Edinburgh Research Explorer

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

OXFORD | Carcinogenesis
UNIVERSITY PRESS

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

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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 Wangsa[1], Rüdiger Meyer[1*], Madison Schiefer[1*], E. Michael Gertz[2*], Daniel Bronder[1*,] Isabel Quintanilla[3] Hesed M. Padilla-Nash[1], Irianna Torres[1], Cynthia Hunn[1], Lidia Warner[1], Floryne O. Buishand[1,4], Yue Hu[1], Daniela Hirsch[5], Timo Gaiser[5], Jordi Camps[1,3], Russell Schwartz[6], Alejandro A. Schäffer[2], Kerstin Heselmeyer-Haddad[1#] and Thomas Ried[1#]

[1]Genetics Branch, Center for Cancer Research, National Cancer Institute/National Institutes of Health, Bethesda, MD

[2]Computational Biology Branch, National Center for Biotechnology Information, National Institutes of Health, Bethesda, MD

[3]Unitat de Biologia Cellular i Genètica Mèdica, Departament de Biologia Cellular, Fisiologia i Immunologia, Facultat de Medicina, Universitat Autònoma de Barcelona, Bellaterra, Spain

[4]Department of Clinical Sciences of Companion Animals, Faculty of Veterinary Medicine, Utrecht University, The Netherlands

[5]Institute of Pathology, University Medical Center Mannheim, University of Heidelberg, Germany

[6]Department of Biological Sciences and Computational Biology Department, Carnegie Mellon University, Pittsburgh, PA

1

[*]Contributed equally

Corresponding Author:

Kerstin Heselmeyer-Haddad, Ph.D. and Thomas Ried, M.D.

Genetics Branch, Center for Cancer Research, NCI

50 South Drive, Rm. 1408

Bethesda, MD 20892

riedt@mail.nih.gov

heselmek@mail.nih.gov

2

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 CRC 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.

SUMMARY

We present a meticulous analysis of the heterogeneity in CRC cells based on single cell cloning using our novel miFISH technique allowing simultaneous visualization of numerical aberrations. We show that, despite random chromosomal instability, specific net genomic imbalances are maintained.

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 and colleagues.(12)

ITH has clinical implications, because it may facilitiate 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 (DCIS) and invasive carcinomas (IDC), using fluorescence in situ hybridization (FISH) of FFPE patient samples. Our study revealed a

5

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% CO_2 at 37°C. The

6

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 FACS 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 approximately 70-80% confluency in a T25 flask, we extracted DNA, RNA, and fixed cells for miFISH using cells from the same passage (Supplementary Figure S1).

Multiplex interphase FISH (miFISH)

Detailed experimental procedures are described in Supplementary Materials and Methods. 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), *ZNF217* (20q13.2). FISH probes were chosen based on tumor supressors and oncogenes known to be involved in CRC. Contigs consisted of two to four overlapping clones spanning 333-687kbp 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

7

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,

8

Santa Clara, CA) and Nanodrop (Nanodrop, Wilmington, DE) instruments. Gene expression measurements using the nCounter PanCancer Pathways Panel were performed using 100ng 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 I). 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 hours for seven days after seeding 20,000 cells per well in a 6 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 $1mm^2$ squares. Cell numbers were normalized to the cell number initially counted 24 hours 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 four days before seeding the cells in three independent growth curve experiments for seven days. To obtain "conditioned" media, we plated the SW480 parental line with the same number of cells, allowed it to grow for four 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,

10

respectively. One hundred thousand cells in Matrigel (Corning, NY, USA) 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 = (length) $\times$ (width)^2/2. Mice were euthanized when the largest subcutaneous tumor reached 750 mm^3. 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 in more detail in the Supplementary Materials and Methods.

11

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 S2). 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% to 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 S3.(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 FACS to establish

12

11 single cell-derived daughter cell lines and propagated them (Supplementary Figure S4) using a process summarized in Supplementary Figure S1 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 cm^2 growth area (Tables I and II).

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 I). 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

13

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 clone A or clone 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 I) 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

14

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.

Compared to HT-29, the SW480 parental and single cell-derived daughter cell lines showed much greater genetic diversity (Table II). 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 II), clone B was near-triploid (Table II), 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 eleven 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 II).

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

15

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, six 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 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.

16

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 3C. 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 clone A and clone 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

17

analyses. In fact, we previously showed that the A clone 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 IV and V).

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

hours), whereas the parental population proliferated profoundly faster (doubling

time ~20 hours) (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 mm^3.

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 5 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 approximately 15 population doublings (1.4×10^4 cells), and 22 population doublings (2.5×10^6 cells) (Supplementary Figure S1). Looking in detail at the

19

parental B clone-derived cell lines (Supplementary Table II), each cell line maintained a similar instability index between the two time points, meaning that the more stable clones, SW480-F11, SW480-2B4 and SW-480-2D4 (Supplementary Table II) 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 II) 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) 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% vs. 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 S5A, Supplementary Table II). 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

20

different in the early time point compared to the later time point (Supplementary

Table II). 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

(Supplementary Table II). 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 II). The trajectory of this evolution is shown in

Supplementary Figure S5B. 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% vs 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

21

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 4B). 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,

22

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 S6). 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 S6). 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).

23

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 4B), 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.

24

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 III) 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 (Figures 4A and 4B), 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.

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

25

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 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 number (Supplementary Tables IV and V). In contrast to SW480, the cell lines derived from HT-29 did not show variations between the clones in ploidy, morphology, or gene expresson.

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

27

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.

29

REFERENCES

1. 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.
2. 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**, 1552-64.
3. Beroukhim, R., *et al.* (2010) The landscape of somatic copy-number alteration across human cancers. *Nature*, **463**, 899-905.
4. 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 U S A*, **93**, 479-84.
5. 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**, 233-40.
6. Bardi, G., *et al.* (1995) Trisomy 7 as the sole cytogenetic aberration in the epithelial component of a colonic adenoma. *Cancer Genet Cytogenet*, **82**, 82-4.
7. 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**, 234-45.
8. 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**, 828-30.
9. Masramon, L., *et al.* (2006) Genetic instability and divergence of clonal populations in colon cancer cells in vitro. *J Cell Sci*, **119**, 1477-82.
10. 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**, 1807-22.
11. Marusyk, A., *et al.* (2012) Intra-tumour heterogeneity: a looking glass for cancer? *Nat Rev Cancer*, **12**, 323-34.
12. Duesberg, P., *et al.* (2011) Is carcinogenesis a form of speciation? *Cell Cycle*, **10**, 2100-14.
13. Burrell, R.A., *et al.* (2013) The causes and consequences of genetic heterogeneity in cancer evolution. *Nature*, **501**, 338-45.
14. Aparicio, S., *et al.* (2013) The implications of clonal genome evolution for cancer medicine. *N Engl J Med*, **368**, 842-51.
15. McGranahan, N., *et al.* (2015) Biological and therapeutic impact of intratumor heterogeneity in cancer evolution. *Cancer Cell*, **27**, 15-26.
16. Gerlinger, M., *et al.* (2012) Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med*, **366**, 883-92.

