Chromothripsis in Human Breast Cancer

Michiel Bolkestein[1], John K.L. Wong[2], Verena Thewes[2,3], Verena Körber[4], Mario Hlevnjak[2,5], Shaymaa Elgaafary[3,6], Markus Schulze[2,5], Felix K.F. Kommoss[7], Hans-Peter Sinn[7], Tobias Anzeneder[8], Steffen Hirsch[9], Frauke Devens[1], Petra Schröter[2], Thomas Höfer[4], Andreas Schneeweiss[3], Peter Lichter[2,6], Marc Zapatka[2], and Aurélie Ernst[1]

ABSTRACT

Chromothripsis is a form of genome instability by which a presumably single catastrophic event generates extensive genomic rearrangements of one or a few chromosomes. Widely assumed to be an early event in tumor development, this phenomenon plays a prominent role in tumor onset. In this study, an analysis of chromothripsis in 252 human breast cancers from two patient cohorts (149 metastatic breast cancers, 63 untreated primary tumors, 29 local relapses, and 11 longitudinal pairs) using whole-genome and whole-exome sequencing reveals that chromothripsis affects a substantial proportion of human breast cancers, with a prevalence over 60% in a cohort of metastatic cases and 25% in a cohort comprising predominantly luminal breast cancers. In the vast majority of cases, multiple chromosomes per tumor were affected, with most chromothriptic events on chromosomes 11

and 17 including, among other significantly altered drivers, *CCND1, ERBB2, CDK12,* and *BRCA1.* Importantly, chromothripsis generated recurrent fusions that drove tumor development. Chromothripsis-related rearrangements were linked with univocal mutational signatures, with clusters of point mutations due to kataegis in close proximity to the genomic breakpoints and with the activation of specific signaling pathways. Analyzing the temporal order of events in tumors with and without chromothripsis as well as longitudinal analysis of chromothriptic patterns in tumor pairs offered important insights into the role of chromothriptic chromosomes in tumor evolution.

Significance: These findings identify chromothripsis as a major driving event in human breast cancer.

Introduction

Chromothripsis is a form of genome instability, whereby one or a few chromosomes are affected by tens to hundreds of clustered DNA rearrangements (1–4). Localized chromosome shattering is considered to occur as a single catastrophic genomic event, followed by inaccurate repair of the resulting fragments. Massive rearrangements generated by this process have been detected across a wide range of tumor types and are associated with poor prognosis in certain entities (5–8). Chromothripsis is believed to promote, and in some cases even cause, cancer development, because it can lead to the simultaneous inactivation of tumor-suppressor genes, formation of oncogenic fusions, and

oncogene amplification (1, 2, 5, 9). Despite initial estimates suggesting a chromothripsis prevalence of 2% to 3% across cancers (2), recent studies showed that chromothripsis likely plays a role in a substantial fraction of human cancers (10–16). The increasing number of sequenced cancer genomes has revealed that previous assessments of chromothripsis prevalence only included the tip of the iceberg, with a massive underestimation of the frequency of occurrence of this phenomenon when using low-resolution methods.

In breast cancer, chromothripsis studies based on whole-genome sequencing data are rare. In 2014, Przybytkowski and colleagues used array comparative genomic hybridization to profile 29 primary tumors from patients with high-risk breast cancer (17). Despite the low resolution of the method, the authors suggested that 41% of high-risk breast cancers may show chromothripsis. Similarly, another study by Chen and colleagues analyzed 42 primary breast cancers on microarrays and described chromothripsis-like patterns in 61% of the cases (18). Based on SNP array data, Li and colleagues identified 15% of cases with hints for chromothriptic events in triple-negative breast cancer (19). In radiation-induced breast carcinomas, Biermann and colleagues reported chromothripsis-like patterns in 9 of 31 cases (29%) on the basis of microarray data (20). Whole-genome sequencing analyses performed by Tang and colleagues for 11 cases with matched primary-relapse pairs (21) pointed to individual examples of tumors with chromothripsis-like patterns, without systematic scoring or comprehensive analysis of chromothripsis. Vasmatzis and colleagues reported chromoanasynthesis as a common mechanism leading to *ERBB2* amplification in early stage HER2-positive breast cancer (22). From 18 analyzed tumors, 15 showed chromoanasynthesis, a form of genome instability that many authors classify as (noncanonical) chromothripsis, due to common features between both processes. The local rearrangements arising from chromoanasynthesis exhibit altered copy numbers due to serial microhomology-mediated template switching during DNA replication (23–25). Resynthesis of fragments from one chromatid and frequent insertions of short sequences

[1]Group Genome Instability in Tumors, DKFZ, Heidelberg, Germany. [2]Division of Molecular Genetics, DKFZ; DKFZ-Heidelberg Center for Personalized Oncology (HIPO) and German Cancer Consortium (DKTK), Heidelberg, Germany. [3]National Center for Tumor Diseases (NCT), University Hospital and DKFZ, Heidelberg, Germany. [4]Division of Theoretical Systems Biology, DKFZ, Heidelberg, Germany. [5]Computational Oncology Group, Molecular Diagnostics Program at the National Center for Tumor Diseases (NCT) and DKFZ, Heidelberg, Germany. [6]Molecular Diagnostics Program at the National Center for Tumor Diseases (NCT) and DKFZ, Heidelberg, Germany. [7]Institute of Pathology, Heidelberg University Hospital, Heidelberg, Germany. [8]Patients' Tumor Bank of Hope (PATH), Munich, Germany. [9]Institute of Human Genetics, Heidelberg University, Heidelberg, Germany.

between the rearrangement junctions are associated with copy-number gains and retention of heterozygosity. As classical/canonical chromothripsis and chromoanasynthesis/noncanonical chromothripsis generate similar oscillating patterns, most studies do not distinguish between these two types of events.

Despite discrepancies between studies due to different methodologies, distinct scoring criteria for chromothripsis, heterogeneous breast cancer subtypes, and relatively small cohorts for most of the above-mentioned studies, these data suggest that chromothripsis may be of unrecognized importance in breast cancer. Therefore, we analyzed chromothriptic patterns and genomic features associated with chromothripsis in 252 patients with breast cancer.

Materials and Methods

Study design and participants

The whole-genome and whole-exome sequencing data were generated within the CATCH and DKFZ-HIPO17 studies. The CATCH trial is a registry trial and analytical platform for prospective, omics-driven stratification of advanced-stage breast cancer. Tumor tissue and matched normal control sample for sequencing (from the patient's whole blood or healthy breast tissue) were obtained after receiving a written informed consent under an Institutional Review Board–approved protocol. The goal of the DKFZ-HIPO17 study is the identification of novel target genes in patients with breast cancer, the development of sequencing-based prognostic and predictive profiles, and their transfer into clinics. This study includes a majority of untreated primary tumors but also local relapses pretreated with endocrine therapy (see Supplementary Table S1 for details on the patient cohorts).

Genome alignment and variant calling for whole-genome sequencing data

Whole-genome sequencing data and whole-exome sequencing data were processed by the DKFZ OTP pipeline (26). The pipeline used BWA-MEM (v0.7.15) for alignment, biobambam (https://github.com/gt1/biobambam) for sorting, and sambamba for duplication marking. The tumor-germline paired alignments were then used by DKFZ indel single-nucleotide variant (SNV) callers for indel and SNV discovery, as described previously (27).

Transcriptome sequencing data processing

Transcriptome sequencing data were processed by the DKFZ OTP pipeline (26). The pipeline used STAR (v2.5.2b; ref. 28) for alignment, biobambam (https://github.com/gt1/biobambam) for sorting, and sambamba for duplication marking. Read counts per gene were summarized by featureCounts (v1.5.1; ref. 29).

Structural variants and copy-number calling from whole-genome sequencing data

We performed copy-number analysis and structural variant calling from whole-genome sequencing data. Two structural variant callers, SvABA v134 (30) and SOPHIA v1.2.16 (https://bitbucket.org/utoprak/sophia/src), were used. SvABA is a structural variant caller based on assembly and discordant read–based approach. SOPHIA is a structural variant caller based on supplementary alignment approach. SOPHIA is integrated in the DKFZ OTP pipeline, where the output was used in combination with alignment files for ploidy estimation and copy-number calling using ACEseq v1.2.8 (31). SvABA outputs were used for structural variant calling for the analysis of microhomologies at the breakpoints. Ploidy estimation was provided by ACEseq.

Copy-number analysis from whole-exome sequencing data

Copy-number analysis from whole-exome sequencing data was performed by EXCAVATOR2 (32), which allows hybrid bin size on captured regions and off-target regions (reads available from the sequencing data but not located in exonic regions). All regions were used for copy-number segmentation and copy-number calling.

Inference of chromothripsis by visual scoring

For visual evaluation of chromothripsis status, the number of switches between copy-number states was counted for each chromosome. Chromosomes containing 10 or more such switches within 50 Mb were scored as chromothripsis-positive with high confidence. Chromosomes with 8 to 9 or 6 to 7 switches within 50 Mb were scored as chromothripsis-positive with intermediate and low confidence, respectively. Within identified chromothripsis-positive regions, the number of distinct copy-number states was counted.

Inference of chromothripsis by algorithm-based scoring

In silico chromothripsis scoring was performed by Shatterseek (12). Copy-number variants from ACEseq (https://github.com/DKFZ-ODCF/ACEseqWorkflow) and structural variants from SOPHIA were used as input. We applied the same criteria as previous studies to define a positive call (33). Only whole-genome sequenced tumors were scored by *in silico* method.

Quantification of indel signatures and indel calling

Indels were called by two software tools, platypus (34) and Mutect2 (35). Mutect2 from GATK v.4.1.2.0 was used. A panel of nontumor sequences from 69 blood samples was provided to Mutect2 to filter technical artifacts. Indels with PASS quality score after running FilterMutectCalls were used. Confidence scores of platypus Indels calls were provided by the DKFZ OTP pipeline (26). Indels with confidence scores from 8 to 10 were used. The filtered outputs of the two tools were intersected to produce a combined set of high-confidence Indels. The combined set was further filtered by a blacklist of artifact Indels from platypus. The filtered output was converted into 83 Indel subclasses by the PCAWG signature preparation tool (36). Finally, the Indel exposures were estimated by sigProfiler (v2.5.1; ref. 36) for each tumor by Indel signatures defined according to the COSMIC signatures V3 (36).

Analyses of SNV mutational signatures

Quality filtered somatic SNVs were used as input for sigProfiler (v2.5.1; ref. 36) to perform a mutational signature analysis and retrieve the exposure of 45 SNV signatures from COSMIC signatures V3. Signatures having less than 5 cases with positive exposures were excluded from the statistical analysis. Cosine similarity of signature exposures was used to measure the relationships across tumors.

