



Structure, Dynamics, and Impact of Replication Stress–Induced Structural Variants in Hepatocellular Carcinoma

Quentin Bayard, Pierre Cordier, Camille Péneau, Sandrine Imbeaud, Theo Z Hirsch, Victor Renault, Jean-Charles Nault, Jean-Frédéric Blanc, Julien Calderaro, Chantal Desdouets, et al.

► To cite this version:

Quentin Bayard, Pierre Cordier, Camille Péneau, Sandrine Imbeaud, Theo Z Hirsch, et al.. Structure, Dynamics, and Impact of Replication Stress–Induced Structural Variants in Hepatocellular Carcinoma. Cancer Research, 2022, 82 (8), pp.1470-1481. 10.1158/0008-5472.CAN-21-3665 . hal-04148002

HAL Id: hal-04148002

<https://enva.hal.science/hal-04148002>

Submitted on 10 Feb 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Structure, Dynamics, and Impact of Replication Stress-Induced Structural Variants in Hepatocellular Carcinoma

Quentin Bayard¹, Pierre Cordier², Camille Péneau¹, Sandrine Imbeaud¹, Theo Z. Hirsch¹, Victor Renault³, Jean-Charles Nault^{1,4,5}, Jean-Frédéric Blanc^{6,7,8}, Julien Calderaro⁹, Chantal Desdouets², Jessica Zucman-Rossi^{1,10}, and Eric Letouze¹

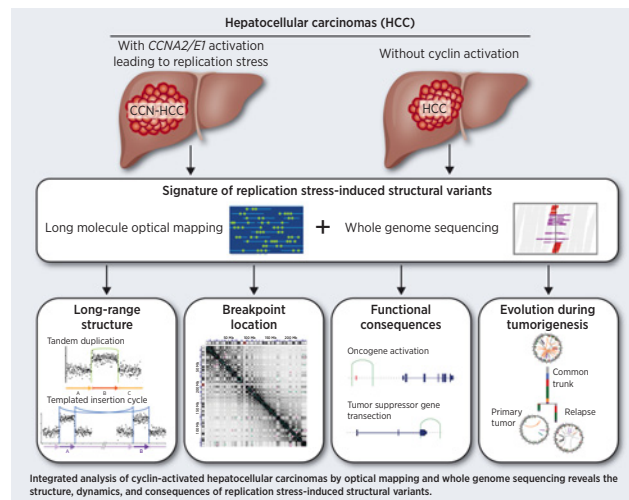


ABSTRACT

Oncogene activation leads to replication stress and promotes genomic instability. Here we combine optical mapping and whole-genome sequencing (WGS) to explore in depth the nature of structural variants (SV) induced by replication stress in cyclin-activated hepatocellular carcinomas (CCN-HCC). In addition to classical tandem duplications, CCN-HCC displayed frequent intra-chromosomal and interchromosomal templated insertion cycles (TIC), likely resulting from template switching events. Template switching preferentially involves active topologically associated domains that are proximal to one another within the 3D genome. Template sizes depend on the type of cyclin activation and are coordinated within each TIC. Replication stress induced continuous accumulation of SVs during CCN-HCC progression, fostering the acquisition of new driver alterations and large-scale copy-number changes at TIC borders. Together, this analysis sheds light on the mechanisms, dynamics, and consequences of SV accumulation in tumors with oncogene-induced replication stress.

Significance: Optical mapping and whole-genome sequencing integration unravels a unique signature of replication stress-

induced structural variants that drive genomic evolution and the acquisition of driver events in CCN-HCC.



¹Centre de Recherche des Cordeliers, Sorbonne Université, Université de Paris, INSERM, Paris, France. ²Laboratory of Proliferation, Stress and Liver Physiopathology, Centre de Recherche des Cordeliers, INSERM, Sorbonne Université, Université de Paris, Paris, France. ³Laboratory for Bioinformatics, Fondation Jean Dausset - CEPH, Paris, France. ⁴Service d'Hépatologie, Hôpital Avicenne, Hôpitaux Universitaires Paris-Seine-Saint-Denis, Assistance-publique Hôpitaux de Paris, Université Sorbonne Paris Nord, Bobigny, France. ⁵Unité de Formation et de Recherche Santé Médecine et Biologie Humaine, Université Paris 13, Communauté d'Universités et Établissements Sorbonne Paris Cité, Paris, France. ⁶Department of Hepato-Gastroenterology and Digestive Oncology, CHU de Bordeaux, Haut-Lévêque Hospital, Bordeaux, Aquitaine, France. ⁷Department of Pathology, CHU de Bordeaux, Pellegrin Hospital, Bordeaux, Aquitaine, France. ⁸Bordeaux Research in Translational Oncology, Université Bordeaux, Bordeaux, Aquitaine, France. ⁹Service d'Anatomopathologie, Hôpital Henri Mondor; Université Paris Est, INSERM U955, Team 18, Institut Mondor de Recherche Biomédicale, Créteil, France. ¹⁰Hôpital Européen Georges Pompidou, Assistance Publique-Hôpitaux de Paris, Paris, France.

Note: Supplementary data for this article are available at Cancer Research Online (<http://cancerres.aacrjournals.org/>).

J. Zucman-Rossi and E. Letouze contributed equally to this article.

Corresponding Authors: Jessica Zucman-Rossi, Centre de Recherche des Cordeliers, Sorbonne Université, Université de Paris, INSERM, Paris F-75006, France. Phone: 336-5072-3700; E-mail: jessica.zucman-rossi@inserm.fr; and Eric Letouze, eric.letouze@inserm.fr

Cancer Res 2022;82:1470-81

doi: 10.1158/0008-5472.CAN-21-3665

©2022 American Association for Cancer Research

Introduction

Structural variants (SV) are an important class of genomic alteration that fuel tumorigenesis by altering the copy number, structure, and regulation of cancer genes. SVs include simple rearrangements such as deletions, tandem duplications, inversions, or interchromosomal translocations, but also complex alterations like chromotripsis (1) or chromoplexy (2), thought to occur during catastrophic chromosomal events. Patterns of SVs are highly heterogeneous across cancers, indicating the existence of various causal mechanisms, more or less active in each tumor (3). Accordingly, SV signatures characterized by a predominance of specific SV categories have been described in several cancers (3-5). Some SV signatures could be attributed to precise molecular etiologies, like the small tandem duplicator phenotype in *BRCA1*-altered breast cancers (4), and the large tandem duplicator phenotype in *CDK12*-altered ovarian (6) and prostate (7) cancers. However, the molecular origins of most SV signatures remain unknown.

Replication stress is a major cause of genomic instability in cancer cells (8). Oncogene activation can induce replication stress by various mechanisms, including premature origin firing and replication-transcription collisions (9). In particular, the overexpression of cyclin E induces precocious S phase entry and firing of unstable ectopic replication origins, leading to replication stress and genomic instability in various cellular models (10-12). Consistently, we recently identified a molecular subgroup of hepatocellular carcinoma (HCC), driven by the activation of cyclin A2 or E1, that displays a massive genomic

instability phenotype (13). In this tumor subgroup, which we named cyclin-activated hepatocellular carcinomas (CCN-HCC), hepatitis B virus (HBV) or adeno-associated virus type 2 (AAV2) insertions, gene fusions or enhancer hijacking activate cyclin A2 or E1, both involved in the G₁–S transition of the cell cycle. CCN-HCC display a transcriptomic signature of replication stress and a characteristic SV signature (named RS1) with hundreds of focal copy-number gains (FCNG), usually between 10 kb and 100 kb, distributed throughout the genome. These focal gains include classical tandem duplications, but also complex events involving several distant regions interconnected by abnormal intra- or interchromosomal junctions, the structure of which could not be resolved by short-read whole-genome sequencing (WGS; ref. 13). CCN-HCC thus constitute a valuable model to explore the complex SVs induced by replication stress.

Here, we aimed at characterizing in depth the SV signature of CCN-HCC using a combination of optical mapping and WGS, exploring the genomic determinants of breakpoint location, and understanding the dynamics and role of this mutational process in tumor evolution.

Materials and Methods

Description of the cohort

A series of 61 HCC samples and their nontumor counterparts were collected from patients surgically treated in four French hospitals located in Bordeaux and Paris regions. The study was approved by Institutional Review Board committees (CCPRB Paris Saint-Louis IRB00003835). Written informed consent was obtained in accordance with French legislation. All samples were immediately frozen in liquid nitrogen and stored at -80°C . Clinical features are summarized in Supplementary Table S1.

