Cell

Leading Edge
Primer

Criteria for Inference of Chromothripsis in Cancer Genomes

Jan O. Korbel[1,*] and Peter J. Campbell[2,3,4,*]
[1]Genome Biology Unit, European Molecular Biology Laboratory, 69117 Heidelberg, Germany
[2]Cancer Genome Project, Wellcome Trust Sanger Institute, Hinxton CB10 1SA, UK
[3]Department of Haematology, Addenbrooke's Hospital, Cambridge CB2 0QQ, UK
[4]Department of Haematology, University of Cambridge, Cambridge CB22XY, UK
*Correspondence: jan.korbel@embl.de (J.O.K.), pc8@sanger.ac.uk (P.J.C.)
http://dx.doi.org/10.1016/j.cell.2013.02.023

Chromothripsis scars the genome when localized chromosome shattering and repair occurs in a one-off catastrophe. Outcomes of this process are detectable as massive DNA rearrangements affecting one or a few chromosomes. Although recent findings suggest a crucial role of chromothripsis in cancer development, the reproducible inference of this process remains challenging, requiring that cataclysmic one-off rearrangements be distinguished from localized lesions that occur progressively. We describe conceptual criteria for the inference of chromothripsis, based on ruling out the alternative hypothesis that stepwise rearrangements occurred. Robust means of inference may facilitate in-depth studies on the impact of, and the mechanisms underlying, chromothripsis.

Introduction

Often described as a disease of the genome, cancer typically results from the acquisition of DNA alterations in somatic cells leading to activation of oncogenes and inactivation of tumor suppressor genes. As a result, cellular processes including cell-cycle control, apoptosis, and DNA repair are impaired, conferring a growth advantage to cells and fomenting tumorigenesis (Stratton et al., 2009). According to a long-standing presumption, a single genetic hit is typically insufficient for a cell to develop into cancer. Instead, several progressive (i.e., gradually acquired or stepwise) DNA alteration events are required, resulting in incremental development and progression of cancer (Knudson, 1971; Stratton et al., 2009).

Recent cancer genome analyses, however, have revisited this presumption by suggesting an alternative process that involves massive de novo structural rearrangement formation in a one-step catastrophic genomic event coined chromothripsis (Stephens et al., 2011) ("chromo" from chromosome; "thripsis" for shattering into pieces; illustrated in Figures 1A and 1B). A key feature of chromothripsis is the formation of tens to hundreds of locally clustered DNA rearrangements through a singular, cataclysmic (one-off) event, resulting in a large number of rearranged fragments (often tens to hundreds) interspersed with widespread losses of sequence fragments (Figure 1B). Occasionally, rearrangements resulting from chromothripsis can lead to the formation of small circular DNA molecules (double-minute chromosomes), which may subsequently become amplified if they harbor oncogenes (Rausch et al., 2012a; Stephens et al., 2011) (Figure 1B). As a result of the massive DNA alterations occurring, chromosomes affected by chromothripsis show a characteristic pattern of copy-number "oscillations," whereby typically only two (or occasionally three) copy-number states are detectable along the chromosome in the context of a large number of rearrangements (Stephens et al., 2011).

This pattern distinguishes chromothripsis from other "punctuated equilibrium"-like mechanisms in which one-off events precipitate multiple successive DNA rearrangements. An example of the latter is the breakage-fusion-bridge cycle (Figure 1C), in which one DNA double-strand break can result in further DNA alterations acquired with each subsequent cell cycle (Bignell et al., 2007; Rudolph et al., 2001). Such processes, although occurring in a short period of time, are conceptually different to chromothripsis because they are associated with DNA replication interspersed with progressive rearrangements, and thus copy-number states can vary extensively across the derivative chromosome.

Impact of Chromothripsis on Cancer Development and Progression

The DNA breakpoints resulting from chromothripsis frequently affect only one or a few chromosomes (Figure 2A). Spectral karyotyping and fluorescent in situ hybridization (FISH) experiments have further shown that only one of the two parental chromosomes (or haplotypes) is typically affected by chromothripsis (Stephens et al., 2011). DNA rearrangements arising through chromothripsis can lead to several simultaneous tumorigenic DNA alterations (Rausch et al., 2012a; Stephens et al., 2011) (illustrated in Figure 1A and Figure 2B). FISH experiments further showed rearrangement outcomes of chromothripsis to be detectable throughout practically all cells in a tumor and not solely in tumor subclones (e.g., Figure 2B), suggesting that chromothripsis occurs as a relatively early tumorigenic event (Rausch et al., 2012a; Stephens et al., 2011). Hence, chromothripsis is thought to contribute to, or even represent a driving force of, cancer development and progression.

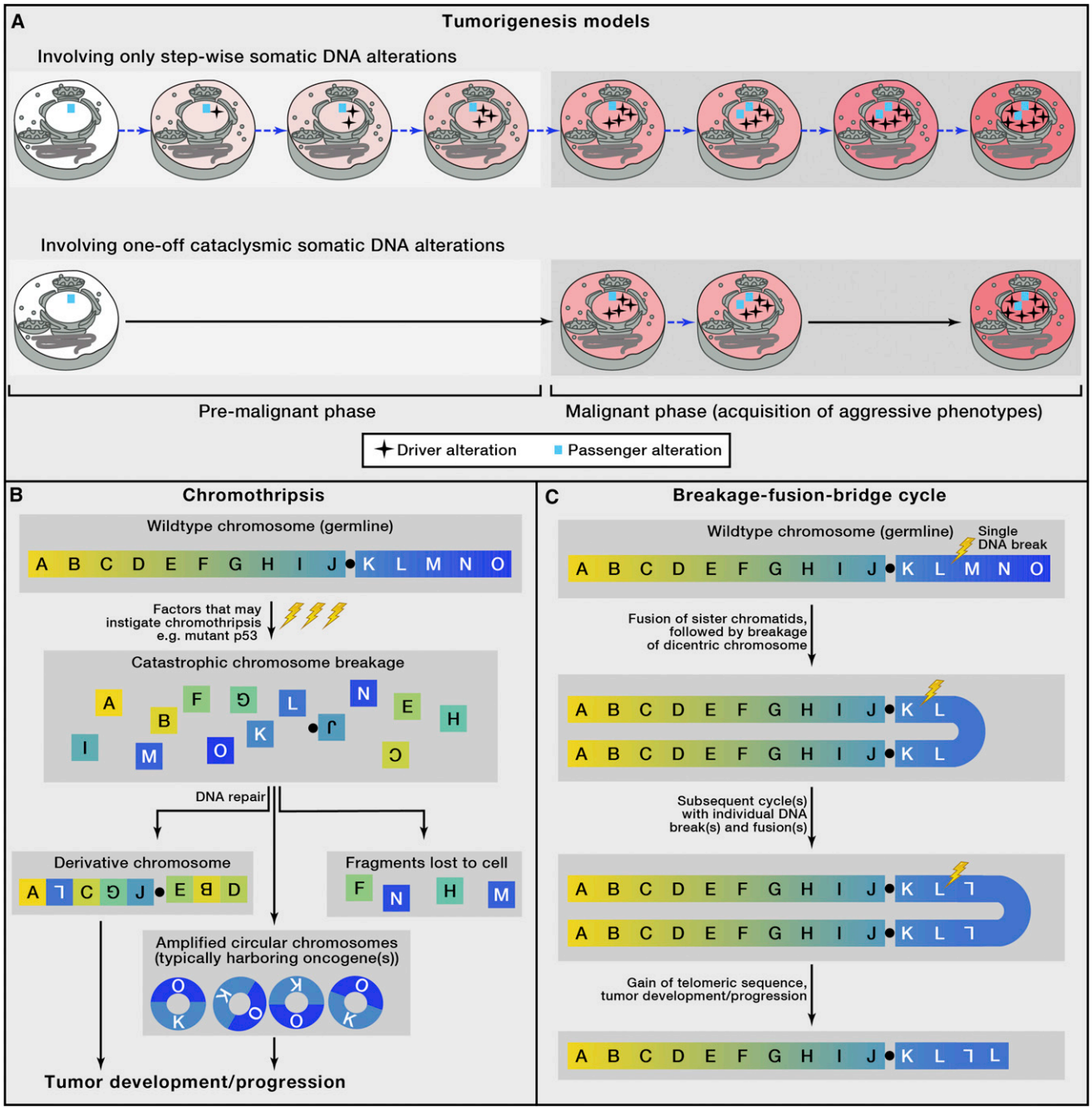

Figure 1. Cataclysmic DNA Rearrangement Processes

(A) Tumorigenesis is classically thought to involve the stepwise acquisition of somatic DNA driver alterations (dashed blue arrows). Cellular "crises," such as chromothripsis, may accelerate this process by resulting in several DNA alterations at once (solid black arrows). The red color symbolizes the acquisition of malignant phenotypes in the cell (white = nonmalignant cell; red = aggressive/highly malignant cell).

