Stabilising selection causes grossly altered but stable karyotypes in metastatic colorectal cancer

William Cross*[1,2], Maximilian Mossner*[1], Salpie Nowinski*[1],George Cresswell*[3], Abhirup Banerjee*[1], Marc Williams*[1,4], Laura Gay*[1], Ann-Marie Baker*[1], Christopher Kimberley[1], Hayley Davis[11], Pierre Martinez[1,5], Maria Traki[1], Viola Walther[1], Kane Smith[1], Giulio Caravagna[3], Sasikumar Amarasingam[1,6], George Elia[1], Alison Berner[1,2], Ryan Changho Choi[1,6,7], Pradeep Ramagiri[8], Ritika Chauhan[8], Nik Matthews[8], Jamie Murphy[9], Anthony Antoniou[6], Susan Clark[6,9], Jo-Anne Chin Aleong[10], Enric Domingo[11], Inmaculada Spiteri[3], Stuart AC McDonald[1], Darryl Shibata[12], Miangela M Lacle[13], Lai Mun Wang[14,15], Morgan Moorghen[6], Ian PM Tomlinson[16], Marco Novelli[2], Marnix Jansen[2,17], Alan Watson[18], Nicholas A Wright[1], John Bridgewater[2,17], Manuel Rodriguez-Justo[2,17], Hemant Kocher[18,10], Simon J Leedham[14], Andrea Sottoriva^[3], Trevor A Graham^[1]

* These authors contributed equally to this manuscript.
^ For correspondence: andrea.sottoriva@icr.ac.uk and t.graham@qmul.ac.uk

1 Centre for Genomics and Computational Biology, Barts Cancer Institute, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, Charterhouse Sq, London, UK.
2 UCL Cancer Institute, UCL, 72 Huntley St, Bloomsbury, London, UK
3 Evolutionary Genomics and Modelling Lab, Centre for Evolution and Cancer, The Institute of Cancer Research, London, UK.
4 Computational Oncology, Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, 321 East 61st St., New York, NY 10065, USA
5 Univ Lyon, Université Claude Bernard Lyon 1, INSERM 1052, CNRS 5286, Centre Léon Bérard, Cancer Research Center of Lyon, Lyon, 69008, France
6 St Mark's Hospital, Harrow, London, UK
7 St. George Hospital, Sydney, Australia
8 Tumour Profiling Unit (TPU), Institute of Cancer Research, London, UK
9 Department of Surgery and Cancer, Imperial College London, London, UK
10 Barts Health NHS Trust, The Royal London Hospital, Whitechapel, London, UK.
11 Department of Oncology, University of Oxford, Oxford
12 Department of Pathology, University of Southern California Keck School of Medicine, Los Angeles, CA 90033, USA
13 Department of Pathology, University Medical Center Utrecht, Utrecht, Netherlands
14 Wellcome Trust Centre Human Genetics, University of Oxford, Roosevelt Drive, Oxford, UK
15 Ludwig Institute, University of Oxford, Oxford, UK
16 Cancer Research UK Edinburgh Centre, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Edinburgh, UK
17 UCL Hospitals NHS Trust, London, UK
18 Department of Gastroenterology, Barts Health, Whipps Cross Hospital, Leytonstone, London, UK.
19 Centre for Tumour Biology, Barts Cancer Institute, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, Charterhouse Sq, London, UK.

Contributions:
Study conception and design: AS & TG
Funding and supervision: SL, HK, NW, AS, TG
Manuscript writing: WC, MM, SN, GC, MW, AS, TG
Bioinformatics: WC, SN, GC, MW, PM, GC, TG
Mathematical modelling: MW, TG
Experimental work: MM, LG, AB, CK, MT, VW, KS, SA, RC, HD, IS, NM
Provision of patient materials: SA, AB, JM, AA, SC, JC, ED, LM, IT, MM, ML, AW, MN, MRJ, JB, HK, SL
Histopathology: MRJ, ML, MJ, NW, MN, LM, MM

Abstract
Aneuploidy, defined as the loss and gain of whole and part chromosomes, is a near-ubiquitous feature of cancer genomes, is prognostic, and likely an important determinant of cancer cell biology. In colorectal cancer (CRC), aneuploidy is found in virtually all tumours, including precursor adenomas. However, the temporal evolutionary dynamics that select for aneuploidy remain broadly uncharacterised. Here we perform genomic analysis of 755 samples from a total of 167 patients with colorectal-derived neoplastic lesions that cross-sectionally represent the distinct stages of tumour evolution, and longitudinally track individual tumours through metastasis and treatment. Precancer lesions (adenomas) exhibited low levels of aneuploidy but high intra-tumour heterogeneity, whereas cancers had high aneuploidy but low heterogeneity, indicating that progression is through a genetic bottleneck that suppresses diversity. Individual CRC glands from the same tumour have similar karyotypes, despite prior evidence of ongoing instability at the cell level. Pseudo-stable aneuploid genomes were observed in metastatic lesions sampled from liver and other organs, after chemo- or targeted therapies, and late recurrences detected many years after the diagnosis of a primary tumour. Modelling indicates that these data are consistent with the action of stabilising selection that 'traps' cancer cell genomes on a fitness peak defined by the specific pattern of aneuploidy. These data show that the initial progression of CRC requires the traversal of a rugged fitness landscape and subsequent genomic evolution, including metastatic dissemination and therapeutic resistance, is constrained by stabilising selection.

Introduction

Most cancer genomes are aneuploid[1]. This is the result of chromosome or part-chromosome gains and losses which are attained through a variety of mechanisms that include chromosome mis-segregation, aberrant double strand break repair, and genome doubling[2]. In colorectal cancer (CRC), approximately 85% of cancers are classified as chromosomally unstable (CIN) by virtue of their having an abnormal karyotype and aneuploidy[3]. The remaining ~15% of cancers exhibit genetic instability at the sequence level[4] and are classified as hypermutant, but nevertheless often also exhibit some chromosome aberrations[3].

Aneuploidy likely has a causal role in cancer evolution. First, the observed ubiquity of aneuploidy in cancers and the prevalence of recurrent chromosome copy number alterations (CNAs) is suggestive that some CNAs are positively selected[5]. Second, aneuploidy is predictive of progression to cancer in premalignant diseases such as Barrett oesophagus[6] and ulcerative colitis[7].Third, in established cancers, individual CNAs have prognostic value over-and-above single nucleotide alterations in key cancer driver genes[8]. Fourth, ploidy has been shown to have prognostic value in many types of established cancer[9], and notably within-tumour heterogeneity of CNAs, as principally measured by the number of clones with distinct karyotypes, is prognostic pan-cancer[10]. Notably, in lung cancer, the diversity of CNAs but not single nucleotide alterations (SNAs) has prognostic value[11].

Aneuploidy should also be subject to negative selection because the loss or gain of large genomic regions may affect the expression of thousands of genes and regulatory elements, some of which are likely to adversely affect tumour cell viability[12,13]. Chromosomal-scale alterations are therefore expected to result in trade-offs between the dosages of positively- and negatively-selected elements[14], potentially leading to stabilising selection on the aneuploid genome. Chromosomal changes may also have a buffering effect against deleterious point mutations[15]. Indeed, clinical studies show that the burden of CNAs is non-monotonically related with prognosis across cancer types[10] (first observed in breast cancer[16]). As such, cancers carrying an intermediate level of alterations ('just right' aneuploidy) are associated with worse prognosis than cancers with lower or higher levels of aneuploidy. Observing purifying selection may be complicated by the fact that less fit variants are effectively removed from the population and do not come to clinical prominence.

