Supplementary Tables

Table S1. Basic patient information

A summary of tumour names, biopsy numbers, sequencing technology and notes for synchronous carcinomas etc.

Table S2. Sequencing data summary

A summary of sequencing information for each biopsy, including; the total number of sequenced reads, the resulting sequencing depth (median depth: exome=63X; genome=25X), and the number of total and exonic mutations as identified through the Platypus caller (see Online Methods).

Table S3. Identified driver mutations

Possible driver mutations in each of the 24 tumours from the CRC-specific and pan-cancer driver gene analysis are listed. Detailed annotations are derived from AnnoVar (see Online Methods). Phylogenetic location (T, B, L) is given for each mutation, corresponding to trunk, branch and leaf, and indicating whether the driver was found ubiquitously or in a subset of biopsies. A small number of homoplasmic variants, entirely in adenomas or MSI+ cancers, are shown as “H”. These variants are typically present in two or more regions of a tumour, but not their common ancestor despite adequate sequencing depth. This phenomenon may reflect the effects of copy number change, of sub-clonal architecture that crosses biopsy boundaries, or sequencing errors.

Table S4. IntOGen colorectal cancer driver mutation list

A summary of the genes used to identify CRC-specific driver mutations downloaded from the IntOGen database (https://www.intogen.org/search?project=COREAD_TCGA, Table S3a). The list also includes the top 15 genes of the cancer genome atlas publication(1) and the top 15 (as of December 2016) from the catalogue of somatic mutations (COSMIC) database(2) under the colon carcinoma section. Pan-cancer driver genes are available from https://www.intogen.org/downloads.
Table S3b lists the top 5% of significantly mutated genes in CRC (non-MSI+ cases) in the TCGA publication, excluding *TTN*.

Table S5. Driver enrichment analysis

A statistical comparison of the frequency of certain driver mutations across CRAs and CRCs. Fisher test used to generate p-values in column 4. Only mutations with valid comparisons (non zero values in either CRA or CRC) were included in the table.

Table S6. Phylogenetic branch statistics

A table of phylogenetic trunk and mean branch lengths as defined by the number of SNAs present. The ratio of branch to trunk lengths and the standard deviation of branch lengths themselves are given. Note it is only possible to accurately estimate the trunk length when more than 2 biopsies are available.

Table S7. Immunohistochemistry

A summary of the counts for Ki67 and β-catenin positive and negative cells in a given CRC histological section (when available). Regions denote sample location (number relates to clock-face sample location). Note the analysis parameters in the last column of the table and the four samples marked in grey that were unable to be analysed due to over-staining of the slide.

Supplementary Figures

Figure S1. Diversity analysis across carcinomas and adenomas

Boxplots show the median and inter quantile range (IQR), upper whisker is 3rd quantile + 1.5*IQR and lower whisker is 1st quantile - 1.5*IQR. **a**: SNA divergence was statistically higher in adenomas compared to carcinomas (p < 2e-16). Y-axis shows the average number of divergent SNAs per megabase between two biopsies (calculated using a sub-sampling process, see Online Methods). Neoplasms with only two sequenced biopsies were excluded from this analysis. **b**: Number of divergent exonic SNAs as function of number biopsies. There was no correlation between genetic divergence and biopsy number (p = 0.5). **c**: Robustness of genetic divergence measurement to biopsy number was also assessed by repeated subsampling of biopsies (with replacement) in three CRCs with a high number of initial biopsies. When only two biopsies were sampled, the error rate of the estimated number of clonal SNAs (blue barcharts) was greater than 5% (compared to the number estimated from eight or more). Four or more biopsies per neoplasm seemed to provide a sufficiently accurate measurement.

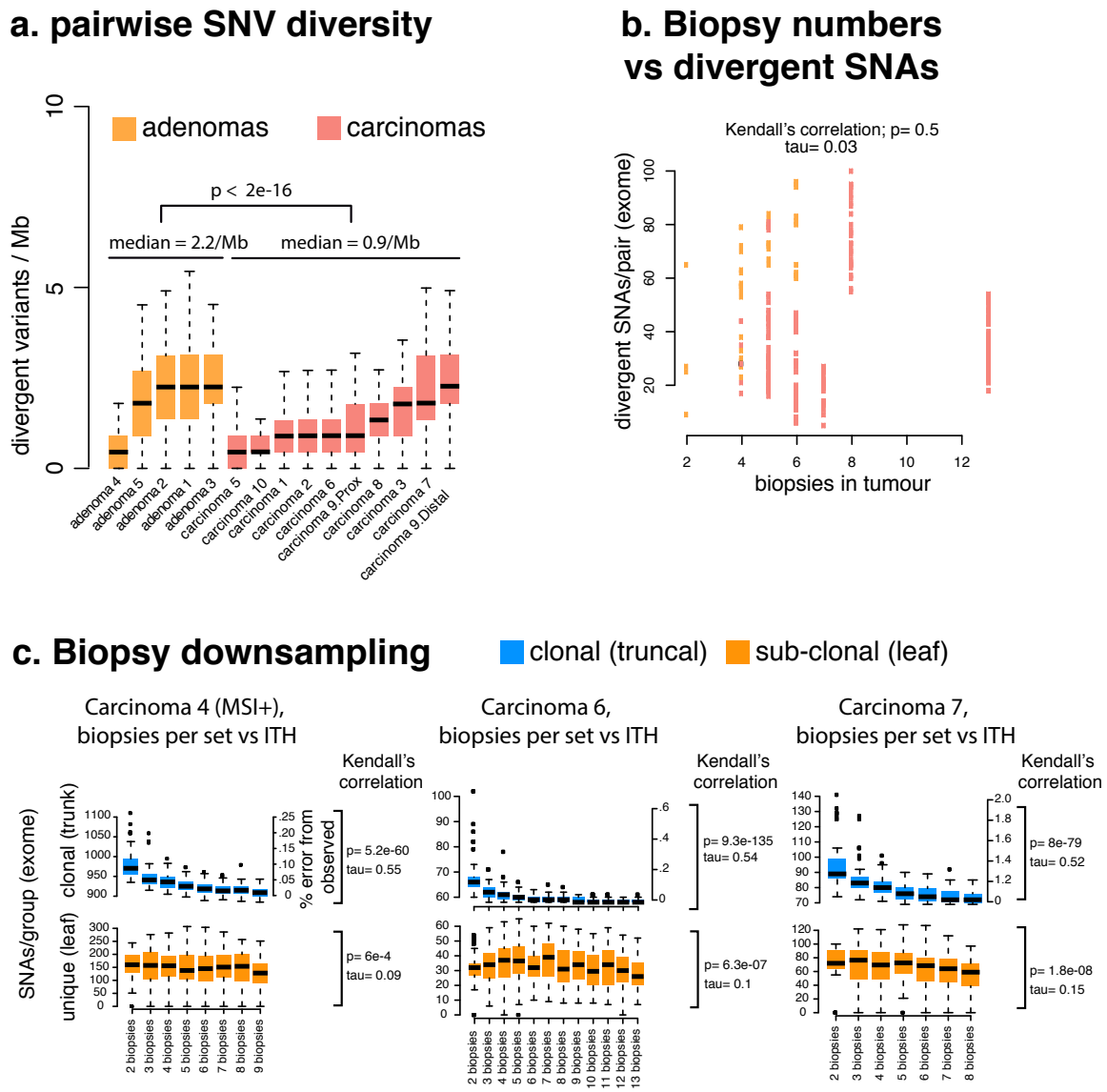


Figure S2. Non-synonymous:synonymous base substitution ratios in adenomas, sporadic carcinomas and MSI+ carcinomas

