4 Analysis of gene traps on chromosome 11

4.1 Introduction

In the previous chapter, I showed that I could efficiently generate homozygous gene-trap mutations on mouse chromosome 11. By Southern analysis, independent gene-trap events from different pools were identified according to the sizes of the proviral/host junction fragments generated by different restriction enzymes. A total of 66 different homozygous gene-trap events have been isolated. 146 ES cell clones representing these 66 gene traps were expanded, and DNA and RNA samples were used to identify the virus integration sites and the trapped exons.

Using the retroviral vector 5' gene-trap strategy, it is possible to identify the gene-trap locus by Splinkerette PCR, Inverse PCR and 5' RACE. All these methods have been tested in large-scale insertion mutagenesis experiments (Mikkers, Allen et al. 2002; Suzuki, Shen et al. 2002; Hansen, Floss et al. 2003). By comparison of the results of a pilot experiment, I decided to use a combination of Splinkerette PCR and 5' RACE for the large-scale experiment.

4.1.1 Splinkerette PCR versus inverse PCR

Inverse PCR is the original method to clone proviral/host junction fragments. First, a restriction endonuclease that cuts only once within the provirus is used to digest the genomic DNA. The completely digested DNA is then self-ligated to form circles at a low DNA concentration and the flanking fragments are amplified using proviral DNA specific primers. Nested PCR is performed to improve the sensitivity of the reaction and specificity of the inverse PCR products. The final products are cloned into plasmids to facilitate sequencing (Fig. 4-1). The inverse PCR conditions can be optimized to amplify fragments as large as 12 kb (Li, Shen et al. 1999). Obtaining large flanking fragments was a great advantage before the mouse genomic sequence was finished. The bigger the fragment, the better chances are that a sequencing result can match a known gene or EST sequence, and determine the proviral insertion site.

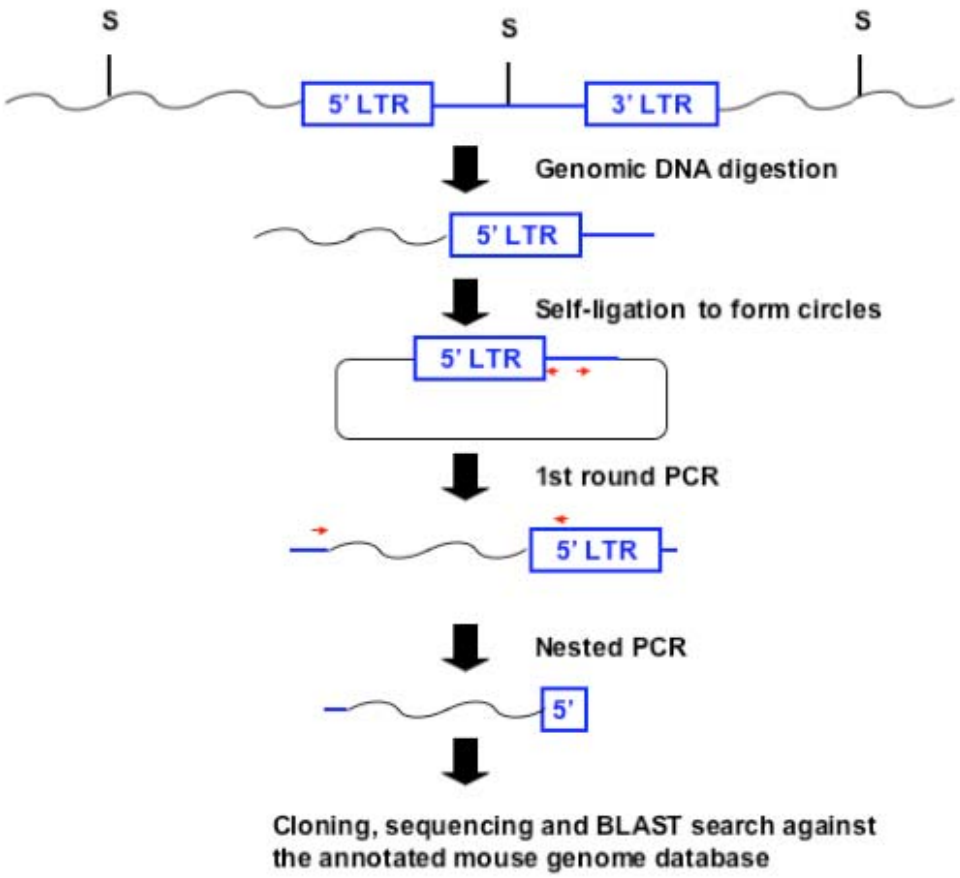


Fig. 4-1 Inverse PCR. First, a restriction enzyme is used to cut the genomic DNA. The completely digested genomic DNA is then self-ligated at low concentration to form circles. The flanking fragments are amplified using proviral DNA specific primers. Nested PCR is performed to improve the sensitivity and specificity of the Inverse PCR products. The final products are cloned into a plasmid to facilitate sequencing.

However, the inverse PCR strategy also has its disadvantages. First, to amplify a fragment big enough to determine the insertion site, a relatively rare cutting restriction enzyme is chosen for digestion of the genomic DNA. Long-range PCR amplification not only limits the recovery rate of proviral flanking sequences, it also significantly increases the cost. Second, the inverse PCR fragments need to be cloned for sequencing. This step limits the speed of the isolation of proviral-flanking sequences.

Splinkerette PCR was recently introduced as an alternative method to clone the proviral flanking fragments. A splinkerette is a pair of oligonucleotides that are partially complementary. One of the oligonucleotides contains a hairpin loop that prevents nonspecific PCR amplification by inhibiting new DNA strand synthesis from the adaptor. The other oligonucleotide contains the bind sites for the two primers used for the two rounds of PCR amplification (Fig. 4-2a). First, a restriction endonuclease that cuts only once within the 5' LTR of the provirus is used to digest the genomic DNA. The completely digested genomic DNA is then ligated to the splinkerette adaptor. The flanking proviral fragments are amplified using a pair of primers homologous to the splinkerette and the 5' LTR, respectively. Nested PCR is performed to improve the sensitivity and specificity of the PCR products (Fig. 4-2b). The final products are purified from the gel and sequenced directly (Mikkers, Allen et al. 2002).

Compared to inverse PCR, splinkerette PCR has some advantages. First, genomic DNA can be digested with a frequent cutter to get smaller fragments for amplification. Because the mouse genome is virtually complete, the size of the amplified flanking fragment is no longer a bottleneck for locus mapping. The development of the SSAHA (Sequence Search and Alignment by Hashing Algorithm) search engine makes it possible to find an exact or "almost exact" match between two sequences, even when the size of matched sequence is very small (Ning, Cox et al. 2001).

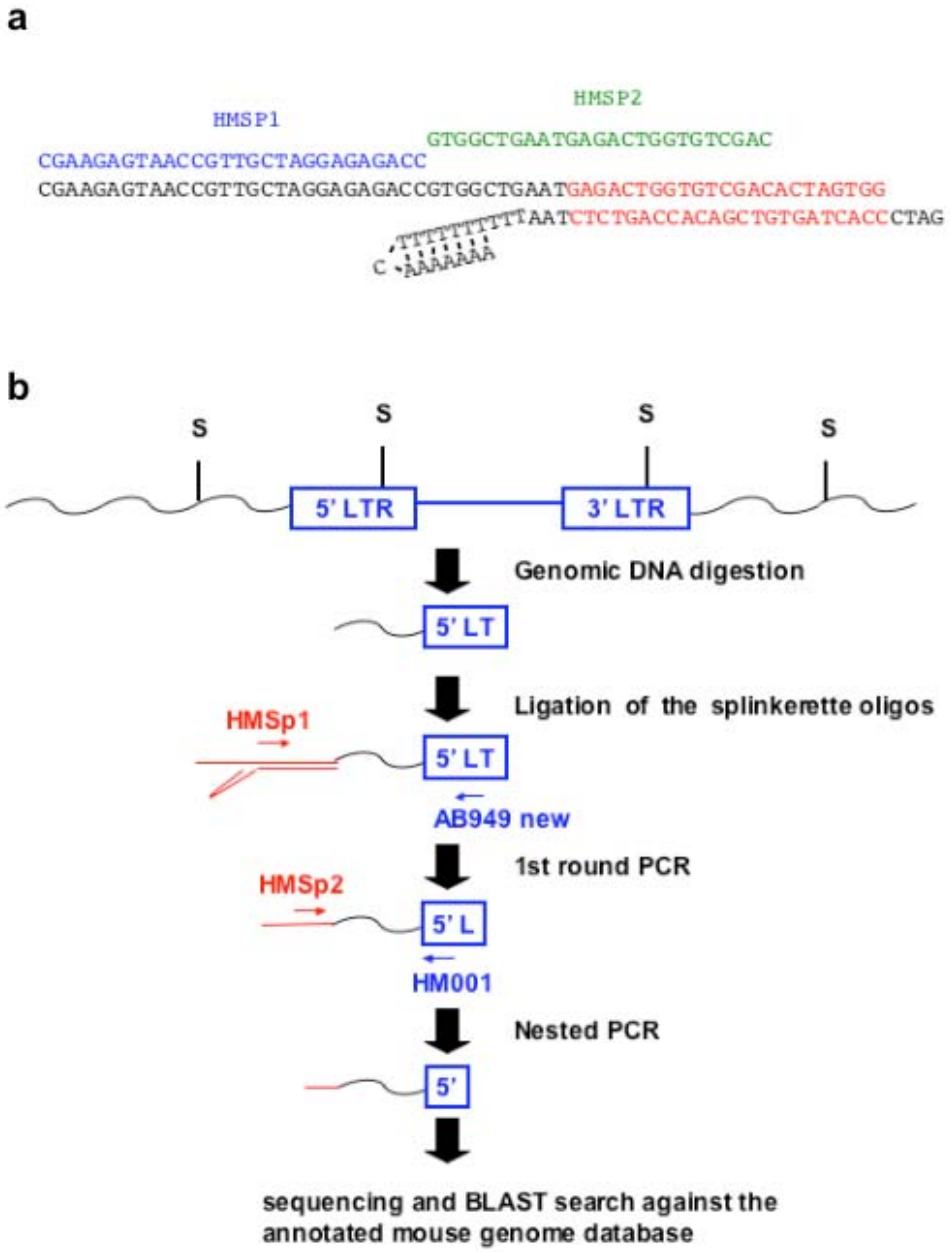

a
HMSP2
HMSP1
GTGGCTGAATGAGACTGGTGTCGAC
CGAAGAGTAACCGTTGCTAGGAGAGACC
CGAAGAGTAACCGTTGCTAGGAGAGACCGTGGCTGAATGAGACTGGTGTCGACACTAGTGG
TTTTTTTTTTAATCTCTGACCACAGCTGTGATCACCCTAG
C AAAAAAA
b
S
S
S
S
5' LTR
3' LTR
Genomic DNA digestion
5' LT
Ligation of the splinkerette oligos
HMSp1
5' LT
AB949 new
1st round PCR
HMSp2
5' L
HM001
Nested PCR
5'
sequencing and BLAST search against the
annotated mouse genome database

Fig. 4-2 Splinkerette PCR. a. The Structure of the splinkerette. A splinkerette is a pair of partially complementary oligos. One of the oligonucleotides contains a hairpin loop that prevents nonspecific PCR amplification by inhibiting new DNA strand synthesis from the adaptor. The other oligonucleotide contains the bind sites for the two primers used for the two rounds of PCR amplification. **b.** Splinkerette PCR. First, a restriction enzyme is used to cut the genomic DNA. The completely digested DNA is then ligated to the splinkerette adaptor. The flanking proviral fragments are PCR amplified. Nested PCR is performed to improve the sensitivity and specificity of the PCR products. The final products are purified from the gel and sequenced directly.

In principle, the PCR products generated by splinkerette PCR can be sequenced directly, but several background amplification products always coexist with the specific proviral insertion product because of endogenous viral sequences. So the PCR products need to be purified from a gel. This step has become the major bottleneck for splinkerette PCR.

4.1.2 5’ RACE

5’ Rapid Amplification of cDNA Ends (5’ RACE) is a method to amplify the 5’ region from an mRNA template between a defined internal site and the 5' end. To specifically amplify a rare template in a complex mixture usually requires two sequence-specific primers flanking the region of interest. This is not compatible with the need to amplify an unknown region with only one known end. 5' RACE methodologies offer a convenient way to solve this problem.

5' RACE, or “anchored” PCR, can be used to isolate and characterize 5' ends of low-copy mRNA templates. Although the 5’ RACE protocols vary from user to user, the general strategy is the same. First, a gene-specific primer is used for first strand cDNA synthesis. This step not only decreases the non-specific amplification, but also increases the possibility of obtaining the 5' end of a long mRNA template. Terminal deoxynucleotidyl transferase (TdT) is then used to add a homopolymeric tail to the 3' end of the cDNA. The 5’ end of the mRNA is then amplified using a pair of primers homologous to the homopolymeric tail and the internal anchor region, respectively. Nested PCR is performed to improve the product yield and specificity of the PCR product (Fig. 4-3). The 5’ RACE procedure can be utilized to amplify and characterize unknown coding sequences in gene-trap mutagenesis. It is an especially important technology for 5’ trapping strategy based on electroporation, because the flanking genomic fragments are difficult to isolate.

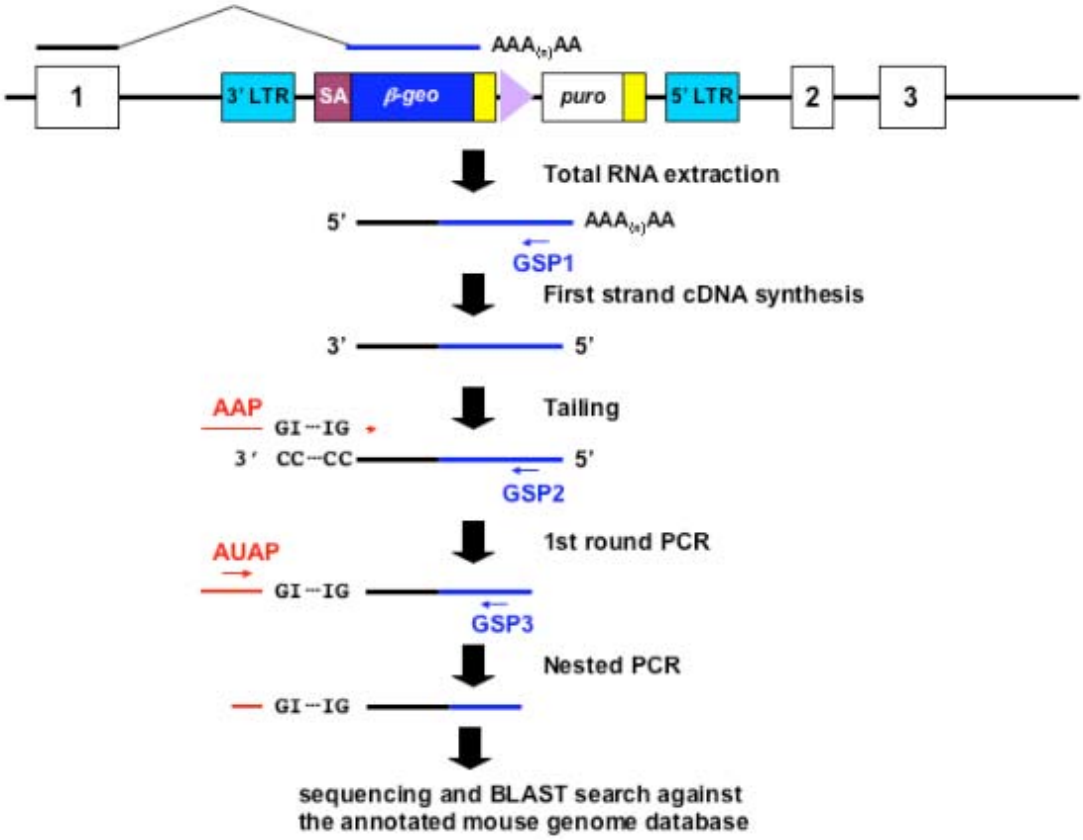


Fig. 4-3 5' RACE. First, a gene-specific primer is used for first strand cDNA synthesis. Terminal deoxynucleotidyl transferase (TdT) is then used to add homopolymeric tails to the 3' ends of the cDNA. The 5' end region of the mRNA is amplified by PCR. Nested PCR is performed to improve the sensitivity and specificity of the PCR products. SA, splicing acceptor; yellow box; bpA; purple arrow, *lox*P site, GSP, gene-specific primer; AAP, abridged anchor primer; AUAP, abridged universal amplification primer.

4.1.3 Distribution of the trapped genes on Chromosome 11

From the drug selection and Southern results, I determined that I had generated homozygous inversions on chromosome 11. Based on the knowledge of Cre efficiency over long genetic distances (Zheng, Sage et al. 2000), it was expected that a large number of the isolated homozygous gene traps would be within 10 Mb proximal and distal to the *E_2DH* locus (20 Mb in total). The results of the original regional trapping experiment (Wentland et al. unpublished data) has shown that 86% of the gene-traps were concentrated on the distal part of chromosome 11, and fell within a 43 Mb region surrounding the *E_2DH* locus. So the range of the regional trapping technology is much higher than expected.

Although I have chosen a different trapping and drug selection strategy for recovering the inversion events, I expected that the regional trapping efficiency should be comparable to the original regional trapping experiment. However, because I used a much larger pool to select out regional trapping events, I expected that I would not be able to isolate big inversions (the largest inversions isolated in the original regional trapping experiment is 82 Mb in size).

4.1.4 Orientation of transcription of the trapped genes

To generate an inversion, the two *lox*P sites must be in opposite orientation on the same chromosome. Therefore, the trapped genes should be transcribed from the antisense strand of the chromosome 11 (from telomere to centromere). The results of the original regional trapping experiment (Wentland et al. unpublished data) has shown that 17 out of the 21 trapped genes on chromosome 11 were transcribed from the antisense strand. However, the other four appeared to be transcribed from the sense strand (from centromere to telomere), but the drug resistance of these clones is the same as the other inversion clones. By fluorescent *in situ* hybridization (FISH), Wentland et al. (unpublished data) has found that the clones that have the expected drug resistance (HAT^R, $Puro^R$ and $G418^R$) but wrong transcription direction of the trapped gene, either have a duplication

chromosome and a wild-type chromosome, or balanced deletion/duplication chromosomes.

These results further confirmed the observations of the previous studies that large heterozygous chromosomal deletions will cause ES cell lethality (Su, Wang et al. 2000; Zheng, Sage et al. 2000). Cells with large heterozygous deletions can only survive if a second genetic change occurs to compensate for the loss of the genetic material caused by the deletion. Unbalanced deletions are rescued by a partial trisomy of two wild-type and one deletion chromosome. The majority of the *trans* recombination products will result in balanced deletion/duplication chromosomes.

4.2 Results

Splinkerette PCR was carried out for each of the expanded clones. To increase the possibility of generating PCR products with suitable size for sequencing, the genomic DNA extracted from the expanded ES cell clones was digested using *Sau*3AI, *Eco*RI or a combination of *Spe*I, *Xba*I and *Nhe*I, respectively. The derived splinkerette PCR products were separated on a 1% agarose gel. The specific PCR fragments were gel purified and sequenced using a pair of primers specific to the splinkerette and the 5' LTR of the retrovirus. The sequences were searched against the annotated mouse genome databases, Ensembl (http://www.ensembl.org/Mus_musculus) and NCBI (http://www.ncbi.nlm.nih.gov/BLAST).

5' RACE was carried out for at least one subclone from each group. RACE products were treated with Exonuclease I and Shrimp Alkaline Phosphatase to destroy the unused primers and dNTPs. The products were sequenced using a primer specific to the Splice Acceptor (SA) region of the trapping cassette. The sequences were also searched against the annotated mouse genome databases, Ensembl (http://www.ensembl.org/Mus_musculus) and NCBI (http://www.ncbi.nlm.nih.gov/BLAST).

5' RACE and/or Splinkerette PCR products were obtained from 49 of the 66 groups of independent recombination events (Table 4-1). The sequences from

44 of the groups matched sequences on chromosome 11. The other 5 groups matched sequences on chromosomes other than chromosome 11. For the remaining 17 groups, either no sequence information was obtained, or the sequence information from 5' RACE and Splinkerette PCR was inconsistent. So the exact identities of these clones are designated as "unknown".

Table 4.1: Gene trap loci.

Group	ES Cell Clone ID	Splinkerette product	RACE product	Invesion / Trapping	Gene Name	Position on the chromosome	Orientation of the transcirption
1	WW103-3A5	ENSMUSG00000018428		Inversion	Akap1 (A kinase (PRKA) anchor protein 1)	88501918 - 88535621 bp (88.5 Mb)	-
	WW103-3B3	ENSMUSG00000018428		Inversion	Akap1 (A kinase (PRKA) anchor protein 1)	88501918 - 88535621 bp (88.5 Mb)	-
	WW103-3B5			Inversion			
	WW103-3D7			Inversion			
2	WW103-3A8			Inversion			
	WW103-3D6	ENSMUSG00000035086		Inversion	Becn1 (coiled-coil, myosin-like BCL2-interacting protein)	100957075 - 100973390 bp (101.0 Mb)	-
3	WW103-3C4	ENSMUSG00000020737	ENSMUSG00000020737	Inversion	Hn1 (hematological and neurological expressed sequence 1)	115168484 - 115185516 bp (115.2 Mb)	-
	WW013-3D5	ENSMUSG00000020737		Inversion	Hn1 (hematological and neurological expressed sequence 1)	115168484 - 115185516 bp (115.2 Mb)	-
4	WW103-4A2			Inversion			
	WW103-4A10		ENSMUSG00000034520	Inversion	Gja7 (gap junction membrane channel protein alpha 7)	102470710 - 102490284 bp (102.5 Mb)	-
	WW103-4B3		ENSMUSG00000034520	Inversion/Trapping	Gja7 (gap junction membrane channel protein alpha 7)	102470710 - 102490284 bp (102.5 Mb)	-
	WW103-4B7			Inversion			
	WW103-4B11		ENSMUSG00000034520	Inversion	Gja7 (gap junction membrane channel protein alpha 7)	102470710 - 102490284 bp (102.5 Mb)	-
	WW103-4C5			Inversion/Trapping			
5	WW103-4A6	ENSMUSG00000020717		Inversion	Pecam1 (platelet/endothelial cell adhesion molecule 1)	106325342 - 106421753 bp (106.3 Mb)	-
6	WW103-4A12						
7	WW103-4B1	ENSMUSG00000010342		Trapping	Tex14 (testis expressed gene 14)	87006730 - 87157487 bp (87.0 Mb)	+
8	WW103-4D11			Inversion			
	WW103-4B11			Inversion			
9	WW103-4D3	ENSMUSG00000020935		Inversion	6720485C15Rik (RIKEN cDNA 6720485C15 gene)	102665181 - 102688272 bp (102.7 Mb)	-
10	WW103-5E2			Trapping			
	WW103-5E3	ENSMUSG00000020923		Inversion	Ubtf (upstream binding transcription factor, RNA polymerase I)	101977115 - 101990113 bp (102.0 Mb)	-
	WW103-5F1	ENSMUSG00000020923		Inversion/Trapping	Ubtf (upstream binding transcription factor, RNA polymerase I)	101977115 - 101990113 bp (102.0 Mb)	-
11	WW103-5E4	ENSMUSG00000013415		Inversion	Igf2bp1 (insulin-like growth factor 2, binding protein 1)	95633855 - 95677035 bp (95.6 Mb)	-
	WW103-5H2			Inversion			
12	WW103-5E5	unknown locuc on chromosome 11		Inversion	unknown	74223628 - 74223809 (74.2 Mb)	-
13	WW103-5F4	ENSMUSG00000013415		Inversion	Igf2bp1 (insulin-like growth factor 2, binding protein 1)	95633855 - 95677035 bp (95.6 Mb)	-
14	WW103-5F9	ENSMUSG00000056649	ENSMUSG00000056649	Inversion	2810410L24Rik (RIKEN cDNA 2810410L24 gene)	119856472 - 119857805 bp (119.9 Mb)	-
15	WW103-5G11			3 Bands?			
16	WW103-5H4			Inversion			
17	WW103-6E2	ENSMUSG00000014195		Inversion/Trapping	Dnajc7 (DnaJ (Hsp40) homolog, subfamily C, member 7)	100253974 - 100291291 bp (100.3 Mb)	-
	WW103-6E11	ENSMUSG00000014195		Inversion	Dnajc7 (DnaJ (Hsp40) homolog, subfamily C, member 7)	100253974 - 100291291 bp (100.3 Mb)	-
	WW103-6F7	ENSMUSG00000014195		Inversion	Dnajc7 (DnaJ (Hsp40) homolog, subfamily C, member 7)	100253974 - 100291291 bp (100.3 Mb)	-
18	WW103-6E5	Chr.X		strange band?			
	WW103-6F8	Chr.X		Trapping			
	WW103-6G4	Chr.X		Trapping			
19	WW103-6F3	ENSMUSG00000020715		Inversion/Trapping	Ern1 (endoplasmic reticulum (ER) to nucleus signalling 1)	106068745 - 106158921 bp (106.1 Mb)	-
	WW103-6H3	ENSMUSG00000020715		Inversion	Ern1 (endoplasmic reticulum (ER) to nucleus signalling 1)	106068745 - 106158921 bp (106.1 Mb)	-
	WW103-6H9	ENSMUSG00000020715		Inversion	Ern1 (endoplasmic reticulum (ER) to nucleus signalling 1)	106068745 - 106158921 bp (106.1 Mb)	-
20	WW103-6G1	ENSMUSG00000013415		Inversion	Igf2bp1 (insulin-like growth factor 2, binding protein 1)	95633855 - 95677035 bp (95.6 Mb)	-
21	WW103-6G2			Inversion/Trapping			
22	WW103-6H5			Inversion			
23	WW103-6G9	ENSMUSG00000001552		Inversion	Jup (junction plakoglobin)	100041744 - 100068862 bp (100.0 Mb)	-
24	WW103-7A2	ENSMUSG00000014195		Inversion	Dnajc7 (DnaJ (Hsp40) homolog, subfamily C, member 7)	100253974 - 100291291 bp (100.3 Mb)	-
	WW103-7A4	ENSMUSG00000014195		Inversion/Trapping	Dnajc7 (DnaJ (Hsp40) homolog, subfamily C, member 7)	100253974 - 100291291 bp (100.3 Mb)	-
	WW103-7C4	ENSMUSG00000014195	ENSMUSG00000014195	Inversion	Dnajc7 (DnaJ (Hsp40) homolog, subfamily C, member 7)	100253974 - 100291291 bp (100.3 Mb)	-
25	WW103-7A7			Inversion			
	WW103-7B5			Inversion/Trapping			
	WW103-7C7			Inversion			
26	WW103-7A9	Chr.19		3 bands			
27	WW103-7B10			Inversion			
28	WW103-8E3	ENSMUSG00000020717		Inversion	Pecam1 (platelet/endothelial cell adhesion molecule 1)	106325342 - 106421753 bp (106.3 Mb)	-
	WW103-8E6	ENSMUSG00000020717		Inversion	Pecam1 (platelet/endothelial cell adhesion molecule 1)	106325342 - 106421753 bp (106.3 Mb)	-
	WW103-8F5	ENSMUSG00000020717		Inversion	Pecam1 (platelet/endothelial cell adhesion molecule 1)	106325342 - 106421753 bp (106.3 Mb)	-
	WW103-8G9	ENSMUSG00000020717	ENSMUSG00000020717	Inversion	Pecam1 (platelet/endothelial cell adhesion molecule 1)	106325342 - 106421753 bp (106.3 Mb)	-
29	WW103-8E10			Inversion/Trapping			
	WW103-8H5			Inversion/Trapping			
	WW103-8H9			Inversion			
30	WW103-8E11			Inversion			
31	WW103-8F2	ENSMUSG00000014195		Inversion	Dnajc7 (DnaJ (Hsp40) homolog, subfamily C, member 7)	100253974 - 100291291 bp (100.3 Mb)	-
	WW103-8F6	ENSMUSG00000014195	ENSMUSG00000014195	Inversion/Trapping	Dnajc7 (DnaJ (Hsp40) homolog, subfamily C, member 7)	100253974 - 100291291 bp (100.3 Mb)	-
	WW103-8G11	ENSMUSG00000014195		Inversion	Dnajc7 (DnaJ (Hsp40) homolog, subfamily C, member 7)	100253974 - 100291291 bp (100.3 Mb)	-
32	WW103-8F8	ENSMUSG00000038268	ENSMUSG00000038268	Inversion	OVCA2 (Ovarian Cancer-associated Gene 2)	74788498 - 74803029 bp (74.8 Mb)	-
33	WW103-8H8			Inversion			
34	WW103-9A1			Inversion/Trapping			
	WW103-9A2						
	WW103-9A9	ENSMUSG00000020530		Inversion	Gqnbp2 (gametogenetin binding protein 2)	84434394 - 84472403 bp (84.4 Mb)	-
	WW103-9B4			Inversion			
	WW103-9D4	ENSMUSG00000020530	ENSMUSG00000020530	Inversion	Gqnbp2 (gametogenetin binding protein 2)	84434394 - 84472403 bp (84.4 Mb)	-
35	WW103-9A5			Trapping			
	WW103-9A10	ENSMUSG00000019173		Inversion/Trapping	Rab5c (RAB5C, member RAS oncogene family)	100386134 - 100409313 bp (100.4 Mb)	-
	WW103-9A12			Inversion			
	WW103-9C6	ENSMUSG00000019173		Inversion	Rab5c (RAB5C, member RAS oncogene family)	100386134 - 100409313 bp (100.4 Mb)	-
	WW103-9D6			Inversion			
36	WW103-9B9			Inversion/Trapping			
	WW103-9C4			Trapping			
37	WW103-9D3			Inversion			
38	WW103-10E1			Inversion			
	WW103-10E2			Inversion			
	WW103-10F5		ENSMUSG00000014195	Inversion	Dnajc7 (DnaJ (Hsp40) homolog, subfamily C, member 7)	100253974 - 100291291 bp (100.3 Mb)	-

Table 4.1 (cont): Gene trap loci.

