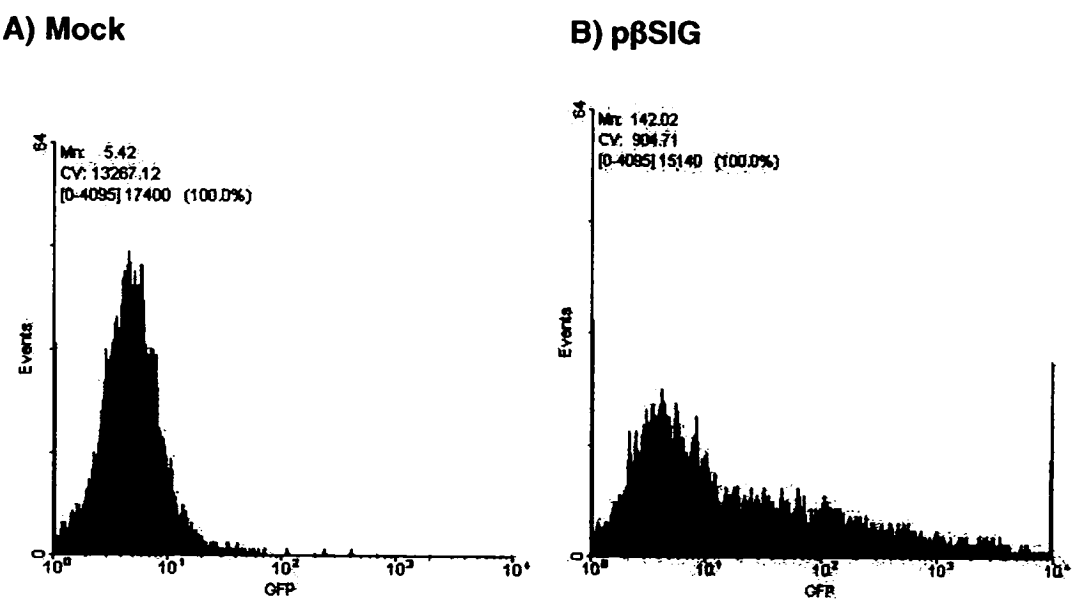



Figure C3. FACS profile of mock and pβSIG transfected cells from clone AZ7. Following transfection with A) no DNA, or B) pβSIG, cells were passed through a FACStar cytometer after 36 hrs. The number of cells is on the vertical axis, increasing intensity of green fluorescence on the horizontal axis. The 10% – 20% of cells from B) with the highest green fluorescence, corresponding to the highest expression of pβSIG, were physically separated from the cell population and cultured, while the remaining 80% – 90% of cells were also kept and maintained in either CM or CM+G418, with or without Ganciclovir selection.

### *3.2.2.5 Plasmid Rescue*

Once all 3 principal criteria are fulfilled, and the clone of interest survives re-transfection well enough to be sorted on FACS, cells must be propagated sufficiently to both freeze an aliquot and harvest genomic DNA. Plasmid rescue (PR) analysis can then begin. The gDNA is digested and religated under dilute conditions to create an intact plasmid from a single digested fragment. We initially employed Hind III, which cleaves once in the plasmid 3' of the HSV-TK cassette, thereby recovering the I-Sce I site(s) and necessary plasmid components with the 5' integration junction. We later switched to Bgl II, which cleaves right next to Hind III but also cuts more frequently in the human genome, generating smaller plasmids which are easier to manipulate. The enzyme Mfe I was also used to recover both

integration junctions at the same time, since it does not cleave within the plasmid but does cut frequently within the genome.

One difficulty with PR, however, is that compound ligation products can still occur. The PR process generates few plasmids, relative to the mass of genomic DNA, and is much less efficient than direct bacterial transformation with, for instance, a Hirt extract. This infrequency introduces an elevated risk of contamination from other plasmids, particularly in a lab such as ours where these procedures are ongoing amongst several projects.

To complicate analysis further, not all plamids which are recovered from a clone of interest necessarily represent DSB misrepair events. Particularly when FACS sorting is not employed (as with pβactin–SceI transfections), the vast majority of plasmids will result from events where the I–Sce I site was never cleaved at all. In this case, "enrichment" is performed by digesting the PR ligation mix with I–Sce I prior to transformation of bacteria. This will cleave plasmids which did not cut I–Sce I *in vivo*, and eliminate them from the bacterial colonies we need to screen.

### 3.2.3 Implementing the Assay

With all of these conditions and requirements identified, we set out to see if the chromosomal DSB assay system could work. The initial process for generating clones was transfection with Eco RI linearized pOdin or pZaphod by electroporation (to favour single integrations), and selection in G418 for the presence of Neomycin resistance. Individual clones were picked, and when enough cells were available to freeze aliquots and harvest genomic DNA, these were then screened by Southern blot in the aim of distinguishing single integrations from multiples. A HSV–TK probe was used initially, to identify: 1) the number of copies integrated; and 2) to ensure the HSV–TK (which is located near one of the linear extremities) had not been degraded upon integration (see Figure C4).

Ultimately, Southern hybridization had to be done with an ampicillin probe, rather than HSV–TK, to establish why some clones could not be successfully plasmid rescued. The

screening of each clone in Ganciclovir was still a necessary step, regardless of the HSV–TK Southern results.

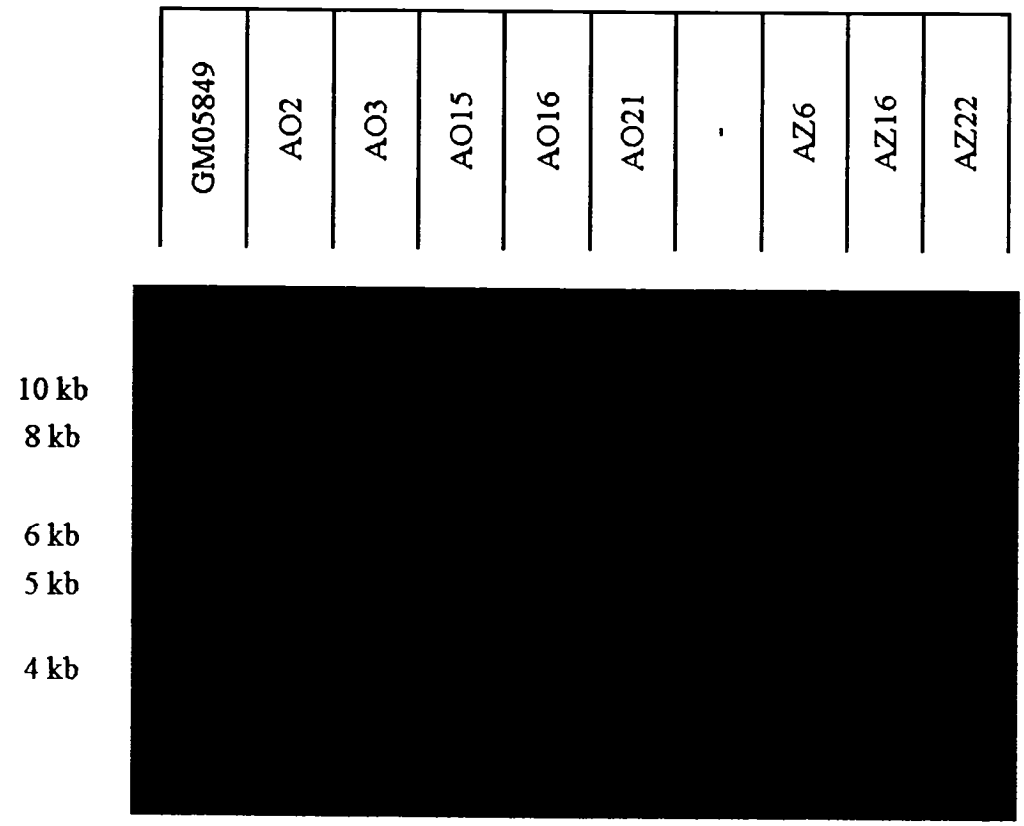



Figure C4. Representative blot showing copy number determination of integrations in mother clones. Genomic DNA (here from GM05849 cells) was harvested from individual clones transfected with pOdin or pZaphod. Samples were digested Hind III (which cleaves once in the plasmid), used for Southern blots, and probed for HSV–TK. Clones with multiple bands (indicating multiple integrations) were eliminated as candidates for our study.

### *3.2.3.1 Efficiency of I–Sce I*

While several groups have applied cellular I–Sce I systems, few even attempt to assess the cutting efficiency of their systems – that is, the proportion of cells which actually incur a DSB [116,172]. Even detecting transfection efficiency is no guarantee a site has been cleaved. The method we have traditionally used in our lab for testing efficiency of I–Sce I cutting *in*

*vivo* is to perform PCR analysis on the site. After digesting the PCR product with I–Sce I, the maintenance of a band which is resistant to cleavage indicates minor misrepair events which took place at the DSB locus in cells.

Obtaining an accurate assessment of I–Sce I cutting is difficult. Just as PCR would not detect the large DSB inserts we wished to study, major misrepair of the I–Sce I site would compromise the annealing sites for a resistant–band PCR evaluation of cleavage efficiency. However, the method was practical as a means of establishing whether a given clone was conducive to any cutting with the pβactin–SceI transfections. And more, to initially test the pβSIG construct, we did do resistant–band PCR's to ensure the I–Sce I ORF was expressed and functional (see Figure C5).

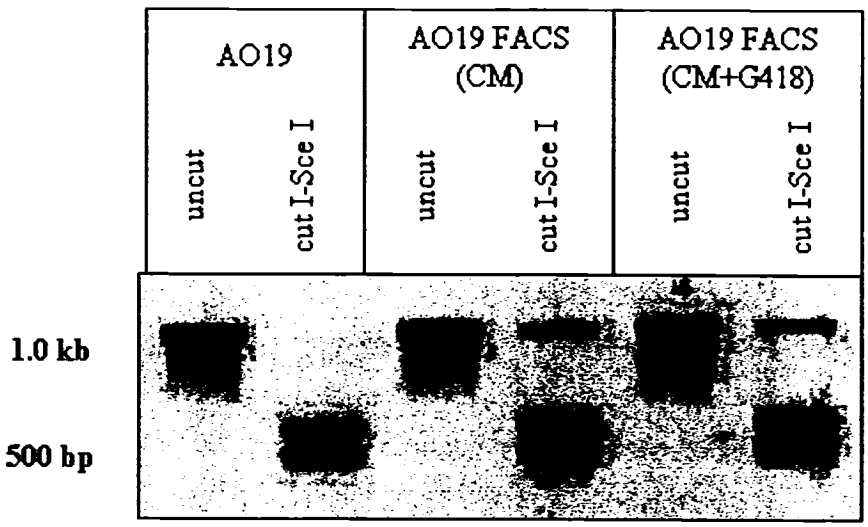



Figure C5. Assessment of I–Sce I cutting efficiency *in vivo* by PCR. Loss of the single I–Sce I site from populations of clone AO19 were determined following transfection with pβSIG and FACS sorting. Ethidium bromide stained gel (inverted colours). PCR with "O.DSB" primers was performed on gDNA (digested Bgl II) harvested from populations of AO19 cells not transfected ("AO19"), or pβSIG transfected cells cultured in CM only ("CM": no G418 from 24 hrs before transfection and thereafter) or CM+G418 ("CM+G418": G418 selection maintained throughout). Cells were transfected with pβSIG, sorted on FACS to isolate the 10% of cells from both populations (CM and CM+G418) expressing the most GFP. Cells were then cultured in CM or CM+G418 (without Ganciclovir) until adequate numbers of cells were available for gDNA extraction. "O.DSB" primers used for PCR on all samples produce bands at 508 bp and 605 bp following digestion of the PCR product with I–Sce I, indicating most cells in the population retain an intact I–Sce I Site. The presence of the 1.1 kb band in the FACS–sorted cell populations shows some I–Sce I cutting took place *in vivo* which resulted in religation of the site without reconstitution, producing a PCR product resistant to further digestion.

### *3.2.3.2 Mother Clones Generated*

We concentrated our efforts on the same cell lines as were used for the extrachromosomal DSB assay. Again, most of the work was done on clones from GM00637 (normal) and GM05849 (AT) human fibroblasts. But we also wished to ensure the system could be applicable in other cells, and generated clones from NT2 (teratocarcinoma), MO59K

(glioma), and MO59J (glioma, deficient in DNA–PKcs). We successfully isolated and propagated a total of 81 mother clones up to the point where sufficient cell material could be screened for integration copy number and HSV–TK activity, and then had to establish which of those could be successfully plasmid rescued (see Table C–A).

| Cell Line | # Clones O: Z | # single integration | # TK active | # plasmid rescue |
|---|---|---|---|---|
| GM00637 | 12: 8 | 9: 3 | 9: 2 | 3: 1 |
| GM05849 | 7: 11 | 5: 6 | 5: 4 | 3: 2 |
| NT2 | 10: 17 | 1: 2 | 1: 1 | 0: 0 |
| MO59J | 9: 6 | 2: 1 | 2: 1 | 0: 0 |
| MO59K | 2: 0 | 1: 0 | 1: 0 | 0: 0 |
| *Sum O: Z (Total)* | *39: 42 (81)* | *18: 12 (30)* | *18: 8 (26)* | *6: 3 (9)* |

Table C–A. Summary of all mother clones characterized which subsequently fulfill all three criteria for use in a chromosomal DSB repair system. Not all clones were analyzed by all three methods, and numbers herein refer only to those clones which were both analyzed and fulfilled the previous criterion (column to the left). "# Clones O: Z": all of the clones generated by transfection with pOdin or pZaphod which were successfully isolated and propagated to the point where genomic DNA could be harvested and used for Southern blot. "# single integration": of all the clones analyzed by Southern, those which produced a single band on Southern blot of genomic DNA hybridized with TK and/or ampicillin resistance probe. "# TK active": of those clones with single integrations and which were tested, died when selected in Ganciclovir. "# plasmid rescue": of those clones with both single integrations and confirmed HSV–TK activity, and were subjected to plasmid rescue, the number which successfully produced discernible non–contaminant plasmids.

### *3.2.3.3 Selection Conditions*

Results from our lab and others have shown that selection for an integrated plasmid (in this case, G418 selection) must be maintained whenever the clone is cultured, otherwise the plasmid can be lost from the cell population [173]. However, dual selection conditions for 2 marker genes at the same time, such as our Neo and HSV–TK, often create the problem of promoter interference [174]. This means that one selectable marker can cease to be expressed, although its sequences remain intact. In our case, that would amount to a clone which survived Ganciclovir selection even though there has been no change at the I–Sce I DSB.

Throughout the process of selecting for DSB misrepair events, we employed combinations of selection conditions to maximize the chance of harvesting DSB inserts. A principal difficulty we encountered was the "bystander effect" of Ganciclovir selection, whereby conversion of the substrate Ganciclovir to a toxic metabolite kills not only the cell expressing HSV–TK, but neighbouring cells as well. This required we establish a fine balance with respect to: 1) achieving adequate levels of I–Sce I cutting to inactivate as many HSV–TK expressing cells as possible; and 2) plating cells at a sufficient density to sustain cell growth, while remaining sparse enough to avoid accumulating excessive toxic metabolite (from cells which remain HSV–TK+) in the culture medium.

Applying Ganciclovir selection to obtain individual colonies (daughter clones) often proved difficult. Each clone displayed a different degree of cell death as a result of the electroporation procedure itself, making it difficult to predict the density at which they should be plated after transfection with the I–Sce I plasmid. Allowing them adequate time to recover and then replating them presented the same difficulty as FACS sorting cells, in that enough cell divisions might take place to generate multiple colonies from the same daughter clone event. While this process could be fine–tuned (and was) successfully for some cell lines, the pooling of transfected cells gave more immediate results.

Following I–Sce I expression, cells could be cultured without Ganciclovir long enough to recover and propagate to the point where we could then plate this "pool" out in several dilutions. Ganciclovir could then be applied, and any cells surviving the selection could then be pooled again to encourage survival. Frozen aliquots of these pooled cells, and genomic DNA samples, then represented multiple daughter clones which could be analyzed.

### 3.2.4 Chromosomal DSB Assay: Results from GM00637 and GM05849

The clones studied most closely are listed in Table C–B, with the integration site of the plasmid and the number of bases degraded from the 5' end of the Eco RI linear plasmid (close to the ampicillin resistance gene).

| Clone | Integration Site (5' junction) | Degradation from Eco RI site |
|---|---|---|
| AO15 | 14q21.2 | 15 bp deleted (+ 11 bp unidentified) |
| AO19 | 17q23.2 | 25 bp |
| AZ3 | Unknown | Unknown (beyond ampicillin) |
| AZ7 | 8q11.1 | 36 bp |
| GO6 | 2p25.3 (MYTL1 intron) | 8 bp |
| GO20 | 1q31.1 | 11 bp |
| GZ1 | 2p11.1 | 38 bp (72 bp inversion, 10 bp duplication) |

Table C–B. Integration sites of mother clones used primarily in our study. Clones are named as follows: from GM00637 ("G") or GM05849 ("A"), transfected with pOdin ("O") or pZaphod ("Z") (linearized Eco RI), and numbered.

#### *3.2.4.1 Southern Analysis of DSB Sites*

One method for screening individual daughter clones, or even pools of cells which survive Ganciclovir selection, is to look for any shifts in the size of the integrated plasmid band by Southern blot analysis. We demonstrate this technique with samples of clone GZ1: pools 1 and 2, as well as individual daughter clones GZ1.18 and GZ1.21 (see Figure C6). Each sample gave a single band at 5.5 kb when digested Hind III and probed for ampicillin resistance (the same size as the plasmid generated by a Hind III PR). This suggested no events of DSB insertion (or deletion) took place in any of these cases – at least none which maintained the ampicillin sequences required for hybridization and PR.

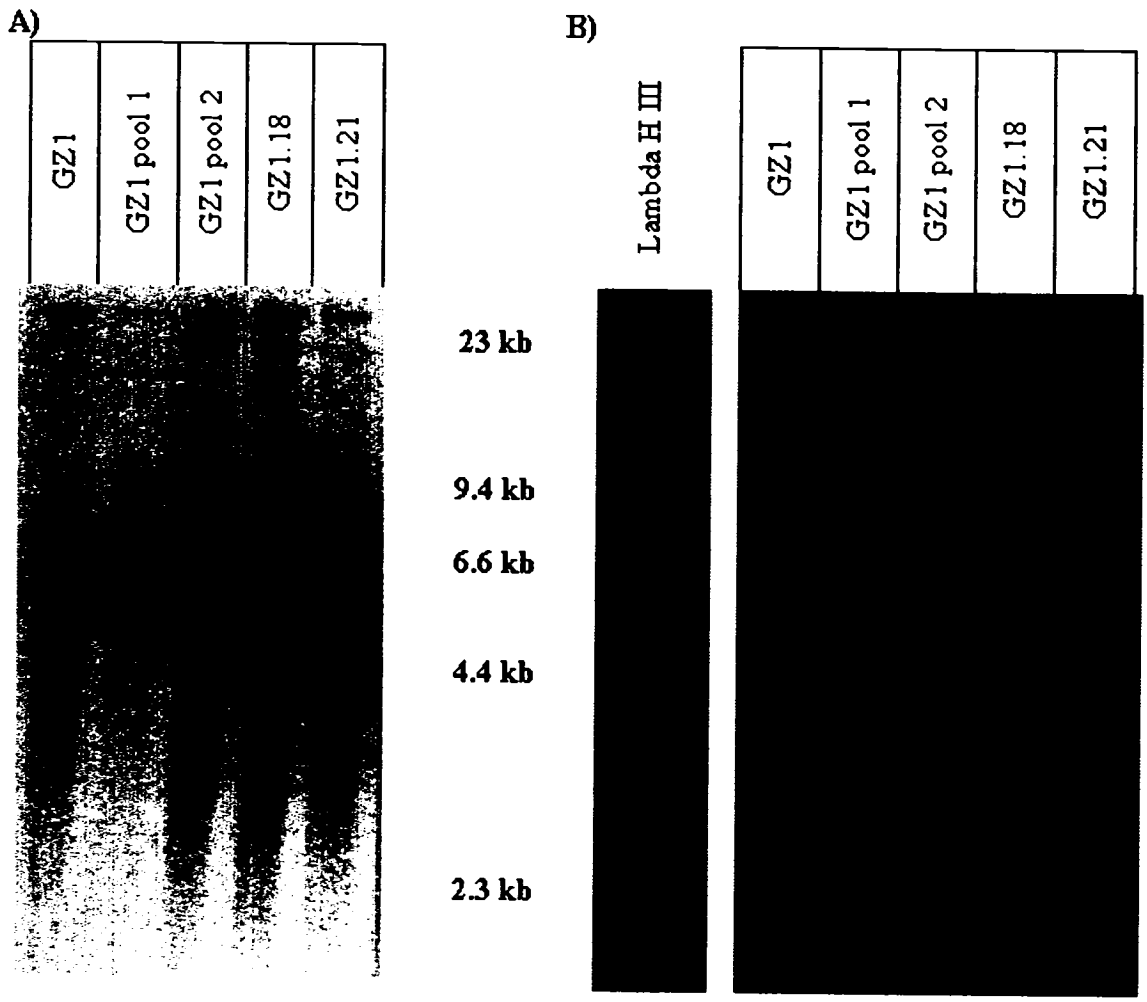



Figure C6. Southern blot analysis of genomic DNA from GZ1 events digested Hind III. A) Ethidium bromide staining (image colours inverted) of mother clone GZ1, transfected cell populations pool 1 and pool 2 (transfected pβactin–SceI), and two daughter clones (GZ1.18 and GZ1.21). B) The same gel, blotted and Southern hybridized with ampicillin resistance gene probe. Ladder is Lambda DNA digested Hind III. Intensity of band in B), including faint band present in sample "GZ1 pool 1", corresponds to amount of DNA loaded as seen in A). Maintenance of band at same size suggests few, if any, repair events occurred following transfection with I–Sce I expression plasmid that would alter Hind III digestion pattern of either pools or individual daughter clones of GZ1.

However, this same "pool 2" of cells transfected with pβactin–SceI did generate a plasmid with an event of DSB misrepair not detected by Southern. Characterization of event "GZ1 H, Gan, n.e. 208" revealed the 1st I–Sce I site of pZaphod was intact but the 2nd I–Sce I site was lost, and a fragment of chr 20p12.1 was in its place. This event was not detected by Southern

analysis, perhaps because it was not represented by enough cells in the pool to hybridize adequately. This is an event which could be a DSB insertion, introducing sequences which alter the Hind III digest pattern and exclude the 3' plasmid sequences. Or, alternatively, this could be a translocation event, with or without deletions which eliminate the 3' plasmid sequences.

The integration of pZaphod in GZ1 involved an inversion of sequences at the 5' end. However, the 3' junction of the pZaphod integration in GZ1 was never established. We cannot rule out the possibility that chr 20p12.1 sequences were already 3' of the plasmid, and that this event represents only deletion following the cut of the 2$^{nd}$ I–Sce I site. This is a valuable illustration of the need to perform PRs which identify both junctions in the mother clone prior to DSB induction. The failure of Southern blot as a means of detecting misrepair events against high background levels of intact plasmids was further motivation to develop the pBSIG system for FACS enrichment of I–Sce I expression.

### *3.2.4.2 PCR Analysis of DSB Sites*

In an attempt to reduce the costs and time needed to culture large numbers of daughter clone events prior to analysis, we developed the "Cell Lysis PCR" (see Methods). By isolating and lysing a few cells from individual clones (generated by Ganciclovir selection), we could perform PCR to screen for loss of the HSV–TK sequences. In addition, being able to isolate and analyze events this early could provide the advantages of: 1) witnessing some DSB misrepair events which might be lethal to the cell after several divisions; and 2) avoiding the often unsuccessful process of propagation and physical isolation, or "picking", of all clones, and enable us to focus our efforts on those of interest. However an internal control would be needed to ensure that absence of signal was not due to lack of DNA.

As an illustration, we developed a duplex PCR system for screening DSB events, and applied it to individual clones of AZ3, from the few cells present in the Cell Lysis solution at this early stage. Clone AZ3 had successfully produced 29 daughter clones resistant to Ganciclovir, but all attempts at plasmid rescue failed (producing nothing but contaminants). It remained useful as a candidate to test PCR as a means of detecting DSB repair. The

duplex PCR used 2 sets of primers in the same reaction tube: the "CFP/CRP" pair (which amplifies a region of the HSV–TK cassette), and another primer set which would amplify either another region of the plasmid, or a specific human genomic sequence to act as the internal control to confirm presence of DNA in the sample (see Figure C7). Attempts were made using second primer sets: 1) "C7–1", which amplifies a specific region of the PTPRE gene at 10q26 (primers used by other lab members in a separate project); or 2) "NCBI Rb", which amplifies a region of human Rb at 13q14. We obtained specific, as well as non–specific bands, sufficient to determine that none of the AZ3 daughter clones had lost the HSV–TK cassette (the original aim of the pZaphod system design).

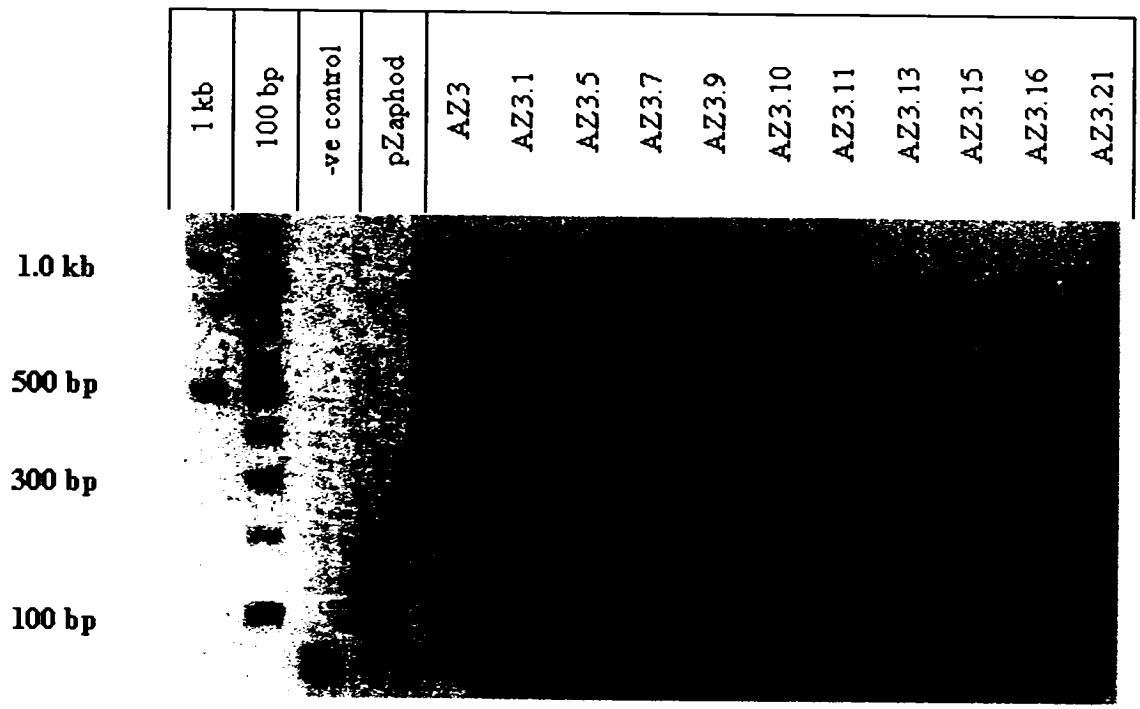



Figure C7. Duplex PCR amplification of HSV–TK and human chromosome 10 sequences on AZ3 and daughter clones. Ethidium bromide stained gel (inverted colours), composite image from 2 lanes of the same gel (same reaction). Primers "CFP/CRP" generate a 126 bp band from pZaphod, while the "C7–1" primers amplify 1031 bp from chromosome 10q26. Despite the strong presence of another band at 300 bp (also present in genomic DNA samples from other human cell lines, not shown), a faint band is detectable at 1031 bp in all samples. Although not a perfectly specific reaction, it does demonstrate the 126 bp band (indicating that portion of HSV–TK between the I–Sce I sites) is intact in all daughter clones, thus cleavage of both I–Sce I sites and subsequent excision of the HSV–TK cassette did not occur.

To confirm the results of the PCR reactions done on small numbers of cells (lysed from the intial selection plate), we propagated the clones and re–analyzed them once sufficient cells were available for genomic DNA extraction. Several of these same clones were screened with primers flanking both I–Sce I sites revealing little, if any changes, had occurred at the DSB sites, and no DSB inserts were identified (see Figure C8).

**PCR Analyses of I–Sce I Sites**

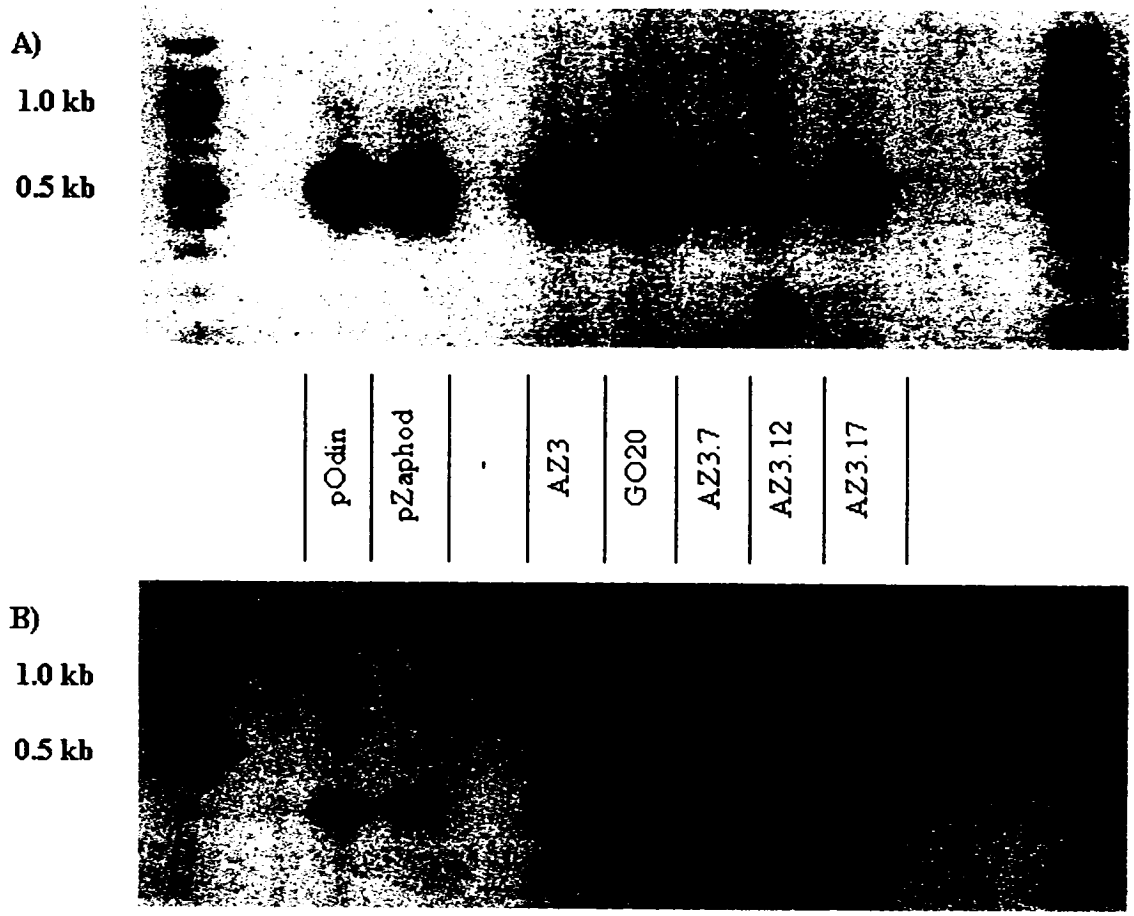



Figure C8. PCR amplification of both I–Sce I sites from AZ3 daughter clones. Ethidium bromide stained gels (inverted colours). Primer sets flanking A) the 1st I–Sce I site, and B) the 2nd I–Sce I site were used to establish if both sites remained intact in several daughter clones of AZ3. Primers "O.DSB F" and "CRP" were used to examine the 1st I–Sce I site, generating fragments of 553 bp in pOdin and 528 bp in pZaphod. Primers "2nd Z SceI" and "Z.DSB R" flank the 2nd I–Sce I site, generating fragments of 320 bp in pOdin and 345 bp in pZaphod. As plasmid rescue failed on AZ3 clones, this screening method revealed no detectable insertions or deletions took place at either I–Sce I site in the AZ3 daughter clones harvested (strong or faint bands are present for all clones).

