Spontaneous Cell Fusions as a Mechanism of Parasexual Recombination in Tumor Cell Populations.

Daria Miroshnychenko[1*], Etienne Baratchart[3*], Meghan C. Ferrall-Fairbanks[3], Robert Vander Velde[1,2], Mark A Laurie[1], Marilyn M. Bui[4], Philipp M. Altrock[3], David Basanta[3,**] and Andriy Marusyk[1,2**].

Affiliations

1. Department of Cancer Physiology, H Lee Moffitt Cancer Center and Research Institute, Tampa, FL, USA.
2. Department of Molecular Medicine, University of South Florida, Tampa, FL, USA.
3. Department of Integrated Mathematical Oncology, H Lee Moffitt Cancer Center and Research Institute, Tampa, FL, USA.
4. Department of Pathology, H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL, 33612, USA

*, ** These authors contributed equally to the study

ABSTRACT

Initiation and progression of cancers reflects the underlying process of somatic evolution, which follows a Darwinian logic, *i.e.*, diversification of heritable phenotypes provides a substrate for natural selection, resulting in the outgrowth of the most fit subpopulations. Whereas somatic evolution can tap into multiple sources of diversification, it is assumed that it lacks access to (para)sexual recombination—one of the most powerful diversification mechanisms throughout all strata of life. Based on observations of spontaneous fusions involving cancer cells, reported genetic instability of polypoid cells, and precedence of fusion-mediated parasexual recombination in fungi, we interrogated whether cell fusions could serve as a source of parasexual recombination in cancer cell populations. Using differentially labelled tumor cells, we found evidence of low-frequency, spontaneous cell fusions in all of the tested breast cancer cell lines both *in vitro* and *in vivo*. While some hybrids remained polyploid, many displayed partial ploidy reduction, generating diverse progeny with heterogeneous combinations of parental alleles. Hybrid cells also displayed elevated levels of phenotypic plasticity, which may further amplify the impact of cell fusions on the diversification of phenotypic traits. Using mathematical modeling, we demonstrated that observed rates of spontaneous somatic cell fusions may allow populations of tumor cells to explore larger areas of the adaptive landscape, relative to strictly asexual populations, which may substantially accelerate a tumor's ability to adapt to new selective pressures.

INTRODUCTION

Cancer is the direct result of somatic clonal evolution, which follows Darwinian-first principles: diversification of heritable phenotypes provides a substrate, upon which natural selection can act, leading to preferential outgrowth of those phenotypes with higher fitness in the specific environment[1,2]. Thus, the ability to generate new heritable diversity is required for the evolvability of populations of tumor cells, both during tumor progression and in response to therapies. Evolving tumors have access to several powerful diversification mechanisms that are considered to be the enabling characteristics within the hallmarks of cancer[3]: genomic instability, elevated mutation rates at nucleotide level, and deregulation of epigenetic mechanisms that control gene expression. At the same time, cancer cells are generally assumed to lack a key evolutionarily-conserved source of diversification—sexual or parasexual (exchange of genetic material without meiosis) recombination. Parasexual process that can dramatically amplify diversity and generate new combinations of beneficial mutations (thus enabling new epistatic interactions), while separating them from disadvantageous ones, hence supporting population fitness and accelerating evolutionary adaptation[4-6].

Cancers are assumed to be strictly asexual, i.e. all the novel genetic and epigenetic solutions "discovered" by tumor cells are thought to be strictly clonal, inheritable only by the direct progeny of (epi) mutated cells. However, occurrences of spontaneous cell fusions within tumors have been documented both *in vitro* and *in vivo*[7-9]. Given the previously reported impact of polyploidization on increased genomic instability[10,11], evidence of ploidy reduction in the progeny of experimentally induced hybrid cells[12,13], and existence of parasexual life cycles involving fusion-mediated recombination in fungi, such as pathogenic yeast C. albicans[14], we decided to examine whether spontaneous cell fusion could serve as a source of diversification in tumor cell populations. We found that, while relatively infrequent, spontaneous cell fusions can be detected in a wide range of breast cancer cell lines both *in vitro* and *in vivo.* A subset of these hybrid cells demonstrated clonogenic viability. While hybrid cells initially consist of the parents' combined genomes, some of the hybrids undergo ploidy reduction that is accompanied by genome recombination, which generates new sub-clonal diversity. Mathematical modeling predictions suggest that this fusion-mediated recombination could augment the evolvability of tumor cell populations even when spatial limitations are considered. Thus, our studies suggest that spontaneous cell fusions may provide populations of tumor cells with a mechanism for parasexual recombination, capable of accelerating diversification an enhancing evolvability.

RESULTS

In the course of multiple experimental studies involving *in vitro* co-cultures of tumor cells that carry different fluorescent protein labels we occasionally noticed double-positive cells upon fluorescent microscopy analyses (**Fig 1A, S1)**. The same observation of double-positive cells was made studying tumor cells and cancer-associated fibroblasts (CAFs) (**Fig. S2 B, C**). Examination of time lapse microscopy images revealed that these double-positive cells can originate from spontaneous cell fusions **(Fig. S2A, B, Videos S1-5**). The phenomenon was not limited to *in vitro* cultures. Confocal microscopy examination of experimental xenograft tumors also revealed occasional presence of cells expressing both fluorescent labels (**Fig. 1B**).

Given the possibility that spontaneous cell fusions between genetically distinct cells might provide evolving populations of tumor cells with new sources of genetic diversification, we decided to systematically investigate this phenomenon. To this end, we labelled panels of breast cancer cell lines and primary breast cancer CAF isolates with lentivirally expressed GFP and mCherry reporters, coupled with resistance genes, to blasticidin and puromycin, respectively. Differentially labelled cells of the same (homotypic) or distinct (heterotypic) cell lines were plated at a 1:1 ratio and, after co-culturing for 3 days, subjected to flow cytometry analysis (**Fig. 1C**). Compared to the separately cultured controls, harvested and admixed no more than 30 min prior to the analysis, all of the heterotypic co-cultures and five out of seven examined homotypic cultures exhibited higher proportions of events in the double-positive gate (two of the homotypic cultures reached statistical significance) (**Fig. 1D, Fig. S2C**).

The significantly higher proportion of putative double-positive cells detected by flow analysis compared to microscopy examination, substantial within group variability, and the detection of double-positive events in some of the negative control samples indicated significant rates of false positives. Therefore, we set to validate the flow cytometry findings using ImageStream, an imaging based platform that combines the high processivity of flow cytometry analysis with the ability to evaluate recorded images of each event[15]. Indeed,

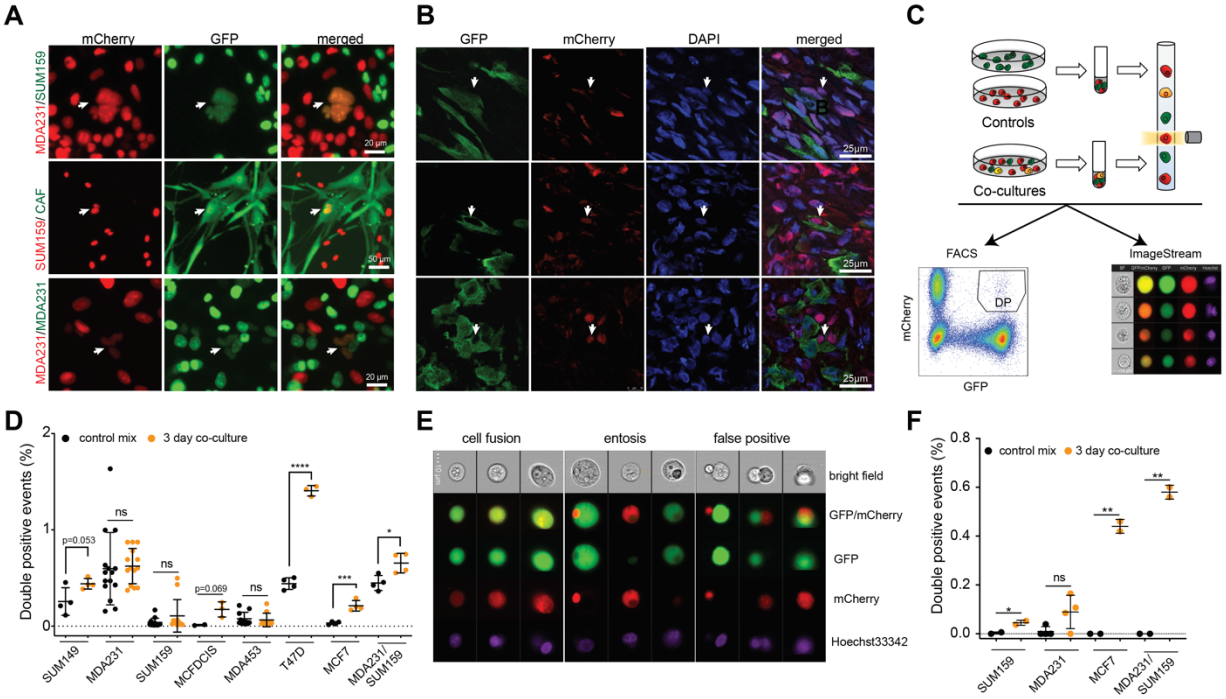


Fig.1 Detection of spontaneous cell fusions *in vitro* and *in vivo*.
A. Live fluorescent microscopy of co-cultures of the indicated differentially labelled cells. Arrowheads indicate cells co-expressing both fluorescent labels. **B.** Confocal immunofluorescent images from a xenograft tumor, initiated with a 50/50 mix of GFP and mCherry-labelled MDA-MB-231 cells. Arrowheads indicate cells co-expressing GFP and mCherry. **C.** Experiment schemata for flow cytometry (conventional and ImageStream) studies. DP in a representative flow cytometry histogram indicates double-positive (GFP and mCherry) populations. **D.** Quantification of FACS-detected frequencies of DP events of *in vitro* cell fusions of the indicated homotypic and heterotypic mixes. Each dot represents a measurement from an independent biological replicate. **E.** Representative images from ImageStream analyses of co-cultures of differentially labelled MCF7 cells. Hoechst33342 was used as a nuclear stain. **F.** Quantification of visually validated DP events from the ImageStream data. *, **, ***, *** indicate p values below 0.05, 0.01, 0.001, and 0.0001, respectively for the to-tailed unpaired t-test.

examination of the images of double-positive gate events (gating logic provided in **Fig.S3A**) revealed significant rates of false positives reflecting cell doublets (**Fig 1E**). Some of the double-positive events were cells-within cells structures, indicating entosis[16] or engulfment of cell fragments (**Fig 1E**). Still, a substantial fraction (~ 20%) of double-positive events were unambiguous mono- or bi-nucleated single cells with clear, overlapping red and green fluorescent signals, indicative of *bona fide* cell fusions (**Fig 1E, F, S3B)**. Direct comparison of flow cytometry and ImageStream analysis of the same sample revealed that true positives represented ~30% of the DP events detected by flow analysis (**Fig. S3C**). Consistent with the expected increase in cell size resulting from fusion of two cells, double-positive cells were significantly larger than cells expressing a single fluorescent marker (**Fig. S3D**). In summary, these results suggest that, while spontaneous cell fusions between cancer cells are relatively infrequent, they occur in a wide range of experimental models.

Next, we asked whether hybrid cells, formed by spontaneous somatic cell fusions, are capable of clonogenic proliferation. To this end, we co-cultured the differentially labelled cells for 3 days, then subjected them to the dual antibiotic selection (**Fig. 2A**). After two weeks of selection, which was sufficient to eliminate cells in the single-labelled negative controls, all of the examined breast cancer cell lines invariably contained viable, proliferating hybrid cells expressing both GFP and mCherry fluorescent markers (**Fig. 2B**). Notably, clonogenic proportions of cells with dual-antibiotic resistance was lower than the frequency of fusion events (**Fig. 1F**

and **2C**), suggesting that only some of the hybrids are capable of sustained proliferation. Similarly, we were able to recover dual-antibiotic-resistant cells with dual fluorescence from *ex vivo* cultures of xenograft tumors initiated by co-injection of differentially labelled cells (**Fig. 2B**). Surprisingly, despite the relatively high rates of fusion detected by flow cytometry (**Fig S2C**), and previous reports on formation of viable hybrids formed by fusions between carcinoma and stromal cells[17,18], we were unable to recover colonies from co-cultures between multiple breast cancer cells and three distinct primary CAF isolates.

Next, we asked whether hybrids formed by spontaneous somatic fusions between cancer cells differ in their proliferative and invasive potential from those of parental cells. At early passages during the post-antibiotic-selection phase, hybrids displayed lower net proliferation rates compared to those of parental cell lines (**Fig. 2D, Fig. S4A**). However, at later passages, most of the examined hybrids matched and, in some cases, exceeded the proliferation rates of the fastest growing parent. This observation is consistent with the elimination of viable but non-proliferative hybrids along with the selection of variants with higher proliferation abilities. (**Fig. 2D, Fig. S4A**). Transmembrane invasion assays revealed that most hybrids displayed invasive rates equal to or exceeding rates of the more invasive fusion parents (**Fig. 2E, Fig. S4B, C**).

Next, we assessed the impact of somatic cell fusions on metastatic colonization potential. To this end, we compared lung colonization potential between a cell line with a relatively weak lung colonization potential (SUM159PT) and a cell line with a strong lung colonization potential (MCF10DCIS), along with their hybrids using the tail vein injection assay. Despite identical initial lung seeding efficiencies, mice injected with SUM159PT cells lost luminescent signal from the lungs (**Fig. 2F**). Histological examination revealed the presence of multiple micro-metastatic nodules, suggesting a microenvironmental growth bottleneck rather than inability to seed lungs *per se* (**Fig. 2G)**. In contrast, luminescent signals in all four mice injected with MCF10DCIS and in two out of four mice injected with the hybrid cells increased over time two out of four animals, while the other two hybrid cell recipients displayed reduced, but detectable luminescent signal. Histological examination revealed that lungs of all of the recipients MCF10DCIS and hybrid cell recipient mice contained macroscopic tumors in the lungs. Surprisingly, despite weaker luminescent signals in two out of three analyzed animals, hybrid tumors grew larger than those from MCF10DCIS tumors (**Fig 2G, H, S4 D**), likely reflecting loss of luciferase gene expression in some of the hybrids (one of the mice with strong luminescent signal denoted by red X in **Fig. 2F** had died prior to euthanasia; necropsy analysis revealed massive tumors in the lungs, but due to poor tissue quality this animal was excluded from the analysis). In summary, these data suggest that that spontaneous cell fusions between neoplastic cells can generate cells with more aggressive oncogenic properties.

