Structural variations as a method for phylogenetic reconstruction of tumour evolution

THE UNIVERSITY OF MELBOURNE

MAREK CMERO[1,4,9], Matthew KH Hong[1,2], Clare Sloggett[7], Sebastian Lunke[8], Jan Schröder[15], Kangbo Mo[9], Andrew Lonie[7], Haroon Naeem[4], Nikhil Sapre[1,2], Pramit Phal[3], Xiaowen Chin[1,2], Michael Kerger[1,2], John S Pedersen[5,6], Andrew Ryan[5], Izhak Haviv[10], Anthony J Costello[1,2], Niall M Corcoran[1,2], Christopher M Hovens[1,2], Cheng Soon Ong[12,13,14], Geoff Macintyre[9,11]

[1]Division of Urology, Department of Surgery, University of Melbourne, [2]Royal Melbourne Hospital and the Australian Prostate Cancer Research Centre Epworth, Richmond VIC, [3]Department of Radiology, Royal Melbourne Hospital, [4]Centre for Neural Engineering, University of Melbourne, [5]TissuPath Specialist Pathology, Mount Waverley, VIC, [6]Monash University Faculty of Medicine, Victoria, Australia, [7]Victorian Life Sciences Computation Initiative, University of Melbourne, [8]Department of Pathology, University of Melbourne, [9]Department of Computing and Information Systems, University of Melbourne, [10]Bhar Ilan University, Israel, [11]Cancer Research UK, Cambrige, [12]NICTA Canberra Research Laboratory, ACT, [13]Department of Computer Science, Australian National University, [14]Department of Electrical Engineering, University of Melbourne, [15]The Walter and Eliza Hall Institute, Parkville

Motivation

Tumour evolution is complex and multi-faceted. Understanding this process is paramount in informing patient outcomes and identifying patterns of progression. Several methods exist to derive tumour phylogenies, for instance single-nucleotide variations (SNVs) and copy-number variations (CNVs). These methods vary in effectiveness depending on the mutational landscape and progression of the cancer. As many prostate cancers are driven by structural variations (SVs), considering rearrangments and other small-scale genomic breaks provides a powerful approach to trace cancer evolution. Our approach allows us to identify groups of variations informative to the progression of metastatic cancers.

Key points summary

- We have developed a novel method to infer tumour phylogeny using structural variations alone (derived from whole-genome sequencing) from multiple samples of the same patient
- Our method allows us to assign particular structural variations to population branching events, which can be compared with the patient's observed disease progression
- This method has the potential to uncover common progression pathways characterised by rearrangements, inform potential methods of resistance and tumour evolution. Ultimately, this will help to inform patient outcomes.

Method overview

Data

- Whole genome sequencing (WGS) data of multiple tumour samples
- In this case study, we present a patient with a primary, 2 different metastatic regions, a local recurrance and a castrate resistant metastasis

Detecting & filtering structural variations

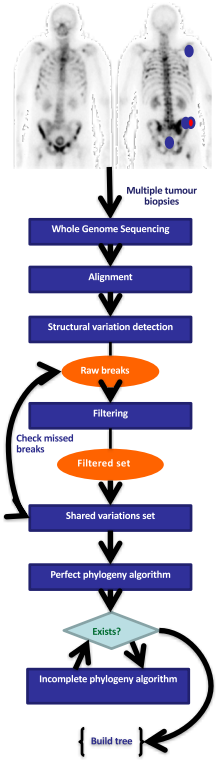
- Use Socrates[1] to detect structural variations (SVs) in the DNA with supporting read evidence at both breakpoint boundaries
- Filter out repeat artifacts, breaks with poor mapping quality, low fraction (vs. normal) of supporting reads, anomalous split-read sizes

Comparison

- Find common set of events in patient tumours, construct a matrix of shared events
- If event is missing in a sample, mine lower confidence data for support

