The evolutionary landscape of colorectal tumorigenesis

William Cross12, Michal Kovac2,3, Ville Mustonen4, Daniel Temko5,1, Hayley Davis6, Ann-Marie Baker1, Sujata Biswas6, Roland Arnold7, Laura Chegwidden6, Chandler Gatenbee9, Alexander R. Anderson6,9, Viktor H. Koelzer2,10, Pierre Martinez1, Xiaowei Jiang11, Enric Domingo5,2, Dan J. Woodcock12, Yun Feng2, Monika Kovacova13, Tim Maughan14, The S:CORT Consortium15, Marnix Jansen16, Manuel Rodriguez-Justo16, Shazad Ashraf17, Richard Guy18, Christopher Cunningham18, James E. East19, David C. Wedge10,12, Lai Mun Wang20, Claire Palles8, Karl Heinimann21,22, Andrea Sottoriva23, Simon J. Leedham6,19, Trevor A. Graham5,1* and Ian P. M. Tomlinson11,24*

The evolutionary events that cause colorectal adenomas (benign) to progress to carcinomas (malignant) remain largely undetermined. Using multi-region genome and exome sequencing of 24 benign and malignant colorectal tumours, we investigate the evolutionary fitness landscape occupied by these neoplasms. Unlike carcinomas, advanced adenomas frequently harbour sub-clonal driver mutations—considered to be functionally important in the carcinogenic process—that have not swept to fixation, and have relatively high genetic heterogeneity. Carcinomas are distinguished from adenomas by widespread aneuploidies that are usually clonal and often accrue in a ‘punctuated’ fashion. We conclude that adenomas evolve across an undulating fitness landscape, whereas carcinomas occupy a sharper fitness peak, probably owing to stabilizing selection.

The classical adenoma–carcinoma sequence of colorectal tumorigenesis postulates that a conventional colorectal adenoma (CRA) is initiated by ‘two hits’ at APC, and typically progresses to colorectal cancer (CRC) through a stepwise accumulation of driver mutations such as KRAS and TP53, and deletion of chromosome 18q. The evolutionary dynamics presumed to underlie this process comprise a series of selective sweeps to (near) fixation, each triggered by an elevation in sub-clone fitness through the occurrence of a new, positively selected driver mutation. In this model, progression to an invasive lesion (carcinoma) is postulated to be prompted by the acquisition of a critical driver mutation burden, implying that adenomas and carcinomas should be distinguishable by specific driver mutations. However, CRCs can develop without the full complement of driver mutations, and some studies have suggested that sub-clonal evolution within established tumours is ‘effectively neutral’, questioning whether selective sweeps occur at all, especially in established CRCs.

As part of a comprehensive assessment of colorectal tumour evolution, here, we have attempted to re-assess the classical model and outline the evolutionary ‘fitness landscape’ of CRAs and CRCs. The fitness landscape—a concept first introduced by Wright in 1932—is an abstraction to help visualize the relationship between genotypes and reproductive success (sub-clone fitness in this context). The x and y axes can be thought of as the genotype ‘space’ (simplified to two dimensions) that can be occupied by adenomas and carcinomas. The z axis, or height, is proportional to genotype fitness: peaks represent particularly fit genotypes, valleys less fit genotypes and ridges/plateaux equally fit genotypes. Individuals sampled
from a population are likely to occupy (local) fitness peaks because less fit individuals have been removed by negative (purifying or stabilizing) selection. Herein, we search for the genotypes associated with the fitness peaks occupied by CRAs and CRCs and probe peak shapes by quantifying intra-tumour heterogeneity (ITH). Transitions around the landscape are measured using phylogenetic and molecular clock analyses. These data provide an understanding of the evolutionary trajectories underpinning the development of CRAs and CRCs.

Results

To map the evolutionary landscape of CRAs and CRCs, we performed multi-region whole-genome sequencing (WGS) or whole-exome sequencing (WES) on 2–16 regions (total 118) from 9 CRAs and 15 CRCs, each with constitutional DNA (see Supplementary Table 1 for sample details and Supplementary Table 2 for sequencing statistics). Five CRCs, including four from Lynch syndrome patients, had microsatellite instability (MSI) owing to defective DNA mismatch repair, and these tumours were analysed as a distinct group unless otherwise stated. The remaining ten CRAs were microsatellite stable (MSS) and, of these, two were synchronous lesions from a single patient. Mutations in a subset of genes were validated using targeted molecular inversion probe sequencing (see Methods).

Somatic single nucleotide alterations (SNAs) do not define CRC fitness peaks. First, we assessed how somatic SNAs defined the coordinates of CRAs and CRCs in the fitness landscape. CRAs tended to have only slightly fewer SNAs than MSS CRCs (CRA median exonic burden = 94, 95% confidence interval (CI): 51–146; MSS CRC median exonic burden = 130, 95% CI: 98–171; P = 0.29, Wilcoxon test; Fig. 1a and Supplementary Table 2). After sequencing coverage normalization, the mutational frequency in CRAs remained very similar to that of MSS CRCs (CRA: 4.1 megabase (Mb)−1; 95% CI: 3.3–4.9; MSS CRC: 4.2 Mb−1; 95% CI: 2.9–6.4; P = 0.9).

Next, we compared the burden of driver mutations across CRAs and CRCs, and included SNAs and indels, as well as copy-neutral loss of heterozygosity (cnLOH) and monosomy (chromosome loss) events that are known to act as ‘second hits’ to the tumour suppressor genes APC and TP53, and also 18q allelic loss/imbalance (Fig. 1c,d and Supplementary Table 3). The burden of tier 1 mutations, which we defined as likely pathogenic changes in known CRC driver genes (see Methods and Supplementary Table 4), was not significantly different in our cohort (CRAs: median = 5 (range: 2–9); MSS CRCs: median = 6 (range: 2–8); P = 0.9). We noted that the difference remained non-significant when comparing drivers across individual biopsies (P = 0.19, Wilcoxon test). Individual tier 1 driver mutations were detected at similar frequencies across CRAs and CRCs, with the exception of TP53, which was more commonly mutated (possessing at least one SNA, indel or copy change) in CRCs (Fisher’s test, P = 0.005; see Fig. 1d and Supplementary Table 5). The frequency of tier 2 driver mutations (uncertain pathogenicity changes in CRCs or pan-cancer driver genes) was also not discernably different in CRAs and MSS CRCs (CRAs: median = 3 (range: 2–4); MSS CRCs: median = 3 (range: 1–7); P = 0.8). Several tier 2 driver mutations were specific to CRAs or CRCs, but most occurred infrequently; only KMT2C was notable, being mutated in four CRAs and no CRCs. The total median driver mutation burdens (tiers 1 and 2 combined: CRAs = 7 (range: 2–12); CRCs = 8 (ranges: 3–15); P = 0.6; Fig. 1c and Supplementary Tables 3 and 4) were also similar. Furthermore, when using an alternative definition of driver genes (the top 15 genes mutated in MSS CRCs (excluding TTN), according to a publication by The Cancer Genome Atlas (TCGA)); Supplementary Table 4b), the burdens remained non-statistically different between CRAs and CRCs (CRAs: median = 5 (range: 2–7); MSS CRCs: median = 5.5 (range: 2–9); P = 0.7).

The power to detect differences in the mutation burden between CRAs and CRCs was limited by sample size, such that differences in the exonic mutation burden smaller than approximately 30 mutations could not be detected with high power (see Supplementary Methods for post-hoc power calculations). With this constraint in mind, the SNA mutation burden (including tier 1 driver mutations) did not distinguish the relative coordinates of CRAs and CRCs in the evolutionary landscape.

