

ORIGINAL ARTICLE

The evolution of single cell-derived colorectal cancer cell lines is dominated by the continued selection of tumor-specific genomic imbalances, despite random chromosomal instability

Darawalee Wangsa¹, Rüdiger Braun^{1,†}, Madison Schiefer^{1,†}, Edward Michael Gertz^{2,†}, Daniel Bronder^{1,†}, Isabel Quintanilla³ Hesed M. Padilla-Nash¹, Irianna Torres¹, Cynthia Hunn¹, Lidia Warner¹, Floryne O. Buishand^{1,4}, Yue Hu¹, Daniela Hirsch⁵, Timo Gaiser⁵, Jordi Camps^{1,3}, Russell Schwartz⁶, Alejandro A. Schäffer², Kerstin Heselmeyer-Haddad^{1,*} and Thomas Ried¹

¹Genetics Branch, Center for Cancer Research, National Cancer Institute/National Institutes of Health, Bethesda, MD 20892, USA, ²Computational Biology Branch, National Center for Biotechnology Information, National Institutes of Health, Bethesda, MD 20892, USA, ³Unitat de Biologia Cel·lular i Genètica Mèdica, Departament de Biologia Cel·lular, Fisiologia i Immunologia, Facultat de Medicina, Universitat Autònoma de Barcelona, Bellaterra, 08193, Spain, ⁴Department of Clinical Sciences of Companion Animals, Faculty of Veterinary Medicine, Utrecht University, Utrecht, 3584, The Netherlands, ⁵Institute of Pathology, University Medical Center Mannheim, University of Heidelberg, Heidelberg, 68167, Germany and ⁶Departments of Biological Sciences and Computational Biology, Carnegie Mellon University, Pittsburgh, PA 15213, USA

*To whom correspondence should be addressed. Tel: +240-760-6120; Fax: +240-541-4441; Email: heselmek@mail.nih.gov
Correspondence may also be addressed to Thomas Ried. Tel: +240-760-6120; Fax: +240-541-4441; Email: riedt@mail.nih.gov

[†]These authors contributed equally to this work.

Abstract

Intratumor heterogeneity is a major challenge in cancer treatment. To decipher patterns of chromosomal heterogeneity, we analyzed six colorectal cancer cell lines by multiplex interphase FISH (miFISH). The mismatch-repair-deficient cell lines DLD-1 and HCT116 had the most stable copy numbers, whereas aneuploid cell lines (HT-29, SW480, SW620 and H508) displayed a higher degree of instability. We subsequently assessed the clonal evolution of single cells in two colorectal carcinoma cell lines, SW480 and HT-29, which both have aneuploid karyotypes but different degrees of chromosomal instability. The clonal compositions of the single cell-derived daughter lines, as assessed by miFISH, differed for HT-29 and SW480. Daughters of HT-29 were stable, clonal, with little heterogeneity. Daughters of SW480 were more heterogeneous, with the single cell-derived daughter lines separating into two distinct populations with different ploidy (hyper-diploid and near-triploid), morphology, gene expression and tumorigenicity. To better understand the evolutionary trajectory for the two SW480 populations, we constructed phylogenetic trees which showed ongoing instability in the daughter lines. When analyzing the evolutionary development over time, most single cell-derived daughter lines maintained their major clonal pattern, with the exception of one daughter line that showed a switch involving a loss of APC. Our meticulous analysis of the clonal evolution and composition of these colorectal cancer models shows that all chromosomes are subject to segregation errors, however, specific net genomic imbalances are maintained. Karyotype evolution is driven by the necessity to arrive at and maintain a specific plateau of chromosomal copy numbers as the drivers of carcinogenesis.

Abbreviations

| | |
|--------|------------------------------------|
| CRC | colorectal carcinomas |
| FISH | fluorescence in situ hybridization |
| ITH | intratumor heterogeneity |
| miFISH | multiplex interphase FISH |

Introduction

One of the defining characteristics of cancer cells is the ability to acquire chromosomal aneuploidies that result in cancer-specific patterns of genomic imbalances (1–3). These imbalances are present in primary tumors and maintained in derived cancer cell lines. For instance, cervical carcinomas and derived cell lines invariably carry additional copies of the long arm of chromosome 3 (4,5) while colorectal carcinomas (CRCs) and derived cell lines are defined by recurrent gains of chromosomes and chromosome arms 7, 8q, 13 and 20q, along with losses of 8p, 17p and 18q (6,7). On the single cell level, however, one can observe considerable chromosomal instability, which results in intratumor heterogeneity (ITH) (8–11). Despite ITH, chromosomes that are frequently gained are rarely lost, and chromosomes that are commonly lost are rarely gained in the cancer cell population. This concept has been called ‘speciation’ by Duesberg *et al.* (12).

ITH has clinical implications, because it may facilitate selection of clones with chromosomal imbalance patterns and gene mutations with the propensity for metastasis and treatment resistance (13–15). It is therefore important to understand the degree of chromosomal instability, the ensuing ITH, the dynamics of its development and the consequences on the tumor population.

We previously studied the clonal composition of synchronous ductal carcinomas *in situ* and invasive carcinomas, using fluorescence in situ hybridization (FISH) of formalin-fixed paraffin-embedded patient samples. Our study revealed a considerable degree of ITH, yet the continued selection for a specific pattern of genomic imbalances in the tumor populations remained (10). Consistent with our findings, other studies show that samples collected from the same tumor display unique gene mutations, which activate different pathways, again confirming the heterogeneity of cancer cell populations that might explain treatment failure and disease recurrence (16–20).

Many functional cancer studies use cell lines established from primary tumors (21,22). Here, we evaluated whether chromosomal heterogeneity is maintained *in vitro* in six CRC cell lines. We subsequently assessed the clonal evolution from a single cell in the two aneuploid cell lines SW480 and HT-29, which differ in their degree of chromosomal instability. To this end, we measured the clonal composition for each single cell-derived clone by multiplex interphase FISH (miFISH) over time and reconstructed its clonal evolution by phylogenetic tree modelling (23,24). Alterations of gene expression, gene mutation patterns, growth rates, morphology and tumorigenicity were additionally assessed for each single cell-derived clone.

Methods**Cell lines and generation of single cells**

All colorectal cell lines (DLD-1, HCT116, H508, SW620, HT-29 and SW480) were purchased from ATCC and cultured with RPMI-1640 or McCoy's medium supplemented with antibiotics, 10% fetal bovine serum and 5% CO₂ at 37°C. The cell lines were tested and authenticated in the past 6 months via spectral karyotyping (SKY). To generate single cell clones, a suspension of bulk parental cells was flow-sorted by fluorescence-activated cell sorting into 96-well plates (*i.e.* side scatter). Each well was then

examined to ensure that only wells harboring a single cell were used for further culturing. After single cell clones were grown to ~70–80% confluency in a T25 flask, we extracted DNA, RNA and fixed cells for miFISH using cells from the same passage (Supplementary Figure 1, available at Carcinogenesis Online).

Multiplex interphase FISH (miFISH)

Detailed experimental procedures are described in Supplementary Materials and Methods, available at Carcinogenesis Online. BAC contigs were assembled for 12 locus-specific identifier probes of the following genes: COX2 (1q31.1), TERC (3q26), APC (5q22), EGFR (7p11), MYC (8q24.21), CCND1 (11q13.3), CDX2 (13q12), CDH1 (16q22.1), TP53 (17p13.1), HER2 (17q12), SMAD4 (18q21) and ZNF217 (20q13.2). FISH probes were chosen based on tumor suppressors and oncogenes known to be involved in CRC. Contigs consisted of 2–4 overlapping clones spanning 333–687 kbp genomic sequences. The probes were combined into three FISH panels (panel 1: TERC-COX2-APC-EGFR, panel 2: CDH1-HER2-TP53-ZNF217, panel 3: CDX2-CCND1-SMAD4-MYC). A total of 300 nuclei were analyzed for each single cell-derived line, and 500 nuclei for the parental lines.

In subsequent analysis, all counts greater than 10 for a FISH probe were treated as if they were exactly 10 for two reasons. First, counting more than 10 copies of a gene is potentially inaccurate. Second, high level copy number gains (amplifications) of a gene may represent a different biological process than simple copy number variation, and thus inferences of unobserved intermediates between a copy number of two and a large copy number are suspect. The censoring was applied to the signal counts for MYC. We established the baseline for the accuracy of the miFISH approach by hybridizing the probe sets to eight cultures of karyotypically normal cells (immortalized normal colon epithelial cells and foreskin keratinocytes). On average, 92.8% of the cells showed two copy numbers for all probes, 3.4% were tetraploid and 3.8 presented with an aberrant pattern.

Instability scores and instability indices are calculated by dividing the number of miFISH patterns multiplied by 100 by the number of cells counted. Instability score was used during the calculation for the six CRC cell lines using two probes at a time (EGFR, CCND1, TERC and CDX2) on 5000 cells. Instability index was used for the miFISH experiments using 12 probes on the same nuclei with 300–500 cells counted in total. Due to the differences in probe and cell numbers counted, instability score should not be compared to instability index, thus the difference in names.

Gene expression profiling

Total RNA was extracted using the RNeasy mini kit (Qiagen, Hilden, Germany). RNA quality and quantity was assessed using both the Bioanalyzer (Agilent, Santa Clara, CA) and Nanodrop (Nanodrop, Wilmington, DE) instruments. Gene expression measurements using the nCounter PanCancer Pathways Panel were performed using 100 ng RNA on the NanoString GEN2 nCounter Analysis System (NanoString Technologies, Seattle, WA). All steps were performed using protocols provided by NanoString. Data from all runs were imported into nSolver version 2.0 (NanoString Technologies), matched to corresponding Reporter Library File (RLF), and subjected to quality control using the software quality metrics with default cutoff criteria. Data were submitted to GEO with the accession number, GSE102647.

