

# Rapid, accurate mapping of transgene integration in viable rhesus macaque embryos using enhanced-specificity tagmentation-assisted PCR

Junghyun Ryu,<sup>1</sup> William Chan,<sup>1</sup> Jochen M. Wettengel,<sup>2,3</sup> Carol B. Hanna,<sup>1</sup> Benjamin J. Burwitz,<sup>3,4</sup> Jon D. Hennebold,<sup>1,5</sup> and Benjamin N. Bimber<sup>4,6</sup>

<sup>1</sup>Division of Reproductive & Developmental Sciences, Oregon National Primate Research Center, Oregon Health & Science University, Beaverton, OR 97006, USA;

<sup>2</sup>Institute of Virology, Technical University of Munich/Helmholtz Zentrum München, München, 81675 Germany; <sup>3</sup>Vaccine & Gene Therapy Institute, Oregon Health & Science University, Beaverton, OR 97006, USA; <sup>4</sup>Division of Pathobiology, Oregon National Primate Research Center, Oregon Health & Science University, Beaverton, OR 97006, USA; <sup>5</sup>Department of Obstetrics & Gynecology, Oregon Health & Science University, Portland, OR 97239, USA; <sup>6</sup>Division of Genetics, Oregon National Primate Research Center, Oregon Health & Science University, Beaverton, OR 97006, USA

Genome engineering is a powerful tool for *in vitro* research and the creation of novel model organisms and has growing clinical applications. Randomly integrating vectors, such as lentivirus- or transposase-based methods, are simple and easy to use but carry risks arising from insertional mutagenesis. Here we present enhanced-specificity tagmentation-assisted PCR (esTag-PCR), a rapid and accurate method for mapping transgene integration and copy number. Using stably transfected HepG2 cells, we demonstrate that esTag-PCR has higher integration site detection accuracy and efficiency than alternative tagmentation-based methods. Next, we performed esTag-PCR on rhesus macaque embryos derived from zygotes injected with piggyBac transposase and transposon/transgene plasmid. Using low-input trophoblast biopsies, we demonstrate that esTag-PCR accurately maps integration events while preserving blastocyst viability. We used these high-resolution data to evaluate the performance of piggyBac-mediated editing of rhesus macaque embryos, demonstrating that increased concentration of transposon/transgene plasmid can increase the fraction of embryos with stable integration; however, the number of integrations per embryo also increases, which may be problematic for some applications. Collectively, esTag-PCR represents an important improvement to the detection of transgene integration, provides a method to validate and screen edited embryos before implantation, and represents an important advance in the creation of transgenic animal models.

## INTRODUCTION

Randomly integrating gene delivery vectors, such as lentivirus- or transposase-based methods (i.e., piggyBac or Sleeping Beauty), are powerful and simple systems for genome engineering.<sup>1–4</sup> They are commonly used for *in vitro* or *in vivo* integration of genetic material into the host genome.<sup>5,6</sup> The integration of an exogenous gene can disrupt endogenous genes, causing unwanted effects.<sup>7</sup> In some settings, such as delivery of a reporter construct, controlling transgene

copy number is important.<sup>7</sup> The relative importance of insertional mutagenesis and copy number vary by application. For *in vitro* experiments with immortalized cells, the implications of unintended disruption of endogenous genes may be minimal; however, controlling unwanted effects may be critical when generating the progenitor for a transgenic animal model. When generating complex transgenic organisms, such as non-human primates, the time to sexual maturity, number of oocytes retrieved, and efficiency of successful pregnancy following implantation are all rate-limiting steps.<sup>8</sup> Thus, it is critical to validate proper transgene delivery or genomic editing of the embryo prior to transplantation and to minimize the risk for deleterious off-target mutations.

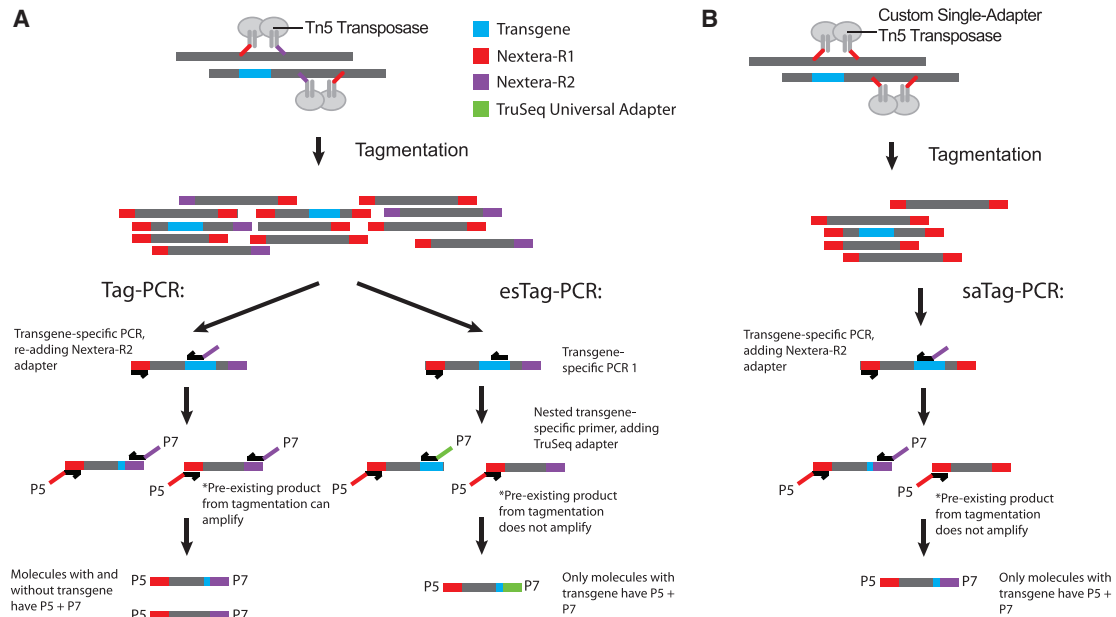
There are many methods for detecting integration events, each with advantages and disadvantages. Most are forms of PCR, including inverse PCR (iPCR), ligation-mediated PCR, and linear amplification PCR (LAM-PCR).<sup>9–14</sup> These methods each involve using restriction enzymes to digest the input DNA, which is then ligated to terminal adapter sequences. Therefore, these methods are limited to detecting integration events proximal to the chosen restriction enzyme(s).<sup>15</sup> A newer version of the method, termed nonrestrictive LAM-PCR (nrLAM-PCR), does not require the use of restriction enzymes and was reported to provide comprehensive mapping of integration events; however, this is a time-consuming protocol.<sup>16</sup> More recently, an alternative protocol was published that uses Tn5 transposase to randomly fragment the DNA (a process termed tagmentation) and add adapters in one step.<sup>17</sup> This protocol, termed tagmentation-assisted PCR (Tag-PCR), is streamlined relative to LAM-PCR; however, the amplification strategy of Tag-PCR

Received 21 August 2021; accepted 16 January 2022;  
<https://doi.org/10.1016/j.omtm.2022.01.009>

**Correspondence:** Benjamin N. Bimber, PhD, Division of Genetics, Oregon National Primate Research Center, Oregon Health & Science University, Beaverton, OR 97006, USA.

E-mail: [bimber@ohsu.edu](mailto:bimber@ohsu.edu)





**Figure 1. Overview of Tag-PCR and esTag-PCR designs**

(A) In both Tag-PCR and esTag-PCR, DNA is incubated with adapter-loaded transposomes, which fragment the DNA and add terminal adapters. Because the transposome can insert in either orientation, the resulting fragments can contain either Nextera-R1 + Nextera-R2, Nextera-R1 + Nextera-R1, or Nextera-R2 + Nextera-R2 adapters. In the original Tag-PCR protocol, PCR is performed using a transgene-specific primer (blue) and a primer complementary to one of the Nextera adapters (red). The transgene-specific primer is designed to add the other Nextera adapter (purple). This will specifically amplify fragments containing the transgene, creating molecules with Nextera-R1 and Nextera-R2 terminal adapters. However, the original pool of tagmentation products will contain non-specific fragments that also have Nextera-R1 and Nextera-R2 adapters from the initial fragmentation. A second round of PCR is performed using primers that bind the Nextera-R1 and Nextera-R2 sequences, adding P5 and P7 adapters. These primers can amplify both transgene-containing fragments and original tagmentation products, creating the potential for sequencing of non-specific fragments. The esTag-PCR protocol was designed to introduce several layers of specificity. The primary PCR is conducted using a transgene-specific primer, but no additional adapter is added. A second PCR is performed using a nested transgene-specific primer, which adds the P5 adapter and TruSeq + P7 adapters (a distinct Illumina-compatible sequencing adapter). (B) Alternative forms of tagmentation-assisted PCR are reported that use custom Tn5, loaded with two copies of a single terminal adapter. Similar to Tag-PCR, PCR is performed using a gene-specific primer and a primer targeting the Nextera-R1 adapter. A second round of PCR is performed using primers that bind the Nextera-R1 and Nextera-R2 sequences, adding P5 and P7 adapters.

can result in the amplification and sequencing of non-target genomic DNA fragments, reducing efficiency and lowering sensitivity. Other tagmentation-based methods have been published that rely on non-commercial transposomes, loaded with custom adapters.<sup>18–20</sup>

Here we present an alternative method for tagmentation-assisted PCR, which we term enhanced-specificity tagmentation-assisted PCR (esTag-PCR). By redesigning the PCR enrichment strategy, we demonstrate considerably lower off-target amplification and higher efficiency than alternative tagmentation-based protocols. We further show that this method can be applied to low-input DNA from trophectoderm (TE) biopsies of edited rhesus macaque embryos, providing a practical method to screen embryos prior to implantation.

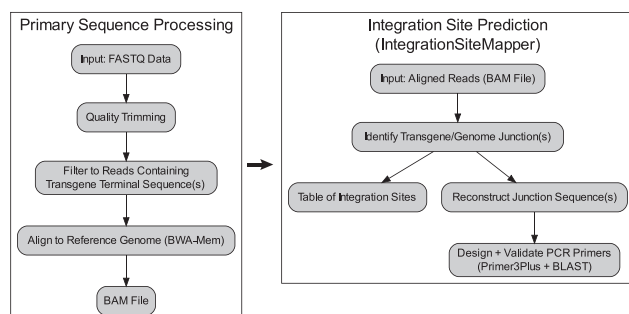
## RESULTS

### Design of esTag-PCR

A schematic of the design of esTag-PCR and comparison with two previously described protocols for tagmentation-based transgene

detection are shown in Figure 1. Tagmentation refers to the process by which Tn5 transposase is loaded with short adapter sequences, then used to randomly fragment DNA and ligate terminal adapters.<sup>21</sup> Commercially available transposases, such as Illumina Nextera (also sold as TDE1), are generally loaded with two distinct adapters. In the case of Nextera/TDE1, these adapters partially match the Illumina Read 1 and Read 2 sequencing primers. As Tn5 can insert in either orientation, this results in DNA fragments with R1/R1, R1/R2, R2/R1, or R2/R2 terminal adapters (illustrated in red and purple in Figure 1).

Tagmentation-assisted PCR first performs PCR on the tagged DNA using a primer that targets one of the terminal tagmentation adapters and a second transgene-specific primer.<sup>17</sup> The transgene-specific primer re-adds the second tagmentation adapter sequence. A second indexing PCR is then performed using primers targeting the two original tagmentation adapter sequences. These secondary primers add the Illumina P5 and P7 sequences. Although this will enrich for transgene-containing fragments, the indexing PCR will also amplify non-specific fragments left over from tagmentation,



**Figure 2. Overview of sequence analysis and integration site mapping**

Data analysis is divided into two main phases: pre-processing and DNA alignment, followed by integration site detection. In the first phase, raw sequence reads are trimmed using quality scores and filtered to retain only reads containing a sequence matching the transgene terminal end(s). The resulting reads are aligned to the host genome using BWA-Mem, creating a BAM file. In the second phase, IntegrationSiteMapper scans the aligned reads for the transgene terminal sequence(s) and determines integration orientation and genomic position of each event. A summary table of unique integration events is created. Additionally, this tool can reconstruct and output the sequence around each genome-transgene junction in the proper orientation. The tool can optionally use these sequences to design amplification primers to span each junction by running Primer3Plus and BLAST.

which do not contain the transgene. One possible mechanism to circumvent this unwanted amplification is to use custom transposomes, loaded with two copies of the same adapter (typically Nextera Read 1), as has been published in multiple studies.<sup>18–20</sup> Because DNA tagmented with these transposomes will contain only the R1 adapter, these methods will eliminate the potential for unwanted amplification of tagmented molecules that contain both R1 and R2 adapters. Although we are not aware of commercially available transposomes pre-loaded with single adapters, adapter-free Tn5 has recently become commercially available, such as Lucigen EZ-Tn5, which could be loaded with custom adapters. As a comparison for this study, we performed a version of Tag-PCR using single-adapter Tn5, which we term single-adapter Tag-PCR (saTag-PCR), that is most similar to TagMap or UDiTaS.<sup>20,22</sup>

