

RNA sequencing of *Sleeping Beauty* transposon-induced tumors detects transposon-RNA fusions in forward genetic cancer screens

Nuri A. Temiz,^{1,7} Branden S. Moriarity,^{1,2,3,4,7} Natalie K. Wolf,^{1,2,4} Jesse D. Riordan,⁵ Adam J. Dupuy,⁵ David A. Largaespada,^{1,2,3,4,6} and Aaron L. Sarver¹

¹Masonic Cancer Center, University of Minnesota, Minneapolis, Minnesota 55455, USA; ²Department of Genetics, Cell Biology and Development, University of Minnesota, Minneapolis, Minnesota 55455, USA; ³Brain Tumor Program, University of Minnesota, Minneapolis, Minnesota 55455, USA; ⁴Department of Pediatrics, University of Minnesota, Minneapolis, Minnesota 55455, USA; ⁵Department of Anatomy and Cell Biology, University of Iowa, Iowa City, Iowa 52242, USA; ⁶Center for Genome Engineering, University of Minnesota, Minneapolis, Minnesota 55455, USA

Forward genetic screens using *Sleeping Beauty* (*SB*)-mobilized T2/Onc transposons have been used to identify common insertion sites (CISs) associated with tumor formation. Recurrent sites of transposon insertion are commonly identified using ligation-mediated PCR (LM-PCR). Here, we use RNA sequencing (RNA-seq) data to directly identify transcriptional events mediated by T2/Onc. Surprisingly, the majority (~80%) of LM-PCR identified junction fragments do not lead to observable changes in RNA transcripts. However, in CIS regions, direct transcriptional effects of transposon insertions are observed. We developed an automated method to systematically identify T2/Onc-genome RNA fusion sequences in RNA-seq data. RNA fusion-based CISs were identified corresponding to both DNA-based CISs (*Cdkn2a*, *Mycl1*, *Nf2*, *Pten*, *Sema6d*, and *Rere*) and additional regions strongly associated with cancer that were not observed by LM-PCR (*Myc*, *Akt1*, *Pth*, *Csflr*, *Fgfr2*, *Wisp1*, *Map3k5*, and *Map4k3*). In addition to calculating recurrent CISs, we also present complementary methods to identify potential driver events via determination of strongly supported fusions and fusions with large transcript level changes in the absence of multitumor recurrence. These methods independently identify CIS regions and also point to cancer-associated genes like *Braf*. We anticipate RNA-seq analyses of tumors from forward genetic screens will become an efficient tool to identify causal events.

[Supplemental material is available for this article.]

Insertional mutagenesis approaches have identified components of metabolic, signaling, and developmental pathways (St Johnston 2002). Slow-transforming retroviruses (Uren et al. 2005) and engineered transposon/transposase systems have been used to implicate driver genes involved in tumor formation (Collier et al. 2005). Tissue-specific expression (Starr et al. 2009) of *Sleeping Beauty* (*SB*) transposase (Ivics et al. 1997) can mobilize genomically integrated T2/Onc, inducing tumor formation in wild-type (WT) mice and accelerating tumor formation in predisposed mice. Sites of transposon insertion are identified via direct sequencing of ligation-mediated PCR (LM-PCR) products generated from restriction enzyme (reLM-PCR) digestion or shearing (shLM-PCR) of tumor DNA. Processing of the sequencing data generates lists of insertion locations, which are further analyzed to identify common insertion sites (CISs) (Sarver et al. 2012). A wide range of tumor tissues including gastrointestinal tract, liver, peripheral nerve, pancreas, and osteosarcoma (Keng et al. 2009; Starr et al. 2009; Pérez-Mancera et al. 2012; Rahrman et al. 2013; Moriarity et al. 2015) have been studied using these approaches.

The T2/Onc transposon contains a bi-functional molecular construct designed to accelerate tumor formation using two separate mechanisms. T2/Onc can activate transcription and utilize an

internal splice donor to generate in-frame fusion mRNA transcripts. T2/Onc also contains two independent splice acceptors capable of disruption of transcript independent of the orientation of transposon insertion (Collier et al. 2005).

While use of reLM-PCR followed by CIS analyses has successfully identified many tumor suppressors and oncogenes important to both human and mouse tumors, this methodology suffers from a number of limitations. Properly positioned restriction sites are necessary for successful cloning. If they are not present in ideal locations, reLM-PCR will fail to identify insertions. The LM-PCR process also utilizes two PCR amplification steps, and as a result the relative clonality of an insertion within a given tumor cannot be clearly defined unless shearing-based approaches are utilized. The use of shearing prior to linker attachment partially alleviates some of these problems because the presence of multiple fragments may indicate that an insertion is clonal. Further, insertions present on the chromosome where the T2/Onc transposon concatemer resides prior to mobilization are often not considered in CIS analyses, due to the “local hopping” phenomenon of *SB* transposition. Additionally, it is often not clear (1) which gene is functionally affected by a set of recurrent insertions and (2) whether the

⁷These authors contributed equally to this work.

Corresponding author: sarver@umn.edu

Article published online before print. Article, supplemental material, and publication date are at <http://www.genome.org/cgi/doi/10.1101/gr.188649.114>.

© 2016 Temiz et al. This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see <http://genome.cshlp.org/site/misc/terms.xhtml>). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

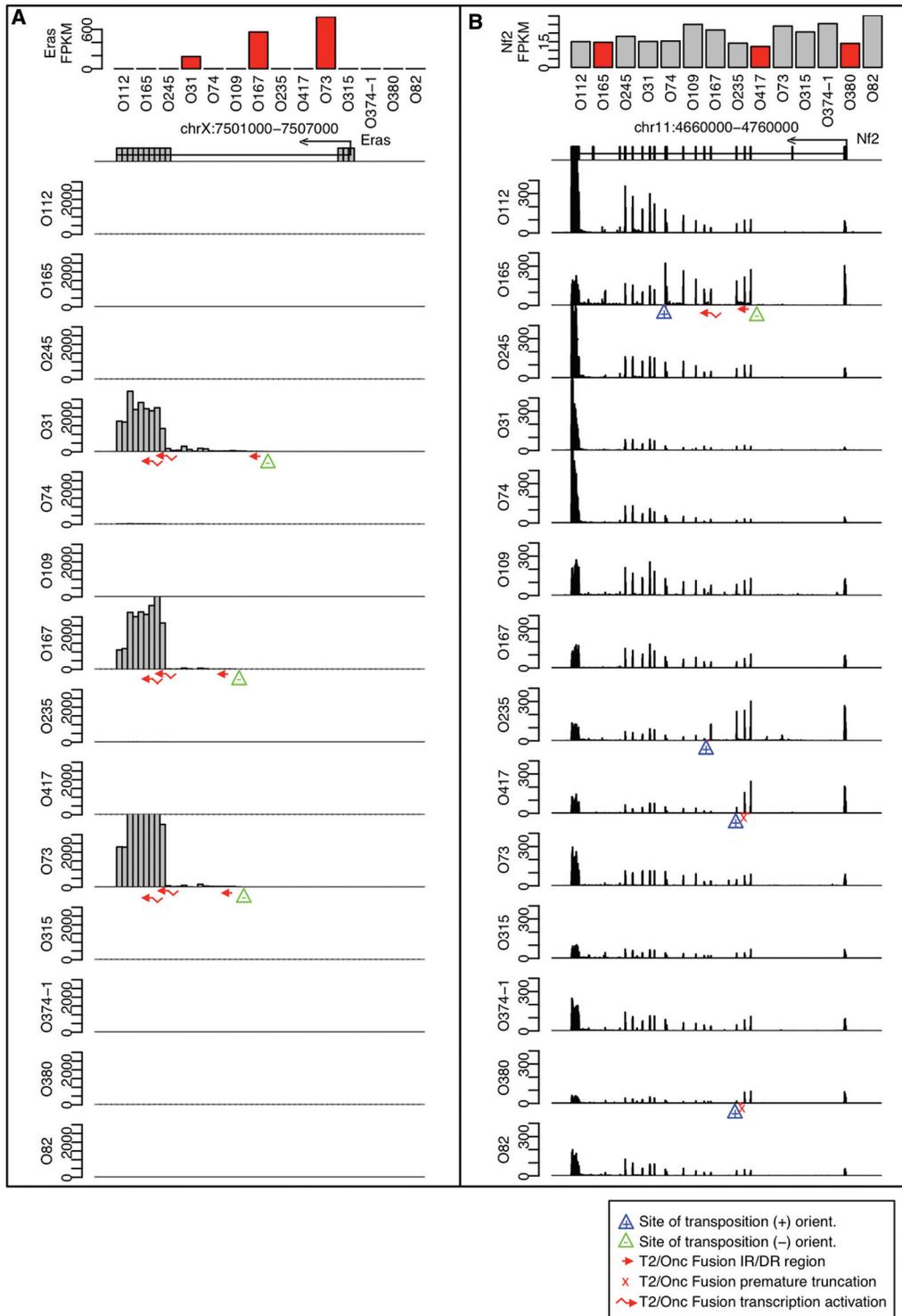


Figure 1. Transcription near CIS insertions. Image showing FPKM values for (A) *Eras* and (B) *Nf2* for each of the 14 *SB* osteosarcoma (OS) tumors. An exon map of the gene region is shown indicating the direction of transcription *below* the FPKM bar plots. For each tumor, a histogram of the raw number of reads observed within each 100-bp window is shown. The locations of LM-PCR-identified insertions are shown with triangles. Blue triangles represent positive orientation insertions that can activate transcription on the + strand, while green triangles represent negative orientation insertions that can activate transcription on the - strand. RNA fusions identified in the RNA-seq data are described in the figure legend (*bottom right*).

