

# Transposon-mediated Insertional Mutagenesis in Gene Discovery and Cancer

---

This dissertation is submitted for the degree of Doctor of Philosophy

By Jun Kong



Team 113, Experimental Cancer Genetics Group

The Wellcome Trust Sanger Institute

Wellcome Trust Genome Campus

Hinxton, Cambridge CB10 1SA



Darwin College

University of Cambridge

Silver Street

Cambridge CB3 9EU

## DECLARATION

I hereby declare that this dissertation is the results of my own work and includes nothing which is the outcome of work done in collaboration, except the tumour watch and disease profiling studies for the *Tel-AML1* mouse project in Chapter 4. This work was a collaboration with Dr. Louise van der Weyden, my colleague in the lab and Dr. Brian Huntly in MRC-CIMR Cambridge. I have included some of their *in vivo* work to prove that this mouse model has been successful for modelling cALL (childhood Acute Lymphoblastic Leukemia) in human, and is included in Figure 4-9.

None of the material presented herein has been submitted previously for the purpose of obtaining another degree. Material included in Chapter 2 has been published as Kong et al. (2008) *Bioinformatics* 24:2923-2925, material from Chapter 3 in Kong et al., (2010) *Nucleic Acids Research* 2010;38;18:e173 and Liang et al., (2009) *Genesis* 47:404-408. These publications are included as appendices. This dissertation does not exceed the word limit for the respective Degree Committee.

Jun Kong

September 20, 2010

## ACKNOWLEDGEMENT

I would like to take this opportunity to extend my sincere thanks to all the people who generously supported me during my PhD study.

I would like to first express my deepest thanks to my supervisor Dr. David Adams. Since the first time I met him in the Sanger institute, he has been given me so much help and support all through the years. While I was working with the mice from the beginning, he opened up my mind to explore other areas based on my interest and under his guidance most efforts afterwards achieved successful results in the end. He was also quite patient with me when I made some mistakes with the experiment. It was a great pleasure and superior experience to have an opportunity to work with David during these four years.

I would like also to give my special thanks to Louise Van Der Weyden, Theodore Whipp, Catherine Wilson, Rebecca McIntyre, Chi Wong and all other members in Team 113 who have provided me enormous help and support during the time. Louise and Cathy showed me how to handle mice, Theo taught me how to grow cell culture and helped me prepare all the routine lab reagents and equipment for my experiment, Rebecca helped discuss and edit my thesis, Chi offered lots of general analysis and suggestions. Most importantly, their friendship has made the Team 113 such a wonderful place for me to carry out my PhD.

The advice and feedback from my second supervisor and the PhD committee members was precious. I am grateful to Dr. Mark Arends from Cambridge University Department of Pathology, Dr. Derek Stemple, Dr. Allan Bradley, Dr. Jos Jonkers from Netherland Cancer Institute (NKI) for their precious time and enlightening advice during my project.

In addition, I received quite a lot of help and advice from other members in Sanger and from the University. I would like to thank Dr. Brian Huntly from MRC-CIMR Cambridge for his advice and collaboration on the *Tel-AML1* project, for Mark Bushell who directed me with the immunoprecipitation experiment design, for Barry Rosen with the construct supply and many suggestions, for Jim Stalker with the informatics support on *iMapper*, for Bee Ling with help and advice on FACS analysis. Without these generous help I could have never achieved so much during my PhD years.

Finally, I would like to thank my parents, Jiefang Kong and Yulin Chen, as well as my fiancée Yi Yao. Their endless and unconditional love and support has accompanied me all through this most exciting period of my life.

## SUMMARY

The advent of DNA sequencing has significantly accelerated biological research and discovery. Complete genomic sequences, together with approximately 20,000–25,000 annotated genes in the human genome, analysed through contemporary bioinformatic technology, must be functionally annotated by up-to-date biological methods to assign genes to pathways and functions. During my PhD, I combined *in vivo* and *in vitro* studies, together with the power of bioinformatics, to dissect gene functions under different contexts.

