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

Daria Myroshnychenko[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**].

Affiliations

1. Department of Cancer Physiology, H Lee Moffitt Cancer Centre 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 Centre and Research Institute, Tampa, FL, USA.
4. Department of Anatomic 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 is underpinned by somatic evolution, which follows 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 to lack 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 yeast, we set to interrogate 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 ploidy reduction, generating diverse progeny with heterogeneous combinations of genetic and phenotypic traits of fusion parents. 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 adaptive landscapes relative to strictly asexual populations, which may substantially accelerate tumor evolution.

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. The ability to generate new heritable diversity is crucial for the evolvability of populations of tumor cells, both during tumor progression and in response to therapies. These tumor cells have access to several powerful diversification mechanisms that are considered to be the hallmarks of cancer: genetic instability, elevated mutation rates and deregulation of epigenetic mechanisms that control gene expression. At the same time, they are generally assumed to lack a key, evolutionarily-conserved source of diversification – (para)sexual recombination – that could dramatically amplify diversity and generate new combinations of beneficial mutations (thus enabling new epistatic interactions) while removing the burden of disadvantageous ones, hence supporting population fitness and accelerating evolutionary adaptation.

While 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), occurrence of spontaneous cell fusions within tumors has been documented both *in vitro* and *in vivo*[1-3]. Given the previously reported impact of polyploidization on increased genomic instability[4, 5], evidence of ploidy reduction in the progeny of experimentally induced hybrid cells[6, 7], and existence of parasexual life cycles involving fusion-mediated recombination in pathogenic yeast C. albicans[8], 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 studies 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.

RESULTS.

In the course of multiple experimental studies involving *in vitro* co-cultures of differentially labelled tumor cells as well as tumor cells and cancer-associated fibroblasts (CAFs) in both 2D and 3D contexts, we occasionally noticed double positive cells (**Fig 1A, S1).** Examination of time lapse microscopy images revealed that these double positive cells can originate from spontaneous cell fusions **(Fig. S2A, B, Videos S1-5**). Observations of double positive cells were not limited to *in vitro* cultures as confocal microscopy examination of experimental xenograft tumors also revealed presence of occasional cells expressing both fluorescent labels (**Fig. 1B**). Given our interest in somatic evolution and the possibility that spontaneous cell fusions between genetically distinct cells might provide 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) types 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 negative controls, admixed ~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 (although only two of the homotypic cultures reached statistical significance) (**Fig. 1D, Fig. S2C**).

The unexpectedly high proportion of putative double-positive cells detected by flow analysis, substantial within group variability, as well as 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[9]. Indeed, examination of the images of double positive gate events (gating logic provided in **Fig.S3A**) revealed significant rates of false

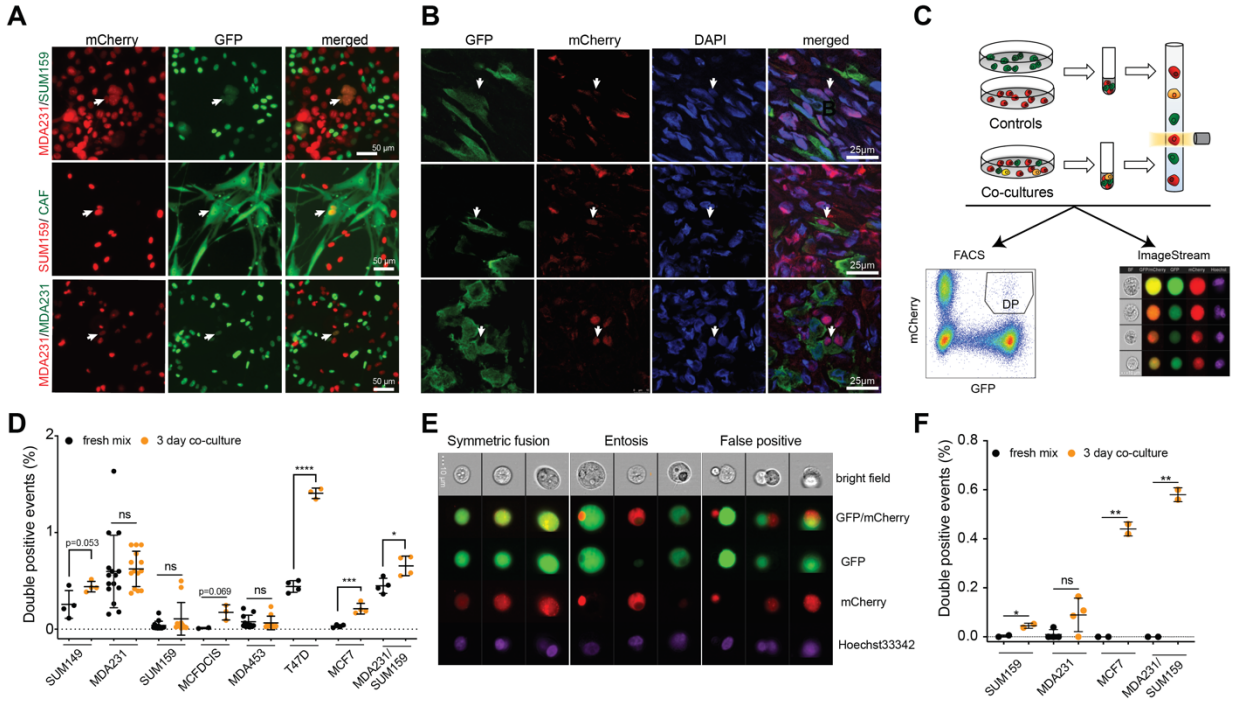


Fig.1 Occurrence of spontaneous cell fusions *in vitro* and *in vivo*.
A. Live fluorescent microscopy of co-cultures of the indicated differentially labelled cells. Arrowheads indicate cells co-expressing both fluorescent labels. **B.** Confocal immunofluorescent images from a xenograft tumor, initiated with a 50/50 mix of GFP and mCherry-labelled image MDA-MB-231 cells. Arrowheads indicate cells co-expressing GFP and mCherry. **C.** Experiment schemata for flow cytometry 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 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.

positives reflecting cell doublets (**Fig 1E**). Some of the double positive events were cells-within cells structures, indicating entosis[10] (**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 balanced 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, the 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 rare events, 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 cells expressing both GFP and mCherry fluorescent markers consistent with outgrowth of hybrid cells - products of somatic cell fusions (**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

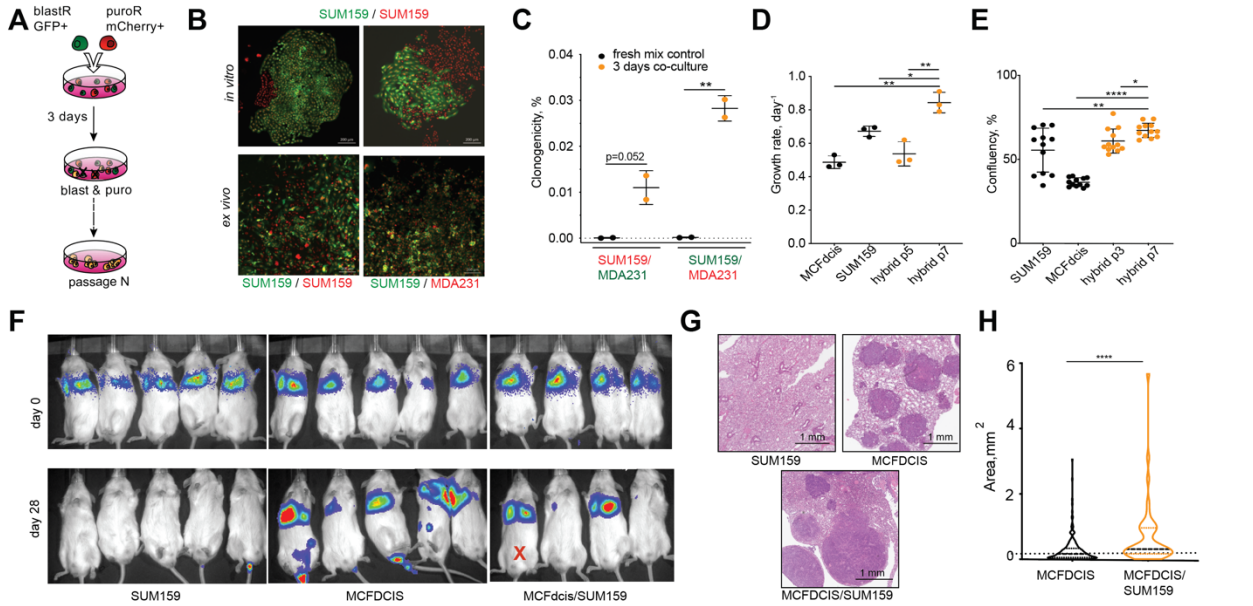


Fig.2 Phenotypic characterization of hybrids.
A. Experiment schemata for the selection of hybrid cells. **B.** Representative images of live fluorescent colonies formed after selection of the indicated *in vitro* co-cultures or *ex vivo* tumors. **C.** Quantitation of clonogenic survival frequencies. **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. **H.** Violin plot of size distribution of metastatic tumors. *, **, ***, *** 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).

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[11, 12], we were unable to recover colonies from co-cultures between multiple breast cancer cells and three distinct primary CAF isolates that we tested.

