Tolerance of Whole-Genome Doubling Propagates Chromosomal Instability and Accelerates Cancer Genome Evolution

Sally M. Dewhurst, Nicholas McGranahan, Rebecca A. Burrell, Andrew J. Rowan, Eva Grönroos, David Endesfelder, Tejal Joshi, Dmitri Mouradov, Peter Gibbs, Robyn L. Ward, Nicholas J. Hawkins, Zoltan Szallasi, Oliver M. Sieber, and Charles Swanton

ABSTRACT

The contribution of whole-genome doubling to chromosomal instability (CIN) and tumor evolution is unclear. We use long-term culture of isogenic tetraploid cells from a stable diploid colon cancer progenitor to investigate how a genome-doubling event affects genome stability over time. Rare cells that survive genome doubling demonstrate increased tolerance to chromosome aberrations. Tetraploid cells do not exhibit increased frequencies of structural or numerical CIN per chromosome. However, the tolerant phenotype in tetraploid cells, coupled with a doubling of chromosome aberrations per cell, allows chromosome abnormalities to evolve specifically in tetraploids, recapitulating chromosomal changes in genomically complex colorectal tumors. Finally, a genome-doubling event is independently predictive of poor relapse-free survival in early-stage disease in two independent cohorts in multivariate analyses [discovery data: hazard ratio (HR), 4.70, 95% confidence interval (CI), 1.04–21.37; validation data: HR, 1.59, 95% CI, 1.05–2.42]. These data highlight an important role for the tolerance of genome doubling in driving cancer genome evolution.

SIGNIFICANCE: Our work sheds light on the importance of whole-genome-doubling events in colorectal cancer evolution. We show that tetraploid cells undergo rapid genomic changes and recapitulate the genetic alterations seen in chromosomally unstable tumors. Furthermore, we demonstrate that a genome-doubling event is prognostic of poor relapse-free survival in this disease type. Cancer Discov; 4(2); 175–85. © 2014 AACR.

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

S.M. Dewhurst and N. McGranahan contributed equally to this work.

Corresponding Author: Charles Swanton, Cancer Research UK London Research Institute, 44 Lincoln’s Inn Fields, London, WC2A 3LY, United Kingdom. Phone: 44-207-269-3463; Fax: 44-207-269-3094; E-mail: charles.swanton@cancer.org.uk

doi: 10.1158/2159-8290.CD-13-0285

©2014 American Association for Cancer Research.
INTRODUCTION

Chromosomal instability (CIN), describing the continual loss and gain of whole and/or parts of chromosomes, is a common feature of most cancers (1). CIN represents a dynamic state that likely contributes to intratumor heterogeneity by creating a genetically diverse pool of tumor cells upon which selection can act. CIN is associated with both poor prognosis and intrinsic multidrug resistance (2, 3). Several different cellular mechanisms are thought to contribute to CIN (1), and it has been suggested that specific “CIN tolerance mechanisms” may be required for these cells to survive elevated genome instability (4, 5). Although mutations in TP53 are likely to contribute to this CIN tolerance, a full mechanistic basis for this phenomenon remains incompletely understood (6).

Another common feature of tumor cells is large-scale alterations in ploidy. Polyploid cells have been observed in multiple cancer types, and tetraploidy, resulting from a genome-doubling event, has been proposed as an intermediate en route to aneuploidy (7–9). Tetraploidy has been shown to be an unstable cellular state, with polyploid yeast exhibiting increased requirements for genome stability maintenance pathways such as homologous recombination repair, as well as showing defects in sister chromatid cohesion (10). Furthermore, artificially generated mammalian tetraploid cell lines display increased segregation errors, due to supernumerary centrosomes (11). It is likely that the p53 pathway limits the proliferation of polyploid cells to protect genomic integrity (12). Several studies have observed tetraploid cancer cells before aneuploidy onset and in the

Figure 1. A relationship between weighted mean chromosome copy number and weighted Genome Instability Index (wGII). Each circle represents one TCGA (The Cancer Genome Atlas) colorectal cancer tumor sample. Red depicts genome-doubled (GD) samples; blue non-genome-doubled (nGD) samples (see Methods). A histogram of weighted mean chromosome copy number for GD (red) and nGD (blue) is shown above. B, i, copy number losses that occur on the background of a diploid genome before genome doubling will result in LOH, whereby one of the parental alleles is lost. In tumors that harbored CIN before genome doubling, the majority of losses will be unbalanced, involving LOH. Unbalanced losses to two copies (AA or BB) are depicted with a purple box with purple dotted lines. ii, losses after genome doubling. In tumors where genome doubling was an early event, before the onset of CIN, the majority of losses to two copies will be balanced without LOH. Balanced losses to two copies (AB or BB) are depicted with a purple box with purple dotted lines. iii, timing of genome doubling estimated using copy number and LOH profiles. Each bar represents one genome-doubled TGA tumor and its height corresponds to the proportion of AB - proportion AA or BB copy number states. Tumor genomes in which the majority of losses to two copies are likely to have occurred after genome doubling are shown in red (n = 130; proportion AB - proportion AA or BB), and those where the majority of losses are likely to have occurred before doubling are shown in blue (n = 66; proportion AB < proportion AA or BB). D, flow cytometry shows a >4N population in the MIN colon cancer cell line HCT-116. 2N, 4N, and >4N populations are indicated on the flow cytometry plot. E, cloning efficiency of 2N and >4N cells is shown with mean and SEM (three experiments). Tetraploid cloning efficiency was significantly lower than diploid cloning efficiency (P = 0.032, Student t test). Diploid cloning efficiency was assessed using CellTiter-Blue (CTB) reagent (colonies were identified as wells with CTB value >1.5x mean average of blank wells) and the Poisson-corrected cloning efficiency was calculated (see Methods). Tetraploid cloning efficiency was calculated by verifying the percentage of surviving tetraploid clones using flow cytometry (flow cytometry data not shown). (continued on following page)
Genome Doubling Accelerates Cancer Evolution

