LETTER

Replication stress links structural and numerical cancer chromosomal instability

Rebecca A. Burrell[1]*, Sarah E. McClelland[1]*, David Endesfelder[1,2], Petra Groth[3], Marie-Christine Weller[3], Nadeem Shaikh[1], Enric Domingo[4], Nnennaya Kanu[1], Sally M. Dewhurst[1], Eva Gronroos[1], Su Kit Chew[1,5], Andrew J. Rowan[1], Arne Schenk[2], Michal Sheffer[6], Michael Howell[1], Maik Kschischo[2], Axel Behrens[1], Thomas Helleday[3], Jiri Bartek[7,8], Ian P. Tomlinson[4] & Charles Swanton[1,5]

Cancer chromosomal instability (CIN) results in an increased rate of change of chromosome number and structure and generates intratumour heterogeneity[1,2]. CIN is observed in most solid tumours and is associated with both poor prognosis and drug resistance[3,4]. Understanding a mechanistic basis for CIN is therefore paramount. Here we find evidence for impaired replication fork progression and increased DNA replication stress in CIN[+] colorectal cancer (CRC) cells relative to CIN[−] CRC cells, with structural chromosome abnormalities precipitating chromosome missegregation in mitosis. We identify three new CIN-suppressor genes (*PIGN* (also known as *MCD4*), *MEX3C* (*RKHD2*) and *ZNF516* (*KIAA0222*)) encoded on chromosome 18q that are subject to frequent copy number loss in CIN[+] CRC. Chromosome 18q loss was temporally associated with aneuploidy onset at the adenoma–carcinoma transition. CIN-suppressor gene silencing leads to DNA replication stress, structural chromosome abnormalities and chromosome missegregation. Supplementing cells with nucleosides, to alleviate replication-associated damage[5], reduces the frequency of chromosome segregation errors after CIN-suppressor gene silencing, and attenuates segregation errors and DNA damage in CIN[+] cells. These data implicate a central role for replication stress in the generation of structural and numerical CIN, which may inform new therapeutic approaches to limit intratumour heterogeneity.

Structural and numerical chromosomal instability are commonly observed together in solid tumours[6] (Supplementary Fig. 1a–c). This co-occurrence can be experimentally induced by defective mitotic checkpoint function or chromosome attachment to the mitotic spindle, or through pre-mitotic defects affecting chromosome structure, such as faulty DNA repair and replication[6–10]. However, the mechanisms underlying CIN in cancer remain unclear.

Colorectal cancers can be broadly classified as CIN[+]/aneuploid or CIN[−]/microsatellite unstable[3]. CIN[+] cells displayed an increased frequency of chromosome segregation errors compared to CIN[−] cells[11] (median 38% versus 18%, $P = 0.0025$, Mann–Whitney U-test; Supplementary Fig. 2a). To address whether mitotic or pre-mitotic mechanisms are responsible for these segregation errors, we analysed high-resolution images of anaphases in a panel of CIN[+] CRC cells. Most of the segregation errors consisted of chromosome fragments without centromeres (acentrics; Fig. 1a) and anaphase bridges (54–81%, median 70%; Fig. 1a, b), indicative of structural chromosome aberrations arising through pre-mitotic defects[12]. By contrast, only 10–43% of segregation errors were lagging chromosomes with centromeres (median 20%; Fig. 1a, b), suggesting that mitotic dysfunction resulting in improper chromosome attachments[6] cannot explain most segregation errors in CIN[+] CRC cells (Supplementary Fig. 2b).

Furthermore, kinetochore distortion of lagging chromosomes (reflecting merotelic attachments[6]) was rare (0–12% of segregation errors, median 8%; Supplementary Fig. 2c). We did not observe differences in mitotic timing, mitotic checkpoint function, sister chromatid cohesion or supernumerary centrioles between CIN[+] and CIN[−] cells, and multipolar spindles were infrequent (0–18%, median 8%; Supplementary Fig. 2d–i). These data suggest that mitotic dysfunction occurs at low frequency in CIN[+] CRC cells, and that most of the observed chromosome segregation errors result from structural chromosome aberrations. Accordingly, 22–71% (median 36%) of CIN[+] cell metaphases had structurally abnormal chromosomes, including acentric chromosomes, dicentric chromosomes and double-stranded DNA breaks (Fig. 1b, c and Supplementary Fig. 2j).

Next we sought a putative cause for these structural chromosome alterations. Activation of the DNA-damage response has been observed in both colorectal adenomas and carcinomas[13,14], and is thought to reflect DNA replication stress[13,15,16]. Pharmacological induction of replication stress in HCT-116 (CIN[−]) cells resulted in structural chromosome aberrations and segregation errors, 82% of which were bridges or acentric chromosomes (Supplementary Fig. 3a–f). Importantly, numerical chromosome changes were also induced (Supplementary Fig. 3g, h), demonstrating that replication stress can result in both structural and numerical chromosomal instability[17].

DNA replication stress results in several cellular phenotypes including DNA damage in prometaphase[10,15,16], ultrafine anaphase DNA bridges (UFBs)[16], and 53BP1-positive nuclear bodies in G1 cells[15,18] (Fig. 1d and Supplementary Fig. 3i–m). CIN[+] cells showed increased prometaphase DNA damage (median 74% CIN[+] versus 34% CIN[−] cells with ≥3 γH2AX foci; $P = 0.033$, Mann–Whitney U-test; Fig. 1d, e) in the absence of increased oxidative DNA damage (Supplementary Fig. 4a–c). γH2AX foci were not confined to telomeres (Supplementary Fig. 4d, e). CIN[+] cells also displayed more 53BP1 nuclear bodies in G1 cells ($P = 0.028$, Mann–Whitney U-test; Fig. 1d, f). Consistent with the hypothesis that replication stress may drive chromosome segregation errors in CIN[+] cells, UFBs were enriched in anaphases with segregation errors compared with anaphases without segregation errors ($P = 0.00018$, paired t-test; Fig. 1d, g).

To assess DNA replication directly, DNA fibre assays were performed for two CIN[−] and four CIN[+] cell lines to measure progression of individual replication forks, fork stalling and asymmetry between sister replication forks. CIN[+] cells showed significantly slower fork rates than CIN[−] cell lines (0.56–0.83 kb min[−1] versus 1.11–1.12 kb min[−1], $P < 0.05$; Fig. 1h, i and Supplementary Fig. 5a, b). Furthermore, there was evidence of increased replication fork stalling and asymmetric sister fork progression in several CIN[+] cell lines

