Method for detection and identification of multiple chromosomal integration sites in transgenic animals created with lentivirus

Elizabeth C. Bryda, Michael Pearson, Yuksel Agca, and Beth A. Bauer
University of Missouri, Columbia, MO, USA

BioTechniques 41:715-719 (December 2006)
doi 10.2144/000112289

Transgene delivery systems, particularly those involving retroviruses, often result in the integration of multiple copies of the transgene throughout the host genome. Since site-specific silencing of transgenes can occur, it becomes important to identify the number and chromosomal location of the multiple copies of the transgenes in order to correlate inheritance of the transgene at a particular chromosomal site with a specific and robust phenotype. Using a technique that combines restriction endonuclease digest and several rounds of PCR amplification followed by nucleotide sequencing, it is possible to identify multiple chromosomal integration sites in transgenic founder animals. By designing genotyping assays to detect each individual integration site in the offspring of these founders, the inheritance of transgenes integrated at specific chromosomal locations can be followed efficiently as the transgenes randomly segregate in subsequent generations. Phenotypic characteristics can then be correlated with inheritance of a transgene integrated at a particular chromosomal location to allow rational selection of breeding animals in order to establish the transgenic line.

INTRODUCTION

Production of genetically engineered animals by pronuclear injection or retroviral delivery systems has been a successful strategy for generating animal models to better understand the functionality of genes. In transgenic animals created by embryo microinjection, the site of integration of the transgene within the genome is random. Thus, when multiple embryos have been injected or infected with the same DNA, the integration site will be different in each founder animal. When the integration events have occurred at the one-cell stage, they should exhibit germline transmission with the potential to be inherited by the founder’s offspring. If the integration occurs at a later stage, the resulting mosaic founder may or may not exhibit germline transmission of the transgene. In the case of pronuclear injection, there is typically one insertion site, although multiple transgene copies are often found in a tandem array at that integration site (1).

Lentivirus transgenesis is becoming an increasingly attractive alternative to pronuclear injection because it is more efficient in terms of successful transgene incorporation into the host genome, less invasive to the embryo, and technically less demanding to perform (2). Lentiviral delivery systems have been used successfully to generate transgenic mice, rats, pigs, and cattle (2–7). The disadvantage of lentivirus is that there are often multiple integration events with random transgene insertions on several chromosomes.

Independent of the method of transgene delivery, the insertion site can have profound effects on transgene expression. This can lead to phenotypic effects in the transgenic animal that are not due to the transgene per se, but are a consequence of the integration site, a phenomenon referred to as position effect (8). It is critical to correlate phenotype with genotype, particularly in animals created via lentivirus transgenesis, since not all copies of the transgene may be contributing equally to the phenotype.

Determining transgene integration sites is challenging. A number of PCR-based methods, often referred to as chromosome walking techniques, have been developed to isolate DNA fragments adjacent to known sequences, including inverse PCR (9), ligation-mediated PCR (LM-PCR) (10), randomly primed PCR (RP-PCR) (11,12), and T-linker PCR (13). The method described in this paper incorporates several elements of these techniques in a unique way that allows the capture of DNA fragments containing the chromosomal region flanking the transgene. Our method enables quick and inexpensive determination of multiple independent transgene integration sites in founder animals and their offspring. Here, we demonstrate that this method is useful for identifying and monitoring multiple transgene integration sites in transgenic animals created using lentivirus.

