

# 1 **Resolving genetic heterogeneity in cancer**

2 Samra Turajlic<sup>1,2</sup>, Andrea Sottoriva<sup>3</sup>, Trevor Graham<sup>4\*</sup> and Charles Swanton<sup>1,5,6\*</sup>

3

4 <sup>1</sup> Cancer Evolution and Genome Instability Laboratory, Francis Crick Institute, London, UK.

5 <sup>2</sup> Skin and Renal Units, The Royal Marsden NHS Foundation Trust, London UK.

6 <sup>3</sup> Evolutionary Genomics and Modelling Laboratory, Centre for Evolution and Cancer,  
7 Institute of Cancer Research, London, UK.

8 <sup>4</sup> Tumour Biology, Evolution and Cancer Laboratory, Barts Cancer Institute, Queen Mary  
9 University of London, London, UK.

10 <sup>5</sup> Cancer Research UK Lung Cancer Centre of Excellence London, University College London  
11 Cancer Institute, London WC1E 6DD, UK;

12 <sup>6</sup> Department of Medical Oncology, University College London Hospitals, London NW1 2BU,  
13 UK.

14

15 These authors contributed equally: Samra Turajlic, Andrea Sottoriva

16 \*email: [t.graham@qmul.ac.uk](mailto:t.graham@qmul.ac.uk); [charles.swanton@crick.ac.uk](mailto:charles.swanton@crick.ac.uk)

17

## 18 **Abstract**

19 To a large extent cancer conforms to evolutionary rules defined by the rates at which clones  
20 mutate, adapt and grow. Compared to species evolution however, cancer is a particular  
21 case, due to the vast population size, chromosomal instability, and the potential for  
22 phenotypic plasticity. Nevertheless, an evolutionary framework is a powerful aid in our  
23 understanding of cancer progression and therapy failure, and could be applied to predict  
24 individual tumour behaviour and aid treatment strategies.

25

## 26 **Introduction**

27 Tumours are composed of subpopulations of cells (subclones) that may be distinguished by  
28 a variety of features impacting their phenotype, including genetic alterations. Genetic  
29 intratumour heterogeneity (ITH) has been documented across most cancers (reviewed in<sup>1</sup>)  
30 and acts as a substrate for clonal evolution. The fundamental biological mechanisms  
31 underlying **clonal evolution [G]** in cancer are similar to those that underpin the evolution of  
32 asexually-reproducing species: replication, heritable variation, **genetic drift [G]**, **selection [G]**  
33 and environmental changes. Central to the neo-Darwinian synthesis of evolutionary biology  
34 is the paradigm of **molecular evolution [G]**, which links Mendelian genetics to Darwinian  
35 adaptation. Molecular evolution is relevant to cancer because the use of genomic  
36 sequencing is a key technology to understand temporal and spatial patterns of **somatic**  
37 **evolution [G]**. At the core of molecular evolution, in turn, is theoretical population genetics,  
38 which has been the fundamental mathematical formalism to describe evolution for the past  
39 90 years<sup>2,3</sup> The same theoretical framework has been used to understand clonal evolution  
40 in cancer<sup>4 5 6 7 8 9 10 11</sup>. The study of the evolutionary dynamics of cancer clones is  
41 fundamentally concerned with the relative frequencies of cancer subpopulations over space

42 and time. Although some peculiarities of cancer evolution distinguish it from classic species  
43 evolution (Box 1), classical evolutionary theory can nevertheless be readily applied to  
44 understand cancer development.

45 Over the past 5 years, a number of next generation sequencing studies have  
46 captured cancer evolution in space and time, illuminating the variety of evolutionary  
47 patterns that shape cancer, and showing their clinical relevance. Here, we provide an  
48 overview of the theoretical models of tumour evolution and the caveats around correctly  
49 interpreting genomic data and inferring evolutionary dynamics. We discuss the relevance of  
50 **chromosome instability [G]** (CIN) as a driver of cancer evolution and, in particular,  
51 metastases; the clinical value of evolutionary classification of cancer; and finally, the role of  
52 clonal evolution in treatment failure.

53

#### 54 **Current models of tumour evolution**

55

56 Cancer as a system is characterized by an astonishing complexity and emergent behaviour.  
57 Nevertheless, this complexity arises from the relatively simple, underlying evolutionary rules  
58 of mutation, genetic drift and selection, involving a large number of interacting agents (for  
59 example, the millions of cancer cells within a single lesion and the surrounding tumour  
60 microenvironment). The emergent behaviour gives rise to different observed ‘modes’ of  
61 evolution (**Figure 1**), which result from different combinations of the aforementioned  
62 fundamental rules in distinct contexts. In other words, since selective pressures change over  
63 time, so can the ‘modes’ of evolution. Here we discuss the principles of selection and  
64 different modes of evolution.

65

#### 66 *Selection*

67

68 Selection, whereby one lineage is ‘favoured’ over another, and produces more surviving  
69 offspring, is arguably the most interesting force in evolution, as it leads to adaptation. In  
70 general, positive selection drives tumour progression, but negative selection (e.g. against  
71 potent neoantigens<sup>12,13</sup>) also contributes to tumour evolution. However, selection is not be  
72 operative at all times. Whereas mutation and drift are continuously occurring, and their rate  
73 depends on cell division and population dynamics, selection is dependent on the  
74 environmental context. For example, if there is no differential survival within a population,  
75 the lack of positive selection would mean that the population evolves neutrally (only  
76 mutation and drift are at play). Consequently, branching (see below) of a tumour  
77 **phylogenetic tree [G]** does not always imply clonal selection, as branching is the natural  
78 product of mutational processes<sup>14 15</sup>. Selection has the effect of ‘pruning’ the tumour tree,  
79 for example, favouring the expansion of some lineages (branches) over others. The  
80 mutation rate itself could also be subject to selection. A higher mutation rate allows for  
81 diversification but also carries the ‘risk’ of increasing the rate at which deleterious  
82 mutations, which perturb cancer growth, are acquired<sup>16,17</sup>. For example, excessive

83 chromosomal instability (CIN) can result in cell autonomous lethality, however a “just-right”  
84 threshold of CIN may be evolutionary advantageous. Mutations in the APC/C subunits may  
85 be selected during the evolution of chromosomally unstable tumour cell populations  
86 resulting in lengthening of mitosis, suppression of chromosome missegregation and  
87 attenuation of excessive CIN<sup>18</sup>.

88

89 Mathematical models suggests that, in a growing population, mutator phenotypes are  
90 selected, because the cells that stochastically acquire positively selected mutations in effect  
91 ‘doubly benefit’ from their own increased fitness and the negative fitness effect of  
92 deleterious mutations on the rest of the population<sup>19</sup>. Relatedly, modelling also suggests  
93 that a mutator phenotype increases the ‘efficiency’ of carcinogenesis by making it more  
94 likely that a necessary set of mutations are acquired for transformation and cancer  
95 progression<sup>17</sup>.

96

#### 97 *Branching evolution*

98 Evolution is always branched, because cell division and mutation continuously produce  
99 divergence at the level of genotypes. This fact is particularly true for cancer genomes, as  
100 cancers often have a mutator phenotype [G]<sup>20</sup>. Hence, in principle, at any given time point a  
101 tumour cell population consists of different cell lineages. Random fluctuations in the birth  
102 and death rates of these distinct lineages can lead to genetic drift, whereby one lineage  
103 produces more surviving offspring than another lineage, and expands by chance. Genetic  
104 drift is referred to as a form of neutral evolution [G], as all lineages are neutral other with  
105 respect to their chance of producing surviving offspring<sup>21,22</sup>. Similar patterns of branching  
106 are also apparent in healthy tissue<sup>23,24</sup>, emphasising branching as a necessary by-product of  
107 proliferation tissues. However, when multiple cancer subclones have increased fitness they  
108 will expand (assuming no other limitations to growth) simultaneously due to selection, as  
109 evidenced by the finding of subclonal cancer driver mutations and their impact upon cancer  
110 progression<sup>25,26</sup>. Selection is also evident in the finding of parallel evolution within the same  
111 tumour where distinct lineages acquire mutations in the same cancer driver gene, leading to  
112 parallel subclonal expansions.

113

#### 114 *Linear evolution*

115 Linear evolution model posits that only one of the lineages survive over time. However, as  
116 with the fossil record, it does not imply that there was only ever a single lineage that  
117 evolved in a step-wise fashion. Any conclusions regarding linear evolution from cancer  
118 genomic data are likely false owing to limited sampling applied to the cancer in question  
119 and limits of resolution by next-generation sequencing technologies.

120

#### 121 *Neutral evolution*

122 Neutral evolution can be regarded as the evolution that occurs in between selection events.  
123 Prior to adaptive mutation occurring, the population evolves neutrally, and when the

124 mutation arises it initiates a rapid **clonal sweep [G]** which can be complete or incomplete. If  
125 the sweep is complete and all the cells in the population carry the adaptive mutation the  
126 dynamics revert to neutral again.

127

### 128 *Punctuated evolution*

129 Punctuated evolution is the opposite of gradual evolution and presents more rapid bursts of  
130 adaptive evolution. If the adaptive mutation is a large-scale alteration of the genome (for  
131 example, loss or gain, translocation or fusion of a chromosome) the adaptive clone has been  
132 referred to as a “hopeful monster”<sup>27,28</sup>. Compared to a small-scale mutation, its genome is  
133 significantly altered, with the ‘hopeful’ referring to the likelihood that the mutation is  
134 adaptive. **Punctuated equilibrium [G]** is a model first proposed by Eldridge and Gould in the  
135 early 1970s for species evolution<sup>29</sup> whereby adaptation occurs in a small spatially-isolated  
136 niche, until the newly-adapted individuals rapidly expand out of the niche and through the  
137 wider population. Because the niche is small, the gradually-adapting population is unlikely  
138 to be sampled before it expands, and so the evolutionary dynamics of the population at  
139 large are ‘punctuated’ by the expansion of the adapted clone. Equilibrium refers to long  
140 periods of clonal stasis during which the adapted clone persists without detectable change.

141

### 142 **Inferring evolutionary mode from genomic data**

143 Although adaptation occurs at the phenotypic level, measuring the tumour cell phenotype  
144 within its original environment is challenging. Surrogate measurements such as gene  
145 expression are informative, but given the complexity and plasticity of the cancer  
146 transcriptome, and the contribution to gene expression signals from cells within the tumour  
147 microenvironment, these are often difficult to interpret in light of evolution. This is why to  
148 date genome profiling has been the preferred tool to study cancer evolution. However,  
149 there are several major caveats when we try to understand the phenotypes from studying  
150 the genotypes, a problem that has been tackled over decades in the field of molecular  
151 evolution. The key issue is that the cancer **genotype–phenotype map [G]**, bar some notable  
152 exceptions, such as treatment resistance mutations, is largely unknown. Therefore,  
153 mapping the tumour phylogenetic tree and the underlying adaptive traits remains difficult.