### *3.2.4.3 I–Sce I Enrichment*

Performing a plasmid rescue and digesting the religated mix with I–Sce I prior to transformation of bacteria eliminates a great deal of background (plasmids from cells which retained intact I–Sce I sites). This enrichment consistently gave a 10–fold (or greater)

reduction in the number of colonies generated by a transformation. The efficiency of enrichment was evaluated by transforming ¼ of the Bgl II PR which was uncut versus ¼ of the Bgl II PR which was digested I–Sce I prior to transformation. The enrichment efficiency of several clones is listed below (see Table C–C).

| CLONE | Conditions | P.R. Bgl II | |
|---|---|---|---|
| | | Non–enriched | Enriched |
| GO6 | FACS | 626 | 99 |
| | non–FACS CM | 681 | 16 |
| | non–FACS G418 | 457 | 17 |
| | | | |
| AZ7 | FACS | 852 | 11 |
| | FACS CM (Gan 14 d) | 167 | 6 |
| | non–FACS CM | 1057 | 13 |
| | non–FACS G418 | 964 | 3 |
| | | | |
| AO19 | 2X pβSIG | 225 | 13 |
| | 3X pβSIG | 266 | 0 |
| | FACS CM (Gan 48 hrs) | 297 | 21 |
| | FACS CM (Gan 14 d) | 143 | 11 |
| | FACS G418 (Gan 48 hrs) | 310 | 120 |
| | FACS G418 (Gan 14 d) | 784 | 11 |

Table C–C. Bacterial colonies obtained with and without I–Sce I enrichment of plasmid rescue. Plasmid rescues were performed with Bgl II on genomic DNA harvested from cells transfected with pβSIG. Some populations were transfected, cultured for 14 days, then re–transfected a second time ("2X"), or the process repeated a third time ("3X"). Some populations were sorted on FACS, where the 10% – 20% of cells expressing the highest levels of GFP ("FACS") were separated from the remainder of the transfected population ("non–FACS"). Cells were cultured in Complete Medium with no selection ("CM") or selecting for Neomycin resistance ("G418"). Some populations were also selected in Ganciclovir for the loss of HSV–TK activity ("Gan"), for either 48 hrs ("48 hrs") or for 14 days ("14 d").

The disadvantage with I–Sce I enrichment is that it excludes events from pZaphod clones which may result from the *in vivo* loss of only one of the two DSBs (although these are unlikely to confer Ganciclovir resistance). Also, comparisons of the enrichment efficiency on PRs of FACS–sorted cells versus non–sorted provides some indication of the advantage

obtained by development of the pβSIG system. Some clones show an increase in the number of colonies resistant to I–Sce I enrichment, suggesting more DSB induction (and subsequent misrepair) took place *in vivo*. For instance, clone GO6 shows a 5–fold increase in the number of colonies obtained with FACS use. There are, however, other clones for which use of FACS does not greatly increase the number of colonies generated. Resistant–band PCR analyses on the same samples confirmed a variation in detectable DSB misrepair amongst clones: some clones generate PCR products which resist digestion more when sorted on FACS, whereas others show similar degrees of undigested PCR product from both FACS and non–FACS populations.

As mentioned previously, the resistant–band PCR is unable to detect any DSB misrepair events which involve insertions and/or deletions which cannot be amplified by PCR. It is therefore not an exact measure of the advantage of the FACS–sorting procedure. Similarly, the colonies tallied in Table C–C do not include analyses of which plasmids are authentic events from each clone, and which are contaminants, and hence are not necessarily an indication of the actual improvement of *in vivo* DSB induction either. Both analyses do, however, demonstrate an overall advantage of I–Sce I enrichment to reducing the number of colonies to screen.

### 3.2.5 Chromosomal DSB Misrepair in GM00637 and GM05849

We established that plasmid rescue was the best method for detecting DSB insert events, despite the high background levels of contaminating plasmids which were generated. This required that each plasmid harbouring a DSB misrepair event was also sequenced at an integration junction to confirm its clonal identity. The total number of DSB misrepair events detected from our analyses, characterized from the most informative clones, are listed in Table C–D.

| Clone | AO15 | GO20 | AZ3 | AZ7 | GZ1 | GO6 | AO19 | TOTAL |
|---|---|---|---|---|---|---|---|---|
| **Total # of PR colonies** | 33 | 25 | 2 | 966 | 2000 | 1858 | 2118 | **7002** |
| **Total # analyzed** | 33 | 25 | 2 | 278 | 508 | 631 | 622 | **2099** |
| **Total # NOT contaminants** | 0 | 0 | 0 | 112 | 382 | 105 | 160 | **759** |
| **Total # chracterized** | – | – | – | 11 | 16 | 24 | 109 | **160** |
| **Characterized with misrepair of DSB** | – | – | – | 5 | 1 | 5 | 56 | **67** |
| **Characterized distinct misrepair outcomes** | – | – | – | 4 | 1 | 1 | 16 | **22** |

Table C–D. Analysis of all colonies obtained by plasmid rescue following DSB induction in mother clones from GM00637 and GM05849 cells. Listed here are the clones studied most extensively in the development of a chromosomal DSB repair assay. "Total # of PR colonies": includes all plasmid rescues performed (with Hind III, Mfe I, Bgl II, or Asc I), following transfections with all I–Sce I expression plamids (pβSIG or pβactin–SceI), with or without I–Sce I enrichment of the plasmid rescue. "Total # analyzed": those colonies subjected to miniprep and diagnostic digest. "NOT contaminant": excludes all colonies which were the common 3 kb (pBluescript) or other plasmid which was not derived from the mother clone listed, or which displayed supplemental bands by diagnostic digest, indicative of complex fragment ligations during the plasmid rescue. "Characterized": all plasmids whose identities were confirmed as an authentic plasmid of a given clone, and for which the I–Sce I site was confirmed as uncut or altered. "Misrepair of DSB": all plasmid which had an I–Sce I site no longer intact. "Distinct misrepair outcome": the number of different alterations to an I–Sce I site observed for a given mother clone.

From over 2000 plasmids which were analyzed, we identified 22 different events of DSB misrepair. These events include some small deletions, duplications, and insertions picked up as a result of various screening techniques for any loss of an I–Sce I site. There were also 2 events which resulted in homologous recombination with the I–Sce I expression plasmid. A total of 8/22 events showed insertions at an I–Sce I DSB, wherein 2/22 events were small insertions of 1 and 4 bp. But most importantly, we identified 6/22 DSB misrepair events which resulted from larger insertions (ranging from at least 240 bp to more than 3.3 kb), of either foreign (bacterial) or human chromosomal DNA at the break site. Only 2/22 DSB misrepairs were observed in GM00637 (normal) cells, both of which were insertion events (of 1 bp, and more than 240 bp). The remaining 20/22 DSB misrepairs were observed in GM05849 (AT) cells, with 11/20 representing deletion events alone. These events are listed in Table C–E, and a select few are described in more detail below.

| DSB Misrepair Event | I-Sce I Site(s) Details |
|---|---|
| **GO6 (chr 2p25.3)** | |
| M, CM, n.e.2 | ins 1 bp "T" |
| **GZ1 (chr 2p11.1)** | |
| H, p2, G4, n.e.208 | 1st I-Sce I site intact, 2nd site lost and ins chr 20p12.1 (240+ bp) |
| **AZ7 (chr 8q11.1)** | |
| B, CM, e.7 | both I-Sce I sites cut, TK excised |
| B, p3, G4, n.e.31 | 4 bp del from 1st I-Sce I, then into pβactin-SceI |
| B, CM, e.1 | 4 bp del from 1st I-Sce I, then into pβSIG |
| B, CM, e.9 | 1st I-Sce I cut, then ins chr 17q22 (616+ bp by sequence) |
| **AO19 (chr 17q23.2)** | |
| 3x pβSIG, B, n.e.10 | del 1 bp |
| 2x pβSIG, B, e.10 | del 4 bp |
| B, CM, 2d, e.13 | del 5 bp |
| 3x pβSIG, B, n.e.34 | del 8 bp |
| B, CM, 2d, e.1 | del 9 bp |
| B, CM, 2d, e.21 | del 21 bp |
| B, G4, Gan, 14d, e.3 | del 324 bp |
| M, CM, Gan, 14d, e.10 | del 587 bp |
| M, CM, Gan, 14d, n.e.46 | del 875 bp |
| M, G4, Gan, 14d, e.1 | del 235 bp at I-Sce I site, (572 bp HSV-TK intact), del 1034 bp HSV-TK |
| B, G4, 2d, e.29 | dup 28 bp |
| B, CM, 2d, e.14 | ins 4 bp "TTCG" |
| M, CM, Gan, 14d, n.e.9 | ins *E. coli* genomic DNA (660+ bp by sequence, 3329+ bp by digests) |
| B, G4, Gan, 14d, e.10 | ins chr 17 (2318+ bp from 26 kb before integration site) |
| B, CM, Gan, 14d, e.9 | ins 18 bp satellite DNA, then chr 17q from 1.5 Mb 5' (360+ bp, PPM1E intron) |
| M, CM, Gan, 14d, n.e.21 | ins chr 8p12 (662 bp sequenced, 3141+ bp by digests) |

Table C-E. Events of DSB misrepair at chromosomal I-Sce I sites in GM00637 (normal) and GM05849 (AT) cells. Events are listed beneath the name and chromosomal integration site of each mother clone. Events are named according to: plasmid rescue enzyme used ("H" = Hind III, "B" = Bgl II, "M" = Mfe I); cell culture conditions ("CM" = complete medium without G418, "G4" = complete medium with G418); use of Ganciclovir selection (="Gan") and duration of Ganciclovir selection ("2d" = 2 days, "14d" = 14 days). "p3" refers to a pool of selected cells, transfected with pβactin-SceI. "2x" and "3x" refer to consecutive transfections with pβSIG without FACS sorting. "e" = I-Sce I enrichment applied to plasmid rescue prior to transformation of bacteria, whereas "n.e." = no enrichment. For details: "del" = deletion, "dup" = duplication, "ins" = insertion.

### *3.2.5.1 Characterized DSB Misrepair Events: Clone GZ1*

Clone GZ1 was derived from the GM00637 (normal) cell line, and has pZaphod integrated at chr 2p11.1 in a ALR/Alpha satellite repeat sequence. The integration into the chromosome was accompanied by a short inversion and small deletion of plasmid sequences at the 5' end, which made initial characterization of events difficult. However, the one DSB misrepair event isolated gave us our first indication that our system could successfully identify insertions of other chromosomal DNA at induced breaks. Event "GZ1 H, p2, G4, n.e.208" showed the 1st I–Sce I site to be intact, but the second site was completely lost, and at least 240 bp of chr 20p12.1 was inserted. The junction of the two sequences begins in a LTR on chr 20 and continues towards the telomere, in a region of relatively high local recombination frequency (2.3 cM/Mb).

### *3.2.5.2 Characterized DSB Misrepair Events: Clone AZ7*

The 4 DSB misrepair events from clone AZ7 illustrated three possible outcomes of chromosomal DSB repair: 1) deletion (resulting from cleavage of both I–Sce I sites and excision of intervening sequences); 2) homologous recombination; and 3) insertion of chromosomal DNA at a DSB.

#### 3.2.5.2.1 Success of pZaphod: Complete HSV–TK Excision

Event "AZ7 B, CM, e.7" illustrates the success of our intial goal for pZaphod: both I–Sce I sites (though in opposite orientations) could be cleaved and the ends ligated together *in vivo* so as not to reconstitute an I–Sce I site, and to remove all HSV–TK sequences to permit Ganciclovir selection. In this case, 2 bp were degraded from the 5' overhang of the first site, and no bases were lost from the 4 bp protruding underhang of the second site. This left 1 bp overlap between the two sites, and the remaining 1 bp gap from the bottom strand of the first site, as well as the 3 bp gap from the top strand of the second site, were filled in (as follows):

a) Cleavage of both I–Sce I sites

5' ... A C G G T T A G G G A T A A  C C C T A C G G A C ... 3'
3' ... T G C C A A T C C C  A A T A G G G A T G C C T G ... 5'

b) Alignment

5' ... A C G G T T A G G G A T - - - C C C T A C G G A C ... 3'
3' ... T G C C A A T C C C - A A T A G G G A T G C C T G ... 5'

c) Synthesis (fill–in)

5' ... A C G G T T A G G G A T t a t C C C T A C G G A C ... 3'
3' ... T G C C A A T C C C t A A T A G G G A T G C C T G ... 5'

However, AZ7 also showed DSB repair events which introduce another complicating factor for an ideal chromosomal DSB repair assay: the need to eliminate all homology from the constructs used, to avoid gene conversion events.

#### 3.2.5.2.2 Homologous Recombination Events

Interestingly, homologous recombination events between the integrated pZaphod and the I–Sce I expression plasmid occurred twice: once with pβactin–SceI (event "AZ7 B, p3, G4, n.e.31", which was represented by several plasmids from the same PR), and another time with pβSIG (event "AZ7 B, CM, e.1"). The 5' integration junction with 8q11.1 was intact (proving it came from AZ7), as were the plasmid sequences of pZaphod up to the 1st I–Sce I site. A deletion of 4 bp took place from the cut I–Sce I site (as well as the 4 bp underhang), and then plasmid sequences were present (at least 690 bp) which were a perfect match to the I–Sce I expression plasmid used in each experiment. The pZaphod sequences around the 2$^{nd}$ I–Sce I site were entirely absent from this plasmid (the primer failed to anneal). The transition from pZaphod into the expression plasmid sequences occurred at the same spot in both, in a region shared between all three plasmids' backbones. These two events then share the same sequence for approximately 300 bp (after the DSB), and then diverge to easily distinguish pβactin from pβSIG sequences.

These two separate events represent a repair involving homology between the DSB site and an incoming sequence, which led to gene conversion at the repair site. These events are real, and not contaminants, since both plasmids show the chr 8q11.1 of AZ7 sequences at their 5' junction. The initial 150 bp of the sequences leading up to the 1st I–Sce I site in pZaphod is the same as the sequencing results from these two events, which contain a 5 bp stretch (40 bp prior to the I–Sce I site) which is missing from either pβSIG or pβactin, thus confirming pZaphod prior (5') to the DSB site.

Both events generated plasmids shorter than the AZ7 plasmid size – which suggests these may be events in which the incoming expression plasmid picked up sequence from the integrated pZaphod vector (see Discussion for possible mechanisms involved). This would explain why the chr 8:pZaphod junction was detected, the ampicillin resistance gene was present, but the gene conversion process did not continue to include the AZ7 "5' wrap" annealing sequences located 4.5 kb upstream of the integration site (beside the 5' gDNA Bgl II site).

#### 3.2.5.2.3 DSB Repair with a Human Chromosomal Fragment

Finally, repair event "AZ7 B, CM, e.9" shows the 1$^{st}$ I–Sce I site was cleaved, 0 bp were lost from the 5' end of the site, and at least 616 bp from chromosome 17q22 were inserted (before reaching a Bgl II site in the chr 17 sequence). This event was only witnessed from PR performed with Bgl II, hence no plasmid sequences 3' of the 1$^{st}$ I–Sce I site were recovered. In contrast to the homologous recombination events, sequencing with the "5' wrap" primer showed this plasmid did include the 4.5 kb of chr 8q11.1 5' of the pZaphod integration site, up to a Bgl II site, and then around the other end of the plasmid continuing back through the Bgl II site in the chr 17q22 sequences. In this case, the event could represent either a large insertion or a translocation.

The genomic context of the insert itself reveals no clues as to why it was used as DSB repair partner. The junction of pZaphod with chr 17q22 occurs within a LINE (L1).The closest gene is 400 kb 3' (telomeric) to the insert, and the recombination rate is moderate (1.5

cM/Mb). There is a region with high MAR potential 13 kb away from the region we sequenced, though we do not know if that sequence is present at the DSB site.

Translocations involving band 17q22 are reported exceptionally infrequently, with only 2 recurrent conditions listed in the Mitelman database: 1) "del (17)(q22)", a deletion in acute megakaryoblastic leukemia; and 2) "der(22)t(17;22)(q22;q13)", a derivative chromosome resulting from fusion with a region of chr 22q13 in a fibroblastoma. As such, there is no obvious reason for the use of chr 17q22 sequences over any other, yet this event confirms other human chromosomal fragments can be used to repair chromosomal DSB sites, and that this phenomenon is detectable using our system.

### *3.2.5.3 Characterized DSB Misrepair Events: Clone AO19*

#### 3.2.5.3.1 Integration Site in AO19

The majority of DSB misrepair events were recovered from clone AO19, leading it to become the best characterized clone as well. It may be that this clone permitted a greater degree of DSB induction than others, as the frequent presence of an uncut band from the resistant–band PCR suggests (data not shown). In the case of AO19, pOdin is integrated with its 5' end towards the centromere, in chr 17q23.2 in the first intron of the USP32 gene (NCBI GeneID: 84669). Through use of sequencing primers that begin in the genomic DNA and continue centromeric (until encountering the plasmid rescue enzyme site, then falling back into the plasmid at the 3' plasmid rescue enzyme site), we established that at least 2871 bp of genomic contig remain intact 5' of the pOdin integration site (up to the Bgl II site 5' of pOdin in the genome).

Sequencing of the plasmid from the only colony recovered from an enriched Mfe I plasmid rescue (which cleaves outside pOdin, giving both junctions) suggests a complex integration of pOdin. The 3' end of pOdin is joined to chr 17q25.3 within an intron of MSF (GeneID: 10801). This sequence is nearly 18 Mb distant from the 5' junction, yet also continuing towards the centromere as the sequence exits pOdin (suggesting a large deletion and/or inversion of flanking sequences upon integration).

As such, both junctions of the AO19 integration are regions of chromosome 17 known to be involved in translocations. USP32 (at the 5' junction of pOdin) and TBC1D3 (at 17q12, GeneID: 84218) genes translocate to form a chimeric fusion protein product, the oncogene TRE2 (GeneID: 9098). MSF (at the 3' junction of pOdin) also creates a fusion protein (in acute myelomonocytic leukemia) by translocation with MLL at 11q23. While the recombination rate at the 5' junction is relatively low (0.6 cM/Mb) the rate at the 3' junction in MSF is exceptionally elevated at 3.0 cM/Mb (by UCSC).

#### 3.2.5.3.2 Deletion Events

Of the 16 different DSB misrepair events detected from AO19, 10 were deletions at the I–Sce I site ranging from 1 bp to 875 bp. A complex deletion occurred in one case, which was recovered from 2 different (Bgl II and Mfe I) plasmid rescues and from several colonies: "AO19 B, G4, Gan, 14d, e.2" which showed 0 bp lost from the 5' side of the I–Sce I site, and 235 bp deleted from the 3' cut site. However, following this deletion there are 572 bp (including the beginning of HSV–TK) present, then another deleted region of 1034 bp of HSV–TK and its Poly–A (pos 2900 – 3934 in linearized pOdin). Following this, the sequence continues to read into the remainder of the pOdin plasmid, and the 3' junction with flanking chr 17 sequences is intact. The deletion at the I–Sce I site itself is easily explained, but why the second deletion from within the HSV–TK? There is a shared microhomology of 2 bp at the junction between the 2 sections, but otherwise no indication why the intervening sequences of HSV–TK were deleted in this plasmid.

#### 3.2.5.3.3 Insertion Events

One event of a 28 bp duplication at a cut I–Sce I site (without reconstitution of the site) was observed, as was a small insert of 4 bp. However large insertion events were observed which involved : 1) a fragment of foreign (bacterial) DNA; 2) other regions of the same chromosome (chr 17); and 3) a region of chr 8p12.

##### 3.2.5.3.3.1 Foreign (Bacterial) DNA Insert

One particular Mfe I PR (not enriched with I–Sce I) produced 19/47 colonies which all represented a loss of 3 bp from the 5' end of the DSB, followed by a 2 bp overlap into an

insertion of 660+ bp of *E. coli* DNA. The Apa LI digestion pattern of these plasmids (as exemplified by event "AO19 M, CM, Gan, 14d, n.e.9") suggests that 3329 bp of *E. coli* are inserted at the I–Sce I breaksite, up to the next Mfe I site. This sequence is from a region of the *E. coli* K–12 genome (NCBI Accession: NC_000913) over 2 Mb away from the fragment we detected as a DSB insert in the extrachromosomal assay. Sequencing with the primer "IRPR (R)" failed, indicating the pOdin sequences 3' of the I–Sce I site were absent from the PR plasmid. Hence, the fragment of *E. coli* present at the chromosomal DSB continues past this Mfe I site, indicating the insertion is at least 3329 bp. As with the extrachromosomal DSB assay, the only probable explanation for this event is that some bacterial DNA was transfected into the cells as contaminants of the plasmid preparations. The fact that so many copies of this event were recovered indicates this integration of over 3.3 kb of bacterial DNA at a chromosomal DSB did not prevent the successful proliferation of generations of daughter cells.

#### 3.2.5.3.3.2 Chromosome 17 Repair Events

Two different DSB misrepair events were recovered from AO19 which involved other portions of the same chromosome arm as the pOdin integration site. Event "AO19 B, G4, Gan, 14d, e.10" showed 0 bp were lost 5' of the cut I–Sce I site, and then (with a 3 bp shared overlap) the sequence is of chr 17q23.2 from a region 26 kb centromeric to the 5' integration junction of pOdin, within the second intron of USP32 (whereas pOdin is integrated in the first intron). This inserted sequence begins in a SINE, and continues towards the telomere, hence towards the point where pOdin is integrated (assuming no rearrangements in the chromosome). Based on restriction digests, and sequencing across the 5' genomic Bgl II site, this inserted sequence continues for at least 2318 bp. The recombination rate of this entire area is very low (0.6 cM/Mb), with one region of very high MAR potential (12 kb telomeric to the beginning of the insert, and 14 kb centromeric to the 5' end of pOdin) located in the middle of these two points but nowhere else in the surrounding 100 kb.

The second insertion involving chromosome 17 (event "AO19 B, CM, Gan, 14d, e.9") also involves a region of 17q23.2 centromeric to the pOdin integration, but this time beginning 1.5 Mb away and continuing centromeric (away from pOdin). At least 340 bp of this region are present at the DSB, beginning in a LINE from within a large intron of PPM1E (GeneID:

22843). Again the local recombination rate is low (0.8 cM/Mb), and there are no areas of high MAR potential closer than 40 kb away.

We have established that the integration of pOdin in chr 17q23 occurred at a major rearrangement between the chromosomal areas 5' and 3' of the plasmid. Thus, there may be no particular reason why these 2 insertion events would employ particular regions of chr 17q23 in a DSB repair event. It could simply be the result of further instability or rearrangement within this chromosome, as captured during repair of a DSB. However, analysis of the exact junction between pOdin and the PPM1E region revealed a near–perfect I–Sce I recognition sequence occurs at that point in the genome. There are 0 bp degraded from the cut I–Sce I site in pOdin, followed by 18 bp of intervening sequence which match satellite DNA (most closely to chr 19q12). Regardless of this small intervening sequence, after this follows the beginning of the chr 17q23 PPM1E sequences which are nearly the junction of another cut I–Sce I site. In fact, the genomic sequences which would precede the beginning of the insert are an exact match to the I–Sce I site, giving a total match of 15/18 bp. In this case, it may have been that expression of I–Sce I resulted not only in cleavage within pOdin, but also of this region 1.5 Mb away, such that two chromosomal DSBs were induced in the same cell. This would greatly increase the chance of this area being broken elsewhere (to create a fragment for DSB insertion), or resolving itself by end–joining and creating a translocation with the pOdin DSB.

#### 3.2.5.3.3.3 Chromosome 8 Insert Event

A number of colonies from different plasmid rescues all produced an event (exemplified by "AO19 M, CM, Gan, 14d, n.e.21") whereby over 3.1 kb of chr 8p12 have inserted at the DSB. The junction is not within a gene or repetitive element, continues towards the telomere, and has a low local recombination rate (0.6 cM/Mb). Of note is that 25 kb from the 8p12 inserted region is the beginning of the NRG1 gene (neuregulin–1, GeneID: 3084). The junction shows 0 bp lost from the cut I–Sce I site, then 2 bp of shared microhomology (overlap) into 8p12, at a point in the genomic DNA where 16/18 bp match the I–Sce I consensus sequence.

This event thus illustrates that portions of other chromosomes can be used in the repair of chromosomal DSBs in human cells. Again, the occurrence of the break junction within a I–Sce I concensus–like sequence strongly implicates DSB induction at other sites as a determining factor in their involvement. NRG1 is found to suffer chromosome breaks in breast and pancreatic cancers, and was recently classified as a novel aphidicolin–induceable fragile site [175]. This region of 8p12 is frequently altered in carcinomas, with translocations resulting in deletion of distal p–arm [176]. The breakpoints are clustered in a small region, with complex rearrangements such as amplifications and deletions of chr 8 often observed [177].

In mapping the 8p breakpoints of 3 breast carcinoma cell lines, derivative chromosomes have been detected with 17q22–25 joined to 8p12 [177]. The event we observe, of DSB repair between our initial DSB site (at 17q23.2) and the endogenous I–Sce I site at 8p12, very near NRG1, provides a good illustration of recurrent chromosomal rearrangements (such as translocations) resulting from concurrent breaks in the genome. It remains to be determined if our event resulted in a simple DSB insertion or some other recombination outcome, such as a derivative chromosome.

#### 3.2.5.3.4 Endogenous I–Sce I Sites in the Human Genome

To our knowledge, this is the first report of induction of I–Sce I cuts at endogenous sites in human chromosomal DNA. The introduction of an artificial sequence (such as a plasmid) into a genome can alter the local chromatin configuration, as is often observed with transgene silencing [178]. As such, it could be argued that repair of a DSB within an integrated plasmid (a foreign DNA) might not be entirely representative of what goes on at a natural site within the genome. In these two cases we report, while the DSB is still induced by a restriction enzyme, the fact it occurs at an endogenous sequence answers much of the criticism which often surrounds DSB repair assays.

We determined if any of the other DSB repair events we observed might have been the result of cleavage at endogenous I–Sce I–like sites. We verified if any other sites in the human genome corresponded to the intact I–Sce I site, or these 2 pseudo–sites which cleaved *in vivo*. As mentioned earlier, the exact 18 bp I–Sce I recognition sequence is not found

anywhere in the human genome. Interestingly, the pseudo–sites we observed cutting at only occur once each in the genome. The results of our search, for all sites matching 15/18 bp (or better) to I–Sce I are listed in Table C–F.

Why would the degenerate 15/18 bp site from 17q23.2, with the first 3/4 bp mismatched, have cut in AO19 when other sites show equal or better matches to the concensus I–Sce I? Both pseudo–sites have their mismatches concentrated towards the ends of the restriction sequence, away from the actual cleavage point, but no more so than most other degenerate sites. We cannot exclude that other endogenous sites may have cut, and not been detected, due to our limited data set. While the 17q23.2 event occurred in an area of relatively low local recombination, it is within a large intron of the PPM1E gene, and as such may have been prone to a more open chromatin conformation, thus more accessible to the restriction enzyme, than some of the other pseudo–sites.