[1]Cancer Research UK London Research Institute, 44 Lincoln's Inn Fields, London WC2A 3LY, UK. [2]University of Applied Sciences Koblenz, RheinAhrCampus, Department of Mathematics and Technology, Joseph-Rovan-Allee 2, 53424 Remagen, Germany. [3]Science for Life Laboratory, Division of Translational Medicine and Chemical Biology, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Box 1031, Stockholm S-171 21, Sweden. [4]Molecular and Population Genetics and NIHR Biomedical Research Centre, The Wellcome Trust Centre for Human Genetics, Oxford OX3 7BN, UK. [5]UCL Cancer Institute, Paul O'Gorman Building, Huntley Street, London WC1E 6BT, UK. [6]Department of Physics of Complex Systems, Weizmann Institute of Science, Rehovot 76100, Israel. [7]Danish Cancer Society Research Center, Strandboulevarden 49, DK-2100 Copenhagen, Denmark. [8]Institute of Molecular and Translational Medicine, Palacky University Olomouc, CZ-775 15, Czech Republic.
*These authors contributed equally to this work.
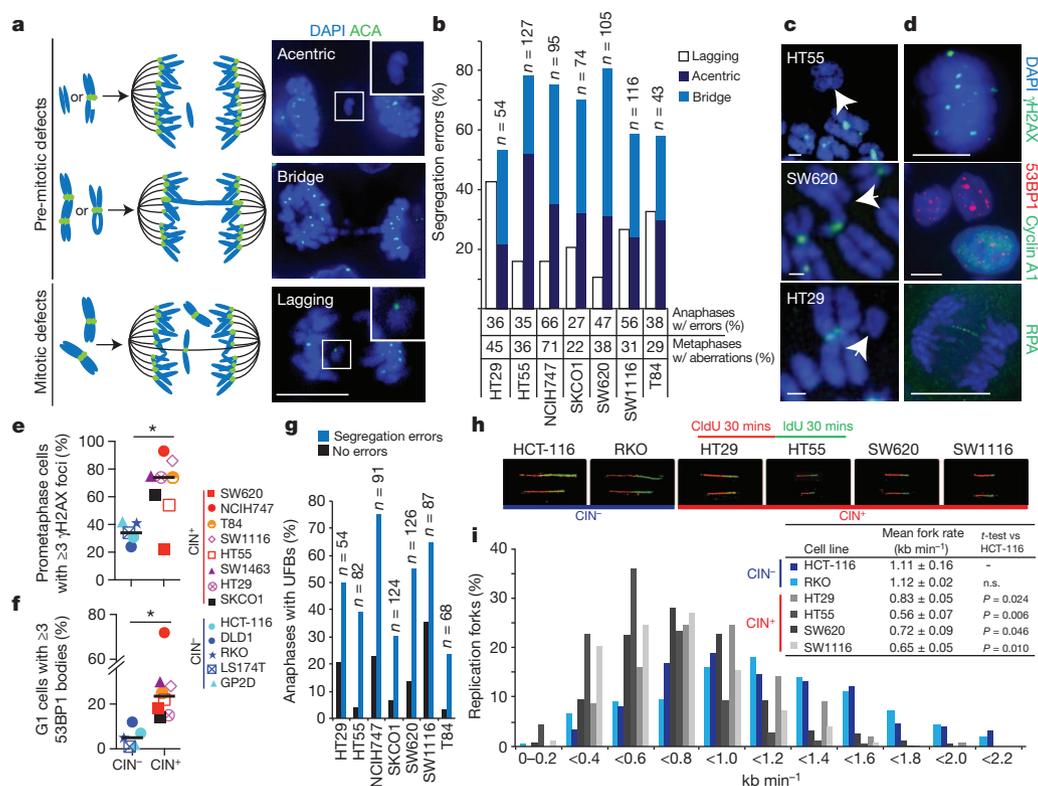
Figure 1 | Replication stress generates chromosome segregation errors in CIN+ cells. a, Schematic illustrating pre-mitotic and mitotic origins of chromosome segregation errors. Right panels denote example images in SW1116 (CIN+) cells stained with 4′,6-diamidino-2-phenylindole (DAPI) and anti-centromere antibodies (ACAs). Scale bar, 10 μm. **b**, Percentage of segregation errors in CIN+ cell lines that are lagging chromosomes, acentrics or anaphase bridges. Bridges extend fully between DNA masses; acentrics and lagging chromosomes were distinguished using ACA staining. Segregation errors not classifiable as bridges, lagging chromosomes or acentrics (<15%) are omitted for clarity. Percentage of anaphases showing segregation errors (n > 100 per cell line), and metaphases displaying structural aberrations (n = 38–84 per cell line) are shown. **c**, Structurally abnormal chromosomes identified on metaphase chromosome spreads, hybridized to an all-centromere probe (green) and stained with DAPI. Scale bars, 1 μm. **d**, Replication-stress-associated cellular phenotypes in NCIH747 (CIN+) cells stained with DAPI

and antibodies as indicated; γH2AX foci in prometaphase (top); 53BP1 bodies in G1 (cyclin A1−) cells (middle); anaphase UFBs, detected with antibodies for the single-stranded DNA-binding protein RPA[27] (bottom). Scale bars, 10 μm. **e**, Percentage of prometaphase DNA damage in CIN+ versus CIN− cells (n > 100 cells per cell line, *P = 0.033). **f**, Percentage of G1 53BP1 bodies in CIN+ versus CIN− cells (n > 250 cells per cell line, *P = 0.028). **g**, Percentage of anaphase UFBs. **h**, **i**, Four CIN+ and two CIN− cell lines were incubated sequentially with 5-chlorodeoxyuridine (CldU) and 5-iododeoxyuridine (IdU) for 30 min each. DNA fibre assays were performed and replication rates at individual replication forks were assessed. Representative fibres from each cell line are shown (**h**, original magnification, ×63). **i**, Distribution of replication fork rates (CldU, n > 300 forks in total per cell line, n = 3 experiments), with mean replication fork rates (CldU, n > 60 forks per experiment, mean ± s.e.m., n = 3 experiments) shown in the key (inset). Two-tailed t-test relative to HCT-116 cells.

(Supplementary Fig. 5c–e), consistent with impaired replication fork progression. These data indicate that increased replication stress, as demonstrated using established hallmarks and direct measurement of replication fork progression, may contribute substantially to chromosome missegregation in CIN+ CRC cells.

We sought a genetic basis for the increased replication stress and chromosome segregation errors in CIN+ cells. Examining whole-exome sequencing data in colorectal tumours from The Cancer Genome Atlas (TCGA)[19], and mutation status for 64 genes in CRC cell lines (COSMIC), showed that the only gene mutated at significantly higher frequencies in CIN+ tumours or cell lines was *TP53*, which is also mutated in 13–33% of CIN− tumours and cell lines (Supplementary Fig. 6a, b). Mutations in *TP53* are thought to be permissive for CIN rather than causative[20]. These data suggest that although mutations in known oncogenes or tumour suppressors may contribute to replication stress in cancer cells[14,21], they are unlikely to account exclusively for the increased replication stress in CIN+ CRC cells. We therefore addressed whether further genetic aberrations may contribute to CIN.

We proposed that regions of consistent somatic copy number loss in CIN+ CRC might encode CIN-suppressor genes, the loss of which could contribute to the induction of chromosome missegregation. To

identify CIN-specific regions of loss, comparative genomic hybridization (CGH) data were analysed for a cohort of 26 aneuploid colorectal tumours and 20 CIN+ cell lines. Chromosome 18q was most frequently subject to copy number loss, observed in 88% of aneuploid tumours and 80% of CIN+ cell lines (Fig. 2a, b and Supplementary Table 1), consistent with published studies in CRC and other tumour types[22–24]. Copy number losses in CIN+ tumours and cell lines were highly concordant (P < 0.001, Fisher's exact test; Supplementary Table 2). Next we examined the temporal sequence and consequences of 18q copy number loss in tumours. In a cohort of 28 adenomas (preinvasive tumour) with carcinoma in the same specimen (Fig. 2c), 18q loss of heterozygosity (LOH; indicative of copy number loss) was observed in 21 out of 28 (75%) carcinomas, compared with 10 out of 28 (35.7%) adenomas, implicating 18q loss in the adenoma–carcinoma transition. Chromosome 18q LOH was significantly associated with aneuploidy in both adenomas and carcinomas (Fig. 2d and Supplementary Fig. 6c).