ITH and phylogenetic analyses suggest that CRCs occupy sharper fitness peaks than CRAs. To broadly assess the shape of the fitness peaks occupied by CRAs and MSS CRCs, we measured the degree of ITH in each tumour. Excluding tumours with only two regions sampled (see Methods), a median of 56% (95% CI: 53–70) of all CRA SNAs were ‘sub-clonal’ (variant not detected in all sampled regions). MSS CRCs had a significantly lower proportion of sub-clonal SNAs (45% (95% CI: 23–77%); P = 0.04; Fig. 2, inset) than CRAs. The average pairwise genetic divergence between the regions of each tumour was then assessed following normalization of the sequencing coverage (Methods). CRAs showed significantly more divergence between biopsies than CRCs (mean divergent SNAs Mb−1: CRA = 2.0; CRC = 1.7; P < 2 × 10−4; Supplementary Fig. 1a) despite having the same average mutation burden. The measured values of ITH were unaffected by the number of biopsies available from each neoplasm (Supplementary Fig. 1b,c).

To further quantify ITH, we used SNAs to construct maximum-parsimony phylogenetic trees (Fig. 2). CRC topologies were often characterized by long trunks (variants ubiquitous across biopsies) with comparatively short branches and leaves (relatively few sub-clonal variants), thus appearing ‘palm tree’ shaped. CRAs had proportionally shorter trunks, and thus longer branches and leaves than CRCs, albeit at borderline significance (average branch and leaf length as a proportion of the trunk: 82% for CRAs versus 50% for MSS CRCs; P = 0.06; Fig. 2 and Supplementary Table 6). The difference remained when the MSI–CRCs were included in the analysis (82% for CRAs versus 45% for all CRCs; P = 0.05). CRAs are thus more genetically diverse than CRCs.

To investigate whether individual CRAs and CRCs occupied single or multiple fitness peaks, we compared the lengths of the phylogenetic tree branches and leaves. Large variations in branch length indicate that mutations accrue faster in some tumour regions than others, which can potentially be caused by selection on a new fitness peak. Average intra-tumour variation in relative branch and leaf length was generally low and similar across CRAs and CRCs (mean s.d. (95% CI): 0.14 (0.06–0.24) for CRAs versus 0.2 (0.06–0.47) for CRCs; P = 0.68; Fig. 3 and Supplementary Table 6). Formal assessment of unbalanced tree topologies could only be performed on one tumour (carcinoma 6) as high numbers of samples are needed for sufficient power. Unbalanced trees occur when some ancestor clones produce more surviving lineages than another—another potential indicator of sub-clonal selection. We did not find any significant asymmetry in this single tumour analysis (Colless’ test, Yule model, P = 0.3). Thus, the available data were consistent with the idea that tumours occupied a single, potentially broad fitness peak.

The SNA-based ITH and phylogenetic analyses suggested that CRAs were more heterogeneous than CRCs, consistent with CRAs occupying a broader fitness peak, under which several distinct genotype–phenotype combinations could co-exist. However, the lower ITH in CRCs could also reflect a more recent selective sweep with a genetic bottleneck during the transition from an adenoma, and/or that CRCs were more spatially mixed than CRAs, causing variants at sub-clonal frequency in multiple samples to appear truncal. We therefore directly sought evidence of stronger selection in CRCs by examining the ratio of non-synonymous to synonymous mutations on tumour trunks and branches/leaves. This showed a reduction in non-synonymous mutations

1662 NATURE ECOLOGY & EVOLUTION | VOL 2 | OCTOBER 2018 | 1661-1672 | www.nature.com/natecolevol
on the branches/leaves of CRCs relative to their trunks (Wilcoxon signed-rank test, $P = 0.01$; Supplementary Fig. 2), but no such reduction for CRAs ($P = 0.9$), possibly representing ongoing positive sub-clonal selection in CRAs. On the reasonable assumption that positive selection acted on the phylogenetic trunk—the location of almost all tier 1 driver mutations—together, these results indicate that sub-clonal selection is absent (neutral dynamics) or weak within the established carcinoma, with possible negative (stabilizing) selection also at play.

**Mutational processes are not detectably associated with fitness advantages.** Mutation signatures were identified de novo using the EMu programme. We recovered aging, MSI-associated and molecular clock signatures (our signatures A, B and C, respectively), as expected (Supplementary Fig. 3a). Our signature D, which resembles Catalogue of Somatic Mutations in Cancer (COSMIC) signature 17 (unknown aetiology; high CTT > CGT frequency) was present at appreciable levels within carcinomas 2, 7, 9P and 10, with its activity often differing between the trunks and branches/leaves of the same
lesion (Supplementary Fig. 3b,c). We explored whether signature D had any effect on sub-clonal evolution in CRCs using WGS. It appeared to increase the mutation burden in two CRCs (2 and 9P), but had no discernible effect on their evolution (see Supplementary Fig. 3d for details). Carcinoma 9D—the synchronous partner of carcinoma 9P—showed low signature 17 activity, despite being located only 10 cm apart in the bowel. These cancers also had different driver mutations, confirming that they essentially behaved as independent neoplasms, with no detectable effect of any shared microenvironment on mutagenic processes (Supplementary Fig. 4).

Major driver mutations can be sub-clonal in CRAs, but are very rarely so in MSS CRCs. Tier 1 driver mutations (defined above) were typically, but not always, clonal in CRAs, whereas in MSS CRCs they were more commonly clonal. However, these distributions were not significantly different between tumour types (CRAs = 39/49 (80%); CRCs = 45/55 (83%); \( P = 0.3 \)). However, the clonal distributions of tier 2 clonal driver mutations were different; CRAs had significantly fewer clonal drivers than MSS CRCs (CRAs = 7/15 (47%); CRCs = 21/26 (80%); Fisher’s exact test, \( P = 0.03 \); Supplementary Table 3). We noted that the clonality of tier...

---

**Fig. 2** | Phylogenetic analysis of CRAs and MSS CRCs. Maximum-parsimony construction of evolutionary trees. For tumours with only two regional biopsies (B), truncal mutations are those shared between the regions. Tier 1 driver mutations (Supplementary Table 3) are shown, illustrating their enrichment on the trunks, especially in CRCs, indicating they were acquired early in evolutionary time. Phylogenetic trees were produced using all available SNAs. Tree shape robustness (branch support) was confirmed by bootstrapping. Branches had greater than 95% support unless otherwise stated (44/55 (80%) of branches had >95% support). The homoplasy index (HI) is shown on each phylogeny. The most parsimonious trees are shown, with more than one region) and leaves (present in only one region). CRAs have a smaller proportion of ubiquitous variants than CRCs (median percentages are 0 50 100 % SNAs on trunk, branch or leaf. HI: 0.25 HI: 0.36 HI: 0.39 HI: 0.43 HI: 0.45 HI: 0.46 HI: 0.47 HI: 0.49 HI: 0.50 HI: 0.51 HI: 0.52 HI: 0.53 HI: 0.54 HI: 0.55 HI: 0.56 HI: 0.57 HI: 0.58 HI: 0.59 HI: 0.60 HI: 0.61 HI: 0.62 HI: 0.63 HI: 0.64 HI: 0.65 HI: 0.66 HI: 0.67 HI: 0.68 HI: 0.69 HI: 0.70 HI: 0.71 HI: 0.72 HI: 0.73 HI: 0.74 HI: 0.75 HI: 0.76 HI: 0.77 HI: 0.78 HI: 0.79 HI: 0.80 HI: 0.81 HI: 0.82 HI: 0.83 HI: 0.84 HI: 0.85 HI: 0.86 HI: 0.87 HI: 0.88 HI: 0.89 HI: 0.90 HI: 0.91 HI: 0.92 HI: 0.93 HI: 0.94 HI: 0.95 HI: 0.96 HI: 0.97 HI: 0.98 HI: 0.99 HI: 1.00
1 driver mutations was the same when using the second definition of driver mutations based on the TCGA publication (37/43 (86%) versus 45/53 (85%); Fisher’s exact test, P > 0.9; see Methods). These findings are consistent with a scarcity of sub-clonal expansions after the most recent common ancestor (MRCA) in CRCs. This trend seems to be similar in CRAs, although CRAs show some evidence of sub-clonal driver mutations.