17. Sottoriva, A.*, et al.* (2013) Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. *Proc Natl Acad Sci U S A*, **110**, 4009-14.

18. de Bruin, E.C.*, et al.* (2014) Spatial and temporal diversity in genomic instability processes defines lung cancer evolution. *Science*, **346**, 251-6.

19. Eirew, P.*, et al.* (2015) Dynamics of genomic clones in breast cancer patient xenografts at single-cell resolution. *Nature*, **518**, 422-6.

20. Hardiman, K.M.*, et al.* (2016) Intra-tumor genetic heterogeneity in rectal cancer. *Lab Invest*, **96**, 4-15.

21. Masters, J.R. (2000) Human cancer cell lines: fact and fantasy. *Nat Rev Mol Cell Biol*, **1**, 233-6.

22. Shoemaker, R.H. (2006) The NCI60 human tumour cell line anticancer drug screen. *Nat Rev Cancer*, **6**, 813-23.

23. Chowdhury, S.A.*, et al.* (2013) Phylogenetic analysis of multiprobe fluorescence in situ hybridization data from tumor cell populations. *Bioinformatics*, **29**, i189-98.

24. Chowdhury, S.A.*, et al.* (2015) Inferring models of multiscale copy number evolution for single-tumor phylogenetics. *Bioinformatics*, **31**, i258-67.

25. International Standing Committee on Human Cytogenomic Nomenclature*, et al.* (2016) *ISCN : an international system for human cytogenomic nomenclature (2016)*. Karger, Basel ; New York.

26. Hirsch, D.*, et al.* (2014) Transcriptome profiling of LGR5 positive colorectal cancer cells. *Genom Data*, **2**, 212-5.

27. Gertz, E.M.*, et al.* (2016) FISHtrees 3.0: Tumor Phylogenetics Using a Ploidy Probe. *PLoS One*, **11**, e0158569.

28. Szakacs, G.*, et al.* (2004) Comparing solid tumors with cell lines: implications for identifying drug resistance genes in cancer. *Mol Interv*, **4**, 323-5.

29. Sharma, S.V.*, et al.* (2010) Cell line-based platforms to evaluate the therapeutic efficacy of candidate anticancer agents. *Nat Rev Cancer*, **10**, 241-53.

30. Hastings, R.J.*, et al.* (1983) Cellular heterogeneity in a tissue culture cell line derived from a human bladder carcinoma. *Br J Cancer*, **47**, 233-44.

31. 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**, 3776-82.

32. Flatow, U.*, et al.* (1987) Tumorigenicity of T24 urinary bladder carcinoma cell sublines. *Int J Cancer*, **40**, 240-5.

33. 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**, 222-8.

34. Brown, J.L.*, et al.* (1990) Clonal analysis of a bladder cancer cell line: an experimental model of tumour heterogeneity. *Br J Cancer*, **61**, 369-76.

31

35.   Eigen, M. (1971) Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften*, **58**, 465-523.

FIGURE LEGENDS

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 5,000 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.

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.

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

32

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 4 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.

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.

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 Clone A and Clone B; (C-D) Cell counts against absolute copy number aggregated over tumor suppressors in SW480 Clone A and Clone B; (E-F) Cell counts against copy number relative to ploidy, aggregated over oncogenes in SW480 Clone A and B; (G-H) Cell counts

33

against copy number relative to ploidy, aggregated over tumor suppressors in

SW480 Clone A and B.

34

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 5,000 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.

DLD-1

	1	2	3	4	5
EGFR (7p)	8	80	8	4	0
CCND1 (11q)	7	86	4	3	0
CDX2 (13)	12	74	9	4	1
TERC (3q)	15	75	7	3	0

HCT116

	1	2	3	4
EGFR (7p)	2	92	3	1
CCND1 (11q)	3	91	4	1
CDX2 (13)	11	73	10	1
TERC (3q)	13	78	5	0

HT29

	1	2	3	4	5	6	7	8	9
EGFR (7p)	0	3	18	73	4	1	0	0	0
CCND1 (11q)	0	1	10	83	4	1	0	0	0
CDX2 (13)	1	4	20	50	12	5	3	2	1
TERC (3q)	1	4	20	67	4	0	0	0	0

SW480

	1	2	3	4	5	6	7
EGFR (7p)	1	4	32	54	7	1	1
CCND1 (11q)	1	16	66	13	3	1	0
CDX2 (13)	3	30	22	36	5	1	1
TERC (3q)	3	25	16	39	11	2	1

SW620

	1	2	3	4	5	6	7
EGFR (7p)	0	5	40	51	1	1	0
CCND1 (11q)	2	25	69	3	0	0	0
CDX2 (13)	62	29	3	1	1	1	1
TERC (3q)	7	81	8	1	1	0	0

H508

	1	2	3	4	5	6	7	8	9	10	11	12	13
EGFR (7p)	0	1	2	13	61	20	1	0	0	0	0	0	0
CCND1 (11q)	1	2	11	42	40	2	0	0	0	0	0	0	0
CDX2 (13)	4	3	3	4	5	9	12	16	18	13	6	2	1
TERC (3q)	7	8	13	19	21	18	7	2	1	0	0	0	0

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.

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 caluster 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. (E) Parental SW480 cells proliferated fastest as compared to single cell-derived clones with and without conditioned media.

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.*

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 Clone A and Clone B; (C-D) Cell counts against absolute copy number aggregated over tumor suppressors in SW480 Clone A and Clone B; (E-F) Cell counts against copy number relative to ploidy, aggregated over oncogenes in SW480 Clone A and B; (G-H) Cell counts against copy number relative to ploidy, aggregated over tumor suppressors in SW480 Clone A and B.*