transition from premalignant to malignant disease (13–15), suggesting that a genome-doubling event can be a driver of tumorigenesis. Consistent with this, tetraploid cells derived through multiple routes have an increased tumorigenic capacity (16–19). Tetraploid subclones have also been observed at later stages of tumor development (20), and whole-genome-doubling events have been inferred to occur both before and after other copy-number alterations across different cancer types (21, 22). However, the effect of a whole-genome-doubling event on CIN and how this could affect genome evolution in human cancer has not been fully explored. Here, we analyze the relationship between ploidy and genomic instability in colorectal cancer, and propose that tolerance of genome doubling might provide tolerance to CIN in this cancer type.

RESULTS

A Relationship between Ploidy and Genomic Complexity in Colorectal Cancer

SNP6.0 data available for 404 stage I–IV colorectal cancer tumors from The Cancer Genome Atlas (TCGA) was used to explore the relationship between ploidy and CIN. To assess structural and numerical CIN, the weighted Genome Instability Index (wGII) was used, which estimates the proportion of the genome with aberrant copy number compared with the median ploidy, weighted on a per-chromosome basis (23). We have previously shown that GII correlates with both numerical and structural CIN in cell lines (3). Significantly higher wGIs were found in polyploid (ploidy \( \geq 3 \)) compared with diploid tumors (\( P < 0.0001 \), Student t test; Fig. 1A), although some

Figure 1. (Continued) F, after single-cell sorting, DNA content was assessed by flow cytometry with Hoescht staining. Two tetraploid clones (TC 3 and TC 4, at passage 3), one diploid clone (DC 8), and HCT-116 are shown. G, flow cytometry of the diploid clone DC 8 also shows a small \( > 4N \) subpopulation. Two further diploid clones (DC-14 and DC-25) and four tetraploid clones (TC-13, TC-16, TC-17, and TC-35) were isolated from DC 8, and their DNA content as assessed by flow cytometry with Hoescht staining is shown (passage 3). H, LOH states of early (passage 5) diploid and tetraploid clones analyzed by SNP6.0. Dark blue LOH events in tetraploid clones are likely to have occurred before genome doubling. A barplot depicting the proportion of the genome displaying LOH is shown above; the black dotted line depicts the mean proportion of LOH in diploids. The majority of LOH events are present in both diploid and tetraploid clones. I, a family tree depicting all diploid and tetraploid clones used in this study. Tetraploid clones are shown in red and diploid clones are shown in blue. MIN, microsatellite instability.
stable tetraploid tumors were observed. In other tumor types, polyploidy was also associated with an increased wGII, suggesting that these two genomic aberrations may be linked in a range of cancers (P < 0.0001 for each cancer type, Student t test; Supplementary Fig. S1A).

Next, we applied an algorithm that identifies tumors that are likely to have undergone a genome-doubling event, even if they are no longer polyploid (adapted from ref. 7; see Methods). Significantly higher wGIs were observed in tumors classified as genome-doubled compared with non–genome-doubled (P < 0.0001, Student t test; Fig. 1A), suggesting a potential relationship between genome doubling and genome complexity. The majority of tumors with a triploid karyotype appeared to have undergone a genome-doubling event (105 of 110 tumors). However, we also observed non–genome-doubled near-diploid tumors with high wGIs, consistent with there being multiple routes to an unstable genome in colorectal cancer.

**Genome Doubling Is an Early Event in the Majority of Colorectal Cancers**

Copy-number losses occurring on the background of a diploid genome will result in LOH. This LOH will leave a permanent footprint in the genome, persisting after a genome-doubling event (Fig. 1B). In contrast, losses occurring after genome doubling are less likely to exhibit LOH (Fig. 1Bi). The types of losses in a genome-doubled sample may thus shed light on the timing of genome doubling relative to copy-number losses in the genome (Fig. 1B; see Methods). In the majority of genome-doubled TCGA colorectal tumors, genome doubling likely occurred as a relatively early event, before the majority of copy-number losses (Fig. 1C). In more than 20% of samples in which genome doubling was classified as an early event, the proportion of the genome exhibiting LOH was less than the mean proportion of LOH in chromosomally stable tumors with microsatellite instability (MIN). These data suggest that CIN can occur after genome doubling in vivo.

