We used mathematical modeling to estimate the relative timings of CNV events in each tumour. The modeling makes the following assumptions in addition to those mentioned explicitly below: (i) The average SNV mutation rate was equal in all of the regions considered (ii) The time between the last common ancestor of the cells in each tumour sample and the tumour removal was small (iii) The copy number state found by CloneHD for the major clone of each tumour sample was shared by all cells within the sample.

1 Growth model

Suppose that during the growth of tumour X an ancestor, c, of the cells in sample S gained an extra copy of region r of the A-allele of chromosome j. We note that any SNV's that occurred in an ancestor of c in region r on the A-allele of j will be present on 2 out of every 3 copies of this region in sample S. However, SNV's that occurred in an ancestor of the cell in region r on the B-allele, and SNV's that occurred in a descendent on the A-allele or B-allele will be present in only 1 of every 3 copies of the region [cite Durinck].

Formalising this reasoning, we model the number of high frequency SNV's in region r, α, and the number of low frequency SNV's in region r, β, as Poisson distributed random variables, with means $\lambda_{alpha} = \mu * l_i * \theta$ and $\lambda_{beta} = 2 * \mu * l * (\mathrm{T} - \theta)$. Where μ is the average SNV mutation rate, l is the length of region r in base pairs, θ is the time in years between the tumour diverging from the normal and the copy gain of r, and T is the time in years between the tumour diverging from the normal and the tumour removal.

We can further generalise to different types of copy number alteration that result in A copies of the A-allele (the more numerous allele) and B copies of the B-allele (the less numerous allele).

In general there are three cases:
Case I: $A \geq 2, B = 0$
Case II: $A = B \geq 2$
Case III: $A \geq 2, B = 1$
leading to Poisson distributions for α and β with mean parameters:
Case I: $\lambda_\alpha = l\mu\theta$, $\lambda_\beta = Al\mu(\mathrm{T} - \theta)$
Case II: $\lambda_\alpha = 2l\mu\theta$, $\lambda_\beta = 2Al\mu(\mathrm{T} - \theta)$
Case III: $\lambda_\alpha = l\mu\theta$, $\lambda_\beta = Al\mu(\mathrm{T} - \theta) + l\mu\mathrm{T}$

1.1 Fitting the model

For each tumour W, for each sample S, we used CloneHD to obtain a copy number state across the genome, giving the number of copies of the A-allele (defined as the more numerous allele), and the B-allele (defined as the less numerous allele) at any given locus. We segmented the genome into R regions so that all samples had uniform copy number states within each region.

For each region r_i with at least one non-diploid sample, we assumed that a single copy number alteration event had occurred in an ancestor of the non-

diploid samples in region r_i. We assumed the event had resulted in A_i copies of the A-allele and B_i copies of the B-allele, where A_i and B_i are the modal number of copies of the A-allele and B-allele respectively across the non-diploid samples.

For those events which conformed to one of the three cases above, we clustered SNV's detected in the region in each non-diploid sample into a high frequency set and a low frequency set, using the R package 'mclust'. We then calculated a consensus number of high frequency mutations α_i and a consensus number of low frequency mutations β_i, for the CNV event, [by taking the average across samples where the clustering was successful. Regions where the clustering was unsuccessful in all non-diploid samples were excluded] [WC: what were the cases where the clustering gives NA?].

For each region i with at least one normal sample, we calculated γ_i, the average number of SNV's along the length l_i of the region, across the diploid samples. We then took the diploid length l_d and the diploid mutation burden γ_d as the sum of these l_i and γ_i respectively.

Define T_0 as the time when W diverged from the normal sample, $\theta_1, \theta_2, ..., \theta_N$ as the times from T_0 to each of the N CNV events considered, and T as the time from T_0 to the point of surgery. Scaling these time parameters by the mutation rate, for each i we define $t_i = \theta_i \mu$, and define $T = \mathrm{T} * \mu$. Further we define $C_i := \{i : \text{CNV } i \text{ falls under Case } i\}$

