Graduate School ETD Form 9
(Revised 12/07)

PURDUE UNIVERSITY
GRADUATE SCHOOL
Thesis/Dissertation Acceptance

This is to certify that the thesis/dissertation prepared

By Yu-Hsiang Chen

Entitled
Analysis of Integration Sites of Transgenic Sheep Generated by Lentiviral Vectors Using Next-Generation Sequencing Technology

For the degree of Master of Science

Is approved by the final examining committee:

Anna Malkova
Chair

Kenneth Cornetta

Stephen Randall

To the best of my knowledge and as understood by the student in the *Research Integrity and Copyright Disclaimer (Graduate School Form 20)*, this thesis/dissertation adheres to the provisions of Purdue University's "Policy on Integrity in Research" and the use of copyrighted material.

Approved by Major Professor(s): Anna Malkova

Approved by: Simon Atkinson 06/28/2013
Head of the Graduate Program Date

ANALYSIS OF INTEGRATION SITES OF TRANSGENIC SHEEP GENERATED BY LENTIVIRAL VECTORS USING NEXT-GENERATION SEQUENCING TECHNOLOGY

A Thesis

Submitted to the Faculty

of

Purdue University

by

Yu-Hsiang Chen

In Partial Fulfillment of the

Requirements for the Degree

of

Master of Science

August 2013

Purdue University

Indianapolis, Indiana

For my beloved family.

獻給我親愛的家人

ACKNOWLEDGEMENTS

First, I would like to thank the team of people in Dr. Cornetta's lab who provided assistance, no matter in experiment part or mental part. They are, in alphabetical order, Aaron, Anna, Aparna, Daniela, Hongyu, Jing, Lisa, Siddharth and Tanveen. I also want to thank the group of people in Dr. Malkova's lab who have always been nice to me and helped me get used to the life in this country. They are Cynthia, Rajula, Sandeep, Soumini and Sreejith.

I also appreciate my committee members Dr. Cornetta, Dr. Malkova and Dr. Randall for their patience and advice to my research. I thank them for all the help and consideration.

My family and friends also gave me strength when I encountered any difficulty. Without your company and encouragement I would never be able to finish this work.

Finally, I would like to thank my girlfriend, Hsing-Hui, for helping me get through all the difficulty studying in the U.S. in these two years. You make me feel I am not alone.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Chen, Yu-Hsiang. M.S., Purdue University, August 2013. Analysis of Integration Sites of Transgenic Sheep Generated by Lentiviral Vectors Using Next-Generation Sequencing Technology. Major Professor: Anna Malkova.

The development of new methods to carry out gene transfer has many benefits to several fields, such as gene therapy, agriculture and animal health[1]. The newly established lentiviral vector systems further increase the efficiency of gene transfer dramatically. Some studies have shown that lentiviral vector systems enhance efficiency over 10-fold higher than traditional pronuclear injection[2], [3]. However, the timing for lentiviral vector integration to occur remains unclear. Integrating in different stages of embryogenesis might lead to different integration patterns between tissues. Moreover, in our previous study we found that the vector copy number in transgenic sheep varied, some having one or more copies per cells while other animals having less than one copy per cell suggesting mosaicism. Here I hypothesized that injection of a lentiviral vector into a single cell embryo can lead to integration very early in embryogenesis but can also occur after several cell divisions. In this study, we focus on investigating integration sites in tissues developing from different germ layers as well as extraembryonic tissues to determine when integration occurs. In addition, we are also interested in insertional mutagenesis caused by viral sequence integration in or near

gene regions. We utilize linear amplification-mediated polymerase chain reaction (LAM-PCR) [4] and next- generation sequencing (NGS) technology[5] to determine possible integration sites. In this study, we found the evidence based on a series of experiments to support my hypothesis, suggesting that integration event also happens after several cell divisions. For insertional mutagenesis analysis, the closest genes can be found according to integration sites, but they are likely too far away from the integration sites to be influenced. A well-annotated sheep genome database is needed for insertional mutagenesis analysis.

CHAPTER 1. INTRODUCTION

1.1 Objectives

The overall goal of this research was to investigate the integration pattern of lentiviral vector after direct injection of lentiviral vectors into single-cell embryo to generate transgenic sheep. So far, no study has demonstrated when the viral vector will integrate into host genome. In a study it was found the vector copy number in transgenic sheep varied, which might suggest that integration events happen after several cell divisions but can also occur very early potentially at the single cell stage. Here I hypothesized that lentiviral vector injected into a single cell embryo can lead to integration very early in embryogenesis but can also occur after several cell divisions. The integration might occur in single-cell stage, resulting in the same integration sites in every organ of the animal; it might also take place in the relatively late stage of the embryogenesis, leading to different integration sites between organs. This research is described with respect to the following specific aims:

1. To evaluate the pattern of LAM-PCR product of organs from different germ layers.
2. To localize exact integration sites by high-throughput sequencing technology.
3. To compare the integration sites between organs.

4. To verify the integration sites by conventional PCR.

5. To examine the genes near integration sites.

CHAPTER 2. LITERATURE REVIEW

2.1 Transgenic Livestock

Gene transfer technology in animals has been developed for over three decades. In 1980, the first transgenic animal was generated by microinjection of foreign DNA into pronulcei of embryos. Since then, microinjection of DNA into zygotes has been a popular method to generate transgenic mice[6]. In 1985, the first transgenic livestock was generated according to this method for the purpose of expressing human growth hormones[7]. The efficiency of generating transgenic livestock, however, was very low (1-5%)[8] due to species differences and inherent technical problems[9]. As a result, obtaining transgenic animals was not only time-consuming but also very costly[10], [11].

Many methods have been developed to overcome this shortage, such as sperm mediated DNA transfer[12], intracytoplasmic injection of sperm heads carrying DNA[13], somatic cell nuclear transfer[14] and injection of viral vectors to embryos[15]. To date, a large number of transgenic animal models have been successfully established to study mechanisms of human diseases in terms of gene-disease relationships, to evaluate gene therapy strategies, and to alter phenotype of farm animals such as increasing growth rates[1], [16], [17].

Among those methods described above, lentivirus-mediated gene transfer systems have become a popular method to accomplish this task due to several features. They share common features with retroviral systems, such as high efficient gene delivery and the ability to integrate permanently into host genome, resulting in long-term transgenic expression. Compared to retroviral rectors, lentiviral vectors can carry larger size of transgenes which can be up to 10 kilobases(kb)[18]. In addition, lentiviral vectors can also infect non-dividing cells[19]. This unique property allows lentiviral vectors to be introduced to more tissues, such as retina, brain, liver and muscle[20–22]. Due to the high efficiency of utilizing lentiviral vector as a gene transfer vehicle, many kinds of transgenic livestock have been generated with high transgenic rate, such as mice[23], pigs[9], cattle[15] and chickens[2], [24].

2.2 Lentiviral Vector

Lentivirus is one of subfamilies of retrovirus. The first isolated lentivirus was equine infectious anemia virus (EIAV). Other lentiviruses were subsequently isolated from other species, such as feline immunodeficiency virus (FIV) from cat, simian immunodeficiency virus (SIV) from nonhuman primates and human immunodeficiency virus type 1 (HIV-1) from human[25]. Lentiviral vectors were developed from the lentiviruses described above. Among these lentiviral vectors, the HIV-1-based vector system is the one which has been studied and applied the most[26].

As one of the subfamilies of retroviruses, lentiviral vectors share many features with retroviruses, such as an RNA genome with *gag, pol*, and *env* genes, which code for internal structure proteins (capsid), viral enzymes (reverse transcriptase and integrase), and envelope glycoproteins, respectively[8]. Usually, the *env* gene would be replaced by vesicular stomatitis virus G protein (VSV-G) gene[27] to broaden host range and to stabilize particles that can be concentrated by ultracentrifugation. Besides this, lentiviral vectors have long terminal repeat (LTR) DNA segmented into U3, R, U5 regions, located at both ends and required for vector integration. Second generation lentiviral vectors have U3 region of 5' LTR replaced by a cytomegalovirus (CMV) promoter to increase transgene expression[28].

2.3 Safety Concern

In spite of the advantages of utilizing lentiviral vector as a gene delivery vehicle, there are still concerns regarding its safety. Although some modifications have been made to ensure safety in designing lentiviral vectors, such as deleting some HIV genes[29], [30], using self-inactivating 3' LTR to eliminate transcriptional ability[31], [32] and separating vector components into three to four different plasmids[30], the possibility of generating replication competent lentivirus (RCL) due to recombination of plasmids and endogenous viral sequences still can not be overlooked. In addition, the tendency of lentiviral vectors to insert sequences semi-randomly into host genome is another concern[33]. This tendency would result in either altering the expression level of nearby genes or disrupting the function of the host genes if the insertion sites are located in

functional domains[19]. Insertional mutagenesis has been observed in trials of X-linked severe combined immunodeficiency (SCID-X1) treated with gammaretroviral vectors. Several SCID-X1 patients developed leukemia after being treated with gene therapy due to the insertion of retroviral vectors into position near *LMO2* proto-oncogene promoter, leading to abnormal expression of *LMO2*[34], [35]. Another concern would be the transfer of vector sequences to non-target tissues, for example, from transgenic embryos to surrogates after embryo transfer[36]. It also could be possible that the transgenic cells migrate through placenta during pregnancy or delivery.