The spatial-temporal evolutionary dynamics that produce the aneuploidy observed in cancer genomes remain poorly understood[17]. In part this is because longitudinal measurement is often not clinically feasible, and is compounded by the lack of effective cancer-specific evolutionary tools to reconstruct ancestral history of chromosome gains and losses[18]. CRC presents a unique opportunity to track clonal evolution over space and time: pre-cancerous lesions (colorectal adenomas, CRAs) are occasionally 'caught in the act' of transformation and are found to contain a small focus of cancer (cancer-in-adenoma, ca-in-ad lesions). CRC frequently metastases to the liver, and repeated hepatic metastasectomy can provide a source of longitudinally sampled tumour material amenable for molecular analysis.

Results

Here we investigated the evolution of aneuploidy in a large cross-sectional and longitudinally followed cohort of CRAs and CRCs. Previously we observed that the degree of aneuploidy was much higher in a small cohort of (malignant) colorectal cancers than (benign) adenomas[19]. To validate this observation, we performed multi-region spatial sampling and genome-wide copy number alteration analysis in n=139 archival colorectal lesions (19 adenomas CRAs, 81 cancers CRCs, and 39 cancer-in-adenoma [ca-in-ad] samples; minimum two regions analysed per neoplasm; Fig 1A). The percentage of genome altered (PGA) was lowest in CRAs, highest in CRCs (8.5% and 27.2% respectively, $p=8.4x10^{-14}$) with ca-in-ad samples comparable to CRCs (25.3%; Fig 1B). Gains were more prevalent

than losses in the CRAs (p=0.0052; Fig S1A) and were seen in similar proportions of gains/losses in the ca-in-ads and CRCs (p=0.75, p=0.32, respectively; Fig S1A).

Intra-tumour heterogeneity of CNAs (hereafter 'divergence'), measured as the fraction of CNAs that were subclonal (Methods), was greatest in CRAs and significantly less in CRCs (p=0.0017; Fig 1C). The heterogeneity of losses and gains changed through progression (Fig S1B). Therefore, progression of colorectal tumours was through an aneuploidy-divergence 'phase space' from low aneuploidy/high divergence to high aneuploidy/low divergence (Fig 1D), suggesting the presence of a selective genetic bottleneck acting on CNAs that tumours pass through during progression from benign to malignant disease. Supporting the hypothesis of a CNA bottleneck at malignant transformation, we observed that within the ca-in-ads, divergence was higher between the adenoma components than between cancer components, and divergence in cancer components of ca-in-ads was significantly lower than in more advanced CRCs (p=0.0077; Fig S1C).

We then investigated the relationship between CNA divergence and clinical outcome. Evolutionary theory predicts that the diversity of pre-existing adaptive phenotypes in a population should be proportional to population evolvability; more diverse populations are more likely to harbour an individual pre-adapted to a new selective pressure. Accordingly, intra-tumour clonal diversity has been observed as a pan-cancer prognostic biomarker[10]. However, CNA divergence was not associated with overall survival in our cohort (comparison of cases above and below median divergence; p=0.98; Fig 1E). Consequently, these data suggested that intra-tumour CNA heterogeneity did not strongly reflect the heterogeneity of adaptive phenotypes in CRC, and was consistent with prior observations of a lack of positive selection experienced by CRC subclones[19–21].

Previous reports using fluorescence *in situ* hybridisation (FISH) to measure CNAs at single cell resolution report widespread chromosomal instability (highly variable copy number) in colorectal cancer cell lines[22] and primary tumours[23]. Given these prior observations, the relatively low levels of CNA diversity that we observed in CRC compared to CRAs was surprising. We hypothesised that large 'bulk' samples could mask heterogeneity between smaller clonal populations. Well-to-moderately differentiated CRCs are made up of a collection of glands, each composed of a few thousand cells recently derived from a common ancestor[23,24]. Glands can reasonably be considered the units of selection in CRC, as gland fission likely expands the cancer cell population[23]. We extracted a total of 307 individual colorectal cancer glands from the notional 'left' and 'right' opposite sides of 6 CRCs (Fig 2A) and performed high-depth exome sequencing (59 glands) or shallow whole genome sequencing (sWGS; 248 glands). Exome sequencing data was used to phase CNAs detected by sWGS (Methods). In addition, we also analysed at least two bulk samples composed of numerous glands from each tumour.

At gland level (individual clone) resolution the pattern of CNAs were broadly homogeneous within cancers (Fig 2B). In all cases, a 'core karyotype' was evident across the glands from that case, with most CNAs being highly frequent across the tumour gland population, and only small numbers of glands showing additional subclonal CNAs (mean CNAs per gland = 7, mean clonal CNAs = 6.5). We note that intra-tumour CNA heterogeneity was detected, but we emphasise the clonality of most CNAs. FISH in two additional cases confirmed ongoing instability at the level of individual cancer cells within glands (Fig S2). But at gland resolution, cancers appeared grossly aneuploid but not genetically unstable.

Single nucleotide alterations (SNAs) were called from individual gland exome sequencing data and were used to reconstruct phylogenetic trees for each case (Methods; Fig 2C). Trees were balanced (ancestral nodes have equal numbers of offspring; Yule Model; p>0.05 for all cases) suggestive of a lack of stringent subclonal positive selection. Individual glands from MSS tumour contained on

average 156 SNAs (range: 68-232) which equated to 40 more SNAs than the most recent common ancestor (MRCA) cell of the cancer (371 SNVs for MSI cases). Assuming SNAs accrued at a constant rate (clock-like) throughout life[25], we calculated the SNV accumulation rate per year as the mean number of SNVs per branch divided by the patient age. We then estimated that for the MSS cases the most recent common ancestor (MRCA) of the tumour was on average 77% (range: 69-88%) of the total age of observed cancer, which equated to cancers being formed an average of 16 years prior to resection (range: 7-22 years). Despite this elapsed time, we observed little subclonal accrual of CNAs within the cancers. CNAs were more homogenous than expected under a null model of random distribution of CNAs (Fig 2D). We hypothesised therefore that the distinct pattern of CNAs observed in each CRC represented a (local) optimum in the fitness landscape, with stabilising selection for the specific pattern of CNAs suppressing diversity at gland and bulk levels.

We then investigated how aneuploidy evolved during metastatic spread and treatment. Colonisation of a new tissue and therapy represent new selective pressures that could change the core karyotype observed in primary CRCs. Even in the absence of positive selection, bottlenecks during metastatic colonisation combined with ongoing aneuploidy at the microscopic level, are likely to lead to gross differences between primary tumour and metastatic deposits. CRC frequently metastasises to the liver and less frequently to other organs, including the lung. We collected tissue from 23 primary sites and 68 metastatic lesions from 22 patients, with the material collected longitudinally from a mode of 2 time-points (range 1-5 time points). Five patients remained completely untreated throughout their studied time course (an additional three patients received no treatment between sampled timepoints), while other patients had experienced a diverse range of treatment regimes. We measured CNAs in each tumour deposit using multi-region sWGS (168 distinct regions analysed).

We frequently observed the maintenance of a core abnormal karyotype in the primary tumour that remained largely unchanged in the metastatic lesions through space, time, metastasis to other tissues and through chemotherapy and/or targeted treatment (Fig 3A-D, Fig S3-S5). For example, LM0027 displayed little divergence in karyotype between the colon primary tumour and a metastasis sample taken a year later in the liver, during this period the patient received no treatment (Fig 3A; mean divergence of 0.26 between timepoints, measured as the fraction of bins altered between samples divided by the total number of aberrant in at least one sample). Maintenance of a core karyotype was also displayed in LM0071 despite a longer period of time elapsing between the primary colon and metastatic liver samples (23 months) and the patient receiving 3-months of oxaliplatin and capecitabine chemotherapy (Fig 3A; mean divergence 0.17). A third patient, LM0029, was followed for 28 months from primary diagnostic biopsy, during which time they had hepatic metastatectomies, one pulmonary metastatectomy, and three challenges of chemotherapy with the anti-epidermal growth factor monoclonal antibody cetuximab. Strikingly, the core karyotype remained relatively stable throughout (Fig 3A; max. mean inter-timepoint divergence of 0.48). Similarly, in patient LM0139, resection of the primary CRC at baseline was followed by by hepatic metastatectomy at (23 months) and pulmonary metastatectomy (at 72 months). Despite 6 years of elapsed evolution from the final clone sampled in the lung had a very similar pattern of CNAs to the initial clone observed in the primary CRC, demonstrating marked stability across time (max. mean inter-timepoint divergence of 0.28; Fig S6A).

