Check for updates

Structural variant evolution after telomere crisis

Sally M. Dewhurst[1,6], Xiaotong Yao[2,3,4,6], Joel Rosiene[3,4], Huasong Tian[3,4], Julie Behr[2,3,4], Nazario Bosco[1,5], Kaori K. Takai[1], Titia de Lange [1✉] & Marcin Imieliński [3,4✉]

Telomere crisis contributes to cancer genome evolution, yet only a subset of cancers display breakage-fusion-bridge (BFB) cycles and chromothripsis, hallmarks of experimental telomere crisis identified in previous studies. We examine the spectrum of structural variants (SVs) instigated by natural telomere crisis. Eight spontaneous post-crisis clones did not show prominent patterns of BFB cycles or chromothripsis. Their crisis-induced genome rearrangements varied from infrequent simple SVs to more frequent and complex SVs. In contrast, BFB cycles and chromothripsis occurred in MRC5 fibroblast clones that escaped telomere crisis after CRISPR-controlled telomerase activation. This system revealed convergent evolutionary lineages altering one allele of chromosome 12p, where a short telomere likely predisposed to fusion. Remarkably, the 12p chromothripsis and BFB events were stabilized by independent fusions to chromosome 21. The data establish that telomere crisis can generate a wide spectrum of SVs implying that a lack of BFB patterns and chromothripsis in cancer genomes does not indicate absence of past telomere crisis.

[1] Laboratory of Cell Biology and Genetics, Rockefeller University, New York, NY, USA. [2] Tri-Institutional Ph.D. Program in Computational Biology and Medicine, Weill Cornell Medicine, New York, NY, USA. [3] Department of Pathology and Laboratory Medicine, Englander Institute for Precision Medicine, Institute for Computational Biomedicine, and Meyer Cancer Center, Weill Cornell Medicine, New York, NY, USA. [4] New York Genome Center, New York, NY, USA. [5] Present address: Department of Biochemistry and Molecular Pharmacology, Institute for Systems Genetics, NYU Langone Health, New York, NY, USA. [6] These authors contributed equally: Sally M. Dewhurst, Xiaotong Yao. ✉email: delange@rockefeller.edu; mski@mskilab.org

Structural variation is a hallmark of cancer genomes. Recent pan-cancer whole-genome sequencing (WGS) studies have revealed a more complete picture of the spectrum of structural variants (SVs) found in cancer genomes, ranging from simple deletions, duplications, and translocations to complex and often multichromosomal rearrangements[1–3]. The PCAWG consortium cataloged WGS variants across >2500 cases spanning 38 tumor types[4] to identify novel classes of complex SVs and cluster these into signatures, mirroring previous work in the categorization of single-nucleotide variants (SNVs) into distinct mutational processes[5–8]. The analysis of genome graphs provides a rigorous and unified framework to classify simple and complex SVs (including chromothripsis, breakage-fusion-bridge (BFB) cycles, and double minutes), identify novel event classes, and study the rearranged structure of aneuploid alleles[3].

However, despite advances in the identification and classification of structural variations, a mechanistic understanding of the underlying causes is often still lacking. SV mutational processes may have a more complex etiology than those driving the formation of SNVs and generate a more complex spectrum of patterns: layers of simple SVs can reshape a locus gradually and across multiple alleles, and complex SVs can rapidly rewire many genomic regions. In addition, multiple underlying causes can lead to the same type of rearrangement, and diverse outcomes can originate from a single cause. Further, it has been expensive and technically challenging to delineate specific mechanisms, although some progress has been made[9–14].

Telomere crisis, which is thought to occur at an early stage of carcinogenesis before a telomere maintenance mechanism is activated[15], has been suggested as a cause of cancer genome SVs. A priori, the genomic consequences of telomere crisis are predicted to be profound: critically short telomeres in human cells can trigger a DNA damage response, and inappropriately engage DNA repair pathways resulting in telomere-to-telomere fusions[16,17]. Subsequent cell divisions in the presence of fused dicentric chromosomes have long been considered a mechanism driving complex chromosomal rearrangements such as BFB cycles in tumors[18,19]. The characteristic fold-back inversions of BFB cycles are known to contribute to tumorigenesis in acute lymphocytic leukemia (ALL)[20], as well as squamous cell cancers and esophageal adenocarcinoma[3]. Modeling of telomere crisis in late-generation telomerase-deficient mice lacking p53 showed that telomere dysfunction engenders cancers with non-reciprocal translocations, as well as focal amplifications and deletions in regions relevant to human cancers[21,22]. Furthermore, mouse models of telomerase reactivation after a period of telomere dysfunction showed that acquisition of specific copy number aberrations and aneuploidy could drive malignant phenotypes[23].

Studies in cultured human cells have also illuminated the genomic consequences of telomere dysfunction. Even a single artificially deprotected telomere can fuse with multiple intra- and inter-chromosomal loci leading to complex fusion products[24] but it is unclear whether these complex rearrangements are compatible with viability and escape from telomere crisis. The resolution of dicentric chromosomes induced by overexpression of a dominant-negative allele of the telomere binding protein TRF2 can lead to the dramatic chromosome shattering phenomenon of chromothripsis[11,14,25]. However, to date, the only study directly investigating the consequences of a sustained period of telomere dysfunction failed to identify any complex rearrangements in HCT116 colon carcinoma cells[26]. This may be because these cells readily escaped from the telomere dysfunction that was induced by the expression of a dominant-negative hTERT (telomerase reverse transcriptase) allele. In genetically unstable HCT116 cells deficient for non-homologous end joining (NHEJ) factors, complex chained SVs were observed after telomere dysfunction, but the relevance of these types of rearrangements to human cancer remains unclear[26].

Given the expanding repertoire of structural variation present in so many cancer types and the potential contribution of telomere dysfunction to some of these aberrations, we set out to characterize the extent and type of structural variation that can be unleashed by telomere crisis and subsequent genome stabilization by telomerase expression. We approached this problem in two ways. First, we performed whole-genome sequencing (WGS) on a panel of nine previously isolated cell lines that had escaped telomere crisis spontaneously through telomerase activation. In the post-crisis immortalized cell line panel, the consequences of telomere crisis were varied, ranging from relatively unperturbed to highly rearranged genomes. Importantly, neither BFB cycles nor chromothripsis was universally observed. Second, we created a controlled in vitro telomere crisis system by engineering an MRC5-derived cell line, in which telomerase could be activated during telomere crisis and analyzed the resulting post-crisis clones by WGS. In this system, telomere crisis often engendered structures reminiscent of BFB cycles and chromothripsis. Together these data establish that the genomic consequences of telomere crisis are not readily predictable and do not invariably include BFBs and chromothripsis. Therefore, it is currently not possible to infer whether telomere crisis occurred in the proliferative history of cancers based on the pattern of SVs.

Results

Genomic complexity after spontaneous telomerase activation.
In order to determine the SVs in post-telomere crisis genomes, we examined nine SV40 large T-transformed cell lines that had undergone spontaneous telomerase activation after passage into telomere crisis (Supplementary Table 1 and Supplementary Fig. 1A). The cell lines represent independent immortalization events in a variety of cell lineages[27–29]. We carried out whole-genome sequencing of these nine post-crisis cell lines and their pre-crisis counterparts to a median depth of 40X (range: 15–51) and generated junction-balanced genome graphs[3] via JaBbA from SvABA[30] and GRIDSS[31] junction calls (see "Methods").

Using short-read WGS data, JaBbA optimally assigns a copy number to both vertices (intervals) and edges (junctions, adjacencies) of genome graphs by fitting a probabilistic model to binned genome-wide read depth. These graphs obey a basic stoichiometric constraint of DNA dosage, namely that every copy of every (interstitial) segment must have a left and a right neighbor. The topology of these genome graphs can be further analyzed to identify simple and complex SV events, including chromothripsis and BFB cycles.

Comparison of ancestral (pre-crisis) and derived (post-crisis) genome graphs showed that eight of nine post-crisis cell lines acquired virtually all (61.9–100%, median 96.6%) of their observed structural variation during or after crisis (Fig. 1a and Supplementary Fig. 1B–E). One cell line (SW13) had acquired significant aneuploidy and genome rearrangement prior to the crisis and was therefore difficult to interpret (Supplementary Fig. 1B). The other eight post-crisis genome graphs demonstrated varying levels of aneuploidy (ploidy ranges: 1.9–3.4) with variable numbers of clonal junctions per genome (range: 5–115, median 25). Analysis of junction-balanced genome graphs[3] revealed complex multichromosomal gains in six samples, with the other two lines harboring only broad arm level losses or gains (Fig. 1a).

Strikingly, besides one instance of chromothripsis (Fig. 1b), genome graph-based categorization of complex SVs[3] identified few classic footprints of chromothripsis or BFB cycles in these genomes. However, several amplified subgraphs were associated with stepwise copy number gains reminiscent of BFB cycles
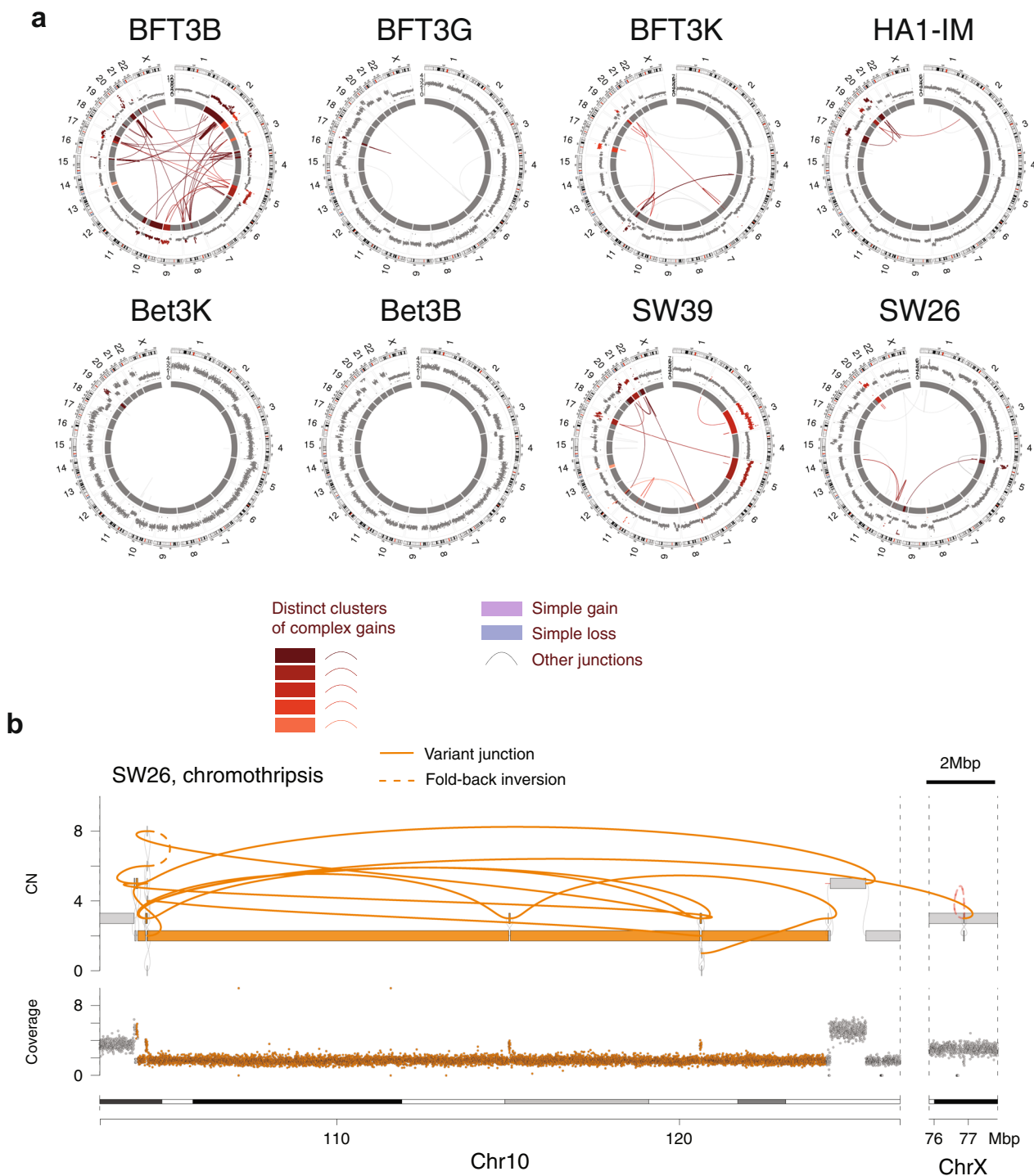
Fig. 1 Genomic complexity after spontaneous telomerase activation. a CIRCOS plots showing eight cell lines that emerged spontaneously from telomere crisis (Supplementary Table 1), five of which show one or more clusters of complex gains. Binned purity- and ploidy-transformed read depth is shown in the periphery, with colored links in the center representing variant (rearrangement) junctions. A series of red colors is used to show junctions and read-depth bins belonging to distinct clusters of complex gains in each cell line. Additional colors describe junctions and bins, including those belonging to simple losses and gains (see "Methods" for details regarding junction and bin classifications). **b** A chromothripsis event in SW26. From bottom to top, chromosomal bands, purity–ploidy-transformed binned coverage data, JaBbA reconstructed copy number with gray and colored edges indicating variant junctions. Dashed colored edges represent fold-back inversion junctions.

(Supplementary Fig. 1D). The majority of copy changes in these subgraphs could not be attributed to fold-back inversion junctions (a hallmark of BFB cycles) but were instead driven by a spectrum of duplication and translocation-like junctions and templated insertion chains. These patterns are exemplified in a 10 Mbp region of 20q of post-crisis cell line BFT3B that is amplified to 10–15 copies, incorporating Mbp scale fragments from 11 other chromosomes at lower copy number, including chromosomes 8 and 19 (Supplementary Fig. 1E). Of note, five of eight cell lines showed modest increases in *TERT* copy number,

providing a possible genomic basis for an escape from telomere crisis (Supplementary Fig. 1C).

In summary, across the eight post-crisis cell lines, spontaneous escape from the crisis was associated with a highly variable spectrum of SV patterns, ranging from relatively unaltered genomes to complex noncanonical patterns of amplification as well as numerical gains and losses. Importantly, BFB-like patterns and chromothripsis were not a general feature of the post-crisis genomes.

An in vitro system for telomerase-mediated escape from natural telomere crisis. To gain a clearer insight into the nature of SVs that arise during telomere crisis, we developed an in vitro system in which we could reproduce telomere crisis and generate a large number of post-crisis clones. MRC5 human lung fibroblasts were chosen to model telomere crisis since they lack telomerase activity and as a consequence have a well-defined in vitro replicative potential determined by telomere attrition. To bypass senescence, the Rb and p21 pathways were inactivated by infecting the population of MRC5 cells with retrovirus-bearing shRNAs targeting the respective transcripts (Supplementary Fig. 2A). This population of MRC5/Rbsh/p21sh was then endowed with an inducible CRISPR activation system (iCRISPRa) to activate the TERT promoter and induce telomerase expression (Supplementary Fig. 2B). The iCRISPRa system employed a doxycycline-inducible nuclease-dead Cas9 fused to a tripartite transcriptional activator (VP64-p65-Rta)[32] and four gRNAs targeting the TERT promoter (Fig. 2a and Supplementary Fig. 2B). The addition of doxycycline (dox) to MRC5/Rbsh/p21sh/iCRISPRa-TERT cells resulted in induction of TERT mRNA within 96 h, whereas without dox, TERT transcripts are undetectable in this cell line ($P < 0.001$, Fig. 2b). A similar dox-induced increase in mRNA expression was noted upon the introduction of sgRNAs to a control gene (Supplementary Fig. 2C). Induction of telomerase activity was readily detectable in a TRAP (telomerase-repeated amplification protocol) assay (Fig. 2c). However, the induced TERT mRNA levels and the TRAP activity were significantly lower than in telomerase-positive control cell lines. The relatively weak telomerase activity in this system harmonizes with recent work showing that cancer-associated TERT promoter mutations initially result in low levels of telomerase activity that is not sufficient to maintain bulk telomere length[33].

At approximately 120 days after the start of the experiment (55 days with dox), the MRC5/Rbsh/p21sh/iCRISPRa-TERT population was proliferating faster than their untreated counterparts (Fig. 2d). Inspection of individual telomere lengths using single-telomere-length analysis (STELA[34]) revealed that although telomerase expression was sufficient to allow the cells to proliferate, it was not sufficient to maintain bulk telomere length (Fig. 2e). After 150 days of continuous culture, the majority (86%) of XpYp telomeres in induced MRC5/Rbsh/p21sh/iCRISPRa-TERT cells were between 1 and 4 kb compared to 40% in uninduced cells (Fig. 2f and Supplementary Fig. 2D). Consistent with this, genomic blotting showed bulk telomere shortening in both induced and uninduced cells (Fig. 2g). These telomere dynamics are consistent with the expectation that in the culture without telomerase, cells with critically short telomeres will preferentially be lost, leading to a surviving population with relatively longer telomeres. In contrast, cells in the induced culture with (low) telomerase activity have the ability to elongate the shortest telomeres. As a result, the induced cells are expected to tolerate telomere attrition better and present with overall shorter telomeres at later time points.

Dissipating telomere crisis in MRC5/Rbsh/p21sh/iCRISPRa-TERT cells. To confirm that the MRC5/Rbsh/p21sh/iCRISPRa-TERT cells experienced telomere crisis before the induction of telomerase increased their proliferation rate, we investigated cells at various time points from the start of the experiment (Fig. 2d). Metaphase spreads showed both induced and uninduced MRC5/Rbsh/p21sh/iCRISPRa-TERT cells contained dicentric and multicentric chromosomes (Fig. 3a) and genomic blots showed high-molecular weight telomere bands consistent with fused telomeres (Fig. 2g). As expected from the ability of telomerase to counteract the formation of critically short telomeres, at 125 days after the start of the experiment, induced cells had significantly fewer fusions than untreated cells (21% vs. 40%, $P < 0.05$; Fig. 3b and Supplementary Fig. 3). PCR-mediated detection of fusions between the Tel Bam 11 family of telomeres[35,36] confirmed these dynamics (Fig. 3c). Quantification of the fusion frequency showed a significant reduction in the number of fusions per haploid genome in the induced population (day 110, $P < 0.01$; Fig. 3d). Consistent with telomerase-mediated genome stabilization, there was a trend toward a lower level of 53BP1-marked DNA damage foci at later time points (Fig. 3e, f) and the percentage of cells with micronuclei (an indicator of genome instability) was significantly reduced at day 110 ($P < 0.05$; Fig. 3g). Taken together, these data indicate that after a period of genomic instability induced by critically short telomeres, iCRISPRa-mediated telomerase activation is sufficient to partially stabilize the genome and allow the MRC5/Rbsh/p21sh/iCRISPRa-TERT cells to navigate the deleterious effects of telomere crisis.

