

Isolation of homozygous mutant mouse embryonic stem cells by selection for copy number increase

Stephen J. Pettitt

Wellcome Trust Sanger Institute

and

Girton College, University of Cambridge

This dissertation is submitted for the degree of Doctor of Philosophy.

June 2, 2011



UNIVERSITY OF
CAMBRIDGE



Acknowledgements

It has been a great privilege to be part of the Wellcome Trust Sanger Institute, and I am very grateful to the Wellcome Trust for funding my time here. I have met many talented and inspirational scientists here, without whom nothing in this thesis would have been possible.

Foremost among these is my supervisor, Allan Bradley, for sharing his wealth of knowledge and experience. He has given me great ideas and opportunities for research projects, as well as a lot of freedom to pursue my own interests. I am very grateful for his trust and support, and consider myself very lucky to have been his student.

I am also very grateful to my other supervisors at the Sanger Institute. Gavin Wright, Christian Söllner, Jon Teague, Adam Butler, Pentao Liu and Mariaestela Ortiz. Haydn Prosser deserves special mention for putting up with my mistakes and questions, while teaching me everything about gene targeting. I was also fortunate to collaborate with Bill Skarnes, who has continued to be a great source of advice. The enthusiasm of my external supervisor, Steve Jackson, has been infectious, and he gave me a lot of useful suggestions for DNA repair experiments. I also thank Junji Takeda (Osaka University) for an interesting discussion of ideas during his visit last year.

Being surrounded by so many successful and bright people in the lab is great motivation. I would particularly like to thank Kosuke, Yue, Amy and Wei, who were always ready to give advice and shared materials and results with me. Frances, James, Alastair and Hiroko all helped my cell culture go smoothly, and Nathalie and Charles read my thesis. Thanks also to EPT for looking after my cells and (sometimes) cooking me curry, Floris for making the espresso machine happen, Mekayla for making me go outside and run around, Roland, who can advise on absolutely anything, and George for some truly terrible jokes.

I thank the fellows, resident and honorary, of the Cavendish Institute for their numerous collaborations in the lower impact, but essential, research fields such as European geography, anthropology, oenology, pyrotechnics, beardomics and bioinformatics. Long may they continue. I am also lucky to have one long-term collaborator in particular, who truly understands the demands of ES cell culture.

Finally, I would like to thank my parents, who have always encouraged and supported me in anything I wanted to do. Their interest in my work means a lot to me; I hope they enjoy reading!

SJP

September 2010

Abstract and declaration

Forward genetic screens are a powerful method to determine which genes are responsible for a particular phenotype in many model organisms. However, a simple method to conduct genetic screens in a mammalian system has been difficult to develop, due to the problem of making random homozygous mutations in the diploid mammalian genome. Mouse embryonic stem (ES) cells provide a convenient model for mammalian cell biology. Previous studies showed that heterozygous mutants in ES cells with mutations in the *Blm* gene segregate homozygous mutants at a low rate, due to an increase in mitotic crossovers.

Using the piggyBac DNA transposon (PB) for initial single copy heterozygous mutagenesis, I describe a method to isolate the rare homozygous cells based on selection for transposon copy number, which increases to two in homozygotes. I successfully isolated homozygous mutants using this system, but my experiments revealed aneuploidy as an alternative copy number gain pathway in ES cells. By extensive engineering of the ES cell line and PB transposase, I developed a method to allow many different homozygous mutants to be generated in a pooled format. This minimises the problem of background from aneuploidy and allows isolation of clonally pure mutants suitable for genetic screens.

I also investigated the properties of the PB transposon. By sequencing and mapping thousands of insertion sites I have investigated the insertion site preferences of PB. This method can also be used to fully define coverage of mutant libraries. I showed that precise excision of PB from the genome depends on the nonhomologous end joining pathway, and present data indicating that transposition can occur throughout the cell cycle.

The methods and tools presented will be useful for study of gene function in mammalian cells, and are also applicable for the study of DNA double strand break repair and copy number stability.

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except where specifically indicated in the text.

