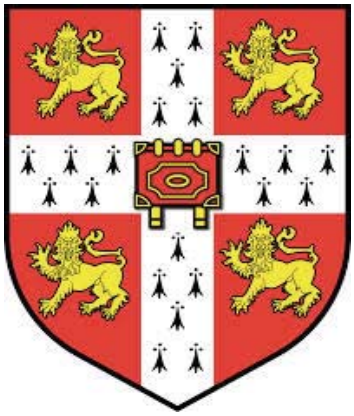
Signatures of Mutational Processes in Human Cancer



Ludmil B. Alexandrov

Darwin College
University of Cambridge

Thesis submitted for the degree of Doctor of Philosophy

July 2014

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 Information derived from the published and unpublished work of others has been acknowledged in the text and a list of references is given in the bibliography. The dissertation does not exceed the stipulated word limit of 60 000 words.

SUMMARY

All cancers originate from a single cell that starts to behave abnormally due to acquired somatic mutations in its genome. These somatic mutations may be the consequence of the intrinsic slight infidelity of the DNA replication machinery, exogenous or endogenous mutagen exposures, enzymatic modification of DNA, or defective DNA repair. In some cancer types, a substantial proportion of somatic mutations are known to be generated by exposures, for example tobacco smoking in lung cancers and ultraviolet light in skin cancers, or by abnormalities of DNA maintenance, for example defective DNA mismatch repair in some colorectal cancers. However, our understanding of the mutational processes that cause somatic mutations in most cancer classes has been remarkably limited.

Different mutational processes often generate different combinations of mutation types, termed “signatures.” There is strong evidence from analyses of known cancer genes in lung cancers and skin cancers that the classes of mutations found and their characteristics match those induced experimentally by tobacco carcinogens and ultraviolet light respectively, the known carcinogenic influences in these cancer types. Thus, the analysis of mutational signatures found in human cancers can provide clues to the processes that have been operative during their development.

In this thesis, I create a theoretical model describing the signatures of mutational processes operative in cancer genomes and develop a systematic computational framework to decipher mutational signatures from mutational catalogues of cancer genomes. The approach is extensively evaluated with simulated data and initially applied to 119 breast cancer whole-genome sequences and 844 breast cancer whole-exome sequences. Novel and known breast cancer mutational signatures are revealed and the contribution of each signature to each cancer sample is estimated.

After this initial application, I use the developed computational framework to perform a comprehensive analysis of cancer genomics data. The approach is applied to 4,938,362 somatic substitutions and insertion/deletions from 7,042 human cancers of 30 classes revealing more than 20 distinct mutational signatures. Some are present in many cancer types, notably a signature attributed to the *APOBEC* family of cytidine deaminases, whereas others are confined to a single cancer class. For some of these processes the underlying biological mechanism is unknown. However, some of

the identified mutational signatures associate to age of cancer diagnosis, smoking, UV light, anticancer drug exposure, presence of *BRCA1* and *BRCA2* mutations, and inactivation of mismatch repair genes.

This thesis provides both a basis for characterizing mutational signatures from cancer-derived somatic mutational catalogues and the first large-scale examination of mutational signatures across multiple cancer types. The results reveal the diversity of mutational processes underlying the development of cancer, with potential implications for understanding of cancer etiology, prevention, and therapy.

ACKNOWLEDGEMENTS

I would like to thank my PhD supervisor, Professor Sir Mike Stratton, for giving me the opportunity to work on this extremely exciting topic and the freedom to explore it from various different angles. Thank you Mike for patiently enduring all my whims and for your support throughout the years! I would have never been able to complete this PhD without your continuous guidance and encouragement. Working with you has been a privilege and an honor!

I would also like to especially thank Dr. Serena Nik-Zainal for the many fruitful discussions about mutational signatures and their biological meaning as well as Dr. Peter Campbell and Dr. David Wedge for all the interesting discussions about mathematical modeling of mutational processes and statistical analysis.

The completion of this PhD project would not have been possible without the advice and encouragement from the members of the Cancer Genome Project. I am also indebted to the hundreds of collaborators throughout the world that were interested in mutational signatures and were willing to share their data with me. I would also like to thank the cancer genomics community (and especially the members of The Cancer Genome Atlas and the International Cancer Genome Consortium) for making all their somatic mutations freely available.

This PhD would not have been possible without the love and support of friends and family. I would not been able to list everyone, but many thanks to Stefan, Elizabeth, Jimmy, Alison, Nicole, and Tessa. I would like to especially thank Oakleigh for her incredible help with proofreading the whole thesis and all her love, care, and patience throughout the past two years. Thank you for being there for me!

I would like to thank my sisters Iliana and Stoyana for always brightening my day and for bringing excitement to my life. I would also like to thank my mother Dora for believing in me and for her love, support, and understanding throughout the past four years. I would like to thank my grandma Lyudmila and my late grandma Lubka for their love and support. Special thanks to my father Boian for showing me the amazing world of science, for enduring me throughout the hardest times, for being there for me no matter what! Гръб до гръб двамина с тебе, бихме се с цял екипаж.

Lastly, I would like to thank the Wellcome Trust for generously funding the last four years and for allowing me to explore the world of cancer genomics.

Table of contents