Genetic and non-genetic clonal diversity in cancer evolution

Authors: James R.M. Black[1,2] and Nicholas McGranahan[1,2†]

1. Cancer Genome Evolution Research Group, University College London Cancer Institute, University College London, London, UK
2. Cancer Research UK Lung Cancer Center of Excellence, University College London Cancer Institute, Paul O'Gorman Building, London, UK

[†]email: nicholas.mcgranahan.10@ucl.ac.uk

Abstract

The observation and analysis of intra- tumour heterogeneity (ITH), particularly in genomic studies, has advanced our understanding of the evolutionary forces that shape cancer growth and development. However, only a subset of the variation observed in a single tumour will have an impact on cancer evolution, highlighting the need to distinguish between functional and non- functional ITH. Emerging studies highlight a role for the cancer epigenome, transcriptome and immune microenvironment in functional ITH. Here, we consider the importance of both genetic and non- genetic ITH and their role in tumour evolution and present the rationale for a broad research focus beyond the cancer genome. Systems- biology analytical approaches will be necessary to outline the scale and importance of functional ITH. By allowing a deeper understanding of tumour evolution this will, in time, encourage development of novel therapies and improve outcomes for patients.

Introduction

In 1859, in his seminal text *On the Origin of Species,* Charles Darwin described a theory of branching evolution by natural selection, based on observations of incredible diversity of phenotypes amongst animals[1]. At the time, the mechanisms underpinning such diversity, including genetic recombination and mutation, were unknown, but such were the strength of his deductions about the function of these variations that this theory remains in the scientific mainstream today.

Cancer is also an evolutionary process[2,3], and it was the observation of phenotypic heterogeneity within tumours that led Nowell to hypothesise that Darwinian clonal evolution underpinned their development[2]. Since then, 'intra-tumour' heterogeneity (ITH), describing diversity within individual tumours, has been defined at multiple different levels, including single point mutations, somatic copy number alterations (SCNAs), epigenetic and transcriptomic changes influencing gene expression, the antitumour immune response and other features of the tumour microenvironment.

An important task that remains is to distinguish between 'functional' variation, conferring a fitness effect that brings about an important change in tumour phenotype, from 'non-functional' variation[4] (Figure 1). Indeed, the extent to which ITH is a result of the stochastic accumulation of mutations following the acquisition of founding 'driver' events, rather than the result of continual clonal evolution and selection through time and space remains an open scientific question and a topic of debate[5]. It is likely that the true extent of selection during cancer evolution varies both between cancer types, and within individual cancers of the same type (Box 1).

This Review will explore what is known about functional and non-functional ITH in cancer, outlining unresolved areas of debate that warrant further study. It will address the diverse causes of ITH and how this might impact cancer treatment and prognosis. The advent of high throughput multi-omics has increased our understanding of the interplay between different cellular processes that contribute to ITH[6]. In order to appreciate in full the importance of ITH in cancer, we must interrogate more than just mutations, SCNAs, or gene expression in isolation; rather, we must seek to link all factors that may influence tumour phenotype. With a systems-biology lens such as this, we may gain the resolution required to better understand cancer evolution and comprehend its origins and vulnerabilities.

ITH and evolution

Cancer develops through clonal evolution[2,3]. Genetic variation acts as a primary substrate for this evolution. This arises through different mechanisms of genomic instability, including endogenous and exogenous processes that generate point mutations, as well as chromosomal instability (CIN)[7,8].

Early studies of tumour heterogeneity developed the notion of cancer as an evolutionary, as well as a genetic, disease. Interphase fluorescence in-situ hybridisation and karyotyping of metaphase chromosomes in the 1990s demonstrated the presence of multiple clones within a tumour[9,10]. Comparative genomic hybridisation microarray analysis enabled accurate characterisation of the copy-number profiles of cancer clones and expanded on these findings. For example, Navin and colleagues classified breast tumours as either 'monogenomic', containing a population of near-homogeneous tumour cells with analogous genomic profiles, or 'polygenomic', containing subpopulations or 'clones' with distinct genomic profiles, and demonstrated that clones in polygenomic tumours were descended from a common ancestor by branched evolution[11]. The advent of next-generation sequencing has enabled this to be characterised with finer granularity. In a seminal study in 2012,

141 Gerlinger and colleagues performed multi-region whole-exome sequencing on tumour
142 samples from a cohort of patients with renal cell carcinoma to describe heterogeneity in
143 putative driver mutations within the same tumour, and ongoing branched evolution over
144 time[12].
145 Since these early observations, a central focus of research has been to distil functional
146 from non-functional somatic variation and to characterise the strength of the evolutionary
147 forces that shape tumour development over time. However, the extent to which cancer is
148 under continuous selection during its development is a contentious topic in the field, echoing
149 a long-standing debate in evolutionary biology[13]. Observed differences between parts of the
150 same tumour do not themselves indicate the presence of competing subclones which are
151 under selection. The accumulation of random somatic alterations over time means that
152 genotypes will diverge even in the absence of selection pressures. As such, some studies have
153 suggested that a subset of tumours may evolve neutrally following the acquisition of
154 necessary driver events[14–18].
155 Low ITH of driver events, potentially indicative of selective 'clonal sweeps' of
156 certain phenotypes early in tumour evolution, is well-described in multiple cancer types[19–23].
157 Conversely, other studies have highlighted the presence of subclonal driver mutations in
158 cancer genes, whereby only a subset of cancer cells, or clones, harbour functional somatic
159 events assumed to confer a fitness advantage[24–29].
160
161 **<u>Methods for evaluating ITH</u>**
162
163 Accurately recapitulating the evolutionary history of a tumour from a genomic
164 snapshot, typically provided at a single time point, and frequently from a single part of a
165 tumour, can be problematic. Despite consensus that one size may not fit all where tumour
166 evolution is concerned, non-uniform terminology and methods of analysis continue to
167 hamper efforts to classify tumours by their patterns of evolutionary development.
168 Many studies, such as a seminal analysis of 21 breast cancer genomes by Campbell
169 and colleagues[30], perform bulk sequencing on a single sample and attempt to infer the
170 evolutionary history of the tumour from the variant allelic frequencies (VAF) of somatic
171 mutations. Such studies must attempt to account for variables such as the amount of non-
172 tumour tissue sampled, the SCNA profile of the tumour, as well as the accuracy and depth of
173 sequencing, all of which can confound accurate interpretation of the data. Another problem
174 among such studies is sampling bias; a variant may be present in all cancer cells sampled, but
175 not all cancer cells in the tumour (Figure 2). In addition, it may not be possible to distinguish
176 the VAF of mutations within subclones that are the product of selection from those within the
177 VAF distribution 'tail' which is a feature of neutral evolution[16,31,32]. This tail-like distribution
178 of read counts of passenger mutations that are not under selection reflects random
179 mutagenesis at each cell division and the expected relationship between the number of
180 mutational events and their clonal frequency over time.
181 Multi-region sampling, while it does not mitigate against neutral tails, may help to
182 identify and classify subclones more accurately and reduce sampling bias. Nevertheless, this
183 does not represent a 'silver bullet'; typically, only a tiny fraction of the total tumour is
184 assessed[33]. Sequencing of every single tumour cell, which remains impossible, would be
185 required to resolve this definitively. Another option is representative sequencing, whereby
186 homogenised fixed tumour material that was not used for pathology is sequenced, thus
187 reducing sampling bias and misclassification rates[33].
188 Even if clonal sweeps are accurately defined, distinguishing between such an event
189 that may have occurred in the recent past, and one that was present at the inception of the
190 tumour is not always possible. Furthermore, negative selection that has eliminated a clone

191 prior to sampling is not detectable using VAF-based approaches. Therefore, other methods
192 are required to disentangle the temporal order of events and to distinguish functional events,
193 subject to selection, from non-functional events, which are not.
194 Copy-number gains and whole-genome duplication events may be used to time
195 somatic alterations and separate early from late events occurring during tumour evolution[29,34].
196 In the Pan Cancer Analysis of Whole Genomes (PCAWG), this approach enabled inference
197 of trends in clonal architecture across tumour types[29]. In colorectal adenocarcinoma, for
198 instance, mutations in *APC, KRAS* and *TP53* were shown to be predominantly early events,
199 while certain copy-number alterations including 15q, 21q and 22q loss, were generally late[29].
200 A method borrowed from evolutionary biology can provide orthogonal evidence of
201 clonal selection. dN/dS, an assessment of the ratio of substitution rates at nonsynonymous vs.
202 synonymous sites, revealed substantial selection pressures acting on apparently normal
203 tissue[35]. In cancer, dN/dS has also been used to reveal positive selection globally, and within
204 specific cancer genes, as well as a near-absence of negative, or purifying selection[36] (Figure
205 3). Whilst dN/dS provides information on global patterns of selection, particularly when large
206 numbers of samples are analysed, it is nevertheless unable to infer selection within specific
207 elements of interest in individual clones (Figure 3C). In addition, further research is needed
208 to enable accurate measurement of selection in the context of other events such as non-coding
209 mutations, indels or structural variants.
210 Robustly describing tumour evolution with a single binary label, such as neutral or
211 branched, punctuated or gradual[37], remains problematic. Indeed, separate analyses of
212 identical data can lead to disparate conclusions: a comparison of VAFs in 904 cancers from
213 14 cancer types, including a minority subjected to multi-region sampling, by Williams and
214 colleagues suggested that the subclonal VAF distributions in 36% (323/904) of these could be
215 explained by neutral evolution[16], but subsequent work using an orthogonal approach (dN/dS),
216 found evidence of subclonal selection within cancer genes in these 323 tumours [16,38].
217
218 Single-cell sequencing platforms may provide further insight into tumour evolution
219 (reviewed elsewhere[39]). This approach presents the opportunity to analyse cellular
220 epigenomes, as assessed through DNA methylation, histone configuration or chromatin
221 accessibility, and cell states, inferred from transcriptomic or protein expression data. This
222 enables explicit evolutionary context to be added to epigenetic or transcriptomic events.
223 Lineage reconstruction at the single cell level has the potential to improve vastly our
224 understanding of tumour phylogeny. An exemplar of this is Direct Library Preparation single-
225 cell whole-genome sequencing, which allows for identification of clonal populations of
226 single cells, pinpointing unique aspects of their genomes[40]. This enables aggregation into
227 'pseudo-bulk' samples from which clonal phylogeny may inferred. At present, however, the
228 accuracy of methods to call mutations from typically shallow single-cell sequencing may be
229 limited by PCR errors and allelic dropout. Efforts to resolve this issue will enhance the future
230 success of this approach.
231
232 **[H1]Copy-number ITH**
233
234 So far, the main form of variation discussed in this review has been point mutations.
235 However, CIN, and structural variation including SCNAs, whole genome duplication and
236 chromothripsis events may engender large fitness effects. The extent to which CIN provides
237 the substrate for ongoing selection and branched evolution in cancer is unclear. Although
238 CIN has long been linked to poor prognosis[7], research has only recently explored the question
239 of whether this process continues throughout tumour evolution, or, conversely, whether
240 SCNAs are predominantly relics of past genomic instability. A single-cell study of 12 triple-

negative breast cancers identified aneuploidy as a punctuated, early event in tumour evolution, preceding clonal expansion of a population of cells with a stable karyotype[20]. This finding was supported by recent work on colorectal cancer reported in a preprint; Cross et al studied CIN within 755 samples, taken from multiple tumour regions or longitudinally in 167 patients and found that certain dominant SCNAs were established early in tumour evolution and persisted through negative selection of karyotype diversity, in spite of ongoing CIN through therapy administration and progression to metastasis[41].

There are also examples of CIN generating functional events late in tumour evolution. For example, PCAWG demonstrated that chromosomal gains occur throughout ‘molecular time’, from early human development to the final stages of tumour growth (median molecular time 0.60; IQR 0.10-0.87). However, within certain cancer types such as lung cancer, papillary renal cancer and melanoma, they are predominantly late events[29]. Furthermore, in a cohort of clear cell renal cell carcinoma (ccRCC), loss of 9p21.3 was subject to subclonal selection and was associated with metastatic progression, as evidenced by the fact that it was a clonal event in just 26% of primary tumour samples versus 64% of metastases [42]. Late arm-level SCNAs have also been described in other cancer types[43–45].

Watkins et al. interrogated pan-cancer multi-region data to assess the degree to which CIN provides a substrate for subclonal phenotypic diversification[46]. This work found evidence of parallel evolutionary events, in which the same genes were affected by different subclonal SCNAs in 37% of tumours analysed; examples included gains at 1q21.3-q44 which encompasses *BCL9, MCL1* and *ARNT*, at 5p15.33 which contains *TERT,* and at 8q24.1 which contains *MYC.* Independent analysis of metastases by multi-region sampling revealed subclonal events within certain tumour types, including gains to *MYC* in ccRCC and *CCND1* in *HER2*$^{+}$ breast cancer, which were enriched in metastases relative to primary tumours. Together, this suggests that late in tumour evolution, CIN engenders extensive subclonal diversity. It also expands on *in vitro* work within patient-derived tumour organoids, where ongoing CIN led to karyotypic heterogeneity over time in models of colorectal carcinoma[47].

It is likely that, even when looking at the mutation and copy number landscapes in parallel, we still fail to capture events with a large influence on cellular fitness and may draw imperfect conclusions as a result. Extending our focus beyond the genome to the epigenome, transcriptome, and the immune microenvironment may allow us a greater understanding of the true extent of ITH. In this way, we can more accurately define the functional evolutionary processes that influence phenotype (Figure 5).

[H1]Epigenetic and transcriptomic ITH

Epigenetic dysregulation influences gene expression and is widespread in cancer, occurring by varied mechanisms such as promoter hypermethylation, altered enhancer activity, and changes in chromatin configuration[48].