(B) Chromothripsis, a cellular crisis altering chromosomes in a one-off burst thought to involve a single cell cycle (adapted from Stephens et al., 2011, Rausch et al., 2012a).

(C) The breakage-fusion-bridge cycle, a prototypic process (McClintock, 1941) involving chromosome end-to-end fusions that lead to clustered breakpoints but not to extensive copy-number state oscillations. This form of "crisis" typically involves several subsequent cell cycles. Though in the classical breakage-fusion bridge cycle only a single DNA break is thought to occur in each cell-division cycle, it is hypothesized that chromosome end-to-end fusions may also lead to chromothripsis events (Stephens et al., 2011).

The characteristic signature of massive DNA rearrangements resulting from chromothripsis has been observed in 2%–3% of cancer samples (Stephens et al., 2011). Distinct malignancies display different rates of chromothripsis (reviewed in Jones and Jallepalli, 2012), and the outcomes of such one-off chromosomal crises have been reported in diverse cancer entities,

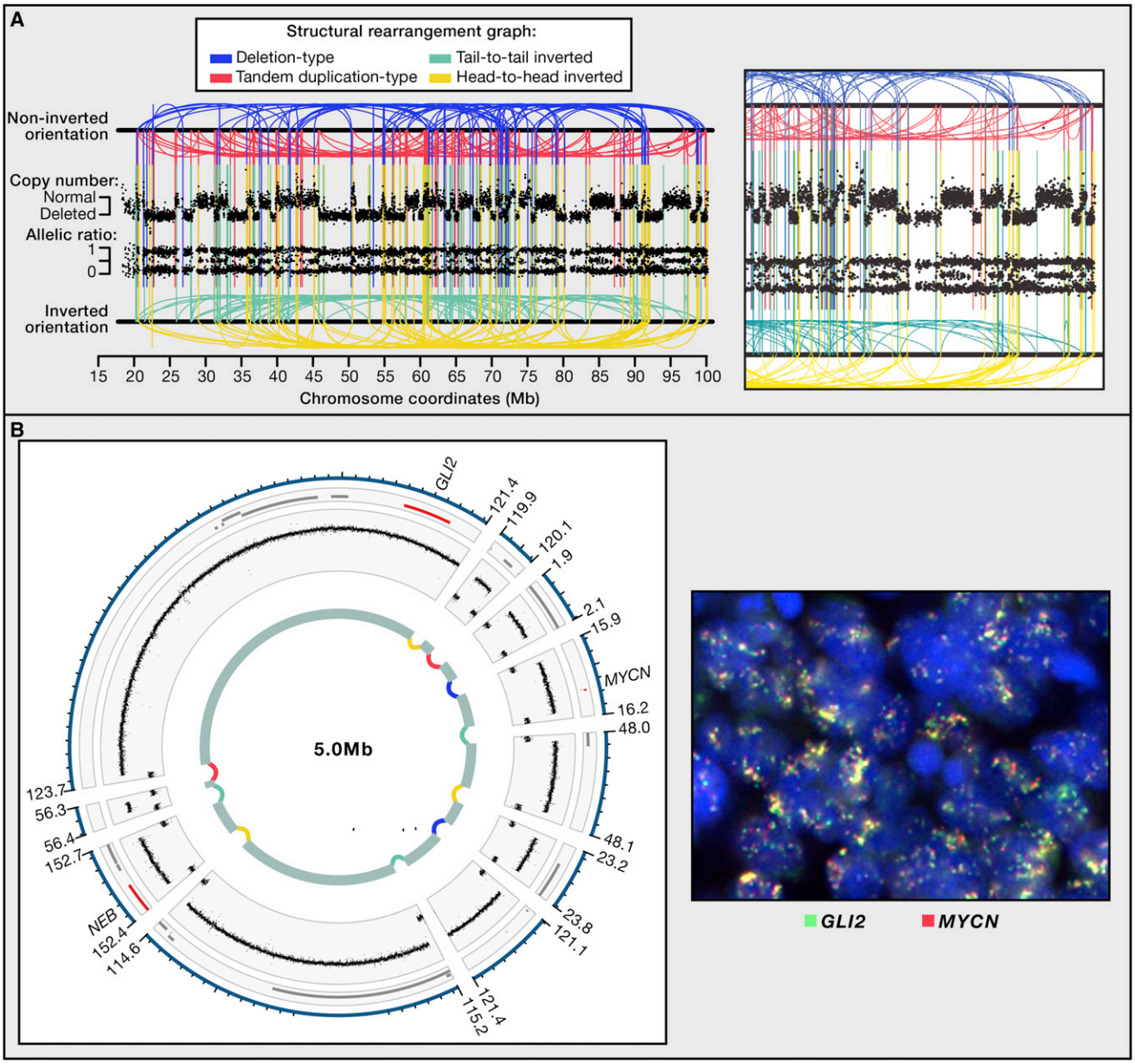

Figure 2. Appearances of Chromothripsis and Progressive DNA Rearrangements

(A) DNA rearrangement pattern of SNU-C1, a tetraploid colorectal cancer cell line, with >200 rearrangements on chromosome 15 associated with widespread DNA fragment loss (reproduced from Stephens et al., 2011). Oscillating copy-number profiles derived from SNP6 microarray data are depicted in the upper panel of points. Allelic ratios for each SNP, depicting segments with retained heterozygosity interspersed with LOH, are shown in the lower panel of dots. Homozygous SNPs cluster at allelic ratios near 0 or 1. Heterozygous SNPs cluster around 0.5. The structural rearrangement graph with intrachromosomal rearrangements of all four possible orientations is depicted as colored lines that connect DNA segments. The box to the right shows a zoomed-in version of the 15q region. Abundant regions with LOH indicate that chromothripsis preceded genome duplication in this cancer cell line.

(B) Chromothripsis in a primary Shh-driven pediatric medulloblastoma sample LFS-MB4 associated with the formation of a circular double-minute chromosome derived from chromosome 2 fragments (reproduced from Rausch et al., 2012a). The outermost rings in the illustrated circular plot depict chromosome coordinates and annotated genes with known oncogenes shown in red. FISH analysis verified the colocalization of the synchronously amplified *MYCN* (red) and *GLI2* (green) oncogenes in the chromothripsis-associated, amplified double-minute chromosomes, and demonstrated their presence throughout virtually all tumor cells (reproduced from Rausch et al., 2012a).

including bone cancer, pediatric medulloblastoma, neuroblastoma, colorectal cancer, melanoma, and hematological malignancies (Hirsch et al., 2012; Kloosterman et al., 2011b; Magrangeas et al., 2011; Molenaar et al., 2012; Northcott et al., 2012; Rausch et al., 2012a; Stephens et al., 2011). Furthermore, chromothripsis has been associated with poor patient survival in several cancers (Hirsch et al., 2012; Magrangeas et al., 2011; Molenaar et al., 2012; Rausch et al., 2012a), indicating its

potential relevance as a prognostic marker, and suggesting chromothripsis as a feature of some particularly aggressive forms of cancer. In sonic hedgehog (Shh)-driven medulloblastoma, chromothripsis has been linked with predisposing (germline) mutations in the gene encoding the p53 tumor suppressor (*TP53*) (Rausch et al., 2012a), and in group-3-subtype medulloblastoma and acute myeloid leukemia with somatic DNA alterations of *TP53* (Northcott et al., 2012; Rausch et al., 2012a). Hence, chromothripsis appears to be prone to occur in specific contexts—i.e., in conjunction with, or even instigated by, progressively acquired DNA alterations.