154

### 155 *Bulk sequencing*

156 The commonly used bulk sequencing, that is profiling of a sample comprised of many cells,  
157 imposes a major limitation on inferences about tumour evolution dynamics. Because the  
158 standard depth of sequencing is many orders of magnitude smaller (100–1000X) than the  
159 number of cells in the sample (10 million – 1 billion), bulk sequencing only recovers  
160 mutations that are either present in all, or the majority of cancer cells in the given sample.  
161 Each doubling of the cancer cell population halves the frequency of new mutations arising in  
162 the population, and hence after just 7 doublings new mutations are undetectable with 100X  
163 sequencing, and after 10 doublings new mutations are undetectable at 1000X sequencing  
164 depth. Thus, detecting selection that resulted in a limited clonal expansion (100s-1000s

165 cells) is problematic. Contamination by stromal cells imposes an additional challenge as it  
166 dilutes out the frequency of variant alleles. Thus, bulk sequencing mostly informs on the  
167 most recent common ancestor (MRCA) of the cells in the sample, but the 'node' in the  
168 phylogenetic tree is extinct in the current malignancy. The more cells in the bulk sample, the  
169 older the MRCA and shorter the 'apparent' branches in the tree. Consequently, different-  
170 sized samples can generate very different portraits of the clonal structure of a tumour.

171

#### 172 *Choice of sequencing assay*

173 The relative abundance of **passenger mutations [G]** (evolutionary neutral; non-adaptive)  
174 over **driver mutations [G]** (positively selected) makes the passengers that hitchhike to a  
175 driver event, very informative vis a vis clonal dynamics. Passenger mutations provide a  
176 genetic mark to distinguish different functional clones, and more specifically the number of  
177 passenger mutations unique to a lineage is a measure of the molecular age of that clone.  
178 The **variant allele frequency [G]** (VAF) determines clone abundance, and the proportion of  
179 passenger mutations shared between clones reveals their ancestry<sup>8,30</sup>. The choice of  
180 sequencing assay (high-depth targeted panel, moderate-depth exome, or lower-depth  
181 whole-genome sequencing) represents a trade-off between the need for high-depth  
182 sequencing to accurately recover clone frequency (or even detect the clone at all) versus  
183 genome-wide detection of passenger mutations that uniquely identify distinct clones.  
184 Moreover, since deeper sequencing provides a broader temporal window on cancer  
185 evolution, the choice of sequencing assays is a compromise between genome sequencing  
186 providing detail on the clonal architecture in only a short and early time window versus  
187 deep targeted sequencing that provides limited clonal information but greater temporal  
188 range. Here, deeper and broader (for example, more of the genome covered) sequencing is  
189 always preferred.

190

#### 191 *Allelic copy-number correction*

192 The study of evolutionary dynamics of cancer clones is fundamentally concerned with the  
193 relative frequencies of cancer clones over space and time. Many bioinformatics tools have  
194 been created to infer clonal frequencies from 'bulk' sequencing data, such as PyClone<sup>31</sup>,  
195 SciClone<sup>32</sup> and PhyloWGS<sup>33</sup>. Broadly, these tools attempt to identify sets of mutations that  
196 are all at the same frequency, and assign them to clones. These tools have been  
197 instrumental to study cancer evolution from cancer bulk data. However, this task requires  
198 many prior inference steps, each one risking introduction of errors, which are then  
199 propagated through the analysis. Structural alterations (loss, gain and rearrangements of  
200 genetic material) are common in cancer genomes and confound the interpretation of  
201 mutation frequency. Because structural alterations typically alter the copy number of a  
202 locus, they also have an impact on the relative frequency of any single nucleotide variant  
203 (SNV) mutations at that locus. Thus, to assign SNVs to clones, it is necessary to 'correct' for  
204 the impact of copy number alteration (CNA), to turn the allelic frequency of an SNV into a  
205 clone frequency. In theory, this is straightforward: the cellular abundance of any individual

206 mutation is simply a product of its frequency and copy number. However, if the allelic copy  
207 number is incorrectly inferred, then the SNVs in that CNA will be scaled to the wrong  
208 frequency, and so potentially erroneously appear as a new clone. In a tumour composed of  
209 50% cancer cells, the difference in frequency of an SNV present on 1 of 3 copies versus 1 of  
210 4 copies is only about 3%, which is a level of accuracy that is rarely achievable with  
211 moderate-depth sequencing (~100X). Moreover, errors can stem from in the initial  
212 inference of the copy number of the locus. Consequently, errors in the allelic copy number  
213 inference propagate to produce an erroneous clone phylogenetic tree, and give a misleading  
214 picture of the clonal structure of a tumour. Considering only SNVs located in diploid regions,  
215 and exploiting the hitchhiking principle<sup>34</sup> helps, but in a highly **aneuploid [G]** genome risks  
216 discarding the majority of SNVs for downstream evolutionary analysis, and potentially  
217 important driver mutations that define a clone may be missed. There remains a need for  
218 higher resolution data (>100x depth at whole-genome resolution) and improved clonal  
219 decomposition methods that effectively handle error propagation from copy number  
220 assignments. Emerging long-read sequencing technology also offers the hope of  
221 circumventing this issue, as long reads intrinsically 'phase' mutations and so directly reveal  
222 their allelic identity.

223

#### 224 *Single-cell sequencing*

225 Single-cell sequencing is an exciting emerging alternative to bulk sequencing for exploring  
226 tumour evolution<sup>35-39 36,38,40-42</sup>. In theory, sequencing individual cells removes the time bias  
227 inherent to bulk sequencing as all the genetic mutations within the sequenced cell,  
228 irrespective of when the mutations arose, should be detectable. Clonal identity also  
229 becomes evident, removing the need for allelic copy-number correction. However, calling  
230 SNVs in single-cell sequencing remains challenging, as it is not possible to distinguish  
231 mutational signal from noise by aggregating information across multiple sequencing reads  
232 (as is the case for bulk sequencing). Combining information from multiple cells addresses  
233 this issue<sup>43</sup>. By contrast, CNAs can be reliably identified<sup>44</sup> but because the background CNA  
234 rate is still not well understood<sup>45</sup> drawing inferences about temporal evolutionary dynamics  
235 from these data is not straightforward. Nevertheless, single-cell sequencing offers a  
236 powerful route to learning how CNAs accrue (since sequencing individual cells means that  
237 some newly born cells can be analysed prior to the effects of selection, informing the  
238 'background' CNA mutation rate<sup>45</sup>). Single-cell sequencing of cells from a large cancer risks  
239 sequencing many cells that are evolutionary 'dead ends' and would not contribute to the  
240 future disease progression. Simply sequencing large numbers of cells would abrogate this  
241 issue, and moreover, gives a direct means to detect and characterize negative selection<sup>46</sup>,  
242 which cannot be identified by bulk sequencing. We expect single-cell sequencing to become  
243 the tool of choice in the future as sequencing costs continue to fall.

244

#### 245 **Detecting selection**

246 Clonal selection drives cancer evolution, and so naturally there is much interest in  
247 identifying the cause of a clone's selective advantage, but detecting selection comes with  
248 several challenges (**Figure 2**). There are two broad approaches for detecting selection: (1)  
249 clone frequency methods and (2) sequence-based methods, and the two approaches are  
250 often used concurrently.

251

#### 252 *Using clone frequency to detect selection*

253 Broadly, frequency-based methods detect selection by looking for lineages that are more  
254 abundant than is expected in the absence of selection under neutral evolution. Frequency-  
255 based methods have been developed that use the clone size distribution, also referred to as  
256 the site frequency spectrum (SFS), which can be measured by the VAF distribution, after  
257 correction for tumour purity and copy number. The appeal of this approach is that the  
258 shape of the clone size distribution under neutral evolution in a well-mixed population is  
259 well known<sup>47 48 6 49 7</sup>. Multi-region sampling can also be used to measure clone frequency:  
260 selection for an ancestral clone causes it to have a disproportionate number of offspring in  
261 the phylogenetic tree constructed from these data<sup>50</sup>. Hybrid methods that simultaneously  
262 consider the VAF distributions from multi-region sampling also exist<sup>51</sup>.

263 Longitudinal sampling of clone abundance provides a particularly powerful method to  
264 detect selection: clones that grow disproportionately faster than others are likely under  
265 selection<sup>52</sup>. However, longitudinal tissue collection and temporal analyses of solid tumours  
266 is rendered challenging by accessibility to tumour tissue. In due course, as sequencing  
267 technologies improve and costs decline, we anticipate circulating free tumour DNA analyses  
268 will help to circumvent some of these challenges<sup>53,54</sup>.

269 Frequency-based methods are limited by the power to detect deviations from the null  
270 neutral model<sup>8</sup>. Weak selection (e.g. a relative selective advantage ~1%) causes only slow  
271 and slight shifts in clone frequency that may go undetected in moderate-depth sequencing.  
272 The spatial architecture of a tumour presents a complication too — selection is invisible if all  
273 the samples are taken within the selected clone. Moreover, frequency-based methods can  
274 only detect selection if the selected clone is sampled when expanding; once the clone has  
275 taken over the population, the new (fitter) population of tumour cells is homogenous with  
276 respect to the selective alteration, and so the within-tumour evolution reverts to neutral. If  
277 clones are very strongly selected then clonal expansion to fixation [G] in the tumour will be  
278 very quick and so unlikely to be detected. In this case, dense longitudinal sampling is  
279 necessary to accurately detect selection. There are therefore multiple caveats in inferring a  
280 neutral evolution model from single low sequencing depth samples.

281

#### 282 *Using mutational patterns to detect selection*

283 Alternative methods use the burden and type of mutations across the genome to detect  
284 selection (collectively we refer to these as “mutational pattern” methods). These methods  
285 exploit the fact that selection causes an over-representation of the mutations that increase  
286 fitness, but do not necessarily increase the frequency of neutral mutations. Indeed,

287 statistical tools to identify cancer driver mutations across tumours work by considering the  
288 frequency at which a gene is found to be mutated across cancers compared to background  
289 expectation<sup>55</sup>. The dN/dS ratio — the ratio of non-synonymous mutations (N mutations) to  
290 synonymous mutations (S mutations) normalized by their respective likelihood of  
291 occurrence — is a popular sequence-based method for detecting selection. The logic of the  
292 method is that non-synonymous mutations will tend to experience selection, whereas  
293 synonymous mutations will be evolutionary neutral, and so positive selection will cause an  
294 over-representation of NS mutations (dN/dS>1) whereas negative selection will cause an  
295 under-representation of NS mutations (dN/dS<1)<sup>56</sup>. Driver mutations have positive dN/dS  
296 values<sup>23</sup> and newly refined powerful methods for dN/dS calculation have been developed  
297 specifically for cancer data<sup>13,57</sup>.

298 For the dN/dS method to work, there has to be a sufficient number of mutations under  
299 selection in the gene or locus to cause a significant deviation of the ratio away from 1.  
300 Hence, a minimum mutation burden is required to calculate the ratio at all, and so the  
301 method is challenging to apply to individual genes that will carry no more than a few  
302 mutations in a single tumour. Whereas dN/dS methods are suitable for looking across  
303 cohorts, they are hard to apply to individual tumour evolutionary dynamics. Population  
304 demographics also influence the dN/dS ratio in a complex manner and potentially confound  
305 its interpretation<sup>58,59</sup>. Moreover, there is evidence that certain synonymous mutations can  
306 be under purifying selection which can impact dN/dS estimates and inference of selection.

