# Cas9-induced on-target genomic damage

**Michał Konrad Kosicki**

Supervisor: Prof. A. Bradley
Dr. S. Teichmann
Dr. M. Hemberg

Advisor: Prof. A. Ferguson-Smith

Wellcome Sanger Institute
University of Cambridge

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

Gonville & Caius College September 2018

# Declaration

I declare that his 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 Acknowledgements and specified in the text. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. This dissertation contains fewer than 60,000 words exclusive of tables, footnotes, bibliography, and appendices.

Michał Konrad Kosicki
September 2018

# Acknowledgements

Results presented in chapter 3 were obtained in collaboration with Rajan Sandeep, the first coauthor of the published book chapter. This work was supervised by Manos Metzakopian and Erik Bennett. Results related to culture of progenitor cells from murine bone marrow presented in chapter 5 were the work of Kärt Tomberg, second author on the published Nature Biotechnology article. Allan Bradley was the primary supervisor. Kärt, Allan and Jorge de la Rosa have read the thesis and offered many useful comments.

I would like to blame the members of Bradley group, in particular Mathias Friedrich and Kärt Tomberg, for making the lab a dangerously attractive place to spend my time in. Thanks to Ross Cook, Haydn Prosser, Alex Strong, Katta Hautaviita and Frances Law for their excellent technical support and putting up with millions of questions. I owe a lot to Manos Metzakopian for keeping my radar straight. Shout out to Lilliana Antunes, my 'thesis buddy', and Dimitris Garyfallos for always keeping my spirits up. Cheers to Rajan Sandeep for invigorating lab's social life outside of workplace. Thanks to everyone who participated in foosball matches, a surprisingly welcome distraction.

There is too many people to personally acknowledge at the Wellcome (Trust) Sanger Institute, who made my time here. Thank you to members of Teichmann lab (my second thesis supervisor) for socilizing and all the shady scientific projects we did together (hey, Johan!), to members of Hemberg lab (my third supervisor) and Trynka lab for lunch-time discussions and to all PhD students, in particular PhD14 cohort, for all the good times.

Two great mentors deserve a special mention: Jason Carroll, who inspired me to make the PhD leap and my External Advisor Anne Ferguson-Smith, who provided crucial advice in the time of need.

I am deeply grateful to all great people at Caius for reminding me that life is not all genetics, it is also linguistics, physics, neuroscience, law and port.

Last, but not least, thanks to Allan for being a great supervisor, for giving me a freedom to learn, for patience with my quasi-scientific ramblings and for supporting me even as I ignored his suggestions (mostly to my detriment).

# Abstract

CRISPR/Cas9 is the gene editing tool of choice in basic research and poised to become one in clinical context. However, current studies on the topic suffer from a number of shortcomings. Mutagenesis is often assessed using bulk methods, which means rare events go undetected, unresolved or are discarded as potential sequencing errors. Many of the genotyping methods rely on short-range PCR, which excludes larger structural variants. Other methods, such as FISH, do not provide basepair resolution, making the genotype assessment imprecise. Furthermore, it is not well understood how Cas9 delivery format influences the dynamics of indel introduction. Finally, many studies of on-target activity were conducted in cancerous cell lines, which do not accurately model the mutagenesis of normal cells in the therapeutic context.

In my thesis, I have investigated on-target lesions induced by Cas9 complexed with single gRNAs and no exogenous template. I have followed the time dynamics of Cas9-induced small indels as a function of reagent delivery methods, established an assay for quantification of Cas9-induced genomic lesions that are not small indels ("complex lesions") and used this assay to isolate and genotype complex lesions, many of which would be missed by standard methods. I found that DNA breaks introduced by single guide RNAs frequently resolved into deletions extending over many kilobases. Furthermore, lesions distal to the cut site and cross-over events were identified. Frequent and extensive DNA damage in mitotically active cells caused by CRISPR/Cas9 editing may have pathogenic consequences.