In a previous study of transgenic sheep[37], no evidence of RCL had been observed in surrogates, fetuses or lambs. RCL had been evaluated by: (1) p24 ELISA, which is performed to screen for HIV-1 viral capsid; (2) high sensitive real-time polymerase chain reaction (qPCR) to detect VSV-G envelope, which is used to pseudotype HIV-1 due to its ability to infect broader cell types.

In a previous study the vector copy number was also evaluated to quantitate gene transfer. Although the majority of the animals had one or more copies per cell, some animals had less than one copy per cell suggesting that there might be mosaicism. This result could occur if the integration happened after several cell divisions. Based on this hypothesis, in this study we focused on identifying lentiviral vector integration sites in transgenic sheep fetal tissues. We evaluated the tissues including placenta and tissues derived from three different germ layers. In addition, we also wanted to further

evaluate insertional mutagenesis caused by viral vector integration. We confirmed the location where the lentiviral vectors integrate to see if the integration sites located in or near important genes.

To identify the integration sites, we conducted LAM-PCR on both sheep fetal and some surrogate tissues. After performing LAM-PCR, we barcoded samples by different index sequences so that we could run multiple samples in one NGS run. After analyzing sequencing data, we verified these integration sites by conventional PCR.

CHAPTER 3. MATERIALS AND METHODS

3.1 Production of Transgenic Embryos

For this portion of the experiment we collaborated with a team led by Dr. Westhusin in the Departments of Veterinary Physiology and Pharmacology, College of Veterinary Medicine, Texas A&M University. Recombinant lentivirus was produced from second generation lentiviral plasmids which contained a green fluorescent protein (GFP) gene as described in the paper of Miyoshi et al.[32] with modifications to enhance titer for embryo microinjection.

Zygotes were obtained surgically from superovulated donor ewes 24 hours post mating. Microinjection was then done by injecting 20 picoliters of High titer (10^9 particles/ml) recombinant lentivirus into perivitelline space of the embryos(Figure 3.1). After injection, the embryos were transferred back to the oviducts of recipient ewes, which received 3-4 embryos for each. At around 70 days of gestation, the pregnant ewe were euthanized to collect tissues from fetuses, placenta and surrogate ewes for analysis.

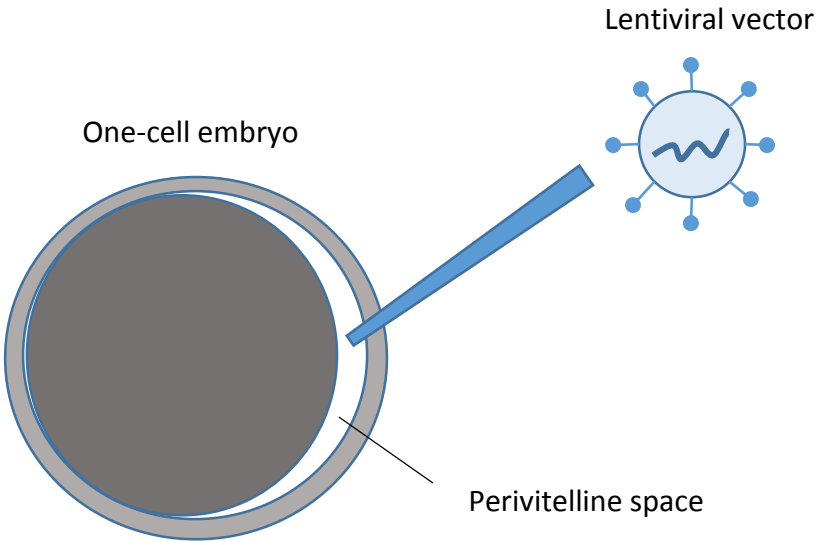

Figure 3.1. Schematic figure of embryo injection of lentiviral vectors into perivitelline space of one-cell sheep embryo

3.2 Tissue Collection and DNA Extraction

Fetuses and surrogate ewes were dissected to collect tissues including heart, liver, lung, kidney, intestine, skeletal muscle, skin, gonad, placentome, uterus, interplacentomal uterus when available. Tissues were cut into 3-5 mm pieces and preserved in All Protect tissue reagent (QIAGEN, Hilden, Germany).

DNA was extracted using DNeasy Blood & Tissue kit (QIAGEN). The procedure was as follows: tissues were cut up to 25 mg and then put into a 1.5 ml microcentrifuge tube. If tissue weight is heavier than 25 mg, the tissue was separated into more than two tubes. To each tube 180 ul of Buffer ATL wad added with 20 ul proteinase K into tube then mix thoroughly by vortexing, and incubated at 56 °C until the tissue is completely lysed. Added 4 ul RNaes A (100 mg/ml, Qiagen) to tube and mixed by vortexing, then incubated at room temperature for 10 minutes. After this 200 ul of Buffer AL was added to a tube and mixed by vortexing. Then 200 ul of ethanol (98-100%) was added to a tube and mixed by vortexing. The mixture was pipetted into DNease Mini spin column placed in a 2 ml collection tube and centrifuged at 8000 rpm for 1 minute. Discarded flow-through and collection tube. Placed DNease Mini spin column in a new 2 ml collection tube, then added 500 ul Buffer AW2, and centrifuged at 14,000 rpm for 3 minutes. Discarded flow-through and collection tube. Placed DNease Mini spin column in a new 1.5 ml tube, then added 200 ul Buffer AE, then incubated at room temperature for 2 minutes. Centrifuged at 8000 rpm for 3 minutes to elute DNA.

3.3 Integration Analysis

3.3.1 LAM-PCR

We took 100 ng DNA from each sample according to the concentration measured from previous step. Linear amplification was performed using labeled LTR-specific primer (LTR Ib-bio, 5'-gaa ccc act gct taa gcc tca-3'). PCR reaction was set up in 0.2 ml tube that contained the following: 5 ul of 10X PCR buffer (Qiagen), 1 ul of 10 mM dNTP, 0.5 ul of 0.5 uM LTR Ib-bio primer (IDT), 0.5 ul of Taq Polymerase (5 units/ul, Qiagen), 100 ng DNA, and ddH_2O to make up total volume of 50 ul. Amplified DNA fragments using the following PCR program: denaturation at 95°C for 5 minutes, followed by 50 cycles of denaturation at 95°C for 1 minute, annealing at 60°C for 45 seconds, and extension at 72°C for 1.5 minutes. A final extension for 10 minutes at 72°C was also included. 1.5 ul of ddH_2O and 0.5 ul of Taq polymerase were added to each tube, then repeated the program above.

20 ul streptavidin-coated magnetic beads (Dynal M-280) was used for each tube to capture PCR products with biotin. Then incubated at room temperature on a shaker for 8-48 hours. Washed beads with 100 ul ddH_2O twice on magnetic stand then discarded all liquid in tube.

Second-stranded synthesis was then performed on single-stranded DNA captured on magnetic beads. The reaction was set up as follows: 2 ul of 10X Hexanucleotide Mix

(Roche), 0.5 ul of 10 mM dNTP, 1 ul of Klenow polymerase (Roche), and 16.5 ul of ddH_2O. Tubes were incubated at 37°C for 1 hour. Beads were washed with 100 ul water twice on magnetic stand then discarded all liquid in tube.

DNA was then digested by Tsp509I. The reaction was set up as follows: 2 ul of 10X Restriction Buffer #1 (NEB), 1 ul of Tsp509I (2.5 units/ul, NEB), and 17 ul of ddH_2O. Tubes were incubated at 65°C for 1 hour. Beads were washed with 100 ul ddH_2O twice on magnetic stand then discarded all liquid in tube.

An adaptor cassette (generated by oligonucleotide 5'-gac ccg gga gat ctg aat tca gtg gca cag cag tta gg-3' and oligonucleotide 5'-aat tcc taa ctg ctg tgc cac gta att cag atc-3') was ligated to the digested end of the captured fragments. The reaction was set up as follows: 1 ul of 10X Incubation Buffer (Epicentre Biotech), 1 ul of ATP (10 mM, Epicentre Biotech), 2 ul of Adaptor cassette (Epicentre Biotech), 1 ul of Fast Link' DNA ligase (2 units/ul, Epicentre Biotech), and 5 ul of ddH_2O. Then incubated at room temperature for 30 minutes. The beads were washed with 100 ul ddH_2O twice on magnetic stand then discarded all liquid in tube. Denatured DNA by 5 ul fresh 0.1 N NaOH. Then incubated at room temperature for 30 minutes followed by using magnetic stand to transfer 5 ul single-strand DNA to a new 1.5 ml tube.

Nested PCR was then performed. For the first round of PCR (primers: LTR II-bio, 5'-agc ttg cct tga gtg ctt ca-3' and LC1, 5'-gac ccg gga gat ctg aat tc-3'), the reaction and were

set up as follows: 5 ul of 10X PCR buffer (Qiagen), 1 ul of 10 mM dNTP, 0.5 ul of 50 uM LTR II-bio primer (IDT), 0.5 ul of 50 uM LC1 primer (IDT), 1 ul of Taq Polymerase (5 units/ul, Qiagen), 2 ul of DNA from previous step, and 40 ul of ddH_2O. Amplified DNA fragments using the following PCR program: denaturation at 95°C for 5 minutes, followed by 35 cycles of denaturation at 95°C for 1 minute, annealing at 60°C for 45 seconds, and extension at 72°C for 1.5 minutes. A final extension for 10 minutes at 72°C was also included.

