Open access • Posted Content • DOI:10.1101/454306

Genomic evidence supports a clonal diaspora model for metastases of esophageal adenocarcinoma — Source link

Ayesha Noorani, Martin Goddard, Jason Crawte, Ludmil B. Alexandrov ...+22 more authors

Institutions: University of Cambridge, Papworth Hospital, Los Alamos National Laboratory, Wellcome Trust Sanger Institute ...+2 more institutions

Published on: 30 Oct 2018 - bioRxiv (Cold Spring Harbor Laboratory)

Genomic evidence supports a clonal diaspora model for metastases of esophageal adenocarcinoma

Ayesha Noorani[1], Martin Goddard[2], Jason Crawte[1], Ludmil B. Alexandrov[3], Xiaodun Li[1], Maria Secrier[4], Matthew D. Eldridge[4], Lawrence Bower[4], Jamie Weaver[1], Pierre Lao-Sirieix[1], Inigo Martincorena[5], Irene Debiram-Beecham[1], Nicola Grehan[1], Shona MacRae[1], Shalini Malhotra[6], Ahmad Miremadi[6], Tabitha Thomas[7], Sarah Galbraith[8], Lorraine Petersen[7], Stephen D. Preston[2], David Gilligan[9], Andrew Hindmarsh[9], Richard H. Hardwick[10], OCCAMS Consortium, Michael R. Stratton[5], David C. Wedge[11, 12]* and Rebecca C. Fitzgerald[1]*

[1]MRC Cancer Unit, University of Cambridge, Biomedical Campus, Cambridge, CB2 OXZ, UK
[2]Department of Histopathology, Papworth Hospital NHS Trust, Cambridge, CB23 3RE, UK
[3]Theoretical Biology and Biophysics (T-6), Los Alamos National Laboratory, New Mexico, 87545, USA
[4]Cancer Research UK Cambridge Research Institute, Cambridge, CB2 0RE, UK
[5]Wellcome Trust Sanger Institute, Cambridge, CB10 1SA, UK
[6]Department of Histopathology, Cambridge University Hospitals NHS Foundation Trust, Cambridge, CB2 0QQ, UK
[7]Arthur Rank Hospice Charity, Cambridge, CB22 3FB, UK
[8]Department of Palliative Care, Cambridge University Hospitals NHS Foundation Trust, Cambridge, CB2 0QQ, UK
[9]Oncology Centre, Cambridge University Hospitals NHS Foundation Trust, Cambridge, CB2 0QQ, UK
[10]Oesophago-Gastric Centre, Cambridge University Hospitals NHS Foundation Trust, Cambridge, CB2 0QQ, UK
[11]Big Data Institute, University of Oxford, Oxford, OX3 7LF, UK
[12]Oxford NIHR Biomedical Research Centre, Oxford, OX4 2PG, UK

*Correspondence to Rebecca Fitzgerald rcf29@mrc-cu.cam.ac.uk or David Wedge david.wedge@bdi.ox.ac.uk

Abstract

Continual evolution of cancer makes it challenging to predict clinical outcomes. Highly varied and unpredictable patient outcomes in esophageal adenocarcinoma (EAC) prompted us to question the pattern and timing of metastatic spread. Whole genome sequencing and phylogenetic analysis of 396 samples across 18 EAC cases demonstrated a stellate pattern on the phylogenetic trees in 90% cases. The age-dependent trinucleotide signature, which can serve as a molecular clock, was absent or reduced in the stellate branches beyond the trunk in most cases ($p<0.0001$). Clustering of lymph nodes and distant metastases (n=250) demonstrated samples sharing a common clonal origin were widely dispersed anatomically. Metastatic subclones at autopsy were present in tissue and blood samples from earlier time-points. We infer that metastasis occurs rapidly across multiple sites, constituting a model of metastatic spread we term clonal diaspora. This has implications for understanding metastatic progression, clinical staging and patient management.

Introduction

In cancer, metastatic spread to distant sites accounts for the majority of deaths (Sporn, 1996). Understanding the anatomical extent of disease is essential to determine the optimum treatment strategy for any given patient. This is difficult in practice since cancer continually evolves at a microscopic scale, often beyond the resolution of clinical imaging techniques. Furthermore, the patterns of metastatic spread are often unpredictable in terms of time-course and anatomical location. Treatments may therefore be unnecessarily toxic (e.g. radical lymphadenectomy and high dose chemotherapy) or lead to under-treatment with high recurrence rates (Lou et al., 2013; Matsuda et al., 2017; Waterman et al., 2004).

Esophageal cancer is the sixth most common cause of cancer-related death worldwide and the current median survival time is still <1 year despite advances in treatment (Smyth et al., 2017). Incidence rates for esophageal adenocarcinoma (EAC) have risen sharply and it is now the predominant form in developed countries. Prognosis is highly variable for EAC patients and with a wide range in the proportion of patients surviving beyond 5 years (18-47% in patients with lymph node involvement) making it difficult to advise patients when embarking on a long course of grueling treatment (Cunningham et al., 2008; Waterman et

58 al., 2004). Despite the clinical classification of EAC as curative or non-curative, depending
59 on the location of associated lymph nodes (Japanese Gastric Cancer, 2011) and involvement
60 of solid organs, controversy exists in the field concerning whether radical lymph node
61 dissection improves outcome (Matsuda et al., 2017; Stiles et al., 2012; Waterman et al.,
62 2004).
63 Intratumor heterogeneity has been widely reported in human cancer with the first formal
64 description of clonal evolution espoused by Peter Nowell in 1976 (Nowell, 1976). Theories of
65 tumor evolution attempt to understand how tumor cell populations respond to selective
66 pressures (Greaves and Maley, 2012). Subsequently there has been much debate about the
67 models of tumor evolution, including linear, branching, neutral and punctuated evolution
68 (Davis et al., 2017; Klein, 2009). With the advent of genome-wide sequencing methods
69 recent large-scale efforts have been made to delineate different models of evolution,
70 summarized in Table S1. Knowledge of how genetic diversity emerges over time as
71 metastases develop remains limited, in part due to the challenge in collecting multiple
72 samples over space and time from cancer patients.
73 To understand the evolution of EAC, we designed a prospective study with extensive
74 sampling over-time including samples from diagnosis, surgery for operative cases and warm
75 autopsy. We used whole genome sequencing at high (50x) and shallow (1x) coverage to
76 interrogate the clonal architecture across time and space. The overall study design, sampling
77 and sequencing strategy are shown in Figure S1.
78

79 ## Results

80 ### Genomic architecture of 18 cases

81 Eighteen cases were included and the clinical demographics of these cases are shown in
82 Table S2 and S3, with details of the individual samples used given in Table S4 and S5. In the
83 first part of the study (Step 1, Figure S1C) we used 50x WGS to construct a phylogenetic tree
84 for each case to understand the relationship between the primary and metastases (Figure 1
85 and Figures S2 and S3). Mutation clustering was performed, and the fractions of tumor cells
86 carrying each set of mutations (Cancer Cell Fraction, CCF) within each sample were used to
87 determine: 1) the clonal and sub-clonal architecture of each tumor (subclonal CCF <95%,
88 clonal CCF ≥ 95%); 2) the hierarchy of events; and 3) the distance of these sub-clonal or
89 clonal clusters from the most recent common ancestor (MRCA) as determined by the

90 number of mutations on each branch of the tree (Figure S1C). The CCF of each clone and
91 subclone is shown in Tables S6 and S7 along with the total number of single nucleotide
92 variants (SNVs) and the tumor purity estimated using the Battenberg algorithm in Table S8.
93 The confidence intervals of the CCFs these clones and subclones are shown in
94 Supplementary Tables S9. All tissue samples undergoing WGS (all snap frozen except
95 diagnostic biopsies which were FFPE archival samples) were also dually scored by expert
96 esophageal histopathologists using the standards set by the International Cancer Genome
97 Consortium (ICGC). Further macro dissection was performed for low cellularity samples to
98 achieve a cellularity >70% to avoid bias (Supplementary Methods).
99 These analyses enabled us to construct phylogenetic trees, in which the trunk represents
100 the mutations common to all samples with length proportional to the number of mutations
101 required for malignant transformation in that case. Branches represent subclones whose
102 mutations are not found in every cancer cell. In all cases we observed a long trunk
103 compared to the rest of the tree (median 19,034 SNVs, IQR 11,299-63,908), consistent with
104 previous studies in EAC (Gerstung et al., 2017; Murugaesu et al., 2015). The median size of
105 clusters across all cases was 3,069 SNVs (IQR1332-63908) and only 2/157 clusters contained
106 fewer than 200 SNVs (case P5 and case S1, Figure S2).
107 The key driver events (Dulak et al., 2012; Secrier et al., 2016) are depicted on each
108 phylogenetic tree (Figure 1 and Figure S2). The most frequent events that have previously
109 been classed as drivers (Dulak et al., 2012; Frankell et al., 2018; Secrier et al., 2016) occurred
110 in the trunks of the phylogenetic trees. TP53 was mutated in the trunk of 16 out of 18 cases,
111 consistent with our knowledge of the disease (Dulak et al., 2013; Nones et al., 2014; Ross-
112 Innes et al., 2015; Secrier et al., 2016; Weaver et al., 2014). Amplifications (gene names in
113 red) were often truncal, but were also observed on the branches of the phylogenetic tree,
114 providing evidence of divergence further down the evolutionary lineage (Figure 1, Figure
115 S2). The majority of events in driver genes were copy number alterations rather than
116 missense variants (Figure 1, Figure S2) (Frankell et al., 2018; Nones et al., 2014; Secrier et al.,
117 2016). There was no significant difference in the overall number of structural variants
118 between primary and metastatic samples (p=0.41, generalized linear model), (Figure S4b).
119 However, a larger proportion of structural variants in metastatic samples were retro-
120 transpositions of mobile elements compared with those in the primary samples (p=0.045,
121 Figure S4c). This contrasts with pancreatic cancer, where deletions and fold-back inversions