Sequence analysis of BRAF, KRAS and NRAS

DNA was extracted using the DNeasy kit (Qiagen) and quantified using Nanodrop. Sequence analysis was done based on pyrosequencing technology. Target regions covered KRAS codons 12, 13, 59, 61, 117 and 146, NRAS codons 12, 13, 59, 61, 117 and 146, and BRAF codon 600. The target regions were separately amplified by PCR (Supplementary Table 1, available at Carcinogenesis Online). The amplicons were then immobilized on Streptavidin Sepharose High Performance beads (GE Healthcare Europe, Freiburg, Germany) and single-stranded DNA was prepared. Sequencing and analysis was done on a PyroMark Q24 system (Qiagen) following the manufacturer's instructions.

Growth curves

Growth curves for the single cell-derived clones and the parental line were established by counting adherent and suspension cells in three

independent experiments every 24 h for 7 days after seeding 20000 cells per well in a six-well plate. The final cell count was established after harvesting the cells by trypsinization in a Neubauer hemacytometer (Superior Marienfeld, Lauda-Königshofen, Germany). The total cell number was determined by averaging the cell count in eight 1 mm² squares. Cell numbers were normalized to the cell number initially counted 24 h after seeding and plotted on a log₁₀ scale.

The subclone combinations were performed by taking an equal number of cells from each clone and mixing them together. We then cultured the mixture for 4 days before seeding the cells in three independent growth curve experiments for 7 days. To obtain 'conditioned' medium, we plated the SW480 parental line with the same number of cells, allowed it to grow for 4 days, and then took the media from the flask.

Tumorigenicity assay

ATHYMIC nude mice were purchased from Charles River Laboratories and bred at NCI's laboratory animal facility. At 7 weeks, each mouse received four subcutaneous injections. SW480 parental cells, an A clone (2G6) and a B clone (1A5) were injected on the left shoulder, the right shoulder and the right flanks, respectively. One hundred thousand cells in Matrigel (Corning, NY) were used per injection site. Negative control samples of Matrigel without cells were injected on the left flanks. Tumor volume (*V*) was measured using the equation $V = (\text{length}) \times (\text{width})^2/2$. Mice were euthanized when the largest subcutaneous tumor reached 750 mm³. The experiments were approved by the Animal Experiments Committee of NCI/NIH (protocol #MB-045-A).

Phylogenetic tree inference

Tree models of tumor progression were computed using the software FISHTrees 3.1 (23) in the weighted, ploidyless mode (24), which models gains and losses of single genes, gains and losses of single chromosomes, and genome doubling as distinct events with different probabilities. In these data, there were two probes, *HER2* and *TP53*, on the same chromosome (chromosome 17); in FISHTrees, a simultaneous gain of one copy or simultaneous loss of one copy of both probes is treated as gain or loss of chromosome 17, respectively. Normally, FISHTrees is run on tumor samples that contain a population of diploid cells and are presumed to have recently evolved from a diploid ancestor. The assumption of a recent diploid ancestor is false for these immortalized cell lines, so FISHTrees was modified to run in with the most abundant clone as the root of the tree.

The counting of gain and loss events inferred by FISHTrees is described more in detail in the [Supplementary Materials and Methods](#).

RESULTS

CRC cell lines exhibit different patterns of chromosomal instability

To decipher patterns of chromosomal heterogeneity, we analyzed the widely used CRC cell lines DLD-1, HCT116, H508, SW620, HT-29 and SW480 using miFISH. For each cell line, we analyzed copy numbers for *EGFR*, *CCND1*, *TERC* and *CDX2* simultaneously on metaphase chromosome preparations and two probes at a time on 5000 interphase nuclei (Figure 1 and Supplementary Figure 2, available at [Carcinogenesis Online](#)). Different sets of cells were used for each probe pair on the interphase nuclei. The analysis of the mismatch-repair-deficient cell lines DLD-1 and HCT116 showed stable diploid karyotypes with 70–90% of the cells having two signals (instability scores 0.665 and 0.531, respectively) (Figure 1). The mismatch-repair-proficient aneuploid cell lines H508, SW620, HT-29 and SW480 showed markedly higher instability (instability scores 3.034, 0.972, 1.663, 0.953, respectively). HT-29 and SW480 cell lines displayed aneuploid karyotypes, which is consistent with the genomic imbalance profiles generated by array comparative genomic hybridization (aCGH) and SKY, as shown in Supplementary Figure 3, available at [Carcinogenesis Online](#) (25).

We assessed the clonal evolution from a single cell in two CRC cell lines, SW480 and HT-29, which have both aneuploid karyotypes but different degrees of chromosomal instability (25). For this purpose, we used fluorescence-activated cell sorting to establish 11 single cell-derived daughter cell lines and propagated them (Supplementary Figure 4, available at [Carcinogenesis Online](#)) using a process summarized in Supplementary Figure 1, available at [Carcinogenesis Online](#), in order to assess their clonal composition and development using miFISH and phylogenetic tree modeling.

Clonal compositions of single cell-derived clones are revealed by multiplex FISH

The miFISH analysis was based on our novel, automated, high-throughput approach that allows enumeration of copy numbers of 12 gene-specific loci in each nucleus (10). This entails the sequential hybridization of three FISH panels comprising four differentially labeled fluorescent probes each, followed by image relocation for 300 individual cells (Figure 2A). We designed three multiplex CRC FISH probe panels targeting *COX2* (1q), *TERC* (3q), *APC* (5q), *EGFR* (7p), *MYC* (8q), *CCND1* (11q), *CDX2* (13q), *CDH1* (16q), *TP53* (17p), *HER2* (17q), *SMAD4* (18q) and *ZNF217* (20q). The results of this comprehensive analysis compared the clonal composition of the parental cell lines with the single cell-derived daughter cell lines grown up to a 25 cm² growth area (Tables 1 and 2).

The composition of the parental HT-29 cell line consisted of three major clones, which we denote by A, B and C. In this context, a 'clone' means a population of cells in which all counts of the miFISH probes are identical. Clone A was present in the majority (66%) of the population (Table 1). The less frequent clones B (8.4%) and C (5.4%) differed from clone A by having lower copy numbers for the two oncogenes *EGFR* or *CCND1*, which could explain why these clones are less abundant. Clone A was the most frequent clone in eight of the 11 single cell-derived cell lines, while clone B dominated the populations of the other three single cell-derived lines. The observation that a larger fraction of the single cell-derived daughter cell lines were dominated by clone A likely reflects the higher frequency of clone A in the parental cell line. The daughter cell lines were relatively stable except for the copy numbers of *EGFR* and *CCND1*, regardless of whether their populations were dominated by clones A or B. This is reflected by the low chromosomal instability indices (see Materials and Methods), indicating either a low basal rate of copy number change or a continuous selection for a genomic aberration profile that defines the parental HT-29 line. The instability index values (Table 1) of the daughters ranged from 3.33 to 15.95 (average 8.76) in the single cell-derived lines, while the parental line had an instability index of 11.40.

EGFR is the gene that shows the greatest copy number variability. *EGFR* was at its highest common copy number, 4, in fewer than 71% of the cells observed in the HT-29 daughter cell lines, whereas all the other genes had more than 95% of cells at their highest common copy number. Interestingly, we observed a small population of the B clone in many of the daughters dominated by clone A, indicating a recapitulation of the parental cell line composition. HT-29 cells are known to have a *BRAF* mutation. We confirmed the presence of this *BRAF* mutation in both the parental cell line and single cell-derived daughter cell lines. Since the *BRAF* mutation activates the *EGFR* signaling pathway, one could speculate that additional *EGFR* copies may not result in further growth advantage which might explain the variability for *EGFR* copy numbers in the HT-29 cell population.

