

## METHOD

## Open Access

# Polygenic *in vivo* validation of cancer mutations using transposons

Su Kit Chew<sup>1,2</sup>, Dong Lu<sup>1,3</sup>, Lia S Campos<sup>1</sup>, Kenneth L Scott<sup>4</sup>, Abdel Saci<sup>5</sup>, Juexuan Wang<sup>1</sup>, Adam Collinson<sup>1,3</sup>, Keiran Raine<sup>1</sup>, Jonathan Hinton<sup>1</sup>, Jon W Teague<sup>1</sup>, David Jones<sup>1</sup>, Andrew Menzies<sup>1</sup>, Adam P Butler<sup>1</sup>, John Gamble<sup>1</sup>, Sarah O'Meara<sup>1</sup>, Stuart McLaren<sup>1</sup>, Lynda Chin<sup>6</sup>, Pentao Liu<sup>1\*</sup> and P Andrew Futreal<sup>1,6\*</sup>

## Abstract

The *in vivo* validation of cancer mutations and genes identified in cancer genomics is resource-intensive because of the low throughput of animal experiments. We describe a mouse model that allows multiple cancer mutations to be validated in each animal line. Animal lines are generated with multiple candidate cancer mutations using transposons. The candidate cancer genes are tagged and randomly expressed in somatic cells, allowing easy identification of the cancer genes involved in the generated tumours. This system presents a useful, generalised and efficient means for animal validation of cancer genes.

## Background

The sequencing of cancer genomes has become a globally coordinated effort to understand the genomic, epigenetic and gene expression changes that occur in cancers [1]. By coupling the output of the catalogues of mutations and alterations in cancer genomes with systematic functional analyses, an overarching aim is to derive both an understanding of the underlying pathophysiological process and improved clinical outcomes [2-4]. The functional validation and characterisation of candidate cancer mutations presents a practical challenge in cancer genomics, due to the diversity of assays and in particular the costs and duration of *in vivo* experiments.

DNA transposons are mobile genetic elements that translocate within the genome via a 'cut-and-paste' mechanism. They are versatile genetic tools used in a variety of genetic models for purposes ranging from transgene delivery to mutagenesis and chromosome engineering [5,6]. The *Sleeping Beauty* (*SB*) transposon is active in mammalian cells and has been used for cancer gene discovery in a variety of tissue contexts by insertional mutagenesis [7-10]. The *piggyBac* (*PB*) transposon is a DNA transposon recently described to be active in mouse ES cells with much higher transposition efficiency than

*Sleeping Beauty* [11,12]. We have previously shown that enhancer/transcript-trapping DNA transposons targeted into the genome can be used to assay the *in vivo* oncogenic potential of the payload cDNA contained in the transposon [13]. Tumourigenesis in this system requires the confluence of appropriate genetic, temporal and microenvironment contexts: first, the transposon must carry an oncogenic cDNA; second it traps a transcript that expresses the cDNA at an appropriate level; third, this occurs in a susceptible tissue and at a proper developmental stage; and finally, the cDNA confers a selective advantage so that expression is maintained by positive selection. A bacterial artificial chromosome vector constructed to carry an array of transposons with different cDNAs was targeted to the *Hprt* locus using recombination mediated cassette exchange in ES cells. When animals generated from these ES cells were crossed to a strain ubiquitously expressing transposase, the progeny developed a broad spectrum of tumours that expressed the oncogenic cDNAs but not the control cDNAs from the transposon array. The transposons stochastically mobilise in somatic cells, expressing their payload cDNA when they insert near enhancers or expressed genes and trap the expression activity. Due to the ubiquitous expression of transposase, transposons can also remobilise and lose payload cDNA expression. Therefore, the expression of an individual or group of transposon cDNA(s) clonally maintained within an expanding tumour over time would

\* Correspondence: pl2@sanger.ac.uk; afutreal@mdanderson.org

<sup>1</sup>Wellcome Trust Sanger Institute, Hinxton, UK

<sup>6</sup>University of Texas MD Anderson Cancer Center, 1515 Holcombe Blvd, Houston, TX 77030, USA

Full list of author information is available at the end of the article

suggest positive clonal selection for those cDNA(s). This approach effectively assays the oncogenic potential of each cDNA payload in multiple tissue contexts. As transposition is somatic and cell-autonomous, each somatic cell would have a unique transposon insertion profile and a single animal line with a library of transposons could develop tumours driven by different combinations of oncogenic cDNAs.

To practically use transposon-mediated *in vivo* tumorigenesis to functionally validate cancer mutations, we required an efficient method for transfecting transposons into the murine genome and a strategy for tracking the transposons. To address those needs, we describe here a nested PB-SB transposon vector that allows pools of multiple transposons to efficiently transduce into the genome of mouse embryonic stem (ES) cells [12]. Individual transposons constructs are tagged with unique 3' UTR sequences to track the exogenous cDNAs and facilitate genotyping. We used this system to assay the *in vivo* oncogenic potential of a set of kinase mutations we have previously observed in human cancers [14], recovering tumours that recurrently expressed a subset of the cDNAs. The tumours show selective retention of certain cDNAs during serial transplantations and generation cell lines. Exome sequencing of the tumours revealed intratumour heterogeneity and evolutionary life histories similar to that observed in human cancers. In two examples of mutations in *Cyclin-dependent kinase 2* (*CDK2*) and *Diacylglycerol kinase, beta* (*DGKB*) genes, we show how the *in vivo* assay reveals novel insights on the functions of these genes in cancer. This system uses standard molecular biology and transgenic protocols, yielding a general mouse model for validating cancer mutations in a wild-type background that is polygenic in throughput. This *in vivo* validation system complements and informs *in vitro* cell-based assays in functional cancer genomics.

## Results and discussion

### Design and generation of transposon constructs

To transfect pools of transposons into the genome of cells without constructing or manipulating large arrays of transposons in bacterial artificial chromosome constructs that were required previously [13], we rationalised that we could take advantage of the high transposition efficiency of *piggyBac* that is two to three orders of magnitude higher than the *Sleeping Beauty* transposon in ES cells [12]. By nesting the rest of the transposon vector design within *PB* terminal repeats (Figure 1a), we can efficiently introduce pools of cargo payload within the *PB* repeats into the genome of cells by co-electroporating transposon constructs with a helper plasmid encoding *PB* transposase. The nested *PB* and *SB* transposon design (Figure 1a) permits us to use either *PB* or *SB* transposase to mobilise the transposon *in vivo*. As our intention was