We can then give a joint-likelihood $\mathscr{L}$ of the data in terms of $\mathbf{t} = (t_1, t_2, ..., t_N), T$:

$$\mathscr{L} = \prod_{i \in C_1} e^{-l_i t_i} \frac{(l_i t_i)^{\alpha}}{\alpha!} e^{-A_i l_i (T - t_i)} \frac{(A l_i (T - t_i))^{\beta_i}}{\beta_i!} \times \prod_{i \in C_2} e^{-2 l_i t_i} \frac{(2 l_i t_i)^{\alpha}}{\alpha!} e^{-2 A_i l_i (T - t_i)} \frac{(2 A_i l_i (T - t_i))^{\beta_i}}{\beta_i!}$$
$$\times \prod_{i \in C_3} e^{-l_i t_i} \frac{(l_i t_i)^{\alpha_i}}{\alpha_i!} e^{-(A_i l_i (T - t_i) + l_i T)} \frac{(A_i l_i (T - t_i) + l_i T)^{\beta_i}}{\beta_i!} \times e^{-2 l_d T} \frac{(2 l_d T)^{\gamma}}{\gamma!} \tag{1}$$

In the next section we seek to maximise this likelihood subject to the constraints $0 \leq \mathbf{t} \leq T$

1.2 Likelihood Maximisation

We seek $\max_{0 \leq \mathbf{t} \leq T} \mathscr{L}$. This leads to the Lagrangian:

$$L = \mathscr{L}(\boldsymbol{t}, T) + \lambda_0 T + \sum_{i \in I} \lambda_i t_i + \sum_{i \in I} \lambda_{N+i} (T - t_i) + \lambda_{2N+1} T \tag{2}$$

and the Kuhn-Tucker conditions

$$\text{For all i} \in \text{I}, \quad \frac{\delta L}{\delta t_i} = 0, t_i \geq 0, \lambda_i \geq 0, \lambda_i t_i = 0$$
$$T - t_i \geq 0, \lambda_{N+i} \geq 0, \lambda_{N+i}(T - t_i) = 0$$
$$\text{and} \quad \frac{\delta L}{\delta T} = 0, T \geq 0, \lambda_{2N+1} \geq 0, \lambda_{2N+1} T = 0$$

For all $i \in I$, $t_i = 0 \implies \mathscr{L} = 0$. There are at least some admissible points where $\mathscr{L} > 0$, so for all $i \in I$, $0 < t_i$ and therefore $\lambda_i = 0$. In addition, since $t_i \leq T$ for all i, $0 < T$ and $\lambda_{2N+1} = 0$.

Similarly, for all $i \in C_1 \cup C_2$, $t_i = T \implies \mathscr{L} = 0$. so in these cases $t_i < T$ and therefore $\lambda_{N+i} = 0$.

So we can simplify 2

$$L = \mathscr{L}(\boldsymbol{t}, T) + \sum_{i \in C_3} \lambda_{N+i}(T - t_i) \tag{3}$$

Our strategy will now be to express each t_i in terms of T and then solve for T.

Consider $i \in C_1$:

$$\frac{\delta L}{\delta t_i} = 0$$
$$\implies \frac{\delta \mathscr{L}}{\delta t_i} = 0$$
$$\implies \frac{\delta log(\mathscr{L})}{\delta t_i} = 0$$

Since $log(\mathscr{L}) = (A_i - 1)l_i + \alpha_i log(t_i) + \beta_i log(T - t_i) + C$, where C does not depend on t_i

$$\implies (A_i - 1)l_i + \frac{\alpha_i}{t_i} - \frac{\beta_i}{(T - t_i)} = 0$$
$$\implies (A_i - 1)l_i(T - t_i)t_i + \alpha(T - t_i) - \beta t_i = 0$$
$$\implies (1 - A_i)l_i t_i^2 + ((A_i - 1)lT - \alpha - \beta)t_i + T\alpha_i = 0$$
$$\implies t_i = \frac{((1 - A_i)lT + \alpha_i + \beta_i) \pm \sqrt{((1 - A_i)lT + \alpha_i + \beta_i)^2 - 4(1 - A_i)lT\alpha_i}}{2(1 - A_i)l_i}$$

Since $-4(1 - A_i)l_i T\alpha_i > 0$, $2(1 - A_i)l_i < 0$ and $t_i > 0$

$$\implies t_i = \frac{((1-A_i)lT + \alpha_i + \beta_i) - \sqrt{((1-A_i)lT + \alpha_i + \beta_i)^2 - 4(1-A_i)lT\alpha_i}}{2(1-A_i)l_i} \tag{4}$$

Consider $i \in C_2$:,

$$\frac{\delta L}{\delta t_i} = 0$$
$$\implies \frac{\delta \mathscr{L}}{\delta t_i} = 0$$
$$\implies \frac{\delta log(\mathscr{L})}{\delta t_i} = 0$$

Similar to the above

$$\implies 2(A_i - 1)l_i + \frac{\alpha_i}{t_i} - \frac{\beta_i}{(T - t_i)} = 0$$