PCR products were captured by 20 ul streptavidin-coated magnetic beads. Washed by 100 ul ddH_2O twice on magnetic stand. Discarded all the liquid in tube. Denatured DNA by 20 ul 0.1 N NaOH. Collected 20 ul denatured DNA to a new 1.5 ul tube then proceeded to second round PCR.

For the second round of PCR (primers: LTRIII, 5'-nnn nnn agt agt gtg tgc ccg tct gt-3' and LCII, 5'-agt ggc aca gca gtt agg), the reaction was set up as follows: 5 ul of 10X PCR buffer (Qiagen), 1 ul of 10 mM dNTP, 0.5 ul of 50 uM LTR III primer (IDT), 0.5 ul of 50 uM LCII primer (IDT), 1 ul of Taq Polymerase (5 units/ul, Qiagen), 2 ul of DNA from previous step, and 40 ul of ddH_2O. PCR program was the same as first round PCR. The resulting products were visualized by gel eletrophoresis.

Figure 3.2. Schematic figure of LAM-PCR. Linear PCR was performed to amplify vector-genome junction region; PCR products were converted to double-stranded, followed by restriction enzyme digestion. Later, linker cassette was ligated to introduce known sequence to the other end of fragments. Nested PCR was performed to amplify the signal so that LAM-PCR products could be seen on a gel.

3.3.2 Next Generation Sequencing and Reads Processing

To sequence LAM-PCR products, individually bar-coded amplicon libraries were generated by using forward fusion primers containing different indices during round 2 nested PCR (Figure 3.3; Table 3.1). Samples were pooled and sequenced on Illumina Miseq instrument by our collaborator in University of Notre Dame. Barcodes and vector sequences were removed from the reads. The rest of the sequence of reads were mapped onto aligning regions in the sheep genome (oviAri1, UCSC Genome Database). Each integration locus was re-examined manually and PCR was done to verify accuracy.

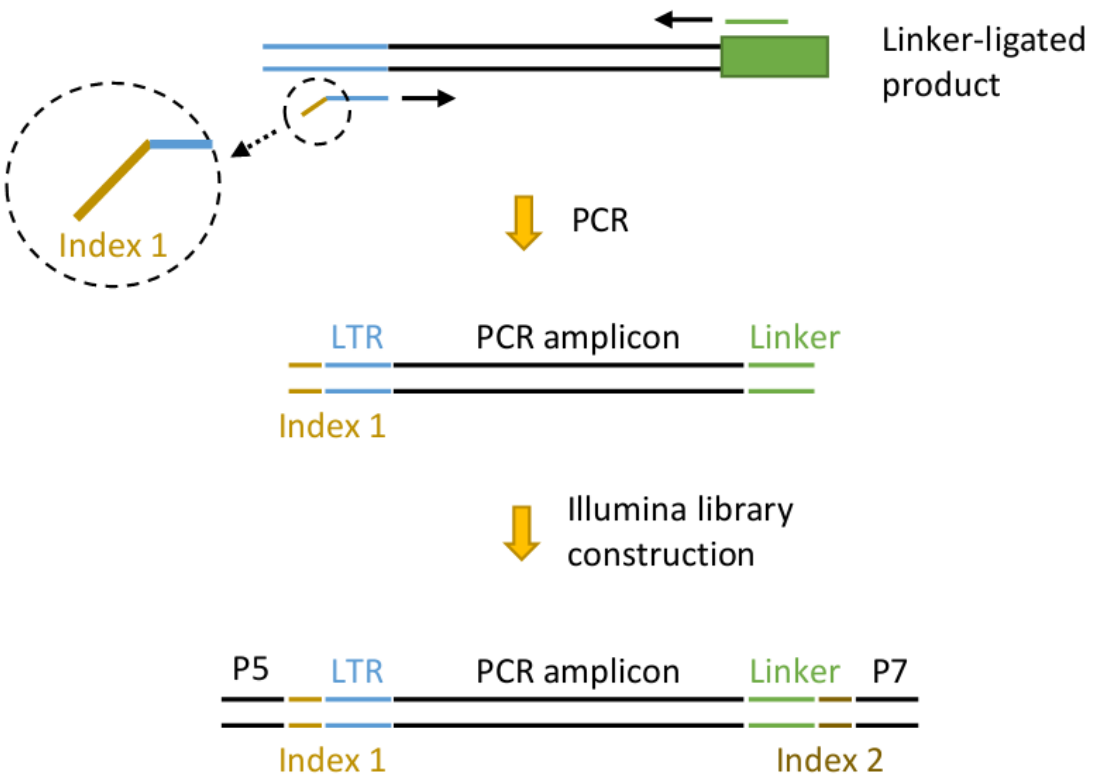

Figure 3.3. Schematic figure of introducing index by fusion primer. Six to eight bases indices were designed at the 5' end primers of round 2 nested PCR. While performing round 2 nested PCR, the first index could be introduced to the LTR end of the amplicon. Second index could be introduced during library preparation. P5 and P7 are the sequences required for next generation sequencing.

Table 3.1. Index sequence corresponding to different animals and tissues

Animal	Organ	Index Sequence
498-1	Kidney	AGTTCC
	Skin	ATGAGC
	Placenta	CGCGTC
	Intestine	GATACA
714-1	Lung	AAGCCGC
	Kidney	CAAGAAC
	Skin	TGACGAC
709-1	Kidney	GGTAGC
	Placenta	TCATTC
	Liver	ATCTTAC
	Skin	TGGTCT
709-2	Intestine	CTCTCTAT
	Uterus	TATCCTCT
	Placentome	AGAGTAGA
	Kidney	AAGGAGTA
	Skin	TGTCGT

CHAPTER 4. RESULTS

4.1 Evaluating the Pattern of LAM-PCR Product from Different Germ Layers

In order to test my hypothesis that the integration can occur after multiple cell divisions, we chose LAM-PCR to evaluate integration sites in organs from different germ layers. LAM-PCR is the common technique for finding integration sites by amplifying the vector-genome junction region. Compared to other methods to track vector insertion sites, such as inverse PCR (IPCR) and ligation-mediated PCR (LM-PCR), LAM-PCR is more sensitive such that the requirement for DNA amount is very low (down to 0.01 ng) for each reaction. LAM-PCR utilizes restriction enzymes resulting in uniquely sized band for each integration site. The products of LAM-PCR can then be visualized on a gel. We can see if there is any different integration site by comparing the LAM-PCR product pattern of each organ. Here we should state that in every LAM-PCR reaction, there will be one internal control band been seen on a gel since the primers used in LAM-PCR was designed to anneal to LTR region, which is identical on both sides of provirus (Figure 4.1). In this study, we conducted LAM-PCR on four transgenic sheep fetuses, which consisted of 34 tissue samples.

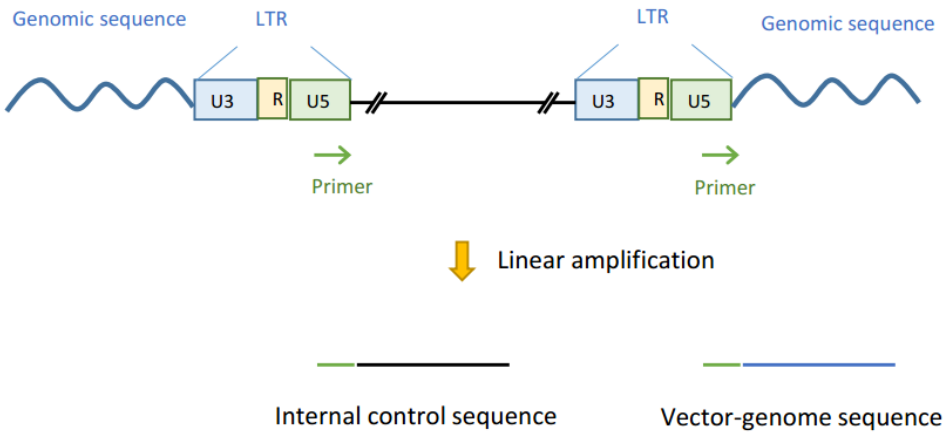

Figure 4.1. Simplified schematic figure of provirus structure. In order to get close to genomic sequence, primers were designed on LTR region, resulting in two kinds of products: (1) internal control sequence, and (2) vector-genome sequence.

In 709-1(Figure 4.2), we observed that all tissues shared the same pattern with three major bands except interplacentomal uterus sample, which was not part of fetal tissues. In 709-2(Figure 4.3), the product patterns of uterus and placentome were different from other samples. These differences were expected because they were not fetal tissues. In animal 498-1 (Figure 4.4), all tissues shared the same pattern that with three major bands in between 200 to 300 bp in size. In animal 714-1(Figure 4.5) we found that all tissues shared the same pattern to each other. Although there was only one major band observed in this animal, there were several faint bands in some tissues, which might indicate other possible integration sites. In this LAM-PCR experiment, we did not see any different pattern among fetal tissues in the same animal, indicating that there were common integration sites in all tissues we examined in the same animal. This might suggest that integration occurred potentially at single-cell stage. Further investigation of exact integration sites is required to confirm this hypothesis.