Bars show the ratio of the non-synonymous and synonymous SNAs across trunks ($T_{s/ns}$) and branches ($B_{s/ns}$) categories. A higher ratio is indicative of increased number of non-synonymous mutations and therefore potentially an increased number of selectable events. **C.** The ratio of $T_{s/ns}$ and $B_{s/ns}$ will indicate whether enrichment of selectable events has occurred earlier in evolutionary time. On average the adenomas do not have a significantly different $T_{s/ns}$ and $B_{s/ns}$ values (Wilcoxon signed-rank test; p=0.9, plotted as a ratio). Likewise, the Lynch carcinomas seem to have little change across trunks and branches (p=0.06). In carcinomas however the ratio is significantly different (p=0.03) indicating that there is an enrichment of non-synonymous SNAs on the phylogenetic trunks of these tumours. In turn this suggests that the positive selection of these mutations may be reduced as carcinomas develop. We note that although the trend is significant, there are individual adenomas and carcinomas that actually show a reduction in $T_{s/ns}$ compared to $B_{s/ns}$.

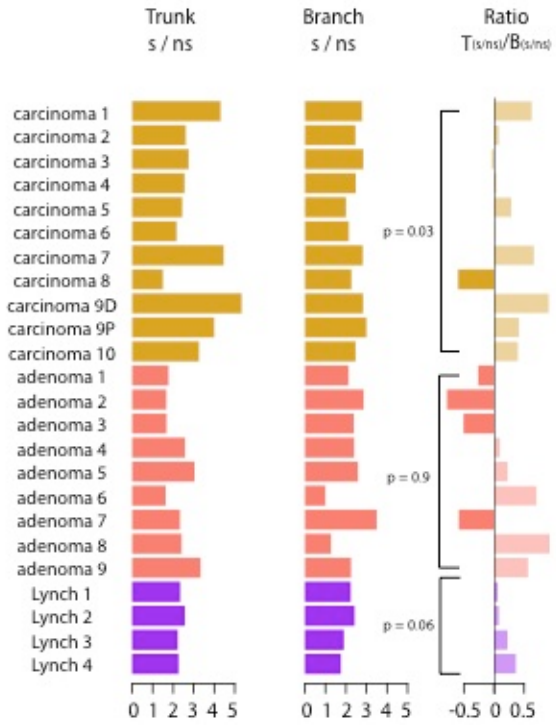
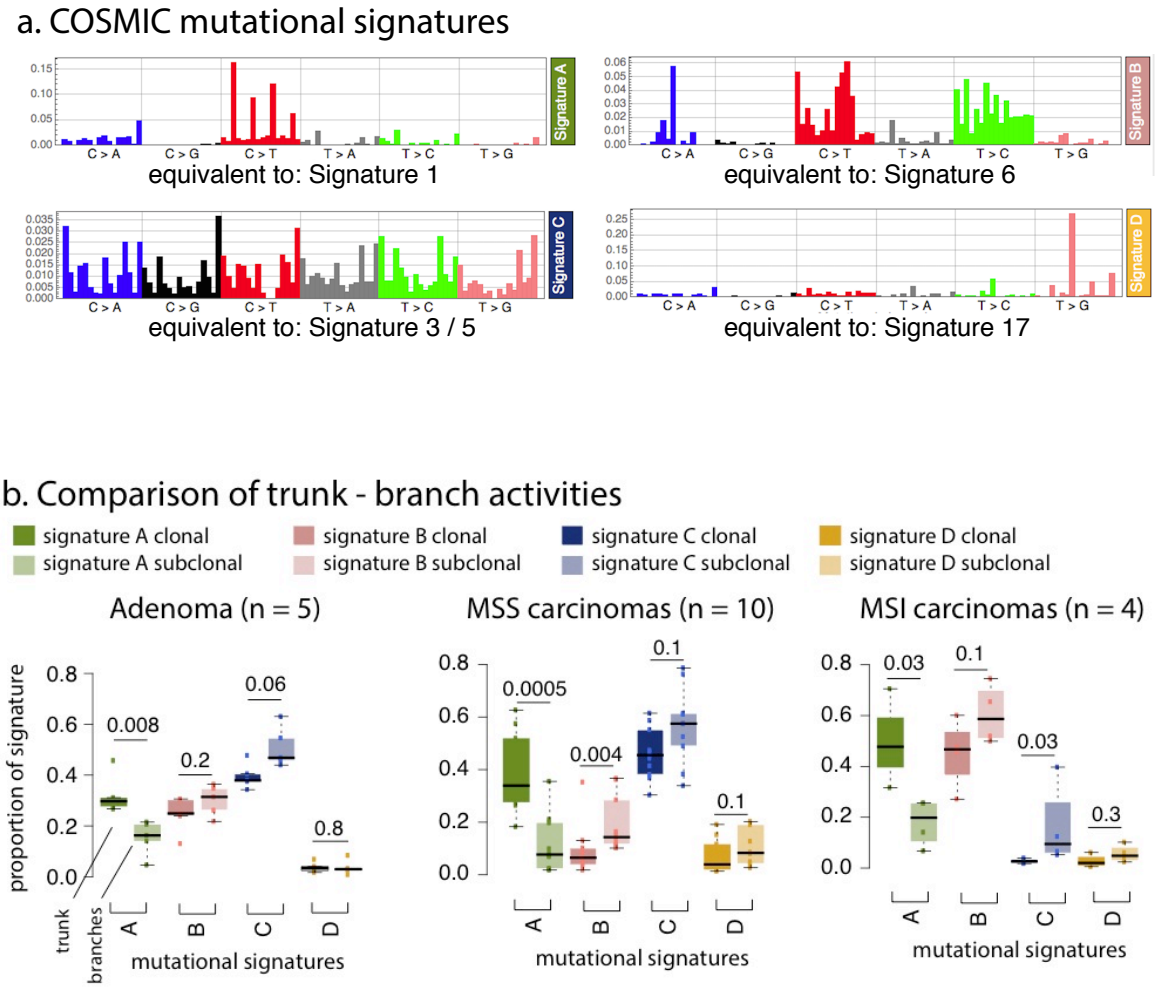
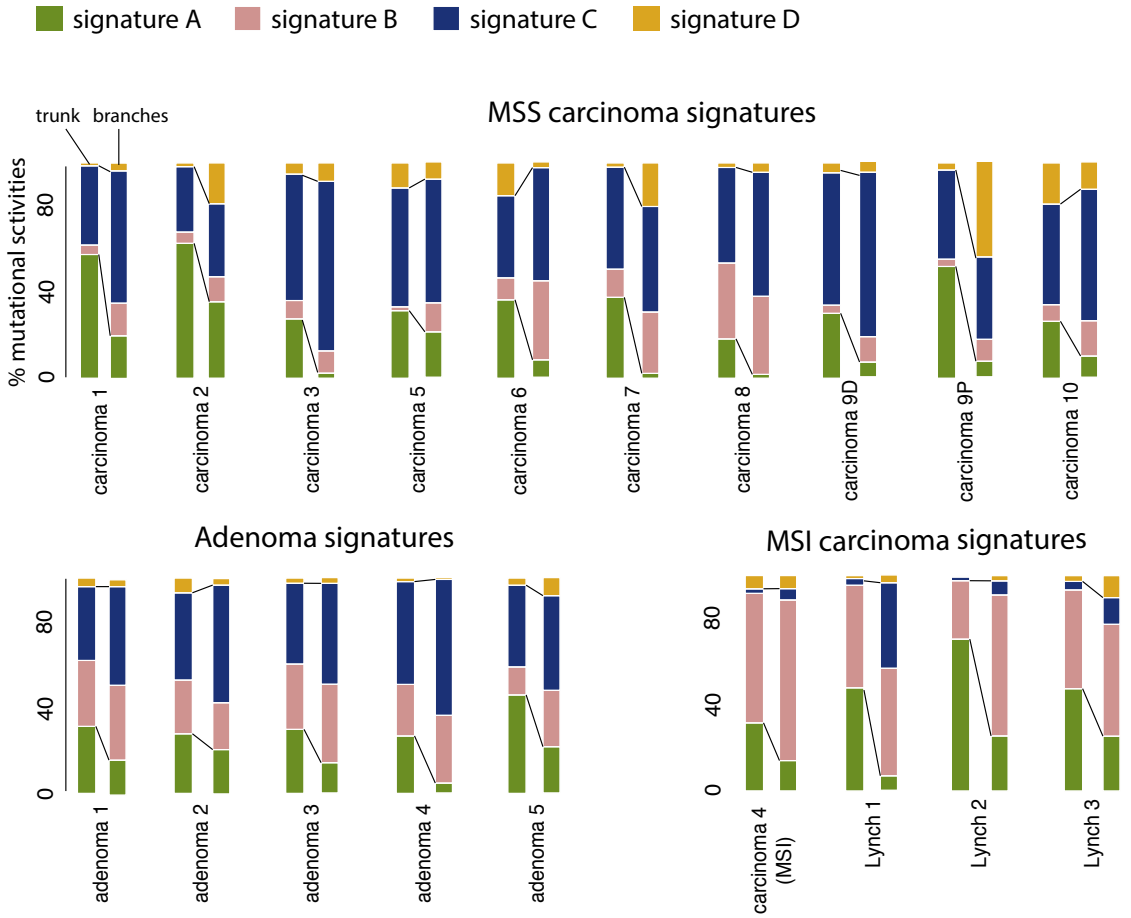