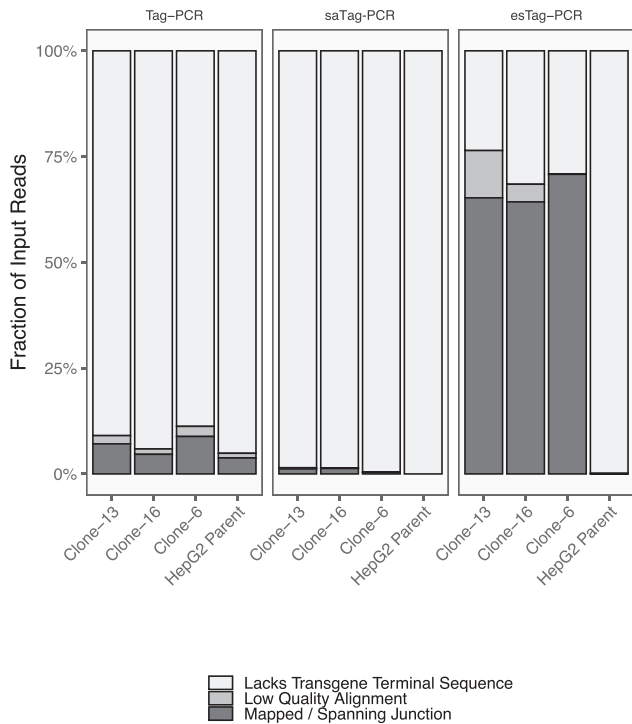
We designed esTag-PCR as an alternative approach to increase amplification specificity, while relying on commercially available reagents. In the esTag-PCR protocol, tagmentation is performed using commercial dual-adapter Tn5 transposase, followed by first-round PCR using a primer targeting one of the tagmentation adapters and a transgene-specific primer; however, this PCR does not re-add the original adapter sequence. A second nested PCR is performed using a separate internal transgene-specific primer and a primer targeting the tagmentation adapter. The transgene-specific primer adds a TruSeq adapter (an alternate Illumina-compatible sequencing primer). Because both PCR steps require transgene-specific priming, this scheme should considerably reduce the possibility of off-target amplification. In this study, all three methods are performed using a cocktail of gene-specific primers targeting both the 5' and 3' ends of the transgene, which allows detection and mapping of both ends of the integrated transgene.

## Data analysis and novel software

The detection of transgene integration into the genome is not a standard type of sequence analysis. Accurate mapping of transgene integration requires determining both the genomic location and orientation of the transgene, because both lentivirus and piggyBac transposase can integrate the transgene in either genomic orientation. Although in certain cases simply creating a table of integration sites is adequate, there are instances when independent validation of specific integration sites is needed. Our analysis scheme, described in detail in [materials and methods](#), is shown in [Figure 2](#). There are two general phases to the analysis: read filtering and alignment, followed by identification and scoring of predicted integration sites. For these analyses, we created two novel tools that are specific for the identification of transgene integration events. As part of pre-alignment processing, we created a utility to filter reads or read pairs on the basis of target sequences, using fuzzy matching (i.e., allowing a limited number of mismatches). For these analyses, we filter the input FASTQ data to retain only reads containing the terminal 15-mer from either end of the transgene, allowing up to 2 mismatches (the read pair is retained if either read contains the sequence). The passing reads, which span the genome-transgene junction (or represent non-integrated vector) are then aligned to the reference genome. We next created another novel tool, IntegrationSiteMapper, which encapsulates many of the steps needed for accurate mapping and validation of integration events. Using a customizable transgene definition (i.e., the sequence and orientation of the transgene terminal regions), the tool inspects the alignments to determine the location, orientation, and number of reads associated with each integration event. The tool also scans and reports reads matching the backbone of the transgene delivery vector, which is customizable, on the basis of the transgene definition. The latter can be useful for plasmid-based delivery systems, as non-integrated plasmid will be detected by esTag-PCR. The tool can optionally reconstruct and output the genome-transgene junctions as an annotated GenBank file and optionally design amplification primers to span each junction. Because this tool accepts a customizable transgene definition, it should be readily adaptable to any delivery system, as well as for mapping of other genomic elements such as transposons.

## Efficiency of esTag-PCR

We first sought to validate the esTag-PCR protocol and compare it with other published tagmentation-based protocols, including the previous-generation Tag-PCR protocol that relies on commercial dual-adapter Tn5 and a related protocol that uses custom Tn5 transposase loaded with a single adapter, termed saTag-PCR in this paper.<sup>20,22</sup> We transfected HepG2 cells with a piggyBac transposase expression plasmid and a piggyBac-compatible transposon/transgene plasmid expressing moxGFP-P2A-NanoLuc luciferase. A panel of clonally derived cell lines that stably express the transgene were established. Because piggyBac-mediated integration is random, each clonal line is expected to have a unique set of integration events. Clones were selected to encompass a range of moxGFP/NanoLuc expression levels. To compare efficiency and accuracy, we performed Tag-PCR, saTag-PCR, and esTag-PCR on these clones and the parent



**Figure 3. Efficiency of Tag-PCR versus esTag-PCR**

The piggyBac transposon system was used to generate HepG2 cells that stably express a *moxGFP/luciferase* expression construct. A panel of clonal cell lines were generated. Tag-PCR, saTag-PCR, and esTag-PCR were performed on these clonal lines, along with the parent HepG2 cells. The graph illustrates the fraction of sequence reads for each sample that fall into each of the following categories, indicated by the color legend: (1) reads lacking the terminal transgene sequence (likely PCR background), (2) reads containing this sequence but producing low-quality alignments, and (3) reads that span the transgene-genome junction. Relative to the other methods, esTag-PCR had a considerably higher fraction of on-target reads.

HepG2 cells (non-transfected). Figure 3 shows a summary of the resulting sequence reads. Reads are first categorized as to whether they contain the transgene-terminal sequence. Reads lacking this sequence likely represent non-specific background amplification. For all samples, esTag-PCR produced considerably less background than other methods, which is expected given the increased amount of gene-specific PCR amplification used in esTag-PCR (Figure 1). For each cell line, esTag-PCR produced 8–13 times more on-target reads (spanning the transgene-genome junction) than Tag-PCR or saTag-PCR. Collectively, this demonstrates superior transgene-specific amplification from esTag-PCR.

#### Accuracy of esTag-PCR

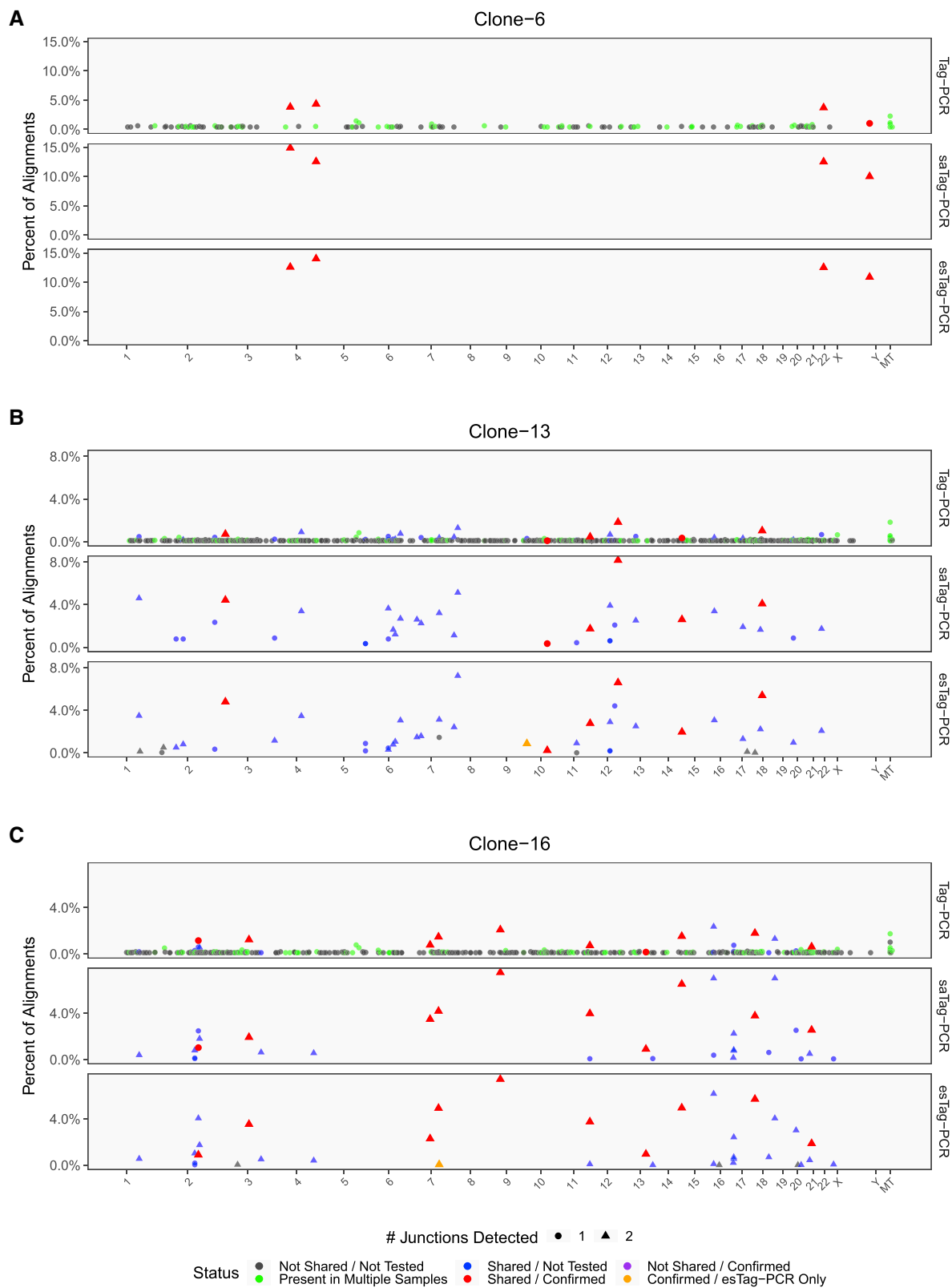
The integration events predicted by Tag-PCR, saTag-PCR, and esTag-PCR were then evaluated in the same clonal HepG2 lines. Each method returns a set of putative integration events (genomic location and orientation) and the number of associated reads. Figure 4 illustrates the results from each method, showing the proportion of reads matching each putative integration site (relative to total junction-

spanning reads). A site is reported if at least four reads are detected for this location, which is a relatively permissive threshold. Clone-6, which contained four predicted sites, had zero background in the esTag-PCR and saTag-PCR data. In contrast, while Tag-PCR also identified the same four sites at somewhat high frequency, it predicted 150 additional sites, most of which are present at low levels. Although the additional sites predicted by Tag-PCR could be removed by stricter filtering, the lowest true integration (10 reads) was only slightly above this threshold, and raising the filter threshold would increase the potential of filtering out true integration sites. For any given integration site, it is possible to detect the 5' terminal junction, 3' terminal junction, or both (Figure 4, triangles versus circles). Sites at which both junctions are detected are presumably of higher confidence, and this might be suitable as a filtering strategy. For the majority of sites predicted by esTag-PCR and saTag-PCR, both 3' and 5' junctions were detected (Figure 4, triangles). In contrast, Tag-PCR detected both junctions in only three of four integration sites for Clone-6 (a single transgene end was detected for chromosome X position 130,165,194), which suggests that detection of both junctions would not be a reliable filter criterion for Tag-PCR data.

Similar patterns were seen in Clone-13 and Clone-16, although each of these clones contained considerably more integration events. The concordance among esTag-PCR, saTag-PCR, and Tag-PCR was strong, with sites predicted by esTag-PCR also detected in saTag-PCR and Tag-PCR (Figure 4, blue icons); however, the Tag-PCR results contained significant low-frequency background, as before. Even though the set of predicted sites was nearly identical between esTag-PCR and saTag-PCR, esTag-PCR achieves this with considerably higher rates of transgene-containing reads, resulting in more efficient sequencing (Figure 3). As a result, the detection limit of esTag-PCR for a given sequencing depth should be greater. In fact, we identified a small number of sites in Clone-13 and Clone-16 predicted by esTag-PCR, but not detected using saTag-PCR (Figure 4, orange and gray dots in esTag-PCR panels). Although it is likely that with sufficient read depth, saTag-PCR would also detect these sites, this highlights the benefit of increased gene-specific amplification.

To validate predicted integration events, 26 sites that spanned a range of frequencies were selected for independent validation (Table S2). For each site, we designed PCR primers to amplify both 3' and 5' transgene-genome junctions. We performed PCR, Sanger-sequenced the amplicons, and used this sequence to validate the sequence of each genome-transgene junction. In every case tested, the site predicted by esTag-PCR was confirmed to match the predicted genomic location (Figure 4, red icons). Notably, esTag-PCR and saTag-PCR, but not Tag-PCR, predicted an integration in Clone-13 at chromosome 10, position 27,036,080, which was detected in 0.2% of reads. Despite the small number of supporting reads, this site was validated by Sanger sequencing, emphasizing the accuracy and lack of background achievable with either of these methods.

Many integration sites predicted by Tag-PCR, but not esTag-PCR, were identical across multiple clones (Figure 4, green dots). True



(legend on next page)



integration at the identical position in multiple samples is unlikely for a randomly integrating vector; however, if the technique produces experimental artifacts, such as amplification of fragments by mispriming, then specific genomic sites might be more prone to amplification. In fact, we found that 23% of sites predicted by Tag-PCR were detected in multiple clonal cell lines at the identical position, which is highly implausible for random integration. These sites included a cluster of relatively high-frequency predicted integration events in mitochondrial DNA (MT; positions 880 and 2,922). In all three clones, these predicted integration sites were present at a higher frequency than many true integrations, which underscores that filtering integrations on the basis of frequency is not adequate to differentiate true integration from noise with the original Tag-PCR protocol.