Group	ES Cell Clone ID	Splinkerette product	RACE product	Invesion / Trapping	Gene Name	Position on the chromosome	Orientation of the transcirption
39	WW103-10E4	Chr.9		Trapping			
40	WW103-10H3			Inversion			
	WW103-10H10			Inversion			
41	WW103-11E2	Chr.6 ?		Inversion			
42	WW103-11E3	ENSMUSG00000051378	ENSMUSG00000051378	Inversion	3000004C01Rik (RIKEN cDNA 3000004C01 gene)	102576655 - 102596217 bp (102.6 Mb)	-
	WW103-11F2	ENSMUSG00000051378	ENSMUSG00000051378	Inversion	3000004C01Rik (RIKEN cDNA 3000004C01 gene)	102576655 - 102596217 bp (102.6 Mb)	-
43	WW103-11G3			Inversion			
44	WW103-11G8	ENSMUSG00000040430		Inversion	Pitpnc1 (phosphatidylinositol transfer protein, cytoplasmic 1)	106882559 - 107008762 bp (106.9 Mb)	-
45	WW103-11H1			Trapping?			
46	WW103-12A1	ENSMUSG00000018537		Inversion	Pcgf2 (polycomb group ring finger 2)	97360948 - 97371469 bp (97.4 Mb)	-
	WW103-12A2	ENSMUSG00000018537		Inversion	Pcgf2 (polycomb group ring finger 2)	97360948 - 97371469 bp (97.4 Mb)	-
	WW103-12A3	ENSMUSG00000018537	ENSMUSG00000018537	Inversion	Pcgf2 (polycomb group ring finger 2)	97360948 - 97371469 bp (97.4 Mb)	-
47	WW103-12A7	ENSMUSG00000034247		Inversion	Plekhm1 (pleckstrin homology domain containing, family M (with RUN domain) member 1)	103036219 - 103083789 bp (103.0 Mb)	-
	WW103-12D1	ENSMUSG00000034247	ENSMUSG00000034247	Inversion	Plekhm1 (pleckstrin homology domain containing, family M (with RUN domain) member 1)	103036219 - 103083789 bp (103.0 Mb)	-
	WW103-12D2	ENSMUSG00000034247	ENSMUSG00000034247	Inversion	Plekhm1 (pleckstrin homology domain containing, family M (with RUN domain) member 1)	103036219 - 103083789 bp (103.0 Mb)	-
48	WW103-12C8	Chr.3		Trapping			
	WW103-12D7	Chr.3		Trapping			
49	WW103-13A8		ENSMUSG00000017119	Inversion	Brca1/Nbr1, bidirectional promoter region	101227295 - 101253068 bp (101.2 Mb)	-
	WW103-13B1			Inversion			
	WW103-13B3			Inversion/Trapping			
	WW103-13B8		ENSMUSG00000017119	Inversion	Brca1/Nbr1, bidirectional promoter region	101227295 - 101253068 bp (101.2 Mb)	-
	WW103-13C4		ENSMUSG00000017119	Inversion	Brca1/Nbr1, bidirectional promoter region	101227295 - 101253068 bp (101.2 Mb)	-
	WW103-13C10		ENSMUSG00000017119	Inversion	Brca1/Nbr1, bidirectional promoter region	101227295 - 101253068 bp (101.2 Mb)	-
	WW103-13D3			Inversion			
50	WW103-13A11	ENSMUSG00000034247	ENSMUSG00000034247	Inversion/Trapping	Plekhm1 (pleckstrin homology domain containing, family M (with RUN domain) member 1)	103036219 - 103083789 bp (103.0 Mb)	-
51	WW103-13C12	Mus musculus hypothetical LOC217071		Trapping	Gm525 (gene model 525, NCBI)	88745074-88759792 bp (88.7 Mb)	+
	WW103-13D10	Mus musculus hypothetical LOC217071		Trapping	Gm525 (gene model 525, NCBI)	88745074-88759792 bp (88.7 Mb)	+
	WW103-13D12	Mus musculus hypothetical LOC217071		Trapping	Gm525 (gene model 525, NCBI)	88745074 - 88759792 bp (88.7 Mb)	+
52	WW103-14E7	ENSMUSG00000014195		Inversion	Dnajc7 (DnaJ (Hsp40) homolog, subfamily C, member 7)	100253974 - 100291291 bp (100.3 Mb)	-
	WW103-14E8	ENSMUSG00000014195	ENSMUSG00000014195	Inversion	Dnajc7 (DnaJ (Hsp40) homolog, subfamily C, member 7)	100253974 - 100291291 bp (100.3 Mb)	-
	WW103-14F1	ENSMUSG00000014195	ENSMUSG00000014195	Inversion/Trapping	Dnajc7 (DnaJ (Hsp40) homolog, subfamily C, member 7)	100253974 - 100291291 bp (100.3 Mb)	-
53	WW103-14F3	ENSMUSG00000001552		Inversion	Jup (junction plakoglobin)	100041744 - 100068862 bp (100.0 Mb)	-
54	WW103-14F11	ENSMUSG00000056649	ENSMUSG00000056649	Inversion	2810410L24Rik (RIKEN cDNA 2810410L24 gene)	119856472 - 119857805 bp (119.9 Mb)	-
55	WW103-16A3			Inversion			
	WW103-16B1	ENSMUSG00000018727		Inversion	D11Ertd636e (DNA segment, Chr 11, ERATO Doi 636, expressed)	113369301 - 113381146 bp (113.4 Mb)	-
	WW103-16B2	ENSMUSG00000018727		Inversion	D11Ertd636e (DNA segment, Chr 11, ERATO Doi 636, expressed)	113369301 - 113381146 bp (113.4 Mb)	-
	WW103-16D2	ENSMUSG00000018727		Inversion	D11Ertd636e (DNA segment, Chr 11, ERATO Doi 636, expressed)	113369301 - 113381146 bp (113.4 Mb)	-
	WW103-16D4	ENSMUSG00000018727		Inversion	D11Ertd636e (DNA segment, Chr 11, ERATO Doi 636, expressed)	113369301 - 113381146 bp (113.4 Mb)	-
56	WW103-16B3	ENSMUSG00000010342	ENSMUSG00000010342	Trapping	Tex14 (testis expressed gene 14)	87006730 - 87157487 bp (87.0 Mb)	+
57	WW103-18E2	ENSMUSG00000037992		Trapping	Rara (retinoic acid receptor, alpha)	98608821 - 98646063 bp (98.6 Mb)	+
	WW103-18G2	ENSMUSG00000037992		Trapping	Rara (retinoic acid receptor, alpha)	98608821 - 98646063 bp (98.6 Mb)	+
	WW103-18G10	ENSMUSG00000037992		Trapping	Rara (retinoic acid receptor, alpha)	98608821 - 98646063 bp (98.6 Mb)	+
	WW103-18G12	ENSMUSG00000037992		Trapping	Rara (retinoic acid receptor, alpha)	98608821 - 98646063 bp (98.6 Mb)	+
58	WW103-18E8		ENSMUSG00000018362	Inversion	Kpna2 (karyopherin (importin) alpha 2)	106659761 - 106670589 bp (106.7 Mb)	-
59	WW103-18F1			Inversion			
	WW103-18F9		ENSMUSG00000020917	Inversion/Trapping	Acly (ATP citrate lyase)	100147484 - 100199024 bp (100.1 Mb)	-
	WW103-18F11		ENSMUSG00000020917	Inversion	Acly (ATP citrate lyase)	100147484 - 100199024 bp (100.1 Mb)	-
60	WW103-18G5	ENSMUSG00000014195		Trapping	Dnajc7 (DnaJ (Hsp40) homolog, subfamily C, member 7)	100253974 - 100291291 bp (100.3 Mb)	-
	WW103-18G7	ENSMUSG00000014195		Inversion	Dnajc7 (DnaJ (Hsp40) homolog, subfamily C, member 7)	100253974 - 100291291 bp (100.3 Mb)	-
61	WW103-18G8	ENSMUSG00000041629		Inversion	D11Wsu99e (DNA segment, Chr 11, Wayne State University 99, expressed)	113332450 - 113355190 bp (113.3 Mb)	-
62	WW103-19A1	ENSMUSG00000006931		Inversion	1110036O03Rik (RIKEN cDNA 1110036O03 gene)	100079579 - 100085819 bp (100.1 Mb)	-
63	WW103-19A2			Inversion			
	WW103-19A9			Inversion			
64	WW103-19A6	ENSMUSG00000020917		Inversion/Trapping	D11Ertd636e (DNA segment, Chr 11, ERATO Doi 636, expressed)	113369301 - 113381146 bp (113.4 Mb)	-
65	WW103-19A8	ENSMUSG00000018160		Inversion	Pparbp (peroxisome proliferator activated receptor binding protein)	97823286 - 97864383 bp (97.8 Mb)	-
	WW103-19B1	ENSMUSG00000018160		Inversion	Pparbp (peroxisome proliferator activated receptor binding protein)	97823286 - 97864383 bp (97.8 Mb)	-
66	WW103-19D3	ENSMUSG00000040430		Inversion	Pitpnc1 (phosphatidylinositol transfer protein, cytoplasmic 1)	106882559 - 107008762 bp (106.9 Mb)	-

4.2.1 Gene trapping hot spots

The 44 unique events selected on chromosome 11 trapped 30 different loci. Several loci on chromosome 11 were trapped more than once. One locus was trapped 6 times (*Dnajc7*), a second locus was trapped 3 times (*Igf2bp1*), and 7 loci were trapped twice (*Pecam*, *2810410L24Rik*, *Tex14*, *Jup*, *Plekhm1*, *D11Erd636e* and *Pitpnc1*). These loci probably represent gene-trap or viral insertion hot spots. The Splinkerette PCR results showed that when different gene-trap insertions occurred at the same locus, they either occurred at different positions in the same intron or in different introns. This result shows that the bias is locus-specific, instead of sequence-specific (Hansen, Floss et al. 2003). One example is given below.

Igf2bp1 (insulin-like growth factor 2, binding protein 1) is also known as *CRD-BP* (c-myc mRNA coding region instability determinant binding protein) (Tessier, Doyle et al. 2004). This protein is a multifunctional RNA-binding protein, which can bind to *c-myc*, insulin-like growth factor II, β*-actin* and H19 mRNAs. By binding to different RNA substrates, this protein can affect their localization, translation, or stability. The protein level of *Igf2bp1* is high during foetal development and almost undetectable in normal adult tissues. But the expression of *Igf2bp1* is reactivated in some adult human tumours including breast, colon, and lung tumours (Tessier, Doyle et al. 2004), though the significance of this is not clear.

Two independent gene-trap events were found in this gene locus. A *Spe*I/*Xba*I/*Nhe*I splinkerette PCR product was obtained from one subclone of the group, WW103-6G1. Sequence of the PCR product matched the second intron of the gene (Fig. 4-4a). *Spe*I/*Xba*I/*Nhe*I and *Sau*3AI splinkerette PCR products were obtained from one subclone of the group, WW103-5F4. Sequences of both products matched the seventh intron of the gene (Fig. 4-4b).

a


IgDp1 Exon 3
EcoRI
PgII
EcoRI
PgII
EcoI
SacI
XbaI
Spliceosetta Fragment
SphI
Spliceosetta Fragment
Spliceosetta Fragment
PgII
Spliceosetta Fragment
Spliceosetta Fragment

b

Fig. 4-4 Two independent gene traps at a same locus, *Igf2bp1*. a. *Spe*I/*Xba*I/*Nhe*I splinkerette PCR product was obtained from one subclone of the first group, WW103-6G1. Sequence of the PCR product mapped to the second intron of *Igf2bp1*. **b.** *Spe*I/*Xba*I/*Nhe*I and *Sau*3A1 splinkerette PCR products were obtained from one subclone of the second group, WW103-5F4. Sequence of both products mapped to the seventh intron of the gene. Splinkerette results were underlined in green, the green arrow represents the direction from the 5' LTR of the virus to the endogenous restriction site. The exons of *Igf2bp1* are underlined in blue, the blue arrow represents the transcription direction of *Igf2bp1*.

4.2.2 Distribution of trapped genes on chromosome 11

Using the Ensembl and NCBI database, the sequences generated from the splinkerette PCR and 5' RACE products were mapped exclusively in a 45.7 Mb distal region of the mouse chromosome 11, surrounding the *E_2DH* locus. About two thirds of the trapped genes (23/33) were clustered within 5 Mb proximal and 5 Mb distal to the *E_2DH* locus (Fig. 4-5). This distribution shows that the efficiency of the Cre inside this 11.3 Mb region is much higher than outside. The biggest inversion distal to the *E_2DH* locus was 19.2 Mb in size. The trapped locus, *2810410L24Rik* (119.9 Mb), is very close to the telomere of the chromosome (121.5 Mb). The biggest inversion proximal to the *E_2DH* locus was 26.5 Mb. So it is reasonable to expect that if a locus in the middle of the chromosome 11 was chosen, the whole region from which homozygous gene-trap clones could be recovered will be even bigger.

4.2.3 Orientation of the transcription of the trapped genes

The orientations of the transcription of the trapped genes were determined according to the Ensembl database (Fig. 4-6). An inversion can only be generated if the trapped genes are transcribed from the antisense strand (from telomere to centromere) of the chromosome 11. In this orientation, the *lox*P site introduced by the retrovirus will be in the opposite orientation to the anchor *lox*P site. Of all the 30 loci mapped to chromosome 11, 27 of them are transcribed from the antisense strand as expected. However, 3 gene traps are transcribed from the sense strand (from centromere to telomere) (Fig. 4-6a).

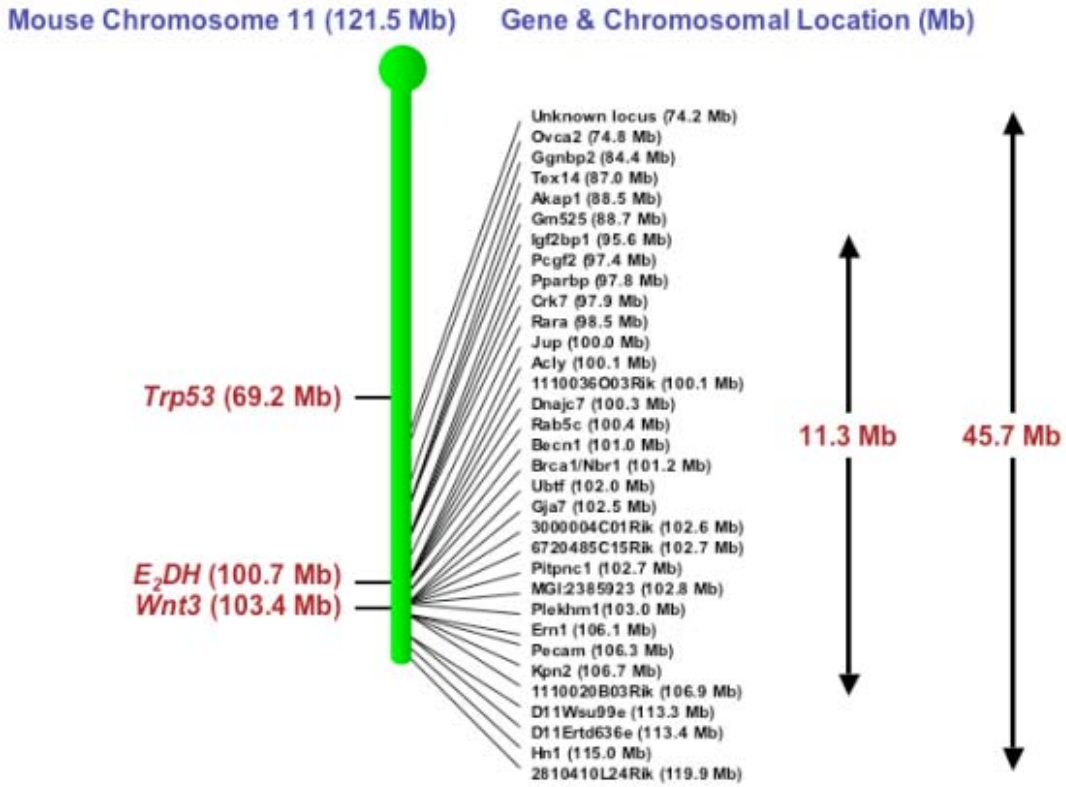


Fig. 4-5 Distribution f gene traps on chromosome 11. Splinkerette PCR and 5' RACE results were mapped on chromosome 11 according to the Ensembl database (http://www.ensembl.org/Mus_musculus/). Gene traps were isolated exclusively in a 45.7 Mb distal region of the mouse chromosome 11, surrounding the *E_2DH* locus. About two third of the trapped genes (23/33) were clustered within 5 Mb proximal and 5 Mb distal to the *E_2DH* locus.

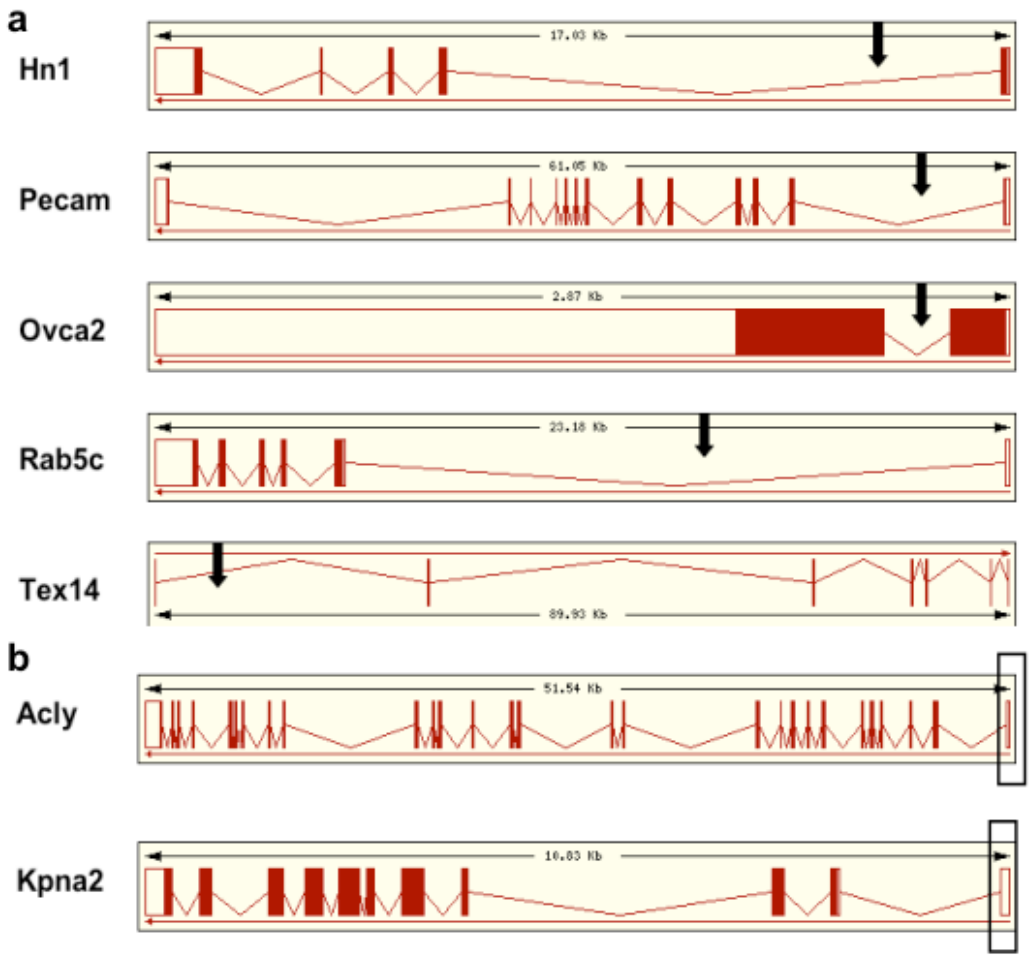


Fig. 4-6 Transcription orientation and the proviral insertion sites of the trapped trapped loci. a. The vector insertion sites of the gene traps at 5 different loci, *Hn1*, *Pecam*, *Ovca2, Rab5c* and *Tex14*, determined by the Splinkerette PCR results. Note that the Tex14is transcribed from the sense strand (from centromere to telomere), and the other four genes are transcribed from the antisense strand (from telomere to centromere). The black arrows point to the proviral insertion site. **b.** The trapped exons of gene trap at 2 different loci, *Acly* and *Kpna2*, determined by the 5' RACE results. The black box highlight the trapped exons. The chromosomal structures of the trapped loci are downloaded from Ensembl (http://www.ensembl.org/Mus_musculus/). The red arrows underneath the exon structures represent the transcription orientation of the genes.

One possibility is a *trans* recombination event between *lox*P sites in direct orientation on the two homologs of chromosome 11 in G1. This will result in a pair of balanced deletion/duplication chromosomes. If the trapped locus is distal to *E_2DH*, the chromosome with the anchor *lox*P site will become a deletion chromosome (Fig. 4-7a). And it is unlikely that such a chromosome can become homozygous after induced mitotic recombination because of the loss of genetic material. If the trapped locus is proximal to *E_2DH*, the chromosome with the anchor *lox*P site will become a duplication chromosome (Fig. 4-7b). This should be able to become homozygous after the induced mitotic recombination. But the double duplication ES cells will lose the *β-geo* cassette after the induced mitotic recombination event and no proviral/host junction fragment should be detected using *lacZ* as a probe. So the balanced deletion/duplication can not be the cause for these clones that are transcribed from the sense strand (from centromere to telomere).

Another possibility is a *trans* recombination event between *lox*P sites in direct orientation on the sister-chromatids of chromosome 11 in G2, which will result into a duplication chromosome and a wild-type chromosome. If the trapped locus is distal to *E_2DH*, the duplication chromosome will carry two *β-geo* cassettes (Fig. 4-8a), one associated with a complete proviral insertion, while the other is a half proviral insertion split by Cre-mediated recombination. But this recombinant can not survive the puromycin selection because the duplication chromosome does not have a functional *Puro* cassette on it. If the trapped locus is proximal to *E_2DH*, the duplication chromosome will carry only one *β-geo* cassette (Fig. 4-8b), which belongs to a complete proviral insertion. The recombinant should be able to survive the puromycin selection because the duplication chromosome has a functional *Puro* cassette on it. The duplication chromosome can also become homozygous after induced mitotic recombination because it does not lose genetic material.

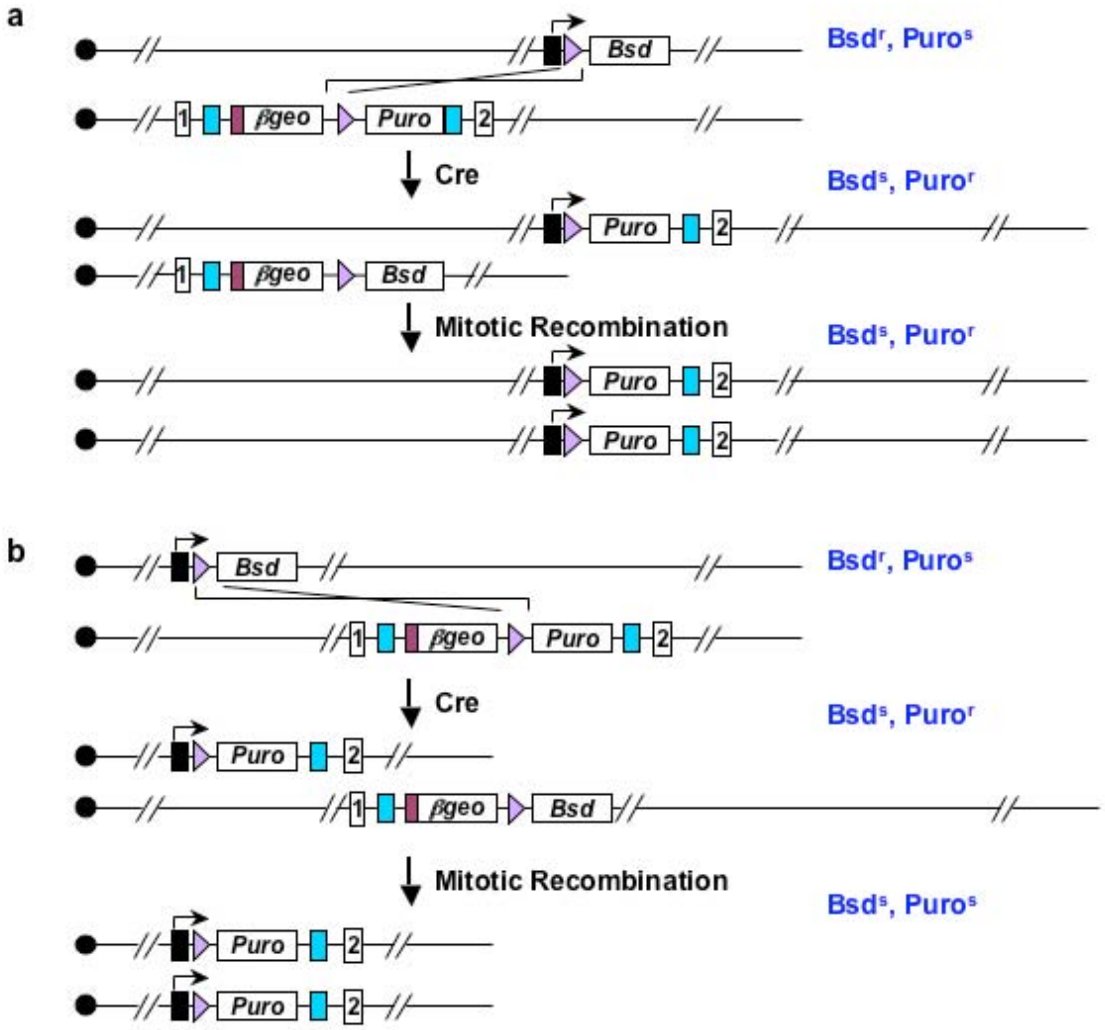

a
Bsd
Bsd^r, Puro^s
1
βgeo
Puro
2
Cre
Bsd^s, Puro^r
Puro
2
1
βgeo
Bsd
Mitotic Recombination
Bsd^s, Puro^r
Puro
2
Puro
2
b
Bsd
Bsd^r, Puro^s
1
βgeo
Puro
2
Cre
Bsd^s, Puro^r
Puro
2
1
βgeo
Bsd
Mitotic Recombination
Bsd^s, Puro^s
Puro
2
Puro
2

Fig. 4-7 G1 *trans* recombination event. a. *lox*P sites in direct orientation on different homologs of chromosome 11 (*trans*) and the trapped locus is proximal to *Hsd17b1*. A pair of balanced deletion/duplication chromosomes will be generated by Cre-mediated recombination (genetic material is not lost at this stage). The resulting cells can survive subsequent puromycin selection. After mitotic recombination, the duplication chromosome can become homozygous. However, the duplication chromosomes do not have the *βgeo* cassette, so these clones can not be picked up by Southern using a *lacZ* probe. **b.** *lox*P sites in direct orientation on different homologs of chromosome 11 (*trans*) and the trapped locus is distal to *Hsd17b1*. A pair of balanced deletion/duplication chromosomes will be generated by Cre-mediated recombination (genetic material is not lost at this stage). The resulting cells can survive subsequent puromycin selection. After mitotic recombination, the deletion chromosome will become homozygous. However, this recombination event was not observed, presumably because the homozygous deletions will be selected against because of homozygous lethality in ES cells. Black box, *PGK* promoter; lavender triangle, wild-type *lox*P site; blue box, virus long terminal repeat; plum box, splice acceptor.

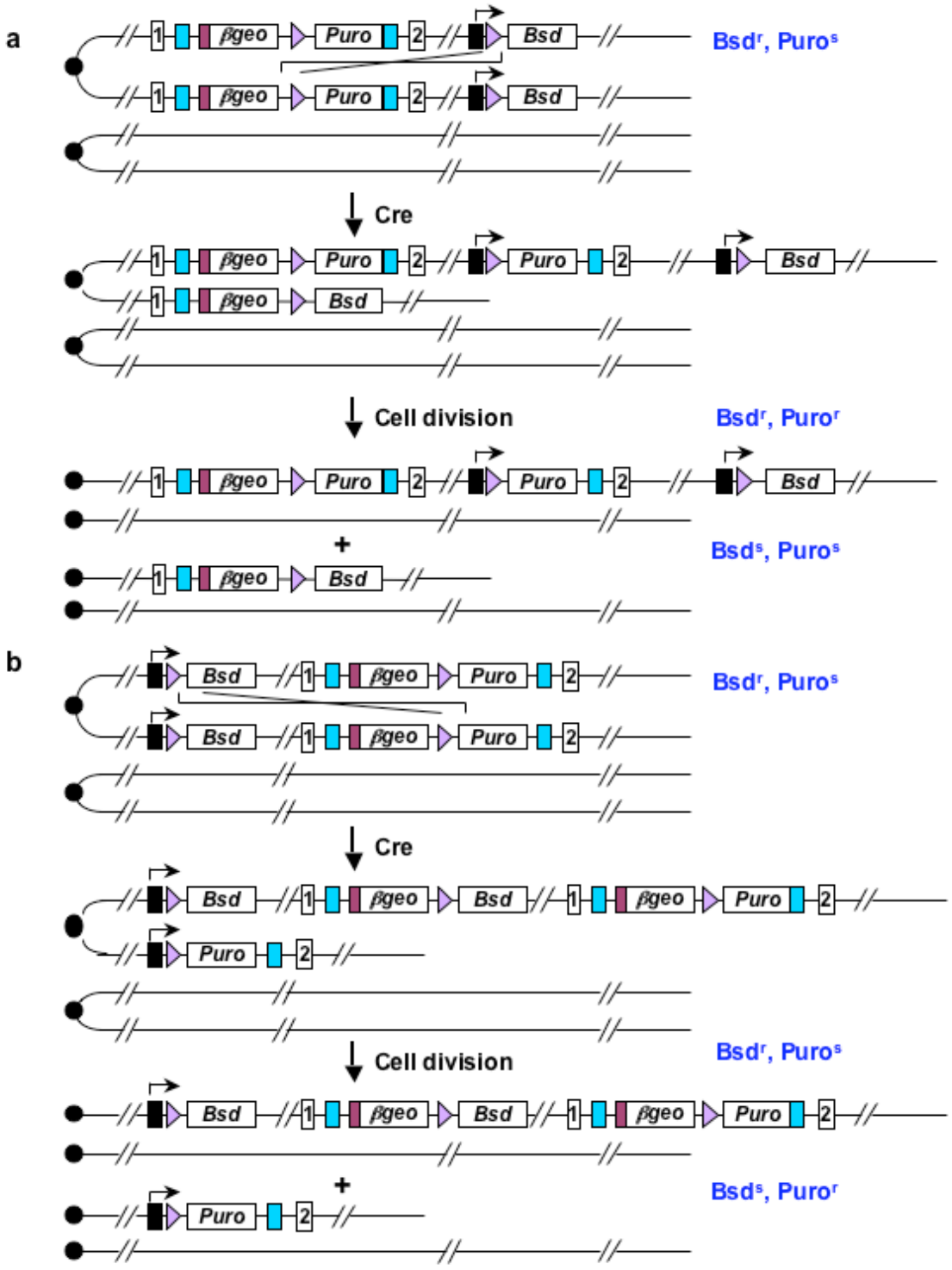

a
βgeo
Puro
Bsd
1
2
Bsd^r, Puro^s
Cre
Cell division
Bsd^r, Puro^r
Bsd^s, Puro^s
b
Bsd^r, Puro^s
Cre
Cell division
Bsd^r, Puro^s
Bsd^s, Puro^r

Fig. 4-8 G2 *cis* recombination event. a. *lox*P sites are in direct orientation on the same homolog of chromosome 11 (*cis*) and the trapped locus is proximal to *Hsd17b1*. A pair of balanced deletion/duplication sister chromatids will be generated by Cre-mediated recombination (G2-*trans*). After cell division, the cells with the duplication chromosome can survive puromycin selection. This duplication chromosome can become homozygous because it does not lose any genetic material. **b.** *lox*P sites in direct orientation on the same homolog of chromosome 11 (*cis*) and the trapped locus is distal to *Hsd17b1*. A pair of balanced deletion/duplication sister chromatids will be generated by Cre-mediated recombination (G2-*trans*). After cell division, the cells with the deletion chromosome can survive the puromycin selection. However, this recombination event was not observed. The heterozygous (before mitotic recombination) and homozygous (after mitotic recombination) deletions will be selected against because of haplosufficiency and/or homozygous lethality in ES cells. Black box, *PGK* promoter; lavender triangle, wild-type *lox*P site; blue box, virus long terminal repeat; plum box, splice acceptor.

These homozygous duplication clones can be distinguished from other homozygous inversion clones by Southern as well as the drug selection. A *Puro* specific probe can detect the 6.0/6.9 kb *Kpn*I fragment representing proviral insertion and 3.4 kb *Kpn*I fragment representing the reconstitute *PGK-Puro-bpA* cassette. A simpler way to confirm the identity of the clones is a sib-selection using M15+puromycin and M15+blasticidine. The clones with two duplication chromosomes should be resistant both to puromycin and blasticidin, while the clones with two inversion chromosomes should be resistant to puromycin but sensitive to blasticidin.

Three trapped loci (*Tex14*, *LOC217071* and *Rara*) transcribed from the sense strand (from centromere to telomere) are all located proximal to the E_2DH locus, and *Tex14* was trapped twice. All the HAT resistant subclones from the three independent events are homozygous for the modified E_2DH locus (*Nde*I digestion, E_2DH 3' probe), and all the clones only carry the 6.9 kb proviral insertion fragment (*Kpn*I digestion, *lacZ* probe). And all these clones are resistant to both puromycin and blasticidin. These results are consistent with these clones carrying two duplication chromosomes.

4.2.4 Proviral insertion sites in trapped loci

For a large portion of the trapped loci, I have cloned the proviral/host junction fragments by Splinkerette PCR. This is informative on the chromosomal structure of the recombinants after regional trapping and induced mitotic recombination. In most of the trapped loci, the proviral insertions occur in the first or second intron, the inversion thus generates a breakpoint between the first one or two exons and the rest of the gene (Fig. 4-6a). That means, in most cases, the transcriptional regulation elements will be several megabases away from the coding region, and in many cases, the open reading frames themselves is also disrupted.

For some clones, the proviral insertion sites were not determined by Splinkerette PCR, for example when the sequence was repetitive or the Splinkerette PCR failed. In these cases, the RACE results mapped the insertion to an exon. In one case, the 5' RACE products mapped to an

unknown locus close to the *Gja7* gene (gap junction membrane channel protein alpha 7 or connexin 45). By searching the NCBI database, the 5' RACE product matched the 5' untranslated region of an alternatively spliced form of connexin 45 (AY390396). So it is likely that I have trapped an alternative spliced form that expresses in the undifferentiated ES cells. In another case, the 5' RACE product mapped to an unknown locus close to the *Brca1* gene (breast cancer 1, early onset). By searching the NCBI database, the 5' RACE product matched *Brca1*/*Nbr1* bidirectional promoter region (AF080589) and some *Brca1* EST sequence. So it is likely that I have trapped an ES cell specific alternative spliced form of *Brca1*.

In at least one case, it appears that more than one gene was disrupted by the retrovirus integration and the subsequent inversion. In this case, the trapped locus, *2810410L24Rik* (119.9 Mb), is very close to the telomere of the chromosome 11. Ensembl predicts this is a single exon gene that is transcribed from the sense strand, but the 5' RACE sequence matches two separate regions from the opposite direction, which suggests that there is another transcript from the opposite strand. By searching the NCBI database, I have identified another cDNA, *D030042H08Rik*, transcribed in a similar way as the 5' RACE product which overlaps with *2810410L24Rik*. The splinkerette results map the retroviral insertion site between the second and third exons of *D030042H08Rik* (Fig. 4-9). Another independent gene-trap was also mapped to a locus very close to *2810410L24Rik*. A hypothetical gene *LOC432619* is also transcribed from the antisense strand. *D030042H08Rik* and *LOC432619* both belong the same mouse UniGene, Mm.269766. Interestingly, the UniGene is named as "RIKEN cDNA 2810410L24 gene (2810410L24Rik)", though the transcription direction of *2810410L24Rik* is opposite to the other transcripts, *D030042H08Rik* and *LOC432619*. In fact, there is another UniGene, Mm.125044 with the same name "2810410L24Rik RIKEN cDNA 2810410L24 gene (2810410L24Rik)". And this one is composed of all the cDNAs and ESTs transcribed from the sense strand. Since *D030042H08Rik* and *LOC432619* have not been mapped in Ensembl, I still name the trapped locus as *2810410L24Rik*. But the proviral integration and the inversion will disrupt transcripts from both directions.


Exon 3
RIKEN 2610410L24 cDNA
RIKEN 2610410L24 cDNA
RIKEN 2610410L24 cDNA
RIKEN 2610410L24 cDNA
Exon 2
RIKEN 2610410L24 cDNA
RIKEN 2610410L24 cDNA
5F'R/3' RACE Product
Exon 2
RIKEN 2610410L24 cDNA
RIKEN 2610410L24 cDNA
5F'R/3' RACE Product
RIKEN 2610410L24 cDNA
RIKEN 2610410L24 cDNA
5F'R/3' RACE Product
RIKEN 2610410L24 cDNA
RIKEN 2610410L24 cDNA
5F'R/3' RACE Product
Exon 1
RIKEN 2610410L24 cDNA
RIKEN 2610410L24 cDNA
Exon 1
RIKEN 2610410L24 cDNA
RIKEN 2610410L24 cDNA

Fig. 4-9 The gene trap insertion at *2810410L24Rik* has disrupted two genes with opposite directions of transcription. *2810410L24Rik* is a single exon gene that is transcribed from the sense strand. The 5' RACE matches two separate regions from the opposite direction, which suggests another transcript from the anti-sense strand. Another gene, *D030042H08Rik*, is transcribed in a similar way as the 5' RACE product and its transcript is overlapping with *2810410L24Rik*. Splinkerette PCR results have mapped the retroviral insertion site between the second and third exons of *D030042H08Rik*. Note that the transcription of *D030042H08Rik* is not the same as the 5' RCAE results, so they might represent different splice forms of the same gene. Another independent gene-trap was also mapped to a locus very close to *2810410L24Rik*, the predicted gene *LOC432619*. This predicted gene is also transcribed from the antisense strand.

4.3 Discussion

In this study, I have mapped 49 of the 66 independent gene traps by sequence analysis of 5' RACE and/or Splinkerette PCR products. Most of the gene traps identified are located on chromosome 11 and are transcribed from the antisense strand (from telomere to centromere). But some possible translocations and other possible chromosomal rearrangements were observed. These events constituted background in the context of the goal of this study, namely to generate homozygously mutated ES cell clones for recessive genetic screens *in vitro*. No matter how these clones survived the stringent selection procedures, through a series of rare recombination events, they have lost or gained a large part of the chromosome 11. The phenotype of these clones can not be attributed to a single gene, and thus they are not suitable for the genetic screens. Importantly, they can be easily identified by the sequence of Splinekerette PCR and 5' RACE products because of the location and transcriptional orientation of the trapped genes.

