Genomic signatures of past and present chromosomal instability in the evolution of Barrett's esophagus to esophageal adenocarcinoma

Authors:

Matthew D. Stachler,[1,2,3,4,#] Chunyang Bao,[1,2,#] Richard W. Tourdot,[2,5,6] Gregory J. Brunette,[5,6] Chip Stewart,[2] Lili Sun,[2,7] Hideo Baba,[8] Masayuki Watanabe,[8] Agoston Agoston,[4] Kunal Jajoo,[9] Jon M. Davison,[10] Katie Nason,[11] Gad Getz,[2] Kenneth K. Wang,[12] Yu Imamura[13], Robert Odze,[4] Adam J. Bass,[1,2,14*] Cheng-Zhong Zhang[2,5,6,7*]

Affiliations:

1. Department of Medical Oncology, Dana-Farber Cancer Institute. Boston MA 02215;

2. Cancer Program, Broad Institute of MIT and Harvard, Cambridge MA 02142;

3. Department of Pathology, University of California, San Francisco. San Francisco CA 94143;

4. Department of Pathology, Brigham and Women's Hospital, Boston MA 02115;

5. Department of Data Science, Dana-Farber Cancer Institute, Boston MA 02215;

6. Department of Biomedical Informatics, Blavatnik Institute of Harvard Medical School, Boston MA 02115;

7. Single-Cell Sequencing Program, Dana-Farber Cancer Institute. Boston MA 02215;

8. Department of Gastroenterological Surgery, Graduate School of Medical Sciences, Kumamoto University, Kumamoto, Japan;

9. Division of Gastroenterology, Department of Internal Medicine, Brigham and Women's Hospital, Boston MA 02115

10. Department of Pathology, University of Pittsburgh School of Medicine, Pittsburgh, PA 15213;

11. Department of Surgery, Baystate Medical Center, University of Massachusetts Medical School, Springfield MA 01107

12. Division of Gastroenterology and Hepatology, Mayo Clinic, Rochester, MN 55905;

13. Department of Gastroenterological Surgery, Cancer Institute Hospital of Japanese Foundation of Cancer Research, Tokyo, Japan

14. Current address: Novartis Institutes of Biomedical Research, Cambridge MA, United States

# These authors contributed equally

Correspondence to: Adam J. Bass (adam.bass@novartis.edu) and Cheng-Zhong Zhang (cheng-zhong_zhang@dfci.harvard.edu)

Conflict of Interest Statement: A.J.B. had received research funding from Bayer, Merck and Novartis, is a consultant to Earli, and HelixNano, a co-founder of Signet Therapeutics, and is now an employee of the Novartis Institutes for Biomedical Research. C.-Z.Zhang co-founded and serves as a scientific advisor to Pillar Biosciences.

Submitted content

Manuscript text, List of references, and Figure Captions

Figures 1-8

Extended Data Figures 1-10

Extended Data Tables 1-8

Methods and captions for Extended Data Figures and Tables

Downloadable Supplementary Data

Haplotype-resolved DNA copy number of bulk BE/EAC samples

Haplotype-resolved DNA copy number of single BE cells from HGD brushing

Total DNA copy number of longitudinal BE sequencing data (re-analysis)

Abstract:

The progression of precancerous lesions to malignancy is often accompanied by increasing complexity of chromosomal alterations but how these alterations arise is poorly understood. Here we performed haplotype-specific analysis of chromosomal copy-number evolution in the progression of Barrett's esophagus (BE) to esophageal adenocarcinoma (EAC) on multiregional whole-genome sequencing data of BE with dysplasia and microscopic EAC foci. We identified distinct patterns of copy-number evolution indicating multigenerational chromosomal instability that is initiated by cell division errors but propagated only after p53 loss. While abnormal mitosis, including whole-genome duplication, underlies chromosomal copy-number changes, segmental alterations display signatures of successive breakage-fusion-bridge cycles and chromothripsis of unstable dicentric chromosomes. Our analysis elucidates how multigenerational chromosomal instability generates copy-number variation in BE cells, precipitates complex alterations including DNA amplifications, and promotes their independent clonal expansion and transformation. In particular, we suggest sloping copy-number variation as a signature of ongoing chromosomal instability that precedes copy-number complexity.

These findings suggest copy-number heterogeneity in advanced cancers originates from chromosomal instability in precancerous cells and such instability may be identified from the presence of sloping copy-number variation in bulk sequencing data.

Large-scale chromosomal rearrangements and copy-number alterations are prevalent in cancer and generally attributed to genomic or chromosomal instability of cancer cells[1-3]. Although much is known about the patterns of genomic rearrangements in fully formed cancers[4,5] and the biological mechanisms of genome instability[6-8], little is understood about what mechanisms are active during cancer evolution and how they generate complex cancer genomes.

Genomic analyses of normal tissues have revealed clonally expanded point mutations but not large structural chromosomal aberrations[9,10]. Early-stage precancerous lesions also show significantly less genome complexity than late-stage dysplasia[11-15] or cancer[4,16,17]. These observations have led to the prevailing view that most chromosomal rearrangements arise late during cancer progression in an episodic manner[18,19], in contrast to the gradual accumulation of short sequence variants (single-nucleotide substitutions or short insertions/deletions)[20,21]. However, the apparently simple genomes of precancerous lesions at the clonal level does not exclude genome instability or complexity at the cellular level. Cells with unstable genomes will generate copy-number variation in the progeny[22,23], but such variation is invisible at the population level due to counterbalancing of random copy-number gains and losses in single cells in the absence of selection (i.e., neutral evolution). Genetic variation is further suppressed by positive selection (e.g., for oncogene amplifications) or negative selection (against large DNA deletions or aneuploidy in general[24]). Based on these considerations, we expect the footprint of genome instability in somatic genome evolution to be most visible in small precancerous lesions with *in situ* clonal expansion of copy-number variation generated by genome instability. This idea has led us to perform multiregional analysis of Barrett's esophagus (BE)[25-27] to dissect the origin of genome complexity in esophageal adenocarcinoma (EAC).

BE is the only known precursor of EAC and estimated to be present in 60-90% of newly diagnosed EAC cases[28]. In contrast to fully formed EACs with complex chromosomal changes[29], BE tissue samples can contain lesions of different histopathological states with varying genomic complexity[30,31]. By analyzing copy-number alterations in concurrent BE (both non-dysplastic and dysplastic) and early EAC (either intramucosal or T1) lesions, we reveal copy-number heterogeneity in BE cells before transformation, relate

44 copy-number evolution patterns in BE cells to those derived from experimental models of chromosomal
45 instability[32-38], and provide mechanistic insight into the evolution of EAC genome complexity.
46 We find that both copy-number heterogeneity and complexity can predate the appearance of
47 cancers or dysplastic lesions and are present in both single BE cells and BE subclones with intact p53.
48 Importantly, p53 loss enables episodic but multigenerational genome evolution initiated by catastrophic
49 events such as whole genome duplication[32,33], chromothripsis[34-36], and dicentric chromosome formation[37,38]:
50 We provide evidence that both copy-number heterogeneity and complex copy-number gains in BE cells
51 reflect multigenerational genome or chromosome instability precipitated by these events. We further
52 demonstrate that ongoing chromosomal instability underlies both progressive DNA deletions in BE cells
53 that result in sloping copy-number variation, and distinct oncogenic amplifications in independently
54 transformed cancers within a single BE field. Together, these findings elucidate how genome instability
55 drives copy-number evolution to promote tumor progression.
56
57 **Results**
58 **Copy-number heterogeneity suggests early onset of chromosomal instability in precancer BE cells**
59 Endoscopic mucosal resection (EMR) is routinely performed in patients with dysplastic BE. In reviewing
60 more than 500 formalin-fixed, paraffin-embedded (FFPE) EMR samples, we identified 14 cases showing
61 unexpected microscopic foci of invasive cancers and one case (Patient 1) with an early cancer removed via
62 esophagectomy. All cancers were either intramucosal or T1 and all samples were collected before treatment.
63 Following independent pathologic re-review by two or more pathologists to confirm the diagnoses
64 (**Methods**), we delineated and performed laser capture microdissection (LCM) to isolate regions
65 corresponding to distinct histopathological states[27] (**Figure 1**), including non-intestinalized columnar
66 metaplasia (COLME), non-dysplastic BE (NDBE), BE indefinite for dysplasia (IND), BE with low-grade
67 dysplasia (LGD) or high-grade dysplasia (HGD), and intramucosal (IMEAC) or early EAC (**Extended**
68 **Data Figure 1)**. We further isolated normal tissue from benign FFPE regions that was used as germline
69 reference.

70 Due to the limited quantity of FFPE DNA from small tissue sections and their lesser quality
71 compared to DNA from fresh or frozen cells, we first performed low-pass whole-genome sequencing
72 (WGS) at ~0.1x mean depth to select libraries with sufficient complexity and then performed deeper
73 sequencing ~20x. The final cohort consisted of 75 BE/EAC (21 COLME/NDBE/IND, 7 LGD, 23 HGD,
74 and 24 IM/EAC) and 15 reference samples from 15 patients (**Extended Data Table 1**). the variant calls
75 generated by standard tools had both high false positive and high false negative detection rates (**Methods).**
76 For single-nucleotide variants (both somatic and germline), short insertions/deletions, and rearrangements,
77 we performed joint variant detection on all samples from each patient to improve variant detection accuracy
78 (**Figure 1**). Although the joint analysis is sufficient to detect mutations shared by multiple samples, the
79 false negative detection of mutations in individual samples due to sequencing dropout still confounds
80 phylogenetic inference (**Methods**). To bypass this challenge, we focused on somatic copy-number
81 alterations (SCNA) for which better accuracy could be achieved.

82 We determined chromosome-specific DNA copy number and copy-number changepoints based on
83 haplotype-specific sequence coverage (**Methods, Supplementary Data**). Parental haplotypes were first
84 inferred by statistical phasing using a reference haplotype panel[39] and then refined based on allelic
85 imbalance across all samples from each patient. We used haplotype-specific sequence coverage to first
86 validate the estimated ploidies and clonal fractions of aneuploid BE/EAC clones and then calculate the
87 integer DNA copy number of parental chromosomes. The determination of long-range parental haplotype
88 both enabled phasing of SCNAs to parental chromosomes and ensured the accuracy of SCNA detection.
89 We further performed segmentation of haplotype-specific DNA copy number and used copy-number
90 changepoints to refine the list of rearrangements. For data presentation clarity, the copy-number plots in
91 the main and extended data figures only show data of the altered homolog, except where stated. The
92 haplotype-specific sequence coverage and copy number of both homologs are provided in **Supplementary**
93 **Data**.

94 We determined the phylogenetic tree of samples from each patient (**Figure 2**) based on haplotype-
95 specific copy-number alterations. SCNAs were first identified independently in each sample and then

96 assigned to phylogenetic branches based on their presence or absence in all samples. The branch length
97 (horizontal distance between nodes) approximately reflects the SCNA burden estimated using the number
98 of altered chromosomes. SCNAs on each branch (labelled in **Extended Data Figure 2)** are summarized in
99 **Extended Data Table 2**; SCNAs that affect esophageal cancer genes or identified more than once in the
100 current cohort are annotated in **Figure 2**. In all but two patients (13 and 14), we identified SCNAs in related
101 BE/EAC genomes affecting a single parental homolog but having distinct changepoints that indicate
102 branching evolution of ancestral chromosomes; these chromosomes are labelled with asterisks near the
103 inferred common ancestor. Whole-genome duplication (WGD) was inferred based on the number of
104 homologous chromosomes with more than one copy[40] and assigned to evolutionary branches based on the
105 WGD status of individual samples (**Methods**). For SCNAs on branches with WGD, their timing relative to
106 WGD was inferred based on the integer copy-number states. Finally, we confirmed the consistency between
107 SCNA-derived phylogenetic trees and genetic similarities estimated from somatic SNVs (**Extended Data**
108 **Figure 2**). The few instances of discrepancy are discussed in **Methods**.

109 The phylogenetic trees of EAC and precursor BE lesions show several recurrent patterns. First, bi-
110 allelic *TP53* inactivation is a truncal event of the evolutionary branches of cancer or high-grade BE lesions
111 (14/15 patients). By contrast, focal deletion near *FHIT* (a common fragile site) is often ancestral to all BE
112 and EAC lesions; bi-allelic inactivation of *CDKN2A* (a frequently inactivated tumor suppressor) can be
113 truncal to either cancer/HGD lesions (Patient 3,5,6,7) or NDBE/LGD lesions (Patient 2,8,9,11,14). Second,
114 evolutionary branches with the highest SCNA burdens are frequently associated with WGD, which is itself
115 also a frequent event (10/15 patients). Third, high-grade dysplastic BE lesions and cancer lesions from the
116 same patient often harbor distinct SCNA breakpoints on single parental chromosomes (13/15 patients) or
117 distinct regions of focal amplification (10/15 patients), indicating copy-number heterogeneity prior to the
118 emergence of aneuploid BE/EAC clones. Finally, we identified more than one early cancer lesion in five
119 patients (Patient 1,2,9,12,15): The distinct cancer foci from each patient often displayed significant genomic
120 divergence but were individually accompanied by precancerous lesions in close proximity (Patient
121 1,9,12,15) and/or showing more genomic similarity (Patients 2,9,12,15). The last observation strongly

122 suggests that the cancer foci had evolved independently from distinct BE cells within the same BE field,
123 i.e., independent malignant transformation.

124 The observation of significant SCNA diversity in BE and EAC subclones suggests highly dynamic
125 copy-number evolution in precancerous BE cells and predicts copy-number diversity at the single-cell level.
126 We directly tested this hypothesis by performing whole-genome sequencing analysis of 68 single cells
127 isolated from a patient with known HGD by endoscopic cytology brushing immediately before
128 radiofrequency ablation. We performed haplotype-specific copy-number analysis and phylogenetic
129 inference using the same strategy as for bulk samples (**Methods**). We identified 12 cells with aneuploid
130 genomes and 56 cells with near diploid genomes. Their phylogeny and selected examples of SCNAs in
131 single BE cells or subclones are shown in **Figure 3;** SCNAs in each cell are listed in **Extended Data Table**
132 **3** and DNA copy-number plots of all cells are available in **Supplementary Data**. All the aneuploid cells
133 share biallelic *TP53* inactivation through a pathogenic R175H mutation and loss-of-heterozygosity
134 generated by 17p loss, but show significant heterogeneity of chromosomal copy-number changes. The onset
135 of genomic heterogeneity in precancer BE cells following bi-allelic *TP53* inactivation recapitulates the
136 pattern seen in bulk samples and provides direct evidence of dynamic precancer genome evolution driven
137 by chromosomal instability. We next discuss specific patterns of copy-number evolution and their
138 mechanistic implications.

139

140 ***TP53* inactivation and the onset of genome instability initiates BE genome evolution**

141 We observed increasing SCNA burden with disease progression (**Figure 4A**,left; **Extended Data Fig. 3A**
142 **and 3B**), but this correlation is mostly attributed to *TP53* mutation status. Samples with *TP53* inactivation
143 show significantly higher SCNA burdens than samples without *TP53* inactivation (**Figure 4A**,middle;
144 **Extended Data Fig. 3C**). In particular, two NDBE samples (from Patient 6 and 15) and four LGD samples
145 (from Patient 6 and 7) with bi-allelic *TP53* inactivation show similar SCNA burdens as HGD and EAC
146 samples; by contrast, NDBE and LGD samples without *TP53* inactivation show significantly fewer SCNAs
147 (**Extended Data Fig. 3A**). These data and the contrasting SCNA burdens in single BE cells with and

148 without intact p53 (**Figure 3A**) both reinforce the association between p53 loss and SCNA evolution[11,31].

149 Prior analyses of ageing esophageal tissues[9,10] by bulk sequencing revealed uniparental disomy
150 (UPD), or copy-neutral loss of heterozygosity, as the only large segmental SCNA. Consistent with this
151 observation, we observed frequent UPDs in both single BE cells (**Extended Data Table 3**) and clones
152 (**Extended Data Table 4**) prior to p53 loss, but only sporadic segmental gains or losses in single BE cells
153 (**Figure 3C,D**) and almost none in BE clones. Remarkably, we identified UPDs on the 9p terminus with
154 varying boundaries in a subclone of 14 single BE cells (**Figure 3E** and **Supplementary Data**). As this
155 variation does not alter total DNA copy number, it can only be revealed by haplotype-resolved copy-number
156 analysis. The varying boundaries of terminal UPD in different cells (arrows in **Figure 3E**) bear an intriguing
157 similarity to our prior observation of varying terminal deletions attributed to ongoing breakage-fusion-
158 bridge cycles[38] (see **Extended Data Figure 7** that will be discussed later). The similarity between varying
159 terminal UPDs and varying terminal deletions suggests a plausible common origin from broken
160 chromosomes generated by breakage-fusion-bridge cycles[23], with deletions resulting from translocations
161 involving other broken ends and UPDs resulting from homology-dependent invasion of broken ends into
162 the intact homolog followed by a half crossover resolution[41] (**Extended Data Fig. 4**, top).

163 In contrast to the simple SCNA landscape in BE cells with intact p53 is the prevalence of arm-level
164 and complex SCNAs in BE cells and clones after p53 loss. Loss of p53 does not directly cause aneuploidy
165 or chromosomal instability in human cells[42], but abolishes p53-dependent arrest after DNA damage[43] or
166 prolonged mitosis[44]. The burst of SCNA complexity after p53 loss is therefore more reflective of an
167 increased frequency of SCNA clonal expansion than an increased rate of SCNA acquisition. Moreover, the
168 observation of sporadic large SCNAs, especially UPDs, in single BE cells with intact p53 indicates that BE
169 cells do acquire DNA breaks, but these breaks do not lead to complex copy-number alterations as seen in
170 BE cells or clones with inactive p53. We next focus on BE cells or clones with inactive p53 and provide
171 evidence supporting that the accumulation of SCNA complexity reflects multigenerational chromosomal
172 instability that is precipitated by sporadic cell division errors but only propagated after p53 inactivation.
173

174 **Whole-genome duplication triggers rapid accumulation of arm-level copy-number changes**

175 The most dramatic change in BE cells is whole-genome duplication (WGD). WGD is inferred to be a
176 frequent event in many epithelial cancers[45,46] and thought to define a particular EAC evolution trajectory[31].
177 We inferred 15 WGD events in bulk BE/EAC lesions from 10/15 patients, including independent WGD
178 occurrences in distinct HGD/EACs from Patient 1,3, and 4 (**Figure 2**). We further inferred two independent
179 WGDs in single BE cells without presence of cancer (**Figure 3A**). These observations suggest that WGD
180 may occur frequently during BE progression before the appearance of cancer.

181 Despite the prevalence of WGD in human cancers[45,46] and its tumor-promoting capacity[47,48], how
182 WGD impacts tumorigenesis remains incompletely understood. One proposal is that tetraploidization (the
183 event that causes WGD) can precipitate additional genome instability including multipolar cell division or
184 chromosome missegregation[6,32,33] that leads to aneuploidy. Consistent with this model, we inferred that
185 more SCNAs in BE/EAC genomes were acquired after WGD than before WGD (**Figure 4A**, right), and
186 evolution branches with WGD acquisition had significantly higher SCNA burdens (30 events/branch) than
187 non-WGD branches (pre-WGD: 7.5/branch; post-WGD: 8.8/branch) (**Figure 4B, Extended Data Table**
188 **2**). Moreover, a majority of post-WGD SCNAs are arm-level changes (302 out of 428 events) and
189 dominated by losses (256) (**Figure 4C**), a pattern also seen in single aneuploid BE cells (**Figure 3A**).

190 The preponderance of chromosome losses after WGD has two implications. First, this pattern
191 cannot be solely explained by increased rates of random chromosome missegregation[32] that generates
192 reciprocal gain and loss in a pair of daughter cells. This pattern could reflect a lower fitness of cells with
193 larger chromosome number due to more frequent mitotic delays and defects[46]. It could arise from multipolar
194 cell divisions that generate three or more progeny cells with predominantly chromosome losses[33]
195 (**Extended Data Fig. 5A**). Future work is needed to test these hypotheses. Second, extensive chromosome
196 losses after WGD may significantly reduce the number of duplicated chromosomes and cause
197 underestimation of WGD incidence in cancer development, especially in cancers with highly aneuploid
198 genomes. Together, our analysis of arm-level SCNAs in BE cells both confirms WGD as a precursor to

199 aneuploidy[49-51] and highlights the diversity of copy-number outcomes[5] generated by post-WGD events
200 including multipolar cell division[33].
201
202 **Segmental copy-number alterations display signatures of dicentric chromosome evolution**
203 In contrast to the prevalence of post-WGD arm-level SCNAs, we inferred a similar number of segmental
204 SCNAs in BE/EAC genomes to have occurred prior to (135) and after WGD (126) in samples with WGD
205 acquisition. The fractions of segmental DNA loss and DNA gain are also comparable among pre-, post-,
206 and WGD branches (**Figure 4C**, right), although branches with WGD acquisition have a higher average
207 SCNA burden (5.9 events) than pre- (1.6) or post-WGD (2.1) branches. These observations indicate that
208 segmental SCNA acquisition is promoted by WGD but also occurs independent of WGD.
209 Segmental SCNAs in BE genomes further display two features of non-randomness. First, SCNA
210 breakpoints are often concentrated on a few chromosomes with complex deletions (chromothripsis) or
211 duplications. Second, distinct SCNAs in related BE/EAC genomes more frequently originate from a single
212 parental chromosome ('mono-allelic') than affect both parental chromosomes ('bi-allelic') (**Figure 4D** and
213 **Extended Data Fig. 3E**). Both features are consistent with one-off or successive SCNA acquisition on
214 individual unstable chromosomes instead of independent SCNA acquisition across the genome. The
215 connection between segmental SCNA acquisition and chromosomal instability is further supported by the
216 observation of larger fractions of deletions (allelic copy number = 0) or duplications (allelic copy number
217 ≥2 in non-WGD samples and ≥3 in WGD samples) in samples with inactive p53 than in samples with intact
218 p53 (**Figure 4E**). Finally, we recognized that many segmental SCNA patterns in BE/EAC genomes are
219 consistent with the outcomes of chromosomal instability from abnormal nuclear structures including
220 micronuclei[34] (**Extended Data Fig. 5B**) and chromosome bridges (**Extended Data Fig. 5C**)[38]. We sought
221 to use the genomic signatures of *in vitro* chromosomal instability to deconvolute segmental copy-number
222 complexity in BE/EAC genomes.
223 The most frequent SCNAs in BE/EAC genomes are gain or loss of large terminal (i.e., spanning a
224 telomere) or internal (with two non-telomeric breakpoints) segments; these alterations are consistent with

225 the outcomes of dicentric chromosome breakage (**Figure 5**). Dicentric chromosomes can result from either
226 end-to-end chromosome fusion or incomplete decatenation of sister chromatids[38] and lead to a 'bridge'
227 between daughter nuclei when the two centromeres segregate to different daughter nuclei. Although
228 dicentric chromosomes can be generated by a variety of mechanisms, the genomic consequences are
229 primarily determined by the formation and breakage of chromosome bridges[37,38]. Breakage of a single
230 dicentric chromosome ('chromatid-type' bridges) will generate reciprocal gain and loss of a telomeric
231 segment ('terminal' SCNAs) (**Figure 5A**). If both sister dicentric chromatids are part of the bridge
232 ('chromosome-type' bridges), their breakage can give rise to large segmental gain or loss within a
233 chromosome arm, hereafter referred to as 'paracentric' SCNAs (**Figure 5B**). Both of these outcomes were
234 directly demonstrated in single-cell experiments[38] but originally described by McClintock (summarized in
235 Ref.[52]) We further observed large SCNAs spanning centromeres ('pericentric' SCNAs) that can result from
236 broken ring chromosomes (**Figure 5C**, first described by McClintock in Ref.[53]) or multicentric
237 chromosomes. The instances of terminal and large internal SCNAs in our BE/EAC cohort are summarized
238 in **Figure 5D** and listed in **Extended Data Table 5:Tab 1**. In total, these events account for ~50% of
239 segmental SCNAs.

240 Although chromosome bridge resolution provides a simple mechanism for single-copy gain or loss
241 of large segments, similar copy-number outcomes may be generated by other processes. For example,
242 terminal deletion or duplication could result from simple chromosomal translocations followed by whole-
243 chromosome losses or gains (**Extended Data Fig. 6A**). This model, however, produces an equal number
244 of terminal gains (including retentions) and losses, and cannot explain the disparity between terminal gains
245 and losses seen in most samples (**Extended Data Fig. 6B**). Moreover, as broken bridge chromosomes can
246 form new dicentrics and undergo breakage-fusion-bridge (BFB) cycles that generate a variety of compound
247 copy-number outcomes, the identification of these compound copy-number patterns in BE/EAC genomes
248 provides stronger evidence of chromosome bridges being involved in BE copy-number evolution.

