

The History of Chromosomal Instability in Genome-Doubled Tumors



Toby M. Baker^{1,2,3}, Siqi Lai², Andrew R. Lynch², Tom Lesluyes¹, Haixi Yan^{1,2}, Huw A. Ogilvie², Annelien Verfaillie¹, Stefan Dentre⁴, Amy L. Bowes¹, Nischalan Pillay^{5,6}, Adrienne M. Flanagan^{5,6}, Charles Swanton^{1,7,8}, Paul T. Spellman³, Maxime Tarabichi^{1,9}, and Peter Van Loo^{1,2,10}

ABSTRACT

Tumors frequently display high chromosomal instability and contain multiple copies of genomic regions. Here, we describe Gain Route Identification and Timing In Cancer (GRITIC), a generic method for timing genomic gains leading to complex copy number states, using single-sample bulk whole-genome sequencing data. By applying GRITIC to 6,091 tumors, we found that non-parsimonious evolution is frequent in the formation of complex copy number states in genome-doubled tumors. We measured chromosomal instability before and after genome duplication in human tumors and found that late genome doubling was followed by an increase in the rate of copy number gain. Copy number gains often accumulate as punctuated bursts, commonly after genome doubling. We infer that genome duplications typically affect the landscape of copy number losses, while only minimally impacting copy number gains. In summary, GRITIC is a novel copy number gain timing framework that permits the analysis of copy number evolution in chromosomally unstable tumors.

SIGNIFICANCE: Complex genomic gains are associated with whole-genome duplications, which are frequent across tumors, span a large fraction of their genomes, and are linked to poorer outcomes. GRITIC infers when these gains occur during tumor development, which will help to identify the genetic events that drive tumor evolution.

See related commentary by Taylor, p. 1766

¹The Francis Crick Institute, London, United Kingdom. ²Department of Genetics, The University of Texas MD Anderson Cancer Center, Houston, Texas. ³Department of Medicine, University of California Los Angeles, Los Angeles, California. ⁴Division of AI in Oncology, German Cancer Research Centre DKFZ, Heidelberg, Germany. ⁵Research Department of Pathology, Cancer Institute, University College London, London, United Kingdom. ⁶Department of Cellular and Molecular Pathology, Royal National Orthopaedic Hospital NHS Trust, Stanmore, United Kingdom. ⁷Cancer Research UK Lung Cancer Centre of Excellence, University College London Cancer Institute, London, United Kingdom. ⁸Department of Oncology, University College London Hospitals, London, United Kingdom. ⁹Institute for Interdisciplinary Research (IRIBHM), Université Libre de Bruxelles, Brussels, Belgium. ¹⁰Department of

Genomic Medicine, The University of Texas MD Anderson Cancer Center, Houston, Texas.

M. Tarabichi and P. Van Loo have jointly supervised the article.

Corresponding Author: Peter Van Loo, Department of Genetics, The University of Texas Anderson Cancer Center, Unit 1010, 1515 Holcombe Boulevard, Houston, TX 77030. E-mail: pvanloo@mdanderson.org

Cancer Discov 2024;14:1810-22

doi: 10.1158/2159-8290.CD-23-1249

This open access article is distributed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) license.

©2024 The Authors; Published by the American Association for Cancer Research

INTRODUCTION

Genomic copy number gains and losses, caused by chromosomal instability (CIN), are common somatic alterations in cancer (1, 2). While somatic single nucleotide variants (SNV) and indels linked to cancer drivers are found in ostensibly healthy tissues, copy number events rarely occur in normal cells (3–6). Identifying when copy number events occur is important for screening purposes and for gaining an understanding of the key molecular mechanisms underlying cancer development.

Of particular interest is the evolution of copy number events in tumors with the most aberrant genomes, as CIN is linked to poorer outcomes (7). Tumors that have undergone whole-genome duplication (WGD) often show elevated numbers of copy number gains and losses (8–10), which may arise through multiple mechanisms, including chromosomal missegregation from centrosomal amplification (11) and a shortage of replication machinery proteins immediately following WGD (12). While copy number gains and losses support further CIN, the temporal relationship between WGDs and CIN is difficult to assess from single-timepoint biopsies in human tumors. The extent of genomic aberration, often used as an indirect proxy for CIN, does not convey the temporal dynamics that define CIN.

To observe the evolution of genomic gains in genome-doubled tumors, the timing of copy number gains and WGDs relative to the accumulation of SNVs can be inferred from whole-genome sequencing data (13–15). Clonal copy number gains, which are present in every tumor cell, can be placed on a timeline from 0 to 1, where 0 represents conception and 1 represents the end of the tumor's clonal evolutionary period. Previous approaches that have used this principle to time copy number gains (16–18) were unable to fully time gains leading to complex copy number states (those with three or more copies of one parental allele). This is due to higher ambiguity in the route history of these complex states relative to simpler states. Either the most parsimonious route history was assumed (17, 18) or these states were not timed at all (16). Recently, two new methods have been developed to time much more complex states than previous approaches but they either only provide bounds on the timing of the first and last gains for a segment (19) or still require an assumption of parsimony (20).

Here, we present Gain Route Identification and Timing In Cancer (GRITIC), a method that can time sequential gains leading to complex clonal copy number states, thereby elucidating the genome-wide evolution of gains in tumors with high CIN. As GRITIC is designed to time clonal copy number gains, it is well suited to unravel the evolution of the earliest genomic events in tumors, those that arise before the emergence of the tumor's most recent common ancestor. After filtering for minimum sample quality and WGD status, we applied GRITIC to a cohort of 1,751 primary tumors from the Pan-Cancer Analysis of Whole Genomes (PCAWG) dataset (21) and 4,340 metastases from the Hartwig Medical Foundation dataset (22). Surprisingly, we observed that the commonly held principle of maximum parsimony (i.e., that copy number states are formed through the simplest possible route) is frequently violated for complex copy number gains in WGD tumors. We found that punctuated bursts of gains, independent of WGD, were common across cancer types. We infer the rate of gains pre- and post-WGD across our cohort and observe that late WGD

causes an immediate increase in the rate of gains, a proxy for the rate of chromosomal instability. By considering the landscape of copy number events before and after a WGD, we found that WGD appears to have a low impact on the landscape of copy number gains but a greater impact on losses.

RESULTS

GRITIC Leverages SNVs to Time Complex Copy Number Gains

Tumors frequently gain additional copies of their genomic regions. In the Hartwig and PCAWG datasets, copy number gains affected an average of 48.6% of the tumor genomes (55.5% and 33.0%, respectively). Complex gains were common, with 27.8% of the gained genome having three or more copies of one parental allele on average (28.8% and 24.1% for Hartwig metastases and PCAWG primary tumors, respectively, Fig. 1A). The frequency of a given complex copy number state was inversely correlated with the largest number of copies of the parental allele for the state, known as the major copy number (Fig. 1A). Metastases also had a higher rate of WGD in our cohorts: 55.3% for Hartwig metastases compared to 31.2% for PCAWG primary tumors (Fig. 1B; Supplementary Fig. S1A), although this appears to be a phenomenon specific to certain cancer types (Supplementary Fig. S1B; refs. 8, 22). Although the difference in complex copy number fraction is largely explained by the higher proportion of WGD tumors, it was still higher in metastases when controlling for WGD frequency (Fig. 1C), likely reflecting increased CIN in metastatic cancers.

Clonal copy number gains can be quantitatively timed by considering SNVs in the gained region. When a gain occurs, all SNVs present on the gained allele are duplicated on the new copy (Fig. 1D). With the reasonable assumption that each base pair in the genome is mutated at most once (23), any SNV on multiple copies in a gained region must have occurred before the copy number gain. This principle can be used to infer the timing of the gain (Fig. 1E; Supplementary Methods; ref. 14). However, for more complex gains, further consideration of the possible routes that lead to these complex states is required. To accomplish this, we developed GRITIC, a new method that can identify, distinguish, and time the gains in these routes.

GRITIC uses a binary tree representation, conceptually similar to an earlier approach (24), to represent the gain history of a given segment. These representations can be used to calculate all possible routes (assuming, at most, a single WGD), resulting in a particular copy number state (Fig. 1F; Supplementary Methods). We found that the number of possible routes increases exponentially with the complexity of the copy-number state (Fig. 1G). Therefore, we limit GRITIC to the timing of copy number gains of segments with no more than 500 possible routes for WGD tumors (Supplementary Table S1). GRITIC uses a Bayesian Markov Chain Monte Carlo (MCMC) approach to infer the posterior probability of all possible route histories and the corresponding set of gain timings from SNV read counts for each gained segment. GRITIC is particularly suited to timing tumors with a WGD, as it uses the simultaneous occurrence of a WGD across all genomic regions as a constraint during inference to improve timing accuracy (Supplementary Fig. S2A and S2B; "Methods").