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 postantibiotic-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

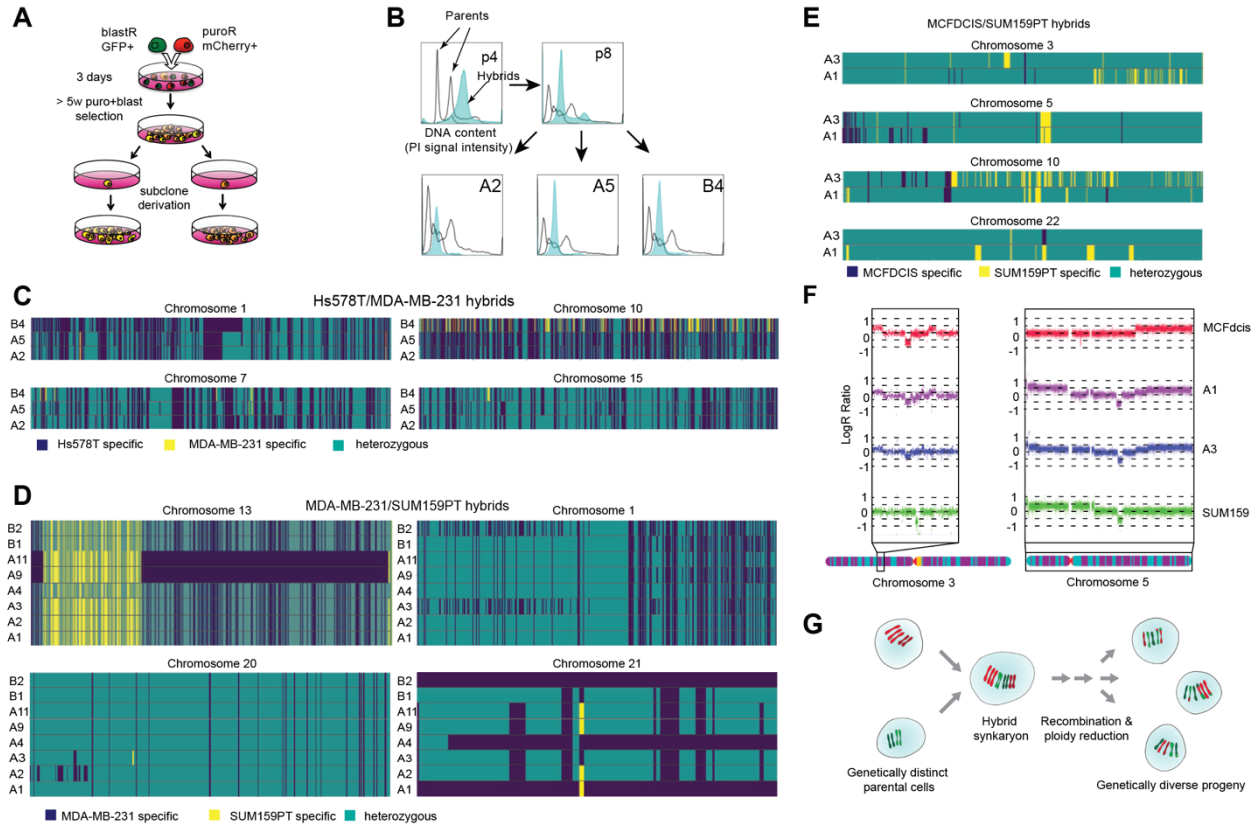


Fig.3 Fusion mediated genetic diversification.
A Experiment schemata for the derivation of single cell derived hybrid subclones.
B DNA content analysis of parental cell lines (HS578T and MDA-MB-231) and hybrids. A2, A5, B4 denote single cell derived subclones. **C – E.** Allelic inheritance of differential SNP markers of the parental cell lines in the subclonal progeny of the indicated hybrids. **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** Hypothetical interpretation of the effect of cell fusion and subsequent chromosome rearrangement on the genetic diversity of cells.

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 (**Fig. 2G**). In contrast, luminescent signals in mice injected with MCFDCIS and the hybrids cells was retained and increased over time in some of the animals. Histological examination revealed that all of these mice contained macroscopic tumors in the lungs. Interestingly, despite weaker luminescent signals, hybrid tumors grew larger than those from MCFDCIS tumors (**Fig 2G, H, S4 D**), most likely reflecting loss of luciferase gene in some of the hybrids. Notably, the hybrid xenograft with strong luminescent signal (denoted by X in Fig. 2F) had died prior to euthanasia; visual examination revealed multiple macroscopic tumors, but due to poor tissue quality this animal was excluded from the analysis. In summary, these data suggest that that spontaneous cell fusions can generate cells with more aggressive oncogenic properties.

In the absence of the pTP53 depended checkpoint function, which is commonly disrupted in cancer cells, polyploidy is known to be associated with increased genomic instability[5]. Consistently, genomic instability[15] and ploidy reduction were reported in experimentally-induced somatic hybrids[6, 7]. Therefore, we decided to examine

the stability of the hybrid cells' DNA content over time. As expected, all of the examined hybrid cell lines immediately following antibiotic selection displayed elevated ploidy that is consistent with the combined 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**). Given that fusion-mediated recombination and stochastic ploidy reduction can serve as the mechanism for parasexual recombination in the pathogenic yeast species C Albicans[8, 16] and that cycles of somatic cell fusions followed by genetic recombination and ploidy reduction have been described to operate in normal hepatocytes[17] and hematopoietic cells[18], 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, different sub-clonal derivatives of the ostensibly stable MCFDCIS/SUM19 and MDA231/MCFDCIS retained similarly elevated ploidy (**Fig S5**). In contrast, genomic sub-clonal derivatives of the hybrids with evidence of ploidy reduction (HS578T/MDA231 and MDA231/SUM159) displayed clear evidence for clonal variation in DNA content (**Fig 3B, Fig. S5)**.

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 the patterns of allelic inheritance between different sub-clones of the genetically unstable HS578T/MDA231 hybrids (**Fig. 3B, Fig. S5**), indicating that ploidy reduction may be accompanied by genetic recombination (**Fig.3C, S6**). Likewise, we observed substantial divergence in the allelic inheritance among different sub-clones from the unstable MDA231/SUM159 hybrids with the Illumina CytoSPN-12 platform (**Fig. 3D, Fig. S7**). Surprisingly, CytoSPN-12-based analysis of the ostensibly stable MCFDCIS/SUM159 hybrids (**Fig. S5**) revealed some divergence in allelic inheritance between the two clones that we have analyzed**,** despite the maintenance of similarly elevated DNA content (**Fig. 3E, Fig. S8)**.

It should be noted that the binary analysis of allelic inheritance misses the potential variability in copy numbers of the differential alleles. Since accurate comparison of copy number differences is complicated in the highly aneuploid genomes of most cancer cell lines, we decided to focus on the ostensibly stable hybrids between the near-diploid cell lines SUM159PT and MCF10DCIS. Analysis of genomic regions retaining both parental alleles revealed copy number differences between the distinct clones, suggesting that genomic diversification impacts not only differential inheritance of the alleles but also their copy number variation (**Fig. 3F**). In summary, these data suggest that cell fusion between genetically distinct tumor cells can serve as a source of significant genetic diversification (**Fig. 3G**).

Genetic diversification is not the only source of heritable phenotypic diversity within tumor cell populations as heterogeneity in biologically and clinically important phenotypes may also be shaped by epigenetic mechanisms. Further, 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[19]. Therefore, we decided to examine the impact of somatic fusions on phenotypic diversification. To this end, we performed single-cell expression profiling with a/the 10x platform to examine the phenotypes of MDA-MB-231, SUM159PT cells and SUM159/MDA231 hybrids at early (2) and late (10) passages following elimination of negative controls with dual -antibiotic selection (**Fig. 4A**). UMAP clustering[20] 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

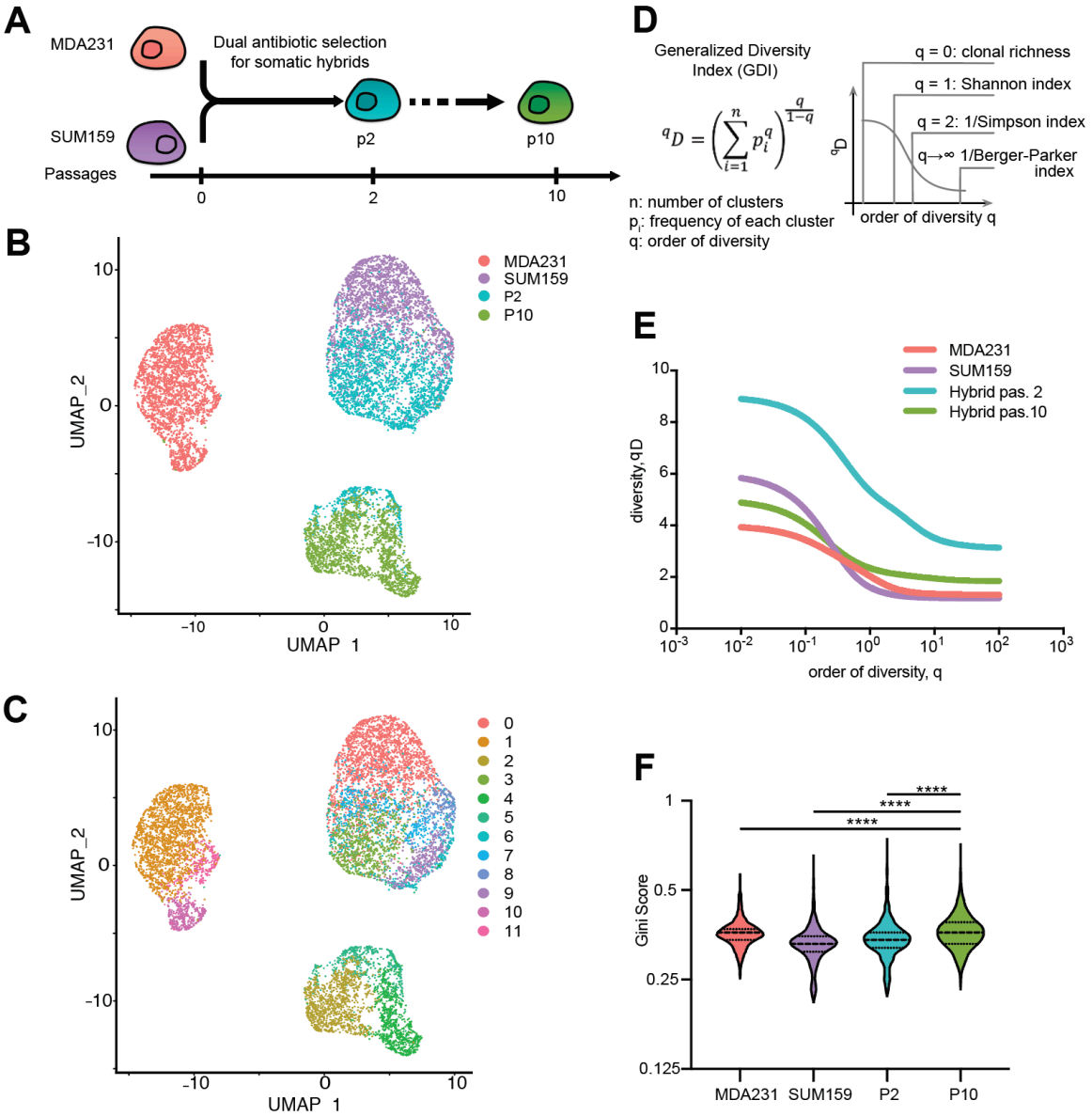


Fig 4. Fusion mediated phenotypic diversification.
A. Experimental schematic for samples used for scRNAseq analyses. **B.** UMAP clustering of analyzed samples and **C**. distinct phenotypic clusters. **D**. Formula for calculation of GDI and mapping to the common diversity indexes. **E**. GDI analysis of phenotypic diversity of the indicated cells based on the clusters identified in (**C**). **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 $p \leq 0.0001$ in a Wilcoxon signed rank test**.**