γH2AX immunostaining and quantification

Manual IHC for γH2AX was performed on 5- μm -thick formalin-fixed paraffin-embedded HCC tissue sections. After dewaxing in xylene and rehydration in serial concentration ethanol, heat-induced epitope retrieval was performed in citrate-based unmasking buffer for 15 minutes at 95°C with a decloaking chamber. Sections were immersed in 3% H_2O_2 for 15 minutes and nonspecific sites were blocked with 1% BSA, 2% goat serum (GS) in 0.1% Tween 20 Tris-buffered saline (TBST1X) for 30 minutes. Sections were subsequently incubated with anti-phospho-histone H2A.X (γH2AX , Ser139) monoclonal primary antibody (Cell Signaling Technology, catalog no. 9718, rabbit clone 20E3) diluted in 1% BSA 2% GS TBST1X (dilution 1/100). After overnight incubation at 4°C , sections were washed in TBST1X and then incubated in biotinylated goat anti-rabbit secondary antibody (Vector Laboratories, catalog no. BA-1000) for 30 minutes in TBST1X followed by Avidin/Biotin-based Horseradish Peroxidase amplification (ABC-HRP) for 30 minutes according to the manufacturer's instructions (Vector Laboratories, catalog no. PK-6100). The revelation system uses the HRP hydrogen peroxide substrate and 3,3'-diaminobenzidine tetrahydrochloride (DAB) as chromogen (Vector Laboratories, catalog no. SK-4100). Sections were counterstained with Mayer's hematoxylin before mounting. γH2AX -positive neoplastic hepatocyte nuclei were counted on 10 random, nonoverlapping high power fields (HPF, $\times 400$ field) for a total tissue area of 2.37 mm^2 (Nikon Eclipse E600 microscope with Plan 40X/0.65 objective and 10X/22 oculars). Positive nuclei or nuclear pyknotic debris from apoptotic/necrotic neoplastic hepatocytes were not considered. Representative HPFs were captured using a Nikon DS-Ri1 microscope camera.

Optical mapping experiments

Optical mapping of tumor and nontumor liver samples from 4 patients (2 CCN-HCC, 2 other HCC) was performed by Bionano Genomics. Ultra-high molecular weight DNA was extracted using the Animal Tissue DNA Isolation Kit on 30 mg of tissue, according to the manufacturer's protocol. DNA was fluorescently labeled with DLE-1 enzyme at the CTTAAG motif using Bionano reagents. The labeled genomic DNA was linearized in the SaphyrChip using NanoChannel arrays and single molecules were imaged and digitized. As molecules are uniquely identifiable by distinct distribution of sequence motif labels, they were then assembled by pairwise alignment into *de novo* genome maps. We obtained a median genomic coverage of 302X (range 276–324X) of DNA molecules at least 150 kb in length (median = 260 kbp) for each sample.

Somatic SV analysis based on optical mapping

Optical mapping was analyzed using tools included in the Bionano Solve software v3.5. *De novo* assembly of single molecules into consensus genome maps was performed with the RefAligner module. Following alignment of the consensus genome maps on the *in silico* digested (DLE-1 enzyme) hg38 human reference, SVs were identified using the two complementary “De Novo” and “Rare Variant” pipelines. In brief, the “De Novo” pipeline identifies SVs by comparing the consensus genome maps assembled in the sample to the hg38 human reference. In contrast, the “Rare Variant” pipeline directly aligns individual molecules on the reference to capture all molecules corresponding to a given variant. These molecules are then assembled and the resulting local assembly is compared with the reference to generate the final variant call. The Variant Annotation pipeline was used to distinguish germline from somatic SVs. For 3 patients, somatic SVs were identified by comparing SVs in the tumor and matched nontumor samples. For the 4th patient (FR-3798T), the nontumor sample did not pass minimal DNA concentration hence somatic SVs were identified by comparing SV calls in the tumor to Bionano's reference database of 59,490 known variants from 57 physiologically normal individuals. Bionano SV calling pipelines produced intra-pipeline and inter-pipeline duplicated SV calls. To filter out redundancy, we first flagged and removed intra-pipeline duplicate calls. We clustered together SVs using following criteria: (i) for insertions, deletions, duplications and inversions, if two calls are located within 10kp and their size are at least 80% of each other; (ii) for translocations, if two or more calls are located within 10kb and share the same orientation.

Among a single cluster, the SV with the lowest percentage of molecules supporting the alteration in the control sample was kept. Selected SVs from the two pipelines were then clustered using the same approach. Only one alteration was finally kept for each cluster, prioritizing alterations (i) with the lowest percentage of molecules supporting the alteration in the control sample, (ii) with the highest “Confidence” score, and (iii) alterations coming from the “DeNovo” rather than “RareVariant” pipeline.

WGS

Whole-genome data from 61 tumors of the LICA-FR series were analyzed in this study, comprising 49 previously published and 12 new cases (Supplementary Table S1). The whole genomes of the 12 new tumor/normal pairs were sequenced at Integrage (Paris, France) on an Illumina HiSeq X Five as paired-end 150 bp reads, with an average depth of ~ 60 -fold for tumors and ~ 30 -fold for matched nontumor liver samples. Sequences were aligned to the

hg38 version of the human genome using BWA version 0.7.12 (14). We used Picard tools version 1.108 (<http://broadinstitute.github.io/picard/>) to remove PCR duplicates and GATK version v3.5 for local indel realignment and base quality recalibration, as recommended in GATK best practices.

WGS somatic structural variation calling and copy-number analysis

We used MANTA software (15) to identify somatic structural rearrangements in WGS data. To keep only the most reliable events, we selected only SVs with the INFO "PASS" flag, supported by at least 15 reads and with a variant allele fraction $\geq 10\%$. We used Facets R package (16) to reconstruct copy-number profiles from WGS bam files of tumor and paired nontumor liver samples. Single nucleotide polymorphisms (SNP) count matrices for tumor and nontumor samples were obtained by processing the bam files with `snp-pileup` (arguments: `-q15; -Q20; -P100; -r20,0`). We used `preProcSample` function (`cval = 35`, `snp.nbhd = 200`, `ndepth = 20`) to generate log-R-ratio (LRR) and B Allele Frequency data from SNP matrices, `ProcSample` (`cval = 350`, `min.nhet = 5`) to estimate the wild-type 2-copy state and `emcnf` (`min.nhet = 5`) to estimate tumor ploidy, purity, allele specific copy numbers and cellular fraction for each segmented chromosome segment. We distinguished isolated SVs from clustered SVs likely to result from a single chromosomal event. We first identified pairs of SVs sharing breakpoints within 1Mb. SVs were grouped within a same cluster if (i) their breakpoints were located within 5 kb or (ii) their breakpoints defined a segment with higher copy number than surrounding regions.

Integration of optical mapping and WGS data

To reconstruct the precise SV landscape in 4 HCC, we integrated long-range information from optical mapping with base resolution information from WGS data. We first matched SV calls identified from both approaches, with a tolerance of 20kb between the breakpoints to account for the lower resolution of optical mapping. At this stage, we rescued SVs that were found in both data types but filtered out during the post-processing of one of the tables (e.g., due to insufficient coverage). On average, 92% of SVs identified by WGS were also found by optical mapping, and 63% of SVs identified by optical mapping were also found by WGS. We next focused on the 644 highly reliable SVs identified by both approaches (34 in FR-1151T, 152 in FR-1994T, 384 in FR-3798T and 74 in FR-1597T), and we used information from both data types to assign the nature of each SV as follows:

- For SVs with the same simple category (deletion, tandem duplication, inversion, and translocation) in WGS and optical mapping calls, we kept this category and the base resolution breakpoint positions identified by WGS.
- Many SVs were identified as "tandem duplications" by WGS and "insertions" by optical mapping. These events correspond to focal tandem duplications without enough duplicated labels to be identified as such by optical mapping. They were classified "tandem duplications" and the base resolution breakpoints identified by WGS were used.
- Other SVs were complex events involving clustered SVs. After visual inspection, we noticed that many of these corresponded to templated insertion cycles (TIC) as described in Fig. 1. We thus designed a pipeline to automatically annotate and validate TIC, as detailed below. Other complex events not corresponding to TIC

were annotated manually after visual inspection of their SVs, copy-number profiles and consensus optical maps with Bionano Access software.