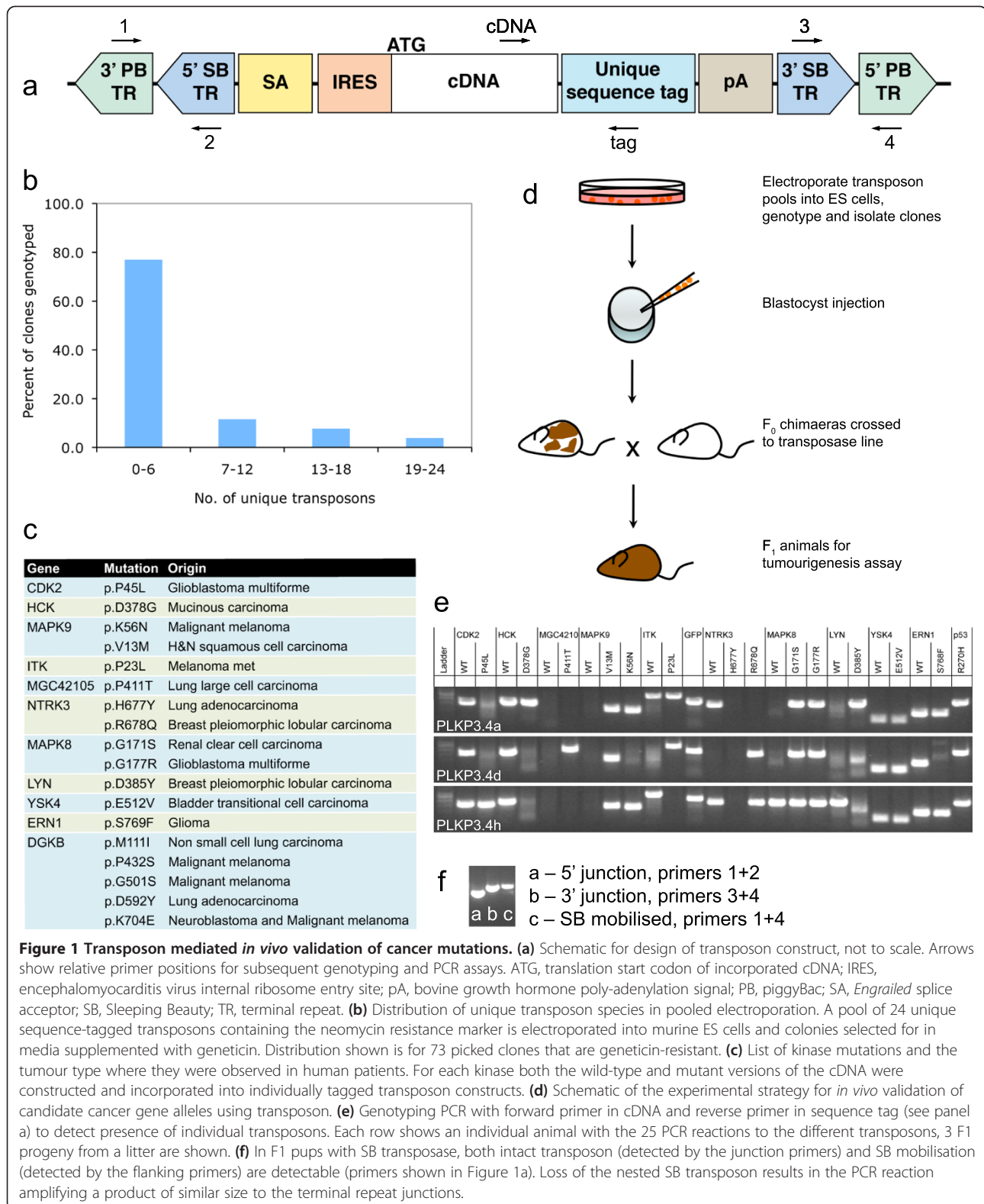
to assay the oncogenic potential of the cDNA payload itself and not the oncogenic potential from disruption of an endogenous loci or novel fusion transcripts, we introduced an internal ribosomal entry sequence (IRES) after the splice acceptor sequence so that the cDNA payload is expressed alone. We incorporated a 60-basepair unique sequence tag between the stop codon and poly-A signal (Figure 1a and see Additional file 1: Table S1). This tag allows us to specifically detect the presence of the transposon in the genome and expression of the cDNA transcript by PCR, and eliminates possible non-specific amplification from the endogenous gene.

To evaluate the efficiency of the pooled electroporation strategy, we combined a pool of 24 unique sequence-tagged *PB* transposons constructs with the transposase helper plasmid for electroporation into mouse ES cells. Co-electroporation led to a genomic distribution of unique transposon insertions where approximately 20% of 73 genotyped Neomycin-resistant clones have greater than six unique transposons using PCR genotyping of the sequence tags (Figure 1b). This frequency distribution demonstrates the utility of pooled transposon electroporation and suggests that the probability of each transposon integrating is independent of the other constructs.

For the candidate cDNA payloads, we curated a list of 18 kinase mutations from a variety of cancers [14] (Figure 1c), chosen for their likelihood to function as dominant gain-of-function point mutations based either on the location of the mutation in the functional domains or the distribution of mutations within the gene. In addition to the mutant alleles, we also constructed transposons with the 11 wild-type cDNA sequences to compare the effects of ectopic expression of the kinases.

### Generation of animals for tumorigenesis assay

The schematic illustrating animal generation for *in vivo* tumorigenesis is outlined in Figure 1d. We electroporated pools of six to 25 transposon constructs together with the *piggyBac* transposase helper plasmid into AB2.2 and JM8A3 mouse ES cell lines [15,16]. Different representations of transposons were identified in the electroporated ES cell clones by PCR genotyping using primers specific to each cDNA and its corresponding unique sequence tag (Figure 1a and Additional file 1: Table S1). We used ES cell clones that each carried four to 21 unique transposons to generate chimaera animals by blastocyst microinjection (Additional file 1: Table S2). The pools were combinations of mutations in different genes, and a pool of different mutations in the same gene (*DGKB*). Potential passenger effects due to insertional mutagenesis by the transposon are avoided by having multiple animal lines generated from different ES cell clones with different pools and initial insertion profiles, so that



recurrent expression of oncogenic cDNA transposons recovered from different tumours would have originated from cells with distinct initial transposon integration sites

(Additional file 1: Table S2). In parallel, we electroporated transposon pools into ES cell lines already expressing either the *PB* or *SB* transposase, the chimaeras generated

from these ES cell clones have mobilised transposons and can be used directly for the tumourigenesis assay in the F0 generation chimaeras. To provide a sensitised background for tumourigenesis, we also co-electroporated a subset of ES cells with a transposon constitutively expressing the dominant-negative R270H allele of the *Trp53* tumour suppressor gene [17] under the control of the CAGG (cytomegalovirus early enhancer and chicken beta-actin promoter hybrid) synthetic promoter.

Chimaera animals with pools of transposons in their genome were subsequently crossed to a mouse line expressing the SB11 transposase from the ubiquitous *Rosa26* promoter [10]. In the F1 progeny of this cross, we are able to detect individual transposons in genomic DNA using primers designed to amplify the cDNA and its unique sequence tag (Figure 1e). While Mendelian segregation of the different transposon insertions in the F1 progeny is expected given the random integration of transposons into the genome, we observed that individuals in the cohort of three littermates shown have similar representations of the transposon pool, this is likely due to multiple copies of each individual transposon being present in the genome at different loci. We are able to verify that the nested transposon is intact, and where *SB11* transposase was used to mobilise the transposon, we can also verify the excision of the *SB* transposon from the flanking *PB* terminal repeats by PCR (Figure 1f).