Alterations in the proportion of genome altered (PGA) were low in sequential timepoints indicating stable aneuploidy (mean PGA increased by 6.5 percentage points on average for consecutive timepoints in the cohort; Fig 3B). Genetic divergence between metastatic regions was not significantly different from divergence between regions of primary regions, within the same time point (Fig 3C; p=0.28, linear mixed model, Fig S6B) and divergence within regions of primary colorectal samples was consistent with previous single timepoint cohorts (mean divergence 0.27). However, genetic divergence between liver and colorectal samples from the same patient were

significantly more divergent than within primary lesions (Fig S6C; p-adjusted= 0.032, Tukey HSD test). This was further supported by CNA divergence changes over time being greater than changes over space (p=4.07e-10, 0.08 lower divergence within timepoints, linear mixed model; Fig S6D). The change in aneuploidy across patients was generally small including in on-treatment intervals (a median mean absolute difference of 11.3 PGA across all timepoints, 16.2 in on-treatment intervals, Fig S6E). Indeed, CNA divergence was not significantly different in on-treatment intervals versus off-treatment intervals (p=0.771, linear mixed model, Fig 3D).

Taken together, these data suggested a model whereby strong positive selection for an abnormal karyotype occurs during transformation from benign to malignant disease, and that thereafter the karyotype is effectively 'fixed' by stabilizing selection acting at the level of individual cells within colorectal glands. The same stabilising selection force persists through metastasis and during treatment. This model is consistent with the 'Big Bang' dynamics that we have previously proposed: we expect positive selected events, both SNAs and CNAs, to be clonal in the tumour and part of the 'trunk' of the phylogenetic tree, and then observed within-tumour evolution is effectively-neutral. Negative selection (for CNAs) acts on individual cells and so is unobservable at gland and bulk levels, and so the persisting tumour cell population evolves effectively-neutrally.

To further investigate the evolutionary dynamics expected under stabilising selection, we constructed a mathematical model of karyotype evolution. Individuals in the model ('cells') had variable fitness (s) (that altered the division (b) and/or death rate (d) of the cell, $s = b - d$) defined by the cell's karyotype according to the relation:

$$s = \frac{s_{max}}{1 + \alpha |G_O - G_i|}$$

Where G_i was the karyotype of cell i, and G_O was a pre-specified 'optimal' karyotype that maximised fitness. A cell's karyotype was represented by a vector of length C, with each entry representing the integer copy number of each locus in the cell's (abstracted) genome. A locus mutated to a new copy number at rate μ. A cell's fitness was then inversely proportional to the absolute distance between a cell's genotype and the maximum fitness genotype (G_O); the parameter α scaled the strength of negative selection experienced by non-optimal genotypes, with $\alpha = 0$ defining a neutral "flat" landscape where all genotypes have the same fitness, $s = s_{max}$ (Fig 4A).

Stochastic simulations showed that ongoing mutation within a growing tumour in the absence of stabilising selection ($\alpha = 0$) generated substantial karyotypic heterogeneity (Fig 4B&C). A growing tumour that initially occupied the fitness maxima (genotype of the first cancer cell = G_O) experienced stabilising selection served to suppress genetic divergence, with stronger negative selection (greater α) more severely limiting divergence (Fig 4B&C), and higher mutation rates generating more diversity (Fig 4B). Thus, modelling suggested that stabilising selection for an optimal CNA genotype was a sufficient explanation of the stability of the aneuploid genome observed in CRCs over space and time.

A cancer initiated by a cell with a sub-optimal karyotype (Fig S7A) experienced directional selection towards the optimal karyotype (Fig S7B-D), and notably CNA diversity of the population increased transiently before the optimal genotype was found by the population. The transient increase in CNA diversity was explained by the coexistence of multiple 'independent' clones with distinct CNA genotypes that all had comparable fitness (equidistant from the fitness peak on the symmetrical fitness landscape; Fig S7C-E). Later in the evolutionary process, these clones are replaced by a single population (potentially composed of multiple convergently-evolved lineages) with the same optimal CNA genotype. This transient increase in diversity and accrual of CNAs driven by directional selection was consistent with our observations of CNA evolution during adenoma to cancer progression.

Discussion

Our study reveals the dynamics of CNA during the evolution of colorectal cancer. We observe large changes in CNAs during the progression from benign to malignant disease, but thereafter the cancer karyotype is effectively fixed, and remains unchanged over many years, through metastasis to distant organs, and despite treatment. These data are suggestive that CNAs define a genotype-phenotype combination that is under selection, with an altered core karyotype necessary for the initiation of malignancy, and sufficient for subsequent metastatic spread. These evolutionary dynamics are consistent with the observation of early metastasis in CRC[26].

We suggest that the fitness landscape traversed by CNAs is 'rugged', with local fitness peaks located at normal-diploid and at multiple aberrant-genome positions. Further, we suggest that moving between fitness peaks is broadly precluded by negative selection. Although there is ongoing chromosomal instability at fitness peak, we propose that the genetic change caused by a single CNA event is insufficiently large to move the cell to a different fitness peak, and instead leaves the cell in a fitness valley where it is removed by negative selection. Given this potentially defining role of CNAs in CRC evolution, we hypothesise that the fixed pattern of CNAs in any individual CRC likely defines the biology of that tumour, and consequently is a major determinant of patient prognosis and treatment response.

That the karyotype is broadly unchanged by treatment is surprising, as presumably the effectiveness of therapy is dependent on it acting as a novel and strong selective pressure. Naively, we expected this new selective pressure to change the fitness landscape and as a result cause wholesale change in the CNAs observed post-therapy. Instead, it may be that treatment alters non-CNA related aspects of the fitness landscape, or does not change the fitness landscape sufficiently to cause traversal to a new CNA-defined fitness peak. An alternative explanation is that chemotherapy response in CRC may be entirely plastic[27], and so does not involves the clonal selection of individual lineages – a phenomenon that may be common across cancer types[28]. In CRC, previous work is suggestive that plasticity is established very early in CRC[20] and is consistent with reports of early metastatic spread[26].

Our data indicate a defining role for copy number alterations in CRC evolution. Positive selection for CNAs in precancer, followed by lack of positive selection and instead stabilising selection in an established malignancy leads to a pseudo-stable karyotype. Leveraging knowledge of these dynamics may lead to prognostic and therapeutic benefit.

Methods

Full methods are provided as supplementary material.

Acknowledgments

MJ, AS and TG acknowledge funding from Cancer Research UK (A22745, A22909 and A19771 respectively). TG, AS and SL are funded by the Wellcome Trust (202778/Z/16/Z, 202778/B/16/Z and 206314/Z/17/Z respectively) and via a Wellcome Trust award to the Centre for Evolution and Cancer at the ICR (105104/Z/14/Z). SL and LMW acknowledge funding from the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre. DS, AS and TG received funding for this project from the NIH via the Cancer Systems Biology Consortium U54 scheme (CA217376). TG received pilot funding from the Barts Charity. AB was supported by Barts Health NHS Trust and Barts Charity grant. ED is supported by the S:CORT consortium which is funded by a grant from the Medical Research Council and Cancer Research UK. JB is in part funded by the UCLH/UCL Biomedical Research Centre. MJ, MN and MRJ receive funding support from Cancer Research UK UCL Experimental Cancer Medicine Centre and the Department of Health's National Institute for Health Research Biomedical Research Centres funding scheme (Reference: C12125/ A15576). The research leading to these results has received funding from the People Programme (Marie Curie Actions) of the European

Union's Seventh Framework Programme (FP7/2007-2013) under REA grant agreement n° 608765 (MM). This research was supported by the National Institute for Health Research (NIHR) Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust and King's College London who performed some of the genome sequencing, and we note that the views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. We are grateful for the support of the QMUL Genome Centre (Charles Mein) and Institute of Cancer Research Tumour Profiling Unit for support with genomic assays. A small project grant from the UK Bowel and Cancer Research (BACR) Charity to TG and SL pump primed this research.