**Table C–F. Distribution and characteristics of I–Sce I–like sites in the human genome**

| SEQUENCE | CHROMOSOME | NCBI CONTIG | POSITION | GENE | REPEAT | cM/Mb |
|---|---|---|---|---|---|---|
| attaccctgttatcccta | I–Sce I site | | | | | |
| **17/18 bp** | | | | | | |
| attaccctgttatccct**C** | 13q12.11 | ref\|NT_024524.13\|Hs13_24680 | 3774213 – 3774196 | – | – | 4.3 |
| attaccctgtt**T**tccta | 7p15.3 | ref\|NT_007819.14\|Hs7_7976 | 23101956 – 23101939 | STK31 | – | 0.9 |
| **16/18 bp** | | | | | | |
| attaccctgttatccc**CC** | 1p13.1 | ref\|NT_004754.15\|Hs1_4911 | 961160 – 961143 | – | – | 1.6 |
| a**C**taccctgttatccc**C**a | 1q23.1 | ref\|NT_079484.1\|Hs1_79549 | 3258801 – 3258784 | – | – | 1.0 |
| attaccctgttatc**AA**ta | 2p11.2 | ref\|NT_022184.13\|Hs2_22340 | 67002455 – 67002438 | – | – | 0.8 |
| attaccctgttatcc**TC**a | 2q22.2 | ref\|NT_005058.14\|Hs2_5215 | 11421926 – 11421909 | LRP1B | – | 1.0 |
| attaccctgttatcc**T**t**T** | 5q14.1 | ref\|NT_006713.13\|Hs5_6870 | 9409210 – 9409227 | MSH3 | – | 1.6 |
| **C**ttaccctgttatccct**C** | 5q14.3 | ref\|NT_023148.12\|Hs5_23304 | 496563 – 496580 | – | – | 0.7 |
| att**G**ccctgttatccc**A**a | * 8p12 * | ref\|NT_007995.13\|Hs8_8152 | 1793001 – 1793018 | – | – | 0.6 |
| at**GT**ccctgttatcccta | 13q14.3 | ref\|NT_024524.13\|Hs13_24680 | 34192331 – 34192348 | – | – | 0.7 |
| at**G**accctgttatccct**C** | 16q22.3 | ref\|NT_024797.14\|Hs16_24953 | 853618 – 853601 | – | – | 2.4 |
| attaccctgttatccc**GT** | 18q21.32 | ref\|NT_025028.13\|Hs18_25184 | 6694501 – 6694518 | – | – | 1.6 |
| attaccctgttatcc**TG**a | 22q11.21 | ref\|NT_011519.10\|Hs22_11676 | 1573105 – 1573088 | – | MIRb (SINE) | 3.0 |
| attaccctgttatc**AA**ta | Xq27.3 | ref\|NT_011786.14\|HsX_11943 | 18728927 – 18728910 | – | – | 3.0 |
| attaccctgttatcc**TG**a | Yq11.221 | ref\|NT_011875.10\|HsY_12032 | 4022622 – 4022605 | – | ERVK (LTR) | – |
| **15/18 bp** | | | | | | |
| a**GC**accctgttatccc**A**a | 1p22.1 | ref\|NT_028050.13\|Hs1_28209 | 2422378 – 2422395 | – | – | 0.8 |
| attaccctgttatc**A**c**GT** | 2q11.2 | ref\|NT_022171.13\|Hs2_22327 | 580100 – 580117 | *D87446* | L1M (LINE) | 0.7 |
| attaccctgttatc**AAA**a | 3q29 | ref\|NT_029928.11\|Hs3_30183 | 1547092 – 1547109 | DLG1 | L1 (LINE) | 2.0 |
| **CC**t**G**ccctgttatcccta | 3q12.1 | ref\|NT_005612.14\|Hs3_5769 | 4716573 – 4716556 | *AF116699* | – | 0.5 |
| a**AtG**ccctgttatccct**C** | 5p15.31 | ref\|NT_023089.13\|Hs5_23245 | 7855465 – 7855482 | MTRR | – | 2.0 |
| attaccctgttatc**AAA**a | 5p14.2 | ref\|NT_006576.14\|Hs5_6733 | 6392191 – 6392174 | – | LIMA3 (LINE) | 1.0 |
| a**AG**accctgttatccct**C** | 6p25.2 | ref\|NT_034880.3\|Hs6_35042 | 3954928 – 3954945 | – | AluJo (SINE) | 2.6 |
| a**CAG**ccctgttatcccta | 6p21.1 | ref\|NT_007592.13\|Hs6_7749 | 32793638 – 32793621 | – | MIRb (SINE) | 1.7 |
| **C**ttaccctgttatccc**CC** | 6q14.1 | ref\|NT_007299.12\|Hs6_7456 | 14493596 – 14493613 | IMPG1 | L1M4 (LINE) | 0.6 |
| attaccctgttatc**TTtG** | 6q23.1 | ref\|NT_025741.13\|Hs6_25897 | 34974430 – 34974447 | – | – | 1.1 |
| **T**ttaccctgttatcc**AtG** | 7q21.13 | ref\|NT_007933.13\|Hs7_8090 | 13531698 – 13531681 | – | – | 0.4 |

**Table C–F (continued). Distribution and characteristics of I–Sce I–like sites in the human genome**

| SEQUENCE 15/18 bp (continued) | CHROMOSOME | NCBI CONTIG | POSITION | GENE | REPEAT | cM/Mb |
|---|---|---|---|---|---|---|
| attaccctgttatc**TTtC** | 7q32.1 | ref\|NT_007933.13\|Hs7_8090 | 52470381 – 52470398 | SND1 | – | 0.4 |
| **G**ttaccctgttatccc**AC** | 8q24.21 | ref\|NT_008046.14\|Hs8_8203 | 43605058 – 43605075 | *MGC27434* | L2 (LINE) | 0.9 |
| attaccctgttatc**TTA**a | 9q31.1 | ref\|NT_008470.16\|Hs9_8627 | 9345477 – 9345494 | ABCA1 | – | 1.5 |
| **T**ttaccctgttatcc**AG**a | 10q25.1 | ref\|NT_030059.11\|Hs10_30314 | 28740989 – 28740972 | – | – | 0.5 |
| **TtG**ccctgttatccc**A**a | 11p15.1 | ref\|NT_009237.16\|Hs11_9394 | 17467007 – 17467024 | – | – | 1.0 |
| **GtG**accctgttatccct**G** | 11q12.1 | ref\|NT_033903.6\|Hs11_34058 | 2910311 – 2910294 | – | – | 0.1 |
| attaccctgttatc**TTtC** | 11q21 | ref\|NT_008984.16\|Hs11_9141 | 8459714 – 8459731 | – | – | 2.0 |
| attaccctgttatc**TTC**a | 12p12.2 | ref\|NT_009714.16\|Hs12_9871 | 13129611 – 13129594 | – | – | 1.6 |
| attaccctgttatcc**AGG** | 12p12.1 | ref\|NT_009714.16\|Hs12_9871 | 15008693 – 15008710 | – | L1PA13 (LINE) | 0.9 |
| attaccctgttatcc**AAG** | 13q32.3 | ref\|NT_009952.14\|Hs13_10109 | 14365067 – 14365084 | *FLJ14624* | – | 1.8 |
| **T**ttaccctgttatccc**AG** | 14q21.1 | ref\|NT_026437.10\|Hs14_26604 | 23055060 – 23055043 | – | AluSq (SINE) | 0.4 |
| **T**ttaccctgttatcc**TG**a | 15q15.1 | ref\|NT_010194.16\|Hs15_10351 | 13544591 – 13544574 | – | – | 0.3 |
| **TC**taccctgttatccct**T** | 17q11.2 | ref\|NT_010799.14\|Hs17_10956 | 6035256 – 6035273 | *AY358653* | HERVP71A (LTR) | 1.8 |
| **TAtG**ccctgttatcccta | * 17q23.2 * | ref\|NT_010783.14\|Hs17_10940 | 15587540 – 15587557 | PRMT5* | L1MC4 (LINE) | 0.6* |
| a**AtG**ccctgttatccct**C** | 22q11.1 | ref\|NT_011519.10\|Hs22_11676 | 681696 – 681649 | – | – | 0.2 |

Table C–F. Distribution and characteristics of I–Sce I–like sites in the human genome. Concensus I–Sce I site (18 bp) is listed in lower case letters, with mismatches in bold capitals. BLAST HUMAN GENOME was performed at NCBI (removing the MEGABLAST and FILTER features, and raising the EXPECT value to 10, to maximize the number of near matches). All hits to the human genome matching 15/18 bp (or better) to the concensus I–Sce I site are listed, verified and characterized using the UCSC Genome Browser (July 2003 Assembly). The list of resulting hits was the same regardless of orientation or strand of the I–Sce I 18 bp site input. "GENE" refers to the presence of the listed site in a known gene, all of which were within introns, with predicted genes listed in italics. "REPEAT" refers to occurrence of the site within a repetitive element, as per UCSC. "cM/Mb" refers to the local recombination rate within that genomic region, as per UCSC. 2 of the hits are listed flanked by asterix, *8p12* and *17q23.2*, which are the 2 sites found to have cleaved during *in vivo* DSB repair events in clone AO19.

## *3.3 Summary: DSB Inserts (Extrachromosomal and Chromosomal)*

We have demonstrated the development and use of systems which detected the insertion of DNA fragments during double–strand break (DSB) repair in human cells, at both extrachromosomal and chromosomal sites. Of importance is the fact that both these systems apply methods other than PCR amplification of a repaired break site, to recover events of insertion at the DSB. As such, both extrachromosomally and chromosomally, we have observed the insertion of kilobases of DNA at a single DSB – events which could not be detected by our analyses with PCR.

### 3.3.1 Frequency of DSB Inserts

In the extrachromosomal assay, we determined that DSB misrepair is much less frequent in normal cells (4% of all repair events) than in AT cells (20% of repair events), but that overall approximately 0.5% – 1.0% of all DSB repair events from both cell lines result in an insertion. In the analysis of chromosomal DSBs, 2099 colonies were analyzed to reveal a total of 22 DSB misrepair events. We cannot, however, accurately establish the frequency of DSB insertional repair at chromosomal sites, the same way we could with extrachromosomal DSBs. This is because we cannot with any certainty establish the frequency of DSB induction with the chromosomal assay, whereas we knew the majority of our input plasmids in the extrachromosomal assay were linearized (and thus represented near 100% DSB efficiency).

As with the extrachromosomal assay, we did find evidence of human chromosomal DNA inserts in the chromosomal assay, though arguably these could also represent reciprocal translocation events. The plasmid rescue technique requires the inserted sequence be cleaved to provide a sequence small enough to replicate as a plasmid. Hence recovery of the plasmid sequences 3' of the DSB site is often impossible, such that we cannot distinguish the insertion of a large lone DNA fragment (which would retain the 3' plasmid sequences and flanking chromosome) from a translocation event which could completely remove the 3' plasmid and chromosome. This phenomenon calls into question the definition of "translocation" somewhat. However, we did detect an event of *E. coli* DNA insertion (from a

region of the K–12 genome distinct from that of the extrachromosomal insert), which is obviously NOT a translocation. And, we detected multiple copies, which (since it cannot replicate autonomously) must have occurred early enough in the cell population to generate multiple clonal cells (and hence, cannot be an artefact of bacterial transformation). This demonstrates that DSB insertional repair can occur at chromosomal sites, and can involve foreign DNA (in this case, most likely a transfected rather than an infectious DNA fragment). These inserts can be at least 3.3 kb in size (as was the *E. coli* insert), while the extrachromosomal assay indicated they could be at least 18 kb.

While the extrachromosomal assay performed in normal cells showed no evidence of human DNA inserting alone at DSBs (without flanking SV40 DNA), the one DSB repair event from clone GZ1 showed chromosomal DSBs in normal cells can be repaired with portions of other chromosomes. However, as with the increased frequency of extrachromosomal DSB misrepair in AT cells, most of the chromosomal DSB misrepair events we isolated also came from those cells. The increased genetic instability and chromosomal breakages of AT cells could account for the greater number of inserts, also, from both chromosomal and extrachromosomal DSB assays. These DSB misrepairs, even those which result in portions of other chromosomes or foreign DNA inserting at the DSB, can restructure the genome of these cells while not proving fatal to the cell. This reveals that DSBs in human cells can be repaired by the uptake of other DNA fragments.

# 4. CONCLUSIONS AND DISCUSSION

## *4.1 Overview*

The objectives of this thesis work were to establish if DNA double–strand breaks (DSBs) could capture other DNA fragments during repair in human cells, and to characterize the sources of any inserted material. To do this we developed two experimental systems, which were applied primarily in both normal human fibroblasts and in cells deficient for the ATM checkpoint and repair kinase. The first system followed the repair of an extrachromosomal DSB, as represented by the recircularization of a linear plasmid, and the analysis of plasmids which were repaired with an increase in size. The second system introduced DSBs into genomic DNA, by integrating a plasmid harbouring the recognition sequence for the rare–cutting I–Sce I endonuclease, and then recovering and analyzing plasmids from cells which had been exposed to I–Sce I expression.

The availability of the human genome sequence enabled us to classify a number of events of DNA insertion during plasmid recircularization (in the extrachromosomal DSB assay), events which had previously fallen into an uncharacterizeable category in plasmid end–joining studies. Both normal and ATM–deficient (AT) cells had the capacity to insert transfected DNA into extrachromosomal DSBs. The AT cells also displayed insertions of human genomic DNA up to 18 kb in size (the probable upper limit of the assay's ability to detect due to bacterial replication restrictions on plasmid size). These human genomic inserts displayed no obvious preferential use of chromosomal regions, repetitive DNA elements, or nucleotide content, although some did span exons and introns demonstrating the capacity to shuffle gene segments into DSBs.

Interestingly, both cell lines were immortalized with SV40 (integrated into the cellular genome), and both made use of SV40 DNA as insert material. Some genomic DNA flanking the integration sites of SV40 also integrated (along with the SV40 DNA) into extrachromosomal DSBs in both normal and AT cells. The normal cells, which harboured a variant episomal SV40 DNA in addition to the integrated form, also made use of this extrachromosomal viral DNA for insertions into extrachromosomal DSBs. Some repair

events from both cell lines had inserted the SV40 origin of replication (ori) sequences, which conferred the ability for these plasmids to replicate in cells expressing Large T–antigen (T–ag). In cases where flanking genomic DNA was present, this constituted an amplification event of chromosomal sequences in the presence of the SV40 ori. In addition, some of these insertions spanned the T–ag coding sequences, enabling the plasmids to replicate in cells which did not previously express T–ag, demonstrating material inserted at extrachromosomal DSBs could also be expressed as functional proteins.

In the context of chromosomal DSBs, our second assay demonstrated events of insertion in both normal and AT cells using transfected or chromosomal DNA in excess of 3 kb, some of which also spanned gene sequences. Surprisingly, we detected two events of DSB repair where the junctions were degenerate I–Sce I recognition sequences in chromosomal DNA, indicating some naturally occurring sites in the human genome will be cut upon exposure to the I–Sce I endonuclease. These events also demonstrated that multiple DSBs in human chromosomes are co–localized during repair.

As a mechanism of DNA repair previously ignored or poorly understood, our demonstration of DSB insertional repair in human cells presents a number of intriguing possibilities for future research, as well as raising some issues of concern for the application of genetic therapies. In the following sections, we discuss how the results of this thesis work compare with previous studies on DSB repair, and reiterate how our methodology encompasses repair events previous groups would have failed to detect. We discuss factors affecting the outcome of our DSB insertional repair events, with particular attention to the influence of ATM. We outline how our work contributes to a better understanding of genome content and evolution, repetitive DNA, symbiosis, infection, horizontal gene transfer, tumour progression and metastasis. We present this mechanism as a means to bridge gaps in genomic DNA which could confer a selective advantage upon the cell. In addition, our discovery that SV40 can amplify genetic material for insertion at DSBs, and that it can integrate into a region associated with a number of genes implicated in SV40–positive cancers, is discussed.

## 4.2 DNA End–joining Assays and DSB Inserts

Our results indicate that normal cells display the ability to insert foreign DNA into an extrachromosomal break, but only in a situation of genomic instability (as in AT) are portions of chromosomes seen to shuffle into extrachromosomal break sites. Interestingly, if both cell types are subjected to the same kind of local instability, as is the case at sites of SV40 integration, then both display insertions of viral:human junction DNA. The limitation therefore is not in the ability of cells to shuffle DNA and repair a break, only the difference in the generation of those fragments afforded by different genetic backgrounds.

A number of previous studies have used plasmid–based DSB assays to assess differences in the efficiency and fidelity of DNA repair in different genetic backgrounds [121,137,138,179–184]. As with our study, many of these characterized repair using cell extracts or human cells derived from patients with AT. Our results confirm this previous work, as we observed both normal and AT cells are efficient at repairing DSBs, but with a 5–fold increase in misrepair events from AT cells. These are primarily deletions at the break site, or insertions of several base pairs (which have been characterized by the previous groups, and which we identified by restriction digests on agarose gel but did not pursue further). Our results also extend these findings, however, by drawing upon the recent wealth of reference sequence data to specifically identify and characterize the source of larger rearrangement and insertion events.

The repair of linear plasmids, accompanied by major rearrangements which could have been insertion events (as evidenced by altered sizes on agarose gel) has been reported previously in normal and AT cells [179,182,184]. These studies primarily assessed the loss of bases at junctions, and the nature of the major rearrangements was not well characterized, perhaps because many were performed prior to the affordable sequencing methods or wealth of sequence data now available. The extent of insert identification was limited to the WI38–VA13 and WI26–VA4 normal human fibroblast cell lines (transformed with SV40), and mentions only that insertions of 3 kb or greater were identified as: one insertion of "an Alu family repetitive sequence"; two insertions of "partial SV40 sequences" (with no sequence details); and "other" insertions made up of portions of the plasmid template [184]. In contrast,

our study identified a large number of insertions in both normal and AT cells (31 and 33, respectively), the majority of which could be sequenced and their source identified. We found insertions of our plasmid template, and insertions of SV40 plus flanking genomic DNA (from the integration sites) in both cell lines. We found 13 distinct insertion events which used only SV40 DNA in normal cells, spanning many portions of the SV40 genome, but no such lone SV40 inserts in AT cells. Conversely, we identified 13 distinct events of human chromosomal DNA insertions in AT cells which could be precisely localized to various regions of the genome (and measuring up to 18 kb long), but no such lone chromosomal inserts in normal cells.

During recircularization of plasmids in a variety of normal human cell lines, the frequency of all reported insertions (including small changes not analyzed by our methods) varies from 13% – 53% of mutants, or 3.8% – 15.37% of all repair events [182,184]. Our results from normal cells show large inserts alone accounted for 23.48% of (unique) mutants, but that this represented only 0.86% of all DSB repair events. Since our method included only those inserts easily distinguishable on gel (no smaller than 200 bp), this would explain why our frequencies for large inserts are lower than those reported previously, which accounted for all sizes of insertions. This highlights that the limitations our assay places on the size of insert which can be detected (200 bp to approximately 18 kb) ignores smaller insertion events, and possibly larger ones as well, and therefore under–estimates the frequency at which DSB insertional repair occurs in human cells.

In companion papers, published during the course of our work, the Waldman lab described the first specific survey of DNA inserts at DSBs in mammalian cells, using mouse fibroblasts [98,116]. They used PCR to amplify the break site (also an I–Sce I induced DSB), and found inserts in 8% of the mutants they studied. They first recorded 21 cases of inserts, 19 of which ranged from 140 bp to 490 bp, whereby they principally captured transfected DNA (their I–Sce I expression plasmid), as well as a few genomic fragments which appeared to be derived fom retrotransposon activity, although were difficult to precisely identify as the mouse genome was incompletely sequenced at the time. They also noted one exceptionally large insertion of 3351 bp, derived from the *E. coli* genome presumably as a contaminant of

their plasmid preparations (much as we detected in both the extrachromosomal and chromosomal assays).

There was also a competition effect noted between different classes of DSB inserts. When transfected alone or in combination with circular DNA, the I–Sce I expression plasmid inserted roughly as frequently as any other sequence, including endogenous DNAs [116]. However, co–transfection of the I–Sce I expression plasmid with linear DNA resulted almost exclusively in insertion of the linear DNA, this time for a total of 13% of their DSB misrepair events characterized [98]. This preference for linear DNA as a DSB insert reinforces the idea that the insertions we observed in human cells result from the capture of free DNA ends.

This also leads to a suggestion that, in our study, the presence of episomal SV40 in the normal cells could potentially have led to competition with other DNAs for DSB capture, and that this was the reason we did not observe lone chromosomal inserts in the normal cells. However, the fact that the other categories of capture events (using pBluescript, and SV40 + genomic DNA) occurred at relatively the same proportions in both cell lines leads us to conclude that a feature other than the presence of episomal SV40 in the normal cell line is responsible for the lack of large genomic DNA fragments being used during DSB repair.

### 4.2.1 ATM Deficiency: Less Repair, or More Breaks?

A fundamental question of our ATM studies is: are the chromosomal DSB inserts we observe only in AT cells the result of reduced repair capacity, or increased genomic instability? The aforementioned end–joining assays, and our own results, indicate AT cells are capable of repairing DSBs as efficiently as normal cells do, but that that there is a lack of repair fidelity which results most often in increased deletions at the repaired DSB. We observe nearly equal overall frequencies of DSB insertion events between normal and AT cells. It is only in the source of insert material that DSB repair differs between these two cell lines. A study employing the same two cell lines as used in our study, normal (GM00637) and AT (GM05849), found that chromosomal instability was 2.6–fold higher in the AT cells,

as evidenced by the spontaneous formation of micronuclei [182]. We suggest that genomic instability, like that inherent in AT cells, can be responsible for generating chromosomal fragments which are then subject to insertion at sites of DSB repair in human cells.

We now direct this discussion towards understanding what characteristics of a genomic region might possibly influence the provision of fragments for DSB insertional repair.

## *4.3 Features of Genomic Insert Sources*

Our results demonstrate the insert process to be a reality of DSB repair in human cell culture assays. We were able to analyze a number of characteristics of the DNA fragments used during DSB insertional repair, such as their nucleotide content, their probability of associating with a nuclear matrix, their localization relative to regions prone to breakage or reportedly involved in genome rearrangements (see Results). However the mechanism which presents a given segment of the genome as a candidate for insertion is uncertain. We now address several features known to create instability and breakage in the genome. We begin with a structure in our DNA that faces loss and addition of sequences regularly: chromosome ends, or telomeres.

### 4.3.1 Telomeres, ATM and DSB Inserts

Telomeres are the repetitive DNA sequences (tandem repeats of "TTAGGG") which cap and protect the ends of eukaryotic chromosomes, distinguishing them from other DSBs. The terminal sequences are progressively lost cell generation after generation as a consequence of replication, shortening the telomeres [185]. They can be replenished or maintained by the activity of telomerase (a reverse transcriptase), or through an alternative homologous recombination mechanism (ALT) with other telomeric sequences [186,187].

ATM was thought to play a role in telomere maintenance, as increased incidences of telomere shortening and chromosome end–fusions were observed in cells of AT patients, but this was only seen in lymphocytes [188]. Previous studies have described the reduced size of telomeres in our AT cell line (GM05849) compared to our normal cell line (GM00637). The

normal cells had exceptionally long telomeres (despite being telomerase–negative), while the AT cells did express telomerase, and maintained sufficently long telomeres within the range of other normal cells [189,190]. Interstitial insertions of telomeric DNA have also been reported at sites of translocation breakpoints [191]. However, we analyzed and found no evidence of telomeric sequences in any of our DSB inserts, supporting the idea that ATM–deficiency does not affect the outcome of DSB insertional repair through any increased exposure to DNA ends brought about by abnormal telomere loss.

### 4.3.2 Gene Amplification and Fragile Sites

The generation of multiple copies of a DNA sequence is one of the most common manifestations of genomic instability seen in tumour cells [192]. Models of gene amplification often invoke breaking chromosomes to resolve amplified structures, making it a feature of the genome which is a good candidate for producing DSB insert fragments.

Some of the earliest studies characterizing gene amplifications in mammalian cells, including some performed in our lab, assessed the activation of replication from integrated polyoma viruses, and their ability to extend replication into adjacent genomic sequences [193,194]. DNA replication is a strictly coordinated process which is initiated from specific origins in human chromosomes, and which are normally limited to acting or "firing" only once per cell cycle [195]. The generation of amplified structures, either within the chromosome or as extrachromosomal molecules, is frequently observed in human tumours, often associated with elevated levels of gene expression, and is believed to initiate via chromosome breakage or deregulated activity of nascent replication origins [145,150,196–198]. Multiple rounds of bidirectional replication origin firing leads to exponential copying on–site, called "onion skin" replication, which must then be resolved by excision and/or homologous recombination to produce tandem arrays [194,199].

Gene amplifications can be selected for *in vitro* and *in vivo* by treatment of cells with chemicals requiring increases in gene copy number (and corresponding expression) to overcome metabolic toxicity. One classic example is exposure to methotrexate, an inhibitor

of DHFR enzymatic activity, which can be used to select for cells which have amplified sufficient copies of the DHFR gene to overcome its toxicity. In neuroblastomas, the original copy of MYCN remains at its location on 2p23 while amplifications of as many as hundreds of copies can be seen in large tandem arrays as: 1) separate extrachromosomal structures called double minutes (DMs); or 2) integrated into various other loci as homogeneously staining regions (HSRs) [200,201].

The generation of multiple copies of a proto–oncogene, such as MYCN can lead to over–expression of normal protein products as a step in the tumorigenic process. Interestingly, amplification has yet to be observed in normal cells, and even cells with deficiencies in DNA repair proteins (such ATM or DNA–PKcs) show little evidence of gene amplification without first being immortalized or transformed [202–204].

One frequent observation accompanying MYCN amplification is an associated increase in copy–number and expression of the MDM2 gene, originally located on another chromosome at 12q13 [205]. Use of an I–Sce I site to induce DSBs within multiple copies of a DHFR amplicon (a HSR) led to the formation of DMs in hamster cells [206]. This and other evidence suggests both DMs and HSRs result from DNA breaks, and are most likely different forms of the same process [197,207].

The induction of breaks at "fragile sites" (see below) may be one mechanism that instigates gene amplifications, or establishes the boundaries of an amplicon [208]. Once an amplification process has begun, these sequences can remain onsite to form a HSR, which then through breakage can become extrachromosomal DMs. This process would create material for use in DSB insertional repair, where the amplifying array can insert elsewhere in the genome at a DSB. This model would explain why many HSRs are non–syntenic, existing at sites other than the original gene locus, which remains intact.

An alternative mechanism for gene amplification is the breakage/fusion/bridge (B/F/B) cycle, first descibed in maize by Barbara McClintock in the 1940's, and only much later demonstrated to occur in rodent and human cells [209,210]. This can result from the loss of a

telomere before or during S–phase, followed by fusion of the exposed ends of the identical sister chromatids. This generates a single dicentric chromomsome which is an inverted repeat, a mirror image at the fusion site. During cell division, the two centromeres are pulled to opposite poles, exposing new broken ends. Provided the shearing does not break the chromosome in precisely the same spot as before, this results in one daughter cell receiving three copies of a particular chromosomal region, while the other daughter receives only one. This model has been used to explain the amplification of the MET oncogene in human gastric carcinoma [211]. This process has also been induced by creating a DSB (with I–Sce I) near telomeres in mouse ES cells, exposing chromatid ends which then become substrates for fusion [212].

In contrast to the formation of DMs and HSRs, where the single–copy original gene remains intact and the amplification is seen elsewhere (either extrachromosomally as a DM, or chromosomally as non–syntenic HSRs), B/F/B amplifications occur on–site as tandem inversions alongside the original copy, and generally cause lower copy–number amplifications [208]. This implicates DSB insertional repair as a major mechanism in the spread of amplified structures throughout the genome, by capturing DMs at break sites, and places DSB insertional repair as one possible step in the progression of tumorigenesis.

We found limited evidence that inserts derive from areas of the genome which may be subject to amplification, such as the case of the extrachromosomal insert "A2–62" in CNOT2 at 12q15. In the chromosomal assay, the chromosome 8 insert event in clone AO19 ("AO19 M, CM, Gan, 14d, n.e.21"), the result of cleavage at an endogenous I–Sce I site, is very near NRG1 which can also become amplified in some cases. However, we have no evidence these regions were amplified in our cells, and as such cannot truly determine if purely chromosomal amplifications play a role in DSB insertional repair. We do, though, have solid evidence of DSB inserts and amplification of chromosomal DNA associated with the SV40 origin (see below).

#### *4.3.2.1 ATM and Amplification*

Previous studies have addressed the specific involvement of ATM in gene amplification. An analysis of the same two cell lines used in our study (GM00637 and GM05849) revealed that amplification of the DHFR gene (by treatment with methotrexate) occurs at the same rate in both. In another system, it was confirmed that loss of ATM alone does not increase gene amplification [204]. These findings are relevant to our results, as we found equal evidence of gene amplification associated with our viral:human DSB inserts from both cell lines. Our normal and AT cells both inserted fragments of SV40 + chromosomal DNA (from the integration sites), and both cell lines showed evidence of replication of repaired plasmids which spanned the SV40 origin, in the presence of T–ag.

The GM05849 cells have been reported to exhibit only one round of replication per cell cycle, like normal cells, following S–phase block with hydroxyurea [213]. This might suggest that the large chromosomal fragments we observed inserting at DSBs in AT cells were not the result of deregulated human origins. However, if a cell is immortalized, and already permissive for gene amplification, then loss of ATM has been shown to increase the probability of a gene amplification event [204]. Since our cells are immortalized already with SV40, we cannot definitively establish whether our inserted material is generated by chromosome breakage alone, or breakage in conjunction with amplification.

#### *4.3.2.2 Fragile Sites, ATM and DSB Inserts*

As analyzed in the Results section, "fragile sites" are regions of the genome which can be induced to break, when a cell is cultured under particular conditions. Fragile sites have been shown to be sites for the preferential integration of transfected DNA in mammalian cells [214]. We found limited evidence for the involvement of fragile sites in generating fragments for use in DSB insertional repair. The one obvious instance was the extrachromosomal insertion event "A3C–72", from FHIT (but outside the region most reported as breaking in the fragile site FRA3B – see Results). From the chromosomal assay, the only possible evidence involves the same repair event in clone AO19 as was mentioned in the amplification section ("AO19 M, CM, Gan, 14d, n.e.21"), involving NRG1 which was recently classified as a novel aphidicolin–induceable fragile site [175].

Comparisons of fragile sites in cells deficient for ATM or ATR found that AT cells have the same degree of fragile site expression as normal cells (i.e. that ATM is not required to maintain fragile site stability) [215]. This provides strong evidence to suggest the source of our DSB inserts is not the result of increased breakage due to some induction of fragile site expression. As was the case with the gene amplification studies, we cannot determine if any fragile sites were induced in our cells at the time of our experiments, and hence do not know to what extent they contribute to providing source material to DSB insertional repair in other situations.

All three features explored here (telomeres, gene amplification, and fragile sites) show no evidence of generating fragments which inserted at DSBs in our assays. Since the characteristics of all three are reported to be the same between normal and AT cells, this suggests some other feature is responsible for the difference in DSB inserts we observe in AT cells (whereby large chromosomal fragments were inserted into extrachromosomal DSBs).

We now turn our attention to another classic feature of genomic instability, one which could possibly be classified as DSB insertional repair (or *vice versa*): chromosome translocations. In conjunction with studies of DNA integration, from laboratory transfections and gene targeting work, translocations provide sound evidence that DSBs in human cells are brought together to undergo repair, which we then discuss in the context of a possible mechanism for DSB insertional repair.

### 4.3.3 DSB Inserts and Translocations

The physical exchange of large chromosomal regions is traditionally classified as a translocation, though it is a term which applies to much larger stretches of DNA than we have characterized as DSB inserts. Two chromosomes which break and exchange large segments, such as whole chromosome arms, undergo a "reciprocal translocation", whereas others display the transfer of such large regions to new derivative chromosomes in a non–reciprocal manner. Another classification which is used to describe cytogenetic

abnormalities in which two or more recipient chromosomes acquire a portion of a single donor chromosome is a "jumping translocation" (JT). Though few cases of JTs have been reported, they include patients with AT, as well as SV40–transformed human fibroblasts [216–218]. The breakpoints of JTs reportedly coincide with fragile sites and known hotspots of viral integration, and are implicated in cases of oncogene amplification [219]. The presence of telomeric sequences at the junctions of JTs has led to the proposal by some that repetitive DNA is involved in the mechanism of donor and recipient site determination [191].

We have already discussed fragile sites, telomeric sequences, and gene amplification in the context of DSB inserts. This, and the characteristics of JTs involving two or more recipient sites, suggests we are investigating a different phenomenon when describing DSB insertional repair. However, translocations in general present an interesting paradigm for how genome organization influences DSB repair.