Alterations to the cancer epigenome can be binary, functioning as ‘on or off’ switches, or induce transient changes in gene expression, forming part of highly plastic gene expression networks. They but can also govern copy number changes. For example, the interplay between lysine demethylases and methyltransferases and their relative activity upon histone H3 lysine 4 (H3K4), H3K9 and H3K27 has been shown to affect copy-number amplification of the *EGFR* oncogene, which encodes the epidermal growth factor receptor[49]. Dysregulation of the cancer epigenome may be global in some cancers. In a pan-cancer analysis, increased global enhancer expression was seen across multiple cancer types, and enhancer activity correlated with the fraction of the genome affected by SCNAs[50]. This intriguing potential consequence of CIN was posited to relate to its impact on chromatin

291 state, although the evolutionary context of such epigenomic dysfunction within tumours
292 requires further elucidation. Interestingly, mutations in genes encoding epigenetic modifiers
293 alone are sufficient to drive some, and an analysis of 24 types of childhood cancer found this
294 group of genes, including *KMT2C, KMT2D* and *SMARCA4,* to be the most commonly
295 mutated (mutations in epigenetic modifiers affected 25% of tumours) across all molecular
296 subtypes[51].
297 Like the epigenome, the cellular transcriptome is frequently aberrant in cancer.
298 Various mechanisms, such as alternate splicing, alternate promoter usage, gene fusions and
299 aberrant oncogenic signalling can underpin this phenomenon[52–56]. The impact of genetic
300 alterations on transcription was explored extensively using bulk whole-genome sequencing
301 with matched RNA-sequencing in the recent PCAWG analysis[55]. In this analysis, SCNAs
302 emerged as the dominant genomic event influencing gene expression, contributing to 17% of
303 gene expression variation, compared to somatic and germline genetic variation in *cis,* which
304 contributed to 1.8% and 1.3% respectively. Cumulatively, non-coding mutations contributed
305 more to variation in allelic expression than coding mutations[55]. This underscores the potential
306 problems of restricting focus to exonic mutations, and is supported by our own analysis of
307 lists of cancer genes (Figure 4). We compared cancer genes identified by Bailey et al[57], who
308 systematically catalogued mutations from 9,423 tumour exomes, with those identified in the
309 COSMIC Cancer Gene Census[58], which incorporates evidence of functional involvement
310 alongside increased mutation frequency when curating cancer gene lists. As would be
311 expected, a whole-exome sequencing approach only identifies a subset (330/657) of
312 COSMIC cancer genes. Intriguingly, it fails to identify the majority of genes affected by
313 translocations and amplifications (Figure 4A). Greater proportions of the COSMIC cancer
314 genes affecting certain cancer types (BRCA, LGG) are identified by Bailey than others
315 (PAAD, SARC) (Figure 4B). This highlights that many cancer genes are likely undiscovered,
316 a substantial proportion of which may be driven by mechansisms beyond point mutations.
317 For example, RNA variants, generated by editing enzymes, are an additional source of
318 diversity within tumours that impact upon protein function in cancer and would be missed by
319 whole-exome sequencing[59,60]
320 The importance of transcriptomic variation in cancer is underlined by the ability of
321 expression-based biomarkers to predict clinical outcome[61]. However, the evolutionary
322 context of this variation is also important: considering ITH in gene expression through multi-
323 region bulk RNA-sequencing can significantly improve the predictive ability of such
324 biomarkers. For example, in non-small cell lung cancer (NSCLC), a prognostic gene
325 expression signature calculated from clonally expressed genes reduced the impact of
326 sampling bias, a problem also highlighted by a transcriptomic analysis of multifocal prostate
327 cancer[62,63]. Multi-region transcriptomics can also shed light on other tumour evolutionary
328 processes. Biswas *et al* underlined the dominant role of SCNAs in influencing gene
329 expression, as ITH of SCNAs was strongly correlated to ITH of gene expression[62]. A multi-
330 region transcriptomic analysis of four patients with advanced bladder cancer identified
331 distinct molecular subtypes thought to arise from different urothelial progenitors in distinct
332 regions of the same tumour, suggesting that tumour subtypes may be somewhat plastic rather
333 than entirely pre-determined[64].
334 An important caveat for bulk transcriptomic and epigenetic analysis is that such
335 datasets comprise tumour and stromal gene expression which, currently, we can only partly
336 deconvolve; tumour purity may therefore introduce bias into transcriptomic analysis of
337 ITH[65]. Integrated analysis of the genome and transcriptome may help to deconvolve tumour-
338 specific gene expression within bulk sequencing samples[66].
339 Single-cell sequencing can capture epigenomic and transcriptomic changes and has
340 helped to shape the concept of cellular transcriptomic promiscuity and its influence on

341 phenotypic plasticity[67]. In lung adenocarcinoma, a highly plastic cellular state is associated
342 with poor prognosis in humans and with treatment resistance in mice; this plastic state
343 mediates transition to more diverse phenotypes and may explain ITH in some tumours[68].
344 Jacks and colleagues studied the epigenome of single lung adenocarcinoma cells en route to
345 metastasis and revealed important changes in chromatin state which characterise gradual loss
346 of cellular identity, and may be controlled by key transcription factors such as those of the
347 RUNX family[69]. The identification of a gene expression signature of this subpopulation that
348 associated with survival was particularly intriguing[69]; the ability to robustly obtain from bulk
349 sequencing evidence of cancer cells with features of stemness, diverse transcriptional
350 landscapes and the ability to influence the evolutionary trajectory of a tumour in the presence
351 of selective pressures, would be of great value in both the research and clinical settings.
352 In studies of haematopoiesis, differences in stemness, cellular states, gene expression
353 profiles and enhancer activity have been explicitly linked to mutations in DNA methylation
354 genes, underscoring the importance of these genes in cancer[70]. However, in general, the
355 heterogeneity of epigenetic events in human cancers, as well as their interplay with the cancer
356 genome and transcriptome, remains poorly understood. This problem is illustrated by debates
357 surrounding evolutionary trajectories in pancreatic cancer, where genetic driver mutations in
358 *TP53, KRAS, CDKN2A* or *SMAD4,* when present, are almost always clonal[71]. In isolation,
359 this might indicate an absence of subclonal selection in this disease, but study of the
360 epigenome in pancreatic cancer evolution suggests that widespread chromatin remodelling
361 might provide the substrate for selection in metastasis[72]. In chronic lymphocytic leukaemia,
362 the epigenetic landscape is significantly disrupted, driving variety in cellular phenotype, and
363 different cancer cell populations may have highly disparate epigenomes[73].
364 Hua *et al.* explicitly compared the ITH of point mutations, SCNAs and DNA
365 methylation in a multi-region study of 84 lung adenocarcinomas[74]. They found that tumour
366 evolutionary trees inferred from SCNAs and DNA methylation were highly similar,
367 demonstrating that patterns of cancer evolution may be agnostic of variant mechanism.
368 Congruency of genomic and epigenomic evolution was also recently described in papillary
369 renal cell carcinoma[75]. Future work should seek to devise tools to define more clearly the
370 relationship between mutations, CIN and epigenetic and transcriptomic states, in both space
371 and time. Integrating whole-genome sequencing with multi-omics is likely to be important to
372 this endeavour.
373 Ultimately, mutations, SCNAs, epigenetic alterations and transcriptional alterations
374 all influence the abundance, structure and function of proteins, the true arbiters of cellular
375 phenotype. Therefore, proteomic studies may be critical to integrating this information.
376 Moreover, a proteogenomics approach facilitates deep analysis of the impact of functional
377 mutations at the pathway level. Such studies have described novel consequences of mutations
378 and SCNAs in breast cancer[76], including the identification of alterations in enzymatic activity
379 which were not visible at the transcriptomic level, as well as in gastric and ovarian
380 cancer[77,78]. In addition, recent publications studying the lung adenocarcinoma proteome with
381 mass spectrometry and phosphoproteomics are an important development in this space[79–81].
382 Proteogenomic analyses have also helped show that the relationship between SCNAs and
383 mRNA and protein abundance may be inconsistent. Intriguingly, in breast cancer, genes in
384 which a correlation was observed between SCNAs and mRNA as well as between SCNAs
385 and protein levels were more likely to be cancer genes than those without a correlation
386 between SCNAs and protein levels[76]. This suggests that in some settings multi-omic
387 approaches that consider the proteome may be more powerful to detect functional events that
388 those that do not. It is hoped that future studies of proteomic ITH, as well as other systems-
389 based approaches, will provide insight into functional clonal diversity, and help to resolve
390 long standing debates about cancer biology.

391
392 **[H1]Microenvironmental ITH**
393
394 A convincing argument to move forward from the reductionist, exclusively genomic,
395 view of tumour evolution stems from the success of immunotherapies that disrupt signalling
396 between cancer and immune cells. If our aim is to understand cancer evolution, we cannot
397 ignore the environment in which a cancer cell evolves. Whether variation is functional or
398 non-functional can be highly context dependent, depending on both the genomic background
399 into which a mutation arises, but also the environment itself. This environmental context has
400 an immune, and a non-immune arm.
401
402 **[H2]** Evolutionary constraints imposed by the anti-tumour immune response
403
404 A critical component of the tumour microenvironment in a growing cancer is the
405 immune infiltrate. Human cellular biology has evolved to combat the threat of cancer in its
406 tissues through a high-fidelity anti-cancer immune response[82]. A prominent part of this
407 response is the adaptive immune response. Here, a tumour, or its associated non-host peptides
408 (neoantigens) are identified and eliminated through the clonal expansion of a highly specific
409 T-cell population and degradation by cytotoxic CD8+ T-cells, or by a CD4+ T-cell-
410 dependent cytotoxic response[83,8485].
411 The anti-tumour immune response equates to a clone-specific negative selection
412 pressure and there are a number of mechanisms that cancers can co-opt in order to evade
413 detection and elimination. Many of these are genomic. For example, mutations in beta-2-
414 microglobulin, *B2M*, a component of the Major Histocompatibility Complex, disrupt antigen
415 presentation in response to immune predation[86]. Similarly, cancer cells undergo loss-of-
416 heterozygosity (LOH) of the Human Leucocyte Antigen (HLA) locus on chromosome 6p. In
417 one study, HLA LOH was detected in 40% of NSCLC and was subclonal in 65% of cases[87].
418 This finding was supported by a recent pan-cancer analysis of multi-region sampled tumours,
419 where 22% of all tumours demonstrated subclonal loss of two copies of the same HLA allele
420 after whole genome doubling[46]. SCNAs outside of the HLA locus can also promote immune
421 evasion, for example through copy-number loss of neoantigens capable of stimulating a
422 functional T cell response[88,89]. In mouse models of ovarian cancer, immune-excluded tumour
423 regions were characterised by copy-number amplification of MYC target genes and increased
424 WNT signalling[90]. An abundance of non-specific SCNAs has also been proposed as a
425 predictor of poor response to immunotherapy[91]; however, disentangling their well-described
426 prognostic role from a predictive one is difficult, and requires further study. There is also
427 evidence of transcriptional neoantigen depletion underpinning immune escape. In NSCLC,
428 downregulation of neoantigenic transcripts was found to occur via promoter
429 hypermethylation (seen in 23% of silenced neoantigen-containing genes versus 11% of the
430 same non neoantigen-containing genes) and was enriched in immune-infiltrated tumours with
431 an intact HLA allele, suggestive of diverse cellular responses to the negative selection
432 pressure imposed by the anti-tumour immune response[89]. Indeed, the remaining repressed
433 neoantigens not subject to promoter hypermethylation may be affected by mechanisms that
434 are yet to be elucidated.
435 Not all neoantigens stimulate a uniform anti-tumour immune response, demonstrating
436 the importance of distinguishing between functional and non-functional ITH. Clonal
437 neoantigens stimulate anti-tumour immunity, and those tumours containing neoantigen-
438 reactive tumour-infiltrating lymphocytes (TILs) are associated with better outcome[92]. In a
439 multi-region study of NSCLC, Chain and colleagues reported a correlation between the
440 number of T-cell clones found in all tumour regions, and the number of clonal, but not

441 subclonal mutations, emphasising the importance of neoantigens in stimulating the immune
442 response early in tumour evolution[93]. Furthermore, the clonality of T-cells within a tumour
443 has also been associated with improved response to anti-programmed cell death 1 (PD1)
444 therapy in melanoma[94,95], although other facets of the T-cell repertoire, such as the diversity
445 of circulating T-cell clones in peripheral blood, also impact response to immune checkpoint
446 blockade[96]. Importantly, the clonal diversity of neoantigens can also influence the anti-
447 tumour immune response. Using an immune-competent mouse model of melanoma, Wolf
448 and colleagues showed a correlation between increased clonal diversity and ineffective
449 rejection of the developing tumour[97]. This builds on earlier findings in mice which suggested
450 that the fraction of cells expressing a clonal, immunogenic peptide is key in determining
451 whether a tumour is eliminated, with small subclones being more able to evade immune
452 rejection[98].

453 As clonal diversity can influence the immune response to a developing tumour, so this
454 response can provide a negative selection pressure on a growing tumour and in turn shape its
455 clonal composition. This is illustrated by recent work in a glioma model, where the immune
456 editing that occurred within immune-competent mice led to the formation of tumours with
457 lower clonality[99]. In high-grade serous ovarian cancer, tumour regions with the highest levels
458 of immune infiltrate were characterised by neoantigen depletion, subclonal HLA LOH, and
459 low clone diversity, indicating predation of that region prior to sampling[100]. These findings
460 are consistent with a study of metastatic colorectal cancer, where those metastases that persist
461 were the least immunogenic, and harboured diverse mechanisms of immune escape[101].
462 Negative selection has also been reported prior to formation of an invasive cancer:
463 histopathological and molecular analysis of pre-invasive lung lesions suggested that immune
464 surveillance is more active in those that regress, relative to those that progress to invasive
465 cancer[102].

466 The extent to which negative selection pressure is exerted on the developing tumour
467 throughout its evolution is unclear: studies have suggested that neoantigen-encoding
468 mutations may be depleted within primary tumours, indicating the impact of negative
469 selection prior to sampling[86,91,103]. However, in a recent analysis which adjusted for single
470 nucleotide substitution mutational signatures, no evidence of negative selection against
471 neoantigens was found, with the exception of NSCLC[104]. A recent preprint reported an
472 orthogonal approach, restricting dN/dS to the immunopeptidome, that detected immune
473 selection in some tumours[105]. Intriguingly, pre-treatment immune selection was detected at
474 increased levels within a subset of patients with metastatic disease who responded poorly to
475 immune checkpoint inhibitor therapy[105].

476 Just as nonsynonymous mutations in a tumour may be immunologically functional or
477 non-functional depending on their ability to elicit a T-cell response, so the functionality and
478 differentiation state of T-cells may also vary between tumours. Work on NSCLC found that
479 dysfunctional and terminally differentiated T-cells expressing PD1 and inducible T-cell
480 costimulatory (ICOS) were more predominant in tumours with a high mutational burden,
481 whilst progenitor-like T-cells in the early stages of differentiation expressing CD27 and
482 CD28, and lacking signs of antigen engagement, were seen in tumours with a low mutational
483 burden[106]. This is clinically significant as a gene signature of differentiation skewing was
484 associated with a worse prognosis across multiple tumour types without immunotherapy[106].

