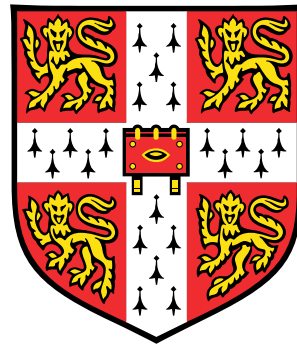


Identifying mediators of malignant transformation in human cancer using genome-wide forward genetic screening approaches



Eleanor Dunstone

Wellcome Sanger Institute

University of Cambridge

This dissertation is submitted for the degree of
Master of Philosophy

I would like to dedicate this thesis to my parents, grandparents and my brother Jack for their ongoing love and support, and to Ben, Lily and Martha for their reassurance and encouragement during the writing process.

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Declaration and specified in the text. Additionally, this thesis does not exceed the prescribed word limit for the relevant Degree Committee.

Eleanor Dunstone
2018

Acknowledgements

I would like to thank my supervisor, Dr. David Adams, for the opportunity to work in his group and for his ongoing help and guidance throughout the project. I would also like to acknowledge the other members of the experimental cancer genetics group for their support, with special thanks going to Dr. Marco Ranzani, Dr. Nicola Thompson and Gemma Turner for their advice and assistance. For their role in teaching me the bioinformatics skills required for this project, I would like to thank Aravind Sankar, Rashid Mamunur and Duncan Berger. I would also like to thank Dr. Vivek Iyer and Rashid Mamunur for their assistance in analysing the sequencing data generated from the whole genome sequencing and CRISPR-Cas9 screen. For sequencing library preparation, I would like to acknowledge the Cancer Genomes Project and Dr. Jonathan Cooper. I would also like to thank the cytogenetics team for performing the M-FISH analysis. Finally, I thank the rest of my friends and colleagues at the Wellcome Sanger Institute for making this year a thoroughly enjoyable and fulfilling experience.

Abstract

Malignant transformation is the transition of a cell from a normal state of proliferative homeostasis to a state of abnormal over-proliferation, acting as the initial step of tumourigenesis. This process is governed by the mutation of genes controlling cellular processes such as cell division, DNA replication and growth signaling. In this project, genome-wide forward genetic screening approaches were used to identify novel candidate genes involved in transformation. The model system used was the transformation-sensitive murine cell line NIH3T3, in which genes were assessed for their ability to initiate the formation of transformed foci of proliferation when mutated *in vitro*.

Firstly, the NIH3T3 genome was sequenced to characterise its genetic background and identify possible reasons for its transformation-sensitive phenotype. This was accompanied by Multiplex - Fluorescence *In Situ* Hybridisation to investigate large-scale genomic alterations. These approaches identified specific indels and single nucleotide variants in known cancer-associated genes, and large-scale genomic alterations, both of which may contribute to the transformation sensitivity of the cell line. The karyotype was discovered to be highly abnormal and heterogeneous, suggesting high chromosomal instability and continuing evolution of the line. This work has provided valuable insight into the limitations of this model and has implications for its use in this project and beyond.

The first genetic screening approach used was a pooled CRISPR-Cas9 genome-wide knock-out screen, identifying candidate tumour suppressor genes by generating loss-of-function mutations. Genes causing an increase in proliferative focus formation when knocked out were identified by sequencing the guide RNA population and identifying those that were overrepresented in the cultured cells using the algorithm Model-based Analysis of Genome-wide CRISPR-Cas9 Knockout. This identified putative genes associated with transformation, which were then compared with existing mutation data from human cancer sequencing efforts. This screen successfully identified some known cancer-associated tumour suppressor genes, along with potential novel candidates. As a complementary approach, a genome-wide transposon-based screen was also conducted, activating genes by inserting the CMV promoter at random throughout the genome using a *PiggyBac*-based transposon. Recovery of the insertion sites in the final cell population to locate sites that are overrepresented is currently in progress, aiming

to identify putative oncogenes.