**An Isogenic Cell Line System to Study the Effects of Genome Doubling**

We next explored the acute effects on genome stability following a genome-doubling event. HCT-116, a diploid MIN colorectal cancer cell line, was found to have a small subpopulation (<2%) with >4N DNA content (Fig. 1D). Single tetraploid and diploid cells were isolated by flow cytometry. The cloning efficiency of tetraploid cells was lower than diploid cells, suggesting that tetraploidy is poorly tolerated in HCT-116 cells under standard culture conditions (2N = 63%, 4N = 6%; Fig. 1E). We expanded one diploid clone (DC 8) and two rare surviving tetraploid clones (TC 3, TC 4; Fig. 1F). It was possible to isolate a second generation of tetraploid clones (TC-13, TC-16, TC-17, and TC-35) as well as two diploid clones (DC-14 and DC-25) from the diploid clone DC 8 (Fig. 1G). Second-generation tetraploid clones had therefore arisen from a single diploid cell spontaneously within the time of the experiment. All tetraploid clones were found to have a seemingly functional p53 response to DNA damage (Supplementary Fig. S1B), and no mutations were found in the coding regions of TP53 or CDKN1A (p21; data not shown), suggesting alternative tolerance routes. Although tetraploids grew slower than diploids at early passages, they grew at approximately the same rate by later passages (∼18 months in culture; Supplementary Fig. S2A–S2C).

All clones were subject to SNP6.0 analysis, and the proportion of the genome showing LOH was determined (Fig. 1H). Tetraploid clones showed limited to no LOH beyond that harbored by the diploid clones (diploid mean, 6.17% (5.96–6.36%) of genome; tetraploid mean, 6.46% (6.16–6.79%); P = 0.181; Kolmogorov–Smirnov test), suggesting that tetraploid clones were not CIN before genome doubling. Analysis of copy-number gains also showed that genome doubling likely occurred before CIN (data not shown). All clones (Fig. 1I) were continuously cultured for over 18 months so that the effects of genome doubling on genome complexity could be assessed over time.

**CIN in Tetraploid Clones**

Clonal FISH was used to assess ploidy and numerical CIN in each clone (Fig. 2A and B). The cell-to-cell variation in chromosome number (percentage of cells deviating from the modal chromosome number of individual colonies) provides a measure of numerical CIN emerging during colony expansion from single cells. Tetraploid colonies displayed significantly higher cell-to-cell variation than diploid colonies [Fig. 2C; passage 5 diploid mean, 7% (0%–23%); tetraploid mean, 28% (5%–57%), P < 0.0001; passage 50 diploid mean = 13% (3%–34%); tetraploid mean, 33% (7%–68%), P < 0.0001, Student t test]. We also isolated diploid and tetraploid clones from a microsatellite stable clone of HCT-116, stably expressing a functional wild-type MLH1 gene (referred to as HCT-116_MLH1; ref. 24; Supplementary Fig. S2D). All four HCT-116_MLH1 tetraploid clones exhibited higher cell-to-cell variation in chromosome number than the diploid clone at passage 5 (P < 0.0001, Student t test; Supplementary Fig. S2E).

We analyzed segregation errors on a per-cell and a per-chromosome basis using immunofluorescence (Fig. 2D). Tetraploid cells exhibited a higher percentage of anaphases displaying segregation errors than diploids across three different passages [diploid mean, 19% (11%–26%); tetraploid mean, 42% (32%–55%), P < 0.005, Student t test; Fig. 2E, top]. However, segregation errors calculated on a per-chromosome basis (dividing fraction of anaphases with errors by number of chromosomes for each clone as determined by SNP6.0 data; see Methods) were not significantly increased in tetraploid clones at any passage (Fig. 2E, bottom). Therefore, in this system, increased segregation errors in tetraploid cells may simply result from the increased number of chromosomes. Tetraploid clones displayed a similar spectrum of segregation errors to diploid clones, includingacentric and centric lagging chromosomes and anaphase bridges (Supplementary Fig. S3A and S3B). These data suggest that there may not be a tetraploid-specific mechanism driving increases in any one type of segregation error in this system.

The prevalence of structural chromosome aberrations in all clones was scored from metaphase chromosome spreads hybridized to an all-centromere fluorescent probe (Fig. 2F). Tetraploid clones displayed more structural abnormalities per cell than diploid clones [Fig. 2G; diploid mean, 0.39 (0.26–0.58) abnormalities; tetraploid mean, 0.93 (0.60–1.62) abnormalities; all passages P < 0.05, Student t test]. However, no significant difference in structural abnormalities on a per-chromosome basis between diploids and tetraploids was observed [chromosomes were counted directly from...
Genome Doubling Accelerates Cancer Evolution