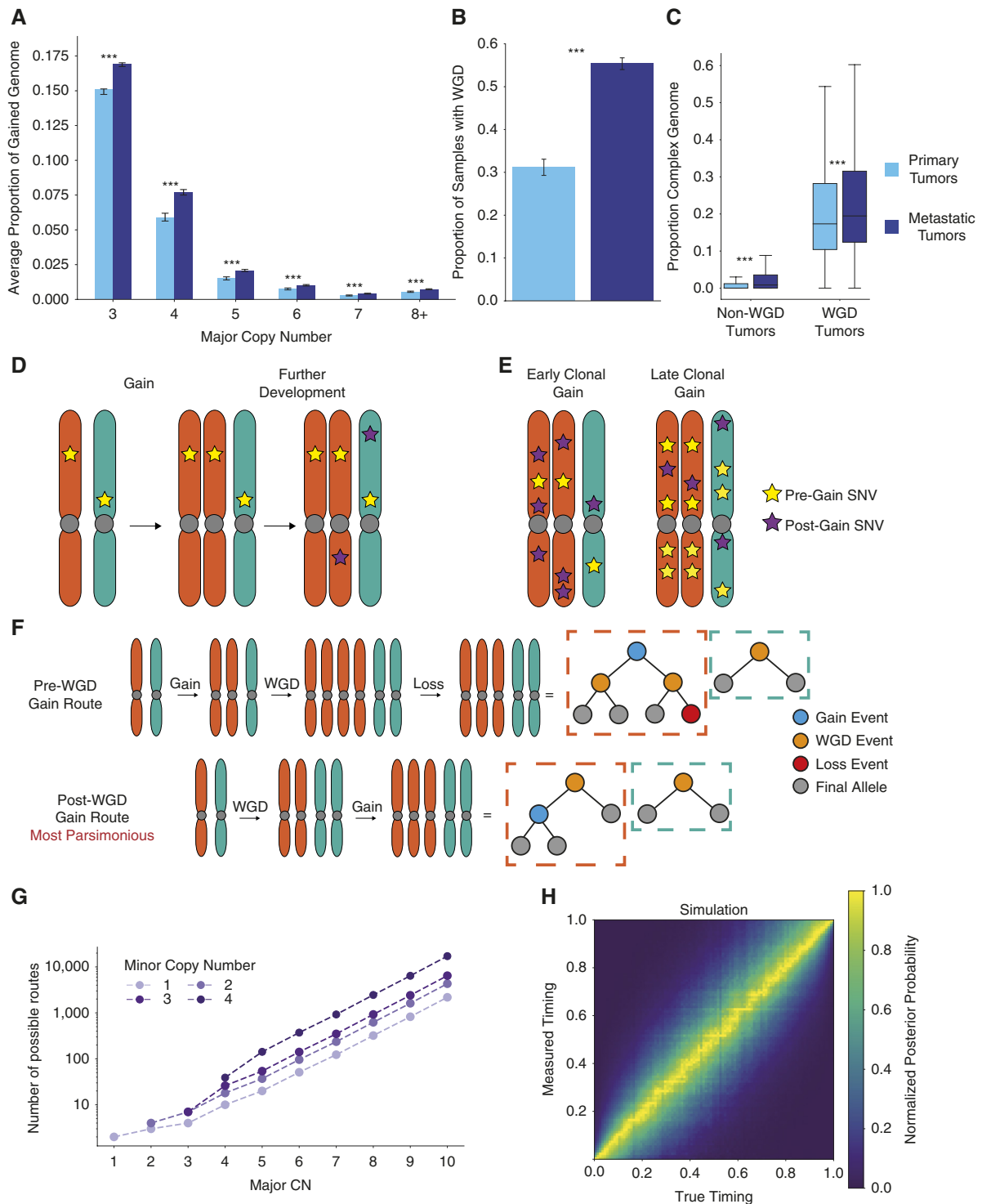


Figure 1. Principles of timing complex copy number gains. **A**, Average proportion of the genome with different major copy number states split by primary and metastatic cohorts. Statistical significance was calculated by permutation test and 95% confidence intervals by bootstrapping over samples. **B**, Proportion of tumor samples identified as WGD in primary and metastatic cohorts. Statistical significance is calculated by proportion test and 95% confidence intervals by normal approximation to a binomial proportion. **C**, Proportion of the genome with a major copy number of at least three in the primary and metastatic cohorts, split by WGD status. Statistical significance was calculated by permutation test and 95% confidence intervals by bootstrapping over samples. **D**, Schematic showing SNVs on a gained allele are duplicated by the gain, the principle underlying copy number gain timing. **E**, Schematic showing the difference in SNVs on multiple copies between a gain that occurs early in the clonal evolutionary period and a gain that occurs later. **F**, Binary tree representation of two possible routes that result in a 3 + 2 copy number state in a WGD tumor. The post-WGD route is the most parsimonious as it involves the fewest events. **G**, The number of theoretically distinguishable unique routes that can result in different allele-specific copy number states given a single WGD. **H**, Distribution of measured posterior probability on gain timing against true gain timing across a representative simulated cohort of complex gains. ***, $P < 0.001$.

We applied GRITIC to a realistic simulated cohort of WGD tumors (“Methods”) and found that with simulated tumor purity and sequencing coverage representative of the tumors in PCAWG and Hartwig, GRITIC can accurately measure the timing of all gains leading to complex states under different parsimony assumptions (Fig. 1H; Supplementary Fig. S3–S7; “Methods”). Although more sensitive to simulation and inference conditions compared to measuring the gain timing itself, GRITIC can also accurately estimate the probabilities of different gain routes (Supplementary Fig. S8–S11). As further validation, we tested GRITIC on patients with multiple samples in the Hartwig cohort and confirmed that shared gains (expected to have occurred earlier) showed earlier timing than gains unique to one sample (expected to have occurred later) in 78.9% (142/180) of cases (Supplementary Fig. S12A and S12B).

Non-Parsimony Is Common in WGD Tumor Gain Evolution

We then applied GRITIC to time the gains that led to 164,062 clonally gained regions across 6,091 tumors in the PCAWG and Hartwig datasets that pass our quality filters (Supplementary Methods). GRITIC reconstructs the timing of both the independent gains and the WGD (if present) in each sample and can time multiple sequential independent gains in the same genomic region (Fig. 2A). GRITIC produces a joint posterior distribution over gain timing and route histories considering all possible routes for each segment. Although different routes have distinct relationships between SNV multiplicity and gain timing, the timing of gains is generally concordant between different routes (Supplementary Figs. S13 and S14).

Parsimonious route histories have often been assumed for the development of copy number states in WGD tumors (17, 18). This is because the total number of allelic copies gained through WGD versus individual independent gains will vary between different routes, as will the number of losses required to make the route self-consistent (Supplementary Methods). Therefore, we sought to evaluate the assumption of parsimonious evolution in WGD tumors.

We first tested this assumption by reanalyzing copy number data from isogenic tetraploid colorectal cancer HCT-116 cell lines obtained from two different passages derived from a diploid progenitor (25, 26). By considering the change in copy number states between the two passages, we found that most events with a major copy number of three in the later passage arose through complex routes that would violate the assumption of parsimony if applied to the later passage in isolation (Fig. 2B). This result suggests that the assumption of parsimonious copy number evolution is often invalid.

We next used GRITIC to test this parsimonious evolution assumption. Owing to the inherent uncertainties in estimating copy number gain routes from SNVs (Supplementary Methods), GRITIC assigns an average posterior probability of 56.2% to non-parsimonious routes from a simulated set of tumors with completely parsimonious evolutionary histories (Supplementary Figs. S15A and S16). Although the model evidence used in GRITIC provides a natural penalty against additional independent gain timing parameters, it does

not penalize the number of losses implied by a given route. Thus, to ensure a conservative estimate of non-parsimony, we applied a penalty term to the number of events required for each route in WGD tumors (“Methods”). This penalty was fitted such that the average posterior probability of non-parsimonious routes was ~5% on a representative cohort of simulated WGD tumors with only parsimonious routes (Supplementary Figs. S17 and S18).

With this penalty term, we evaluated non-parsimonious evolution across genome-doubled tumors. Surprisingly, we found that non-parsimony was common: 29.8% of the total posterior probability on gained segments in WGD tumors was on non-parsimonious route histories, with 5.6% on routes with two or more additional events compared to the simplest route (Fig. 2C; Supplementary Figs. S15B and S16; $P < 0.001$, permutation test). Owing to our conservative penalty term, these are likely underestimates of non-parsimony in copy number evolution.

Non-parsimony occurs in agreement with known phenomena. Gains on chromosome 5q are known to be an initiating event in clear cell renal cell carcinoma, combined with the loss of 3p (27). Indeed, we found that 52/54 of the genome-doubled clear cell renal cell carcinomas had LOH loss of 3p, indicating that it likely occurred pre-WGD. In line with this, GRITIC inferred that gains on chromosome 5 with a major copy number of 3 are significantly more non-parsimonious (i.e., earlier, Supplementary Fig. S15C; Supplementary Methods) than the background for clear cell renal cell carcinoma (Fig. 2D; Supplementary Fig. S19; $P < 0.001$, permutation test). Conversely, gains on chromosome 5 with a major copy number of four were more likely parsimonious (i.e., earlier, Fig. 2D; Supplementary Fig. S15C; $P < 0.05$, permutation test). This effect was also observed when a non-parsimony penalty term was not applied (Supplementary Fig. S15D and S15E). More generally, we found that the frequency of pre-WGD gains for major copy number 3 and 4 states was correlated, both for segments within the same chromosome and across different samples (Supplementary Figs. S20 and S21).