Figure S3. Analysis of mutational signatures

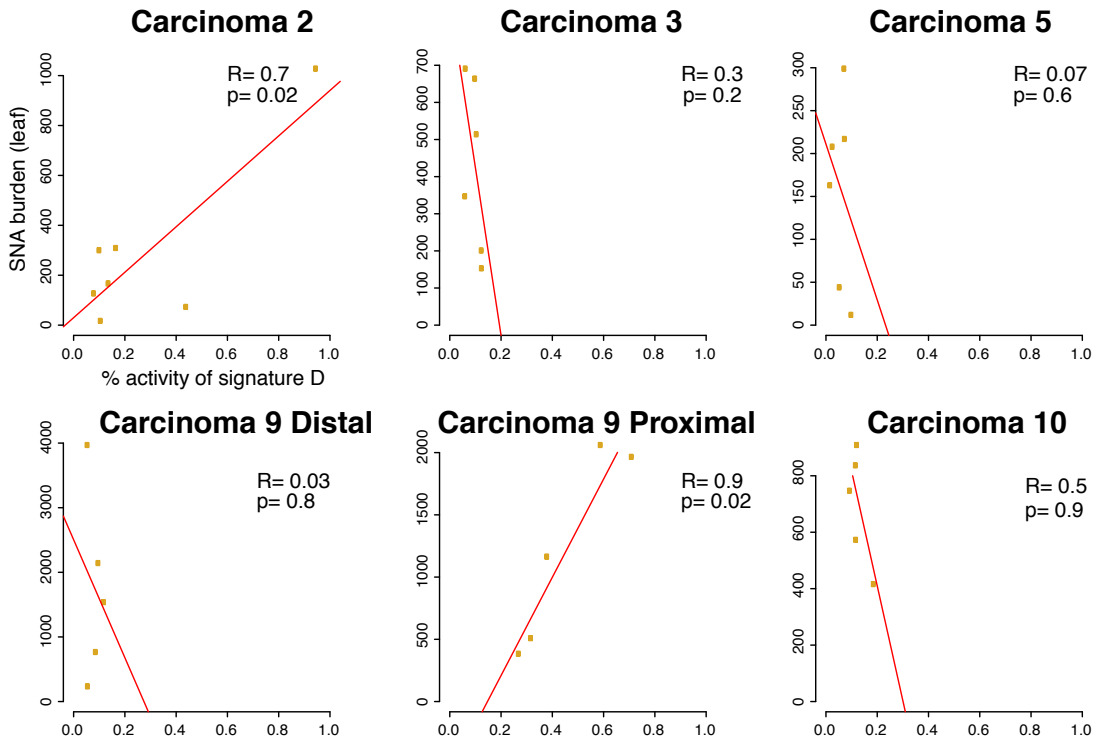
A summary of the *de novo* mutational signature analysis, based on 96-channel mutation spectra in Emu, is shown. Boxplots show the median and inter quantile range (IQR), upper whisker is 3rd quantile + 1.5*IQR and lower whisker is 1st quantile - 1.5*IQR. **a**. Signature A resembles COSMIC signature 1 (C>T changes at methylated CpGs caused by deamination of methyl-C and related to ageing). Our signature B is similar to MSI-associated COSMIC signature 6) and signature C resembles COSMIC Signatures 3 (double strand break repair) and 5 (tumour-specific molecular clock) respectively. Signature D resembles COSMIC signature 17 (unknown aetiology). **b.** In both CRAs and CRCs, our signature A was universally present and, as expected, comprised a significantly larger proportion of mutations on the trunk than the branches and leaves. The relative activity of this signature was not significantly different between the trunks of CRCs and CRAs (mean CRA: 32% versus CRC: 37%; p=0.4), perhaps owing to the bulk of mutations occurring prior to tumorigenesis. CRC Lynch 4 is not shown as it has only 2 biopsies. **c.** The relative activity of each signature on the trunk and branches/leaves of each tumour type is shown via barcharts. **d**. Increased signature D activity on branches/leaves relative to the trunk might occur secondary to a decrease in the importance of another signature (typically ageing-related Signature A), or owing to a true increase in the activity of that signature. The latter is expected to lead to an increased burden of signature-specific mutations, whereas that is not necessarily the case in the former. We found a significant, positive correlation between regional signature D activity and total SNA burden in carcinomas 2 and 9P, indicating an underlying hypermutational process in these lesions. **e.** The burden of non-signature D-specific mutations correlated against signature D activity, where no relationship was found with between non-D mutations and D activity for any cancer, suggesting that signature D activity has a subtle effect, if any, on tumour evolution.