122 are more common, and breast cancer where tandem duplications dominate (Yates et al.,
123 2015). Furthermore, the proportion of SVs found uniquely in metastases or in primary sites
124 was higher than that of SNVs (Figure 1, Figure S4a), suggesting an increase in genomic
125 instability in later stages of the disease. However, it cannot be ruled out that some SVs have
126 not been identified in every sample as a result of lower sensitivity in the detection of SVs
127 than SNVs.
128 Across the eighteen cases, 8 mutational signatures were observed, with varying prevalence
129 and consistent with previous studies (Figure S5 (Ajani et al., 2015; Mariette et al., 2003;
130 Sottoriva et al., 2013; Yachida et al., 2010). None of the signatures observed have been
131 associated with treatment with alkylating antineoplastic agents (Alexandrov et al., 2013),
132 platinum therapy (Liu et al., 2017) or radiation therapy (Behjati et al., 2016).
133
134 **Sub-cohort analysis of cases with local and distant spread**
135 Ten of eighteen patients (S3, S4, P1-4, P6, P8-10) had nodal and solid organ metastases,
136 allowing a direct comparison of the genomic architecture between different metastatic sites
137 (Figure 1).
138 In four of these ten cases, an isolated clone or subclone confined to distant metastases
139 shared the highest congruence to the most recent common ancestor (MRCA), depicted as a
140 dashed black node on the first branch of the phylogenetic tree (P1, P4, P10, S3 in Figure 1).
141 In P1, this subclone was shared between the primary tumor and a pleural metastasis. In S3
142 and P4, the clone involved in this isolated seeding was identified at a single distant site and
143 not in the primary tumor (S3: liver metastasis (D1), P4: para-aortic lymph node (L3)); both
144 events would render a patient's management palliative. Interestingly, in P9 a subclone was
145 found in a premalignant area of Barrett's esophagus and a pleural metastasis but not in any
146 of four areas of the primary tumor subject to 50X WGS. This lineage shares no variants with
147 the main lineage and appears to be an independent second cancer. While undetected in any
148 of the adenocarcinoma samples, it is plausible that this arose from an unsampled area of the
149 primary tumor (Figure 1). The high CCF (>0.9) in both these cases suggests that these
150 mutations developed *de novo* in metastatic sites soon after dissemination. In P10, the early
151 seeding cluster was shared between a distant para-aortic node and a sub-clonal metastasis
152 in the right hemi-diaphragm. This isolated seeding event showed little divergence from the

153 MRCA (median 1913 SNVs, IQR 1540-1421) and suggests early seeding to distant
154 metastases.

155 A striking observation was that 9/10 cases had a clone (outlined in red on the phylogenetic
156 tree) that was followed by a dispersion of multiple subclones from the primary to discrete
157 metastatic sites in a stellate pattern on the phylogenetic tree. The subclones forming this
158 distinct pattern were located in both primary and metastatic tissue in eight cases (P1, P2,
159 P3, S4, P4, P6, P8, P10) and in P9 were unique to metastases (Figure 1). P9 harbored sub-
160 clonal CCFs in multiple sites, which could also indicate metastasis-to-metastasis seeding and
161 further evidence for this was sought in the following parts of our study. The only case
162 lacking a stellate pattern on the phylogenetic tree was S3, a non-autopsy case with limited
163 tissue sampling. The early distant seeding in S3 is consistent with a pattern of parallel
164 evolution (Figure 1).

165

166 **Shallow whole genome sequencing to assess spatial spread of EAC**

167 In the second step of the study we aimed to elucidate the relative timing of metastatic
168 events and to do this we performed 1X WGS in a further 250 tissue samples from 6 autopsy
169 cases (Figure 2, Figure S1B). We did not call new mutations, as this would not be possible at
170 1X sequencing, but used this method to detect the spread of clones and subclones
171 previously identified using 50X WGS (validation of methods in Figures S6 and S7). The
172 samples used for this part of the study are outlined in Table S10. The median size of clusters
173 (identified at 50X WGS) that we aimed to detect using 1X WGS was 3,784 (IQR 1966-49955).
174 Sample sites were grouped according to their similarity based on the presence of subclones
175 and clones previously detected with 50X WGS (Supplementary Methods, Shallow Whole
176 Genome Sequencing for Subclone Identification). The resulting groups of samples are color
177 coded and numbered, and the distribution of sample sites in these groups is shown on the
178 adjacent body map. (Figure 2, see also Supplementary Methods). The most striking
179 observation is that samples that grouped together based on shared clonal origins were
180 widely dispersed anatomically.
181 Four out of six cases with extensive spatial sampling (Figure 2) had liver metastases
182 evaluated and 3 of these contained samples that were more similar to local lymph node
183 metastases than neighboring liver metastases (P4, P6, P8 but not P10). The high number of

184 groups within the liver (up to 4 in P6) suggested seeding by multiple subclones (seen in P4,
185 P6, P8), whereas the single group (orange, number 3) in the liver of P10 indicated a single
186 clonal expansion.
187 A comparison of lymph node location and genomic contiguity revealed no evidence of
188 tropism, i.e. genomically similar lymph nodes did not occupy nearby anatomical locations.
189 Lymph nodes above and below the diaphragm were frequently seeded from common
190 events (P2: clusters 1, 3; P4: clusters 5, 6; P6: cluster 5; P8: clusters 2, 3, 6; P10: cluster 4), at
191 odds with a progression from local to distant nodes. Similarly, a comparison of lymph node
192 and solid organ metastases revealed scant evidence for tropism, with the exception of P1
193 (Supplementary Methods). In this cancer, separate subclones seeded lymph node and
194 pleural metastases (Figures 1 and 2). However, the distant metastasis (D1) was seen to
195 branch earlier than the lymph node metastasis in the evolutionary tree in our 50X WGS
196 analysis (Figure 1).
197 We further traced regions of the primary tumor at autopsy with similar subclonal
198 compositions to each of the groups of metastases, shown as adjacent tumor maps (Figure 2,
199 bottom left of each case). Subclones occupied discrete, spatially distinct areas in the primary
200 tumor.
201

202 **Timing of metastatic spread**

203 To examine the timing of metastatic spread we analyzed the mutational signatures as well
204 as comparing the mutations present at autopsy with those observed in the diagnostic biopsy
205 samples and longitudinally collected plasma samples (Figure S1C).
206 Signature 1 arises from the enzymatic deamination of methylated cytosines which is an
207 endogenous process that occurs continuously in both healthy and cancerous cells. This has
208 been shown to act as a molecular clock, (Alexandrov et al., 2015; Alexandrov et al., 2013;
209 Blokzijl et al., 2016; Gao et al., 2016; Letouze et al., 2017; Lodato et al., 2018), and was
210 therefore used here as a method to examine the temporal relationship between
211 metastases. Using a previously described method for deconvolving mutational signatures
212 (Alexandrov et al., 2015), we observed that signature 1 was present in the trunk but absent
213 in all subclones that constituted the stellate pattern on the phylogenetic tree (following the
214 red clone in Figure 1) for P2, P4, P6, P9, P10, S4 and it was significantly reduced for P1 (21%
215 to 3%) and P3 (16% to 9%) (Wilcoxon signed rank test p=0.039, Figure S15). Suspecting that

