Chapter Two – Materials and methods

1. Vectors

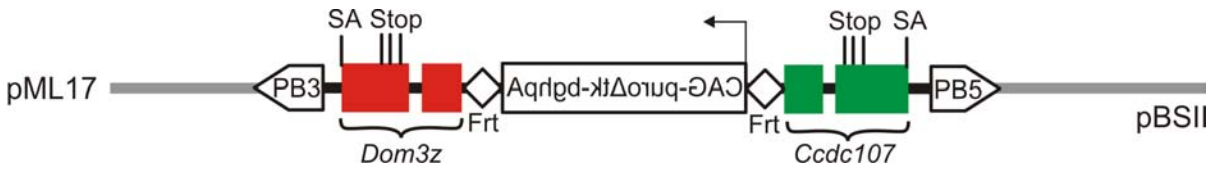
1.1. PB transposon and transposase vectors

piggyBac Transposase plasmids

A CMV promoter-driven mammalian codon-optimised *piggyBac* transposase (mPBase) expressing plasmid was obtained from Juan Cadinanos (CMV-mPBase). The plasmid was modified with the Neomycin resistant cassette removed (pmPBΔNeo). A CMV promoter driven hyperactive piggyBac transposase (HyPBase) expressing plasmid was obtained from Kosuke Yusa (CMV-hyPBase).

To construct PGK-Puro and PBase co-expressing plasmids, a PGK-Puromycin resistant cassette, the *Eco*RI/*Not*I fragment from plasmid pPGKPuro, was blunt ligated to the CMV promoter driven mPBase or HyPBase plasmid, *Na*eI/*Sap*I fragment, to give CMV-mPBase-PGK-Puro and CMV-hyPBase-PGK-Puro. The control plasmid CAG-eGFP-bGHpA contains the *PGK-PuroΔtk* cassette.

Construction of the mutagenic PB (pML17)



The PB inverted terminal repeats (PBITRs) were obtained from Qi Liang. pML6 Plasmid was generated by inserting PCR amplified PB5 ITR, using Bsd-PB5 ITR as template, into *Sal*I/*ECoR*V digested vector plasmid pQL2+PB3 (obtained from Qi Liang). The PGK promoter and bgh-PolyA signal were removed from the pML6 by PCR amplification of the EM7-Neo cassette and ligated into *Nhe*I/*BamH*I-digested pML6 to generate pML9 plasmid with a floxed EM7-Neo cassette. Two recombineering arms were PCR amplified from the human *HPRT* gene intron 2,690 bp away from the 3' end of the *hprt* exon 2. The 5' and 3' recombineering arms were

sequentially inserted 5' to the PB3'ITR (*Kpn*I/*Xho*I) and 3' to the PB5'ITR (*BamH*I/*Not*I) of pML9 to give rise to pML12. A *Not*I site was also introduced at the 5' of the 5' recombineering arm during PCR.

Dom3z and *Ccdc107* last two exons were PCR-amplified from BAC bMQ-365E12 and bMQ-242N12 respectively, and cloned into pML5 using primers Dom3zF and Dom3zR for *Dom3z*-exon cloning and Ccdc107F and Ccdc107R for *Ccdc107*-exon cloning.

Primers

Ccdc107F: GCATTTAGGCCGGCCGAGCCAAGGAGACAGACTGG

CCdc107R: GGAATCGGCGCGCCTTTATTTCCCCACTGGATCTT

Dom3zF: GCATTTAGGCCGGCCCCAAGTCCTCAGACCCAGTG

Dom3zR: GGAATCGGCGCGCCGCCAGCCTCTACACCCAGTA

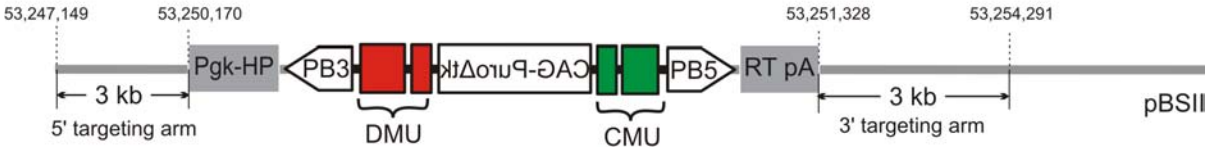
Dom3Z mutagenic unit was extracted from *Age*I/*ECoR*I digested pML5-Dom3Z and it was ligated into pML5-Ccdc107 (*NgoMI*V/*ECoR*I) to generate a PB transposon with a bidirectional pair of mutagen units pML5-Dom3Z-Ccdc107. *EcoR*V/*Sfo*I digestion of pML5-Dom3Z-Ccdc107 followed with blunt ligation was done to remove the residual Neo cassette in the vector. Additional stop codons were introduced to two other reading frames for both Dom3Z and Ccdc107 penultimate exons by site directed mutagenesis.

An *Fse*I/*Sac*II fragment containing the pair of mutagen units from pML5-Dom3Z-Ccdc107 was blunt ligated into *Fse*I/*ECoR*I digested vector pML12 to give rise to pML13. *Nhe*I/*Age*I digested pML10+FRT plasmid containing the FRT flanked CAGG-PuroΔTK cassette was blunt ligated into *Asc*I-linearised vector pML13 to complete all the modules in the PB transposon and to give rise to pML17.

1.2. Targeting vectors

Construction of targeting vector Gdf9TVPB (pGDF9T-ML4)

pGDF9T-ML4

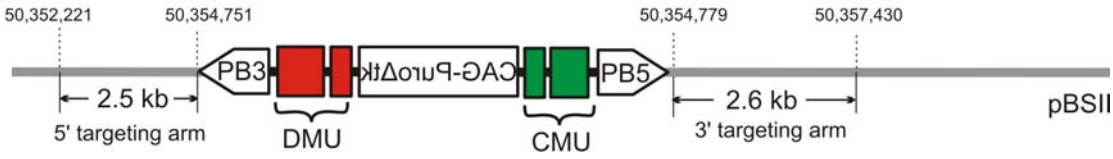


The coordinates are genomic fragments used for the targeting arms and are based on NCBI Build 37.

The *Not*I fragment of pML17 containing the mutagenic PB was used as the linear fragment to insert the PB onto plasmid GDF9T-ML2 using recombineering in the EL350 bacterial strain. The resultant plasmid GDF9T-ML3 was subject to *in vitro* Cre recombinase-mediated excision of floxed kanamycin resistant cassette, giving rise to final targeting vector pGDF9T-ML4. The targeting vector was linearised with *Pme*I before electroporation.

Construction of targeting vector HprtTVIn2PB

HprtTVIn2PB



The coordinates are genomic fragments used for the targeting arms and are based on NCBI Build 37.

The *Asc*I site in pML17 was destroyed by digesting the plasmid with *Asc*I followed by klenow treatment and self-ligation to give pML18. *Xho*I/*Hind*III and *Hind*III/*Pst*I fragments from the pML18 were ligated with *Xho*I/*Pst*I fragment of the ARM1 (Haydn Prosser) to give pML20. The hprt targeting arms were extracted from RMCE cassette plasmid (pCEI-3, Haydn Prosser), 2.5 kb *Fse*I/*Xho*I fragment from pCEI-3 containing the hprt 5' homology arm was cloned into *Fse*I/*Xho*I digested pML20 to give phprtTV-left. The 3' homology arm was PCR amplified from pCEI with the KOD system using primers ML45f and ML45r. The PCR fragment was then digested with PacI and AscI and cloned into phprtTV-left to give the final targeting vector

HprtTVIn2PB. This targeting vector was designed to insert PB transposon into intron 2 of the *Hprt* locus. The homology arms were AB1 origin. The vector was linearised with *Asc*I for gene targeting.

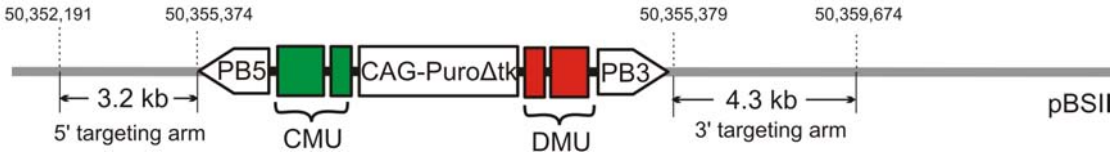
Primers

ML45f: ACCTTAATTAAGATAGATGTTATAGTGTACTCTCCTCTCC

ML45r: AAAGGCGCGCCAGGCACTCAAGATGATCCATATACT

Construction of targeting vector HprtTVE3PB

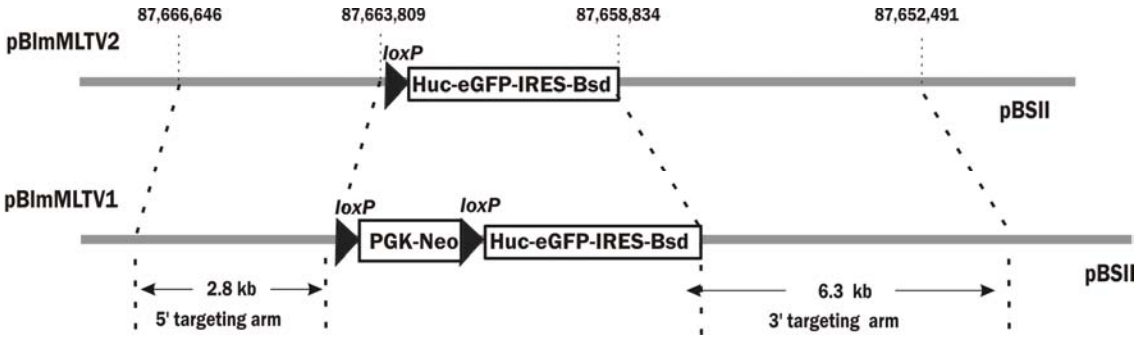
HprtTVE3PB



The coordinates are genomic fragments used for the targeting arms and are based on NCBI Build 37.

The *BspE*I/*BsiW*I fragment and *BspE*I/*Nsi*I fragment of the pML20 was ligated into *Nsi*I/*BsiW*I fragment of the targeting vector pDTAhprtE3RCAG (Kosuke Yusa). This vector contains the homology arms retrieved from B6 strain BAC, for targeted insertion of PB into a TTAA site within the exon 3 of the *hprt* locus. The vector was linearised with *Pme*I for gene targeting.

Construction for targeting vector BlmTV (BlmTV1 and BlmTV2)



The coordinates are genomic fragments used for the targeting arms and are based on NCBI Build 37.

Human ubiquitin C (Huc) promoter was PCR-amplified from human BAC RP11 214K23 using primers ML67f and ML67r with Phusion PCR system and cloned into pML27 to give pML28. pML38 (pCAG-IRES-Bsd) was made by three-way ligation of the *EcoR*I fragment from pCAG-IRES-Neo, the *EcoR*I/*Nco*I fragment from pCAG-IRES-Neo and the *EcoR*I/*Nco*I fragment from PCR amplified Bsd resistant gene from pML313. IRES-Bsd was PCR amplified with ML79f and ML78r and cloned into *BsrG*I site in pPB-Huc-eGFP to give pML40.