485 A structured immune microenvironment, manifest as spatial differences between
486 areas of the same tumour, can be important to both the trajectory of a tumour and patient
487 outcome. In triple-negative breast cancer, highly multiplexed imaging enabled classification
488 of tumours by their extent of tumour-immune mixing, as either ‘cold’, ‘mixed’ or
489 ‘compartmentalized’[107]. Intriguingly, this feature correlated with expression of the
490 immunotherapy targets indoleamine 2,3-dioxygenase 1 (IDO1) and PD1 ligand 1 (PDL1) on

tumour or non-tumour cells: in the compartmentalized tumours, IDO-1 and PDL1 were expressed by non-tumour cells, such as monocytes, and these tumours were associated with improved prognosis[107]. Within lung adenocarcinoma, the presence of more than one 'immune-cold' region in which immune evasion might have occurred, predicted poor patient outcome irrespective of the immune phenotypes observed in the rest of the tumour[108]. Immune-cold regions from the same tumour were more likely to share subclonal mutations than immune-hot regions, raising the possibility that 'functional' events, which mostly remain to be elucidated, underpin the ability of an evolving tumour to evade immune rejection.

Taken together, these studies demonstrate that tumours, or parts of tumours, may have different relationships with the anti-tumour immune response and can behave differently, thereby highlighting the importance of heterogeneity in the microenvironment of a tumour towards shaping its evolution.

The non-immune microenvironment and tumour evolution

Cancers exploit local signalling networks that can profoundly impact cellular identity, to co-opt stromal cells and foster a microenvironment that is favourable to tumour growth via the angiogenic switch, epithelial-to-mesenchymal transition and other processes (reviewed elsewhere[109]). This creates regional spatial differences between parts of tumours, which may be observed by analysis of pH gradients, hypoxia and growth factor concentration[110]. Moreover, ITH may be driven in part by microenvironmental factors: this is supported by modelling data[111] and implied in a study of ccRCC, where regional variations observed on imaging could not be accounted for by genetic variation[112]. Interactions between cancer cells and their stromal counterparts can profoundly influence the trajectory of a tumour. p53-dependent senescence in hepatic stellate cells may act in a non-cell-autonomous manner to promote macrophage differentiation and an anti-tumourigenic microenvironment[113]. In contrast, cancer cells may co-opt the systemic environment, outside the confines of the tumour, such as in the formation of the pre-metastatic niche through tumour-secreted factors and tumour-shed vesicles[114]. Incoming cancer cells can then interact with this environment. For example, a co-culture system demonstrated that, in the early stages of lung metastasis, interaction between alveolar epithelial cells and disseminated breast cancer cells has been shown to influence behaviour of metastatic cells, enabling them to remain indolent and survive for long periods of time[115].

Distinct clones within a tumour may also interact; clonal cooperativity within a cancer has been described in mouse models of breast cancer[116]. Furthermore, in a study of human colorectal cancer tissue, Schurch and colleagues used co-detection by indexing (CODEX) imaging of formalin-fixed, paraffin-embedded tissues to profile the 'cellular neighbourhoods', and found that features of their relationships, such as coupling of the immune and tumour neighbourhoods and disruption of inter-neighbourhood communication, correlated with poor prognosis[117].

Understanding the extent of cooperation and 'task sharing' between cancer clones is an important area of further study. Highly multiplexed imaging modalities such as mass cytometry are increasingly facilitating the accurate phenotyping of diverse cell types obtained from tumour samples. This has given greater granularity to early observations of clonal cooperation, and relationships between cells can be described in detail[118–120]. Multi-omic studies that can be mapped spatially to the tumour will help to further define functional examples of tumour heterogeneity.

Clinical impact of ITH

ITH and assessments of tumour evolutionary trajectories, as well as the interplay between the cancer cell and the immune microenvironment are all important in the context of clinical questions. Will a tumour metastasise? Will it respond to therapy, and how durable will any response be? What is the prognosis of this cancer? In order to answer such questions, we must have as clear a picture as possible of the phenotype of the tumour: for this, we must comprehend the evolutionary forces that have fostered it.

Metastasis

Understanding functional ITH may provide insights into the process of metastasis. The impact of cancer evolutionary dynamics on metastasis is an active area of research and our current understanding is reviewed elsewhere[121].

Unpaired analysis of genomes from primary and metastatic tumours suggests that metastasis-specific genomically-encoded driver mutations are rare[122]. Nonetheless, the abundance of certain driver events in metastases may exceed that within the primary tumour, such as the enrichment for loss of 9p (containing the tumour suppressor *CDKN2A*), in metastatic ccRCC[42], or mutations in *MLK4* (also known as *MAP3K21*) pan-cancer[122]. An enrichment within metastatic cancers for mutations of epigenetic regulators has also been characterised, adding weight to the notion that transcriptional promiscuity and phenotypic diversity may be a prerequisite for spread in some settings[45,72,123]. Importantly, transcriptional promiscuity need not be genomically underpinned, emphasizing the need for analysis beyond the metastatic cancer genome.

As ITH of primary tumours provides insight into their evolutionary history, so analysis of intra-tumour, and inter-tumour heterogeneity in metastases can shape our understanding of their biology. For example, in using genomics to attempt to understand the timing of metastatic dissemination, an appreciation of genomic heterogeneity within both primary tumours and metastases is essential. Under-sampling of a primary tumour may exaggerate the degree of genetic divergence between primary and metastatic lesions and hence convey metastatic divergence occurring earlier in evolutionary time, potentially contributing to a lack of consensus in the field. In a cohort of 17 patients with disseminated breast cancer, metastatic divergence was estimated to occur relatively late, at 87% of molecular time[123]. Conversely, a recent analysis by Curtis and colleagues, using mathematical modelling, estimated metastatic seeding to occur 2-4 years before diagnosis in colorectal, breast and lung cancers[124]. This study also highlighted a role for systemic anti-cancer treatment in promoting clonal evolution and thus influencing ITH at relapse: 57% of treated metastases showed private driver events, compared to 20% of those that were untreated. This work also serves to highlight the fact that an as-yet-undefined degree of the genetic diversity demonstrated in other studies of post-therapy tumour metastases may reflect ongoing evolution in response to treatment.

The problem with sampling also extends to questions regarding the mode of dissemination to the metastatic site. To robustly assess this, multi-region sampling of both primary and metastatic lesions is required, and there are few studies that have done this to date[121]. Nonetheless, many studies have attempted to quantify the relative contributions of monoclonal (a single subclone seeds every metastasis) and polyclonal (multiple subclones seed one or more metastases) dissemination in cancer. Of note, Reiter and colleagues performed an analysis of 317 regions from 20 patients across lymph node and distant metastatic sites, to investigate the relative preponderance of metastatic seeding patterns[125]. They identified a higher genetic diversity among lymph node metastases than distant metastases, suggestive of polyclonal seeding patterns and more relaxed selection pressures in

lymph node seeding clones relative to distant sites. The potential pitfalls of inferring seeding patterns from bulk sequencing is underlined by a study which presents a tool that combines inference of clonal lineage with migration histories[126]. By using this tool to re-analyse published sequencing data, Raphael and colleagues showed how seeding patterns might often be misclassified; for example, in a study that had originally suggested a polyclonal seeding pattern, it was posited that monoclonal seeding was in fact a more parsimonious explanation for the data[126].

Whilst multi-region sampling of a resected solid primary tumour is relatively straightforward, ethical constraints prevent simultaneous sampling of multiple metastases from patients. Two strategies that may help to combat this problem and provide insights into the evolutionary history and diversity of cells within the same or different metastases are autopsy studies and circulating tumour DNA (ctDNA) analyses.

Autopsy studies allow for sampling of multiple metastases from the same patient. One such study of 76 untreated metastases from 20 patients suggested that driver mutations in this setting are typically found within all metastases and have likely occurred before metastatic dissemination[127]. Another analysis underlined the potential limitations of inferring phylogeny from a single metastatic sample[128]; this revealed pervasive branched evolution in 6 out of 7 melanoma patients and suggested that metastases in different sites may have entirely different clonal histories, as well as different active mutational processes, than the primary tumour. This is an active area of research and larger autopsy studies, such as the Posthumous Evaluation of Advanced Cancer Environment (PEACE; NCT03004755)[129], may provide greater insights in the future.

Studies measuring ctDNA can help to build a picture of functional tumour heterogeneity. In a cohort of 42 patients with gastrointestinal cancers, ctDNA was demonstrated to be superior to a single metastatic biopsy at capturing mechanisms of resistance to targeted therapy in the majority of patients[130]. In one patient where extensive sampling was undertaken following autopsy, parallel mechanisms of resistance had evolved across different metastatic sites. This diversity could be captured in ctDNA sampling but would have been missed in any single metastatic biopsy. ctDNA collection can also help to determine the pattern and timing of metastatic spread. In a study of oesophageal adenocarcinoma, extensive sampling, including of the primary tumour, blood plasma and metastatic sites at autopsy revealed 'clonal diaspora' as the predominant mode of spread, in which multiple subclones rapidly seeded multiple metastatic sites[131]. In studies seeking to track pre-defined genetic events over time, targeted panel sequencing of ctDNA represents a promising avenue of research.

Immune editing is known to shape the evolution of metastasis in colorectal cancer[101], and analysis of a mouse model of breast cancer metastasis by Lo and colleagues suggests the immune-microenvironment can influence the modality of dissemination[132]. In this study, dissemination within mice lacking natural killer (NK) cell immunity was more likely to be monoclonal, and cells that spread in clusters had lower expression of NK-activating genes, and increased expression of NK-inhibitory genes. Defining functional ITH in the context of metastasis will be a multi-faceted endeavour that will help to answer outstanding questions about its biology

[H2]Informing prognosis

The balance and degree of functional and non-functional ITH within a tumour may have important prognostic implications. ITH of driver mutations, which confer a significant fitness impact upon the cancer cell, could be considered a proxy for 'functional' subclonal diversity. Several studies have described a relationship between clonal diversity of driver

mutations in cancer and poor outcome[133–138]. In pre-invasive lesions such as Barrett's esophagus, most measures of clonal genetic diversity have been found to correlate with early progression to invasive disease[139]. However, the relationship between ITH and outcome is nuanced. Early acquisition and clonal sweeps of driver events, resulting in a relatively homogeneous tumour, are also associated with progression to metastasis and poor outcome[23], as seen in ccRCC[42].

CIN is also a source of functional and non-functional diversity in some tumours, and gene expression correlates of CIN have long been associated with poor outcome across multiple cancer types[61]. In NSCLC, subclonal diversity of SCNAs was prospectively correlated with poor prognosis, whilst copy number aberrations also conferred poor survival in ccRCC[45,136]. Conversely, in some settings extreme diversity may be associated with improved prognosis. For example, in some malignancies, a high mutational burden may be associated with improved patient outcomes, likely owing to increased immune surveillance[137,140]. Similarly, patients whose tumours have extreme levels of CIN tend to have better outcomes[134,141,142]. The fundamental role of CIN in tumourigenesis is underlined by the relative scarcity of SCNAs, when compared to somatic mutations, within normal human tissue[143,144]. Clonal expansion of mutant cells seems seldom to lead to cancer in the absence of SCNAs. A recent study of SCNAs within pre-invasive lesions also highlights their importance: shallow whole-genome sequencing in a longitudinal study of 88 patients with Barrett's esophagus demonstrated that copy-number profiles of lesions were able to predict patients' risk of progression[145].

A possible consensus may emerge: ITH of driver events, in the context of tolerable amounts of CIN, is associated with poor outcomes in patients (Figure 6). This may result from ongoing subclonal selection of functional variation. However, in the presence of extreme CIN, ITH does not predict poor outcome. This paradox may relate to an inability of cells in such tumours to maintain a high-level fitness, or potentially to such tumours being exceptionally susceptible to anti-cancer therapy that may make retaining fitness impossible. A recent evolutionary model of the prognostic impact of ITH supports the view that clonal diversity, in the presence of genomic instability, is associated with faster tumour growth and reduced survival for patients[146].

[H2]Therapy response and resistance

Understanding the clonal architecture of a cancer can be crucial for effective treatment. For example, clonal, rather than subclonal, neoantigens, appear to stimulate an effective immune response, and their abundance has been associated with survival in multiple cohorts treated with immune checkpoint inhibitors[92,147]. Also, in gastric cancers treated with an experimental fibroblast growth factor receptor (FGFR) inhibitor, response was conditional on a clonal *FGFR2* amplification, and patients with a subclonal amplification failed to respond[148].

Treatment resistance, either *de-novo*[149] or pre-existing, is unfortunately a near-ubiquitous feature of cancer treatment. This may arise through varied mechanisms: for example, in NSCLC, patients may develop resistance to first-generation EGFR inhibitors via the *EGFR*^{T790M} point mutation[150]; in glioma, resistance to temozolamide can arise through a fusion structural variant in *MGMT*[151]; and in acute myeloid leukaemia (AML), transcriptional plasticity may drive resistance to bromodomain and extraterminal (BET) inhibitor therapy[152]. An understanding of resistance mechanisms may reveal opportunities for further treatment; examples include osimertinib, which binds irreversibly to mutated EGFR-T790M, for treatment of NSCLC, or targeting of enhancer switching with lysine demethylase 1A (KDM1A) inhibition for AML treatment[150,152].

Resistance is rarely attributable to a single event; indeed, extensive CIN, promiscuous transcriptional signalling and epigenetic plasticity, all conferring unstable cellular phenotypes, may each fuel non-genetic resistance and confound multiple lines of treatment[153–157]. In such circumstances, targeted therapies may have limited value. Treatments that co-opt the anti-cancer immune response such as immune checkpoint blockade, adoptive T-cell therapy or vaccine therapy, however, might be less vulnerable to such resistance. Nevertheless, treatment resistance remains a significant problem[158,159] and approaches to delay, or even prevent this are urgently required.

Adaptive, evolutionary-aware strategies may hold promise in improving outcomes for patients. Resistant cells can have a relative fitness disadvantage compared to their sensitive neighbours in the absence of treatment[160–162], and in such a scenario, temporarily withholding treatment could cause a net growth in the sensitive population and a decline in the resistant population. Non-destructive modelling, as demonstrated in NSCLC whereby dead cells are collected from culture without disrupting the live population, suggests this approach is effective in controlling the clonal composition of tumours over time[163]. However, preliminary results from certain clinical trials of adaptive treatments have been disappointing[164], highlighting the need for further work and a deeper understanding of the fitness costs of resistance and how this can be measured over time.

[H1] Conclusions and Perspectives

Our understanding of cancer evolution has increased exponentially in the last decade. The advent of next-generation sequencing has shed light on extensive genomic heterogeneity within cancers and given insight into the evolutionary pressures governing tumour growth.