249 The most common outcome of BFB cycles is the presence of DNA duplications near the
250 boundaries of large segmental deletions (**Figure 6A,B**) or large segmental gains. Instances of these patterns

251 in BE/EAC genomes are listed in **Extended Data Table 5:Tab 2** and also summarized in **Figure 5D**. The
252 identification of interchromosomal rearrangements between both simple and compound SCNA breakpoints
253 (**Figure 6A,B** and **Extended Data Figure 6C,D**) also suggests that these broken ends were generated
254 simultaneously, most likely from the resolution of multichromosomal bridges as seen in experimental
255 models of telomere crisis[37] or chromosome bridge resolution[38].

256 Successive DNA duplications at the broken ends of chromosomes can generate focal amplifications
257 (**Figure 6C**, top). Remarkably, the amplification on 7q in IMEAC shares a common SCNA boundary with
258 the terminal deletion in HGD. (The same pattern of reciprocal DNA retention and loss is also seen in 17q
259 of these two clones.) This pattern of reciprocal DNA retention and deletion directly recapitulates the
260 outcome of broken bridge chromosomes between daughter nuclei (**Figure 5A**) that is only visible by
261 multiregional sequencing. Based on this observation, we inferred that the HGD and the IMEAC clones were
262 independently derived from sibling cells each having inherited a broken piece of a dicentric Chr.7 with
263 amplified DNA that was present in their common ancestor.

264 Besides DNA duplications at broken termini, BFB cycles can also generate progressive DNA losses
265 from either sequential breakage or deficient replication of bridge chromatin[38]. As each new deletion erases
266 the boundary of preceding deletions, progressive DNA losses can only be revealed in different progeny
267 clones (**Extended Data Figure 7**) but not in a single clone. We observed 11 instances of terminal or
268 paracentric SCNAs with distinct breakpoints in different BE/EAC lesions from the same patient that are
269 consistent with progressive DNA losses (**Extended Data Table 6:Tab1**). One example of varying 4q-
270 terminal losses (boundaries marked by black arrows) in five lesions from Patient 2 is shown in **Figure 6D**.

271 In summary, we identified frequent duplications or deletions of large terminal, paracentric, and
272 pericentric segments in BE genomes and attributed them to the formation and breakage of dicentric
273 chromosomes (**Figure 5**). This mechanistic association is further supported by the observation of (1)
274 additional duplications or progressive DNA losses at SCNA boundaries (**Figure 6**) reflecting successive
275 BFB cycles (**Extended Data Fig. 7**); and (2) interchromosomal translocations between SCNA boundaries
276 indicating simultaneous generation of broken chromosome ends. In particular, the observation of reciprocal

277 DNA loss and gain in distinct BE/EAC clones from the same patient that directly recapitulate the outcome
278 of dicentric bridge resolution between daughter cells (**Figure 6C**) provides the most compelling evidence
279 of BFB cycles during BE evolution.
280
281 **Contemporaneous chromothripsis and BFB cycles generate EAC copy-number complexity**
282 Besides simple DNA loss and gain, dicentric chromosomes can also undergo DNA fragmentation[37,38] either
283 from chromosome bridge resolution or in micronuclei from chromosome missegregation. These processes
284 generate chromothripsis with distinct oscillating DNA copy number patterns. For chromothripsis from
285 bridge resolution, fragmentation of the bridge chromatin creates oscillating copy number in a fraction of
286 the chromosome arm that was in the bridge, and the region with oscillating copy number is usually adjacent
287 to the boundaries of large terminal or internal SCNAs corresponding to termini of broken bridge
288 chromosomes (**Extended Data Fig. 8A**). We inferred that 35 instances of chromothripsis were consistent
289 with this pattern (**Extended Data Table 7:Tab1**, 'direct' in Column N) and show representative examples
290 in **Extended Data Fig. 8B-D**. For chromothripsis resulting from fragmentation of dicentric chromosomes
291 partitioned into micronuclei, the oscillating copy-number pattern should span whole chromosome arms
292 ("chromosome/arm") (**Extended Data Fig. 8E**). We inferred that 25 instances of chromothripsis were
293 consistent with this evolution sequence (**Extended Data Table 7:Tab1**, 'downstream' in Column N). The
294 second scenario is best demonstrated in the example shown in **Extended Data Fig. 8F**, where the three-
295 state oscillating copy-number pattern (CN=0,1,2) spanning both Chr.17q and 18p together with inter-
296 chromosomal rearrangements indicated chromothripsis of a dicentric translocated chromosome. We
297 additionally identified 40 instances of chromothripsis spanning entire chromosomes or arms that are
298 consistent with micronucleation and 7 instances of regional chromothripsis without a clear relationship to
299 large terminal/internal SCNAs.
300 We further analyzed DNA rearrangements related to chromothripsis but restricted this analysis to
301 ancestral chromothripsis shared by three or more samples for which joint rearrangement detection can
302 achieve good accuracy (see **Methods**). We identified two examples of chromothripsis involving sub-

303 chromosomal regions (including arms) from multiple chromosomes (**Extended Data Fig. 8F,G**) that are
304 consistent with multichromosomal bridge resolution. In two instances of chromothripsis, we further
305 identified clustered rearrangement breakpoints near single SCNA boundaries (**Extended Data Fig. 8D,G**)
306 that resemble the tandem-short-templates rearrangement pattern observed in chromothripsis from bridge
307 resolution[38] and micronucleation[34]. These rearrangement patterns provide additional evidence supporting
308 the connection between chromothripsis and chromosomal bridges or subsequent micronuclei.

309 The comparison of SCNAs in related BE/EAC genomes provides further evidence for BFB cycles
310 in BE genome evolution. In the example shown in **Figure 7A**, the ancestral paracentric deletion shared by
311 all three genomes (LGD2/HGD3/EAC) was followed by regional chromothripsis and BFB amplifications
312 near the centromeric break end in the LGD2 clone and a terminal duplication near the telomeric break end
313 in the EAC clone; both downstream alterations likely arose from secondary BFB cycles after the ancestral
314 paracentric deletion. In the example shown in **Figure 7B**, the (mostly) non-overlapping segments retained
315 by the HGD and IMEAC genomes is consistent with a random distribution of DNA fragments from a single
316 micronuclear chromosome into a pair of daughter cells[34]. Other examples of chromothripsis as one of the
317 branching outcomes of BFB cycles are listed in **Extended Data Table 6** and **Figure 6E**.

318 The examples in **Figure 7A** and **7B** illustrate how copy-number breakpoints with either identical
319 (**Figure 7A**, dotted line) or complementary (**Figure 7B**, dashed lines) DNA retention and loss in related
320 genomes can inform about the evolutionary sequence of the observed copy-number alterations. This is
321 further demonstrated in the Chr5 example in **Figure 7C**. The shared copy-number breakpoint (dotted line)
322 with complementary DNA retention and deletion in IMEAC2 and EAC1 indicates a reciprocal distribution
323 of broken chromosome fragments into their ancestors; the paracentric loss in IMEAC2 further suggests a
324 chromosome-type bridge breakage event (**Figure 5B**). Therefore, the chromothripsis alteration with three
325 oscillating copy-number states in EAC1 must have arisen downstream of the ancestral breakage event.

326 The combination of chromothripsis and successive DNA duplications in BFB cycles can explain
327 complex segmental gains and amplifications. Whereas simple BFB cycles generate duplications flanked by
328 large segmental deletions (**Figure 5D, Extended Data Fig. 8H**), BFB cycles following chromothripsis

329 generate segmental gains or amplifications with interspersed DNA deletions (**Extended Data Fig. 8I**).
330 Several copy-number patterns in Patient 1 indicate contemporaneous chromothripsis and BFB
331 amplifications (**Figure 7C**). On both Chr.1p and Chr.16p, the oscillation between DNA deletion and
332 amplification in EAC1 suggests an evolution sequence of ancestral chromothripsis followed by downstream
333 BFB amplifications; the same regions in IMEAC2 display terminal duplications (Chr.1p) and a simple
334 terminal deletion (Chr.16p). The presence of a shared copy-number breakpoint on Chr.1p and a common
335 region of terminal deletion on Chr.16p between the EAC1 and IMEAC2 genomes suggests that the distinct
336 copy-number patterns reflect divergent evolutionary outcomes of a single ancestral broken chromosome.
337 Interestingly, the amplified regions on 16p in the EAC1 genome do not contain known oncogenes but are
338 co-amplified with a region on 18q containing *GATA6*, a recurrently amplified EAC oncogene. By contrast,
339 the IMEAC2 genome harbors neither amplification but has more amplified *GATA4* on Chr.8p. Moreover,
340 the shared boundaries of amplified regions on 8p in both EAC1 and IMEAC2 indicates that the *GATA4*
341 amplification was ancestral to both genomes but underwent different downstream evolution. The distinct
342 *GATA4* and *GATA6* amplifications in these two genomes, likely reflective of positive selection for their
343 combined expression[54], highlights how persistent chromosomal instability rapidly generates copy-number
344 heterogeneity and fuels the acquisition of oncogenic amplifications.

345 As DNA amplification is only one out of many possible outcomes of multigenerational copy-
346 number evolution (we operationally defined focally amplified regions to have allelic copy number ≥ 8 that
347 can be attained with at least three rounds of duplications), clonally fixated amplifications are likely
348 reflective of positive selection and expected to contain oncogenes. Among 45 focally amplified regions
349 each spanning one or multiple loci on a chromosome (**Extended Data Table 7:Tab2**), 24 encompass
350 putative oncogenes and 29 overlap with regions that are recurrently amplified in cancer. The significance
351 of focal amplification as a mechanism of oncogenic activation during EAC transformation[30,31] is further
352 supported by the observation of both recurrent amplifications of EAC oncogenes, including *ERBB2* on 17q
353 (5/15 patients) (**Extended Data Fig. 8H,I**) and *GATA6* on 18q (4/15 patients), and sporadic oncogene
354 amplifications that are exclusive to cancer lesions but not their precursors, including *IGF1R* (Patient 3),

355 *MET* (Patient 4), and *KRAS* (Patient 10).

356 In summary, we found that many complex segmental copy-number alterations in BE/EAC
357 genomes, including focal amplifications, can be deconvoluted into different evolution sequences of
358 sequence duplications generated by BFB cycles and chromothripsis from DNA fragmentation (**Figure 7D**).
359 Together with observations of terminal/internal SCNAs reflecting simple copy-number outcomes of BFB
360 cycles, these data provide *in vivo* evidence for the involvement of abnormal nuclear structures including
361 micronuclei[34-36] and chromosome bridges[37,38] in the generation of EAC genome complexity.

362

363 **Chromosomal instability generates continuous copy-number variation prior to discrete changes**

364 Our analysis of BE/EAC genomes reveals both copy-number complexity and copy-number heterogeneity
365 in BE subclones that indicate multigenerational evolution of unstable chromosomes. Importantly,
366 chromosomal instability first generates copy-number variation in single BE cells. We wondered whether
367 such instability in single BE cells can be discerned prior to copy-number heterogeneity or complexity in
368 BE subclones.

369 If chromosome breakage only generates reciprocal DNA retention and loss between sibling cells,
370 such changes are not visible at the clonal level as there is not net DNA gain or loss. However, we previously
371 demonstrated that chromosomes in both micronuclei and bridges undergo deficient DNA replication
372 leading to net DNA losses[34,38]. If broken chromosomes remain mitotically unstable for multiple generations,
373 successive under-replication of the broken termini can generate varying terminal losses in the progeny
374 population (**Figure 8A**) that lead to 'sloping' copy number variation (**Extended Data Figure 7**). We
375 identified sloping copy-number variation on three chromosomes in the HGD sample from Patient10 (**Figure**
376 **8B**). The constant DNA copy number of the intact homolog (gray) establishes that the sloping copy-number
377 pattern reflects genetic variation instead of technical variability (e.g., due to FFPE DNA degradation).
378 Moreover, the observation of clonal ('discrete') copy-number changes on both Chr.9 and Chr.11 in the
379 IMEAC genome within the same regions of sloping copy number in HGD suggests that the IMEAC
380 ancestor was a subclone of HGD. Remarkably, the IMEAC genome does not show clonal copy-number

381 alterations on 12q that would have been derived from an HGD subclone with varying 12q loss, but contains
382 a high-level amplification spanning *KRAS* on the 12p arm; the amplification was inferred to have originated
383 from the same parental chromosome with sloping copy number variation on the 12q-terminus in HGD. It
384 is tempting to speculate that the *KRAS* amplification had evolved from an unstable Chr12 missing the q-
385 terminus by chromothripsis and subsequent duplications.

386 To further explore the possibility that sloping copy-number variation in early-stage BE samples
387 precedes clonal SCNAs in late-stage BE subclones, we analyzed the sequencing data of longitudinal BE
388 samples released in a recent study[55] (**Extended Data Figure 9A**). We first confirmed the presence of large
389 segmental SCNAs in both non-dysplastic and dysplastic BE samples prior to transformation and the
390 presence of distinct copy-number alterations in aneuploid BE or early cancer clones indicating copy-number
391 evolution (**Extended Data Figure 9B and 10, Extended Data Table 8, Supplementary Data**). The
392 observation of extensive copy-number evolution in longitudinal BE samples provides orthogonal evidence
393 of persistent chromosomal instability in BE cells that complements the observation of widespread copy-
394 number heterogeneity in multifocal BE samples. We further identified sloping copy-number variation in 9
395 patients. (Due to the limited sequencing depth, this inference was based on total DNA sequence coverage
396 instead of haplotype-specific coverage.) In Patient 86, we observed sloping copy-number variation on the
397 1q arm in the NDBE sample indicating varying terminal gains (**Figure 8C**, top); the same region shows a
398 clonal terminal retention in a late-stage HGD sample (**Figure 8C**, middle). In contrast to the sloping DNA
399 copy number of 1p, the 1q arm contains a subclonal paracentric gain that may be related to the
400 chromothripsis at the same 1q-terminal region in another NDBE lesion (**Figure 8C**, bottom). Together, the
401 observations in both longitudinal and multifocal BE samples suggest ongoing evolution of unstable BE
402 genomes prior to the emergence of EAC clones. As sloping copy-number variation precedes clonal SCNAs,
403 it may ultimately serve as a prognostic marker of BE progression or ongoing genome instability.

404

405 **Discussion**

406 We here studied precancer genome evolution in a unique sample set of incipient esophageal
407 adenocarcinomas and adjacent Barrett's esophagus lesions by haplotype-specific copy-number analysis.
408 We identified recurrent copy-number evolutionary patterns related to both gross karyotype changes and
409 complex segmental alterations including focal amplifications that indicate continuous genome instability in
410 BE cells.

411 We find that arm-level copy-number changes often accumulate in episodic bursts and are consistent
412 with the outcome of whole-genome duplication (WGD) and downstream events including multipolar cell
413 division and micronucleation[32,33]. WGD is frequently followed by extensive chromosome losses, giving
414 rise to highly aneuploid genomes, but can also generate near complete genome duplication. For example,
415 the EAC genome in Patient 7 is a near complete duplication of the LGD2 genome (with odd copy-number
416 states on 4q, 5, and 9q indicating post-WGD losses); the D5 cell in the single-cell collection is close to a
417 complete duplication of the F12 cell (with odd copy-number states on 2p, 9q and post-WGD gains of 17q
418 and 18p). When and how duplicated genomes re-establish stable karyotypes *in vitro* and *in vivo* require
419 further investigation.

420 We find several patterns of segmental copy-number alterations in BE/EAC genomes that are
421 consistent with an origin from dicentric chromosome breakage and evolution[38]. These include simple
422 segmental copy-number gains and losses consistent with the outcome of a single BFB cycle (**Figure 5**),
423 compound copy-number gains consistent with successive BFB cycles (**Figure 6A-C**), and distinct copy-
424 number alterations to a single parental chromosome in related BE/EAC genomes that are consistent with
425 copy-number variation generated by multigenerational BFB cycles (**Figure 6C-E**). The mechanistic
426 association between BE/EAC genome complexity and BFB cycles is further supported by the presence of
427 regional or arm-level chromothripsis (**Figure 7A,C** and **Extended Data Fig. 8A-F**), interchromosomal
428 translocations (**Figure 6A,B, Extended Data Fig. 8F,G**), and tandem-short-templates rearrangements
429 (**Extended Data Fig. 8D,G**), all of which were previously identified *in vitro*[37,38]. Finally, the patterns of
430 progressive DNA deletions (**Figure 6D**) and sloping copy-number variation (**Figure 8B,C**) provide strong
431 evidence for ongoing BFB cycles[38] in BE cells. The sloping copy-number pattern is most simply explained

432 by the under-replication of a broken chromosome over multiple generations that generates a polyclonal
433 mixture of cells with varying DNA losses. This pattern of polyclonal copy-number variation may be
434 regarded as a signature of ongoing or 'present' genome instability that precedes clonal SCNAs that indicate
435 'past' genome instability (**Figure 8A**).

436 We observe nearly ubiquitous bi-allelic *TP53* inactivation preceding the emergence of aneuploid
437 BE cells or BE clones. This result reinforces prior observations in BE cells[50] or from comparative studies
438 of BEs and late EACs[11,30,31,56]. However, cells with intact p53 do occasionally acquire large copy-number
439 alterations. This is demonstrated by the observation of infrequent arm-level or large segmental SCNAs in
440 single BE cells (**Figure 3**) and even instances of chromothripsis in BE clones (e.g., on Chr9p in Patient 8
441 BE1-3, Patient 11 LGD, and Patient 6, all samples) inferred to have occurred prior to *TP53* inactivation. In
442 contrast to BE cells with intact p53, the most distinguishing features of p53-null BE cells include (1)
443 massive aneuploidy including whole-genome duplication; and (2) complex segmental gains (with copy-
444 number states above two) that require multiple generations of chromosome breakage and recombination.
445 This observation suggests that the dominant tumor suppressive mechanism of p53 may be the suppression
446 of cell proliferation after chromosome missegregation[44].

447 The abrogation of p53-dependent cell cycle arrest after chromosome missegregation has two
448 implications (**Figure 8D**). First, arm-level or large segmental SCNAs generated by chromosome
449 missegregation events are more likely to undergo clonal expansion and become visible at the clonal level.
450 Second, and more importantly, it allows single cell division errors such as whole-genome duplication or
451 chromosome bridge formation to precipitate multigenerational instability that both generates copy-number
452 heterogeneity and fuels the acquisition of oncogenic amplifications. Therefore, even without an apparent
453 increase in the rate of events that generate unstable chromosomes, p53 loss marks the onset of rapid
454 accumulation of copy-number heterogeneity and complexity that contrasts with continuous SNV
455 accumulation. This explains the significant differences between SCNAs in ageing esophagus or BEs with
456 intact p53 and in BEs with deficient p53. Interestingly, we observed a novel pattern of copy-number
457 variation in BE cells with intact p53 reflecting uniparental disomy (UPD) alterations with varying

458 boundaries (**Figure 3E**). How large segmental UPDs arise in mammalian cells is unknown. The similarity
459 of progressive DNA breakpoints in varying UPDs to those in progressive DNA losses (**Figure 6D**) suggests
460 that these two patterns may reflect different DNA repair outcomes of broken chromosomes generated by
461 successive BFB cycles (**Extended Data Fig. 4**). If this model were true, it further implies that cells with
462 intact p53 do tolerate certain types of chromosomal instability but raises the question of how p53 or other
463 selection factors impact the rearrangement outcomes of such instability.

464 The early onset of genome instability during BE progression revealed in our analysis challenges
465 the prevailing view that chromosomal aberrations are exclusive to advanced cancers or only arise late during
466 tumor development. Analyses of advanced tumors by either bulk[5] or single-cell[57] sequencing usually reveal
467 only truncal or late subclonal alterations, indicating relatively late divergence of different cancer subclones.
468 As late-stage cancers are often dominated by the most aggressive clones, analyses of late-stage cancers
469 cannot reveal copy-number heterogeneity in single cells prior to transformation. By contrast, genetic
470 diversity is more visible in precancerous lesions due to the lack of dominant clones. This explains the
471 observation of significant copy-number differences in multifocal BE clones (**Figure 2**), copy-number
472 evolution in longitudinal BE samples (**Extended Data Fig. 9,10**), and sloping copy-number variation in
473 single BE lesions (**Figure 8B,C**). Moreover, the generation of complex copy-number gains, including focal
474 amplifications, necessitates multigenerational chromosomal instability that invariably creates copy-number
475 heterogeneity (**Figures 3, 6, 7**). Therefore, complex DNA gains in EACs or dysplastic BEs can be regarded
476 as a signature of 'past' chromosomal instability in their ancestor cells.

477 Oncogenic amplifications are a hallmark of advanced EACs. Our analyses demonstrate that these
478 events are frequently present in both early EACs and dysplastic BEs with deficient p53 (**Figures 2** and **3**).
479 We further identified distinct oncogenic amplifications in different dysplastic BEs or early EACs from the
480 same patient (**Figure 2** and **7C**), some of which were associated with independently transformed EAC foci.
481 As independent EAC clones may grow into each other to form a single tumor mass or seed different
482 metastatic lesions, both intratumor and primary/metastasis oncogenic amplification heterogeneity[58] may be
483 the inherent outcome of chromosomal instability after p53 loss that could have been initiated in precancer

484 BE cells and persist after transformation.

485 Our model of chromosomal-instability driven copy-number evolution makes several predictions.
486 First, segmental copy-number complexity at the clonal level is preceded by copy-number heterogeneity at
487 the single cell level. This is demonstrated in our study (**Figure 3** and **8**) but should be further tested by
488 single-cell DNA sequencing of precancerous or ageing tissues. Second, p53 loss enables the accumulation
489 of copy-number heterogeneity in precancer lesions that may differ from late-stage cancers due to the lack
490 of clonal sweep. This prediction can be tested in other cancers with early p53 inactivation and precursor
491 conditions, including serous ovarian cancers[16], basal breast cancers, uterine serous endometrial cancers,
492 pancreatic cancers[59], and colitis-associated colorectal cancers[15]. Finally, our analysis of SCNAs in BE/EAC
493 genomes suggests a mechanism-based classification of copy-number patterns. Extending this analysis to
494 cancers both with and without *TP53* inactivation will generate new knowledge of tumor evolution dynamics
495 with both diagnostic and therapeutic implications.

496

497 **Data availability:** All sequencing data generated in the current study were uploaded to Genotypes and
498 Phenotypes (dbGaP) (accession phs002706) with controlled access according to the Protocol approved
499 by the Institutional Review Board of the Brigham and Women's Hospital. Third-party data that were re-
500 analyzed were obtained from the European Genome-phenome Archive (EGA) through data access
501 agreement approved by the International Cancer Genome Consortium. Processed DNA copy number data
502 and compiled copy-number plots are available at https://github.com/chunyangbao/NG_ESAD75.

503

504 **Code availability:** Usage of published or public bioinformatic packages is stated in Methods with
505 references to either the publications or the repositories of the software packages. All the algorithms and
506 bioinformatic pipelines implemented in this study are described in Methods; scripts and codes are uploaded
507 to https://github.com/chunyangbao/NG_ESAD75.