216 the number of signature 1 mutations in branch subclones was below the resolution of our
217 deconvolution methods, we identified the number of mutations with the characteristic
218 feature of signature 1, i.e. C>T mutations at a CpG context, along the trunk to the stellate
219 pattern and on the longest branch leading from the stellate pattern. With the exception of
220 P8, the proportion of mutations with this feature was significantly lower post stellate
221 pattern ($p < 9.1e-5$, Chi-squared test) and the median proportion of such mutations
222 occurring prior to the stellate pattern was 0.911 (Figure 3B). Thus, in the majority of cases
223 one might deduce that very little time has elapsed between the appearance of the cell that
224 is ancestral to disseminating cells and the individual cells that seeded each of the
225 metastases. The substantial number of mutations arising from other mutational processes
226 later in the evolutionary history (Table S11) suggests an increase in the activity of other
227 processes. Of note, there was an increase in the proportion of signature 3 in subclonal SNVs
228 compared to clonal SNVs, with this signature being associated with the failure of DNA
229 double strand break repair (Wilcoxon signed rank test $p=0.019$, Figure S15).
230 Next we investigated eight cases (P1-4, P6, P8-10) for which the esophageal diagnostic
231 biopsy (FFPE tissue) was available, with a median time prior to autopsy of 12 months (range
232 5-30 months, Figure 3A). 1X WGS identified metastatic subclones detected at autopsy in the
233 diagnostic sample, ranging from 100% of subclones in 6 cases (P2, P4, P6, P8, P9, P10) to
234 75% in P3 and 57% in P1. This analysis also clarified that all metastatic subclones in P9
235 (Figure 1 identified from 50X WGS in the clonal discovery part of the study) arose from the
236 primary site and were present at the earliest presentation of disease.
237 Tumor reseeding from metastatic sites to the primary tumor (Kim et al., 2009) or from
238 metastasis to metastasis (Gundem et al., 2015) have been suggested as possible modes of
239 spread. However, in this study the direction of seeding from primary to metastases, rather
240 than *vice versa*, is clearly indicated by two observations. Firstly, the founder clone of the
241 diaspora was observed in the primary tumor for all cases. If this clone originated in a
242 metastasis, we would expect to observe additional subclones shared between multiple
243 metastases and not the primary, but such subclones are not observed. Secondly, the
244 founder clone of each diaspora and the great majority of subclones below it in the
245 phylogenetic tree were identified in primary diagnostic samples, indicating that this clone
246 was already present in each primary tumor at diagnosis. Metastasis to metastasis seeding
247 would result in a subclone present in multiple metastases and absent from the primary

248 tumor. When we included all 1X WGS and 50X WGS samples (8 and 4 samples, respectively,
249 from the primary), no such subclones were identified in our cases.
250 In eight cases, plasma was available from rapid autopsy and 1X WGS of circulating tumor
251 DNA (ctDNA) demonstrated that in all but one case (P1), every subclone from autopsy was
252 also represented in plasma (Figure S8). In P1, the three subclones not found in the plasma
253 were distal sub-clonal branches on the phylogenetic tree. We also assessed the clonal
254 composition of ctDNA at earlier time-points in five available cases and assessed the TP53
255 fraction using digital PCR (Figure 3C, Figure S9, Table S12). All clones and subclones from the
256 50X WGS phylogenetic tree were detected at earlier time points in S4, P6, P10, (Figure 3C),
257 while 83% and 29% of metastatic subclones were detected in S3 and P1 (Figure S9),
258 respectively. Interestingly, P6 was a patient being treated with curative intent and had no
259 radiological evidence of distant nodal or organ metastases at the time of clinical staging.
260 However, at the time of diagnosis all subclones later found in the metastases were already
261 present in the blood plasma (Figure 3C). Case S4 is noteworthy as the brain metastases (D1,
262 D2 in Figure 1) appeared to have originated from a subclone shared between the primary
263 and a local lymph node, both of which were removed at the time of surgery (Figure 3C).
264

265 **Discussion**

266 We have gathered multiple lines of evidence which suggest that for the majority of EACs
267 metastasis occurs rapidly to multiple sites. These lines of evidence can be summarized as
268 follows. We observe multiple subclones each seeding multiple metastatic sites. These
269 subclones are frequently derived from a single parental clone, often resulting in a stellate
270 pattern on the phylogenetic tree. Metastases in solid organs can bypass nodal involvement.
271 Samples within solid organ sites frequently resemble distant metastases more closely than
272 neighboring metastases within the same organ, i.e. no tropism is observed. All metastases
273 appear to have spread directly from the primary site, with little or no evidence of
274 metastasis-to-metastasis seeding. One interesting possibility is that because the esophagus
275 is highly vascularized, it may be particularly subject to hematogenous spread.
276

277 These features differ in some respects from previously described models of metastasis and
278 we propose that they may constitute a distinct model of evolution. We suggest that this
279 pattern could be referred to as a 'diaspora', by extension of the anthropological term to

280 cancer (Pienta et al., 2013). Within this context, it is associated with the observation that
281 multiple cell populations in metastatic sites are directly linked to the primary site of origin
282 and that individual subclones seed multiple tissue types, analogous to a diaspora crossing
283 multiple national boundaries.
284
285 A number of features were frequently associated with this phenomenon (Figure 4), with 9 of
286 the cases (all except S3) displaying at least 2 of the 4 following features: i) stellate pattern
287 on the phylogenetic tree; ii) lack of signature 1 mutations post MRCA or post-diaspora; iii)
288 spread of subclones to multiple organs of different type; iv) evidence for selection in post
289 diaspora genotypes. Regarding the latter, we looked for driver amplifications post MRCA or
290 post diaspora on a per case basis and identified selection in 6/10 cases. However, this is
291 likely to be an under-estimate, since there may be non-copy number drivers present in
292 additional cases. The ratio of non-synonymous to synonymous mutations (dN/dS) analyzed
293 across all cases as a whole, (Dentro et al., 2018), indicated positive selection in both clonal
294 and subclonal genomes, albeit with lower levels of selection within subclones (Figure S10).
295
296 Until recently the genomic architectures of metastatic samples have not been defined with
297 enough resolution to discern temporal patterns of metastatic spread. Several distinct
298 patterns are now emerging which are not necessarily mutually exclusive or cancer-type
299 specific. In pancreatic cancer, Yachida et al. demonstrated that distant organ seeding was a
300 late event consistent with a linear progression model (Yachida et al., 2010). In prostate
301 cancer, linear progression is often succeeded by multiple waves of seeding (Gundem et al.,
302 2015). The same study further demonstrated widespread subclonal evolution in metastases
303 and metastasis-to-metastasis spread in keeping with the relatively long longevity of prostate
304 cancer. Strikingly, a stellate pattern was not observed in any of the cases in that study,
305 despite using a similar design to that used in this study.
306
307 In Table S13 we compare the features of our proposed Diaspora model to the previously
308 posited linear (Foulds, 1954) and parallel (Klein, 2009) models and consider the implications
309 for clinical practice. Whereas the linear model predicts that linear progression will lead to a
310 single subclone seeding lymph node sites followed by transmission to distant organs, the
311 diaspora model posits simultaneous seeding of multiple sites directly from the primary.

312 Unlike the parallel model, the diaspora model implies that metastasis formation occurs after
313 the majority of evolution has occurred in the primary tumor, resulting in multiple subclones
314 found in common between primary and metastatic tumors. Contemporaneous with this
315 study, lymphatic and distant metastases in colon cancer have been shown to arise from
316 independent subclones in the primary tumor with disparate evolutionary trajectories
317 (Naxerova et al., 2017). In contrast, in EAC we find that individual subclones frequently seed
318 both lymph node and distant organs suggesting that disparate trajectories for nodal and
319 solid organ metastases do not exist for this disease (Figure 1, 2). Of note we acknowledge
320 that, despite the extensive and systematic sampling across all autopsy cases, further
321 sampling may add further branches to our phylogenetic tree although this is unlikely to
322 affect the diaspora event itself.
323
324 In common with the Big Bang Model proposed for colorectal cancer (Sottoriva et al., 2015),
325 our model predicts the occurrence of highly branching phylogenies. However, the Big Bang
326 Model proposes neutral dynamics, whereas we observe strong evidence for selection in
327 subclonal populations in the form of dN/dS ratios and the occurrence of subclonal driver
328 amplifications (Figure 1, Figures S10 and S11). Moreover, the clonal maps of the primary
329 tumor demonstrate subclones that occupy spatially discrete areas of the primary tumor
330 (Figure 2), in contrast to the intermixed subclones predicted by the Big Bang Model
331 (Sottoriva et al., 2015).
332
333 The sequence of events in metastatic progression has far-reaching clinical implications
334 (Table S13). Clonal architecture in EAC defies anatomical location of lymph node stations
335 and distant sites, which is the current basis for the TNM staging and determines whether
336 curative therapy is appropriate. It has been suggested that the high recurrence rate, 52%
337 within one year, results from seeding of distant metastases that are not detected at the
338 time of diagnosis (Mariette et al., 2003). This study provides molecular evidence for this
339 observation and highlights the need for different systemic approaches to disease
340 management, including more aggressive adjuvant therapy which is not currently the
341 mainstay of treatment (Burt et al., 2017; Gabriel et al., 2017; Pasquali et al., 2017; Sjoquist
342 et al., 2011). Furthermore, the presence of subclones identified from autopsy in diagnostic
343 blood and tissue samples suggests that there is the potential for earlier detection.