In the absence of the TP53 depended checkpoint function, which is commonly disrupted in cancer cells, polyploidy is known to be associated with increased genomic instability[11]. Consistently, genomic

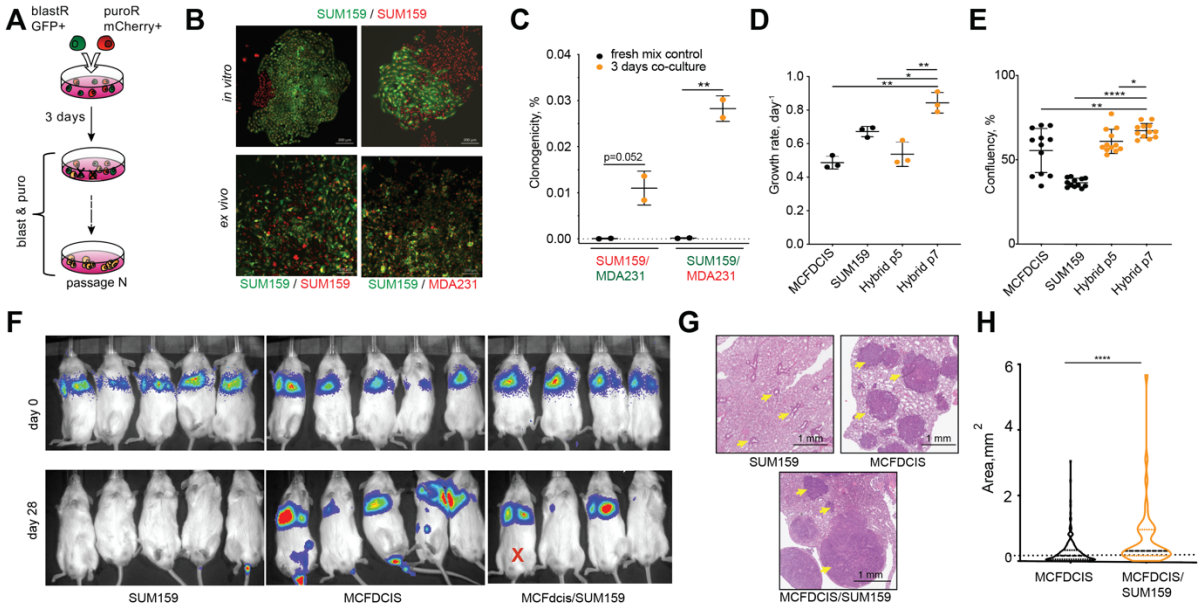


Fig.2 Phenotypic characterization of hybrids.
A. Experiment schemata for the selection of hybrid cells. **B.** Representative images of live fluorescent colonies formed after selection of the indicated *in vitro* co-cultures or *ex vivo* tumors. **C.** Quantitation of clonogenic survival frequencies. **D.** Growth rates of the indicated parental cell lines and hybrids at the indicated passages. **E.** Quantification of transwell cell migration assays of the indicated cells. **F**. Bioluminescence imaging of animals, injected with the indicated cells via the tail vein at the indicated times. Red X denotes a mouse that died prior to euthanasia. **G**. Representative images of H&E stains of lungs from the indicated xenograft transplants. Yellow arrows point to example tumors. **H.** Violin plot of size distribution of individual metastatic tumors from H&E stains as in G. *, **, ***, *** indicate p values below 0.05, 0.01, 0.001 and 0.0001, respectively for the to-tailed unpaired t-test (C, D & E) and Mann-Whitney U test (H).

instability[19] and ploidy reduction were reported in experimentally-induced somatic hybrids[12,13]. Therefore, we decided to examine the stability of hybrid cells' DNA content over time. As expected, immediately following antibiotic selection all of the examined hybrid cell lines displayed elevated ploidy, is consistent with thecombined DNA content of two parents (**Fig S5**). However, three out of seven examined hybrids displayed clear evidence of ploidy reduction with passaging, while the ploidy of the remaining four hybrids remained ostensibly stable (**Fig S5**).

Since fusion-mediated recombination and stochastic loss of parental DNA accompanying ploidy reduction can serve as the mechanism for parasexual recombination in the pathogenic yeast species C Albicans[14,20] and that cycles of somatic cell fusions followed by genetic recombination and ploidy reduction have been described to operate in normal hepatocytes[21] and hematopoietic cells[22], we decided to compare the genomes of single-cell-derived sub-clones of somatic hybrids (**Fig. 3A**). Consistent with the lack of ploidy reduction in mixed populations, all of the examined sub-clonal derivatives of MCF10DCIS/SUM159PT and MDA-MB-231/MCF10DCIS hybrids retained elevated ploidy (**Fig S5**). In contrast, genomic sub-clonal derivatives of the hybrids with evidence of ploidy reduction (HS578T/MDA-MB-231 and MDA-MB-231/SUM159PT) displayed clear evidence for clonal variation in DNA content (**Fig 3B, Fig. S5)**.

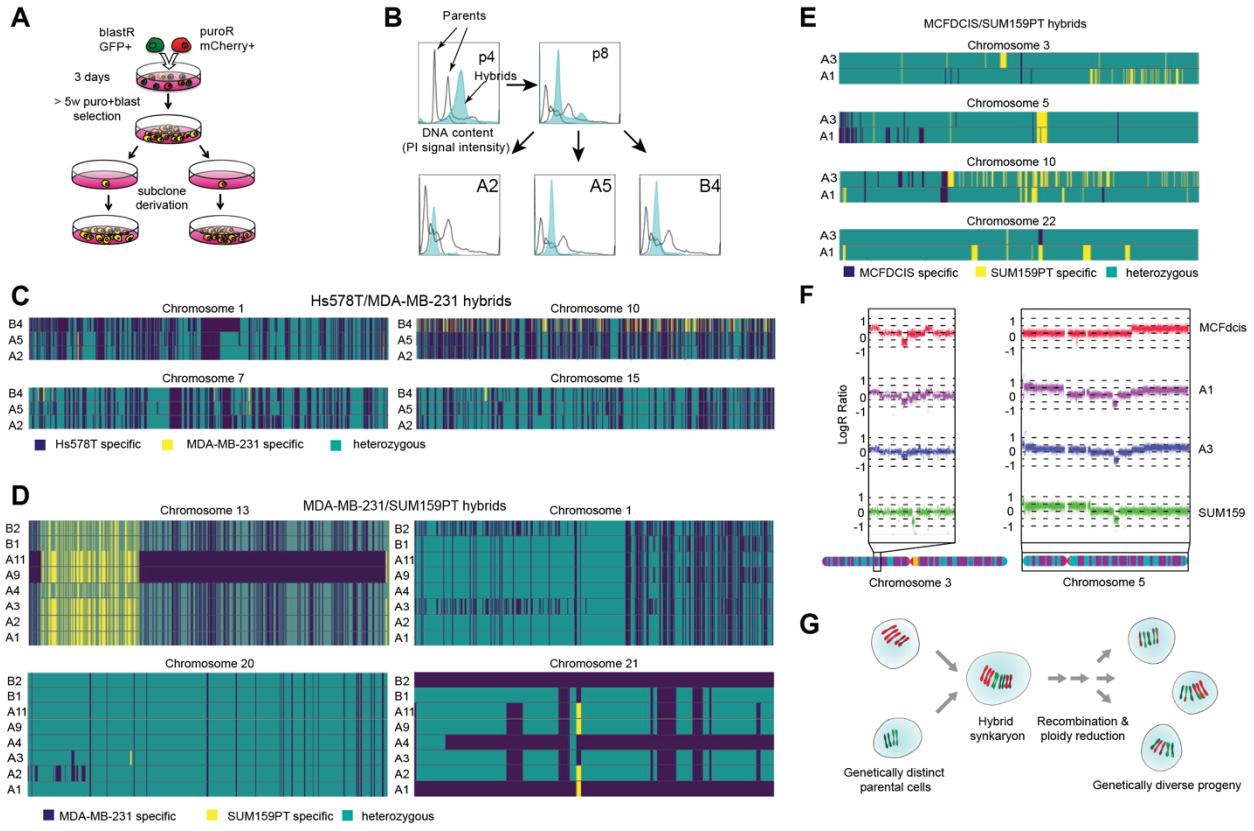


Fig.3 Fusion mediated genetic diversification.
A. Experiment schemata for the derivation of single cell derived hybrid subclones. **B.** DNA content analysis of parental cell lines (HS578T and MDA-MB-231) and hybrids. A2, A5, B4 denote single cell derived subclones. **C – E.** Allelic inheritance of differential SNP markers of the parental cell lines in the subclonal progeny of the indicated hybrids. Rows depict distinct sub-clones shown in **Fig. S5**; columns indicate parent-specific alleles, as described in the color key. **F.** LogR Ratio plots for Chromosome 5 and a fragment of chromosome 3 depict different copy numbers for different fragments of the chromosomes. Chromosome graphics were created by the U.S. Department of Energy Genomic Science program (https://genomicscience.energy.gov). **G.** Scheme illustrating the proposed fusion mediated recombination mechanism.

In order to gain deeper insights into the impact of hybridization on diversification of heritable phenotypes, we compared patterns of inheritance of cell-line-specific alleles. Using the Affymetrix CytoScan SNP (single nucleotide polymorphism) platform, we found evidence of significant divergence in patterns of allelic inheritance between different sub-clones of the genetically unstable HS578T/MDA-MB-231 hybrids (**Fig. 3B, Fig. S5**). These differences indicate that ploidy reduction may be accompanied by genetic recombination (**Fig.3C, S6**). Likewise, we observed substantial divergence in allelic inheritance among different sub-clones from the unstable MDA-MB-231/SUM159PT hybrids using the Illumina CytoSPN-12 platform (**Fig. 3D, Fig. S7**). Surprisingly, CytoSPN-12-based analysis of the ostensibly stable MCF10DCIS/SUM159PT hybrids (**Fig. S5**) revealed some divergence in allelic inheritance between the two clones, despite maintenance of similarly elevated DNA content (**Fig. 3E, Fig. S8)**.

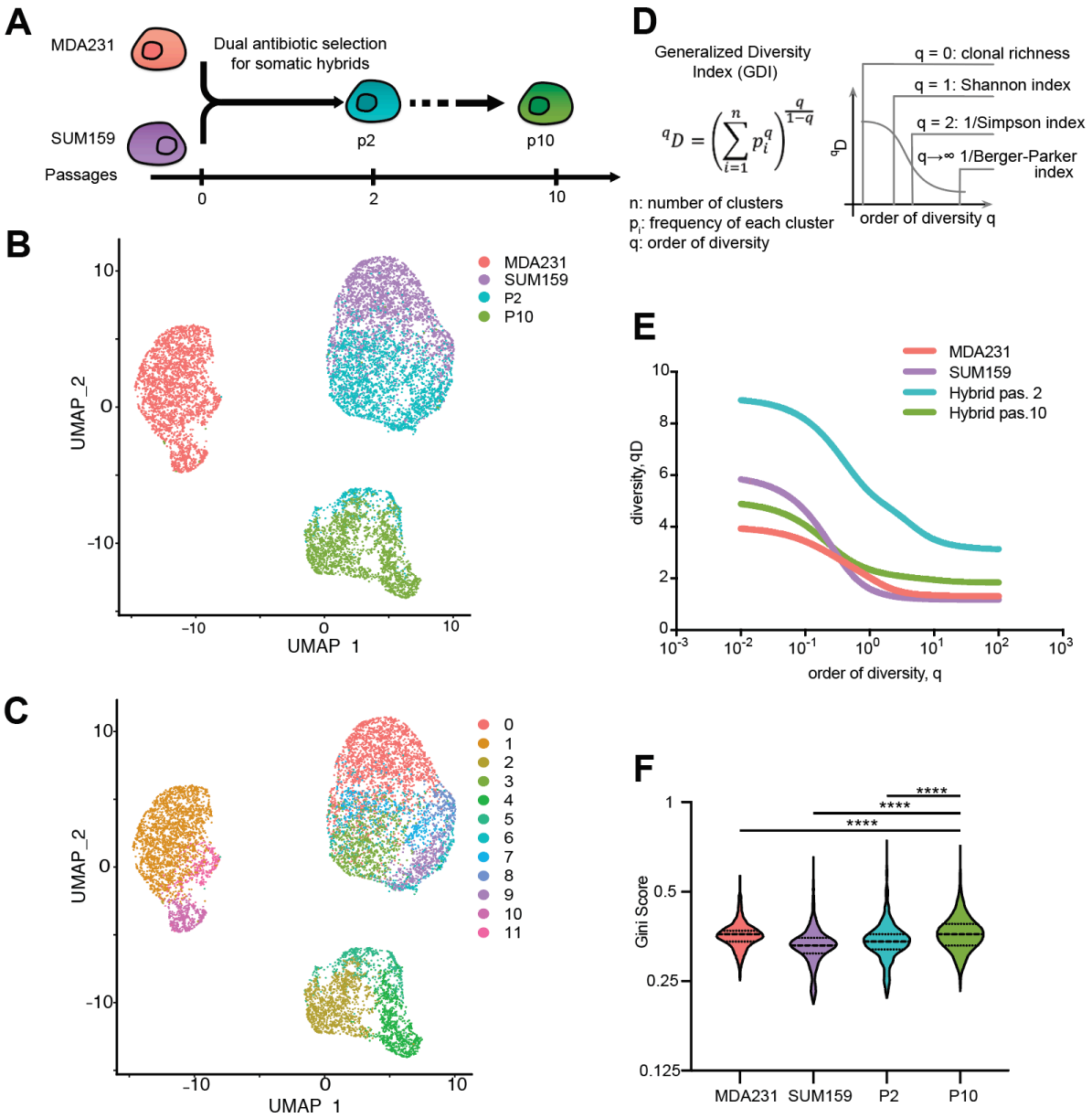


Fig 4. Fusion mediated phenotypic diversification.

A. Experimental schematic for samples used for scRNAseq analyses. **B.** UMAP distribution of cell phenotypes of parental cells and hybrids at the indicated passages from single cell expression analysis data. **C**. UMAP-defined distinct phenotypic clusters, used in GDI analyses. **D**. Formula for calculation of a generalized diversity index (GDI), and mappings to common diversity indexes that are special cases. **E**. GDI analysis of phenotypic diversity of parental and hybrid cells. **F.** Comparison of Gini scores across all of the genes with expression value >1 read in all four cell lines, between the indicated cells. Dashed lines represent medians and dotted lines represent quartiles. **** denotes p ≤ 0.0001 in a Wilcoxon signed rank test.