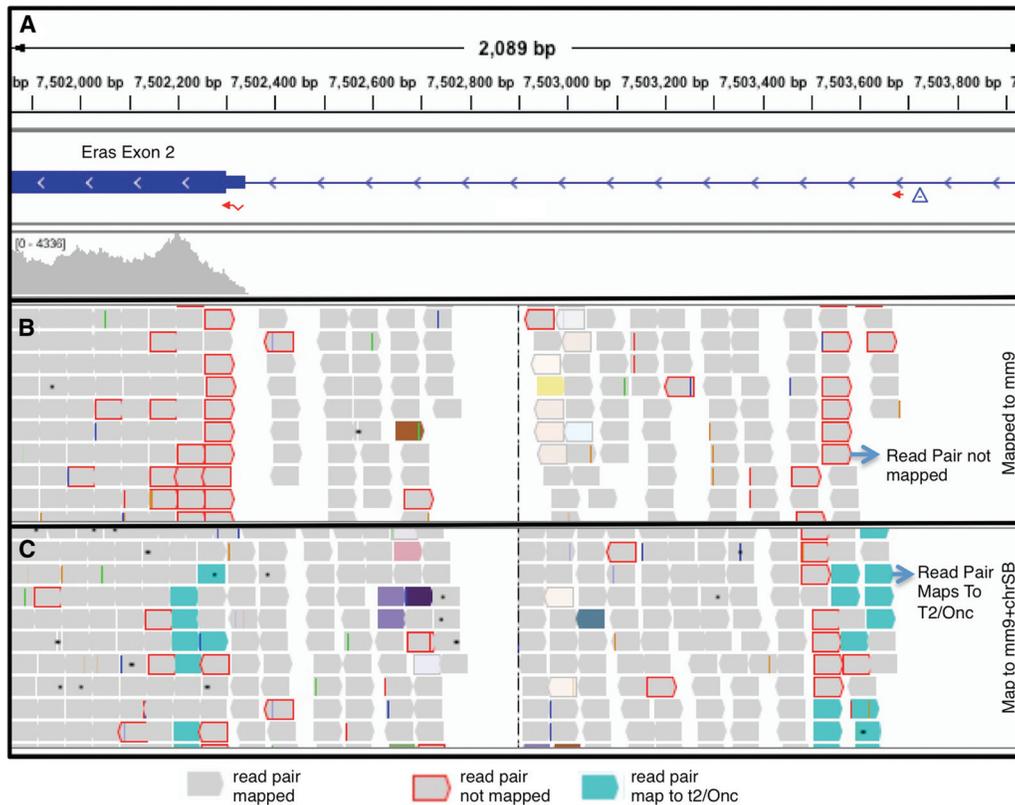


Figure 2. Mapping to a modified mm9 genome. (A) IGV view of *Eras* genomic region from a tumor containing a transposon insertion. (B) Within the *Eras* transcript Exon 2, many unmapped read pairs were observed in the regions of fusions following mapping to the standard mm9 genome. (C) Remapping to the mm9 genome with added T2/Onc sequence reveals the presence of T2/Onc-genome fusion sequences. Many of the previously unmapped pairs now map to the T2/Onc sequence.

transposon is activating transcription or generating premature truncations.

Given these limitations, we set out to determine whether we could use RNA-seq data obtained from *SB*-mutagenized tumors to identify how transposon insertion in a given tumor was affecting RNA transcripts of nearby loci. We show that RNA-seq data can be used to identify insertions directly via the presence of transposon-endogenous gene fusion transcripts and provide a pipeline for this analysis.

Results

Direct visualization of transcriptional modulation by transposon insertion

RNA-seq data was obtained from *SB*-induced WT ($n = 5$) and *SB*-accelerated LSL-*Trp53*^{R270H} ($n = 9$) osteosarcoma tumors from 14 unique mice (Moriarity et al. 2015). We hypothesized that transcriptional modifications driven by T2/Onc insertion would be directly observable by examining the sequence mapping footprints in regions of the genome containing CISs. Functionally relevant insertions should modify transcript levels near genes with transposon insertions, and these modifications should not be present in tumors without insertions. To test this, the RNA-seq data sets were mapped to the mouse genome, and the genomic regions around CISs were visualized. Strong evidence for transposon-mediated transcriptional modification was present. For example, *Eras* is

a putative oncogene activating the PI3-kinase pathway in undifferentiated stem cells (Takahashi et al. 2003). Three out of 14 mice had *SB* insertions detected by LM-PCR between the first and second exons in orientations expected to activate transcription of *Eras*. In the three tumors with insertions, large increases in transcript are visually apparent corresponding to insertion-mediated transcriptional activation. The three tumors with insertions have FPKM values above 100, while the remaining 11 have values close to zero ($P < 3.97 \times 10^{-5}$) (Fig. 1A). *Myc* (2/14) and *Mycl1* (2/14) were also observed to be significantly increased in tumors with evidence for transposon insertions (Supplemental Fig. S1).

As an example of a loss-of-function insertion, *Nf2* is a tumor suppressor gene (Morrison et al. 2001) and 4/14 tumors had LM-PCR insertions within *Nf2*. In each of the four tumors, decreases in the transcript levels are observable directly following the transposon insertion site consistent with the premature truncation of a single allele of the *Nf2* gene. Decreased transcript levels were not observable in the remaining 10 tumors that did not contain transposon insertions (Fig. 1B).

Paired-end sequences derived from mRNA near sites of T2/Onc insertion contain fusion transcripts

Paired-end reads that mapped near the site of insertion in each of the exemplary CIS cases described above were examined using the IGV browser. In a typical RNA-seq experiment, the vast majority of paired-end sequences will map near each other in genomic

location, taking into consideration the complexities of splicing. In regions near transposon insertion, sequences were present where one read pair mapped to the genome while the other read pair did not. The majority of the single mapping sequences pointed in the same direction, and this was often associated with a change in overall transcript level. We hypothesized that fusion transcripts composed of murine genome-derived mRNA fused to transposon-derived mRNA were present. To test this hypothesis, we mapped the sequence data to the mm9 genome where the T2/Onc transposon sequence (position 1–2153) was added as an extra chromosome. Approximately 50% of the previously single-end mapping sequences now mapped as fusion sequences, where one end of

the pair mapped to a genomic location and the other mapped to the T2/Onc transposon (Fig. 2).

Fusion sequences were observed to map to five distinct regions of the T2/Onc transposon (Fig. 3A). The location of fusion pair mapping within T2/Onc indicates which genetic element of T2/Onc is being utilized. Fusions generated from the T2/Onc transcriptional activator mapped to the splice donor region following the transcriptional activator. Tumor O245 had an insertion between Exon 6 and Exon 7 of *Capzb* that led to transcriptional activation starting at Exon 7 (Fig. 3B). In addition to initiating known transcripts, the T2/Onc transcriptional activator was also observed to trigger previously unobserved transcripts. Tumor O165 had an