Identification of fusion genes

Fusion genes from the RNA sequencing (RNA-seq) data were identified by Arriba (Arriba: Fast and accurate gene fusion detection from RNA-seq data, https://github.com/suhrig/arriba). Candidate fusions from medium and high confidence cases were further validated by analyzing structural variants from the whole-genome sequencing identified by SOPHIA. These structural variants called by SOPHIA within 200 kb of fusion calls were combined into a high confidence set. We performed a regression analysis to compare the number of fusions per breakpoint in tumors with chromothripsis as compared with tumors without chromothripsis (see **Fig. 2**; Supplementary Fig. S3; Supplementary Table S5).

Significance of chromothriptic events per chromosome

We evaluated the likelihood of the observed number of chromothriptic events per chromosome (see Supplementary Table S3). Random and nonoverlapping regions were sampled from chromosome 1 to chromosome X. Size of the resampled regions is identical to the size of the chromothriptic regions per tumor. Resampling was performed 50,000 times, evaluating the number of random samples per chromosome. The total number of successes is counted as the peak number of events per chromosome exceeding or equal to the observed peak of chromothriptic events.

Microhomologies at the breakpoints and DNA repair processes

Structural variants were called by SvABA (30), an assembly and discordant read–based approach for structural variants discovery. The HOMO field was retrieved for each structural variant called by the assembly method of SvABA. To estimate the contribution of different homology sizes, the structural variants with homology information were binned for analysis and visualization. There are 5 bins for homology usage: blunt end to 1 bp, 2 bp, 3–5 bp, 6–9 bp, and >10 bp. The proportions of each bin were normalized by the total number of structural variants, where significance was assessed by beta-regression.

Pathogenic germline variants in cancer predisposition genes

SNVs and indels were called in the tumor sample and subsequently annotated as germline variants in case they were detected in the control sample derived from the patient's whole blood or normal breast tissue. Rare germline SNVs and indels in a list of cancer predisposition genes were filtered and assessed according to the AMP-ACMG guidelines.

Differential gene expression analysis

Differential gene expression analyses were performed independently on DKFZ-HIPO17 and CATCH due to different library preparation protocols. The analyses were performed on groups of tumors stratified by their breast cancer subtype and the site of metastasis for metastatic tumors. Groups with less than 20 tumors were excluded. The analyses were performed by DESeq2 (37), contrasting chromothripsis-positive tumors and chromothripsis-negative tumors. Differentially expressed genes were taken at the significance level of P-adjusted smaller than or equal to 0.05. Gene set enrichment analyses were performed by GSEA (38) using ranks of signed test statistics provided by DESeq2.

Statistical analysis

Statistical analyses and visualizations were performed using R, karyoploteR (39), pheatmap, wesanderson, ComplexHeatmap (40), and ggplot2 (41). For comparison of mutational signatures, Wilcoxon rank-sum test was applied on $\log_2$ absolute exposures for statistical testing. Family-wise correction of P values was performed according to Bonferroni on statistic contrasting mutational signatures, microhomologies, and chromothripsis occurrence per chromosome. False discovery rate of less than 5% was applied on gene-expression data analysis.

Association between polyploidy and chromothripsis

See Supplemental Methods.

Mutation timing

Data preprocessing

Mutation timing was based on variant allele frequencies (VAF) along with combined estimates of tumor cell content and segment-wise copy numbers as determined with ACEseq, excluding sex chromosomes, segments with $<10^7$ bp and, in order to avoid bias due to kataegis, segments with mutation densities above the upper 95% quantile of mutation densities along the genome. To avoid misclassification of mutations as clonal or subclonal due to uncertainties in tumor ploidy and purity, the output of ACEseq was validated by visual inspection and, for one case with unclear ploidy-purity-solution (OE5B_primary), by FISH. Based on this validation, five tumors (46DP_second, 39867J_first, 39867J_second, CZ6A_first, and T6Z1_first) were excluded from the analysis. Furthermore, ploidies and purities were manually corrected for two tumors, B2HF_first (corrected ploidy: 3; corrected tumor cell content: 0.45) and 0E5B_first (corrected ploidy: 2; corrected tumor cell content: 0.38), and copy-number estimates were adjusted accordingly.

Inferring mutation densities of clonal mutations with weighted binomial clustering

Clonal and subclonal mutations were distinguished based on their VAFs as outlined in the following. Measured VAFs vary around their true value, which, for clonal mutations, is expected at

$$f_{\text{clonal}} = \frac{k\rho}{\pi\rho + 2(1-\rho)}, \qquad (A)$$

where k denotes the number of chromosomal copies carrying the mutation, ρ the tumor cell content, and π the copy number at the given locus. On segments with a heterozygous deletion as well as on copy-number–neutral segments without LOH, k will typically be 1; here we neglect the very small probability of mutating the same position on both alleles (infinite sites hypothesis). By contrast, k may take higher integer values on segments gained by (one or multiple rounds of) duplication, depending on whether the mutation was acquired before or after the gain. For simplicity, we here assume that gains are predominantly caused by a single genomic alteration and that mutations therefore lie either on all A-alleles, on all B-alleles, or on a single copy. It follows that $k \in \{1, \alpha, \beta\}$, $\alpha, \beta, \pi \in \mathbb{N}$, where α and β are the number of A- and B-alleles, and $\alpha + \beta = \pi$. Subclonal mutations on a clonally gained segment gain are acquired after the most recent common ancestor (MRCA) and are therefore present on a single copy with VAF $< \frac{\rho}{\pi\rho + 2(1-\rho)}$.

To estimate densities of somatic SNVs (sSNV) that are clonal on distinct copy numbers, measured VAFs were classified into low- and high-order clonal peaks by weighted binomial clustering. Here, low-order clonal peaks arise from sSNVs on single chromosomal copies and high-order clonal peaks by sSNVs on multiple copies. In order to avoid contamination with subclonal mutations at near-clonal VAFs, this classification was restricted to high-confidence clonal sSNVs by requiring VAF $\geq \frac{\rho}{\pi\rho + 2(1-\rho)}$. Consequently, the size of the low-order peak is quantified on its right-hand side only and thus needs to be multiplied by 2 afterward. Due to finite sequencing depth, observed VAFs are expected to be binomially distributed around their true VAF and distinct clonal peaks have relative sizes, which we quantify by weights $w = (w_1, w_\alpha, w_\beta)$. Then, the probability of measuring the i-th sSNV at VAF_i can be computed to yield

$$P(\text{VAF}_i | f, w) = \sum_k B\big(n_{\text{Var,I}}; n_{\text{Var,i}} + n_{\text{Ref,I}}, f_{\text{clonal}}(k)\big) P(k|w), \quad (B)$$

where $B\big(n_{\text{Var,I}}; n_{\text{Var,i}} + n_{\text{Ref,I}}, f_{\text{clonal}}(k)\big)$ is the binomial probability for drawing $n_{\text{Var,i}}$ variant reads at sequencing depth $n_{\text{Var,i}} + n_{\text{Ref,I}}$ from the k-th order clonal peak, and $P(k|w)$ is the relative size of this

peak, according to the weights w. The posterior probability is, using Eq. B, given up to normalization by

$$P(f, w | \{\text{VAF}_i\}_i^N) \propto \prod_{i=1}^{N} P(\text{VAF}_i | f, w) P(w), \qquad \text{(C)}$$

where $P(w)$ is the prior probability for the weights, and N is the total number of SNVs. Using a uniform distribution for $P(w)$, high-confidence clonal mutations were assigned to distinct clonal peaks at $f_{\text{clonal}}(k)$ according to the weights at the maximum a posteriori probability (MAP). The peak size of the low-order clonal peak was multiplied by 2, thus correcting for the conservative selection of high-confidence clonal mutations, which excluded clonal mutations on the left-hand side of the low-order peak. The number of low- and high-order clonal mutations, $n_{k,l}$, according to the weights $\text{MAP}(w_{k,l})$ at MAP on segment l accordingly read

$$\begin{aligned} n_{1,l} &= \text{MAP}(w_{1,l}) N_l \\ n_{\alpha,l} &= \text{MAP}(w_{\alpha,l}) N_l \qquad \text{(D)} \\ n_{\beta,l} &= \text{MAP}(w_{1,l}) N_l, \end{aligned}$$

where N_l is the total number of mutations on segment l.

Timing of earliest and most recent common ancestors

To estimate mutation densities (sSNVs/bp) at the MRCA, we computed the number of clonal mutations (n_l) that would have been acquired on a single genomic copy if no copy-number change had occurred, separately for each genomic segment l (as classified by ACEseq). Specifically, n_l is obtained by adding the number of low-order clonal mutations that had been elevated to higher clonal orders, to the mutational load on each gained segment, and subsequently dividing by the copy number of the segment:

$$n_l = \frac{n_{1,l} + n_{\alpha,l}\alpha + n_{\beta,l}\beta}{\alpha_l + \beta_l}. \qquad \text{(E)}$$

To account for false positive clonal mutations due to incomplete tissue sampling (i.e., mutations that are clonal in the analyzed sample but not in the tumor), n_l was further corrected by comparing primary and relapse samples from the two tumor pairs, for which reliable ploidy-purity estimates were available (B2HF, 3DUGSZ). The fraction of mutations that were classified as high-confidence clonal sSNVs using the aforementioned criteria, but undetected in the relapse sample was taken as the false positive rate, FP. With this correction, the mutation density at the MRCA, $\tilde{m}_{\text{MRCA}}$, was then estimated by dividing the sum of all corrected lower-order clonal sSNVs by the length of the analyzed genomic fraction, $\sum_l g_l$, as $\tilde{m}_{\text{MRCA}} = \frac{\sum_l n_l (1-FP)}{\sum_l g_l}$. Lower and upper 95% confidence bounds for $\tilde{m}_{\text{MRCA}}$ were estimated by bootstrapping the genomic fragments 1,000 times.

Mutation densities at the earliest common ancestor (ECA) were estimated from segments on which higher-order clonal mutations were significantly less frequent than expected at the MRCA, indicative of an earlier origin. Adjusted P values (Holm-corrected for multiple testing, FDR ≤ 0.01) were computed according to a negative binomial distribution, thus accounting for heterogeneous mutation rates along the genome. Because mutation densities at the ECA are computed from higher-order clonal mutations, $\tilde{m}_{\text{ECA}}$ directly results as $\tilde{m}_{\text{ECA}} = \frac{\sum_l n_{\alpha,l} + n_{\beta,l}}{\sum_l g_l}$. Lower and upper 95% confidence bounds were estimated by bootstrapping, as before.