The occurrence of metastasis is a pivotal event in the life history of a cancer. Understanding the mechanism behind such an event would have potential relevance to predicting and preventing metastatic spread. From the relatively modest number of cases within this study, we were not able to identify aberrations in specific genes or recurrent copy number changes associated with the occurrence of metastasis (Schumacher et al., 2017; Stoecklein et al., 2008). In a number of cases, diaspora was coincident with an increase in the proportion of signature 3 mutations, associated with failure of DNA double-strand break-repair by homologous recombination. Our findings are in keeping with the failure of DNA repair driving the appearance of genomic heterogeneity. Whether the heterogeneity observed is itself the driver of diaspora or merely a symptom is an important area for future study. Our investigations of the potential drivers of diaspora were limited to genomic factors, and further multi-platform studies looking at epigenetic and transcriptomic factors are other important avenues of future research. We anticipate that analyses of single cells or small clusters from primary sites, disseminated tumor cells and circulating tumor cells will also yield finer resolution of the processes of dissemination and metastasis. In addition, understanding the timing of this event will be of value in planning key events in patient management such as surgery and oncological therapy.

In cancer generally there are currently very few in-depth studies examining the spatial and temporal evolution of metastases as noted in a recent comprehensive study of metastases from multiple primary sites (Robinson et al., 2017). Further studies are required to ascertain whether our diaspora theory also pertains to other cancer types.

Acknowledgements

We would like to thank the following individuals for their help with study set-up, patient liaison and tissue collection, Ben Smith, Nyrai Chinyama, Vijay Sujendran, Peter Safranek, Athanosios Xanthos, Tara Nuckcheddy-Grant, Rachel de la Rue, Sebastian Zeki, Rachael Fels Elliott, Peter Collins, Kitty Puttock, Sophie Rabey and staff at Arthur Rank Hospice and Luke A Wylie for scientific discussion and contribution. We are grateful to Professor Simon Tavaré, FRS for his guidance and support for the esophageal whole genome sequencing project as a part of the International Cancer Genome Consortium (ICGC). We would like to

thank Jo Westmoreland, LMB visual aids for her graphic art expertise. Thanks also go to the Cancer Research UK Cambridge Institute Genomics Core for their technical expertise. Above all, we are indebted to the patients who donated tissue samples to this project, and their families who supported them through it.

Funding

Ayesha Noorani was funded through an MRC Clinical Research Fellowship. The work was funded through the above and an MRC core grant (RG84369) and an NIHR Research Professorship (RG67258) to Rebecca Fitzgerald. Funding for sample sequencing (50X WGS) was through the Oesophageal Cancer Clinical and Molecular Stratification (OCCAMS) Consortium as part of the International Cancer Genome Consortium and was funded by a programme grant from Cancer Research UK (RG66287). All OCCAMS samples which were part of the surgical/endoscopy cohort were obtained from Cambridge patients. We thank the Human Research Tissue Bank, which is supported by the National Institute for Health Research (NIHR) Cambridge Biomedical Research Centre, from Addenbrooke's Hospital. Additional infrastructure support was provided from the CRUK funded Experimental Cancer Medicine Centre in Cambridge. David Wedge is funded by the Li Ka Shing foundation and the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre.

Author Contributions

AN designed the study, implemented the rapid autopsy study, performed the experiments, analyzed data and wrote the manuscript. MG and S.D.P contributed expertise in pathology and sample collection for the rapid autopsy study. ID-B and NG assisted in study implementation, and along with JC, assisted with sample collection at autopsy. M.D.E performed genomic data generation and QC. LB conducted data management. XL, PL-S and JW were involved with autopsy sample collection, advice on experiments and data analysis, and XL contributed to paper writing. LA and IM assisted with data analysis. NG assisted with study Implementation. SMac coordinated the sequencing of samples from the OCCAMS project and contributed to paper writing. SM and AM provided pathology data. TT, SG, LP and DG assisted in implementation and ethical conduct of the autopsy study. R.H.H and AH were involved in surgical sample collection and providing surgical expertise. M.R.S

407 contributed to critical evaluation of the study data and manuscript. D.C.W was responsible
408 for data analysis, paper writing, and assuring integrity of data. The OCCAMS consortium was
409 the vehicle through which the infrastructure and funding was obtained to support the study
410 and the consortium contributed to discussions on the ICGC data and the clinical
411 ramifications. R.C.F provided grant funding and was responsible for study design,
412 supervision of the project, writing the paper and assuring integrity of the data.
413
414 The authors declare no competing interests.

415 **References**

416

417 Ajani, J.A., D'Amico, T.A., Almhanna, K., Bentrem, D.J., Besh, S., Chao, J., Das, P.,
418 Denlinger, C., Fanta, P., Fuchs, C.S., *et al.* (2015). Esophageal and esophagogastric junction
419 cancers, version 1.2015. J Natl Compr Canc Netw *13*, 194-227.
420 Alexandrov, L.B., Jones, P.H., Wedge, D.C., Sale, J.E., Campbell, P.J., Nik-Zainal, S., and
421 Stratton, M.R. (2015). Clock-like mutational processes in human somatic cells. Nat Genet *47*,
422 1402-1407.
423 Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Aparicio, S.A., Behjati, S., Biankin, A.V.,
424 Bignell, G.R., Bolli, N., Borg, A., Borresen-Dale, A.L., *et al.* (2013). Signatures of
425 mutational processes in human cancer. Nature *500*, 415-421.
426 Behjati, S., Gundem, G., Wedge, D.C., Roberts, N.D., Tarpey, P.S., Cooke, S.L., Van Loo,
427 P., Alexandrov, L.B., Ramakrishna, M., Davies, H., *et al.* (2016). Mutational signatures of
428 ionizing radiation in second malignancies. Nat Commun *7*, 12605.
429 Blokzijl, F., de Ligt, J., Jager, M., Sasselli, V., Roerink, S., Sasaki, N., Huch, M., Boymans,
430 S., Kuijk, E., Prins, P., *et al.* (2016). Tissue-specific mutation accumulation in human adult
431 stem cells during life. Nature *538*, 260-264.
432 Bolli, N., Avet-Loiseau, H., Wedge, D.C., Van Loo, P., Alexandrov, L.B., Martincorena, I.,
433 Dawson, K.J., Iorio, F., Nik-Zainal, S., Bignell, G.R., *et al.* (2014). Heterogeneity of genomic
434 evolution and mutational profiles in multiple myeloma. Nat Commun *5*, 2997.
435 Burt, B.M., Groth, S.S., Sada, Y.H., Farjah, F., Cornwell, L., Sugarbaker, D.J., and
436 Massarweh, N.N. (2017). Utility of Adjuvant Chemotherapy After Neoadjuvant
437 Chemoradiation and Esophagectomy for Esophageal Cancer. Ann Surg *266*, 297-304.
438 Chen, X., Schulz-Trieglaff, O., Shaw, R., Barnes, B., Schlesinger, F., Kallberg, M., Cox,
439 A.J., Kruglyak, S., and Saunders, C.T. (2016). Manta: rapid detection of structural variants
440 and indels for germline and cancer sequencing applications. Bioinformatics *32*, 1220-1222.
441 Cunningham, D., Starling, N., Rao, S., Iveson, T., Nicolson, M., Coxon, F., Middleton, G.,
442 Daniel, F., Oates, J., Norman, A.R., *et al.* (2008). Capecitabine and oxaliplatin for advanced
443 esophagogastric cancer. N Engl J Med *358*, 36-46.
444 Davis, A., Gao, R., and Navin, N. (2017). Tumor evolution: Linear, branching, neutral or
445 punctuated? Biochim Biophys Acta *1867*, 151-161.
446 Dentro, S.C., Leshchiner, I., Haase, K., Tarabichi, M., Wintersinger, J., Deshwar, A.G., Yu,
447 K., Rubanova, Y., Macintyre, G., Vazquez-Garcia, I., *et al.* (2018). Portraits of genetic intra-
448 tumour heterogeneity and subclonal selection across cancer types. bioRxiv.
449 Dulak, A.M., Schumacher, S.E., van Lieshout, J., Imamura, Y., Fox, C., Shim, B., Ramos,
450 A.H., Saksena, G., Baca, S.C., Baselga, J., *et al.* (2012). Gastrointestinal adenocarcinomas of
451 the esophagus, stomach, and colon exhibit distinct patterns of genome instability and
452 oncogenesis. Cancer Res *72*, 4383-4393.
453 Dulak, A.M., Stojanov, P., Peng, S., Lawrence, M.S., Fox, C., Stewart, C., Bandla, S.,
454 Imamura, Y., Schumacher, S.E., Shefler, E., *et al.* (2013). Exome and whole-genome
455 sequencing of esophageal adenocarcinoma identifies recurrent driver events and mutational
456 complexity. Nat Genet *45*, 478-486.
457 Foulds, L. (1954). The experimental study of tumor progression: a review. Cancer Res *14*,
458 327-339.
459 Frankell, A.M., Jammula, S., Contino, G., Killcoyne, S.S., Abbas, S., Perner, J., Bower, L.,
460 Devonshire, G., Grehan, N., Mok, J., *et al.* (2018). The landscape of selection in 551
461 Esophageal Adenocarcinomas defines genomic biomarkers for the clinic. bioRxiv.
462 Gabriel, E., Attwood, K., Shah, R., Nurkin, S., Hochwald, S., and Kukar, M. (2017). Novel
463 Calculator to Estimate Overall Survival Benefit from Neoadjuvant Chemoradiation in
464 Patients with Esophageal Adenocarcinoma. J Am Coll Surg *224*, 884-894 e881.