Genomic screening of post-crisis clones. To assess the genome structure of proliferating post-crisis cells, single-cell clones were isolated from induced MRC5/Rbsh/p21sh/iCRISPRa-TERT cells at day 120 ("Y clones") and day 150 ("Z clones") (Fig. 4a). The clonal yield at day 150 was greater than at day 120 in induced cells, but no clones could be isolated from the uninduced population at either time point. The lower clonal yield at day 120 may be due to incomplete stabilization of the telomeres since clones from this time point showed a higher burden of fused telomeres than those derived from day 150 (Supplementary Fig. 4A). Post-crisis clones from both time points showed evidence of ultrashort telomeres and reduced telomere length (Supplementary Fig. 4B, C). Telomerase activity in post-crisis clones was comparable to the parental induced population, indicating that clone viability was not due to selection for increased telomerase activity (Supplementary Fig. 4D). To generate control clones that had not passed through a period of telomere crisis, early-passage MRC5 cells were infected with a retrovirus expressing hTERT and single-cell clones were isolated (Supplementary Fig. 4E). Genome profiling with a low-pass (~5X) WGS was performed on eight hTERT-expressing control clones (CT clones), 36 Y clones from day 120, and 82 Z clones from day 150 (Supplementary Table 2).

Analysis of genome-wide read depth across 118 clones from both day 120 (Y clones) and day 150 (Z clones) demonstrated predominantly diploid genomes with a striking enrichment of clones with DNA loss on most of chromosome 12p (63%, 74/118, Fig. 4b). Within the other 44 samples, we observed a subset of clones (5%, 6/118) with gains of chromosome 21q. As expected, control CT clones showed no evidence of SVs or copy number variants (Supplementary Fig. 4F). Hierarchical clustering of all clones by their coverage on chromosomes 12p and 21q revealed six distinct clusters (Fig. 4c). A minority of clones were diploid on chromosomes 12 and 21 and elsewhere in the genome and are therefore designated as "unrearranged" (32% of clones, 38/118). Of note, the unrearranged group was enriched in day 120 (Y) samples compared to day 150 (Z) samples ($P = 1.79 \times 10^{-9}$, odds ratio 14.7, Fisher's exact test; Fig. 4c), suggesting that these clones may have largely avoided crisis prior to telomerase induction. The cluster of clones with 21q gain was diploid on 12p.
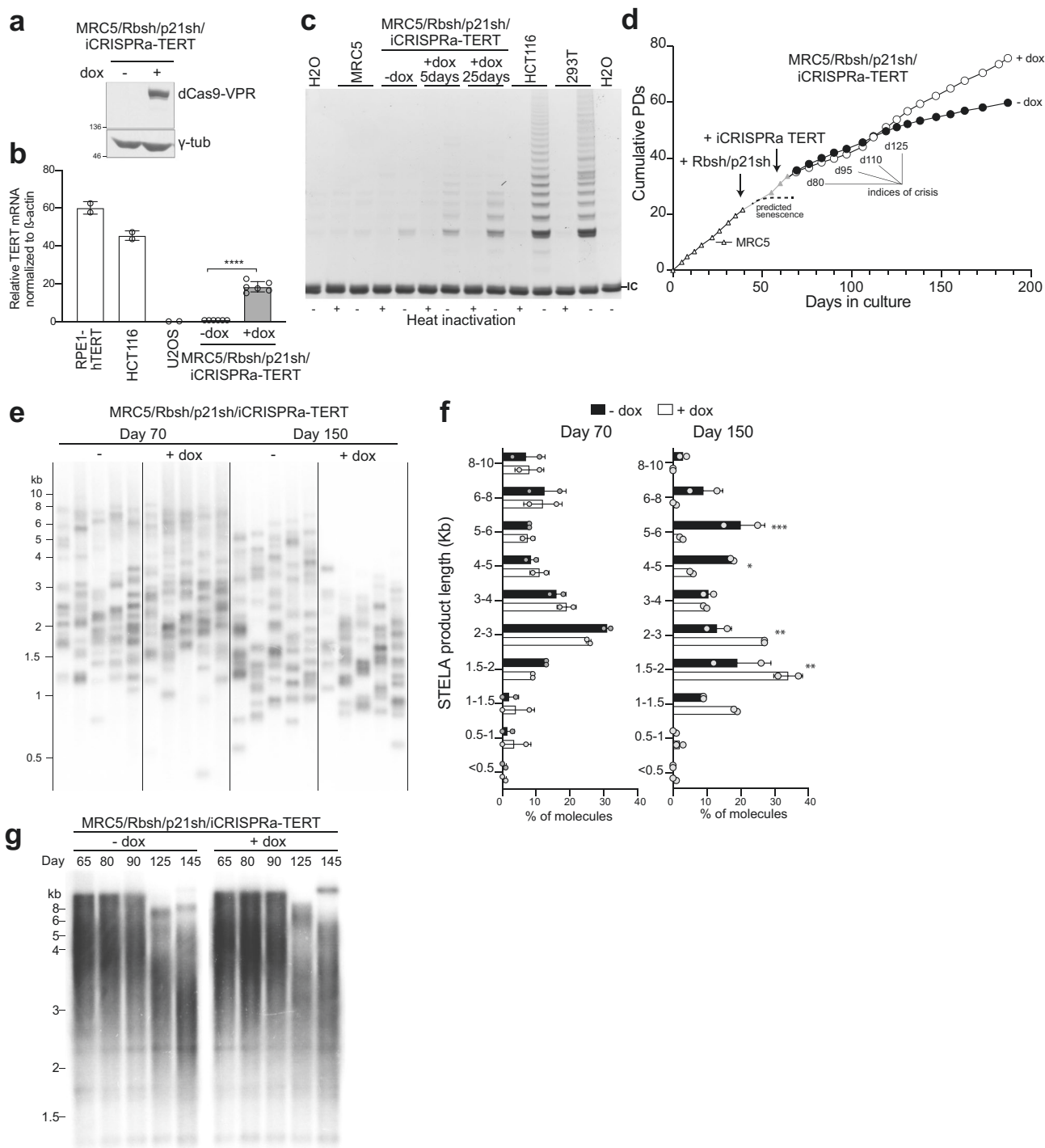
The remaining 74 clones (63%) all showed a heterogeneous pattern of copy number alterations targeting 12p (Fig. 4c). One out of the 118 clones (0.8%) displayed the singular pattern of distinct interspersed losses that resembled chromothripsis. Complete loss of one copy of 12p ("arm loss") was found in a cluster of six clones (6/118, 5%). The second cluster of 67 clones all shared a breakpoint near the distal end of 12p and a large deletion starting ~9 Mbp from the centromere. These clones were differentiated into two clusters by the presence or absence of an amplification ~8–9 Mbp from the 12p telomere. In the 47 clones that contained this amplification, aggregated consensus read-depth profiles revealed stepwise gains at the distal end of 12p, a pattern reminiscent of BFB cycles (Supplementary Fig. 5A). This cluster was therefore labeled "BFB-like", a designation that is further supported by the data presented below. The 20 clones (17%) that lack the amplicon ~8–9 Mbp harbored varying boundaries of the shared larger deletion; based on the analysis described below, we designate these as "early BFB-like". In summary, these low-pass WGS copy number profiles indicated a limited set of distinct lineages surviving telomere crisis, with at least two lineages independently converging on 12p.

High-resolution reconstruction and lineage of post-crisis genomes. To gain further insight into structural variant evolution

5

Fig. 2 An in vitro system for a telomerase-mediated escape from natural telomere crisis. a Immunoblot for dCas9-VPR (using a Cas9 Ab) in MRC5/
Rbsh/p21sh/iCRISPRa-TERT cells with or without doxycycline treatment for 96 h (see also Supplementary Fig. 2A, B). The blot shown is representative of
at least two experiments. **b** qPCR of TERT mRNA expression in RPE-1, HCT116, U2OS ($n = 2$), and MRC5/Rbsh/p21sh/iCRISPRa-TERT cells with and
without doxycycline treatment ($n = 6$). Values are normalized to β-actin mRNA. Error bars represent means ± SDs; P value from two-tailed Student's t test;
****$P < 0.0001$. **c** TRAP assay on MRC5 and MRC5/Rbsh/p21sh/iCRISPRa-TERT cells with and without doxycycline treatment for indicated time periods.
HCT116 and 293T (Phoenix) cells are included as positive controls. IC = internal control PCR product at 36 bp. The gel shown is representative of at
least two experiments. **d** Growth curve of parental MRC5 cells, MRC5/Rbsh/p21sh cells, and MRC5/Rbsh/p21sh/iCRISPRa-TERT cells grown with or
without doxycycline. Arrows indicate when each construct was introduced. Days in culture represent total time in culture from parental MRC5 cells to late
passage MRC5/Rbsh/p21sh/iCRISPRa-TERT cells. Time points for telomere analysis (presented in Fig. 3) and the approximate onset of senescence in
the parental MRC5 cells are indicated. **e** STELA of XpYp telomeres in MRC5/Rbsh/p21sh/iCRISPRa-TERT cells with or without doxycycline treatment
at 70 and 150 days of culture. **f** Quantification of band intensity in **e**, with background signal subtracted. Data from two independent experiments
(see also Supplementary Fig. 2D) were analyzed with two-way ANOVA with multiple comparisons; all points at day 70 are not significant, day 150; 5–6 kb
$P = 0.0005$; 4–5 kb $P = 0.0199$; 2–3 kb $P = 0.005$; 1.5–2 kb $P = 0.0025$. Biological replicates represent cells at approximately the same days in culture
(±5 days). **g** Genomic blot of telomeric MboI/AluI fragments in MRC5/Rbsh/p21sh/iCRISPRa-TERT cells grown with or without doxycycline at the
indicated time points. The blot shown is representative of at least two experiments.

along these lineages, we chose 13 representative clones spanning
the five clusters with rearrangements involving 12p for high-
depth WGS to a median read depth of 50× (range: 30–88).
Phylogenies derived from genome-wide SNV patterns demon-
strated a median branch length of 551 SNVs (range: 9–2409), a
low mutation density (<1 SNV/Mbp) that is consistent
with previous WGS studies of clones in cell culture[37]. This
analysis revealed four major clades (Fig. 5a). These clades had
good concordance with copy number alteration and rearrange-
ment junction patterns in the same 12p region, suggesting these
clones represent distinct post-crisis evolutionary lineages
(Fig. 5b).

In order to further reconcile the shared and distinct
rearrangement junctions present in the evolution of these clones,
we carried out a local assembly of rearrangement junctions and
junction balance analysis (see "Methods"[3]), which revealed seven
distinct junction-balanced genome graphs spanning 12p (Fig. 5c).
With the exception of the chromothriptic lineage (see below),
each of these distinct lineages was represented by more than one
post-crisis clone.

To reconstruct a set of linear alleles that parsimoniously
explain these different genome graph patterns[3] (Fig. 5c), we
applied gGnome to the data (see the section "Joint Reconstruction
of allelic evolution in MRC5"). We constrained our model to
contain one intact allele of chromosome 12 for the following
reasons: (1) karyotypes and chromosome painting showed a
single copy of chromosome 12 was altered in the post-crisis
clones (see Fig. 6); and (2) rearrangement of one allele is more
likely than rearrangement across two alleles. Application of this
constraint to the full set of MRC5 clones in a joint inference
revealed a parsimonious set of rearranged alleles that explained
the observed collection of clonally related junction-balanced
graphs (Fig. 5c).

Analyzing the clonal evolution of these rearranged 12p alleles,
we identified eight clones demonstrating progressive stages of a
BFB cycle. This complex variant evolved after a long-range
inversion junction (j1) joined a distal end of 12p to its peri-
centromere. This junction was followed by subsequent fold-back
inversion junctions (j2, j3, j4), clustered at the 8–9 Mbp focus on
12p, which are present in two different sets of post-crisis clones
(Early BFB, BFB, Fig. 5c). The earliest of the fold-back inversion
junctions (j2) in the BFB lineage was associated with a cluster of
3G or C mutations within 2 kbp of each other, consistent with
APOBEC-mediated mutagenesis[25] (Supplementary Fig. 5B). The
most complex locus in the BFB lineage (Z43, late BFB, Fig. 5c),
contained six variant junctions in *cis*, including two late tandem
duplications (j5, j6). Although j6, which connects the distal
portion of 12p to the 12p centromere, was not directly observed

in the short-read WGS data, it was imputed (dashed line, j6,
Fig. 5c) to resolve the duplication of j1 in clone Z43, as well as two
allelic ends in the genome graph. Remarkably, the vast majority
(97%) of SNVs detected in this BFB lineage (Fig. 5a) were either
shared by all clones or private to a single clone, indicating that
these stages of BFB evolution occurred rapidly in the history of
the experiment.

We confirmed a chromothripsis event in an independent
lineage (Y8), which lacked j1 and all subsequent junctions of the
BFB lineage, further supporting the idea that this is an
independent lineage (Fig. 5c). Integration of copy number data
with the SNV phylogeny showed clones from the unrearranged
lineage (Y1 and Y4) and one of the 12p arm loss clones (Y11) to
be mutationally distant (>2000 SNVs) from the chromothripsis
(Y8) and BFB lineages, which shared over 1583 SNVs (Fig. 5a).
Supporting this, a small (~21.5 kbp) simple deletion junction was
shared across Y8, Y15, and all the BFB lineage samples, yet was
absent in Y11 (Supplementary Fig. 5C).

This comparison established that the 12p loss in Y11 could not
have occurred after j1 and indicates that a second independent
arm loss must have given rise to Y15. Interestingly, the Y15 arm
loss clone was clustered in the BFB/Y8 clade in the SNV
phylogeny, sharing 30 SNVs with the BFB lineage which it did not
share with Y8 (Fig. 5a). This indicates that the 12p arm loss in
Y15 may have arisen either before or after j1. Although the
breakpoints of the Y11 and Y15 arm losses could not be mapped
due to their location in the 12 centromeric region, based on the
SNV phylogeny, they likely represent distinct events. Taken
together, these results support a model whereby at least three
lineages independently rearranged a previously wild type 12p
during telomere crisis (Fig. 5c). Our data appear to have captured
sequential steps in the formation of an increasingly complex BFB-
like event. Each of these stages must represent a stabilized allele
since the post-crisis lines are clonal, and multiple clones share the
same rearrangement junctions (Fig. 5b). This necessarily raises
the question as to what caused the on-going instability, and how
and where these complex alleles are terminated.

Resolution of BFB cycles in telomere crisis. Analysis of
junction-balanced genome graphs allows for the nomination of
"loose ends" (or allelic ends), representing copy number changes
that cannot be resolved through assembly or mapping of short
reads. We identified three distinct loose ends across the four
variant graphs spanning the eight clones in the BFB lineage
(Fig. 5c and Supplementary Fig. 5D). Each of these loose ends was
placed at the terminus of their respective reconstructed allele, and
we posit they represent the new "ends" of the derivative alleles of
the BFB lineage. Distinct ends for each of these rearranged
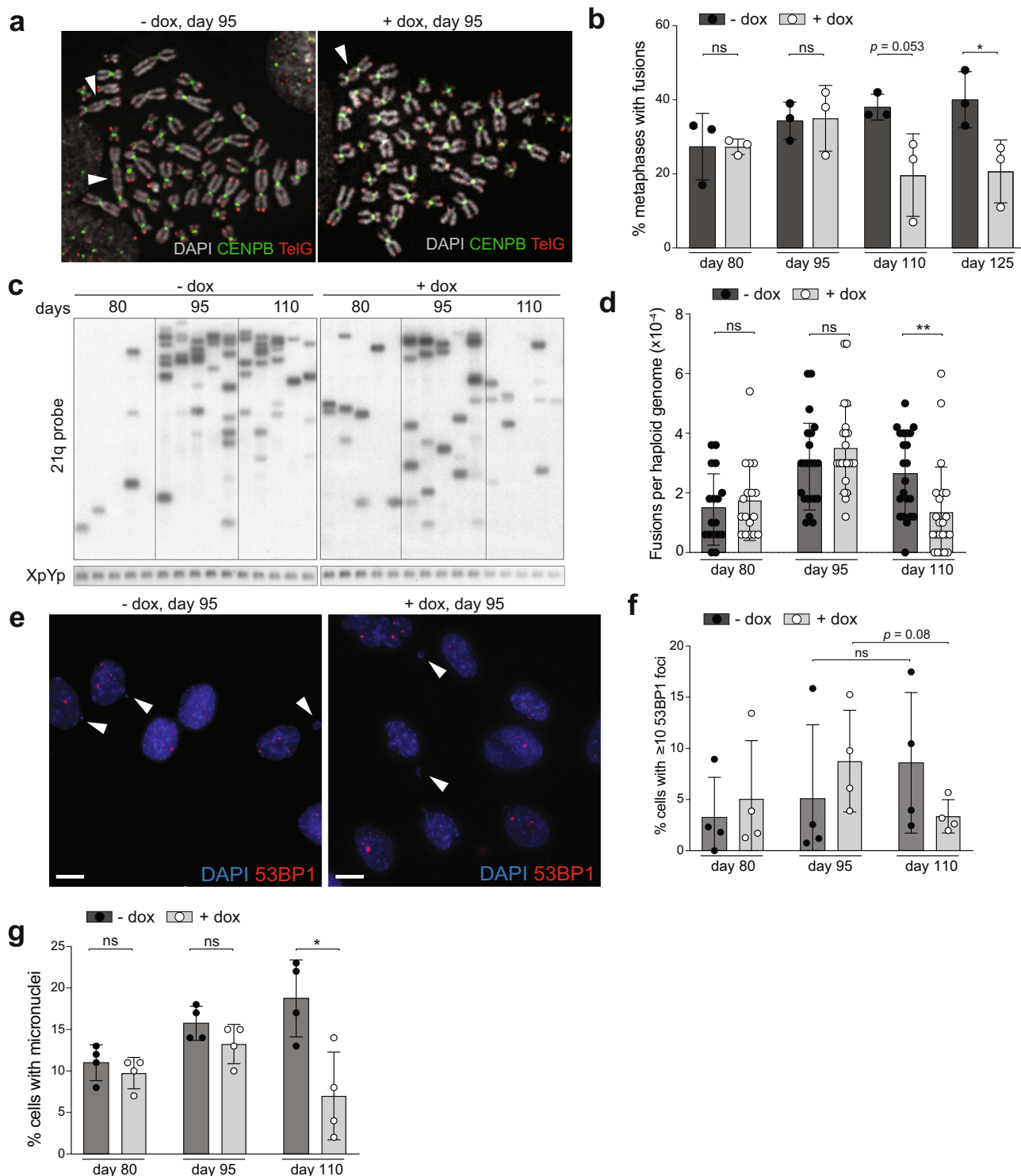
lineages suggest the derivative 12p allele could have been stabilized independently. We did not observe telomere repeat-containing reads mated to these loose ends, arguing against neo-telomere formation at these loci. Instead, loose reads represented highly repetitive unmappable sequences which may be a result of the junctions being in close proximity to centromeric regions (see below).

