The tandem duplicator phenotype as a distinct genomic configuration in cancer

Francesca Menghi[a,1], Koichiro Inaki[a,1,2], XingYi Woo[b], Pooja A. Kumar[a], Krzysztof R. Grzeda[a], Ankit Malhotra[a], Vinod Yadav[a], Hyunsoo Kim[a], Eladio J. Marquez[a], Duygu Ucar[a], Phung T. Shreckengast[a], Joel P. Wagner[a], George MacIntyre[a], Krishna R. Murthy Karuturi[a], Ralph Scully[c], James Keck[d], Jeffrey H. Chuang[a], and Edison T. Liu[b,3]

[a]The Jackson Laboratory for Genomic Medicine, Farmington, CT 06032; [b]The Jackson Laboratory, Bar Harbor, ME 04609; [c]Division of Hematology Oncology, Department of Medicine, and Cancer Research Institute, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, MA 02215; and [d]The Jackson Laboratory, Sacramento, CA 95838

Edited by Tom Walsh, University of Washington, Seattle, WA, and accepted by the Editorial Board March 2, 2016 (received for review October 8, 2015)

Next-generation sequencing studies have revealed genome-wide structural variation patterns in cancer, such as chromothripsis and chromoplexy, that do not engage a single discernable driver mutation, and whose clinical relevance is unclear. We devised a robust genomic metric able to identify cancers with a chromotype called tandem duplicator phenotype (TDP) characterized by frequent and distributed tandem duplications (TDs). Enriched only in triple-negative breast cancer (TNBC) and in ovarian, endometrial, and liver cancers, TDP tumors conjointly exhibit tumor protein p53 (*TP53*) mutations, disruption of breast cancer 1 (*BRCA1*), and increased expression of DNA replication genes pointing at rereplication in a defective checkpoint environment as a plausible causal mechanism. The resultant TDs in TDP augment global oncogene expression and disrupt tumor suppressor genes. Importantly, the TDP strongly correlates with cisplatin sensitivity in both TNBC cell lines and primary patient-derived xenografts. We conclude that the TDP is a common cancer chromotype that coordinately alters oncogene/tumor suppressor expression with potential as a marker for chemotherapeutic response.

tandem duplications | cisplatin | triple-negative breast cancer | BRCA1 | TP53

Cancer evolution is generally thought to result from the progressive accumulation of genomic lesions affecting key regulatory components of physiological cellular functions (1, 2). Oncogenic changes can manifest as single-nucleotide mutations; copy number alterations, such as deletions or duplications; and balanced rearrangements, including chromosomal translocations and inversions (3).

More recently, the systematic application of whole-genome sequencing (WGS) to the study of human cancer genomes has uncovered more complex scenarios, where large portions of the genome are affected by a multitude of somatic structural variations, which either originate from a few unique catastrophic events [e.g., chromothripsis, chromoplexy (4–6)] or result from the derangement of key molecular mechanisms leading to specific mutator phenotypes (7, 8). Although not always associated with a discernible driver mutation, these genome-wide structural variation patterns have the potential to deregulate several oncogenic elements simultaneously, and have been clearly associated with malignant phenotypes (4, 5, 9, 10).

Despite their relevance to the tumorigenic process, the causes of these genome-wide chromotypes, the cancer-driving oncogenic elements induced by these structural changes, and the clinical implications of these configurations remain unclear. Although recent advances have been made in understanding the mechanisms underlying chromothripsis, no specific therapeutic intervention has yet been identified for chromothriptic cancers or for other chromotypes (11–14).

Here, we study one of these genomic configurations, the tandem duplicator phenotype (TDP), which is characterized by the presence of a large number of somatic head-to-tail DNA segmental duplications [i.e., tandem duplications (TDs)] homogeneously distributed throughout the cancer genome (10, 15). In a meta-analysis of over 3,000 cancer genomes, we identify the most prevalent genetic features associated with this phenotype and those genetic features that may be responsible for its tumorigenic drive. Furthermore, we show an association between the extent of TDP and sensitivity to platinum-based chemotherapy in cell and primary xenograft models of triple-negative breast cancer (TNBC), providing a first indication of the potential utility of the TDP chromotype as a predictive genomic biomarker in a clinical setting.

Results

Homogeneous Distribution of TDs Across Cancer Genomes as a Systematic Measure of the TDP. To address the lack of a systematic approach to identify and score the TDP, we developed a reproducible metric of TD genomic distribution, which we refer to as the TDP score. For each tumor sample, we tally the total number of TDs mapped by breakpoint analysis, and compare the

Significance

In this study, we provide the first detailed molecular characterization, to our knowledge, of a distinct cancer genomic configuration, the tandem duplicator phenotype (TDP), that is significantly enriched in the molecularly related triple-negative breast, serous ovarian, and endometrial carcinomas. We show here that TDP represents an oncogenic configuration featuring (*i*) genome-wide disruption of cancer genes, (*ii*) loss of cell cycle control and DNA damage repair, and (*iii*) increased sensitivity to cisplatin chemotherapy both in vitro and in vivo. Therefore, the TDP is a systems strategy to achieve a protumorigenic genomic configuration by altering a large number of oncogenes and tumor suppressors. The TDP arises in a molecular context of joint genomic instability and replicative drive, and is consequently associated with enhanced sensitivity to cisplatin.

Author contributions: F.M., K.I., and E.T.L. designed research; F.M., P.A.K., P.T.S., G.M., and J.K. performed research; K.R.G., A.M., D.U., and J.H.C. contributed new reagents/analytic tools; F.M., K.I., X.W., K.R.G., A.M., V.Y., H.K., E.J.M., D.U., J.P.W., K.R.M.K., and J.H.C. analyzed data; and F.M., R.S., and E.T.L. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. T.W. is a guest editor invited by the Editorial Board.

Freely available online through the PNAS open access option.

Data deposition: Whole-genome sequencing data are freely available from the Sequence Read Archive (SRA) database, www.ncbi.nlm.nih.gov/sra (Project ID SRP057179).

[1]F.M. and K.I. contributed equally to this work.

[2]Present address: Functional Genomics and Proteomics Research Group, Discovery Science and Technology Department, Daiichi Sankyo RD Novare Co., Ltd, Tokyo 103-8426, Japan.

[3]To whom correspondence should be addressed. Email: edison.liu@jax.org.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1520010113/-/DCSupplemental.

observed (Obs_i) and expected (Exp_i) numbers of TDs for each chromosome i:

$$\text{TDP Score} = -\frac{\sum_i |Obs_i - Exp_i|}{\text{TD}} + k,$$

where k equals the threshold value, which normalizes all values to the subsequently determined threshold for the TDP configuration (discussed below).

This metric is easily able to distinguish between a genomic configuration characterized by localized segmental amplifications with TDs vs. the TDP, in which TDs are evenly distributed across all chromosomes (Fig. 1*A*).

To address the incidence and genomic properties of the TDP, we combined WGS data from 277 human genomes representing 11 cancer types, including 96 breast tumors and cancer cell lines (4, 9, 10, 15–22) (Dataset S1). We observed that the TDP score distribution in this dataset follows a trimodal pattern (*Materials and Methods* and Fig. S1*A*), suggesting that cancers can be separated into distinct groups based on their propensity for TD formation. Upon visual inspection of tumors within a range of TDP scores by Circos plots, those tumors with the highest scores show the characteristic TD distribution of the TDP (Fig. 1*A*). In order to derive an unbiased threshold for classifying TDP tumors, we identified the threshold as the score that corresponds to 2 SDs from the second modal peak (−0.71; Fig. S1*A*). To simplify data presentation, we then set the TDP score to 0 at this defining threshold (k), resulting in positive and negative scores for TDP and non-TDP tumors, respectively (Fig. 1*B*). Using this threshold, 18.1% of the tumors analyzed are classified as TDPs, each showing a high number of TDs (average number of TDs per sample = 112.2, range: 23–416, modal TDP score = 0.19) that are broadly distributed throughout the genome (Fig. 1*A* and Fig. S1 *A* and *B*). By contrast, non-TDP samples are either associated with an intermediate number of TDs (10 to ∼100, modal TDP score = −0.50) that are invariably clustered in specific genomic regions or have a low number of TDs altogether (<20) indicative of a more stable genome (Fig. 1*A* and Fig. S1 *A* and *B*).

We applied a similar scoring method to the three other basic rearrangements (deletion, inversion, and interchromosomal translocation), but found no evidence for distinct groups to manifest in multimodal score distributions as seen for TDs (Fig. S1*C*). This finding suggests that the TDP is not merely an indicator of genomic instability but, instead, represents a unique tumor subgroup with a distinct structural phenotype.

Previous evidence has suggested a higher frequency of the TDP in TNBC and ovarian (OV) cancer (19). Using our more precise and quantitative TDP measure based on only the WGS dataset, we confirmed that the TDP occurs statistically more frequently in TNBC (P = 2.16E-04), OV carcinoma (P = 4.95E-02), and hepatocellular carcinoma (P = 2.92E-02) but that it is significantly depleted in non-TNBC (P = 5.27E-02), glioblastoma (P = 4.10E-02), and prostate cancer (P = 1.77E-03) (Table 1). Indeed, we rarely observed TDP samples in prostate cancer, in which chromoplexy and chromothripsis appear to be the predominant whole-genome rearrangement patterns (4). This finding suggests that different mechanisms are active in different tumor types to produce specific dominant cancer genomic configurations.