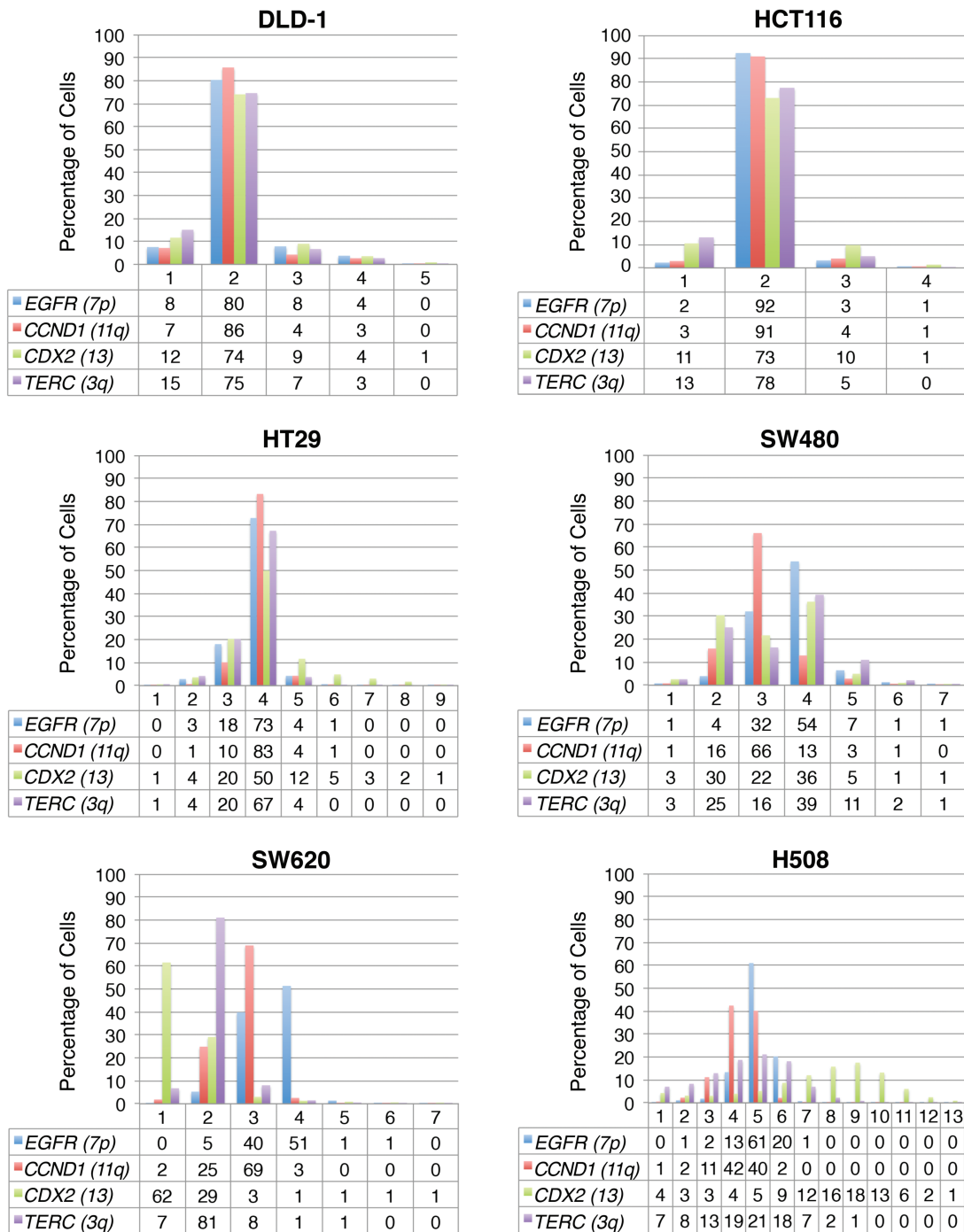


Figure 1. Cytogenetic analysis by interphase FISH of the colorectal cancer cell lines DLD-1, HCT116, H508, SW620, HT-29 and SW480. Note the increased chromosomal instability in aneuploid cell lines. The results were based on the enumeration of 5000 interphase nuclei. The color scheme for the different probes is indicated. The numbers below the columns indicate the copy numbers. Y-axis, percentage of cells with a given count.

Compared to HT-29, the SW480 parental and single cell-derived daughter cell lines showed much greater genetic diversity (Table 2). The parental cell line revealed two major and two minor clones, which we denote by A, B, C and D. The most frequent clone A was present in 42.2% of the cells, clone B occurred in 15.4%, while clones C and D comprised 4.2 and 4%, respectively (Figure 2B). The four lettered clones displayed distinct copy number patterns and differed with respect to their overall ploidy: clone A showed a hyperdiploid pattern (Table 2), clone B was near-triploid (Table 2), while clones C and D had a

hypertetraploid baseline. The most common clone in the parental cell line, the hyperdiploid clone A, is the dominant clone in four of the 11 single cell-derived daughter cell lines. The second most common clone in the parental population, the near-triploid clone B, became the most frequent clone in six daughter cell lines. The hypertetraploid clone C was the most frequent in one of the daughter cell lines (2H7) (Table 2).

Compared to HT-29, the single cell-derived daughter cell lines for SW480 showed on average (12.64) a higher instability index (Table 2), ranging from 4.33 to 37.0. Cell line 2H7, which

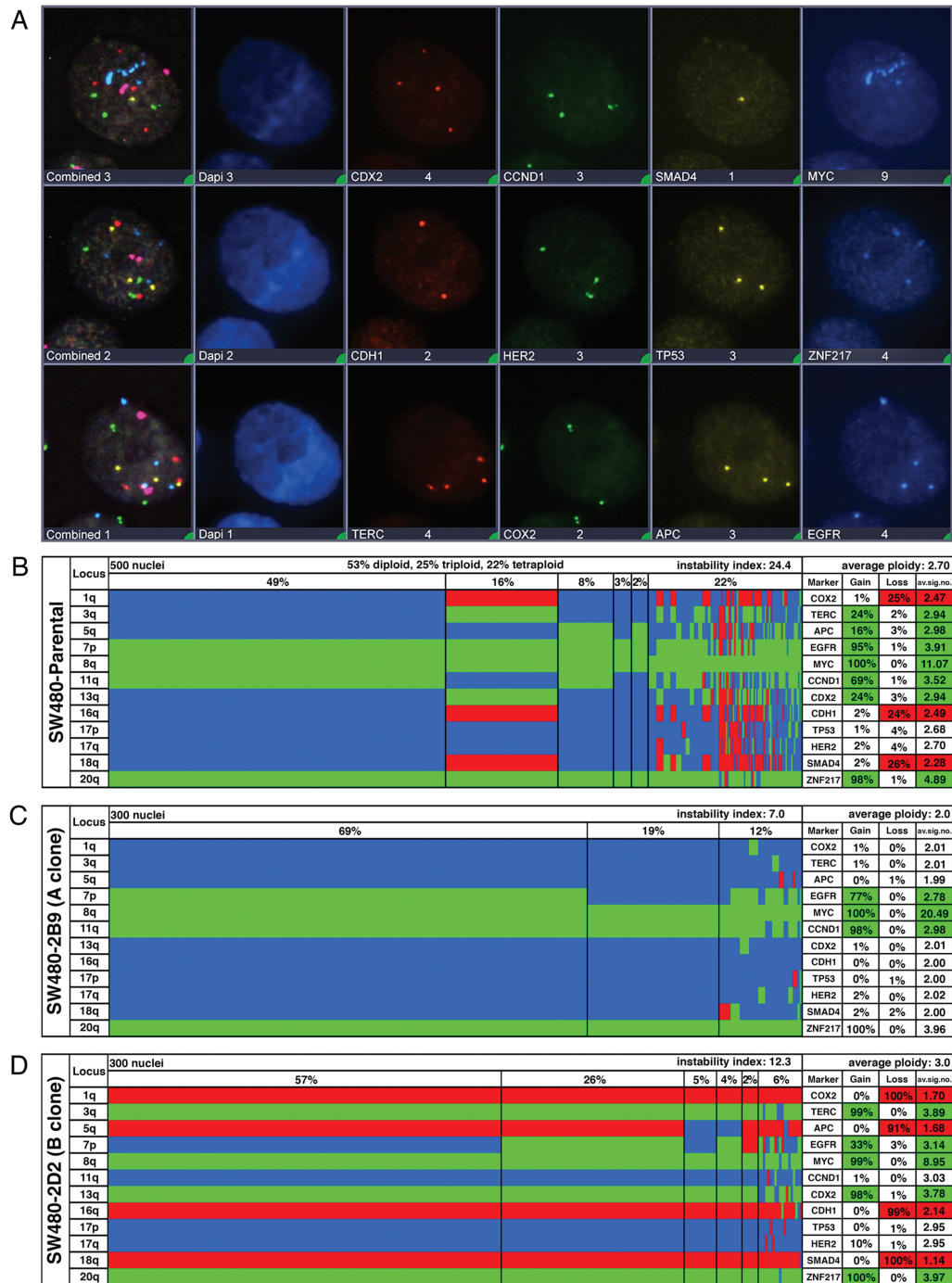


Figure 2. miFISH analysis with 12 gene-specific probes. (A) Composite image of all 12 individual probes and combined images for each panel. (B) Summary of clonal imbalance according to miFISH for the SW480 parental, (C) clone A and (D) clone B cell lines. The color scheme is as follows: green, gains; red, losses; blue, unchanged. The 'Locus' column depicts the specific chromosome arm for each probe. Each vertical line discerns specific signal patterns in the clones and how prevalent they are in the population.

was dominated by the tetraploid parental clone C, was the most unstable with an instability index of 37.0, which was substantially higher than that of the parental cell line (24.4). For the other single cell-derived daughter cell lines, the instability index was lower than in the parental population ranging from 4.33 to 15.33. Interestingly, one of the single cell-derived daughter cell lines (2C8) dominated by clone A duplicated its genome during propagation, resulting in a minor clone D population, similar to what was observed in the parental cell line.

While the single cell-derived daughter cell lines of HT-29 showed signal patterns identical to the parental cell clones, 6 of 11 daughter cell lines in SW480 displayed a major clone that had similar, but not identical gene copy numbers to that seen in the major clones of the parental line, reflecting the higher instability observed in SW480. A gain of 20q is common in CRC, but evidently was not required for daughter cell line 2A9 to survive and proliferate since 2A9 did not carry extra copies as seen in the parental and all other daughter cell lines. While all

Table 1. FISH probe copy number counts and major clones in the HT-29 single cell-derived cell lines

