

Cytogenetic signatures favoring metastatic organotropism in colorectal cancer

Received: 28 December 2022

Accepted: 21 March 2025

Published online: 05 April 2025

 Check for updates


Mariola Monika Golas^{1,2} , Bastian Gunawan^{3,9}, Angelika Gutenberg⁴, Bernhard C. Danner⁵, Jan S. Gerdes^{10,11}, Christine Stadelmann⁶, Laszlo Füzesi³, Torsten Liersch⁷ & Bjoern Sander⁸ 

Colorectal carcinoma (CRC) exhibits metastatic organotropism, primarily targeting liver, lung, and rarely the brain. Here, we study chromosomal imbalances (CIs) in cohorts of primary CRCs and metastases. Brain metastases show the highest burden of CIs, including aneuploidies and focal CIs, with enrichment of +12p encoding *KRAS*. Compared to liver and lung metastases, brain metastases present with increased co-occurrence of *KRAS* mutation and amplification. CRCs with concurrent *KRAS* mutation and amplification display significant metabolic reprogramming with upregulation of glycolysis, alongside upregulation of cell cycle pathways, including copy number gains of *MDM2* and *CDK4*. Evolutionary modeling suggests early acquisition of many organotropic CIs enriched in both liver and brain metastases, while brain-enriched CIs preferentially emerge later. Collectively, this study supports a model where cytogenetic events in CRCs favor site-specific metastatic colonization. These site-enriched CI patterns may serve as biomarkers for metastatic potential in precision oncology.

Colorectal cancer (CRC) is responsible for about 10% of cancer-related deaths worldwide, which corresponds to ~900,000 deaths annually¹. The vast majority of cancer-associated deaths are related to metastatic spread of the primary malignancy². About one in four to five patients with CRC presents with distant metastasis at initial cancer staging³, and about one in five patients diagnosed with CRC develops metachronous metastatic disease in the clinical course⁴. Metastatic CRC spread can occur very early in the progression of disease, at a time-point when the primary tumor is not yet necessarily clinically manifest^{5–7}. In general, metastatic seeding can follow a polyclonal pattern, in which metastatic cells spread from multiple tumor clones, but it can also follow a monoclonal pattern^{7–9}. Previous work indicates that in about two-thirds of CRC cases, lymph node and distant metastases appear to be

derived from independent subclones¹⁰. In the course of CRC disease, the main site for metastasis is the liver, followed by the lung¹¹. Less commonly, peritoneal spread is seen in CRC, and rare sites of metastasis include other gastrointestinal organs, bones, and the brain/central nervous system (CNS)¹¹. Thus, the spread of CRC cells follows a specific organ pattern, a phenomenon referred to as metastatic organotropism^{12,13}.

Accumulating evidence supports a model in which successful colonization of cancer cells at a secondary site does not result merely from passive mechanical constraints, but rather involves active shaping of the microenvironment at the secondary site to create a permissive niche that facilitates cancer cell survival and growth at distant sites^{14–18}. Cancer cells appear to act on the prospective secondary site

¹Human Genetics, Faculty of Medicine, University of Augsburg, Augsburg, Germany. ²Comprehensive Cancer Center Augsburg, University Medical Center Augsburg, Augsburg, Germany. ³Institute of Pathology, University Medical Center Göttingen, Göttingen, Germany. ⁴Department of Neurosurgery, Asklepios Hospital Harburg, Hamburg, Germany. ⁵Department of Cardiac, Thoracic and Vascular Surgery, University Medical Center Göttingen, Göttingen, Germany. ⁶Department of Neuropathology, University Medical Center Göttingen, Göttingen, Germany. ⁷Department of General, Visceral and Pediatric Surgery, University Medical Center Göttingen, Göttingen, Germany. ⁸Institute of Pathology, Hannover Medical School, Hannover, Germany. ⁹Present address: Institute of Pathology Northern Hesse, Kassel, Germany. ¹⁰Present address: Epilepsy Center Hamburg, Evangelical Hospital Alsterdorf, Neurology and Epileptology, Hamburg, Germany.  e-mail: monika.golas@med.uni-augsburg.de; sander.bjoern.pat@mh-hannover.de

to induce a microenvironment that favors invasion and colonization of the secondary organ^{14–18}. The formation of this pre-metastatic niche is hypothesized to be induced by factors and extracellular vesicles, including exosomes released from the primary cancer cells^{15,19,20}. In addition, evidence is accumulating that hematopoietic stem cell-derived cells shape the metastatic niche. For example, cancer cells appear to use tenascin C-activated macrophages to induce formation of a vascular niche in the target organ¹⁸. However, the genetic processes involved in metastatic organotropism remain poorly understood.

Modeling the tumor evolution may identify genetic determinants that confer metastatic organotropism. The process of tumor evolution is driven by the acquisition of a diverse array of genetic alterations, ranging from single nucleotide variants and small insertions/deletions (indels), to focal chromosomal aberrations and structural variants such as translocations and inversions, as well as chromosomal arm aneuploidies and whole chromosomal aneuploidies^{21,22}. In particular, primary CRCs commonly carry a set of defined DNA copy number variations (CNVs), including gains of 8q, 13q, and 20q, as well as losses of 18q^{23–28}. These CNVs often result from whole chromosomal arm and whole chromosome aneuploidies^{24,29}. In addition to CNVs, mutations in a number of genes have been described: collaborative next-generation sequencing (NGS) efforts²³ have led to the description of a temporal sequence of driver mutation acquisition in the adenomatous polyposis coli gene (*APC*), the tumor protein gene P53 (*TP53*), the Kirsten rat sarcoma virus oncogene (*KRAS*), and the SMAD family mothers against decapentaplegic homolog 4 gene (*SMAD4*), among others³⁰. This sequence is consistent with previous studies^{31–34} that culminated in the proposal of a sequential progression model, which has become known as the Vogelgram³⁵. However, the specific sequence of genetic events may not be as critical for CRC development as the overall accumulation of genetic alterations over time³⁵.

Following the Vogelgram model, a number of additional models of tumor evolution in CRC have been proposed. These evolutionary models integrate the acquisition of genetic changes in the form of gradual processes or bursts, combined with different extents of clonal selection³⁶. In the linear model, of which the Vogelgram is a prototype, genetic events are acquired sequentially. In this model, genetic changes with a growth advantage lead to preferential survival of the fittest clone(s)³⁵. The branched evolution³⁷ and the neutral model³⁸ share the idea of a parallel, i.e., branched, expansion of clones, but differ in the extent of selective pressure. While the former model assumes some degree of selection³⁷, the latter model favors unrestricted co-expansion of clones in the absence of relevant selection pressure³⁸. In the fourth model, the punctuated model³⁹, also known as the big bang model⁴⁰, massive genetic changes occur in bursts during cancer development, followed by the expansion of clones with selective advantage³⁹ or under neutral evolutionary conditions^{40,41}.

Cytogenetic aberrations in tumor cells, such as those arising from chromoplexy of several chromosomes³⁹ and chromothripsis of single chromosomes⁴² as well as chromosomal CNVs⁴³ have been linked to the punctuated model. In other tumors, it has been suggested that the acquisition of CNVs followed a more gradual evolutionary path⁴⁴. Thus, while we have a detailed idea of possible acquisition patterns of genetic alterations, an understanding of the chronology of metastatic capacity acquisition, and of the genetic alterations driving metastatic organotropism, is only beginning to emerge^{28,45}.

Primary CRC and its metastases share a majority of somatic mutations with few exclusive mutations private to the primary tumor or metastasis^{46,47}. In a study analyzing primary CRCs and their metastases, the vast majority of which colonized the liver, CRC metastasis-enriched mutations were identified in genes linked to the extracellular matrix, PI3K signaling, and stellate cell activation⁴⁷. Also, involvement of TGF- β signaling has been suggested in liver metastasis⁴⁵, as has RAS signaling with *KRAS* mutations in lung and brain metastases^{48,49}.

However, whether these mutations represent organotropic genetic alterations or a general expression of the metastatic potential remained largely unknown. Recent studies have shown no significant differences between the tumor mutational burden (TMB) in primary CRC and its metastases²⁸. Along these lines, *KRAS* mutations were the only organotropic mutations in lung and brain metastases of CRC that have been repeatedly reported^{28,45}, yet the complexity of metastatic behavior, especially in less-studied sites like the brain, remains to be explored in detail.

At the same time, previous CRC studies on CIs have focused on genomic changes in primary tumors and single metastatic sites, which facilitated the definition of common chromosomal aberrations in primary CRC and specific metastases^{23–28,50–63}. CRC is an example of a cancer entity characterized by a complex pattern of chromosomal imbalances (CIs)^{23–28,50–63}, and it has been suggested that chromosomal aberrations are acquired early in disease development⁹. Since CIs typically alter the expression of a (large) group of genes simultaneously⁶⁴, CIs can potentially have additive or even synergistic effects and be further potentiated by co-occurrence of multiple CIs.

In this work, we examine the cytogenetic profile of CRCs and their metastatic spread to the liver, lung, and brain by combining independent cohorts. The first cohort, termed CRCTropism, employs a dedicated cytogenetic approach, while the second cohort, the Memorial Sloan Kettering (MSK) MetTropism cohort²⁸, is based on NGS data. In the CRCTropism cohort, the vast majority of patients, including all cases with brain metastases, underwent tumor sampling prior to receiving any targeted therapy. This approach enables the identification of genetic signatures that support intrinsic organotropy rather than reflecting potential consequences of targeted therapeutic interventions. We demonstrate that metastases are characterized by site-enriched patterns of cytogenetic events, which translate into cytogenetic signatures for secondary sites. Using a third cohort from The Cancer Genome Atlas (TCGA)^{65,66}, we show that concurrent *KRAS* mutation and amplification, enriched in brain metastases, is associated with enhanced metabolic reprogramming and typically coupled with *MDM2* and *CDK4* co-amplifications already in primary tumors. Organotropic patterns reveal the enrichment of specific CNVs in liver and brain metastases, their rarity in lung metastases, and a temporal dimension defined by the order in which these alterations are acquired during tumor evolution.

Results

Characterization of the CRCTropism cohort

The CRCTropism cohort comprised a total of 314 tumors derived from 191 patients diagnosed with metastatic CRC. Demographic and clinicopathological data of this cohort are summarized in Supplementary Table 1, and an overview of the analyses is given in Supplementary Fig. 1. The cohort consists of 80 primary CRCs as well as 117, 78, and 39 metastases to the liver, lung, and brain, respectively. In some cases, multiple tumor tissue samples were analyzed, including samples from different areas of the primary tumor, recurrences, and/or one or more metastases. For 80 patients, at least one primary CRC/metastasis (P/M) pair was available, totaling 97 P/M pairs. The comparative analysis of the primary CRCs reported in this study and TCGA cohort^{65,66}, which exclusively comprises primary CRCs, revealed no statistically significant differences, except for the subgroup of primary CRCs associated with lung metastases. This subgroup exhibited a significantly lower prevalence of +7q (0.426 in the TCGA cohort vs. 0.143 in primary CRC with lung metastases [log2 ratio -1.57]; $p_{\text{adjust}} = 0.013$; Fisher test with Benjamini-Hochberg correction of multiple testing) and -18p (0.568 vs. 0.194 [log2 ratio -1.55]; $p_{\text{adjust}} = 4.17 \times 10^{-3}$) and +20p (0.402 vs. 0.074 [log2 ratio -2.44]; $p_{\text{adjust}} = 0.013$), concomitant with a higher frequency of -20p (0.068 vs. 0.296, [log2 ratio 2.12]; $p_{\text{adjust}} = 0.013$; Supplementary Fig. 2).

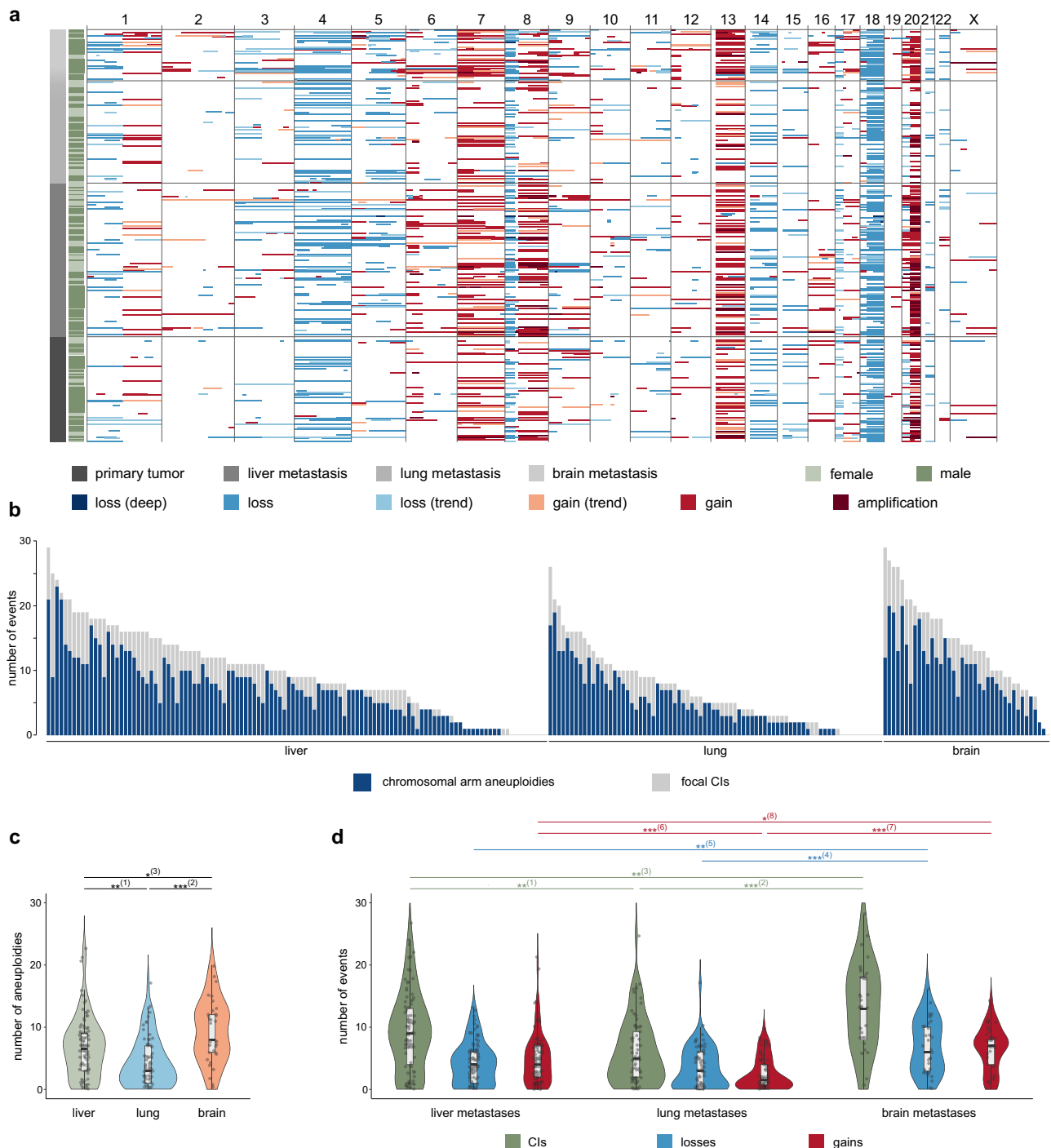


Fig. 1 | Pattern of CIs observed in the CRCTropism cohort. a Chromosomal losses and gains in shades of blue and red, respectively, observed in the cohort with 314 tumors, including $n = 80$ primary CRC and $n = 234$ distant CRC metastases in the liver, lung, and brain. On the left, tumor site and sex are color-encoded according to the legend. **b** Number of apparent chromosomal arm aneuploidies in relation to focal CIs, visualized as a stacked bar chart ($n = 234$). **c** Distribution of chromosomal arm aneuploidies in CRC metastases ($n = 191$). (1) $p_{\text{adjust}} = 2.96 \times 10^{-3}$; (2) $p_{\text{adjust}} = 6.94 \times 10^{-6}$; (3) $p_{\text{adjust}} = 1.31 \times 10^{-2}$. **d** Distribution of CIs including losses and gains for CRC metastases ($n = 191$). (1) $p_{\text{adjust}} = 2.46 \times 10^{-3}$; (2) $p_{\text{adjust}} = 1.37 \times 10^{-6}$;

(3) $p_{\text{adjust}} = 5.57 \times 10^{-3}$; (4) $p_{\text{adjust}} = 3.70 \times 10^{-4}$; (5) $p_{\text{adjust}} = 5.96 \times 10^{-3}$; (6) $p_{\text{adjust}} = 1.44 \times 10^{-4}$; (7) $p_{\text{adjust}} = 4.97 \times 10^{-7}$; (8) $p_{\text{adjust}} = 1.75 \times 10^{-2}$. **c, d** Data are presented as combined violin and box plots (thick line in the box corresponds to median; lower and upper edges of the box indicating the first (Q1) and third (Q3) quartiles; whiskers extend to 1.5 times the IQR). Individual data points are presented using the jitter method. Kruskal-Wallis tests with Dunn posthoc tests (two-sided) and Benjamini-Hochberg corrections for multiple comparisons were used to test for statistical significance. *, $p_{\text{adjust}} < 0.05$; **, $p_{\text{adjust}} < 0.01$, and ***, $p_{\text{adjust}} < 0.001$. Source data are provided as a Source Data file.