#### Validating cancer mutations by tumourigenesis *in vivo*

For *in vivo* tumourigenesis, we aged the mice with daily health monitoring for malignancy. As detailed in Table 1, chimaeric mice with both transposons and transposase expression presented histopathology-verified tumours at a higher rate than control mice that had transposons but no constitutive transposase expression (details of individual animals in Additional file 1: Table S2). In F1 progeny animals, the presence of a transposon expressing the dominant negative allele of p53 reduced the median tumour latency by 33.5 weeks; interestingly, these had lower incidence of tumours. This might be due to a lower general fitness in animals with this transposon, resulting in more animals being culled for non-malignancies. Given that transposon mobilisation is a cell independent somatic event and individual transposons are also represented in different pool combination and ES cell clones, we scored the occurrence of each candidate cDNA expression in

different tumour types as a way to assess the potential of a given cDNA in contributing to tumourigenesis. We detected cDNA expression from the transposons in approximately half of the tumours assayed. In tumours without detectable cDNA expression, we are not able to distinguish between tumours that arose because of a transient effect of an oncogene (for example, the expression of a mutant kinase cDNA from a transposon might have contributed to the establishment of a clone but is then lost subsequently), versus tumours that arose due to other mutations occurring spontaneously. While the potential for remobilisation is an inevitable effect of the constitutive transposase activity, we reasoned that we should avoid any *a priori* assumptions on the suitable developmental stage or timing for mobilisation of the transposons. Also, the ability to incorporate *in vivo* selection pressure to retain transposon expression would provide a more stringent assay for validating the candidate cDNAs, while decreasing the potential passenger transposon insertions that might occur after a transient transposase mobilisation. Table 2 shows the occurrence for expression of each transposon in each tumour type.

Each occurrence of a cDNA's expression in a tumour represents a potential context in which its expression contributed to tumourigenesis. In tumours, we often see loss of transposons and transposon cDNA expression compared to normal untransformed tissues. This bias can arise from clonal competition in dividing tumour cells, as transformed cells are likely to be represented by a few successful subclones. In contrast, the mixture of clones within normal tissues with different transposon insertion and expression profiles results in a net representation of most transposons from the germline. As an example, in an F<sub>0</sub> *PB* transposase expressing chimaera (animal id PLKD4.1a) with hepatocellular carcinoma (Figure 2a and b), the expression pattern in normal tissues suggests that stochastic and ectopic expression of many kinase alleles can be tolerated in normal tissues. In the tumour, only a subset of kinase transposon and cDNA expression is positively selected for and retained (Figure 2b).

To examine whether tumours retain transposon payload expression and have serial engraftment capacity, we transplanted fragments of 23 primary tumours by subcutaneous injection into immune-compromised NOD.Cg-*Prkdc<sup>scid</sup> Il2rg<sup>tm1Wjl/Sz</sup>* (NSG) mice [18]. We observed

**Table 1 Summary of *in vivo* tumourigenesis assay**

Generation	Transposase	Transposons	Animals (n)	Animals with tumours (n)	Median tumour latency (weeks)	Animals with tumours (%)
F0 chimaeras	+	+	21	8	85.5	38
F0 chimaeras	-	+	20	3	89.9	15
F1 animals	+	+	65	44	108.4	67
F1 animals	+	+(p53.R270H)	20	6	74.9	30

**Table 2 Occurrence of transposon cDNA expression in different tumour types**

cDNA	Lymphoma	Carcinoma	Blastoma	Sarcoma	Total
CDK2.P45L	11	5	0	0	16
ERN1.S768F	8	4	0	0	12
LYN	5	6	0	0	22
ERN1	6	4	0	0	10
ITK	6	3	0	0	9
MAPK8.G171S	5	4	0	0	9
NTRK3	3	6	0	0	9
HCK.D378G	4	4	0	0	8
MAPK8	4	4	0	0	8
CDK2	4	3	0	0	7
MAPK8.G177R	3	4	0	0	7
MGC42105.P411T	5	2	0	0	7
DGKB	4	1	1	0	6
HCK	2	4	0	0	6
NTRK3.H677Y	4	2	0	0	6
YSK4	4	2	0	0	6
YSK4.E512V	3	2	0	0	5
MAPK9	3	1	0	0	4
MGC42105	3	1	0	0	4
ITK.P23L	2	1	0	0	0
NTRK3.R678Q	1	2	0	1	1
DGKB.D592Y	0	0	0	1	1
DGKB.G501S	0	0	0	1	1
DGKB.P432S	0	0	0	1	1
MAPK9.K56N	0	1	0	0	1
DGKB.K704E	0	0	0	0	0
DGKB.M111I	0	0	0	0	0
LYN.D385Y	0	0	0	0	0
MAPK9.V13M	0	0	0	0	0

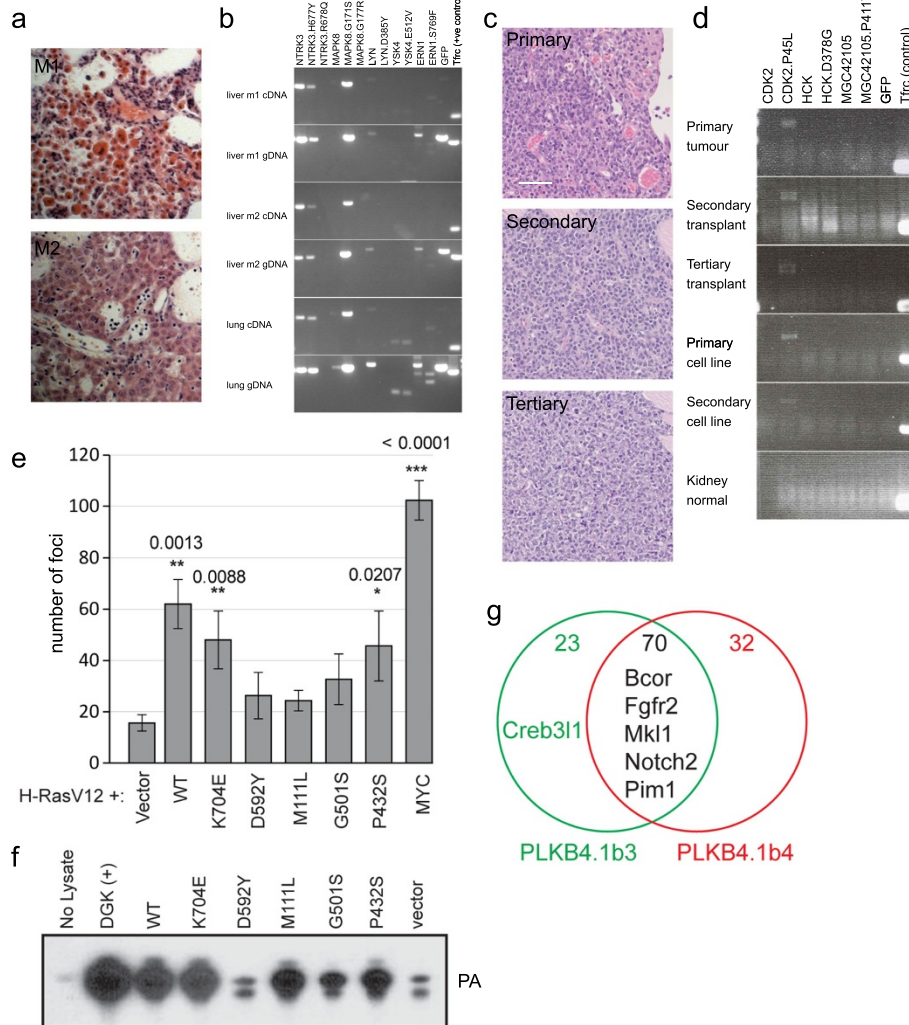
that five (22%) of the tumours successfully engrafted in NSG mice. Of the five engrafted tumours, we attempted serial transplantation and four engrafted successfully again in NSG mice. In addition, efforts to create tumour cell lines from 34 primary tumours led to 13 tumour cell lines (38% success rate). As an example, from an F<sub>1</sub> animal (id PLKH1.4b) with lymphoma, the tumour engrafted in NSG mice serially and maintained consistent histopathological features (Figure 2c). This animal had no transposon cDNA expression in sampled normal tissues and only CDK2.P45L expression in the lymphoma, the expression of the transposon cDNA was maintained through the serial transplants and also in cell lines derived from the tumour and the secondary engraftment (Figure 2d). Of the four tumours that underwent two serial engraftments, three