307

### 308 *Stochasticity versus determinism*

309 In small populations, both in cancer and species, stochasticity can dominate the evolution of  
310 even strongly selected mutations<sup>3</sup>, but a large clone in a large population can behave more  
311 deterministically<sup>60</sup>. The threshold between stochastic and becoming deterministic is  
312 inversely proportional to the selective advantage of the mutant<sup>52,61,22</sup>. This ‘evolutionary  
313 rule’ about the transition from stochasticity to determinism has implications for the  
314 predictability of cancer evolution: small, stochastically evolving clones have unpredictable  
315 evolution, whereas large clones evolve more predictably. In other words, we are likely to be  
316 able to accurately predict the evolution of clones that have already grown large enough to  
317 be detected, but an accurate prediction about the emergence of specific minor clones will  
318 be more challenging.

319

### 320 **Chromosome instability in cancer evolution**

321

#### 322 *CIN and clonal fitness*

323 Alterations in copy number affect a greater proportion of the cancer genome than any other  
324 mutations<sup>62</sup>, and can act as “hopeful monsters”<sup>63</sup>, offering potentially high adaptive  
325 advantage to evolving cancers. They result from CIN, a consequence of ongoing errors in  
326 chromosome segregation during mitosis and errors of DNA replication and repair<sup>64,65</sup>. The  
327 end result is aneuploidy (an unbalanced chromosome complement) involving entire



328 chromosomes (whole-chromosome aneuploidy) or parts of chromosomes (partial or  
329 segmental aneuploidy). Aneuploidy can also occur independently of CIN if a single event of  
330 chromosome missegregation leads to expansion of the aneuploid clone followed by clonal  
331 stasis, that is, without detectable ongoing CIN. Such tumours are homogeneously or clonally  
332 aneuploid, whereas tumours with ongoing CIN are heterogeneously or subclonally  
333 aneuploid<sup>66,67</sup>. In addition, aneuploidy can result from single catastrophic events, termed  
334 **chromoplexy [G]** (if affecting multiple chromosomes) or **chromothripsis [G]** (affecting 1–2  
335 chromosomes), the relevance of which has become increasingly evident across different  
336 cancer types<sup>68</sup>. Whatever the mechanism, aneuploidy can alter the somatic copy number,  
337 and therefore expression, of many genes at the same time. Although the background  
338 alteration rate varies substantially across chromosomes<sup>45</sup> it does not account for evidence  
339 of recurrent chromosomal level or arm level aberrations in tumours<sup>69</sup>, which can best be  
340 explained by selection (positive or purifying). Location of tumour suppressor genes and  
341 oncogenes re-capitulates the patterns of aneuploidy observed across different cancers<sup>70,71</sup>  
342 also shows the adaptive potential provided by CIN. In a mouse model of acute lymphoblastic  
343 leukaemia and hepatocellular carcinoma, induction of CIN in T cells and hepatocytes  
344 resulted in tumour-specific patterns of chromosome copy alterations, suggesting that  
345 selective pressure is tissue context-dependent<sup>72</sup>. CIN can also provide means of disease  
346 escape following curative treatment with surgery or disease control with targeted therapy.  
347 Induction of CIN in the KRAS model of lung cancer resulted in rapid relapse, with recurrent  
348 tumours showing high levels of aneuploidy<sup>73</sup>, with emergent independence from the  
349 original oncogenic stimulus. In chronic myeloid leukaemia patients who developed  
350 resistance to BCR-ABL targeting imatinib developed additional chromosomal alterations<sup>74</sup>.

351

352 Some effects of CIN are independent of gene-specific alterations, including reduced  
353 proliferation, proteotoxic stress, metabolic changes, upregulation of the stress response and  
354 further genome instability. The latter in particular has a profound impact as aneuploid cells  
355 continue to create more genetic diversity<sup>75,76</sup>. The fact that aneuploidy (or CIN) can be both  
356 detrimental and advantageous highlights the importance of determining the selective  
357 landscape. This is well illustrated in yeast where aneuploidy provides a fitness advantage  
358 under severe environmental conditions, acting as “a first evolutionary line of defence”<sup>77</sup>, but  
359 does not persist upon reversion to normal conditions. In a systematic study of the oncogenic  
360 potential of aneuploidy in mouse embryonic fibroblasts (MEFs), trisomy failed to induce  
361 transformation under any conditions and the cells grew poorly compared to matched  
362 euploid cells, consistent with a fitness penalty. However, during long-term growth, triploid  
363 cells acquired other aneuploidies that conferred improved fitness<sup>78</sup>. The authors suggest  
364 that low levels of aneuploidy may be tumour-protective, but that the genome-destabilizing  
365 effects of aneuploidy are tumour-promoting under certain growth conditions. Thus, the rare  
366 growth-promoting aneuploidies expand and rise to clonal levels, whilst growth-inhibitory  
367 aneuploidies are selected against. Consistent with this notion, aneuploid cells grew better  
368 than euploid cells under conditions of environmental stress such as hypoxia and

369 chemotherapy<sup>79</sup>. Addition of a single chromosome increased the tolerance to  
370 environmental stresses and was not chromosome-specific, suggesting that overexpression  
371 of particular genes is not the only contributor to adaptive potential.

372

### 373 *CIN and metastases*

374 Complex processes of metastatic spread, which require a multitude of cellular phenotypes  
375 could be well served by the karyotypic and phenotypic heterogeneity generated by CIN.  
376 Comparative studies of matched primary tumour–metastasis pairs have reported  
377 enrichment for aneuploidy in metastatic lesions from prostate, pancreatic, breast and colon  
378 cancers (reviewed in<sup>80</sup>). Through a detailed clonal resolution of matched clear cell renal cell  
379 cancer (ccRCC) primary and metastatic tumours, we recently reported that a critical  
380 difference between tumour clones that are metastasis-competent compared to those that  
381 fail to metastasize is the degree of aneuploidy and chromosome complexity (measured by  
382 fluorescence-activated cell sorting (FACS) and weighted genome instability index<sup>81 82</sup>).  
383 Furthermore, we observed that specific somatic CNAs, loss of 9p and loss of 14q, were  
384 highly enriched within the metastasizing clones, reflecting active selection. We found no  
385 evidence of selection for the smaller scale mutations such as SNVs<sup>82</sup>. Beyond altering the  
386 expression of many genes simultaneously potential mechanism by which chromosomal  
387 alterations contribute to metastasis include the induction of mesenchymal transition  
388 through changes in expression of intercellular junction proteins<sup>83</sup>, activation of cGAS–  
389 stimulator of interferon genes (STING) pathway by cytosolic DNA from chromosome  
390 missegregation<sup>84 85,86</sup>, and immune evasion<sup>87</sup>.

391

### 392 *CIN and clinical outcomes*

393 The role of CIN in cancer evolution and progression is evidenced by its association with poor  
394 clinical outcomes in a number of retrospective studies<sup>88,89</sup>. More recently, analyses in a  
395 prospective cohort of early stage non-small cell lung cancer (NSCLC) evolution (TRACERx-  
396 Lung study) showed that CIN confers an increased risk of recurrence and death  
397 independently of known predictive markers<sup>26</sup>. In TRACERx-Renal, a similarly prospective  
398 study of clear cell renal cell carcinoma (ccRCC), increase in aneuploidy was associated with  
399 shorter progression-free and overall survival<sup>25</sup>. Intriguingly, the level of CIN has a bearing on  
400 its overall impact on prognosis. In a pan-cancer analysis of >2,000 samples, only moderate  
401 levels of CIN (>25% and <75%) were associated with decreased survival, concordant with  
402 previous studies showing that excessive levels of CIN confer an improved prognosis<sup>90,91 92</sup>.  
403 These observations are consistent with a fitness cost of CIN, with the selective advantage of  
404 karyotypic heterogeneity negated by excessive levels of aneuploidy.

405

406 CIN is also linked to resistance to anti-cancer treatment, including chemotherapy<sup>93,94</sup>, and  
407 CTLA4 and PD1 immune checkpoint inhibitors<sup>87 95</sup>. In NSCLC, CIN can lead to subclonal loss  
408 of heterozygosity (LOH) in the genes encoding the human leukocyte antigen (HLA)<sup>96</sup>, with  
409 pervasive evidence of positive selection for this event in tumours. In this context, HLA LOH

410 facilitates accumulation of subclonal neoantigens, and further clonal evolution<sup>96</sup>. In ccRCC,  
411 we observed increased rates of HLA LOH in primary tumour subclones that were selected in  
412 metastatic sites, highlighting again the role of immune evasion in metastasis<sup>82</sup>.

413

#### 414 **Evolutionary patterns and patient outcomes**

415 A critical question is whether understanding a tumour's evolutionary trajectory and  
416 evolutionary potential can help to predict patient outcomes. In particular, the presence of  
417 clonal diversity is expected to provide a rich repertoire of alterations that could be adaptive  
418 under selective pressure of therapy, alterations in tumour environment or metastatic  
419 colonisation of distant sites. In a prospective study of Barrett's oesophagus, a premalignant  
420 condition, progression to adenocarcinoma correlated with clonal diversity independently of  
421 other genetic risk factors<sup>97</sup>. Multiple studies have demonstrated the link between subclonal  
422 diversification and adverse clinical outcomes in chronic lymphocytic leukaemia<sup>98 99</sup>, head  
423 and neck cancer<sup>100</sup> ovarian cancer<sup>101</sup> and across other cancer types<sup>102</sup>. Subclonal  
424 diversification of somatic CNA and mutational drivers was associated with adverse  
425 prognostic features in ccRCC, and independently associated with reduced progression-  
426 free and overall survival<sup>25</sup>. In NSCLC, diversity of SCNAs but not SNVs correlated with the  
427 risk of relapse and death<sup>26</sup>. In patients with breast cancer, intratumour heterogeneity of  
428 HER2 copy number, detected at single-cell resolution, was associated with shorter  
429 survival<sup>103</sup>.

