Science Advances
AAAS

advances.sciencemag.org/cgi/content/full/6/24/eaay9691/DC1

Supplementary Materials for

The clonal evolution of metastatic colorectal cancer

Ha X. Dang, Bradley A. Krasnick, Brian S. White, Julie G. Grossman, Matthew S. Strand, Jin Zhang, Christopher R. Cabanski, Christopher A. Miller, Robert S. Fulton, S. Peter Goedegebuure, Catrina C. Fronick, Malachi Griffith, David E. Larson, Brian D. Goetz, Jason R. Walker, William G. Hawkins, Steven M. Strasberg, David C. Linehan, Kian H. Lim, A. Craig Lockhart, Elaine R. Mardis, Richard K. Wilson, Timothy J. Ley, Christopher A. Maher, Ryan C. Fields*

*Corresponding author. Email: rcfields@wustl.edu

Published 10 June 2020, *Sci. Adv.* **6**, eaay9691 (2020)
DOI: 10.1126/sciadv.aay9691

The PDF file includes:

Other Supplementary Material for this manuscript includes the following:

(available at advances.sciencemag.org/cgi/content/full/6/24/eaay9691/DC1)

Supplementary Methods

Patient consent and tissue processing

Patients with a pathologically-confirmed diagnosis of adenocarcinoma of the colon or rectum were identified from the medical and surgical oncology clinics at the Alvin J. Siteman Comprehensive Cancer Center at Washington University School of Medicine, St. Louis, Missouri, USA. Under an IRB-approved protocol, patients were consented to a blood and tumor collection protocol. Fresh tumor specimens were procured from the operating room in cold RPMI medium supplemented with 1% penicillin, streptomycin, and amphotericin. Tissue was processed using scalpels and divided into pieces of 50-100 mg before flash freezing in liquid nitrogen. Nucleic acids were purified from frozen tumor sample pieces: total RNA was extracted using RNeasy Mini kits (Qiagen) following the manufacturer's protocol. RNA samples were treated with RNase free DNase (RQ1, Promega), and RNA samples were stored at -80^{o}C. Genomic DNA was extracted from both tissue samples and PBMC using the Wizard Genomic DNA Purification kit (Promega). Additionally, in several patients, the archived tumor specimens (in most cases, the previously resected primary tumors) in the form of formalin-fixed parafin-embeded (FFPE) were obtained from internal and external pathology departments under the same protocol. Tumor blocks of archived samples were sectioned by the Digestive Disease Research Core Center at our institution, and sections were transferred to the Tissue Procurement Core for H&E staining and isolation of nucleic acids. The H&E slides were evaluated by a trained Pathologist for tumor content. Tumor samples were isolated via 21 gauge punch or curl biopsies from any area of the tumor that contained >70% tumor with <20% necrosis that included both areas of the tumor center or leading edge/tumor front. Tumor tissues, collected through punches or slide scraping were subjected to deparaffinization and subsequent nucleic acid extraction using QIAamp DNA kits (Mini or Micro, dependent on the amount of tissue) and High Pure RNA Paraffin kits (Roche) following the manufacturer's instructions. All patients had clinical genomic sequencing performed on one of their clinical specimens by a CLIA-certified, commercial vendor as part of their standard-of-care treatment. Demographic, pre-treatment, treatment, follow-up, and other clinical metadata were prospectively obtained from the medical records.

Procedure for patient derived xenografts

A portion of the resected, fresh tumor specimens described above were used for PDX creation. Multiple 1-2 mm pieces from different regions of the tumor were dissociated into a single cell suspension using mechanical dissociation and enzymatic digest *(32)*. For single cell suspension creation, tumors were dissociated sharply into multiple 5-10 pieces (1-2 mm each), with all pieces placed in enzymatic digest (RPMI, Gibco by Fisher Scientific, St. Louis, MO; DNAse I, Sigma, St. Louis, MO, USA; HEPES, Corning, Corning, NY, USA; Hyaluronidase, Sigma). Pieces in enzymatic digest (total volume ~6 ml) were placed into a gentleMACS C Tube® (Miltenyi Biotec, Bergish Gladbach, Germany), followed by 1 minute of mechanical digestion (GentleMACS Dissociator®, Miltenyi Biotec). The tube was placed on a shaker at 37^{o}C for ~15 minutes. Suspensions were then quenched in complete media (RPMI + 10% FBS, Gibco), and run through a 70 micron filter (Fisher Scientific). Cells were then counted and resuspended in 1:1 PBS and Matrigel® Matrix (Corning), for a volume of 50 µl per 1 million cells. This was then injected subcutaneously (50 µl per injection) into the right and/or left flank of recipient NOD/SCID(NOD.CB17−Prkdcscid/J) or Nude (nu/J) mice to produce a solid tumor xenograft. Between 4−24 weeks post inoculation (tumor size 1-2 cm), recipient mice (P1) were euthanized and the tumors were removed. A portion of each tumor was dissociated, as above, and used to generate second (P2), third (P3), and beyond xenografts.

Sequencing strategies for clonal evolution analysis

One primary region, all metastasis and all xenograft samples were selected for the discovery phase. In the discovery phase, the primary and metastasis samples were sequenced using both whole genome and deep exome sequencing, and the xenograft samples were sequenced using deep exome sequencing. Variants and copy numbers were called

for individual samples using WGS and exome data. Next, a hybrid targeted validation panel was designed to target all non-silent SNVs and indels, and silent SNVs in diploid heterozygous non-repetitive regions identified from WGS and exome data from individual samples (see targeted validation design and sequencing). Finally, all samples including the discovery samples and additional multi region primary samples were included in the targeted validation phase and only validated variants identified from targeted sequencing were used for clonal evolution analysis. This strategy guaranteed that all variants used for clonal evolution analysis were evenly covered in all samples.

Library construction and whole genome sequencing

Single indexed libraries were constructed with 50-250 ng of genomic DNA utilizing the KAPA HTP library prep kit (KAPA Biosystems) per sample. The samples were fragmented on the Covaris LE220 (Covaris) targeting a size range between 300-700 bp. PCR cycle optimization was performed to prevent over cycling during the enrichment PCR. Eight PCR reactions were amplified to enrich for proper adaptor ligated fragments. The final size selection of the library was achieved by running the enriched library fragments on a Caliper XT chip (Perkin Elmer). Depending on the sample quality, two to three fractions were collected per sample: 375 bp, 475 bp, and 675 bp. The concentration of each library fraction was accurately determined through qPCR (Kapa Biosystems) in order to generated cluster counts appropriate for the Illumina HiSeq 2000 platform (Illumina). 2 x 101 bp sequence data was generated per library. Sample identity was confirmed for each sample by comparing sequence data with the Illumina Human OmniExpress genotype array data.