Relative timing of likely driver mutations

Nonsynonymous SNVs and frameshift insertions/deletions in likely driver genes (selected according to IntOGen, v2019.11.12) were classified as subclonal if the probability of observing $\leq n_{\text{Var},i}$ reads according to a binomial distribution with success probability f_{clonal} (1; Eq. A) was smaller than 5% and else as clonal. In regions with LOH or copy-number gains, clonal mutations were further classified as early or late clonal mutations according to their sampling probability computed from binomial distributions with success probabilities $f_{\text{clonal}}(\alpha)$ and $f_{\text{clonal}}(\beta)$, respectively.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request. Sequence data have been deposited at the European Genomephenome Archive (EGA), under accession number EGAS00001004662.

Results

Chromothripsis is a major event in human breast cancer

We analyzed chromothriptic patterns based on paired-end Illumina sequencing data for 252 patients with breast cancer from the CATCH and DKFZ-HIPO17 cohorts, including 171 whole-genome sequences (median coverage $81\times$) and 114 whole-exome sequences ($323\times$). Tumor and matched germline samples (i.e., blood or normal breast tissue) were processed with standardized pipelines to detect copy-number variants and other structural variants, SNVs, short insertions and deletions (indels), and ploidy status. The respective patient cohorts are described in Supplementary Table S1. From these 252 patients, we analyzed 149 advanced metastatic breast cancer samples (CATCH cohort, see **Fig. 1A**), 63 primary tumors, 29 pretreated local relapses, and 11 pairs with two longitudinal tumor samples each (DKFZ-HIPO17 cohort).

To infer chromothripsis in cancer genomes, we applied established criteria (e.g., ≥ 10 changes in copy-number states on an individual chromosome, see Materials and Methods for all details on the scoring procedure and on the inference of the timing between chromothripsis and polyploidization; ref. 33). We distinguished between (i) canonical chromothripsis involving two or three copy-number states and (ii) noncanonical chromothripsis involving more than three copy-number states (**Fig. 1B**). To warrant stringent criteria with respect to the clustering of the breakpoints, we required a minimum of 10 switches in segmental copy number within 50 Mb for high-confidence scoring, as outlined previously (42). We confirmed the accuracy of the chromothripsis inference by comparing visual scoring and algorithm-based scoring, with a validation rate of 84% (percentage of matching scores between both methods, see Supplementary Table S2). This combined scoring confirmed the hallmarks of chromothripsis, such as clustering of breakpoints and randomness of fragment order and orientation, as defined by Korbel and Campbell (33) and as applied in our previous studies (43, 44). In addition to tumors for which chromothripsis was scored with high confidence, we also scored intermediate- and low-confidence chromothriptic events, with 8 to 9 and 6 to 7 oscillations between copy-number states, respectively.

The overall prevalence of chromothripsis was close to 60% (high confidence events), with 65% in the CATCH cohort ($n = 97/149$) and 55% in the subset of DKFZ-HIPO17 cases with available whole-genome sequences ($n = 6/11$ cases with longitudinal sampling, with only the first tumor of each patient counted in the prevalence; **Fig. 1C**).
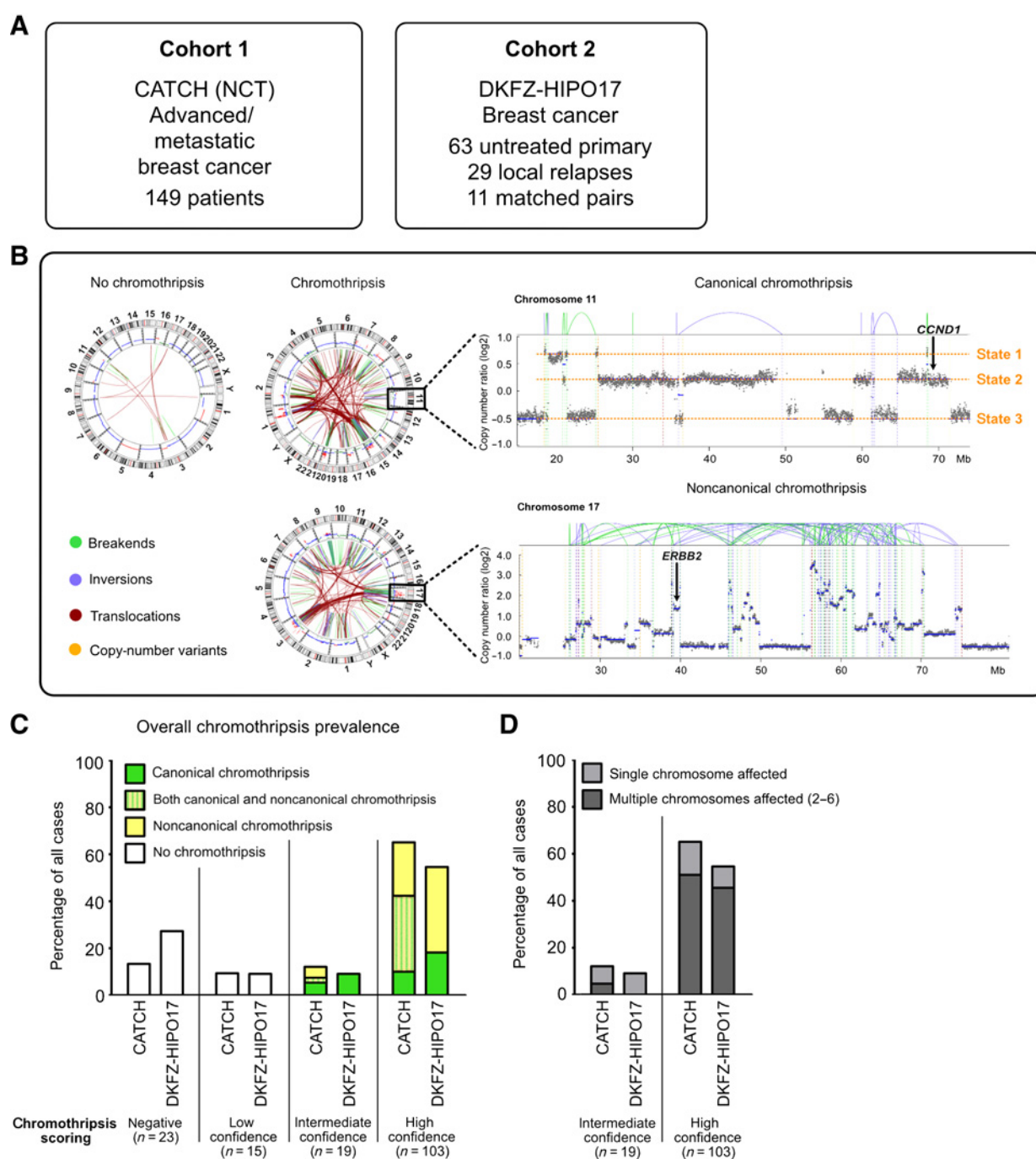
Figure 1.
Chromothripsis is a major driving event in breast cancer. **A,** Overview of cohorts of patients with breast cancer. **B,** Different chromothriptic patterns: canonical chromothripsis (oscillations between two or three copy-number states) and noncanonical chromothripsis (more than three copy-number states). Representative CIRCOS plots and copy-number plots are shown. **C** and **D,** Chromothriptic patterns and prevalence in 160 whole-genome sequences from the CATCH cohort ($n = 149$) and the DKFZ-HIPO17 cohort ($n = 11$). Chromothripsis prevalence (high confidence chromothripsis, 10 or more switches between copy-number states; intermediate confidence, 8 or 9 switches; low confidence, 6 or 7 switches; **C**) and percentage of tumors for which either single or multiple chromosomes show chromothripsis (**D**).

This remarkably high prevalence suggests that chromothripsis is a key driver event in breast cancer. We detected a majority of noncanonical chromothriptic events, with more than two thirds of chromothriptic cases displaying more than three copy-number states on at least one chromothriptic chromosome. In 80% of the tumors, multiple chromosomes were affected by chromothripsis (**Fig. 1D**). Frequent interchromosomal rearrangements between the chromothriptic chromosomes (seen on the CIRCOS plots on **Fig. 1B**) suggested one chromothriptic

event affecting multiple chromosomes, rather than independent chromothriptic events.

In addition, we analyzed 92 cases from the DKFZ-HIPO17 cohort for which whole-exome sequencing was performed (see Materials and Methods for all details on the scoring procedure based on whole-exome sequences). The chromothripsis prevalence in this cohort was 25% (Supplementary Fig. S1A). Scoring for chromothriptic events for 22 DKFZ-HIPO17 cases for which both whole-genome and whole-exome sequences were available showed a very high concordance (77% matching scores, $n = 17/22$). Therefore, the lower chromothripsis prevalence in the DKFZ-HIPO cohort is likely due to the clinical characteristics of these patients, rather than to technical issues (e.g., lower sensitivity of whole-exome sequencing as compared with whole-genome sequencing). As the majority of the tumors from the DKFZ-HIPO17 cohort were from less advanced stages as compared with the CATCH cohort, a lower chromothripsis prevalence in the DKFZ-HIPO cohort is conceivable. In contrast, the high chromothripsis prevalence for the advanced metastatic, highly aggressive tumors from the CATCH cohort goes along with the known association between chromothripsis and poor prognosis in a number of other tumor entities (5, 6, 8). Within the CATCH cohort, the low number of cases without chromothripsis made it challenging to test for links between chromothripsis and clinical outcome (Supplementary Fig. S1B). We observed a slight trend toward younger age both at diagnosis and at metastasis as well as larger metastasis size for cases with chromothripsis (Supplementary Fig. S1C, non-significant). Altogether, at least one out of four breast cancers may be due to chromothripsis.

