MODELS OF CANCER

Harnessing transposons for cancer gene discovery

Neal G. Copeland and Nancy A. Jenkins

Abstract | Recently, it has become possible to mobilize the Tc1/mariner transposon, Sleeping Beauty (*SB*), in mouse somatic cells at frequencies high enough to induce cancer. Tumours result from *SB* insertional mutagenesis of cancer genes, thus facilitating the identification of the genes and signalling pathways that drive tumour formation. A conditional *SB* transposition system has also been developed that makes it possible to limit where *SB* mutagenesis occurs, providing a means to selectively model many types of human cancer. *SB* mutagenesis has already identified a large collection of known cancer genes in addition to a plethora of new candidate cancer genes and potential drug targets.

The development of agents, such as monoclonal antibodies, small molecules, peptide memetics and antisense oligonucleotides, which enable the specific modulation of aberrant signalling pathways and mutated genes that cause cancer, has sometimes had great success in the clinic without causing the deleterious side effects that are often associated with conventional cancer chemotherapeutics[1]. This has spurred a worldwide effort to identify all of the genes and signalling pathways that cause cancer with the hope that this will lead to the identification of new drug targets that are amenable to molecular therapies. Although still in their infancy, these efforts have already provided surprising results[2]. These efforts suggest that the cancer genome is composed of hundreds of infrequently mutated genes in addition to a few commonly mutated genes. They also show that there is much heterogeneity in the genes that cause cancer, even in two cancers of the same type. Some researchers have argued, however, that many of these infrequently mutated genes are not true cancer-causing genes but are, instead, statistical artefacts[3–5]. It will therefore be important to validate a role for each of these genes in cancer, a difficult task given the large number of infrequently mutated genes. One way of doing this is to sequence the entire genome of hundreds of tumours of each type of human cancer. This is the approach currently used by the International Cancer Genome Consortium (see Further information), which aims to study more than 25,000 cancer genomes, representing 50 different cancer types and/or subtypes at the genomic, epigenomic and transcriptomic levels. Another method is through comparative genomics; for example, by identifying whether these genes are also mutated in cancers that arise in mice.

Genomics and Genetics Division, Institute of Molecular and Cell Biology, Agency for Science, Technology and Research, 61 Biopolis Drive, Proteos, Singapore 138673. Correspondence to N.A.J.
e-mail:
njenkins@imcb.a-star.edu.sg
doi:10.1038/nrc2916
Published online
16 September 2010

Insertional mutagenesis

A valuable, high-throughput method for cancer gene discovery in the mouse has been retroviral insertional mutagenesis (RIM)[6]. Retroviruses can induce cancer as a result of their normal life cycle: the proviral DNA integrates into the mouse genome and can deregulate the expression of an oncogene or inactivate the expression of a tumour suppressor gene. The retroviral integration sites in tumours thus mark the location of candidate cancer (CAN) genes. High-throughput methods have also been developed for PCR-amplifying and sequencing of retroviral integration sites from tumours. This technology, and the availability of the mouse genome sequence, has made it possible to identify hundreds of CAN genes using RIM[7].

There are, however, several limitations to the use of RIM. First, retroviruses preferentially target the 5′ end of expressed genes[8,9], which limits the number of CAN genes accessible to RIM. Second, retroviruses also carry strong enhancers in their long terminal repeats (LTRs) and can deregulate the expression of oncogenes located hundreds of kilobases (kb) away[10], which can greatly complicate the identification of CAN genes using RIM. Last, retroviruses induce haematopoietic and mammary tumours but little else. RIM is therefore not of great general use for modelling most solid cancers, which are the most common cancers that affect humans.

Transposon-based insertional mutagenesis (TIM) provides an alternative high-throughput platform for cancer gene discovery. Transposable elements are discrete DNA elements that have the unique ability to change their genomic position. Excluding transposable elements like retrotransposons, the vast majority of

At a glance

- Transposon-based insertional mutagenesis (TIM) can be used to model many types of cancer in mice and facilitates the rapid identification of genes that cause cancer.
- Comparison of the genes mutated by TIM with the genes mutated in human cancer can help to discriminate between those that are driver and those that are passenger mutations in human cancer.
- TIM can identify new human cancer genes that have so far been missed in human cancer genome sequencing studies, that are difficult to identify because they are epigenetically regulated and that are located in large amplified or deleted regions.
- TIM induces tumours at all stages of tumour progression, sometimes in a single animal. By identifying the genes mutated by TIM at each stage of progression it might be possible to determine the order in which the mutations were acquired and whether they are involved in tumour initiation, progression and metastasis.
- Mouse cancer models induced by TIM provide a plethora of new models for studying the biology of cancer and for testing new cancer therapeutics before they go into the clinic.

transposable elements use a 'cut and paste' mechanism for transposition, in which the transposable element-encoded transposase catalyses the excision of the transposon from its original location in the genome and then promotes its reintegration elsewhere. There are two types of transposons: autonomous transposons, which encode an active transposase and are thus capable of transposing on their own, and nonautonomous transposons, which lack a functional transposase but retain the *cis*-acting DNA sequences that are necessary for transposition. Nonautonomous transposons are therefore active only when transposase is supplied in *trans*, which in principle allows one to control the tissue in which TIM occurs by limiting where the transposase is expressed. Therefore, TIM can be used to selectively model many types of cancer, including solid cancers.

The awakening of Sleeping Beauty (*SB*)

Tc1/mariner transposons are the most widespread transposons found in nature[11], suggesting that they might be less dependent on the host factors required for transposition than other transposable elements and could therefore potentially provide an alternative insertional mutagen for cancer gene discovery in the mouse. Although a few active Tc1/mariner elements have been isolated from invertebrates[11], no active Tc1/mariner elements have been isolated from vertebrate cells. This is probably because transposons can be detrimental to the host and are not subject to positive selection. Mutations in transposons thus accumulate in a time-proportional manner[12], limiting their use for genetic studies in vertebrates.

The best-characterized Tc1/mariner elements are those found in teleost fish. These transposable elements are classified into three major types: zebrafish, salmonoid and *Xenopus laevis TXr*[13], of which the salmonoid subfamily is probably the youngest and the most recently active[14]. These elements are small, ranging from 1.3 kb to 2.4 kb in length, and encode a single transposase gene that is flanked by inverted terminal repeats (FIG. 1a). Each inverted terminal repeat contains two imperfect direct repeats, which are the binding sites for the transposase[15]. Tc1/mariner elements have little overt target site preference — with the exception that they always integrate into a TA dinucleotide (FIG. 1b).

In a landmark paper, Ivics and colleagues[16] compared the sequences of 12 partial salmonoid-type transposable element sequences from 8 different fish species, and used the accumulated phylogenetic data to predict the consensus sequence of an active transposase. Genetic engineering was then used to construct a synthetic active transposase. This transposase, designated SB10, could bind to inverted terminal repeats of salmonoid transposons when supplied in *trans* in a substrate-specific manner, and could mediate its precise cut and paste transposition in cells from all the major vertebrate classes[17]. This transposition system, which was 'awakened' from 10 million years of evolutionary sleep[14], was aptly named Sleeping Beauty (*SB*)[16].

SB can transpose in mouse somatic[18], embryonic stem (ES)[19] and germ cells[20–23], although the transposition frequency varies greatly among these cell types. In ES cells, the frequency of *SB* excision, a measure of *SB* transposition, is 3.5×10^{-5} excisions per cell per generation, and the frequency of *SB* transposition in mouse spermatids is orders of magnitude higher (0.2 to 2.0 transposition events per spermatid). One explanation for these differences is the differential expression of the host factors required for transposition, such as the DNA-bending, high-mobility group B1 (HMGB1) protein. In mouse cells deficient in HMGB1, *SB* transposition has been shown to be severely reduced compared with that in mouse cells that overexpress HMGB1, in which *SB* transposition is increased[24]. HMGB1 stimulates the preferential binding of the transposase to the direct repeat furthest from the cleavage site at the end of the transposon and promotes bending of DNA fragments containing the transposon inverted terminal repeat[24] (FIG. 1c). This is thought to help ensure that transposase–transposon complexes are first formed at the internal direct repeats, which subsequently promote the juxtaposition of functional sites in transposon DNA, thereby assisting the formation of synaptic complexes that are required for transposon excision. Another host factor that has been shown to affect *SB* transposition is the MYC-interacting protein zinc finger 1 (MIZ1). MIZ1 downregulates cyclin D1 expression through its interaction with the transposase, resulting in a prolonged G1 phase of the cell cycle and retarded growth of transposase-expressing cells[25]. This temporary G1 arrest enhances *SB* transposition at a phase in the cell cycle in which the non-homologous end joining (NHEJ) pathway of DNA repair is preferentially active. This is the major cellular pathway that repairs the free 3′ overhangs that remain after *SB* excision[26,27] (FIG. 1b). This transposase-induced G1 slowdown is thought to be a transposon function that has evolved to maximize the chance for a successful transposition event[25].

Another factor that affects the frequency of *SB* transposition is methylation of the transposon. Increased DNA methylation has been shown to cause the transposon to adopt a heterochromatic confirmation, which has a greatly increased affinity for transposase[28,29].