It should be noted that the actual genetic variability within progeny of cell fusions could be higher than suggested by the binary analysis of allelic inheritance, as it misses potential differences in copy numbers of the parent-specific alleles. Since accurate comparison of copy number differences in highly aneuploid genomes of most cancer cell lines is complicated, we decided to focus on hybrids between the near-diploid cell lines

SUM159PT and MCF10DCIS, which did not show evidence of ploidy reduction based on flow cytometry analyses (**Fig. S5**). Analysis of genomic regions that retain both parental alleles revealed copy number differences between the distinct clones, suggesting that genomic diversification impacts not only differential inheritance of alleles but also their copy number variation (**Fig. 3F**). In summary, these data suggest that cell fusion between genetically distinct tumor cells can be a source of significant genetic diversification (**Fig. 3G**).

In addition to genetic diversification, heterogeneity in biologically and clinically important phenotypes may also be shaped by epigenetic mechanisms. Theoretical studies have suggested that cell fusions between genetically identical but phenotypically distinct cells could create significant diversity due to the resultant collision of gene expression networks[23]. Therefore, we decided to examine the impact of somatic fusions on phenotypic diversification. To this end, we performed single-cell expression profiling (10x Genomics platform) to examine the phenotypes of MDA-MB-231, SUM159PT cells and SUM159PT/MDA-MB-231 hybrids at early (2d) and late (10d) passages with dual-antibiotic selection (**Fig. 4A**). Clustering[24] of single cell expression profiles revealed that phenotypes of the hybrids were distinct from both parents. Interestingly, we observed a substantial shift in the phenotypes of hybrids at the later passage, which is consistent with the selection of a fit sub-population of hybrid cells and additional diversification (**Fig. 4B, C**).

In order to quantify phenotypic diversity within parental cell types and the hybrids from the single-cell profiling data, we decided to take advantage of a general diversity index (GDI) **(Fig. 4D**)[25]. GDI enables characterization of diversity across a spectrum of orders of diversity[26], ranging from clonal richness (low order of diversity reveals the number of distinct subpopulations) to classic measures of species diversity, such as Shannon or Simpson indices[27] (intermediate orders of diversity), and to high orders of diversity that giving increased weight to the highly abundant sub-populations[25]. We considered individual (UMAP) clusters identified as sub-populations ("species") and found that, at the early passage, hybrids displayed higher diversity across all orders of diversity (**Fig. 4E**). At the later passage, diversity at low orders ("species richness") decreased. Yet, at intermediate and high orders ("species evenness") diversity of late passages remained higher than in either those of the parental cells.

Our GDI analyses rely on grouping phenotypes into clusters based on high throughput single cell analyses. These analyses might miss rare subpopulations and general elevation of cell-to-cell variability. Therefore, we decided to interrogate the dispersion of transcript reads across cells using the Gini dispersion index, which captures variability of gene expression across all of the transcriptome and has been recently applied towards characterization of phenotypic diversification in cancer cell populations[28]. We found that SUM159PT/MDA-MB-231 hybrids displayed elevated Gini indexes compared to both parents (**Fig. 4F**). This finding further supports the notion that somatic hybridization can lead to phenotypic diversification.

Despite the substantial genetic and phenotypic diversification observed in hybrid cells and their progeny, the impact of this diversification on diversity of large populations of tumor cells is less intuitively

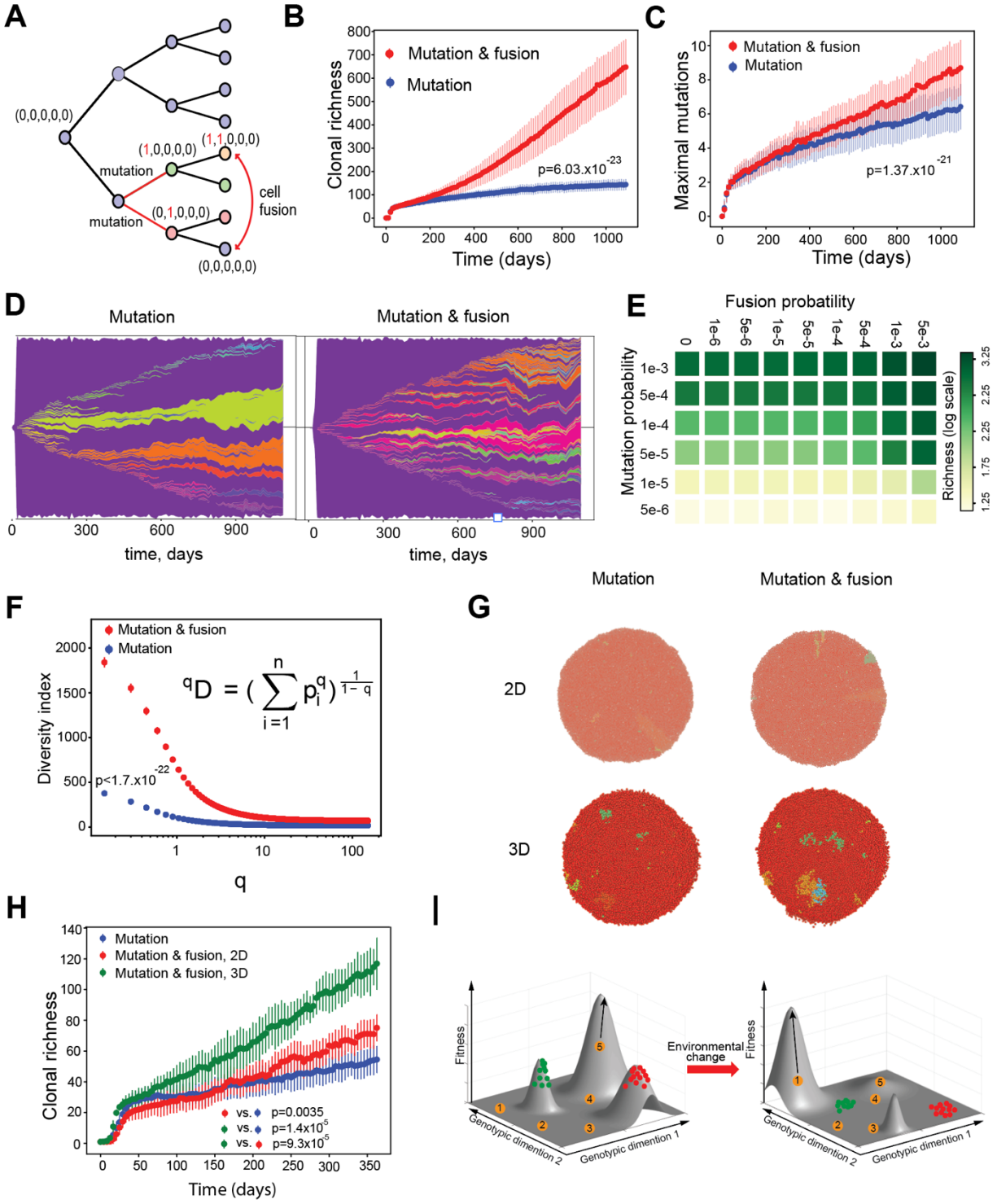


Fig.5. Impact of fusion mediated recombination on genetic diversity in tumor cell populations.
A. Model schemata of a birth-death process. A binary vector represents the genotype of the cells, with (0,0,…,0) representing the initial genotype. Cells can stochastically acquire mutations during cell division and randomly exchange mutations during fusion-mediated recombination. Dynamics of unique mutation acquisition (**B**) and highest number of unique mutations within a single lineage (**C**) in the presence and absence of fusion mediated recombination (fusion probability pf=0.005, mutational probability $\mu=5x10^{-5}$) over 1095 days. For both (B) and (C) the solid lines represent the means over 50 seeds and the error bars represent standard deviations. **D.** Muller plot depicting the impact of fusion-mediated recombination on clonal dynamics, with pf either 0 or 0.01, and $\mu=1x10^{-5}$. **E.**

Impact of mutational and fusion probability on clonal richness. **F**. Generalized Diversity Index after 1095 days of simulation; pf=0.001, $\mu=1x10^{-4}$. **G**. Visualization of the spatial distribution of sub-populations carrying unique genotypes at the end of in silico simulation of growth for 60 days with pf=0.00035, $\mu=5x10^{-3}$ **H**. Number of distinct mutants over time for the spatial simulations for 365 days, pf=0.00035, $\mu=5x10^{-3}$, carrying capacity 10000 cells; The solid lines represent the means over 10 seeds and the error bars are the standard deviation. **I**. Putative impact of fusion mediated recombination on the ability of populations of tumor cells to explore adaptive landscape. Left - under constant conditions, we assume that major genetic sub-clones occupy local fitness peaks and that most new variants are disadvantageous. However, some of the new mutants might "discover" a distinct fitness peak, leading to amplification through positive selection. Right – environmental change (such as initiation of therapy) changes adaptive landscape. Some of the new variants produced by fusion-mediated recombination might "discover" new fitness peaks. All of the p values show the results of KS tests for the indicated comparisons

clear given the relatively low frequency of spontaneous cell fusions. To interrogate the impact of fusion at the population level, we used *in silico* simulations based on a birth-death branching model of tumor growth (**Fig. 5A and Mathematical Supplement**). We started by deriving the probability of clonogenic cell fusions and cell proliferation from our experimental data (**Mathematical Supplement**). Using these estimates together with values of mutation rates from the literature[29], we compared diversification rates between the scenarios of populations of tumor cells evolving through mutations only, and through mutation with fusion mediated recombination.

We found that fusion-mediated recombination substantially increases accumulation of clonal richness (of sub-populations defined by unique mutational combinations), and also increases maximum numbers of mutations observed within a single lineage (**Fig. 5B-D**). This effect was observed across a biologically feasible range of fusion and mutation rates (**Fig. 5E, Fig. S9**). The impact of fusion was more pronounced at higher fusion probabilities. Less intuitively, the impact of cell fusions was also elevated by higher mutation rates.

In order to better understand stronger impact of fusions at higher mutation rates, we examined the ability of fusion-mediated recombination to generate new mutational variants as a function of pre-existing clonal richness, given fixed mutational and fusion probabilities. We found that higher levels of mutational heterogeneity dramatically enhanced the impact of fusion-mediated recombination (**Fig S10**), suggesting that biological impact of spontaneous cell fusions may be higher in tumors with higher levels of genetic intratumor heterogeneity. Even though the impact of fusion-mediated recombination was clearly reflected in the commonly used Shannon and Simpson diversity indexes (**Fig. S11A, B**), GDI enabled a more informative capture of the of the phenomenon, showing the highest impact at the lowest orders of diversity (**Fig. 5F, Fig. S11C**).

In our modeling so far, we assumed a well-mixed population. However, this assumption is violated in solid tumors, where spatial restrictions can have a profound impact on selective pressures and expansion of mutant subpopulations[30-33]. Thus, we asked whether fusion-mediated recombination could still have a substantial impact in spatially explicit contexts, and whether fusions between genetically dissimilar cells should have a higher impact on diversification. Using spatial agent-based simulations (**Fig. 12 and Mathematical Supplement**), we compared diversification between spatially restricted populations with evolving through mutations and cell fusions versus those evolving through mutations only, in 2D and 3D (**Fig. 5G and Videos S6-9**). We found that diversity was still acquired faster in the presence of fusion-mediated recombination and that the impact of cell fusions was substantially higher in 3D compared to 2D, reflecting higher number of neighbors that increase the probability to have a genetically distinct neighbor (**Fig. 5H**). In summary, despite the relatively low frequency of spontaneous somatic cell fusion and the impact of spatial constraints, fusion-mediated recombination can have a

profound impact on somatic evolution, through accelerated diversification of tumor cell populations and generation of rare mutational variants capable of exploring larger swathes of adaptive landscapes (**Fig. 5I**).

DISCUSSION

Our findings reveal a common occurrence of spontaneous cell fusions across multiple experimental cell line models of breast cancers, demonstrate the impact of fusion-mediated recombination diversification of cancer cell genotypes and phenotypes, and suggest that, despite relatively low frequency, this phenomenon can have a significant impact on somatic evolution. These results question the prevalent assumption of the strict asexuality of cancers. Given that somatic evolution has access to powerful diversification mechanisms, including nucleotide-level mutation rates, chromosomal instability, and epigenetic dysregulations, fusion-mediated recombination is unlikely to be essential for the ability of populations of tumor cells to evolve. However, this mechanism could amplify the pre-existing genetic and phenotypic heterogeneity, augmenting the ability of cancer cell populations to access larger areas of adaptive landscapes through exploring different combinations of mutations, created by replication error and other mutational mechanisms, thus augmenting their evolvability.

Spontaneous cell fusions have been implicated in directly increasing the ability of cancer cells to invade, migrate and establish metastases[18,34,35], which underscores the importance of cell fusion for cancer initiation and progression[7-9]. Our results are consistent with these prior findings and support the prior arguments for the potential importance of cell fusions. In addition to fusions between neoplastic cells, multiple prior studies have described the occurrence of spontaneous, clonogenic cell fusions between neoplastic cells and non-neoplastic cells within the tumor microenvironment, including fibroblasts[17,36], endothelia[37] and myeloid[18,35] cells both *in vitro* and *in vivo*. While we observed significant rates of cell fusions between breast carcinoma cells and CAFs (**Fig. S2C**), we failed to derive clonogenic progeny from these heterotypic hybrids. Since non-immortalized primary CAFs remain genetically normal[38], lack of proliferation might reflect the dominant impact of the intact TP53 checkpoint to limit proliferation of polyploid cells[11,39]. Indeed, examination of time lapse microscopy images indicated that, in contrast to hybrids between carcinoma cells that frequently underwent cell division, hybrids involving fusion with primary CAFs either failed to proliferate or stopped proliferation after a single cell division (data not shown). While these results do not exclude the possibility of the formation of proliferatively viable hybrids between cancer and non-cancer cells within the tumor microenvironment, which has been documented in multiple studies[17,35-37], they suggest that spontaneous somatic cell fusions between cancer cells that have lost intact checkpoint mechanisms to preserve genomic integrity, are more likely generating proliferatively viable progeny.

The biological significance of direct phenotypic impact of cell fusions is often met with skepticism due to the low frequency of these events, as well as lack of evidence that cell fusions are required for the acquisition of metastatic and invasive phenotypes. However, our results suggest an additional, indirect impact with potentially profound implications. We posit that spontaneous cell fusions, followed by genetic recombination due to chromosomal instability and partial ploidy reduction, constitute a parasexual recombination mechanism similar to fusion-mediated facultative parasexual recombination observed in pathogenic yeast C. albicans[14]. Whereas genetic and phenotypic diversification that results from cell fusion has been previously proposed to underlie cell transformation and tumor initiation[13,19], our results suggest the importance of fusion-mediated recombination for the ongoing diversification

within tumor cell populations. Notably, observed frequency of clonogenic cell fusion is similar to the frequency of parasexual recombination in yeast species with a facultative (para)sexual recombination life cycle[40], supporting the notion of potential importance of fusion-mediated recombination in somatic evolution. Indeed, *in silico* modeling experiments suggest that observed rates of cell fusions significantly accelerate diversification within tumor cell populations, once an initial diversity is established.