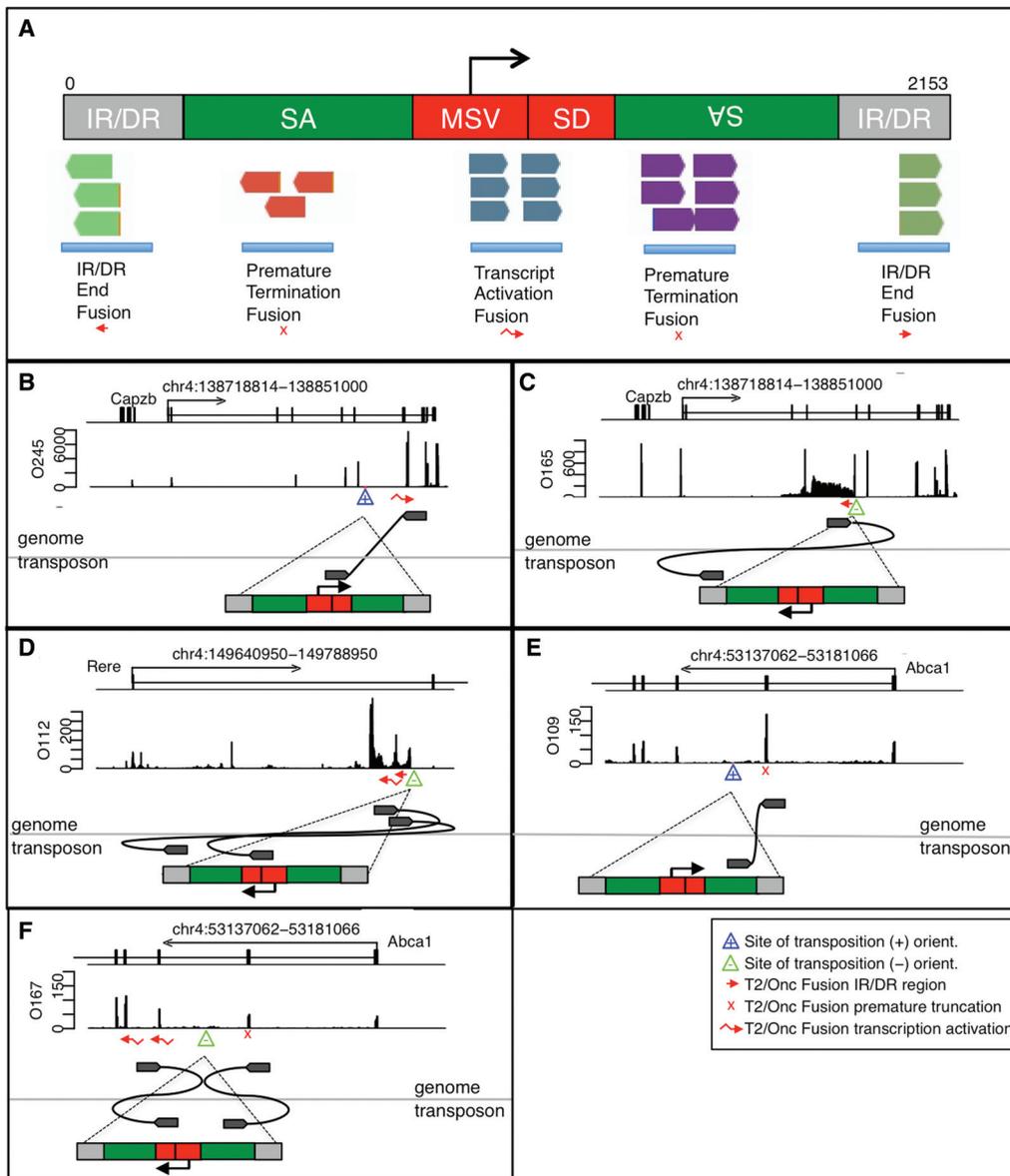


Figure 3. Location of fusion within T2/Onc sequence indicates transposon mechanism of transcriptional alteration. (A) Map of a single copy of the T2/Onc transposon showing the locations of the various fusion products. The location of the fusion pair mapping to T2/Onc indicates how T2/Onc is modifying the transcript in order to generate the observed fusion event. (B) Transcriptional activation within the *Capzb* gene. (C) Transcriptional activation of the opposite strand of the *Capzb* gene. (D) Transcriptional activation on the opposite strand with apparent splicing in the *Rere* gene. (E) Premature truncation in *Abca1*. (F) Premature truncation and transcriptional activation happening at the same time in *Abca1*. Locations of LM-PCR junction fragments and fusion products are shown as described for Figure 1.

insertion in an orientation predicted to activate transcription of a novel transcript on the opposite strand relative to the *Capzb* gene (Fig. 3C). In additional cases, splicing of the novel opposite strand transcript was observed. Tumor O112 had an insertion between Exon 1 and Exon 2 of *Rere* in an orientation predicted to lead to premature disruption of the transcript (Fig. 3D).

Fusions generated from prematurely truncated sequences mapped to the splice acceptor regions. An insertion was observed between Exon 2 and Exon 3 in *Abca1* in the same orientation as the insertion in *Rere*, but instead of observing opposite strand transcriptional activation, we observed a truncation fusion product and a decrease in transcript (Fig. 3E). An insertion in the opposite orientation in *Abca1* resulted in both a premature truncation and a transcriptional activation within the same gene (Fig. 3F). In addition to spliced products, fusions generated from transcriptional read-through from the inverted repeat/direct repeat (IR/DR) were observed in *Eras* and *Rere* (Figs. 1A, 3C). These were usually associated with the site of insertion identified by LM-PCR and were present at a lower rate than the spliced products, when spliced products were also present.

RNA fusions can be systematically identified

We identified all sequences where one end of the paired read mapped to one of the five T2/Onc sequence regions and the other mapped to the genome. Across the 14 tumors, 1806 fusions were observed (Supplemental Table S1). Many of the fusions are in close proximity to each other and describe alternatively spliced products. Each tumor had 28 to 324 distinct fusions (Table 1).

Libraries were sequenced to varying depths from 18 to 241 million paired-end reads to identify the optimal read depth. The depth of sequencing showed a relationship with the total number of fusion regions observed ($r^2 = 0.56$). The slope of the trend line (1.4) indicates that, for each million additional reads in depth, on average an additional 1.4 fusions will be identified. The y -intercept of the trend line (31) indicates that, even at lower sequence depth, a significant number of fusions will be identified (Fig. 4A). To normalize the number of fusion events across such a wide depth of sequencing, we calculated counts per million (CPM) values for each fusion. A CPM of 0.5 would mean that in the smallest library at least 10 fusion-spanning paired-end reads were present.

In the *SB* tumors, apparent fusions were observed in the *Foxf2* gene, as well as the *En2* gene, both of which were used in the construction of the T2/Onc transposon. In addition, we mapped RNA-seq data from tumors derived from a traditional genetic mouse tumor model (without *SB* mutagenesis) and did not observe fusions.

Many reLM-PCR insertions are not supported by RNA fusions, and insertions with fusion support have increased read counts

Across these 14 tumors, 2223 insertion sites were identified by reLM-PCR followed by TAPDANCE analysis (Supplemental Table S3). Three hundred seventy-six of the 2223 insertion sites (~17%) have an RNA fusion event within $\pm 100,000$ bp of the LM-PCR site (Table 1). While reLM-PCR amplification is not linear, we suspected reLM-PCR insertions with increased read numbers are generally more likely to be clonal driver events, therefore generating observable fusions. If this was the case, reLM-PCR reads with RNA fusion support should have increased read numbers relative to reLM-PCR reads without RNA fusion support. reLM-PCR insertions supported by RNA-transposon fusions were observed, on average, 5142 times across all 14 libraries, while unsupported insertions were observed an average of 1026 times (Table 1). The increase in read number for fusion-supported insertions holds true independently for each tumor library and is highly statistically significant ($P = 0.00004$, two-group t -test).

Clonal insertions identified by shLM-PCR have increased fusion support

Shearing has been used as an alternative to restriction enzymes to fragment DNA prior to LM-PCR and has been reported as a semi-quantitative method to identify clonal insertions based on the presence of multiple supporting fragments (Koudijs et al. 2011; Riordan et al. 2014). If the observed fusions were generated primarily from clonal insertions, shLM-PCR insertion locations with more supporting fragments would have increased fusion support. We generated shLM-PCR data from six tumors for which we also had RNA-seq data and examined the support (Supplemental Table S4). As the number of distinct supporting fragments increases, the likelihood of fusion support also increases (Fig. 4C). These results indicate that shLM-PCR junctions with many shear

Table 1. Summary of tumor RNA-seq, LM-PCR, and fusion data

Tumor	Genotype	Reads (millions)	Mapping	reLM-PCR	% LM-PCR with fusion support	Fusions	% fusions with LM-PCR support	Avg count # supported	Avg count # no support	Fusions with CPM >0.5	Genes with CPM >0.5
O112	WT	242	93%	63	59%	324	48%	6702	835	42	23
O31	WT	96	91%	134	36%	293	42%	4264	1123	38	24
O74	WT	95	93%	172	15%	99	47%	7723	1028	7	7
O245	WT	91	94%	192	32%	283	61%	6634	2204	55	32
O73	WT	58	87%	189	25%	206	76%	6399	288	74	38
O165	p53	83	95%	411	10%	126	70%	3919	994	11	8
O109	p53	62	89%	98	6%	28	32%	1729	914	3	2
O235	p53	60	95%	346	7%	48	67%	3944	540	14	7
O167	p53	57	84%	40	63%	121	71%	4960	1437	30	18
O417	p53	54	94%	248	10%	70	40%	2385	1190	13	12
O374-1	p53	31	96%	42	19%	79	37%	12,650	2842	33	13
O82	p53	24	95%	110	11%	30	67%	6634	2204	8	5
O380	p53	20	96%	53	9%	49	20%	15,321	2279	15	10
O315	p53	18	96%	125	8%	50	36%	5129	1125	22	15

