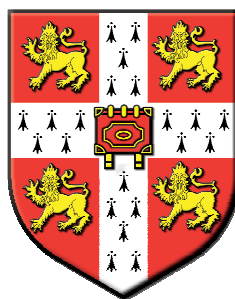


**A recessive genetic screen to discover components in miRNA  
pathways using *piggyBac*-mediated insertional mutagenesis in  
*Blm*-deficient mouse embryonic stem cells**

**Meng Li**

**A dissertation submitted in partial fulfilment of the requirements for the award of  
the degree of**

**Doctor of Philosophy  
of the  
University of Cambridge**



The Wellcome Trust Sanger Institute  
Sidney Sussex College, Cambridge

## **Declarations**

This thesis is submitted on partial fulfilment of the requirements for the degree of Doctor of Philosophy of the University of Cambridge. It describes the work carried out at the Wellcome Trust Sanger Institute from October 2006 to August 2010. Unless otherwise indicated, the research is my own and not the product of collaboration.

This thesis does not exceed the word limit of 60,000 as set by the Degree Committee for the Faculty of Biology.

Meng Li

25<sup>th</sup> August 2010

## Acknowledgements

My foremost gratitude goes to Prof. Allan Bradley for being an inspirational mentor with his broad vision to science. For the past four years, he has provided me with guidance, support and patience, helping me through the difficulties and frustrations. I am aware of the fact that without his open-minded approach to science and his encouragement, I would not be able to have the freedom to pursue my own ideas and build up confidence in conducting independent research.

Many thanks to the Bradley group, current members and alumni, for being great people to work with, and always being there for help and advice. Special acknowledgement goes to Kosuke Yusa, for making his resources available to me and from whom I have learnt a great deal and had so many instructive discussions. Also thanks to Qi Liang, for introducing me to many basic techniques and providing me with much of the starting materials to start my projects. Thanks to Stephen Pettitt, Yue Huang and Wei Wang (Mr.), who are also working on *piggyBac* and *Blm*-deficient ES cell systems, for all the instructive discussions and sharing of resources. I am very grateful to Frances Law, Alastair Beasley and James Cooper for tissue culture support; Haydn Prosser and Hiroko Yusa for providing me with useful plasmids; Roland Rad and George Vassiliou for useful discussions and having laughs with. I would like also to thank many great people from other labs in the Sanger Institute; Daniel Turner and Sabine Eckert for their expertise in conducting Illumina sequencing of my samples, Zeming Ning for helping me with bioinformatics analysis and Tomas Fitzgerald for array CGH analysis.

Thanks to my thesis committee members Julie Ahringer, Bill Skarnes, and David Adams for their comments and advice throughout my PhD. Thanks too to Seth Grant, Jyoti Choudhary and Jürg Bähler for hosting me in their labs during my first PhD year rotations and to Mark Colin and Marcelo Coba for guiding me through the rotations.

Robert, you are always here for me and this work would not be as it is without your constant support. I look forward to our life together as husband and wife. Thank you for proof-reading of this thesis too.

Mum and Dad, thank you so much for all your love and support emotionally and financially in whatever I chose to do and always being there for me even you are thousands of miles away in China when I am faced with difficult decisions and feeling sad.

Thanks to my future in-laws, Andy and Julie Phipps for being so welcoming, generous and treating me as a true family member for the past six years.

**Abstract**

Genome-wide recessive genetic screens complement reverse genetic approaches to ascribe gene function in biological pathways of interest without prior knowledge. The ability to conduct such screens in cultured embryonic stem cells requires a tractable mutagenic system combined with a *Blm*-deficient background to efficiently generate and convert heterozygous mutations to homozygosity in parallel on a genome-wide scale.

The first part of this thesis describes the establishment of a novel mutagenic strategy based on remobilisation of a single copy *piggyBac* transposon targeted within the genome. This strategy has the significant advantage over conventional co-transfection methods for transposon delivery in their ease of mutant library construction while maintaining a single copy of the mutagen per cell for the subsequent establishment of the genotype-phenotype causality.