References:

1. Cleal, K. & Baird, D.M. Catastrophic Endgames: Emerging Mechanisms of Telomere-Driven Genomic Instability. *Trends Genet* **36**, 347-359 (2020).
2. Knouse, K.A., Davoli, T., Elledge, S.J. & Amon, A. Aneuploidy in Cancer: Seq-ing Answers to Old Questions. *Annual Review of Cancer Biology* **1**, 335-354 (2017).
3. Levine, M.S. & Holland, A.J. The impact of mitotic errors on cell proliferation and tumorigenesis. *Genes Dev* **32**, 620-638 (2018).
4. Li, Y. *et al.* Patterns of somatic structural variation in human cancer genomes. *Nature* **578**, 112-121 (2020).
5. Watkins, T.B.K. *et al.* Pervasive chromosomal instability and karyotype order in tumour evolution. *Nature* **587**, 126-132 (2020).
6. Holland, A.J. & Cleveland, D.W. Boveri revisited: chromosomal instability, aneuploidy and tumorigenesis. *Nat Rev Mol Cell Biol* **10**, 478-87 (2009).
7. Ganem, N.J. & Pellman, D. Linking abnormal mitosis to the acquisition of DNA damage. *J Cell Biol* **199**, 871-81 (2012).
8. Tanaka, H. & Watanabe, T. Mechanisms Underlying Recurrent Genomic Amplification in Human Cancers. *Trends Cancer* **6**, 462-477 (2020).
9. Martincorena, I. *et al.* Somatic mutant clones colonize the human esophagus with age. *Science* **362**, 911-917 (2018).
10. Yokoyama, A. *et al.* Age-related remodelling of oesophageal epithelia by mutated cancer drivers. *Nature* **565**, 312-317 (2019).
11. Weaver, J.M.J. *et al.* Ordering of mutations in preinvasive disease stages of esophageal carcinogenesis. *Nat Genet* **46**, 837-843 (2014).
12. Shain, A.H. *et al.* The Genetic Evolution of Melanoma from Precursor Lesions. *N Engl J Med* **373**, 1926-36 (2015).
13. Choi, C.R., Bakir, I.A., Hart, A.L. & Graham, T.A. Clonal evolution of colorectal cancer in IBD. *Nat Rev Gastroenterol Hepatol* **14**, 218-229 (2017).
14. Stachler, M.D. *et al.* Detection of Mutations in Barrett's Esophagus Before Progression to High-Grade Dysplasia or Adenocarcinoma. *Gastroenterology* **155**, 156-167 (2018).
15. Baker, A.M. *et al.* Evolutionary history of human colitis-associated colorectal cancer. *Gut* **68**, 985-995 (2019).
16. Patch, A.M. *et al.* Whole-genome characterization of chemoresistant ovarian cancer. *Nature* **521**, 489-94 (2015).
17. Waddell, N. *et al.* Whole genomes redefine the mutational landscape of pancreatic cancer. *Nature* **518**, 495-501 (2015).
18. Navin, N. *et al.* Tumour evolution inferred by single-cell sequencing. *Nature* **472**, 90-4 (2011).
19. Davis, A., Gao, R. & Navin, N. Tumor evolution: Linear, branching, neutral or punctuated? *Biochim Biophys Acta Rev Cancer* **1867**, 151-161 (2017).
20. Helleday, T., Eshtad, S. & Nik-Zainal, S. Mechanisms underlying mutational signatures in human cancers. *Nat Rev Genet* **15**, 585-98 (2014).
21. Turajlic, S., Sottoriva, A., Graham, T. & Swanton, C. Resolving genetic heterogeneity in cancer. *Nat Rev Genet* **20**, 404-416 (2019).

22. McClintock, B. Spontaneous Alterations in Chromosome Size and Form in Zea Mays. *Cold Spring Harb Symp Quant Biol* **9**, 72-81 (1941).
23. McClintock, B. The Stability of Broken Ends of Chromosomes in Zea Mays. *Genetics* **26**, 234-82 (1941).
24. Siegel, J.J. & Amon, A. New insights into the troubles of aneuploidy. *Annu Rev Cell Dev Biol* **28**, 189-214 (2012).
25. Reid, B.J., Li, X., Galipeau, P.C. & Vaughan, T.L. Barrett's oesophagus and oesophageal adenocarcinoma: time for a new synthesis. *Nat Rev Cancer* **10**, 87-101 (2010).
26. Contino, G., Vaughan, T.L., Whiteman, D. & Fitzgerald, R.C. The Evolving Genomic Landscape of Barrett's Esophagus and Esophageal Adenocarcinoma. *Gastroenterology* **153**, 657-673 e1 (2017).
27. Peters, Y. *et al.* Barrett oesophagus. *Nat Rev Dis Primers* **5**, 35 (2019).
28. Tan, M.C. *et al.* Systematic review with meta-analysis: prevalence of prior and concurrent Barrett's oesophagus in oesophageal adenocarcinoma patients. *Aliment Pharmacol Ther* **52**, 20-36 (2020).
29. Nones, K. *et al.* Genomic catastrophes frequently arise in esophageal adenocarcinoma and drive tumorigenesis. *Nat Commun* **5**, 5224 (2014).
30. Ross-Innes, C.S. *et al.* Whole-genome sequencing provides new insights into the clonal architecture of Barrett's esophagus and esophageal adenocarcinoma. *Nat Genet* **47**, 1038-1046 (2015).
31. Stachler, M.D. *et al.* Paired exome analysis of Barrett's esophagus and adenocarcinoma. *Nat Genet* **47**, 1047-55 (2015).
32. Ganem, N.J., Godinho, S.A. & Pellman, D. A mechanism linking extra centrosomes to chromosomal instability. *Nature* **460**, 278-82 (2009).
33. Bollen, Y. *et al.* Reconstructing single-cell karyotype alterations in colorectal cancer identifies punctuated and gradual diversification patterns. *Nat Genet* **53**, 1187-1195 (2021).
34. Zhang, C.Z. *et al.* Chromothripsis from DNA damage in micronuclei. *Nature* **522**, 179-84 (2015).
35. Ly, P. *et al.* Chromosome segregation errors generate a diverse spectrum of simple and complex genomic rearrangements. *Nat Genet* **51**, 705-715 (2019).
36. Shoshani, O. *et al.* Chromothripsis drives the evolution of gene amplification in cancer. *Nature* **591**, 137-141 (2021).
37. Maciejowski, J., Li, Y., Bosco, N., Campbell, P.J. & de Lange, T. Chromothripsis and Kataegis Induced by Telomere Crisis. *Cell* **163**, 1641-54 (2015).
38. Umbreit, N.T. *et al.* Mechanisms generating cancer genome complexity from a single cell division error. *Science* **368**(2020).
39. Loh, P.R. *et al.* Reference-based phasing using the Haplotype Reference Consortium panel. *Nat Genet* **48**, 1443-1448 (2016).
40. Carter, S.L. *et al.* Absolute quantification of somatic DNA alterations in human cancer. *Nat Biotechnol* **30**, 413-21 (2012).
41. Smith, C.E., Lam, A.F. & Symington, L.S. Aberrant double-strand break repair resulting in half crossovers in mutants defective for Rad51 or the DNA polymerase delta complex. *Mol Cell Biol* **29**, 1432-41 (2009).

42. Bunz, F. *et al.* Targeted inactivation of p53 in human cells does not result in aneuploidy. *Cancer Res* **62**, 1129-33 (2002).
43. Bunz, F. *et al.* Requirement for p53 and p21 to sustain G2 arrest after DNA damage. *Science* **282**, 1497-501 (1998).
44. Uetake, Y. & Sluder, G. Activation of the apoptotic pathway during prolonged prometaphase blocks daughter cell proliferation. *Mol Biol Cell* **29**, 2632-2643 (2018).
45. Zack, T.I. *et al.* Pan-cancer patterns of somatic copy number alteration. *Nat Genet* **45**, 1134-1140 (2013).
46. Quinton, R.J. *et al.* Whole-genome doubling confers unique genetic vulnerabilities on tumour cells. *Nature* **590**, 492-497 (2021).
47. Fujiwara, T. *et al.* Cytokinesis failure generating tetraploids promotes tumorigenesis in p53-null cells. *Nature* **437**, 1043-7 (2005).
48. Davoli, T. & de Lange, T. Telomere-driven tetraploidization occurs in human cells undergoing crisis and promotes transformation of mouse cells. *Cancer Cell* **21**, 765-76 (2012).
49. Shackney, S.E. *et al.* Model for the genetic evolution of human solid tumors. *Cancer Res* **49**, 3344-54 (1989).
50. Galipeau, P.C. *et al.* 17p (p53) allelic losses, 4N (G2/tetraploid) populations, and progression to aneuploidy in Barrett's esophagus. *Proc Natl Acad Sci U S A* **93**, 7081-4 (1996).
51. Barrett, M.T. *et al.* Evolution of neoplastic cell lineages in Barrett oesophagus. *Nat Genet* **22**, 106-9 (1999).
52. McClintock, B. Chromosome organization and genic expression. *Cold Spring Harb Symp Quant Biol* **16**, 13-47 (1951).
53. McClintock, B. The Production of Homozygous Deficient Tissues with Mutant Characteristics by Means of the Aberrant Mitotic Behavior of Ring-Shaped Chromosomes. *Genetics* **23**, 315-76 (1938).
54. Xin, M. *et al.* A threshold of GATA4 and GATA6 expression is required for cardiovascular development. *Proc Natl Acad Sci U S A* **103**, 11189-94 (2006).
55. Killcoyne, S. *et al.* Genomic copy number predicts esophageal cancer years before transformation. *Nat Med* **26**, 1726-1732 (2020).
56. Newell, F. *et al.* Complex structural rearrangements are present in high-grade dysplastic Barrett's oesophagus samples. *BMC Med Genomics* **12**, 31 (2019).
57. Minussi, D.C. *et al.* Breast tumours maintain a reservoir of subclonal diversity during expansion. *Nature* **592**, 302-308 (2021).
58. Pectasides, E. *et al.* Genomic Heterogeneity as a Barrier to Precision Medicine in Gastroesophageal Adenocarcinoma. *Cancer Discov* **8**, 37-48 (2018).
59. Notta, F. *et al.* A renewed model of pancreatic cancer evolution based on genomic rearrangement patterns. *Nature* **538**, 378-382 (2016).
60. Smeby, J. *et al.* Transcriptional and functional consequences of TP53 splice mutations in colorectal cancer. *Oncogenesis* **8**, 35 (2019).
61. Chui, M.H. *et al.* Somatic intronic TP53 c.375+5G mutations are a recurrent but under-recognized mode of TP53 inactivation. *J Pathol Clin Res* **8**, 14-18 (2022).

Acknowledgements

We would like to thank all patients who were willing to contribute samples to this study, the DFCI Center for Cancer Genomics and the Genomics Platform at the Broad Institute of MIT and Harvard for assistance with library preparation and sequencing, and Dr. David Pellman for a critical reading of the manuscript.

Funding sources Doris Duke Charitable Foundation (MDS), National Institutes of Health (AJB:U54CA163060; MDS:K08DK109209; CZZ: K22CA216319), Claudia Adams Barr Program for Innovative Cancer Research (CZZ).

Author contributions Conception of study: MDS, AJB, CB, CZZ; Patient selection and clinical data collection: KW, JMD, KKM, MDS, JK; Histologic review: AA, RO, MDS; Sequencing experimental design and data generation: MDS, LS; Fluorescence in-situ hybridization analysis: HB, MW, YI; Bioinformatic analysis: CB, MDS, CZZ with help from RT, GB, CS, GG; Manuscript preparation: CZZ, CB, MDS, AJB; Manuscript review: KKW, YI

Figure Legends

Figure 1: Overview of experimental design and bioinformatic analysis. Fifteen patients whose Barrett's esophagus tissue samples presented early invasive esophageal adenocarcinomas (EAC) were selected. After histological review, 75 samples of early cancer (EAC) and precancerous lesions, including non-dysplastic Barrett's Esophagus (NDBE), low-grade dysplasia (LGD) and high-grade dysplasia (HGD), were collected via laser capture microdissection and subjected to whole-genome sequencing. We perform joint variant detection on samples from each patient and then determine their phylogeny based on genetic alterations shared by two or more samples (filled triangles). Based on the phylogeny, we then infer the timing and evolution of copy-number alterations (both shared and private), including distinct copy-number changes on a single parental chromosome in related BE/EAC genomes generated by branching evolution.

Figure 2: Phylogeny of early EAC and precursor BE lesions determined by haplotype-specific copy-number alterations. Phylogenetic trees are grouped based on the timing of whole-genome duplication (WGD, thick solid line) events. Samples are colored based on their histopathology grading: non-dysplastic (blue), low-grade dysplasia (orange), high-grade dysplasia (red), carcinoma (magenta). The branch length (horizontal distance between nodes) approximately reflects the number of altered chromosomes. For a complete list of alterations along each phylogenetic branch, see **Extended Data Table 2** and **Extended Data Figure 2**. Annotated alterations include: (1) recurrent alterations or those affecting known EAC drivers; (2) focally amplified regions or oncogenes (magenta); (3) chromosomes or chromosome arms (with asterisks) with divergent copy-number alterations in more than one progeny clones. Note that Patient 13 contained a splice-site mutation (c.375+5G>C) in *TP53* that was assessed to produce truncated p53[60] and also reported to be a recurrent hospot in cancers in a recent study[61]. The colors of annotated chromosomes reflect the complexity of copy-number alterations: simple deletion/duplication, uniparental disomy, arm-level gain/loss (blue), large segmental (terminal or internal) copy-number changes or their combinations (orange), complex copy-number alterations (red), focal amplifications (magenta). For classification of copy-number alterations, see **Extended Data Figure 4**.

Figure 3: Copy-number evolution in 56 near diploid and 12 aneuploid BE cells from a high-grade dysplastic Barrett's esophagus determined by single-cell sequencing. **A**. Phylogenetic tree with annotated haplotype-specific copy number alterations (blue for losses, red for gains). Open circles represent single cells; large filled circles represent subclones of cells (with annotated cell counts) with identical copy number; small filled circles represent inferred intermediate

34 states (gray for pre-WGD, black for post-WGD). Aneuploid cells are separated into two
35 branches each inferred to have undergone an independent whole-genome duplication (WGD)
36 event. **B-H**. Examples of copy-number alterations before (B-E) and after (F-H) p53
37 inactivation. Gray and black dots represent haplotype-specific DNA copy number of parental
38 chromosomes. **B**. Ancestral 3p uniparental disomy (UPD) shared by all but four cells. **C**.
39 Sporadic 3p terminal gain after 3p UPD in one cell. **D**. Large paracentric deletion on 1p and
40 UPD at the 1q-terminus shared by five cells. **E**. Progressive 9p UPD in a subclone of 14 cells.
41 Only four cells are shown, see **Supplementary Data** for the others. **F**. Terminal duplication
42 after terminal deletion on 9p shared by cell G1 and D11 that is consistent with two rounds of
43 breakage-fusion-bridge cycles. **G**. Chromothripsis of Chr.22q shared by cell C5, F2, and F7. **H**.
44 Focal amplification spanning the *ERBB2* gene on Chr.17 (~40Mb) in cell C5 and F7 (red
45 circles) that displays the signature copy-number pattern of breakage-fusion-bridge cycles. For a
46 detailed list of alterations in each cell, see **Extended Data Table 3**.

47 **Figure 4**: Landscape of somatic copy-number alterations (SCNA) in BE and EAC clones. **A**. Mean
48 SCNA burden in samples grouped by disease stage (*left*), *TP53* mutation status (*middle*), and
49 timing relative to whole-genome duplication (*right*). The SCNA burden is measured by the
50 total number of altered autosomes (both parental homologs, maximum 44) and subdivided into
51 local deletions or duplications (gray), uniparental disomies (light gray), arm-level SCNAs (dark
52 gray), and segmental SCNAs (black). In the middle panel, the 'intact' *TP53* group ("*TP53*")
53 only includes NDBE/LGD samples without detectable *TP53* alterations, but not HGD/EAC
54 samples. See **Extended Data Figure 3** for additional information including the SCNA burden
55 in each sample. **B**. SCNA burden along ancestral (having more than one progeny clone) and
56 terminal (only one progeny clone) phylogenetic branches. The bottom shows the *TP53*
57 mutation status and the relative timing to WGD of each branch. **C**. Total counts of arm-level
58 (left) and segmental (right) SCNAs (filled bars for gains, open bars for losses) in evolutionary
59 branches preceding, concurrent with, or after WGD. Segmental SCNAs only include large
60 internal/terminal SCNAs but not complex SCNAs that can generate both DNA gain and loss.
61 The significantly higher burden of arm-level SCNAs in WGD-concurrent branches than pre-
62 WGD branches ($p < 10^{-4}$, Mann-Whitney) that is dominated by chromosome losses ($p < 10^{-4}$,
63 two-sided Fisher's exact test) is consistent with episodic chromosome losses after
64 tetraploidization. WGD is also associated with a modest but significant increase of segmental
65 SCNA burden ($p = 0.01$, Mann-Whitney; WGD-concurrent vs pre-WGD) and of post-WGD
66 arm-level SCNAs relative to pre-WGD branches ($p = 0.01$, Mann-Whitney). These can be

67 explained by elevated rates of chromosome missegregation after tetraploidization that may also
68 lead to complex segmental changes (e.g., from micronucleation). **D**. Allelic distribution of
69 segmental SCNAs identified in all samples from each patient. Shown are the number of
70 chromosomes (Chrs.1-22 and X) with single SCNAs (open bars), multiple SCNAs on a single
71 parental homolog ('mono-allelic'), and multiple SCNAs affecting both homologs ('bi-allelic').
72 Mono-allelic and bi-allelic SCNAs with multiple breakpoints are further divided into
73 subcategories based on whether SCNA breakpoints are found in a single BE/EAC genome, or
74 in multiple related BE/EAC genomes. SCNAs (or SCNA breakpoints) concentrating on single
75 parental chromosomes (mono-allelic) is consistent with either single catastrophic events (e.g.,
76 chromothripsis) or successive SCNA acquisition on single unstable chromosomes, whereas
77 SCNAs affecting both parental chromosomes (bi-allelic) are consistent with independent
78 SCNA acquisition. **E**. Fraction of the germline genome at different copy-number states (from
79 100kb-level allelic copy number). Deletion (dark blue), subclonal deletion/loss (light blue),
80 subclonal gain (light red), or duplication (dark red). There is a marked increase in the fractions
81 of both deleted (CN = 0) and duplicated DNA (CN ≥ 2) in BE/EAC genomes with inactive p53
82 compared to BE genomes with intact p53. Samples with WGD also have a larger fraction of the
83 genome at the single-copy state reflecting DNA loss after WGD.

84 **Figure 5**: Segmental copy-number alterations in BE/EAC genomes that match the outcomes of dicentric
85 chromosome bridge resolution. **A-C.** (*Left*) Different types of dicentric chromosome breakage
86 and their copy-number outcomes: (A) terminal; (B) paracentric; or (C) pericentric segmental
87 copy number changes. The open and filled chromatids may be sister chromatids or different
88 chromosomes. Both A and B were demonstrated *in vitro* in Umbreit et al. (2020). The model
89 that pericentric copy-number changes may arise from broken dicentric ring chromosomes (C)
90 or multicentric chromosomes (not shown) has not been demonstrated *in vitro* but is plausible as
91 telomere crisis may lead to multiple critically shortened telomeres. Examples of ring
92 chromosomes (Umbreit et al., 2020) or rearrangements involving multiple chromosomes
93 (Maciejowski et al., 2015) were also observed in the progeny populations of cells that have
94 undergone telomere crisis or bridge induction. (*Right*) Examples from BE/EAC genomes that
95 recapitulate the copy-number outcomes of bridge resolution. The allelic copy-number plots
96 (25kb bins) show the DNA copy number of the altered chromosome; the intact homolog is not
97 shown. Examples of gain and loss in each group are unrelated. See **Downloadable**
98 **Supplementary Data** for the copy-number plots of both homologs in each sample. **D.**
99 Summary of terminal/internal SCNAs in BE/EAC genomes. The number of instances is shown

100 next to the copy-number pattern generated by different copy-number outcomes of BFB cycles.
101 See **Extended Data Table 5** for the complete list.

102 **Figure 6**: Segmental copy-number patterns consistent with multigenerational breakage-fusion-bridge
103 cycles. Arabic numbers represent different BFB outcomes that are also labelled in **Extended**
104 **Data Fig. 5D**. Schematic diagrams of altered chromosomes are drawn according to the
105 segmental DNA copy number. **A**. (*Top*) Terminal deletion -> terminal duplication; (*bottom*)
106 paracentric deletion -> two duplications near the centromeric break end. **B**. (*Top*) Pericentric
107 retention -> terminal duplication at the q-terminus; (*middle*) paracentric deletion -> whole-
108 chromosome duplication of the centromeric segment; (*bottom*) terminal gain or pericentric
109 deletion after whole-chromosome gain. Magenta lines represent translocations between broken
110 fragments. See **Extended Data Figure 6** for more examples. **C**. Complementary copy-number
111 gain and loss at a single breakpoint (dashed line) in HGD and IMEAC reflect two broken
112 pieces of a single dicentric chromosome. The focally amplified region on the telomeric end in
113 IMEAC is consistent with preceding BFB amplifications. **D**. A series of terminal deletions on
114 the same parental chromosome seen in five lesions from Patient 2. The proximal boundaries of
115 the subclonal DNA loss near the 4q-terminus in HGD2 and clonal DNA loss in IMEAC2
116 suggest that IMEAC2 may have evolved from a subclone in HGD2. See **Extended Data**
117 **Figure 7** for an example of the same pattern revealed in experimental BFB evolution. **E**.
118 Summary of SCNAs in related BE/EAC genomes reflecting divergent/branching BFB
119 outcomes. See **Downloadable Supplementary Data** for the copy-number plots of each
120 instance.

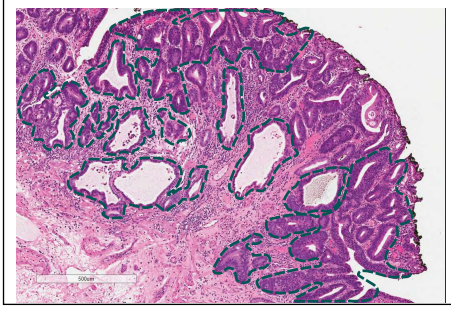
121 **Figure 7**: Complex SCNAs in BE/EAC genomes indicating successive chromothripsis and BFB cycles.
122 Arabic numbers represent different BFB outcomes that are also labelled in **Extended Data Fig.**
123 **5D**. **A**. Divergent chromothripsis (in LGD2) and terminal duplication (EAC) occurring
124 downstream of an ancestral paracentric deletion in Patient 6. The dotted line represents the
125 ancestral breakpoint shared by all three genomes; dashed lines represent private SCNA
126 breakpoints. **B**. Reciprocal distribution of Chr.14q in HGD and IMEAC lesions from Patient
127 11. The bottom shows an enlarged view of the outlined region (dashed box). Except for a small
128 segment near 30Mb, all the other segments retained in the IMEAC genome are lost from the
129 HGD genome. Dashed lines denote SCNA breakpoints with opposite retention and loss in the
130 two genomes. **C**. Five subchromosomal regions with distinct copy-number patterns in two
131 cancer lesions from Patient 1. For regions on 5p, 1p, and 8p, we infer the SCNAs evolved from
132 a single unstable ancestor chromosome based on shared SCNA breakpoints (dotted lines). For

133 16p, the SCNAs are related by a common region of terminal deletion with adjacent boundaries
134 (dashed lines). The amplified regions on 16p in EAC1 are joined to the amplified region on 18q
135 spanning *GATA6*. The order of chromothripsis and amplification is determined based on
136 whether the amplified regions are interrupted by deletions (indicating chromothripsis before
137 amplification) or peppered with DNA losses (indicating chromothripsis after amplification). **D**.
138 Summary of chromothripsis and DNA amplification instances grouped by copy-number
139 features and the inferred evolutionary sequences. The inference of chromothripsis arising either
140 directly from or downstream of dicentric chromosome breakage is based on the span of
141 oscillating copy-number pattern relative to entire chromosomes; instances with less certainty
142 are annotated accordingly ("possibly downstream" or "likely direct").

143 **Figure 8**: Chromosomal instability creates copy-number heterogeneity prior to copy-number complexity.
144 **A**. Successive BFB cycles can generate progressive DNA losses at the broken ends of
145 chromosomes resulting in a gradual attenuation (sloping) of DNA copy number towards either
146 telomeric (top) or centromeric (bottom) boundaries. Individual broken ends in single cells may
147 acquire terminal duplications that become visible after clonal expansion, but the population
148 average will accrue DNA loss due to deficient DNA replication. **B**. Sloping DNA copy number
149 on Chrs.9, 11, and 12 (black dots) in the HGD sample from Patient 10. The constant DNA copy
150 number of the other homolog is shown in gray. In the regions of sloping copy-number variation
151 on Chrs.9 and 11 in HGD, we observe clonal copy-number changepoints in IMEAC, suggesting
152 clonal expansion of a subclone/single cell in the HGD sample. **C**. BE copy-number evolution
153 revealed in longitudinal BE sequencing data published by Killcoyne et al. (2020). In this patient
154 (Patient 86), the NDBE sample at 0 month displays sloping (1p terminus) and subclonal (1q
155 terminus) copy-number variation. A subsequent HGD lesion (at 60 months) from the same
156 patient shows a (sub)clonal paracentric loss on 1p; another NDBE lesion (timing unspecified)
157 showed chromothripsis at the 1q-terminus in the same region of subclonal copy-number gain in
158 the NDBE lesion at 0 month. Both examples indicate copy-number heterogeneity. See
159 **Extended Data Fig. 10** for additional examples. **D**. Evolutionary dynamics of local sequence
160 changes (single-nucleotide variants, short sequence deletions/duplications) and chromosomal
161 structural aberrations during esophageal cancer evolution. Prior to p53 loss, the suppression of
162 clonal expansion of chromosomal structural alterations implies that only alterations that do not
163 disrupt chromosomal instability (local sequence changes, focal deletions/duplications, or
164 uniparental disomies) are detectable at the clonal level. After p53 loss, there is a rapid increase
165 of SCNA burden per cell that is due to both clonal expansion of ancestral SCNAs and SCNA

accumulation during multigenerational evolution of unstable chromosomes, which generates both copy-number heterogeneity and DNA duplications. Although the average mutational burden per cell (of both local and structural alterations) and the total genetic diversity of the tumor clone continue to increase during cell proliferation, the acquisition of cancer drivers can cause clonal dominance or sweep that make minor subclones harder to detect by bulk or even single-cell sequencing. Therefore, analyses of precancer lesions with limited clonal expansion can reveal ancestral genetic heterogeneity that may be undetectable in advanced cancers.