Overall, the CIs observed in the CRCTropism cohort resulted in a characteristic pattern of gains and losses, enriched in whole chromosomal arm aneuploidies and certain focal CNVs (Fig. 1a and Supplementary Fig. 3). In line with previous results⁶⁷, the CRCTropism cohort showed the typical CI pattern associated with CRC, including the loss

of the short arm of chromosome 8 accompanied by a gain of the long arm of this chromosome, with typical breakpoints located in chromosomal bands 8p11 or 8p12 (Supplementary Fig. 4).

Focusing on primary CRCs, the most common DNA copy number aberrations were +13q, -18q, and +20q (Fig. 1a). A similar pattern of CIs

was seen in liver, lung, and brain metastases (Fig. 1a). However, notable differences emerged in the frequencies of specific alterations at secondary sites. For example, brain metastases exhibited a higher proportion of gains at 7p (liver 0.433 [log2 ratio -0.48], lung 0.265 [log2 ratio -1.19] vs. brain 0.606; $p_{\text{adjust}} = 0.025$; Fisher exact test with Benjamini-Hochberg correction), and 12p (liver 0.067 [log2 ratio -1.85], lung 0.015 [log2 ratio -4.01] vs. brain 0.242; $p_{\text{adjust}} = 7.80 \times 10^{-3}$). Moreover, lung metastases showed a significantly lower frequency of +20p (liver 0.267 [log2 ratio -0.18], lung 0.044 [log2 ratio -2.78] vs. brain 0.303; $p_{\text{adjust}} = 4.22 \times 10^{-3}$), among other aberrations.

Chromosomal arm aneuploidies as major source of DNA copy number aberrations

To address the CI profile in detail, a series of analyses was conducted. First, we asked about the type of CI (i.e., chromosomal arm aneuploidies vs. focal subchromosomal arm CNVs). Overall, apparent chromosomal arm aneuploidies contributed as the major source of CIs, accounting for about three-quarters of all CIs (Fig. 1b). Brain metastases exhibited the highest number of apparent chromosomal arm aneuploidies, while lung metastases showed the lowest (Fig. 1c). Specifically, the median number of aneuploidies was 6.5 (interquartile range [IQR] 6) in liver metastases, 3 (IQR 6) in lung metastases, and 8 (IQR 6) in brain metastases. Statistical analysis using the Kruskal-Wallis test revealed significant differences ($p_{\text{adjust}} = 1.35 \times 10^{-18}$ with Benjamini-Hochberg correction; Supplementary Table 2). Subsequent pairwise comparisons with Dunn test and Benjamini-Hochberg correction indicated adjusted p -values of 2.96×10^{-3} for liver vs. lung metastasis, 1.31×10^{-2} for liver vs. brain metastasis, and 6.94×10^{-6} for lung vs. brain metastasis.

Profile of whole chromosomal arm and subchromosomal arm copy number aberrations

We next explored the overall pattern of CIs by integrating data from both whole chromosomal arms and subchromosomal regions. Lung metastases had the fewest CIs, with a median count of 5 (IQR 7.25; Fig. 1d), while brain metastases exhibited the highest count, with a median of 13 CIs (IQR 10). Liver metastases fell in between, showing a median of 9 CIs (IQR 9). The Kruskal-Wallis test demonstrated significant differences among these groups ($p_{\text{adjust}} = 6.10 \times 10^{-23}$; Supplementary Table 2). Subsequent pairwise comparisons using Dunn test with Benjamini-Hochberg correction indicated that brain metastases had a significantly greater number of CIs compared to both liver metastases ($p_{\text{adjust}} = 5.57 \times 10^{-3}$) and lung metastases ($p_{\text{adjust}} = 1.37 \times 10^{-6}$). These significant differences in total CI counts were consistent with the observed chromosomal arm aneuploidies.

This pattern was also evident when analyzing both losses and gains separately. The Kruskal-Wallis tests revealed significant findings for both categories ($p_{\text{adjust}} = 5.86 \times 10^{-8}$ for losses and $p_{\text{adjust}} = 1.86 \times 10^{-5}$ for gains). In particular, liver metastases had a median of 4 losses (IQR 5), lung metastases had a median of 3 losses (IQR 6), while brain metastases had the highest median of 6 losses (IQR 7). For gains, liver metastases showed a median of 4 gains (IQR 5), lung metastases had 1.5 gains in median (IQR 3.25), whereas brain metastases demonstrated a higher median of 7 gains (IQR 4). With the exception of losses in liver vs. lung metastases, all pairwise comparisons yielded statistically significant results in Dunn tests with Benjamini-Hochberg correction (Supplementary Table 2). Thus, we conclude that CRC metastases exhibit distinct variations in the counts of individual CI types.

Organotropic mapping identifies site-specific CI patterns in CRC metastases

Our analysis of DNA copy number alterations in the cohort suggested distinct patterns of CIs in metastatic CRC dependent on the site of metastasis. To test whether certain CIs are enriched at specific metastatic sites, we used a ternary plot, which highlights significant

differences in the proportion of lesions harboring particular imbalances (Fig. 2a). We refer to this visualization as organotropic map. This map projects the frequency of observed CIs into a ternary diagram, with statistical significance assessed by Fisher exact test with Benjamini-Hochberg correction for multiple comparisons. In this representation, the size of each circle corresponds to the frequency of the respective CI within the cohort. The corners of the organotropic map denote the secondary sites. Organotropic CIs are positioned in proximity to one of the corners (i.e., metastasis organs), if they are predominantly associated with a single metastatic site; those enriched at two sites are placed in the middle between two corners. Conversely, genetic changes uniformly associated with all three sites are situated at the center of the organotropic map.

A total of sixteen statistically significant CNVs were identified (Fig. 2a and Supplementary Fig. 5). Notably, none of these organotropic CIs were associated with lung metastasis. Nine CIs were enriched in both liver and brain metastasis, specifically +7p, -8p, +8q, +13q, +16p, +16q, -18p, +20p, and +20q (Fig. 2a,b and Supplementary Fig. 5). These CIs largely correspond to some of the most frequently observed chromosomal aberrations in CRC. In addition, seven organotropic CIs were specifically enriched in brain metastasis (Fig. 2a), with four CIs identified in more than 15% of brain metastases (Supplementary Fig. 5). These CIs include -3q, +5q, -6q, and +12p. Overall, we conclude that CRC exhibits a range of organotropic CIs that are preferentially observed in specific metastatic organs. This organotropic signature of CRC metastases is visually summarized in Fig. 2c, highlighting the chromosomal imbalances across autosomes.

Gene-Level Mapping of CNVs in CRC Metastases

To address CNVs in an independent cohort, we took advantage of the MSK MetTropism cohort²⁸, which comprises a total of 3,548 CRC samples including 2,401 primary tumors, 624 liver metastases, 146 lung metastases, and 22 brain/CNS metastases (hereinafter referred to as brain metastases). The demographics of this cohort are given in Supplementary Table 3, and an overview of the analyses performed is provided in Supplementary Fig. 1. The overall pattern of chromosomal arm level gains and losses in the MSK MetTropism cohort was consistent with that observed in the CRCTropism cohort, with +13q, -18, and +20q being the most common aneuploidies (Fig. 3a). However, as observed for the CRCTropism cohort, brain metastases showed a higher number of CIs. Amongst others, -3p, -3q, +5q, +7p, +8q, +12p, and +13q were enriched in brain metastases. In line with these observations, the fraction genome altered (FGA), a measure of the total proportion of the genome affected by CNVs, was significantly higher in brain metastases compared to liver and lung metastases (Kruskal-Wallis test, $p = 8.23 \times 10^{-9}$, followed by Dunn test with Benjamini-Hochberg correction: brain vs. liver, $p_{\text{adjust}} = 4.66 \times 10^{-9}$; brain vs. lung, $p_{\text{adjust}} = 5.50 \times 10^{-9}$). In contrast, there were no statically significant differences in TMB between the three metastatic sites ($p = 0.113$, Kruskal-Wallis test; Fig. 3b).

To further characterize these genomic alterations, we performed a gene-level analysis of the MSK MetTropism cohort. This analysis revealed a distinct pattern of organotropic CNVs, predominantly enriched in brain metastases – consistent with the higher FGA observed in this group and corroborating the findings from our CRCTropism cohort (Fig. 3c). These enriched genes spanned various functional categories, including the following (note that genes often have multiple cellular functions; for simplicity, each gene has been assigned to only one category):

- (I) signaling: e.g., *PDGFRB* (chromosomal band 5q32), *FGFR4* (5q35.2), *FLT4* (5q35.3), *BRAF* (7q34), *RET* (10q11.21), *KRAS* (12p12.1), *ERBB3* (12q13.2), *FLT3* (13q12.2), *FLT1* (13q12.3), and *MAP2K1* (15q22.31);
- (II) chromatin organization/transcription: e.g., *ARID1A* (1p36.11), *SETD2* (3p21.31), *PBRM1* (3p21.1), *TERT* (5p15.33), *CDK8* (13q12.13), and *CTCF* (16q22.1);

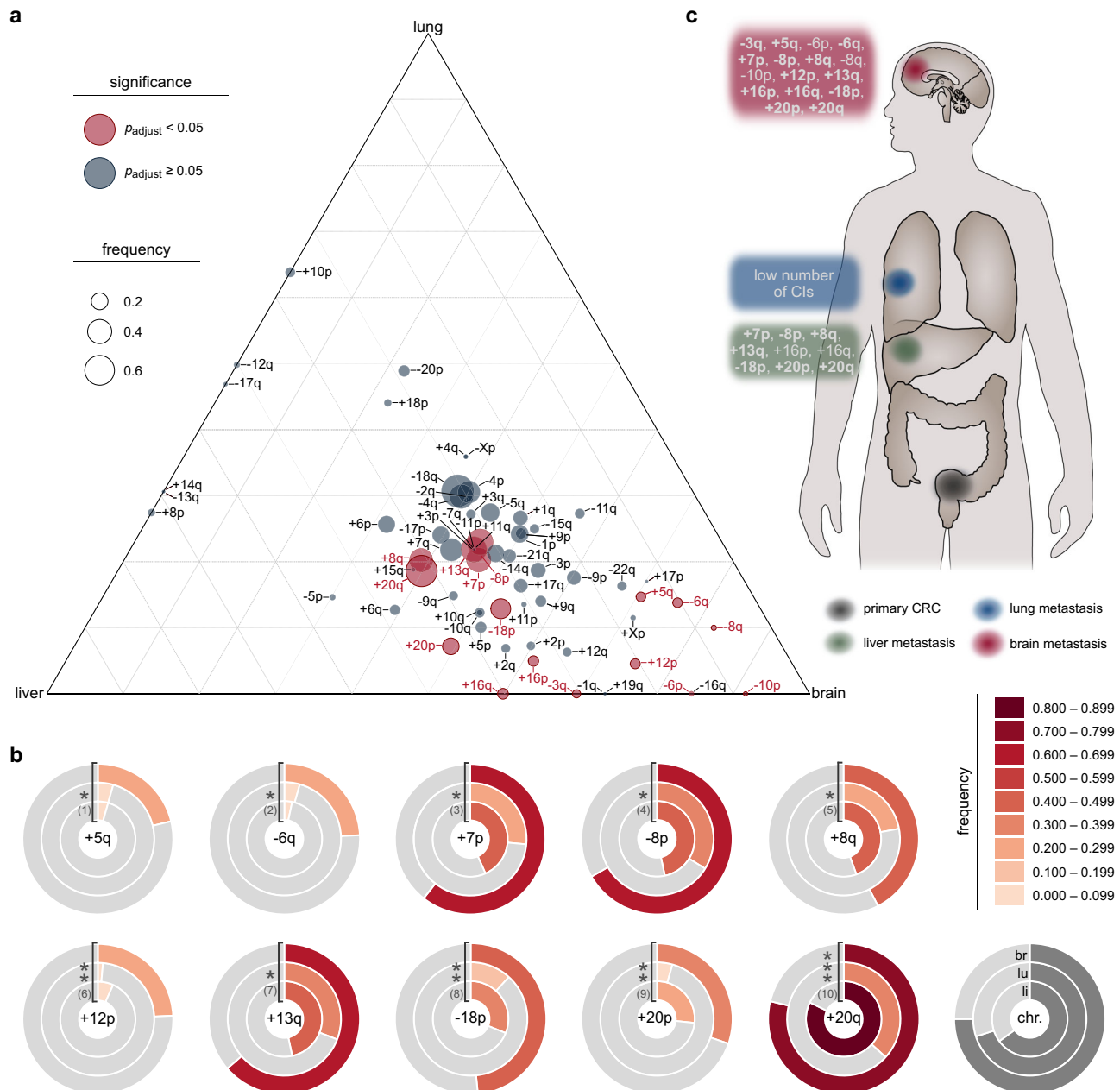


Fig. 2 | Site-specific pattern of DNA copy number aberrations in CRC metastases based on the CRCTropism cohort. **a** Organotrophic map presenting CIs with site-specific enrichment ($n = 191$). All p -values were obtained by two-sided Fisher exact tests followed by Benjamini-Hochberg adjustment for multiple comparison (legend). The size of the bubbles encodes the frequency a given CI was observed in the cohort (legend). **b** Doughnut plots showing the fraction of metastases ($n = 191$) that harbor selected organotrophic CIs (see legend for the assignment of secondary organs: li, liver; lu, lung; br, brain; chr., chromosomal arm; see also Supplementary Fig. 5 for a full representation of all organotrophic CIs). The frequency of occurrence

is color-encoded according to the legend shown on the right. Two-sided Fisher exact tests with Benjamini-Hochberg adjustment for multiple comparison were used (*, $p_{\text{adjust}} < 0.05$; **, $p_{\text{adjust}} < 0.01$, and ***, $p_{\text{adjust}} < 0.001$). (1) $p_{\text{adjust}} = 3.80 \times 10^{-2}$; (2) $p_{\text{adjust}} = 1.14 \times 10^{-2}$; (3) $p_{\text{adjust}} = 2.48 \times 10^{-2}$; (4) $p_{\text{adjust}} = 3.80 \times 10^{-2}$; (5) $p_{\text{adjust}} = 4.10 \times 10^{-2}$; (6) $p_{\text{adjust}} = 7.80 \times 10^{-3}$; (7) $p_{\text{adjust}} = 3.80 \times 10^{-2}$; (8) $p_{\text{adjust}} = 4.85 \times 10^{-3}$; (9) $p_{\text{adjust}} = 4.22 \times 10^{-3}$; (10) $p_{\text{adjust}} = 3.08 \times 10^{-7}$. **c** Graphical summary of autosomal organotrophic CIs with arbitrary intra-organ sites of lesions. CIs observed in more than 15% of the respective metastases are highlighted in bold. Source data are provided as a Source Data file.

