Title: Origins of cancer genome complexity revealed by haplotype-resolved genomic analysis of 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, 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: Herbert Irving Comprehensive Cancer Center, Division of Hematology and Oncology, Department of Medicine, Columbia University, NY 10032, United States

#,* These authors contributed equally

Correspondence to: Adam J. Bass (ab5147@cumc.columbia.edu) and Cheng-Zhong Zhang (cheng-zhong_zhang@dfci.harvard.edu)

Conflict of Interest Statement: A.J.B. has received research funding from Bayer, Merck and Novartis, is a consultant to Earli, and HelixNano and a co-founder of Signet Therapeutics. C.-Z.Zhang co-founded and serves as a scientific advisor to Pillar Biosciences.

Submitted content:

Manuscript text

Figures 1-7

Supplementary Figures 1-8

Tables S1-7

Downloadable Supplementary Data

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

Patient 1 (https://www.dropbox.com/s/w2qortyhnw6292b/Patient1_CNPlot.pdf?dl=0)

Patient 2 (https://www.dropbox.com/s/ub88obbgg99vdxh/Patient2_CNPlot.pdf?dl=0)

Patient 3 (https://www.dropbox.com/s/5uykg7fxyobprbi/Patient3_CNPlot.pdf?dl=0)

Patient 4 (https://www.dropbox.com/s/kyoq2rhvn2k59ya/Patient4_CNPlot.pdf?dl=0)

Patient 5 (https://www.dropbox.com/s/ngqq6psj723qg0q/Patient5_CNPlot.pdf?dl=0)

Patient 6 (https://www.dropbox.com/s/18x9ecs7k2z19ag/Patient6_CNPlot.pdf?dl=0)

Patient 7 (https://www.dropbox.com/s/m8jmfu8obvpv0r5/Patient7_CNPlot.pdf?dl=0)

Patient 8 (https://www.dropbox.com/s/on75pogh3311gb3/Patient8_CNPlot.pdf?dl=0)

Patient 9 (https://www.dropbox.com/s/j50d8londwomqk4/Patient9_CNPlot.pdf?dl=0)

Patient 10 (https://www.dropbox.com/s/hn3g1zortavimg6/Patient10_CNPlot.pdf?dl=0)

Patient 11 (https://www.dropbox.com/s/uoz6ammrqzalk6v/Patient11_CNPlot.pdf?dl=0)

Patient 12 (https://www.dropbox.com/s/dy0h85bo69maub9/Patient12_CNPlot.pdf?dl=0)

Patient 13 (https://www.dropbox.com/s/8lnjopmslvvw0ni/Patient13_CNPlot.pdf?dl=0)

Patient 14 (https://www.dropbox.com/s/5ptz07mg9bzyjnm/Patient14_CNPlot.pdf?dl=0)

Patient 15 (https://www.dropbox.com/s/p8wolef592mlrpt/Patient15_CNPlot.pdf?dl=0)

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

https://www.dropbox.com/s/u84u1zg02s8luwz/SingleCellDataPlot.pdf?dl=0

Abstract:

Complex chromosomal alterations are a hallmark of advanced cancers but rarely seen in normal tissue. The progression of precancerous lesions to malignancy is often accompanied by increasing complexity of chromosomal alterations that can drive their transformation through focal oncogenic amplifications. However, the etiology and evolution dynamics of these alterations are poorly understood. Here we study chromosomal copy-number evolution in the progression of Barrett's esophagus (BE) to esophageal adenocarcinoma (EAC) by multi-regional whole-genome sequencing analysis of BE samples with dysplasia and microscopic EAC foci. Through haplotype-specific copy-number analysis of BE genome evolution, we identified distinct patterns of episodic copy-number evolution consistent with the outcomes of abnormal mitosis and dicentric chromosome breakage. While abnormal mitosis, including whole-genome duplication, accounts for most chromosome or arm-level copy-number changes, segmental copy-number alterations display signatures of multi-generational evolution of unstable dicentric chromosomes. Continuous evolution of dicentric chromosomes through breakage-fusion-bridge cycles and chromothripsis rapidly increases genomic complexity and diversity among BE cells, culminating in the generation of distinct focal amplifications. These mutational processes enable multiple subclones within small dysplastic areas to undergo parallel transformation to cancer following acquisition of distinct oncogenic amplifications. Our results demonstrate how chromosomal instability drives clonal diversification in precancer evolution and promotes tumorigenesis in primary human samples.

Main Text:

Introduction

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 cancer genomic rearrangements(*4, 5*) and the biological mechanisms of genome instability(*6-9*), little is understood about what mechanisms are active in cancer evolution and how they generate cancer-promoting alterations such as focal amplifications.

Genomic analyses of normal tissues have revealed clonal expansion of point mutations but not of large structural chromosomal aberrations(*10, 11*). Early-stage premalignant lesions also show significantly less genome complexity than late-stage dysplasia or cancer(*12-16*). By contrast, the genomes of fully fledged cancers often show remarkable complexity reflecting the compound outcomes of many rounds of rearrangements(*4, 17, 18*). These observations have led to the prevailing view that most chromosomal rearrangements arise late during the progression from premalignant dysplasia to cancer and their acquisition appears to be episodic in contrast to gradual accumulation of short sequence variants (single-nucleotide substitutions or short insertions/deletions)(*19*).

A plausible explanation for the abrupt emergence of complex cancer genomes is that complex chromosomal changes arise from the rapid evolution of unstable chromosomes. For example, multi-generational evolution of dicentric chromosomes through breakage-fusion-bridge cycles and chromothripsis can generate a variety of genomic consequences, including arm-level and segmental copy-number changes, complex rearrangements, and focal amplifications(*20*). The generation of complex chromosomal changes through breakage or fragmentation of unstable chromosomes can further give rise to progeny cells with divergent or complementary copy-number changes(*20, 21*) that can give rise to "punctuated-equilibrium" copy-number evolution and genetic heterogeneity seen in many types of cancers(*22-25*). However, there has been no direct evidence from cancer genomes supporting this model.

A critical challenge in dissecting the evolution of chromosomal alterations is to gain knowledge of the intermediate steps of chromosomal changes before their footprints are masked by subsequent

events. To overcome this challenge, we have developed and applied novel experimental and analytical strategies to elucidate the evolution roadmap of early esophageal adenocarcinoma (EAC) from its precursor Barrett's Esophagus(*26-28*). Barrett's Esophagus (BE) is the replacement of the native squamous epithelium with columnar epithelial cells that develop intestinal differentiation(*29*). Instead of sampling fully formed EACs that likely have overgrown the most closely related precursors, we utilize a collection of predominately endoscopic mucosal resections (EMRs) (1-3cm in size) each consisting of regions of non-dysplastic BE, dysplasia, and early EAC (either intramucosal or T1) lesions isolated by laser capture microdissection. The application of haplotype-resolved genome analysis to this unique sample collection enables us to trace the evolution of individual parental chromosomes during disease progression and provides unprecedented resolution of the phylogeny of BE evolution.

Our analysis provides two lines of evidence relating unstable chromosomes to the generation of complex esophageal cancer genomes. First, we identify sequential evolution patterns of chromosomal and segmental copy-number changes in BE genomes that display hallmarks of genome instability due to whole-genome duplication and breakage-fusion-bridge (BFB) cycles of dicentric chromosomes. Second, we uncover extensive copy-number heterogeneity in premalignant BE cells, both at the subclonal level and at the single-cell level, that recapitulates the outcomes of branching evolution of unstable chromosomes. Notably, copy-number complexity and heterogeneity can predate the appearance of cancers or dysplastic lesions, but only becomes apparent at the clonal level after *TP53* loss. Therefore, *TP53* loss enables BE cells to acquire a mutator phenotype fueled by events such as whole genome duplication, chromothripsis, and BFB cycles, which not only generate SCNAs but further lead to elevated rates of secondary SCNA acquisition including focal amplifications. We further demonstrate that rapid accumulation of genome complexity and diversity in precancerous BE cells due to chromosomal instability can lead to multiple independent transformations within a single BE field and create markedly heterogenous tumors harboring distinct oncogenic amplifications. Together, these findings reveal the mutational processes that give rise to genomic complexity and diversity in precancer evolution and provide insight into the etiology of tumor heterogeneity and transition from premalignancy to cancer.

Results

Copy-number heterogeneity suggests early onset of chromosomal instability in precancer BE cells

Endoscopic mucosal resection (EMR) is generally performed in patients with dysplastic BE. In reviewing more than 500 EMR samples, we identified 14 cases showing unexpected microscopic foci of invasive cancers and one case (Patient 1) in which an early cancer was removed via esophagectomy. The cancers were either intramucosal or T1 and all samples were collected before treatment (chemo/radiotherapy or endoscopic ablation). Following independent pathologic re-review by two or more pathologists to confirm the diagnoses, we delineated and performed laser capture microdissection to isolate distinct regions from formalin fixed-paraffin embedded (FFPE) samples (**Figure 1**) corresponding to distinct histopathological states(*26*) including non-intestinalized columnar metaplasia (COLME), non-dysplastic BE (NDBE), BE indefinite for dysplasia (IND), BE with low-grade dysplasia (LGD) or high-grade dysplasia (HGD), and intramucosal (IMEAC) or early EAC (**Figure S1)**. We isolated normal tissue from either benign FFPE blocks or EMRs with non-epithelial normal tissue. The final collection consisted of 75 BE/EAC (21 COLME/NDBE/IND, 7 LGD, 23 HGD, and 24 cancer samples) and 15 reference samples.

We performed whole-genome sequencing on all the samples (mean sequencing depth 23x, **Table S1**) and joint variant detection on samples from each patient (**Methods**). To trace the evolution of somatic copy-number alterations (SCNA) to each homologous chromosome, we identified haplotype-specific copy-number breakpoints using the haplotypes of parental chromosomes. Parental haplotypes were first inferred by statistical phasing using a reference haplotype panel(*30*) and then refined based on allelic imbalance across all samples from each patient (**Figure S2**). The determination of long-range parental haplotype enabled phasing of segmental copy-number changes and copy-number breakpoints across entire chromosomes/arms. (The complete haplotype-resolved copy-number data are provided in **Supplementary data**.) For data presentation clarity, copy-number plots in the main and supplementary figures only show the copy number of the altered homolog, except where stated.

We detected single-nucleotide variants, short insertion/deletion variants, and chromosomal rearrangements using standard methods (**Methods**). The overall mutational burden is similar to those reported based on WGS in fresh BE and EAC tissue(*31-33*) (**Figure S3**). Due to technical artifacts generated in FFPE DNA library construction, we only included rearrangements with both breakpoints associated with haplotype-specific copy-number changes and used them for SCNA classification.

We determined the phylogenetic tree of all lesions from each patient (**Figure 2**) based on haplotype-specific copy-number breakpoints that were independently identified in each lesion; the phylogenetic relationship was then validated using shared single-nucleotide mutations (**Methods**). Individual SCNAs associated to each phylogenetic branch, including whole-genome duplication events (WGD, thick solid line), are summarized in **Table S2**. The branch length approximately reflects the number of altered chromosomes. For branches with WGD acquisition, we separated alterations inferred to have occurred pre- (gray branch) and post-WGD (black branch). In each phylogenetic tree, we labelled alterations to genes that are recurrently altered in esophageal cancers and further annotated parental chromosomes that acquired divergent segmental copy-number alterations in different progeny clones indicating branching evolution (These chromosomes are labelled with asterisks near the common ancestor).

The phylogenetic trees of EAC and precursor BE lesions show several recurrent patterns. First, bi-allelic *TP53* inactivation is a truncal event of the evolutionary branches that eventually lead to cancer (14/15 patients). By contrast, local segmental deletion near *FHIT* (a common fragile site) or bi-allelic inactivation of *CDKN2A* (a frequently inactivated tumor suppressor) are not exclusive to the cancer branch, often occurring exclusively within a branch unrelated to the cancer sample(s). Second, the number of SCNAs between clonal fixation events (evolutionary branch length) shows significant variation; in particular, branches with the highest SCNA burden are frequently associated with WGD, which is itself also a frequent event (10/15 patients). Third, high-grade dysplastic BE lesions and cancer lesions from the same patient often harbor distinct complex copy-number alterations (13/15 patients) or distinct regions of focal amplification (10/15 patients), suggesting clonal expansion of independently

evolved complex chromosomal changes. Finally, we identified more than one early cancer lesion in 5/15 patients; different cancer foci from the same patient often show significant genomic divergence but are individually accompanied by genomically similar precancer lesions. This result suggests independent EAC transformations, *i.e.*, more than one genomically distinct BE cell became invasive and formed independent cancers.

The above observations suggest highly dynamic genome evolution in precancerous BE cells driven by chromosomal instability. We expected such evolution to also generate copy-number diversity at the single-cell level and directly tested this hypothesis by single-cell genomic analysis of pre-EAC tissue. We isolated 68 single cells from a patient with known HGD via endoscopic cytology brushing immediately before radiofrequency ablation and performed haplotype-specific copy-number analysis and phylogeny construction by the same strategy used for bulk samples (**Methods**). The final phylogeny is shown in **Figure 3A** with chromosomal copy-number changes annotated for each branch (**Table S3**). We identified 12 cells with aneuploid genomes and 56 cells with a mostly diploid genome. In comparison to diploid cells, the aneuploid cells share bi-allelic *TP53* loss through a pathogenic R175H mutation and loss of heterozygosity generated by 17p loss but show significant heterogeneity of chromosomal copy-number changes. The onset of genome heterogeneity characterized by chromosomal alterations in precancer BE cells following bi-allelic *TP53* inactivation recapitulates the pattern seen in bulk samples and provides direct evidence of dynamic precancer genome evolution driven by chromosomal instability. We next discuss specific patterns of copy-number evolution and their mechanistic implications.