However, we also noted that the most frequently mutated CRC driver genes, apart from the probable tumour-initiating mutations in APC, were sub-clonal in at least one CRA. Notably, in adenoma 2, KRAS Q61H and an ARID2 frameshift mutation were present in one region, which was separate from the three regions of this tumour that contained a TP53 E219X mutation. Adenoma 3 had a PIK3CA E545K mutation in two tumour regions, GNAS R201H in another.

**Fig. 3 | CNAs in CRAs and MSS CRCs.** a. A genome-wide view of CNAs is shown for each region of CRAs (top) and CRCs (bottom). CRCs show a greater CNA burden than CRAs, and most CNAs are clonal in CRCs, whereas CRAs show more frequent sub-clonal CNAs. A copy number ≥5 is shown as ‘polysomy’. b. Estimated ploidy, with a summary of the proportion of each tumour at different copy states. Black bars show the range of biopsy copy numbers. c. Size distributions of ubiquitous and sub-clonal (branch and leaf) CNAs demonstrate the preference of CRCs to have larger events. Box plots show the median and interquartile range (IQR), the upper whisker is the third quantile + 1.5x the IQR and the lower whisker is the first quantile – 1.5x the IQR. The colour coding of copy-number states (top right) applies to all panels.
Articles NaTUre eCology & evolUTioN

Fig. 4 | Geography of CRCs. Photographs of the tumour specimens from histopathology, with the biopsy locations marked (numbered B1, B2, and so on). The sporadic MSI+ cancer 4 is included here. The corresponding phylogenetic relationship between tumour regions is shown below the photograph of each tumour. The regression plots show pairwise physical and genetic separation for each biopsy from that cancer (line represents the regression itself, fitted to the origin, and 95% CI). There was a significant positive correlation between the phylogenetic (mutational) distance and physical distance in every case.

and an AKA9P frame-shift in another. SMAD4 R496H in adenoma 4 was also present in a single region. There was no evidence from the phylogenetic analysis that these proven driver mutations were associated with differential sub-clonal expansion, suggesting that their selective benefits were relatively modest (Fig. 2). NRAS G60V and PIK3CA H1047R were present in both regions of adenomas 7 and 8, respectively, but were putatively sub-clonal since their corrected allele frequencies were significantly lower (P < 0.05) than those of other driver mutations, suggesting that biopsies crossed sub-clonal boundaries. In contrast, only one sub-clonal mutation with high-confidence pathogenicity (CHD1 R619X in carcinoma 1) was found in the MSS CRCs. There was no evidence for parallel evolution of sub-clones based on recurrent known or novel drivers (details not shown).

Next, we wished to relate the heterogeneity of mutational burdens to fundamental molecular processes. Immunohistochemistry for Ki67 (proliferation) and β-catenin (activated Wnt signalling; Supplementary Fig. 5 and Methods) showed positive cell fractions of 53% (2–80%) and 82% (3–97%), respectively, with considerable variability between and within CRCs (Supplementary Table 7). Neither Ki67 nor β-catenin expression was associated with regional SNA burden or ploidy (SNA burden: coefficient of determination, R² = 0.2, P = 0.2; ploidy: R² = 0.9, P = 0.08; Supplementary Fig. 5).

Genetic and spatial relationships between CRC sub-clones. In all CRCs, physical and phylogenetic distances between biopsies were strongly correlated (R² = 0.81–0.93, P < 10⁻⁴ for all carcinomas measured; Fig. 4). The invasive edge of CRCs and central regions had similar mutational burdens (exonic SNAs, edge versus central; P = 0.76). We looked further for sub-clonal mixing within the sampled regions of the MSS CRCs with WGS data by clustering SNA cancer cell fractions (CCFs) across related samples using a Dirichlet process-based model (Supplementary Fig. 6). Only 10% of biopsy samples showed evidence of ≥1 sub-clonal population. While we do not exclude a degree of sub-clonal intermingling, these results suggest that, given the depth of our sequencing data, sub-clonal expansions broadly occurred in a spatially contiguous, uniform and discrete fashion.

Copy-number changes differ between CRAs and CRCs. Next, we assessed whether somatic copy-number alterations (CNAs) might define the fitness peaks occupied by CRAs and MSS CRCs. Every region of every tumour carried at least one CNA, including cnLOH (see Methods). In a combined analysis of all regions from each tumour, as expected, CRAs had fewer CNAs (number of discrete CNA segments >1 Mb) than CRCs16,17 (CRAs: median = 13 (95% CI: 7–11); CRCs: median = 40 (95% CI: 15–42); P = 0.003; Fig. 1b). Correspondingly, the overall average proportion of the genome disrupted by CNAs (copy number ≠ 2; allelic ratio ≠ 1) was higher in CRCs (72% for CRs; 40% for CRAs; P = 0.05; Fig. 3). These data show that CRCs display higher CNA levels than CRAs.

It is currently difficult to identify driver CNAs with certainty in cancer18. In colorectal tumours, losses (deletions or cnLOH) on chromosomes 5q, 17p and 18q are often thought to be second hits involving the tumour suppressors APC, TP53 and SMAD4, respectively (although 18q loss is more common than SMAD4 mutation).
The status of the other recurrent changes, such as 1q gain, 7 gain, 8p deletion, 13q gain and 20 gain, as drivers or passengers is less clear. Many recurrent, and hence potential driver, CNAs were present at significantly higher frequencies in CRCs compared with CRAs (Fig. 3). Notably, 17p loss occurred in 9/10 MSS CRCs, but only 2/9 CRAs (Fisher's exact test, \(P = 0.005\)), paralleling the TP53 SNA data. By comparison, loss at the APC locus (8/10 CRCs versus 5/9 CRAs; \(P = 0.35\)) and SMAD4 locus (7/10 CRCs versus 4/9 CRAs; \(P = 0.37\); Fig. 2) occurred at similar frequencies in both lesion types.

Every tumour had at least two clearly sub-clonal CNAs (non-ubiquitous, present versus absent changes; Fig. 3a) and no chromosome aberration was exclusively ubiquitous or sub-clonal across the tumours. Overall, 75 and 48% of gains were sub-clonal in CRAs and CRCs, respectively (\(P = 0.002\)), compared with 57 and 27% of losses/cnLOH (\(P = 0.007\); Supplementary Fig. 7). Thus, a greater proportion of CNAs were sub-clonal in CRAs than in CRCs.

We compared the size distribution of large (>1 Mb) CNAs in early (truncal) versus late (sub-clonal) tumour evolution. In CRCs, sub-clonal CNAs were smaller than ubiquitous CNAs (\(P < 0.001\); Supplementary Fig. 3c), but this difference was not present in CRAs (\(P = 0.45\)). The lower frequency of large CNAs later in evolutionary time in CRAs may suggest that the cancers have obtained a near-optimal level of aneuploidy, with further large-scale CNAs subjected to negative or stabilizing selection. In adenomas, since the overall CNA burden is lower, new large CNAs may still be tolerated.