Some gene-trap hot spots were found in all of the trapped loci. 9 of the 30 mapped gene-trap loci on chromosome 11 were hit more than once. One of them, *Dnajc7*, was hit 6 times. Interestingly, this locus (100.3 Mb) is very close to the anchor locus, *E_2DH* (100.7 Mb). The recombination efficiency over such small distance will be extremely high (Zheng, Sage et al. 2000). One would predict that if a gene-trap hot spot with the correct transcriptional orientation happens to be close to the anchor point, the small inversions will dominate the pool of inversion clones, and consequently the induced mitotic recombination clones. This will significantly decrease the possibility of identifying other recombination events, especially those relatively rare events (large inversions) from the same pool. It is almost impossible to screen against these clones by Southern analysis even if the hot spots are already known, because the proviral insertion bias is locus specific, not sequence specific, which means that even if the same locus is hit multiple times, the proviral insertion will occur in different introns or different positions in the same intron.

A similar situation was also noticed in other genome-wide gene trapping programs (Zambrowicz, Friedrich et al. 1998; Hansen, Floss et al. 2003; Skarnes, von Melchner et al. 2004). The efficiency of trapping new genes is not linear, and it will drop with the increase of gene-trap tags. Within the first 100,000 tags, the rate of capturing new genes declines to about one new gene every 35 tags (Skarnes, von Melchner et al. 2004). If our strategy is applied to generate genome-wide homozygous gene-trap clones, there will also be a balance point beyond which new genes can not be mutated economically. However, if various plasmid and retroviral vectors are used, it will help to overcome the bias of gene trap insertions of a single vector and increase the efficiency of gene trapping (Hansen, Floss et al. 2003).

One way to control this bias is to limit the size of the pool of original gene traps. As discussed in the previous chapter, between zero and one inversion events are expected after Cre-mediated recombination in a pool of 100 trapping events. If the starting cell number and the electroporation conditions are the same, it is possible to predict the type of recombination (inversion or translocation) and the size of the inversion (small or large) simply by counting the number of puromycin resistant colonies on each plate.

On the other hand, 21 of the 30 mapped gene-trap loci on chromosome 11 were hit only once. So the regional trapping experiment is far from reaching saturation. If the same experiment is repeated, many new homozygous gene traps on chromosome 11 will be recovered. Also, 27 of the 30 mapped gene-trap loci are transcribed from the antisense strand (from telomere to centromere), which proves that our strategy is highly efficient for generating homozygous gene-trap mutations transcribing from one strand. Simply by changing the orientation of the *lox*P site of the anchor point targeting vector, the genes in the same region but are transcribed from the other strand will be trapped.

I have compared the gene traps identified in my experiment with the ones identified in the regional trapping experiment by Meredith Wentland (Wentland et al. unpublished data). It is interesting to notice that the distribution pattern

of the gene traps is extremely similar between the two experiments. In both experiments, most gene traps occurred within a region of 40 Mb in distal region of the mouse chromosome 11. A large portion of the gene traps clustered within 5 Mb proximal and 5 Mb distal of the *E_2DH* locus (18/21 for Wentland's experiment and 20/30 for our experiment). This distribution pattern of the gene traps is expected based on the relationship between the efficiency of Cre in recombining *lox*P sites over different distances (Liu, Zhang et al. 1998; Zheng, Sage et al. 2000).

Though the distribution pattern is very similar between the two different experiments, the gene traps isolated are totally different. None of the trapped loci identified in Wentland's experiment were hit in my experiment or vice versa. Of the 21 gene traps isolated in Wentland's experiment, only 2 matched known mouse genes, 4 matched predicted transcripts, 4 matched ESTs and 11 matched unknown loci. In my experiment, most of the gene traps matched known mouse genes or transcripts, only one of them appeared to be an unknown locus on chromosome 11. This difference probably reflects the different "trappable" sets of genes for 5' and 3' trapping. 5' trapping is dependent on the expression of the trapped locus in undifferentiated ES cells, whilst 3' trapping is not. Considering that the purpose of my experiment is to mutate genes for *in vitro* screen, 5' trapping is more likely to disrupt a functional gene in ES cells and thus more likely to result in a phenotype *in vitro.*

In both experiments, only genes transcribed from one strand can be selected. However, some gene traps transcribed from the opposite strand have also survived the selection (4/21 for Wentland's experiment and 3/30 for my experiment). Wentland et al. (unpublished data) have carried out Fluorescence In situ Hybridization (FISH) to identify the alternative recombination events. She found that these clones were either cells with balanced deletion/duplication chromosomes derived from a G1 *trans* recombination event, or cells with one wild-type chromosome and one duplication chromosome derived from a G2 *trans* recombination event. By Southern analysis and drug sib-selection, the three ES cell clones in which

the trapped loci are transcribed on the sense strand are all clones that carry two duplication chromosomes. These alternative recombination events caused some background, but their efficiency is relatively low and thus did not present a serious problem for the genetic screen. Clones with these events can be easily identified from the sequence of their Splinkerette PCR and 5' RACE products or drug selection.

Cells with translocation chromosomes and duplication chromosomes are by-products of my products. They are not useful for the *in vitro* genetic screen because in these clones, a large genomic region is either deleted or duplicated, and the phenotype is very hard to be associated with a certain gene. But on the other hand, these rearranged chromosomes might be a useful resource for other experiments. ES cells with large deletions may be selected against if the deletion affects cell viability or growth (Zheng, Sage et al. 2000). For example, a 22 cM deletion distal to the *E_2DH* locus (*E_2DH* - *D11Mit69*) was found to be haploinsufficient in ES cells. In the rare cells that survived selection, the remaining wild-type chromosome was duplicated. In some of the translocation events identified in my experiment, the modified *E_2DH* locus became homozygous. That means that the resulting ES cell clones are partially trisomic for the genomic region translocated from other chromosomes, but they have lost one copy of the genomic region on chromosome 11 distal to the *E_2DH* locus, which is approximately 20 Mb in size. The Southern analysis using the *E_2DH* 3' probe has shown that these clones do not have wild-type chromosome 11. It is possible that the gain of a genomic region from another chromosome somehow can compensate the haploinsufficiency caused by the loss of the distal region of chromosome 11. Another possibility is that the chromosome which the distal region on chromosome 11 was translocated to was duplicated. I have not carried out FISH experiment to determine the exact genomic structure of the clones. But these clones can be useful to study the functional relationships between different genomic regions.

There are several reports that gene-trap insertions do not completely disrupt the normal transcription of the endogenous gene and the mutagenicity of

gene-trapping is still a controversial issue. Mitchell et al. (2001) have generated sixty mouse lines with secretory gene-trap vectors. Twenty-five of them showed visible embryonic or adult phenotypes. For 11 of the 25 gene traps that showed observable phenotypes, alleles generated by gene-targeting have also been reported. Ten of these strains showed exactly the same phenotype as the gene-targeted mutations. The remaining strain had a less severe phenotype than the gene-targeted allele but still caused embryonic lethality (Mitchell, Pinson et al. 2001). Stanford et al. (2001) has reviewed one hundred additional gene-trap insertions that have been described in the literature. Sixty percent of these insertions show “obvious” phenotypes, and 40% are recessive lethal mutations. The frequency of recessive lethal mutations and “obvious” phenotypes generated by gene-trapping is comparable to the results obtained from gene-targeting (Stanford, Cohn et al. 2001). Nevertheless, trapping alone is not always sufficient to completely block transcription. Leaky expression of wild-type transcripts can partially rescue some phenotypes and thus complicate analysis.

Our strategy not only inserts a trapping cassette in the gene, but also introduces a breakpoint at the proviral insertion site. So this technique should be more mutagenic than the other mutagenesis methods. At the same time, long-range chromosomal rearrangements can disrupt more than the trapped locus. In my experiment, in one gene-trap clone, the inversion disrupted two partially overlapped genes that are transcribed in opposite orientations. In another case, the inversion created a breakpoint at a bidirectional promoter. So the expression of genes around the breakpoint need to be checked to avoid misinterpretation of any observed phenotypes.

The identification of the gene-trap loci has proved that this strategy is useful to generate homozygous mutations in a genomic region of interest. This strategy can easily be applied to other mouse chromosomes to generate homozygous mutant ES cells in other regions of the genome. This resource will facilitate large-scale *in vitro* genetic screens.

5 Genetic screen on homozygous gene traps

5.1 Introduction

During *in vitro* differentiation, ES cells can form cystic embryo-like aggregates, embryoid bodies (EB), that contain cells of endodermal, ectodermal and mesodermal lineages, which can further differentiate into more specialized cell types. The morphological changes of embryoid bodies are accompanied, at the molecular level, by the changes in the expression of a set of lineage-specific and tissue-specific markers. By comparing the dynamic changes in the expression of these markers *in vivo* and *in vitro*, different stages of EB differentiation *in vitro* can be linked to different stages of embryogenesis *in vivo* (Leahy, Xiong et al. 1999). These properties allow us to use ES cell *in vitro* differentiation as an *in vitro* model to study early embryogenesis and this facilitates genetic approaches.

5.1.1 *In vitro* differentiation protocols

There are three main protocols for ES cell *in vitro* differentiation: the hanging drop method (Wobus, Wallukat et al. 1991); the mass culture method (Doetschman, Eistetter et al. 1985); and the methylcellulose method (Wiles and Keller 1991). All three of these have been widely used for making embryoid bodies (EB) for different purposes.

The advantage of the hanging drop method is that the starting number of ES cells in an embryoid body is defined, so the size and the differentiation pattern of the EBs generated by this method is more consistent than with the other two methods. This characteristic is particularly important for developmental studies, which require the comparisons between EBs under different culture conditions and/or with different mutations (Wobus, Guan et al. 2002). However, this method is also more complicated than the other two methods.

On the other hand, the mass culture method is useful for differentiating a large number of ES cells. By plating undifferentiated ES cells onto bacteriological Petri dishes, the cells automatically form cell aggregates, and the aggregates can differentiate into a variety of different cell types. However, the size and the differentiation pattern can vary significantly between plates or between

experiments, even when the same ES cell line is used. The methylcellulose method is used specifically for the differentiation of haematopoietic lineages, and is not suitable for other purposes.

In this project, it was necessary to compare the *in vitro* differentiation potential of a number of homozygous mutant ES cell clones. Therefore the hanging drop method was the most appropriate *in vitro* differentiation protocol to use.

5.1.2 Parameters influencing *in vitro* differentiation of ES cells

The developmental potency of ES cells in culture is dependent on a number of intrinsic and extrinsic parameters. These include the number of ES cells used to make the EBs; the composition of the differentiation medium; cellular growth factors and differentiation inducers added to the culture medium; the ES cell lines; as well as the genetic changes in the ES cell genome.

Compared to *in vivo* differentiation in the mouse, the parameters for *in vitro* differentiation are more controllable. Whichever differentiation protocol is chosen, extrinsic parameters can be effectively controlled by using defined medium and culture conditions. Variations caused by intrinsic parameters can be eliminated by choosing an appropriate control ES cell line. Thus loss-of-function or gain-of-function studies using *in vitro* differentiation can be an ideal alternative to study the phenotypes of mutations on embryogenesis and early development (Wobus, Guan et al. 2002).

5.1.3 Recessive genetic screens using ES cell *in vitro* differentiation

Genetic analysis of recessive mutations in ES cells is informative on possible functions *in vivo*, especially for mutations that result in embryonic lethality. A recessive genetic screen using ES cell *in vitro* differentiation can be used to identify important genes in the differentiation process.

The bottleneck of recessive genetic screens in ES cells is the difficulty of obtaining enough homozygous mutant ES cells. If a genetic screen is performed to identify genes involved in ES cell *in vitro* differentiation, pure homozygous mutant ES cell clones need to be differentiated individually to

check for their differentiation potential. Existing methods to generate homozygous mutations in ES cells are not ideal for this purpose. In the previous chapters, I have demonstrated that a strategy combining regional trapping and inducible mitotic recombination can be used to generate homozygous mutations in a genomic region of interest. By Splinkerette PCR and 5' RACE, proviral/host flanking genomic sequences and/or cDNA sequence were isolated to identify the proviral insertion sites and inversion breakpoints of these mutant clones.

A total of 30 different gene-trap loci on chromosome 11 that are homozygously mutated were isolated. These homozygous gene-trap clones can be used to perform a small-scale genetic screen to identify the mutations that will disrupt the normal ES *in vitro* differentiation process. Each gene-trap clone has been differentiated individually, and a set of important lineage-specific and tissue-specific markers have been checked to determine the differentiation potential of each of the homozygous mutant ES cell lines. Mutant cell lines that show an abnormal differentiation pattern have been confirmed using independent methods.

5.2 Results

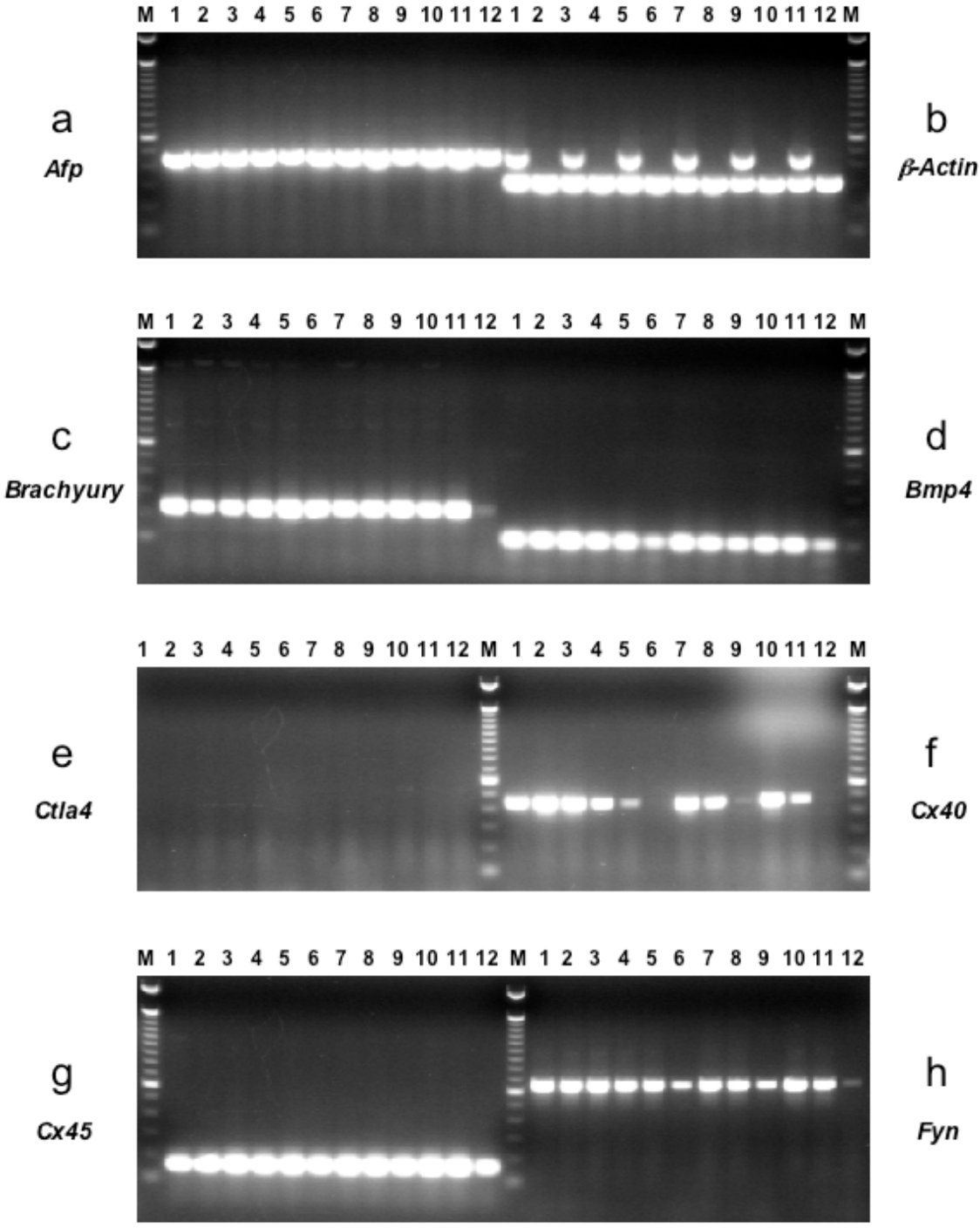
5.2.1 Primary screen

For each of the 33 mapped gene-trap loci, at least one subclone was chosen for the primary *in vitro* differentiation screen. Embryoid bodies were made and cultured as described before (Wobus, Guan et al. 2002). In brief, undifferentiated ES cells were maintained on feeder layers until they were used for *in vitro* differentiation. To setup the assay, ES cells were trypsinized and diluted to a final concentration of approximately 600 cells in 20 μl Differentiation Medium (see material and methods). 20 μl drops of the ES cell suspension were laid onto the bottom of 100-mm bacteriological Petri dishes. The Petri dishes were inverted and the ES cell aggregates were cultured in the resulting hanging drops for two days. After this, the Petri dishes were turned the right way up and Differentiation Medium was added into each dish to rinse the aggregates. The aggregates were cultured in suspension for

another three days. The sample was harvested at Day 5. At the same time, the EBs on the remaining dishes were plated out onto gelatinized 90-mm tissue culture plates. The plated EBs were subsequently cultured in Differentiation Medium supplemented with 10^{-8} M retinoic acid (RA) and the medium was changed every two days. Subsequent samples were taken at Day 8 and Day 11.

When all the samples were taken, RNA was extracted from each sample and quantified. 5 μg total RNA was used for first strand cDNA synthesis. The resulting cDNA was used as a template for RT-PCR. In the primary of screen, 16 pairs of primers were used (*Afp*, *β-Actin*, *Brachyury*, *Bmp4*, *Ctla4*, *Cx40*, *Cx45*, *Fyn*, *Gata4*, *Goosecoid*, *Hnf4*, *Nodal*, *Oct3/4*, *Pecam*, *Tie2* and *vHNF1*). All the homozygous mutant cell lines that showed abnormal expression (significant up-regulation or down-regulation compared to the WW93A12 control line) for one or more markers in the primary screen were selected for the second round screen (Fig. 5-1 and Fig. 5-2).

All the mitotic recombination clones (WW103) used in the screen carry two homologs of chromosome 11 from the same parent, either bi-paternal or bi-maternal. It is possible that because of the imprinting, the *in vitro* differentiation pattern of ES cells carrying bi-paternal or bi-maternal homologs of chromosome 11 will be different from that of the wild-type ES cells that have one paternal and one maternal homologs of chromosome 11. Also, the *in vitro* differentiation potential of ES cells homozygous for the targeted E_2DH allele has not been assessed. So an ideal control cell line for this experiment will carry two homologs of chromosome 11 from the same parent as the WW103 clones, and this control cell line is also homozygous for the targeted E_2DH allele.


M 1 2 3 4 5 6 7 8 9 10 11 12 1 2 3 4 5 6 7 8 9 10 11 12 M
a
Afp
b
β-Actin
M 1 2 3 4 5 6 7 8 9 10 11 12 1 2 3 4 5 6 7 8 9 10 11 12 M
c
Brachyury
d
Bmp4
1 2 3 4 5 6 7 8 9 10 11 12 M 1 2 3 4 5 6 7 8 9 10 11 12 M
e
Ctla4
f
Cx40
M 1 2 3 4 5 6 7 8 9 10 11 12 M 1 2 3 4 5 6 7 8 9 10 11 12
g
Cx45
h
Fyn

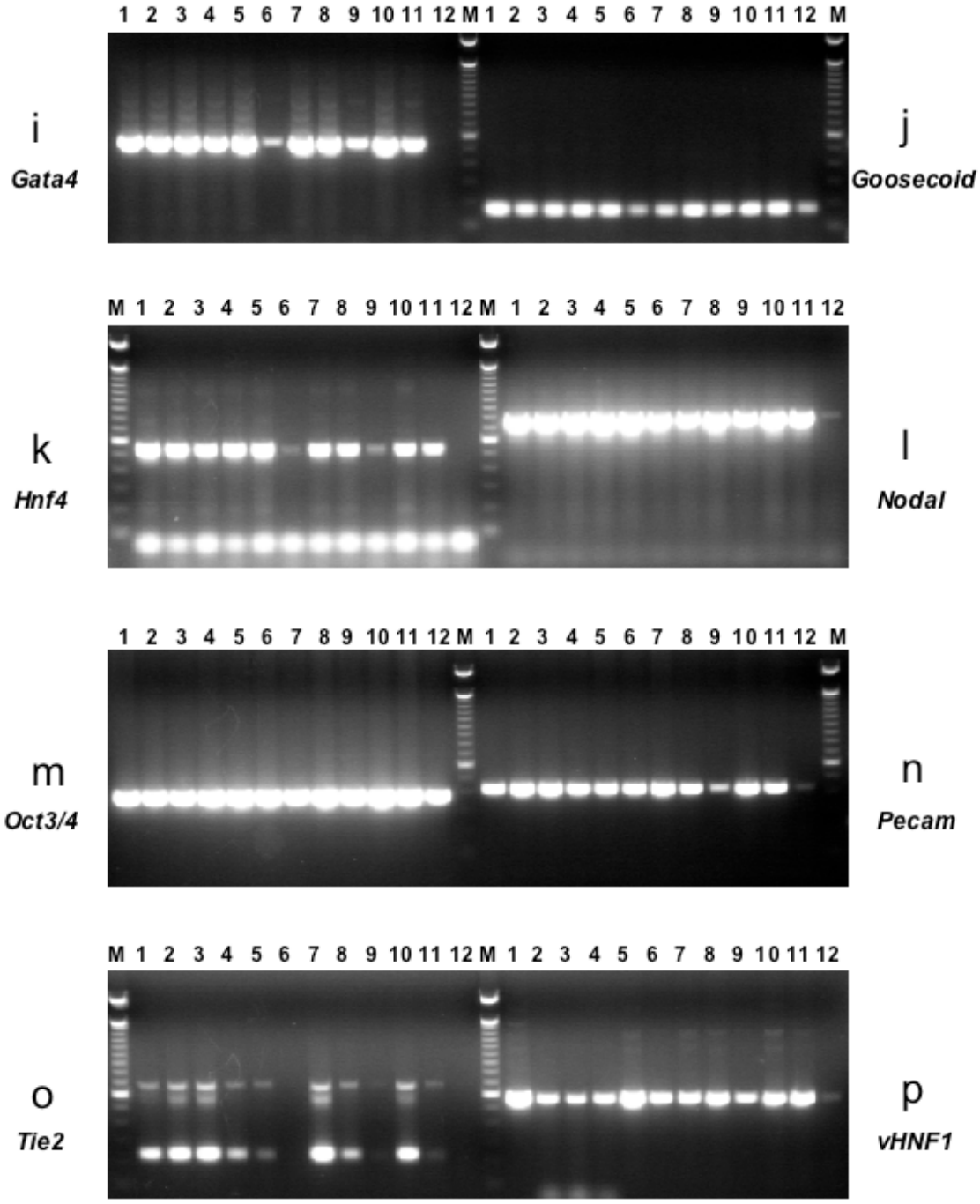
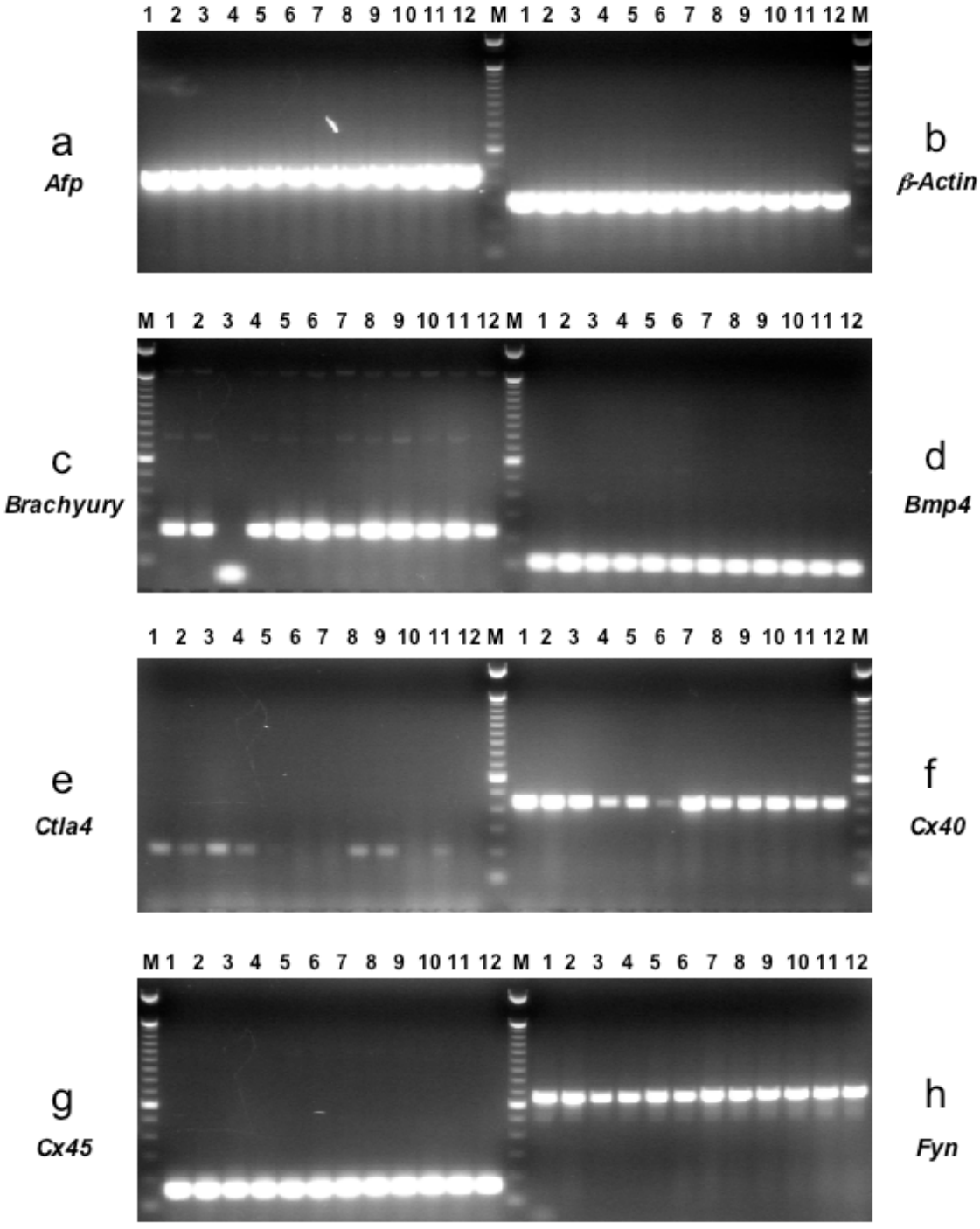
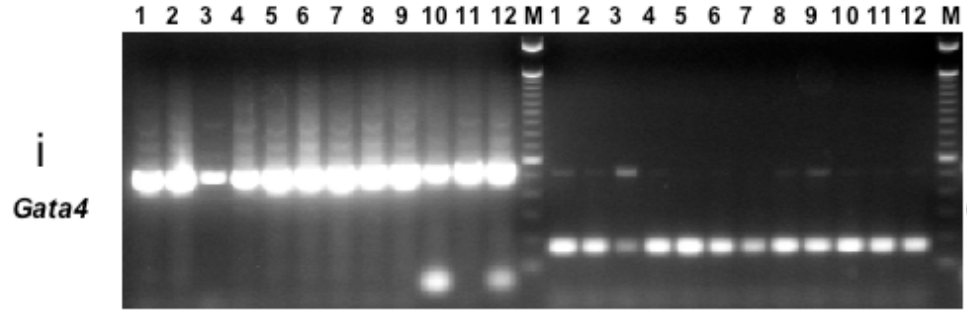

1 2 3 4 5 6 7 8 9 10 11 12 M 1 2 3 4 5 6 7 8 9 10 11 12 M
i
Gata4
j
Goosecoid
M 1 2 3 4 5 6 7 8 9 10 11 12 M 1 2 3 4 5 6 7 8 9 10 11 12
k
Hnf4
l
Nodal
1 2 3 4 5 6 7 8 9 10 11 12 M 1 2 3 4 5 6 7 8 9 10 11 12 M
m
Oct3/4
n
Pecam
M 1 2 3 4 5 6 7 8 9 10 11 12 M 1 2 3 4 5 6 7 8 9 10 11 12
o
Tie2
p
vHNF1

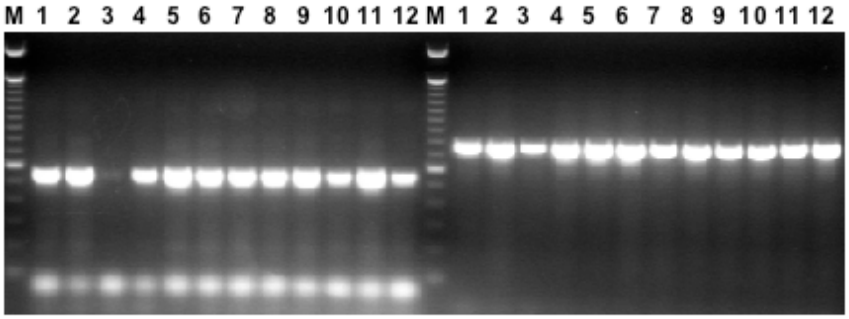
Fig. 5-1 RT-PCR results of Day 5 embryoid bodies. a. *Afp*. **b.** *β-Actin*. Note that line 1, 3, 5, 7, 9 and 11 were mistakenly mixed with PCR products from **a**. *β-Actin* was used as a loading control. **c.** *Brachyury*. **d.** *Bmp4*. **e.** *Ctla4*. **f.** *Cx40*. **g.** *Cx45*. **h.** *Fyn*. **i.** *Gata4*. **j.** *Goosecoid*. **k.** *Hnf4*. **l.** *Nodal*. **m.** *Oct3/4*. **n.** *Pecam*. **o.** *Tie2*. **p.** *vHNF1*. Lane 1, WW103-16B3; 2, WW103-16D2; 3, WW103-17F2; 4, WW103-17G6; 5, WW103-18E8; 6, WW103-18F11; 7, WW103-18G10; 8, WW103-19A1; 9, WW103-19A2; 10, WW103-19A6; 11, WW103-19A8; 12, WW103-19D3; M, 100 bp ladder (Invitrogen)


1 2 3 4 5 6 7 8 9 10 11 12 M 1 2 3 4 5 6 7 8 9 10 11 12 M
a
Afp
b
β-Actin
M 1 2 3 4 5 6 7 8 9 10 11 12 M 1 2 3 4 5 6 7 8 9 10 11 12
c
Brachyury
d
Bmp4
1 2 3 4 5 6 7 8 9 10 11 12 M 1 2 3 4 5 6 7 8 9 10 11 12 M
e
Ctla4
f
Cx40
M 1 2 3 4 5 6 7 8 9 10 11 12 M 1 2 3 4 5 6 7 8 9 10 11 12
g
Cx45
h
Fyn



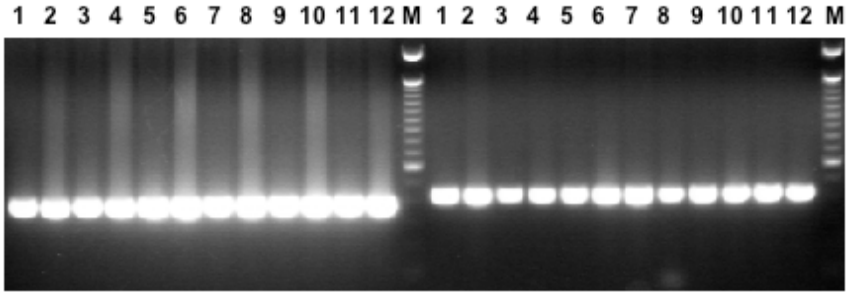
i
Gata4

j
Goosecoid



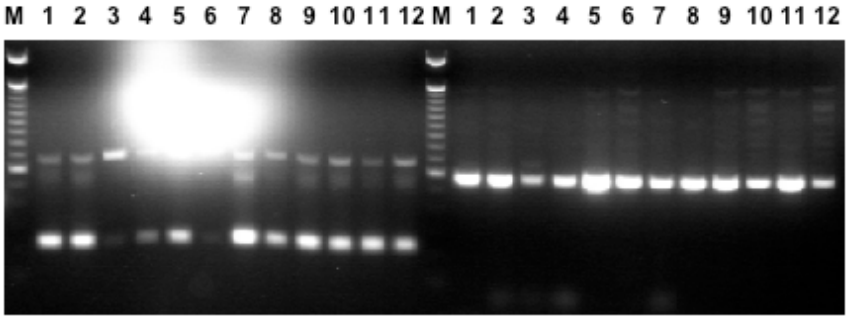
k
Hnf4

l
Nodal



m
Oct3/4

n
Pecam



o
Tie2

p
vHNF1

Fig. 5-2 RT-PCR result of Day 8 embryoid bodies. a. *Afp*. **b.** *β-Actin*. **c.** *Brachyury*. **d.** *Bmp4*. **e.** *Ctla4*. **f.** *Cx40*. **g.** *Cx45*. **h.** *Fyn*. **i.** *Gata4*. **j.** *Goosecoid*. **k.** *Hnf4*. **l.** *Nodal*. **m.** *Oct3/4*. **n.** *Pecam*. **o.** *Tie2*. **p.** *vHNF1*. Lane 1, WW103-16B3; 2, WW103-16D2; 3, WW103-17F2; 4, WW103-17G6; 5, WW103-18E8; 6, WW103-18F11; 7, WW103-18G10; 8, WW103-19A1; 9, WW103-19A2; 10, WW103-19A6; 11, WW103-19A8; 12, WW103-19D3; M, 100 bp ladder (Invitrogen)

To generate this control ES cell line, a Cre expression plasmid was electroporated into the WW69-D6 cell line and mitotic recombination clones were selected in M15 supplemented with HAT. The clones with the desired phenotype were identified both by sib-selection and by Southern analysis using an *E_2DH* 3' probe. The G2-X recombinants are resistant to HAT and blasticidin, but sensitive to G418 and puromycin. Southern analysis of *Nde*I digested genomic DNA will generate a 9.6 kb targeted fragment instead of the 13.1 kb wild-type fragment. One clone with the desired genotype, WW93-A12 and its subclones were used as controls in the ES cell *in vitro* differentiation screen.

5.2.2 Secondary screen

Mutant cell lines that showed an abnormal expression pattern for the markers checked in the primary screen were subcloned and single colonies were picked to avoid cross-contamination by ES cells that did not have the correct genotype. The control cell line, WW93-A12 was also subcloned. Southern analysis was performed on all the subclones to confirm their identities (Fig. 5-3).

The *in vitro* differentiation protocol for the second round screen is essentially the same as that of the first round. But more time points were taken and more molecular markers were checked using RT-PCR. The clones that still showed abnormal expression for the markers checked were characterized individually.