References

1. Taylor, A. M. *et al.* Genomic and Functional Approaches to Understanding Cancer Aneuploidy. *Cancer Cell* **33**, 676-689.e3 (2018).
2. Sansregret, L., Vanhaesebroeck, B. & Swanton, C. Determinants and clinical implications of chromosomal instability in cancer. *Nature Reviews Clinical Oncology* vol. 15 139–150 (2018).
3. Network, T. C. G. A. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330–337 (2012).
4. Grasso, C. S. *et al.* Genetic mechanisms of immune evasion in colorectal cancer. *Cancer Discov.* **8**, 730–749 (2018).
5. Mermel, C. H. *et al.* GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.* **12**, R41 (2011).
6. Li, X. *et al.* Assessment of esophageal adenocarcinoma risk using somatic chromosome alterations in longitudinal samples in Barrett's esophagus. *Cancer Prev. Res.* **8**, 845–856 (2015).
7. Rubin, C. E. *et al.* DNA aneuploidy in colonic biopsies predicts future development of dysplasia in ulcerative colitis. *Gastroenterology* **103**, 1611–1620 (1992).
8. Smith, J. C. & Sheltzer, J. M. Systematic identification of mutations and copy number alterations associated with cancer patient prognosis. *Elife* **7**, (2018).
9. Hieronymus, H. *et al.* Tumor copy number alteration burden is a pan-cancer prognostic factor associated with recurrence and death. *Elife* **7**, (2018).
10. Andor, N. *et al.* Pan-cancer analysis of the extent and consequences of intratumor heterogeneity. *Nat. Med.* (2016) doi:10.1038/nm.3984.
11. Jamal-Hanjani, M. *et al.* Tracking the Evolution of Non–Small-Cell Lung Cancer. *N. Engl. J. Med.* **376**, (2017).
12. Sheltzer, J. M. & Amon, A. The aneuploidy paradox: Costs and benefits of an incorrect karyotype. *Trends in Genetics* vol. 27 446–453 (2011).
13. Sheltzer, J. M. *et al.* Single-chromosome Gains Commonly Function as Tumor Suppressors. *Cancer Cell* **31**, 240–255 (2017).
14. Li, L. *et al.* Cancer-causing karyotypes: chromosomal equilibria between destabilizing aneuploidy and stabilizing selection for oncogenic function. *Cancer Genet. Cytogenet.* **188**, 1–25 (2009).
15. López, S. *et al.* Interplay between whole-genome doubling and the accumulation of deleterious alterations in cancer evolution. *Nat. Genet.* **52**, 283–293 (2020).
16. Birkbak, N. J. *et al.* Paradoxical relationship between chromosomal instability and survival outcome in cancer. *Cancer Res.* **71**, 3447–3452 (2011).
17. Ben-David, U. & Amon, A. Context is everything: aneuploidy in cancer. *Nature Reviews Genetics* (2019) doi:10.1038/s41576-019-0171-x.
18. Schwartz, R. & Schäffer, A. A. The evolution of tumour phylogenetics: Principles and practice. *Nature Reviews Genetics* vol. 18 (2017).
19. Cross, W. *et al.* The evolutionary landscape of colorectal tumorigenesis. *Nat. Ecol. Evol.* **2**, 1661–1672 (2018).
20. Sottoriva, A. *et al.* A big bang model of human colorectal tumor growth. *Nat. Genet.* **47**, 209–

216 (2015).
21. Williams, M. J. *et al.* Quantification of subclonal selection in cancer from bulk sequencing data. *Nat. Genet.* (2018) doi:10.1038/s41588-018-0128-6.
22. Lengauer, C., Kinzler, K. W. & Vogelstein, B. Genetic instability in colorectal cancers. *Nature* **386**, 623–627 (1997).
23. Sottoriva, A. *et al.* A big bang model of human colorectal tumor growth. *Nat. Genet.* **47**, (2015).
24. Siegmund, K. D., Marjoram, P., Woo, Y. J., Tavaré, S. & Shibata, D. Inferring clonal expansion and cancer stem cell dynamics from DNA methylation patterns in colorectal cancers. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 4828–4833 (2009).
25. Alexandrov, L. B. *et al.* Clock-like mutational processes in human somatic cells. *Nat. Genet.* **47**, 1402–1407 (2015).
26. Hu, Z. *et al.* Quantitative evidence for early metastatic seeding in colorectal cancer. *Nat. Genet.* (2019) doi:10.1038/s41588-019-0423-x.
27. Kreso, A. *et al.* Variable clonal repopulation dynamics influence chemotherapy response in colorectal cancer. *Science (80-.).* **339**, 543–548 (2013).
28. Boumahdi, S. & de Sauvage, F. J. The great escape: tumour cell plasticity in resistance to targeted therapy. *Nature Reviews Drug Discovery* (2019) doi:10.1038/s41573-019-0044-1.

Figures

Figure 1: Aneuploidy increases through colorectal cancer progression, and clonal diversity decreases.

A: Heatmap showing copy number alterations (CNAs) in adenomas (orange; top), adenomas containing a focus of cancer (ca-in-ads; green; middle) and established cancers (blue; bottom). Thin black lines delineate cases, within black lines are multiple regions of each case. Red=gain, blue=loss.

B: The percentage genome altered (PGA) increases through progression from CRAs, to ca-in-ads and then CRCs.

C: Genetic divergence decreases through progression from CRAs to CRCs.

D: Phase space of PGA versus divergence. CRAs (orange) have low PGA but high divergence, whereas CRCs (blue) have high PGA but low divergence. Ca-in-ads (green) are more disparate in PGA and divergence.

E: Divergence is not significantly associated with overall survival in CRCs (Kaplan Meier analysis).