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

We identified bi-allelic *TP53* inactivation as a truncal event of the evolutionary branches that eventually lead to cancer in all but one patient (Patient 9). Samples with intact *TP53* showed significantly less SCNA burden than samples with *TP53* inactivation (**Figure S4A, Table S2**). The connection between *TP53* inactivation and SCNA acquisition is further supported by the disparity of SCNA burden in single cells (**Figure 3A**) and in bulk NDBE and LGD clones with and without intact p53 (**Figure S4A)**. Two NDBE

samples (in Patient 6 and 15) and four LGD samples (in Patient 6 and 7) with bi-allelic *TP53* inactivation show similar SCNA burdens as HGD and EAC samples; by contrast, NDBE and LGD samples without *TP53* inactivation show significantly fewer SCNAs (**Figure 4A,** left and middle). These data reinforce the model that *TP53* inactivation promotes BE progression through copy-number evolution(*14, 34*).

Uniparental disomy (UPD), or copy-neutral loss of heterozygosity, is the only type of large segmental SCNAs with frequent clonal presence in non-cancerous esophageal tissue(*10, 11*). We observed frequent UPDs in both single BE cells (**Table S3**) and clones (**Table S4**). Remarkably, UPD events are more prevalent in lesions with intact p53 (**Table S4**), which contrasts with the significantly higher burden of other large SCNAs in lesions with inactive p53 (**Figure 4A**, middle). We further identified an intriguing pattern of varying UPD on the 9p terminus in a subclone of 14 single BE cells (**Figure 3D**), indicating a new form of genetic heterogeneity that can only be revealed by haplotype-resolved copy-number analysis. The varying boundaries of terminal UPD in different cells (arrows in **Figure 3D**) bear a striking similarity to our prior observation of varying boundaries of terminal deletions attributed to ongoing breakage-fusion-bridge cycles seen in *in vitro* studies(*20*). The similarity between varying terminal UPD and varying terminal deletion suggests that they could have both originated from broken chromosomes with varying terminal deletion generated by breakage-fusion-bridge cycles: The prevalence of UPDs in cells with intact p53 could reflect selection for stable chromosomes with intact telomeres resulting from homologous recombination between the broken chromosome and the intact homolog.

Contrasting the simple SCNA landscape in BE cells or lesions with intact p53 is the prevalence of arm-level and complex SCNAs in BE cells or lesions with *TP53* inactivation (**Figure 3A,F-H**, **Figure 4A,** middle). Therefore, p53-dependent suppression of chromosomal instability(*35*) accounts for the absence of large SCNAs in normal tissue and its abrogation in BE cells enables clonal expansion and accumulation of both chromosomal and segmental SCNAs even in the absence of histologic dysplasia. We next turn our focus to the evolutionary dynamics of large SCNAs during BE progression.

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

Whole-genome duplication (WGD) is inferred to be a frequent event in many epithelial cancers(*5*) and thought to define a particular EAC evolution trajectory(*34*). Based on haplotype-specific copy number data, we inferred the presence of 15 WGD events in bulk BE/EAC lesions from 10/15 patients (**Figure 2**). Moreover, based on the phylogeny of BE/EAC lesions, we observed that WGD events can arise at different times during BE progression (early/truncal:4/15; intermediate: 4/15; late 2/15) and arise independently in separate evolutionary branches (3/15 patients). We even identified two independent WGDs in single BE cells from a dysplastic lesion without presence of cancer (**Figure 3A**). The prevalence of WGD events throughout BE progression suggests that its spontaneous rate during tumorigenesis is significantly higher than its prevalence inferred from terminal cancer genomes.

Despite the prevalence of WGD in human cancers(*36, 37*) and multiple lines of evidence supporting a tumor-promoting role of WGD(*38, 39*), how WGD impacts tumorigenesis remains incompletely understood. One proposal is that cytokinesis failure or tetraploidization can precipitate additional genome instability including multipolar cell division or chromosome mis-segregation(*6, 40, 41*) that leads to additional chromosomal copy-number changes. Consistent with this model, we observed that WGD is frequently followed with the acquisition of extensive segmental or whole-chromosomal SCNAs (**Figure 4A**, right). In particular, the evolution branches with WGD acquisition show significantly higher SCNA burden (mostly post-WGD) than non-WGD branches (**Figure 4B**). The observation of similar SNV burden but significantly higher SCNA burden in WGD branches than non-WGD branches (**Figure S4E**) suggests an episodic increase in SCNA acquisition rates relative to SNV accumulation that more closely follows the number of cell divisions.

We noticed that post-WGD SCNAs are dominated by chromosomal loss (both arm-level and whole-chromosome): among 22 samples inferred to have undergone one round of WGD, 208 chromosomes (or chromosome arms) were present at the allelic single-copy state, indicating chromosomal loss after WGD; by contrast, only 46 chromosomes were present at the triple-copy state that is most parsimoniously explained by chromosomal gain after WGD. The preponderance of chromosomal

loss after WGD was also observed in single aneuploid BE cells (**Figure 3A**). The disparity between chromosome loss and gain after WGD has two implications. First, this pattern cannot be solely explained by increased rates of random chromosome mis-segregation(*40*), but most likely reflect multi-polar cell divisions that predominantly causes chromosome loss (**Figure S5A**). Future work is needed to test this model. Second, extensive chromosome loss may obscure the signature of the ancestral WGD event and cause underestimation of WGD prevalence.

In summary, our analysis of arm-level copy-number evolution suggests that arm-level copy-number changes often accumulate in episodic bursts that are consistent with the outcome of WGD and downstream events including multi-polar cell divisions. These mutational processes can generate progeny cells of highly aneuploid genomes with distinct chromosomal copy-number changes that eventually become tumor subclones with significant genomic differences.

Segmental copy-number changes display signatures of dicentric chromosome evolution

In contrast to the significant increase of chromosomal copy-number changes after WGD, we observed similar burdens of segmental SCNAs before and after WGD (**Figure 4A**, right). This result is consistent with observations in pancreatic cancer evolution(*25, 41*) and suggests that segmental SCNAs are generated by different mutational processes than arm-level SCNAs.

We recognized that many segmental copy-number changes observed in BE and EAC genomes are consistent with the copy-number outcomes of dicentric chromosome breakage. Dicentric chromosomes can result from end-to-end chromosome fusion after telomere erosion or recombination between broken chromatids. The most common copy-number patterns resulting from breakage of a single dicentric chromosome are telomere-bound copy-number changes (**Figure 5A,S5B**)(*20*). We identified 84 instances (41 gain, 43 loss) of such alterations (hereafter referred to as “terminal” SCNAs). Breakage of sister dicentric chromatids can give rise to large segmental gain or loss within a chromosome arm (**Figure 5B**)(*20*). We identified 53 instances (14 gain, 39 loss) of this pattern (hereafter referred to as “paracentric SCNAs”). We further observed five instances of SCNAs (2 gain, 3 loss) that span the centromere

("pericentric SCNAs"). This pattern can result from breakage of ring chromosomes (**Figure 5C**) or multi-centric chromosomes. (The observations of terminal, paracentric, and pericentric SCNAs are summarized in **Table S5.)** The identification of rearrangements joining terminal/paracentric/pericentric SCNA breakpoints on different chromosomes (**Figure S6**) suggests that these events were generated at the same time and likely by a similar mechanism due to the formation of dicentric chromosomes.

The mechanistic connection between terminal/paracentric/pericentric SCNAs and chromosome bridge resolution is further supported by the observation of "compound" copy-number patterns generated by sequential breakage-fusion-bridge (BFB) cycles after the broken chromatids in daughter cells form new dicentrics (**Figure S5C**). The simplest example is gain of a terminal segment at the break end either next to a terminal loss (**Figure 5D**, 18 instances) or next to an internal loss (11 instances). We further observed examples of two sequential gains at a broken end (**Figure 5E**, 4 instances) or focal amplifications after multiple sequential gains (**Figure 5F,** 8 instances) over generations of BFB cycles.

Besides copy-number patterns generated by sequential BFB cycles, we further observed 17 instances of divergent copy-number alterations to a single ancestral chromosome in related BE or EAC lesions that are consistent with branching evolution of BFB cycles. In eight instances, the SCNAs in related lesions show the same pattern (terminal or paracentric) of alterations but with distinct breakpoints; in seven instances, the SCNAs in related lesions share copy-number breakpoints consistent with ancestral terminal/paracentric alterations but contain distinct downstream alterations. Both of these patterns were observed in the *in vitro* evolution of broken chromosomes(*20*). One representative example is shown in **Figure 5G** of the copy-number profiles of Chr.2 in the LGD and EAC lesions from Patient 6. Both genomes share an ancestral paracentric copy-number loss followed by terminal gains near the centromeric break end, but the EAC genome had acquired additional copy-number gain near the telomeric break end: the varying copy-number further indicates a polyclonal mixture that was also seen *in vitro*(*20, 42*).

Finally, we observed opposite copy-number changes at the same breakpoint on two chromosome arms (7q and 17q) in two related lesions in Patient 4 (the example on Chr.7q is shown in **Figure 5H**). This copy-number pattern exactly recapitulates the outcome of complementary broken pieces of a single

dicentric chromosome (**Figure 5A**). The most plausible explanation is that the HGD and IMEAC clones descended from a pair of daughter cells each having inherited complementary pieces of a dicentric Chr.7 after a single breakage event. Remarkably, the IMEAC genome also contains an amplified copy-number pattern at the distal terminus of 7q with the signature of BFB amplification; both this amplification and the reciprocal distribution of 7q fragments into the ancestors of HGD and IMEAC were likely generated by continuous BFB cycles initiated by an ancestral terminal loss.

In summary, segmental copy-number alterations in BE/EAC genomes display multiple signature copy-number patterns of ongoing breakage-fusion-bridge (BFB) cycles initiated by the formation and breakage of dicentric chromosome bridges. These patterns include (1) terminal, paracentric, and pericentric SCNAs that can result from one BFB cycle; (2) complex copy-number patterns including focal amplifications resulting from sequential BFB cycles; and (3) branching copy-number patterns in different progeny clones generated by continuous evolution of BFB cycles. These results suggest the formation and evolution of dicentric chromosomes as a major mutational process of segmental SCNA acquisition.

Contemporaneous breakage-fusion-bridge cycles and chromothripsis create complex copy-number changes including focal amplifications

Complex chromosomal alterations are a prevalent feature of esophageal cancer genomes(*43*) and also account for a large fraction of segmental copy-number alterations in cancer genomes(*4*). We identified 76 instances of complex segmental copy-number patterns (including chromothripsis) containing at least two SCNAs on a chromosome arm that cannot be explained as combinations of terminal, paracentric, or pericentric SCNAs generated by sequential BFB cycles (**Table S6**). These events were identified in 14 out of 15 patients and mostly present in HGD or cancer lesions. Three instances of chromothripsis resulting in focal deletions spanning the *CDKN2A* locus on 9p were observed in non-dysplastic (Patient 8, **Figure 6A**) and LGD (Patient 11) lesions with intact *TP53* or inferred to have occurred prior to *TP53* inactivation (Patient 6), indicating that complex chromosomal changes can arise and undergo rare clonal expansion even with intact p53.

Of the 76 instances of complex copy-number changes, 26 display oscillating copy-number patterns between two or three states that span entire chromosome arms (**Figure 6A**), including three instances of complementary oscillating copy-number patterns in different BE/EAC lesions from the same patient. We inferred these patterns to be standalone chromothripsis and may have arisen from micronuclei (**Figure S5D**)(*21*). The remaining 50 instances show both oscillating copy-number patterns and additional segmental SCNAs with features of BFB cycles. Although the connection between chromothripsis and BFB cycles was both implicated by observations in cancer genomes(*44, 45*) and corroborated by *in vitro* experiments(*20, 42, 46*), a complete survey of copy-number signatures associated with different evolution sequences of chromothripsis and BFB cycles in cancer genomes is lacking. Here we address this question by relating copy-number patterns revealed from *in vitro* experiments (**Figure S5C, 6B, 6C**) to those observed in BE/EAC genomes and elucidate the etiology of complex copy-number changes in EAC genomes.

We identified 26 instances of complex copy-number alterations consistent with the evolution sequence of chromothripsis arising secondary to dicentric chromosome formation. When chromothripsis is generated directly from breakage of a dicentric bridge chromosome, only the chromosomal segment in the bridge undergoes fragmentation(*20*). The oscillating copy-number pattern generated by this mechanism will be restricted to a fraction of the chromosome arm ("regional") and accompanied by terminal or internal SCNAs from asymmetric distribution of the dicentric chromosome (**Figure 6B**) (*20, 42*). Chromothripsis can also result from fragmentation of dicentric chromosomes partitioned into micronuclei(*20*). This process can generate oscillating copy-number patterns extending to whole chromosome arms ("chromosome/arm") (**Figure 6C**)(*20, 42*). Examples corresponding to each scenario (**Figure 6D,E** for regional chromothripsis, **Figure 6F,G** for arm-level chromothripsis) are described below.

