nature ecology & evolution

ARTICLES

https://doi.org/10.1038/s41559-020-01367-y

Check for updates

Spontaneous cell fusions as a mechanism of parasexual recombination in tumour cell populations

Daria Miroshnychenko[1,6], Etienne Baratchart[2,6], Meghan C. Ferrall-Fairbanks[2], Robert Vander Velde[1,3], Mark A. Laurie[1], Marilyn M. Bui[4], Aik Choon Tan[5], Philipp M. Altrock[2], David Basanta[2,6] and Andriy Marusyk[1,3,6]✉

The initiation and progression of cancers reflect the underlying process of somatic evolution, in which the diversification of heritable phenotypes provides a substrate for natural selection, resulting in the outgrowth of the most fit subpopulations. Although somatic evolution can tap into multiple sources of diversification, it is assumed to lack access to (para)sexual recombination—a key diversification mechanism throughout all strata of life. On the basis of observations of spontaneous fusions involving cancer cells, the reported genetic instability of polypoid cells and the precedence of fusion-mediated parasexual recombination in fungi, we asked whether cell fusions between genetically distinct cancer cells could produce parasexual recombination. Using differentially labelled tumour cells, we found evidence of low-frequency, spontaneous cell fusions between carcinoma cells in multiple cell line models of breast cancer both in vitro and in vivo. While some hybrids remained polyploid, many displayed partial ploidy reduction, generating diverse progeny with heterogeneous inheritance of parental alleles, indicative of partial recombination. 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 modelling, we demonstrated that the observed rates of spontaneous somatic cell fusions may enable populations of tumour cells to amplify clonal heterogeneity, thus facilitating the exploration of larger areas of the adaptive landscape (relative to strictly asexual populations), which may substantially accelerate a tumour's ability to adapt to new selective pressures.

Cancer is the direct result of somatic clonal evolution, which follows Darwinian principles: the diversification of heritable phenotypes provides a substrate on which natural selection can act, leading to the preferential outgrowth of phenotypes with higher fitness in the specific environment[1,2]. The ability to generate new heritable diversity is thus required for the evolvability of populations of tumour cells, both during tumour progression and in response to therapies. Evolving tumours have access to several powerful diversification mechanisms that are considered the enabling characteristics within the hallmarks-of-cancer framework[3]: genomic instability, elevated mutation rates and the 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. In genetically diverse populations, (para)sexual recombination can dramatically amplify diversity and generate new mutational combinations (thus enabling new epistatic interactions), while unlinking advantageous mutations from disadvantageous ones, hence supporting population fitness and accelerating evolutionary adaptation[4–6].

Populations of tumour cells are assumed to be strictly asexual—that is, all novel genetic and epigenetic solutions 'discovered' by tumour cells are thought to be strictly clonal, inheritable only by the direct progeny of (epi-)mutated cells. However, occurrences of spontaneous cell fusions involving tumour cells have been documented both in vitro and in vivo[7–9]. Given the previously reported impact of genome duplication on increased genomic instability[10,11], the evidence of ploidy reduction in the progeny of experimentally induced hybrid cells[12,13], the reported genetic recombination in the asexual ploidy cycle of cancer cells[14] and the existence of parasexual life cycles involving fusion-mediated recombination in fungi (such as the pathogenic yeast *Candida albicans*[15]), we decided to examine whether spontaneous cell fusion involving genetically distinct cells could lead to parasexual diversification in tumour 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 are clonogenically viable. Whereas cell fusion simply combines two genomes, some of the hybrids undergo ploidy reduction that is accompanied by genome recombination, which generates new subclonal diversity. Our in silico modelling suggests that this fusion-mediated recombination could augment the evolvability of tumour cell populations even when spatial limitations are considered. Our studies thus suggest that spontaneous cell fusions may provide populations of tumour cells with a mechanism for parasexual recombination and make them capable of exploring combinations of mutations from different clonal lineages, thus accelerating diversification and enhancing evolvability.

[1]Department of Cancer Physiology, H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL, USA. [2]Department of Integrated Mathematical Oncology, H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL, USA. [3]Department of Molecular Medicine, University of South Florida, Tampa, FL, USA. [4]Department of Pathology, H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL, USA. [5]Department of Biostatistics and Bioinformatics, H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL, USA. [6]These authors contributed equally: Daria Miroshnychenko, Etienne Baratchart, David Basanta, Andriy Marusyk. ✉e-mail: andriy.marusyk@moffitt.org

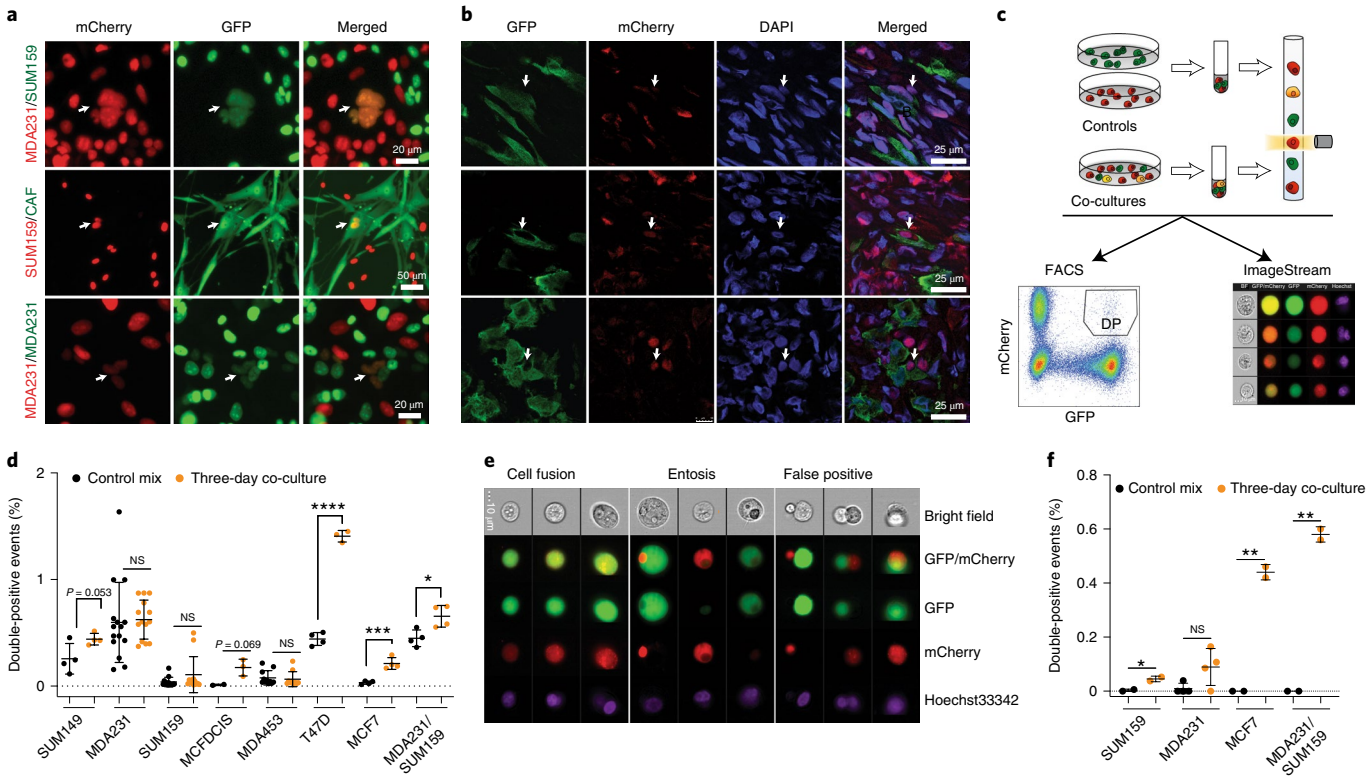


Fig. 1 | Detection of spontaneous cell fusions in vitro and in vivo. a, Live fluorescence microscopy images of co-cultures of the indicated differentially labelled cells. The arrows indicate cells co-expressing both fluorescent labels. **b**, Confocal immunofluorescent images from a xenograft tumour, initiated with a 50/50 mix of GFP and mCherry-labelled MDA-MB-231 (MDA231) cells. The arrows 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 fluorescence-activated cell sorting (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. $^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$; $^{****}P < 0.0001$; NS, not significant; two-tailed unpaired *t*-test. The error bars represent the standard deviations (s.d.) from the mean values of biological replicates, represented by dots.

Results

During multiple experimental studies involving in vitro co-cultures of tumour cells that carry different fluorescent protein labels, we occasionally noticed double-positive cells on fluorescence microscopy analyses (Fig. 1a and Extended Data Fig. 1a). Similarly, we observed double-positive cells in co-cultures of tumour cells and cancer-associated fibroblasts (CAFs) (Extended Data Fig. 1b). The examination of time-lapse microscopy images revealed that these double-positive cells can originate from spontaneous cell fusions (Extended Data Fig. 1c,d and Supplementary Videos 1–5). The phenomenon was not limited to in vitro cultures. Confocal microscopy examination of experimental xenograft tumours also revealed the 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 tumour cells with a new source of genetic diversification, we decided to systematically investigate this phenomenon. To this end, we labelled panels of breast cancer cell lines and primary breast CAF isolates with lentiviral vectors expressing green fluorescent protein (GFP) and mCherry reporters and co-expressing the antibiotic resistance markers 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 three days, subjected to flow cytometry analysis (Fig. 1c). Compared with the separately cultured controls, collected and admixed no more than 30 min before 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 and Extended Data Fig. 1e).

The notably higher proportion of double-positive events detected by flow analysis than that detected by 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. We therefore 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[16]. Indeed, the examination of the images of double-positive gate events (the gating logic is provided in Extended Data Fig. 2a) revealed substantial rates of false positives reflecting cell doublets (Fig. 1e). Some of the double-positive events were cell-within-cell structures, indicating entosis[17] 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 and Extended Data Fig. 2b). Direct comparison of flow cytometry and ImageStream analysis of the same sample revealed that true positives represented ~30% of the double-positive events detected by flow analysis (Extended Data Fig. 2c). Consistent with the expected increase in cell size resulting from the fusion of two

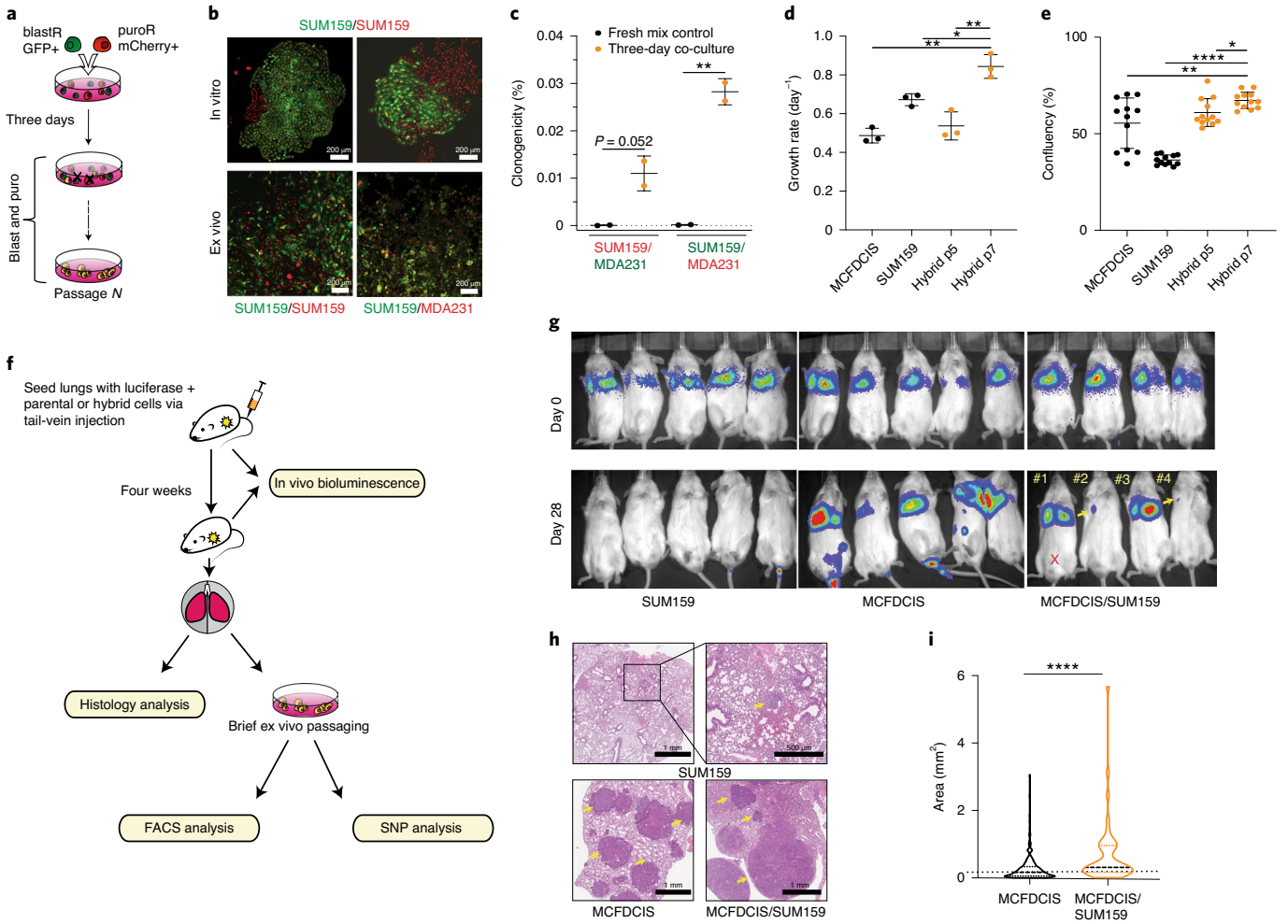


Fig. 2 | Phenotypic characterization of hybrids. a, Experiment schema for the selection of hybrid cells. **b**, Representative images of live fluorescent colonies formed after selection of the indicated in vitro co-cultures or ex vivo tumours. **c**, Quantitation of the frequency of fusions leading to clonogenically viable hybrid cells. **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**, Experiment schema for the analyses of lung colonization. **g**, In vivo bioluminescence imaging of animals injected with mixed SUM159PT/MCFDCIS hybrids or parental cells via tail-vein injection to seed lung metastases. #1–4 indicate individual animals used for subsequent analyses. The red X denotes a mouse that died prior to euthanasia. **h**, Representative images of haematoxylin and eosin (H&E) stains of lungs from the indicated xenograft transplants. The yellow arrows point to example tumours. A magnified image of a micrometastasis is shown in the top right. **i**, Violin plots of the size distributions of individual macrometastatic lung tumours from the analyses of H&E stained histology slides. $N = 215$ (MCFDCIS); $N = 44$ (SUM159/MDFDCIS); the lines are at the medians (dashed) and quartiles (dotted). $^{*}P < 0.05$; $^{**}P < 0.01$; $^{****}P < 0.0001$; two-tailed unpaired *t*-test (**c**–**e**) and Mann–Whitney *U* test (**h**). The error bars represent s.d.; each dot represents a biological replicate.

cells, the double-positive cells were notably larger than the cells expressing a single fluorescent marker (Extended Data Fig. 2d). 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.

