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 Caner 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

TABLE OF TABLES

TABLE OF CONTENTS