Representative examples of regional chromothripsis arising directly from dicentric chromosome breakage (**Figure 6B**) include Chr.7q in the HGD1 genome from Patient 3 (oscillating copy number next to the boundary of terminal loss), Chr.15q in the IMEAC genome from Patient 3 (oscillating copy number

in the region of terminal gain), Chr.7q in the IMEAC genomes from Patient 1 (oscillating copy number restricted to the region of paracentric gain), and Chr.5q in the HGD2/IMEAC genomes from Patient 3 (oscillating copy number next to the boundary of paracentric loss) (**Supplementary Data**). We further observed examples of regional chromothripsis with evidence of preceding terminal or internal SCNAs based on the timing of SCNA breakpoints. In the example shown in **Figure 6D**, the chromothripsis event that generated the single-copy oscillating pattern was inferred to have occurred after the terminal deletion with a two-copy difference; both the chromothripsis and the earlier terminal deletion were likely generated by BFB cycles involving a dicentric Chr.14. In another example shown in **Figure 6E**, the ancestral paracentric loss shared by all three genomes (LGD2/HGD3/EAC) was followed by regional chromothripsis and BFB amplifications near the centromeric break end in the LGD2 genome and a simple terminal duplication near the telomeric break end in the EAC genome; the branching evolution pattern suggests that both private alterations were downstream events of an ancestral chromosome-type BFB cycle that generated the paracentric loss.

We found at least six instances of chromothripsis involving whole chromosome arms that likely resulted from DNA fragmentation in micronuclei (**Figure 6C**). In the example shown in **Figure 6F**, both 17q and 18p show three-state oscillating copy-number patterns (CN=0,1,2) and also contain multiple inter-chromosomal rearrangements, indicating origin from a single chromothripsis event. The presence of multiple foldback rearrangements in the focally amplified region on 17p implies preceding BFB cycles leading to 17p-terminal loss. Taking these together, we infer the following evolution sequence giving rise to the observed alterations: (1) multiple BFB cycles created a broken Chr.17 (missing p-terminus); (2) the broken Chr.17 joined a broken Chr.18 (missing q-terminus) to form a dicentric chromosome; (3) the dicentric chromosome was partially replicated and underwent chromothripsis to create three-state oscillating copy-number patterns seen on both arms. In another example shown in **Figure 6G**, we identified one copy-number breakpoint on 5p (dotted line) associated with opposite copy-number changes in the EAC genome and the IMEAC genome, suggesting that they were both derived from one broken chromatid; the observation of a paracentric loss in the IMEAC genome indicates this breakage to be

generated by a chromosome-type BFB cycle (**Figure 5B**). We therefore infer that the chromothripsis event generating three-state oscillating copy-number pattern (CN=0,2,4) on Chr.5q in the EAC genome was a downstream event (likely direct consequence) of the ancestral chromosome-type BFB cycle. Additional examples reflecting this evolution sequence include Chr.5 and Chr.11 in Patient 4, Chr.10 in Patient 6 and Chr.6 in Patient 8 (**Supplementary Data**), all of which show oscillating copy-number patterns spanning both chromosome arms accompanied by ancestral terminal/internal SCNAs reflecting dicentric chromosome breakage. Together, these observations suggest dicentric chromosome formation as a frequent initiating event leading to chromothripsis of both sub-chromosomal regions and entire chromosome arms.

Chromothripsis is also a frequent event leading to sequence duplication or amplification through subsequent BFB cycles or extra-chromosomal DNA formation(*44*, *45*, *47*, *48*). We identified 32 instances of complex copy-number patterns reflecting this evolution sequence, including eight instances where chromothripsis was also preceded by terminal/internal SCNAs generated by dicentric chromosome breakage. Two representative examples are shown in **Figure 6H**. In the first example (Chr.1p), we infer that an ancestral chromothripsis event generating the copy-number breakpoint shared by the EAC genome and the IMEAC genome (dotted line) was followed by branching evolution of BFB cycles creating one focally amplified region in the IMEAC genome and multiple focally amplified regions in the EAC genome, both with multiple intermediate copy-number states consistent with BFB amplification. In the second example (Chr.16p), we inferred that the terminal loss in the IMEAC genome and the complex amplified copy-number pattern in the EAC genome were both initiated by the formation of a dicentric Chr.16, with chromothripsis and subsequent BFB amplification in the EAC genome as secondary events.

In summary, we were able to explain all complex copy-number patterns observed in BE/EAC genomes based on the copy-number outcomes of different evolution sequences of BFB cycles and chromothripsis derived from *in vitro* experiments. Together with observations of terminal/internal SCNAs reflecting simple copy-number outcomes of BFB cycles, our results elucidate how the evolution of dicentric chromosomes with intermittent chromothripsis and segmental gain and loss can explain the

generation of segmental copy-number complexity in BE evolution. Based on this insight, we subdivided segmental SCNAs into three categories (simple, complex, and focal amplification) and showed their relative prevalence in BE evolution in **Figure S4**. The appearance of all three categories of SCNAs after *TP53* inactivation indirectly supports their common etiology from unstable dicentric chromosomes.

Focally amplified regions continue to evolve during BE progression to EAC

Gene amplification is a common mechanism of oncogenic activation in EAC(*33, 34*). In total, we identified 45 focally amplified regions (allelic copy number ≥ 8) each spanning one or multiple loci on a chromosome (**Table S7**). The focally amplified regions often display multiple intermediate copy-number states consistent with BFB cycles (38 cases) and are frequently accompanied by interspersed deletions or oscillating copy-number patterns indicating chromothripsis (27 instances). We further identified two focally amplified regions that are exclusive to the NDBE lesion in Patient 6 with *TP53* inactivation. These observations suggest that focal amplifications can arise spontaneously after *TP53* inactivation and do not always promote disease progression.

Due to the dynamic nature of chromothripsis and BFB cycles that are related to focal amplifications, we expect focal amplifications to arise and evolve continuously during BE evolution. This implies that (1) amplifications that become clonal are likely driven by positive selection and contain oncogenes; and (2) different clones can acquire different amplifications. Consistent with the first prediction, we found that 29 focally amplifications overlap with regions that are recurrently amplified in esophageal cancers or other cancers and 24 encompass putative oncogenes. The significance of focal amplification as a mechanism of oncogenic activation is further underscored by the observation of independent amplifications of EAC oncogenes including *ERBB2* on 17q (5/15 patients) and *GATA6* on 18q (4/15 patients) among others. Consistent with the second prediction, we identified focal amplifications that are private to one or a subset of BE/EAC lesions in 10 out of 15 patients (**Figure 2** and **S7**), including focal amplifications of *IGF1R* (Patient 3), *MET* (Patient 4), and *KRAS* (Patient 10) that are exclusive to the cancer samples. We also identified distinct focal amplifications in different cancer lesions

within individual patients (Patients 1, 9, and 15) that were likely promoting the independent transformation events.

The dynamic nature of focal amplification is also reflected in the observation of focally amplified regions with distinct copy-number breakpoints (six instances) or copy-number levels (ten instances) in related BE/EAC lesions from the same patient. In three patients, we found that the cancer lesion had acquired additional gains of oncogenic amplifications already existing in the precursor lesion (*GATA4* in Patient1, *ERBB2* and *GATA6* in Patient 7, *CCND1* in Patient 11) (**Figure S7**, **Table S6**), suggesting gain of fitness with the dosage of amplification(*47, 49*).

In summary, our analysis of focal amplifications in BE and EAC genomes recapitulates their connection to chromothripsis and BFB cycles suggested by *in vitro* experiments(*20, 47, 48*) and further demonstrates how continuous genome evolution leads to spontaneously arising oncogenic amplifications that promote BE progression, generate genetic heterogeneity, and fuel independent transformation of dysplastic lesions.

Discussion

We here studied the evolutionary dynamics of chromosomal copy number changes and rearrangements using a unique sample set of incipient EACs and paired BE lesions, including both samples that are proximal and thus more closely related to the cancer and those that are more distantly related to the cancer. Through haplotype-specific copy-number analysis, we first determine the phylogeny of BE genome evolution, then dissect the dynamics of chromosomal and segmental copy-number changes, and further identify recurrent evolutionary patterns that give rise to rapid karyotype changes and complex chromosomal alterations including focal amplifications that are a major source of oncogenic signaling in many cancer types.

We find that arm-level copy-number changes often accumulate in episodic bursts and are consistent with the outcome of whole-genome duplication (WGD) followed by extensive chromosome loss generated by downstream events including multi-polar cell division and micronucleation.

Interestingly, we identified two instances of near complete genome duplication: the genome of the EAC lesion from Patient 7 is a near complete duplication of the LGD2 genome (with two post-WGD SCNAs); the D5 cell in the single-cell collection is close to a complete duplication of the F12 cell (with four post-WGD SCNAs). This observation raises the question of how duplicated genomes re-establish stable karyotypes *in vivo*, which may reveal new genetic dependencies associated with aneuploid genomes.

We identify multiple patterns of segmental copy-number alteration and evolution in BE/EAC genomes that are consistent with the formation and evolution of dicentric bridge chromosomes revealed by *in vitro* experiments(*20*). Multiple lines of evidence strongly support breakage-fusion-bridge (BFB) cycles and chromothripsis of dicentric chromosomes as a major mechanism generating segmental copy-number complexity in BE/EAC genomes. First, many segmental copy-number changes in single BE/EAC genomes are consistent with the outcomes of a single breakage-fusion-bridge cycle: terminal, paracentric, or pericentric SCNAs, sometimes in combination with regional chromothripsis. Second, we identified many instances of complex copy-number patterns in single BE/EAC genomes corresponding to compound SCNAs generated by consecutive BFB cycles, including BFB amplifications. Third, we observed multiple instances of divergent copy-number patterns of a single parental chromosome in distinct BE/EAC genomes from the same patient reflecting branching evolution of BFB cycles and chromothripsis. In particular, we observed complementary copy-number changes that exactly recapitulate reciprocal distribution of a single broken dicentric chromosome into daughter cells in one BFB cycle. Both the sequential and branching patterns of copy-number evolution recapitulate findings from *in vitro* evolution(*20, 42*) and highlight the dynamic nature of BFB cycles and chromothripsis first described in cancer genomes(*44, 45*). Finally, we identified inter-chromosomal rearrangements joining breakpoints associated with different copy-number outcomes associated with chromosome bridge resolution (**Figure S6**). The most parsimonious explanation is that these breakpoints all result from the resolution of a single chromatin bridge consisting of multiple dicentric chromosomes generated by telomere crisis(*42, 50, 51*).

We observe that bi-allelic *TP53* inactivation is a near ubiquitous early event preceding aneuploidy and complex chromosomal abnormalities, both at the single-cell level and in clonal

populations. This result confirms the evolution sequence of early *TP53* inactivation during the progression of BE to EAC inferred from comparative studies of BEs and late EACs (*14, 33, 34, 43, 52, 53*). The strong association between chromosomal aberrations and *TP53* inactivation further suggests that the suppression of cell proliferation after chromosome mis-segregation (*54*) may be the dominant tumor suppressive mechanism in BE progression. Interestingly, uniparental disomy (UPD) alterations appear to be more prevalent in samples with intact *TP53* (both clonal populations and single cells) than in samples with inactive *TP53*. The observation of UPDs with varying boundaries in single BE cells that resemble the pattern of terminal loss observed *in vitro* after dicentric chromosome induction(*20*) indicates that UPDs may have resulted from broken chromosomes generated by BFB cycles that undergo homologous recombination with the intact homolog (**Figure S8**).

By demonstrating how specific instigating chromosomal processes, including tetraploidization and telomeric loss followed by dicentric chromosome formation, trigger chromosomal instability and episodic SCNA accumulation, we provide a mechanism-based explanation of episodic accumulation of copy-number complexity and heterogeneity. This model has two implications for BE progression. First, branching evolution of unstable chromosomes or genomes rapidly increases genomic diversity in BE cells and subclonal heterogeneity in EAC. Second, continuous shuffling and reciprocal distribution of DNA sequence by chromothripsis and BFB cycles can efficiently sample the space of all possible segmental copy-number alterations and generate focal amplifications of oncogenes that may promote cancer transformation. These two features explain how the onset of chromosomal instability enabled by *TP53* inactivation leads to a rapid accumulation of genome diversity and complexity in precancer BE cells. One consequence of such dynamic evolution is the emergence of cells within a single BE field with shared *TP53* mutations but distinct oncogenic amplifications that independently lead to the development of an invasive phenotype (**Figure 7**). Polyclonal expansion of these lesions may eventually create a single large tumor mass with oncogenic driver heterogeneity as revealed by multi-regional or paired primary-metastatic analyses of advanced EAC(*55*). Therefore, the appearance of copy-number heterogeneity of potentially targetable oncogenic drivers in EAC may reflect the intrinsic evolutionary dynamics of

chromosomal instability throughout precancer evolution(*56, 57*) rather than being specific to advanced EACs. This model may be also applicable to other cancers with early *TP53* inactivation and high SCNA burdens, including serous ovarian cancers(*17*), basal breast cancers (*58*), uterine serous endometrial cancers(*59*), pancreatic cancers(*25*), and colitis-associated colorectal cancers(*13*).