We then 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 three days and 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 (Fig. 2b). Notably, the clonogenic proportion of cells with dual antibiotic resistance was lower than the frequency of fusion events (Figs. 1f and 2c), suggesting that only some of the hybrids were capable of sustained proliferation. Similarly, we were able to recover dual-antibiotic-resistant cells expressing both fluorescent markers from ex vivo cultures of xenograft tumours initiated by the co-injection of differentially labelled cells (Fig. 2b). Despite the relatively high rates of fusion detected by flow cytometry (Extended Data Fig. 1e) and microscopy, as well as previous reports on viable hybrids formed by fusions between carcinoma and stromal cells[18,19], we were unable to recover colonies from co-cultures between multiple breast cancer cells and three distinct primary, non-immortalized CAF isolates.

Next, given the prior reports of fusion-mediated increase in invasive and metastatic potential[19,20], we asked whether hybrids formed by spontaneous somatic fusions between cancer cells differ from parental cells in their proliferative and invasive potential. At early passages during the post-antibiotic-selection phase, hybrids displayed lower net proliferation rates than did the parental cell lines (Fig. 2d and Supplementary Fig. 4a). 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 of an increase in proliferation rates with passaging is consistent with the elimination of viable but non-proliferative hybrids along with the selection of variants with higher proliferative abilities (Fig. 2d and Extended Data Fig. 3a). Transmembrane invasion assays revealed that most hybrids displayed invasive rates equal to or exceeding rates of the more invasive fusion parents (Fig. 2e and Extended Data Fig. 3b,c).

We then 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 potential (MCF10DCIS), along with their hybrids, using the tail-vein injection assay (Fig. 2f). Despite identical initial lung seeding efficiencies, mice injected with SUM159PT cells lost luminescent signal from the lungs (Fig. 2g), although post-mortem histological examination revealed the presence of multiple micrometastatic nodules, suggesting a microenvironmental growth bottleneck rather than an inability to seed lungs per se (Fig. 2h). In contrast, luminescent signals in all four mice injected with MCF10DCIS cells and in two out of four mice injected with the hybrid cells increased over time, while the other two hybrid cell recipients displayed a reduced but detectable luminescent signal (Fig. 2g). Histological examination revealed that the lungs of all of the MCF10DCIS recipient and hybrid cell recipient mice contained macroscopic tumours. Surprisingly, despite weaker luminescent signals in two out of three analysed animals, histological examination revealed larger tumours than those in mice transplanted with MCF10DCIS cells (Fig. 2h,i and Extended Data Fig. 3d), probably reflecting the loss of luciferase gene expression in some of the hybrids. One of the mice with a strong luminescent signal (#1 in Fig. 2f) had died prior to euthanasia; necropsy analysis revealed massive tumours in the lungs, but due to poor tissue quality, this animal was excluded from the analysis. Notably, flow cytometry analysis of lungs recovered from the recipients of hybrid cells revealed that the majority of fluorescent cells expressed both GFP and mCherry, including lungs of the animals that displayed a reduction in the luminescent signal (#2 and #4, Fig. 2g and Extended Data Fig. 4b). In summary, consistent with previously reported observations, these data suggest that spontaneous cell fusions between neoplastic cells can generate cells with more aggressive oncogenic properties.

In the absence of the TP53-dependent checkpoint function, which is commonly disrupted in cancer cells, polyploidy is known to be associated with increased genomic instability[11]. Consistently, genomic instability[21] and ploidy reduction[12,13] were reported in experimentally induced somatic hybrids. We therefore decided to examine whether spontaneously formed hybrid cells can maintain stable genomes over time. As expected, at early (1–4) passages, counted after the complete elimination of singe-antibiotic-resistant control cells, all of the examined hybrid cell lines displayed elevated DNA content, consistent with the combined genomes of two parents (Fig. 3b and Extended Data Fig. 4a). However, the average ploidy of three out of seven examined hybrids was evidently reduced with additional passaging (passages 4–10), while the average ploidy of the remaining four hybrids remained seemingly unchanged (Extended Data Fig. 4a). Given that the genomic instability of tumour cells can be enhanced by genome doubling[22], that fusion-mediated recombination and the stochastic loss of parental DNA accompanying ploidy reduction can serve as the mechanism for parasexual recombination in the pathogenic yeast species *C. albicans*[15,23], and that cycles of somatic cell fusions followed by genetic recombination and ploidy reduction have been described to operate in normal hepatocytes[24] and hematopoietic cells[25], we decided to compare the genomes of single-cell-derived subclones of somatic hybrids (the derivation schema is shown in Fig. 3a). Consistent with the maintenance of polyploid genomes in mixed populations, all of the examined subclonal derivatives of SUM159PT/MCF10DCIS and MDA-MB-231/MCF10DCIS hybrids retained elevated DNA content (Extended Data Fig. 4a). In contrast, the genomes of individual subclones from the hybrids with reduced ploidy (MDA-MB-231/Hs578T and MDA-MB-231/SUM159PT) displayed substantial variation in DNA content (Fig. 3b and Extended Data Fig. 4a), suggestive of genomic diversification.

To gain deeper insights into the impact of hybridization on genomic diversification, we compared patterns of inheritance of cell-line-specific alleles. To this end, we characterized single nucleotide polymorphism (SNP) profiles of the genomically unstable MDA-MB-231/Hs578T hybrids and the parental cell lines using the Affymetrix CytoScan SNP platform. Focusing on homozygotic cell-line-specific SNPs that differentiate the parental cells, we characterized their inheritance in individual single-cell-derived hybrid subclones, isolated from mixed populations of hybrid cells (Fig. 3c). Genome-wide analysis of allelic inheritance revealed that for the majority of the SNPs that discriminate the parental cell lines, the subclones showed mixed inheritance. However, in $44.9 \pm 0.4\%$ of the analysed loci, only MDA-MB-231-specific alleles were detected; in a smaller fraction ($0.62 \pm 0.26\%$), only Hs578T unique alleles could be detected (Fig. 3d and Supplementary Table 1). Importantly, differential inheritance was not confined to whole chromosomes or large chromosomal regions. Instead, we observed a notable mosaicism in the SNP inheritance within individual chromosomes, with substantial variability in the degree of mosaicism between individual chromosomes (selected examples are shown in Fig. 3e, and a complete set of chromosomes is shown in Supplementary Fig. 2). Individual subclones displayed distinct patterns of inheritance of parent-specific alleles. Although a high degree of genomic rearrangements within parental cell lines complicates the analyses, mosaic inheritance of parental SNPs across individual chromosomes and substantial variability between distinct subclones strongly suggest that ploidy reduction has been accompanied by partial genomic recombination.

We observed similar patterns of mosaic loss of parent-specific alleles as well as variability in patterns of inheritance between distinct subclones in MDA-MB-231/SUM159PT hybrids, analysed with the Illumina CytoSPN-12 platform (Fig. 3d,f and Supplementary Fig. 3). Notably, the analysis of the SUM159PT/MCF10DCIS hybrids (where both fusion parents have relatively stable, near-diploid genomes and the hybrid populations do not show obvious signs of ploidy reduction; Extended Data Fig. 4a) also revealed the loss of some of the parent-specific alleles, as well as divergence in allelic inheritance between the two distinct subclones that we have analysed, although to a lower extent than that in the MDA-MB-231/Hs578T and MDA-MB-231/SUM159PT hybrids (Fig. 3d,g and Supplementary Fig. 4). Interestingly, the analysis of hybrid cells (recovered from the tumour-bearing lungs of a mouse (#3 shown in Fig. 2g) injected with pooled SUM159/MCF10DCIS hybrids) revealed distinct and more extensive patterns of inheritance of parental SNPs (Supplementary Fig. 1), potentially reflecting the impact of distinct selective pressures experienced by cells in vivo.

While we observed a substantial interclonal variability in patterns of mosaic SNP inheritance, a large fraction of SNPs displayed identical patterns of inheritance in distinct subclones. Given that the numbers of distinct SNP alleles can vary between highly aneuploid genomes of different cancer cell lines, this unequal contribution could be the most parsimonious explanation for the observed similarities in the patterns of SNP inheritance. Should this be the case, more numerous alleles would be more likely to be retained in hybrids under stochastic ploidy reduction. Additionally, this numeric inequality could lead to detection issues, where numerically superior alleles from one parent could mask the signal from less numerous alleles from another parent. Differences in the numbers of cell-line-specific alleles inherited from different

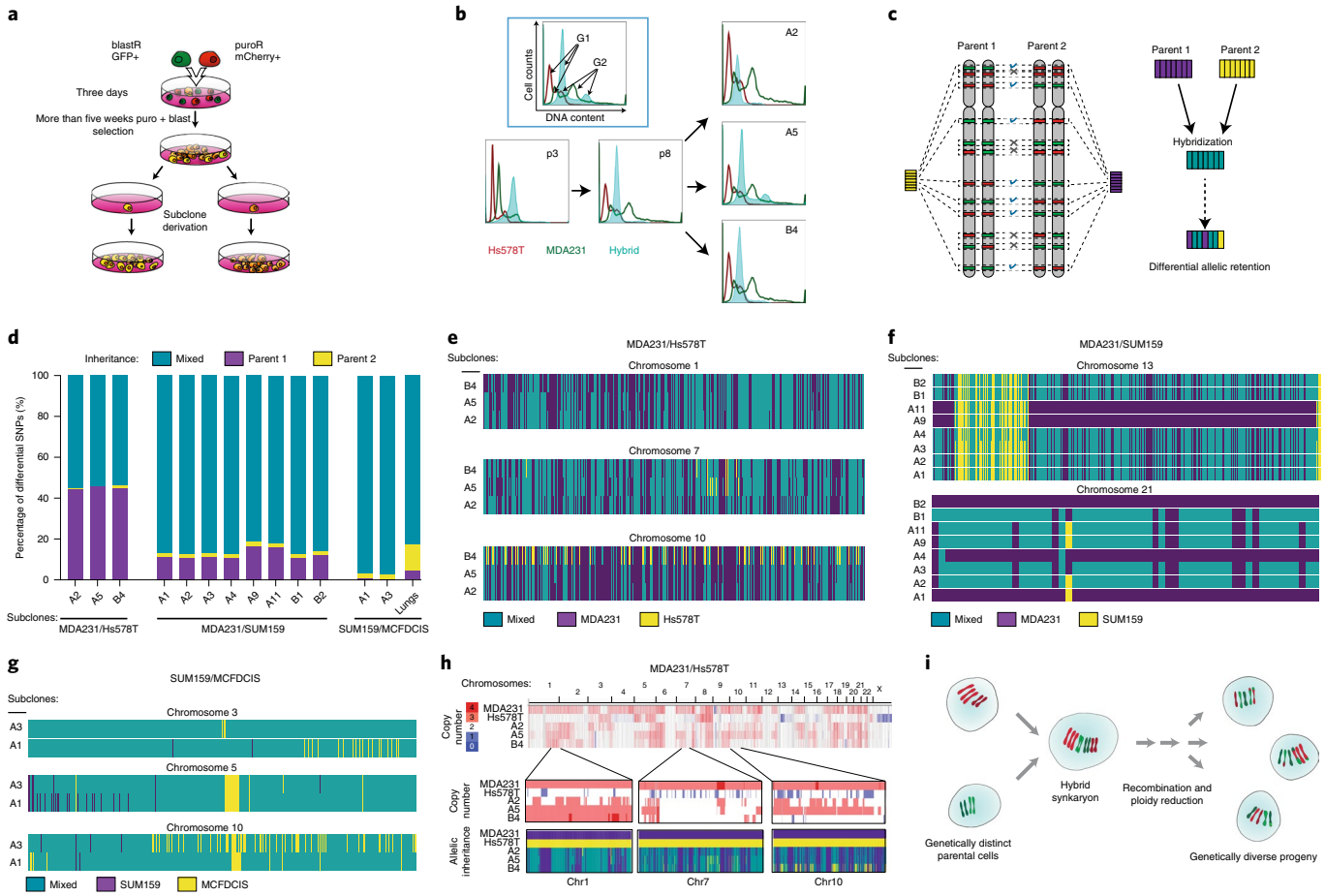


Fig. 3 | Fusion-mediated genetic diversification. a, Experiment schema for the derivation of hybrid subclones. **b**, DNA content analysis of parental cell lines (in this example, HS578T and MDA-MB-231) and hybrids. Propidium iodide (PI)-based DNA content profiles of the hybrids are superimposed with profiles of the parent cells. The inset describes the axes and indicates the positions of the G1 and G2 peaks. A2, A5 and B4 denote distinct hybrid subclones; p3 and p8 refer to the passage number of the mixed hybrid population. The shift of the DNA content profile to the left with extended passages indicates a reduction of the average cell ploidy in the hybrids. The differences in the profiles of individual hybrid subclones indicate diversification in DNA content. **c**, Experiment schema of the pipeline for the analysis and visualization of SNP inheritance. **d**, Summary of analyses of inheritance of cell-line-specific SNP alleles across the indicated hybrid subclones. 'Lungs' refers to the mixed population, isolated from the colonized lungs of mouse #3 from Fig. 2g and Supplementary Fig. 1. **e**–**g**, Inheritance of cell-line-specific alleles mapped to specific chromosomes in the hybrid subclones of the indicated hybrids. The columns represent individual alleles, and the colours indicate mixed or parent-specific inheritance. Maps for all of the individual chromosomes are provided in Supplementary Figs. 1–5. **h**, Analyses of copy numbers for the cell-line-specific alleles in parental cell lines and hybrid subclones. The selected zoom-ins of illustrative chromosomal regions depict the correspondence between copy numbers and inheritance of cell-line-specific SNPs. **i**, Schema of a proposed fusion-mediated diversification.

parental cell lines might also explain the apparent dominance of the MDA-MB-231 cell lines in the two hybrids that we have analysed by SNP arrays. To address whether distinct copy numbers of differential SNP alleles could explain the observed inheritance patterns, and to examine the variegation in allelic copy numbers between distinct subclones, we contrasted copy number data with allelic inheritance data across several chromosomal regions in MDA-MB-231/Hs578T hybrids. The dominance of MDA-MB-231-exclusive and mixed allelic inheritance in MDA-MB-231/Hs578T hybrids was generally consistent with higher copy numbers of differential SNPs inherited from MDA-MB-231 cells (Fig. 3h). Importantly, we observed substantial variability in allelic copy numbers between distinct subclones of the hybrids, indicating additional diversification. Some of this variability was consistent with differences in allelic inheritance. For example, lower copy numbers from the region of chromosome 10 in subclone B4 were linked with distinct patterns and higher proportions of Hs578T alleles (Fig. 3h). However, the majority of similarities and dissimilarities in the patterns of allelic inheritance within different hybrid subclones could not be fully explained by differences in allelic copy numbers between the genomes of different parental cell lines, suggesting the contribution of additional factors. The analysis of the more stable SUM159/MCFDCIS hybrids revealed similar clonal variegation in allelic copy numbers and patterns of allelic inheritance that could be partially explained by unequal copy numbers of SNP alleles in the parental cells (Extended Data Fig. 5a).