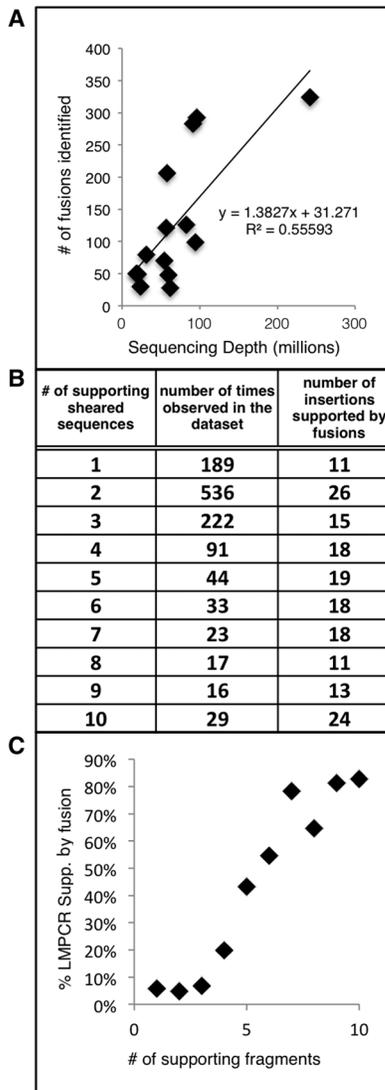


Figure 4. Interplay between sequencing depth, LM-PCR depth, and RNA fusions. (A) Total number of fusions observed in each sample plotted against the sequencing depth. (B) The number of fragments in six tumors using the shLM-PCR technique, ordered in increasing number of fragments which is a semiquantitative indication of clonality and the number of fusions which support these shLM-PCR junction fragments. (C) Relationship between shLM-PCR number of supporting fragments and RNA fusion support. As more supporting shear fragments are observed, the percentage of insertions with observed fusions increases.

fragments are more likely to be clonal and generate fusion transcripts than shLM-PCR junctions with fewer shear fragments.

RNA fusions are observed that are not supported by LM-PCR

Of the 1806 potential fusions identified across 14 tumors, 974 (53%) are within 100,000 bases of a reLM-PCR-identified insertion. Eight hundred thirty-two fusions without LM-PCR support map to 514 unique genes, including a fusion to *Myc*. Of the 574 fusions identified across six tumors, where shLM-PCR was available, 407 (71%) were supported by shLM-PCR. One hundred sixty-seven fusions without shLM-PCR map to 120 genes and include fusions to *Rb1*, *Wrm*, *Notch2*, and *Csf1r*. These results indicate that a signif-

icant number of transposon insertions are being missed by LM-PCR based methodology but can be inferred by the presence of RNA fusions. RT-PCR was used to validate that transposon-derived fusion sequences were present in the tumors in six examples where reLM-PCR did not identify transposon insertions (Supplemental Figs. S2–S9).

RNA fusions can define precise RNA CIS (RCIS) genes

RCISs were calculated from the RNA fusion locations and compared to DNA CISs (DCISs) generated with reLM-PCR junction fragments. Thirty-five DCIS regions (Supplemental Table S5) and 31 RCIS regions (Supplemental Table S6) were identified. Many regions were present in both lists, including *Eras*, *Nf2*, *Pten*, *Mycl1*, *Cdkn2a*, *Sema6d*, *Rere*, *Ammecr1*, *Heatr3*, *Abca1*, *Lims2*, and *Cd2ap* (Supplemental Table S7). Additional genomic regions strongly associated with cancer were only found in the RCIS list, including *Myc*, *Akt1*, *Pth*, *Csf1r*, *Fgfr2*, *Wisp1*, *Map3k5*, and *Map4k3*. Significantly more RCISs were observed in *SB*-induced WT tumors than in *SB*-accelerated *LSL-Trp53^{R270H}* tumors ($P = 0.0008$).

In many cases, CISs have multiple genes in close proximity, and RNA-seq data can discern which transcripts are directly modified. We identified insertions between *Ammecr1* and *Rgag1* and the insertions are closer to the *Ammecr1* gene, while *Rgag1* transcript level is strongly increased (Fig. 5A). RNA-seq data can also allow identification of the specific transcriptional modification occurring in each tumor. The *Cdkn2a* (p16) locus has reLM-PCR insertions in four tumors, two within the gene predicted to disrupt the transcript and one in an orientation suggesting transcriptional activation (Fig. 5B). The tumor with the highest level of *Cdkn2a* (O73) has a reLM-PCR insertion and fusion indicating transcriptional activation of *Cdkn2a*. The tumor with the lowest level of *Cdkn2a* transcript (O74) has an insertion activating transcription of a noncanonical transcript on the opposite strand. In summary, RNA-seq data complement reLM-PCR and allow direct identification of the transposon-mediated transcriptional changes within CISs that were not clear from the reLM-PCR data alone.

Tumor-centric models of recurrent and nonrecurrent driver events

Identifying driver events that are present in specific tumors is necessary to understand heterogeneous tumor formation, and nonrecurrent fusions may be important drivers of tumor formation (Table 2). Two distinct approaches were utilized with the RNA-seq data to identify nonrecurrent drivers. First, fusions with strong support defined by fusion CPM > 0.5 were identified. Many of these were previously identified via CIS analyses in our larger reLM-PCR (Moriarity et al. 2015). This analysis also identified nonrecurrent oncogenes, such as *Ksr1* (Müller et al. 2001), *Ptpn13* (Dromard et al. 2007), *Wisp1* (Xu et al. 2001), and *Ptk2* (also known as *FAK*) (Mitra and Schlaepfer 2006).

As a second approach, insertion-modified transcripts (IMTs), with outlier expression levels and RNA fusions, were identified. IMTs with increased expression included many genes identified as CISs, such as *Eras*, *Myc*, and *Mycl1* (Table 2). For example, increases in 3' exons were observed in *Cngb3*, *Poln*, and *Braf*, which were not previously seen as a DCIS or RCIS (Fig. 6A–C). IMTs with decreased levels included many genes described as CISs, such as *Nf2*, *Pten*, *Cdkn2a*, and *Rere*. In addition to truncation, transposon insertion also leads to opposite strand transcription as shown for *Cdc42*, *Xpo4*, and *Max* (Fig. 6D–F).

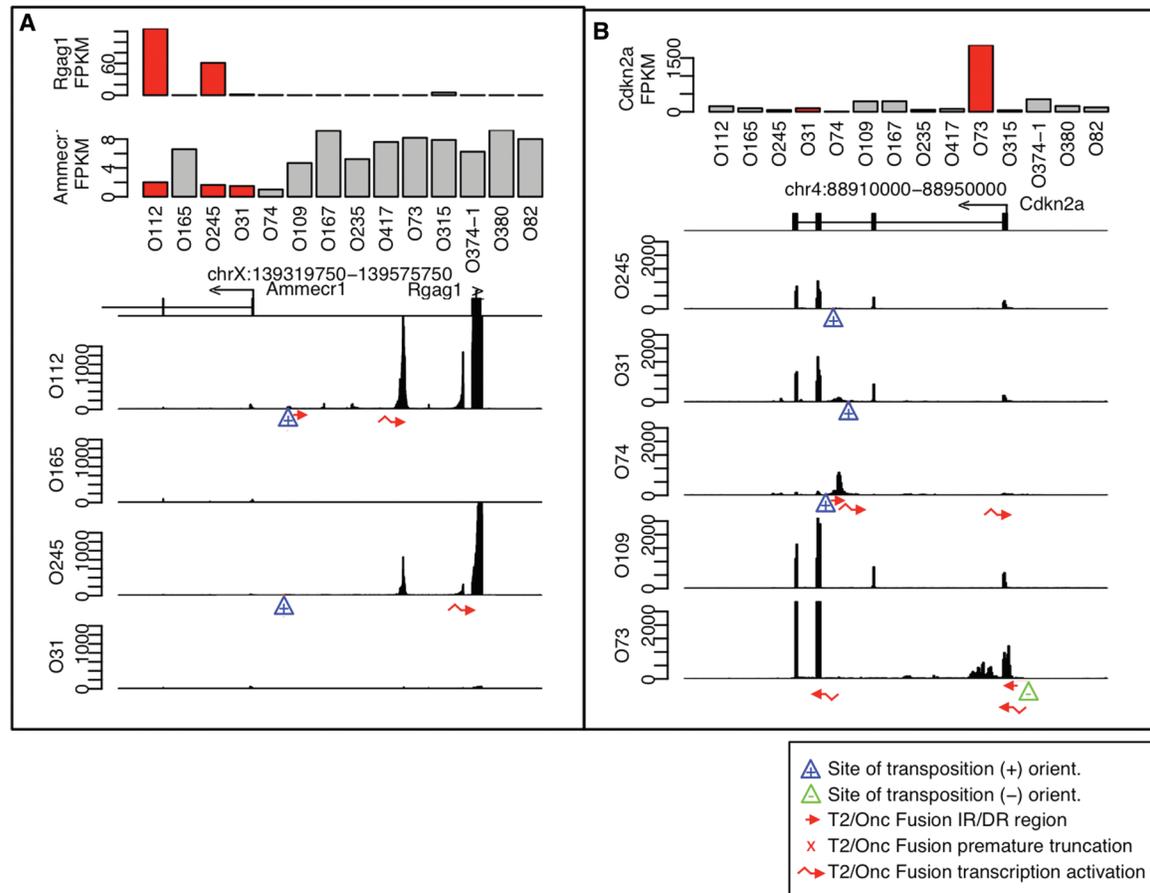


Figure 5. Precise identification of CIS transposon targeting. Histograms, similar to Figure 1, allow precise identification of the molecular mechanism of insertion in CIS near (A) *Ammecr1* and *Rgag1*, and (B) *Cdkn2a*. Locations of LM-PCR junction fragments and fusion products are shown as described for Figure 1.