Library construction and whole exome sequencing

Single indexed libraries were constructed with 50-250 ng of genomic DNA utilizing the KAPA HTP library prep kit (KAPA Biosystems) per sample. The samples were fragmented on the Covaris LE220 (Covaris) targeting 250 bp inserts. Two libraries were constructed per tumor/metastasis/xenograft and a single library was constructed for the normals. Nine libraries were pooled pre-capture generating a 5 µg library pool. Each library pool was hybridized with the SeqCap EZ Human Exome Kit v3.0 (Roche Nimblegen) that targets over 20,000 genes spanning ~64 Mb of the human genome. The libraries were hybridized for 72 hours at 47°C followed by stringent washing. Enriched ssDNA library fragments were amplified with KAPA HiFi HotStart polymerase and 200 nM primers prior to sequencing. The concentration of each captured library pool was accurately determined through qPCR according to the manufacturer's protocol (KAPA Biosystems) to produce cluster counts appropriate for the Illumina HiSeq2000 platform. 2x100 bp sequence data was generated per capture pool. Sample identity was confirmed for each sample by comparing sequence data with the Illumina Human OmniExpress genotype array data.

Targeted validation design and sequencing

Targeted validation design

A NimbleGen EZ-Seq Hybrid Capture was designed to target all non-coding silent SNVs in copy-neutral non-repetitive regions plus all non-silent SNVs and small indels in both copy-neutral and copy-altered regions that were called by at least one caller using either WGS or exome data in the discovery phase. Additionally, the exon regions of 152 cancer genes frequently mutated in CRC *(2)* and pan-cancers *(35)* and the break points of selected gene fusions were also targeted. We also spiked-in IDT Lockdown probes to target 309 selected variants in regions of subclonal copy number variations affecting *APC*, *PTEN*, and *TCF7L2* identified from patient CRC8 and 92 gene fusion breakpoints. This resulted in a targeted capture panel that cover ~12.6 Mbp throughout the genome (104,404 specific sites and 152 genes).

Library construction and sequencing

Manual dual indexed libraries were constructed with 20-100 ng of genomic DNA utilizing the Swift Accel NGS 2S PCR Free Kit (Swift Biosciences). The samples were fragmented on the Covaris LE220 (Covaris) targeting 250 bp inserts. Libraries were amplified with KAPA HiFi HotStart Ready Mix. (KAPA Biosystems). FFPE, Fresh Frozen Tissue, and blood libraries were pooled separately for capture. Each library pool consisted of 9-11 libraries. The library pools were hybridized with a custom Nimblegen targeted probe set (Roche) and IDT Lockdown probes (Integrated DNA Technologies). The IDT probes were added at an equal molar equivalent to the probe concentration of the Nimblegen probes, thus representing the targeted regions in similar quantity. The libraries were hybridized for 72 hours at 47°C followed by stringent washing. Enriched ssDNA library fragments were amplified with KAPA HiFi HotStart and 200 nM primers prior to sequencing. The concentration of each captured library pool was accurately determined through qPCR according to the manufacturer's protocol (KAPA Biosystems) to produce cluster counts appropriate for the Illumina HiSeq2500 1T platform. 2x125 bp reads were generated targeting 500-800x mean depth of coverage per sample.

Genome alignment

We obtained an average of ~830 million pair-ended reads for whole genome, ~136 million pair-ended reads for whole exome, and ~65 million pair-ended reads for targeted sequencing. Reads were aligned to the human genome assembly version GRCh37 using BWA v0.5.9 *(36)* (parameters -t 4 -q 5). The average alignment rates were ~95%, ~98%, and ~99% for whole genome, exome, and targeted sequencing. Aligned reads were marked for duplicated using Picard v1.46 (https://github.com/broadinstitute/picard). The mean deduplicated coverage for WGS was ~67x for the primary, metastasis, and xenograft samples, and ~36x for the normal samples. The mean deduplicated coverage for whole exome was ~215x for the primary, metastasis, and xenograft samples, and ~76x for the normal samples. The mean deduplicated coverage for targeted sequencing was ~251x for the primary, metastasis, and xenograft samples, and 243x for the normal samples.

Mutation detection pipeline

We used a production variant calling pipeline in the Genome Modeling System (GMS) developed at the McDonnell Genome Institute *(37)* (https://github.com/genome/gms) that incorporated multiple variant callers. For single nucleotide variant calling, the pipeline included variants reported by Samtools r963 (parameters: -A -B), SomaticSniper v1.0.2 (https://github.com/genome/somatic-sniper, parameters: -F vcf -q 1 -Q 15), VarScan2 v2.2.6 (http://dkoboldt.github.io/varscan/, default parameters), and Strelka v0.4.6.2 (https://github.com/Illumina/strelka, parameters: isSkipDepthFilters = 0). For insertions and deletions, the pipeline included variants reported by GATK r5336 (https://gatk.broadinstitute.org/), Pindel v0.5 (https://github.com/genome/pindel, parameters --variant-freq-cutoff=0.2), VarScan2 v2.2.6 (default parameters), Strelka v0.4.6.2 (parameters isSkipDepthFilters = 0), and Mutect 1.1.4 (https://software.broadinstitute.org/cancer/cga/mutect, parameters: --number-of-chunks 50). SNVs and indels were filtered to remove false positive using the GMS false-positive filter v1 (parameters: -min-base-quality 15) and somatic-score-mapping-quality (parameters: --min-mapping-quality 40 --min-somatic-score 40). Variants matching those found in dbSNP (https://www.ncbi.nlm.nih.gov/snp/) with a global minor allele frequency (GMAF) of >0.1% were flagged as germline variants and removed. Additionally, noncoding variants in blacklisted and low mappability regions (http://hgdownload.soe.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeMapability/, mappability score < 0.75) were removed. We also performed a panel of normal filtering to remove variants found with at least 5 reads supported in > 5% of ~400 unrelated whole genome normal samples.

In the validation phase, all non-coding silent SNVs in copy-neutral regions plus all non-silent SNVs and small indels in both copy-neutral and copy-altered regions that were called by at least one caller were chosen for validation using targeted sequencing (see targeted panel design). Upon validation, SNVs or indels were retained as somatic events only when they are called by at least three callers in either targeted sequencing or WGS/exome sequencing and

passed all above filters. Only variants called by at least three callers in targeted sequencing were used in subsequent clonal evolution analysis.

Mutation calling in xenografts

In the discovery phase, the same mutation calling pipeline for patient tumor was applied to the xenograft, additionally followed by a filter to remove variants from mouse. This filter used the UCSC liftOver tool (https://genome.ucsc.edu/cgi-bin/hgLiftOver) to identify mouse genome regions homologous to human genome and removed any variants that matched the mouse homologous genome sequence. In the validation phase, only variants confirmed in patient tumors were kept for further analysis.