To resolve the genomic architecture at these loci, we generated karyotypes from metaphase spreads for representative rearranged clones (Fig. 6a and Supplementary Fig. 6A), which revealed that in the BFB and chromothripsis (Y8) lineages, the chromosome 12

derivative was likely linked to a copy of chromosome 21 with an intact long arm (Supplementary Fig. 6A). These observations were confirmed with chromosome painting, demonstrating a derivative chromosome transitioning between 12 and 21 (Fig. 6b, c and Supplementary Fig. 6B). Two possible events can explain these findings: the 12–21 fusion could have occurred as an early event during telomere crisis, preceding the divergence of Y8 (chromothriptic) and the BFB lineage; alternatively, independent 21 fusion events stabilized the derivative chromosome 12 following the formation of the distinct junction lineages in Fig. 5c. We consider the first possibility unlikely since the creation

Fig. 3 Dissipating telomere crisis in MRC5/Rbsh/p21sh/iCRISPRa-TERT cells. a Metaphase spreads from MRC5/Rbsh/p21sh/iCRISPRa-TERT cells with and without doxycycline at day 95. Telomeres are detected with a telomeric repeat PNA probe (TelG, red), and centromeres are detected with a probe for CENPB (green). DNA was stained with DAPI (gray). Chromosome fusions are indicated by white arrowheads. **b** Quantification of the percentage of metaphase spreads with at least one fusion after the indicated days of continuous culture for MRC5/Rbsh/p21sh/iCRISPRa-TERT cells with and without doxycycline (see also Supplementary Fig. 3A), two-tailed Student's t test; ns not significant, $*P = 0.0422$. **c** Gel showing products of telomere fusion PCR on MRC5/Rbsh/p21sh/iCRISPRa-TERT cells cultured with and without doxycycline for the indicated time. Each lane represents an independent replicate PCR reaction. Telomere fusion products are detected by hybridization with a probe for the 21q telomere (see "Methods"), and the control XpYp PCR product is detected with ethidium bromide staining. **d** Quantification of the number of telomere fusion products per haploid genome using the assay shown in panel **c**. Each dot represents a single PCR reaction. Reactions from two independent biological replicates are shown, two-tailed Student's t test; ns not significant; $**P = 0.0061$. **e** Detection of micronuclei (arrowheads) and DNA damage foci using indirect immunofluorescence for 53BP1 (red) in the indicated cells. DNA is stained with DAPI (blue); scale bar (white) = 10 μm. **f** Quantification of the percentage of cells with >10 53BP1 foci at the indicated time points; two-tailed Student's t test; ns not significant. **g** Quantification of the percentage of cells with micronuclei after the indicated days in culture, two-tailed Student's t test; ns not significant; $*P = 0.0157$. In panels **b**, **f**, and **g**, error bars indicate means and standard deviations from three independent biological replicates.
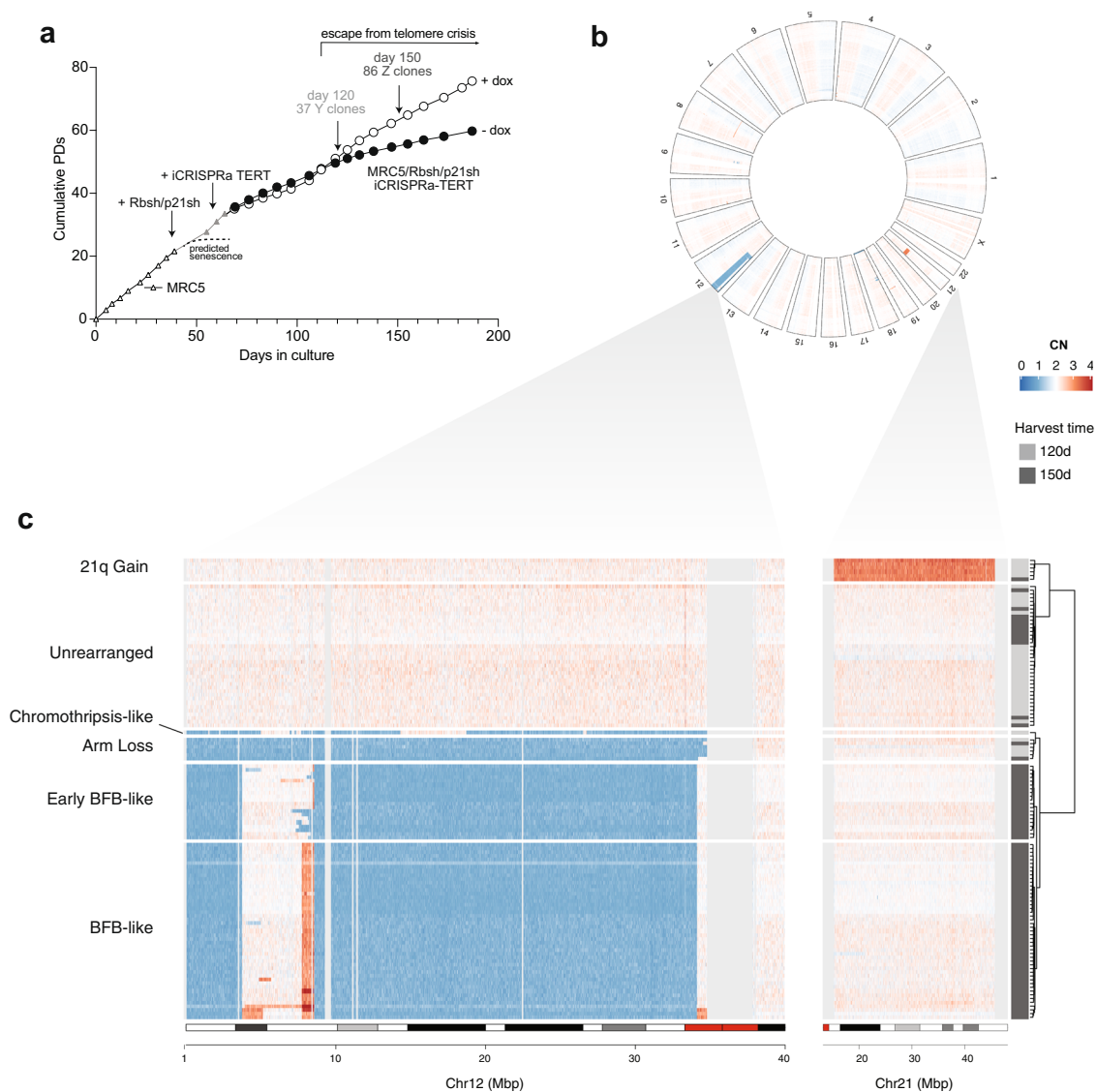

Fig. 4 Genomic screening of post-crisis clones. a Growth curve of MRC5/Rbsh/p21sh/iCRISPRa-TERT cells with and without doxycycline, indicating the time points at which single-cell clones were derived (day 120 and day 150). **b** Circular heatmap showing genome-wide binned purity- and ploidy-transformed read depth (in units of CN across 118 low-pass WGS-profiled clones. Heatmap rows correspond to concentric rings in the heatmap. Clones are clustered with respect to genome-wide copy number profile similarity (see "Methods"). **c** Zoomed-in portion of chromosomes 12 and 21 that underwent copy number alterations in a majority of the clones, clustered based on their coverage across these regions. Clusters are named with respect to their consensus copy number pattern, and on the basis of high-depth WGS analyses presented in Fig. 5. Chromosome 21 gain $n = 6$ clones; unrearranged $n = 38$; chromothripsis-like $n = 1$; arm loss $n = 6$; early BFB-like $n = 20$; BFB-like $n = 47$.
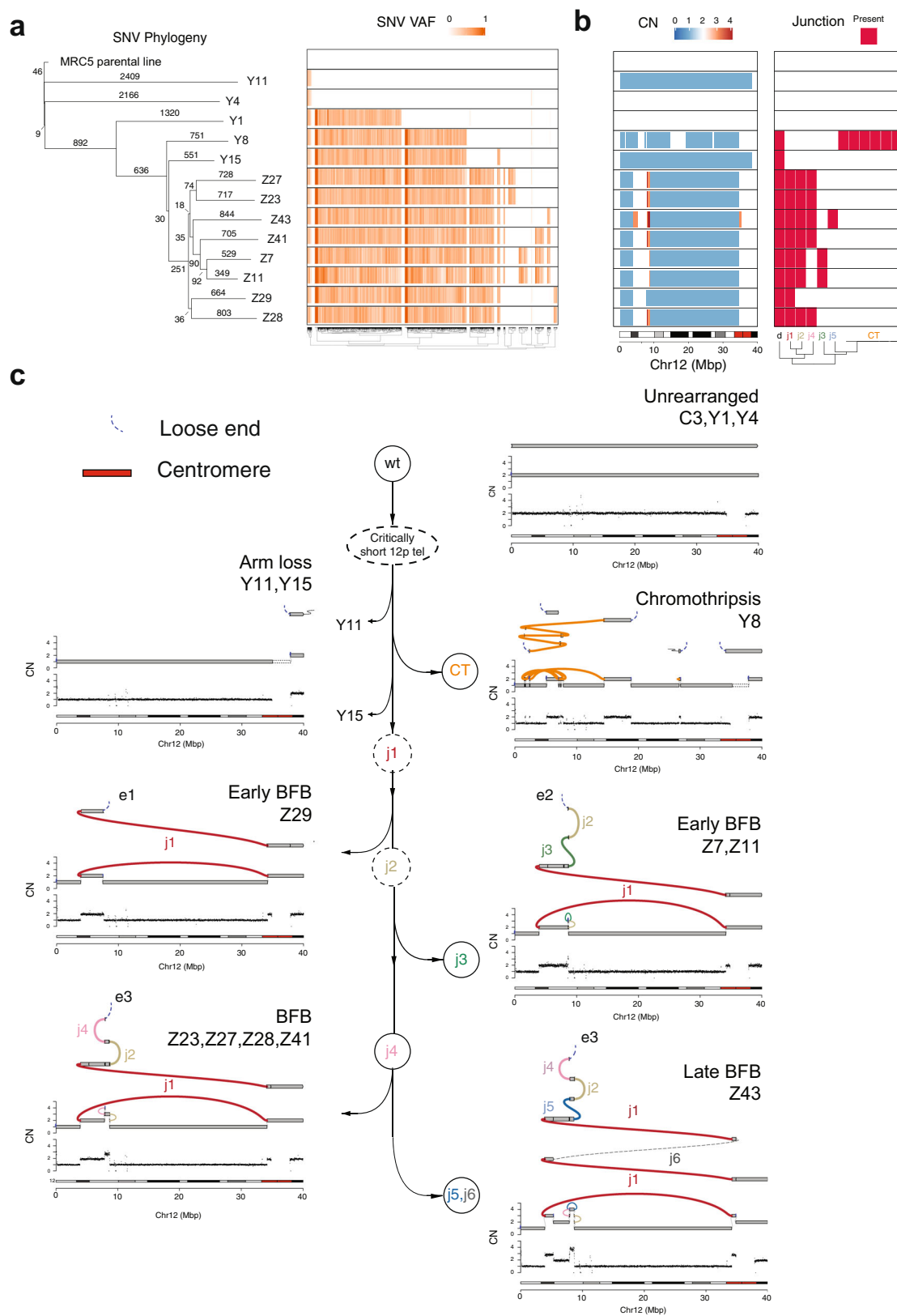
of the long-range inversion (j1) and subsequent fold-back junctions in the BFB lineage would require the formation of interstitial 12p breaks on a 12p-21 derivative chromosome. Such breaks are predicted to result in the loss of 21, which would be distal to these junctions on the fusion allele. Furthermore, the acrocentric nature of chromosome 21 would make it more likely

to stabilize the overall chromosome architecture, suggesting that an early 12–21 derivative chromosome would be unlikely to engage in the additional SV events observed in the BFB lineage. We, therefore, consider it likely that each of the BFB cycles and chromothripsis clones was independently resolved through subsequent fusion to 21 (Fig. 6c).

Fig. 5 High-resolution reconstruction and lineage of post-crisis genomes. a SNV-based phylogeny inferred across 13 high-depth WGS clones and heatmap of variant allele fractions (VAF) for SNVs detected among two or more clones. For simplicity, private SNVs (those found only in a single clone) are not shown. **b** Heatmap of chromosome 12p copy numbers and variant junction patterns in chromosome 12 (see the text and "Methods"). **c** Proposed tree showing distinct trajectories of structural variant evolution following 12p attrition and subsequent telomere crisis. Each terminal node in the tree is associated with a unique 12p profile comprising a representative binned read-depth pattern (bottom track) from one or more clones mapping to an identical junction-balanced genome graph (second track from bottom). The top track in each profile represents a reconstruction of the rearranged allele. Each allele is a walk of genomic intervals and reference/variant junctions that, in combination with an unrearranged 12p allele (not shown), sum to the observed genome graph (see "Methods"). Two distinct arrows linking Y11 and Y15 demonstrate that these clones are distinct lineages (based on divergent SNV patterns, see panel **a**), that converge to identical WGS 12p CN profiles (although with likely distinct breakpoints inside the 12p centromere unmappable by WGS, see the text).

Unlike the BFB and chromothripsis clusters, one of the two 12p arm loss lineages (Y11) did not appear to be fused to chromosome 21. In this clone, the derivative chromosome 12 appears to contain a distinct fusion (with a longer p-arm) (Supplementary Fig. 6C). This is consistent with our analysis of the SNV phylogeny, showing Y11 to be mutationally distant from the BFB lineage (Fig. 5a and Supplementary Fig. 5C). We were unable to further resolve the nature of the stabilization event in this clone. It would be necessary to perform long molecule DNA sequencing across different lineages in order to confirm the distinct nature of the fusion junction in each of the post-crisis clones.

A short telomere renders 12p vulnerable to telomere attrition. The convergent evolution patterns observed in our system suggest either 12p vulnerability to rearrangements or selection for 12p loss during telomere crisis. We believe strong selection is unlikely, given the existence of day 150 clones with diploid 12p (15.8%, 13/82, with or without 21q gain). The preferential rearrangement of the short arm of chromosome 12 in the post-crisis system could be explained if one of the two 12p telomeres is among the shortest telomeres in the MCR5 parental cells. Attrition of the shortest telomeres is predicted to generate the first telomere fusions and associated rearrangements in the culture.

We first asked whether the same parental allele was targeted across the chromosome 12-associated events in our cohort. Such allele specificity would argue against a selection for loss of 12p sequences since such selection should have occurred without allele preference. We phased heterozygous SNPs on 12p on the basis of whether they belonged to the lost (L) or retained (R) allele on the early 12p arm loss clone, Y11 (Fig. 7a). Analyzing phased SNP patterns across all the high- and low-pass MRC5 clone WGS profiles in our dataset demonstrated that the L allele of 12p was the exclusive target of all chromosome 12 structural variants (Fig. 7b and Supplementary Fig. 7A). This included the clones from the chromothripsis (Y8) and BFB (Z43) lineages (Fig. 7c), which our phylogenetic clustering suggested to be likely independent events on a previously unrearranged chromosome 12 (Fig. 5c). On the basis of these results, we concluded that the short arm of the L allele of 12 was the most vulnerable to rearrangement in the MRC5 parental line.

We next tested whether the preferential 12p events could be due to the presence of a short telomere on one of the 12p alleles. To this end, we combined telomeric FISH with BAC probes specific for chromosome 12 and two other chromosomes (6 and 8) that did not show evidence for structural variants in WGS (Figs. 7d and 4b). Comparing the ratio of the telomeric signal of the shortest 12p telomeres to the signal of all other telomeres in individual metaphase spreads revealed that one of the 12p telomeres was significantly shorter (Fig. 7e). The shortest telomeres of 6 and 18 (Supplementary Fig. 7B) were also shorter than the median but not to the same extent as 12p. The relative telomere length of the shortest 21p allele showed a heterogeneous

distribution that overall was significantly longer than 12p in the parental cells (Fig. 7e). This does not exclude the possibility of 21 becoming critically short at later time points, and indeed the observation of a low percentage of clones in the 5X WGS screening with amplification of 21q could indicate that this chromosome end did occasionally become deprotected in this population (Fig. 4c). Such deprotection of a chromosome 21 telomere is consistent with chromosome 21 preferentially stabilizing the derivative chromosome 12 (Fig. 6b and Supplementary Fig. 6A).

To look for evidence of chromosome 12 being involved in the initial fusion events in this system, we combined a chromosome 12 BAC probe with a centromere probe in MRC5/Rbsh/p21sh/iCRISPRa-TERT cells in crisis (at day 90). Strikingly, we observed a number of instances of chromosome 12 within chromosome fusion events (Fig. 7f). The fraction of chromosome fusions involving chromosome 12 is higher than expected (~50% observed versus ~4% expected, Supplementary Fig. 7C). Collectively, these data support the hypothesis that a short telomere on one allele of 12p increased the chance of 12p partaking in a fusion event that preceded subsequent rearrangement lineages.

Discussion

We have described the first whole-genome profiles of cells emerging from natural telomere crisis, both in the setting of spontaneous and controlled telomerase activation. Analysis of a variety of post-crisis genomes from divergent lineages and independent immortalization events uncovered highly complex patterns of copy number amplification and rearrangement. However, many of the genomes showed minimal genomic alterations. The rearrangements we did observe were not typified by the expected predominance of fold-back inversions that are indicative of BFB cycles or low amplitude copy number oscillations associated with chromothripsis. These cell lines spent a varying amount of time in telomere crisis, potentially with very different numbers of chromosome fusions, which is hard to quantify with limited historical data available. Due to the limited similarities between these cell lines, we constructed an in vitro system that allowed us to sequence high numbers of post-crisis genomes.