The role that DSBs play in causing translocations is widely suspected, but sparsely investigated. Indirect evidence in rodent cells suggests bleomycin–induced DSBs may induce translocations [220]. Companion studies by Maria Jasin's group in mouse ES cells give evidence of recurrent translocations induced by DSBs in mammalian cells. While the repair of a single I–Sce I induced DSB was carried out by homologous recombination (gene conversion) with a heterologous sequence, there was a total absence of reciprocal translocations [221]. However, inducing DSBs with two I–Sce I sites on different chromosomes led to frequent translocations [172].

Our discovery of two different repair events in the chromosomal DSB assay which appeared to have cut at endogenous I–Sce I–like sequences extends these findings to human cells, and employs two sets of chromosomal loci wherein one of the two partners in each case is a completely natural site with no artifical construct. This provides convincing evidence that DSBs can initiate translocations in human cells. It also indicates that DSBs in chromosomes are brought together to undergo repair in human cells. There is, of course, a great deal of evidence that DNA molecules are physically brought together during DSB repair, but from a

source frequently overlooked: "illegitimate" DNA integration, and gene targeting. We now discuss these, in the context of a mechanism for DSB insertional repair.

## *4.4 Possible Mechanisms for DSB Insertions*

Our group and others have demonstrated that DNA transfected into mammalian cells will integrate (often exclusively) at sites of known or induced DSBs [99,101,117,168]. The introduction of a DSB in the genome greatly elevates the frequency of homologous recombination (100–fold increase in hamster fibroblasts), a phenomenon which thus increases the efficiency of gene targeting up to 5000–fold in mouse embryonic stem cells [114,222]. This demonstrates that the nucleus is capable of bringing DNA molecules together to jointly undergo recombination and repair.

The possibility exists that sequences are not directly patched into a DSB, but rather are copied from a template and the copy inserted at the break site. This "break–induced replication" may involve the invasion of one or both DNA ends, and has been proposed to explain the absence of sequence loss, and the involvement of microhomologies, at break:insert junctions [117,223,224]. This process has been described by our lab and others as ectopic gene targeting or "copy–join", whereby regions of homology between an incoming vector and a genomic site, including a genomic DSB, lead to strand invasion and copying of genomic sequences onto the vector, prior to its integration elsewhere in the genome [99,100,225]. This process could account for the two events of homologous recombination we observed in clone AZ7. However, for insert events which do involve a great deal of degradation from the ends of the inserted molecule, or a complete lack of any microhomology, it is entirely plausible that the DSB repair machinery might simply "paste" a fragment which happens to be in the vicinity into the break.

All our results point to chromosome breakage as the basis for generating fragments, which are then used during DSB repair. A very recent report of I–Sce I induced DSBs in hamster cells demonstrated small fragments could be excised from the genome (by flanking I–Sce I sites) and then undergo insertion into a third I–Sce I site elsewhere in the genome [226]. This

provides strong evidence for our model whereby increased genomic instability, causing DSBs, generates fragments with breaks on both ends, which then insert into another DSB.

The mechanism for physically moving DSB repair substrates to the same locus in the human nucleus is completely unknown. However, as we saw in the Introduction, the great affinity some DNA repair proteins have for binding directly to broken ends, as well as clustering into foci, suggests they may play roles in bringing DNA fragments and chromosomal loci together. We now have at our disposal DSB labelling and immunofluorescence techniques (such as GFP–coupled γ–H2AX, or DNA intercalating dyes) which are used to study DSB repair complexes or plasmid uptake by cells [75,227]. Combining these methods would enable real–time determinations of the dynamics of association between DNA regions and repair proteins or structural proteins (like those of the nuclear matrix), as well as provide the ability to trace their movements within cells. In this way, we could better define the principal players and mechanistic causes of co–localization of DNA fragments for DSB insertional repair.

## *4.5 Inserts and Genome Evolution*

Human vanity leads many to believe we are the pinnacle of evolution – when in truth we represent the shoot at the end of one branch, on a huge tree with countless other organisms, all supported by the same roots. Nature displays myriad possible outcomes from similar genetic beginnings, and the human genome is one adequately successful example.

> "So we know that the human genome is a large and disordered jumble of ancient viruses punctuated by a modest collection of genes, some from bacteria, and that it has come to do much more than it could possibly have been designed to do."[228]
>
> —Henry Gee

This quote, which accompanied the landmark issue of Nature describing the release of the draft sequence of the human genome, confirmed a concept that has long been accepted: that

the human genome evolved, in part, as a patchwork of DNA from different sources. The studies described in this thesis outline perhaps the first direct comprehensive demonstration of this process, capturing events of DNA shuffling in human cells. Our results show that human cells patch a site of DNA damage with (amongst others) fragments of human DNA which span repetitive elements, exons, and introns.

Comparative genomics tells us the ability to begin shuffling exons, in order to re–use existing protein domains in different combinations, coincides with the great diversification of multi–cellular organisms, the metazoan (Cambrian) "explosion" [229]. In order to observe this shuffling, we had to accelerate the evolutionary process, and follow the repair of artifically–induced DSBs. But ultimately what we have demonstrated is a process, taking place in the genomes of our cells, that can account for rapid change in an instant.

We have seen that breaks do occur naturally, and frequently. The fact that DSBs occur more readily in highly expressed regions means any recruitment of fragments to a DSB puts the new pieces into dynamic loci, more likely to be open and expressed, upon which evolutionary forces can act to "try out" this new genome addition. This means that active regions of the genome are both more likely to generate a fragment which could be used for insertion, as well as to become sites of insertion. This implies that Nature may have even more opportunity to test DNA inserts in coding regions than elsewhere, which (as we have observed) could bring exons and introns from other genes to experiment with, and encourage even more rapid evolutionary changes.

Traditional Darwinian evolutionary theory is based on a few simple principles. These dictate that a cell or organism is successful if it is able to survive in its environment long enough to reproduce. Any traits which increase that reproductive output are then passed on to more offspring, and those which prove detrimental to reproduction face "selective pressure", and are eliminated from the population. But environments can change, or cells and organisms can differentiate and give rise to new species, accompanied by new traits. The source of this novelty, according to the early molecular Darwinists, is the gradual change of single nucleotide bases from time to time as errors of replication or repair. Over millenia, these are

predicted to add up to new genes and new genomes. The concept that large–scale changes can occur, leading to rapid or "saltatory" evolution, is still considered by many theorists to be a great risk (perhaps too great) for a cell or organism to take.

On an evolutionary timescale, the alterations which a single DSB insertional repair event can impart upon a genome are cosmic leaps beyond the single nucleotide changes that Darwinian geneticists invoke as the source of diversity upon which natural selection can act. While it is true that some DSB insert events might prove detrimental to survivial, should they disrupt an exon for instance, we propose that the majority of such inserts are likely to be at least neutral in terms of selective pressures, and even advantageous. We have evidence from some of our chromosomal DSB assay clones that large inserts at DSBs had no visibly negative effects on cell growth and survival. We saw in the Introduction the degree to which our genome is composed of DNA elements, encoding protein motifs, which are re–used time and again in novel combinations. We now address the role of DSB insertional repair in repetitive DNA and genome organization.

### 4.5.1 Repetitive Elements and DSB Inserts

With the evidence we present, DSB insertional repair could be an ongoing process which is not always detrimental. To argue that the human (or any other) genome is too delicate to support radical changes of this nature takes us right back to the evidence which sparked this entire study in the first place: the predominance of repetitive DNA in the genome.

LTRs and LINE elements are ancient retroviruses which successfully infiltrated our genome without creating an evolutionary dead–end. Not only did they sneak in once, but they managed time and again (as evidenced by multiple families) to establish many colonies which then spread to generate hundreds of thousands of copies. If taking on new genetic material, or shuffling around existing DNA, were so horribly detrimental, then how and why would we have been able to survive to a point where we are mostly composed of this material?

There is some, albeit limited, evidence that mobile genetic elements can lead to disease in humans. A total of 13 cases of L1 insertions, and 18 Alu insertions, have been reported as arising either *de novo* or in the recent evolutionary past, which result in a disease–causing mutation [106]. LINE retrotransposition into the Factor VIII gene has been recorded in 2 cases, resulting in haemophilia A [230]. Although there is ample evidence that mobile elements can contribute to genomic rearrangements in a non–detrimental manner also. Though long suspected, it was only recently demonstrated that LINE reverse transcriptase is responsible for the retrotranposition of human SINE (Alu) and pseudogene mRNAs, as well as the probable cause of spread of the SVA class of repeats [231–233].

One mechanism for exon shuffling has been the ability of retrotransposable elements to transcribe beyond their 3' ends, carrying flanking DNA with them to new locations as in LINE–mediated transduction [234]. Our lab has shown another way to bring existing sequences to new areas is for homologous recombination between LINEs in different genomic loci to extend their gene conversion tracts into flanking DNA [89]. However, the vast majority of LINEs in the genome are not the complete 6.1 kb actively retrotransposing type, but rather are truncated and incompetent for retrotransposition, averaging only 900 bp long. Recent work dissecting the machinery necessary for LINE integration into the genome of rodent cells shows that, in the absence of specific integrase activity (which induces a break and hence creates the integration site), LINE cDNAs still integrate into the genome [235]. The presumed sites of non–integrase mediated insertion are naturally occurring DSBs.

On a global genome scale, there is a concentration of LINE DNA in AT–rich (gene poor) regions of the genome, while SINEs seem to accumulate in GC–rich regions (gene rich) [236]. At greater resolution, however, one particular analysis of the genome sequence indicates that a great deal of transcribed RNAs actually contain truncated LINE and SINE sequences in the final translated proteins [237]. These LINE segments were not part of the original protein's open reading frame, but came to be incorporated as a result of alternative splicing.

We can envision a DSB in an intron being repaired by insertion of DNA fragments, and should one of those contain appropriate sequences from a LINE, establishing a novel

transcript through protein domain addition and alternative splicing. This model fits with the fact that many repetitive elements carry several potential splice sites [238]. This is one way in which repetitive DNA elements could provide selective advantages to a genome, using DSB insertional repair (and other mechanisms) to spread and thereby increase the combinatorial complexity of existing genes.

## 4.6 Endosymbiosis and Inheritance of DSB Inserts

The endosymbiotic theory of evolution posits that the bacterial ancestors of mitochondria inhabited an early eukaryotic cell, and in exchange for housing provided energy in the form of oxidative respiration (the electron transport chain) [239]. While they maintained their own genome, there was a gradual transfer of some genetic material (still vital to mitochondrial function) to the nuclear genome. It has now been observed that the mechanism of DSB insertional repair can account for this transfer – taking fragments of mitchondrial DNA and using this to patch genomic DSBs [111,112]. But to have an impact beyond the host organism, and to be of evolutionary importance, this must take place in germ cells.

Mouse germ cells display chromosomal rearrangements in association with γ–H2AX (DSB) foci, and end–joining activites are certainly active in these cells as this is the basis for the integration of foreign DNA by pronuclear microinjection when generating (non–targeted) transgenic mouse lines [240,241]. However, spontaneous insertions in the human germ line have been estimated to be relatively infrequent (based primarily on studies of LINE elements), accounting for about 1 in 600 mutations [242]. A survey of gross chromosomal rearrangements in humans lists 28 instances of germline deletion events which are accompanied by detectable "short" insertions [148].

Analysis of one familial translocation breakpoint t(9;11)(p24;q23) revealed the concurrent insertion of 41 bp of 12S mitochonrial DNA during the translocation event [243]. While perhaps not large enough an insertion to demonstrate the entirety of endosymbiotic transfer of mitochondrial genes to the nucleus in humans, this does reveal mitochondrial DNA can be used to repair sites of DNA damage in human germ cells, and shows the transmission of a DSB insert event between generations. With our attention now drawn to this mechanism of

repair, and the genome sequence at our disposal, we are likley to see further identification and better characterization of germline DSB inserts and their inheritance.

## *4.7 Horizontal Gene Transfer and the Evolution of Genomics*

The transfer of genes from one organism to another (non–reproductively), and from one species to another, is a phenomenon frequently discussed in microbiology, but rarely applied to "higher" eukaryotes. This is known as horizontal, or lateral, gene transfer (HGT) [244]. The initial analyses of the human genome, based on studies of sequence divergence and orthologous sequences (shared between species), concluded that as many as 223 of our genes were the product of direct transfer from bacteria [1]. This sparked a heated debate amongst evolutionary biologists, rarely seen in modern science. The detractors pointed out the paucity of sequenced genomes and the failure to draw on phylogeny, which could explain the loss of some genes from "intermediate" evolutionary stages [245]. Revisions and more annotations may continue, but the proponents maintain that, even if they had over–estimated the proportion of the contribution lateral gene transfer had made to the make–up of the human genome, the process had occurred nonetheless.

### 4.7.1 Viral Evolution

> "We as humans are not an organism which is encased in glass or concrete. We interact with our environment. So in a sense, we play ping pong with viruses and bacteria, and this drives evolution...."[246]
>
> — Josef Penninger

In human cells, we observed that DSB inserts spanning the SV40 origin of replication conferred replicative capacity upon the new circularized repair product, in cells known to express T–ag (COS–7 and GM00637 or GM05849). In some cases of DSB repair, where both the origin and sequences coding for T–ag are mobilized and used to patch DSBs, these newly formed molecules have the capability of replicating in human cells (HeLa) which

previously lacked T–ag expression. This autonomous replication also occurred in repair products which included intact origin and T–ag regions of SV40 along with flanking cellular DNA, demonstrating the spontaneous amplification of human genomic sequences via association with SV40 DNA in both normal and AT cells.

Conversely, if the SV40 fragments we detected were indeed linear fragments prior to the repair, then one could look at the subset of SV40 DSB repair events as situations in which these extrachromosomal viral DNA molecules (with or without adjacent cellular DNA) patched their own random DSBs with the transfected linear substrate DNA from pBluescript. This then implicates DSB insertions in the evolution of viral genomes. Recircularization events of SV40 genomes in mammalian cells have previously been described, but due to the lack of reference sequence data and the tools to compare them, only deletions at the repair junctions could be characterized [247]. Variant SV40 genomes which acquire small repetitive host DNAs can arise in monkey and human cell lines, but again the paucity of sequence information available at the time of study limited the degree to which insertions and complex rearrangements could be identified [248–250]. The ability for mammalian genomes to integrate SV40, and for SV40 to take up other DNA, paints an interesting picture of genetic ping–pong between host and pathogen.

#### *4.7.1.1 The Origins of Retroviruses*

The impact that a single retrovirus strain has on a species can be devastating, to say the least. Particularly a retrovirus, like the human immunodeficiency virus (HIV), which transmits and kills its host and offspring through the fundamental stage of species survival: reproduction. (Please note this abstraction of HIV is used to illustrate an evolutionary concept with a relevant human example. However, the burden that HIV/AIDS places on individuals, and our society in general, cannot be understated. We express our utmost sensitivity and respect for this horrible affliction, in the hopes this discussion will contribute to a useful dialogue).

A fascinating bit of phylogenetic comparison has led to the proposal that LTR retrotransposons, which are found in nearly all species, are in fact the ancestors of

retroviruses and not just their derelict successors [251]. Both elements encode *gag* and *pol* genes for proteases, reverse transcriptases, and integrases. The difference is that retroviruses have acquired a third gene for an envelope, *env*, which permits them to transmit between cells. This step has occurred independently on several occasions, creating novel retroviruses from a variety of vertebrate and invertebrate endogenous retrotransposons [252]. The mechanism for this acquisition has yet to be demonstrated but the characterization of DSB insertional repair in plants, yeast, rodents, and now (in this thesis) in human cells provides one compelling possibility.

If DNA from a virus can be used to patch a DSB, as we have seen in the case of SV40, then the envelope genes of an infecting strain could find their way into the genome at sites of retrotransposons. The Human Genome Sequencing Consortium, in their initial analysis of the draft human genome sequence, state that while DNA transposons have become completely inactive, LTR retrotransposons may remain somewhat active [1]. This is intriguing, as we find 15% of our extrachromosomal inserts (of human DNA alone) have a junction which lies within an LTR, which make up only 8% of the genome [1]. This indicates these regions are exceptionally prone to involvement in DSB insertional repair in the human genome. Perhaps these elements continue to play an important role in our genome despite a loss of active retrotranposition, as sites which preferentially generate fragments capable of inserting into DSBs. They may also be even more likely than other genomic loci to insert DNA from other sources, such as a viral *env* fragment, contributing to the evolution of new retroviruses. One can envision developing a system for observing insertions at an I–Sce I site in LTRs, using human cells which could be infected with *env*–containing viruses, to determine if we could observe this process in action.

In all, these examples provide a basis for the involvement of DSB insertional repair in a process of genetic exchange between viruses and their hosts. As the evidence for HGT accumulates, it challenges us to find the mechanisms capable of vindicating it. We question the belief that genomes operate as closed systems, unwilling to permit drastic additions for fear of upsetting a delicate balance. The results of this thesis work have demonstrated how, in a single act of DSB repair, large portions of coding and non–coding chromosomal DNA

can "hop" to a new location. And given the environments in which we live, and have evolved, there is no shortage of opportunites for our ancestral and current genomes to be exposed to other types of pathogenic or foreign DNA.

## *4.8 DSB Inserts, SV40 and Disease*

### 4.8.1 SV40 and Cancer

Simian Virus 40 (SV40) was discovered in 1960, a contaminant of the polio vaccine which was eventually administered to millions of people between 1955 and 1963 [253–255]. Though its preferred host is the rhesus monkey (*Macaca mulatta*), it is closely related to the human papova viruses, and is known to establish variable levels of infection in certain human tissues, resulting in the generation of neutralizing antibodies [256]. Since its discovery, SV40 has been shown to induce an array of tumours when injected into rodents, but it is only in recent years that groups have screened the equivalent human cancers and positively identified the virus. These include ependymoma and choroid plexus tumours, osteosarcomas and sarcomas, and a well characterized association with mesotheliomas [257–263]. And now, in the past two years, several studies have identified SV40 sequences in human lymphomas [264–267]. Though controversial, some studies reveal SV40 is absent from tumours in countries which did not use contaminated polio vaccines [268,269].

### 4.8.2 Traditional Views of SV40 Mode of Action

The mechanisms by which viral–encoded proteins, in particular the SV40 large tumour antigen (T–ag), can transform mammalian cells have been extensively described [270]. T–ag inhibits both p53 and Rb tumour suppressor proteins to promote excessive cell proliferation, and has recently been shown to extend the proliferative lifespan of human cells by activating expression of telomerase [271]. While most SV40 in human cells is believed to persist extrachromosomally, the shift from episomal to chromosomally–integrated SV40 DNA is the laboratory standard for defining "stable" transformation of a cell [141,257]. Human tissues (and particularly patient specimens) are typically screened for the presence or

absence of SV40 by PCR amplification of T–ag sequences, with little or no attention paid to other portions of the viral genome, such as the origin of replication [257,272].

### 4.8.3 Replication Origins

The regulatory region of SV40 consists of enhancer/promoter elements, as well as the origin of replication, and can be found to rearrange and recombine in laboratory strains and human tumours, sometimes producing exceptionally fast replicating viral genomes [166,273,274]. The SV40 origin of replication maintains its efficiency despite a variety of rearrangements in its core and auxiliary sequence elements [267]. Both mammalian and SV40 origins of replication have a core region (which dictates where replication begins), and auxiliary elements which stimulate replication [275]. In order to infect a cell and produce high copy–numbers of the viral genome, the viral origin of replication must be capable of overcoming the licensing restrictions placed on the host cell's genomic origins [276,277]; that is, a viral origin must be able to replicate more than once per cell cycle.

Though most studies show replication from the SV40 origin requires T–ag expression, some evidence suggests this is not exclusively the case [278,279]. Under certain conditions, the origin is competent for replication in the absence of T–ag, as are the flanking promoter elements capable of strong expression of associated coding regions [280]. There are even reports of mammalian amplification promoting sequences capable of activating the SV40 core origin in the absence of T–ag [275]. Interestingly, the same DNA end–binding proteins (Ku70/80) implicated in integration of retroviral cDNA and repair of DNA damage also bind to mammalian origins of replication [281,282].

### 4.8.4 Gene Amplification and Expression

The ability for SV40 to amplify both its own, and other associated DNA, in mammalian cells is well established. Early studies with CV–1 monkey kidney cells, permissive for SV40 infection, identified populations of presumably linear DNA molecules derived from both SV40 and host DNA, which were believed to be packaged into pseudovirions and able to replicate [283]. Rearrangements in SV40 transformed rodent cell lines have described the

appearance of small, viral DNA–containing "microchromosomes" upon irradiation or chemical exposure [284–286]. And amplification of SV40 and flanking host DNA can be induced in hamster cells by infection with other viruses, such as herpes simplex virus (HSV) [287]. What is novel in our work is not only the fact that we observed this amplification process in human cells, but that we have evidence these molecules can travel to new loci via insertion into DSBs (see Figure D1).

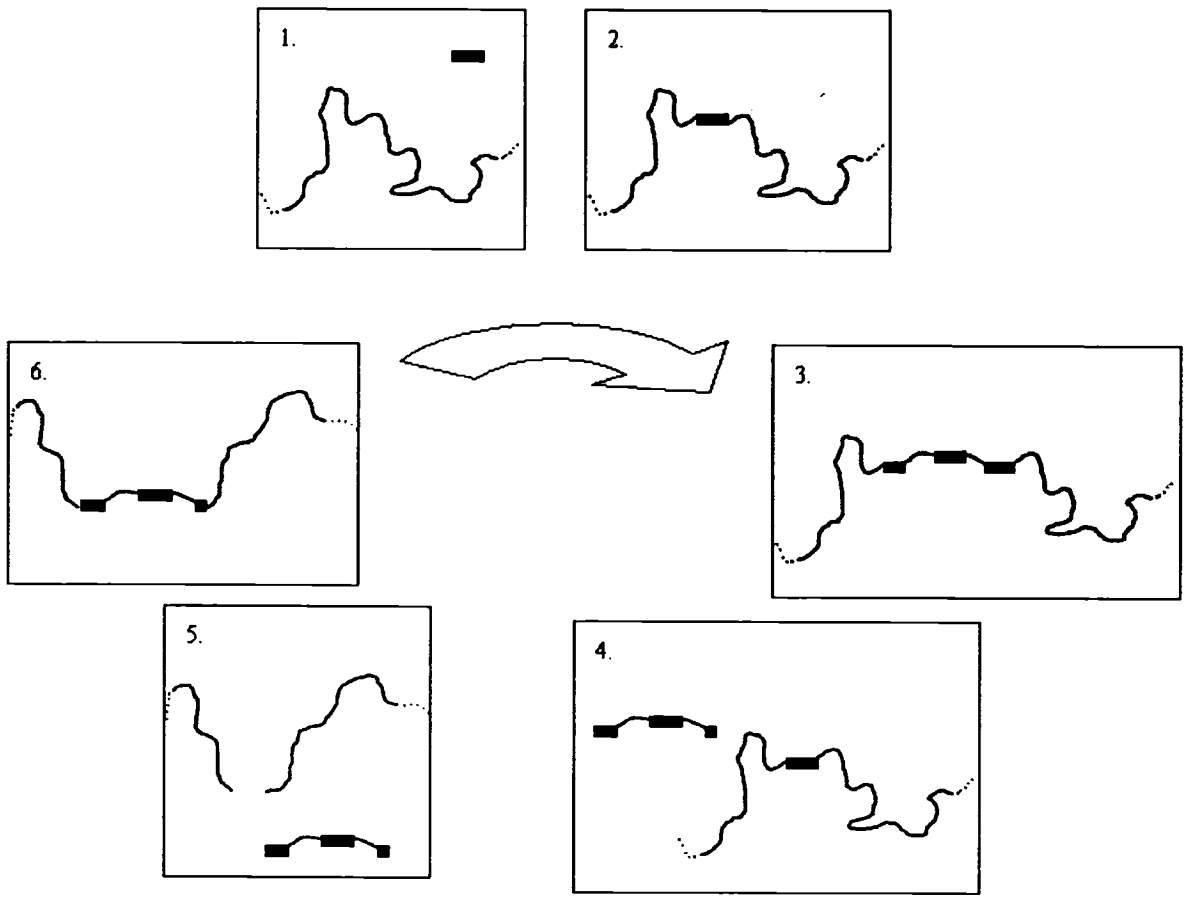



Figure D1. Model for amplification and spread of genomic DNA associated with the SV40 origin of replication. Genomic DNA is shown in thin lines, SV40 DNA in thick lines. Step 1: SV40 DNA enters the cell nucleus. Step 2: A region of SV40 spanning (at least) the origin of replication integrates into the genome, possibly at an endogenous break site. Step 3: Replication is initiated from the SV40 origin by presence of T–antigen (either encoded on site or elsewhere in the cell), out of schedule with the genomic origins, leading to extra copies of some SV40 DNA and flanking genomic DNA. Step 4: A portion of the copied DNA (SV40 and flanking genomic DNA) is freed from the genome, possibly due again to a break. Step 5: The free–floating SV40/genomic DNA fragment co–localizes with an endogenous break elsewhere in the genome. Step 6: The free–floating fragment inserts at the genomic break, and the process can begin again. This may involve further amplification (a return to Step 3), and the shuffling of still more flanking genomic DNA to new locations.

Amplification of the DHFR locus in rodent and human cell lines, as well as the MYCN oncogene in human tumours, results in a gene dosage imbalance [192,288,289]. The repeated units or "amplicons" have been key indicators of the presence and abberant activity of replication origins [290]. These amplified units can each be hundreds of kilobases in length, manifesting as both large blocks of homogeneously staining regions (HSRs) within a chromosome, and/or more often as extrachromosomal repeats called double minutes (DMs) [196,291]. HSRs and DMs can be detected by FISH analysis in metaphase and interphase [292,293].

Human tumours associated with SV40 can show profound chromosomal abberations [294]. Leukemic cell lines and primary human tumours have exhibited amplification of large non–contiguous chromosomal fragments [295,296]. Plasmids containing the SV40 origin and MAR sequences (directing them to the nuclear matrix) were capable of inducing genetic amplification in human cells [297]. But while most studies of SV40 focus on the ability of its protein products to interfere with normal cell functioning in human cell lines, few details exist on the particular integration of SV40 sequences and the resulting structure in the context of the whole human genome [141].

### 4.8.5 SV40 Integration at 18q21

The integration sites of SV40 in several human cell lines have been characterized and reviewed, displaying no apparent preference for any one particular chromosomal site [141,154–160]. Although it is a well–documented carcinogen in rodent cells, human cells are generally considered non–permissive for SV40, despite increasing reports of association between SV40 sequences and human cancers equivalent to those observed in infected rodents including lymphoma, osteosarcoma, and mesothelioma [257.258,265.268,298].

Our observation of SV40 integration in the AT cell line draws our attention to this particular area of chromosome 18q21. This chromosomal region is frequently deleted or translocated in a variety of cancers (including carcinomas, osteosarcoma, mesothelioma, and colorectal cancer), many of which are cancers beginning to demonstrate strong associations with SV40 infection in patients [299]. As well as deletions, other cancers such as diffuse large B–cell

lymphomas and follicular lymphomas often show what is referred to as "add(18)(q21)"; that is, "additional material of unknown origin" in that region on the chromosome (as reported in the Mitelman Database of Chromosome Aberrations in Cancer) [149].

In our extrachromosomal DSB assay, we identified DNA inserts from the GM05849 cell line which came from the integration of SV40 into chr 18q21 in the WDR7 gene (NCBI Locus ID: 23335). This is a TGF–beta resistance associated gene, reported to be over–expressed in a number of cell lines according to the CGH (comparative genomic hybridization) database at NCBI [293]. A number of putative tumour suppressor genes, some also implicated as transcription factors in TGF–beta signalling, lie nearby on chr 18q21: DCC (deleted in colorectal carcinoma, ID: 1630), MADR2/Smad2 (ID: 4087), DPC4/Smad4 (deleted in pancreatic carcinoma, ID: 4089). Within the 4 Mb range of the chr 18q/SV40 inserts, at 18q21.2 is MALT1 (mucosa associated lymphoid tissue lymphoma translocation gene 1, ID: 10892), frequently translocated in low–grade B–cell lymphomas [300]. Within 2 Mb outside our span of insertions lie FVT1 at 18q21.3 (follicular lymphoma variant translocation 1, ID:2531), which is highly overexpressed in some T–cell malignancies [133], as well as BCL2 (ID: 596). BCL2 overexpression is detected in the great majority of non–Hodgkin's lymphomas, as well as in other myeloid and lymphoid cell lines and tumours not exclusively associated with the translocation t(14;18)(q32;q21), which places BCL2 under the control of an immunoglobulin enhancer element [301–305]. It is interesting to note that the frequency of instability in this region, as demonstrated by translocations, increases with age as does lymphoma incidence [306], though any correlation with cohorts exposed to SV40 from contaminated polio vaccines remains unexplored.

### 4.8.6 Summary: SV40

While many of the cancers described above derive from cell types other than the fibroblast cells we have studied, the cytogenetic and sequencing technologies available today warrant a renewed investigation into the possibility of preferred sites for the integration of SV40 in human cells. We have seen indications that SV40 could play a role in disrupting the replication and expression of associated DNA, while SV40 is increasingly detected in human cancer patients. The discovery of widespread SV40 infections amongst monkeys suffering

from simian immunodeficiency syndrome raises even more concern for situations of compound infection [167,307,308]. The close association of some types of lymphoma, particularly systemic non–Hodgkin's lymphoma, with human acquired immunodeficiency syndrome (AIDS) and HIV infection, warrant further investigation into the role SV40 plays in the onset of human disease [309,310]. Current research priorities and therapies being devised to screen or treat SV40 association in cancers are all focused on vaccination against T–ag expression. Our results raise concerns that these priorities ignore other potential risks of contact with the virus, often increasing patients' exposure to SV40 DNA through vaccination strategies [255].

The tide of public opinion is shifting to acknowledge that, under natural conditions, SV40 exposure may lead to cancer in humans [311]. Despite the evidence, however, no sufficient causal link has been firmly established between SV40 and human cancer. A recent commentary concludes that "the resistance to implicate SV40...in human carcinogenesis lies in the inability to convincingly define a biologically functional role at a cellular level" [312]. A heated and very public argument wages in the literature over the possibility that SV40 infection causes cancer in humans – much like the controversy discussed previously between phylogeneticists and some other evolutionary genomics groups over horizontal gene transfer and the evolution of the human genome. What is striking is that both debates stem from lack of solid evidence for a mechanism, and that our confirmation of DSB insertional repair taking place in human cells provides a mechanistic foundation which could explain both phenomena.

We have mentioned the possibility, worthy of investigation, that SV40 DNA may be used in DSB insertional repair not just in cell culture assays, but possibly also in some of the millions of people exposed to SV40. Let us turn now to another natural source of DSB inserts, which could also bestow a selective advantage upon its host cells: apoptosis.