- (III) DNA damage response and repair: e.g., *MSH6* (2p16.3), *BARD1* (2q35), *MLH1* (3p21.2), *BAP1* (3p21.1), *ATR* (3q23), *MSH3* (5q14.1), *MRE11A* (11q21), and *ATM* (11q22.3), *RADS1B* (14q24.1), *BRCA1* (17q21.31), *RADS1C* (17q22), *CHEK2* (22q12.1);
- (IV) cell cycle: e.g., *CDK4* (12q14.1), *MDM2* (12q15); and
- (V) microRNA biology: e.g., *DROSHA* (5p13.3).

Some organotrophic genes, such as *AGO2* (8q24.3), were enriched in both liver and brain metastases. Gains in a single gene, the

transcription factor *JUN* (1p32.1), were uniquely enriched in both lung and brain metastases.

Co-occurrence pattern of organotrophic *KRAS* amplifications and mutations

One of the organotrophic genes identified in our analysis is *KRAS*, located on 12p12.1, with amplifications significantly enriched in brain metastases. Organotrophic mapping of our CRCTropism cohort demonstrated a significant enrichment of +12p in brain metastases

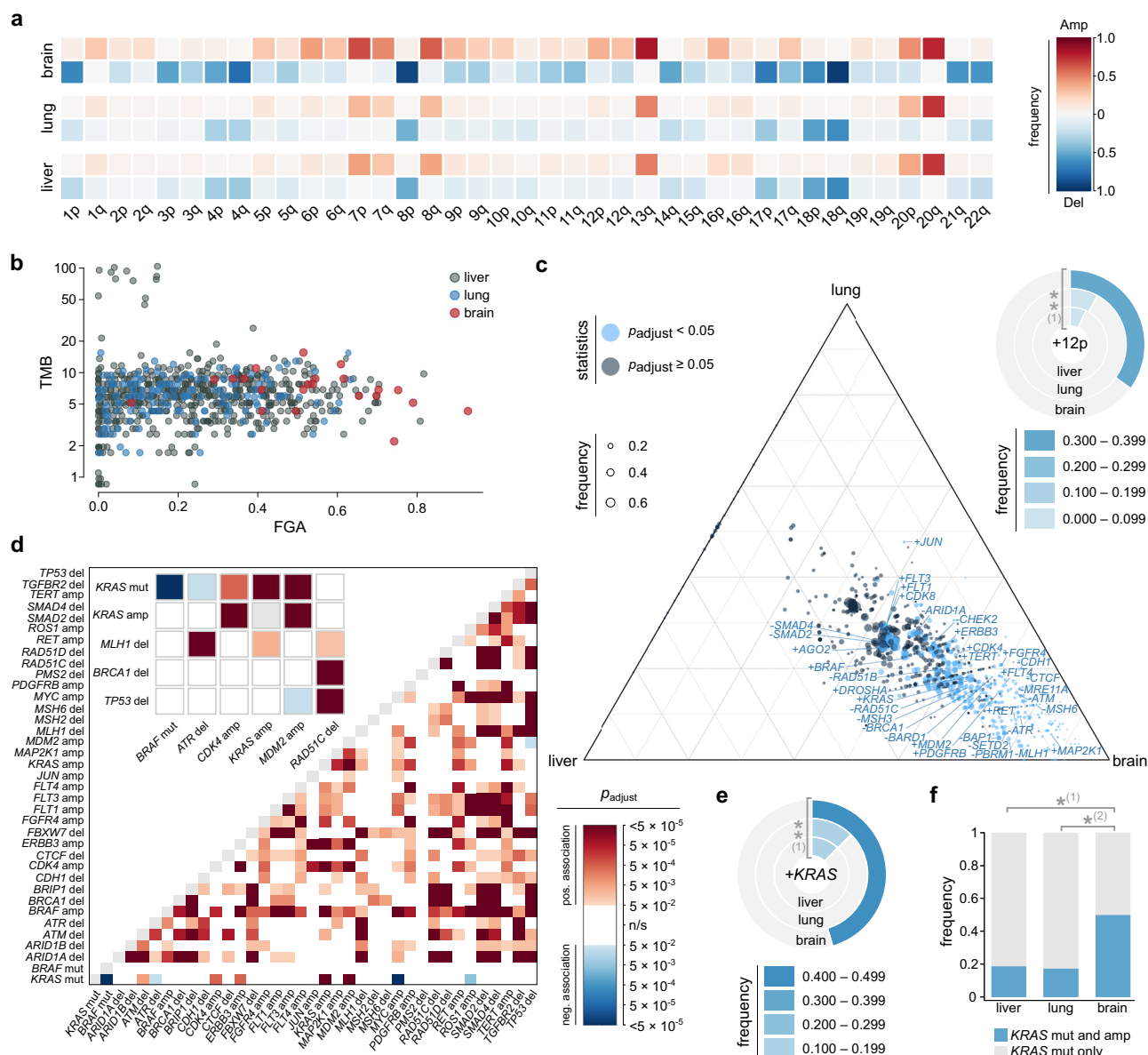


Fig. 3 | Organotropic genetic signatures of metastatic CRC in the MetTropism cohort. **a** Pattern of chromosomal arm aneuploidies in liver, lung, and brain metastases determined using ASCETS, with $n = 792$ metastases subjected to analysis. For each metastatic site, the frequency of the chromosomal arm aneuploidy (bottom, deletion; top, amplification) is color-coded as shown in the color scale at the right. **b** Profile of tumor mutational burden (TMB) and fraction genome altered (FGA) for each liver, lung, and brain metastasis (FGA: $p = 8.23 \times 10^{-9}$ [Kruskal-Wallis test], $p_{\text{adjust}} = 4.66 \times 10^{-9}$ for brain metastases vs. liver and $p_{\text{adjust}} = 5.50 \times 10^{-9}$ for brain metastases vs. lung metastases [Dunn posthoc test with Benjamini-Hochberg correction]; TMB: $p = 0.113$ [Kruskal-Wallis test]; $n = 792$). **c** Organotropic map of gene-level gains and losses derived from GISTIC 2.0 ($n = 792$). Significant CIs are indicated in blue, with the event frequency represented by the bubble size (legend). The doughnut plot in the top right shows the fraction of metastases with +12p as determined using ASCETS ($n = 758$), where the occurrence frequency is color-

encoded according to the legend on the right. (1) $p_{\text{adjust}} = 5.19 \times 10^{-3}$. **d** Co-occurrence probability mapping of genetic changes observed in CRC metastases ($n = 792$), with red shades indicating significant positive associations and blue shades depicting significant negative associations, based on co-occurrence modeling (two-sided), followed by Benjamini-Hochberg adjustment. The inset presents close-ups of selected genetic changes (amp, amplification; del, deletion; mut, mutation). **e** Fraction of metastases ($n = 792$) harboring *KRAS* amplifications determined using GISTIC 2.0 (color-coded according to the legend). (1) $p_{\text{adjust}} = 3.49 \times 10^{-3}$. **f** Bar plot depicting the frequency of *KRAS* amplifications among metastases ($n = 352$) with oncogenic *KRAS* mutations. (1) $p_{\text{adjust}} = 0.019$; (2) $p_{\text{adjust}} = 0.019$. **c, e, f** Adjusted p -values were obtained by two-sided Fisher exact test with Benjamini-Hochberg adjustment for multiple comparisons. *, $p_{\text{adjust}} < 0.05$; **, $p_{\text{adjust}} < 0.01$. Source data are provided as a Source Data file.

(Fig. 2a). This finding was independently corroborated in the MSK MetTropism cohort, which also showed a significant enrichment of +12p chromosomal arm aneuploidy in brain metastases (7%, 8%, and 35% for liver, lung and brain metastases, respectively; $p_{\text{adjust}} = 5.19 \times 10^{-3}$, Fisher exact test followed by Benjamini-Hochberg correction; Fig. 3c, right).

Given that oncogenic *KRAS* mutations are known drivers in CRC, we next asked whether *KRAS* amplifications are associated with

oncogenic *KRAS* mutations. To this end, we analyzed the co-occurrence of event pairs in the lesions and took advantage of a probabilistic model that accounts for the frequencies of the respective events. Based on this model, positive and negative co-occurrences, defined as two events observed together with a significantly higher or lower frequency, respectively, than assuming a random association, were visualized in a color-encoded triangular heatmap of Benjamini-Hochberg adjusted p -values. A significant positive association between

KRAS mutations and *KRAS* amplifications was observed in the co-occurrence analysis ($p_{\text{adjust}} < 5 \times 10^{-5}$; Fig. 3d, inset). Notably, *KRAS* gains were identified in nearly half of the brain metastases, but only in about 12% in liver and lung metastases ($p_{\text{adjust}} = 3.49 \times 10^{-3}$, Fisher exact test followed by Benjamini-Hochberg correction; Fig. 3e), in line with the analysis of +12p chromosomal arm aneuploidy. Further analysis revealed that the proportion of CRC metastases with simultaneous *KRAS* mutation and amplification was significantly higher in brain metastases compared to liver and lung metastases (liver, 19%; lung, 17%; brain, 50%; $p_{\text{adjust}} = 0.019$, each; Fisher exact test with Benjamini-Hochberg correction; Fig. 3f).

Oncogenic *KRAS* mutations were found to co-occur with several other CIs, such as *CDK4* amplifications and *MDM2* amplifications (Fig. 3d, inset). As anticipated, a significant negative association was observed between *KRAS* mutations and *BRAF* mutations, and *ATR* deletions (Fig. 3d, inset). Similarly, a negative association was noted between *TP53* deletions and *MDM2* amplifications. These findings collectively indicate a significant co-occurrence pattern, suggesting a unique genomic landscape that may contribute to CRC metastasis.

KRAS CNVs translate into increased *KRAS* pathway activity

We hypothesized that the co-occurrence of *KRAS* mutations and amplifications may facilitate metastatic spread through the rewiring of the cancer cell pathways, reflected by changes in gene expression patterns. To test this hypothesis, we analyzed the TCGA cohort, which provides genomic, transcriptomic and proteomic data^{65,66,68}. Patient characteristics are given in Supplementary Table 4. We identified 64 CRC cases exhibiting simultaneous *KRAS* mutation and amplification (Fig. 4a). Transcriptome analysis revealed significantly higher *KRAS* expression levels in CRCs with both *KRAS* mutation and amplification compared to those with *KRAS* mutation alone ($p_{\text{adjust}} = 1.32 \times 10^{-3}$, Wilcoxon rank sum test with Benjamini-Hochberg correction; Fig. 4b), confirming that the *KRAS* amplification is linked to increased *KRAS* expression.

To test whether the significant increase in *KRAS* expression in *KRAS*-mutated and amplified CRCs is also reflected in pathway activity, we analyzed the levels of activated (phosphorylated) downstream effector proteins mitogen-activated protein kinase kinase 1 and 2 (MAP2K1/2, also known as MEK1 and MEK2) in CRCs harboring oncogenic *KRAS* mutations. This analysis utilized reverse-phase protein array (RPPA) data on 160 CRCs, which provide information on both total and phosphorylated protein levels for selected cancer-associated proteins⁶⁸. Oncogenic *KRAS* signals through RAF, resulting in the phosphorylation of MAP2K1/2 at two serine residues within the activation loop⁶⁹. We hypothesized that, in addition to *KRAS*, other factors, including the protein levels of downstream target proteins, might influence the amount of phosphorylated MAP2K1/2. Consequently, we employed multiple linear regression to model MAP2K1/2 phosphorylation.

*KRAS*_CNV level correlated with *KRAS* mRNA expression, necessitating the exclusion of the latter from the model. The remaining variables showed no relevant multicollinearity, with variance inflation factors (VIF) of 1.008 for *KRAS*_CNV, 1.009 for pBRAF_S445 (a measure of constitutively phosphorylated BRAF), and 1.016 for pMAP2K1 (total protein). Data on total MAP2K2 were not available. Our multiple linear regression analysis confirmed that the *KRAS*_CNV is a significant predictor of phosphorylated MAP2K1/2 ($p = 0.041$, likelihood ratio test). In CRC with oncogenic *KRAS* mutations, *KRAS*_CNV positively contributed to predicting levels of phosphorylated MAP2K1/2, alongside pBRAF_S445 and pMAP2K1, all of which demonstrated similar importance scores (Fig. 4c). These findings suggest that the copy number of *KRAS* translates into increased *KRAS* pathway activity at the functional protein level.

Rewiring of CRC cells with simultaneous *KRAS* mutation and amplification

Building on our findings that concurrent *KRAS* mutations and amplifications result in increased pathway activity, we sought to further

investigate the functional consequences of these genetic alterations. Specifically, we hypothesized that the concurrent *KRAS* mutation and amplification may lead to a more accentuated reprogramming of cellular pathways, ultimately influencing the aggressive behavior of CRC. To explore these functional implications, we performed gene set enrichment analysis (GSEA) utilizing oncogenic signature definitions. We were particularly interested in the M2900 gene set, which represents genes upregulated in cancer cells overexpressing an oncogenic *KRAS* form, as it closely reflects our cohort. Comparing CRC samples with both *KRAS* mutation and amplification to those with *KRAS* mutation alone, we observed significant enrichment of this gene set in the former (normalized enrichment score, $\text{NES}_{\text{GSEA}} = 1.95$, $p < 1 \times 10^{-4}$, $p_{\text{adjust}} = 0.040$; Fig. 4d). This finding suggests that *KRAS* mutation and amplification together contribute to a more pronounced *KRAS*-driven oncogenic profile, potentially enhancing the aggressive phenotype of these tumors.

Further analysis of the transcriptome identified 200 significantly upregulated genes and 156 significantly downregulated genes in CRCs with concurrent *KRAS* mutation and amplification. The volcano plot of expressed genes showed a unique peak-like concentration at a fold change of about 1.25 in the CRC group with both *KRAS* mutation and amplification (Fig. 4e). Notably, most upregulated genes shared a chromosomal location on chromosome 12 (Fig. 4e, inset), suggesting that other loci on this chromosome may be co-amplified alongside *KRAS* due to larger CNVs.

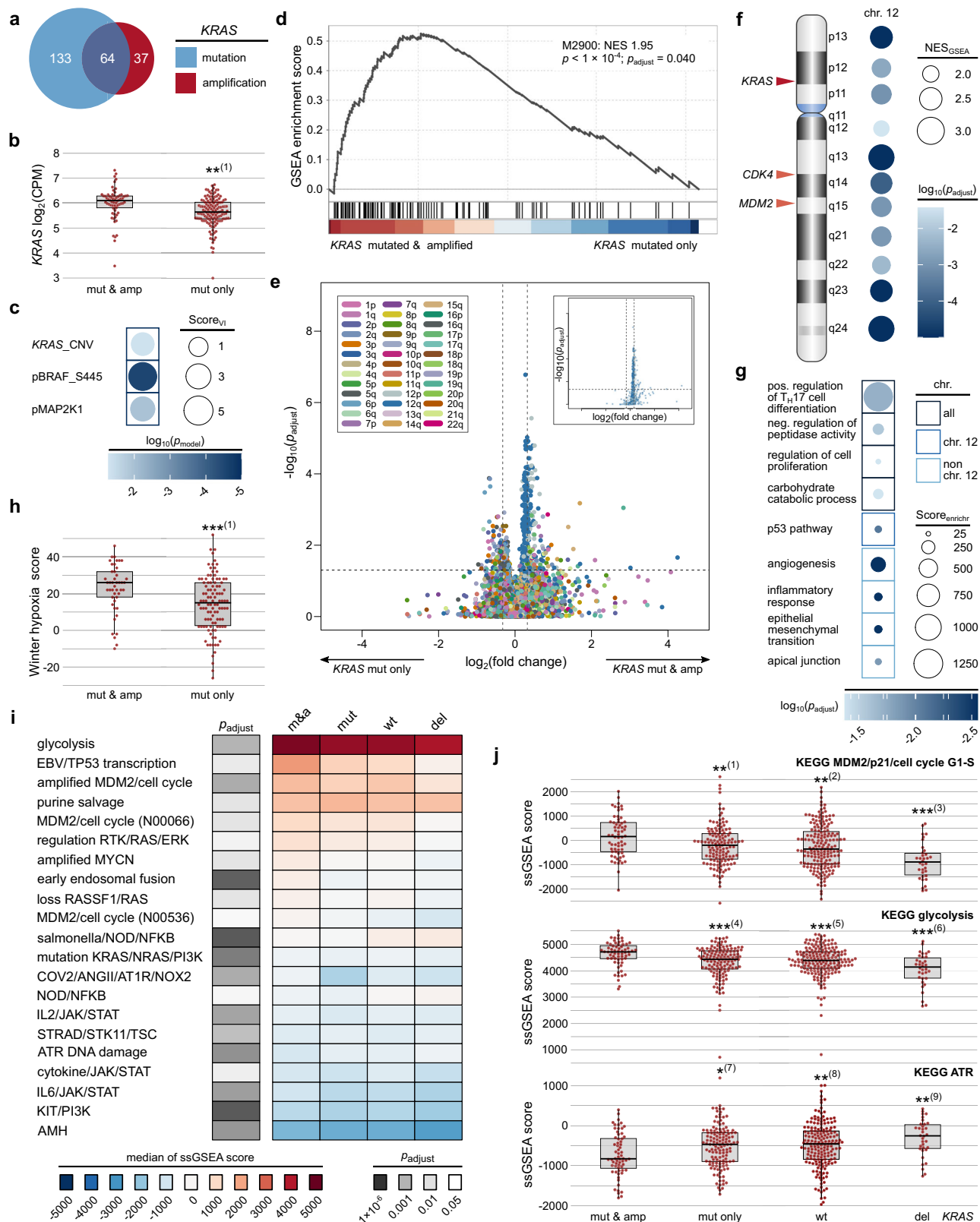
To validate our findings, we performed GSEA using C1 gene set definitions from MSigDB, which correspond to chromosomal bands. This analysis confirmed that genes enriched in CRCs with *KRAS* mutation and amplification are predominantly located on both arms of chromosome 12 (Fig. 4f; NES_{GSEA} ranging from 2.80 to 1.95; $p_{\text{adjust}} < 9 \times 10^{-5}$ to 0.038), with no other chromosomal bands showing significance. Enriched genes included key oncogenes such as *CDK4* and *MDM2* (both encoded on 12q), further supporting the notion that the peak-like asymmetry in the volcano plot is due to co-amplification of multiple loci on chromosome 12 along with *KRAS*.