To identify candidate CIN-suppressor genes encoded within regions of recurrent copy number loss, HCT-116 cells were transfected with pools of four short interfering RNAs (siRNAs) targeting the most frequently lost genes, present at ≤1 copy in at least 30% of CIN+ cell lines (94 genes encoded on chromosome 18q; Supplementary
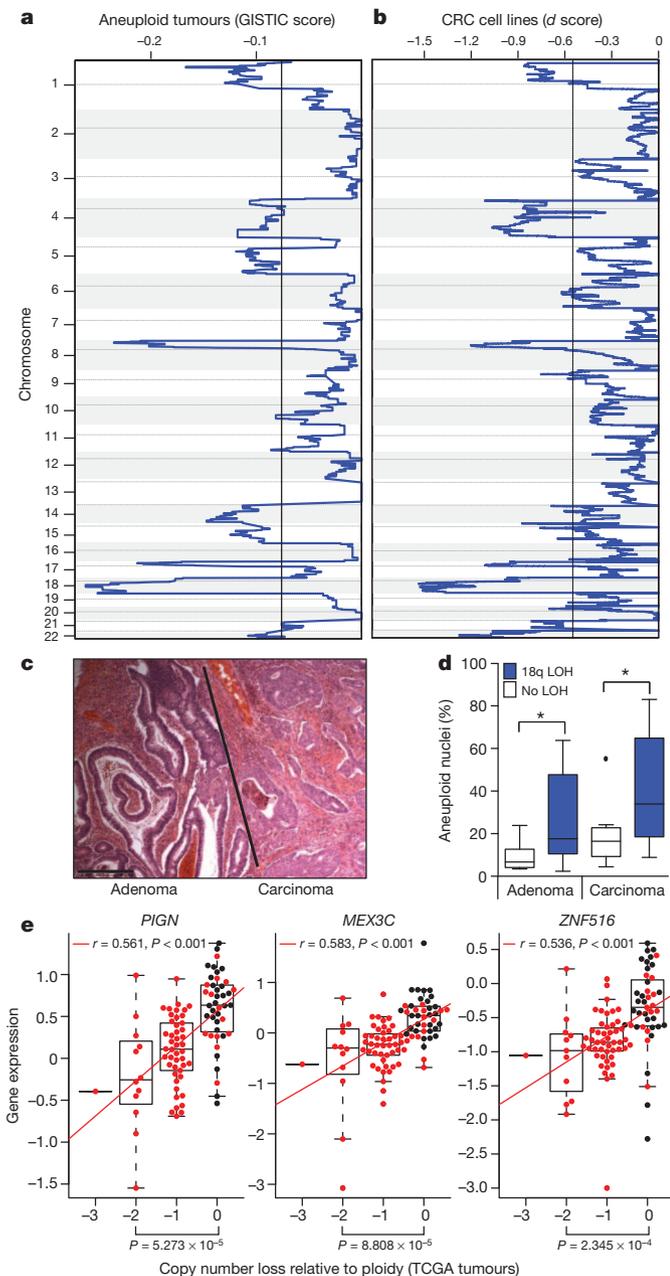
Figure 2 | Somatic copy number loss of chromosome 18q in CIN⁺ CRC.
a, GISTIC analysis for somatic copy number loss in 26 aneuploid colorectal tumours. $Q = 0.25$ determines significance (black line). **b**, Significance analysis of microarrays (SAM)-based d scores for somatic copy number loss in 20 CIN⁺ CRC cell lines compared to 9 CIN⁻ cell lines. $Q = 0.25$ determines significance, black line. **c**, Haematoxylin and eosin-stained tumour specimen, showing adenoma with adjacent carcinoma. Scale bar, 300 μm. **d**, Percentage of aneuploid nuclei, measured by DNA image cytometry, in paired adenomas and carcinomas ($n = 20$) with/without 18q LOH (Tukey box plot with outliers shown; *$P < 0.05$, two-tailed t-test). **e**, Spearman's rank correlation between mRNA expression and DNA copy number for *PIGN*, *MEX3C* and *ZNF516* in CIN⁻ ($n = 28$, black dots) versus CIN⁺ ($n = 74$, red dots) tumours (TCGA). Tumours were defined as CIN⁺ based on a weighted genome integrity index > 0.2 (see Methods). Statistic: Mann–Whitney U-test.

Table 3). After 48 h, anaphase segregation errors were quantified (Supplementary Fig. 7a). Importantly, this approach would identify both mitotic and pre-mitotic defects. siRNA pools inducing segregation errors to ≥3 standard deviations above the frequency in control-transfected cells were assessed in validation assays. All sequences targeting *PIGN*, *MEX3C* and *ZNF516* induced segregation errors and efficiently depleted messenger RNA levels (Supplementary Fig. 7b–d).

Additional independent siRNA sequences targeting each of the three genes also induced segregation errors (Supplementary Fig. 7e), and these genes were prioritized for further analysis, although we cannot exclude a contribution of other genes encoded on 18q to CIN. Of note, *MEX3C* is the only protein-coding gene located between two genes implicated in CRC carcinogenesis, *DCC* and *SMAD4*. Copy number loss of *PIGN*, *MEX3C* or *ZNF516* is observed in 85% of the 20 CIN⁺ cell lines and 84% of aneuploid tumours ($n = 103$) in the independent TCGA cohort[19], with loss of all three genes in 70% of CIN⁺ cell lines and 79% of aneuploid tumours (Supplementary Table 4). Importantly, reduced copy number correlated significantly with reduced mRNA expression for all three genes, both in the TCGA tumour cohort and CRC cell lines (Fig. 2e, Supplementary Fig. 8a and Supplementary Table 5).

We next ensured that off-target siRNA effects reported against the mitotic checkpoint protein MAD2 and the DNA repair protein RAD51 were not causing chromosome missegregation[25] (Supplementary Fig. 8b–e). Two siMEX3C sequences partially depleted MAD2 protein, and were excluded from further analysis. Furthermore, expression of exogenous green fluorescent protein (GFP)-tagged MEX3C rescued segregation errors induced after silencing endogenous MEX3C using a 3′-untranslated region (UTR)-targeted siRNA, and expression of siRNA-resistant PIGN–GFP and ZNF516–GFP reduced segregation errors induced by depletion of endogenous PIGN or ZNF516 (Supplementary Fig. 8f). Silencing each CIN suppressor also induced segregation errors in two further CIN⁻ CRC cell lines, DLD1 and RKO, and in 18q-normal CIN⁺ NCIH508 cells (Supplementary Fig. 9a–c). Induction of segregation errors was independent of oncogenic KRAS signalling in HCT-116 cells, as segregation errors were also induced after CIN-suppressor gene silencing in an isogenic KRAS^{wild-type/−} cell line (Supplementary Fig. 9d).