465 Gao, Z., Wyman, M.J., Sella, G., and Przeworski, M. (2016). Interpreting the Dependence of
466 Mutation Rates on Age and Time. PLoS Biol *14*, e1002355.
467 Genomes Project, C., Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin,
468 R.M., Handsaker, R.E., Kang, H.M., Marth, G.T., and McVean, G.A. (2012). An integrated
469 map of genetic variation from 1,092 human genomes. Nature *491*, 56-65.
470 Gerstung, M., Jolly, C., Leshchiner, I., Dentro, S.C., Gonzalez, S., Mitchell, T.J., Rubanova,
471 Y., Anur, P., Rosebrock, D., Yu, K., *et al.* (2017). The evolutionary history of 2,658 cancers.
472 bioRxiv.
473 Greaves, M., and Maley, C.C. (2012). Clonal evolution in cancer. Nature *481*, 306-313.
474 Gundem, G., Van Loo, P., Kremeyer, B., Alexandrov, L.B., Tubio, J.M.C., Papaemmanuil,
475 E., Brewer, D.S., Kallio, H.M.L., Hognas, G., Annala, M., *et al.* (2015). The evolutionary
476 history of lethal metastatic prostate cancer. Nature *520*, 353-357.
477 Japanese Gastric Cancer, A. (2011). Japanese classification of gastric carcinoma: 3rd English
478 edition. Gastric Cancer *14*, 101-112.
479 Jiao, W., Vembu, S., Deshwar, A.G., Stein, L., and Morris, Q. (2014). Inferring clonal
480 evolution of tumors from single nucleotide somatic mutations. BMC Bioinformatics *15*, 35.
481 Kim, M.Y., Oskarsson, T., Acharyya, S., Nguyen, D.X., Zhang, X.H., Norton, L., and
482 Massague, J. (2009). Tumor self-seeding by circulating cancer cells. Cell *139*, 1315-1326.
483 Klein, C.A. (2009). Parallel progression of primary tumours and metastases. Nat Rev Cancer
484 *9*, 302-312.
485 Kuipers, J., Jahn, K., Raphael, B.J., and Beerenwinkel, N. (2017). Single-cell sequencing data
486 reveal widespread recurrence and loss of mutational hits in the life histories of tumors.
487 Genome Res *27*, 1885-1894.
488 Letouze, E., Shinde, J., Renault, V., Couchy, G., Blanc, J.F., Tubacher, E., Bayard, Q., Bacq,
489 D., Meyer, V., Semhoun, J., *et al.* (2017). Mutational signatures reveal the dynamic interplay
490 of risk factors and cellular processes during liver tumorigenesis. Nat Commun *8*, 1315.
491 Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler
492 transform. Bioinformatics *25*, 1754-1760.
493 Liu, D., Abbosh, P., Keliher, D., Reardon, B., Miao, D., Mouw, K., Weiner-Taylor, A.,
494 Wankowicz, S., Han, G., Teo, M.Y., *et al.* (2017). Mutational patterns in chemotherapy
495 resistant muscle-invasive bladder cancer. Nat Commun *8*, 2193.
496 Lodato, M.A., Rodin, R.E., Bohrson, C.L., Coulter, M.E., Barton, A.R., Kwon, M., Sherman,
497 M.A., Vitzthum, C.M., Luquette, L.J., Yandava, C.N., *et al.* (2018). Aging and
498 neurodegeneration are associated with increased mutations in single human neurons. Science
499 *359*, 555-559.
500 Lou, F., Sima, C.S., Adusumilli, P.S., Bains, M.S., Sarkaria, I.S., Rusch, V.W., and Rizk,
501 N.P. (2013). Esophageal cancer recurrence patterns and implications for surveillance. J
502 Thorac Oncol *8*, 1558-1562.
503 Mariette, C., Balon, J.M., Piessen, G., Fabre, S., Van Seuningen, I., and Triboulet, J.P.
504 (2003). Pattern of recurrence following complete resection of esophageal carcinoma and
505 factors predictive of recurrent disease. Cancer *97*, 1616-1623.
506 Martincorena Inigo, R.K.M., Gerstung Moritz, Dawson Kevin J, Haase Kerstin, Van Loo
507 Peter, Davies Helen, Michael R. Stratton Michael R, Campbell Peter J. (2017). Universal
508 Patterns Of Selection In Cancer And Somatic Tissues. Cell.
509 Matsuda, S., Takeuchi, H., Kawakubo, H., and Kitagawa, Y. (2017). Three-field lymph node
510 dissection in esophageal cancer surgery. J Thorac Dis *9*, S731-S740.
511 Murugaesu, N., Wilson, G.A., Birkbak, N.J., Watkins, T., McGranahan, N., Kumar, S.,
512 Abbassi-Ghadi, N., Salm, M., Mitter, R., Horswell, S., *et al.* (2015). Tracking the genomic
513 evolution of esophageal adenocarcinoma through neoadjuvant chemotherapy. Cancer Discov
514 *5*, 821-831.

515 Naxerova, K., Reiter, J.G., Brachtel, E., Lennerz, J.K., van de Wetering, M., Rowan, A., Cai,
516 T., Clevers, H., Swanton, C., Nowak, M.A., *et al.* (2017). Origins of lymphatic and distant
517 metastases in human colorectal cancer. Science *357*, 55-60.
518 Nik-Zainal, S., Van Loo, P., Wedge, D.C., Alexandrov, L.B., Greenman, C.D., Lau, K.W.,
519 Raine, K., Jones, D., Marshall, J., Ramakrishna, M., *et al.* (2012). The life history of 21
520 breast cancers. Cell *149*, 994-1007.
521 Nones, K., Waddell, N., Wayte, N., Patch, A.M., Bailey, P., Newell, F., Holmes, O., Fink,
522 J.L., Quinn, M.C., Tang, Y.H., *et al.* (2014). Genomic catastrophes frequently arise in
523 esophageal adenocarcinoma and drive tumorigenesis. Nat Commun *5*, 5224.
524 Nowell, P.C. (1976). The clonal evolution of tumor cell populations. Science *194*, 23-28.
525 Pasquali, S., Yim, G., Vohra, R.S., Mocellin, S., Nyanhongo, D., Marriott, P., Geh, J.I., and
526 Griffiths, E.A. (2017). Survival After Neoadjuvant and Adjuvant Treatments Compared to
527 Surgery Alone for Resectable Esophageal Carcinoma: A Network Meta-analysis. Ann Surg
528 *265*, 481-491.
529 Pienta, K.J., Robertson, B.A., Coffey, D.S., and Taichman, R.S. (2013). The cancer diaspora:
530 Metastasis beyond the seed and soil hypothesis. Clin Cancer Res *19*, 5849-5855.
531 Robinson, D.R., Wu, Y.M., Lonigro, R.J., Vats, P., Cobain, E., Everett, J., Cao, X., Rabban,
532 E., Kumar-Sinha, C., Raymond, V., *et al.* (2017). Integrative clinical genomics of metastatic
533 cancer. Nature *548*, 297-303.
534 Ross-Innes, C.S., Becq, J., Warren, A., Cheetham, R.K., Northen, H., O'Donovan, M.,
535 Malhotra, S., di Pietro, M., Ivakhno, S., He, M., *et al.* (2015). Whole-genome sequencing
536 provides new insights into the clonal architecture of Barrett's esophagus and esophageal
537 adenocarcinoma. Nat Genet *47*, 1038-1046.
538 Saunders, C.T., Wong, W.S., Swamy, S., Becq, J., Murray, L.J., and Cheetham, R.K. (2012).
539 Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs.
540 Bioinformatics *28*, 1811-1817.
541 Scheinin, I., Sie, D., Bengtsson, H., van de Wiel, M.A., Olshen, A.B., van Thuijl, H.F., van
542 Essen, H.F., Eijk, P.P., Rustenburg, F., Meijer, G.A., *et al.* (2014). DNA copy number
543 analysis of fresh and formalin-fixed specimens by shallow whole-genome sequencing with
544 identification and exclusion of problematic regions in the genome assembly. Genome Res *24*,
545 2022-2032.
546 Schumacher, S., Bartenhagen, C., Hoffmann, M., Will, D., Fischer, J.C., Baldus, S.E., Vay,
547 C., Fluegen, G., Dizdar, L., Vallbohmer, D., *et al.* (2017). Disseminated tumour cells with
548 highly aberrant genomes are linked to poor prognosis in operable oesophageal
549 adenocarcinoma. Br J Cancer *117*, 725-733.
550 Secrier, M., Li, X., de Silva, N., Eldridge, M.D., Contino, G., Bornschein, J., MacRae, S.,
551 Grehan, N., O'Donovan, M., Miremadi, A., *et al.* (2016). Mutational signatures in esophageal
552 adenocarcinoma define etiologically distinct subgroups with therapeutic relevance. Nat Genet
553 *48*, 1131-1141.
554 Sjoquist, K.M., Burmeister, B.H., Smithers, B.M., Zalcberg, J.R., Simes, R.J., Barbour, A.,
555 Gebski, V., and Australasian Gastro-Intestinal Trials, G. (2011). Survival after neoadjuvant
556 chemotherapy or chemoradiotherapy for resectable oesophageal carcinoma: an updated meta-
557 analysis. Lancet Oncol *12*, 681-692.
558 Smyth, E.C., Lagergren, J., Fitzgerald, R.C., Lordick, F., Shah, M.A., Lagergren, P., and
559 Cunningham, D. (2017). Oesophageal cancer. Nat Rev Dis Primers *3*, 17048.
560 Sottoriva, A., Kang, H., Ma, Z., Graham, T.A., Salomon, M.P., Zhao, J., Marjoram, P.,
561 Siegmund, K., Press, M.F., Shibata, D., *et al.* (2015). A Big Bang model of human colorectal
562 tumor growth. Nat Genet *47*, 209-216.
563 Sottoriva, A., Spiteri, I., Piccirillo, S.G., Touloumis, A., Collins, V.P., Marioni, J.C., Curtis,
564 C., Watts, C., and Tavare, S. (2013). Intratumor heterogeneity in human glioblastoma reflects
565 cancer evolutionary dynamics. Proc Natl Acad Sci U S A *110*, 4009-4014.

