Tumour phylogeny reconstruction from copy number profiles of temporal multi-region samples

Bingxin Lu, Kit Curtius, Kane Smith, Ibrahim Al Bakir, Trevor A. Graham, Chris P. Barnes

Introduction

Chromosomal instability (CIN) is a hallmark of cancer cells, causing structural or numerical chromosomal abnormalities and intra-tumour heterogeneity (ITH) [1] (Fig. 1). Understanding CIN evolution will enable fundamental understanding of the evolution of somatic genomes, which has wide-reaching health implications.

Normal genome

Genome with CIN

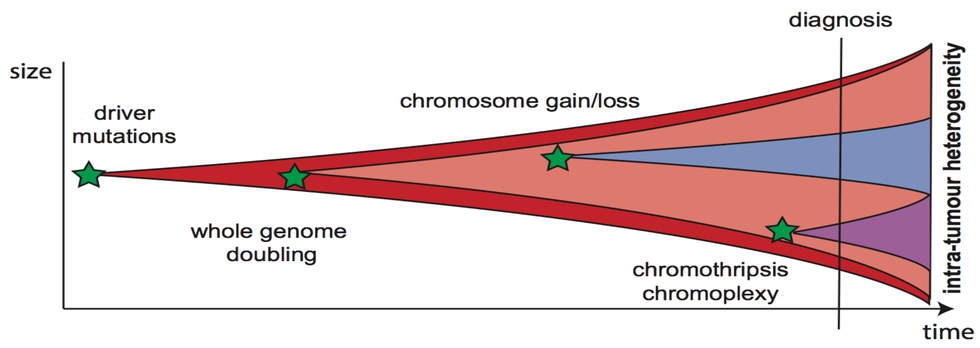

Figure 1. CIN can enable cancer genomes to evolve rapidly and generate ITH.

In this work, we aim to infer the evolutionary history of longitudinal healthy, dysplastic and tumour samples from total copy number profiles, which are called from multi-region low-pass WGS patient data.

Previous methods to address this problem are mostly distance matrix methods or maximum parsimony methods [2,3]. Very few methods used model-based approach [4]. To get more accurate and reliable inferences of tumour phylogeny, we are developing maximum likelihood and Bayesian methods based on a novel model of evolution.

Method

Given the total copy number profiles and sampling time information of all samples, our methods will infer the sample phylogeny and mutation rates of somatic copy number alteration (Fig. 2).

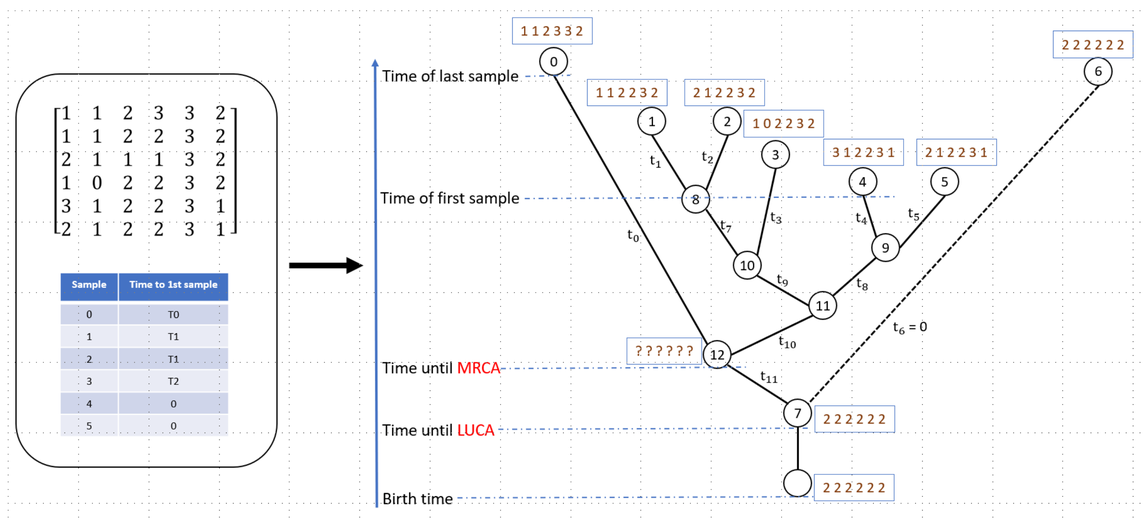

Figure 2. Illustration of our phylogeny inference method.

The key component in our inference of phylogeny is a novel model of evolution. We assume the profiles of copy number states follow a basic continuous-time non-reversible Markov model.