We consider our in vitro system to be a good representation of the telomere crisis for a number of reasons. Telomeres in this system have been eroded through replicative attrition, rather than being subject to acute deprotection by the removal of TRF2. This is an important distinction since telomeres lacking TRF2 are repaired by c-NHEJ, whereas other DNA repair pathways are active at naturally eroded telomeres[35,38,39]. Furthermore, the number of dicentric chromosomes in our system is low (generally 1–2 per metaphase spread), which is similar to the frequency observed in other natural telomere crisis systems[29]. Apart from the abrogation of the Rb/p21 pathways, which is considered likely to occur before telomere crisis in vivo[40–43], these cells contain intact DNA repair pathways, and we make no assumptions as to what the predominant repair mechanisms will be in this context.
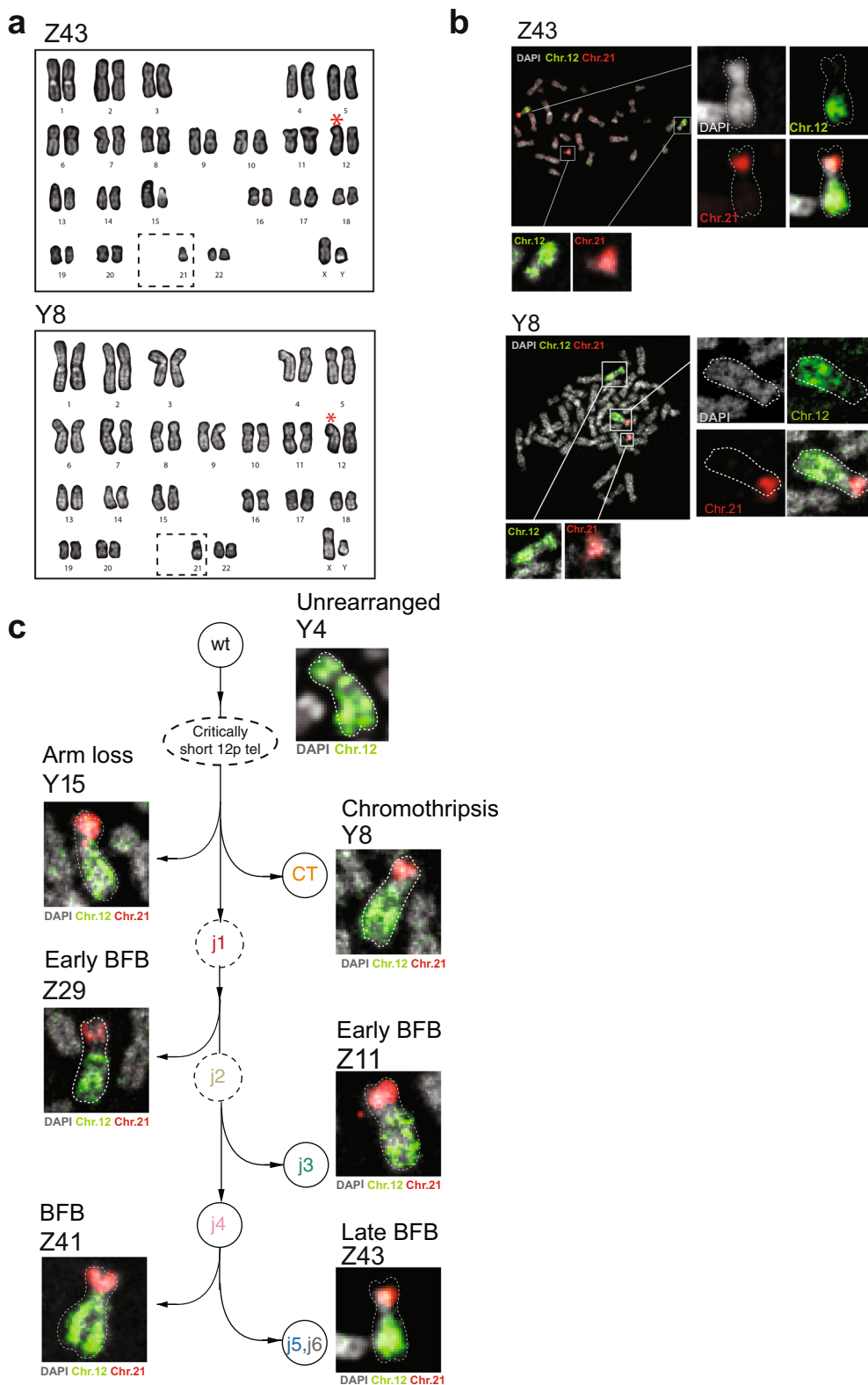
Fig. 6 Resolution of BFB cycles in telomere crisis. a DAPI banded karyotypes of post-crisis clones Z43 and Y8 showing a rearranged chromosome 12 (red star) and loss of one copy of chromosome 21 (dashed box) (see also Supplementary Fig. 6A). **b** Representative metaphase spreads of clone Z43 and Y8 hybridized with whole-chromosome pairs for chromosomes 12 (green) and 21 (red). DNA was stained with DAPI (gray). Insets show enlarged images of the 12–21 derivative marker chromosome and intact copies of the sister alleles (see also Supplementary Fig. 6B). **c** Images of derivative chromosome 12:21 from representative clones from each branch of the evolution of chromosome 12 post-crisis (according to the analysis in Fig. 5c). Metaphases were hybridized with whole-chromosome pairs for 12 (green) and 21 (red). DNA was stained with DAPI (gray).
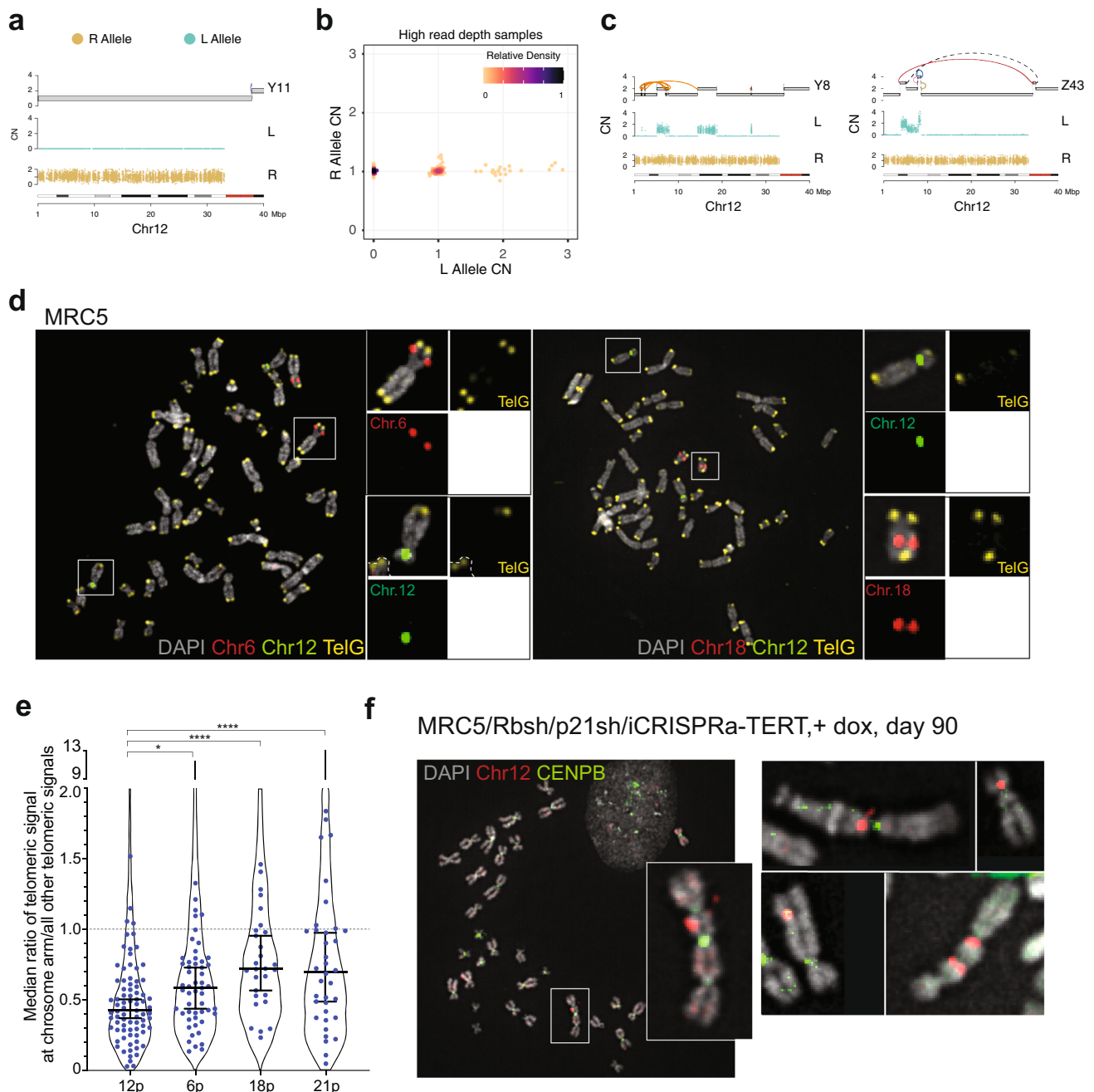
Fig. 7 A short telomere on one allele of 12p. a Genomic track plots of parental alleles phased into lost ("L") and retained ("R") haplotypes (see "Methods") on chromosome 12p of clone Y11. **b** Scatter plot showing purity- and ploidy-transformed L and R haplotype-specific allelic read depth across 12p segments in high-pass WGS-profiled post-crisis clones. **c** Genomic track plots of allelic read counts on the L and R allele of clones Y8 and Z43, two post-crisis clones that independently acquired structural variants on an otherwise unrearranged chromosome 12p allele. **d** Metaphase spreads of early-passage MRC5 cells hybridized with BAC probes to chromosome 12 (green) combined with probes for chromosome 6 or 18 (red) and a PNA probe for telomeres (TelG, yellow). Insets of white-boxed chromosomes are shown with each channel individually. **e** Quantification of the relative length of the shortest of the two 12p telomeres. Each dot shows the median ratio of the TelG signal of the shortest telomeres of the indicated chromosome arm to all other telomeres in each metaphase spread. Violin plots show the data from all telomeres analyzed. Chromosome 12 was identified using a specific BAC probe (Chr.12p11.2) in 79 metaphase spreads with a total of 3992 telomeres. Chromosomes 6 and 18 were identified based on BAC probe hybridization (Chr.6p21.2–21.3, Chr.18q12.3–21.1) in 53 and 28 metaphases, respectively (2629 and 1497 telomeres, respectively). Chromosome 21 was identified from DAPI banding patterns in 36 metaphases (1757 telomeres). P values were derived from a two-sided Student's t test; $*P = 0.0147$; $****P < 0.0001$. Error bars show median with 95% CI. **f** Dicentric chromosomes containing chromosome 12 in telomere crisis. Metaphase spreads from MRC5/Rbsh/p21sh/iCRISPRa-TERT cells cultured with doxycycline at day 90 (during crisis) were hybridized with a BAC probe for Chr.12p11.2 (red) and a CENPB PNA probe (green) to identify centromeres. A full spread is shown with white box inset zoom in. Further examples from other spreads are also shown (see also Supplementary Fig. 7B).

Further, the relatively weak telomerase activity that can only sustain the shortest telomeres within the population is similar to what occurs in cancer, since many tumors maintain very short telomeres despite activation of telomerase[44–47].

This system revealed striking convergent evolution of rearrangements on chromosome 12p, for which we consider the most plausible explanation to be a short telomere on one of the 12p alleles driving early chromosome fusions in telomere crisis. Although the rearrangement events on 12p are likely specific to this cell line, we can draw valid conclusions about the consequence of short telomeres across other systems. It seems likely that the first events during the telomere crisis are driven by the shortest telomere(s) within a cell population. These data suggest a minimal set of events that can occur as a result of a single deprotected telomere. We document clean patterns of BFB-like events that represent progressive stages in the evolution of more complex genome architectures. These data have provided an important snapshot into the events that occur during a relatively short time period of telomere crisis. The comparatively flat genomes in the majority of post-crisis clones suggest that the consequences of telomere crisis do not have to be spectacular. It may be that in this system there is selection against complex events involving multiple chromosomes. The more complex events that can be observed in the immediate aftermath of dicentric chromosome resolution[14] may not lead to viable post-crisis clones. Our data also point to a surprising role for acrocentric chromosomes in stabilizing fusion events, which has also been suggested and observed in other studies[14,48]. Since it was technically challenging to resolve these stabilizing events directly through assembly or mapping of short reads, it is possible that these types of events have been overlooked in large-scale WGS analyses and could be an important hallmark of post-crisis genomes.

In conclusion, our data reveal that telomere crisis can instigate a wide spectrum of structural variations in the viable descendants of this genomic trauma. First, our results indicate that cells can emerge from a telomere crisis with minimally altered genomes. Second, BFB cycles and chromothripsis are not a universal hallmark of post-crisis cell lines. Third, our results indicate that natural telomere crisis can manifest as a focal and diverse cascade of SV events converging on a single chromosome arm. Since no single class of structural variation appears to be a hallmark of past telomere crisis, other genomic insignia will have to be identified in order to determine whether a given cancer has experienced telomere dysfunction in its proliferative history.

Methods
Cell lines. MRC5 human lung fibroblasts (CCL-171), Phoenix-ampho (CRL-3213), RPE-1 hTERT (CRL-4000), HCT116 (CCL-247), and U2OS (HTB-96) cells were obtained from ATCC for this study. 293-FT cells were obtained from Thermo-Fisher. MRC5 cells and derivatives thereof were grown in EMEM media (ATCC) supplemented with 15% fetal bovine serum (FBS; Gibco) and 100 U/mL of penicillin and 100 μg/mL streptomycin (PenStrep, Gibco) at 37 °C, 5% CO$_2$. hTERT RPE-1 cells were grown in DMEM:F12 media (Gibco) with 10% FBS and PenStrep at 37 °C, 5% CO$_2$. HCT116 colorectal carcinoma cells and U2OS cells were grown in DMEM with 10% FBS and PenStrep at 37 °C, 5% CO$_2$.

Immortalized cell line panel. Details of the post-crisis immortalized cell line panel are provided in Supplementary Table 1. HA-1M cells were a kind gift of Silvia Bacchetti[29,49], SW13/26/39 cells were a kind gift of Jerry Shay[28], and Bet-3B/3 K and BFT3B/G/K cells were a kind gift of Roger Reddel[27].

Cloning and plasmids. A dual-shRNA vector LM2PshRB.698-p21.890-PURO was used to knockdown Rb and p21[50]. The inducible dCas9-VPR (pCW57-dCas9-VPR) construct was created by Gibson assembly of the dCas9-VPR insert from SP-dCas9-VPR (Addgene #63798) into pCW57-MCS1-P2A-MCS2-Neo (Addgene #89180). Retroviral pLVX-hTERT was a kind gift of Teresa Davoli. Activating *TERT* gRNAs were targeted up to 1000 bp upstream of the *TERT* promoter transcriptional start site and designed using online software from the Broad Institute (portals.broadinstitute.org/gpp/public/analysis-tools/sgrna-design). gRNA

sequences were cloned into a modified version of lentiGuide-Puro (Addgene #52963) in which the selection cassette had been swapped for Zeocin resistance. Activating *TERT* gRNA sequences are shown in Supplementary Table 3. *TTN* gRNA sequences were used as described[32].

Viral gene delivery. Retroviral constructs were transfected into Phoenix amphitropic cells using calcium phosphate precipitation. Lentiviral constructs were transfected with appropriate packaging vectors using calcium phosphate precipitation into 293-FT cells. Viral supernatants were collected and filtered before addition to target cells, supplemented with 4 μg/ml polybrene. For activating gRNA constructs, multiple viral supernatants were collected and concentrated using PEG-it Virus Precipitation Solution (System Biosciences LV810A-1). Cells were infected two to three times at 12-h intervals before selection in the appropriate antibiotic.

Immunoblotting. For immunoblotting, cell pellets were directly lysed in 1× Laemmli buffer (2% SDS, 5% β-mercaptoethanol, 10% glycerol, 0.002% bromophenol blue, and 62.5 mM Tris-HCl pH 6.8) at a concentration of 10^7 cells/ml. Lysates were denatured at 100 °C, and DNA was sheared with a 28½ gauge insulin needle. Lysates were resolved on SDS/PAGE gels (Life Technologies), transferred to nitrocellulose membranes, and blocked with 5% milk in TBS with 0.1% Tween-20. Primary antibodies (anti-Cas9 7A9-3A3, Cell Signaling Technology #14697S 1:1000, anti-γ-tubulin Sigma #T5326 1:1000, anti- Human Retinoblastoma protein BD Pharmigen #554136 1:500, anti-p21 F-5 Santa Cruz sc-6246 1:200) were incubated overnight, before membrane washing and incubation with appropriate HRP-conjugated secondary antibodies (Amersham, NA934 and NXA931, 1:20,000) and detection with SuperSignal ECL West Pico PLUS chemiluminescence (ThermoFisher).

Immunofluorescence. Cells were grown on glass coverslips and fixed in 3% paraformaldehyde and 2% sucrose. Coverslips were permeabilized in 0.5% Triton-X-100/PBS, and blocked in goat block (0.1% BSA, 3% goat serum, 0.1% Triton-X-100, 2 mM EDTA) in PBS. Primary and secondary antibodies (Rabbit anti-53BP1 Abcam #ab-175933 1:1000, F(ab')2-goat anti-rabbit IgG (H + L) Cross-adsorbed Alexa Fluor 488 ThermoFisher A-11070, 1:500) were diluted in goat block. Slides were counter-stained with DAPI and mounted using prolong gold antifade medium. Images were acquired on a DeltaVision microscope (Applied Precision) equipped with a cooled charge-coupled device camera (DV Elite CMOS Camera), with a PlanApo ×60 1.42 NA objective (Olympus America), and SoftWoRx software. Images were analyzed for foci numbers using a custom-made algorithm written for FIJI, courtesy of Leonid Timashev[51].