the selection of a fit sub-population of hybrid cells, as well as 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 the general diversity index (GDI) **(Fig. 4D)**[21]. GDI enables characterization of diversity across a spectrum of scales (orders of diversity)[22],

ranging from clonal richness (low order of diversity reveals the number of distinct subpopulations) to classic measures, such as Shannon or Simpson indices (intermediate order of diversity) and to the high orders of diversity, giving more weight to more highly abundant populations[21]. We found that, at the early passage, hybrids displayed dramatically higher diversity across all scales (**Fig. 4E**). However, at the later passage, diversity at low orders went down while remaining higher than either those of the parental cells at intermediate and high orders of diversity. The GDI analyses rely on grouping phenotypes into clusters, which 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 Gini index, which has been recently applied towards characterization of phenotypic diversification in cancer cell populations[23]. We found that SUM159/MDA231 hybrids displayed elevated Gini indexes compared to both parents (**Fig. 4F**), further supporting the notion that somatic hybridization might lead to phenotypic diversification.

Despite the substantial genetic and phenotypic diversification observed in hybrid cells and their progeny, the impact of this diversification on population diversity is less intuitively clear given the relatively low frequency of spontaneous cell fusions. To interrogate the impact of the phenomena at the population level, we decided to use *in silico* simulations based on a stochastic Poisson mutational branching model of tumor growth (**Fig. 5A and Mathematical Supplement**). We started by deriving the probability of clonogenic cell fusions, cell proliferation and cell death from our experimental data (**Mathematical supplement**). Using these estimates, as well as estimates of mutation rates from the literature[24], we compared diversification rates between populations of tumor cells where genetic diversity is generated through mutation-mediated and fusion-mediated recombination and those with genetic diversity sourced from mutation-mediated recombination only. We found that fusion-mediated recombination substantially increases the emergence of unique mutants as well as the maximal numbers of mutations within a single lineage (**Fig. 5B-D**). This effect was observed across a range of fusion and mutation probabilities (**Fig. 5E, Fig. S9**). As expected, the impact was more pronounced at higher fusion probabilities. Additionally, the impact of fusion-mediated recombination was less intuitively elevated by higher mutation rates. In order to better understand this observation, we examined the ability of fusion-mediated recombination to generate new mutational variants as a function of pre-existing mutational heterogeneity 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**).

The above modeling results were obtained with an assumption of a well-mixed population, which is clearly violated in solid tumors. Thus, we ask whether fusion-mediated recombination could still have a substantial impact when spatial constraints are considered and whether fusions between genetically dissimilar cells should have a higher impact on diversification. Using spatial agent-based simulations (**Mathematical Supplement**), we compared diversification between populations with only mutations (which are not impacted by spatial considerations) and spatially restricted populations with mutations and cell fusions in 2D and 3D (**Fig. 5G**). Even though spatial constraints blunted the impact of fusions, diversity was still acquired faster in the

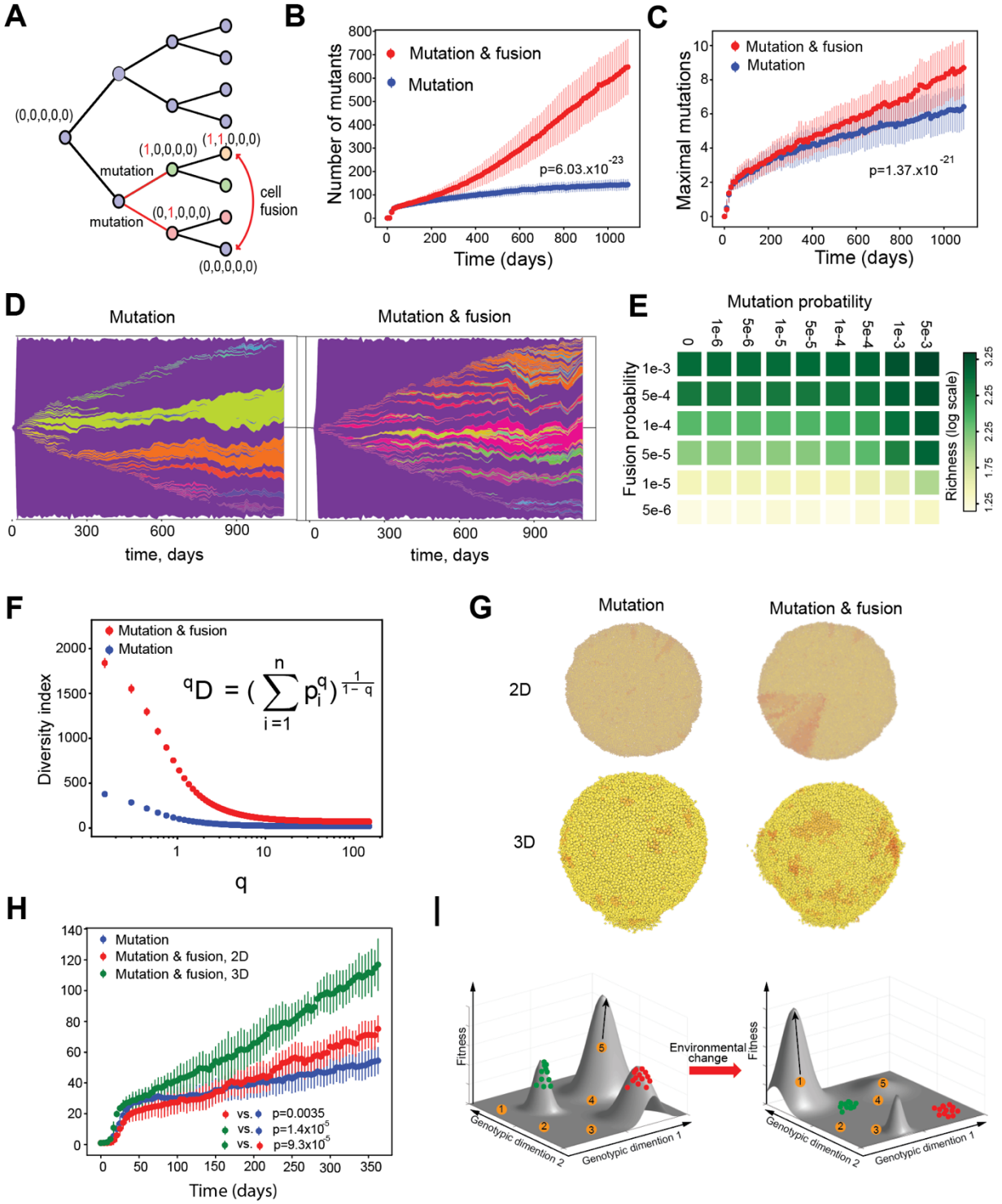


Fig.5. Impact of fusion mediated recombination on genetic diversity in tumor cell populations.
A. Model schemata of a birth-death process. A binary vector represents the genotype of the cells, with (0,0,…,0) representing the initial genotype. Cells can stochastically acquire mutations during cell division and randomly exchange mutations during fusion-mediated recombination. Dynamics of unique mutation acquisition (**B**) and highest number of unique mutation 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; p values show the result of KS tests. **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 numbers of distinct mutational variants (clonal richness). **F**. Generalized Diversity Index after 1095 days of simulation; pf=0.001, $\mu=1x10^{-4}$. **G**. Visualization of the end result (350 days) of spatial agent-based simulations examining how spatial constraints modify the impact of fusion mediated recombination; pf=0.00035, $\mu=5x10^{-3}$. **H**. Number of distinct mutants over time for the spatial simulation shown in (G); p values show the results of KS tests for the indicated comparisons. **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.

presence of fusion-mediated recombination. Notably, the impact of cell fusions was substantially higher in 3D compared to 2D scenarios, reflecting higher 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 explore larger areas of adaptive landscapes, thus augmenting their evolvability.

Multiple prior studies have described the occurrence of spontaneous, clonogenic cell fusions between neoplastic cells as well as fusions between neoplastic and non-neoplastic cells within the tumor microenvironment, including fibroblasts[11, 25], endothelia[26] and myeloid cells[27] 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 in from these heterotypic hybrids. Since non-immortalized primary CAFs remain genetically normal[13], lack of proliferation might reflect the dominant impact of the intact TP53 checkpoint to limit proliferation of polyploid cells[5, 14]. Indeed, examination of time lapse microscopy images indicated that, in contrast to hybrids between carcinoma cells, which 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, they suggest that spontaneous somatic cell fusions between cancer cells, which have lost intact checkpoint mechanisms to preserve genomic integrity, may be more capable of generating proliferatively viable progeny.

Spontaneous cell fusions have been implicated in directly increasing the ability of cancer cells to invade, migrate and establish metastases[12, 28]. Our results, documenting

occurrence of low-frequency spontaneous cell fusions across a panel of breast cancer cell lines (**Fig. 1**), and demonstrating an increased invasive and metastatic potential of somatic hybrids (**Fig. 2**) are consistent with these prior findings and support the prior argument for the potential importance of cell fusions. However, the relatively low frequency of spontaneous cell fusions as well as the lack of evidence that cell fusions are required to facilitate invasive phenotypes tamper interest towards this phenomenon in the field.