Whereas the TDP score is based on the identification of TDs through the assignment of breakpoints, and relies on the availability of WGS data, Ng et al. (15) estimated the prevalence of the TDP by counting the number of TD-like features from array-based copy number profiling in high-grade serous OV carcinoma. We wanted to compare the performance of our TDP scoring algorithm when applied to sequence- vs. array-based detection systems. We therefore analyzed Affymetrix SNP 6.0 array segmented copy number data from a subset of 81 tumor genomes profiled as part of The Cancer Genome Atlas (TCGA) project to compute copy number (array)-derived TDP scores and compare them with TDP scores obtained using paired-end WGS data (Fig. S2 *A* and *B*). Using SNP array copy number data alone, we could identify TDP samples with high specificity (0.95; Fig. S2 *C* and *D*) but lower sensitivity (0.57), likely due to the lower resolution of array data in detecting short segmental duplications. To increase the discrimination power of the SNP array-based TDP classification, we set a more stringent threshold to categorize non-TDP samples (Fig. S2*E*) and improve the sensitivity of the technology to 0.80 (Fig. S2*F*).

The advantage of analyzing array-based data is the availability of a larger number of cancer samples. When we classified 2,987 primary tumors from several TCGA datasets profiled using the Affymetrix SNP 6.0 array, we were able to reproduce our previous findings that the TDP is significantly enriched in TNBC (P = 1.23E-08) and OV cancer (P = 4.16E-94), whereas it is depleted in non-TNBC (P = 2.41E-20) (Table 1 and Dataset S2). In addition, because of the greater number of available tumors in the TCGA array dataset, we found that uterine corpus endometrial carcinoma (UCEC) also is enriched in TDPs (P = 2.80E-09). Interestingly, most of the UCEC samples classified as TDPs belong to the recently described cluster 4 endometrial carcinoma subtype, which is characterized by an extensive degree of copy number variations (CNVs) and has been shown to share a similar

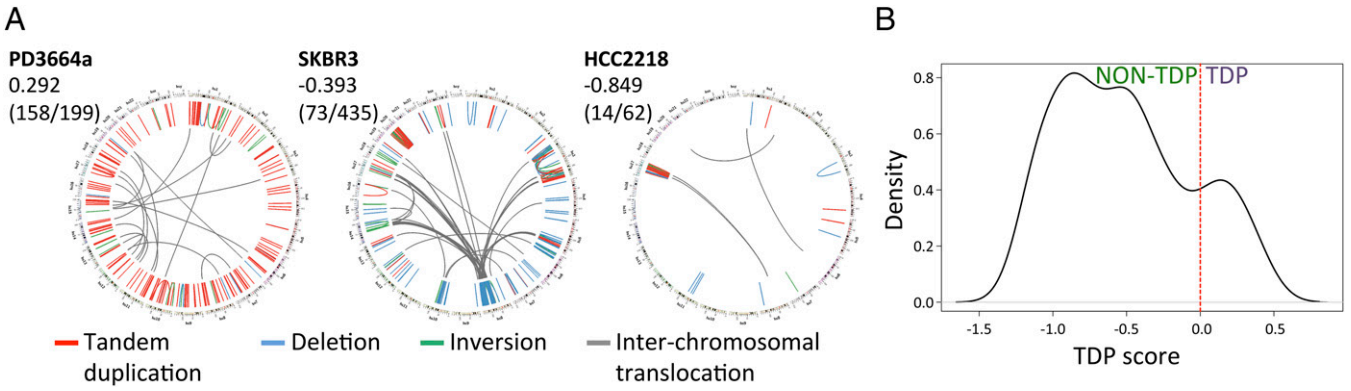

Fig. 1. TDP scoring and sample classification. (*A*) Circos plots showing structural variations of representative cancer genomes with different levels of TDP scores. For each plot, sample identification number, the TDP score, and number of TDs over the total number of detected rearrangements are indicated (top to bottom). Structural variations were classified based on the four basic discordant paired-end mappings as TDs (red), deletions (blue), unpaired inversions (green), or interchromosomal translocations (gray). (*B*) Trimodal distribution of the TDP score values across the 277 cancer samples examined.

Table 1. Prevalence of the TDP among different tumor types

	WGS					SNP array*				
Cancer type	Total no.	TDP no.	%	*P*	Status	Total no.	TDP no.	%	*P*	Status
TNBC	40	17	42.5	2.16E-04	E	94	37	39.4	1.23E-08	E
Other breast cancers (non-TNBC)	56	6	10.7	5.27E-02	D	594	22	3.7	2.41E-20	D
Colorectal adenocarcinoma	14	0	0.0	6.11E-02	ns	545	6	1.1	3.36E-31	D
Glioblastoma	16	0	0.0	4.10E-02	D	18	2	11.1	2.50E-01	ns
Hepatocellular carcinoma	19	7	36.8	2.92E-02	E	NA	—	—	—	
Kidney renal clear cell carcinoma	3	0	0.0	5.49E-01	ns	509	2	0.4	4.61E-34	D
Lung adenocarcinoma	25	3	12.0	1.69E-01	ns	NA	—	—	—	
Lung squamous cell carcinoma	18	5	27.8	1.24E-01	ns	364	31	8.5	3.43E-05	D
Multiple myeloma	7	0	0.0	2.47E-01	ns	NA	—	—	—	
OV	26	8	30.8	4.95E-02	E	382	236	61.8	4.16E-94	E
Prostate cancer	43	1	2.3	1.77E-03	D	NA	—	—	—	
Endometrial carcinoma	10	3	30.0	1.76E-01	ns	481	123	25.6	2.80E-09	E
Total	277	50	18.1			2,987	459	15.4		

TDP status was assigned based on either WGS data (n = 277 tumor samples) or Affymetrix SNP 6.0 array data (SNP array, n = 2,987 tumor samples). *P* values were computed using the binomial test. D, depletion; E, enrichment; ns, nonsignificant.
*Tumor samples were classified based on the stringent thresholds described in Fig. S2*E*.

molecular phenotype with TNBC and OV cancer (23). The consistent observation of TDP enrichment/depletion across alternative cancer datasets, generated via diverse genomic technologies and analysis protocols, suggests that our scoring approach is reproducible and generalizable.

TD Breakpoints Occur in Regions of Open Chromatin and Active Transcription. To investigate possible molecular mechanisms for the generation of the TDP, we examined the genetic, epigenetic, and transcriptional configurations of the chromosomal coordinates affected by TD events in TDP genomes. We focused our analysis on breast cancer [TNB and non-TNB (NTNB) WGS datasets, n = 23 TDP tumor genomes], because this type of cancer was the best-represented tumor type in our WGS sample cohort, and therefore provided adequate statistical power. We first asked whether TDs in TDP occurred in functional regions of the genome enriched for genes. We observed a highly significant positive correlation between the number of TD breakpoints and the number of genes in local windows along the genome (R = 0.5, P = 1.8E-178; 10-Mb sliding windows, 1-Mb offset; Fig. S3*A*). Furthermore, TD breakpoints were biased to occur within gene bodies (exons + introns) as opposed to intergenically (P < 1.0E-04; Fig. S3*B*). We assessed the physiological expression levels of genes that are frequently affected by breast cancer TD breakpoints in normal breast tissue. Based on a collection of 106 normal breast epithelium samples from the TCGA breast cancer dataset, genes located at the boundaries of TDs show significantly higher levels of activity in the normal breast when compared with the entire gene population (P < 2.2E-16; Fig. S3*C*). This observation is consistent with the positioning of TD boundaries near genes with antioncogenic signals, which would subsequently be disrupted during TD formation. However, it also suggests that TD formation requires transcriptional activity. Indeed, we observed a significant enrichment of Pol2 binding sites as well as histone modification marks associated with an open chromatin configuration (H3K4me3, H3K4me1, and H3K27ac) in the proximity of TD breakpoints (P < 1.0E-04; Fig. S3 *D–F*). This finding is in agreement with recent findings describing a strong affinity of structural variation breakpoints for genomic regions characterized by protein binding and euchromatin (22). By contrast, H3K9me3 signals, which mark heterochromatin, were depleted from TD breakpoint regions (P < 1.0E-04; Fig. S3 *E* and *F*). Concordant results were obtained by testing different nonoverlapping symmetrical windows around the TD breakpoints, showing that significant associations between functionalized chromatin regions and TDs are maintained up to ∼200–500 kb from the TD breakpoints (Fig. S3*G*). Overall, these results concordantly indicate a significantly higher likelihood for TD breakpoints to affect transcriptionally active, easily accessible chromatin regions. The mechanistic underpinnings of this relationship are unclear. One possibility is that if TDs are related to replisome stalling, collisions between the replisome and ongoing transcription might be more common in highly transcribed genes. Alternatively, the TDs embedded within certain highly transcribed genes may be preferentially selected during tumor evolution (discussed below).

Genomic Features of TDs in TDP and Non-TDP Tumors. A comparison between the genomic properties of structural rearrangements occurring in TDP and non-TDP samples shows a striking difference in the per-sample median TD span size, with TDP samples having significantly smaller median spans (median span size = 89.9 kb for TDPs and 1,189.7 kb for non-TDPs; P = 7.23E-09; Fig. 2*A*). More specifically, by plotting the distribution of the collection of all individual TD spans for TDP and non-TDP genomes (WGS dataset; n = 50 and n = 227, respectively), we observed that whereas non-TDP tumors feature a continuum range of very large TDs reaching a plateau at around 1 Mb, TDP samples are characterized by two sharper TD span distribution modes at ∼10 kb and ∼250 kb (Fig. 2*B*). This finding suggests that in TDP tumors, the mechanism for generating TDs may be different than for non-TDP tumors.