The fundamental basis of my PhD research is utilizing a system called transposon mutagenesis. Transposons are mobile genetic elements that represent a large portion of the repetitive sequences in the human genome. In *in vitro* cell culture studies, I developed a novel system called ‘Slingshot’ that is based on the *piggyBac* transposon system, which is capable of randomly mutagenizing the genome of many cell types in a ‘gain-of-function’ or ‘loss-of-function’ manner. Using this system, I performed drug resistance screens in the mouse embryonic stem cells. Subsequently, several drug transporter genes were identified in these screens that provide drug resistance to puromycin and the anti-cancer drug vincristine. I have also validated the efficiency of this transposon system using human somatic cell lines. In the *in vivo* studies, in collaboration with other colleagues, a *Tel-AML1* knockin mouse was generated to model childhood acute lymphoblastic leukaemia (cALL) that is characterized by a chromosomal translocation which results in the expression of a TEL-AML1 fusion protein. When crossed with the *Sleeping Beauty* transposon mice for cooperative mutations, some of these *Tel-AML1* mice derived the appropriate type of B cell leukaemia under tumour watch analysis. Another conditional *Brd4-NUT* mouse model for human midline carcinoma was generated using a similar knockin approach. Although this model did not transmit through the germ line for *in vivo* studies, *in vitro* experiments have revealed a strong cell growth arrest phenotype associated with the *Brd4-NUT* expression.

In addition, to provide better analysis of insertion sites for the transposon studies, I have developed an online web-based tool called ‘*iMapper*’, which analyzes large numbers of transposon integration sites from sequence reads and maps them to the appropriate Ensembl genome. I have successfully used this bioinformatics tool to analyze the insertion sites in sequence reads generated by my own experiments. This online resource is freely accessible and could facilitate the analyses of sequence reads and mapping of insertion sites for mutagenesis studies performed world-wide.

## TABLE OF FIGURES

|  |     |
|--|-----|
| Figure 1-1. Retrovirus genome structure and recombinant retrovirus production.....   | 10  |
| Figure 1-2. Schematic of three basic gene trap strategies.....   | 12  |
| Figure 1-3. Molecular structure of the Sleeping Beauty transposon and the SB transposition system .....                        | 17  |
| Figure 1-4. Molecular structure of the PB transposon and the PB transposition system.....                                      | 19  |
| Figure 1-5. The Cre-loxP system and three applications in the eukaryote genome .....   | 32  |
| Figure 1-6. An outline of my PhD projects.....   | 39  |
| Figure 2-1. Schematic diagram of linker-based PCR procedure and contaminating sequences .....                                  | 43  |
| Figure 2-2. The workflow of <i>iMapper</i> .....   | 47  |
| Figure 2-3. The interface of <i>iMapper</i> .....  | 50  |
| Figure 2-4. The output of <i>iMapper</i> .....   | 54  |
| Figure 2-5. Performance of <i>iMapper</i> : analysis of 1920 piggyBac traces .....   | 57  |
| Figure 3-1. Plasmid based PB Transposon system in cell culture .....   | 65  |
| Figure 3-2. Schematic diagram of the Slingshot plasmid and the mutagenesis scenarios possible with this system .....           | 68  |
| Figure 3-3. Functional test for the Slingshot PB plasmid.....  | 74  |
| Figure 3-4. Testing of transposition activity and re-integration sites in Slingshot PB integrated ES cell lines.....           | 76  |
| Figure 3-5. Drug resistance screen in Slingshot PB cell line PB/PB-1 using puromycin .....                                     | 79  |
| Figure 3-6. Screen for vincristine resistance in the Slingshot PB cell line PB/PB-1.....                                       | 81  |
| Figure 3-7. Slingshot is a self-inactivating transposon system.....  | 83  |
| Figure 3-8. Slingshot is functional in three human somatic cell lines .....  | 85  |
| Figure 3-9. Improvement of the Slingshot PB donor colony formation efficiency using chicken insulator sequence cHS4.....       | 87  |
| Figure 4-1. Balanced chromosome translocation and cancer cell malignancy initiated by chromosome translocation. ....           | 92  |
| Figure 4-2. Schematic diagrams of <i>Tel-AML1</i> targeting constructs and targeted alleles .....                              | 96  |
| Figure 4-3. Crossing strategy for tumour watch and subsequent characterization studies in the <i>Tel-AML1</i> mouse model..... | 98  |
| Figure 4-4. Characterization of <i>Tel-AML1</i> targeted ES cells.....   | 105 |
| Figure 4-5. Characterization of the mouse-human TEL-AML1 protein by <i>in vitro</i> studies..                                  | 107 |