**MSS CRC evolution can involve either ‘punctuated’ or more gradual CNA acquisition.** Since CNAs were the principal genetic feature distinguishing between CRAs and CRCs, we investigated their role in the transition between the benign and malignant fitness peaks. Using a similar strategy to Durinck et al.\(^a\) and Newman et al.\(^a\) (details in Methods), we used the SNAs within informative chromosomal segments (copy-number gains and cnLOH) as a molecular clock to time the occurrence of that CNA. SNAs present on a chromosome before its gain increase in variant allele frequency (VAF) following the copy-number change, whereas SNAs that accrue after the gain are present only on one allele and so remain at lower VAF. The ratio of higher to lower VAF SNAs therefore estimates the time of CNA occurrence.

Sufficient SNAs for molecular clock analysis were only present in WGS data. Of the five MSS CRCs analysed by WGS, carcinomas 3, 9P and 10 showed a clustering of CNA timings shortly before the MRCA (Kolmogorov–Smirnov test against a uniform distribution of CNA timings, \(P < 0.02\) for all; Fig. 5). A similar, borderline significant CNA cluster occurred in carcinoma 9D. Carcinoma 5 showed a more gradual accumulation of CNAs.

Since the timing method demonstrated a form of ‘punctuated’ CNA evolution (rejection of the null hypothesis of uniform accumulation), but did not distinguish between multiple gains of individual chromosomes and genome doubling followed by chromosomal gain or loss, we searched heuristically for evidence of genome doubling using a score based on the number of chromosome centromeres present at copy number 4 or above, with extra weight for allelic balance (Supplementary Fig. 8). Based on this measure, all of the CRAs with significantly or borderline significantly clustered CNA timings (\(n = 4\)) were genome doubled on this measure, as was the untyped carcinoma 8. The CNAs in these tumours are typically trisomies, judged to have arisen by chromosome (arm) loss subsequent to allele-balanced genome doubling. The other tumours (including carcinomas 1, 2, 4, 5, 6 and 7) were scored as non-genome doubled. Sub-clonal genome doubling was present in one CRA (adenoma 2). This tumour carried a TP53 mutation in its genome-doubled regions, and overall TP53 mutations (SNAs and/or CNAs) were associated with genome doubling in MSS CRCs (Fisher's exact test, \(P = 0.018\)). In addition, genome-doubled cancer regions had higher Ki67 expression (see above; \(P = 0.04\); Supplementary Fig. 5), hinting at the existence of a selective benefit of doubling.

**The evolutionary landscape of MSI+ CRCs.** The overall SNA burden of the five MSI+ CRCs was, as expected, far higher than in MSS CRCs (Supplementary Fig. 9a). More pointedly, the number of tier 1 CRC driver mutations was also higher (median = 12 (range: 4–14)) than in MSS CRCs (median = 3; \(P = 0.042\); Supplementary Fig. 9b), while the CNA burden was lower (Supplementary Fig. 9c). Of note, in MSI+ CRCs, the great majority of driver SNAs were truncal, the number of sub-clonal tier 1 drivers was only a little greater (median = 1 (range: 0–7)) than in MSS CRCs, and the proportion of all sub-clonal SNAs was not significantly increased (median: 34% for MSI+ CRCs versus 42% for MSS CRCs; \(P = 0.13\); Supplementary Fig. 9d). In phylogenetic analysis, neither the average branch/leaf length as a proportion of the trunk nor its variability differed significantly between MSI+ and MSS CRCs (Supplementary Fig. 9e). Our signature B (COSMIC signature 6) predominated in MSI+ CRCs, especially on the branches/leaves, but the other COSMIC MSI-associated signatures\(^b\) were not detected. Overall, the data suggest that MSI+ CRCs evolve in a similar way to MSS CRCs, albeit with some limited evidence of sub-clonal selection.

**Discussion**

Here, we have contrasted the patterns of evolution in CRCs and their classical adenomatous precursor lesions, and our data begin to reveal the shape of the fitness landscape over which CRCs grow. CRAs tend to evolve through the acquisition of major driver mutations in genes such as APC, KRAS and TP53 and via 18q loss, as per the Vogelstein model\(^c\). More recently discovered cancer driver mutations are also present in many adenomas (Supplementary Tables 3 and 4). In fact, CRAs can harbour mutations in any of the major CRC driver genes, but these mutations do not necessarily occur in a stereotypical order. Driver mutation acquisition also does not necessarily cause selective sweeps (leading to ‘stepwise’ evolution of the tumour cell population), since sub-clones with additional major driver mutations may not displace sub-clones lacking those mutations, but instead may co-exist in spatially discrete areas. It follows that many driver mutations probably confer a relatively small selective advantage. This is reflected in several observations in CRAs, including a relatively high level of genetic diversity (both SNAs and CNAs), variation in the major driver mutation complement in different regions of individual tumours, and phylogenetic trees with relatively long branches/leaves. It is even possible that SNA accumulation is not an essential feature of tumorigenesis before malignancy, and we speculate that carcinomas need not arise from the sub-clone with the greatest number of driver mutations, thus explaining why some CRCs have a very small driver mutation complement\(^d\).

MSS CRCs have longer phylogenetic tree trunks than branches/leaves compared with CRAs. These findings may reflect the influence of several factors, including not only selective constraints, but also time from the MRCA after an additional selective sweep, ploidy, sample purity and genomic instability. Overall, the lack of sub-clonal driver SNAs and reduction in non-synonymous SNAs on the branches and leaves of CRCs suggest that there is not strong positive sub-clonal selection for SNAs after the MRCA. In contrast, CRAs show sub-clonal drivers and relatively high ITH, together providing evidence of (perhaps relatively weak) sub-clonal selection.

Although present in CRAs, large CNAs and genome doubling are much more common in CRCs. CNAs on CRC tree branches/leaves are smaller than those on trunks. While negative or stabilizing selection remains difficult to measure, this is consistent with the relatively low genetic diversity in CRCs, based on SNAs and large CNAs. For most MSS CRCs, a near-triploid karyotype seems optimal, either through genome doubling followed by loss of some chromosomes or through a gain of chromosomes that mostly
occurs within a putatively short time window between malignant progression and the MRCA. In each case, one or more selective sweeps seem to occur, rendering the driver SNAs and most CNAs clonal. We do not exclude additional positive selection for specific sub-clonal CNAs in CRCs, but this remains unproven and, indeed, our data showed no evidence of sub-clonal selection. Although every CRC had at least one sub-clonal CNA, we found no evidence of parallel CNA evolution.

In all our MSI+ cases, defective MMR and most major driver mutations arose on the phylogenetic trunk, and the relative branch/leaf length was similar to that of MSS CRCs. Although the sporadic MSI+ cancer had a low driver mutation burden, as expected if driven in part by a methylator phenotype, its evolution was otherwise similar to the Lynch syndrome CRCs. We speculate that MSI+ CRCs experience either multiple selective sweeps driven by individual SNAs or, more intriguingly, by co-occurring or epistatically acting non-canonical driver SNAs (such as CTNNB1, SOX9, NF1 and CASP8).

A small number of ITH studies have been undertaken previously in CRC. Kim et al. performed multi-region WES of 5 primary and metastatic CRCs and 9 CRCs, respectively, and Suzuki et al. performed deep targeted sequencing of 799 genes in 4 CRCs. Similar to our study, these studies reported that major driver mutations, affecting APC, KRAS and TP53, were truncal, with the exception of PIK3CA. Uchi et al. also reported that large copy gains were common on the trunk of the evolutionary tree, with focal deletions on branches.

Fewer studies of CRA evolution exist. Kim et al. used WES to compare malignant and benign regions of four mixed cancer-in-adenoma polyps. They reported similar SNA burdens in cancer and adenoma regions, and thus suggested that the regions evolved in parallel, rather than the carcinoma progressing from a late adenoma. However, we note that it is extremely difficult to distinguish between benign and malignant components of these lesions, since malignancy is defined not by cytology but by invasion, and hence the location of tumour cells. The different neoplastic components of such polyps may therefore, in reality, both be ‘cancerous’. For these reasons, in this study, we compared advanced CRAs and CRCs that were distinct lesions.