We directly sequenced the rearrangement junctions of 122 TDs from 11 different TNBC cell lines of both TDP and non-TDP types, and analyzed the sequences at the breakpoint junctions for patterns indicative of specific DNA repair mechanisms (10, 21, 24). We classified the validated breakpoint junctions into those junctions characterized by the presence of short (<10 bps) or long insertions; short (<5 bps), long, or no microhomology (MH); or long-range imperfect homology (Fig. 2*C*). The large majority of TDs in TDP tumors (72%, range: 46–82%) show overlapping MH between the two DNA segments contributing to the rearrangement junction, suggesting that the underlying mechanism entails MH-mediated end-joining or MH-mediated break-induced replication (MMBIR) (13, 24). Significantly, only 40% (range: 27–86%) of TDs found in non-TDP tumors show a similar profile [odds ratio (OR) = 3.6, P = 6E-04; Fig. 2*C* and Dataset S3]. By contrast, TD rearrangements characterized by long-range imperfect homology, a signature indicative of nonallelic homologous repair (NAHR) (24), are prevalent in non-TDP tumors [23% (range: 0–50%) vs. 7% (range: 0–31%) in TDPs; OR = 0.25,

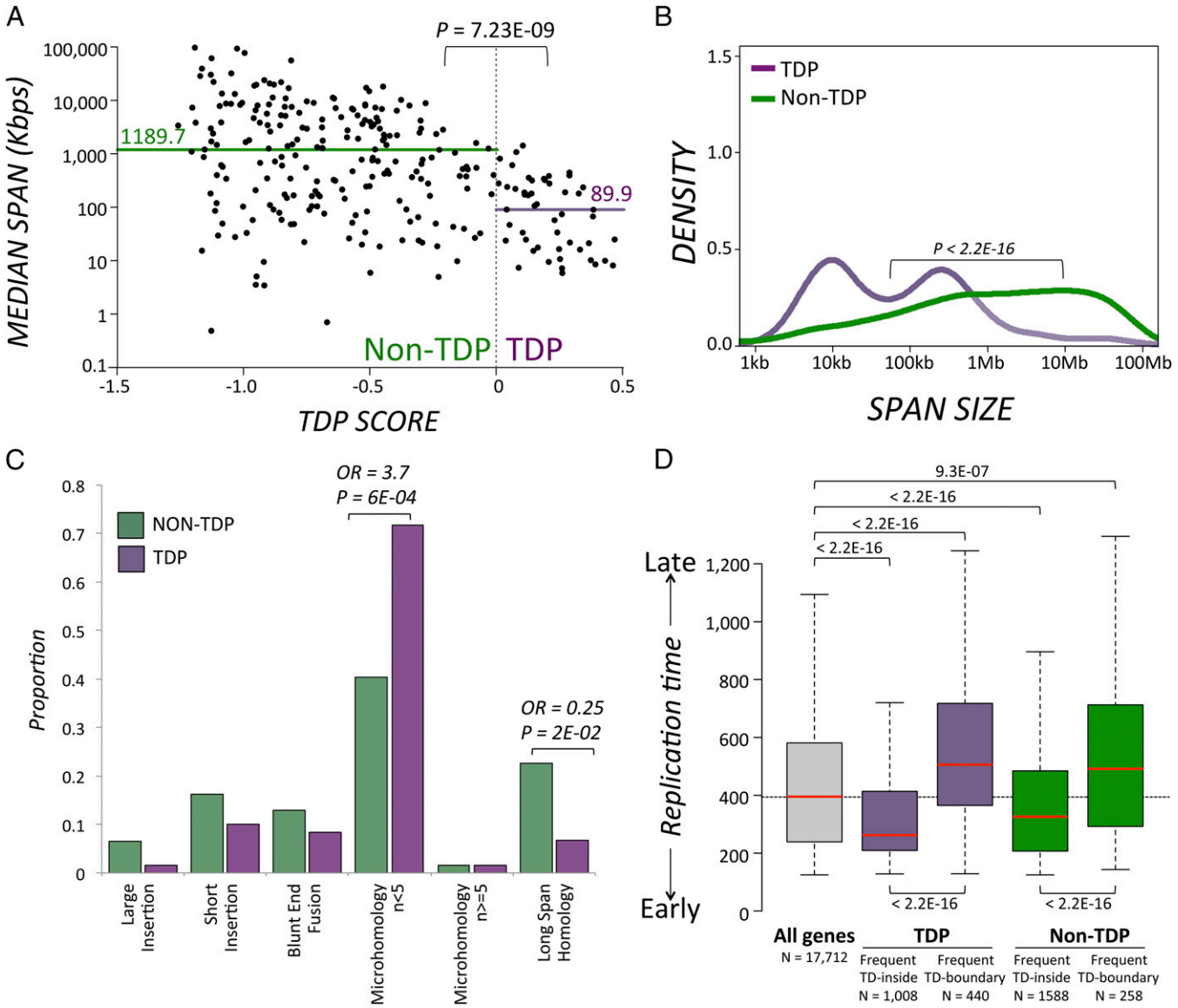

Fig. 2. Genomic features of TDs in TDP and non-TDP tumors. (*A*) Correlation of TDP score and median TD span size across the 277 tumor genomes analyzed by WGS. Horizontal lines indicate the overall median span size for the TDP and non-TDP sample subgroups. A *P* value was computed using Student's *t* test. (*B*) TD span distributions for the TDP and the non-TDP sample groups. TDP samples feature TDs with span peaks at ∼10 kb and ∼150 kb. Non-TDP samples feature a much larger TD span range, which homogeneously ranges from ∼1 to ∼10 Mb. A *P* value for the distance between the two empirical distributions was generated using the two-sample Kolmogorov–Smirnov test. (*C*) Sequence analysis of TD breakpoints across TDP ($n = 4$) and non-TDP ($n = 7$) TNBC cell line genomes. ORs and *P* values were computed using Fisher's exact test. (*D*) Replication time (RT) of genes located inside or on the boundary of TDs in TDP and non-TDP samples based on the breast cancer dataset. RT is expressed on a scale of 100 (early) to 1,500 (late). *P* values were computed based on the Mann–Whitney *U* test.

P = 2E-02; Fig. 2*C* and Dataset S3]. These differences further support the idea that distinct DNA repair mechanisms may underlie the formation of TDs in TDP and non-TDP tumors.

Recent evidence has revealed meaningful correlations between DNA replication timing, genomic instability, and the emergence of DNA mutations (25, 26). Indeed, we found a significant association between TD-affected genes and replication timing (27). Genes truncated by TD boundaries are found in late replication regions, and genes spanned by TDs are enriched in early replicating regions ($P < 2.2E\text{-}16$ and $P < 2.2E\text{-}16$ for the TDP set; $P = 9.3E\text{-}07$ and $P < 2.2E\text{-}16$ for the non-TDP set; Fig. 2*D*). This specific pattern of replication timing is consistent across all samples (TDPs and non-TDPs), and it may reflect a shortage of DNA repair opportunities in late S phase, leading to an increased incidence of misrepaired double-strand breaks resulting in CNVs (25). However, given that DNA replication typically encompasses ∼400- to 800-kb chromosomal domains, it is plausible that the shorter TDs found in TDP genomes are generated within intrareplication timing domains, whereas, the larger, non-TDP TDs are more likely to result from the spatial proximity of distinct replication domains through the tridimensional looping of chromatin structures.

TDP Is Characterized by the Coordinated Perturbation of Several Cancer Genes. One of the most direct consequences of DNA segmental duplication is the increased expression of the genes that are entirely comprised within the rearrangement, whose copy number is thus augmented. We hypothesize that a genomic configuration generating a large number of segmental duplications would represent a cancer genomic mechanism for the

modulation of hundreds of potential oncogenic signals, providing a selective advantage for the TDP cancer cell. To assess this possibility, we first compared changes in gene expression between normal and tumor breast samples, with respect to the genes found to be most frequently affected by TDs in the TDP breast cancer WGS dataset (n = 23; Dataset S4). As hypothesized, genes that are frequently found inside TDs are generally overexpressed in breast cancers when compared with the normal breast epithelium (median $\log_2$-fold change = 0.17, P = 4.0E-16). In contrast, genes frequently located at the boundaries of TDs appear to be down-regulated in breast cancers (median $\log_2$-fold change = −0.3, P = 5.0E-05) (Fig. 3*A*). Moreover, genes frequently encompassed by TD segments are enriched in known oncogenes (P = 1.2E-02) and genes whose increased expression levels are associated with a poor prognosis for patients with breast cancer (P = 3.3E-05), whereas genes that map to TD boundaries are most significantly associated with known (P = 5.9E-05) and putative tumor suppressor genes (STOP genes, P = 5.1E-04; good prognosis genes, P = 4.6E-12; Fig. 3*B*). We confirmed these findings by identifying the genes affected by TD-like features predicted using SNP array data, which provided a significantly larger dataset (n = 418 TDP tumor samples; Fig. S4*A*). Indeed, well-known oncogenes, such as paired box 8 (*PAX8*), erb-b2 receptor tyrosine kinase 2 (*ERBB2*), and *MYC*, are among the most recurrent genes that are spanned by a TD across TDP samples, whereas known tumor suppressor genes, such as *RAD51L*, *PTEN*, and *RB1*, populate the list of the top genes affected by TD breakpoints (Fig. S4*B* and Dataset S5).