Applying a penalty to the number of events provides a conservative estimate on the non-parsimonious evolution of complex gains. However, we found that, as expected, this causes the inferred probability of non-parsimonious evolution to be inaccurate for cohorts simulated to contain non-parsimonious routes (Supplementary Fig. S9). Therefore, for all subsequent analyses, we show the results of applying GRITIC without this non-parsimony penalty term and display the results with a penalty term in the Supplementary Information. In general, despite different route probabilities, we find that the results are highly consistent.

We find that the initial gains that occur independently of the WGD tend to occur earlier as the major copy number increases (“Methods”; Fig. 2E; Supplementary Fig. S22A–S22C). In contrast, the timing of all copy number gains that occur independently of WGD is much more uniformly distributed across mutation time for all copy number states (Fig. 2F; Supplementary Fig. S22). This suggests that moderate-level amplifications in tumor development generally begin early but accumulate further gains throughout the clonal evolutionary period.

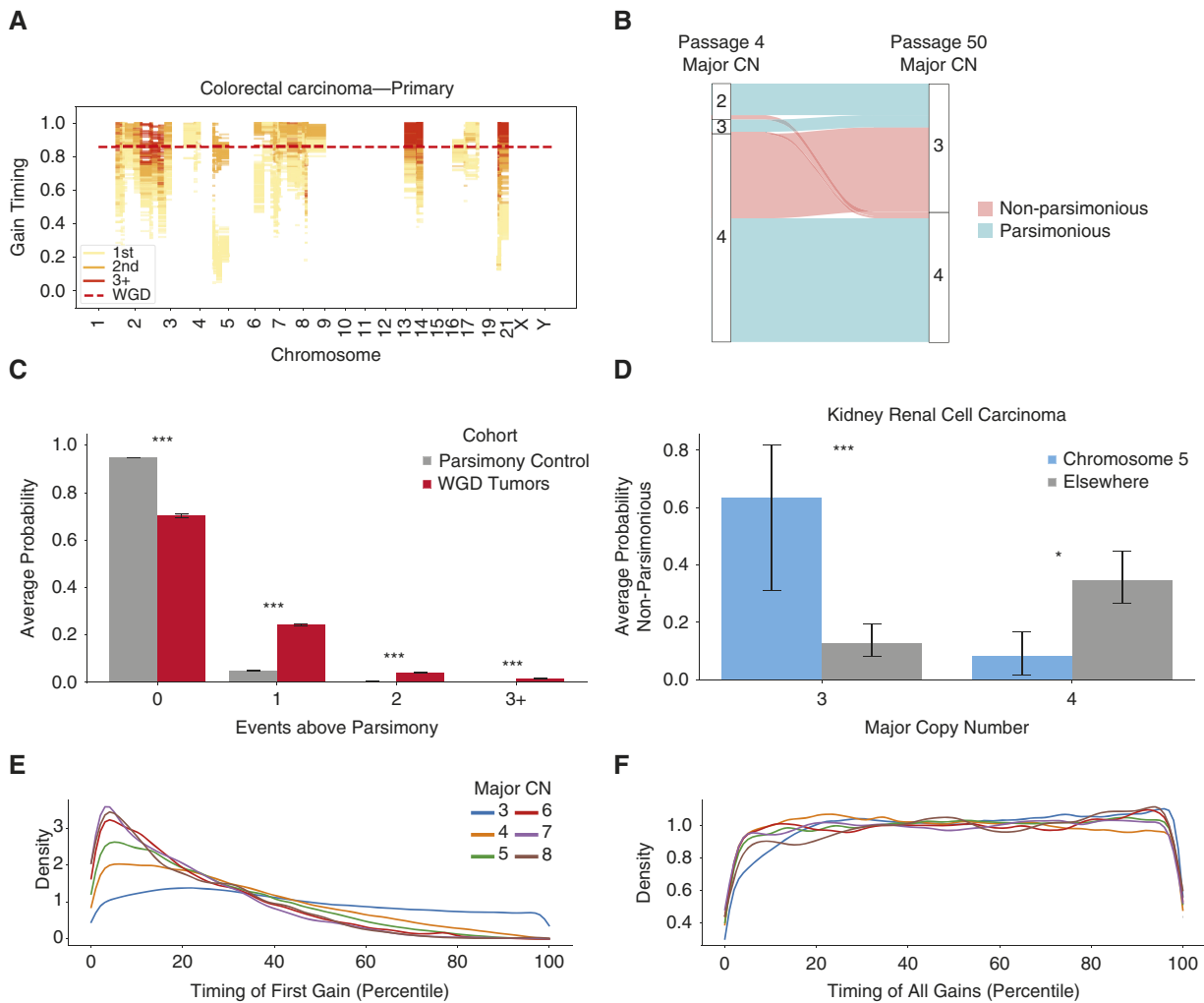


Figure 2. Non-parsimonious copy number evolution in cancer. **A**, Example posterior distribution of copy number gain timing in a whole-genome duplicated sample from Hartwig with GRITIC. 100 independent draws from the posterior sample for each gained segment are shown. **B**, The proportion of different copy number states at passage 4 that result in copy number states with a major copy number of 3 or 4 at passage 50 in four tetraploid clones in a colorectal cancer cell line. The flows are colored by whether the transition would be considered a parsimonious route by the copy number at passage 50. **C**, The average posterior probability on the number of additional events required to reach the final state over the most parsimonious route for complex gained states in the PCAWG and Hartwig cohort, compared to a simulated control where only parsimonious routes were included. A penalty on non-parsimony was applied during inference. Statistical significance is calculated with a permutation test and 95% confidence intervals by bootstrapping over samples. **D**, The average probability on non-parsimonious routes for gained segments in clear cell renal cell carcinoma, split by major copy number and gain location. A penalty on non-parsimony was applied during inference. Statistical significance was calculated with a permutation test and 95% confidence intervals were calculated by bootstrapping over samples. **E**, The timing distribution of the first gains in complex segments relative to other gains in the same sample, as defined by their quantile ranking within each sample, split by major copy number. **F**, The timing distribution of all first gains in complex segments relative to other gains in the same sample, as defined by their quantile ranking within each sample, split by major copy number. *, $P < 0.05$; ***, $P < 0.001$.

The Effect of Genome Doubling on the Rate of Chromosomal Instability

Next, we evaluated the effect of genome doubling on the rate of CIN. We analyzed the copy number profiles of 260 individual tumor cells in an undifferentiated soft tissue sarcoma that had undergone consecutive subclonal WGDs, as shown experimentally (Supplementary Methods). We found progressively higher rates of inter-copy number diversity in cells with each round of WGD (Fig. 3A; Supplementary Fig. S23A–S23C; $P < 0.001$, Mann–Whitney U Test; “Methods”), suggesting that WGD increases CIN even within the same tumor.

We used GRITIC to calculate the rate of independent gains, a proxy for the rate of CIN, relative to the occurrence of WGD across our cohort (“Methods”). We found that the gain rate increased after WGD in tumors with late genome doubling (Fig. 3B; Supplementary Fig. S24). Interestingly, the tumors with the earliest genome doubling show a different trend, exhibiting a higher gain rate before early WGD and a subsequent lower post-WGD gain frequency (Fig. 3B), an effect greater than expected from samples simulated with a uniform rate of gains (Supplementary Fig. S25; Supplementary Methods). This behavior may be driven primarily by breast tumors, which were enriched for early genome doubling (Supplementary Fig. S24),