|   |     |
|---|-----|
| Figure 4-6. Characterization of the transposon system in <i>Tel-AML1</i> mouse model.....                                   | 109 |
| Figure 4-7. <i>In vivo</i> validation of the <i>Tel-AML1</i> expression strategy and real-time qPCR.....                    | 111 |
| Figure 4-8. Investigation of cryptic splicing in <i>Tel-AML1</i> knockin mice by intercrossing ...                          | 113 |
| Figure 4-9. Tumour progression and histological analysis in <i>Tel-AML1</i> knockin mice.....                               | 116 |
| Figure 4-10. The targeting construct of Rosa 26 GFP-TEL-AML1 mouse model.....   | 117 |
| Figure 5-1. Karyotype of t(15;19) BRD4-NUT translocation in a 30-year-old midline carcinoma patient. ....                   | 123 |
| Figure 5-2. Schematic diagrams of Brd4-NUT targeting constructs and targeted alleles. ....                                  | 124 |
| Figure 5-3. Crossing strategy for tumour watch and subsequent characterization studies in <i>Brd4-NUT</i> mouse model. .... | 126 |
| Figure 5-4. Characterization of the <i>Brd4-NUT</i> targeted ES cells. ....   | 130 |
| Figure 5-5. Evaluation of the <i>Brd4-NUT</i> expression level by qPCR.....   | 132 |
| Figure 5-6. Cell plating assay in <i>Brd4-NUT</i> ES cells. ....  | 137 |
| Figure 5-7. Colony formation assay with <i>Brd4-NUT</i> ES cells .....  | 139 |
| Figure 5-8. Cell cycle analysis of <i>Brd4-NUT</i> ES cells .....   | 141 |

## TABLE OF TABLES

|  |     |
|--|-----|
| Table 2-1. Time required for <i>iMapper</i> to analyze two datasets with different settings.....       | 58  |
| Table 3-1. Evaluation of PB/PB-1 transposition efficiency using a time Series of 4-OHT treatment ..... | 77  |
| Table 5-1. Brd4-NUT clone microinjection information.....  | 133 |

# TABLE OF CONTENTS

|   |     |
|---|-----|
| DECLARATION .....   | I   |
| ACKNOWLEDGEMENT .....   | II  |
| SUMMARY .....   | III |
| TABLE OF FIGURES .....  | IV  |
| TABLE OF TABLES .....   | V   |
| Chapter 1. General introduction .....   | 10  |
| 1.1 Introduction to genetic screening and transposons .....   | 10  |
| 1.1.1 The human genome and functional studies .....   | 10  |
| 1.1.2 Genetic screening is a powerful tool for functional studies .....   | 2   |
| 1.1.3 Dominant and recessive genetic screens .....  | 2   |
| 1.1.4 Forward and reverse genetic screens .....   | 3   |
| 1.1.5 Reverse genetic screening for functional studies .....  | 4   |
| 1.1.6 Classical forward genetic screens .....   | 5   |
| 1.1.7 Insertional mutagenesis screens .....   | 8   |
| 1.1.8 Transposon-mediated mutagenesis .....   | 14  |
| 1.2 Insertional mutagenesis screen for cancer gene identification .....   | 24  |
| 1.2.1 A summary of cancer .....   | 24  |
| 1.2.2 Methods for cancer gene identification .....  | 26  |
| 1.3 The experimental mouse as a model organism .....  | 28  |
| 1.3.1 The mouse and human genome .....  | 28  |
| 1.3.2 Mouse as a model organism .....   | 29  |
| 1.3.3 Mouse embryonic stem cell as a genetic tool .....   | 29  |
| 1.3.4 Strategies used in ES cells for mouse genetics study .....  | 30  |
| 1.3.5 Mouse as a model for human cancer .....   | 33  |
| 1.4 Insertional sites analysis and mapping .....  | 34  |
| 1.4.1 Isolation of the insertion sites .....  | 35  |
| 1.4.2 Sequence mapping .....  | 36  |
| 1.5 My PhD project overview .....   | 38  |
| Chapter 2. <i>iMapper</i> : A web server for the automated analysis and mapping of insertional<br>mutagenesis sequence data against Ensembl genomes ..... | 41  |
| 2.1 Introduction .....  | 41  |
| 2.2 Aim and summary of the project .....  | 44  |