This systems strategy to generating the cancer state supposes that many different combinations of oncogenic signals would suffice as opposed to a single dominant oncogenic cassette such as the cassette proposed for genes associated with *ERBB2* amplification (9). To test this strategy, we examined the frequency of specific one-gene and multiple-gene combinations affected by TDs across 418 TDP genomes assessed using SNP array data (TNB, NTNB, OV, and UCEC datasets) and found that only up to a maximum of 15.5% of tumors share TD-like features affecting a single common tumor suppressor gene (*RAD51L1* and, at lower frequencies, *WWOX*, *NF1*, *RB1*, *PTEN*, and *BRCA1*; Fig. S4*B* and Dataset S5), and, even less frequently, an oncogene (i.e., *PAX8*, duplicated in 10.5% of tumors, followed by *ERBB2*, *ERBB3*, *TERC*, *STAT2*, *CDK2*, and *MYC;* Fig. S4*B* and Dataset S5). In addition, two-gene combinations are relatively rare, with the top-scoring gene pairs being those pairs that map within a short distance of each other, and are therefore affected by the same TDs (e.g., *PAX8*, *PSD4*, which are coordinately duplicated in 8.9% of the tumors examined, or *PAX8*, *CBWD2*, *IL1RN*, which are coordinately duplicated in 6% of the tumors examined (Fig. S4 *C–E*). Much rarer are two-gene combinations comprising frequent TD-boundary genes (Fig. S4*C*), arguing against the presence of a dominant TD-affected cancer gene or small gene set.

Intriguingly, we observed that the shorter span TDs seen exclusively in TDP (~10 kb) do not cause the segmental duplication of full-length genes but disrupt gene body integrity. We found that 38.2% (1,181 of 3,086) of the short-span TDs (span < 100 kb) present in the 50 TDP cancer genomes analyzed by WGS are completely embedded within a gene body, often disrupting the intron/exon structure (P < 0.001; Fig. S5). Moreover, we observed that the genes affected by these short TDs are more likely to function as anticancer as opposed to procancer genes, because they are enriched in TSGs (P = 1.5E-03) and putative TSGs (P = 1.8E-11), while being depleted for oncogenes (known oncogenes, P = 7.3E-03; poor prognosis genes, P = 3.1E-04; Fig. 3*B*).

Taken together, these results strongly suggest that the consequence of generating many TDs is a genome-wide mechanism that simultaneously augments (albeit moderately) the expression of many oncogenes and suppresses the expression of anti-oncogenes/antitumor suppressors. In this model, there is no obvious genetic driver by virtue of levels of expression or the frequency of occurrence. Given these findings, and the fact that the TDP characteristic is presumably established in the preneoplastic cells prior to the generation of TDs, we searched for genetic alterations that might cause a cell to adopt a TDP.

Insights into the Molecular Background Favoring TDP Formation. In the first analysis, we could not find enrichment of specific TDs in the TDP tumors that could explain the unique genomic features associated with the phenotype. This result suggested that there may be intrinsic molecular differences between TDP and non-TDP tumors that induce the TDP and that the changes in gene expression arising from tandem duplicons are a consequence of the TDP.

To identify factors that may correlate with the molecular mechanisms underlying the phenotype, we investigated the characteristics

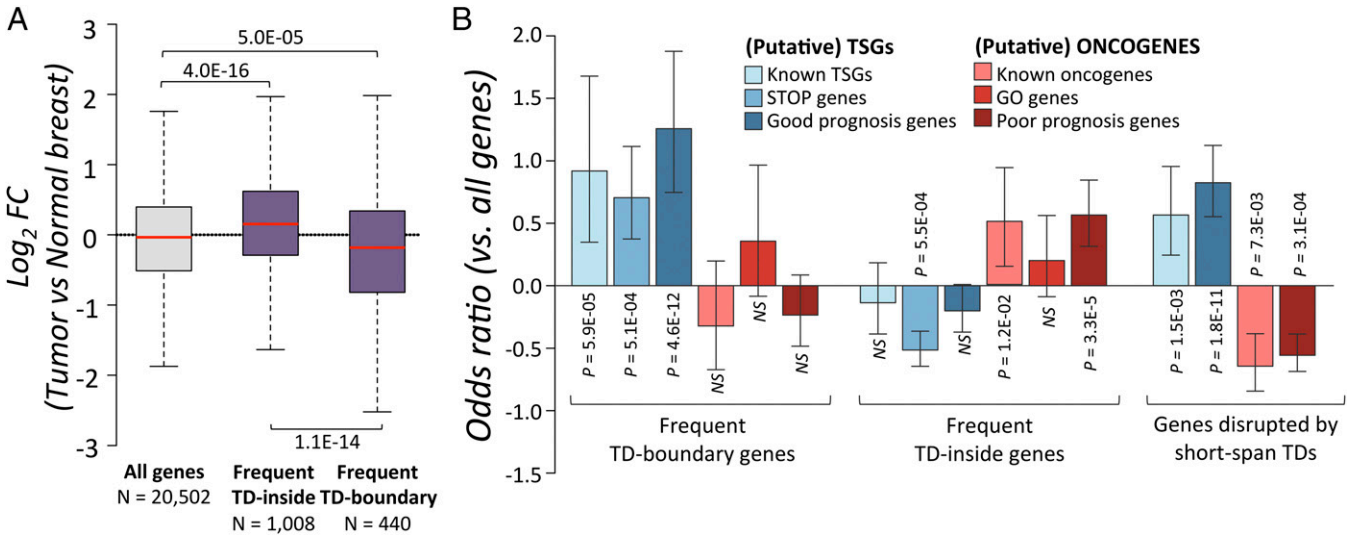

Fig. 3. TDP is characterized by the coordinated perturbation of several cancer genes. (*A*) Fold change (FC) in gene expression (breast tumor/normal breast) for genes frequently located inside or at the boundary of TDs in TDP tumors (P values determined by the Mann–Whitney U test). (*B*) Genes frequently affected by a TD breakpoint are enriched in anticancer genes (*Left*), whereas genes frequently spanned by a TD are enriched in procancer genes (*Middle*). (*Right*) Short-span TDs appear to interfere with anticancer most frequently as opposed to procancer gene integrity. (P values determined by Fisher's exact test).