c. Composition of signatures across mutations



d. Correlation of signature D activity and total SNA burden



e. Correlation of signature D activity and non-signature D SNA burden

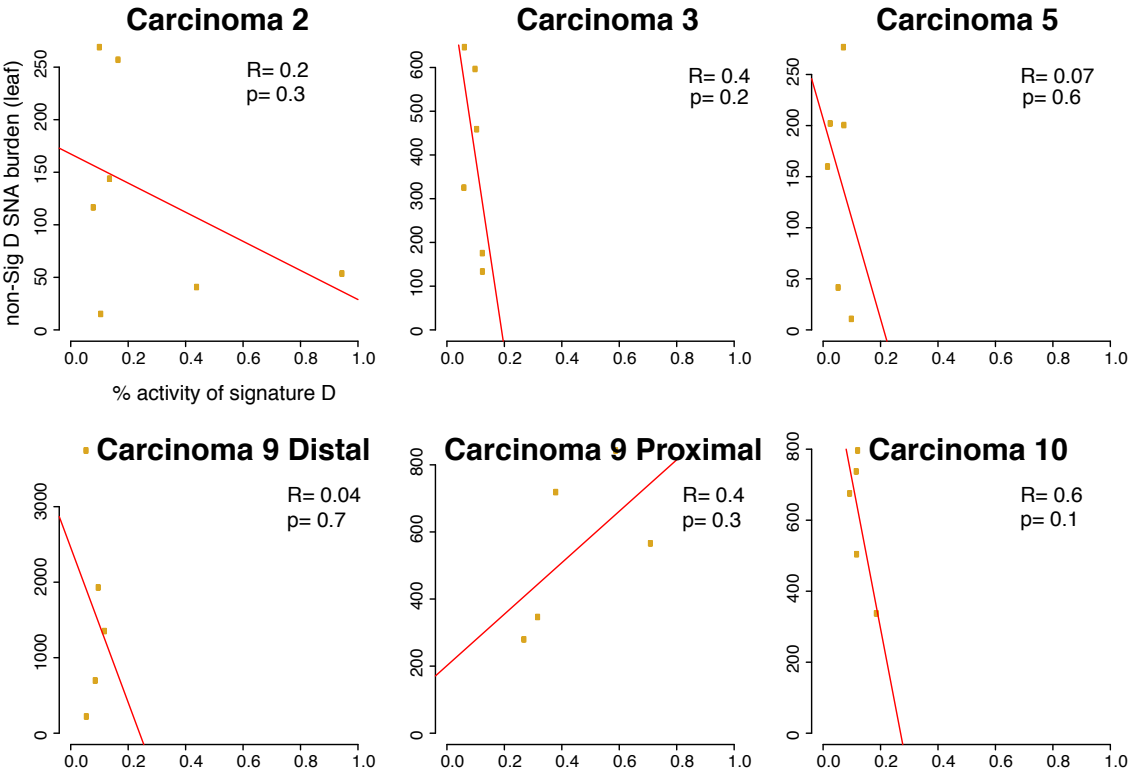


Figure S4. Mutational signatures of synchronous carcinomas

The synchronous lesions (carcinoma 9 distal and proximal) had presumably developed in a similar environment for most of their existence, given that they were located only 10cm apart in the bowel. There was a significant enrichment of signature D on the branches of the proximal lesion, but the truncal mutations of the distal (9D) and proximal (9P) cancers resulted from very different signature A and C activities. The truncal mutational signatures activities in 9D matched those in the biopsies of normal-appearing mucosa from between the two cancers, whereas 9P seemed to have developed a different composite of signatures. Despite these differences, the total mutational burdens of the two CRCs were remarkably similar (total SNAs, distal: 31,633, proximal: 29,257, see Table S8). Driver mutations, including distinct *APC* mutations, accrued independently in the trunk of each CRC (Figure 2). A total of 261 SNAs, none of which appeared pathogenic, was shared between intervening, morphologically normal epithelium and both CRCs (see Table S8), suggesting that field cancerisation was an unlikely scenario and these cancers evolved independently along different trajectories. This analysis suggests that non-aging-related mutational signatures can arise early in CRC development and, whilst a transient major mutagenic event cannot be excluded, probably result from non-macroenvironmental influences.