The above analyses were performed on subclones derived from the same pool of hybrids. Therefore, the distinct subclones could have been the progeny of same original fusion. To test whether similar patterns of allelic inheritance could be observed in independently derived hybrids, we derived three new hybrid subclones from each of the two independent mixed populations of MDA-MB-231/SUM159 hybrids. Despite the substantial variegation in allelic inheritance and copy numbers within distinct subclones, most of the conserved patterns of allelic inheritance were shared between subclones, derived from distinct hybrid parents (Extended Data Fig. 5b

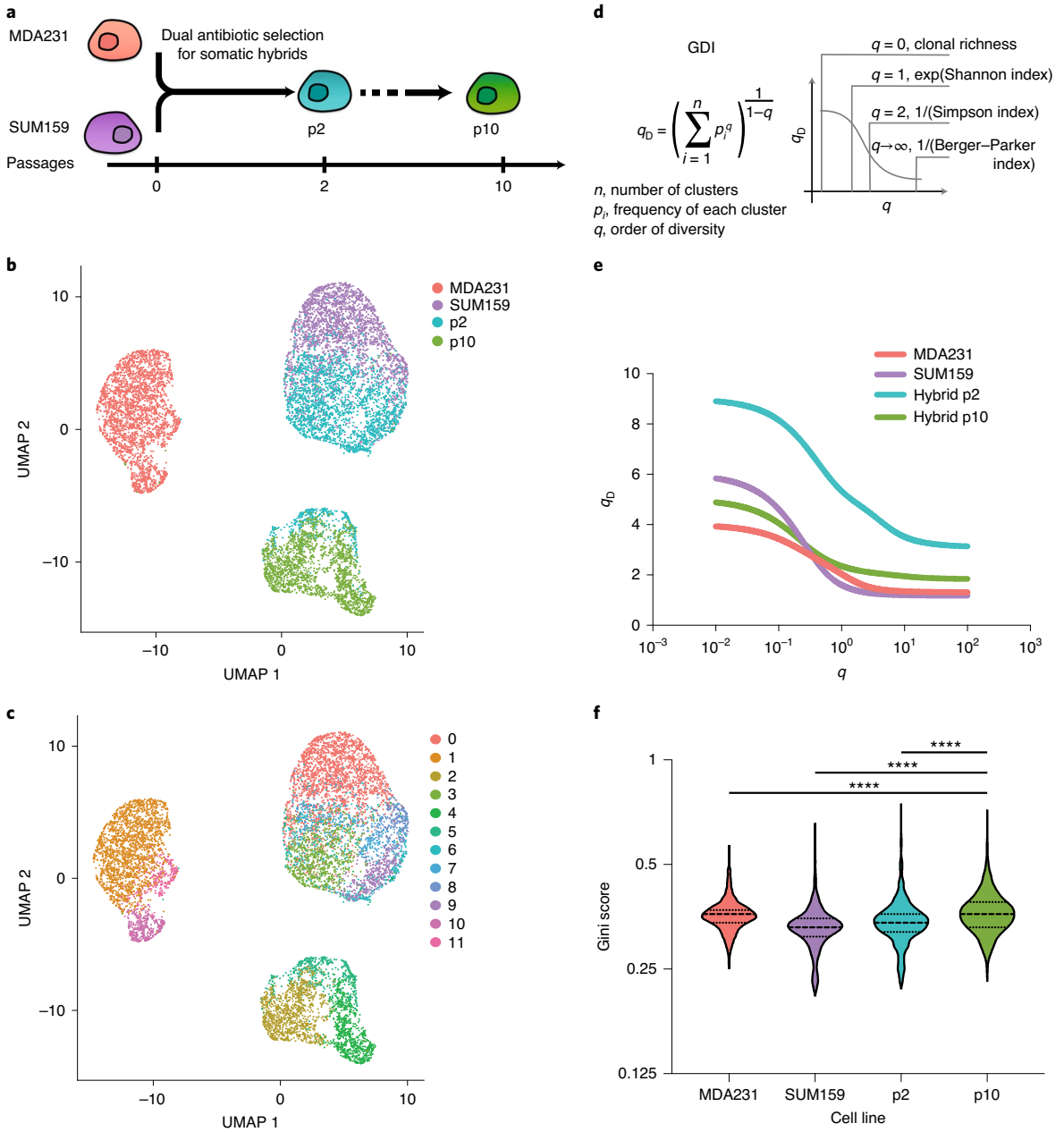


Fig. 4 | Fusion-mediated phenotypic diversification. a, Experimental schema for the samples used for the scRNAseq analyses. **b**, UMAP distribution of cell phenotypes of parental cells and hybrids at the indicated passages from the single-cell expression analysis data. **c**, UMAP-defined distinct phenotypic clusters (0–11) used in the GDI analyses. **d**, Formula for the calculation of a GDI, and mappings to common diversity indexes that are special cases. **e**, GDI analysis of the 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. The dashed lines represent the medians, and the dotted lines represent the quartiles. $N = 1{,}249$; $^{****}P \leq 0.0001$; Wilcoxon signed-rank test.

and Supplementary Fig. 5). These recurrent patterns were generally consistent with the unequal contribution of SNP copy numbers from parental genomes. However, similar to the observations in the MDA-MB-231/Hs578T hybrids, the unequal contribution of allelic copy numbers could not fully explain the observed patterns (such as mosaic patterns of mixed and SUM159-specific inheritance within chromosome 13 in Extended Data Fig. 5b), suggesting the contribution of additional mechanisms.

In addition to genetic diversification, heterogeneity in biologically and clinically important phenotypes of cancer cells is shaped by epigenetic mechanisms[26]. Theoretical studies have suggested that cell fusions between genetically identical but phenotypically distinct cells could create remarkable diversity due to the resultant collision of gene expression networks[27]. We therefore 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 cells, SUM159PT cells and MDA-MB-231/SUM159PT hybrids at early (2) and extended (10) passages under dual antibiotic selection (Fig. 4a). Uniform manifold approximation and projection (UMAP) clustering[28] of single-cell expression profiles revealed that the phenotypes of hybrid cells were distinct from those of 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 subpopulation of hybrid cells and additional diversification (Fig. 4b,c).

To quantify the phenotypic diversity within parental cell types and hybrids from the single-cell profiling data, we used a generalized diversity index (GDI) (Fig. 4d)[29]. The GDI enables the characterization of diversity across a spectrum of orders of diversity[30], ranging from clonal richness (a low order of diversity reveals the number of distinct subpopulations) to classic measures of species diversity, such as the Shannon and Simpson indexes[31] (intermediate orders of diversity) and to high orders of diversity that give increased weight to the highly abundant subpopulations[29]. Considering individual UMAP-defined clusters as subpopulations ('species'), we found that at the early passage, hybrids displayed higher diversity across all orders of diversity (Fig. 4e). However, at passage 10, the diversity at low orders ('species richness') decreased. Yet, at intermediate and high orders ('species evenness'), the diversity of late passages remained higher than in either of the parental cells.

Our GDI analyses rely on grouping phenotypes into distinct clusters. However, these analyses might miss lower-level cell-to-cell phenotypic variability. We therefore interrogated the dispersion of transcript reads across cells using the Gini dispersion index, which captures the variability of gene expression across all of the transcriptome and has been recently applied towards the characterization of phenotypic diversification in cancer cell populations[32]. We found that the MDA-MB-231/SUM159PT hybrids displayed elevated Gini indexes compared with both parents; in contrast to the GDI metrics, lower-level phenotypic diversity increased at the later passage (Fig. 4f). These findings further support the notion that somatic hybridization can lead to phenotypic diversification.

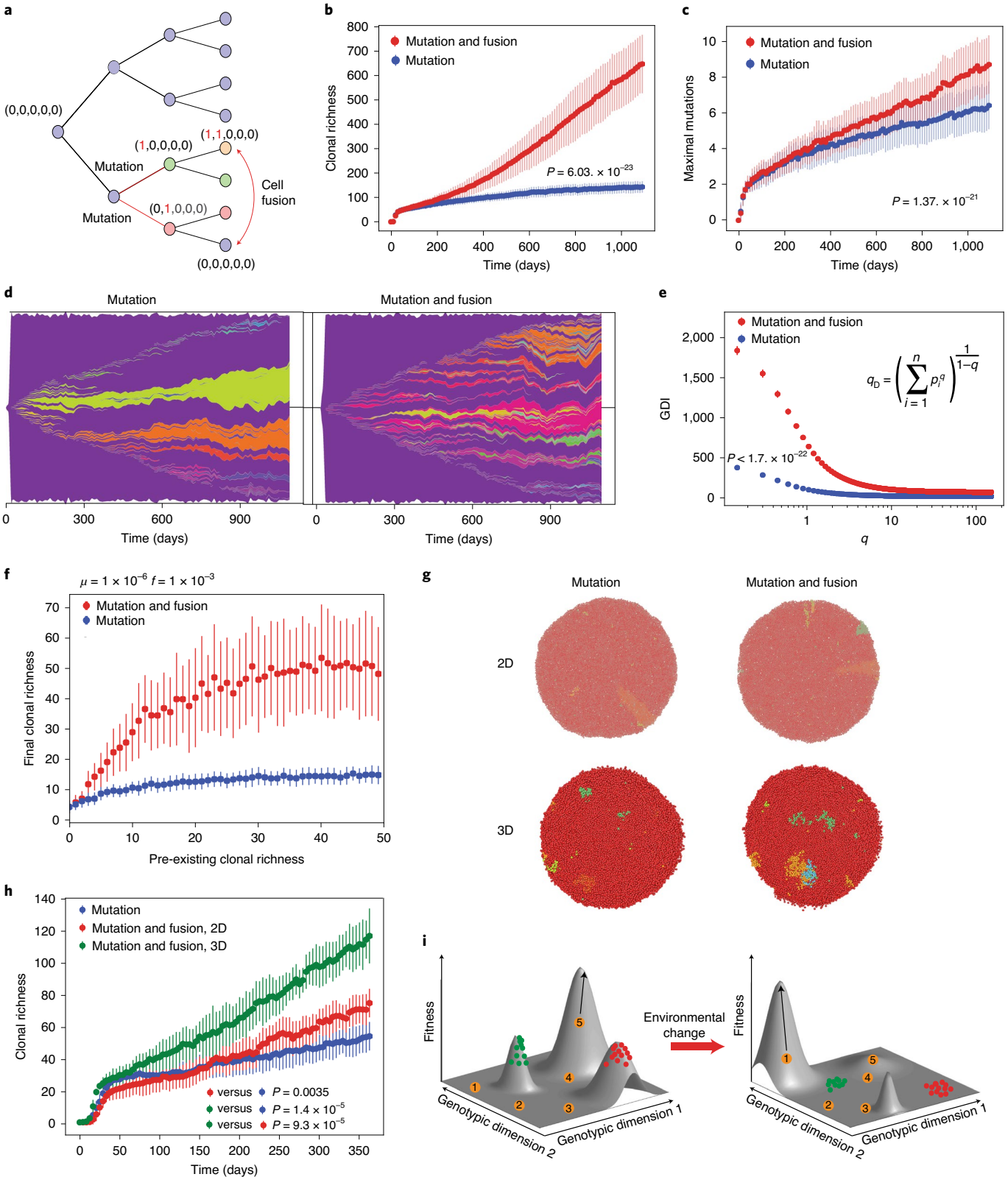
Despite the substantial genetic and phenotypic diversification observed in hybrid cells and their progeny, spontaneous fusion events are relatively rare. It is easy to intuit the potential impact of rare events creating cells with dramatically enhanced oncogenic properties, but the impact of low-frequency fusion-mediated recombination events on mutational diversity within tumour cell populations is less obvious. To evaluate this impact, we used in silico simulations based on a birth–death branching model of tumour growth (Fig. 5a and Supplementary Mathematical Methods). We started by estimating the probabilities of clonogenic cell fusions and cell proliferation from our experimental data. In these estimations, we accounted for the fact that only fusions between cells labelled with distinct markers could be detected, and that in some experiments, the proportions of cells with distinct labels were unequal (Supplementary Mathematical Methods). Our analyses suggest that the median probability of a cell fusing with another cell is 6.6×10^{-3} (range, 1.6×10^{-4}–1.36×10^{-1}) in vitro and 6.6×10^{-5} (range 3.8×10^{-15}–4.1×10^{-4}) in vivo (Extended Data Fig. 6a,b and Supplementary Tables 2 and 3). Using these estimates, together with values of genetic mutation rates of 10^{-6} to 10^{-3} described in the literature[33] as well as the number of potential driver genes of 300, we compared the accumulation of diversity between the scenarios of populations of tumour cells evolving through mutations only and those evolving through mutations and fusion-mediated recombination. We found that fusion-mediated recombination can substantially enhance the increase in clonal richness (groups of tumour cells defined by unique mutational combinations) while also increasing the maximum numbers of mutations observed within a single lineage (Fig. 5b–d). Although the impact of fusion-mediated recombination was clearly captured by the commonly used Shannon and Simpson diversity indexes (Extended Data Fig. 6c,d), the GDI enabled a more informative evaluation of the effect, showing the highest impact at the lowest orders of diversity (Fig. 5e and Extended Data Fig. 6e).

This effect was observed across a biologically feasible range of fusion and mutation rates (Extended Data Fig. 7). While larger population sizes and higher rates of proliferation and turnover of tumour cells predictably led to increased clonal diversity, they did not substantially modulate the impact of fusion-mediated recombination (Extended Data Fig. 8a,b). As expected, in all cases, 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. To better understand the 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. 5f and Extended Data Fig. 8c). The biological impact of fusion-mediated recombination should therefore be more pronounced in tumours with higher levels of mutational clonal heterogeneity.