By a similar argument to the above

$$\implies t_i = \frac{(2(1-A_i)lT + \alpha_i + \beta_i) - \sqrt{(2(1-A_i)lT + \alpha_i + \beta_i)^2 - 8(1-A_i)lT\alpha_i}}{4(1-A_i)l_i} \tag{5}$$

Consider $i \in C_3$:

$$\frac{\delta L}{\delta t_i} = 0$$
$$\implies \mathscr{L}((A_i - 1)l + \frac{\alpha_i}{t_i} - \frac{A_i\beta_i}{(A_i+1)T - A_i t_i}) - \lambda_{N+i} = 0$$

Since $\mathscr{L} > 0$

$$\implies (A_i - 1)l + \frac{\alpha_i}{t_i} - \frac{A_i\beta_i}{(A_i+1)T - A_i t_i} = \frac{\lambda_{N+i}}{\mathscr{L}}$$

If $t_i < T$, since $\lambda_{N+i}(T - t_i) = 0$, we have $\lambda_{N+i} = 0$. So

$$(A_i - 1)l_i + \frac{\alpha_i}{t_i} - \frac{A_i\beta_i}{(A_i+1)T - A_i t_i} = 0$$

and similar to Cases 1 and 2

$$t_i = \frac{((1-A_i)(A_i+1)lT + A_i(\alpha_i + \beta_i))}{2(1-A_i)A_i l_i} - \frac{\sqrt{((1-A_i)(1+A_i)lT + A_i(\alpha_i+\beta_i))^2 - 4(1-A_i)A_i(A_i+1)lT\alpha_i}}{2(1-A_i)A_i l_i} \tag{6}$$

Moreover

$$0 = \frac{\lambda_{N+i}}{\mathscr{L}} = (A_i - 1)l_i + \frac{\alpha_i}{t_i} - \frac{A_i\beta_i}{(A_i+1)T - A_i t_i} > (A_i - 1)l_i + \frac{\alpha_i - A_i\beta_i}{T}$$

$$\implies T < \frac{A_i\beta_i - \alpha_i}{(A_i - 1)l_i}$$

Whereas if $t_i - T = 0$, then

$$t_i = T \tag{7}$$

$$\frac{\lambda_{N+i}}{\mathscr{L}} = (A_i - 1)l_i + \frac{\alpha_i}{t_i} - \frac{A_i\beta_i}{(A_i+1)T - A_i t_i} = (A_i - 1)l_i + \frac{\alpha_i - A_i\beta_i}{T} \tag{8}$$

Moreover, since $\lambda_{N+i} \geq 0$

$$0 \leq \frac{\lambda_{N+i}}{\mathscr{L}} = (A_i - 1)l_i + \frac{\alpha_i}{t_i} - \frac{A_i\beta_i}{(A_i+1)T - A_i t_i} = (A_i - 1)l_i + \frac{\alpha_i - A_i\beta_i}{T}$$

$$\implies T \geq \frac{A_i\beta_i - \alpha_i}{(A_i - 1)l_i}$$

So defining $T_{crit_i} := \frac{A_i\beta_i - \alpha_i}{(A_i-1)l_i}$; if $T < T_{crit_i}$ then 6 holds, and if $T \geq T_{crit_i}$ then 7 and 8 hold.

Defining $C_{3_1}(T) := \{i : T < T_{crit_i}\}$ and $C_{3_2}(T) := \{i : T \geq T_{crit_i}\}$, we can now give a piecewise determination of $\frac{\delta\mathscr{L}}{\delta T}$ in terms of T only, with the determination depending on which of the possible $|C_3| + 2$ intervals, defined by the $|C_3|$ values of T_{crit} , contains T

$$\begin{aligned}\frac{\delta\mathscr{L}}{\delta T} = \mathscr{L}(&-(\sum_{i\in C_1} A_i l_i + \sum_{i\in C_2} 2A_i l_i + \sum_{i\in C_3}(A_i+1)l_i + 2l_d) \\ &+ (\sum_{i\in C_1}\frac{\beta_i}{T - t_i} + \sum_{i\in C_2}\frac{\beta_i}{T - t_i} + \sum_{i\in C_3}\frac{(A_i+1)\beta_i}{(A_i+1)T - A_i t_i} + \frac{\gamma}{T})) \\ &+ \sum_{i\in C_{3_1}} \lambda_{N+i}\end{aligned} \tag{9}$$