5.2.3 WW103-8E6 (*Pecam*)

One of the mitotic recombination clones, WW103-8E6, have overtly impaired *in vitro* differentiation potential. When the EBs were plated onto the gelatinized tissue culture plates, the EBs could not form cystic three-dimensional structures. When RT-PCR was performed using a series of molecular markers, the expression of some markers in the day 8 EBs was significantly down-regulated compared to the wild-type control, WW93-A12 (Fig. 5-4).

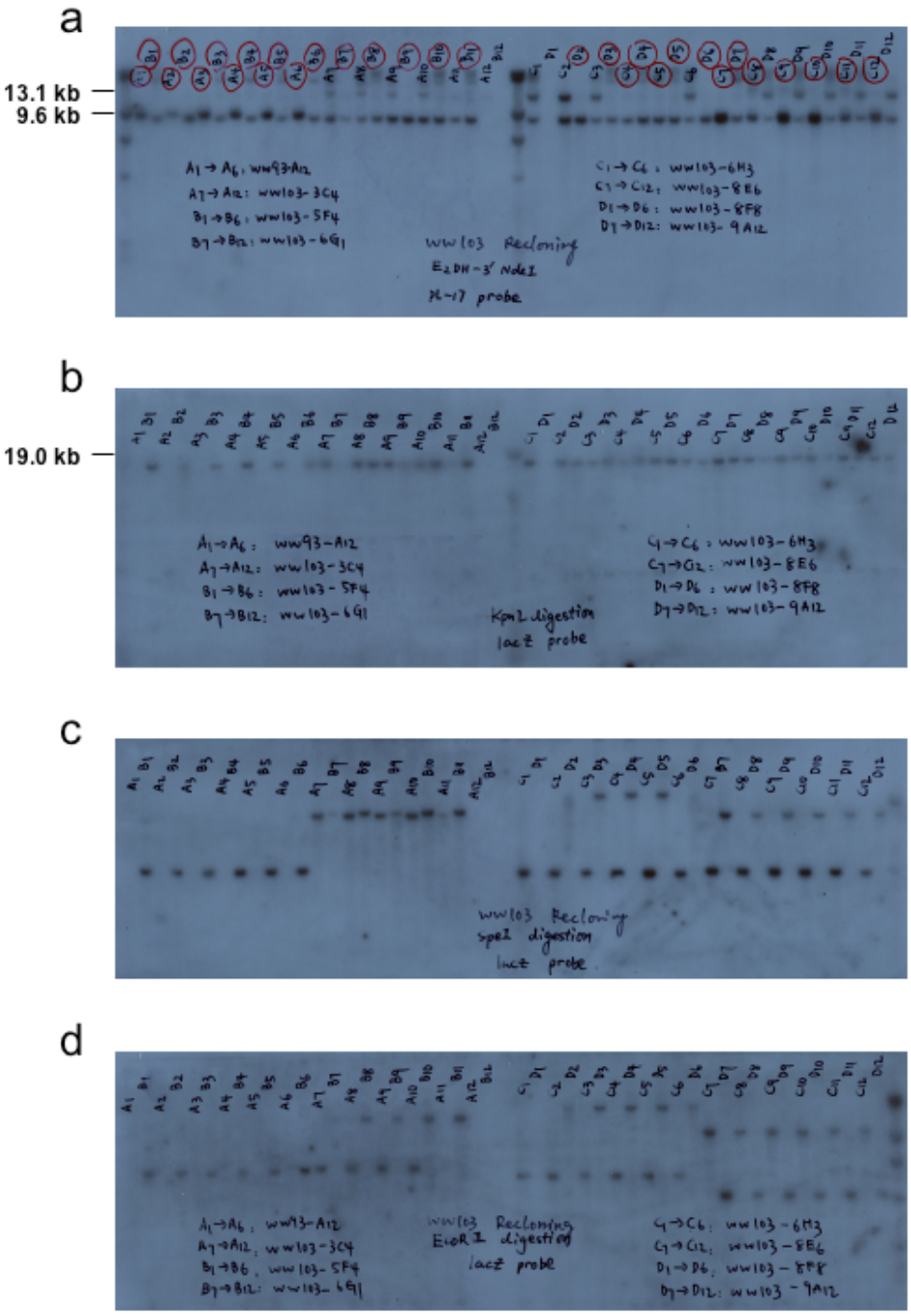
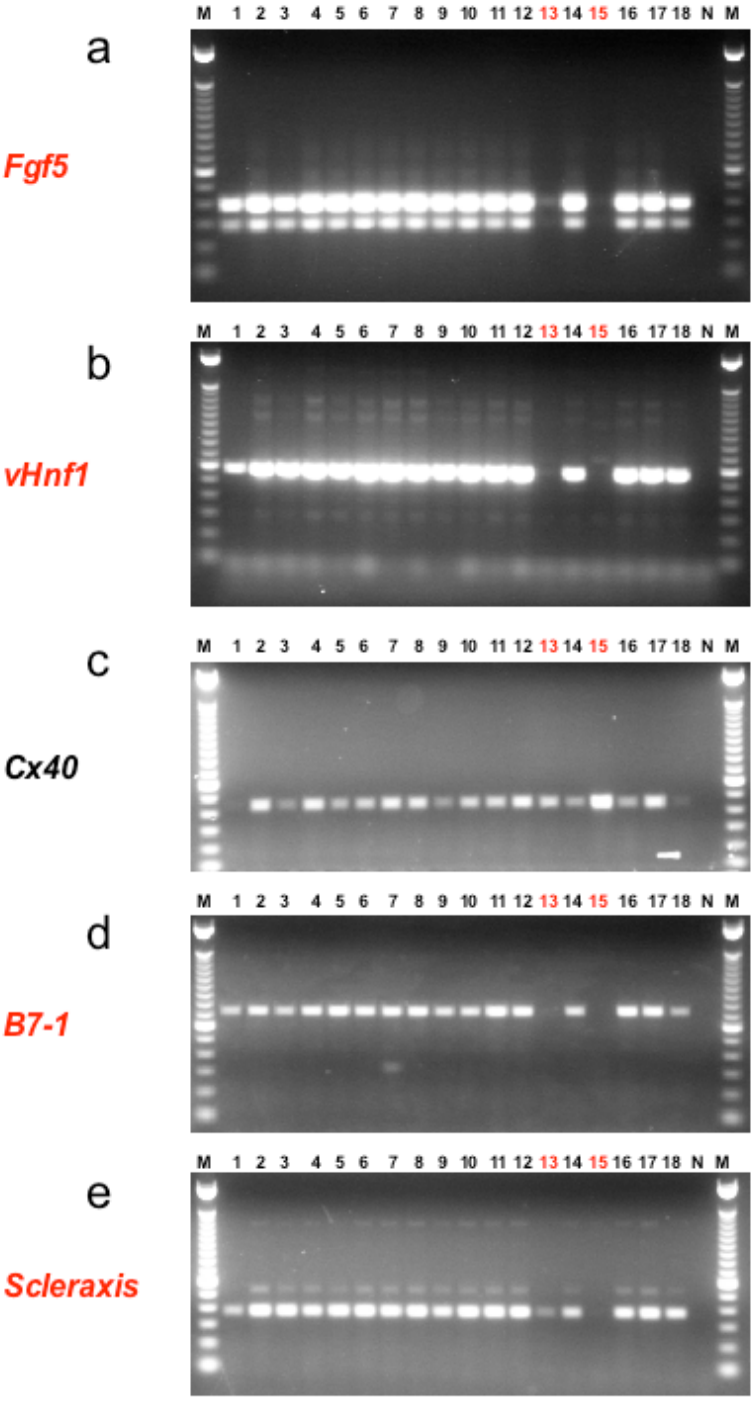
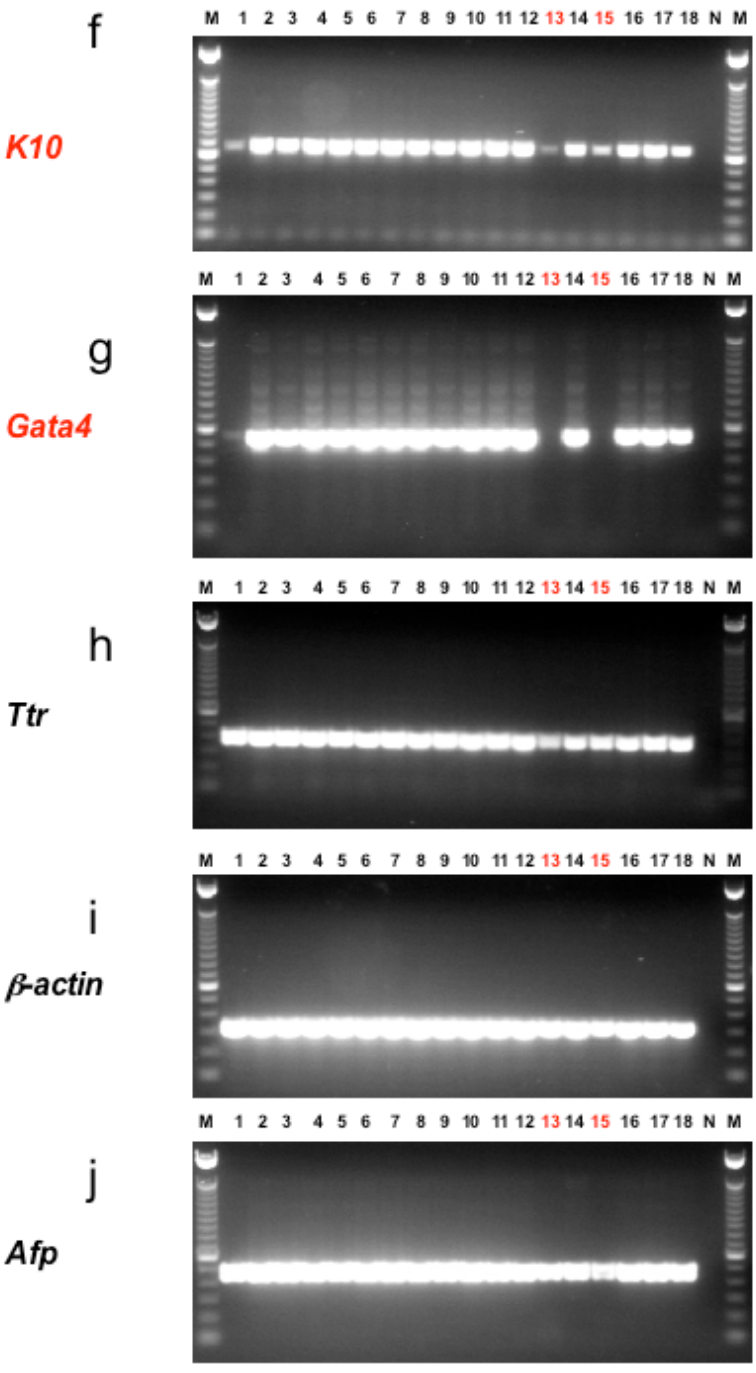

a
13.1 kb
9.6 kb
A1→A6: ww93-A12
A7→A12: ww103-3C4
B1→B6: ww103-5F4
B7→B12: ww103-6G1
ww103 Recloning
E2DH-3' NdeI
PL-17 probe
C1→C6: ww103-6H3
C7→C12: ww103-8E6
D1→D6: ww103-8F8
D7→D12: ww103-9A12
b
19.0 kb
A1→A6: ww93-A12
A7→A12: ww103-3C4
B1→B6: ww103-5F4
B7→B12: ww103-6G1
KpnI digestion
lacZ probe
C1→C6: ww103-6H3
C7→C12: ww103-8E6
D1→D6: ww103-8F8
D7→D12: ww103-9A12
c
ww103 Recloning
SpeI digestion
lacZ probe
d
A1→A6: ww93-A12
A7→A12: ww103-3C4
B1→B6: ww103-5F4
B7→B12: ww103-6G1
ww103 Recloning
EcoRI digestion
lacZ probe
C1→C6: ww103-6H3
C7→C12: ww103-8E6
D1→D6: ww103-8F8
D7→D12: ww103-9A12

Fig. 5-3 Southern analysis of subclones of homozygous mutant ES cell lines. a. Genomic DNA from the subclones was cut with *Nde*I and hybridized to an E_2DH 3' probe. The targeted restriction fragment is 9.6 kb and the wild type fragment is 13.1 kb. Note that some subclones of WW103-6H3 (c1, c2, c3 and c6) are heterozygous. **b.** Genomic DNA from the subclones was cut with *Kpn*I and hybridized to a *lacZ* probe. A 19.0 kb inversion restriction fragment was detected for all the subclones except those of the control cell line, WW93-A12. **c.** Genomic DNA from the subclones was cut with *Spe*I and hybridized to a *lacZ* probe. Proviral/host junction fragments of different sizes were detected for all the subclones except those of the control cell line, WW93-A12. **d.** Genomic DNA from the subclones was cut with *Eco*RI and hybridized to a *lacZ* probe. Proviral/host junction fragments of different sizes were detected for all the subclones except those of the control cell line, WW93-A12.

The *Spe*I/*Xba*I/*Nhe*I Splinkerette PCR product from this clone mapped the proviral insertion site to the first intron of *Pecam* (Platelet endothelial cell Adhesion Molecule Precursor, CD31) (Fig. 5-5a). The 5' RACE product matched an alternative spliced exon (Exon 1b) (Fig. 5-5b). In the Ensembl browser, there are at least three different spliced forms at the 5' end of this gene. *Pecam* transcripts can start from Exon 1a, Exon 1b or a site just 5' to Exon2 (Fig. 5-5c). The open reading frame (ORF) of *PECAM* starts from Exon 2. So the breakpoint in intron 1 created by the inversion would disrupt the transcripts starting from Exon1a and 1b, but it may not affect the transcripts starting from Exon 2. RT-PCR primers were used to determine the expression of different alternative spliced forms of *Pecam* in undifferentiated WW93-A12 and WW103-8E6 ES cells. This analysis revealed that none of the transcripts in undifferentiated ES cells started from Exon 1a (data not shown). In undifferentiated WW93-A12 ES cells, most *Pecam* transcripts start from Exon 1b. However, in undifferentiated WW103-8E6 ES cells, *Pecam* transcripts starting from Exon 2 and Exon 1b were both detected, implying that the inversion did not completely block the transcription across the breakpoint (Fig. 5-5d).

The *in vitro* differentiation of another *Pecam* gene-trap clone, WW103-4A6, showed that the differentiation of this clone was not impaired by the proviral insertion and the breakage caused by inversion. The *Sau*3A1 Splinkerette PCR product fro this clone has mapped the proviral insertion site to the third intron of *Pecam* (Fig. 5-6a and b). RT-PCR analysis of WW103-4A6 during the process of differentiation showed the expression of all the molecular markers during differentiation which was the same as the control cell line, WW93-A12 (data not shown). RT-PCR using a pair of primers specifically designed to amplify Exons 6, 7 and 8 of *Pecam* showed that the Pecam expression was completely blocked in WW103-4A6. On the other hand, WW103-8E6 and WW103-8G9, subclones in the same group as 8E6, showed normal *Pecam* expression (Fig. 5-6c).


M 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 N M
a
Fgf5
M 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 N M
b
vHnf1
M 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 N M
c
Cx40
M 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 N M
d
B7-1
M 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 N M
e
Scleraxis


f
K10
M 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 N M
g
Gata4
M 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 N M
h
Ttr
M 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 N M
i
β-actin
M 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 N M
j
Afp
M 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 N M

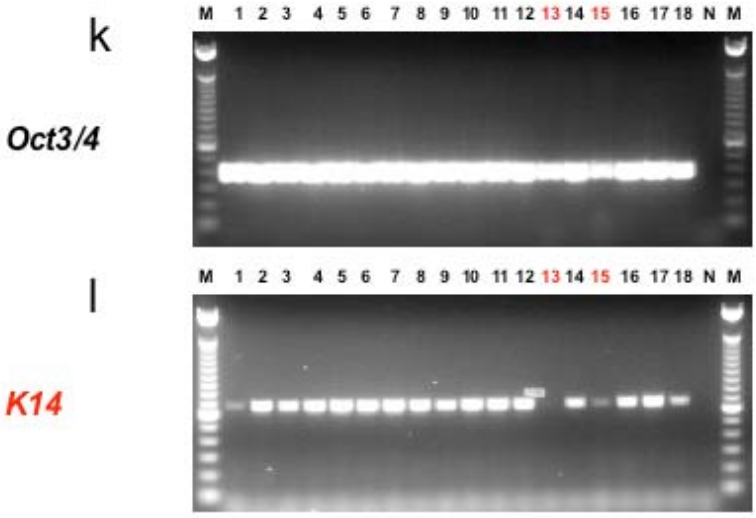
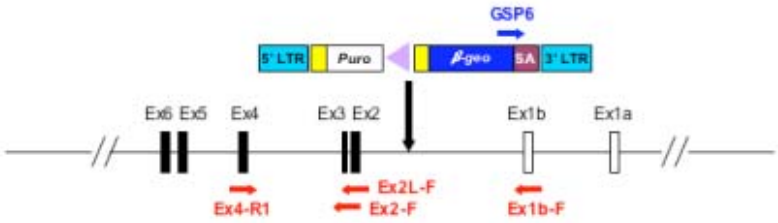


Fig. 5-4 RT-PCR results of Day 8 embryoid bodies. a. *Fgf5*. **b.** *vHnf1*. **c.** *Cx40*. **d.** *B7-1*. **e.** *Scleraxis*. **f.** *K10*. **g.** *Gata4*. **h.** *Ttr*. **i.** *β-actin*. **j.** *Afp*. **k.** *Oct3/4*. **l.** *K14*. Note that lane 13 is WW103-8E6 (marked in red). Lane 1 is the AB2.2 wild type ES cell line. Lane 2 is the WW93-A12 control line, all the other lanes are irrelevant mutant cell lines. At day 8, EBs derived from WW103-8E6 didn't express *Fgf5, vHnf, B7-1, Gata4 and K14*. The expression of *Scleraxis* and *K10* was also down-regulated compared to WW93-A12 control. Markers that are down-regulated are marked in red. Interestingly, the expression of various markers differs between EBs from AB2.2 (lane 1) and those from WW93A12 (lane 2).The expression pattern of the mutant cell lines is more similar to WW93-A12. In the whole screen, WW93A12 and its subclones were used as control.

a

b

c


GSP6
5' LTR
Puro
β-geo
SA
3' LTR
Ex6 Ex5
Ex4
Ex3 Ex2
Ex1b
Ex1a
Ex4-R1
Ex2L-F
Ex2-F
Ex1b-F

d

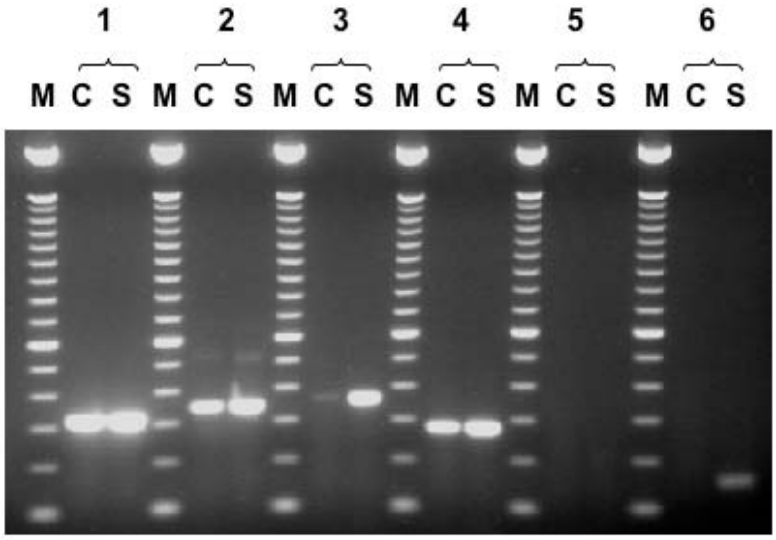
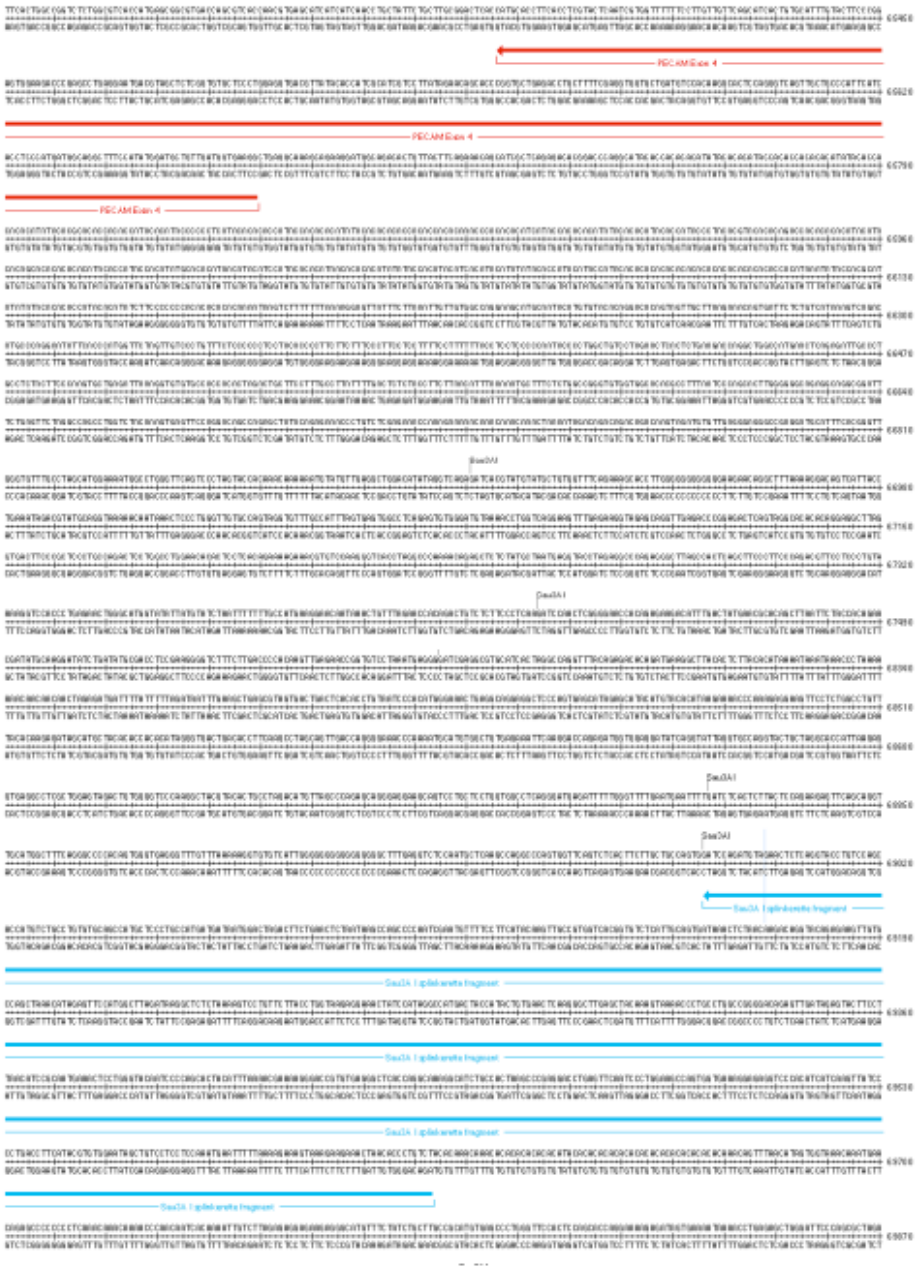

1
2
3
4
5
6
M C S M C S M C S M C S M C S M C S

Fig. 5-5 WW103-8E6. a. Splinkerette PCR result. *Spe*I/*Xba*I/*Nhe*I Splinkerette PCR product mapped the proviral insertion site to the first intron of the *Pecam* gene. The second and the third exons of *Pecam a*re marked in red, and the splinkerette PCR fragment is marked in blue. **b.** 5'RACE result. The 5' RACE product was mapped to an alternative exon of *Pecam*.The splice acceptor of the SA-βgeo cassette is marked in black, the *Pecam* alternative exon 1b is marked in red, the polyC tail is marked in green. **c.** Schematic illustration of the proviral insertion in WW103-8E6. *Pecam* gene specific primers were designed to decide the structure of the clone. PECAM-Ex1b-F, PECAM-Ex2L-F, PECAM-Ex2-F and PECAM-Ex4-R1 primers are shown as red arrows, SA-βgeo cassette specific primer GSP6 is shown as a blue arrow, the coding exons are marked in black. **d.** RT-PCR result. 1: positive control, *β-actin* RT primers; 2: PECAM-Ex1b-F/PECAM-Ex4-R1; 3: PECAM-Ex2L-F/PECAM-Ex4-R1; 4: PECAM-Ex2-F/PECAM-Ex4-R1; 5: negative control, no primers were added. 6: PECAM-Ex1b-F/GSP6. M, 100 bp ladder (Invitrogen); C: RNA extracted from undifferentiated WW93-A12 ES cells; S: RNA extracted from undifferentiated WW103-8E6 ES cells. Note that in undifferentiated WW93-A12 ES cells, most *Pecam* transcripts start from Exon 1b. In undifferentiated WW103-8E6 ES cells, a large portion of Pecam transcripts start from Exon 2. But significant amount of transcripts started from Exon 1b were still detected.

a


PECAM Exon 4
PECAM Exon 4
PECAM Exon 4
Sau3AI
Sau3AI
Sau3AI
Sau3A I gDNA enrich fragment
Sau3A I gDNA enrich fragment
Sau3A I gDNA enrich fragment
Sau3A I gDNA enrich fragment

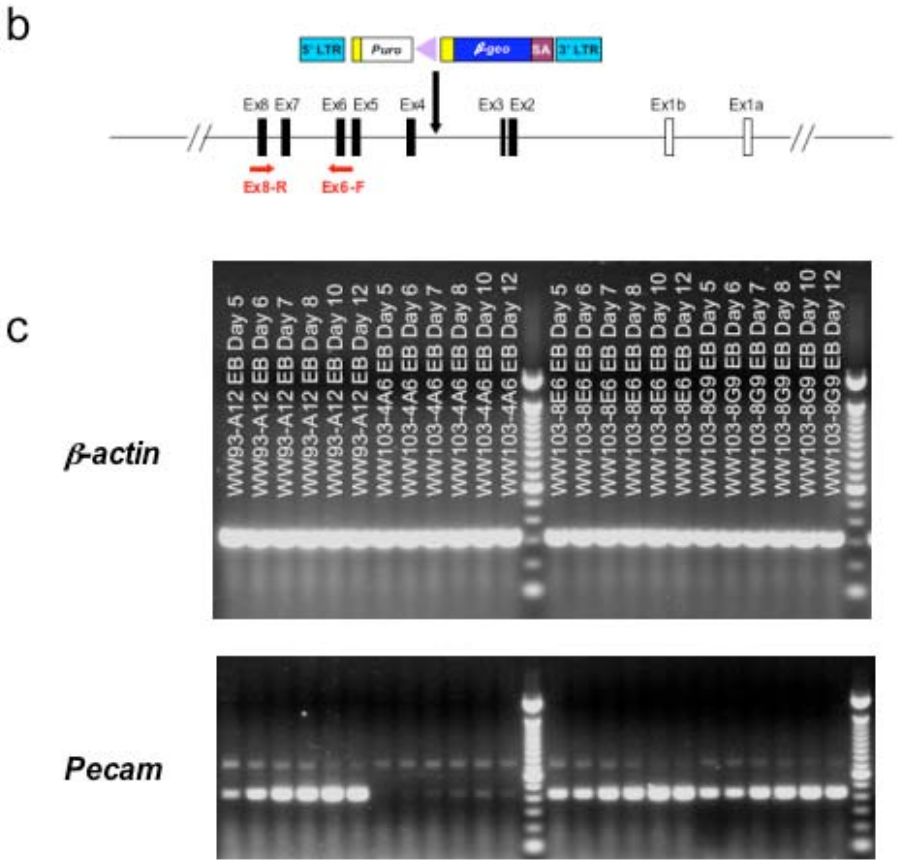


Fig. 5-6 WW103-4A6. a. Splinkerette PCR result. *Sau*3AI Splinkerette PCR product mapped the proviral insertion site to the third intron of the *Pecam* gene. The forth exon of *Pecam* is marked in red, and the splinkerette PCR fragment is marked in blue. **b.** Schematic illustration of the proviral insertion in WW103-4A6. *Pecam* gene specific primers were designed to confirm the structure of the allele. PECAM-Ex6-F and PECAM-Ex8-R primers are shown as red arrows, the coding exons are marked in black. **c.** RT-PCR result. The expression of *Pecam* during *in vitro* differentiation of WW93-A12, WW103-4A6, WW103-8E6 and WW103-8G9 was examined by RT-PCR using PECAM-Ex6-F and PECAM-Ex8-R primers. Note that the *Pecam* expression was completely blocked in WW103-4A6, but not in WW103-8E6 and WW103-8G9.

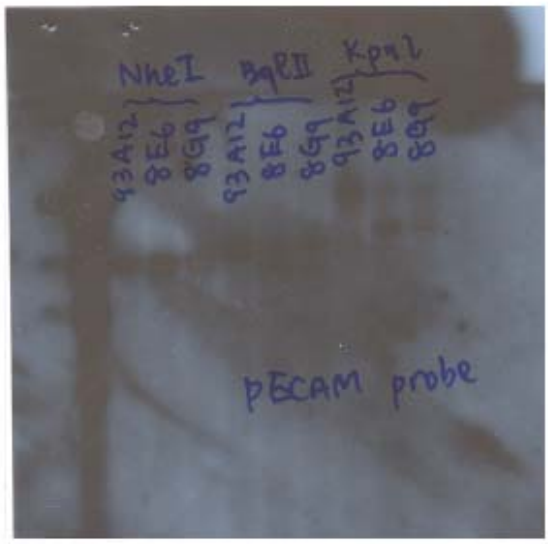
In an attempt to resolve how this situation could have occurred, Southern analysis was performed using a *Pecam* specific probe (Fig. 5-7). This revealed that both WW103-8E6 and WW103-8G9 were heterozygous for *Pecam* locus. But interestingly, the ratio between the targeted restriction fragment and the wild-type restriction fragment is not 1:1. For WW103-8E6, the ratio is around 2:1, while for WW103-8G9 the ratio is around 1:2. The unexpected Southern result suggested that both clones might be trisomic. If so, it is most likely that the trisomy appeared after the end point cassette targeting and before the retrovirus infection. In this case, the original trisomy would contain two 3' *Hprt* chromosomes targeted with the end point cassette, and one 5' *Hprt* wild-type chromosome. After regional trapping, the puromycin resistant trisomy will have one 3' *Hprt* chromosome with an inversion, one 3' *Hprt* chromosome with targeted end point cassette and one 5' *Hprt* wild-type chromosome. Induced mitotic recombination can generate two different products: clones with two inversion chromosomes and one chromosome with the end point cassette (WW103-8E6), or clones with one inversion chromosome and two chromosomes with the end point cassette (WW103-8G9). In both cases, the clones will carry three targeted E_2DH alleles (end point targeting), thus Southern analysis using E_2DH probe can not distinguish these trisomies from the homozygous inversion clones.

Therefore the impaired differentiation potential of WW103-8E6 does not have any direct connection with the *Pecam* trapping and the subsequent inversion. This may be the result of the up-regulation of the chromosome 11 genes caused by the extra chromosome.