Mechanisms Hypothesized to be Involved in Chromothripsis

Although we can find evidence of these cataclysmic events in genomes, the mechanisms that give rise to them are still being worked out. Computational analyses of breakpoint junction sequences performed at nucleotide resolution have provided initial clues on the mechanism for rejoining the shattered DNA fragments. Abundant 2–4 nt long repeating sequences (i.e., observed "microhomology") at the respective rearrangement breakpoints (Stephens et al., 2011) are consistent with the repair of shattered DNA fragments by nonhomologous end-joining (NHEJ). Simulation-based computational analyses, described in more detail below, have further provided compelling evidence that the complex chromosome aberrations resulting from chromothripsis result from singular, catastrophic DNA rearrangement event (Rausch et al., 2012a; Stephens et al., 2011).

Several hypothetical mechanisms have been proposed to lead to the massive DNA rearrangements observed in conjunction with chromothripsis (recently reviewed in Forment et al., 2012; Jones and Jallepalli, 2012; Maher and Wilson, 2012). Most proposed mechanisms assume that chromothripsis acts on condensed chromosomes in association with mitosis, which may explain the highly localized nature of DNA breakpoints on a single (or few) chromosomes (Stephens et al., 2011)—although localized DNA shattering could also occur in the context of the regular spatial organization of interphase chromosomes (Lichter et al., 1988; Rausch et al., 2012a). In brief, the following mechanistic hypotheses have been presented and discussed: ionizing radiation acting upon condensed chromosomes (Stephens et al., 2011); critical telomere shortening followed by chromosome end-to-end fusions and subsequent massive DNA breakage (Stephens et al., 2011); abortive apoptosis events (Tubio and Estivill, 2011); "premature chromosome compaction," in which chromosomes condense before completing DNA replication and may consequently shatter (Johnson and Rao, 1970; Meyerson and Pellman, 2011); and DNA damage associated with the packaging of mitotically "delayed" chromosomes into separate cellular compartments known as micronuclei (Crasta et al., 2012). In this regard, a particularly relevant observation made by Crasta and coworkers is that of DNA fragmentation affecting isolated chromosomes packaged into micronuclei, which addresses the conceptual problem of how highly localized DNA shattering, in the context of chromothripsis, might be achieved at the molecular level.

Beyond reports of chromothripsis in many cancers, there is evidence that a similar (or perhaps identical) process may act upon germline DNA, resulting in constitutional disorders (Chiang et al., 2012; Kloosterman et al., 2011a; Liu et al., 2011). Nucleotide resolution analyses of the DNA breakpoint junctions of "constitutional chromothripsis" events revealed the presence of microhomology compatible with NHEJ in some patients (Kloosterman et al., 2011a; Kloosterman et al., 2012). In others, sequence-based evidence for replication-associated structural rearrangements involving the proposed microhomology-mediated break-induced replication (MMBIR) mechanism was reported (Liu et al., 2011), with MMBIR thought to be frequently associated with duplication events and with the insertion of short DNA-template-derived sequences (i.e., templated insertions) at the respective breakpoint junctions. The frequent association of chromothripsis in cancer with sequence loss (Stephens et al., 2011), rather than with duplication, and the lack of template-derived insertions at the respective DNA breakpoints in medulloblastoma (Rausch et al., 2012a) suggests that they may differ mechanistically from "constitutional chromothripsis" events. As is the case for chromothripsis in cancer, the molecular mechanism driving "constitutional chromothripsis" has not yet been experimentally elucidated.

Challenges in the Assessment of Chromothripsis in Cancer Genomes

Accurate inference of chromothripsis is crucial for further characterization of the underlying molecular process. However, the genomic signature left by other processes can resemble that of chromothripsis potentially resulting in misclassification of chromothripsis events that may hamper research on the mechanistic basis of chromothripsis and impede attempts to exploit chromothripsis as a biomarker for disease prognosis. To robustly and reproducibly identify DNA rearrangements arising from chromothripsis, those alterations underlying a one-off event must be distinguished from DNA alterations occurring in a stepwise manner.

Different operational definitions have been applied for inferring chromothripsis in microarray based copy-number profiling data. These operational definitions have been geared toward recognizing oscillating copy-number profiles, by requiring the detection of at least 10, 20, or 50 copy-number alterations (i.e., identifiable shifts in the copy-number profile) on a particular chromosome, with these alterations oscillating between only two or three copy-number states (Hirsch et al., 2012; Jones et al., 2012; Magrangeas et al., 2011; Molenaar et al., 2012; Northcott et al., 2012; Rausch et al., 2012a; Stephens et al., 2011). In addition to requiring a fixed number of copy-number alterations (such as 50) as a threshold, the number of DNA breakpoints associated with oscillating copy-number alterations has been put in relation to the total number of breakpoints on a chromosome to define a threshold for inferring chromothripsis in microarray data (Kim et al., 2013).

Marked differences in the spatial distribution, number, and types of somatically acquired DNA rearrangements observed between cancer entities (Yates and Campbell, 2012), however, limit the utility of a defined threshold in terms of identified copy-number alterations for ascertaining chromothripsis. Specifically, cancers displaying pronounced genomic instability, such as ovarian cancer (Cancer Genome Atlas Research

