Somatic evolution in human blood and colon

Henry Lee-Six
Trinity College
University of Cambridge

September, 2018

Dissertation submitted for the degree of Doctor of Philosophy

Summary

Somatic evolution in human blood and colon

Henry Lee-Six

All cancers were once normal cells. They became cancerous through the chance acquisition of particular somatic mutations that gave them a selective advantage over their neighbours. Thus, the mutations that initiate cancer occur in normal cells, and the normal clonal dynamics of the tissue determine a mutant cell's ability to establish a malignant clone; yet these remain poorly understood in humans. One tissue was selected for the exploration of each of these two facets of somatic evolution: blood for clonal dynamics; colon for mutational processes.

Blood presents an opportunity to study normal human clonal dynamics, as clones mix spatially and longitudinal samples can be taken. We isolated 140 single haematopoietic stem and progenitor cells from a healthy 59 year-old and grew them *in vitro* into colonies that were whole genome sequenced. Population genetics approaches were applied to this dataset, allowing us to elucidate for the first time the number of active haematopoietic stem cells, the rate at which clones grow and shrink, and the cellular output of stem cell clones.

Colonic epithelium is organised into crypts, at the base of which sit a small number of stem cells. All cells in a crypt ultimately share an ancestor in one stem cell that existed recently, and consequently share the mutations that were present in this ancestor. We exploited this natural clonal unit, isolating single colonic crypts through laser capture microdissection. 570 colonic crypts from 42 individuals were whole genome sequenced. We describe the burden and pattern of somatic mutations in these genomes and their variability across and within different people, identifying some mutational processes that are ubiquitous and others that are sporadic. Targeted sequencing of an additional 1,500 crypts allowed us to quantify the frequency of driver mutations in normal human colon.

Together, these two studies inform on the somatic evolution of normal tissues, describing new biology in human tissue homeostasis and providing a window into the processes that govern cancer incidence.

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as 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 other qualification at the University of Cambridge or any other university. It does not exceed the prescribed limit of 60,000 words.

Description of contributions

This work was collaborative in nature. Here, I outline the contributions that others have made to work presented in this dissertation. I further draw attention to their contributions at relevant points throughout the text.

Results Chapter 1: Clonal dynamics of normal blood.

The general experimental design, as displayed in figure 1.1a, was formulated by Peter Campbell, David Kent, and Tony Green. The bone marrow aspirate was performed by Brian Huntly. Flow sorting of the bone marrow aspirate and culturing colonies was performed by the Kent lab. Subsequent peripheral blood draws and flow sorting of peripheral blood were performed by members of the Green and Kent labs.
All bioinformatic analyses were performed by me, with the following exceptions:

- Sebastian Grossman ran the Shearwater and HipSTR algorithms, and most iterations of the SCITE tree building algorithm.
- Robert Osborne wrote and ran an algorithm to tabulate reads from deep sequencing of peripheral blood.
- Peter Campbell wrote the mixed effects model to separate signal from noise in the targeted sequencing data from peripheral blood.

My contributions to the project were the following:

- Contributing to decisions about which peripheral blood draws should be sequenced.
- Devising the filters to clean the whole genome sequencing data and the method of calling mutation.
- Analyses of mutational signatures and embryology of blood, and the relationship between blood cells on the phylogeny.
- Conceiving the method of estimating stem cell numbers by performing a capture-recapture experiment of different peripheral blood samples.
- With substantial guidance from Peter Campbell and Kevin Dawson, writing the approximate Bayesian computation for estimating the number of active haematopoietic stem cells.

- Comparing the clonal contributions to granulocytes and lymphocytes.

Results Chapter 2: Mutational landscape of normal colon.

The idea of using laser capture microdissection to isolate normal crypts for sequencing was Mike Stratton's. Collaborators kindly provided frozen biopsies of colonic tissue. The organoid component of this study was set up by Sam Behjati and Sophie Roerink in collaboration with the Clevers lab, and the organoids were derived by the Clevers lab. A protocol to sequence small amounts of laser capture microdissected material was devised by Peter Ellis.