5.2.4 WW103-14F11 (*2810410L24Rik*)

As described in the previous chapter, the WW103-14F11 subclone has a proviral insertion at the *2810410L24Rik* locus (119.9 Mb) (Fig. 5-8a), which is close to the telomere of the chromosome 11. But instead of trapping the *2810410L24Rik* gene, which is transcribed from the sense strand (from centromere to telomere), the retrovirus trapped another transcript transcribed from the anti-sense strand (from telomere to centromere), *D030042H08Rik*.

a


NheI
BglII
KpnI
93A12
8E6
8G9
93A12
8E6
8G9
93A12
8E6
8G9
pECAM probe

b

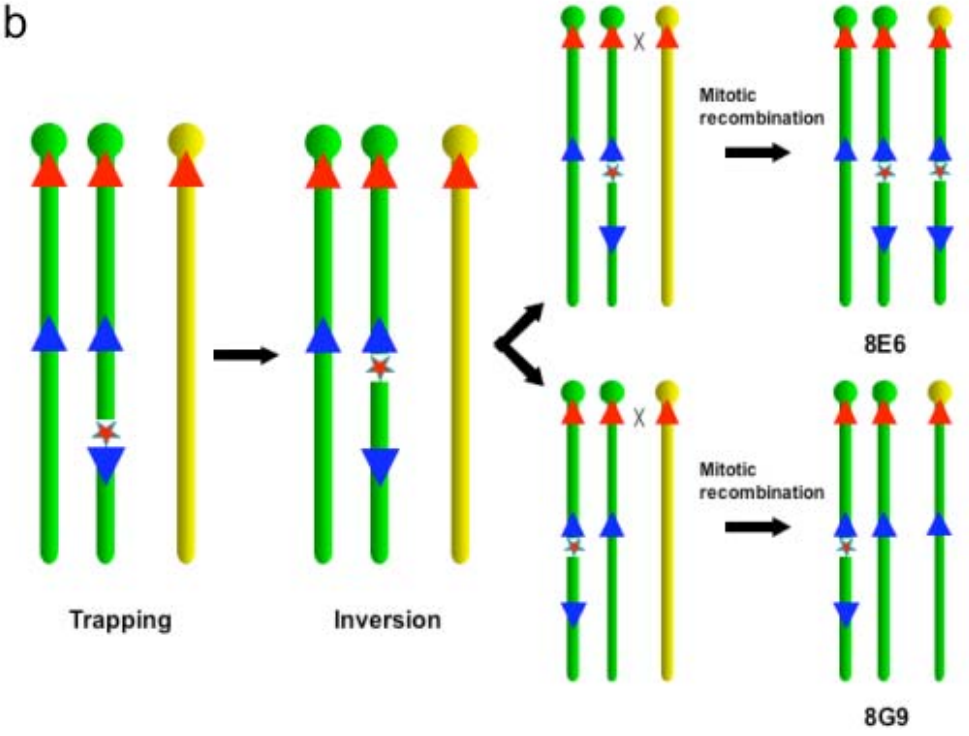
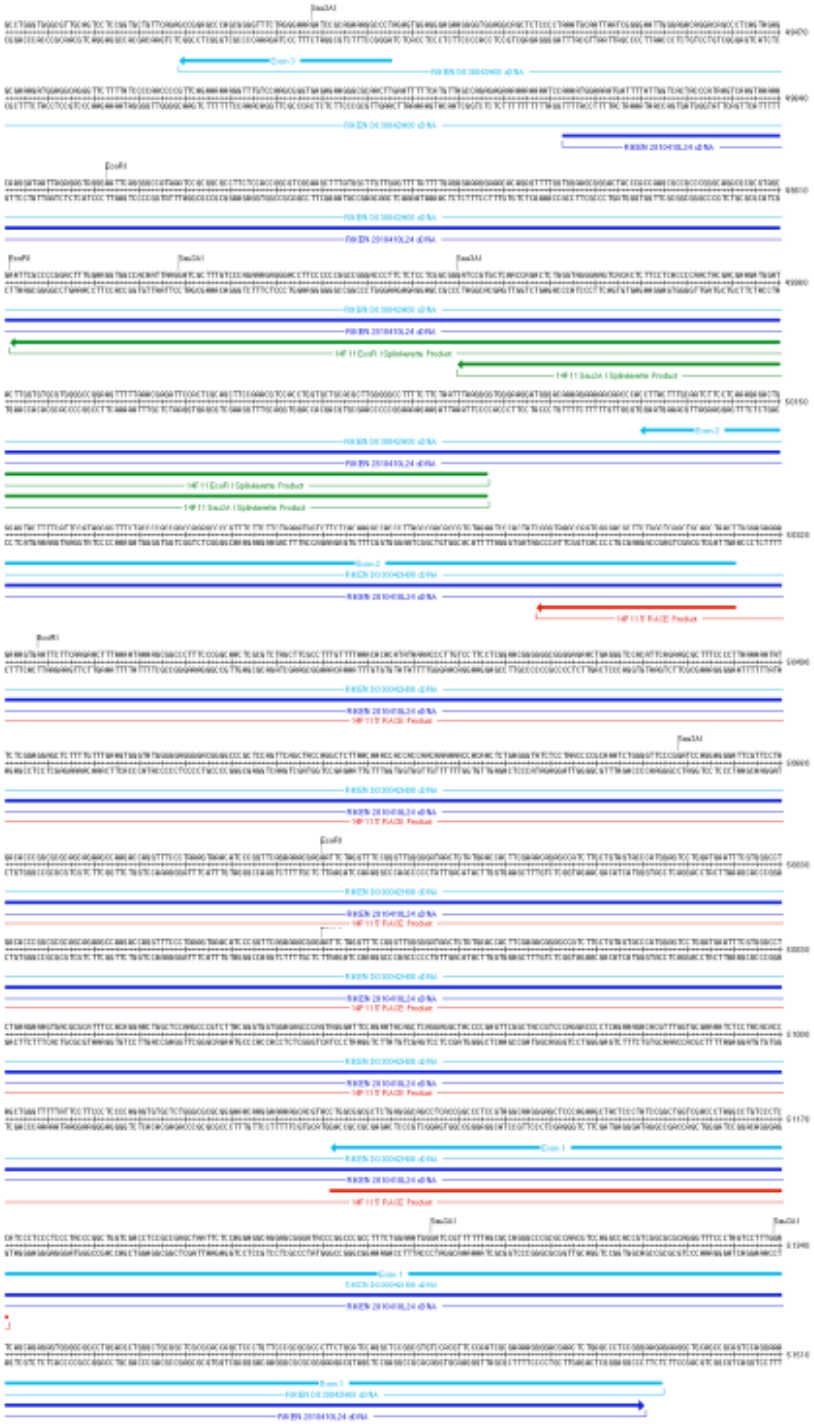
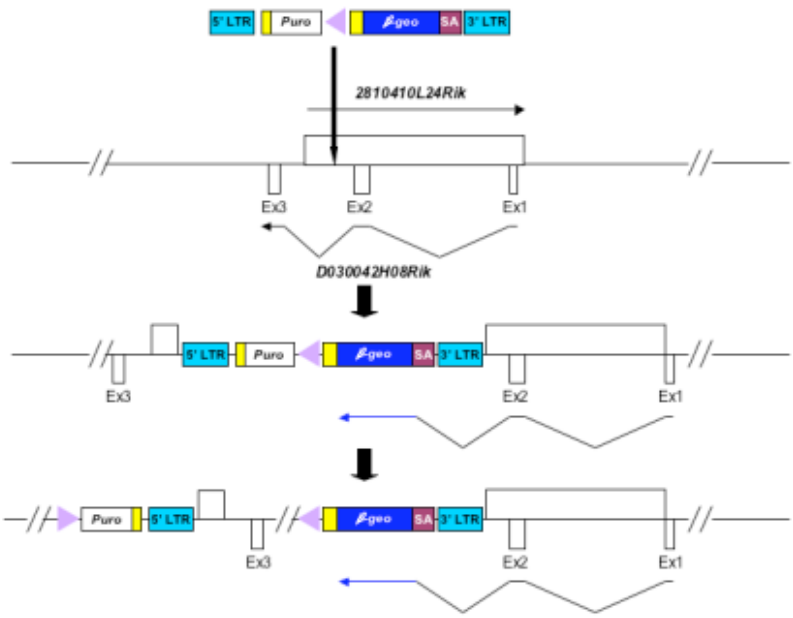

Trapping
Inversion
Mitotic recombination
8E6
Mitotic recombination
8G9

Fig. 5-7 WW103-8E6 and 8G9 are both chromosome 11 trisomies. a. Southern analysis of clones, WW103-8E6 and WW103-8G9. A *Pecam* specific probe was used to detect a 10.6 kb *Bgl*II wild type fragment and a 9 kb *Bgl*I targeted fragment. The same probe was also used to detect a 30 kb *Kpn*I wild type fragment and a 19 kb targeted fragment. Both WW103-8E6 and WW103-8G show a wild type fragment for both digestions. But the ratio between the targeted restriction fragment and the wild type restriction fragment is not 1:1. For WW103-8E6, the ratio is around 2:1, while for WW103-8G9 the ratio is around 1:2. The Southern result suggested that both clones might be trisomic. **b.** A schematic illustration of possible recombination in WW103-8E6 and WW103-8G. The original trisomy probably contained two 3' *Hprt* chromosomes targeted with the end point cassette, and one 5' *Hprt* wild type chromosome. After regional trapping, the puromycin resistant trisomy has one 3' *Hprt* chromosome with the inversion, one 3' *Hprt* chromosome with the end point cassette and one 5' *Hprt* wild type chromosome. Induced mitotic recombination can generate two different products, depending on which 3' *Hprt* chromosome participates in the recombination process. Clones can be generated with two inversion chromosomes and one chromosome with the targeted end point cassette (WW103-8E6), or clones with one inversion chromosome and two chromosomes with the targeted end point cassette (WW103-8G9).

a



b


5' LTR
Puro
βgeo
SA
3' LTR
2810410L24Rik
Ex3
Ex2
Ex1
D030042H08Rik

c

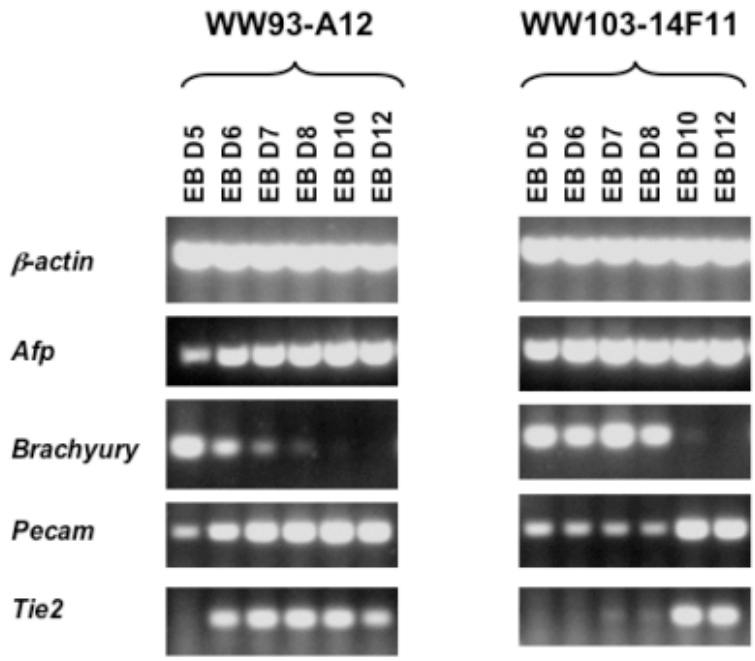

WW93-A12
WW103-14F11
EB D5
EB D6
EB D7
EB D8
EB D10
EB D12
β-actin
Afp
Brachyury
Pecam
Tie2

Fig. 5-8 WW103-14F11. a. Splinkerette PCR and 5' RACE.results. *Eco*RI and *Sau*3AI Splinkerette PCR products have both mapped the proviral insertion site to the *2810410L24Rik* locus. The 5' RACE result matched another transcript *D030042H08Rik*.The *2810410L24Rik* transcript is marked in dark blue, the *D030042H08Rik* transcript is marked in light blue, the splinkerette PCR fragments are marked in green and the 5'RACE product is marked in red. **b.** Schematic illustration of the structure of proviral insertion and the subsequent inversion in WW103-14F11. **c.** RT-PCR results of EBs derived from WW013-14F11 and WW93-A12. Note that *Pecam* and *Tie2* RT-PCR results showed that the up-regulation of these two markers in the differentiation process was significantly delayed. On the other hand, the early mesoderm marker, *Brachyury*'s down-regulation was also delayed.

The splinkerette results mapped the proviral insertion site between the second and third exons of *D030042H08Rik*. However, the 5' RACE result did not match the *D030042H08Rik* cDNA sequence perfectly, although the transcript structure is similar. It is possible that the 5' RACE result and the *D030042H08Rik* cDNA sequence represent two different alternative splice forms of the same gene.

Nevertheless, the gene-trap retrovirus insertion and the subsequent inversion will disrupt the transcripts from both strands (Fig. 5-8b). The *in vitro* differentiation results showed that EBs derived from WW103-14F11 have impaired potential to develop into endothelial cells. RT-PCR using *Pecam* and *Tie2* primers has shown that the up-regulation of the expression of these two markers during the differentiation process was significantly delayed. On the other hand, the early mesoderm marker, *Brachyury*'s down-regulation was also delayed (Fig. 5-8c).

Further confirmation of this subclone is still undergoing. One way to directly confirm the defective endothelial cell differentiation is to use collagen IV coated dishes to induce undifferentiated ES cells to first differentiate into $Flk1^+$ cells (Yamashita, Itoh et al. 2000). When FACS sorted $Flk1^+$ cells were cultured with the addition of VEGF, these cells will further differentiate into PECAM1+ sheets of endothelial cells, which also express other endothelial cell-specific markers, such as *VE-cadherin* and *CD34*. By comparing the endothelia cell differentiation of the WW103-14F11 cells and the wild-type control cells, it will be possible to identify the molecular mechanism underlying the defective phenotype and determine at which stage the differentiation into endothelial lineage is blocked. However, it will be difficult to distinguish the phenotypes of the two genes transcribed from the opposite directions.

5.2.5 WW103-13D10 (*LOC217071*)

*Sau*3A1 and SpeI/*Xba*I/*Nhe*I Splinkerette PCR products mapped the proviral insertion site in the WW103-13D10 clones to the second intron of a hypothetical mouse gene, *LOC217071* (Fig. 5-9a and b). This gene is transcribed from the sense strand (from the centromere to telomere), and

Southern analysis using an *E_2DH* 3' probe has confirmed that this clone is homozygous for the targeted *E_2DH* allele. Southern analysis using a *LacZ* probe has shown that it only carried a 6.9 kb *Kpn*I restriction fragment which suggested that WW103-13D10 contains an intact proviral insertion. As discussed in the previous chapter, this might be caused by a G2 *trans* recombination event. The duplication chromosome has both a functional *Puro* and a functional *Bsd* cassette, and it can become homozygous after induced mitotic recombination because it has not lost any genetic material.

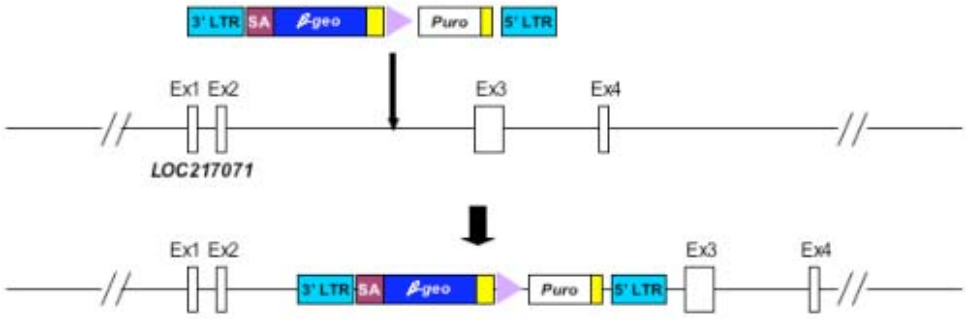
The homozygous duplication clone showed an obvious abnormality in *in vitro* differentiation. The undifferentiated WW103-13D10 ES cells expressed high levels of markers for differentiated cell types, such as *Afp*, *Gata4* and *Hnf4*. The expression of undifferentiated ES cell markers, like *Nodal* and *Oct3/4*, was significantly down-regulated, compared to the WW93-A12 control (Fig. 5-9c).

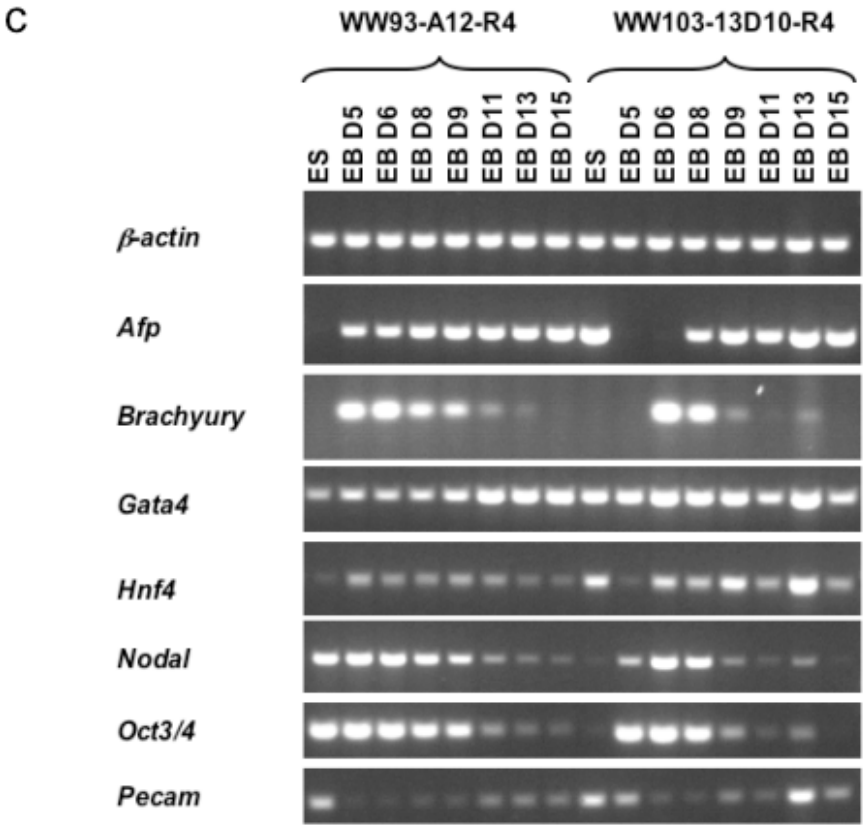
Interestingly, during the process of differentiation, the EBs made from WW103-13D10 ES cells seemed to differentiate normally. At day 5, they lost the expression of *Afp*, but regained the expression of *Nodal* and *Oct3/4*. After this, various markers showed expression patterns similar to those which were observed in the WW93-A12 control. But *Hnf4* and *Gata4* expression were still significantly up-regulated compared to the control.

Sib-selection was carried out on two subclones each from WW93-A12 and WW103-13D10. The same number of undifferentiated ES cells were plated into the wells of a gelatinized 24-well plate and selected in M15, M15+G418, M14+puromycin, M15+blasticidin and M15+HAT, respectively. WW103-13D10 was resistant to both puromycin and blasticidin, which suggested that this clone have two duplication chromosomes, instead of two inversion chromosomes (Fig. 5-9d). Most likely, the phenotype observed in WW103-13D10 was caused by the duplication, instead of the disruption of the *LOC217071* locus.

a

b


3' LTR
SA
β-geo
Puro
5' LTR
Ex1 Ex2
Ex3
Ex4
LOC217071
Ex1 Ex2
3' LTR
SA
β-geo
Puro
5' LTR
Ex3
Ex4


c
WW93-A12-R4
WW103-13D10-R4
ES
EB D5
EB D6
EB D8
EB D9
EB D11
EB D13
EB D15
β-actin
Afp
Brachyury
Gata4
Hnf4
Nodal
Oct3/4
Pecam

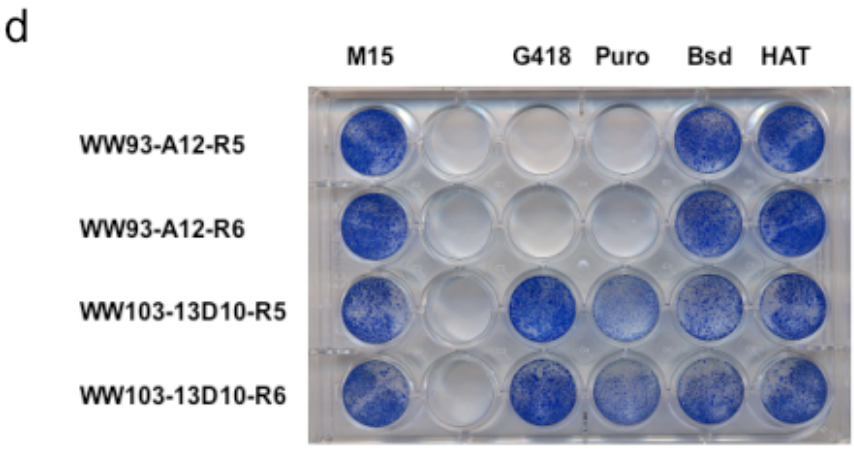

d
M15
G418
Puro
Bsd
HAT
WW93-A12-R5
WW93-A12-R6
WW103-13D10-R5
WW103-13D10-R6

Fig. 5-9 WW103-13D10 a. Splinkerette PCR results. *Spe*I/*Xba*I/*Nhe*I and *Sau*3AI Splinkerette PCR products have mapped the proviral insertion site to the *LOC217071* locus. The splinkerette PCR fragments are marked in green. **b.** Schematic illustration of the structure of proviral insertion in WW103-14F11. **c.** RT-PCR results of EBs derived from WW013-14F11 and WW93-A12. Note undifferentiated WW103-13D10 ES cells express high amounts of *Afp*, *Hnf4* and *Gata4,* but low amount of *Oct3/4* and *Nodal*. Despite the altered expression of these markers,the EBs derived from WW103-13D10 ES cells seem to differentiate normally. **d.** Sib-selection of WW93-A12 and WW103-13D10 clones. Two subclones of each cell line were used for the sib selection. Same amount of undifferentiated ES cells were plated onto a gelatinized 24-well plate and selected with M15, M15+G418, M14+Puromycin, M15+Blasticidine and M15+HAT, respectively. WW103-13D10 was resistant to both Puromycin and Blasticidine, which suggested that this clone is a homozygous duplication, instead of a homozygous inversion.

5.2.6 WW103-18F11 (*Acly*)

One of the mitotic recombination clones, WW103-18F11, showed impaired *in vitro* differentiation potential. After EBs made of WW103-18F11 ES cells were plated on the gelatinized tissue culture plates at Day 5, the EBs did not form cystic three-dimensional structures.

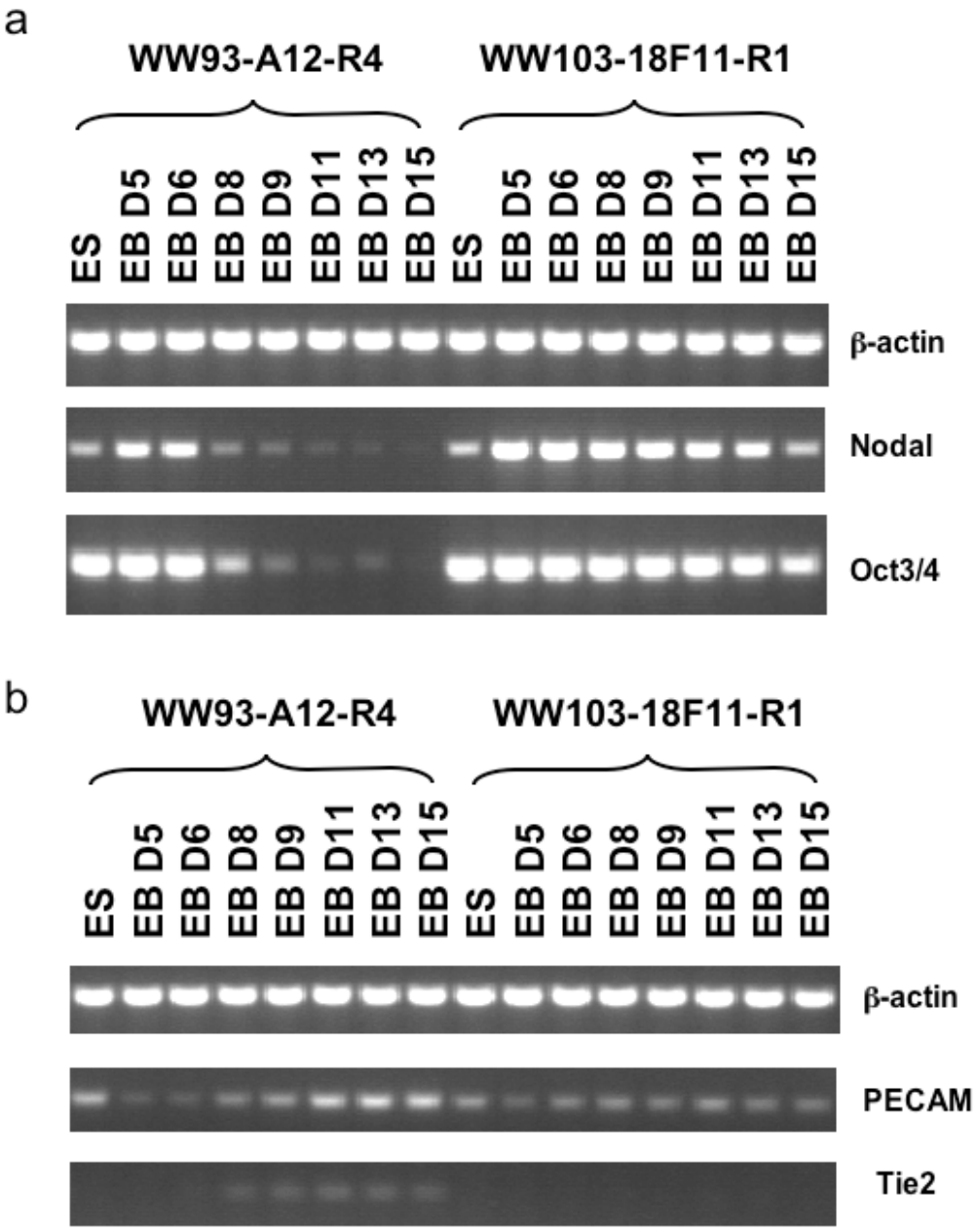
When RT-PCR was performed on RNA extracted from WW103-18F11 embryoid bodies collected at different time points, these EBs were found to express high levels of the undifferentiated ES cell markers, *Oct3/4* and *Nodal*, as late as Day 18 of the *in vitro* differentiation protocol. The expression of *Oct3/4* and *Nodal* still decreased a little during the differentiation process, but down-regulation was not as rapid as that in the control cell line (Fig. 5-10a and e).

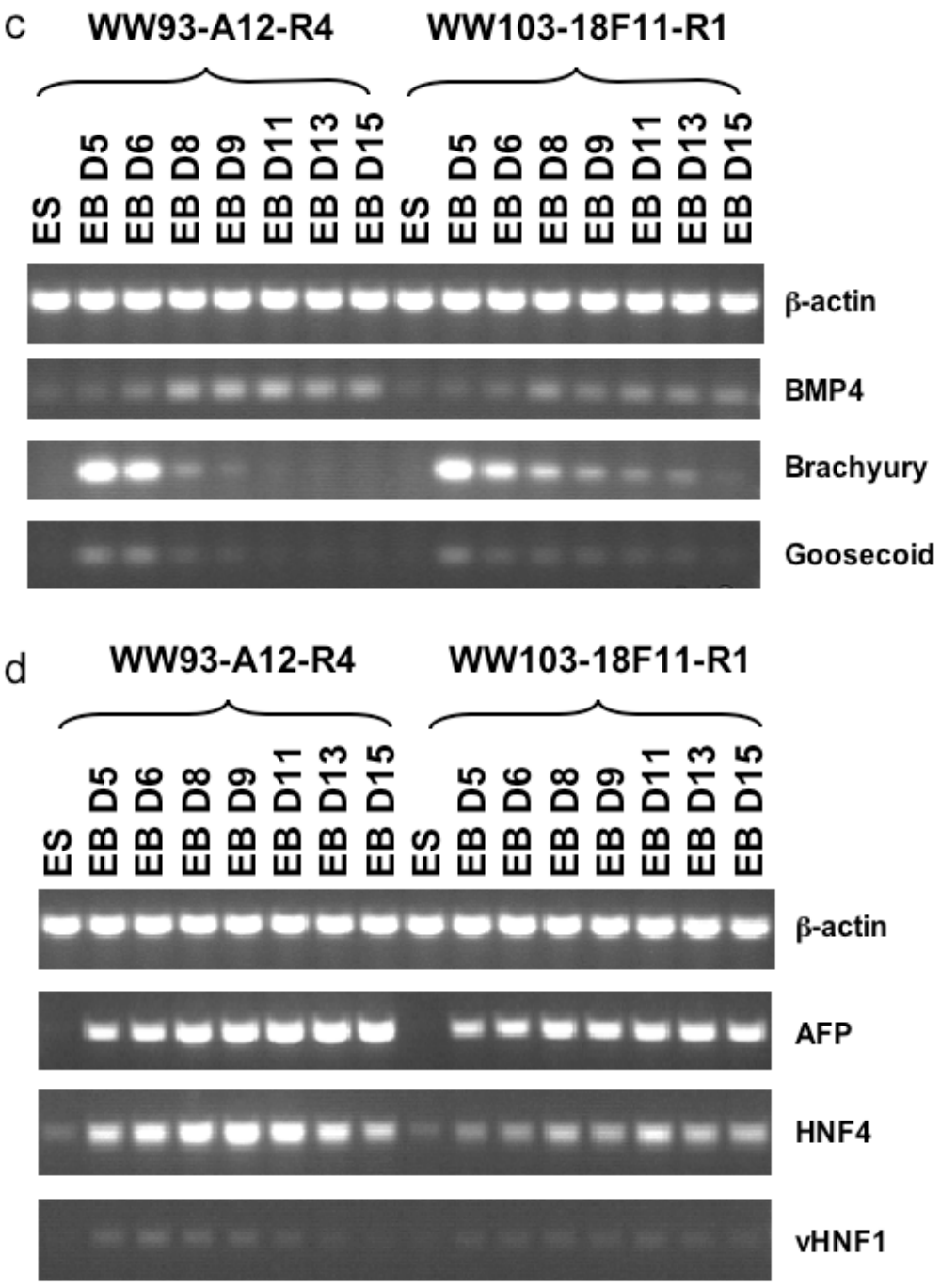
Tie2 expression was not detected during the whole process of differentiation of WW103-18F11 cells. The expression of *Pecam* was maintained at a constant basal level, instead of being up-regulated, as was observed in the WW93-A12 control cell line (Fig. 5-10b). Both of the markers are endothelial cell-specific proteins expressed during the formation of vascular structures in ES-derived EBs. The *Tie2* gene encodes a growth factor receptor, while the *Pecam* protein is an endothelial cell specific antigen. Vittet *et al*. (1996) has shown that both genes are expressed at low levels in undifferentiated ES cells. Normally, in the process of *in vitro* differentiation, the expression of both genes is absent at day 0-3 and is detected again from day 4. After this, the expression level of both genes is consistently up-regulated, as detected by Northern blotting and/or Immunofluorescence. However, in that experiment, only EBs from Day 3 to Day 7 were checked (Vittet, Prandini et al. 1996). In my experiment, I have observed the expression of *Tie2* and *Pecam* in the control line throughout the 15-day differentiation process. Thus, my observation suggested that the differentiation of endothelial cells in the mutant cell line was significantly impaired over the entire 15-day differentiation process.

The expression pattern of the early mesodermal markers (*Bmp4*, *Brachyury*, *Goosecoid*) is similar between WW103-18F11 and WW93-A12 (Fig. 5-10c). However, low levels of expression of *Brachyury* and *Goosecoid* were still detected in WW103-18F11 derived EBs collected at later stages of the differentiation process, while no expression of these markers were detected in later stage EBs derived from WW93A12. The expression of one of the endodermal markers, *Hnf4*, in WW103-18F11 was much lower than that in the control. Apart from these changes, no major differences were observed in the levels of expression of the other markers (Fig. 5-10d).

5' RACE results revealed that the gene-trap retrovirus trapped Exon 1 of ATP-citrate lyase (*Acly*) (Fig. 5-11a). *Acly* is one of two cytosolic enzymes in eukaryotes that synthesize acetyl-coenzyme A (acetyl-CoA), the other enzyme is acetyl-CoA synthetase 1. *Acly* catalyzes the formation of acetyl-coenzyme A (CoA) from citrate and CoA, and hydrolyzes ATP to ADP and phosphate. Because acetyl-CoA is an essential component for cholesterol and triglycerides synthesis, *Acly* is believed to be a potential therapeutic target for hyperlipidemias ad obesity (Beigneux, Kosinski et al. 2004).