Whereas the significance of the direct phenotypic impact of spontaneous somatic cell fusions has yet to be properly interrogated, 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[8]. Whereas genetic and phenotypic diversification resulting from cell fusion has been proposed to underlie cell transformation and tumor initiation[7, 15], our results suggest the importance of fusion-mediated recombination for the ongoing diversification within tumor cell populations. Even though the experimentally observed frequency of spontaneous cell fusions is relatively low, it is notably similar to the frequency of parasexual recombination in yeast species with facultative (para)sexual recombination life cycle[29] , suggesting biological significance of this phenomenon. Indeed, *in silico* modeling experiments suggest that observed rates of cell fusions significantly accelerate diversification within tumor cell populations once 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[19]. In support of this notion, cell fusions have been associated with nuclear reprogramming[27] and increased phenotypic plasticity in multiple contexts[30] while increased migration and invasiveness of hybrid cells (**Fig. S4**) has been linked with stemness[31].

At this point, we cannot completely exclude the possibility that our observations of spontaneous cell fusions within *in vitro* and xenograft models might represent an experimental artefact. On the other hand, occurrence of spontaneous cell fusions between xenografted breast carcinoma cells and host stroma has been documented in an animal model[25]. Additionally, spontaneous fusions between cancer cells and myeloid cells have been rigorously described in animal models and human malignancies[12]. Detecting spontaneous cell fusions in primary human cancers is notoriously difficult as most tumors are thought to initiate from a single transformed cell. Even though giant polyploid cells consistent with spontaneous fusions can be observed in many human neoplasms, the absence of genetic markers makes it 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[32, 33] – providing direct clinical evidence for a potential clinical relevance of spontaneous cell fusions. Unfortunately, these cases are too rare for systematic interrogation. Still, given the potentially profound impact of spontaneous cell fusions and 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 the need to develop methodological approaches to either validate or refute clinical relevance of the phenomena

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[34]. 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. 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), 0.2mg/ml RNAse (Qiagen) and 20ug/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 SUM159/MDA231 hybrids CytoSNP-12 v2.1 BeadChip array from Illumina was used and data analyzed with GenomeStudio 2.0 software (Illumina). For MCFDCIS/SUM159 and HS578T/MDA231 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 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 SUM159/MDA231. Libraries were constructed according to manufacturer's protocol, sequenced on a 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/) [35]. 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 [36]. 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 [37]. The data was then visualized by running the uniform manifold approximation and projection (UMAP) algorithm [20, 38] and either coloring individual cells based on their cluster identity or cell type identity. Gini coefficients were calculated with the package DescTools in R [39]. 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.

ACKNOWLEDGMENTS

We thank Dr. Aaron Goldman for sharing his unpublished observations of frequent occurrence of putative cell fusions in experimental studies, which confirmed our observations from multiple lines of work thus motivating this study. We thank the Flow Cytometry, Analytic Microscopy, Tissue Histology, Biostatistics and Bioinformatics Shared Resource and Molecular Genomics 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 winning team award.

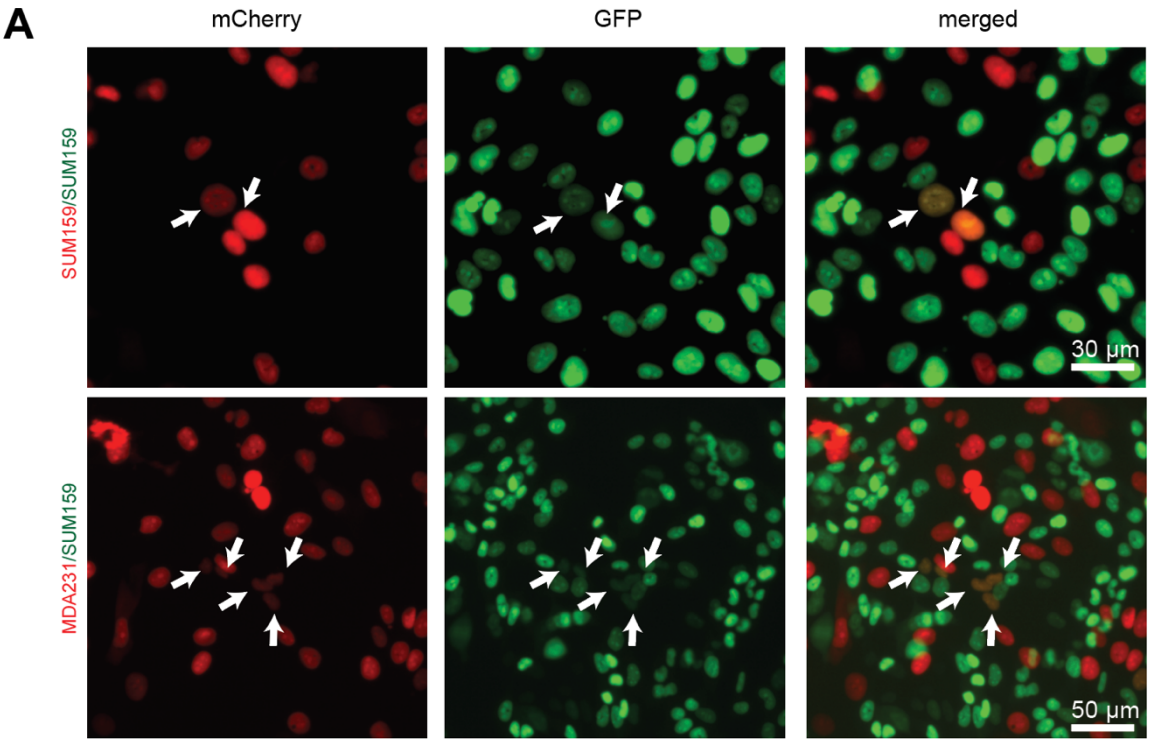
CITED REFERENCES.

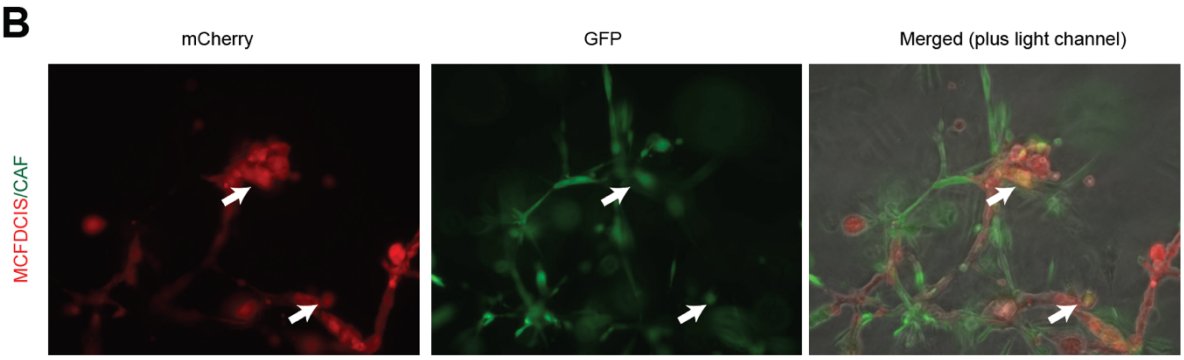
1. Duelli, D. & Lazebnik, Y. Cell fusion: a hidden enemy? *Cancer Cell* **3**, 445-8 (2003).
2. Lu, X. & Kang, Y. Cell fusion as a hidden force in tumor progression. *Cancer Res* **69**, 8536-9 (2009).
3. Platt, J.L., Zhou, X., Lefferts, A.R. & Cascalho, M. Cell Fusion in the War on Cancer: A Perspective on the Inception of Malignancy. *Int J Mol Sci* **17** (2016).
4. Kuznetsova, A.Y., Seget, K., Moeller, G.K., de Pagter, M.S., de Roos, J.A., Durrbaum, M., Kuffer, C., Muller, S., Zaman, G.J., Kloosterman, W.P. & Storchova, Z. Chromosomal instability, tolerance of mitotic errors and multidrug resistance are promoted by tetraploidization in human cells. *Cell Cycle* **14**, 2810-20 (2015).
5. Fujiwara, T., Bandi, M., Nitta, M., Ivanova, E.V., Bronson, R.T. & Pellman, D. Cytokinesis failure generating tetraploids promotes tumorigenesis in p53-null cells. *Nature* **437**, 1043-7 (2005).
6. Su, Y., Subedee, A., Bloushtain-Qimron, N., Savova, V., Krzystanek, M., Li, L., Marusyk, A., Tabassum, D.P., Zak, A., Flacker, M.J., Li, M., Lin, J.J., Sukumar, S., Suzuki, H., Long, H., Szallasi, Z., Gimelbrant, A., Maruyama, R. & Polyak, K. Somatic Cell Fusions Reveal Extensive Heterogeneity in Basal-like Breast Cancer. *Cell Rep* **11**, 1549-63 (2015).
7. Zhou, X., Merchak, K., Lee, W., Grande, J.P., Cascalho, M. & Platt, J.L. Cell Fusion Connects Oncogenesis with Tumor Evolution. *Am J Pathol* **185**, 2049-60 (2015).
8. Bennett, R.J. The parasexual lifestyle of Candida albicans. *Curr Opin Microbiol* **28**, 10-7 (2015).
9. 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**, 279-90 (2007).