Identification and validation of TICs

Using WGS data, we annotated as TIC the closed chains of SVs sharing breakpoints within 1Mb, linked by segments with higher copy number than surrounding regions. Higher copy number was defined based on copy-number calls or a LRR threshold defined as follows: $LRR_{threshold} = \max(\text{med}(LRR_L) + \text{sd}(LRR_L), \text{med}(LRR_R) + \text{sd}(LRR_R))$, with LRR_L (resp. LRR_R) the LRR values 100 kb upstream (resp. downstream) the segment. We called "intrachromosomal" TIC involving two segments on a same chromosome, "interchromosomal" TIC involving two segments on different chromosomes, and "complex" TIC involving more than two regions. In the 4 tumors with matched WGS and optical mapping, we used optical mapping data to validate TIC and identify their host chromosomes. We reconstructed the long rearranged sequences resulting from each TIC based on the breakpoint locations and short-read junction orientations. For each event, two hypothetical sequences were reconstructed corresponding to each possible acceptor locus. Rearranged sequences were *in silico* digested with DLE-1 enzyme using `fa2 cmap_multi_color` tool (Bionano Solve software) to create the corresponding references. Bionano molecules were then aligned on these hypothetical references using `RefAligner` module. The number of molecules spanning each hypothetical reference allowed to validate the TIC and define the acceptor locus in 94% of cases, as illustrated in Supplementary Fig. S3.

Hi-C data analysis

We downloaded raw FASTQ files of published normal liver tissue Hi-C data (GEO sample GSM1419086) (17). We used HiC-pro software (18) to generate normalized contact maps with a bin size of 200 kb, `cpro2juicebox` (<https://github.com/nservant/HiC-Pro/blob/master/bin/utis/hicpro2juicebox.sh>) to convert `allValidPair` file to `hic` format, and Juicebox tool (19) for data visualization. To define active and inactive chromatin compartments, we performed an eigenvector analysis of each chromosome's contact matrix using Juicer Tools, with Knight-Ruiz matrix balancing normalization (20) and 50k bins. Compartments with the earliest replication timing (based on ENCODE RepliSeq data in HepG2 cell line) were annotated as active and the others inactive. We then assessed whether TIC preferentially involved regions that are close in the 3D genome organization. To do so, we quantified the proximity of regions associated by TIC as the mean number of valid interaction pairs in HiC-pro results, and we compared these values with an *in silico* background model generated as follows. For each pair of chromosome segments S1 and S2 involved in a TIC, we randomly picked up 100 pairs of segments S1' and S2' meeting the following criteria: (i) S1' and S2' belong to the same type of chromatin compartment (active/inactive) as S1 and S2, respectively; (ii) S1' and S2' have the same length as S1 and S2, respectively; (iii) for intrachromosomal TIC, the distance between S1' and S2' is the same as between S1 and S2; and (iv) for interchromosomal TIC, 50 simulations are generated with S1' = S1 and 50 with S2' = S2.

Finally, we compared the distributions of valid interaction pairs between real TIC and *in silico* simulations using Wilcoxon tests.

Data availability

WGS and optical mapping data generated for this study have been deposited to the European Genome Archive under study accession number (EGAS00001005629).

Results

CCN-HCC undergo replication stress

We previously showed that CCN-HCC overexpress genes of the ATR pathway implicated in the response to replication stress (13). To assess directly the presence of DNA damage induced by replication stress, we performed an immunostaining of γ H2AX in 6 CCN-HCC (3 with *CCNA2* and 3 with *CCNE1* activation) and 4 control HCC without cyclin activation. H2AX phosphorylation was rare in control HCC, mostly limited to occasional scattered positive nuclei (median = 2.5 positive nuclei in 10 HPF). By contrast, CCN-HCC displayed a significant increase in mild to moderate pan-nuclear γ H2AX labeling ($P = 0.019$, Wilcoxon test) with a median of 50 positive nuclei in 10 HPF (Supplementary Fig. S1A and S1B). These data confirm that CCN-HCC undergo replication stress.

Integrated analysis of SVs in HCC by optical mapping and WGS

We combined Bionano optical genome mapping and WGS to reconstruct the long-range structure and base resolution junctions of SVs in 4 HCC, comprising 2 CCN-HCC with HBV insertions activating *CCNE1* (FR-1994T) or *CCNA2* (FR-3798T), and 2 HBV-related HCC without cyclin activation (FR-1151T and FR-1597T), as well as their matched nontumor liver tissues. Bionano optical mapping technology uses fluorescently labeled sequence motifs to generate single-molecule optical maps of long DNA fragments (21). We successfully extracted high molecular weight DNA for every sample except the nontumor counterpart of FR-3798T for which insufficient DNA was available, and we generated between 1,242 Gbp and 1,360 Gbp of raw data per sample considering only molecules >150 kbp. This corresponds to a median genomic coverage of 302X (range 276–324X), with a median molecule size of 260 kbp (Supplementary Table S2). Using two complementary SV detection methods we identified a total of 1,016 somatic SVs from optical mapping data (mean = 254 per sample). Whole genomes were sequenced as paired-end 150 bp reads with a coverage ~60X in tumors and 30X in matched nontumor tissues. Using MANTA, we identified on average 165 somatic SVs per sample from WGS. SVs identified by both approaches were highly consistent: 92% of SVs identified by WGS were also detected with optical mapping, which revealed a median of 1.4 times more SVs than WGS (Supplementary Fig. S2A). Notably, optical mapping identified new SVs validated by copy-number changes, including centromeric events, subclonal alterations, and complex intrachromosomal SVs (Supplementary Fig. S2B), some of which could be rescued in raw WGS SV calls. We next focused on a set of 644 high-quality events identified by both approaches, which we characterized using information from both short-read junctions and optical maps (Supplementary Table S3). The two CCN-HCC displayed a high genomic instability, with on average 202 FCNGs (median size = 25 kb) of low amplitude (1 extra copy), bordered by abnormal intra- or interchromosomal junctions, consistent with the RS1 signature (Fig. 1A). This pattern was not seen in the two other HCC: FR-1151T displayed few SVs and FR-1597T displayed mostly SVs clustered on chromosomes 1 and 5 resulting from a chromotripsis event.

TICs are hallmarks of CCN-HCC

We next explored the mechanisms at the origin of FCNG, characteristic of CCN-HCC. 73% of FCNG corresponded to classical tandem duplications, validated by both WGS-derived abnormal junctions and long-range optical maps (Fig. 1B). 12% of FCNG

involved two segments A and B from different chromosomes, with an extra copy of the B segment inserted between two copies of the A segment, a rearrangement called interchromosomal TIC (Fig. 1C). 6% of FCNG were intrachromosomal TIC involving two distant segments from a same chromosome. These events appeared in short-read WGS as pairs of “deletion” and “duplication” junctions (Fig. 1D), or pairs of “inversion” junctions (Fig. 1E) depending on the orientation of the B segment. Finally, 9% of FCNG corresponded to complex TIC involving more than two segments located on one or several chromosomes. The example shown in Fig. 1F illustrates a complex TIC involving 4 segments (A, B, C, D), with one extra copy of the B, C, and D segments inserted between two copies of segment A. This is to our knowledge the first demonstration of the existence of TIC using long molecule imaging. Overall, 112 TIC were identified in the 2 CCN-HCC whereas these events were rare in the 2 other HCC (2 in FR-1151T, 1 in FR-1597T). We next developed a computational framework to predict TIC from WGS data only, using abnormal junctions and copy number, which we validated in samples with matched optical mapping (Supplementary Fig. S3). We then quantified the diverse types of TIC in 61 HCC genomes including 21 CCN-HCC (Fig. 2A). CCN-HCC displayed a significant increase of all TIC categories as compared with other HCC, with more events in *CCNA2*- than *CCNE1*-activated tumors (Fig. 2B). The burden of tandem duplication and TIC was correlated with a gene expression signature of cell proliferation, an association that remained significant in HCC without cyclin activation ($P = 0.027$). By contrast, these focal SVs were not associated with increased ploidy. Tandem duplication and TIC are thus hallmarks of CCN-HCC and their numbers are correlated (Fig. 2C), suggesting that they result from the same biological mechanism. Indeed, all these SVs are consistent with a replication-based mechanism involving fork stalling followed by one or more template switching events until the replication fork moves back to the original DNA template upstream the initial stalling (Supplementary Fig. S4).