To characterize this mutant cell line further, pure subclones of WW103-18F11 were derived by low density plating to form single colonies. Six subclones were picked and expanded. To confirm chromosomal structure of these subclones, sib-selection was performed on two of the WW103-18F11 subclones, WW103-18F11-R1 and WW103-18F11-R6, as well as two subclones of the control cell line, WW93-A12-R4 and WW93-A12-R5. An equal number of ES cells from each subclone were plated onto multiple 6-well plates and selected with M15, M15+puromycin, M15+blasticidin, M15+G418 and M15+HAT, respectively. As expected, the two WW103-18F11 subclones are $Puro^R$, Neo^R, Bsd^S and HAT^R, and the two WW93-A12 subclones are $Puro^S$, Neo^S, Bsd^R and HAT^R. The drug resistance pattern of WW103-18F11 suggests that WW103-18F11 have undergone correct recombination (Fig. 5-11b).


a
WW93-A12-R4
WW103-18F11-R1
ES
EB D5
EB D6
EB D8
EB D9
EB D11
EB D13
EB D15
β-actin
Nodal
Oct3/4
b
WW93-A12-R4
WW103-18F11-R1
ES
EB D5
EB D6
EB D8
EB D9
EB D11
EB D13
EB D15
β-actin
PECAM
Tie2


c
WW93-A12-R4
WW103-18F11-R1
ES
EB D5
EB D6
EB D8
EB D9
EB D11
EB D13
EB D15
β-actin
BMP4
Brachyury
Goosecoid
d
WW93-A12-R4
WW103-18F11-R1
ES
EB D5
EB D6
EB D8
EB D9
EB D11
EB D13
EB D15
β-actin
AFP
HNF4
vHNF1

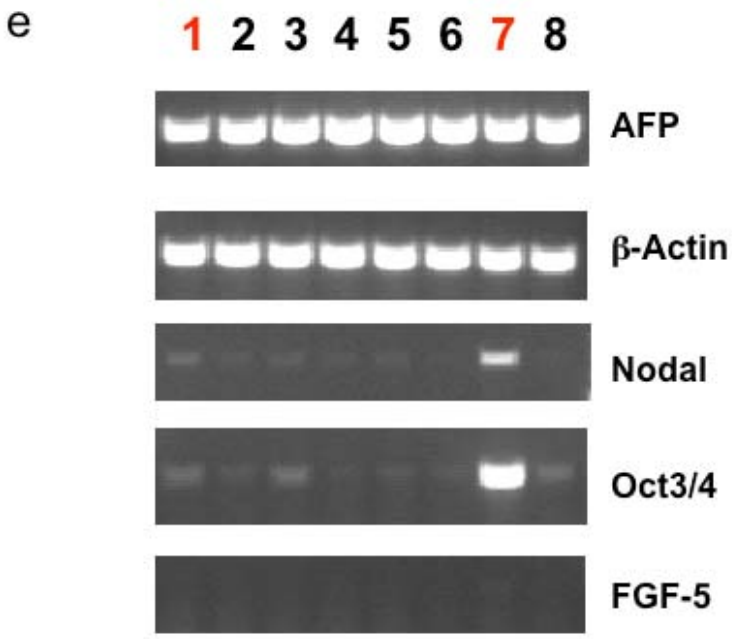
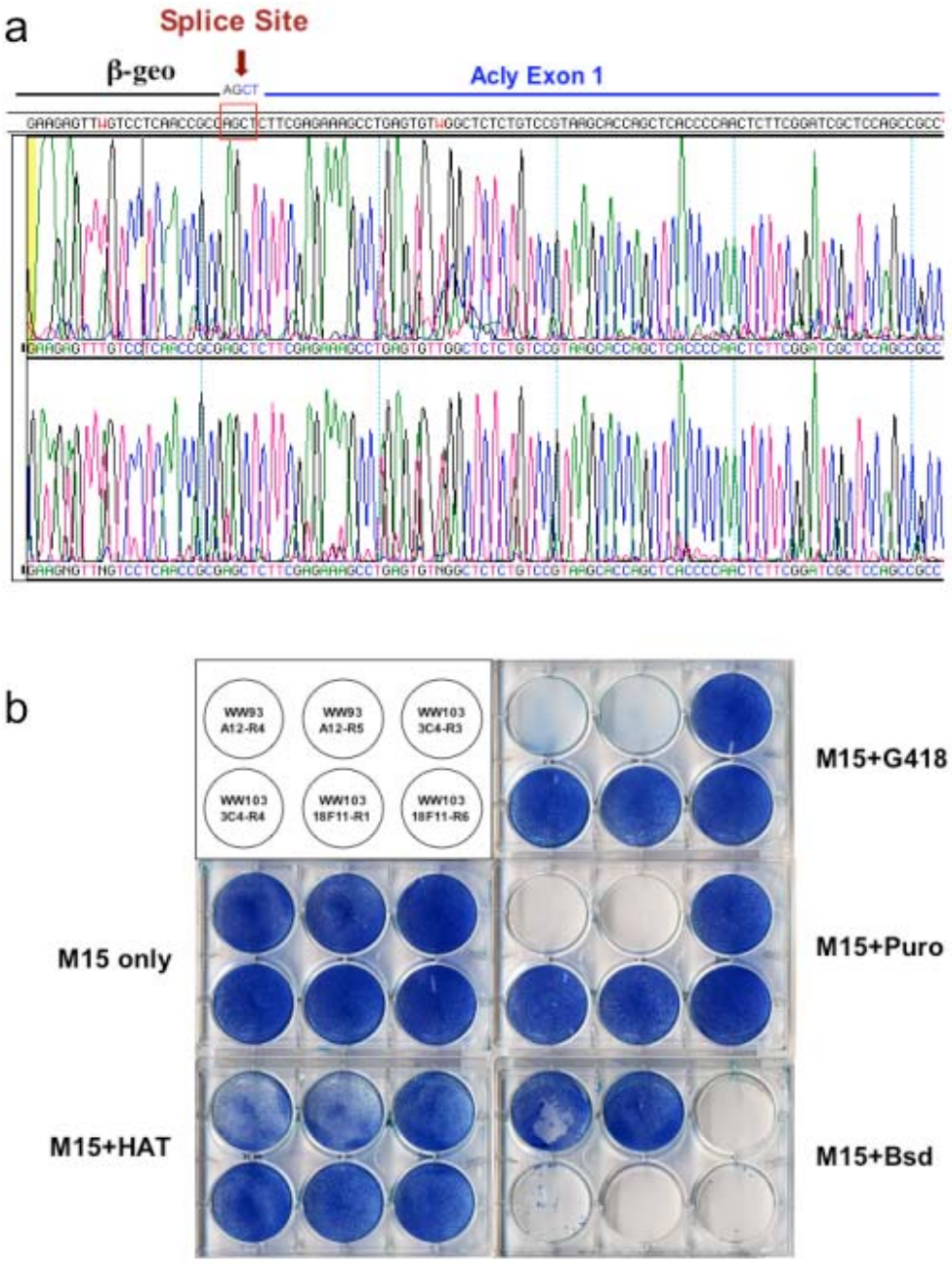
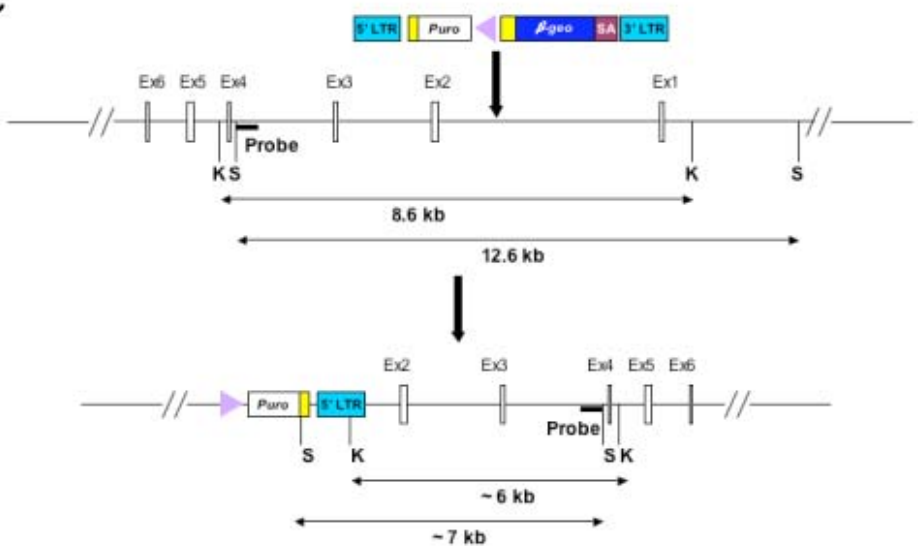


Fig. 5-10 RT-PCR results of WW103-18F11. a. Primitive ectoderm markers. The expression of *Oct3/4* and *Nodal* decreased slightly in the differentiation process. **b.** Endothelial markers. *Tie2* expression was not detected throughout the differentiation process of WW103-18F11 cells. The expression of *Pecam* was not up-regulated as observed in WW93-A12 control cell line. **c.** Early mesoderm markers. Low levels of expression of *Brachyury* and *Goosecoid* were still detected in WW103-18F11 derived EBs collected at later stages of the differentiation process, while no expression of these markers were detected in later stage EBs derived from WW93A12. **d.** Endoderm markers. The expression of most endoderm markers appears quite similar between the two cell lines. However, the expression of *Hnf4* in WW103-18F11 was much lower than that in the control. **e.** Day 18 embryoid bodies RT-PCR results. Note that WW103-18F11 EBs still express high amount of *Nodal* and *Oct3/4* at day 18. Lane 1, WW93-A12 control cell line; lane 7, WW103-18F11. Lanes 2-6 and 8 are day 18 EBs derived from other irrelevant mutant cell lines.


a
Splice Site
β-geo
AGCT
Acly Exon 1
b
WW93 A12-R4
WW93 A12-R5
WW103 3C4-R3
WW103 3C4-R4
WW103 18F11-R1
WW103 18F11-R6
M15+G418
M15 only
M15+Puro
M15+HAT
M15+Bsd

c


5' LTR
Puro
βgeo
SA
3' LTR
Ex6
Ex5
Ex4
Ex3
Ex2
Ex1
Probe
K S
K
S
8.6 kb
12.6 kb
Puro
5' LTR
Ex2
Ex3
Ex4
Ex5
Ex6
Probe
S
K
S K
~ 6 kb
~ 7 kb

d

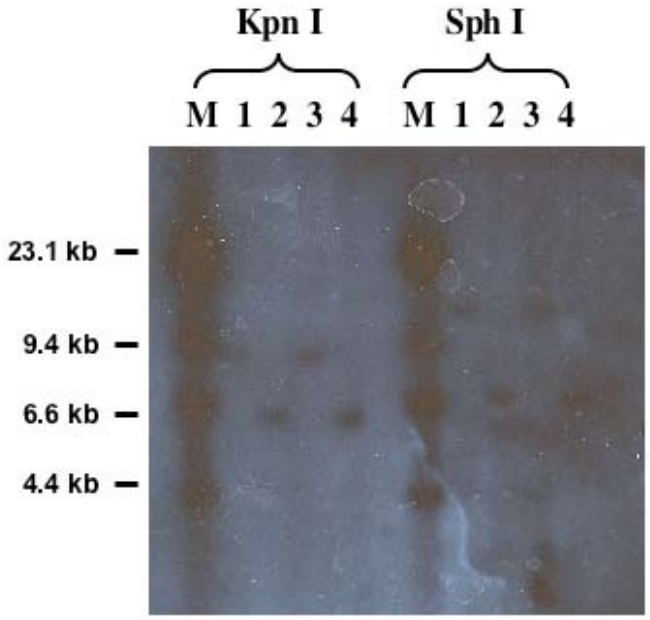

Kpn I
Sph I
M 1 2 3 4
M 1 2 3 4
23.1 kb
9.4 kb
6.6 kb
4.4 kb

Fig. 5-11 Identification of WW103-18F11. a. 5' RACE result. The sequence of the 5' RACE product matched the first exon of *Acly* gene. **b.** Sib-selection of WW103-18F11 and WW93-A12. Sib-selection was carried out on two of the WW103-18F11 subclones, WW103-18F11-R1 and WW103-18F11-R6, as well as two subclones of control cell line, WW93-A12-R4 and WW93-A12-R5. The same number of ES cells from two subclones each of WW93-A12 and WW103-18F11 were plated onto gelatinized 6-well plates and selected with M15, M15+Puromycin, M15+Blasticidine, M15+G418 and M15+HAT, respectively. The two WW103-18F11 subclones are $Puro^R$, Neo^R, Bsd^S and HAT^R, and the two WW93-A12 subclones are $Puro^S$, Neo^S, Bsd^R and HAT^R. The drug resistance pattern of WW103-18F11 suggested that it is a homozygous inversion clone. **c.** Schematic illustration of the structure of the proviral insertion and the subsequent inversion in WW103-18F11. **d.** Southern analysis of WW103-18F11. Southern analysis has been carried out using an *Acly* gene specific probe. This probe can detect an 8.6 kb *Kpn*I restriction fragment for the wild type allele. For WW103-18F11, the probe only detected an 6 kb *Kpn*I restriction fragment for the gene trap allele. Also, the probe only detected an 7 kb *Sph*I restriction fragment for the gene trap allele, instead of the 12.6 kb *Sph*I restriction fragment for the wild type allele. This Southern result has confirmed that both *Acly* alleles have been disrupted by the gene trap insertion. M, λ *Hin*dIII marker (New England Biolabs); lane 1, WW93-A12-R4; lane 2, WW103-18F11-R1; lane 3, WW93-A12-R5; lane 4, WW103-18F11-R6.

Southern analysis was carried out using an *Acly* gene specific probe. When this probe was hybridized to *Kpn*I digested genomic DNA, it detects an 8.6 kb wild-type fragment and an approximately 7 kb gene-trap fragment. When this probe was hybridized to *Sph*I digested genomic DNA, it detects a 12.6 kb wild-type fragment and an approximately 8 kb gene-trap fragment. As expected, only the gene-trap fragment was detected in the WW103-18F11 subclones. This Southern result confirms that both alleles of the *Acly* gene have been disrupted by the gene-trap insertion.

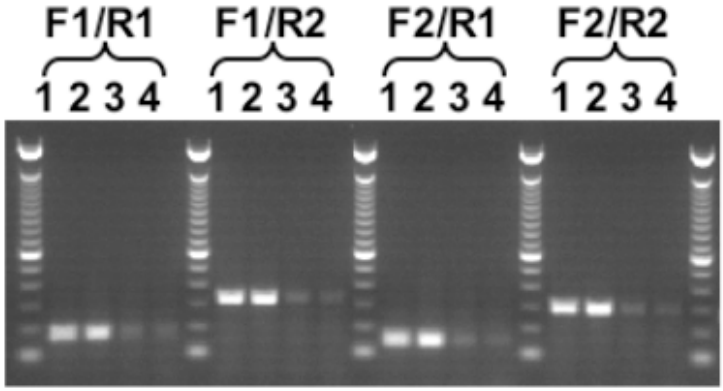
To see whether the gene-trap insertion and the subsequent inversion has disrupted transcription of the locus, PCR primers were used to specifically amplify cDNA fragments from Exon 1 to Exon 2 (F1/R1 and F2/R1) and Exon 1 to Exon 3 (F1/R2 and F2/R2). First strand cDNA was synthesised using total RNA extracted from WW103-18F11 and WW93-A12 ES cells. The RT-PCR results showed that transcription from Exon 1 to downstream exons was blocked. Weak PCR bands were detected for the *Acly*-deficient cell lines, which are likely to be contamination from feeder cells (Fig. 5-12a). Primer pairs F1/GSP4 and F2/GSP4 were used to specifically amplify the Exon 1/β-geo fusion transcript from the trapped allele. As expected, specific bands were only detected for WW103-18F11 subclones, but not for WW93-A12 control (Fig. 5-12c).

To see whether the gene-trap insertion and the subsequent inversion has affected the transcription of downstream exons, PCR primers were designed to specifically amplify cDNA fragments from Exon 24 to Exon 28 (F3/R3 and F3/R4) and Exon 25 to Exon 28 (F4/R3 and F4/R4). Specific PCR bands were detected in the WW103-18F11 mutant cell line and the WW93-A12 control cell line. Therefore it is likely that there is an alternative transcription start point between the retroviral insertion point and Exon 24, but the precise location of the mutant transcript start in WW103-18F11 is not known (Fig. 5-12b).

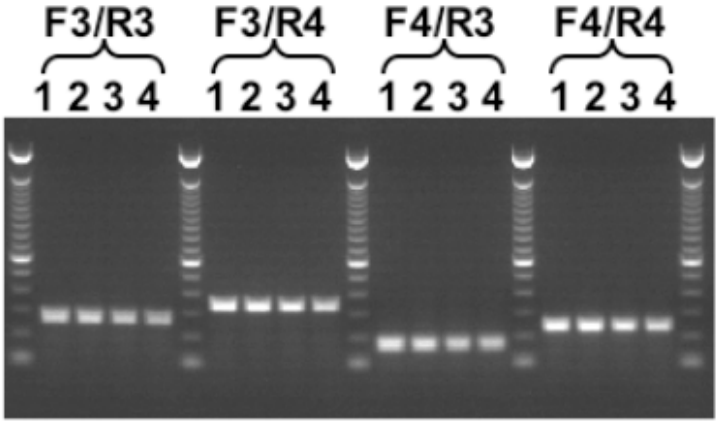
These RT-PCR primer pairs have also been used to check *Acly* expression during *in vitro* differentiation. In the WW93-A12 control cell line, *Acly* was highly expressed in undifferentiated ES cells, as well as throughout the whole differentiation process. In the WW103-18F11 mutant cell line, the F1/R1 and F2/R2 primer pairs did not detect the expression of the *Acly* upstream exons during *in vitro* differentiation, but the F3/R2 and F4/R4 primer pairs did detect expression of the *Acly* downstream exons (Fig. 5-12d).

To identify a causal link between the gene-trap insertion and inversion at the *Acly* locus and the severely impaired differentiation potential, a BAC rescue experiment was carried out to reverse the phenotype of the WW103-18F11 ES cell clone. A 129 S7 BAC clone, BMQ-290J5 was identified in Ensembl and confirmed to contain the complete *Acly* gene by PCR (Fig. 5-13a and data not shown). A *PGK-EM7-Bsd-bpA* cassette (pL313) was inserted into the *SacB* gene on the backbone of this BAC clone by *E. coli* recombination (Liu, Jenkins et al. 2003). The correct insertion of the *Bsd* cassette into the BAC backbone was confirmed by Southern using a *SacB* specific probe (Fig. 5-13b and c).

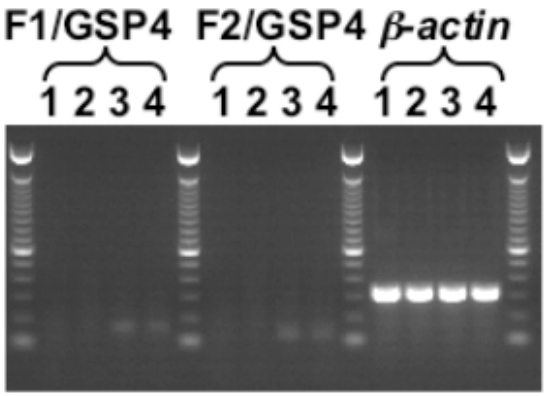
a


F1/R1
F1/R2
F2/R1
F2/R2
1 2 3 4
1 2 3 4
1 2 3 4
1 2 3 4

b


F3/R3
F3/R4
F4/R3
F4/R4
1 2 3 4
1 2 3 4
1 2 3 4
1 2 3 4

c


F1/GSP4
F2/GSP4
β-actin
1 2 3 4
1 2 3 4
1 2 3 4

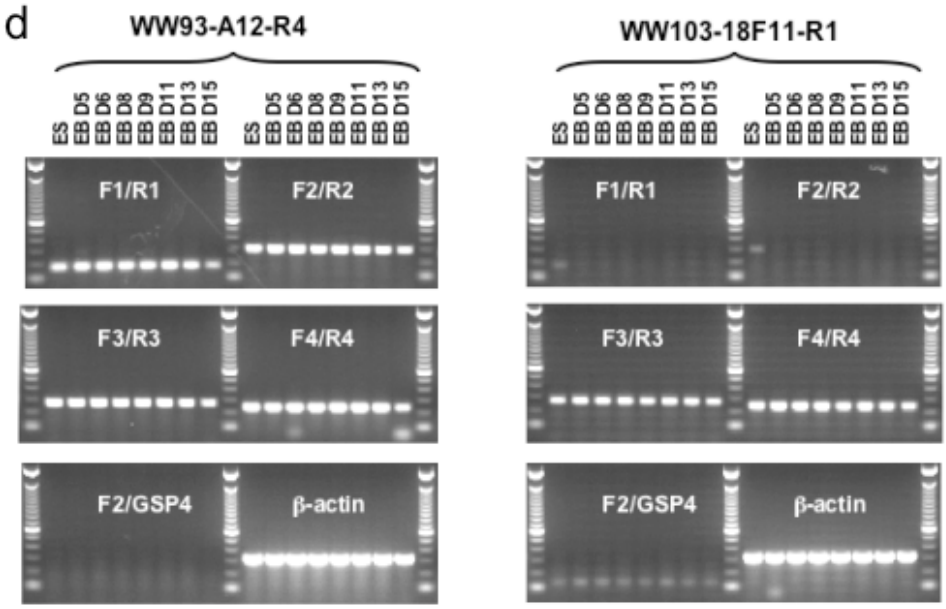


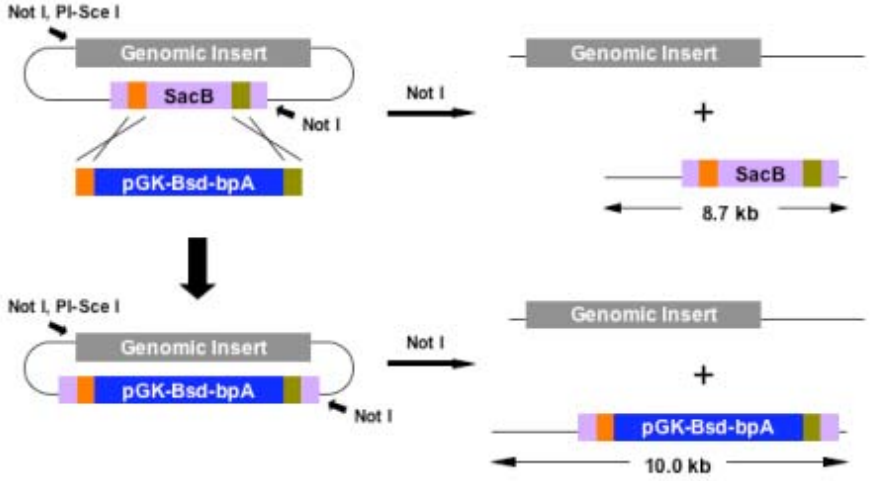
Fig. 5-12 Confirmation of WW103-18F11. a. RT-PCR to detect transcripts across the inversion breakpoint. PCR primers were designed to specifically amplify cDNA fragments from Exon 1 to Exon 2 (F1/R1 and F2/R1) and Exon 1 to Exon 3 (F1/R2 and F2/R2). RNA extracted from undifferentiated ES cells was used as template. lane 1, WW93-A12-R4; lane 2, WW93-A12-R5; lane 3, WW103-18F11-R1; lane 4, WW103-18F11-R6. RT-PCR results showed that *Acly* transcription across the inversion breakpoint was blocked. **b.** RT-PCR to detect the transcription of downstream exons. PCR primers were designed to specifically amplify cDNA fragments from Exon 24 to Exon 28 (F3/R3 and F3/R4) and Exon 25 to Exon 28 (F4/R3 and F4/R4). RT-PCR results showed that the transcription of downstream exons was not affected by the inversion. This suggests the existence of alternative transcript start points. **c.** RT-PCR to detect the Exon 1/β-geo fusion transcript. Primer pairs F1/GSP4 and F2/GSP4 were used to specifically amplify the gene trap fusion transcript. **d.** RT-PCR to detect *Acly* expression during ES cell *in vitro* differentiation. High *Acly* expression was detected in the whole process of differentiation of WW93-A12 control cell line, but not in WW103-18F11.

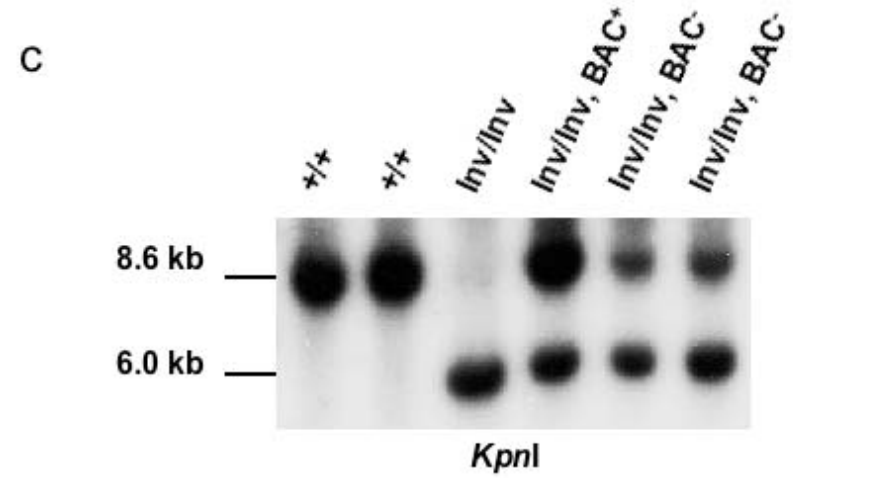
The modified BAC clone was linearized by I-*Sce*I and electroporated into WW103-18F11 ES cells (HAT^R, Neo^R, $Puro^R$, Bsd^S). 12 blasticidin resistant clones were picked and Southern analysis was carried out using an *Acly* gene specific probe to identify ES cell clones with a wild-type restriction fragment (Fig. 5-13c). One of the clones, WW113-2-8 has the wild-type restriction fragment and the ratio between the wild-type restriction fragment and the targeted restriction fragment is about 1:1, suggesting that this is likely to be a complemented clone which contains two wild-type copies of *Acly* gene. Another two clones, WW113-2-10 and WW113-2-11 also have the wild-type restriction fragment. But the ratio between the wild-type restriction fragment and the targeted restriction fragment is about 1:2, which suggests that both clones might contain a single copy of the BAC DNA, which can be randomly truncated and are likely to be incomplete. Western analysis was performed on whole-cell lysates extracted from undifferentiated wild-type control, *Acly*-deficient and BAC-rescued ES cells using a polyclonal rabbit anti-*Acly* antibody. *Acly* protein was not detected in the lysates from the WW103-18F11 cells. However, one of the BAC-rescued clones (WW113-8) expressed similar level of the *Acly* protein as the WW93-A12 wild-type control cells, indicating that this clone (WW113-2-8) is a rescued clone (Fig. 5-13d). The other two clones, WW113-2-10 and 2-11, which did not express *Acly* protein, are likely to only contain a truncated form of the BAC DNA and thus they were used as negative controls.

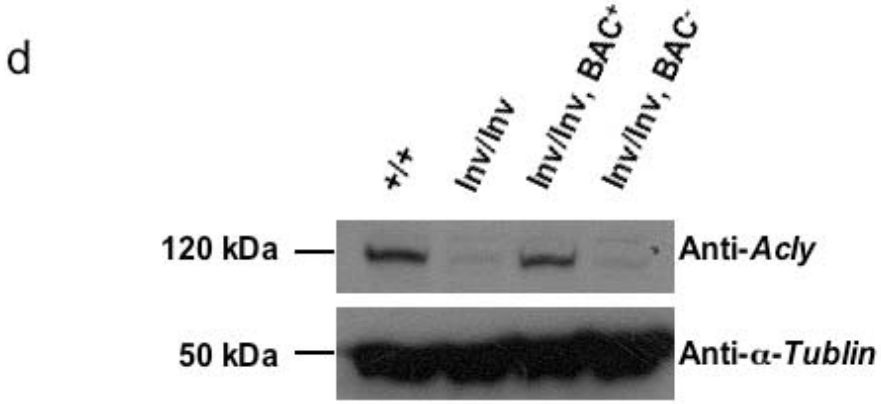
These three clones were expanded and induced to differentiate *in vitro*. After EBs were plated on the gelatinized tissue culture plates at Day 5, the EBs derived from WW113-2-8 ES cells could form cystic three-dimensional structures, while the EBs derived from WW113-2-10 and 2-11 ES cells could not (Fig. 5-13e).

a

b


Not I, PI-Sce I
Genomic Insert
SacB
Not I
pGK-Bsd-bpA
Not I
Genomic Insert
+
SacB
8.7 kb
Not I, PI-Sce I
Genomic Insert
pGK-Bsd-bpA
Not I
Not I
Genomic Insert
+
pGK-Bsd-bpA
10.0 kb


c
+/+
+/+
Inv/Inv
Inv/Inv, BAC+
Inv/Inv, BAC-
Inv/Inv, BAC-
8.6 kb
6.0 kb
KpnI


d
+/+
Inv/Inv
Inv/Inv, BAC+
Inv/Inv, BAC-
120 kDa
Anti-Acly
50 kDa
Anti-α-Tublin

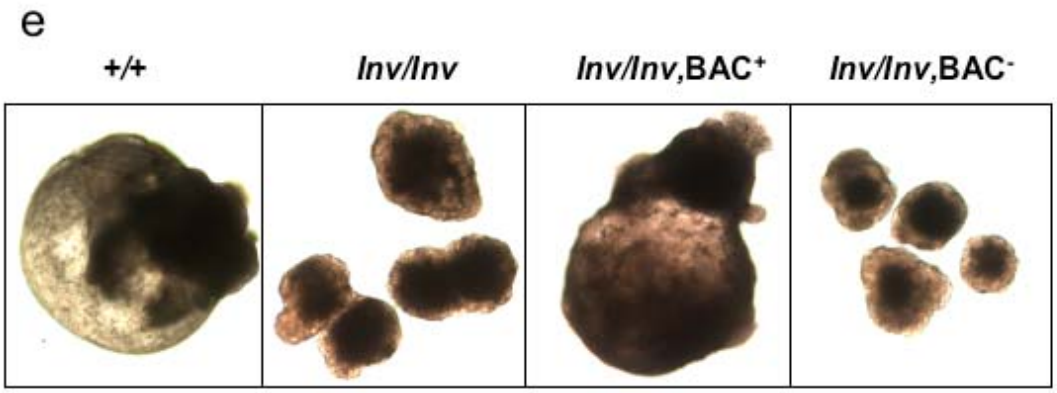

e
+/+
Inv/Inv
Inv/Inv,BAC+
Inv/Inv,BAC-

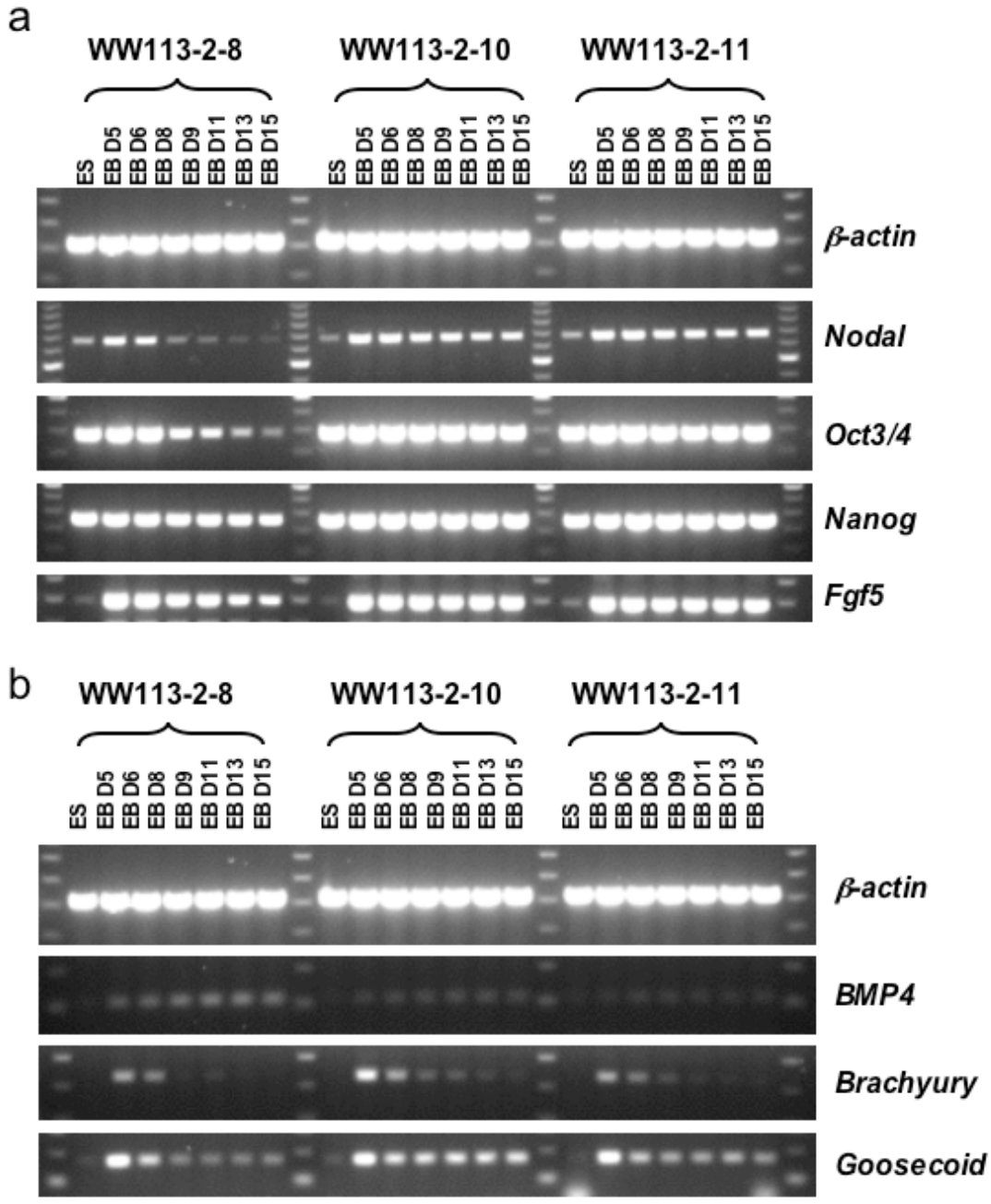
Fig. 5-13 BAC rescue. a. BAC clone BMQ-290J5. An 129 S7 BAC clone, BMQ-290J5 was picked according to the BAC end mapping information of this clone in the Ensembl database. This clone was confirmed to contain the complete *Acly* gene by PCR. BAC clone BMQ-290J5 is highlighted in blue. **b.** Insertion of a *PGK-Bsd-bpA* cassette into the BAC backbone. To facilitate the drug selection of BAC integration in ES cells, a *PGK-Bsd-bpA* cassette (pL313) was inserted into the *SacB* gene on the backbone of this BAC clone by *E. coli* recombination. **c.** Southern analysis of modified BAC clones. The correct insertion of the *Bsd* cassette into the BAC backbone was confirmed by Southern using a *SacB* specific probe. BAC DNA from the modified clones was digested with *Not*I. The *SacB* probe detected a 8.7 kb restriction restriction fragment for the unmodified BAC clone, and it detected a 10 kb fragment for the modified clones. **d.** Southern analysis of BAC inserted ES cell clones. Genomic DNA from 12 blasticidin resistant clones were digested with *Kpn*I, and Southern analysis was carried out using an *Acly* gene specific probe to identify ES cell clones with a wild type restriction fragment. WW113-2-8 (red) has the wild type restriction fragment and the ratio between the wild type restriction fragment and the targeted restriction fragment is about 1:1, suggesting that this is likely to be a complemented clone which contains two wild type copies of Acly gene. Another two clones, WW113-2-10 and WW113-2-11 (blue) have the wild type restriction fragment, but the ratio between the wild type restriction fragment and the targeted restriction fragment is about 1:2, which suggests that both clones might contain a single BAC clone which is likely to be incomplete. These two clones were used as control for the in vitro differentiation experiment.

When RT-PCR was performed on RNA extracted from WW113-2-8 embryoid bodies collected at different time points, the expression of *Oct3/4*, *Nodal* and *Nanog* were down-regulated rapidly as observed in WW93-A12 derived EBs. However, WW113-2-10 and 2-11 still expressed high level of these primitive ectoderm markers at late stages of their differentiation process (Fig. 5-14a).

The expression pattern of the early mesodermal markers (*Bmp4*, *Brachyury* and *Goosecoid*) also became normal in the EBs derived from WW113-2-8 ES cells. The expression of *Brachyury* and *Goosecoid* was down-regulated much quicker in the WW113-2-8 derived EBs than in the WW113-2-10 or 2-11 derived EBs. The expression pattern of these markers in WW113-2-8 derived EBs was similar to that observed in the control WW93-A12 derived EBs (Fig. 5-14b).

Acly gene-specific RT-PCR primer pairs have also been used to check *Acly* expression during *in vitro* differentiation of the BAC rescue cell line, WW113-2-8. In the WW113-2-8 cell line, significantly higher *Acly* expression than the other two control cell lines was observed in undifferentiated ES cells, as well as throughout the whole differentiation process. RT-PCR using the F1/GSP4 primer pair confirmed that gene-trap transcripts were still present in the rescued cell line, WW113-2-8. So the phenotype observed in the WW103-18F11 mutant cell line was caused by the loss of normal *Acly* transcription, instead of the dominant-negative effects, as the phenotypes could be reversed by re-introducing a wild-type copy of *Acly* gene (Fig. 5-14c).

All these data suggest a direct link between the reduction in *Acly* expression and the impaired *in vitro* differentiation potential observed in WW103-18F11 derived EBs.


a
WW113-2-8
WW113-2-10
WW113-2-11
ES
EB D5
EB D6
EB D8
EB D9
EB D11
EB D13
EB D15
β-actin
Nodal
Oct3/4
Nanog
Fgf5
b
WW113-2-8
WW113-2-10
WW113-2-11
β-actin
BMP4
Brachyury
Goosecoid

5.3 Discussion

In this chapter, I have described how I have used ES cell *in vitro* differentiation to screen a set of homozygous mutant ES clones. A panel of 16 markers was used to carry out the primary screen. To increase the throughput of the screen, I only took samples at three time points. If a homozygous mutant ES cell clone showed abnormal expression for one or more markers, the clone was subsequently tested in the second round screen. In the second round screen, more samples were taken at different time points, and additional markers were checked by RT-PCR to confirm the authenticity of the phenotype and also try to explain the phenotype at the molecular level by the gain or loss of specific differentiation markers.

5.3.1 Throughput of the screen

In this experiment, only a limited number of homozygous mutant ES clones were used for the *in vitro* differentiation screen. Therefore, it is possible to make a large number of EBs for each cell line and take samples at multiple time points. However, if the number of cell lines for screening increases to several hundred or several thousand, it would be necessary to make tens of thousands of plates of EBs. To make hundreds of thousands of “hanging drops”, transfer them to gelatinized tissue culture plates and change media regularly will be a labour-intensive work.