430

431 However, lack of detectable clonal diversity does not always correlate with improved clinical  
432 outcome. In multiple myeloma, detection of neutral evolution dynamics correlated with  
433 progression-free and overall survival<sup>106</sup> and associated with the presence of a strong clonal  
434 oncogenic driver, which might explain the lack of ongoing selection. It is also increasingly  
435 apparent that some tumours acquire multiple and/or strong drivers in a short period of time  
436 (punctuated evolution), whereas others show a more steady rate of driver acquisition (gradual  
437 evolution)<sup>10,107-109</sup>. The result of punctuated evolution is a rapid clonal sweep and a fairly  
438 homogenous tumour mass. In ccRCC, these tumours are characterized by low driver  
439 intratumour heterogeneity and high levels of clonal aneuploidy which became fixed early  
440 on in tumour evolution. These tumours proliferated faster, disseminated rapidly to many  
441 different sites (Figure 3a), and had worse outcome, compared to those characterized by  
442 clonal diversity and subclonal aneuploidy<sup>25</sup>. Metastases from rapidly evolving tumours  
443 were seeded by the same dominant clone found at the primary site resulting in limited  
444 inter-metastatic heterogeneity in untreated patients (Figure 3a). By contrast, tumours  
445 with subclonal aneuploidy, evolving in a Darwinian fashion and gradually accumulating  
446 driver alterations, grew more slowly and over longer periods of time. In some cases  
447 metastases were seeded by multiple clones resulting in inter-metastatic heterogeneity (in  
448 untreated patients). In line with this, a mathematical model of metastases formation  
449 suggests that the probability of observing inter-metastatic heterogeneity (which results  
450 from distinct clones in the primary tumour seeding different metastatic sites) increases

451 when the primary tumour grows slowly<sup>110</sup>. Intriguingly, gradually evolving tumours were  
452 also associated with a specific pattern of metastatic progression, termed  
453 “oligometastases”<sup>82</sup> (Figure 3b). Oligometastases, defined as a small number of lesions  
454 confined to a single site, are conceptualized as an “intermediate state of metastatic  
455 capacity”<sup>111,112</sup> with an important clinical implication for directed, potentially curative  
456 treatment for such lesions. Reduced metastatic efficiency of clonally diverse tumours  
457 could be a result of clonal interference (inter-clonal competition at the primary tumour  
458 site) or a reflection of weak clonal drivers, with subclonal driver events providing  
459 additional fitness required for metastases.

460 Pancreatic cancer has traditionally been viewed as following gradual evolution  
461 with sequential acquisition of driver events. However, some pancreatic cancers show  
462 punctuated equilibrium as the principle evolutionary trajectory, whereby multiple driver  
463 events are acquired, sometimes through a single ‘catastrophic’ event results in complex  
464 genomic rearrangements<sup>113</sup>. Consistent with our observations in renal cancer, such  
465 evolutionary trajectories result in limited inter-metastatic heterogeneity, as all  
466 metastases are seeded by the dominant primary tumour clone<sup>114</sup>. Another example is  
467 uveal melanoma, characterized by aggressive though latent liver metastases in a  
468 proportion of patients, especially those whose primary tumour harbours *BAP1* mutations.  
469 *BAP1* mutations and chromosomal complexity were shown to arise in a short burst early  
470 on in tumouregensis<sup>107</sup>, implying that metastatic potential can be acquired at the earliest  
471 stages of cancer evolution. Similar observations have been made in triple-negative breast  
472 cancer<sup>109</sup>, while chromoplexy and chromotripsis were shown to fuel rapid evolution in  
473 prostate cancer and colorectal cancer<sup>108,115</sup>, respectively.

474

475 Finally, the temporal order in which mutations are acquired during tumour evolution  
476 impacts the clinical phenotype and outcome in myeloproliferative neoplasms<sup>104</sup>, ccRCC  
477<sup>25</sup>, NSCLC and breast cancer<sup>105</sup>. These observations are consistent with determinism, and  
478 suggest that evolutionary trajectories could potentially be predicted for patient benefit.

479

480 The observation of the wide spectrum of evolutionary patterns in cancer begins to  
481 reconcile the diverse clinical phenotypes and varied outcomes seen in the clinic. In  
482 particular, the occurrence of punctuated genomic evolution highlights the challenge of  
483 managing cancers that acquire metastatic competency early, cancers that are ‘born to be  
484 bad’. Supporting this notion are pre-clinical models which show metastatic dissemination  
485 before frank malignancy is detected histologically<sup>116</sup>. These observations are especially  
486 relevant for cancer screening approaches. As the latency between the emergence of the  
487 invasive clone and metastatic spread can be short the window for early detection could  
488 be very limited<sup>117</sup>. Many questions about evolutionary trajectories remain including the  
489 environmental conditions which favour gradual evolution (gradual accumulation of driver  
490 mutations), or punctuated evolution (large-scale rearrangements of the genome leading  
491 to many drivers acquired at once) and how these may be altered for therapeutic benefit.

492

## 493 **Origin of the treatment-resistant clone**

494

### 495 *Resistance to targeted therapies*

496 Targeting oncogenic drivers in both blood and solid malignancies has brought about a  
497 remarkable change in the cancer treatment landscape. Notable examples include BCR/ABL  
498 translocation in chronic myeloid leukaemia (CML), where the use of imatinib has resulted in  
499 10-year survival rates of ~85%<sup>118</sup>; KIT mutations in gastrointestinal stromal tumours (GISTs),  
500 HER2 amplification in breast cancer, EGFR mutations in NSCLC, and BRAF mutations in  
501 melanoma. However, with the exception of CML, disease control afforded by targeted  
502 agents is fairly short-lived, and treatment rarely results in long-term survival for the patient.  
503 Mutational complexity of solid cancers may be a contributing factor to inevitability of  
504 resistance, as every additional mutation could provide a pathway to treatment resistance.  
505 Accordingly, higher tumour mutational burden (TMB) correlates with shortened benefit  
506 from EGFR-tyrosine kinase inhibitor (TKI) in metastatic EGFR-mutant NSCLC<sup>119</sup>. Although  
507 resistance mutations can arise de-novo<sup>120</sup>, they frequently pre-exist as minor subclones  
508 (Figure 4a)<sup>121,122</sup> though the ability to detect them in pre-treatment samples is limited by  
509 the breadth of sampling and depth of sequencing. Modelling of tumour growth suggests  
510 that detectable metastatic lesions can harbour ten or more resistant subclones<sup>123</sup>. Although  
511 there are limitations to these models (reviewed in<sup>124</sup>), the predictions are consistent with  
512 the observations in clinical and genomic data. In a recent study of patients with chronic  
513 lymphoid leukaemia treated with ibrutinib, resistance was attributable to the emergence  
514 of mutations in BTK and/or PLCG2 which were detected with a high-sensitivity method up  
515 to 15 months prior to clinical progression, with some patients evolving multiple  
516 resistance mutations<sup>125</sup>. Polyclonal treatment resistance has been described in other  
517 tumour types, with evidence of parallel expansion of clones harbouring distinct mechanisms  
518 of resistance under selective pressure of therapy<sup>126-128</sup>. Upfront evaluation of the resistant  
519 clones can also be used to forecast the duration of therapeutic benefit, as recently  
520 demonstrated in metastatic colorectal cancer using frequent time-course liquid biopsies and  
521 mathematical modelling<sup>129</sup>.

522 Thus, a comprehensive catalogue of resistant mutations could inform appropriate  
523 combinatorial strategies, while dynamic monitoring of emerging and resolving alterations  
524 can facilitate adaptive treatment strategies. This approach was well illustrated by the  
525 example of EGFR inhibition in colorectal cancer and the waxing and waning of the resistant  
526 RAS-mutant alleles in the blood in response to treatment initiation and withdrawal<sup>130</sup>.  
527 These observations also highlight the issue of fitness penalty associated with resistant  
528 mutations: KRAS mutations were detected in cell-free DNA from patients who developed  
529 resistance to EGFR inhibition; however, when therapy was withdrawn they remained  
530 undetectable, suggesting that they require ongoing therapy for their maintenance and that  
531 resistance comes at a cost. The higher the fitness cost, the harder it is for the resistant clone  
532 to emerge as modelled in **xenografts** derived from patients (**[G]** (PDXs) with BRAF-V600E

533 mutant melanoma or NSCLC, who developed resistance to BRAF inhibition. PDXs were  
534 exposed to ERK inhibition (downstream of BRAF), which resulted in multiple BRAF-amplified  
535 clones being selected and propagated. When BRAF, MEK and ERK inhibition were combined  
536 in an intermittent schedule, the fitness disadvantage prevented the emergence of the BRAF-  
537 amplified subclones <sup>131</sup>. Finally, clonal complexity may impact the drug target itself.  
538 Although frequently clonal by virtue of being founder alterations, drug targets can also be  
539 found in tumour subclones. In a recent clinical trial FGFR inhibitor responders harboured a  
540 clonal FGFR amplification, whereas non-responders harboured subclonal amplifications <sup>132</sup>.

541

#### 542 *Resistance to immune checkpoint inhibition*

543 Another important development in cancer therapeutics has been the advent of **immune**  
544 **checkpoint blockade** [G]. The efficacy of checkpoint inhibitors (CPIs) is contingent on pre-  
545 existing recognition of the tumour by the immune system, through presentation of  
546 neoantigens which result from somatic mutations accumulated by the tumour. Accordingly,  
547 the best responses are observed in tumours with an abundance of somatic mutations (that  
548 is, a high TMB), which increases the likelihood of a potent neoantigen being presented to  
549 the immune system. Initially, it was expected that CPIs might circumvent the clonal diversity  
550 faced by targeted therapies; however, it has become apparent that clonal evolution has a  
551 profound impact on immunotherapy success and failure. Subclonal neoantigens do not  
552 stimulate an adequate tumour response, as shown by reduced sensitivity to checkpoint  
553 blockade in melanoma and NSCLC tumours that have a significant proportion of subclonal  
554 mutations <sup>133</sup>. This pattern has been confirmed across additional tumour types <sup>134</sup>.  
555 Neoantigen evolution, or immune-editing, underlies some aspects of acquired resistance to  
556 CPIs. Both clonal and subclonal neoantigens loss under selective pressure of CPI treatment  
557 have been reported. Clonal neoantigens are lost through deletion of the chromosome  
558 region that harbours the alteration, whereas subclonal neoantigens are lost through  
559 outgrowth of alternative subclones <sup>135</sup>. Critically, peptides generated from the lost  
560 neoantigens elicited clonal T-cell expansion in autologous T-cell cultures, suggesting that  
561 they generated functional immune responses <sup>135</sup>. Neoantigen immune editing has also been  
562 reported in the context of adoptive transfer of autologous lymphocytes that specifically  
563 target proteins encoded by cancer-specific mutations, another area of active clinical  
564 development which holds much promise <sup>136</sup>. T-cell recognized neoantigens were selectively  
565 lost over time in metastatic melanomas treated by adoptive T-cell transfer <sup>137</sup>, accompanied  
566 by development of neoantigen-specific T-cell reactivity in tumour-infiltrating lymphocytes,  
567 indicating immunediting.

568 Inactivation of antigen presentation is another important mechanism of acquired CPI  
569 resistance. For example, point mutations, deletions or LOH in *B2M*, which encodes an  
570 essential component of MHC class I antigen presentation, and in the genes encoding  
571 interferon-receptor-associated Janus kinase 1 (JAK1) or JAK2, have all been reported as  
572 common mechanisms <sup>138 37</sup>. Just as with the drivers of resistance to targeted therapy, these  
573 alterations were selected and expanded under therapy. Vaccine strategies are also

574 vulnerable to these alterations. In a trial of an RNA-based vaccine against a spectrum of  
575 cancer mutations, neo-epitope-specific killing was demonstrated in a patient who initially  
576 responded, but developed resistance owing to the outgrowth of  $\beta$ 2-microglobulin-deficient  
577 melanoma cells <sup>139</sup>. Another mechanism of immune evasion occurs through selection of  
578 tumour populations where HLA is either mutated or lost. In a recent report of adoptive T  
579 cell transfer in a patient with colorectal cancer, profiling of a progressive lesion revealed loss  
580 of the chromosome 6 haplotype encoding the HLA allele that recognized the targeted  
581 mutant KRAS <sup>140</sup>.