To further characterize the functional landscape of the differentially expressed genes, we first tested the upregulated genes, independent of their chromosomal location, for enrichment using Enrichr with two complementary gene set definitions, Gene Ontology (GO) and MSigDB Hallmark. This analysis identified several significant biological processes, including 'Positive Regulation Of T-helper 17 Cell Differentiation' (GO:2000321; $p_{\text{adjust}} = 0.019$), 'Negative Regulation Of Peptidase Activity' (GO:0010466; $p_{\text{adjust}} = 0.027$), 'Regulation Of Cell Population Proliferation' (GO:0042127; $p_{\text{adjust}} = 0.042$), and 'Carbohydrate Catabolic Process' (GO:0016052; $p_{\text{adjust}} = 0.042$; Fig. 4g).

We further analyzed the upregulated genes separately according to their chromosomal location, performing separate Enrichr analyses for genes on chromosome 12 and those outside of it using MSigDB Hallmark definitions. The enriched gene set for differentially expressed genes on chromosome 12 was the 'p53 Pathway' that includes *MDM2* ($p_{\text{adjust}} = 8.18 \times 10^{-3}$), while those outside chromosome 12 included 'Angiogenesis' ($p_{\text{adjust}} = 2.87 \times 10^{-3}$), 'Inflammatory Response' ($p_{\text{adjust}} = 2.87 \times 10^{-3}$), 'Epithelial Mesenchymal Transition' (EMT; $p_{\text{adjust}} = 2.87 \times 10^{-3}$) and 'Apical Junction' ($p_{\text{adjust}} = 0.014$; all p -values adjusted the Benjamini-Hochberg method for multiple testing). These biological functions are enriched in processes associated with hypoxia^{70–72}.

To investigate this association, we examined whether CRCs with *KRAS* mutation and amplification demonstrate enhanced hypoxia adaptation by analyzing the Winter hypoxia score as a measure of hypoxic response⁷³. Our results indicate that CRCs with concurrent *KRAS* mutation and amplification exhibited significantly higher hypoxia scores compared to those with *KRAS* mutation alone ($p = 8.75 \times 10^{-4}$, Wilcoxon rank sum test; Fig. 4h).

Given our observation of increased hypoxia adaptation in *KRAS* mutated and amplified CRC, we explored the potential implications of



these alterations on hypoxia-related pathways, complemented by additional medically relevant pathways as implemented in the Kyoto Encyclopedia of Genes and Genomes (KEGG) MEDICUS gene set definitions. We hypothesized that highly *KRAS*-responsive gene sets would show a systematic increase or decrease in enrichments across the following groups: (1) CRC with simultaneous *KRAS* mutation and amplification, (2) CRC with *KRAS* mutation only, (3) CRC with wild-type

KRAS, and (4) CRC with *KRAS* deletion (no *KRAS* mutation). To test this hypothesis, we performed single-sample gene set enrichment analysis (ssGSEA), which calculates an enrichment score for each sample-gene set pairing. In total, 94 gene sets were statistically significant using the Kruskal-Wallis test with Benjamini-Hochberg correction. Of these, 21 showed statistically significant differences in Dunn test with Benjamini-Hochberg correction when comparing CRC with *KRAS*

Fig. 4 | Rewiring of *KRAS* mutated and amplified CRCs as indicated by the TCGA cohort. **a** Venn diagram showing the relationship between CRCs with *KRAS* mutations (blue circle) and amplifications (red circles). **b** Box plot of normalized *KRAS* expression levels, with *p*-values from two-sided Wilcoxon rank-sum test and Benjamini-Hochberg adjustment for multiple comparison ($n = 64$ and $n = 133$ for CRC with *KRAS* mutation and amplification and CRC with *KRAS* mutation, respectively). (1) $p_{\text{adjust}} = 1.32 \times 10^{-3}$. **c** Multiple linear regression indicating *KRAS* CNV as a significant predictor of phosphorylated MAP2K1/2, along with pBRAF_S445, and pMAP2K1 with similar variable importance scores (Score_{vi}), with $n = 160$. **d** Two-sided GSEA using the M2900 gene set, derived from cells overexpressing oncogenic *KRAS*. The normalized enrichment score (NES), nominal *p*-value and false discovery rate (FDR)-adjusted *p*-value are shown ($n = 197$). **e** Volcano plot of differential gene expression, colored by chromosomal arm (see legend). The inset highlights genes on chromosome 12. Statistical significance was assessed using two-sided Wilcoxon rank-sum tests with Benjamini-Hochberg correction for multiple comparisons ($n = 197$). The horizontal dashed line indicates $p_{\text{adjust}} = 0.05$, while vertical dashed lines correspond to a 1.25-fold change. **f** Two-sided GSEA using C1 positional gene sets on chromosomal bands enriched in *KRAS* mutated and amplified CRCs vs. CRC with *KRAS* mutation ($n = 197$); circle size indicates NES and color the FDR-adjusted *p*-value. An ideogram of chromosome 12 is provided for orientation. **g** Enrichr analysis of upregulated genes in CRCs with *KRAS* mutation and amplification vs. CRC with *KRAS* mutation ($n = 197$),

using enriched genes on chromosome 12 and other chromosomes. Circle size reflects the Enrichr score and color indicates the Benjamini-Hochberg-adjusted *p*-value. The Gene Ontology and MSigDB Hallmark genes sets were tested. **h** Winter hypoxia score comparison using two-sided Wilcoxon rank-sum test ($n = 45$ and $n = 106$ for CRC with *KRAS* mutation and amplification and CRC with *KRAS* mutation, respectively). (1) $p_{\text{adjust}} = 8.75 \times 10^{-4}$. **i** Heatmap of median ssGSEA scores using KEGG_MEDICUS gene sets, with Benjamini-Hochberg adjusted *p*-values from Kruskal-Wallis tests encoded in grayscale ($n = 412$). m&a, *KRAS* mutated and amplified CRCs; mut, *KRAS* mutated CRCs; wt, *KRAS* wildtype CRCs; del, *KRAS* deletion CRCs (without *KRAS* mutation). **j** Box plots of ssGSEA scores for selected gene sets, analyzed by Kruskal-Wallis tests with Dunn posthoc test (two-sided) and Benjamini-Hochberg adjustment ($n = 64$ for *KRAS* mutated and amplified CRCs; $n = 133$ for *KRAS* mutated CRCs; $n = 181$ for *KRAS* wildtype CRCs; $n = 34$ for *KRAS* deletion CRCs (without *KRAS* mutation); (1) $p_{\text{adjust}} = 6.78 \times 10^{-3}$; (2) $p_{\text{adjust}} = 1.69 \times 10^{-3}$; (3) $p_{\text{adjust}} = 1.59 \times 10^{-7}$; (4) $p_{\text{adjust}} = 6.19 \times 10^{-4}$; (5) $p_{\text{adjust}} = 5.27 \times 10^{-4}$; (6) $p_{\text{adjust}} = 3.20 \times 10^{-6}$; (7) $p_{\text{adjust}} = 3.12 \times 10^{-2}$; (8) $p_{\text{adjust}} = 7.63 \times 10^{-3}$; (9) $p_{\text{adjust}} = 1.17 \times 10^{-3}$). Analyses compare CRCs with *KRAS* mutation and amplification to those with *KRAS* mutation alone (**b**, **d–h**) as well as to wildtype CRC and CRC with *KRAS* deletions without *KRAS* mutation (**i**, **j**). (**b**, **d**, **h**, **j**) Data are presented with individual data points and a box-and-whisker-plot (thick line, median; box edges, Q1/Q3; whiskers, $1.5 \times \text{IQR}$). *, $p_{\text{adjust}} < 0.05$; **, $p_{\text{adjust}} < 0.01$, and ***, $p_{\text{adjust}} < 0.001$. Source data are provided as a Source Data file.

mutation and amplification to CRC with *KRAS* mutation alone (Fig. 4i). The identified pathways and processes included, among others, critical signaling cascades, cell cycle regulation, transcription, DNA damage response, as well as glycolysis, a key metabolic response to hypoxia.

Notably, some pathways exhibited a significant gradual increase or decrease across the four groups, with *KRAS* mutated and amplified CRCs consistently showing the highest or lowest enrichment (Fig. 4j). These pathways included glycolysis (Kruskal-Wallis test $p_{\text{adjust}} = 2.33 \times 10^{-4}$; Dunn test $p_{\text{adjust}} = 6.19 \times 10^{-4}$ [CRC with *KRAS* mutation and amplification vs. CRC with *KRAS* mutation], 5.27×10^{-4} [CRC with *KRAS* mutation and amplification vs. wildtype *KRAS* CRC], and 3.20×10^{-6} [CRC with *KRAS* mutation and amplification vs. CRC with *KRAS* deletion]), MDM2/P21/cell cycle G1-S (Kruskal-Wallis test $p_{\text{adjust}} = 5.61 \times 10^{-5}$; Dunn test $p_{\text{adjust}} = 6.78 \times 10^{-3}$, 1.69×10^{-3} , and 1.59×10^{-7} , respectively), and ATR signaling (Kruskal-Wallis test $p_{\text{adjust}} = 1.32 \times 10^{-2}$; Dunn test $p_{\text{adjust}} = 3.12 \times 10^{-2}$, 7.63×10^{-3} , and 1.17×10^{-3} , respectively), among others. Collectively, these results indicate that CRCs with both *KRAS* mutation and amplification exhibit a distinct gene expression pattern characterized by enhanced glycolytic activity, disrupted cell cycle control, and impaired DNA damage response, among other factors, which may, at least in part, be facilitated by co-amplified loci on chromosome 12.

Upregulation of *HOX* genes

Having established the distinct molecular landscape associated with CRCs exhibiting concurrent *KRAS* mutation and amplification, we next explored other genetic alterations implicated in metastatic organotropism, specifically focusing on CRCs with deletions of the DNA repair genes *MLH1* and *BRCA1*, respectively, both of which exhibit enrichment in brain metastases (Fig. 3c). For *MLH1* deletion CRCs, we identified 18 significantly upregulated and 209 significantly downregulated genes in comparison to control CRC without microsatellite instability (MSI) phenotype (*MLH1* control CRC). The volcano plot revealed a distinct asymmetric pattern linked to downregulated genes on the 3p arm, where *MLH1* is located (Fig. 5a) – similar to the pattern observed in CRCs with *KRAS* mutation and amplification (Fig. 4e). GSEA confirmed significant enrichment of genes located in bands 3p14 to 3p26 that are downregulated in CRC with *MLH1* deletion (Fig. 5b; NES_{GSEA} ranging from -2.88 to -2.15 ; $p_{\text{adjust}} < 10^{-3}$ to 3.42×10^{-3}).

However, we also identified several genes that are upregulated in CRCs with *MLH1* deletion. A g:Profiler analysis of these differentially upregulated genes showed significant enrichment of pathways related to DNA-binding transcription factor activity, embryonic morphogenesis,

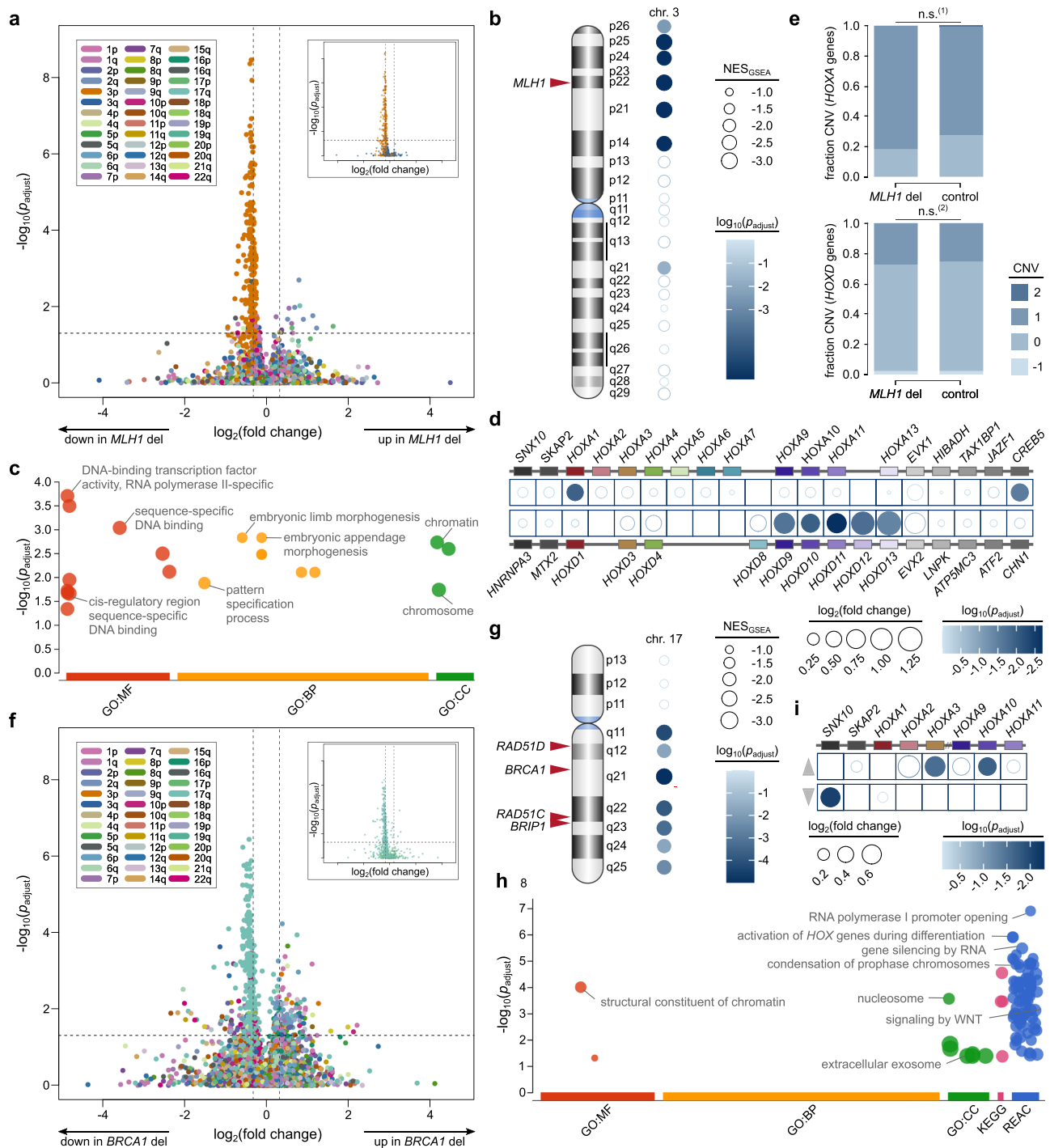
and chromatin dynamics (Fig. 5c). In alignment with our g:Profiler findings, differentially expressed genes from the *HOX* family, including *HOXA1* (located on 7p15.2) and *HOXD9* to *HOXD13* (located on 2q31.1), were significantly upregulated in *MLH1*-deleted CRCs, along with *CREB5* (Fig. 5d). Further analysis of the copy number at the *HOXA* and *HOXD* loci indicated CNVs are not the primary cause of the enhanced expression of these genes ($p_{\text{adjust}} = 0.757$ and 0.941 , respectively; Fisher exact test with Benjamini-Hochberg correction; Fig. 5e). This suggests that the observed upregulation of *HOX* gene expression in CRCs with *MLH1* deletion is likely attributable to transcriptional regulation rather than alterations in gene dosage.