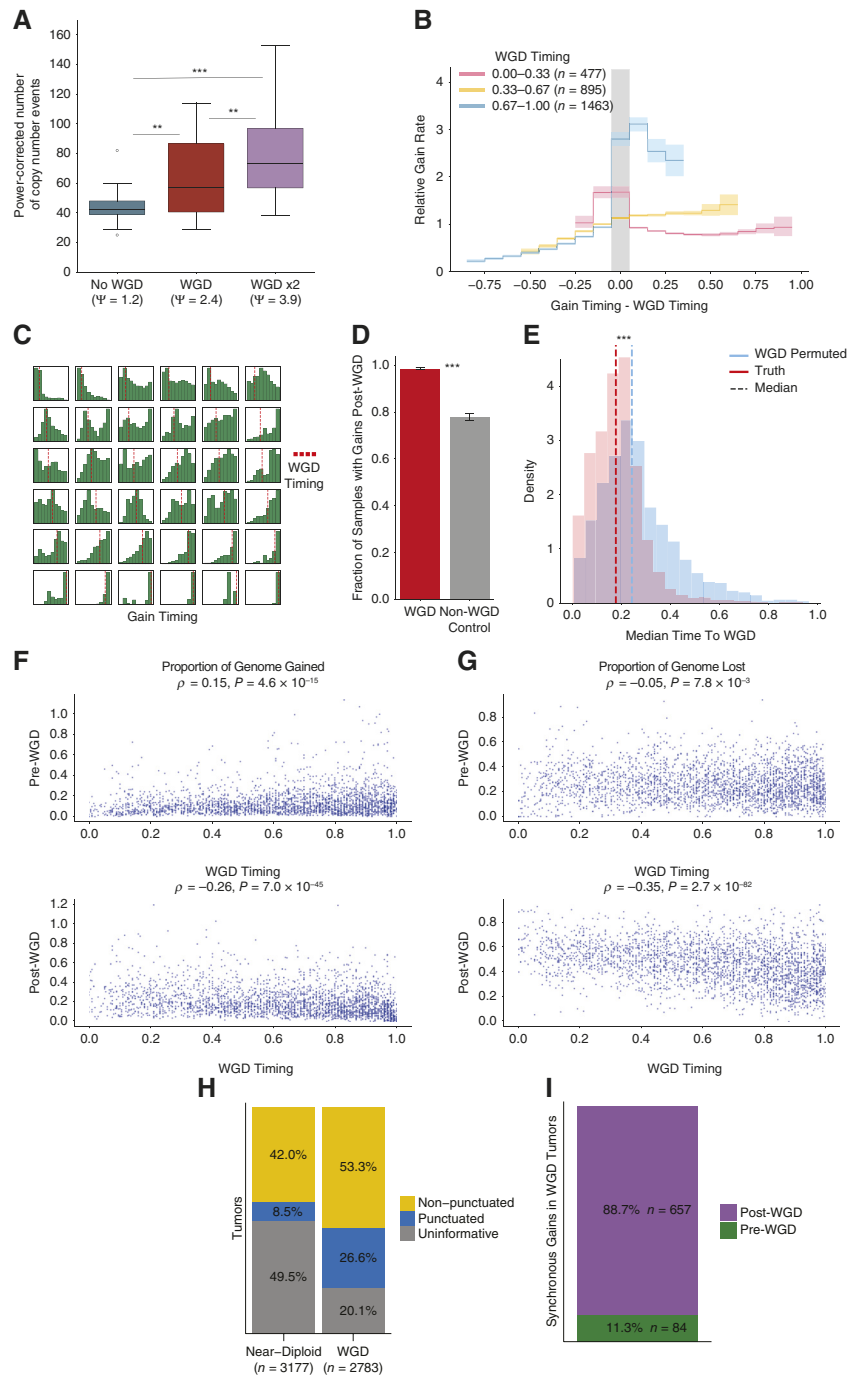


Figure 3. The history of chromosomal instability over tumor evolution. **A**, The number of copy number events required to reach the final copy number state from the most recent common ancestor from a collection of tumor cells with different ploidy states from a single undifferentiated sarcoma. The number of events is normalized for ploidy and statistical significance is calculated with a Mann-Whitney U test. Ψ denotes population ploidy as determined by FACS. **B**, The normalized rate of gains relative to the WGD timing across cancer types. Statistical significance is calculated with a permutation test and 95% confidence intervals are calculated by bootstrapping over samples. **C**, Histograms of gain timing in a random selection of tumors with a WGD. Samples are ordered by independent gain timing from left to right and WGD timing from top to bottom. **D**, Proportion of samples with gains post-WGD for WGD tumors and a cohort of control non-WGD tumors with a pseudo-WGD timing randomly sampled from WGD tumors with the same cancer type. Statistical significance is calculated with a permutation test and 95% confidence intervals are calculated by bootstrapping over samples. **E**, Distribution of the mean mutation time between the timing of all independent gains and the median WGD timing for each genome duplicated sample. Two cohorts are displayed, one with the correct WGD timing and another where the WGD timing is permuted between samples of the same cancer type. Statistical significance calculated by Mann-Whitney U test. **F**, Proportion of genome gained before and after WGD against WGD timing for genome-doubled tumors. **G**, Proportion of genome lost before and after WGD against WGD timing for genome-doubled tumors. **H**, Proportion of tumors with clonal gains identified as occurring in a punctuated burst, or uninformative where the number of gains was too low to classify, split by WGD status. **I**, Proportion of punctuated gains occurring in WGD samples, classified by whether they occurred pre- or post-WGD. **, $P < 0.01$; ***, $P < 0.001$.

consistent with reports of highly aneuploid karyotypes of early precursor lesions of breast cancers (28). However, both trends were generally conserved when considering individual cancer types with a sufficient number of samples (Supplementary Fig. S24). While difficult to observe in individual tumors (Fig. 3C; Supplementary Fig. S26), the effect of WGD on CIN in aggregate was clear.

The rate of copy number gain accumulation increases over the clonal evolutionary period for both WGD and non-WGD tumors (Supplementary Fig. S27). Nevertheless, post-WGD gains occurred more frequently than could be explained by the increase in CIN over tumor development that is also present in non-WGD tumors. The proportion of WGD tumors that had a gain post-WGD (98.5%) was significantly higher than in a control cohort of non-WGD tumors each given a realistic pseudo-WGD timing (78.0%, $P < 0.001$, permutation test; Fig. 3D; Supplementary Fig. S28A and S28B; “Methods”). Moreover, the mean mutational time between independent gains and WGD was significantly lower than expected from permuting WGD timing across samples from the same cancer type ($P < 0.001$, Mann-Whitney U Test; Fig. 3E; Supplementary Figs. S28C, S28D, and S29; “Methods”).

The later a WGD occurs, the less time there is for gains to accumulate post-WGD. Correspondingly, the amount of genome gained post-WGD was negatively correlated with WGD timing (Fig. 3F; Supplementary Fig. S30A and S30B). Similarly, the genomic material gained pre-WGD was weakly positively correlated with WGD timing (Fig. 3F). There was also a stronger negative correlation between WGD timing and the proportion of genome lost post-WGD (Fig. 3G; Supplementary Fig. S30C). Generally, these trends are conserved when cancer types are considered separately and after applying the non-parsimony penalty (Supplementary Figs. S31–S33). The high frequency of post-WGD losses is consistent with the widespread hypothesis that, in many cancers, WGD serves as an evolutionary intermediate to fitter sub-tetraploid karyotypes (29).

In contrast, there was very little correlation between WGD timing and the proportion of the genome that was lost pre-WGD. This suggests that such losses pre-WGD either do not accumulate steadily with respect to SNV accumulation or are linked to WGD itself (Fig. 3G). This could support a model whereby a major advantage of WGDs is the mitigation of the deleterious effect of mutations in regions with copy number losses, as reported by Lopez and colleagues (26). Compared to non-WGD tumors, WGD tumors have a higher proportion of genomic LOH (26). This suggests that rather than SNV accumulation, a high level of genomic loss could lead to WGD, which may explain our results.

Punctuated Gain Evolution in WGD Tumors

Next, we sought to better understand the distribution of gain timing in WGD samples. We previously found that copy number gains in non-WGD tumors often occur as punctuated events (17). Indeed, using GRITIC, we found that gains in 16.9% of informative non-WGD samples (those with gains affecting at least three separate chromosomes) occurred over significantly shorter timespans than expected under a permutation model (Fig. 3H; Supplementary Fig. S34A–S34D; “Methods”).

Using GRITIC, we can now study the timing of gains that arise independently of the WGD. We found that gains in 33.3% of informative WGD tumors occurred significantly closer in time than expected from permutations (Fig. 3H; Supplementary Fig. S34A–S34E). Most of these (88.7%) occurred post-WGD (Fig. 3I; Supplementary Fig. S34F), suggesting that WGD may increase the likelihood of or tolerance to punctuated bursts of gains (25). Together, these results suggest that copy number gains occur frequently in punctuated bursts, even in the most chromosomally unstable samples. These punctuated bursts likely explain why some tumors with late WGD still accumulate most clonal gains post-WGD (Supplementary Fig. S30A and S30B). We observed no significant association between tumors with chromothripsis and those with punctuated gains (Supplementary Fig. S35A and S35B), suggesting that the two processes are unrelated. We also observed similar segment size distributions for punctuated and non-punctuated gains (Supplementary Fig. S36A–S36D; “Methods”), suggesting similar underlying mechanisms.

Measuring the Impact of Genome Doubling on the Copy Number Landscape

The landscape of gains and losses along the genome is similar for WGD and non-WGD tumors across cancer types (9). Therefore, we sought to determine whether the landscape of copy number events differs before versus after WGD. First, we compared the relative frequency of arm-level copy number gains across different cancer types pre- and post-WGD and found that they were highly positively correlated (Fig. 4A; Supplementary Figs. S37 and S38).

We then investigated how arm loss frequencies were affected by WGD. As a pre-WGD loss and post-WGD loss cannot both be inferred for a given genomic region in the same sample, we applied a correction to the post-WGD loss frequency (“Methods”). With this correction, we also observed clear positive correlations ($P < 0.001$; Fig. 4B; Supplementary Fig. S39). This suggests that losses without LOH observed post-WGD are mostly derived from a continuation of the same processes that lead to pre-WGD LOH losses. However, there are outliers where the pre-WGD loss frequency is much higher than expected given the corresponding post-WGD loss frequency.