|            |   |    |
|------------|---|----|
| 2.3        | Materials and Methods .....   | 45 |
| 2.3.1      | Architecture of the program .....   | 45 |
| 2.3.2      | Sequence Processing.....  | 46 |
| 2.3.3      | Performance Test .....  | 48 |
| 2.3.4      | Chromosome <i>KaryoView</i> graph.....  | 48 |
| 2.4        | Results .....   | 48 |
| 2.4.1      | Program interface .....   | 48 |
| 2.4.2      | Program Output .....  | 52 |
| 2.4.3      | Performance of <i>iMapper</i> .....   | 55 |
| 2.5        | Discussion .....  | 59 |
| Chapter 3. | Slingshot: A <i>piggyBac</i> based transposon system for tamoxifen-inducible ‘self-inactivating’ insertional mutagenesis..... | 62 |
| 3.1        | Introduction .....  | 62 |
| 3.2        | Aim and summary of the project.....   | 66 |
| 3.3        | Materials and Methods .....   | 69 |
| 3.3.1      | Plasmids construction .....   | 69 |
| 3.3.2      | Cell culture media.....   | 69 |
| 3.3.3      | Generation of an cell lines with stable integration of the Slingshot plasmid .....  | 70 |
| 3.3.4      | Trapping efficiency test and mobilisation assay.....  | 70 |
| 3.3.5      | Excision PCR on DNA for 4-OHT treated cells.....  | 70 |
| 3.3.6      | Drug resistance screen using puromycin.....   | 71 |
| 3.3.7      | Drug resistance screen using vincristine .....  | 71 |
| 3.3.8      | Splinkerette PCR and insertion sites analysis.....  | 71 |
| 3.3.9      | Western blotting .....  | 72 |
| 3.4        | Results .....   | 72 |
| 3.4.1      | Generation of the Slingshot PB system and Slingshot ES cell lines .....   | 72 |
| 3.4.2      | Evaluating the jumping efficiency of the Slingshot transposon from a stable donor   | 75 |
| 3.4.3      | Drug resistance screens using the Slingshot system.....   | 78 |
| 3.4.4      | Self-inactivation of the transposon after transposition .....   | 82 |
| 3.4.5      | The Slingshot transposon system is active in somatic cell lines .....   | 84 |
| 3.4.6      | Increasing the integration efficiency using chicken insulator sequence cHS4.....  | 86 |
| 3.5        | Discussion .....  | 88 |