10. Fais, S. & Overholtzer, M. Cell-in-cell phenomena in cancer. *Nat Rev Cancer* **18**, 758-766 (2018).
11. 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**, 2504-15 (2012).
12. Gast, C.E., Silk, A.D., Zarour, L., Riegler, L., Burkhart, J.G., Gustafson, K.T., Parappilly, M.S., Roh-Johnson, M., Goodman, J.R., Olson, B., Schmidt, M., Swain, J.R., Davies, P.S., Shasthri, V., Iizuka, S., Flynn, P., Watson, S., Korkola, J., Courtneidge, S.A., Fischer, J.M., Jaboin, J., Billingsley, K.G., Lopez, C.D., Burchard, J., Gray, J., Coussens, L.M., Sheppard, B.C. & Wong, M.H. Cell fusion potentiates tumor heterogeneity and reveals circulating hybrid cells that correlate with stage and survival. *Sci Adv* **4**, eaat7828 (2018).
13. Qiu, W., Hu, M., Sridhar, A., Opeskin, K., Fox, S., Shipitsin, M., Trivett, M., Thompson, E.R., Ramakrishna, M., Gorringe, K.L., Polyak, K., Haviv, I. & Campbell, I.G. No evidence of clonal somatic genetic alterations in cancer-associated fibroblasts from human breast and ovarian carcinomas. *Nat Genet* **40**, 650-5 (2008).
14. Vitale, I., Senovilla, L., Jemaa, M., Michaud, M., Galluzzi, L., Kepp, O., Nanty, L., Criollo, A., Rello-Varona, S., Manic, G., Metivier, D., Vivet, S., Tajeddine, N., Joza, N., Valent, A., Castedo, M. & Kroemer, G. Multipolar mitosis of tetraploid cells: inhibition by p53 and dependency on Mos. *EMBO J* **29**, 1272-84 (2010).
15. Duelli, D.M., Padilla-Nash, H.M., Berman, D., Murphy, K.M., Ried, T. & Lazebnik, Y. A virus causes cancer by inducing massive chromosomal instability through cell fusion. *Curr Biol* **17**, 431-7 (2007).
16. Forche, A., Alby, K., Schaefer, D., Johnson, A.D., Berman, J. & Bennett, R.J. The parasexual cycle in Candida albicans provides an alternative pathway to meiosis for the formation of recombinant strains. *PLoS Biol* **6**, e110 (2008).
17. Duncan, A.W., Taylor, M.H., Hickey, R.D., Hanlon Newell, A.E., Lenzi, M.L., Olson, S.B., Finegold, M.J. & Grompe, M. The ploidy conveyor of mature hepatocytes as a source of genetic variation. *Nature* **467**, 707-10 (2010).
18. 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**, 2837-43 (2012).
19. Koulakov, A.A. & Lazebnik, Y. The problem of colliding networks and its relation to cell fusion and cancer. *Biophys J* **103**, 2011-20 (2012).
20. Becht, E., McInnes, L., Healy, J., Dutertre, C.A., Kwok, I.W.H., Ng, L.G., Ginhoux, F. & Newell, E.W. Dimensionality reduction for visualizing single-cell data using UMAP. *Nat Biotechnol* (2018).
21. Ferrall-Fairbanks, M.C., Ball, M., Padron, E. & Altrock, P.M. Leveraging single cell RNA sequencing experiments to model intra-tumor heterogeneity. *bioRxiv* (2018).
22. Hill, M.O. Diversity and Evenness: A Unifying Notation and Its Consequences. *Ecology* **54**, 427-432 (1973).
23. Hinohara, K., Wu, H.J., Sebastien, V., McDonald, T.O., Igarashi, K.J., Yamamoto, K.N., Madsen, T., Fassl, A., Egri, S.B., Papanastasiou, M., Ding, L., Peluffo, G., Cohen, O., Kales, S.C., Lal-Nag, M., Rai, G., Maloney, D.J., Jadhav, A., Simeonov, A., Wagle, N., Brown, M., Meissner, A., Sicinski, P., Jaffe, J.D., Jeselsohn, R., Gimelbrant, A.A., Michor, F. & Polyak, K. KDM5 Histone Demethylase Activity Links Cellular Transcriptomic

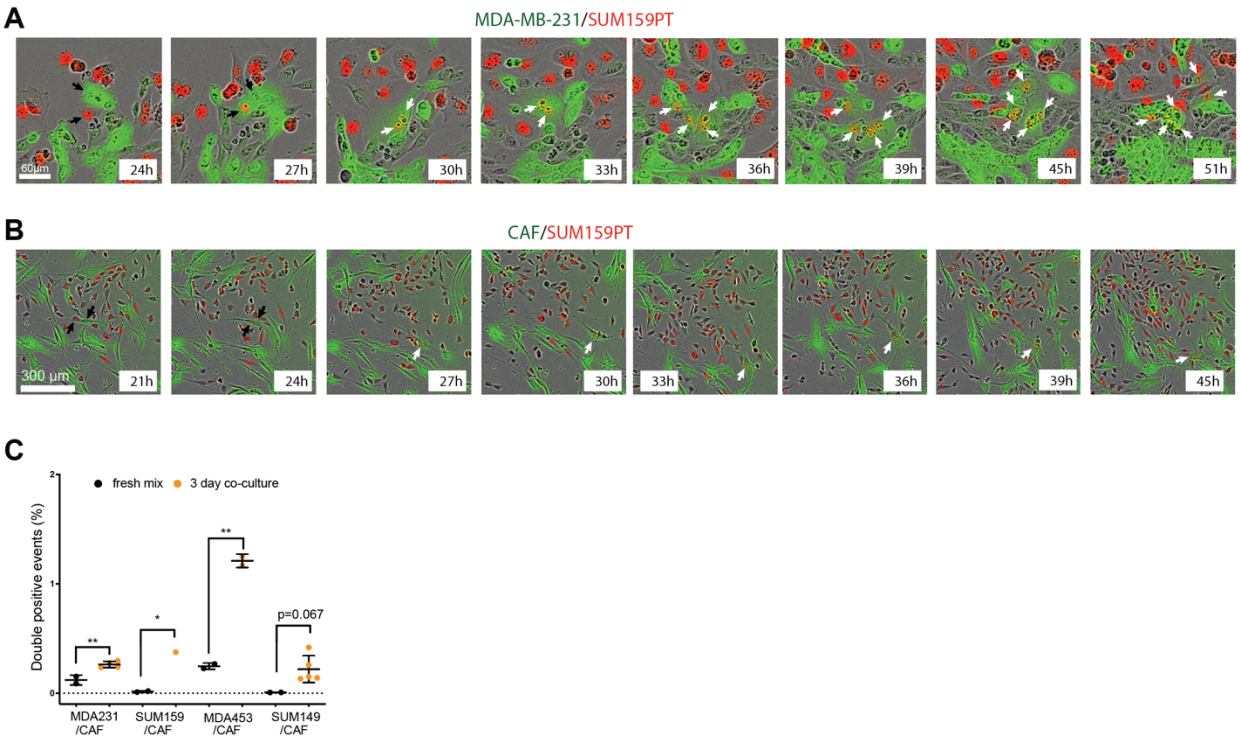
Heterogeneity to Therapeutic Resistance. *Cancer Cell* **35**, 330-332 (2019).
24. Loeb, L.A. Human Cancers Express a Mutator Phenotype: Hypothesis, Origin, and Consequences. *Cancer Res* **76**, 2057-9 (2016).
25. Jacobsen, B.M., Harrell, J.C., Jedlicka, P., Borges, V.F., Varella-Garcia, M. & Horwitz, K.B. Spontaneous fusion with, and transformation of mouse stroma by, malignant human breast cancer epithelium. *Cancer Res* **66**, 8274-9 (2006).
26. Mortensen, K., Lichtenberg, J., Thomsen, P.D. & Larsson, L.I. Spontaneous fusion between cancer cells and endothelial cells. *Cell Mol Life Sci* **61**, 2125-31 (2004).
27. Powell, A.E., Anderson, E.C., Davies, P.S., Silk, A.D., Pelz, C., Impey, S. & Wong, M.H. Fusion between Intestinal epithelial cells and macrophages in a cancer context results in nuclear reprogramming. *Cancer Res* **71**, 1497-505 (2011).
28. 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**, 9385-90 (2009).
29. Nieuwenhuis, B.P. & James, T.Y. The frequency of sex in fungi. *Philos Trans R Soc Lond B Biol Sci* **371** (2016).
30. Lazebnik, Y. The shock of being united and symphiliosis. Another lesson from plants? *Cell Cycle* **13**, 2323-9 (2014).
31. Mani, S.A., Guo, W., Liao, M.J., Eaton, E.N., Ayyanan, A., Zhou, A.Y., Brooks, M., Reinhard, F., Zhang, C.C., Shipitsin, M., Campbell, L.L., Polyak, K., Brisken, C., Yang, J. & Weinberg, R.A. The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell* **133**, 704-15 (2008).
32. Lazova, R., Laberge, G.S., Duvall, E., Spoelstra, N., Klump, V., Sznol, M., Cooper, D., Spritz, R.A., Chang, J.T. & Pawelek, J.M. 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 (2013).
33. 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 (2017).
34. Marusyk, A., Tabassum, D.P., Janiszewska, M., Place, A.E., Trinh, A., Rozhok, A.I., Pyne, S., Guerriero, J.L., Shu, S., Ekram, M., Ishkin, A., Cahill, D.P., Nikolsky, Y., Chan, T.A., Rimawi, M.F., Hilsenbeck, S., Schiff, R., Osborne, K.C., Letai, A. & Polyak, K. Spatial Proximity to Fibroblasts Impacts Molecular Features and Therapeutic Sensitivity of Breast Cancer Cells Influencing Clinical Outcomes. *Cancer Res* **76**, 6495-6506 (2016).
35. 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 (2018).
36. 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 (2018).
37. Newman, M.E. Modularity and community structure in networks. *Proc Natl Acad Sci U S A* **103**, 8577-82 (2006).
38. McInnes, L., Healy, J., Saul, N. & Großberger, L. UMAP: Uniform Manifold Approximation and Projection. *Journal of Open Source Software* **3** (2018).
39. Signorell, A. DescTools: Tools for descriptive statistics. *R package version 0.99* **18** (2016).