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

Cell Lines	Clones	Ploidy	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 population (%)	Cells within Parental population (%)	Comparison to Parental Clone	Instability Index
HT29 Parental	A	3.36	3	4	3	4	10	4	4	3	2	3	2	5	66.40			11.40
	B	3.27	3	4	3	3	10	4	4	3	2	3	2	5	8.40			
	C	3.27	3	4	3	4	10	3	4	3	2	3	2	5	5.4			
HT29 1A4	1	3.36	3	4	3	4	10	4	4	3	2	3	2	5	69.33	66.40	Same as A	12.67
	2	3.27	3	4	3	4	10	4	3	3	2	3	2	5	5.67	0.40	Similar to A	
	3	3.27	3	4	3	3	10	4	4	3	2	3	2	5	4.00	8.40	Same as B	
	4	3.27	3	4	3	4	10	3	4	3	2	3	2	5	0.67	5.40	Same as C	
HT29 2B1	1	3.36	3	4	3	4	10	4	4	3	2	3	2	5	69.67	66.40	Same as A	14.67
	2	3.27	3	4	3	3	10	4	4	3	2	3	2	5	1.67	8.40	Same as B	
	3	3.27	3	4	3	4	10	3	4	3	2	3	2	5	1.00	5.40	Same as C	
HT29 2C6	1	3.27	3	4	3	4	10	4	4	3	2	3	2	5	71.43	8.40	Same as B	9.63
	2	3.36	3	4	3	4	10	4	4	3	2	3	2	5	0.66	66.40	Same as A	
HT29 2C12	1	3.36	3	4	3	4	10	4	4	3	2	3	2	5	74.00	66.40	Same as A	8.00
	2	3.27	3	4	3	4	10	4	4	3	2	3	2	4	5.00	1.40	Similar to A	
	3	3.27	3	4	3	3	10	4	4	3	2	3	2	5	3.00	8.40	Same as B	
	4	3.27	3	4	3	4	10	3	4	3	2	3	2	5	2.67	5.40	Same as C	
HT29 2D3	1	3.27	3	4	3	3	10	4	4	3	2	3	2	5	67.77	8.40	Same as B	15.95
	2	3.36	3	4	3	4	10	4	4	3	2	3	2	5	1.33	66.40	Same as A	
HT29 2D8	1	3.36	3	4	3	4	10	4	4	3	2	3	2	5	88.08	66.40	Same as A	4.30
	2	3.45	3	4	3	4	10	4	4	3	3	3	2	5	5.30	0.60	Similar to A	
	3	3.27	3	4	3	4	10	3	4	3	2	3	2	5	1.99	5.40	Same as C	
HT29 2E1	1	3.36	3	4	3	4	10	4	4	3	2	3	2	5	81.19	66.40	Same as A	8.91
	2	3.27	3	4	3	3	10	4	4	3	2	3	2	5	1.32	8.40	Same as B	
	3	3.27	3	4	3	4	10	3	4	3	2	3	2	5	0.66	5.40	Same as C	
HT29 2E2	1	3.36	3	4	3	4	10	4	4	3	2	3	2	5	90.33	66.40	Same as A	6.67
	2		3	4	3	3	10	4	4	3	2	3	2	5	0.67	8.40	Same as B	
HT29 2E4	1	3.27	3	4	3	3	10	4	4	3	2	3	2	5	91.00	8.40	Same as B	3.33
HT29 2G1	1	3.36	3	4	3	4	10	4	4	3	2	3	2	5	90.00	66.40	Same as A	4.67
	2	3.27	3	4	3	3	10	4	4	3	2	3	2	5	4.67	8.40	Same as B	
	3	3.27	3	4	3	4	10	3	4	3	2	3	2	5	0.33	5.40	Same as C	
HT29 2G5	1	3.36	3	4	3	4	10	4	4	3	2	3	2	5	91.06	66.40	Same as A	4.97
	2	3.27	3	4	3	3	10	4	4	3	2	3	2	5	3.97	8.40	Same as B	

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%

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

	Probes		COX2	TERC	APC	EGFR	MYC	CCND1	CDX2	CDH1	TP53	HER2	SMAD4	ZNF217	Cells within population (%)	Cells within Parental population (%)	Comparison to Parental Clone	Instability Index
	Chromosome		1q31.1	3q26	5q22	7p11	8q24.21	11q13.3	13q12	16q22.1	17p13.1	17q12	18q21	20q13.2				
Cell Lines	Clones	Ploidy																
SW480 Parental	A	2.36	2	2	2	3	10	3	2	2	2	2	2	4	42.20			24.40
	B	3.00	2	4	3	4	10	3	4	2	3	3	1	4	15.40			
	C	4.82	4	4	6	6	10	5	4	4	4	4	4	8	4.20			
	D	4.73	4	4	4	6	10	6	4	4	4	4	4	8	4.00			
SW480 1A5	1	3.18	3	4	3	4	10	3	4	3	3	3	1	4	37.33	0.00	Similar to B	14.00
	2	3.09	2	4	3	4	10	3	4	3	3	3	1	4	36.00	0.00	Similar to B	
	3	3.00	2	4	3	4	10	3	4	2	3	3	1	4	0.33	15.40	Same as B	
SW480 2A9	1	2.82	2	4	3	4	10	3	4	2	3	3	1	2	56.33	0.00	Similar to B	9.67
	2	2.73	2	4	2	4	10	3	4	2	3	3	1	2	22.33	0.00	Similar to B	
	3	2.91	2	4	4	4	10	3	4	2	3	3	1	2	6.33	0.00	Similar to B	
SW480 2B4	1	3.00	2	4	3	4	10	3	4	2	3	3	1	4	84.33	14.60	Same as B	8.00
SW480 2B9	1	2.36	2	2	2	3	10	3	2	2	2	2	2	4	66.00	42.2	Same as A	7.00
	2	2.27	2	2	2	2	10	3	2	2	2	2	2	4	18.00	1.00	Similar to A	
SW480 2C8	1	2.36	2	2	2	3	10	3	2	2	2	2	2	4	68.33	40.00	Same as A	15.33
	2	4.73	4	4	4	6	10	6	4	4	4	4	4	8	5.00	4.00	Same as D	
SW480 2D2	1	2.73	2	4	1	3	10	3	4	2	3	3	1	4	32.00	0.00	Similar to B	12.33
	2	2.82	2	4	2	3	10	3	4	2	3	3	1	4	20.00	0.00	Similar to B	
	3	2.91	2	4	2	4	10	3	4	2	3	3	1	4	13.00	0.00	Similar to B	
	4	2.82	2	4	1	4	10	3	4	2	3	3	1	4	11.67	0.00	Similar to B	
	5	3.00	2	4	3	4	10	3	4	2	3	3	1	4	3.33	15.40	Same as B	
SW480 2D4	1	2.91	2	4	3	4	10	3	3	2	3	3	1	4	81.00	0.80	Similar to B	6.00
	2	3.00	2	4	3	4	10	3	4	2	3	3	1	4	0.33	15.40	Same as B	
SW480 2F11	1	3.00	2	4	3	4	10	3	4	2	3	3	1	4	90.67	14.60	Same as B	4.35
SW480 2G4	1	2.27	2	2	2	2	10	3	2	2	2	2	2	4	93.33	1.00	Similar to A	4.33
	2	2.36	2	2	2	3	10	3	2	2	2	2	2	4	0.33	42.40	Same as A	
SW480 2G6	1	2.27	2	2	2	2	10	3	2	2	2	2	2	4	83.11	1.00	Similar to A	9.27
SW480 2H7	1	4.82	4	4	6	6	10	5	4	4	4	4	4	8	39.67	3.80	Same as C	37.00
	2	4.73	4	4	5	6	10	5	4	4	4	4	4	8	6.67	0.00	Similar to C	

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 MATERIALS AND METHODS

Fluorescence in situ Hybridization

Metaphase chromosomes were prepared using standard protocols. Cells for interphase FISH were prepared as described(1). Bacterial artificial chromosome (BAC) contigs were constructed for the analysis of six CRC cell lines for four probes: *EGFR*, *CCND1*, *TERC* and *CDX2*. Our standard FISH protocol was used for hybridization and detection (https://ccr.cancer.gov/Genetics-Branch/thomas-ried under resources). Slides were mounted with DAPI/antifade (Vector Laboratories, Burlingame, CA) and imaged using a Zeiss Axioplan with 40X objective using the Metafer4-5.2.19 (Metasystems, Newton, MA).