Previous work from our group examined single glands from 11 CRCs and 4 CRAs for CNAs, and for AmpliSeq panels of SNAs that had been derived from bulk tumour WES. Although that manuscript and our present study had very different focuses, the findings are consistent. For example, one feature of the ‘Big Bang’ model of sub-clonal intermixing expounded in the previous study is that after the MRCA, CRC sub-clones radiate outwards without notable differential sub-clonal expansion or selection of further advantageous variants. Our present study is broadly consistent with those data. However, there are also some differences between the studies that allow refinement of the ‘Big Bang’ model. For example, our current
study, which benefits from the significantly increased genomic resolution of WGS and WES, emphasizes that large sub-clones after the MRCA remain spatially restricted in CRCs, and consequently that the previously observed, widespread clonal ‘intermixing’ in some CRCs may reflect the shape, size and boundary location of discrete sub-clones, as well as outwards radiation of low-frequency clones (which were undetectable in our study). Furthermore, while the ‘Big Bang’ was broadly consistent with ITH measured in CRAs, our present study finds that sub-clonal driver mutations in the absence of selective sweeps occur commonly in these tumours.

In a study analogous to ours, Stacherl et al. 19 exome sequenced 5–11 samples of oesophageal carcinoma and its precursor, Barrett’s oesophagus from 5 patients. Comparing the two studies reveals both similarities and differences. Barrett’s oesophagus is not a discrete tumour and is generally a highly polyclonal lesion, reflected in multiple ‘initiating’ deletion mutations in CDKN2A and a series of clonal expansions without selective sweeps. In comparison, CRAs are discrete and probably have mononclonal origins usually caused by biallelic APC mutation, followed either by selective sweeps or by polyclonal expansions reminiscent of Barrett’s oesophagus. We note that in both Barrett’s oesophagus and CRA, data are consistent with malignant progression sometimes occurring from a sub-clone that does not have the largest driver mutation burden.

In summary, we have used measurements of ITH to reveal the evolutionary trajectories of colorectal tumour cell populations across what appears to be a rather flat fitness landscape for adenomas, with a higher, sharper peak occupied by cancers. Our data refine the Vogelstein model of CRC progression by showing that driver mutations do not necessarily lead to hard selective sweeps, and that progression to CRC can involve punctuated evolution.

Methods

Sample acquisition and processing. Oxfordshire Research Ethics Committee B gave permission for the study (protocol 05/Q1605/66), and all samples were collected with informed patient consent, which was obtained by the local clinicians before tissue collection. Fresh-frozen biopsies from 24 CRAs or CRCs were obtained from the John Radcliffe Hospital, Oxford or from University Hospital, Basel. In the case of CRAs, the biopsies were obtained from endoscopic resections, whereas the CRC biopsies were obtained from surgical resections. Any residual bulk cancer surplus to diagnostic requirements was also acquired. Paired normal biopsies were also taken from regions clearly separate from the tumour.

Library preparation and sequencing. DNA was extracted from tumour regions and morphologically normal tissue using the QiaGen DNeasy kit. The sequencing library preparations were performed using either the NENxt Exome kit or, in the case of exome sequencing, the Illumina TruSeq Exome kit. Sequencing of these biopsies was carried out using standard protocols on the Illumina HiSeq 2000 by the Genomics Core at the Wellcome Trust Centre for Human Genetics, Oxford. The FastQC programme was used to assess the raw sequencing quality, and coverage and depth were assessed using the GATK package (specifically, the DepthOfCoverage module).

Pre-processing and nucleotide variant calling. Reads in FASTQ format were aligned to version 19 of the human genome reference using BWA version 0.7.5 (ref. 20). The Picard package was used to identify duplicate reads and the SAMtools package was used to count the number of reads in the binary alignment map (.bam) files. SNA calling was performed using the platypus tool. Variant call format (.vcf) files were annotated with ANNOVAR and converted to tab-delimited file format using snp2hts. SNAs were categorized as somatic if they were present in at least 1 tumour sample and, either the normal sample had <40% coverage and 0 mutant reads or ≥40% depth and ≤1 mutant read. To obtain high confidence and consistent variants for phylogenetic analysis, only variants called with depths consistently ≥10x and an allele frequency ≥1% in 1 or more regions were retained.

Calling small insertions and deletions. Indel calling remains problematic and highly inaccurate. For this reason, we focused mainly on potential driver events. We identified a set of reliable indels by performing a first pass using the Scalpel tool then verifying the presence in the Platypus call sets, and also by visual inspection using the IGV browser. On visual inspection, we found that many ubiquitous indels were actually called incompletely across the biopsy sets by Platypus and sometimes missed all together by Scalpel, highlighting the inconsistency of the currently available tools.

Driver mutation identification. We used two classifications of driver genes (Supplementary Table 3). First, we identified CRC drivers using the IntOGen database 2016.5. Driver mutations were then classified using a two-tier system. Tier 1 driver mutations were considered very likely to be involved in colorectal carcinogenesis and included canonical mutations such as the most protein-truncating APC SNAs, BRF1 V600E and KRAS codon 12, 13, 61, 117 and 116 changes. For non-canonical mutations, tier 1 status was assigned to protein-truncating mutations in tumour suppressor genes or recurrently occurring mutations in CRAs or CRCs in the COSMIC database (http://cancer.sanger.ac.uk/cosmic). Tier 2 mutations, considered to be of lower confidence as drivers comprised all other coding or splice site changes in the same set of genes. These genes are defined in Supplementary Table 3a. Second, we used a more restrictive definition of the top 5% of significantly mutated genes in the TCGA publication exclusively (non-MSI+ carcinomas). These genes are defined in Supplementary Table 3b.

Coverage normalization. We normalized for differences in sequencing coverage to avoid bias in mutation calling and ITH measurements due to unequal coverage between samples. To do this, we identified a subset of 110,533 exonic regions consistently sequenced at ≥10x in all biopsies across all tumours. To normalize coverage between samples, we individually sub-sampled each bam file such that each contained a roughly equal number of reads. This equated to around 17,000,000 reads per biopsy, covering a total of 25 Mb of exonic regions.

We then generated 100 sets of ‘mini-bams’ (one bam from each sample), where each bam contained the same 10,000 randomly selected regions from the initial set. For each set of regions, we repeated the joint Platypus calling procedure (as per the above). The mutation frequency was calculated by simply taking the resulting number of variants and the total length of the 10,000 regions in each iteration.

Deep sequencing. To validate the exome and genome sequencing, and to search for additional low-allele-frequency somatic mutations, we sequenced a panel of 50 CRC driver genes in selected tumour regions. These genes encompassed all well-established driver events, including the top 15 genes from the IntOGen list (see Supplementary Table 3), and included both tier 1 and 2 driver mutations. Coding regions were captured using molecular inversion probes and sequenced using the Illumina NextSeq. Molecule tagging was used to ensure that reads were derived from unique tumour DNA molecules. Further details of the gene panel and sequencing protocol are available on request.

All 44 of the subset of driver mutations identified by WES or WGS from 3 carcinomas and 5 adenomas were validated. We were also able to identify two additional ubiquitous APC mutations that had not been called in the WGS or WES data.

CNA calling. We used CloneHD to call absolute copy-number and LOH profiles for each sample. First, we collected raw-read-depth data across the genomes using SAMTools. For WGS samples, we used 1-kilobase (kb) non-overlapping windows (excluding difficult regions such as centromeric regions). For WES samples, bed files corresponding to the exome-capture kit were used and the data were further placed onto 20-kb bins. For each set, we used the corresponding normal sample to identify outlier bins, which we removed. Furthermore, variant call files for each sample set were used to identify germline heterogeneous loci to collect B-allele data informative for LOH and/or unbalanced aberrations in the tumour samples.