MATERIALS AND METHODS

Animals

Lewis rat lines carrying an enhanced green fluorescent protein (EGFP) transgene were created by using the EGFP DNA construct and experimental protocol described by Lois et al. (2). Transgene positive founder animals were identified using an EGFP PCR assay (14). To assess GFP expression, tail biopsies were examined for fluorescence under a Nikon SMZ1500 UV dissecting scope (Nikon Instruments, Melville, NY, USA). Founders (F0) were bred to wild-type Lewis rats obtained from Harlan Sprague Dawley (Indianapolis, IN, USA) to generate the N1. N1 animals were genotyped using integration site-specific PCR assays. Selected N1 animals were mated to wild-type Lewis rats to generate the N2. N2 animals were genotyped using integration site-specific genotyping assays, and GFP expression was confirmed. Several of these lines (F455.5, F456.9, F458.7, F463.1, and F463.5) demonstrated stable transmission of the integration site-specific transgene coupled with robust GFP expression and were donated to the Rat Research and Resource Center (RRRC) at the University of Missouri. All other rat lines and mouse strains used are available either through the RRRC (www.nrrrc.missouri.edu) for rats or...
the University of Missouri/Harlan Mutant Mouse Regional Resource Center (MMMRC; www.mmmrc.org) for mice.

**Preparation of DNA**

DNA was isolated from tail biopsies using the DNeasy® Tissue kit (Qiagen, Valencia, CA, USA). Restriction endonuclease digestion was performed with *Pst*I and *Hha*I (Figure 1, step 1). These enzymes were chosen because they created 3′ overhangs and their recognition sites were not present within the transgene sequence. Three micrograms genomic DNA were digested with 20 U enzyme in a total reaction volume of 30 μL as recommended by the manufacturer. Reactions were incubated for 2 h for partial digestion.

**PCR Amplification and Linker Ligation**

Three transgene-specific nested primers were designed to both the 5′ and 3′ regions of the transgene using PrimerQuest available from Integrated DNA Technologies (IDT; Coralville, IA, USA). Primer 1 was farthest from the junction between the known transgene sequence and the unknown integration site, whereas primer 3 was closest (Figure 1). All gene-specific primers were designed to have an optimum melting temperature (Tm) of 60°C, optimum primer length of 24 bp, and an optimum % GC content of 50. Y-linker and primer sequences for the Y-linker and their relationships are provided in Figure 2. Y-linker A (Figure 2) contains a terminal inverted T at the 3′ end to inhibit extension by DNA polymerases. The Y-linker was prepared by combining equal volumes of 8 μM Y-linker A and 8 μM Y-linker E, incubating at 95°C for 5 min, and then letting the reaction cool on the benchtop. The sequences of the gene-specific primers used for the lentivirus-generated rat lines are available upon request or are listed under the genotyping assay for the specific lines at www.nrrrc.missouri.edu. Primers and linkers were synthesized by IDT.

All PCRs were performed in a 25-μL volume containing 2.5 μL 10× FastStart Taq with 20 mM MgCl2 (Roche Diagnostics, Indianapolis, IN, USA), 0.2 mM each dNTP, 1.6 μM each primer, and 1.25 U FastStart Taq buffer. In the first PCR, 1 μL restriction digest from above and transgene-specific primer 1 were used, and a single round of PCR was performed with the following thermal cycling conditions: 94°C for 10 min, 60°C for 1 min, and 72°C for 10 min (Figure 1, step 1). Following amplification, a 10-μL reaction containing 7 μL this PCR, 1 μM Y-linker, 400 U T4 DNA Ligase (New England Biolabs, Ipswich, MA, USA), and 1 μL 10× T4 DNA Ligase buffer supplied with the enzyme was incubated at 16°C for 16 h (Figure 1, step 3). Transgene-specific primer 2, Y-primer D, and 1 μL ligase reaction were used in the second PCR with the following cycling conditions: 94°C for 5 min, 20 cycles of 94°C for 30 s, 60°C for 30 s, 72°C for 1.5 min, and one cycle of 72°C for 10 min (Figure 1, step 4). The same cycling conditions were used in the third PCR. This final reaction utilized transgene-specific primer 3, Y-primer G, and 1 μL from the second PCR (Figure 1, step 5).