B6 strain BAC RP24-180J10 was used to construct the *Blm* targeting vector. Following the *Blm* targeting strategy conducted by Luo and co-workers (Luo et al., 2000), the targeting vector was designed to replace the ATG-containing exon 2 of the *Blm* gene with the *Huc-eGFP-IRES-Bsd* reporter to give rise to null *Blm* alleles. *Blm* mini homology arms for driving recombineering of the reporter construct to the BAC were sequentially PCR amplified and cloned into pML40, flanking the Huc-eGFP-IRES-Bsd together with a floxed PGK-EM7-Neo cassette (pML65). A retrieval vector (pML54) was constructed by PCR amplification of the retroviral homology arms from the *Blm*-containing BAC and subsequent cloning into pBSKSII+. The *Asc*I/*Pac*I fragment of pML65 was purified and used for engineering the *Blm*-containing BAC, for replacing the *Blm* exon 2 with the reporter. The correctly targeted BAC-containing bacteria were further used for retrieval of the BAC region containing the 5' and 3' homology arms together with the Huc-eGFP-IRES-Bsd and floxed PGK-EM7-Neo cassette. The retrieval fragment was prepared by linearising pML54 with *Sac*I. The correctly retrieved plasmid gave rise to BlmTVML1. The floxed PGK-EM7-Neo cassette was popped out in EL350 and the resulting plasmid became BlmTVML2. Both BlmTVML1 and BlmTVML2 were linearised with *Pme*I before electroporations.

Primers:

ML67f: AATGGCGCGCCATCGATGGCCTCCGCGCCGGGTTTTGGCGCC

ML67r: AATGCTAGCATCGATGTCTAACAAAAAAGCCAAAAACGGC

ML79f: GACGAGCTGTACAAGTAACCGCCCCCCCCCCCTCTCCCTCCCCCCCCCCTAACGTTAC

ML78r: AACTGAATTCTGTACATTAGCCCTCCCACACATAACCAGAG

Construction of targeting vector GDF9TV-GFP-IRES-Bsd (pML46)

The coordinates are genomic fragments used for the targeting arms and are based on NCBI Build 37.

The *Xho*I/*Not*I fragment of pGDF9TML2, which contained the 3' Gdf9 targeting arm, was ligated into the *Sal*I/*Not*I fragment of the pML40, to give pML43. The 5' Gdf9 targeting arm was PCR-amplified from the pGDF9ML2 using primers ML83f and ML83r with Phusion PCR system. The PCR product was digested with *Asc*I and *Xho*I and cloned into pML43 to give the final targeting vector pML46. The targeting vector was linearised with *Not*I before electroporation.

Primers:

ML83f: TAAAGGCGCGCCCTGACTCCAGAGCACTCTACTACAT

ML83r: TAAACTCGAGCATGGCAGTCACCCGGTCCAGGTTA

1.3. Vectors for miRNA reporters and their targeting vectors

Construction of miR-eGFP

The mouse miR-155 backbone based mirshRNAs

The mouse miR-155 precursor stem loop surrounding sequences were constructed by annealing oligos ML123f and ML123r. The annealed oligos were treated with Klenow fragment 3'-5' exo$^-$ (NEB) to form the full mirshRNA unit with the the original miR-155 BIC 134–283 region replaced with a synthetic polylinker containing two inverted Bbs1 sites (Chung et al., 2006). The treated oligos were then digested with *Mlu*I and *Eag*I and ligated into pBS-LR digested with the same enzymes, to give pML64.

The target specific sequences recognising eGFP were constructed by annealing of two 64 nt oligos. The eGFP targeting sequences were either from published work or designed using

BLOCK-iT™ Pol II miR RNAi Designer (Invitrogen, online software). Five different eGFP sequences were constructed in total using five sets of oligos, ML124f and ML124r, ML125f and ML125r, ML126f and ML126r, ML127f and ML127r, and ML128f and ML128r. The miR-155 based the mirshRNA has the effective sequence residing in the 5' strand. The annealed oligos were treated with T4 PNK (NEB) and ligated into Bbs1 digested pML64, to give pML64-124, pML64-125, pML64-126, pML64-127, and pML64-128 (respectively to the oligo sequences). The constructs were verified by DNA sequencing.

The single unit of the target-specific mirshRNA was PCR amplified with four sets of primers (ML131f and Ml131r, ML132f and ML132r, ML133f and ML133r, ML134f and ML134r) containing restriction sites that allowed five-way ligation of four copies of the target-specific mirshRNA into *Avr*II/*Nde*I digested pML29, giving rise to pML79-124, pML79-125, pML79-126, pML79-127, and pML79-128 respectively. The mirshRNAs with the same or different eGFP target sequences were placed within the intron of the human ubiquitin promoter, within the Huc-EM7-Neo-bghpA cassette flanked by PBITRs. The final constructs were verified by DNA sequencing. pML79-128 was further modified by cloning four copies more of the identical mirshRNA to give pML88, which contains a Huc-EM7-Neo cassette with 8 copies of the mirshRNA recognising eGFP polycistronically expressed from the intron 1 of the Huc promoter.

Oligos for constructing the miR-155 backbone:

ML123f:

AATTACGCGTTGGAGGCTTGCTGAAGGCTGTATGCTGTTGTCTTCAAGATCTGGAAGACACAGGACAC
AAGGCCT

ML123r:

AATTCGGCCGTTGTCATCCTCCCACGGTGGCCATTTGTTCCATGTGAGTGCTAGTAACAGGCCTTGTGT
CCTGTG

Target sequence specific oligo pairs:

ML124f: GCTGtagttgtactccagcttgtgcGTTTTGGCCACTGACTGACgcacaagctagtacaactac

ML124r: TCCTgtagttgtactagcttgtgcGTCAGTCAGTGGCCAAAACgcacaagctggagtacaacta

ML125f: GCTGcatgatatagacgttgtggctGTTTTGGCCACTGACTGACagccacaacctatatcatgc
ML125r: TCCTGcatgatataggttgtggctGTCAGTCAGTGGCCAAAACagccacaacgtctatatcatg

ML126f: GCTGtgaagaagtcgtgctgcttcaGTTTTGGCCACTGACTGACtgaagcagcgacttcttcac
ML126r: TCCTgtgaagaagtcgctgcttcaGTCAGTCAGTGGCCAAAACtgaagcagcacgacttcttca

ML127f: GCTGtgtagttgtactccagcttgtGTTTTGGCCACTGACTGACacaagctggtacaactacac
ML127r: TCCTgtgtagttgtaccagcttgtGTCAGTCAGTGGCCAAAACacaagctggagtacaactaca

ML128f: GCTGttgaagttcaccttgatgccgGTTTTGGCCACTGACTGACcggcatcagtgaacttcaac
ML128r: TCCTgttgaagttcactgatgccgGTCAGTCAGTGGCCAAAACcggcatcaaggtgaacttcaa

ML124F~ML128r: the sequences shown in upper case are the backbone sequence and the loop sequence; the sequences shown in lower case are the target-specific sequences to different parts of the eGFP mRNA.

5-way ligation primers for constructing the multimers of miR-155:
ML131f: AATT CCTAGG (*Avr*II) TGGAGGCTTGCTGAAGGCTGTATGC
ML131r: AATT TGTACA (*BsrG*I) TTGTCATCCTCCCACGGTGGCCATT
ML132f: AATT TGTACA (*BsrG*I) TGGAGGCTTGCTGAAGGCTGTATGC
ML132r: AATT GAATTC (*EcoR*I) TTGTCATCCTCCCACGGTGGCCATT
ML133f: AATT GAATTC (*EcoR*I) TGGAGGCTTGCTGAAGGCTGTATGC
ML133r: AATT GGATCC (*BamH*I) TTGTCATCCTCCCACGGTGGCCATT
ML134f: AATT GGATCC (*BamH*I) TGGAGGCTTGCTGAAGGCTGTATGC
ML134r: AATT CATATG (*Nde*I) TTGTCATCCTCCCACGGTGGCCATT

The human miR-30 backbone based mirshRNAs

For the human miR-30 backbone-based construction, miR-30 surrounding context sequences were PCR amplified from the plasmid LMP (Dickins et al., 2005) using primers ML47f and ML47r and inserted into pML21 to give pML23. Two sets of oligos ML44f and ML44r and ML91f and ML91r were used to anneal to form the eGFP target-specific miR-30 stem-loop sequences. For miR-30, the 3' strand is the effector strand of the miRNA. The annealed oligos were treated with Klenow fragment 3'-5' exo$^-$ (NEB) to form the full mirshRNA and were subsequently cut with *Xho*I/*EcoR*I to be ligated with pML23 digested with the same enzymes, to give pML23-eGFP (ML44f and ML44r) and pML23-eGFP2 (ML91f and ML91r). The constructs were verified by DNA sequencing. The target-sequence containing miR-30 were PCR amplified with four sets primers containing restriction sites that allowed four-way ligation of three copies of the target-specific mirshRNA into *Avr*II/*Nde*I digested pML29 as before.

Primers

ML47f: AATTCTAGAAGATCTCTCgacTAGGGATAACAGG

ML47r: ATATCTAGATTGAAAAAagtgatttaatttataccattt

Target sequence specific oligo pairs:

ML44f: ACTCTCGAGTGCTGTTGACAGTGAGCGAgcacaagctggagtacaactaTAGTGAAGCCACAG

ML44r: ATTGAATTCCGAGGCAGTAGGCAGgcacaagctggagtacaactaTACATCTGTGGCTTC

ML91f:

AAATCTCGAGAAGGTATATGCTGTTGACAGTGAGCGaagccacaacgtctatatcatgTAGTGAAGCCACAG

ML91r: ATTGAATTCCGAGGCAGTAGGCACagccacaacgtctatatcatGTACATCTGTGGCTTC

ML44f~ML91r: the sequences shown in upper case are the backbone sequence and the loop sequence; the sequences shown in lower case are the target-specific sequences to different parts of the eGFP mRNA.

Construction of mir290C reporters

pML81 Was constructed by inserting three copies of the oligo-annealed ESCC miRNA target sequence and its surrounding sequence (up to 48 bp in total, using 3 sets of oligos, ML151f and ML151r, ML152f and ML152r, ML153f and ML153r) within the 3'UTR of Cdk1na (Wang et al., 2008c) to pML80, which contains the wild type Cdk1na 3'UTR, amplified from BAC RP23-73N16. The target recognition sites were confirmed by sequencing. The 3x recognition sites were further PCR-amplified using two sets of primers (ML182f and Ml182r, ML183f and ML183r) with different restriction site combinations and ligated into the *Xba*I/*Pac*I fragment of the pML82C, between the Neo resistant gene and the rbgpA site to give rise to pML82M, a PBITR flanked Huc-EM7-Neo cassette with 6 copies of the ESCC miRNA recognition sites.