We then ran filterHD (part of the CloneHD tool) to identify changes in the read-depth and B-allele tracks that go beyond the noise resulting from the finite sequencing depth for each set. Such ‘jump’ locations were used as input for cloneHD copy-number calling for the tumour samples. For the WES samples, we also used the corresponding normal samples to correct for platform-related bias in the read-depth tracks. The bias correction is important for WES samples, and not using it would result in a large number of loci with high jump probabilities; that is, over-segmentation (for further details see the discussion in ref. 21).

We validated the general patterns of large CNAs to calls from single nucleotide polymorphism (SNP) arrays (Illumina Global Screening Array) using the OncoSNP programme on the majority of tumours. Overall, 93% of the larger CNAs were in congruence.

For this analysis, we did not report structural variants or chromosomal hypermutation events such as chromothripsis and chromoplexy.

Exclusion of tumours from analysis of heterogeneity. The ability to accurately measure ITH depends on the number of biopsies available from the tumour. We assessed the relationship between biopsy number and genetic divergence and found no correlation (Supplementary Fig. 1b). A sub-sampling bootstrap analysis of three neoplasms each with a high number of biopsies showed that at least four biopsies per neoplasm provided a suitably accurate measurement of ITH (Supplementary Fig. 1c). Hence, neoplasms with fewer than four available biopsies were excluded from ITH analysis.

Analysis on sub-clonal populations. Since larger sub-clones might disturb the overall phylogenetic analysis, we investigated the existence of such in the samples using WGS with the Battenberg algorithm (https://github.com/cancerit/
Briefly, the algorithm phases heterozygous SNPs with use of the 1000 Genomes genotypes as a reference panel. The resulting haplotypes are corrected for occasional errors in phasing in regions with low linkage disequilibrium. Among the randomization of the results to estimate the B-allele frequency (BAF) values, t-tests are performed on the BAFs of each copy-number segment to identify whether they correspond to the value resulting from a fully clonal copy-number change. If not, the copy-number segment is represented as a mixture of two different copy-number states, with the fraction of cells bearing each copy-number state estimated from the average BAF of the heterozygous SNPs in that segment. Clusters of sub-clonal substitutions were identified in the WGS data using a Bayesian Dirichlet process in a n-dimensional space, where n is the number of related samples, as previously described. For each mutation, the allele frequency was converted to a CCF before clustering, allowing for purity and copy-number estimates obtained from the Battenberg algorithm, as described previously. Clusters were identified as local peaks in the posterior mutation density obtained from the Bayesian Dirichlet process. For each cluster, a region representing a ‘basin of attraction’ was defined by a set of planes running through the point of minimum density between each pair of cluster positions. Mutations were assigned to the cluster in whose basin of attraction they were most likely to fall, using posterior probabilities from the Bayesian Dirichlet process.

We investigated the geographical spread of sub-clones by analysing the presence of each sub-clone across samples from each patient. Sub-clones may be: clonal (defined as CCF ≥ 0.9) in all samples; clonal in some samples and absent (defined as CCF < 0.1) from others; present sub-clonally in a single sample; or present sub-clonally and sub-clonal categories represent clones/sub-clones that are in defined geographical regions of various sizes, whereas the last category represents sub-clones that are more diffuse, indicative of sub-clone mixing across regions covering multiple biopsies. The percentage of mutations in this last category was low (median 12%; range 5–26%), indicating that where sub-clones were present they were generally in defined geographical regions.

Phylogenetic analysis. We built phylogenetic trees from the SNA sets for each tumour using PAUP* software (http://phylosolutions.com/paup-test/). First, we converted each variant set into a binary matrix, where the rows related to a particular biopsy or the normal sample and the columns related to a specific variant. The binary encoding (0/1) designated the absence or presence of a variant. A nexus file was used to specify the parsimony parameters needed for the tree construction along with the variant matrix. The following functions and parameters were used: (1) the outgroup function was used to root all resulting trees to the normal sample (effectively, a column on the mutation matrix containing only zeros); (2) the heuristic function was used to perform a heuristic search of 10,000,000 trees from the given tree space, with 1,000 of the shortest trees output for the main analysis; (3) the bootstrap function was used to perform a subsampling procedure 10,000 times that involved randomly selecting a set of mutations from the binary matrix (with replacement), with the proportion of each branch instance reported in a log file; and (4) the all trees function was used in the cases where fewer than 10 biopsies were present. This made it possible to perform an extended ‘brute-force’ run to acquire the deepest shortest tree(s) from the total search space, at the expense of computational time. The resulting .trefiles were visualized and converted to .pdf format using FigTree software (http://tree.bio.ed.ac.uk/software/figtree/). The homoplasy index for the most parsimonious tree in a given set was automatically calculated and output to the PAUP* log file.

To obtain the shortest and thus most parsimonious tree, an Rscript using the ape Treeshape package was used to calculate the number of taxa for a given set, thus enabling a probability to be determined for obtaining a given Penny and Hendy index.

To perform tree balance analysis (only appropriate for carcinoma 6, which possessed >15 samples), we used the Colless' test function as implemented in the ape package. We tested the topology of the apeTree package (a user specified) against a balanced 'Yule' tree; hence, the P-value represents the likelihood of imbalance given this branching process.

Diversity analysis: adenoma versus cancer. To compare the SNA diversity of adenomas and carcinomas, we performed two main analyses. First, we wanted to compare the proportion of trunal SNAs from the total SNAs called in adenomas and carcinomas, but since certain tumours had many more biopsies, we performed a biopsy-wise down-sampling procedure. Here, in tumours with more than four biopsies (which was considered a reasonable minimum), we randomly selected various four-biopsy combinations and calculated the proportion of variants classified as trunk, branch and leaf. The distributions of these proportions were then plotted with the quartiles for the four-biopsy samples.

Second, the down-sampled call sets were used to assess the SNA diversity by measuring the SNA divergence of a pair of randomly selected biopsies from a given tumour. The diverging (non-shared) somatic variants were given as a number per Mb of the genome across the 100 down-sampled iterations.

Mutational signatures with EMu. For each tumour set, we classified all SNAs based on their flanking trinucleotide sequence context and whether they fell onto the X-axis (transcribed) or branched (not transcribed) of the phylogenetic tree. Inference of mutational processes and their activities was then determined using the EMu algorithm. Through this analysis, we identified four signatures (A, B, C and D). The resemblance to the COSMIC signatures was determined by comparing 96-channel mutation frequencies, and by visual inspection.

Comparison of spatial and genetic distances. To obtain estimates of the physical distances between each biopsy from any tumour, we produced normalized measurements of the number of pixels on a straight line drawn between the centre points of the biopsy locations. The photographs of the tumour and biopsy locations were used as a reference. We measured the physical distance between each biopsy and the centre of an overlapping 1000 Genomes genotype as a reference panel. The resulting haplotypes were determined using a bootstrapping analysis of the median values (1,000 repeats of 20 samples (10 adenoma and 10 carcinoma) with replacement). Data in contingency tables were analysed using Fisher's exact test. Researchers were not blinded to sample type.

Power calculation for detecting a difference in mutation burden. We calculated the power to detect a difference in the mutation burden of nine adenomas versus ten carcinomas. Comparing the measured burdens, we assessed 90 mutations per adenoma (s.d.: 55 mutations) and considered the power of a t-test to detect a difference between the carcinomas with a variety of higher burdens, requiring a (one-tailed) significance level of P<0.05. Standard formulae were used to calculate power. Our data had very good power to detect a 50% increase in burden in carcinomas and good/fair power to detect a 33% increase (Supplementary Fig. 10).

CNA timing model. Full details of the model are presented as a Supplementary Note.