582

### 583 **Conclusions and perspective**

584 An understanding of the dynamics of cancer evolution might lead to improvement in clinical  
585 outcomes, as it enables prognoses to be accurately determined and 'evolution-aware'  
586 patient management to be applied. Genomic analysis provides a quantitative measurement  
587 of evolutionary dynamics and evolutionary potential. There is tremendous value still to be  
588 gleaned from analyses of the rapidly-growing public repository of cancer genomic data;  
589 particular insight can be gained from the large sample numbers and the inter-comparison of  
590 evolutionary dynamics between cancer types. However, we caution that our inferences are  
591 severely restricted by the limitations of single-biopsy, bulk-sequenced data sets. As  
592 sequencing costs continue to fall, deeper sequencing will allow more accurate  
593 determination of clonal fractions (reducing error on inferences derived from these data) and  
594 enable the resolution of smaller clones. Single-cell sequencing technology promises to  
595 circumvent much of the complexity of 'bulk' sequencing data, and this maturing technology  
596 promises the concurrent measurement of genotypes and phenotypes in individual cells <sup>141</sup>,  
597 together with a characterization of their in-situ microenvironment <sup>36</sup>.

598 Improving the availability of samples from which to study cancer evolutionary  
599 dynamics also presents a bottleneck: we hope initiatives such as our TRACERx <sup>142</sup> and PEACE  
600 <sup>143</sup> studies, which provide infrastructure for longitudinal and post-mortem collection of  
601 tumour samples, will become more common. Even at a single time-point, these studies  
602 provide greater representative tumour sampling relative to single-tumour biopsies, which  
603 under-represent tumour bulk, leading to the risk of clonal illusion. Quantitative genomic  
604 analysis of 'liquid biopsies' (the analysis of tumour DNA from peripheral blood samples) may  
605 overcome this issue and provide an amenable route for minimally-invasive longitudinal  
606 disease monitoring as well as predictions on disease course and treatment response  
607 <sup>53,129,144-146</sup>. In summary, evolutionary genomics provides an ever-improving lens to reveal  
608 the clonal dynamics of cancer and impact patient outcomes.

609 **Box 1. Is cancer a special case of evolution?**

610 Despite major overlaps between evolutionary biology and cancer biology, there are a few  
611 aspects of cancer evolution that indicate tumours may be a special case of evolutionary  
612 systems. First, tumours are extremely large populations, much larger than most common  
613 ecosystems and more akin to bacteria colonies, with populations in the order of 100s of  
614 billions of cells. This implies that the total diversity is astounding. Another special feature of  
615 cancers is that chromosomal instability, which is central to cancer evolution. Chromosomal  
616 instability allows for the generation of true 'hopeful monsters' — grossly altered clones that  
617 may be adaptive — a phenomena thought to be very rare in species evolution. Cancer cell  
618 plasticity, or phenotypic change that does not require underlying heritable variation, is also  
619 a fundamental force that guides tumour adaptation and makes the system rather 'non-  
620 Darwinian' in some contexts.

621

622 **Figure 1. Modes of cancer evolution.** Cancers evolve according to Darwinian rules: mutation  
623 and selection of beneficial new mutations drives the expansion of subclones, and between  
624 and within selected clones, the cellular populations experience neutral drift. Different  
625 'modes' of evolution appear depending on when and how the evolutionary process is  
626 sampled.

627

628 **Figure 2. Challenges in detecting selection.** **a.** Limited sampling in time and space  
629 confounds measurement of evolutionary dynamics. (i) Sampling within a clone shows  
630 neutral dynamics. (ii) Non-uniform spatial sampling can look like selection when it is absent  
631 because of genetic divergence, or *vice versa*. (iii) If driver mutations accrue rarely but exhibit  
632 a strong effect, most evolutionary time shows only neutral dynamics. (iv) Selection occurs  
633 within a small niche that is below the detection limit, so evolution appears neutral because  
634 selected subclones are undetectable. (v) Using frequency/phylogenetic methods, selection  
635 can only be detected when a clone boundary is sampled. **b.** Bulk sequencing data has a  
636 profound time bias, allowing only the earliest – and so highest frequency – mutations to be  
637 detected. As a tumour doubles its cell number, new mutations that arise represent an  
638 exponentially smaller fraction of the tumour, and so rapidly fall below detectable frequency.  
639 **c.** Error in copy-number assignment propagates and confounds the identification of tumour  
640 subclones. Limited depth sequencing (say 100X) causes dispersion in the true VAF of a  
641 variant, and true VAF is determined by clonal abundance and underlying copy-number state  
642 (coloured shapes on plot). This leads to mutations in different clones, or at different copy-  
643 number states, being erroneously misassigned clonal identities (red boxes). The  $1/f^2$  tail of  
644 low frequency mutations is an inevitable consequence of tumour growth, and further  
645 complicates clonal inference on VAF data.

646

647 **Figure 3. Clonal evolution and metastases.** Different modes of evolution in the primary  
648 tumour can impact the mode of metastatic progression<sup>25</sup>. Metastatic capacity is associated  
649 with increased chromosome complexity<sup>82</sup>. **A.** Tumours that evolve in a punctuated fashion



650 with early onset of clonal chromosome complexity grow rapidly and metastasise early and  
651 widely. Metastases are monophyletic (single dominant clone seeds all the metastatic sites)  
652 and monoclonal (single clone seeds single site), and there is limited inter-metastatic  
653 heterogeneity. **B.** Tumours that evolve in a branched/Darwinian fashion grow more slowly  
654 are composed of distinct subpopulations of cells with differential metastatic capacity and  
655 chromosome complexity is acquired late. They can be associated with solitary or oligo-  
656 metastases. When they spread to multiple sites they may do so in a polyphyletic fashion  
657 (different subclones seed different sites), which may include organ-specific patterns and  
658 result in inter-metastatic heterogeneity<sup>110</sup>. If multiple clones seed the same site the  
659 metastasis is polyclonal.

660

661 **Figure 4. Clonal evolution of treatment resistance. A.** Resistant mutations can be present in  
662 the tumour population before the start of therapy, usually as a minor subclone<sup>123,125</sup>. They  
663 may evade detection in the baseline sample if they are present at very low frequency or a  
664 restricted to an unsampled region of the tumour. They may be even neutral or deleterious  
665 before therapy. Under the selective pressure of therapy, the treatment-sensitive population  
666 diminishes leaving the resistant population to expand under positive selection. Multiple  
667 subclones bearing distinct resistant mutations can emerge at the same time, indicating  
668 parallel evolution of resistance<sup>126-129</sup>. **B.** Treatment resistance can be a result of a de novo  
669 mutation which carries a selection advantage under therapy and becomes fixed in the  
670 tumour population. In this case resistance takes longer to emerge<sup>120</sup>.

671

## 672 References

- 673 1 McGranahan, N. & Swanton, C. Biological and therapeutic impact of intratumor  
674 heterogeneity in cancer evolution. *Cancer Cell* **27**, 15-26,  
675 doi:10.1016/j.ccell.2014.12.001 (2015).
- 676 2 <[https://en.wikipedia.org/wiki/The\\_Genetical\\_Theory\\_of\\_Natural\\_Selection](https://en.wikipedia.org/wiki/The_Genetical_Theory_of_Natural_Selection)> (  
677 3 Lynch, M. *et al.* Genetic drift, selection and the evolution of the mutation rate. *Nat*  
678 *Rev Genet* **17**, 704-714, doi:10.1038/nrg.2016.104 (2016).
- 679 4 Bozic, I. *et al.* Evolutionary dynamics of cancer in response to targeted combination  
680 therapy. *Elife* **2**, e00747, doi:10.7554/eLife.00747 (2013).
- 681 5 Bozic, I. *et al.* Accumulation of driver and passenger mutations during tumor  
682 progression. *Proc Natl Acad Sci U S A* **107**, 18545-18550,  
683 doi:10.1073/pnas.1010978107 (2010).
- 684 6 Durrett, R. Population Genetics of Neutral Mutations in Exponentially Growing  
685 Cancer Cell Populations. *Ann Appl Probab* **23**, 230-250 (2013).
- 686 7 Williams, M. J., Werner, B., Barnes, C. P., Graham, T. A. & Sottoriva, A.  
687 Identification of neutral tumor evolution across cancer types. *Nat Genet* **48**, 238-244,  
688 doi:10.1038/ng.3489 (2016).
- 689 8 Williams, M. J. *et al.* Quantification of subclonal selection in cancer from bulk  
690 sequencing data. *Nat Genet* **50**, 895-903, doi:10.1038/s41588-018-0128-6 (2018).
- 691 9 Iwasa, Y., Nowak, M. A. & Michor, F. Evolution of resistance during clonal  
692 expansion. *Genetics* **172**, 2557-2566, doi:10.1534/genetics.105.049791 (2006).
- 693 10 Tsao, J. L. *et al.* Genetic reconstruction of individual colorectal tumor histories. *Proc*  
694 *Natl Acad Sci U S A* **97**, 1236-1241 (2000).