Oligo pairs for ESCC miRNA recognition site construction:

*EcoR*I to *Hind*III

ML151f: AATTCGTGTGATCCTCAGACCTGAATAGCACTTTGGAAAAATGAGTAGGACTTTGA

ML151r: AGCTTCAAAGTCCTACTCATTTTTCCAAAGTGCTATTCAGGTCTGAGGATCACACG

*Hind*III to *Spe*I

ML152f: AGCTTGTGATCCTCAGACCTGAATAGCACTTTGGAAAAATGAGTAGGACTTTGA

ML152r: CTAGTCAAAGTCCTACTCATTTTTCCAAAGTGCTATTCAGGTCTGAGGATCACA

*Spe*I to *Pst*I

ML153f: CTAGTGTGATCCTCAGACCTGAATAGCACTTTGGAAAAATGAGTAGGACTTTGCTGCA

ML153r: GCAAAGTCCTACTCATTTTTCCAAAGTGCTATTCAGGTCTGAGGATCACA

Primers for multimerisation of the ESCC miRNA recognition sites:

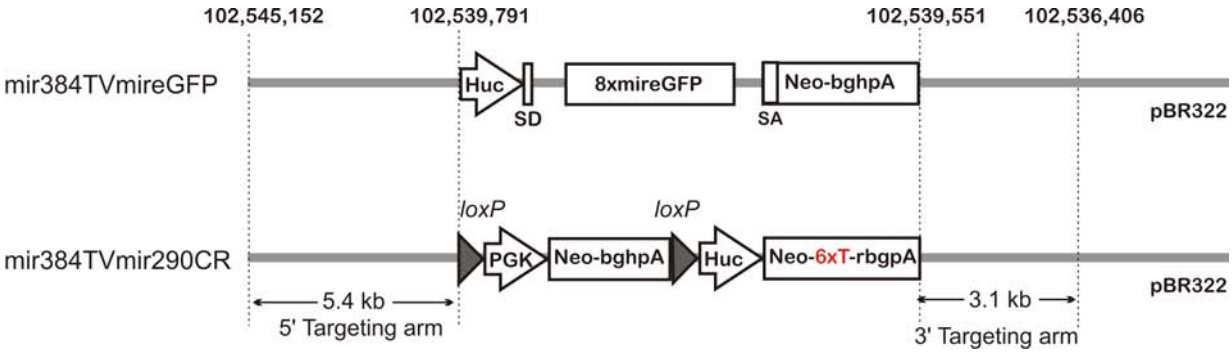
ML182f: AATT TCTAGA (*Xba*I) GTCCCCACTAGGACTTTGGAATTCG

ML182r: AATT CCATGG (*Nco*I) TAGATATGCCTCCTGCCGACCTGCA

ML183f: AATT CCATG (*Nco*I) GTCCCCACTAGGACTTTGGAATTCG

ML183r: AATT TTAATTAA (*Pac*I) TAGATATGCCTCCTGCCGACCTGCA

Construction of targeting vector for miR-384 locus with miR-eGFP (miR-384TVmiR-eGFP) or mir290CR reporter (miR-384TVmir290CR)



The coordinates are genomic fragments used for the targeting arms and are based on NCBI Build 37.

pML88 Was modified by PCR cloning of mini homology arms from miR-384TVPTK (Haydn Prosser) outside of the reporter construct to give pML91. The linear *Kpn*I/*Age*I fragment of pML91 was used to drive homologous recombination of the reporter construct onto the miR-384TV to replace the PGK-EM7-PuroΔtk in the recombineering strain EL350. The correct plasmid DNA was extracted and re-transformed into Stbl2 strain (Invitrogen) to isolate the correct the plasmid from the mixture. The resultant vector is miR-384TVmiR-eGFP. The targeting vector was linearised with *Sbf*I for gene targeting.

The mir290CR reporter, the ESCC miRNA-regulated Neo resistant reporter was modified similarly to miR-eGFP by PCR cloning of mini homology arms into pML82M to give pML93. The *Kpn*I/*Age*I fragment of pML93 was used to conduct recombineering experiments to generate the miR-384TVmir290CR targeting vector. This targeting vector was also linearised with *Sbf*I for gene targeting.

1.4. BAC vectors for giant PB transposons (PB-HPRT-28, PB-HPT-70 and PB-HPRT-100)

The human BAC (RP11-674A04) was obtained from the Wellcome Trust Sanger Institute BAC clone Archives. The *loxP* site on the BAC backbone was replaced by the *EM7-Zeomycin* resistant cassette (gift from Junji Takeda at Osaka University) by recombineering(Lee et al.,

2001). The plasmid containing both PB ITRs was a gift from Xiaozhong Wang at Northwestern University. PL451, PL452 and PL313 are gifts from Pentao Liu at Wellcome Trust Sanger Institute. The PB5'ITR was PCR-amplified using primers Neo-PB5-F and Neo-PB5-R to insert the PB5'ITR 5' upstream of a *loxP*-flanked *PGK-EM7-Neo* cassette in PL452. The PB3'ITR was PCR-amplified with primers Bsd-PB3-F and Bsd-PB3-R to insert PB3'ITR 3' downstream of an EM7 driven Blasticydin (Bsd) cassette in PL313. The PB5'ITR together with the *loxP* flanked *PGK-EM7-Neo* cassette was PCR-amplified with 110 bp chimeric PCR primers (PB3-forward and PB3-reverse) and inserted 10 kb downstream of the *HPRT* stop codon on the BAC by recombineering. The *loxP*-flanked Neomycin cassette was excised by *L*-arabinose induced Cre expression in the bacterial strain EL350(Lee et al., 2001). *PGK-PuroΔtk* was cloned from YTC37(You-Tzung and Allan, 2000) into PL451. A second round of recombineering was carried out to introduce a *PGK-PuroΔTK/Frt-PGK-EM7-Neo-Frt* using chimeric primers ML157f and ML157r to insert it immediately downstream of the PB5'ITR in the BAC. The neomycin resistant cassette was removed by *L*-arabinose induction in the EL250 strain(Lee et al., 2001). The *EM7-Bsd-PB3' ITR* fragment was inserted either 10 or 40 kb upstream of the *HPRT* start codon to generate PB transposons with a cargo capacity of 70 kb (using primers 70kb-F and 70kb-R) and 100 kb (using chimeric primers ML160f and ML160r), respectively. To construct the 28 kb PB vector, a plasmid was made by PCR with primers (ML162fx and ML162r) to introduce a *loxP* flanked *PGK-EM7-Neo* cassette 3' downstream of the human HPRT mini-gene(Ramirez-Solis et al., 1995). The DNA fragment excised from this plasmid containing the exon 3-9 portion of the *hHPRT* mini-gene together with the *loxP* flanked *PGK-EM7-Neo* cassette was used to replace the exon 3 to 9 by recombineering on the 70 kb PB-containing BAC using primers ML163f and 60kb-R. The sequences of all primers are shown below.

Name	Sequence
Neo-PB5-F	5'- CCTCGAGGGGATCCATCGTCTAAAGAACTACCCATTTTA -3'
Neo-PB5-R	5'- GAGCTCCACCGCGGAAAGTTTAGGTCGAGTAAAGCGCAA -3'
Bsd-PB3-F	5'- GTTCTAGAGCGGCCGCCTGAATAATAAAAAAATTAGAAACT -3'
Bsd-PB3-R	5'- GAGCTCCACCGCGGATTTTTGTAAAAGAGAGAATGTTTA -3'
PB3-forward	5'- CAGAATATTTACACAGTCTCAAACTATCATTCCTCGTATTAATTACAAAGGGAAAAATAGTA TGTTTATACTGGAGATCTTCGACCTGCATCCATCTAGATCCAC -3'
PB3-reverse	5'- ATATTAGGAGGCCCGAGATGTCAATTTGTTTCATTAATGATGATGTTAATCTGGATTACTTTG TCAAGACAGCCTCTGCCATTTTTGTAAAAGAGAGAATGTTTA -3'
ML157f	5'- GATTTGCGCTTTACTCGACCTAAACTTTAGTAGCTATGTTTGGACCTCATTTGGATCCCAA TTCAAACAAACTTTAAAATAATTCTACCGGGTAGGGGAGGCGCT -3'
ML157r	5'- AATAATTCCTTAATATCAATGGATATTCAAGTCAGTGTTCAAATCTCCAATTGTCTCATAA GTGTCATAAATGGGTGTGTTTGGATCCATGTACCTGACTGATGA -3'
70kb-F	5'- CTGGTATCTTCAACAAATAAATTGCAGGGAATAAAGACATGATGGAGGAACCTAGACTA AAAGAGACTTAAAGATGAACCCAATTCCGATCATATTCAATAACCC -3'
70kb-R	5'- GTCTCATAAGTGTCATAAATGGGTGTGTATTTTAAAGTTTGTTTGAATTGGGATCCAAATG AGGTCCAAACATAGCTACTAAAGTTTAGGTCGAGTAAAGCGCAA -3'
ML160f	5'- CCAGTAATAACTAGTTAGAAAATGTAAAGGAAAACCGCCTGCCATACCCCGCCAAACACA CATACACTAAAGATGAATTTTCGACCTGCATCCATCTAGATCCAC -3'
ML160r	5'- GCCTGGGTGACAGAGTGAGACCCTGTCTCCCCAACCACACCCCAAAACAAAGCTTTGTGTG TTTTTTTTTTAAATGGAAACACCGCGGATTTTTGTAAAAGAGAG -3'
ML162fx	5'- AATTGGATCCGATCATATTCAATAACCCTTAAT -3'
ML162r	5'- AATTCCGCGGACGATGGATCCCCTCGAGGGACCT -3'
ML163f	5'- TGATAGATCCATTCCTATGACTGTAGA -3'
60kb-R	5'- GGTTCATCTTTAAGTCTCTTTTAGTCTAGGTTCCTCCATCATGTCTTTATTCCCTGCAATTTAT TTGTTGAAGATACCAGCTAGTCGAGCCCCAGCTGGTTCTTT -3'

Table 2-1: primers used to construct giant PB transposons.

2. ES cell lines

2.1. Genotype of the precursor ES cell lines

AB1: 129S7/SvEv$^{Brd\text{-}Hprt+}$, XY (Evans and Kaufman, 1981; Kuehn et al., 1987; Simpson et al., 1997)

AB2.2: 129S7/SvEv$^{Brd\text{-}Hprtb\text{-}m2}$, XY (Kuehn et al., 1987; Simpson et al., 1997)

NN5: AB2.2, *Blm*$^{tm3/tm4}$ (Luo *et al.*, 2000; Guo, 2004)

Jm8.F6: C57BL/6N, XY (Pettitt et al., 2009)

2.2. ES cell lines generated

Table 2-2 summarises all the cell lines generated for this thesis work.

I.D.	Cell line	Locus	Vector	Long range PCR for genotyping	Comments
1	AB1- R26$^{mPBaseERT2/+}$	*Rosa26*	N/A	N/A	Juan Cadinanos generated the ES cell line AB1-R26$^{LSLmPBaseERT2/+}$. Lox-Stop-lox cassette was poped out from this cell line.
2	AB1- Hprt^{PBint2}-R26$^{mPBaseERT2/+}$	*Hprt*	hprtTVPB	5' arm: ML51f and PB5-1: 3.6 kb ML51f: CCCATATCTCACTATAATTCACTGTCA 3' arm: PB3-1 and ML50r: 4.2 kb ML50r: ACTAACAACCCTTTCTCTCAAGGTCTAGTT	Gene targeting of cell line #1. Long range PCR was achieved with LA taq.
3	AB1-R26$^{mPBaseERT2/+}$-Hprt^{M}	*Hprt*	N/A	N/A	6-TG resisitant spontanous hprt mutant of cell line #1 selected with 6-TG.
4	AB1- R26$^{mPBaseERT2/+}$ - Hprt^{M} - Gdf9$^{hprtmimiPB/+}$-	*Gdf9*	Gdf9TVML4	5' arm, ML30f-ML24r: 3.9 kb ML30f: CCAAGTCATGGCACTTCCCAGCAACTTCCT ML24r: GGCGGAGCAGAGGAGGAGGCGGAGG	Gene targeting of cell line #3.
5	NN5- Gdf9$^{hprtminiPB/+}$	*Gdf9*	Gdf9TVML4	Same as cell line #4.	Gene targeting of cell line NN5.
6	NN5- Gdf9$^{eGFP/+}$	*Gdf9*	pML46	3' arm: IresF- ML31r, 4.0kb IresF: ACCCCATTGTATGGGATCTG	Gene targeting of cell line NN5.
7	Jm8.F6-Blm$^{e/e}$	*Blm*	BlmTVML1 BlmTVML2	5' arm, ML108f-ML102rx: 3.6 kb (BlmTV2) ML108f-ML102rx: 5.5 kb (BlmTV1) ML108f: AGGGACTACTTTTCCTCGCCTGTTC ML102rx: AGACCATCACATACACTCCCATTTTT 3' arm, ML103f-IRES-F: 7.8 kb ML103f: AGTCACATTAGGCTGTACTGGAAAG IRES-F: ACCCCATTGTATGGGATCTG	Gene targeting of cell line Jm8.F6 sequentially with BlmTVML2 and BlmTVML1. Subsequently, the floxed PGK-Neo-pA cassette was popped out with Cre transient expression.
8	Jm8.F6-Blm$^{e/e}$- Hprt^{PBint2}	*Hprt*	hprtTVPB	ML48f-ML30r: 3.4 kb ML48f: GCAAATTTGGTTAACTAGGTTCAATGAGAG ML30r: AACCTCGATATACAGACCGATAAAACACAT ML31f-ML48r: 3.3 kb ML48r: GAAAGTTACAGAGCACTAGGGACAGAAAAA ML31f: TTGCGCTTTACTCGACCTAAACTTT	Gene targeting of cell line #7.
9	Jm8.F6-Blm$^{e/e}$- Hprt^{PBex3}	*Hprt*	B6PBhprtE3TV	Kosuke Yusa supplied primers.	Gene targeting of cell line #7.
10	Jm8.F6-Blm$^{e/e}$- miR-384$^{miR-eGFP}$	*miR-384*	miR-384TV-miR-eGFP	ML173f-ML173r: 3.5 kb ML173f: TAGCCAGATTTTTCCTCCTCTCCTG ML173r: TTTCCCAAGGTCACAGAACTGAAAA ML172f-Neo-rev2: 8.3 kb ML172f: AATGTCTCATGTGGCTAATGGCACT Neo-rev2: AGGTCGGTCTTGACAAAAAGAAC	Gene targeting of cell line #7.
11	Jm8.F6-Blm$^{e/e}$- Hprt^{PBex3}-miR-384$^{miR-eGFP}$	*Hprt*	B6PBhprtE3TV	Same as cell line #9.	Gene targeting of cell line #9.
12	Jm8.F6-Blm$^{e/e}$- miR-384$^{miR-eGFP}$ Hprt^{PBint2}	*Hprt*	hprtTVPB	Same as cell line #2.	Gene targeting of cell line #8.
13	Jm8.F6-Blm$^{e/e}$- miR-384^{mir290CR}	*miR-384*	miR-384TV-mir290CRNeo	Same as cell line#10.	Gene targeting of cell line #7.