CIN-suppressor gene silencing (using siPIGN, siMEX3C and siZNF516 siRNAs) in HCT-116 cells primarily induced acentric chromosomes and anaphase bridges (Fig. 3a) in the absence of gross mitotic defects (data not shown), similar to observations in CIN⁺ CRC cell lines with 18q loss (Fig. 1b). This suggested a pre-mitotic origin for these chromosome segregation errors and, accordingly, we observed an increased frequency of structurally abnormal chromosomes (Fig. 3a, b and Supplementary Fig. 10a). To assess chromosome non-disjunction, HCT-116 cell lines stably expressing short hairpin RNAs (shRNAs) were constructed. Silencing each CIN-suppressor gene (using shPIGN, shMEX3C and shZNF516) increased segregation error frequency (Supplementary Fig. 10b), and single cell clones grown from each cell line displayed significantly increased intracolony deviation for chromosomes 2 and 15 (Fig. 3c, d and Supplementary Fig. 10c), indicating that CIN-suppressor gene silencing induces both structural and numerical instability.

We next examined replication stress after CIN-suppressor gene silencing. Increased prometaphase DNA damage was observed (Fig. 3e), and this increased concomitantly with, rather than after, the rise in segregation error frequency (Supplementary Fig. 10d), supporting the hypothesis that the observed DNA damage reflects a cause, rather than a consequence, of segregation errors. Silencing each CIN-suppressor gene also resulted in a increased number of 53BP1 bodies in G1 cells (Fig. 3f), which were unaffected by cytokinesis inhibition, and are therefore unlikely to reflect cytokinesis-induced chromosome damage[7,15] (Supplementary Fig. 10e–g). Silencing *PIGN* and *ZNF516* also significantly increased the frequency of UFBs (Fig. 3g). Consistent with loss of CIN-suppressor genes contributing to replication stress in CIN⁺ cells, transient co-expression of PIGN, MEX3C and ZNF516 resulted in a partial reduction in G1 53BP1 bodies in three CIN⁺ cell lines with 18q loss (Supplementary Fig. 11a–d). DNA fibre assays revealed a shift in distribution towards reduced replication fork speeds after silencing each of the CIN-suppressor genes in HCT-116 cells, with reduced average fork rates after silencing *PIGN* and *ZNF516* (Fig. 3h, i and Supplementary Fig. 12a, b). Furthermore, we observed
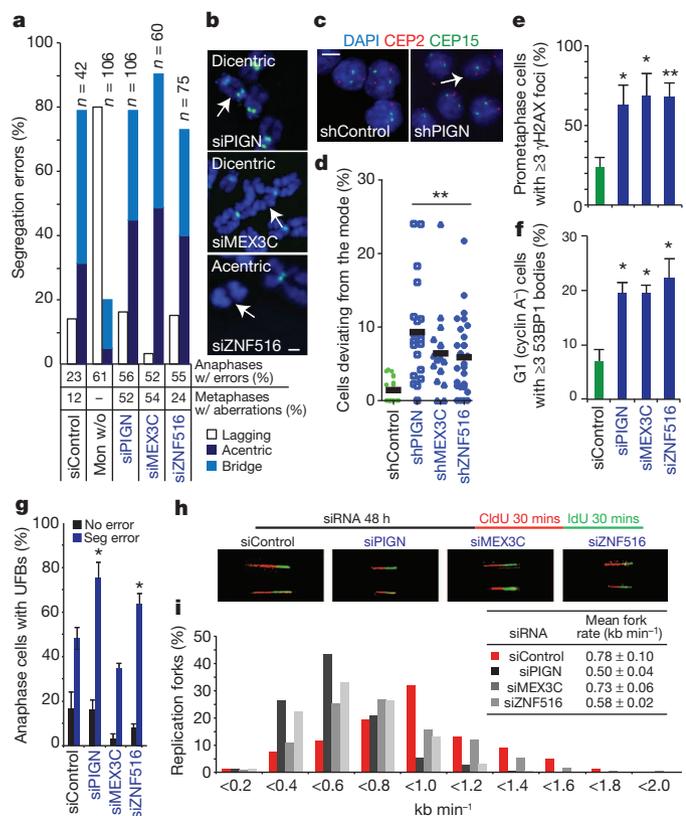
Figure 3 | Candidate suppressors of replication stress and CIN encoded on chromosome 18q. a, b, CIN-suppressor genes were silenced in HCT-116 (CIN$^-$) cells using siRNAs for 48 h. siControl denotes a non-targeting siRNA control. **a,** Percentage of segregation errors accounted for by lagging chromosomes, acentrics and anaphase bridges. Other segregation errors (<15%) are omitted for clarity. For comparison, segregation errors arising via improper chromosome attachments were induced by monastrol (mon) treatment (100 μM, 1 h, 75-min release). Percentages of anaphases displaying segregation errors and metaphases displaying ≥1 structurally abnormal chromosome (n > 100) are shown. **b,** Examples of structurally abnormal chromosomes as indicated. Scale bar, 1 μm. **c, d,** Cell lines stably expressing shRNAs as indicated were seeded at low density on glass slides to allow colony formation. Slides were fixed and hybridized to DNA probes for centromeres 2 and 15. **c,** Example images of control and shPIGN cells. Arrow indicates cell with loss of chromosome 15. Scale bar, 10 μm. **d,** Percentage deviation from the modal centromere copy number per colony (mean of two probes (CEP2 and CEP15)). Lines are median values; **P < 0.01, Dunn's multiple comparison test. **e–g,** HCT-116 cells were scored for replication-stress-associated phenotypes after siRNA-mediated CIN-suppressor gene silencing: percentage of prometaphases with ≥3 γH2AX foci (n > 100 per experiment) (**e**); percentage of G1 cells with ≥3 53BP1 bodies (n > 150 per experiment) (**f**); percentage of anaphases with UFBs (n = 100 per experiment) (**g**). Data in **e–g** are mean ± s.e.m., n = 3; *P < 0.05, **P < 0.01, two-tailed t-test. **h, i,** DNA fibre assays were performed after siRNA transfection as indicated. Representative fibre images for siRNA transfections are indicated as shown (**h**, original magnification, ×63). **i,** Distribution of replication fork rates (n > 200 forks in total per siRNA transfection, n = 2 experiments), with mean fork rates (n > 70 forks per experiment, mean ± s.d., n = 2 experiments) shown in the key (inset).

increased sister fork asymmetry after silencing each of the three genes (Supplementary Fig. 12c), consistent with impaired replication fork progression.