Copy number analysis

CopyCat v1.6.9 (https://github.com/chrisamiller/copycat) was used for copy number analysis of the whole genome data using read depth per 1kb window generated using bam-window v0.7 (https://github.com/genome-vendor/bam-window). VarScan2 was used for copy number calling from whole exome data. CopywriteR *(38)* was used for copy number calling from targeted sequencing data. Loss of heterozygosity was called using VarScan2. All copy number calls from CopyCat, VarScan2, and CopywriteR were converted to absolute copy number for subsequent clonal evolution analysis using SciClone and ClonEvol. Regions with absolute copy number >2.75 were defined as a gain event and regions with absolute copy number <1.5 were defined as a loss event.

Manual review of somatic mutations and copy number variations

Nonsilent mutations hitting cancer genes were manually reviewed by investigating the alignment using IGV Genome Browser to flag for false positive SNV and indel calls and rescue calls presented at low frequency in a sample with strong evidence in another sample from the same patient. Copy number variation calls hitting cancer genes were also manually reviewed to recover calls for individual samples based on calls from other samples from the same patient.

Clonal evolution analysis

Identification of subclonal populations

Heterozygous somatic SNVs and small indels in non-repetitive regions of autosomal chromosomes validated in targeted sequencing and with at least 50x depth of combined coverage were used for clonal evolution analysis (except for patients CRC5 and CRC7 the depth cutoff was 30x due to lower number of variants with >50x coverage). We found that reducing the depth cutoff did not affect variant clustering results but improved clonal evolution inference and provided better estimate of clonal cellular fractions. First, the variant allele frequency (VAF) was calculated based on aggregated read counts generated using bam-readcounts v0.4 (https://github.com/genome/bam-readcount) from whole genome, whole exome, and targeted sequencing data. Read counts were aggregated for all variants across all samples from the same patients to ensure that variant robustly detected in one sample but presented at low frequency in another sample still receive appropriate read counts. Next, to identify subpopulations of cancer cells, variants were first clustered using sciClone *(33)* (https://github.com/genome/sciclone) across all samples from the same patient including primary regions, metastasis, and PDX samples with the copy number margin of 0.75 utilizing all copy number values from WGS, exome, and targeted validation analyses. This approach of clustering variants based on their frequency has been widely used in clonal evolution analysis *(39)*. It is likely that some events such as whole genome duplication (masked from total copy number analysis) could result in dispersed bimodal VAFs and over-clustering. SciClone could mitigate this effect as its variational Bayesian beta mixture model could fit clusters of variants with dispersed VAFs. Additionally, all clustering results were manually investigated and clusters that had similar presence across samples but demonstrated dispersed VAFs were merged. This hybrid semi-automatic approach could mitigate the potential effect of whole genome duplication on our clonal evolution analysis.

Inferring consensus clonal evolution trees and clonal admixtures

Upon clustering of variants, the variant clusters and VAF of the variants were used in clonal evolution analysis using ClonEvol *(34)* (https://github.com/hdng/clonevol) to infer the consensus clonal evolution trees for each patient, estimate tumor purity and clonal cellular fractions and admixtures for individual samples. CRC11 was excluded from tree construction due to low purity primary tumor sample. Upon the consensus clonal evolution trees were constructed, multi regions from the primary tumors were combined using ClonEvol to represent the primary tumor. In brief, read counts for individual variants were summed from those of the variants in all primary regions and aggregated cellular fractions of individual variant clusters and clones were calculated.

Mapping cancer gene events onto the clonal evolution trees

Putative cancer genes were defined as genes frequently mutated in CRC *(2)* and other cancer in COSMIC database (https://cancer.sanger.ac.uk/cosmic) and genes associated with CRC curated from MalaCards database (https://www.malacards.org/). Heterozygous non-silent mutations in cancer genes were included in the clustering analysis by sciClone. Other non-silent mutations and copy number alteration in cancer genes were mapped to the clonal marker variant clusters by matching the presence/absence status of the mutations and copy number events with that of the clusters across samples. Non-silent mutations and copy number alterations targeting cancer genes were defined as potential driver events.

Visualization

Visualizations including clonal evolution trees, fishplots, clonal admixtures were generated using ClonEvol *(34)* and FishPlot *(40)* (https://github.com/chrisamiller/fishplot) packages. Other figures were generated using R statistical software package (https://www.r-project.org/) and ggplot2 (https://ggplot2.tidyverse.org/).