Since somatic chromosomal alteration events operate on different scales of the genome, we start with a substitution model of segment duplication and deletion and then add chromosomal gain, chromosomal loss and whole genome doubling (Fig. 3).

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 2e & -2(e+u) & 2u & 0 & 0 \\ 0 & 4e & -4(e+u) & 4u & 0 \\ 0 & 0 & 6e & -6(e+u) & 6u \\ 0 & 0 & 0 & 8e & -8e \end{bmatrix}$$

$$p(X \to Y) = p(\Delta g = 0)\prod_j \Big[\sum_z p(\Delta c_j = z)\prod_i \sum_{x\in k} p(x_{ij} + z \to y_{ij})\Big] + p(\Delta g = 1)\prod_j \Big[\sum_z p(\Delta c_j = z)\prod_i \sum_{x\in k} p(2x_{ij} + z \to y_{ij})\Big]$$

Figure 3. Top: The substitution rate matrix that specifies rate of copy number change. Bottom: the probability of copy number profile changing from X to Y.

From this model, we compute the likelihood of tree and find the tree with either maximum likelihood via heuristic method or highest posterior probability via MCMC approach.

Results

To validate our methods, we developed a program to simulate real data (Fig. 4). We inferred trees from the simulated copy number profiles and sampling times. Then we compared the simulated (true) trees and mutation rates with the inferred trees and mutation rates to see their differences (Fig. 5).

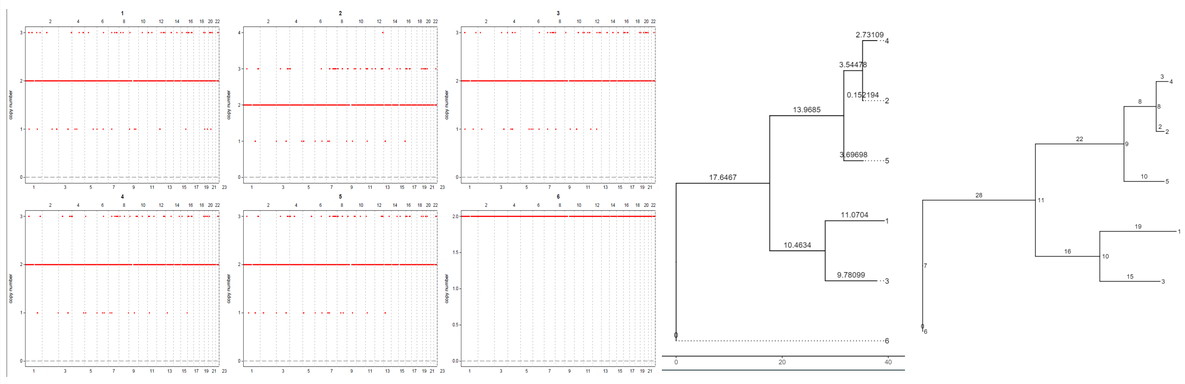

Figure 4. A simulated dataset with 5 samples. The segment duplication and deletion rate used for simulation are 6e-5 and 3e-5 respectively.

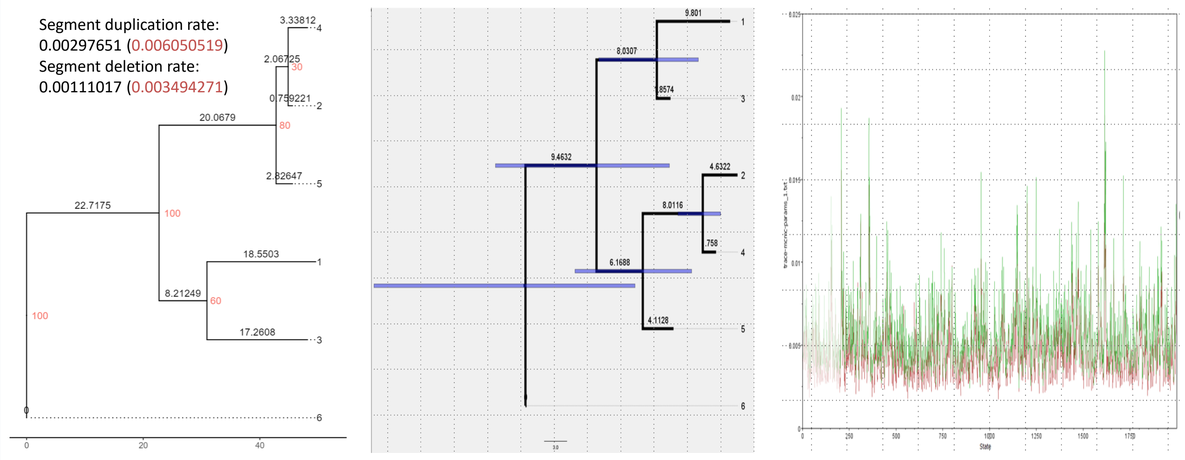

Figure 5. Phylogeny inference on the simulated dataset.