566 Sporn, M.B. (1996). The war on cancer. Lancet *347*, 1377-1381.
567 Stiles, B.M., Nasar, A., Mirza, F.A., Lee, P.C., Paul, S., Port, J.L., and Altorki, N.K. (2012).
568 Worldwide Oesophageal Cancer Collaboration guidelines for lymphadenectomy predict
569 survival following neoadjuvant therapy. Eur J Cardiothorac Surg *42*, 659-664.
570 Stoecklein, N.H., Hosch, S.B., Bezler, M., Stern, F., Hartmann, C.H., Vay, C., Siegmund, A.,
571 Scheunemann, P., Schurr, P., Knoefel, W.T.*, et al.* (2008). Direct genetic analysis of single
572 disseminated cancer cells for prediction of outcome and therapy selection in esophageal
573 cancer. Cancer Cell *13*, 441-453.
574 Waterman, T.A., Hagen, J.A., Peters, J.H., DeMeester, S.R., Taylor, C.R., and Demeester,
575 T.R. (2004). The prognostic importance of immunohistochemically detected node metastases
576 in resected esophageal adenocarcinoma. Ann Thorac Surg *78*, 1161-1169; discussion 1161-
577 1169.
578 Weaver, J.M., Ross-Innes, C.S., Shannon, N., Lynch, A.G., Forshew, T., Barbera, M.,
579 Murtaza, M., Ong, C.A., Lao-Sirieix, P., Dunning, M.J.*, et al.* (2014). Ordering of mutations
580 in preinvasive disease stages of esophageal carcinogenesis. Nat Genet *46*, 837-843.
581 Yachida, S., Jones, S., Bozic, I., Antal, T., Leary, R., Fu, B., Kamiyama, M., Hruban, R.H.,
582 Eshleman, J.R., Nowak, M.A.*, et al.* (2010). Distant metastasis occurs late during the genetic
583 evolution of pancreatic cancer. Nature *467*, 1114-1117.
584 Yates, L.R., Gerstung, M., Knappskog, S., Desmedt, C., Gundem, G., Van Loo, P., Aas, T.,
585 Alexandrov, L.B., Larsimont, D., Davies, H.*, et al.* (2015). Subclonal diversification of
586 primary breast cancer revealed by multiregion sequencing. Nat Med *21*, 751-759.
587

588 **Figure Legends**

589 **Figure 1. Phylogenetic Analysis of ten cases with nodal and distant metastases**

590 Patient body maps (S for surgical case and P for rapid autopsy) are shown. Green circles
591 denote lymph node metastases and yellow circles distant metastases. The labels within
592 each circle describe the specific location (see Table S4 and S5 for precise anatomical
593 descriptions). An organ is shown in color if metastases were sequenced from that case. The
594 adjacent wedged semi-circle depicts the clinical timelines for each patient. Each wedge
595 corresponds to one month; blue wedges indicate the total lifetime of the patient and red
596 wedges indicate periods of therapy. Phylogenetic trees for each patient are shown and
597 details of how these trees were constructed are provided in Supplementary methods and
598 Supplementary Figures S12A and B; pink = truncal events shared by all samples, purple =
599 branch events shared by more than one sample, yellow = leaves, events unique to a sample.
600 The circle at the end of a trunk, branch or leaf represents a clone or subclone. Each clone
601 or subclone is annotated to show which samples it is present in, where E1-E4 are samples
602 from the primary esophageal tumor, L1-L4 are lymph nodes, and D1-8 are distant
603 metastases - the numbering corresponds to the adjacent body map. A subclone annotated
604 with E1, L2 for example indicates that this subclone is seen only in samples E1 and L2. The
605 precise CCF of each subclone and clone (barring the MRCA) is shown in supplementary
606 Tables S6 and S7. The length of the branches of the tree are reflective of the number of
607 SNVs in the subclone/clone. The scales adjacent to each case are relative, given the
608 variable number of SNVs per case. Trees are annotated with potential driver events, black:
609 missense variants, red: amplifications. Gray dots outlined with a black dashed line denote
610 the first subclone/clone to metastasize that would be classified as non-curative based on
611 anatomical location. Red dots mark the stellate pattern on the phylogenetic tree.

612

613 **Figure 2. 1X WGS and similarity matrix clustering of 250 further tissue samples from six**
614 **cases**

615 1X WGS was performed at an average depth of 1x to track subclones and clones previously
616 discovered using 50X WGS. Pearson correlation similarity matrix clustering was performed
617 on all samples for each case (plotted against each other) with red indicating sample
618 similarity (r=1) and blue indicating dissimilarity (r=-1). Sample sites used in this part of the
619 study are shown in Table S10 and the entire organ is highlighted if solid organ sites_were

620 sequenced. For example, liver metastases were only seen in P4,P6,P8,P10. Similarly, P2
621 had lymph nodes only (only colored dots are seen which represent lymph nodes, no solid
622 organs are highlighted). Clustering was performed based on the presence of subclones
623 and clones already detected using 50X WGS and distinct clusters were identified for each
624 case as demonstrated by the adjacent key per case (each group is both colored and
625 numbered). Samples are displayed on the adjoining body maps for which the color coding
626 corresponds to the genomic clustering in the adjacent heatmap. Sites with multiple
627 samples are magnified and the division of samples shown. Maps of the primary tumor
628 with representation of metastatic subclones are shown with each case, with the colors of
629 the subclones being the same as those in the matrix and body map. Areas shaded red in
630 the primary tumor represent subclones that were not detected in the metastatic samples
631 that underwent 1X WGS and were instead confined to areas of the primary tumor.
632
633 **Figure 3. Temporal tracing of metastatic EAC using multiple lines of evidence**
634 A) Proportion of metastatic subclones present at earlier time-points in archival formalin
635 fixed paraffin embedded (FFPE) samples. The case ID is shown on the Y-axis along with the
636 time-point that the sample was taken, and the % of metastatic subclones present on the
637 X- axis. The n represents the total number of metastatic subclones. B) Mutational
638 signature analysis of ageing signature (signature 1) pre-and post- diaspora in all 8 cases
639 with local and distant spread ($p<1.18e-90$ for all cases). Chi Squared test was used to
640 determine the p value. Survival is shown in months from the point of diagnosis *= cases
641 which underwent surgery. C) Plasma ctDNA 1X WGS and digital droplet PCR (ddPCR)
642 analysis for *TP53* mutant allele fraction (MAF). The MAF of *TP53* (%) is shown on the Y-axis
643 and days from diagnosis are shown on the X-axis. The shaded areas represent time periods
644 of therapy. 1X WGS at select time-points was performed and the clonal composition of
645 these samples is shown as pie-charts. The color of each subclone corresponds to the color
646 of the corresponding node on the adjacent phylogenetic tree.
647
648 **Figure 4. Diaspora model of metastatic spread and associated features.** Panel A depicts
649 clonal diaspora with colored circles representing clones and subclones. *= evidence of
650 selection. Panel B explains the five features seen in diaspora (one is defining, and the
651 other are associated with diaspora) and whether these are present (✓) or absent (x) in

652 each case. *✓ implies that the feature is present, and that the evidence was from 1X
653 WGS.
654

655 ## Methods

656 ### Patients and tumor samples

657 We collected 396 samples from surgery and endoscopy (part of esophageal ICGC) as well as
658 from a rapid autopsy programme called PHOENIX. Patients were eligible if they were at least
659 18 years of age and had received a confirmed diagnosis of EAC following central pathology
660 review. Patients were only approached for the PHOENIX study following a palliative
661 diagnosis, with the full involvement of the multidisciplinary team. All demographic and
662 clinical data was anonymized and stored on a central study database (OpenClinica and
663 Labkey).