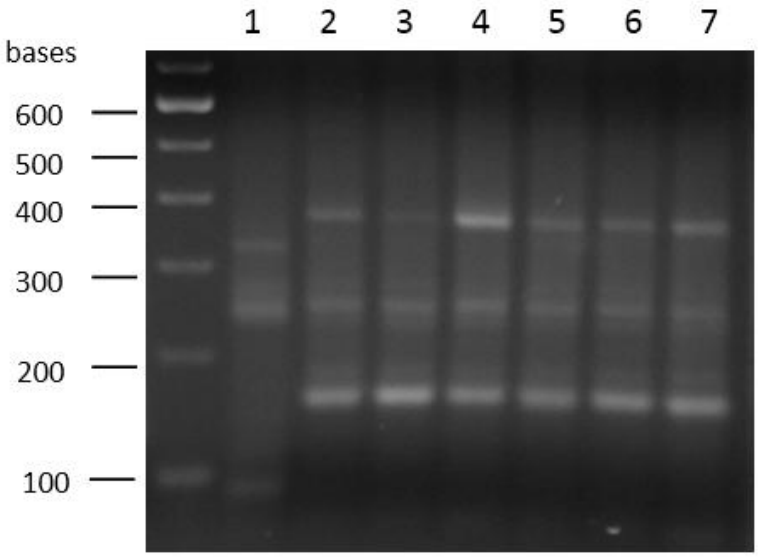

Figure 4.2.
LAM-PCR products of transgenic sheep fetal tissues-animal 709-1. (1)interplacentomal uterus, (2)liver, (3)placenta, (4)placentome, (5)gonad, (6)kidney, (7)heart.

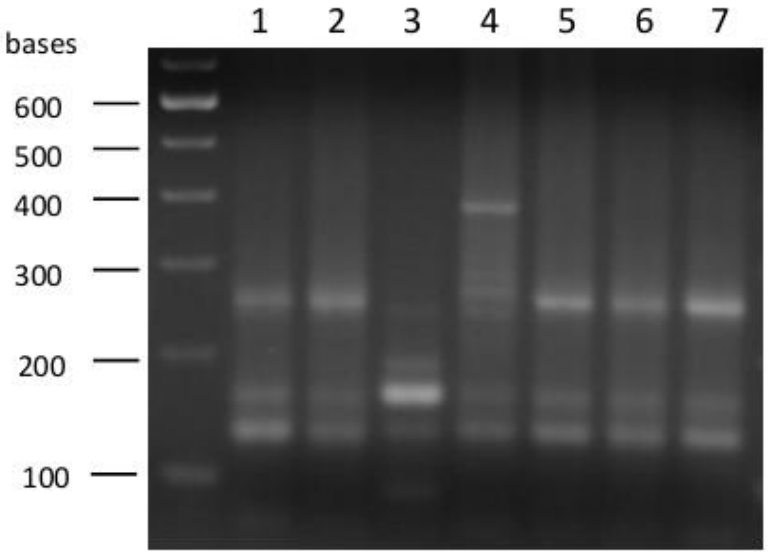

Figure 4.3. LAM-PCR products of transgenic sheep fetal tissues-animal 709-2. (1)lung, (2)intestine, (3)uterus, (4)placentome, (5)heart, (6)liver, (7)kidney.

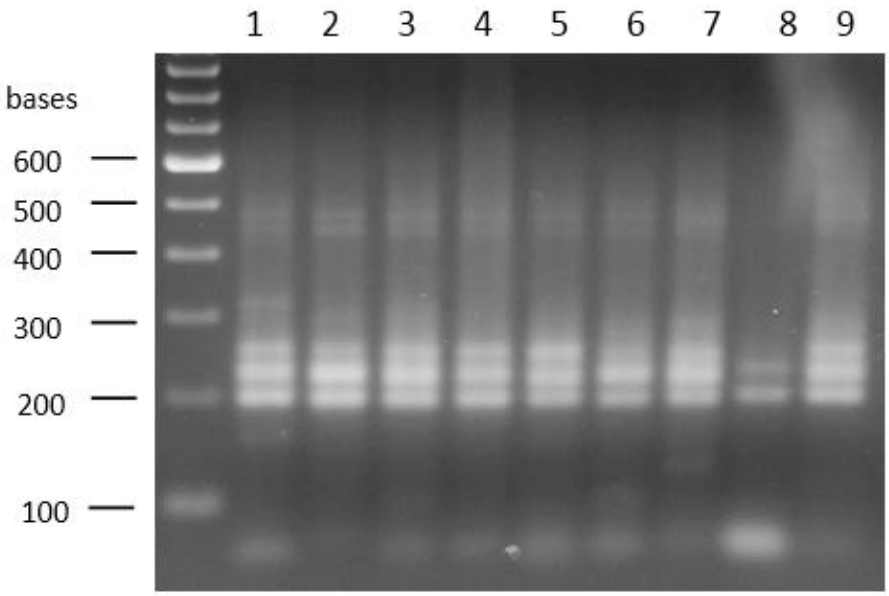

Figure 4.4. LAM-PCR products of transgenic sheep fetal tissues-animal 498-1.
(1)intestine, (2)placenta, (3)skin, (4)testis, (5)skeletal muscle, (6)lung, (7)liver, (8)kidney, (9)heart.

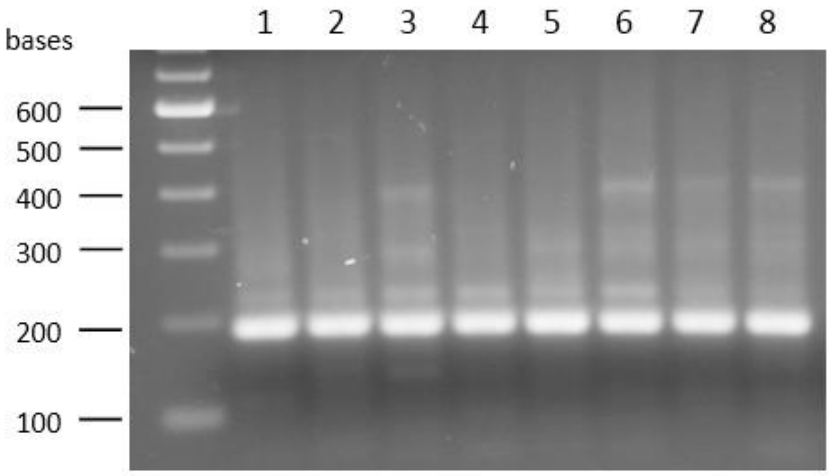

Figure 4.5. LAM-PCR products of transgenic sheep fetal tissues-animal 714-1. (1)heart, (2)liver, (3)lung, (4)kidney, (5)intestine, (6)skeletal muscle, (7)testis, (8)blood.

4.2 Localizing Exact Integration Sites by High-Throughput Sequencing Technology

In order to further identify the exact location of the integration sites, we utilized high-throughput sequencing technology. Among many platforms, we chose Miseq, launched by Illumina, to carry out this work. Miseq is a powerful platform that can yield sequences up to 2 Giga bases per run. In addition, the low error rate (0.8 %) is another attractive feature. Moreover, up to 96 samples can be run at once when samples are barcoded. This is the main reason why we chose Miseq as the high-throughput sequencing platform.

Among these LAM-PCR products, we collected samples with different patterns as well as samples derived from different germ layers. In order to add barcodes to DNA samples, we performed PCR using primers with different barcodes on intermediate products in the step before final exponential PCR in LAM-PCR. After quantifying, we mixed the same amount (100 ng) of final PCR products into one tube for next-generation sequencing. The library preparation, cluster generation and sequencing were done by our collaborator at the University of Notre Dame.

4.3 Comparing the Integration Sites between Organs

After getting sequencing data back, we sorted sequencing reads by barcodes to identify the data for a specific sample. We removed the reads without LTR sequence in the 5' end of the reads, which might have been the product of non-specific amplification. In order to get the genome sequence adjacent to vector sequence, we trimmed LTR

sequences as well as adapter sequences that were added in by the LAM-PCR experiment. To identify the integration sites, we applied "BLAT" using qualified reads against sheep genome database (version: oviAri1) on UCSC genome browser. Here we listed the top two groups of sequencing reads of every sample we sent for sequencing (Table 4.1).

For 709-1, the most abundant group of reads in all the tissues except interplacentomal uterus could be mapped to Chromosome 8 with 94% identity. For 709-2, the reads of intestine, kidney and skin could be mapped to Chromosome 1 with 100% identity. For uterus in 709-2, although the most abundant group of reads could be mapped to ChrX: 76405016, the undefined sequences ("N" base) near this position indicated this region of sheep genome was not well confirmed by enough sequencing data.

In animal 498-1, we found that the most abundant group of reads (from 33% to 55%) in every tissue could not be mapped to any genomic region in database. The second abundant group of reads had lower identity (81%) against the position ChrX: 50970245. Animal 714-1, 709-1 and 709-2 also had high percentages of reads with low identity against the same position as 498-1. Since the integration of lentiviral vectors is relatively random, it was less likely that these four animals had the same integration site. In addition, the reads with low identity against position ChrX:50970245 also suggested that these reads might be the result of non-specific amplification.

For animal 714-1, we observed the same phenomenon as 498-1 that the same group of reads in all the samples could not be mapped to genome. The result indicated that the

integration sites were the same in tissues derived from different layers of the same animal. In addition, the sheep genome database is not well-developed so that some of our samples had integration sites but the exact location of the integration could not be mapped to the genome.