Collectively, these data demonstrate high accuracy of integration detection using esTag-PCR or saTag-PCR, with minimal background. Although the set of predicted integration events was nearly identical between esTag-PCR and saTag-PCR, the on-target efficiency of esTag-PCR was considerably higher. This might be expected because esTag-PCR contains two rounds of gene-specific nested PCR, in contrast to one round of gene-specific PCR in saTag-PCR. The higher efficiency should reduce the sequencing requirements for esTag-PCR and should increase the sensitivity of esTag-PCR over saTag-PCR for a given amount of raw sequence reads. Although it is possible that saTag-PCR efficiency could be improved with optimization, such as switching to two rounds of gene-specific amplification as with esTag-PCR, the fact that esTag-PCR can be performed using commercially available reagents represents an advantage.

#### Single-end versus paired-end sequencing for esTag-PCR

For these studies, esTag-PCR was performed using single-end 150 bp Illumina sequencing, which was selected over paired-end sequencing because Read 1 contains the complete transgene/genome junction, and single-end sequencing offers reduced cost and data size. Because paired-end sequencing should generate more sequence coverage over the flanking genomic region, it could improve the ability of the DNA aligner to uniquely place some reads within the genome and might increase detection or accuracy. To test this, we sequenced Clone-6, Clone-13, and Clone-16 using paired-end sequencing and then performed analysis using either Read 1 alone or both reads. The results were virtually identical (data not shown), which suggests that at least in these samples, the genomic sequence provided by single-end 150 bp reads was adequate for placement within the genome. We

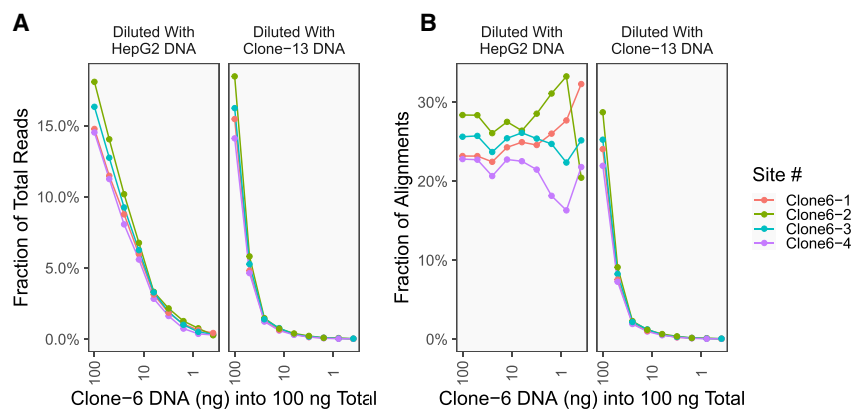
cannot rule out the possibility that integrations in certain genomic loci, such as highly repetitive or duplicated regions, would benefit from the additional information provided by paired-end data. Furthermore, as paired-end data would provide information about the length of the original tagged fragment, on the basis of the alignment position of the paired end, paired data can also provide information about the number of unique input molecules for each predicted integration site. This extra information could be especially useful for the evaluation of rare events or those with limited sequence read support.

#### Sensitivity of integration site detection

To evaluate the sensitivity of esTag-PCR, we performed a serial dilution experiment. DNA from Clone-6, which has four integration sites, was serially diluted using DNA from either non-transduced HepG2 cells or Clone-13, which has a large number of integrations. Each tagmentation reaction used a total of 100 ng input DNA, with the amount of Clone-6 DNA at each dilution step ranging from 100 to 0.2 ng. We performed esTag-PCR using each dilution step, then quantified the fraction of reads or alignments that matched each of the four Clone-6 integration sites (Figure 5). When the dilution was performed using HepG2 DNA, which lacks transgene, the fraction of total reads harboring the Clone-6 integration sites decreased with the decreased amount of Clone-6 DNA, although all four sites were detected at all dilution steps except the lowest (0.2 ng Clone-6 DNA into 100 ng total DNA). This mirrors Figure 3 and indicates that if there are not sufficient on-target molecules, only background sequences will be obtained. Nonetheless, even with very low input DNA from Clone-6, the fraction of aligned reads from each of the four integration sites was high and remained fairly stable across dilution steps. This indicates that a rare integration event in a background of wild-type cells, such as a cell population with low-efficiency transduction, might be detected. When Clone-6 DNA was diluted with Clone-13 DNA, a sample with many integration events that will compete with the Clone-6 integration events, detection of the Clone-6 events was reduced; however, we could detect reads from all four junctions at all dilution steps except the lowest. Unlike the dilution using HepG2 DNA, the fraction of alignments matching Clone-6 decreased substantially with each step, which is due to competition from the Clone-13 transgene-containing reads. This result is intuitive and suggests that if a given integration event is rare within a mixed population of transgene-containing cells, it will be more difficult to detect.

#### Figure 4. Accuracy of Tag-PCR versus esTag-PCR

Tag-PCR, saTag-PCR, and esTag-PCR were performed on a panel of HepG2 clonally derived cell lines, where each clone contains a unique set of piggyBac-mediated integration events. (A–C) Each graph illustrates the predicted integration events for one clone, displaying the results from Tag-PCR (top), saTag-PCR (middle), and esTag-PCR (bottom). The y-axis indicates the fraction of reads for each predicted site. Sites are colored on the basis of several categories. If the site is predicted in a given sample by two methods, it is categorized as “shared.” For a subset of sites, primers were designed to span both the predicted 5′ and 3′ junction sites, followed by PCR amplification of the transgene-genome junction and Sanger sequencing. All of these sites were confirmed to be true integrations (red dots). Sites detected in esTag-PCR alone and validated by PCR and Sanger sequencing are colored orange. Any remaining shared sites (for which sequence validation was not performed) are colored blue. Sites were further categorized as to whether a site was predicted at the identical genomic position in multiple samples, which is unlikely for a randomly integrating vector (green dots). Remaining sites, which are only predicted in one sample and by one method, are colored gray. The shape of the glyph indicates whether both 5′ and 3′ transgene-genome junctions were detected (triangles) or whether only a single junction was detected (circles).



**Figure 5. Sensitivity of esTag-PCR**

To test the sensitivity of esTag-PCR, DNA from Clone-6, which has 4 verified integration events, was serially diluted using DNA from either untransduced parent HepG2 cells or DNA from Clone-13, which has many integration events. Each tagmentation reaction used 100 ng input DNA, with the amount of Clone-6 DNA at each dilution step ranging from 100 to 0.2 ng. (A) The graphs display the fraction of total reads matching each of the four integration sites from Clone-6 (y-axis), relative to the amount of Clone-6 DNA (x-axis). (B) Similar graph as (A), except that the y-axis displays the fraction of alignments matching each of the four integration sites from Clone-6.

### Transgene expression relative to copy number

Because our panel of clonal HepG2 lines was generated by random integration, each clone contains a variable transgene copy number, with randomly spaced integration events. The transgene encodes both moxGFP and a secreted form of NanoLuc luciferase (Figure S1). This provides the opportunity to compare transgene expression level with copy number. Using a panel of eight clones, we measured luciferase expression relative to transgene copy number (Figure 6). Although there is a clear trend between copy number and luciferase levels, as might be expected, this pattern is absent for clones with a smaller transgene copy number (<5). This discrepancy is most likely due to unequal expression across integration events, based on the genomic context of each site. It is also possible that some integration events are partially silenced or otherwise not capable of expressing the reporter protein. There was no clear correlation between the locations of integration events relative to the nearest gene and expression level (data not shown). Irrespective of the cause, this demonstrates that reporter gene expression is not strictly correlated with copy number, at least with lower transgene copy numbers. Reporter gene expression by itself may also not be a reliable screening tool for integration, particularly if transgene copy number is a concern.

### Adaptation of esTag-PCR for lentiviral transduction

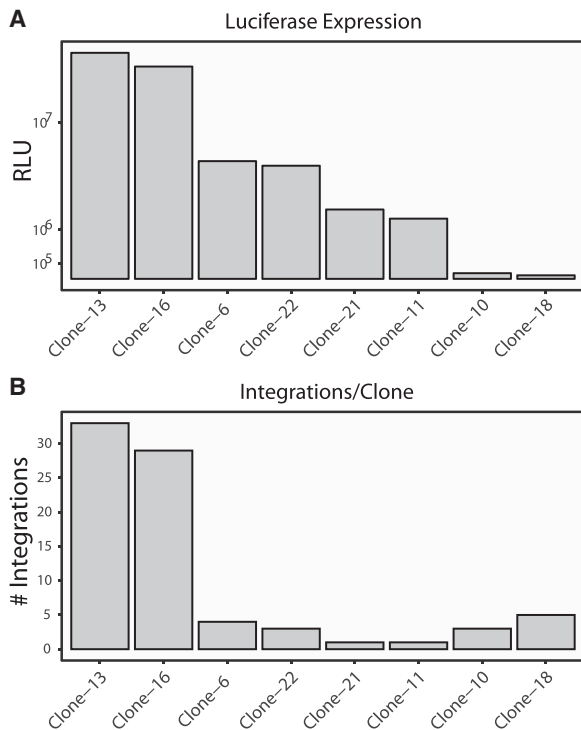
The esTag-PCR protocol can easily be adapted for any gene delivery system, provided that transgene-specific primers can be designed. To demonstrate applicability to lentiviral systems, we transduced HepG2, HEK293, and primary normal human dermal fibroblast (NHDF) cells with a lentiviral vector encoding the identical moxGFP-P2A-NanoLuc luciferase cassette as the prior piggyBac experiments (Figure S1). The sequences of the transgene-specific primers are listed in Table S1. Although lentiviral transduction efficiency was low in all cases, on the basis of direct visualization using fluorescent microscopy, moxGFP-expressing cells were detected. Bulk cells were collected, which represent a mosaic population with primarily non-transduced cells, and assessed using esTag-PCR. These data returned a high rate of on-target (genome-transgene-spanning) reads. Our analysis software was easily adapted to the structure of the

lentiviral transgene, and we successfully mapped integration sites in all cell types (Figure 7).

### Integration site mapping in transgenic macaque embryos

Because of the time, effort, and cost associated with generation and manipulation of transgenic embryos, accurate characterization of transgene integration sites, while preserving embryo viability, is critical. To demonstrate the suitability of esTag-PCR for screening transgenic embryos, we generated a panel of rhesus macaque blastocysts from zygotes that were injected with piggyBac transposase mRNA and a transgene/transposon plasmid (Figure 8A). Presumptive zygotes were injected with piggyBac mRNA (30 ng/μL) and transgene-encoding vector (30 ng/μL). TE biopsies were obtained from embryos that developed into blastocysts, with the biopsies being used for whole-genome amplification (WGA) and esTag-PCR. The results for two representative samples are shown in Figures 8B and 8C. Because esTag-PCR returns a high fraction of genome-transgene-spanning reads (in contrast to Tag-PCR or saTag-PCR), these data are presented as the fraction of reads per integration site relative to total sequence reads, as opposed to only junction-spanning reads. We further increased filtering stringency to require at least 10 reads from a given site. For E1118, two sites were predicted by esTag-PCR. Both of these were verified by PCR and Sanger sequencing of the vector-genome junctions (primers listed in Table S2). Importantly, these data demonstrate the ability of esTag-PCR to accurately detect integrations using small DNA input, while preserving the viability of the embryo itself.

We next sought to test whether the concentration of transgene/transposon vector used for injection would affect integration efficiency, measured using the transgene copy number per embryo. Presumptive zygotes were injected with piggyBac mRNA (30 ng/μL) and transgene/transposon plasmid at either 30 ng/μL (n = 175) or 100 ng/μL (n = 95). Injection volume was estimated to be 50–100 pL on the basis of pipette dimensions and calculated flow rate. As such, embryos injected with more concentrated transgene/transposon plasmid should receive more copies of the plasmid. The transgene used in these experiments constitutively expresses the red fluorescent protein mCherry.



**Figure 6. Transgene expression in clonal cell lines**

A panel of eight clonal HepG2 cells were generated, each stably expressing a moxGFP/luciferase expression construct, using piggyBac. (A) The graph displays relative luminescence for each clone. Values are background-subtracted relative to the parent HepG2 cells. (B) The graph shows the number of transgene integration events for each clone. RLU, relative luminescence unit.

Embryos were cultured for 7–9 days post-injection and sorted on the basis of red fluorescence. Because non-integrated plasmid can express mCherry, red fluorescence does not necessarily indicate stable transgene integration; however, fluorescence provides a screening method to identify successful injection and the potential for transgene integration. Embryos injected with more concentrated transgene/transposon plasmid had nearly a 2-fold increase in mCherry-positive embryos (11 embryos [6.2%] for 30 ng/ $\mu$ L and 13 embryos [13.7%] for 100 ng/ $\mu$ L). TE biopsy was performed on all mCherry-positive blastocysts, followed by esTag-PCR. A summary of the number of integrations per embryo is shown in Figure 8D. As might be predicted, injection with more concentrated transgene/transposon plasmid resulted in a higher fraction of embryos with stable transgene integration. It should be noted that the number of integrations per embryo was also higher when injected with more concentrated plasmid (mean 1.1 for 30 ng/ $\mu$ L versus 2.8 for 100 ng/ $\mu$ L). These data demonstrate that although increasing plasmid concentration will increase the number of embryos with stable transgene integration, if single integration or low copy number is required, then reduced plasmid concentration may be preferable. After injection, some embryos arrest before reaching the blastocyst stage. We performed esTag-PCR on 19 arrested embryos to compare transgene copy number relative to em-

bryos that developed into blastocysts (Figure 8E). On average, arrested embryos had more integrations/cell (mean 9.6 integrations).