This has been shown to lead to an 11-fold to 100-fold increase in the frequency of *SB* transposition compared with the transcriptionally active euchromatic transposon DNA found in ES cells. Contrary to this, *SB* transposons often undergo post-integrative silencing owing to increased DNA methylation and histone deacetylation once they have been incorporated into the germ line, which is not observed in ES cells[30–32]. This is thought to be part of a host survival pathway that targets invading transposons for transcriptional silencing[33,34]. These two seemingly contradictory results are difficult to reconcile. One possibility is that the increased affinity of the transposase for heterochromatic DNA is an attempt to overcome the silencing effects of the host survival pathway.

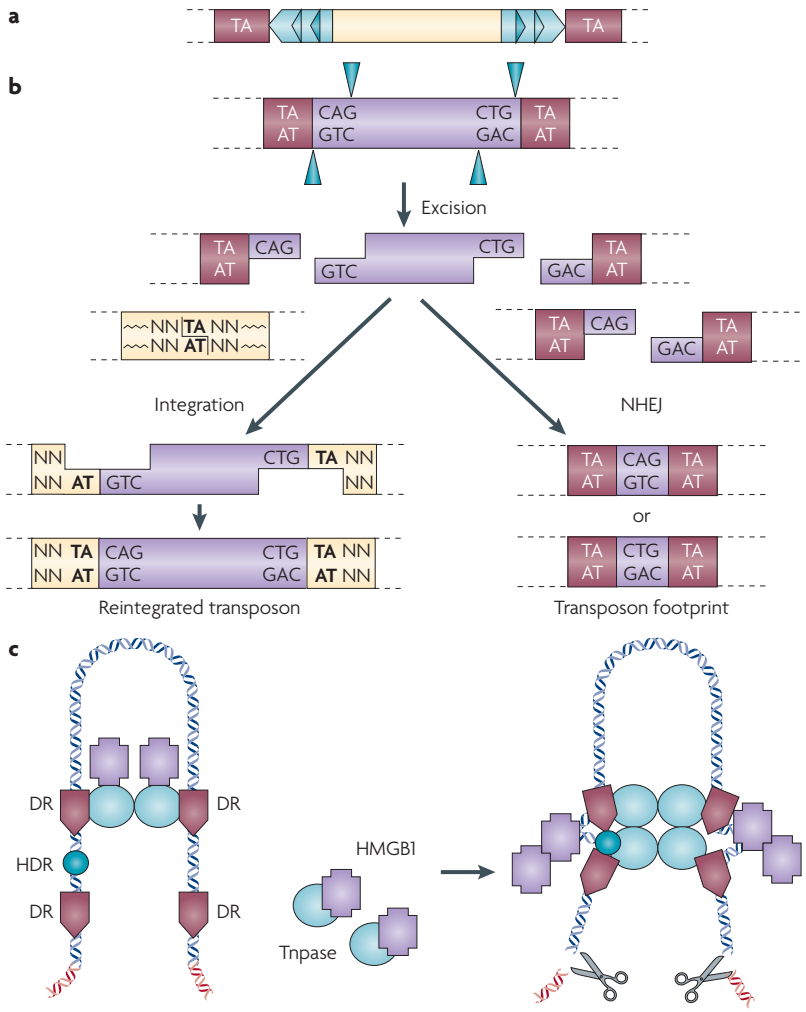

Figure 1 | ***SB* structure and mechanism of transposition. a** | The Sleeping Beauty (*SB*) inverted terminal repeats (light blue arrows) each contain two direct repeats (DRs) (light and dark blue triangles) that are the binding sites for the transposase. The genomic TA dinucleotide that is duplicated during integration and flanks the transposon is shown in red. **b** | Cut and paste mechanism of *SB* transposition. The transposase (blue triangles) initiates transposition by cleaving both ends of the transposon to generate 3 base pair (bp) 5′ overhangs (purple) and also cleaves a genomic TA dinucleotide at the integration site (yellow) to create a gap with a 3′ TA overhang at both ends. The host non-homologous end joining (NHEJ) DNA repair machinery then repairs the single-stranded gaps at the integration site and the double-strand breaks in the donor DNA. A small 5 bp footprint (one TA and 3 bps from the end of the transposon) remains at the excision site. **c** | Proposed model for the role of DNA-bending, high-mobility group B1 (HMGB1) in *SB* synaptic complex formation. *SB* transposase (blue circle) recruits HMGB1 (purple cross) to the transposon DRs furthest from the transposon ends. HMGB1 stimulates binding of the transposase to the inner DRs. Once in contact with DNA, HMGB1 bends the spacer region between the DRs, thereby helping to correctly position the outer DRs for binding by the transposase. Cleavage occurs only if complex formation is complete. The complex includes four transposase binding sites (red, pointed squares), the HMGB1 direct repeat (HDR) enhancer sequence (dark blue circle) and four transposase subunits. Figure is reproduced, with permission, from REF. 24 © (2003) Oxford University Press.

SB insertion site preference

Sequence analysis of *SB* integration sites from somatic and germ cells showed that they are AT-rich; with the sequence ANNTANNT favoured over other TA dinucleotides[35–37]. Moreover, target site sequences tend to have a bendable structure and a palindromic pattern of potential hydrogen-bonding sites in the major groove of DNA. These features are conserved in the Tc1/mariner family of transposons and could be preferred either because they favour the formation and stability of the integration complex or because they promote *SB*-mediated cleavage through the distortion and twisting of the target DNA. These results confirm earlier suggestions that *SB* transposons do not integrate randomly on a micro-scale, even though, at the genome level, integration, with respect to genes and intergenic regions, is fairly random[36,38,39].

Although random genomic insertion of *SB* is observed when transposition takes place from transfected plasmids, *SB* demonstrates a strong propensity for local hopping when transposition takes place from a chromosomal site[19,21,22]. Local hopping refers to a process in which the transposon, once excised from the genome, preferentially reintegrates near its original location. In germline mutagenesis screens, 50–80% of *SB* transposons reinsert within ~6 megabases (Mb) either side of the donor locus[36,38,39]. This local hopping interval is much larger than that for other transposons, such as P elements, for which the local hopping interval is ~100 kb[40]. Outside the local hopping interval, *SB* insertions are distributed across the genome and no chromosome is favoured.

SB transposition induces cancer in mice

Because of the low transposition frequency of *SB* in somatic cells and the requirement for multiple cooperating mutations to produce cancer[41], it was clear early on that several modifications would need to be made to the *SB* transposition system before it could be used to induce cancer in mice. The first modification was the development of a new mutagenic transposon, T2/Onc2, which could deregulate the expression of a proto-oncogene or inactivate the expression of a tumour suppressor gene[42] (FIG. 2). In addition, T2/Onc2 was also reduced in size to ~2.0 kb, its optimal size for transposition[43]. T2/Onc2 has optimized transposase binding sites that increase *SB* transposition compared with other *SB* transposons that have been used previously[44]. T2/Onc2 was introduced into the mouse germ line by microinjection, and three transgenic lines carrying high copy numbers of the transposon linked together at a single site in the genome were selected for further experiments (FIG. 3). Transgenic lines with high copy numbers of T2/Onc2 were selected because *SB* is a cut and paste transposon and not a copy and paste transposon,

and there is only a 40% to 50% probability that an excised transposon will reintegrate into the genome[30,45]. Therefore, once transposition is initiated the number of transposons integrated in the genome will decrease over time. In order to insertionally mutate enough cooperating cancer genes to induce cancer before all transposons disappear from the genome, a large starting number of transposons would theoretically be needed.

To mobilize T2/Onc2 in all somatic cells, a genetically enhanced transposase, SB11 (REF. 43), was knocked in to the mouse *Rosa26* locus (*Rosa26–SB11*)[42] (FIG. 4). Genes targeted to *Rosa26* are widely expressed during development and in adult tissues[46,47], but are not subject to the epigenetic silencing that often occurs with conventional transgenes. Hopefully, therefore, this would drive SB11 expression in all somatic cells beginning early in development and in adult tissues.

When *Rosa26–SB11* mice were crossed to high copy T2/Onc2 transgenic mice, half to three-quarters of the double transgenic embryos died from developmental defects at embryonic day 10–16 (REF. 42). One possible cause for these developmental defects is double-strand DNA breaks resulting from *SB* excision that are left unrepaired because the NHEJ repair pathway is overwhelmed by the massive mobilization of so many *SB* transposons. Genomic rearrangements located near the site of the transposon concatamer have also been reported following mobilization of *SB* transposons from concatamers[48] and this could also account for some of these developmental defects, although the rearrangements reported to date have all been shown to induce recessive phenotypes.