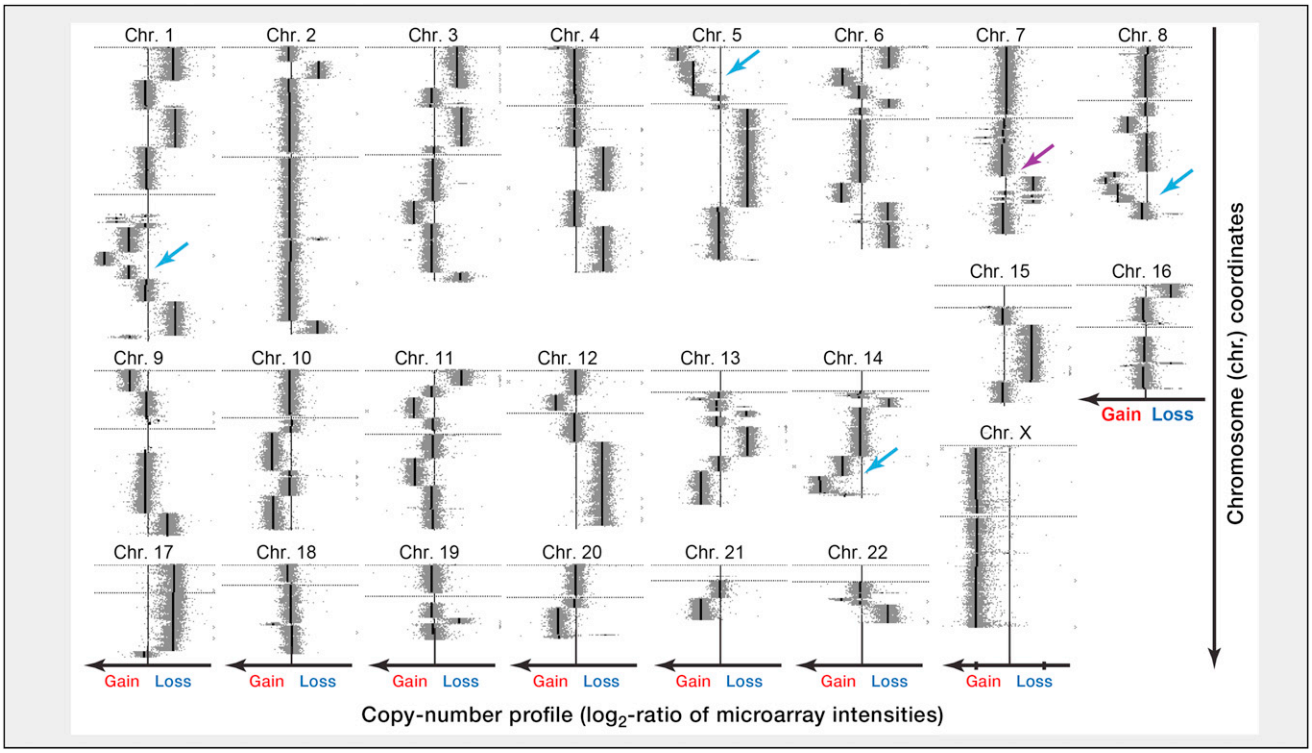

Figure 3. Amalgam of DNA Rearrangements in a Cancer Genome from an Ovarian Cancer Patient

The large number and diversity of DNA rearrangements detectable in this cancer genome highlight the necessity to use rigorous statistics for distinguishing chromothripsis events from progressive DNA alterations. Ovarian cancers show widespread DNA copy-number alterations throughout the genome, most of which involve progressive rearrangements (depicted by light blue arrows). Although chromosome 7 may potentially have undergone chromothripsis (purple arrow), the large genome-wide number of alterations limits the utility of operational definitions for inference—hence calling for rigorous statistical testing. This cancer genome copy-number alteration profile was determined using microarrays (Cancer Genome Atlas Research Network, 2011). Array data were reanalyzed with Nexus 6v10 (Biodiscovery) copy-number software, as described in Rausch et al., 2012a. Scales corresponding to array $\log_2$ ratios of 1 (gain) and −1 (loss) are indicated beneath the axis corresponding to the X chromosome.

Network, 2011), can harbor such a high number of progressively acquired somatic DNA alterations per chromosome (Figure 3) that based on operational definitions, those cancers may mistakenly be suspected to have undergone chromothripsis. Additionally, accumulations of DNA alterations on the same chromosome can be achieved by multistep processes, rather than one-off events, e.g., through successional breakage-fusion-bridge cycles (Bignell et al., 2007; Rudolph et al., 2001) or through consecutive deletions that originate from fragile sites or are driven by positive selection (Bignell et al., 2010). Thus, although operational definitions can facilitate the screening for chromothripsis in microarray copy-number profiling data, from which copy-number state information but not the relative order or orientation of rearrangements can be reconstructed, their utility is noticeably limited—and because operational definitions are prone to subjectivity, they can interfere with reproducibility.

Criteria for Statistical Assessment of Chromothripsis

A more robust and accurate distinction between DNA rearrangements arising from chromothripsis and those occurring in a stepwise fashion can be achieved by applying criteria that enable rigorous statistical evaluation of cancer genome sequencing data (Rausch et al., 2012a; Stephens et al., 2011). The aim of these criteria is to evaluate the model that a particular set of DNA rearrangements resulted from stepwise somatic DNA alterations as compared to the alternative model that the rearrangements arose through a single catastrophic event (i.e., chromothripsis).

The following sections outline the rationale behind several different criteria, each of which can facilitate the statistical inference of chromothripsis, allowing for more reproducible and accurate ascertainment of chromothripsis than otherwise possible using solely operational definitions. Most of these criteria take into account the entire set of structural rearrangements that have occurred on a chromosome in question, including the relative order and orientation of rearranged segments, which are typically detected using whole-genome paired-end DNA sequencing data, and which can be represented in the form of a structural rearrangement graph (Figure 2A and Box 1).

Clustering of Breakpoints

DNA breakpoints occurring in conjunction with chromothripsis typically show pronounced clustering (depicted in Figure 4A). Often, 5–10 breaks can be observed within 50 kb, followed by

Box 1. Construction of DNA Structural Rearrangement Graphs

A crucial prerequisite for the inference of chromothripsis is the accurate mapping of somatically acquired DNA structural rearrangements in samples of interest to obtain a structural rearrangement graph, which represents the set of somatic rearrangements that occurred on a chromosome, comprising copy-number state information and data on the relative order and orientation of segments subsequent to rearrangement (see e.g., Figure 2A). Accurate structural rearrangement graphs can be obtained using sequence variant discovering approaches in massively parallel DNA sequencing data. These approaches include paired-end mapping, which is based on sequencing the ends of size-selected DNA fragments, and detecting DNA rearrangements by identifying paired ends that map abnormally onto the human reference assembly (Campbell et al., 2008; Korbel et al., 2007; Mills et al., 2011). Deletion-type rearrangements (tail-to-head) are inferred based on the abnormal distance of mapped ends, tandem duplication-type (head-to-tail) alterations based on their abnormal relative mapping order, and inversion-type alterations (head-to-head or tail-to-tail) based on their abnormal relative mapping orientation. The sensitivity of paired-end mapping for detecting DNA alterations is improved when DNA sequencing libraries with different library insert sizes are used (Mills et al., 2011; Rausch et al., 2012a). Read-depth analysis (Campbell et al., 2008; Chiang et al., 2009), an approach based on identifying copy-number alterations by analyzing the DNA read depth of coverage, can also be used to discover structural rearrangements and to infer the copy-number status of segments. Split-read (or clipped-read) analysis, which is based on evaluating gapped read alignments onto the human reference genome assembly, enables the fine-mapping of DNA rearrangement breakpoints (Rausch et al., 2012b; Wang et al., 2011; Ye et al., 2009). In theory, DNA sequence assembly can further improve the detection of structural rearrangement events, although recent analyses suggest that assembly using short DNA read data displays low sensitivity compared to the aforementioned sequence variant discovery approaches (Mills et al., 2011).

Data from several of these rearrangement discovery approaches are typically combined to describe the somatic DNA structural rearrangement graph. This graph serves as the starting point for the described criteria for inferring chromothripsis.

long tracts of intact chromosomal sequence. Breakpoints can be confined to individual chromosome arms with the clustering presumably resulting from whatever process drives the chromosome fragmentation (Stephens et al., 2011). Thus, an analysis of breakpoint clustering can be used as means to obtain evidence for chromothripsis (Rausch et al., 2012a; Stephens et al., 2011), as outlined in Box 2.

Under a progressive rearrangements model, tendencies of breakpoints to cluster substantially imply a "memory" of previous rearrangements from one cell division to the next. Although less pronounced than in chromothripsis, local accumulation of breakpoints can be observed in progressive rearrangement scenarios where it may be driven by either chromosomal fragility or selection for particular genes within a chromosomal region (Campbell et al., 2010). As a consequence, under progressive rearrangement scenarios, breakpoint clustering tends to be recurrent across patients because both the locations of cancer genes and fragile sites represent intrinsic features of the human genome (a priori information that can be taken into consideration for rigorous statistical evaluation of breakpoint clustering).

Regularity of Oscillating Copy-Number States

The aforementioned oscillating behavior of copy-number states resulting from chromothripsis (e.g., as evident from the chromothripsis example shown in Figure 2A) can be evaluated rigorously, as illustrated in Figure 4B, by simulating a gradual process in which each of the structural rearrangements detected on a chromosome, according to the rearrangement graph, are introduced onto an in silico (modeled) chromosome one-after-another (Rausch et al., 2012a; Stephens et al., 2011). By introducing these rearrangements in a stepwise fashion using Monte Carlo simulations, we can assess the ability of the progressive rearrangement null model to reproduce the regular (oscillating) nature of copy-number state switches characteristic for chromothripsis. Support for chromothripsis is obtained in cases where the null model is ruled out based on these simulations.