Figure 2. A, diagram of a clonal FISH slide, showing the two measures of chromosome number deviation that can be scored: cell-to-cell variation in chromosome number is the percentage of cells that deviate from the modal chromosome number of each individual colony. Colony-to-colony variation reflects differences in the modal chromosome copy number between colonies. B, example images of colonies from four clones with chromosome 2 (CEP2) shown in red, and chromosome 8 (CEP8) in green. Individual cells have been highlighted, and their copy-number state for these two chromosomes is shown in the inset. Scale bar (in white) = 10 μm. C, cell-to-cell variation in chromosome number. The average percentage deviation of two chromosomes, chromosome 2 and 8, is shown for all clones at both passage 5 and 50 (clones all scored using an Ariol automated microscope system, except DC 8, TC 3, and TC 4 at passage 50, which were scored using a DeltaVision microscope). Passage numbers shown throughout are correct to within four passages. Colonies with <10 cells were excluded from analysis. Each point represents one colony. Median number of cells: passage 5 = 2,479, passage 50 = 2,105. D, chromosome segregation errors in anaphase were visualized by immunofluorescence. Representative single z-stack images show types of segregation errors that were scored. Blue, 4′,6-diamidino-2-phenylindole (DAPI); red, crest. Side panels show each channel individually, and inset shows a close-up of the segregation error. Scale bars (in white) = 3 μm. E, chromosome segregation errors per cell. Fifty anaphases were scored for each cell line; only data from bipolar anaphases is shown. On a per-cell basis are shown on top graph (colored bars represent individual clones, see key above). P values refer to comparisons between diploids and tetraploids at each passage (Student’s t test). On a per-chromosome basis (bottom, all gray bars, representing the same clones as in graph immediately above), there is no significant difference in segregation errors per chromosome. P values are indicated above bars. F, representative images of normal and abnormal metaphase chromosomes are shown, stained with DAPI and probed with an all-human centromere probe. Scale bars (in white) = 2.5 μm. G, structural abnormalities on a per-cell and per-chromosome basis in diploid and tetraploid clones. Number of structural abnormalities per cell is shown on the top colored graph (P values refer to comparisons between diploids and tetraploids at each passage), and the number of structural abnormalities per chromosome is shown on the graph below with gray bars representing exactly the same clones as in graph above (P values for comparisons between diploid and tetraploids at each passage are indicated in bars above). Median number of spreads scored at each passage: passage 5 = 25, passage 25 = 29, passage 50 = 27, and HCT-116 = 37. ns, not significant.

Published OnlineFirst January 19, 2014; DOI: 10.1158/2159-8290.CD-13-0285

Downloaded from cancerdiscovery.aacrjournals.org on June 20, 2017. © 2014 American Association for Cancer Research.
Dewhurst et al.

**Tolerance of CIN in Tetraploid Clones**

We hypothesized that, despite no increase in CIN in tetraploid clones on a per-chromosome basis (Fig. 2E and G), increased tolerance to chromosomal segregation errors might contribute to the association between ploidy and genomic complexity (Fig. 1A). In clonal FISH experiments, the presence of colonies with modes differing from either two for diploid or four for tetraploid clones (colony-to-colony variation) suggests tolerance of an unbalanced aneuploid genome. Individual colonies with aneuploid chromosome numbers were observed in all tetraploid clones [Fig. 3A and B, passage 5 mean, 25% of colonies (8%–44%); passage 50 mean, 30% (18%–43%)]. In contrast, we observed only a single aneuploid colony in HCT-116 (1.7% of all colonies). One tetraploid

**Figure 3.** A and B, colony-to-colony variation in modal chromosome copy number for chromosome 2 (A) and 8 (B). Frequency of different colony modes from clonal FISH data is shown from all clones at passage 5 and at passage 50. Median number of colonies scored: passage 5 = 44, passage 50 = 39. C, live-cell imaging of H2B-mRFP-expressing cell reveals different daughter cell fates after segregation errors. The percentage frequency of each cell fate [mitosis vs. death or arrest (arrest = interphase > 48 hours after division; see Methods)] in long-term live-cell imaging studies of all diploid and tetraploid clones either after no error or after a segregation error is shown. Example images of mitoses are shown above each panel. Data shown are an amalgamation of all clones (for individual results and n numbers see Supplementary Fig. S4G and also see Supplementary Movies SA–SF). D, wGII at different passages for diploid and tetraploid clones. Dashed line indicates wGII = 0.2, a threshold separating MIN and CIN cell lines (23). E, weighted mean chromosome copy number versus wGII for colorectal cancer tumors from the TCGA (gray), diploid clones (blue), and tetraploid clones (red) at different passages. Diploid clones at all passages overlay the same point. Lighter colors represent later passages for tetraploid clones. F, genome-wide copy number losses and gains for all clones at passage 5, 25, and 50 (and passage 75 for DC-14, DC-25, TC-13, TC-16, TC-17, and TC-35). Blue sections represent loss and red sections represent gain (relative to ploidy). ns, not significant; mRFP, monomeric red fluorescent protein.
Genome Doubling Accelerates Cancer Evolution

We assessed the correlation between loss of genes on chromosome 4q and wGII in the TCGA dataset, controlling for the increased likelihood of losses in chromosomally unstable tumors (see Methods). Loss of genes on chromosome 4q was significantly correlated with increasing wGII ($P < 0.001$; Fig. S4A), and chromosome 4 was one of the chromosomes whose loss was most strongly correlated with genomic instability (Supplementary Fig. S7).

**Genome Doubling Is Associated with Poor Prognosis in Colorectal Cancer Tumors**

Given the established relationship between genomic instability and poor clinical outcome, we reasoned that genome doubling could be a useful prognostic marker in early-stage colorectal cancer. Survival data were obtained from 150 TCGA patients with stage I–III colorectal cancer and an Australian validation cohort comprising 389 patients with stage II–III colorectal cancer. Relapse-free survival was used as the endpoint for both datasets, censored at 2 years, given that $80\%$ of recurrences occur within this period (25), and also due to the paucity of survival data available in the TCGA cohort beyond this point. A genome-doubling event was significantly associated with relapse in both the TCGA [Fig. 4B; $P = 0.019$; hazard ratio (HR), 5.1; 95% confidence interval (CI), 1.1–22.8] and the validation cohort (Fig. 4C; $P = 0.0022$; HR, 1.80; 95% CI, 1.2–2.8).