Our model of branching evolution in genomically unstable BE cells leading to distinct oncogenic amplifications contrasts with the lack of heterogeneity of gain-of-function mutations in oncogenes(*60*). We think this difference reflects the different spontaneous mutation rates of copy-number alterations and point mutations. Although the *in vivo* mutation rates cannot be measured directly, the frequency of spontaneous gene amplification in cultured cells is known to be quite high in genomically unstable cancer cells (10^{-2}-10^{-4})(*61, 62*), but very low in normal immortalized cells(*63*). By contrast, the rate of spontaneous single-nucleotide substitutions (10^{-8}-10^{-9}/generation)(*64*) may be insufficient to produce heterogeneous gain-of-function mutations in oncogenes leading to parallel evolution and transformation of subclones. The disparity between copy-number and point-mutation heterogeneity is also reflected in the observation of almost completely diploid genomes in ultra-hypermutated cancers(*65*). Interestingly, a recent study revealed that point mutations can propagate asymmetrically into daughter cells due to unrepaired damaged nucleotides generated by exogenous mutagens(*66*). Asymmetric propagation of point mutations is conceptually similar to complementary copy-number changes generated by reciprocal distribution of broken chromosomes(*20, 21*), both giving rise to progeny cells with combinatorial genetic variation that more efficiently sample the genotype space. We expect this model to be a general paradigm of unstable somatic genome evolution that promotes tumor progression and gives rise to tumor heterogeneity.

Data Accession: All raw sequencing data will be uploaded to dbGap.

References:

1. K. Cleal, D. M. Baird, Catastrophic Endgames: Emerging Mechanisms of Telomere-Driven Genomic Instability. *Trends Genet* **36**, 347-359 (2020).
2. K. A. Knouse, T. Davoli, S. J. Elledge, A. Amon, Aneuploidy in Cancer: Seq-ing Answers to Old Questions. *Annual Review of Cancer Biology* **1**, 335-354 (2017).
3. M. S. Levine, A. J. Holland, The impact of mitotic errors on cell proliferation and tumorigenesis. *Genes Dev* **32**, 620-638 (2018).
4. Y. Li *et al.*, Patterns of somatic structural variation in human cancer genomes. *Nature* **578**, 112-121 (2020).
5. T. B. K. Watkins *et al.*, Pervasive chromosomal instability and karyotype order in tumour evolution. *Nature* **587**, 126-132 (2020).
6. A. J. Holland, D. W. Cleveland, Boveri revisited: chromosomal instability, aneuploidy and tumorigenesis. *Nature reviews. Molecular cell biology* **10**, 478-487 (2009).
7. N. J. Ganem, D. Pellman, Linking abnormal mitosis to the acquisition of DNA damage. *The Journal of cell biology* **199**, 871-881 (2012).
8. C. Z. Zhang, D. Pellman, From Mutational Mechanisms in Single Cells to Mutational Patterns in Cancer Genomes. *Cold Spring Harb Symp Quant Biol* **80**, 117-137 (2015).
9. H. Tanaka, T. Watanabe, Mechanisms Underlying Recurrent Genomic Amplification in Human Cancers. *Trends Cancer* **6**, 462-477 (2020).
10. I. Martincorena *et al.*, Somatic mutant clones colonize the human esophagus with age. *Science* **362**, 911-917 (2018).
11. A. Yokoyama *et al.*, Age-related remodelling of oesophageal epithelia by mutated cancer drivers. *Nature* **565**, 312-317 (2019).
12. A. H. Shain *et al.*, The Genetic Evolution of Melanoma from Precursor Lesions. *N Engl J Med* **373**, 1926-1936 (2015).
13. A. M. Baker *et al.*, Evolutionary history of human colitis-associated colorectal cancer. *Gut* **68**, 985-995 (2019).
14. J. M. J. Weaver *et al.*, Ordering of mutations in preinvasive disease stages of esophageal carcinogenesis. *Nat Genet* **46**, 837-843 (2014).
15. C. R. Choi, I. A. Bakir, A. L. Hart, T. A. Graham, Clonal evolution of colorectal cancer in IBD. *Nat Rev Gastroenterol Hepatol* **14**, 218-229 (2017).
16. M. D. Stachler *et al.*, Detection of Mutations in Barrett's Esophagus Before Progression to High-Grade Dysplasia or Adenocarcinoma. *Gastroenterology* **155**, 156-167 (2018).
17. A. M. Patch *et al.*, Whole-genome characterization of chemoresistant ovarian cancer. *Nature* **521**, 489-494 (2015).
18. N. Waddell *et al.*, Whole genomes redefine the mutational landscape of pancreatic cancer. *Nature* **518**, 495-501 (2015).
19. T. Helleday, S. Eshtad, S. Nik-Zainal, Mechanisms underlying mutational signatures in human cancers. *Nature reviews. Genetics* **15**, 585-598 (2014).
20. N. T. Umbreit *et al.*, Mechanisms generating cancer genome complexity from a single cell division error. *Science* **368**, (2020).
21. C. Z. Zhang *et al.*, Chromothripsis from DNA damage in micronuclei. *Nature* **522**, 179-184 (2015).
22. R. Gao *et al.*, Punctuated copy number evolution and clonal stasis in triple-negative breast cancer. *Nat Genet* **48**, 1119-1130 (2016).
23. N. Navin *et al.*, Tumour evolution inferred by single-cell sequencing. *Nature* **472**, 90-94 (2011).
24. S. Nik-Zainal *et al.*, The life history of 21 breast cancers. *Cell* **149**, 994-1007 (2012).
25. F. Notta *et al.*, A renewed model of pancreatic cancer evolution based on genomic rearrangement patterns. *Nature* **538**, 378-382 (2016).
26. Y. Peters *et al.*, Barrett oesophagus. *Nat Rev Dis Primers* **5**, 35 (2019).

27. B. J. Reid, X. Li, P. C. Galipeau, T. L. Vaughan, Barrett's oesophagus and oesophageal adenocarcinoma: time for a new synthesis. *Nat Rev Cancer* **10**, 87-101 (2010).
28. G. Contino, T. L. Vaughan, D. Whiteman, R. C. Fitzgerald, The Evolving Genomic Landscape of Barrett's Esophagus and Esophageal Adenocarcinoma. *Gastroenterology* **153**, 657-673 e651 (2017).
29. V. Giroux, A. K. Rustgi, Metaplasia: tissue injury adaptation and a precursor to the dysplasia-cancer sequence. *Nat Rev Cancer* **17**, 594-604 (2017).
30. P. R. Loh *et al.*, Reference-based phasing using the Haplotype Reference Consortium panel. *Nat Genet* **48**, 1443-1448 (2016).
31. N. Cancer Genome Atlas Research *et al.*, Integrated genomic characterization of oesophageal carcinoma. *Nature* **541**, 169-175 (2017).
32. A. M. Dulak *et al.*, Exome and whole-genome sequencing of esophageal adenocarcinoma identifies recurrent driver events and mutational complexity. *Nat Genet* **45**, 478-486 (2013).
33. C. S. Ross-Innes *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).
34. M. D. Stachler *et al.*, Paired exome analysis of Barrett's esophagus and adenocarcinoma. *Nat Genet* **47**, 1047-1055 (2015).
35. E. R. Kastenhuber, S. W. Lowe, Putting p53 in Context. *Cell* **170**, 1062-1078 (2017).
36. C. M. Bielski *et al.*, Genome doubling shapes the evolution and prognosis of advanced cancers. *Nat Genet* **50**, 1189-1195 (2018).
37. T. I. Zack *et al.*, Pan-cancer patterns of somatic copy number alteration. *Nat Genet* **45**, 1134-1140 (2013).
38. T. Davoli, T. de Lange, Telomere-driven tetraploidization occurs in human cells undergoing crisis and promotes transformation of mouse cells. *Cancer Cell* **21**, 765-776 (2012).
39. T. Fujiwara *et al.*, Cytokinesis failure generating tetraploids promotes tumorigenesis in p53-null cells. *Nature* **437**, 1043-1047 (2005).
40. N. J. Ganem, S. A. Godinho, D. Pellman, A mechanism linking extra centrosomes to chromosomal instability. *Nature* **460**, 278-282 (2009).
41. F. Notta, S. A. Hahn, F. X. Real, A genetic roadmap of pancreatic cancer: still evolving. *Gut* **66**, 2170-2178 (2017).
42. J. Maciejowski, Y. Li, N. Bosco, P. J. Campbell, T. de Lange, Chromothripsis and Kataegis Induced by Telomere Crisis. *Cell* **163**, 1641-1654 (2015).
43. K. Nones *et al.*, Genomic catastrophes frequently arise in esophageal adenocarcinoma and drive tumorigenesis. *Nat Commun* **5**, 5224 (2014).
44. D. W. Garsed *et al.*, The architecture and evolution of cancer neochromosomes. *Cancer Cell* **26**, 653-667 (2014).
45. Y. Li *et al.*, Constitutional and somatic rearrangement of chromosome 21 in acute lymphoblastic leukaemia. *Nature* **508**, 98-102 (2014).
46. K. Cleal, R. E. Jones, J. W. Grimstead, E. A. Hendrickson, D. M. Baird, Chromothripsis during telomere crisis is independent of NHEJ, and consistent with a replicative origin. *Genome Res* **29**, 737-749 (2019).
47. O. Shoshani *et al.*, Chromothripsis drives the evolution of gene amplification in cancer. *Nature*, (2020).
48. P. Ly *et al.*, Chromosome segregation errors generate a diverse spectrum of simple and complex genomic rearrangements. *Nat Genet* **51**, 705-715 (2019).
49. J. K. Cowell, Double minutes and homogeneously staining regions: gene amplification in mammalian cells. *Annual review of genetics* **16**, 21-59 (1982).
50. T. De Lange, Telomere-related genome instability in cancer. *Cold Spring Harb Symp Quant Biol* **70**, 197-204 (2005).

51. S. E. Artandi *et al.*, Telomere dysfunction promotes non-reciprocal translocations and epithelial cancers in mice. *Nature* **406**, 641-645 (2000).
52. X. Li *et al.*, Temporal and spatial evolution of somatic chromosomal alterations: a case-cohort study of Barrett's esophagus. *Cancer Prev Res (Phila)* **7**, 114-127 (2014).
53. F. Newell *et al.*, Complex structural rearrangements are present in high-grade dysplastic Barrett's oesophagus samples. *BMC Med Genomics* **12**, 31 (2019).
54. Y. Uetake, G. Sluder, Activation of the apoptotic pathway during prolonged prometaphase blocks daughter cell proliferation. *Mol Biol Cell* **29**, 2632-2643 (2018).
55. E. Pectasides *et al.*, Genomic Heterogeneity as a Barrier to Precision Medicine in Gastroesophageal Adenocarcinoma. *Cancer Discov* **8**, 37-48 (2018).
56. C. C. Maley *et al.*, Genetic clonal diversity predicts progression to esophageal adenocarcinoma. *Nat Genet* **38**, 468-473 (2006).
57. P. Martinez *et al.*, Dynamic clonal equilibrium and predetermined cancer risk in Barrett's oesophagus. *Nat Commun* **7**, 12158 (2016).
58. N. Cancer Genome Atlas, Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61-70 (2012).
59. N. Cancer Genome Atlas Research *et al.*, Integrated genomic characterization of endometrial carcinoma. *Nature* **497**, 67-73 (2013).
60. J. G. Reiter *et al.*, Minimal functional driver gene heterogeneity among untreated metastases. *Science* **361**, 1033-1037 (2018).
61. R. N. Johnston, S. M. Beverley, R. T. Schimke, Rapid spontaneous dihydrofolate reductase gene amplification shown by fluorescence-activated cell sorting. *Proceedings of the National Academy of Sciences of the United States of America* **80**, 3711-3715 (1983).
62. T. D. Tlsty, B. H. Margolin, K. Lum, Differences in the rates of gene amplification in nontumorigenic and tumorigenic cell lines as measured by Luria-Delbruck fluctuation analysis. *Proceedings of the National Academy of Sciences of the United States of America* **86**, 9441-9445 (1989).
63. T. D. Tlsty, Normal diploid human and rodent cells lack a detectable frequency of gene amplification. *Proceedings of the National Academy of Sciences of the United States of America* **87**, 3132-3136 (1990).
64. I. Martincorena, P. J. Campbell, Somatic mutation in cancer and normal cells. *Science* **349**, 1483-1489 (2015).
65. A. Shlien *et al.*, Combined hereditary and somatic mutations of replication error repair genes result in rapid onset of ultra-hypermutated cancers. *Nat Genet* **47**, 257-262 (2015).
66. S. J. Aitken *et al.*, Pervasive lesion segregation shapes cancer genome evolution. *Nature* **583**, 265-270 (2020).

Acknowledgements: We would like to thank all of the patients who were willing to contribute samples to the manuscript, the DFCI Center for cancer genomics and BROAD Institute for help with library preparation and sequencing, and Dr. David Pellman for his insightful review of the manuscript.
Funding: Doris Duke Charitable Foundation (MDS), National Institutes of Health (AJB:U54CA163060; MDS:K08DK109209)
Author contributions: Data analysis: CB, MDS, AJB, CZ; Computational analysis: RT, GB, CB, CZ, CS, GG; Patient selection and clinical data collection: KW, JMD, KKM, MDS, JK; Pre-sequencing ‘wet lab’: MDS, LS; FISH: HB, MW, YI; Histologic review: AA, RO, MDS; Conceived of study: MDS, AJB, CZ; Manuscript preparation: MDS, CB, AJB, CZ; Manuscript review: MDS, CB, KKW, YI, AJB, CZ
Competing interests: A.J.B. has received research funding from Bayer, Merck and Novartis, is a consultant to Earli, and HelixNano and a co-founder of Signet Therapeutics. C.-Z. Zhang co-founded and serves as a scientific advisor to Pillar Biosciences.
Data Accession: All raw sequencing data will be uploaded to dbGap.