Of the 23 arms across different cancer types that had a higher loss proportion pre-WGD than post-WGD, 14 were 9p ($n = 5$) and 17p ($n = 9$; Fig. 4B). Chromosome arms 9p and 17p contain the frequently hit tumor suppressor genes *CDKN2A* and *TP53*, respectively. Thus, while gains and losses are broadly unaffected by a WGD, we find specific arm-level losses that occur disproportionately pre-WGD. We note that most arms had higher event rates post-WGD than pre-WGD, reflective of increased CIN post-WGD. In agreement with previous findings (10, 30), we found that the frequencies of chromosome arm events maintained similar levels of (anti) correlation with the overall tumor suppressor and oncogene density (30) across pre-WGD, post-WGD, and non-WGD copy number events (Supplementary Figs. S40 and S41). Despite certain arms having significantly higher pre- than post-WGD loss, the frequency of both event types showed a similar correlation with driver gene density across chromosome arms. This suggests that the disproportionate pre-WGD loss frequency of

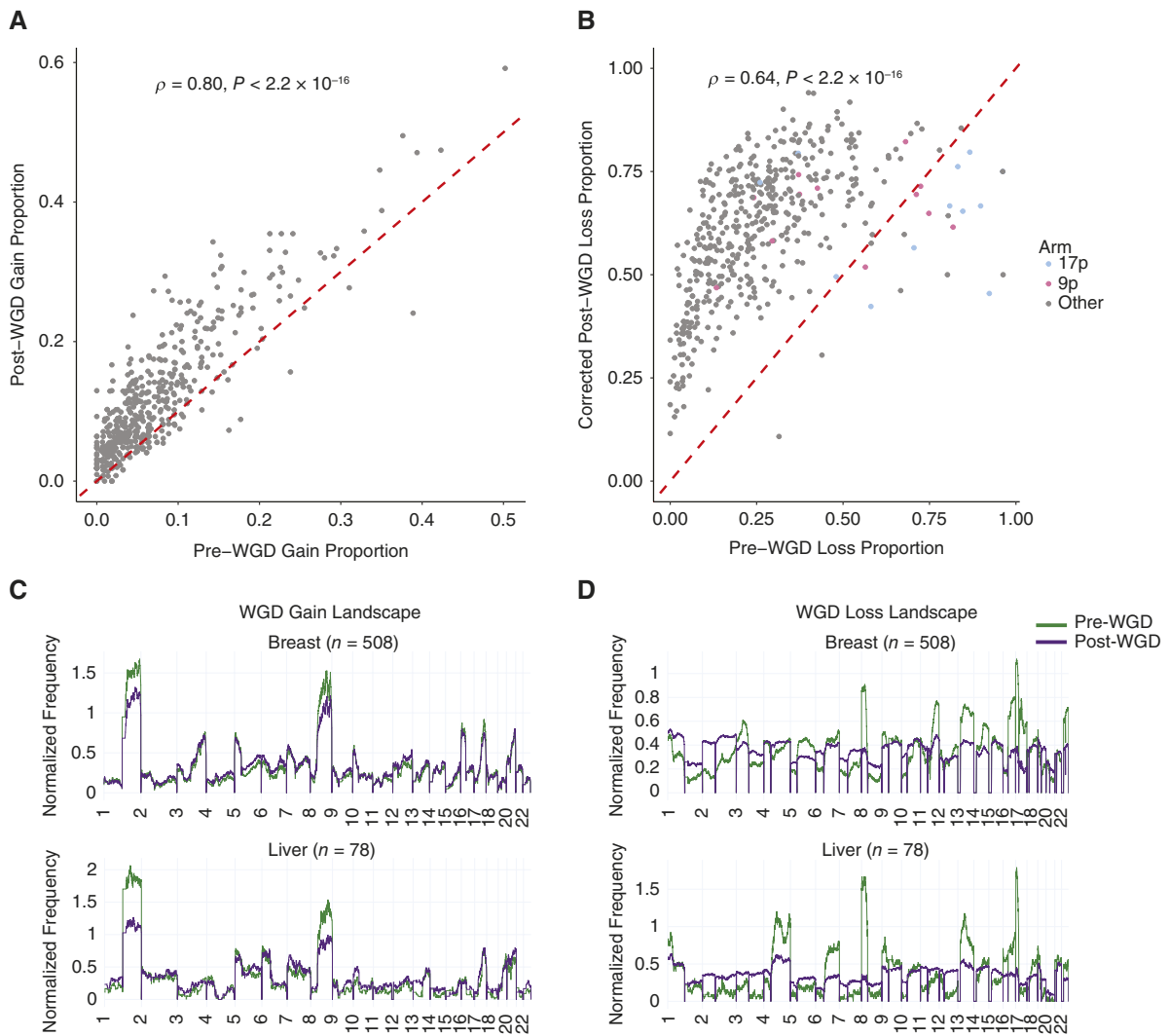


Figure 4. The landscape of pre- and post-WGD copy number events. **A**, Frequency of arm-level gain events occurring before and after a WGD in different cancer types. Each point corresponds to the frequency of arm gain relative to WGD in an individual cancer type. **B**, Frequency of arm-level loss events occurring before and after a WGD in different cancer types. Each point corresponds to the frequency of arm loss relative to WGD in an individual cancer type. The post-WGD arm loss frequency is corrected for mutual exclusivity in measuring pre-WGD and post-WGD arm losses in the same region. **C**, Frequency of pre- and post-WGD gains for breast and liver tumors. **D**, Frequency of pre- and post-WGD losses for breast and liver tumors. The post-WGD loss frequency is corrected for mutual exclusivity. Both frequencies are normalized so that the pre- and post-WGD frequencies integrate to the same arbitrary constant.

certain chromosome arms may be due to specific genes, perhaps in the context of a second inactivating hit, rather than the overall tumor suppressor density.

We then normalized the relative rates of pre- and post-WGD events for both gain and loss across the genome. We found that the rates of pre- and post-WGD gains were similar when aggregated across cancer types (Supplementary Fig. S42). However, there are notable differences in individual tumor types. For example, gains on 1q and chromosome 8 were disproportionately likely to occur pre-WGD relative to other events (Fig. 4C; Supplementary Figs. S42–S46). Gains on chromosome 3 are often pre-WGD for small cell lung cancers and upper respiratory tract carcinomas (Supplementary Figs. S45 and S46). Chromosome 7 is commonly

gained pre-WGD in glioblastomas, in agreement with previous observations that these events occur very early (Supplementary Fig. S43; ref. 17).

The differences were much larger pre- versus post-WGD for losses, both at the aggregate level and in individual cancer types (Fig. 4D; Supplementary Figs. S47–S49). For example, we observed high levels of pre-WGD loss of chromosome 18 in colorectal and pancreatic adenocarcinomas and 8p in liver, prostate, and colorectal tumors (Fig. 4D; Supplementary Fig. S47 and S48). Together, our results suggest a model in which the landscape of copy number gains remains broadly similar post-WGD, although several chromosome arms are predominantly lost pre-WGD. The frequencies of copy number changes in tumors are dictated by a combination of physical

factors that affect how often they occur and the selective impact that they confer (31). As the mechanisms that result in copy number gains and losses are similar, this suggests that the change in loss frequencies post-WGD is driven by changes in selective impact.

DISCUSSION

GRITIC is a Bayesian framework for genome-wide timing of both simple and complex gains in cancer evolution. It leverages the relationship between alternate read counts, copy number, and purity to determine the sequence and timing of gains. GRITIC theoretically allows the timing of gains from any copy number state, although in practice for computational efficiency and timing accuracy, we focus on segments with no more than 500 possible routes and at least 20 SNVs, respectively.

By applying GRITIC to the PCAWG and Hartwig datasets, we measured the genome-wide timing of gains relative to WGD, describing the effect of WGD on chromosomal instability. We found that late WGDs tended to induce a spike in gain activity, which remained elevated compared with the pre-WGD gain rate. Conversely, early WGDs were preceded by an elevated gain rate, which then decreased post-WGD. This increase in genomic instability likely enables WGD tumors to have greater adaptivity in response to therapy, contributing to their poor prognosis (29). It is worth noting that losses may affect these results, as they can make previously occurring gains unobservable. Similarly, because losses cannot be quantitatively timed, the gain rate over mutation time was not corrected for total genomic content. However, the contrasting patterns observed for tumors with different WGD timing suggest that an increase in CIN post-WGD does not solely result from more chromosomes missegregating at the same rate as pre-WGD.

We found that the landscapes of gains occurring before and after genome duplication were similar, which parallels the similarity in gain landscape between tumors with versus without WGD (9). Therefore, we hypothesize that WGD only has a very moderate impact on the fitness landscape of gains. Although the pre- and post-WGD landscapes are also broadly similar for losses, there are selective pressures to lose certain chromosome arms encoding well-known tumor suppressor genes before WGD, leading to LOH.