In addition to genetic diversification (**Fig. 3**), we have also observed an increase in phenotypic diversity in hybrid cells (**Fig. 4**). Whereas increased phenotypic variability might simply reflect genetic diversification, a recent theoretical study has suggested that cell fusions between genetically identical cells with distinct phenotypic states can trigger destabilization of gene regulatory networks, increasing phenotypic entropy and enabling cells to reach phenotypic states distinct from those of the two parents[23]. In support of this notion, cell fusions have been associated with nuclear reprogramming[35] and increased phenotypic plasticity in multiple contexts[41], while increased migration and invasiveness of hybrid cells (**Fig. S4**) has been linked with stemness[42]. Therefore, cell fusions might have a profound impact on tumor cell heterogeneity and cancer progression even in the absence of genetic diversification.

At this point, we cannot completely exclude the possibility that our observations of spontaneous cell fusions within *in vitro* and xenograft models might not be relevant to primary human cancers. On the other hand, occurrence of spontaneous cell fusions between xenografted breast carcinoma cells and host stroma has been documented in an animal model[36]. Additionally, spontaneous fusions between cancer cells and myeloid cells have been rigorously described in animal models and human malignancies[18]. Detecting spontaneous cell fusions in primary human cancers is notoriously difficult, as in most cases all of the neoplastic cells descend from the same clonal origin. Even though giant polyploid cells consistent with spontaneous fusions can be observed in many human neoplasms[43,44], the absence of genetic markers makes it currently impossible to discriminate cell fusions from endoreduplication. Strikingly, two case reports have documented occurrence of cancers that combine genotypes of donor and recipient genomes in patients that have received bone marrow transplantations[45,46]. These findings provide direct evidence for potential clinical relevance of spontaneous cell fusions. Unfortunately, these cases are too rare for systematic interrogation. Still, given the abovementioned studies documenting spontaneous cell fusions in human malignancies, we posit that our results warrant the suspension of the notion that cancers are strictly asexual and that we need to develop methodological approaches to either validate or refute clinical relevance of the phenomena. Should fusion-mediated genetic recombination and phenotypic diversification prove to be relevant for human malignancies, they might have profound implication for evolvability of tumor cell populations, which underlies both clinical progression and acquisition of therapy resistance (**Fig. 5I**).

MATERIALS AND METHODS

Cell lines and tissue culture conditions.

Breast cancer cell lines were obtained from the following sources: MDA-MB-231, HCC1937, HS578T, T47D, MDA-MB-453 from ATCC MCF10DCIS.com from Dr. F. Miller (Karmanos Cancer Institute, Detroit, MI), and SUM149PT from Dr. S. Ethier (University of Michigan, Ann Arbor, MI). Identity of the cell lines was confirmed by short tandem repeats (STR) analysis. CAFs were derived from primary tumors and cultured in SUM medium as described before[47]. All cell lines were tested for

mycoplasma infection routinely. Breast cancer cell lines: MDA-MB-231, SUM159PT, MDA-MB-453 were grown in McCoy medium (ThermoScientific) with 10% FBS (Life Technologies); T-47D, HCC1937 were grown in RPMI-1640 medium (ThermoScientific) with 10% FBS and 10 µg/ml human recombinant insulin (ThermoScientific); MCF10DCIS were grown in DMEM/F12 supplemented with 5% horse serum (ThermoScientific), 10 µg/ml human recombinant insulin, 20 ng/ml EGF (PeproTech), 100 ng/ml cholera toxin (ThermoFisher), 5 µg/ml hydrocortisone (Sigma); MCF7, HS578T were grown in DMEM/F12 with 10% FBS and 10 µg/ml human recombinant insulin, SUM149PT were grown in SUM medium (1:1 mix of DMEM/F12 and Human Mammary Epithelial Cell Growth Medium (Sigma), 5% FBS, 5 µg/ml). Fluorescently labelled derivates of carcinoma cell lines and fibroblasts were obtained by lentiviral expression of pLV[exp]-CAG-NLS-GFP-P2A-puro, pLV[exp]-CAG-NLS-mCherry-P2A-puro, pLV[Exp]-CAG>Bsd(ns):P2A:EGFP (custom vectors from VectorBuilder), or mCherry/Luciferase (obtained from Dr. C. Mitsiades, DFCI). For fusion assays, 50/50 mixes of cells expressing differentially expressed fluorescent and antibiotic resistance markers were seeded into 6cm or 10 cm tissue culture dishes (Sarstedt) in SUM media. Following co-culture for three days, cells were harvested and subjected to flow cytometry analysis, or replated for dual antibiotic selection with 10 µg/ml blasticidin and 2.5 µg/ml puromycin for 7-14 days.

Flow cytometry analyses

For the detection of hybrids, cells from monocultures or co-cultures of differentially labelled cells were harvested with 0.25% Trypsin (ThermoFisher) and resuspended in PBS with 0.1 µg/mL DAPI (Sigma). For the negative controls, cells from monocultures were mixed immediately after harvesting and kept on ice. Flow cytometry analyses were performed using MACSQuant VYB cytometer (Milteniy Biotec), with data analyzed by FlowJo software. Average number of collected events was 80,000. For the image-based flow cytometry analyses cells were incubated for 20 min in PBS with 5 µg/ml Hoechst33342 (ThermoFisher), prior to harvest. The analyses were performed with Amnis ImageStream X Mark II imaging flow cytometer (Amnis, Luminex), using IDEAS software (Amnis, Luminex). Average number of collected events was 10,000.

For the DNA content analyses 106 of cells was resuspended in 500 µl of PBS and then 4.5 ml ice cold 70% ethanol was added dropwise. Cells were kept at -20C for at least 2 hours. Then cells were washed in PBS twice and resuspended in 300 µl PBS with 0.1% Triton X-100 (ThermoFisher), 0.2 mg/ml RNAse (Qiagen) and 20 µg/ml PI (ThermoFisher). Flow cytometry analysis was performed using MACSQuant VYB instrument (Milteniy Biotec). For each of the tested pairs of co-cultures and controls, FACS analyses were performed over two or more distinct experiments.

Microscopy studies

Live cell images were acquired with Axioscope microscope with A-Plan 10x/0.25 Ph1 objective and AxioCam ICm1 camera (Zeiss), using ZEN software (Zeiss). Time-lapse videos were generated with IncuCyte live cell imaging system (Sartorius) using ZOOM 10X objective for breast cancer cell mixes and ZOOM 4x objective for breast cancer cell- fibroblast mixes. Images were acquired in red and green fluorescent channels as well as visible light channel every 3 hours for 4-5 days. For immunofluorescent detection of GFP and mCherry in xenograft tumors, formalin fixed, paraffin embedded tumors were cut at 5 microns sections. Deparaffinized tissue slices were blocked in PBS with 10% goat serum for 30 min at room temperature (RT), then incubated at RT for 1 hour with primary antibodies and 1 hour with secondary antibodies and 0.1 µg/mL DAPI (Sigma) with 3x10 min washes after each incubation. Vector TrueVIEW (Vector Labs) autofluorescence quenching reagent was used prior to mounting the slides. Anti dsRed (1:100, Clonetech #632496) was used for the

detection of mCherry, anti-GFP 4B10 (1:100, CST #632496) was used for detection of GFP. Alexa Fluor Plus 647 goat anti-rabbit (1:1000, Invitrogen # A32733), and Alexa Fluor Plus 488 goat anti-mouse (1:1000, Invitrogen #A32723) were used as secondary antibodies.
Confocal immunofluorescent images were acquired with Leica TCS SP5 system with 63x objective (Leica).

Mouse xenograft studies.
To detect fusion *in vivo*, parental GFP and mCherry expressing cells were harvested and mixed at 50/50 ratio in culture media mixed with 50% Matrigel (BD Biosciences). The mixtures were injected into fat pads of 8 weeks old female NSG mice on both flanks, with $1x10^6$ cells in 100 µl volume per injection. Tumor growth was monitored weekly by palpation. When tumor diameter reached 1 cm or animals become moribund with symptoms of reduced mobility, hunching and labored breathing, mice were euthanized. Xenograft tumors were fixed in PFA and paraffin embedded.
For the lung metastasis assays, luciferase-labelled parental cell lines (MCF10DCIS and SUM159PT) or their hybrids at passage 8 post antibiotic selection were harvested and resuspended in DMEM/F12 medium with 10% FBS. $2x10^5$ cells were injected per animal through the tail vein into 8 weeks old NSG mice. Tumor growth was monitored by bioluminescence capture with the IVIS-200 imaging system (PerkinElmer) under isoflurane anesthesia. Imaging was performed immediately after injection and then weekly. One month after injection, mice were euthanized, and lungs were extracted and fixed in PFA and paraffin-embedded. Tissue slices were stained with hematoxylin and eosin (H&E) and scanned with Aperio ScanScope XT Slide Scanner (Leica). Metastasis number and area were annotated and analyzed using QuPath software (https://qupath.github.io/). All animal experiments were performed in accordance with the guidelines of the IACUC of the H. Lee Moffitt Cancer Center.

Growth and Invasion/migration assays.
To determine growth rates, $5x10^4$ cells were seeded in triplicates into 6 cm culture dishes. Upon reaching ~90% confluency, cells were harvested by trypsinization, counted and re-seeded at the same starting density. Growth rates were calculated as average ln(cell number fold change)/(number of days in culture) over three passages. Invasion/migration assay was performed in 12 well ThinCert plates with 8 µm pore inserts (Greiner Bio-One, #665638). Parental cells and hybrids were plated in appropriate FBS-free medium containing 0.1% BSA (Sigma) on transwell insert membrane covered with 30% Matrigel (BD Biosciences) in PBS. Lower well contained medium with 10% FBS. 10^4 cells were plated and cultured for 72 h. Matrigel was removed from the transwell insert and cells migrated through the membrane fixed in 100% methanol for 10 min on ice and stained with 0.5% crystal violet solution in 25% methanol. Membranes were cut off and mounted on glass slides with xylene based Cytoseal mounting media (ThermoFisher). Slides were scanned Evos Auto imaging system (ThermoFisher) and images were analyzed with ImageJ software (https://fiji.sc/). 16 tiles per sample were analyzed.

Colony formation assays.
For *in vitro* clonogenic assays, 50/50 mixes of GFP and mCherry expressing parental cell lines were grown separately (controls) or in co-cultures for 3 days. For *ex vivo* clonogenic assays, tumors initiated from implantation of single labelled cells or 50/50 mixes were harvested 4 weeks post implantation, and digested with the mixture of 2 mg/ml collagenase I (Worthington Biochem) and 2 mg/ml hyalurinidase (Sigma H3506) at 37°C for 3 hours. $1x10^6$ cells were seeded in 10 cm culture dishes in corresponding growth media with 10 µg/ml blasticidin and 2.5 µg/ml puromycin; for negative controls, $5x10^5$ of each of the parental cells were seeded. For clonogenicity rate controls, parental cells were seeded at 100 cells per 6 cm dish in corresponding antibiotic. Following 2 weeks of incubation, with selection media replaced

twice per week, and acquisition of fluorescent images of representative colonies, media was removed, plates were washed twice with PBS, then colonies were fixed in a solution of 12.5% acetic acid and 30% methanol for 15 min, and stained with 0.1% crystal violet solution in water for 4 h. Colonies with approximately 50 or more cells were manually counted, and the percentage of plated cells that formed colonies was determined. In order to account for reduced clonogenic survival of cells, freshly isolated from xenograft tumors, clonogenic data from *ex vivo* isolates were normalized to clonogenicity in the absence of antibioitic selection.

SNP and copy number analyses.
DNA from ~$3x10^6$ cells was extracted with DNeasy Blood&Tissue kit (Qiagen). For SUM159PT/MDA-MB-231 hybrids CytoSNP-12 v2.1 BeadChip array from Illumina was used and data analyzed with GenomeStudio 2.0 software (Illumina). For MCF10DCIS/SUM159PT and HS578T/MDA-MB-231 hybrids CytoScan array from Affimetrix was used. SNP data analyzed and LogR Ratio plots were generated with ChAS software (Affimetrix). Data were analyzed and visualized using R (version 3.3.2) code. First, genotypes of parental cell lines were compared and differential homozygous SNPs were selected. Then, for each of the differential SNP an identity score was assigned for a hybrid sample: 0 if SNP genotype was homozygous and the same as parent #1, 1 if SNP genotype was homozygous and the same as parent #2 and 0.5 if SNP genotype was heterozygous. The data were plotted as a heatmap where rows represent hybrid samples and columns are aligned by position number of each SNP, while color reflects the genotype identity score.

Single cell RNA sequencing and analysis.
Chromium Single Cell 3' Library, Gel Bead & Multiplex Kit and Chip Kit (10X Genomics) was used to encapsulate and barcode for a target of 10,000 single cells total for cDNA preparation of parental (SUM159PT and MDA-MB-231) and passaged fusion cells (passage 2 and 10). Targeted cell population sizes were 2500 cells for each parental SUM159PT and MDA-MB-231 lines, as well as 2500 cells for each passage 2 and passage 10 of hybrid SUM159PT/MDA-MB-231. Libraries were constructed according to manufacturer's protocol, sequenced on an Illumina NovaSeq, and mapped to the human genome (GRCh38) using CellRanger (10X Genomics) with an extended reference to include GFP and mCherry proteins that the parental lines were engineered to expressed.

Raw gene expression matrices generated per sample using CellRanger (version 3.0.1) were combined in R (version 3.5.2), and converted to a Seurat object using the Seurat R package (https://satijalab.org/seurat/) [48]. From this, all cells were removed that had over two-standard deviations of the mean UMIs derived from the mitochondrial genome (between 5-12% depending on sample) and only kept cells with gene expression within two-standard deviations of the mean gene expression [49]. From the remaining 10,059 gene expression matrices were log-normalized and scaled to remove variation due to total cellular read counts. To reduce dimensionality of this dataset, the first 200 principal components were calculated based on the top 2000 variable genes. Jack Straw analysis confirmed that the majority of the dataset variation were captured in these first principal components. All the cells were then clustered using the Louvain algorithm implemented by Seurat by creating a graph representation of the dataset with 10,059 nodes and 413,785 edges and optimizing based on modularity. With a resolution of 0.6, the algorithm identified 12 communities with a maximum modularity of 0.8931, which gave us confidence in this clustering [50]. The data was then visualized by running the uniform manifold approximation and projection (UMAP) algorithm [24,51] and either coloring individual cells based on their cluster identity or cell type identity. Gini coefficients were calculated with the package DescTools in R [52]. Zero values were

excluded, and reads were used from 10x genomics outputs.

Statistical analyses.
Statistical analyses of *in vitro* and *in vivo* experimental data were performed using GraphPad Prism and Matlab software, using statistical tests indicated in figure legends.