Templated insertions involve active topologically associated domains that are close in the 3D genome organization

We previously showed that RS1 breakpoints are strongly enriched in early-replicated, highly transcribed regions with active chromatin marks, and slightly enriched in repeated regions (13). This might reflect more frequent replication fork collapse in these regions. Here we explored the genomic properties favoring the “landing” of presumed template switching events in specific distant DNA regions. We first examined microhomologies at abnormal junctions. Short microhomologies were identified in 69% of RS1 breakpoints, with a median size of 2 bp, a distribution comparable with that of other SV signatures in HCC (Supplementary Fig. S5). We next hypothesized that template switching may preferentially involve regions that are close in the 3D conformation of the hepatocyte genome. Using published Hi-C data (17), we reconstructed the contact maps in normal adult liver tissue and we identified active and inactive topologically associated domains (TAD). While tandem duplications mostly occurred within a same TAD, most TIC junctions involved regions located in two different active TADs (Fig. 3A). In addition, distant regions bound by TIC junctions displayed significantly more contacts in Hi-C data than expected by chance from the background, both for intra-chromosomal ($P = 8.0 \times 10^{-9}$) and interchromosomal ($P < 2.2 \times 10^{-16}$) TIC (Fig. 3B). Thus, TIC preferentially involve active TADs that can be distant in genomic space but are close in the 3D genome conformation, as nicely illustrated by the TIC junctions on chromosomes 2 and 3 (Fig. 3C).

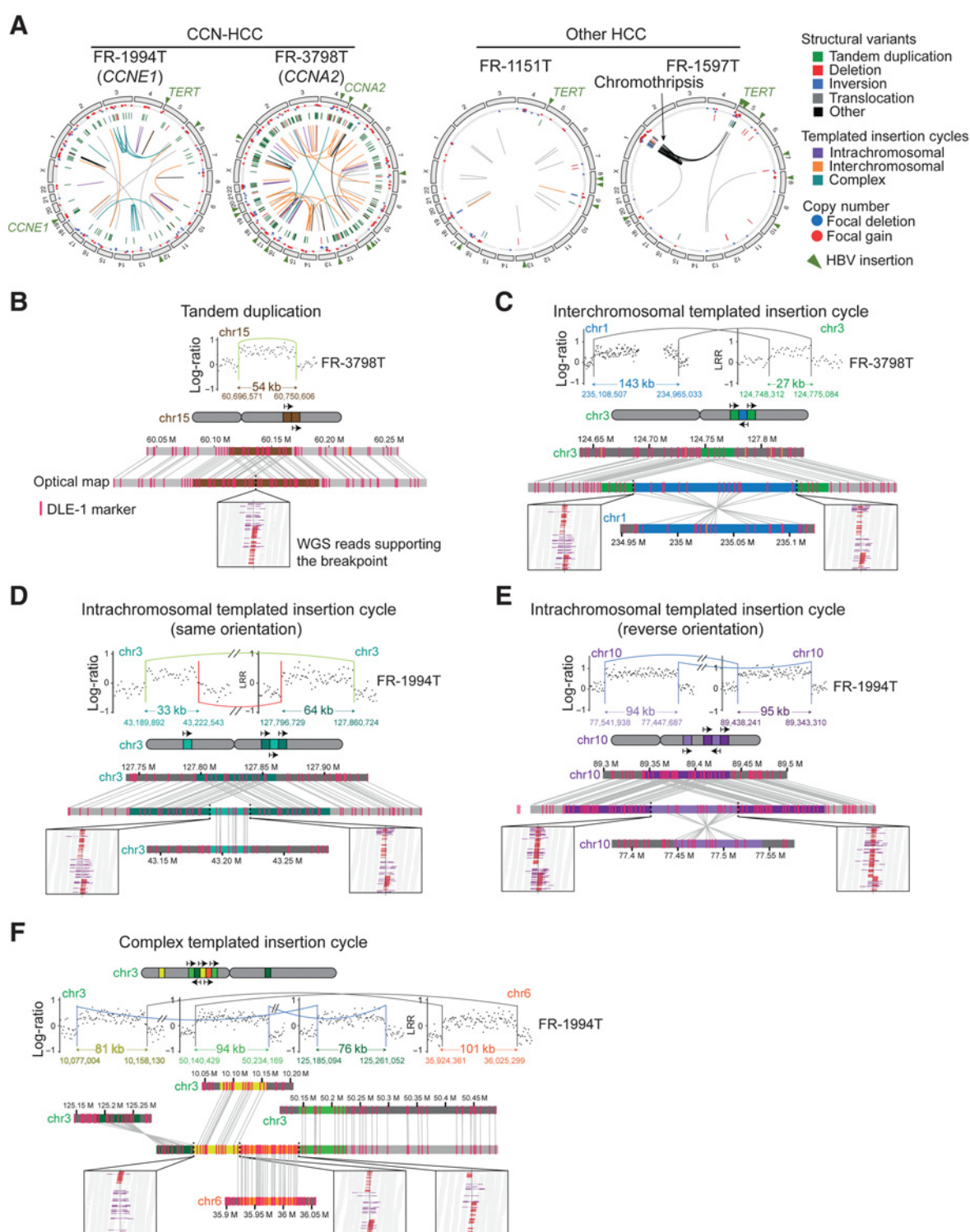


Figure 1.

Optical mapping reveals complex template switching events in CCN-HCC. **A**, CIRCOS plots showing the SV landscapes of four HCC reconstructed from optical mapping and WGS data. Left, CCN-HCC display a massive accumulation of TICs and tandem duplications. **B–F**, Illustrate the reconstruction of the complex structure of replication stress-induced SVs in CCN-HCC. Top panels display abnormal junctions identified in WGS data overlaid on the coverage log-ratio profiles. Bottom panels show the long-range optical maps with the alignment of DLE-1 motifs on reference chromosomes, as well as WGS reads (red) and read pairs (purple) that support the breakpoint site. **B**, Tandem duplication. **C**, Interchromosomal TIC. **D**, Intrachromosomal TIC with inserted template in the same orientation. **E**, Intrachromosomal TIC with inserted template in reverse orientation. **F**, Complex TIC involving four chromosome segments on two different chromosomes.

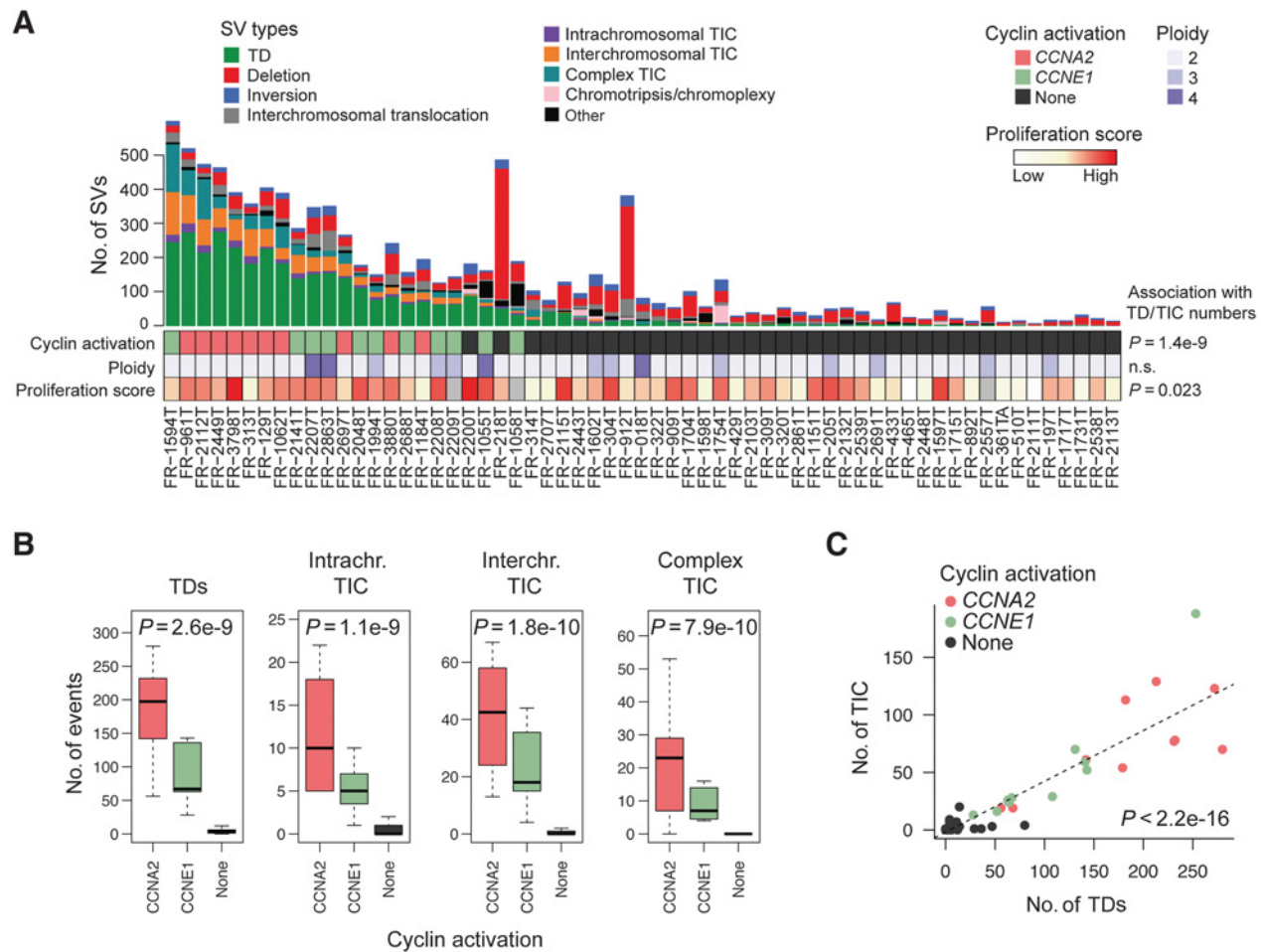


Figure 2.