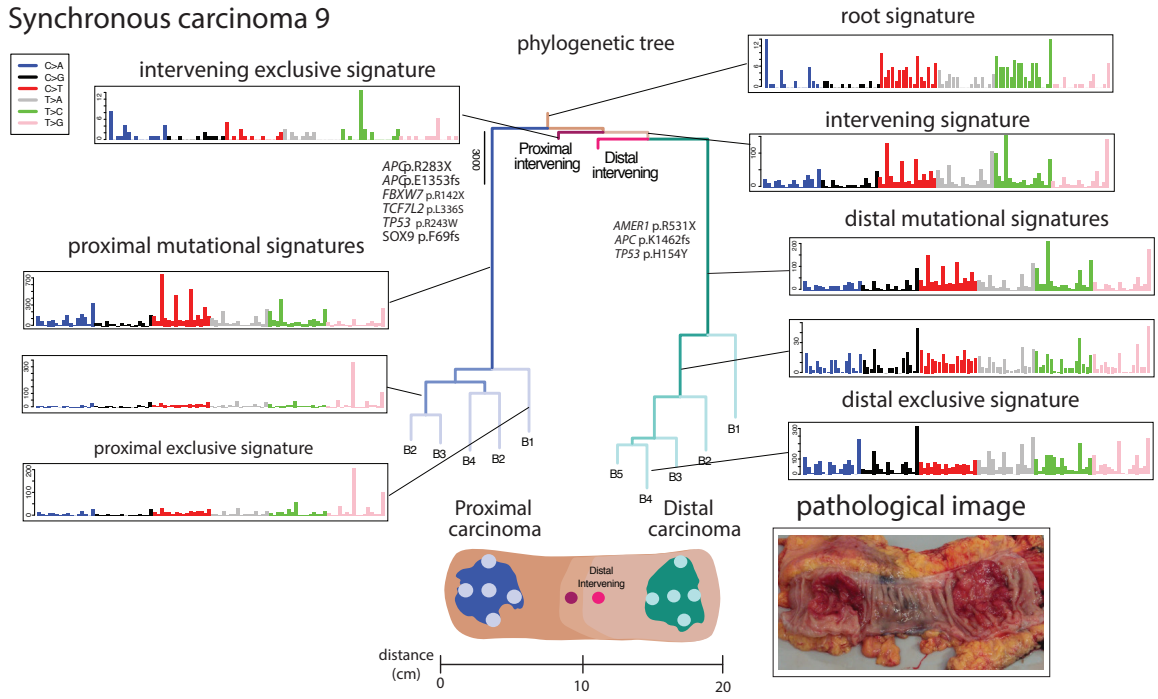


Figure S5. CRC immunohistochemistry

a. Full tissue sections were stained for Ki-67 and β-catenin (i & ii) by immunohistochemistry. The HALO ™ classifier machine learning algorithm was trained to categorise tumour tissue (red), stroma (green) and background regions (yellow) for each stain (iii, iv). Goodness of classification was visually confirmed on all samples. Cell counting was conducted in the cancer cell compartment using the HALO ™ cytonuclear algorithm (see Online Methods). Positive cancer cells are shown in red, negative cells are shown in blue (v, vi). The total counts and percentage of marker-positive cells in the tumour cell compartment were recorded. **b.** Correlations of exonic mutational burden and ploidy with Ki-67 and Beta-catenin expression for each tumour region are shown in the left plots (p-values from multivariate linear regression by tumour region, correcting for inter-tumour differences; none significant). Boxplots right illustrate the differences in expression in carcinomas that are suspected to have been genome doubled. Here, Ki-67 expression is significantly higher in genome doubled carcinomas.

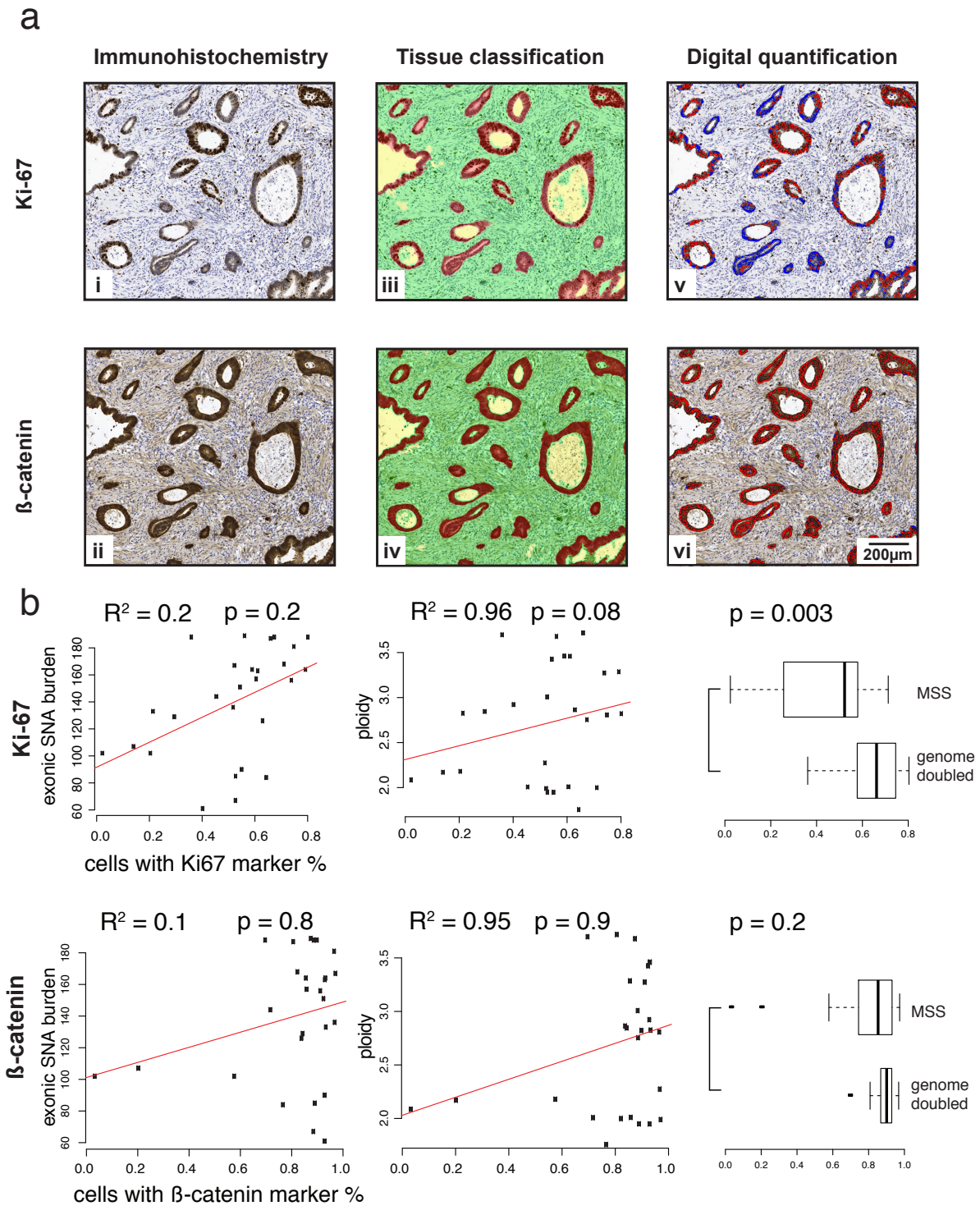

a
Immunohistochemistry
Tissue classification
Digital quantification
Ki-67
ß-catenin
i
ii
iii
iv
v
vi
200µm
b
R² = 0.2 p = 0.2
R² = 0.96 p = 0.08
p = 0.003
Ki-67
exonic SNA burden
ploidy
MSS
genome doubled
cells with Ki67 marker %
R² = 0.1 p = 0.8
R² = 0.95 p = 0.9
p = 0.2
ß-catenin
cells with ß-catenin marker %