Computational methods.
Muller's plots were obtained using the RTool Evofreq[53]. The agent-based model was implemented in the JAVA framework HAL[54]. To visualize cells carrying unique genotypes (results presented in Fig. 5G and supplementary videos S6-9), the scalar value of the genotype binary vector value in base 2 was mapped to rainbow color scale. Detailed description of mathematical modeling is provided in the Mathematical Supplement.

ACKNOWLEDGMENTS

We thank Dr. Aaron Goldman for sharing his unpublished observations of frequent occurrence of cell fusions in his experimental studies, which confirmed our observations and helped motivating this work. We thank Dr. Yuri Lazebnik and Dr. Aaron Goldman for providing thoughtful feedback. We thank Rafael Bravo for the help he provided for using the framework HAL We thank the Flow Cytometry, Analytic Microscopy, Tissue Histology, Biostatistics and Bioinformatics Shared Resource and Molecular Genomic Core Facilities at the H. Lee Moffitt Cancer Center & Research Institute; an NCI designated Comprehensive Cancer Center (P30-CA076292). We thank Moffitt SPARK program for supporting internship for MAL. This work was supported by Susan G. Komen Breast Cancer Foundation CCR17481976 (A.M.), Moffitt Cancer Biology and Evolution program pilot award (AM), and Integrative Mathematical Oncology workshop award (AM and DB).

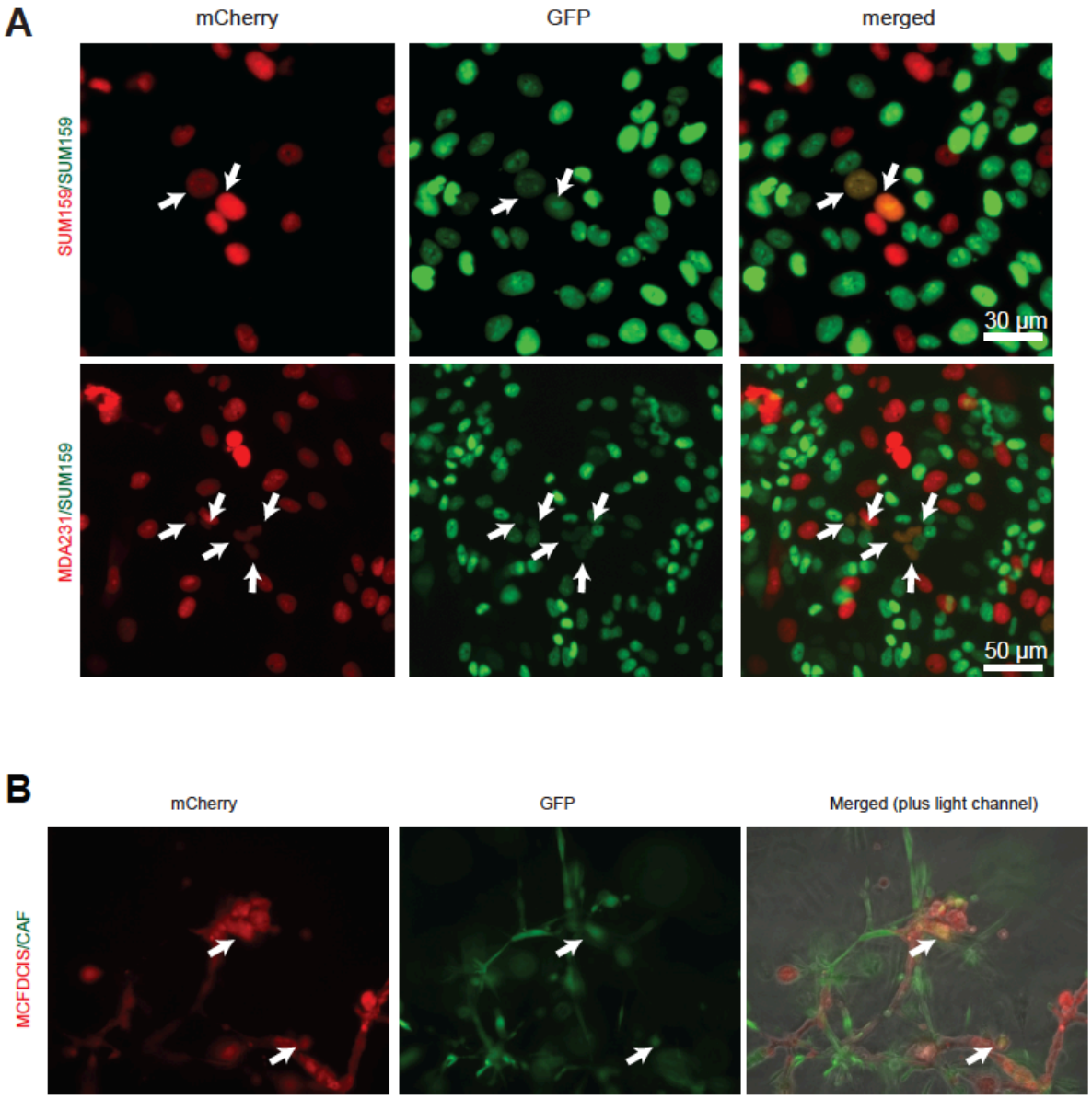
CITED REFERENCES

1 Greaves, M. & Maley, C. C. Clonal evolution in cancer. *Nature* **481**, 306-313, doi:10.1038/nature10762 (2012).
2 Scott, J. & Marusyk, A. Somatic clonal evolution: A selection-centric perspective. *Biochim Biophys Acta*, doi:10.1016/j.bbcan.2017.01.006 (2017).
3 Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646-674, doi:10.1016/j.cell.2011.02.013 (2011).
4 Becks, L. & Agrawal, A. F. The evolution of sex is favoured during adaptation to new environments. *PLoS biology* **10**, e1001317, doi:10.1371/journal.pbio.1001317 (2012).
5 McDonald, M. J., Rice, D. P. & Desai, M. M. Sex speeds adaptation by altering the dynamics of molecular evolution. *Nature* **531**, 233-236, doi:10.1038/nature17143 (2016).
6 Sinai, S., Olejarz, J., Neagu, I. A. & Nowak, M. A. Primordial sex facilitates the emergence of evolution. *J R Soc Interface* **15**, doi:10.1098/rsif.2018.0003 (2018).
7 Duelli, D. & Lazebnik, Y. Cell fusion: a hidden enemy? *Cancer cell* **3**, 445-448 (2003).
8 Lu, X. & Kang, Y. Cell fusion as a hidden force in tumor progression. *Cancer research* **69**, 8536-8539, doi:10.1158/0008-5472.CAN-09-2159 (2009).
9 Platt, J. L., Zhou, X., Lefferts, A. R. & Cascalho, M. Cell Fusion in the War on Cancer: A Perspective on the Inception of Malignancy. *Int J Mol Sci* **17**, doi:10.3390/ijms17071118 (2016).
10 Kuznetsova, A. Y. *et al.* Chromosomal instability, tolerance of mitotic errors and multidrug resistance are promoted by tetraploidization in human cells. *Cell Cycle* **14**, 2810-2820, doi:10.1080/15384101.2015.1068482 (2015).
11 Fujiwara, T. *et al.* Cytokinesis failure generating tetraploids promotes tumorigenesis in p53-null cells. *Nature* **437**, 1043-1047, doi:10.1038/nature04217 (2005).
12 Su, Y. *et al.* Somatic Cell Fusions Reveal Extensive Heterogeneity in Basal-like Breast Cancer. *Cell Rep* **11**, 1549-1563, doi:10.1016/j.celrep.2015.05.011 (2015).
13 Zhou, X. *et al.* Cell Fusion Connects Oncogenesis with Tumor Evolution. *Am J Pathol* **185**, 2049-2060, doi:10.1016/j.ajpath.2015.03.014 (2015).

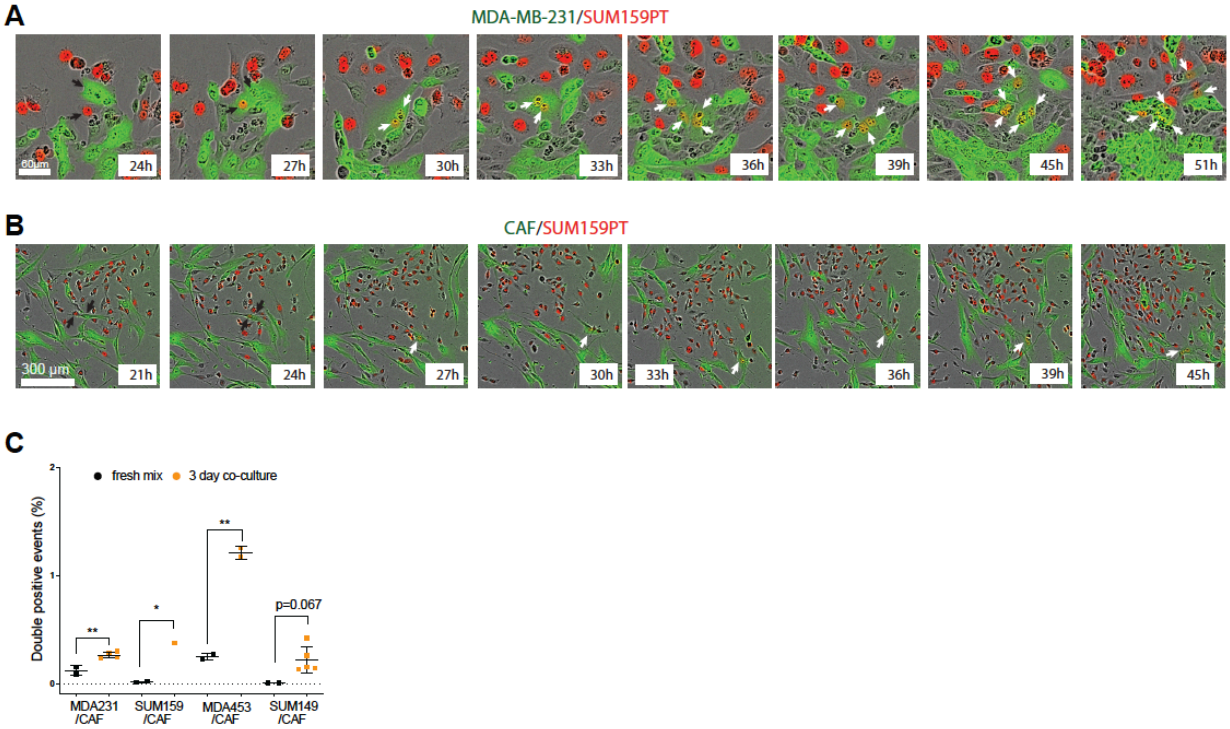
14 Bennett, R. J. The parasexual lifestyle of Candida albicans. *Curr Opin Microbiol* **28**, 10-17, doi:10.1016/j.mib.2015.06.017 (2015).
15 Zuba-Surma, E. K., Kucia, M., Abdel-Latif, A., Lillard, J. W., Jr. & Ratajczak, M. Z. The ImageStream System: a key step to a new era in imaging. *Folia Histochem Cytobiol* **45**, 279-290 (2007).
16 Fais, S. & Overholtzer, M. Cell-in-cell phenomena in cancer. *Nature reviews. Cancer* **18**, 758-766, doi:10.1038/s41568-018-0073-9 (2018).
17 Rappa, G., Mercapide, J. & Lorico, A. Spontaneous formation of tumorigenic hybrids between breast cancer and multipotent stromal cells is a source of tumor heterogeneity. *Am J Pathol* **180**, 2504-2515, doi:10.1016/j.ajpath.2012.02.020 (2012).
18 Gast, C. E. *et al.* Cell fusion potentiates tumor heterogeneity and reveals circulating hybrid cells that correlate with stage and survival. *Sci Adv* **4**, eaat7828, doi:10.1126/sciadv.aat7828 (2018).
19 Duelli, D. M. *et al.* A virus causes cancer by inducing massive chromosomal instability through cell fusion. *Curr Biol* **17**, 431-437, doi:10.1016/j.cub.2007.01.049 (2007).
20 Forche, A. *et al.* The parasexual cycle in Candida albicans provides an alternative pathway to meiosis for the formation of recombinant strains. *PLoS biology* **6**, e110, doi:10.1371/journal.pbio.0060110 (2008).
21 Duncan, A. W. *et al.* The ploidy conveyor of mature hepatocytes as a source of genetic variation. *Nature* **467**, 707-710, doi:10.1038/nature09414 (2010).
22 Skinner, A. M., Grompe, M. & Kurre, P. Intra-hematopoietic cell fusion as a source of somatic variation in the hematopoietic system. *J Cell Sci* **125**, 2837-2843, doi:10.1242/jcs.100123 (2012).
23 Koulakov, A. A. & Lazebnik, Y. The problem of colliding networks and its relation to cell fusion and cancer. *Biophys J* **103**, 2011-2020, doi:10.1016/j.bpj.2012.08.062 (2012).
24 Becht, E. *et al.* Dimensionality reduction for visualizing single-cell data using UMAP. *Nat Biotechnol*, doi:10.1038/nbt.4314 (2018).
25 Ferrall-Fairbanks, M. C., Ball, M., Padron, E. & Altrock, P. M. Leveraging Single-Cell RNA Sequencing Experiments to Model Intratumor Heterogeneity. *JCO Clin Cancer Inform* **3**, 1-10, doi:10.1200/CCI.18.00074 (2019).
26 Hill, M. O. Diversity and Evenness: A Unifying Notation and Its Consequences. *Ecology* **54**, 427-432, doi:10.2307/1934352 (1973).
27 Rosenzweig, M. L., L, R. M. & Press, C. U. *Species Diversity in Space and Time*. (Cambridge University Press, 1995).
28 Hinohara, K. *et al.* KDM5 Histone Demethylase Activity Links Cellular Transcriptomic Heterogeneity to Therapeutic Resistance. *Cancer cell* **35**, 330-332, doi:10.1016/j.ccell.2019.01.012 (2019).
29 Loeb, L. A. Human Cancers Express a Mutator Phenotype: Hypothesis, Origin, and Consequences. *Cancer research* **76**, 2057-2059, doi:10.1158/0008-5472.CAN-16-0794 (2016).
30 Waclaw, B. *et al.* A spatial model predicts that dispersal and cell turnover limit intratumour heterogeneity. *Nature* **525**, 261-264, doi:10.1038/nature14971 (2015).
31 Kimmel, G. J., Gerlee, P. & Altrock, P. M. Time scales and wave formation in non-linear spatial public goods games. *PLoS Comput Biol* **15**, e1007361, doi:10.1371/journal.pcbi.1007361 (2019).
32 Gallaher, J. A., Enriquez-Navas, P. M., Luddy, K. A., Gatenby, R. A. & Anderson, A. R. A. Spatial Heterogeneity and Evolutionary Dynamics Modulate Time to Recurrence in Continuous and Adaptive Cancer Therapies. *Cancer research* **78**, 2127-2139, doi:10.1158/0008-5472.CAN-17-2649 (2018).
33 Noble, R., Burri, D., Kather, J. N. & Beerenwinkel, N. Spatial structure governs the mode of tumour evolution. *bioRxiv*, 586735, doi:10.1101/586735 (2019).
34 Lu, X. & Kang, Y. Efficient acquisition of dual metastasis organotropism to bone and lung through stable spontaneous fusion between MDA-MB-231 variants. *Proc Natl Acad Sci U S A* **106**, 9385-9390, doi:10.1073/pnas.0900108106 (2009).
35 Powell, A. E. *et al.* Fusion between Intestinal epithelial cells and macrophages in a cancer context results in nuclear reprogramming. *Cancer research* **71**, 1497-1505, doi:10.1158/0008-5472.CAN-10-3223 (2011).
36 Jacobsen, B. M. *et al.* Spontaneous fusion with, and transformation of mouse stroma by, malignant human breast cancer epithelium. *Cancer research* **66**, 8274-8279, doi:10.1158/0008-5472.CAN-06-1456 (2006).
37 Mortensen, K., Lichtenberg, J., Thomsen, P. D. & Larsson, L. I. Spontaneous fusion between cancer cells and endothelial cells. *Cell Mol Life Sci* **61**, 2125-2131, doi:10.1007/s00018-004-4200-2 (2004).