Table 2-2: A summary of ES cell lines generated for this thesis work.

3. Media and chemicals used for ES cell culture

Phosphate Buffered Saline (PBS): PBS (1x) was prepared in 10 L quantities and stored at room temperature. NaCl (80.0 g), KCl (2.0 g), $Na_2HPO_4.7H_2O$ (10.72 g) and KH_2PO_4 (2.0 g) were dissolved in 10 litres of Milli-Q water. The PH was adjusted to 7.2 with saturated solution of $Na_2HPO_4.7H_2O$. Add phenol red until a peach colour is achieved.

Trypsin: tryspin solution was prepared in 5 L quantities and filter-sterilised and stored at -20°C. NaCl (35.0 g), D-glucose (5.0 g), $Na_2HPO_4.7H_2O$ (0.9 g), KCl (1.85 g), KH_2PO_4 (1.2 g), EDTA (2.0 g), trypsin (12.5 g, Invitrogen, 840-725IL) and Tris Base (15.0 g) were mixed in 5 L MilliQ water and the PH was adjusted to 7.6 with HCl.

β-mercaptoethanol (BME): 10^{-2} M stock solution (100X) was prepared by adding 72 µl of 14 M β-mercaptoethanol to 100 ml PBS and filter-sterilised.

Gelatin: 0.1 % (W/V) gelatin solution was prepared in MilliQ water and sterilized by autoclaving.

Cell staining solution: 2 % (W/V) methylene blue (Sigma) in methanol.

ES cell lysis buffer: 50 mM Tris-HCl [pH 8.0], 5 mM EDTA [pH 8.0], 200 mM NaCl, 1 % [w/v] SDS. Upon cell lysis, a final concentration of 0.1 mg/mL proteinase K (10x stock) was added.

Proteinase K (Roche, 03 115 879 001) stock: 100 mg lyophilised proteinase K was dissolved in 10 ml ddH20, to give 10x stock. The stock was aliquoted and stored in -20°C.

Blasticidin S HCl (Invitrogen, R210-01): 5 mg/ml stock (1000x) was made in PBS. After the powder was dissolved completely, the solution was sterilized by filtering through a 0.2 µm syringe filter and stored at -20°C.

FIAU (1-(2'-deoxy-2'fluoro-β-D-arabinofuranosyl)-5-iodouracil): 200 μM stock (1000x) was made in PBS and 5 M NaOH was added drop-wise until all the powder dissolved. After mixing, the solution was sterilised by filtering through a 0.2 μm syringe filter and stored at -20°C.

G418 (or Geneticin, Invitrogen, 10131): 50 mg/ml solution The working concentration is 180 μg/ml.

Puromycin (C22H29N7O5.2HCl, Sigma): 3 mg/ml stock (1000X) was made in milliQ water, then sterilised and stored at -20°C.

HAT (Invitrogen, 21060-17): 50X liquid stock, containing 5 mM Hypoxanthine, 20 μM Aminopterin and 0.8 mM Thymidine. The stock was stored at -20°C.

HT (Invitrogen, 11067030): 50Xliquid stock, containing 5 mM Hypoxanthine and 0.8 mM Thymidine.

4. AB1, AB2.2 and B6 derived ES cell culture conditions

The culture conditions of AB1, AB2.2 and B6 derivatives have been described in details previously by Ramirez and co-workers (Ramirez-Solis et al., 1993). These cells were maintained on γ-irradiated (60 Gray) or mitomycin C (Sigma) inactivated monolayer SNL 76/7 STO feeder cells. The ES were cultured at 37°C with 5 % CO_2 in M15 medium, which consists of Knockout^{TM} D-MEM (Invitrogen 10829-018) and 15 % foetal bovine serum (FCS, Hyclone), 2 mM L-glutamine (Sigma), 50 units/ml penicillin (Sigma), 40 mg/ml streptomycin (Sigma) and 0.1 mM β-mercaptoethanol (Sigma).

4.1. Passaging, freezing down and thawing ES cells

For passaging, the medium was changed to M15 two hours prior to passage. The plate was washed with PBS twice before the addition of trypsin. The ES cells were trypsinised for 5 min at 37°C before the addition of M15 medium to inactivate the trypsin. The cells were suspended by pipetting the cell-containing medium up and down vigorously. The single cell

suspension was plated again on fresh feeder-containing plate. For freezing down, the ES cells were spun down after resuspension in M15 medium post trypsinisation, and resuspended in freezing medium, which consists of Knockout D-MEM, FCS, and DMSO (7:2:1, v/v). The vials were left in polystyrene boxes or a freezing pot at -80°C for at least 24 hours and transferred to liquid nitrogen subsequently. For thawing ES cells, the cells were taken out of liquid nitrogen and immediately thawed at 37°C. Cells were then spun down and the freezing medium was aspirated. The cells were re-suspended in fresh culture medium and transferred to a feeder-containing plate.

4.2. ES cell transfection by electroporation

70 % - 90 % confluent ES cells were fed 2~4 hours pre-electroporation. The cells were trypsinised as described before and the cell suspension was spun down at 1200 rpm for 3 min at room temperature. The cells were then resuspended in PBS and washed twice and finally counted to make a final concentration of 1.4 x10^7 cell/ml. 0.7 ml of cells were mixed with DNA and transferred into 0.4 cm gap cuvette (Bio-Rad). The gene pulser (Bio-Rad) was configured at 230 V and 500 µF to electroporate the cells. Cells were recovered in M15 medium and plated into 90 mm feeder-containing plates. For gene targeting, 20 µg of linearised targeting vector was used. The next day, drug selection was applied and the drug-containing medium was changed daily to allow ES cell colonies to form. For BAC transfection, 5 µg of BAC DNA purified using Qiagen large construct purification kit (Optional protocol without endonuclease treatment) was used. For PB mobilisation, the conditions were described in a separate section.

4.3. ES cell transfection by lipofection and pulse puromycin selection

Transient transfection of PGK-puro coexpressing PBase and hyPBase plasmids was achieved with Lipofectamine 2000 (Invitrogen). For large scale experiment, 3 x 10^6 cells in suspension were transfected with 24 µg of PBase or control plasmids (CMV-hyPBase-PGK-Puro, CMV-mPBase-PGK-Puro) mixed with 60 µl Lipofectamine 2000 reagent. The transfected cells were plated onto 90 mm feeder-containing tissue culture plate in 3 ml OptiMEM medium. One hour post plating, 7 ml of M15 was added. The next day, fresh M15 was replaced. The

transfection experiment can be scaled down according to Invitrogen's standard protocol. For pulse puromycin enrichment strategy, M15 containing 1 or 1.5 µg/ml puromycin was supplemented 16 hours post lipofection and was sustained for 48 hours.

4.4. Picking ES cell colonies

50 µl of trypsin was added to each well of the 96-well round bottom plate. After washing the 90 mm tissue culture plate with PBS, ES cell colonies were picked using a P20 pipette manually and transferred to the trypsin-containing wells. After completing the full plate, the cells were incubated at 37°C for 15 min. 150 µl of M15 was added per well to inactivate the trypsin. The colonies were broken up into single-cell suspension by continuous pipetting the cell suspension up and down vigorously. The cells were then transferred to a feeder-containing 96-well plate.

4.5. Cre or Flp mediated recombination to pop-out selection cassettes

20 µg of Cre expressing plasmid (pCAGGS-Cre) or Flp expressing plasmid (PGK-FlpO) was electroporated into $1x10^7$ cells. The electroporated cells were serial diluted in M15 and 1, 000 cells were plated onto a 90 mm feeder-containing plate in duplicates. The next day, drug-containing medium was added to one of the plates while the other was grown under M15. Upon eight days selection, the differential colony number reflected how efficient the pop-out reaction had proceeded. According to the ratio, corresponding number of colonies were picked. Usually a third of the colonies contained the pop-out event.

5 *piggyBac* mobilisation in ES cells

5.1. Plasmid-to-genome mobilisation

The condition described was established for obtaining predominantly single-copy *piggyBac* integration per cell using a promoter-driven selection cassette to select for genomic integration events. Both *piggyBac* transposon-containing plasmid and transposase-expressing plasmid were prepared using commercial column based maxi-purification and dissolved in sterile Tris-EDTA buffer (TE). The transposon-containing plasmid was diluted to a final concentration 10 ng/µl, whereas the transposase-expressing plasmid was adjusted to a final

concentration of 1 µg/µl. 100 ng of transposon donor was mixed with 10 µg mPBase expressing plasmid. The DNA was mixed with $1x10^7$ ES cells and electroporated under the condition as described before. One tenth of the electroporated ES cells were plated on a 90 mm feeder-containing plate and the rest on the other. The next day, the drug-containing M15 medium was supplied to the cells. After eight days selection, the plate with a tenth of the electroporated cells was stained using the methylene-based cell staining solution to estimate the mutant complexity. The ES cell colonies could also be picked to extract DNA for *piggyBac* copy number per cell analysis using the PB5'ITR sequence as the probe.

5.2. Intra-genomic *piggyBac* mobilisation

When the transposon is pre-integrated in the genome, transfection of the transposase-containing plasmid mediates the mobilisation of the transposon from the original location to elsewhere in the genome. A selection scheme is necessary to eliminate the cells with the transposon still residing in the donor locus. This can be achieved by engineering the transposon into the *X*-linked *Hprt* locus or a bipartite hprt mini gene that is engineered in any selected genomic location. When the transposon is present in *Hprt*, it disrupts its expression; therefore such cells are sensitive to HAT. Upon excision, the function of Hprt is restored; therefore cells become resistant to HAT. Additionally, a positive selection marker, which is present within the *piggyBac* transposon, can be used to select for transposon re-integration events. The excision efficiency using *Hprt* as the donor locus is measured to be around 0.1% ~ 1 % of total electroporated cells using mPBase (Wang *et al.*, 2008b; Liang *et al.*, 2009).