While further work to validate the candidates identified is needed, this work has made some progress towards identifying novel transformation-associated genes. If these genes can be validated, they may provide useful insights into the biology of early tumourigenesis, informing further research and possible therapeutic targets.

Contents

Contents	xi
List of Figures	xv
List of Tables	xvii
Nomenclature	xviii
1 Introduction	1
1.1 Cancer genetics: a brief overview	1
1.2 Malignant transformation	2
1.3 Models of malignant transformation	3
1.3.1 NIH3T3 cells	5
1.4 Genetic screening tools	5
1.4.1 CRISPR-Cas9	5
1.4.2 Transposons	7
1.5 Overall aims	8
1.6 Abbreviations	9
1.7 Thesis Overview	9
2 Genomic analysis of NIH3T3 and NIH3T3-Cas9 cells	11
2.1 Introduction	11
2.1.1 Aims	12
2.2 Materials and methods	13
2.2.1 Materials	13
2.2.2 Methods	13
2.3 Results	17
2.3.1 Summary of variants in NIH3T3 wild-type	17

2.3.2	Comparison of NIH3T3 wild-type variants with Cancer Gene Census genes	17
2.3.3	Comparison of NIH3T3 wild-type variants with COSMIC	19
2.3.4	Comparison of NIH3T3 wild-type and NIH3T3-Cas9	24
2.3.5	Multiplex Fluorescence <i>In Situ</i> Hybridisation (M-FISH) of NIH3T3-Cas9	24
2.4	Discussion	28
2.4.1	Comparison of NIH3T3 wild-type genome with human cancer genome data	28
2.4.2	Large scale genomic alterations in NIH3T3-Cas9	29
2.4.3	Comparison of single nucleotide variants and indels in NIH3T3 wild-type and NIH3T3-Cas9	31
3	Identifying mediators of malignant transformation in cancer using genome-wide CRISPR-Cas9 knockout screening	33
3.1	Introduction	33
3.1.1	Aims	35
3.2	Materials and Methods	35
3.2.1	Materials	35
3.2.2	Methods	37
3.3	Results	47
3.3.1	Focus formation assay with pBabe-puro Ras-V12	47
3.3.2	Genome-wide CRISPR-Cas9 knockout screen for mediators of malignant transformation	49
3.3.3	Screen quality	50
3.3.4	Candidate genes	51
3.3.5	Prioritising genes for validation using existing cancer genome data	51
3.4	Discussion	58
4	Identifying mediators of malignant transformation in cancer using genome-wide transposon-based gene activation	63
4.1	Introduction	63
4.1.1	Aims	64
4.2	Materials and Methods	64
4.2.1	Materials	64
4.2.2	Methods	65
4.3	Results	67

4.4	Discussion	69
5	Conclusions and further directions	71
5.1	Characterisation of the genetic background of NIH3T3 and NIH3T3-Cas9 . .	71
5.2	Identification of candidate genes associated with malignant transformation .	73
	References	75
A	Software and databases	93
B	Supplementary information	95
B.1	Generation of the NIH3T3-Cas9 cell line	95
B.2	Cas9 activity determination in NIH3T3-Cas9	96
B.3	NIH3T3 wild-type variants with coding consequences overlapping mouse ho- mologues of CGC genes	97
B.4	Verification of amplified Genome-wide Knockout CRISPR Library v2	102
B.5	Primer sequences for CRISPR-Cas9 gRNA insert library preparation	103
B.5.1	1st round PCR - Genome-wide CRISPR-Cas9 knockout screen	103
B.5.2	1st round PCR - Validation (pooled gRNA lentivirus)	103
B.5.3	2nd round PCR	104
B.6	Genes identified by the genome-wide CRISPR-Cas9 knockout screen	104
B.7	Determination of NIH3T3 transfection efficiency	106