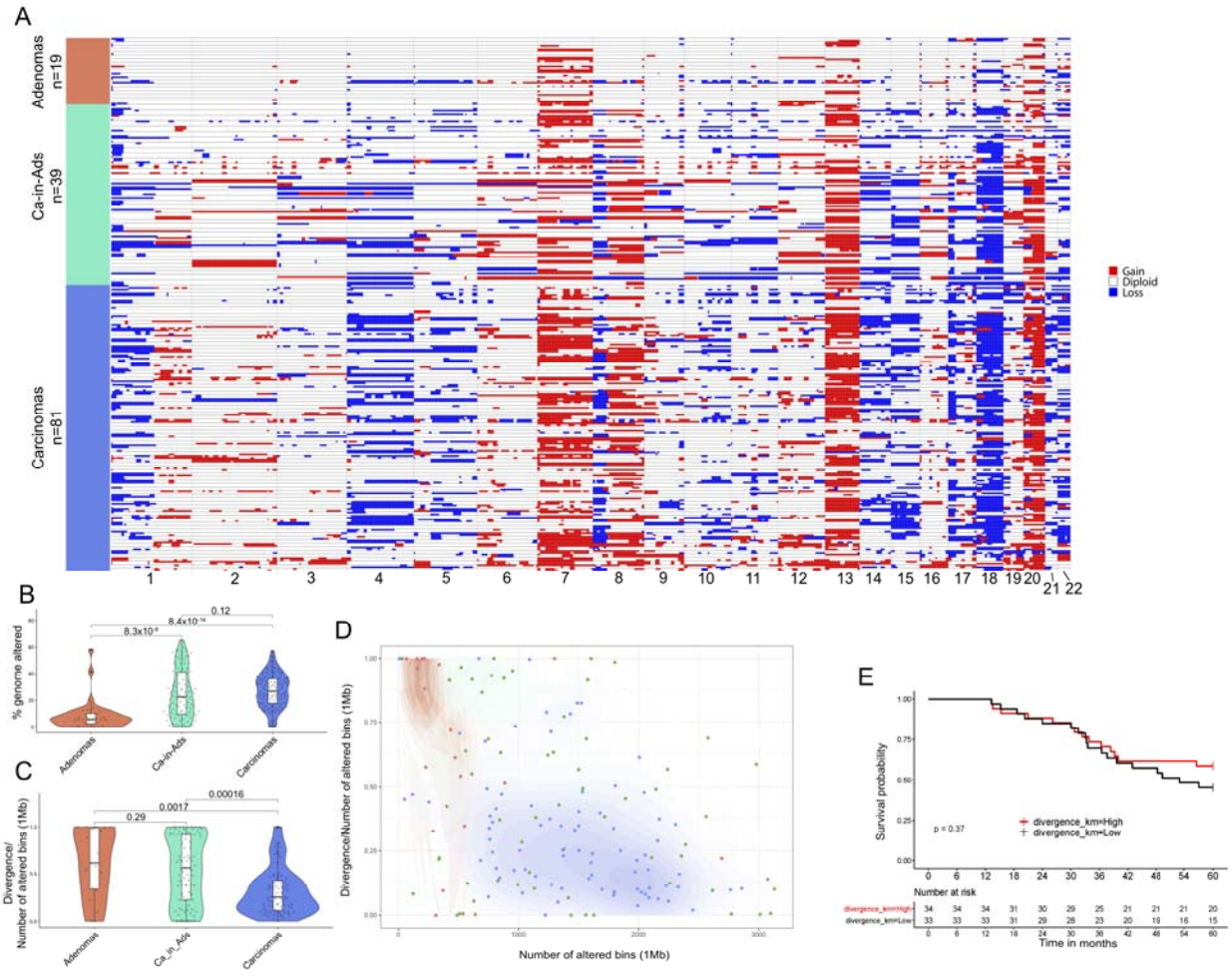


Figure 2: Lack of CNA heterogeneity at the level of individual colorectal cancer glands

A. Schematic showing strategy for extraction of individual cancer glands for genomic analysis.

B. (left column) Heatmaps showing individual CNAs in individual glands in each of six CRCs. Glands were sequenced from arbitrary opposite sides of each CRC (side 1, 2, etc). Red hue colours indicate gains, blue losses.

C: (centre column) Phylogenetic trees reconstructed from individual gland single nucleotide alteration calls from exome sequenced glands. Trees are balanced (comparison with Yule model). Note the trunk length is not shown to scale, number above trunk represent number of clonal SNAs. HI=homoplasy index.

D: (right column) Proportion of gland pairs (y-axis) by genetic divergence (red line) versus a null model where the same burden of CNAs observed in the case are randomly placed around the genome (green model). Clonality of the observed CNAs is significantly different to the null model.

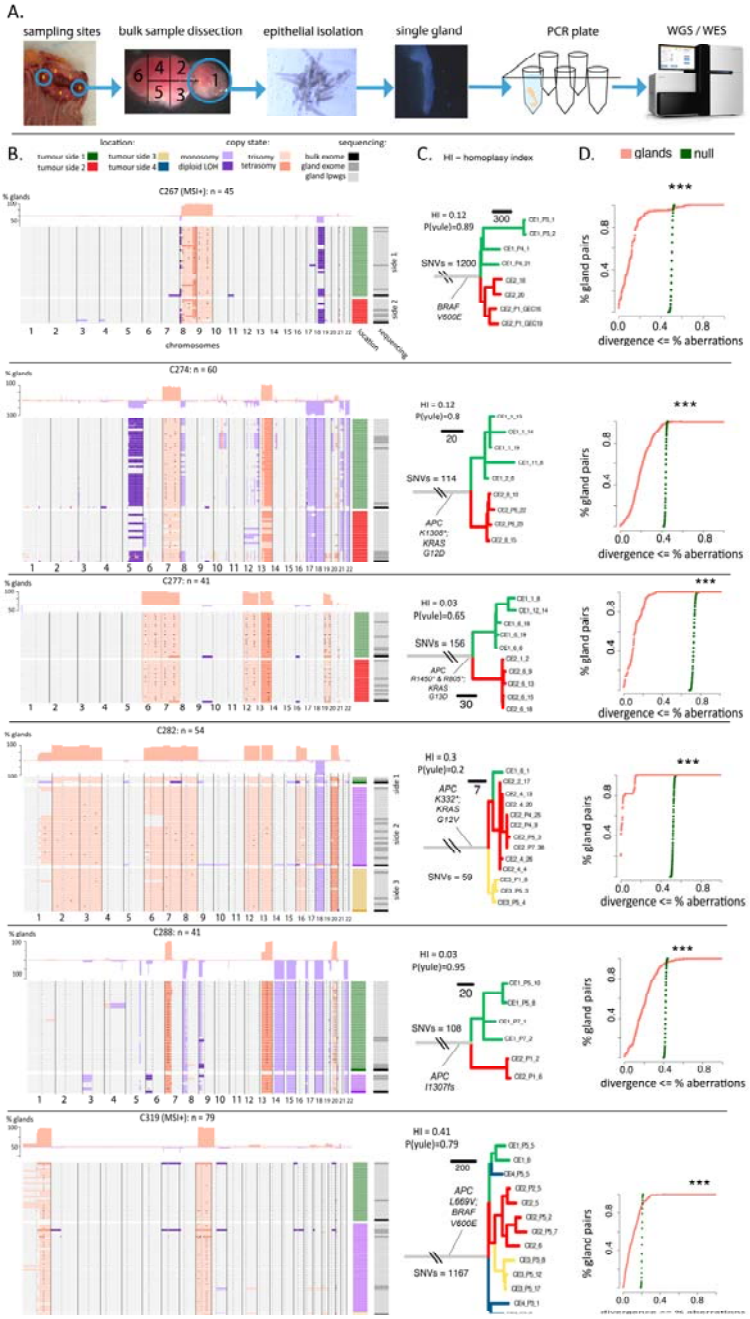


Figure 3: Stability of CNAs across space, time, tissue and treatment in colorectal cancer.
A: Strikingly similar CNA profiles were observed over space time, tissue and through various treatment courses. Timelines of four representative metastatic patients are shown.
B: The mean percent genome altered (PGA) remained broadly constant for over time in each patient (x-axis is on a log scale).
C: Genetic divergence within primary tumours was not significantly different to genetic divergence within matched the metastatic lesions.
D: Within-lesion genetic divergence was not significantly different after treatment with various chemotherapy and/or targeted agents.