We applied our methods on total copy number profiles called from low-pass WGS patient data which were sampled at different time points and regions [5] (Fig. 6).

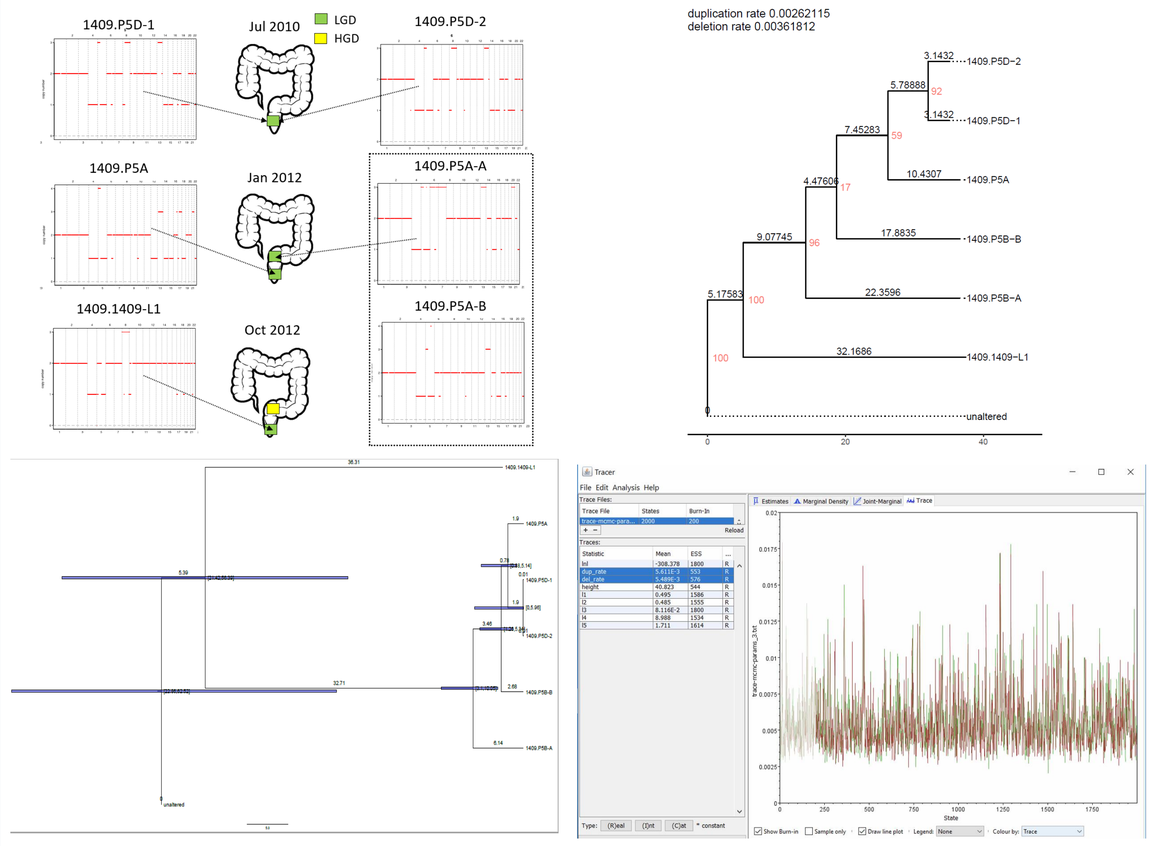

Figure 6. Phylogeny inference on a real dataset.

Conclusion

Our new method of phylogeny inference based on total copy number profiles of longitudinal multi-region samples may help to elaborate the evolutionary history of patient samples in terms of CIN.

The novel model of evolution provides a good way of integrating multiple levels of somatic chromosomal alteration events.

In future, we plan to apply our methods on total copy number profiles called from single cell sequencing data.

Reference

[1] Bakhoum and Cantley. Cell 174(6) (2018): 1347-1360.
[2] Schwarz et al. PLoS computational biology 10(4) (2014): e1003535.
[3] Letouzé et al. Genome biology 11(7) (2010): R76.
[4] Martinez et al. Nature communications 9(1) (2018): 794.
[5] Baker et al. *Gut* 68(6) (2019): 985-995.