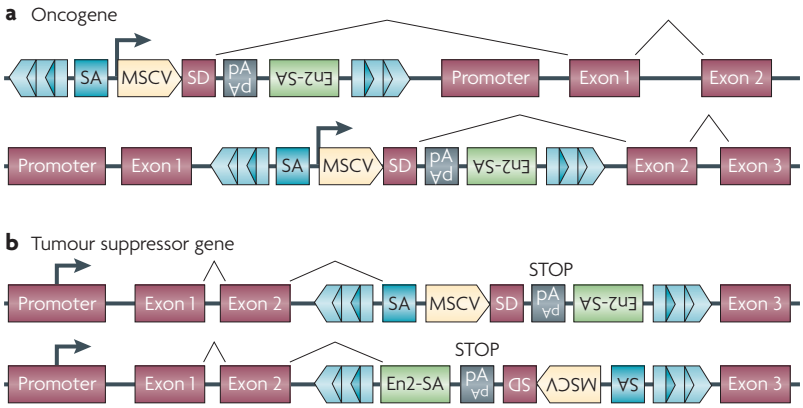

Figure 2 | **T2/Onc2 can deregulate the expression of oncogenes or inactivate expression of tumour suppressor genes. a** | T2/Onc2 contains a murine stem cell virus (MSCV) 5′ long terminal repeat (LTR) derived from the MSCVneo vector (Clontech) and a splice donor (SD) site derived from exon 1 of the mouse *Foxf2* gene[50]. T2/Onc2 can thus promote the expression of an oncogene when integrated upstream of or within the gene in the same transcriptional orientation. T2/Onc2 is flanked by optimized *SB* transposase binding sites (light and dark blue triangles) that are located within the transposon inverted terminal repeats (light blue arrows), which increase the frequency of *SB* transposition[44]. **b** | T2/Onc2 also contains two splice acceptors and a bi-directional polyA (pA) and can thus prematurely terminate transcription of a tumour suppressor gene when integrated within the tumour supressor gene in either orientation. One splice acceptor is derived from exon 2 of mouse engrailed 2 (En2-SA) and the other from the carp β-actin gene (SA)[50].

The double transgenic mice that made it to birth were monitored for tumour development and began to show signs of illness by 7 weeks of age[42]. By 17 weeks, all of these mice had died from cancer. Haematopoietic tumours largely of T cell origin predominated, but some solid tumours, including medulloblastoma, pituitary and intestinal neoplasia, were also detected at low frequencies. Southern blot analysis of tumour DNA showed multiple clonal and subclonal *SB* integrations, which were consistent with an insertional mutagenesis mechanism of tumour induction.

To confirm that the tumours were induced by insertional mutagenesis, the transposon-mouse DNA junction fragments from 16 tumours (primarily T cell lymphomas) were amplified by PCR and sequenced (FIG. 5). In total, 781 unique *SB* insertion sites were sequenced from these 16 tumours (49 insertions per tumour). Several factors account for the amplification of so many unique insertions from a small number of tumours. First, tumour cells contain multiple clonal and subclonal disease-causing insertions (driver mutations), which mark the location of cooperating cancer-causing genes. Clonal insertions are thought to be involved in tumour initiation and subclonal insertions in tumour progression. Second, tumour cells contain multiple non-disease-causing insertions (passenger mutations). Last, tumours contain contaminating normal cells that also harbour different *SB* insertion patterns.

Genomic regions in tumour cells that contained a higher frequency of *SB* insertions than predicted by chance were then identified by statistical means (BOX 1) and CAN genes in these regions were noted as common insertion sites (CIS). Unlike retroviruses, there is little evidence to indicate that T2/Onc2 contains strong enhancer activity even though T2/Onc2 contains a murine stem cell virus (MSCV) LTR. Therefore, *SB* insertions located in or near oncogenes are almost always oriented in the same transcriptional orientation as the oncogene, near the 5′ end or in an intron, whereas insertions located in tumour suppressor genes show little directional bias and are scattered throughout the coding region (FIG. 2b). It is therefore much easier to identify CAN genes that are mutated by *SB* as opposed to retroviruses, and the pattern of *SB* insertions at each locus can provide a strong predictor of whether a CAN gene functions as an oncogene or a tumour suppressor gene.

Among the 20 CAN genes identified in these 16 tumours, 7 are validated human cancer genes and another 7 are mutated by retroviral insertional mutagenesis in mouse leukaemias[42]. This demonstrated for the first time the power of *SB* mutagenesis for cancer gene discovery. Another six CAN genes were not implicated in human cancer. These are the most interesting, as they are the genes that represent potential new therapeutic targets. Of the ten T cell leukaemias analysed, six (60%) contained insertions in intron 27 of *Notch1*. All insertions were oriented in the same transcriptional direction as *Notch1* and resulted in the expression of a truncated *Notch1* transcript containing the MSCV promoter and

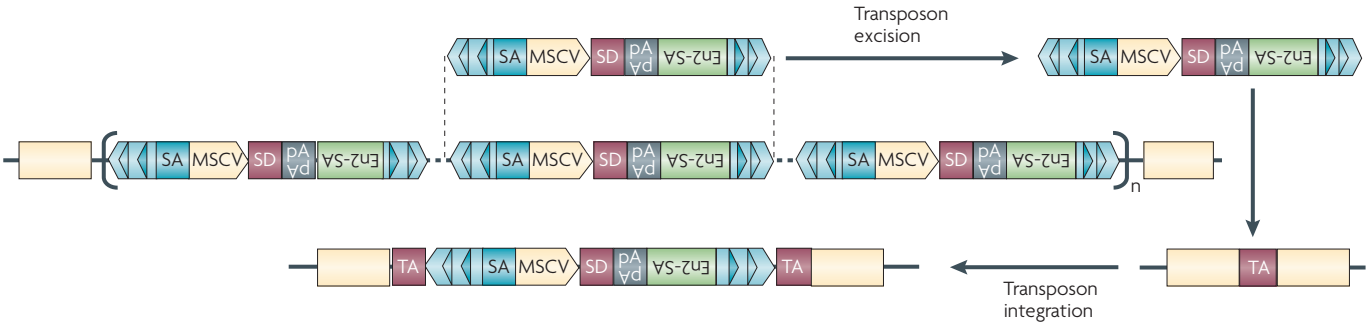

Figure 3 | **High copy T2/Onc2 transgenic lines.** T2/Onc2-containing plasmid DNA was linearized by cleavage with *ScaI*, which cuts only within the plasmid sequences (dashed line)[42], and microinjected into (B6C3) F_2 hybrid embryos using standard techniques. Three high copy T2/Onc2 transgenic founder mice (with 148, 214 and 358 copies) were subsequently identified and transgenic lines established. The T2/Onc2 transposon concatamers within each transgenic line were transmitted at normal Mendelian frequencies, mapped to different mouse chromosomes, and had no heterozygous phenotypes[42]. Mobilization of an individual transposon from the concatamer leaves behind flanking plasmid sequences, which are then joined to create a concatamer of n-1. The excised transposon is then free to reintegrate at a TA dinucleotide elsewhere in the genome, either near the concatamer (local hopping) or on a different chromosome. Mouse genomic DNA is highlighted in yellow. MSCV, murine stem cell virus; pA, polyA; SD, splice donor.

the 3′ end of *Notch1*. This was remarkable as *NOTCH1* is also mutated in ~60% of human T cell acute leukaemias (T-ALL). The fusion transcript induced by the *SB* insertions also mimicked that seen in patients with T-ALL with t(7;9), in which the translocation drives the overexpression of an activated *NOTCH1* carboxy-terminal protein fragment[49].

There were also several instances in which two or more mouse CAN genes were mutated in the same tumour, suggesting that they cooperate to induce cancer. In addition, there were many genes that were mutated in a single tumour only and thus did not meet the statistical criteria for inclusion on the CAN gene list. However, pathway analysis showed that many of these genes function in a limited number of signalling pathways (such as receptor recycling, cellular transport, tumour necrosis factor (TNF), Ras, Jak–signal transducer and activator of transcription (STAT), Ets transcription factor and Wnt), suggesting that some might also be disease causing. By analogy, human tumours are also known to contain a large number of infrequently mutated genes that function in a restricted number of signalling pathways[2].

SB mutagenesis identifies ARF cooperating genes

Collier and colleagues[50] showed that *SB* mutagenesis can also accelerate the formation of soft tissue sarcomas in *Arf*-deficient mice. In these studies, *Arf*-deficient mice carrying an *SB10* transposase transgene under the control of the ubiquitous CAGGS promoter (*CAGGS–SB10*) were crossed with *Arf*-deficient mice carrying a low copy T2/Onc2 transposon concatamer (~25 copies). Although mobilization of T2/Onc2 by *CAGGS–SB10* was too low to induce tumours in wild-type mice, it was sufficient to accelerate the formation of sarcomas in *Arf*-deficient mice. The most frequently mutated gene was *Braf*, which was mutated in 22 of 28 tumours. All *SB* insertions were located in intron 9 and were oriented in the same transcriptional direction as *Braf*. The truncated BRAF produced by these insertions contained the kinase domain but lacked the amino-terminal negative regulatory elements of wild-type BRAF, and was capable of morphological transformation of NIH 3T3 cells. Activating *BRAF* point mutations have also been identified in 9% of human sarcoma cell lines and 0.5–5% of primary human sarcomas[51,52], confirming that *SB* mutagenesis can be used to identify cancer genes in solid tumours.

CAGGS–SB10 also failed to induce tumours when crossed to high copy T2/Onc2 transposon transgenic mice[53]. Epigenetic silencing of the *CAGGS–SB10* transgene[53], which was not observed in *Rosa26–SB11* mice[53], probably caused this. This shows the value of the knock-in approach used by Dupuy and colleagues for expressing the transposase. Interestingly, *Rosa26–SB11* could induce tumours when crossed with low copy T2/Onc2 transgenic mice, although the median latency of tumours was 6–7 months compared with 68 days for high copy T2/Onc2 transgenic lines[53]. Tumours were primarily haematopoietic, although histological examination revealed focal areas of epithelial hyperplasia in the anterior and dorsolateral prostate in some mice[54], as well as astrocytomas in the brains of other mice[53]. Sequence analysis of the *SB* insertion sites in prostate lesions identified several prostate CAN genes, including phosphodiesterase 4D, cAMP-specific (*Pde4d*)[54]. *PDE4D* is overexpressed in human prostate cancer and shows changes in its isoform expression. Knock down of *PDE4D* also reduces the growth and proliferation rate of prostate cancer xenografts *in vivo*[54], demonstrating that *SB* mutagenesis can also identify genes involved in the control of prostate cancer cell growth. Analysis of the *SB* insertion sites present in the astrocytomas revealed a high frequency of colony stimulating factor 1 receptor (*Csf1r*) insertions, clustered in introns 5 and 8, which resulted in increased levels of *Csf1r* mRNA and protein[53]. *CSF1* and *CSF1R* are also overexpressed in high-grade human astrocytomas, suggesting a causal link between increased *CSR1R* expression and astrocytoma formation.