| Cell lines | Probes | | COX2 | TERC | APC | EGFR | MYC | CCND1 | CDX2 | CDH1 | TP53 | HER2 | SMAD4 | ZNF217 | Cells within parental population (%) | Comparison to parental clone | Cells within population (%) | Instability index |
|---------------|------------|--------|------|------|-----|------|-----|-------|------|------|------|------|-------|--------|--------------------------------------|------------------------------|-----------------------------|-------------------|
| | Chromosome | Ploidy | | | | | | | | | | | | | | | | |
| HT29 parental | A | 3.36 | 3 | 4 | 3 | 4 | 10 | 4 | 4 | 3 | 2 | 3 | 2 | 5 | 66.40 | | 66.40 | 11.40 |
| | B | 3.27 | 3 | 4 | 3 | 3 | 10 | 4 | 4 | 3 | 2 | 3 | 2 | 5 | 8.40 | | 8.40 | |
| | C | 3.27 | 3 | 4 | 3 | 4 | 10 | 3 | 4 | 3 | 2 | 3 | 2 | 5 | 5.4 | | 5.4 | |
| | 1 | 3.36 | 3 | 4 | 3 | 4 | 10 | 4 | 4 | 3 | 2 | 3 | 2 | 5 | 69.33 | Same as A | 69.33 | 12.67 |
| HT29 1A4 | 2 | 3.27 | 3 | 4 | 3 | 4 | 10 | 4 | 3 | 3 | 2 | 3 | 2 | 5 | 0.40 | Similar to A | 0.40 | |
| | 3 | 3.27 | 3 | 4 | 3 | 3 | 10 | 4 | 4 | 3 | 2 | 3 | 2 | 5 | 8.40 | Same as B | 8.40 | |
| | 4 | 3.27 | 3 | 4 | 3 | 4 | 10 | 3 | 4 | 3 | 2 | 3 | 2 | 5 | 0.67 | Same as C | 0.67 | |
| | 5 | 3.36 | 3 | 4 | 3 | 4 | 10 | 4 | 4 | 3 | 2 | 3 | 2 | 5 | 69.67 | Same as A | 69.67 | 14.67 |
| HT29 2B1 | 1 | 3.36 | 3 | 4 | 3 | 4 | 10 | 4 | 4 | 3 | 2 | 3 | 2 | 5 | 1.67 | Same as B | 1.67 | |
| | 2 | 3.27 | 3 | 4 | 3 | 3 | 10 | 4 | 4 | 3 | 2 | 3 | 2 | 5 | 1.00 | Same as C | 1.00 | |
| | 3 | 3.27 | 3 | 4 | 3 | 4 | 10 | 3 | 4 | 3 | 2 | 3 | 2 | 5 | 71.43 | Same as A | 71.43 | 9.63 |
| | 4 | 3.27 | 3 | 4 | 3 | 3 | 10 | 4 | 4 | 3 | 2 | 3 | 2 | 5 | 0.66 | Same as B | 0.66 | |
| HT29 2C6 | 1 | 3.36 | 3 | 4 | 3 | 4 | 10 | 4 | 4 | 3 | 2 | 3 | 2 | 5 | 74.00 | Same as A | 74.00 | 8.00 |
| | 2 | 3.36 | 3 | 4 | 3 | 4 | 10 | 4 | 4 | 3 | 2 | 3 | 2 | 5 | 5.00 | Similar to A | 5.00 | |
| | 3 | 3.27 | 3 | 4 | 3 | 4 | 10 | 4 | 4 | 3 | 2 | 3 | 2 | 5 | 3.00 | Same as B | 3.00 | |
| | 4 | 3.27 | 3 | 4 | 3 | 4 | 10 | 3 | 4 | 3 | 2 | 3 | 2 | 5 | 2.67 | Same as C | 2.67 | |
| HT29 2D3 | 1 | 3.27 | 3 | 4 | 3 | 3 | 10 | 4 | 4 | 3 | 2 | 3 | 2 | 5 | 67.77 | Same as B | 67.77 | 15.95 |
| | 2 | 3.36 | 3 | 4 | 3 | 4 | 10 | 4 | 4 | 3 | 2 | 3 | 2 | 5 | 1.33 | Same as A | 1.33 | |
| | 3 | 3.36 | 3 | 4 | 3 | 4 | 10 | 4 | 4 | 3 | 2 | 3 | 2 | 5 | 88.08 | Same as A | 88.08 | 4.30 |
| | 4 | 3.45 | 3 | 4 | 3 | 4 | 10 | 4 | 4 | 3 | 3 | 3 | 2 | 5 | 5.30 | Similar to A | 5.30 | |
| HT29 2D8 | 1 | 3.36 | 3 | 4 | 3 | 4 | 10 | 3 | 4 | 3 | 2 | 3 | 2 | 5 | 1.99 | Same as C | 1.99 | |
| | 2 | 3.27 | 3 | 4 | 3 | 4 | 10 | 4 | 4 | 3 | 2 | 3 | 2 | 5 | 81.19 | Same as A | 81.19 | 8.91 |
| | 3 | 3.27 | 3 | 4 | 3 | 4 | 10 | 3 | 4 | 3 | 2 | 3 | 2 | 5 | 1.32 | Same as B | 1.32 | |
| | 4 | 3.27 | 3 | 4 | 3 | 4 | 10 | 4 | 4 | 3 | 2 | 3 | 2 | 5 | 0.66 | Same as C | 0.66 | |
| HT29 2E1 | 1 | 3.36 | 3 | 4 | 3 | 4 | 10 | 4 | 4 | 3 | 2 | 3 | 2 | 5 | 90.33 | Same as A | 90.33 | 6.67 |
| | 2 | 3.27 | 3 | 4 | 3 | 4 | 10 | 4 | 4 | 3 | 2 | 3 | 2 | 5 | 8.40 | Same as A | 8.40 | |
| | 3 | 3.27 | 3 | 4 | 3 | 3 | 10 | 4 | 4 | 3 | 2 | 3 | 2 | 5 | 0.67 | Same as B | 0.67 | |
| | 4 | 3.27 | 3 | 4 | 3 | 4 | 10 | 4 | 4 | 3 | 2 | 3 | 2 | 5 | 91.00 | Same as B | 91.00 | 3.33 |
| HT29 2E2 | 1 | 3.36 | 3 | 4 | 3 | 4 | 10 | 4 | 4 | 3 | 2 | 3 | 2 | 5 | 90.00 | Same as A | 90.00 | 4.67 |
| | 2 | 3.27 | 3 | 4 | 3 | 3 | 10 | 4 | 4 | 3 | 2 | 3 | 2 | 5 | 4.67 | Same as B | 4.67 | |
| | 3 | 3.27 | 3 | 4 | 3 | 4 | 10 | 3 | 4 | 3 | 2 | 3 | 2 | 5 | 0.33 | Same as C | 0.33 | |
| | 4 | 3.36 | 3 | 4 | 3 | 4 | 10 | 4 | 4 | 3 | 2 | 3 | 2 | 5 | 91.06 | Same as A | 91.06 | 4.97 |
| HT29 2E4 | 1 | 3.36 | 3 | 4 | 3 | 4 | 10 | 4 | 4 | 3 | 2 | 3 | 2 | 5 | 3.97 | Same as B | 3.97 | |
| | 2 | 3.27 | 3 | 4 | 3 | 4 | 10 | 4 | 4 | 3 | 2 | 3 | 2 | 5 | 8.40 | Same as B | 8.40 | |
| | 3 | 3.27 | 3 | 4 | 3 | 4 | 10 | 4 | 4 | 3 | 2 | 3 | 2 | 5 | 66.40 | Same as A | 66.40 | |
| | 4 | 3.27 | 3 | 4 | 3 | 4 | 10 | 3 | 4 | 3 | 2 | 3 | 2 | 5 | 8.40 | Same as B | 8.40 | |
| HT29 2G1 | 1 | 3.36 | 3 | 4 | 3 | 4 | 10 | 4 | 4 | 3 | 2 | 3 | 2 | 5 | 66.40 | Same as A | 66.40 | |
| | 2 | 3.27 | 3 | 4 | 3 | 3 | 10 | 4 | 4 | 3 | 2 | 3 | 2 | 5 | 4.67 | Same as B | 4.67 | |
| | 3 | 3.27 | 3 | 4 | 3 | 4 | 10 | 3 | 4 | 3 | 2 | 3 | 2 | 5 | 0.33 | Same as C | 0.33 | |
| | 4 | 3.36 | 3 | 4 | 3 | 4 | 10 | 4 | 4 | 3 | 2 | 3 | 2 | 5 | 91.06 | Same as A | 91.06 | |
| HT29 2G5 | 1 | 3.36 | 3 | 4 | 3 | 4 | 10 | 4 | 4 | 3 | 2 | 3 | 2 | 5 | 8.40 | Same as B | 8.40 | |
| | 2 | 3.27 | 3 | 4 | 3 | 4 | 10 | 4 | 4 | 3 | 2 | 3 | 2 | 5 | 66.40 | Same as A | 66.40 | |
| | 3 | 3.27 | 3 | 4 | 3 | 4 | 10 | 4 | 4 | 3 | 2 | 3 | 2 | 5 | 8.40 | Same as B | 8.40 | |
| | 4 | 3.27 | 3 | 4 | 3 | 4 | 10 | 4 | 4 | 3 | 2 | 3 | 2 | 5 | 66.40 | Same as A | 66.40 | |

Major parental clone in blue, second major parental clone in pink, third major clone in green. Clones in white are infrequent in parental. Signal numbers in red differ from major clones observed in parental cell line. Signal numbers highlighted in yellow differ from major clone in each cell line. Parental threshold >4%. Single cell clones cell line threshold >5%.

daughters dominated by clones A or C were copy number neutral for SMAD4, the clone B dominated daughters showed a consistent loss of this tumor suppressor. Two daughter cell lines, 2G4 and 2G6, harbored only two copies of EGFR unlike the parental A clone that had three copies. Like HT-29, SW480 carries a mutation (in this case in the gene KRAS) that activates the EGFR pathway, possibly reducing the typical growth advantage of cells with extra copies of EGFR. The parental KRAS mutation was present in all SW480 daughter cell lines.

Single cell-derived clones showed profound differences in phenotypes and gene expression

The morphologies of the single cell-derived daughter cell lines of SW480 were strikingly different (Figure 3A). The single cell-derived lines dominated by clone A revealed a growth pattern reminiscent of spheroids, while the daughters dominated by clone B grew adherently, suggesting that there is a genetic basis for the different growth patterns. Both growth patterns were observed in the parental cultures. Daughters of HT-29, in contrast, did not exhibit differences in morphology depending on the dominant clone (A or B).

To assess to what extent the genomic aberration patterns observed by the miFISH analysis of the single cell-derived daughter cell lines correlated with gene expression profiles, we used the NanoString nCounter PanCancer Pathways platform, on which we measured the expression values of 770 cancer-related genes in the parental cell lines and all daughter cell lines. The results are presented as an unsupervised cluster analysis in Figure 3B and C. The gene expression analyses did not distinguish HT-29 single cell-derived daughter cell lines, but separated SW480 daughter lines in two distinct clusters. The gene expression analysis of the SW480 daughter cell lines followed the separation between hyperdiploid and near-triploid lines (Figure 3C). This indicates that the genetic differences between clones A and B, observable by miFISH copy-number patterns, are reflected in distinct gene expression profiles. These differences might have implications when selecting the cell lines for functional analyses. In fact, we previously showed that the clone A had higher Notch signaling and was depleted from the culture when inhibiting the stem cell marker LGR5 using RNA interference (26). Of note, only copy numbers of SMAD4 correlated significantly with gene expression, while no correlation was observed for any of the other genes (Supplementary Tables 4 and 5, available at *Carcinogenesis* Online).