For miFISH, all probes were labeled with fluorophores by nick translation. BAC clones for *COX2*, *CCND1,* and *HER2* were labeled in green with DY-505-dUTP (Dyomics, Jena, Germany), those for *APC*, *TP53,* and *SMAD4* in gold with DY-547P1-dUTP (Dyomics), *TERC*, *CDX2*, and *CDH1* were labeled red with DY-590-dUTP (Dyomics), whereas *EGFR*, *MYC,* and *ZNF217* were labeled in aqua with DY-415-dUTP (Dyomics). Cells were collected as suspensions on Cell-Tak™ slides (Corning, Corning, NY), and fixed in methanol/acetic acid (3:1) for 10 minutes. The fixed slides were sequentially hybridized with the three probe panels. Hybridization and detection was performed as previously described(1).

Ploidy Assignment

1

Since *MYC* and *ZNF217* were amplified, and *APC, EGFR, CDX2* and *SMAD4* were

quite variable among the different cell lines, ploidy was assigned with the help of

calculating the average signal numbers of the remaining six gene markers: *COX2,*

TERC, CCND1, CDH1, TP53 and *HER2*. The following thresholds and rules were then

applied to assign ploidies from 2 to 8 to each cell using this six-marker average.

<2.5	Ploidy 2
=2.5	Ploidy 2 if more markers show 2 signals vs. 3 signals
	Ploidy 3 if more or an equal amount of markers show 3 signals vs. 2 signals
>2.5<3.5	Ploidy 3
=3.5	Ploidy 3 if more markers show 3 signals vs. 4 signals,
	Ploidy 4 if more or an equal amount of markers show 4 signals vs. 3 signal
>3.5<5.3	Ploidy 4
≥5.3<7.3	Ploidy 6
≥7.3	Ploidy 8

NanoString data analysis

Where the required routines to perform the analysis were available, analysis was

performed using NanoString's nSolver [3.0] software or one of its advanced analysis

modules. When computing the correlation between FISH probes and NanoString

probes, the required functionality was not available in nSolver, so a pipeline using Linux

system tools, Perl scripts and SQLite was developed, and statistical analysis was

ultimately performed using GNU R.

The raw Nanostring counts were scaled by the geometric mean of NanoString's

positive control probes as normalization for technical variation. Background subtraction

was not used, either for the heat plots or correlation computations. For the heatplots,

background subtraction censored low values unnecessarily and caused conceptual and

practical difficulties when logarithms were to be taken of zero or negative values. For

the sake of consistency with the heat plots, correlations are reported without

background subtraction. For completeness, we also computed correlations with

background subtraction, but background subtraction did not improve correlation. The

correlation of *SMAD4* remains significant (nominal p < 0.001) with background

subtraction.

Because the total amount of mRNA was precisely quantified, housekeeping

genes were not used as a control for total mRNA. For completeness, normalization via

housekeeping genes was attempted, with effectiveness assessed by ability to improve

correlation between FISH and NanoString probes, although it proved ineffective.

Phylogenetic tree inference

Tree models of tumor progression were computed using the software FISHtrees 3.1(2)

in the weighted, ploidyless mode(3), which models gains and losses of single genes,

gains and losses of single chromosomes, and genome doubling as distinct events with

different probabilities. Because *MYC* was so often censored at a count of 10, *MYC* was

not used when forming the trees.

Normally, FISHtrees is run on tumor samples, which generally 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 a mode in which the most abundant clone

was taken to be the root of the tree.

Sample complexity measures

Sample instability was computed using the FISH probe counts for genes measured, omitting *MYC*. The instability index was calculated by the number of signal patterns multiplied by 100, divided by the number of counted nuclei.

Simpson's index (also known as Gini-Simpson's index) was calculated as in prior work.(4,5) Simpson's index is bounded above by one, though this bound is not strict. If the configurations being measured can take on k different values, then the maximum value is $1 - \frac{1}{k}$. Numbers closer to one indicate more complexity. We applied Simpson's index to the distribution of observed cell count patterns. As with instability index, *MYC* was not used when computing Simpson's index.

Counts of gain and loss events inferred by FISHtrees

We counted the distribution copy numbers over the daughter cell lines derived from SW480. For each integer j from 0 to 9 and each gene g, we counted the number of cells that had copy number j for gene g. We then repeated the count of cells, but instead of grouping counts per gene by copy number, we grouped the counts by the difference k between the copy number of the gene and the ploidy of the cell. The relevant values of k are from -2 to 5.

We used FISHtrees to compute phylogenetic trees for the daughter cell lines derived from SW480. Trees consist of nodes and edges, each node representing a specific pattern of copy numbers, and each edge connecting a *parent* node to a *child* node by inferred mutational events. We collected the set of edges for which cells

4

having copy number configuration of the parent node of the edge were observed in the

daughter cell line. (FISHtrees may insert unobserved evolutionary intermediates.)

Then, for each possible parent p and each gene g, we recorded the number of edges

emanating from p that changed the copy number of g. Losses and gains were recorded

separately. It is common that not all genes are changed by an edge, and in such cases

neither gains nor losses are recorded for unchanged genes. For each copy number j

and each gene g, we summed the count of loss and gain edges for g over all parents

with copy number j for gene g. We performed a similar sum of gains and losses over

all ploidy differences k and gene g.

When accumulating data for oncogenes and tumor suppressors, *EGFR*, *CCND1*

HER2, and *ZNF217* were classified as oncogenes; and *APC*, *CDX2*, *CDH1*, *TP53* and

SMAD4 were classified as tumor suppressors. *COX2* and *TERC* were left unclassified,

as they are not typically known to be active oncogenes in CRC. *MYC* was not used in

any FISHtrees analysis.