Chromothripsis generates breast cancer drivers

Chromothripsis promotes cancer development by disrupting tumor-suppressor genes and by activating oncogenes in one catastrophic event (1, 2, 5). We identified distinct chromosomes, chromosome regions, and loci including breast cancer drivers for which significantly more chromothriptic events were detected than expected by chance (permutation test, see **Fig. 2A**; see Supplementary Table S3 for P values associated with the enrichment of specific chromosomes and driver genes). Chromosomes 11 and 17 were significantly enriched for chromothriptic events in both breast cancer cohorts. Driver genes statistically enriched within frequent chromothriptic regions included among others *CCND1* on chromosome 11 and *CDK12, BRCA1*, and *ERBB2* on chromosome 17 (**Fig. 2A**). Interestingly, specific chromosome regions showed statistically more frequent chromothriptic-related rearrangements in only one of the two breast cancer cohorts, possibly reflecting differences in clinical features between the two cohorts. For instance, chromosomes 6 and 12 showed recurrent chromothriptic events specifically in the DKFZ-HIPO17 cohort, whereas chromosomes 8, 19, and 20 were significantly enriched for chromothriptic events in the CATCH cohort (**Fig. 2A**). Subdividing by breast cancer subtype did not show any major difference regarding the chromosome regions affected by chromothripsis, at least between ER-, HER- tumors and ER+, HER- tumors of the CATCH cohort, for which the number of cases was sufficient to address this question (Supplementary Fig. S1D).

We next investigated how chromothripsis affects the copy-number landscape. A subset of the frequent copy-number gains and losses were located in the same chromosome regions in tumors with or without chromothripsis (Supplementary Fig. S2A). This suggests that chromothripsis-independent events lead to copy-number alterations in the same regions as those affected by chromothripsis. Different initial events, chromothripsis-driven or chromothripsis-independent,

lead to the subsequent selection of identical cancer drivers. However, for specific genomic regions, the differences in the proportions of copy-number alterations between tumors with and without chromothripsis were significantly different (Supplementary Fig. S2B; Supplementary Table S4). For instance, gains on chromosome 20q or losses on chromosome 17 (including the *TP53* locus) were significantly more frequent in tumors with chromothripsis ($P < 9.10^{-5}$ and $P < 3.10^{-4}$, respectively), possibly pointing to factors facilitating the chromothriptic event itself or the survival of a clone after such an event.

Importantly, tumors with chromothripsis showed significantly more gene fusions (**Fig. 2B**), as shown by the identification of fusion transcripts from RNA-seq (**Fig. 2C**). To avoid issues arising from the reliability of fusion gene predictions, we focused on gene fusions detected with high confidence and with supporting reads from the matching DNA sequencing data, as outlined previously (42). In addition, we also investigated whether the fusion transcripts were in frame (for potential oncogenic fusions) or not in frame (disruption of tumor-suppressor genes). Regression analysis showed that the increased number of fusions in tumors with chromothripsis was not merely due to the number of structural variants but also to the chromothripsis status itself, with 70% more fusion genes in tumors with chromothripsis for a given number of structural variants (Supplementary Fig. S3; Supplementary Table S5). This finding, consistent with what we showed in other tumor entities (42), may have implications for the search for druggable targets in tumors with chromothripsis, as a number of fusion genes offer druggable events. Beyond the subset of fusions that are druggable and/or used for diagnostic purposes, recurrent fusions generated by chromothripsis or by other processes drive tumor development. Notably, we identified inactivating gene fusions of *NF1*, generated by chromothriptic events (**Fig. 2C**) or by independent rearrangements (Supplementary Table S5). In addition, we detected *ESR1* fusions, which have been described as drivers of endocrine therapy resistance and metastasis in breast cancer (44).

Genomic features of tumors with chromothripsis

Compromised function of essential checkpoints or DNA repair factors has been linked with chromothripsis (1, 43). We asked how the inactivation of *TP53, BRCA1*, or *BRCA2* may relate to chromothripsis in breast cancer. For these essential guardians of genome integrity, we scored pathogenic germline variants, truncating somatic variants and copy-number losses to identify cases with two intact copies and one or two hits, respectively (CATCH cohort, **Fig. 3A**). For *TP53* and for *BRCA1*, the proportion of tumors without any alteration was significantly lower in tumors with chromothripsis as compared with tumors without chromothripsis (Fisher exact test, $P < 0.005$ for *TP53* and $P < 0.05$ for *BRCA1*). Germline mutations in *TP53* are strongly linked with chromothripsis (1), and inactivation of essential checkpoints is thought to be an essential prerequisite for the survival of a clone with chromothripsis. Our cohort did not include any patient with germline mutation in *TP53*, in line with the majority of reported *TP53* mutations in breast cancer being somatic. As chromosome 17 and in particular the *TP53* locus are significantly enriched for chromothriptic events in breast cancer (see **Fig. 2**), loss of one copy of *TP53* by copy-number alteration during the chromothriptic event itself likely plays a major role in the survival of chromothriptic clones in mammary cells. As a second mechanism leading to compromised p53 function, *TP53* mutations are predominantly early and clonal in breast cancer (45), also facilitating the survival of chromothriptic clones in a number of cases by inactivating this essential checkpoint.
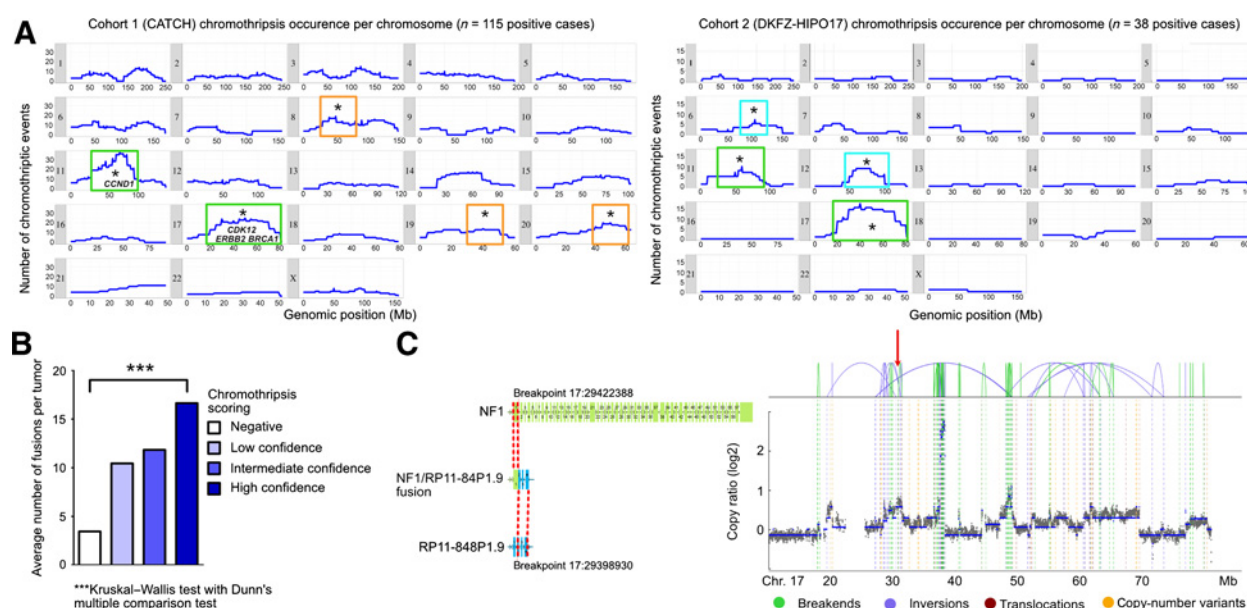
Figure 2.
Chromothripsis generates breast cancer drivers. **A,** Frequency of chromothriptic events on each chromosome for both breast cancer cohorts. The *y*-axis shows the number of chromothriptic events affecting each chromosomal fragment from all high and intermediate confidence chromothriptic cases (*n* = 115 in the CATCH cohort and *n* = 38 in the DKFZ-HIPO17 cohort). Location of known driver genes frequently affected by chromothriptic events is indicated. Stars indicate chromosomes that are significantly enriched for chromothriptic events (permutation test, see Supplementary Table S3). Chromothripsis scoring was done based on whole-genome sequences in the CATCH cohort and whole-exome sequences in the DKFZ-HIPO17 cohort. **B,** Tumors with chromothripsis show significantly more fusion genes. Fusion genes were detected by combining fusion detection from RNA-seq data and structural variant calling from whole-genome sequencing data (CATCH cohort). **C,** One representative example of a disruptive fusion leading to the inactivation of the *NF1* tumor-suppressor gene by a chromothriptic event on chromosome 17. ***, *P* < 0.001.

Distinct DNA repair processes are active in tumors with chromothripsis

After a DNA break, different repair processes, some more error-prone than others, can repair the damage (46). Dissecting the length of the microhomologies at the chromosome breakpoints allows inferring which repair processes were presumably involved in the rejoining of the segments. Blunt ends and short microhomologies (1–2 bp), most common after repair by nonhomologous end joining, as well as microhomologies of 3–5 bp, frequent after alternative end joining, were significantly enriched in tumors with chromothripsis (**Fig. 3B**, left plot). These differences in microhomology length were significant when comparing tumors with versus without chromothripsis (case wise, left plot) but also when comparing breakpoints on chromothriptic chromosomes versus the rest of the genome (region wise, middle plot). Conversely, long homologies (>10 bp) characteristic of repair by homologous recombination were significantly less frequent in tumors with chromothripsis. This supports the link between chromothripsis and homologous recombination deficiency that we reported previously (43) and highlights the role of nonhomologous end joining and alternative end joining in the restitching of chromothriptic chromosomes in breast cancer. Surprisingly, tumors with biallelic inactivation of either *BRCA1* or *BRCA2* were not dramatically different from tumors without compromised *BRCA* in terms of repair patterns. Biallelic mutant tumors showed significantly higher fractions of 2 bp microhomologies at the breakpoints, characteristic of nonhomologous end joining, but only a minor difference with respect to longer homologies, typical of homologous recombination. Due to the low number of tumors with biallelic inactivation of either *BRCA1* or *BRCA2* (*n* = 7), this question is challenging to address in this cohort nevertheless.

To identify mutational processes active in mammary tumors with chromothripsis, we compared the contributions of mutational signatures between tumors with and without chromothripsis-linked rearrangements. COSMIC mutational signatures (36) ID4 and ID9 (both of unknown etiology) as well as SBS2 (linked with APOBEC activity; ref. 47) were significantly more pronounced in tumors with chromothripsis (**Fig. 3C** and **D**). In line with this, de Lange and colleagues reported that chromatin bridges (occasionally leading to chromothripsis) contain extensive single-strand DNA, which represents one of the target substrates for APOBEC enzymes (48). The authors showed in cultured cells that the regions caught up in chromatin bridges harbored clusters of point mutations (48), known as kataegis (49). Interestingly, we observed similar clusters of point mutations in close proximity to the genomic breakpoints (**Fig. 3E**). These mutation clusters were principally in association with chromothripsis-related rearrangements (highlighted by green boxes, **Fig. 3E**), although occasional clusters were also found in proximity of chromothripsis-independent structural variants. Mechanistically, this suggests a role for APOBEC in a subset of chromothripsis-driven tumors.