Growth rates and tumorigenicity of SW480 daughter cell lines

Since we observed profound differences between clones A and B in SW480 by miFISH, gene expression and morphology, we determined the growth characteristics of the parental SW480 population and of five single cell-derived daughter cell lines that were, based on the signal patterns of their major clones, most likely derived from clones A (2C8, 2G6), B (1A5, 2F11) and C (2H7). All five tested daughter cell lines revealed similar proliferation rates (doubling time ~ 31 h), whereas the parental population proliferated profoundly faster (doubling time ~20 h) (Figure 3D). This was consistent with the potential of the respective cell lines to establish tumors following injection into nude mice. The parental cell line revealed increased tumorigenicity. Parental SW480 cells formed tumors in all mice, while clones 2G6 and 1A5 did not form tumors before the mice had to be sacrificed since the tumors from the SW480 parental line reached a critical size of 750 mm³.

To test whether the presence of both clones A and B in the parental cell line could explain faster growth, we recombined two lines of each clone in equal parts. The mixed population of clones 2C8, 2G6, 1A5 and 2F11 was cultured for five passages to allow an adjustment of the proportion of each clone in the population. However, combining those single cell-derived clones resulted in a similar growth pattern as seen for the single clones and did not restore the faster proliferation of the parental population (Figure 3D). To understand whether including more clones would result in a faster proliferation, we mixed all 11 subclones. While this mixture grew faster than the single cell-derived cell clones, it still did not grow as fast as the parental line. Lastly, we investigated whether the secretion of an important growth factor by a subset of cells in the parental line could explain the faster growth. We therefore cultured the single cell-derived clones and the mixture of the 11 clones with medium collected from the culture of the parental line. However, the use of such 'conditioned' medium did not result in faster growth of the single cell-derived lines, nor did it result in faster growth of the combined single cell-derived clones (Figure 3E).

Clonal development in SW480 daughter cell lines over time

To follow the clonal evolution of the SW480 single cell-derived lines, we harvested cells for miFISH from two specific points (early and late) during culture: after ~15 population doublings (1.4×10^4 cells), and 22 population doublings (2.5×10^6 cells) (Supplementary Figure 1, available at *Carcinogenesis* Online). Looking in detail at the parental B clone-derived cell lines (Supplementary Table 2, available at *Carcinogenesis* Online), each cell line maintained a similar instability index between the two time points, meaning that the more stable clones, SW480-F11, SW480-2B4 and SW480-2D4 (Supplementary Table 2, available at *Carcinogenesis* Online) stayed with one exception below an index of 9 (range 4.3–11), while the more unstable clones SW480-2D2, SW480-2A9 and SW480-1A5 (Supplementary Table 2, available at *Carcinogenesis* Online) had indices higher than 9 (range 9.7–19.3) for both time points, indicating that the instability level might be intrinsic to the cell from which they are derived.

The phylogenetic consensus miFISHtree (Supplementary Figure 5A, available at *Carcinogenesis* Online) of clone SW480-2F11 compared two time points and is an example of a stable cell line. The major clone observed was identical to the parental B clone and was present at both time points (92 versus 91%). Similar percentages of different clones were found when comparing early (96 well; blue circle) versus late (T25 plate; red circle) time points in SW480-2F11, indicating similar levels of heterogeneity between both measurements (Supplementary Figure 5A, Supplementary Table 2, available at *Carcinogenesis* Online). The other two stable cell lines, SW480-2B4 and SW480-2D4 behaved similarly. Interestingly, two of the stable cell lines featured the identical parental B clone as the major clone, while the third one, SW480-2D4 had different CDX2 copy number as the only difference. This did not appear to have any negative influence on the stability of the clone.

SW480-2D2 is an example of an unstable daughter cell line, which interestingly was the only cell line that had a major clonal pattern that was different in the early time point compared to the later time point (Supplementary Table 2, available at *Carcinogenesis* Online). At the early time point, this cell line featured the exact parental B clone in 61% of the cell population, but also had a minor clone (19%) with an EGFR copy number change, suggestive of increased chromosomal instability

Table 2. FISH probe copy number counts and major clones in the SW480 single cell-derived cell lines

| Probes | COX2 1q31.1 | TERC 3q26 | APC 5q22 | EGFR 7p11 | MYC 8q24.21 | CCND1 11q13.3 | CDX2 13q12 | CDH1 16q22.1 | TP53 17p13.1 | HER2 17q12 | SMAD4 18q21 | ZNF217 20q13.2 | Cells within parental population (%) | Comparison to parental clone | Instability index |
|-------------------|----------------|--------------|-------------|--------------|----------------|------------------|---------------|-----------------|-----------------|---------------|----------------|-------------------|-----------------------------------------------|------------------------------------|----------------------|
| Chromosome | Clones | | | | | | | | | | | | Ploidy | | |
| SW480 Parental | A | 2 | 2 | 2 | 3 | 10 | 3 | 2 | 2 | 2 | 2 | 4 | 42.20 | | 24.40 |
| | B | 2 | 4 | 3 | 4 | 10 | 3 | 4 | 2 | 3 | 3 | 1 | 15.40 | | |
| | C | 4 | 4 | 6 | 6 | 10 | 5 | 4 | 4 | 4 | 4 | 4 | 4.20 | | |
| | D | 4 | 4 | 4 | 6 | 10 | 6 | 4 | 4 | 4 | 4 | 4 | 4.00 | | |
| SW480 1A5 | 1 | 3 | 4 | 3 | 4 | 10 | 3 | 4 | 3 | 3 | 3 | 1 | 37.33 | Similar to B | 14.00 |
| | 2 | 2 | 4 | 3 | 4 | 10 | 3 | 4 | 3 | 3 | 1 | 4 | 36.00 | Similar to B | |
| | 3 | 2 | 4 | 3 | 4 | 10 | 3 | 4 | 3 | 3 | 1 | 4 | 15.40 | Same as B | |
| SW480 2A9 | 1 | 2 | 4 | 3 | 4 | 10 | 3 | 4 | 2 | 3 | 3 | 1 | 56.33 | Similar to B | 9.67 |
| | 2 | 2 | 4 | 2 | 4 | 10 | 3 | 4 | 2 | 3 | 3 | 1 | 22.33 | Similar to B | |
| | 3 | 2 | 4 | 4 | 4 | 10 | 3 | 4 | 2 | 3 | 3 | 1 | 6.33 | Similar to B | |
| SW480 2B4 | 1 | 2 | 4 | 3 | 4 | 10 | 3 | 4 | 2 | 3 | 3 | 1 | 84.33 | Same as B | 8.00 |
| | 2 | 2 | 4 | 2 | 3 | 10 | 3 | 2 | 2 | 2 | 2 | 4 | 66.00 | Same as A | 7.00 |
| SW480 2B9 | 1 | 2 | 2 | 2 | 3 | 10 | 3 | 2 | 2 | 2 | 2 | 4 | 18.00 | Similar to A | |
| | 2 | 2 | 2 | 2 | 2 | 10 | 3 | 2 | 2 | 2 | 2 | 4 | 68.33 | Same as A | 15.33 |
| SW480 2C8 | 1 | 2 | 2 | 2 | 3 | 10 | 3 | 2 | 2 | 2 | 2 | 4 | 5.00 | Same as D | |
| | 2 | 4 | 4 | 4 | 6 | 10 | 6 | 4 | 4 | 4 | 4 | 8 | 32.00 | Similar to B | 12.33 |
| SW480 2D2 | 1 | 2 | 4 | 1 | 3 | 10 | 3 | 4 | 2 | 3 | 3 | 1 | 20.00 | Similar to B | |
| | 2 | 2 | 4 | 2 | 3 | 10 | 3 | 4 | 2 | 3 | 3 | 1 | 13.00 | Similar to B | |
| | 3 | 2 | 4 | 2 | 4 | 10 | 3 | 4 | 2 | 3 | 3 | 1 | 11.67 | Similar to B | |
| | 4 | 2 | 4 | 1 | 4 | 10 | 3 | 4 | 2 | 3 | 3 | 1 | 3.33 | Same as B | |
| | 5 | 2 | 4 | 3 | 4 | 10 | 3 | 4 | 2 | 3 | 3 | 1 | 81.00 | Similar to B | 6.00 |
| SW480 2D4 | 1 | 2 | 4 | 3 | 4 | 10 | 3 | 3 | 2 | 3 | 3 | 1 | 15.40 | Same as B | |
| | 2 | 2 | 4 | 3 | 4 | 10 | 3 | 4 | 2 | 3 | 3 | 1 | 15.40 | Same as B | |
| SW480 2F11 | 1 | 2 | 4 | 3 | 4 | 10 | 3 | 4 | 2 | 3 | 3 | 1 | 14.60 | Same as B | 4.35 |
| | 2 | 2 | 4 | 3 | 4 | 10 | 3 | 4 | 2 | 3 | 3 | 1 | 93.33 | Similar to A | 4.33 |
| SW480 2G4 | 1 | 2 | 2 | 2 | 2 | 10 | 3 | 2 | 2 | 2 | 2 | 4 | 0.33 | Same as A | |
| | 2 | 2 | 2 | 2 | 3 | 10 | 3 | 2 | 2 | 2 | 2 | 4 | 83.11 | Similar to A | 9.27 |
| SW480 2G6 | 1 | 2 | 2 | 2 | 2 | 10 | 3 | 2 | 2 | 2 | 2 | 4 | 39.67 | Same as C | 37.00 |
| | 2 | 4 | 4 | 6 | 6 | 10 | 5 | 4 | 4 | 4 | 4 | 8 | 6.67 | Similar to C | |
| SW480 2H7 | 1 | 4 | 4 | 4 | 6 | 10 | 5 | 4 | 4 | 4 | 4 | 8 | 0.00 | | |
| | 2 | 4 | 4 | 5 | 6 | 10 | 5 | 4 | 4 | 4 | 4 | 8 | 0.00 | | |

Blue: hyperdiploid clone, pink: near-triploid clone, green: first hypertetraploid clone, grey: second hypertetraploid clone. Red digits: signal numbers that differ from parental clone, yellow highlights: signal numbers that differ between the clones of a single cell clone. Parental threshold >4%; Single cell clones cell line threshold >5%.