Figure S6. Evidence of mixing of sub-clones within biopsy regions

WGS CRCs (carcinomas 2, 3, 4, 5, 9D, 9P and 10) were analysed using the Battenberg algorithm (see Online Methods). The below pairwise regional plots show shared variant allele frequencies corrected for copy number. Red shading shows cluster nodes of SNAs. Nodes on y=0 or x=0 are consistent with sub-clonal SNAs (branches) between the biopsy pairs. Clonal SNAs (between pairs of biopsies) form a node that is at or close to (1,1). Nodes on y=x (but not near origin) may be chance deviations from clonality and/or low coverage sites. Nodes elsewhere in the space may represent mixing of clones across tumours or 'clonal patches' that are partially sampled across biopsy boundaries. There was some evidence for sub-clonality within certain biopsies in a few CRCs. Biopsy 4 (leading edge, shown in comparison to biopsies 2,3 & 4) of Carcinoma 3 had some sub-clonal SNAs that were clonal in all other biopsies from that cancer; biopsy 6 (central tumour) of Carcinoma 5 also possessed sub-clonal variants that were clonal in two leading edge samples (biopsies 3 and 4); and biopsy 1 of Carcinoma 10 seemed to possess many sub-clonal variants when compared to all four other biopsies from that tumour (illustrated below). In this cancer in particular some SNAs appeared to represent clonal mixing (where the same SNA is found in disparate regions across a tumour, see Figure 4).

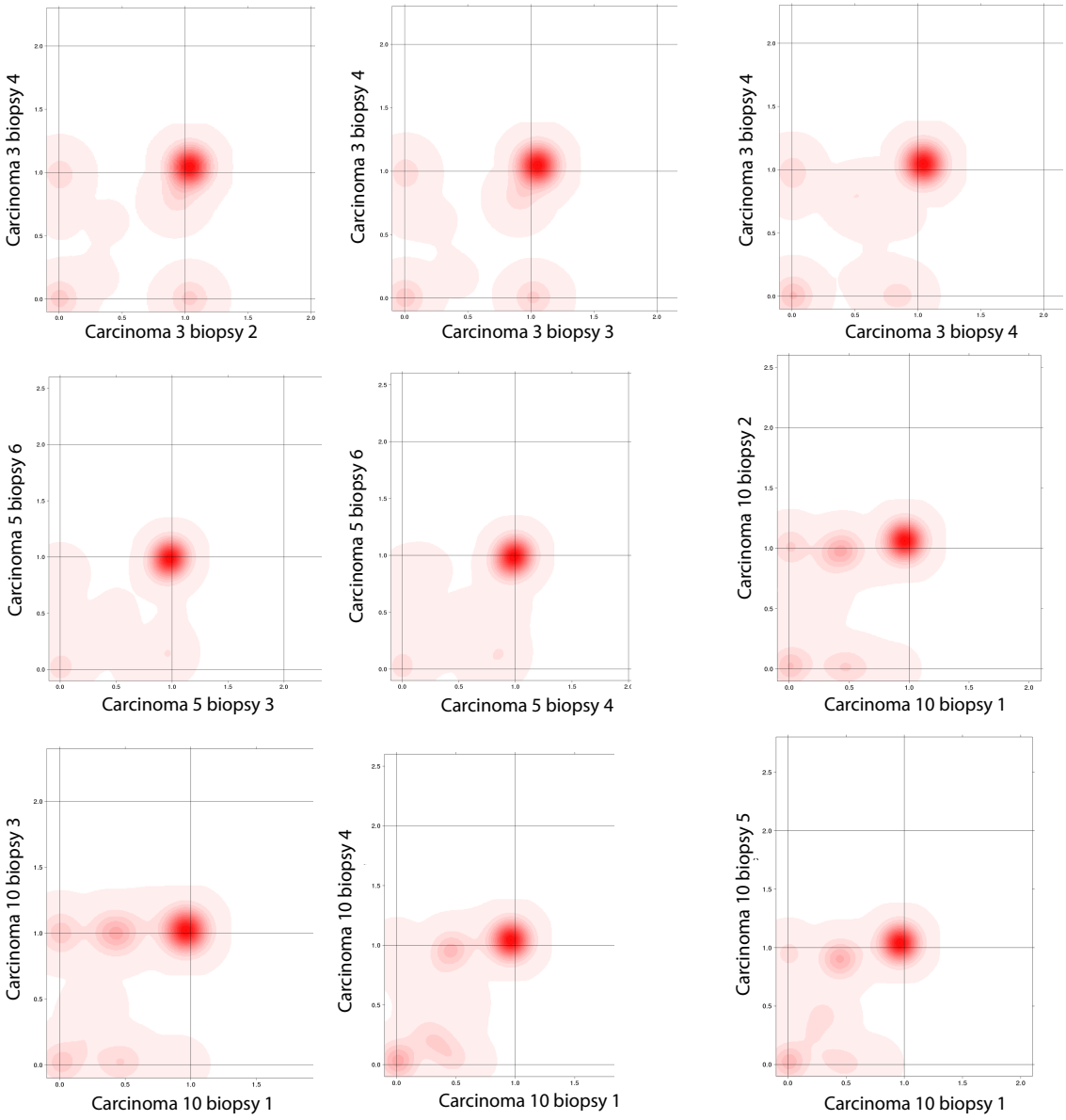


Figure S7. Histogram of CNA frequencies across MSS CRCs and CRAs

The histograms of all CNAs across adenomas and carcinomas was produced by binning all segmentation profiles into 10kb windows across the genome, noting the copy state and regional distribution of the events. The resulting frequency was normalised for the number of biopsies analysed. It is possible to observe common chromosomal aberrations as previously noted by the TCGA (1), which includes 1p, 5q, 8p, 14, 15, 18 and 22 loss and gain of several chromosomes including 2, 7, 12, 13 and 20. In carcinomas, deletions/cnLOH are more likely to be ubiquitous than gains, whereas in adenomas, most events tend to be sub-clonal/regional.