FIGURE 1 Archival Search in Endoscopic Mucosal Resections (EMR)

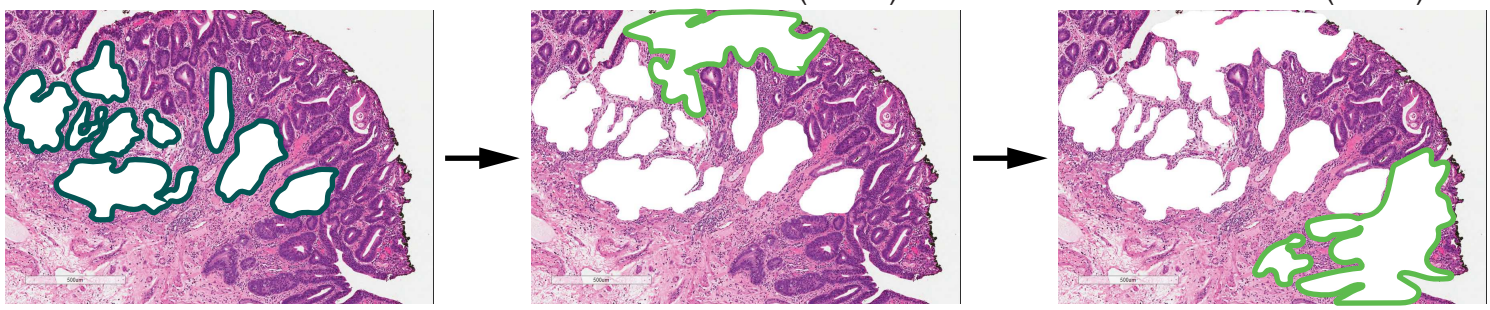


Histologic review ←

Identified 15 patients with
- T1 Esophageal AdenoCarcinoma (EAC) or IntraMucosal Esophageal AdenoCarcinoma (IMEAC)
- no prior therapy

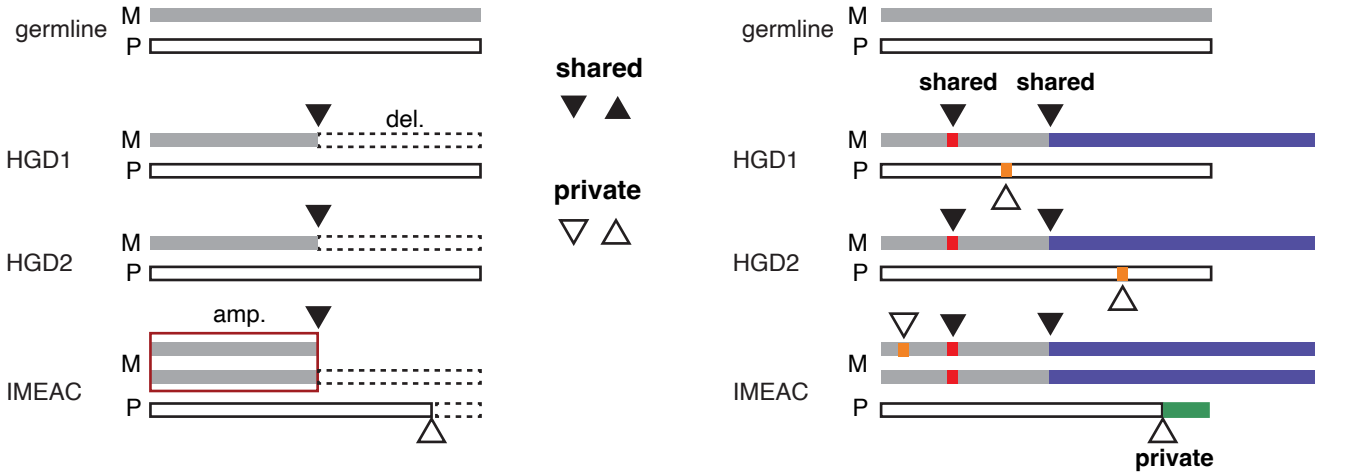
▼ Sequential Laser Capture Microdissection (IM/EAC > HGD > LGD > NDBE)

Cut IMEAC → Cut 1st HGD (HGD1) → Cut 2nd HGD (HGD2)



▼ DNA library construction, quality assessment, whole-genome sequencing (75 samples/15 patients)

Somatic variant detection



Inference of phylogenetic relationship and order of genetic alterations

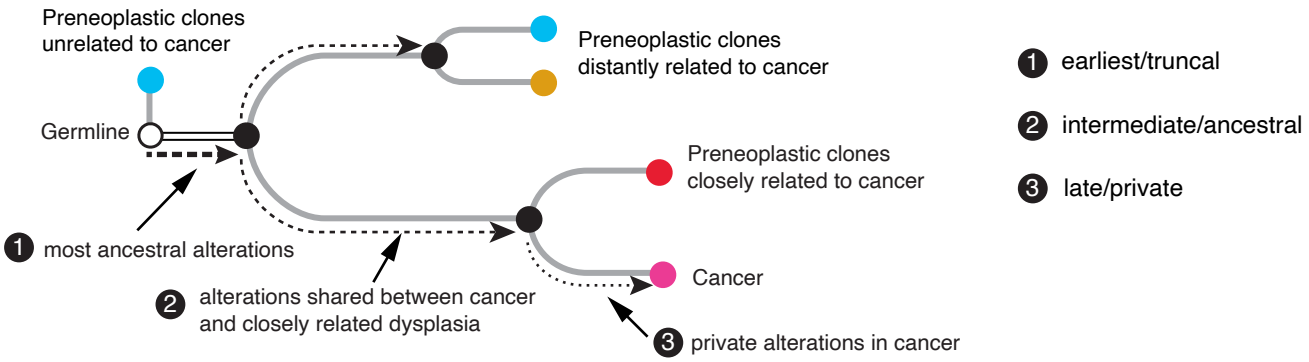


FIGURE 2

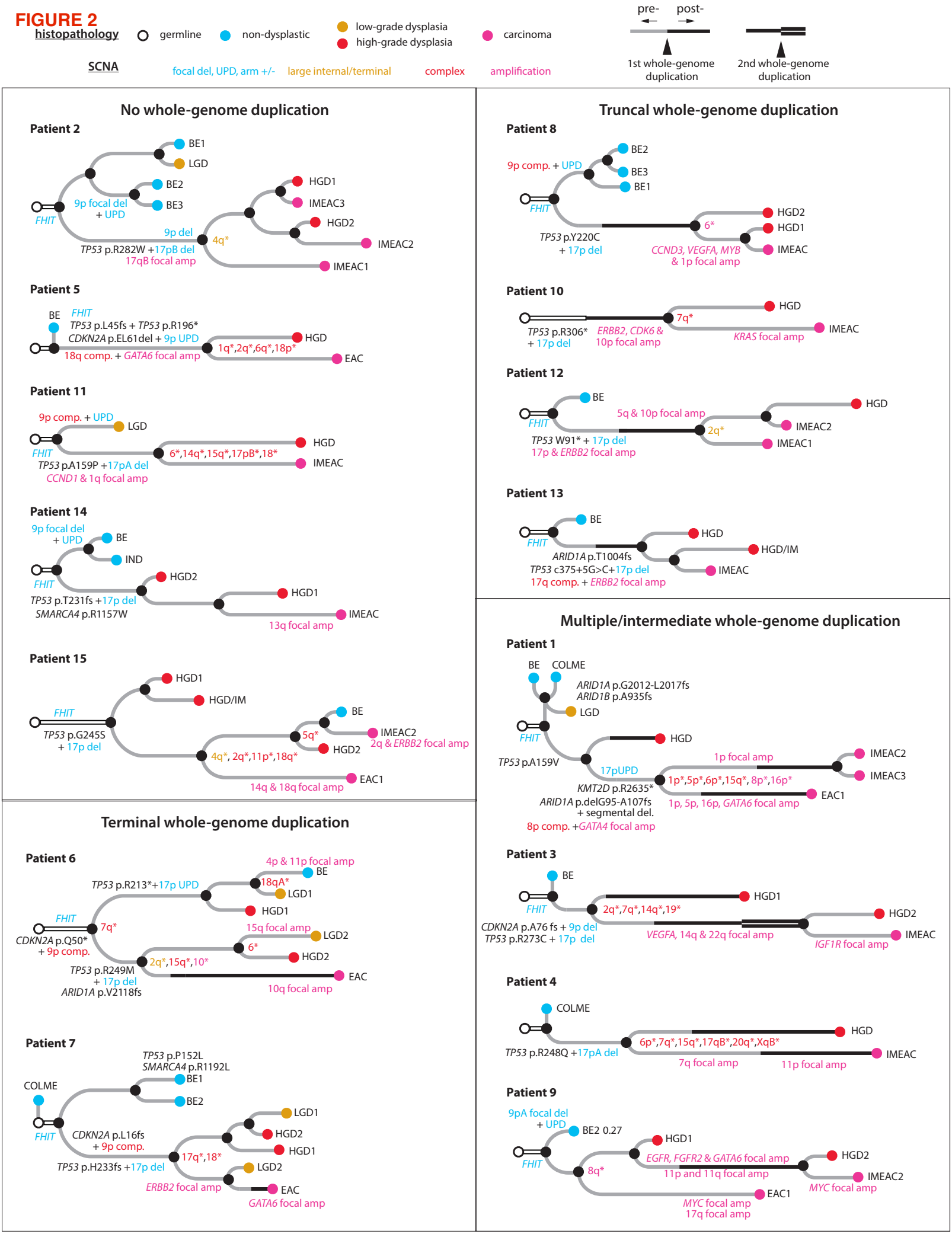
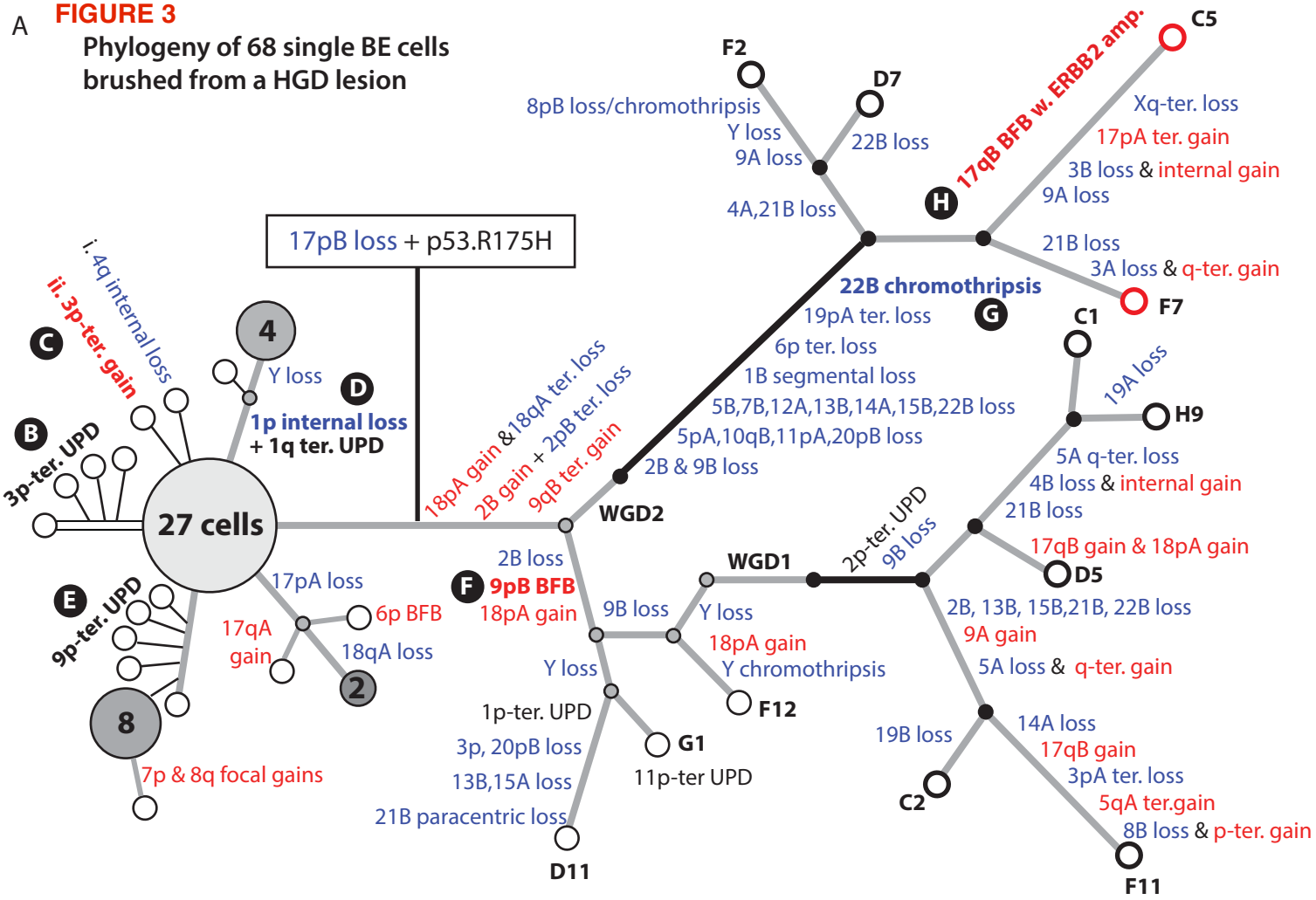

histopathology
germline
non-dysplastic
low-grade dysplasia
high-grade dysplasia
carcinoma
SCNA
focal del, UPD, arm +/-
large internal/terminal
complex
amplification
pre-
post-
1st whole-genome duplication
2nd whole-genome duplication
No whole-genome duplication
Patient 2
BE1
LGD
BE2
BE3
9p focal del + UPD
FHIT
9p del
TP53 p.R282W +17pB del
17qB focal amp
4q*
HGD1
IMEAC3
HGD2
IMEAC2
IMEAC1
Patient 5
BE
FHIT
TP53 p.L45fs + TP53 p.R196*
CDKN2A p.EL61del + 9p UPD
18q comp. + GATA6 focal amp
1q*,2q*,6q*,18p*
HGD
EAC
Patient 11
9p comp. + UPD
LGD
FHIT
TP53 p.A159P +17pA del
CCND1 & 1q focal amp
6*,14q*,15q*,17pB*,18*
HGD
IMEAC
Patient 14
9p focal del + UPD
BE
IND
FHIT
TP53 p.T231fs +17p del
SMARCA4 p.R1157W
HGD2
HGD1
IMEAC
13q focal amp
Patient 15
HGD1
HGD/IM
FHIT
TP53 p.G245S + 17p del
BE
5q*
IMEAC2
2q & ERBB2 focal amp
HGD2
4q*, 2q*,11p*,18q*
EAC1
14q & 18q focal amp
Terminal whole-genome duplication
Patient 6
4p & 11p focal amp
BE
TP53 p.R213*+17p UPD
18qA*
LGD1
HGD1
FHIT
7q*
CDKN2A p.Q50* + 9p comp.
15q focal amp
LGD2
6*
HGD2
2q*,15q*,10*
TP53 p.R249M + 17p del
ARID1A p.V2118fs
EAC
10q focal amp
Patient 7
TP53 p.P152L
SMARCA4 p.R1192L
BE1
BE2
COLME
FHIT
CDKN2A p.L16fs + 9p comp.
LGD1
HGD2
HGD1
17q*,18*
TP53 p.H233fs +17p del
LGD2
ERBB2 focal amp
EAC
GATA6 focal amp
Truncal whole-genome duplication
Patient 8
BE2
9p comp. + UPD
BE3
BE1
FHIT
HGD2
6*
HGD1
TP53 p.Y220C + 17p del
CCND3, VEGFA, MYB & 1p focal amp
IMEAC
Patient 10
HGD
7q*
TP53 p.R306* + 17p del
ERBB2, CDK6 & 10p focal amp
IMEAC
KRAS focal amp
Patient 12
BE
HGD
FHIT
5q & 10p focal amp
IMEAC2
2q*
TP53 W91* + 17p del
17p & ERBB2 focal amp
IMEAC1
Patient 13
BE
FHIT
HGD
HGD/IM
ARID1A p.T1004fs
TP53 c375+5G>C+17p del
IMEAC
17q comp. + ERBB2 focal amp
Multiple/intermediate whole-genome duplication
Patient 1
BE
COLME
ARID1A p.G2012-L2017fs
ARID1B p.A935fs
LGD
FHIT
HGD
TP53 p.A159V
1p focal amp
IMEAC2
IMEAC3
17pUPD
1p*,5p*,6p*,15q*, 8p*,16p*
KMT2D p.R2635*
ARID1A p.delG95-A107fs + segmental del.
EAC1
1p, 5p, 16p, GATA6 focal amp
8p comp. +GATA4 focal amp
Patient 3
BE
FHIT
HGD1
2q*,7q*,14q*,19*
HGD2
CDKN2A p.A76 fs + 9p del
TP53 p.R273C + 17p del
VEGFA, 14q & 22q focal amp
IMEAC
IGF1R focal amp
Patient 4
COLME
HGD
6p*,7q*,15q*,17qB*,20q*,XqB*
TP53 p.R248Q +17pA del
IMEAC
7q focal amp
11p focal amp
Patient 9
9pA focal del + UPD
BE2 0.27
HGD1
FHIT
EGFR, FGFR2 & GATA6 focal amp
HGD2
8q*
11p and 11q focal amp
IMEAC2
MYC focal amp
EAC1
MYC focal amp
17q focal amp

FIGURE 3

A

Phylogeny of 68 single BE cells brushed from a HGD lesion



Selected examples of SCNAs in single BE cells

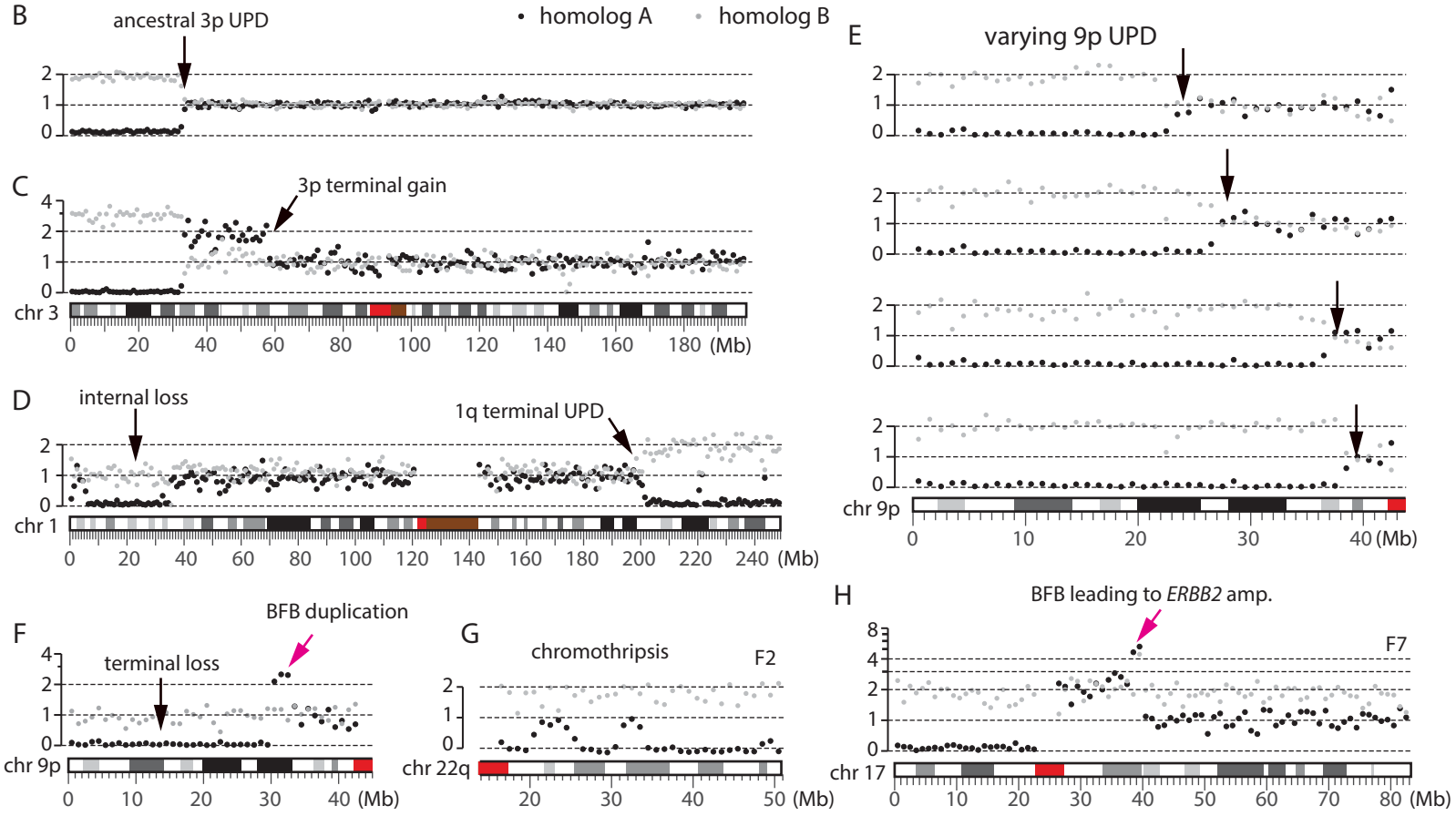


FIGURE 4

A

mean SCNA burden in each sample

disease stage

# of autosomes with SCNAs

40
30
20
10
0

UPD

focal amp
complex
terminal
large internal
arm-level
local

COLME NDBE LGD HGD EAC

TP53 status

TP53 *TP53*$^{-/-}$

relative timing to whole-genome duplication

no WGD pre-WGD post-WGD

NDBE LGD HGD EAC HGD EAC HGD EAC

B

SCNA burden in each phylogenetic branch

ancestral branches

terminal/private branches

SCNA burden

60
40
20
0

TP53
WGD

WGD post-WGD

C

mean SCNA burden in pre-, WGD, and post-WGD branches

arm-level segmental

total SCNAs

300
200
100
0

$p < 0.0001$ (Fisher's)

loss
gain

pre-WGD (36) WGD (13) post-WGD (20)

pre-WGD (36) WGD (13) post-WGD (20) (# branches)