showed retention of the transposon cDNA expression observed in the primary tumours.

The transposon construct with expression detected most frequently in tumours was the mutant allele of *Cyclin-dependent kinase 2* (CDK2.P45L, Table 1). The CDK2.P45L-expressing tumours were from animals generated by six independent ES cell clones with different pools of transposon insertions, pointing to the recurrence of CDK2.P45L expression as a result of positive selection. The CDK2.P45L mutation is a point substitution we observed once in a single glioblastoma patient where the conserved proline residue in the PSTAIRE interaction motif is replaced with leucine [14]. The PSTAIRE motif is the central helical motif in the interaction of CDK2 with its regulatory partner CYCLIN E1. Biochemical assays suggest that CDK2.P45L is unlikely to drive oncogenesis through an activating gain of function or deregulation mechanism, as the mutation disrupts the interaction between CDK2 and its regulatory binding partner CYCLIN E1 even though its kinase activity is retained [19]. Nevertheless, the recurrence and selective retention of expression in tumours (Figure 2a-d) observed here suggests that it can contribute a positive selective advantage to tumours *in vivo*.

In a separate pool, six transposons with alleles of *DGKB* (5 mutants, 1 wild-type) were introduced into the genomes of the ES cells, these mutations were observed in lung and melanoma cancer genomes (Figure 1c). The wild-type, K704E and P432S alleles have varying activity in transforming Ink/Arf-null mouse embryonic fibroblast (MEF) cells when expressed together with HRAS.V12 (Figure 2e). This transformation activity does not correlate with kinase activity of the protein, as the M111L allele retains kinase activity but does not cooperate with HRAS.V12 in transformation (Figure 2f). Remarkably, only the expression of the wild-type *DGKB* cDNA was recurrently recovered in multiple tumours (Table 1). While both substrate and product of *DGKB* (diacylglycerol and phosphatidic acid, respectively) play diverse roles in intracellular signalling [20], our data suggest that overexpressed *DGKB* can have oncogenic activity that is unrelated to its kinase activity, possibly through a scaffold or complex recruitment function given that the substitution mutants do not appear to contribute to tumorigenesis *in vivo* in this experimental context.

Bi-functional transposons designed with strong promoter/enhancer elements to drive ectopic gene expression and splice-acceptor-polyA signals to disrupt expression have been successfully used as insertional mutagens to identify cancer genes [7-10]. Unlike the insertional mutagenesis transposons used in those screens, the design of our cDNA delivery transposon does not contain any promoter/enhancer elements. While there is a possibility



**Figure 2 Tumours generated in animals with transposons.** (a) An F<sub>0</sub> chimaera (animal id PLKD4.1a) that presented with hepatocellular carcinoma in two liver lobes (m1 and m2), histological sections from both tumours visualised with haematoxylin and eosin (H&E) stains. (b) Transposon expression (RT-PCR using cDNA) and representation (PCR using genomic DNA (gDNA)) in two tumour samples and normal lung. (c) An F<sub>1</sub> animal (id PLKH1.4b) that presented with a solid lymphoma tumour (top panel), H&E histological section. Serial subcutaneous transplants of the tumour in NSG mice gave rise to secondary (middle panel) and tertiary tumours (lower panel). (d) Transposon expression of PLKH1.4b in the serially transplanted tumour, cell lines derived from the primary and secondary transplanted tumours, and normal kidney tissues. (e) Colony forming assay of different DGKB alleles in cooperation with HRAS.V12 using Ink/Arf mutant MEFs. MYC is a positive control. Error bars denote standard deviation, *P* value from two-tailed T-test compared to HRAS.V12 alone. (f) Relative activity of different DGKB alleles as measured by phosphatidic acid (PA) production. (g) Number of mutated genes unique to and shared between two regions of a single lymphoma that was exome-sequenced. Known cancer genes with mutations are shown.

that splice acceptor's promoter/enhancer-trapping activity can disrupt the expression of an endogenous tumour suppressor gene locus in a particular tumour sample, possibly resulting in hypomorphic gene function, the complete disruption of any particular gene function in a cell requires either a disrupting insertion followed by spontaneous loss of heterozygosity at the other copy, or two independent disrupting insertions at both copies of the gene in the chromosome pair. The basis of validating a candidate cDNA payload is based on observing recurrent expression across tumours from different animals, different founder

animals derived from multiple ES cell clones, and with each tumour having a unique somatic profile of transposon insertions. Nevertheless, to assess whether the transposon-mediated delivery of the cDNAs contributes to a potential mutational load and to understand the co-operating mutations that could act together with the expressed exogenous kinases to drive tumourigenesis, we performed paired-end exome sequencing on tumour and normal matched samples from two animals that presented with lymphoma (animal ids PLKB4.1b and PLKE5.1a). Variants were called using customised CaVEMan