Prevalence of Regions with Interspersed Loss and Retention of Heterozygosity

Chromothripsis frequently leads to massive loss of segments on the affected chromosome with segmental losses being interspersed with regions displaying normal (disomic) copy-number (e.g., copy-number states oscillating between copy-number = 1 and copy-number = 2). Although monosomic regions have evidently lost heterozygosity, the key feature of chromothripsis is that the segments in the higher (disomic) copy-number state have retained heterozygosity (Stephens et al., 2011). The result is a highly regular (oscillating) pattern of segments with retained heterozygosity interspersed with loss-of-heterozygosity (LOH) (see Figure 2A, and illustration in Figure 4C). Once lost to the cell through deletion, heterozygosity cannot be regained. Hence, in the presence of an abundance of copy-number states oscillating between the states "1" and "2," perfect concordance between disomic regions and heterozygous regions will be unlikely in the event of gradually acquired rearrangements (Figure 4C). A simulation in which rearrangements are randomly and sequentially drawn from the available structural rearrangement graph (Box 1) can be employed to assess this concordance and hence to evaluate the hypothesis that DNA rearrangements were gradually acquired.

It is worth noting that if chromothripsis occurs in the context of polyploidy, the lower copy-number state may not display LOH, but instead may reflect the resulting allelic contribution in lost genomic segments (e.g., alternating between allelic ratios of 1:1 and 2:1, if genome duplication precedes chromothripsis). Nonetheless, in the case of chromothripsis, the resulting allelic ratios will oscillate between segments that are lost and retained, and evaluation of concordance of this oscillating behavior with the segmental copy-number state changes can hence facilitate the discrimination of one-off from progressive rearrangements in the context of polyploidy.

Prevalence of Rearrangements Affecting a Single Haplotype

When chromothripsis occurs, fragments resulting from chromosomal DNA shattering typically originate from a single parental

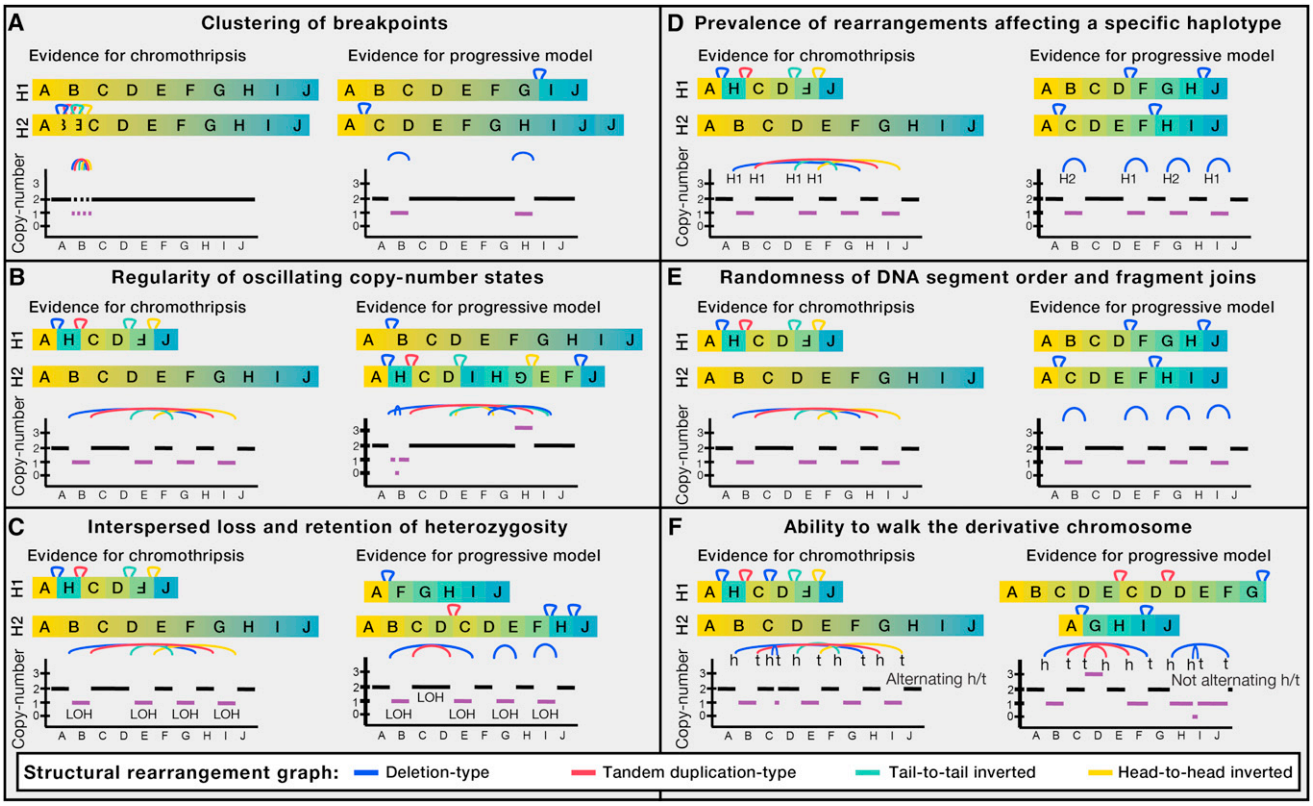

Figure 4. Criteria for the Inference of Chromothripsis

(A) Breakpoint clustering can yield evidence for chromothripsis (left), as stepwise alterations (right) do not typically lead to a similar level of clustering of DNA breaks. Curved colored lines depict individual rearrangements.

(B) Oscillating copy-number profiles. The left panel depicts a particular set of rearrangements resulting in oscillating copy number, indicative of chromothripsis. The null hypothesis of stepwise alterations can be rejected if simulations making use of all rearrangements depicted in the rearrangement graph fail to result in oscillations involving so few (in this case two) copy-number states. This is illustrated in the right panel, where the copy-number profile displays four different states.

(C) Interspersed regions with loss and retention of heterozygosity often result from chromothripsis (left) and can be used for statistical testing as in the presence of stepwise alterations (right) such regularity of patterns is unlikely to occur.

(D) Chromothripsis-associated rearrangements are typically detectable on a single parental copy (haplotype) of affected chromosomes (referred to as H1 in the left), whereas stepwise alterations do not typically show such preference.

(E) Because fragments are randomly joined following DNA shattering (left), it follows that the relative order of rearranged fragments and the type of fragment joins should be uniformly distributed. By comparison, clustered stepwise alterations often show biases toward certain rearrangement forms (right), and are thus not expected to result in such uniform joining and ordering of segments.

(F) In a region of chromothripsis, each fragment is either retained in or lost from the derivative chromosome, enabling an unambiguous walk through the rearrangements created. As a result, when viewed on the reference genome, adjacent reads demarcating breakpoints inferred by paired-end mapping show perfect alternations between head ("h") and tail ("t") paired-end reads (left). In contrast, most progressive DNA alteration scenarios that result in nested rearrangements (right) do not have this property.

chromosome (or haplotype). Considering that DNA rearrangements can be associated with a specific haplotype using phasing (Box 3), the extent to which rearrangements are biased toward a single haplotype, rather than occurring on both haplotypes (assuming disomy), can be used to obtain further evidence for chromothripsis (Figure 4D). Under the assumption that progressive rearrangements affect each haplotype randomly, a statistical test can provide evidence for chromothripsis by defining the extent to which rearrangements are concentrated on a single haplotype—for example, by using the Poisson assumption that in the presence of progressive rearrangements structural rearrangements occur on both haplotypes (null hypothesis), rather than only on a single one.