Metaphase spread preparation and staining. Metaphase spreads were prepared by treatment of cells with 0.1 μg/ml colcemid (Roche) for 3 h before trypsinization and swelling at 37 °C for 5–10 min in 0.075 M KCl. Cells were fixed in a freshly prepared 3:1 mixture of methanol to acetic acid and stored at 4 °C overnight or longer. Spreads were prepared by dropping cell solution onto cold glass slides exposed to steam from a 75 °C water bath, flooding slide with acetic acid, before exposure of the dropped cells for 3–5 s in steam. Slides were dried overnight before storage in 100% ethanol at −20 °C. For visualization of fusions, slides were rinsed in PBS, fixed in 4% formaldehyde/PBS for 5 min, and dehydrated in an ethanol series before co-denaturation of the slide and PNA probes (TelG-Cy3 PNA Bio F1006, CENPB-AF488 PNA Bio F3004) for 3 min at 80 °C in hybridization solution (10 mM Tris-HCl pH 7.2, 70% formamide, 0.5% Roche 11096176001 blocking reagent). Hybridization was carried out for 2 h at RT in the dark, before washing twice in 10 mM Tris-HCl pH 7.2, 70% formamide and 0.1% BSA, then washing three times in 0.1 M Tris-HCl pH 7.2, 0.15 M NaCl, 0.08% Tween-20. DAPI was included in the second wash. Slides were dehydrated through an ethanol series before mounting with Prolong Gold antifade medium (Invitrogen).

For chromosome painting, slides were prepared as above for chromosome fusions, and co-denaturation of chromosome-specific paints (XCP-12 Metasystems D-0312-050-FI, XCP-21 Metasystems D-0321-050-OR) was carried out at 75 °C for 2 min, before hybridization overnight at 37 °C. Post hybridization washes were 0.4× SSC for 2 min at 72 °C, 2× SSC, 0.05% Tween-20 for 30 s, followed by counterstaining in DAPI for 15 min, and a rinse in ddH$_2$O before mounting in Prolong Gold antifade medium (Invitrogen). For karyotyping, slides were prepared as above, and analysis was carried out on DAPI stained chromosomes.

BAC probes. To identify individual chromosomes on metaphase spreads, BAC probes were ordered from BACPAC Genomics (Chr.12p11.2 RP11-90H7, Chr.18q12.3–21.1 RP11-91K12, Chr.6p21.2–21.3 RP11-79J17). Probe DNA was nick-translated with either Digoxigenin-11-UTP or Biotin-16-UTP (Roche) using DNase I (Roche) and DNA polymerase I (NEB) overnight at 15 °C. Probes were precipitated with Cot1 human DNA (Invitrogen) and salmon sperm DNA (Invitrogen) and resuspended in 50% formamide, 2× SSC, and 10% dextran sulfate before denaturation for 8 min at 80 °C. Metaphase spreads were prepared as above, and slides were denatured with 70% formamide, 2× SSC for 2 min at 80 °C before dehydration through an ethanol series. Slides were co-denatured for 2 min at 80 °C with TelG-647 (PNA Bio F1014) in hybridization solution (10 mM Tris-HCl pH 7.2, 70% formamide, 0.5% Roche 11096176001 blocking reagent) followed by a 2-h

hybridization at RT. Denatured BAC probes were then applied and hybridized overnight at 37 °C. Slides were washed for 3 × 5 min in 1× SSC at 60 °C, followed by a blocking step in 30 μg/ml BSA, 4× SSC and 0.1% Tween-20 for 30 min at 37 °C. BAC probes were detected with anti-digoxigenin–rhodamine (Roche 11207750910, 1:400) and Avadin-FITC antibodies (VWR CAP21221, 1:400) in 10 μg/ml BSA, 1× SSC, and 0.1% Tween-20 by incubating for 30 min at 37 °C, before washing twice for 5 min in 4× SSC and 0.1% Tween-20 at 42 °C. Counterstaining with DAPI was carried out for 15 min at RT, before a further wash at 42 °C in 4× SSC, 0.1% Tween-20, and mounting in Prolong Gold antifade (Invitrogen). Images were acquired on the Deltavision microscope equipment detailed above. Images were analyzed using FIJI[52]. Briefly, individual chromosomes were detected with the BAC probes. Intensity measurements of the TelG signal were quantified for each p-arm and q-arm of identified chromosomes. Measurements were also taken for all other telomeres in the same spread. Background subtracted measurements for all telomeres were compared to the shortest (lowest intensity) of the Chr.12p (or Chr.18p or Chr.6p) telomeres on each spread, and the results expressed as a ratio. For identification of fusions containing chromosome 12, the same protocol was carried out using a centromere probe (CENPB-AF488, PNA Bio F3004).

qPCR. RNA was isolated from cell pellets using a Qiagen RNeasy kit, according to the manufacturer's instructions. cDNA was synthesized using Superscript IV first-strand synthesis (ThermoFisher). qPCR was carried with SYBER Green reagents (ThermoFisher) and run on a Life Technologies QuantStudio 12 K machine. qPCR primer sequences are shown in Supplementary Table 3. Expression was quantified using the standard ΔΔCT method relative to β-actin.

TRAP assay. Telomerase activity was assessed using the TRAPeze kit (EMD Millipore S7700) according to the manufacturer's instructions. Amplification products were resolved on 12% PAGE gels and visualized with EtBr staining.

STELA, fusion PCR, and telomeric blots. High-molecular-weight DNA was extracted from cell pellets using a MagAttract HMW DNA kit (Qiagen) and solubilized by overnight digestion with *Eco*RI (for STELA and Fusion PCR) or a combination of *Alu*I and *Mbo*I (for telomeric blots). STELA was carried out essentially as described[34]. Briefly, 10 ng of DNA was ligated to a mixture of six telorette linkers (Supplementary Table 3) overnight at 35 °C, before dilution with water to a concentration of 200 pg/μl. Multiple PCR reactions for each sample were carried out with 200 pg of annealed DNA using the XpYpE2 and teltail primers (Supplementary Table 3) and FailSafe PCR reagents (Epicenter). PCR conditions were as follows: 94 °C for 15 s, 27 cycles of 95 °C for 15 s, 58 °C for 20 s, 68 °C for 10 min, and a final extension at 68 °C for 9 min. PCR products were resolved on 0.8% TAE gels, denatured, and transferred to the Hybond membrane via southern blotting. Products were detected with a randomly primed α-^{32}P DNA probe created by amplification of the telomere-adjacent region of the XpYp telomere (using XpYpE2 and XpYpB2 primers, Supplementary Table 3). For quantification, FIJI was used to measure the relative signal between indicated molecular weight markers relative to the background signal for each sample.

Fusion PCR was carried out essentially as described[11,35]. Subtelomeric primers (Supplementary Table 3) used for amplification of telomeric fusions were XpYpM, 17p6, and 21q4. The control primer XpYpc2tr was included for control amplification of XpYp subtelomeric DNA and detected using EtBr. Fusion products were detected with a random primed α-^{32}P-labeled (Klenow) DNA probe (21q probe) specific for the TelBam11 telomere subfamily[36,53], which was created with the 21q4 primer and 21q-seq-rev2. The number of fusions per haploid genome (6 pg) is calculated based on the amount of input DNA in each PCR reaction.

Telomere length was assessed using telomeric restriction fragment analysis. Briefly, *Alu*I/*Mbo*I-digested genomic DNA was run on 0.8% TAE gels, before denaturation, neutralization, and transfer onto a Hybond membrane according to standard Southern blotting procedures. Telomeric DNA was detected using a TTAGGG repeat primed α-^{32}P (Klenow)-labeled Sty11 telomeric repeat probe[45].

WGS library preparation. Genomic DNA was extracted from cell pellets using a QIAGEN QIAamp DNA mini kit and sheared using a Covaris Ultrasonicator (E220) to ~300 bp fragments. DNA concentration was measured using Qbit 4.0 reagents (ThermoFisher), and 200 ng of fragmented DNA was used for library preparation. End repair and A-tailing was carried out with NEBNext End repair reaction enzyme mix and buffer (E7442), and KAPA dual-indexed adapters (Roche) were ligated using the T4 DNA ligase kit from NEB (M0202). Post-ligation size selection was performed with AMPure XP beads (Beckman Coulter) before washing two times in 80% ethanol. Libraries were amplified using KAPA HiFi HotStart ready mix (Roche) and P5 and P7 primers (IDT). PCR program was as follows: 98 °C for 45 s, five cycles of 98 °C for 15 s, 60 °C for 30 s, 72 °C for 30 s, and a final extension at 72 °C for 5 min. A further size selection and washing step was carried out after library amplification, and library quality was confirmed on Bioanalyzer chips (Agilent) and using a KAPA Library Quantification kit (Roche). Libraries were pooled and submitted for sequencing on NovaSeq 6000 at the New York Genome Center.

WGS basic data processing. Reads were aligned to GRCh37/hg19 using the Burroughs-Wheeler aligner (bwa mem v0.7.8, http://bio-bwa.sourceforge.net/)[54]. Best practices for post-alignment data processing were followed through use of Picard (https://broadinstitute.github.io/picard/) tools to mark duplicates, the GATK (v.2.7.4) (https://software.broadinstitute.org/gatk/) IndelRealigner module, and GATK base quality recalibration.

Variant rearrangement junctions were identified using SvABA[30] (https://github.com/walaj/SvABA) and GRIDSS[31] (https://github.com/PapenfussLab/gridss) with standard settings. For MRC5 samples, the somatic variant setting of each tool was used, with the ancestral MRC5 line as the matched normal. For post-crisis SV40 transformed cell lines the respective pre-crisis clone was used as the matched normal. 1-kbp binned read depth was computed and corrected for GC and mappability using fragCounter (https://github.com/mskilab/fragcounter). Systematic read-depth bias was subsequently removed using dryclean (https://github.com/mskilab/dryclean)[55].

Low-pass WGS clustering. Genome-wide binned read depth was aggregated across 118 low-pass WGS clones across 10-kbp bins by taking the median of 1-kbp binned normalized read depth from dryclean (see above). To minimize read-depth noise in unmappable regions, recurrent (>10% of the cohort) low-quality coverage regions (defined by Hadi et al.[3]) were combined with regions bearing consistently high variance in our high-pass sequencing dataset (standard deviation >0.3 for bin value over the mean in 100-kbp windows). Hierarchical clustering was then applied on the genome-wide Euclidean distance of bins, with "method = ward.D2" option. Six clusters were identified following dendrogram inspection.

Junction balance analysis. Preliminary junction-balanced genome graphs were generated for MRC5 and SV40T cell lines from binned read depth and junction calls (see above) using JaBbA (https://github.com/mskilab/JaBbA)[3]. Briefly, 1-kbp binned read-depth output from dryclean was collapsed to 5 kbp and JaBbA was run with slack penalty 500 for MRC5 clones and 100 for SV40T cell lines. gGnome (https://github.com/mskilab/gGnome) was used to identify complex structural variant patterns. Genome graphs and corresponding genomic data (e.g., binned coverage, allelic bin counts) were visualized using gTrack (https://github.com/mskilab/gTrack).

Joint inference of junction balance in MRC5. To chart structural variant evolution across sub-clades of MRC5 clones, a procedure was developed to jointly infer junction-balanced genome graphs in a lineage (e.g., BFB lineage in Fig. 5c). This co-calling algorithm augmented the existing JaBbA model, described in detail by Hadi et al.[3], enabling the application to a compendium of genome graphs by minimizing the total number of unique loose ends assigned a nonzero copy number across the graph compendium.

To describe this algorithm, we extend the notation introduced in Hadi et al.[3]. Formally, we define a collection $\{G^i\}_{i \in 1,...,n}$ of identical genome graphs across n clones, each a replica of a "prototype" genome graph G^0. The mapping p maps each vertex $v \in V(G^i)$ and edge $e \in E(G^i)$, $i \in 1,...,n$ to its corresponding vertex $p(v) \in V(G^0)$ and edge $p(e) \in E(G^0)$ in the prototype graph. We then jointly infer unique copy number assignments κ^i to the vertices and edges of each genome graph G^i by solving the mixed-integer program:

$$minimize_{\kappa^i:V_I(G^i) \cup E(G^i) \to \mathbb{N}, \, i \in 1,...,n} \, \lambda \mathcal{R}\left(\{G^i\}_{i \in 0,...,n}, \{\kappa^i\}_{i \in 1,...,n}, p\right) + \sum_{i \in 1,...,n} \mathcal{V}\left(G^i, \kappa^i, x^i, J^i\right) \quad (1)$$

$$subject \; to:$$

$$\kappa^i(v) = \kappa^i(\bar{v}), \quad \forall_{v \in V_I(G^i), \, i \in 1,...,n}$$

$$\kappa^i(e) = \kappa^i(\bar{e}), \quad \forall_{e \in E(G^i), \, i \in 1,...,n}$$

$$\kappa^i(v) = \sum_{e \in E^-(v,G^i)} \kappa^i(e) = \sum_{e \in E^+(v,G^i)} \kappa^i(e)$$

$$\kappa^i(e) \leq u^i(e), \quad e \in E(G^i), \quad i \in 1,...,n$$

where x^i and J^i represent the binned read-depth data and bin-node mappings for clone i and $\mathcal{V}(G^i, \kappa^i, x^i, J^i)$ is the read-depth residual for genome graph i, analogous to Hadi et al.[3]. An additional term in this new joint formulation is $u^i:E(G^i) \to \{0,\infty\}$, which is a data-derived mapping that constrains the upper bound of each edge $e \in E(G^i)$, e.g., on the basis of whether that junction has read support in clone i. In addition, a joint complexity penalty $\mathcal{R}$ couples the collection $\{\kappa^i\}_{i \in 1,...,n}$ of copy number mappings across the collection of graphs $\{G^i\}_{i \in 1,...,n}$ to each other by jointly penalizing loose ends at all vertices that map to the same prototype graph vertex $v \in V(G^0)$. Formally,

$$\mathcal{R}\left(\{G^i\}_{i \in 0,...,n}, \{\kappa^i\}_{i \in 1,...,n}, p\right) = \sum_{e \in E_L(G_0)} \left[\sum_{\{\hat{e} \in E_L(G^i) \, | \, p(\hat{e}) = e, \, i \in 1,...,n\}} \kappa^i(\hat{e}) \right] \quad (2)$$

in Hadi et al.[3], the hyperparameter λ in Eq. (1) controls the relative contribution of the read-depth residual and complexity penalty to the objective function. It is important to note that while each of the graphs G^i have an identical structure, the

constraints imposed by the upper bounds u^i and bin profiles x^i couple each graph to its junction and read-depth data, and hence lead to a unique fit κ^i on the basis of this data. The ℓ_0 penalty (defined using the Iversion bracket $[\![\ ]\!]$ operator) in Eq. (2) couples the solutions κ^i by adding an exponential prior on the number of unique loose ends across the entire graph compendium, where uniqueness is defined by the mapping p to the prototype graph G^0.

This joint mixed-integer programming model in Eq. (1) is implemented in the "balance" function of gGnome. The model was applied to a collection of genome graphs representing the structure of chromosome 12 across 13 clones. The prototype graph for this genome graph collection was built from the disjoint union of intervals of the 13 preliminary graphs (via the GenomicRanges "disjoin" function) and the union of junction calls fit across those graphs (via gGnome "merge.Junction" function). Each graph was associated using the read-depth data and bin-to-node mappings as per Hadi et al.[3]. The mapping u^i for each reference edge was set to ∞ while variant edges were assigned ∞ on the basis of bwa mem realignments of read pairs in each clone.bam file to the corresponding junction contig via rSeqLib (https://github.com/mskilab/rSeqLib)[56], otherwise they were assigned 0.

Equation (1) was then solved using the IBM CPLEX (v12.6.2) MIQP optimizer within the gGnome package after setting the hyperparameter λ to 100. This value was chosen after a parameter sweep observing for the visual concordance of genome graphs, loose ends, and read-depth profiles in the region.

Joint reconstruction of allelic evolution in MRC5. Evolving 12p alleles were jointly reconstructed across 13 MRC5 clones through the analysis of junction-balanced genome graphs (G^i, κ^i) (see "Joint inference of junction balance in MRC5" section above). The procedure for joint allelic phasing described in Hadi et al.[3] was extended to identify the most parsimonious collection of linear and/or cyclic walks and associated walk copy numbers that summed to the vertex and edge copy numbers in the compendium (G^i, κ^i).

Formally, the subgraph of vertices and edges with a nonzero copy number in each (G^i, κ^i) were exhaustively traversed to derive all minimal paths and cycles H^i, where for each walk $h \in H^i$ maps to subsets $V(h) \subseteq V(G^i)$ and $E(h) \subseteq V(G^i)$ of vertices and edges in the graph G^i. The nodes and vertices of these walks were then projected via the mapping p to define a unique set of walks H^0 in the prototype graph G^0. We extend our notation p (see the previous section) so that for a walk $h \in H^i$ the mapping $p(h) \in H^0$ denotes the walk formed by projecting the vertices and edges of h via p to H^0. With these definitions, the single graph haplotype inference defined in Hadi et al.[3] was extended to a joint inference by solving the following mixed-integer linear program to assign a copy number $\phi^i(h) \in \mathbb{N}$ to each walk $h \in H^i$.

$$minimize_{\phi^i : H^i \to \mathbb{N}, \ i \in 1,...,n} \sum_{h \in H^0} \Big[\sum_{\hat{h} \mid \hat{h} \in H^i, \ p(\hat{h})=h, \ i \in 1,...,n} \phi^i\big(\hat{h}\big) \Big] \quad (3)$$

subject to:

$$\kappa^i(v) = \sum_{h \in H^i} \phi^i(h)\delta(v,h), \quad \forall_{v \in V_I(G^i), \ i \in i,...,n}$$

$$\kappa^i(e) = \sum_{h \in H^i} \phi^i(h)\delta(e,h), \quad \forall_{e \in E(G^i), \ i \in i,...,n}$$

where the function $\delta(v,h)$ and $\delta(e,h)$ is 1 if vertex v and edge e belong to walk h and 0 otherwise. The Iverson bracket ($[\![\ ]\!]$) operator in the objective function Eq. (3) minimizes the total number of unique walks used across the compendium, hence identifying a jointly parsimonious assignment of copy number to walks across the compendium of graphs. Equation (3) was solved using the IBM CPLEX (v12.6.2) MIQP optimizer within the gGnome package. Variant cycles and paths from the resulting solution were manually combined to yield a set of consistent linear paths, i.e., somatic haplotypes, to yield allelic reconstructions in Fig. 5c.