# Table of contents


**1 Introduction** **1**
1.1 Precise and efficient modification of DNA . . . 1
1.2 DSB repair . . . 3
1.2.1 Repair pathways . . . 4
1.2.2 Cell-cycle arrest, apoptosis and controlled DSB induction . . . 5
1.2.3 Diversity in cellular DNA repair . . . 6
1.3 Precision nucleases . . . 7
1.3.1 CRISPR – biology and applications . . . 7
1.3.2 Cas9 off-target problem . . . 9
1.3.3 Cas9 on-target damage . . . 10
1.4 Outstanding issues . . . 11

**2 Materials and Methods** **13**
2.1 Cell lines and cell culture . . . 13
2.2 Vectors . . . 13
2.3 Transfections, flow cytometry and sequencing . . . 13
2.4 Comparison of delivery methods in HEK cells . . . 15
2.5 PacBio sequencing and analysis . . . 15
2.6 Bioinformatics . . . 16

**3 Indel dynamics** **17**
3.1 Introduction . . . 17
3.1.1 Cas9 mutagenesis of genes . . . 17
3.1.2 Genotyping of small indels in bulk cell populations . . . 17
3.1.3 Factors influencing Cas9 mutagenesis and genotyping . . . 18
3.2 Results . . . 19
3.2.1 Transfection and construct integration dynamics . . . 19
3.2.2 Mutagenesis efficiency over time . . . 21
3.2.3 Indel profiles over time . . . 23
3.2.4 Dynamic effect of gRNA stability on indel efficiency . . . 23
3.2.5 Comparison of IDAA and TIDE methods . . . 24
3.3 Discussion . . . 24
3.3.1 Causes and consequences of mutagenesis efficiency fluctuation . . . 24
3.3.2 Delayed large indel formation . . . 24
3.3.3 Caveats of indel profiling . . . 26

**4 Detection and quantification of complex lesions** **27**
4.1 Introduction . . . 27
4.1.1 Mutation reporters . . . 27

4.2 Results . . . 28
4.2.1 Assay design . . . 28
4.2.2 Mouse *PigA* and human *PIGA* loci . . . 29
4.2.3 Autosomal *Cd9* locus . . . 32
4.3 Discussion . . . 34

**5 Genotyping of complex lesions** **39**
5.1 Introduction . . . 39
5.2 Results . . . 39
5.2.1 Deletions underlying loss of gene expression caused by intronic gRNAs . . . 39
5.2.2 Deletions in primary bone marrow cells . . . 44
5.2.3 Insertions . . . 45
5.2.4 Non-contiguous lesions . . . 48
5.2.5 Unexpected genotypes of inconsistent clones . . . 48
5.2.6 Diversity of resolved alleles at the *PigA* locus . . . 49
5.2.7 Diversity of deletion fingerprints at the *PigA* locus . . . 51
5.3 Discussion . . . 52
5.3.1 Consequences of large deletions . . . 52
5.3.2 Consequences of other complex lesions . . . 53
5.3.3 Stochasticity of large deletions . . . 54
5.3.4 Other considerations . . . 54

**6 Discussion** **57**
6.1 Causes of complex lesions . . . 57
6.2 Ways to avoid complex lesions . . . 58
6.3 Ways to exploit complex lesions . . . 59
6.4 Probing protein isoform diversity using CRISPR/Cas9-based assays . . . 60
6.5 New methods for genotyping of complex lesions . . . 60
6.6 Complex lesions and risk management . . . 60

**References** **63**

# Glossary

**BIR** break-induced replication.

**BL6** *Mus musculus*.

**CAST** *Mus musculus castaneus*.

**CRISPR** Clustered Regularly Interspaced Short Palindromic Repeats.

**DDR** DNA damage repair.

**DSB** double-stranded break.

**dsDNA** double-stranded DNA.

**ES** embryonic stem.

**gRNA** guide RNA.

**HR** homologous recombination.

**IDAA** Indel Detection by Amplicon Analysis.

**LOH** loss of heterozygosity.

**MMEJ** microhomology-mediated end-joining.

**NAHR** non-allelic homologous recombination.

**NGS** Next-Generation Sequencing.

**NHEJ** non-homologous end-joining.

**NMD** nonsense-mediated decay.

**PAM** protospacer adjacent motif.

**RNP** ribonucleoprotein.

**SNP** single-nucleotide polymorphism.

**SSA** single-strand annealing.

**ssDNA** single-stranded DNA.

**SSTR** single-strand template repair.

**TIDE** Tracking of Indels by DEcomposition.