However, it is necessary to acknowledge that further advances in this field will require not only more extensive sampling of the cancer genome across space and time, but also a more in-depth exploration of the cancer cell and its environment, moving beyond the cancer genome. Recent work, analysing the impact of structural variation, epigenomic and transcriptomic dysregulation and the immune microenvironment on cancer evolution have highlighted extensive functional variation within tumours, with profound impacts on phenotype.

Genetic and non-genetic divergence is a feature of every tumour, in part simply reflecting the random acquisition of mutations during cell division. Thus, rather than simply cataloguing diversity, future work must distinguish between functional and non-functional ITH, identifying events that might be subject to negative or positive selection during tumour evolution. Indeed, this is a fundamental issue for cancer research: notwithstanding the possibility of cure by complete surgical resection of all cancer cells present in a patient, the genetic and non-genetic trajectories of the cells within a tumour have a profound impact on disease prognosis. A thorough understanding of the relative weights exerted by various biological pulleys and levers during this process might enable us to fine-tune anti-cancer treatments and more effectively control the evolutionary fate of cancer cells.

ToC blurb

This review discusses the role of functional (impacting tumour phenotype) and non-functional intra-tumour heterogeneity (ITH) in cancer evolution. It highlights the importance of considering genetic and non-genetic factors and their impact on patient outcomes.

741 **Acknowledgements**
742 The authors would like to thank Chris Bailey, Alex Frankell, Emma Colliver, Carlos-
743 Martinez Ruiz and Jonas Demeulemeester for their thoughtful review of the manuscript.
744 Figures created with Biorender.com.
745
746 **Conflict of Interest**
747 The authors declare no conflicts of interest.
748 **Competing interests**
749 The authors declare no competing interests.
750
751 **Author contributions**
752 JRMB and NM both researched data for the article and made a substantial contribution to discussion of content, writing, reviewing and
753 editing the article.
754
755 **Peer review information**
756 *Nature Reviews Cancer* thanks Moritz Gerstung and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.
757
758 **Glossary:**
759 **Chromosomal instability (CIN): A defect in which cells can gain, lose or rearrange parts**
760 **of or whole chromosomes during cell division; this is a source of variation in cancer.**
761 **Chromothripsis: A mutational process in which large numbers of clustered structural**
762 **rearrangements occur in single or multiple chromosomes.**
763 **Molecular time: An estimate of the timing of an event, from the first cell division**
764 **following fertilisation to a cell division that occurred only recently before sampling.**
765 **Enhancer: A short genomic region that influences the expression of another gene in *cis*.**
766
767
768 **References**
769
770 1. Darwin, C. *The origin of species by means of natural selection: or, the preservation of*
771 *favored races in the struggle for life*. (AL Burt New York, 1859).
772 2. Nowell, P. C. The clonal evolution of tumor cell populations. *Science* **194**, 23–28
773 (1976).
774 3. Greaves, M. & Maley, C. C. Clonal evolution in cancer. *Nature* **481**, 306–313 (2012).
775 4. Williams, M. J., Werner, B., Graham, T. A. & Sottoriva, A. Functional versus non-
776 functional intratumor heterogeneity in cancer. *Mol. Cell. Oncol.* **3**, e1162897 (2016).
777 5. Turajlic, S., Sottoriva, A., Graham, T. & Swanton, C. Resolving genetic heterogeneity
778 in cancer. *Nat. Rev. Genet.* **20**, 404–416 (2019).
779 6. Marusyk, A., Janiszewska, M. & Polyak, K. Intratumor Heterogeneity: The Rosetta
780 Stone of Therapy Resistance. *Cancer Cell* **37**, 471–484 (2020).
781 7. McGranahan, N., Burrell, R. A., Endesfelder, D., Novelli, M. R. & Swanton, C.
782 Cancer chromosomal instability: therapeutic and diagnostic challenges. *EMBO Rep.*
783 **13**, 528–538 (2012).
784 8. Alexandrov, L. B. *et al.* Clock-like mutational processes in human somatic cells. *Nat.*
785 *Genet.* **47**, 1402–1407 (2015).
786 9. Coons, S. W., Johnson, P. C. & Shapiro, J. R. Cytogenetic and Flow Cytometry DNA
787 Analysis of Regional Heterogeneity in a Low Grade Human Glioma. *Cancer Res.* **55**,
788 1569 LP – 1577 (1995).
789 10. Teixeira, M. R., Pandis, N., Bardi, G., Andersen, J. A. & Heim, S. Karyotypic
790 Comparisons of Multiple Tumorous and Macroscopically Normal Surrounding Tissue
791 Samples from Patients with Breast Cancer. *Cancer Res.* **56**, 855 LP – 859 (1996).
792 11. Navin, N. *et al.* Inferring tumor progression from genomic heterogeneity. *Genome Res.*

793 **20**, 68–80 (2010).
794 12. Gerlinger, M. *et al.* Intratumor Heterogeneity and Branched Evolution Revealed by
795 Multiregion Sequencing. *N. Engl. J. Med.* **366**, 883–892 (2012).
796 13. OHTA, T. Slightly Deleterious Mutant Substitutions in Evolution. *Nature* **246**, 96–98
797 (1973).
798 14. Ling, S. *et al.* Extremely high genetic diversity in a single tumor points to prevalence
799 of non-Darwinian cell evolution. *Proc. Natl. Acad. Sci.* **112**, E6496 LP-E6505 (2015).
800 15. Sottoriva, A. *et al.* A Big Bang model of human colorectal tumor growth. *Nat. Genet.*
801 **47**, 209–216 (2015).
802 16. Williams, M. J., Werner, B., Barnes, C. P., Graham, T. A. & Sottoriva, A.
803 Identification of neutral tumor evolution across cancer types. *Nat. Genet.* **48**, 238–244
804 (2016).
805 17. Sun, R. *et al.* Between-region genetic divergence reflects the mode and tempo of
806 tumor evolution. *Nat. Genet.* **49**, 1015–1024 (2017).
807 18. Cross, W. *et al.* The evolutionary landscape of colorectal tumorigenesis. *Nat. Ecol.*
808 *Evol.* **2**, 1661–1672 (2018).
809 19. Baca, S. C. *et al.* Punctuated Evolution of Prostate Cancer Genomes. *Cell* **153**, 666–
810 677 (2013).
811 20. Gao, R. *et al.* Punctuated copy number evolution and clonal stasis in triple-negative
812 breast cancer. *Nat. Genet.* **48**, 1119–1130 (2016).
813 21. Notta, F. *et al.* A renewed model of pancreatic cancer evolution based on genomic
814 rearrangement patterns. *Nature* **538**, 378–382 (2016).
815 22. Field, M. G. *et al.* Punctuated evolution of canonical genomic aberrations in uveal
816 melanoma. *Nat. Commun.* **9**, 116 (2018).
817 23. Reiter, J. G. *et al.* An analysis of genetic heterogeneity in untreated cancers. *Nat. Rev.*
818 *Cancer* 639–650 (2019) doi:10.1038/s41568-019-0185-x.
819 24. Papaemmanuil, E. *et al.* Clinical and biological implications of driver mutations in
820 myelodysplastic syndromes. *Blood* **122**, 3616–27; quiz 3699 (2013).
821 25. Brastianos, P. K. *et al.* Genomic Characterization of Brain Metastases Reveals
822 Branched Evolution and Potential Therapeutic Targets. *Cancer Discov.* **5**, 1164–1177
823 (2015).
824 26. Landau, D. A. *et al.* Mutations driving CLL and their evolution in progression and
825 relapse. *Nature* **526**, 525–530 (2015).
826 27. Yates, L. R. *et al.* Subclonal diversification of primary breast cancer revealed by
827 multiregion sequencing. *Nat. Med.* **21**, 751–759 (2015).
828 28. Mitchell, T. J. *et al.* Timing the Landmark Events in the Evolution of Clear Cell Renal
829 Cell Cancer: TRACERx Renal. *Cell* **173**, 611-623.e17 (2018).
830 29. Gerstung, M. *et al.* The evolutionary history of 2 ,658 cancers. *Nature* **578**, (2020).
831 30. Nik-Zainal, S. *et al.* The life history of 21 breast cancers. *Cell* **149**, 994–1007 (2012).
832 31. Kessler, D. A. & Levine, H. Large population solution of the stochastic Luria–
833 Delbrück evolution model. *Proc. Natl. Acad. Sci.* **110**, 11682–11687 (2013).
834 32. Caravagna, G. *et al.* Subclonal reconstruction of tumors by using machine learning and
835 population genetics. *Nat. Genet.* **52**, 898–907 (2020).
836 33. Litchfield, K. *et al.* Representative Sequencing: Unbiased Sampling of Solid Tumor
837 Tissue. *Cell Rep.* **31**, 107550 (2020).
838 34. Durinck, S. *et al.* Temporal dissection of tumorigenesis in primary cancers. *Cancer*
839 *Discov.* **1**, 137–143 (2011).
840 35. Martincorena, I. *et al.* High burden and pervasive positive selection of somatic
841 mutations in normal human skin. *Science (80-.).* **348**, 880–886 (2015).
842 36. Martincorena, I. *et al.* Universal Patterns of Selection in Cancer and Somatic Tissues.

843 *Cell* **171**, 1029-1041.e21 (2017).
844 37. McGranahan, N. & Swanton, C. Clonal Heterogeneity and Tumor Evolution: Past,
845 Present, and the Future. *Cell* **168**, 613–628 (2017).
846 38. Tarabichi, M. *et al.* Neutral tumor evolution? *Nat. Genet.* **50**, 1630–1633 (2018).
847 39. Nam, A. S., Chaligne, R. & Landau, D. A. Integrating genetic and non-genetic
848 determinants of cancer evolution by single-cell multi-omics. *Nat. Rev. Genet.* (2020)
849 doi:10.1038/s41576-020-0265-5.
850 40. Laks, E. *et al.* Clonal Decomposition and DNA Replication States Defined by Scaled
851 Single-Cell Genome Sequencing. *Cell* **179**, 1207-1221.e22 (2019).
852 41. Cross, W. *et al.* Stabilising selection causes grossly altered but stable karyotypes in
853 metastatic colorectal cancer. *bioRxiv* 2020.03.26.007138 (2020)
854 doi:10.1101/2020.03.26.007138.
855 42. Turajlic, S. *et al.* Tracking Cancer Evolution Reveals Constrained Routes to
856 Metastases: TRACERx Renal. *Cell* **173**, 581-594.e12 (2018).
857 43. Gundem, G. *et al.* The evolutionary history of lethal metastatic prostate cancer. *Nature*
858 **520**, 353–357 (2015).
859 44. Suzuki, H. *et al.* Mutational landscape and clonal architecture in grade II and III
860 gliomas. *Nat. Genet.* **47**, 458–468 (2015).
861 45. Jamal-Hanjani, M. *et al.* Tracking the evolution of non-small-cell lung cancer. *N. Engl.*
862 *J. Med.* **376**, 2109–2121 (2017).
863 46. Watkins, T. B. K. *et al.* Pervasive chromosomal instability and karyotype order in
864 tumour evolution. *Nature* (2020) doi:10.1038/s41586-020-2698-6.
865 47. Bolhaqueiro, A. C. F. *et al.* Ongoing chromosomal instability and karyotype evolution
866 in human colorectal cancer organoids. *Nat. Genet.* **51**, 824–834 (2019).
867 48. Flavahan, W. A., Gaskell, E. & Bernstein, B. E. Epigenetic plasticity and the
868 hallmarks of cancer. *Science (80-.).* **357**, (2017).
869 49. Clarke, T. L. *et al.* Histone Lysine Methylation Dynamics Control EGFR DNA Copy-
870 Number Amplification. *Cancer Discov.* **10**, 306–325 (2020).
871 50. Chen, H. *et al.* A Pan-Cancer Analysis of Enhancer Expression in Nearly 9000 Patient
872 Samples. *Cell* **173**, 386-399.e12 (2018).
873 51. Gröbner, S. N. *et al.* The landscape of genomic alterations across childhood cancers.
874 *Nature* **555**, 321–327 (2018).
875 52. Kahles, A. *et al.* Comprehensive Analysis of Alternative Splicing Across Tumors from
876 8,705 Patients. *Cancer Cell* **34**, 211-224.e6 (2018).
877 53. Shiraishi, Y. *et al.* A comprehensive characterization of cis-acting splicing-associated
878 variants in human cancer. *Genome Res.* **28**, 1111–1125 (2018).
879 54. Demircioğlu, D. *et al.* A Pan-cancer Transcriptome Analysis Reveals Pervasive
880 Regulation through Alternative Promoters. *Cell* **178**, 1465-1477.e17 (2019).
881 55. Calabrese, C. *et al.* Genomic basis for RNA alterations in cancer. *Nature* **578**, 129–136
882 (2020).
883 56. Zhang, Y. *et al.* High-coverage whole-genome analysis of 1220 cancers reveals
884 hundreds of genes deregulated by rearrangement-mediated cis-regulatory alterations.
885 *Nat. Commun.* **11**, 736 (2020).
886 57. Bailey, M. H. *et al.* Comprehensive Characterization of Cancer Driver Genes and
887 Mutations. *Cell* **173**, 371-385.e18 (2018).
888 58. Sondka, Z. *et al.* The COSMIC Cancer Gene Census: describing genetic dysfunction
889 across all human cancers. *Nat. Rev. Cancer* **18**, 696–705 (2018).
890 59. Baysal, B. E., Sharma, S., Hashemikhabir, S. & Janga, S. C. RNA Editing in
891 Pathogenesis of Cancer. *Cancer Res.* **77**, 3733–3739 (2017).
892 60. Chen, L. *et al.* Recoding RNA editing of AZIN1 predisposes to hepatocellular