Loose-end classification. Each loose end in each MRC5 genome graph was analyzed to identify a clone-specific (i.e., absent in the ancestral MRC5 line) origin for the mates of high mapping quality (MAPQ = 60) reads mapping to the location and strand of the loose end. These mates were assessed for neo-telomeric sequences by counting instances of 11 permutations of a 12-bp telomere repeat motif (TTAGGGTTAGGG) (using the R/Bioconductor Biostrings package) in the mates. The mates were also assembled into contigs using fermi[57] aligned using bwa mem[58] via the RSeqLib R package[56] to hg38 (which contains a more highly resolved centromere build) to characterize novel repeat (e.g., centromere) fusions (GATK Human reference genome, hg38, data bundle including Homo_-sapiens_assembly38.fasta, gs://gcp-public-data--broad-reference). The loose-end loci were also assessed through overlap with the hg19 repeatMasker database (human_g1k_v37_decoy.repeatmasker) for the presence of reference annotated repeats that might explain the absence of a mappable junction explaining the copy number change.

SNV phylogeny. To compute an SNV phylogeny across MRC clones, we first identified SNV that were acquired in MRC5 clones relative to the ancestral MRC5 line using Strelka2[59] (https://github.com/Illumina/strelka) under paired (i.e., tumor/normal) mode with the clone as the "tumor" and the MRC ancestral line as the "normal" sample and default parameters and GATK hg19 resource bundle (Genome Analysis Toolkit GATK Resource Bundle for hg19; gs://gatk-legacy-bundles). Acquired SNVs were first filtered according to the Strelka2 PASS filter as well as additional filters (MQ = 60, SomaticEVS>12, total ALT count >4) yielding 27,220 total unique variants across the 13 MRC5 clones. Reference and variant allelic read counts were assessed at each SNV site (via the R/Bioconductor Rsamtools package, version 3.6.1, http://www.r-project.org/) across all 13 clones. We then further required a >0.5 posterior probability of a variant being present in a sample, by assuming the Binomial likelihood of variant read count and using the aggregated allele frequency in all samples as the prior, resulting in the final 14,970 unique mutations. The binary matrix of clones by SNV loci was then used to derive a neighbor-joining phylogenetic tree using the R/Bioconductor package ape. Following tree construction, we associated each SNV with its most likely phylogenetic tree branch by comparing the binary incidence vector associated with each SNV with the binary incidence vector associated with each tree branch, and finding the closest branch using Jaccard distance, only linking SNV to branches when the SNV was within <0.1 Jaccard distance of the closest branch, thus producing the groupings of SNVs in Fig. 5a.

Parental SNP allelic phasing and imbalance. Germline heterozygous sites in the parental MRC5 line were identified by computing allelic counts at HapMap sites (GATK human reference genome, hg19 data bundle, hapmap_3.3.b37.vcf) and identifying loci with variant allele fraction >0.3 and <0.7. Y11, a clone with loss of a single allele at 12p, was chosen to phase parental SNPs on 12p. At each locus, the allele (reference or alternate) with a 0 read count was assigned to the "L" (lost) haplotype and the other allele was assigned to the "R" (retained) haplotype. (All heterozygous SNP loci in the region contained exactly one allele with a 0 read count). L and R allelic counts were then computed at these sites across all 13 high-pass WGS and 131 low-pass WGS samples. These counts were divided by the genome-wide mean of heterozygous SNP allele counts (in these 100% pure and nearly diploid samples) to derive the absolute allelic copy number[60].

SNV clustering. Inter-SNV distances were computed for all pairs of reference adjacent acquired SNVs associated with each MRC5 clone and visualized as rainfall plots. Runs of two or more SNVs with inter-SNV distances <2 kbp were nominated as clusters. Two distinct SNV clusters were identified on chromosome 12p across the 13 clones.

Statistical analysis. Statistical analysis for in vitro experiments was carried out using Prism software (GraphPad Software). All relevant statistical experimental details (n numbers, SD) are provided in the figure legends. Statistically significant associations between binary variables were determined using a two-tailed Fisher's exact test. Significance was assessed on the basis of Bonferroni-corrected P values <0.05. Effect sizes (odds ratios) are reported alongside 95% confidence intervals for each test. Details of all other quantitative analyses (e.g., read-depth processing, clone clustering, genome graph inference, allelic reconstructions, phylogenetic reconstruction, SNV clustering, parental SNP phasing) are described above.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Whole-genome sequencing data have been deposited to the sequence read archive as aligned.bam files (https://www.ncbi.nlm.nih.gov/sra) under BioProject accession PRJNA693405. Reasonable requests for any other data pertaining to this study should be directed to and will be fulfilled by the corresponding authors. Source data are provided with this paper.

Code availability

Custom software packages referenced in this study are available at https://github.com/mskilab (JaBbA build 7926cc7, gGnome build 7f5bf56, dryclean build 6d2bced, fragCounter build 575af99, dryclean build 6d2bced, rSeqLib build 23fbaf0, skitools build 61187fa, gUtils build 449ab2a, gTrack build 947c35c). Analysis code to generate the figures in the paper is available on request.

References

1. Drier, Y. et al. Somatic rearrangements across cancer reveal classes of samples with distinct patterns of DNA breakage and rearrangement-induced hypermutability. *Genome Res.* **23**, 228–235 (2013).
2. Yang, L. et al. Diverse mechanisms of somatic structural variations in human cancer genomes. *Cell* **153**, 919–929 (2013).
3. Hadi, K. et al. Distinct classes of complex structural variation uncovered across thousands of cancer genome graphs. *Cell* **183**, 197–210.e32 (2020).
4. The ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium. Pan-cancer analysis of whole genomes. *Nature* **578**, 82–93 (2020).
5. Li, Y. et al. Patterns of somatic structural variation in human cancer genomes. *Nature* **578**, 112–121 (2020).
6. Alexandrov, L. B. et al. Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
7. Nik-Zainal, S. et al. Mutational processes molding the genomes of 21 breast cancers. *Cell* **149**, 979–993 (2012).
8. Menghi, F. et al. The tandem duplicator phenotype is a prevalent genome-wide cancer configuration driven by distinct gene mutations. *Cancer Cell* **34**, 197–210.e5 (2018).
9. Zhang, C. Z. et al. Chromothripsis from DNA damage in micronuclei. *Nature* **522**, 179–184 (2015).
10. Willis, N. A. et al. Mechanism of tandem duplication formation in BRCA1-mutant cells. *Nature* **551**, 590–595 (2017).
11. Maciejowski, J., Li, Y., Bosco, N., Campbell, P. J. & de Lange, T. Chromothripsis and kataegis induced by telomere crisis. *Cell* **163**, 1641–1654 (2015).
12. Ly, P. et al. Chromosome segregation errors generate a diverse spectrum of simple and complex genomic rearrangements. *Nat. Genet.* **51**, 705–715 (2019).
13. Ghezraoui, H. et al. Chromosomal translocations in human cells are generated by canonical nonhomologous end-joining. *Mol. Cell* **55**, 829–842 (2014).
14. Umbreit, N. T. et al. Mechanisms generating cancer genome complexity from a single cell division error. *Science* **368**, eaba0712 (2020).
15. Shay, J. W. & Wright, W. E. Senescence and immortalization: role of telomeres and telomerase. *Carcinogenesis* **26**, 867–874 (2005).
16. Artandi, S. E. & DePinho, R. A. A critical role for telomeres in suppressing and facilitating carcinogenesis. *Curr. Opin. Genet Dev.* **10**, 39–46 (2000).
17. Maciejowski, J. & de Lange, T. Telomeres in cancer: tumour suppression and genome instability. *Nat. Rev. Mol. Cell Biol.* **18**, 175 (2017).
18. McClintock, B. The behavior in successive nuclear divisions of a chromosome broken at meiosis. *Proc. Natl Acad. Sci. USA* **25**, 405–416 (1939).
19. Gisselsson, D. et al. Telomere dysfunction triggers extensive DNA fragmentation and evolution of complex chromosome abnormalities in human malignant tumors. *Proc. Natl Acad. Sci. USA* **98**, 12683–12688 (2001).
20. Li, Y. et al. Constitutional and somatic rearrangement of chromosome 21 in acute lymphoblastic leukaemia. *Nature* **508**, 98–102 (2014).
21. Artandi, S. E. et al. Telomere dysfunction promotes non-reciprocal translocations and epithelial cancers in mice. *Nature* **406**, 641–645 (2000).
22. O'Hagan, R. C. et al. Telomere dysfunction provokes regional amplification and deletion in cancer genomes. *Cancer Cell* **2**, 149–155 (2002).
23. Ding, Z. et al. Telomerase reactivation following telomere dysfunction yields murine prostate tumors with bone metastases. *Cell* **148**, 896–907 (2012).
24. Liddiard, K. et al. Sister chromatid, but not NHEJ-mediated inter-chromosomal telomere fusions, occur independently of DNA ligases 3 and 4. *Genome Res.* **26**, 588–600 (2016).
25. Maciejowski, J. et al. APOBEC3-dependent kataegis and TREX1-driven chromothripsis during telomere crisis. *Nat. Genet.* **52**, 884–890 (2020).
26. Cleal, K., Jones, R. E., Grimstead, J. W., Hendrickson, E. A. & Baird, D. M. Chromothripsis during telomere crisis is independent of NHEJ, and consistent with a replicative origin. *Genome Res.* **29**, 737–749 (2019).
27. Bryan, T. M., Englezou, A., Gupta, J., Bacchetti, S. & Reddel, R. R. Telomere elongation in immortal human cells without detectable telomerase activity. *EMBO J.* **14**, 4240–4248 (1995).
28. Shay, J. W. & Wright, W. E. Quantitation of the frequency of immortalization of normal human diploid fibroblasts by SV40 large T-antigen. *Exp. Cell Res.* **184**, 109–118 (1989).
29. Counter, C. M. et al. Telomere shortening associated with chromosome instability is arrested in immortal cells which express telomerase activity. *EMBO J.* **11**, 1921–1929 (1992).
30. Wala, J. A. et al. SvABA: genome-wide detection of structural variants and indels by local assembly. *Genome Res.* **28**, 581–591 (2018).
31. Cameron, D. L. et al. GRIDSS: sensitive and specific genomic rearrangement detection using positional de Bruijn graph assembly. *Genome Res.* **27**, 2050–2060 (2017).
32. Chavez, A. et al. Highly efficient Cas9-mediated transcriptional programming. *Nat. Methods* **12**, 326–328 (2015).
33. Chiba, K. et al. Mutations in the promoter of the telomerase gene TERT contribute to tumorigenesis by a two-step mechanism. *Science* **357**, 1416–1420 (2017).
34. Baird, D. M., Rowson, J., Wynford-Thomas, D. & Kipling, D. Extensive allelic variation and ultrashort telomeres in senescent human cells. *Nat. Genet.* **33**, 203–207 (2003).
35. Capper, R. et al. The nature of telomere fusion and a definition of the critical telomere length in human cells. *Genes Dev.* **21**, 2495–2508 (2007).
36. Riethman, H. et al. Mapping and initial analysis of human subtelomeric sequence assemblies. *Genome Res.* **14**, 18–28 (2004).
37. Petljak, M. et al. Characterizing mutational signatures in human cancer cell lines reveals episodic APOBEC mutagenesis. *Cell* **176**, 1282–1294.e20 (2019).
38. Smogorzewska, A., Karlseder, J., Holtgreve-Grez, H., Jauch, A. & de Lange, T. DNA ligase IV-dependent NHEJ of deprotected mammalian telomeres in G1 and G2. *Curr. Biol.* **12**, 1635 (2002).
39. Letsolo, B. T., Rowson, J. & Baird, D. M. Fusion of short telomeres in human cells is characterized by extensive deletion and microhomology, and can result in complex rearrangements. *Nucleic Acids Res.* **38**, 1841–1852 (2010).
40. Shay, J. W., Pereira-Smith, O. M. & Wright, W. E. A role for both RB and p53 in the regulation of human cellular senescence. *Exp. Cell Res.* **196**, 33–39 (1991).
41. d'Adda di Fagagna, F. et al. A DNA damage checkpoint response in telomere-initiated senescence. *Nature* **426**, 194–198 (2003).
42. Brown, J. P., Wei, W. & Sedivy, J. M. Bypass of senescence after disruption of p21CIP1/WAF1 gene in normal diploid human fibroblasts. *Science* **277**, 831–834 (1997).
43. Jacobs, J. J. & de Lange, T. Significant role for p16INK4a in p53-independent telomere-directed senescence. *Curr. Biol.* **14**, 2302–2308 (2004).
44. Furugori, E. et al. Telomere shortening in gastric carcinoma with aging despite telomerase activation. *J. Cancer Res. Clin. Oncol.* **126**, 481–485 (2000).
45. de Lange, T. et al. Structure and variability of human chromosome ends. *Mol. Cell Biol.* **10**, 518–527 (1990).
46. Mehle, C., Ljungberg, B. & Roos, G. Telomere shortening in renal cell carcinoma. *Cancer Res.* **54**, 236–241 (1994).
47. Barthel, F. P. et al. Systematic analysis of telomere length and somatic alterations in 31 cancer types. *Nat. Genet.* **49**, 349–357 (2017).
48. Stimpson, K. M. et al. Telomere disruption results in non-random formation of de novo dicentric chromosomes involving acrocentric human chromosomes. *PLoS Genet.* **6**, e1001061 (2010).
49. Stewart, N. & Bacchetti, S. Expression of SV40 large T antigen, but not small t antigen, is required for the induction of chromosomal aberrations in transformed human cells. *Virology* **180**, 49–57 (1991).
50. Chicas, A. et al. Dissecting the unique role of the retinoblastoma tumor suppressor during cellular senescence. *Cancer Cell* **17**, 376–387 (2010).
51. Doksani, Y. & de Lange, T. Telomere-internal double-strand breaks are repaired by homologous recombination and PARP1/Lig3-dependent end-joining. *Cell Rep.* **17**, 1646–1656 (2016).
52. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
53. Brown, W. R. et al. Structure and polymorphism of human telomere-associated DNA. *Cell* **63**, 119–132 (1990).
54. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
55. Deshpande, A., Walradt, T., Hu, Y., Koren, A. & Imielinski, M. Robust foreground detection in somatic copy number data. Preprint at https://www.biorxiv.org/content/10.1101/847681v1 (2019).
56. Wala, J. & Beroukhim, R. SeqLib: a C++ API for rapid BAM manipulation, sequence alignment and sequence assembly. *Bioinformatics* **33**, 751–753 (2017).
57. Li, H. Exploring single-sample SNP and INDEL calling with whole-genome de novo assembly. *Bioinformatics* **28**, 1838–1844 (2012).
58. Li, H. Toward better understanding of artifacts in variant calling from high-coverage samples. *Bioinformatics* **30**, 2843–2851 (2014).
59. Kim, S. et al. Strelka2: fast and accurate calling of germline and somatic variants. *Nat. Methods* **15**, 591–594 (2018).
60. Carter, S. L. et al. Absolute quantification of somatic DNA alterations in human cancer. *Nat. Biotechnol.* **30**, 413–421 (2012).

Author contributions

Conceptualization: S.M.D., M.I., and T.d.L. Methodology: S.M.D., X.Y., and M.I. Software and formal analysis: X.Y., J.R., J.B., and M.I. Investigation: S.M.D., N.B., K.T., and H.T. Writing—original draft: S.M.D., M.I., and T.d.L. Writing—review and editing: S.M.D., X.Y., M.I., and T.d.L. Supervision and funding acquisition: T.d.L. and M.I.

Competing interests

T.d.L. is on the SAB of Calico Life Sciences, LLC. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-021-21933-7.

Correspondence and requests for materials should be addressed to T.d.L. or M.Ińs.

Peer review information *Nature Communications* thanks Ryan Layer and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Structural variant evolution after telomere crisis
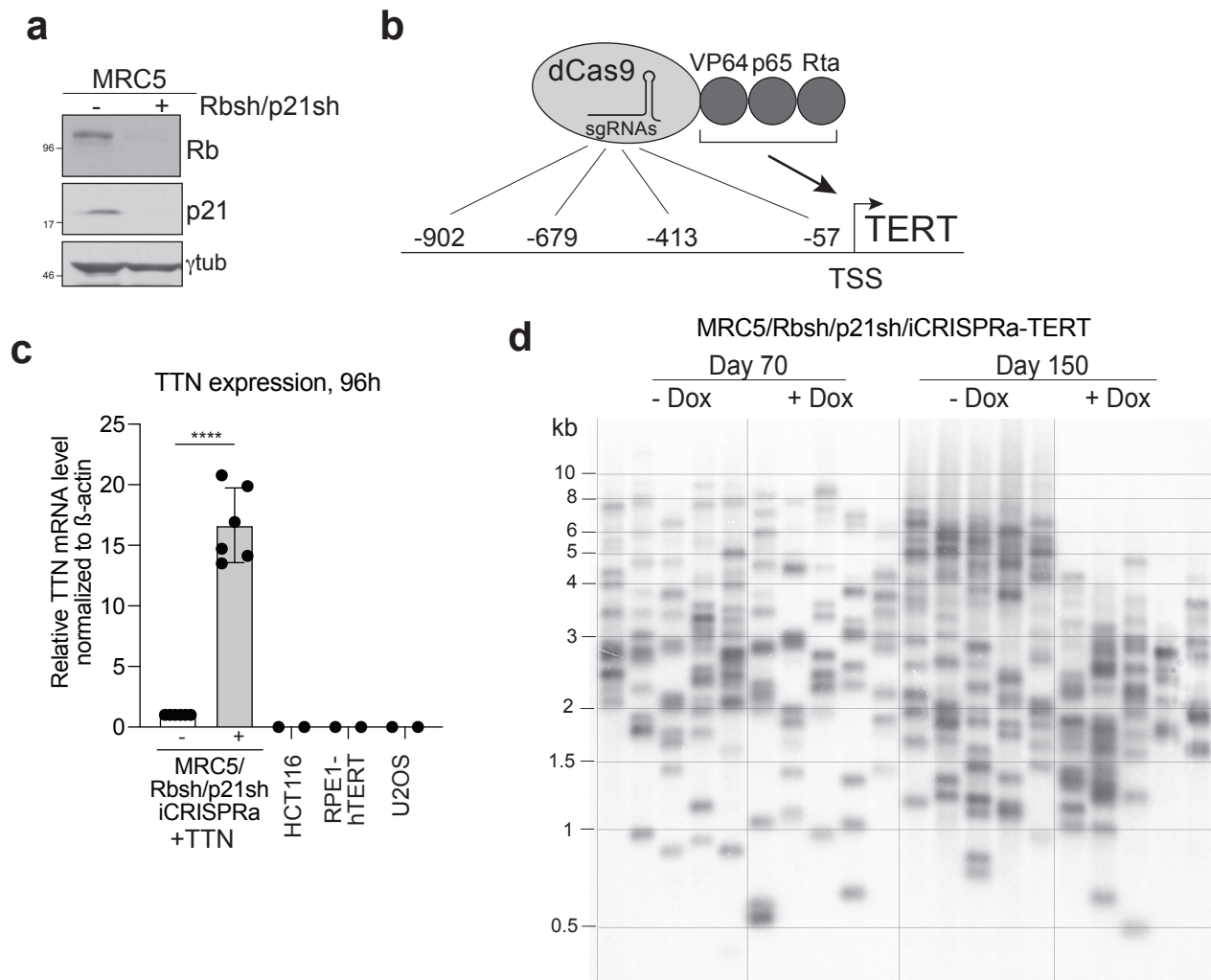
Supplementary Information
- Supplementary Figures
  - Supplementary Figure 1
  - Supplementary Figure 2
  - Supplementary Figure 3
  - Supplementary Figure 4
  - Supplementary Figure 5
  - Supplementary Figure 6
  - Supplementary Figure 7
- Supplementary Tables
  - Supplementary Table 1
  - Supplementary Table 2
  - Supplementary Table 3