and Pindel with output filtering (see Materials and methods). To validate the variants, we used Sanger sequencing for indels and pyrosequencing for substitutions. We validated 125 somatic mutations in PLKB4.1b and 28 somatic mutations in PLKE5.1a (Additional file 1: Tables S3, S4). Both animals were F<sub>0</sub> chimaeras generated from transposase expressing ES cells, *PB* transposase in PLKB4.1b and *SB* transposase in PLKE5.1a. As *PB* transposition has been described to be largely precise and error-free in multiple species with no defined molecular signature [6,12,21,22], we are unable to quantify the relative contribution of transposon activity in the generation of the additional mutations in PLKB4.1b. However, given that *SB* excision results in a characteristic insertion of a short motif comprising the end of the transposon terminal repeat and a duplication of the TA dinucleotide insertion site [23,24], it is possible to estimate if transposition represents a significant mutation load. From the PLKE5.1a exome, none of the validated mutations exhibit the molecular footprint arising from *SB* transposition, suggesting that other additional drivers of mutations in tumours do occur and play the major role in the generation of these additional mutations. As the PLKB4.1b tumour was macroscopically heterogeneous, we sampled two different regions of the tumour for exome sequencing (sample ids PLKB4.1b3 and PLKB4.1b4). The two regions had 70 common mutations (75% and 68% of each region's mutations respectively) that included five of the six known cancer genes mutated in both regions (Figure 2g, [25]). Both regions express CDK2.P45L cDNA. The regional exome sequencing shows that while a core set of mutations is likely to have contributed to the tumour, intra-tumour heterogeneity and genotypic divergence occurred in the life history of the tumour.

The *CDK2* and *DGKB* mutations highlight as examples that *in vivo* validation can provide novel information on the oncogenic activity of mutant alleles that would not be gleaned from biochemical or cell biology assays alone. In the case of CDK2.P45L, even though the mutant is unable to form a stable association with its cognate activating cyclin [19], the *in vivo* data showing positive selection for the expression of the mutant cDNA suggest that there could potentially be additional assembly factors that facilitate transient association between CDK2 and CYCLIN E1. For the *DGKB* mutant alleles, the uncoupling of kinase activity, transformation activity and *in vivo* representation in tumours underscores the critical importance for the multiplicity and depth of validation assays in functional cancer genomics. Without complementary functional characterization with *in vitro* and cell-based assays for the other mutations listed in Figure 1c, we are intentionally cautious in refraining from asserting whether the other mutations are validated or non-functional passenger mutations in the cancer genome. Even so, we note that our

data presented here can serve to direct future investigations. Expression of both the wild-type and S768F mutant allele of *endoplasmic reticulum to nucleus signalling 1* (*ERN1*, also *IRE1*) are the second and fourth most frequently recovered transposon cDNAs in our study (Table 2), corroborating recent description of *ERN1* as playing key roles in tumour angiogenesis, growth and invasion [26,27]. In contrast, for the *v-yes-1 Yamaguchi sarcoma viral related oncogene homologue* (*LYN*) kinase that is implicated in a wide variety of cancers [28-30], while expression of the wild-type kinase is the third most frequent occurrence, we never recovered any tumours expressing the D385Y mutant (Table 2). This would suggest that the D385Y mutation is potentially a loss-of-function mutation, though this has to be characterised molecularly with biochemical and cell-based activity assays.

In this study, lymphomas and carcinomas were the predominant types of tumours generated. While the CDK2.P45L mutation was originally discovered in glioblastoma multiforme, its corresponding cDNA expression was detected in 11 lymphomas and five carcinomas in the *in vivo* tumourigenesis assay. This probably reflects a susceptibility of tissues to transformation and the underlying propensities of the promoter-transposase and transposon species combinations used here. A similar predisposition towards haematopoietic malignancies was observed in the first whole-body insertional mutagenesis screens with the *SB* transposon [7,8], and the predisposition can be varied by using a different transposon such as *PB* or by expressing the transposase from different promoters [31]. Indeed, from our relatively small cohort of F<sub>0</sub> chimaeras derived from ES cells expressing transposase, a single *SB100* transposase-expressing animal presented with lymphoma. In comparison, six animals derived from ES cell clones expressing the *PB* transposase presented with lymphomas, hepatocellular carcinomas and a Wilms-like blastoma. Future iterations of this validation system could incorporate tissue-specific promoters or conditional 'Lox-stop-Lox' cassettes in the transposon to facilitate granular control for validating candidate cancer mutations in a tissue, developmental and temporal-specific manner.

With a larger animal cohort and tumour numbers, it also would be informative to explore in further studies whether co-expression of multiple transposon cDNA payloads could be indicative of synergistic genetic interactions in the pool. In this study, we have not taken into account the presence of any other transposon cDNA expression when looking at the recurrence of individual candidate cDNA payload, including the expression of the dominant-negative *Trp53* in some of our pools. We reasoned that it should be the selective retention of cDNA expression that is an indication of a candidate's contribution to clonal selection and oncogenic potential.

In addition to the selection forces we observed acting on the exogenous kinase cDNAs, the exome sequencing of the tumours show that these bona fide tumours can have additional mutations that are either prevalent or divergent within the tumour, exhibiting intratumour genotypic diversity and heterogeneity similar to that observed in the clonal evolutionary history of human tumours [32,33].

## Conclusions

We have shown here an efficient *in vivo* system for validating cancer genes that uses standard molecular biology and transgenic techniques. Several systems for elegant forward and reverse genetic screens have been engineered in mice that look at genes in specific tumour tissue types (that is, liver or brain) or in specific cancer processes such as metastasis [34-38]. While those systems allow for specific studies in particular tumour types and biological processes, the system described here allows for generalised whole-body *in vivo* validation of gain of function or neomorphic alleles that are observed in human cancer genomes. Given the crucial importance of balancing the requirements of *in vivo* validation with the resource-intensive nature of animal experiments, the system described here is advantageous as a generalised *in vivo* validation step in functional cancer genomics. The identification or validation of genes *in vivo* should be complemented by molecular and cellular characterisation. With our collection of kinase mutations that have not been previously characterised *in vivo*, we show here for the P45L mutation in *CDK2* and the *DGKB* alleles that *in vivo* contexts can reveal phenotypic divergences and nuances when compared to standard cell-based or biochemical assays. Compared to our previous experimental system that required the construction of large bacterial artificial chromosomes for the transposon gene arrays [13], the procedures for generating these animals are relatively simple, providing an easy means of polygenic validation in multiple somatic tissue types per animal line that is not previously accessible. This transposon-based *in vivo* system allows us to isolate somatic tumours where the oncogene candidates are expressed in a susceptible tissue microenvironment and at levels that are clonally advantageous, complementing existing *in vitro* and cell-based assays for validating and characterising genes. By using a system where the candidate mutation has to be selectively retained, this system also provides a means of generating cell lines useful for further characterising functional requirements and roles of the mutations. This is particularly advantageous as cancer genes and mutation candidates from large scale sequencing studies are often not preserved within a functional context in patient-derived cell lines. While the overall throughput of validation described here averages a little less than a dozen candidate cDNAs per animal line,

the relative simplicity of the techniques described here is scalable with the number of candidates for validation by increasing the number of animal lines. The *in vivo* validation system described here should prove to be a useful addition to the current suite of functional cancer genomics tools.