These data suggest that segregation errors resulting from CIN-suppressor gene silencing are driven by replication stress. To test this hypothesis further, HCT-116 cells transfected with siRNAs against CIN-suppressor genes were supplemented with nucleosides, previously shown to reduce replication-induced DNA damage[5]. Nucleosides significantly reduced the induction of segregation errors after silencing of *PIGN* (62% to 32%), *MEX3C* (57% to 36%) and *ZNF516* (43%

to 34%), whereas control segregation errors were unaffected (Fig. 4a). We then tested whether nucleoside supplementation could attenuate chromosome missegregation in CIN$^+$ cell lines with 18q loss. Nucleoside supplementation significantly reduced segregation error frequency by 45–55% and attenuated prometaphase DNA damage by 28–43% in four CIN$^+$ cell lines (Fig. 4b, c and Supplementary Fig. 13a), indicating suppression of replication-induced DNA damage and subsequent chromosome missegregation. The extent of the nucleoside-mediated reduction in segregation errors indicates that *de novo* generation of structural chromosome aberrations is responsible for a large proportion of chromosome missegregation events in CIN$^+$ cells. By contrast, nucleoside supplementation did not affect the segregation error frequency in four CIN$^-$ cell lines or 18q-normal CIN$^+$ NCIH508 cells (Supplementary Fig. 13b, c). Nucleoside supplementation did not affect proliferation, cell cycle distribution, or cellular ATP levels[5] (Supplementary Fig. 14a–f).

Our findings implicate replication stress as a major driver of chromosomal instability in CRC. In addition to impaired replication fork progression, CIN$^+$ cells exhibit replication-stress-associated DNA damage and structurally abnormal chromosomes that missegregate during mitosis, linking structural and numerical instability. Supplementing CIN$^+$ cells with nucleosides reduced both DNA damage and segregation errors, supporting a direct role for replication stress in driving CIN. Given the complex nature of replication stress and CIN, it is likely that there are several genetic aberrations contributing to these phenotypes within an individual tumour. Here, we suggest a putative genetic mechanism that may contribute to CIN in CRC, through the recurrent loss of a region on chromosome 18q, encoding three newly identified suppressors of replication stress and chromosome missegregation. The loss of chromosome 18q in many solid tumour types suggests the potential contribution of this locus to CIN in tumours beyond CRC. Furthermore, reports of DNA replication stress across several tumour types suggest replication stress may be a common route to CIN and intratumour heterogeneity[13,26]. Efforts to target or restrain replication stress may therefore provide a rational approach to limit tumour diversity, genome evolution and adaptation.
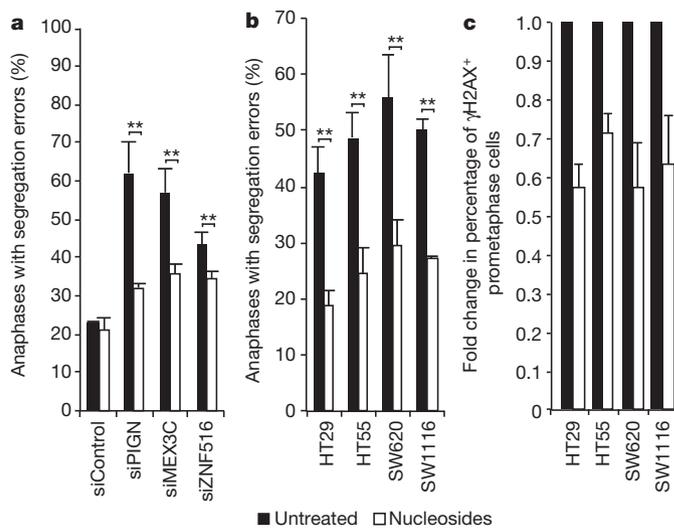
Figure 4 | Nucleoside supplementation reduces segregation error frequency and prometaphase DNA damage. a, Percentage of HCT-116 anaphases with segregation errors, with/without 48 h nucleoside supplementation 24 h after siRNA transfection (n = 30). **b,** Percentage of anaphases with segregation errors with/without 48 h nucleoside supplementation in CIN$^+$ cell lines (n = 30). **c,** Fold change in the percentage of prometaphases exhibiting ≥3 γH2AX foci with/without 48 h nucleoside supplementation in CIN$^+$ cell lines (n ≥ 100 cells per condition per cell line). Data are mean ± s.e.m. n = 3 experiments; **P < 0.01, two-tailed t-test. Unnormalized data are shown in Supplementary Fig. 13a.

METHODS SUMMARY

Cells were maintained at 37 °C at 5% CO_2. FISH was performed as described[11]. The all-centromere probe was used as per manufacturer's instructions. Image acquisition was performed using a Deltavision microscope in a temperature and CO_2-controlled chamber. siRNA transfections were at 40 nM, using Lipofectamine RNAi-max (Invitrogen). Bioinformatics analysis was performed using R.

Full Methods and any associated references are available in the online version of the paper.

1. Gerlinger, M. *et al.* Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N. Engl. J. Med.* **366,** 883–892 (2012).
2. Ding, L. *et al.* Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature* **481,** 506–510 (2012).
3. Lengauer, C., Kinzler, K. W. & Vogelstein, B. Genetic instabilities in human cancers. *Nature* **396,** 643–649 (1998).
4. McGranahan, N., Burrell, R. A., Endesfelder, D., Novelli, M. R. & Swanton, C. Cancer chromosomal instability: therapeutic and diagnostic challenges. *EMBO Rep.* **13,** 528–538 (2012).
5. Bester, A. C. *et al.* Nucleotide deficiency promotes genomic instability in early stages of cancer development. *Cell* **145,** 435–446 (2011).
6. Thompson, S. L. & Compton, D. A. Chromosomes and cancer cells. *Chromosome Res.* **19,** 433–444 (2011).
7. Janssen, A., van der Burg, M., Szuhai, K., Kops, G. J. & Medema, R. H. Chromosome segregation errors as a cause of DNA damage and structural chromosome aberrations. *Science* **333,** 1895–1898 (2011).
8. Crasta, K. *et al.* DNA breaks and chromosome pulverization from errors in mitosis. *Nature* **482,** 53–58 (2012).
9. Pampalona, J., Soler, D., Genesca, A. & Tusell, L. Whole chromosome loss is promoted by telomere dysfunction in primary cells. *Genes Chromosom. Cancer* **49,** 368–378 (2010).
10. Ichijima, Y. *et al.* DNA lesions induced by replication stress trigger mitotic aberration and tetraploidy development. *PLoS ONE* **5,** e8821 (2010).
11. Thompson, S. L. & Compton, D. A. Examining the link between chromosomal instability and aneuploidy in human cells. *J. Cell Biol.* **180,** 665–672 (2008).
12. Gisselsson, D. Classification of chromosome segregation errors in cancer. *Chromosoma* **117,** 511–519 (2008).
13. Bartkova, J. *et al.* DNA damage response as a candidate anti-cancer barrier in early human tumorigenesis. *Nature* **434,** 864–870 (2005).
14. Tort, F. *et al.* Retinoblastoma pathway defects show differential ability to activate the constitutive DNA damage response in human tumorigenesis. *Cancer Res* **66,** 10258–10263 (2006).
15. Lukas, C. *et al.* 53BP1 nuclear bodies form around DNA lesions generated by mitotic transmission of chromosomes under replication stress. *Nature Cell Biol.* **13,** 243–253 (2011).
16. Chan, K. L., Palmai-Pallag, T., Ying, S. & Hickson, I. D. Replication stress induces sister-chromatid bridging at fragile site loci in mitosis. *Nature Cell Biol.* **11,** 753–760 (2009).
17. Kawabata, T. *et al.* Stalled fork rescue via dormant replication origins in unchallenged S phase promotes proper chromosome segregation and tumor suppression. *Mol. Cell* **41,** 543–553 (2011).
18. Harrigan, J. A. *et al.* Replication stress induces 53BP1-containing OPT domains in G1 cells. *J. Cell Biol.* **193,** 97–108 (2011).
19. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487,** 330–337 (2012).
20. Bunz, F. *et al.* Targeted inactivation of p53 in human cells does not result in aneuploidy. *Cancer Res.* **62,** 1129–1133 (2002).
21. Halazonetis, T. D., Gorgoulis, V. G. & Bartek, J. An oncogene-induced DNA damage model for cancer development. *Science* **319,** 1352–1355 (2008).
22. Rowan, A. *et al.* Refining molecular analysis in the pathways of colorectal carcinogenesis. *Clin. Gastroenterol. Hepatol.* **3,** 1115–1123 (2005).
23. Pasello, G. *et al.* DNA copy number alterations correlate with survival of esophageal adenocarcinoma patients. *Mod. Pathol.* **22,** 58–65 (2009).
24. Yatsuoka, T. *et al.* Association of poor prognosis with loss of 12q, 17p, and 18q, and concordant loss of 6q/17p and 12q/18q in human pancreatic ductal adenocarcinoma. *Am. J. Gastroenterol.* **95,** 2080–2085 (2000).
25. Sigoillot, F. D. *et al.* A bioinformatics method identifies prominent off-targeted transcripts in RNAi screens. *Nature Methods* **9,** 363–366 (2012).
26. Dereli-Oz, A., Versini, G. & Halazonetis, T. D. Studies of genomic copy number changes in human cancers reveal signatures of DNA replication stress. *Mol. Oncol.* **5,** 308–314 (2011).
27. Chan, K. L. & Hickson, I. D. On the origins of ultra-fine anaphase bridges. *Cell Cycle* **8,** 3065–3066 (2009).