Supplementary Figures

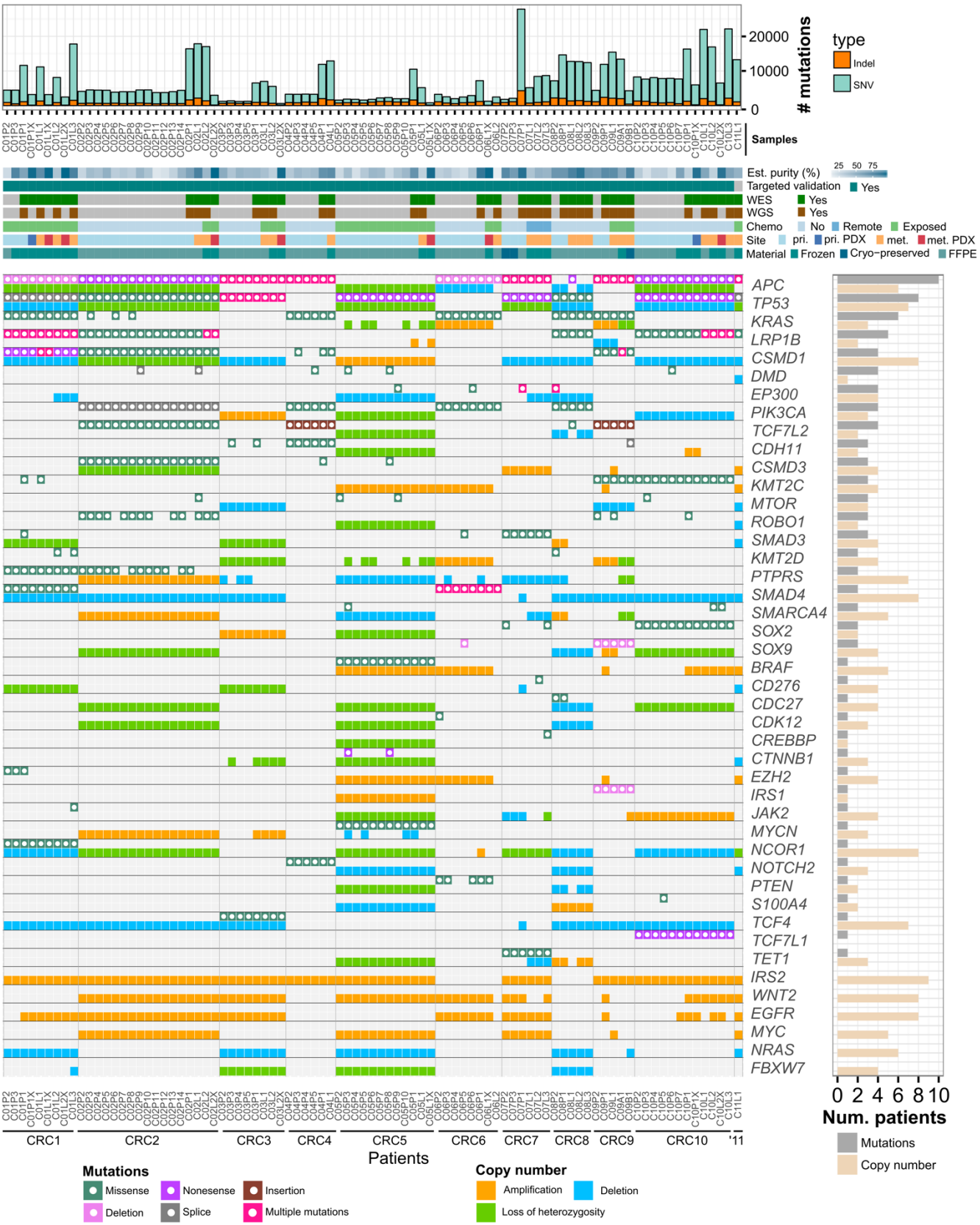

Fig. S1. Frequently mutated colorectal cancer genes in metastatic colorectal cancer patients. Each gene is presented with two rows: top – SNVs and small indels; bottom – copy number alteration. Pri. – primary, met. – metastasis

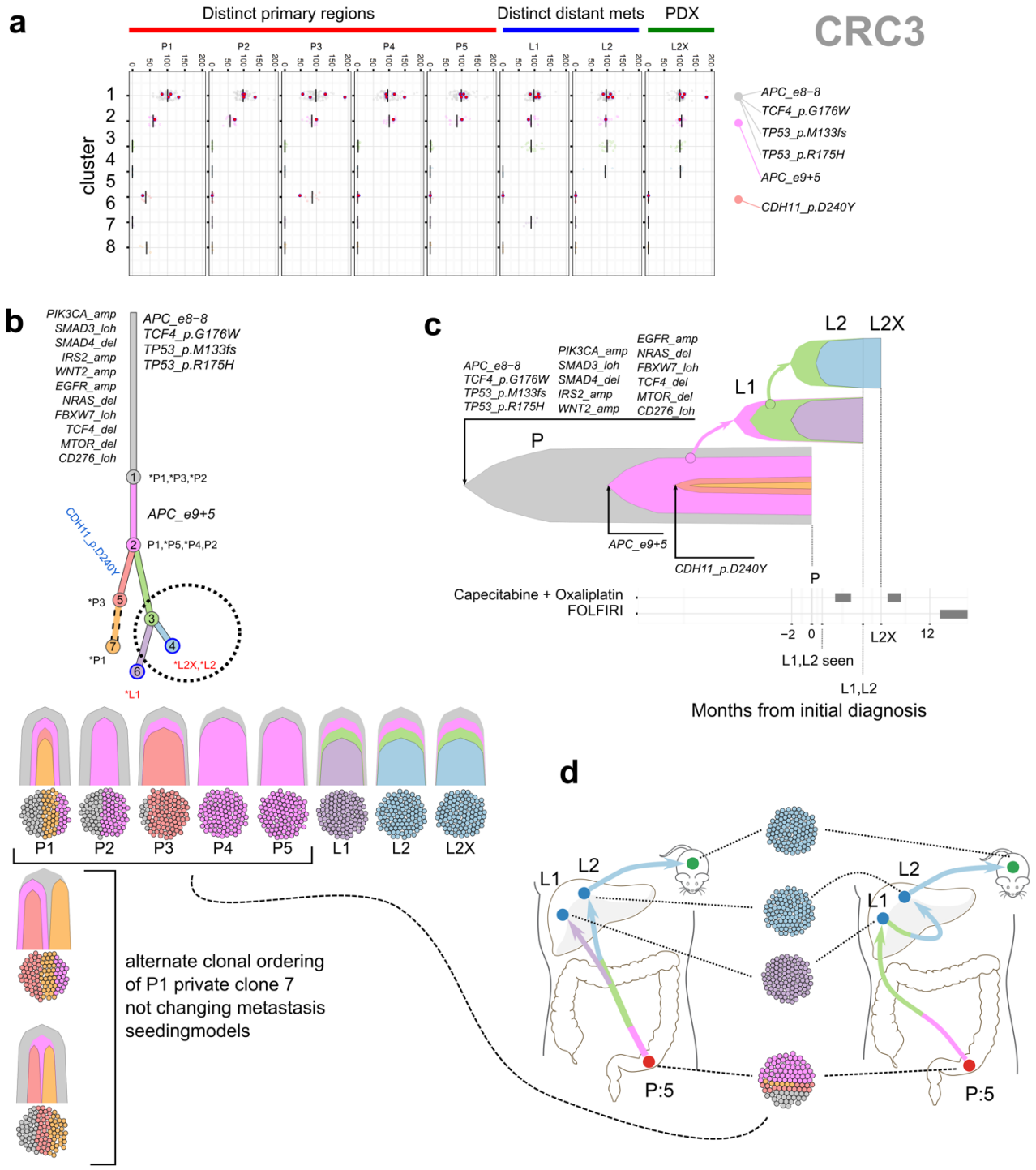

Fig. S2. Clonal evolution of patient CRC3. (a) Clustering of variants with y-axis showing cluster identity and x-axis showing estimated purity-corrected cancer cell fraction estimated by ClonEvol. Potential driver variants are highlighted as red dots whose details are on the far right of the plot; **(b)** Predicted clonal evolution tree (top) and subclonal population of individual samples including multi-region primary samples (bottom). Dotted circle highlights the branch suggestive of metastasis seeding by metastasis; **(c)** Fishplot interpretation of evolution from primary (multi-regions merged) to metastasis with clinical timeline. The right-end of each fishplot represents the time the primary tumor/metastasis was resected. Treatment courses are presented as boxes underneath the time axis. All primary regions are combined and aggregated cellular fractions of clones are presented. Metastasis seedings are indicated by arrows connecting primary and metastasis tumors. L2X engraftment refers to time that L2 was transferred to the mouse to establish a PDX. Cancer genes whose alterations are clonal markers of a clone are indicated with arrows pointing to the tips of the corresponding clone; **(d)** Anatomic representation of tumor location and metastatic progression. In this patient, 8 clones were detected. The clonal evolution of the primary tumor P (with 5 regions sequenced, P1-P5) was predicted to be linear. Liver metastasis L1 was seeded by subclone 2 (pink) from the primary, and liver metastasis L2 was seeded by subclone 3 (light green) whose mutations were not seen in any primary regions but present in L1. This raises a possibility that metastasis L2 is seeded by metastasis L1 (c, d - right), however, subclone 3 was not present in any samples and therefore a model where an unknown site evolved from the primary and seeded the two metastases L1 and L2 is also possible (d - left). Lastly, the xenograft L2X maintained the only clone (4, light blue) from metastasis L2.