In our modelling thus far, we assumed a well-mixed population. However, this assumption is clearly violated in solid tumours, where spatial restrictions can have a profound impact on selective pressures and expansions of mutant subpopulations[34–37]. We thus asked whether fusion-mediated recombination could still have a substantial impact in spatially restricted contexts, where fusions between genetically dissimilar cells are less likely. Using spatial agent-based simulations (Extended Data Fig. 9 and Supplementary Mathematical Methods), we compared the diversification of spatially restricted populations that evolve through mutations and cell fusions versus those evolving through mutations only, in both two-dimensional (2D) and 3D contexts (Fig. 5g and Supplementary Videos 6–9). We found that even in spatially restricted contexts, diversity was still acquired faster in the presence of fusion-mediated

Fig. 5 | Impact of fusion-mediated recombination on genetic diversity in tumour cell populations. a, Schema of the in silico model using 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. **b**,**c**, Dynamics of unique mutation acquisition (**b**) and the highest number of unique mutations within a single lineage (**c**) in the presence and absence of fusion-mediated recombination (fusion probability, $f=0.005$; mutational probability, $\mu=5\times10^{-5}$) over 1,095 days. In **b** and **c**, the solid lines represent the means and the error bars represent the standard deviations of 50 simulations. **d**, Muller plots depicting the impact of fusion-mediated recombination on clonal dynamics, with f either 0 (no fusions) or 0.01, and $\mu=1\times10^{-5}$. **e**, GDI after 1,095 days of simulation; $f=0.005$; $\mu=1\times10^{-4}$. **f**, Impact of the initial genetic heterogeneity (clonal richness) on the ability of fusion-mediated recombination to further diversify tumour cell populations after 1,095 days of in silico tumour growth at the indicated mutation and fusion rates. **g**, Visualization of the spatial distribution of subpopulations carrying unique genotypes at the end of an in silico simulation of growth for 60 days; $f=0.00035$; $\mu=2.5\times10^{-4}$; carrying capacity, 10,000 cells. **h**, Number of distinct mutants over time for the spatial simulations for 365 days; $f=0.00035$; $\mu=2.5\times10^{-4}$; carrying capacity, 10,000 cells. The solid lines represent the means over 10 seeds, and the error bars are the standard deviations. All P values show the results of Kolmogorov–Smirnov tests for the indicated comparisons. **i**, A model schema illustrating the proposed impact of fusion-mediated recombination on the ability of populations of tumour cells to explore adaptive landscapes. Under constant conditions, we assume that major genetic subclones 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 (left). Environmental change (such as the initiation of therapy) changes the adaptive landscape. Some of the new variants produced by fusion-mediated recombination might 'discover' new fitness peaks (right).

recombination. Notably, the impact of cell fusions was substantially higher in 3D than in 2D contexts, reflecting the higher number of neighbours, which increased the probability of having a genetically distinct neighbour (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 the accelerated diversification of tumour cell populations and the generation of rare mutational variants capable of exploring larger swathes of adaptive landscapes (Fig. 5i).



Discussion

Our findings reveal a surprisingly common occurrence of spontaneous cell fusions across multiple experimental cell-line models of breast cancer. Spontaneous cell fusions between cancer cells and normal cells in the tumour microenvironment have been described by multiple prior studies from many groups[18,19,38–41]. Although the subject of fusions between cancer cells has received less attention, spontaneous fusions between cells have also been reported[20]. These studies have documented the impact of cell fusions on creating and enhancing malignant phenotypes, such as invasion, migration and metastatic dissemination, thus providing the basis for the argument that cell fusions can be an important contributor to cancer initiation and progression[7–9]. Our results are consistent with these prior findings. However, in addition to the previously proposed ability to generate cells with novel oncogenic properties, our results suggest the presence of a further, indirect impact with potentially profound implications. Given the evidence of genetic recombination and phenotypic diversification resulting from spontaneous cell fusions between cancer cells, we posit that this fusion-mediated recombination and partial ploidy reduction can provide populations of tumour cells with a parasexual recombination mechanism, similar to the fusion-mediated facultative parasexual recombination observed in the pathogenic yeast *C. albicans*[15]. Using mathematical modelling, we demonstrate that despite its relatively low frequency, this fusion-mediated recombination can have a strong impact on promoting evolution in cancer cell populations.

Cell fusions have been previously proposed to induce oncogenic transformation and tumour initiation through genome destabilization[13,21,41]. A large fraction of primary cancers show evidence of whole-genome doubling followed by partial genome loss[42], although the contribution of cell fusions towards these tetraploid intermediates is unclear, as genome doubling can be also achieved by endoreduplication and aborted cell cycle[22]. Experimental studies have linked increased ploidy with elevated genomic instability, reflecting a higher tolerance of chromosome aberrations[43]. Therefore, regardless of whether genome doubling is the result of cell fusions, endoreduplication or aborted cell cycle, it can lead to increased genome diversification, thus fuelling cancer evolution. Indeed, whole-genome doubling in primary tumours has been associated with significantly worse outcomes[43]. Nevertheless, fusions between genetically different tumour cells could have additional consequences to those arising from the duplication of the same genome or fusions between cancer and normal cells. Spontaneous fusions between genetically different tumour cells can combine mutations that have been acquired within independent evolutionary trajectories. Subsequently, the stochastic loss of some of the genetic material during ploidy reduction could create a variety of combinations of mutations from the genomes of fusion partners (Fig. 3i).

The extent of this additional diversification depends on the genetic divergence of parental cells, as well as on the extent of genomic recombination between the parental genomes. The latter can theoretically range from a complete lack of recombination, where ploidy reduction is achieved through losses of whole chromosomes, to an extensive blending observed in meiotic processes. Our analyses of the inheritance of cell-line-specific SNPs reveal mosaicism across chromosomes, which strongly indicates recombination at a level within these two extremes. Intriguingly, the extent of this mosaicism is highly variable between individual chromosomes (for example, compare chromosomes 13 and 21 in the MDA-MB-231/SUM159PT hybrids depicted in Fig. 3f). This extensive mosaicism, confined to a subset of chromosomes, as well as the oscillating DNA copy number patterns in genomes of hybrid subclones (Fig. 3h) indicate a potential occurrence of chromothripsis—a single mutational event that involves the shattering a single or several chromosomes into a large number of pieces and the rejoining of these pieces in different patterns[44], which would further increase genomic diversity following cell fusions. Our results thus support the notion of partial recombination with unequal impacts in different areas of the genome.

Sexual and parasexual recombination, mediated by a plethora of diverse, independently evolved mechanisms, is near universal across all strata of life, highlighting its essential importance for species evolution[45]. Of note, the frequency of clonogenic cell fusion that we have documented in vitro is similar to the frequency of parasexual recombination in yeast species with a facultative parasexual recombination life cycle[46], supporting the notion of the potential importance of fusion-mediated recombination in somatic evolution. Indeed, our in silico modelling experiments suggest that the cell fusions might significantly accelerate diversification within tumour cell populations. As fusion-mediated recombination can break, combine and reshuffle mutations from different subclonal lineages, its impact should be the strongest in tumours with higher levels of clonal mutational heterogeneity. By enabling cancer cell populations to explore epistatic interactions between mutations that are strictly confined to clonal lineages in asexual populations, fusion-mediated recombination facilitates the exploration of larger areas of the adaptive landscape, thus augmenting the population's evolvability.

In our in silico simulations, the impact of fusion-mediated recombination on genetic heterogeneity was observed over a wide range of mutation rates and fusion probabilities. However, models present simplifications and abstractions of biological reality (for example, we assumed diploid genomes, no explicit fitness effect of mutations and so on) as well as uncertainty over model assumptions and parameterization (including fusion probabilities, numbers of potentially relevant mutations and mutation rates). These results should therefore be treated as theoretical suggestions rather than quantitative proofs. It should be noted that our estimations of fusion probabilities in vivo were substantially lower than the estimations of fusion probabilities in vitro. This discrepancy might reflect methodological differences, but it might also reflect biological reality. Still, we observed a significant impact of fusions on diversification even with fusion probabilities within the in vivo estimated range (Extended Data Figs. 6 and 7). Notably, our simulations did not consider the ability of fusion-mediated recombination to unlink deleterious mutations from beneficial ones. Additionally, our simulations were initiated from single tumour cells and ran over limited time scales, thus probably underestimating the degree of clonal mutational heterogeneity that can feed fusion-mediated diversification. Our modelling might therefore have under- or overestimated the potential impact of fusions on evolutionary dynamics. A more rigorous qualitative analysis would require refinement of the model's biological assumptions and higher confidence parameterization.

In addition to genetic diversification (Fig. 3), we have observed an increase in the phenotypic diversity of hybrid cells (Fig. 4). Although 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 the destabilization of gene regulatory networks, increasing phenotypic entropy and enabling cells to reach phenotypic states distinct from those of the two parents[27]. In support of this view, cell fusions have been associated with nuclear reprogramming[47] and increased phenotypic plasticity in multiple contexts[48], while increased migration and invasiveness of hybrid cells (Extended Data Fig. 3) have been linked with stemness[49]. Given the emergent importance of epigenetic mechanisms in metastatic colonization, we speculate that this increase in epigenetic plasticity (rather than genetic diversification or the inheritance of invasion/metastatic properties from one of the fusion parents) probably underlies the increased metastatic potential of hybrid cells. Therefore, cell fusions might strongly enhance tumour cell heterogeneity and cancer progression independently of genetic diversification.

Multiple prior studies have described the occurrence of spontaneous, clonogenic cell fusions between neoplastic cells and non-neoplastic cells within the tumour microenvironment, including fibroblasts[18,38,40] and endothelial[39] and myeloid[19,47] cells both in vitro and in vivo. While we observed relatively high (compared to fusions between carcinoma cells) rates of cell fusions between breast cancer cells and CAFs (Extended Data Fig. 1e), we failed to derive clonogenic progeny from these heterotypic hybrids. Since non-immortalized primary CAFs remain genetically normal[50], the lack of proliferation might reflect the dominant impact of the intact TP53 checkpoint to limit the proliferation of polyploid cells[11,51]. Indeed, the 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 (an example of the latter is shown in Extended Data Fig. 1d). While these results do not necessarily contradict the possibility of the formation of proliferatively viable hybrids between cancer and non-cancer cells in the tumour microenvironment (as documented in multiple prior studies[18,19,38–41]), they suggest that spontaneous somatic cell fusions between cancer cells that have lost intact checkpoint mechanisms to preserve genomic integrity might be more likely to generate proliferatively viable progeny and thus more likely to contribute to diversification within cancer cell populations (Fig. 5i).

At this point, we cannot completely exclude the possibility that our observations of spontaneous cell fusions in in vitro and xenograft models might be irrelevant to primary human cancers. Nevertheless, the occurrence of viable spontaneous cell fusions has been documented in vivo in multiple animal models[19,38,39,41]. Cells with phenotypes consistent with hybrids between cancer cells and leukocytes have been detected in circulation in human malignancies[19]. 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. Although giant polyploid cells consistent with spontaneous fusions can be observed in many human neoplasms[52,53], the absence of genetic markers makes it currently impossible to discriminate cell fusions from endoreduplication or aborted cell division. Strikingly, two case reports have documented the occurrence of cancers that combine genotypes of donor and recipient genomes in patients that have received bone marrow transplantations[54,55]. These findings provide direct evidence for the potential clinical relevance of spontaneous cell fusions. Unfortunately, these cases are too rare for systematic interrogation at this point. Still, given the abovementioned studies documenting spontaneous cell fusions in human malignancies, we postulate that our results warrant the suspension of the notion that cancers are strictly asexual. Rigorous testing of the clinical relevance of spontaneous cell fusions would require the development of not only new detection approaches but also new experimental and bioinformatical pipelines to understand the degree of fusion-mediated recombination. These efforts might be well warranted, as they might have profound implications for the evolvability of tumour cell populations should fusion-mediated genetic recombination and phenotypic diversification prove to be relevant for human malignancies, and evolvability underlies both clinical progression and the acquisition of therapy resistance.

Methods

Cell lines and tissue culture conditions. The breast cancer cell lines were obtained from the following sources: MDA-MB-231, HCC1937, HS578T, T47D and MDA-MB-453 from ATCC MCF10DCIS.com (MCFDCIS) from F. Miller (Karmanos Cancer Institute, Detroit, MI), and SUM149PT from S. Ethier (University of Michigan, Ann Arbor, MI). The identities of the cell lines were confirmed by short tandem repeats analysis. The CAFs were derived from primary tumours and cultured in SUM medium as described before[56]. All cell lines were routinely tested for mycoplasma infection. The cell lines MDA-MB-231, SUM159PT and MDA-MB-453 were grown in McCoy medium (ThermoScientific) with 10% FBS (Life Technologies); T-47D and HCC1937 were grown in RPMI-1640 medium (ThermoScientific) with 10% FBS and 10 μg ml^{-1} human recombinant insulin (ThermoScientific); MCF10DCIS was grown in DMEM/F12 supplemented with 5% horse serum (ThermoScientific), 10 μg ml^{-1} human recombinant insulin, 20 ng ml^{-1} EGF (PeproTech), 100 ng ml^{-1} cholera toxin (ThermoFisher) and 5 μg ml^{-1} hydrocortisone (Sigma); MCF7 and HS578T were grown in DMEM/F12 with 10% FBS and 10 μg ml^{-1} human recombinant insulin; and SUM149PT was grown in SUM medium (1:1 mix of DMEM/F12 and Human Mammary Epithelial Cell Growth Medium (Sigma), 5% FBS, 5 μg ml^{-1}). Fluorescently labelled derivates of the 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 C. Mitsiades, DFCI).

For the fusion assays, 50/50 mixes of cells expressing differentially expressed fluorescent and antibiotic resistance markers were seeded into 6 cm or 10 cm tissue culture dishes (Sarstedt) in a 50/50 mixture of DMEM-F12 and 10% FBS/MEGM with supplements. Following co-culture for three days, the cells were collected and subjected to flow cytometry analysis or replated for dual antibiotic selection with 10 μg ml^{-1} blasticidin and 2.5 μg ml^{-1} puromycin. Parental cells expressing a single antibiotic resistance marker were used as negative controls. During the 7–14 days required to eliminate all of the cells in the single-resistance control plates, cells from mixed populations were replated once to relieve contact inhibition. After surviving cells from the mixed cultures approached confluency under double antibiotic selection, they were considered as passage 0. For each subsequent passage, 5×10^5 cells were replated into 10 cm dishes and grown until ~90% confluency (5–7$\times10^6$ per dish). Antibiotic selection was maintained for these additional passages. For the derivation of subclones, cells from mixed hybrid cultures were seeded into 96-well plates at a seeding density of <1 cell per well. The day after seeding, the wells were examined, and those containing >1 cells were excluded. On reaching >50 cells, the colonies were trypsinized and replated into 6-well plates and then into 10 cm dishes. On reaching confluency, the cells were divided for the DNA content and SNP inheritance analyses.