- 695 11 Altrock, P. M., Liu, L. L. & Michor, F. The mathematics of cancer: integrating  
696 quantitative models. *Nat Rev Cancer* **15**, 730-745, doi:10.1038/nrc4029 (2015).
- 697 12 Marty, R., Thompson, W. K., Salem, R. M., Zanetti, M. & Carter, H. Evolutionary  
698 Pressure against MHC Class II Binding Cancer Mutations. *Cell* **175**, 416-428 e413,  
699 doi:10.1016/j.cell.2018.08.048 (2018).
- 700 13 Zapata, L. *et al.* Negative selection in tumor genome evolution acts on essential  
701 cellular functions and the immunopeptidome. *Genome Biol* **19**, 67,  
702 doi:10.1186/s13059-018-1434-0 (2018).
- 703 14 Donnelly, P. & Tavaré, S. The population genealogy of the infinitely--many neutral  
704 alleles model. *J Math Biol* **25**, 381-391 (1987).
- 705 15 Griffiths, R. C. The frequency spectrum of a mutation, and its age, in a general  
706 diffusion model. *Theor Popul Biol* **64**, 241-251 (2003).
- 707 16 McFarland, C. D., Mirny, L. A. & Korolev, K. S. Tug-of-war between driver and  
708 passenger mutations in cancer and other adaptive processes. *Proc Natl Acad Sci U S A*  
709 **111**, 15138-15143, doi:10.1073/pnas.1404341111 (2014).
- 710 17 McFarland, C. D., Korolev, K. S., Kryukov, G. V., Sunyaev, S. R. & Mirny, L. A.  
711 Impact of deleterious passenger mutations on cancer progression. *Proc Natl Acad Sci*  
712 *U S A* **110**, 2910-2915, doi:10.1073/pnas.1213968110 (2013).
- 713 18 Sansregret, L. *et al.* APC/C Dysfunction Limits Excessive Cancer Chromosomal  
714 Instability. *Cancer Discov* **7**, 218-233, doi:10.1158/2159-8290.CD-16-0645 (2017).
- 715 19 R, S. D., Gutteridge, A., Swanton, C., Maley, C. C. & Graham, T. A. Modelling the  
716 evolution of genetic instability during tumour progression. *Evol Appl* **6**, 20-33,  
717 doi:10.1111/eva.12024 (2013).
- 718 20 Loeb, L. A. Mutator phenotype in cancer: origin and consequences. *Semin Cancer*  
719 *Biol* **20**, 279-280, doi:10.1016/j.semcancer.2010.10.006 (2010).
- 720 21 Kimura, M. *The Neutral Theory of Molecular Evolution*. (Cambridge University  
721 Press, 1983).
- 722 22 Hughes, A. L. Near neutrality: leading edge of the neutral theory of molecular  
723 evolution. *Ann N Y Acad Sci* **1133**, 162-179, doi:10.1196/annals.1438.001 (2008).
- 724 23 Martincorena, I. *et al.* Tumor evolution. High burden and pervasive positive selection  
725 of somatic mutations in normal human skin. *Science* **348**, 880-886,  
726 doi:10.1126/science.aaa6806 (2015).
- 727 24 Lee-Six, H. *et al.* Population dynamics of normal human blood inferred from somatic  
728 mutations. *Nature* **561**, 473-478, doi:10.1038/s41586-018-0497-0 (2018).
- 729 25 Turajlic, S. *et al.* Deterministic Evolutionary Trajectories Influence Primary Tumor  
730 Growth: TRACERx Renal. *Cell* **173**, 595-610 e511, doi:10.1016/j.cell.2018.03.043  
731 (2018).
- 732 26 Jamal-Hanjani, M. *et al.* Tracking the Evolution of Non-Small-Cell Lung Cancer. *N*  
733 *Engl J Med* **376**, 2109-2121, doi:10.1056/NEJMoa1616288 (2017).
- 734 27 Gerlinger, M. *et al.* Cancer: evolution within a lifetime. *Annu Rev Genet* **48**, 215-236,  
735 doi:10.1146/annurev-genet-120213-092314 (2014).
- 736 28 Markowitz, F. A saltationist theory of cancer evolution. *Nat Genet* **48**, 1102-1103,  
737 doi:10.1038/ng.3687 (2016).
- 738 29 Eldredge, N. & Gould, S. J. On punctuated equilibria. *Science* **276**, 338-341 (1997).
- 739 30 Nik-Zainal, S. *et al.* The life history of 21 breast cancers. *Cell* **149**, 994-1007,  
740 doi:10.1016/j.cell.2012.04.023 (2012).
- 741 31 Roth, A. *et al.* PyClone: statistical inference of clonal population structure in cancer.  
742 *Nat Methods* **11**, 396-398, doi:10.1038/nmeth.2883 (2014).

- 743 32 Miller, C. A. *et al.* SciClone: inferring clonal architecture and tracking the spatial and  
744 temporal patterns of tumor evolution. *PLoS Comput Biol* **10**, e1003665,  
745 doi:10.1371/journal.pcbi.1003665 (2014).
- 746 33 Deshwar, A. G. *et al.* PhyloWGS: reconstructing subclonal composition and evolution  
747 from whole-genome sequencing of tumors. *Genome Biol* **16**, 35, doi:10.1186/s13059-  
748 015-0602-8 (2015).
- 749 34 Smith, J. M. & Haigh, J. The hitch-hiking effect of a favourable gene. *Genet Res* **89**,  
750 391-403, doi:10.1017/S0016672308009579 (2007).
- 751 35 Kim, C. *et al.* Chemoresistance Evolution in Triple-Negative Breast Cancer  
752 Delineated by Single-Cell Sequencing. *Cell* **173**, 879-893 e813,  
753 doi:10.1016/j.cell.2018.03.041 (2018).
- 754 36 Casasent, A. K. *et al.* Multiclonal Invasion in Breast Tumors Identified by  
755 Topographic Single Cell Sequencing. *Cell* **172**, 205-217 e212,  
756 doi:10.1016/j.cell.2017.12.007 (2018).
- 757 37 Gao, J. *et al.* Loss of IFN-gamma Pathway Genes in Tumor Cells as a Mechanism of  
758 Resistance to Anti-CTLA-4 Therapy. *Cell* **167**, 397-404 e399,  
759 doi:10.1016/j.cell.2016.08.069 (2016).
- 760 38 Eirew, P. *et al.* Dynamics of genomic clones in breast cancer patient xenografts at  
761 single-cell resolution. *Nature* **518**, 422-426, doi:10.1038/nature13952 (2015).
- 762 39 Xu, X. *et al.* Single-cell exome sequencing reveals single-nucleotide mutation  
763 characteristics of a kidney tumor. *Cell* **148**, 886-895, doi:10.1016/j.cell.2012.02.025  
764 (2012).
- 765 40 Zhang, K. Stratifying tissue heterogeneity with scalable single-cell assays. *Nat*  
766 *Methods* **14**, 238-239, doi:10.1038/nmeth.4209 (2017).
- 767 41 McPherson, A. *et al.* Divergent modes of clonal spread and intraperitoneal mixing in  
768 high-grade serous ovarian cancer. *Nat Genet* **48**, 758-767, doi:10.1038/ng.3573  
769 (2016).
- 770 42 Leung, M. L. *et al.* Highly multiplexed targeted DNA sequencing from single nuclei.  
771 *Nat Protoc* **11**, 214-235, doi:10.1038/nprot.2016.005 (2016).
- 772 43 Roth, A. *et al.* Clonal genotype and population structure inference from single-cell  
773 tumor sequencing. *Nat Methods* **13**, 573-576, doi:10.1038/nmeth.3867 (2016).
- 774 44 Zahn, H. *et al.* Scalable whole-genome single-cell library preparation without  
775 preamplification. *Nat Methods* **14**, 167-173, doi:10.1038/nmeth.4140 (2017).
- 776 45 Worrall, J. T. *et al.* Non-random Mis-segregation of Human Chromosomes. *Cell Rep*  
777 **23**, 3366-3380, doi:10.1016/j.celrep.2018.05.047 (2018).
- 778 46 Emma Laks, H. Z., Daniel Lai, Andrew McPherson, Adi Steif, Jazmine Brimhall,  
779 Justina Biele, Beixi Wang, Tehmina Masud, Diljot Grewal, Cydney Nielsen,  
780 Samantha Leung, Viktoria Bojilova, Maia Smith, Oleg Golovko, Steven Poon, Peter  
781 Eirew, Farhia Kabeer, Teresa Ruiz de Algara, So Ra Lee, M. Jafar Taghiyar, Curtis  
782 Huebner, Jessica Ngo, Tim Chan, Spencer Vattrt-Watts, Pascale Walters, Nafis Abrar,  
783 Sophia Chan, Matt Wiens, Lauren Martin, R. Wilder Scott, Michael T. Underhill,  
784 Elizabeth Chavez, Christian Steidl, Daniel Da Costa, Yusanne Ma, Robin J. N.  
785 Coope, Richard Corbett, Stephen Pleasance, Richard Moore, Andy J. Mungall, CRUK  
786 IMAXT Consortium, Marco A. Marra, Carl Hansen, Sohrab Shah, Samuel Aparicio.  
787 *Resource: Scalable whole genome sequencing of 40,000 single cells identifies*  
788 *stochastic aneuploidies, genome replication states and clonal repertoires* (BIORXIV,  
789 2018).
- 790 47 Luria, S. E. & Delbruck, M. Mutations of Bacteria from Virus Sensitivity to Virus  
791 Resistance. *Genetics* **28**, 491-511 (1943).

792 48 Maruvka, Y. E., Kessler, D. A. & Shnerb, N. M. The birth-death-mutation process: a  
793 new paradigm for fat tailed distributions. *PLoS One* **6**, e26480,  
794 doi:10.1371/journal.pone.0026480 (2011).

795 49 Kessler, D. A. & Levine, H. Large population solution of the stochastic Luria-  
796 Delbruck evolution model. *Proc Natl Acad Sci U S A* **110**, 11682-11687,  
797 doi:10.1073/pnas.1309667110 (2013).

798 50 Bozic, I., Gerold, J. M. & Nowak, M. A. Quantifying Clonal and Subclonal Passenger  
799 Mutations in Cancer Evolution. *PLoS Comput Biol* **12**, e1004731,  
800 doi:10.1371/journal.pcbi.1004731 (2016).

801 51 Sun, R. *et al.* Between-region genetic divergence reflects the mode and tempo of  
802 tumor evolution. *Nat Genet* **49**, 1015-1024, doi:10.1038/ng.3891 (2017).

803 52 Levy, S. F. *et al.* Quantitative evolutionary dynamics using high-resolution lineage  
804 tracking. *Nature* **519**, 181-186, doi:10.1038/nature14279 (2015).

805 53 Abbosh, C. *et al.* Phylogenetic ctDNA analysis depicts early-stage lung cancer  
806 evolution. *Nature* **545**, 446-451, doi:10.1038/nature22364 (2017).

807 54 Murtaza, M. *et al.* Multifocal clonal evolution characterized using circulating tumour  
808 DNA in a case of metastatic breast cancer. *Nat Commun* **6**, 8760,  
809 doi:10.1038/ncomms9760 (2015).

810 55 Lawrence, M. S. *et al.* Mutational heterogeneity in cancer and the search for new  
811 cancer-associated genes. *Nature* **499**, 214-218, doi:10.1038/nature12213 (2013).

812 56 Wu, C. I., Wang, H. Y., Ling, S. & Lu, X. The Ecology and Evolution of Cancer: The  
813 Ultra-Microevolutionary Process. *Annu Rev Genet* **50**, 347-369, doi:10.1146/annurev-  
814 genet-112414-054842 (2016).

815 57 Martincorena, I. *et al.* Universal Patterns of Selection in Cancer and Somatic Tissues.  
816 *Cell* **171**, 1029-1041 e1021, doi:10.1016/j.cell.2017.09.042 (2017).

817 58 Rocha, E. P. *et al.* Comparisons of dN/dS are time dependent for closely related  
818 bacterial genomes. *J Theor Biol* **239**, 226-235, doi:10.1016/j.jtbi.2005.08.037 (2006).

819 59 Kryazhimskiy, S. & Plotkin, J. B. The population genetics of dN/dS. *PLoS Genet* **4**,  
820 e1000304, doi:10.1371/journal.pgen.1000304 (2008).

821 60 Clark, D. L. H. a. A. G. *Principles of Population Genetics*. 4th edn, (2006).

822 61 Kimura, M. *Neutral theory of molecular evolution*. (1983).

823 62 Beroukhim, R. *et al.* The landscape of somatic copy-number alteration across human  
824 cancers. *Nature* **463**, 899-905, doi:10.1038/nature08822 (2010).

825 63 Goldschmidt, R. *he Material Basis of Evolution*. second edn, (Yale University,  
826 1982).

827 64 Burrell, R. A. *et al.* Replication stress links structural and numerical cancer  
828 chromosomal instability. *Nature* **494**, 492-496, doi:10.1038/nature11935 (2013).

829 65 Bakhoun, S. F. *et al.* The mitotic origin of chromosomal instability. *Curr Biol* **24**,  
830 R148-149, doi:10.1016/j.cub.2014.01.019 (2014).