When extending outcome to beyond 2 years, genome doubling remained significant in the larger validation cohort, but not the TCGA cohort (Supplementary Fig. S8). Genome doubling remained significant in both cohorts in multivariate analysis when tumor stage, age, and MIN status were included (TCGA data: $P = 0.045$; HR, 4.70; 95% CI, 1.04–21.37; validation data: $P = 0.028$; HR, 1.59; 95% CI, 1.05–2.42; Supplementary Table S2).

In contrast, wGII was only significant in univariate analysis for one of the cohorts (TCGA data: $P = 0.1296$; HR, 6.09; 95% CI, 0.57–64.93; validation data: $P = 0.00649$; HR, 3.81; 95% CI, 1.44–10.03), and was not significant in multivariate analysis when including tumor stage, age, and MSI status in either cohort (Supplementary Table S2). In the larger validation cohort, genome doubling was significant when restricting to just diploid tumors ($P = 0.001$; Supplementary Fig. S8C), and also when including polyploidy (ploidy $\geq 3$) in a multivariate analysis ($P = 0.0209$; Supplementary Table S2). A genome-doubling event may therefore provide prognostic relevance with a greater sensitivity than aneuploidy to detect high-risk tumors. Furthermore, genome doubling was significant in predicting overall 5-year survival when restricting to just early stage I–II tumors (available for the TCGA cohort only; Supplementary Table S2).

Within genome-doubled tumors, subtetraploid (ploidy < 4) samples were genomically more complex than tetraploid samples, and enriched for higher tumor stage (Fig. 4D; $P = 0.0062$, Cochran–Armitage test). This supports a model in which genome doubling is an early event in some colorectal cancers, permitting the evolution of more genomically complex, subtetraploid, higher-stage tumors.

**DISCUSSION**

Through long-term culture of naturally occurring, rare surviving tetraploid clones, we observed the evolution of genomic complexity over time specifically in tetraploid genomes. A

colony was observed in a diploid clone, consistent with our findings that tetraploid cells can emerge as a rare event in diploid clones (Fig. 1G). The outcome of this assay is unlikely to be affected by differing proliferation rates, as colonies are grown from sparsely seeded single cells. These data suggest that tolerance of aneuploidy is enhanced in tetraploid clones, but a rare event in diploid cells. HCT-116_MLH1 tetraploid clones also displayed similar colony-to-colony variation in modal chromosome number (Supplementary Fig. S3C).

We used live-cell imaging to track the fate of histone2B monomorphic red fluorescent protein (mRFP)-tagged cells following a chromosome segregation error. In diploid clones, daughter cells derived from a parental cell that had undergone a segregation error frequently died or underwent cell-cycle arrest (death or arrest—diploid mean, 58%; HCT-116, 43%; DC 8, 68%; DC-14, 55%; DC-25, 68%; Fig. 3C), whereas almost all cells that underwent a normal division continued through a subsequent mitosis (Fig. 3C; Supplementary Movies SA–SF and Supplementary Fig. S4A–S4G). However, daughter cells arising from tetraploid clones after a segregation error (including anaphase bridges and lagging chromatin; data not shown) died or arrested less frequently, with the majority continuing through a normal mitosis in the subsequent cell cycle (death or arrest—tetraploid mean, 16%; TC 3, 18%; TC 4, 12%; TC-13, 11%; TC-16, 10%; TC-17, 34%; TC-35, 12%; $P = 0.0002$, Student t test; Fig. 3C). These data indicate that tetraploid progeny have a greater tolerance of chromosome segregation errors relative to diploids.

We used SNP6.0 data for all the clones at multiple different passages over 18 months and calculated genomic complexity using the wGII. Overall, wGII significantly increased from passage 5 to 75 in tetraploid clones, but remained stable in diploid clones after a segregation error (including anaphase bridges and lagging chromatin; data not shown) died or arrested less frequently, with the majority continuing through a normal mitosis in the subsequent cell cycle (death or arrest—tetraploid mean, 16%; TC 3, 18%; TC 4, 12%; TC-13, 11%; TC-16, 10%; TC-17, 34%; TC-35, 12%; $P = 0.0002$, Student t test; Fig. 3C). These data indicate that tetraploid progeny have a greater tolerance of chromosome segregation errors relative to diploids.

**Tetraploid Clones Evolve Specific Chromosome Losses**

In every clone, chromosomal gains and losses relative to the cell line ploidy were assessed over long-term culture (Fig. 3F). Chromosomal aberrations present in parental HCT-116 were observed in all clones. No novel losses were common to all early-passage tetraploids. However, a noncontiguous region of chromosome 4q containing 362 genes was lost to three copies in all tetraploid clones by passage 50 (~1 year), consistent with selection for loss of this region during prolonged culture (Fig. 3F and Supplementary Table S1). The copy-number loss of chromosome 4q occurs after genome doubling, as it did not occur before passage 25 in any clone, and it did not display LOH (Fig. 3F and Supplementary Fig. S6A). A similar pattern of chromosome losses and increasing genomic complexity was found in HCT-116_MLH1 tetraploid clones (Supplementary Fig. S6B). This suggests that elevated genome complexity and selection of chromosome 4q loss in HCT-116 tetraploid clones is not driven by MIN.