The RT-PCR method is not sensitive enough for the high-throughput analysis either. Approximately 20 μg of total RNA can be extracted from a plate of 40 EBs after Day 10. But for the early EBs (Day 5 to Day 10), sometimes two or three plates of EBs need to be combined together to get enough RNA. The cDNA synthesized from 5 μg of total RNA is only enough for about 20 RT-PCR reactions. If 10 cell lines are checked at the same time, 8 time points are taken for each cell line, and 16 markers are screened, this will require 1,280 PCR reactions. Any clones that do not show an obvious abnormality in these 16 markers will be discarded which is not a thorough analysis of the differentiation potential. Also, RT-PCR is a semi-quantitative approach to

assess gene expression, which makes it difficult to detect minor changes in expression.

An alternative approach to RT-PCR is to use cDNA and oligonucleotide microarray technology, which has been well characterized and proven to be a powerful tool for large-scale screens. This technology enables one to check the expression of all the genes in the mouse genome simultaneously. The development of array technology has made it possible to use very small amounts of starting RNA template. However, the downside of this technology is that it is still very expensive and the high cost makes it impractical to screen a lot of samples. Another problem of using microarray analysis to study ES cell *in vitro* differentiation is the complexity of the input material. It would be necessary to perform many control experiments to define the normal ranges of expression levels during differentiation, before comparisons can be made with samples from the mutant lines. Fluorescent reporters and FACS can also be used to screen the mutants in a high-throughput manner. It will be further discussed in the final chapter.

Considerable data has accumulated on the expression pattern of various markers characterizing the development of the three germ layers and other differentiated cell types during the ES cell *in vitro* differentiation process. However, this data is scattered throughout the literature and is far from being systematic or comprehensive. The results in these publications were generated by various methods, including RT-PCR, Northern, *in situ* hybridization or immunohistochemistry. Different ES cell lines (feeder-free or feeder-dependent), different differentiation protocols, and different lengths of observation periods make the data generated from these different experiments difficult to compare.

So before ES cell *in vitro* differentiation is used for a large-scale *in vitro* recessive screen, a systematic, quantitative study should be performed to determine the expression pattern of important developmental and differentiation markers in the differentiation process of the wildly used ES cell lines (AB2.2, D3, R1 and E14.1, etc.). Ideally, this data needs to be compared

to the expression pattern of these markers *in vivo* to link the *in vitro* differentiation with its *in vivo* counterpart.

5.3.2 Alternative recombination

Interestingly, two clones that have shown an abnormality during *in vitro* differentiation both contain either an extra chromosome 11 (WW103-8E6) or two partial duplication chromosome 11s (WW103-13D10). As discussed in the previous chapter, some clones can undergo a G2 *trans* recombination event and the resulting duplication chromosome can become homozygous by induced mitotic recombination. These homozygous duplication clones have as many as four copies of all the genes in the chromosomal region between the gene-trap locus and the end point targeting locus (*E_2DH*, 100.7 Mb). For WW103-13D10 (*LOC217071*, 88.7Mb), the duplication region is 12 Mb. It is reasonable to expect that such a big chromosomal rearrangement will cause an abnormality in differentiation. The WW103-8E6 clone has accumulated an extra chromosome before regional trapping. The subsequent mitotic recombination has duplicated the inversion chromosome, but a wild-type chromosome with the end point targeting cassette is still present. So the phenotypes of these clones with alternative recombination events are not related to the gene-trap loci, and are caused by the duplication of a part of or the whole chromosome.

5.3.3 *Acly* deficiency and the impaired differentiation potential

Acly is an important enzyme involved in fatty acid biosynthesis. Its product, acetyl-CoA, is the key building block for *de novo* lipogenesis (Beigneux, Kosinski et al. 2004). There are at least three principal sources of acetyl-CoA: 1) amino acid degradation produces cytosolic acetyl-CoA, 2) fatty acid oxidation produces mitochondrial acetyl-CoA, 3) Glycolysis produces pyruvate, which is converted to mitochondrial acetyl-CoA by pyruvate dehydrogenase (Garrett and Grisham 1999). The acetyl-CoA from amino acid degradation is not sufficient for fatty acid biosynthesis, and the acetyl-CoA produced by fatty acid oxidation and by pyruvate dehydrogenase can not cross the mitochondrial membrane. So cytosolic acetyl-CoA is mainly generated from citrate which is transported from the mitochondria to the

cytosol. ATP-citrate lyase converts the citrate to acetyl-CoA and oxaloacetate. Acetyl-CoA provides the substrate for cytosolic fatty acid synthesis, while the oxaloacetate is converted to malate which is transported back into the mitochondria where it can be converted back into citrate (Fig. 5-15).

5.3.3.1 *Acly* deficiency in the mouse

To investigate the phenotype of *Acly* deficiency in the mouse, an *Acly* knockout has been examined. This mouse line was generated from the Bay Genomics gene-trap resource (Stryke, Kawamoto et al. 2003). In this clone, a β-galactosidase marker is expressed from *Acly* regulatory sequences. Beigneux *et al*. (2004) have found that *Acly* is required for embryonic development, because no viable homozygous embryos were identified after 8.5 dpc. The early embryonic lethality suggested that the alternative pathways to produce acetyl-CoA in the cytosol are not sufficient to support development in the absence of *Acly* during development (Beigneux, Kosinski et al. 2004).

Northern and Western analysis of *Acly* mRNA and protein showed that in all the tissues examined (liver, heart, kidney, brain, and white adipose tissue), heterozygous mice expressed half of the normal amount of *Acly* mRNA and protein. But the heterozygous mice were healthy, fertile, and normolipidemic on both normal and high fat diets. The expression of another acetyl-CoA enzyme, Acetyl-CoA synthetase 1, was not up-regulated. Thus it seems that *Acly* is synthesized in adequate quantities and half-normal amount of the enzyme is enough for providing sufficient acetyl-CoA (Beigneux, Kosinski et al. 2004).

One interesting finding is that *Acly* is expressed at high levels in the neural tube at 8.5 dpc. The fact that *Acly* is not expressed in other foetal tissues suggests that *Acly* might not function as a house-keeping gene during development. Otherwise, widespread expression of *Acly* will be detected in all the cell lineages. Instead, it might have a tissue-specific function in embryogenesis, apart from producing Acetyl-CoA for lipogenesis.

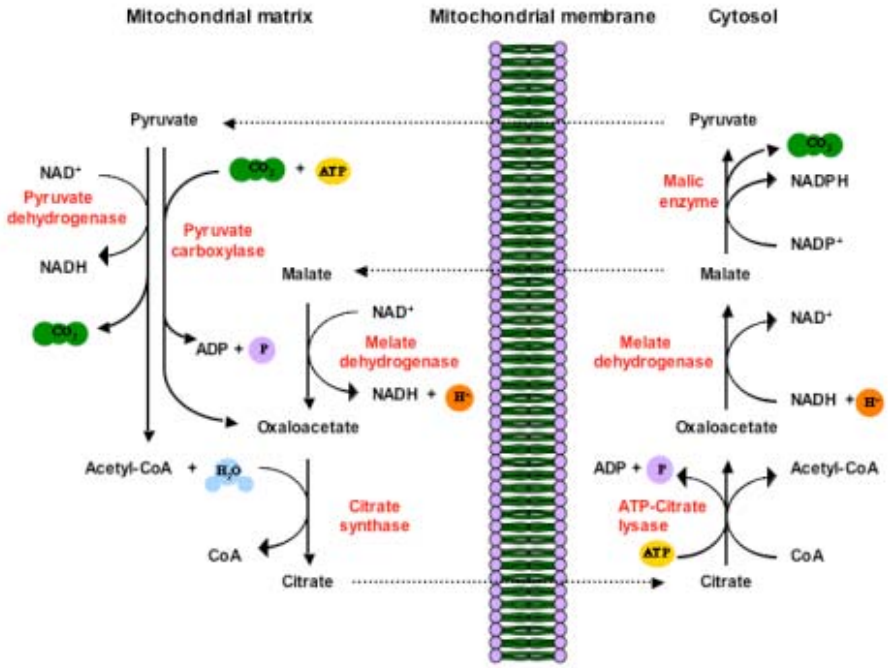


Fig. 5-15 The function of ATP-citrate lyase gene. The only known function of *Acly in vivo* is to generate acetyl-CoA by the ATP-driven conversion of citrate and CoA into oxaloacetate and acetyl-CoA. This is the first step for the *de novo* biosynthesis of sterol and fatty acid.

5.3.3.2 *Acly* and cell differentiation during sexual development

Acly has been shown to be involved in the sexual development of the fungus *Sordaria macrospora*. The fruiting body formation of filamentous ascomycetes involves the formation of the outer structures, as well as the development of mature ascospores within the fruiting body itself. Since this process requires the differentiation of several specialized tissues and some dramatic morphological and physiological changes, fruiting body maturation has been used as a model system to study multicellular development in eukaryotes (Nowrousian, Masloff et al. 1999).

Norwrousian *et al*. (1999) has used UV mutagenesis to screen for mutants with defects in fruiting body formation. One of the sterile mutants, per5, showed normal vegetative growth. But the fruiting body neck of the mutant strain was much shorter than that of the wild-type control. Most importantly, the fruiting bodies of the mutant strain only contain immature asci with no ascospores. DAPI staining showed that the immature asci still have eight nuclei within them, which suggests that there is no impairment in karyogamy or meiotic and postmeiotic divisions (Nowrousian, Masloff et al. 1999).

An indexed cosmid library was used to rescue the phenotype. A single complementing cosmid was isolated and sequence analysis has identified an ORF which has significant homology with higher eukaryotic *Acly*s. Analysis of the mutant *Acly* gene has identified a single nucleotide exchange (T to A), which altered a codon for aspartic acid into one for glutamic acid. The cloned mutant *Acly* gene can not rescue the phenotype of the mutant strain. Therefore, the mutation in *Acly* gene is responsible for the sterile phenotype (Nowrousian, Masloff et al. 1999).

As the mutant strain showed normal vegetative growth, it seems that sufficient acetyl-CoA is still produced for lipogenesis, either by residual *Acly* activity and/or expression of other acetyl-CoA-producing enzymes. But the attenuated *Acly* production can not satisfy the demand of acetyl-CoA during sexual development. So, the house-keeping functions of *Acly* can be circumvented to

a certain degree, but are essential under specific physiological conditions, such as sexual differentiation (Nowrousian, Masloff et al. 1999).

5.3.3.3 *Acly* as an *Brachyury* downstream notochord gene

Acly appears to be a downstream target of *Brachyury* in *Ciona intestinalis*. It is expressed specifically in the notochord in the embryogenesis process in *Ciona intestinalis*. The notochord has two major functions during chordate embryogenesis, providing inductive signals for the patterning of the neural tube and paraxial mesoderm and supporting the larval tail. The *Brachyury* gene encodes a transcription factor which contains a T DNA-binding domain. In vertebrates, *Brachyury* is first expressed in the presumptive mesoderm, and its expression is gradually restricted to the developing notochord and tailbud. *Brachyury* is believed to be one of the determinants for posterior mesoderm formation and notochord differentiation (Hotta, Takahashi et al. 2000).

By expressing the *Ciona intestinalis Brachyury* gene, *Ci-Bra*, in endoderm cells, Hotta *et al*. (1999) have isolated cDNA clones for 501 independent genes that were activated by *Ci-Bra* mis- and/or overexpression. By *in situ* hybridization, nearly 40 genes were found to be specifically or predominantly expressed in notochord, and therefore suggested to be *Brachyury*-downstream genes involved in notochord formation and function (Hotta, Takahashi et al. 1999). One of these genes, *Ci-Acl*, was found to share a high degree of homology with human ATP-citrate lyase (*ACLY*). The expression of this gene was first detected at the neural plate stage by *in situ* hybridization and its expression is restricted to notochord cells (Hotta, Takahashi et al. 2000). The fact that *Ci-Acl* only begins to express at the neural plate stage suggests that this gene might not be the immediate or direct target of *Ci-Bra*. Instead, it might be regulated by transcription factors, which in turn are regulated by *Ci-Bra* (Hotta, Takahashi et al. 2000).

It is interesting to notice that during embryogenesis of *Ciona intestinalis*, the expression of *Acly* is also highly restricted, similar to its expression pattern in murine embryogenesis. Considering the function of notochord in the

patterning of neural tube and paraxial mesoderm, it is likely that in both organisms, *Acly* plays some roles in neural tube and mesoderm differentiation.

5.3.3.4 Radicicol binds and inhibits mammalian *Acly*

Radicicol was first isolated from *Monosporium bonorden* as an antifungal antibiotic. But recently, this chemical was found to be able to reverse the transformed phenotype in *src*, *ras*, *mos*, *raf*, *fos*, and SV40-transformed cell lines. It can also cause cell cycle arrest and inhibit *in vivo* angiogenesis. So radicicol and its derivative are considered to be potential anti-cancer drugs (Ki, Ishigami et al. 2000).

To identify the *in vivo* target molecule of radicicol, Ki *et al*. (2000) used an affinity matrix to isolate radicicol-binding protein. Radicicol was biotinylated at various positions, and these variant compounds were then tested for their activity of morphological reversion of *src*-transformed phenotype. Two of the compounds, BR-1 and BR-6 were found to retain the activity. BR-6 was found to bind a 90-kDa protein, which was identified to be *Hsp90* by immunoblotting. BR-1 was shown to bind another 120-kDa protein, whose internal amino acid sequence was identical to human and rat ATP-citrate lyase. The identity of this 120-kDa protein was then confirmed by immunoblotting. Kinetic analysis showed that the activity of rat ATP-citrate lyase was inhibited by radicicol and BR-1, but not by BR-6. Radicicol was also found to be a non-competitive inhibitor of ATP-citrate lyase (Ki, Ishigami et al. 2000).

The fact that two radicicol derivatives, BR-1 and BR-6, bind two different proteins *in vivo* suggests that radicicol can bind different targets through different portions of its molecular structure. But the K_i value for ATP-citrate lyase was higher than the effective concentration of radicicol to reverse the transformed phenotype in *src*-tranformed cells, which suggests that this enzyme might not be directly involved in this process (Ki, Ishigami et al. 2000). Ki *et al*. (2000) hypothesized that BR-1 might be not stable and may be cleaved *in vivo* to generate free radicicol (Ki, Ishigami et al. 2000). But it could also be possible that radicicol is modified or cleaved *in vivo* to generate more

potent molecules to inhibit the enzyme activity of ATP-citrate lyase. Thus the phenotypes of radicicol, especially the ability to reverse the transformed phenotype in the cancer cell lines, might be partially associated with its binding and subsequent inhibition of *Acly* protein.

5.3.3.5 *Acly* is an important component of cell growth and transformation

Stable knockdown of *Acly* leads to impaired glucose-dependent lipid synthesis and also impaired *Akt*-mediated tumorigenesis (Bauer, Hatzivassiliou et al. 2005). Mammalian cells can not autonomously utilize the environmental nutrients to sustain their growth. Instead, constant extracellular signalling is needed to regulate the cellular metabolism of nutrients. However, cancer cells gain the autonomous ability to utilize nutrients by constitutively activating the normal signalling pathway without extracellular signals.

PI3K/*Akt* signalling pathway is critical for the cytokine-stimulated glucose metabolism, and its constitutive activation is commonly observed in cancer cells. In mammalian cells, glucose can either be oxidized to generate bioenergy, or be converted into other macromolecules to support biosynthesis. PI3K/*Akt* pathway can regulate the conversion of glucose to lipid and thus is essential for channeling the glucose into biosynthesis pathways. *Acly* is the main enzyme for producing cytosolic Acetyl-CoA for lipogenesis, and it is phosphorylated by *Akt in vivo* (Berwick, Hers et al. 2002). So it is possible that *Akt*-dependent cell transformation depends on *Acly* for *de novo* lipogenesis.

Bauer et al. (2005) used a shRNA construct to stably knock down the expression of *Acly* in a *Akt*-transformed cell line, FL5.12. *Akt*-expressing cells with or without *Acly* knockdown were injected into nude mice intravenously, and the mice were monitored for Akt-dependent leukemogenesis. Mice administrated *Acly* knockdown cells exhibited a significant delay or even a complete resistance to leukemogenesis.

5.3.3.6 A possible explanation of the phenotype of *Acly* deficient ES cells

Our *in vitro* differentiation results and works published before (Hotta, Takahashi et al. 1999; Nowrousian, Masloff et al. 1999; Hotta, Takahashi et al. 2000; Ki, Ishigami et al. 2000; Bauer, Hatzivassiliou et al. 2005) all suggested a pivotal function of *Acly* in cell differentiation and transformation. The only known function of *Acly in vivo* is to generate acetyl-CoA by the ATP-driven conversion of citrate and CoA into oxaloacetate and acetyl-CoA. This serves as the first step for the *de novo* biosynthesis of sterol and fatty acid. So the housekeeping function of the gene should be important for cell survival. But our observation and other published works (Nowrousian, Masloff et al. 1999; Beigneux, Kosinski et al. 2004) suggested that the house-keeping function of this gene can be circumvented to some degree either by the residual *Acly* activity or other alternative acetyl-CoA producing pathways.

In this study, there is no apparent difference in the growth rate, colony formation ability or ES cell/colony morphology in *Acly*-deficient ES cells compared to the wild-type control (data not shown). Microarray analysis using RNA extracted from undifferentiated WW103-18F11 and WW93-A12 ES cells showed that the expression levels of most mouse genes are similar (this work is still ongoing), which suggests that *Acly* deficiency does not cause observable phenotype in ES cells and the acetyl-CoA production in *Acly*-deficient ES cells seems to be sufficient to sustain the normal growth and division of ES cells.

However, when the *Acly*-deficient ES cells were differentiated *in vitro*, they could not form the typical three-dimensional cystic structures. In addition, the expression of some germ layer and cell type specific markers had changed. The RT-PCR results suggested that most cells in the cell aggregates were still undifferentiated ES cells. It is possible that the transition from the normal ES cell growth/division to the drastic re-programming and cell fate determination in the differentiation process demands higher than normal amounts of acetyl-CoA. A similar situation might accompany the transition from vegetative to sexual development in *S. macrospora* (Nowrousian, Masloff et al. 1999). It is

possible that this energetic demand can not be fulfilled by the alternative metabolic pathways, which might be partly due to different metabolic costs of lipogenesis.

Another possible explanation is that in order for some ES cells in an EB to be differentiated into a certain cell type, these cells must gain “competence” before the differentiation process is induced. The differentiation competence might involve as one component a threshold in acetyl-CoA concentration, which might be much higher than the level that is necessary for the ES cell growth and division.

The exact mechanism by which the acetyl-CoA production can influence the potential of ES cells to differentiate *in vitro* is unknown. Acetyl-CoA can be used to produce fatty acids, sterols and other important molecules which need the acetyl base, such as acetylcholine (Beigneux, Kosinski et al. 2004). Therefore, ATP-citrate lyase may either control the overall cytosolic acetyl-CoA concentration to indirectly regulate the pathways that need acetyl-CoA, or it could directly interact with various acetyltransferases or lipid/sterol synthetases to form an enzyme complex to provide acetyl-CoA. Nevertheless, *Acly* seems to play an important role in development and differentiation of certain cell types.

The difficulty to determine the primary locus of action of *Acly* make it hard to link this gene directly with any known genetic pathways controlling ES cell differentiation. It is not unexpected for mutants identified by such a genetic screen. However, if more homozygous ES cell mutants are generated in the future and screened using the same strategy, it will be possible to group the mutants by their apparent defects and study the relationships between the mutants with similar phenotypes. The importance of a genetic screen is that it can not only fill in the gaps in a known pathway, but also identify new pathways that are not necessarily overlapping with the known ones.

5.3.3.7 Future experiments to identify the function of *Acly* in ES cell *in vitro* differentiation

The RT-PCR results detected transcription of *Acly* downstream exons in the mutant line. Since *Acly* is a large gene (51.54 kb), it is possible that there are other alternative transcription start points. The proviral insertion and the subsequent inversion might not completely block all the *Acly* transcripts, so the mutation generated in the homozygous mutant cell line might not a null allele. To resolve this, a homozygous *Acly* gene targeted ES cell line can be constructed and these ES cells can be differentiated to confirm the function of the gene in ES cell *in vitro* differentiation.

To investigate the *in vivo* differentiation potential of the WW103-18F11 ES cell line, 1X 10^7 undifferentiated WW103-18F11 and control WW93A12 ES cells were injected subcutaneously into both flanks of 8-week old F1 hybrid mice (129 S7/SvEv$^{\text{Brd-}Hprtb\text{-}m2}$ X C57$^{\text{TyrBrdC1}}$ female). The animals were examined periodically over 4 weeks for the appearance and growth of tumours. 4 weeks after injection of ES cells, the mice were sacrificed, and the size of each tumor was measured after dissection. Tumor samples were cut into two halves, one half was fixed in 10% formalin for histopathological analysis, and the other half was dissected into several pieces (depending on its size) and snap-frozen in liquid nitrogen for subsequent RNA and DNA extraction.

For the WW93-A12 ES cell line, tumours were found at every ES cell injection site (8/8). Though the size of the tumours varied, all the tumours collected were dark red and highly vascular. When the tumours were bisected, a fluid filled central cavity was always found in the centre of the tumour. In contrast, for the WW103-18F11 ES cell line, only 3 tumours were found 4 weeks after the injection (3/8). All three tumours were very small and pale. No blood vessels were found on their surface. When the tumours were bisected, no fluid filled cavities were present.

Histopathology results of three tumours generated from WW103-18F11 cells and eight cases of WW93-A12 teratocarcinomas have confirmed that the differentiation potential of WW103-18F11 clones were greatly impaired

(pathology analysis was performed by Dr. Madhuri Warren). The WW103-18F11 tumours are circumscribed mixed ganglier/neuroepithelial tumours plus embryonal carcinoma composed predominantly of nests of mature glial cells with scarce neuroepithelial differentiation in the form of Homer-Wright rosettes. Nests of undifferentiated embryonal carcinoma (ES cells) are also seen. There was no evidence of differentiation into other germ cell lineages.

All the WW93-A12 tumour are circumscribed immature teratocarcinomas composed predominantly of immature glial tissue and tissues from all three germ layers: simple cuboidal epithelium, columnar epithelium, ciliated respiratory type epithelium, mucin secreting gastrointestinal epithelium; cartilage, osteoid, immature neuroepithelium, smooth muscle; and stratified squamous epithelium. In some samples, nests of immature embryonal carcinoma (undifferentiated ES cells) and isolated syncytiotrophoblast cells were also found.

We have also injected the BAC rescued ES cells into the F1 hybrid mice and are now waiting for the pathology results of the teratocarcinomas generated by subcutaneous injection. For the rescued cell line, tumours were found at every ES cell injection site (8/8). All of the tumours were dark red and highly vascular. Some of these tumours have a fluid filled central cavity in the centre of the tumour.

From the initial result, we can conclude that the differentiation potential of the *Acly*-deficient ES cells is also impaired *in vivo*. But complete pathology results of the tumours derived from the rescued cells are needed to confirm that the *in vivo* differentiation potential is fully recovered in these cells.

Because the phenotype of the *Acly*-deficient ES cells may depend on some mutations or silencing in a second gene, inactivation of *Acly* in an independent ES cell clone is necessary to prove that the *Acly* gene is solely responsible for the differentiation defect we observed in WW103-18F11 deficient cell line. Although BAC rescue experiment can make a causal link between the mutation in *Acly* gene and the defective phenotypes, it is

possible that other genes or transcriptional elements also play some roles in the differentiation defects. If over-expression of *Acly* cDNA can also rescue the defective phenotypes, it will effectively exclude the involvement of other genes or transcriptional elements.

5.3.4 Summary

In this chapter, I have described the strategy used to screen for an *in vitro* differentiation phenotype in homozygous mutant ES cell lines. Restricted by the detection method, I checked the expression of a limited number of markers in the differentiation process. In spite of this limitation, I successfully identified several clones with a reproducible *in vitro* phenotype. By Southern analysis and sib-selection using different drugs, I found some of these clones are the products of alternative recombination events. But two of the clones, WW103-14F11 and WW103-18F11, are products of regional trapping and subsequent inversion. Detailed expression analysis and functional studies have been carried out on WW103-18F11. The impaired *in vitro* differentiation potential observed in this clones was caused by the disruption of the ATP-Citrate lyase (*Acly*) gene. Therefore, this strategy has proved to be able to identify *in vitro* differentiation mutants and facilitate regional screens for genes involved in the early embryogenesis in the mouse genome.

6 Summary, significance and future goals

In the previous chapters, I have shown that localized gene-trap mutagenesis can be achieved by regional trapping and that the gene-trap mutations generated can be made homozygous by inducible mitotic recombination. A genetic screen has been carried out on the isolated homozygous mutant clones using an ES cell *in vitro* differentiation assay. Clones that show abnormal morphological and gene expression changes during the differentiation process were identified. Other experiments were carried out to confirm these findings. Therefore, I have demonstrated that I can use this strategy to generate homozygous mutant clones in a given region of the mouse genome and use these clones for an *in vitro* recessive genetic screen. In principle, this strategy can be applied to other chromosomes in the mouse genome to create genome-wide homozygous mutant ES cells. This will be a valuable resource for *in vitro* recessive genetic screens.

Before I discuss the potential application of this strategy, I would like to describe some of the latest advancements in mutagenesis techniques, because no single mutagenesis method can completely replace the other methods, and mouse genetics will depend on a combination of these methods as a whole.

6.1 Chemical mutagenesis

Regional and genome-wide ENU mutagenesis in the mouse is a powerful way to generate dominant and recessive mutations for phenotype-driven genetic screens. Such screens can provide a large amount of information about a phenotype of interest or even a certain genetic pathway in a relatively short period of time.

A recent development in this field is to generate ENU- or EMS-induced alleles in mouse ES cells (Chen, Yee et al. 2000; Munroe, Bergstrom et al. 2000). Conventional germ cell mutagenesis with ENU is compromised by the inability to easily determine the mutation rate, strain and interlocus variation in mutation induction, as well as the extensive mouse husbandry requirements (Munroe, Bergstrom et al. 2000). Genome-wide recessive mutations

transmitted by ENU treated males can only be rendered homozygous after three generations of breeding, at which time phenotype screens can be performed. Chen *et al*. (2000) and Munroe *et al*. (2000) have both used the mouse *Hprt* locus to determine that the mutation rate in ES cell is comparable to the mutation rate in spermatogonia in adult male mice. By using ENU mutated ES cells, one generation can be eliminated from the complicated breeding strategy. Also storing ES cells is more convenient than cryopreserving sperm.

ENU/EMS mutagenesis in ES cells can be used for two different purposes, to screen for an allelic series of mutations of a target gene *in vitro* (Vivian, Chen et al. 2002; Greber, Lehrach et al. 2005) or to perform genome-wide recessive genetic screens *in vivo* (Munroe, Ackerman et al. 2004). Vivian *et al*. (2002) has used an RT-PCR based high throughput mutation detection technology to identify mutations in *Smad2* and *Smad4*, which are both embryonic lethal when the genes are knocked out. Of the five non-silent mutations that were transmitted through the germline and bred to homozygosity, one was a severe hypomorph, one was a dominant-negative allele, and the other three did not show any phenotype (Vivian, Chen et al. 2002). Munroe *et al*. (2004) have demonstrated the feasibility of performing genome-wide mutation screens with only two generations of breeding. This strategy was possible because chimeras derived from a single EMS treated ES cell clone transmit variations of the same mutagenized diploid genome, whereas ENU-treated males transmit numerous unrelated genomes (Munroe, Ackerman et al. 2004).

ENU mutagenesis has also been used to generate bi-allelic mutations in ES cells deficient in the Bloom's syndrome gene (*Blm*) (Yusa, Horie et al. 2004). Yusa *et al*. (2004) used a combination of ENU mutagenesis and transient loss of *Blm* expression to generate an ES cell library with genome-wide homozygous mutations. This library was evaluated by screening for mutants in a known pathway, glycosylphosphatidylinositol (GPI)-anchor biosynthesis. Mutants in12 out of 23 known genes involved in this pathway have been obtained, and two unknown mutants were also isolated (Yusa, Horie et al. 2004). Though ENU mutagenesis is proved to be an efficient tool to generate

mutants in ES cells, it is still a difficult task to identify the mutated gene. In cases when little is known about the pathway, this can only be achieved by expression cloning.

6.2 Transposon mutagenesis

Retroviral and plasmid-based vectors are the two main approaches for insertional mutagenesis. Mutagenesis rates for these vectors are improved by ensuring that vector insertions coupled with actuation of a selectable marker, a concept known as a "gene trap". Different gene-trap vector designs are needed to achieve broad genome coverage in large-scale genetic screens. The synthetic *Sleeping Beauty* (SB) transposon system provides a promising alternative delivery method for gene-trap vectors (Ivics, Hackett et al. 1997).

Sleeping Beauty (SB) belongs to the *Tc1/mariner* superfamily of transposons. Ivics *et al*. (1997) reconstructed the transposon and transposase, *SB10*, from endogenous transposons inactivated by mutations accumulated in evolution. Both the reconstructed transposon and the transposase were shown to be active in mouse and human cell lines (Ivics, Hackett et al. 1997). It is composed of the SB transposon element and the separately expressed transposase. The SB transposon element contains two terminal inverted repeats (IR). The excision and re-insertion of the SB transposon element into the host genome occurs by a cut-and-paste process mediated by the transposase which binds to the terminal IRs. The insertion of the SB transposon itself could cause an insertional mutation if the expression of host gene is interrupted.

The SB system was first used as an insertional mutagen in mouse ES cells (Luo, Ivics et al. 1998). But in ES cells, the transposition efficiency is quite low (3.5×10^{-5} events/ cell per generation). Though there is still room to improve the efficiency of SB system *in vitro*, this system does not appear to be suitable for a genome-wide mutagenesis effort in ES cells. However, efficient transposition has been observed in the mouse germline, either by crossing males doubly transgenic for *SB10* transposase and a gene-trap transposon to wild-type females (Dupuy, Fritz et al. 2001), or by injecting transposon vectors

and SB10 mRNA together into one-cell mouse embryos (Dupuy, Clark et al. 2002). In these studies, on average, 1.5 to 2 transposon insertion were found in each of the offspring.

To determine sequence preferences and mutagenicity of SB-mediated transposition, Carlson *et al*. (2003) have cloned and analyzed 44 gene-trap transposon insertion sites from a panel of 30 mice. 19 of the 44 mapped transposon insertion sites were mapped to chromosome 9 where the transposon concatomer was located. The remaining insertion occurred on other chromosomes without obvious preference for chromosome or region. The local transposition interval appears to be between 5 to 15 Mb. Analysis of the transposon/host flanking sequence has shown that transposition sites are AT-rich and the favoured sequence is “ANNTANNT”. 27% transposon insertions were in transcription units. Of the 6 insertions in heterozygous animals which were bred in attempts to generate homozygous mice for the insertions, two were found to be homozygously lethal (Carlson, Dupuy et al. 2003). The transposition and gene insertion frequencies mean that *Sleeping Beauty* is still not efficient enough for a genome-wide mutagenesis screen.

The transposon and a transposase-expression vector can be electroporated into host cells where they co-exist episomally for a short period of time during which transposition is catalysed from the vector to the genome. Although this episomal method is very efficient in cultured somatic cells and in somatic cells *in vivo*, the transposition efficiency in mouse ES cells is very low (Luo, Ivics et al. 1998). Therefore it is not currently efficient enough for genome-wide mutagenesis in ES cells without a significant improvement of its efficiency in ES cells.

6.3 RNA interference

RNA interference (RNAi) was first noticed in *C.elegans* as a response to exogenous double strand RNA (dsRNA), which induce sequence specific knockdown of an endogenous gene’s function. Double strand RNA mediated gene inactivation is a highly conserved process. The basic mechanism of RNAi includes three major steps: first, a double strand RNA is cleaved by

Dicer protein into 21-25 nucleotides (nt) double strand RNAs; second, these small interfering RNAs (siRNA) associate with a complex (RISC, RNA-induced silencing complex) which has RNA nuclease activity; third, RISC unwinds siRNA and uses it as the template to capture and destroy endogenous transcript (Hannon 2002).

The RNAi phenomenon was quickly adopted for large-scale genome-wide genetic screens in *C. elegan*s. In *C. elegans*, this form of post-transcriptional gene silencing (PTGS) only requires a few molecules of double strand RNA in one cell to initiate the process. It can spread to all the cells in the body of the worm and pass through the germ line for several generations with almost complete penetrance (Kamath, Fraser et al. 2003). The delivery of dsRNA in *C. elegans* is also very simple, it can be achieved either by soaking the worms in dsRNA solution or feeding the worm with dsRNA-expressing *E. coli*.

Naturally, the success of RNAi technology in *C. elegans* inspired many to apply it to more complex mammalian systems. However at the beginning, this technology has encountered some problems. First, dsRNA becomes diluted in subsequent cell divisions, and the silencing phenotype can not be inherited unless a dsRNA-expressing construct is stably integrated in the genome. Second, dsRNA triggers a non-specific global translation inhibition by activating the RNA-dependent protein kinase (PKR) pathway (Hannon 2002). A way to bypass this problem is to express short hairpin RNA (shRNA) in mammalian cells

Elbashir *et al*. (2001) showed that 21 or 22 nucleotides double strand RNA could strongly induce gene-specific inactivation without eliciting the non-specific translation inhibition effect observed with longer dsRNAs (Elbashir, Harborth et al. 2001). However, the shRNA mediated RNAi effect in mammalian cells is not inherited nor can it spread to adjacent cells. Brummelkamp *et al*. (2002) developed a mammalian expression vector to synthesize short hairpin-structured RNA transcripts (shRNA) *in vivo*. The shRNA can be recognized and cleaved by the endogenous PTGS machinery and can trigger the RNAi process. With these developments, shRNA

technology has become a practical tool to study gene function in mammalian cells.

Recently, two groups have reported the construction and initial application of shRNA expressing libraries targeting human and mouse genes (Berns, Hijmans et al. 2004; Paddison, Silva et al. 2004). Berns *et al*. (2004) constructed a library of 23,472 distinct shRNAs targeting 7,914 human genes. They obtained on average 70% inhibition of expression for approximately 70% of the genes in the library. A screen using this library has successfully identified one known and five unknown modulators of the p53-dependent proliferation arrest (Berns, Hijmans et al. 2004). Paddison *et al*. (2004) targeted 9,610 human genes and over 5,563 mouse genes in their library. One quarter of this library was used to screen for shRNAs that interfere with 26S proteasome function. Nearly half of the shRNA clones that were expected to target proteosomal proteins were recovered as positive in the screen (Paddison, Silva et al. 2004). These experiments have shown that RNAi has become a practical tool for recessive genetic screens in mammalian cells in culture.