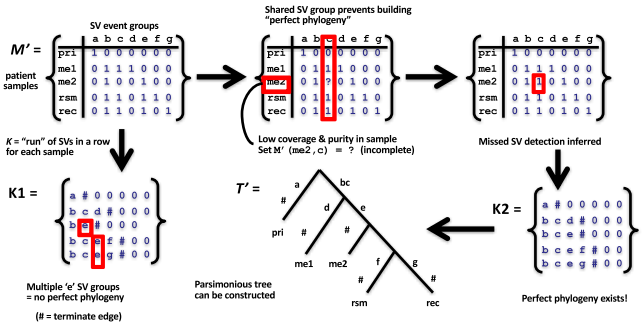
Phylogeny

- Collapse the events matrix to unique groups of breaks, remove any where the pattern is shared with <2% of breaks
- Construct a matrix of shared breakpoint events and use the Gusfield algorithm[2] to test if a 'perfect phylogeny' exists
- If no such phylogeny exists, look for 'incompatible edges' (groups of SVs that contradict a possible phylogeny)
- Where no phylogeny exists, run a heuristic to consider certain SV group fields as 'incomplete' for certain samples, and run incomplete phylogeny algorithm[3]
- Check if tree exists, if so, build tree



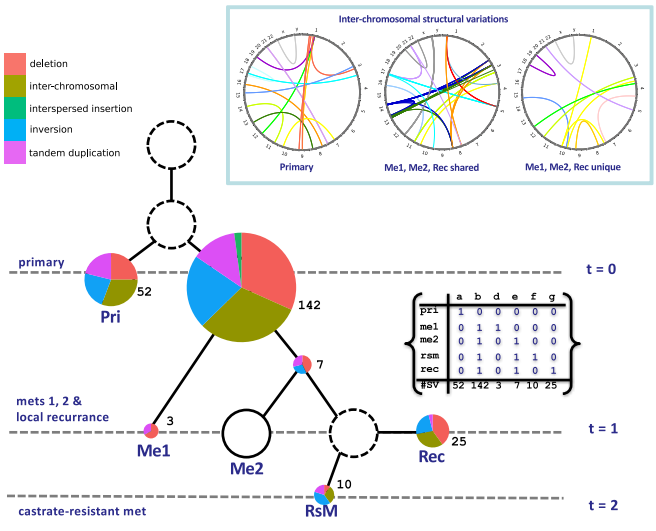
Phylogeny building algorithm

We collapse our shared SVs into a binary matrix of sample rows m and variation columns *n*. Then use the Gusfield[2] algorithm to check for perfect phylogeny and the Pe'er[3] algorithm to infer the presence of SVs (which may have been missed) in cells using a heuristic to identify SV groups of "conflict".



Preliminary results

We ran the algorithm on 5 tumour samples from the same patient, biopsied at 3 different time-points. The method allows us to group structural variations to distinct branches of the phylogenetic tree.



Future work & considerations

- As tumour samples can be made up of several tumour populations, we are working on a method to account for separating SVs as belonging to particular sub-clonal populations
- We plan to compare and contrast this method of tree building with SNV and CNV based methods

References

1. Schroeder, J., Hsu, A., Boyle, S. E., Macintyre, G., Cmero, M., Tothill, R.W., Johnstone, R. W. , Shackleton, M, Papenfuss, A. T. (2013). Socrates : Identification of genomic rearrangements in tumour genomes by re-aligning soft clipped reads
2. Gusfield, D. (1991). Efficient Algorithms for Inferring Evolutionary Trees, *95616*(2), 19–28.
3. Pe'er, I., Pupko, T., Shamir, R., & Sharan, R. (2004). Incomplete Directed Perfect Phylogeny. *SIAM Journal on Computing*, *33*(3), 590–607. doi:10.1137/S0097539702406510

Contact: Marek Cmero, Centre for Neural Engineering,
The University of Melbourne Parkville Victoria 3010, Australia