Within genome-doubled tumors, subtetraploid (ploidy < 4) samples were genomically more complex than tetraploid samples, and enriched for higher tumor stage (Fig. 4D; $P = 0.0062$, Cochran–Armitage test). This supports a model in which genome doubling is an early event in some colorectal cancers, permitting the evolution of more genomically complex, subtetraploid, higher-stage tumors.

**DISCUSSION**

Through long-term culture of naturally occurring, rare surviving tetraploid clones, we observed the evolution of genomic complexity over time specifically in tetraploid genomes. A...
year after genome doubling, the HCT-116 genome was markedly altered relative to the genomically stable diploid progenitors, suggesting that tolerance of genome doubling permits rapid genome evolution (Fig. 3F).

Thompson and Compton (4) have shown that increasing the segregation error rates in diploid MIN colorectal cancer cells does not result in the propagation of chromosomally unstable progeny, suggesting that CIN is a complex phenotype requiring both initiation of segregation errors and tolerance of the ensuing altered genomic content in daughter cells. Although tetraploid cells displayed elevated structural and numerical CIN relative to diploid cells, diploids and tetraploids had similar segregation errors and structural abnormalities on a per-chromosome basis (Fig. 2E and G). These data suggest that there is no additional instability initiated by tetraploidization on a per-chromosome basis in our system, in which segregation defects are already observed in diploid cells. However, long-term live-cell imaging revealed increased tolerance to chromosome missegregation events in tetraploids (Fig. 3C), and we observed many aneuploid colonies in tetraploid clones (Fig. 3A). Thus, in rare tetraploid cells that survive cloning, the same MIN

Figure 4. A, relationship between copy-number loss and wGII in the TCGA cohort for genes identified as recurrently lost in tetraploid clones (see Supplementary Table S1). Blue represents loss and white represents no loss (relative to ploidy). Colorectal cancer tumors (columns) are ordered according to increasing wGII score, from left to right. Every gene (rows) shows a significant correlation between copy-number loss and wGII, even when taking into account an increased likelihood of loss in high-wGII tumors (P < 0.001; see Methods). Chromosome schematic shows where genes reside on chromosome 4; genes within regions shown in red are not depicted in the plot as they are not recurrently lost in all tetraploid clones. B, Kaplan-Meier relapse-free survival curves, censored at 2 years for genome-doubled (GD; red) and non-genome-doubled (nGD; blue) TCGA cohort colorectal cancer tumors (n = 150). P = 0.019, log-rank test (for full survival curves see Supplementary Fig. S8A). C, Kaplan-Meier relapse-free survival curves, censored at 2 years for genome-doubled (GD; red) and non-genome-doubled (nGD; blue) validation cohort colorectal cancer tumors (n = 389). P = 0.0022, log-rank test (for full survival curves see Supplementary Fig. S8B). D, relationship between wGII, ploidy, and tumor stage in genome-doubled tumors. Each circle represents one genome-doubled tumor. The barplot shows the proportion of different tumor stages for tetraploid and subtetraploid samples. P = 0.0062, Cochran-Armitage test for trend.
cell line that cannot propagate artificially induced (4) or endogenous segregation errors is now able to propagate an unstable genome.

Within the limits of SNP6.0-based analyses, we conclude that tetraploid clones were unlikely to have been CIN before genome doubling (Fig. 1H); however, it is unlikely that a genome-doubling event alone is sufficient for cells to tolerate CIN. It is possible that chromosome segregation error tolerance mechanisms can be activated in diploid cells that may or may not undergo genome doubling. Whole-genome doubling may therefore represent a consequence of a prior somatic event permitting the tolerance of genome instability, which can then exacerbate genome evolution due to elevated chromosomal aberrations occurring on a per-cell basis. Defining the mechanistic basis for such tolerance mechanisms in the presence of wild-type p53 is clearly a high priority.

Strikingly, all tetraploid clones displayed convergent loss of regions of chromosome 4q, which we also found to be commonly lost in colorectal cancer tumors with elevated genomic complexity (Fig. 4A). Interestingly, chromosome 4q loss has previously been suggested as a predictor of outcome in early-stage colorectal cancer (26). Conceivably, tetraploidy can provide a permissive genetic background for selection of high-risk genomic copy number aberrations over time.

Consistent with the effect of tetraploidization upon emerging genome instability, a genome-doubling event is an independent predictor of poorer relapse-free survival in colorectal cancer from the TCGA and in a larger validation cohort in both univariate and multivariate analyses (Fig. 4B and C and Supplementary Table S2). These data support studies that have linked aneuploidy with disease outcome in colorectal cancer (27), and genome doubling with poor prognosis in ovarian cancers (7). Genome doubling may forecast the onset and tolerance of elevated CIN, which has previously been shown to be associated with both poor prognosis and intrinsic drug resistance (2, 3).