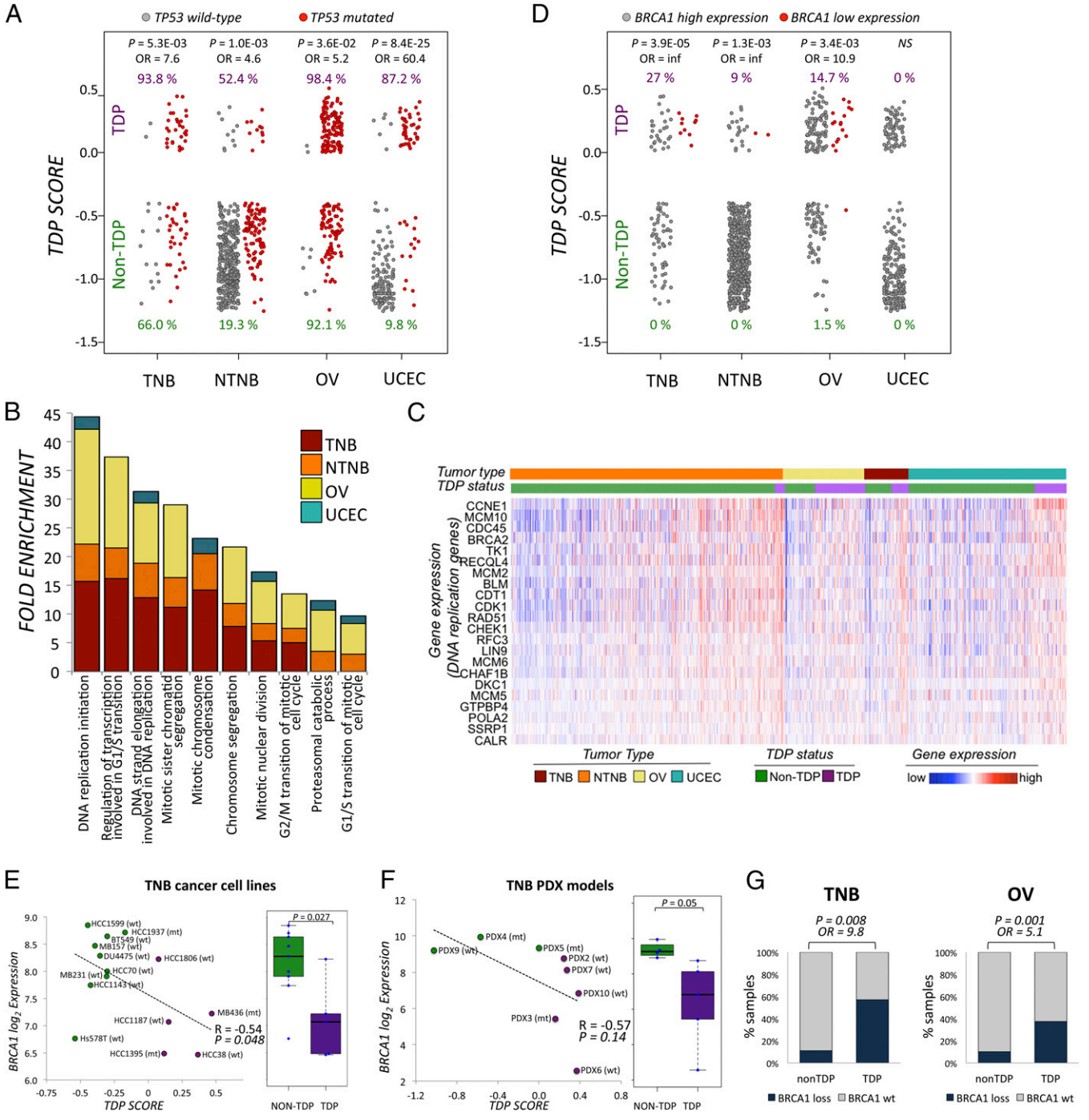

Fig. 4. Loss of the *TP53* and *BRCA1* tumor suppressor genes in the context of abnormal DNA replication may provide a permissive background for the insurgence of the TDP. (*A*) *TP53* mutation rate is recurrently higher in TDP samples compared with non-TDP samples. ORs and corresponding *P* values refer to the enrichment of TDP samples for samples with gene disruption. Percentages of TDP and non-TDP samples carrying the gene disruption are indicated in purple and green, respectively. (*B* and *C*) DNA replication genes are consistently up-regulated in TDP vs. non-TDP samples. (*B*) Top 10 GO terms significantly enriched in up-regulated genes (TDP vs. non-TDP) across the four different datasets analyzed. (*C*) Heat map of individual gene expression levels. Tumor samples are sorted based on tumor type and increasing TDP score. Only the 23 DEGs closely involved in DNA replication are shown. (*D*) TDP samples are significantly enriched in *BRCA1* low expressors across different tumor types. The threshold for low *BRCA1* expression was defined based on the bimodal distribution of *BRCA1* transcriptional levels in each individual dataset. Graph annotations are as in *A*. Expression levels of the *BRCA1* gene in TDP (purple) and non-TDP (green) TNBC cell lines (*E*) and PDXs (*F*) are shown. TDP scores for these genomes were computed based on WGS data. The *BRCA1* somatic mutational status is indicated in brackets. mt, mutated; na, not available; wt, wild type. Pearson correlation coefficients (*R*) and their corresponding *P* values are reported in each graph. (*Right*) Box plots of *BRCA1* expression values for TDP and non-TDP sample groups, $\log_2$-fold changes and Student's *t* test *P* values are shown. (*G*) TDP samples are enriched for *BRCA1*-deficient tumors in both the TNB and OV datasets. *BRCA1* loss is defined by the presence of germline or somatic mutations, or promoter methylation.

of TDP as compared with non-TDP samples within each of the three most highly TDP-enriched tumor datasets: TNB, OV, and UCEC. In addition, we extended our analysis to non-TNBCs (NTNB dataset), which, although depleted in TDPs as a cancer group, comprised a sufficient number of TDP and non-TDP samples to perform statistical comparisons. We first computed the overall mutation burden as the total number of genes per sample that are affected by at least one nonsilent mutation as assessed by exome sequencing (23, 28, 29). Although the TNB, NTNB, and, to a lesser extent, OV datasets showed a significantly higher mutation burden in the TDP subgroup (P = 3.7E-05, P = 9.4E-06, and P = 4.0E-02, respectively), this trend was not consistent in the remaining dataset (UCEC; Fig. S6).

We therefore focused on individual gene mutations to search for genes that, when mutated, are associated with the TDP. For each cancer dataset analyzed, we compiled a list of frequently mutated genes (i.e. mutated in at least 15% of cases within either the TDP or non-TDP sample subgroup). Somatic mutation frequencies were then compared between TDP and non-TDP tumors using Fisher's exact test, and significant differences were assessed across cancer datasets (Dataset S6). Of a total of 56 frequently mutated genes, the *TP53* gene is the only one whose somatic mutation rate is recurrently higher in TDP relative to non-TDP samples across different tumor types, with all of the four examined datasets showing a significant enrichment (TNB, OR = 7.6; NTNB, OR = 4.6; OV, OR = 5.2; UCEC, OR = 60.4) (Fig. 4*A* and Dataset S6).

We then asked whether TDP and non-TDP tumors show profiles of differential gene expression that distinguish these two states. Following the identification of differentially expressed genes (DEGs) between TDP and non-TDP tumors within each tumor-type dataset, we performed a gene ontology (GO) enrichment analysis of the lists of up- and down-DEGs to identify biological processes most commonly perturbed in association with the TDP. Up-regulation of genes engaged in biological processes relevant to cell proliferation and DNA replication appeared to be the most robustly and consistently enriched across all four analyzed datasets (Fig. 4*B*). This finding strongly suggests that TDPs are more prone to increased/perturbed DNA replication. Among the DNA replication genes most frequently up-regulated (in at least three of the four datasets examined), *CCNE1* was the one with the highest cumulative fold change, followed by several critical DNA replication initiation factors, including *CDT1*, *MCM2*, *MCM6*, and *MCM10* (30, 31) (Fig 4*C* and Dataset S7).

Although no multigene cassettes engaged in specific biological processes appeared to be consistently down-modulated in the TDP datasets, we observed in the cancer subgroup of TNBC that the *BRCA1* gene is among the most significantly down-regulated genes, with a greater than two-fold decrease in TDP vs. non-TDP tumors (P = 0.03; Fig. S7 *A* and *B*). Indeed, we found a highly significant enrichment for TDP tumors in *BRCA1* low expressors (27% of all TDP samples compared with 0% of non-TDP TNBC samples; P = 3.9E-05). We validated the strong association between low *BRCA1* expression and the TDP score in the NTNB (P = 1.3E-03) and OV (P = 3.4E-03) datasets (Fig. 4*D*), and in two other independent TNBC datasets (P = 0.027 and P = 0.05), all showing an overall negative correlation between *BRCA1* expression level and TDP score (Fig. 4 *E* and *F*). Furthermore, we found a significant association between *BRCA1* promoter methylation status and reduced *BRCA1* expression levels in the TNB (R = −0.61, P = 2.3E-07) and OV (R = −0.74, P < 1.0E-05E) datasets (Fig. S7*C*), pointing at epigenetic silencing as a key mechanism of transcriptional inactivation of *BRCA1* in TDP tumors.

Whereas we did not find any enrichment in *BRCA1* somatic mutations that distinguishes TDP, when we combined somatic and germline mutations and promoter hypermethylation, we did observe a significant increase in the frequency of *BRCA1* disruption in TDP vs. non-TDP tumors in the TNB and OV datasets (OR = 9.8 and OR = 5.1, P = 8.0E-03 and P = 1.0E-03, respectively; Fig. 4*G*). On the contrary, *BRCA2* mutation rates did not show any association with the TDP and, instead, appeared to be modestly but consistently higher in the non-TDP tumor sets (Fig. S7*D*), raising the hypothesis that the TDP is an exquisite feature of *BRCA1* loss and not of *BRCA2* loss.

When taken together, these results suggest that a combination of *TP53* loss-of-function mutation, *BRCA1* reduced expression/activity, and overexpression of DNA replication and cell cycle genes may be required for TDP generation.

We have established that certain multigene expression changes are strongly associated with TDP tumors. We asked whether changes were a result of the TDs or preceded the induction of these structural mutations. Of the 23 DEGs involved in DNA replication and cell cycle associated with TDP, only four (*CALR*, *CCNE1*, *RAD51*, and *TK1*) were also found inside TDs in multiple TDP samples but at modest frequencies of <5% (Dataset S5). If we removed the TDP tumor samples harboring physical TDs spanning these four differentially expressed DNA replication and cell cycle genes, the association of these genes with the TDP remains statistically significant (P < 1.0E-04; Fig. S8). This observation suggests that their overexpression is likely to be engaged in the establishment of the TDP and is not simply a consequence of the phenotype.

TDP as a Genomic Marker for Drug Sensitivity. We explored whether the TDP could represent a marker for drug sensitivity by searching the Genomics of Drug Sensitivity in Cancer database (32) for drugs and compounds that differed in their effect between the TDP and non-TDP breast cancer cell lines. Interestingly, cisplatin was among six drugs showing a significant positive correlation between TDP scores (computed based on available WGS data) and IC_{50} values (Dataset S8). Given the utility of platinum-based therapeutics as neoadjuvants in the clinical management of patients with TNBC (33–35), and the reported association between platinum-based treatment clinical success and a "BRCA-ness" molecular profile (36–38), we hypothesized that the TDP subset of TNBCs may be characterized by a better response to platinum-based chemotherapy. We therefore tested a total of 14 genomically characterized TNBC cell lines and found significant negative correlations between IC_{50} values relative to both cisplatin and carboplatin treatments and the TDP score (R = −0.57, P = 0.032 for cisplatin; R = −0.58, P = 0.029 for carboplatin) (Fig. 5*A*). By contrast, olaparib, an inhibitor of the poly ADP ribose polymerase (PARP) shown to have antitumor activity in patients with BRCA-mutated cancer (39, 40), did not show any significant association with the TDP score when tested on our panel of TNBC cell lines (Dataset S9), suggesting that the sensitivity of TDP tumors to cisplatin may not be exclusively related to the mutational status of *BRCA1* or *BRCA2*.

Remarkably, although the levels of *BRCA1* expression correlate with the TDP score in the TNBC cell lines examined, we did not observe any significant association between *BRCA1* levels and either cisplatin or carboplatin IC_{50} values (Dataset S9). This finding indicates that platinum sensitivity correlates more directly with the TDP score than with *BRCA1* expression levels and that the TDP score, which is modulated by other genes in addition to *BRCA1*, may be a key genomic predictor of cisplatin sensitivity in TNBC.