E

fraction of genome by copy-number states

NDBE/LGD with intact p53

BE/EAC with inactive p53

no WGD with WGD

1
0.8
0.6
0.4
0.2
0

deletion duplication

deletion loss duplication

0 1 2 0 1 2 3 0 2 4

allelic DNA copy number

D

allelic distribution of segmental SCNAs

# of chromosomes with segmental SCNAs

20
10
0

single SCNA

Patient 1 → 15

multiple SCNAs

mono-allelic bi-allelic

multiple BE/EACs

both

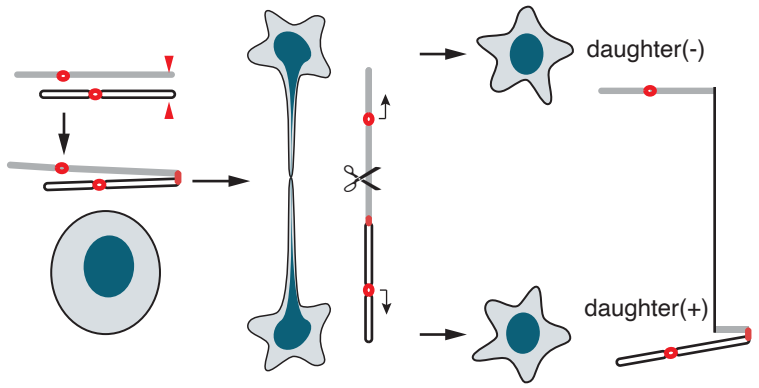
single BE/EAC

multiple BE/EACs

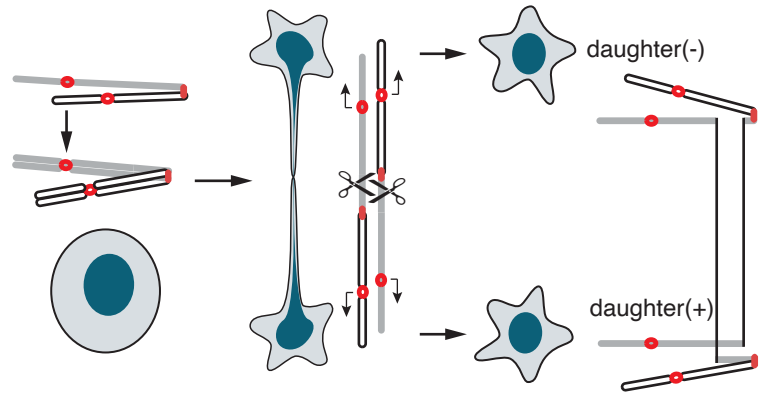
single BE/EAC

FIGURE 5 Models of dicentric bridge resolution

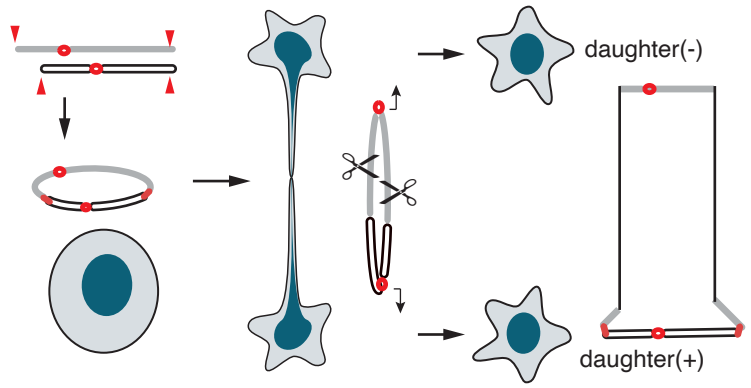
A chromatid-type fusion



B chromosome-type fusion

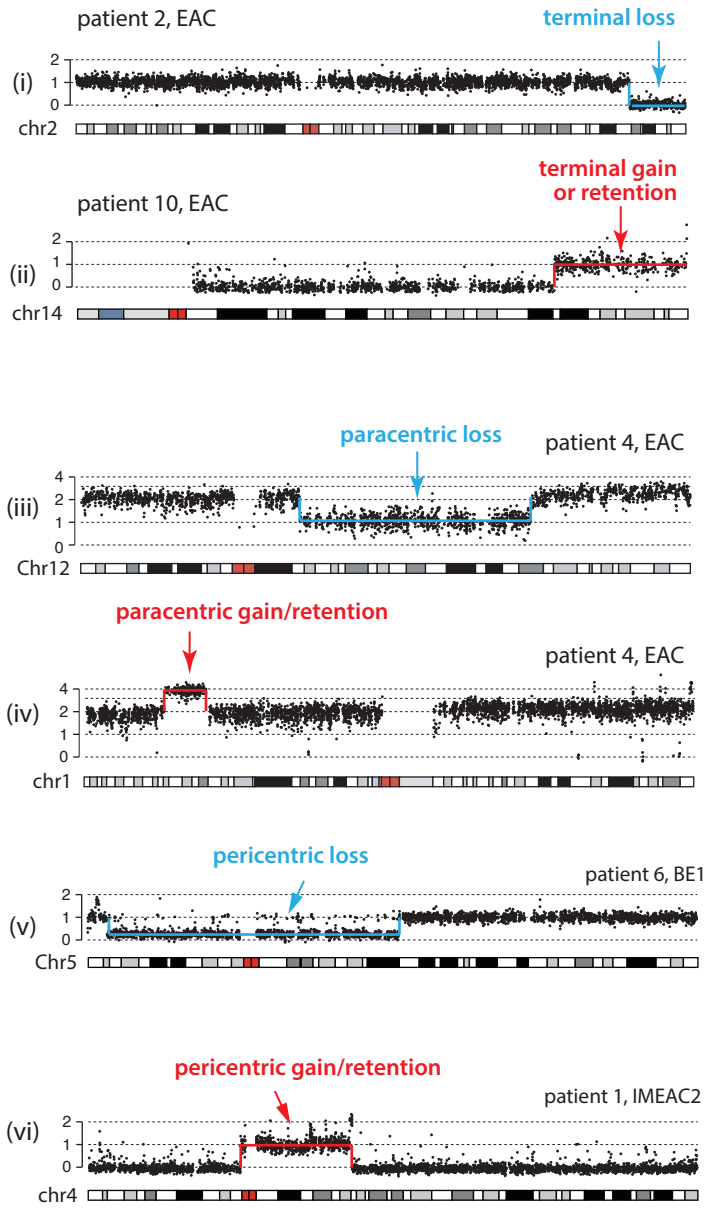


C ring fusion



Observations in esophageal cancers

- altered homolog only (other homolog not shown)



D

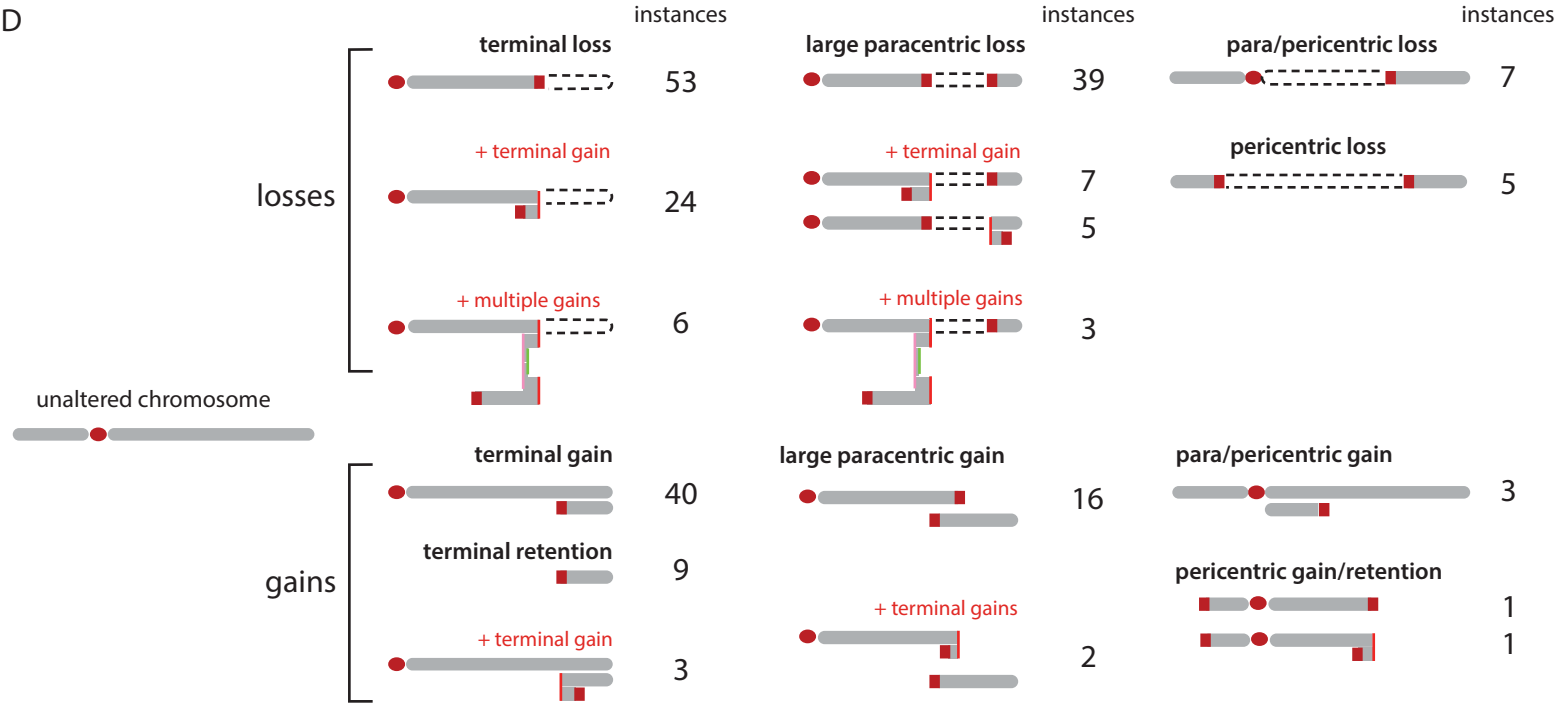
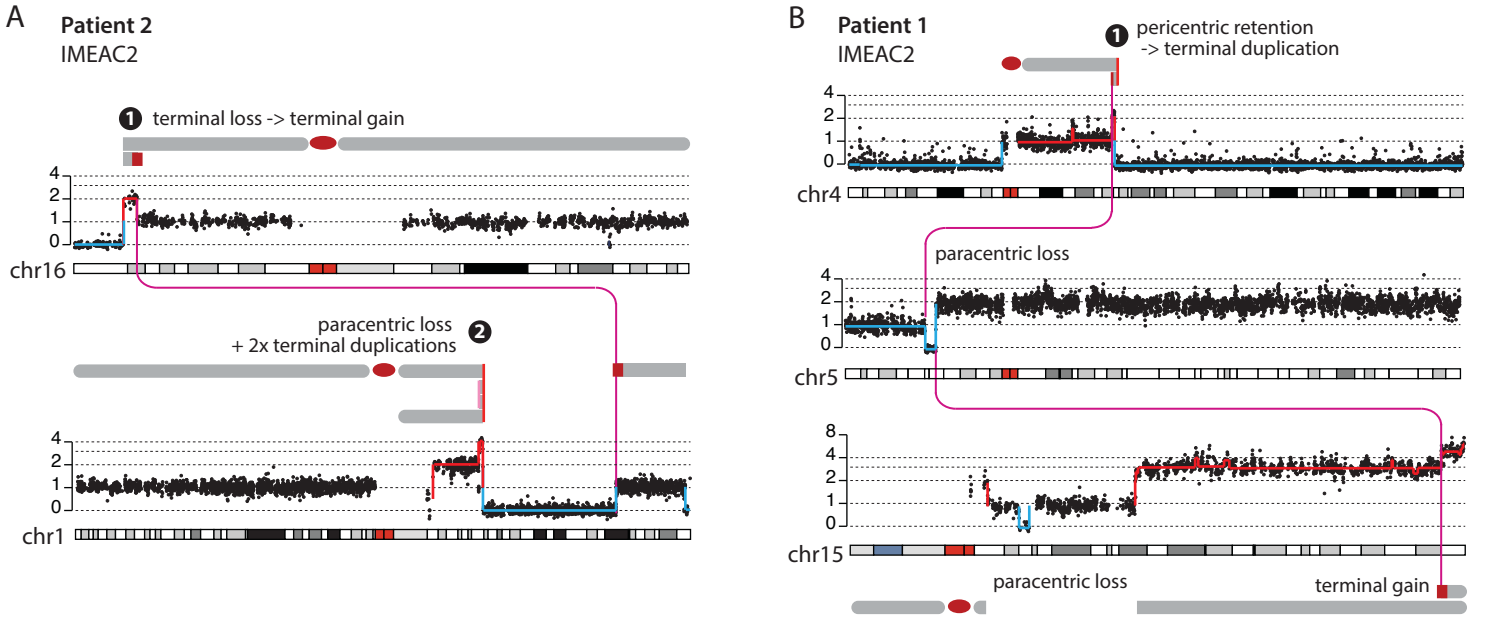
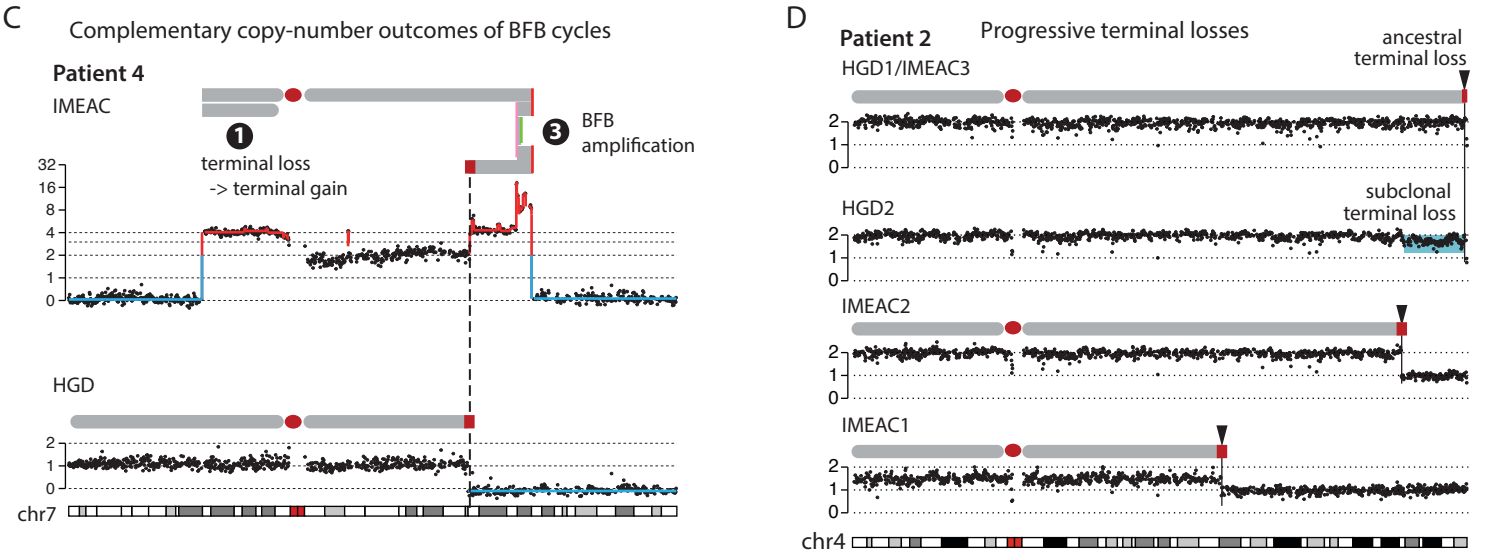


FIGURE 6

Compound copy-number outcomes of BFB cycles



Divergent/branching copy-number outcomes of BFB cycles



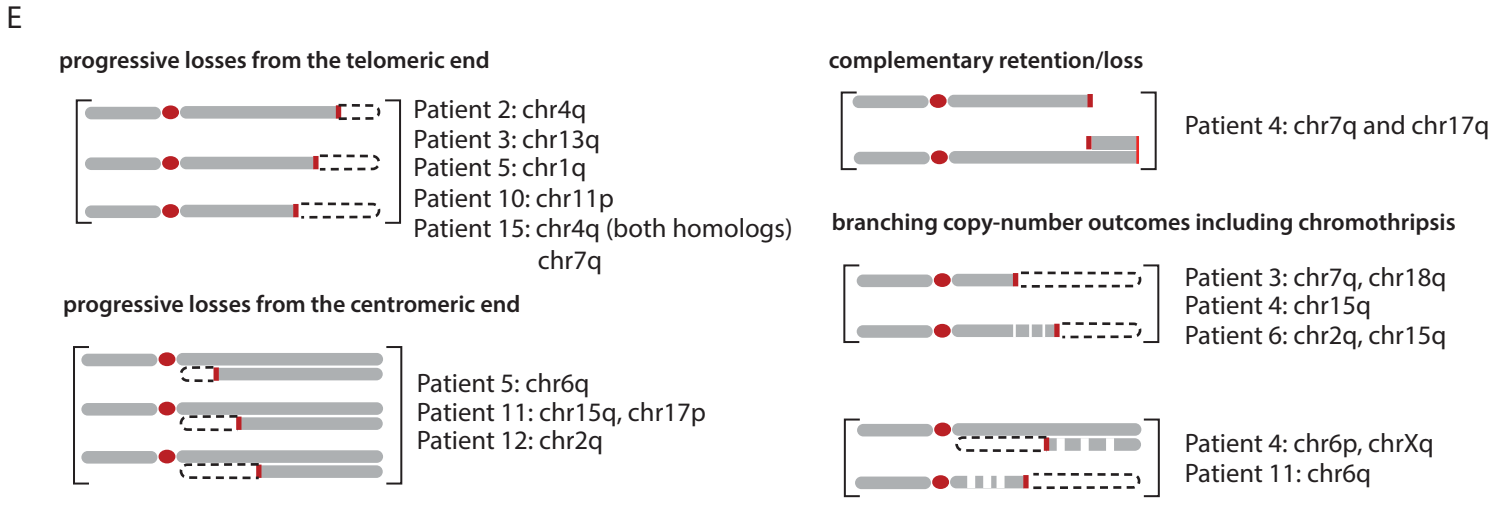
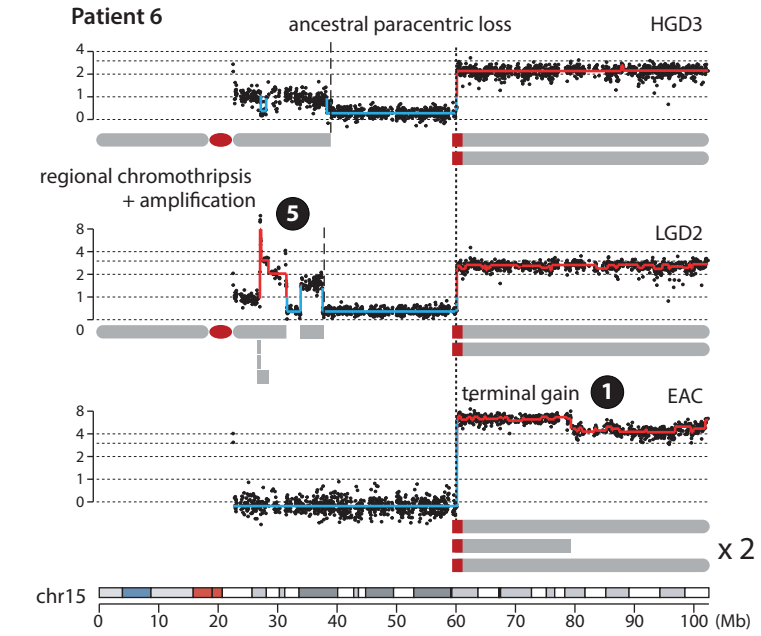
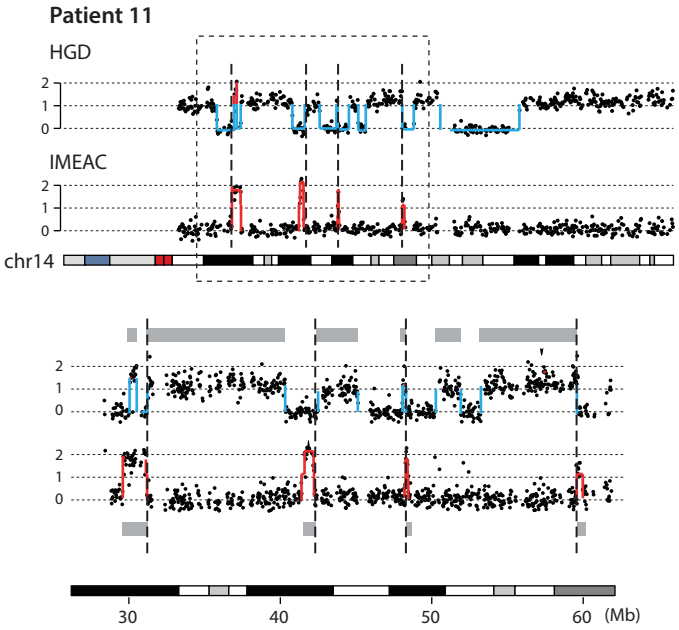


FIGURE 7

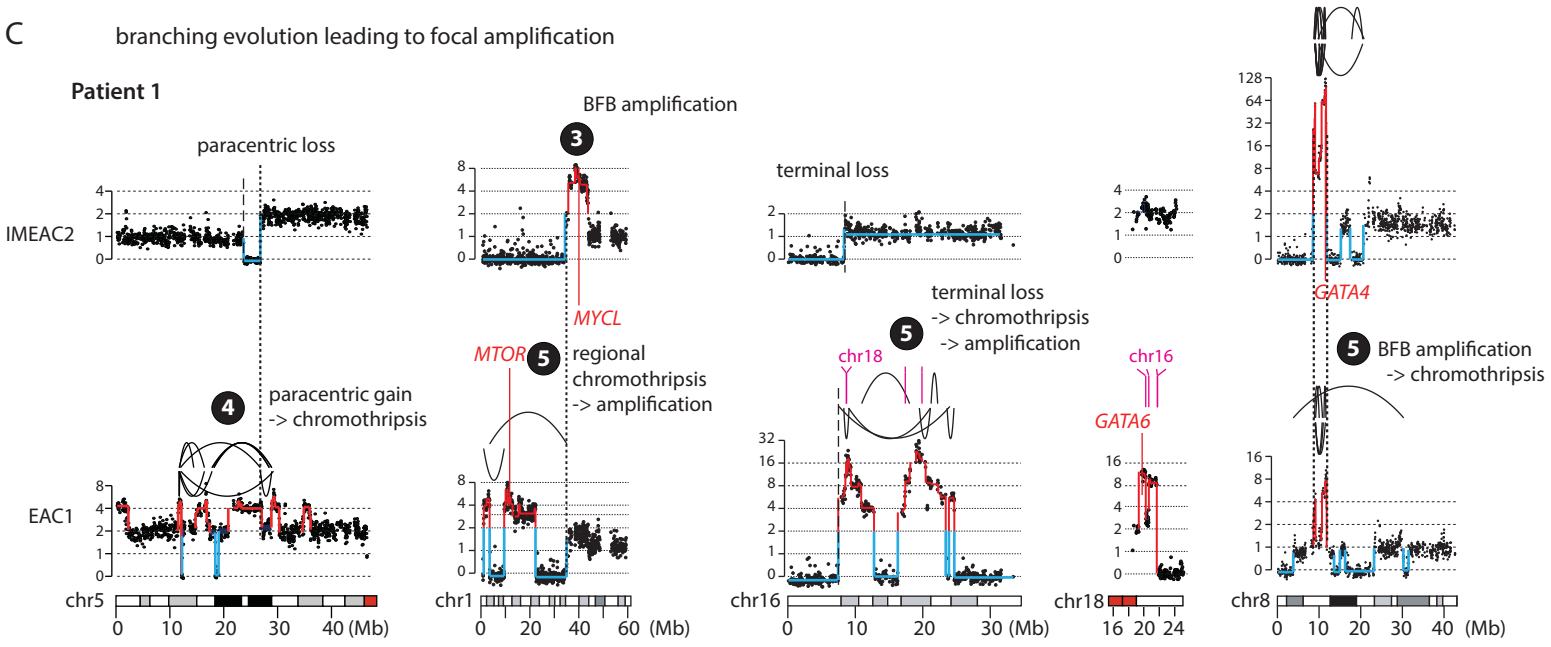
A chromothripsis arising downstream of BFB cycle