If the purpose of gene delivery is to create the progenitor for a transgenic line, unintended consequences from transgene integration may be of great importance. A benefit of esTag-PCR data is that embryos can be screened on the basis of the location and putative impact of integrations. From the panel of macaque embryos, we categorized 443 piggyBac-mediated integration events according to their predicted effects on transcription and protein coding (Figure 8F). Most integration events occurred in intergenic regions (43.0%), although the next highest category was integration into an intron (41.8%). The two categories that might be predicted to have the greatest effect on protein expression, integration into an exon or immediately upstream of a gene (near enhancers and promoters), represented 2.8% and 4.2% of sites, respectively. Although the location of integrations mediated by piggyBac or lentivirus cannot be controlled, these esTag-PCR data demonstrate that embryos can be categorized on the basis of the location and number of integration events, providing a means to screen and select embryos for implantation.

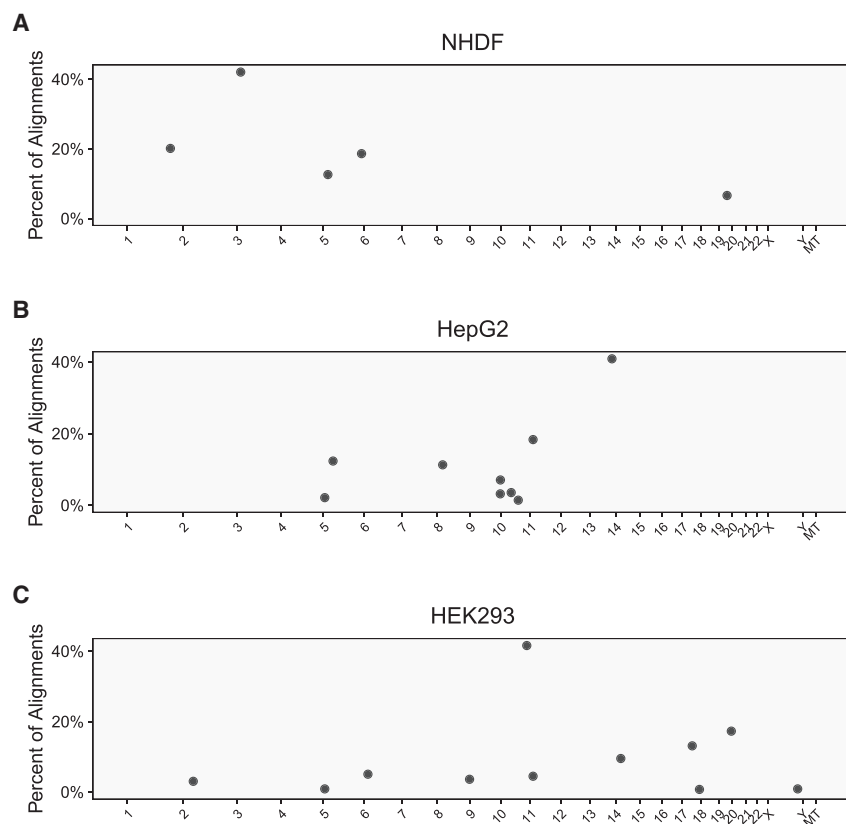
## DISCUSSION

Here we present a novel strategy for precise mapping of genomic integration sites. We demonstrate high efficiency and accuracy, with the ability to detect integration location, orientation, and copy number. By using two rounds of gene-specific PCR, esTag-PCR provides a high rate of on-target reads, thereby reducing sequence requirements, reducing cost per experiment, and providing easier discrimination between true integration events and false positives. We validated this method using piggyBac transposase and lentiviral delivery systems.

We further demonstrate that esTag-PCR provides an effective means to characterize and screen putative transgenic embryos. Following injection of piggyBac-based editing machinery, embryos are allowed to develop to the blastocyst stage, followed by a TE biopsy and vitrification of the embryo. We show that DNA amplified from TE cells can be used for esTag-PCR, providing the complete set and location of transgene integrations. Although piggyBac and lentiviral integration are quasi-random and can insert into genes, the data obtained by esTag-PCR can be used to exclude embryos with potentially disruptive transgene integrations, and prioritize which embryos are used for implantation.

One of the key advantages of esTag-PCR is the relative simplicity and speed of the protocol. Starting with purified DNA, the sequence-ready library can be created in 4–5 h. Because of the high on-target efficiency we demonstrate from esTag-PCR, fewer total sequence reads are required per sample. As such, it is practical to multiplex and sequence esTag-PCR libraries on small-format instruments, such as the Illumina MiSeq or iSeq, further improving data turnaround. Because freezing embryos prior to implantation is a common technique, it is therefore practical to bank and screen panels of putatively edited embryos. The results of esTag-PCR can be used to select optimal embryos for implantation, on the basis of the quantity and





**Figure 7. Adaptation of esTag-PCR for lentiviral transduced cells**

Primary normal human dermal fibroblasts (NHDF), HepG2, HEK293, and cells were each transduced with a lentiviral vector. We performed esTag-PCR on each heterogeneous population using primers specific to the lentiviral LTRs. (A–C) Graphs denote the integration sites detected in each cell population, indicated on the graph, with the y-axis indicating the fraction of sequence reads detected per site.

in the Tn5 adapters, as demonstrated by UDiTaS.<sup>22</sup> In this system, the Tn5 adapters each encode a unique molecular index. This sequence is incorporated once during the tagmentation step, thereby tagging each fragment. All molecules amplified from a given template molecule will share this UMI. The UMI can be used to collapse the resulting sequence reads on the basis of template molecule, thereby providing more accurate quantitation of each integration site. Finally, versions of this protocol have been published that employ a biotin-tagged oligos and a streptavidin pull-down to increase sensitivity.<sup>18</sup> Although a strength of esTag-PCR is the relative simplicity and ability to use 100% commercially available reagents, modifications such as these could enhance detection and might be advantageous in some applications.

location of integration events. Furthermore, because results from esTag-PCR could theoretically be obtained in less than 24 h, it may even be possible to screen samples while maintaining the embryo in culture, thereby avoiding vitrification and allowing transfer of fresh embryos, if desired.

Although we demonstrate that esTag-PCR can detect transgene integration mediated by either piggyBac or lentivirus, esTag-PCR can easily be adapted to other delivery systems, or completely different applications. For example, mapping of transposable elements or the integration of viruses (such as HIV/SIV) could be accomplished with esTag-PCR. Furthermore, the analysis software created for this study is powerful and highly adaptable. This software could also be used to analyze data generated by methods besides esTag-PCR.

Although not developed specifically for transgene detection, many tagmentation-based protocols have been published in recent years that could be adapted for transgene detection. For example, tagmentation-based tag integration site sequencing (TTISS) is a method to identify Cas-mediated cleavage sites.<sup>19</sup> Similar to GUIDE-seq, TTISS uses donor DNA to tag double-stranded DNA breaks. After tagging, tagmentation is performed using single-adaptor Tn5, followed by two rounds of nested PCR targeting the donor sequence. Another potentially useful modification is the use of unique molecular index (UMI)

Altogether, this study validates a novel method to accurately map integration events into the genome, with broad utility for *in vitro* and *in vivo* genome editing. The system is sensitive and versatile, with potential utility for detecting other genomic alterations, such as mapping transposable elements or endogenous retroviruses. We further show direct applicability to the generation of genome-edited non-human primate embryos, for which it enables rapid screening of embryos prior to implantation while preserving viability, providing a critical tool for the creation of novel transgenic animal models.

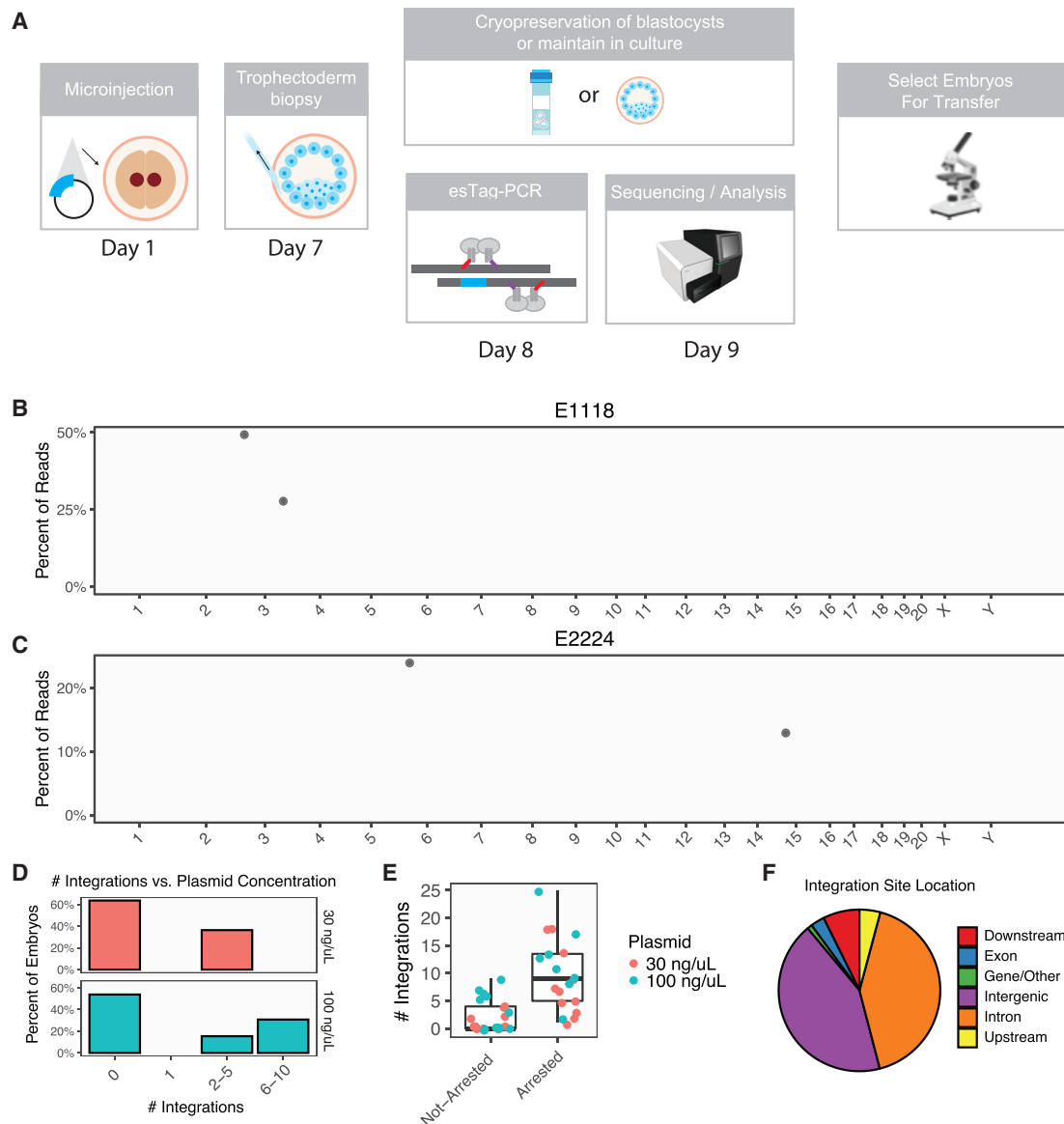
## MATERIALS AND METHODS

### Cell lines and cell culture

Experiments used HepG2 cells, an immortalized human hepatocyte cell line (ATCC HB-8065), and HEK293 cells, an immortalized human kidney cell line (American Type Culture Collection [ATCC] CRL-1573). Cells were cultured in DMEM-F12 supplemented with 10% fetal bovine serum (FBS) and 1% penicillin/streptomycin.