METHODS

Cell lines. Cells were maintained at 37 °C with 5% CO_2 in DMEM with L-glutamine (Gibco) or RPMI 1640 media (NCIH747 cells; Gibco), supplemented with 10% FBS, and 1/10,000 units of penicillin–streptomycin (Sigma). Cell-line CIN status was defined as described previously[28].

Cell treatments. All compounds were from Sigma. Nocodazole: 50 or 100 ng ml^{-1}; aphidicolin: 0.2 µM, 24 h; monastrol wash-out: 100 µM, 1 or 16 h, washed three times in fresh medium before 75 min recovery; blebbistatin: 100 µM, 4 h. The nucleosides adenosine, cytidine, guanosine and uridine were freshly prepared for each experiment, filter sterilized, and used at 0.3 or 30 µM. H_2O_2: 350 µM, 4 h, before 16 h recovery.

Metaphase spreads and clonal FISH preparation. Metaphase spreads: cells were collected after 1-h treatment with 10 mM colcemid (Gibco), and swelled with KCl (0.4%, 37 °C, 7 min) before fixation in 3:1 methanol:acetic acid. Cells were dropped onto glass slides and aged for ~2 weeks. All-human centromere probe (Poseidon) was used according to the manufacturer's instructions.

Clonal FISH: 500 cells were expanded into colonies of 30–60 cells on glass slides before KCl treatment and fixation in 3:1 methanol:acetic acid. Slides were denatured (70 °C in 2× sodium citrate (SSC)/75% formamide, 2 min, quenched in ice-cold 70% ethanol) and dehydrated through an ethanol series. Probes (CEP2 (D2Z1 Spectrum orange), CEP15 (D15Z1 Spectrum green; Abbott Molecular Probes)) were denatured (90 °C, 6 min) and hybridized to slides (16 h, 37 °C), and then washed (50% formamide/2× SSC and 2× SSC at 42 °C, followed by 4× SSC and PBS washes at room temperature). Slides were dehydrated and mounted in Vectashield hardset plus DAPI mounting medium (H-1500).

Immunofluorescence. Cells grown on coverslips were fixed in: 10% Triton X-100, 1 M PIPES, 0.5 M EGTA, 1 M $MgCl_2$ and 4% formaldehyde. Mouse primary antibodies: α-tubulin (1:1,000; Sigma T6074), NDC80 (1:800; Abcam Ab3613), centrin3 (1:1,000; Abcam Ab54531), cyclin A1 (1:350; Santa Cruz sc-56299), phospho-histone H2A.X Ser 139 (1:500; Millipore 05-636), and RPA (1:500; Neomarkers MS-691-P0). Rabbit primary antibodies: 53BP1 (1:500; Santa Cruz sc-22760), β-tubulin (1:1,000, Abcam), and human ACAs (1:250; Antibodies Incorporated). Secondary antibodies (1:500; Molecular Probes): goat anti-mouse conjugated to AlexaFluor 488 (A11017), goat anti-rabbit AF594 (A11012), and goat anti-human AF647 (A21445). DNA was stained with DAPI (Roche) and coverslips mounted in Vectashield (Vector H-1000). Preparation and immuno-staining of metaphase spreads was performed as described previously[29]. Most images were acquired using an Olympus DeltaVision RT microscope (Applied Precision, LLC) equipped with a Coolsnap HQ camera. Three-dimensional image stacks were acquired in 0.2-µm steps, using an Olympus ×100 or ×60 1.4 numerical aperture UPlanSApo oil immersion objective. Deconvolution of image stacks and quantitative measurements was performed with SoftWorx Explorer (Applied Precision, LLC).

DNA fibre assays. Cells were plated (with/without siRNA transfection) 48 h before sequential pulse-labelling with CldU and IdU (Sigma Aldrich, 30 min each) and collection for DNA fibre assays as described previously[30].

RNA interference. siRNA transfections were performed at 40 nM by reverse transfection with Lipofectamine RNAiMax (Invitrogen). See Supplementary Table 6 for sequences used. The screen was performed in 12-well plates (0.5×10^5 HCT-116 cells per well on coverslips) using siGENOME SMARTpools (Dharmacon) with one control well per plate (Dharmacon control no.2). After 48 h, cells were fixed and stained for α-tubulin. Thirty anaphases per siRNA pool were scored manually for segregation errors.

DNA transfections and GFP-tagged construct mutagenesis. H2B-mRFP cells: cells were transfected with pH2B-mRFP (gift from A. Straube) using Fugene 6.0 (Promega), and selected in 1 mg ml^{-1} G418 before flow-sorting for mRFP expression. Cells were maintained in 500 µg ml^{-1} G418.

shRNA cell-line synthesis: HCT-116 cells were transfected with shRNA plasmids (Open Biosystems, see Supplementary Table 7) using Fugene 6.0 (Promega) and selected in 0.5 µM puromycin for 2–3 weeks. Three colonies per shRNA were amplified, RNA was extracted and silencing assessed by quantitative PCR (qPCR). Colonies with efficient silencing were selected. Cell lines were maintained in 0.5 µM puromycin.