## *4.9 Apoptosis, Wounds, and Genometastasis: New Life for Old DNA*

With the burgeoning field of apoptosis (programmed cell death) research came an appreciation for, or perhaps misinterpretation of, this cellular suicide as an altruistic act. The important hallmark of apoptosis, and the target of much cancer therapy, is to induce the fragmentation of the genome. This is believed to remove a cell's capacity to do harm. However, several lines of evidence suggest apoptosis could play another role – that of provider of DNA patches – which may contribute to rescuing cells in need of DSB repair.

As apoptosis begins, the genome becomes cleaved and the nuclear membrane breaks down into smaller vesicles, encapsulating the fragmented genome into apoptotic corpses. Phagocytosis of these corpses is generally (but not exclusively) a receptor–mediated event carried out by macrophages, which co–opt the endoplasmic reticulum to gain additional membrane material for engulfment [313,314]. Interestingly, this recycling process of membrane renewal is such that the endoplasmic reticulum and the nuclear membrane can also be extensions of one another, believed to facilitate protein and ion transport into the nucleus [315]. Apoptotic bodies containing fragmented DNA from dying cells, instead of being removed from posing a threat, become an ideal system for sending ready–made linear DNA fragments to neighbouring cells, packaged for delivery in the same membranes as their target cells and organelles.

This concept of "apoptotic corpse gene delivery" is closer to reality than one might initially suspect. A comparison of mouse and rat embryonic fibroblasts (whose labelled chromosomes could be readily distinguished) showed that, upon induction of apoptosis to one set of cells, exposing them to the other set of healthy cells resulted in the rapid uptake of DNA to the nucleus, and its incorporation into host chromosomes [316]. This situation required that the recipient cells be deficient for p53 (hence, unable to trigger their own apoptotic pathway), and that the donor cells have oncogenes such as c–myc activated to drive cell cycle progression in the absence of p53. Similar uptake of apoptotic bodies containing HIV

or Epstein–Barr Virus (EBV) DNA, and the shuttling of that DNA into nuclei, has been reported by the same group [317,318].

Non–viral DNA uptake can also be mediated by cell–surface receptors, such as by the FGF2 receptor [316]. Physical traumas suffered by tissues rapidly increase expression levels of FGF2, resulting in rapid uptake of exogenous DNA into the nuclei of wounded cells. As such, there is evidence that mammalian cells (under certain conditions) are capable of taking up and expressing retroviral, viral, and oncogenic material as a result of natural "transfection" processes, apoptosis and mechanical wounds. This is important, as DNA can resist degradation (under certain conditions) and circulate in the plasma for years [319]. The sources of this circulating DNA include activated oncogenes, leading to a risk of secondary–site tumour formation by a process dubbed "genometastasis" [320].

The mechanism of DSB insertional repair explains how DNA from apoptotic bodies becomes incorporated into a new host cell genome. It also describes the basis for the establishment of secondary sites in genometastases. This possibility should be of increasing concern, now that a new wave of cancer vaccine technologies are being developed, and in clinical trial, which inoculate patients with transformed irradiated cells. While the aim of these vaccines is to train the host to recognize cancer–specific antigens, the process exposes them to fragmented oncogenic DNA, and should be investigated with respect to the risks posed from DSB insertional repair.

## *4.10 The Advantage of DNA Patches: A Quick Fix, a Long Reach*

We mention previously the possibility that some DSB inserts do not display any detrimental effects upon the cell phenotype (as is the case of our chromosomal assay inserts which permitted clonal expansion). We also point out the huge proportion of the genome (over 98%) which is made up of non–exonic DNA, which we suggest could have accumulated in large part by DSB insertional repair. We propose that this mechanism of repair confers a selective advantage, not only by providing a means to rapidly repair DSBs, but also in the

context of increased complexification of the genome both by shuffling of functional units and perhaps even providing greater range of motion to chromosomal territories within the nucleus.

To picture precisely what is going on in an interphase nucleus when a DSB occurs in a chromosome, recall the packaging of DNA from the Introduction. Imagine the incredible degree of tension carried along the length of the DNA molecule (and the potential energy stored within) as it is rolled about histones then twisted, folded, coiled, threaded through protein complexes, and likely anchored into loops along a protein scaffold. To suddenly endure a cut in such a strand would be like winding an elastic band between two pencils and then clipping it in the middle. Any two ends which weren't fully anchored in place directly adjacent the break site, could spin out into the nuclear ether.

Recent technical developments suggest this can be the case in cells, as measurements of DSB ends show they can become separated in S/G2 phase [321]. But if the tension suddenly released when the coiled chromosome is severed results in two ends being separated, there might be a further advantage to using some other piece of DNA (which happens to be in the vicinity, or which can be actively brought to the break site) to bridge the gap. The large concentrations of repair proteins into factories can bring together multiple DSBs, but may also permit DSB insertional repair to bridge gaps between broken ends, facilitating repair.

Bridging or not, a cell which demonstrates the ability to generate or obtain DNA fragments, which it can then use to quickly insert into a DSB, displays a selective advantage because it can more efficiently rescue itself from cell cycle arrest and apoptosis. Survival for a multi–cellular organism might sometimes mean suppression of major genomic rearrangements is favourable, such that the organism does not develop a cancer which would limit its reproductive success. But survival for a cell means avoidance of death, and any mechanism which permits the cell to quickly re–establish its cell cycle following a DSB would be a selective advantage (see Figure D2).

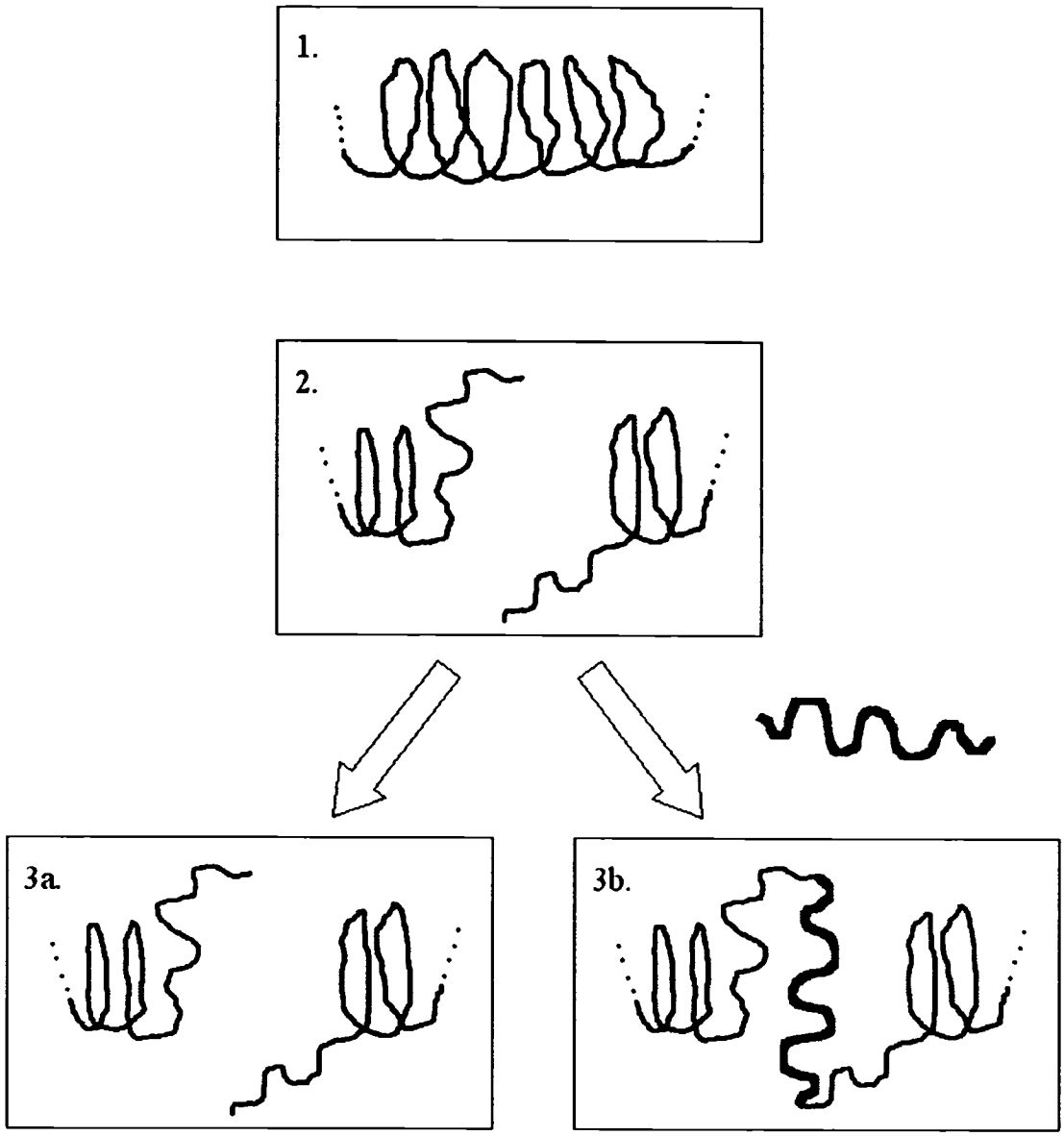



Figure D2. Model for the selective advantage conferred by bridging of distant DSB ends by insertional repair. Step 1: A chromosome (black line) is tightly wound and packaged into chromatin. Step 2: A DSB occurs, releasing some of the torsional stress in the chromosome and distancing the broken DNA ends. Step 3: a) Repair of the DSB does not take place at all, or not quickly enough to avoid cell death by apoptosis, because the ends are not close enough together; or b) a fragment of DNA (thick grey line) which is able to co–localize with the DSB will insert during repair, and the cell resumes cycling.

At the level of a multi–cellular (metazoan) organism, we have seen DSB insertional repair does permit shuffling of coding and non–coding sequences, which could lead to novel genetic combinations. We have suggested this mechanism could help explain the accumulation of "junk" DNA in our genomes. Beyond the selectively neutral occurrence of mobile DNA inserting without disrupting a coding sequence, or even the seemingly disadvantageous metabolic cost of replicating so much DNA, we propose that padding a genome in this way is in fact advantageous.

We have discussed how chromosomes move within the nucleus, and how this positioning influences gene expression. Accumulating more DNA, as by DSB insertional repair, may increase the bulk of the genome in such a way as to provide a wider range of motion to coding regions. Having genes separated by long stretches of linker DNA enables a longer reach, such that different combinations of genes on different chromosomes might be able to co–localize and be subject to common replication, repair, and transcriptional control. This could enable more combinations of interaction between loci, which may be responsible for the differential expression of genes seen in development and cellular differentiation (see Figure D3). We might just owe our phenotypic complexity to the stuffing and shuffling of our genome by DSB insertional repair.

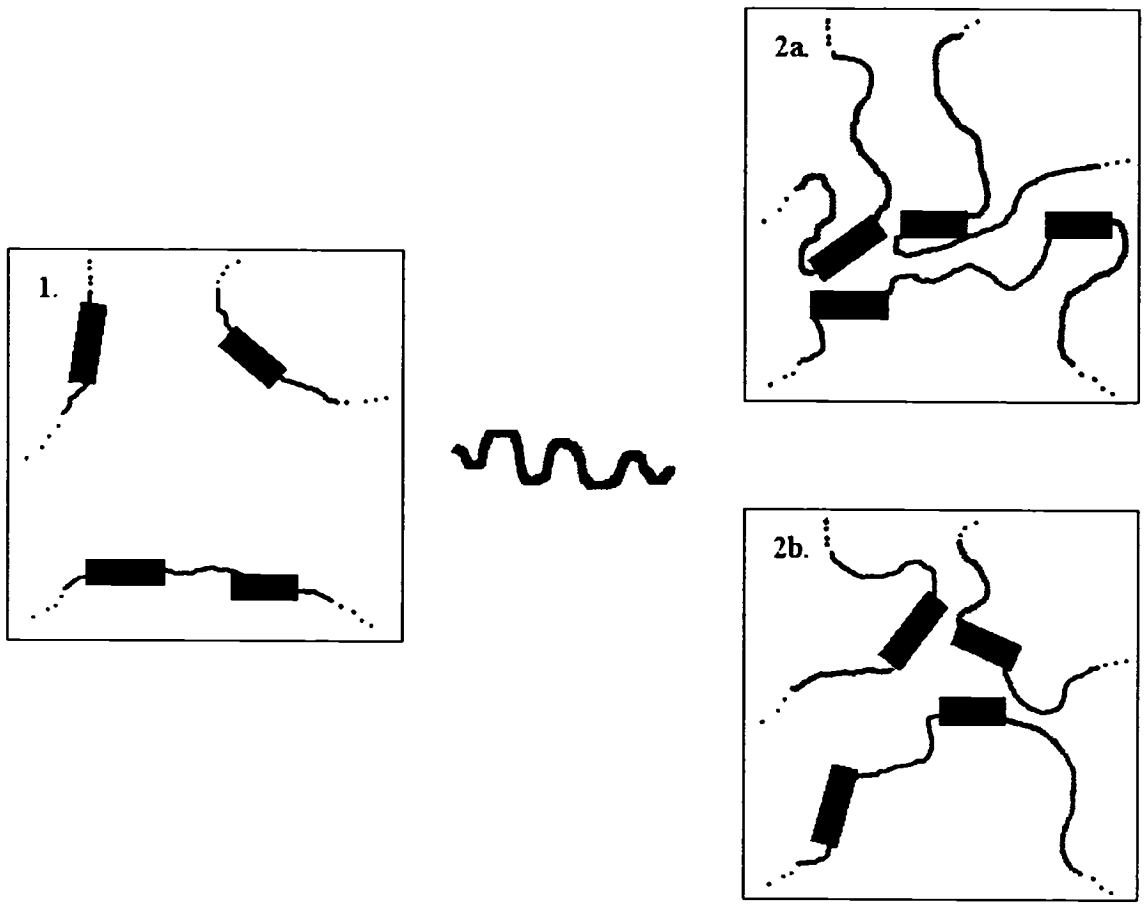



Figure D3. Model for the increased complexity of gene associations in genomes with inserted DNA. Genome 1: A gene–rich, "junk"–poor genome. The three distinct chromosomes of this genome (thin black lines) show little DNA separating their genes (black boxes), and limited ranges of motion. Genome 2: A genome which has succeeded in incorporating other DNA fragments (thick grey line) will achieve greater range of motion for its chromosomes within the nucleus. Images 2a. and 2b. show different spatial relationships of co–localizing genes from distinct chromosomes. This repositioning within the nucleus could result in genes coming under common influence from replication, repair, or transcription "factories", producing a greater diversity of phenotypes.

## 4.11 Summary: DNA Breaks and Disease

Altogether, these examples provide compelling evidence for the need to consider assessing the DSB insertional repair capacity of some patients prior to administering therapy. In particular, we need to address the fact that most therapies cause breaks in the DNA of cancerous cells, using radiation or chemotherapeutic agents, in the hopes of inducing apoptosis. Recently, high densities of DNA damage in irradiated primary human fibroblasts were proposed to explain cases of presumed DSB insertion of genomic fragments during rearrangement and deletion of the HPRT locus [322]. If fragmented oncogenes can act as DSB patches, and then their delivery mediated by apoptosis, the very treatments we hope are eliminating cancerous cells by acting as magic bullets could instead be scatter bombs – therapeutic hydra whose heads, when severed, cause even more to spring forth.

From a cytogenetic standpoint, evaluating the progression from normal to cancerous cell is a challenge. Typically, the complex rearrangements seen in solid tumours (which make up 95% of all cancers) makes it difficult to discern a primary from a "bystander" event, and the assumption is that loss of a cellular gatekeeper function leads to genomic instability [323]. We have found that a background of genomic instability leads to the insertion of various chromosomal fragments at DSB sites, suggesting this mechanism of DSB insertional repair could be, at least in part, responsible for the progressive complexity observered in tumour karyotypes. The recent advent of chromosome painting techniques provides unprecedented resolution at which to analyze the relocation of chromosomal segments, and one could envision applying this technique to a system such as ours, where ATM deficiency in a cell results in genomic instability, to follow the outcome of DSB repair and ascertain the contribution that insertional repair makes to these complex karyotypes.

The efficiency of DNA repair appears normal in most cancer cells that have been studied, which has long posed a problem for understanding the hypermutability required for a cell to acquire sufficient genetic changes to become cancerous. Our study suggests that, as in the case of AT cells, it is perhaps not the efficiency of DNA repair, but the kind of infidelity of repair that a genetic background permits, that can rapidly change a cell's genome. Even

though inserts during DSB repair account for a small proportion of the overall DNA repair events, each insert event results in a dramatic alteration of DNA content in the region of the DSB. This can include the insertion of functional genetic elements, such as entire exons (as in the case of several AT insertion events), or a viral origin of replication. Moreover, the insertion of coding sequences (such as for T–ag, which conferred replicative capacity to repaired plasmids in HeLa cells), demonstrates that inserted material can be transcriptionally active and expressed from its new locus in human cells. This DSB insertional repair may be one mechanism which helps to explain the paradox of genomic instability and rearrangement (hypermutability) in cancer cells with efficient DNA repair capacities.

## 4.12 Future Directions

The extrachromosomal DSB assay, while simple, proved incredibly robust. It could be applied to any cell system (or organ system) which permits the transfection and subsequent harvesting of plasmids, to gain a better understanding of how cellular phenotypes influence the rate and spectrum of DSB insertion events.

A chromosomal DSB insert assay, while considerably more complex to establish and employ successfully, is now a reality. With the cell lines we have generated, there remain a number of mother clones not yet fully characterized, which potentially meet the criteria we set out (single vector integrations, upon which we can perform plasmid rescue, and apply Ganciclovir selection). The refinement of the GFP FACS–sorting of high I–Sce I expression levels, which we have demonstrated, means these existing clones could be subjected to a proven methodology with little development time, and the results compared to our existing data to better understand DSB insertional repair in a chromosomal context.

While we provide strong support implicating NHEJ as the principal mechanism responsible for DSB insertional repair, the exact components of the repair machinery involved and their individual contributions to the process remain unexplored. As mentioned previously, we chose the MO59 cell pair as a model system because MO59J was the only human cell line available at the time of our work with any deficiency in a component of NHEJ (and an incomplete one, at that, with reduced DNA–PKcs activity). Recently, widespread use of inhibitory RNA technologies has afforded a means of conditionally suppressing a given protein's level in the cell. Using these techniques, the DSB insert assays could be repeated to dissect what influence each NHEJ repair protein (such as Ku70, Ku80, DNA–PKcs, XRCC4, LigaseIV, or Artemis) or other repair proteins (Rad 51, Rad 50, Mre11, Nbs1) have on the efficiency, fidelity, and source of insertions at a DSB.

Also of interest for further investigation is the involvement of SV40 integration in the tumorigenic process. While current clinical analyses of SV40 and cancer concentrate on T–ag expression and its role in disrupting normal cellular responses, our results suggests a role

for the viral origin of replication in gene amplification and genomic instability. It would be important to investigate tumour specimens (particularly from cancers suspected to be associated with SV40) for the presence of integrated viral DNA. This should assess where integration occurs, and specifically if the viral origin is present. Establishing if any amplification or rearrangement at the cytogenetic level has occurred, or altered expression levels of associated genes, could be detected by Comparative Genomic Hybridization (CGH). These would be the first steps in exploring the involvement, in clinical cancer onset or progression, of this phenomenon which we observe in transformed cell cultures.

Our discovery of endogenous sites in the human genome which can be cut upon expression of the I–Sce I endonuclease introduces many possibilities for future studies. We observed at least two sites which cut, 8p12 and 17q23.2, thus identifying specific loci in the human genome which can now be used to study DSB repair in an entirely natural context. There is no need to introduce artificial bacterial or viral sequences, which can disrupt the chromatin context of the DSB site.

The question is often asked "why do organisms continue to support such a metabolically costly process as DNA replication of all this 'junk DNA'"? The assumption is that the more unnecessary DNA one can afford to carry around, the less likely a mutational event will affect the vital components of the genome. But how did they get this way in the first place? Obviously many genome insertion events have no serious selective disadvantage, as the persistence (or success) of we and other "junk DNA" heavy genome carriers can attest.

But what of organisms with smaller gene–rich genomes? Do they display DSB insertional repair capabilities to a similar extent? We have mentioned reports that yeast, when prevented from using homologous recombination, can employ DSB insertional repair. But to what extent has it played a role in the reshaping of other genomes? The ongoing sequencing projects will provide the necessary reference material against which, not only phylogenetic comparisons will be possible, but our questions of evolution and speciation through DSB insertional repair capacities will have the tools to be addressed.

# 5. APPENDICES

5.1 Appendix A: List of Primers

5.2 Appendix B: Vector Construction

## 5.1 Appendix A : List of Primers

| Oligo Name | Sequence (5' to 3') | Description |
|---|---|---|
| **Extrachromosomal DSB Assay** | | |
| pBS Insert (F) | actgttgggaagggcgatcg | To sequence or amplify the polylinker (DSB) in pBluescript |
| pBS Insert (R) | cagtgagcgcaacgcaatta | |
| Ampi (F) | aacccactcgtgcacccaactgat | To sequence across ampicillin gene in pBluescript |
| Ampi (R) | aaacgctggtgaaagtaaaagatgctgaa | |
| pBS 120 (F) | ccgagatagggttgagtgttgttc | To sequence polylinker from further back 5' |
| pBS R2 | gggggaaacgcctggtatc | To sequence polylinker from further back 3' |
| | | |
| pBluSce (F) | gtcccattagggggatccactagttctagagcggcc | For inverse PCR on pBluescript, introducing I–SceI site in–frame in polylinker |
| pBluSce (R) | aatagggatatcaagcttatcgataccgtcgacctcga | |
| pBluSce Inv2 (F) | tatccctagggggatccactagttctagagcg | For inverse PCR on pBluescript, introducing I–SceI site in–frame in polylinker |
| pBluSce Inv2 (R) | acagggtaatatcaagcttatcgataccgtcg | |
| pBluSce linker top | agcttgatatccctattgtcccatta | Forms an oligo to introduce I–Sce I site to polylinker of pBluescript |
| pBluSce linker bottom | taatgggacaatagggatatca | |
| | | |
| known A216 probe (F) | tgggaaaatgagggctgggtgag | To sequence further within specific pBS insertion events |
| known A216 probe (R) | gcacattaggactttggctttga | |
| A2 16 R2 | gccaaagtcctaatgtgcagtca | |
| A2B16–F2 | ccccctaaagaaaaaaaaatgatg | |
| 18inSV40 (F) | aaggggtgggggtgtttaac | |
| 18inSV40 (R) | cccccacccctttttctcattaa | |
| SmaB9 F1 | tgagcctttcccttgcctaaca | |
| GM1 9B | gcgcggtggctcatgtctgtaat | |
| | | |
| SV40 436 (F) | acgccttttgtgtttgttttagag | To sequence from given positions in SV40 genome |
| SV40 3382 (F) | tcccctccagtgccctttac | |
| SV40 4091 (F) | acctggcaaactttcctcaataac | |
| SV40 4933 (F) | atcccagaagcctccaaagtcag | |
| SV40 5217 (F) | ctccaaaaaagcctcctcactact | |

(cont.)

| Oligo Name | Sequence (5' to 3') | Description |
|---|---|---|
| **Chromosomal DSB Assay** | | |
| Sce/RsrII (F) | gaccgattaccctgttatccctacg | Forms an oligo with I–Sce I site and ends cohesive to Rsr II |
| Sce/RsrII (R) | gtccgtagggataacagggtaatcg | |
| Sac to Not (F) | ggcatatgggcgcgccctcgaggc | Forms an oligo to eliminate Sal I site, and introduce Asc I in pNTKV |
| Sac to Not (R) | ggccgcctcgagggcgcgcccatatgccgc | |
| Rsr KO (F) | cagggtaatccgtagagtcgagcagtgtggttttcaaga | For inverse PCR to introduce an I–Sce I site while eliminating Rsr II from pos 2 of pNTKV |
| Rsr KO (R) | ttatccctaaccgtaaaaaggccgcgttgctg | |
| NCBI Rb (F) | cagatgcaattgtttgggtg | To amplify a fragment from human Rb (control for duplex PCR) |
| NCBI Rb (R) | tgaatgggcagtcaatcaaa | |
| C7–1 (F) | tgaggagtccaaggaacaca | To amplify a fragment from human PTPRE (control for duplex PCR) |
| C7–1 (R) | caccagggaaacaccacatt | |
| pSceTA Inverse F | gcggggagaggcggtttgcgtattg | To amplify I–Sce I ORF from pSce TA |
| SceI ORF R wBglII | ctagatctttattatttcaggaaagtttcggaggagat | |
| IRES–GFP F | cgggtggcggtcggggttgt | To amplify IRES–GFP cassette from pAdenoVator |
| IRES–GFP R (Bgl II) | gaagatctgggggcatgcgcgttgtcaaatat | |
| IRPR (F) | ttccgtgtcgcccttattccctttttg | To amplify 4.1 kb internal to linear pNTKV |
| IRPR (R) | gggctgaccgcttcctcgtgctttac | |
| CFP1 | ggcgcggtcccaggtccacttc | To amplify I–Sce I site in pOdin |
| CRP1 | cgggagtttcacgccaccaagatgt | |
| O.DSB (F) | agcggcagggtcggaacaggagagc | To amplify I–Sce I site in pOdin, or 1st site in pZaphod |
| O.DSB (R) | cgcgcggccgggtagcacagg | |
| 2nd Z SceI | gggcgttgcgtggggtcagtc | To sequence the 2nd I–Sce I site in pZaphod (forward) |
| Z.DSB R | ggggatgcggtgggctctatggcttctg | To sequence the 2nd I–Sce I site in pZaphod (reverse) |
| 5' jxn OZ | ccccgaaaagtgccacctgacg | To sequence genomic junctions of pNTKV (linearized Eco RI) insertion sites |
| 3' jxn OZ | ggcgcctaccggtggatgtggaatgtg | |
| 1stZDSBwalk1 | ccccggccatcagcacgcgtct | To sequence (walk) towards 2nd I–SceI site in pZaphod |
| off 3' Z | ccggtggatgtggaatgtgtg | To sequence across the 3' junction of Eco RI cut pNTKV |

(cont.)

| Oligo Name | Sequence (5' to 3') | Description |
|---|---|---|
| **Chromosomal DSB Assay (cont.)** | | |
| GO6 5' wrap | agggtgcggaattcagaagttt | To sequence from within genomic DNA 5' of integrated plasmid, across genomic Bgl II sites (into 3' end of Bgl II plasmid rescue) |
| AO19 5' wrap | accgctgtaagtttctgt | |
| AZ7 5' wrap | ccaggctagtgaaagtacggaca | |
| GZ1 5'jxn A | gcctatccttttgcgcattgctctgtt | To sequence (walk) along the inversion at 5' junction of clone GZ1 |
| 5'jxn inwards | cggccgccatggtcatagctgtt | To sequence inwards from 5' jxn of pNTKV (for inversion in clone GZ1) |

Table AA. List of oligonucleotide primers used in this study. Primers were used for sequencing, for PCR amplification of plasmid segments used in vector construction (including inverse PCR), or annealed to make a double-stranded linker "oligo" for insertion of restriction sites in plasmid constructs. Nomenclature "(F)" for forward primer and "(R)" for reverse primer used to denote direction of product generated, relative to plasmid template sequence.

## 5.2 Appendix B : Vector Construction

### 5.2.1 Insertion of Plasmid Rescue Sites in DSB Template Plasmids

Once integrated into the cellular genome, our DSB substrate needed to be recuperable for analysis of the DSB repair event. As we wished to observe events of insertion of unpredictable sequence, we decided to insert rare restriction sites into the DSB template which would permit later excision of an intact plasmid from the genomic DNA while minimizing the risk of cutting within any inserted sequences. To this end, we introduced the "Sac to Not" oligonucleotide linker, which replaced the Sma I and Sal I sites adjacent to the Eco RI linearization site with Nde I, Asc I, and Xho I sites. The oligos were designed and ordered as follows:

5' ... G G C A T A T G G G C G C G C C C T C G A G G C ... 3'
3' ... C G C C G T A T A C C C G C G C G G G A G C T C C G C C G G ... 5'

The oligos were resuspended in $ddH_2O$, annealed at room temperature, and the ends phosphorylated with T4 Polynucleotide Kinase. Plasmid pNTKV–1901 was digested with both Sac II and Not I, gel purified, dephosphorylated with CIP, and ligated with the oligo in a 1:3 molar ratio.

### 5.2.2 Insertion of I–Sce I Site Between Promoter and TK in pOdin

The construction of pOdin required a unique restriction site between the promoter and TK where a linker could be ligated. Initial digestion of pNTKV–1901 was done with Aat II, producing fragments of 2498 and 3110 bp.

The larger (3110 bp) fragment, harbouring the Ampicillin resistance gene, ColE1 ori and most of TK, was recircularized with T4 DNA Ligase to generate a circular plasmid capable of replicating in bacteria, which now had a single Hinc II site located between the MCI Promoter and TK. This 3110 bp plasmid was linearized with Hinc II (blunt), the ends dephosphorylated with CIP to prevent religation upon itself, and gel purified.

To introduce the I–Sce I site, "Sce/Rsr" oligos harbouring the I–Sce I site (in bold), with ends cohesive to an Rsr II digest, were ordered as follows:

5' ... G T C C G **T A G G G A T A A C A G G G T A A** T C G ... 3'
3' ... G C A T C C C T A T T G T C C C A T T A G C C A G ... 5'

The oligos were resuspended in $ddH_2O$, annealed at room temperature, the overhangs filled in with DNA Polymerase I (large Klenow fragment), and the ends phosphorylated with T4 Polynucleotide Kinase. This blunt, double–stranded oligo was then added to the purified 3110 bp fragment in a 3:1 molar ratio, ligated with T4 DNA Ligase, and electroporated into DH10B bacteria.

Following screening of colonies by alkaline lysis miniprep and restriction digests (to ensure the insertion of an I–Sce I site into the Hinc II), the new plasmid was linearized again by digestion with Aat II. This fragment was mixed in a 1:1 molar ratio with the CIP–treated 2498 bp remaining fragment of pNTKV–1901, and colonies screened as before to give a pNTKV–1901 plasmid which now contained the blunt Sce/Rsr oligo at the Hinc II site between the MCI promoter and TK. Large quantities of plasmid were generated by maxiprep (Qiagen), and linearization done with Eco RI (followed by phenol purification) prior to transfection into human cells.