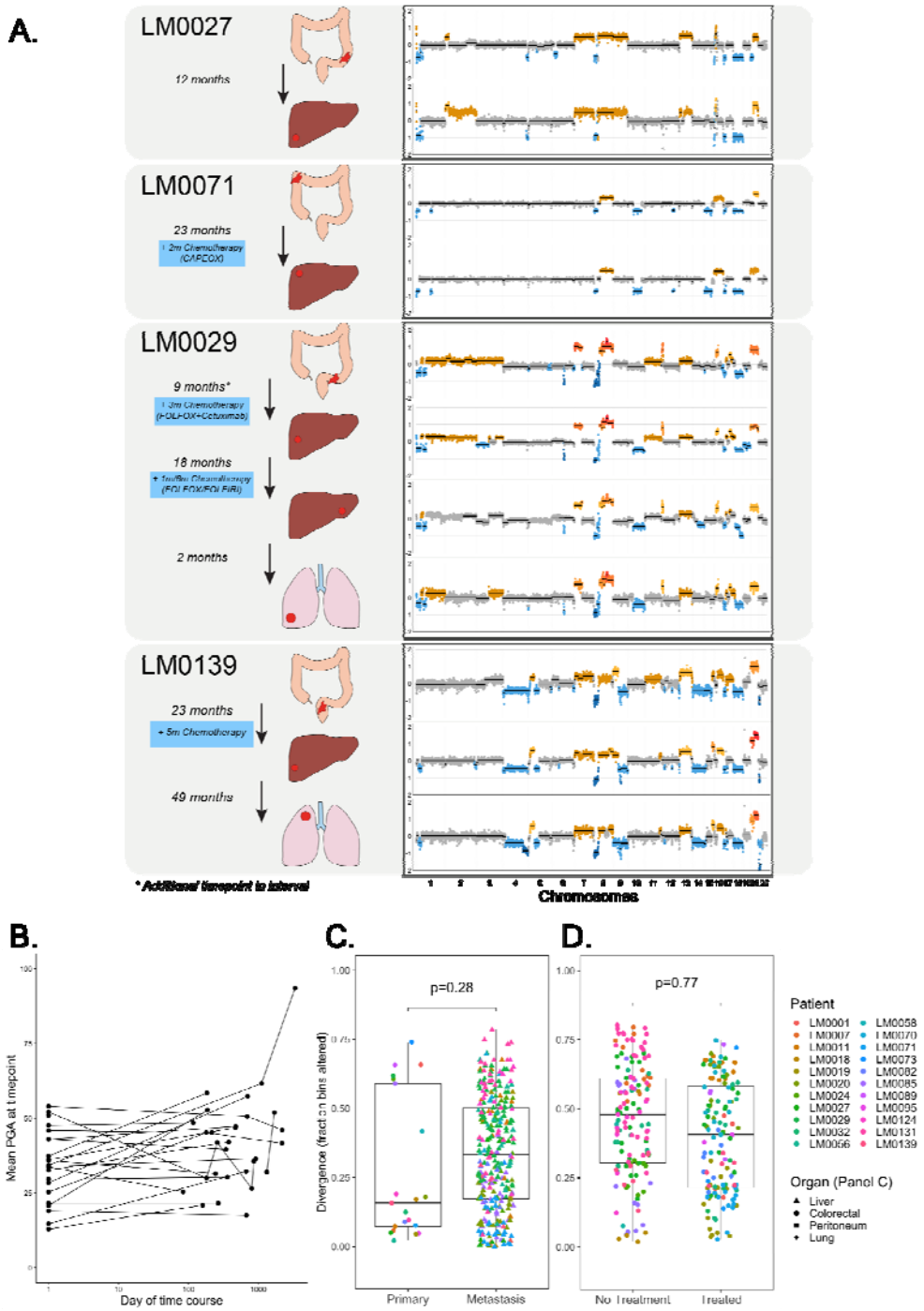
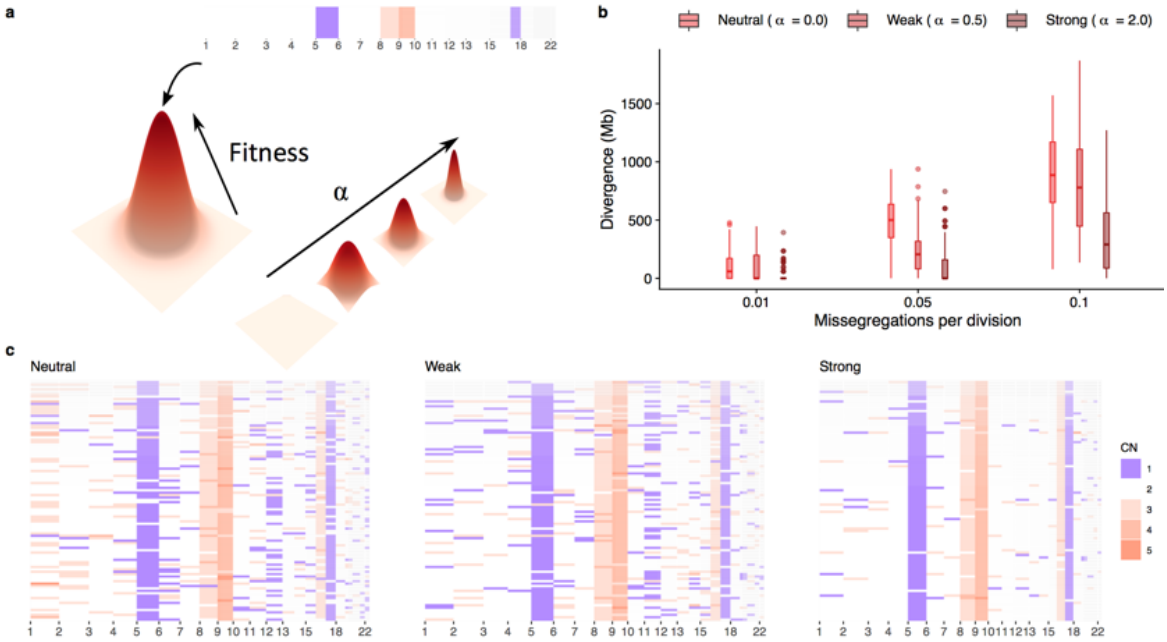


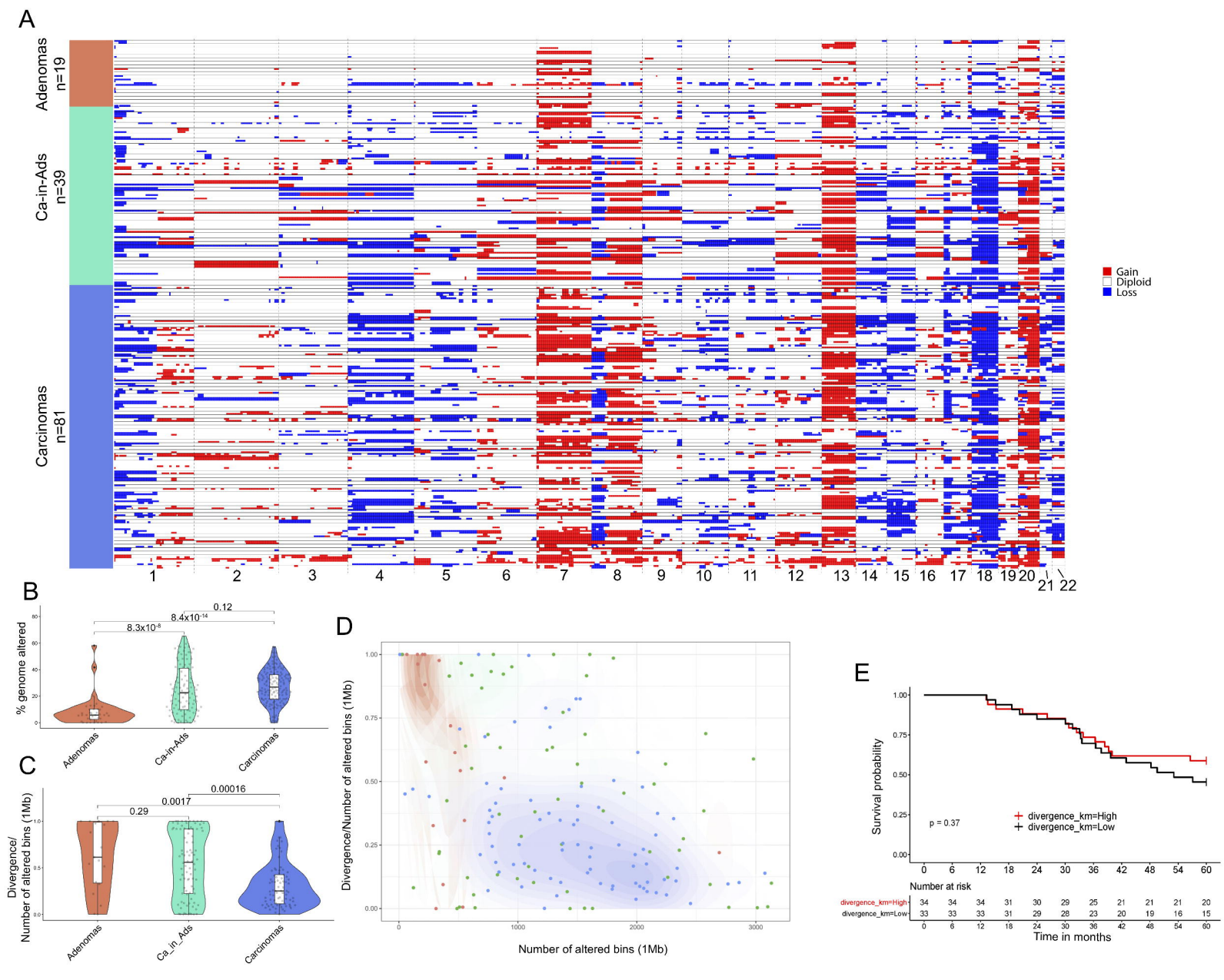
Figure 4: Simulations demonstrate negative selection suppresses karyotypic diversity in a growing cancer.