We explored this hypothesis further by testing the in vivo response to cisplatin treatment in eight independent patient-derived xenograft (PDX) models of TNBC. Following a 3-wk-long cisplatin regimen, four of the five TDP PDX models showed remarkable levels of tumor shrinkage, including two complete responses (>80% average tumor shrinkage across all animals in the treatment arm, PDX3 and PDX6) and two partial responses (30–80% average tumor shrinkage, PDX7 and PDX10) (Fig. 5*B*). On the contrary, none of the three non-TDP models analyzed exhibited a reduction of the original tumor volume after 3 wk of treatment, and all three responses to the cisplatin regimen were classified as progressive disease (>20% average tumor

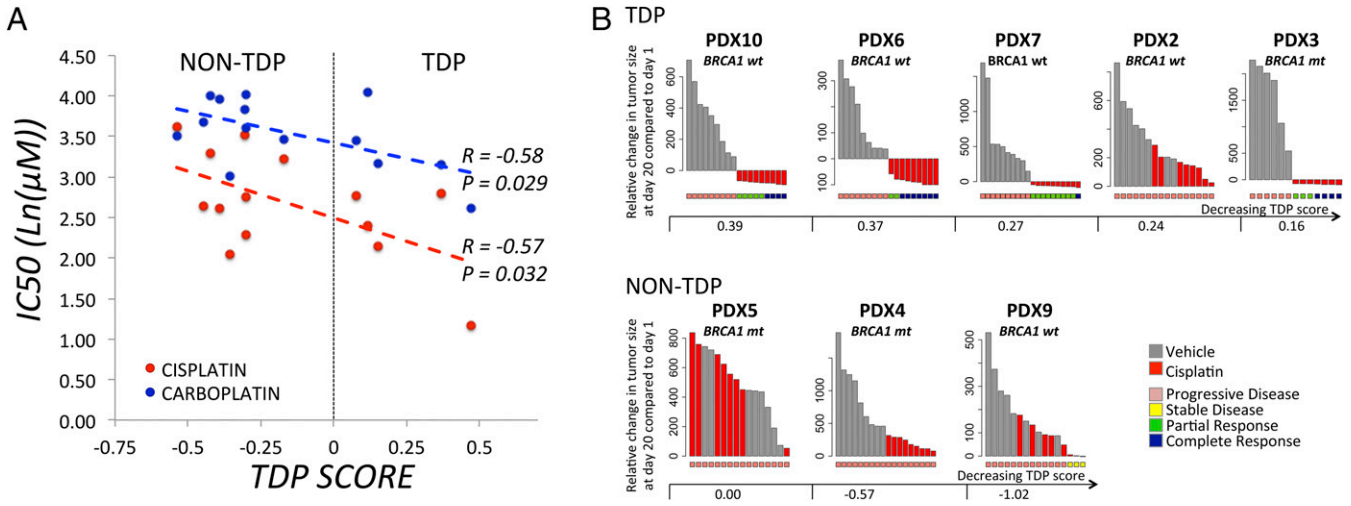

Fig. 5. TDP as a genomic marker for drug sensitivity. (*A*) TDP scores correlate with cisplatin or carboplatin sensitivities in TNBC cell lines. Pearson correlation coefficients (*R*) and their corresponding *P* values are reported in the graph. Ln, natural logarithm. (*B*) TDP scores associate with cisplatin sensitivity in vivo. Waterfall plots representing cisplatin response for eight TNB PDX models sorted by decreasing values of TDP scores are shown. Response calls are indicated underneath each bar and were computed based on adapted Response Evaluation Criteria in Solid Tumors (RECIST) criteria as described in *SI Materials and Methods*.

growth; Fig. 5*B*). Thus, in both established cell lines and in vivo PDXs, TDP status is strongly associated with cisplatin sensitivity.

Discussion

Recent studies have described previously unrecognized massive structural aberration events occurring on a genome-wide scale in human cancer (4, 5, 10, 41). A fundamental challenge is to define a quantitative metric to identify these global genomic configurations in cancer samples systematically and to investigate the role they play in tumorigenesis (42). Here, we devised a mathematical approach to the unbiased recognition of the TDP (9, 10, 15, 19). By applying this TDP scoring metric to a collection of ∼3,000 tumors with genomic data (WGS and/or SNP array), we provide statistical evidence that the TDP is enriched in specific tumor types, suggesting a distinct biological mechanism underlying this phenotype that cuts across histological subtypes (Fig. S9*A*).

The mechanisms for the generation of segmental TDs have been previously explored in *Saccharomyces cerevisiae* (43). Green et al. (43) have shown how defects in the molecular machinery responsible for preventing DNA rereplication can result in head-to-tail segmental duplications in yeast. Notably, the TDs in this study were mediated by NAHR between yeast transposon repetitive elements, a mechanism distinct from the MH-mediated mechanisms that dominate in the TDP tumors we have analyzed here. The authors proposed a mechanism by which the increase in copy number of chromosomal segments can result from the molecular repair of stalled rereplication bubble structures emerging in a permissive molecular background (e.g., following the deregulation of DNA replication proteins). They name this process rereplication-induced gene amplification (RRIGA) and speculate that it may play a critical role in oncogenesis (43, 44). Koszul et al. (45) and Koszul and coworkers (46) identified an MH-mediated *POL32*-dependent replicative mechanism underlying segmental TD formation in *S. cerevisiae*. A genetic analysis of MH-mediated TD formation in *Escherichia coli* by Slack et al. (47) implicated stalled replication as a trigger to the formation of these TDs. These observations are today collectively termed MMBIR (13, 14). Costantino et al. (48) demonstrated the significant enrichment of short-span copy number gains (<200 kb) in an artificial model of DNA replication stress induced through the ectopic overexpression of the *CCNE1* gene in the U2OS human osteosarcoma cell line. Our work supports these observations in a spontaneous human cancer setting. Indeed, the size range of the DNA duplications generated via RRIGA in yeast and via *CCNE1* overexpression in the U2OS cell line matches the size range of the TDs found in our TDP samples, which we have shown to be characterized by the significant overexpression of replication initiation genes, including *CDT1* and *CCNE1* (Fig. 4*C*). We therefore speculate that the mechanism of TD formation in the TDP chromotype may entail replicative mechanisms, such as MMBIR.

Whereas cancers with high amplification of a single locus in non-TDP tumors depend on a dominant driver oncogene, such as *ERBB2* or *MYC*, the TDP is unusual in that there does not appear to be a discernable single cancer driver gene targeted by the TDP. Rather, different combinations of many potential drivers appear to be affected by the widespread genomic distribution of TDs. Indeed, in our analysis of genes perturbed by TDs in TDP, we could not find any individual gene that appears to be affected in more than 15.5% of the samples examined (Dataset S5). However, the TDP configuration generates changes that affect the expression and function of hundreds of genes in a distributed manner within each tumor. Thus, TDP tumors may derive selective growth advantage from a systemic process, namely, genome-wide segmental TD formation, which simultaneously targets many cancer genes distributed across the genome. In seeking to uncover the root genetic aberrations that may underlie the induction of the TDP, we looked at the gene expression and mutational profiles that are frequently found and most strongly associated with the TDP across a number of tumor types. Our findings suggest that the TDP is induced by specific combinations of gene perturbations that (*i*) cause the loss of genome integrity (i.e., loss of *TP53* and *BRCA1*) and (*ii*) drive the augmented expression of cell cycle and DNA replication genes (e.g., increased activity of *CCNE1*, *CDT1*). In fact, combinations of these TDP-associated gene perturbations occur remarkably more frequently in TDP than in non-TDP TNBC tumors (OR = 17.2, $P = 2.1E\text{-}05$; Fig. S9*B*). Earlier reports have suggested a BRCA1-independent mechanism for the TDP, based on the absence of *BRCA1* mutations in samples (breast and OV carcinomas) with a large number of TDs (15, 19). However, we observed a strong negative correlation between *BRCA1* gene expression and the TDP score, as well as the enrichment for BRCA1-defective tumors (assessed by the presence of somatic or germline mutations, or promoter hypermethylation) in TNBC and OV cancer (Fig. 4 *D*–*G*). This finding strongly supports a

previously unrecognized critical role for *BRCA1* loss of function in the induction of the TDP.

Finally, we find that the quantitative assessment of the TDP may have clinical relevance. We describe an association between the extent of TDP and greater sensitivity to platinum-based chemotherapy both in cell lines and in PDXs. It has been reported that breast tumors with perturbations of *BRCA1* respond better to cisplatin treatment (37). Although our observations in vitro suggest that cisplatin sensitivity is better correlated with the TDP score than *BRCA1* levels or mutational status, we suggest that the TDP score integrates multiple genetic factors, such as *TP53* status and select driver gene expression (e.g., *CDT1*, *CCNE1*), which may be the genetic components needed for the sensitivity phenotype. Whereas recent neoadjuvant studies suggest that the effectiveness of cisplatin in TNBC is associated with loss of *BRCA1* by mutation or low expression (34, 36, 38), it may be that the TDP score is a more robust predictor of response to platinum-based chemotherapies independent of tumor type. Indeed, high TDP scores are enriched in TNBC, in OV cancer, and in the recently described cluster 4 endometrial carcinoma, which have been shown to share a similar transcriptional and molecular profile (23). Given the specific molecular determinants associated with the TDP across tumor types, it will be interesting to investigate the possible benefit of a cisplatin and PARP inhibitor combination in TDP tumors.