Supplementary figures

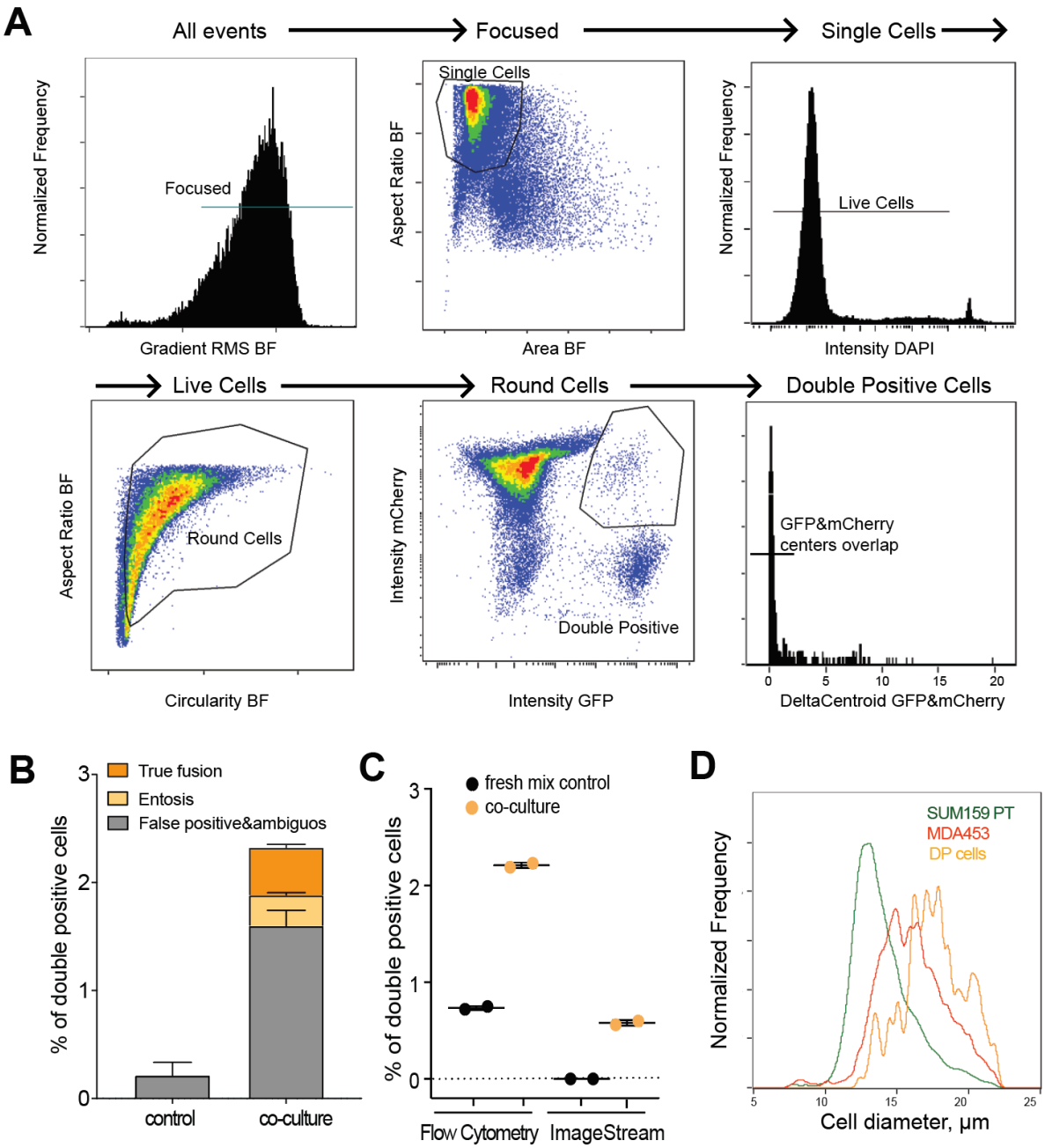




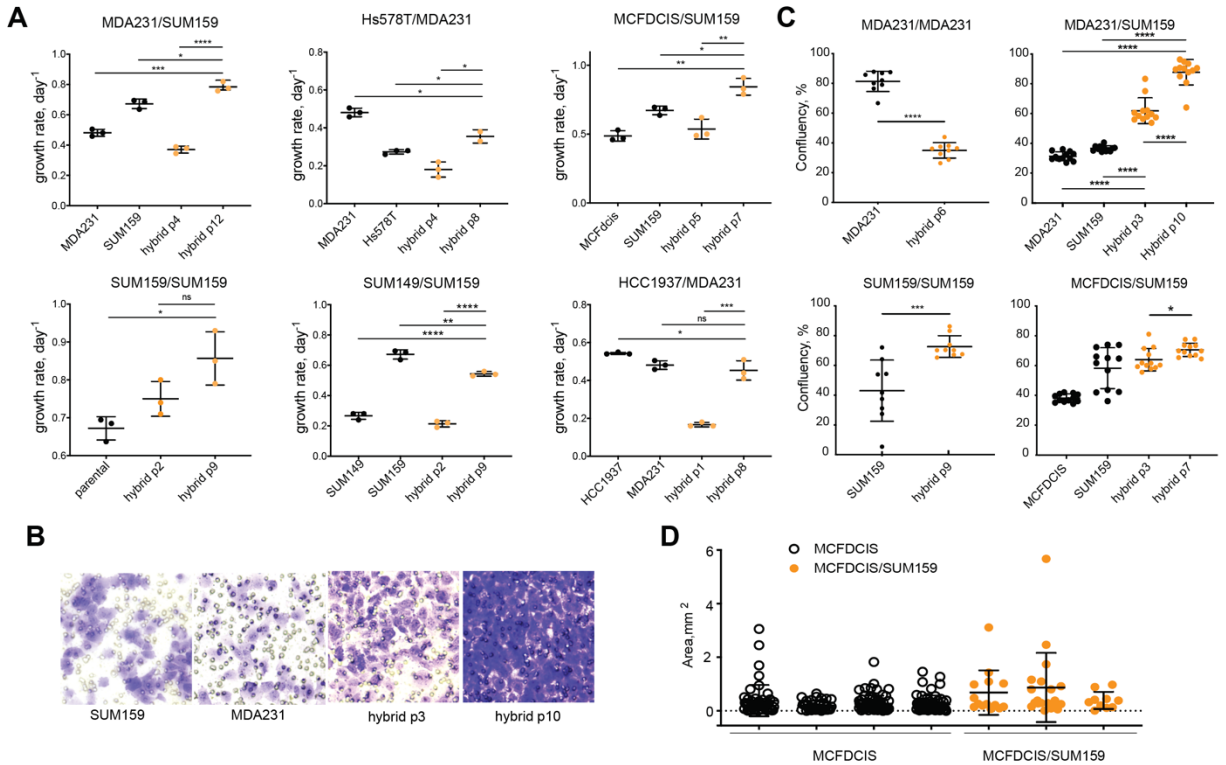
S1. A. Live cell fluorescent microscopy images of 2d co-cultures between differentially labelled (nuclear GFP and mCherry) cell lines. Arrowheads indicate cells that express both labels. **B.** Live-cell fluorescent microscopy images of 3D Matrigel co-cultures between differentially labelled (cytoplasmic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



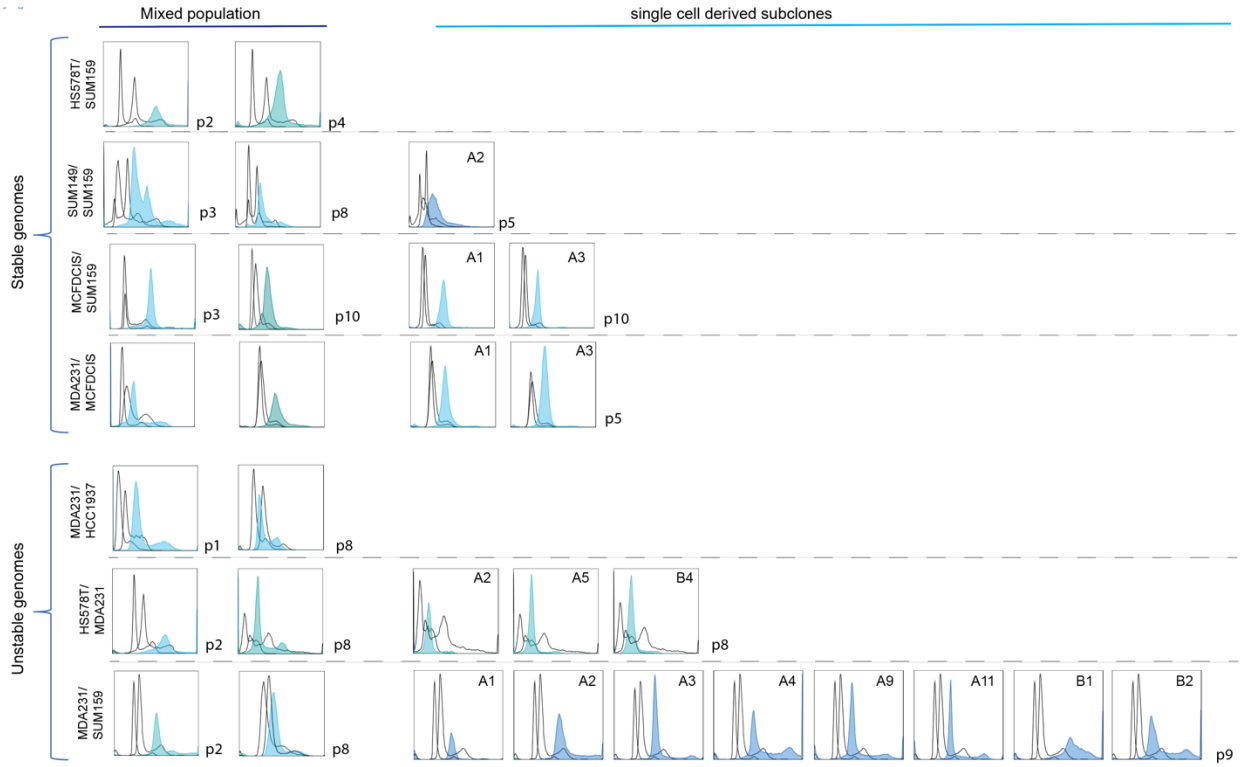
S2. Time-lapse live fluorescent microscopy images of mCherry+ SUM159PT cells cultured with GFP+ MDA-MB-231 (**A**) and CAF (**B**). The labels indicate time after plating. Black arrowheads show fusion parents, white arrowheads show double positive hybrid cells after fusion and after cell division. **C.** 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.