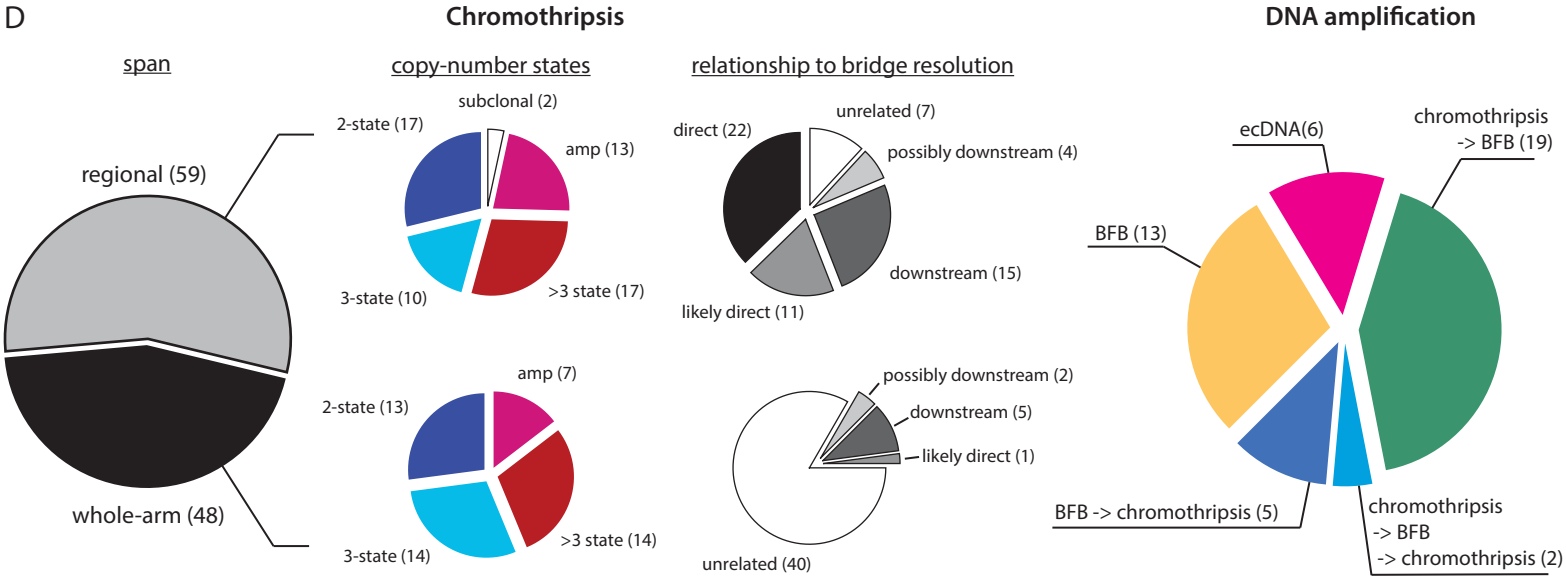
B chromothripsis creating reciprocal DNA retention and loss



C branching evolution leading to focal amplification

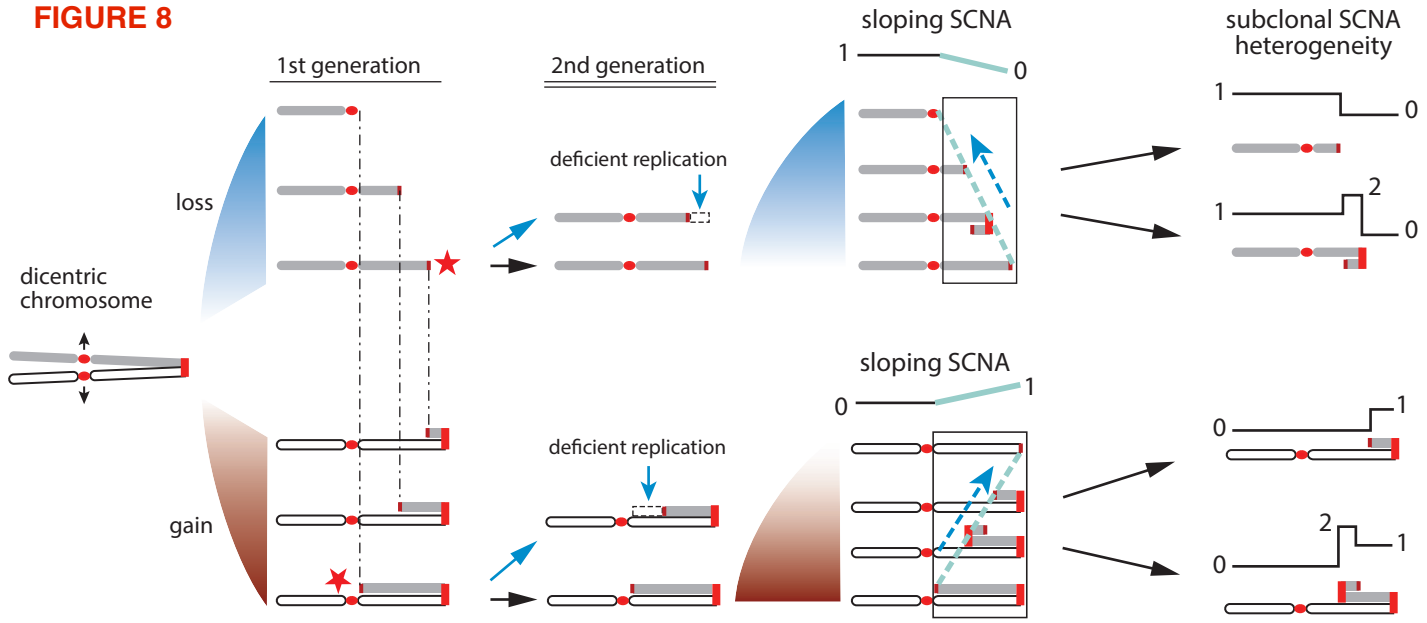


D

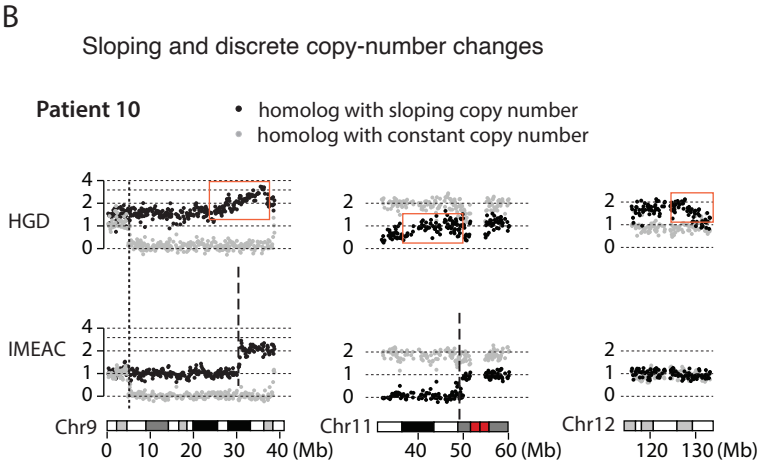


A

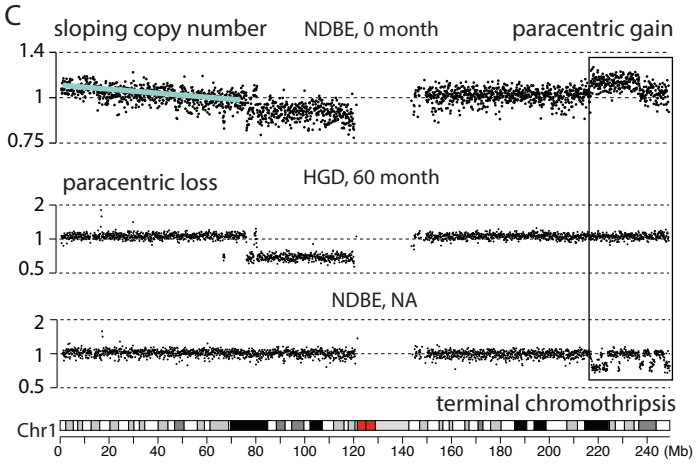
FIGURE 8



B



C



D

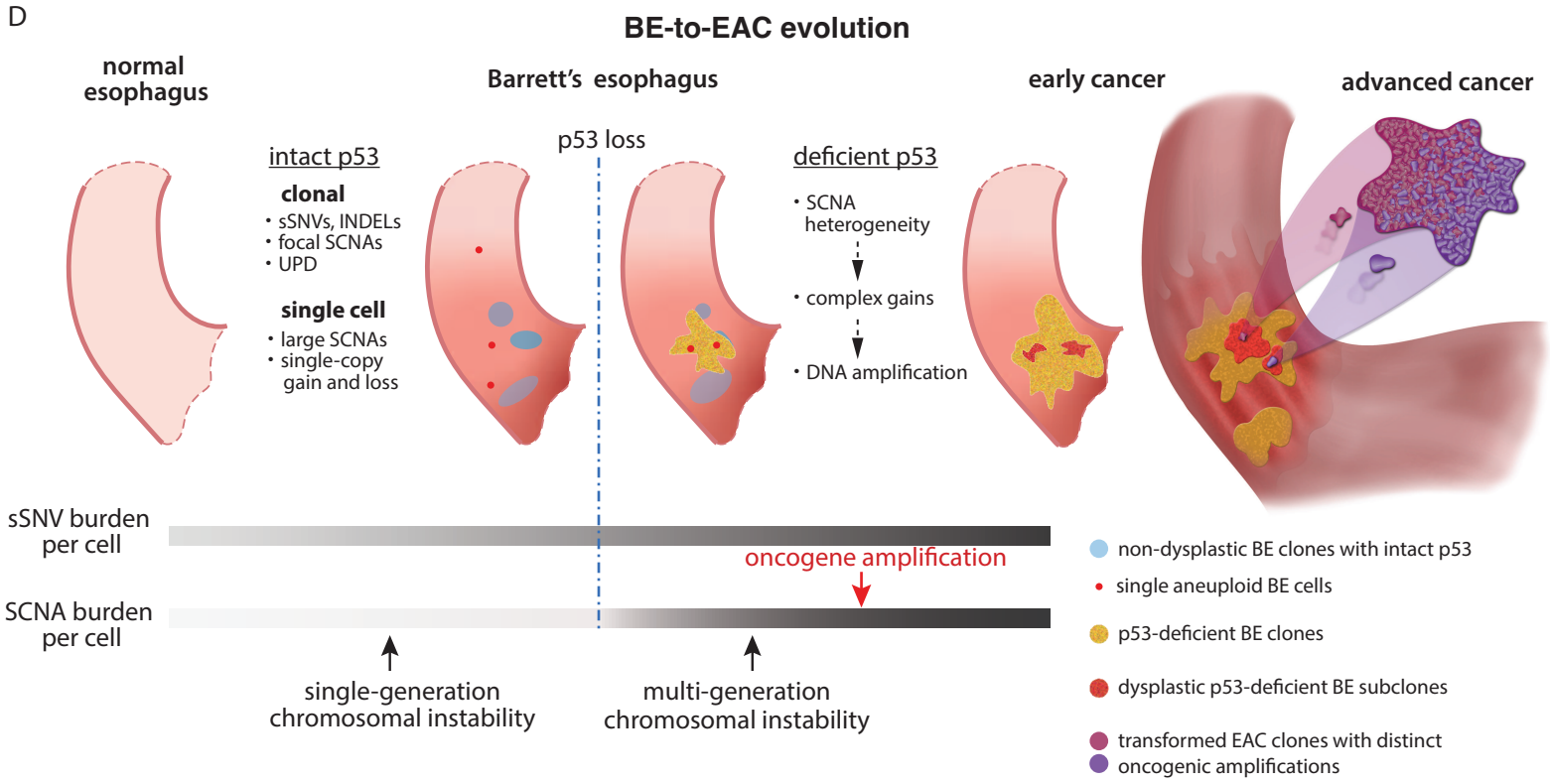
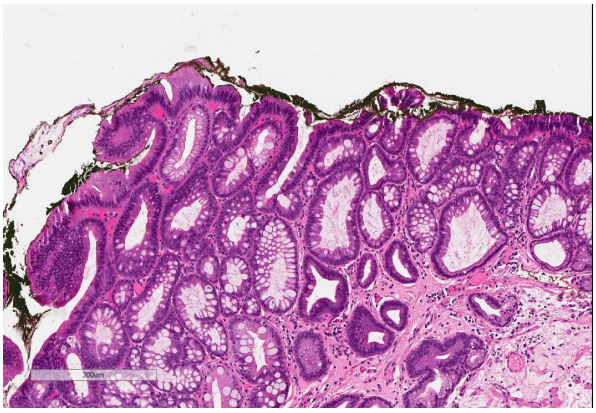


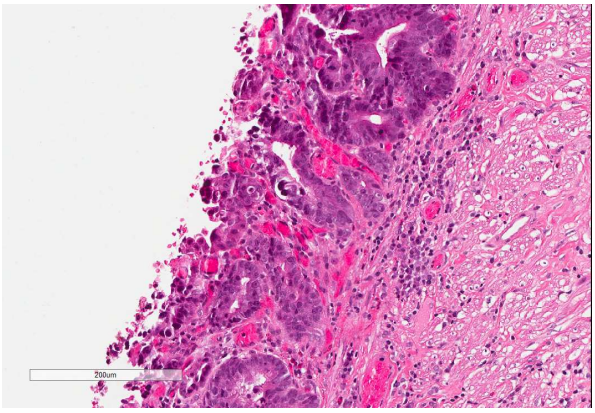
FIGURE S1

Barrett's esophagus

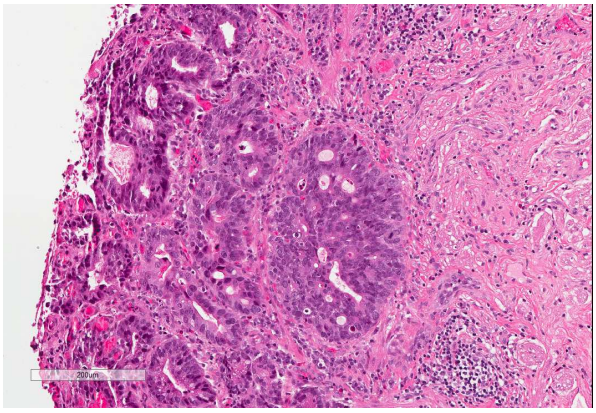
Low grade dysplasia



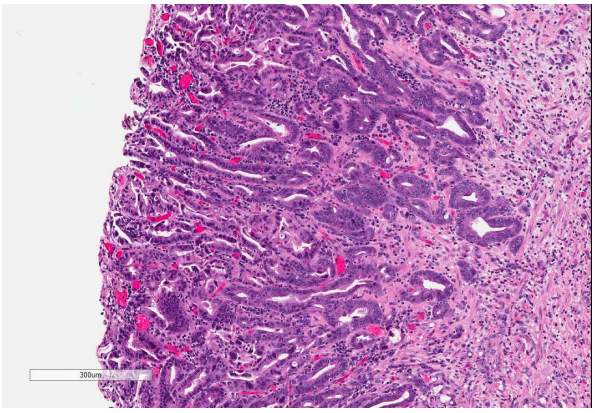
High grade dysplasia



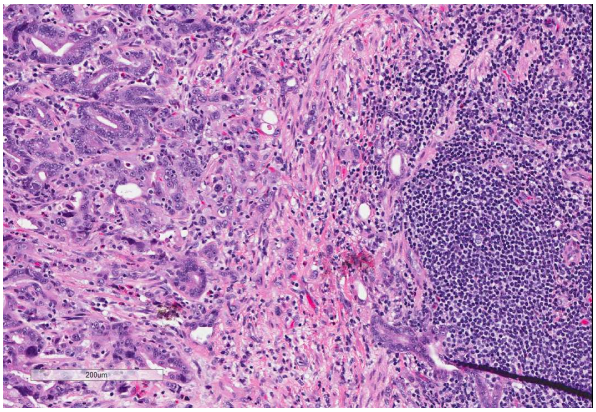
Small Intramucosal adenocarcinoma



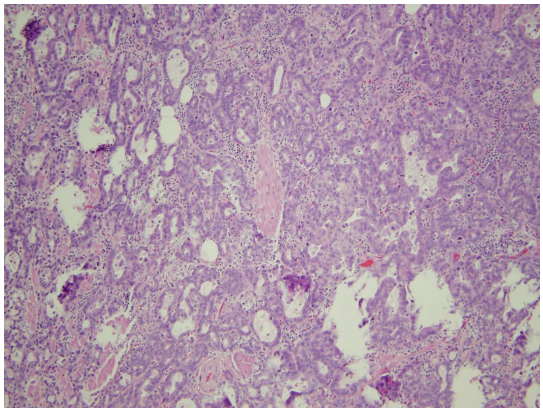
Intramucosal adenocarcinoma approaching fully invasive



Early adenocarcinoma
(small glad invading through muscularis mucosa)



Adenocarcinoma



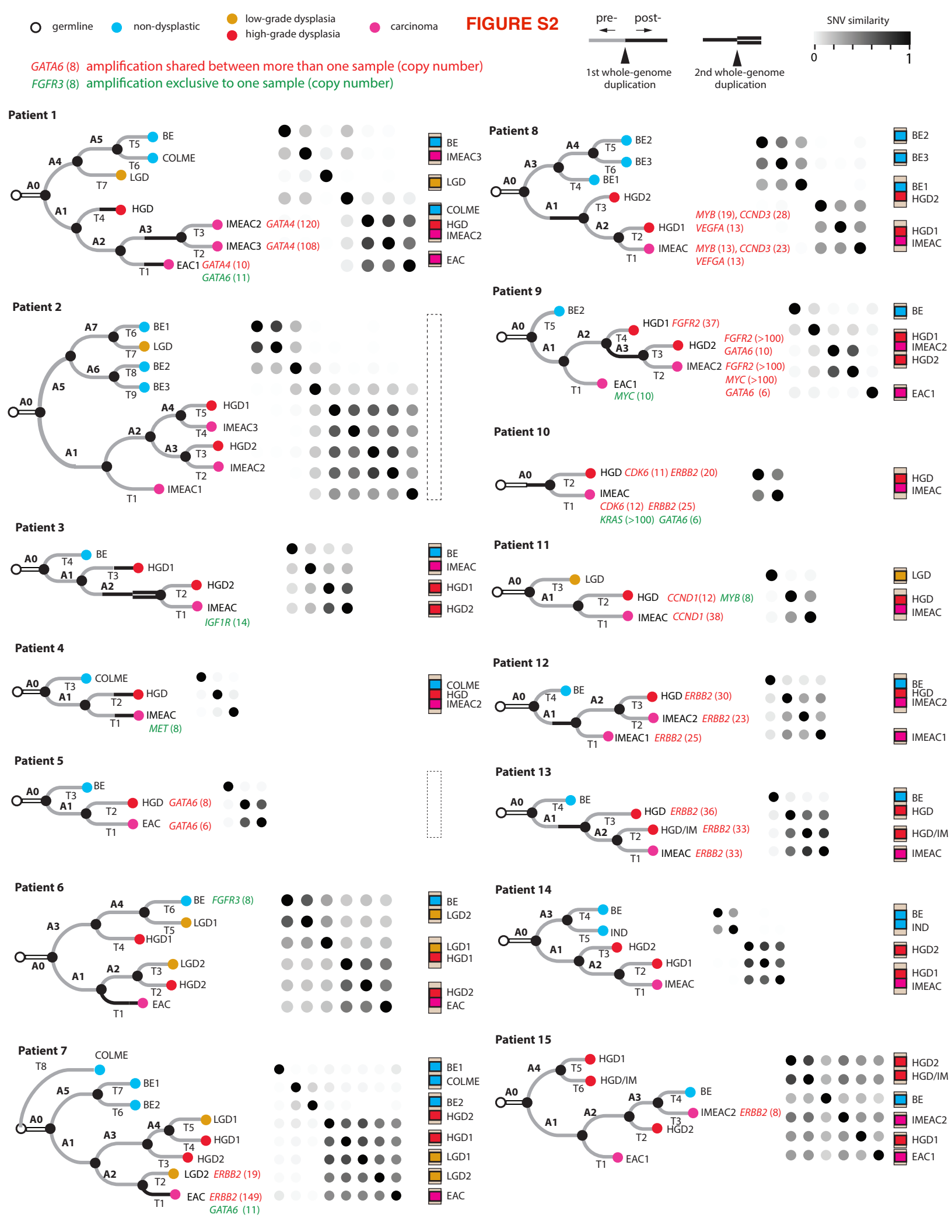
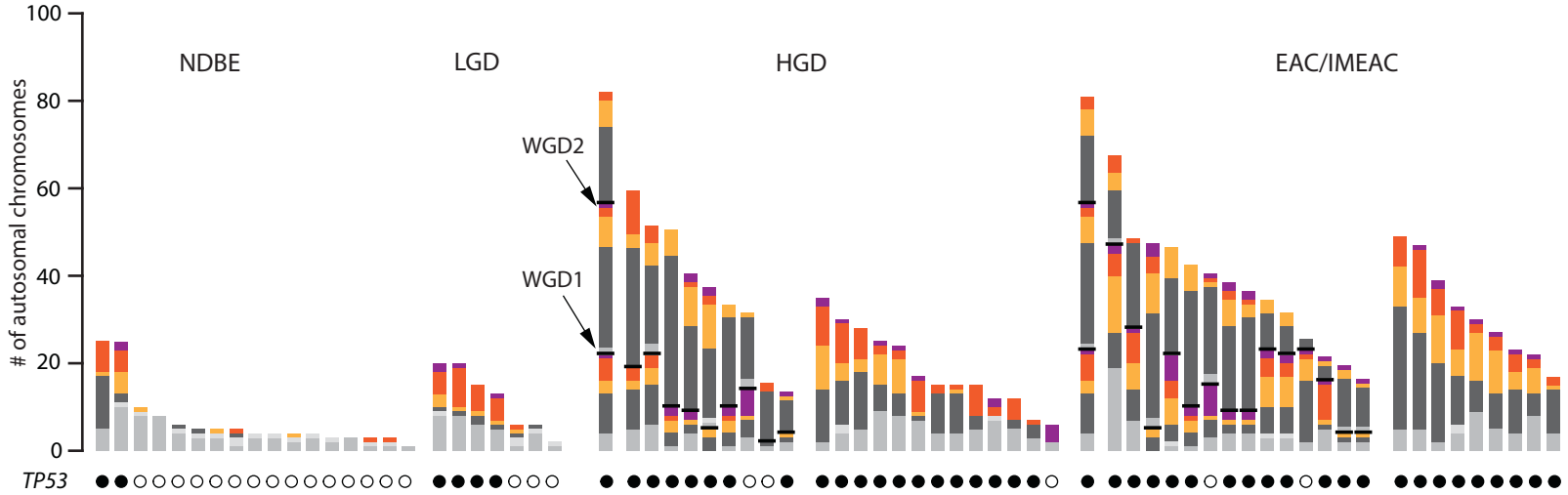