893 carcinoma. *Nat. Med.* **19**, 209–216 (2013).
894 61. Carter, S. L., Eklund, A. C., Kohane, I. S., Harris, L. N. & Szallasi, Z. A signature of
895 chromosomal instability inferred from gene expression profiles predicts clinical
896 outcome in multiple human cancers. *Nat. Genet.* **38**, 1043–1048 (2006).
897 62. Biswas, D. *et al.* A clonal expression biomarker associates with lung cancer mortality.
898 *Nat. Med.* **25**, 1540–1548 (2019).
899 63. Salami, S. S., Tomlins, S. A. & Palapattu, G. S. Transcriptomic Heterogeneity in
900 Multifocal Prostate Cancer. *JCI* **201**, 666–667 (2019).
901 64. Thomsen, M. B. H. *et al.* Comprehensive multiregional analysis of molecular
902 heterogeneity in bladder cancer. *Sci. Rep.* **7**, 11702 (2017).
903 65. Aran, D., Sirota, M. & Butte, A. J. Systematic pan-cancer analysis of tumour purity.
904 *Nat. Commun.* **6**, 8971 (2015).
905 66. Wang, F. *et al.* Integrated transcriptomic–genomic tool Texomer profiles cancer
906 tissues. *Nat. Methods* **16**, 401–404 (2019).
907 67. Teschendorff, A. E. & Enver, T. Single-cell entropy for accurate estimation of
908 differentiation potency from a cell's transcriptome. *Nat. Commun.* **8**, 15599 (2017).
909 68. Marjanovic, N. D. *et al.* Emergence of a High-Plasticity Cell State during Lung Cancer
910 Evolution. *Cancer Cell* (2020) doi:https://doi.org/10.1016/j.ccell.2020.06.012.
911 69. Lafave, L. M. *et al.* Epigenomic state transitions characterize tumor progression in
912 lung adenocarcinoma. *Cancer Cell* 212–228 (2020) doi:10.1016/j.ccell.2020.06.006.
913 70. Izzo, F. *et al.* DNA methylation disruption reshapes the hematopoietic differentiation
914 landscape. *Nat. Genet.* **52**, 378–387 (2020).
915 71. Raphael, B. J. Integrated Genomic Characterization of Pancreatic Ductal
916 Adenocarcinoma. *Cancer Cell* **32**, 185-203.e13 (2017).
917 72. McDonald, O. G. *et al.* Epigenomic reprogramming during pancreatic cancer
918 progression links anabolic glucose metabolism to distant metastasis. *Nat. Genet.* **49**,
919 367–376 (2017).
920 73. Pastore, A. *et al.* Corrupted coordination of epigenetic modifications leads to diverging
921 chromatin states and transcriptional heterogeneity in CLL. *Nat. Commun.* **10**, 1874
922 (2019).
923 74. Hua, X. *et al.* Genetic and epigenetic intratumor heterogeneity impacts prognosis of
924 lung adenocarcinoma. *Nat. Commun.* 2459 (2020) doi:10.1038/s41467-020-16295-5.
925 75. Zhu, B. *et al.* The genomic and epigenomic evolutionary history of papillary renal cell
926 carcinomas. *Nat. Commun.* **11**, 3096 (2020).
927 76. Mertins, P. *et al.* Proteogenomics connects somatic mutations to signalling in breast
928 cancer. *Nature* **534**, 55–62 (2016).
929 77. Zhang, H. *et al.* Integrated Proteogenomic Characterization of Human High-Grade
930 Serous Ovarian Cancer. *Cell* **166**, 755–765 (2016).
931 78. Mun, D.-G. *et al.* Proteogenomic Characterization of Human Early-Onset Gastric
932 Cancer. *Cancer Cell* **35**, 111-124.e10 (2019).
933 79. Chen, Y. J. *et al.* Proteogenomics of Non-smoking Lung Cancer in East Asia
934 Delineates Molecular Signatures of Pathogenesis and Progression. *Cell* **182**, 226-
935 244.e17 (2020).
936 80. Gillette, M. A. *et al.* Proteogenomic Characterization Reveals Therapeutic
937 Vulnerabilities in Lung Adenocarcinoma. *Cell* **182**, 200-225.e35 (2020).
938 81. Xu, J.-Y. *et al.* Integrative Proteomic Characterization of Human Lung
939 Adenocarcinoma. *Cell* **182**, 245-261.e17 (2020).
940 82. Hanahan, D. & Weinberg, R. A. Hallmarks of Cancer: The Next Generation. *Cell* **144**,
941 646–674 (2011).
942 83. Marty, R. *et al.* MHC-I Genotype Restricts the Oncogenic Mutational Landscape. *Cell*

943 **171**, 1272-1283.e15 (2017).
944 84. Marty Pyke, R. *et al.* Evolutionary Pressure against MHC Class II Binding Cancer
945 Mutations. *Cell* **175**, 416-428.e13 (2018).
946 85. DuPage, M., Mazumdar, C., Schmidt, L. M., Cheung, A. F. & Jacks, T. Expression of
947 tumour-specific antigens underlies cancer immunoediting. *Nature* **482**, 405–409
948 (2012).
949 86. Rooney, M. S., Shukla, S. A., Wu, C. J., Getz, G. & Hacohen, N. Molecular and
950 Genetic Properties of Tumors Associated with Local Immune Cytolytic Activity. *Cell*
951 **160**, 48–61 (2015).
952 87. McGranahan, N. *et al.* Allele-Specific HLA Loss and Immune Escape in Lung Cancer
953 Evolution. *Cell* **171**, 1259-1271.e11 (2017).
954 88. Anagnostou, V. *et al.* Evolution of neoantigen landscape during immune checkpoint
955 blockade in non-small cell lung cancer. *Cancer Discov.* **7**, 264–276 (2017).
956 89. Rosenthal, R. *et al.* Neoantigen-directed immune escape in lung cancer evolution.
957 *Nature* **567**, 479–485 (2019).
958 90. Jiménez-Sánchez, A. *et al.* Unraveling tumor–immune heterogeneity in advanced
959 ovarian cancer uncovers immunogenic effect of chemotherapy. *Nat. Genet.* **52**, 582–
960 593 (2020).
961 91. Davoli, T., Uno, H., Wooten, E. C. & Elledge, S. J. Tumor aneuploidy correlates with
962 markers of immune evasion and with reduced response to immunotherapy. *Science*
963 *(80-.).* **355**, (2017).
964 92. McGranahan, N. *et al.* Clonal neoantigens elicit T cell immunoreactivity and
965 sensitivity to immune checkpoint blockade. *Science (80-.).* **351**, 1463–1470 (2016).
966 93. Joshi, K. *et al.* Spatial heterogeneity of the T cell receptor repertoire reflects the
967 mutational landscape in lung cancer. *Nat. Med.* **25**, 1549–1559 (2019).
968 94. Roh, W. *et al.* Integrated molecular analysis of tumor biopsies on sequential CTLA-4
969 and PD-1 blockade reveals markers of response and resistance. *Sci. Transl. Med.* **9**,
970 eaah3560 (2017).
971 95. Tumeh, P. C. *et al.* PD-1 blockade induces responses by inhibiting adaptive immune
972 resistance. *Nature* **515**, 568–571 (2014).
973 96. Snyder, A. *et al.* Contribution of systemic and somatic factors to clinical response and
974 resistance to PD-L1 blockade in urothelial cancer: An exploratory multi-omic analysis.
975 *PLOS Med.* **14**, e1002309 (2017).
976 97. Wolf, Y. *et al.* UVB-Induced Tumor Heterogeneity Diminishes Immune Response in
977 Melanoma. *Cell* **179**, 219-235.e21 (2019).
978 98. Gejman, R. S. *et al.* Rejection of immunogenic tumor clones is limited by clonal
979 fraction. *Elife* **7**, e41090 (2018).
980 99. Maire, C. L. *et al.* Glioma escape signature and clonal development under immune
981 pressure. *J. Clin. Invest.* **130**, 5257–5271 (2020).
982 100. Zhang, A. W. *et al.* Interfaces of Malignant and Immunologic Clonal Dynamics in
983 Ovarian Cancer. *Cell* **173**, 1755-1769.e22 (2018).
984 101. Angelova, M. *et al.* Evolution of Metastases in Space and Time under Immune
985 Selection. *Cell* **175**, 751-765.e16 (2018).
986 102. Pennycuick, A. *et al.* Immune surveillance in clinical regression of pre-invasive
987 squamous cell lung cancer. *Cancer Discov.* CD-19-1366 (2020) doi:10.1158/2159-
988 8290.CD-19-1366.
989 103. Zapata, L. *et al.* Negative selection in tumor genome evolution acts on essential
990 cellular functions and the immunopeptidome. *Genome Biol.* **19**, 67 (2018).
991 104. Eynden, J. Van Den, Jiménez-sánchez, A., Miller, M. L. & Larsson, E. Lack of
992 detectable neoantigen depletion signals in the untreated cancer genome. *Nat. Genet.*

993 (2019) doi:10.1038/s41588-019-0532-6.
994 105. Zapata, L. *et al.* dN/dS dynamics quantify tumour immunogenicity and predict
995 response to immunotherapy. *bioRxiv* 2020.07.21.215038 (2020)
996 doi:10.1101/2020.07.21.215038.
997 106. Ghorani, E. *et al.* The T cell differentiation landscape is shaped by tumour mutations
998 in lung cancer. *Nat. Cancer* **1**, 546–561 (2020).
999 107. Keren, L. *et al.* A Structured Tumor-Immune Microenvironment in Triple Negative
1000 Breast Cancer Revealed by Multiplexed Ion Beam Imaging. *Cell* **174**, 1373-1387.e19
1001 (2018).
1002 108. Abduljabbar, K. *et al.* Geospatial immune variability illuminates differential evolution
1003 of lung adenocarcinoma. *Nat. Med.* **26**, 1054–1062 (2020).
1004 109. Wang, M. *et al.* Role of tumor microenvironment in tumorigenesis. *J. Cancer* **8**, 761–
1005 773 (2017).
1006 110. Korenchan, D. E. & Flavell, R. R. Spatiotemporal pH Heterogeneity as a Promoter of
1007 Cancer Progression and Therapeutic Resistance. *Cancers (Basel).* **11**, 1026 (2019).
1008 111. Lloyd, M. C. *et al.* Darwinian Dynamics of Intratumoral Heterogeneity: Not Solely
1009 Random Mutations but Also Variable Environmental Selection Forces. *Cancer Res.*
1010 **76**, 3136–3144 (2016).
1011 112. Hoefflin, R. *et al.* Spatial niche formation but not malignant progression is a driving
1012 force for intratumoural heterogeneity. *Nat. Commun.* **7**, 11845 (2016).
1013 113. Lujambio, A. *et al.* Non-Cell-Autonomous Tumor Suppression by p53. *Cell* **153**, 449–
1014 460 (2013).
1015 114. Peinado, H. *et al.* Pre-metastatic niches: organ-specific homes for metastases. *Nat.*
1016 *Rev. Cancer* **17**, 302–317 (2017).
1017 115. Montagner, M. *et al.* Crosstalk with lung epithelial cells regulates Sfrp2-mediated
1018 latency in breast cancer dissemination. *Nat. Cell Biol.* **22**, 289–296 (2020).
1019 116. Marusyk, A. *et al.* Non-cell-autonomous driving of tumour growth supports sub-clonal
1020 heterogeneity. *Nature* **514**, 54–58 (2014).
1021 117. Schürch, C. M. *et al.* Coordinated Cellular Neighborhoods Orchestrate Antitumoral
1022 Immunity at the Colorectal Cancer Invasive Front. *Cell* **182**, 1341-1359.e19 (2020).
1023 118. Wagner, J. *et al.* A Single-Cell Atlas of the Tumor and Immune Ecosystem of Human
1024 Breast Cancer. *Cell* **177**, 1330-1345.e18 (2019).
1025 119. Ali, H. R. *et al.* Imaging mass cytometry and multiplatform genomics define the
1026 phenogenomic landscape of breast cancer. *Nat. Cancer* **1**, 163–175 (2020).
1027 120. Jackson, H. W. *et al.* The single-cell pathology landscape of breast cancer. *Nature* **578**,
1028 615–620 (2020).
1029 121. Birkbak, N. J. & McGranahan, N. Cancer Genome Evolutionary Trajectories in
1030 Metastasis. *Cancer Cell* **37**, 8–19 (2020).
1031 122. Priestley, P. *et al.* Pan-cancer whole-genome analyses of metastatic solid tumours.
1032 *Nature* **575**, 210–216 (2019).
1033 123. Yates, L. R. *et al.* Genomic Evolution of Breast Cancer Metastasis and Relapse.
1034 *Cancer Cell* **32**, 169-184.e7 (2017).
1035 124. Hu, Z., Li, Z., Ma, Z. & Curtis, C. Multi-cancer analysis of clonality and the timing of
1036 systemic spread in paired primary tumors and metastases. *Nat. Genet.* **52**, 701–708
1037 (2020).
1038 125. Reiter, J. G. *et al.* Lymph node metastases develop through a wider evolutionary
1039 bottleneck than distant metastases. *Nat. Genet.* **52**, 692–700 (2020).
1040 126. El-Kebir, M., Satas, G. & Raphael, B. J. Inferring parsimonious migration histories for
1041 metastatic cancers. *Nat. Genet.* **50**, 718–726 (2018).
1042 127. Reiter, J. G. *et al.* Minimal functional driver gene heterogeneity among untreated