Discussion

Analyses of RNA-seq data identified changes in transcripts and RNA fusions consistent with LM-PCR-identified insertion sites. However, only 20% of the LM-PCR insertions had supporting fusions. Many fusions were observed in the absence of LM-PCR-defined insertion sites, which were validated by RT-PCR and direct Sanger sequencing. In addition to previously described DCIS, recurrent and nonrecurrent osteosarcoma-specific oncogenes and tumor suppressors were identified by RCIS, high CPM fusions, and IMTs from the RNA-seq data.

Functionally relevant insertion events may be more observable in RNA-seq-based analyses of transposon fusions than in reLM-PCR analyses of junction fragments because fusion events must make up a significant part of the transcript population. In reLM-PCR, no estimate is made of the “normal” population at a given insertion locus, while in RNA-seq this is implicit in the methodology. As a result, reLM-PCR techniques may identify false positive insertion sites that are not clonal. Shearing prior to LM-PCR can act as a semiquantitative assessment of clonality, but the removal of insertions that may appear to be nonclonal via shearing may come at the cost of generating false negatives where fusion support exists.

Statistical analyses of association between transposon insertion and overall transcript level clearly showed transcriptional ac-

tivation was occurring as a result of transposon insertion near *Myc*, *Mycl1*, and *Eras*. However, tumors with T2/Onc fusions in *Nf2* showed a trend toward decreased transcript levels ($P < 0.06$). Rationale for the lack of significance include (1) that *Nf2* transcript levels vary for each sample, (2) that the presence of prematurely truncated polyadenylated product may increase the overall FPKM levels to a level near those observed in nontruncated products, and (3) that sequences leading to mRNA degradation may be removed in prematurely truncated transcripts masking decreases. RNA-seq fusion-based methods are better at identifying transcriptional activation events compared to LM-PCR, because the dynamic range of transcriptional activation signal is much greater than heterozygous premature transcript termination. LM-PCR will not show signal bias based on transcriptional activation since it is a DNA-based method.

Decreases in transcript observed near insertions activating opposite strand sequences may be driving RNA silencing interactions via the degradation of complementary mRNA as is commonly utilized to knock down genes in cells utilizing small interfering RNA. While unexpected, this functionality is not surprising due to the harnessing of evolutionary tumor selection in forward genetic screens.

Pth, *Ptprd*, and *Tgfb1* were identified only after RNA-seq-based analysis approaches described here, and each gene has clear roles in osteosarcoma biology. Parathyroid Hormone (*Pth*) has been