Dewhurst, Yao *et al.* Supplementary Figure 1
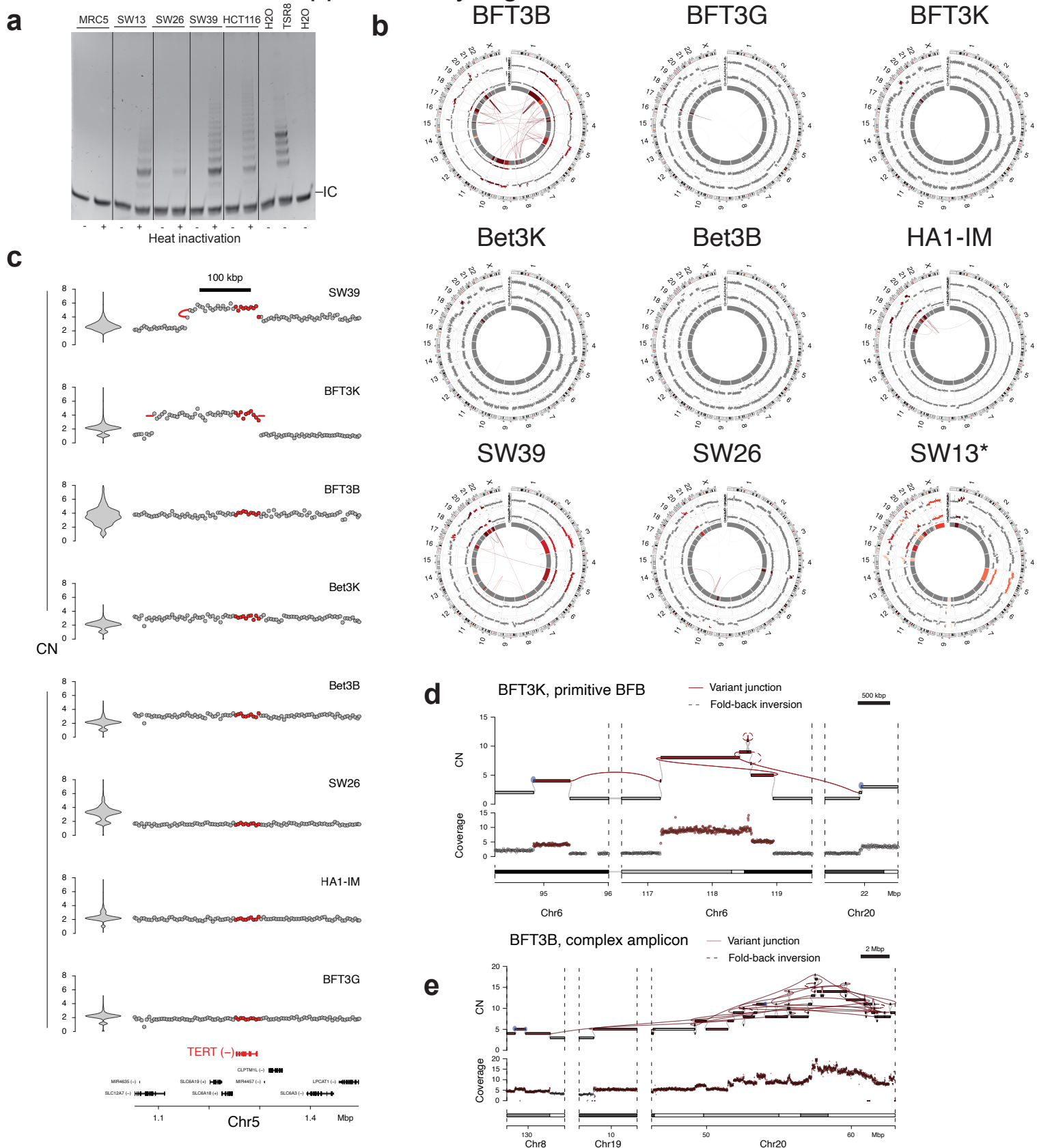
Supplementary Figure 1. Telomerase activity and complex gains after spontaneous telomere crisis resolution. Related to Figure 1.
a) TRAP assay showing telomerase activity in SV40 immortalized clones SW13, SW26 and SW39. MRC5 is included as a negative control, and HCT116 as a positive control. TSR8 is the positive control template. IC = internal control at 36 bp. Gel shown is representative, experiment was carried out at least twice. **b)** CIRCOS plots of all nine SV40 immortalized clones together with their respective pre-crisis clones. From innermost circle, variant junctions present only in post-crisis sample, colored by complex amplicon clusters or grey if not classified into any, JaBbA inferred copy number, purity-ploidy transformed binned coverage of the pre-crisis clone, of the post-crisis clone, and chromosome bands. SW13 is marked with an asterisk due to the presence of extensive pre-crisis structural rearrangements. **c)** Read depth and junction patterns at the TERT locus across the seven post-crisis clones. Each track shows binned purity- and ploidy-transformed read depth in units of CN, with variant (rearrangement) junctions and loose ends plotted as red arcs. The bottom track highlights the TERT gene among genes on chromosome 5p15. Violin plots to the left of tracks show the genome-wide distribution of read depth in units of CN, demonstrating that 6 of 7 clones have elevated CN at the TERT locus. **d-e)** Example clusters of complex gains from the cell lines shown in Figure 1A, each showing binned purity- and ploidy- transformed read-depth, with the top track showing the associated junction-balanced genome graph (see Methods3, with y-axis representing units of per cell copy number (CN) across bins and graph nodes (i.e. intervals). Grey and colored edges represent reference and variant junctions, respectively. Dashed colored edges represent fold-back inversion junctions. Blue edges represent loose ends (see Methods for further details). Bins and junctions are colored as per Figure 1A. In e, only an excerpt of the entire complex amplicon cluster proximal to chromosome 20q peak is represented.

a

MRC5
- | + Rbsh/p21sh

96 — Rb

17 — p21

46 — γtub

b

dCas9 — VP64 p65 Rta

sgRNAs

-902   -679   -413   -57   TERT

TSS

c

TTN expression, 96h

****

Relative TTN mRNA level normalized to ß-actin

25
20
15
10
5
0

MRC5/
Rbsh/p21sh
iCRISPRa
+TTN

HCT116

RPE1-
hTERT

U2OS

d

MRC5/Rbsh/p21sh/iCRISPRa-TERT

Day 70 | Day 150
- Dox | + Dox | - Dox | + Dox

kb
10
8
6
5
4
3
2
1.5
1
0.5

Supplementary Figure 2. A system for controlled escape from telomere crisis. Related to Figure 2.
a) Immunoblot for Rb and p21 in MRC5 cells and MRC5/Rbsh/p21sh cells. A representative blot is shown, experiment was carried out at least twice. **b)** Schematic diagram illustrating the iCRISPRa-TERT system. The positions of the TERT activating gRNAs relative to the transcriptional start site (TSS) of the TERT gene are shown. **c)** qPCR for the TTN control gene, activated with a combination of four sgRNAs in MRC5/Rbsh/p21sh/iCRISPRa cells (with or without dox, 96 hrs) n=6. HCT116, RPE1-hTERT and U2OS cells are included as negative controls (n=2). Data for dox- and dox+ samples were compared with a two-tailed student's t-test, p<0.0001. Error bars indicate mean ±SD. **d)** STELA of the XpYp telomere in MRC5/Rbsh/p21sh/iCRISPRa-TERT cells after 70 days or 150 days of continuous culture with or without doxycycline. This is a biological replicate of the STELA in Figure 2E.

MRC5/Rbsh/p21sh iCRISPRa-TERT
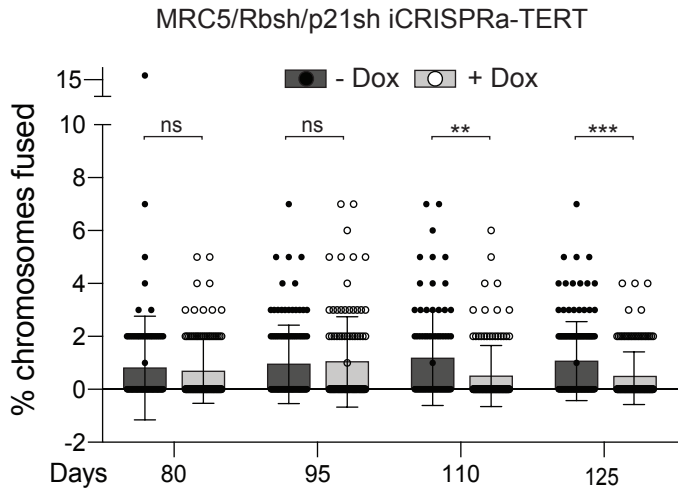
**Supplementary Figure 3. Telomere fusions after controlled escape from crisis. Related to Figure 3.**
Quantification of percentage of fused chromosomes after the indicated days of continuous culture for MRC5/Rbsh/p21sh/iCRISPRa-TERT cells with and without doxycycline. Error bars represent means ±SDs from three independent biological replicates, bar at mean, each dot represents one metaphase spread, n=84-127. P values derived from a two-sided student's t-test. ns, not significant; **, p=0.0012; ***; p=0.0004.

Dewhurst, Yao *et al*. Supplementary Figure 4
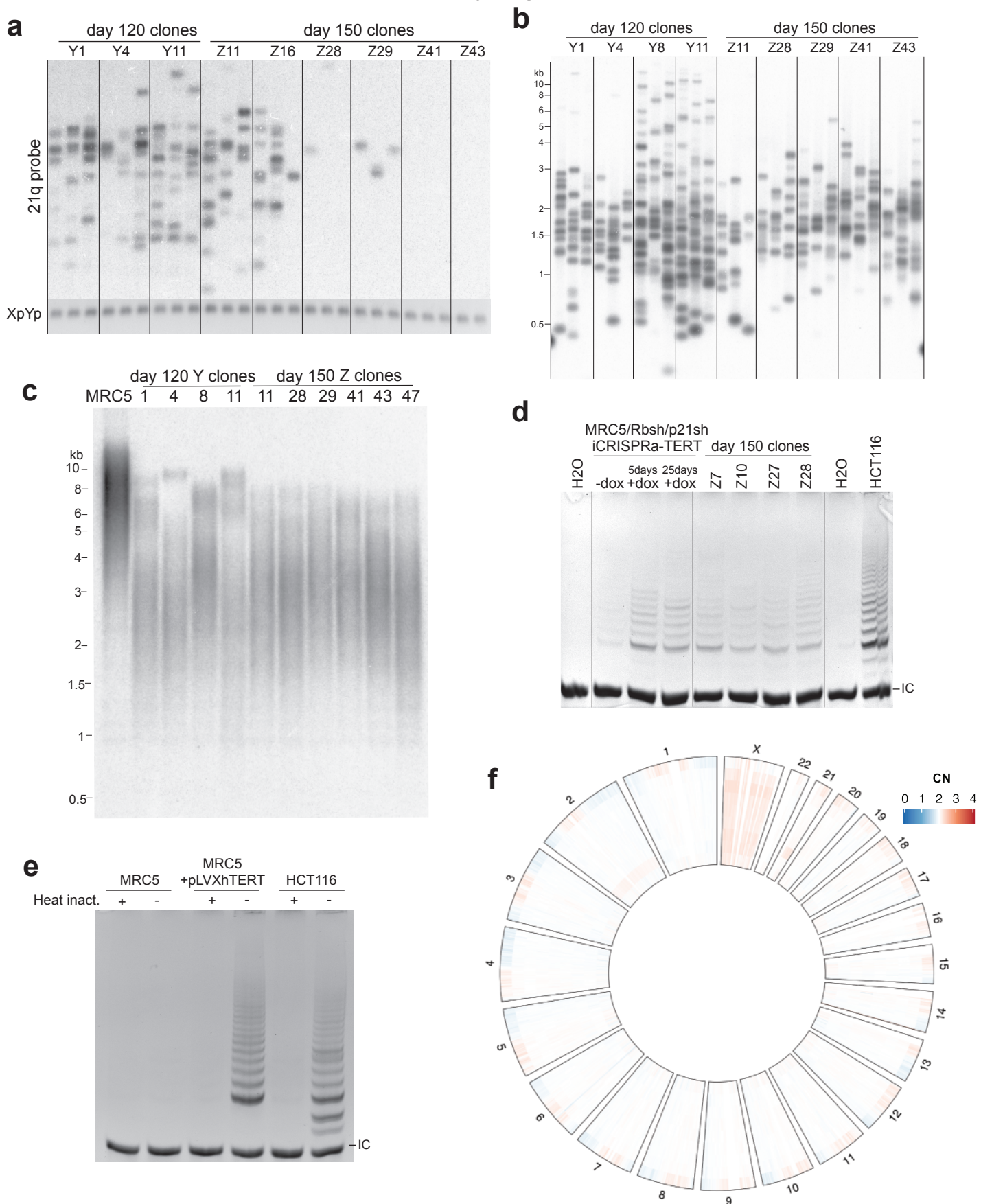
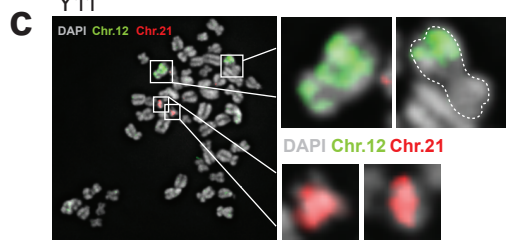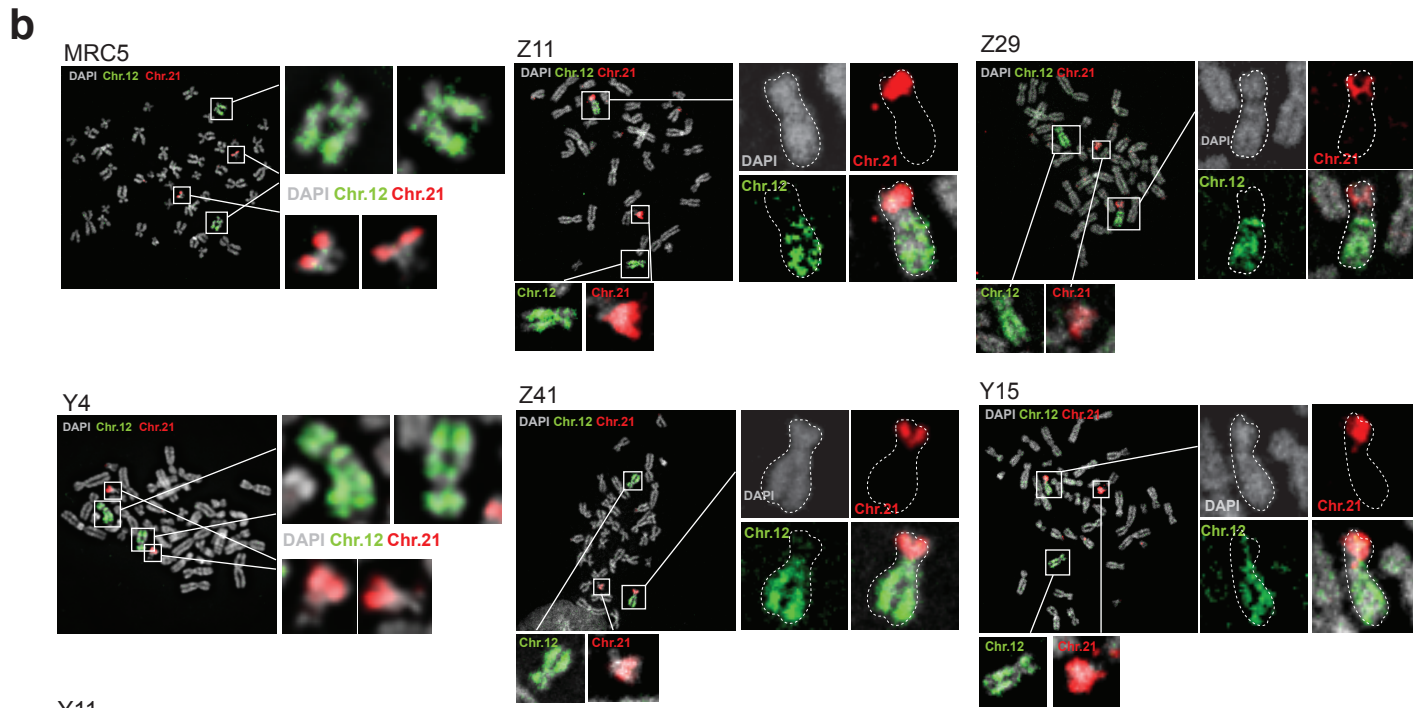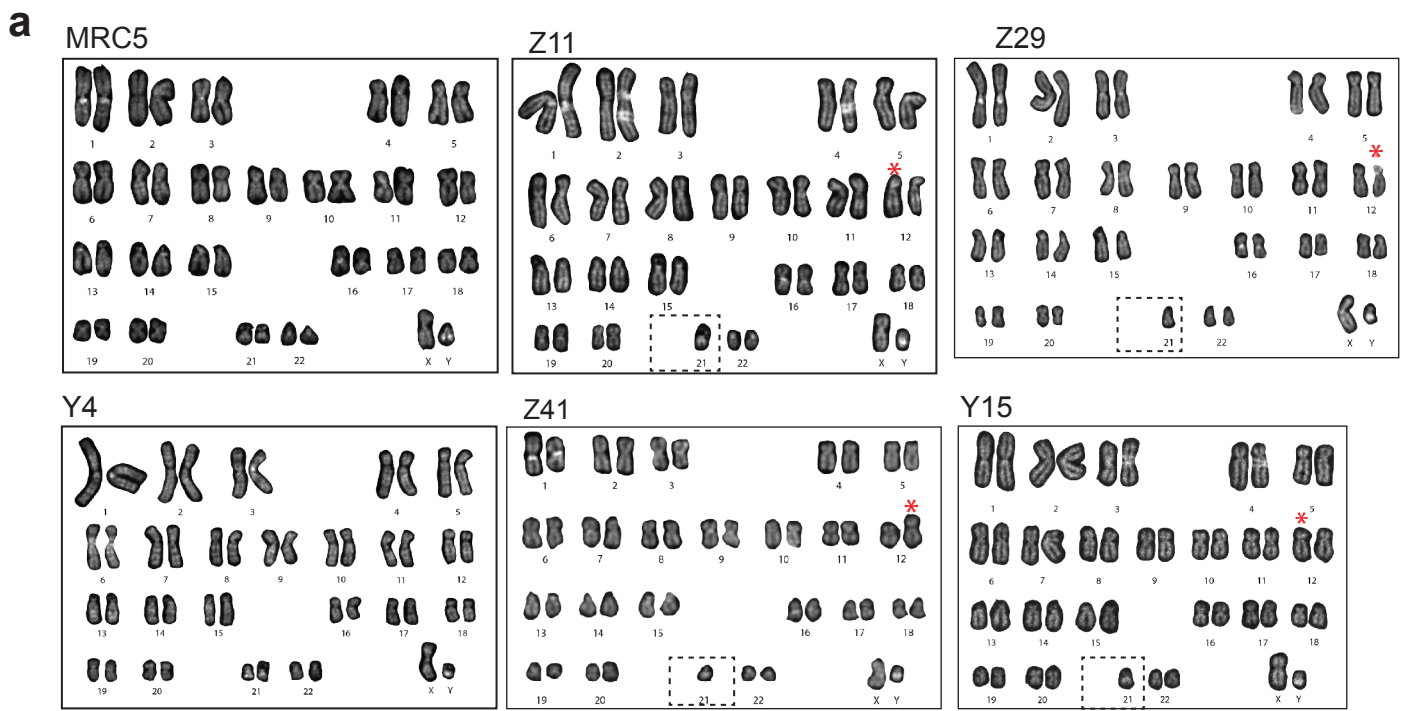**Supplementary Figure 4. Post-crisis telomere dynamics. Related to Figure 4.**
a) Products of telomere fusion PCR on a panel of post-crisis clones (see Figure 4A). Telomere fusions are detected by hybridization to the 21q probe. The control XpYp band is detected with Ethidium bromide staining. Supe blot is representative of at least two experiments. **b)** STELA products from a panel of post-crisis clones from both time points. The blot is a representative of at least two experiments. **c)** Telomeric restriction fragment blot of DNA from MRC5 cells and a panel of post-crisis clones from both time points. The blot is a representative of at least two experiments. **d)** TRAP assay showing telomerase activity in MRC5/Rbsh/p21sh/iCRISPRa-TERT treated with doxycycline for the indicated number of days and a selection of post-crisis clones from the day 150 timepoint. HCT116 is included as a positive control. IC= internal control, 36 bp. The gel is a representative of at least two experiments. **e)** TRAP assay showing robust telomerase activity in MRC5 cells infected with retroviral pLVX-hTERT. IC= internal control, 36bp The gel is a representative of at least two experiments. **f)** Circular heatmap showing genome-wide binned purity- and ploidy-transformed read depth (in units of CN across 8 low pass WGS-profiled control CT clones). Heatmap rows correspond to concentric rings in the heatmap.
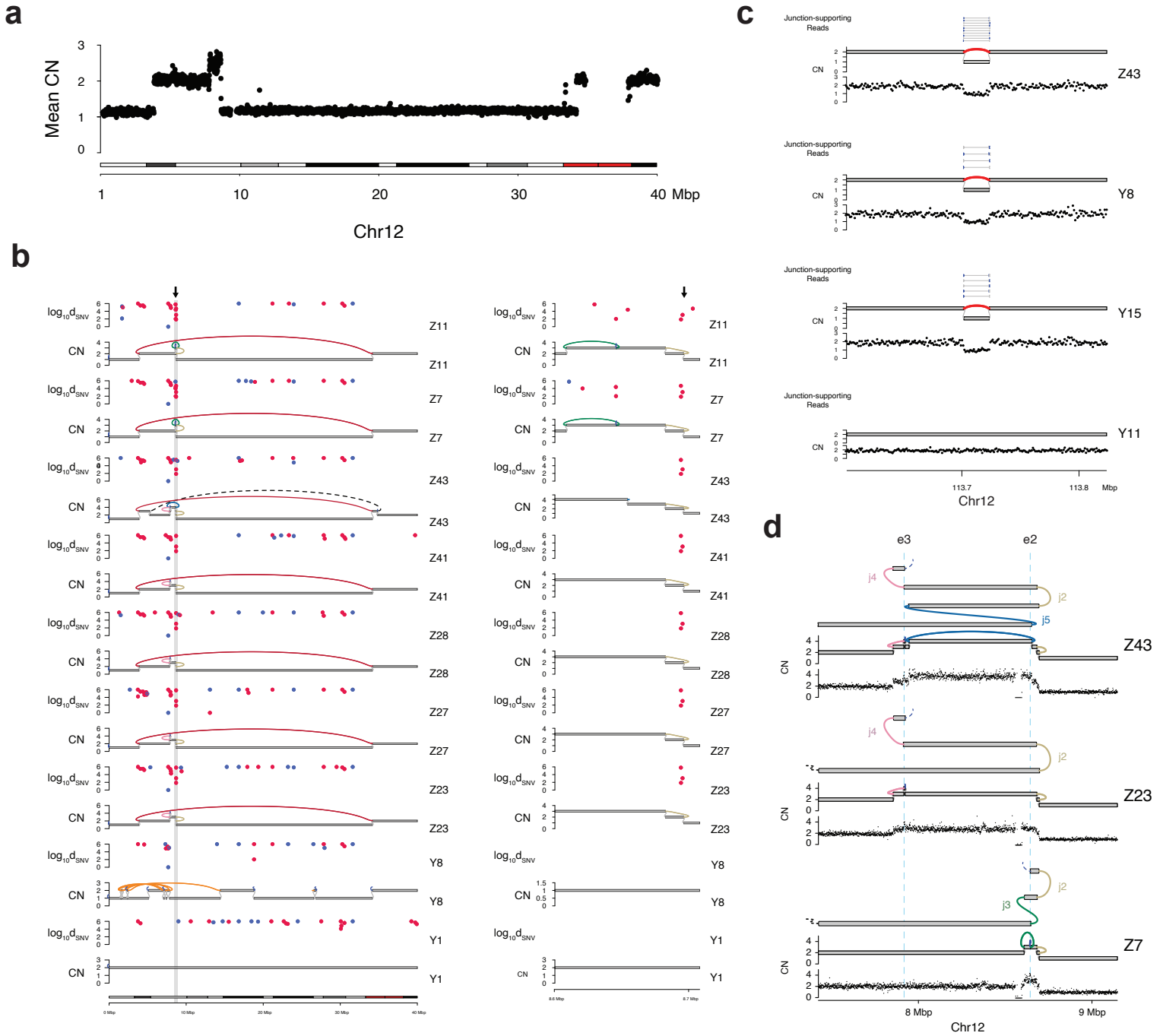
Supplementary Figure 5. WGS analysis showing genome alterations post crisis clones. Related to Figure 4 and 5.
a) Consensus copy number profile of 47 complex low pass WGS clones targeting chromosome 12p. **b)** Rainfall plot demonstrating SNV patterns as a function of GC vs AT reference nucleotide context, with y-axis showing the logarithm of the inter SNV distance. Highlighted region represents a GC strand coordinated cluster that is found across all four clones in the BFB cluster which harbour junction j2. **c)** Junction supporting read pairs and corresponding drop in coverage at the small deletion on chromosome 12q in representative BFB-like (Z43), chromothripsis-like (Y8), close-relative Y15, and lack of such evidence in distant Y11. **d)** Detailed JaBbA models and linear allele reconstruction at two distinct loose ends, e2 and e3.