In our analysis of CRCs with *BRCA1* deletion, we identified 152 significantly upregulated and 313 significantly downregulated genes compared to control CRCs. The volcano plot revealed a similar peak-like pattern associated with genes on chromosome 17q, where *BRCA1* is located (Fig. 5f). GSEA confirmed significant downregulation of genes encoded on 17q, including other homologous recombination repair (HRR) genes such as *RAD51C*, *RAD51D*, and *BRIPI* (Fig. 5g; NES_{GSEA} ranging from -2.63 to -2.04 ; $p_{\text{adjust}} < 5 \times 10^{-5}$ to 0.019). A g:Profiler analysis of the upregulated genes revealed significant enrichment in pathways involved in transcriptional regulation, gene silencing, chromosome condensation, WNT signaling, and activation of *HOX* genes during differentiation (Fig. 5h). Specifically, we observed significant upregulation of *HOXA3* and *HOXA10*, alongside a notable downregulation of *SNX10* (Fig. 5i). Overall, our findings highlight complex molecular alterations associated with CRCs harboring deletions of *MLH1* and *BRCA1*, respectively, and point towards an activation of *HOX* genes in these cancers.

Oncogenetic modeling identifies sets of early and late cytogenetic events in CRC

The definition of site-specific CIs (Figs. 2a, 3c) suggests that CRCs may share common patterns of cytogenetic evolution, despite the inherent individuality of each tumor's evolution. We therefore reconstructed the cytogenetic evolution of CRC, utilizing two approaches: (I) reconstruction of oncogenetic trees using maximum likelihood estimation, a cross-sectional approach to model common phylogenetic relationships within the entire cohort or its subsets (Fig. 6) and (II) phylogenetic tree reconstruction for individual patients through maximum parsimony using multi-sampling (Fig. 7 and Supplementary Fig. 6).

Oncogenetic modeling was applied to the three metastatic subgroups studied here. Events were classified into nine categories,



ranging from class 1 (initial events) to class 9 (late events), based on their distance from the root of the oncogenetic tree. All metastatic sites shared the loss of 18q as an initial event (Fig. 6). Likewise, losses of 8p and gains of 13q and 20q were among the earliest events in cytogenetic tumor evolution of CRC, regardless of the metastatic site. In contrast, losses of 1p, 5q, and 14q, amongst others, were positioned further away from the root in the oncogenetic trees, indicating their occurrence as intermediate to late events in cancer development.

The oncogenetic models indicated site-specific differences in the relative timing and nature of cytogenetic events. For example, gain of 20p was modeled as intermediate event in liver and brain metastasis, whereas this event represented a late event in lung metastasis. Conversely, loss of 20p was modeled as an intermediate event in lung metastases but late event in liver metastases. Organotropic CI enriched

in the brain, such as -3q, +5q, -6q, and +12p, were modeled to be acquired later during tumor evolution. In contrast, those enriched in both liver and brain metastases, including +7p, -8p, +8q, +13q, -18p, +20q, were predominantly positioned early in the oncogenetic tree. Thus, the CI patterns observed in the CRC metastasis are also reflected in cytogenetic evolutionary pathways, revealing both common and distinct patterns across different metastatic sites.

Cytogenetic tumor evolution in individual patients

Given the individual nature of tumor evolution, we explored whether the event divisions shown in the oncogenetic trees—reflecting shared patterns across the cohort—are supported by phylogenetic trees derived from multiple samples of individual patients. To this end, we reconstructed phylogenetic trees for individual patients with multiple

Fig. 5 | Transcriptional profile of CRCs with deletions of *MLH1* and *BRCA1* in the TCGA cohort. a Volcano plot showing differential gene expression between *MLH1* deletion CRC and *MLH1* control CRCs, with genes colored by chromosomal arm (inset highlights genes on chromosome 3; horizontal dashed line, $p_{\text{adjust}} = 0.05$; vertical dashed lines, 1.25-fold change). Significance was determined by two-sided Wilcoxon rank-sum tests with Benjamini-Hochberg adjustment ($n = 276$). **b** Two-sided GSEA using C1 positional gene sets indicating that genes encoded on chromosomal arm 3p (ideogram on the left) are significantly downregulated in CRC with *MLH1* deletion compared to *MLH1* control CRCs ($n = 276$). Circle size represents NES, color indicates FDR-adjusted p -value. **c** G:profiler analysis of genes upregulated in *MLH1* deletion CRCs, categorized by Gene Ontology (GO), including molecular function (MF), biological process (BP), and cellular component (CC) ($n = 276$). P -values are adjusted using the g:SCS method. **d** Differential gene expression of the *HOXA* and *HOXD* loci and flanking genes, with circle size based on \log_2 (fold change) and color reflecting Benjamini-Hochberg-adjusted p -values from two-sided Wilcoxon rank-sum tests (open circles indicate $p_{\text{adjust}} > 0.05$; $n = 276$). **e** CNV analysis of the *HOXA* and *HOXD* loci (two-sided Fisher exact test with Benjamini-Hochberg correction; n.s. not significant, (1) $p_{\text{adjust}} = 0.757$; (2)

$p_{\text{adjust}} = 0.941$; $n = 276$). **f** Volcano plot visualizing differential gene expression for *BRCA1* deletion CRC vs. *BRCA1* control CRC ($n = 213$). Inset shows genes on chromosome 17. Significance determined by two-sided Wilcoxon rank-sum tests with Benjamini-Hochberg adjustment. **g** Two-sided GSEA using C1 positional gene sets for chromosome 17 (ideogram) reveals that genes encoded on 17q are significantly downregulated in CRCs with *BRCA1* deletion compared to *BRCA1* control CRC ($n = 213$). Circle size represents NES, color indicates FDR-adjusted p -value. **h** G:profiler analysis of genes upregulated in *BRCA1* deletion CRCs compared to *BRCA1* control CRC ($n = 213$) with GO, KEGG (Kyoto Encyclopedia of Genes and Genomes), and Reactome (REAC) pathways (p -values adjusted by g:SCS method). **i** Differential gene expression of the *HOXA* locus and surrounding genes, with the top showing upregulated genes in *BRCA1* deletion CRCs and the bottom showing downregulated ones; circle size reflects \log_2 (fold change), while color indicates Benjamini-Hochberg-adjusted p -values from two-sided Wilcoxon rank-sum test (open circles denote $p_{\text{adjust}} > 0.05$; $n = 213$). Panels **a–e** compare CRCs with *MLH1* deletion to control CRCs, while (**f–i**) compare CRCs with *BRCA1* deletion to control CRCs. Source data are provided as a Source Data file.

tumor samples available for multi-regional cytogenetic analysis (Fig. 7 and Supplementary Fig. 6). The coloring of events was based on the event classes introduced in Fig. 6.

The inferred phylogenetic trees suggested that cytogenetic divergence between the primary CRC and its metastatic spread may occur at various time points (e.g., Figure 7c, g, h). In line with the oncogenetic models (Fig. 6), the loss of the long arm of chromosome 18 is shared among different tumors as a truncal chromosomal aberration in the individual patients' phylogenetic trees (Fig. 7), independent of the site of metastasis. Brain-tropic CIs identified as late events in oncogenetic modeling (Fig. 6) were preferentially seen in the branches of the phylogenetic trees of individual patients, supporting the overall tumor evolution pattern.

To quantify these observations, we divided each phylogenetic tree into trunk/early, branches/intermediate, and sub-branches/late categories and determined a CI score for each individual patient. The analysis revealed that events positioned in the trunk of the individual phylogenetic trees were enriched in events classified as early in the oncogenetic models, while events in the branches were enriched in those attributed as later events in the oncogenetic trees. To assess the statistical significance of these observations, we employed linear mixed-effect modeling, followed by post-hoc testing. This model confirmed significant differences in CI score distributions between the categories ($p = 3.14 \times 10^{-5}$, likelihood ratio test; $p_{\text{adjust}} = 9.27 \times 10^{-3}$ for trunk/early vs. branches/intermediate; $p_{\text{adjust}} = 9.27 \times 10^{-3}$ branches/intermediate vs. sub-branches/late; and $p_{\text{adjust}} = 2.86 \times 10^{-7}$ trunk/early vs. sub-branches/late, posthoc test with Benjamini-Hochberg correction; Supplementary Fig. 7). Thus, while the evolution of each tumor is inherently individual, our analyses suggest that specific chromosomal aberrations are preferentially observed at distinct stages of tumor evolution, including early, intermediate, and late phases.

Discussion

Our study shows how site-enriched patterns of DNA copy number aberrations shape metastatic behavior, providing insights into CRC progression. Collectively, our findings demonstrate that the DNA copy number aberrations are not randomly distributed; rather, they exhibit distinct patterns at different metastatic sites in terms of (I) the total number of CIs, (II) the frequency and specificity of CNVs, and (III) the sequential order in which specific CIs are acquired during tumor evolution. The organotrophic CIs we identified showed significant co-occurrences in probabilistic models. Such co-occurrences would in principle allow for additive or even synergistic effects in promoting metastatic spread.

Most organotrophic CIs we identified were either brain-and-liver-tropic or brain-tropic. In this respect, metastatic colonization of the

brain poses a number of challenges, as cancer cells must overcome barriers such as the blood-brain barrier and adapt to the unique tissue environment characterized by hypoxia and reduced glucose concentrations⁷⁴. To successfully establish tumors in the brain, metastatic cells likely require additional capabilities beyond those needed for primary tumor growth and peripheral metastasis. The organotrophic CIs identified in our analysis correspond to larger CNVs, often involving whole chromosomal arm or whole chromosomal aneuploidies, that impact numerous genes simultaneously. Such extensive alterations may disrupt or deregulate multiple critical regulatory networks, potentially granting cancer cells a selective advantage in adapting to various pressures⁷⁵. In particular, it has been shown that aneuploidy inversely correlates with immune infiltration of tumor tissue^{29,75}, and chromosomal copy number changes affect immune cell function⁷⁶. These observations suggest that aneuploidy facilitates escape from immune surveillance^{29,75}. In addition, aneuploidy appears to modify drug response in cancer cells⁷⁷.

One of the organotrophic CIs identified in our study is the gain of chromosome 12p, which we modeled as a late event in oncogenetic progression and found to be preferentially associated with brain metastases. This chromosomal arm encodes several critical genes, including the oncogene *KRAS*, which is a driver in CRC development³³. A recent meta-analysis indicated that the majority of brain metastases examined harbor *KRAS* mutations⁴⁸. Furthermore, previous studies on CRCs have established an association between gains of 12p and somatic *KRAS* mutations⁷⁸, reinforcing our observations. This suggests that the +12p event – particularly in the presence of oncogenic *KRAS* mutations – contributes to the metastatic organotropism observed in brain metastases.

At the level of protein function, our analysis shows that the *KRAS* CNV in *KRAS* mutant CRCs translates into enhanced *KRAS* pathway activity. This finding aligns with our observation showing that *KRAS* amplifications further intensify metabolic rewiring in *KRAS* mutated CRCs, leading to an upregulation of the glycolytic pathway, amongst others. Although glycolysis is less efficient than oxidative phosphorylation in terms of ATP production, it generates crucial biomolecular intermediates necessary for cell proliferation, thereby supporting tumor growth^{79,80}.

A pivotal player in this metabolic adaptation is Hypoxia Inducible Factor 1 Alpha (HIF1 α), a master regulator of glycolysis⁸¹ that upregulates the expression of glycolytic enzymes⁸². In this respect, mutant *KRAS* enhances HIF1 α expression at the transcriptional level⁸³ and also stabilizes HIF1 α by inhibiting prolyl hydroxylases (PHDs), which mark HIF1 α for degradation⁸⁴. This dual mechanism not only amplifies glycolytic flux but also promotes a tumor microenvironment conducive to survival under hypoxic conditions.

This adaptative capacity appears to be particularly relevant in the context of brain metastases, as previous studies have shown that HIF1 α levels are significantly elevated in CRC brain metastases compared to paired primary CRC samples⁸⁵. In line with increased HIF1 α in brain metastases, we observed a significantly higher Winter hypoxia score in *KRAS* mutated and amplified CRC, suggesting that these tumors possess an enhanced capacity for hypoxic adaptation. Given that the partial pressure of oxygen (PtO₂) in the human brain is relatively low – estimated at ~20–25 mmHg – compared to around 55–60 mmHg in the large bowel, 50–55 mmHg in the liver, and over 100 mmHg in the lung alveolus⁸⁶, we propose that the enhanced hypoxic adaptation in *KRAS* mutated and amplified CRC may facilitate colonization and growth of cancer cells in the brain, thereby contributing to the organotropism towards the brain.

Moreover, oncogene activation such as resulting from *RAS* mutations are well-established inducers of replication stress, which in turn activates the ATR DNA damage response pathway⁸⁷. This stress response serves as an adaptive mechanism, enabling *KRAS*-driven cancer cells to tolerate the genomic instability associated with aberrant proliferation. However, our findings appear to reveal a seemingly paradoxical relationship between mutated and amplified *KRAS* and the ATR damage response pathway in CRC. In particular, our analysis revealed that the ATR pathway is modestly yet significantly down-regulated in CRC harboring *KRAS* mutations and amplifications. The extent of this ATR pathway suppression may be a crucial factor influencing the behavior of *KRAS*-mutant cancer cells. When ATR levels are strongly reduced, the resultant genomic instability can lead to synthetic lethality, wherein the accumulation of unrepaired DNA damage triggers cell death in these cancer cells⁸⁸. Conversely, when ATR suppression is moderate, genomic instability may promote tumor development in *KRAS*-mutant cells⁸⁸. In this scenario, residual ATR activity may provide sufficient capacity for DNA repair, allowing the cells to tolerate chromosomal aberrations and eventually benefit from additional driver alterations for continuation of their malignant progression. This delicate balance between DNA damage and repair capacity may drive the evolution of more aggressive cancer phenotypes. Consistently, we observed the highest number of CIs and the greatest FGA in brain metastases.

Organotropism in cancer is likely influenced by a variety of factors, with *KRAS* alterations being a significant contributor. Of particular interest are *MDM2* and *CDK4*, both of which, like *KRAS*, are encoded on chromosome 12. We observed a significant enrichment of gains in *MDM2* and *CDK4* in brain metastases, suggesting that these alterations may play a role in organotropism alongside *KRAS* mutations. This observation is further supported by our pathway analyses, which reveal a notable upregulation of the *MDM2* pathway in CRC with *KRAS* mutations and amplifications. The amplification of *MDM2* leads to dysregulation of the p53 pathway⁸⁹, while *CDK4* amplification enhances cell proliferation⁹⁰. Collectively, these genomic alterations may contribute to the aggressive phenotype observed in brain metastases.