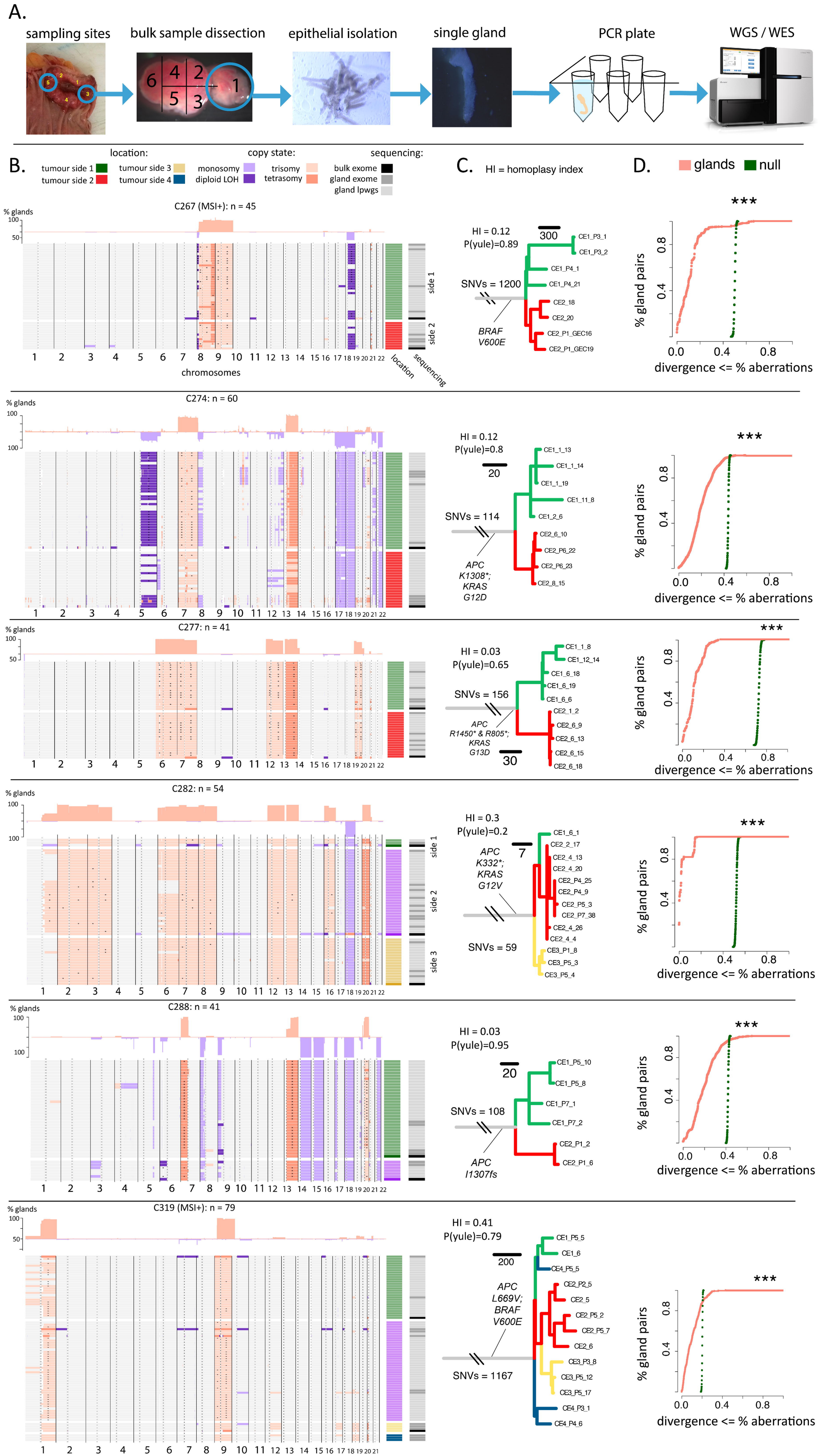
A: Schematic showing shape of fitness landscape. The model parameter α determines the steepness of the single fitness peak in the landscape, at position defined by the above heatmap illustrating the optimal CNA pattern.

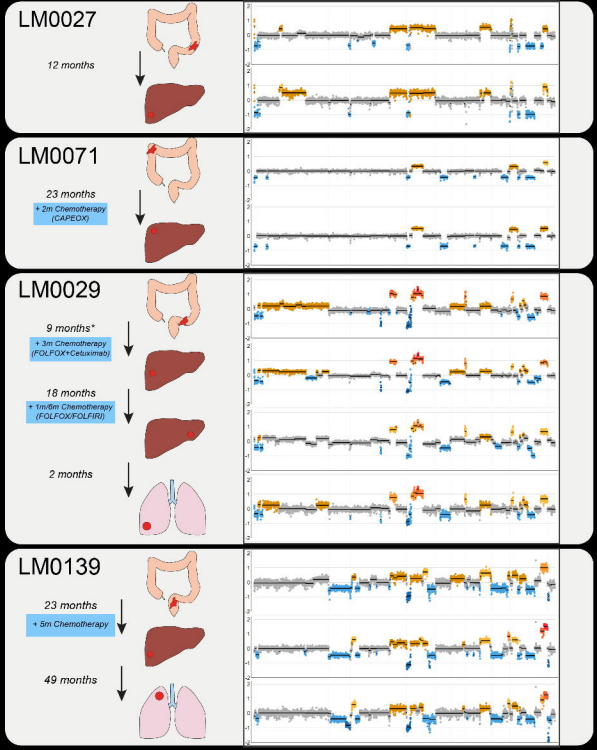
B: Genetic divergence as a function of CNA alteration rate and selection strength α. Strong selection on an initially optimal clone supresses diversity and higher mutation rates lead to more diversity.

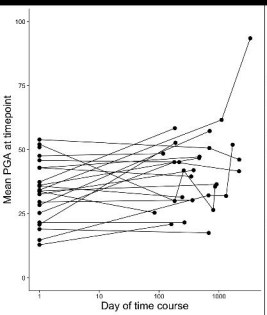
C: Heatmaps showing CNAs of 100 randomly sampled individual simulated cells after 20 population doublings as function of negative selective strength α; stronger selection suppresses within-tumour diversity.