Figure and Table Legends

Figure 1: Overview of experimental design and analytical workflow. From more than 500 endoscopic mucosal resections, we have identified 15 patients present with early invasive esophageal cancers. Through laser capture microdissection, 75 samples of early EAC and precancer lesions, including non-dysplastic Barrett's Esophagus (NDBE), low-grade dysplasia (LGD) and high-grade dysplasia (HGD), are collected and subjected to whole-genome sequencing. We first determine the phylogeny of samples from each patient based on genomic alterations shared by two or more samples. Based on the phylogeny, we then infer the timing and evolution of copy-number alterations, including branching evolution of a single ancestral chromosome generating distinct copy-number changes in related BE/EAC genomes.

Figure 2: Phylogenetic trees of EAC and precursor non-dysplastic and dysplastic lesions determined by haplotype-specific copy-number alterations. Fifteen patients 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 approximately reflects the number of altered chromosomes. Annotated alterations include: (1) recurrent alterations or those involving known EAC drivers; (2) focally amplified regions or oncogenes (magenta); (3) chromosomes or chromosome arms with divergent copy-number alterations in more than one progeny clones. The colors of annotated chromosomes reflect the complexity of copy-number alterations: simple deletion/duplication, uniparental disomy, arm-level copy-number changes (blue), large segmental (terminal or internal) copy-number changes or their combinations (orange), complex copy-number alterations (red), focal amplifications (magenta). For a detailed list of alterations along each phylogenetic branch, see **Table S2** and **Figure S7**. See **Figure S8** for examples of different types of copy-number alterations.

Figure 3: Phylogeny of diploid and aneuploid BE cells from a high-grade dysplastic Barrett's Esophagus revealed by single-cell sequencing. **A**. Phylogenetic tree of 68 single BE cells determined by haplotype-specific copy number. Each open circle represents a single cell. Large filled circles represent subclones of cells with the same copy-number profile. Small filled circles represent inferred intermediate states (gray for pre-WGD, black for post WGD). Aneuploid cells are separated into two branches each inferred to have undergone an independent whole-genome duplication (WGD) event. **B-H**. Examples of copy-number alterations before (B-E) and after (F-H) *TP53* inactivation. Gray and black dots represent haplotype-specific DNA copy number of parental chromosomes. **B**. Ancestral 3p uniparental disomy (UPD) shared by all but four

cells. **C**. Sporadic 3p terminal gain after 3p UPD. **D**. Two segmental copy number alterations on Chr.1 (large paracentric deletion on 1p and UPD at the q-terminus) shared by a subclone. **E**. Variable boundaries of 9p UPD in a subclone. **F**. Chromothripsis of the Y chromosome in cell F12. **G**. Chromothripsis of Chr.22 that is shared by cell F2, C5 and F7. **H**. Focal amplification spanning the *ERBB2* gene on Chr.17 in cell C5 and F7 (red circles) that displays the signature copy-number pattern of breakage-fusion-bridge cycles. For a detailed list of alterations in each cell, see **Table S3**.

Figure 4: Landscape of somatic copy-number alterations (SCNA) in BE and EAC lesions. **A**. Mean SCNA burden in samples grouped by disease stage (left), *TP53* mutation status (middle), and timing relative to whole-genome duplication (right). The SCNA burden is measured by the total number of altered autosomes (maximum 44) and subdivided into local deletions or duplications (gray), uniparental disomies (light gray), arm-level SCNAs (dark gray), and complex segmental SCNAs (black). In the middle panel, the "intact" *TP53* group ("*TP53*") only includes NDBE/LGD samples without detectable *TP53* alterations, but not HGD/EAC samples. See **Figure S4** for the SCNA burden in each sample. **B**. SCNA burden along ancestral (having more than one progeny clone) and private (only one progeny clone) evolutionary branches. For each branch, we also show the *TP53* mutation status, the highest grade of its progeny lesions, and its relationship to WGD events.

Figure 5: Patterns of segmental copy-number changes observed in BE/EAC genomes that display signatures of dicentric chromosome bridge resolution. **A-C.** (*Left*) Different types of dicentric chromosome breakage and the copy-number outcomes: (A) terminal; (B) paracentric; or (C) pericentric segmental copy number changes. The open and filled chromatids and may be sister chromatids or different chromosomes. Both A and B were demonstrated *in vitro* in Umbreit et al. (2020). The model that pericentric copy-number changes may arise from breakage of dicentric ring chromosomes (C) or multi-centric chromosomes (not shown) has not been demonstrated *in vitro* but is plausible as telomere crisis *in vivo* may lead to multiple critically shortened telomeres. Examples of ring chromosomes (Umbreit et al., 2020) or rearrangements involving multiple chromosomes (Maciejowski et al., 2015) were also observed in the progeny populations of cells that underwent telomere crisis or bridge induction. (*Right*) Examples from BE/EAC genomes. The copy-number plots (25kb bins) show the DNA copy number of the altered chromosome but not the intact homolog. Examples of gain and loss in each group are unrelated. **D-H**. Examples of SCNAs in BE/EAC genomes that can be explained by consecutive or multiple BFB cycles. Arabic numbers represent different evolutionary outcomes

of BFB cycles labelled in **Figure S5C**. **D**. Terminal loss -> terminal gain. **E**. Paracentric loss -> two duplications near the centromeric break end. **F**. BFB amplification. **G**. Paracentric loss followed by copy-number gain at the centromeric break end (shared) and copy number gain near the telomeric break end in the EAC sample. The variable copy number of the black homolog contrasts with the constant copy-number of the intact gray homolog, indicating genetic heterogeneity due to ongoing breakage-fusion-bridge cycles. **H**. Reciprocal copy-number changes in the HGD1 and IMEAC1 lesions at a single breakpoint reflect the outcome of breakage of a single dicentric chromosomes. The focally amplified region on the telomeric end in IMEAC1 is also consistent with BFB amplifications.

Figure 6: Examples of complex copy-number patterns in BE/EAC genomes reflecting different evolution sequences of chromothripsis and BFB cycles. Arabic numbers represent different evolutionary outcomes of BFB cycles labelled in **Figure S5C**. **A**. Standalone arm-level chromothripsis followed by arm-level UPD. **B**. A schematic model of regional chromothripsis resulting from dicentric chromosome breakage. **C**. A schematic model of arm-level chromothripsis resulting from fragmentation of a dicentric chromosome in a micronucleus. **D**. Regional chromothripsis occurring downstream of an ancestral terminal loss inferred by the timing of copy-number changes. **E**. Chromothripsis and BFB amplification occurring downstream of an ancestral paracentric loss inferred by multi-sample analysis. **F**. Chromothripsis of a dicentric chromosome generated by inter-chromosomal rearrangement leading to three-state copy-number oscillation in two chromosome arms. **G**. Arm-level chromothripsis in EAC1 inferred to be a downstream event of a chromosome-type BFB that also generated a paracentric deletion in the IMEAC genome. The common origin of the chromothripsis event and the paracentric loss is established by a shared copy-number breakpoint (dotted line) with complementary copy-number changes in each genome. **H**. Examples of chromothripsis followed by BFB amplification. The copy-number breakpoint (1p, dotted line) and terminal loss (16p, dashed lines) shared between the EAC and IMEAC genomes suggests that the divergent copy-number patterns in these genomes originate from a single unstable (dicentric) chromosome.

Figure 7: Cartoon showing concurrent evolution and clonal expansion of BE cells driven by chromosomal instability leading to independent cancers in a single BE field.

Figure S1: Example images of Barrett's esophagus without dysplasia (NDBE), low-grade dysplasia (LGD), high-grade dysplasia (HGD), intramucosal esophageal adenocarcinoma (IMEAC), and esophageal adenocarcinoma (EAC).

Figure S2: Haplotype inference and haplotype-specific copy-number analysis. **A**. Overall workflow of haplotype-specific copy-number analysis in a single sample. 1. Normalized sequence coverage is calculated from raw read depth coverage after GC bias normalization. 2. Local allelic fraction (black and grey dots) is calculated as the average allelic fraction of phased genotypes at SNP sites in 25kb bins. 3. Local allelic DNA copy-number is calculated from unphased sequence coverage and phased allelic fraction. 4. Copy-number changepoints are phased based on local allelic copy number (alleleA and alleleB CN). The requirement of an intact homolog with constant copy number at each copy-number changepoint implies that only vertical or horizontal (double arrows), but not off-diagonal transitions in copy-number states are allowed. (Dashed box) When the average allelic fraction is calculated from unphased genotypes, there is no separation of discrete copy-number states. **B**. Phasing of copy-number changepoints. (*Left*) Allowed copy-number transitions: At each copy-number changepoint there should be one intact homolog with constant copy number; the only exception is uniparental disomy (UPD) resulting in loss of one homolog and gain of the other homolog. (*Right*) Forbidden copy-number transitions: Incorrect local haplotype inference can lead to copy-number changes of both homologs. The constraint of having one constant homolog (except UPD events) is used to correct long-range switching errors in statistical haplotype inference. **C**. Joint haplotype inference from related samples. Allelic imbalance in one sample is used to derive the complete chromosomal haplotype that is then used to calculate haplotype-specific copy number in other samples with incomplete allelic depth separation.

Figure S3: Average burden of single-nucleotide variants and structural variants in BE samples grouped by histopathological grading.

Figure S4: Landscape of somatic copy-number alterations (SCNA) in BE and EAC lesions. **A**. Total SCNA burden in each sample broken down by SCNA type. Local deletion/duplication, UPD, and arm-level SCNAs are colored the same as in Figure 4. Complex segmental copy-number changes are further subdivided to (1) terminal, paracentric, and pericentric gain/loss (orange); (2) complex alterations (red); and (3) focal amplifications (purple). Samples are grouped based on pathological grading (non-dysplastic BE, low-grade dysplasia, high-grade dysplasia, and carcinoma). The mutation status of *TP53* is shown below each sample: black circles for bi-allelic inactivation, open circles for no identifiable alterations or mono-allelic inactivation. **B**. Mean SCNA burden for samples in each pathological group. **C**. Mean SCNA burden in samples with and without bi-allelic *TP53* inactivation. The SCNA burdens in five HGD/EAC samples (four from Patient 9, one from Patient 1) without evidence of bi-allelic *TP53* loss are

shown separately from NDBE/LGD samples without *TP53* inactivation. **D**. Mean SCNA burden in samples without whole-genome duplication (WGD), or inferred to have occurred prior to or after WGD. **E**. Mutation burden along ancestral (having more than one progeny clone) and private (only one progeny clone) evolutionary branches of SCNAs (top) and SNVs (bottom). The SNV burden is measured by the number of 100 kb bins with at least one mutation. As the median SNV frequency in EAC is less than 10 per Mb, the average number of mutations in 100 kb bins is less than 1. The number of 100kb bins with shared mutations is a better measure of gradual mutagenesis as it avoids overcounting of mutations due to local hypermutation such as kataegis. SNV burden on private branches is not analyzed due to the high rate of false positive mutation detection from a single sample.

Figure S5: Copy-number outcomes of tetraploidization (**A**), single BFB cycle of a dicentric chromosome (**B**), multi-generational evolution (including BFB cycles and chromothripsis) of a dicentric chromosome (**C**), and micronucleation (**D**). **A**. Cytokinesis failure leads to a binucleated cell that can give rise to tetraploid cells (4N) after bipolar mitosis or aneuploid cells (less than 4N) after multi-polar mitosis. **B**. Resolution of dicentric bridge chromosomes results in reciprocal gain and loss of a telomere-bound segment (chromatid-type fusion) or a large internal segment (chromosome-type fusion, **Figure 5B**), with occasional DNA fragmentation near the breakage site leading to local chromothripsis (**Figure 6B**). **C**. Multi-generational evolution of a broken chromosome through concurrent breakage-fusion-bridge (BFB) cycles and chromothripsis. (*Top*) Copy-number outcomes of breakage-fusion-bridge cycles without chromosome fragmentation. (*Bottom*) Copy-number outcomes of breakage-fusion-bridge cycles after ancestral chromothripsis. The unstable chromosome generated by BFB cycles can become mitotically stable after acquiring a telomere through *de novo* telomere addition, translocation to a telomeric segment, or chromothripsis of a pair of sister chromatids. The broken chromosome can also give rise to an unstable ring chromosome (after loss of both telomeres) or acentric extrachromosomal circles (after centromeric loss), both of which can generate high-level gene amplification. **D**. Chromosome fragmentation in micronuclei results in chromothripsis with oscillating DNA copy number in both daughter cells.

Figure S6: Examples of rearrangements joining breakpoints of terminal, paracentric, or pericentric copy-number alterations. **A**. (*left*) Two breakpoints related to the paracentric loss on 5p respectively join breakpoints related to the pericentric retention on 4q and the terminal gain of 15q; (*right*) the breakpoint of the 10q-terminal gain joins the telomeric breakpoint of the paracentric loss on 12q. **B**. The breakpoint of the 16p-terminal gain joins a breakpoint related to the paracentric

loss in 1q. **C**. The breakpoint of the 11p-terminal gain joins a breakpoint in the 15q terminal region with complex gains.

Figure S7: Phylogenetic trees with branch labels and annotated oncogene amplifications. Somatic alterations on each branch are listed in **Table S2**.