A genome-doubling event could represent a macroevolutionary leap in tumors, analogous to saltation in ecol- ogy, which both precipitates and sustains extensive chromosomal rearrangements. Other examples of punctuated evolutionary events occurring in tumors have been proposed, such as chromothripsis and chromoplexy, both involving complex chromosomal rearrangement events (28, 29). Whole genome-doubling events can drastically alter the evolution of whole organisms, for example by facilitating subfunctionalization of duplicated genes (30), so that there may be additional benefits to a genome-doubling event beyond the propagation of CIN. It will be important to investigate mechanisms leading to the emergence of tetraploid cells in tumors, such as cellular stress, cytokinesis failure, or telomere shortening (8).

On the basis of these data, we suggest that the tolerance of genome doubling combined with an elevated chromosome segregation error rate on a per-cell basis provides the fuel for rapid genomic change, accelerating evolution of tumors from a karyotypically stable to a more complex state. Deciphering the cause of CIN tolerance is likely to have important therapeutic implications in guiding efforts to limit tumor diversity, evolution, and adaptation.

**METHODS**

**Cell Culture**

HCT-116 cells were obtained from the European Collection of Animal Cell Cultures (ECACC) by Cancer Research UK (CRUK) cell services, and short tandem repeat (STR) fingerprinted on 10/10. The HCT-116_MLH1.3 clone was a gift from Françoise Praz (Saint-Antoine Research Centre, Paris, France), and was STR fingerprinted on 02/10, with similar results to HCT-116 except at vWA. Cells were maintained at 37°C in 5% CO2 in Dulbecco’s Modified Eagle Medium (DMEM) with high glucose and l-glutamine (Invitrogen), supplemented with 10% FBS and 1× PenStrep (Sigma). Clones were passaged approximately once a week and were split at the same dilution. Passage numbers represented in figures are within <4 passages of the passage used. CellTiter-Blue (Promega) assays were performed following the manufacturer’s instructions. Cloning efficiency was estimated using the Poisson distribution: efficiency = (−100) × ln(of wells with no colony/total # of wells)% (as in ref. 31).

**Fluorescence-Activated Cell Sorting Analysis**

Cells were stained with 10 μg/ml Hoechst 33342 (Sigma) for 1 hour at 37°C. A MoLo (Beckman Coulter) cell sorter was used to sort single cells into 96-well plates with 20% FBS media. Assessment of DNA content was carried out by flow cytometry using propidium iodide (Sigma) with RNase (Life Technologies), or Hoechst 33342, after fixation in 70% ethanol.

**Clonal FISH, Metaphase Spreads, and Immunofluorescence**

FISH probing for chromosomes 2 (CEP2 D271, SO) and 8 (CEP8, D872, SGn, both Abbott Molecular probes) was performed as described previously (23). Slides were scored semi-automatically using the Ariol system (Leica Microsystems). Colonies were scanned at 40× magnification with z-stacks of 9 × 0.7 μm, and analyzed with the automated SPOT assay before manual curation. Three slides were scored using an Olympus DeltaVision RT microscope (Applied Precision, LLC) equipped with a Coolscope HQ camera with an Olympus 40× 1.3 numerical aperture UPlanSapo oil immersion objective. Metaphase spreads were prepared and probed with an all-human centromere probe (Poseidon) as described previously (23). Immunofluorescence and segregation error classification were performed as described previously (23). One hundred z-stacks of 0.2 μm were captured for each image for scoring types of segregation error; single z-stacks are represented in figures for clarity.

**H2B-mRFP Transfection and Live-Cell Imaging Analysis**

Cells were transfected with pH2B–mRFP (gift from A. Straube, Warwick Medical School, University of Warwick, Coventry, UK) by using Fugene 6.0 (Promega) and were selected in 1 mg/mL G418 (Life technologies) before flow sorting for mRFP expression. Cells were maintained in 500 μg/mL G418 and imaged in an 8-well chamber (LabTek) using the same DeltaVision microscope in 5% CO2 at 37°C; 14-μm z-stacks were taken every 3 minutes for 6 hours and every 15 minutes thereafter, for approximately 60 hours, and analyzed using softWoRx Explorer (Applied Precision, LLC). Cells were scored as arrested if they failed to divide within 48 hours of the parental division. Multipolar divisions were excluded from analysis.

**SNP Array Processing**

Cell lines were analyzed with Affymetrix SNP6.0 arrays. PICNIC (32) was used for normalization and integer copy-number estimation. The TCGA Affymetrix SNP6.0 data were downloaded for 422 colon adenocarcinomas, 898 breast, 391 lung (adenocarcinoma), 407 lung (squamous cell carcinoma), 506 ovarian, and 503 renal cancers (https://tcga-data.nci.nih.gov/tcga/). Samples that failed Affymetrix
Genotyping Console quality control (QC) were excluded. LogR and B allele frequencies (BAF) were obtained using the aroma R package (33). Integer copy numbers were estimated using OncoSNP (34). Validation cohort analysis was performed on Illumina 610 Quad arrays. LogRs and BAFs were obtained using GenomeStudio V2011.1 and Genotyping Module V1.9.4. For QC, samples with moving SD > 0.28 were discarded. Integer copy numbers were estimated using OncoSNP (34).

Ploidy was estimated for each sample by summing the weighted median integer copy for each chromosome and dividing by the number of chromosomes analyzed (n = 22). The number of chromosomes in each sample was estimated by summing the modal copy numbers from the segmented copy-number profile of each chromosome. Each segment was weighted according to the number of base pairs it covered. Copy-number segments of loss and gain were defined relative to ploidy. wGII was calculated as in ref. 23.