1043 metastases. *Science* **361**, 1033–1037 (2018).
1044 128. Rabbie, R. *et al.* Multi-site clonality analysis uncovers pervasive heterogeneity across
1045 melanoma metastases. *Nat. Commun.* **11**, 4306 (2020).
1046 129. NCT03004755. The PEACE (Posthumous Evaluation of Advanced Cancer
1047 Environment) Study (PEACE). *ClinicalTrials.gov*
1048 https://clinicaltrials.gov/ct2/show/NCT03004755.
1049 130. Parikh, A. R. *et al.* Liquid versus tissue biopsy for detecting acquired resistance and
1050 tumor heterogeneity in gastrointestinal cancers. *Nat. Med.* **25**, 1415–1421 (2019).
1051 131. Noorani, A. *et al.* Genomic evidence supports a clonal diaspora model for metastases
1052 of esophageal adenocarcinoma. *Nat. Genet.* **52**, 74–83 (2020).
1053 132. Lo, H. C. *et al.* Resistance to natural killer cell immunosurveillance confers a selective
1054 advantage to polyclonal metastasis. *Nat. Cancer* (2020) doi:10.1038/s43018-020-
1055 0068-9.
1056 133. Landau, D. A. *et al.* Evolution and Impact of Subclonal Mutations in Chronic
1057 Lymphocytic Leukemia. *Cell* **152**, 714–726 (2013).
1058 134. Andor, N. *et al.* Pan-cancer analysis of the extent and consequences of intratumor
1059 heterogeneity. *Nat. Med.* **22**, 105–113 (2016).
1060 135. Rye, I. H. *et al.* Intratumor heterogeneity defines treatment-resistant HER2+ breast
1061 tumors. *Mol. Oncol.* **12**, 1838–1855 (2018).
1062 136. Turajlic, S. *et al.* Deterministic Evolutionary Trajectories Influence Primary Tumor
1063 Growth: TRACERx Renal. *Cell* **173**, 595-610.e11 (2018).
1064 137. Cerrano, M. *et al.* Prognostic Impact of Clonal Diversity in Acute Myeloid Leukemia
1065 (AML) Treated with Intensive Chemotherapy (IC). *Blood* **134**, 2700 (2019).
1066 138. Iacobuzio-Donahue, C. A., Litchfield, K. & Swanton, C. Intratumor heterogeneity
1067 reflects clinical disease course. *Nat. Cancer* **1**, 3–6 (2020).
1068 139. Martinez, P. *et al.* Dynamic clonal equilibrium and predetermined cancer risk in
1069 Barrett's oesophagus. *Nat. Commun.* **7**, 12158 (2016).
1070 140. Wei, S. C., Duffy, C. R. & Allison, J. P. Fundamental Mechanisms of Immune
1071 Checkpoint Blockade Therapy. *Cancer Discov.* **8**, 1069–1086 (2018).
1072 141. Birkbak, N. J. *et al.* Paradoxical relationship between chromosomal instability and
1073 survival outcome in cancer. *Cancer Res.* **71**, 3447–3452 (2011).
1074 142. Jamal-Hanjani, M. *et al.* Extreme chromosomal instability forecasts improved outcome
1075 in ER-negative breast cancer: a prospective validation cohort study from the TACT
1076 trial. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.* **26**, 1340–1346 (2015).
1077 143. Lawson, A. R. J. *et al.* Extensive heterogeneity in somatic mutation and selection in
1078 the human bladder. *Science (80-.).* **370**, 75–82 (2020).
1079 144. Moore, L. *et al.* The mutational landscape of normal human endometrial epithelium.
1080 *Nature* **580**, 640–646 (2020).
1081 145. Killcoyne, S. *et al.* Genomic copy number predicts esophageal cancer years before
1082 transformation. *Nat. Med.* (2020) doi:10.1038/s41591-020-1033-y.
1083 146. Noble, R. & Burley, J. T. When , why and how tumour clonal diversity predicts
1084 survival. *Evol. Appl.* 1–11 (2020) doi:10.1111/eva.13057.
1085 147. Riaz, N. *et al.* Tumor and Microenvironment Evolution during Immunotherapy with
1086 Nivolumab. *Cell* **171**, 934-949.e16 (2017).
1087 148. Pearson, A. *et al.* High-Level Clonal FGFR Amplification and Response to FGFR
1088 Inhibition in a Translational Clinical Trial. *Cancer Discov.* **6**, 838–851 (2016).
1089 149. Russo, M. *et al.* Adaptive mutability of colorectal cancers in response to targeted
1090 therapies. *Science (80-.).* **366**, 1473–1480 (2019).
1091 150. Mok, T. S. *et al.* Osimertinib or platinum-pemetrexed in EGFR T790M-Positive lung
1092 cancer. *N. Engl. J. Med.* **376**, 629–640 (2017).

1093 151. Oldrini, B. *et al.* MGMT genomic rearrangements contribute to chemotherapy
1094 resistance in gliomas. *Nat. Commun.* **11**, 3883 (2020).
1095 152. Bell, C. C. *et al.* Targeting enhancer switching overcomes non-genetic drug resistance
1096 in acute myeloid leukaemia. *Nat. Commun.* **10**, 2723 (2019).
1097 153. Fong, C. Y. *et al.* BET inhibitor resistance emerges from leukaemia stem cells. *Nature*
1098 **525**, 538–542 (2015).
1099 154. Hinohara, K. *et al.* KDM5 Histone Demethylase Activity Links Cellular
1100 Transcriptomic Heterogeneity to Therapeutic Resistance. *Cancer Cell* **34**, 939-953.e9
1101 (2018).
1102 155. McGranahan, N. & Swanton, C. Neoantigen quality, not quantity. *Sci. Transl. Med.*
1103 **11**, eaax7918 (2019).
1104 156. Rambow, F. *et al.* Toward Minimal Residual Disease-Directed Therapy in Melanoma.
1105 *Cell* **174**, 843-855.e19 (2018).
1106 157. Pleasance, E. *et al.* Pan-cancer analysis of advanced patient tumors reveals interactions
1107 between therapy and genomic landscapes. *Nat. Cancer* **1**, 452–468 (2020).
1108 158. Gao, J. *et al.* Loss of IFN-γ Pathway Genes in Tumor Cells as a Mechanism of
1109 Resistance to Anti-CTLA-4 Therapy. *Cell* **167**, 397-404.e9 (2016).
1110 159. Sahin, U. *et al.* Personalized RNA mutanome vaccines mobilize poly-specific
1111 therapeutic immunity against cancer. *Nature* **547**, 222–226 (2017).
1112 160. Siravegna, G. *et al.* Clonal evolution and resistance to EGFR blockade in the blood of
1113 colorectal cancer patients. *Nat. Med.* **21**, 795–801 (2015).
1114 161. Xue, Y. *et al.* An approach to suppress the evolution of resistance in BRAF(V600E)-
1115 mutant cancer. *Nat. Med.* **23**, 929–937 (2017).
1116 162. Gopal, P. *et al.* Clonal selection confers distinct evolutionary trajectories in BRAF-
1117 driven cancers. *Nat. Commun.* **10**, 5143 (2019).
1118 163. Acar, A. *et al.* Exploiting evolutionary steering to induce collateral drug sensitivity in
1119 cancer. *Nat. Commun.* **11**, 1923 (2020).
1120 164. Algazi, A. *et al.* Abstract CT013: SWOG S1320: Improved progression-free survival
1121 with continuous compared to intermittent dosing with dabrafenib and trametinib in
1122 patients with BRAF mutated melanoma. *Cancer Res.* **80**, CT013 LP-CT013 (2020).
1123

1124 Highlighted Refs (not included in the actual References section because Mendeley deletes
1125 them)

1126 1. Williams #15
1127 a. This study demonstrates that a subset of tumours evolve without clear
1128 evidence of subclonal selection
1129 2. Gerstung #28
1130 a. As part of the Pan Cancer Analysis of Whole Genomes, this study provides an
1131 overview of evolutionary patterns across cancer types, identifying different
1132 driver events that typically occur early or late in cancer.
1133 3. Watkins #46
1134 a. This multi-region, pan cancer analysis of tumour karyotype uncovers parallel
1135 evolution of events within different subclones in one third of tumours
1136 sampled, and identifies the important role of chromosomal instability in
1137 generating subclonal diversity in cancer.
1138 4. Calabrese #56
1139 a. This pan-cancer study of paired whole genomes and transcriptomes illustrates
1140 the variety of transcriptomic alterations in cancer, and underlines the influence
1141 of copy number events and non-coding mutations on gene expression.
1142 5. Marjanovic #67

1143 a. This study of a mouse model of lung adenocarcinoma illustrates that highly
1144 plastic stem-like cells with diverse transcriptional states drive resistance to
1145 therapy and poor clinical outcome.
1146 6. McDonald #70
1147 a. This study of pancreatic cancer is an example of the important potential role of
1148 non-genetic variation in cancer evolution.
1149 7. Rosenthal #87
1150 a. This work highlights the role of immune editing in shaping early cancer
1151 evolution by negative selection, as well as the diversity of mechanisms of
1152 immune evasion.
1153 8. Wolf #95
1154 a. Here, a mouse model of melanoma is used to illustrate that increased clonal
1155 diversity of a developing tumour is associated with evasion of the anti-cancer
1156 immune response
1157 9. Rambow #154
1158 a. This work highlights the role played by stem-like cancer cells in non-genetic
1159 mechanisms of resistance to cancer therapy.
1160
1161

Figure 1: Functional and non-functional intra-tumour heterogeneity in tumour evolution

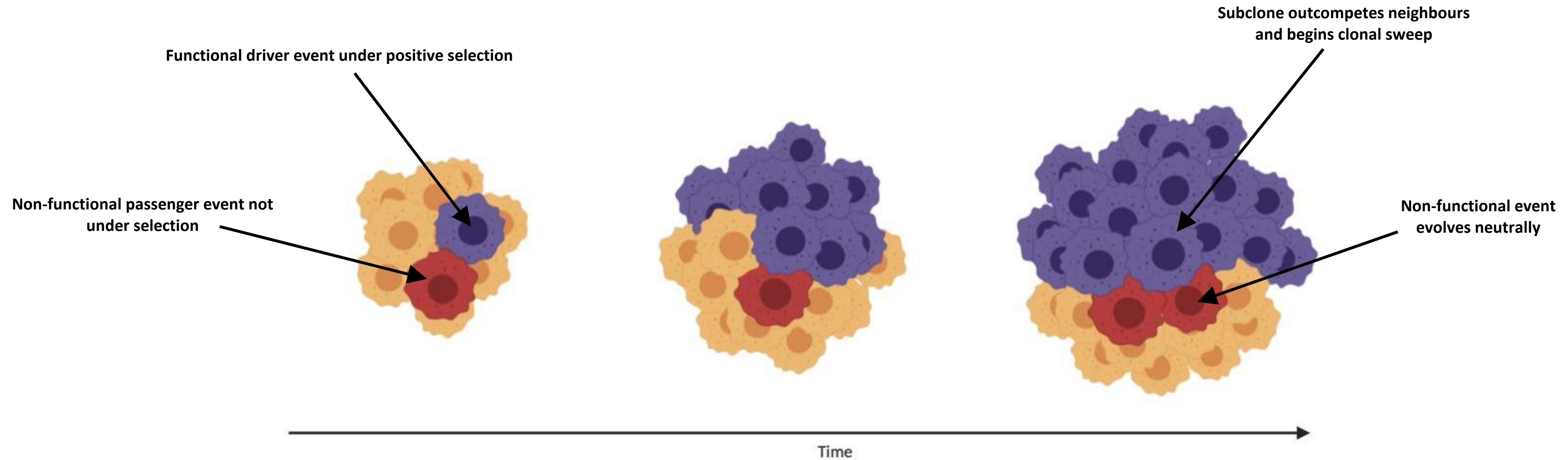


Fig. 1 | **Functional and non-functional intra-tumour heterogeneity in tumour evolution.** The increased rate of phenotypic variation in cancers compared with normal tissues means that new subclones arise and compete. A minority contain a driver event, such as a genetic mutation or copy number alteration, that grants a selective advantage. These subclones may grow at a faster rate than their neighbours and outcompete them in a 'selective sweep'.

Figure 2: Methods of assessing tumour evolution: clonal frequency inference

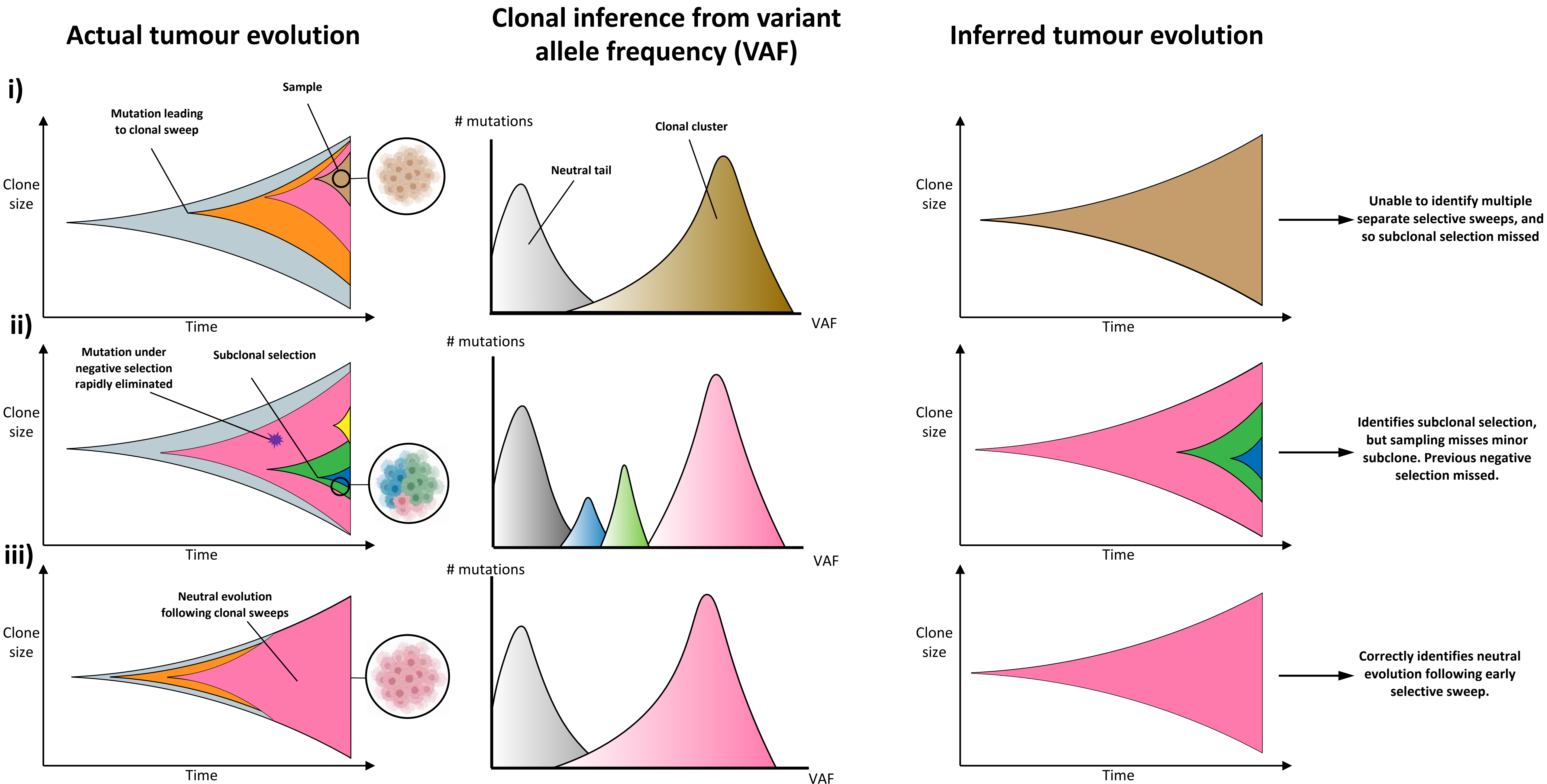
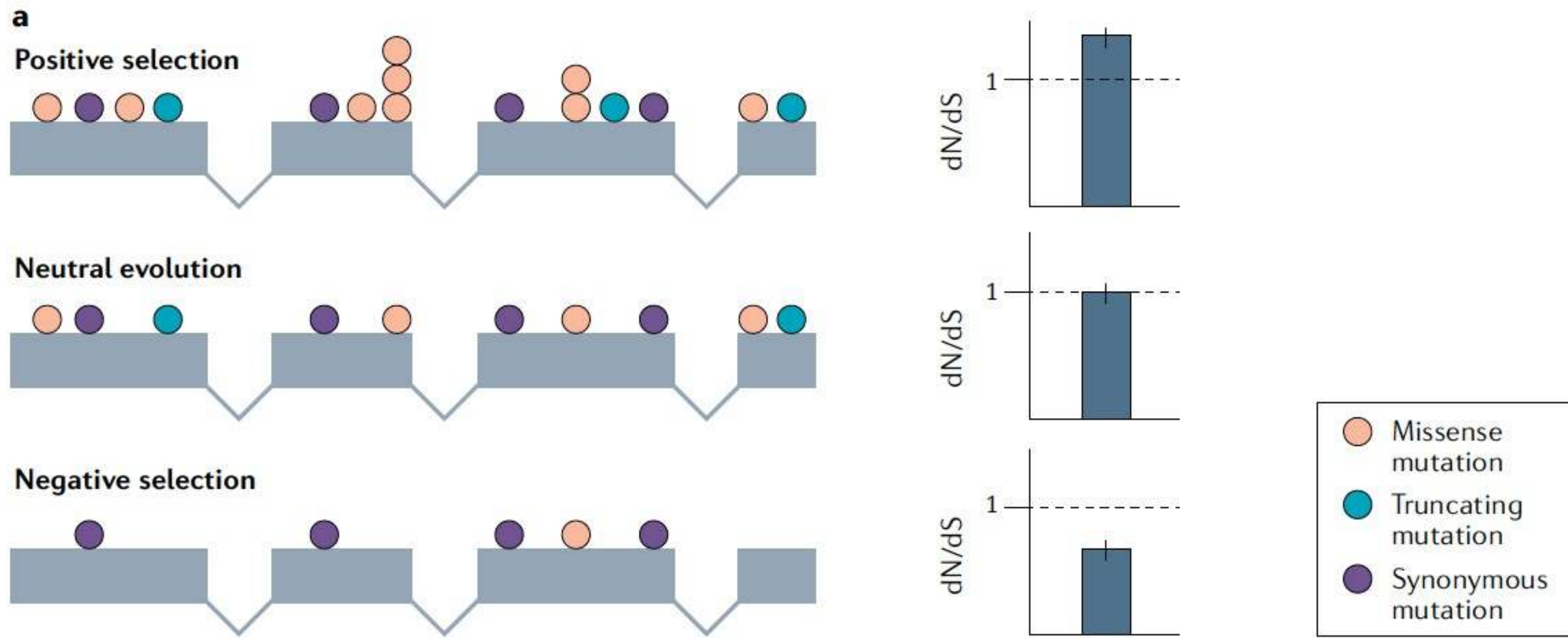
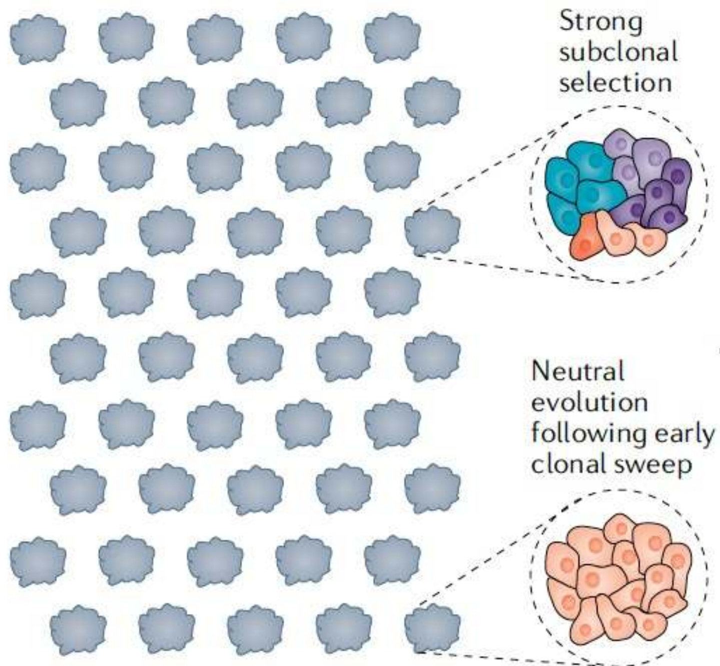