38 Qiu, W. *et al.* No evidence of clonal somatic genetic alterations in cancer-associated fibroblasts from human breast and ovarian carcinomas. *Nat Genet* **40**, 650-655, doi:10.1038/ng.117 (2008).
39 Vitale, I. *et al.* Multipolar mitosis of tetraploid cells: inhibition by p53 and dependency on Mos. *EMBO J* **29**, 1272-1284, doi:10.1038/emboj.2010.11 (2010).
40 Nieuwenhuis, B. P. & James, T. Y. The frequency of sex in fungi. *Philos Trans R Soc Lond B Biol Sci* **371**, doi:10.1098/rstb.2015.0540 (2016).
41 Lazebnik, Y. The shock of being united and symphiliosis. Another lesson from plants? *Cell Cycle* **13**, 2323-2329, doi:10.4161/cc.29704 (2014).
42 Mani, S. A. *et al.* The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell* **133**, 704-715, doi:10.1016/j.cell.2008.03.027 (2008).
43 Amend, S. R. *et al.* Polyploid giant cancer cells: Unrecognized actuators of tumorigenesis, metastasis, and resistance. *Prostate* **79**, 1489-1497, doi:10.1002/pros.23877 (2019).
44 Islam, S. *et al.* Drug-induced aneuploidy and polyploidy is a mechanism of disease relapse in MYC/BCL2-addicted diffuse large B-cell lymphoma. *Oncotarget* **9**, 35875-35890, doi:10.18632/oncotarget.26251 (2018).
45 Lazova, R. *et al.* A Melanoma Brain Metastasis with a Donor-Patient Hybrid Genome following Bone Marrow Transplantation: First Evidence for Fusion in Human Cancer. *PloS one* **8**, e66731, doi:10.1371/journal.pone.0066731 (2013).
46 LaBerge, G. S., Duvall, E., Grasmick, Z., Haedicke, K. & Pawelek, J. A Melanoma Lymph Node Metastasis with a Donor-Patient Hybrid Genome following Bone Marrow Transplantation: A Second Case of Leucocyte-Tumor Cell Hybridization in Cancer Metastasis. *PloS one* **12**, e0168581, doi:10.1371/journal.pone.0168581 (2017).
47 Marusyk, A. *et al.* Spatial Proximity to Fibroblasts Impacts Molecular Features and Therapeutic Sensitivity of Breast Cancer Cells Influencing Clinical Outcomes. *Cancer research* **76**, 6495-6506, doi:10.1158/0008-5472.CAN-16-1457 (2016).
48 Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* **36**, 411-420, doi:10.1038/nbt.4096 (2018).
49 AlJanahi, A. A., Danielsen, M. & Dunbar, C. E. An Introduction to the Analysis of Single-Cell RNA-Sequencing Data. *Mol Ther Methods Clin Dev* **10**, 189-196, doi:10.1016/j.omtm.2018.07.003 (2018).
50 Newman, M. E. Modularity and community structure in networks. *Proc Natl Acad Sci U S A* **103**, 8577-8582, doi:10.1073/pnas.0601602103 (2006).
51 McInnes, L., Healy, J., Saul, N. & Großberger, L. UMAP: Uniform Manifold Approximation and Projection. *Journal of Open Source Software* **3**, doi:10.21105/joss.00861 (2018).
52 Signorell, A. DescTools: Tools for descriptive statistics. *R package version 0.99* **18** (2016).
53 Gatenbee, C. D., Schenck, R. O., Bravo, R. R. & Anderson, A. R. A. EvoFreq: visualization of the Evolutionary Frequencies of sequence and model data. *BMC Bioinformatics* **20**, 710, doi:10.1186/s12859-019-3173-y (2019).
54 Bravo, R. R. *et al.* Hybrid Automata Library: A flexible platform for hybrid modeling with real-time visualization. *PLoS Comput Biol* **16**, e1007635, doi:10.1371/journal.pcbi.1007635 (2020).

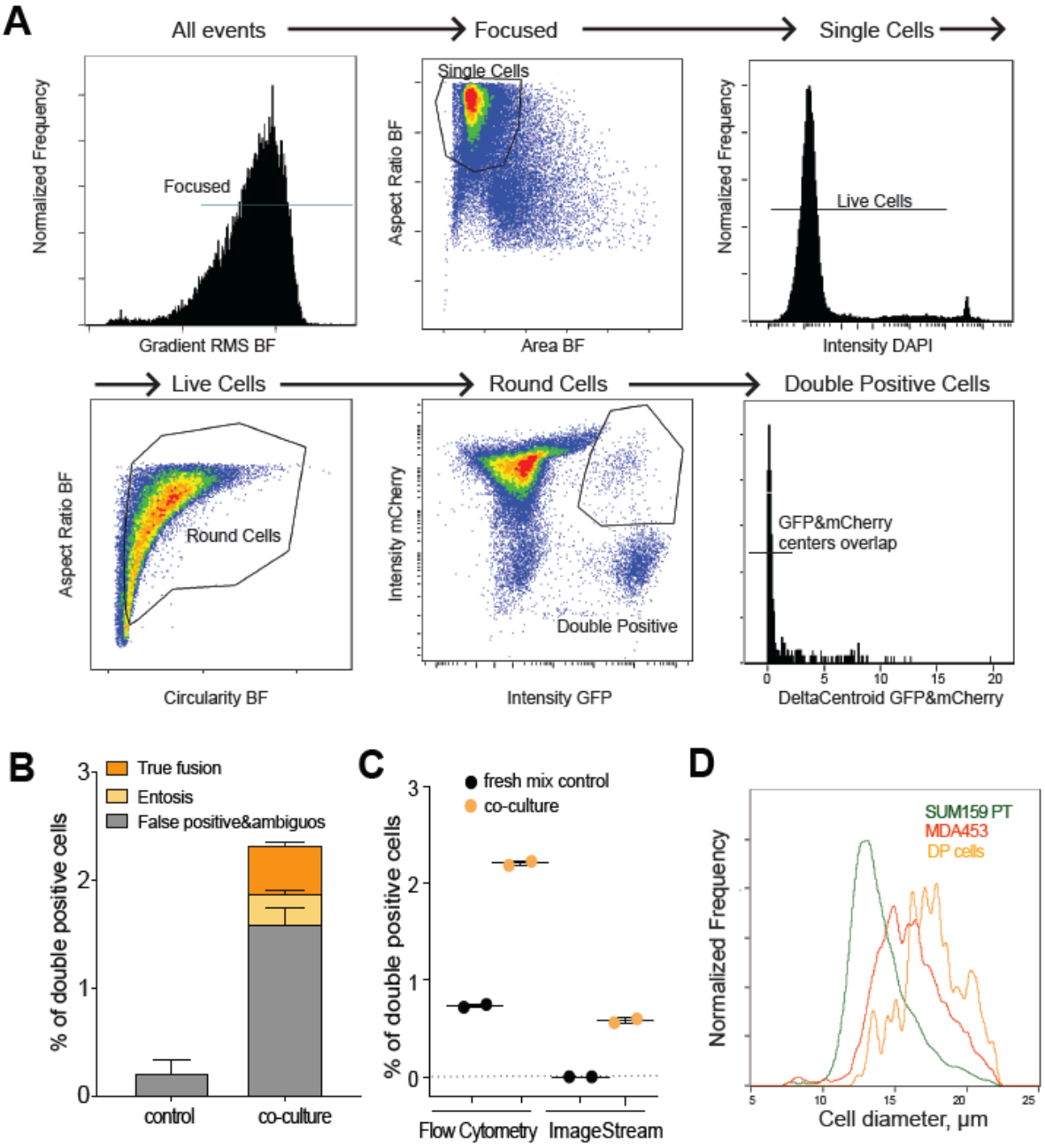
Supplementary materials



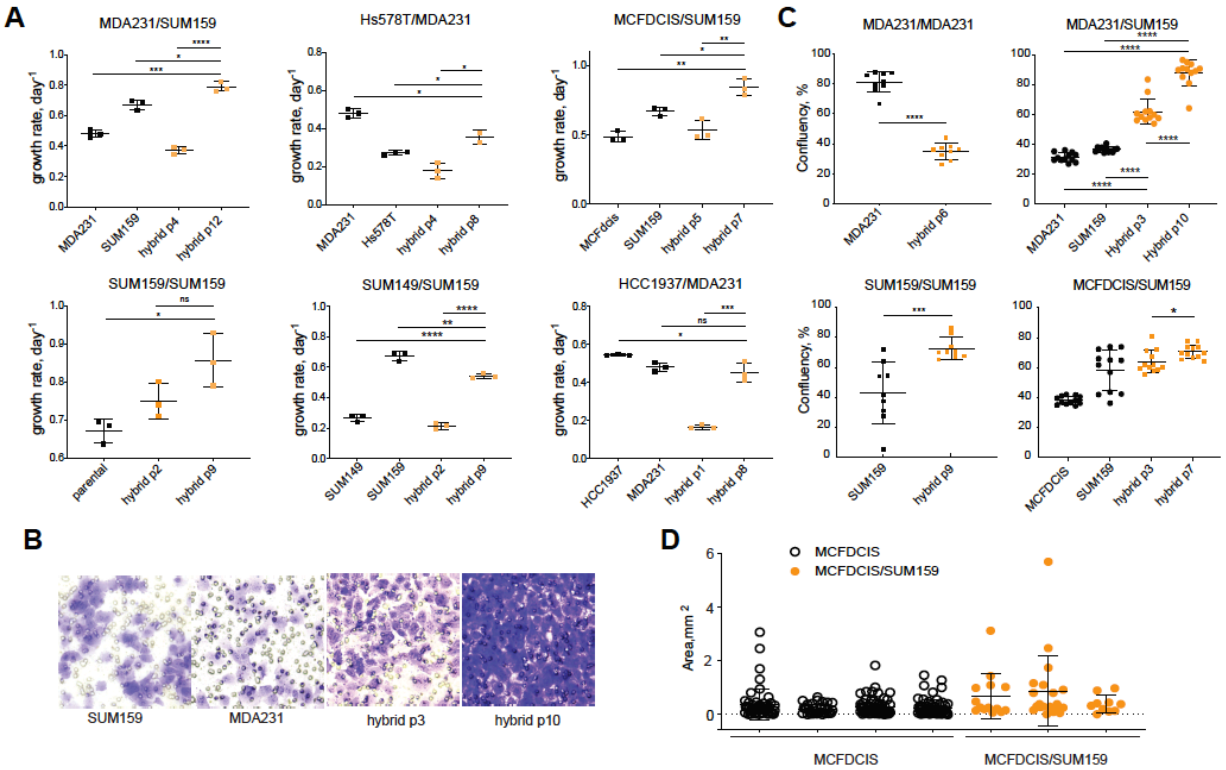
S1. A. **Detection of putative spontaneous cell fusions**. Live cell fluorescent microscopy images of 2d co-cultures between differentially labelled (nuclear GFP and mCherry) cell lines. Arrowheads indicate cells that express both labels. **B.** Live-cell fluorescent microscopy images of 3D Matrigel co-cultures between differentially labelled (cytoplasmic GFP and dsRED) MCF10DCIS breast carcinoma cells and primary cancer associated fibroblasts isolate.



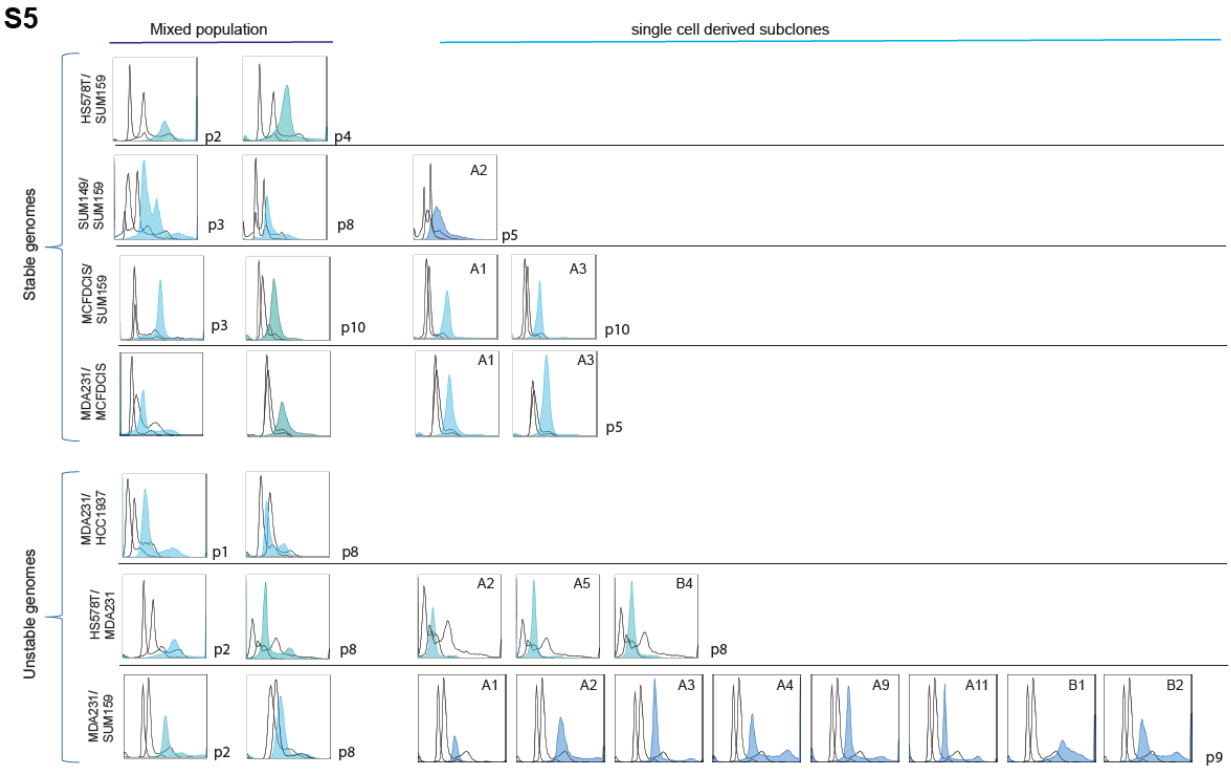
S2. Time-lapse live fluorescent microscopy images of mCherry+ SUM159PT cells cultured with GFP+ MDA-MB-231 (**A**) and CAF (**B**). The labels indicate time after plating. Black arrowheads show fusion parents, white arrowheads show double-positive hybrid cells after fusion and after cell division. **C.** Quantification of flow cytometry detection of double-positive events in the co-cultures between GFP+ CAFs and indicated breast cancer cell lines labelled with mCherry.
*and ** denote $p<0.05$ and $p<0.01$of 2-tailed unpaired t-test respectively.