Table 2. Tumor-centric view of potential driver events

	RNA CIS	CPM >0.5	Increased IMTs	Decreased IMTs
O112	<i>Ammecr1</i> <i>Heatr3</i> <i>Ints8</i> <i>Mycl1</i> <i>Nmnat2</i> <i>Reep3</i> / <i>Jmjd1c</i> <i>Rere</i> <i>Sema6d</i>	AC103621.1 AC125270.2 <i>Bcar3</i> <i>Chka</i> <i>Creb1</i> <i>Gm10779</i> <i>Heatr3</i> <i>lars</i> <i>Ints8</i> <i>Krt222</i> <i>Mycl1</i> <i>Poln</i> <i>Prkag2</i> <i>Rab2a</i> <i>Ralb</i> <i>Serpina12</i> <i>Setd2</i> <i>Smarce1</i> <i>Sorcs2</i> <i>Tfb2m</i> <i>Tmtc2</i> <i>Zmynd11</i> <i>mmu-mir-760</i>	1700007G11Rik <i>Asxl3</i> <i>Cnst</i> <i>Dctd</i> <i>Fgr</i> <i>Gab1</i> <i>Gm10754</i> <i>Gm5591</i> <i>Heatr3</i> <i>lars</i> <i>Jmjd1c</i> <i>Krt222</i> <i>Krt24</i> <i>Mfsd2a</i> <i>Mycl1</i> <i>Pcdha6</i> <i>Plcd3</i> <i>Poln</i> <i>Prkag2</i> <i>Rgag1</i> <i>Sema6d</i> <i>Serpina12</i> <i>Setd2</i> <i>Sh3rf3</i> <i>Skint5</i> <i>Ufm1</i> <i>Utrn</i>	<i>Arhgef3</i> <i>Bcar3</i> <i>Fry</i> <i>Gphn</i> <i>Maf</i> <i>Mbd5</i> <i>Mll3</i> <i>Napi1/1</i> <i>Rab2a</i> <i>Reep3</i> <i>Rere</i> <i>Sdk2</i> <i>Sepp1</i> <i>Setd3</i> <i>Smarce1</i> <i>Spred2</i> <i>Tapt1</i> <i>Tmtc2</i> <i>Ube2v2</i> <i>Wwc2</i> <i>Xpo5</i> <i>Zmynd11</i>
O31	<i>Akt1</i> <i>Ammecr1</i> <i>Cd2ap</i> <i>Cdkn2a</i> <i>Cnot7</i> <i>Eras</i> <i>Grif1</i> <i>Gtf2i</i> <i>Ints8</i> <i>Nmnat2</i>	1700112E06Rik <i>Akt1</i> <i>Cercam</i> <i>Chd7</i> <i>Eros</i> <i>Fam13b</i> <i>Fars2</i> <i>Flna</i> <i>Gbx1</i> <i>Gcnt1</i> <i>Kdm5c</i> <i>Mrpl51</i> <i>Mylk3</i> <i>Myo1e</i> <i>Ncapd2</i> <i>Pcsk1n</i> <i>Ptprd</i> <i>RP23-103L13.1</i> <i>Rcan2</i> <i>Rere</i> <i>Rsnb11</i> <i>Smad5</i> <i>Stim2</i> <i>Thsd7a</i>	1700112E06Rik <i>Ano6</i> <i>Atg10</i> <i>Atp2b1</i> <i>Btg2</i> <i>Cercam</i> <i>Cngb3</i> <i>Eras</i> <i>Fars2</i> <i>Gatad2a</i> <i>H2-M1</i> <i>Herc1</i> <i>Il1r2</i> <i>Iqch</i> <i>Mkl2</i> <i>Mrgpre</i> <i>Mrpl51</i> <i>Mylk3</i> <i>Nmnat2</i> <i>Rcan2</i> <i>Reck</i> <i>Scml2</i> <i>Sec23ip</i> <i>Sema4d</i> <i>Slc35e2</i> <i>Slc4a5</i> <i>Stim2</i> <i>Tbx21</i> <i>Tgfb1</i> <i>Tmem60</i>	<i>Ammecr1</i> <i>Bptf</i> <i>Emp1</i> <i>Fam13b</i> <i>Grif1</i> <i>Ints8</i> <i>Lrp6</i> <i>Mll5</i> <i>Nadk</i> <i>Ncapd2</i> <i>Pofut1</i> <i>Shq1</i> <i>Smad4</i> <i>Spred1</i> <i>St13</i> <i>Wdfy2</i>
O74	<i>Cdkn2a</i> <i>Fgfr2</i> <i>Map3k5</i> <i>Pten</i> <i>Pth</i>	<i>Fgfr2</i> <i>Pten</i> <i>Serbp1</i> <i>Tnfrsf21</i> <i>Tshz1</i> <i>Zbtb1</i> <i>Zbtb25</i>	<i>Gga1</i> <i>Gstm3</i> <i>Igfbp1</i> <i>Pitx3</i> <i>Poli</i> <i>Pth</i> <i>Sema4d</i> <i>Tiam2</i> <i>Zbtb25</i>	<i>Ahdel</i> <i>Arid1a</i> <i>Cdkn2a</i> <i>Chd7</i> <i>Crtc3</i> <i>Csnk1a1</i> <i>Ino80</i> <i>Itch</i> <i>Macf1</i> <i>Ptk2</i> <i>Rab12</i> <i>Serbp1</i> <i>Slc26a2</i> <i>Tshz1</i> <i>Uba3</i>
O245	<i>Ammecr1</i> <i>Bicc1</i> <i>Cd2ap</i> <i>Cnot7</i> <i>Grif1</i> <i>Gtf2i</i> <i>Ints8</i> <i>Rere</i> <i>Wisp1</i>	AC134581.1 AC145610.2 <i>Camk2g</i> <i>Capzb</i> <i>Cd2ap</i> <i>Crkrs</i> <i>Dicer1</i> <i>Doc2b</i> <i>Dpagt10</i> <i>Frs3</i> <i>Galnt2</i> <i>Gmds</i> <i>Grif1</i> <i>Ino80</i> <i>Ints8</i> <i>Map3k14</i> <i>Pik3r1</i> <i>Prps1</i> <i>Rab3b</i> <i>Runx1</i> <i>Satb2</i> <i>Sfrs3</i> <i>Snx7</i> <i>Sox13</i> <i>Stam</i> <i>Tax1bp1</i> <i>Tbck</i> <i>Tmem57</i> <i>Ube2r2</i> <i>Zdhhc2</i> <i>Zeb2</i> <i>Zswim6</i>	<i>Aak1</i> <i>Amot1l</i> <i>Brdw1</i> <i>Capzb</i> <i>Cngb3</i> <i>Doc2b</i> <i>Dpagt1</i> <i>Frs3</i> <i>Gmds</i> <i>Ino80</i> <i>Ints8</i> <i>Kalrn</i> <i>Lrrc40</i> <i>Lrrfip1</i> <i>Mef2b</i> <i>Mllt10</i> <i>Nfib</i> <i>Rab3b</i> <i>Rgag1</i> <i>Rims3</i> <i>Sema5a</i> <i>Snx7</i> <i>Sox13</i> <i>Stam</i> <i>Ube2r2</i> <i>Usp49</i>	<i>Ammecr1</i> <i>Camk2g</i> <i>Csad</i> <i>Csnk2a2</i> <i>Cul3</i> <i>Dicer1</i> <i>Fbxo11</i> <i>Gtf2i</i> <i>Kdm2a</i> <i>Map3k14</i> <i>Mkl1</i> <i>Pds5a</i> <i>Pik3r1</i> <i>Plcb4</i> <i>Rere</i> <i>Xpo4</i> <i>Zeb2</i> <i>Zswim6</i>
O73	<i>Cdkn2a</i> <i>Csf1r</i> <i>Eras</i> <i>Lims2</i> <i>Map3k5</i>	AC101777.1 <i>Acsl4</i> <i>Acvr1</i> <i>Braf</i> <i>Cdc42</i> <i>Cdkn2a</i> <i>Dcaf12</i> <i>Eras</i> <i>Gsk3b</i> <i>Hdac9</i> <i>Hectd1</i> <i>Helt</i> <i>Hnrnpm</i> <i>Hsn2</i> <i>Igfn1</i> <i>Il3ra</i> <i>Kalrn</i> <i>Keap1</i> <i>Lef1</i> <i>Lhfp</i> <i>Lims2</i> <i>Lrrc19</i> <i>Mam13</i> <i>Map3k5</i> <i>Mcph1</i> <i>Mef2a</i> <i>Phactr1</i> <i>Pparg</i> <i>Prune2</i> <i>RP23-9013.8</i> <i>Rapgef2</i> <i>Robo2</i> <i>Rpgrip1</i> <i>SNORA70</i> <i>Smg6</i> <i>Spcs1</i> <i>Zfand3</i> <i>Zfp710</i>	<i>Acsl4</i> <i>Acvr1</i> <i>Braf</i> <i>Cdkn2a</i> <i>Eras</i> <i>Frm4a</i> <i>Gsk3b</i> <i>Hdac9</i> <i>Hectd1</i> <i>Hnrnpm</i> <i>Igfn1</i> <i>Jarid2</i> <i>Keap1</i> <i>Lhfp</i> <i>Lims2</i> <i>Lrrc19</i> <i>Mam13</i> <i>Med13l</i> <i>Mef2a</i> <i>Pcsk1n</i> <i>Pex7</i> <i>Phactr1</i> <i>Pparg</i> <i>Prune2</i> <i>Rapgef2</i> <i>Robo2</i> <i>Rpgrip1</i> <i>Rm3</i> <i>Smg6</i> <i>Spcs1</i> <i>Zfp710</i>	<i>Ambra1</i> <i>AU040320</i> <i>Baz2a</i> <i>Cdc42</i> <i>Csf1r</i> <i>Eif4g3</i> <i>Galnt2</i> <i>Glt8d1</i> <i>Max</i> <i>Mcph1</i> <i>Mdga1</i> <i>Ogdh</i> <i>Ppp2r5a</i> <i>Sh3pxd2a</i> <i>Sifn5</i> <i>SNORA70</i> <i>Wdr33</i> <i>Zfand3</i>
O165	<i>Capzb</i> <i>Heatr3</i> <i>Map4k3</i> <i>Nf2</i>	<i>Aktip</i> <i>Atp2a2</i> <i>Frm4a</i> <i>Nfs1</i> <i>Rbm12</i> <i>Sfrs5</i> <i>Snrpc</i> <i>Zfc3h1</i>	<i>Aktip</i> <i>Frm4a</i> <i>Gzmc</i> <i>Muc20</i> <i>Rbm12</i> <i>Sec61a2</i> <i>Slc17a7</i> <i>Zfc3h1</i>	<i>Capzb</i> <i>Drg1</i> <i>Myo1e</i> <i>Rplp2</i> <i>Tgfb1</i> <i>Ube2r2</i> <i>Usp47</i>
O109	<i>Abca1</i> <i>Pth</i>	<i>Pth</i> <i>RP23-204P1.2</i>	<i>Fam129a</i> <i>Pth</i> <i>Sox6</i> <i>Wnt9a</i>	<i>Itga5</i> <i>Pdlim1</i> <i>Rbm28</i> <i>Rfwd2</i>
O235	<i>Bicc1</i> <i>Myc</i> <i>Reep3</i> / <i>Jmjd1c</i>	<i>Bicc1</i> <i>Dlgap4</i> <i>Myc</i> <i>Pogz</i> <i>RP23-344H23.1</i> <i>Reep3</i> <i>Tdp1</i>	2700049A03Rik <i>Bicc1</i> <i>Hook1</i> <i>Myc</i> <i>Tdp1</i>	<i>Dlg1</i> <i>Dlgap4</i> <i>Efna5</i> <i>Hps4</i> <i>Pogz</i> <i>Prpf6</i> <i>Wrrn</i>
O167	<i>Abca1</i> <i>Eras</i> <i>Mycl1</i> <i>Nsmce2</i> <i>Sema6d</i>	1700003F12Rik 1700123I01Rik <i>Aco1</i> <i>Aff4</i> <i>Alpi</i> <i>Eras</i> <i>Fam120a</i> <i>Foxo3</i> <i>Lgr4</i> <i>Mycl1</i> <i>Nsmce2</i> <i>Pkp4</i> <i>RP23-335N12.3</i> <i>Rilpl2</i> <i>Sema6d</i> <i>Smad3</i> <i>Thrap3</i> <i>Zbtb8a</i>	1700003F12Rik 1700123I01Rik <i>Abca1</i> <i>Brdt</i> <i>Eras</i> <i>Fam120a</i> <i>Gp5</i> <i>Lgr4</i> <i>Mycl1</i> <i>Nsmce2</i> <i>Pcsk1n</i> <i>Per3</i> <i>Rilpl2</i> <i>Srrm3</i> <i>Zbtb8a</i> <i>Zfp516</i>	<i>Aco1</i> <i>Cspp1</i> <i>Med16</i> <i>Smad3</i> <i>St3gal4</i> <i>Ylpm1</i>
O417	<i>Nf2</i> <i>Sash1</i> <i>Wisp1</i>	2310036022Rik <i>Arid2</i> <i>Asap1</i> <i>Ksr1</i> <i>Nf2</i> <i>Pnkdk</i> <i>Ptpn13</i> <i>Qrich1</i> <i>Rpl18a</i> <i>Smarcc1</i> <i>St3gal4</i> <i>Wisp1</i>		<i>Ddx42</i> <i>Nf2</i> <i>Rsc1a1</i> <i>Ssh2</i>
O374-1	<i>Fgfr2</i> <i>Myc</i>	2310008H09Rik <i>Al314180</i> <i>Fgfr2</i> <i>Gm9992</i> <i>Hadha</i> <i>Mto1</i> <i>Myc</i> <i>Rnf19a</i> <i>SNORA17</i> <i>Sema5a</i> <i>Slbp</i> <i>Tas2r119</i> <i>Unc93a</i>	2310008H09Rik <i>Al314180</i> <i>Gm9992</i> <i>Hadha</i> <i>Mto1</i> <i>Myc</i> <i>Rnf19a</i> <i>Slbp</i> <i>Slc47a1</i> <i>SNORA17</i> <i>SNORD123</i> <i>Tas2r119</i> <i>Unc93a</i>	<i>Spryd3</i>
O82	<i>Lims2</i>	<i>Anp32a</i> <i>Asap1</i> <i>Bear3</i> <i>Kif5b</i> <i>Tial1</i> <i>Zfxh4</i>	<i>Csnk1d</i> <i>Sik3</i>	
O380	<i>Csf1r</i> <i>Nf2</i> <i>Nsmce2</i>	AC102874.1 <i>Csf1r</i> <i>Ndufs5</i> <i>Nf2</i> <i>Nptn</i> <i>Nsmce2</i> <i>Ptp4a2</i> <i>Sfrs9</i> <i>Tarbp2</i> <i>Zbtb20</i>	<i>Nkain3</i> <i>Samd12</i>	<i>Nf2</i>
O315	<i>Akt1</i> <i>Capzb</i> <i>Map4k3</i> <i>Pten</i> <i>Sash1</i>	9030625A04Rik <i>Akt1</i> <i>Arl4c</i> <i>Capzb</i> <i>Igfb8</i> <i>Kcnb1</i> <i>Map4k3</i> <i>Notch2</i> <i>Pdgfra</i> <i>Ppp2r2b</i> <i>RP23-191F22.5</i> <i>Rpl6</i> <i>S1pr1</i> <i>Sash1</i> <i>Sme1a</i>	<i>Atp2b3</i> <i>Gm3435</i> <i>Kcnb1</i> <i>Pdgfra</i> <i>Sash1</i>	<i>Deaf15</i> <i>Pten</i> <i>Rpl6</i>