Comparison of parental to child trees

For each of HT-29 and SW480, we also computed trees for the union of all single cell

daughter clones and the parent. To evaluate the extent to which each daughter follows

its own evolutionary trajectory, we computed the reconstruction error, as defined in,(3)

between the tree derived from each daughter sample alone and that derived from the

union the parental and daughter data for that cell line, a *union tree*. The lower the

reconstruction error, the more nearly the daughter tree matches a subtree of the union

tree. As a second test of independence in the daughter cells, we used FISHtrees to

5

compute consensus trees(6) comparing the tree for c) (the parent combined with each daughter) to the tree for each daughter alone. A lower value of weighted shared nodes suggests that the daughter follows a trajectory more independent of that of the parent. These measures additionally provide partial validation for tree inferences based on the known relationships of parental and daughter lines.

REFERENCES

1. 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**, 1807-22.

2. Chowdhury, S.A., *et al.* (2013) Phylogenetic analysis of multiprobe fluorescence in situ hybridization data from tumor cell populations. *Bioinformatics*, **29**, i189-98.

3. Chowdhury, S.A., *et al.* (2015) Inferring models of multiscale copy number evolution for single-tumor phylogenetics. *Bioinformatics*, **31**, i258-67.

4. Park, S.Y., *et al.* (2010) Cellular and genetic diversity in the progression of in situ human breast carcinomas to an invasive phenotype. *J Clin Invest*, **120**, 636-44.

5. Wangsa, D., *et al.* (2016) Phylogenetic analysis of multiple FISH markers in oral tongue squamous cell carcinoma suggests that a diverse distribution of copy number changes is associated with poor prognosis. *Int J Cancer*, **138**, 98-109.

6. Gertz, E.M., *et al.* (2016) FISHtrees 3.0: Tumor Phylogenetics Using a Ploidy Probe. *PLoS One*, **11**, e0158569.

6
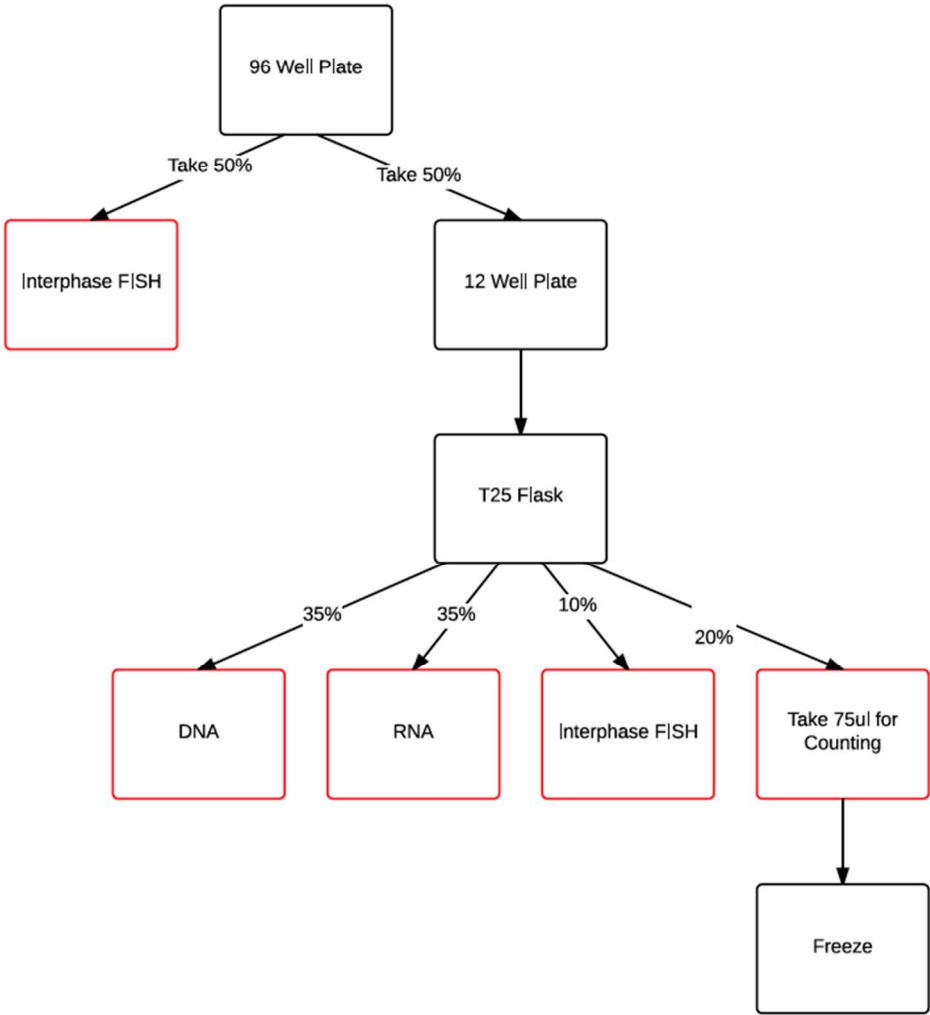
Supplementary Figure S1: Experimental strategy for the analysis of single cell-derived clones.

DLD-1

	1	2	3	4	5
EGFR (7p)	0	80	1	18	1
CCND1 (11q)	1	77	2	20	0
CDX2 (13)	1	78	2	19	0
TERC (3q)	1	79	1	19	0

HCT116

	1	2	3	4
EGFR (7p)	0	97	3	0
CCND1 (11q)	2	94	2	2
CDX2 (13)	3	93	1	3
TERC (3q)	3	93	1	3

HT29

	1	2	3	4	5	6	7	8
EGFR (7p)	0	0	11	87	1	1	0	0
CCND1 (11q)	1	0	6	91	2	0	0	0
CDX2 (13)	0	1	1	93	0	4	0	1
TERC (3q)	0	1	0	94	3	1	0	1

SW480

	1	2	3	4	5	6	7	8	9	10
EGFR (7p)	0	1	10	46	4	22	4	14	1	1
CCND1 (11q)	0	4	52	5	21	19	1	0	0	0
CDX2 (13)	0	53	2	46	1	0	0	0	0	0
TERC (3q)	1	51	3	43	3	0	0	0	0	0

SW620

	0	1	2	3	4	5	6	7	8
EGFR (7p)	0	0	8	27	62	2	1	0	1
CCND1 (11q)	0	0	4	9	79	6	0	0	2
CDX2 (13)	3	88	8	0	1	0	0	0	0
TERC (3q)	1	9	85	3	2	0	0	0	0

H508

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
EGFR (7p)	0	0	2	12	83	2	0	0	1	0	0	0	0	0
CCND1 (11q)	0	0	3	17	73	5	1	0	0	1	0	0	0	0
CDX2 (13)	0	0	0	1	1	1	3	3	23	61	6	0	0	1
TERC (3q)	0	0	0	3	12	80	1	3	1	0	0	0	0	0