Table 4.1. Potential integration sites in different tissues of each animal

498-1	Hyp. IS	# of reads	% of Total	Chromosome	Position	Identity
Kidney	IS1	55144	55.24		no match	
	IS2	13820	13.84	X	50970245	81%
Skin	IS1	21912	33.56		no match	
	IS2	10638	16.29	X	50970245	81%
Placenta	IS1	7781	52.02		no match	
	IS2	2320	15.51	X	50970245	81%
Intestine	IS1	13105	41.22		no match	
	IS2	5070	15.95	X	50970245	81%

714-1	Hyp. IS	# of reads	% of Total	Chromosome	Position	Identity
Lung	IS1	75033	65.39	X	50970245	81%
	IS2	8710	7.59		no match	
Kidney	IS1	37089	59.43	X	50970245	81%
	IS2	4727	7.57		no match	
Skin	IS1	22788	68.99	X	50970245	81%
	IS2	2780	8.42		no match	

709-1	Hyp. IS	# of reads	% of Total	Chromosome	Position	Identity
Kidney	IS1	11312	45.48	8	23921255	94%
	IS2	3812	15.33	X	50970245	81%
Placenta	IS1	11037	48.67	8	23921255	94%
	IS2	3407	15.02	X	50970245	81%
Inter uterus	IS1	5306	23.37	X	50970245	81%
	IS2	3507	15.44		no match	
Skin	IS1	13105	41.22	8	23921522	94%
	IS2	5070	15.95	X	50970245	81%

Table 4.1. Cont.

709-2	Hyp. IS	# of reads	% of Total	Chromosome	Position	Identity
Intestine	IS1	1189	16.9	X	50970245	81%
	IS2	1095	15.57	1	26891712	100%
Uterus	IS1	50703	52.28	X	76405016	*
	IS2	10068	10.38	X	50970245	81%
Placentome	IS1	7361	19.63	X	50970245	81%
	IS2	5577	14.87		no match	
Kidney	IS1	12161	16.49	1	26891712	100%
	IS2	11106	15.06	X	50970245	81%
Skin	IS1	3085	19.38	X	50970245	81%
	IS2	2512	15.78	1	26891712	100%

4.4 Verifying the Integration Sites by Conventional PCR

In order to confirm the integration sites we found from NGS data, we designed the primers for PCR that were located near (several hundred bases away) possible integration site (Table 4.2). After finding possible integration sites in tissue samples, we performed PCR on all available samples in the same animal. For the hypothetical common integration site in all four animals, we did not see any amplified PCR product on gel (data not shown). Despite the identity of the integration site on Chromosome 8 in 709-1 is only 94%, we could still observe the expected band (579 bp) in all the tissues except interplacentomal uterus (Figure 4.6, Table 4.3). We could also observe the expected band (487 bp) in all tissues except uterus and placentome of 709-2 (Figure 4.7, Table 4.3). Besides this, we also conducted PCR on 709-2's another integration site we found in a previous study, and the expected band (288 bp) could also be seen in all tissues except uterus (Fig. 4.8, Table 4.3). In addition, we also screened the integration sites which could be found in other animals (411-1, 536-1) in previous study. PCR product of all tissues of 411-1 could be observed (435 bp)(Figure 4.9, Table 4.3). Interestingly, in 536-1 we found that there were two bands in the placenta sample and a single band in other samples (Figure 4.10). The unique band in placenta was expected(372 bp), but not the band that appeared in all tissues.

Table 4.2. Confirmation primer list

Animal	Chromosome	Strand	Position	Confirmation Primer Sequence	Expected Length (bp)
714-1	ChrX	+	50970245	TCTTAGCATATAATCAGGCAATGG	252
709-1	Chr8	+	23921254	GGTCCTGAGGGGAGTATGGT	579
709-2	Chr1	+	26891711	TGGGAAAACTGAGGATTTGG	487
709-2	ChrX	-	83102460	GTGTCAAGACCCGGTAGGAA	288
411-1	Chr6	-	125302589	GGAAGACTCTGGGAGTGCTG	435
536-1	Chr4	+	49573928	ATTTAAGGCGGGGGTTCAGT	372

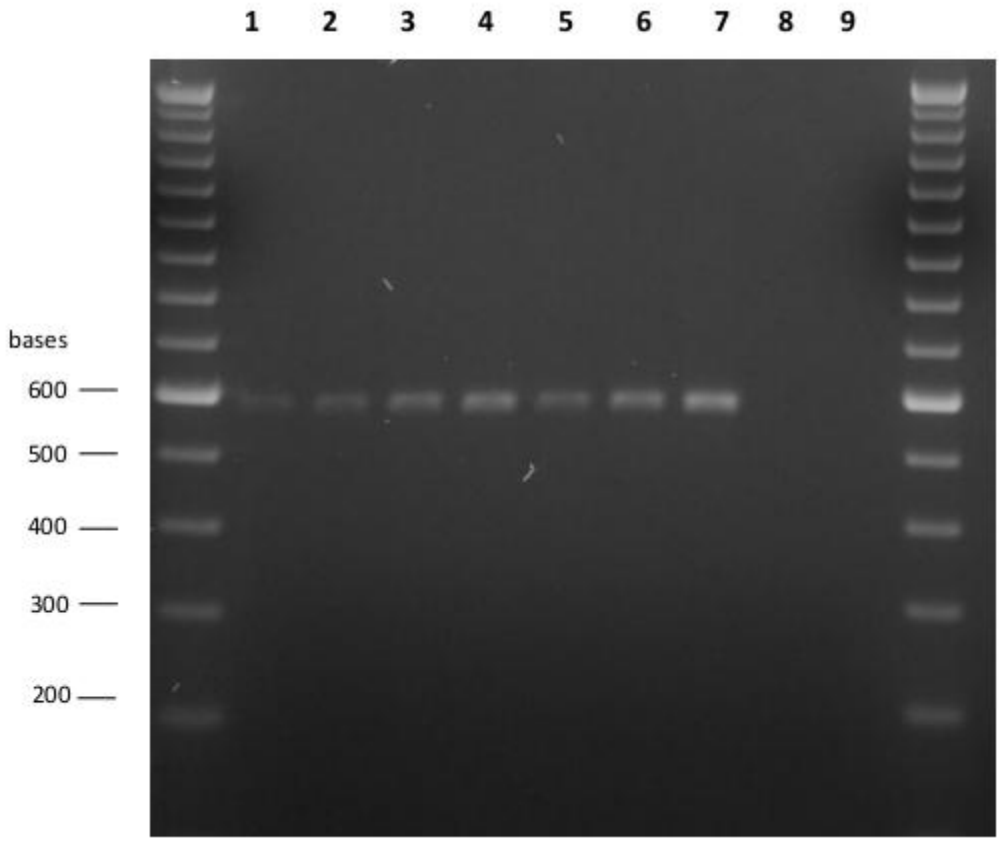

Figure 4.6. PCR to confirm integration site-animal 709-1. (1)heart, (2)kidney, (3)skin, (4)gonad, (5)placentome, (6)placenta, (7)liver, (8)interplacentomal uterus, (9)negative control.

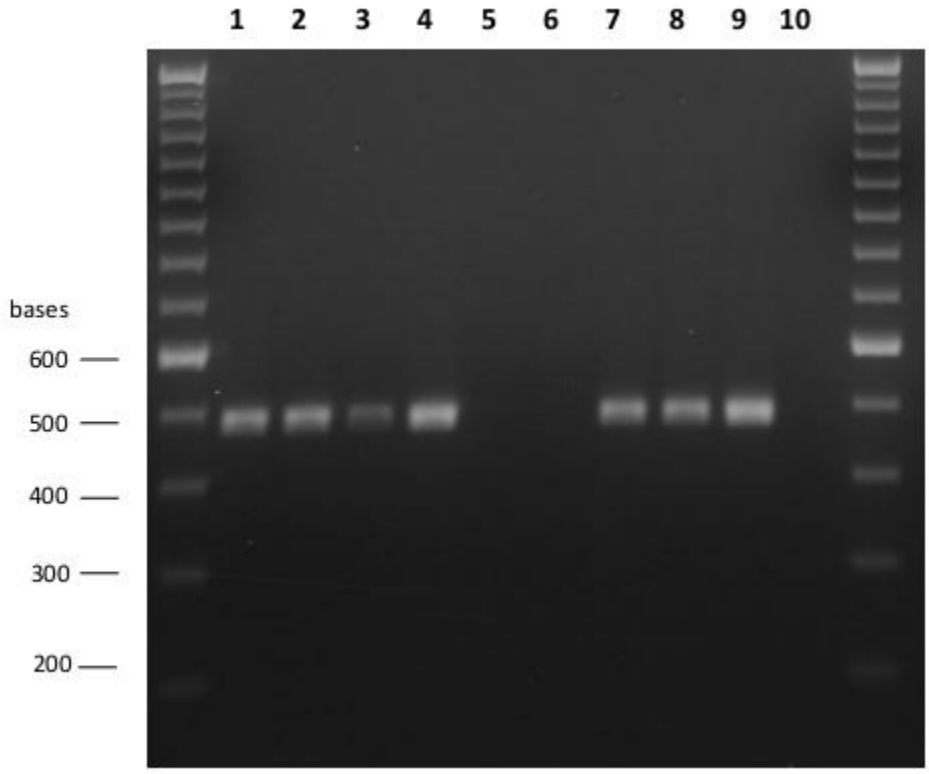

Figure 4.7. PCR to confirm integration site-animal 709-2(IS1). (1)lung, (2)intestine, (3)skin, (4)placenta, (5)uterus, (6)placentome, (7)heart, (8)liver, (9)kidney, (10)negative control.