Selection for particular genes within a chromosomal region can bias progressive rearrangements to occur preferentially on only one rather than on both haplotypes. However, the rearranged genomic regions would be recurrent across patients if driven by selection, providing a possible rationale to account for such potentially confounding factor.

Randomness of DNA Fragment Joins

The assumption underlying the chromothripsis theory is that the chromosome fragments are randomly stitched together ("joined"), involving a DNA double-strand repair process. The implication is that at each join, the orientation of the two DNA fragment ends should be random (illustrated in Figure 4E), in

Box 2. Outline of Statistical Algorithms for Inferring Chromothripsis

Guidelines for evaluating the following four criteria are outlined:

1. *Clustering of breakpoints*: Let $\{x_{(1)}, x_{(2)}, \ldots, x_{(n)}\}$ be the set of breakpoint locations on a given chromosome, ordered from the lowest to the highest (as positioned on the reference genome). The null model of random breakpoint locations implies that the distances between adjacent breakpoints, $\{x_{(2)} - x_{(1)}, x_{(3)} - x_{(2)}, \ldots, x_{(n)} - x_{(n-1)}\}$, should be distributed according to an exponential distribution with mean $\sum_1^{n-1}(x_{(i+1)} - x_{(i)})/(\mathrm{n}-1)$ which can be readily evaluated using a goodness-of-fit test. In our experience, chromothripsis is typically associated with a strong departure from this null distribution, although some situations of progressive rearrangements (e.g., rearrangements arising through successive breakage fusion bridge cycles; Figure 1C) are too.

2. *Randomness of DNA fragment joins*: Let $\{r_{\mathrm{Del}}, r_{\mathrm{TD}}, r_{\mathrm{H2H}}, r_{\mathrm{T2T}}\}$ be the counts of observed rearrangements that have a deletion-type, tandem duplication-type, head-to-head-inverted, and tail-to-tail-inverted orientation respectively. If more than one chromosome is involved, then interchromosomal rearrangements can be interpreted in the same four categories using orientation of the strands at the breakpoint. Then, in a region of chromothripsis, we would expect these counts to be distributed as a multinomial distribution with parameters $n = \sum r_i$ and probability $p_i = 1/4$. A departure from this distribution can be employed as evidence against the rearrangements arising from a chromothripsis process.

3. *Randomness of DNA fragment order*: In a chromothripsis event, the presumption is that the original position of a fragment on the reference genome carries no information about the origins of the fragments it is joined to at either end. To test this, let $\{x_{(1)}, x_{(2)}, \ldots, x_{(n)}\}$ be the set of breakpoint locations, ordered from the lowest to the highest (as positioned on the reference genome). Each observed rearrangement consists of two DNA breaks joined together and can be denoted as $\{(I_1, I_2)\}$, where I refers to the index of the ordered breakpoints $\{x_{(I)}\}$. Under the chromothripsis model, the paired indices should be random draws without replacement from $\{1, 2, \ldots, n\}$. There are suites of tools available for statistically assessing randomness that could be adapted here. One possibility, for example, would be to calculate the mean of $\{|I_2 - I_1|\}$ and compare this to 1,000 Monte Carlo simulations. When we have tested this in practice, the fragment order from a chromothripsis process is not entirely random, implying some spatial structure to the DNA repair process but considerably more random than most progressive rearrangement scenarios.

4. *Ability to walk the derivative chromosome*: As can be seen in Figure 4F, the ability to walk the derivative chromosome implies that adjacent DNA reads demarcating breakpoints inferred by paired-end mapping (Box 1) must alternate between the head of the paired-end fragment and the tail of that fragment. Let $\{x_{(1)}, x_{(2)}, \ldots, x_{(n)}\}$ be the set of breakpoint locations, ordered from the lowest to the highest (as positioned on the reference genome), and let $\{s_{(1)}, s_{(2)}, \ldots, s_{(n)}\}$ be the paired-end DNA read (head or tail) associated with each of these breakpoints. In a region of chromothripsis, if all rearrangements were observed, $\{s_{(1)}, s_{(2)}, \ldots, s_{(n)}\}$ would be a perfect alternating sequence of heads and tails when ordered along the reference genome assembly (Figure 4F). Because some rearrangements are likely to be missed in the sequencing, the problem is one of whether there are longer runs of alternating heads and tails than expected by chance. This circumstance could be assessed by adapting the Wald-Wolfowitz test for runs. Note that some progressive rearrangement processes could give similar runs, such as a series of deletions on a given chromosome, but that many processes, especially those associated with amplification, will not.

Box 3. Application of Haplotype Phasing to Improve Structural Rearrangement Analysis

A normal human genome is diploid (2n), and cancer genomes can display different karyotype configurations (e.g., tetraploidy, 4n). According to the theory of chromothripsis, structural rearrangements arising should normally display a bias toward occurring on a single chromosome homolog (i.e., haplotype), rather than on both haplotypes for disomic karyotypes (or all four in the case of 4n). Hence, the ability to relate rearrangements to a specific haplotype would allow inferring chromothripsis events with increased power. Using short read DNA sequencing, haplotype phases of 300–400 kb could be used to monitor whether adjacent DNA breakpoints arose on a single DNA molecule, using the 1000 Genomes Project integrated haplotype reference panel (1000 Genomes Project Consortium et al., 2012) in conjunction with computational approaches based on imputation (Browning and Browning, 2011). Chromosome-wide phasing data can be obtained when germline whole-genomic sequencing data from both parents or somatic genome sequencing data from aneuploid secondary tumors (which are common in the context of hereditary disorders such as Li-Fraumeni syndrome; Li and Fraumeni, 1969) are available for a patient sample in question.

analogy to a pearl necklace that after being disrupted is put together, with the pearls added to the chain in random order and orientation.

To evaluate rearrangement patterns for chromothripsis, the uniformity of orientation of joined DNA fragments can be inferred by interpreting the structural rearrangement graph (Box 1)—that is, the number of tail-to-head (deletion-type), head-to-tail (tandem-duplication-type), head-to-head and tail-to-tail (inversion-type) rearrangements observed should be broadly equal (Box 2). This criterion applies whether the rearrangements are intrachromosomal or interchromosomal. In contrast, for many other types of clustered rearrangements, this property does not apply. For example, in regions affected by recurrent breakage-fusion-bridge cycles, there will be predominance of head-to-head and tail-to-tail inverted rearrangements, whereas for chromosomal fragile sites, deletions tend to dominate among the spectrum of rearrangements (Campbell et al., 2010).

Randomness of DNA Fragment Order

Because chromosome fragments are randomly joined, their relative order, namely their position on the derivative chromosome, also should be approximately random (Figure 4E) provided that there is no preference for joining particular ends together, such as maintaining a centromere or telomere. Hence, as an extension to the criterion to evaluate the randomness of DNA fragment orientation, an assessment of the randomness of DNA fragment order can be used to obtain further evidence for chromothripsis (Box 2). This criterion applies to both intrachromosomal and interchromosomal rearrangements.

"Walking" the Derivative Chromosome

If all DNA rearrangements in a region with chromothripsis are detectable, it should be possible to reconstruct the relative order in which segments are joined based on the structural rearrangement graph. Computational approaches for piecing together

such ''digital karyotypes'' are being developed (Greenman et al., 2012). For our purposes here, the chromothripsis model means that each DNA segment included in the derivative chromosome resulting from chromothripsis has consistent orientation (namely it has a head at one end and a tail at the other). The derivative chromosome then forms a single, coherent chain of segments with the constraint that either end of each segment must have consistent configuration. Each DNA segment retained in the derivative chromosome must be demarcated at either end by genomic rearrangements—when viewed from the perspective of the reference genome, each separate segment will start with a rearrangement from the head of the segment and finish with a rearrangement from the tail of the segment. This constraint will lead to an alternating head/tail sequence of DNA rearrangements, detected by paired-end mapping (Box 1), when paired-ends demarcating breakpoints are represented along the reference genome (see illustration in Figure 4F).