We modelled the last common ancestor of WGS tumour samples by performing an Rscript using the apeTree package (a user specified) against a balanced 'Yule' tree; hence, the P-value represents the likelihood of imbalance given this branching process.

Diversity analysis: adenoma versus cancer. To compare the SNA diversity of adenomas and carcinomas, we performed two main analyses. First, we wanted to compare the proportion of trunal SNAs from the total SNAs called in adenomas and carcinomas, but since certain tumours had many more biopsies, we performed a biopsy-wise down-sampling procedure. Here, in tumours with more than four biopsies (which was considered a reasonable minimum), we randomly selected various four-biopsy combinations and calculated the proportion of variants classified as trunk, branch and leaf. The distributions of these proportions were then plotted with the quartiles for the four-biopsy samples.
The premise of the timing model is to use the information from all CNAs under consideration in a joint likelihood maximization of their respective timing. The individual timings are estimated using the assumption that the SNAs in a CNA region will accumulate following a Poisson process, based itself on the assumption of a given mutation rate. The SNAs in a given CNA region were partitioned into those on the non-amplified and amplified DNA strands (referred to as alpha and beta variants, respectively, in the Supplementary Note) using a Gaussian mixture model, implemented using the R package ‘mixtools’ and the ‘normalmixEM’ function. Here, the parameter mu (the starting cluster means) was inferred from the cellularity and copy state of the CNA. For example, in a biopsy with 100% tumour content, the expected cluster means of a tri-region are 0.33 and 0.67, representing the non-amplified and amplified chromosomes. We performed a heuristic search (using 10,000 iterations of the normalmixEM function) to assess the confidence in the inferred cluster centres. Only cases where clustering was deemed successful, as determined by the final means of the two clusters being located within one s.d. of the expected cluster means, were passed forward for timing analysis. Note that ‘complex CNAs’—those CNAs where there was reason to suspect that multiple alterations occurred at the same locus (for example, focal gain after an arm gain)—were excluded from the analysis.

The statistical assessment of the distribution of the CNA timings was performed by comparing the distribution of timings for each tumour against a uniform distribution using the Kolmogorov–Smirnov test. The duration of the comparative test was set to the interval between the inferred initiation of the adenoma (the timing of the 5q LOH event) and the time since the MRCA. In the cases where no 5q event was detected, the time of initiation of adenoma growth was taken as the average time of 5q LOH in the two cases where this event could be timed (adenoma 4 and carcinoma 5).

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

**Code availability.** MATLAB code to calculate the CNA timing is available from the GitHub project page: https://github.com/daniel-temko/CNVTiming.

**Data availability**
Raw data are available via the European Genome-Phenome Archive (https://ega-archive.org/) accession code: EGAS00001003066.

Received: 17 April 2018; Accepted: 12 July 2018; Published online: 31 August 2018

**References**
1. Morson, B. C. Evolution of cancer of the colon and rectum. *Cancer* **34**, 845–849 (1974).
2. Ashton-Rickardt, P. G. et al. High frequency of APC loss in sporadic colorectal carcinoma due to breaks clustered in 5q21–22. * Oncogene **4**, 1169–1174 (1994).
3. Powell, S. M. et al. APC mutations occur early during colorectal tumorigenesis. *Nature* **359**, 235–237 (1992).
4. Fearon, E. R. & Vogelstein, B. A genetic model for colorectal tumorigenesis. *Cell **61**, 759–767 (1990).
5. Jones, S. et al. Comparative lesion sequencing provides insights into tumor evolution. *Proc. Natl Acad. Sci. USA **105**, 4283–4288 (2008).
6. Smith, G. et al. Mutations in APC, Kirsten-ras, and p53—alternative genetic pathways to colorectal cancer. *Proc. Natl Acad. Sci. USA **99**, 9433–9438 (2002).
7. Muzny, D. M. et al. Comprehensive molecular characterization of human colon and rectal cancer. *Nature **487**, 330–337 (2012).
8. Sottoriva, A. et al. A Big Bang model of human colorectal tumor growth. *Nat. Genet. **47**, 209–216 (2015).
9. Williams, M. J., Werner, B., Barnes, C. P., Graham, T. A. & Sottoriva, A. Identification of neutral tumor evolution across cancer types. *Nat. Genet. **48**, 238–244 (2016).
10. Wright, S. The roles of mutation, inbreeding, crossingbreed and selection in evolution. In *Proc. 6th International Congress of Genetics Vol. 1* 356–366 (1932).
11. Yap, T. A., Gerlinger, M., Futreal, P. A., Pusztai, L. & Swanton, C. Intrapitumor heterogeneity: seeing the wood for the trees. *Sci. Transl. Med. **4**, 127ps10 (2012).
12. Blum, M. G. B. & François, O. On statistical tests of phylogenetic tree imbalance. *Math. Biosci. **195**, 141–153 (2005).
13. Fischer, A., Illingworth, C. J., Campbell, P. J. & Mustonen, V. EMu: probabilistic inference of mutational processes and their localization in the cancer genome. *Genome Biol. **14**, B39 (2013).
14. Alexandrov, L. B. et al. Signatures of mutational processes in human cancer. *Nature **500**, 415–421 (2013).
15. Katainen, R. et al. CTCF/cohesin-binding sites are frequently mutated in cancer. *Nat. Genet. **47**, 818–821 (2015).
16. Quirke, P. et al. DNA aneuploidy in colorectal adenomas. *Br. J. Cancer **53**, 477–481 (1986).
17. Jones, A. M. et al. Analysis of copy number changes suggests chromosomal instability in a minority of large colorectal adenomas. *J. Pathol. **213**, 249–256 (2007).
18. Wang, H., Liang, L., Fang, J.-Y. & Xu, J. Somatic gene copy number alterations in colorectal cancer: new quest for cancer drivers and biomarkers. *Oncogene **35**, 2011–2019 (2016).
19. Durinck, S. et al. Temporal dissection of tumorigenesis in primary cancers. *Cancer Discov. **1**, 137–143 (2011).
20. Newman, S. et al. The relative timing of mutations in a breast cancer genome. *PLoS ONE **8**, e64991 (2013).
21. Toyota, M. et al. CpG island methylphyton phenotype in colorectal cancer. *Proc. Natl Acad. Sci. USA **96**, 8681–8686 (1999).
22. Kim, T.-M. et al. Subclonal genomic architectures of primary and metastatic colorectal cancer based on intratumoral genetic heterogeneity. *Clim. Cancer Res. **21**, 4461–4472 (2015).
23. Uchi, R. et al. Integrated multiregional analysis proposing a new model of colorectal cancer evolution. *PLoS Genet. **12**, e1005778 (2016).
24. Vuroski, X. et al. Multiregion ultra-deep sequencing reveals early intermixing and variable levels of intratumoral heterogeneity in colorectal cancer. *Mol. Oncol. **11**, 124–139 (2017).
25. Kim, T.-M. et al. Clonal origins and parallel evolution of regionally synchronous colorectal adenoma and carcinoma. *Oncotarget **6**, 27725–27735 (2015).
26. Stachler, M. D. et al. Paired exome analysis of Barrett’s esophagus and adenoacarcinoma. *Nat. Genet. **47**, 1047–1055 (2015).
27. Maley, C. C. et al. Genetic clonal diversity predicts progression to esophageal adenocarcinoma. *Nat. Genet. **38**, 468–473 (2006).
28. Ross-Innes, C. S. et al. Whole-genome sequencing provides new insights into the clonal architecture of Barrett’s esophagus and esophageal adenocarcinoma. *Nat. Genet. **47**, 1038–1046 (2015).
29. Andrews, S. FastQC: a quality control tool for high throughput sequence data (Babraham Bioinformatics, 2013); http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
30. McKenna, A. et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res. **20**, 1297–1303 (2010).
31. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at http://arxiv.org/abs/1303.3997 (2013).
32. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics **25**, 2078–2079 (2009).
33. Rimmer, A. et al. Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. *Nat. Methods **46**, 912–918 (2014).
34. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res. **38**, e160 (2014).
35. Cingolani, P. et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin) **6**, 80–92 (2012).
36. Hasan, M. S., Wu, X. & Zhang, I. Performance evaluation of indel calling tools using real short-read data. *Hum. Genom. **9**, 20 (2015).
37. Narzisi, G. et al. Accurate de novo and transmitted indel detection in exome-capture data using microarray. *Nat. Methods **11**, 1033–1036 (2014).
38. Thörvaldsdóttir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genome data visualization and exploration. *Brief Bioinform. **14**, 178–192 (2013).
39. Fischer, A., Vázquez-García, I., Illingworth, C. J., R. J. & Mustonen, V. High-definition reconstruction of clonal composition in cancer. *Cell Rep. **7**, 1740–1752 (2014).
40. Werner, B., Traulsen, A., Sottoriva, A. & Dingli, D. Detecting truly clonal evolution of subclonal gene copy number alterations in colorectal cancer. *PLoS ONE **9**, 8681–8686 (2014).
exzellenter Nachwuchsforschende’). A-M.B. also acknowledges funding from Cancer Research UK (A14895). D.C.W. is supported by the Li Ka Shing Foundation. X.J. and I.P.M.T. are supported by an ERC advanced grant (EVOCAN-340560). The S:CORT study is funded by the MRC and Cancer Research UK. K.H is supported by Krebsliga Zentralschweiz. A.S. is supported by the Wellcome Trust (202778/B/16/Z), Cancer Research UK (A22909) and the Chris Roos Fellowship in Evolution and Cancer. This work was also supported a Wellcome Trust award to the Centre for Evolution and Cancer (105104/Z/14/Z). J.E.E. was funded by the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre (BRC). V.M. was funded by the Swiss National Science Foundation (P2SKP3_168322 / 1 and P2SKP3_168322 / 2). D.T. acknowledges funding from the EPSRC (grant no.: EP/F500351/1).

