Abstract

Determining when cancer associated DNA mutations arise, and what their consequences are, is critical for a variety of clinical needs, such as the early detection of cancer. For a particular class of mutations, copy number alterations (CNAs), it appears that these often occur in a short time window, years before the onset of cancer. Why this happens is unclear, as is the growth advantage imbued by different CNAs to cancer cells. Using a probabilistic modelling approach, this project will investigate why CNAs occur rapidly in some cancer types, and measure the effect they have on tumour growth. This will involve simulation, analytic methods, and bioinformatic analyses. The computational modelling will be combined with state-of-the-art single cell DNA sequencing datasets.

Introduction

Copy number alterations (CNAs) are a class of mutations detected in 90% of tumours, with significance for cancer initiation [1] and treatment success [2]. In this project we will combine probabilistic modelling with cancer sequencing data to investigate two key aspects of CNA accumulation: 1) when CNAs occur throughout life, and 2) the oncogenic effect of specific CNAs.

Aim 1. Recent work has estimated that CNAs often arise decades before tumours are detected [3]. Curiously, certain cancers appear to acquire CNAs rapidly in a small time window, termed punctuated evolution (PE), while other cancers show more uniform CNA timings, termed gradual evolution. Patterns of PE could be inferred due to transient molecular mechanisms (e.g. breakage-fusion-bridge cycles), but could also be due to specific selective pressures, e.g. the rapid accumulation of an optimal genotype with stabilising selection [4].

Aim 2. We, and others, have recently shown that CNAs accumulate during tumour growth [5]. However, the growth advantage these mutations provide to cancer cells remains unclear, prohibiting their use as prognostic biomarkers.

Research Challenge

Aim 1: Infer CNA timings in single cell cancer data and assess whether clustered timings are driven by molecular mechanisms or selective effects.

Aim 2: Infer the growth advantages provided by CNAs *in-vivo* in a cancer type specific manner.

Data & Methodology

Data: we will compare against single cell DNA sequencing of ~1000 cells per patient, from 95 patients across 7 cancer types. Collaborators have generated and processed all data.

Methodology

Aim 1: We hypothesise that comparing CNA timing patterns versus the diversity of cancer cells, will allow us to discern between mutational mechanisms, or selection driven, PE patterns. Preliminary CNA timing for the data has already been performed. Various diversity measures will be computed on the data. To compare against the data, a stochastic model tracking lineages from conception to tumour sampling, modelling CNA accumulation with a mixed Poisson-process/branching process framework will be

developed. Through analytic and simulation based methods, CNA timing versus tumour diversity will be investigated for both variable CNA rates and stabilising selection. Via either approximate Bayesian computation, or likelihood methods based on the analytic results, model selection analysis will be carried out to assess whether the model with variable CNA rates, selection, or a combination of both best fits the data. This analysis will be carried out per tumour sample, and trends across cancer types will be sought.

Aim 2: Single cell DNA sequencing allows phylogenetic trees of the cancer cells to be inferred. Selectively advantageous CNAs will distort the balance of phylogenies, inflating the number of descendants carrying the alteration [6]. To quantify this signal, probabilistic models of cancer will be developed, incorporating cells dividing, dying, and the accumulation of CNAs which may lead to selective advantages. Single cells will be sampled from simulated tumour generated, which will be used for phylogenetic inference. Simulation output will be compared to data using approximate Bayesian computation, focusing on tree balance summary statistics. This will enable (i) inference of selection parameters in human cancers, (ii) identification of tree edges associated with the largest selective gains. Using this second inference, and CNAs that are assigned to tree edges, selective advantages will be related to specific CNAs. Orthogonal approaches will be used to verify the approach: CNAs associated with selective advantages are expected to contain amplification/deletions of tissue-specific oncogenes/tumour-suppressor-genes, public cancer mutation datasets will be used to annotate such mutations to CNAs gained down the trees.

RRI/Ethical Considerations

Care will be taken when reporting any clinically informative results, following recommended guidelines [7].

Expected Outcome & Impact

Bioinformatic methodology to quantify punctuated copy number evolution in state of the art genomic data and cancer specific classification of the mode of CNA evolution. Bioinformatic methodology to quantify tissue-dependent selective effects of CNAs. All work has potential impact on early cancer detection through insight on (i) when to look for certain oncogenic mutations, (ii) prioritising which mutations have largest oncogenic effect.

References

[1] Cross, W., Kovac, M., Mustonen, V. *et al.* The evolutionary landscape of colorectal tumorigenesis. *Nat Ecol Evol* **2**, 1661–1672 (2018). https://doi.org/10.1038/s41559-018-0642-z

[2] Jamal-Hanjani, M. *et al.* Tracking the Evolution of Non–Small-Cell Lung Cancer. *N Engl J Med 2017;376:2109-2121*

[3] Gerstung, M., Jolly, C., Leshchiner, I. *et al.* The evolutionary history of 2,658 cancers. *Nature* **578**, 122–128 (2020). https://doi.org/10.1038/s41586-019-1907-7

[4] Cross, W. Stabilising selection causes grossly altered but stable karyotypes in metastatic colorectal cancer (2020) bioRxiv doi: https://doi.org/10.1101/2020.03.26.007138

[5] Minussi, D.C., Nicholson, M.D., Ye, H. *et al.* Breast tumours maintain a reservoir of subclonal diversity during expansion. *Nature* **592**, 302–308 (2021). https://doi.org/10.1038/s41586-021-03357-x

[6] Househam, J., Heide, T., Cresswell, G.D. *et al.* Phenotypic plasticity and genetic control in colorectal cancer evolution. *Nature* **611**, 744–753 (2022). https://doi.org/10.1038/s41586-022-05311-

[7] Marilyn M. Li *et al.* Standards and Guidelines for the Interpretation and Reporting of Sequence Variants in Cancer A Joint Consensus Recommendation of the Association for Molecular Pathology, American Society of Clinical Oncology, and College of American Pathologists (2017) *The Journal of Molecular Diagnostics.* https://doi.org/10.1016/j.jmoldx.2016.10.002