In summary, we envision that the TDP assessment may provide a unique genome sequence-based predictive marker for platinum-based drug sensitivity and allow for detailed interrogation of more precise mechanisms of cisplatin sensitivity.

Materials and Methods

WGS Datasets and TDP Classification. A catalog of somatic structural variation data was compiled from a number of WGS studies, comprising a total of 277 tumor samples, as listed in Dataset S1 (4, 9, 10, 15–22). We manually curated the available structural variation information (relative orientation and mapping coordinates of the discordant mate-pair or paired-end read clusters) from every individual study to classify each reported somatic event into one of the four basic rearrangements: deletion, TD, inversion, or interchromosomal translocation (49). For studies that reported structural variation coordinates relative to the hg18 reference human genome, a lift over to hg19 was performed using the Galaxy Lift-Over tool (https://usegalaxy.org). Previous attempts at describing the genomic features of the TDP have relied on a basic TD count or on the proportion of TDs relative to the total number of structural variations in a cancer genome (10, 15). These approaches lack in robustness, because they are prone to be influenced by observer and technical biases, such as sequencing coverage, and are not able to discriminate between the genome-wide TD prevalence that characterizes the TDP vs. abnormal TD accumulation in a few functional genomic loci, previously described in association with focal amplification (9, 16). Our proposed metric to calculate the TDP score is described in the main text. A visualization of the TDP score distribution density plot across all samples suggested a multimodal distribution (Fig. S1*A*). We used the *normalmixEM* function of the mixtools package in R to fit different numbers of mixture components (up to five) to the TDP score value distribution (50), using default estimates as the starting values for the iterative procedure. We compared the resulting mixture model estimates using the Bayesian information criterion and found that a trimodal distribution corresponded to the optimal fit.

TCGA Genomic Datasets. Affymetrix SNP 6.0 CNV datasets for primary tumor tissues were downloaded from the TCGA Data Portal in the form of level 3 CNV data type (CNV segments). Primary tumor samples from the TCGA breast invasive carcinoma dataset were classified as TNBC (TNB) or non-TNBC (NTNB), according to TCGA clinical annotations (28) (https://tcga-data.nci.nih.gov/tcga/).

TCGA somatic mutation data for the TNB, NTNB, OV, and UCEC datasets were downloaded from the UCSC Cancer Genomic Browser (https://genome-cancer.ucsc.edu) as gene-based somatic mutation calls generated by the TCGA PANCANCER Analysis Working Group. For each sample, any gene affected by at least one nonsilent somatic mutation (nonsense, missense, short insertion/deletion, splice site mutation, stop codon read-through) was considered somatically mutated.

RNA-sequencing (RNA-seq) gene expression data for the TNB, NTNB, OV, and UCEC datasets were downloaded from the TCGA Data Portal in the form of level 3 RSEM raw expression estimates, generated using the TCGA RNA Sequencing Version 2 analysis pipeline. Raw gene read counts were then scale-normalized using the trimmed mean of M-values normalization method before being converted into log counts per million with associated precision weights using the voom transformation included in the limma package in R (51).

Detection of TD-Like Features Based on Copy Number Profiling. Based on the assumption that an isolated TD within any given genomic locus will result in a chromosomal segment with uniform, increased copy number compared with its two adjacent genomic regions, we scanned SNP array genomic data for CNV profiles indicative of TD-like features (i.e., copy number segments with a length ranging from 1 kb to 2 Mb, characterized by a copy number increase of one or more units and flanked by segments of equal copy number) (15) (Fig. S2*A*). The identified TD-like features were then used to compute TDP scores following the same metric and threshold applied for WGS data (as described in *Results*).

Analysis of Differential Gene Expression. To identify DEGs between any two given groups of samples, the RNA-seq expression dataset was first filtered and only genes whose expression value was >1 in at least $n - 1$ samples [with n = number of samples in the smallest sample group (i.e., TDP, non-TDP)] were retained for further analysis. Sample group comparisons were carried out using the moderated t statistic of the limma package in R (51). A false discovery rate-adjusted P-value threshold of 0.05 was used to identify DEGs.

GO Enrichment Analysis. Gene enrichment analyses for GO terms were carried out using the topGO package in R (52). Briefly, predefined lists of interesting genes were tested for their enrichment in GO terms against the all-gene background using Fisher's exact test as the test statistic and the eliminating genes (*elim*) algorithm as the method for GO graph structure. GO terms with less than 10 annotated genes were removed from the analysis.

Cell Culture and IC_{50} Determination. All of the cell lines were purchased from the American Type Culture Collection. They were authenticated by short tandem repeat DNA profiling and regularly tested for *Mycoplasma* contamination using the MycoAlert PLUS Mycoplasma Detection Kit (Lonza). MB436, HCC38, HCC1187, HCC1395, MDA-MB231, HCC1937, HCC1599, HCC1143, HCC70, DU4475, MDA-MB157, and HCC1806 were maintained in RPMI with 10% (vol/vol) FBS. BT549 was maintained in DMEM with 10% (vol/vol) FBS, and Hs578T was maintained in DMEM with 10% (vol/vol) FBS and 0.01 mg/mL bovine insulin. IC_{50} value determinations were obtained by plating target cells in 96-well plates at a density of 1–5 $\times$ 10^3 cells per well. After 24 h, cisplatin (Santa Cruz Biotechnology, Inc.) or carboplatin (Selleck Chemicals) was added in triplicate wells to the culture medium in half-log serial dilutions in the range of 3.3 nM to 100 μM. Cells were incubated for 72 h before assessing cell viability using a WST-8 assay (Dojindo Molecular Technologies, Inc.). Absorbance values were normalized to control wells (medium only), and IC_{50} values were calculated using the IC_{50} R package (53). Two independent replicate experiments were carried out for each cell line and each treatment, and the average IC_{50} value from the two experiments was used for the analysis.

WGS of TNBC Cell Lines. Cell line genomic DNA was isolated from ~1 $\times$ 10^6 cells using a DNeasy Kit (Qiagen) and fragmented using Covaris E220 (Covaris) to a range of sizes centered on 500 bp. Paired-end DNA libraries were constructed using a NEBNext DNA Library Prep Master Mix set for Illumina (New England BioLabs), including a bead-based size selection to select for inserts with an average size of 500 bp and 10 cycles of PCR. The resulting libraries were quantified by quantitative PCR and pooled in groups of two before being sequenced on one lane of an Illumina HiSeq 2500 platform. Fastq files were paired and run through the next-generation sequencing (NGS) quality control (QC) Toolkit (version 2.3; IlluQC_PRLL.pl) with a quality control cutoff of 30, before alignment to the human reference genome (National Center for Biotechnology Information Build 37 from the 1000 Genomes Project) using bwa (version 0.7.4) and default parameters (bwa mem). The HydraMulti algorithm (54) was used to predict structural variation events. All datasets were analyzed at the same time, and structural variation events were filtered as described by Malhotra et al. (55). Only structural variations exclusive to individual datasets were considered for further analysis. WGS data are freely available from the Sequence Read Archive database (www.ncbi.nlm.nih.gov/sra) under project ID SRP057179.

Animal Work. All animal procedures were approved by The Jackson Laboratory Institutional Animal Care and Use Committee (IACUC) under protocol number 12027.

Additional methods information can be found in *SI Materials and Methods*.

ACKNOWLEDGMENTS. WGS library preparation and sequence data analysis were performed by Scientific Services at The Jackson Laboratory, Bar Harbor, ME. Research reported in this publication was partially supported by the National Cancer Institute under Award P30CA034196. J.H.C. was supported by the National Human Genome Research Institute and National Cancer Institute of the NIH under Awards R21HG007554 and R21CA184851.