The second part of this thesis concerns the development of reporter systems to conduct genetic screens using the established mutant pools, for the identification of novel factors in the miRNA biogenesis pathways and dissecting possible differential regulators in the two branches of the miRNA downstream effector pathways, i.e. miRNA mediated mRNA decay and translational repression. To our knowledge, this is the first attempt to use non-hypothesis driven genetic approach to identify novel components in this pathway in mammals. Preliminary screening with one of the reporter system has revealed a homozygous mutant in a known effector in the miRNA-mediated repression, *Ago2*.

A final part of this thesis presents a separate part of the research during the PhD to advance the *piggyBac* transposon technology in large genomic DNA delivery. This work has demonstrated a giant cargo capacity of up to 100 kb for *piggyBac* transposons. The integrations of giant *piggyBac* transposons are intact, they can be expressed stably and can be remobilised from the genome. Giant *piggyBac* transposons open new doors to many applications in basic mammalian genetics and gene therapy, which are not possible with existing methods.

**Contents**

Abbreviations .....	xi
List of figures .....	xiii
List of tables .....	xvi

**Chapter 1 – Introduction**

Chapter overview .....	1
1. Reverse genetics.....	2
1.1. Homologous recombination based gene targeting.....	2
1.2. Reverse genetics using RNA interference.....	3
1.1.3. Zinc finger nuclease mediated genetic alternations .....	5
2. Forward genetics and different screen designs .....	6
2.1. Germline and somatic mutagenesis .....	7
2.2. Dominant, recessive and genetic interaction screens.....	8
2.2.1. Dominant genetic screens.....	8
2.2.2. Recessive genetic screens .....	9
2.2.2.1. Germline recessive mutagenesis .....	9
2.2.2.2. Mitotic recombination mediated recessive mutagenesis in the soma .....	11
2.2.3. Genetic-interaction screens .....	14
3. Means of mutagenesis for forward genetics screens .....	17
3.1. Chemical agents.....	17
3.2. Physical agents.....	19
3.3. Biological insertional agents.....	19
3.3.1. Retroviral vectors .....	19
3.3.2. DNA transposons.....	20
3.3.2.1. <i>Sleeping Beauty</i> and <i>piggyBac</i> possess different characteristics .....	22
3.3.2.2. Transposon-mediated germline mutagenesis in mice .....	24
3.3.2.3. Transposon-mediated somatic mutagenesis for cancer gene discovery .....	26
3.4. Insertional mutagen designs .....	29

---

3.4.1. Loss-of-function designs.....	29
3.4.1.1. Promoter-trap designs.....	29
3.4.1.2. PolyA-trap designs .....	31
3.4.2. Gain-of-function designs .....	32
4. Mammalian cells as genetic models .....	35
4.1. The mouse and rat as mammalian model experimental organisms.....	35
4.2. Mammalian cells as experimental models .....	36
4.3. Mouse ES cells as an attractive mammalian cell-based model.....	37
5. Strategies for recessive genetic screens in mammalian cell culture .....	39
5.1. Loss-of-heterozygosity-based strategies.....	39
5.1.1. Induced mitotic recombination using Cre/ <i>loxP</i> system.....	40
5.1.2. High G418 selection .....	42
5.1.3. <i>Blm</i> -deficient ES cell system and further developments.....	43
5.2. Haploid mammalian cell lines for recessive genetic screens .....	49
6. miRNAs and their biogenesis pathways.....	50
6.1. The discovery of miRNAs.....	50
6.2. miRNAs and siRNAs .....	52
6.3. miRNA biogenesis.....	53
6.3.1. The canonical biogenesis pathway .....	53
6.3.2. Different roles of Dicer homologues .....	55
6.3.3. Strand selection of the miRNA: miRNA* duplex.....	56
6.3.4. Choice of Argonaute association .....	56
6.3.5. Non-Canonical biogenesis pathways .....	57
6.4. Regulation of miRNA biogenesis .....	58
6.5. Wider implications of miRNA biogenesis .....	60
6.5.1. miRNA biogenesis and cancer .....	60
6.5.2. Hijacking miRNA biogenesis by viruses.....	61
6.5.3. miRNA biogenesis pathway mutants as tools to studying miRNA functions .....	62
7. Thesis project design.....	64