Flow cytometry analyses. For the detection of hybrids, cells from monocultures or co-cultures of differentially labelled cells were collected with 0.25% Trypsin (ThermoFisher) and resuspended in PBS with 0.1 μg ml^{-1} DAPI (Sigma). For the negative controls, cells from monocultures were mixed immediately after collection and kept on ice. The flow cytometry analyses were performed using a MACSQuant VYB cytometer (Milteniy Biotec), and the data were analysed using FlowJo 10.5.0 software. The 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^{-1} Hoechst33342 (ThermoFisher) before collection. The analyses were performed with an Amnis ImageStream X Mark II imaging flow cytometer (Amnis, Luminex) using IDEAS 6.0 software (Amnis, Luminex). The average number of collected events was 10,000.

For the DNA content analyses, 10^6 cells were resuspended in 500 μl of PBS, and then 4.5 ml of ice-cold 70% ethanol was added dropwise. The cells were kept at −20 °C for at least two hours, then washed in PBS twice and resuspended in 300 μl of PBS with 0.1% Triton X-100 (ThermoFisher), 0.2 mg ml^{-1} RNAse (Qiagen) and 20 μg ml^{-1} propidium iodide (ThermoFisher). The flow cytometry analysis was performed using a 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. The live cell images were acquired with an Axioscope microscope with an A-Plan ×10/0.25 Ph1 objective and AxioCam ICm1 camera (Zeiss) using ZEN 2.1 software (Zeiss). The time-lapse videos were generated with the IncuCyte live cell imaging system (Sartorius) using a ZOOM ×10 objective for the breast cancer cell mixes and a ZOOM ×4 objective for the breast cancer cell and fibroblast mixes. Images were acquired in the red and green fluorescent channels as well as the visible light channel every three hours for four to five days. For the immunofluorescent detection of GFP and mCherry in xenograft tumours, formalin-fixed, paraffin-embedded tumours were cut at 5 μm sections. Deparaffinized tissue slices were blocked in PBS with 10% goat serum for 30 min at room temperature, then incubated at room temperature for one hour with primary antibodies and one hour with secondary antibodies and 0.1 μg ml^{-1} DAPI (Sigma), with 3×10 min washes after each incubation. Vector TrueVIEW (Vector Labs) autofluorescence quenching reagent was used before mounting the slides. Anti dsRed (1:100, Clonetech #632496) was used for the detection of mCherry, and anti-GFP 4B10 (1:100, CST #632496) was used for the 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 a ×63 objective (Leica).

For the histology analyses of metastatic outgrowth, formalin-fixed, paraffin-embedded slides were sectioned at 5 μm, stained with hematoxylin and eosin and scanned with an Aperio ScanScope XT Slide Scanner (Leica). The

images were segmented into lung and metastatic regions. The segmentation and annotation of the numbers and areas of metastatic lesions were done using QuPath software (https://qupath.github.io/) in consultation with a pathologist.

Mouse xenograft studies. To detect fusion in vivo, parental GFP- and mCherry-expressing cells were collected and mixed at a 50/50 ratio in DMEME-F12 culture media mixed with 50% Matrigel (BD Biosciences). The mixtures were injected into fat pads of eight-week-old female NSG mice on both flanks, with 1×10^6 cells in a 100 µl volume per injection. Tumour growth was monitored weekly by palpation. When the tumour diameter reached 1 cm or the animals became moribund with symptoms of reduced mobility, hunching and laboured breathing, the mice were euthanized. The xenograft tumours were fixed in formaldehyde and embedded in paraffin.

For the lung metastasis assays, luciferase-labelled parental cell lines (MCF10DCIS and SUM159PT) or their hybrids at passage 8 after antibiotic selection were collected and resuspended in DMEM/F12 medium with 10% FBS. Approximately 2×10^5 cells were injected per animal through the tail vein into eight-week-old NSG mice. Tumour growth was monitored by bioluminescence capture with the IVIS-200 imaging system (PerkinElmer) under isoflurane anaesthesia. The imaging was performed immediately after cell injection and then weekly. One month after injection, the mice were euthanized, and their lungs were extracted. One lobe was fixed in 10% formalin for subsequent paraffin embedding and histology analysis, while the remaining tissue was cut into small fragments and incubated in collagenase/hyaluronidase/BSA solution (5 mg ml^{-1} each) at 37 °C under constant mixing, until the disappearance of visible chunks of tissue. Following two washes with ice-cold PBS, the cell suspension was filtered using 500 µm cell strainers (Puriselect), pelleted and plated into 10 cm dishes for a brief (three days) ex vivo culture before the FACS analyses. All animal experiments were performed in accordance with the approved procedures of IACUC protocol #IS00005557 of the H. Lee Moffitt Cancer Center.

Growth and invasion/migration assays. To determine growth rates, 5×10^4 cells were seeded in triplicate into 6 cm culture dishes. On reaching ~90% confluency, the cells were collected by trypsinization, counted and reseeded at the same starting density. The growth rates were calculated as average ln(cell number fold change)/(number of days in culture) over three passages. The invasion/migration assays were 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. The lower wells contained medium with 10% FBS. Approximately 10^4 cells were plated and cultured for 72 hours. Matrigel was removed from the transwell insert, and the 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. The membranes were cut off and mounted on glass slides with xylene-based Cytoseal mounting media (ThermoFisher). The slides were scanned using an Evos Autoimaging system (ThermoFisher), and the images were analysed with ImageJ software (https://fiji.sc/). Sixteen tiles per sample were analysed.

Colony formation assays. For the in vitro clonogenic assays, 50/50 mixes of GFP- and mCherry-expressing parental cell lines were grown separately (controls) or in co-cultures for three days. For the ex vivo clonogenic assays, tumours initiated from the implantation of single-labelled cells or 50/50 mixes were collected four weeks after implantation and digested with a mixture of 2 mg ml^{-1} collagenase I (Worthington Biochem) and 2 mg ml^{-1} hyaluronidase (Sigma H3506) at 37 °C for three hours. Approximately 1×10^6 cells were seeded in 10 cm culture dishes in corresponding growth media with 10 µg ml^{-1} blasticidin and 2.5 µg ml^{-1} puromycin; for the negative controls, 5×10^5 of each of the parental cells were seeded. For the clonogenicity rate controls, parental cells were seeded at 100 cells per 6 cm dish in corresponding antibiotic. After two weeks of incubation, with the selection media replaced twice per week and the acquisition of fluorescent images of representative colonies, the medium was removed, the plates were washed twice with PBS and the 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 four hours. Colonies with approximately 50 or more cells were manually counted. Clonogenic frequency was calculated using numbers of colonies as percentages of seeded cells. To account for the reduced clonogenic survival of cells freshly isolated from xenograft tumours, the clonogenic data for the ex vivo samples were normalized to the clonogenicity of ex vivo cells isolated from tumours seeded with parental cells and plated in the absence of antibiotic selection.

SNP and copy number analyses. DNA from ~3×10^6 cells was extracted with a DNeasy Blood&Tissue kit (Qiagen). For the SUM159PT/MDA-MB-231 hybrids and MCFDCIS/SUM159PT from lung metastases, a CytoSNP-12 v.2.1 BeadChip array from Illumina was used, and the data were analysed with GenomeStudio v.2.0 software (Illumina). For the MCF10DCIS/SUM159PT and HS578T/MDA-MB-231 hybrids, a CytoScan array from Affymetrix was used. The SNP data were analysed and the LogR Ratio plots were generated with ChAS 4.2 software (Affymetrix). The data were analysed and visualized using R (v.3.3.2). First, the genotypes of parental cell lines were compared, and homozygous SNPs that were distinct between the two parental cell lines were selected. Then, for each of the differential SNPs, an identity score was assigned for a hybrid sample: 0 if only SNPs from parent 1 were detected, 1 if only SNPs from parent 2 were detected and 0.5 if both were detected. The data were plotted as a heat map, where the rows represent hybrid samples and the columns are aligned by the position number of each SNP, with the colours conveying the identity scores. For the copy number and genotype analyses and the visualization of the data from the CytoScan and CytoSNP-12 SNP arrays, we used Chromosome Analysis Suite v.4.1 (ChAS 4.1). We used Genome Studio to analyse the copy numbers and genotypes of the samples. The segmented copy numbers of the samples were visualized in the Integrative Genome Browser (IGV v.2.6.2). The Morpheus online tool (https://software.broadinstitute.org/morpheus) was used to visualize the differential allelic inheritance data.

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 cDNA preparation of parental (SUM159PT and MDA-MB-231) and hybrid cells. The targeted cell population sizes were 2,500 cells for each of the parental or hybrid samples. The libraries were constructed according to the 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 the GFP and mCherry proteins that the parental lines were engineered to express.

The raw gene expression matrices generated per sample using CellRanger (v.3.0.1) were combined in R (v.3.5.2) and converted to a Seurat object using the Seurat R package (https://satijalab.org/seurat/)[57]. All cells that had over two standard deviations of the mean unique molecular identifier (UMIs) count derived from the mitochondrial genome (between 5% and 12% depending on sample) were then removed, and we kept only cells with gene expression within two standard deviations of the mean gene expression[58]. From the remaining 10,059 cells, the gene expression matrices were log-normalized and scaled to remove variation due to total cellular read counts. To reduce the dimensionality of this dataset, the first 200 principal components were calculated on the basis of the top 2,000 variable genes. A Jack Straw analysis confirmed that the majority of the dataset variation was 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 on the basis of 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[59]. The data were then visualized by running the UMAP algorithm[28], colouring individual cells on the basis of either their cluster identity or their cell type identity. Gini coefficients were calculated with the package DescTools in R (ref. [60]). Zero values were excluded, and reads were used from 10x Genomics outputs.

Statistical analyses. The statistical analyses of the in vitro and in vivo experimental data were performed using GraphPad Prism 8.1.2 and MATLAB R2017a software, using the statistical tests indicated in the figure legends.

Computational methods. Muller's plots were obtained using the RTool Evofreq[61]. The agent-based model was implemented in the JAVA framework HAL[62]. To visualize cells carrying unique genotypes (the results are presented in Fig. 5g and Supplementary Videos 6–9), the scalar value of the genotype binary vector value in base 2 was mapped to a rainbow colour scale. A detailed description of the mathematical modelling is provided in the Supplementary Mathematical Methods.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All data supporting the findings of this study are available in the article and its Supplementary Information files. All SNP microarray data files along with their associated metadata have been deposited in the GEO database under the accession codes GSE159681 and GSE157845.

Code availability

The code for running the non-spatial and spatial simulations as well as for quantifying single-cell diversity is available online at https://github.com/ebaratch/FusionProject.

Received: 24 March 2020; Accepted: 26 November 2020;
Published online: 18 January 2021

References

1. Greaves, M. & Maley, C. C. Clonal evolution in cancer. *Nature* **481**, 306–313 (2012).
2. Scott, J. & Marusyk, A. Somatic clonal evolution: a selection-centric perspective. *Biochim. Biophys. Acta, Rev. Cancer* **1867**, 139–150 (2017).

3. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674 (2011).
4. Becks, L. & Agrawal, A. F. The evolution of sex is favoured during adaptation to new environments. *PLoS Biol.* **10**, e1001317 (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 (2016).
6. Sinai, S., Olejarz, J., Neagu, I. A. & Nowak, M. A. Primordial sex facilitates the emergence of evolution. *J. R. Soc. Interface* **15**, 20180003 (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 Res.* **69**, 8536–8539 (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**, 1118 (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 (2015).
11. Fujiwara, T. et al. Cytokinesis failure generating tetraploids promotes tumorigenesis in p53-null cells. *Nature* **437**, 1043–1047 (2005).
12. Su, Y. et al. Somatic cell fusions reveal extensive heterogeneity in basal-like breast cancer. *Cell Rep.* **11**, 1549–1563 (2015).
13. Zhou, X. et al. Cell fusion connects oncogenesis with tumor evolution. *Am. J. Pathol.* **185**, 2049–2060 (2015).
14. Erenpreisa, J. & Cragg, M. S. MOS, aneuploidy and the ploidy cycle of cancer cells. *Oncogene* **29**, 5447–5451 (2010).
15. Bennett, R. J. The parasexual lifestyle of *Candida albicans*. *Curr. Opin. Microbiol.* **28**, 10–17 (2015).
16. 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).
17. Fais, S. & Overholtzer, M. Cell-in-cell phenomena in cancer. *Nat. Rev. Cancer* **18**, 758–766 (2018).
18. 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 (2012).
19. 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 (2018).
20. 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. USA* **106**, 9385–9390 (2009).
21. Duelli, D. M. et al. A virus causes cancer by inducing massive chromosomal instability through cell fusion. *Curr. Biol.* **17**, 431–437 (2007).
22. Storchova, Z. & Pellman, D. From polyploidy to aneuploidy, genome instability and cancer. *Nat. Rev. Mol. Cell Biol.* **5**, 45–54 (2004).
23. Forche, A. et al. The parasexual cycle in *Candida albicans* provides an alternative pathway to meiosis for the formation of recombinant strains. *PLoS Biol.* **6**, e110 (2008).
24. Duncan, A. W. et al. The ploidy conveyor of mature hepatocytes as a source of genetic variation. *Nature* **467**, 707–710 (2010).
25. 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 (2012).
26. Marusyk, A., Janiszewska, M. & Polyak, K. Intratumor heterogeneity: the Rosetta Stone of therapy resistance. *Cancer Cell* **37**, 471–484 (2020).
27. Koulakov, A. A. & Lazebnik, Y. The problem of colliding networks and its relation to cell fusion and cancer. *Biophys. J.* **103**, 2011–2020 (2012).
28. Becht, E. et al. Dimensionality reduction for visualizing single-cell data using UMAP. *Nat. Biotechnol.* **37**, 38–44 (2018).
29. 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 (2019).
30. Hill, M. O. Diversity and evenness: a unifying notation and its consequences. *Ecology* **54**, 427–432 (1973).
31. Rosenzweig, M. L. *Species Diversity in Space and Time* (Cambridge Univ. Press, 1995).
32. Hinohara, K. et al. KDM5 histone demethylase activity links cellular transcriptomic heterogeneity to therapeutic resistance. *Cancer Cell* **35**, 330–332 (2019).
33. Loeb, L. A. Human cancers express a mutator phenotype: hypothesis, origin, and consequences. *Cancer Res.* **76**, 2057–2059 (2016).
34. Waclaw, B. et al. A spatial model predicts that dispersal and cell turnover limit intratumour heterogeneity. *Nature* **525**, 261–264 (2015).
35. 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 (2019).
36. 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 Res.* **78**, 2127–2139 (2018).
37. Noble, R., Burri, D., Kather, J. N. & Beerenwinkel, N. Spatial structure governs the mode of tumour evolution. Preprint at *bioRxiv* https://doi.org/10.1101/586735 (2019).
38. Jacobsen, B. M. et al. Spontaneous fusion with, and transformation of mouse stroma by, malignant human breast cancer epithelium. *Cancer Res.* **66**, 8274–8279 (2006).
39. 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 (2004).
40. Melzer, C., von der Ohe, J. & Hass, R. In vivo cell fusion between mesenchymal stroma/stem-like cells and breast cancer cells. *Cancers* https://doi.org/10.3390/cancers11020185 (2019).
41. Searles, S. C., Santosa, E. K. & Bui, J. D. Cell–cell fusion as a mechanism of DNA exchange in cancer. *Oncotarget* **9**, 6156–6173 (2018).
42. Zack, T. I. et al. Pan-cancer patterns of somatic copy number alteration. *Nat. Genet.* **45**, 1134–1140 (2013).
43. Dewhurst, S. M. et al. Tolerance of whole-genome doubling propagates chromosomal instability and accelerates cancer genome evolution. *Cancer Discov.* **4**, 175–185 (2014).
44. Stephens, P. J. et al. Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell* **144**, 27–40 (2011).
45. Birdsell, J. A. & Wills, C. in *Evolutionary Biology* (eds Macintyre, R. J. & Clegg, M. T.) 27–138 (Springer US, 2003).
46. Nieuwenhuis, B. P. & James, T. Y. The frequency of sex in fungi. *Phil. Trans. R. Soc. B* https://doi.org/10.1098/rstb.2015.0540 (2016).
47. Powell, A. E. et al. Fusion between intestinal epithelial cells and macrophages in a cancer context results in nuclear reprogramming. *Cancer Res.* **71**, 1497–1505 (2011).
48. Lazebnik, Y. The shock of being united and symphiliosis: another lesson from plants? *Cell Cycle* **13**, 2323–2329 (2014).
49. Mani, S. A. et al. The epithelial–mesenchymal transition generates cells with properties of stem cells. *Cell* **133**, 704–715 (2008).
50. 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 (2008).
51. Vitale, I. et al. Multipolar mitosis of tetraploid cells: inhibition by p53 and dependency on Mos. *EMBO J.* **29**, 1272–1284 (2010).
52. Amend, S. R. et al. Polyploid giant cancer cells: unrecognized actuators of tumorigenesis, metastasis, and resistance. *Prostate* **79**, 1489–1497 (2019).
53. 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 (2018).
54. 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 (2013).
55. 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).
56. Marusyk, A. et al. Spatial proximity to fibroblasts impacts molecular features and therapeutic sensitivity of breast cancer cells influencing clinical outcomes. *Cancer Res.* **76**, 6495–6506 (2016).
57. 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).
58. 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).
59. Newman, M. E. Modularity and community structure in networks. *Proc. Natl Acad. Sci. USA* **103**, 8577–8582 (2006).
60. Signorell, A. et al. DescTools. R package v. 0.99.38 (2020).
61. 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 Bioinform.* **20**, 710 (2019).
62. Bravo, R. R. et al. Hybrid Automata Library: a flexible platform for hybrid modeling with real-time visualization. *PLoS Comput. Biol.* **16**, e1007635 (2020).