### piggyBac-mediated stable transfection of HepG2 cells and generation of clonal cell lines

HepG2 cells were transfected with a piggyBac-compatible vector containing an expression cassette that is under the control of the CAG promoter (GenBank: OK413188). This cassette allows the expression of moxGFP-NLS and NanoLuc luciferase linked by a P2A site, which is



**Figure 8. esTag-PCR using rhesus macaque embryos**

Rhesus macaque embryos were injected with a piggyBac-compatible plasmid and piggyBac RNA. (A) Schematic of transgene delivery and integration site mapping in rhesus macaque embryos. (B and C) Representative plots are shown for each of two embryos, illustrating the genomic location of transgene integration events, detected using esTag-PCR. (D) esTag-PCR was performed on TE biopsies from embryos injected with transposon/transgene plasmid at either 30 or 100 ng/ $\mu$ L. The graph illustrates the number of integration events obtained for each plasmid concentration, demonstrating that the number of integration events is reduced with lower plasmid concentration. (E) A boxplot is shown comparing the number of integrations per cell for embryos that arrested prior to reaching the blastocyst stage ( $n = 19$ ) compared with embryos that developed to the blastocyst stage ( $n = 24$ ). (F) The pie chart displays a summary of the location of integration events relative to genes. Sites are categorized as within an exon, within an intron, intergenic, downstream (within 5,000 bp of the transcription start site), upstream (within 5,000 bp of the transcript end), or gene/other (within the gene body, but none of the other categories).

followed by an internal ribosome entry site (IRES) region and a puromycin resistance gene (see plasmid map in Figure S1A). Cells were co-transfected with the transgene/transposon plasmid and a plasmid encoding codon-optimized hyperactive piggyBac transposase at a ratio of 3:1 (transposon/transposase).<sup>23</sup> Transfection was performed using Lipofectamine 3000 (L3000008; Thermo Fisher Scientific), according

to the manufacturer's instructions. After 3 days, cells were trypsinized and plated at low density. After an additional 3 days, cells were placed on puromycin selection. After 10–14 days, moxGFP<sup>+</sup> cell-single clusters were identified on fluorescent microscopy, and individual clusters were transferred to wells of a 96-well plate by detaching and aspirating them carefully with a 1 mL filtered tip. After 2 days of attachment, cell

colonies were broken up into single cells using trypsin. These clonal populations were further expanded and analyzed.

### Lentivirus production and lentiviral-mediated transduction

The identical moxGFP-P2A-NanoLuc cassette was cloned into a lentiviral vector (see plasmid map in [Figure S1B](#) and GenBank: OK413189). Lentiviral particles were generated by co-transfecting the plasmid with packaging plasmids according to published protocols.<sup>24</sup> Aliquots of lentiviral containing media were thawed and added directly to HepG2 cells in a 6-well plate. Expression of moxGFP was monitored using fluorescent microscopy. Normal human dermal fibroblast cells were obtained from ATCC (PCS-201-012).

### Genomic DNA extraction, WGA, and TE biopsies

For experiments with immortalized cells, genomic DNA (gDNA) was extracted using the GeneJET Genomic DNA Purification Kit (K0722; Thermo Fisher Scientific) according to the manufacturer's protocol. For experiments with rhesus macaque embryos, TE biopsies were performed (described below), followed by WGA using REPLiG (150345; Qiagen) according to the manufacturer's instructions. The WGA product was purified using Ampure XP beads (Beckman Coulter) at a bead-to-sample ratio of 0.9.

### Tagmentation-assisted PCR and single-adaptor tagmentation-assisted PCR

Tagmentation-assisted PCR was performed according to the published protocol.<sup>17</sup> A modified Tag-PCR protocol using custom Tn5, loaded with two copies of the Nextera Read 1 adapter, was used. Custom Tn5 was generated as previously described, loaded with Nextera Read 1 (TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG), and tagmentation was performed as described.<sup>12</sup> Briefly, tagmentation was performed with 100 ng gDNA in a tagmentation reaction that included 5.5  $\mu$ L NIB-HEPES, 2.5  $\mu$ L 4X TTD buffer, 2  $\mu$ L DNA, and 1.5  $\mu$ L Tn5. After preheating the thermocycler to 55°C, the reaction mixture was incubated for 10 min at 55°C, followed by a 10°C hold. After incubation, 1.15  $\mu$ L 10% SDS (1% final concentration) was added. The fragmented DNA was purified using Ampure XP beads with a bead-to-sample ratio of 0.8 and eluted into 30  $\mu$ L H<sub>2</sub>O. Following tagmentation, the first-round PCR was then performed using 25  $\mu$ L KAPA HiFi HotStart ReadyMix (KK2602; Kapa Biosystems), 1.25  $\mu$ L of each primer at 10  $\mu$ M (see [Table S1](#)), and 21.25  $\mu$ L tagmentation product. The following PCR conditions were used: 95°C for 3 min, followed by 30 cycles of 98°C for 30 s, 63°C for 30 s, and 72°C for 1 min, followed by 72°C for 10 min and a 4°C hold. Index PCR was performed using commercial Illumina Nextera XT indexes (product number FC-131-2001). PCR conditions were as follows: 12.5  $\mu$ L KAPA HiFi HotStart ReadyMix, 1.25  $\mu$ L of each index primer at 10  $\mu$ M, 2  $\mu$ L of the first-round PCR, and 8  $\mu$ L H<sub>2</sub>O. PCR conditions for the nested PCR were as follows: 98°C for 45 s, followed by 10 cycles of 98°C for 20 s, 54°C for 30 s, and 72°C for 20 s, ending with 72°C for 1 min and a 4°C hold. The PCR was cleaned using Ampure XP beads at a bead-to-sample ratio of 0.8. Purified samples were quantified using a Qubit Fluorometer (Invitrogen).

### Enhanced-specificity Tag-PCR

For esTag-PCR, the initial tagmentation step was performed with 100 ng gDNA or purified WGA product in a tagmentation reaction that included 25  $\mu$ L TD buffer, 2.5  $\mu$ L Illumina TDE1 enzyme, 100 ng DNA, and H<sub>2</sub>O to final volume of 50  $\mu$ L. After preheating the thermocycler to 58°C, the reaction mixture was incubated for 5 min at 58°C, followed by a 10°C hold. The fragmented DNA was purified using Ampure XP beads with a bead-to-sample ratio of 0.8 and eluted into 30  $\mu$ L H<sub>2</sub>O.

Following tagmentation, the first-round PCR was then performed using 25  $\mu$ L KAPA HiFi HotStart ReadyMix, 1.25  $\mu$ L of each primer at 10  $\mu$ M (see [Table S1](#)), and 21.25  $\mu$ L tagmentation product. The following PCR conditions were used: 95°C for 3 min, followed by 30 cycles of 98°C for 30 s, 63°C for 30 s, and 72°C for 1 min, followed by 72°C for 10 min and a 4°C hold. A secondary nested PCR was then performed using 12.5  $\mu$ L KAPA HiFi HotStart ReadyMix, 1.25  $\mu$ L index primer at 10  $\mu$ M, 0.75  $\mu$ L of each transgene-specific primer at 10  $\mu$ M (see [Table S1](#)), 2  $\mu$ L of the first-round PCR, and 7.5  $\mu$ L H<sub>2</sub>O. PCR conditions for the nested PCR were 98°C for 45 s, followed by 10 cycles of 98°C for 20 s, 54°C for 30 s, and 72°C for 20 s, ending with 72°C for 1 min and a 4°C hold. The second PCR adds a unique index to each sample. The PCR was cleaned using Ampure XP beads at a bead-to-sample ratio of 0.8. Purified samples were quantified using a Qubit Fluorometer. Of note, although esTag-PCR was performed in this study using a single Illumina index, the forward primer could be easily adapted to include an index between the P5 and Read 1 sequence. Using the primer PB-3TR-Inner-P5-TruSeq-R1 ([Table S1](#)) as an example, an i5 index could be inserted into the TruSeq universal adapter sequence as indicated: AATGATACGGCGACCACCGAGATCTACAC-[i5]-TCTTTCCCTACACGACGCTCTTCCGATCT-ATTTCAGAATGCATGCGTCA. Although this would require unique primers to be used for each sample in the second PCR, it would significantly increase the number of index combinations and the level of multiplexing that could be performed.

### Next-generation sequencing and data analysis

The sequence libraries from Tag-PCR and saTag-PCR were sequenced on an Illumina MiSeq instrument using paired-end 150 bp reads, while esTag-PCR libraries were sequenced using single-end 150 bp reads. The resulting FASTQ data were quality trimmed using Trimmomatic.<sup>25</sup> Next, reads were filtered to retain only reads or read pairs where at least one read contains the terminal 15-mer nucleotide sequence from either terminal end of the transgene, allowing an edit distance of 2 (based on Hamming distance). This was performed using the tool PrintReads-Containing, which we created for this project and distribute as part of the DISCVR-seq software package (<https://github.com/bimberlab/discvrseq>). For piggyBac experiments, the search sequences AGAC-TATCTTTCTAGGGTTAA, TTAACCCTAGAAAGATAGTCT, GA TTATCTTTCTAGGGTTAA, and TTAACCCTAGAAAGATAATC were used. For lentiviral experiments, the search sequences TGGAA GGGCTAATTCACCTCC, AGTGTGGAAAATCTCTAGCA, GGAGT GAATTAGCCCTTCCA, and TGCTAGAGATTTTCCACACT were used. The passing reads were aligned to the reference genome using

BWA-mem.<sup>26,27</sup> Human data were aligned to the GRCh38.p13 genome build (release 98, assembly ID GCA\_000001405.28), and rhesus macaque data were aligned to the MMul\_10 genome build (release 98, assembly ID GCA\_003339765.3). The location and orientation of integration events were then mapped using the tool IntegrationSiteMapper, which we created and made available as part of the DISCVr-seq software package (<https://github.com/bimberlab/discvrseq>). This tool inspects each alignment, creating a summary table of all unique integration events. Alignments are first grouped by read, and then filtered by mapping quality (MAPQ < 20). Next, alignments are scanned for the sequence of the transgene terminal region(s) (see Table S1) to identify the precise location and orientation of the genome-transgene junctions. Optionally, the tool can export an annotated GenBank-format file containing the reconstructed sequences of the transgene-genome regions. Optionally, it can also design amplification primers for each junction using Primer3Plus.<sup>28</sup> SnpEff was used to categorize the location of integration events, using National Center for Biotechnology Information (NCBI) gene annotations (build 103).<sup>29</sup> All raw sequence data have been submitted to the NCBI Sequence Read Archive (SRA) database under BioProject PRJNA750488. A mapping of sample to SRA ID is available in Table S3. Documentation for IntegrationSiteMapper can be found at [https://bimberlab.github.io/DISCVRSeq/toolDoc/com\\_github\\_discvrseq\\_walkers\\_tagpcr\\_IntegrationSiteMapper.html](https://bimberlab.github.io/DISCVRSeq/toolDoc/com_github_discvrseq_walkers_tagpcr_IntegrationSiteMapper.html).

#### PCR validation of transgene integration

A subset of integration events was selected for validation by PCR and Sanger sequencing. The list of sites, primers, and annealing temperatures is available in Table S2. For each integration event, PCR was performed in a 25  $\mu$ L PCR that included 12.5  $\mu$ L 2X KAPA HiFi HotStart ReadyMix, 10.7  $\mu$ L ultrapure water, 0.75  $\mu$ L of each 10  $\mu$ M forward or reverse primer, and 0.3  $\mu$ L of 100 ng/ $\mu$ L template gDNA. The cycling protocol was as follows: 95°C for 3 min, then 35 cycles of 98°C for 20 s, amplicon-specific annealing temperature for 15 s, 72°C for 30 s, followed by 72°C for 1 min, and a 4°C hold. Annealing temperatures were gradient-optimized and generally ranged between 68°C and 72°C (see Table S2). PCR products were imaged on a 2% TBE agarose gel. Single bands were PCR-purified using GeneJET PCR Purification Kit (K0702; Thermo Fisher Scientific) and Sanger-sequenced. If there were multiple bands, the band of the correct size was excised and gel-purified using PureLink Quick Gel Extraction Kit (K210012; Thermo Fisher Scientific) and Sanger-sequenced. The resulting sequence traces were analyzed using either SnapGene or Geneious software and mapped to the appropriate species-specific genome using NCBI BLAST+.<sup>30</sup>

#### Quantification of moxGFP and luciferase expression

Quantification of mean moxGFP fluorescence was performed using flow cytometry. Luciferase expression was monitored by a luminescence-based assay. For this assay, 10  $\mu$ L cell culture supernatant was plated on a white 96-well plate and combined with 100  $\mu$ L PBS-T (0.1% Tween 20) containing a 1:1,000 dilution of 1 mM Coelenterazine N stock solution (PJK Biotech) dissolved in acidified methanol. Luminescence was detected directly afterward using a VICTOR X Light Multilabel plate reader (PerkinElmer).

#### Animals

Rhesus macaques were socially housed at the Oregon National Primate Research Center (ONPRC) in animal biosafety level 2 rooms with autonomously controlled temperature, humidity, and lighting. Rhesus macaques were fed commercially prepared primate chow twice daily and received supplemental fresh fruit or vegetables daily. Fresh, potable water was provided via automatic water systems. Animal care and all experimental protocols and procedures were approved by the ONPRC Institutional Animal Care and Use Committee (IACUC). The ONPRC is a Category I facility. The Laboratory Animal Care and Use Program at the ONPRC is fully accredited by the American Association for Accreditation of Laboratory Animal Care and has an approved assurance (A3304-01) for the care and use of animals on file with the National Institutes of Health Office for Protection from Research Risks. ONPRC adheres to national guidelines established in the Animal Welfare Act (7 U.S. Code §§ 2131–2159) and the *Guide for the Care and Use of Laboratory Animals* (8th ed.), as mandated by U.S. Public Health Service policy.