Quantification of TIC in 61 HCC genomes. **A**, Number of SVs per category in 21 CCN-HCC and 40 other HCC. Samples are ordered by decreasing number of TIC and tandem duplications (TD). **B**, Box-and-whisker plots showing the number of TDs and TIC in HCC with activation of CCNA2, CCNE1, or none. **C**, Correlation between the number of tandem duplications and TIC in 61 HCC.

Template sizes depend on the activated cyclin and are conserved within TIC

The median size of the templates (gained segments) in TIC was 30 kb, slightly larger than the median size of tandem duplications (23 kb). Templates were significantly smaller in CCNA2 (median = 24 kb) than in CCNE1-activated tumors (median = 40 kb, $P = 1.3 \times 10^{-6}$, Fig. 3D). Thus, these two oncogenes induce a similar SV signature but with subtle differences in terms of the number and size of events. Strikingly, we observed a close correlation between template sizes within a given TIC event (Fig. 3E; Supplementary Fig. S6A, $P = 1.0 \times 10^{-45}$). This correlation was confirmed within individual samples, indicating that it does not result from tumor-specific properties (Supplementary Fig. S6B). Rather, it seems that the size of the first template determines the length of DNA that will be replicated at the new location after template switching. In the most frequent case of TIC involving two regions (with an extra copy of segment B inserted between two copies of segment A), this suggests a mechanism in which the replication fork would move back to the original template x kb upstream the initial fork stalling event, with x being determined by the length of segment B.

Functional consequences of replication stress-induced SVs

To explore the functional consequences of replication stress-induced SVs, we mapped all events affecting genes from the cancer gene census (22) and known liver cancer genes (23). Overall, 70 cancer genes were duplicated, 102 were transected, and 220 had their promoter region (50 kb upstream transcription start site) altered by RS1 SVs in this series of 21 CCN-HCC genomes (Fig. 4A). After *TERT*, the most frequently duplicated oncogenes were CCNE1, MYCL and STAT6 (Fig. 4B). MYCL is amplified in several cancers (24) and its paralog MYC is a hotspot of gains and focal amplifications in HCC (23). JAK/STAT is a key pathway activated in 9% of HCC mostly by *IL6ST* mutations or *JAK3* amplifications (23). At CCNE1 locus, RS1 SVs may reinforce the initial activation of CCNE1 by viral insertion. In FR-1994T, a tandem duplication led to duplicate the HBV sequence inserted in CCNE1 promoter (Fig. 4C). Similarly, nested duplications generated multiple copies of AAV2 sequence upstream CCNE1 in FR-2141T. Due to the small size of replication stress-induced SVs, gene transections (where one of the breakpoints is intragenic) were 1.5 times more frequent than duplications. Recurrent transections were identified in the tumor suppressor *RAD51B* but they were

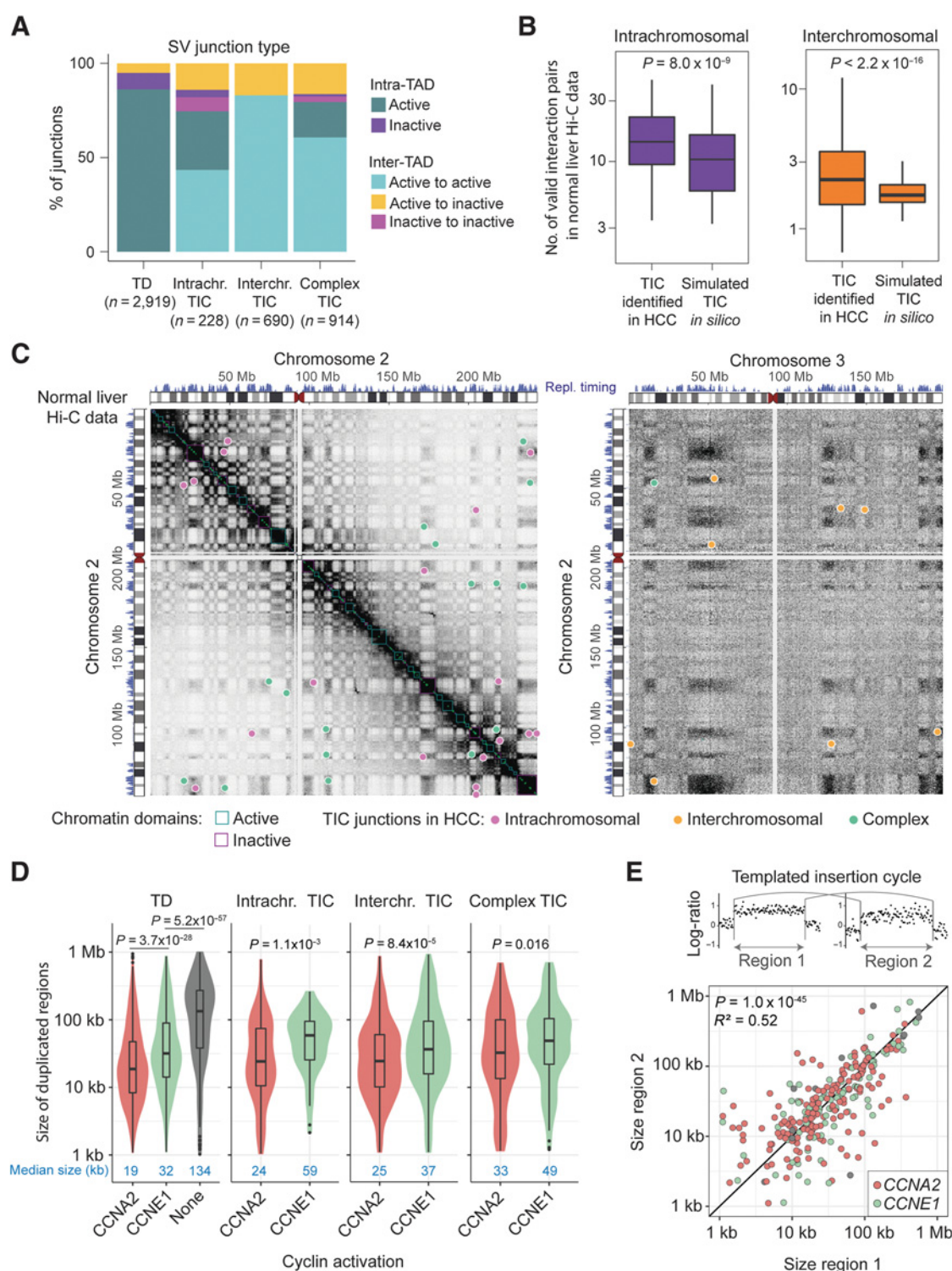


Figure 3.

Determinants of template insertion size and location. **A**, Proportion of tandem duplications (TD) and TIC involving template switching within a same TAD or between two active, two inactive, or one active and one inactive TAD. **B**, Number of contacts between chromosome locations linked by TIC identified in HCC as compared with random TIC simulated *in silico*. **C**, Projection of TIC junctions within the contact maps of two chromosomes derived from Hi-C data of normal liver. Left panel indicates intrachromosomal contacts within chromosome 2. Right panel indicates interchromosomal contacts between chromosomes 2 and 3. Colored dots indicate the location and type of TIC. **D**, Size distribution of tandem duplications (TD) and TIC in tumors with activation of CCNA2, CCNE1, or none. **E**, Correlation between templates sizes within the same TIC event.