PIGN/MEX3C/ZNF516–GFP transfections: for siRNA rescue experiment in Supplementary Fig. 8f, 1 µg (siRNA-insensitive PIGNins–GFP and MEX3C–GFP) or 0.5 µg (ZNF516ins–GFP) DNA was either co-transfected with siRNA using Lipofectamine RNAiMax for 48 h (MEX3C) or transfected 24 h after siRNA transfection using Lipofectamine 2000 (Invitrogen) for 24 h (PIGN and ZNF516). For co-transfection into CIN$^+$ cell lines, 2 µg total DNA was transfected using Lipofectamine LTX plus (Invitrogen), according to the manufacturer's instructions. PIGN–GFP and MEX3C–GFP were from Geneocopoeia, ZNF516–GFP was from Origene.

siRNA-insensitive mutagenesis: the Quikchange XL site-directed mutagenesis kit (Stratagene) was used to create three base mismatches in siRNA target sequences for PIGN oligo 3 and ZNF516 oligo 3 in PIGN–GFP and ZNF516–GFP, respectively. Mutagenesis was confirmed by sequencing.

RNA extraction and reverse transcriptase qPCR. RNA was extracted using the Qiagen RNeasy kit. Reverse transcription was performed using the first-strand cDNA synthesis kit (Amersham). qPCR was performed in 96-well plates using pre-designed TaqMan probe/primers on a ABI 7900HT system (Applied Biosystems). All reactions were performed in duplicate. The relative amount of mRNA was calculated using the comparative C_t method after normalization to *GAPDH* expression.

Western blotting. Cellular protein extracts were separated on NuPAGE 4–12% Bis-Tris gels (Invitrogen) then transferred to polyvinylidene fluoride membrane (Millipore). Membranes were incubated with antibodies: MAD2 (mouse 1:1,000; BD-Biosciences 610678), RAD51 (mouse 1:1,000; Abcam ab213), GFP (mouse 1:1,000; Santa Cruz sc-9996), and Turbo–GFP (rabbit 1:1,000; Evrogen) in 5% milk in TBS, and detected with a horseradish peroxidase (HRP)-conjugated secondary antibody (1:10,000; Dako) and chemiluminescence (ECL, Amersham Biosciences). Loading was quantified with HRP-conjugated anti-β-actin (1:100,000; Sigma).

Time-lapse microscopy. H2B-mRFP-labelled cells were grown in 8-well chamber slides (LabTek). Fourteen-micrometre z-stacks (seven images) were acquired using an Olympus ×40 1.3 numerical aperture UPlanSApo oil immersion objective every 3 min for 6 h using a DeltaVision microscope in a temperature and CO_2-controlled chamber. Analysis was performed using Softworx Explorer.

Flow cytometry. Mitotic index: cells were fixed in 70% ethanol and stained with a mouse anti-MPM2 antibody (3:500, overnight at 4 °C; Millipore 05-368) and then a goat anti-mouse AF647 (Molecular Probes; A21463) and DAPI.

8-oxo-guanine: cells (with/without H_2O_2 treatment) were fixed in 4% formaldehyde, DNA denatured in 2 M HCl for 20 min, and stained with mouse anti-8-oxoguanine (1:200; Abcam ab62623). Data were processed using FlowJo software.

Immunohistochemistry. Paraffin-embedded cell pellets (with/without H_2O_2 treatment), were sectioned at 4 µm, de-waxed in xylene, then rehydrated through ethanol series to water. Endogenous peroxidase was blocked (1.6% H_2O_2, 10 min), followed by incubation with 10% normal horse serum (30 min). Sections were incubated with mouse anti-8-oxoguanine (1:1,000, 1 h), washed three times in PBS, then incubated with biotinylated horse anti-mouse antibodies (1:400, 35 min; Vector Labs). After washing, peroxidase substrate (DAB) was added (2 min), slides were washed in water and counterstained with haematoxylin. Slides were washed, dehydrated and mounted in DPX-type mountant.

Proliferation assays. Plates were imaged using an IncuCyte long-term in-situ cell imaging system, within an incubator. Phase-contrast images were acquired every 2 h for 72 h and the percentage of cell monolayer confluence was determined automatically. Outlying wells were excluded, and growth curves constructed. Growth rates were calculated by measuring the gradient of the linear growth phase.

ATP measurement. After nucleoside supplementation, cells were treated with Cell Titer Glo reagent (Promega). Control measurements were taken from wells containing media only with/without nucleosides. ATP levels were normalized to the biomass per well (*in vitro* toxicology assay kit sulforhodamine B solution, Sigma).

SNP 6.0 array processing. Cell lines: Affymetrix single nucleotide polymorphism (SNP) 6.0 data were obtained for 20 CIN$^+$ and 9 CIN$^-$ cell lines (Wellcome Trust Sanger Institute). Integer copy numbers were estimated for each SNP probe using the PICNIC algorithm[31].

TCGA: Affymetrix SNP 6.0 data were downloaded for 365 CRC samples and logR ratios and allelic differences were estimated by the Affymetrix Genotyping Console. Samples that failed the Affymetrix quality control parameters were excluded. All tumours with <60% tumour nuclei (based on pathological estimates of adjacent sections) were excluded. To estimate copy number, the GAP algorithm was used[32]. Ploidy was estimated by calculating the weighted median copy number across all copy number segments, with weights equal to the segment length. Copy number segments of loss and gain were defined relative to the ploidy status of each sample by subtracting the ploidy estimate from the estimated copy number of the segment.

Karyotypic complexity scores. The structural complexity score was defined as the sum of all structurally aberrant regions. Regions of intrachromosomal gain and loss were defined relative to the modal copy number of the chromosome, and each region counted as one structural aberration. To avoid overestimation, aberrant regions <1 megabase were excluded. The numerical complexity score was the sum of all whole chromosome gains and losses (chromosomes with >75% of SNP copy number values higher or lower than the ploidy of the sample were counted as whole chromosome gains or losses, respectively). Several copy number events affecting the same chromosome were scored separately (for example, −2 copies =

2 chromosome losses). The structural and numerical complexity scores were divided by 1.5 for triploid cell lines, and by 2 for tetraploid cell lines, to account for the increased likelihood of karyotypic abnormalities in polyploid genomes.

Weighted genome instability index: because FACS-based DNA index measures were not available for the TCGA tumours, and information about the microsatellite instability status was unavailable for a sufficient number of tumours, an alternative means of classification was required. The genome instability index (GII)[33] is the percentage of SNPs across the genome present at an aberrant copy number, relative to the baseline ploidy of the sample. We adapted the GII to account for variation in chromosome size, so that large chromosomes do not have a greater effect on the score than small chromosomes: the percentage of aberrant SNPs for each chromosome was calculated separately, and the mean percentage aberration was then calculated across all 22 chromosomes. To define a threshold for CIN^- versus CIN^+, the weighted GII was calculated for the cell lines. A threshold of 0.2 accurately distinguished CIN^+ from CIN^-, as previously defined[28]. The same threshold was then applied to the TCGA cohort of tumours.