1. Hanahan D, Weinberg RA (2011) Hallmarks of cancer: The next generation. *Cell* 144(5):646–674.
2. Yates LR, Campbell PJ (2012) Evolution of the cancer genome. *Nat Rev Genet* 13(11): 795–806.
3. Stratton MR, Campbell PJ, Futreal PA (2009) The cancer genome. *Nature* 458(7239): 719–724.
4. Baca SC, et al. (2013) Punctuated evolution of prostate cancer genomes. *Cell* 153(3): 666–677.
5. Stephens PJ, et al. (2011) Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell* 144(1):27–40.
6. Zhang CZ, Leibowitz ML, Pellman D (2013) Chromothripsis and beyond: Rapid genome evolution from complex chromosomal rearrangements. *Genes Dev* 27(23): 2513–2530.
7. Zhang F, Carvalho CM, Lupski JR (2009) Complex human chromosomal and genomic rearrangements. *Trends Genet* 25(7):298–307.
8. Gisselsson D, et al. (2000) Chromosomal breakage-fusion-bridge events cause genetic intratumor heterogeneity. *Proc Natl Acad Sci USA* 97(10):5357–5362.
9. Inaki K, et al. (2014) Systems consequences of amplicon formation in human breast cancer. *Genome Res* 24(10):1559–1571.
10. Stephens PJ, et al. (2009) Complex landscapes of somatic rearrangement in human breast cancer genomes. *Nature* 462(7276):1005–1010.
11. Crasta K, et al. (2012) DNA breaks and chromosome pulverization from errors in mitosis. *Nature* 482(7383):53–58.
12. Maciejowski J, Li Y, Bosco N, Campbell PJ, de Lange T (2015) Chromothripsis and Kataegis Induced by Telomere Crisis. *Cell* 163(7):1641–1654.
13. Willis NA, Rass E, Scully R (2015) Deciphering the Code of the Cancer Genome: Mechanisms of Chromosome Rearrangement. *Trends Cancer* 1(4):217–230.
14. Zhang CZ, et al. (2015) Chromothripsis from DNA damage in micronuclei. *Nature* 522(7555):179–184.
15. Ng CK, et al. (2012) The role of tandem duplicator phenotype in tumour evolution in high-grade serous ovarian cancer. *J Pathol* 226(5):703–712.
16. Hillmer AM, et al. (2011) Comprehensive long-span paired-end-tag mapping reveals characteristic patterns of structural variations in epithelial cancer genomes. *Genome Res* 21(5):665–675.
17. Natrajan R, et al. (2012) A whole-genome massively parallel sequencing analysis of BRCA1 mutant oestrogen receptor-negative and -positive breast cancers. *J Pathol* 227(1):29–41.
18. Nik-Zainal S, et al.; Breast Cancer Working Group of the International Cancer Genome Consortium (2012) Mutational processes molding the genomes of 21 breast cancers. *Cell* 149(5):979–993.
19. McBride DJ, et al. (2012) Tandem duplication of chromosomal segments is common in ovarian and breast cancer genomes. *J Pathol* 227(4):446–455.
20. Imielinski M, et al. (2012) Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell* 150(6):1107–1120.
21. Yang L, et al. (2013) Diverse mechanisms of somatic structural variations in human cancer genomes. *Cell* 153(4):919–929.
22. Grzeda KR, et al. (2014) Functional chromatin features are associated with structural mutations in cancer. *BMC Genomics* 15:1013.
23. Kandoth C, et al.; Cancer Genome Atlas Research Network (2013) Integrated genomic characterization of endometrial carcinoma. *Nature* 497(7447):67–73.
24. Hastings PJ, Lupski JR, Rosenberg SM, Ira G (2009) Mechanisms of change in gene copy number. *Nat Rev Genet* 10(8):551–564.
25. De S, Michor F (2011) DNA replication timing and long-range DNA interactions predict mutational landscapes of cancer genomes. *Nat Biotechnol* 29(12):1103–1108.
26. Sima J, Gilbert DM (2014) Complex correlations: Replication timing and mutational landscapes during cancer and genome evolution. *Curr Opin Genet Dev* 25:93–100.
27. Chen CL, et al. (2010) Impact of replication timing on non-CpG and CpG substitution rates in mammalian genomes. *Genome Res* 20(4):447–457.
28. Cancer Genome Atlas N; Cancer Genome Atlas Network (2012) Comprehensive molecular portraits of human breast tumours. *Nature* 490(7418):61–70.
29. Cancer Genome Atlas Research Network (2011) Integrated genomic analyses of ovarian carcinoma. *Nature* 474(7353):609–615.
30. Caillat C, Perrakis A (2012) Cdt1 and geminin in DNA replication initiation. *Subcell Biochem* 62:71–87.
31. Powell SK, et al. (2015) Dynamic loading and redistribution of the Mcm2-7 helicase complex through the cell cycle. *EMBO J* 34(4):531–543.
32. Yang W, et al. (2013) Genomics of Drug Sensitivity in Cancer (GDSC): A resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res* 41(Database issue): D955–D961.
33. Sikov WM, et al. (2015) Impact of the addition of carboplatin and/or bevacizumab to neoadjuvant once-per-week paclitaxel followed by dose-dense doxorubicin and cyclophosphamide on pathologic complete response rates in stage II to III triple-negative breast cancer: CALGB 40603 (Alliance). *J Clin Oncol* 33(1):13–21.
34. von Minckwitz G, Martin M (2012) Neoadjuvant treatments for triple-negative breast cancer (TNBC). *Ann Oncol* 23(Suppl 6):vi35–vi39.
35. von Minckwitz G, et al. (2014) Neoadjuvant carboplatin in patients with triple-negative and HER2-positive early breast cancer (GeparSixto; GBG 66): A randomised phase 2 trial. *Lancet Oncol* 15(7):747–756.
36. Davis SL, Eckhardt SG, Tentler JJ, Diamond JR (2014) Triple-negative breast cancer: Bridging the gap from cancer genomics to predictive biomarkers. *Ther Adv Med Oncol* 6(3):88–100.
37. Stefansson OA, Villanueva A, Vidal A, Martí L, Esteller M (2012) BRCA1 epigenetic inactivation predicts sensitivity to platinum-based chemotherapy in breast and ovarian cancer. *Epigenetics* 7(11):1225–1229.
38. Silver DP, et al. (2010) Efficacy of neoadjuvant Cisplatin in triple-negative breast cancer. *J Clin Oncol* 28(7):1145–1153.
39. Fong PC, et al. (2009) Inhibition of poly(ADP-ribose) polymerase in tumors from BRCA mutation carriers. *N Engl J Med* 361(2):123–134.
40. Farmer H, et al. (2005) Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature* 434(7035):917–921.
41. Liu P, et al. (2011) Chromosome catastrophes involve replication mechanisms generating complex genomic rearrangements. *Cell* 146(6):889–903.
42. Korbel JO, Campbell PJ (2013) Criteria for inference of chromothripsis in cancer genomes. *Cell* 152(6):1226–1236.
43. Green BM, Finn KJ, Li JJ (2010) Loss of DNA replication control is a potent inducer of gene amplification. *Science* 329(5994):943–946.
44. Finn KJ, Li JJ (2013) Single-stranded annealing induced by re-initiation of replication origins provides a novel and efficient mechanism for generating copy number expansion via non-allelic homologous recombination. *PLoS Genet* 9(1):e1003192.
45. Koszul R, Caburet S, Dujon B, Fischer G (2004) Eucaryotic genome evolution through the spontaneous duplication of large chromosomal segments. *EMBO J* 23(1):234–243.
46. Payen C, Koszul R, Dujon B, Fischer G (2008) Segmental duplications arise from Pol32-dependent repair of broken forks through two alternative replication-based mechanisms. *PLoS Genet* 4(9):e1000175.
47. Slack A, Thornton PC, Magner DB, Rosenberg SM, Hastings PJ (2006) On the mechanism of gene amplification induced under stress in Escherichia coli. *PLoS Genet* 2(4):e48.
48. Costantino L, et al. (2014) Break-induced replication repair of damaged forks induces genomic duplications in human cells. *Science* 343(6166):88–91.
49. Medvedev P, Stanciu M, Brudno M (2009) Computational methods for discovering structural variation with next-generation sequencing. *Nat Methods* 6(11, Suppl):S13–S20.
50. Benaglia T, Chauveau D, Hunter DR, Young DS (2009) mixtools: An R Package for Analyzing Finite Mixture Models. *J Stat Softw* 32(6):1–29.
51. Smyth GK (2004) Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* 3:e3.
52. Alexa A, Rahnenführer J, Lengauer T (2006) Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics* 22(13):1600–1607.
53. Frommolt P, Thomas RK (2008) Standardized high-throughput evaluation of cell-based compound screens. *BMC Bioinformatics* 9:475.
54. Lindberg MR, Hall IM, Quinlan AR (2015) Population-based structural variation discovery with Hydra-Multi. *Bioinformatics* 31(8):1286–1289.
55. Malhotra A, et al. (2013) Breakpoint profiling of 64 cancer genomes reveals numerous complex rearrangements spawned by homology-independent mechanisms. *Genome Res* 23(5):762–776.
56. Futreal PA, et al. (2004) A census of human cancer genes. *Nat Rev Cancer* 4(3):177–183.
57. Forbes SA, et al. (2008) The Catalogue of Somatic Mutations in Cancer (COSMIC). *Curr Protoc Hum Genet* Chapter 10:Unit 10.11.
58. Higgins ME, Claremont M, Major JE, Sander C, Lash AE (2007) CancerGenes: A gene selection resource for cancer genome projects. *Nucleic Acids Res* 35(Database issue): D721–D726.
59. Zhao M, Sun J, Zhao Z (2013) TSGene: A web resource for tumor suppressor genes. *Nucleic Acids Res* 41(Database issue):D970–D976.
60. Santarius T, Shipley J, Brewer D, Stratton MR, Cooper CS (2010) A census of amplified and overexpressed human cancer genes. *Nat Rev Cancer* 10(1):59–64.
61. Solimini NL, et al. (2012) Recurrent hemizygous deletions in cancers may optimize proliferative potential. *Science* 337(6090):104–109.
62. Consortium EP; ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* 489(7414):57–74.
63. Bernstein BE, et al. (2010) The NIH Roadmap Epigenomics Mapping Consortium. *Nat Biotechnol* 28(10):1045–1048.
64. Feng J, Liu T, Qin B, Zhang Y, Liu XS (2012) Identifying ChIP-seq enrichment using MACS. *Nat Protoc* 7(9):1728–1740.
65. Zhang Y, et al. (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 9(9):R137.
66. Shultz LD, et al. (2014) Human cancer growth and therapy in immunodeficient mouse models. *Cold Spring Harb Protoc* 2014(7):694–708.
67. Conway T, et al. (2012) Xenome–a tool for classifying reads from xenograft samples. *Bioinformatics* 28(12):i172–i178.
68. Rausch T, et al. (2012) DELLY: Structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* 28(18):i333–i339.
69. Li B, Dewey CN (2011) RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12:323.
70. Eisenhauer EA, et al. (2009) New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). *Eur J Cancer* 45(2):228–247.
71. Cerami E, et al. (2012) The cBio cancer genomics portal: An open platform for exploring multidimensional cancer genomics data. *Cancer Discov* 2(5):401–404.
72. Gao J, et al. (2013) Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* 6(269):pl1.