Acknowledgements

We thank A. Goldman for sharing his unpublished observations of the frequent occurrence of cell fusions in his experimental studies, which confirmed our observations and helped motivate this work. We thank Y. Lazebnik, A. Goldman and D. Tabassum for providing thoughtful feedback. We thank R. 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 and Research Institute, an NCI designated Comprehensive Cancer Center (P30-CA076292). We thank the Moffitt SPARK programme for supporting an internship for M.A.L. This work was supported by the Susan G. Komen Breast Cancer Foundation grant no. CCR17481976 (A.M.), a Moffitt Cancer Biology and Evolution programme pilot award (A.M.), an Integrative Mathematical Oncology Workshop award (A.M. and D.B.), an American Cancer Society ACS-IRG award (P.M.A.), A PSON U01 CA244101 (D.B.), an MSM U01 CA202958 (D.B. and E.B.) and the Bankhead-Coley Cancer Research Program, award no. 20B06 (P.M.A., D.B. and A.M.).

Author contributions

A.M. and D.M. designed the studies. D.M. and M.A.L. performed the experimental studies and analysed the data. E.B., M.C.F.-F., P.M.A. and D.B. developed the mathematical models. M.C.F.-F., R.V.V. and A.C.T. performed the bioinformatical analyses. M.M.B. provided the pathology analysis. A.M., E.B., P.M.A., D.M. and D.B. wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41559-020-01367-y.

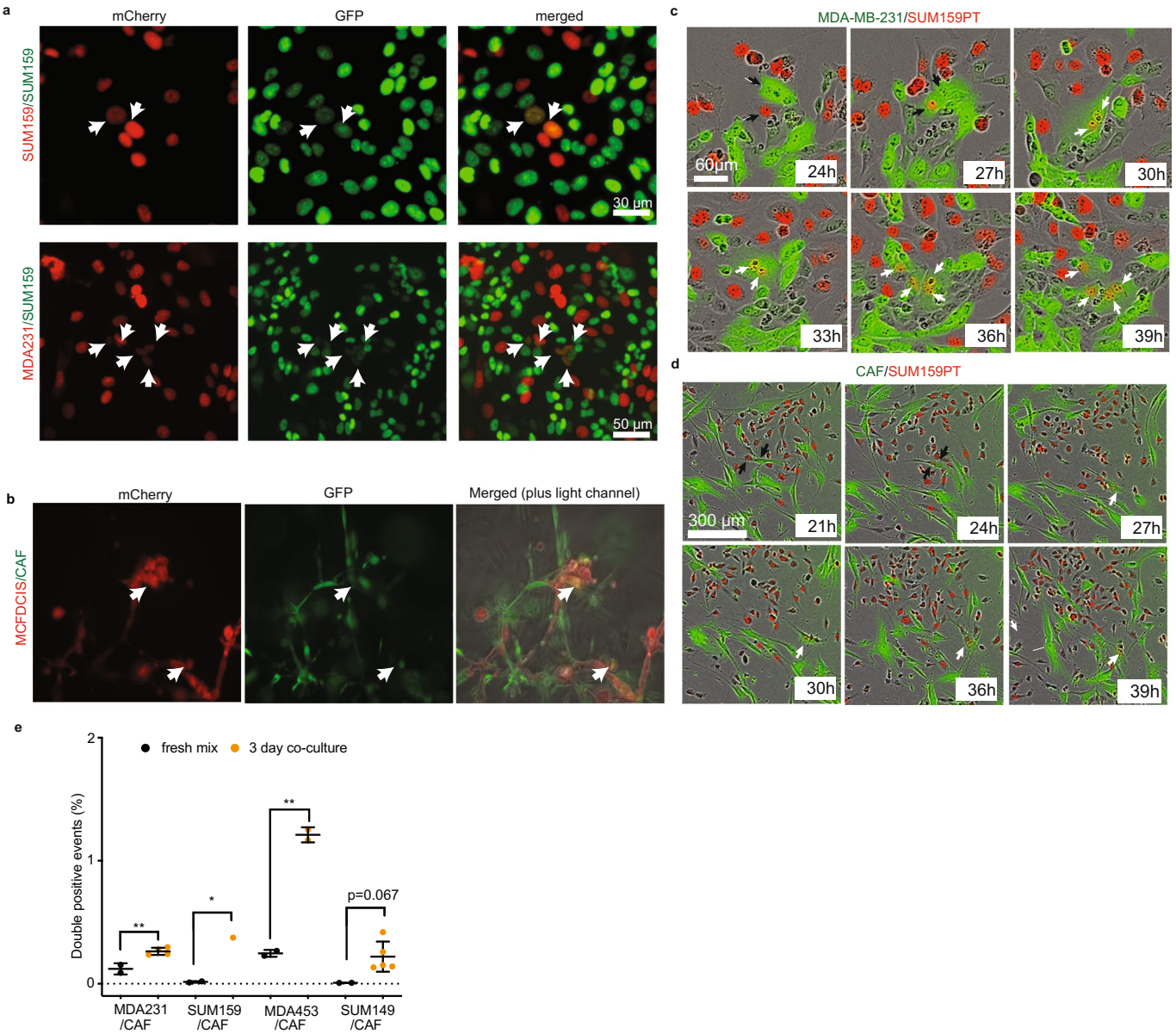
Supplementary information is available for this paper at https://doi.org/10.1038/s41559-020-01367-y.

Correspondence and requests for materials should be addressed to A.M.

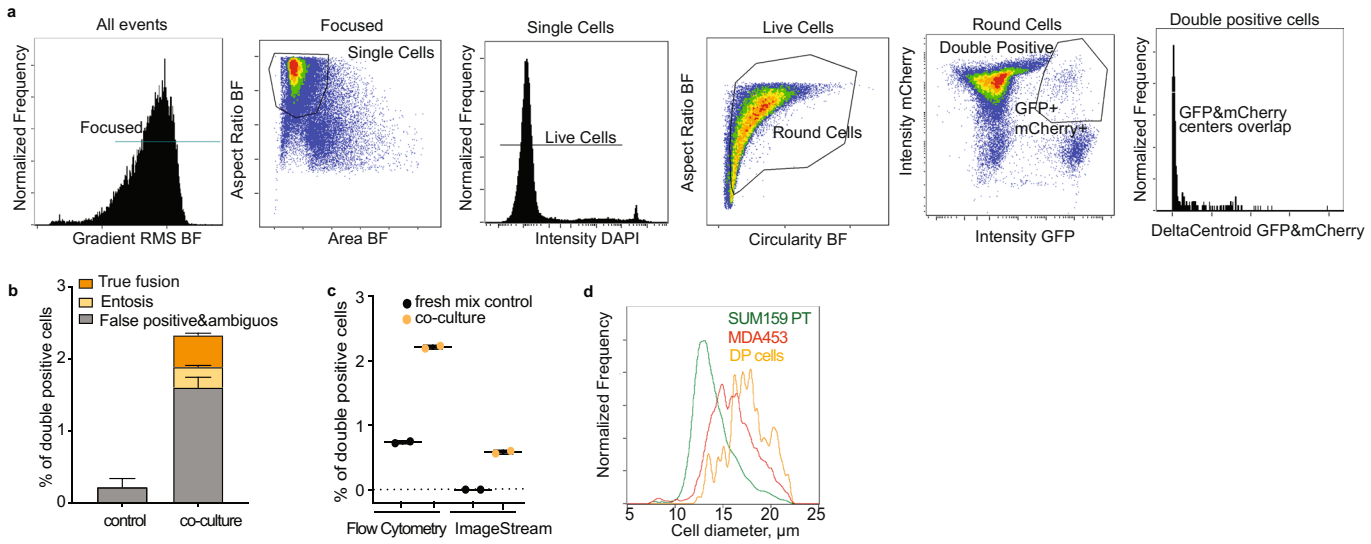
Peer review information *Nature Ecology and Evolution* thanks James DeGregori and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

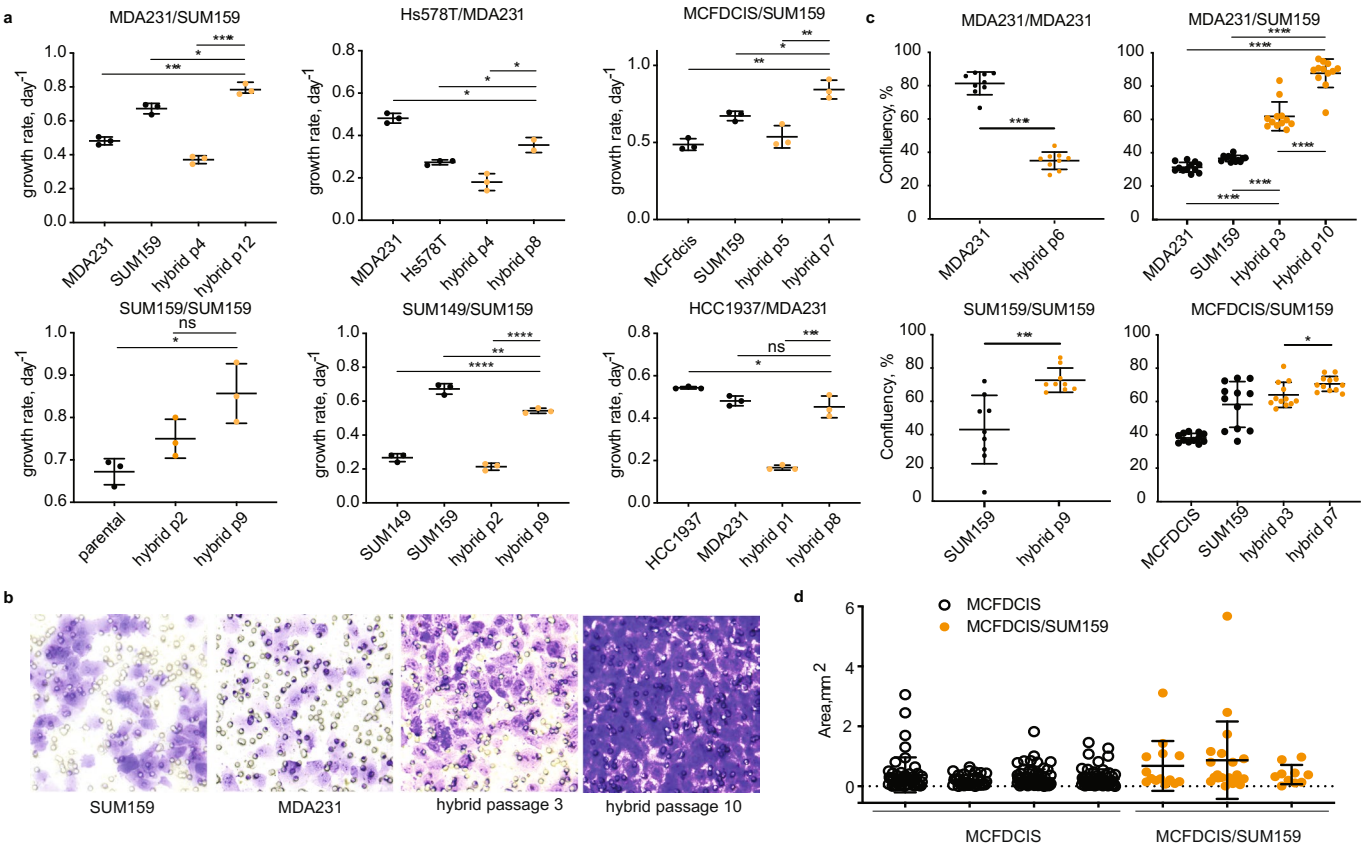
Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