Signaling pathways active in tumors with chromothripsis

To identify signaling pathways and biological processes linked with chromothripsis, we analyzed differentially expressed genes between tumors with and without chromothripsis. In the CATCH cohort, we restricted this analysis to RNA-seq data of liver metastases and to the subtype of ER$^+$/HER2$^-$ tumors, to exclude any bias due to the tumor site or subtype. Unsupervised clustering analysis showed a strong effect of the chromothripsis status on the clustering in both cohorts (Supplementary Fig. S4A and S4B). Gene set enrichment analysis identified genes involved in SRC, MYC, mTOR, and ATM signaling as
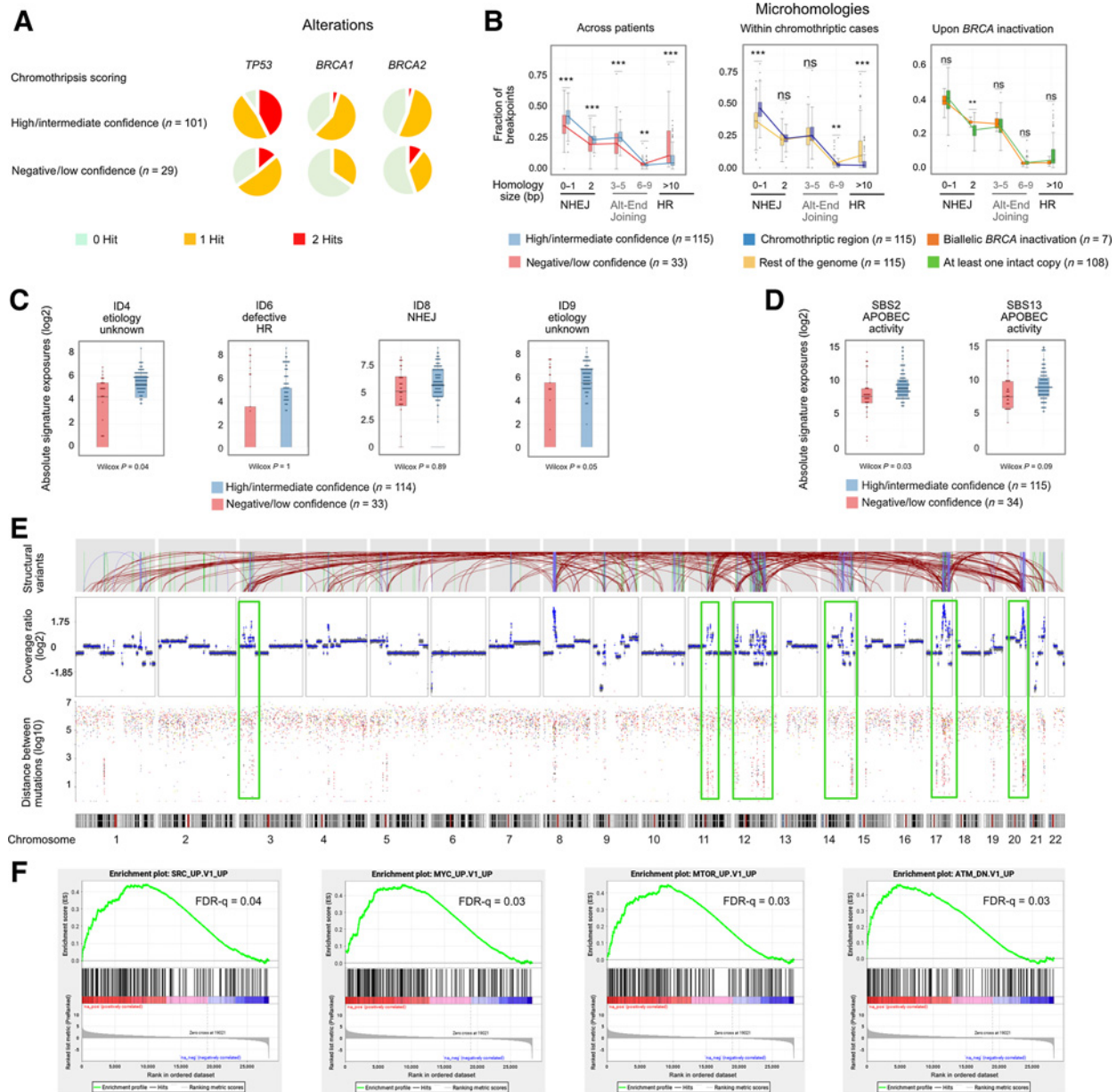
Figure 3.
Genomic features and processes linked with chromothripsis in metastatic breast cancer. **A,** Proportions of alterations in *TP53, BRCA1*, and *BRCA2* in cases with or without chromothripsis in the CATCH cohort for all patients for which evaluation of pathogenic germline variants was available. In addition to pathogenic germline variants, alterations include copy-number loss and truncating somatic variants. **B,** Microhomology at the breakpoints in cases with or without chromothripsis (left; across patients), in chromothriptic regions as compared with the rest of the genome for chromothriptic cases (middle; within chromothriptic cases), and in *BRCA1* or *BRCA2* mutant cases as compared with cases with at least one intact copy (right). **C,** Small insertion and deletion (ID) signature analyses in cases with or without chromothripsis. Wilcoxon tests were applied, and multiple testing correction was performed. **D,** Single base substitution (SBS) mutational signatures in cases with or without chromothripsis. Wilcoxon tests were applied, and multiple testing correction was performed. **E,** Structural variants, copy-number variants, and rainfall plot mapping the intermutational distance showing clusters of mutations (kataegis patterns). Green boxes highlight chromothriptic chromosomes. **F,** Enrichment plots from GSEA conducted with differentially expressed genes between tumors with or without chromothripsis in the CATCH cohort based on RNA-seq data. To exclude tumor site and subtype bias, we restricted the analysis to liver metastases and to ER$^+$/HER2$^-$ tumors. ns, nonsignificant; **, $P < 0.01$; ***, $P < 0.001$.

significantly overrepresented in tumors with chromothripsis (**Fig. 3F**). To detect genes linked with chromothripsis across cohorts, we identified genes that are differentially expressed between tumors with and without chromothripsis both in the ER$^+$/HER2$^-$ liver metastases of the CATCH cohort as well as in the luminal tumors

of the DKFZ-HIPO17 cohort. Among these common differentially expressed genes, the carboxypeptidase B1 (*CPB1*) was strongly enriched in tumors with chromothripsis ($P < 0.002$, Supplementary Table S6). Importantly, overexpression of *CPB1* was suggested as a putative biomarker to identify patients with breast cancer with

low-grade tumors who are at higher than expected risk of recurrence (50). Even though it is not possible to distinguish causative links from correlations, comparative analyses of RNA-seq data in tumors with or without chromothripsis may identify genes and biological processes involved in this form of genome instability.

Longitudinal analysis of chromothriptic patterns

To understand the role of chromothriptic chromosomes in tumor evolution, we analyzed chromothriptic patterns in 11 tumor pairs of the DKFZ-HIPO17 cohort with two longitudinal tumor samples for each patient. Four matched pairs (three primary-relapse pairs and one primary-metastasis pair) showed very stable chromothriptic patterns between the first and the second tumors, with the same major clone at both time points (**Fig. 4A**, left plot, and Supplementary Fig. S5). This was reflected by the high proportion of shared structural variants (visualized by green arcs on the CIRCOS plots), common SNVs, and mutational signatures between both tumor samples for each patient (**Fig. 4B–E**). In such tumors, chromothripsis was likely an early and causative event, providing a strong selective advantage to the resulting clone, which resisted treatment. This also suggests that the surgical resection of the initial tumors was not complete, with the remaining cells giving rise to a local relapse, apparently by monoclonal seeding.

Surprisingly, we also identified seven cases for which the second tumors were newly developed tumors (**Fig. 4A**, right plot, and **E**), genetically independent from the first tumors. The median number of SNVs for these seven cases was 2,773 at the first time point and 1,976 at second time point. None of the SNVs was shared between the first and the second samples for these seven pairs when single-nucleotide polymorphisms were filtered out. Even in the case of a very early clonal divergence, at least one SNV would have been shared between the first and the second tumors. In line with this, the structural variants for these seven cases were also private to each time point (**Fig. 4A**, right plot), which altogether indicates a very good therapeutic management of the first tumor (the initial clone was undetectable at the second time point) and an independently developed second tumor. For these 7 patients, the average time between the first and the second tumors was 4.2 years, as compared with 2.5 years for the 4 patients with bona fide primary-relapse or primary-metastasis pairs. We considered three hypotheses to explain the independent development of two genetically distinct tumors in these 7 patients.

First, we searched for pathogenic variants in germline predisposition genes, as cancer prone syndromes could potentially explain the development of multiple tumors. Only 1 of the 7 patients showed a pathogenic germline variant in *ERCC2* (stop gain mutation, see **Fig. 4C**). However, carriers of heterozygous *ERCC2* mutations do not have a higher cancer risk.

Second, we hypothesized that the second tumors might potentially be induced by the therapy received to treat the first tumors, as chemotherapy and radiotherapy were suggested to induce chromothripsis (20, 51, 52). Behjati and colleagues described mutational signatures of ionizing radiation in second malignancies (53), and in particular a significant excess of deletions relative to insertions in radiation-associated second malignancies. Interestingly, the ratios of genome-wide deletions/insertions were higher in the second tumor for all 7 patients (**Fig. 4D**). As an exception, in 1 of the 4 patients for which the same major clone was present at both time points (OE5B/T2VO), the deletions/insertions ratio was high in both tumors. However, the high ratio already before radiation may be linked with the *ATM* germline mutation of this patient, as germline mutations in DNA repair genes were associated with an excess of deletions (53). Mutational footprints of chemotherapy were also identified (54) and

might potentially play a role in second tumors (e.g., signature SBS37 in the second tumor of patient J63LAV, as this signature is linked with oxaliplatin treatment, see **Fig. 4B**). However, the contribution of chemotherapy-associated mutational signatures is challenging to assess here, due to the different chemotherapy treatments received by these patients. Altogether, radiation and chemotherapy may have played a role in the development of the second tumors, even though we cannot quantify to which extent.

Third, we calculated the probability of developing two independent breast tumors due to bad luck. Based on a breast cancer incidence of 12%, the probability of developing two independent tumors for a woman is 0.0144. Therefore, in a cohort of 100 patients, it is likely to encounter at least 1 patient with two independent tumors. The tumor pairs for this study were collected by specifically searching for longitudinal pairs within a collection of more than 5,000 breast cancer samples. Therefore, the possibility of two independent tumors having developed due to bad luck in these 7 patients is very well conceivable.