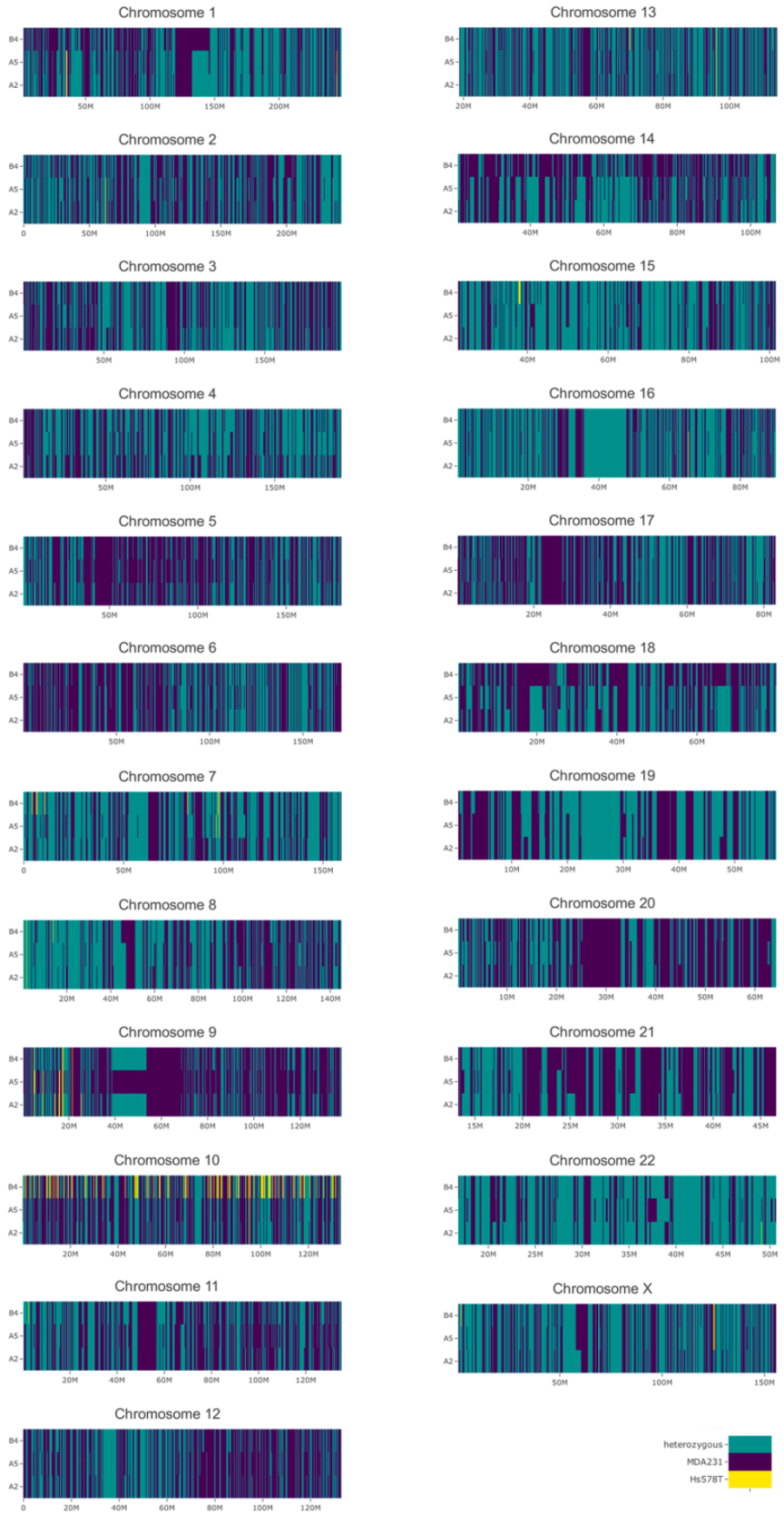
S3. A. Gating strategy for the detection of double positive cells with Image Stream imaging flow cytometry platform. **B.** Quantitation of different classes of double positive events in 3-day co-cultures between 50/50 mixes of GFP/mCherry labelled MCF7 cells. **C**. Direct comparison of frequency of double positive events detected in co-cultures of differentially labelled MDA-MB-231 (mCherry+) and SUM159PT (GFP+) cells and freshly mixed controls between FACs and Image Stream platforms (validated true positives). **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).



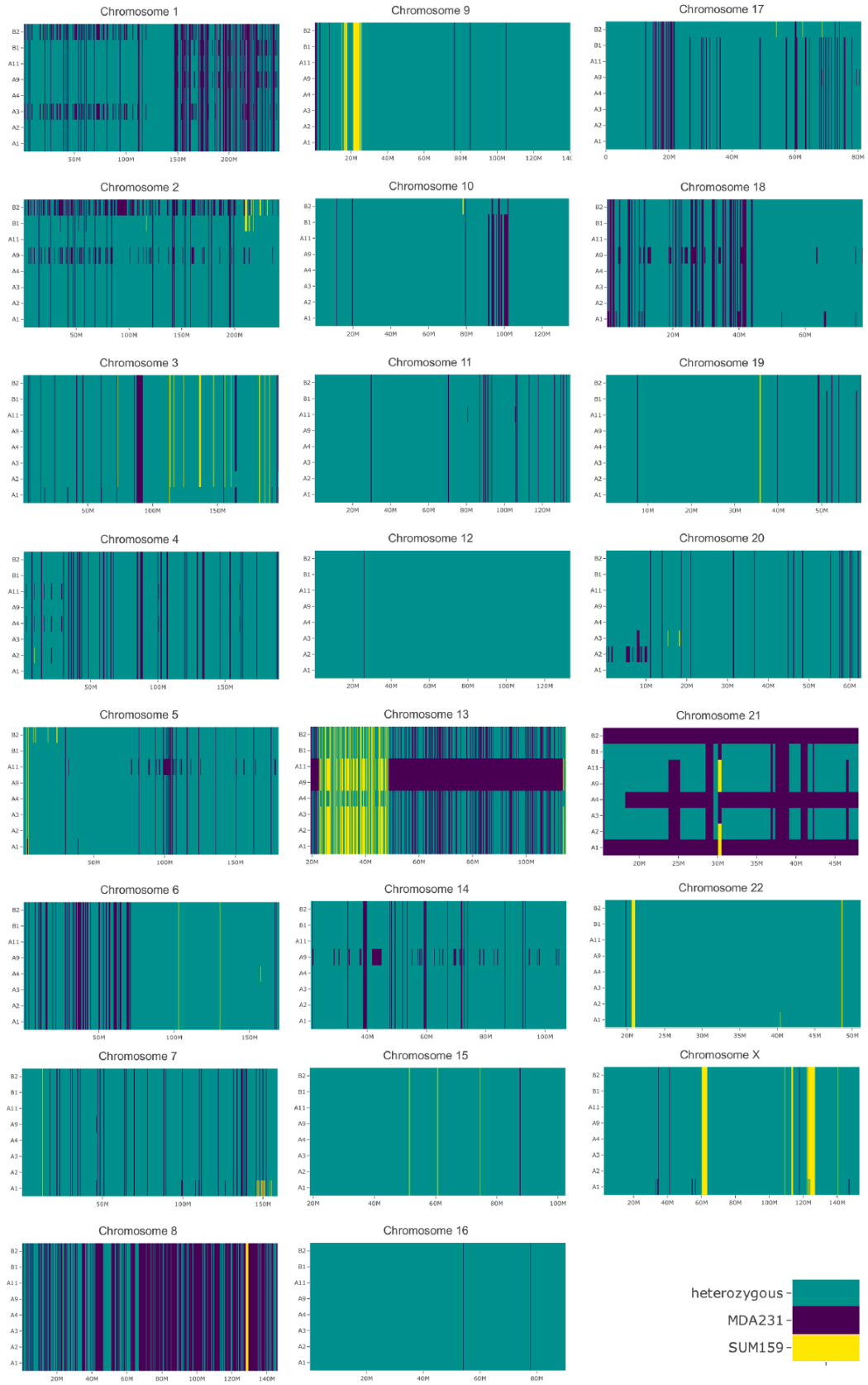
S4. A. Growth rates of the indicated cell lines and their hybrids, at the indicated passages post-antibiotic-selection. **B.** Representative images of stained membranes from Boyden chamber cell invasion/migration assay. **C**. Quantitation of Boyden chamber cell invasion/migration assay data. **D.** Quantification of area of lung metastases formed after tail vein injection of MCF10DCIS and MCF10DCIS/SUM159 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.