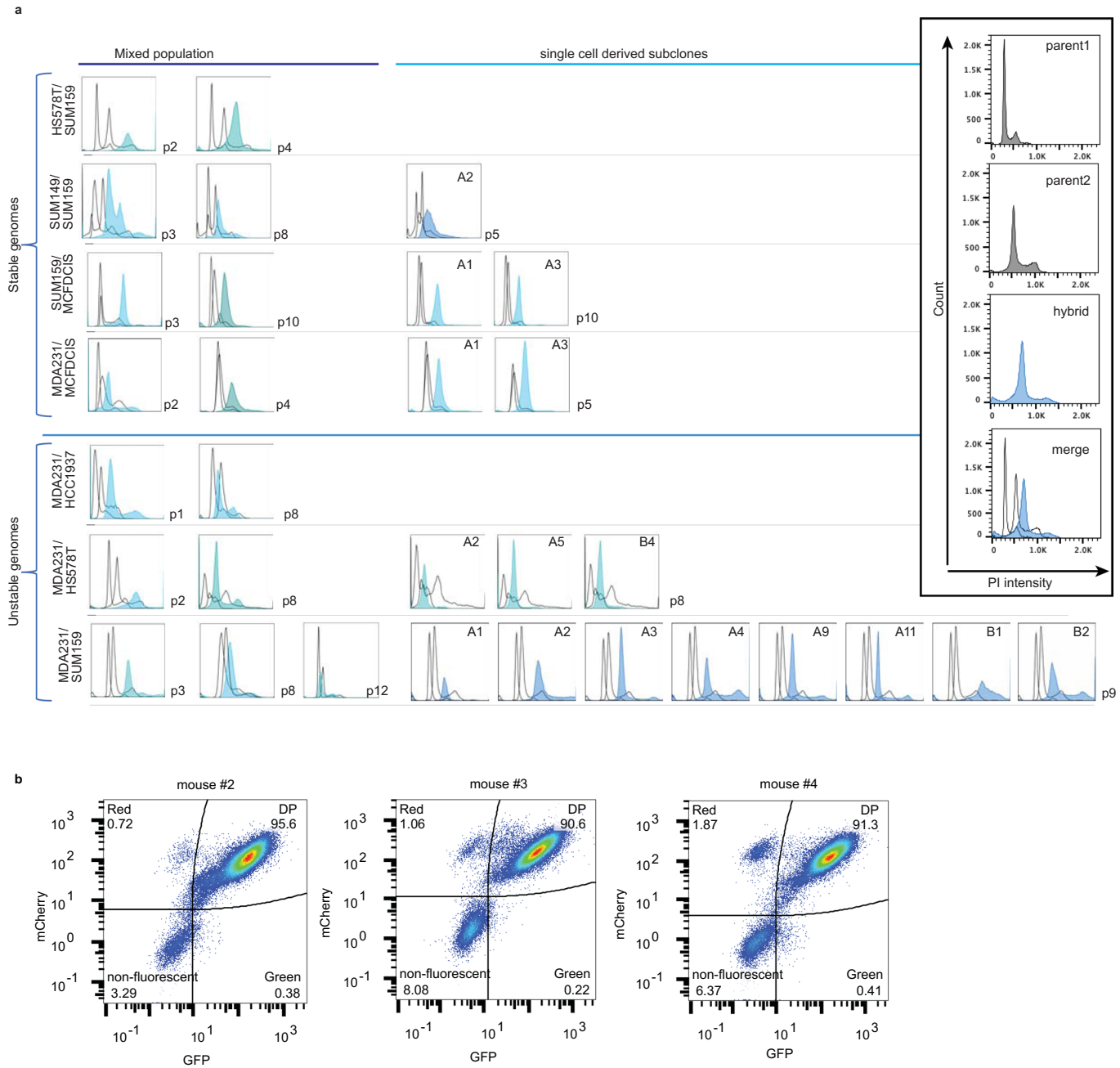
Extended Data Fig. 1 | Detection of putative spontaneous cell fusions. a. Live-cell fluorescence microscopy images of 2D co-cultures between differentially labelled (nuclear GFP and mCherry) cell lines. Arrowheads indicate cells that express both labels. **b**. Live-cell fluorescence microscopy images of 3D Matrigel co-cultures between MCF10DCIS breast carcinoma cells and a primary breast CAF isolate labelled with cytoplasmic GFP and dsRED, respectively. Arrowheads indicate cells that express both labels. **c,d** Time-lapse live fluorescence microscopy images of mCherry+ SUM159PT cells co-cultured with GFP+ MDA-MB-231 cells (**c**) and CAFs (**d**). The labels indicate time after plating. Black arrowheads show fusion parents, and white arrowheads show double-positive hybrid cells and their progeny. **e**. 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. Error bars represent SD, each dot represents an independent biological replicate. *and ** denote p values below 0.05 and 0.01 for two-tailed unpaired t-test, respectively.



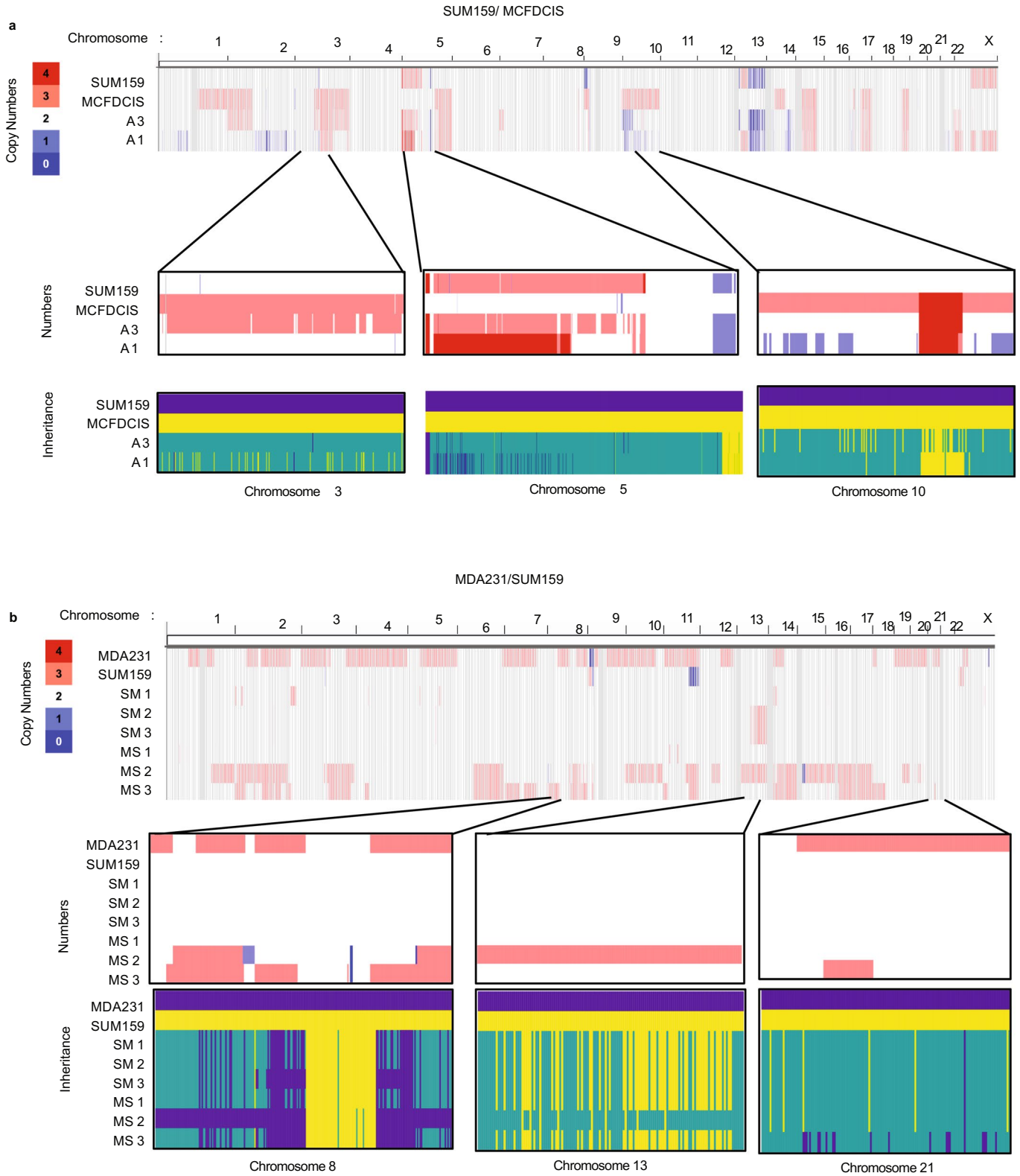
Extended Data Fig. 2 | ImageStream detection of spontaneous cell fusions. a. Gating strategy for the detection of double-positive cells with ImageStream imaging-based flow cytometry platform. **b**. Quantitation of different classes of double-positive events in 3-day co-cultures between 50/50 mixes of GFP/mCherry labelled MCF7 cells, with examples of diffedrent 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 ImageStream platforms (validated true positives percentages are plotted for ImageStream analyses). Error bars represent SD, dots represent biological replicates. **d**. Distribution of cell diameters of the parental and double-positive cells from ImageStream data shown in (**c**) measured in bright field and plotted using IDEAS software (ImageStream).



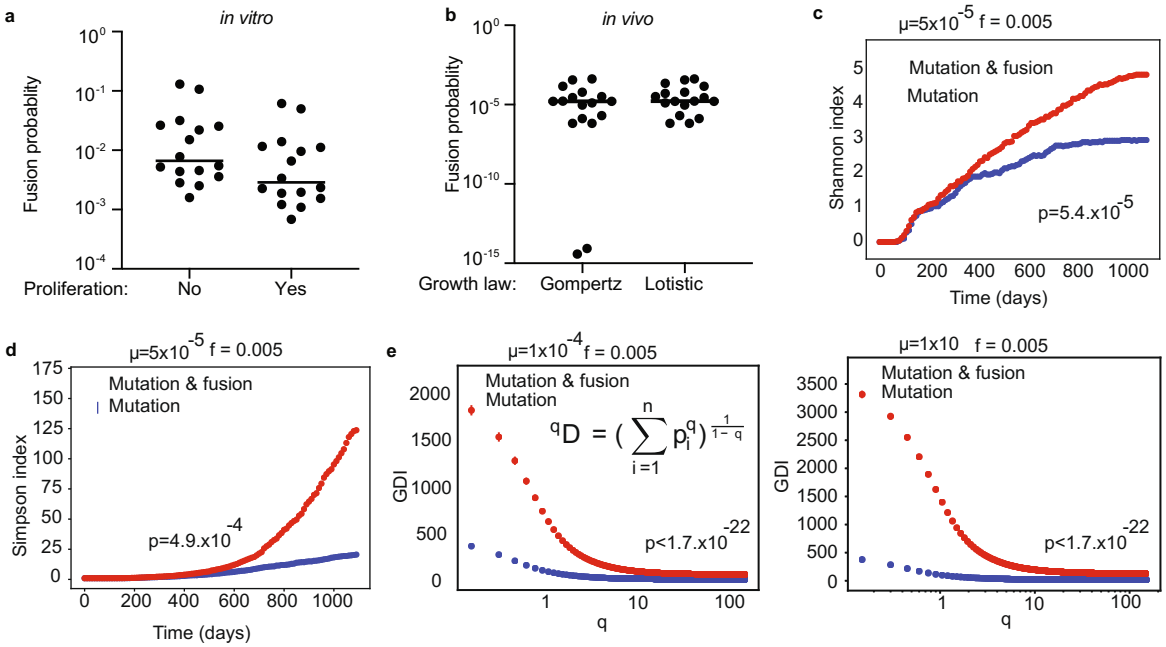
Extended Data Fig. 3 | Phenotypic characterization of hybrid cells. a. Growth rates of the indicated cell lines and their hybrids, at the indicated passages post-antibiotic-selection. **b**. Representative images of stained membranes from Boyden chamber cell invasion/migration assay. **c**. Quantitation of Boyden chamber cell invasion/migration assay data. **d**. Quantification of area of lung metastases formed after tail vein injection of MCFDCIS and MCFDCIS/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 two-tailed unpaired t-test. Error bars represent SD, dots represent biological replicates (a,c,d).



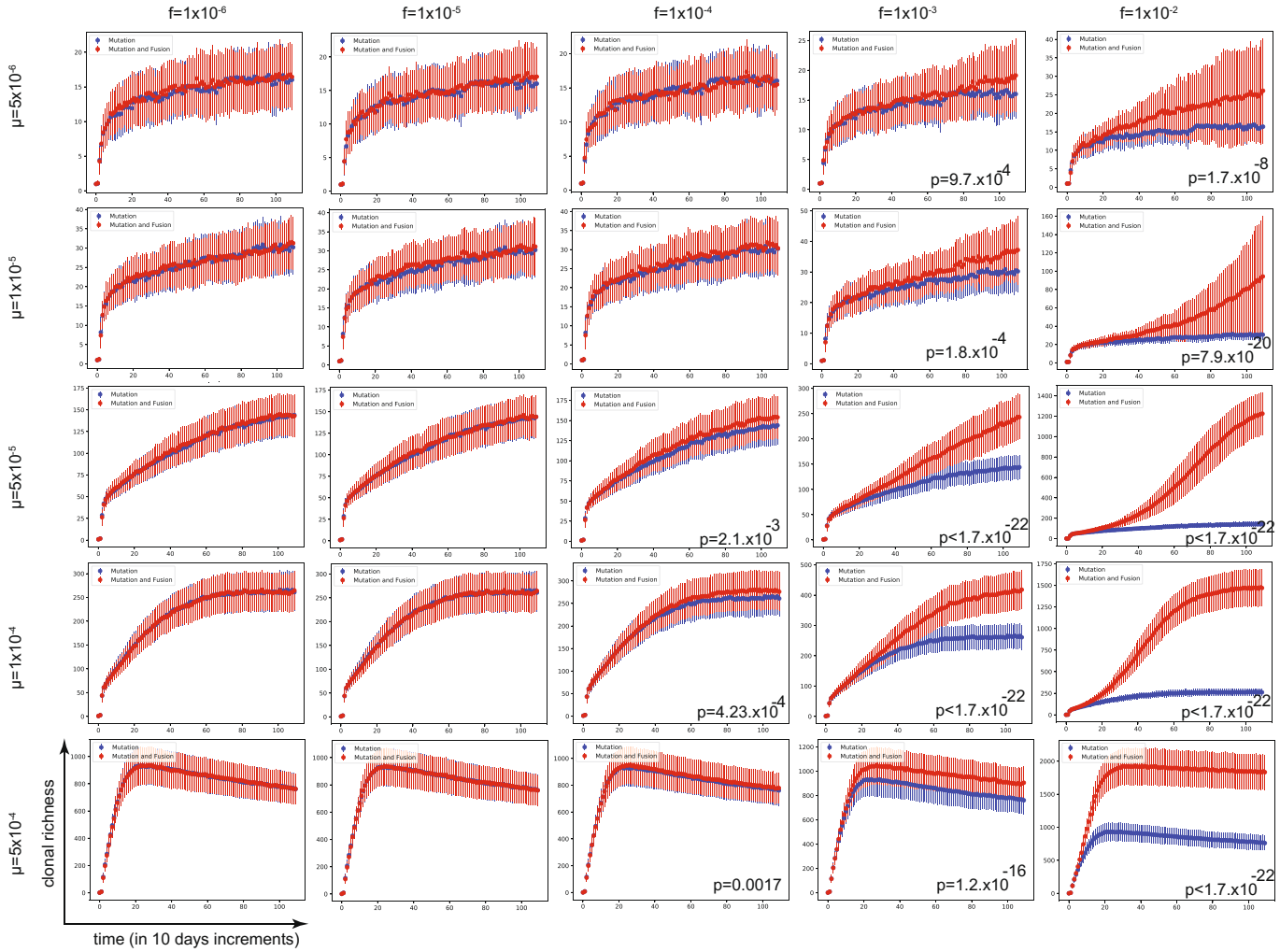
Extended Data Fig. 4 | DNA content analysis of hybrid cells. a. FACS analysis of DNA content of the indicated parental cell lines and their hybrids. P1-12 indicate passage number of the mixed hybrid populations; for hybrid subclones the number indicates passage of the mixed culture, used for isolation of single cell subclones. Black contours indicate DNA content profiles for parental cell lines and their G1 and G2 peaks are used as reference points, filled histograms indicate DNA content profiles of the hybrids. "Unstable genomes" and "stable genomes" refer to hybrids that, respectively, did and did not show evidence of reduced DNA content between early and extended passages. Hybrids with stable genomes demonstrate G1 peak closer to G2 peak of parents and no observed difference in DNA content between passages and single cell clones. Inset indicates axes and shows components of the merged plots. **b**. FACS analysis of retention of GFP and mCherry fluorescence in hybrid cells, recovered from metastasis-bearing lungs of the indicated animals depicted in Fig. 2g.