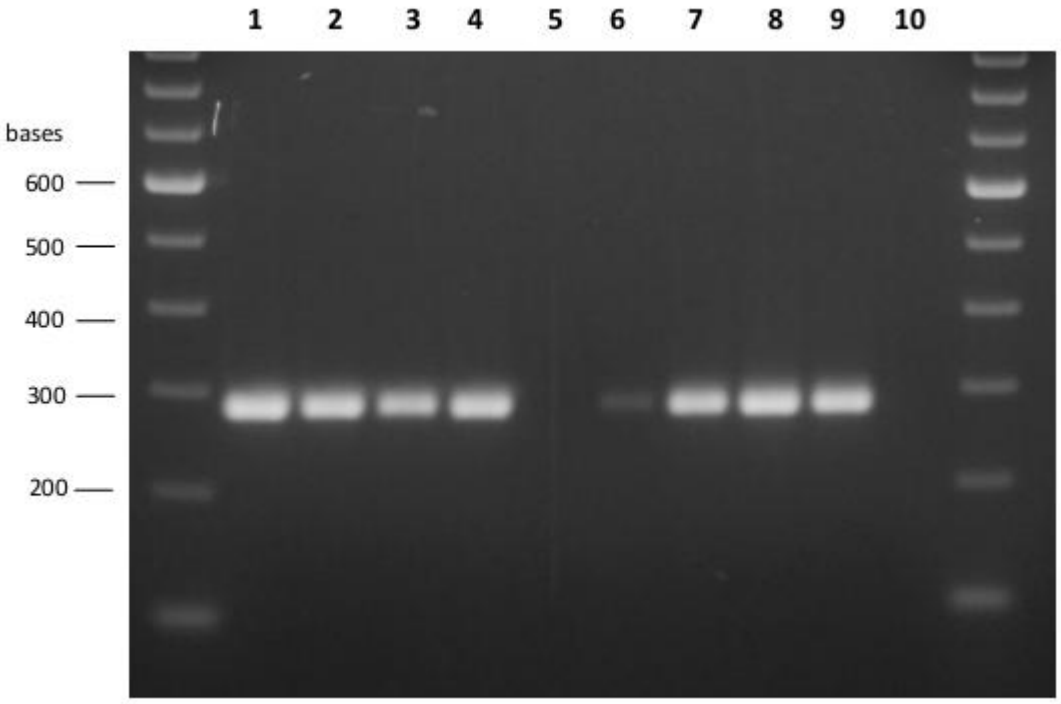

Figure 4.8. PCR to confirm integration site-animal 709-2(IS2). (1)lung, (2)intestine, (3)skin, (4)placenta, (5)uterus, (6)placentome, (7)heart, (8)liver, (9)kidney, (10)negative control.

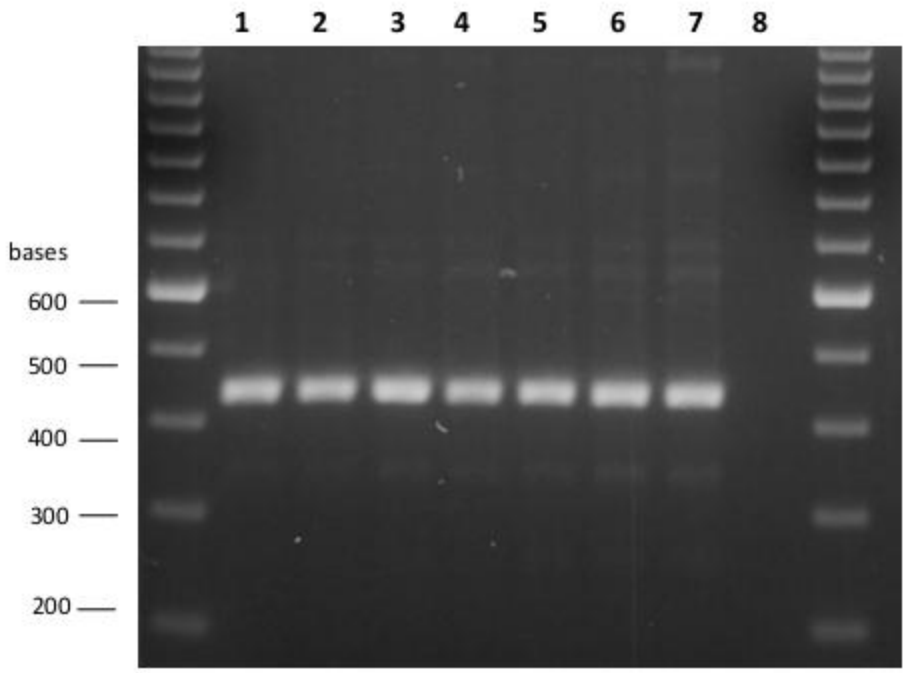

Figure 4.9. PCR to confirm integration site-animal411-1. (1)heart, (2)liver, (3)lung, (4)kidney, (5)intestine, (6)skin, (7)placenta, (8)negative control.

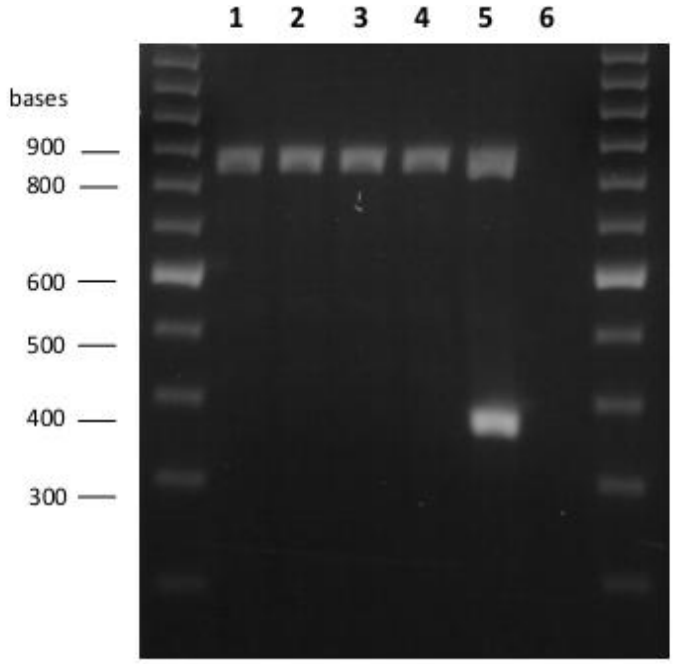

Figure 4.10. PCR to confirm integration site-animal 536-1. (1)heart, (2)lung, (3)skeletal muscle, (4)gonad, (5)intestine, (6)negative control.

Table 4.3. Integration sites confirmed by conventional PCR

Animal	Heart	Liver	Lung	Kidney	Intestine	Sk. Muscle	Skin	Gonad	Blood	Placenta	Placentome	Inter. uterus	Uterus	Chr.	Position
498-1	-	-	-	-	-	-	-	-	NA	-	NA	NA	NA	X	50970245
714-1	-	-	-	-	-	-	-	-	-	NA	NA	NA	NA	X	50970245
709-1	+	+	NA	+	NA	NA	+	+	NA	+	+	NA	-	8	23921255
709-1	-	-	NA	-	NA	NA	-	-	NA	-	-	NA	-	X	50970245
709-2	-	-	-	-	-	NA	-	NA	NA	-	-	-	NA	X	50970245
709-2	+	+	+	+	+	NA	+	NA	NA	+	-	-	NA	1	26891712
709-2	+	+	+	+	+	NA	+	NA	NA	+	+	-	NA	X	83102460
411-1	+	+	+	+	+	NA	+	NA	NA	+	NA	NA	NA	6	125302589
536-1	-	NA	-	NA	NA	-	NA	-	NA	+	NA	NA	NA	4	49573928

4.5 Examining the Genes near Integration Sites

In order to examine for possible bias in the site of integration, we searched sheep database to compare integration site in different animals. Since one of 709-2 integration sites is on Chromosome X, which has no gene information on the database, we could not tell if this integration site was located in a gene region. Then, we searched the genes closest to integration sites from either direction. We found that all of the genes were very far away from integration sites (> 2 Mega bases), suggesting that the provirus would not influence expression of those genes (Table 4.4). We need to notice, however, that there might be some genes located even closer to integration site but not yet annotated in the sheep genome database.

Table 4.4. Gene ontology analysis of confirmed integration sites

Animal	Chromosome	Strand	Position	Gene name	Strand	Position	Distance
411-1	chr6	-	125302589	RAB28	-	123049372	2.3 MB
				CNO	+	127642942	2.3 MB
536-1	chr4	+	49573928	RPL23A	-	45570293	4 MB
				SLC26A3	-	51618165	2 MB
709-1	chr8	+	23921254	ASF1A	-	21684001	2.2 MB
				FOXO3	-	31032239	7.2 MB
709-2	chr1	+	26891711	UQCRH	+	20795962	6.1 MB
				PRKAA2	+	31664555	4.8 MB
709-2	chrX	-	83102460				N/A

CHAPTER 5. DISCUSSION

The applications of gene transfer technology to farm animals have many benefits to agriculture and animal health. Recently developed lentiviral vector systems can undoubtedly contribute to this field dramatically due to their high efficient gene transfer. Many studies have proven that utilizing lentiviral vectors as delivery vehicles can increase the efficiency from several folds even up to several tens of times[2]. The major concern of applying this technology is safety. In previous studies[37] in our lab, it has been confirmed that there is no RCL observed in transgenic fetuses, lambs as well as surrogate mothers. However, the timing for integration event to occur after microinjection remains mysterious. We evaluated the vector copy number analysis and found that although the majority of animals had one or more copy numbers, some animals had less than one copy. This phenomenon implies that the vector integration might occur after several cell divisions at least in some animals.