#### Oocyte collection and *in vitro* fertilization

The ONPRC Assisted Reproductive Technologies (ART) Core provided gametes and performed *in vitro* fertilization (IVF) according to published protocols. Sexually mature rhesus monkeys with normal ovarian cyclicity were treated with a standard 10 day controlled ovarian stimulation cycle regimen as previously described<sup>31</sup> to produce multiple pre-ovulatory follicles containing mature ova. Prior to ovulatory events, ultrasound-guided percutaneous follicle aspiration was performed, and recovered oocytes were isolated into warmed TALP-HEPES containing 0.3% BSA with 5 IU/mL heparin. Cumulus enclosed metaphase II (MII) ova were washed into pre-equilibrated 100  $\mu$ L drops of BO-IVF (IVF Bioscience, Falmouth, Cornwall, UK) under oil with 5 ova/drop and incubated for approximately 4 h at 37°C in humidified 5% CO<sub>2</sub> in air until insemination. Rhesus macaque semen was collected from a male at the same day of oocyte collection and the sperm washed in warmed TALP-HEPES to a final concentration of 20 million sperm/mL.<sup>32</sup> A standard *in vitro* fertilization protocol was followed with sperm activated by 1 mM caffeine + 1 mM dibutyryl-cAMP for 15 min before adding 10  $\mu$ L to each culture drops containing ova.<sup>31</sup>

#### Injection of plasmid and piggyBac mRNA into rhesus macaque zygotes

Approximately 14 h post-insemination, presumptive zygotes were washed to remove sperm attached to the zona pellucida and transferred into a warmed TALP-HEPES drop covered with oil. Injection material containing 30 or 100 ng/ $\mu$ L piggyBac-compatible transgene plasmid, along with 30 ng/ $\mu$ L of piggyBac mRNA (Hera BioLabs) was back-loaded into a glass-pulled micro-injection pipette. Zygotes were stabilized by gentle suction onto a glass holding pipette (Cooper Surgical), and material was injected under continuous positive flow into the cytoplasm of zygotes, using a Narishige micro-manipulator (Narishige International, Amityville, NY). Injected embryos were then transferred into embryo culture media (BO-IVC; IVF Bioscience),



covered with oil, and cultured in a 6%, 5%, 89% mixture of CO<sub>2</sub>, O<sub>2</sub>, and N<sub>2</sub> at 37°C in humidified air, respectively.

### TE cell biopsy from piggyBac-injected blastocysts

Once embryos developed into expanded blastocysts, between days 7 and 9 post-insemination, a TE biopsy was performed as previously described.<sup>33</sup> Briefly, individual blastocysts were placed in a 20 µL drop of warmed TALP-HEPES under oil and stabilized with a holding pipette. An objective-mounted laser (ZYRCOS, Hamilton Thorne, Beverly, MA) was used to create a rent in the zona pellucida, large enough to pass the glass biopsy pipette (Cooper Surgical) to aspirate TE cells following laser dissection of 10–15 cells. Biopsied TE cells were transferred into a PCR tube to conduct WGA and to identify the integration of plasmid. Embryos are subjectively assessed before and during biopsy and graded as poor quality, medium quality, or good quality. Embryos are subjectively assessed again after thawing, mainly on the basis of re-expansion of the blastocoel. The biopsied blastocyst was then vitrified in a 0.25 mL straw (IMV Technologies, Maple Grove, MN) using the DMSO Blastocyst Vitrification Kit (Life-Global Group, Guilford, CT) according to the manufacturer's guidelines. Embryos that arrested during mitosis and failed to develop to the blastocyst stage underwent complete zona pellucida removal by brief exposure to acidic Tyrode's solution (prepared in house according to Ramsey and Hanna<sup>31</sup>) and were stored for subsequent analysis of transgene integration into the genome.

### SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.omtm.2022.01.009>.

### ACKNOWLEDGMENTS

We thank the laboratory of Dr. Andrew Adey for providing single-index Tn5 transposase. This work was supported by National Institutes of Health Somatic Cell Genome Editing grant U24OD026631 (to J.D.H.) and the Office of the Director, National Institutes of Health, award P51OD011092 to the Oregon National Primate Research Center. It would not be possible without the Oregon National Primate Research Center Division of Comparative Medicine, the Endocrinology Services Core, and the Assisted Reproductive Core. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

### AUTHOR CONTRIBUTIONS

J.R., W.C., J.M.W., and C.B.H. conducted experiments. B.N.B. drafted the manuscript, and all authors contributed to editing. B.N.B., B.J.B., and J.D.H. conceptualized the experiments and acquired funding.

### DECLARATION OF INTERESTS

The authors declare no competing interests.

### REFERENCES

- Sandoval-Villegas, N., Nurieva, W., Amberger, M., and Ivics, Z. (2021). Contemporary transposon tools: a review and guide through mechanisms and applications of sleeping beauty, piggyBac and Tol2 for genome engineering. *Int. J. Mol. Sci.* 22. <https://doi.org/10.3390/ijms22105084>.
- Kebrlari, P., Izsvak, Z., Narayanavari, S.A., Singh, H., and Ivics, Z. (2017). Gene therapy with the sleeping beauty transposon system. *Trends Genet.* 33, 852–870. <https://doi.org/10.1016/j.tig.2017.08.008>.
- Anguela, X.M., and High, K.A. (2019). Entering the modern era of gene therapy. *Annu. Rev. Med.* 70, 273–288. <https://doi.org/10.1146/annurev-med-012017-043332>.
- Sakuma, T., Barry, M.A., and Ikeda, Y. (2012). Lentiviral vectors: basic to translational. *Biochem. J.* 443, 603–618. <https://doi.org/10.1042/BJ20120146>.
- Park, F. (2007). Lentiviral vectors: are they the future of animal transgenesis? *Physiol. Genomics* 31, 159–173. <https://doi.org/10.1152/physiolgenomics.00069.2007>.
- Blesch, A. (2004). Lentiviral and MLV based retroviral vectors for ex vivo and in vivo gene transfer. *Methods* 33, 164–172. <https://doi.org/10.1016/j.ymeth.2003.11.005>.
- Cain-Hom, C., Splinter, E., van Min, M., Simonis, M., van de Heijning, M., Martinez, M., Asghari, V., Cox, J.C., and Warming, S. (2017). Efficient mapping of transgene integration sites and local structural changes in Cre transgenic mice using targeted locus amplification. *Nucleic Acids Res.* 45, e62. <https://doi.org/10.1093/nar/gkw1329>.
- Bavister, B.D., Boatman, D.E., Collins, K., Dierschke, D.J., and Eisele, S.G. (1984). Birth of rhesus monkey infant after in vitro fertilization and nonsurgical embryo transfer. *Proc. Natl. Acad. Sci. U S A* 81, 2218–2222. <https://doi.org/10.1073/pnas.81.7.2218>.
- Uemura, S., Nagaoka, T., Yokoyama, M., Igarashi, M., and Kishi, M. (2014). A simple and highly efficient method to identify the integration site of a transgene in the animal genome. *Neurosci. Res.* 80, 91–94. <https://doi.org/10.1016/j.neures.2013.11.007>.
- Mueller, P.R., and Wold, B. (1989). In vivo footprinting of a muscle specific enhancer by ligation mediated PCR. *Science* 246, 780–786. <https://doi.org/10.1126/science.2814500>.
- Schmidt, M., Schwarzwald, K., Bartholomae, C., Zaoui, K., Ball, C., Pilz, I., Braun, S., Glimm, H., and von Kalle, C. (2007). High-resolution insertion-site analysis by linear amplification-mediated PCR (LAM-PCR). *Nat. Methods* 4, 1051–1057. <https://doi.org/10.1038/nmeth1103>.
- Picelli, S., Bjorklund, A.K., Reinius, B., Sagasser, S., Winberg, G., and Sandberg, R. (2014). Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res.* 24, 2033–2040. <https://doi.org/10.1101/gr.177881.114>.
- Ochman, H., Gerber, A.S., and Hartl, D.L. (1988). Genetic applications of an inverse polymerase chain reaction. *Genetics* 120, 621–623. <https://doi.org/10.1093/genetics/120.3.621>.
- Ochman, H., Ajioka, J.W., Garza, D., and Hartl, D.L. (1990). Inverse polymerase chain reaction. *Biotechnology (N Y)* 8, 759–760. <https://doi.org/10.1038/nbt0890-759>.
- Harkey, M.A., Kaul, R., Jacobs, M.A., Kurre, P., Bovee, D., Levy, R., and Blau, C.A. (2007). Multiarm high-throughput integration site detection: limitations of LAM-PCR technology and optimization for clonal analysis. *Stem Cells Dev.* 16, 381–392. <https://doi.org/10.1089/scd.2007.0015>.
- Paruzynski, A., Arens, A., Gabriel, R., Bartholomae, C.C., Scholz, S., Wang, W., Wolf, S., Glimm, H., Schmidt, M., and von Kalle, C. (2010). Genome-wide high-throughput integrome analyses by nrLAM-PCR and next-generation sequencing. *Nat. Protoc.* 5, 1379–1395. <https://doi.org/10.1038/nprot.2010.87>.
- Hamada, M., Nishio, N., Okuno, Y., Suzuki, S., Kawashima, N., Muramatsu, H., Tsubota, S., Wilson, M.H., Morita, D., Kataoka, S., et al. (2018). Integration mapping of piggyBac-mediated CD19 chimeric antigen receptor T cells analyzed by novel tagmentation-assisted PCR. *EBioMedicine* 34, 18–26. <https://doi.org/10.1016/j.ebiom.2018.07.008>.
- Fueller, J., Herbst, K., Meurer, M., Gubicza, K., Kurtulmus, B., Knopf, J.D., Kirmmaier, D., Buchmuller, B.C., Pereira, G., Lemberg, M.K., and Knop, M. (2020). CRISPR-Cas12a-assisted PCR tagging of mammalian genes. *J. Cell Biol.* 219. <https://doi.org/10.1083/jcb.201910210>.
- Schmid-Burgk, J.L., Gao, L., Li, D., Gardner, Z., Strecker, J., Lash, B., and Zhang, F. (2020). Highly parallel profiling of Cas9 variant specificity. *Mol. Cell* 78, 794–800.e8. <https://doi.org/10.1016/j.molcel.2020.02.023>.



20. Stern, D.L. (2017). Tagmentation-based mapping (TagMap) of mobile DNA genomic insertion sites. *bioRxiv*, 037762. <https://doi.org/10.1101/037762>.
21. Adey, A., Morrison, H.G., Asan, X., Kitzman, J.O., Turner, E.H., Stackhouse, B., MacKenzie, A.P., Caruccio, N.C., Zhang, X., and Shendure, J. (2010). Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. *Genome Biol.* 11, R119. <https://doi.org/10.1186/gb-2010-11-12-r119>.
22. Giannoukos, G., Ciulla, D.M., Marco, E., Abdulkarim, H.S., Barrera, L.A., Bothmer, A., Dhanapal, V., Gloskowski, S.W., Jayaram, H., Maeder, M.L., et al. (2018). UDiTaS, a genome editing detection method for indels and genome rearrangements. *BMC Genomics* 19, 212. <https://doi.org/10.1186/s12864-018-4561-9>.
23. Cadinanos, J., and Bradley, A. (2007). Generation of an inducible and optimized piggyBac transposon system. *Nucleic Acids Res.* 35, e87. <https://doi.org/10.1093/nar/gkm446>.
24. Lever, A.M. (2000). HIV RNA packaging and lentivirus-based vectors. *Adv. Pharmacol.* 48, 1–28. [https://doi.org/10.1016/s1054-3589\(00\)48002-6](https://doi.org/10.1016/s1054-3589(00)48002-6).
25. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
26. Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26, 589–595. <https://doi.org/10.1093/bioinformatics/btp698>.
27. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>.
28. Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R., and Leunissen, J.A. (2007). Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res.* 35, W71–W74. <https://doi.org/10.1093/nar/gkm306>.
29. Cingolani, P., Platts, A., Wang le, L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., and Ruden, D.M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* 6, 80–92. <https://doi.org/10.4161/fly.19695>.
30. Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics* 10, 421. <https://doi.org/10.1186/1471-2105-10-421>.
31. Ramsey, C., and Hanna, C. (2019). In vitro culture of rhesus macaque (*Macaca mulatta*) embryos. *Methods Mol. Biol.* 2006, 341–353. [https://doi.org/10.1007/978-1-4939-9566-0\\_23](https://doi.org/10.1007/978-1-4939-9566-0_23).
32. Houser, L.A., Ramsey, C., de Carvalho, F.M., Kolwitz, B., Naito, C., Coleman, K., and Hanna, C.B. (2021). Improved training and semen collection outcomes using the closed box chair for macaques. *Animals (Basel)* 11. <https://doi.org/10.3390/ani11082384>.
33. McArthur, S.J., Leigh, D., Marshall, J.T., de Boer, K.A., and Jansen, R.P. (2005). Pregnancies and live births after trophoctoderm biopsy and preimplantation genetic testing of human blastocysts. *Fertil. Steril* 84, 1628–1636. <https://doi.org/10.1016/j.fertnstert.2005.05.063>.