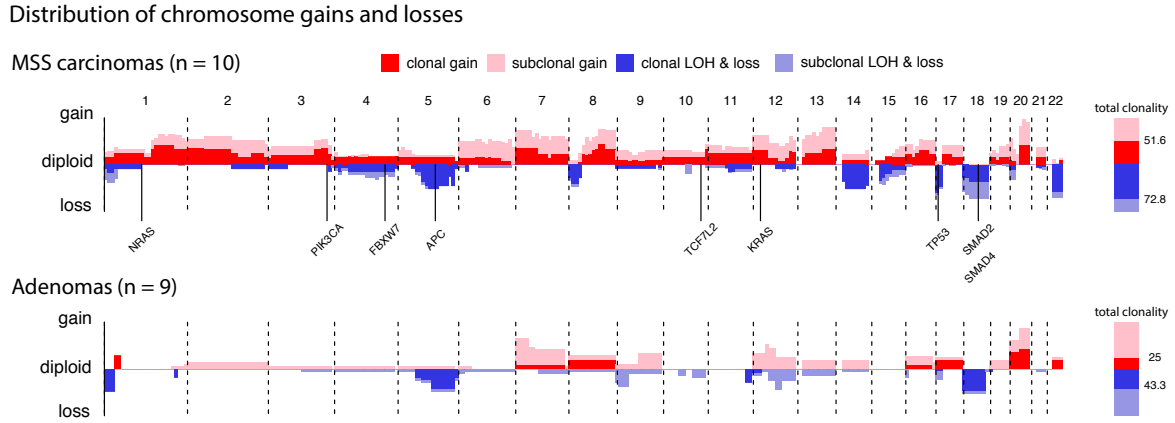


Figure S8. Genome doubling assessment

We reasoned that the most specific indicators of genome doubling would be the presence of chromosome centromeres at copy number 4 or above, especially if derived equally from both parental homologues. Furthermore, doubling should be assessed independent of chromosome size. We therefore derived a heuristic, pre-specified genome doubling score (GDS) from

GDS =
(2 x no. chromosomes at CN4 and allelic balance) +
(2 x no. chromosomes at CN5) +
(no. chromosomes at CN4 and allelic balance)

For improved sensitivity, we included copy number changes present in any tumour region, and also separately computed regional GDSs, showing evidence of heterogeneity in adenoma 2 (three doubled and one non-doubled region). The score distribution from all CRAs and MSS CRCs is shown below, with tumours assigned as doubled if they have score of ≥10. Although this threshold is subject to a degree of uncertainty, there is general agreement with ploidy scores (Figure 3B), as would be expected ($p<0.001$, linear regression).

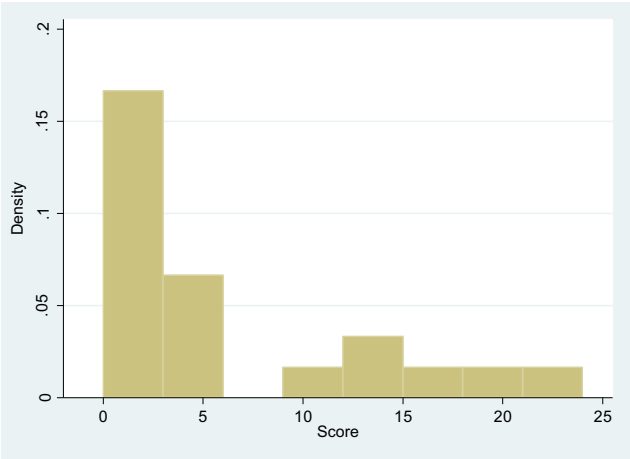
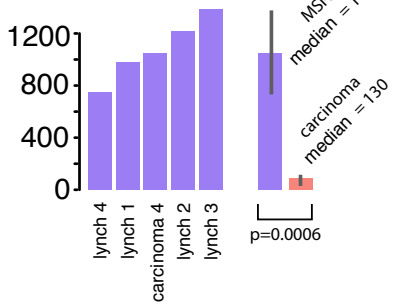


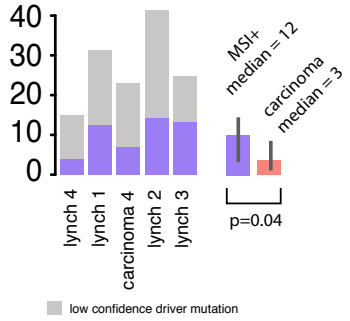
Figure S9. Evolutionary features of MSI+ lesions

a. Exonic SNA mutation burdens are significantly increased in MSI+ CRCs compared with MSS CRCs, as expected, but the sporadic MSI CRC is not an outlier. Bars show ranges. **b**. Driver mutations are shown as for Figure 1. **c**. MSI+ CRCs contain few CNAs and correspondingly have near-diploid genomes. **d**. The proportions of ubiquitous and sub-clonal mutations are similar between MSI+ and MSS CRCs. Bars show ranges. Boxplots show the median and inter quantile range (IQR), upper whisker is 3rd quantile + 1.5*IQR and lower whisker is 1st quantile - 1.5*IQR. **e**. Phylogenetic trees for MSI+ CRCs have a similar topology to MSS CRCs, with most tier 1 driver mutations truncal.

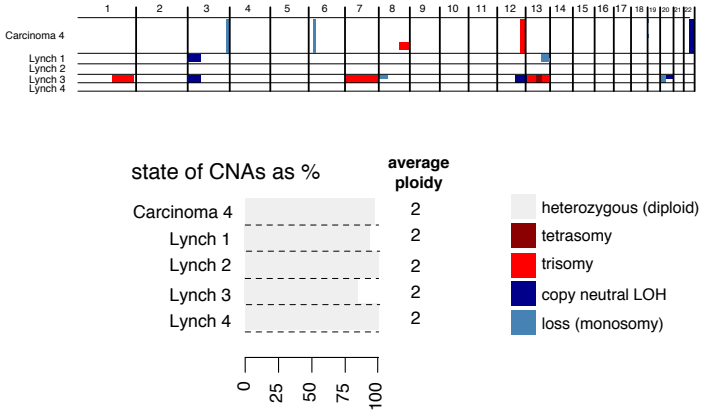
a. Exonic SNV burdens



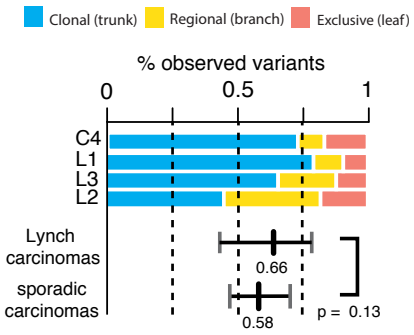
b. Driver mutation burdens



c. MSI+ chromosome number aberrations



d. Clonal proportions

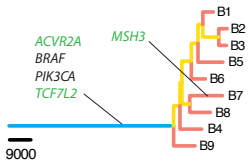


e. Phylogenies MSI+ carcinomas

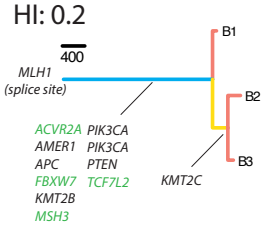
* = bootstrap value of >= 65, < 95
\+ = homoplasic variant