Taken together, this cohort is extremely informative with respect to the longitudinal analysis of chromothriptic patterns. All 4 patients with bona fide primary-relapse pairs showed identical chromothriptic patterns in both tumors. This suggests that, in such cases, chromothripsis was an early driver event leading to a major selection advantage and resistance to treatment. In the seven cases with independent tumors, we saw different scenarios, including (i) a first tumor without chromothripsis, but a second tumor with chromothripsis (e.g., patients 45RV, 46DP) or (ii) chromothriptic events on different chromosomes for the two time points (e.g., patients 39867, T6Z1, QYXQ) or (iii) chromothripsis on the same chromosomes but with different patterns (e.g., patients T6Z1, 39867), pointing to independent chromothriptic events affecting the same chromosomes at both time points. Interestingly, from 11 chromothriptic chromosomes detected in the second tumors but not in the first tumors, kataegis patterns appeared on seven of these, together with the chromothripsis-related rearrangements (Supplementary Fig. S5).

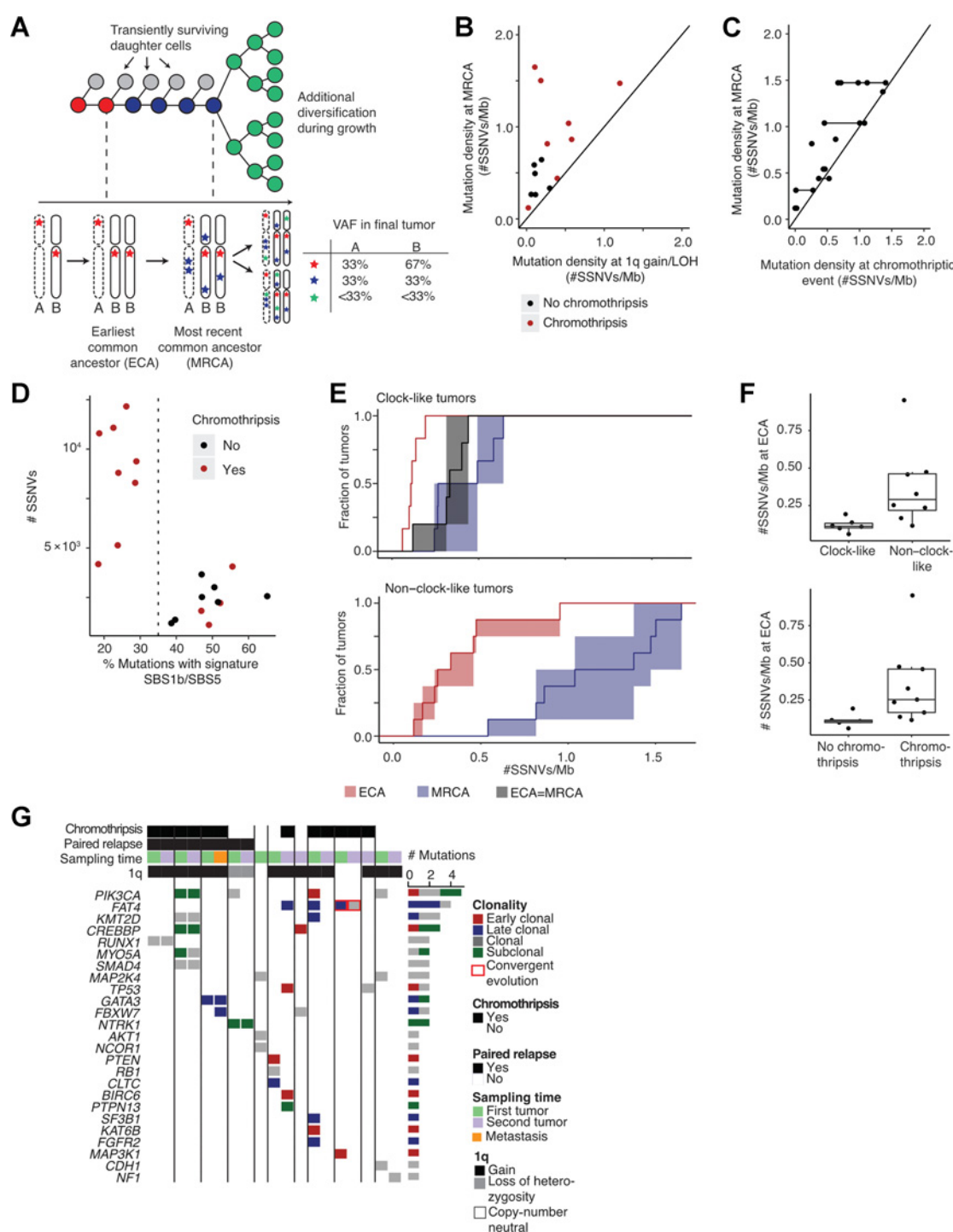### Temporal order of events in tumors with and without chromothripsis

Next, we used the accumulation of mutations during tumorigenesis to time large-scale chromosomal gains, by analyzing mutation densities on amplified and nonamplified alleles separately (45, 55). To this end, we classified mutations on the amplified DNA segment according to their VAFs as (i) early clonal, being present on all copies of a gained allele and thus timing the ECA, (ii) late clonal, being present only on one copy of a gained allele and timing the MRCA, and (iii) subclonal (**Fig. 5A**). 1q gains were coincident with the ECA in 64% of tumors (Supplementary Fig. S6A; sometimes co-occurring with gains in other chromosomes) and generally took place prior to the MRCA in tumors with and without chromothripsis (**Fig. 5B**). Thus, we identify 1q gain, previously reported as a frequent alteration in breast cancer (41), as an early event. To time chromothripsis, we focused on chromothriptic chromosomes with copy numbers ≤4 (as analysis becomes ambiguous at higher copy numbers) and with sufficiently long segments ($>10^7$ bp, allowing reliable estimation of sSNV density). The density of clonal mutations placed the majority of chromothriptic events before the MRCA (**Fig. 5C**). At least in a subset of cases, chromothripsis and 1q gain appear to have occurred in temporal proximity (Supplementary Fig. S6B). In sum, these analyses reveal the early origin of common early copy-number alterations such as 1q gain and rearrangements linked with chromothripsis.

By quantifying overall mutational load, we found that tumors fell into two categories (**Fig. 5D**): one with high proportion of sSNVs with

Figure 4.
Longitudinal analysis of chromothriptic patterns for 11 cases with two tumors. **A,** CIRCOS plots for 11 pairs, with two tumor samples for each patient (subset of the DKFZ-HIPO17 cohort). The left plot shows CIRCOS plots for four patients with bona fide primary-relapse or primary-metastasis pairs. Green lines on CIRCOS plots show structural variants common between the primary and the relapsed tumors. The right plot shows CIRCOS plots for 7 patients with two independent tumors each. Red lines show structural variants private to either the first or the second tumor sample. Orange marks highlight chromosome regions affected by chromothripsis. **B** and **C,** Mutational signatures (single base substitution, SBS, shown in **B**, and small insertion and deletion signatures, ID, shown in **C**) for both tumors for each of the 11 patients. All signatures with a contribution higher than 5% were considered. **D,** Annotation of the chromothripsis status, the ratios of deletions over insertions (high in radiation-induced tumors and in tumors with homologous recombination deficiency; ref. 53), and Venn diagrams showing the proportion of common SNVs shared between the first and the second tumors. **E,** Signature exposure cosine similarity heatmap for 11 tumor pairs. The first heatmap shows the cosine similarity between clonal evolution stages. The annotation stripes (top three rows and left columns) indicate whether the tumor pairs are genetically similar or independent, clonal evolution timing, and tumor IDs of the specimen, respectively. Clonal stages with less than 20 somatic SNVs are not shown. The second heatmap shows the cosine similarity between pairs. The annotation stripes (top three rows and left columns) indicate whether the tumor pairs are genetically similar or independent, whether each tumor is the first or the second tumors and the IDs of the cases, respectively. Matched pairs mean matched primary-relapse pairs for OE5B and DUGSZ but matched primary-metastasis for HWX7 and B2HF. The cosine similarity calculations were performed on 43 COSMIC SBS V3 signatures excluding clock like signatures (SBS1b and SBS5). Normalized signature exposures were estimated by sigProfiler.
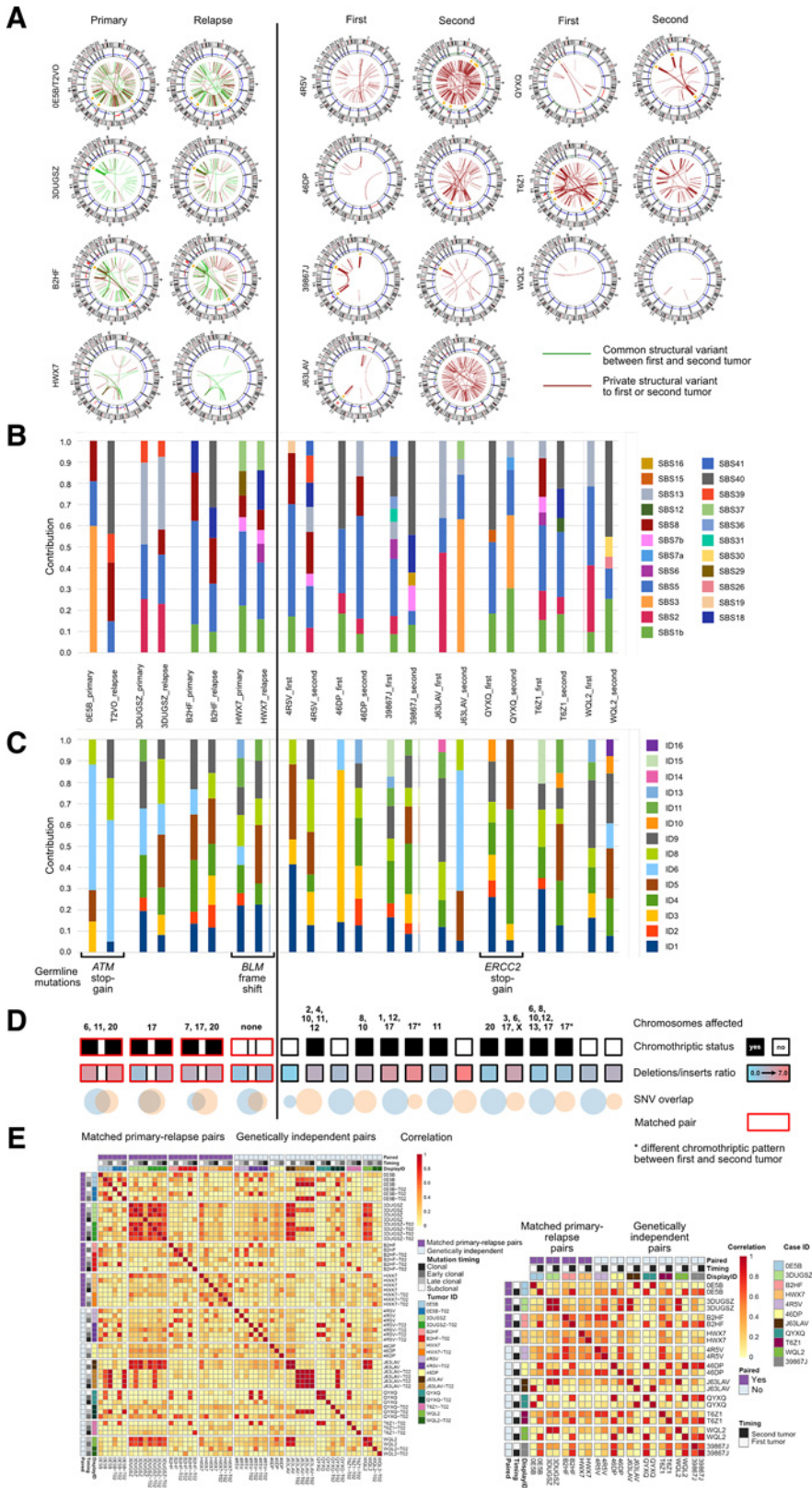
Figure 5.
Temporal order of events in tumors with or without chromothripsis. **A,** Timing copy-number gains using point mutations. Mutations acquired prior to a chromosomal gain are found at 67% VAF. Mutations acquired after a gain are found at 33% VAF if clonal and at VAFs < 33% if subclonal. **B,** Segment-wise timing of copy-number variants via weighted binomial clustering identifies 1q gain as an early event in tumors with or without chromothripsis (tumors with copy numbers > 4 at 1q were excluded from the analysis; points represent mutation densities with MAP). **C,** Segment-wise timing of copy-number variants associated with chromothripsis. Each point corresponds to the mutation density with MAP on one chromothriptic chromosome; horizontal lines combine chromosomes from a single tumor. Shown are data from 9 tumors for which chromothriptic timing was possible. **D,** Mutational burden and proportion of mutations explained by clock-like processes (COSMIC signatures SBS1b/SBS5) in tumors with or without chromothripsis. **E,** Mutational burden at ECAs and MRCA, with lines corresponding to mutation densities at MAP and shaded areas to 95% confidence intervals of estimated mutation densities. **F,** Mutational burden with MAP at tumor onset in clock-like and non–clock-like tumors. **G,** Oncoprint of driver mutations grouped by early clonal, late clonal, clonal, and subclonal mutations. From the 11 tumor pairs shown in **Fig. 4**, only tumors with complete information related to ploidy were used for the analysis of the temporal order of events.