In addition to pathways directly linked to genes encompassed by CIs, our analysis uncovered also alterations in other biological processes, including those involving gene members encoded in regions not covered by CIs. Notably, we found that deletions of the DNA repair genes *MLH1* and *BRCA1* are organotropically enriched in brain metastases. While haploinsufficiency of these genes may contribute to genomic instability^{91,92}, we also observed significant upregulations of specific *HOX* gene sets. In CRCs with *MLH1* deletions, there was significant upregulation of *HOXA1* and *HOXD9* – *HOXD13*, whereas in instances of *BRCA1* deletions, we identified increased expression of *HOXA3* and *HOXA10*. This suggests a complex regulatory landscape, in which the transcriptional activation of *HOX* genes may function as a metastasis promoting mechanism, operating independent of apparent gene dosage effects.

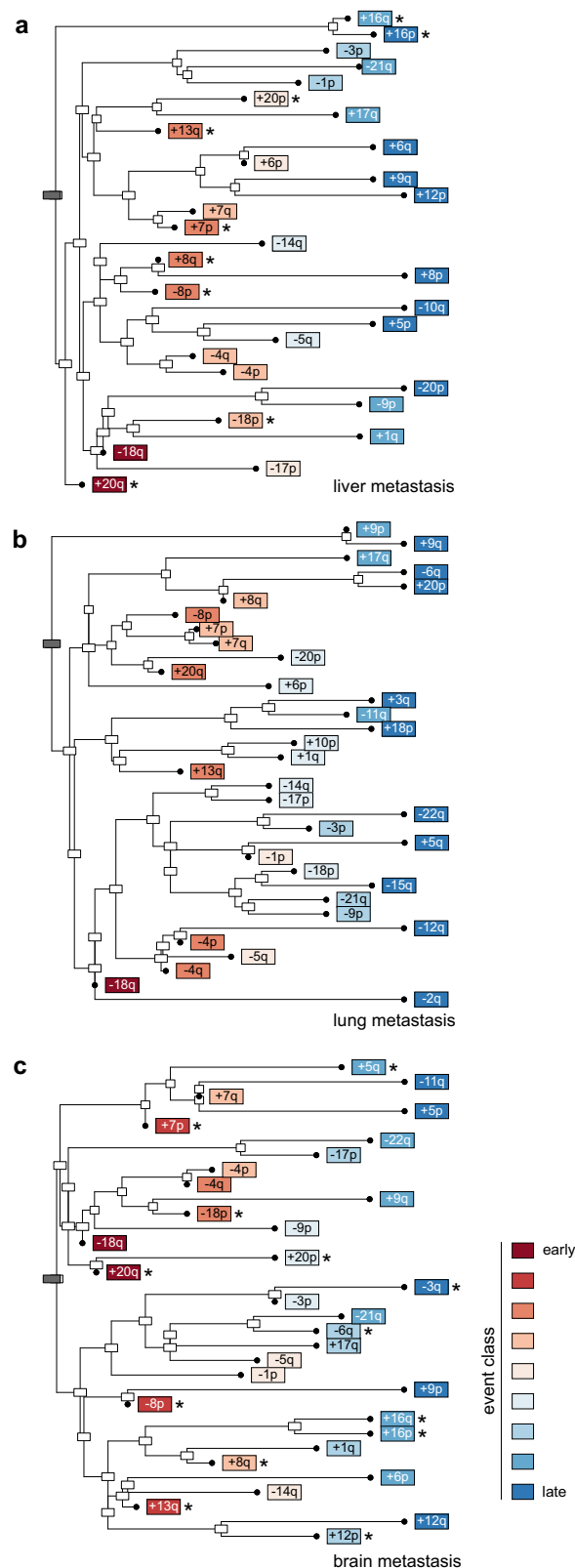


Fig. 6 | Oncogenetic tree models of metastatic CRC based on maximum likelihood estimation. Analyses were done using the CRCTropism cohort for the following metastasis sites: **a.** liver ($n = 98$), **b.** lung ($n = 69$), and **c.** brain ($n = 33$). CIs were classified into nine categories from early to late according to proximity to the root (gray boxes). Organotropic events are marked by asterisks. Source data are provided as a Source Data file.

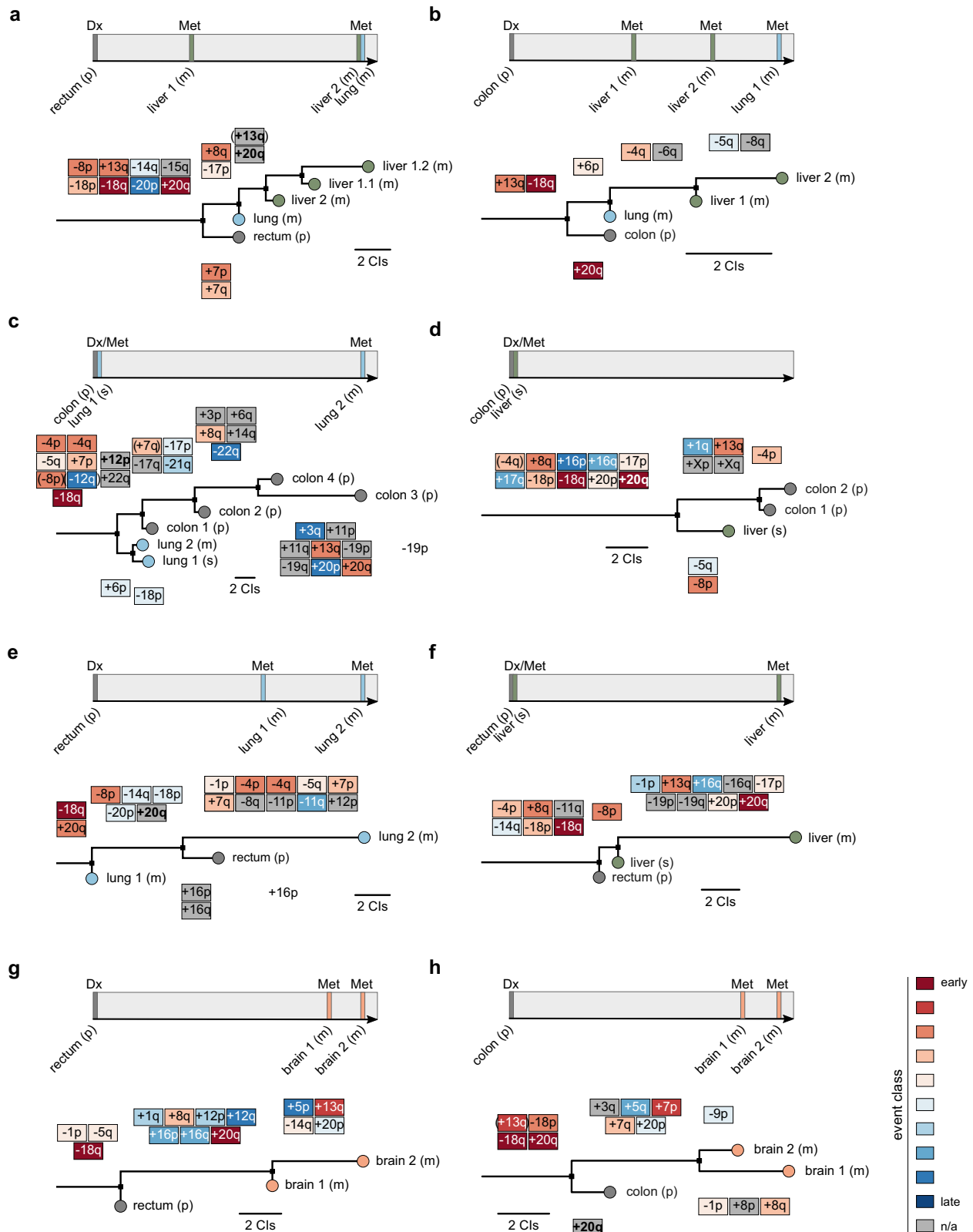


Fig. 7 | Reconstruction of phylogenetic trees using multi-sample cytogenetic data from individual CRC patients of the CRCtropism cohort. a–h The clinical course (top) and the reconstructed phylogenetic tree (bottom) are shown, with the tree scaled according to the number of Cls. Maximum parsimony was used for tree inference. Cl events are colored based on the classification introduced in Fig. 6. If a

particular event could not be uniquely linked (indicated by brackets), it is displayed at the earliest suggested occurrence. High-level amplifications are indicated in bold. The site of the primary tumor (p) and metastases (m, metachronous metastasis; s, synchronous metastasis) is also listed. Dx, diagnosis of primary cancer; Met, metastasis. Source data are provided as a Source Data file.

The *HOX* gene family, known for its role in embryonic development, appears to be repurposed in the context of tumorigenesis. A growing body of literature highlights the involvement of *HOX* genes in orchestrating aggressive tumor phenotypes and enhancing metastatic capabilities across various malignancies. For instance, *HOXD9* has been shown to significantly enhance critical malignant characteristics of CRC cells, including proliferation as well as migration and invasion⁹³. The authors revealed a bidirectional relationship wherein TGF- β 1 elevates *HOXD9* expression, while *HOXD9* promotes TGF- β 1-triggered EMT, a pivotal process that empowers cancer cells to invade and metastasize⁹³. Similarly, *HOXA10* has been implicated in promoting EMT through the TGF- β signaling cascade⁹⁴, and *HOXD13* is linked to proliferation and invasion⁹⁵, reinforcing the notion that *HOX* gene dysregulation may contribute to the aggressive behavior of CRC.

In contrast to liver and brain metastases, the only lung-enriched CNV is a gain of *JUN*, which is also enriched in brain metastases. *JUN* codes for a pro-proliferative transcription factor⁹⁶ that interacts with the RAS signaling pathway⁹⁷. This seems remarkable given the previously reported increased rate of *KRAS* mutations in CRC lung metastases⁹⁸. The marked underrepresentation of organotropic events in lung metastases is in line with a recent study⁴⁵, and it has also been reported that *SRC*, localized in chromosomal band 20q11.23, is less frequently amplified in lung metastases²⁸. Overall, lung metastases showed the lowest number of CIs in our study. Taken together, these observations suggest that site-specificity in pulmonary seeding – besides a potential role of the JUN pathway – may follow other mechanisms that are not (strictly) dependent on the presence of specific CIs¹⁸.

The concept of early metastatic spread in CRC^{5–7} provides a framework for these observations. We suggest a model in which the early spread inherently involves tumor cells at early cytogenetic developmental stages, setting the foundation for distinct metastatic trajectories. When these early-disseminating tumor cells encounter the oxygen-rich pulmonary microenvironment, they appear capable of establishing metastases without requiring extensive chromosomal alterations. The favorable conditions of the lung environment could enable clonal expansion at a relatively early cytogenetic stage, resulting in lung metastases that can maintain this early cytogenetic profile. Although this oxygen-rich environment initially seems to offer conditions for metastatic growth, the notable enrichment of *KRAS* mutations in lung metastases²⁸ points to a more nuanced process. Specifically, *KRAS* mutations could thereby reduce the selective pressure for additional chromosomal imbalances. This stands in contrast to liver and brain metastases, where microenvironmental conditions might necessitate additional genomic alterations for successful colonization, in line with the more complex cytogenetic profiles. While our findings point toward minimal requirements for lung colonization, they do not exclude the possibility that metastatic cells continue to evolve and acquire additional chromosomal alterations within the lung microenvironment, nor do they preclude the successful establishment of lung metastases by tumor cells with more complex cytogenetic profiles. Indeed, our cohort and the MetTropism cohort²⁸ includes lung metastases displaying such advanced cytogenetic profiles.

Overall, the exploration of organotropic colonization signatures in metastatic CRC reveals a complex interplay of genetic factors that influence tumor behavior. Whereas the identified organotropic CIs suggest a predisposition for growth at specific metastatic sites, they do not provide a distinction between CRCs that metastasize and those that remain localized. This observation aligns with a growing body of evidence indicating that the genomic profiles of primary CRCs, regardless of metastatic spread, exhibit remarkable similarities in both mutation and CNV patterns^{28,99}. Notably, only a limited subset of alterations—such as *SMAD4* mutations and *MYC* amplifications—has been significantly associated with metastatic CRC^{28,99}. This observation underscores the need for further investigation of the mechanisms driving metastasis beyond invasion depths.

One potential approach for distinguishing metastatic potential in CRCs lies in examining the timing and sequence of acquiring specific CIs, which may play a critical role in shaping the biological behavior of cancer cells, including their metastatic potential. This notion parallels models proposed in the study of myeloproliferative neoplasms, where the order of mutations influences disease characteristics¹⁰⁰. Supporting this perspective, recent research focusing exclusively on primary tumors has identified chromosomal gains at 7p and 8q as early events in the development of metastatic CRC²⁴. In contrast, these same alterations were observed as late cytogenetic events in primary CRCs lacking clinical evidence of synchronous or metachronous metastases²⁴. This temporal distinction suggests that the early acquisition of chromosomal gains at these loci may favor metastatic dissemination and could serve as biomarker for assessing metastatic potential in CRC. In light of these findings, exploring how early genetic changes might confer a selective advantage to tumor cells during the metastatic cascade represents a direction for future research.

In conclusion, our results support a model in which a site-specific pattern of cytogenetic events and the order in which particular CIs are acquired contribute to the potential of CRC to colonize particular secondary sites. Involvement of genes linked to signaling, chromatin/transcription, EMT, DNA repair and cell cycle may assist cancer cells to initiate and maintain a cascade that ultimately favors their survival in potentially challenging environments at the secondary site. The observed patterns of co-occurring genomic alterations and gene expression changes provide insights into the molecular mechanisms underlying CRC progression and metastasis, potentially offering approaches for targeted therapies and prognostic markers that predict the metastatic potential of CRC. Thus, the differential CIs described here add a cytogenetic layer to Stephen Paget's seed and soil hypothesis¹³ and link cytogenetic events to the phenomenon of metastatic organotropism.

Methods

Cohorts

The study was approved by the local ethics committees (University Medical Center Göttingen, Ludwig Maximilian University of Munich) and included data published previously^{60,63}. For the original data, the ethics committee of the University Medical Center Göttingen waived the requirement for informed consent. Data collection from the MSK MetTropism²⁸ and TCGA (PanCancer Atlas)^{65,66} cohorts were collected by their respective research teams and made publicly available. The original studies reported that data collection was approved by their respective ethics committees, with informed consent obtained from patients. For the CRCTropism cohort, data comprised CI information, supplemented with the patient's sex and age at diagnosis as well as metastasis site. The sex was determined through self-reporting, combined with results from cytogenetic analyses. CIs in primary CRCs and metastases were determined by comparative genomic hybridization (CGH)¹⁰¹. Signal profiles were assigned for each chromosome by hierarchical cluster analysis. Threshold criteria were adjusted for the signal-to-background ratio to account for differences including tumor purity. We excluded the chromosomal regions 1p32pter, 13p, 14p, 15p, 21p, 22p, telomeres, and constitutive heterochromatic regions at 1q, 9q, 16q, and Yq from the analysis due to insufficient reliability as described previously¹⁰¹, as well as individual chromosomal regions that did not pass quality criteria. Chromosomal arm aneuploidy was defined as chromosomal gain or loss that apparently encompassed the entire chromosomal arm²⁹. Short arms of chromosomes 13, 14, 15, 21, and 22 were not considered for the aneuploidy count. Except where otherwise stated, aneuploidies and focal CIs were combined in the analyses. For statistical analysis, only one metastasis per patient was selected. In cases where multiple metastatic lesions were analyzed, we prioritized the selection based on their relative frequency within our dataset. Specifically, brain metastases were given the highest priority

due to their lower occurrence in the cohort, followed by lung metastases, and subsequently liver metastases. This prioritization ensured that our analysis reflected the characteristics of less common metastatic sites. When multiple metastases were present in the same location, we selected the one with the lowest number of CIs for analysis. This selection was based on the hypothesis that the metastasis with the fewest CIs might represent a state closer to the minimum genomic alterations potentially associated with metastatic spread.