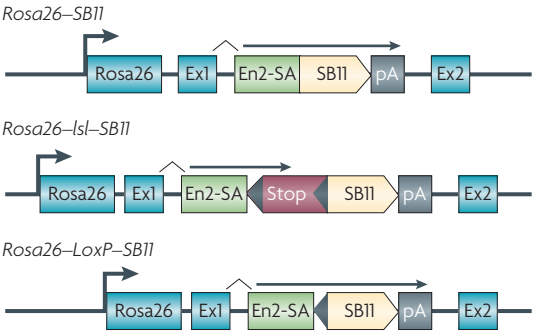

Figure 4 | **Two Rosa26 *SB* transposase knock-in alleles.** A genetically enhanced Sleeping Beauty (*SB*) transposase (SB11) expression cassette consisting of an En2 splice acceptor (En2-SA), SB11 complimentary DNA and an SV40 polyA (pA) signal, was knocked in to intron 1 of the mouse *Rosa26* locus to generate the *Rosa26–SB11* allele. In a second round of targeting, a transcriptional stop cassette flanked by *loxP* sites (black triangles), containing two SV40 pA sites and one bovine growth hormone pA site inserted downstream of the green fluorescent protein cDNA (Clontech), was inserted between the En2-SA and SB11 to generate the conditional *Rosa26–lsl–SB11* allele. Cre recombinase can be used to remove the transcriptional stop cassette to generate the *Rosa26–LoxP–SB11* allele and activate transposase expression.

T2/Onc3 induces epithelial tumours

Although several factors could explain the haematopoietic tumour bias observed in *Rosa26–SB11*;T2/Onc2 mice, one obvious potential contributing factor is the choice of the MSVC promoter included in T2/Onc2 to drive oncogene overexpression[42] (FIG. 2). This promoter is derived from a cloned Moloney murine leukaemia virus[55] and is therefore likely to drive oncogene expression in haematopoietic cells at a higher rate than in other cell types. To investigate this possibility, a new transposon, T2/Onc3, was created, which is identical to T2/Onc2, except that the MSCV LTR is replaced with a CAGGS promoter[56]. CAGGS contains the cytomegalovirus immediate early enhancer fused to a chicken β-actin minimal promoter and is expressed strongly in various cell types, including epithelial cells[57,58]. T2/Onc3 was introduced into the mouse germ line by microinjection and mobilized by breeding to *Roas26–SB11* mice. As the mice aged they developed tumours; however, the average age of death of these mice was 55 weeks compared with 68 days for *Rosa26–SB11*;T2/Onc2 mice[56]. Most tumours were epithelial in origin and each animal had on average three independent types of cancer at the time of euthanasia. Squamous cell carcinomas of the skin and hepatocellular carcinomas of the liver were the most predominant classes of tumours; however, many other cancers were observed at lower frequencies, including cancers of the colon, lung, brain, breast, nasal cavity and ovary, as well as the salivary, preputial, parathyroid and adrenal glands. These results show that the nature of the promoter included in the transposon matters and suggest that it will be important to tailor the structure of the transposon to the type of cancer being modelled.

Preputial
Exocrine glands that produce pheromones found in front of the genitals in some mammals.

A conditional *SB* transposition system

In order to better control the types of tumours induced by *SB*, a conditional *SB* transposition system was developed by introducing a floxed transcriptional stop cassette (lox–stop–lox; lsl) into the 5′ end of SB11, which itself had been knocked in to the *Rosa26* locus (*Rosa26–lsl–SB11*)[56,59,60] (FIG. 4). When the stop cassette is present the transposase is not expressed but can be reactivated by breeding *Rosa26–lsl–SB11* mice with mice carrying a tissue-specific Cre transgene. Cre will recombine out the stop cassette to produce *Rosa26–LoxP–SB11* mice and allow the transposase to be expressed (FIG. 4). This makes it possible to restrict transposase expression to virtually any target tissue and eliminate the embryonic lethality caused by ubiquitous transposase expression. One can also incorporate a sensitizing mutation into the cross, such as activated *Kras* or mutant *Trp53* allele, in order to drive the formation of certain types of tumours or accelerate the formation of tumours naturally induced by *SB* (FIG. 6).

Colorectal cancer. To evaluate the potential of this conditional transposition system for selectively modelling cancer, *Rosa26–lsl–SB11*;T2/Onc2 mice were bred with *Villin–Cre* mice to activate *SB* transposition in the epithelial cells of the gastrointestinal tract[60]. As expected, *Rosa26–lsl–SB11*;T2/Onc2;*Villin–Cre* mice did not develop haematopoietic tumours, as transposition was not triggered in the haematopoietic compartment. Instead, the mice developed intraepithelial neoplasias, adenomas and a low frequency of adenocarcinomas in the gastrointestinal tract late in life (>300 days). Analysis of 16,690 PCR-amplified, sequenced and mapped nonredundant insertions from 135 gastrointestinal tract tumours identified 77 mouse gastrointestinal tract CAN genes[60] (BOX 1). There are 38 genes with human homologues in the catalogue of somatic mutations in human cancer (COSMIC) database (see Further information), of which 18 (47%) have documented non-silent mutations in human cancer, which is significantly higher than expected by chance ($p < 0.05$). If the data comparison is limited to COSMIC genes mutated in human colorectal cancer (CRC), the overlap is more significant ($p < 0.001$).

By eliminating genes that were not mutated, amplified, deleted or misregulated in human CRC, the list of mouse CAN genes could be narrowed to 15 genes, which are the most likely to be driver genes in human CRC[60]. The most frequent gene was *Apc* (61 insertions). *APC* is a member of the Wnt signalling pathway and is the most commonly mutated gene in human CRC (70% to 80% of human CRCs have *APC* mutations)[61]. Three other genes on the list, *Bmpr1a*, *Smad4* and *Pten*, are mutated in juvenile polyposis syndrome, juvenile intestinal polyposis and Cowden disease, and another gene, *Fbxw7*, is a component of the SCF ubiquitin ligase complex that is mutated in 11.5% of human CRC[62]. Therefore, 5 of the 15 genes identified are validated human CRC genes and together represent some of the most commonly mutated genes identified in human CRC.

Ingenuity pathway analysis
A curated database and analysis system used to explain how proteins work together to cause cellular changes.

Hepatocellular carcinoma. *Rosa26–lsl–SB11*;T2/Onc2 mice were mated with *albumin–Cre* transgenic mice to activate *SB* transposition in the liver[59]. A conditional dominant-negative *Trp53* transgene (*p53–lsl–R270H*)[63] was also included as a sensitizing mutation as *Trp53* mutations are common in human hepatocellular carcinoma (HCC)[64,65]. Characterization of 67 preneoplastic liver nodules isolated from 6 mice identified 19 HCC CAN genes comprising homologues of genes already implicated in tumour formation and apoptosis, such as epidermal growth factor receptor (*EGFR*), hypoxia induced factor 1α (*HIF1A*), *MAP2K4*, *MET*, *PAK4*, *VRK2*, *TRPM7* and *TAOK3*.

Egfr was the most frequently mutated gene (58 of 67 nodules). All *Egfr* insertions were located in intron 24 where they promoted the expression of a truncated EGFR protein containing the kinase domain but lacking the C-terminal domain. C-terminal deletions in human *EGFR* increase its autokinase activity and transforming ability, and have tumorigenic properties[66–68]. *EGFR* is also overexpressed in 15–40% of human HCCs, and EGF signalling is activated in ~50% of human HCCs[69,70]. Finally, truncated *EGFR* was linked to fumaryl acetoacetate hydrolase (*Fah*), which allows selective repopulation of the livers of FAH-deficient mice, and shown to induce hyperplastic nodules in the livers of FAH-deficient mice, which were visible by 43 days after injection[59]. So far, however, the type of activating somatic mutations found in the tyrosine kinase domain of *EGFR* in human non-small-cell lung cancer[71] have not been reported for *EGFR* in human HCC.

Moreover, the human orthologues of four mouse HCC CAN genes also showed increases in copy number in human HCC, including *EGFR*, *SLC25A13*, *MET* and *UBE2H*[59]. Overexpression of *MET* in human HCC is well characterized[72] and mice carrying mutations in *Met* are also prone to liver tumours[73]. The human orthologues of four mouse HCC CAN genes showed decreases in copy number in human HCC, including *MARCH1*, *PSD3*, *MAP2K4* and *NFIB*[73]. *MAP2K4* is a putative tumour suppressor gene in tumours of the breast, prostate and pancreas, and may thus have a similar function in the liver[74–76].