In WGD tumors, many copy number segments arise through routes that violate the principle of maximum parsimony, even after this was substantially penalized. It is worth noting that we inferred the most parsimonious routes from single biopsies only. Event histories that appear non-parsimonious, as measured in one biopsy, may be parsimonious when the full heterogeneity of copy number in the tumor is considered. These findings call into question the validity of the maximum parsimony assumption in cancer evolution, particularly in the context of inference from a single biopsy.

GRITIC considers the gains in different segments separately. Future work incorporating structural variant information into timing analyses would enable more integrated evolutionary analyses of copy number changes across the genome (24). Similarly, as our inference is limited by ambiguities in resolving routes from SNV read counts, phasing SNVs

either to haplotypes or ideally to specific copies will substantially reduce this ambiguity and allow greater resolution of evolution. Currently, we are unable to time events in tumors with multiple WGDs, which we estimate to be 5.8% of the patients in the PCAWG and Hartwig cohorts. While this is theoretically possible in our framework, more work is required to build an inference pipeline that can link the complex gained states across segments.

We have restricted our analysis to the timing of gains using a mutation-based timescale. Although it preserves the true order of events, it does not have a linear mapping to real-time (Supplementary Methods). By only timing gains using clock-like mutations (32), it is possible to time events in absolute time (17). However, as this greatly reduces the number of mutations per segment, only the largest events such as WGDs can be timed in this manner and thus we are unable to consider the timing of individual copy number gains in real time.

We have considered the timing of loss events when they can be compared to the gain timing results obtained from GRITIC. This is principally by contrasting the landscapes of copy number events pre- and post-WGD. Making additional comparisons is limited as we are unable to quantitatively time clonal copy number losses. Future methods could use structural variants to quantitatively time losses that are linked to copy number gains.

In summary, GRITIC is a novel computational framework for inferring copy number gain evolution from a single bulk whole-genome sequencing experiment. GRITIC can be applied across cohorts to reconstruct more complete evolutionary timelines (17) and to better understand CIN, particularly in relation to WGD.

METHODS

Data Collection

Whole-genome sequencing, alignment, mutation calling, and copy number data were obtained from the PCAWG (21) and Hartwig Medical Foundation datasets (33) uniformly processed using the Hartwig Medical foundation pipeline. Only samples with a number of reads per clonal copy of at least five were considered, a measure of sequencing coverage corrected for tumor purity (Supplementary Methods). This threshold was found to be sufficient from simulations (Supplementary Fig. S50).

To obtain clonal copy number profiles, PURPLE copy number outputs were rounded to the nearest integer for each parental allele. SNV clustering information was obtained by using the default settings of DPCLust (15). DPCLust was run for 2,000 iterations with 1,000 burn-in steps.

The cancer type classifications for each sample were obtained from a unified annotation of the PCAWG and Hartwig cohorts (33). The cancer type information for 143 PCAWG and 775 Hartwig tumors was not available from this set of unified annotations. It was obtained by mapping between the cancer type information for each cohort and the unified annotations (Supplementary Methods). This allowed us to obtain unified cancer type information for all PCAWG tumors and an additional 694 Hartwig tumors. The remaining 81 Hartwig samples were removed from our analysis.

SNVs that were identified as part of kataegis events were identified using the PCAWG kataegis detection pipeline (21). These were filtered from the data because kataegis is a localized hypermutation process that leads to sets of SNVs that violate the assumption of constant relative mutation rates across the genome.

GRITIC

GRITIC enumerates the binary tree structures that represent all routes to a given gained copy number state for a particular segment, accounting for the presence of up to one WGD. These representations are used to sample the SNV multiplicity proportions that correspond to the range of possible gain and WGD timing for each route. The relative probability of each route and gain timing was calculated using a uniform prior and the likelihood of the SNV read counts for the segment given each sampled multiplicity proportion. GRITIC outputs a posterior distribution over gain timing and routes. A full description of the GRITIC method and its principles is provided in the Supplementary Methods.

GRITIC WGD Calling

To identify WGD in our cohort, we calculated the cumulative number of base pairs spanned by each clonal major copy number state and identified the major copy number state spanning the highest number of total base pairs, i.e., the mode of the major allele. If the mode was one, the sample was identified as non-WGD. If it was two, we calculated the individual timing of all segments with major copy number two. A core principle of GRITIC is that WGD causes simultaneous gain across the genome. Therefore, if at least 60% of the base pairs spanned by the major copy number two segments had posterior gain timing distributions with overlapping 90% credible intervals, then the sample was identified as WGD.

This provided WGD calls consistent with those provided by the datasets using copy number profiles alone (Supplementary Fig. S51A and S51B). Notably, the samples with a major copy number mode of two but with less than 60% timing overlap were enriched in lung and skin tumors, which are tumor types known to have late copy number gains (Supplementary Fig. S51C). Indeed, the timing of the maximum overlap was later in mutation time compared to tumors with greater than 60% overlap in the timing of major copy number two gains ($P < 0.001$; Mann-Whitney U Test; Supplementary Fig. S51D).

Penalty on Non-Parsimony

As discussed in this paper earlier, the model evidence used in the computation of the probability for each route provides a natural penalty against additional independent gains because the evidence is integrated over the extra parameters required to model the timing of these additional gains. However, this provides no penalty for the additional loss events required to make each route consistent. Therefore, we applied a penalty term P to the route posterior probability based on the number of events n implied by each route.

$$P = e^{-nl}$$

We tuned the penalty parameter l on a representative simulated cohort of simulated tumors to have only parsimonious routes. We set l such that the total probability of non-parsimony across the cohort was approximately 5%. As this will depend on the exact setup of the simulation, we did not tune l precisely; instead, we found that $l = 2.7$ was a reasonable penalty, giving ~5% (5.3%) total non-parsimonious probabilities.

Measuring Performance of Gain Timing Inference using Simulated Data

We used a probabilistic approach to compare the gain timing inferred from GRITIC with the simulated ground truth. For each simulated segment and route, we sorted all inferred gain timings and all true timings and compared these timings pairwise. Although the number of independent gains differs between copy number states, the total number of gains, including those inferred to arise through WGD, is the same across all routes for a given copy number state.

Only comparisons corresponding to independent gains in the sorted true cohort were collected. The pairwise comparisons for each route were stored in a histogram and each comparison from every route was weighted according to the inferred route probability from GRITIC.

Measuring Non-Parsimony Probabilities

We assessed non-parsimony across the simulated and PCAWG and Hartwig cohorts by examining the total probability assigned to routes for each gained segment that had more events than the route with the minimum number of events for the segment.

Comparing the Timing of Major Copy Number Gains

We sought to compare the relative timing of gains leading to different major copy number states across tumors. We calculated the relative percentile rank of each posterior sample of gain timing across the combined posterior distribution over all segments for a given tumor. The percentile ranks were directly compared between the samples. We computed and compared two percentile rank distributions: one with only the initial gain that leads to each complex state and a second with all gains.

Calculating the Rate of Gains Relative to WGD

The rate of gain relative to WGD was measured by summing the posterior density multiplied by the segment base pair length across evenly sized bins of *gain timing* - *WGD timing* across the cohort. The bins each had a size of 0.1 in mutation time and were distributed across the 1 to 99th percentiles of the *gain timing* - *WGD timing* distribution. A small offset was applied to the bin start points such that one bin ranged from -0.05 to 0.05.

The binned gain-timing distribution needs to be normalized to account for the distribution of WGD timing in each cohort, as the maximum possible time for a gain to occur before and after the WGD is dependent on the WGD timing. Therefore, we normalized the binned posterior density by dividing it by a second binned distribution that used the same samples and the corresponding WGD timing distribution. However, in this normalizing distribution, the gain timing was uniformly distributed across mutation time, and the gains in each sample were weighted by segment base pair length such that each sample contributed equally. This normalizing distribution therefore represented the change in *gain timing* - *WGD timing* that would occur purely from differences in WGD timing alone.

Measuring the Proportion of Samples with Gains Post-WGD

We sought to identify the proportion of tumors with post-WGD gains. A tumor was identified as having post-WGD gains if at least 50% of the samples from the posterior gain timing distribution for any segment had at least one post-WGD gain.

As the rate of copy number gains generally increases over tumor development independent of WGD, we sought to compare the proportion of tumors with post-WGD gains to that expected from a control cohort of non-WGD tumors.

Each non-WGD tumor was given a pseudo-WGD timing distribution randomly sampled from WGD tumors of the same cohort. The fraction of non-WGD tumors that had at least one gained region that occurred after their randomly assigned WGD timing distribution was then calculated as a control in the same manner as the WGD cohort. Only cancer types with at least 10 WGD and 10 non-WGD samples were considered.

Measuring the Average Timing Proximity between Independent Gains and WGD

We measured the median difference in timing between the posterior gain timing distribution over all the gained segments and the WGD timing sampled from the WGD timing distribution for the tumor.