To confirm our hypothesis, we selected four animals with different copy numbers (498-1: 0.4 copier/cell; 714-1: 0.45 copier/cell; 709-1: 3 copier/cell; 709-2: 1.3 copier/cell) calculated based on qPCR result to perform LAM-PCR. The preliminary data of LAM-PCR products on the gel showed no difference between fetal tissues in each animal. This

indicated that the integration event happened in an early stage, especially for 709-1 and 709-2 that it might occur in one-cell embryo stage.

We further collected LAM-PCR products for next-generation sequencing to find out the integration sites in each animal. In the integration site analysis, we found the all four animals had a high proportion of reads against the position ChrX: 50970245, suggesting that this might be the result of non-specific amplification against homologous sequences on sheep genome. We confirmed this hypothesis by PCR using primer annealing to this region along with vector primer. No PCR product was observed. For the other possible integration sites which were able to be analyzed, only one integration site with an identity higher than 90% could be found in animal 709-1 and 709-2, which is different from the copy number (709-1: 3 copier/cell; 709-2: 1.3 copier/cell) we observed in qPCR result of a previous study. It should be stated here that the LAM-PCR method had some limitations that should be considered when explaining our data. First, the use of restriction enzyme; LAM-PCR utilizes restriction enzyme to cut the flanking genomic sequence outside the vector sequence. If the flanking sequence is too short (<20 bp) after cutting, the length of the sequence would be insufficient to identify genomic location. If the distance of the restriction site is too far from the LTR-genomic junction the amplification reaction would likely not reach the genomic cut site and the fragment would not be amplified[38]. Secondly, the biased distributions of CpG in mammalian DNA is another factor to influence discovery of integration site by LAM-PCR. The percentage of GC content near integration site will influence restriction enzyme

efficiency as well as the PCR efficiency[39]. In addition, the PCR efficiency will also be influenced by inverted repeat sequences within the amplicon[40]. These limitations all influence the detection of an integration site, so that only a portion of integration sites can be found when utilizing LAM-PCR.

Nevertheless, both integration sites in 709-1 and 709-2 can be confirmed by PCR (Figure 4.6, 4.7). In addition, another integration site in 709-2 had been found in a previous study has also been confirmed by PCR in this study (Figure 4.8). Since the copy number is greater than 1 copy per cell and the integration sites can be detected in all fetal tissues and the placenta, this finding suggesting that the integration occurred early in embryogenesis, possibly at the one-cell embryo stage.

Furthermore, we also verified integration sites in two other animals (411-1, 536-1) which had been found in a previous study. Although in 411-1 we can detect integration site in all fetal tissues, the copy number (1 copy/10 cells) estimated by previous qPCR result suggested that the integration occurred after several cell divisions and led to mosaicism. Interestingly, we found that in animal 536-1 (Figure 4.10) there were two bands from the placenta tissue but only one band in other tissues. Based on the size of plasmid we used to transduce cells, it is less likely that there were two integration sites so close to each other. We then analyzed these two bands by Sanger sequencing. We found that the larger band was a false negative since the BLAST result showed that the primers we used had homologous sequences within the target region and the size of

this region was similar to the PCR product on gel (~850 bp, Figure 5.1). We later analyzed the smaller band. We found that this was the true integration site and solely in placenta. In mammalian development, embryonic cells separate into two groups of cells-the inner cell mass and trophectoderm, the latter will give rise to placenta after implantation[41]. This might be able to explain why in 536-1 only placenta had this integration site. This result supports my hypothesis that lentiviral vector integration can also occur after several cell divisions. Another study to generate transgenic rabbits by lentiviral vectors also supported this hypothesis[42]. They observed that the transgenic founder rabbits showed mosaic pattern of transgene expression. It should be noticed that in the early embryonic development of a rabbit it took only 11 hours after fertilization to reach 4-cell embryo stage, compared to rodent, sheep, swine and monkey embryos for which it took 30-40 hours. The phenomenon of mosaic transgenic founder rabbits combined with rapid early embryonic development might suggest that the transgene integration occurred in certain time after fertilization, probably after several cell divisions.

For gene ontology analysis, although we found some genes are cancer-related genes, the distance between integration sites and genes are too far away for any interaction to be considered. There should be some genes located closer to the integration sites and the closest genes identified in our analysis, but these were not evident due to the incomplete annotation of the sheep genome. A well-annotated sheep genome database is needed for more accurate gene ontology analysis.

Figure 5.1. Primers homology sequence on sheep genome. The homology sequence of primers to genome led to false positive result of confirmation PCR.

In summary, LAM-PCR and sequencing data identified a common integration in the tissues of the animals studied, each integration site being unique for each animal. In certain animals (709-1 and 709-2) it appears that integration occurred shortly after injection into the single cell embryo. Our finding in animal 411-1 and 536-1 supported the hypothesis that the integration can also occurs after several cell divisions. It should be taken into consideration that the mosaicism means that only a portion of the cells will contain the vector and if the desired phenotype requires all of the offspring to express the vector then careful screening will be required to insure all cells in the animal contain the vector. This is particularly important if the animal is used to generate offspring (founder animal) to ensure the animal carries the transgene of interest in germ line cells. We also know from human gene therapy work that vector can integrate preferentially into different gene regions and can influence surrounding gene expression. The stage of differentiation (hematopoietic stem cells versus differentiated T cells) also influence the effect integration may play on altering cell growth. To determine if lentiviral vectors have a preferential site of integration, or how they alter cell growth in an embryo, additional animals and improved annotation of the sheep genome database will be required.

LIST OF REFERENCES

LIST OF REFERENCES

[1] W. a Kues and H. Niemann, “Advances in farm animal transgenesis.,” *Preventive Veterinary Medicine*, vol. 102, no. 2, pp. 146–56, Nov. 2011.

[2] M. J. McGrew, A. Sherman, F. M. Ellard, S. G. Lillico, H. J. Gilhooley, A. J. Kingsman, K. a Mitrophanous, and H. Sang, “Efficient production of germline transgenic chickens using lentiviral vectors.,” *EMBO Reports*, vol. 5, no. 7, pp. 728–33, Jul. 2004.

[3] A. Pfeifer, “Lentiviral transgenesis.,” *Transgenic Research*, vol. 13, no. 6, pp. 513–22, Dec. 2004.

[4] M. Schmidt, K. Schwarzwaelder, C. Bartholomae, K. Zaoui, C. Ball, I. Pilz, S. Braun, H. Glimm, and C. Von Kalle, “High-resolution insertion-site analysis by linear amplification – mediated PCR (LAM-PCR),” *Nature Methods*, vol. 4, no. 12, pp. 1051–1057, 2007.

[5] M. a Quail, M. Smith, P. Coupland, T. D. Otto, S. R. Harris, T. R. Connor, A. Bertoni, H. P. Swerdlow, and Y. Gu, “A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers.,” *BMC Genomics*, vol. 13, no. 1, p. 341, Jan. 2012.

[6] J. W. Gordon, G. a Scangos, D. J. Plotkin, J. a Barbosa, and F. H. Ruddle, “Genetic transformation of mouse embryos by microinjection of purified DNA.,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 77, no. 12, pp. 7380–4, Dec. 1980.

[7] R. L. B. R E Hammer, V G Pursel, C E Rexroad, R J Wall, D J Bolt, K M Ebert, R D Palmiter, “Production of transgenic rabbits, sheep and pigs by microinjection,” *Nature*, vol. 315, no. 20, pp. 680–683, 1985.

[8] C. B. a Whitelaw, “Transgenic livestock made easy.,” *Trends in Biotechnology*, vol. 22, no. 4, pp. 157–9; discussion 159–60, Apr. 2004.

[9] A. Hofmann, B. Kessler, S. Ewerling, M. Weppert, B. Vogg, H. Ludwig, M. Stojkovic, M. Boelhauve, G. Brem, E. Wolf, and A. Pfeifer, “Efficient transgenesis in farm animals by lentiviral vectors,” *EMBO Reports*, vol. 4, no. 11, pp. 1054–1058, Nov. 2003.

[10] A S. Moffat, “Improving gene transfer into livestock.,” *Science (New York, N.Y.)*, vol. 282, no. 5394, pp. 1619–20, Nov. 1998.

[11] K. Wells, K. Moore, and R. Wall, “Transgene vectors go retro.,” *Nature Biotechnology*, vol. 17, no. 1, pp. 25–6, Jan. 1999.

[12] K. Chang, J. Qian, M. Jiang, Y. Liu, C. Wu, C. Chen, C. Lai, H. Lo, C. Hsiao, L. Brown, J. B. Jr, H. Huang, P. Ho, P. Y. Shih, C. Yao, W. Lin, C. Chen, F. Wu, Y. Lin, J. Xu, and K. Wang, “Effective generation of transgenic pigs and mice by linker based sperm-mediated gene transfer.,” *BMC Biotechnology*, vol. 13, pp. 1–13, 2002.

[13] A. C. Perry, “Mammalian Transgenesis by Intracytoplasmic Sperm Injection,” *Science*, vol. 284, no. 5417, pp. 1180–1183, May 1999.

[14] A. E. Schnieke, “Human Factor IX Transgenic Sheep Produced by Transfer of Nuclei from Transfected Fetal Fibroblasts,” *Science*, vol. 278, no. 5346, pp. 2130–2133, Dec. 1997.