SNV driver
indel driver

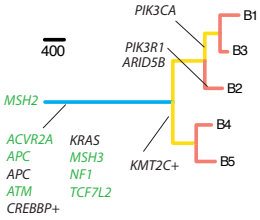
Carcinoma 4
HI: 0.46



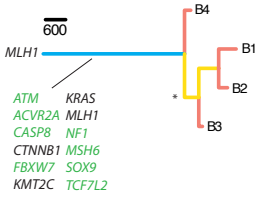
Lynch carcinoma 1
HI: 0.2



Lynch carcinoma 2
HI: 0.32



Lynch carcinoma 3
HI: 0.34



Lynch carcinoma 4
HI: -

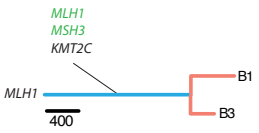
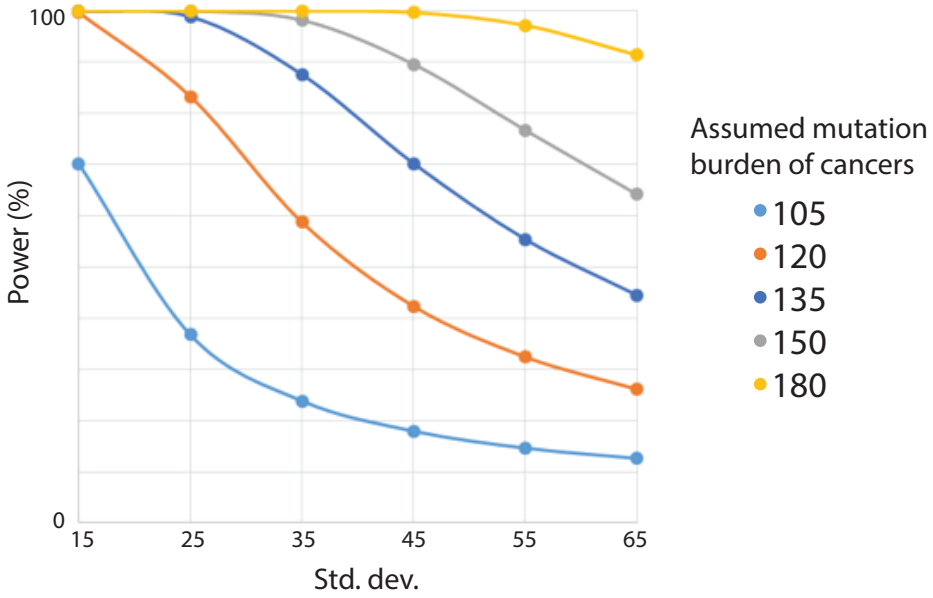


Figure S10. Power to detect difference in mutation burden between adenomas and carcinomas.

Power (y-axis) to detect a difference between a mean of 90 mutations in adenomas and specified higher number of mutations in carcinomas (coloured lines), as a function of the standard deviation of mutation burden across tumours (x-axis). Measured standard deviation was about 35 mutations (see Methods).



Supplementary Note

We used mathematical modeling to estimate the relative timings of SCNA events (CNV's) in each tumour along the cell lineage between divergence of tumour samples from the normal and the last common ancestor of tumour samples. In addition to those assumptions mentioned explicitly below the modeling assumes that the average SNA mutation rate was equal in all of the chromosomal regions considered.

1 Growth model

Suppose that during the growth of cell lineage L an ancestral cell, c, gained an extra copy of region r of the A-allele of chromosome j. We note that any SNA'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, SNA's that occurred in an ancestor of the cell in region r on the B-allele, and SNA'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.

Formalising this reasoning, we model the number of high frequency SNA's in region r, α, and the number of low frequency SNA'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 SNA 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: $A = 2, B = 0$
Case II: $A = B = 2$
Case III: $A = 2, B = 1$
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. The genome was segmented into R regions so that all samples had uniform copy number states within each region.

We identified clonal CNV events by selecting the genomic regions which shared the same non-diploid copy number samples across all tumour samples. For CNV events of length greater than 10Mb which conformed to one of the

three cases above, we clustered SNA's detected in the region in each sample into a high frequency set and a low frequency set, using the R package 'mixtools'. For these purposes we inferred the centres of the allele frequency distributions for each sample based on the cellularity and copy number state. 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 samples were excluded.

We also considered the possibility that there are regions of the genome inferred to be diploid in the last common ancestor of tumour samples of length l_i, containing γ_i SNA's. However, for the present purposes both these values were considered to be zero.

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:

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

and Kuhn-Tucker conditions:

$$
\begin{aligned}
&\text{For all i} \in \text{I} \quad \frac{\partial L}{\partial t_i} = 0, \quad \frac{\partial L}{\partial T} = 0 \\
&\text{For all i} \in \text{I} \quad t_i \geq 0 \quad T - t_i \geq 0 \\
&\text{For all i} \in \text{I} \quad \lambda_i t_i = 0 \quad \lambda_{N+i}(T - t_i) = 0 \\
&\text{For all i} \in \text{I} \quad \lambda_i \geq 0, \quad \lambda_{N+i} \geq 0
\end{aligned}
$$

We will now find expressions for $(\mathbf{t}, T)$ that must hold at a global maximum, and show that there is at most one solution.

For all $i \in I$, since $\alpha_i > 0$, $t_i = 0 \implies \mathscr{L} = 0$. There are at least some admissible points where $\mathscr{L} > 0$, so at a global maximum, $0 < t_i$ and therefore $\lambda_i = 0$.

Similarly, for all $i \in C_1 \cup C_2$, since $\beta_i > 0$, $t_i = T \implies \mathscr{L} = 0$. 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 $i \in C_1$:

$$
\begin{aligned}
&\frac{\partial L}{\partial t_i} = 0 \\
\implies &\frac{\partial \mathscr{L}}{\partial t_i} = 0 \\
\implies &\frac{\partial log(\mathscr{L})}{\partial t_i} = 0
\end{aligned}
$$

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

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

Since $-4(1-A_i)l_iT\alpha_i > 0$, $2(1-A_i)l_i < 0$ 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 $i \in C_2$:,

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

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 $i \in C_3$:

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

Since $\mathscr{L} > 0$

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

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

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

and similar to Cases 1 and 2

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

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

$$t_i = T \quad (7)$$

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

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 $\frac{\partial \mathscr{L}}{\partial T}$ 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

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

So we have

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

Substituting from 7 and 8

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

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

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

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

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 $\frac{\partial(T-t_i)}{\partial T} < 0$

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

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.