831 66 Heng, H. H. *et al.* Chromosomal instability (CIN): what it is and why it is crucial to  
832 cancer evolution. *Cancer Metastasis Rev* **32**, 325-340, doi:10.1007/s10555-013-9427-  
833 7 (2013).

834 67 Heng, H. H., Regan, S. M., Liu, G. & Ye, C. J. Why it is crucial to analyze non clonal  
835 chromosome aberrations or NCCAs? *Mol Cytogenet* **9**, 15, doi:10.1186/s13039-016-  
836 0223-2 (2016).

837 68 Leibowitz, M. L., Zhang, C. Z. & Pellman, D. Chromothripsis: A New Mechanism  
838 for Rapid Karyotype Evolution. *Annu Rev Genet* **49**, 183-211, doi:10.1146/annurev-  
839 genet-120213-092228 (2015).

840 69 Zack, T. I. *et al.* Pan-cancer patterns of somatic copy number alteration. *Nat Genet*  
841 **45**, 1134-1140, doi:10.1038/ng.2760 (2013).

842 70 Davoli, T. *et al.* Cumulative haploinsufficiency and triplosensitivity drive aneuploidy  
843 patterns and shape the cancer genome. *Cell* **155**, 948-962,  
844 doi:10.1016/j.cell.2013.10.011 (2013).

845 71 Solimini, N. L. *et al.* Recurrent hemizygous deletions in cancers may optimize  
846 proliferative potential. *Science* **337**, 104-109, doi:10.1126/science.1219580 (2012).

847 72 Fojter, F. *et al.* Deletion of the MAD2L1 spindle assembly checkpoint gene is  
848 tolerated in mouse models of acute T-cell lymphoma and hepatocellular carcinoma.  
849 *Elife* **6**, doi:10.7554/eLife.20873 (2017).

850 73 Sotillo, R., Schwartzman, J. M., Socci, N. D. & Benezra, R. Mad2-induced  
851 chromosome instability leads to lung tumour relapse after oncogene withdrawal.  
852 *Nature* **464**, 436-440, doi:10.1038/nature08803 (2010).

853 74 Hochhaus, A. *et al.* Molecular and chromosomal mechanisms of resistance to imatinib  
854 (STI571) therapy. *Leukemia* **16**, 2190-2196, doi:10.1038/sj.leu.2402741 (2002).

855 75 Targa, A. & Rancati, G. Cancer: a CINful evolution. *Curr Opin Cell Biol* **52**, 136-  
856 144, doi:10.1016/j.ceb.2018.03.007 (2018).

857 76 Tang, Y. C. & Amon, A. Gene copy-number alterations: a cost-benefit analysis. *Cell*  
858 **152**, 394-405, doi:10.1016/j.cell.2012.11.043 (2013).

859 77 Yona, A. H. *et al.* Chromosomal duplication is a transient evolutionary solution to  
860 stress. *Proc Natl Acad Sci U S A* **109**, 21010-21015, doi:10.1073/pnas.1211150109  
861 (2012).

862 78 Sheltzer, J. M. *et al.* Single-chromosome Gains Commonly Function as Tumor  
863 Suppressors. *Cancer Cell* **31**, 240-255, doi:10.1016/j.ccell.2016.12.004 (2017).

864 79 Rutledge, S. D. *et al.* Selective advantage of trisomic human cells cultured in non-  
865 standard conditions. *Sci Rep* **6**, 22828, doi:10.1038/srep22828 (2016).

866 80 Turajlic, S. & Swanton, C. Metastasis as an evolutionary process. *Science* **352**, 169-  
867 175, doi:10.1126/science.aaf2784 (2016).

868 81 Endesfelder, D. *et al.* Chromosomal instability selects gene copy-number variants  
869 encoding core regulators of proliferation in ER+ breast cancer. *Cancer Res* **74**, 4853-  
870 4863, doi:10.1158/0008-5472.CAN-13-2664 (2014).

871 82 Turajlic, S. *et al.* Tracking Cancer Evolution Reveals Constrained Routes to  
872 Metastases: TRACERx Renal. *Cell* **173**, 581-594 e512,  
873 doi:10.1016/j.cell.2018.03.057 (2018).

874 83 Gao, C. *et al.* Chromosome instability drives phenotypic switching to metastasis. *Proc*  
875 *Natl Acad Sci U S A* **113**, 14793-14798, doi:10.1073/pnas.1618215113 (2016).

876 84 Bakhom, S. F. *et al.* Chromosomal instability drives metastasis through a cytosolic  
877 DNA response. *Nature* **553**, 467-472, doi:10.1038/nature25432 (2018).

878 85 Mackenzie, K. J. *et al.* cGAS surveillance of micronuclei links genome instability to  
879 innate immunity. *Nature* **548**, 461-465, doi:10.1038/nature23449 (2017).

880 86 Umbreit, N. T. & Pellman, D. Cancer biology: Genome jail-break triggers lockdown.  
881 *Nature* **550**, 340-341, doi:10.1038/nature24146 (2017).

882 87 Davoli, T., Uno, H., Wooten, E. C. & Elledge, S. J. Tumor aneuploidy correlates with  
883 markers of immune evasion and with reduced response to immunotherapy. *Science*  
884 **355**, doi:10.1126/science.aaf8399 (2017).

885 88 Carter, S. L., Eklund, A. C., Kohane, I. S., Harris, L. N. & Szallasi, Z. A signature of  
886 chromosomal instability inferred from gene expression profiles predicts clinical  
887 outcome in multiple human cancers. *Nat Genet* **38**, 1043-1048, doi:10.1038/ng1861  
888 (2006).

889 89 Walther, A., Houlston, R. & Tomlinson, I. Association between chromosomal  
890 instability and prognosis in colorectal cancer: a meta-analysis. *Gut* **57**, 941-950,  
891 doi:10.1136/gut.2007.135004 (2008).

- 892 90 Roylance, R. *et al.* Relationship of extreme chromosomal instability with long-term  
893 survival in a retrospective analysis of primary breast cancer. *Cancer Epidemiol*  
894 *Biomarkers Prev* **20**, 2183-2194, doi:10.1158/1055-9965.EPI-11-0343 (2011).
- 895 91 Birkbak, N. J. *et al.* Paradoxical relationship between chromosomal instability and  
896 survival outcome in cancer. *Cancer Res* **71**, 3447-3452, doi:10.1158/0008-  
897 5472.CAN-10-3667 (2011).
- 898 92 Jamal-Hanjani, M. *et al.* Extreme chromosomal instability forecasts improved  
899 outcome in ER-negative breast cancer: a prospective validation cohort study from the  
900 TACT trial. *Ann Oncol* **26**, 1340-1346, doi:10.1093/annonc/mdv178 (2015).
- 901 93 Swanton, C. *et al.* Chromosomal instability determines taxane response. *Proc Natl*  
902 *Acad Sci U S A* **106**, 8671-8676, doi:10.1073/pnas.0811835106 (2009).
- 903 94 Duesberg, P., Stindl, R. & Hehlmann, R. Explaining the high mutation rates of cancer  
904 cells to drug and multidrug resistance by chromosome reassortments that are  
905 catalyzed by aneuploidy. *Proc Natl Acad Sci U S A* **97**, 14295-14300,  
906 doi:10.1073/pnas.97.26.14295 (2000).
- 907 95 Roh, W. *et al.* Integrated molecular analysis of tumor biopsies on sequential CTLA-4  
908 and PD-1 blockade reveals markers of response and resistance. *Sci Transl Med* **9**,  
909 doi:10.1126/scitranslmed.aah3560 (2017).
- 910 96 McGranahan, N. *et al.* Allele-Specific HLA Loss and Immune Escape in Lung Cancer  
911 Evolution. *Cell* **171**, 1259-1271 e1211, doi:10.1016/j.cell.2017.10.001 (2017).
- 912 97 Maley, C. C. *et al.* Genetic clonal diversity predicts progression to esophageal  
913 adenocarcinoma. *Nat Genet* **38**, 468-473, doi:10.1038/ng1768 (2006).
- 914 98 Landau, D. A. *et al.* Evolution and impact of subclonal mutations in chronic  
915 lymphocytic leukemia. *Cell* **152**, 714-726, doi:10.1016/j.cell.2013.01.019 (2013).
- 916 99 Nadeu, F. *et al.* Clinical impact of the subclonal architecture and mutational  
917 complexity in chronic lymphocytic leukemia. *Leukemia* **32**, 645-653,  
918 doi:10.1038/leu.2017.291 (2018).
- 919 100 Mroz, E. A. *et al.* High intratumor genetic heterogeneity is related to worse outcome  
920 in patients with head and neck squamous cell carcinoma. *Cancer* **119**, 3034-3042,  
921 doi:10.1002/cncr.28150 (2013).
- 922 101 Schwarz, R. F. *et al.* Spatial and temporal heterogeneity in high-grade serous ovarian  
923 cancer: a phylogenetic analysis. *PLoS Med* **12**, e1001789,  
924 doi:10.1371/journal.pmed.1001789 (2015).
- 925 102 Andor, N. *et al.* Pan-cancer analysis of the extent and consequences of intratumor  
926 heterogeneity. *Nat Med* **22**, 105-113, doi:10.1038/nm.3984 (2016).
- 927 103 Rye, I. H. *et al.* Intra-tumor heterogeneity defines treatment-resistant HER2+ breast  
928 tumors. *Mol Oncol*, doi:10.1002/1878-0261.12375 (2018).
- 929 104 Ortmann, C. A. *et al.* Effect of mutation order on myeloproliferative neoplasms. *N*  
930 *Engl J Med* **372**, 601-612, doi:10.1056/NEJMoa1412098 (2015).
- 931 105 Caravagna, G. *et al.* Detecting repeated cancer evolution from multi-region tumor  
932 sequencing data. *Nat Methods* **15**, 707-714, doi:10.1038/s41592-018-0108-x (2018).
- 933 106 Johnson, D. C. *et al.* Neutral tumor evolution in myeloma is associated with poor  
934 prognosis. *Blood* **130**, 1639-1643, doi:10.1182/blood-2016-11-750612 (2017).
- 935 107 Field, M. G. *et al.* Punctuated evolution of canonical genomic aberrations in uveal  
936 melanoma. *Nat Commun* **9**, 116, doi:10.1038/s41467-017-02428-w (2018).
- 937 108 Baca, S. C. *et al.* Punctuated evolution of prostate cancer genomes. *Cell* **153**, 666-  
938 677, doi:10.1016/j.cell.2013.03.021 (2013).
- 939 109 Gao, R. *et al.* Punctuated copy number evolution and clonal stasis in triple-negative  
940 breast cancer. *Nat Genet* **48**, 1119-1130, doi:10.1038/ng.3641 (2016).

941 110 Reiter, J. G. *et al.* Minimal functional driver gene heterogeneity among untreated  
942 metastases. *Science* **361**, 1033-1037, doi:10.1126/science.aat7171 (2018).

943 111 Hellman, S. & Weichselbaum, R. R. Oligometastases. *J Clin Oncol* **13**, 8-10,  
944 doi:10.1200/JCO.1995.13.1.8 (1995).