Supplementary Figure S2: Cytogenetic analysis by metaphase FISH for six CRC cell lines using four probes (*EGFR*, *CCND1*, *CDX2* and *TERC*) simultaneously. The results are based on 100 metaphases for all cell lines except SW480 (n=200). The numbers indicate the copy number (X-axis) and next to each probe, the percentage of cells is shown.
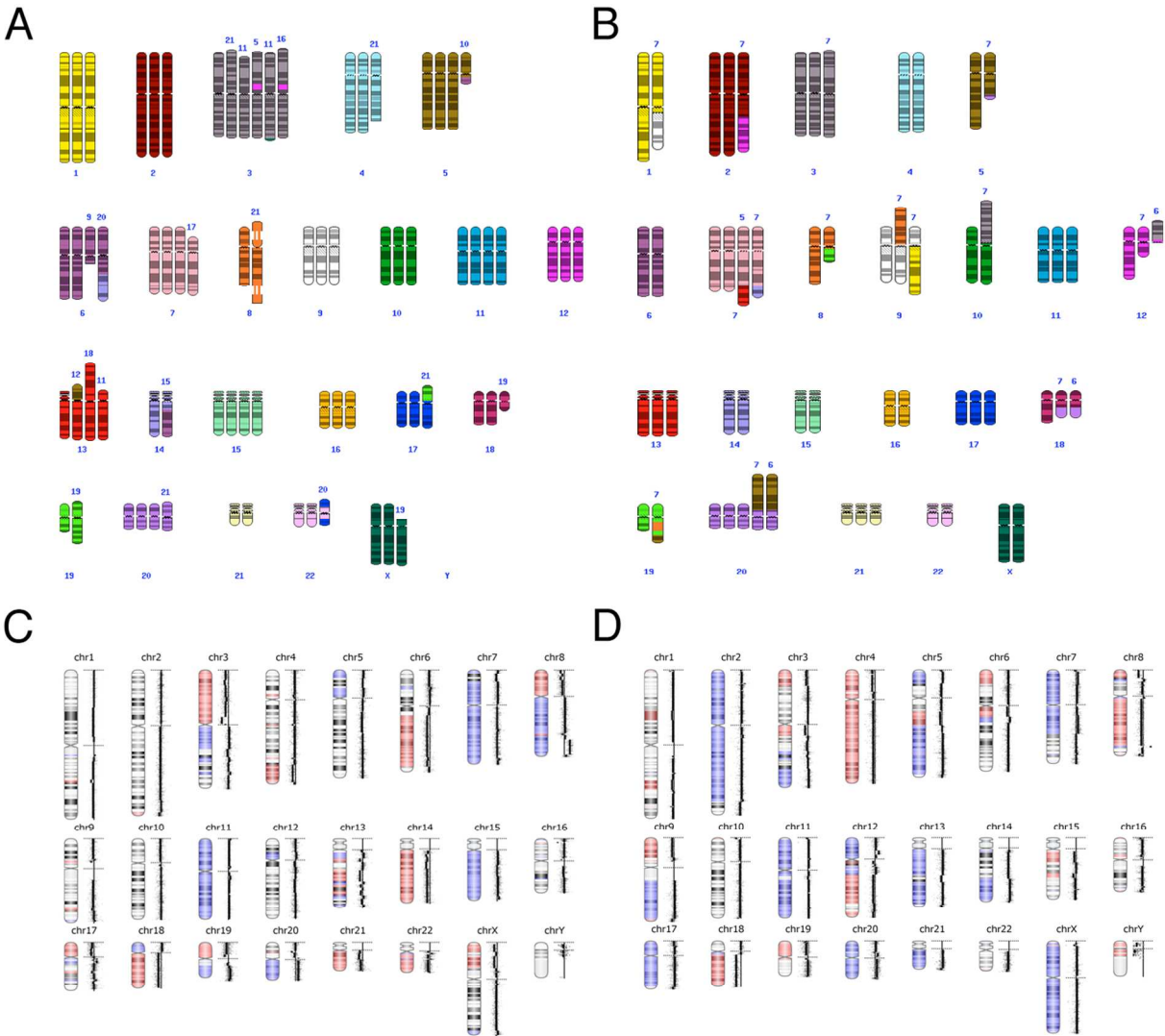
Supplementary Figure S3: SKY and aCGH results of the cell lines HT-29 (A and C) and SW480 (B and D).
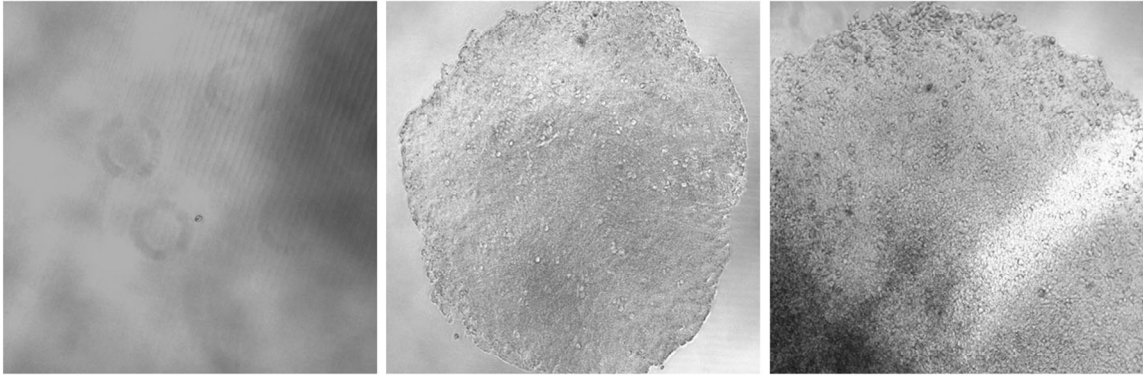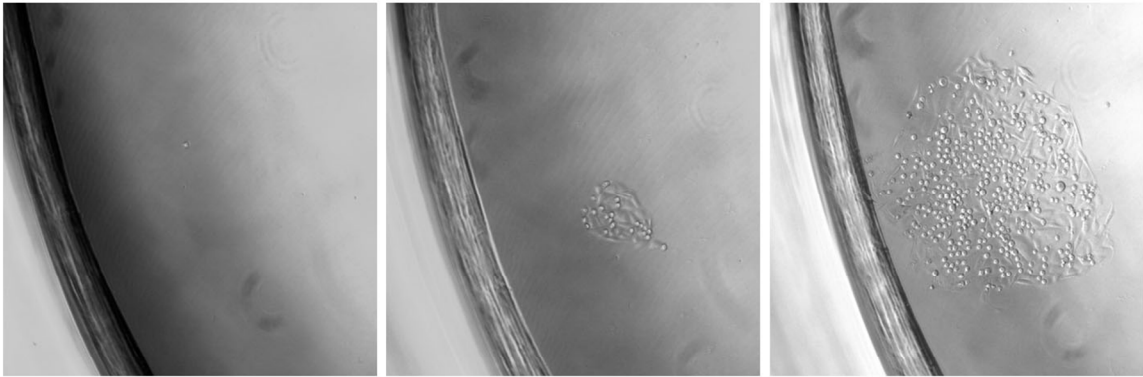
Supplementary Figure S4: Examples of single cell-derived clones before and after propagation for HT29 (A) and SW480 (B).