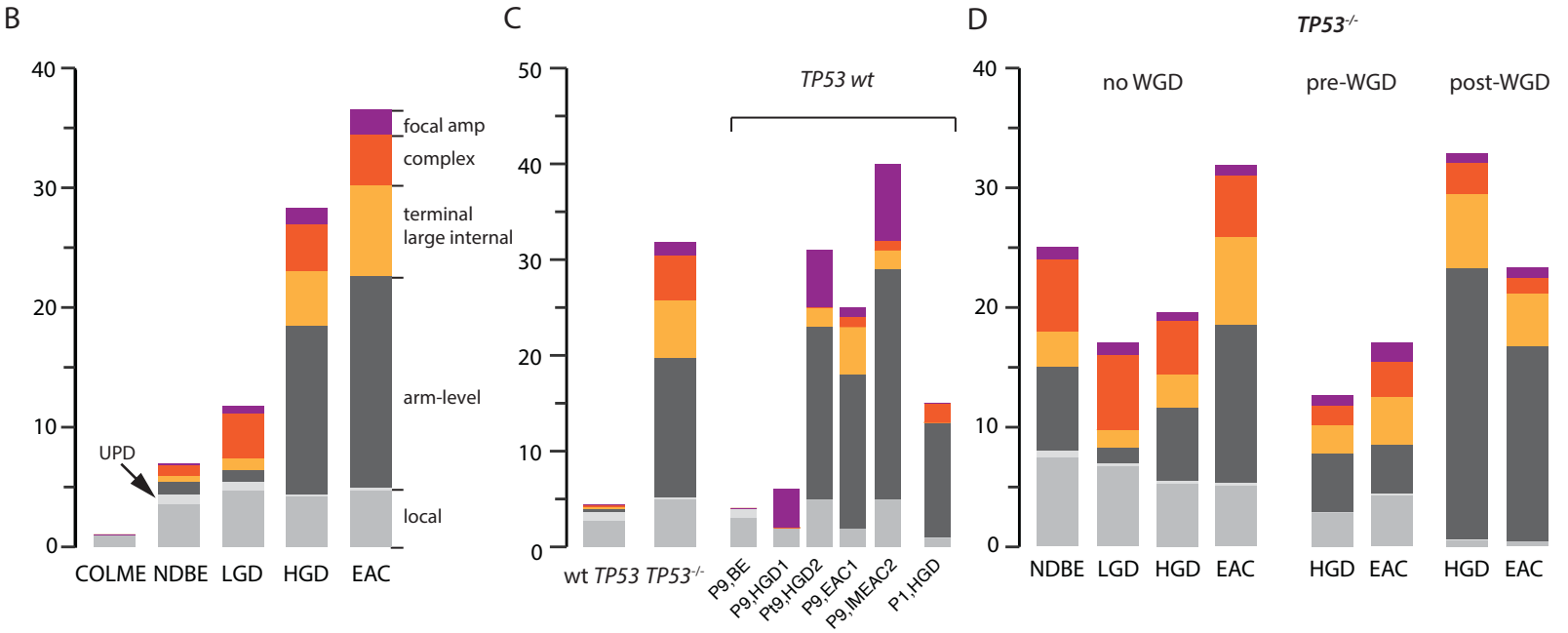
FIGURE S2
germline
non-dysplastic
low-grade dysplasia
high-grade dysplasia
carcinoma
GATA6 (8) amplification shared between more than one sample (copy number)
FGFR3 (8) amplification exclusive to one sample (copy number)
pre-
post-
1st whole-genome duplication
2nd whole-genome duplication
SNV similarity
0
1
Patient 1
BE
IMEAC3
LGD
COLME
HGD
IMEAC2
EAC
IMEAC2 GATA4 (120)
IMEAC3 GATA4 (108)
EAC1 GATA4 (10) GATA6 (11)
Patient 2
BE1
LGD
BE2
BE3
HGD1
IMEAC3
HGD2
IMEAC2
IMEAC1
Patient 3
BE
HGD1
HGD2
IMEAC
IGF1R (14)
Patient 4
COLME
HGD
IMEAC
MET (8)
Patient 5
BE
HGD GATA6 (8)
EAC GATA6 (6)
Patient 6
BE FGFR3 (8)
LGD1
HGD1
LGD2
HGD2
EAC
Patient 7
COLME
BE1
BE2
LGD1
HGD1
HGD2
LGD2 ERBB2 (19)
EAC ERBB2 (149) GATA6 (11)
Patient 8
BE2
BE3
BE1
HGD2
HGD1
IMEAC
MYB (19), CCND3 (28) VEGFA (13)
MYB (13), CCND3 (23) VEGFA (13)
Patient 9
BE2
HGD1 FGFR2 (37)
HGD2 FGFR2 (>100) GATA6 (10)
IMEAC2 FGFR2 (>100) MYC (>100) GATA6 (6)
EAC1 MYC (10)
Patient 10
HGD CDK6 (11) ERBB2 (20)
IMEAC CDK6 (12) ERBB2 (25) KRAS (>100) GATA6 (6)
Patient 11
LGD
HGD CCND1(12) MYB (8)
IMEAC CCND1 (38)
Patient 12
BE
HGD ERBB2 (30)
IMEAC2 ERBB2 (23)
IMEAC1 ERBB2 (25)
Patient 13
BE
HGD ERBB2 (36)
HGD/IM ERBB2 (33)
IMEAC ERBB2 (33)
Patient 14
BE
IND
HGD2
HGD1
IMEAC
Patient 15
HGD1
HGD/IM
BE
IMEAC2 ERBB2 (8)
HGD2
EAC1

A **FIGURE S3**

Total SCNA burden in each sample

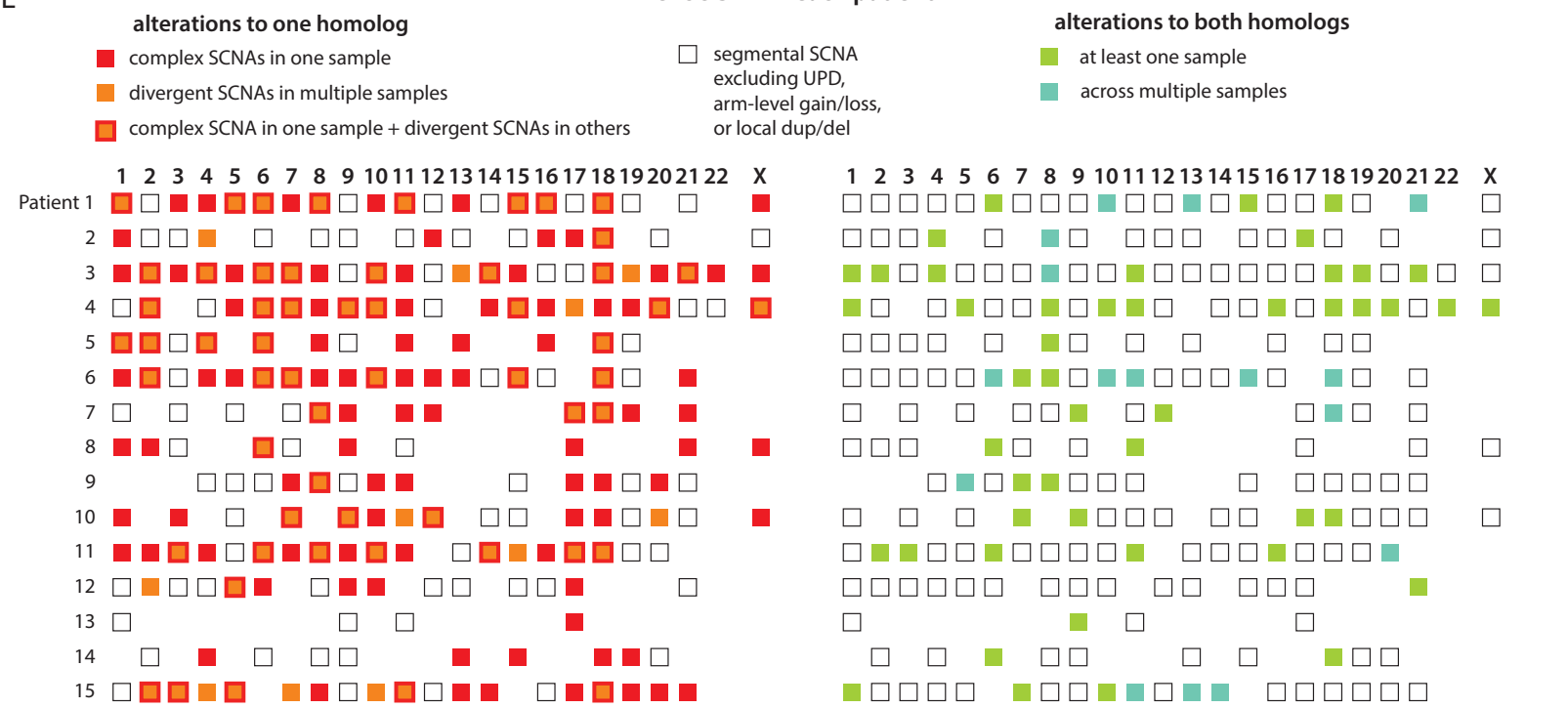


mean SCNA burden per sample



E

Allelic SCNA in each patient



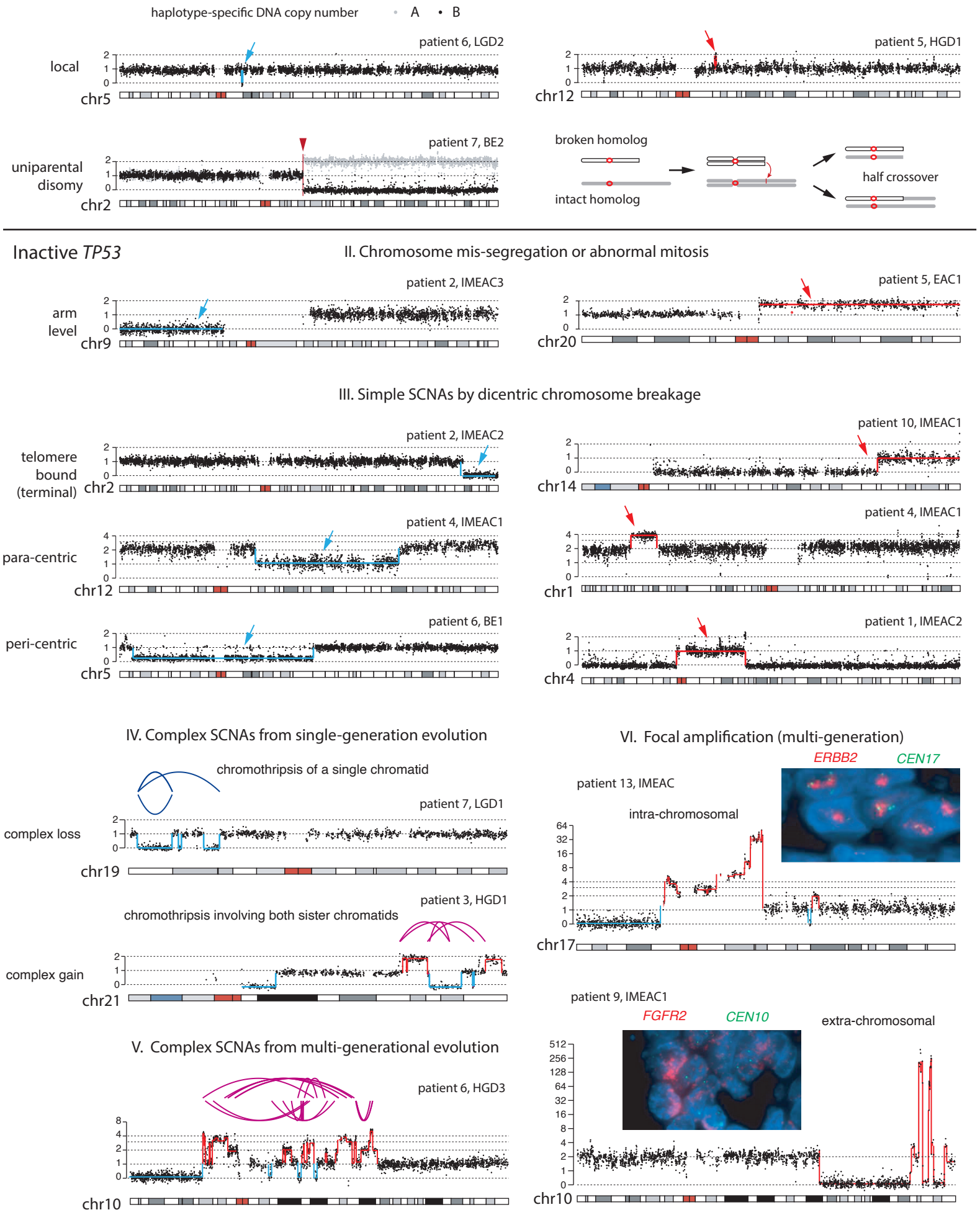


FIGURE S4
Intact TP53
I. Do not generate unstable chromosomes
haplotype-specific DNA copy number
A
B
local
patient 6, LGD2
chr5
patient 5, HGD1
chr12
uniparental disomy
patient 7, BE2
chr2
broken homolog
intact homolog
half crossover
Inactive TP53
II. Chromosome mis-segregation or abnormal mitosis
arm level
patient 2, IMEAC3
chr9
patient 5, EAC1
chr20
III. Simple SCNAs by dicentric chromosome breakage
telomere bound (terminal)
patient 2, IMEAC2
chr2
patient 10, IMEAC1
chr14
para-centric
patient 4, IMEAC1
chr12
patient 4, IMEAC1
chr1
peri-centric
patient 6, BE1
chr5
patient 1, IMEAC2
chr4
IV. Complex SCNAs from single-generation evolution
chromothripsis of a single chromatid
complex loss
patient 7, LGD1
chr19
chromothripsis involving both sister chromatids
patient 3, HGD1
complex gain
chr21
V. Complex SCNAs from multi-generational evolution
patient 6, HGD3
chr10
VI. Focal amplification (multi-generation)
ERBB2
CEN17
patient 13, IMEAC
intra-chromosomal
chr17
patient 9, IMEAC1
FGFR2
CEN10
extra-chromosomal
chr10

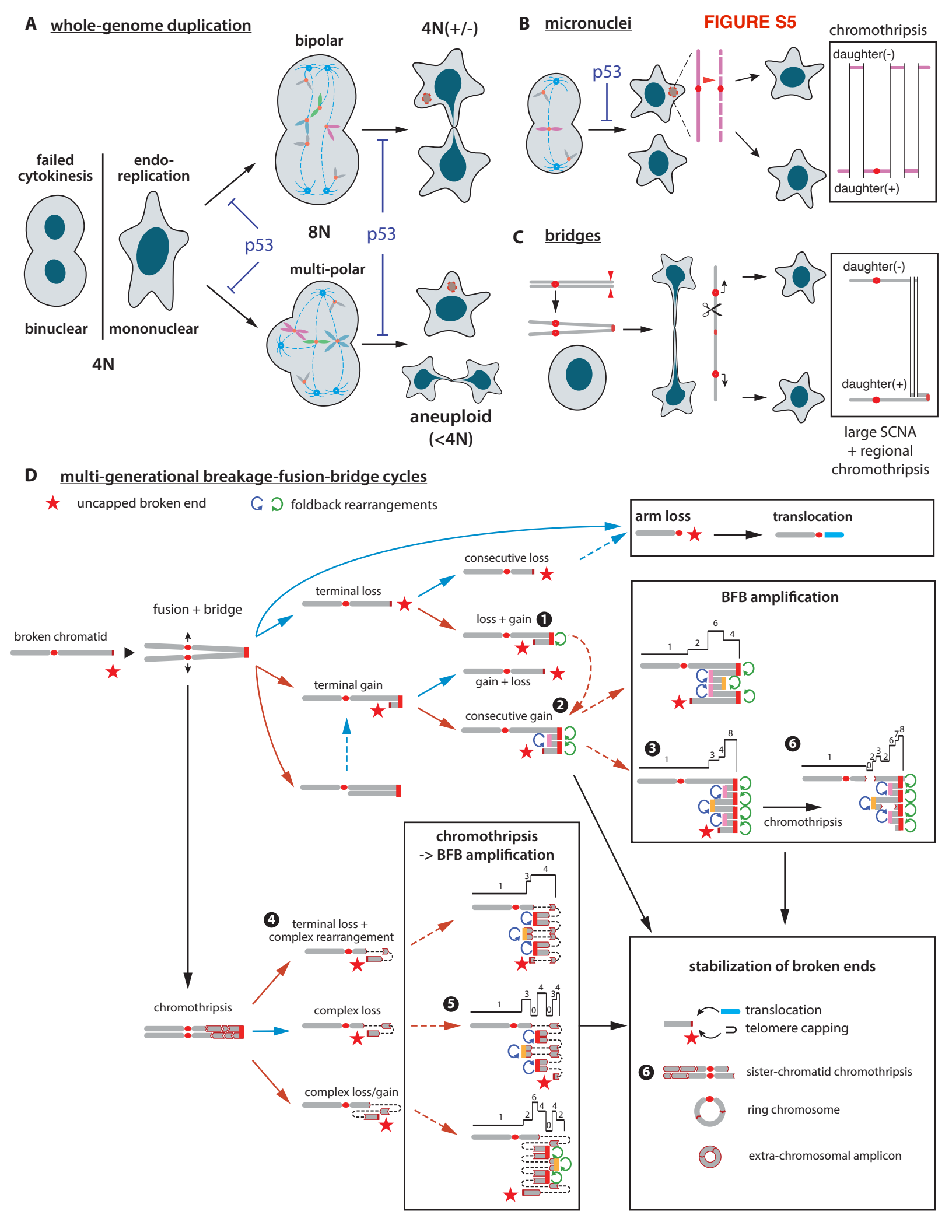
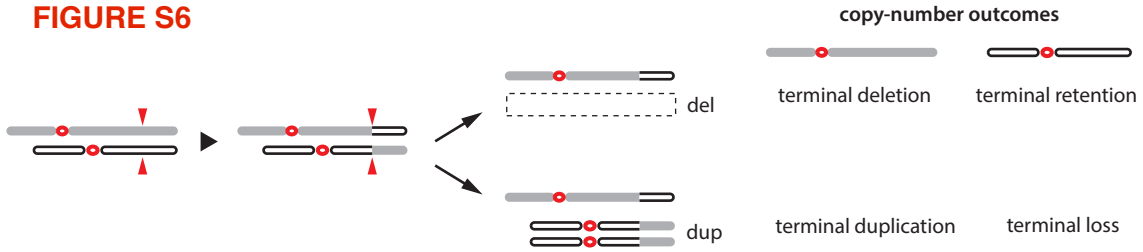
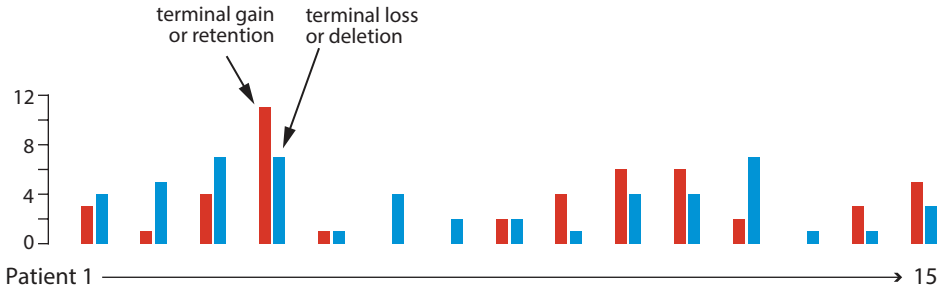


FIGURE S5
A whole-genome duplication
failed cytokinesis
endo-replication
binuclear
mononuclear
4N
bipolar
8N
4N(+/-)
p53
p53
multi-polar
aneuploid (<4N)
B micronuclei
p53
chromothripsis
daughter(-)
daughter(+)
C bridges
daughter(-)
daughter(+)
large SCNA + regional chromothripsis
D multi-generational breakage-fusion-bridge cycles
uncapped broken end
foldback rearrangements
broken chromatid
fusion + bridge
terminal loss
consecutive loss
loss + gain
terminal gain
gain + loss
consecutive gain
arm loss
translocation
BFB amplification
chromothripsis
chromothripsis -> BFB amplification
terminal loss + complex rearrangement
complex loss
complex loss/gain
stabilization of broken ends
translocation
telomere capping
sister-chromatid chromothripsis
ring chromosome
extra-chromosomal amplicon