For mobilising a genomic-residing *piggyBac* transposon, $1x10^7$ ES cells were electroporated with 25 µg transposase (either mPBase or HyPBase), using the ES cell transfection by electroporation protocol as described before. One tenth of the electroporated ES cells were plated on a 90 mm feeder-containing plate and the rest on another. The next day, the drug containing M15 medium was supplied to the cells. After eight days selection, the plate with a tenth of the electroporated cells was stained using the methylene-based cell staining solution to estimate the mutant complexity.

5.3. PB Transposon excision for phenotypic reversion

The *piggyBac* transposon can be removed from the genome in the phenotypic reversion experiments in order to establish the causal link between the genotype and phenotype. The piggyBac transposon used this thesis contains a *PuroΔTK* selection cassette, which enables the negative selection of ES cells with the transposons removed from the genome. $3x10^6$ ES cells were electroporated with 25 μg transposase (either mPBase or HyPBase), using the ES cell transfection by electroporation protocol described before. As a control, $3x10^6$ ES cells were electroporated with 25 μg pBSKSII+ DNA. The cells were maintained in M15 medium for three days after the electroporation, allowing the decay of TK mRNA and protein. At day four post electroporation, the ES cells were trypsinised and $1x10^5$ cells were plated on a fresh feeder-containing 90mm plate. The next day, FIAU selection was commenced and the medium was changed daily for the first six days. At day eight to ten, the colony number between the control and the PBase transfected cells were compared in order to estimate the reversion efficiency. The colonies from the PBase transfected cells were picked for further molecular analysis of the genotype and phenotype assessments.

6. Homozygote mutant generation using *Blm*-deficient ES cells

The heterozygote mutant libraries were generated by the *piggyBac* intra-genomic mobilisation method for obtaining around 1,000 colonies per 90 mm plate using both the HAT and positive selection (Puromycin, 3 μg/ml) to obtain genome-wide *piggyBac* reintegration events as described previously. Twenty independent pools were conducted to give a total complexity of 20,000 mutants. The ES cell colonies from each pool were trypsinised for 15 min at 37°C to obtain a single cell suspension with vigorous pipetting, and plated onto a fresh 90 mm feeder-containing plate. The mutant pools were expanded until the 90 mm plate was confluent, before they were subjected to the phenotypic selection.

The total number of cells required to obtain a few homozygote mutants for each initial heterozygote mutant was calculated based on the LOH rate in *Blm*-deficient ES cells, as shown in the following equation: $N = C \times 10^4 \times m \times 4.2^{-1}$; where N is the total cell number, C is the complexity of the heterozygote mutant library, and m is the number of homozygote mutants

expected. When the mixed mutant pools were confluent on 90 mm plate, it would reach 3 x10^7 cells. Therefore, the total cell numbers for all twenty mutant pools would be 6x10^8 cells, thus the number of homozygote mutants for each initial heterozygote mutant would be around 12. In practice, the heterozygote-to-homozygote conversion is a stochastic process and the LOH rate differs across the lengths of the chromosomes with the LOH rate lower than the measured LOH rate towards centromeric regions. Therefore, the calculation is only a guide to the expansion process.

7. Fluorescent Activated Cell sorting (FACs) of eGFP expressing ES cells

ES cells either transfected with eGFP or having eGFP expressing cassette knocked-in the ES cell genome were analysed live for eGFP expression. ES cells were harvested from wells of the 24-well plate by trypsinisation and were washed twice with PBS. The cells were resuspended in a 1 % (V/V) FCS-containing PBS solution to a final density of 1x10^5 cells/ml. The cells were passed through a cell-strainer-capped round bottom FACs tube (BD Falcon, 352235) to remove of the cell clumps. The tubes were left on ice until FACs analysis. For FACs analysis, either Beckman Coulter FC-500 or BD LSRII FACs machines were used.

A two-parameter dot-plot of Forward Light Scatter (FLS) vs. Side Scatter (SS) was plotted and adjusted to give the best separation of the cells. Another two-parameter dot-plot of FL1 and FLS was plotted, and the voltage of the FL1 channel was adjusted to the best signal to noise ratio for detecting green fluorescent cells. Unless specified, 10,000 ungated events were acquired for each sample. Analysis of the FACs data was conducted using FlowJo software.

8. Cytogenetic analysis

8.1. Preparation of metaphase spread

ES cells were grown in a well of a 24 well plate until 80% - 90 % confluent. Two hours prior to trypsinisation, the medium was changed to M15 with 1mg/ml cocemid (1:100 from the stock solution, Sigma). The cells were trypsinised after cocemid treatment and washed with PBS twice. The cell pellet was then resuspended in the residual PBS and 1 ml of 1 % (w/v) trisodium citrate was added to further resuspend the cells and the solution was incubated at

37°C for 20 min. The solution was cooled down on ice and the ice-chilled fixing agent was added dropwise for the first 5 min. The fixing agent contains methanol and glacial acidic acid (3:1, v/v). Then 7 ml of fixing agent was gradually added and the solution was slowly mixed with a glass pipette. The solution was washed with fixing agent twice and was spun down at 1,000 rpm for 5 min. A final residue 100 µl of fixing agent was left to resuspend the cells and one drop was applied onto a microscope slide (positively charged superfrost or normal glass slide with 100 % ethanol pretreated) which was placed in a pre-wetted paper towel to allow spreading. The slide can be stained with Dapi to visualise the chromosomes using a fluorescent microscope or directly visualised under a light microscope.

8.2. Sister chromatid exchange analysis

Sister chromatid exchanges (SCEs) can be visualised in a proliferating cells with the sister chromatids differentially labelled and stained. When proliferating cells are incubated with the thymidine analogue BrdU (5-bromo-2'-deoxyuridine), BrdU is incorporated into the newly synthesised DNA strand. Upon cell division after the first round of DNA replication, the sister chromatids are equally lableled. In the second round of DNA synthesis and at the metaphase, the chromatid containing the original template strand is darker than the chromatid with both strands of the DNA containing BrdU, arising from two rounds of cell cycle. Therefore, any SCEs can be visualised at this stage. Further cell cycles will accumulate sister chromatids with both strands labelled with BrdU, therefore the ratio of differentially labeled to undifferentially labelled chromatids will significantly reduce. Therefore for SCE analysis, it is important to know the doubling time of the cell line under investigation to in order to optimise the BrdU labelling time.

Freshly thawed BrdU stock solution (2 mM) was diluted in ES cell medium to give a final 10 µM BrdU. The medium was was added to ES cells grown on a well of the 24 well feeder-containing plate. For JM8.F6 ES cells, after 34 hours BrdU incubation, the cells were prepared for metaphase spread as described before. BrdU is light sensitive; therefore the slides were kept in dark to dry. 0.01 % (w/v) Acridine orange staining solution was prepared in Sorensen's buffer (0.06 M, pH 6.5). Sorensen's buffer was made with 2.51 g of KH_2PO_4 and 5.87 g of

Na_2HPO_4 in 1 L of distilled water and the pH was adjusted. The metaphase slide was stained in Acridine orange solution for 5 min in the dark and was rinsed briefly with Sorensen's buffer. The slide was mounted with the same buffer and the excess fluid was blotted. The slide was examined using fluorescence microscope with the green channel. Acridine orange produces a green fluorescence when bound to double-stranded DNA and red fluorescence with single stranded DNA and RNA. Regions of BrdU-labelled DNA produce red fluorescence due to photo-degradation of the BrdU.

9. DNA methods

9.1. Recombineering technology

9.1.1. E. *coli* strains (Lee et al., 2001)

DH10B: F^- endA1 recA1 galE15 galK16 nupG rpsL ΔlacX74 Φ80lacZΔM15 araD139 Δ(ara,leu)7697 mcrA Δ(mrr-hsdRMS-mcrBC) λ^-

EL250: DH10B [λ*cl*857 (*cro–bio*A) <> *ara*C–P_{BAD}*flpe*]

EL350: DH10B [λ*cl*857 (*cro–bio*A) <> *ara*C–P_{BAD}*cre*]

The DH1B strain lacks *recA*, which provides a good host environment for BAC or plasmid containing repetitive sequences. EL250 and EL350 were modified based on DY380 (DH10B [λ*cl*857 (*cro–bio*A), *tet*]), which contains the defective λ prophage. The λ *cl857* repressor provides a temperature regulatable expression. At 32°C, the *P*L operon encoding *exo* and the red recombination genes, *bet* and *gam* are repressed by the *cl*857 repressor. At 42°C, the *cl*857 repressor is inactivated and the *P*L operon can be expressed. Gam inhibits the *E. coli* RecBCD nuclease from attacking the electroporated linear DNA, whereas Exo and Beta mediate the recombination activity. Therefore, incubation of DY380 at 42°C permits the homologous recombination between a linear exogenous DNA and the *E. coli* genomic DNA. EL250 and EL350 were generated based on DY380 by homologous recombination based replacement of the *tet* gene with *araC* and the arabinose-inducible *flpe* and *cre* genes, respectively.

The procedues described below can be used for BAC targeting, retrieving DNA fragment from the BAC, and plasmid fragment replacement.

9.1.2. Transformation of E. *coli* by electroporation

Either DH10B or the recombingenic strains were inoculated o/n from glycerol stock at 32°C in 5 ml LB broth. The next day, 0.5 ml of culture was transferred into 9.5 ml of the low salt LB (per 1 L, 10 g Bacto-tryptine, 5 g Bactor-yeast extract, 5 g NaCl, 1 L ddH_2O, and the pH is adjusted to 7.0 with NaOH, sterilised by autoclaving) and was inoculated until the OD_{600} reached 0.6. The bacteria were cooled down on ice with gentle shaking and harvested by spinning down at 5,000 rpm for 5 min at 4°C. The pellet was then resuspended in 1 ml of ice cold double distilled water and washed twice with ice-cold water by spinning at 9,600 rpm for 30 seconds each time twice in the cold room. The pellet was finally resuspended in 180 µl of ice-cold water and mixed with the 10 ng plasmid or 100 ng BAC from a mini DNA preparation. The mixture was transferred to a pre-chilled 0.1 cm cuvette (BioRad Gene Pulser) and electroporated using a Gene Pulser (Bio-Rad) at 1750 V, 25 µF and 200 Ω. After electroporation, the bacteria were immediately transferred into SOC medium (2 % bacto-tryptine, 0.5 % Bacto-yeast extract, 10mM NaCl, 2.5 mM KCl, 10mM $MgCl_2$, 10 mM $MgSO_4$ and 20 mM Glucose in ddH_2O, pH =7.0, sterilised by autoclaving with Glucose added after autoclaving) and recovered at 32°C for one hour. The recovered culture was then plated on a 10 cm agar plate with the appropriate antibiotics and incubated at 32°C o/n.

9.1.3. DNA fragment preparation for recombineering

DNA fragments for recombineering could be either cut out of a plasmid or generated by PCR using chimeric PCR primers, as homologous sequences over 50 bp are sufficient for recombineering. Longer homologous sequences can improve the efficiency further. For the PCR-based method, each primer in the set was designed to contain 80 bp of sequence homologous to the target BAC or plasmid, followed by a 25 bp sequence identical to the target DNA fragment to be amplified. The KOD polymerase based PCR system (Novagen) was found to be the most reliable system to amplify target DNA fragments with long chimeric primers without the formation of primer dimers. The reaction set-up was as described by the

manufacture's protocol with addition of 5 % (V/V) final concentration of glycerol in difficult-to-amply cases. Usually 20 ng plasmid DNA or 100 ng BAC DNA was used as the template for the PCR reaction. The PCR condition was 98°C 2 min, and followed by 30 cycles of 98°C 20 sec, 68°C 1 min (depending on the target length), and 72°C 5 min. The PCR product was column purified to remove salt and the concentration was estimated by running 2 µl on an agarose gel. For the release of fragment from a plasmid, the restriction digested fragments were separated on agarose gel and the correct fragment was extracted from the gel and column purified to remove the agarose.