### *5.2.2.1 pOdin Control*

To ensure that the HSV–TK was still functional after insertion of the I–Sce I site between the MCI promoter and TK, the plasmid was transfected into mouse LTA (TK–) cells, and selected for TK activity in HAT medium. Colonies were obtained at an adequate frequency to suggest TK activity. Additionally, we began to contruct pOdin derivatives with inserts of different sizes at the I–Sce I site, to establish how large an insertion would need to occur *in vivo* in order to inactivate the TK, and permit detection of clones following selection in Ganciclovir. The initial pool of DNA fragments used as inserts came from commercial molecular weight ladders (NEB), which proved ineffective as some of these fragments were plasmid in origin, and their ligation (with pOdin linearized I–Sce I) produced a number of different plasmids, other than pOdin insertions, which complicated analyses. The next pool

of DNA fragments used as inserts came from digested Lambda DNA. However, by this point, we were obtaining clones from some of the mothers, such as AO19, and decided to concentrate on the events before us, foregoing further attempts to establish these insertion size requirements for HSV–TK inactivation.

### 5.2.3 Insertion of Two I–Sce I Sites in pZaphod

In order to obtain a unique restriction site into which the Sce/Rsr linker could be ligated, while at the same time preserving the polylinker sites both 5' and 3' of the TK, we eliminated the Rsr II site 5' of TK and at the same time introduced an I–Sce I site by inverse PCR. The Rsr II site 3' of TK could then be uniquely cleaved, and the Sce/Rsr linker inserted.

#### *5.2.3.1 I–Sce I Site 5' of TK in pZaphod*

The inverse PCR reaction was performed with the Expand High Fidelity PCR System (Roche Applied Science – Laval, QC) on circular template plasmid (modified pNTKV–1901 containing the new "Sac to Not" polylinker 3' of TK, as above). Oligos "RsrKO Forward" and "RsrKO Reverse" were designed which annealed such that the Rsr II site 5' of TK was abolished, and each oligo contained one half of the I–Sce I recognition sequence. Ligation of the resulting 5.6 kb PCR product produced a circular plasmid with an intact I–Sce I site (see Figure AB.1). Oligos were phosphorylated with T4 Kinase prior to reaction. PCR conditions were as follows: 1.25 ng template plasmid, 50 nM each primer, 0.3 µL Expand High Fidelity polymerase, 4 µL dNTP mix (2.5 µM each dATP, dCTP, dGTP, dTTP), 5 µL buffer, total reaction volume 50 µL in $ddH_2O$. 94C for 2 min (1 cycle); 94C for 15 sec, 57C for 30 sec, 68C for 4 min (10 cycles); 94C for 15 sec, 57C for 30 sec, 68C for 4 min + 10 sec added each cycle (15 cycles); 68C for 7 min (1 cycle). The PCR product was then gel purified and resuspended before ligation (T4 Ligase), bacterial transformation, and miniprep. Candidate plasmids were screened by Rsr II and I–Sce I digestion to ensure each enzyme cut only once.

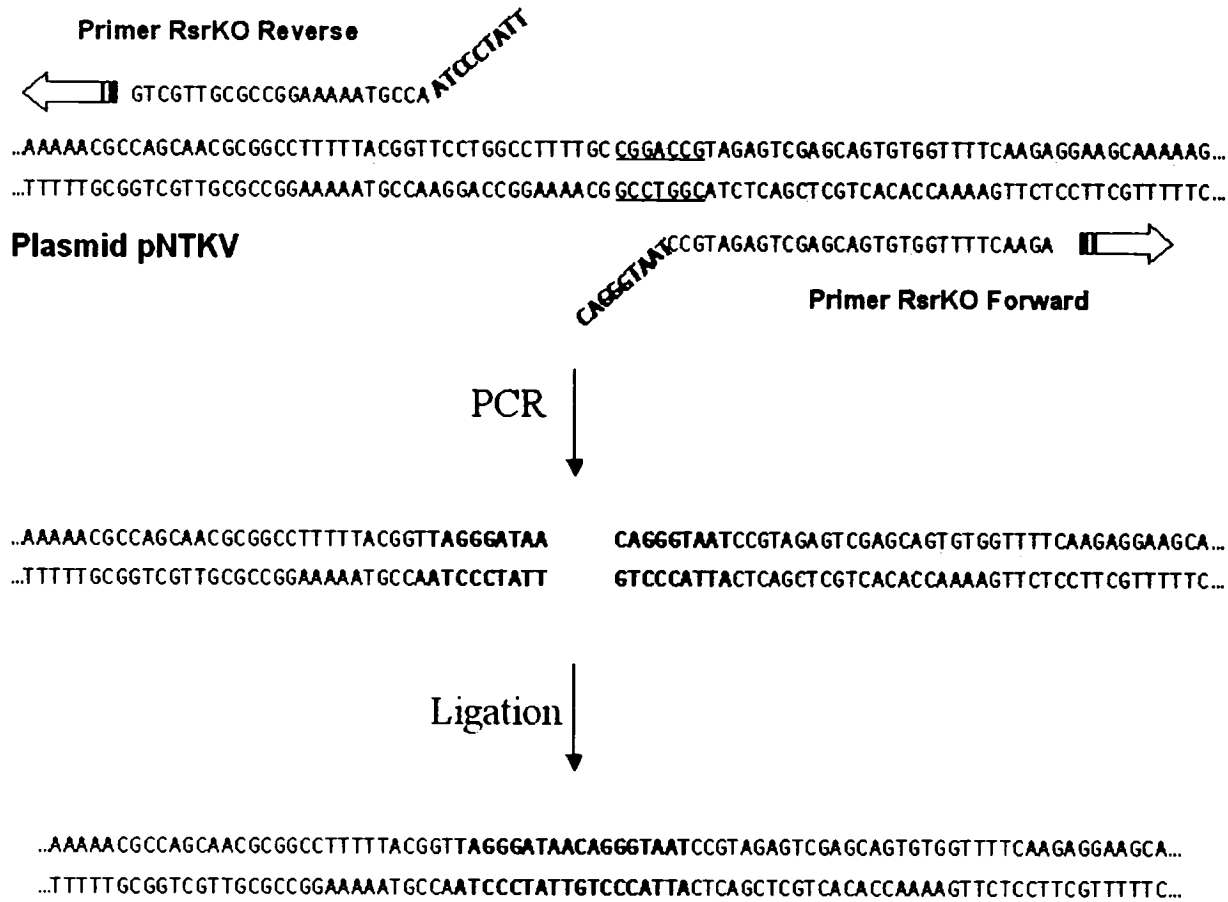



Figure AB.1. "RsrKO" Inverse PCR to generate a 5.6 kb plasmid with an I–Sce I site in place of the Rsr II site at pos 0 in pNTKV. Primers with 5' ends nonhomologous to the template pNTKV each introduced one half of the I–Sce I recognition sequence, which created an intact I–Sce I site upon ligation of the resulting PCR product. Rsr II recognition sequence is underlined, I–Sce I recognition sequence in bold.

### *5.2.3.2 I–Sce I Site 3' of TK in pZaphod*

The plasmid produced from the RsrKO Inverse PCR was linearized by digestion with Rsr II, dephosphorylated with CIP, and gel purified. "Sce/Rsr" oligos were phosphorylated and annealed (producing cohesive Rsr II ends), and ligated together with the linear plasmid. The resulting plasmid pZaphod now has a 2$^{nd}$ I–Sce I site 3' of the TK cassette, is immediately flanked by Rsr II sites on both sides, and is in the reverse complement orientation to the I–Sce I site produced from the RsrKO Inverse PCR (see Figure AB.2).

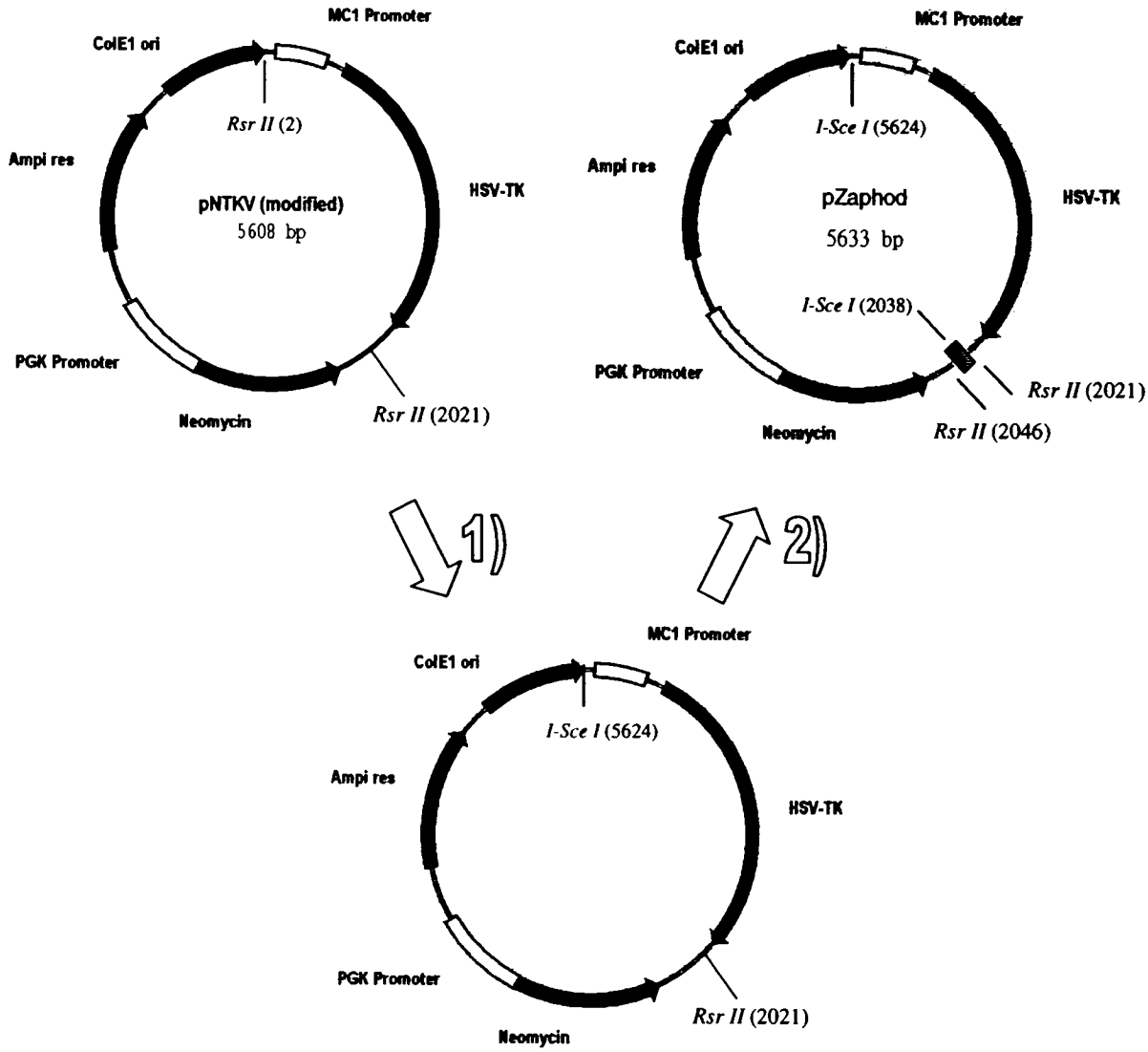



Figure AB.2. Construction of pZaphod. 1) The pNTKV plasmid (modified by inclusion of the "Sac to Not" oligo with restriction sites) was used as template for the RsrKO Inverse PCR (see Fig AB.1). 2) This plasmid then had a unique Rsr II site after the HSV–TK, which was cut, the plasmid dephosphorylated, and then ligated with the "Sce/Rsr" oligo to create the modified pNTKV with two I–Sce I sites flanking the PGK promoter and HSV–TK, now called pZaphod.

The goal of pZaphod was to produce a plasmid which would not easily ligate upon itself following cleavage with I–Sce I, and favour DSB misrepair events (which might include

insertions). As such, when the two I–Sce I sites in pZaphod are cleaved to produce a complete excision of the TK cassette, the non–cohesive ends in the remaining plasmid would be as follows:

Prior to TK cassette..... ..... After TK cassette

5' ... A C G G T T A G G G A T A A C C C T A C G G A C ... 3'
3' ... T G C C A A T C C C A A T A G G G A T G C C T G ... 5'

### 5.2.4 I-Sce I Expression Plasmid (Using FACS)

Previous work in our lab which used I–Sce I expression to generate DSBs employed the pβactin–SceI plasmid (a gift of Dr. Maria Jasin – Memorial Sloan–Kettering Cancer Center, NY) [89,100]. These studies were principally done in rodent cells with adequate cutting efficiency, but when we began applying the plasmid to human cell cultures we encountered greatly reduced efficiency of I–Sce I cutting. Repeated rounds of transfection followed by plasmid rescue and PCR assessment of a band resistant to further I–Sce I digestion, all gave negative results, suggesting little or no cutting was taking place in the human cells.

Our desire was to improve the efficiency of expression of I–Sce I in human cells. This was done in 2 stages: 1) an adenovirus–based plasmid (pAdSIG), which ultimately failed to express I–Sce I effectively, then 2) a modification of the pβactin–SceI plasmid (pβSIG) to include green fluorescent protein (GFP) expression, that enabled adequate expression of I–Sce I as well as enrichment of the transfected cell population by FACS.

#### *5.2.4.1 pAdSIG Construction*

The initial system we chose was to develop a new expression vector based on adenoviral sequences (see Figure AB.3). The pAdenoVator system (Qbiogene – Montreal, QC) was chosen for 2 reasons: 1) it includes a construct into which the I–Sce I ORF could be inserted upstream of an internal ribosomal entry site (IRES) and GFP, such that FACS could be used to enrich the cell population for those which were successfully transfected and expressing I–Sce I, and 2) we could use the cloning vector directly, with the I–Sce I ORF, and continue

transfection by electroporation, but failing that we could purchase the remainder of the kit to generate an adenovirus that expressed I–Sce I and GFP.

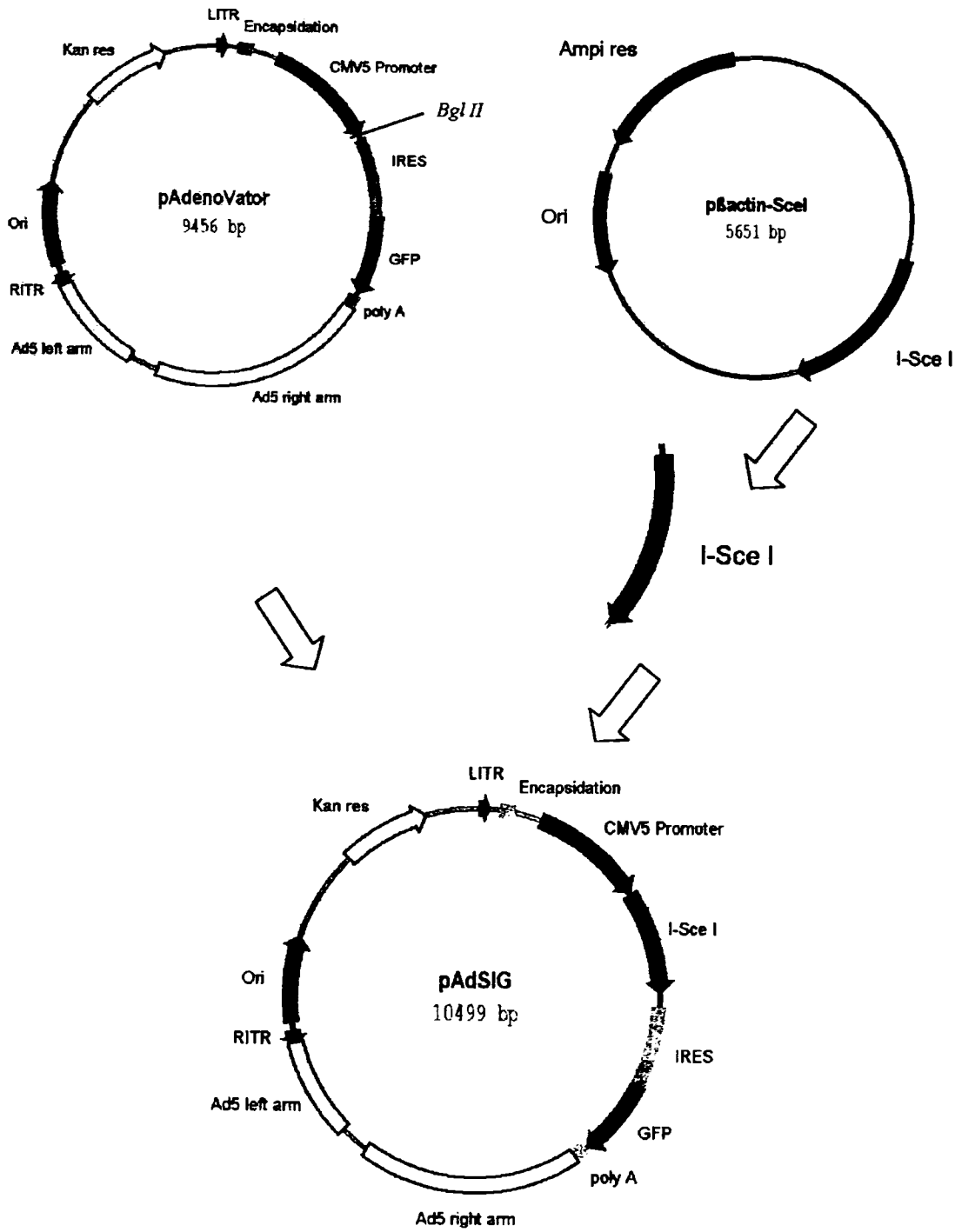



Figure AB.3. Construction of pAdSIG. The pAdenoVator CMV5–IRES–GFP plasmid was linearized Bgl II. The I–Sce I gene was isolated from the original pβactin–SceI plasmid by PCR amplification with primers which introduced Bgl II sites on either side. Cohesive ligation of the Bgl II–digested PCR product into Bgl II–linearized and CIP–treated pAdenoVator CMV5–IRES–GFP produced pAdSIG.

### *5.2.4.2 pβSIG Construction*

The second method to improve I–Sce I expression by sorting on FACS was to take the IRES–GFP components from the pAdenoVator and insert them into pβactin–SceI (see Figure AB.4). In this way, we were assured the I–Sce I would be in–frame (as it was left untouched), and any GFP expression guaranteed the I–Sce I was being expressed (as the GFP and I–Sce I would have to be encoded by the same RNA). The original pβactin–SceI plasmid conferred ampicillin resistance to host bacteria, which would have interfered with our plasmid rescue of pOdin or pZaphod. Therefore we replaced ampicillin with tetracycline resistance by excising the majority of the ampicillin resistance gene and placing the tetracycline resistance gene in between.

The CMV5–IRES–GFP plasmid was digested Pst I, and the ends blunted by treatment with T4 Polymerase, before extracting the 5.1 kb band on gel and phenol cleaning. Once cleaned, this 5.1 kb fragment was digested Bgl II, and the 1.8 kb band encoding the IRES–GFP cassette gel purified.

The original pβactin–SceI plasmid is a pUC–based plasmid which codes for ampicillin resistance (as well as I–Sce I). It had previously been modified in our lab by Hélène D'Anjou, whereby she amplified the tetracycline resistance gene from pBr322 by PCR, and cloned it into the middle of the ampicillin resistance gene in pβactin–SceI. Another lab member, Hugo Wurtele, had subsequently excised the majority of the ampicillin resistance gene from both sides via digestion with Ssp I and Ppu MI/Ahd I, then religation (named pSceTdelA) .

To insert the IRES–GFP cassette, pSceTdelA was digested Bgl II and Stu I (to excise 800 bp directly following the I–Sce I gene and leaving a Bgl II cohesive end 5' with a blunt end 3' at the break site), and the ends dephosphorylated with CIP. The IRES–GFP cassette, which has a Bgl II cohesive site 5' and a blunt site 3', was then ligated into the cut pSceTdelA, to produce a modified pβactin–SceI with I–Sce I followed by the IRES–GFP, which we called pβSIG.

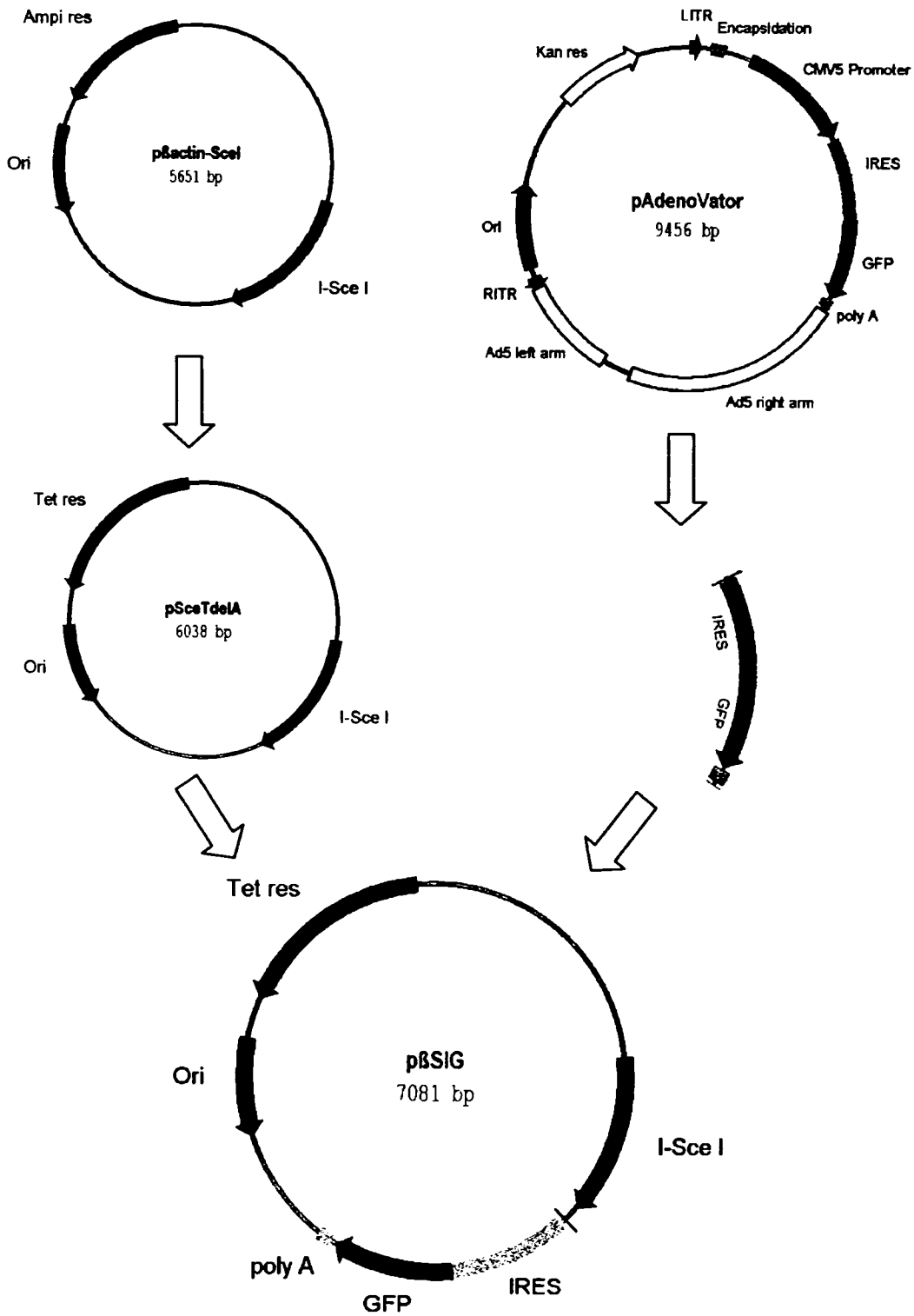



Figure AB.4. Construction of pβSIG. The original pβactin–SceI plasmid was modified by replacing ampicillin resistance with tetracycline resistance, to give pSceTdelA. The IRES–GFP encoding fragment was excised from pAdenoVator CMV5–IRES–GFP, and ligated into the Bgl II site after the I–Sce I gene in pSceTdelA, to produce pβSIG.

### *5.2.4.3 Comparison of pAdSIG and pβSIG*

Transfections were done in parallel to compare the efficiency of GFP expression, at different DNA concentrations, for the two plasmids. In 6–well dishes, $3.5x10^5$ cells per sample (GM05849) were each electroporated with 10, 25, and 50 μg pβSIG, 15 μg of pAdSIG, 15 μg of the original pAdenoVator CMV5–IRES–GFP, and 25 μg pEGFP–N1 (BD Biosciences). The pEGFP–N1 plasmid was frequently used in our lab as a control to assure we could properly detect GFP expression from our other constructs. pEGFP–N1 expresses an enhanced GFP and fluoresces more intensely than our constructs. (Only in retrospect did we realize pEGFP–N1 also contains the SV40 origin of replication, hence this plasmid will replicate in GM05849 cells, which accounts for the much higher levels of expression observed – see fluorescence microscopy image in the chromosomal DSB assay Results section, Figure C2).

A comparison between the stock pAdenoVator and pAdSIG was performed by co–transfection of 20 μg of each with 2 μg of pZaphod into GM00637 and GM05849 cells, followed by 48 hrs incubation and then Hirt extraction. By fluorescence microscopy, approximately 10% of cells showed any degree of GFP expression, with very weak signal overall and slightly less from pAdSIG than from pAdenoVator. The Hirt extracts were enriched by digestion with Nru I (which cleaves within the HSV–TK) prior to transformation into bacteria and plating on ampicillin (for pZaphod). Analysis of 128 colonies from the GM00637 Hirt and another 128 colonies from GM05849 Hirt (transfected pAdSIG + pZaphod and enriched Nru I) revealed none had lost the HSV–TK cassette between the two I–Sce I sites, hence no proof that this first version of pAdSIG could express functional I–Sce I.

The next test was transfection of clone GO6 with 40 μg pAdenoVator, 40 μg of pAdSIG, or 21 μg of pβactin–SceI (approximately molar equivalents).

The pAdSIG candidates from the original ligation of the Sce ORF PCR product with pAdenoVator were taken and 8 candidates were co–transfected (20 μg of each) into GM05849 along with 2 μg of pOdin.

A series of three further pAdSIG candidates (generated from the same original ligation of the Sce ORF PCR product with pAdenoVator) were tested, by co–transfection into GM05849 cells with pZaphod or pOdin, followed by Hirt extraction at 96 hrs, I–Sce I enrichment, transformation into bacteria, and plating on ampicillin + LB (for pZaphod or pOdin). Alkaline miniprep and restriction digests on 32 colonies from each of the three candidates showed none of these had lost the I–Sce I site. We were informed from the technical assistance of the pAdenoVator provider that, without a gene cloned into the site before the IRES, the GFP would not be expressed. However, the near total lack of any GFP detection from either the pAdenoVator or pAdSIG candidates, concurrent with the lack of any I–Sce I site loss from the Hirt extracted pOdin or pZaphod, or from any of the mother clones transfected with pAdSIG, led to the conclusion that this plasmid would not be useful. We thus pursued the development and use of pβSIG further.

Co–transfections of pβSIG and pBluSce were also performed, to test for both the functionality of the GFP expression (by fluorescence microscopy) and I–Sce I cleavage capability (by Hirt recovery of pBluSce plasmids and screening for loss or modification of the I–Sce I site). The plasmid pBluSce was developed as a pBluescript plasmid with an I–Sce I site in the polylinker, to be used to better correlate the findings of the extrachromosomal DSB assay with the chromosomal one, by generating the same kind of DSB (an I–Sce I cut) *in vivo*, rather than introducing an already–linear plasmid to the cells. The testing of pβSIG on pBluSce occurred at the same time as testing on AO19, which quickly showed pβSIG to be functional. As many results were also being obtained at this time from the extrachromosomal assay with pBluescript (cleaved Sma I), the use of pBluSce was not pursued any further.

# BIBLIOGRAPHY


1. Lander,E.S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).

2. Venter,J.C. *et al.* The Sequence of the Human Genome. *Science* **291**, 1304–1351 (2001).

3. Doolittle,W.F. & Sapienza,C. Selfish genes, the phenotype paradigm and genome evolution. *Nature* **284**, 601–603 (1980).

4. Orgel,L.E. & Crick,F.H.C. Selfish DNA: The ultimate parasite. *Nature (London)* **284**, 604–607 (1980).

5. Bork,P. & Copley,R. The draft sequences. Filling in the gaps. *Nature* **409**, 818–820 (2001).

6. Celniker,S.E. The Drosophila genome. *Curr. Opin. Genet. Dev.* **10**, 612–616 (2000).

7. The C.elegans Sequencing Consortium. Genome Sequence of the Nematode *C.elegans*: A Platform for Investigating Biology. *Science* **282**, 2012–2018 (1998).

8. Copley,R.R., Doerks,T., Letunic,I. & Bork,P. Protein domain analysis in the era of complete genomes. *FEBS Lett.* **513**, 129–134 (2002).

9. Rubin,G.M. The draft sequences. Comparing species. *Nature* **409**, 820–821 (2001).

10. Baltimore,D. Our genome unveiled. *Nature* **409**, 814–816 (2001).

11. Cho,K.S., Elizondo,L.I. & Boerkoel,C.F. Advances in chromatin remodeling and human disease. *Curr. Opin. Genet. Dev.* **14**, 308–315 (2004).

12. Ehrenhofer–Murray,A.E. Chromatin dynamics at DNA replication, transcription and repair. *Eur. J. Biochem.* **271**, 2335–2349 (2004).

13. Workman,J.L. & Kingston,R.E. Alteration of nucleosome structure as a mechanism of transcriptional regulation. *Annu. Rev. Biochem.* **67**, 545–579 (1998).

14. Wolffe,A.P. & Guschin,D. Review: Chromatin Structural Features and Targets That Regulate Transcription. *J. Struct. Biol.* **129**, 102–122 (2000).

15. Grewal,S.I.S. & Moazed,D. Heterochromatin and Epigenetic Control of Gene Expression. *Science* **301**, 798–802 (2003).

16. Schubeler,D. *et al.* Nuclear localization and histone acetylation: a pathway for chromatin opening and transcriptional activation of the human beta –globin locus. *Gene. Dev.* **14**, 940–950 (2000).

17. Lin,C.M., Fu,H., Martinovsky,M., Bouhassira,E. & Aladjem,M.I. Dynamic alterations of replication timing in mammalian cells. *Curr. Biol.* **13**, 1019–1028 (2003).

18. Koering,C.E. *et al.* Human telomeric position effect is determined by chromosomal context and telomeric chromatin integrity. *EMBO Rep.* **3**, 1055–1061 (2002).

19. Henikoff,S., Furuyama,T. & Ahmad,K. Histone variants, nucleosome assembly and epigenetic inheritance. *Trends Genet.* **20**, 320–326 (2004).

20. Jablonka,E. & Lamb,M.J. The inheritance of acquired epigenetic variations. *J Theor. Biol.* **139**, 69–83 (1989).

21. Jones,P.A. & Laird,P.W. Cancer epigenetics comes of age. *Nat. Genet.* **21**, 163–167 (1999).

22. Jablonka,E.V.A. & Lamb,M.J. The Changing Concept of Epigenetics. *Ann. NY Acad. Sci.* **981**, 82–96 (2002).

23. Fahrner,J.A., Eguchi,S., Herman,J.G. & Baylin,S.B. Dependence of Histone Modifications and Gene Expression on DNA Hypermethylation in Cancer. *Cancer Res.* **62**, 7213–7218 (2002).