**Author contributions**

I.P.M.T., T.A.G. and S.J.L conceived and designed the study. R.G., J.E.E., L.M.W., K.H., S.J.L and I.P.M.T. provided the samples. H.D., A.-M.B., S.B. and L.C. performed the experiments. W.C., M. Kovac, V.M., R.A. and D.C.W. performed the bioinformatics analysis. W.C. and D.T. performed the mathematical analysis. C.G., A.R.A. and V.H.K. performed the image analysis. M.J., M.R.-J. and L.M.W. performed the pathology assessment. E.D., T.M. and the S:CORT consortium provided reference data. W.C., M. Kovac, V.M., D.T., R.A., V.H.K., X.J., D.C.W., Y.F., M. Kovacova, S.A., A.S., S.J.L., T.A.G. and I.P.M.T. analysed the data. W.C., A.S., S.J.L., T.A.G. and I.P.M.T. performed the evolutionary analysis. W.C., T.A.G. and I.P.M.T. wrote the manuscript with input from all authors.

**Competing interests**

The authors declare no competing interests.

**Additional information**

Supplementary information is available for this paper at [https://doi.org/10.1038/s41559-018-0642-z](https://doi.org/10.1038/s41559-018-0642-z).

Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints).

Correspondence and requests for materials should be addressed to T.A.G. or I.P.M.T.

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

---

**The S:CORT Consortium**

Richard Adams25, Simon Bach17, Andrew Beggs17, Louise Brown26, Francesca Buffa14, Jean-Baptiste Cazier7, Enric Domingo14, Andrew Blake14, Che-Hsi Wu27, Ekaterina Chatzpili28, Susan Richman29, Philip Dunne30, Paul Harkin30, Geoff Higgins14, Jim Hill31, Chris Holmes12,27, Denis Horgan32, Rick Kaplan26, Richard Kennedy30, Mark Lawler30, Simon Leedham6,19, Tim Maughan14, Ultan McDermott28, Gillies McKenna14, Gary Middleton33, Dion Morton17, Graeme Murray34, Phil Quirke29, Manuel Salto-Tellez30, Les Samuel34, Anna Schuh14, David Sebag-Montefiore29, Matt Seymour29, Ricky Sharma14, Richard Sullivan35, Ian Tomlinson2,11, Nicholas West29 and Richard Wilson30

---

25Centre for Trials Research, Cardiff University, Cardiff, UK. 26Clinical Trials Unit, University College London, London, UK. 27Department of Statistics, University of Oxford, Oxford, UK. 28Wellcome Trust Sanger Institute, Hinxton, UK. 29Institute of Cancer and Pathology, Faculty of Medicine and Health, University of Leeds, Leeds, UK. 30School of Medicine, Queens University Belfast, Belfast, UK. 31Department of General Surgery, Manchester Royal Infirmary, Manchester, UK. 32European Alliance for Personalised Medicine, Brussels, Belgium. 33Institute of Immunology and Immunotherapy, University of Birmingham, Birmingham, UK. 34Institute of Medical Sciences, University of Aberdeen, Aberdeen, UK. 35Institute of Cancer Policy, Kings College London, London, UK.
### 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 [Authors & Referees](https://www.nature.com/nature/authors) and the [Editorial Policy Checklist](https://www.nature.com/nature/authors/editors-policy-checklist).

#### Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

- [ ] **n/a**
- [x] Confirmed

- [ ] The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- [ ] An indication of 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 statistics 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)
- [x] 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.
- [x] For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- [x] For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- [x] Estimates of effect sizes (e.g. Cohen’s d, Pearson’s r), indicating how they were calculated
- [ ] Clearly defined error bars
  - State explicitly what error bars represent (e.g. SD, SE, CI).

Our web collection on [statistics for biologists](https://www.nature.com/nature/research-statistics) may be useful.

#### Software and code

**Policy information about availability of computer code**

- **Data collection**: Samples were acquired from fresh frozen pathological specimens. Patients were anonymised.
- **Data analysis**: All core analyses performed using published open-source software (see methods). Custom software used in the manuscript has been made publically available (via GitHub).

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/reviewers upon request. 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](https://www.nature.com/nature/authors/data-access). 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

All raw data (bam files, and processed data including vcf files) are deposited in the EGA archive under accession number EGAS00001003066. Data are available subject to written application and approval from the data access committee. All figures make use of these raw data.
Field-specific reporting

Please select 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/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

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

| Sample size | No power calculations were performed prior to commencement of the study. Power was assessed post hoc (see Methods). In total 24 patient derived tumours were analysed |
| Data exclusions | No specific sample inclusion/exclusion criteria were specified (see methods). |
| Replication | A subset of mutation calls were replicated using an alternative methodology (reported in the Results). |
| Randomization | Randomisation was not applicable. |
| Blinding | No blinding was used. |

Reporting for specific materials, systems and methods

Materials & experimental systems

| n/a | Involved in the study |
| ☑ Unique biological materials |
| ☑ Antibodies |
| ☑ Eukaryotic cell lines |
| ☑ Palaeontology |
| ☑ Animals and other organisms |
| ☑ Human research participants |

Methods

| n/a | Involved in the study |
| ☑ ChiP-seq |
| ☑ Flow cytometry |
| ☑ MRI-based neuroimaging |

Human research participants

Policy information about studies involving human research participants

Population characteristics

Patients were selected on their basis of having colorectal cancer or adenoma, and age >18. No other selection criteria were applied.

Recruitment

Patients were consented to the study (for tissue donation) prior to their clinical procedure. The consenting process was performed by a member of the clinical team based at the hospital.