9.1.4. Recombineering procedure

Strain EL250 or EL350 was inoculated o/n in 5 ml LB at 32°C. The next day, 0.5 ml of the culture was transferred to 9.5 ml fresh low salt LB and cultivated until the OD_{600} reached 0.6 (approx. 2 hours). The bacterial culture was incubated at 42°C for 15 min to induce the *gam*, *bet* and *exo* expression and was spun down at 4 °C at 5,000 rpm for 5 min. The pellet was re-suspended in 1 ml ice-cold ddH_20. The bacterial culture was further washed with ice-cold ddH_20 in the cold room (4 °C) by spinning it down at 9,800 rpm for 30 sec. The pellet was re-suspended in 200 µl ice-cold ddH_20. 100 ng of the DNA fragment was mixed with the bacteria and the mixture was electroporated using the Gene-pulser as described in the previous section. After one hour recovery in SOC medium at 32°C, one tenth of the culture was plated on a 90 mm agar plate with the appropriate antibiotics, and the rest onto another plate. The plates were incubated at 32°C o/n. The colonies were picked and plasmid DNA was extracted which was subjected to diagnostic analysis either by restriction digestion or junction PCR from the target BAC or plasmid to the recombineering DNA fragment.

9.1.5. Cre or Flp mediated cassette pop-out in E. *coli*

EL250 or EL350 containing a BAC or plasmid with a *FRT* or *loxP* flanked cassette respectively were cultured in 5 ml LB at 32°C o/n with vigorous shaking. The next day, 0.5 ml of culture was transferred into 9.5 ml fresh low salt LB to shake for a further 2 hours or until the OD_{600} reached 0.6 and 100 µl of 10 % (v/v) *L*-arabinose solution was added to the culture to give a final concentration of 0.1 % (v/v) *L*-arabinose. The culture was further shaken for 1 hour and

the bacterial culture was serial diluted to 10^{-5} of the original culture in LB media, and 100 µl was plated on one 90 mm agar plate with the appropriate antibiotics and an equivalent amount on another plate without the antibiotics. The presence of the *loxP* or *FRT* flanked cassette confers the resistance to the antibiotics used; therefore, the antibiotic containing plate could be used to define the background level of bacteria without "pop-out" events. The plates were incubated at 32°C o/n and the next day, colonies could be picked from the plate without the antibiotics for further diagnostic analysis by either PCR or restriction digest of the BAC or plasmid DNA.

9.2. Southern blotting and hybridisation

4 µg of genomic DNA was digested with 20 units of appropriate restriction enzyme over night. The digested DNA was resolved on a 0.8 % agarose gel in TAE buffer either at 45 V for 10 hours or 15V o/n. The gel was rinsed with MilliQ water first and then was placed in denaturing buffer (1 M NaCl and 0.4 N NaOH in MilliQ water) with gentle agitation for 20 min at room temperature. An alkaline based transfer system was set up for o/n transfer as described in the original Southern method with the denaturing buffer as the transfer buffer. A gel tray was placed in a tank containing the transfer buffer. Three layers of the 3MM paper was soaked in the transfer buffer and was layered on top of the gel tray with their ends submerged under the transfer buffer in the tank. These Whatman 3MM paper was act as the liquid supply for the capillary transfer. The agarose gel was placed on top of the Whatman 3MM paper, followed by the Hybond XL (Amersham) membrane, which was presoaked in distilled water before placed on the agarose gel. Three layers of Whatman 3MM paper was placed on the membrane and finally a deck of tissue towels were layered on the assembly to drive the buffer upwards for the DNA transfer to the membrane. Bubbles were eliminated using a plastic pipette to roll over each layer during the assembly. At last a weight was placed on top of the tissue deck. The next day, the membrane was rinsed in the neutralising buffer (0.5 M Tris HCl, 1M NaCl in MilliQ water) for 1 min and was baked at 80°C for at least one hour for cross-linking.

The dried membrane was pre-hybridised in a hybridisation tube with 15 ml hybridisation buffer (Sigma, H7033) supplemented with 150 µl of the denatured Salmon sperm SSDNA (Ambion) at 65°C for 1 hour while the probe was labelled. 50 ng probe DNA was labelled with [γ-^{32}P]dCTP using Prime-It®II Random Primer labelling kit (Stratagene product, Agilent, 300385) for one hour, following the manufacturer's protocol. The labelled probe was filtered to eliminate the unlabeled isotopes using ProbeQuant G-50 column (GE). The probe was denatured by heating at 100°C and snap cooled on ice for 10 min before it was added to the pre-hybridised membrane. Hybridisation was conducted in a rotating oven o/n at 65°C.

The next day, the membrane was rinsed with low stringency wash buffer (2xSSC, 0.1 % (v/v) SDS; 20xSSC buffer stock, 3M NaCl and 300 mM Na_3citrate.2$H_2$0 with pH adjusted to 7.0 with 14N HCl) and washed with this buffer for 10 min at 65°C. According to the signal strength of the membrane, it was further washed with higher stringency wash buffer (1xSSC, 0.1 % SDS or 0.5 x SSC, 0.1 % SDS) until the signal was roughly around 10-20 counts per second on the mini monitor Geiger counter. The membrane was sealed and placed in a cassette with an intensifying screen and an X-ray film was placed in the cassette to be exposed o/n at -80°C.

9.3. Southern blotting probes

Table 2-3: southern blotting probes used in this thesis work.

Probe	Size (bp)	Enzyme(s) to digest gDNA	Expected pattern	Probe construction
Neo	773	N/A	N/A	*Eag*I/*Xba*I digest from plasmid pML4
Blm	437	*EcoR*I/*Nhe*I	WT: 9kb +Neo: 6.2 kb ΔNeo: 7.4 kb	PCR from wild type genomic DNA with: ML147f: TGGACTCAACACAGTGGAGGCCCTT ML147r: GTGAAGGTCAGAGGACAACTTGAAG
Msh6	385	*Spe*I	WT: 4.7 kb Mutant: 6.8 kb	PCR from wild type genomic DNA with: ML72f: GGTGGATGCGATCCTTGTAGGCGCT ML72r: CACGGAGCAGCCTCCCCCTCCTTCC
Gdf9	700	*Nco*I	WT: 6 kb Targeted: 6.5 kb	*Sal*I/*BamH*I digest of pGDF9-P1(Dong et al., 1996).
PB5'ITR	235	N/A	N/A	*EcoR*V/*Nsi*I digestion of pML5.

9.4. Splinkerette PCR

9.4.1. Adaptor preparation

The methodology of the Splinkerette is depicted in Figure2-1. The Splinkerette adaptor was prepared by mixing 150 pmol of HMSpAa and 150 pmol of HMSpBb with 5µl of NEB buffer 2 (10x stock contains 500 mM NaCl, 100 mM Tris-HCl, 100 mM $MgCl_2$, 10 mM Dithiothreitol with pH =7.9 at 25°C) in double-distilled water with a total volume of 100 µl. The mixture was heated at 95°C for 10 min and gradually cooled down to room temperature, allowing the oligos to anneal. The adaptors can then be stored at -20°C. The HMSpBb oligo can be designed with different restriction enzyme recognition sequences to suit the need. The 4 bp cutter *Sau*3AI was used in this project. There primers were not phosphorylated at the 5'.

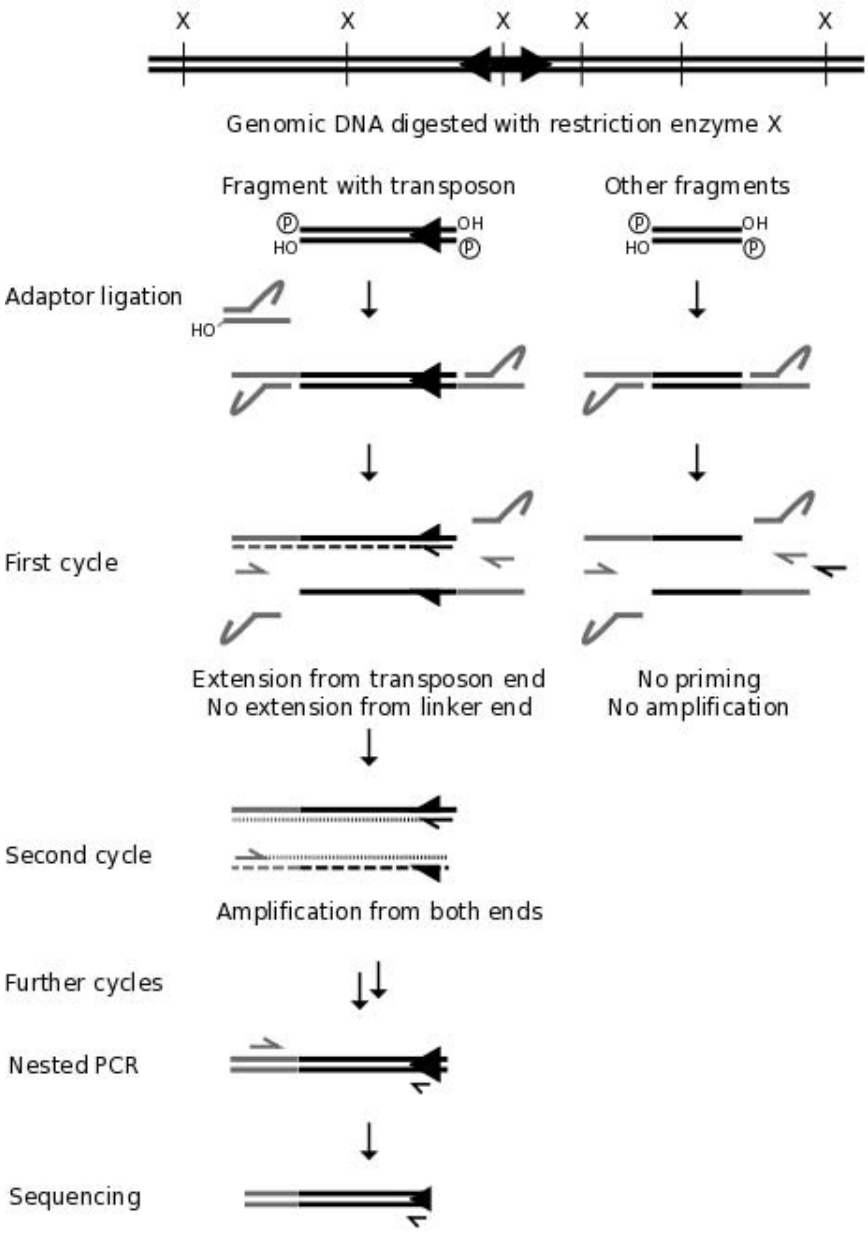
HMSpAa:

5'CGAAGAGTAACCGTTGCTAGGAGAGACCGTGGCTGAATGAGACTGGTGTCGACACTAGTGG

HMSpBb-*Sau*3AI:

5'gatcCCACTAGTGTCGACACCAGTCTCTAATTTTTTTTTTCAAAAAAA

Figure 2-1: Schematic representation of the Splinkerette -PCR method.