clock-like signatures (SBS1b and SBS5, reflecting mutational processes that may have operated continuously, in a clock-like manner, generating mutations at a steady rate) and low total number of sSNVs ("clock-like tumors"), and the other with opposite characteristics ("non-clock-like tumors"). For most patients (9/11), both tumors fell into the same category (Supplementary Fig. S6C). Notably, all non–clock-like tumors (8/8) had chromothriptic chromosomes, whereas only 4 of 11 clock-like tumors exhibited chromothripsis (**Fig. 5D**). To gain insight into the rate of acquisition of sSNVs, we evaluated the number of sSNVs in the ECA and at the MRCA (**Fig. 5E**). In non–clock-like tumors, mutation densities were already elevated at the tumor's ECA and on average more sSNVs were acquired between ECA and MRCA (**Fig. 5E** and **F**). These data imply that the non–clock-like, and thus the majority of chromothriptic, tumors started with a higher sSNV burden and subsequently had a higher rate of sSNV accumulation. Although mutation signatures in the non–clock-like category did not exhibit a common pattern, three tumors showed enrichment for signature SBS3 associated with a defect in homologous recombination and another three were enriched in signatures related to elevated APOBEC activity (see above). These analyses point to distinct mutational processes acting already early in the majority of tumors with chromothripsis compared with nonchromothriptic tumors.

We then analyzed the enrichment of functional driver gene mutations in paired tumors. We assessed the functionality or driver role of nonsynonymous sSNVs in putative driver genes according to IntOGen (see Materials and Methods for details on driver gene classification). In total, 44 functional driver sSNVs were detected across all patients (**Fig. 5G**), including early or late clonal drivers as well as subclonal drivers. The vast majority of the drivers were clonal, independently of the chromothripsis status. For the four primary-relapse pairs, the same drivers were detected at both time points. This finding is therapeutically relevant, as the regrowth of the major clone after an incomplete tumor eradication (with the same potential targets and no major clonal drift) would provide useful information to consider therapy options. For the 7 patients with independently arising second tumors, we detected different drivers between both time points. In this scenario, the new tumors harbor different genomic profiles, distinct therapeutic targets, and potentially belong to different breast cancer subgroups as compared with the first tumors. In some cases, second tumors may be clinically diagnosed as relapses, but the possibility of a genetically independent new tumor is important to consider, as it may occur more frequently than currently estimated and goes along with major therapeutic implications.

Discussion

We showed that chromothripsis is a frequent event in breast cancer, with 65% of the tumors from the advanced breast cancer cohort CATCH displaying at least one chromothriptic chromosome. Even in the luminal subtype–enriched DKFZ-HIPO17 cohort, at least one tumor out of four may be due to chromothripsis. The lower chromothripsis prevalence in the DKFZ-HIPO17 cohort, enriched for untreated luminal breast cancer, raises the question whether tumors with chromothripsis in this cohort may be linked with more dismal prognosis for these patients as compared with patients from the same cohort with nonchromothriptic tumors. Among carcinomas, breast cancer belongs to the tumor types with the highest chromothripsis prevalence, with few others such as esophagus and prostate carcinomas reaching a similar range (72% and 56%, respectively; ref. 10). Differences in chromothripsis prevalence between tumor types may be linked to the susceptibility of the cell of origin to catastrophic events,

to the levels of replication stress, the efficiency of the apoptotic response, and of the response to DNA damage.

Chromothripsis is commonly described as an early event in tumor development, with a causative role (1, 2). We showed that in breast cancer, chromothriptic events frequently lead to the inactivation of tumor-suppressor genes and to the activation of oncogenes, with a statistical enrichment for breast cancer drivers in the chromosome regions affected by chromothripsis. This supports the simultaneous, rather than sequential, inactivation of preneoplastic genetic drivers leading to clonal expansion. As a subset of chromosomes frequently affected by chromothripsis were specific to each breast cancer cohort, this suggests a selection advantage provided by drivers playing an essential role in one given subtype. Altogether, our findings question the progressive model for these breast cancers and support the data from Gao and colleagues showing that the majority of copy-number aberrations are acquired at the earliest stages of breast cancer evolution (56).

We showed that APOBEC activity and kataegis are linked with chromothripsis in breast cancer. This finding supports data derived from the analysis of cultured cells by de Lange and colleagues (48), as well as observations reported by Park and colleagues in colorectal cancer (57) and by us in leukemia (58). Mechanistically, this association suggests that chromatin bridges containing single-strand breaks processed by APOBEC may lead to mutational clusters at the chromothriptic breakpoints in a subset of tumors with chromothripsis.

We identified signaling pathways significantly linked with chromothripsis in metastatic breast cancer, such as SRC, MYC, MTOR, and ATM signaling, with some of these pathways previously identified by us and by others as linked with chromothripsis in other tumor entities (43, 59). Even though it is unclear at this stage whether these pathways may offer actionable targets and whether causative links with chromothripsis exist, it will be essential to investigate the role of the activation of these signaling pathways in the context of chromothripsis.

The longitudinal analysis of chromothriptic patterns in patients with two tumors revealed a surprising discovery. For 7 out of 11 patients, the second tumors were not relapses of the first tumors, but newly developed genetically independent tumors. Therefore, in a number of breast cancer cases, second tumors may be clinically diagnosed as relapses, but it is important to consider the possibility of genetically independent new tumors, with different molecular features and requiring different therapeutic approaches as compared with the first tumors.

By analyzing chromothriptic patterns based on sequencing data from 252 patients with breast cancer, we identified chromothripsis as a major driver event in breast cancer.

Disclosure of Potential Conflicts of Interest

Authors' Contributions

M. Bolkestein: Formal analysis, investigation. **J.K.L. Wong:** Formal analysis, investigation, methodology. **V. Thewes:** Data curation, writing-review and editing. **V. Körber:** Formal analysis, investigation, writing-review and editing. **M. Hlevnjak:** Formal analysis. **S. Elgaafary:** Data curation, formal analysis. **M. Schulze:** Formal analysis. **F.K.F. Kommoss:** Investigation. **H.-P. Sinn:** Investigation. **T. Anzeneder:**