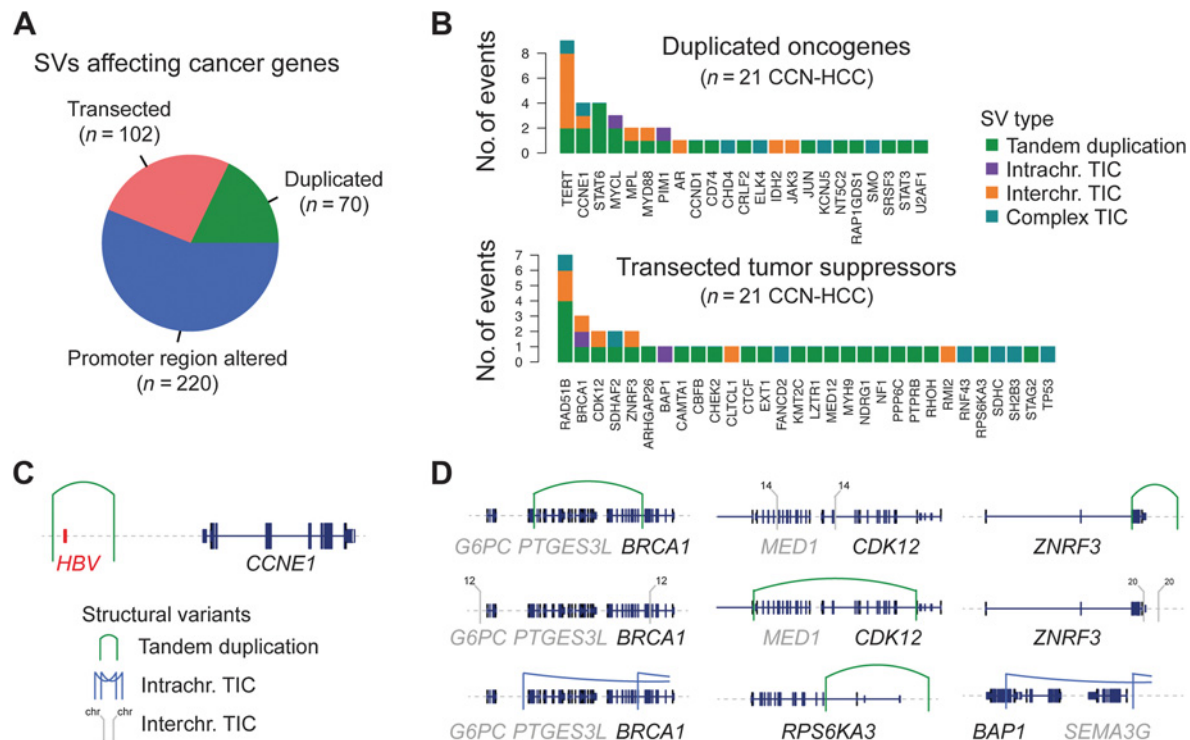


Figure 4.

Functional consequences of replication stress-induced SVs. **A**, Number of SVs leading to the transection, duplication, or promoter alteration of cancer genes across 21 CCN-HCC genomes. Cancer genes include genes from the Cancer Gene Census (tier 1) and known liver cancer genes. **B**, Duplicated oncogenes (top) and transected tumor suppressors (bottom) due to TIC and tandem duplications in CCN-HCC. **C**, Tandem duplication including an HBV insertion in the promoter region of CCNE1 in FR-1994T. **D**, Examples of TIC and tandem duplications transecting tumor suppressor genes in CCN-HCC.

mostly located within a single intron. These events are unlikely to be functional and may correspond to a locus prone to replication fork collapse. By contrast, transections disrupting the coding sequence were identified in DNA repair genes (*BRCA1*, *CDK12*) and well-known liver cancer tumor suppressors (*ZNRF3*, *BAP1*, *RPS6KA3*; **Fig. 4D**). Thus, tandem duplications and TIC induced by replication stress can lead to the activation of oncogenes and disruption of tumor suppressors.

Replication stress is a continuous mutational process during CCN-HCC evolution

To investigate how the SV landscape of CCN-HCC evolves during tumorigenesis, we analyzed multiple samples from 5 patients comprising synchronous CCN-HCC nodules (2 cases) and pairs of primary CCN-HCC and relapses (3 cases). In patient 1602, only the primary tumor displayed *CCNA2* activation. The second tumor, operated 10 years later, was an independent tumor with a completely different genomic landscape and a paucity of SVs. This contrast illustrates within a single patient the massive genomic instability associated with *CCNA2* activation in CCN-HCC. In the 4 other patients, samples had a monoclonal origin and we could reconstruct their molecular evolution (**Fig. 5**). Oncogenic events activating *CCNA2* or *CCNE1* (gene fusion, AAV2 insertion or high-level amplification) were always trunk, consistent with their role of initiating events in CCN-HCC. By contrast replication stress-induced SVs appeared at any stage of tumor progression. In patient 2207 for example, 187 tandem duplications and TIC were common to both samples, 35 were specific to the primary tumor, and 40 were specific to the relapse. Shared and private

replication stress-induced SVs were identified in every patient and involved driver genes like *TERT*, *TP53*, *BRCA1*, and *ZNRF3*. Thus, replication stress is a continuous source of genomic instability, leading to the progressive accumulation of SVs and additional driver events during CCN-HCC progression.

Templated insertions foster large-scale chromosome rearrangements

By copying and pasting long DNA sequences into distant DNA regions, template switching events generate new homologous sequences that may favor the occurrence of secondary chromosome rearrangements. In patient #2208 for example, an interchromosomal TIC between 4p14 and 6p21.1 was followed by large gains of the 6pter→6p21.1 and 4p14→4qter regions in one of the two nodules (**Fig. 6A**). The breakpoints of these gains coincided precisely with the TIC junctions, strongly supporting a link between the two events. By integrating WGS and optical mapping, we reconstructed the long-range structure of two similar events in FR-1994T, with large chromosome gains bordered by interchromosomal TIC involving two (**Fig. 6B**) or three (Supplementary Fig. S7) segments. In both cases, optical maps revealed chromosome chimeras consistent with reciprocal translocations involving TIC regions. These data suggest a two-step process in which (1) an interchromosomal TIC leads to copy-and-paste a long DNA sequence into another chromosome and (2) a reciprocal translocation occurs between the inserted template and its homologous sequence at the original locus (**Fig. 6B**). Subsequent duplication of one of the chromosome chimeras explains the large gains bordered by TIC junctions.

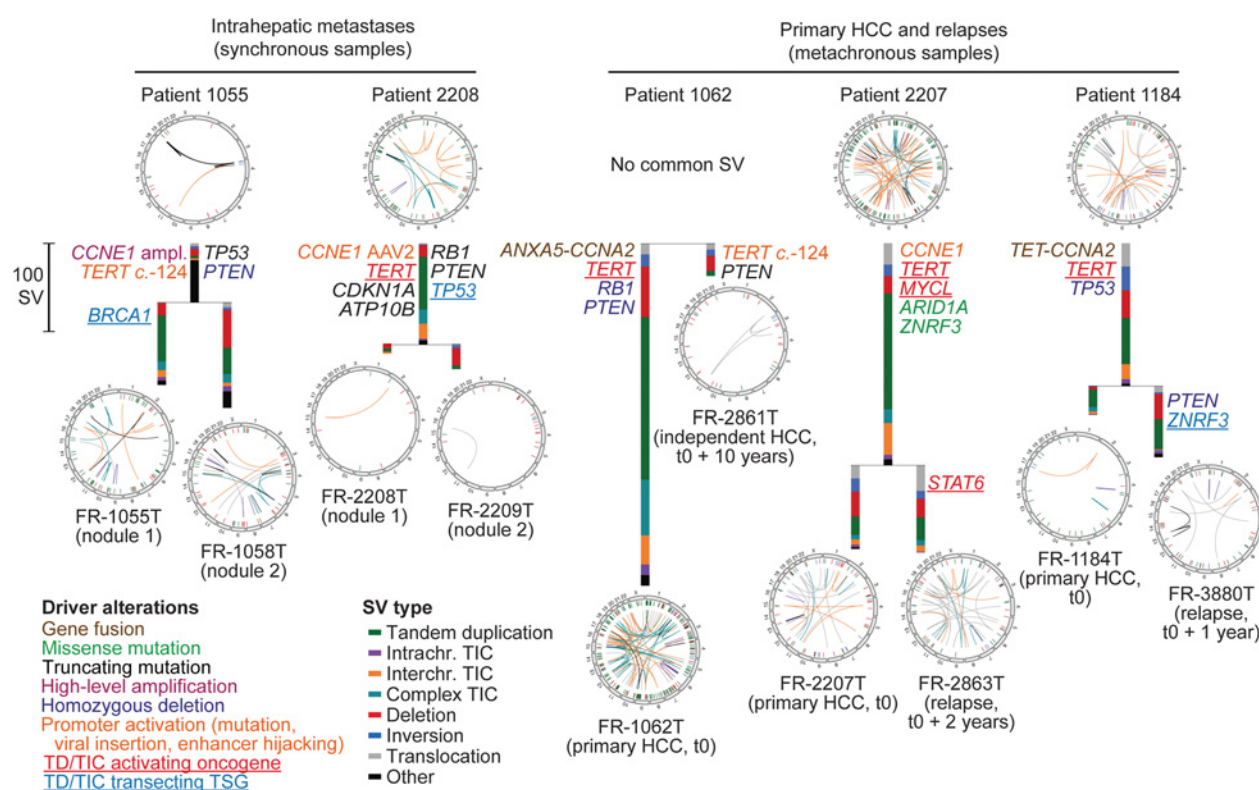


Figure 5.