## Materials and methods

### Transposon constructs

The piggyBac transposase construct and components of the nested PB-SB transposon vector were cloned as previously described [12,13]. Kinase mutations were generated using the QuikChange Lightning Site-Directed Mutagenesis Kit (Stratagene) and cDNA clones from the IMAGE consortium. *NTRK3* cDNA was a gift from B.D. Nelkin [39]. The cDNAs and the 60-basepair unique sequence tag were added by PCR cloning (KOD, Takara) into the transposon vector (See Additional file 1: Table S1 for tag sequences).

### Cell culture

Both AB2.2 and JM8A3 mouse ES cells were grown in standard M15 media on SNL76/7 feeder cells [40]. M15 comprises Knockout DMEM (GIBCO), 15% Fetal Bovine Serum (Invitrogen), 1× Pencillin, Streptomycin and Glutamine (GIBCO), 1× Non-Essential Amino Acids (GIBCO), 0.1 mM  $\beta$ -mercaptoethanol (Sigma), 1,000 U/mL human LIF (Millipore). Tumour derived cell lines were cultured in RPMI with 10% Fetal Bovine Serum (Invitrogen), 1× Pencillin, Streptomycin and Glutamine (GIBCO), 1× Non-Essential Amino Acids (GIBCO).

### Manipulation of mouse ES cells

To electroporate mouse ES cells, cells were harvested by trypsinisation, washed in PBS and resuspended to  $10^7$  cells/mL. For each mL of cells, 30  $\mu$ g total of pooled transposon constructs of equal quantity were co-electroporated with 10  $\mu$ g of CAGG-PBase transposase helper plasmid and 30 ng of *PB-SB-pGK-Neo* co-selection marker on a Biorad GenePulser. The electroporation settings are 230 V, 500  $\mu$ F with expected time constant between 5.6 and 8.0. In all transposon pools, *PB-SB-ires-eGFP* is a negative control in the pool. Selection for clones containing transposon using *PB-SB-pGK-Neo* co-selection was done in M15 media supplemented with 125  $\mu$ g/mL Geneticin (GIBCO).

### Animal work

All animal-related protocols and care was provided in accordance with the Animal (Scientific Procedures) Act 1986. To generate animals, chimaera animals were generated by injecting ES cells into blastocysts as per standard protocols. For the tumorigenesis assay, animals are aged and monitored daily for signs of malignancy. Animals



exhibiting signs of poor health or distress are euthanized by exposure to rising CO<sub>2</sub> concentration, followed by necropsy and histopathological analyses of tissues.

For serial tumour transplant, tumour samples were harvested immediately after the animal was culled. Tumour tissue is briefly surface sterilised in 70% ethanol and immediately washed in PBS, followed by mincing with scalpels. Fine tumour mince is injected subcutaneously into the flanks of NSG mice with a 21-gauge needle and syringe.

### Histology

Tissues were fixed overnight in 10% formalin (Sigma), dehydrated and cleared for paraffin embedment as per standard histology protocols. The paraffin blocks were cut at 5 µm thickness and stained with haematoxylin and eosin as per standard protocols. All culled animals were scored for presence of tumour and tumour type(s) as applicable, independently of the detection of transposon and cDNA expression without randomisation or blinding.

### Detecting transposon and payload cDNA expression

Both genomic DNA and RNA are isolated from cells or tissues using an AllPrep kit (Qiagen) where fresh tissues were available, this was not done in cases such as found dead animals where the tissues were degraded. For reverse transcriptase reaction, cDNA was synthesised from extracted RNA using Superscript III (Invitrogen). PCR was done on standard conditions using Extensor (Thermo), primers pairs are listed in Additional file 1: Table S5. The primer pair for *Transferrin receptor (Tfrc)* positive control spans an intron, allowing detection of any genomic DNA contamination in cDNA PCR and vice versa.

### Cell line derivation

Fresh tumour mince is divided between three 35 mm tissue culture dishes and incubated for 16 h in RPMI media supplemented with 30, 100 and 300 U/mL collagenase II (GIBCO). Cells from all three dishes are collected the next day, centrifuged at 200 rcf for 5 min and washed in RPMI media to remove the collagenase. The resulting cell suspension is seeded into 75 cm<sup>2</sup> flasks, media changed every 2 to 4 days until the cells are approximately 80% confluent and ready for subculturing.

### Kinase assay

DGKB kinase activity was assessed by measuring the abundance of phosphatidic acid (PA) as a reflection of cellular DGK activity in protein lysates, as described previously [41]. Briefly, lysates from 293 cells transfected with the wild-type and indicated DGKB mutants were separated by thin layer chromatography, PA abundance was analysed by autoradiography. Empty vector and

Bacterial DGK were used as negative and positive controls, respectively.

### Transformation assay

*Ink4a/Arf*-deficient primary murine embryonic fibroblasts (MEFs) were plated in DMEM containing 10% FBS at a density of  $8 \times 10^5$  cells per 10 cm, 16 h before transfection. For RAS cooperation, 1.5 µg HRAS(Val 12) vector was co-transfected with 6.5 µg pEF-Dest51-LacZ control vector, MYC or the indicated DGKB variants in pEF-Dest51 using Lipofectamine2000 (Invitrogen) following the manufacturer's instructions. The total amount of transfected DNA was kept constant at 7.5 µg, and transfections were done in duplicate three times. At 48 h after transfection, each transfected 10-cm plate was equally split into three 10-cm plates and incubated for 10 days, during which media was refreshed twice. Cells were washed, fixed in 10% formalin and stained with Giemsa solution (Sigma) for 10 min at room temperature for foci quantification. Two-tailed *t*-test calculations were performed using Prism 4 (Graphpad).

### Exome sequencing

Sequencing was based on exome capture and performed using 76 basepair paired-end reads on the Illumina GAIIX platform. The sequencing data are deposited into the European Nucleotide Archive under study accession number ERP000896. Mapping of the sequence data was done using BWA [42]. Variant calling was done using CaVE-Man (Cancer Variants through Expectation Maximisation) [43] for substitutions and Pindel [44] for insertions and deletions, modifications to the variant callers and subsequent filtering were as previously described [45]. These algorithms identified somatic variants in the tumour samples compared to matched normal sample for the same animal.

### Additional file

**Additional file 1: Table S1.** Tag sequences. **Table S2.** Details of tumour bearing mice. **Table S3.** Validated mutations from animal PLKB4.1b. **Table S4.** Validated mutations from animal PLKE5.1a. **Table S5.** Genotyping primers.

### Competing interests

The authors declare that they have no competing interests.

### Authors' contributions

SKC, DL, LSC, JW and AC designed, performed and analysed results of the transposon and animal experiments. KLS, AS and LC designed, performed and analysed results of the *in vitro* characterisation experiments. SKC, KR, JH, JWT, DJ, AM, APB and JG analysed the sequencing data with SO and SM providing sample processing and technical support. SKC, PL and PAF wrote the manuscript with discussion and input from all authors. PL and PAF directed the research. All authors read and approved the final manuscript.

## Acknowledgements

We thank the Wellcome Trust Sanger Institute's Research Support Facility, Mouse Genome Project, Histology and Sequencing teams for their technical support; Dr BD Nelkin (Johns Hopkins Medicine) for the NTRK3 cDNA plasmid. This work was supported by the Wellcome Trust.