Ingenuity pathway analysis identified two pathways that were over-represented among the 19 mouse HCC CAN genes. One network includes the NFIB and HIF1A transcription factors, which transduce EGFR-initiated phosphorylation signalling cascades. The second network comprises genes that interact with *TNF*. TNF can induce tyrosine phosphorylation and internalization of *EGFR*, and activate NF-κB, which in turn regulates apoptosis during liver tumorigenesis[77]. These results provide additional evidence for a role for *Egfr* signalling in HCC.

HCC is commonly found in people who have contributing factors for HCC, such as cirrhosis of the liver or who are chronically infected with the hepatitis B or C virus[78–81]. These agents induce chronic inflammation and associated reactive hyperplasia in the liver, which is thought to be important for HCC development. In the future, *SB*-induced HCC should be examined in mice that are sensitized by one or more of these contributing factors.

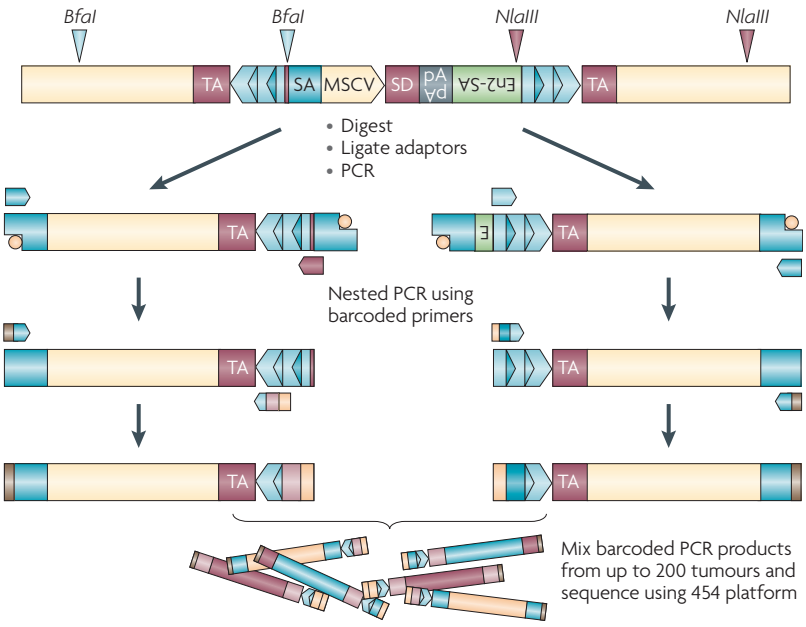

Figure 5 | **Barcoded splinkerette PCR amplification and deep sequencing.** A modified splinkerette PCR method using barcoded primers was used to amplify the Sleeping Beauty (*SB*) transposon insertions from mouse tumours so that they could be directly sequenced using a GS FLX Genome Sequencer[59,60]. Briefly, tumour DNA was digested with *BfaI* or *NlaIII* to generate left or right transposon-mouse DNA junction fragments, respectively. *NlaIII* cleaves the transposon within the En2-SA sequences and *BfaI* cleaves within the transposon sequences located between the inverted terminal repeats and the carp β-actin SA (red rectangle). Double-stranded splinkerettes (blue rectangle) containing a C3 spacer blocking group (yellow circle) at the 3′ end of the splinkerette, which helps to prevent mispriming[101], were then ligated onto the ends of each fragment and the transposon-mouse junction fragments PCR amplified using primers complimentary to transposon sequences (blue pointed rectangle) and the splinkerette (red pointed rectangle). The resulting PCR products were then used as a template in a nested PCR, in which the primers incorporated the A/B tags at their 5′ ends (brown and orange rectangles) that are used for sample processing on the FLX Genome Sequencer. Additionally, a 10 base pair (bp) barcode (red and dark blue rectangles)[102] was added so that it would be sequenced just before the transposon tag. This makes it possible to amplify transposon junction from up to 200 different *SB*-induced tumours, pool the resulting products and then sequence all the PCR products on a single run. Each sequence is then assigned to the tumour it came from using the barcode. SD, splice donor.

The range of genes mutated by SB varies among different cancer types. The range of genes mutated by *SB* varies widely in different types of mouse cancer[42,50,53,54,59,60], similar to that observed in human cancer[2]. Therefore, it will probably be necessary to examine all the different cancers that affect humans in order to obtain a complete picture of the genes and signalling pathways that cause cancer. This will be a massive undertaking as by some estimates there are more than 200 types of cancer affecting more than 60 different organs in our body (see the CancerHelp UK website; Further information).

Most mutations are located in oncogenes. Nearly 90% of transposon-based mutations seem to represent gain-of-function mutations in oncogenes, although there is still an appreciable number (~10%) that seem to represent

Kernel function
A weighting function used in nonparametric function estimation. It gives the weights of the nearby data points in making an estimate.

Bonferroni correction
A method used to address the problem of multiple comparisons.

loss-of-function mutations in tumour suppressor genes (N.G.C. and N.A.J., unpublished observations). Some are well-known tumour suppressor genes such as *Apc*, *Pten* and *Nf1*, but others seem to represent new tumour suppressor genes. In most instances, only one allele is mutated, although there are some genes such as *Ikzf1* in which both alleles are often mutated (N.G.C. and N.A.J., unpublished observations). It remains to be determined whether these tumour suppressor genes are functioning as haploinsufficient genes or whether the second allele is inactivated by some other genetic event, such as spontaneous mutation or epigenetic silencing.

Local hopping in cancer cells

SB is not locked in place after transposition and can continuously transpose to new sites. This can effectively reduce the amount of local hopping observed in cancer cells as by the time a tumour is formed each *SB* transposon will have undergone multiple rounds of transposition and many transposons will have been lost from the genome. The reduction in local hopping provides an increased chance that *SB* will mutate a cancer gene locus outside the local hopping interval during the course of tumour progression. In germline mutagenesis screens, the frequency of local hopping varies between 50% and 80%[36,38,39], whereas in sarcomas and gastrointestinal tract tumours it is 23% and 45%, respectively[50,60]. The lowest frequency is found in haematopoietic tumours, with only 6–11% showing evidence of local hopping[42]. *SB*, like most cut and paste transposons, requires DNA synthesis for transposition[82]. These differences may therefore reflect the number of times a cancer-initiating cell has divided, and so the number of rounds of transposition that have occurred, before a cancer-initiating cells is converted into a tumour cell. In addition, the insertions recovered from tumours are by definition selected, and selection could also mask the true insertional patterns of the transposons. This could also introduce a bias into the observed insertion patterns for insertions that are recovered from tumours.

Box 1 | Identification of CAN genes

Two statistical methods have been developed for identifying genomic regions in cancer cells that contain a higher number of Sleeping Beauty (*SB*) insertions than predicted by random chance (common insertion sites; CIS) and are therefore likely to contain candidate cancer (CAN) genes. The first uses Monte Carlo simulations to count the expected number of CIS of any given size in a specified window, taking into account the exact distribution of TA dinucleotides across the genome, unfinished regions that are represented by long tracts of Ns and the number of insertions mapped in the study[53,59,60]. Once the total count of insertions is randomly distributed among the real TA sites across the genome, a summary of the number of CIS of size ≥ 3, ≥ 4… $\geq n$ is recorded within windows of different sizes (that is, 10 kilobases (kb), 30 kb, 60 kb and 100 kb). This process is then repeated hundreds of times and the average counts of those reiterations computed. The values obtained can be interpreted as expected values (E-values) as they indicate the number of CIS of a given size that would be expected in a window of a given size by random chance. E-value thresholds of $E<1$ are typically used, which means that on average one CIS or fewer will have occurred by random chance. Small window sizes bias against discovering large genes that have disruptive insertions evenly spaced throughout the gene, and large window sizes bias against discovering small genes with tightly clustered insertions. It is therefore best to repeat the analysis using different window sizes.

The second method makes use of a Gaussian Kernel Convolution framework[98,99]. A Kernel function is positioned at every insertion site in the data set. For any position in the genome, an estimate of the number of insertions can be obtained by summing all the Kernel functions. The analysis is carried out repeatedly using different Kernel sizes (that is, 10 kb, 30 kb, 60 kb and 100 kb). The significance of the peaks is then evaluated using a null-distribution computed by means of a random permutation of the data, taking into account the true distribution of TA dinucleotides in the genome. The probability of detecting false CIS is controlled by applying the Bonferroni correction[100]. The kernel convolution method offers the advantage that it evaluates the insertion data set on multiple scales (using different kernel sizes), allowing the identification of a higher number of CAN genes in a single analysis.

Both statistical approaches generate CIS lists that are similar but not identical (N.G.C. and N.A.J., unpublished observations). Currently, it seems best to carry out both types of statistical analysis and then combine the two gene lists.

An alternative transposition system

Since the awakening of *SB*, several other transposons have been identified that can transpose in mouse cells such as Minos, piggyBac (PB) and Tol2 (REFS 83–86). Among these transposons, the one that has received the most attention is PB, which was first identified when it jumped from its insect host, the cabbage looper moth *Trichoplusia ni*, into the baculovirus genome[87]. Similar to *SB*, PB transposes through a cut and paste mechanism, but its preferred insertion site is TTAA, which is duplicated on insertion[88,89].