Supplementary Figure S5: Phylogenetic consensus FISHtrees of a stable single cell-derived line (A) SW480-2F11 and an unstable single cell-derived line (B) SW480-2D2 by weight. The consensus tree is specified by weight in percentage. Colors indicate 15 population doubling (blue), 22 population doubling (red) and common nodes (black). Common nodes are depicted as diamonds whereas those found only in either early (15 doublings) or late (22 doubling) cell culture passages are depicted as ellipses.
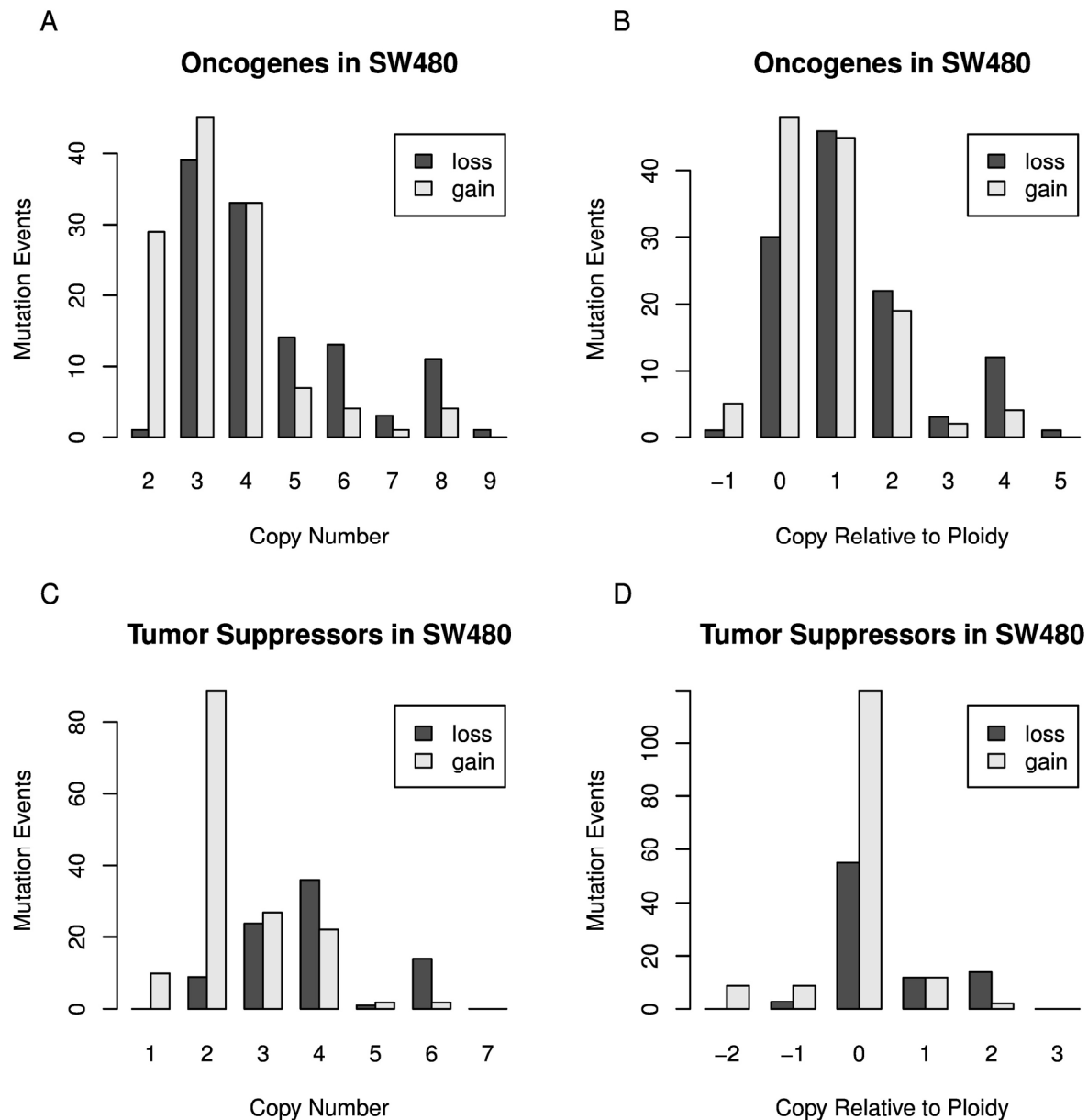
Supplementary Figure S6: Inferred gains and losses plotted against copy number counts, aggregated over genes of a given type. (A) Gains and losses against absolute copy number aggregated over oncogenes; (B) Gains and losses against copy number relative to ploidy, aggregated over oncogenes; (C) Gains and losses against absolute copy number aggregated over tumor suppressors; (D) Gains and losses against copy number relative to ploidy, aggregated over tumor suppressors.

Target region	PCR primer (forward)	PCR primer (reverse)
KRAS exon 2 codons 12&13	GGCCTGCTGAAAATGACTG	BIOTIN-TAGCTGTATCGTCAAGGCACTCT
KRAS exon 3 codons 59&61	GGCCATTTGTCCGTCATCT	BIOTIN-TCCTCATGTACTGGTCCCTCATT
KRAS exon 4 codon 117	TGGACAGGTTTTGAAAGATATTTG	BIOTIN-GTCCTGAGCCTGTTTTGTGTCTA
KRAS exon 4 codon 146	AGGCTCAGGACTTAGCAAGAAGTT	BIOTIN-GCCCTCTCAAGAGACAAAAACAT
NRAS exon 2 codons 12&13	GAGTACAAACTGGTGGTGGTTG	BIOTIN-GGATTGTCAGTGCGCTTTTC
NRAS exon 3 codons 59&61	AGATGGTAAACCTGTTTGTTG	BIOTIN-TATTGGTCTCTCATGGCACTGTAC
NRAS exon 4 codon 117	AGACTCGGATGATGTACCTATGG	BIOTIN-GCACAAATGCTGAAAGCTGTAC
NRAS exon 4 codon 146	ACAAAACAAGCCCACGAACTG	BIOTIN-TGAAAGCTGTACCATACCTGTCTG
BRAF exon 15 codon 600	TAGGTGATTTTGGTCTAGCTACA	BIOTIN-AGGGCCAAAAATTTAATCAGTG

Target region	Sequencing primer
KRAS exon 2 codons 12&13	AAACTTGTGGTAGTTGGA
KRAS exon 3 codons 59&61	TTGTTGGACATACTGGAT
KRAS exon 4 codon 117	TGGTGCTAGTGGGAAA
KRAS exon 4 codon 146	TTCCATTCATTGAAACC
NRAS exon 2 codons 12&13	GTGGTGGTTGGAGCA
NRAS exon 3 codons 59&61	TTGTTGGACATACTGGAT
NRAS exon 4 codon 117	TGGTGCTAGTGGGAAA
NRAS exon 4 codon 146	TTCCATTCATTGAAACC
BRAF exon 15 codon 600	TGATTTTGGTCTAGCTACA