## Author details

<sup>1</sup>Wellcome Trust Sanger Institute, Hinxton, UK. <sup>2</sup>Current address: UCL Cancer Institute, London, UK. <sup>3</sup>Current address: Babraham Institute, Cambridge, UK. <sup>4</sup>Baylor College of Medicine, Houston, TX, USA. <sup>5</sup>Novartis Institutes for Biomedical Research, Cambridge, MA, USA. <sup>6</sup>University of Texas MD Anderson Cancer Center, 1515 Holcombe Blvd, Houston, TX 77030, USA.

Received: 10 March 2014 Accepted: 27 August 2014

Published online: 27 September 2014

## References

- Hudson TJ, Anderson W, Artez A, Barker AD, Bell C, Bernabe RR, Bhan MK, Calvo F, Eerola I, Gerhard DS, Guttmacher A, Guyer M, Hemsley FM, Jennings JL, Kerr D, Klatt P, Kolar P, Kusada J, Lane DP, Laplace F, Youyong L, Nettekoven G, Ozenberger B, Peterson J, Rao TS, Remacle J, Schafer AJ, Shibata T, Stratton MR, Vockley JG, *et al*: **International network of cancer genome projects.** *Nature* 2010, **464**:993–998.
- Boehm JS, Hahn WC: **Towards systematic functional characterization of cancer genomes.** *Nat Rev Genet* 2011, **12**:487–498.
- Chin L, Gray JW: **Translating insights from the cancer genome into clinical practice.** *Nature* 2008, **452**:553–563.
- Stratton MR, Campbell PJ, Futreal PA: **The cancer genome.** *Nature* 2009, **458**:719–724.
- Candela H, Hake S: **The art and design of genetic screens: maize.** *Nat Rev Genet* 2008, **9**:192–203.
- Thibault ST, Singer MA, Miyazaki WY, Milash B, Dompe NA, Singh CM, Buchholz R, Demsky M, Fawcett R, Francis-Lang HL, Ryner L, Cheung LM, Chong A, Erickson C, Fisher WW, Greer K, Hartouni SR, Howie E, Jakkula L, Joo D, Killpack K, Laufer A, Mazzotta J, Smith RD, Stevens LM, Stuber C, Tan LR, Ventura R, Woo A, Zakrajsek I, *et al*: **A complementary transposon tool kit for *Drosophila melanogaster* using P and piggyBac.** *Nat Genet* 2004, **36**:283–287.
- Collier LS, Carlson CM, Ravimohan S, Dupuy AJ, Largaespada DA: **Cancer gene discovery in solid tumours using transposon-based somatic mutagenesis in the mouse.** *Nature* 2005, **436**:272–276.
- Dupuy AJ, Akagi K, Largaespada DA, Copeland NG, Jenkins NA: **Mammalian mutagenesis using a highly mobile somatic Sleeping Beauty transposon system.** *Nature* 2005, **436**:221–226.
- Ivics Z, Hackett PB, Plasterk RH, Ivics Z: **Molecular reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells.** *Cell* 1997, **91**:501–510.
- Starr TK, Allaei R, Silverstein KA, Staggs RA, Sarver AL, Bergemann TL, Gupta M, O'Sullivan MG, Matise I, Dupuy AJ, Collier LS, Powers S, Oberg AL, Asmann YW, Thibodeau SN, Tessarollo L, Copeland NG, Jenkins NA, Cormier RT, Largaespada DA: **A transposon-based genetic screen in mice identifies genes altered in colorectal cancer.** *Science* 2009, **323**:1747–1750.
- Ding S, Wu X, Li G, Han M, Zhuang Y, Xu T: **Efficient transposition of the piggyBac (PB) transposon in mammalian cells and mice.** *Cell* 2005, **122**:473–483.
- Wang W, Lin C, Lu D, Ning Z, Cox T, Melvin D, Wang X, Bradley A, Liu P: **Chromosomal transposition of PiggyBac in mouse embryonic stem cells.** *Proc Natl Acad Sci U S A* 2008, **105**:9290–9295.
- Su Q, Prosser HM, Campos LS, Ortiz M, Nakamura T, Warren M, Dupuy AJ, Jenkins NA, Copeland NG, Bradley A, Liu P: **A DNA transposon-based approach to validate oncogenic mutations in the mouse.** *Proc Natl Acad Sci U S A* 2008, **105**:19904–19909.
- Greenman C, Stephens P, Smith R, Dalgliesh GL, Hunter C, Bignell G, Davies H, Teague J, Butler A, Stevens C, Edkins S, O'Meara S, Vastrik I, Schmidt EE, Avis T, Barthorpe S, Bhamra G, Buck G, Choudhury B, Clements J, Cole J, Dicks E, Forbes S, Gray K, Halliday K, Harrison R, Hills K, Hintson J, Jenkinson A, Jones D, *et al*: **Patterns of somatic mutation in human cancer genomes.** *Nature* 2007, **446**:153–158.
- Bradley A, Zheng B, Liu P: **Thirteen years of manipulating the mouse genome: a personal history.** *Int J Dev Biol* 1998, **42**:943–950.
- Pettitt SJ, Liang Q, Rairdan XY, Moran JL, Prosser HM, Beier DR, Lloyd KC, Bradley A, Skarnes WC: **Agouti C57BL/6 N embryonic stem cells for mouse genetic resources.** *Nat Methods* 2009, **6**:493–495.
- Olive KP, Tuveson DA, Ruhe ZC, Yin B, Willis NA, Bronson RT, Crowley D, Jacks T: **Mutant p53 gain of function in two mouse models of Li-Fraumeni syndrome.** *Cell* 2004, **119**:847–860.
- Shultz LD, Lyons BL, Burzenski LM, Gott B, Chen X, Chaleff S, Kotb M, Gillies SD, King M, Mangada J, Greiner DL, Handgretinger R: **Human lymphoid and myeloid cell development in NOD/LtSz-scid IL2R gamma null mice engrafted with mobilized human hemopoietic stem cells.** *J Immunol* 2005, **174**:6477–6489.
- Child ES, Hendrychova T, McCague K, Futreal A, Otyepka M, Mann DJ: **A cancer-derived mutation in the PSTAIRE helix of cyclin-dependent kinase 2 alters the stability of cyclin binding.** *Biochimica et biophysica acta* 2010, **1803**:858–864.
- Caricasole A, Bettini E, Sala C, Roncarati R, Kobayashi N, Caldara F, Goto K, Terstappen GC: **Molecular cloning and characterization of the human diacylglycerol kinase beta (DGKbeta) gene: alternative splicing generates DGKbeta isoforms with different properties.** *J Biol Chem* 2002, **277**:4790–4796.
- Fraser MJ, Ciszczon T, Elick T, Bauser C: **Precise excision of TTAAspecific lepidopteran transposons piggyBac (IFP2) and tagalong (TFP3) from the baculovirus genome in cell lines from two species of Lepidoptera.** *Insect Mol Biol* 1996, **5**:141–151.
- Grossman GL, Rafferty CS, Fraser MJ, Benedict MQ: **The piggyBac element is capable of precise excision and transposition in cells and embryos of the mosquito, *Anopheles gambiae*.** *Insect Biochem Mol Biol* 2000, **30**:909–914.
- Plasterk RH, Ivicsk Z, Ivics Z: **Resident aliens: the Tc1/mariner superfamily of transposable elements.** *Trends Genet* 1999, **15**:326–332.
- Luo G, Ivics Z, Ivicsk Z, Bradley A: **Chromosomal transposition of a Tc1/mariner-like element in mouse embryonic stem cells.** *Proc Natl Acad Sci U S A* 1998, **95**:10769–10773.
- The Cancer Gene Census** [http://www.sanger.ac.uk/genetics/CGP/Census/]
- Auf G, Jabouille A, Guerit S, Pineau R, Delugin M, Boucheccareilh M, Magnin N, Favereaux A, Maitre M, Gaiser T, von Deimling A, Czabanka M, Vajkoczy P, Chevet E, Bikfalvi A, Moenner M: **Inositol-requiring enzyme 1alpha is a key regulator of angiogenesis and invasion in malignant glioma.** *Proc Natl Acad Sci U S A* 2010, **107**:15553–15558.
- Drogat B, Auguste P, Nguyen DT, Boucheccareilh M, Pineau R, Nalbantoglu J, Kaufman RJ, Chevet E, Bikfalvi A, Moenner M: **IRE1 signaling is essential for ischemia-induced vascular endothelial growth factor-A expression and contributes to angiogenesis and tumor growth in vivo.** *Cancer Res* 2007, **67**:6700–6707.
- Choi YL, Bocanegra M, Kwon MJ, Shin YK, Nam SJ, Yang JH, Kao J, Grandis JR: **Lyn kinase mediates cell motility and tumor growth in EGFRvIII-expressing head and neck cancer.** *Clin Cancer Res* 2012, **18**:2850–2860.
- Rad R, Rad L, Wang W, Cadinanos J, Vassiliou G, Rice S, Campos LS, Yusa K, Banerjee R, Li MA, de la Rosa J, Strong A, Lu D, Ellis P, Conte N, Yang FT, Liu P, Bradley A: **PiggyBac transposon mutagenesis: a tool for cancer gene discovery in mice.** *Science* 2010, **330**:1104–1107.
- Gerlinger M, Rowan AJ, Horswell S, Larkin J, Endesfelder D, Gronroos E, Martinez P, Matthews N, Stewart A, Tarpey V, Varela I, Phillimore B, Begum S, McDonald NQ, Butler A, Jones D, Raine K, Latimer C, Santos CR, Nohadani M, Eklund AC, Spencer-Dene B, Clark G, Pickering L, Stamp G, Gore M, Szallasi Z, Downward J, Futreal PA, Swanton C: **Intratumor heterogeneity and branched evolution revealed by multiregion sequencing.** *N Engl J Med* 2012, **366**:883–892.
- Park SY, Gonen M, Kim HJ, Michor F, Polyak K: **Cellular and genetic diversity in the progression of in situ human breast carcinomas to an invasive phenotype.** *J Clin Invest* 2010, **120**:636–644.