---

---

**Chapter 2 – Materials and methods**

1. Vectors.....	67
1.1. PB transposon and transposase vectors .....	67
1.2. PB transposon targeting vectors .....	69
1.3. Vectors miRNA reporters and their targeting vectors .....	72
1.4. BAC vectors for giant PB transposons .....	77
2. ES cell lines .....	79
2.1. Genotype of the precursor ES cell lines .....	79
2.2. ES cell lines generated.....	79
3. Media and chemicals used for ES cell culture.....	81
4. AB1, AB2.2 and B6 derived ES cell culture and manipulations.....	82
4.1. Passaging, freezing down and thawing ES cells.....	82
4.2. ES cell transfection by electroporation .....	83
4.3. ES cell transfection by lipofection and pulse puromycin selection.....	83
4.4. Picking ES cell colonies .....	84
4.5. Cre or Flp mediated recombination to pop-out selection cassettes.....	84
5. <i>piggyBac</i> mobilisation in ES cells.....	84
5.1. Plasmid-to-genome mobilisation .....	84
5.2. Intra-genomic <i>piggyBac</i> mobilisation.....	85
5.3. <i>piggyBac</i> Transposon excision for phenotypic reversion.....	86
6. Homozygote mutant generation using <i>Blm</i> -deficient ES cells.....	86
7. FACs analysis of eGFP expressing ES cells .....	87
8. Cytogenetic analysis .....	87
8.1. Preparation of metaphase spread.....	87
8.2. Sister chromatid exchange analysis.....	88
9. DNA methods .....	89
9.1. Recombineering technology.....	89
9.1.1. <i>E. coli</i> strains .....	89
9.1.2. Transformation of <i>E. coli</i> by electroporation.....	90

9.1.3. DNA fragment preparation for recombineering.....	90
9.1.4. Recombineering procedure .....	91
9.1.5. Cre or Flp mediated cassette pop-out in <i>E. coli</i> .....	91
9.2. Southern blotting and hybridization .....	92
9.3. Southern blotting probes used in this thesis.....	93
9.4. Splinkerette PCR.....	94
9.4.1. Adaptor preparation .....	94
9.4.2. Genomic DNA digestion and adaptor ligation .....	96
9.4.3. Nested-PCR amplification .....	96
9.5. Genomic DNA triple-primer competition PCR.....	97
10. RNA method: RT-PCR .....	98
11. Protein method: western blotting .....	101
11.1. Whole-cell protein extraction .....	101
11.2. Protein blotting and antibody hybridisation .....	101
11.3. Antibodies used in this thesis.....	102
12. Regional high density Genomic comparative hybridisation (CGH) array.....	102
13. Illumina sequencing for PB integration sites analysis.....	103
14. Bioinformatic analysis of Illumina sequencing data.....	105

### **Chapter 3 – Generation of an efficient mutagenic strategy using *piggyBac***

1. Introduction.....	107
1.1. Design principles of the mutagen.....	108
1.2. DNA mismatch repair pathway as a screening model for proof-of-principle .....	110
2. Results .....	113
2.1. Mutagenic exon-pair selection .....	113
2.2. Validations of mutagen units using an <i>Hprt</i> trapping assay.....	115
2.2.1. <i>Hprt</i> trapping assay with 6-TG and G418 dual selection .....	117
2.2.2. 4-OHT induction optimisation .....	118
2.2.3. <i>Hprt</i> trapping assay with sequential 6-TG and G418 selection .....	122



2.3. Validations of mutagen units using a random trapping assay .....	125
2.3.1. Generation of a cell line with an autosomal single-copy PB transposon .....	126
2.3.2. The random trapping assay .....	127
2.4. Local hopping of PB around the <i>Gdf9</i> locus .....	131
2.5. Proof-of-principle of my mutagenic strategy using DNA mismatch repair screen .....	133
2.5.1. PB reintegration efficiency estimation in NN5-Gdf9 <sup>hprtminiPB/+</sup> cells.....	133
2.5.2. Library construction and screening .....	134
2.5.3. Mutant analysis and validation.....	138
3. Discussion .....	141
3.1. Molecular design of the gene-inactivating PB transposon .....	141
3.2. <i>piggyBac</i> possesses a fast transposition kinetics .....	142
3.3. Local hopping characteristics of <i>piggyBac</i> .....	144
3.4. A recessive genetic screen using the established mutagenesis strategy.....	146
3.5. Complete genotype reversion using PBase with FIAU selection.....	147