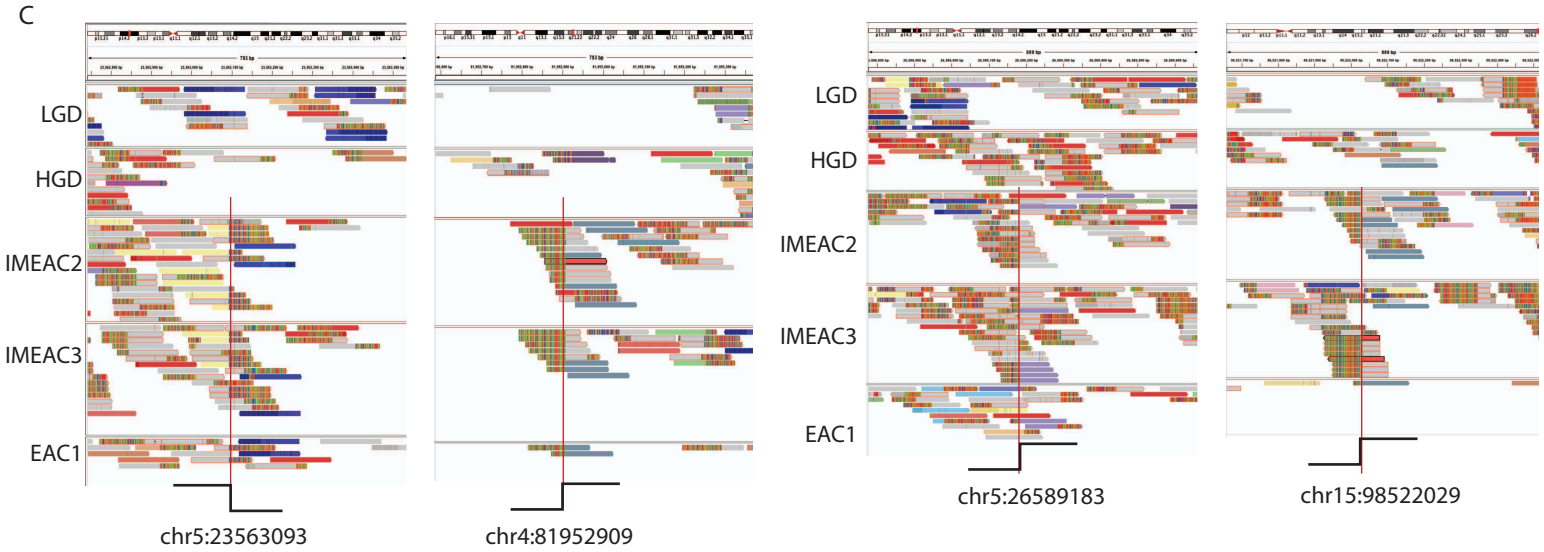
A **FIGURE S6**



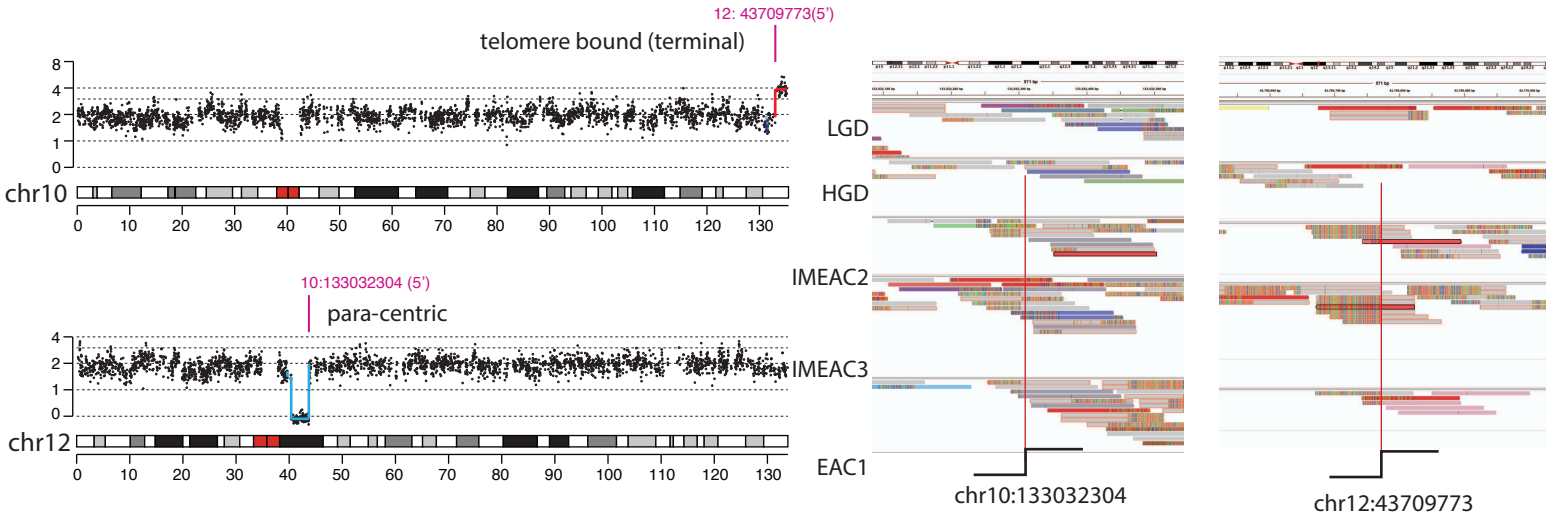
B



C

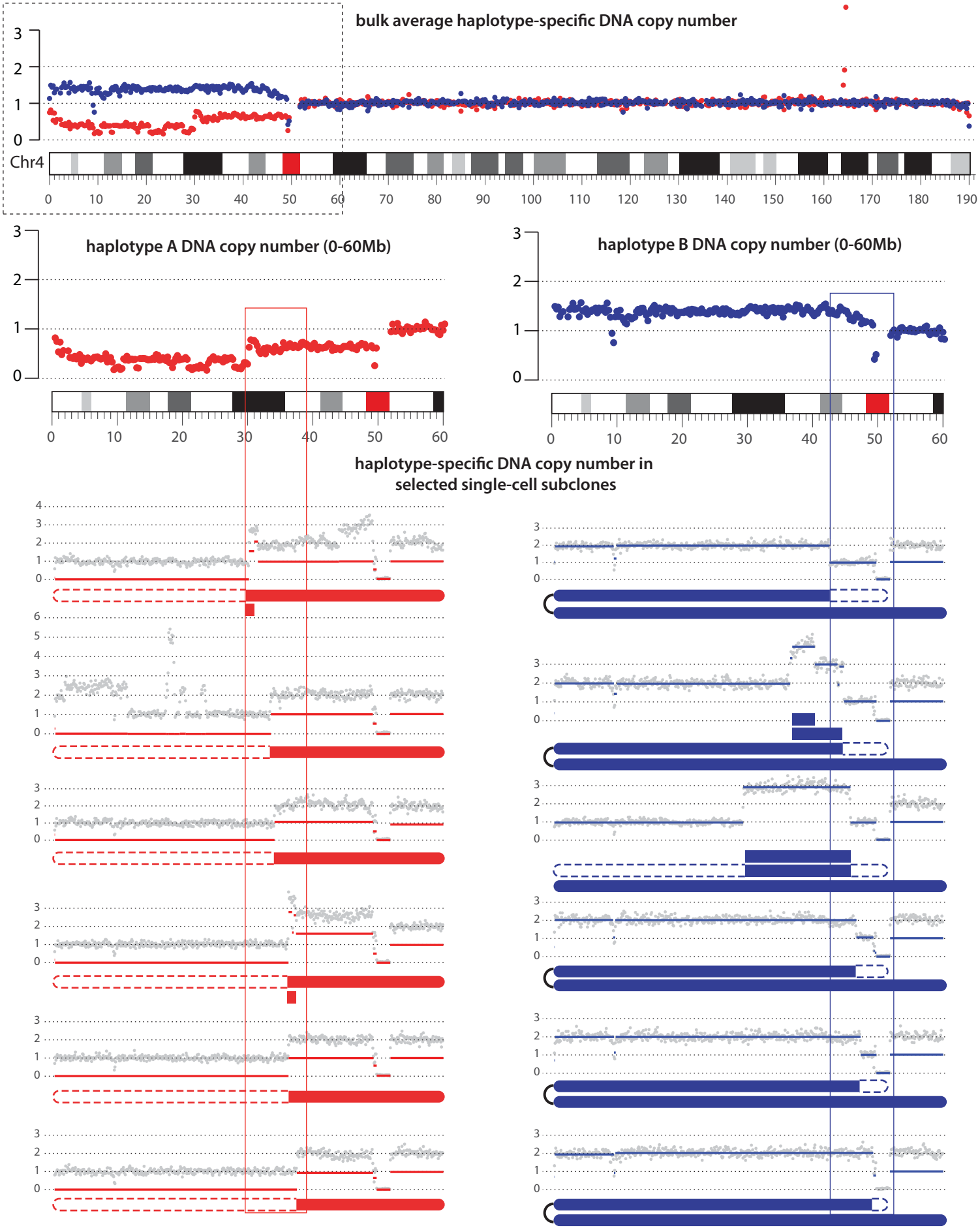


D



Sloping copy-number variation from in vitro BFB cycles

FIGURE S7



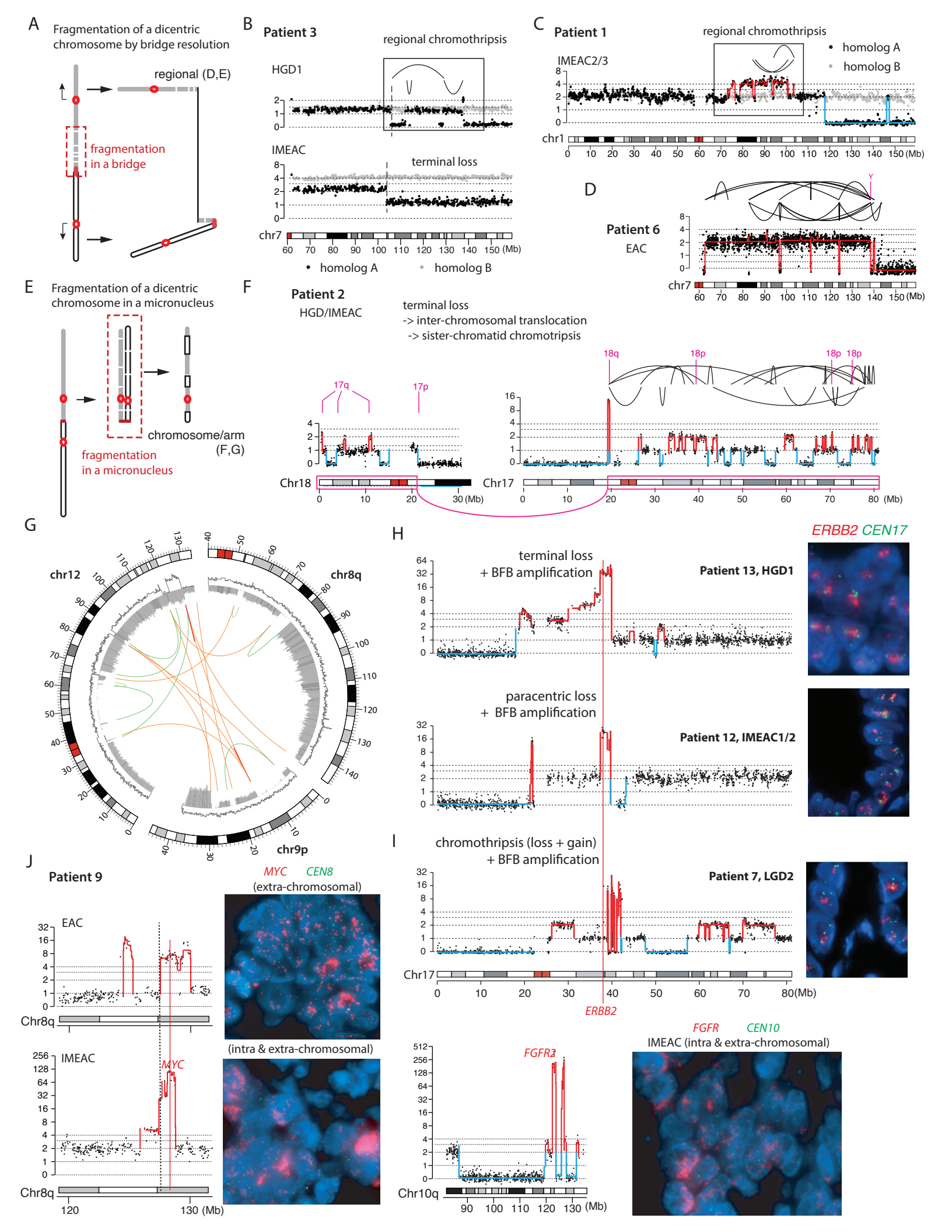
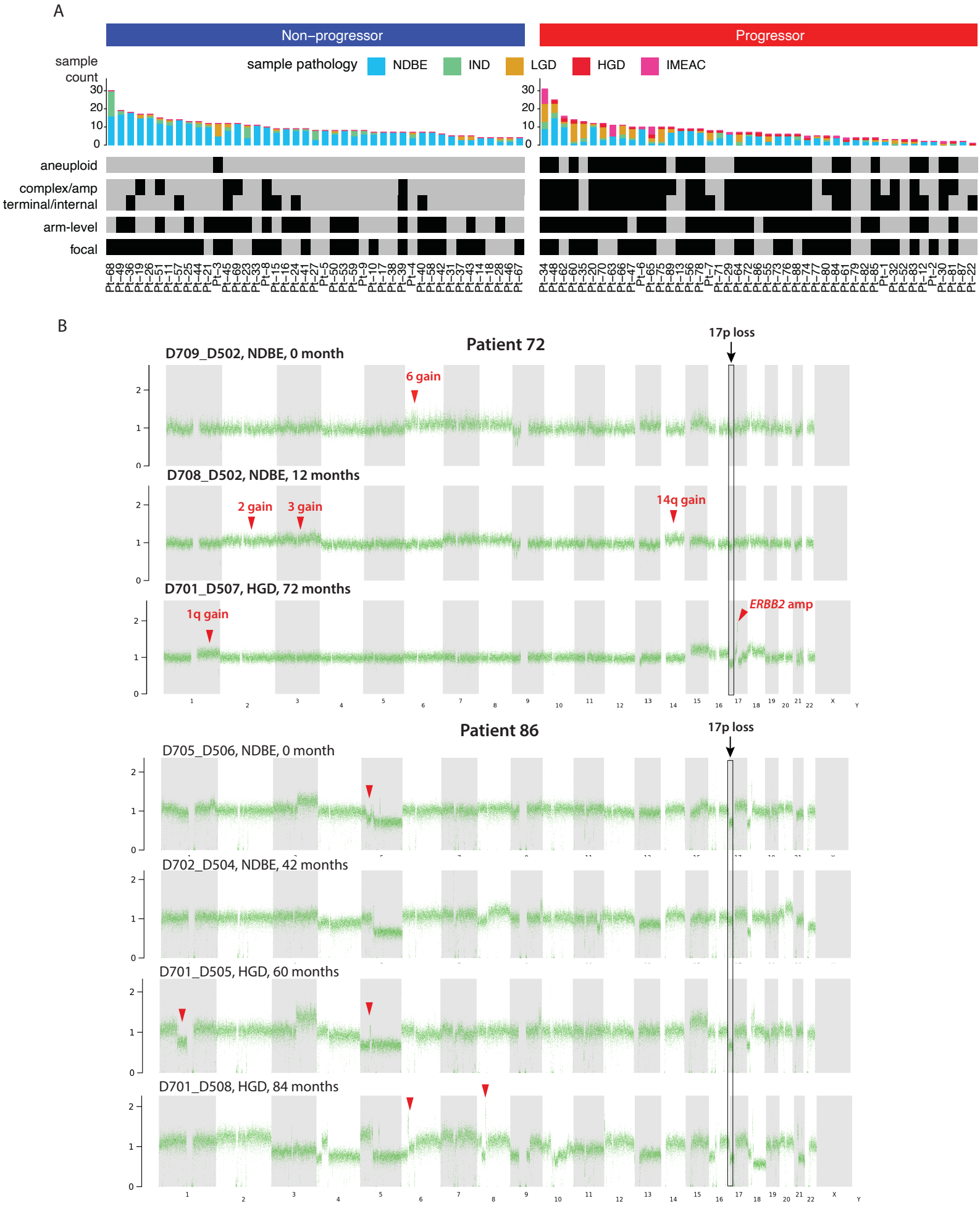

A
Fragmentation of a dicentric chromosome by bridge resolution
regional (D,E)
fragmentation in a bridge
B
Patient 3
regional chromothripsis
HGD1
IMEAC
terminal loss
chr7
homolog A
homolog B
C
Patient 1
regional chromothripsis
IMEAC2/3
homolog A
homolog B
chr1
D
Patient 6
EAC
chr7
E
Fragmentation of a dicentric chromosome in a micronucleus
chromosome/arm (F,G)
fragmentation in a micronucleus
F
Patient 2
HGD/IMEAC
terminal loss
-> inter-chromosomal translocation
-> sister-chromatid chromotripsis
17q
17p
18q
18p
18p
18p
Chr18
Chr17
G
chr12
chr8q
chr9p
H
ERBB2 CEN17
terminal loss
+ BFB amplification
Patient 13, HGD1
paracentric loss
+ BFB amplification
Patient 12, IMEAC1/2
I
chromothripsis (loss + gain)
+ BFB amplification
Patient 7, LGD2
Chr17
ERBB2
J
Patient 9
MYC CEN8
(extra-chromosomal)
EAC
Chr8q
(intra & extra-chromosomal)
IMEAC
MYC
Chr8q
FGFR CEN10
IMEAC (intra & extra-chromosomal)
FGFR2
Chr10q

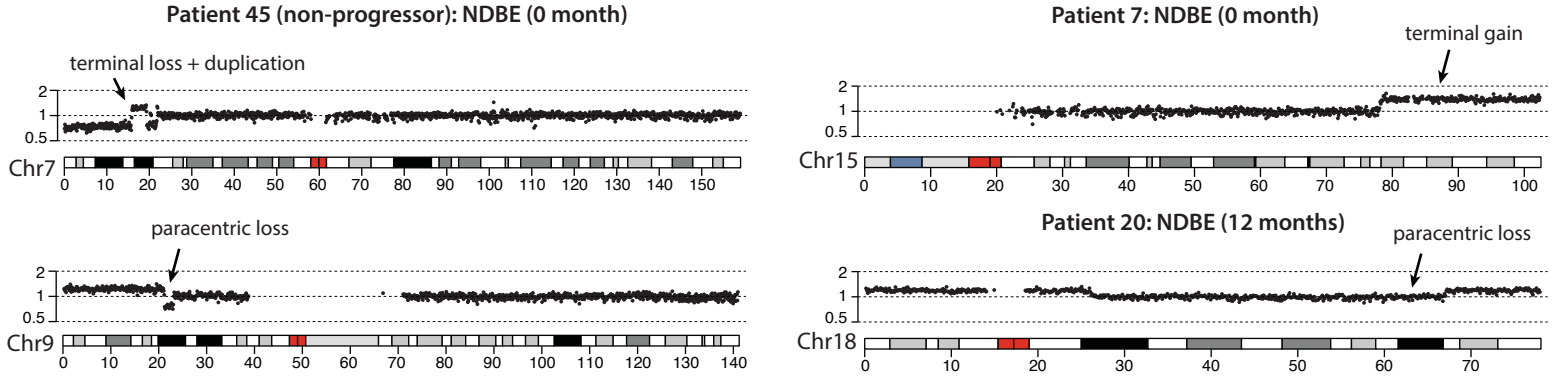
FIGURE S9


A
Non–progressor
Progressor
sample count
sample pathology
NDBE
IND
LGD
HGD
IMEAC
aneuploid
complex/amp
terminal/internal
arm-level
focal
B
Patient 72
17p loss
D709_D502, NDBE, 0 month
6 gain
D708_D502, NDBE, 12 months
2 gain
3 gain
14q gain
D701_D507, HGD, 72 months
1q gain
ERBB2 amp
Patient 86
17p loss
D705_D506, NDBE, 0 month
D702_D504, NDBE, 42 months
D701_D505, HGD, 60 months
D701_D508, HGD, 84 months

A

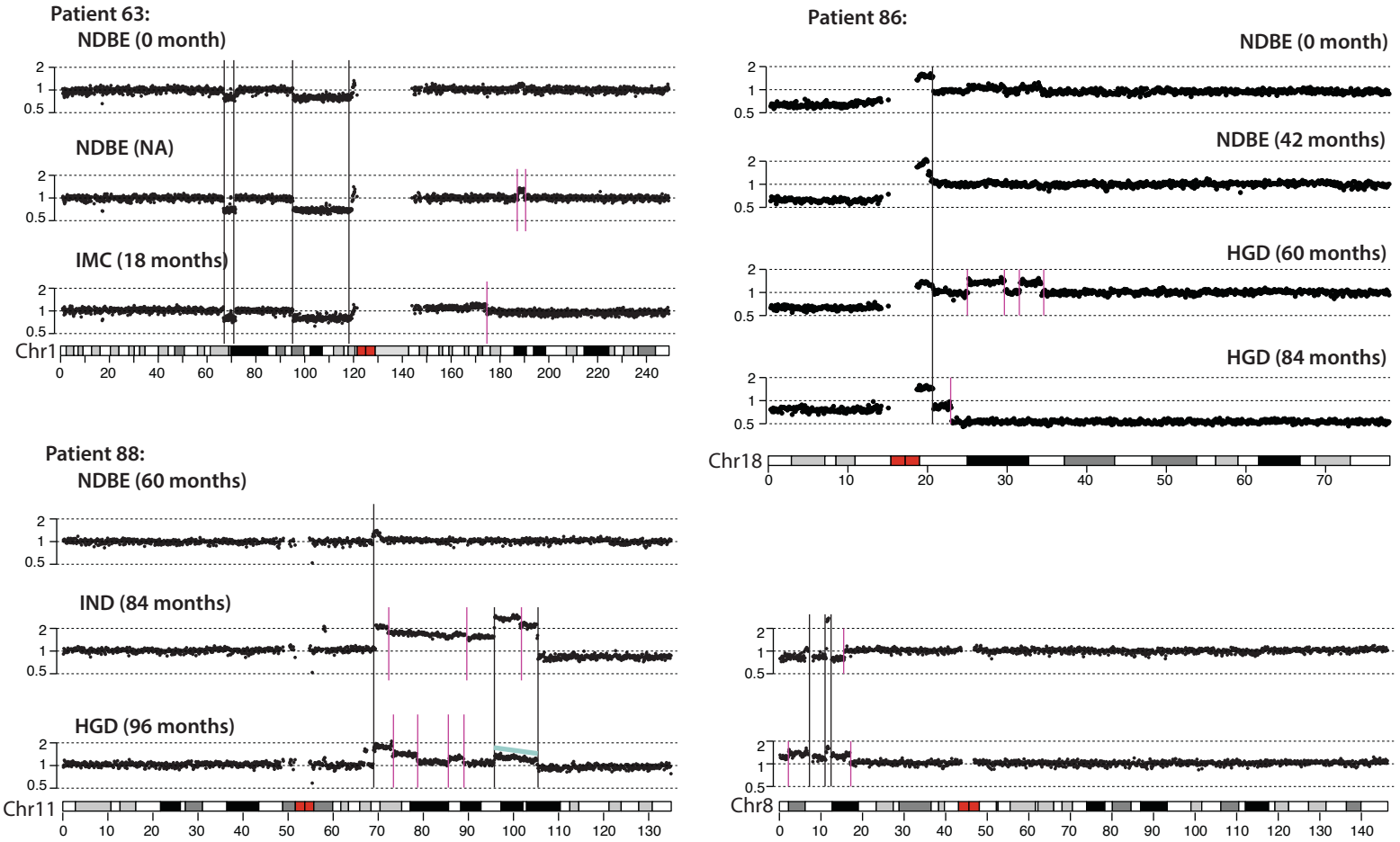
Large terminal and internal SCNAs

FIGURE S10



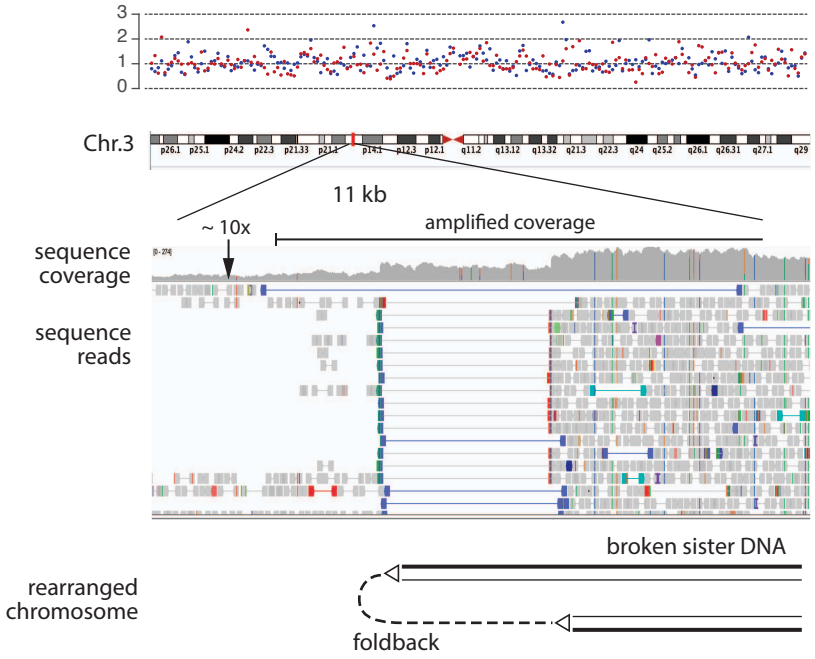
B

Divergent segmental SCNAs in longitudinal BE/EACs



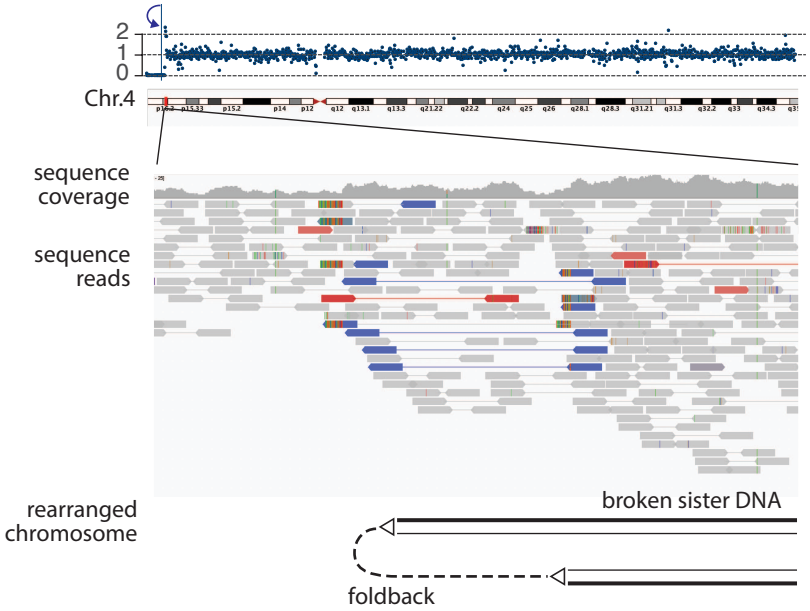
A Example of foldback rearrangements in a single cell that did not undergo breakage-fusion-bridge cycle

Additional Figure 1



This cell contained a micronucleus that was induced by nocodazole block, but did not undergo BFB cycle. The cell was sequenced to 10x mean coverage.

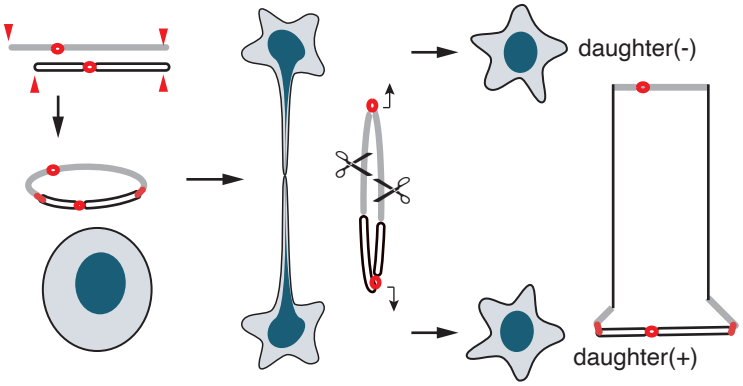
B Example of foldback rearrangements in a single-cell derived subclone that is consistent with BFB cycle

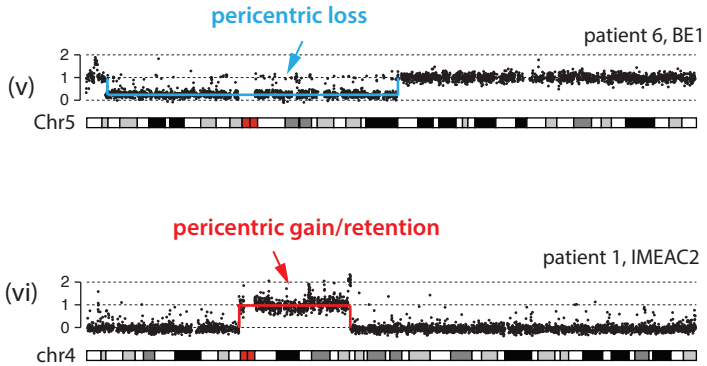
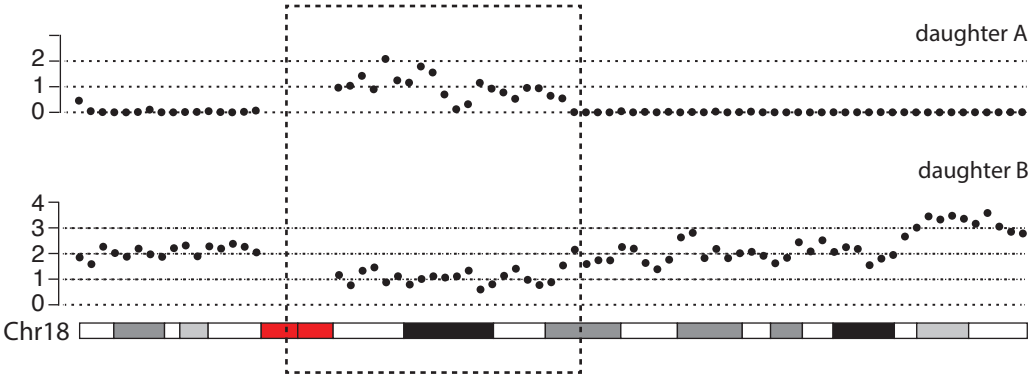


The primary reason for which the foldback rearrangement results from a BFB cycle is because the terminal duplication is right next to a terminal deletion, but not because of the foldback rearrangement itself.

Additional Figure 2

A





B