(Supplementary Table 2, available at *Carcinogenesis* Online). At the later time point, we observed a switch from the exact parental B clone to clones similar to B, with variable copy numbers in APC and EGFR, resulting in a new major clone with a copy number loss of APC (Supplementary Table 2, available at *Carcinogenesis* Online). The trajectory of this evolution is shown in Supplementary Figure 5B, available at *Carcinogenesis* Online. The two other unstable daughters, SW480-1A5 and SW480-2A9 did not have a major clonal pattern change. In SW480-1A5, the frequency of the major clone, which featured losses of COX2 and CDH1, became less frequent in the later time point (48 versus 37%). The loss of APC in SW480-2A9 did not lead to a shift of major clones between the two time points.

FISHtrees models depict patterns of clonal evolution

To evaluate clonal evolution on a single cell level, we recently developed algorithms to reconstruct evolutionary trajectories of cancer cell populations from single cell FISH data. These algorithms are encoded in the software FISHtrees (27). We compared each SW480 daughter cell line to its parental line to assess the degree to which the tree derived from the daughter follows a similar evolutionary trajectory to a subtree of the parent. The analysis showed a high degree of concordance of tree inferences between single-cell derived clones and the full set of parental and child clones on shared clonal patterns (mean tree reconstruction error 6.2%). However, in many instances we could identify only few shared clonal patterns between parental and daughter clones (mean 37% overlap in clonal patterns), indicating substantial ongoing acquisition of new copy number changes in each single cell-derived cell line over about 25 population doublings. The copy number and edge profiles are consistent with undirected copy number gains and losses. However, the mean imbalance profiles found in the daughter lines remained the ones also observed in the parent line, i.e. chromosomes that are commonly gained are rarely lost, and vice versa (Figure 4A and B). This profile—in which the clonal composition yields a persistent average profile of imbalances despite a high degree of ongoing diversification—is therefore consistent with selective pressure for maintaining these imbalances despite ongoing genomic instability.

FISHtrees shows ongoing instability In SW480

The distribution of copy number counts for the SW480 child cell lines, aggregated separately for oncogenes and tumor suppressors, is shown in Figure 5. In absolute copy numbers, oncogenes usually have a copy number of two and four. Tumor suppressors also have a sizeable number of cells in which the copy number of the gene is two. Interestingly, when cell counts are plotted against copy number relative to the ploidy, the distributions are more strongly peaked, with an obvious skew of oncogenes being gained, and rarely lost. The preference for tumor suppressors is to be at or below ploidy.

The copy number and edge profiles are consistent with undirected copy number gains and losses, with constraints, around a fitness peak. In particular, oncogenes are not inferred to participate in substantially more event changes resulting in a gain, nor are tumor suppressors inferred to participate in more events resulting in a loss (Supplementary Figure 6, available at *Carcinogenesis* Online). The population distribution of copy numbers, however, shows selective pressure on the resulting cells for the population to prefer net gains and losses of specific genes. There is a clear constraint on this process to favor events that do not reduce the copy number below two (Supplementary Figure 6, available at *Carcinogenesis* Online). We suggest this is due to the

majority of copy number changes observed in the daughter cell lines being chromosome missegregation events, rather than focal losses. It is possible that such large-scale losses are incompatible with life for SW480, which has not evolved to have chromosomes with copy number of one.

In summary, the FISHtrees analysis of ITH in single cell-derived clones of HT-29 and SW480 revealed ongoing chromosomal instability without evidence that specific chromosomes are more prone to copy number changes than others. However, the genomic imbalance pattern observed in the parental clones was in general maintained even in single cell clones and is consistent with the copy number changes observed in the majority of primary sporadic CRCs (7).

Discussion

Cell lines established from primary tumors are widely used in studies of cancer genetics and cancer cell biology (28,29). Cell lines provide a unique window into the intrinsic ability of tumor cells to establish genomic and phenotypic heterogeneity relative to a single progenitor cell, a phenomenon also observed in primary tumors.

We analyzed six CRC cell lines, two diploid (DLD-1 and HCT116) and four aneuploid cell lines (HT-29, SW480, SW620 and H508). We chose two aneuploid CRC cell lines (HT-29 and SW480) to explore how genomic heterogeneity is established after single cell cloning. The results show that in the single cell-derived daughter cell lines, mean relative copy number changes of specific genes persist (Figure 4A and B), and that the aggregate loss and gain patterns observed in the parental lines are maintained.

There are conserved patterns of evolution (mechanisms of diversification and selective biases) that are intrinsic to particular cell lines and lead to recurrent patterns of heterogeneity that partly distinguish the progeny of one cell line from another. Our data suggest that this propensity is intrinsic to a tumor cell line. Despite these general propensities, regrowth is a stochastic process, which can be substantially shaped by chance events in one clone versus another in seeding a new population, which in turn shapes the idiosyncrasies of the emerging population.

A key feature of cancers is ITH, i.e. a different genetic make-up in terms of cancer-related mutations and copy number changes, despite the fact that in the tumor population as a whole, genomic imbalances are conserved. To characterize the pattern of ITH and its evolution, we established 11 single cell-derived clones of the CRC cell lines SW480 and HT-29. Substantial heterogeneity appeared within a few generations, as indicated by the Simpson index (Supplementary Table 3, available at *Carcinogenesis* Online) of the gene copy numbers. FISHtrees analysis (24,27) of the phylogeny of each daughter clone suggested that each one developed largely independently, suggesting random segregation errors, consistent with genetic drift. Regardless of this apparent randomness, genes that are commonly gained are rarely lost, and vice versa (Figure 4A and B), leading to mean copy number profiles largely recapitulating those of the parental cell line.

Sampling early and later time points during the growth of the single cell-derived daughter cells revealed the same major clone at both time points for almost all of the daughter cell lines indicating the maintenance of genomic imbalances despite ongoing instability. However, one SW480 daughter cell line evolved from the paternal B clone to a pattern sporting a copy number loss of APC indicating that this change might confer a growth advantage and therefore produced a clone that was able to outcompete the paternal clone.