Figure S8: Classification of somatic copy-number changes based on the mutational mechanisms. I. Local sequence deletion/duplication and uniparental disomy do not generate unstable (acentric/dicentric) chromosomes. These events make up the majority of alterations seen in cells with intact p53. II-VI. Copy-number alterations resulting from chromosomal instability undergo clonal expansion only after p53 loss. II. Arm-level copy-number changes are generated by chromosome mis-segregation or abnormal mitosis, including cytokinesis failure and multi-polar cell division. III. Terminal, paracentric, and pericentric segmental copy-number alterations can result from dicentric chromosome breakage and undergo multi-generational evolution, generating copy-number gains or focal amplifications. IV. Dicentric chromosome breakage or DNA damage in micronuclei can lead to chromosome fragmentation and create two-state (single-chromatid fragmentation) or three-state (sister-chromatid fragmentation) oscillating copy-number pattern. V. Chromothripsis can also lead to telomere loss and create unstable dicentric chromosomes, generating complex copy-number gains or focal amplifications through breakage-fusion-bridge cycles. VI. Focally amplified sequences can be either intra-chromosomal (e.g., generated by intra-chromosomal BFBs) or extra-chromosomal (e.g., acentric fragments from chromothripsis).

Table S1: Summary of sequencing coverage and metrics of each sample.

Table S2: List of somatic copy-number alterations by evolutionary branches (Tab 1) and individual sample (Tab 2)

Table S3: List of large segmental or arm-level SCNAs detected in single cells from a high-grade dysplastic Barrett's Esophagus (related to **Figure 3**).

Table S4: List of local SCNAs (<1Mb) (Tab 1) and uniparental disomy (Tab 2) alterations detected in bulk BE and EAC samples.

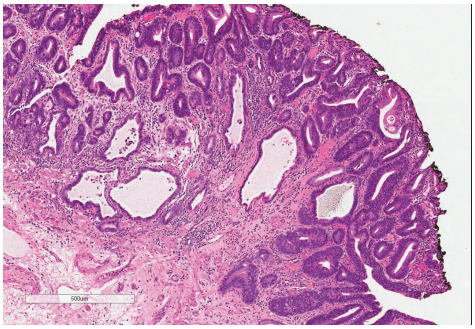
Table S5: List of segmental SCNAs that can be explained by single (Tab 1) or multi-generational (Tab 2,3) BFB cycles

Table S6: List of chromothripsis (Tab 1) detected in bulk BE/EAC samples and subdivided into

standalone chromothripsis (Tab 2,3), chromothripsis arising secondary to large segmental SCNAs (Tab 4), and chromothripsis followed by sequence duplication or amplification (Tab 5).

Table S7: List of focal amplifications (Tab 1) detected in bulk BE/EAC samples and selected validation by Fluorescence in-situ hybridization (Tab 2).

Archival Search in Endoscopic Mucosal Resections (EMR)

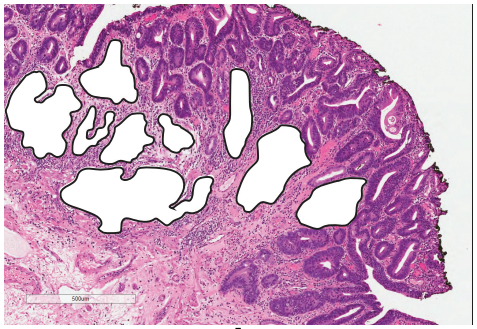


Histologic review

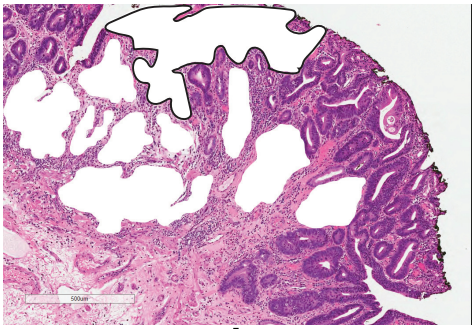
Identified 15 patients with
- T1 Esophageal AdenoCarcinoma (EAC) or IntraMucosal Esophageal AdenoCarcinoma (IMEAC)
- no prior therapy

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

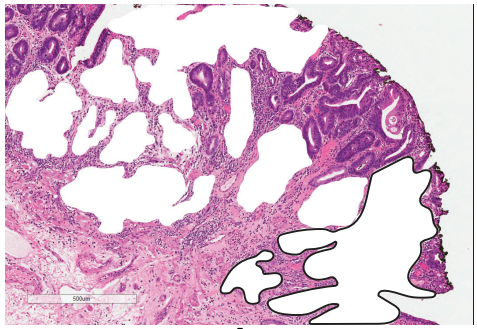
Cut IMEAC



Cut first HGD



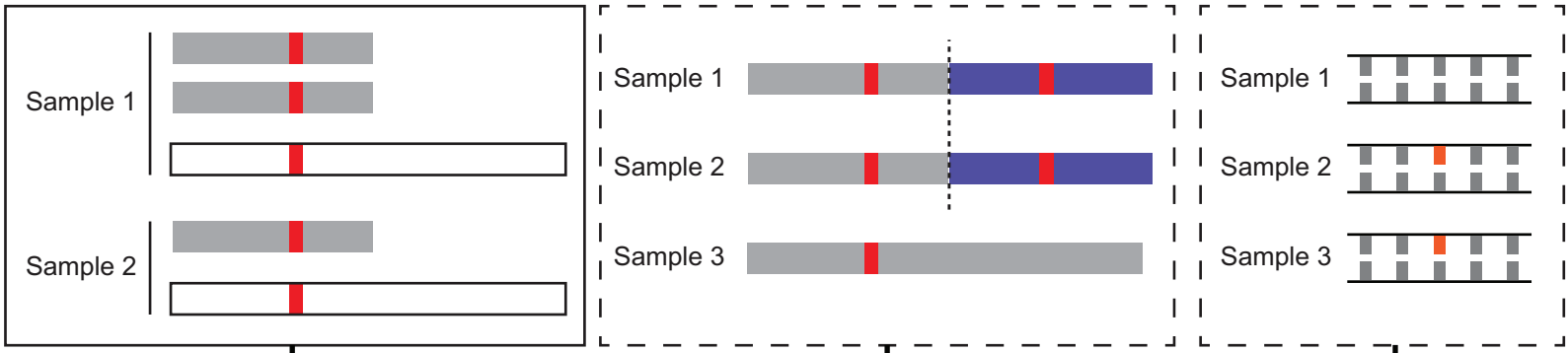
Cut second HGD



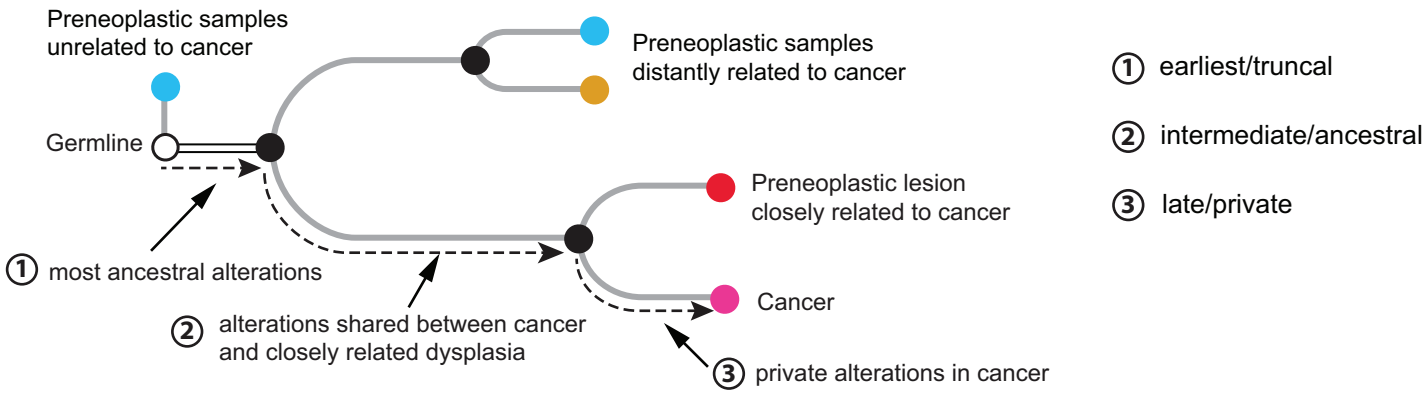
Whole-Genome Sequencing Analysis (75 samples from 15 patients)

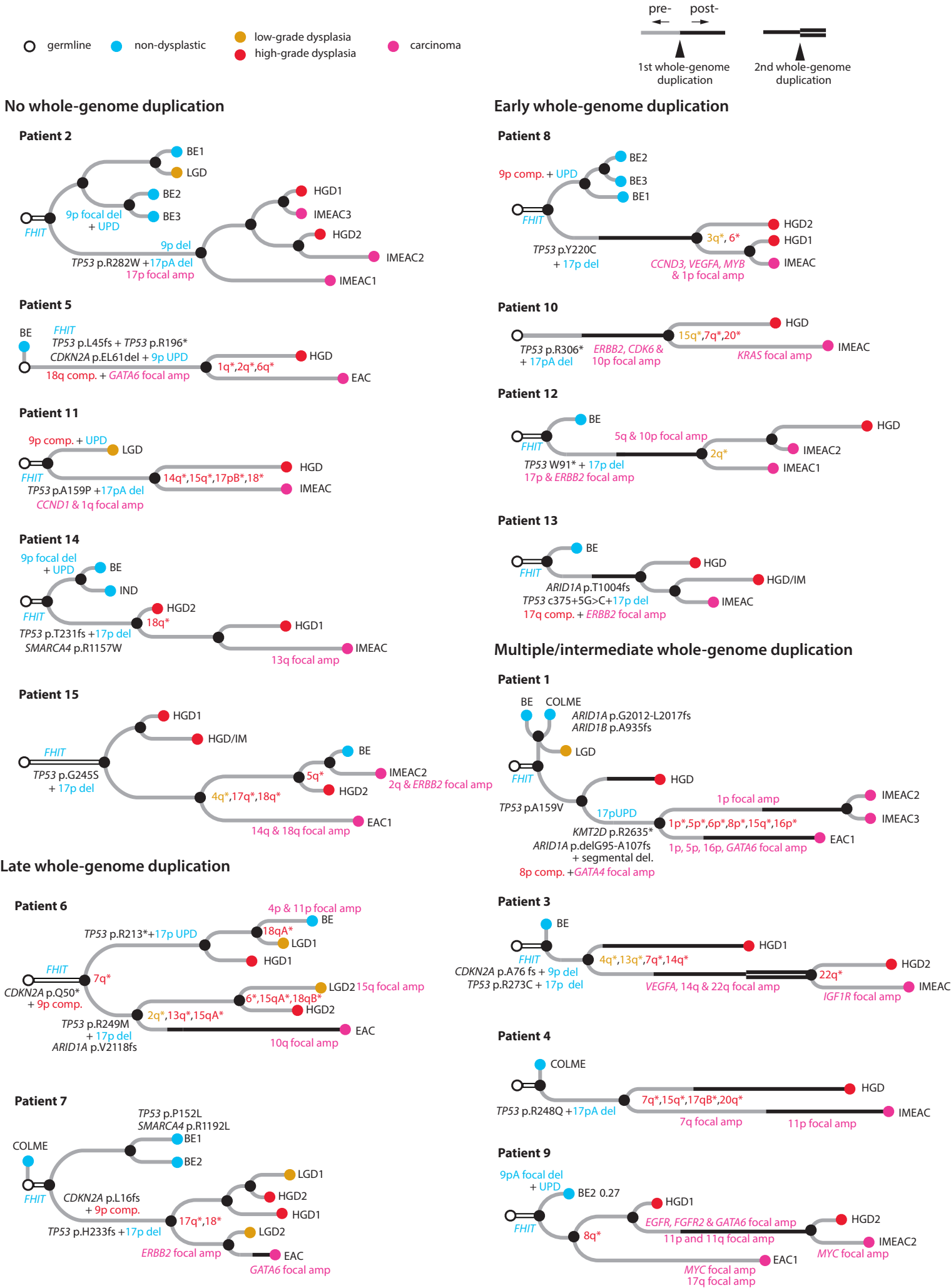
Sample-specific **copy-number analysis** after joint haplotype inference