RNAi technology still has some limitations. First, it can only knockdown the expression of a gene. Incomplete inhibition will cause a hypomorphic phenotype in many cases. If the residual expression of the target gene is still enough for its normal function, it will be missed in large-scale genetic screens. An example of this is illustrated by a systematic function analysis of the *C. elegans* genome using RNAi. Although this screen targeted about 86% of the 19,427 predicted genes, mutant phenotypes were only identified for 1,722 genes (Kamath, Fraser et al. 2003). Another example of this limitation is that just 22 out of 55 shRNAs targeting 26S proteasome components were identified as positive in the screen. Another 14 shRNAs scored above background in the second focused assay in the same study (Paddison, Silva et al. 2004). Second, the design of an shRNA-expressing construct requires prior knowledge of its target, which is greatly limited by the annotation of the mouse genome. That means a genetic screen using this technology is always going to be a forward genetics screen. Any genes not in the library will never

be identified in the screen. So although shRNA screens are potentially powerful, they lack the coverage of a screen performed with a random mutagen like ENU.

6.4 Forward genetics versus reverse genetics

Forward genetics refers to the techniques used to identify mutations that produce a certain phenotype. A mutagen is often used to accelerate this process. Once mutants have been isolated, the mutated gene can be molecularly identified. Reverse genetics refers to the method to determine the phenotype that results from mutating a given gene, usually by deleting the gene of interest.

Historically, forward genetic screens have been the main method for gene function discovery in various model organisms. But in the mouse, the development of mouse gene knockout technology has made reverse genetics the most powerful and widely used functional genomics tool. The distinction between these two approaches is no longer so clear. For example, gene-trap insertional mutagenesis is a typical forward genetics approach that has been widely used in *in vitro* and *in vivo* forward genetic screens. But the development of 5' RACE technology has made the identification of the insertion site much easier than before, so a large number of mutant clones can be generated and identified in a high-throughput way (Skarnes, von Melchner et al. 2004), and reverse genetic screens can be carried out on these ES cell clones or the mice derived from them.

The completion of the mouse and human genome has provided an unprecented opportunity for both forward and reverse genetics studies. For forward genetics, it is now much easier to map and identify the causative genetic change. For reverse genetics, the availability of the sequence information for each mouse gene has made it possible to knockout any gene in the mouse genome by gene-targeting or it can be knocked down by RNAi.

Though reverse genetics is more straightforward, and the phenotype can be quickly linked to the mutation, forward genetics has its own advantages. First,

it is quick to generate a lot of mutations for phenotype analysis. Second, it is an unbiased, phenotype-driven approach and no previous knowledge of the pathway involved is needed. It is not surprising that even a screen for a well-characterized pathway can still identify unknown components. Third, a variety of allelic mutations can be generated and they might affect a gene's function in different ways. So forward genetics will play an increasingly important role in mouse functional genomics.

6.5 Selection versus screening

Most of the genetic screens performed in mammalian cells are in fact selections. The distinction between a selection and a screen depends on the method used to detect the phenotype of the mutants. A selection requires a strategy to distinguish those mutant cells that show a given phenotype from the rest of the cell population. This can be achieved by two ways, either by accumulating the cells that carry the desired mutations, or more often, by selectively killing the rest of the cells that do not carry the relevant mutations (Grimm 2004).

On the other hand, in a screen, mutants must be examined one by one to determine whether and to what extent they have the desired phenotype. So for a selection or a screen conducted on the same scale, a screen will require much more time and labour. Geneticists always prefer to perform a selection whenever it is possible. But screens are particularly useful when a broad dynamic range of gene activity is examined (Shuman and Silhavy 2003), for example the mutations that affect ES cell *in vitro* differentiation in our study.

The development of FACS technology has made it possible to turn a screen into a selection by selectively accumulating the mutants that show a certain phenotype. For example, if we want to carry out a screen on ES cell differentiation into mesodermal lineages, mutant ES cells can first be differentiated on collagen IV coated dishes, and *Flk1*$^{+}$ cells derived from embryonic stem cells can then be sorted by FACS (Yamashita, Itoh et al. 2000), while the undifferentiated mutant ES cells can be sorted by ES cell specific markers, such as SSEA-1. If a cell lineage-specific cell surface

marker is not available, a fluorescence reporter can be used to tag an intra-cellular lineage-specific gene. Examples for this strategy is the use of *Sox1*-GFP knock-in to track the differentiation of ES cells into neuroectodermal precursors (Ying, Stavridis et al. 2003) and the use of a *Gsc*-GFP reporter to investigate the differentiation course of mesendodermal cells (Tada, Era et al. 2005). Random mutations can then generated in this modified cell line. The mutant cells are induced to differentiate under optimized conditions, and the cells that do not express the reporter can be sorted out by FACS and further analyzed. Fluorescent cells can also be screened in a high-throughput anner using live cell imaging machines.

6.6 The future of genetic screens in mouse ES cells

As I discussed before, mouse ES cells are a unique experimental system that not only has the potential to be a model for mouse early embryogenesis but also sheds the light on how to manipulate their human counterparts to treat human diseases. However the factors and the pathways that direct their differentiation are still not well understood. So genetic screens for discrete differentiation steps can provide an immense amount of data and information to elucidate the regulation of pathways underlying this process (Grimm 2004).

The biggest obstacle for a genetic screen in ES cells is the generation of recessive mutations. We have demonstrated that we can use a strategy which combines regional trapping and inducible mitotic recombination to generate recessive mutations in a region of interest. A genetic screen using these homozygous clones has identified genes that are involved in ES cell *in vitro* differentiation. Thus we have shown that a genetic screen of a complex pathway like *in vitro* differentiation is feasible in ES cells.

Other mutagenesis methods in ES cells can also be combined with inducible mitotic recombination to generate homozygous mutations, such as ENU, irradiation, transposons and gene targeting. RNAi can also be used to perform recessive genetic screens *in vitro*. Because of the limitations of every existing mutagenesis method, it is likely that a combination of different methods is needed to saturate the mouse genome.

To use mouse ES cell *in vitro* differentiation in a genetic screen, a lot of fundamental work still needs to be done. For example, it would be an advantage to know how the expression of each mouse gene changes during the whole differentiation process. This will not only provide a background control for mutant phenotyping, it will also provide a set of markers for each of the differentiation steps and cell lineages, which will be more reliable than just monitoring a few markers.

The limiting factor for a high throughput genetic assay in mammalian cells is always the read-out, or the detection of the cellular changes (Grimm 2004). The use of cDNA and oligonucleotide microarrays is one of the solutions. FACS sorting based on different cell lineage specific markers is another promising way to determine ES cell *in vitro* differentiation potential. Or florescence reporters can be knocked into cell lineage marker genes and these can be used to monitor the expression of these markers in the differentiation process.

The International mouse knockout project has already proposed to systematically knockout every mouse gene (Austin, Battey et al. 2004; Auwerx, Avner et al. 2004). Known or predicted human disease genes will likely be high priority candidates. But how to decide the priority of other genes, especially those genes that no biological function has ever been attributed, will be a challenge for the organizers of this international program. *In vitro* data can provide some useful information about the function of these unknown genes. For example, it will be helpful for the researchers to decide which targeting strategy to use (for example, conventional or conditional knockout) and even which phenotypes to expect. So an ES cell *in vitro* differentiation screen can serve as a pre-screen for the analysis of gene function in whole animals in a large-scale knockout project.

To make such a genetic screen possible, it is necessary to make a library of homozygous mutant ES cells. It can be achieve by generating a library of mutants of a mixture of different genotypes (Guo, Wang et al. 2004; Yusa,

Horie et al. 2004). The advantage of this strategy is that the library is easy to make and maintain. However, this strategy has limited the application of the library to genetic screens in which mutants are identified by their resistance to a specific mutagen. It is impossible to select for mutants that are sensitive to the same mutagen which can be equally important to elucidate a complicated genetic pathway. On the other hand, a genetic screen can also be performed on an array of homozygous ES cells mutants. These homozygous mutants, which can be maintained in a format convenient for high-throughput screens, can be exposed to a range of different concentrations of a specific mutagen, which can not only identify mutants that are sensitive or resistant to this mutagen, but also determine the levels of resistance or sensitivity of these mutants, which can be informative to their role in the interested genetic pathway. Pure homozygous mutant ES cell clones are particularly important for genetic screens on ES cell differentiation because mutants are difficult to be identified by drug selection. Homozygous mutant ES cell clones can be exposed to different differentiation inducers to analysis their differentiation into a variety of cell lineages.

In this study, we have demonstrated that inducible mitotic recombination can be used to generate homozygous gene-trap mutations in mouse embryonic stem cells in a high-throughput way. Homozygous mutant ES cells lines produced by this strategy can be used for genetic screens. However, the genetic instability of ES cells in culture and the epigenetic changes caused by induced mitotic recombination might interfere with the phenotype-driven screens. Care need be taken to choose appropriate positive and negative control cell lines to keep the background of the screens to a reasonable level. On the other hand, genetic and epigenetic instabilities also exist in the other existing high-throughput method to generate homozygous mutant ES cells using *Blm*-deficient ES cells. *Blm*-deficient ES cells have already been successfully used for phenotype-driven screens (Guo, Wang et al. 2004; Yusa, Horie et al. 2004), so it is reasonable to predict these background interferences can be controlled by a good experimental design.

Inducible mitotic recombination is also compatible with other mutagenesis methods, including ENU (Chen, Yee et al. 2000; Munroe, Bergstrom et al. 2000), transposon mutagenesis (Ivics, Hackett et al. 1997; Luo, Ivics et al. 1998) and gene targeting (Thomas and Capecchi 1987). RNAi is another way to knock down gene expression for recessive screens in ES cells (Berns, Hijmans et al. 2004; Paddison, Silva et al. 2004). The limitations of the existing mutagenesis methods suggest that the most effective way to saturate the genome with recessive mutations is to use a combination of these methods. Recessive genetic screens in mouse ES cells will accelerate functional studies of genes in the mouse, as well as provide a foundation for applied research to differentiate human ES cells into cell types that can be potentially used to treat the human diseases.

References

(1998). "Genome sequence of the nematode C. elegans: a platform for investigating biology." Science **282**(5396): 2012-8.
Abuin, A. and A. Bradley (1996). "Recycling selectable markers in mouse embryonic stem cells." Mol Cell Biol **16**(4): 1851-6.
Adams, M. D., S. E. Celniker, et al. (2000). "The genome sequence of Drosophila melanogaster." Science **287**(5461): 2185-95.
Ankeny, R. A. (2001). "The natural history of Caenorhabditis elegans research." Nat Rev Genet **2**(6): 474-9.
Araki, K., M. Araki, et al. (1995). "Site-specific recombination of a transgene in fertilized eggs by transient expression of Cre recombinase." Proc Natl Acad Sci U S A **92**(1): 160-4.
Austin, C. P., J. F. Battey, et al. (2004). "The knockout mouse project." Nat Genet **36**(9): 921-4.
Auwerx, J., P. Avner, et al. (2004). "The European dimension for the mouse genome mutagenesis program." Nat Genet **36**(9): 925-7.
Bauer, D. E., G. Hatzivassiliou, et al. (2005). "ATP citrate lyase is an important component of cell growth and transformation." Oncogene.
Beigneux, A. P., C. Kosinski, et al. (2004). "ATP-citrate lyase deficiency in the mouse." J Biol Chem **279**(10): 9557-64.
Berns, K., E. M. Hijmans, et al. (2004). "A large-scale RNAi screen in human cells identifies new components of the p53 pathway." Nature **428**(6981): 431-7.
Berwick, D. C., I. Hers, et al. (2002). "The identification of ATP-citrate lyase as a protein kinase B (Akt) substrate in primary adipocytes." J Biol Chem **277**(37): 33895-900.
Bessereau, J. L., A. Wright, et al. (2001). "Mobilization of a Drosophila transposon in the Caenorhabditis elegans germ line." Nature **413**(6851): 70-4.
Bonaldo, P., K. Chowdhury, et al. (1998). "Efficient gene trap screening for novel developmental genes using IRES beta geo vector and in vitro preselection." Exp Cell Res **244**(1): 125-36.
Bradley, A., M. Evans, et al. (1984). "Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines." Nature **309**(5965): 255-6.
Brenner, S. (1974). "The genetics of Caenorhabditis elegans." Genetics **77**(1): 71-94.
Capecchi, M. R. (1989). "Altering the genome by homologous recombination." Science **244**(4910): 1288-92.
Carlson, C. M., A. J. Dupuy, et al. (2003). "Transposon mutagenesis of the mouse germline." Genetics **165**(1): 243-56.
Carthew, R. W. (2001). "Gene silencing by double-stranded RNA." Curr Opin Cell Biol **13**(2): 244-8.
Cattanach, B. M. and J. Jones (1994). "Genetic imprinting in the mouse: implications for gene regulation." J Inherit Metab Dis **17**(4): 403-20.
Chambers, I., D. Colby, et al. (2003). "Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells." Cell **113**(5): 643-55.

Chambers, I. and A. Smith (2004). "Self-renewal of teratocarcinoma and embryonic stem cells." Oncogene **23**(43): 7150-60.
Chen, Y., D. Yee, et al. (2000). "Genotype-based screen for ENU-induced mutations in mouse embryonic stem cells." Nat Genet **24**(3): 314-7.
Chester, N., F. Kuo, et al. (1998). "Stage-specific apoptosis, developmental delay, and embryonic lethality in mice homozygous for a targeted disruption in the murine Bloom's syndrome gene." Genes Dev **12**(21): 3382-93.
Copeland, N. G., N. A. Jenkins, et al. (2001). "Recombineering: a powerful new tool for mouse functional genomics." Nat Rev Genet **2**(10): 769-79.
Czyz, J., C. Wiese, et al. (2003). "Potential of embryonic and adult stem cells in vitro." Biol Chem **384**(10-11): 1391-409.
Doetschman, T. C., H. Eistetter, et al. (1985). "The in vitro development of blastocyst-derived embryonic stem cell lines: formation of visceral yolk sac, blood islands and myocardium." J Embryol Exp Morphol **87**: 27-45.
Dupuy, A. J., K. Clark, et al. (2002). "Mammalian germ-line transgenesis by transposition." Proc Natl Acad Sci U S A **99**(7): 4495-9.
Dupuy, A. J., S. Fritz, et al. (2001). "Transposition and gene disruption in the male germline of the mouse." Genesis **30**(2): 82-8.
Elbashir, S. M., J. Harborth, et al. (2001). "Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells." Nature **411**(6836): 494-8.
Evans, M. J. and M. H. Kaufman (1981). "Establishment in culture of pluripotential cells from mouse embryos." Nature **292**(5819): 154-6.
Farley, F. W., P. Soriano, et al. (2000). "Widespread recombinase expression using FLPeR (flipper) mice." Genesis **28**(3-4): 106-10.
Forsburg, S. L. (2001). "The art and design of genetic screens: yeast." Nat Rev Genet **2**(9): 659-68.
Friedman, R. and A. L. Hughes (2001). "Pattern and timing of gene duplication in animal genomes." Genome Res **11**(11): 1842-7.
Friedrich, G. and P. Soriano (1991). "Promoter traps in embryonic stem cells: a genetic screen to identify and mutate developmental genes in mice." Genes Dev **5**(9): 1513-23.
Gajovic, S., K. Chowdhury, et al. (1998). "Genes expressed after retinoic acid-mediated differentiation of embryoid bodies are likely to be expressed during embryo development." Exp Cell Res **242**(1): 138-43.
Garrett, H. G. and C. M. Grisham (1999). "Biochemistry (Second edition)." **Chapter 25**: 802-805.
Goffeau, A., B. G. Barrell, et al. (1996). "Life with 6000 genes." Science **274**(5287): 546, 563-7.
Goldstein, J. L. (2001). "Laskers for 2001: knockout mice and test-tube babies." Nat Med **7**(10): 1079-80.
Golic, K. G. (1991). "Site-specific recombination between homologous chromosomes in Drosophila." Science **252**(5008): 958-61.
Golic, K. G. and S. Lindquist (1989). "The FLP recombinase of yeast catalyzes site-specific recombination in the Drosophila genome." Cell **59**(3): 499-509.

Goss, K. H., M. A. Risinger, et al. (2002). "Enhanced tumor formation in mice heterozygous for Blm mutation." Science **297**(5589): 2051-3.
Gossler, A., A. L. Joyner, et al. (1989). "Mouse embryonic stem cells and reporter constructs to detect developmentally regulated genes." Science **244**(4903): 463-5.
Greber, B., H. Lehrach, et al. (2005). "Mouse splice mutant generation from ENU-treated ES cells--a gene-driven approach." Genomics **85**(5): 557-62.
Grignani, F., T. Kinsella, et al. (1998). "High-efficiency gene transfer and selection of human hematopoietic progenitor cells with a hybrid EBV/retroviral vector expressing the green fluorescence protein." Cancer Res **58**(1): 14-9.
Grimm, S. (2004). "The art and design of genetic screens: mammalian culture cells." Nat Rev Genet **5**(3): 179-89.
Gu, H., Y. R. Zou, et al. (1993). "Independent control of immunoglobulin switch recombination at individual switch regions evidenced through Cre-loxP-mediated gene targeting." Cell **73**(6): 1155-64.
Guo, G., W. Wang, et al. (2004). "Mismatch repair genes identified using genetic screens in Blm-deficient embryonic stem cells." Nature **429**(6994): 891-5.
Hamilton, D. L. and K. Abremski (1984). "Site-specific recombination by the bacteriophage P1 lox-Cre system. Cre-mediated synapsis of two lox sites." J Mol Biol **178**(2): 481-6.
Hannon, G. J. (2002). "RNA interference." Nature **418**(6894): 244-51.
Hansen, J., T. Floss, et al. (2003). "A large-scale, gene-driven mutagenesis approach for the functional analysis of the mouse genome." Proc Natl Acad Sci U S A **100**(17): 9918-22.
Hartwell, L. H., J. Culotti, et al. (1970). "Genetic control of the cell-division cycle in yeast. I. Detection of mutants." Proc Natl Acad Sci U S A **66**(2): 352-9.
Hotta, K., H. Takahashi, et al. (2000). "Characterization of Brachyury-downstream notochord genes in the Ciona intestinalis embryo." Dev Biol **224**(1): 69-80.
Hotta, K., H. Takahashi, et al. (1999). "Temporal expression patterns of 39 Brachyury-downstream genes associated with notochord formation in the Ciona intestinalis embryo." Dev Growth Differ **41**(6): 657-64.
Hrabe de Angelis, M. H., H. Flaswinkel, et al. (2000). "Genome-wide, large-scale production of mutant mice by ENU mutagenesis." Nat Genet **25**(4): 444-7.
Humpherys, D., K. Eggan, et al. (2001). "Epigenetic instability in ES cells and cloned mice." Science **293**(5527): 95-7.
Ivics, Z., P. B. Hackett, et al. (1997). "Molecular reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells." Cell **91**(4): 501-10.
Jaenisch, R. (1976). "Germ line integration and Mendelian transmission of the exogenous Moloney leukemia virus." Proc Natl Acad Sci U S A **73**(4): 1260-4.
Jansen, G., E. Hazendonk, et al. (1997). "Reverse genetics by chemical mutagenesis in Caenorhabditis elegans." Nat Genet **17**(1): 119-21.

Jorgensen, E. M. and S. E. Mango (2002). "The art and design of genetic screens: caenorhabditis elegans." Nat Rev Genet **3**(5): 356-69.
Justice, M. J. (2000). "Capitalizing on large-scale mouse mutagenesis screens." Nat Rev Genet **1**(2): 109-15.
Kamath, R. S., A. G. Fraser, et al. (2003). "Systematic functional analysis of the Caenorhabditis elegans genome using RNAi." Nature **421**(6920): 231-7.
Kawasaki, H., K. Mizuseki, et al. (2000). "Induction of midbrain dopaminergic neurons from ES cells by stromal cell-derived inducing activity." Neuron **28**(1): 31-40.
Ki, S. W., K. Ishigami, et al. (2000). "Radicicol binds and inhibits mammalian ATP citrate lyase." J Biol Chem **275**(50): 39231-6.
Kile, B. T., K. E. Hentges, et al. (2003). "Functional genetic analysis of mouse chromosome 11." Nature **425**(6953): 81-6.
Koller, B. H., P. Marrack, et al. (1990). "Normal development of mice deficient in beta 2M, MHC class I proteins, and CD8+ T cells." Science **248**(4960): 1227-30.
Kuehn, M. R., A. Bradley, et al. (1987). "A potential animal model for Lesch-Nyhan syndrome through introduction of HPRT mutations into mice." Nature **326**(6110): 295-8.
Kumar, A. and M. Snyder (2001). "Emerging technologies in yeast genomics." Nat Rev Genet **2**(4): 302-12.
Lakso, M., B. Sauer, et al. (1992). "Targeted oncogene activation by site-specific recombination in transgenic mice." Proc Natl Acad Sci U S A **89**(14): 6232-6.
Lander, E. S., L. M. Linton, et al. (2001). "Initial sequencing and analysis of the human genome." Nature **409**(6822): 860-921.
Leahy, A., J. W. Xiong, et al. (1999). "Use of developmental marker genes to define temporal and spatial patterns of differentiation during embryoid body formation." J Exp Zool **284**(1): 67-81.
Lefebvre, L., N. Dionne, et al. (2001). "Selection for transgene homozygosity in embryonic stem cells results in extensive loss of heterozygosity." Nat Genet **27**(3): 257-8.
Li, J., H. Shen, et al. (1999). "Leukaemia disease genes: large-scale cloning and pathway predictions." Nat Genet **23**(3): 348-53.
Lindsay, E. A., F. Vitelli, et al. (2001). "Tbx1 haploinsufficieny in the DiGeorge syndrome region causes aortic arch defects in mice." Nature **410**(6824): 97-101.
Lindsley, D. L., L. Sandler, et al. (1972). "Segmental aneuploidy and the genetic gross structure of the Drosophila genome." Genetics **71**(1): 157-84.
Liu, L. X., J. M. Spoerke, et al. (1999). "High-throughput isolation of Caenorhabditis elegans deletion mutants." Genome Res **9**(9): 859-67.
Liu, P., N. A. Jenkins, et al. (2002). "Efficient Cre-loxP-induced mitotic recombination in mouse embryonic stem cells." Nat Genet **30**(1): 66-72.
Liu, P., N. A. Jenkins, et al. (2003). "A highly efficient recombineering-based method for generating conditional knockout mutations." Genome Res **13**(3): 476-84.

Liu, P., H. Zhang, et al. (1998). "Embryonic lethality and tumorigenesis caused by segmental aneuploidy on mouse chromosome 11." Genetics **150**(3): 1155-68.

Liu, X., H. Wu, et al. (1997). "Trisomy eight in ES cells is a common potential problem in gene targeting and interferes with germ line transmission." Dev Dyn **209**(1): 85-91.

Luo, G., Z. Ivics, et al. (1998). "Chromosomal transposition of a Tc1/mariner-like element in mouse embryonic stem cells." Proc Natl Acad Sci U S A **95**(18): 10769-73.

Luo, G., I. M. Santoro, et al. (2000). "Cancer predisposition caused by elevated mitotic recombination in Bloom mice." Nat Genet **26**(4): 424-9.

Maine, E. M. (2001). "RNAi As a tool for understanding germline development in Caenorhabditis elegans: uses and cautions." Dev Biol **239**(2): 177-89.

Mann, J. R., I. Gadi, et al. (1990). "Androgenetic mouse embryonic stem cells are pluripotent and cause skeletal defects in chimeras: implications for genetic imprinting." Cell **62**(2): 251-60.

McDaniel, L. D., N. Chester, et al. (2003). "Chromosome instability and tumor predisposition inversely correlate with BLM protein levels." DNA Repair (Amst) **2**(12): 1387-404.

McMahon, A. P. and A. Bradley (1990). "The Wnt-1 (int-1) proto-oncogene is required for development of a large region of the mouse brain." Cell **62**(6): 1073-85.

Mikkers, H., J. Allen, et al. (2002). "High-throughput retroviral tagging to identify components of specific signaling pathways in cancer." Nat Genet **32**(1): 153-9.

Mitchell, K. J., K. I. Pinson, et al. (2001). "Functional analysis of secreted and transmembrane proteins critical to mouse development." Nat Genet **28**(3): 241-9.

Mortensen, R. M., D. A. Conner, et al. (1992). "Production of homozygous mutant ES cells with a single targeting construct." Mol Cell Biol **12**(5): 2391-5.

Munroe, R. J., S. L. Ackerman, et al. (2004). "Genomewide two-generation screens for recessive mutations by ES cell mutagenesis." Mamm Genome **15**(12): 960-5.

Munroe, R. J., R. A. Bergstrom, et al. (2000). "Mouse mutants from chemically mutagenized embryonic stem cells." Nat Genet **24**(3): 318-21.

Nagy, A. (2000). "Cre recombinase: the universal reagent for genome tailoring." Genesis **26**(2): 99-109.

Nagy, A., C. Moens, et al. (1998). "Dissecting the role of N-myc in development using a single targeting vector to generate a series of alleles." Curr Biol **8**(11): 661-4.

Nagy, A., J. Rossant, et al. (1993). "Derivation of completely cell culture-derived mice from early-passage embryonic stem cells." Proc Natl Acad Sci U S A **90**(18): 8424-8.

Nakano, T., H. Kodama, et al. (1994). "Generation of lymphohematopoietic cells from embryonic stem cells in culture." Science **265**(5175): 1098-101.

Nichols, J., E. P. Evans, et al. (1990). "Establishment of germ-line-competent embryonic stem (ES) cells using differentiation inhibiting activity." Development **110**(4): 1341-8.

Ning, Z., A. J. Cox, et al. (2001). "SSAHA: a fast search method for large DNA databases." Genome Res **11**(10): 1725-9.

Nishikawa, S. I., S. Nishikawa, et al. (1998). "Progressive lineage analysis by cell sorting and culture identifies FLK1+VE-cadherin+ cells at a diverging point of endothelial and hemopoietic lineages." Development **125**(9): 1747-57.

Nolan, P. M., J. Peters, et al. (2000). "A systematic, genome-wide, phenotype-driven mutagenesis programme for gene function studies in the mouse." Nat Genet **25**(4): 440-3.

Nowrousian, M., S. Masloff, et al. (1999). "Cell differentiation during sexual development of the fungus Sordaria macrospora requires ATP citrate lyase activity." Mol Cell Biol **19**(1): 450-60.

Nurse, P. (1975). "Genetic control of cell size at cell division in yeast." Nature **256**(5518): 547-51.

Nusslein-Volhard, C. and E. Wieschaus (1980). "Mutations affecting segment number and polarity in Drosophila." Nature **287**(5785): 795-801.

Orban, P. C., D. Chui, et al. (1992). "Tissue- and site-specific DNA recombination in transgenic mice." Proc Natl Acad Sci U S A **89**(15): 6861-5.

Paddison, P. J., J. M. Silva, et al. (2004). "A resource for large-scale RNA-interference-based screens in mammals." Nature **428**(6981): 427-31.

Ramalho-Santos, M., S. Yoon, et al. (2002). ""Stemness": transcriptional profiling of embryonic and adult stem cells." Science **298**(5593): 597-600.

Ramirez-Solis, R., A. C. Davis, et al. (1993). "Gene targeting in embryonic stem cells." Methods Enzymol **225**: 855-78.

Ramirez-Solis, R., P. Liu, et al. (1995). "Chromosome engineering in mice." Nature **378**(6558): 720-4.

Rohwedel, J., K. Guan, et al. (2001). "Embryonic stem cells as an in vitro model for mutagenicity, cytotoxicity and embryotoxicity studies: present state and future prospects." Toxicol In Vitro **15**(6): 741-53.

Rorth, P. (1996). "A modular misexpression screen in Drosophila detecting tissue-specific phenotypes." Proc Natl Acad Sci U S A **93**(22): 12418-22.

Rorth, P., K. Szabo, et al. (1998). "Systematic gain-of-function genetics in Drosophila." Development **125**(6): 1049-57.

Rossant, J. and A. Spence (1998). "Chimeras and mosaics in mouse mutant analysis." Trends Genet **14**(9): 358-63.

Rubin, G. M. and E. B. Lewis (2000). "A brief history of Drosophila's contributions to genome research." Science **287**(5461): 2216-8.

Sauer, B. and N. Henderson (1988). "Site-specific DNA recombination in mammalian cells by the Cre recombinase of bacteriophage P1." Proc Natl Acad Sci U S A **85**(14): 5166-70.

Schwartzberg, P. L., E. J. Robertson, et al. (1990). "Targeted gene disruption of the endogenous c-abl locus by homologous recombination with DNA encoding a selectable fusion protein." Proc Natl Acad Sci U S A **87**(8): 3210-4.

Sharov, A. A., Y. Piao, et al. (2003). "Transcriptome analysis of mouse stem cells and early embryos." PLoS Biol **1**(3): E74.
Shuman, H. A. and T. J. Silhavy (2003). "The art and design of genetic screens: Escherichia coli." Nat Rev Genet **4**(6): 419-31.
Simon, M. A. (1994). "Signal transduction during the development of the Drosophila R7 photoreceptor." Dev Biol **166**(2): 431-42.
Skarnes, W. C., J. E. Moss, et al. (1995). "Capturing genes encoding membrane and secreted proteins important for mouse development." Proc Natl Acad Sci U S A **92**(14): 6592-6.
Skarnes, W. C., H. von Melchner, et al. (2004). "A public gene trap resource for mouse functional genomics." Nat Genet **36**(6): 543-4.
Smith, A. J., M. A. De Sousa, et al. (1995). "A site-directed chromosomal translocation induced in embryonic stem cells by Cre-loxP recombination." Nat Genet **9**(4): 376-85.
St Johnston, D. (2002). "The art and design of genetic screens: Drosophila melanogaster." Nat Rev Genet **3**(3): 176-88.
Stanford, W. L., J. B. Cohn, et al. (2001). "Gene-trap mutagenesis: past, present and beyond." Nat Rev Genet **2**(10): 756-68.
Stoykova, A., K. Chowdhury, et al. (1998). "Gene trap expression and mutational analysis for genes involved in the development of the mammalian nervous system." Dev Dyn **212**(2): 198-213.
Stryke, D., M. Kawamoto, et al. (2003). "BayGenomics: a resource of insertional mutations in mouse embryonic stem cells." Nucleic Acids Res **31**(1): 278-81.
Su, H., A. A. Mills, et al. (2002). "A targeted X-linked CMV-Cre line." Genesis **32**(2): 187-8.
Su, H., X. Wang, et al. (2000). "Nested chromosomal deletions induced with retroviral vectors in mice." Nat Genet **24**(1): 92-5.
Sulston, J. E. (1976). "Post-embryonic development in the ventral cord of Caenorhabditis elegans." Philos Trans R Soc Lond B Biol Sci **275**(938): 287-97.
Sulston, J. E. and H. R. Horvitz (1977). "Post-embryonic cell lineages of the nematode, Caenorhabditis elegans." Dev Biol **56**(1): 110-56.
Sulston, J. E., E. Schierenberg, et al. (1983). "The embryonic cell lineage of the nematode Caenorhabditis elegans." Dev Biol **100**(1): 64-119.
Suzuki, T., H. Shen, et al. (2002). "New genes involved in cancer identified by retroviral tagging." Nat Genet **32**(1): 166-74.
Tada, S., T. Era, et al. (2005). "Characterization of mesendoderm: a diverging point of the definitive endoderm and mesoderm in embryonic stem cell differentiation culture." Development **132**(19): 4363-74.
te Riele, H., E. R. Maandag, et al. (1990). "Consecutive inactivation of both alleles of the pim-1 proto-oncogene by homologous recombination in embryonic stem cells." Nature **348**(6302): 649-51.
Tessier, C. R., G. A. Doyle, et al. (2004). "Mammary tumor induction in transgenic mice expressing an RNA-binding protein." Cancer Res **64**(1): 209-14.
Theodosiou, N. A. and T. Xu (1998). "Use of FLP/FRT system to study Drosophila development." Methods **14**(4): 355-65.

Thomas, K. R. and M. R. Capecchi (1987). "Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells." Cell **51**(3): 503-12.

Tsien, J. Z., D. F. Chen, et al. (1996). "Subregion- and cell type-restricted gene knockout in mouse brain." Cell **87**(7): 1317-26.

Venter, J. C., M. D. Adams, et al. (2001). "The sequence of the human genome." Science **291**(5507): 1304-51.

Vittet, D., M. H. Prandini, et al. (1996). "Embryonic stem cells differentiate in vitro to endothelial cells through successive maturation steps." Blood **88**(9): 3424-31.

Vivian, J. L., Y. Chen, et al. (2002). "An allelic series of mutations in Smad2 and Smad4 identified in a genotype-based screen of N-ethyl-N-nitrosourea-mutagenized mouse embryonic stem cells." Proc Natl Acad Sci U S A **99**(24): 15542-7.

von Melchner, H. and H. E. Ruley (1989). "Identification of cellular promoters by using a retrovirus promoter trap." J Virol **63**(8): 3227-33.

Waterston, R. H., K. Lindblad-Toh, et al. (2002). "Initial sequencing and comparative analysis of the mouse genome." Nature **420**(6915): 520-62.

Wiles, M. V. and G. Keller (1991). "Multiple hematopoietic lineages develop from embryonic stem (ES) cells in culture." Development **111**(2): 259-67.

Wobus, A. M. (2001). "Potential of embryonic stem cells." Mol Aspects Med **22**(3): 149-64.

Wobus, A. M., K. Guan, et al. (2002). "Embryonic stem cells as a model to study cardiac, skeletal muscle, and vascular smooth muscle cell differentiation." Methods Mol Biol **185**: 127-56.

Wobus, A. M., G. Wallukat, et al. (1991). "Pluripotent mouse embryonic stem cells are able to differentiate into cardiomyocytes expressing chronotropic responses to adrenergic and cholinergic agents and Ca2+ channel blockers." Differentiation **48**(3): 173-82.

Wood, V., R. Gwilliam, et al. (2002). "The genome sequence of Schizosaccharomyces pombe." Nature **415**(6874): 871-80.

Xu, T. and S. D. Harrison (1994). "Mosaic analysis using FLP recombinase." Methods Cell Biol **44**: 655-81.

Xu, T. and G. M. Rubin (1993). "Analysis of genetic mosaics in developing and adult Drosophila tissues." Development **117**(4): 1223-37.

Xu, T., W. Wang, et al. (1995). "Identifying tumor suppressors in genetic mosaics: the Drosophila lats gene encodes a putative protein kinase." Development **121**(4): 1053-63.

Yamashita, J., H. Itoh, et al. (2000). "Flk1-positive cells derived from embryonic stem cells serve as vascular progenitors." Nature **408**(6808): 92-6.

Ying, Q. L., M. Stavridis, et al. (2003). "Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture." Nat Biotechnol **21**(2): 183-6.

Yu, Y. and A. Bradley (2001). "Engineering chromosomal rearrangements in mice." Nat Rev Genet **2**(10): 780-90.