[15] A. Hofmann, V. Zakhartchenko, M. Weppert, H. Sebald, H. Wenigerkind, G. Brem, E. Wolf, and A. Pfeifer, “Generation of transgenic cattle by lentiviral gene transfer into oocytes.,” *Biology of Reproduction*, vol. 71, no. 2, pp. 405–9, Aug. 2004.

[16] G. B. K. Sathasivam, C. Hobbs, L. Mangiarini, A. Mahal, M. Turmaine, P. Doherty, S. Davies, “Transgenic models of Huntington’s disease,” *Philosophical Transactions of the Royal Society B*, vol. 354, pp. 963–969, 1999.

[17] C. B. Whitelaw, E. Farini, and J. Webster, “The changing role of cell culture in the generation of transgenic livestock.,” *Cytotechnology*, vol. 31, no. 1–2, pp. 3–8, Sep. 1999.

[18] J. Mátrai, M. K. L. Chuah, and T. VandenDriessche, “Recent advances in lentiviral vector development and applications.,” *Molecular Therapy : the Journal of the American Society of Gene Therapy*, vol. 18, no. 3, pp. 477–90, Mar. 2010.

[19] Y. Yi, M. J. Noh, and K. H. Lee, “Current advances in retroviral gene therapy.,” *Current Gene Therapy*, vol. 11, no. 3, pp. 218–28, Jun. 2011.

[20] L. Jarrosson-wuilleme, C. Goujon, J. Bernaud, D. Rigal, J. Darlix, and A. Cimarelli, “Transduction of Nondividing Human Macrophages with Gammaretrovirus-Derived Vectors,” *Journal of Virology*, vol. 80, no. 3, pp. 1152–1159, 2006.

[21] T. Nguyen, “Highly Efficient Lentiviral Vector-Mediated Transduction of Nondividing, Fully Reimplantable Primary Hepatocytes,” *Molecular Therapy*, vol. 6, no. 2, pp. 199–209, Aug. 2002.

[22] L. Naldini, “Lentiviruses as gene transfer agents for delivery to non-dividing cells.,” *Current Opinion in Biotechnology*, vol. 9, no. 5, pp. 457–63, Oct. 1998.

[23] C. Lois, E. J. Hong, S. Pease, E. J. Brown, and D. Baltimore, “Germline transmission and tissue-specific expression of transgenes delivered by lentiviral vectors.,” *Science (New York, N.Y.)*, vol. 295, no. 5556, pp. 868–72, Feb. 2002.

[24] H. Sang, “Prospects for transgenesis in the chick.,” *Mechanisms of Development*, vol. 121, no. 9, pp. 1179–86, Sep. 2004.

[25] F. W.-S. H. Tang, K. L. Kuhen, “Lentivirus replication and regulation.,” *Annual Review Genetics*, vol. 33, pp. 133–170, 1999.

[26] S. Lillico, D. Vasey, T. King, and B. Whitelaw, “Lentiviral transgenesis in livestock.,” *Transgenic Research*, vol. 20, no. 3, pp. 441–2, Jun. 2011.

[27] C. E. Thomas, A. Ehrhardt, and M. a Kay, “Progress and problems with the use of viral vectors for gene therapy.,” *Nature Reviews. Genetics*, vol. 4, no. 5, pp. 346–58, May 2003.

[28] I. Gruh, S. Wunderlich, M. Winkler, K. Schwanke, J. Heinke, U. Bl, R. Li, A. Haverich, and U. Martin, “Human CMV immediate-early enhancer : a useful tool to enhance cell-type-specific expression from lentiviral vectors,” *the Journal of Gene Medicine Research*, no. August 2007, pp. 21–32, 2008.

[29] D. T. R. Zufferey, D. Nagy, R. Mandel, L. Naldini, “Multiply attenuate lentiviral vector achieves efficient gene delivery in vivo,” *Nature Biotechnology*, vol. 15, pp. 871–875, 1997.

[30] T. Dull, R. Zufferey, M. Kelly, R. J. Mandel, M. Nguyen, D. Trono, and L. Naldini, “A third-generation lentivirus vector with a conditional packaging system.,” *Journal of Virology*, vol. 72, no. 11, pp. 8463–71, Nov. 1998.

[31] R. Zufferey, T. Dull, R. J. Mandel, a Bukovsky, D. Quiroz, L. Naldini, and D. Trono, “Self-inactivating lentivirus vector for safe and efficient in vivo gene delivery.,” *Journal of Virology*, vol. 72, no. 12, pp. 9873–80, Dec. 1998.

[32] H. Miyoshi, U. Blömer, M. Takahashi, F. H. Gage, and I. M. Verma, “Development of a self-inactivating lentivirus vector.,” *Journal of Virology*, vol. 72, no. 10, pp. 8150–7, Oct. 1998.

[33] T. Sakuma, M. a Barry, and Y. Ikeda, “Lentiviral vectors: basic to translational.,” *The Biochemical Journal*, vol. 443, no. 3, pp. 603–18, May 2012.

[34] M. G. Ott, M. Schmidt, K. Schwarzwaelder, S. Stein, U. Siler, U. Koehl, H. Glimm, K. Kühlcke, A. Schilz, H. Kunkel, S. Naundorf, A. Brinkmann, A. Deichmann, M. Fischer, C. Ball, I. Pilz, C. Dunbar, Y. Du, N. a Jenkins, N. G. Copeland, U. Lüthi, M. Hassan, A. J. Thrasher, D. Hoelzer, C. von Kalle, R. Seger, and M. Grez, "Correction of X-linked chronic granulomatous disease by gene therapy, augmented by insertional activation of MDS1-EVI1, PRDM16 or SETBP1.,"*Nature Medicine*, vol. 12, no. 4, pp. 401–9, May 2006.

[35] S. Hacein-Bey-Abina, C. Von Kalle, M. Schmidt, M. P. McCormack, N. Wulffraat, P. Leboulch, a Lim, C. S. Osborne, R. Pawliuk, E. Morillon, R. Sorensen, a Forster, P. Fraser, J. I. Cohen, G. de Saint Basile, I. Alexander, U. Wintergerst, T. Frebourg, a Aurias, D. Stoppa-Lyonnet, S. Romana, I. Radford-Weiss, F. Gross, F. Valensi, E. Delabesse, E. Macintyre, F. Sigaux, J. Soulier, L. E. Leiva, M. Wissler, C. Prinz, T. H. Rabbitts, F. Le Deist, a Fischer, and M. Cavazzana-Calvo, "LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1.," *Science (New York, N.Y.)*, vol. 302, no. 5644, pp. 415–9, Oct. 2003.

[36] K. Cornetta, J. Yao, A. Jasti, S. Koop, M. Douglas, D. Hsu, L. a Couture, T. Hawkins, and L. Duffy, "Replication-competent lentivirus analysis of clinical grade vector products.," *Molecular Therapy : the Journal of the American Society of Gene Therapy*, vol. 19, no. 3, pp. 557–66, Mar. 2011.

[37] K. Cornetta, K. Tessanne, C. Long, J. Yao, C. Satterfield, and M. Westhusin, "Transgenic sheep generated by lentiviral vectors: safety and integration analysis of surrogates and their offspring.," *Transgenic Research*, Nov. 2012.

[38] M. a Harkey, R. Kaul, M. a Jacobs, P. Kurre, D. Bovee, R. Levy, and C. A. Blau, "Multiarm high-throughput integration site detection: limitations of LAM-PCR technology and optimization for clonal analysis.," *Stem Cells and Development*, vol. 16, no. 3, pp. 381–92, Jun. 2007.

[39] R. Gabriel, R. Eckenberg, A. Paruzynski, C. C. Bartholomae, A. Nowrouzi, A. Arens, S. J. Howe, A. Recchia, C. Cattoglio, W. Wang, K. Faber, K. Schwarzwaelder, R. Kirsten, A. Deichmann, C. R. Ball, K. S. Balaggan, R. J. Yáñez-Muñoz, R. R. Ali, H. B. Gaspar, L. Biasco, A. Aiuti, D. Cesana, E. Montini, L. Naldini, O. Cohen-Haguenauer, F. Mavilio, A. J. Thrasher, H. Glimm, C. von Kalle, W. Saurin, and M. Schmidt, "Comprehensive genomic access to vector integration in clinical gene therapy.," *Nature Medicine*, vol. 15, no. 12, pp. 1431–6, Dec. 2009.

[40] G. P. Wang, A. Garrigue, A. Ciuffi, K. Ronen, J. Leipzig, C. Berry, C. Lagresle-Peyrou, F. Benjelloun, S. Hacein-Bey-Abina, A. Fischer, M. Cavazzana-Calvo, and F. D. Bushman, “DNA bar coding and pyrosequencing to analyze adverse events in therapeutic gene transfer.,” *Nucleic Acids Research*, vol. 36, no. 9, p. e49, May 2008.

[41] A. Ralston and J. Rossant, “The genetics of induced pluripotency.,” *Reproduction (Cambridge, England)*, vol. 139, no. 1, pp. 35–44, Jan. 2010.

[42] L. Hiripi, D. Negre, F.-L. Cosset, K. Kvell, T. Czömpöly, M. Baranyi, E. Gócza, O. Hoffmann, B. Bender, and Z. Bosze, “Transgenic rabbit production with simian immunodeficiency virus-derived lentiviral vector.,”*Transgenic Research*, vol. 19, no. 5, pp. 799–808, Oct. 2010.