**Supplemental information**

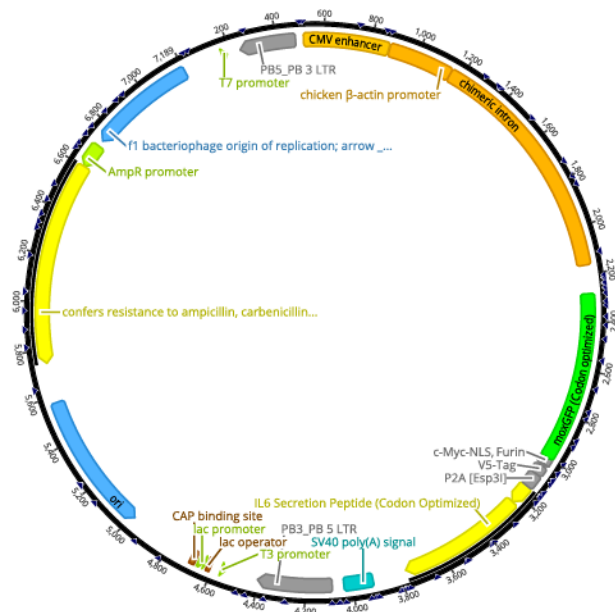
**Rapid, accurate mapping of transgene integration  
in viable rhesus macaque embryos using  
enhanced-specificity tagmentation-assisted PCR**

**Junghyun Ryu, William Chan, Jochen M. Wettengel, Carol B. Hanna, Benjamin J. Burwitz, Jon D. Hennebold, and Benjamin N. Bimber**

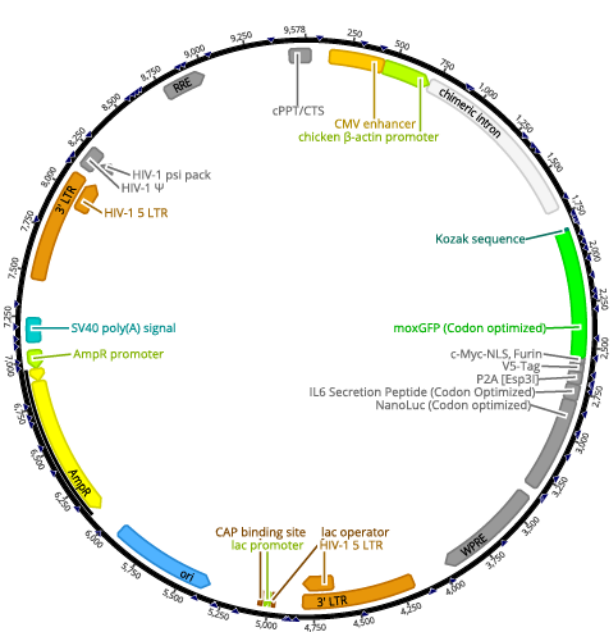
Supplemental Text

**Supplemental Figure 1. A)** Vector map of plasmid pB-CAG\_moxGFP-P2A-NanoLuc (#20). **B)** Vector map of plasmid pLV-CAG\_moxGFP-P2A-NanoLuc (#53).

A. Plasmid map of pB-CAG\_moxGFP-P2A-NanoLuc (#020)



B. Plasmid map of pLV-CAG\_moxGFP-P2A-NanoLuc (#053)



**Supplemental Table I.** Table of primers used for esTag-PCR amplification.

Name	Assay	Purpose	Sequence
PB-3TR-Outer	esTag-PCR/piggyBac	piggyBac Integration, PCR1	GCGACGGATTTCGCGCTATTT
PB-5TR-Outer	esTag-PCR/piggyBac	piggyBac Integration, PCR1	GACCGATAAAACACATGCGTCA
LV-3TR-Outer	esTag-PCR/Lentivirus	Lentiviral Integration, PCR1	GAGAGCTGCATCCGGAGTAC
LV-5TR-Outer	esTag-PCR/Lentivirus	Lentiviral Integration, PCR1	TCCTCTGGTTTCCCTTTTCGC
PB-3TR-Inner-P5-TruSeq-R1	esTag-PCR/piggyBac	piggyBac Integration, PCR2	AATGATACGGCGACCAACCGAGATCTACACTCTTTCCCTACACGACGCTCTCCGATCT ATTTCAAGAAATGCATGCGTCA
PB-5TR-Inner-P5-TruSeq-R1	esTag-PCR/piggyBac	piggyBac Integration, PCR2	AATGATACGGCGACCAACCGAGATCTACACTCTTTCCCTACACGACGCTCTCCGATCT CACATGATTATCTTTAACGTACGTCAC
LV-3TR-Inner-P5-TruSeq-R1	esTag-PCR/Lentivirus	Lentiviral Integration, PCR2	AATGATACGGCGACCAACCGAGATCTACACTCTTTCCCTACACGACGCTCTCCGATCT TAGTGTGTGCCCCGTCTGTTG
LV-5TR-Inner-P5-TruSeq-R1	esTag-PCR/Lentivirus	Lentiviral Integration, PCR2	AATGATACGGCGACCAACCGAGATCTACACTCTTTCCCTACACGACGCTCTCCGATCT AAGCAGTGGGTTCCCTAGTT
P7-N721-NexteraR2	esTag-PCR	Nested/Index PCR	CAAGCAGAAGACGGCATAACGAGAT GCAGCGTA GTCTCGTGGGCTCGGAGATG
P7-N722-NexteraR2	esTag-PCR	Nested/Index PCR	CAAGCAGAAGACGGCATAACGAGAT CTGCGCAT GTCTCGTGGGCTCGGAGATG
P7-N723-NexteraR2	esTag-PCR	Nested/Index PCR	CAAGCAGAAGACGGCATAACGAGAT GAGCGCTA GTCTCGTGGGCTCGGAGATG
P7-N724-NexteraR2	esTag-PCR	Nested/Index PCR	CAAGCAGAAGACGGCATAACGAGAT CGCTCAGT GTCTCGTGGGCTCGGAGATG
P7-N726-NexteraR2	esTag-PCR	Nested/Index PCR	CAAGCAGAAGACGGCATAACGAGAT GTCTTAGG GTCTCGTGGGCTCGGAGATG
P7-N727-NexteraR2	esTag-PCR	Nested/Index PCR	CAAGCAGAAGACGGCATAACGAGAT ACTGATCG GTCTCGTGGGCTCGGAGATG
P7-N728-NexteraR2	esTag-PCR	Nested/Index PCR	CAAGCAGAAGACGGCATAACGAGAT TAGCTGCA GTCTCGTGGGCTCGGAGATG
P7-N729-NexteraR2	esTag-PCR	Nested/Index PCR	CAAGCAGAAGACGGCATAACGAGAT GACGTGGA GTCTCGTGGGCTCGGAGATG
R2-Stub-PB-3TR-Outer	Tn5/Anchor-Seq	piggyBac Integration, PCR1	GTCTCGTGGGCTCGGAGATG-GCGACGGATTTCGCGCTATTT
R2-Stub-PB-5TR-Outer	Tn5/Anchor-Seq	piggyBac Integration, PCR1	GTCTCGTGGGCTCGGAGATG-GACCGATAAAACACATGCGTCA
NTSR1-F1	Tn5/Anchor-Seq	piggyBac Integration, PCR1	TCGTCGGCAGCGTCAGAT

**Supplemental Table II.** Table of vector-genome junctions selected for PCR amplification and Sanger sequence validation. For each junction, the PCR amplification primers are listed, along with the PCR annealing temperature.

Sample	Chr	Position	Orientation	JunctionName	Primer-F-Name	Primer-F-Seq	Primer-R-Name	Primer-R-Seq	Expected Size	Annealing Temp	Status
Clone-6	3	171,771,812	+	PB-3TR	Tag-PCR-21	CATTGACAAGCAGCCCTCAGC	Tag-PCR-13	ATCCTCTTGAGCCCTGACCT	862	70	Verified Integration
Clone-6	3	171,771,812	+	PB-5TR	Tag-PCR-05	AAGGGGATACAGCTCGCCCT	Tag-PCR-63	AGACCGATAAAACACATGGGTCA	269	68	Verified Integration
Clone-6	4	77,863,325	+	PB-3TR	Tag-PCR-21	CATTGACAAGCAGCCCTCAGC	Tag-PCR-49	TGCACAAGAAAGGGCTGTCA	279	65-68	Verified Integration
Clone-6	4	77,863,325	+	PB-5TR	Tag-PCR-19	CATTCTGGAGAGCTGCACCC	Tag-PCR-5TR	CAAAATTAATAAAACCTCGATATACAGACCGATAAAACACAT	375	70	Verified Integration
Clone-6	21	41,732,490	-	PB-3TR	Tag-PCR-24	CCACCAGCAGACCCCAACA	Tag-PCR-3TR	GCATGCGTCAATTTTACGCAGACTATCTTTCTAG	451	70	Verified Integration
Clone-6	21	41,732,490	-	PB-5TR	Tag-PCR-5TR	CAAAATTAATAAAACCTCGATATACAGACCGATAAAACACATG	Tag-PCR-54	TGCTTTACAGTGGAGGTGCAA	221	68	Verified Integration
Clone-6	X	130,165,194	+	PB-3TR	Tag-PCR-22	CATTTTGACTCAGCGCGTCTG	Tag-PCR-27	CCTTTTCCCGGAGCGTT	329	68	Verified Integration
Clone-6	X	130,165,194	+	PB-5TR	Tag-PCR-23	CCACCAGAAGTACACGGCT	Tag-PCR-62	ACACATGATTAATCTTTAACGTACGTAC	272	62	Verified Integration
Clone-16	2	42,485,324	-	PB-3TR	Tag-PCR-44	TCACGACAGACACAAAGC	Tag-PCR-34	GCATTTGACAGCAGCGCTCAC	290	66	Verified Integration
Clone-16	2	42,485,324	-	PB-5TR	Tag-PCR-63	AGACCCGATAAAACACATCGCTCA	Tag-PCR-02	AAATGCTCTTTGCTGGCAGC	210	68	Verified Integration
Clone-16	6	167,123,583	-	PB-3TR	Tag-PCR-50	TGCACCTGGAGCTTCGGTG	Tag-PCR-29	CTCAGGGGAGCTCCAAGC	262	75	Verified Integration
Clone-16	6	167,123,583	-	PB-5TR	Tag-PCR-5TR	CAAAATTAATAAAACCTCGATATACAGACCGATAAAACACATG	Tag-PCR-25	CCCTCACAAAGCTCTTCTCTG	214	70	Verified Integration
Clone-16	7	30,289,372	-	PB-3TR	Tag-PCR-32	GATGCAGCAGGATGGGGTA	Tag-PCR-3TR	GCATGCGTCAATTTTACGCAGACTATCTTTCTAG	434	68	Verified Integration
Clone-16	7	30,289,372	-	PB-5TR	Tag-PCR-5TR	CAAAATTAATAAAACCTCGATATACAGACCGATAAAACACATG	Tag-PCR-35	GCCACTCCCCATAAAAGCCT	251	68	Verified Integration
Clone-16	8	120,860,954	-	PB-3TR	Tag-PCR-36	GCCAGCTACCCGTGAACCA	Tag-PCR-3TR	GCATGCGTCAATTTTACGCAGACTATCTTTCTAG	452	70	Verified Integration
Clone-16	8	120,860,954	-	PB-5TR	Tag-PCR-5TR	CAAAATTAATAAAACCTCGATATACAGACCGATAAAACACATG	Tag-PCR-37	GGAGCAGCAGGATTCACAGT	455	70	Verified Integration
Clone-16	9	66,315,816	-	PB-3TR	Tag-PCR-55	TGGCTTCATTGCAAAAGGTTG	Tag-PCR-3TR	GCATGCGTCAATTTTACGCAGACTATCTTTCTAG	284	68	Verified Integration
Clone-16	9	66,315,816	-	PB-5TR	Tag-PCR-5TR	CAAAATTAATAAAACCTCGATATACAGACCGATAAAACACATG	Tag-PCR-38	GGCAGTAGGCTTTGTGACCT	316	68	Verified Integration
Clone-16	11	65,556,905	+	PB-3TR	Tag-PCR-20	CATTGACAAGCAGCCCTCAC	Tag-PCR-12	AGGAGGAGCTAAGGTCAAGG	411	72	Verified Integration
Clone-16	11	65,556,905	+	PB-5TR	Tag-PCR-33	GCACACCCCATCTCATGACA	Tag-PCR-5TR	CAAAATTAATAAAACCTCGATATACAGACCGATAAAACACAT	330	68	Verified Integration
Clone-16	13	23,753,309	-	PB-3TR	Tag-PCR-06	ACACGCTGTACACACACCC	Tag-PCR-3TR	GCATGCGTCAATTTTACGCAGACTATCTTTCTAG	477	70	Verified Integration
Clone-16	13	23,753,309	-	PB-5TR	Tag-PCR-5TR	CAAAATTAATAAAACCTCGATATACAGACCGATAAAACACATG	Tag-PCR-11	AGCTAGAGGGCAACACCTTA	512	70	Verified Integration
Clone-16	14	54,325,742	-	PB-3TR	Tag-PCR-17	CAGCCGGTGGAGGTTGATG	Tag-PCR-29	CTCAGGGAGGCTCCAAGC	295	70	Verified Integration
Clone-16	14	54,325,742	-	PB-5TR	Tag-PCR-5TR	CAAAATTAATAAAACCTCGATATACAGACCGATAAAACACATG	Tag-PCR-07	ACCTAGCTCTGTTGGCATCC	388	68	Verified Integration
Clone-16	20	57,699,496	+	PB-3TR	Tag-PCR-3TR	GCATGCGTCAATTTTACGCAGACTATCTTTCTAG	Tag-PCR-14	CACCAGCGACATCCGACC	411	70	Verified Integration
Clone-16	20	57,699,496	+	PB-5TR	Tag-PCR-43	TACCATCGGCACCAACCTC	Tag-PCR-5TR	CAAAATTAATAAAACCTCGATATACAGACCGATAAAACACAT	307	68	Verified Integration
Clone-16	3	4,927,456	+	PB-3TR	Tag-PCR-3TR	GCATGCGTCAATTTTACGCAGACTATCTTTCTAG	Tag-PCR-40	GGGTGTGGCCTGGTTCTC	472	70	Verified Integration
Clone-16	3	4,927,456	+	PB-5TR	Tag-PCR-01	AAACCTTGGGCCGAGAGAAG	Tag-PCR-5TR	CAAAATTAATAAAACCTCGATATACAGACCGATAAAACACAT	370	70	Verified Integration
Clone-16	17	507,201,24	+	PB-3TR	PB-3TR-Nested	ATTTCAGAATGCATGCGCTGA	Clone-16 413-R	TCCCGTATAAGCGTCTATGC	219	60	Verified Integration
Clone-16	17	507,201,24	+	PB-5TR	Clone-16 413-F	TAGGAGGGATTTCGGCCTGA	PB-5TR-Nested	CACATGATTATCTTTAACGTACGTAC	554	60	Verified Integration
Clone-16	7	33060938	-	PB-3TR	PB-3TR-Nested	ATTTCAGAATGCATGCGCTGA	Clone-16 14-R	GATGGGCAGAGCAGACCTAC	243	60	Verified Integration
Clone-16	7	33060938	-	PB-5TR	Clone-16 14-F	CACCTCAGCCTGGGAAACAT	PB-5TR-Nested	CACATGATTATCTTTAACGTACGTAC	562	60	Verified Integration
Clone-13	1	121,201,016	-	PB-3TR	Tag-PCR-48	TCTGGGTGTTCGGGAAAGTG	Tag-PCR-3TR	GCATGCGTCAATTTTACGCAGACTATCTTTCTAG	594	68	Verified Integration
Clone-13	1	121,201,016	-	PB-5TR	Tag-PCR-5TR	CAAAATTAATAAAACCTCGATATACAGACCGATAAAACACATG	Tag-PCR-04	AAGAGCTGGCCCTTCGGAC	551	68	Verified Integration
Clone-13	2	151,750,189	-	PB-3TR	Tag-PCR-42b	AGCCCTTTCTGCTATCTGCTT	Tag-PCR-3TR	GCATGCGTCAATTTTACGCAGACTATCTTTCTAG	400	68	Verified Integration
Clone-13	2	151,750,189	-	PB-5TR	Tag-PCR-5TR	CAAAATTAATAAAACCTCGATATACAGACCGATAAAACACATG	Tag-PCR-15	CACCCTGCAGGAAAGCCTAA	343	68	Verified Integration
Clone-13	11	66,620,830	-	PB-3TR	Tag-PCR-45b	GGAGGGAAACAGACAGTGGG	Tag-PCR-3TR	GCATGCGTCAATTTTACGCAGACTATCTTTCTAG	155	68	Verified Integration
Clone-13	11	66,620,830	-	PB-5TR	Tag-PCR-5TR	CAAAATTAATAAAACCTCGATATACAGACCGATAAAACACATG	Tag-PCR-41	GTGGAAGGATGGCTTGAGCT	398	70	Verified Integration
Clone-13	14	55,102,774	-	PB-3TR	Tag-PCR-57	TGCTTGGCAAGCAGCCT	Tag-PCR-3TR	GCATGCGTCAATTTTACGCAGACTATCTTTCTAG	446	68	Verified Integration
Clone-13	14	55,102,774	-	PB-5TR	Tag-PCR-5TR	CAAAATTAATAAAACCTCGATATACAGACCGATAAAACACATG	Tag-PCR-61b	ACCTGCTTCTCTGTGCTT	420	68	Verified Integration
Clone-13	17	80,019,664	-	PB-3TR	Tag-PCR-03	AACCTCGACGGGTGAGAC	Tag-PCR-3TR	GCATGCGTCAATTTTACGCAGACTATCTTTCTAG	345	70	Verified Integration
Clone-13	17	80,019,664	-	PB-5TR	Tag-PCR-5TR	CAAAATTAATAAAACCTCGATATACAGACCGATAAAACACATG	Tag-PCR-30	CTTGCTGTAGTGGGCATCA	553	68	Verified Integration
Clone-13	12	44,243,842	+	PB-3TR	Tag-PCR-34	GCATTGACAAGCAGCCTCAC	Tag-PCR-46	TCAGGCGCAGAAAGACAAGT	205	70	Verified Integration
Clone-13	12	44,243,842	+	PB-5TR	Tag-PCR-08	ACTGTGCTGTGCCATTCTGA	Tag-PCR-5TR	CAAAATTAATAAAACCTCGATATACAGACCGATAAAACACAT	258	68	Verified Integration
Clone-13	9	83093891	-	PB-3TR	PB-3TR-Nested	ATTTCAGAATGCATGCGCTCA	Clone-13 28-R	CAAGCTGAGGGATCTGGCTC	233	60	Verified Integration
Clone-13	9	83093891	-	PB-5TR	Clone-13 28-F	TGTGGAACAGGAATGGCATCT	PB-5TR-Nested	CACATGATTATCTTTAACGTACGTAC	338	60	Verified Integration
E1118	2	127,571,832	+	PB-3TR	Tag-PCR-3TR	GCATGCGTCAATTTTACGCAGACTATCTTTCTAG	Tag-PCR-18	CAGCTCACTGCAAGGCGATTC	152	69	Verified Integration
E1118	2	127,571,832	+	PB-5TR	Tag-PCR-16	CAGCAGAGAGGGAGACATGG	Tag-PCR-5TR	CAAAATTAATAAAACCTCGATATACAGACCGATAAAACACAT	275	68	Verified Integration
E1118	1	218,457,763	-	PB-3TR	Tag-PCR-31	GAACCATATGGGCCCCCG	Tag-PCR-5TR	GCATGCGTCAATTTTACGCAGACTATCTTTCTAG	461 if no integration	68	Verified Lack of Integration
E1118	1	218,457,763	-	PB-5TR	Tag-PCR-5TR	CAAAATTAATAAAACCTCGATATACAGACCGATAAAACACATG	Tag-PCR-56	TGGTCAGTGGCGTTGATTCA	461 if no integration	68	Verified Lack of Integration
E1118	3	61,259,373	-	PB-3TR	Tag-PCR-58b	GGCTCCTAGAAGGTGTTGGC	Tag-PCR-3TR	GCATGCGTCAATTTTACGCAGACTATCTTTCTAG	216	68	Verified Integration
E1118	3	61,259,373	-	PB-5TR	Tag-PCR-5TR	CAAAATTAATAAAACCTCGATATACAGACCGATAAAACACATG	Tag-PCR-26	CCTCTGAGTAGAAGGGCCT	423	68	Verified Integration