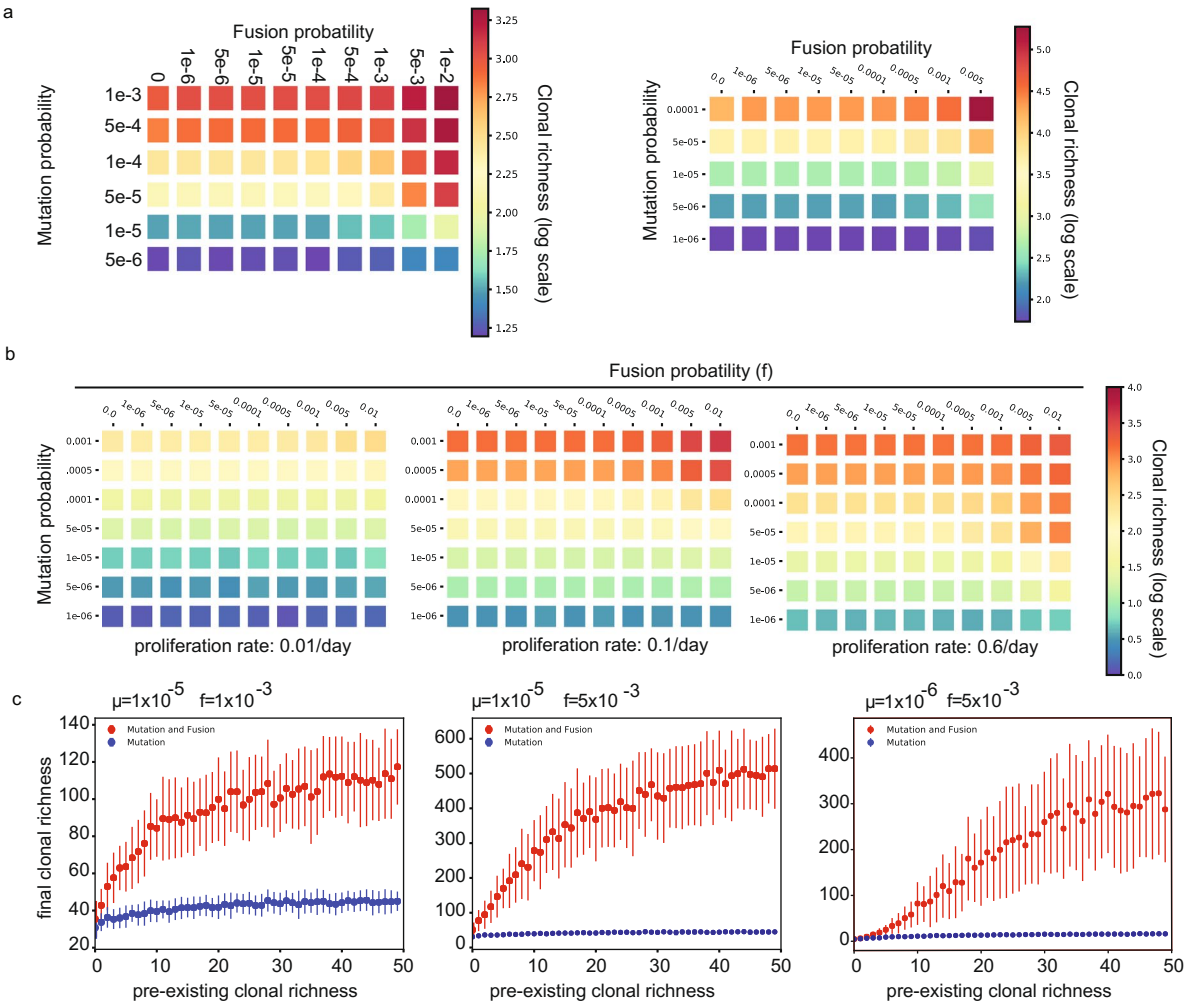
Extended Data Fig. 5 | Copy number analyses of cell line-specific SNP alleles. SNP copy number and inheritance data were obtained with Illumina CytoSPN-12 platform. Copy number data status is shown for all of the cell line-specific SNP alleles across the genome for the indicated sublones. For the indicated selected zoom-in insets, correspondence between SNP copy numbers and detected inheritance is plotted. Turquoise color indicates mixed inheritance. **a**. Analyses of the SUM159/MCFDCIS hybrids shown in Fig. 3g and Supplementary Fig. 4. **b**. Analyses of subclones from MDA-MB-231/SUM159 hybrids, seeded from 1-t passage. SM1-3 and MS1-3 represent subclones isolated from two distinct hybrid populations, using opposite fluorophore/antibiotic labels of the parental cells.



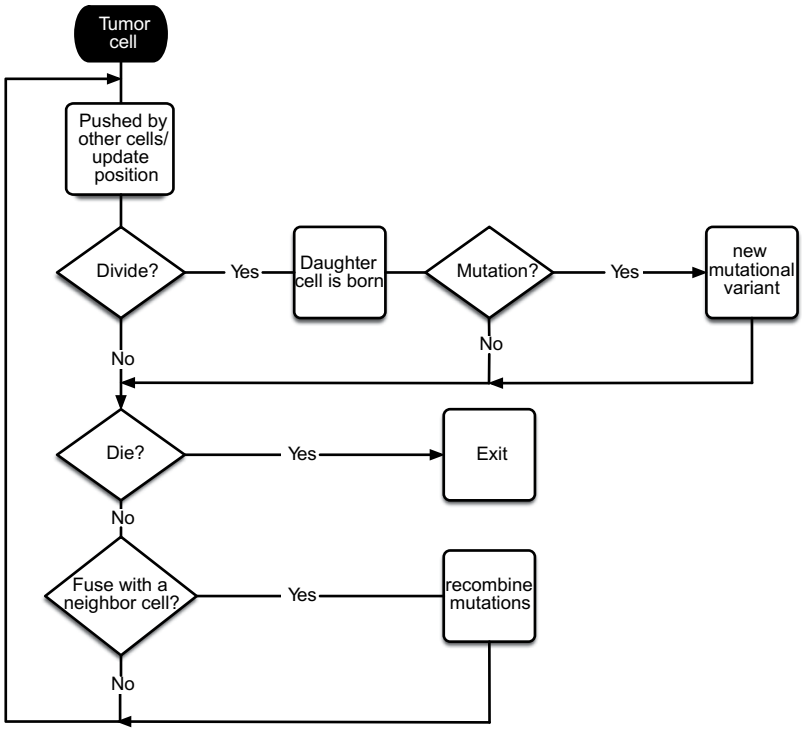
Extended Data Fig. 6 | Impact of fusions on diversification. a. Estimates of fusion probabilities from in vitro data, with assumptions of no proliferation of hybrids, or proliferation at the same rates as the parental cells. **b**. Estimates of fusion probabilities from in vivo data for Gompertz and Logistic growth. **c-e**. Comparisons are drawn between results of *in silico* simulations involving mutations only versus mutation and fusion. Clonal diversity is captured by Shannon (**c**), Simpson (**d**) and GDI (**c**) diversity indexes. Mutation and fusion rates are indicated in the figures. Indicated p values denote the results for Kolmogorov-Smirnov test for the final timepoint of simulations (**c**, **d**), or for the all of the <1 q values (**e**); growth rate a=0.67/day; carrying capacity K=10^4 cells; number of genes G=300.



Extended Data Fig. 7 | Parameter sweep analysis 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 (f) rates; p values indicate the results of Kolmogorov-Smirnov test comparing clonal richness at the end of the simulation (only shown for differences that has reached the 0.05 significance threshold); growth rate a=0.67/day; carrying capacity K=10^4 cells; number of genes G=300.



Extended Data Fig. 8 | Impact of population size, cell turnover and starting genetic heterogeneity on diversification by fusion-mediated recombination. **a**. Exploration of the impact of maximal tumor population size (carrying capacity) on diversification over indicated mutation and fusion rates. Left: carrying capacity K=10^4 cells; Right: carrying capacity K=10^6 cells; growth rate a=0.67/day; number of genes G=300 **b**. Exploration of the impact of tumor proliferation/turnover rates on diversification over indicated mutation and fusion rates; carrying capacity K=10^4 cells; number of genes G=50 **c**. Relationship between the initial clonal richness, and clonal richness at the indicated mutation and fusion rates. In all the panels, clonal richness at the end of the 1095 days of *in silico* growth is depicted; growth rate a=0.67/day; carrying capacity K=10^4 cells; number of genes G=300.



Extended Data Fig. 9 | Flowchart for agent based model for spatial simulations.

nature research

Corresponding author(s): Andriy Marusyk, PhD

Last updated by author(s): Nov 16, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	GenomeStudio 2.0 software (Illumina) and ChAS software (Affimetrix) were used for analysis of SNP data. ImageJ software (https://fiji.sc/) was used for the quantification of numbers and size of metastatic lesions.
Data analysis	Graph Pad Prism and R software were used for data analyses.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

SNP microarray data files along with their associated metadata have been deposited in the GEO database under the accession code GSE159681, GSE157845.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical methods were used to determine sample size. For fusion detection by FACS we collected ~80-100k events aiming to detect 0.5-1% of double positive events. Fusion rates were verified by using image-based cytometry where we collected ~10k events that was restricted by instrument running time. Our experiments show statistically significant difference between co-cultured cells and mix of cells that is proof of principal rather than statements on the exact fusion rates for cancer cell lines. Sample size for scRNA seq was defined by instrument maximum capacity and was 10000 cells per run that included 4 samples (~2500cells per sample)
Data exclusions	No data was excluded from the analysis.
Replication	All of the in vitro experiments were performed with 2-10 biological replicates, as indicated in the figures. The animal studies were performed using 4-5 mice per group, which is widely used in the field and defined by practical consideration and the reduction principle of animal studies.
Randomization	Animals were randomly assigned to experimental groups
Blinding	No blinding during data collection was performed; image analyses were performed using machine learning based segmentation.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	RRID:AB_10013483, RRID:AB_1196614, RRID:AB_2633282, RRID:AB_2633275A
Validation	Immunofluorescent staining was validated using positive and negative control xenograft tissue sections with known overexpression of protein markers

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	RRID:CVCL_0062, RRID:CVCL_0290, RRID:CVCL_0332, RRID:CVCL_0553, RRID:CVCL_0418, RRID:CVCL_5552, RRID:CVCL_5423, RRID:CVCL_3422
Authentication	Identity of cell lines were confirmed by STR analysis
Mycoplasma contamination	Cell lined were tested negative for mycoplasma contamination
Commonly misidentified lines (See ICLAC register)	*Name any commonly misidentified cell lines used in the study and provide a rationale for their use.*

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	NOD.Cg-Prkdcscid Il2rgtm1Wjl/SzJ JAX #005557
Wild animals	*Provide details on animals observed in or captured in the field; report species, sex and age where possible. Describe how animals were caught and transported and what happened to captive animals after the study (if killed, explain why and describe method; if released, say where and when) OR state that the study did not involve wild animals.*
Field-collected samples	*For laboratory work with field-collected samples, describe all relevant parameters such as housing, maintenance, temperature, photoperiod and end-of-experiment protocol OR state that the study did not involve samples collected from the field.*
Ethics oversight	All animal experiments were performed in accordance with the guidelines of the IACUC #IS00005557 of the H. Lee Moffitt Cancer Center

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	For fusion detection cells were harvested with 0.25% Trypsin and resuspended in PBS with 0.1 µg/mL DAPI. For DNA ploidy 10E6 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, 0.2 mg/ml RNAse and 20 µg/ml PI.
Instrument	MACSQuant VYB; Image Stream
Software	FlowJo; Image Stream software
Cell population abundance	For fusion detection average number of collected events was 80,000. Percentage of double positive events (fusion rate) was defined as % of fluorescent DAPI-negative events ((double positive/(DAPI-negative- mCherry/GFP -negative))*100. Depending on cell confluency of the samples, DAPI-negative group was 60-95% of total, mCherry/GFP-negative was 1-20%.
Gating strategy	All data were plotted using pseudocolor visualization. FSC/SSC gate included ~95% of events in the range of 10k-150k for SSC-A and 10k-160k for FSC-A. All events with intensity lower that 10k for SSC-A and FSC-A were excluded. Next, we gated our all DAPI-negative cells in SSC-A/DAPI plot selecting all events with lower than 1.1E3 DAPI intensity in the range 0-10E5 biexponential axis. Than in GFP/mCherry intensity plot we applied quadrant gating with the boundary at ~10E3 for both axis in range of 0-10E5, biexponential.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.