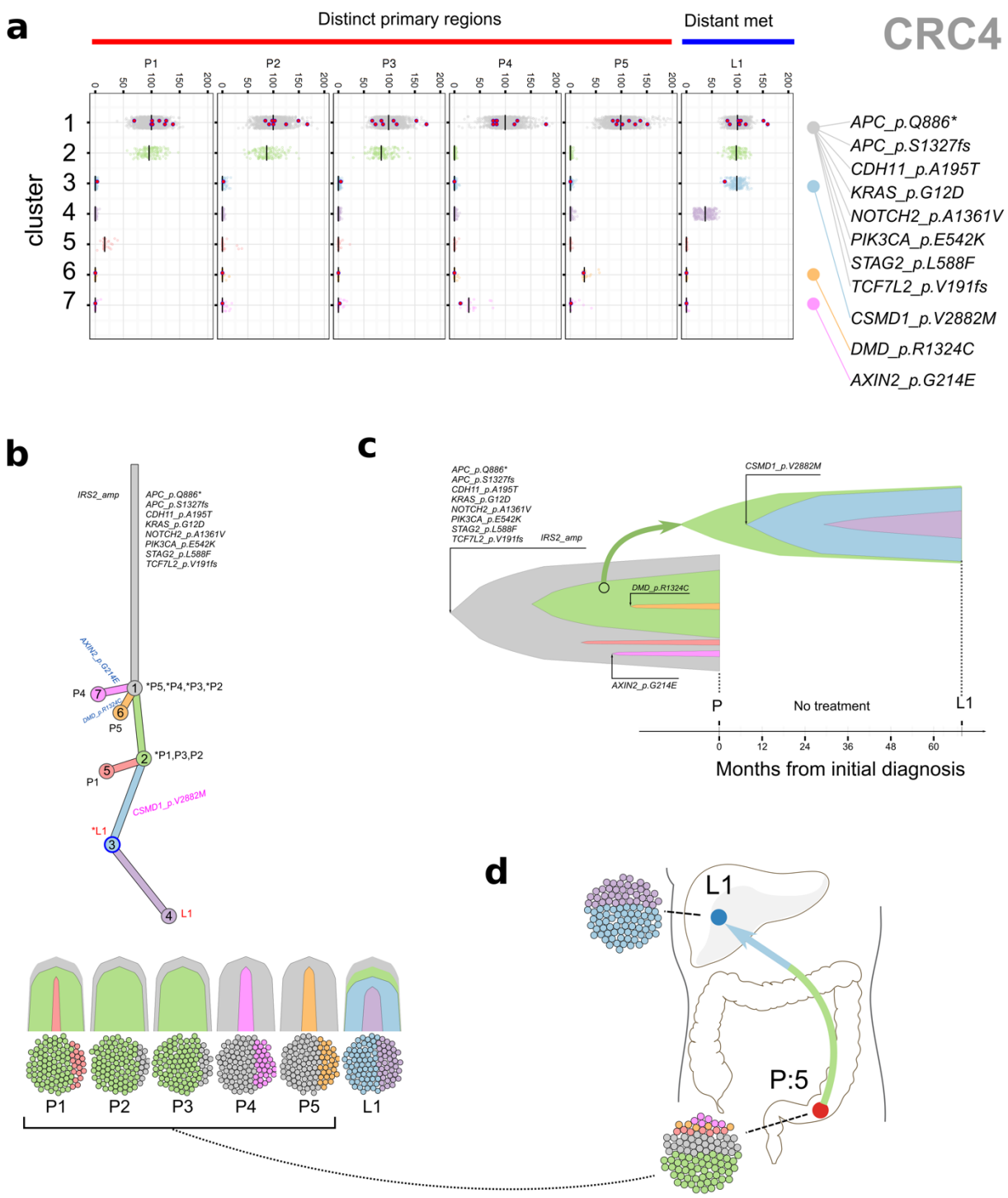

Fig. S3. Clonal evolution of patient CRC4. Panels a-d are organized similar to Fig. S2. In this patient, 7 clones were identified. Metastasis L1 was seeded by subclone 2 (light green) that was present in 3/5 primary regions.