PB has several features that make it an attractive alternative insertional mutagen. A mouse codon-optimized PB transposase (mPB) has been developed that is 300 times more efficient than SB11 (REFS 45,90), and 13 times more efficient than a recently described hyperactive *SB* transposase, SB100, in mouse ES cells[45,91]; however, the opposite has been reported in human HeLa and CD34 cells[91,92]. In addition, mPB has been fused to a mutated oestrogen receptor, ERT2 (mPB-ERT2), so that it provides a high level of transposase activity[90]. The inducible ERT2 domain makes it possible to tightly regulate the activity of the transposase by the addition of tamoxifen or 4-hydroxytamoxifen (4-OHT). In the absence of tamoxifen, heat-shock proteins sequester mPB-ERT2 in the cytoplasm, which prevents it from functioning[93]. In the presence of tamoxifen, mPB-ERT2 is released and can translocate to the nucleus and activate transposition[94]. This should make it possible to turn off PB transposition after tumours are initiated and so potentially reduce the number of background PB insertions in tumour cells. PB can also carry much larger cargo than *SB*. For every 1 kb increase in size from its original 1.7 kb length, *SB* shows around a 30% decrease in transposition frequency[17]. PB, conversely, can carry up to 9.1 kb of foreign DNA without significantly reducing its integration frequency[84]. In addition, PB shows little propensity for local hoping and preferentially integrates into genes[45], which should increase the ability of PB to insertionally mutate cancer genes. This is perhaps the most important difference to *SB*, as the different target site selection properties of PB are predicted to result in a different range of genes that can be mutated by PB and *SB*. Finally, canonical PB excises cleanly from the host genome, unlike *SB*, which leaves behind a 5 bp mutagenic footprint following excision, although certain PB excisions are still associated with mutations at the excision

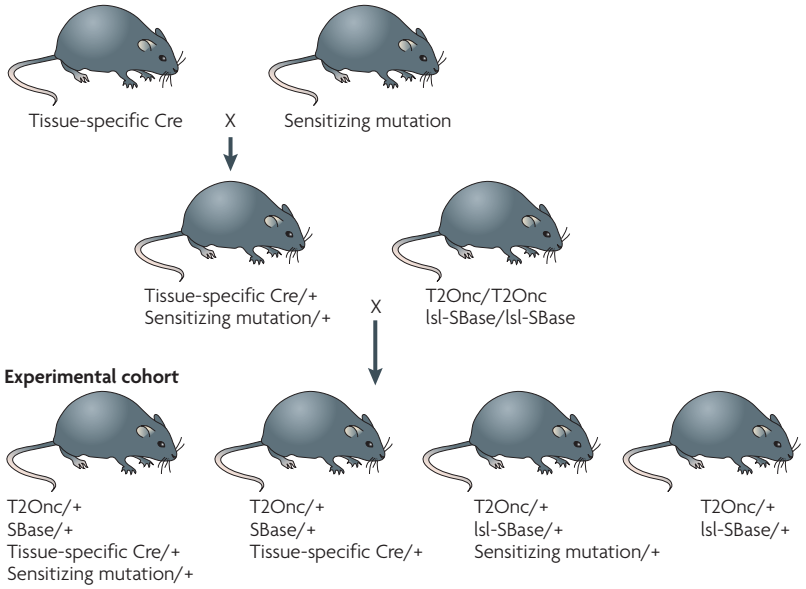

Figure 6 | **Breeding scheme used for mobilizing *SB* in a sensitized mouse genetic background.** Mice carrying a tissue-specific Cre transgene are crossed to mice carrying a sensitizing mutation, such as activated *Kras* transgene or inactivated *Trp53* gene. The resulting double heterozygotes are then crossed to mice that are homozygous for T2/Onc2 or T2/Onc3 (T2Onc) and the conditional transposase allele, *Rosa26–lsl–SB11* (lsl–SBase). Progeny mice carrying the four expected resultant genotypes (experiment cohort) are then assessed for tumour development as they age. *SB*, Sleeping Beauty.

sites[84,95] (FIG. 1b). PB transposition should therefore induce fewer cancer-causing mutations that cannot be detected because they are no longer tagged by the transposon.

Although there is currently no published data showing that PB can function as an effective cancer gene mutagen, there are data showing that PB is a versatile cancer gene mutagen, which is capable of identifying cancer genes not yet identified in retroviral or *SB* transposon screens (R. Rand, P. Liu and A. Bradley, personal communication). PB thus seems to represent an attractive alternative mutagen for cancer gene discovery in the mouse.

Using transposons to aid drug discovery

Although primarily generated for insertional mutagenesis, transposon-based systems also harbour huge potential for studying many other aspects of cancer biology. For example, a recent PB-based screen in ES cells[96] has identified a new genetic interaction between BRCA1 and 53BP1 that is relevant to breast cancer. Notably, loss of 53BP1 was shown to partially restore the homologous-recombination defect of *BRCA1*-depleted cells and revert their hypersensitivity to DNA-damaging agents. Transposon-based systems can also be used to identify genes that provide resistance or sensitivity to chemotherapeutic agents, as demonstrated for retroviral insertional mutagenesis, which was used to identify candidate genes involved in the resistance to MEK inhibitors in leukaemias that were initiated by hyperactive Ras[97].

Future perspectives

Although it has only been 5 years since it was first shown that *SB* can be mobilized at frequencies high enough to induce cancer[42,50], it has already become apparent that *SB* can be used to model many types of human cancer in mice. This, in turn, should allow the rapid identification of many of the mutant genes and signalling pathways that drive cancer in mice. Using laser capture microdissection, it should also be possible to isolate and clone the *SB* insertions from preneoplastic, neoplastic and metastatic lesions from the same animal. Because each cancer-initiating cell is marked by a unique collection of *SB* insertions, it may then be possible to determine whether the individual lesions are descended from a single cancer-initiating cell or from multiple cancer-initiating cells. Moreover, this might also make it possible to determine the order in which the mutations were acquired and whether they are involved in the initiation, progression and metastasis of the tumour. *SB*-induced mouse cancer models will also provide a plethora of new models for furthering our understanding of the biology of cancer, for identifying new drug targets and for preclinical testing of new cancer therapeutics before they go into the clinic.

1. Imai, K. & Takaoka, A. Comparing antibody and small-molecule therapies for cancer. *Nature Rev. Cancer* **6**, 714–727 (2006).
2. Copeland, N. G. & Jenkins, N. A. Deciphering the genetic landscape of cancer-from genes to pathways. *Trends Genet.* **25**, 455–462 (2009).
3. Forrest, W. F. & Cavet, G. Comment on "The consensus coding sequences of human breast and colorectal cancers". *Science* **317**, 1500 (2007).
4. Getz, G. *et al.* Comment on "The consensus coding sequences of human breast and colorectal cancers". *Science* **317**, 1500 (2007).
5. Rubin, A. F. & Green, P. Comment on "The consensus coding sequences of human breast and colorectal cancers". *Science* **317**, 1500 (2007).
6. Kool, J. & Berns, A. High-throughput insertional mutagenesis screens in mice to identify oncogenic networks. *Nature Rev. Cancer* **9**, 389–399 (2009). **This paper reviews the advantages and the limitations of retroviral insertional mutagenesis in mice for identifying genes and signalling pathways that are important for cancer.**
7. Akagi, K., Suzuki, T., Stephens, R. M., Jenkins, N. A. & Copeland, N. G. RTCGD: retroviral tagged cancer gene database. *Nucleic Acids Res.* **32**, D523–D527 (2004).
8. Mitchell, R. S. *et al.* Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. *PLoS Biol.* **2**, e234 (2004).
9. Wu, X., Li, Y., Crise, B. & Burgess, S. M. Transcription start regions in the human genome are favored targets for MLV integration. *Science* **300**, 1749–1751 (2003).
10. Neil, J. C. & Cameron, E. R. Retroviral insertion sites and cancer: fountain of all knowledge? *Cancer Cell* **2**, 253–255 (2002).
11. Plasterk, R. H., Izsvak, Z. & Ivics, Z. Resident aliens: the Tc1/mariner superfamily of transposable elements. *Trends Genet.* **15**, 326–332 (1999).
12. Lohe, A. R., Moriyama, E. N., Lidholm, D. A. & Hartl, D. L. Horizontal transmission, vertical inactivation, and stochastic loss of mariner-like transposable elements. *Mol. Biol. Evol.* **12**, 62–72 (1995).
13. Izsvak, Z., Ivics, Z. & Hackett, P. B. Repetitive elements and their genetic applications in zebrafish. *Biochem. Cell Biol.* **75**, 507–523 (1997).
14. Ivics, Z., Izsvak, Z., Minter, A. & Hackett, P. B. Identification of functional domains and evolution of Tc1-like transposable elements. *Proc. Natl Acad. Sci. USA* **93**, 5008–5013 (1996).
15. Zayed, H., Izsvak, Z., Walisko, O. & Ivics, Z. Development of hyperactive sleeping beauty transposon vectors by mutational analysis. *Mol. Ther.* **9**, 292–304 (2004).
16. Ivics, Z., Hackett, P. B., Plasterk, R. H. & Izsvak, Z. Molecular reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells. *Cell* **91**, 501–510 (1997). **A landmark paper that constructed the first synthetic highly active transposon (SB) for use in mammalian cells.**
17. Izsvak, Z., Ivics, Z. & Plasterk, R. H. Sleeping Beauty, a wide host-range transposon vector for genetic transformation in vertebrates. *J. Mol. Biol.* **302**, 93–102 (2000).
18. Yant, S. R. *et al.* Somatic integration and long-term transgene expression in normal and haemophilic mice using a DNA transposon system. *Nature Genet.* **25**, 35–41 (2000).
19. Luo, G., Ivics, Z., Izsvak, Z. & Bradley, A. Chromosomal transposition of a Tc1/mariner-like element in mouse embryonic stem cells. *Proc. Natl Acad. Sci. USA* **95**, 10769–10773 (1998).
20. Dupuy, A. J. *et al.* Mammalian germ-line transgenesis by transposition. *Proc. Natl Acad. Sci. USA* **99**, 4495–4499 (2002).
21. Dupuy, A. J., Fritz, S. & Largaespada, D. A. Transposition and gene disruption in the male germline of the mouse. *Genesis* **30**, 82–88 (2001).