List of Figures

2.1	Schematic of NIH3T3 and NIH3T3-Cas9 whole genome sequence analysis	15
2.2	Putative non-germline variants in NIH3T3 wild-type	18
2.3	Multiplex Fluorescence In Situ Hybridisation of NIH3T3-Cas9	25
2.4	NIH3T3 karyotyping by murine multicolour banding	27
3.1	Schematic of genome-wide CRISPR-Cas9 knockout screen using NIH3T3-Cas9 cells	40
3.2	Focus formation assay using NIH3T3 cells transfected with pBabe-puro Ras-V12	48
3.3	Receiver Operating Characteristic curve for CRISPR-Cas9 knockout screen	50
3.4	Arrayed focus formation assays for genome-wide CRISPR-Cas9 screen validation	55
3.5	Comparisons between normalised gRNA read counts in genome-wide CRISPR-Cas9 knockout screen	59
3.6	Comparison between normalised gRNA read counts in the plasmid library and focus formation samples	60
3.7	Mean normalised read counts for gRNAs against <i>Kdsr</i> , <i>Nup160</i> and <i>Smu1</i>	60
4.1	Focus formation assay for negative and positive controls of the genome-wide transposon-based activation screen	68
B.1	Histogram of gRNA read counts from sequencing of amplified Genome-wide Knockout CRISPR Library v2	102

List of Tables

1.1	Abbreviations	9
2.1	Reagents used in the methods described in Chapter two	13
2.2	PCR programme for the library preparation for whole genome sequencing of NIH3T3 wild-type and NIH3T3-Cas9	14
2.3	NIH3T3 wild-type variants by consequence	19
2.4	NIH3T3 wild-type variants in mouse homologues of CGC genes	20
2.5	Indels and truncation mutations in mouse homologues of Cancer Gene Census genes in NIH3T3 wild-type	21
2.6	Missense SNVs in mouse equivalents of Cancer Gene Census genes in NIH3T3 wild-type	23
2.7	NIH3T3-Cas9-specific variants by consequence	26
2.8	NIH3T3-Cas9-specific variants overlapping mouse CGC gene homologues	26
3.1	Reagents used in Chapter 3	37
3.2	Transfection reagent quantities for transfection with Ras and GFP plasmids	38
3.3	Transfection reagent quantities for Genome-wide Knockout CRISPR Library v2 lentivirus production	39
3.4	Reagent quantities for the 1st round PCR in the CRISPR-Cas9 gRNA insert library preparation	42
3.5	PCR programme for the 1st round PCR in the CRISPR-Cas9 gRNA insert library preparation	42
3.6	Reagent quantities for the 2nd round PCR in the CRISPR-Cas9 gRNA insert library preparation	42
3.7	PCR programme for the 2nd round PCR in the CRISPR-Cas9 gRNA insert library preparation	43
3.9	Transfection reagent quantities for validation focus formation assays	44

3.8	DNA sequences encoding complementary gRNAs from the arrayed plasmid library used in validation	45
3.10	Transfection reagent quantities for validation virus production	46
3.11	Genes containing recurrent deletions in multiple tumour types	53
3.12	Genes for individual validation	53
3.13	Mean normalised read counts for gRNAs against <i>Kdsr</i> , <i>Nup160</i> and <i>Smu1</i>	58
4.1	Reagents used in chapter 4	65
4.2	Transfection reagent quantities used in genome-wide transposon-based gene activation screen	67
A.1	Software and databases used in analyses	93
B.1	Reagents used in the generation of the NIH3T3-Cas9 cell line	95
B.2	NIH3T3 wild-type coding variants in mouse homologues of CGC genes	97
B.3	Primer sequences for the 1st round PCR in the CRISPR-Cas9 gRNA insert library preparation for the genome-wide CRISPR-Cas9 knockout screen	103
B.4	Primer sequences for the 1st round PCR in the CRISPR-Cas9 gRNA insert library preparation for the validation using a pooled gRNA virus	103
B.5	Primer sequences for the 2nd round PCR in the CRISPR-Cas9 gRNA insert library preparation	104
B.6	Putative transformation-associated genes identified by the genome-wide CRISPR-Cas9 knockout screen	105
B.7	Reagents used in the determination of NIH3T3 transfection efficiency	107
B.8	Transfection reagent quantities for determination of NIH3T3 transfection efficiency	107