24. Palstra,R.J. *et al.* The beta–globin nuclear compartment in development and erythroid differentiation. *Nat. Genet.* **35**, 190–194 (2003).

25. Alberts,B. *et al.* Molecular Biology of the Cell. Garland Science, New York, NY (2002).

26. Bednar,J. *et al.* Nucleosomes, linker DNA, and linker histone form a unique structural motif that directs the higher–order folding and compaction of chromatin. *Proc. Natl. Acad. Sci. USA* **95**, 14173–14178 (1998).

27. Manuelidis,L. & Chen,T.L. A unified model of eukaryotic chromosomes. *Cytometry* **11**, 8–25 (1990).

28. Yokota,H., van den Engh,G., Hearst,J.E., Sachs,R.K. & Trask,B.J. Evidence for the organization of chromatin in megabase pair–sized loops arranged along a random walk path in the human G0/G1 interphase nucleus. *J. Cell Biol.* **130**, 1239–1249 (1995).

29. Pederson,T. The spatial organization of the genome in mammalian cells. *Curr. Opin. Genet. Dev.* **14**, 203–209 (2004).

30. Cremer,M. *et al.* Three dimensional analysis of histone methylation patterns in normal and tumor cell nuclei. *Eur. J. Histochem.* **48**, 15–28 (2004).

31. Nagele,R.G. *et al.* Chromosomes exhibit preferential positioning in nuclei of quiescent human cells. *J. Cell Sci.* **112 ( Pt 4)**, 525–535 (1999).

32. Leitch,A.R., Brown,J.K.M., Mosgoller,W., Schwarzacher,T. & Heslop–Harrison,J.S. The spatial localization of homologous chromosomes in human fibroblasts at mitosis. *Hum. Genet.* **93**, 275–280 (1994).

33. Pederson,T. Half a century of 'the nuclear matrix'. *Mol. Biol. Cell* **11**, 799–805 (2000).

34. Hozak,P., Sasseville,A.M., Raymond,Y. & Cook,P.R. Lamin proteins form an internal nucleoskeleton as well as a peripheral lamina in human cells. *J. Cell Sci.* **108**, 635–644 (1995).

35. Goldman,R.D., Gruenbaum,Y., Moir,R.D., Shumaker,D.K. & Spann,T.P. Nuclear lamins: building blocks of nuclear architecture. *Gene. Dev.* **16**, 533–547 (2002).

36. Hozak,P., Hassan,A.B., Jackson,D.A. & Cook,P.R. Visualization of replication factories attached to nucleoskeleton. *Cell* **73**, 361–373 (1993).

37. Haaf,T., Golub,E.I., Reddy,G., Readdin,C.M. & Ward,D.C. Nuclear Foci of Mammalian Rad51 Recombination Protein in Somatic cells After DNA Damage and its Localization in Synaptonemal Complexes. *Proc. Natl. Acad. Sci. USA* **92**, 2298–2302 (1995).

38. Lamond,A.I. & Earnshaw,W.C. Structure and Function in the Nucleus. *Science* **280**, 547–553 (1998).

39. Wei,X. *et al.* Segregation of Transcription and Replication Sites Into Higher Order Domains. *Science* **281**, 1502–1505 (1998).

40. Jackson,D.A. & Pombo,A. Replicon Clusters Are Stable Units of Chromosome Structure: Evidence That Nuclear Organization Contributes to the Efficient Activation and Propagation of S Phase in Human Cells. *J. Cell Biol.* **140**, 1285–1295 (1998).

41. Hozak,P. & Cook,P.R. Replication factories. *Trends Cell Biol.* **4**, 48–52 (1994).

42. Stauffer,M.E. & Chazin,W.J. Structural Mechanisms of DNA Replication, Repair, and Recombination. *J. Biol. Chem.* **279**, 30915–30918 (2004).

43. Croft,J.A. *et al.* Differences in the Localization and Morphology of Chromosomes in the Human Nucleus. *J. Cell Biol.* **145**, 1119–1131 (1999).

44. Bridger,J.M., Boyle,S., Kill,I.R. & Bickmore,W.A. Re–modelling of nuclear architecture in quiescent and senescent human fibroblasts. *Curr. Biol.* **10**, 149–152 (2000).

45. Sun,H.B., Shen,J. & Yokota,H. Size–dependent positioning of human chromosomes in interphase nuclei. *Biophys. J.* **79**, 184–190 (2000).

46. Parada,L.A., Roix,J.J. & Misteli,T. An uncertainty principle in chromosome positioning. *Trends Cell Biol.* **13**, 393–396 (2003).

47. Cremer,M. *et al.* Inheritance of gene density–related higher order chromatin arrangements in normal and tumor cell nuclei. *J. Cell Biol.* **162**, 809–820 (2003).

48. Gerlich,D. & Ellenberg,J. Dynamics of chromosome positioning during the cell cycle. *Curr. Op. Cell Biol.* **15**, 664–671 (2003).

49. Kurz,A. *et al.* Active and inactive genes localize preferentially in the periphery of chromosome territories. *J. Cell Biol.* **135**, 1195–1205 (1996).

50. Verschure,P.J. *et al.* Large–scale Chromatin Organization and the Localization of Proteins Involved in Gene Expression in Human Cells. *J. Histochem. Cytochem.* **50**, 1303–1312 (2002).

51. Mahy,N.L., Perry,P.E. & Bickmore,W.A. Gene density and transcription influence the localization of chromatin outside of chromosome territories detectable by FISH. *J. Cell Biol.* **159**, 753–763 (2002).

52. Mahy,N.L., Perry,P.E., Gilchrist,S., Baldock,R.A. & Bickmore,W.A. Spatial organization of active and inactive genes and noncoding DNA within chromosome territories. *J. Cell Biol.* **157**, 579–589 (2002).

53. Tanabe,H. *et al.* Evolutionary conservation of chromosome territory arrangements in cell nuclei from higher primates. *Proc. Natl. Acad. Sci. USA* **99**, 4424–4429 (2002).

54. Nikiforova,M.N. *et al.* Proximity of Chromosomal Loci That Participate in Radiation–Induced Rearrangements in Human Cells. *Science* **290**, 138–141 (2000).

55. Kozubek,S. *et al.* The topological organization of chromosomes 9 and 22 in cell nuclei has a determinative role in the induction of t(9,22) translocations and in the pathogenesis of t(9,22) leukemias. *Chromosoma* **108**, 426–435 (1999).

56. Parada,L.A., McQueen,P.G., Munson,P.J. & Misteli,T. Conservation of Relative Chromosome Positioning in Normal and Cancer Cells. *Curr. Biol.* **12**, 1692–1697 (2002).

57. Roix,J.J., McQueen,P.G., Munson,P.J., Parada,L.A. & Misteli,T. Spatial proximity of translocation–prone gene loci in human lymphomas. *Nat. Genet.* **34**, 287–291 (2003).

58. Hernandez–Verdun,D., Roussel,P. & Gebrane–Younes,J. Emerging concepts of nucleolar assembly. *J. Cell Sci.* **115**, 2265–2270 (2002).

59. Haaf,T. & Ward,D.C. Inhibition of RNA Polymerase II Transcription Causes Chromatin Decondensation, Loss of Nucleolar Structure, and Dispersion of Chromosomal Domains. *Exp. Cell Res.* **224**, 163–173 (1996).

60. Khanna,K.K. & Jackson,S.P. DNA double–strand breaks: signaling, repair and the cancer connection. *Nat. Genet.* **27**, 247–254 (2001).

61. Roth,D.B. Restraining the V(D)J recombinase. *Nat. Rev. Immunol.* **3**, 656–666 (2003).

62. Wang,J.C. Cellular roles of DNA topoisomerases: a molecular perspective. *Nat. Rev. Mol. Cell. Biol.* **3**, 430–440 (2002).

63. Vilenchik,M.M. & Knudson,A.G. Endogenous DNA double-strand breaks: production, fidelity of repair, and induction of cancer. *Proc. Natl. Acad. Sci. U. S. A* **100**, 12871–12876 (2003).

64. Huang,L., Clarkin,K.C. & Wahl,G.M. Sensitivity and selectivity of the DNA damage sensor responsible for activating p53–dependent G1 arrest. *Proc. Natl. Acad. Sci. USA* **93**, 4827–4832 (1996).

65. Rich,T., Allen,R.L. & Wyllie,A.H. Defying death after DNA damage. *Nature* **407**, 777–783 (2000).

66. Merrick,C.J., Jackson,D. & Diffley,J.F.X. Visualization of Altered Replication Dynamics after DNA Damage in Human Cells. *J. Biol. Chem.* **279**, 20067–20075 (2004).

67. Durocher,D. & Jackson,S.P. DNA–PK, ATM and ATR as sensors of DNA damage: variations on a theme? *Curr. Op. Cell Biol.* **13**, 225–231 (2001).

68. Shiloh,Y. ATM and related protein kinases: safeguarding genome integrity. *Nat. Rev. Cancer* **3**, 155–168 (2003).

69. Panta,G.R. *et al.* ATM and the Catalytic Subunit of DNA–Dependent Protein Kinase Activate NF–{kappa}B through a Common MEK/Extracellular Signal–Regulated Kinase/p90rsk Signaling Pathway in Response to Distinct Forms of DNA Damage. *Mol. Cell. Biol.* **24**, 1823–1835 (2004).

70. Aleyasin,H. *et al.* Nuclear Factor–{kappa}B Modulates the p53 Response in Neurons Exposed to DNA Damage. *J. Neurosci.* **24**, 2963–2973 (2004).

71. Smith,G.C.M. *et al.* Purification and DNA binding properties of the ataxia–telangiectasia gene product ATM. *Proc. Natl. Acad. Sci. USA* **96**, 11134–11139 (1999).

72. Lukas,C., Falck,J., Bartkova,J., Bartek,J. & Lukas,J. Distinct spatiotemporal dynamics of mammalian checkpoint regulators induced by DNA damage. *Nat. Cell Biol* **5**, 255–260 (2003).

73. Ahn,J., Urist,M. & Prives,C. The Chk2 protein kinase. *DNA Repair* **3**, 1039–1047 (2004).

74. Bakkenist,C.J. & Kastan,M.B. DNA damage activates ATM through intermolecular autophosphorylation and dimer dissociation. *Nature* **421**, 499–506 (2003).

75. Rogakou,E.P., Boon,C., Redon,C. & Bonner,W.M. Megabase Chromatin Domains Involved in DNA Double–Strand Breaks In Vivo. *J. Cell Biol.* **146**, 905–916 (1999).

76. Celeste,A. *et al.* Histone H2AX phosphorylation is dispensable for the initial recognition of DNA breaks. *Nat. Cell Biol* **5**, 675–679 (2003).

77. Burma,S., Chen,B.P., Murphy,M., Kurimasa,A. & Chen,D.J. ATM Phosphorylates Histone H2AX in Response to DNA Double-strand Breaks. *J. Biol. Chem.* **276**, 42462–42467 (2001).

78. Ward,I.M. & Chen,J. Histone H2AX Is Phosphorylated in an ATR-dependent Manner in Response to Replicational Stress. *J. Biol. Chem.* **276**, 47759–47762 (2001).

79. Stiff,T. *et al.* ATM and DNA-PK Function Redundantly to Phosphorylate H2AX after Exposure to Ionizing Radiation. *Cancer Res.* **64**, 2390–2396 (2004).

80. Nelms,B.E., Maser,R.S., MacKay,J.F., Lagally,M.G. & Petrini,J.H. In Situ Visualization of DNA Double-Strand Break Repair in Human Fibroblasts. *Science* **280**, 590–592 (1998).

81. Aten,J.A. *et al.* Dynamics of DNA Double-Strand Breaks Revealed by Clustering of Damaged Chromosome Domains. *Science* **303**, 92–95 (2004).

82. Christmann,M., Tomicic,M.T., Roos,W.P. & Kaina,B. Mechanisms of human DNA repair: an update. *Toxicology* **193**, 3–34 (2003).

83. Kruger,I., Rothkamm,K. & Lobrich,M. Enhanced fidelity for rejoining radiation-induced DNA double-strand breaks in the G2 phase of Chinese hamster ovary cells. *Nucl. Acids. Res.* **32**, 2677–2684 (2004).

84. Flygare,J., Benson,F. & Hellgren,D. Expression of the human RAD51 gene during the cell cycle in primary human peripheral blood lymphocytes. *Biochim. Biophys. Acta* **1312**, 231–236 (1996).

85. Saleh-Gohari,N. & Helleday,T. Conservative homologous recombination preferentially repairs DNA double-strand breaks in the S phase of the cell cycle in human cells. *Nucl. Acids. Res.* **32**, 3683–3688 (2004).

86. Rothkamm,K. & Lobrich,M. From the Cover: Evidence for a lack of DNA double-strand break repair in human cells exposed to very low x-ray doses. *Proc. Natl. Acad. Sci. USA* **100**, 5057–5062 (2003).

87. Allen,C., Miller,C.A. & Nickoloff,J.A. The mutagenic potential of a single DNA double-strand break in a mammalian chromosome is not influenced by transcription. *DNA Repair (Amst)* **2**, 1147–1156 (2003).

88. Leuther,K.K., Hammarsten,O., Kornberg,R.D. & Chu,G. Structure of DNA-dependent protein kinase: implications for its regulation by DNA. *EMBO J.* **18**, 1114–1123 (1999).

89. Tremblay,A., Jasin,M. & Chartrand,P. A double-strand break in a chromosomal LINE element can be repaired by gene conversion with various endogenous LINE elements in mouse cells. *Mol. Cell. Biol.* **20**, 54–60 (2000).

90. Valerie,K. & Povirk,L.F. Regulation and mechanisms of mammalian double–strand break repair. *Oncogene* **22**, 5792–5812 (2003).

91. Dyck,E.V., Stasiak,A.Z., Stasiak,A. & West,S.C. Binding of double–strand breaks in DNA by human Rad52 protein. *Nature* **398**, 728–731 (1999).

92. Gottlieb,T.M. & Jackson,S.P. The DNA–dependent protein kinase: requirement for DNA ends and association with Ku antigen. *Cell* **72**, 131–142 (1993).

93. Novac,O., Matheos,D., Araujo,F.D., Price,G.B. & Zannis–Hadjopoulos,M. In Vivo Association of Ku with Mammalian Origins of DNA Replication. *Mol. Biol. Cell* **12**, 3386–3401 (2001).

94. Coffey,G. & Campbell,C. An alternate form of Ku80 is required for DNA end–binding activity in mammalian mitochondria. *Nucl. Acids. Res.* **28**, 3793–3800 (2000).

95. Mirzoeva,O.K. & Petrini,J.H.J. DNA Replication–Dependent Nuclear Dynamics of the Mre11 Complex. *Mol. Cancer Res.* **1**, 207–218 (2003).

96. Kobayashi,J. *et al.* NBS1 Localizes to [gamma]–H2AX Foci through Interaction with the FHA/BRCT Domain. *Curr. Biol.* **12**, 1846–1851 (2002).

97. Roth,D.B. & Wilson,J.H. Nonhomologous recombination in mammalian cells: role for short sequence homologies in the joining reaction. *Mol. Cell Biol.* **6**, 4295–4304 (1986).

98. Lukacsovich,T., Yang,D. & Waldman,A.S. Repair of a specific double–strand break generated within a mammalian chromosome by yeast endonuclease I–SceI. *Nucl. Acids. Res.* **22**, 5649–5657 (1994).

99. Dellaire,G., Lemieux,N., Belmaaza,A. & Chartrand,P. Ectopic gene targeting exhibits a bimodal distribution of integration in murine cells, indicating that both intra– and interchromosomal sites are accessible to the targeting vector. *Mol. Cell Biol.* **17**, 5571–5580 (1997).

100. Dellaire,G., Yan,J., Little,K.C., Drouin,R. & Chartrand,P. Evidence that extrachromosomal double–strand break repair can be coupled to the repair of chromosomal double–strand breaks in mammalian cells. *Chromosoma* **111**, 304–312 (2002).

101. Dellaire,G. & Chartrand,P. Direct evidence that transgene integration is random in murine cells, implying that naturally occurring double–strand breaks may be distributed similarly within the genome. *Radiat. Res.* **149**, 325–329 (1998).

102. Gusew,N., Nepveu,A. & Chartrand,P. Linear DNA must have free ends to transform rat cells efficiently. *Mol. Gen. Genet.* **206**, 121–125 (1987).

103. Belmaaza,A., Wallenburg,J.C., Brouillette,S., Gusew,N. & Chartrand,P. Genetic exchange between endogenous and exogenous LINE–1 repetitive elements in mouse cells. *Nucl. Acids. Res.*. **18**, 6385–6391 (1990).

104. Belmaaza,A., Milot,E., Villemure,J.F. & Chartrand,P. Interference of DNA sequence divergence with precise recombinational DNA repair in mammalina cells. *EMBO J.* **13**, 5355–5360 (1994).

105. Richard,M., Belmaaza,A., Gusew,N., Wallenburg,J.C. & Chartrand,P. Integration of a vector containing a repetitive LINE–1 element in the human genome. *Mol. Cell. Biol.* **14**, 6689–6695 (1994).

106. Ostertag,E.M. & Kazazian,H.H., Jr. Biology of mammalian L1 retrotransposons. *Annu. Rev. Genet.* **35**, 501–538 (2001).

107. Rouet,P., Smih,F. & Jasin,M. Introduction of double–strand breaks into the genome of mouse cells by expression of a rare–cutting endonuclease. *Mol. Cell Biol.* **14**, 8096–8106 (1994).

108. Rouet,P., Smih,F. & Jasin,M. Expression of a Site–Specific Endonuclease Stimulates Homologous Recombination in Mammalian Cells. *Proc. Natl. Acad. Sci. USA* **91**, 6064–6068 (1994).

109. Gorbunova,V. & Levy,A.A. Non–homologous DNA end joining in plant cells is associated with deletions and filler DNA insertions. *Nucl. Acids. Res.*. **25**, 4650–4657 (1997).

110. Salomon,S. & Puchta,H. Capture of genomic and T–DNA sequences during double–strand break repair in somatic plant cells. *EMBO J.* **17**, 6086–6095 (1998).

111. Ricchetti,M., Fairhead,C. & Dujon,B. Mitochondrial DNA repairs double–strand breaks in yeast chromosomes. *Nature* **402**, 96–100 (1999).

112. Yu,X. & Gabriel,A. Patching broken chromosomes with extranuclear cellular DNA. *Mol. Cell* **4**, 873–881 (1999).

113. Moore,J.K. & Haber,J.E. Capture of retrotransposon DNA at the sites of chromosomal double–strand breaks. *Nature* **383**, 644–646 (1996).

114. Sargent,R.G., Brenneman,M.A. & Wilson,J.H. Repair of site–specific double–strand breaks in a mammalian chromosome by homologous and illegitimate recombination. *Mol. Cell Biol.* **17**, 267–277 (1997).

115. Liang,F., Han,M., Romanienko,P.J. & Jasin,M. Homology–directed repair is a major double–strand break repair pathway in mammalian cells. *Proc. Natl. Acad. Sci. USA* **95**, 5172–5177 (1998).

116. Lin,Y. & Waldman,A.S. Capture of DNA sequences at double–strand breaks in mammalian chromosomes. *Genetics* **158**, 1665–1674 (2001).

117. Lin,Y. & Waldman,A.S. Promiscuous patching of broken chromosomes in mammalian cells with extrachromosomal DNA. *Nucl. Acids. Res.* **29**, 3975–3981 (2001).

118. Pandita,T.K. *et al.* Ionizing radiation activates the ATM kinase throughout the cell cycle. *Oncogene* **19**, 1386–1391 (2000).

119. Duker,N.J. Chromosome breakage syndromes and cancer. *Am. J Med. Genet.* **115**, 125–129 (2002).

120. Meyn,M.S. Ataxia–telangiectasia, cancer and the pathobiology of the ATM gene. *Clin. Genet.* **55**, 289–304 (1999).

121. Lobrich,M., Kuhne,M., Wetzel,J. & Rothkamm,K. Joining of correct and incorrect DNA double–strand break ends in normal human and ataxia telangiectasia fibroblasts. *Gene. Chromosome. Canc.* **27**, 59–68 (2000).

122. Pandita,T.K. & Dhar,S. Influence of ATM function on interactions between telomeres and nuclear matrix. *Radiat. Res.* **154**, 133–139 (2000).

123. Zhou,C., Yang,Y. & Jong,A.Y. Mini–prep in ten minutes. *Biotechniques* **8**, 172–173 (1990).

124. Gilad,S. *et al.* Predominance of null mutations in ataxia–telangiectasia. *Hum. Mol. Genet.* **5**, 433–439 (1996).

125. Meyn,M.S. High spontaneous intrachromosomal recombination rates in ataxia–telangiectasia. *Science* **260**, 1327–1330 (1993).

126. Andrews,P.W. Retinoic acid induces neuronal differentiation of a cloned human embryonal carcinoma cell line in vitro. *Dev. Biol.* **103**, 285–293 (1984).

127. Skowronski,J., Fanning,T.G. & Singer,M.F. Unit–length line–1 transcripts in human teratocarcinoma cells. *Mol. Cell Biol.* **8**, 1385–1397 (1988).

128. Hata,K. & Sakaki,Y. Identification of critical CpG sites for repression of L1 transcription by DNA methylation. *Gene* **189**, 227–234 (1997).

129. Galloway,A.M., Spencer,C.A., Anderson,C.W. & Allalunis–Turner,M.J. Differential stability of the DNA–activated protein kinase catalytic subunit mRNA in human glioma cells. *Oncogene* **18**, 1361–1368 (1999).

130. Boggs,S.S., Gregg,R.G., Borenstein,N. & Smithies,O. Efficient transformation and frequent single–site, single–copy insertion of DNA can be obtained in mouse erythroleukemia cells transformed by electroporation. *Exp. Hematol.* **14**, 988–994 (1986).

131. Rodriguez,A. & Flemington,E.K. Transfection–mediated cell–cycle signaling: considerations for transient transfection–based cell–cycle studies. *Anal. Biochem.* **272**, 171–181 (1999).

132. Reardon,J.E. Herpes simplex virus type 1 and human DNA polymerase interactions with 2'–deoxyguanosine 5'–triphosphate analogues. Kinetics of incorporation into DNA and induction of inhibition. *J. Biol. Chem.* **264**, 19039–19044 (1989).

133. Hirt,B. Selective extraction of polyoma DNA from infected mouse cell cultures. *J. Mol. Biol.* **26**, 365–369 (1967).

134. Grant,S.G.N., Jessee,J., Bloom,F.R. & Hanahan,D. Differential Plasmid Rescue From Transgenic Mouse DNAs Into *Escherichia coli* Methylation–Restriction Mutants. *Proc. Natl. Acad. Sci. USA* **87**, 4645–4649 (1990).

135. Little,K.C.E. & Chartrand,P. Genomic DNA is captured and amplified during double–strand break (DSB) repair in human cells. *Oncogene* **23**, 4166–4172 (2004).

136. Smith,J., Baldeyron,C., De Oliveira,I., Sala–Trepat,M. & Papadopoulo,D. The influence of DNA double–strand break structure on end–joining in human cells. *Nucl. Acids. Res.* **29**, 4783–4792 (2001).

137. Roscheisen,C., Haupter,S., Zechner,U. & Speit,G. Characterization of spontaneous and induced mutations in SV40–transformed normal and ataxia telangiectasia cell lines. *Somat. Cell Mol. Genet.* **20**, 493–504 (1994).

138. Cox,R., Debenham,P.G., Masson,W.K. & Webb,M.B. Ataxia–telangiectasia: a human mutation giving high–frequency misrepair of DNA double–stranded scissions. *Mol. Biol. Med.* **3**, 229–244 (1986).

139. Paull,T.T. & Gellert,M. A mechanistic basis for Mre11–directed DNA joining at microhomologies. *Proc. Natl. Acad. Sci. USA* **97**, 6409–6414 (2000).

140. Paulson,J.R. & Laemmli,U.K. The structure of histone–depleted metaphase chromosomes. *Cell* **12**, 817–828 (1977).

141. Shera,K.A., Shera,C.A. & McDougall,J.K. Small tumor virus genomes are integrated near nuclear matrix attachment regions in transformed cells. *J. Virol.* **75**, 12339–12346 (2001).

142. Singh,G.B., Kramer,J.A. & Krawetz,S.A. Mathematical model to predict regions of chromatin attachment to the nuclear matrix. *Nucl. Acids. Res.* **25**, 1419–1425 (1997).

143. Gumus,G., Sunguroglu,A., Tukun,A., Sayin,D.B. & Bokesoy,I. Common fragile sites associated with the breakpoints of chromosomal aberrations in hematologic neoplasms. *Cancer Genet. Cytogenet.* **133**, 168–171 (2002).

144. Simonic,I. & Gericke,G.S. The enigma of common fragile sites. *Hum. Genet.* **97**, 524–531 (1996).

145. Popescu,N.C. Genetic alterations in cancer as a result of breakage at fragile sites. *Cancer Lett.* **192**, 1–17 (2003).

146. Corbin,S. *et al.* Identification of Unstable Sequences within the Common Fragile Site at 3p14.2: Implications for the Mechanism of Deletions within Fragile Histidine

Triad Gene/Common Fragile Site at 3p14.2 in Tumors. *Cancer Res.* **62**, 3477–3484 (2002).

147. Becker,N.A., Thorland,E.C., Denison,S.R., Phillips,L.A. & Smith,D.I. Evidence that instability within the FRA3B region extends four megabases. *Oncogene* **21**, 8713–8722 (2002).

148. Abeysinghe,S.S., Stenson,P.D., Krawczak,M. & Cooper,D.N. Gross Rearrangement Breakpoint Database (GRaBD). *Hum. Mutat.* **23**, 219–221 (2004).

149. Mitelman Database of Chromosome Aberrations in Cancer (2004). Mitelman, F, Johansson, B, and Mertens, F. http://cgap.nci.nih.gov/Chromosomes/Mitelman. 2004.

150. Schwab,M. Amplification of oncogenes in human cancer cells. *Bioessays* **20**, 473–479 (1998).

151. Hecht,F., Tajara,E.H., Lockwood,D., Sandberg,A.A. & Hecht,B.K. New common fragile sites. *Cancer Genet. Cytogenet.* **33**, 1–9 (1988).

152. Kussel–Andermann,P. *et al.* Vezatin, a novel transmembrane protein, bridges myosin VIIA to the cadherin–catenins complex. *EMBO J.* **19**, 6020–6029 (2000).

153. Vogelstein,B. *et al.* Genetic alterations during colorectal–tumor development. *N. Engl. J. Med.* **319**, 525–532 (1988).

154. Croce,C.M. Assignment of the integration site for simian virus 40 to chromosome 17 in GM54VA, a human cell line transformed by simian virus 40. *Proc. Natl. Acad. Sci. U. S. A* **74**, 315–318 (1977).

155. Hara,H. & Kaji,H. Random integration of SV40 in SV40–transformed, immortalized human fibroblasts. *Exp. Cell Res.* **168**, 531–538 (1987).

156. Hwang,S.P. & Kucherlapati,R. Localization and organization of integrated simian virus 40 sequences in a human cell line. *Virology* **105**, 196–204 (1980).

157. Kao,C. *et al.* Carcinogen–induced amplification of SV40 DNA inserted at 9q12–21.1 associated with chromosome breakage, deletions, and translocations in human uroepithelial cell transformation in vitro. *Gene. Chromosome. Canc.* **8**, 155–166 (1993).

158. Marlhens,F. *et al.* Karyotype evolution of the human HBL–100 cell line and mapping of the integration site of SV40 DNA. *Ann. Genet.* **31**, 81–86 (1988).

159. Rabin,M., Uhlenbeck,O.C., Steffensen,D.M. & Mangel,W.F. Chromosomal sites of integration of simian virus 40 DNA sequences mapped by in situ hybridization in two transformed hybrid cell lines. *J. Virol.* **49**, 445–451 (1984).

160. Yano,O. *et al.* Cloning and sequencing of viral integration site in human fibroblasts immortalized by simian virus 40. *Cell Struct. Funct.* **16**, 63–71 (1991).

161. Wurtele,H., Little,K.C. & Chartrand,P. Illegitimate DNA integration in mammalian cells. *Gene Ther.* **10**, 1791–1799 (2003).

162. Mishmar,D. *et al.* Molecular characterization of a common fragile site (FRA7H) on human chromosome 7 by the cloning of a simian virus 40 integration site. *Proc. Natl. Acad. Sci. USA* **95**, 8141–8146 (1998).

163. Maulbecker,C., Mohr,I., Gluzman,Y., Bartholomew,J. & Botchan,M. A deletion in the simian virus 40 large T antigen impairs lytic replication in monkey cells in vivo but enhances DNA replication in vitro: new complementation function of T antigen. *J. Virol.* **66**, 2195–2207 (1992).

164. Huang,K.C., Yamasaki,E.F. & Snapka,R.M. Maintenance of episomal SV40 genomes in GM637 human fibroblasts. *Virology* **262**, 457–469 (1999).

165. Murnane,J.P., Fuller,L.F. & Painter,R.B. Establishment and characterization of a permanent pSV ori—transformed ataxia–telangiectasia cell line. *Exp. Cell Res.* **158**, 119–126 (1985).

166. Lednicky,J.A. & Butel,J.S. Tissue culture adaptation of natural isolates of simian virus 40: changes occur in viral regulatory region but not in carboxy–terminal domain of large T–antigen. *J. Gen. Virol.* **78**, 1697–1705 (1997).

167. Lednicky,J.A. *et al.* Natural Isolates of Simian Virus 40 from Immunocompromised Monkeys Display Extensive Genetic Heterogeneity: New Implications for Polyomavirus Disease. *J. Virol.* **72**, 3980–3990 (1998).

168. Lin,Y., Lukacsovich,T. & Waldman,A.S. Multiple Pathways for Repair of DNA Double–Strand Breaks in Mammalian Chromosomes. *Mol. Cell. Biol.* **19**, 8353–8360 (1999).

169. Bergsma,D.J., Olive,D.M., Hartzell,S.W. & Subramanian,K.N. Territorial limits and functional anatomy of the simian virus 40 replication origin. *Proc. Natl. Acad. Sci. U. S. A* **79**, 381–385 (1982).

170. Dujon,B. *et al.* Mobile introns: definition of terms and recommended nomenclature. *Gene* **82**, 115–118 (1989).

171. Jasin,M. Genetic manipulation of genomes with rare–cutting endonucleases. *Trends Genet.* **12**, 224–228 (1996).

172. Richardson,C. & Jasin,M. Frequent chromosomal translocations induced by DNA double–strand breaks. *Nature* **405**, 697–700 (2000).