proposed as a treatment for osteoporosis and has a shown clear effect in increasing bone density in humans (Neer et al. 2001). Rats showed dose-dependent increases in osteosarcoma tumors following daily treatment with *Pth* (Vahle et al. 2002), independently validating that *Pth* has a role in osteosarcoma development. PTPRD is a protein tyrosine phosphatase, and a frameshift mutation was described in a human osteosarcoma patient (Egas-Bejar et al. 2014). Finally, a *TGFBR1* variant was associated with osteosarcoma susceptibility in a Chinese population (Hu et al. 2010).

Many papers have identified differentially expressed transcripts in tumors. What makes RNA-seq applied to *SB*-mutagenized tumors unique is that transcriptional changes in tumors in

response to transposon insertion can be directly observed. Spontaneous tumor formation utilizes a multitude of overlapping mechanisms including mutation, copy number variation, epigenetic variation, and genetic background for which complete information is difficult to obtain and often highly noisy. As a result, it is often difficult to identify causal events. Using RNA-seq to study *SB*-mutagenized tumor model systems provides identification of causal events with reduced noise. Our data indicate that transposon insertion will have effects that are less predictable and more complex than previously imagined due to an interplay between the strength of the native promoter compared to the strength of the T2/Onc promoter, the presence of cryptic transcription and

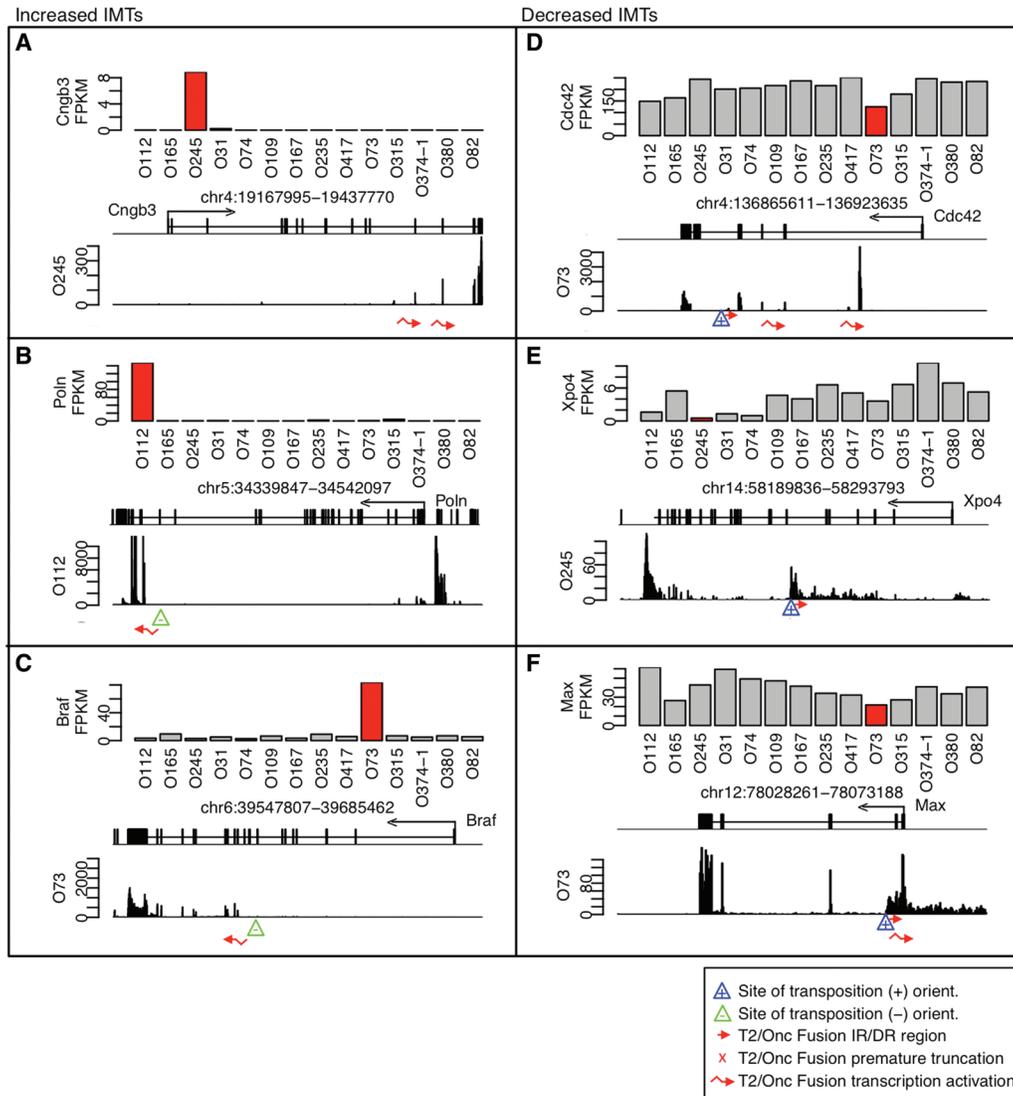


Figure 6. Insertion-modified transcripts (IMTs) in osteosarcoma tumors. Examples of increased IMTs via transcriptional activation for (A) *Cngb3*, (B) *Poln*, and (C) *Braf*, and for decreased IMTs with apparent activation of opposite strand transcript (D) *Cdc42*, (E) *Xpo4*, and (F) *Max*. FPKM, locations of reLM-PCR junction fragments and fusion products are shown as described for Figure 1.

splicing elements, the availability and strength of native splice donors and acceptors, and the bifunctional capabilities of T2/Onc. Thus, direct visualization with RNA-seq is essential to understand the mechanistic effects of transposon insertion. Our analysis method will be of value beyond cancer applications as this methodology can be applied to any forward genetic screen that utilizes transcript-modifying transposon mutagenesis.

Methods

SB-mutagenized osteosarcomas

The generation and analysis of these tumors has been previously described (Moriarity et al. 2015). Briefly, mice undergoing transposon mutagenesis targeted to osteoblasts and their precursors, via *Osterix1-Cre* transgene, were generated on a WT or conditional LSL-*Trp53*^{R270H} background (Olive et al. 2004) and aged for osteosarcoma development. Upon necropsy, tumors were snap-frozen

in RNAlater (Sigma). reLM-PCR (Rahrmann et al. 2013) and shLM-PCR (Riordan et al. 2014) were performed as previously described. RNA was extracted using a RNA Midi extraction kit (Qiagen) and quality-tested using the Bioanalyzer 2100 (Agilent Technologies). The TruSeq RNA Sample Preparation kit (Illumina) was used to generate genomic sequences of ~200 bp, which were sequenced using the HiSeq 2000 (Illumina).

RNA-seq mapping

TopHat (Trapnell et al. 2009) was used to map paired-end sequences to the mm9 genome modified to include 2153 bases of the T2/Onc transposon sequence (Collier et al. 2005) as Chromosome chrSB. The BEDTools (Quinlan et al. 2010) function maskfasta was used to remove the *En2* gene, a portion of which was used to generate the T2/Onc transposon. The total numbers of mapped and unmapped sequences were obtained using the SAMtools (Li et al. 2009) function faidx. Fragments per kilobase

of exon per million fragments mapped (FPKM) values for each gene were generated using Cuffdiff (Trapnell et al. 2010). IGV (Thorvaldsdóttir et al. 2013) and R (R Core Team 2012) were used to visualize mapped sequence data.

Identification of transposon fusions from RNA-seq data

Mapped, sorted, indexed BAM files were used to identify paired-end sequences where one end mapped to an endogenous genomic location and the other end mapped to the T2/Onc sequence using the SAMtools view command. A Perl script was used to summarize T2/Onc fusions based on (1) the total number of reads that support the fusion, (2) the number of unique fusions, and (3) the total number of reads mapping to that location (Fusion_Finder.pl). The CPM for each fusion was calculated by dividing the number of reads that support the fusion by the total number of reads sequenced and multiplying the number by 1,000,000.