#### **Chapter 4 – Generation of a *Blm*-deficient ES cell line with single-copy PB in *Hprt* locus**

1. Introduction.....	149
2. Results .....	149
2.1. Generation of a new <i>Blm</i> -deficient ES cell line .....	149
2.2. Phenotypic assessments of the <i>Blm</i> <sup>e/e</sup> cell line .....	153
2.3. Gene targeting of the mutagenic PB to the <i>Hprt</i> locus in <i>Blm</i> <sup>e/e</sup> cells .....	154
2.4. PB-remobilisation assessment of the newly generated cell lines.....	156
3. Discussion .....	159

#### **Chapter 5 – Development of reporter systems for probing the miRNA pathway**

1. Introduction.....	162
1.1. Why studying the miRNA biogenesis and effector pathways .....	162
1.2. Challenges in using <i>Blm</i> -deficient ES cell system to study the miRNA pathway.....	164
2. Results .....	164

2.1. Development of the artificial miR-eGFP system .....	166
2.1.1. Generation of an eGFP knock-in reporter cell line .....	166
2.1.2. miR-30-based miR-eGFP generation and analysis .....	167
2.1.3. Construct optimisation for the artificial miR-eGFP construct .....	171
2.1.4. Generation of an artificial miR-eGFP knockin in the <i>Blm<sup>e/e</sup>Hprt<sup>PBin2</sup></i> cell line .....	175
2.1.5. Validation of the miR-eGFP knockin <i>Blm<sup>e/e</sup>Hprt<sup>PBin2</sup></i> cell line.....	177
2.2. Development of an endogenous miRNA target reporter.....	179
2.2.1. Artificial ESCC-miRNA target generation .....	180
2.2.2. Generation of an ESCC-miRNA target reporter knockin <i>Blm<sup>e/e</sup>Hprt<sup>PBin2</sup></i> cell line .....	186
3. Discussion .....	189

## **Chapter 6 – Preliminary screen and condition optimisation for the miR-eGFP system**

1. Introduction.....	193
2. Results .....	194
2.1. Generation of the mutant library .....	194
2.2. Screening strategies and optimisation .....	196
2.3. Mutant validation .....	202
3. Discussion .....	204
3.1. Future improvements on the miR-eGFP screening system.....	204
3.2. <i>Ago2</i> , Argonaute proteins and the small RNA mediated pathways.....	206

## **Chapter 7 – Mobilisation of giant *PiggyBac* transposons in the mouse genome**

1. Introduction.....	210
2. Results .....	211
2.1. The generation of giant PB transposons .....	211
2.2. Transposition detection of giant PB transposons in ES cells.....	213
2.3. Individual clone analysis of giant PB genomic integrations .....	218
2.4. Analysis of chromosomal excision capability of giant PB.....	221
3. Discussion .....	223

**Chapter 8 – Conclusions and future work**

8.1. Conclusions..... 225

8.2. Future work..... 226

**References** ..... 229

**Abbreviations**

°C	degrees centigrade
4-OHT	4-hydroxyltamoxifen
6-TG	6-thioguanine
Ago	Argonaute
BAC	bacterial artificial chromosome
<i>Blm</i>	Bloom syndrome gene
bp	base pair
<i>bghpA</i>	bovine growth hormone poly-adenylation signal
BrdU	5-bromo-2-deoxyuridine
<i>Bsd</i>	blasticidin resistant gene
CAG	the hybrid cytomegalovirus enhancer/chicken $\beta$ actin promoter
CMV	cytomegalovirus promoter
CGH	comparative genomic hybridisation
DMSO	dimethylsulfoxide
ESCC	ES cell-specific cell cycle-regulating
FIAU	1-(2-deoxy-2-fluoro-1- D-arabinofuranosyl)-5-iodouracil
<i>Huc</i>	human ubiquitin C promoter
<i>IRES</i>	Internal ribosome entry site
LB	Luria-Bertani broth
LOH	loss of heterozygosity
miRNA	microRNA
MMR	mismatch repair
MOPS	3-morpholinopropane-1-sulfonic acid, $C_7H_{15}NO_4S$
<i>Neo</i>	neomycin resistant gene
nt	nucleotides
o/n	over night
PB	<i>piggyBac</i> transposon
PB5 or PB3	<i>piggyBac</i> 5' (or 3') inverted terminal repeat