a

b

c

Supplementary Figure 6. Karyotype evolution in post-crisis clones. Related to Figure 6.
a) DAPI banded karyotypes from the MRC5 parental cell line and a selection of post-crisis clones. Marker chromosomes (derivative 12) are indicated with a red star. Dashed box indicates the absence of an intact copy of chromosome 21. **b)** Representative images of metaphase spreads from post-crisis clones hybridized with both 12 (green) or 21 (red) chromosome paints. DNA stained with DAPI (grey). **c)** Post-crisis clone Y11 hybridized with both 12 (green) or 21 (red) chromosome paints. DNA stained with DAPI (grey).

a

b

c
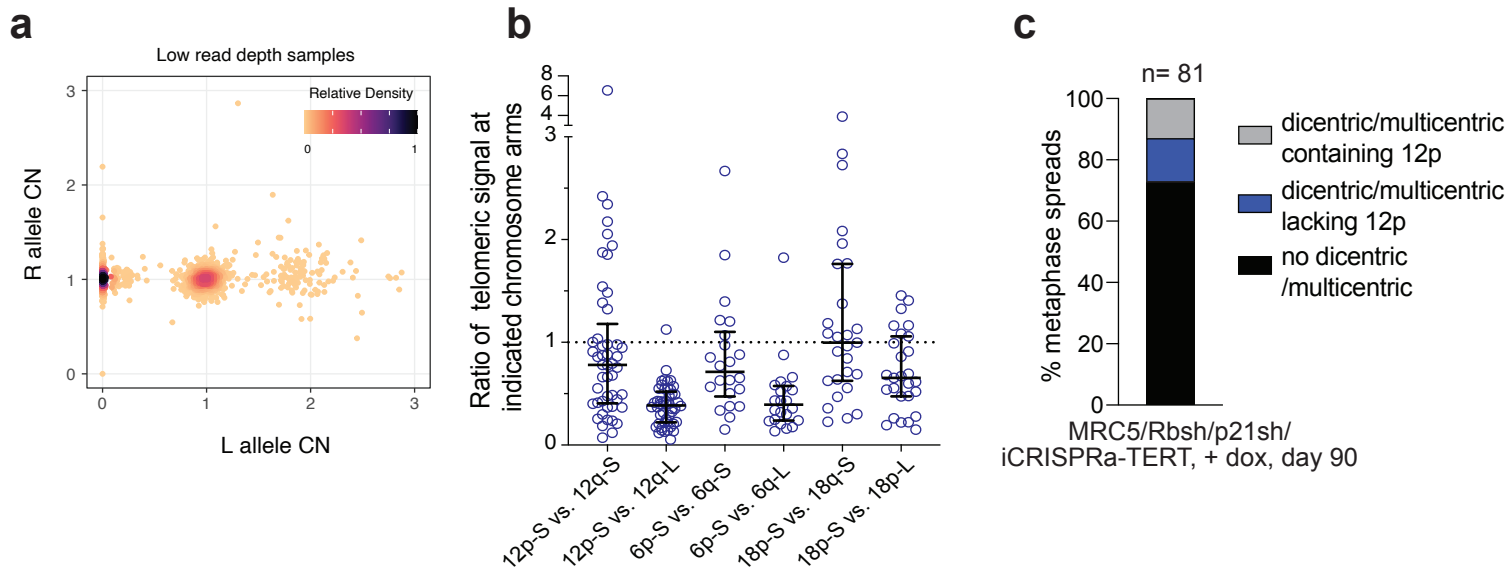
Supplementary Figure 7. Chromosome 12p allelic imbalance. Related to Figure 7.
a) Scatter plot showing purity- and ploidy-transformed L and R haplotype specific allelic read depth across 12p segments in low pass WGS-profiled post-crisis clones. **b)** Analysis showing that 12p, 6p and 18p arms contain the shortest telomere of those chromosomes. Comparison of the shortest (S) telomere end from each arm (p or q) to each other allele of that chromosome (i.e. q-short allele vs. p-short allele) across chromosome 12, 6 and 18 alleles. Ratio of telomeric intensity from TelG hybridization, as in Figure 7D (S= short allele, L= long allele, based on TelG intensity). Error bars represent median and interquartile ranges. Chromosome 12: n= 49 cells, chr.18: n=27, chr.6: n= 22. **c)** Fraction of dicentric or multicentric chromosomes containing 12p in MRC5/Rbsh/p21sh/iCRISPRa-TERT cells in crisis (with doxycycline at day 90) among all dicentric and multicentric chromosomes. Based on analysis of images as in Figure 7E. n=81 metaphases scored from two independent experiments.

Supplementary Table 1: SV40T immortalized pre-crisis and post-crisis telomerase positive cell lines

Name	Cell Line/Tissue Source	Telomerase status	Reference	Kind gift of
HA1 p15	Human Embryonic Kidney cells	Negative	(Counter et al., 1992)	Silvia Bacchetti/AdVec
SW13 PD73	IMR90 lung fibroblasts	Negative	(Shay and Wright, 1989)	Jerry Shay, UTSW
SW26 PD68.5	IMR90 lung fibroblasts	Negative	(Shay and Wright, 1989)	Jerry Shay, UTSW
SW39 PD72.5	IMR90 lung fibroblasts	Negative	(Shay and Wright, 1989)	Jerry Shay, UTSW
Bet3B p3	NHBE-10 bronchial epithelial cells	Negative	(Bryan et al., 1995)	Roger Reddel, CMRI
Bet3K p7	NHBE-10 bronchial epithelial cells	Negative	(Bryan et al., 1995)	Roger Reddel, CMRI
BFT3B p10	BF-10 bronchial fibroblasts	Negative	(Bryan et al., 1995)	Roger Reddel, CMRI
BFT3G p7	BF-10 bronchial fibroblasts	Negative	(Bryan et al., 1995)	Roger Reddel, CMRI
BFT3K p9	BF-10 bronchial fibroblasts	Negative	(Bryan et al., 1995)	Roger Reddel, CMRI
HA1-IM PD216	Human Embryonic Kidney cells	Positive (see ref.)	(Counter et al., 1992)	Silvia Bacchetti/AdVec
SW13 PD184	IMR90 lung fibroblasts	Positive (Fig.S1A)	(Shay and Wright, 1989)	Jerry Shay, UTSW
SW26 PD130+	IMR90 lung fibroblasts	Positive (Fig.S1A)	(Shay and Wright, 1989)	Jerry Shay, UTSW
SW39 PD130+	IMR90 lung fibroblasts	Positive (Fig.S1A)	(Shay and Wright, 1989)	Jerry Shay, UTSW
Bet3B p25 post-crisis	NHBE-10 bronchial epithelial cells	Positive (see ref.)	(Bryan et al., 1995)	Roger Reddel, CMRI
Bet3K p25 post-crisis	NHBE-10 bronchial epithelial cells	Positive (see ref.)	(Bryan et al., 1995)	Roger Reddel, CMRI
BFT3B p28 post-crisis	BF-10 bronchial fibroblasts	Positive (see ref.)	(Bryan et al., 1995)	Roger Reddel, CMRI
BFT3G p28 post-crisis	BF-10 bronchial fibroblasts	Positive (see ref.)	(Bryan et al., 1995)	Roger Reddel, CMRI
BFT3K p34 post-crisis	BF-10 bronchial fibroblasts	Positive (see ref.)	(Bryan et al., 1995)	Roger Reddel, CMRI

Supplementary Table 2: Number of clones analyzed by high and low-pass WGS

Name	Number Sequenced	
	Low Pass	High Pass
Parental cell lines	3	1
Day 120 (Y) clones	37	5
Day 150 (Z) clones	83	8
Control (CT) clones	8	0

Supplementary Table 3: Oligos used in this study

OLIGO NAME and SEQUENCE	SOURCE
TERT_gRNA_1- AGTCGCGGGGAAGTGTTGCA	This study
TERT_gRNA_2- ATCTGCCAGACAGAGTGCCG	This study
TERT_gRNA_3- TCGAATCGGCCTAGGCTGTG	This study
TERT_gRNA_4- GAAACTCGCGCCGCGAGGAG	This study
TTN_gRNA_1- CCTTGGTGAAGTCTCCTTTG	(Chavez et al., 2015)
TTN_gRNA_2- ATGTTAAAATCCGAAAATGC	(Chavez et al., 2015)
TTN_gRNA_3- GGGCACAGTCCTCAGGTTTG	(Chavez et al., 2015)
TTN_gRNA_4- ATGAGCTCTCTTCAACGTTA	(Chavez et al., 2015)
TERT qPCR forward- GGAGCAAGTTGCAAAGCATTG	This study
TERT qPCR reverse- TCCCACGACGTAGTCCATGTT	This study
TTN qPCR forward- TGTTGCCACTGGTGCTAAAG	This study
TTN qPCR reverse- ACAGCAGTCTTCTCCGCTTC	This study
β-actin qPCR forward-TGGATCAGCAAGCAGGAGTATG	This study
β-actin qPCR reverse- GCATTTGCGGTGGACGAT	This study
STELA telorette1- TGCTCCGTGCATCTGGCATCCCCTAAC	(Baird et al., 2003)
STELA telorette2- TGCTCCGTGCATCTGGCATCTAACCCT	(Baird et al., 2003)
STELA telorette3- TGCTCCGTGCATCTGGCATCCCTAACC	(Baird et al., 2003)
STELA telorette4-TGCTCCGTGCATCTGGCATCCTAACCC	(Baird et al., 2003)
STELA telorette5- TGCTCCGTGCATCTGGCATCAACCCTA	(Baird et al., 2003)
STELA telorette6- TGCTCCGTGCATCTGGCATCACCCTAA	(Baird et al., 2003)
STELA XpYpE2- TTGTCTCAGGGTCCTAGTG	(Baird et al., 2003)
STELA teltail - TGCTCCGTGCATCTGGCATC	(Baird et al., 2003)
STELA XpYpB2- TCTGAAAGTGGACCTATCAG	(Baird et al., 2003)
Fusion PCR XpYpM- ACCAGGTTTTCCAGTGTGTT	(Capper et al., 2007)
Fusion PCR 17p6- GGCTGAACTATAGCCTCTGC	(Capper et al., 2007)
Fusion PCR 21q4- GGGACATATTTTGGGGTTGC	(Capper et al., 2007)
Fusion PCR XpYpc2tr- GCTATGGCTTCTTGGGGC	(Capper et al., 2007)
Fusion PCR 21q-seq-rev2-ACACAGAAGGTTGATATACACAG	(Capper et al., 2007)