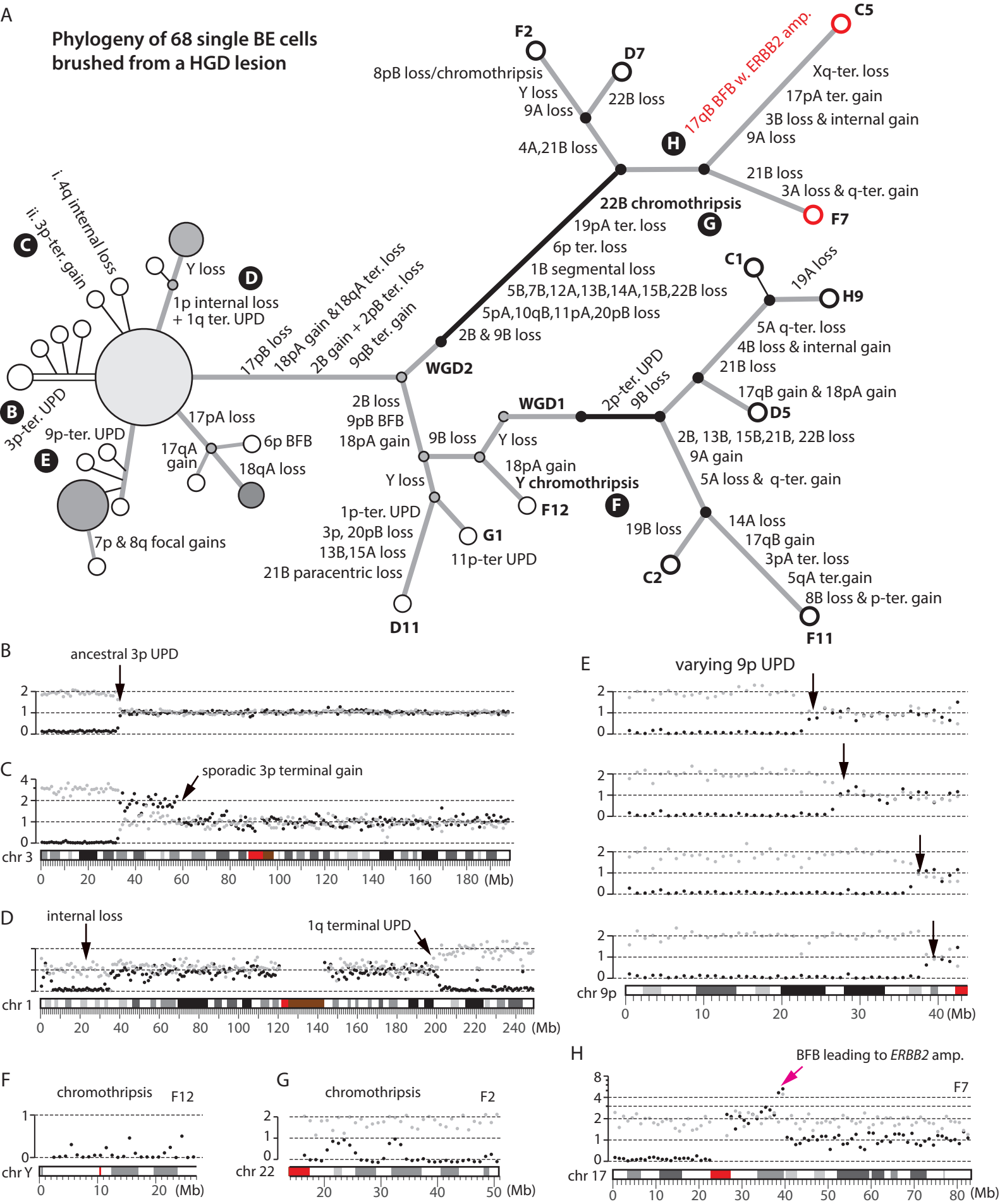
Joint detection of **structural rearrangements** and **local variants**

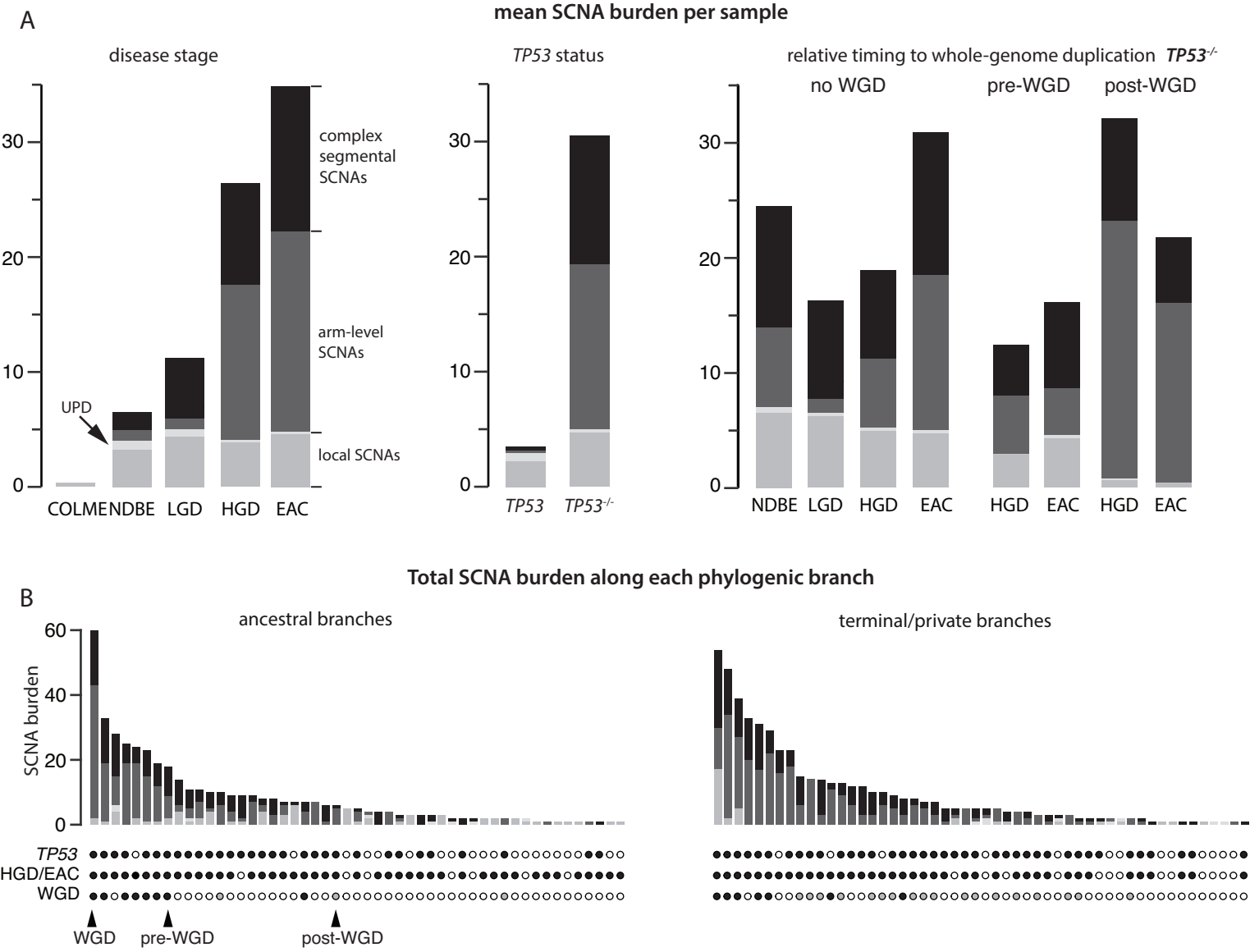


Phylogenetic Inference and Analysis of Somatic Copy-Number Evolution



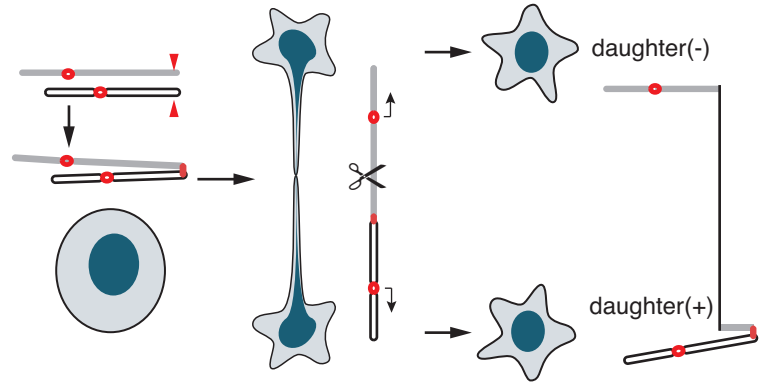

germline
non-dysplastic
low-grade dysplasia
high-grade dysplasia
carcinoma
pre-
post-
1st whole-genome duplication
2nd whole-genome duplication
No whole-genome duplication
Patient 2
BE1
LGD
BE2
BE3
9p focal del + UPD
FHIT
9p del
TP53 p.R282W +17pA del
17p focal amp
HGD1
IMEAC3
HGD2
IMEAC2
IMEAC1
Patient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


A
Phylogeny of 68 single BE cells brushed from a HGD lesion
F2
D7
8pB loss/chromothripsis
Y loss
9A loss
22B loss
4A,21B loss
H
17qB BFB w. ERBB2 amp.
C5
Xq-ter. loss
17pA ter. gain
3B loss & internal gain
9A loss
21B loss
3A loss & q-ter. gain
F7
22B chromothripsis
G
19pA ter. loss
6p ter. loss
1B segmental loss
5B,7B,12A,13B,14A,15B,22B loss
5pA,10qB,11pA,20pB loss
2B & 9B loss
WGD2
i. 4q internal loss
ii. 3p-ter. gain
C
Y loss
D
1p internal loss
+ 1q ter. UPD
17pB loss
18pA gain &18qA ter. loss
2B gain + 2pB ter. loss
9qB ter. gain
B
3p-ter. UPD
9p-ter. UPD
E
17pA loss
17qA gain
6p BFB
18qA loss
7p & 8q focal gains
2B loss
9pB BFB
18pA gain
9B loss
Y loss
1p-ter. UPD
3p, 20pB loss
13B,15A loss
21B paracentric loss
D11
G1
11p-ter UPD
WGD1
Y loss
18pA gain
Y chromothripsis
F
F12
2p-ter. UPD
9B loss
C1
19A loss
H9
5A q-ter. loss
4B loss & internal gain
21B loss
17qB gain & 18pA gain
D5
2B, 13B, 15B,21B, 22B loss
9A gain
5A loss & q-ter. gain
19B loss
C2
14A loss
17qB gain
3pA ter. loss
5qA ter.gain
8B loss & p-ter. gain
F11
B
ancestral 3p UPD
C
sporadic 3p terminal gain
chr 3
0 20 40 60 80 100 120 140 160 180 (Mb)
D
internal loss
1q terminal UPD
chr 1
0 20 40 60 80 100 120 140 160 180 200 220 240 (Mb)
E
varying 9p UPD
chr 9p
0 10 20 30 40 (Mb)
F
chromothripsis
F12
chr Y
0 10 20 (Mb)
G
chromothripsis
F2
chr 22
20 30 40 50 (Mb)
H
BFB leading to ERBB2 amp.
F7
chr 17
0 10 20 30 40 50 60 70 80 (Mb)

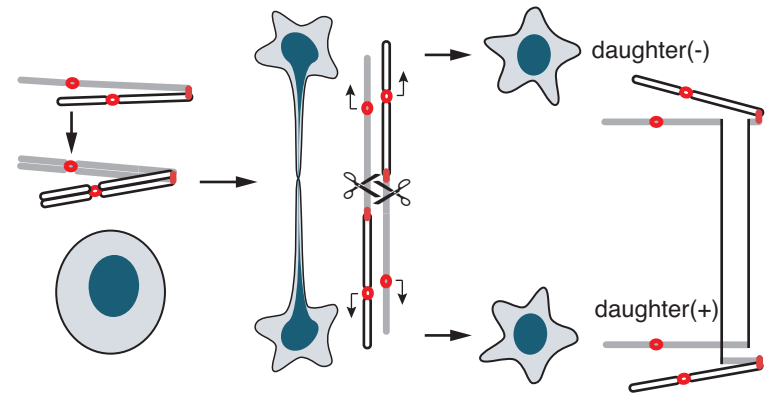

A
mean SCNA burden per sample
disease stage
TP53 status
relative timing to whole-genome duplication TP53-/-
no WGD
pre-WGD
post-WGD
complex segmental SCNAs
arm-level SCNAs
local SCNAs
UPD
0
10
20
30
COL
MEN
DBE
LGD
HGD
EAC
TP53
TP53-/-
NDBE
B
Total SCNA burden along each phylogenic branch
ancestral branches
terminal/private branches
SCNA burden
0
20
40
60
TP53
HGD/EAC
WGD
WGD
pre-WGD
post-WGD