Importantly, this organization of alternating heads and tails need not be the case under the alternative model because when the rearrangements occur sequentially, some segments can be ''reused'' in the derivative chromosome. This would generally break the perfectly alternating head/tail series. Consider, for example, two tandem duplications, one nested entirely within the other. There are four breakpoints from two DNA rearrangements. The two breakpoints at the lowest genomic reference coordinates are both demarcated by tails (see Figure 4F, right). This would break the alternating head/tail sequence and would be inconsistent with chromothripsis. As described in Box 2, it is relatively straightforward to test for consistency with this criterion.

Summary and Outlook

In this primer, we describe the characterization of chromothripsis within a genome as a statistical question geared toward discriminating rearrangements resulting from chromothripsis from those that result from subsequent stepwise DNA alterations. Approaches for inferring the presence of chromothripsis in genomes harboring appreciable levels of gradually acquired alterations can be viewed as conceptually similar to detecting driver alterations among the tumult of passenger mutations and structural abnormalities typically observed in a cancer genome. In this analogy, the discrimination of driver from passenger alterations in studies focusing on generating ''cancer gene catalogs'' benefits from statistical approaches for rejecting the hypothesis that an event corresponds to a stochastically occurring, inconsequential, passenger alteration (Dees et al., 2012).

Not all of the aforementioned criteria for inferring chromothripsis can be applied to each cancer sample. In cancers harboring extreme levels of genomic instability, the characteristic stamp of chromothripsis may be hidden behind the mass of stepwise alterations in such a way that it may not be confidently detectable with the approaches described here. Tumor heterogeneity and ploidy may affect the inference of chromothripsis. Heterogeneity, as a confounding factor, can be partially dealt with by focusing analyses on those DNA alterations that affected the same subset of cells based on haplotype-specific analyses of subclonal alterations (e.g., using approaches described in Nik-Zainal et al., 2012). Exome sequencing or microarray based copy-number profiles cannot be used to infer order and orientation of rearranged segments, limiting criteria that can be used for inferring chromothripsis to the evaluation of breakpoint clustering, or to operational definitions (such as the enumeration of copy-number state changes). Even the most widely used massively parallel DNA sequencing techniques have remaining limitations, with short DNA reads ($\leq$150 nt) and the most commonly used paired-end library (Box 1) insert sizes (<400 bp) remaining ineffective for ascertaining sequence variation in highly repetitive DNA (Onishi-Seebacher and Korbel, 2011). This technological constraint inevitably limits analyses to ''mappable'' genomic regions, which have been estimated to comprise ~90% of the human reference assembly (1000 Genomes Project Consortium et al., 2012). Hence, the available data may in some cases not be sufficient to infer chromothripsis reliably, in which case the criteria we describe may be biased toward presuming that progressive DNA rearrangements occurred.

Despite these challenges, the criteria described here will enable researchers to ascertain chromothripsis in cancer genomes in a rigorous, and more reliable, fashion than feasible on the basis of operational definitions. We recommend assessment of each of the criteria we described on cancer samples harboring rearrangements that can be clearly attributed to chromothripsis as well as on such harboring DNA alterations that undoubtedly underlie a stepwise process, because this will facilitate identifying optimal parameters for discriminating one-off from progressive alterations, which may depend on sequencing depth and protocol used. With massively parallel DNA sequencing technology increasingly prevailing over microarray-based approaches for cancer genome analysis, we propose that future studies should verify the occurrence of chromothripsis by using sequencing data, and by demonstrating the applicability of different—e.g., at least two—criteria as ''minimal evidence'' for discriminating stepwise from one-off events. We foresee that using robust, reproducible criteria for classification, future research will reveal electrifying insights into the functional consequences and mechanistic basis of chromothripsis.

ACKNOWLEDGMENTS

J.O.K. acknowledges funding from the European Commission (HEALTH-F2-2010-260791). P.J.C. is a Wellcome Trust Senior Clinical Fellow. We thank Tobias Rausch, Stephanie Sungalee, Balca Mardin, Joachim Weischenfeldt, Christopher Buccitelli, and Wolfgang Huber for their thoughtful comments.

REFERENCES

1000 Genomes Project Consortium, Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T., and McVean, G.A. (2012). An integrated map of genetic variation from 1,092 human genomes. Nature *491*, 56–65.

Bignell, G.R., Santarius, T., Pole, J.C., Butler, A.P., Perry, J., Pleasance, E., Greenman, C., Menzies, A., Taylor, S., Edkins, S., et al. (2007). Architectures of somatic genomic rearrangement in human cancer amplicons at sequence-level resolution. Genome Res. *17*, 1296–1303.

Bignell, G.R., Greenman, C.D., Davies, H., Butler, A.P., Edkins, S., Andrews, J.M., Buck, G., Chen, L., Beare, D., Latimer, C., et al. (2010). Signatures of mutation and selection in the cancer genome. Nature *463*, 893–898.

Browning, S.R., and Browning, B.L. (2011). Haplotype phasing: existing methods and new developments. Nat. Rev. Genet. *12*, 703–714.

Campbell, P.J., Stephens, P.J., Pleasance, E.D., O'Meara, S., Li, H., Santarius, T., Stebbings, L.A., Leroy, C., Edkins, S., Hardy, C., et al. (2008). Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. Nat. Genet. *40*, 722–729.

Campbell, P.J., Yachida, S., Mudie, L.J., Stephens, P.J., Pleasance, E.D., Stebbings, L.A., Morsberger, L.A., Latimer, C., McLaren, S., Lin, M.L., et al. (2010). The patterns and dynamics of genomic instability in metastatic pancreatic cancer. Nature *467*, 1109–1113.

Cancer Genome Atlas Research Network. (2011). Integrated genomic analyses of ovarian carcinoma. Nature *474*, 609–615.

Chiang, D.Y., Getz, G., Jaffe, D.B., O'Kelly, M.J., Zhao, X., Carter, S.L., Russ, C., Nusbaum, C., Meyerson, M., and Lander, E.S. (2009). High-resolution mapping of copy-number alterations with massively parallel sequencing. Nat. Methods *6*, 99–103.

Chiang, C., Jacobsen, J.C., Ernst, C., Hanscom, C., Heilbut, A., Blumenthal, I., Mills, R.E., Kirby, A., Lindgren, A.M., Rudiger, S.R., et al. (2012). Complex reorganization and predominant non-homologous repair following chromosomal breakage in karyotypically balanced germline rearrangements and transgenic integration. Nat. Genet. *44*, 390–397, S1.

Crasta, K., Ganem, N.J., Dagher, R., Lantermann, A.B., Ivanova, E.V., Pan, Y., Nezi, L., Protopopov, A., Chowdhury, D., and Pellman, D. (2012). DNA breaks and chromosome pulverization from errors in mitosis. Nature *482*, 53–58.

Dees, N.D., Zhang, Q., Kandoth, C., Wendl, M.C., Schierding, W., Koboldt, D.C., Mooney, T.B., Callaway, M.B., Dooling, D., Mardis, E.R., et al. (2012). MuSiC: identifying mutational significance in cancer genomes. Genome Res. *22*, 1589–1598.

Forment, J.V., Kaidi, A., and Jackson, S.P. (2012). Chromothripsis and cancer: causes and consequences of chromosome shattering. Nat. Rev. Cancer *12*, 663–670.

Greenman, C.D., Pleasance, E.D., Newman, S., Yang, F., Fu, B., Nik-Zainal, S., Jones, D., Lau, K.W., Carter, N., Edwards, P.A., et al. (2012). Estimation of rearrangement phylogeny for cancer genomes. Genome Res. *22*, 346–361.