A fragment of genomic DNA containing a *piggyBac* inverted terminal repeat (black double arrow) is shown. After digestion (cleavage sites marked with X), splinkerette adaptors (shown in grey) are ligated to all fragments. If the fragment contains the transposon end (left), the transposon-specific primer (black arrow) can hybridize and extend in the first round. Extension into the long-strand of the adaptor provides the template for the adaptor primer to anneal and extend. In the subsequent cycles of PCR, the transposon-specific and adaptor primers can amplify the transposon-genomic junction. A nested PCR is used to improve specificity. This scheme is adapted from review (Li et al., 2010).

9.4.2. Genomic DNA digestion and adaptor ligation

1-2 µg of genomic DNA was digested with 4 units of *Sau*3AI (NEB) in a 30 µl reaction at 37°C for 3 hours and the enzyme was heat-inactivated at 65°C for 20 min. 5µl of digested genomic DNA was mixed with 3 µl adaptor and 4 units of T4 ligase (NEB) in a 20 µl reaction and the ligation was carried out at 16°C overnight. The next day, the ligase was heat-inactivated at 65°C for 15 min. The ligation mixture was diluted ten-fold for the PCR amplification.

9.4.3. Nested-PCR amplification

A *Taq* polymerase-based Platmium®PCR supermix (Invitrogen) was used for both rounds of the PCR. The first round PCR was set with 0.5 µl of the diluted ligation mixture, o.5 µl of the 10 µM adaptor specific primer SP1 and 0.5 µl of the 10 µM *piggyBac* transposon specific primers PB5-1 or PB3-1 (corresponding to the PB5'ITR and PB3'ITR) and 18 µl of the supermix. The PCR condition was 94°C for 2 min, followed by 2 cycles of 94°C 20 sec, 68°C 30 sec, 72°C 1min; then followed by 30 cycles of 94°C 20 sec, 60°C 30 sec, 72°C 1 min; and a final extension of 72°C 5 min.

The second round of PCR was set up similar to the first round with 1 µl of the first round PCR reaction as the template, and 0.5 µl of the 10 µM nested-adaptor primer SP2 and 0.5 µl of the 10 µM nested *piggyBac* specific primer PB5-2 or PB3-2 (corresponding to the PB3'ITR and PB5'ITR) were used. The PCR condition was 94°C for 2 min, followed by 30 cycles of 94°C 20 sec, 60°C 30 sec, 72°C 1 min; and a final extension of 72°C 5 min. 2 µl of the reaction mixed was used to run on a 1 % agarose gel to assess the PCR products. When a single product was generated, the PCR reaction mix was diluted in water and was subjected directly to capillary sequencing. When multiple products appeared, the different products were resolved by gel electrophoresis and the PCR products were extracted individually by gel extraction before capillary sequencing. The sequencing primers were further nested *piggyBac* transposon ITR specific primers PB5-seq and PB3-seq.

Primer sequences for nested PCR and sequencing:

Adaptor-specific primers:

HMSp1: 5'CGAAGAGTAACCGTTGCTAGGAGAGACC

HMSp2: 5'GTGGCTGAATGAGACTGGTGTCGAC

PB5'ITR-sepcific primers:

PB5-1: 5'CAAAATCAGTGACACTTACCGCATTGACAA

PB5-2: 5'CTTACCGCATTGACAAGCACGCCTCACGGG

PB5-seq: 5'TTAGAAAGAGAGAGCAATATTTCAAGAATG

PB3'ITR-sepcific primers:

PB3-1: 5'TAAATAAACCTCGATATACAGACCGATAAA

PB3-2: 5'ATATACAGACCGATAAAACACATGCGTCAA

PB3-seq: 5'TTTTACGCATGATTATCTTTAACGTACGTC

9.5. Genomic DNA triple-primer competition PCR

To confirm the homozygosity of the mutant alleles, triple-primer PCR reactions were conducted on a locus-specific basis, based on determination of the mutagen integration site by the Splinkerette PCR. Two of the primers are locus specific and can amplify a wild-type product that spans the transposon insertion site. The third primer is mutagen-specific (primers in the PBITRs are used) and amplifies the mutagen-host junction together with one of the locus-specific primers. These three primers are used in equimolar amounts in a conventional PCR reaction set-up.

Msh6 locus:

ML71f: 5'TTTTGCCTCCGTGTATGTATGTGTG

ML71r-2: 5'CAGCAAGTTTTAGTACCTGGGTAAA

Wild type allele detection: ML71f + ML71r-2: 860bp

Mutant allele detection: Ml71r-2 + PB5-1: 585bp

Ago2 locus:

Ago2F: 5'GCAGCCCCAGCCTCTTTACTTGT

Ago2R: 5'GAAAGCTCTGCCAACAGCTGGAAC

Wild type allele detection: Ago2F + Ago2R: 224bp

Mutant allele detection: PB5seq + Ago2R: 163bp

Primers for locus specific PCRs in Chapter 7

Actin locus (control):

Actin-F: 5'GTTTGGACAAAGACCCAGAGG

Actin-R: 5'CTCTCTCGTGGCTAGTACCTCAC

Giant PB integrated clones:

Bf6f: 5'GGGCCTCTTAAGATACTTTTAGCTG (for clone Bf6)

Bf6r: 5'AACCCAATTTCAAAACAATTTATCA (for clone Bf6)

Cd8f: 5'TTATGGCCAAGGACATATACAAGTT (for clone Cd8)

Cd8r*: 5'AACTATGTAAACGATAACACGAAACG (for clone Cd8)

Ch2f: 5'GTACACAAGAGCTGATTCCATCC (for clone Ch2)

Ch2r*: 5'CTCCTTTGGCTGTTAGATGAACTTA (for clone Ch2)

Dc7f: 5'GACTTCTACCTGTTTGGTTGTGTTT (for clone Dc7)

Dc7r*: 5'AAAAGCTAGACCCGGATTTAAAAG (for clone Dc7)

Dc11f: 5'TGAGCTGACTAACAGAACAGAGTCA (for clone Dc11)

Dc11r*: 5'CCTAAACACACAAAACACGTAACAG (for clone Dc11)

* primers were used with PB3-seq to PCR amplify the PB5 ITR to genomic junction for each individual ES cell lines.

10. RNA method: RT-PCR

Total RNA was extracted from ES cells grown in a well of a 24-well gelatinised plate, either using a RNeasy Mini extraction kit (Qiagen) or using Trizol reagent with chloroform extraction (Invitrogen), following the manufacturer's protocol. The concentration of the RNA was determined with a NanoDrop ND-1000 machine by measuring the absorbance at 260 nm. 1 µg of the RNA was mixed with 1 µl of either random hexamer primers (10 µM stock) or an oligo-dT primer (10 µM stock) in 10 µl DEPC water (Ambion) to give a final volume of 12 µl. The mixture was heated at 100°C for 2 min and snap-cooled on ice for 2 min. The reverse transcription reaction was set up with a Superscript^{TM}II cDNA synthesis kit (Invitrogen), by

adding 1 µl 10 mM dNTP, 4 µl of 5x synthesis buffer (250 mM Tris-HCl, 375 mM KCl, 15 mM $MgCl_2$, pH 8.3), 2 µl 0.1 M DTT and 1 µl (200 U) of Supercript^{TM}II reverse transcriptase to the RNA and oligonucleotide mixture. The reverse transcription was conducted at 42°C for 1 hour and was followed by heat-inactivation at 70°C for 15 min. The cDNA reaction mixture was diluted ten-fold for use as the PCR template.

The PCR reaction was set up with 1 µl of the tenfold diluted cDNA mixture, 1 µl of gene specific primer mix (10 µM stock) and 18 µl of PCR supermix (Platmium® with anti-taq antibody in the mixture to permit ambient temperature PCR-setup, Invitrogen). The PCR condition was 94°C for 2 min, followed by 30 cycles of 94°C 20 sec, 60°C 30 sec, 72°C 1 min; and a final extension of 72°C 5 min.

Primer sequences

Oligo-dT primer: 5'-GGC CAC GCG TCG ACT AGT AC $(T)_{17}$-3'

Random hexamer primer: Invitrogen, cat no. 48190011

Gene specific primers:

Msh6

Msh6Ex1F: 5'CTTGTACAGCTTCTTCCCCAAGTCC

Msh6Ex3R: 5'TTACTTAAGGCCTCATCTGCACGTT

Ago2

Ago2Ex1F: 5'CTCGGCAACGCCACCATGTACT

Ago2Ex4R: 5'CGACACCCACTTGATGGATACCTTA

Ago2Ex13F: 5'CCACAGACCCTATCCAATCTCTGCT

Ago2Ex16R: TGATAGTCCTTCTCCAGCTTGATGC

Dgcr8

Dgcr8ex2F: 5'GTGGATGAAGAGGCCTTGAATTTCT

Dgcr8ex5R: 5'TTGTTCTCGTTCCTCATTGTCCTTC

Dgcr8ex9F: 5'CCTGGAAATTCTCATCCCTGACTTT

Dgcr8ex12R: 5'ACATGCGTAGTAAGGAACCCCAGTT

Hprt

Hprt-Ex1F: 5'TCAGACCGCTTTTTGCCGCG

Hprt-Ex3R 5'TTATAGCCCCCCTTGAGCACACAG

Hprt-Ex7F 5'GCTGGTGAAAAGGACCTCT

Hprt-Ex9R 5'CACAGGACTAGAACACCTGC

*Ccdc7**

Ccdc107-rev-1: 5'GGGACTCTTTTTACTAAACCTTCAGTGGTT

Ccdc107-rev-2: 5'TTTAGTTTTGTGTCTAGAAGTTCCCTCTGG

*Dom3z**

Dom3z-rev-1: 5'GAGACCCACTTAGAAACACGCCTTTATTAT

Dom3z-rev-2: 5'CTAAGGAAGGCAGCACAGAAGTTCAT

*: Mutagenic exons used for gene inactivation. Rev-1 is from the last exon and rev-2 is from the penultimate exon of the mutagens. The reverse primers were paired with locus- specific primers for detecting trapping events, and triple-primer genomic PCR was used to assess homozygosity.

Table 2-4: Locus-specific primers used for the random trapping assay.

Gene	Chr	Mutagen	Location	Locus-specific primer
Tmem131	1	*Ccdc107*	Intron 4	d7-Tmem131-ex2: 5'TTCAGTCGGAGAGCATAATAGAAGT
Dut	2	*Ccdc107*	Intron 4	g10-Dut-ex1: 5'TGCCGGCTACGACCTATTCAG
Fstl4	11	*Dom3z*	2nd last intron	e8-Fstl4-ex12: 5'AGGTCCAGAGACGCTTGAAACCTAC
A24Rik	15	*Dom3z*	Intron 17-18	c4-A24Rik-ex15: 5'GGTGGCTACTTTCACCATCAACATC
Sfrs3	17	*Ccdc107*	Intron 2	e4-Sfrs3-ex1: 5'ACCGAGAATCTGTAGGAGCAGAACC
Sema6a	18	*Dom3z*	Intron 1	f8-Sema6a-ex1: 5'TTTCTTGAGCATTTACCCTGGTCCT
Undcd3	11	*Ccdc107*	Intron 1	h8-Undcd3-ex1: 5'CTTTATGACCAGGCCCTGTTGG
Stk22s1	7	*Dom3z*	Intron 6-7	a8-stk22s1-ex6: 5'GAAGCTAAGATACCTCCAGCAGCAA
Ran	5	*Dom3z*	Intron 4	f2-Ran-ex1: 5'GACAGGCGCGGAGACTCTTCT
Tmem50a	4	*Ccdc107*	Intron 1	b3-Tmem50a-ex1: 5'GGCTGTTTTGTTTTCTTGCAGAACT

These primers were used in conjunction with *CCdc7* and *Dom3z* reverse primers to confirm trapping events.