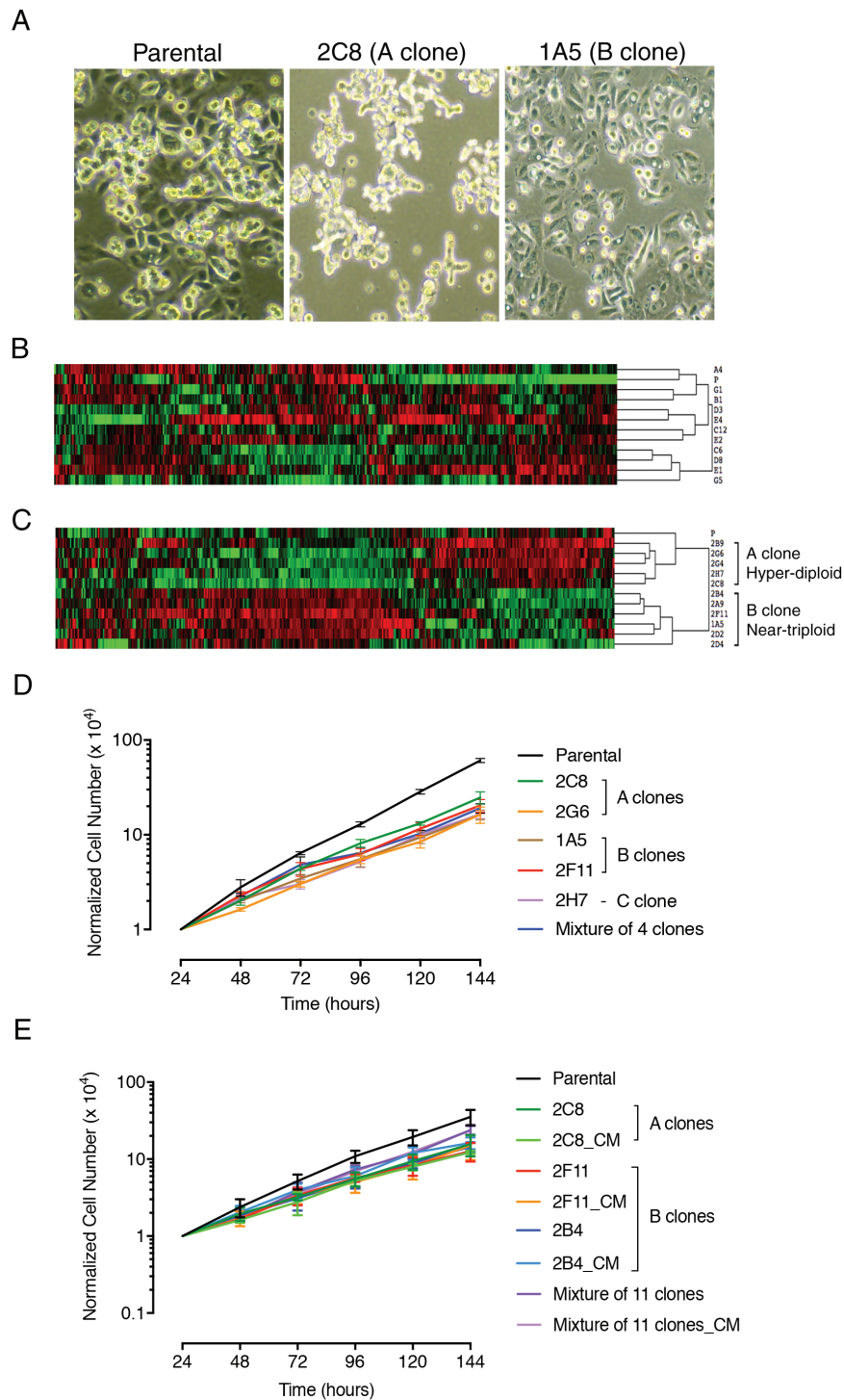


Figure 3. Phenotypic differences (A). Morphological differences in the SW480 parental cell line, single cell-derived A clone (1A5) and single cell-derived B clone (2C8) (B, C). Gene expression profiling using the NanoString technology of parental cell lines and single cell-derived clones presented as an unsupervised hierarchical cluster analysis in HT-29 (B) and SW480 (C). In the panel, the parental cell line is depicted by 'P'. Note that the single cell-derived daughter cell lines derived from parental SW480 show two distinct clusters, which matches the hyperdiploid and near-triploid clones observed by miFISH in the daughter lines. (D) All single cell-derived daughter cell lines proliferated equally fast but slower than the parental line for SW480. A mixture of four single cell-derived clones had the same proliferation rate as the single cell-derived clones. (E) Parental SW480 cells proliferated faster than the A or B clones, with and without conditioned media. A mixture of 11 single cell-derived clone with and without conditioned media grew faster than the single A and B clones however, it still did not proliferate as fast as the parental SW480 cells.

Having observed intra-cell line heterogeneity and seemingly random aberrations, we asked what processes drive the growth of each cell line. One could hypothesize that each cell line optimizes for survival and growth rate. To explore this, we measured the

growth rates of five daughter cell lines derived from single cells of SW480, as well as the parental line and a mixture of four daughter lines. Surprisingly, all daughters and the mixture of four daughter cell lines had similar growth rates, which were lower than the

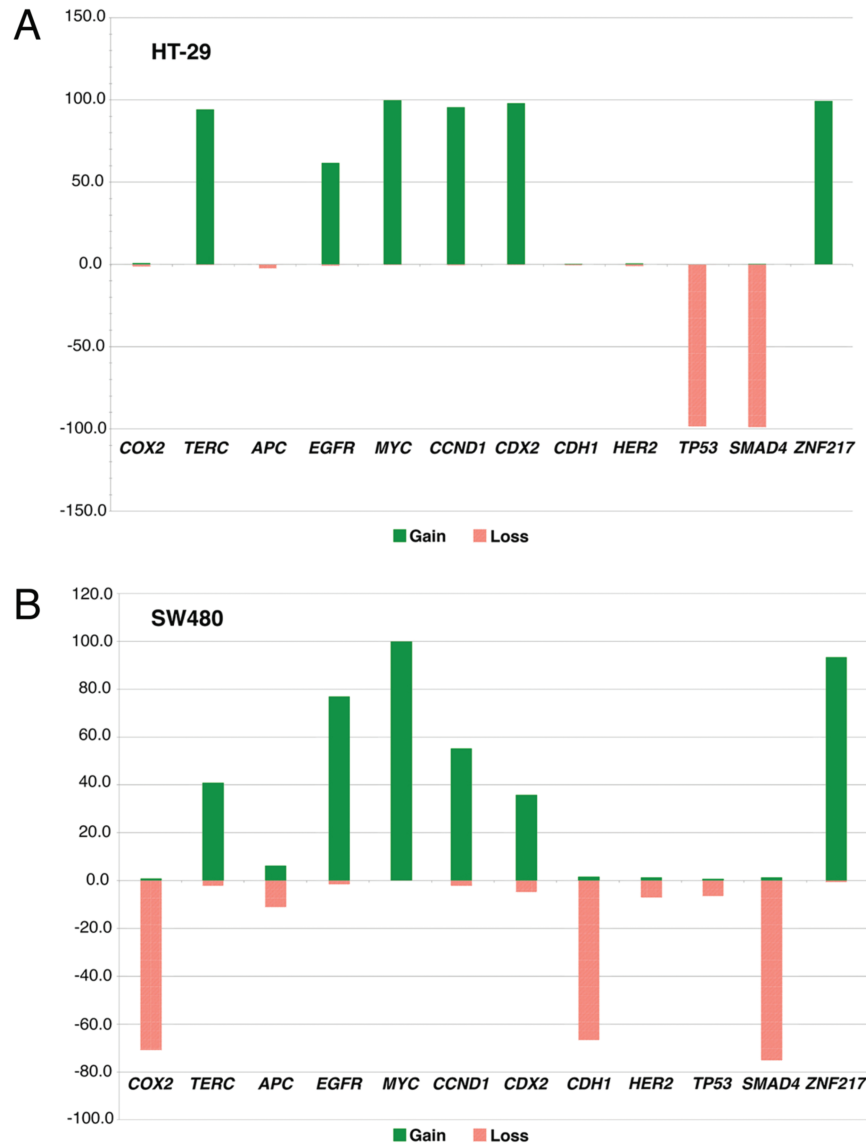


Figure 4. Average gain and loss frequencies for HT-29 (A) and SW480 (B) single cell-derived clones for all gene markers. The percentage of cells with gains and losses are shown above and below the 0% line.

growth rate of the parental SW480 line. Testing the hypothesis that a subset of the cells in the parental cells might secrete an important growth factor, we used ‘conditioned medium’, rendered from the parental cell line, to grow the single cell-derived daughter lines. However, neither the single daughters nor mixtures of them showed improved proliferation when cultured in the ‘conditioned medium’, indicating that no important growth factor was secreted from the parental cells. To further explore if including more single cell clones into a mixture would restore parental cell line growth rates, we mixed all 11 single cell-derived daughters and found that the mixture still grew consistently slower than the parental line. However, the mixture of 11 daughter cell lines grew faster than the individual daughters and the mixture of four daughter cell lines, indicating that including more single cell-derived lines might eventually restore the higher parental growth rate.

The parental line also showed greatly increased tumorigenicity than any of the daughter lines upon injection into nude mice. The phenomenon that cell line clones can have similar growth rates despite differences in ploidy has been described previously (30). It is more typical, however, for different clones

to have different growth rates (31–33). The phenomenon of daughter clones having a lower growth rate than the parental cell line has also been described previously (34). One possible explanation for the lower growth rates in daughter cell lines is that the parental cell line achieved a faster growth rate via cooperation between clones that cannot be achieved in a more homogeneous population. Another intriguing possibility would be to interpret cell line clonal composition based on quasi-species theory, which suggests that the parental cell line could proliferate faster than any single cell-derived clone because any single cells in a small sample are unlikely to be exactly at the fitness peak (35). Of note, the parental line also showed greatly increased tumorigenicity compared to any of the daughter lines upon injection into nude mice.

Finally, we assessed how genomic heterogeneity is reflected in functional heterogeneity in terms of morphology, gene expression and growth. For SW480, miFISH revealed two dominant clones, which were distinct in copy number profiles and gene expression profiles. With the exception of SMAD4, gene expression levels were not correlated with genomic copy