22. Fischer, S. E., Wienholds, E. & Plasterk, R. H. Regulated transposition of a fish transposon in the mouse germ line. *Proc. Natl Acad. Sci. USA* **98**, 6759–6764 (2001).
23. Horie, K. *et al.* Efficient chromosomal transposition of a Tc1/mariner- like transposon Sleeping Beauty in mice. *Proc. Natl Acad. Sci. USA* **98**, 9191–9196 (2001).
24. Zayed, H., Izsvak, Z., Khare, D., Heinemann, U. & Ivics, Z. The DNA-bending protein HMGB1 is a cellular cofactor of Sleeping Beauty transposition. *Nucleic Acids Res.* **31**, 2313–2322 (2003).
25. Walisko, O. *et al.* Sleeping Beauty transposase modulates cell-cycle progression through interaction with Miz-1. *Proc. Natl Acad. Sci. USA* **103**, 4062–4067 (2006).
26. Izsvak, Z. *et al.* Healing the wounds inflicted by sleeping beauty transposition by double-strand break repair in mammalian somatic cells. *Mol. Cell* **13**, 279–290 (2004).
27. Yant, S. R. & Kay, M. A. Nonhomologous-end-joining factors regulate DNA repair fidelity during Sleeping Beauty element transposition in mammalian cells. *Mol. Cell. Biol.* **23**, 8505–8518 (2003).
28. Ikeda, R. *et al.* Sleeping beauty transposase has an affinity for heterochromatin conformation. *Mol. Cell. Biol.* **27**, 1665–1676 (2007).
29. Yusa, K., Takeda, J. & Horie, K. Enhancement of Sleeping Beauty transposition by CpG methylation: possible role of heterochromatin formation. *Mol. Cell. Biol.* **24**, 4004–4018 (2004).
30. Garrison, B. S., Yant, S. R., Mikkelsen, J. G. & Kay, M. A. Postintegrative gene silencing within the Sleeping Beauty transposition system. *Mol. Cell. Biol.* **27**, 8824–8833 (2007).
31. Park, C. W., Kren, B. T., Largaespada, D. A. & Steer, C. J. DNA methylation of Sleeping Beauty with transposition into the mouse genome. *Genes Cells* **10**, 763–776 (2005).
32. Park, C. W., Park, J., Kren, B. T. & Steer, C. J. Sleeping Beauty transposition in the mouse genome is associated with changes in DNA methylation at the site of insertion. *Genomics* **88**, 204–213 (2006).
33. Walsh, C. P. & Bestor, T. H. Cytosine methylation and mammalian development. *Genes Dev.* **13**, 26–34 (1999).
34. Yoder, J. A., Walsh, C. P. & Bestor, T. H. Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet.* **13**, 335–340 (1997).
35. Liu, G. *et al.* Target-site preferences of Sleeping Beauty transposons. *J. Mol. Biol.* **346**, 161–173 (2005).
36. Vigdal, T. J., Kaufman, C. D., Izsvak, Z., Voytas, D. F. & Ivics, Z. Common physical properties of DNA affecting target site selection of sleeping beauty and other Tc1/mariner transposable elements. *J. Mol. Biol.* **323**, 441–452 (2002).
37. Yant, S. R. *et al.* High-resolution genome-wide mapping of transposon integration in mammals. *Mol. Cell. Biol.* **25**, 2085–2094 (2005).
38. Carlson, C. M. *et al.* Transposon mutagenesis of the mouse germline. *Genetics* **165**, 243–256 (2003).
39. Horie, K. *et al.* Characterization of Sleeping Beauty transposition and its application to genetic screening in mice. *Mol. Cell. Biol.* **23**, 9189–9207 (2003).
40. Tower, J., Karpen, G. H., Craig, N. & Spradling, A. C. Preferential transposition of *Drosophila* P elements to nearby chromosomal sites. *Genetics* **133**, 347–359 (1993).
41. Hanahan, D. & Weinberg, R. A. The hallmarks of cancer. *Cell* **100**, 57–70 (2000).
42. Dupuy, A. J., Akagi, K., Largaespada, D. A., Copeland, N. G. & Jenkins, N. A. Mammalian mutagenesis using a highly mobile somatic Sleeping Beauty transposon system. *Nature* **436**, 221–226 (2005).
This was the first paper showing that SB transposon-based mutagenesis can induce cancer when mobilized in somatic cells of wild-type mice.
43. Geurts, A. M. *et al.* Gene transfer into genomes of human cells by the sleeping beauty transposon system. *Mol. Ther.* **8**, 108–117 (2003).
44. Cui, Z., Geurts, A. M., Liu, G., Kaufman, C. D. & Hackett, P. B. Structure-function analysis of the inverted terminal repeats of the sleeping beauty transposon. *J. Mol. Biol.* **318**, 1221–1235 (2002).
45. Liang, Q., Kong, J., Stalker, J. & Bradley, A. Chromosomal mobilization and reintegration of Sleeping Beauty and PiggyBac transposons. *Genesis* **47**, 404–408 (2009).
46. Soriano, P. Generalized lacZ expression with the ROSA26 Cre reporter strain. *Nature Genet.* **21**, 70–71 (1999).
47. Zambrowicz, B. P. *et al.* Disruption of overlapping transcripts in the ROSA β geo 26 gene trap strain leads to widespread expression of β-galactosidase in mouse embryos and hematopoietic cells. *Proc. Natl Acad. Sci. USA* **94**, 3789–3794 (1997).
48. Geurts, A. M. *et al.* Gene mutations and genomic rearrangements in the mouse as a result of transposon mobilization from chromosomal concatemers. *PLoS Genet.* **2**, e156 (2006).
49. Ellisen, L. W. *et al.* TAN-1, the human homolog of the *Drosophila* notch gene, is broken by chromosomal translocations in T lymphoblastic neoplasms. *Cell* **66**, 649–661 (1991).
50. Collier, L. S., Carlson, C. M., Ravimohan, S., Dupuy, A. J. & Largaespada, D. A. Cancer gene discovery in solid tumours using transposon-based somatic mutagenesis in the mouse. *Nature* **436**, 272–276 (2005).
This was the first paper showing that SB transposon-based mutagenesis can identify genes involved in solid cancer when mobilized in somatic cells of sensitized mice.
51. Davies, H. *et al.* Mutations of the BRAF gene in human cancer. *Nature* **417**, 949–954 (2002).
52. Seidel, C. *et al.* Alterations of cancer-related genes in soft tissue sarcomas: hypermethylation of RASSF1A is frequently detected in leiomyosarcoma and associated with poor prognosis in sarcoma. *Int. J. Cancer* **114**, 442–447 (2005).
53. Collier, L. S. *et al.* Whole-body sleeping beauty mutagenesis can cause penetrant leukemia/lymphoma and rare high-grade glioma without associated embryonic lethality. *Cancer Res.* **69**, 8429–8437 (2009).
54. Rahrmann, E. P. *et al.* Identification of PDE4D as a proliferation promoting factor in prostate cancer using a Sleeping Beauty transposon-based somatic mutagenesis screen. *Cancer Res.* **69**, 4388–4397 (2009).
55. Hawley, R. G., Fong, A. Z., Burns, B. F. & Hawley, T. S. Transplantable myeloproliferative disease induced in mice by an interleukin 6 retrovirus. *J. Exp. Med.* **176**, 1149–1163 (1992).
56. Dupuy, A. J. *et al.* A modified sleeping beauty transposon system that can be used to model a wide variety of human cancers in mice. *Cancer Res.* **69**, 8150–8156 (2009).
This paper describes the development of a modified SB transposon, T2/Onc3, which selectively induces epithelial tumours when mobilized in somatic cells of wild-type mice.
57. Niwa, H., Yamamura, K. & Miyazaki, J. Efficient selection for high-expression transfectants with a novel eukaryotic vector. *Gene* **108**, 193–199 (1991).
58. Okabe, M., Ikawa, M., Kominami, K., Nakanishi, T. & Nishimune, Y. 'Green mice' as a source of ubiquitous green cells. *FEBS Lett.* **407**, 313–319 (1997).
59. Keng, V. W. *et al.* A conditional transposon-based insertional mutagenesis screen for genes associated with mouse hepatocellular carcinoma. *Nature Biotech.* **27**, 264–274 (2009).
This paper describes the development of a conditional transposition system that can be used to model many different kinds of human cancer in mice. This transposition system was also used in a proof-of-principle experiment to model HCC in mice.
60. Starr, T. K. *et al.* A transposon-based genetic screen in mice identifies genes altered in colorectal cancer. *Science* **323**, 1747–1750 (2009).
This paper describes the development of a conditional transposition system that can be used to model many different kinds of human cancer in mice. This transposition system was also used in a proof-of-principle experiment to model CRC in mice.
61. Segditsas, S. & Tomlinson, I. Colorectal cancer and genetic alterations in the Wnt pathway. *Oncogene* **25**, 7531–7537 (2006).
62. Rajagopalan, H. *et al.* Inactivation of hCDC4 can cause chromosomal instability. *Nature* **428**, 77–81 (2004).
63. de Vries, A. *et al.* Targeted point mutations of p53 lead to dominant-negative inhibition of wild-type p53 function. *Proc. Natl Acad. Sci. USA* **99**, 2948–2953 (2002).
64. Bressac, B., Kew, M., Wands, J. & Ozturk, M. Selective G to T mutations of p53 gene in hepatocellular carcinoma from southern Africa. *Nature* **350**, 429–431 (1991).
65. Buendia, M. A. Genetics of hepatocellular carcinoma. *Semin. Cancer Biol.* **10**, 185–200 (2000).
66. Boerner, J. L., Danielsen, A. & Maihle, N. J. Ligand-independent oncogenic signaling by the epidermal growth factor receptor: v-ErbB as a paradigm. *Exp. Cell Res.* **284**, 111–121 (2003).
67. Chang, C. M. *et al.* A minor tyrosine phosphorylation site located within the CAIN domain plays a critical role in regulating tissue-specific transformation by erbB kinase. *J. Virol.* **69**, 1172–1180 (1995).
68. Frederick, L., Wang, X. Y., Eley, G. & James, C. D. Diversity and frequency of epidermal growth factor receptor mutations in human glioblastomas. *Cancer Res.* **60**, 1383–1387 (2000).
69. Villanueva, A. *et al.* Pivotal role of mTOR signaling in hepatocellular carcinoma. *Gastroenterology* **135**, 1972–1983 (2008).
70. Villanueva, A., Newell, P., Chiang, D. Y., Friedman, S. L. & Llovet, J. M. Genomics and signaling pathways in hepatocellular carcinoma. *Semin. Liver Dis.* **27**, 55–76 (2007).
71. Paez, J. G. *et al.* EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science* **304**, 1497–1500 (2004).
72. Salvi, A. *et al. In vitro* c-met inhibition by antisense RNA and plasmid-based RNAi down-modulates migration and invasion of hepatocellular carcinoma cells. *Int. J. Oncol.* **31**, 451–460 (2007).
73. Takami, T. *et al.* Loss of hepatocyte growth factor/c-Met signaling pathway accelerates early stages of *N*-nitrosodiethylamine induced hepatocarcinogenesis. *Cancer Res.* **67**, 9844–9851 (2007).
74. Nakayama, K. *et al.* Homozygous deletion of MKK4 in ovarian serous carcinoma. *Cancer Biol. Ther.* **5**, 630–634 (2006).
75. Su, G. H., Song, J. J., Repasky, E. A., Schutte, M. & Kern, S. E. Mutation rate of MAP2K4/MKK4 in breast carcinoma. *Hum. Mutat.* **19**, 81 (2002).
76. Xin, W. *et al.* MAP2K4/MKK4 expression in pancreatic cancer: genetic validation of immunohistochemistry and relationship to disease course. *Clin. Cancer Res.* **10**, 8516–8520 (2004).
77. Arsura, M. *et al.* Transient activation of NF-κB through a TAK1/IKK kinase pathway by TGF-β1 inhibits AP-1/SMAD signaling and apoptosis: implications in liver tumor formation. *Oncogene* **22**, 412–425 (2003).
78. Anzola, M. Hepatocellular carcinoma: role of hepatitis B and hepatitis C viruses proteins in hepatocarcinogenesis. *J. Viral Hepat.* **11**, 383–393 (2004).
79. Beasley, R. P. Hepatitis B virus. The major etiology of hepatocellular carcinoma. *Cancer* **61**, 1942–1956 (1988).
80. Perz, J. F., Armstrong, G. L., Farrington, L. A., Hutin, Y. J. & Bell, B. P. The contributions of hepatitis B virus and hepatitis C virus infections to cirrhosis and primary liver cancer worldwide. *J. Hepatol.* **45**, 529–538 (2006).
81. Yang, H. I. *et al.* Hepatitis B e antigen and the risk of hepatocellular carcinoma. *N. Engl. J. Med.* **347**, 168–174 (2002).
82. Mizuuchi, K. *In vitro* transposition of bacteriophage Mu: a biochemical approach to a novel replication reaction. *Cell* **35**, 785–794 (1983).
83. de Wit, T. *et al.* Tagged mutagenesis by efficient Minos-based germ line transposition. *Mol. Cell. Biol.* **30**, 68–77 (2010).
84. Ding, S. *et al.* Efficient transposition of the piggyBac (PB) transposon in mammalian cells and mice. *Cell* **122**, 473–483 (2005).
This paper shows that PB is a highly active transposon in mammalian cells and the mouse germ line and an excellent alternative transposition system to SB for insertional mutagenesis screens in mice.
85. Keng, V. W. *et al.* Efficient transposition of Tol2 in the mouse germline. *Genetics* **183**, 1565–1573 (2009).
86. Wang, W. *et al.* Chromosomal transposition of PiggyBac in mouse embryonic stem cells. *Proc. Natl Acad. Sci. USA* **105**, 9290–9295 (2008).
87. Smith, G. E., Summers, M. D. & Fraser, M. J. Production of human β interferon in insect cells infected with a baculovirus expression vector. *Mol. Cell. Biol.* **3**, 2156–2165 (1983).
88. Fraser, M. J., Cary, L., Boonvisudhi, K. & Wang, H. G. Assay for movement of Lepidopteran transposon IFP2 in insect cells using a baculovirus genome as a target DNA. *Virology* **211**, 397–407 (1995).
89. Fraser, M. J., Ciszczon, T., Elick, T. & Bauser, C. Precise excision of TTAA-specific lepidopteran transposons piggyBac (IFP2) and tagalong (TFP3) from the baculovirus genome in cell lines from two species of Lepidoptera. *Insect Mol. Biol.* **5**, 141–151 (1996).
90. Cadinanos, J. & Bradley, A. Generation of an inducible and optimized piggyBac transposon system. *Nucleic Acids Res.* **35**, e87 (2007).