References

1. Rausch T, Jones DT, Zapatka M, Stutz AM, Zichner T, Weischenfeldt J, et al. Genome sequencing of pediatric medulloblastoma links catastrophic DNA rearrangements with TP53 mutations. Cell 2012;148:59–71.
2. Stephens PJ, Greenman CD, Fu B, Yang F, Bignell GR, Mudie LJ, et al. Massive genomic rearrangement acquired in a single catastrophic event during cancer development. Cell 2011;144:27–40.
3. Crasta K, Ganem NJ, Dagher R, Lantermann AB, Ivanova EV, Pan Y, et al. DNA breaks and chromosome pulverization from errors in mitosis. Nature 2012;482: 53–8.
4. Hatch EM, Hetzer MW. Linking micronuclei to chromosome fragmentation. Cell 2015;161:1502–4.
5. Rode A, Maass KK, Willmund KV, Lichter P, Ernst A. Chromothripsis in cancer cells: an update. Int J Cancer 2016;138:2322–33.
6. Fontana MC, Marconi G, Feenstra JDM, Fonzi E, Papayannidis C, Ghelli Luserna di Rora A, et al. Chromothripsis in acute myeloid leukemia: biological features and impact on survival. Leukemia 2018;32:1609–20.
7. Rucker FG, Dolnik A, Blatte TJ, Teleanu V, Ernst A, Thol F, et al. Chromo-thripsis is linked to TP53 alteration, cell cycle impairment, and dismal outcome in acute myeloid leukemia with complex karyotype. Haematologica 2018;103:e17–e20.
8. Molenaar JJ, Koster J, Zwijnenburg DA, van Sluis P, Valentijn LJ, van der Ploeg I, et al. Sequencing of neuroblastoma identifies chromothripsis and defects in neuritogenesis genes. Nature 2012;483:589–93.
9. Li Y, Schwab C, Ryan SL, Papaemmanuil E, Robinson HM, Jacobs P, et al. Constitutional and somatic rearrangement of chromosome 21 in acute lym-phoblastic leukaemia. Nature 2014;508:98–102.
10. Cortes-Ciriano I, Lee JJ, Xi R, Jain D, Jung YL, Yang L, et al. Comprehensive analysis of chromothripsis in 2,658 human cancers using whole-genome sequencing. Nat Genet 2020;52:331–41.
11. Cai H, Kumar N, Bagheri HC, von Mering C, Robinson MD, Baudis M. Chromothripsis-like patterns are recurring but heterogeneously distributed features in a survey of 22,347 cancer genome screens. BMC Genomics 2014; 15:82.
12. Cancer Genome Atlas N. Genomic classification of cutaneous melanoma. Cell 2015;161:1681–96.
13. Lee JJ, Park S, Park H, Kim S, Lee J, Lee J, et al. Tracing oncogene rearrangements in the mutational history of lung adenocarcinoma. Cell 2019;177:1842–57.
14. Quigley DA, Dang HX, Zhao SG, Lloyd P, Aggarwal R, Alumkal JJ, et al. Genomic hallmarks and structural variation in metastatic prostate cancer. Cell 2018;174: 758–69.
15. Mitchell TJ, Turajlic S, Rowan A, Nicol D, Farmery JHR, O'Brien T, et al. Timing the landmark events in the evolution of clear cell renal cell cancer: TRACERx renal. Cell 2018;173:611–23.
16. Kloosterman WP, Hoogstraat M, Paling O, Tavakoli-Yaraki M, Renkens I, Vermaat JS, et al. Chromothripsis is a common mechanism driving genomic rearrangements in primary and metastatic colorectal cancer. Genome Biol 2011; 12:R103.
17. Przybytkowski E, Lenkiewicz E, Barrett MT, Klein K, Nabavi S, Greenwood CM, et al. Chromosome-breakage genomic instability and chromothripsis in breast cancer. BMC Genomics 2014;15:579.
18. Chen H, Singh RR, Lu X, Huo L, Yao H, Aldape K, et al. Genome-wide copy number aberrations and HER2 and FGFR1 alterations in primary breast cancer by molecular inversion probe microarray. Oncotarget 2017;8: 10845–57.
19. Li Z, Zhang X, Hou C, Zhou Y, Chen J, Cai H, et al. Comprehensive identification and characterization of somatic copy number alterations in triplenegative breast cancer. Int J Oncol 2020;56:522–30.
20. Biermann J, Langen B, Nemes S, Holmberg E, Parris TZ, Werner Ronnerman E, et al. Radiation-induced genomic instability in breast carcinomas of the Swedish hemangioma cohort. Genes Chromosomes Cancer 2019;58:627–35.
21. Tang MH, Dahlgren M, Brueffer C, Tjitrowirjo T, Winter C, Chen Y, et al. Remarkable similarities of chromosomal rearrangements between primary human breast cancers and matched distant metastases as revealed by whole-genome sequencing. Oncotarget 2015;6:37169–84.
22. Vasmatzis G, Wang X, Smadbeck JB, Murphy SJ, Geiersbach KB, Johnson SH, et al. Chromoanasynthesis is a common mechanism that leads to ERBB2 amplifications in a cohort of early stage HER2(+) breast cancer samples. BMC Cancer 2018;18:738.
23. Holland AJ, Cleveland DW. Chromoanagenesis and cancer: mechanisms and consequences of localized, complex chromosomal rearrangements. Nat Med 2012;18:1630–8.
24. Zhang CZ, Leibowitz ML, Pellman D. Chromothripsis and beyond: rapid genome evolution from complex chromosomal rearrangements. Genes Dev 2013;27:2513–30.
25. Pellestor F, Gatinois V. Chromoanasynthesis: another way for the formation of complex chromosomal abnormalities in human reproduction. Hum Reprod 2018;33:1381–7.
26. Reisinger E, Genthner L, Kerssemakers J, Kensche P, Borufka S, Jugold A, et al. OTP: an automatized system for managing and processing NGS data. J Biotechnol 2017;261:53–62.
27. Jones DT, Hutter B, Jager N, Korshunov A, Kool M, Warnatz HJ, et al. Recurrent somatic alterations of FGFR1 and NTRK2 in pilocytic astrocytoma. Nat Genet 2013;45:927–32.
28. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 2013;29:15–21.
29. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 2014;30: 923–30.
30. Wala JA, Bandopadhayay P, Greenwald NF, O'Rourke R, Sharpe T, Stewart C, et al. SvABA: genome-wide detection of structural variants and indels by local assembly. Genome Res 2018;28:581–91.
31. Kleinheinz K, Bludau I, Hübschmann D, Heinold M, Kensche P, Gu Z, et al. ACEseq – allele specific copy number estimation from whole genome sequenc-ing. BioRXiv 2017:210807. DOI: https://doi.org/10.1101/210807.
32. D'Aurizio R, Pippucci T, Tattini L, Giusti B, Pellegrini M, Magi A. Enhanced copy number variants detection from whole-exome sequencing data using EXCAVATOR2. Nucleic Acids Res 2016;44:e154.
33. Korbel JO, Campbell PJ. Criteria for inference of chromothripsis in cancer genomes. Cell 2013;152:1226–36.
34. Rimmer A, Phan H, Mathieson I, Iqbal Z, Twigg SRF, Wilkie AOM, et al. Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. Nat Genet 2014;46:912–8.
35. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol 2013;31:213–9.
36. Alexandrov LB, Kim J, Haradhvala NJ, Huang MN, Tian Ng AW, Wu Y, et al. The repertoire of mutational signatures in human cancer. Nature 2020; 578:94–101.

37. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 2014;15:550.

38. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005;102:15545–50.

39. Gel B, Serra E. karyoploteR: an R/Bioconductor package to plot customizable genomes displaying arbitrary data. Bioinformatics 2017;33:3088–90.

40. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics 2016;32:2847–9.

41. Ginestet C. ggplot2: elegant graphics for data analysis. J R Stat Soc A Stat 2011; 174:245.

42. Voronina N, Wong JKL, Hubschmann D, Hlevnjak M, Uhrig S, Heilig CE, et al. The landscape of chromothripsis across adult cancer types. Nat Commun 2020; 11:2320.

43. Ratnaparkhe M, Wong JKL, Wei PC, Hlevnjak M, Kolb T, Simovic M, et al. Defective DNA damage repair leads to frequent catastrophic genomic events in murine and human tumors. Nat Commun 2018;9:4760.

44. Lei JT, Gou X, Ellis MJ. ESR1 fusions drive endocrine therapy resistance and metastasis in breast cancer. Mol Cell Oncol 2018;5:e1526005.

45. Gerstung M, Jolly C, Leshchiner I, Dentro SC, Gonzalez S, Rosebrock D, et al. The evolutionary history of 2,658 cancers. Nature 2020;578:122–8.

46. Schipler A, Iliakis G. DNA double-strand-break complexity levels and their possible contributions to the probability for error-prone processing and repair pathway choice. Nucleic Acids Res 2013;41:7589–605.

47. Nik-Zainal S, Alexandrov LB, Wedge DC, Van Loo P, Greenman CD, Raine K, et al. Mutational processes molding the genomes of 21 breast cancers. Cell 2012; 149:979–93.

48. Maciejowski J, Li Y, Bosco N, Campbell PJ, de Lange T. Chromothripsis and kataegis induced by telomere crisis. Cell 2015;163:1641–54.

49. Nik-Zainal S, Davies H, Staaf J, Ramakrishna M, Glodzik D, Zou X, et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. Nature 2016;534:47–54.

50. Bouchal P, Dvorakova M, Roumeliotis T, Bortlicek Z, Ihnatova I, Prochazkova I, et al. Combined proteomics and transcriptomics identifies carboxypeptidase B1 and nuclear factor kappaB (NF-kappaB) associated proteins as putative biomarkers of metastasis in low grade breast cancer. Mol Cell Proteomics 2015;14: 1814–30.

51. Mardin BR, Drainas AP, Waszak SM, Weischenfeldt J, Isokane M, Stutz AM, et al. A cell-based model system links chromothripsis with hyperploidy. Mol Syst Biol 2015;11:828.

52. Morishita M, Muramatsu T, Suto Y, Hirai M, Konishi T, Hayashi S, et al. Chromothripsis-like chromosomal rearrangements induced by ionizing radiation using proton microbeam irradiation system. Oncotarget 2016;7: 10182–92.

53. Behjati S, Gundem G, Wedge DC, Roberts ND, Tarpey PS, Cooke SL, et al. Mutational signatures of ionizing radiation in second malignancies. Nat Commun 2016;7:12605.

54. Pich O, Muinos F, Lolkema MP, Steeghs N, Gonzalez-Perez A, Lopez-Bigas N. The mutational footprints of cancer therapies. Nat Genet 2019;51:1732–40.

55. Durinck S, Ho C, Wang NJ, Liao W, Jakkula LR, Collisson EA, et al. Temporal dissection of tumorigenesis in primary cancers. Cancer Discov 2011;1: 137–43.

56. Gao R, Davis A, McDonald TO, Sei E, Shi X, Wang Y, et al. Punctuated copy number evolution and clonal stasis in triple-negative breast cancer. Nat Genet 2016;48:1119–30.

57. Yang L, Wang S, Lee JJ, Lee S, Lee E, Shinbrot E, et al. An enhanced genetic model of colorectal cancer progression history. Genome Biol 2019;20:168.

58. Ratnaparkhe M, Hlevnjak M, Kolb T, Jauch A, Maass KK, Devens F, et al. Genomic profiling of acute lymphoblastic leukemia in ataxia telangiectasia patients reveals tight link between ATM mutations and chromothripsis. Leukemia 2017;31:2048–56.

59. Mueller S, Engleitner T, Maresch R, Zukowska M, Lange S, Kaltenbacher T, et al. Evolutionary routes and KRAS dosage define pancreatic cancer phenotypes. Nature 2018;554:62–8.