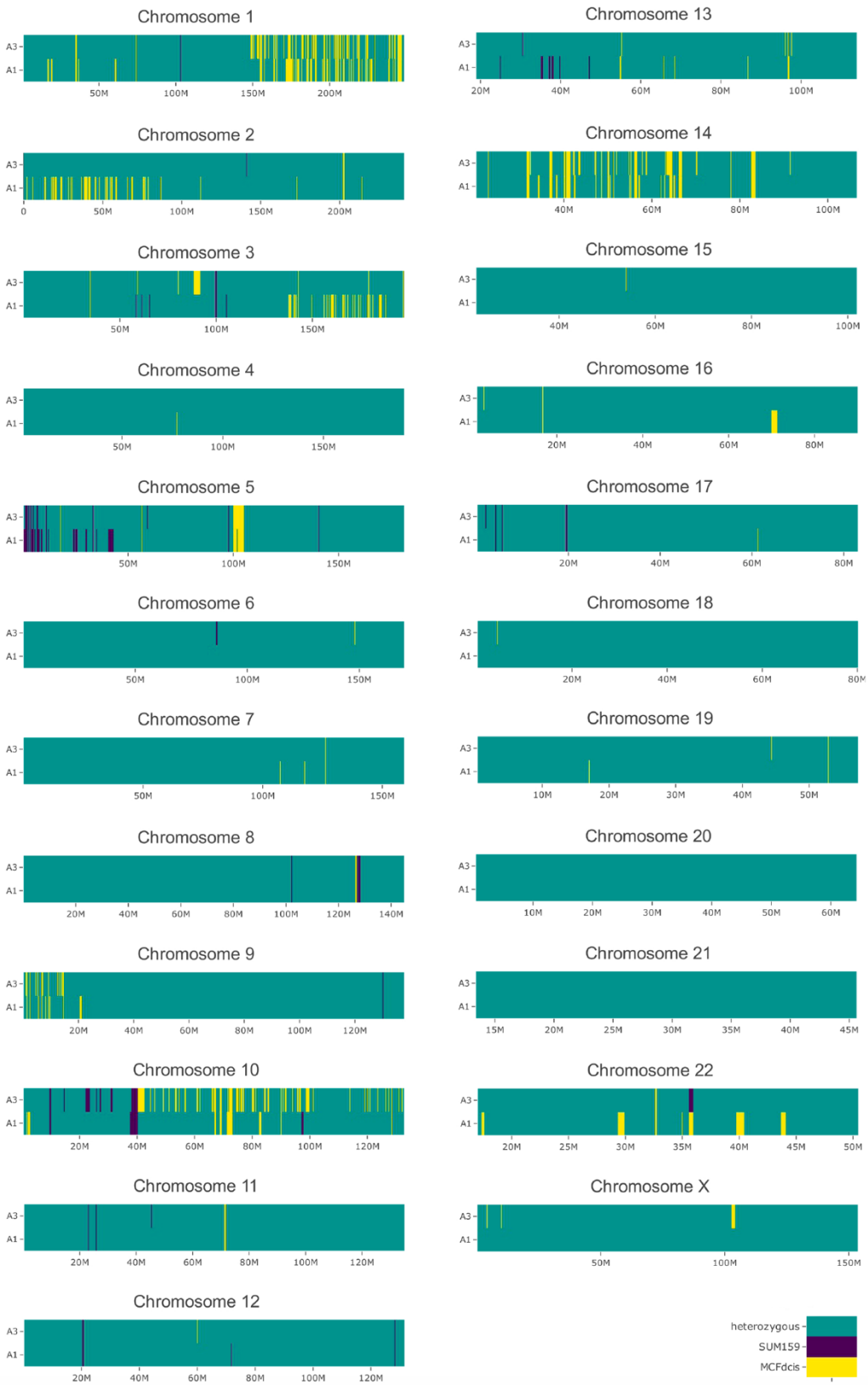
S5 Ploidy analysis of the indicated parental cell lines and their hybrids at indicated passages post-antibiotic-selection. Black contours indicate DNA content profiles for parental cells, – parental cell lines, filled histograms indicate DNA content profiles of the hybrids.



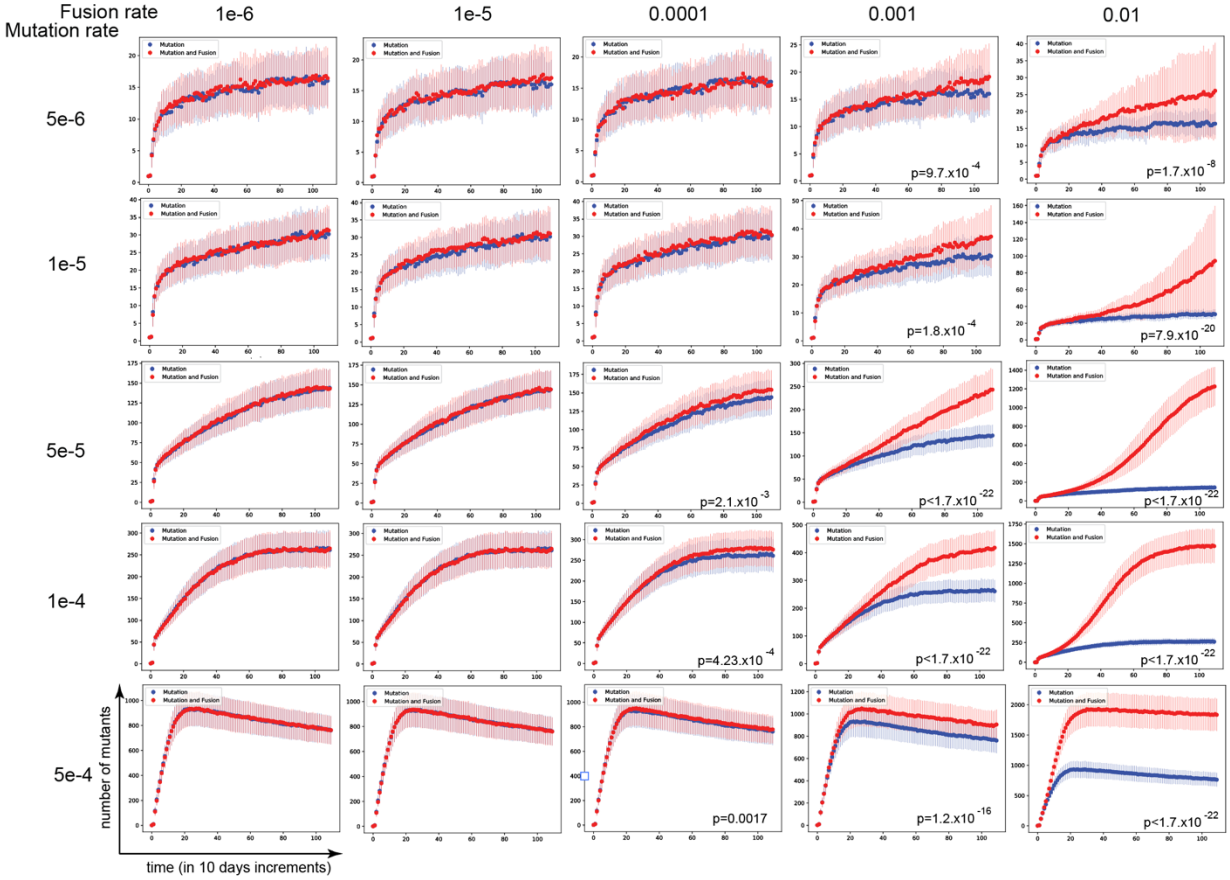
S6. SNP allelic inheritance analyses for hybrids between HS578T and MDA-MB-231 cell lines.



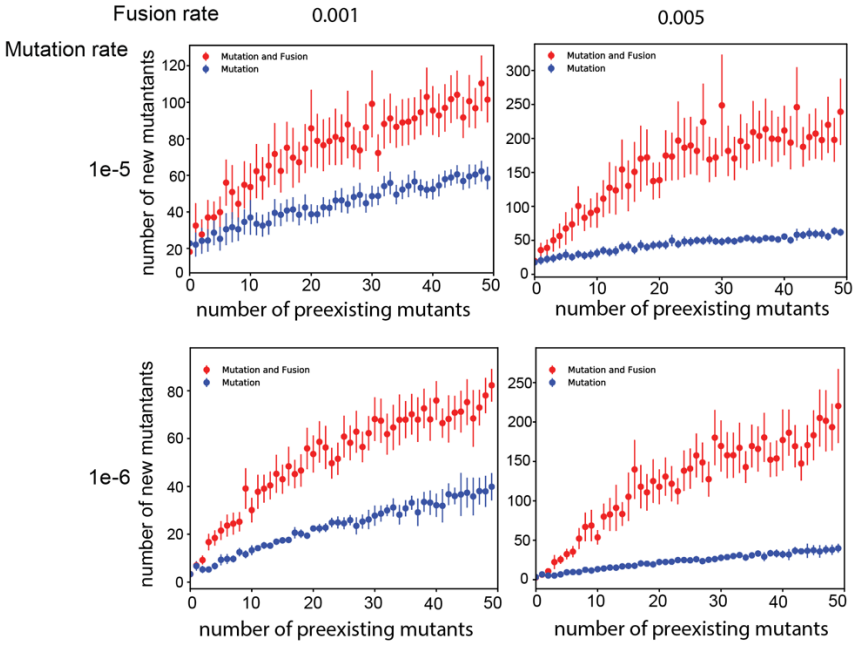
S7. SNP allelic inheritance analyses for hybrids between MDA-MB-231 and SUM159PT cell lines.



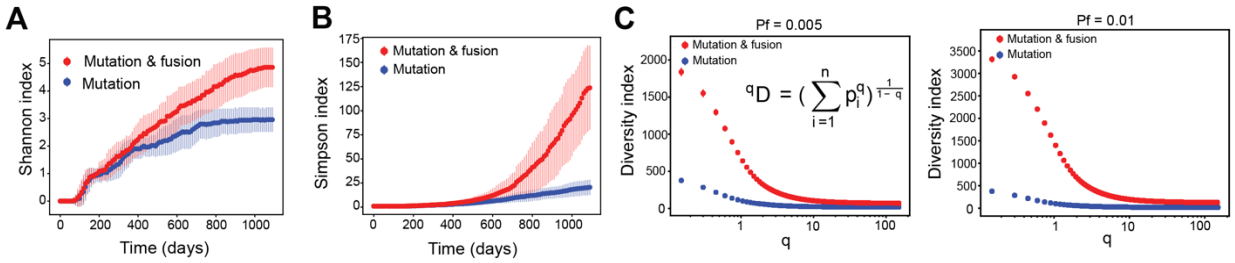
S8. SNP allelic inheritance analyses for hybrids between SUM159PT and MCF10DCIS cell lines.



S9. Plots depicting change in number of unique mutational variants over time for **Fig. 5E** for the indicated fusion and mutation rates.



S10. Impact of the initial clonal richness on the numbers of new mutational variants, after 1095 days of simulations at the indicated mutation and fusion rates. Indicated p values denote the results for Kolmogorov-Smirnov test for those panels exceeding $p<0.05$ statistical significance threshold.



S11. Changes in Shannon (**A**) and Simpson (**B**) diversity indexes for the simulations shown in **Fig. 5B, C**. GDI indexes for the indicated fusion rates; $\mu=1x10^{-4}$.