91. Mates, L. *et al.* Molecular evolution of a novel hyperactive Sleeping Beauty transposase enables robust stable gene transfer in vertebrates. *Nature Genet.* **41**, 753–761 (2009).
92. Grabundzija, I. *et al.* Comparative analysis of transposable element vector systems in human cells. *Mol. Ther.* **18**, 1200–1209 (2010).
93. Chambraud, B., Berry, M., Redeuilh, G., Chambon, P. & Baulieu, E. E. Several regions of human estrogen receptor are involved in the formation of receptor-heat shock protein 90 complexes. *J. Biol. Chem.* **265**, 20686–20691 (1990).
94. Feil, R., Wagner, J., Metzger, D. & Chambon, P. Regulation of Cre recombinase activity by mutated estrogen receptor ligand-binding domains. *Biochem. Biophys. Res. Commun.* **237**, 752–757 (1997).
95. Mitra, R., Fain-Thornton, J. & Craig, N. L. piggyBac can bypass DNA synthesis during cut and paste transposition. *EMBO J.* **27**, 1097–1109 (2008).
96. Bouwman, P. *et al.* 53BP1 loss rescues BRCA1 deficiency and is associated with triple-negative and BRCA-mutated breast cancers. *Nature Struct. Mol. Biol.* **17**, 688–695 (2010).
97. Lauchle, J. O. *et al.* Response and resistance to MEK inhibition in leukaemias initiated by hyperactive Ras. *Nature* **461**, 411–414 (2009).
98. de Ridder, J., Uren, A., Kool, J., Reinders, M. & Wessels, L. Detecting statistically significant common insertion sites in retroviral insertional mutagenesis screens. *PLoS Comput. Biol.* **2**, e166 (2006).
99. Uren, A. G. *et al.* Large-scale mutagenesis in p19(ARF)- and p53-deficient mice identifies cancer genes and their collaborative networks. *Cell* **133**, 727–741 (2008).
This paper shows how large-scale screens in mice carrying sensitizing cancer causing mutations can be used to identify genes that are involved in collaborative cancer networks.
100. Yang, Y. H. *et al.* Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res.* **30**, e15 (2002).
101. Devon, R. S., Porteous, D. J. & Brookes, A. J. Splinkerettes-improved vectorettes for greater efficiency in PCR walking. *Nucleic Acids Res.* **23**, 1644–1645 (1995).
102. Gavin, A. J. *et al.* Pooled library tissue tags for EST-based gene discovery. *Bioinformatics* **18**, 1162–1166 (2002).

Acknowledgements
The authors would like to thank S. C. Lee for help with the figures and members of their laboratory for helpful comments on the Review. Support was provided by the Biomedical Research Council (BMRC), Agency for Science and Technology and Research (A*STAR), Singapore.

Competing interests statement
The authors declare no competing financial interests.

FURTHER INFORMATION
Nancy A. Jenkins's homepage: http://www.imcb.a-star.edu.sg/php/nn.php
CancerHelp UK: http://www.cancerhelp.org.uk
COSMIC database: http://www.sanger.ac.uk/genetics/CGP/cosmic
International Cancer Genome Consortium: http://icgc.org/

ALL LINKS ARE ACTIVE IN THE ONLINE PDF