173. Goring,D.R., Gupta,K. & DuBow,M.S. Analysis of spontaneous mutations in a chromosomally located HSV–1 thymidine kinase (TK) gene in a human cell line. *Somat. Cell Mol. Genet.* **13**, 47–56 (1987).

174. Villemure,J.F., Savard,N. & Belmaaza,A. Promoter suppression in cultured mammalian cells can be blocked by the chicken beta–globin chromatin insulator 5'HS4 and matrix/scaffold attachment regions. *J. Mol. Biol.* **312**, 963–974 (2001).

175. Birnbaum,D. *et al.* Chromosome arm 8p and cancer: a fragile hypothesis. *Lancet Oncol.* **4**, 639–642 (2003).

176. Adelaide,J. *et al.* A recurrent chromosome translocation breakpoint in breast and pancreatic cancer cell lines targets the neuregulin/NRG1 gene. *Gene. Chromosome. Canc.* **37**, 333–345 (2003).

177. Courtay–Cahen,C., Morris,J.S. & Edwards,P.A.W. Chromosome Translocations in Breast Cancer with Breakpoints at 8p12. *Genomics* **66**, 15–25 (2000).

178. Martin,D.I.K. & Whitelaw,E. The vagaries of variegating transgenes. *Bioessays* **18**, 919–923 (1996).

179. Dar,M.E., Winters,T.A. & Jorgensen,T.J. Identification of defective illegitimate recombinational repair of oxidatively–induced DNA double–strand breaks in ataxia–telangiectasia cells. *Mut. Res./DNA Repair* **384**, 169–179 (1997).

180. Li,Y. *et al.* Expression of ATM in ataxia telangiectasia fibroblasts rescues defects in DNA double–strand break repair in nuclear extracts. *Environ. Mol. Mutagen.* **37**, 128–140 (2001).

181. Powell,S., Whitaker,S., Peacock,J. & McMillan,T. Ataxia telangiectasia: an investigation of the repair defect in the cell line AT5BIVA by plasmid reconstitution. *Mutat. Res.* **294**, 9–20 (1993).

182. Runger,T.M., Poot,M. & Kraemer,K.H. Abnormal processing of transfected plasmid DNA in cells from patients with ataxia telangiectasia. *Mutat. Res* **293**, 47–54 (1992).

183. Tachibana,A. & Sasaki,M.S. Characteristics of the end–joining of DNA double–strand breaks by the ataxia–telangiectasia nuclear extract. *Biochem. Biophys. Res Commun.* **297**, 275–281 (2002).

184. Tatsumi–Miyajima,J., Yagi,T. & Takebe,H. Analysis of mutations caused by DNA double–strand breaks produced by a restriction enzyme in shuttle vector plasmids propagated in ataxia telangiectasia cells. *Mutat. Res.* **294**, 317–323 (1993).

185. Baird,D.M. & Kipling,D. The extent and significance of telomere loss with age. *Ann. N.Y. Acad. Sci.* **1019**, 265–268 (2004).

186. Bailey,S.M., Brenneman,M.A. & Goodwin,E.H. Frequent recombination in telomeric DNA may extend the proliferative life of telomerase–negative cells. *Nucl. Acids. Res.* **32**, 3743–3751 (2004).

187. Dunham,M.A., Neumann,A.A., Fasching,C.L. & Reddel,R.R. Telomere maintenance by recombination in human cells. *Nat. Genet.* **26**, 447–450 (2000).

188. Metcalfe,J.A. *et al.* Accelerated telomere shortening in ataxia telangiectasia. *Nat. Genet.* **13**, 350–353 (1996).

189. Xia,S.J., Shammas,M.A. & Shmookler Reis,R.J. Reduced telomere length in ataxia–telangiectasia fibroblasts. *Mut. Res./DNA Repair* **364**, 1–11 (1996).

190. Sprung,C.N., Bryan,T.M., Reddel,R.R. & Murnane,J.P. Normal telomere maintenance in immortal ataxia telangiectasia cell lines. *Mut. Res./Fund. Mol. M.* **379**, 177–184 (1997).

191. Vermeesch,J.R. *et al.* Interstitial telomeric sequences at the junction site of a jumping translocation. *Hum. Genet.* **99**, 735–737 (1997).

192. Schwab,M. Amplified MYCN in human neuroblastoma: paradigm for the translation of molecular genetics to clinical oncology. *Ann. N.Y. Acad. Sci.* **963**, 63–73 (2002).

193. Chartrand,P., Gusew–Chartrand,N. & Bourgaux,P. Integrated polyoma genomes in inducible permissive transformed cells. *J. Virol.* **39**, 185–195 (1981).

194. Baran,N., Neer,A. & Manor,H. "Onion skin" replication of integrated polyoma virus DNA and flanking sequences in polyoma–transformed rat cells: termination within a specific cellular DNA segment. *Proc. Natl. Acad. Sci. U. S. A* **80**, 105–109 (1983).

195. Zannis–Hadjopoulos,M. & Price,G.B. Eukaryotic DNA replication. *J. Cell. Biochem.* Suppl 32–33, 1–14 (1999).

196. Benner,S.E., Wahl,G.M. & Von Hoff,D.D. Double minute chromosomes and homogeneously staining regions in tumors taken directly from patients versus in human tumor cell lines. *Anti–Cancer Drug.* **2**, 11–25 (1991).

197. Singer,M.J., Mesner,L.D., Friedman,C.L., Trask,B.J. & Hamlin,J.L. Amplification of the human dihydrofolate reductase gene via double minutes is initiated by chromosome breaks. *Proc. Natl. Acad. Sci. U. S. A* **97**, 7921–7926 (2000).

198. Zhou,Y. *et al.* Genome–wide Identification of Chromosomal Regions of Increased Tumor Expression by Transcriptome Analysis. *Cancer Res.* **63**, 5781–5784 (2003).

199. Syu,L.J. & Fluck,M.M. Site–specific in Situ amplification of the integrated polyomavirus genome: A case for a context–specific over–replication model of gene amplification. *J. Mol. Biol.* **271**, 76–99 (1997).

200. Corvi,R., Amler,L.C., Savelyeva,L., Gehring,M. & Schwab,M. MYCN is retained in single copy at chromosome 2 band p23–24 during amplification in human neuroblastoma cells. *Proc. Natl. Acad Sci U. S. A* **91**, 5523–5527 (1994).

201. Kohl,N.E. *et al.* Transposition and amplification of oncogene–related sequences in human neuroblastomas. *Cell* **35**, 359–367 (1983).

202. Tlsty,T.D. Normal Diploid Human and Rodent Cells Lack a Detectable Frequency of Gene Amplification. *Proc. Natl. Acad. Sci. USA* **87**, 3132–3136 (1990).

203. Mondello,C. *et al.* Increased Gene Amplification in Immortal Rodent Cells Deficient for the DNA–dependent Protein Kinase Catalytic Subunit. *Cancer Res.* **61**, 4520–4525 (2001).

204. Mondello,C. *et al.* Gene amplification in fibroblasts from ataxia telangiectasia (AT) patients and in X–ray hypersensitive AT–like Chinese hamster mutants. *Carcinogenesis* **22**, 141–145 (2001).

205. Corvi,R. *et al.* Non–syntenic amplification of MDM2 and MYCN in human neuroblastoma. *Oncogene* **10**, 1081–1086 (1995).

206. Coquelle,A., Rozier,L., Dutrillaux,B. & Debatisse,M. Induction of multiple double–strand breaks within an hsr by meganucleaseI–SceI expression or fragile site activation leads to formation of double minutes and other chromosomal rearrangements. *Oncogene* **21**, 7671–7679 (2002).

207. Windle,B., Draper,B.W., Yin,Y.X., O'Gorman,S. & Wahl,G.M. A central role for chromosome breakage in gene amplification, deletion formation, and amplicon integration. *Genes Dev.* **5**, 160–174 (1991).

208. Coquelle,A., Toledo,F., Stern,S., Bieth,A. & Debatisse,M. A new role for hypoxia in tumor progression: induction of fragile site triggering genomic rearrangements and formation of complex DMs and HSRs. *Mol. Cell* **2**, 259–265 (1998).

209. McClintock,B. The stability of broken ends of chromosomes in *Zea mays*. *Genetics* **26**, 234–282 (1941).

210. Lo,A.W. *et al.* DNA amplification by breakage/fusion/bridge cycles initiated by spontaneous telomere loss in a human cancer cell line. *Neoplasia.* **4**, 531–538 (2002).

211. Hellman,A. *et al.* A role for common fragile site induction in amplification of human oncogenes. *Cancer Cell* **1**, 89–97 (2002).

212. Lo,A.W.I. *et al.* Chromosome Instability as a Result of Double–Strand Breaks near Telomeres in Mouse Embryonic Stem Cells. *Mol. Cell. Biol.* **22**, 4836–4850 (2002).

213. Hahn,P., Kapp,L.N. & Painter,R.B. Establishment and characterization of two human cell lines with amplified dihydrofolate reductase genes. *Exp. Cell Res.* **168**, 89–94 (1987).

214. Rassool,F.V. *et al.* Preferential Integration of Marker DNA into the Chromosomal Fragile Site at 3p14: An Approach to Cloning Fragile Sites. *Proc. Natl. Acad. Sci. USA* **88**, 6657–6661 (1991).

215. Casper,A.M., Nghiem,P., Arlt,M.F. & Glover,T.W. ATR regulates fragile site stability. *Cell* **111**, 779–789 (2002).

216. Taylor,A.M.R., Oxford,J.M. & Metcalfe,J.A. Spontaneous cytogenetic abnormalites in lymphocytes from thirteen patients with ataxia telangiectasia. *Int. J. Cancer* **27**, 311–319 (1981).

217. Kazmierczak,B., Stern,C., Bartnitzke,S. & Bullerdiek,J. Non–random jumping translocations as a result of SV40 large T–antigen expression in benign human tumor cells. *Cell Biol. Int.* **19**, 315–322 (1995).

218. Hoffschir,F. *et al.* Jumping translocations originate clonal rearrangements in SV40–transformed human fibroblasts. *Int. J. Canc.* **52**, 130–136 (1992).

219. Padilla–Nash,H.M. *et al.* Jumping translocations are common in solid tumor cell lines and result in recurrent fusions of whole chromosome arms. *Gene. Chromosome. Canc.* **30**, 349–363 (2001).

220. Wang,P. *et al.* Gene rearrangements induced by the DNA double–strand cleaving agent neocarzinostatin: conservative non–homologous reciprocal exchanges in an otherwise stable genome. *Nucl. Acids. Res.* **30**, 2639–2646 (2002).

221. Richardson,C., Moynahan,M.E. & Jasin,M. Double–strand break repair by interchromosomal recombination: suppression of chromosomal translocations. *Gene. Dev.* **12**, 3831–3842 (1998).

222. Donoho,G., Jasin,M. & Berg,P. Analysis of Gene Targeting and Intrachromosomal Homologous Recombination Stimulated by Genomic Double–Strand Breaks in Mouse Embryonic Stem Cells. *Mol. Cell. Biol.* **18**, 4070–4078 (1998).

223. Belmaaza,A. & Chartrand,P. One–sided invasion events in homologous recombination at double–strand breaks. *Mutat. Res.* **314**, 199–208 (1994).

224. Haber,J.E. DNA repair. Gatekeepers of recombination. *Nature* **398**, 665–667 (1999).

225. Merrihew,R.V., Marburger,K., Pennington,S.L., Roth,D.B. & Wilson,J.H. High–frequency illegitimate integration of transfected DNA at preintegrated target sites in a mammalian genome. *Mol. Cell. Biol.* **16**, 10–18 (1996).

226. Guirouilh–Barbat,J. *et al.* Impact of the KU80 Pathway on NHEJ–Induced Genome Rearrangements in Mammalian Cells. *Mol. Cell* **14**, 611–623 (2004).

227. Golzio,M., Teissie,J. & Rols,M.P. Direct visualization at the single–cell level of electrically mediated gene delivery. *Proc. Natl. Acad. Sci. USA* **99**, 1292–1297 (2002).

228. Gee,H. A journey into the genome: what's there. Nature Science Update (Online) . 12–2–2004. Nature News Service / Macmillan Magazines Ltd.

229. Patthy,L. Genome evolution and the evolution of exon–shuffling — a review. *Gene* **238**, 103–114 (1999).

230. Kazazian,H.H., Jr. *et al.* Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. *Nature* **332**, 164–166 (1988).

231. Esnault,C., Maestre,J. & Heidmann,T. Human LINE retrotransposons generate processed pseudogenes. *Nat. Genet.* **24**, 363–367 (2000).

232. Ostertag,E.M., Goodier,J.L., Zhang,Y. & Kazazian,H.H., Jr. SVA elements are nonautonomous retrotransposons that cause disease in humans. *Am. J. Hum. Genet.* **73**, 1444–1451 (2003).

233. Dewannieux,M., Esnault,C. & Heidmann,T. LINE-mediated retrotransposition of marked Alu sequences. *Nat. Genet.* **35**, 41–48 (2003).

234. Moran,J.V., DeBerardinis,R.J. & Kazazian,H.H., Jr. Exon Shuffling by L1 Retrotransposition. *Science* **283**, 1530–1534 (1999).

235. Morrish,T.A. *et al.* DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. *Nat. Genet.* **31**, 159–165 (2002).

236. Smit,A.F. Interspersed repeats and other mementos of transposable elements in mammalian genomes. *Curr. Opin. Genet. Dev.* **9**, 657–663 (1999).

237. Li,W.H., Gu,Z., Wang,H. & Nekrutenko,A. Evolutionary analyses of the human genome. *Nature* **409**, 847–849 (2001).

238. Nekrutenko,A. & Li,W.H. Transposable elements are found in a large number of human protein-coding genes. *Trends Genet.* **17**, 619–621 (2001).

239. Lang,B.F., Gray,M.W. & Burger,G. Mitochondrial genome evolution and the origin of eukaryotes. *Annu. Rev. Genet.* **33**, 351–397 (1999).

240. Forand,A., Dutrillaux,B. & Bernardino-Sgherri,J. {gamma}-H2AX Expression Pattern in Non-Irradiated Neonatal Mouse Germ Cells and after Low-Dose {gamma}-Radiation: Relationships Between Chromatid Breaks and DNA Double-Strand Breaks. *Biol. Reprod.* **71**, 643–649 (2004).

241. Rulicke,T. & Hubscher,U. Germ line transformation of mammals by pronuclear microinjection. *Exp. Physiol.* **85**, 589–601 (2000).

242. Kazazian,H.H., Jr. An estimated frequency of endogenous insertional mutations in humans. *Nat. Genet.* **22**, 130 (1999).

243. Willett-Brozick,J.E., Savul,S.A., Richey,L.E. & Baysal,B.E. Germ line insertion of mtDNA at the breakpoint junction of a reciprocal constitutional translocation. *Hum. Genet.* **109**, 216–223 (2001).

244. Bushman,F.D. Lateral DNA transfer: Mechanisms and consequences. Cold Spring Harbor Laboratory Press, Woodbury, NY, 11797–2924, USA. (2002).

245. Salzberg,S.L., White,O., Peterson,J. & Eisen,J.A. Microbial Genes in the Human Genome: Lateral Transfer or Gene Loss? *Science* **292**, 1903–1906 (2001).

246. Ingram,J. Josef Penninger (Interview: Video Recording). Penninger, J. 20–12–2001. Daily Planet, Discovery Channel.

247. Woodworth–Gutai,M. Recombination in SV40–infected cells: nucleotide sequences at viral–viral recombinant joints in naturally arising variants. *Virology* **109**, 344–352 (1981).

248. Dhruva,B.R., Shenk,T. & Subramanian,K.N. Integration in vivo into simian virus 40 DNA of a sequence that resembles a certain family of genomic interspersed repeated sequences. *Proc. Natl. Acad. Sci. USA* **77**, 4514–4518 (1980).

249. Sheflin,L., Celeste,A. & Woodworth–Gutai,M. Recombination in simian virus 40–infected cells. Structure of naturally arising variants ev–2114, ev–2102, and ev–1110. *J. Biol. Chem.* **258**, 14315–14321 (1983).

250. Stary,A. & Sarasin,A. Molecular analysis of DNA junctions produced by illegitimate recombination in human cells. *Nucl. Acids. Res.* **20**, 4269–4274 (1992).

251. Doolittle,R.F. & Feng,D.F. Tracing the origin of retroviruses. *Curr. Top. Microbiol. Immunol.* **176**, 195–211 (1992).

252. Malik,H.S., Henikoff,S. & Eickbush,T.H. Poised for Contagion: Evolutionary Origins of the Infectious Abilities of Invertebrate Retroviruses. *Genome Res.* **10**, 1307–1318 (2000).

253. Sweet,B.H. & Hilleman,M.R. The vacuolating virus, SV40. *Proc. Soc. Exp. Biol. Med.* **105**, 420–427 (1960).

254. Butel,J.S. & Lednicky,J.A. Cell and molecular biology of simian virus 40: implications for human infections and disease. *J. Natl. Cancer Inst.* **91**, 119–134 (1999).

255. Ferber,D. Virology. Monkey virus link to cancer grows stronger. *Science* **296**, 1012–1015 (2002).

256. Li,R.M. *et al.* Molecular identification of SV40 infection in human subjects and possible association with kidney disease. *J Am. Soc. Nephrol.* **13**, 2320–2330 (2002).

257. Mendoza,S.M., Konishi,T. & Miller,C.W. Integration of SV40 in human osteosarcoma DNA. *Oncogene* **17**, 2457–2462 (1998).

258. Carbone,M. *et al.* SV40 and human brain tumors. *Int. J. Cancer* **106**, 140–142 (2003).

259. Jasani,B. *et al.* Association of SV40 with human tumours. *Semin. Cancer Biol.* **11**, 49–61 (2001).

260. Shi,L.M., Ho,J., Norling,L.A., Roy,M. & Xu,Y. A real time quantitative PCR–based method for the detection and quantification of simian virus 40. *Biologicals* **27**, 241–252 (1999).

261. Cerrano,P.G. *et al.* Simian virus 40 and malignant mesothelioma (Review). *Int. J. Oncol.* **22**, 187–194 (2003).

262. Lednicky,J.A., Garcea,R.L., Bergsagel,D.J. & Butel,J.S. Natural Simian Virus 40 Strains Are Present in Human Choroid Plexus and Ependymoma Tumors. *Virology* **212**, 710–717 (1995).

263. Bergsagel,D.J., Finegold,M.J., Butel,J.S., Kupdky,W.J. & Garcea,R.L. DNA sequences similar to those of simian virus 40 in ependymomas and choroid plexus tumors of childhood. *N. Engl. J. Med.* **326**, 988–993 (1992).

264. Vilchez,R.A. *et al.* Association between simian virus 40 and non–Hodgkin lymphoma. *Lancet* 817–823 (2002).

265. Shivapurkar,N. *et al.* Presence of simian virus 40 DNA sequences in human lymphomas. *Lancet* **359**, 851–852 (2002).

266. Nakatsuka,S.i. *et al.* Simian Virus 40 Sequences in Malignant Lymphomas in Japan. *Cancer Res.* **63**, 7606–7608 (2003).

267. Butel,J.S., Vilchez,R.A., Jorgensen,J.L. & Kozinetz,C.A. Association between SV40 and non–Hodgkin's lymphoma. *Leuk. Lymphoma* **44** Suppl 3, S33–S39 (2003).

268. Gazdar,A.F., Butel,J.S. & Carbone,M. SV40 and human tumours: myth, association or causality? *Nat. Rev. Cancer* **2**, 957–964 (2002).

269. Capello,D., Rossi,D., Gaudino,G., Carbone,A. & Gaidano,G. Simian virus 40 infection in lymphoproliferative disorders. *Lancet* **361**, 88–89 (2003).

270. Saenz–Robles,M.T., Sullivan,C.S. & Pipas,J.M. Transforming functions of Simian Virus 40. *Oncogene* **20**, 7899–7907 (2001).

271. Foddis,R. *et al.* SV40 infection induces telomerase activity in human mesothelial cells. *Oncogene* **21**, 1434–1442 (2002).

272. Zimmerman,R.J. & Cerutti,P.A. A comparison of markers of human fibroblast transformation induced by chemical carcinogen treatment or by transfection of an origin–defective SV40–containing plasmid. *Mutat. Res.* **199**, 449–459 (1988).

273. David,H., Mendoza,S., Konishi,T. & Miller,C.W. Simian virus 40 is present in human lymphomas and normal blood. *Cancer Lett.* **162**, 57–64 (2001).

274. Lednicky,J.A. & Butel,J.S. Simian virus 40 regulatory region structural diversity and the association of viral archetypal regulatory regions with human brain tumors. *Semin. Cancer Biol.* **11**, 39–47 (2001).

275. Staib,C., Wegner,M. & Grummt,F. Activation of SV40 DNA replication in vivo by amplification–promoting sequences of the mouse ribosomal gene cluster. *Chromosoma* **107**, 33–38 (1998).

276. Nishitani,H. & Lygerou,Z. Control of DNA replication licensing in a cell cycle. *Genes Cells* **7**, 523–534 (2002).

277. Bandura,J.L. & Calvi,B.R. Duplication of the genome in normal and cancer cell cycles. *Cancer Biol. Ther.* **1**, 8–13 (2002).

278. Baiker,A. *et al.* Mitotic stability of an episomal vector containing a human scaffold/matrix–attached region is provided by association with nuclear matrix. *Nat. Cell Biol.* **2**, 182–184 (2000).

279. Iguchi–Ariga,S.M., Itani,T., Yamaguchi,M. & Ariga,H. c–myc protein can be substituted for SV40 T antigen in SV40 DNA replication. *Nucl. Acids. Res.* **15**, 4889–4899 (1987).

280. Strayer,D.S. Effective gene transfer using viral vectors based on SV40. *Methods Mol. Biol.* **133**, 61–74 (2000).

281. Li,L. *et al.* Role of the non–homologous DNA end joining pathway in the early steps of retroviral infection. *EMBO J.* **20**, 3272–3281 (2001).

282. Novac,O., Matheos,D., Araujo,F.D., Price,G.B. & Zannis–Hadjopoulos,M. In vivo association of Ku with mammalian origins of DNA replication. *Mol. Biol. Cell* **12**, 3386–3401 (2001).

283. Gruss,P. & Sauer,G. Infectious linear DNA sequences replicating in simian virus 40–infected cells. *J. Virol.* **21**, 565–578 (1977).

284. Gurney,T., Jr. & Gurney,E.G. Spontaneous rearrangement of integrated simian virus 40 DNA in nine transformed rodent cell lines. *J. Virol.* **63**, 165–174 (1989).

285. Lavi,S. Carcinogen–mediated amplification of viral DNA sequences in simian virus 40–transformed Chinese hamster embryo cells. *Proc. Natl. Acad. Sci. USA* **78**, 6144–6148 (1981).

286. Vleminckx,C. & Moens,W. Induction of microchromosomes by chemical carcinogens correlates with SV40–DNA amplification in SV40–transformed Chinese hamster cells. *Carcinogenesis* **7**, 229–234 (1986).

287. Gerspach,R. & Matz,B. Herpes simplex virus–directed overreplication of chromosomal DNA physically linked to the simian virus 40 integration site of a transformed hamster cell line. *Virology* **165**, 282–285 (1988).

288. Combaret,V. *et al.* Circulating MYCN DNA as a tumor–specific marker in neuroblastoma patients. *Cancer Res.* **62**, 3646–3648 (2002).

289. Solovei,I. *et al.* Topology of double minutes (dmins) and homogeneously staining regions (HSRs) in nuclei of human neuroblastoma cell lines. *Gene. Chromosome. Canc.* **29**, 297–308 (2000).

290. Leu,T.H. & Hamlin,J.L. Activation of a mammalian origin of replication by chromosomal rearrangement. *Mol. Cell Biol.* **12**, 2804–2812 (1992).

291. Trask,B.J. & Hamlin,J.L. Early dihydrofolate reductase gene amplification events in CHO cells usually occur on the same chromosome arm as the original locus. *Genes Dev.* **3**, 1913–1925 (1989).

292. Sait,S.N. *et al.* Double minute chromosomes in acute myeloid leukemia and myelodysplastic syndrome: identification of new amplification regions by fluorescence in situ hybridization and spectral karyotyping. *Gene. Chromosome. Canc.* **34**, 42–47 (2002).

293. Sanders,S., Keck–Waggoner,C.L., Zimonjic,D.B., Popescu,N.C. & Thorgeirsson,S.S. Assignment of WDR7 (alias TRAG, TGF–beta resistance associated gene) to orthologous regions of human chromosome 18q21.1—>q22 and mouse chromosome 18D.1–E.3 by fluorescence in situ hybridization. *Cytogenet. Cell Genet.* **88**, 324–325 (2000).

294. Lee,W.C. & Testa,J.R. Somatic genetic alterations in human malignant mesothelioma (review). *Int. J. Oncol.* **14**, 181–188 (1999).

295. Feo,S., Di Liegro,C., Mangano,R., Read,M. & Fried,M. The amplicons in HL60 cells contain novel cellular sequences linked to MYC locus DNA. *Oncogene* **13**, 1521–1529 (1996).

296. O'Neill,S. *et al.* MYCN amplification and 17q in neuroblastoma: evidence for structural association. *Gene. Chromosome. Canc.* **30**, 87–90 (2001).

297. Shimizu,N., Miura,Y., Sakamoto,Y. & Tsutsui,K. Plasmids with a mammalian replication origin and a matrix attachment region initiate the event similar to gene amplification. *Cancer Res.* **61**, 6987–6990 (2001).

298. Vilchez,R.A., Kozinetz,C.A., Arrington,A.S., Madden,C.R. & Butel,J.S. Simian virus 40 in human cancers. *Am. J. Med.* **114**, 675–684 (2003).

299. Takebayashi,S. *et al.* Identification of new minimally lost regions on 18q in head and neck squamous cell carcinoma. *Cancer Res.* **60**, 3397–3403 (2000).

300. Akagi,T. *et al.* A novel gene, MALT1 at 18q21, is involved in t(11;18) (q21;q21) found in low–grade B–cell lymphoma of mucosa–associated lymphoid tissue. *Oncogene* **18**, 5785–5794 (1999).

301. Aisenberg,A.C., Wilkes,B.M. & Jacobson,J.O. The bcl–2 gene is rearranged in many diffuse B–cell lymphomas. *Blood* **71**, 969–972 (1988).

302. Akagi,T., Kondo,E. & Yoshino,T. Expression of Bcl–2 protein and Bcl–2 mRNA in normal and neoplastic lymphoid tissues. *Leuk. Lymphoma* **13**, 81–87 (1994).

303. Ngan,B.Y., Chen–Levy,Z., Weiss,L.M., Warnke,R.A. & Cleary,M.L. Expression in non–Hodgkin's lymphoma of the bcl–2 protein associated with the t(14;18) chromosomal translocation. *N. Engl. J Med.* **318**, 1638–1644 (1988).

304. Papakonstantinou,G., Verbeke,C., Hastka,J., Bohrer,M. & Hehlmann,R. bcl–2 expression in non–Hodgkin's lymphomas is not associated with bcl– 2 gene rearrangements. *Br. J. Haematol.* **113**, 383–390 (2001).

305. Pezzella,F. *et al.* Expression of the bcl–2 oncogene protein is not specific for the 14;18 chromosomal translocation. *Am. J. Pathol.* **137**, 225–232 (1990).

306. Liu,Y., Hernandez,A.M., Shibata,D. & Cortopassi,G.A. BCL2 translocation frequency rises with age in humans. *Proc. Natl. Acad. Sci. USA* **91**, 8910–8914 (1994).

307. Ilyinskii,P.O., Daniel,M.D., Horvath,C.J. & Desrosiers,R.C. Genetic analysis of simian virus 40 from brains and kidneys of macaque monkeys. *J. Virol.* **66**, 6353–6360 (1992).

308. Horvath,C.J. *et al.* Simian virus 40–induced disease in rhesus monkeys with simian acquired immunodeficiency syndrome. *Am. J. Pathol.* **140**, 1431–1440 (1992).

309. Martini,F. *et al.* Simian–virus–40 footprints in human lymphoproliferative disorders of HIV– and HIV+ patients. *Int. J. Cancer* **78**, 669–674 (1998).

310. Vilchez,R.A. *et al.* Detection of polyomavirus simian virus 40 tumor antigen DNA in AIDS–related systemic non–Hodgkin lymphoma. *J. Acquir. Immune. Defic. Syndr.* **29**, 109–116 (2002).

311. Vilchez,R.A. & Butel,J.S. Emergent Human Pathogen Simian Virus 40 and Its Role in Cancer. *Clin. Microbiol. Rev.* **17**, 495–508 (2004).

312. Malkin,D. Simian virus 40 and non–Hodgkin lymphoma. *Lancet* **359**, 812–813 (2002).

313. Krieser,R.J. & White,K. Engulfment mechanism of apoptotic cells. *Curr. Opin. Cell Biol.* **14**, 734–738 (2002).

314. Gagnon,E. *et al.* Endoplasmic reticulum–mediated phagocytosis is a mechanism of entry into macrophages. *Cell* **110**, 119–131 (2002).

315. Subramanian,K. & Meyer,T. Calcium–induced restructuring of nuclear envelope and endoplasmic reticulum calcium stores. *Cell* **89**, 963–971 (1997).

316. Bergsmedh,A. *et al.* Horizontal transfer of oncogenes by uptake of apoptotic bodies. *Proc. Natl. Acad. Sci. USA* **98**, 6407–6411 (2001).

317. Spetz,A.L., Patterson,B.K., Lore,K., Andersson,J. & Holmgren,L. Functional Gene Transfer of HIV DNA by an HIV Receptor–Independent Mechanism. *J. Immunol.* **163**, 736–742 (1999).

318. Holmgren,L. *et al.* Horizontal Transfer of DNA by the Uptake of Apoptotic Bodies. *Blood* **93**, 3956–3963 (1999).

319. Invernizzi,P. *et al.* Presence of fetal DNA in maternal plasma decades after pregnancy. *Hum. Genet.* **110**, 587–591 (2002).

320. Garcia–Olmo,D.A. & Garcia–Olmo,D.C. Functionality of Circulating DNA: The Hypothesis of Genometastasis. *Ann. N.Y. Acad. Sci.* **945**, 265–275 (2001).

321. Lisby,M., Mortensen,U.H. & Rothstein,R. Colocalization of multiple DNA double–strand breaks at a single Rad52 repair centre. *Nat. Cell Biol.* **5**, 572–577 (2003).

322. Singleton,B.K., Griffin,C.S. & Thacker,J. Clustered DNA Damage Leads to Complex Genetic Changes in Irradiated Human Cells. *Cancer Res.* **62**, 6263–6269 (2002).

323. Elliott,B. & Jasin,M. Double–strand breaks and translocations in cancer. *Cell Mol. Life Sci.* **59**, 373–385 (2002).