Natural history of five CCN-HCC. Oncogenetic trees were reconstructed from WGS data of patients with synchronous ($n = 2$) or metachronous ($n = 3$) CCN-HCC samples. The length of each branch is proportional to the number of acquired SVs with a color code indicating their types. Driver alterations are annotated on each branch and colored by alteration type. CIRCOS plots on the top of the trees represent early SVs common to both samples, and those at the bottom indicate late SVs specific to each sample.

Overall, we identified 87 instances of large copy-number changes bordered by TIC in our series of 21 CCN-HCC genomes (mean = 4 by tumor). Other overlapping events in CCN-HCC included nested tandem duplications and TIC (mean = 7 by tumor, **Fig. 6C–E**), which may indicate regions particularly prone to replication fork stalling or a destabilizing effect of prior SVs. Thus, TIC favor the occurrence of secondary alterations including large-scale chromosome rearrangements and copy-number changes.

Discussion

Replication-based mechanisms are emerging as a major cause of SVs. Such mechanisms were initially proposed to explain complex SVs in the germline (25–27). More recently, pan-cancer analysis of whole genomes revealed “copy-and-paste” classes of somatic SVs thought to arise from similar replication-based mechanisms (3). However, the precise structure of these rearrangements could only be predicted from short-read sequencing data, and the link with replication has not been formally established. By combining WGS and optical mapping, we could reconstruct the structure of complex SVs both at large scale and single-base resolution in a subgroup of HCC with cyclin-induced replication stress. We identified a variety of SVs in these tumors ranging from classical tandem duplications to complex TIC with multiple template switching events. Tandem duplication and TIC are correlated both in numbers and size, and are likely generated by the same process, tandem duplication being

simple instances of local insertions of an extra copy of a genomic template. Replication fork collapse followed by repair via break-induced replication (BIR) or microhomology-mediated BIR have been proposed to explain the formation of TIC (26, 28). Consistently, BIR was involved in the formation of genomic duplications in human cells overexpressing cyclin E (11). However, microhomologies were not particularly enriched at RS1 breakpoints. Besides, BIR is a highly inaccurate process, with a mutation rate up to 2,800-fold higher than normal replication (29). In CCN-HCC, we did not observe a higher mutation rate in duplicated templates than in surrounding sequences. Thus, the formation of tandem duplication and TIC upon replication stress in CCN-HCC might result from a different process, such as the replication Fork Stalling and Template Switching mechanism (25). Further mechanistic studies are required to demonstrate that replication stress-induced SVs result from template switching and unravel the precise molecular mechanism at their origin.

The enrichment of RS1 breakpoints in early-replicated active TADs is in agreement with experimental data showing that cyclin E overexpression leads to the firing of unstable replication origins in actively transcribed regions (12). Using Hi-C data, we showed that TIC junctions preferentially occur between active TADs that are close in the 3D organization of the nucleus. Thus, the spatial segregation of active chromatin compartments within the same chromosome territories (30) favors template switching between active TADs. Another striking observation was the strong correlation of template sizes within

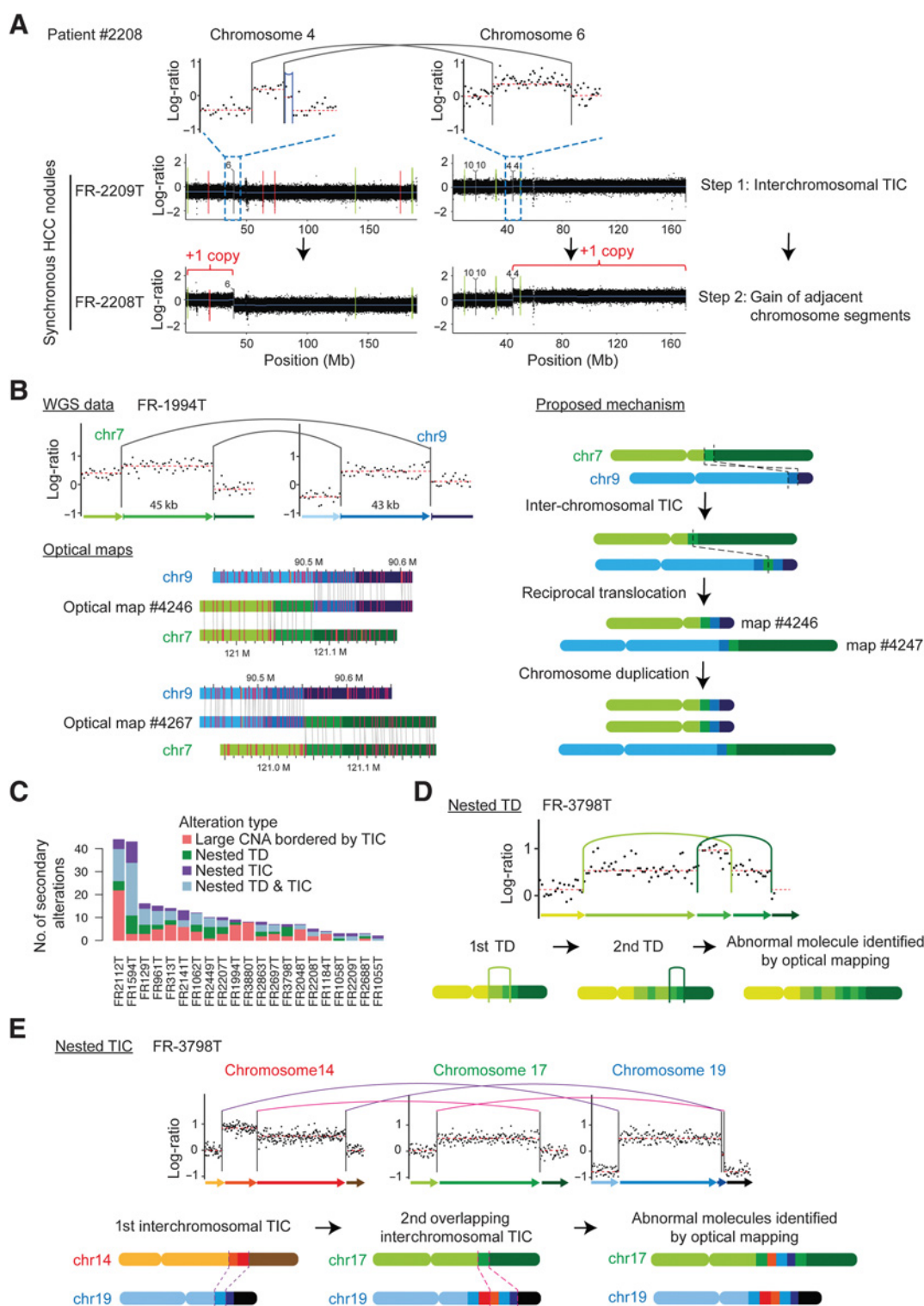


Figure 6.

Secondary chromosome rearrangements and nested SVs. **A**, Example of a TIC followed by large gains of adjacent chromosome segments. **B**, Optical mapping reconstruction of large copy-number gains bordered by an interchromosomal TIC and proposed mechanism. **C**, Number of secondary alterations per CCN-HCC genome. **D**, Example of nested tandem duplications (TD). **E**, Example of nested interchromosomal TIC.

TIC. This might be due to physical constraints on the DNA molecule, for example if the repair process involves loop formation between the newly synthesized strand and the previous template.

In contrast to chromotripsis or chromoplexy that generate a sudden burst of SVs (1, 2), template switching events induce a progressive accumulation of SVs during tumor progression, as evidenced by our multi-sample analysis. Thus, replication stress is a continuous mutational process in CCN-HCC, leading to the acquisition of secondary driver events by transecting tumor suppressor genes, duplicating oncogenes and altering regulatory sequences in promoter regions. In addition to these focal SVs spanning tens to hundreds of kb, TIC promote secondary large-scale chromosome rearrangements and copy-number changes. These events, reminiscent of chromosome rearrangements occurring at HBV insertion sites (31), are frequent in CCN-HCC and will affect the transcriptome of hundreds of genes. Thus, TIC have both direct and indirect oncogenic consequences during CCN-HCC progression.