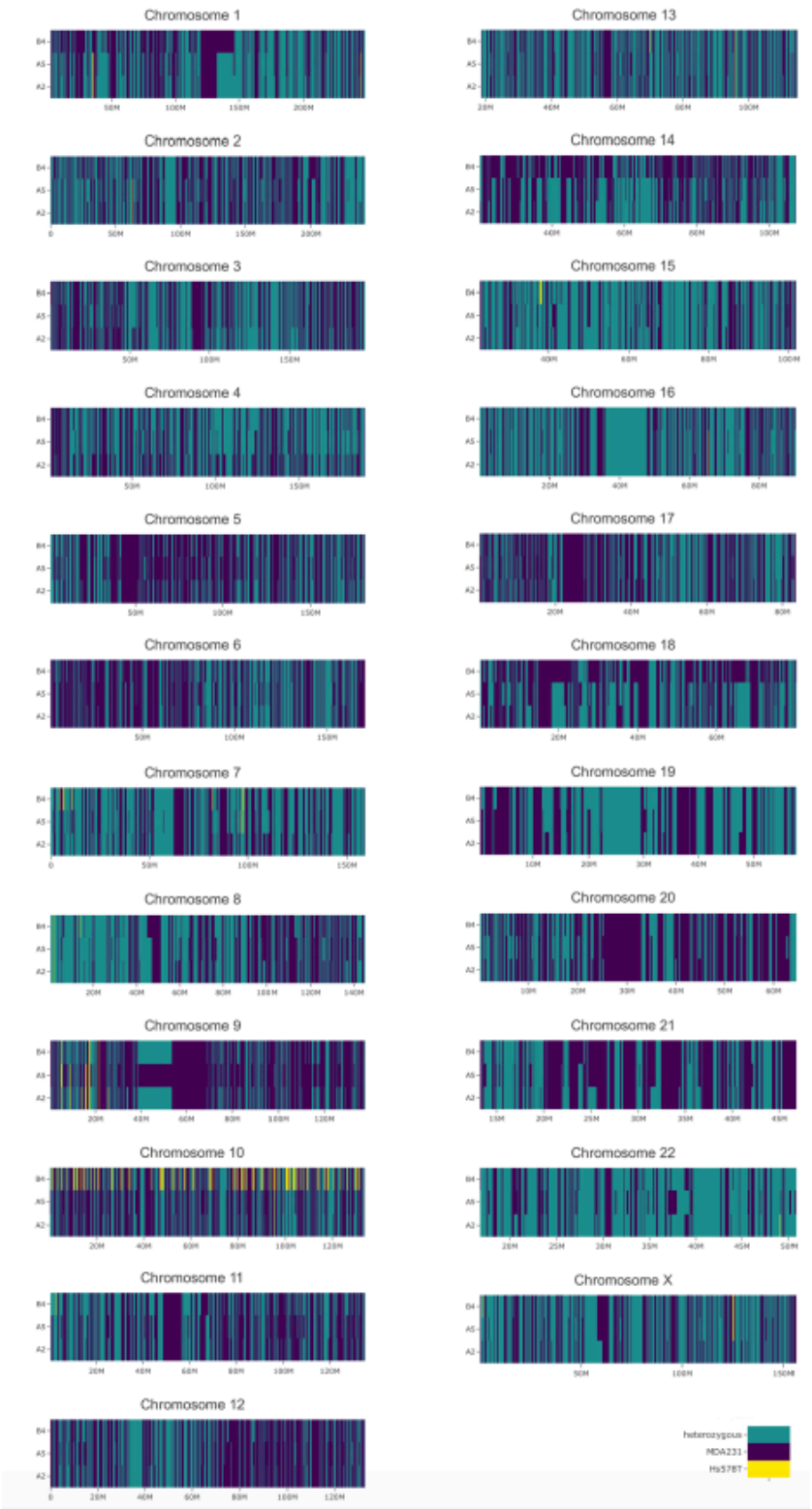
S3. Image Stream detection of spontaneous cell fusions. **A.** Gating strategy for the detection of double-positive cells with Image Stream imaging flow cytometry platform. **B.** Quantitation of different classes of double-positive events in 3-day co-cultures between 50/50 mixes of GFP/mCherry labelled MCF7 cells, with examples of different classes of events provided in **Fig. 1E**. **C**. Comparison of frequency of double-positive events detected from the same samples of co-cultures of differentially labelled MDA-MB-231 (mCherry+) and SUM159PT (GFP+) cells and freshly mixed controls using FACS and Image Stream platforms (validated true positives percentages are plotted for Image Stream analyses). **D.** Distribution of cell diameters of the parental and double-positive cells from Image stream data shown in (**C**) measured in bright field and plotted using IDEAS software (ImageStream).



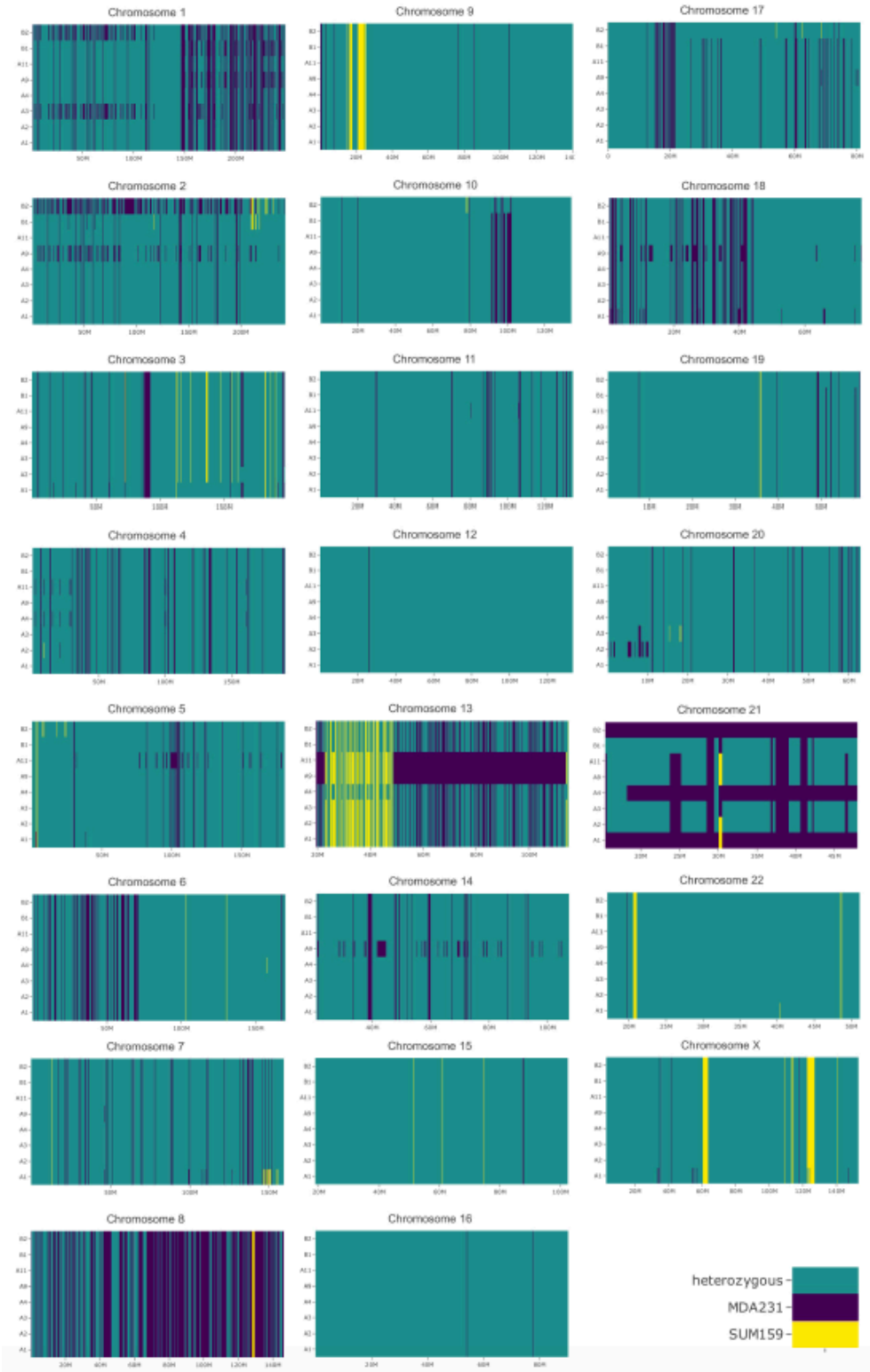
S4. Phenotypic characterization of hybrid cells. A. Growth rates of the indicated cell lines and their hybrids, at the indicated passages post-antibiotic-selection. **B.** Representative images of stained membranes from Boyden chamber cell invasion/migration assay. **C**. Quantitation of Boyden chamber cell invasion/migration assay data. **D.** Quantification of area of lung metastases formed after tail vein injection of MCF10DCIS and MCF10DCIS/SUM159PT hybrids. Data from individual mice are plotted separately. *, **, ***, *** denote p values below 0.05, 0.01, 0.001 and 0.0001, respectively, for the to-tailed unpaired t-test.



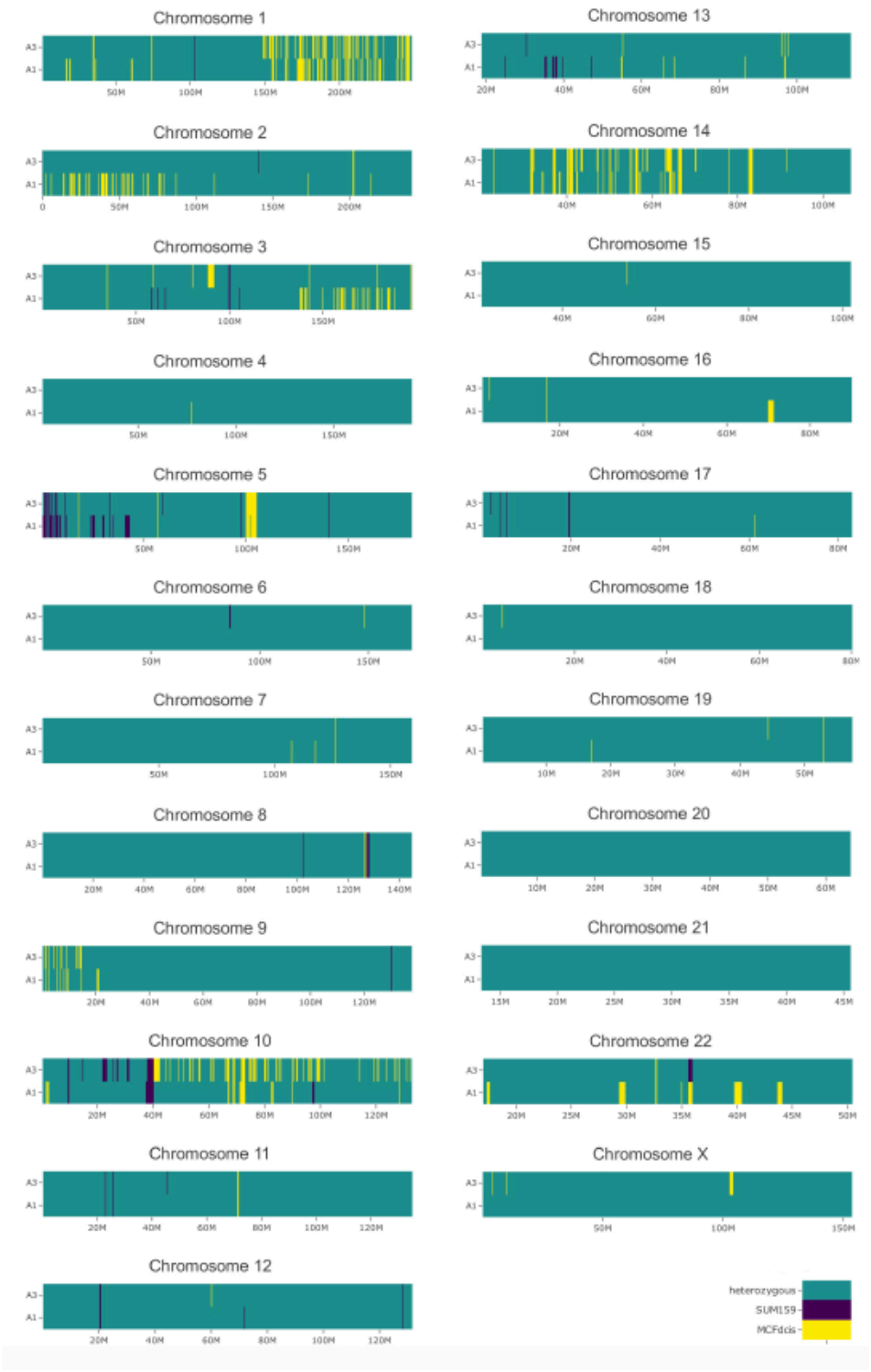
S5 DNA content analysis of hybrid cells. FACS analysis of DNA content of the indicated parental cell lines and their hybrids at indicated passages post-antibiotic-selection. Black contours indicate DNA content profiles for parental cell lines, filled histograms indicate DNA content profiles of the hybrids.