11. Protein methods: Western blotting

11.1. Whole-cell protein extraction

ES cells were trypsinised from a gelatinised plate and washed with PBS twice. The cell number was determined and $1x10^5$ cells were used per lane. Usually $1x10^6$ cells were resuspended in 50 µl of PBS. A 2x sample buffer was prepared with 25 µl 4x NuPAGE® LDS sample buffer (Invitrogen, the stock contains 40 % glycerol, 4 % lithium docecyl sulfate, 0.8 M triethanolamine-Cl pH 7.6, 4 % Ficoll-400, 0.025 % phenol red, 0.025 % brilliant blue G250 and 2 mM EDTA-disodium) supplemented with 5 µl of 14 M β-mercaptoethanol and 20 µl PBS. 50 µl of 2x sample buffer was mixed with the cell suspension and the mixture was heated at 95°C for 5 min and vigorously vortexed for 5 min for whole cell lysis.

11.2. Protein blotting and antibody hybridisation

10 µl of the sample was loaded in each lane of the pre-cast 4 – 12 % Bis-Tris gel (Invitrogen) and a tenfold diluted sample was loaded for detecting β-actin as the loading control. The samples were resolved by running using 1x SDS MOPS running buffer (Invitrogen, 20x stock contains, 0.6M MOPS, 1.2 M Tris, 2 % SDS, and 50 mM sodium bisulfite) at 140 V at room temperature. The protein was transferred to Hybond ECL membrane (Amersham) using the Bio-rad transfer system with chilled transfer buffer (100 ml methanol, 50 ml 20x transfer buffer (Invitrogen), the rest MilliQ water to make up 1 L of the 1x transfer buffer) at 4°C with 90 V for a least one hour.

After transfer, the membrane was soaked with 2.5 % (w/v) milk powder in 1x TBST buffer (blocking buffer) for 30 min at room temperature with gentle shaking. TBST was made by adding 100 ml of 10x TBS (500 mM Tris.HCl, pH 7.4 and 1500 mM NaCl), 0.5 ml Tween 20 to up to 1 L of MIlliQ water. The primary antibody was diluted in 1 ml of 0.5x blocking buffer (Table 2-5 shows the dilutions of each antibody used) and used to cover the protein containing site of the membrane, and covered with parafilm. The membrane was incubated with the primary antibody at 4°C o/n. The next day, the membrane was washed with TBST for 30 min with gentle shaking twice. The horseradish peroxidase (HRP)-conjugated secondary antibody (1 µl) was diluted in 1 ml of 0.5x blocking buffer and incubated with the membrane

for 1 hour at room temperature. The membrane was then washed as before. The chemilluminescence (ECL) detection mix (Amersham) was applied to the membrane and left at room temperature for 1 min before the ECL film (Amersham) was exposed to the membrane. The exposure time varied depending on the strength of the signal, typically 10 sec, 30 sec, 1 min and 5 min exposures were used to give the best signal.

11.3. Antibodies used in this thesis

Table 2-5: Antibodies used in this thesis work.

Name	Type	Species	Clonal	Dilution	Source	Cat. No.
Anti-eGFP	Primary	Mouse	Mono	1: 1000	Roche	11 814 460 001
Anti-Dgcr8	Primary	Rabbit	Poly	1: 300	ProteinTech Group	10996-1-AP
Anti-Hprt	Primary	Rabbit	Poly	1: 500	Abcam	Ab10479-200
Anti-Blm*	Primary	Rabbit	Poly	1: 500	Abcam	Ab476-100
Anti-β actin	Primary	Mouse	Mono	1: 1000	Sigma	A5441
HRP conjugated anti-mouse IgG	Secondary	Horse	Mono	1: 1000	Cell Signalling	# 7076
HRP conjugated anti-rabbit IgG	Secondary	Horse	Mono	1: 1000	Cell Signalling	# 7074

* Blm detection only worked with homemade polyacrylamide gel, not the pre-casted gels from Invitrogen.

12. Regional high density Genomic comparative hybridisation (CGH) array

Genomic comparative hybridisation experiment was conducted in Chapter Seven to detect the copy number gain after PB-mediated cargo integration. A customer regional CGH array was designed to serve this purpose. The 230 kb human genomic region Chr*X*: 133,358,379-133,591,045 (hg18), covering the whole BAC (RP11-674A04) was used to design the hybridization probes for an Agilent regional CGH array (8x15K), with the criteria that the probes must pass a similarity score filter to exclude probes with secondary genomic alignments and exclusion of repetitive genomic regions. Additional criteria were adopted to avoid mouse-human cross species hybridization. The rules were: a, reject probes that have more than 90 % identity to the mouse genome; b, reject probes which have 20 bp or more of uninterrupted sequence matching to the mouse genome. In total, 1773 probes were selected from this region to provide an average detection resolution of 130 bp, and they were printed

in triplicate on the array. The remaining 9,600 probes were a random selection of probes from Agilent catalogue mouse CGH HD probes to provide the baseline normalization.

The DNA was extracted using Puregene kit (Qiagen). The DNA with the large-cargo PB integrated was compared to the DNA extracted from AB2.2. Both samples and the control were mixed in equal amount with pooled genomic DNA from human male primary cell lines. In this array, within the high density human probe region, the copy number increase of one on the Log_2 scale represents the gain of an extra copy. The raw array data was normalized using a robust cubic spline interpolation method contained inside the R® package aCGH.Spline (http://cran.r-project.org/web/packages/aCGH.Spline/index.html) to adjust for dye biases. A custom wavelet transform was applied to remove the presence of genomic waves and the true baseline was estimated using the median value reported by the 9,600 randomly selected probes.

The detailed information on the array design and the experimental procedures as well as the raw data was deposited on ArrayExpress (www.ebi.ac.uk/aerep/login). The ArrayExpress accession for the array design is A-MEXP-1849 and the accession for the experimental data is E-MEXP-2788. The login details are as follows: username, A-MEXP-1849; password, 1276737775749.

13. Illunima sequencing for PB integration sites analysis

For giant PB transposon integration analysis described in Chapter Seven, Illumina sequencing was conducted. The sequencing work was conducted by Daniel J. Turner and Sabine Eckert in the Next-generation sequencing R&D department at the Sanger Institute. Genomic DNA was extracted from pooled ES cells as described before. The DNA was sheared acoustically, and paired-end Illumina shotgun libraries were prepared, up to the point of adapter ligation. Ligated libraries were quantified by qPCR, relative to a concentration standard, using primers Ad_T_qPCR1 and Ad_B_qPCR2, which annealed to the Illumina adapters, allowing equimolar amounts of each library to be used in a semi-specific enrichment PCR. qPCR reactions were set up using a SybrGreen system following the manufacturer's recommendation (Applied

Biosystems). Approximately 750 ng of ligated library was used in the subsequent enrichment PCR. Enrichment PCR was optimized by performing a gradient PCR to find an appropriate annealing temperature, and by qPCR to establish the minimum number of cycles of amplification. In the first round of PCR, primer PB5pr_1 that was specific for the PB5'ITR and an adaptor specific primer PCR_V4, which was tailed with an Illumina-compatible sequence(Lander et al., 2001), were used. Phusion enzyme was used for the PCR and the cycling condition is as follows: 1 cycle of 94°C 2 minutes; 18 cycles of 94°C 20 seconds, 62°C 20 seconds, 72°C 40 seconds; and 1 cycle of 72°C 10 minutes. In the nested PCR round, 5 µl of first round PCR was used as the template. A PB5'ITR specific primer PB5prP5_2 was used along with primer PCR_V4.

Index sequences were designed to be error-correcting, by the use of Hamming codes(Hamming, 1950; Mamanova et al., 2010), and they were included in primer PCR_V4. The cycling condition was the same as before except that an annealing temperature of 60°C was used and 12 cycles were performed. Following PCR, reactions were cleaned up using a spin column, and were run in a 2 % agarose gel. Gel slices corresponding to a fragment size of 250-350 bp were cut, and the DNA was extracted without heating (Waterston et al., 2002). Following qPCR quantification, libraries were amplified onto an Illumina paired-end flowcell, following the manufacturer's recommended conditions, and flowcells were sequenced, using customized PB5-ITR sequencing primer PB5prSeqR1 for read 1, TranSeqR2 for read 2 and an indexing sequencing primer RInv4 to decode the barcode of each library. The sequences of all primers in this section are shown in Table 2-6.

Table 2-6: Primers used for the multiplex Illumina sequencing to identify PB integration sites.

Name	Sequence
Ad_T_qPCR1 (Sigma, desalted)	5'- CTTTCCCTACACGACGCTCTTC -3'
Ad_B_qPCR2 (Sigma, desalted)	5'- ATTCCTGCTGAACCGCTCTTC -3'
PB5pr_1	5'- GACGGATTCGCGCTATTTAGAAAGAGAG -3'
Primer PE_PCR_V4	5'- CAAGCAGAAGACGGCATACGAGATCGGT [INDEX]
(Index has below 6 versions)	ACACTCTTTCCCTACACGACGCTCTTCCGATCT -3'
Index V4.1	5'- ACCTAG -3'
Index V4.2	5'- ACTGCT -3'
Index V4.3	5'- ATAGTG -3'
Index V4.4	5'- ATCAGA -3'
Index V4.5	5'- CAAGGT -3'
Index V4.6	5'- CCAACA -3'
PB5prP5_2	5'- AATGATACGGCGACCACCGAGATCTACACATGCGTCAATTTT ACGCAGACTATC -3'
PB5prSeqR1	5'- ATGCGTCAATTTTACGCAGACTATCTTTC -3'
TranSeqR2	5'- ACACTCTTTCCCTACACGACGCTCTTCCGATCT -3'
RInv4	5'- AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT -3'

14. Bioinformatic analysis of Illumina sequencing data

The Illumina sequencing analysis was conducted by Zeming Ning from the Sanger Institute. One lane of Illumina pair-end reads of 2 x 70 bp was used to sequence all six sample libraries in a multiplexing manner. Raw sequences were grouped into six bins based on the barcode sequences incorporated in primer PE_PCR_V4. Each bin was analyzed separately. Read 1 sequences that contained the end of the PB5' ITR joining directly to the BAC vector sequence, representing the BAC random integration events, were excluded. Duplicated sequences are generated in Illumina sequencing by the PCR reactions. Duplicate reads was defined as sequences that possess identical start and end mapping locations on the reference mouse genome. The code ssaha_duply (ftp://ftp.sanger.ac.uk/pub/zn1/transposon/codes/) was developed to detect and remove duplicated segments from the raw read files. It combines two paired-end read files into one file, while two paired-ends are merged into one sequence.

This program then sorts the merged sequences to check the occurrence number and only one copy is maintained while other copies are removed.

SSAHA2 (http://www.sanger.ac.uk/resources/software/ssaha2/) was used to map paired-end Illumina reads against the mouse reference sequence NCBI_M37.fa. Before assembly, possible chimeric reads caused by ligation of two or more genomic sequences was filtered. For a true integration event, a genomic sequence joins immediately downstream of the PB5'ITR, and the genomic sequence should start with TTAA. The condition was set so that the query start matching point must be base position 6 or 5 of the read 1 (in genomic sequence, it is possible to have GTTAA just by chance). A full read match contains read lengths bigger or equal to 65 bp. The Cleaned and uniquely mapped reads were assembled using the pile-up utility of ssaha_pileup (ftp://ftp.sanger.ac.uk/pub/zn1/ssaha_pileup) to extract information, such as read coverage.

For a clean and contamination-free library, the results from the pileup file are reliable and accurate. However, cross-contamination can occur during the sample preparation steps prior to the PCR incorporation or during the barcode-primer synthesis and purification at source. Such contaminations can be directly reflected as cross-contamination in the sequences. Therefore, cross-library check for identical reads after PCR duplication filtering is an important step for the quality control.