Authors' Disclosures

T.Z. Hirsch reports grants from CARPEM during the conduct of the study. J.C. Nault reports grants from Bayer and grants from Ipsen during the conduct of the study. J. Calderaro reports personal fees from Owkin, Keen Eye and personal fees from Crosscope outside the submitted work. No disclosures were reported by the other authors.

Authors' Contributions

Q. Bayard: Software, investigation, visualization, methodology, writing—original draft. **P. Cordier:** Investigation, visualization, writing—review and editing. **C. Pèneau:** Investigation, writing—review and editing. **S. Imbeaud:** Software, investigation, writing—review and editing. **T.Z. Hirsch:** Investigation, writing—review and editing. **V. Renault:** Software, investigation, writing—review and editing. **J.C. Nault:** Resources, writing—review and editing. **J.F. Blanc:** Resources, writing—review and

editing. **J. Calderaro:** Resources, writing—review and editing. **C. Desdouets:** Supervision, writing—review and editing. **J. Zucman-Rossi:** Conceptualization, supervision, funding acquisition, writing—review and editing. **E. Letouze:** Conceptualization, supervision, funding acquisition, writing—original draft, project administration.

Acknowledgments

The authors thank Tatiana Popova and Chunlong Chen (Institut Curie) for fruitful discussions, the Bionano team (Karl Hong and Andy Pang) for their help with optical mapping data analysis, and Barkha Gupta for the processing of histologic slides. They thank the clinicians, surgeons, pathologists, hepatologists, and oncologists who contributed to the tissue collection and clinical annotations. The authors also thank the Réseau national CRB Foie (BB-0033-0085), the tumor banks of CHU Bordeaux (BB-0033-00036), Jean Verdier Hospital (APHP), and CHU Henri Mondor (APHP) for contributing to the tissue collection. This work was supported by France Génomique, ITMO Cancer AVIESAN (Alliance Nationale pour les Sciences de la Vie et de la Santé, National Alliance for Life Sciences & Health) within the framework of the Cancer Plan ("HTE program-HetColi network"), ANRS, and the French Liver Biobanks network – INCa, BB-0033-00085. The group is supported by the Ligue Nationale Contre le Cancer (Equipe Labellisée), Labex OncoImmunology (investissement d'avenir), Coup d'Élan de la Fondation Bettencourt-Schueller and the SIRIC CARPEM. Q. Bayard was supported by fellowship from the HOB doctoral school, the Ministry of Education and Research and Association pour la Recherche contre le Cancer (ARC). P. Cordier is a recipient of Plan Cancer INSERM (programme « Soutien pour la formation à la recherche fondamentale et translationnelle en cancérologie »). C. Pèneau was supported by a fellowship from ANRS (French national agency for research on AIDS and viral hepatitis) and T.Z. Hirsch by a fellowship from Cancéropole Ile de France and Fondation d'Entreprise Bristol-Myers Squibb pour la Recherche en Immuno-Oncologie and CARPEM.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received October 27, 2021; revised January 4, 2022; accepted February 2, 2022; published first February 25, 2022.

References

- 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.
- Baca SC, Prandi D, Lawrence MS, Mosquera JM, Romanel A, Drier Y, et al. Punctuated evolution of prostate cancer genomes. *Cell* 2013;153:666–77.
- Li Y, Roberts ND, Wala JA, Shapira O, Schumacher SE, Kumar K, et al. Patterns of somatic structural variation in human cancer genomes. *Nature* 2020;578:112–21.
- 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.
- Letouze E, Shinde J, Renault V, Couchy G, Blanc J-F, Tubacher E, et al. Mutational signatures reveal the dynamic interplay of risk factors and cellular processes during liver tumorigenesis. *Nat Commun* 2017;8:1315.
- Popova T, Manié E, Boeva V, Battistella A, Goundiam O, Smith NK, et al. Ovarian cancers harboring inactivating mutations in CDK12 display a distinct genomic instability pattern characterized by large tandem duplications. *Cancer Res* 2016;76:1882–91.
- Wu Y-M, Cieślak M, Lonigro RJ, Vats P, Reimers MA, Cao X, et al. Inactivation of CDK12 delineates a distinct immunogenic class of advanced prostate cancer. *Cell* 2018;173:1770–82.
- Gaillard H, García-Muse T, Aguilera A. Replication stress and cancer. *Nat Rev Cancer* 2015;15:276–89.
- Kotsantis P, Petermann E, Boulton SJ. Mechanisms of oncogene-induced replication stress: jigsaw falling into place. *Cancer Discov* 2018;8:537–55.
- Jones RM, Mortusewicz O, Afzal I, Lorvellec M, García P, Helleday T, et al. Increased replication initiation and conflicts with transcription underlie Cyclin E–induced replication stress. *Oncogene* 2013;32:3744–53.
- Costantino L, Sotiropoulos SK, Rantala JK, Magin S, Mladenov E, Helleday T, et al. Break-induced replication repair of damaged forks induces genomic duplications in human cells. *Science* 2014;343:88–91.
- Macheret M, Halazonetis TD. Intragenic origins due to short G1 phases underlie oncogene-induced DNA replication stress. *Nature* 2018;555:112–6.
- Bayard Q, Meunier L, Peneau C, Renault V, Shinde J, Nault J-C, et al. Cyclin A2/E1 activation defines a hepatocellular carcinoma subclass with a rearrangement signature of replication stress. *Nat Commun* 2018;9:5235.
- Li H, Durbin R. Fast and accurate short-read alignment with Burrows–Wheeler transform. *Bioinformatics* 2009;25:1754–60.
- Chen X, Schulz-Trieglaff O, Shaw R, Barnes B, Schlesinger F, Källberg M, et al. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics* 2016;32:1220–2.
- Shen R, Seshan VE. FACETS: allele-specific copy-number and clonal heterogeneity analysis tool for high-throughput DNA sequencing. *Nucleic Acids Res* 2016;44:e131.
- Leung D, Jung I, Rajagopal N, Schmitt A, Selvaraj S, Lee AY, et al. Integrative analysis of haplotype-resolved epigenomes across human tissues. *Nature* 2015;518:350–4.
- Servant N, Varoquaux N, Lajoie BR, Viara E, Chen C-J, Vert J-P, et al. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol* 2015;16:259.
- Durand NC, Robinson JT, Shamim MS, Machol I, Mesirov JP, Lander ES, et al. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst* 2016;3:99–101.
- Knight PA, Ruiz D. A fast algorithm for matrix balancing. *IMA J Numer Anal* 2013;33:1029–47.
- Cao H, Hastie AR, Cao D, Lam ET, Sun Y, Huang H, et al. Rapid detection of structural variation in a human genome using nanochannel-based genome mapping technology. *Gigascience* 2014;3:34.
- Sondka Z, Bamford S, Cole CG, Ward SA, Dunham I, Forbes SA. The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. *Nat Rev Cancer* 2018;18:696.

23. Schulze K, Imbeaud S, Letouzé E, Alexandrov LB, Calderaro J, Rebouissou S, et al. Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets. *Nat Genet* 2015;47:505–11.
24. Beroukhi R, Mermel CH, Porter D, Wei G, Raychaudhuri S, Donovan J, et al. The landscape of somatic copy-number alteration across human cancers. *Nature* 2010;463:899–905.
25. Lee JA, Carvalho CMB, Lupski JR. A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. *Cell* 2007;131:1235–47.
26. Hastings PJ, Ira G, Lupski JR. A microhomology-mediated break-induced replication model for the origin of human copy-number variation. *PLoS Genet* 2009;5:e1000327.
27. Carvalho CMB, Ramocki MB, Pehlivan D, Franco LM, Gonzaga-Jauregui C, Fang P, et al. Inverted genomic segments and complex triplication rearrangements are mediated by inverted repeats in the human genome. *Nat Genet* 2011;43:1074–81.
28. Smith CE, Llorente B, Symington LS. Template switching during break-induced replication. *Nature* 2007;447:102–5.
29. Deem A, Keszthelyi A, Blackgrove T, Vayl A, Coffey B, Mathur R, et al. Break-induced replication is highly inaccurate. *PLoS Biol* 2011;9:e1000594.
30. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 2009;326:289–93.
31. Péneau C, Imbeaud S, La Bella T, Hirsch TZ, Caruso S, Calderaro J, et al. Hepatitis B virus integrations promote local and distant oncogenic driver alterations in hepatocellular carcinoma. *Gut* 2022;71:616–26.