**Supplemental Table III.** Table of SRA datasets with raw sequence reads for each sample.

Sample	Type	SRA Run(s)
Clone-6	esTag-PCR	SRR15289274, SRR15289282, SRR15289287
Clone-6	Tag-PCR	SRR15289268, SRR15289271, SRR15289290, SRR16766766
Clone-6	Tn5/Anchor-Seq	SRR16766761
Clone-10	esTag-PCR	SRR15289284
Clone-11	esTag-PCR	SRR15289283
Clone-13	esTag-PCR	SRR15289273, SRR15289286, SRR15289294
Clone-13	Tag-PCR	SRR15289270, SRR15289289, SRR15289292
Clone-13	Tn5/Anchor-Seq	SRR16766760
Clone-16	esTag-PCR	SRR15289272, SRR15289285, SRR15289293
Clone-16	Tag-PCR	SRR15289269, SRR15289288, SRR15289291, SRR16766765
Clone-16	Tn5/Anchor-Seq	SRR16766759
Clone-18	esTag-PCR	SRR15289281
Clone-21	esTag-PCR	SRR15289280
Clone-22	esTag-PCR	SRR15289279
E1118	esTag-PCR	SRR15289161
E1119	esTag-PCR	SRR15289160
E2164	esTag-PCR	SRR15289149
E2165	esTag-PCR	SRR15289138
E2166	esTag-PCR	SRR15289127
E2167	esTag-PCR	SRR15289169
E2168	esTag-PCR	SRR15289165
E2169	esTag-PCR	SRR15289164
E2170	esTag-PCR	SRR15289163
E2171	esTag-PCR	SRR15289162
E2172	esTag-PCR	SRR15289159
E2173	esTag-PCR	SRR15289158
E2174	esTag-PCR	SRR15289157
E2175	esTag-PCR	SRR15289156
E2204	esTag-PCR	SRR15289151
E2205	esTag-PCR	SRR15289154
E2207	esTag-PCR	SRR15289153
E2208	esTag-PCR	SRR15289152
E2209	esTag-PCR	SRR15289148
E2210	esTag-PCR	SRR15289155
E2211	esTag-PCR	SRR15289150
E2224	esTag-PCR	SRR15289147
E2225	esTag-PCR	SRR15289145
E2235	esTag-PCR	SRR15289146
E2301	esTag-PCR	SRR15289137
E2302	esTag-PCR	SRR15289136
E2304	esTag-PCR	SRR15289139
E2306	esTag-PCR	SRR15289134
E2307	esTag-PCR	SRR15289133
E2308	esTag-PCR	SRR15289135
E2309	esTag-PCR	SRR15289132
E2310	esTag-PCR	SRR15289130
E2311	esTag-PCR	SRR15289131
E2312	esTag-PCR	SRR15289123
E2314	esTag-PCR	SRR15289121
E2316	esTag-PCR	SRR15289126
E2317	esTag-PCR	SRR15289129
E2318	esTag-PCR	SRR15289122
E2374	esTag-PCR	SRR15289167
E2376	esTag-PCR	SRR15289117
E2377	esTag-PCR	SRR15289166
E2378	esTag-PCR	SRR15289168
HEK293	esTag-PCR-LV	SRR15289277, SRR15289278
HepG2	esTag-PCR-LV	SRR15289275, SRR15289276
NHDF	esTag-PCR-LV	SRR16028388, SRR16028387
Clone-6 Clone-13 Dilution Step 1	esTag-PCR Dilution Expt	SRR16775679
Clone-6 Clone-13 Dilution Step 2	esTag-PCR Dilution Expt	SRR16775678
Clone-6 Clone-13 Dilution Step 3	esTag-PCR Dilution Expt	SRR16775667
Clone-6 Clone-13 Dilution Step 4	esTag-PCR Dilution Expt	SRR16775666
Clone-6 Clone-13 Dilution Step 5	esTag-PCR Dilution Expt	SRR16775665
Clone-6 Clone-13 Dilution Step 6	esTag-PCR Dilution Expt	SRR16775664
Clone-6 Clone-13 Dilution Step 7	esTag-PCR Dilution Expt	SRR16775663
Clone-6 Clone-13 Dilution Step 8	esTag-PCR Dilution Expt	SRR16775662
Clone-6 Clone-13 Dilution Step 9	esTag-PCR Dilution Expt	SRR16775661
Clone-6 Clone-13 Dilution Step 10	esTag-PCR Dilution Expt	SRR16775660
Clone-6 HepG2 Dilution Step 1	esTag-PCR Dilution Expt	SRR16775677
Clone-6 HepG2 Dilution Step 2	esTag-PCR Dilution Expt	SRR16775676
Clone-6 HepG2 Dilution Step 3	esTag-PCR Dilution Expt	SRR16775675
Clone-6 HepG2 Dilution Step 4	esTag-PCR Dilution Expt	SRR16775674
Clone-6 HepG2 Dilution Step 5	esTag-PCR Dilution Expt	SRR16775673
Clone-6 HepG2 Dilution Step 6	esTag-PCR Dilution Expt	SRR16775672
Clone-6 HepG2 Dilution Step 7	esTag-PCR Dilution Expt	SRR16775671
Clone-6 HepG2 Dilution Step 8	esTag-PCR Dilution Expt	SRR16775670
Clone-6 HepG2 Dilution Step 9	esTag-PCR Dilution Expt	SRR16775669
Clone-6 HepG2 Dilution Step 10	esTag-PCR Dilution Expt	SRR16775668
100ng Arrested Embryo-1	esTag-PCR	SRR15289124
100ng Arrested Embryo-2	esTag-PCR	SRR15289128
100ng Arrested Embryo-3	esTag-PCR	SRR15289125
100ng Arrested Embryo-4	esTag-PCR	SRR15289144
100ng Arrested Embryo-5	esTag-PCR	SRR15289143
100ng Arrested Embryo-6	esTag-PCR	SRR15289142
100ng Arrested Embryo-7	esTag-PCR	SRR15289140
100ng Arrested Embryo-8	esTag-PCR	SRR15289141
30ng Arrested Embryo-1	esTag-PCR	SRR15430997
30ng Arrested Embryo-2	esTag-PCR	SRR15430993
30ng Arrested Embryo-3	esTag-PCR	SRR15430990
30ng Arrested Embryo-4	esTag-PCR	SRR15430998
30ng Arrested Embryo-5	esTag-PCR	SRR15430992
30ng Arrested Embryo-6	esTag-PCR	SRR15430999
30ng Arrested Embryo-7	esTag-PCR	SRR15431000
30ng Arrested Embryo-8	esTag-PCR	SRR15430991
30ng Arrested Embryo-9	esTag-PCR	SRR15430989
30ng Arrested Embryo-10	esTag-PCR	SRR15430996
30ng Arrested Embryo-11	esTag-PCR	SRR15430988