Sequencing data (TCGA). Preprocessed level 2 somatic mutation data was obtained from the TCGA[19] for 101 colorectal tumours for which SNP 6.0 copy number data was also available. All genes with ≤ 5 somatic mutations were excluded from the analysis. Colorectal tumours were classified as CIN^+ using weighted GII > 0.2. Somatic mutation data for the 29 colorectal cancer cell lines were obtained from the COSMIC database (http://cancer.sanger.ac.uk/cancergenome/projects/cosmic/). The association of CIN status with somatic mutation status of *TP53*, *APC*, *SMAD4* and *KRAS* was tested with one-sided Fisher's exact tests, and the *P* values were adjusted for multiple testing with the Benjamini and Hochberg method[34].

Defining somatic copy number losses in CIN^+ versus CIN^- tumours and cell lines. Aneuploid tumours: bacterial artificial chromosome (BAC) array-CGH data were obtained for 26 aneuploid tumour samples, and segmented by circular binary segmentation (R package DNAcopy). The genomic identification of significant targets in cancer (GISTIC) algorithm[35] was used to identify regions of consistent gain and loss, with thresholds of 0.1 or -0.1 for gain or loss, respectively, and a *Q* value threshold of 0.25. Aneuploidy was defined by flow cytometry (DNA index >1.2).

Cell lines: minimum consistent regions of genomic alteration across all cell lines were assessed for DNA copy number. Each region in each cell line was normalized to the ploidy baseline of the cell line to give X_{norm} and was then defined as either lost (copy number $<$ ploidy baseline) or not lost (copy number $\geq$ ploidy baseline) and set to 0. Each region was assessed for gain in the same manner. To test for statistical significance between CIN^+ and CIN^- cell lines, a *d* score for each lost region was computed by calculating the mean normalized copy number X_{norm} across CIN^+ (mean($X_{norm,C}$)) and CIN^- (mean($X_{norm,M}$)) cell lines, thereby accounting for both amplitude and frequency of genomic aberrations. SAM (R package siggenes) was then performed with a modified two-sample t-statistic:

$$d(i) = \frac{(\text{mean}(X_{norm,C}(i)) - \text{mean}(X_{norm,M}(i))}{s(i) + s_0}$$

The parameter $s(i)$ defines the region-specific standard deviation[36]. In contrast to a standard two-sample *t*-statistic, SAM includes an additional parameter s_0, which decreases the influence of high sample variance. This was empirically set to 0.5, resulting in a balanced weighting of frequency and amplitude. To detect significant regions, we randomly permuted ($N = 10,000$) SNP probes for each sample separately. To save computation time, we randomly drew copy numbers for each sample, setting the probability for a given copy number to the percentage of SNP probes showing this copy number level across the genome. For each tested region, *P* values were estimated by counting the percentage of permutation *d* scores greater or equal than the observed *d* score. To adjust for multiple testing, *Q* values were estimated with the R-package qvalue and genes with $Q < 0.25$ were called significant. To ensure selection of genes consistently altered across CIN^+ cell lines, genomic changes not seen in $\geq 50\%$ of cell lines were excluded from

further analysis. Genes were then mapped to regions using the R package BioMart[37]. All genes present at ≤ 1 copies in $\geq 30\%$ of CIN^+ cell lines, and no more than 1 CIN^- cell line, were selected for functional investigation.

Carcinoma-in-adenoma samples. Twenty archival formalin-fixed and paraffin-embedded tumours showing adjacent but discrete colorectal carcinoma and adenoma were identified. Samples and records were used in accordance with UK research ethics approval (MREC06/Q1702/99). Haematoxylin and eosin slides of the samples were reviewed, regions marked as normal (if present)/adenoma/carcinoma, and used as a guide to take tissue from each region from unstained slides by needle microdissection. DNA was extracted with standard proteinase K digestion, and purified (DNeasy kit, Qiagen). Ploidy analysis was performed using automated image-based cytometry (Fairfield Imaging) as previously described[38,39].

LOH analysis: SNP arrays: the Illumina Sentrix Beadarray human linkage mapping panel version IVb was used according to the Goldengate genotyping assay instructions (Illumina). DNA was amplified, fragmented and hybridized to the Beadchip, followed by single-base extension. Beadchips were stained, dried and scanned using a Beadarray reader (Illumina). Image data were visualized using Genomestudio (Illumina). All samples had call rates above 0.97. Adenomas and carcinomas were marked as having LOH or no LOH in 18q according to the absence or presence of heterozygous alleles respectively.

LOH analysis using microsatellites: standard PCR conditions were used with the forward primer fluorescently labelled with the fluorescent dyes HEX or FAM. At each marker, LOH was considered present when a peak area in the adenoma or the carcinoma was reduced to 50% of the other allele, relative to the normal paired DNA. Up to five microsatellites in 18q21 (D18S46, D18S1110, D18S35, D18S69 and D18S1407) were analysed. All PCRs were performed twice and all samples analysed with SNP arrays had concordant results.

Gene expression analysis. TCGA: Agilent 244 K custom gene expression (G4502A-07-3) data were downloaded for 154 CRC samples and print-tip normalized with the R-package marray. Duplicated probes were averaged.

Cell lines: Affymetrix HT-HGU133A microarray data for 20 CIN^+ and 9 CIN^- cell lines were obtained from the Wellcome Trust Sanger Institute. The data were RMA-normalized with the R-package affy. Spearman's rank correlation coefficient was used for the correlation of copy number and expression data.

28. Lee, A. J. *et al.* Chromosomal instability confers intrinsic multidrug resistance. *Cancer Res.* **71,** 1858–1870 (2011).
29. Cesare, A. J. *et al.* Spontaneous occurrence of telomeric DNA damage response in the absence of chromosome fusions. *Nature Struct. Mol. Biol.* **16,** 1244–1251 (2009).
30. Groth, P. *et al.* Methylated DNA causes a physical block to replication forks independently of damage signalling, O^6-methylguanine or DNA single-strand breaks and results in DNA damage. *J. Mol. Biol.* **402,** 70–82 (2010).
31. Greenman, C. D. *et al.* PICNIC: an algorithm to predict absolute allelic copy number variation with microarray cancer data. *Biostatistics* **11,** 164–175 (2010).
32. Popova, T. *et al.* Genome Alteration Print (GAP): a tool to visualize and mine complex cancer genomic profiles obtained by SNP arrays. *Genome Biol.* **10,** R128 (2009).
33. Chin, S. F. *et al.* High-resolution aCGH and expression profiling identifies a novel genomic subtype of ER negative breast cancer. *Genome Biol.* **8,** R215 (2007).
34. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Stat. Soc.* **57,** 289–300 (1995).
35. Beroukhim, R. *et al.* Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma. *Proc Natl Acad Sci USA* **104,** 20007–20012 (2007).
36. Storey, J. D. & Siegmund, D. Approximate *P*-values for local sequence alignments: numerical studies. *J. Comput. Biol.* **8,** 549–556 (2001).
37. Durinck, S. *et al.* BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics* **21,** 3439–3440 (2005).
38. Thirlwell, C. *et al.* Clonality assessment and clonal ordering of individual neoplastic crypts shows polyclonality of colorectal adenomas. *Gastroenterology* **138,** 1441–1454 (2010).
39. Leedham, S. J. *et al.* Clonality, founder mutations, and field cancerization in human ulcerative colitis-associated neoplasia. *Gastroenterology* **136,** 542–550 (2009).