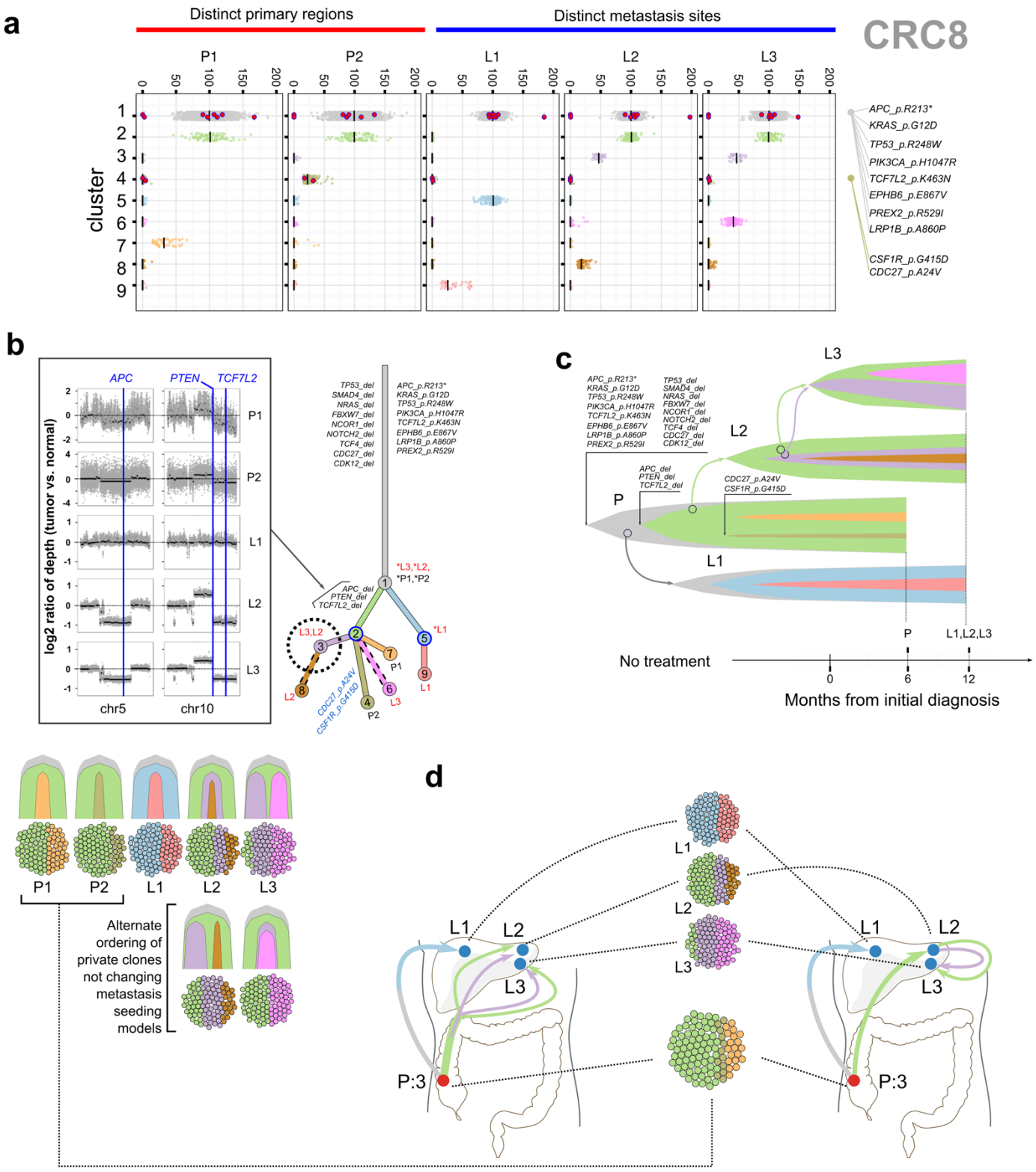

Fig. S4. Clonal evolution of patient CRC8. Panels a-d are organized similar to Fig. S2. In this patient, 9 clones were identified. A branching evolution was predicted for the primary tumor. Liver metastasis L1 was seeded by clone 1 (light gray) from the primary. Liver metastasis L2 was seeded by subclone 2 (light green) from the primary tumor. Panel b (left) shows loss of *APC, PTEN, TCF7L2* in two liver metastases L2 and L3 but not detected from liver metastasis L1. These events were also detected in the primary regions (P1, P2) with lower log2 ratio suggesting they are subclonal events. Subclone 3 (purple) and its clonal marker variants were found to present at subclonal levels in both L2 and L3 but absent from the primary tumor. This raises a distinct possibility of a polyclonal model of metastasis seeding another metastasis and a model where L2 seeded L3 (c, d - right) is presented. A model where an unknown site evolved from the primary and seeded both L2 and L3 was also presented (d - left).

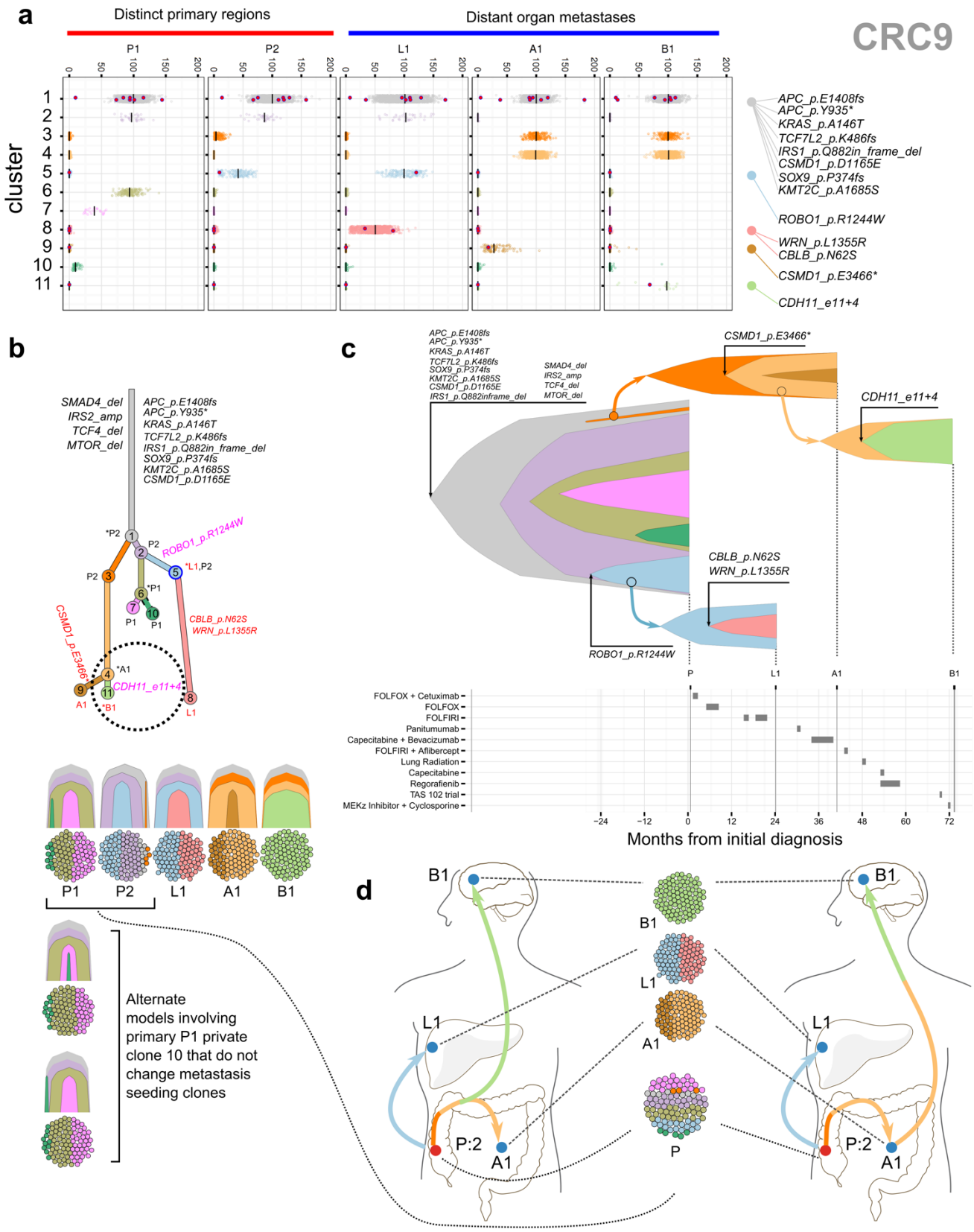

Fig. S5. Clonal evolution of patient CRC9. Panels a-d are organized similar to Fig. S2. In this patient, 11 clones were identified. A branching evolution was predicted for the primary tumor. Liver metastasis L1 was seeded by clone 5 (light blue) from the primary. Abdominal wall metastasis A1 was seeded by rare subclone 3 (orange) from the primary tumor. Subclone 4 (yellow) was predicted to seed brain metatasis B1 and its clonal marker variants was found in A1 but absent from the primary tumor. This raises a distinct possibility that abdominal wall metastasis A1 seeded brain metastasis B1 (c, d - right). A model where an unknown site evolved from the primary and seeded both A1 and B1 was also possible (d - left).