**PCR Product Purification and Analysis**

Amplification products from the third PCR were separated by gel electrophoresis on 1%–3% agarose gels, and individual products were gel-purified using QIAquick® Gel Extraction kit (Qiagen). When multiple bands were detected, all bands were isolated. The nucleotide sequence of each amplification product was determined using the transgene-specific primer 3 as a sequencing primer (Figure 1, step 6). Nucleotide sequencing was performed either at our DNA Core (University of Missouri-Columbia) or by SeqWright (Houston, TX, USA). The nucleotide sequences were aligned and examined to confirm the presence of the expected known transgene sequence and determine the flanking sequence repre-
senting the insertion site. The insertion site sequence was analyzed using Basic Local Alignment Search Tool (BLAST) (15) to determine the chromosomal location of the transgene.

RESULTS AND DISCUSSION

Lentivirus was used as a delivery system to create Lewis rat lines carrying an EGFP transgene (2). A total of nine transgene positive founder (F0) animals were recovered. While all nine founders were positive for the presence of the EGFP transgene, eight founders expressed EGFP based on epifluorescent microscopy of tail biopsies, while one founder (456) had no detectable fluorescence (Table 1).

Our method for determining chromosomal integration sites was used to identify the chromosomal location(s) of the transgene in each of the nine founders. PCR products ranged in size from 100–1000 bp. Longer PCR fragments were generally necessary for determining the insertion site when the region contained repetitive elements. Several sites were successfully identified with recovered products as small as 110 bp, which included as little as 20 bp genomic sequence. For seven of the founders, we identified one to four integration sites depending on the founder. We failed to identify the integration site in two founders using PstI and HhaI digestion, and we did not pursue these further. It is possible that by using other restriction enzymes with 4–6 bp recognition sites or increasing PCR extension times to generate larger products, we would have successfully identified the integration sites in these founder lines. For seven lines, site-specific genotyping assays were developed for every integration site identified in the founder. In one case (463), we recovered genomic sequence that matched a sequence found on many chromosomes, so we could not assign a chromosomal location.

<table>
<thead>
<tr>
<th>Founder ID (no.)</th>
<th>Chromosome Insertion Sites</th>
<th>GFP Fluorescence</th>
<th>F0</th>
<th>Chromosome Insertion Sites</th>
<th>PCR + (no.)</th>
<th>GFP Fluorescence</th>
<th>N1</th>
<th>Chromosome Insertion Sites</th>
<th>PCR + (no.)</th>
<th>GFP Fluorescence</th>
<th>N2</th>
<th>Chromosome Insertion Sites</th>
<th>PCR + (no.)</th>
<th>GFP Fluorescence</th>
</tr>
</thead>
<tbody>
<tr>
<td>452</td>
<td>12</td>
<td>+++</td>
<td></td>
<td></td>
<td></td>
<td>FS</td>
<td></td>
<td></td>
<td></td>
<td>NC</td>
<td></td>
<td></td>
<td></td>
<td>NC</td>
</tr>
<tr>
<td>455</td>
<td>1, 5</td>
<td>+</td>
<td></td>
<td>1</td>
<td>1</td>
<td>−</td>
<td></td>
<td>5</td>
<td>1</td>
<td>+++</td>
<td></td>
<td>5</td>
<td>1</td>
<td>+++</td>
</tr>
<tr>
<td>456</td>
<td>9, 17</td>
<td>−</td>
<td></td>
<td>9</td>
<td>1</td>
<td>+++</td>
<td></td>
<td>9</td>
<td>1</td>
<td>+++</td>
<td></td>
<td>5</td>
<td>1</td>
<td>+++</td>
</tr>
<tr>
<td>458</td>
<td>7, 10, 12</td>
<td>+++</td>
<td></td>
<td>7</td>
<td>3</td>
<td>+++</td>
<td></td>
<td>7</td>
<td>3</td>
<td>+++</td>
<td></td>
<td>10</td>
<td>3</td>
<td>+++</td>
</tr>
<tr>
<td>459</td>
<td>2</td>
<td>++</td>
<td></td>
<td>2</td>
<td>2</td>
<td>−</td>
<td></td>
<td>2</td>
<td>2</td>
<td>−</td>
<td></td>
<td>2</td>
<td>2</td>
<td>−</td>
</tr>
<tr>
<td>462</td>
<td>8</td>
<td>+++</td>
<td></td>
<td>8</td>
<td>1</td>
<td>+++</td>
<td></td>
<td>8</td>
<td>1</td>
<td>+++</td>
<td></td>
<td>NC</td>
<td></td>
<td>NC</td>
</tr>
<tr>
<td>463</td>
<td>UN, 1, 2, 5</td>
<td>+++</td>
<td></td>
<td>UN, 5</td>
<td>3</td>
<td>+++</td>
<td></td>
<td>UN, 5</td>
<td>3</td>
<td>+++</td>
<td></td>
<td>UN, 5</td>
<td>3</td>
<td>+++</td>
</tr>
</tbody>
</table>