Contents

1	Introduction	1
1.1	The human and mouse genomes	1
1.1.1	New genetic approaches	1
1.1.2	Importance of the mouse genome	1
1.1.3	Experimental approaches to analyse gene function	2
1.2	Reverse genetics in mice	2
1.2.1	Embryonic stem cells	3
1.2.2	Gene targeting	3
1.3	Forward genetics in mice	5
1.3.1	Inbred strains	5
1.3.2	ENU mutagenesis	6
1.3.3	Irradiation	7
1.3.4	Insertional Mutagenesis	7
1.3.5	Transposons active in mammalian cells	8
1.3.6	Comparison of transposons	10
1.4	Genetic screens in embryonic stem cells	11
1.4.1	Practicality of genome-wide screens in mice	11
1.4.2	Suitability of embryonic stem cells as a model	11
1.4.3	Dominant and recessive screens	12
1.4.4	Making homozygous mutations in ES cells	15
1.4.5	Biology of cells with mutations in the <i>BLM</i> gene	18
1.5	Isolation of homozygous mutants by selection for copy number increase	22
2	Materials and Methods	25
2.1	Embryonic stem cell lines	25
2.1.1	Wild-type cell lines	25
2.1.2	<i>Blm</i> -deficient cell lines	25
2.1.3	Other mutant cell lines	25
2.2	Cell culture	25
2.2.1	Culture conditions	25
2.2.2	Selective media	25
2.2.3	Transfection of ES cells	26
2.2.4	Cellular analysis	26
2.2.5	Isolation of nucleic acids and proteins	27
2.3	ES cell genotyping	28
2.3.1	PCR and long range PCR	28
2.3.2	Mapping transposon integration sites by splinkerette PCR	28
2.3.3	Southern blot	28
2.3.4	RT-PCR	30
2.3.5	Western blot	30
2.4	Molecular biology	30
2.4.1	Recombineering	30
2.4.2	Conventional cloning	31

3	A vector to make homozygous mutations with high genome coverage	33
3.1	Introduction	33
3.1.1	Estimating library coverage	33
3.1.2	Illumina sequencing technology	33
3.1.3	Mutagens	34
3.1.4	Isolation of homozygous mutants	34
3.2	Results	35
3.2.1	An insertional mutagen for non-selectable mutagenesis	35
3.2.2	Dual selection cassette for copy number based selection	37
3.2.3	Genome coverage and insertion preferences of the TNP vector	40
3.3	Discussion	55
3.3.1	The TNN/TNP transposon vector—mutagenesis	55
3.3.2	The TNN/TNP transposon vector—copy number selection	55
3.3.3	Coverage and insertion site preferences of PB	56
3.3.4	Conclusions	57
4	The rate of loss of heterozygosity in <i>Blm</i>-deficient ES cells	59
4.1	Introduction	59
4.1.1	Using fluctuation analysis to measure the rate of rare events in cell culture	59
4.2	Results	60
4.2.1	Choice of loci	60
4.2.2	Calculation of mutation rate	61
4.3	Discussion	61
4.3.1	Comparison with previously calculated rates	61
4.3.2	Implications for library coverage	61
5	Isolation of homozygous mutants in <i>Blm</i>-deficient ES cells based on copy number	65
5.1	Introduction	65
5.1.1	Copy number based selection	65
5.2	Results	65
5.2.1	Generation of single copy insertions	65
5.2.2	Mapping of insertion sites	68
5.2.3	Generation of double resistant clones	68
5.2.4	Genotyping double-resistant clones	68
5.2.5	Two classes of mutants are present in the double resistant population	70
5.2.6	Summary of isolated double resistant clones	73
5.2.7	Double resistant clones retaining a wild type locus	73
5.2.8	Karyotype of wild type retaining clones	78
5.2.9	DNA content analysis of wild type retaining subclones	78
5.2.10	The transposon disrupts transcription of genes when inserted into introns	81
5.3	Discussion	81
5.3.1	Paths to increase transposon copy number in <i>Blm</i> cells	81
5.3.2	Clones for which double resistant cells were not isolated	81
5.3.3	Implications for creation of homozygous mutant libraries	82
5.3.4	Conclusions	84
6	Isolating large numbers of homozygous mutants in parallel	87
6.1	Introduction	87
6.1.1	Aims	87
6.1.2	Clonal expansion	87
6.2	Results	87
6.2.1	C57BL/6 targeting vector to insert the transposon at the <i>Hprt</i> locus.	87
6.2.2	Generating libraries with the LGP cell line	91
6.2.3	Sources of background in double-resistant population	93