945 112 Weichselbaum, R. R. & Hellman, S. Oligometastases revisited. *Nat Rev Clin Oncol* **8**,  
946 378-382, doi:10.1038/nrclinonc.2011.44 (2011).

947 113 Notta, F. *et al.* A renewed model of pancreatic cancer evolution based on genomic  
948 rearrangement patterns. *Nature* **538**, 378-382, doi:10.1038/nature19823 (2016).

949 114 Makohon-Moore, A. P. *et al.* Limited heterogeneity of known driver gene mutations  
950 among the metastases of individual patients with pancreatic cancer. *Nat Genet* **49**,  
951 358-366, doi:10.1038/ng.3764 (2017).

952 115 Stephens, P. J. *et al.* Massive genomic rearrangement acquired in a single catastrophic  
953 event during cancer development. *Cell* **144**, 27-40, doi:10.1016/j.cell.2010.11.055  
954 (2011).

955 116 Rhim, A. D. *et al.* EMT and dissemination precede pancreatic tumor formation. *Cell*  
956 **148**, 349-361, doi:10.1016/j.cell.2011.11.025 (2012).

957 117 Baker, A. M. *et al.* Evolutionary history of human colitis-associated colorectal cancer.  
958 *Gut*, doi:10.1136/gutjnl-2018-316191 (2018).

959 118 Hochhaus, A. *et al.* Long-Term Outcomes of Imatinib Treatment for Chronic Myeloid  
960 Leukemia. *N Engl J Med* **376**, 917-927, doi:10.1056/NEJMoa1609324 (2017).

961 119 Offin, M. *et al.* Tumor Mutation Burden and Efficacy of EGFR-Tyrosine Kinase  
962 Inhibitors in Patients with EGFR-Mutant Lung Cancers. *Clin Cancer Res*,  
963 doi:10.1158/1078-0432.CCR-18-1102 (2018).

964 120 Hata, A. N. *et al.* Tumor cells can follow distinct evolutionary paths to become  
965 resistant to epidermal growth factor receptor inhibition. *Nat Med* **22**, 262-269,  
966 doi:10.1038/nm.4040 (2016).

967 121 Misale, S. *et al.* Emergence of KRAS mutations and acquired resistance to anti-EGFR  
968 therapy in colorectal cancer. *Nature* **486**, 532-536, doi:10.1038/nature11156 (2012).

969 122 Diaz, L. A., Jr. *et al.* The molecular evolution of acquired resistance to targeted EGFR  
970 blockade in colorectal cancers. *Nature* **486**, 537-540, doi:10.1038/nature11219  
971 (2012).

972 123 Bozic, I. & Nowak, M. A. Timing and heterogeneity of mutations associated with  
973 drug resistance in metastatic cancers. *Proc Natl Acad Sci U S A* **111**, 15964-15968,  
974 doi:10.1073/pnas.1412075111 (2014).

975 124 Pogrebniak, K. L. & Curtis, C. Harnessing Tumor Evolution to Circumvent  
976 Resistance. *Trends Genet* **34**, 639-651, doi:10.1016/j.tig.2018.05.007 (2018).

977 125 Ahn, I. E. *et al.* Clonal evolution leading to ibrutinib resistance in chronic  
978 lymphocytic leukemia. *Blood* **129**, 1469-1479, doi:10.1182/blood-2016-06-719294  
979 (2017).

980 126 Bettegowda, C. *et al.* Detection of circulating tumor DNA in early- and late-stage  
981 human malignancies. *Sci Transl Med* **6**, 224ra224, doi:10.1126/scitranslmed.3007094  
982 (2014).

983 127 Juric, D. *et al.* Convergent loss of PTEN leads to clinical resistance to a PI(3)Kalpha  
984 inhibitor. *Nature* **518**, 240-244, doi:10.1038/nature13948 (2015).

985 128 Shi, H. *et al.* Acquired resistance and clonal evolution in melanoma during BRAF  
986 inhibitor therapy. *Cancer Discov* **4**, 80-93, doi:10.1158/2159-8290.CD-13-0642  
987 (2014).

988 129 Khan, K. H. *et al.* Longitudinal Liquid Biopsy and Mathematical Modeling of Clonal  
989 Evolution Forecast Time to Treatment Failure in the PROSPECT-C Phase II

990 Colorectal Cancer Clinical Trial. *Cancer Discov* **8**, 1270-1285, doi:10.1158/2159-  
991 8290.CD-17-0891 (2018).

992 130 Siravegna, G. *et al.* Clonal evolution and resistance to EGFR blockade in the blood of  
993 colorectal cancer patients. *Nat Med* **21**, 827, doi:10.1038/nm0715-827b (2015).

994 131 Xue, Y. *et al.* An approach to suppress the evolution of resistance in BRAF(V600E)-  
995 mutant cancer. *Nat Med* **23**, 929-937, doi:10.1038/nm.4369 (2017).

996 132 Pearson, A. *et al.* High-Level Clonal FGFR Amplification and Response to FGFR  
997 Inhibition in a Translational Clinical Trial. *Cancer Discov* **6**, 838-851,  
998 doi:10.1158/2159-8290.CD-15-1246 (2016).

999 133 McGranahan, N. *et al.* Clonal neoantigens elicit T cell immunoreactivity and  
1000 sensitivity to immune checkpoint blockade. *Science* **351**, 1463-1469,  
1001 doi:10.1126/science.aaf1490 (2016).

1002 134 Miao, D. *et al.* Genomic correlates of response to immune checkpoint blockade in  
1003 microsatellite-stable solid tumors. *Nat Genet*, doi:10.1038/s41588-018-0200-2 (2018).

1004 135 Anagnostou, V. *et al.* Evolution of Neoantigen Landscape during Immune Checkpoint  
1005 Blockade in Non-Small Cell Lung Cancer. *Cancer Discov* **7**, 264-276,  
1006 doi:10.1158/2159-8290.CD-16-0828 (2017).

1007 136 Zacharakis, N. *et al.* Immune recognition of somatic mutations leading to complete  
1008 durable regression in metastatic breast cancer. *Nat Med* **24**, 724-730,  
1009 doi:10.1038/s41591-018-0040-8 (2018).

1010 137 Verdegaal, E. M. *et al.* Neoantigen landscape dynamics during human melanoma-T  
1011 cell interactions. *Nature* **536**, 91-95, doi:10.1038/nature18945 (2016).

1012 138 Zaretsky, J. M. *et al.* Mutations Associated with Acquired Resistance to PD-1  
1013 Blockade in Melanoma. *N Engl J Med* **375**, 819-829, doi:10.1056/NEJMoa1604958  
1014 (2016).

1015 139 Sahin, U. *et al.* Personalized RNA mutanome vaccines mobilize poly-specific  
1016 therapeutic immunity against cancer. *Nature* **547**, 222-226, doi:10.1038/nature23003  
1017 (2017).

1018 140 Tran, E. *et al.* T-Cell Transfer Therapy Targeting Mutant KRAS in Cancer. *N Engl J*  
1019 *Med* **375**, 2255-2262, doi:10.1056/NEJMoa1609279 (2016).

1020 141 Macaulay, I. C. *et al.* G&T-seq: parallel sequencing of single-cell genomes and  
1021 transcriptomes. *Nat Methods* **12**, 519-522, doi:10.1038/nmeth.3370 (2015).

1022 142 <http://tracerx.co.uk/>.

1023 143 [https://www.cancerresearchuk.org/about-cancer/find-a-clinical-trial/a-study-looking-](https://www.cancerresearchuk.org/about-cancer/find-a-clinical-trial/a-study-looking-at-blood-and-tissue-samples-to-learn-more-about-advanced-cancer-peace)  
1024 [at-blood-and-tissue-samples-to-learn-more-about-advanced-cancer-peace](https://www.cancerresearchuk.org/about-cancer/find-a-clinical-trial/a-study-looking-at-blood-and-tissue-samples-to-learn-more-about-advanced-cancer-peace).

1025 144 Gray, E. S. *et al.* Circulating tumor DNA to monitor treatment response and detect  
1026 acquired resistance in patients with metastatic melanoma. *Oncotarget* **6**, 42008-  
1027 42018, doi:10.18632/oncotarget.5788 (2015).

1028 145 Spina, V. *et al.* Circulating tumor DNA reveals genetics, clonal evolution, and  
1029 residual disease in classical Hodgkin lymphoma. *Blood* **131**, 2413-2425,  
1030 doi:10.1182/blood-2017-11-812073 (2018).

1031 146 O'Leary, B. *et al.* Early circulating tumor DNA dynamics and clonal selection with  
1032 palbociclib and fulvestrant for breast cancer. *Nat Commun* **9**, 896,  
1033 doi:10.1038/s41467-018-03215-x (2018).

1034

1035 **Glossary**

1036

1037 **Clonal evolution**



1038 A process by which genetic and epigenetic alterations create diversity that acts as substrate  
1039 for natural selection.

1040

1041 **Subclone**

1042 A populations of cells in the tumor that harbour the same set of genomic alterations

1043

1044 **Genetic drift**

1045 A stochastic random process that changes subclone frequency

1046

1047 **Selection**

1048 A non-random process shaped by environmental and tumour properties that changes  
1049 subclone frequency

1050

1051 **Molecular evolution**

1052 Evolutionary change at the level of DNA sequence.

1053

1054 **Somatic evolution**

1055 Accumulation of genomic alterations in somatic cells

1056

1057 **Chromosome instability**

1058 A type of genomic instability that involves parts of or entire chromosomes.

1059

1060 **Mutator phenotype**

1061 Increase in mutation rates in cancer

1062

1063 **Neutral evolution**

1064 Clonal diversity not caused by selection

1065

1066 **Phylogenetic tree**

1067 A branching diagram showing the hierarchy of clones within the tumour

1068

1069 **Clonal sweep**

1070 Reduction of diversity due to the fixation of a variant due to strong positive selection.

1071

1072 **Punctuated equilibrium**

1073 Rapid speciation events with long periods of intervening stasis.

1074

1075 **Hopeful monster**

1076 The generation of an individual with a grossly-altered genome compared to its ancestor,  
1077 which may be adaptive. A hopeful monster is the result of punctuated change in the  
1078 genome.

1079  
1080 **Passenger mutation**  
1081 A mutation that has no effect on clone fitness  
1082  
1083 **Driver mutation**  
1084 A mutation that increases clone fitness  
1085  
1086 **Variant Allele Frequency**  
1087 Relative frequency of a variant in a tumour sample, expressed as a percentage  
1088  
1089 **Aneuploid**  
1090 The presence of an abnormal chromosome complement  
1091  
1092 **Fixation**  
1093 Rise of a variant in frequency in the population to 100%  
1094  
1095 **Chromoplexy**  
1096 A complex rearrangement of the cancer genome that involves a number of chromosomes  
1097  
1098 **Chromothripsis**  
1099 A complex rearrangement of the cancer genome that involves a single chromosome  
1100  
1101 **Patient-derived xenografts**  
1102 A tumour model where the tissue from patient's tumour is implanted in an immunodeficient  
1103 mouse.  
1104  
1105 **Immune checkpoint blockade**  
1106 Therapies that target immune checkpoints such as CTLA4 and PD1 which tumours can use to  
1107 escape anti-tumour immune responses