Table 1. Phenotype and Genotype Data Over Three Generations

Nine founder lines (F0) and two generations of offspring for some lines were evaluated for chromosome insertion sites and green fluorescent protein (GFP) fluorescence. For the N1 and N2 generations, the number of offspring that were PCR positive using the various insertion site-specific genotyping assays are indicated along with their GFP expression levels. ++++, high expression of GFP fluorescence; +, low expression of GFP fluorescence; −, no expression of GFP fluorescence; FS, founder was sterile, no offspring were recovered; NC, breeding was not continued; UN, integration site sequence was identified, but chromosomal location could not be determined.

Figure 2. Alignment of linkers and linker primers. Y-linker A (blue) contains a nonextendable base (3InvdT) and, when annealed to Y-linker E (green), forms the complete Y-linker that is partially double-stranded. This linker has a single-stranded region (green bold) to which Y-linker primer D (pink) is designed. There is a single T (underlined) at the 3' end of the Y-linker E sequence, which is included to facilitate pairing with the 3'A-overhang on the PCR product generated using specific gene primer 1 (see Figure 1, step 2). Y-linker primer G (red) is used in the final round of PCR amplification. The orientation of each sequence (5' or 3') is indicated to the left.
to this integration site. Nonetheless, the genotyping assay based on this sequence allowed the transgene integration site to be followed not only in the founder but in his offspring.

To follow segregation of the integration sites, the offspring from matings between each of the seven founders and wild-type Lewis rats were monitored for presence and expression of the transgene at each integration site using the site-specific genotyping assays and microscopic examination for GFP fluorescence. One founder (452) was infertile. For the remaining six founder lines, offspring were obtained, and both the insertion sites and the GFP expression were determined.

Founder 455 had two independent transgene insertion sites: one on chromosome 1, and a second on chromosome 5. When this founder was mated to a wild-type Lewis animal, two N1 offspring were recovered. One N1 carried the chromosome 1 integration site, but did not have detectable GFP expression. Lack of expression may have been due to positional effects. The other N1 carried the chromosome 5 integration site and did have GFP expression. The chromosome 5 transgene positive N2 offspring (n = 5) continued to have high fluorescence, demonstrating that GFP expression in this line was associated with the chromosome 9 transgene. With the chromosome 9 transgene. We speculate that GFP expression was suppressed at the chromosome 17 insertion due to a position effect and that the founder was mosaic for the chromosome 9 insertion, resulting in undetectable GFP expression. In subsequent generations, inheritance of the chromosome 9 insertion was germline, and expression occurred in all cells leading to detectable fluorescence.

For line 458, three insertion sites were detected. By correlating GFP expression with inheritance of the various transgene insertion sites, it was possible to show that the chromosome 7 integration gave good GFP expression. This was confirmed in N2 animals (n = 10) carrying only the chromosome 7 integration. It should be noted that while none of the N1 animals carried the chromosome 10 integration site, this site was confirmed in the original founder by integration site-specific genotyping. It is possible that the founder was mosaic for the chromosome 10 integration site.