We repeated this process for 25 samples of the WGD distribution for each tumor and calculated the average median difference in timing. This resulted in the distribution of the average difference between WGD and gain timing for all WGD tumors in our cohort.

This distribution was compared to a control distribution calculated identically, except that all WGD distributions were randomly permuted between WGD tumors of the same cancer type.

Inferring Punctuated Gains

We applied a permutation-based approach (34) to identify samples that had pan-genome gains occurring in a punctuated burst using a method similar to that of Gerstung and colleagues (17). For each sample, we compared the gain timing variation across segments to permuted samples with gains obtained from across tumors with the same cancer time. A tumor was defined as punctuated if its gain timing variation was lower than 95% of permuted samples. A full description of the punctuated gains inference method is provided in the Supplementary Methods.

Determining the Relative Order of Copy Number Events and WGD

The relative order of independent copy number gains and WGD was determined from the joint posterior distribution over gain timing and routes. Each posterior sample for a gained segment in a WGD tumor contains the sampled timing of all the independent gains and the sampled WGD timing. The average number of gains that occur pre- and post-WGD for each segment was then computed from the posterior samples.

The assignment of gains pre- and post-WGD is sensitive to parsimony considerations, particularly for gains that occur at a time close to the WGD. This is because each gain pre-WGD is equivalent to two post-WGD gains in terms of how much it increments the copy number and therefore affects the total number of events required for a route. This is why we present all relevant results with and without a penalty on non-parsimony. The overall proportion of gains pre-WGD and post-WGD are highly correlated with and without the penalty (Supplementary Fig. S52A and S52B).

A region was identified as having a pre-WGD loss if its minor copy number was zero and a post-WGD loss if its minor copy number was one. This involves a weak assumption of parsimony as a minor copy number of zero could arise from two post-WGD losses, though without any other gains, a minor copy number of one cannot arise from a pre-WGD loss.

To correct for mutual exclusivity when measuring pre- and post-WGD losses, a corrected post-WGD loss proportion was calculated by dividing the post-WGD loss proportion by $1 - \text{the pre-WGD loss proportion}$, thereby changing the denominator to only just the samples that did not have a pre-WGD loss.

The implied losses from complex gained segments are not included in the analysis of the WGD loss landscapes as they comprise a minority of overall losses and cannot straightforwardly be combined with the proportions corrected for mutually exclusivity in measuring pre- and post-WGD losses of the minor allele.

Calculating Arm Pre- and Post-WGD Event Rates

We assessed a chromosome arm as being pre- or post-WGD gained in a WGD tumor if at least 50% of the total base pairs in the arm belonged to segments with at least 50% of posterior gain timing samples and at least one gain pre- or post-WGD, respectively. Similar to the classifications made at a segment level, an arm was classified as having a pre-WGD loss if at least 50% of the arm had a minor copy number of zero and as a post-WGD loss if at least 50% of the arm had a minor copy number of one.

Pan-Genome Copy Number Event Landscapes

We also produced pre- and post-WGD gain and loss proportions at 1 kb resolution across the genome. For certain cancer types with a low number of samples, the pre-WGD loss proportion was 1.0, leading to division by zero when calculating the corrected post-WGD loss proportion. Therefore, we clipped the pre-WGD loss proportion to a maximum of 0.95 when calculating the corrected post-WGD loss proportion. To compare relative rates, pan-genome event proportions were normalized such that the integral of the proportions over the genome was equal to 10^9 , chosen to provide a normalized event frequency with a magnitude of approximately one.

Data availability

An access request for sequencing data and metadata from the Hartwig Medical Foundation can be found at <https://www.hartwigmedicalfoundation.nl/en/data/data-access-request/>. Researchers with ICGC access can obtain Hartwig pipeline output for the ICGC subset of the PCAWG cohort by following instructions at <https://docs.icgc-argo.org/docs/data-access/icgc-25k-data>. Similarly, researchers with The Cancer Genomic Atlas (TCGA) access can obtain Hartwig pipeline output for TCGA subset of the PCAWG cohort at <https://icgc.bionimbus.org/files/5310a3ac-0344-458a-88ce-d55445540120>. GRITIC output files for the PCAWG, Hartwig, and simulated samples are deposited on Zenodo at <https://zenodo.org/records/12010145> (doi: 10.5281/zenodo.12010144). GRITIC is available at <https://github.com/VanLoo-lab/gritic>. The scripts used to run GRITIC on PCAWG, Hartwig, and simulated samples, as well as to produce the figure for this manuscript can be found at https://github.com/VanLoo-lab/gritic_analysis.

Authors' Disclosures

C. Swanton reports grants, personal fees, and other support from AstraZeneca, grants and personal fees from Boehringer Ingelheim, Bristol Myers Squibb, Pfizer, and Roche Ventana, grants from Invitae (formerly Archer Dx), Ono Pharmaceuticals, and Personalis, personal fees from Amgen, Illumina, GlaxoSmithKline, MSD, Achilles Therapeutics, Bicycle Therapeutics, Genentech, GRAIL, and Medixci, personal fees and other support from Relay Therapeutics and Saga Diagnostics, personal fees from Sarah Canon Research Institute and China Innovation Centre of Roche (CICOR) during the conduct of the study; in addition, C. Swanton has a patent for PCT/GB2017/053289 issued, a patent for PCT/EP2016/059401 issued, a patent for PCT/EP2016/071471 issued, a patent for PCT/GB2018/052004 issued, a patent for PCT/GB2020/050221 issued, a patent for PCT/GB2018/051912 issued, a patent for PCT/US2017/28013 issued, a patent for PCT/GB2018/051912 issued, a patent for PCT/GB2018/051892 issued, and a patent for PCT/EP2022/077987 issued; and C. Swanton is a Royal Society Napier Research Professor (RSRP\R\210001). His work is supported by the Francis Crick Institute that receives its core funding from Cancer Research UK (CC2041), the UK Medical Research Council (CC2041), and the Wellcome Trust (CC2041). For the purpose of Open Access, the author has applied a CC BY public copyright license to any Author Accepted Manuscript version arising from this submission. C. Swanton is funded by Cancer Research UK [TRACERx (C11496/A17786), PEACE (C416/A21999), and CRUK Cancer Immunotherapy Catalyst Network]; Cancer Research UK Lung Cancer Centre of Excellence (C11496/A30025); the Rosetrees Trust, Butterfield and Stonegate Trusts; NovoNordisk Foundation (ID16584); Royal Society Professorship Enhancement Award (RP/EA/180007 & RF\ERE\231118); National Institute for Health Research (NIHR) University College London Hospitals Biomedical Research Centre; the Cancer Research UK-University College London Centre; Experimental Cancer Medicine Centre; the Breast Cancer Research Foundation (US; BCRF-22-157); Cancer Research UK Early Detection and Diagnosis Primer

Award (Grant EDDPMA-Nov21/100034); and The Mark Foundation for Cancer Research Aspire Award (Grant 21-029-ASP) and ASPIRE Phase II award (Grant 23-034-ASP). C. Swanton is in receipt of an ERC Advanced Grant (PROTEUS) from the European Research Council under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 835297). C. Swanton is co-Chief investigator for the NHS Galleri Trial and Chief investigator for AstraZeneca's MeRmaid one and two clinical trials. P. Van Loo reports grants from Cancer Research UK, UK Medical Research Council, Wellcome Trust, and Cancer Prevention and Research Institute of Texas during the conduct of the study. No disclosures were reported by the other authors.

Authors' Contributions

T.M. Baker: Conceptualized and developed GRITIC under the supervision of **M. Tarabichi** and **P. Van Loo**. **H.A. Ogilvie:** Contributed to the development of the mathematical formulation of GRITIC. **T.M. Baker:** Carried out the majority of the analyses. **S. Lai:** Conducted the punctuated gains analysis. **A.R. Lynch:** Performed analysis of genomic features of punctuated gains. **T. Lesluyes** and **S. Dentre:** Conducted the SNV clustering analysis. **H. Yan** and **A.L. Bowes:** Performed computational single-cell analysis. **A. Verfaillie:** Performed wet-lab experiments on undifferentiated sarcomas. **T.M. Baker, M. Tarabichi,** and **P. Van Loo:** Wrote the manuscript with contributions from **A.R. Lynch** and **H.A. Ogilvie**. **N. Pillay, A.M. Flanagan, C. Swanton,** and **P.T. Spellman** Provided samples and expertise and contributed to manuscript writing. All authors have read and approved the final manuscript.