664 All samples were collected according to a strict SOP. Post-mortems were completed within 6
665 hours of death to ensure tissue integrity for WGS. The clinical characteristics of the patients
666 are provided in Tables S2 and S3.

667 ### Whole genome sequencing and data analysis

668 We used the Illumina HiSeq platform to perform WGS on multiple regions collected from
669 each primary tumor, lymph node and/or solid organ metastasis (Figure S1A, B, Tables S4 and
670 S5). All DNA extractions and WGS conformed with ICGC quality control standards and
671 required ≥70% cellularity and a matched germline sample. WGS was performed at high
672 depth (median coverage 66.3, IQR 56.1-87.2) to discover mutations in 122 samples from 18
673 patients (Tables S3a and b). In addition, low depth WGS (median coverage 1, IQR 1-5) was
674 performed to track these mutations spatially in up to 48 solid tissue samples per case,
675 (total=250) and 8 ctDNA samples at autopsy. Temporal tracking was performed in cases with
676 archival biopsy material, and where historical bloods were available (Table S12, Figure 3A,
677 C). For each patient the number of subclones and the cancer cell fraction within each
678 subclone was inferred using an extension of a previously described Bayesian Dirichlet
679 process (Nik-Zainal et al., 2012) and we applied a set of previously described rules to derive
680 a phylogenetic tree (Additional Methods; (Jiao et al., 2014). All sequencing data have been

681 deposited in the European Genome-Phenome Archive under accession number
682 EGAD00001003403. TP53 analysis in cell free tumor DNA (ctDNA) was performed using
683 Digital PCR on the Bio-rad platform (Bio-rad, California) using validated TP53 assays (Table
684 S14).