The NGS-based MSK MetTropism cohort²⁸ was selected as independent cohort. Identified single/multi-nucleotide variants and indels, segmented copy number alterations (with subtraction from matched normals), the TMB, FGA, MSI score and clinical data were retrieved from cBioPortal^{102,103}. The resource OncoKB¹⁰⁴ was used to classify mutations according to oncogenic potential. Only mutations listed as ‘oncogenic’ or ‘likely oncogenic’ were considered. Aneuploidies were determined by ASCETS¹⁰⁵ version 1.0 using default parameters. Gains and losses were analyzed using GISTIC 2.0¹⁰⁶ with a threshold for copy number amplifications and deletions of 0.3 and a *q*-value threshold of 0.05, conducted using GenePattern¹⁰⁷. Only genes included in the MSK gene panel were considered in the analyses on the gene level. In addition, we obtained genomic and transcriptomic data from the TCGA project (PanCancer Atlas)^{65,66}. The data, encompassing cancer subtype data, somatic mutations, CNVs, aneuploidies, transcriptomic profiles derived from RNA sequencing, mRNA expression z-scores relative to diploid samples, methylation data, MSIsensor¹⁰⁸ scores, as well as Winter hypoxia scores were accessed through the Genomic Data Commons Data (GCD) Portal¹⁰⁹, cBioPortal^{102,103} as well as from a previous study¹¹⁰. CRCs classified as *POLE* subtype or unknown subtype were excluded.

Functional proteomics data

RPPA data⁶⁸ on TCGA cancer samples were downloaded from the GDC data portal. Protein data included in the analysis were required to have successfully passed quality control (validation status ‘valid’).

Co-occurrence modeling

The R¹¹¹ package cooccur¹¹² (version 1.3) was used to derive a probabilistic model of co-occurring CIs in CRC metastases. *P*-values obtained from the analyses were adjusted for multiple testing using the Benjamini-Hochberg method¹¹³. Co-occurrences are represented using a color scale: significant positive associations are indicated with shades of red, while significant negative associations are depicted in shades of blue. Associations that were not statistically significant and combinations that were unavailable are left blank.

Organotropic mapping

Data were visualized in a ternary plot using the R package ggtern¹¹⁴ built 3.5.0. A minimum occurrence threshold of three times was established for CI events to be included in the analysis. The relative frequency of each event observed in the cohort was calculated for the given metastasis site, and used to position the data point for each chromosomal arm in the organotropic map. The relative frequencies of each chromosomal aberration in the cohort were used to set the diameter of the respective bubble. To assess the statistical significance, two-sided Fisher exact tests, followed by *p*-value correction for multiple testing using the Benjamini-Hochberg method¹¹³ were computed. Statistically significant chromosomal aberrations were defined as organotropic aberrations.

Transcriptome and pathway analysis

For identification of differentially expressed genes, transcriptomic data were processed using edgeR¹¹⁵ package version 4.0.3. For chromosome annotation, the package Organism.dplyr version 1.30.1 was used. Differential gene expression analysis was analyzed using two-

sided Wilcoxon rank-sum tests with Benjamini-Hochberg adjustment¹¹⁶ to compare the following CRC groups:

- (I) *KRAS* mutation and amplification;
- (II) *KRAS* mutation;
- (III) *KRAS* wildtype;
- (IV) *KRAS* deletion (but no oncogenic mutation);
- (V) *MLH1* del: *MLH1* deletion, but no pathogenic mutation in *MLH1*, *MSH2*, *MSH6*, *PMS2*, no deletion of *MSH2*, *MSH6*, *PMS2*, no *MLH1* methylation defined as beta value < 0.3¹¹⁷, no low expression of *MSH2*, *MSH6*, and *PMS2* defined as mRNA expression z-scores ≥ mean – one standard deviation (SD);
- (VI) *MLH1* control: no pathogenic mutation and no deletion in the genes *MLH1*, *MSH2*, *MSH6*, and *PMS2*, no methylation defined as beta value < 0.3¹¹⁷ and no low expression defined as mRNA expression z-scores ≥ mean – one SD, MSIsensor score < 3.5, not classified as MSI in cBioPortal;
- (VII) *BRCA1* del: *BRCA1* deletion, but no pathogenic mutation in *BRCA1*, *BRCA2*, *PALB2*, *BARD1*, no deletion of *BRCA2*, *PALB2*, *BARD1*, no low expression of *BRCA2*, *PALB2*, *BARD1* defined as mRNA expression z-scores ≥ mean – one SD, MSIsensor score < 3.5, no *MLH1* methylation defined as beta value < 0.3¹¹⁷;
- (VIII) *BRCA1* control: in the genes *BRCA1*, *BRCA2*, *PALB2*, *RAD51C*, *BARD1* no pathogenic mutation and no deletion, no low expression defined as mRNA expression z-scores ≥ mean – one SD, MSIsensor score < 3.5, no *MLH1* methylation defined as beta value < 0.3¹¹⁷.

Differential gene expression data were visualized using volcano plots, with data points colored according to their respective chromosomal arms. Differentially expressed genes were defined using the following cut-offs for the log₂(fold change) and statistical significance: 1.25-fold upregulation or downregulation and *p*_{adjust} < 0.05. For enrichment analyses, the following packages and parameters were used:

- (I) Enrichr¹¹⁸: background gene list used, gene set definitions: GO^{119,120}, and MSigDB Hallmark^{121,122};
- (II) GSEA version 20.4.0¹²³, conducted in GenePattern¹⁰⁷: permutation type: phenotype, scoring scheme: weighted, metric for ranking genes: Signal2Noise, normalization mode: meandiv, randomization mode: no_balance, gene set definitions: MSigDB^{121,122};
- (III) ssGSEA version 10.1.0¹²⁴, conducted in GenePattern¹⁰⁷: weighting exponent 0.75, gene set definitions: MSigDB^{121,122};
- (IV) g:profiler version e111_eg58_p18_f463989d¹²⁵: organism: *homo sapiens*, gene set definitions: GO^{119,120}, KEGG¹²⁶, and Reactome¹²⁷.

Oncogenetic tree models

Oncogenetic tree models derived by maximum likelihood estimation were computed using the R¹¹¹ package oncomodel version 1.0¹²⁸. The most common CIs found in the respective CRC metastasis type were considered for tree modeling. Chromosomal aberrations were grouped in nine classes according to their distance from the root of the oncogenetic tree (equal distances).

Phylogenetic tree reconstruction using multi-region analyses

For patients with available CI data from multiple samples (primary CRC and at least one metastasis), CI events were encoded as discrete data. These data were then analyzed to infer phylogenetic relationships using the maximum parsimony method, as implemented in the software suite PHYLIP¹²⁹ version 3.697. The resultant phylogenetic trees were visualized using FigTree (version 1.4.4; <http://tree.bio.ed.ac.uk/software/figtree/>). CIs were color-encoded according to their classification within the oncogenetic trees.

Data modeling

To investigate the relationship between *KRAS* copy number and MAP2K1/2 phosphorylation in *KRAS*-mutated CRCs, we employed a multiple linear regression model. The response variable in our model was the level of phosphorylated MAP2K1/2, which serves as a marker for MAP2K1/2 activation. Our primary predictor of interest was the copy number of *KRAS* (*KRAS_CNV*). Additionally, we included two other predictors: the total protein level of MAP2K1 (pMAP2K1) and the level of phosphorylated BRAF at serine 445 (pBRAF_S445), which serves as an indicator of BRAF. For linear regression modeling, the `lm` function implemented in the R package `stats` (built 4.3.1) was used. The significance of the *KRAS_CNV* effect was assessed using a likelihood ratio test comparing the full model to a reduced model without the *KRAS_CNV* term. We checked for multicollinearity among predictors using VIFs calculated with the `car` package (version 3.1.2). Variable importance was assessed using the `vip` package (version 0.4.1). Model diagnostics, including residual plots and tests for normality and homoscedasticity, were performed using base R functions and the `lmtest` package (version 0.9.40). To assess the evolutionary trajectories, we calculated CI scores, which correspond to the average of the CI event categories derived from the phylogenetic trees. Linear mixed-effect modeling was conducted using the package `lme4`¹³⁰ (version 1.1.35.5), with tumor evolutionary step as fixed effect and tumors as random effect to analyze CI scores. Post-hoc analyses with Benjamini-Hochberg adjustment¹¹³ were performed following the modeling.

Further statistics, reproducibility and data visualization

Statistical analyses were performed using two-sided tests. The Wilcoxon rank-sum test with continuity correction, Kruskal-Wallis tests with Dunn post-hoc tests and Fisher exact test for contingency tables were used, as indicated. To address multiple testing, *p*-values were corrected using the Benjamini-Hochberg procedure¹¹³. The significance level was set at *p* < 0.05 (when applicable, after correction for multiple testing). The statistical environment *R*¹¹¹ (version 4.3.3) was used for statistical analysis and data visualization using `ggplot2` version 3.5.1. No statistical method was used to predetermine sample size. No data that passed quality control were excluded from the analyses. The experiments were not randomized. The Investigators were not blinded to allocation during experiments and outcome assessment.

Reporting summary

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

Data availability

The previously published CGH data^{60,63} comprise chromosomal imbalance profiles indicating net clonal changes, characterized by regions of chromosomal losses, gains, and amplifications at the chromosomal band level, accompanied by basic demographic information including sex and age. The MSK MetTropism publicly available data²⁸ used in this study are available in the cBioPortal database (<https://www.cbioportal.org>) under accession code MSK MetTropism. The TCGA publicly available data^{65,66,68} used in this study are available at the GDC Portal (<https://portal.gdc.cancer.gov>) under accession codes TCGA-COAD and TCGA-READ and cBioPortal under accession code Colorectal Adenocarcinoma (TCGA, PanCancer Atlas). The remaining data are available within the Article, Supplementary Information or Source Data file. Source data are provided with this paper.

References

- Global Burden of Disease Cancer Collaboration. et al. *Global, Regional, and National Cancer Incidence, Mortality, Years of Life Lost, Years Lived With Disability, and Disability-Adjusted Life-Years for 29 Cancer Groups, 1990 to 2017: A Systematic Analysis for the Global Burden of Disease Study*. *JAMA Oncol* **5**, 1749–1768 (2019).
- Chaffer, C. L. & Weinberg, R. A. A perspective on cancer cell metastasis. *Science* **331**, 1559–1564 (2011).
- Brouwer, N. P. M. et al. An overview of 25 years of incidence, treatment and outcome of colorectal cancer patients. *Int. J. Cancer* **143**, 2758–2766 (2018).
- Elferink, M. A., de Jong, K. P., Klaase, J. M., Siemerink, E. J. & de Wilt, J. H. Metachronous metastases from colorectal cancer: a population-based study in North-East Netherlands. *Int. J. Colorectal Dis.* **30**, 205–212 (2015).
- Hu, Z. et al. Quantitative evidence for early metastatic seeding in colorectal cancer. *Nat. Genet* **51**, 1113–1122 (2019).
- Hu, Z., Li, Z., Ma, Z. & Curtis, C. Multi-cancer analysis of clonality and the timing of systemic spread in paired primary tumors and metastases. *Nat. Genet* **52**, 701–708 (2020).
- Alves, J. M., Prado-Lopez, S., Cameselle-Teijeiro, J. M. & Posada, D. Rapid evolution and biogeographic spread in a colorectal cancer. *Nat. Commun.* **10**, 5139 (2019).
- Wei, Q. et al. Multiregion whole-exome sequencing of matched primary and metastatic tumors revealed genomic heterogeneity and suggested polyclonal seeding in colorectal cancer metastasis. *Ann. Oncol.* **28**, 2135–2141 (2017).
- Dang, H. X. et al. The clonal evolution of metastatic colorectal cancer. *Sci. Adv.* **6**, eaay9691 (2020).
- Naxerova, K. et al. Origins of lymphatic and distant metastases in human colorectal cancer. *Science* **357**, 55–60 (2017).
- Riihimäki, M., Hemminki, A., Sundquist, J. & Hemminki, K. Patterns of metastasis in colon and rectal cancer. *Sci. Rep.* **6**, 29765 (2016).
- Fidler, I. J. & Nicolson, G. L. Organ selectivity for implantation survival and growth of B16 melanoma variant tumor lines. *J. Natl Cancer Inst.* **57**, 1199–1202 (1976).
- Paget, S. The distribution of secondary growths in cancer of the breast. *Lancet* **133**, 571–573 (1889).
- Kaplan, R. N. et al. VEGFR1-positive haematopoietic bone marrow progenitors initiate the pre-metastatic niche. *Nature* **438**, 820–827 (2005).
- Hoshino, A. et al. Tumour exosome integrins determine organotropic metastasis. *Nature* **527**, 329–335 (2015).
- Fong, M. Y. et al. Breast-cancer-secreted miR-122 reprograms glucose metabolism in premetastatic niche to promote metastasis. *Nat. Cell Biol.* **17**, 183–194 (2015).
- Peinado, H. et al. Melanoma exosomes educate bone marrow progenitor cells toward a pro-metastatic phenotype through MET. *Nat. Med.* **18**, 883–891 (2012).
- Hongu, T. et al. Perivascular tenascin C triggers sequential activation of macrophages and endothelial cells to generate a pro-metastatic vascular niche in the lungs. *Nat. Cancer* **3**, 486–504 (2022).
- Kang, Y. et al. A multigenic program mediating breast cancer metastasis to bone. *Cancer Cell* **3**, 537–549 (2003).
- Minn, A. J. et al. Genes that mediate breast cancer metastasis to lung. *Nature* **436**, 518–524 (2005).
- Li, Y. et al. Patterns of somatic structural variation in human cancer genomes. *Nature* **578**, 112–121 (2020).
- ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium. Pan-cancer analysis of whole genomes. *Nature* **578**, 82–93 (2020).
- Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330–337 (2012).
- Golas, M. M. et al. Evolutionary patterns of chromosomal and microsatellite instability in proximal and distal colorectal cancer. *Colorectal Dis.* **24**, 157–176 (2022).
- Ried, T. et al. Comparative genomic hybridization reveals a specific pattern of chromosomal gains and losses during the genesis of colorectal tumors. *Genes Chromosomes Cancer* **15**, 234–245 (1996).