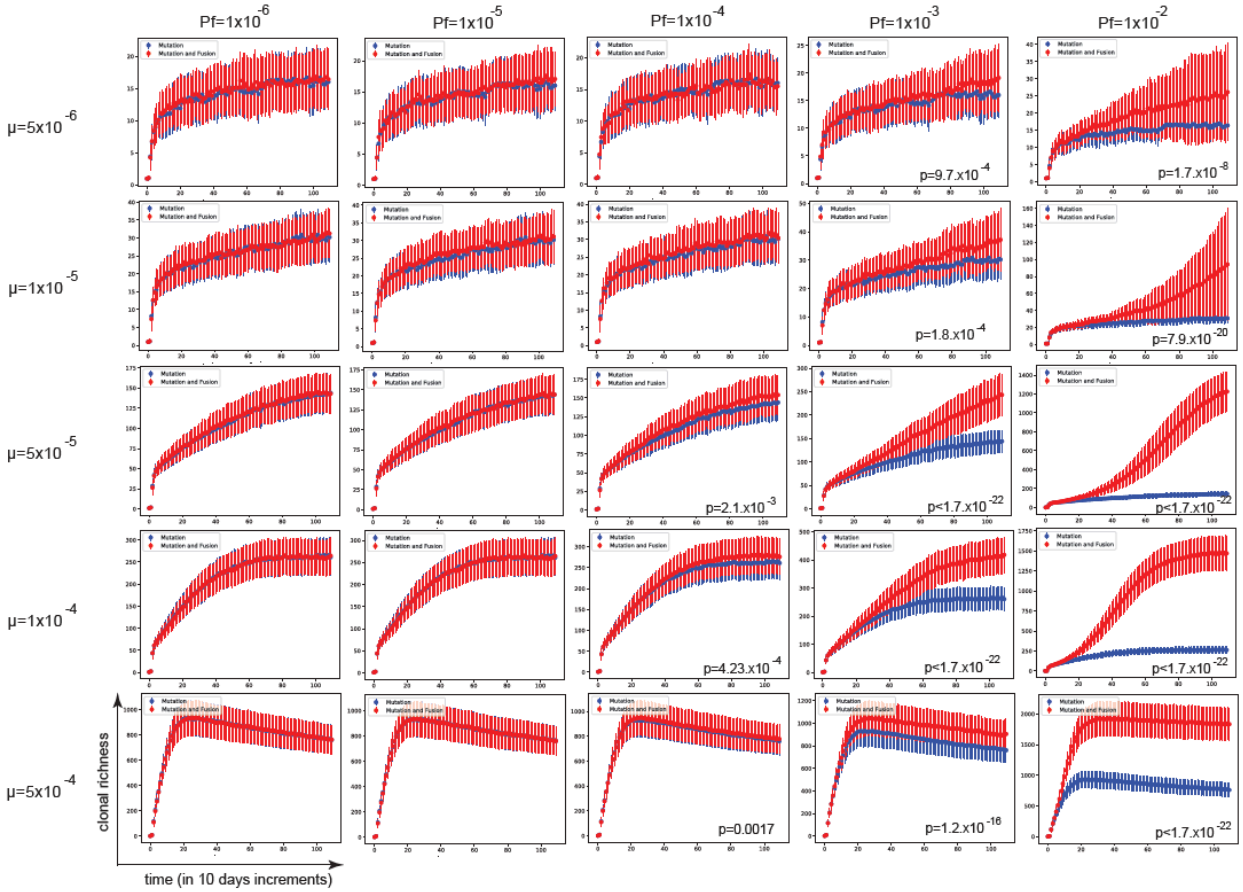
S6. Allelic inheritance analyses of cell line specific alleles for HS578T/MDA-MB-231 hybrids. From the analyses of Affymetrix CytoScan SNP array. Each chromosome is depicted in a separate panel. Rows depict distinct sub-clones shown in Fig. S5; columns indicate parent-specific alleles, as described in the color key. color code indicates



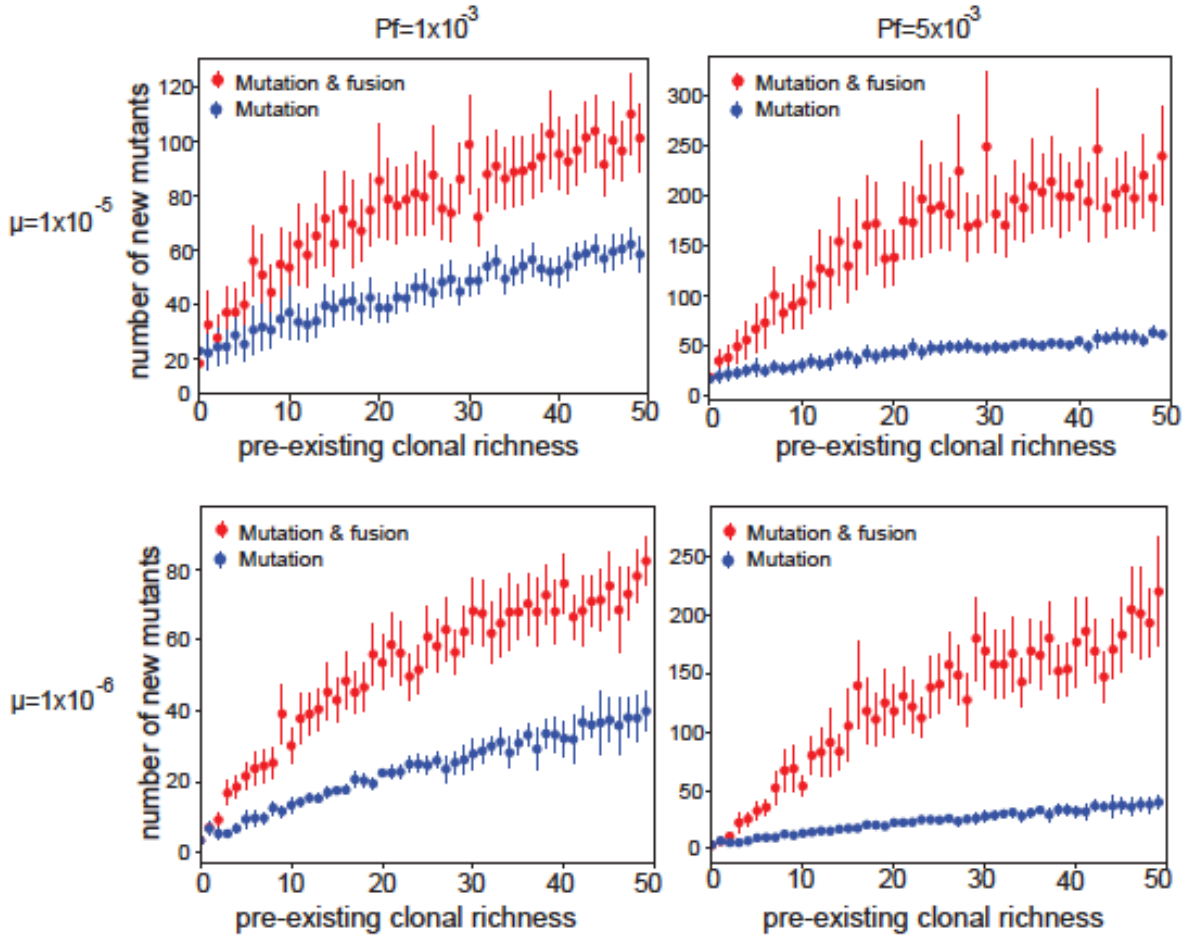
S7. Allelic inheritance analyses of cell line specific alleles for SUM159PT/MDA-MB-231 hybrids. From the analyses using Illumina CytoSPN-12 platform. Each chromosome is depicted in a separate panel. Rows depict distinct sub-clones shown in Fig. S5; columns indicate parent-specific alleles, as described in the color key.



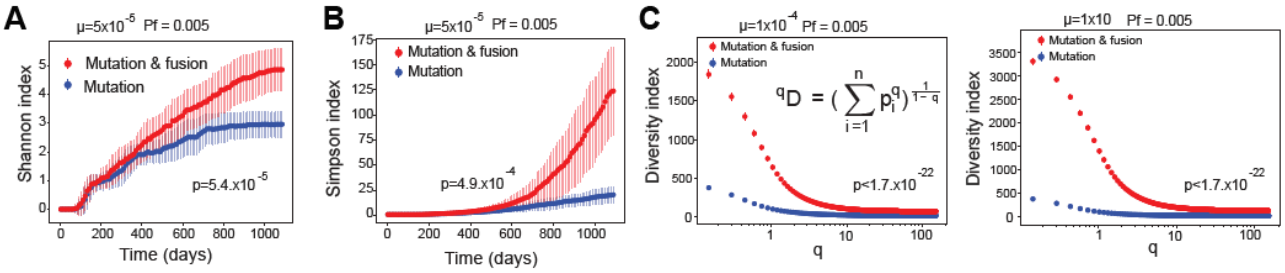
S8. Allelic inheritance analyses of cell line specific alleles for MCF10DCIS/SUM159PT hybrids. From the analyses using Illumina CytoSPN-12 platform. Each chromosome is depicted in a separate panel. Rows depict distinct sub-clones shown in Fig. S5; columns indicate parent-specific alleles, as described in the color key.



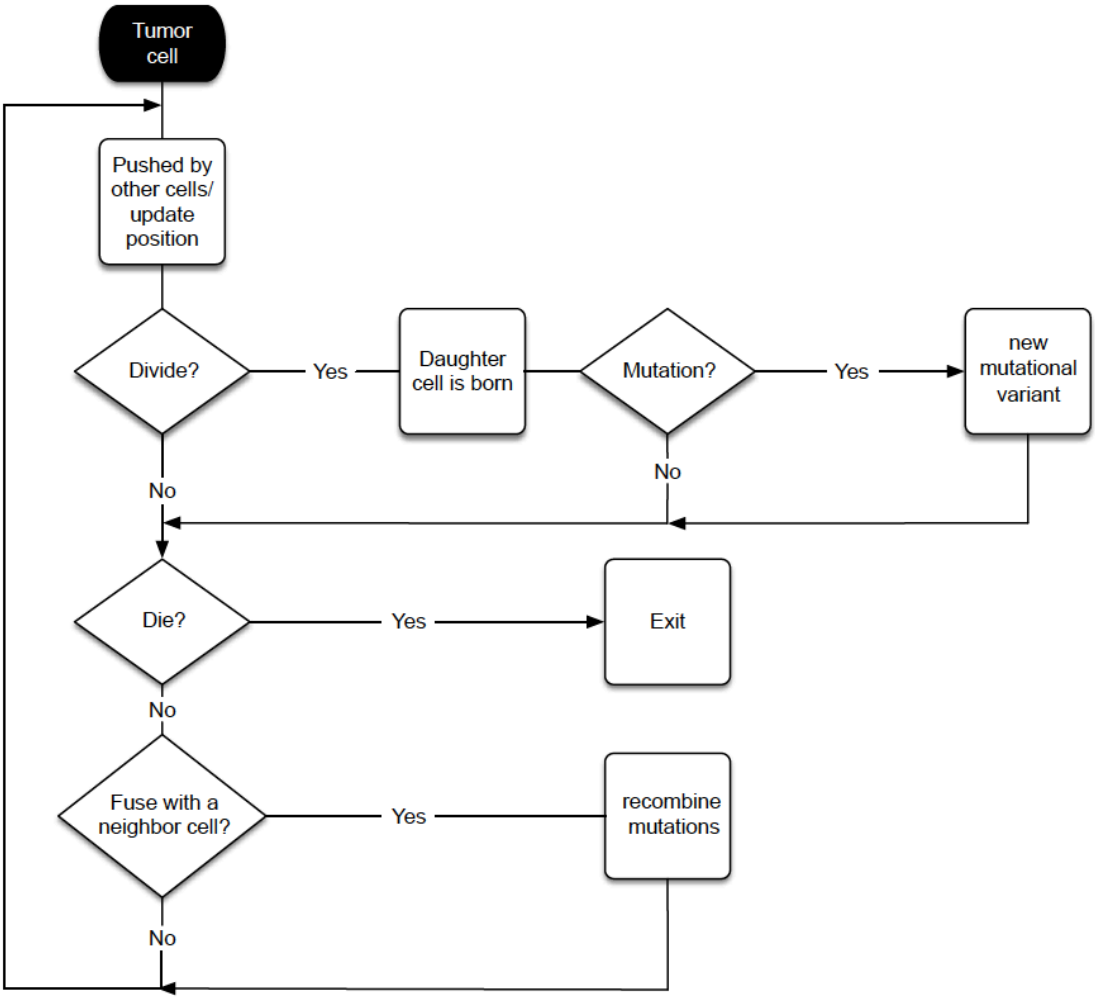
S9. **Parameter sweep for the impact of mutation and fusion rates on clonal richness**. Graphs depict results of *in silico* simulations with branching birth-death model showing clonal richness over time for the indicated mutation (μ) and fusion (Pf) rates; p values indicate the results of Kolmogorov-Smirnov test (only shown for differences that has reached the 0.05 significance threshold).



S10. **Impact of the initial clonal richness on acquisition of new mutational variants**. In silico simulations depicting relationship between the initial clonal richness, and clonal richness after 1095 days of simulations at the indicated mutation and fusion rates.



S11. **Impact of fusions on diversification**. Comparisons are drawn between results of in silico simulations involving mutations only versus mutation and fusion. Clonal diversity is captured by Shannon (**A**), Simpson (**B**) and GDI (**C**) diversity indexes. Mutation and fusion rates are indicated in the figures. Indicated p values denote the results for Kolmogorov-Smirnov test.



S12. Flowchart for agent based model for spatial simulations.

Supplementary Videos:

S1, S2. Time lapse videos of co-cultures between MDA-MB-231cells labelled with cytoplasmic GFP and nuclear mCherry. Images were acquired every 3 hours over 5 days.

S3, S4. Time lapse videos of co-cultures between MDA-MB-231cells labelled with cytoplasmic GFP and SUM159PT cells labelled with nuclear mCherry. Images were acquired every 3 hours over 5 days.

S5. Time lapse videos of co-cultures between SUM159PT labelled with nuclear mCherry and CAFS labelled with nuclear GFP. co-cultured with CAFs with nuclear GFP. Images were acquired every 3 hours over 5 days.

S6 Video of the entire simulation corresponding to Fig5G (2D mutation only).

S7 Video of the entire simulation corresponding to Fig5G (2D mutation and fusion)

S8 Video of the entire simulation corresponding to Fig5G (3D mutation only)

S9 Video of the entire simulation corresponding to Fig5G (3D mutation and fusion)

S10 Video of 3 days *in vitro* spatial simulation for the inferences of fusion rates described in the Mathematical Supplement.