Models of dicentric bridge resolution

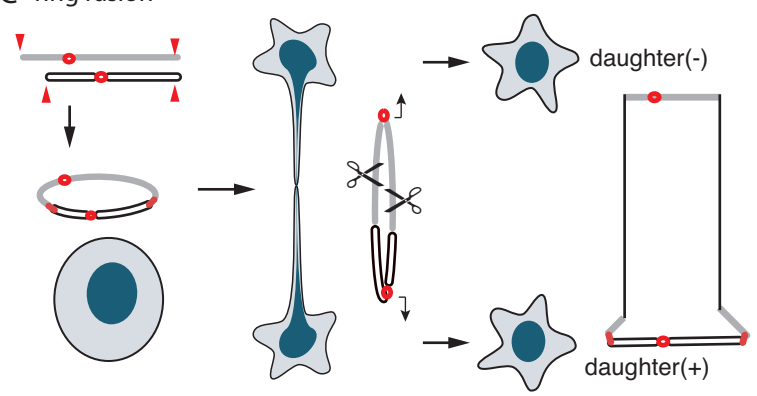
A chromatid-type fusion



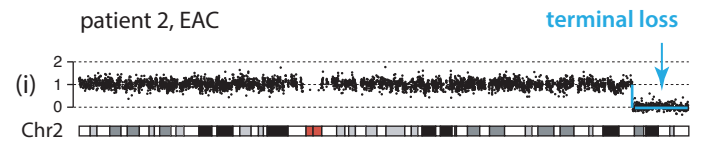
B chromosome-type fusion

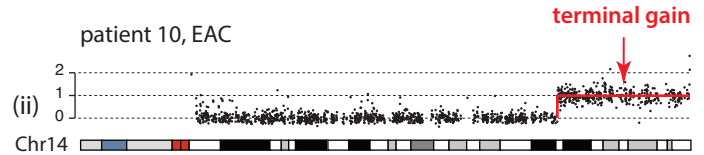


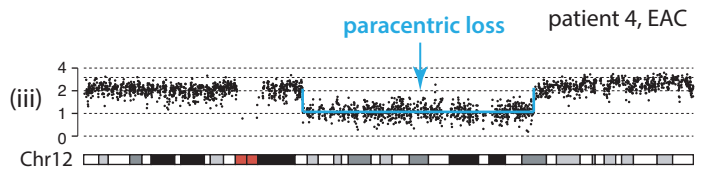
C ring fusion

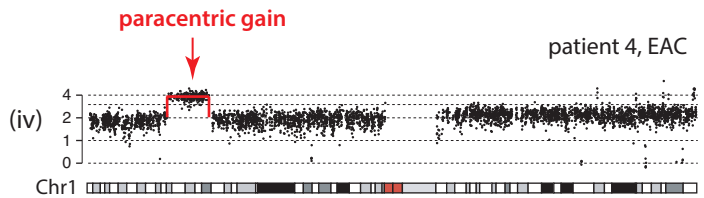


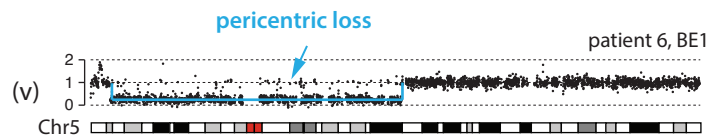
Observations in esophageal cancers

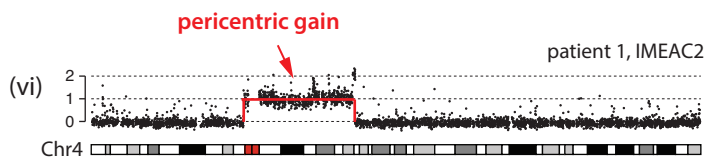




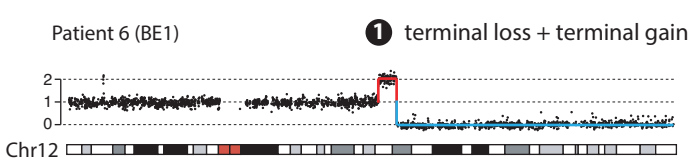




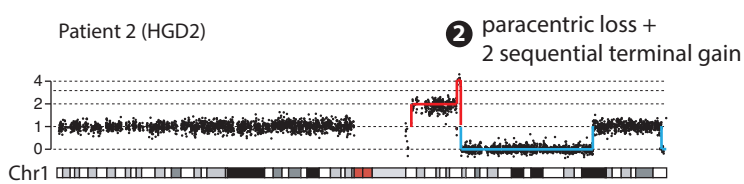




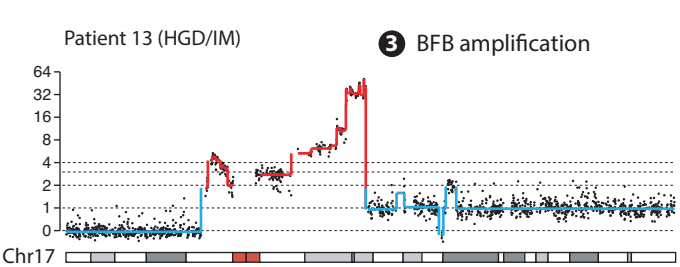
D



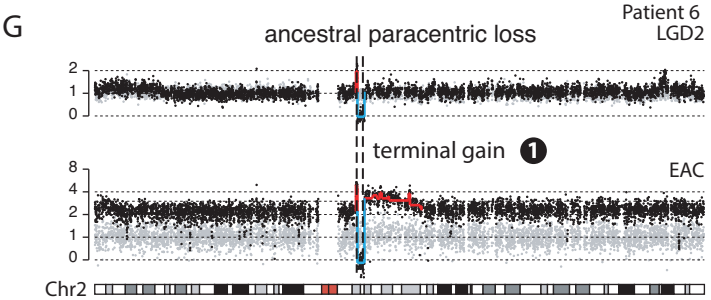
E



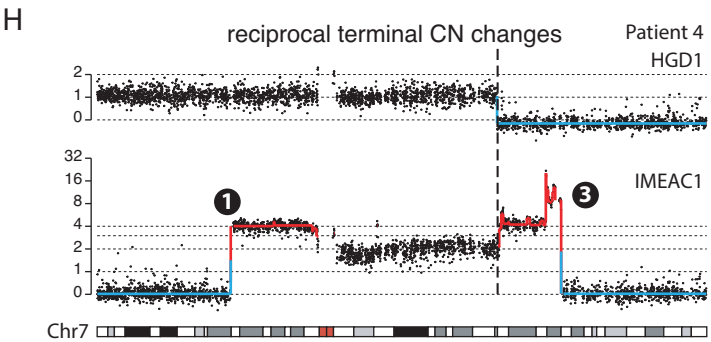
F

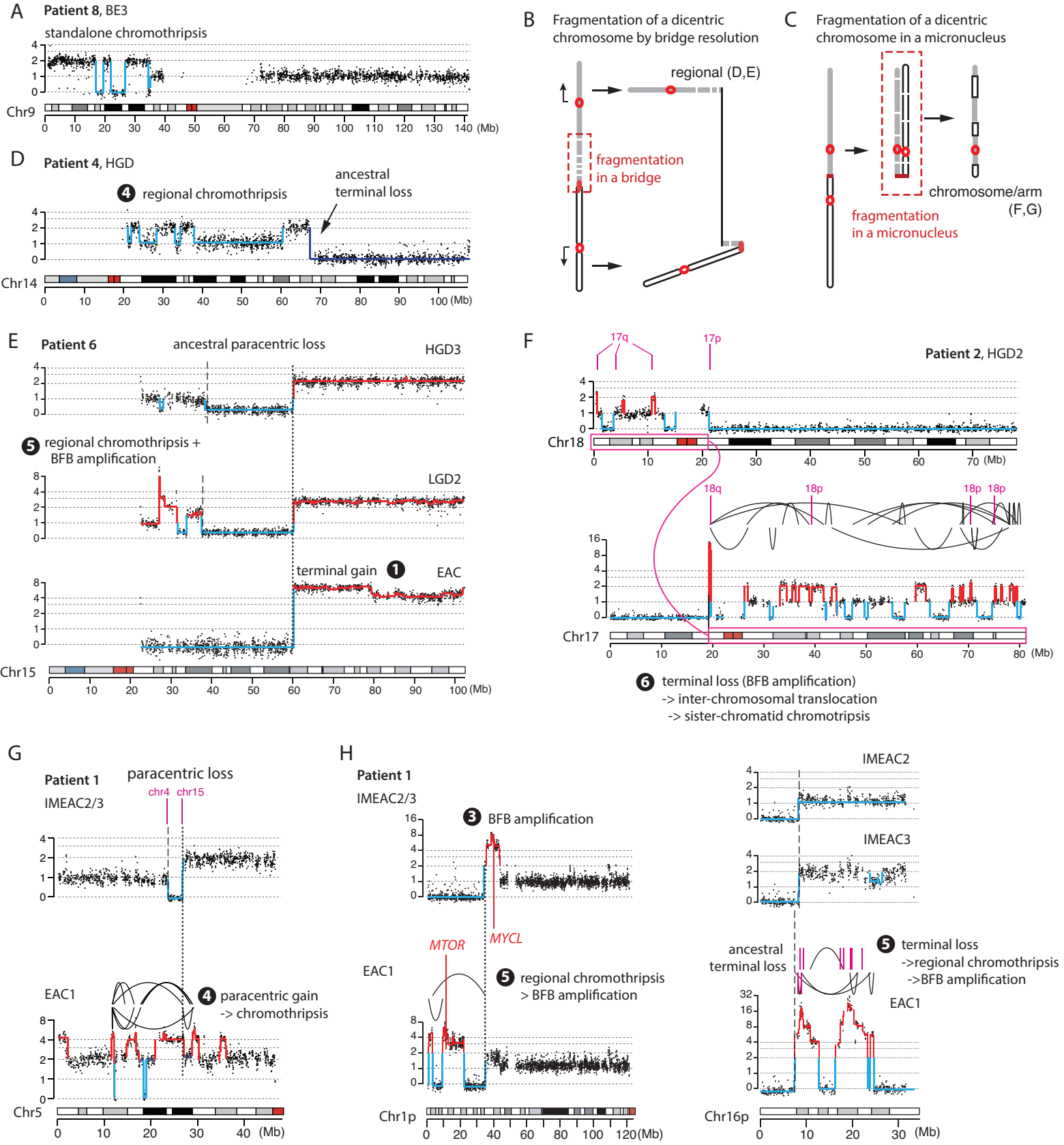


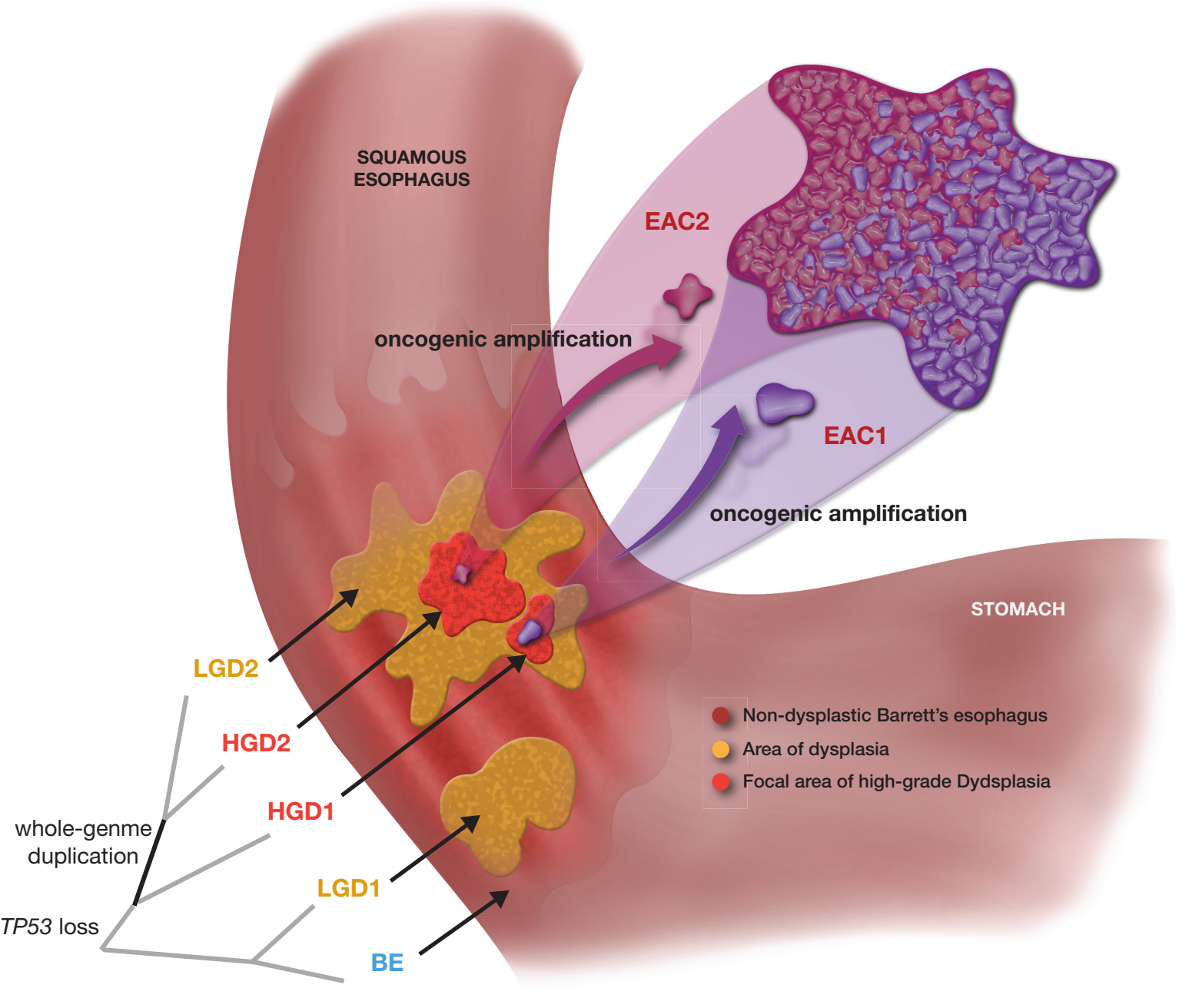
G



H




A
Patient 8, BE3
standalone chromothripsis
Chr9
B
Fragmentation of a dicentric chromosome by bridge resolution
regional (D,E)
fragmentation in a bridge
C
Fragmentation of a dicentric chromosome in a micronucleus
chromosome/arm (F,G)
fragmentation in a micronucleus
D
Patient 4, HGD
4 regional chromothripsis
ancestral terminal loss
Chr14
E
Patient 6
ancestral paracentric loss
HGD3
5 regional chromothripsis + BFB amplification
LGD2
terminal gain 1
EAC
Chr15
F
Patient 2, HGD2
17q
17p
Chr18
18q
18p
18p 18p
Chr17
6 terminal loss (BFB amplification)
-> inter-chromosomal translocation
-> sister-chromatid chromotripsis
G
Patient 1
IMEAC2/3
paracentric loss
chr4 chr15
EAC1
4 paracentric gain
-> chromothripsis
Chr5
H
Patient 1
IMEAC2/3
3 BFB amplification
MTOR
MYCL
EAC1
5 regional chromothripsis
> BFB amplification
Chr1p
IMEAC2
IMEAC3
ancestral terminal loss
5 terminal loss
->regional chromothripsis
->BFB amplification
EAC1
Chr16p


SQUAMOUS
ESOPHAGUS
EAC2
oncogenic amplification
EAC1
oncogenic amplification
STOMACH
LGD2
HGD2
HGD1
LGD1
BE
whole-genme
duplication
TP53 loss
Non-dysplastic Barrett's esophagus
Area of dysplasia
Focal area of high-grade Dydsplasia