WELLCOME TRUST SANGER INSTITUTE
THE UNIVERSITY OF CAMBRIDGE

# Exploring mutational signatures in twenty-one breast cancers

Serena Nik-Zainal

Murray Edwards College

**This dissertation is submitted for the degree of Doctor of Philosophy**
**on the**
**26th of September 2012**

**DECLARATION OF ORIGINALITY**

I declare that this thesis is my own work and has not been submitted in any form for another degree or diploma at any university or other institute of tertiary education. Information derived from the published and unpublished work of others has been acknowledged in the text and a list of references is given in the bibliography.

**CONTENTS**

**SUMMARY**

The set of somatic mutations observed in a cancer genome is the aggregate outcome of one or more biological processes that have been operative over the lifetime of a patient. Each biological process is characterised by the pattern of mutations that it leaves on the cancer genome or "signature" which is determined by the underlying mechanisms of DNA damage and of DNA repair that constitute the biological process.

In this thesis, I set out to extract the mutational signatures characterising the biological processes that have been operative in breast cancers. Catalogues of all classes of somatic mutation were generated from twenty-one whole-genome sequenced breast cancers using an integrated suite of bioinformatic algorithms which had been optimised for producing complete datasets with high sensitivity and specificity.

Mathematical methods were applied in order to extract underlying mutational signatures. Multiple distinct single-substitution, double-substitution and deletion signatures were unearthed by these analyses. Remarkably, these signatures were able to distinguish breast cancers from women with germline mutations in *BRCA1* and *BRCA2*, indicating how defects in homologous recombination leaves its mutagenic imprint on cancer genomes. Furthermore, an intriguing phenomenon of localised hypermutation characterised almost exclusively by cytosine mutations at TpC dinucleotides, demonstrating marked co-localisation with somatic rearrangements, was uncovered. These clusters of regional hypermutation were a frequent observation, occurring in thirteen out of the twenty-one breast cancers studied and have been termed *kataegis* (greek for showers/thunderstorms/towards the earth). The mechanism underlying this mutational signature is unknown. However, a role for the *APOBEC* family of cytidine deaminases in kataegis is proposed. Finally, integrated analysis of substitution mutations and expression data revealed the past operation of transcription-dependent mechanisms in generating the mutational profiles in these cancers.

This study harnesses the full scale of whole-genome sequencing demonstrating how detailed analyses of genomic data can provide biological understanding into hitherto unrecognised mutational signatures present in breast cancer genomes. In the future, the analyses of vast numbers of catalogues of somatic mutation from numerous worldwide cancer sequencing projects may herald further insights into mutational processes that underpin the development of cancers.

**ACKNOWLEDGEMENTS**

I would like to thank Professor Mike Stratton for welcoming me into his laboratory and for the patient manner in which he has supervised this work and guided me through research training. I would also like to acknowledge the advice and encouragement from the members of the Cancer Genome Project, in particular, Dr Peter Campbell and Dr Andy Futreal. I am indebted to the large number of people who have helped with technical wet-bench matters, provided informatic and administrative support, both within the Cancer Genome Project and the Wellcome Trust Sanger Institute. I would also like to acknowledge several key people who have helped me to in the intellectual development of this project: Ludmil B. Alexandrov, David Wedge and Peter Van Loo for informatic development and Michael Neuberger of the MRC Laboratory of Molecular Biology, Cambridge, for guidance on mutational signatures. Finally, I would like to thank my family, in particular, my husband who has been endlessly supportive, patient and cheerful, and who has counselled me on so many occasions, my children who have kept me smiling and my mother who has been ceaseless in her affection and her assistance.

This work is supported by the Wellcome Trust Clinical Research Training Programme in Cambridge.