6.2.4	A G1-specific transposase to conserve copy number during transposition	99
6.2.5	Mobilisation using G1-specific transposase and FIAU counterselection	100
6.2.6	Results of double selection with PB-CDT1 and FIAU counterselection	103
6.2.7	Some allelic mutants retain the wild type locus	107
6.3	Discussion	107
6.3.1	Sources of background in mutants isolated from complex pools	107
6.3.2	PB transposition and the cell cycle	110
6.3.3	Generation and uses of pooled mutant libraries	110
6.3.4	Conclusions	113
7	Repair of DNA double strand breaks caused by piggyBac transposition	115
7.1	Introduction	115
7.1.1	Excision of transposons	115
7.1.2	Cellular double strand break repair pathways	115
7.1.3	Experimental induction of DNA double strand breaks	119
7.1.4	Aims	120
7.2	Results	120
7.2.1	Reporter cell lines with DNA repair deficiencies	120
7.2.2	<i>Xrcc4</i> and <i>Xlf</i> are required for survival after transposition	120
7.2.3	Mutations at the donor locus in <i>Xrcc4</i> mutants	124
7.2.4	Low frequency of mutations at the donor locus in <i>Xlf</i> mutants	124
7.2.5	No evidence for larger deletions related to transposition in <i>Xrcc4</i> mutants	130
7.2.6	PARP inhibition does not affect repair in the absence of <i>Xrcc4</i>	130
7.2.7	Excision and reintegration are not affected by inhibitors of PARP, ATM or DNA-PKcs in wild type cells	130
7.2.8	Homologous recombination repair of PB-induced breaks	134
7.3	Discussion	134
7.3.1	Requirement for host repair pathways in repair of PB-induced breaks	134
7.3.2	Differential requirement for <i>Xrcc4</i> and <i>Xlf</i> at PB-induced breaks	136
7.3.3	DNA repair requirements in V(D)J recombination and PB transposition	136
7.3.4	DNA repair requirements in SB transposition	137
7.3.5	Advantages of PB for programming double strand breaks	137
8	Discussion	141
8.1	Enrichment for homozygous mutants in <i>Blm</i> -deficient ES cells	141
8.1.1	Future improvements to library generation	141
8.2	Using enriched libraries for screens	141
8.2.1	New technologies applicable to genetic screens	141
8.2.2	Comparison to other systems for recessive genetic screens	142
8.3	Other uses of the copy number selection transposon	143
8.4	Conclusions	143
	Bibliography	155
	A Protocol: Generating libraries using the LGN cell line	157
	B Primer sequences	159

List of Figures

1.1	Insertion and replacement targeting vectors	5
1.2	Types of gene trap vector	9
1.3	Mitotic recombination leading to LOH in heterozygous cells	17
1.4	Formation and resolution of double Holliday junctions	21
1.5	Screens for selectable phenotypes in <i>Blm</i> -deficient cells	23
2.1	Targeting NN5 cells with a <i>Rosa26</i> :Cre-ERT2 construct	25
2.2	Splinkerette PCR method	29
3.1	Copy number gain during loss of heterozygosity	36
3.2	Features considered in design of the mutagen	37
3.3	Cloning scheme for PiggyBac mutagenesis vector	38
3.4	loxP sites are functional in the inverter construct	40
3.5	Cloning the inverter construct	41
3.6	Cloning of the TNN plasmid	42
3.7	Fixing a mutation in loxP site	43
3.8	Function of the transposon construct	43
3.9	Background resistance from leaky ERT2-Cre activity	44
3.10	G418 kill curve for puro ⁺ cells	45
3.11	Sensitivity of <i>Xrcc4</i> and <i>Xlf</i> mutant cells to bleomycin	46
3.12	Setup of pilot experiment for Illumina sequencing and dropout screens	46
3.13	Detection of loss of a tagged <i>Xrcc4</i> mutant by PCR	48
3.14	Distribution of paired ends in Illumina sequencing of PB insertions	49
3.15	Venn diagram showing effect of applying coverage filter	52
3.16	Graph of associations of PB insertions with genes and chromatin features	54
3.17	Venn diagrams illustrating insertion sites that overlap multiple features	55
4.1	Measuring rare events in cell culture.	60
4.2	Number of LOH events observed for three loci on chromosome 11	62
4.3	Plot of distance from centromere for all mouse genes.	64
5.1	Experimental scheme for clone-by-clone isolation of homozygous mutants	66
5.2	Generation of clones with single copy transposon insertions	67
5.3	Mapping of insertion sites by splinkerette PCR	69
5.4	Predicted expansion time required	70
5.5	Results of double drug selection	71
5.6	Southern blot to identify potential homozygous clones	72
5.7	Selection background due to lack of L-glutamine	73
5.8	Genotyping homozygous <i>Dym</i> mutants	74
5.9	No selection background under normal selection conditions	75
5.10	PCR genotyping of double resistant subclones	77
5.11	Southern blots of double resistant clones using locus-specific probes	79
5.12	Karyotypes of wild type retaining clones	80
5.13	DNA content analysis of wild type retaining subclones	82
5.14	Mutagenicity of the TNN/TNP construct	83
5.15	Effect of expansion time	85
6.1	Problems caused by clonal expansion in heterogeneous pools of cells	88
6.2	Retrieval of a fragment of the C57BL/6 <i>Hprt</i> locus	89