PBase	<i>piggyBac</i> transposase
PB ITRs	<i>piggyBac</i> transposon inverted terminal repeats
PBS	Phosphate buffered saline
PCR	polymerase chain reaction
<i>PGK</i>	phosphoglycerate kinase
Pri-miRNA	Primary microRNA transcript
Pre-miRNA	Precursor microRNA transcript
<i>Puro</i>	puromycin resistant gene
<i>PuroΔtk</i>	fusion gene of puromycin <i>N</i> -acetyltransferase gene ( <i>puro</i> ) and truncated herpes simplex viral type 1 thymidine kinase gene ( $\Delta tk$ )
<i>rbgpA</i>	rabbit $\beta$ -globin poly-adenylation signal
RISC	RNA-induced silencing complex
RNAi	RNA interference
r.p.m	revolutions per minute
RT-PCR	reverse transcripton- PCR
SCE	sister chromatid exchange
SDS	Sodium lauryl sulfate, $C_{12}H_{25}SO_4Na$
siRNA	small interfering RNA
TBS	Tris buffered saline

## List of figures

### Chapter 1

Figure 1-1: Germline and somatic mutagenesis .....	11
Figure 1-2: Somatic mosaic recessive screens in <i>Drosophila</i> eye .....	13
Figure 1-3: Synthetic suppression screens.....	15
Figure 1-4: Synthetic enhancement screens.....	17
Figure 1-5: Transposon-mediated mutagenesis .....	25
Figure 1-6: Molecular designs for loss- and gain-of-function insertional mutagens.....	34
Figure 1-7: Mitotic recombination following G2-X and G2-Z segregation.....	41
Figure 1-8: Generation of homozygous mutant from heterozygous counterparts.....	46
Figure 1-9: Four steps involved in conducting recessive genetic screens using <i>Blm</i> -deficient ES Cells .....	47
Figure 1-10: Canonical miRNA and endo-siRNA biogenesis pathway.....	55
Figure 1-11: Mirtron biogenesis pathway, bypassing the <i>Drosha</i> processing step .....	58
Figure 1-12: Schematic representation of the mutagenic strategy employed for this project...	65
Figure 1-13: A schematic representation of a reporter strategy to screen for miRNA biogenesis mutants .....	66

### Chapter 2

Figure2-1: Schematic representation of the Splinkerette-PCR method .....	95
--	----

### Chapter 3

Figure 3-1: Design of an inactivating PB transposon targeted into the <i>Hprt</i> locus of <i>Blm</i> -deficient ES cells. ....	107
Figure 3-2: The mechanistic basis of 6-TG mediated cytotoxicity through MMR system.....	112
Figure 3-3: Mutagen <i>Dom3z</i> and <i>Ccdc107</i> exon-pair structures .....	115
Figure 3-4: Schematic representations of the <i>Hprt</i> trapping assay .....	116
Figure 3-5: Reversion analysis of the 6-TG resistant colonies .....	118
Figure 3-6: 4-OHT induction protocol optimisation.....	121

Figure 3-7: PB transposon mediated <i>Hprt</i> trapping.....	124
Figure 3-8: Schematic representation of the random trapping assay .....	125
Figure 3-9: Gene targeting of the <i>Gdf9</i> locus to generate the NN5- <i>Gdf9</i> <sup>hprrminiPB/+</sup> cell line .....	127
Figure 3-10: Characterisation of the trapping events mediated by the mutagenic units .....	130
Figure 3-11: Local hopping observed with PB transposon re-mobilised from the <i>Gdf9</i> locus ....	132
Figure 3-12: PB re-integration efficiency excising from the <i>Gdf9</i> locus .....	134
Figure 3-13: Schematic representation of the experimental procedures for isolating homozygous MMR mutants .....	136
Figure 3-14: <i>Msh6</i> mutant validation.....	138
Figure 3-15: <i>Msh6</i> mutant rescue analysis .....	140
Figure 3-16: A model to account for the PB transposition kinetics .....	143
Figure 3-17: Possible intra-genomic mobilisation kinetics of PB transposon.....	146