Annotation, CIS, and IMTs

The nearest gene was identified using the BEDTools function closest with mm9 annotation. The TAPDANCE CIS calling algorithm (Sarver et al. 2012) was utilized to identify and annotate (1) DCISs from the list of all LM-PCR insertions with no chromosome exclusion, and (2) RCISs from the list of all RNA fusions. For each tumor, increased IMTs were defined as genes that had FPKM levels more than 1.5 times the interquartile range above the third quartile and had a fusion. Tumors with decreased IMTs were defined as genes that had FPKM levels below the first quartile where the first quartile had a value greater than 1 and had a fusion.

Fusion-spanning RT-PCR

RNA was extracted from tumors using the RNA Mini kit (Invitrogen) and reverse-transcribed into cDNA using the SuperScript III reverse-transcriptase kit (Life Technologies), following the manufacturer's instructions. PCR was performed using 2× GoTaq (Promega) with an initial denaturation step at 95°C for 2 min, 35 cycles of denaturing at 95°C for 25 sec, annealing at 55°C for 35 sec, extension at 72°C for 1.5 min, and a final extension step at 72°C for 5 min. PCR amplicons were resolved on 1% agarose gels. PCR products were subcloned into PCR-Blunt II-TOPO (Life Technologies) following the manufacturer's instructions and subsequently sequenced using standard Sanger sequencing. Primer sequences can be found in Supplemental Table S9.

Data access

The sequencing data from this study have been submitted to the NCBI Sequence Read Archive (SRA; <http://www.ncbi.nlm.nih.gov/sra/>) under accession number SRP047432. The Fusion Finder script is included in the Supplemental Material and can also be obtained from http://www.oncomir.umn.edu/fusion_finder.

Acknowledgments

This research was supported by the Sobiech Osteosarcoma Fund Award, Children's Cancer Research Fund, American Cancer Center Research Professor Grant #123939 (to D.A.L.), National Cancer Institute, R01 CA113636 (to D.A.L.), and the comprehensive Masonic Cancer Center US National Institutes of Health grant P30 CA077598. We extend our thanks to the University of Minnesota resources involved in our project. The University of Minnesota Genomics Center provided services for RNA sequencing and the Minnesota Supercomputing Institute provided access to computational resources.

References

- Collier LS, Carlson CM, Ravimohan S, Dupuy AJ, Largaespada DA. 2005. Cancer gene discovery in solid tumours using transposon-based somatic mutagenesis in the mouse. *Nature* **436**: 272–276.
- Dromard M, Bompard G, Glondu-Lassis M, Puech C, Chalbos D, Freiss G. 2007. The putative tumor suppressor gene *PTPN13/PTPL1* induces apoptosis through insulin receptor substrate-1 dephosphorylation. *Cancer Res* **67**: 6806–6813.
- Egas-Bejar D, Anderson PM, Agarwal R, Corrales-Medina F, Devarajan E, Huh WW, Brown RE, Subbiah V. 2014. Theranostic profiling for actionable aberrations in advanced high risk osteosarcoma with aggressive biology reveals high molecular diversity: the human fingerprint hypothesis. *Oncoscience* **1**: 167–179.
- Hu YS, Pan Y, Li WH, Zhang Y, Li J, Ma BA. 2010. Association between TGFBR1*6A and osteosarcoma: a Chinese case-control study. *BMC Cancer* **10**: 169.
- Ivics Z, Hackett PB, Plasterk RH, Izsvák Z. 1997. Molecular reconstruction of *Sleeping Beauty*, a *Tc1*-like transposon from fish, and its transposition in human cells. *Cell* **91**: 501–510.
- Keng VW, Villanueva A, Chiang DY, Dupuy AJ, Ryan BJ, Matise I, Silverstein KA, Sarver A, Starr TK, Akagi K, et al. 2009. A conditional transposon-based insertional mutagenesis screen for genes associated with mouse hepatocellular carcinoma. *Nat Biotechnol* **27**: 264–274.
- Koudijs MJ, Klijn C, van der Weyden L, Kool J, ten Hoeve J, Sie D, Prasetyanti PR, Schut E, Kas S, Whipp T, et al. 2011. High-throughput semiquantitative analysis of insertional mutations in heterogeneous tumors. *Genome Res* **21**: 2181–2189.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**: 2078–2079.
- Mitra SK, Schlaepfer DD. 2006. Integrin-regulated FAK–Src signaling in normal and cancer cells. *Curr Opin Cell Biol* **18**: 516–523.
- Moriarty BS, Otto GM, Rahrmann EP, Rathe SK, Wolf NK, Weg MT, Manlove LA, LaRue RS, Temiz NA, Molyneux SD, et al. 2015. A *Sleeping Beauty* forward genetic screen identifies new genes and pathways driving osteosarcoma development and metastasis. *Nat Genet* **47**: 615–624.
- Morrison H, Sherman LS, Legg J, Banine F, Isacke C, Haipek CA, Gutmann DH, Ponta H, Herrlich P. 2001. The NF2 tumor suppressor gene product, merlin, mediates contact inhibition of growth through interactions with CD44. *Genes Dev* **15**: 968–980.
- Müller J, Ory S, Copeland T, Piwnicka-Worms H, Morrison DK. 2001. C-TAK1 regulates Ras signaling by phosphorylating the MAPK scaffold, KSR1. *Mol Cell* **8**: 983–993.
- Neer RM, Arnaud CD, Zanchetta JR, Prince R, Gaich GA, Reginster JY, Hodsmann AB, Eriksen EF, Ish-Shalom S, Genant HK, et al. 2001. Effect of parathyroid hormone (1–34) on fractures and bone mineral density in postmenopausal women with osteoporosis. *N Engl J Med* **344**: 1434–1441.
- Olive KP, Tuveson DA, Ruhe ZC, Yin B, Willis NA, Bronson RT, Crowley D, Jacks T. 2004. Mutant *p53* gain of function in two mouse models of Li-Fraumeni syndrome. *Cell* **119**: 847–860.
- Pérez-Mancera PA, Rust AG, van der Weyden L, Kristiansen G, Li A, Sarver AL, Silverstein KA, Grützmann R, Aust D, Rümmele P, et al. 2012. The deubiquitinase *USP9X* suppresses pancreatic ductal adenocarcinoma. *Nature* **486**: 266–270.
- Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841–842.
- R Core Team. 2012. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>.
- Rahrmann EP, Watson AL, Keng VW, Choi K, Moriarty BS, Beckmann DA, Wolf NK, Sarver A, Collins MH, Moertel CL, et al. 2013. Forward genetic screen for malignant peripheral nerve sheath tumor formation identifies new genes and pathways driving tumorigenesis. *Nat Genet* **45**: 756–766.
- Riordan JD, Drury LJ, Smith RP, Brett BT, Rogers LM, Scheetz TE, Dupuy AJ. 2014. Sequencing methods and datasets to improve functional interpretation of *sleeping beauty* mutagenesis screens. *BMC Genomics* **15**: 1150.
- Sarver AL, Erdman J, Starr T, Largaespada DA, Silverstein KAT. 2012. TAPDANCE: an automated tool to identify and annotate transposon insertion CISs and associations between CISs from next generation sequence data. *BMC Bioinformatics* **13**: 154.
- St Johnston D. 2002. The art and design of genetic screens: *Drosophila melanogaster*. *Nat Rev Genet* **3**: 176–188.
- Starr TK, Allaei R, Silverstein KA, Staggs RA, Sarver AL, Bergemann TL, Gupta M, O'Sullivan MG, Matise I, Dupuy AJ, et al. 2009. A transposon-based genetic screen in mice identifies genes altered in colorectal cancer. *Science* **323**: 1747–1750.

- Takahashi K, Mitsui K, Yamanaka S. 2003. Role of ERAs in promoting tumour-like properties in mouse embryonic stem cells. *Nature* **423**: 541–545.
- Thorvaldsdóttir H, Robinson JT, Mesirov JP. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* **14**: 178–192.
- Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**: 1105–1111.
- Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* **28**: 511–515.
- Uren AG, Kool J, Berns A, Van Lohuizen M. 2005. Retroviral insertional mutagenesis: past, present and future. *Oncogene* **24**: 7656–7672.
- Vahle JL, Sato M, Long GG, Young JK, Francis PC, Engelhardt JA, Westmore MS, Linda Y, Nold JB. 2002. Skeletal changes in rats given daily subcutaneous injections of recombinant human parathyroid hormone (1-34) for 2 years and relevance to human safety. *Toxicol Pathol* **30**: 312–321.
- Xu L, Corcoran RB, Welsh JW, Pennica D, Levine AJ. 2001. *WISP-1* is a Wnt-1- and β -catenin-responsive oncogene. *Genes Dev* **14**: 585–595.

Received December 17, 2014; accepted in revised form November 4, 2015.