|   |     |
|---|-----|
| Chapter 4. Modelling <i>Tel-AML1</i> oncogenic translocation using knockin mice and transposon-mediated insertional mutagenesis.....          | 91  |
| 4.1 Introduction .....  | 91  |
| 4.2 Aims and summary of the project .....   | 94  |
| 4.3 Materials and Methods .....   | 99  |
| 4.3.1 Targeting construct generation .....  | 99  |
| 4.3.2 ES cell transfection and selection.....   | 100 |
| 4.3.3 Generation of pMSCV expression constructs .....   | 100 |
| 4.3.4 Immunocytochemistry .....   | 101 |
| 4.3.5 Immunoprecipitation of Flag Tagged Proteins.....  | 101 |
| 4.3.6 RNA isolation and cDNA preparation.....   | 102 |
| 4.3.7 Quantitative PCR.....   | 102 |
| 4.4 Results .....   | 103 |
| 4.4.1 Generating the <i>Tel-AML1</i> knockin mouse model and characterizing the targeted ES cells .....                                       | 103 |
| 4.4.2 Detection of the Tel-AML1 fusion protein by immunoprecipitation .....   | 104 |
| 4.4.3 <i>In vitro</i> characterization of the mouse TEL – human AML1 fusion protein.....  | 106 |
| 4.4.4 Analysis of the <i>Sleeping Beauty</i> Transposon system in the knockin mouse .....   | 108 |
| 4.4.5 Validation of the <i>Tel-AML1</i> expression level by real-time qPCR.....   | 110 |
| 4.4.6 Analysis of cryptic splicing in <i>Tel-AML1</i> knockin mice.....   | 112 |
| 4.4.7 Tumour watch study in <i>Tel-AML1</i> knockin mice .....  | 114 |
| 4.4.8 An alternative mouse model of TEL-AML1 .....  | 117 |
| 4.5 Discussion .....  | 118 |
| 4.5.1 Advantages of knockin mouse model for characterizing human <i>TEL-AML1</i> oncogenic translocation .....                                | 118 |
| 4.5.2 Expression level of oncogenic fusion protein.....   | 118 |
| 4.5.3 The choice of mouse or human AML1 .....   | 119 |
| 4.5.4 The <i>Tel-AML1</i> knockin mice as a model for human ALL .....   | 119 |
| Chapter 5. Modelling the consequences of <i>Brd4-NUT</i> oncogenic translocation in mouse ES cells using a conditional knockin strategy ..... | 121 |
| 5.1 Introduction .....  | 121 |
| 5.2 Aims and summary of the project .....   | 123 |
| 5.3 Materials and Methods .....   | 127 |
| 5.3.1 Targeting vector construction.....  | 127 |

|             |   |     |
|-------------|---|-----|
| 5.3.2       | Immunoprecipitation of FLAG Tagged Proteins.....  | 127 |
| 5.3.3       | Cell proliferation and colony formation assay .....   | 128 |
| 5.3.4       | Cell Cycle Analysis by Flow Cytometry.....  | 128 |
| 5.4         | Results .....   | 128 |
| 5.4.1       | Generating “conditional” <i>Brd4-NUT</i> knockin mouse model and characterizing targeted ES cells .....   | 128 |
| 5.4.2       | Analysis of Brd4-NUT expression using Quantitative PCR .....  | 131 |
| 5.4.3       | Deriving germ line transmission with the Brd4-NUT knockin ES cell lines .....   | 133 |
| 5.4.4       | Expression of <i>Brd4-NUT</i> fusion impaired cell growth in ES cells.....  | 134 |
| 5.4.5       | Expression of <i>Brd4-NUT</i> fusion blocks colony formation ability in ES cells.....   | 138 |
| 5.4.6       | Expression of <i>Brd4-NUT</i> arrested cell cycle at G2/M phase.....  | 140 |
| 5.5         | Discussion .....  | 142 |
| 5.5.1       | Models of choices for characterizing human carcinogenic translocations .....  | 142 |
| 5.5.2       | Germ line transmission of the <i>Brd4-NUT</i> knockin ES cells.....   | 142 |
| 5.5.3       | <i>Brd4-NUT</i> induced cell growth arrest .....  | 143 |
| Chapter 6.  | Summary and future directions .....   | 144 |
| 6.1         | Generation of a high-efficient insertional mutagenesis pipeline.....  | 144 |
| 6.2         | Slingshot PB system in cell culture mutagenesis screen.....   | 145 |
| 6.3         | Transposon-mediated insertional mutagenesis for cancer mouse model.....   | 146 |
| 6.4         | iMapper: improvements and future directions .....   | 147 |
| REFERENCES  | .....   | 149 |
| Appendix A. | Primers and linker sequences used for Splinkerette PCR .....  | 160 |
| Appendix B. | ABBREVIATIONS .....   | 161 |
| Appendix C. | Publication 1 - Chromosomal mobilization and reintegration of <i>Sleeping Beauty</i> and <i>piggyBac</i> transposons.....                                       | 163 |
| Appendix D. | Publication 2 - <i>iMapper</i> : a web application for the automated analysis and mapping of insertional mutagenesis sequence data against Ensembl genomes..... | 164 |
| Appendix E. | Publication 3 - Slingshot: a <i>piggyBac</i> based transposon system for tamoxifen-inducible 'self-inactivating' insertional mutagenesis.....                   | 165 |