Fig. 2 | **Methods of assessing tumour evolution and clonal frequency inference.** Three examples of tumour evolution (left panels), with inferred clonal variant allele frequencies (VAFs) (middle panels) and tumour evolution (right panels). Clusters of mutations are ordered according to VAFs; the clonal cluster contains many high-VAF mutations, whereas the 'neutral tail' (coloured grey) contains mutations with lower VAFs. Interpretation can be confounded by sampling. a | Ongoing subclonal selection leads to formation of subclones (blue/grey, orange and blue). All blue/grey, orange and blue mutations appear clonal within the sample, and so have indistinguishable VAFs. Previous selective sweeps are not distinguishable, and so subclonal selection may not be identified.
b | Mutations (orange, green and purple) have different VAFs, so subclones are identifiable. Previous negative selection in this tumour is not identified.
c | Tumour evolution is reconstructed relatively accurately, with a previous clonal sweep and neutral ongoing evolution, although previous clonal sweeps that occurred very early are not distinguishable.





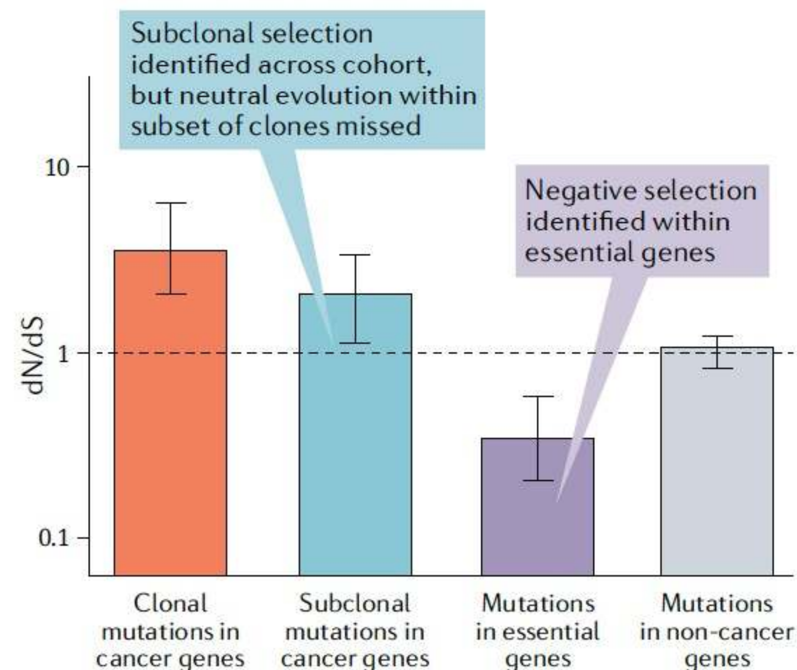


Fig. 3 | **Methods of assessing tumour evolution: dN/dS.** a | This method uses the ratio of non- synonymous mutations at non- synonymous sites to synonymous mutations at synonymous sites to infer selection. Ratio of substitution rates at non- synonymous sites versus synonymous sites (dN/dS) > 1 indicates that the rate of non- synonymous mutations is above the expected background, and signifies positive selection within a given gene or locus. dN/dS = 1 suggests mutations in that gene are neutral. dN/dS < 1 suggests negative or purifying selection. This technique works on a cohort level rather than in individual tumours, and ignores selection due to copy number alterations. b | An example cohort comprising tumours with varying evolutionary histories. c | dN/dS identifies clonal and subclonal selection of driver events within cancer genes. However, this approach is unable to infer selection within individual clones, and so the group of tumours with no subclonal selection is not identified. Previous negative selection is identified, as would be expected within genes essential for cellular function. Passenger mutations are not under positive or negative selection.

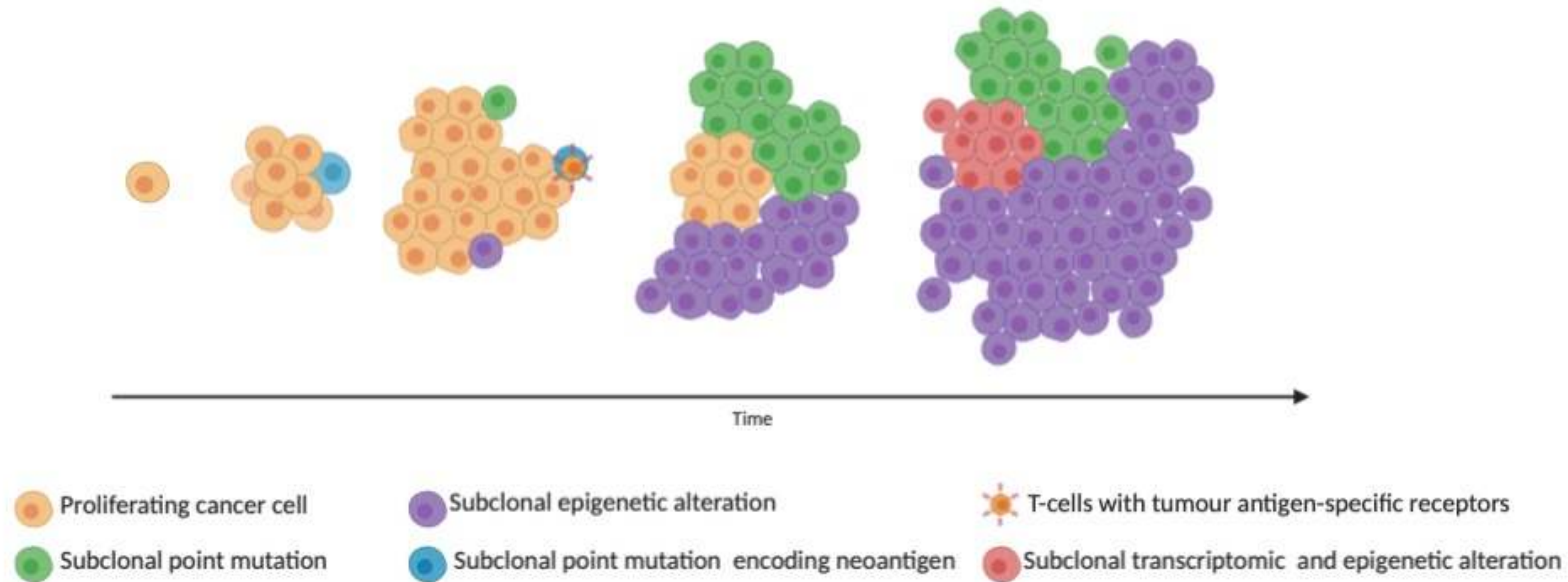
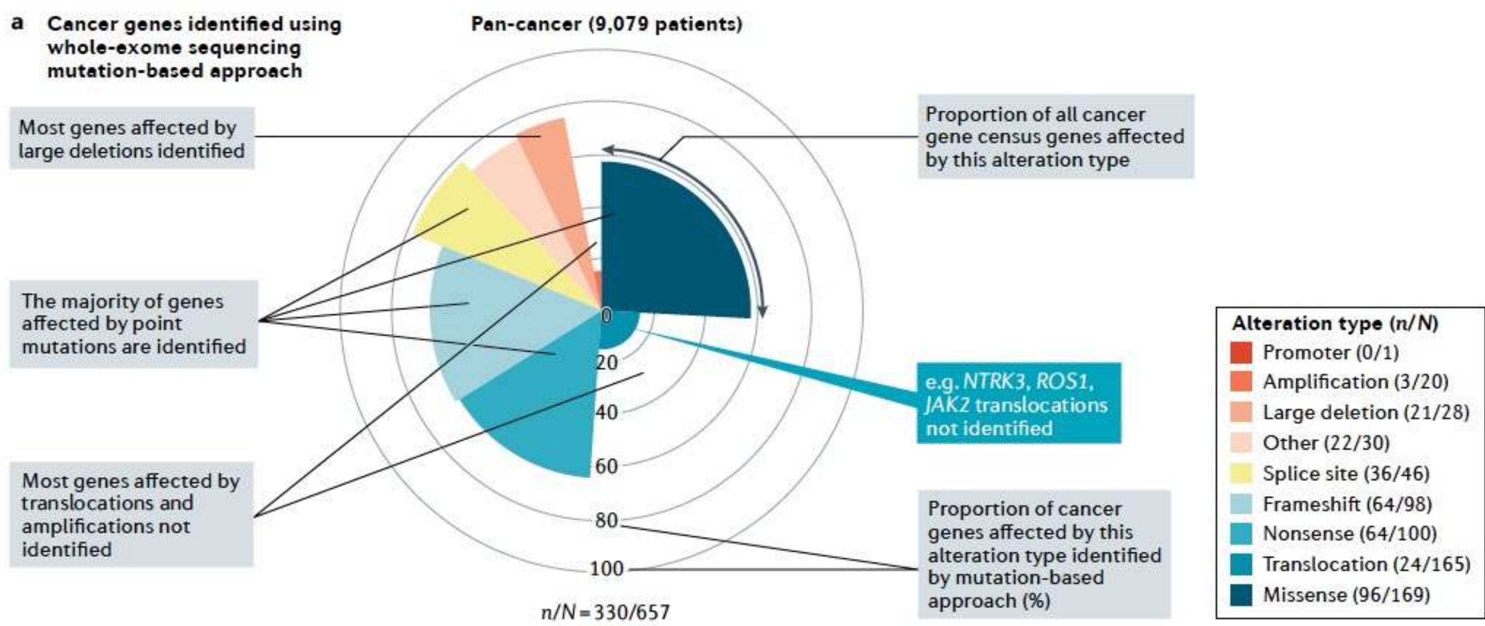


Fig. 4 | **Tumour evolution may be incorrectly classified using an exclusively genomic approach.** The evolving tumour acquires traits conveying selective advantages. In this example, subclones arise that contain genetic, epigenetic or transcriptomic alterations that give a selective advantage. The subclone containing an advantageous epigenetic alteration begins a selective sweep of the tumour. The subclone containing a neoantigen that stimulates the anticancer immune response is eliminated by neoantigen- reactive T cells. Here, an exclusively genomic approach would fail to identify all of the functional events shaping the evolution of this tumour.