Validation Cohort
Validation cohort colorectal cancer patients were recruited from the Royal Melbourne Hospital (Parkville, VIC, Australia), Western Hospital Footscray (Footscray, VIC, Australia), and St. Vincent’s Hospital Sydney (Darlinghurst, NSW, Australia). The study was ethics approved, and patients gave informed consent.

Genome Doubling Algorithm
A modified version of a published algorithm (7) was used. Each sample, s, was represented as an aberration profile of major and minor allele copy numbers at chromosome arm resolution. From this profile, we calculated Ns, the total number of aberrations (relative to diploid) and Ps, the probabilities of loss/gain for each allele at each chromosome arm. Of note, 10,000 simulations were run for each sample. In each simulation, Ns sequential aberrations, based on Ps, were applied to a diploid profile. A P value for genome doubling was obtained by counting the percentage of simulations in which the proportion of chromosome arms with a major allele copy number ≥ 2 was higher than that observed in the sample. For samples with ploidy ≤ 3, a P value threshold of 0.001 was used. To avoid underestimating genome doubling in high-ploidy samples, P ≤ 0.05 was used if ploidy = 4, and all samples in which ploidy ≥ 5 were classified as genome-doubled.

Estimating Timing of Genome Doubling
Each genome-doubled sample was represented as an array of genotype proportions reflecting copy numbers ranging from zero to eight. Sixteen possible genotypes were discriminated: zero copies, A B (one copy); AA/BB and AB (two copies); AAA/BBB and AAB/ABB (three copies), etc. (where A and B represent the two parental alleles). Only losses to two copies (AA, BB, and AB) were used, as these can reflect losses either before (AA or BB) or after genome doubling (AB). Samples with a higher-proportion AB compared with AA/BB were classified as having genome doubled before the majority of losses, whereas those where AA/BB > AB were classified as having genome doubled after the majority of losses.

Significance of Correlation between wGII and Copy-Number Loss
On the basis of the observed probability for loss, given by the percentage of genome that is lost in that sample, we generated an aberration state (loss or no loss) for each sample separately. A point-biserial correlation between aberration state and wGII was then calculated across samples. This process was repeated 10,000 times and a P value was obtained for each gene by counting the percentage of simulations showing a greater correlation coefficient than that observed for that gene.

Statistical Analyses
Survival curves were plotted according to the Kaplan–Meier method. Log-rank test statistics were used to assess significance for univariate analysis. Cox proportional hazards regression models were conducted for multivariate survival analysis (R package survival). Survival times were censored at 2 years unless otherwise stated. All statistical analyses were carried out using the R statistical environment or GraphPad Prism.

Disclosure of Potential Conflicts of Interest
No potential conflicts of interest were disclosed.

Authors’ Contributions
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): S.M. Dewhurst, N. McGranahan, E. Grönroos, T. Joshi, P. Gibbs, R.L. Ward, N.J. Hawkins, O.M. Sieber, C. Swanton
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): S.M. Dewhurst, N. McGranahan, R.A. Burrell, D. Endesfelder, T. Joshi, D. Mouradov, O.M. Sieber, C. Swanton
Writing, review, and/or revision of the manuscript: S.M. Dewhurst, N. McGranahan, R.A. Burrell, D. Endesfelder, R.L. Ward, O.M. Sieber, C. Swanton
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): R.A. Burrell, A.J. Rowan, P. Gibbs, N.J. Hawkins, C. Swanton
Study supervision: R.A. Burrell, A.J. Rowan, C. Swanton

Acknowledgments
The authors thank the laboratory of Tim Hunt for the p21 antibody, Anne Straube for the H2B-mRFP plasmid, and François Praz for the HCT-116, MLH1 cell line. The authors are very grateful to the cell services facility at CRUK.

The authors acknowledge the Victorian Cancer Biobank for the provision of patient specimens for the validation cohort and BioGrid Australia for providing de-identified clinical data.

The results published here are in part based on data generated by the TCGA pilot project established by the NCI and NHGRI (information about the TCGA and the investigators and institutions that constitute the TCGA research network can be found at http://cancergenome.nih.gov/). The data were retrieved through database of Genotypes and Phenotypes (dbGaP) authorization (Accession No. phs000178.v5.p5).

Grant Support
This project was funded by grants from CRUK (to S.M. Dewhurst, N. McGranahan, R.A. Burrell, and C. Swanton), the Medical Research Council (to A.J. Rowan and C. Swanton; ID:G0701935/2), EU Framework 7 (to E. Grönroos, C. Swanton, T. Joshi, and Z. Szallasi; projects PREDICT and RESPONSIFY), The Prostate Cancer Foundation (to C. Swanton), The Rosetree Trust (to C. Swanton), The Breast Cancer Foundation (to C. Swanton), and the NHMRC through a Project Grant (Application ID 489418; to O.M. Sieber, P. Gibbs, and R.L. Ward), and supported by researchers at the National Institute for Health Research University College London Hospitals Biomedical Research Centre.

Received June 13, 2013; revised December 2, 2013; accepted December 4, 2013; published OnlineFirst January 19, 2014.
REFERENCES