## Chapter 4

Figure 4-1: Double-allele targeting of <i>Blm</i> locus of the JM8.F6 ES cell line.....	152
Figure 4-2: Functional validation of the new <i>Blm</i> -deficient ES cell line <i>Blm</i> <sup>e/e</sup> .....	154
Figure 4-3: Genetic variations with the <i>Hprt</i> targeting region.....	155
Figure 4-4: The <i>Hprt</i> targeting of the mutagenic PB transposon.....	157
Figure 4-5: PB re-mobilisation efficiency assessments .....	158

## Chapter 5

Figure 5-1: Schematic representation of the artificial miR-eGFP system.....	165
Figure 5-2: Targeting <i>Gdf9</i> locus with an eGFPiRESBsd reporter .....	166
Figure 5-3: The construction of miR-30 backbone based single-unit miR-eGFP .....	168
Figure 5-4: miR-30-based artificial miR-eGFP evaluation .....	170
Figure 5-5: miR-155 backbone based miR-eGFP multimer generation .....	172
Figure 5-6: Optimisation of miR-eGFP to knockdown eGFP .....	173
Figure 5-7: Gene targeting of miR-eGFP multimer construct to the <i>mir-384</i> locus.....	176
Figure 5-8: mireGFP-knockin cell line validation.....	178

Figure 5-9: An endogenous miRNA-based reporter system .....	180
Figure 5-10: ESCC-miRNA eGFP reporter analysis.....	181
Figure 5-11: <i>Dgcr8</i> <sup>gt1/tm1</sup> ES cell validation .....	183
Figure 5-12: Causality establishment of the ES cell miRNAs and the ESCC-miRNA reporter .....	184
Figure 5-13: ESCC-miRNA Neo reporter analysis .....	186
Figure 5-14: The <i>mir384</i> -locus targeting with the ESCC-miRNA reporter construct .....	188

## Chapter 6

Figure 6-1: A hypothetical estimation of generating homozygous mutant in miRNA pathway..	194
Figure 6-2: Schematic representation of the mutant library construction.....	195
Figure 6-3: Screening strategies for the miR-eGFPsystem.....	196
Figure 6-4: Southern detection for the PB transposon reintegration patterns.....	201
Figure 6-5: <i>Ago2</i> mutant molecular analysis .....	203
Figure 6-6: Multiple roles of <i>Ago2</i> in the non-coding small RNA biogenesis, regulation and effector pathways .....	209

## Chapter 7

Figure 7-1: Giant PB construction and selection scheme incorporated to detect their transposition in mouse ES cells.....	212
Figure 7-2: Transfection efficiency of the AB2.2 ES cells determined by the eGFP expression after pulse-puromycin selection .....	213
Figure 7-3: Outline of the experimental scheme .....	214
Figure 7-4: Schematic representation of the experimental platform and bioinformatics analysis to identify transposition events using the Illumina sequencing .....	217
Figure 7-5: Giant PB transposition with single copy integration per cell .....	219
Figure 7-6: Regional high density CGH array analysis to determine the PB cargo integrity .....	220



**List of tables**

Table 2-1: primers used to construct giant PB transposons .....	79
Table 2-2: A summary of ES cell lines generated for this thesis work .....	80
Table 2-3: Southern blotting probes used in this thesis work. ....	93
Table 2-4: Locus-specific primers used for the random trapping assay .....	100
Table 2-5: Antibodies used in this thesis work .....	102
Table 2-6: Primers used for the multiplex Illumina sequencing to identify PB integration sites	105
Table 3-1: Estimate of number of cells required to obtain <i>Hprt</i> trapping events .....	119
Table 3-2: Ten PB reintegration clones selected from the random trapping .....	128
Table 3-3: Summary of seven independent integration sites identified from the MMR screen.	137
Table 3-4: Local-hopping comparison among different genomic loci .....	144
Table 5-1: Summary of different miR-eGFP constructs generated.....	173
Table 6-1: Efficiency comparison of screening protocols to identify eGFP-positive clones .....	200
Table 6-2: Summary of the identities of the eGFP-positive clones.....	200
Table 7-1: Transposition efficiency of different sized PB transposons and versions of the PBBase .....	216