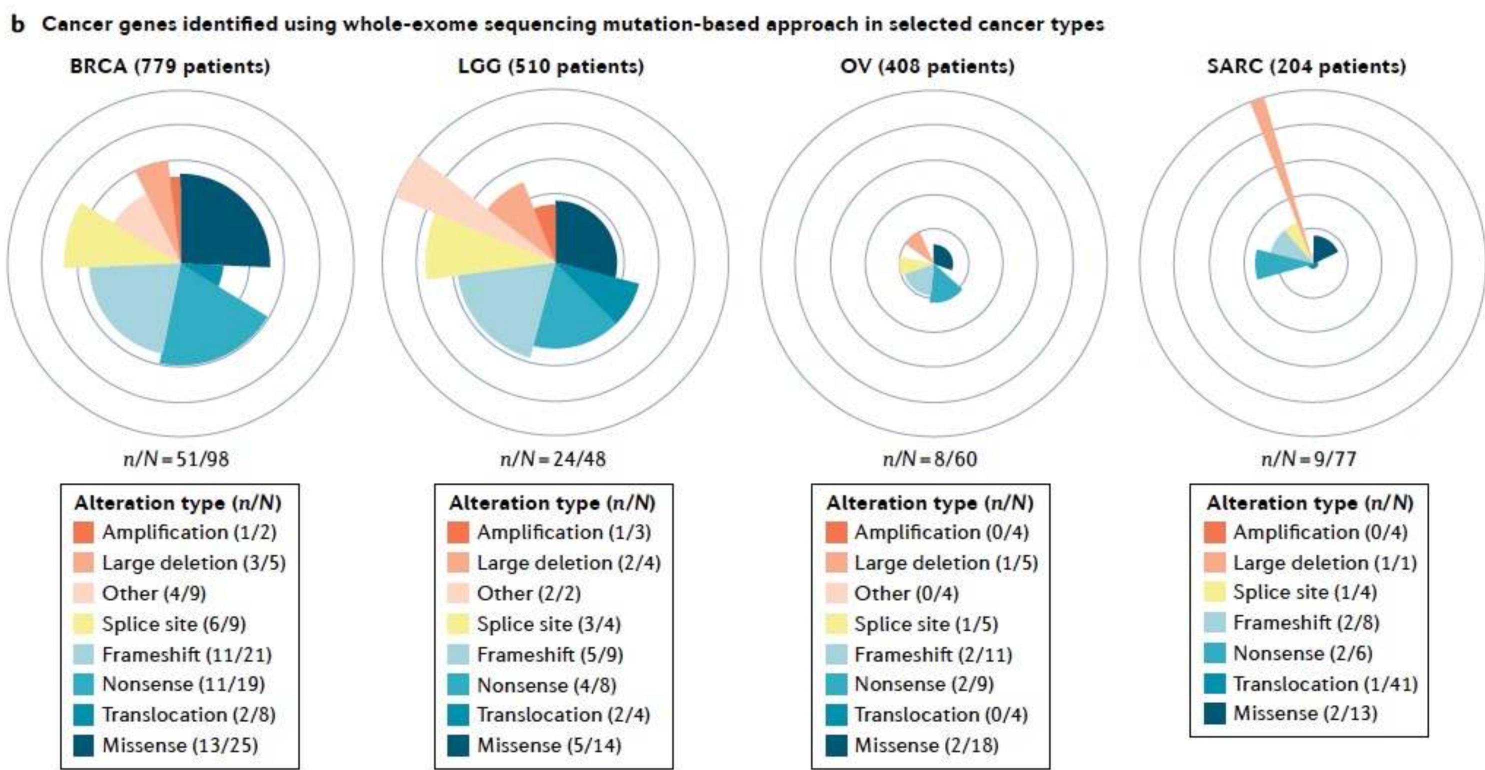
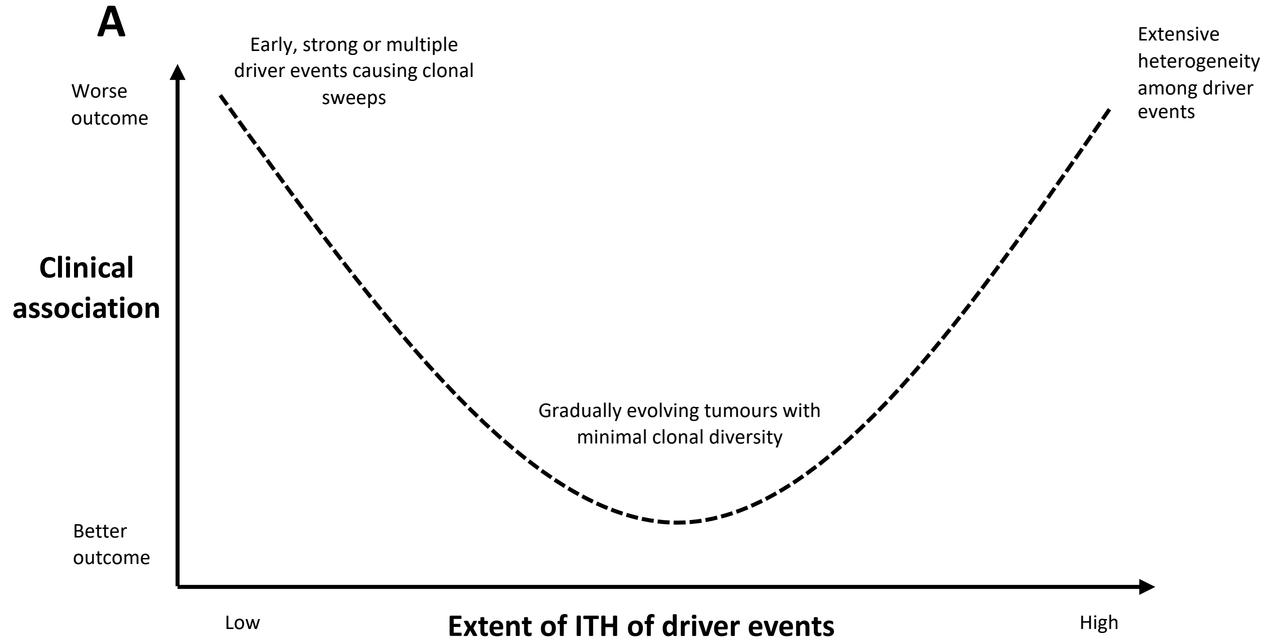


Fig. 5 | **Comparison of cancer genes defined by the COSMIC Cancer Gene Census and by a systematic pan-cancer whole-exome sequencing mutation-based approach.** The COSMIC Cancer Gene Census[60] curates lists of 'Tier 1' cancer driver genes that have a cancer- related functional role as well as documented evidence of recurrent cancer- causing mutations, and the types of alteration that affect these genes. a | Chart displaying the alteration types affecting COSMIC (release v90) cancer genes that overlap with cancer genes identified in a systematic mutation- based approach by Bailey et al.[59] based on pan- cancer whole- exome sequencing. In this analysis, genes that were rescued following manual curation by Bailey et al.[59] were excluded, and only genes listed in COSMIC as affecting cancer types studied by Bailey et al.[59] were considered. In some cases, COSMIC cancer genes are included more than once if they are affected by multiple alteration types. n is the number of cancer genes that were identified by the approach of Bailey et al.[59], while N is the total number of COSMIC cancer genes. With the whole- exome sequencing mutation- based approach, many COSMIC cancer genes were not identified, in particular, those affected by translocation and amplification. This underlines the importance of the role of cancer genes that are not frequently affected by exonic mutations as functional driver events in cancer. The systemic approach of Bailey et al. identified the majority of genes affected by large deletions; in such genes, there may be functional overlap of deletions and loss- of- function mutations, and so they may be more easily identified as drivers. Clearly, COSMIC lists are not exhaustive, and it is probable that many important driver events have not yet been identified. b | For four selected cancer types included in the analysis by Bailey et al.[59], the proportion of COSMIC cancer type- specific driver genes that were identified by a cancer type- specific whole- exome sequencing mutation- based analysis is shown. In the example, breast invasive carcinoma (BRCA) and brain lower grade glioma (LGG) have a relatively large proportion (52% and 50%, respectively) of known cancer type- specific cancer genes identified by the systematic mutation- based approach, whereas in ovarian serous cystadenocarcinoma (OV) and sarcoma (SARC) many known cancer genes are missed (13% and 12% identified, respectively). Cancer genes with translocations and amplifications are frequently missed; in sarcoma, for example, only 1 of 41 COSMIC cancer genes affected by a translocation is identified by Bailey et al.[59]. Data for each cancer type included in the analysis by Bailey et al. are shown in Supplementary Fig. 1.



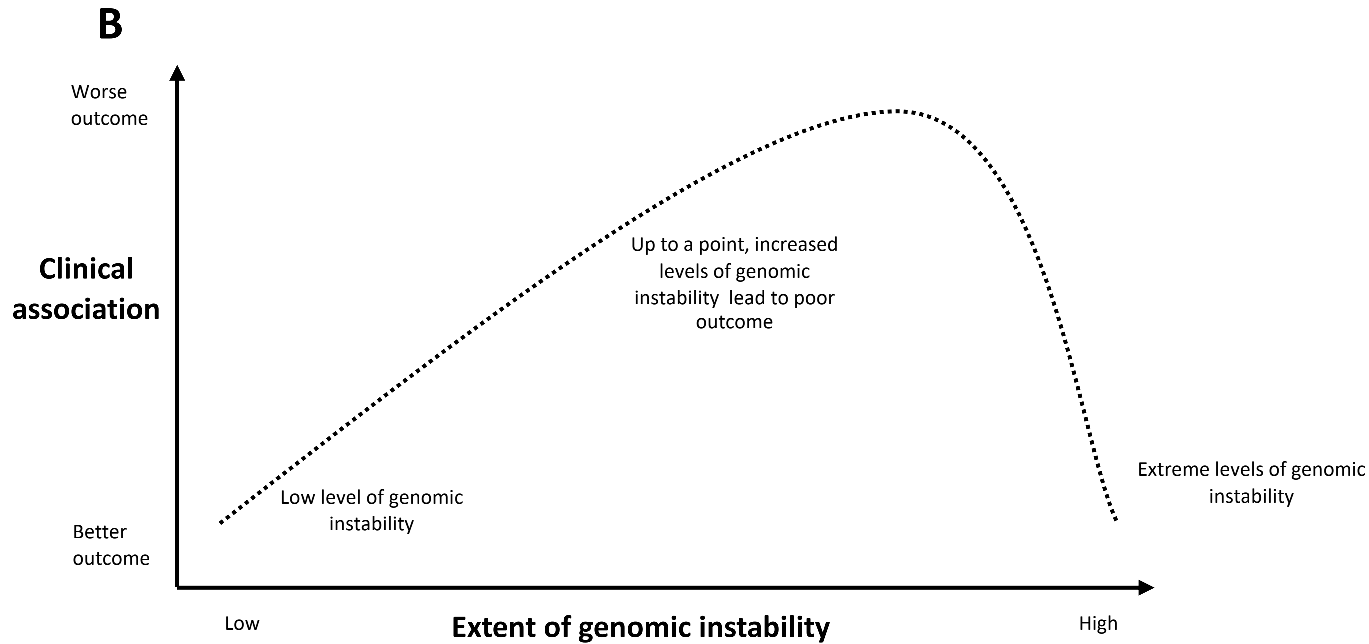
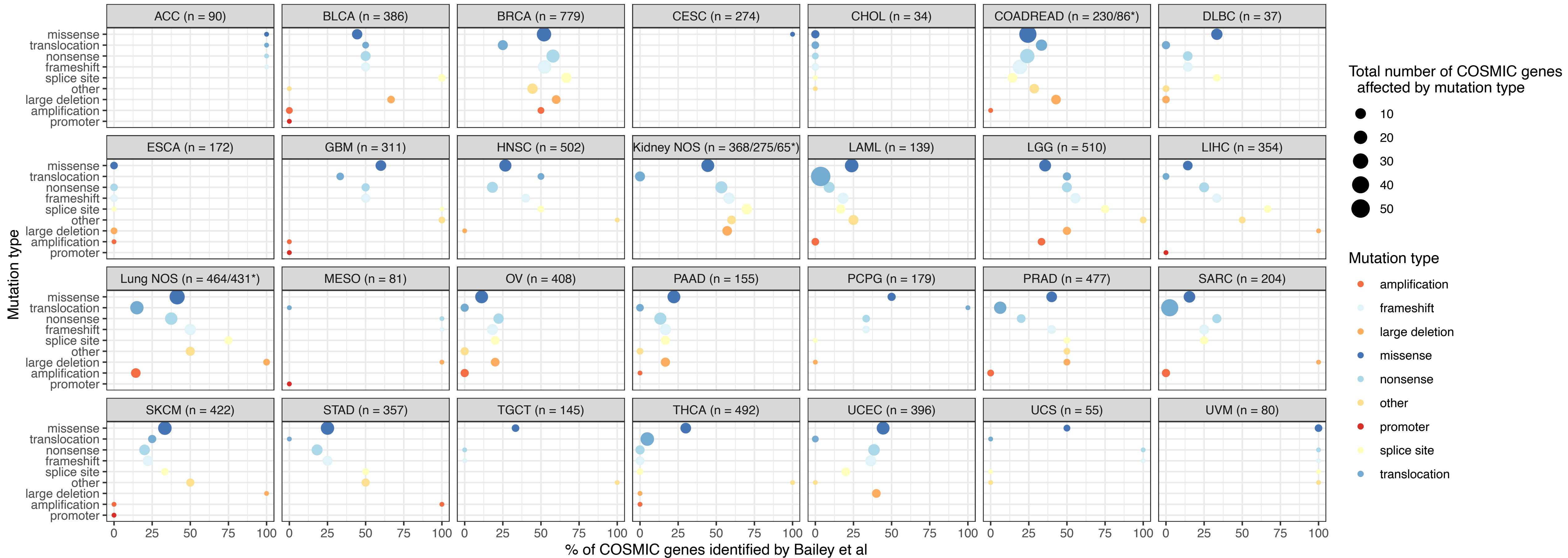


Fig. 6 | **Prognostic impact of intra-tumour heterogeneity.** The relationship between intra-tumour heterogeneity (ITH) and clinical outcome is complex, and is influenced by the degree of clonal diversity among driver events within the tumour as well as the extent of genomic instability. Hypothesized relationships between ITH of driver events and clinical outcome (part a) and between genomic instability and clinical outcome (part b).



Supplementary Figure 1: Extended comparison of mutation types within cancer genes defined by the COSMIC Cancer Gene Census and pan-cancer whole-exome sequencing. For all cancer types included in the analysis of Bailey et al that overlap with those in the COSMIC Cancer Gene Census, the proportion of cancer type-specific cancer genes, affected by certain mutation types, that are identified by a cancer-type specific analysis in Bailey et al is shown. In some cases, genes are included more than once if they can be mutated by multiple mechanisms. The number of samples of each tumour type analysed in Bailey et al is specified.

*Lung NOS & Renal NOS (not otherwise specified) - COSMIC Cancer Gene Census v90 does not specify relevant cancer subtypes studied in Bailey et al (LUAD, LUSC and KIRC, KIRP, KICH respectively); COAD and READ were analysed separately but consensus genes merged.