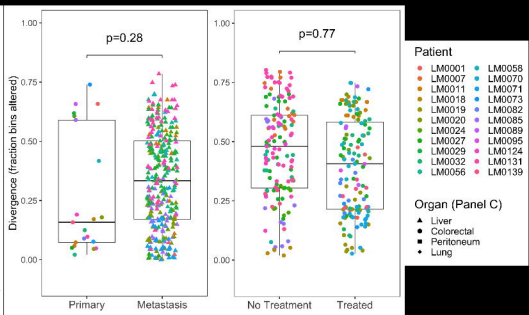


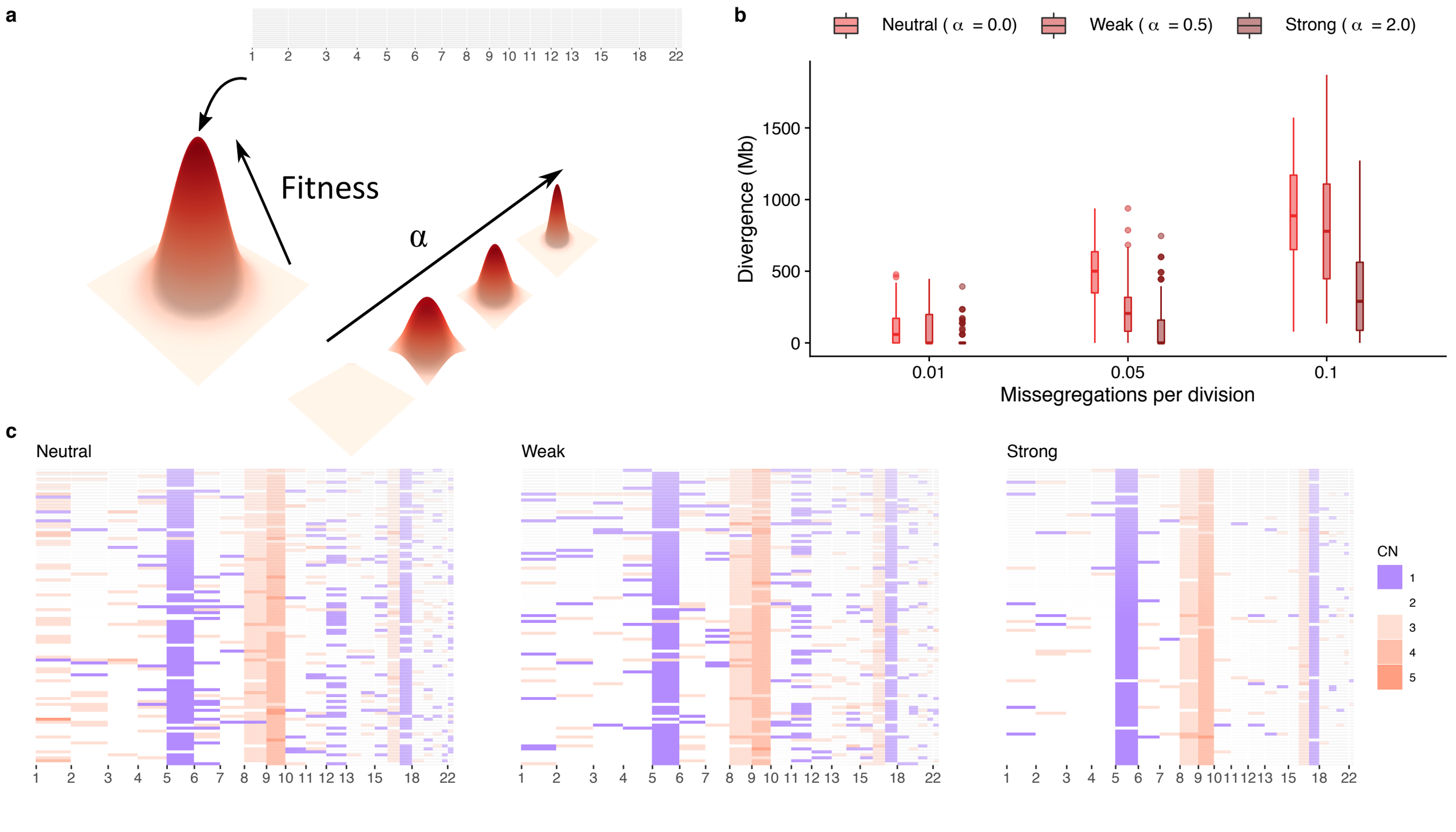

A
Adenomas n=19
Ca-in-Ads n=39
Carcinomas n=81
Gain
Diploid
Loss
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
B
% genome altered
0.12
8.4x10-14
8.3x10-8
Adenomas
Ca-in-Ads
Carcinomas
C
Divergence/ Number of altered bins (1Mb)
0.00016
0.0017
0.29
Adenomas
Ca_in_Ads
Carcinomas
D
Divergence/Number of altered bins (1Mb)
Number of altered bins (1Mb)
E
Survival probability
p = 0.37
divergence_km=High
divergence_km=Low
Number at risk
divergence_km=High 34 34 34 31 30 29 25 21 21 21 20
divergence_km=Low 33 33 33 31 29 28 23 20 19 16 15
0 6 12 18 24 30 36 42 48 54 60
Time in months


A.
sampling sites
bulk sample dissection
epithelial isolation
single gland
PCR plate
WGS / WES
B.
location:
tumour side 1
tumour side 2
tumour side 3
tumour side 4
copy state:
monosomy
diploid LOH
trisomy
tetrasomy
sequencing:
bulk exome
gland exome
gland lpwgs
C. HI = homoplasy index
D. glands null
C267 (MSI+): n = 45
HI = 0.12
P(yule)=0.89
SNVs = 1200
BRAF V600E
C274: n = 60
HI = 0.12
P(yule)=0.8
SNVs = 114
APC K1308*; KRAS G12D
C277: n = 41
HI = 0.03
P(yule)=0.65
SNVs = 156
APC R1450* & R805*; KRAS G13D
C282: n = 54
HI = 0.3
P(yule)=0.2
APC K332*; KRAS G12V
SNVs = 59
C288: n = 41
HI = 0.03
P(yule)=0.95
SNVs = 108
APC I1307fs
C319 (MSI+): n = 79
HI = 0.41
P(yule)=0.79
SNVs = 1167
APC L669V; BRAF V600E
% glands
chromosomes
location
sequencing
side 1
side 2
side 3
% gland pairs
divergence <= % aberrations
***


LM0027
12 months
LM0071
23 months
+ 2m Chemotherapy (CAPEOX)
LM0029
9 months*
+ 3m Chemotherapy (FOLFOX+Cetuximab)
18 months
+ 1m/6m Chemotherapy (FOLFOX/FOLFIRI)
2 months
LM0139
23 months
+ 5m Chemotherapy
49 months


Mean PGA at timepoint
Day of time course


p=0.28
p=0.77
Divergence (fraction bins altered)
Primary
Metastasis
No Treatment
Treated
Patient
LM0001
LM0007
LM0011
LM0018
LM0019
LM0020
LM0024
LM0027
LM0029
LM0032
LM0056
LM0058
LM0070
LM0071
LM0073
LM0082
LM0085
LM0089
LM0095
LM0124
LM0131
LM0139
Organ (Panel C)
Liver
Colorectal
Peritoneum
Lung


a
1 2 3 4 5 6 7 8 9 10 11 12 13 15 18 22
Fitness
α
b
Neutral (α = 0.0)
Weak (α = 0.5)
Strong (α = 2.0)
Divergence (Mb)
1500
1000
500
0
0.01
0.05
0.1
Missegregations per division
c
Neutral
Weak
Strong
CN
1
2
3
4
5