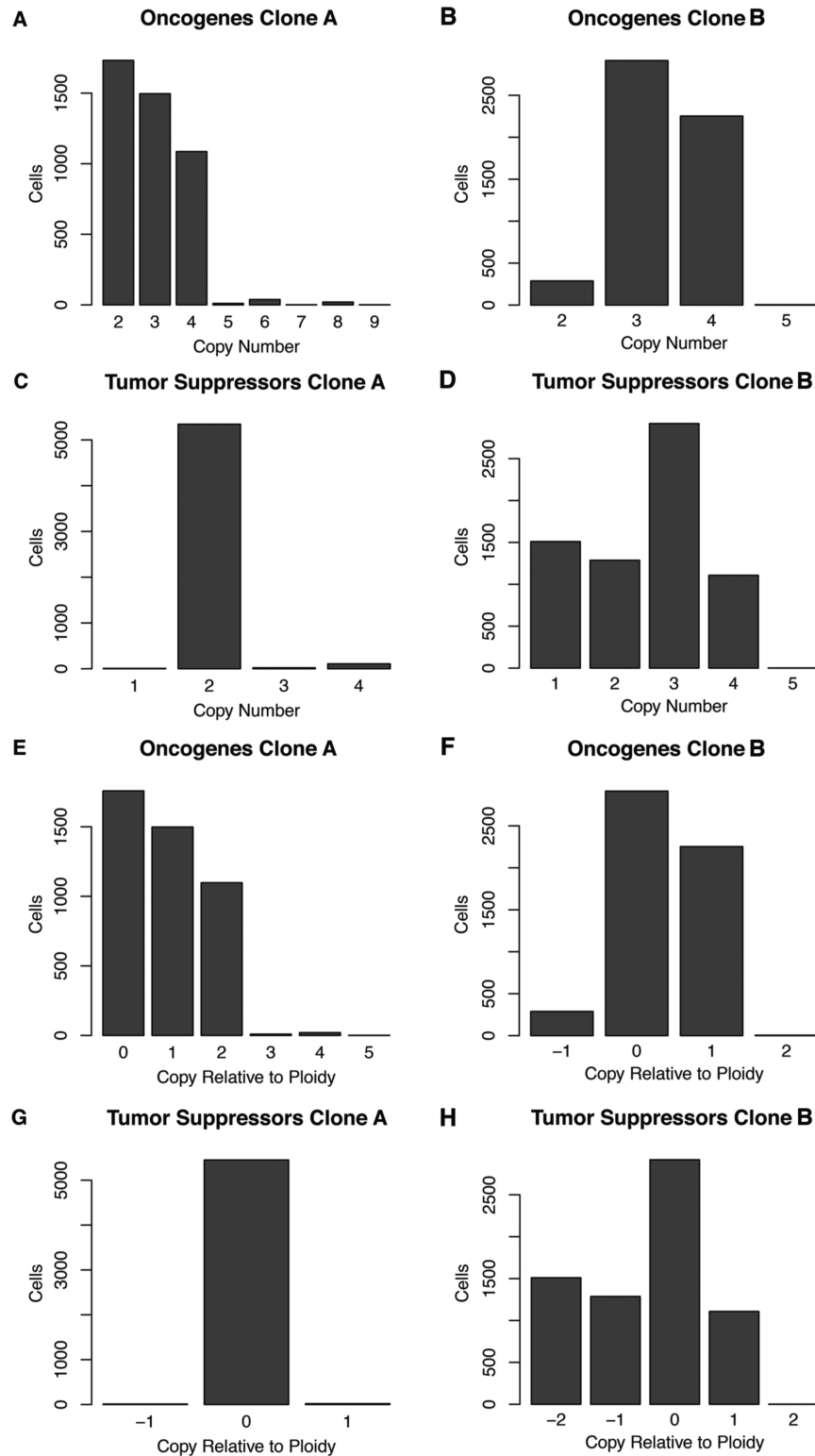


Figure 5. Number of observed cells plotted against copy number counts, aggregated over genes of a given type. (A, B) Cell counts against absolute copy number aggregated over oncogenes in SW480 Clones A and B; (C, D) Cell counts against absolute copy number aggregated over tumor suppressors in SW480 Clones A and B; (E, F) Cell counts against copy number relative to ploidy, aggregated over oncogenes in SW480 Clones A and B; (G, H) Cell counts against copy number relative to ploidy, aggregated over tumor suppressors in SW480 Clones A and B.

number (Supplementary Tables 4 and 5, available at Carcinogenesis Online). In contrast to SW480, the cell lines derived from HT-29 did not show variations between the clones in ploidy, morphology or gene expression.

We comprehensively analyzed the clonal composition of established and widely used CRC cell lines applying advanced molecular cytogenetic techniques (SKY and aCGH), gene expression profiling and miFISH which allows simultaneous

measurements of copy number changes of multiple genes in individual cells, and an algorithm developed to reconstruct the clonal evolution of cancer cell populations (FISHtrees (27)). When new cell lines are grown from single daughter cells they missegregate chromosomes haphazardly and reestablish a heterogeneous cell population. These heterogeneous mixtures nonetheless remain largely populated by clones with copy number configurations similar to those of the parental lines. Karyotype evolution is driven by the necessity to arrive at and maintain a specific plateau of chromosomal copy numbers as the drivers of carcinogenesis.

Supplementary material

Supplementary material can be found at *Carcinogenesis* online.

Funding

This research was supported in part by the Intramural Research Program of the National Institutes of Health, NCI and NLM. R.S. was supported by U.S. National Institutes of Health awards 1R01CA140214, 1R21CA216452 and Pennsylvania Dept. of Health Grant GBMF4554 #4100070287. The Pennsylvania Department of Health specifically disclaims responsibility for any analyses, interpretations or conclusions. F.O.B. is supported by a Rubicon grant #190.2015.2.310.005 of The Netherlands Organization for Scientific Research. R.B. is supported by a Mildred Scheel postdoctoral scholarship of the German Cancer Aid (Deutsche Krebshilfe).

Acknowledgements

Gene expression profiling using the NanoString analysis system was conducted at the CCR Genomics Core at the National Cancer Institute. The authors are grateful to Drs. Reinhard Ebner and Yuri Lazebnik for critical comments on the manuscript, and Buddy Chen for editorial assistance.

Conflict of Interest Statement: None declared.

References

- Ried, T. et al. (1999) Genomic changes defining the genesis, progression, and malignancy potential in solid human tumors: a phenotype/genotype correlation. *Genes Chromosomes Cancer*, 25, 195–204.
- Habermann, J.K. et al. (2009) The gene expression signature of genomic instability in breast cancer is an independent predictor of clinical outcome. *Int. J. Cancer*, 124, 1552–1564.
- Beroukhi, R. et al. (2010) The landscape of somatic copy-number alteration across human cancers. *Nature*, 463, 899–905.
- Heselmeyer, K. et al. (1996) Gain of chromosome 3q defines the transition from severe dysplasia to invasive carcinoma of the uterine cervix. *Proc. Natl. Acad. Sci. USA*, 93, 479–484.
- Heselmeyer, K. et al. (1997) Advanced-stage cervical carcinomas are defined by a recurrent pattern of chromosomal aberrations revealing high genetic instability and a consistent gain of chromosome arm 3q. *Genes Chromosomes Cancer*, 19, 233–240.
- Bardi, G. et al. (1995) Trisomy 7 as the sole cytogenetic aberration in the epithelial component of a colonic adenoma. *Cancer Genet. Cytogenet.*, 82, 82–84.
- Ried, T. et al. (1996) Comparative genomic hybridization reveals a specific pattern of chromosomal gains and losses during the genesis of colorectal tumors. *Genes Chromosomes Cancer*, 15, 234–245.
- Camps, J. et al. (2005) Comprehensive measurement of chromosomal instability in cancer cells: combination of fluorescence in situ hybridization and cytokinesis-block micronucleus assay. *FASEB J.*, 19, 828–830.
- Masramon, L. et al. (2006) Genetic instability and divergence of clonal populations in colon cancer cells *in vitro*. *J. Cell Sci.*, 119, 1477–1482.
- Heselmeyer-Haddad, K. et al. (2012) Single-cell genetic analysis of ductal carcinoma in situ and invasive breast cancer reveals enormous tumor heterogeneity yet conserved genomic imbalances and gain of MYC during progression. *Am. J. Pathol.*, 181, 1807–1822.
- Marusyk, A. et al. (2012) Intra-tumour heterogeneity: a looking glass for cancer? *Nat. Rev. Cancer*, 12, 323–334.
- Duesberg, P. et al. (2011) Is carcinogenesis a form of speciation? *Cell Cycle*, 10, 2100–2114.
- Burrell, R.A. et al. (2013) The causes and consequences of genetic heterogeneity in cancer evolution. *Nature*, 501, 338–345.
- Aparicio, S. et al. (2013) The implications of clonal genome evolution for cancer medicine. *N. Engl. J. Med.*, 368, 842–851.
- McGranahan, N. et al. (2015) Biological and therapeutic impact of intra-tumor heterogeneity in cancer evolution. *Cancer Cell*, 27, 15–26.
- Gerlinger, M. et al. (2012) Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N. Engl. J. Med.*, 366, 883–892.
- Sottoriva, A. et al. (2013) Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. *Proc. Natl. Acad. Sci. USA*, 110, 4009–4014.
- de Bruin, E.C. et al. (2014) Spatial and temporal diversity in genomic instability processes defines lung cancer evolution. *Science*, 346, 251–256.
- Eirew, P. et al. (2015) Dynamics of genomic clones in breast cancer patient xenografts at single-cell resolution. *Nature*, 518, 422–426.
- Hardiman, K.M. et al. (2016) Intra-tumor genetic heterogeneity in rectal cancer. *Lab. Invest.*, 96, 4–15.
- Masters, J.R. (2000) Human cancer cell lines: fact and fantasy. *Nat. Rev. Mol. Cell Biol.*, 1, 233–236.
- Shoemaker, R.H. (2006) The NCI60 human tumour cell line anticancer drug screen. *Nat. Rev. Cancer*, 6, 813–823.
- Chowdhury, S.A. et al. (2013) Phylogenetic analysis of multiprobe fluorescence in situ hybridization data from tumor cell populations. *Bioinformatics*, 29, i189–i198.
- Chowdhury, S.A. et al. (2015) Inferring models of multiscale copy number evolution for single-tumor phylogenetics. *Bioinformatics*, 31, i258–i267.
- International Standing Committee on Human Cytogenomic Nomenclature, et al. (2016) ISCN: An International System for Human Cytogenomic Nomenclature (2016). Karger, Basel, New York.
- Hirsch, D. et al. (2014) Transcriptome profiling of LGR5 positive colorectal cancer cells. *Genom. Data*, 2, 212–215.
- Gertz, E.M. et al. (2016) FISHtrees 3.0: tumor phylogenetics using a ploidy probe. *PLoS One*, 11, e0158569.
- Szakács, G. et al. (2004) Comparing solid tumors with cell lines: implications for identifying drug resistance genes in cancer. *Mol. Interv.*, 4, 323–325.
- Sharma, S.V. et al. (2010) Cell line-based platforms to evaluate the therapeutic efficacy of candidate anticancer agents. *Nat. Rev. Cancer*, 10, 241–253.
- Hastings, R.J. et al. (1983) Cellular heterogeneity in a tissue culture cell line derived from a human bladder carcinoma. *Br. J. Cancer*, 47, 233–244.
- Sweeney, F.L. et al. (1982) Heterogeneity of the growth and metastatic behavior of cloned cell lines derived from a primary rhabdomyosarcoma. *Cancer Res.*, 42, 3776–3782.
- Flatow, U. et al. (1987) Tumorigenicity of T24 urinary bladder carcinoma cell sublines. *Int. J. Cancer*, 40, 240–245.
- Jamasbi, R.J. et al. (1990) Biological heterogeneity and radiation sensitivity of *in vitro* propagated lung metastatic lines originated from a transplantable squamous cell carcinoma of BALB/c mouse. *In Vitro Cell. Dev. Biol.*, 26, 222–228.
- Brown, J.L. et al. (1990) Clonal analysis of a bladder cancer cell line: an experimental model of tumour heterogeneity. *Br. J. Cancer*, 61, 369–376.
- Eigen, M. (1971) Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften*, 58, 465–523.