Figure 1

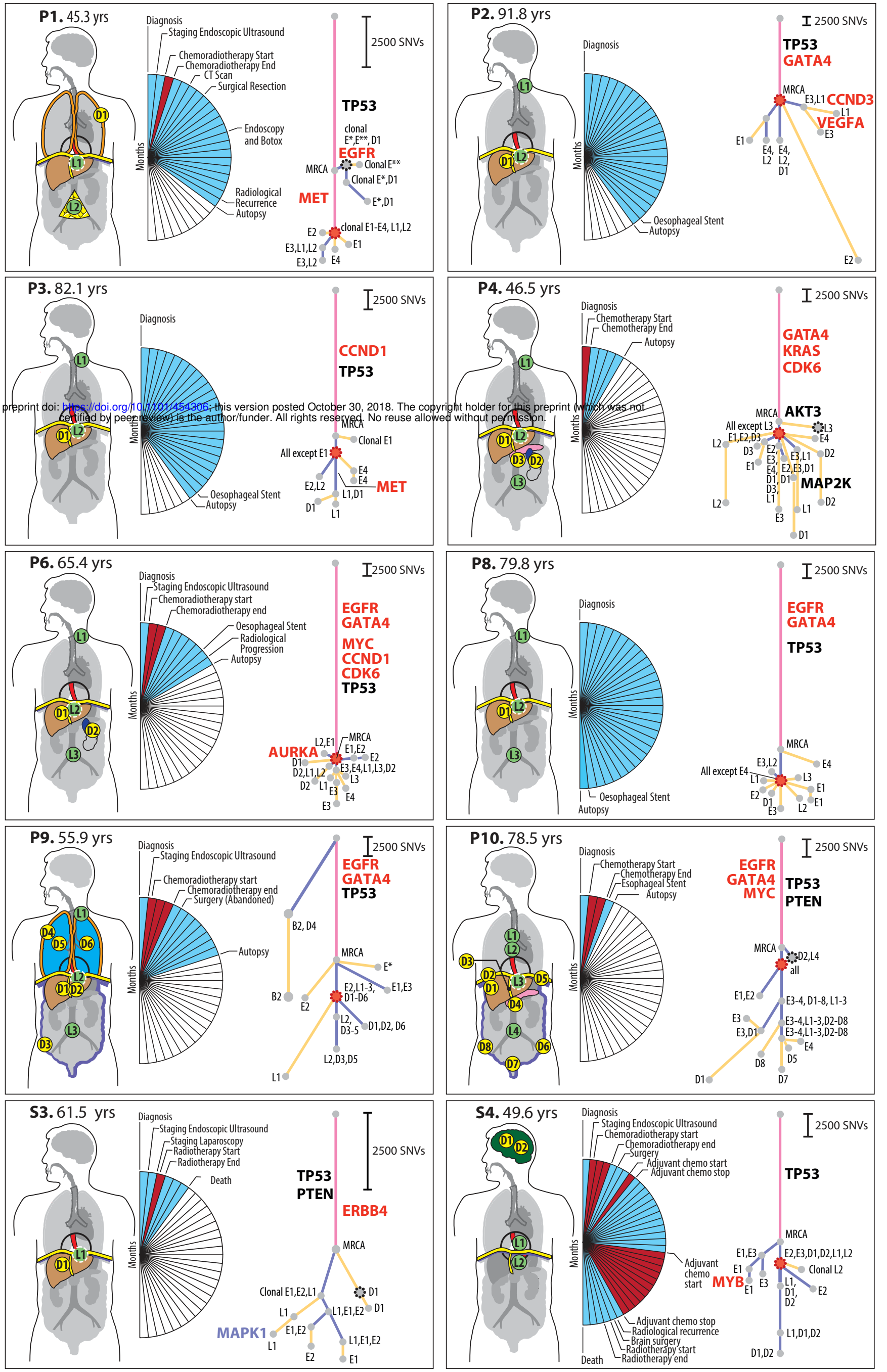


Figure 2

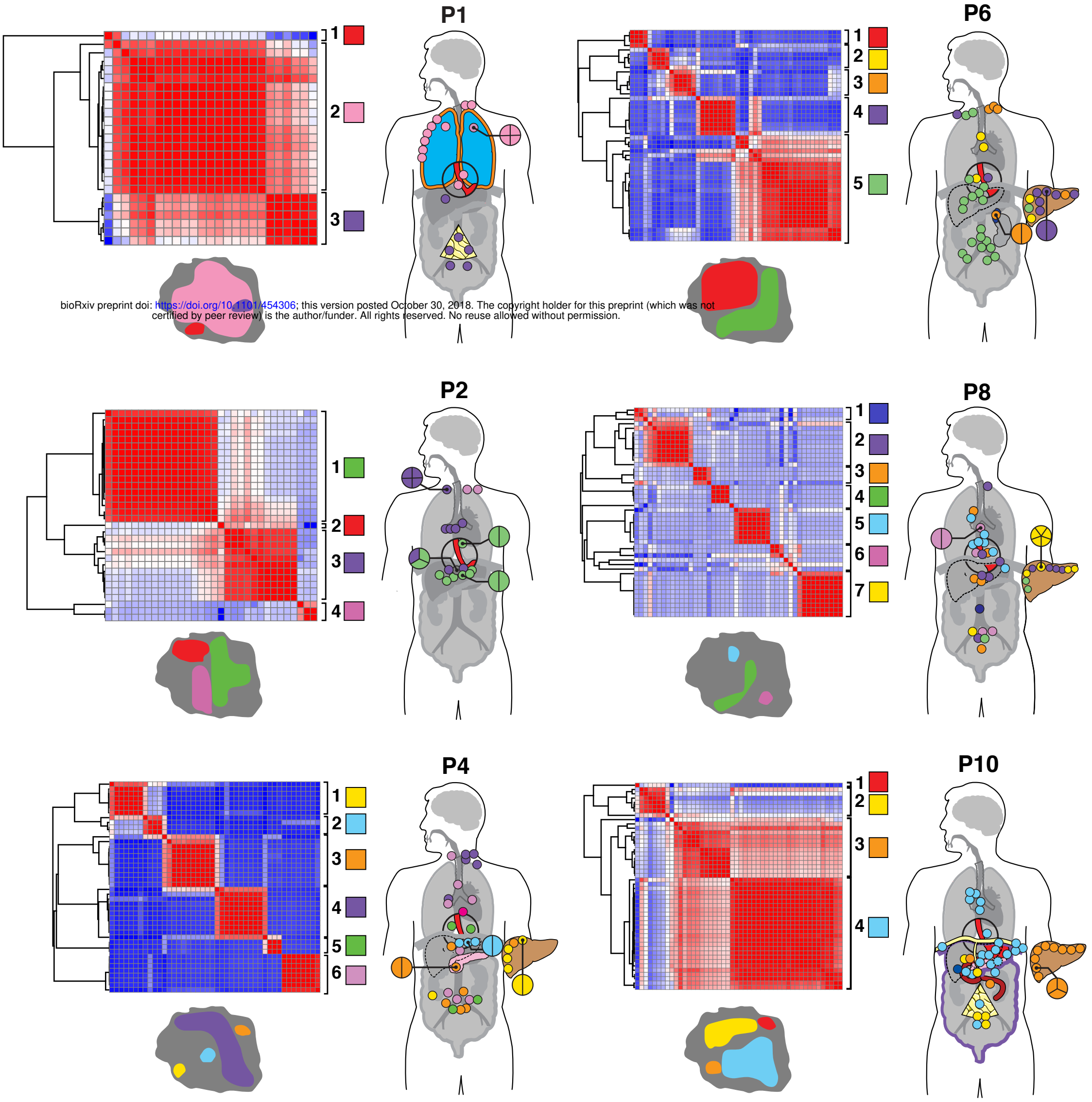
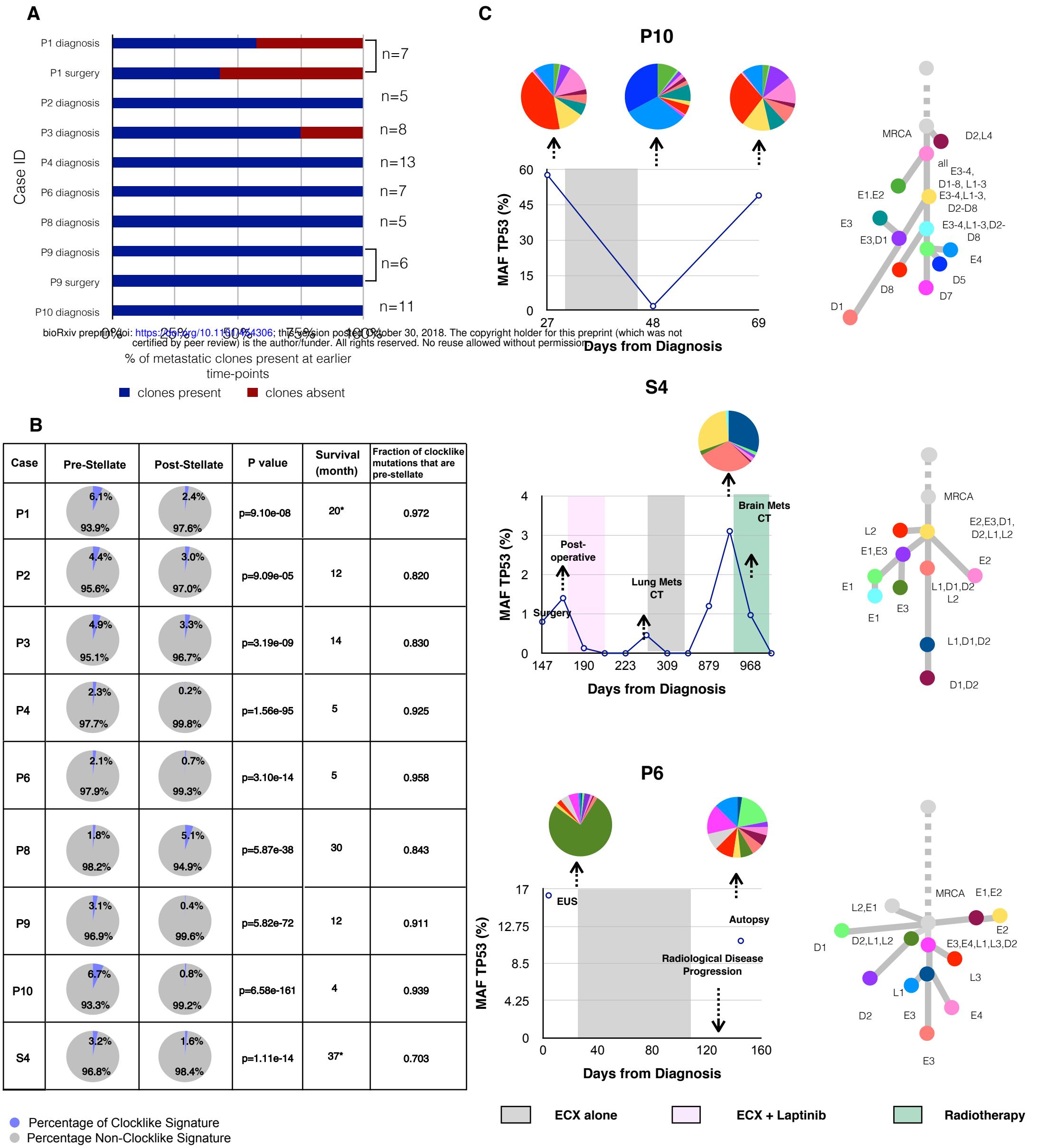


Figure 3



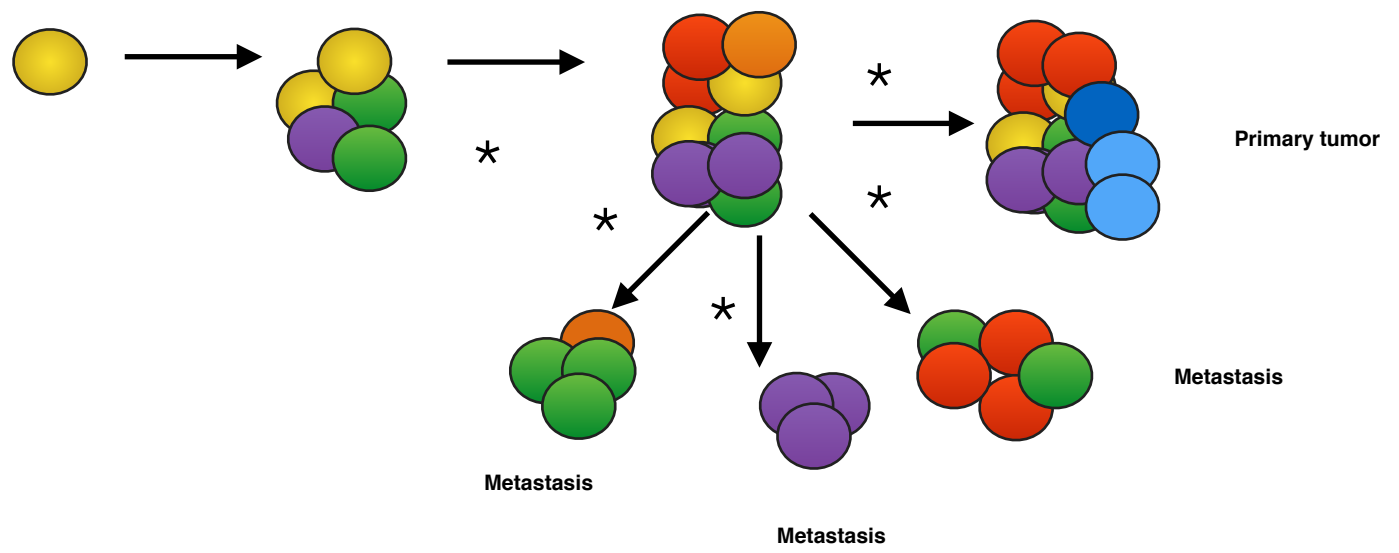
B

Case	Pre-Stellate	Post-Stellate	P value	Survival (month)	Fraction of clocklike mutations that are pre-stellate
P1	6.1% / 93.9%	2.4% / 97.6%	p=9.10e-08	20*	0.972
P2	4.4% / 95.6%	3.0% / 97.0%	p=9.09e-05	12	0.820
P3	4.9% / 95.1%	3.3% / 96.7%	p=3.19e-09	14	0.830
P4	2.3% / 97.7%	0.2% / 99.8%	p=1.56e-95	5	0.925
P6	2.1% / 97.9%	0.7% / 99.3%	p=3.10e-14	5	0.958
P8	1.8% / 98.2%	5.1% / 94.9%	p=5.87e-38	30	0.843
P9	3.1% / 96.9%	0.4% / 99.6%	p=5.82e-72	12	0.911
P10	6.7% / 93.3%	0.8% / 99.2%	p=6.58e-161	4	0.939
S4	3.2% / 96.8%	1.6% / 98.4%	p=1.11e-14	37*	0.703

Percentage of Clocklike Signature
Percentage Non-Clocklike Signature

Figure 4

A



B

Features of Diaspora

	DEFINING	ASSOCIATED			
Case	Multiple subclones from primary spread to multiple metastatic sites	Stellate pattern of three or more subclones derived from the same ancestor found in metastatic sites	Lack of Signature 1 mutations, indicating rapid accumulation of mutations and near-synchronous spread	Spread of at least one subclone to organs of different types, including both lymph nodes and distant organs	Evidence for selection of subclones within the diaspora, indicative of an evolutionary niche (driver amplifications)
P1	✓	×	✓	*✓	×
P2	✓	✓	✓	✓	✓
P3	✓	×	×	✓	✓
P4	✓	✓	✓	✓	✓
P6	✓	✓	✓	✓	✓
P8	✓	✓	×	*✓	×
P9	✓	✓	×	✓	×
P10	✓	×	✓	✓	×
S3	×	×	×	×	✓
S4	✓	✓	✓	✓	✓