26. Smeets, D. et al. Copy number load predicts outcome of metastatic colorectal cancer patients receiving bevacizumab combination therapy. *Nat. Commun.* **9**, 4112 (2018).
27. Zhao, Q. et al. Comprehensive profiling of 1015 patients' exomes reveals genomic-clinical associations in colorectal cancer. *Nat. Commun.* **13**, 2342 (2022).
28. Nguyen, B. et al. Genomic characterization of metastatic patterns from prospective clinical sequencing of 25,000 patients. *Cell* **185**, 563–575.e511 (2022).
29. Taylor, A. M. et al. Genomic and functional approaches to understanding cancer aneuploidy. *Cancer Cell* **33**, 676–689.e673 (2018).
30. Gerstung, M. et al. The evolutionary history of 2,658 cancers. *Nature* **578**, 122–128 (2020).
31. Powell, S. M. et al. APC mutations occur early during colorectal tumorigenesis. *Nature* **359**, 235–237 (1992).
32. Baker, S. J. et al. Chromosome 17 deletions and p53 gene mutations in colorectal carcinomas. *Science* **244**, 217–221 (1989).
33. Forrester, K., Almoguera, C., Han, K., Grizzle, W. E. & Perucho, M. Detection of high incidence of K-ras oncogenes during human colon tumorigenesis. *Nature* **327**, 298–303 (1987).
34. Thiagalingam, S. et al. Evaluation of candidate tumour suppressor genes on chromosome 18 in colorectal cancers. *Nat. Genet* **13**, 343–346 (1996).
35. Fearon, E. R. & Vogelstein, B. A genetic model for colorectal tumorigenesis. *Cell* **61**, 759–767 (1990).
36. Davis, A., Gao, R. & Navin, N. Tumor evolution: Linear, branching, neutral or punctuated? *Biochim Biophys. Acta Rev. Cancer* **1867**, 151–161 (2017).
37. Gerlinger, M. et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N. Engl. J. Med* **366**, 883–892 (2012).
38. 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).
39. Baca, S. C. et al. Punctuated evolution of prostate cancer genomes. *Cell* **153**, 666–677 (2013).
40. Sottoriva, A. et al. A Big Bang model of human colorectal tumor growth. *Nat. Genet* **47**, 209–216 (2015).
41. Sun, R., Hu, Z. & Curtis, C. Big bang tumor growth and clonal evolution. *Cold Spring Harb Perspect Med* **8**, a028381 (2018).
42. Stephens, P. J. et al. Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell* **144**, 27–40 (2011).
43. Gao, R. et al. Punctuated copy number evolution and clonal stasis in triple-negative breast cancer. *Nat. Genet* **48**, 1119–1130 (2016).
44. Cross, W. et al. The evolutionary landscape of colorectal tumorigenesis. *Nat. Ecol. Evol.* **2**, 1661–1672 (2018).
45. Jiang, B. et al. Machine learning of genomic features in organotrophic metastases stratifies progression risk of primary tumors. *Nat. Commun.* **12**, 6692 (2021).
46. Mendelaar, P. A. J. et al. Whole genome sequencing of metastatic colorectal cancer reveals prior treatment effects and specific metastasis features. *Nat. Commun.* **12**, 574 (2021).
47. Ishaque, N. et al. Whole genome sequencing puts forward hypotheses on metastasis evolution and therapy in colorectal cancer. *Nat. Commun.* **9**, 4782 (2018).
48. Müller, S. et al. Brain metastases from colorectal cancer: a systematic review of the literature and meta-analysis to establish a guideline for daily treatment. *Cancers (Basel)* **13**, 900 (2021).
49. Yaeger, R. et al. RAS mutations affect pattern of metastatic spread and increase propensity for brain metastasis in colorectal cancer. *Cancer* **121**, 1195–1203 (2015).
50. Ghadimi, B. M. et al. Distinct chromosomal profiles in metastasizing and non-metastasizing colorectal carcinomas. *Cell Oncol.* **28**, 273–281 (2006).
51. Diep, C. B. et al. The order of genetic events associated with colorectal cancer progression inferred from meta-analysis of copy number changes. *Genes Chromosomes Cancer* **45**, 31–41 (2006).
52. Mekenkamp, L. J. et al. Chromosome 20p11 gains are associated with liver-specific metastasis in patients with colorectal cancer. *Gut* **62**, 94–101 (2013).
53. Al-Mulla, F., Keith, W. N., Pickford, I. R., Going, J. J. & Birnie, G. D. Comparative genomic hybridization analysis of primary colorectal carcinomas and their synchronous metastases. *Genes Chromosomes Cancer* **24**, 306–314 (1999).
54. Parada, L. A. et al. Cytogenetic analyses of secondary liver tumors reveal significant differences in genomic imbalances between primary and metastatic colon carcinomas. *Clin. Exp. Metastasis* **17**, 471–479 (1999).
55. Yamamoto, S. et al. Identification of chromosomal aberrations of metastatic potential in colorectal carcinoma. *Genes Chromosomes Cancer* **49**, 487–496 (2010).
56. Sayagues, J. M. et al. Mapping of genetic abnormalities of primary tumours from metastatic CRC by high-resolution SNP arrays. *PLoS One* **5**, e13752 (2010).
57. Bruin, S. C. et al. Specific genomic aberrations in primary colorectal cancer are associated with liver metastases. *BMC Cancer* **10**, 662 (2010).
58. Korn, W. M. et al. Chromosome arm 20q gains and other genomic alterations in colorectal cancer metastatic to liver, as analyzed by comparative genomic hybridization and fluorescence in situ hybridization. *Genes Chromosomes Cancer* **25**, 82–90 (1999).
59. Sayagues, J. M. et al. Intratumoural cytogenetic heterogeneity of sporadic colorectal carcinomas suggests several pathways to liver metastasis. *J. Pathol.* **221**, 308–319 (2010).
60. Danner, B. C. et al. Comparison of chromosomal aberrations in primary colorectal carcinomas to their pulmonary metastases. *Cancer Genet* **204**, 122–128 (2011).
61. Jiang, J. K., Chen, Y. J., Lin, C. H., Yu, I. T. & Lin, J. K. Genetic changes and clonality relationship between primary colorectal cancers and their pulmonary metastases—an analysis by comparative genomic hybridization. *Genes Chromosomes Cancer* **43**, 25–36 (2005).
62. Knosel, T., Schluns, K., Dietel, M. & Petersen, I. Chromosomal alterations in lung metastases of colorectal carcinomas: associations with tissue specific tumor dissemination. *Clin. Exp. Metastasis* **22**, 533–538 (2005).
63. Gutenberg, A. et al. High chromosomal instability in brain metastases of colorectal carcinoma. *Cancer Genet Cytogenet* **198**, 47–51 (2010).
64. Fehrmann, R. S. et al. Gene expression analysis identifies global gene dosage sensitivity in cancer. *Nat. Genet* **47**, 115–125 (2015).
65. Liu, Y. et al. Comparative molecular analysis of gastrointestinal adenocarcinomas. *Cancer Cell* **33**, 721–735.e728 (2018).
66. Cancer Genome Atlas Research Network et al. The cancer genome atlas pan-cancer analysis project. *Nat. Genet.* **45**, 1113–1120 (2013).
67. Palin, K. et al. Contribution of allelic imbalance to colorectal cancer. *Nat. Commun.* **9**, 3664 (2018).
68. Akbani, R. et al. A pan-cancer proteomic perspective on the cancer genome atlas. *Nat. Commun.* **5**, 3887 (2014).
69. Alessi, D. R. et al. Identification of the sites in MAP kinase kinase-1 phosphorylated by p74raf-1. *EMBO J.* **13**, 1610–1619 (1994).
70. Dang, E. V. et al. Control of T(H)17/T(reg) balance by hypoxia-inducible factor 1. *Cell* **146**, 772–784 (2011).
71. Semenza, G. L., Roth, P. H., Fang, H. M. & Wang, G. L. Transcriptional regulation of genes encoding glycolytic enzymes by hypoxia-inducible factor 1. *J. Biol. Chem.* **269**, 23757–23763 (1994).

72. Yang, M. H. et al. Direct regulation of TWIST by HIF-1 α promotes metastasis. *Nat. Cell Biol.* **10**, 295–305 (2008).
73. Winter, S. C. et al. Relation of a hypoxia metagene derived from head and neck cancer to prognosis of multiple cancers. *Cancer Res* **67**, 3441–3449 (2007).
74. Yuzhalin, A. E. & Yu, D. Brain metastasis organotropism. *Cold Spring Harb Perspect Med* **10**, a037242 (2020).
75. Davoli, T., Uno, H., Wooten, E. C. & Elledge, S. J. Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy. *Science* **355**, eaaf8399 (2017).
76. Xian, S. et al. The unfolded protein response links tumor aneuploidy to local immune dysregulation. *EMBO Rep.* **22**, e52509 (2021).
77. Shukla, A. et al. Chromosome arm aneuploidies shape tumour evolution and drug response. *Nat. Commun.* **11**, 449 (2020).
78. Leslie, A. et al. Mutations of APC, K-ras, and p53 are associated with specific chromosomal aberrations in colorectal adenocarcinomas. *Cancer Res* **63**, 4656–4661 (2003).
79. Hu, Y. et al. K-ras(G12V) transformation leads to mitochondrial dysfunction and a metabolic switch from oxidative phosphorylation to glycolysis. *Cell Res* **22**, 399–412 (2012).
80. Racker, E., Resnick, R. J. & Feldman, R. Glycolysis and methylenoisobutyrate uptake in rat-1 cells transfected with ras or myc oncogenes. *Proc. Natl Acad. Sci. USA* **82**, 3535–3538 (1985).
81. Seagroves, T. N. et al. Transcription factor HIF-1 is a necessary mediator of the Pasteur effect in mammalian cells. *Mol. Cell Biol.* **21**, 3436–3444 (2001).
82. Mole, D. R. et al. Genome-wide association of hypoxia-inducible factor (HIF)-1 α and HIF-2 α DNA binding with expression profiling of hypoxia-inducible transcripts. *J. Biol. Chem.* **284**, 16767–16775 (2009).
83. Wang, Y., Lei, F., Rong, W., Zeng, Q. & Sun, W. Positive feedback between oncogenic KRAS and HIF-1 α confers drug resistance in colorectal cancer. *Onco Targets Ther.* **8**, 1229–1237 (2015).
84. Liu, H. et al. Mutant KRAS triggers functional reprogramming of tumor-associated macrophages in colorectal cancer. *Signal Transduct. Target Ther.* **6**, 144 (2021).
85. Berghoff, A. S. et al. Differential role of angiogenesis and tumour cell proliferation in brain metastases according to primary tumour type: analysis of 639 cases. *Neuropathol. Appl. Neurobiol.* **41**, e41–e55 (2015).
86. Ortiz-Prado, E., Dunn, J. F., Vasconez, J., Castillo, D. & Viscor, G. Partial pressure of oxygen in the human body: a general review. *Am. J. Blood Res* **9**, 1–14 (2019).
87. Di Micco, R. et al. Oncogene-induced senescence is a DNA damage response triggered by DNA hyper-replication. *Nature* **444**, 638–642 (2006).
88. Gilad, O. et al. Combining ATR suppression with oncogenic Ras synergistically increases genomic instability, causing synthetic lethality or tumorigenesis in a dosage-dependent manner. *Cancer Res* **70**, 9693–9702 (2010).
89. Momand, J., Zambetti, G. P., Olson, D. C., George, D. & Levine, A. J. The mdm-2 oncogene product forms a complex with the p53 protein and inhibits p53-mediated transactivation. *Cell* **69**, 1237–1245 (1992).
90. Matsushime, H. et al. Identification and properties of an atypical catalytic subunit (p34PSK-J3/cdk4) for mammalian D type G1 cyclins. *Cell* **71**, 323–334 (1992).
91. Konishi, H. et al. Mutation of a single allele of the cancer susceptibility gene BRCA1 leads to genomic instability in human breast epithelial cells. *Proc. Natl Acad. Sci. USA* **108**, 17773–17778 (2011).
92. Shrestha, K. S., Aska, E. M., Tuominen, M. M. & Kauppi, L. Tissue-specific reduction in MLH1 expression induces microsatellite instability in intestine of Mlh1(+/-) mice. *DNA Repair (Amst.)* **106**, 103178 (2021).
93. Liu, M. et al. HOXD9 promote epithelial-mesenchymal transition and metastasis in colorectal carcinoma. *Cancer Med* **9**, 3932–3943 (2020).
94. Song, C. & Zhou, C. HOXA10 mediates epithelial-mesenchymal transition to promote gastric cancer metastasis partly via modulation of TGF β 2/Smad/METTL3 signaling axis. *J. Exp. Clin. Cancer Res* **40**, 62 (2021).
95. Yin, J. & Guo, Y. HOXD13 promotes the malignant progression of colon cancer by upregulating PTPRN2. *Cancer Med* **10**, 5524–5533 (2021).
96. Wisdom, R., Johnson, R. S. & Moore, C. c-Jun regulates cell cycle progression and apoptosis by distinct mechanisms. *EMBO J.* **18**, 188–197 (1999).
97. Mechta, F., Lallemand, D., Pfarr, C. M. & Yaniv, M. Transformation by ras modifies AP1 composition and activity. *Oncogene* **14**, 837–847 (1997).
98. Cejas, P. et al. KRAS mutations in primary colorectal cancer tumors and related metastases: a potential role in prediction of lung metastasis. *PLoS One* **4**, e8199 (2009).
99. Li, C. et al. Integrated Omics of Metastatic Colorectal Cancer. *Cancer Cell* **38**, 734–747.e739 (2020).
100. Ortmann, C. A. et al. Effect of mutation order on myeloproliferative neoplasms. *N. Engl. J. Med* **372**, 601–612 (2015).
101. Kallioniemi, O. P. et al. Optimizing comparative genomic hybridization for analysis of DNA sequence copy number changes in solid tumors. *Genes Chromosomes Cancer* **10**, 231–243 (1994).
102. Cerami, E. et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2**, 401–404 (2012).
103. Gao, J. et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal* **6**, pl1 (2013).
104. Chakravarty, D. et al. OncoKB: a precision oncology knowledge base. *JCO Precis. Oncol.* **2017** <https://doi.org/10.1200/po.17.00011> (2017).
105. Spurr, L. F. et al. Quantification of aneuploidy in targeted sequencing data using ASCETS. *Bioinformatics* **37**, 2461–2463 (2021).
106. 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).
107. Reich, M. et al. GenePattern 2.0. *Nat. Genet.* **38**, 500–501 (2006).
108. Niu, B. et al. MSIsensor: microsatellite instability detection using paired tumor-normal sequence data. *Bioinformatics* **30**, 1015–1016 (2014).
109. Grossman, R. L. et al. Toward a shared vision for cancer genomic data. *N. Engl. J. Med* **375**, 1109–1112 (2016).
110. Shih, J. et al. Cancer aneuploidies are shaped primarily by effects on tumour fitness. *Nature* **619**, 793–800 (2023).
111. Ihaka, R. & Gentleman, R. R: a language for data analysis and graphics. *J. Computational Graph. Stat.* **5**, 299–314 (1996).
112. Griffith, D. M., Veech, J. A. & Marsh, C. J. cooccur: probabilistic species co-occurrence analysis in R. *J. Stat. Softw.* **69**, 1–17 (2016).
113. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* **57**, 289–300 (1995).
114. Hamilton, N. E. & Ferry, M. ggtern: ternary diagrams using ggplot2. *J. Stat. Softw.* **87**, 1–17 (2018).
115. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
116. Li, Y., Ge, X., Peng, F., Li, W. & Li, J. J. Exaggerated false positives by popular differential expression methods when analyzing human population samples. *Genome Biol.* **23**, 79 (2022).
117. Hoadley, K. A. et al. Cell-of-origin patterns dominate the molecular classification of 10,000 tumors from 33 types of cancer. *Cell* **173**, 291–304.e296 (2018).

118. Kuleshov, M. V. et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* **44**, W90–W97 (2016).
119. Ashburner, M. et al. Gene ontology: tool for the unification of biology. the gene ontology consortium. *Nat. Genet* **25**, 25–29 (2000).
120. Gene Ontology Consortium et al. The gene ontology knowledgebase in 2023. *Genetics* **224**, iyad031 (2023).
121. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
122. Liberzon, A. et al. Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–1740 (2011).
123. Mootha, V. K. et al. PGC-1 α -responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat. Genet* **34**, 267–273 (2003).
124. Barbie, D. A. et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* **462**, 108–112 (2009).
125. Kolberg, L. et al. g:Profiler-interoperable web service for functional enrichment analysis and gene identifier mapping (2023 update). *Nucleic Acids Res* **51**, W207–W212 (2023).
126. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28**, 27–30 (2000).
127. Gillespie, M. et al. The reactome pathway knowledgebase 2022. *Nucleic Acids Res* **50**, D687–D692 (2022).
128. von Heydebreck, A., Gunawan, B. & Füzesi, L. Maximum likelihood estimation of oncogenetic tree models. *Biostatistics* **5**, 545–556 (2004).
129. Felsenstein, J. PHYLIP - phylogeny inference package. *Cladistics* **5**, 164–166 (1989).
130. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67**, 1–48 (2015).

Acknowledgements

We thank Christina Enders and Cynthia Bunker for technical support, and thank the researchers at the Memorial Sloan Kettering Cancer Center, New York, the TCGA network and the Functional Proteomics Reverse Phase Protein Array Core at the University of Texas MD Anderson Cancer Center for making their data available. This study was supported by the guest professorship program of the University of Augsburg and the professorship of the Faculty of Medicine at the University of Augsburg to MMG.

Author contributions

Conceptualization: M.M.G., L.F., B.S. Methodology: M.M.G, B.S. Formal analysis: M.M.G., B.S. Investigation: M.M.G., B.S. Resources: M.M.G.,

B.G., A.G., B.C.D., C.S., L.F., T.L., B.S. Data curation: B.G., J.S.G., L.F. Writing – Original Draft: M.M.G. Writing – Review & Editing: all authors. Visualization: M.M.G., B.S. Supervision: M.M.G., L.F., B.S. Project administration: M.M.G. Funding acquisition: M.M.G.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

The authors are not aware of any conflict of interest related to this study.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-58413-1>.

Correspondence and requests for materials should be addressed to Mariola Monika Golas or Bjoern Sander.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

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

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025