6.3	Cloning the TV28 targeting vector	90
6.4	Targeting the transposon to <i>Hprt</i> using the TV28 vector	92
6.5	Confirmation of targeting and testing transposition from the <i>Hprt</i> ^{PB} locus.	93
6.6	Targeting an inducible Cre gene in <i>Blm</i> ^{e/e} cells	94
6.7	Gene targeting combined with 4-OHT treatment to insert the transposon	95
6.8	Confirmation of targeting in LGP cells	95
6.9	General scheme for library generation using the LGP cell line.	96
6.10	Colonies stained at various stages of the library generation process	97
6.11	Background in double-resistant clones from pooled libraries	98
6.12	Sources of background in pooled libraries	100
6.13	Analysis of clones from the LGP A2 library	101
6.14	Structure and function of PBase-CDT1 fusion protein.	102
6.15	Western blot using anti-CDT1 antibody	103
6.16	Mobilisation using the PB-CDT1 fusion protein preserves copy number	104
6.17	Use of the fusion protein and counterselection does not compromise isolation of transposition events	105
6.18	Verification of 4-OHT sensitivity and specificity in the LGNL1 library	105
6.19	Analysis of double-resistant clones generated with G1-specific mobilisation and FIAU counterselection	106
6.20	Analysis of clonal relationships between LGNL1 double-resistant clones	107
6.21	PCR genotyping of allelic LGNL1 subclones	109
6.22	Uses of enriched libraries	112
7.1	Double strand break repair pathways in mammalian cells.	117
7.2	TV28 reporter locus for excision	121
7.3	Targeting the <i>Hprt</i> ^{PB} locus in NHEJ-deficient cells	121
7.4	Example of transposition assay	122
7.5	Survival in HAT medium following transfection of NHEJ reporter cell lines	123
7.6	PCR amplification of donor locus after transposition	125
7.7	Examples of mutations at the donor site in <i>Xrcc4</i> mutants	125
7.8	Expected and observed microhomology use at repair sites	127
7.9	Structure of insertions at excision site in NHEJ mutants	128
7.10	Sequence of insertions with clear structure	128
7.11	Junction sequence of a repair event retaining part of the transposon	129
7.12	PCR amplification of donor locus from <i>Xlf</i> mutants	129
7.13	FIAU+6-TG selection to detect large deletions.	131
7.14	Results of transposition in <i>Xrcc4</i> mutant cells treated with PARP inhibitor	132
7.15	Graph showing frequency of repair event classes in different mutants	133
7.16	Distribution of total deletion size at repair site in NHEJ mutants	133
7.17	Transposition in wild type cells treated with ATM, DNA-PKcs and PARP inhibitors	135
7.18	Requirements of NHEJ factors in V(D)J recombination and PB transposition	137

List of Tables

1.1	Comparison of mutagens	12
2.1	Drugs used in cell culture	26
3.1	Filtering Illumina sequencing reads	48
3.2	Many insertion sites are unique to one sample and have low sequence coverage	51
3.3	Effect of minimum coverage filtering on agreement between samples	51
3.4	Search for mapped insertion sites in <i>Xrcc4</i> and <i>Xlf</i> mutants.	52
3.5	Association of PB integrations with genes and chromatin features	54
4.1	Calculation of LOH rate	62
5.1	Results of genotyping for all double resistant clones	76
6.1	Mapping data for LGNL1 clones with allelic insertions	108
7.1	Transposition outcomes using the TV28 reporter locus	121
7.2	Targeting efficiency in NHEJ-deficient cell lines	122
7.3	Sequencing of the repair site in <i>Xrcc4</i> mutants	126
7.4	Mutations at donor locus in <i>Xlf</i> mutants	126
7.5	Mutations at the site of repair in <i>Xrcc4</i> cells treated with a PARP inhibitor	131
7.6	Summary of types of event observed in different mutants	132
7.7	Comparison of different site specific nuclease systems for causing experimental DSBs	139