Hirsch, D., Kemmerling, R., Davis, S., Camps, J., Meltzer, P.S., Ried, T., and Gaiser, T. (2012). Chromothripsis and focal copy number alterations determine poor outcome in malignant melanoma. Cancer Res., in press. Published online December 27, 2012. http://dx.doi.org/10.1158/0008-5472.

Johnson, R.T., and Rao, P.N. (1970). Mammalian cell fusion: induction of premature chromosome condensation in interphase nuclei. Nature *226*, 717–722.

Jones, M.J., and Jallepalli, P.V. (2012). Chromothripsis: chromosomes in crisis. Dev. Cell *23*, 908–917.

Jones, D.T., Jäger, N., Kool, M., Zichner, T., Hutter, B., Sultan, M., Cho, Y.J., Pugh, T.J., Hovestadt, V., Stütz, A.M., et al. (2012). Dissecting the genomic complexity underlying medulloblastoma. Nature *488*, 100–105.

Kim, T.M., Xi, R., Luquette, L.J., Park, R.W., Johnson, M.D., and Park, P.J. (2013). Functional genomic analysis of chromosomal aberrations in a compendium of 8000 cancer genomes. Genome Res. *23*, 217–227.

Kloosterman, W.P., Guryev, V., van Roosmalen, M., Duran, K.J., de Bruijn, E., Bakker, S.C., Letteboer, T., van Nesselrooij, B., Hochstenbach, R., Poot, M., and Cuppen, E. (2011a). Chromothripsis as a mechanism driving complex de novo structural rearrangements in the germline. Hum. Mol. Genet. *20*, 1916–1924.

Kloosterman, W.P., Hoogstraat, M., Paling, O., Tavakoli-Yaraki, M., Renkens, I., Vermaat, J.S., van Roosmalen, M.J., van Lieshout, S., Nijman, I.J., Roessingh, W., et al. (2011b). Chromothripsis is a common mechanism driving genomic rearrangements in primary and metastatic colorectal cancer. Genome Biol. *12*, R103.

Kloosterman, W.P., Tavakoli-Yaraki, M., van Roosmalen, M.J., van Binsbergen, E., Renkens, I., Duran, K., Ballarati, L., Vergult, S., Giardino, D., Hansson, K., et al. (2012). Constitutional chromothripsis rearrangements involve clustered double-stranded DNA breaks and nonhomologous repair mechanisms. Cell Rep. *1*, 648–655.

Knudson, A.G., Jr. (1971). Mutation and cancer: statistical study of retinoblastoma. Proc. Natl. Acad. Sci. USA *68*, 820–823.

Korbel, J.O., Urban, A.E., Affourtit, J.P., Godwin, B., Grubert, F., Simons, J.F., Kim, P.M., Palejev, D., Carriero, N.J., Du, L., et al. (2007). Paired-end mapping reveals extensive structural variation in the human genome. Science *318*, 420–426.

Li, F.P., and Fraumeni, J.F., Jr. (1969). Soft-tissue sarcomas, breast cancer, and other neoplasms. A familial syndrome? Ann. Intern. Med. *71*, 747–752.

Lichter, P., Cremer, T., Borden, J., Manuelidis, L., and Ward, D.C. (1988). Delineation of individual human chromosomes in metaphase and interphase cells by in situ suppression hybridization using recombinant DNA libraries. Hum. Genet. *80*, 224–234.

Liu, P., Erez, A., Nagamani, S.C., Dhar, S.U., Kołodziejska, K.E., Dharmadhikari, A.V., Cooper, M.L., Wiszniewska, J., Zhang, F., Withers, M.A., et al. (2011). Chromosome catastrophes involve replication mechanisms generating complex genomic rearrangements. Cell *146*, 889–903.

Magrangeas, F., Avet-Loiseau, H., Munshi, N.C., and Minvielle, S. (2011). Chromothripsis identifies a rare and aggressive entity among newly diagnosed multiple myeloma patients. Blood *118*, 675–678.

Maher, C.A., and Wilson, R.K. (2012). Chromothripsis and human disease: piecing together the shattering process. Cell *148*, 29–32.

McClintock, B. (1941). The Stability of Broken Ends of Chromosomes in Zea Mays. Genetics *26*, 234–282.

Meyerson, M., and Pellman, D. (2011). Cancer genomes evolve by pulverizing single chromosomes. Cell *144*, 9–10.

Mills, R.E., Walter, K., Stewart, C., Handsaker, R.E., Chen, K., Alkan, C., Abyzov, A., Yoon, S.C., Ye, K., Cheetham, R.K., et al.; 1000 Genomes Project. (2011). Mapping copy number variation by population-scale genome sequencing. Nature *470*, 59–65.

Molenaar, J.J., Koster, J., Zwijnenburg, D.A., van Sluis, P., Valentijn, L.J., van der Ploeg, I., Hamdi, M., van Nes, J., Westerman, B.A., van Arkel, J., et al. (2012). Sequencing of neuroblastoma identifies chromothripsis and defects in neuritogenesis genes. Nature *483*, 589–593.

Nik-Zainal, S., Van Loo, P., Wedge, D.C., Alexandrov, L.B., Greenman, C.D., Lau, K.W., Raine, K., Jones, D., Marshall, J., Ramakrishna, M., et al.; Breast Cancer Working Group of the International Cancer Genome Consortium. (2012). The life history of 21 breast cancers. Cell *149*, 994–1007.

Northcott, P.A., Shih, D.J., Peacock, J., Garzia, L., Morrissy, A.S., Zichner, T., Stütz, A.M., Korshunov, A., Reimand, J., Schumacher, S.E., et al. (2012). Subgroup-specific structural variation across 1,000 medulloblastoma genomes. Nature *488*, 49–56.

Onishi-Seebacher, M., and Korbel, J.O. (2011). Challenges in studying genomic structural variant formation mechanisms: the short-read dilemma and beyond. Bioessays *33*, 840–850.

Rausch, T., Jones, D.T., Zapatka, M., Stütz, A.M., Zichner, T., Weischenfeldt, J., Jäger, N., Remke, M., Shih, D., Northcott, P.A., et al. (2012a). Genome sequencing of pediatric medulloblastoma links catastrophic DNA rearrangements with TP53 mutations. Cell *148*, 59–71.

Rausch, T., Zichner, T., Schlattl, A., Stütz, A.M., Benes, V., and Korbel, J.O. (2012b). DELLY: structural variant discovery by integrated paired-end and split-read analysis. Bioinformatics *28*, i333–i339.

Rudolph, K.L., Millard, M., Bosenberg, M.W., and DePinho, R.A. (2001). Telomere dysfunction and evolution of intestinal carcinoma in mice and humans. Nat. Genet. *28*, 155–159.

Stephens, P.J., Greenman, C.D., Fu, B., Yang, F., Bignell, G.R., Mudie, L.J., Pleasance, E.D., Lau, K.W., Beare, D., Stebbings, L.A., et al. (2011). Massive

genomic rearrangement acquired in a single catastrophic event during cancer development. Cell *144*, 27–40.

Stratton, M.R., Campbell, P.J., and Futreal, P.A. (2009). The cancer genome. Nature *458*, 719–724.

Tubio, J.M., and Estivill, X. (2011). Cancer: When catastrophe strikes a cell. Nature *470*, 476–477.

Wang, J., Mullighan, C.G., Easton, J., Roberts, S., Heatley, S.L., Ma, J., Rusch, M.C., Chen, K., Harris, C.C., Ding, L., et al. (2011). CREST maps somatic structural variation in cancer genomes with base-pair resolution. Nat. Methods *8*, 652–654.

Yates, L.R., and Campbell, P.J. (2012). Evolution of the cancer genome. Nat. Rev. Genet. *13*, 795–806.

Ye, K., Schulz, M.H., Long, Q., Apweiler, R., and Ning, Z. (2009). Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. Bioinformatics *25*, 2865–2871.