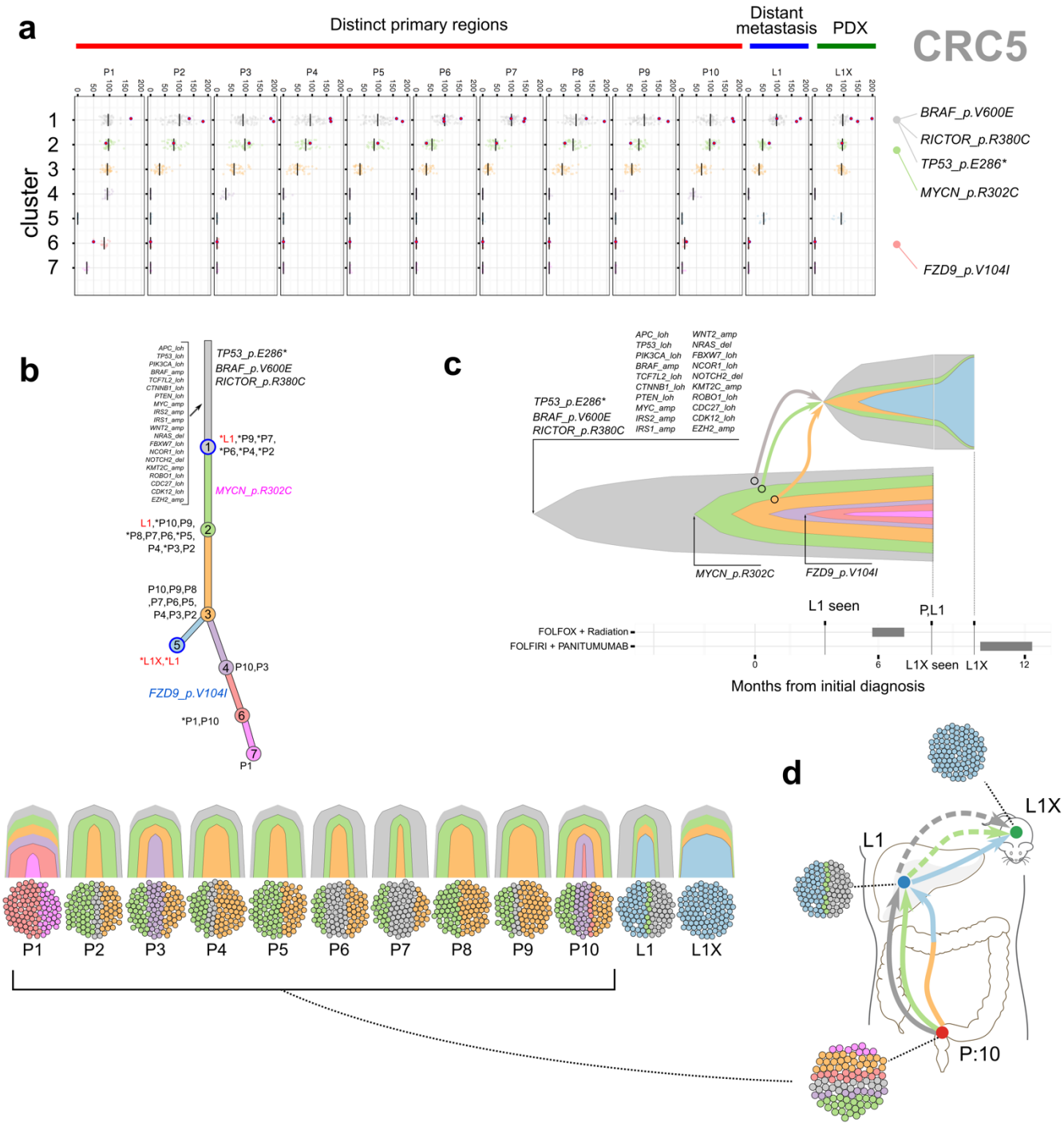

Fig. S6. Clonal evolution of patient CRC5. Panels a-d are organized similar to Fig. S2. In this patient, 7 clones were identified. A linear evolution was predicted for the primary tumor. Liver metastasis L1 was seeded by three clones (1/gray, 2/light green, 3/light orange) from the primary following a polyclonal model. Only 1/3 of clones (clone 5, light blue) from liver metastasis L1 was recapitulated in the xenograft L1X.