Acknowledgments

This work was supported by the Francis Crick Institute that receives its core funding from Cancer Research UK (CC2008, CC2041), the UK Medical Research Council (CC2008, CC2041), and the Wellcome Trust (CC2008, CC2041). T.M. Baker was supported by a PhD fellowship from Boehringer Ingelheim Fonds. A.R. Lynch is a TRIUMPH Fellow in the CPRIT Training Program (RP210028). M. Tarabichi was supported as a postdoctoral researcher of the F.R.S.-FNRS. C. Swanton is Royal Society Napier Research Professor (RP150154). His work is funded by Cancer Research UK (TRACERx, C11496/A17786; Cancer Research UK Lung Cancer Centre of Excellence C11496/A30025), the Rosetrees Trust, Butterfield and Stonegate Trusts, NovoNordisk Foundation (ID16584), a Royal Society Research Professorship Enhancement Award (RP/EA/180007), the National Institute for Health Research (NIHR) Biomedical Research Centre at University College London Hospitals, the Breast Cancer Research Foundation (BCRF 20-157), the CRUK-UCL Experimental Cancer Medicine Centre. His research is supported by a Stand Up To Cancer-LUNGevity-American Lung Association Lung Cancer Interception Dream Team Translational Research Grant (SU2C-AACR-DT23-17). C. Swanton also receives funding from the European Research Council (FP7-THESEUS-617844, FP7-PloidyNet 607722, PROTEUS 835297 and Chromavision 665233). A.M. Flanagan, N. Pillay, and P. Van Loo acknowledge support from Sarcoma UK (SUKG01.2018). A.M. Flanagan and N. Pillay were also supported by the National Institute for Health Research (NIHR) Biomedical Research Centre at University College London Hospitals, the CRUK-UCL Centre, Experimental Cancer Medicine Centre and Bone Cancer Research Trust infrastructure grants (2019 to 2023). N. Pillay has received funding from Cancer Research UK (award number 18387) and holds a Cancer Research UK Career Establishment award (RCCCEA-Nov23/100003). P. Van Loo is a Winton Group Leader in recognition of the Winton Charitable Foundation's support toward the establishment of The Francis Crick Institute. P. Van Loo is a CPRIT Scholar in Cancer

Research and acknowledges CPRIT grant support (RR210006). The results shown here are in part based upon data generated by the TCGA Research Network: <https://www.cancer.gov/tcga>. This publication and the underlying study have been made possible partly based on data that the Hartwig Medical Foundation and the Center of Personalised Cancer Treatment (CPCT) have made available to the study through the Hartwig Medical Database. We thank Chris Yau and Chris Barnes for their helpful comments and feedback on the Supplementary Methods.

Note

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

Received October 25, 2023; revised April 22, 2024; accepted June 27, 2024; published first June 28, 2024.

REFERENCES

- Zack TI, Schumacher SE, Carter SL, Cherniack AD, Saksena G, Tabak B, et al. Pan-cancer patterns of somatic copy number alteration. *Nat Genet* 2013;45:1134–40.
- Storchova Z, Kuffer C. The consequences of tetraploidy and aneuploidy. *J Cell Sci* 2008;121:3859–66.
- Martincorena I, Fowler JC, Wabik A, Lawson ARJ, Abascal F, Hall MWJ, et al. Somatic mutant clones colonize the human esophagus with age. *Science* 2018;362:911–7.
- Martincorena I, Roshan A, Gerstung M, Ellis P, Van Loo P, McLaren S, et al. Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science* 2015;348:880–6.
- Lawson ARJ, Abascal F, Coorens THH, Hooks Y, O'Neill L, Latimer C, et al. Extensive heterogeneity in somatic mutation and selection in the human bladder. *Science* 2020;370:75–82.
- Knouse KA, Wu J, Whittaker CA, Amon A. Single cell sequencing reveals low levels of aneuploidy across mammalian tissues. *Proc Natl Acad Sci U S A* 2014;111:13409–14.
- Ben-David U, Amon A. Context is everything: aneuploidy in cancer. *Nat Rev Genet* 2020;21:44–62.
- Bielski CM, Zehir A, Penson AV, Donoghue MTA, Chatila W, Armenia J, et al. Genome doubling shapes the evolution and prognosis of advanced cancers. *Nat Genet* 2018;50:1189–95.
- Prasad K, Bloomfield M, Levi H, Keuper K, Bernhard SV, Baudoin NC, et al. Whole-genome duplication shapes the aneuploidy landscape of human cancers. *Cancer Res* 2022;82:1736–52.
- Watkins TBK, Lim EL, Petkovic M, Elizalde S, Birkbak NJ, Wilson GA, et al. Pervasive chromosomal instability and karyotype order in tumour evolution. *Nature* 2020;587:126–32.
- Ganem NJ, Godinho SA, Pellman D. A mechanism linking extra centrosomes to chromosomal instability. *Nature* 2009;460:278–82.
- Gemble S, Wardenaar R, Keuper K, Srivastava N, Nano M, Macé A-S, et al. Genetic instability from a single S phase after whole-genome duplication. *Nature* 2022;604:146–51.
- Durinck S, Ho C, Wang NJ, Liao W, Jakkula LR, Collisson EA, et al. Temporal dissection of tumorigenesis in primary cancers. *Cancer Discov* 2011;1:137–43.
- Jolly C, Van Loo P. Timing somatic events in the evolution of cancer. *Genome Biol* 2018;19:95.
- Nik-Zainal S, Van Loo P, Wedge DC, Alexandrov LB, Greenman CD, Lau KW, et al. The life history of 21 breast cancers. *Cell* 2012;149:994–1007.
- Purdum E, Ho C, Grasso CS, Quist MJ, Cho RJ, Spellman P. Methods and challenges in timing chromosomal abnormalities within cancer samples. *Bioinformatics* 2013;29:3113–20.
- Gerstung M, Jolly C, Leshchiner I, Dentre SC, Gonzalez S, Rosebrock D, et al. The evolutionary history of 2,658 cancers. *Nature* 2020;578:122–8.

18. Leshchiner I, Mroz EA, Cha J, Rosebrock D, Spiro O, Bonilla-Velez J, et al. Inferring early genetic progression in cancers with unobtainable premalignant disease. *Nat Cancer* 2023;4:550–63.
19. Wang Z, Xia Y, Mills L, Nikolakopoulos AN, Maeser N, Dehm SM, et al. Evolving copy number gains promote tumor expansion and bolster mutational diversification. *Nat Commun* 2024;15:2025.
20. Jakobsdottir GM, Dentre SC, Bristow RG, Wedge DC. AmplificationTimeR: an R package for timing sequential amplification events. *Bioinformatics* 2024;40:btac281.
21. ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium. Pan-cancer analysis of whole genomes. *Nature* 2020;578:82–93.
22. Priestley P, Baber J, Lolkema MP, Steeghs N, de Bruijn E, Shale C, et al. Pan-cancer whole-genome analyses of metastatic solid tumours. *Nature* 2019;575:210–6.
23. Demeulemeester J, Dentre SC, Gerstung M, Van Loo P. Biallelic mutations in cancer genomes reveal local mutational determinants. *Nat Genet* 2022;54:128–33.
24. Greenman CD, Pleasance ED, Newman S, Yang F, Fu B, Nik-Zainal S, et al. Estimation of rearrangement phylogeny for cancer genomes. *Genome Res* 2012;22:346–61.
25. Dewhurst SM, McGranahan N, Burrell RA, Rowan AJ, Grönroos E, Endesfelder D, et al. Tolerance of whole-genome doubling propagates chromosomal instability and accelerates cancer genome evolution. *Cancer Discov* 2014;4:175–85.
26. López S, Lim EL, Horswell S, Haase K, Huebner A, Dietzen M, et al. Interplay between whole-genome doubling and the accumulation of deleterious alterations in cancer evolution. *Nat Genet* 2020;52:283–93.
27. Mitchell TJ, Turajlic S, Rowan A, Nicol D, Farmery JHR, O'Brien T, et al. Timing the landmark events in the evolution of clear cell renal cell cancer: TRACERx renal. *Cell* 2018;173:611–23.e17.
28. Casasent AK, Schalck A, Gao R, Sei E, Long A, Pangburn W, et al. Multiclonal invasion in breast tumors identified by topographic single cell sequencing. *Cell* 2018;172:205–17.e12.
29. Laughney AM, Elizalde S, Genovese G, Bakhoun SF. Dynamics of tumor heterogeneity derived from clonal karyotypic evolution. *Cell Rep* 2015;12:809–20.
30. Davoli T, Xu AW, Mengwasser KE, Sack LM, Yoon JC, Park PJ, et al. Cumulative haploinsufficiency and triplosensitivity drive aneuploidy patterns and shape the cancer genome. *Cell* 2013;155:948–62.
31. Baker TM, Waise S, Tarabichi M, Van Loo P. Aneuploidy and complex genomic rearrangements in cancer evolution. *Nat Cancer* 2024;5:228–39.
32. Alexandrov LB, Jones PH, Wedge DC, Sale JE, Campbell PJ, Nik-Zainal S, et al. Clock-like mutational processes in human somatic cells. *Nat Genet* 2015;47:1402–7.
33. Martínez-Jiménez F, Movasati A, Brunner SR, Nguyen L, Priestley P, Cuppen E, et al. Pan-cancer whole-genome comparison of primary and metastatic solid tumours. *Nature* 2023;618:333–41.
34. Strona G, Nappo D, Boccacci F, Fattorini S, San-Miguel-Ayanz J. A fast and unbiased procedure to randomize ecological binary matrices with fixed row and column totals. *Nat Commun* 2014;5:4114.