Supplementary Table I. PCR primers and sequencing primers

Cell Lines	Time Point	Clones	Ploidy	COX2	TERC	APC	EGFR	MYC	CCND1	CDX2	CDH1	TP53	HER2	SMAD4	ZNF217	Cells within population (%)	Cells within Parental population (%)	Comparison to Parental Clone	Instability Index
SW480 Parental		A	2.36	2	2	2	3	10	3	2	2	2	2	2	4	42.20			
		B	3.00	2	4	3	4	10	3	4	2	3	3	1	4	15.40			
		C	4.82	4	4	6	6	10	5	4	4	4	4	4	8	4.20			24.40
		D	4.73	4	4	4	6	10	6	4	4	4	4	4	8	4.00			
SW480 1A5	96 well	1	3.18	3	4	3	4	10	3	4	3	3	3	1	4	48.00	0.00	Similar to B	19.33
		2	3.09	2	4	3	4	10	3	4	3	3	3	1	4	20.67	0.00	Similar to B	
SW480 1A5	T25	1	3.18	3	4	3	4	10	3	4	3	3	3	1	4	37.33	0.00	Similar to B	
		2	3.09	2	4	3	4	10	3	4	3	3	3	1	4	36.00	0.00	Similar to B	14.00
		3	3.00	2	4	3	4	10	3	4	2	3	3	1	4	0.33	15.40	Same as B	
SW480 2A9	96 well	1	2.82	2	4	3	4	10	3	4	2	3	3	1	2	42.33	0.00	Similar to B	16.00
		2	2.73	2	4	2	4	10	3	4	2	3	3	1	2	19.00	0.00	Similar to B	
SW480 2A9	T25	1	2.82	2	4	3	4	10	3	4	2	3	3	1	2	56.33	0.00	Similar to B	
		2	2.73	2	4	2	4	10	3	4	2	3	3	1	2	22.33	0.00	Similar to B	9.67
		3	2.91	2	4	4	4	10	3	4	2	3	3	1	2	6.33	0.00	Similar to B	
SW480 2B4	96 well	1	3.00	2	4	3	4	10	3	4	2	3	3	1	4	94.00	15.40	Same as B	2.67
SW480 2B4	T25	1	3.00	2	4	3	4	10	3	4	2	3	3	1	4	84.33	14.60	Same as B	8.00
SW480 2D2	96 well	1	3.00	2	4	3	4	10	3	4	2	3	3	1	4	61.33	5.40	Same as B	11.33
		2	2.92	2	4	3	3	10	3	4	2	3	3	1	4	19.00	0.00	Similar to B	
SW480 2D2	T25	1	2.73	2	4	1	3	10	3	4	2	3	3	1	4	32.00	0.00	Similar to B	
		2	2.82	2	4	2	3	10	3	4	2	3	3	1	4	20.00	0.00	Similar to B	
		3	2.91	2	4	2	4	10	3	4	2	3	3	1	4	13.00	0.00	Similar to B	12.33
		4	2.82	2	4	1	4	10	3	4	2	3	3	1	4	11.67	0.00	Similar to B	
		5	3.00	2	4	3	4	10	3	4	2	3	3	1	4	3.33	15.40	Same as B	
SW480 2D4	96 well	1	2.91	2	4	3	4	10	3	3	2	3	3	1	4	81.67	0.80	Similar to B	9.33
		2	3.00	2	4	3	4	10	3	4	2	3	3	1	4	0.67	15.40	Same as B	
SW480 2D4	T25	1	2.91	2	4	3	4	10	3	3	2	3	3	1	4	81.00	0.80	Similar to B	6.00
		2	3.00	2	4	3	4	10	3	4	2	3	3	1	4	0.33	15.40	Same as B	
SW480 2F11	96 well	1	3.00	2	4	3	4	10	3	4	2	3	3	1	4	91.67	15.40	Same as B	5.00
SW480 2F11	T25	1	3.00	2	4	3	4	10	3	4	2	3	3	1	4	90.67	14.60	Same as B	4.35

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

Threshold for parental cell lines: >4%

Genes	Sample	Simpson
(CDX2,TERC)	DLD1	0.614
(CDX2,TERC)	H508	0.974
(CDX2,TERC)	HCT116	0.584
(CDX2,TERC)	HT-29	0.811
(CDX2,TERC)	SW480	0.876
(CDX2,TERC)	SW620	0.733
(EGFR,CCND1)	DLD1	0.475
(EGFR,CCND1)	H508	0.820
(EGFR,CCND1)	HCT116	0.243
(EGFR,CCND1)	HT-29	0.560
(EGFR,CCND1)	SW480	0.791
(EGFR,CCND1)	SW620	0.752

Supplementary Table III. Simpson index for cell lines analyzed using pairs of FISH

probes.

FISH Probe	Test	Correlation Coefficient	Nominal P-value
APC	Pearson	-0.30	0.34
CCND1	Pearson	0.43	0.16
CDH1	Pearson	-0.20	0.53
EGFR	Pearson	0.59	0.044
HER2	Pearson	0.15	0.63
MYC	Pearson	0.17	0.60
SMAD4	Pearson	-0.065	0.84
TP53	Pearson	0.096	0.77
APC	Spearman	0.098	0.77
CCND1	Spearman	0.47	0.13
CDH1	Spearman	0.0035	0.99
EGFR	Spearman	0.30	0.34
HER2	Spearman	0.35	0.27
MYC	Spearman	0	1
SMAD4	Spearman	0.12	0.71
TP53	Spearman	0.43	0.16

Supplementary Table IV Correlation between NanoString and FISH probes for HT29.
The P-value shown is not corrected for multiple testing.

FISH Probe	Test	Correlation Coefficient	Nominal P-value
APC	Pearson	0.52	0.084
CCND1	Pearson	-0.37	0.24
CDH1	Pearson	0.00043	1.00
EGFR	Pearson	0.16	0.71
HER2	Pearson	0.37	0.24
MYC	Pearson	-0.24	0.45
SMAD4	Pearson	0.88	0.00086
TP53	Pearson	0.42	0.17
APC	Spearman	0.67	0.020
CCND1	Spearman	-0.30	0.35
CDH1	Spearman	-0.26	0.42
EGFR	Spearman	0.35	0.40
HER2	Spearman	0.45	0.14
MYC	Spearman	-0.43	0.16
SMAD4	Spearman	0.83	0.0028
TP53	Spearman	0.70	0.015

Supplementary Table V. Correlation between NanoString and FISH probes for SW480.
The P-value shown is not corrected for multiple testing.