My contributions to the project were the following:

- Designing, through discussions with Mike Stratton, the experiment in terms of the number of crypts from each patient on which to perform targeted and whole genome sequencing.
- Setting up collaborations to obtain tissue for microdissection.
- Devising a method for fixing, staining, sectioning, and dissecting crypts.
- Microdissecting all the crypts.
- All analyses of the crypts and organoids, including, principally:
    - Calling mutations.
    - Driver analysis.
    - Mutational signature analysis.
    - Comparisons between normal and cancer.

For both projects, conclusions are either entirely my own or have arisen through frequent discussions over three years with many lab members, but principally Peter Campbell and Mike Stratton.

Acknowledgements

With warm and sincere gratitude, I thank my doctoral supervisors Peter Campbell and Mike Stratton. They gave me incredibly exciting projects to work on and put everything in place that I might carry them out. Together, we pored over novel findings, and spent, cumulatively, hundreds of hours discussing their implications and the next steps. Observing and imitating their distinctive styles of engaging with data was the best teaching I have ever had. I thank them for allowing me both to explore my own ideas and for guiding me back in the right direction when I had lost my way.

This thesis would not have been the same without the hundreds of conversations, in corridors and over coffee as much as in formal meetings, with many people. Inigo Martincorena was, in my first year, my 'go-to post doc' for understanding the deeper implications of what we were attempting to do; later, as a principal investigator, he was always willing to provide thoughtful and thought-provoking explanations for our discoveries. David Kent helped me navigate the complex literature of haematopoiesis when I was at sea, and so appreciate the broader implications of our findings. Rob Osborne was a friendly source of encouragement and level-headed advice who gave me a structure to follow when nothing was working in the lab; without him I may have given up on the approach that was ultimately fruitful. Kevin Dawson and I spent many hours on the blue sofa seemingly going around in circles over the mathematics of determining the number of haematopoietic stem cells; in retrospect, we were spiralling upwards to a conclusion. I learnt from Moritz Gerstung in our joint weekly statistics meetings (as well as from Peter, Inigo, and Kevin) how to look at statistics critically and constructively. I thank Grace Collord for her unfailing encouragement and advice (and cake), but especially for painstakingly reading this thesis; her comments improved it substantially. All were also good companions.

I am indebted to all those with whom I had the pleasure of working over the last three years. Their enthusiasm made me look forward to going into the lab. In no particular order, I thank: Seb, whose rational approach was invaluable to the construction of the stem cell phylogeny; Luiza, whose impetus drove the laser capture microdissection protocol; Simon and Heather, both good sounding boards and running companions; Sophie, who generously allowed

me to play with her data in what turned out to be a fruitful collaboration; Sam, who encouraged me to apply to the Sanger and set the bar high for what one should achieve in a PhD; Gene, a friend and purveyor of late-night biscuits; Phil, who engaged with the colon project and helped with microdissection; Nicola and Dominik, who helped me start coding; Tom, who taught me some of the more complex statistical analyses; Jyoti, Emily and Francesco, for our discussions about haematopoiesis; Kenichi, for good-naturedly relieving me of other work so that I could focus on my thesis; Nicos, Sandro, Federico, Tim C., Dan, and Mathijs for statistical advice; Alex, Tim B., Yvette, and Liz for their advice with laser capture microdissection; Wendy, Jo, Calli, Laura, and Claire, for their incredibly efficient, reliable, and helpful administrative support; and others that I cannot list here due to space constraints.

Finally, I could not have written this dissertation without the support of my family. My parents, brother, and partner were always willing to lend an ear to my trials and tribulations. They encouraged me constantly, and, at the busiest times, supported me in the reproduction of daily life and allowed me to put my work first. My mother and brother Edward were dedicated and effective proof-readers of this dissertation. I am, above all, deeply grateful to my partner, Tess. In many ways, she bore the brunt of the PhD, enduring my erratic hours and crises of confidence, and, throughout, helped me to keep my sanity. This dissertation is also theirs.

Contents