For line 459, which had a single detected transgene insertion on chromosome 2, two offspring were recovered that inherited the chromosome 2 transgene insertion site. However, neither of these animals exhibited GFP fluorescence. This could be due to a positional effect associated with the particular insertion site on chromosome 2 coupled with inheritance through the male lineage, or alternatively, we may have missed an insertion site in founder 459 that was associated with the fluorescence seen in the founder, which was not inherited by these two offspring.

Founder 463 had the greatest number of insertion sites; by continuing to correlate GFP expression with inheritance of specific transgene insertion sites, it was possible by the N2 generation to identify animals with single transgene insertion sites that maintained high GFP expression.

In addition to the lentivirus experiment described above, we have successfully used this technique to determine the lentivirus integration sites and to generate rat lines with single transgene integration sites for a Sprague-Dawley GFP transgenic (RRRC: 0065), derived from the line created by Lois et al. (2), and a rat model containing the human presenilin-1 gene (RRRC: 0061). To test whether our method for identifying insertion sites had broader application beyond just determining chromosomal locations of lentivirus transgene integration, we attempted to identify the transgene integration site of three additional rodent strains: (i) a mouse transgenic strain (MMRRC: 000366-MU/H) carrying the EGFP gene on the FVB inbred genetic background (16); (ii) a knockout mouse line (MMRRC: 000352-MU/H) involving the acetyl-coenzyme A dehydrogenase long chain gene (Acadl), which carries a duplication of exons 3 and 4 with insertion of the neocassette into the approximately 11 kb intron 4 (17); and (iii) a transgenic rat line (RRRC: 0043), which contains a mutated version of the human HLA-B2705 gene on a Lewis genetic background (18). In the case of the FVB-EGFP strain and the Acadl knockout strain, we were able to determine precisely the insertion site on chromosome 3 and within the large Acadl intron 4, respectively. Our method failed for the HLA-B2705 transgenic rat line, which has a high transgene copy number (24 copies) in homozygous animals (18). The insertions are integrated at a single locus (18), and the multiple copies have probably integrated as concatamers. A major advantage of determining the precise chromosomal integration site in animals created via pronuclear injection is that genotyping assays can be developed that allow animals heterozygous for a transgene to be easily distinguished from homozygous animals. Without this type of information, PCR-based genotyping assays can distinguish only whether animals carry the transgene. While it is possible to use Southern blot analysis and densitometry to measure transgene copy number as an alternative method to distinguish homozygotes from heterozygotes, it is laborious, time-consuming, and not practical for many labs. In summary, we have described a quick and straightforward method for determining the location of multiple chromosomal integration sites in animals created by lentivirus transgenesis. Our studies underline the need to carefully correlate a particular
integration site with gene expression over multiple generations in order to create stable models. We have also found that this method has applicability for detecting transgene integration sites in other cases of random integration, but is probably limited to situations where very few tandem copies of the transgene have been integrated at the insertion site.

ACKNOWLEDGMENTS

E.C.B. and M.P. contributed equally to this work. This work was supported in part by grants from the National Institutes of Health (NIH; U42 RR014821 and P40 RR16939). We thank James Sparks for technical assistance and Howard Wilson for assistance with graphics.

COMPETING INTERESTS STATEMENT

The authors declare no competing interests.

REFERENCES


Received 6 July 2006; accepted 22 August 2006.

Address correspondence to Beth A. Bauer or Yuksel Agca, University of Missouri-Columbia, Department of Veterinary Pathobiology, 1600 E. Rollins St., Columbia, MO 65211, USA, e-mail: bauereb@missouri.edu; agcay@missouri.edu

To purchase reprints of this article, contact: Reprints@BioTechniques.com