So we have

$$\begin{aligned}0 &= \frac{\delta\mathscr{L}}{\delta T} \\ \implies 0 &= \sum_{i\in C_1}\frac{\beta_i}{T - t_i} + \sum_{i\in C_2}\frac{\beta_i}{T - t_i} + \sum_{i\in C_3}\frac{(A_i+1)\beta_i}{(A_i+1)T - a_i t_i} + \frac{\gamma}{T} \\ &- \big(\sum_{i\in C_1} A_i l_i + \sum_{i\in C_2} 2A_i l_i + \sum_{i\in C_3}(A_i+1)l_i + 2l_d\big) + \sum_{i\in C_{3_2}}\frac{\lambda_{N+i}}{\mathscr{L}}\end{aligned}$$

Substituting from 7 and 8

$$\implies 0 = \sum_{i\in C_1} \frac{\beta_i}{T-t_i} + \sum_{i\in C_2} \frac{\beta_i}{T-t_i} + \sum_{i\in C_3} \frac{(A_i+1)\beta_i}{(A_i+1)T - A_i t_i} + \frac{\gamma}{T}$$
$$- \big(\sum_{i\in C_1} A_i l_i + \sum_{i\in C_2} 2A_i l_i + \sum_{i\in C_3} (A_i+1)l_i + 2l_d \big) + \sum_{i\in C_{3_2}} ((A_i-1)l_i + \frac{\alpha_i - A_i\beta_i}{T})$$

$$\implies 0 = \sum_{i\in C_1\cup C_2} \frac{\beta_i}{T-t_i} + \sum_{i\in C_{3_1}} \frac{(A_i+1)\beta_i}{(A_i+1)T - A_i t_i} + \sum_{i\in C_{3_2}} \frac{\alpha+\beta_i}{T} + \frac{\gamma}{T}$$
$$- \big(\sum_{i\in C_1} A_i l_i + \sum_{i\in C_2} 2A_i l_i + \sum_{i\in C_{3_1}} (A_i+1)l_i + \sum_{i\in C_{3_2}} 2l_i + 2l_d \big) \quad (10)$$

It is easily shown that this function is continuous by showing it is continuous at the piecewise breakpoints. Moreover we now show it is decreasing in T so that it has at most one solution.

The right summand is a negative constant. The left summand is a sum of fractions with constant numerators. Therefore it suffices to show that denominators of the fractions in the left summand are all increasing.

Consider $i \in C_1$

From 4

$$T - t_i = \frac{((1-A_i)lT - \alpha_i - \beta_i) + \sqrt{((1-A_i)lT + \alpha_i + \beta_i)^2 - 4(1-A_i)lT\alpha_i}}{2(1-A_i)l_i}$$

Let $x = (1-A_i)l$, then

$$T - t_i = \frac{(xT - \alpha_i - \beta_i) + \sqrt{(xT + \alpha_i + \beta_i)^2 - 4xT\alpha_i}}{2x}$$

Suppose for a contradiction that $\frac{\delta(T-t_i)}{\delta T} < 0$

$$\implies \frac{1}{2} + \frac{(2x(xT + \alpha_i + \beta_i) - 4x\alpha_i)}{4x\sqrt{(xT + \alpha_i + \beta_i)^2 - 4xT\alpha_i}} < 0$$
$$\implies \frac{(xT + \alpha_i + \beta_i) - 2\alpha_i}{\sqrt{(xT + \alpha_i + \beta_i)^2 - 4xT\alpha_i}} < -1$$
$$\implies \sqrt{(xT + \alpha_i + \beta_i)^2 - 4xT\alpha_i} < -(xT - \alpha_i + \beta_i)$$
$$\implies x^2T^2 + (\beta_i + \alpha_i)^2 + 2xT(\beta_i - \alpha_i) < x^2T^2 + (\beta_i - \alpha_i)^2 + 2xT(\beta_i - \alpha_i)$$
$$\implies 4\alpha_i\beta_i < 0$$

Which contradicts the fact that in all cases α_i and β_i are greater than 0

The other denominators can all be shown to be increasing by similar reasoning.

In all cases we are able to find a root of 10, numerically, and can thus be confident that it is the only root.

2 Confidence intervals

We used bootstrapping to calculate mean square errors for each of the estimated parameters $(\mathbf{t}, T)$. Using the mathematical model described above parameterised by the estimates for $(\mathbf{t}, T)$, we generated 100 simulated mutation data-sets, and in each case used the pipeline to re-estimate $(\mathbf{t}, T)$ from the simulated data. We then calculated the mean square error of these results compared to the original estimates used for the simulation.