34. Bric A, Miething C, Bialucha CU, Scuoppo C, Zender L, Krasnitz A, Xuan Z, Zuber J, Wigler M, Hicks J, McCombie RW, Hemann MT, Hannon GJ, Powers S, Lowe SW: **Functional identification of tumor-suppressor genes through an in vivo RNA interference screen in a mouse lymphoma model.** *Cancer Cell* 2009, **16**:324–335.
35. Gumireddy K, Sun F, Klein-Szanto AJ, Gibbins JM, Gimotty PA, Saunders AJ, Schultz PG, Huang Q: **In vivo selection for metastasis promoting genes in the mouse.** *Proc Natl Acad Sci U S A* 2007, **104**:6696–6701.
36. Wangenstein KJ, Wilber A, Keng WW, He Z, Matise I, Wangenstein L, Carson CM, Chen Y, Steer CJ, McIvor RS, Largaespada DA, Wang X, Ekker SC: **A facile method for somatic, lifelong manipulation of multiple genes in the mouse liver.** *Hepatology* 2008, **47**:1714–1724.
37. Wiesner SM, Decker SA, Larson JD, Ericson K, Forster C, Gallardo JL, Long C, Demorest ZL, Zamora EA, Low WC, SantaCruz K, Largaespada DA, Ohlfest JR: **De novo induction of genetically engineered brain tumors in mice using plasmid DNA.** *Cancer Res* 2009, **69**:431–439.
38. Zender L, Xue W, Zuber J, Semighini CP, Krasnitz A, Ma B, Zender P, Kubicka S, Luk JM, Schirmacher P, McCombie WR, Wigler M, Hicks J, Hannon GJ, Powers S, Lowe SW: **An oncogenomics-based in vivo RNAi screen identifies tumor suppressors in liver cancer.** *Cell* 2008, **135**:852–864.
39. McGregor LM, Baylin SB, Griffin CA, Hawkins AL, Nelkin BD: **Molecular cloning of the cDNA for human TrkC (NTRK3), chromosomal assignment, and evidence for a splice variant.** *Genomics* 1994, **22**:267–272.
40. McMahon AP, Bradley A: **The Wnt-1 (int-1) proto-oncogene is required for development of a large region of the mouse brain.** *Cell* 1990, **62**:1073–1085.
41. Saci A, Carpenter CL: **RhoA GTPase regulates B cell receptor signaling.** *Mol Cell* 2005, **17**:205–214.
42. Li H, Durbin R: **Fast and accurate long-read alignment with Burrows-Wheeler transform.** *Bioinformatics* 2010, **26**:589–595.
43. Do CB, Batzoglu S: **What is the expectation maximization algorithm?** *Nat Biotechnol* 2008, **26**:897–899.
44. Ye K, Schulz MH, Long Q, Apweiler R, Ning Z: **Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads.** *Bioinformatics* 2009, **25**:2865–2871.
45. Varela I, Tarpey P, Raine K, Huang D, Ong CK, Stephens P, Davies H, Jones D, Lin ML, Teague J, Bignell G, Butler A, Cho J, Dalgliesh GL, Galappaththige D, Greenman C, Hardy C, Jia M, Latimer C, Lau KW, Marshall J, McLaren S, Menzies A, Mudie L, Stebbings L, Largaespada DA, Wessels LF, Richard S, Kahnoski RJ, Anema J, *et al*: **Exome sequencing identifies frequent mutation of the SWI/SNF complex gene PBRM1 in renal carcinoma.** *Nature* 2011, **469**:539–542.

doi:10.1186/s13059-014-0455-6

Cite this article as: Chew *et al*: Polygenic in vivo validation of cancer mutations using transposons. *Genome Biology* 2014 **15**:455.

Submit your next manuscript to BioMed Central  
and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