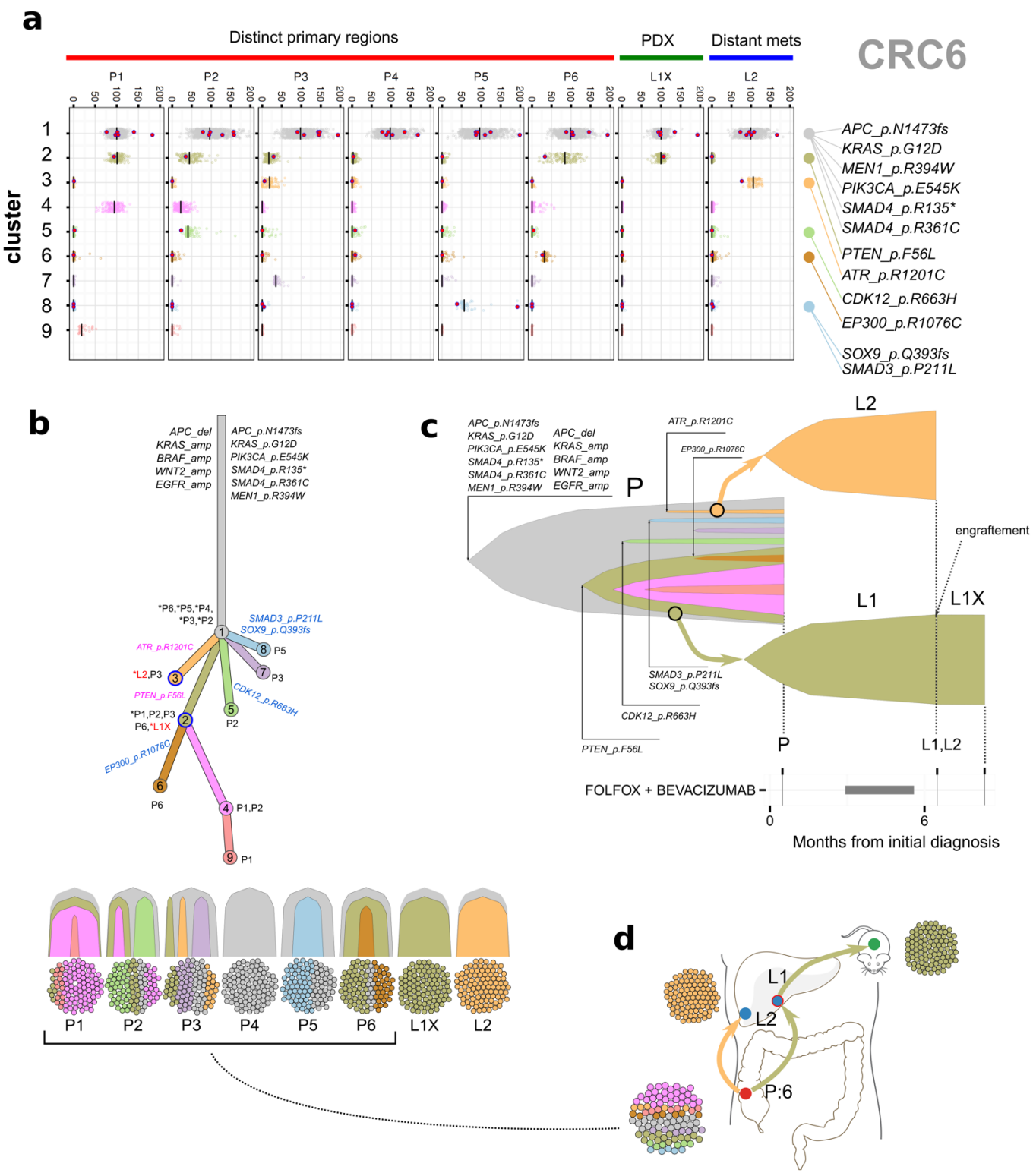

Fig. S7. Clonal evolution of patient CRC6. Panels a-d are organized similar to Fig. S2. In this patient, 9 clones were identified. A branching evolution was predicted for the primary tumor. Liver metastasis L2 was seeded by rare subclone 3 (light orange) that was found in 1/6 primary regions. Liver metastasis L1 had low quality and was not analyzed but the PDX derived from L1 was analyzed and it was predicted that L1 was seeded by subclone 2 (dark green). It is unknown if other clones existed in L1 and the model presented in panel c is based on an assumption that clone 2 was the only clone in L1.

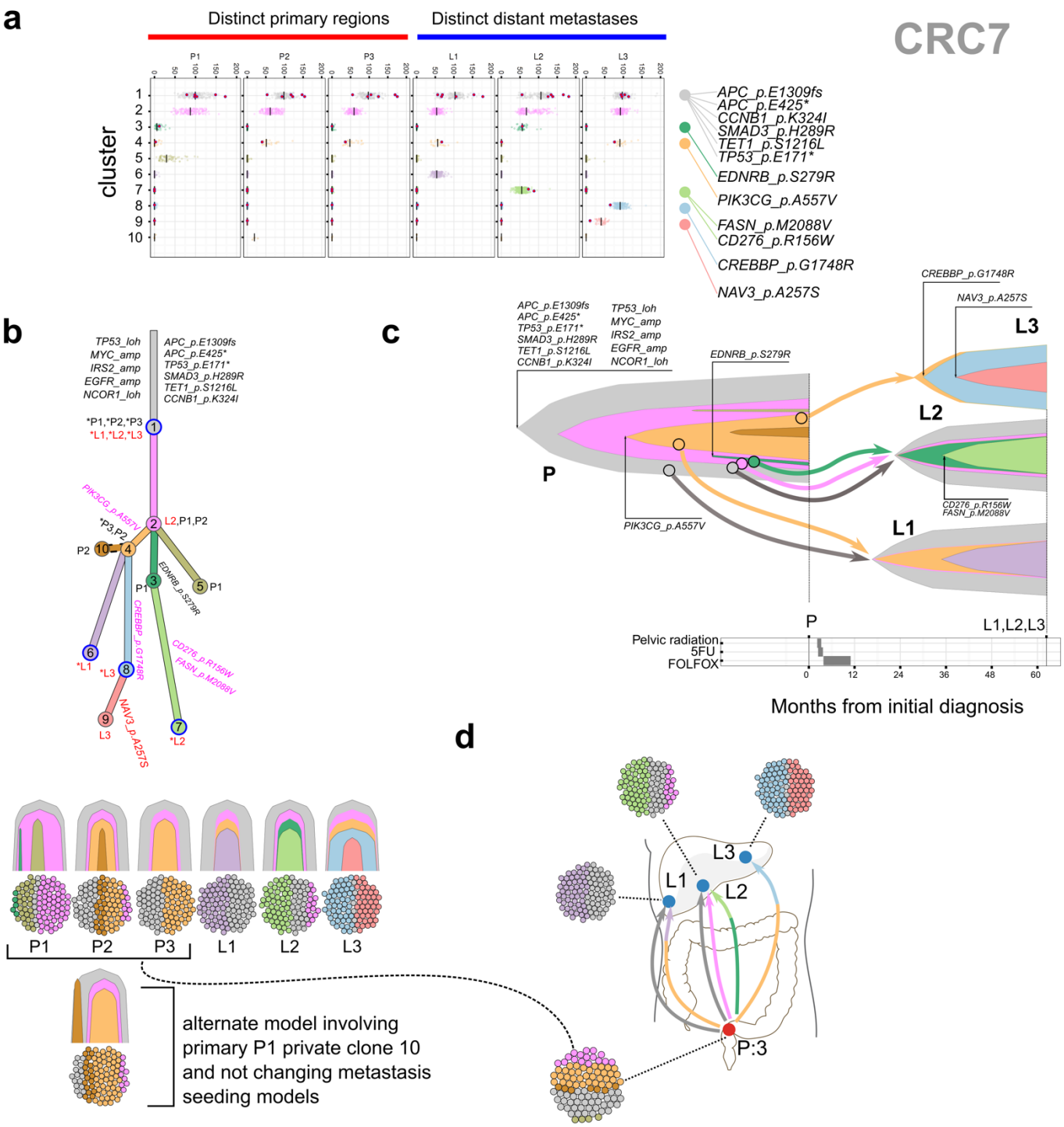

Fig. S8. Clonal evolution of patient CRC7. Panels a-d are organized similar to Fig. S2. In this patient, 10 clones were identified. A branching evolution was predicted for the primary tumor. Liver metastasis L1 was seeded by two clones (1/gray, 4/light orange) from the primary following a polyclonal model. Similarly, liver metastasis L2 was seeded by three clones from the primary tumor (1/gray, 2/pink, 3/dark green). Metastasis L3 was seeded by subclone 4 (light orange) from the primary tumor.

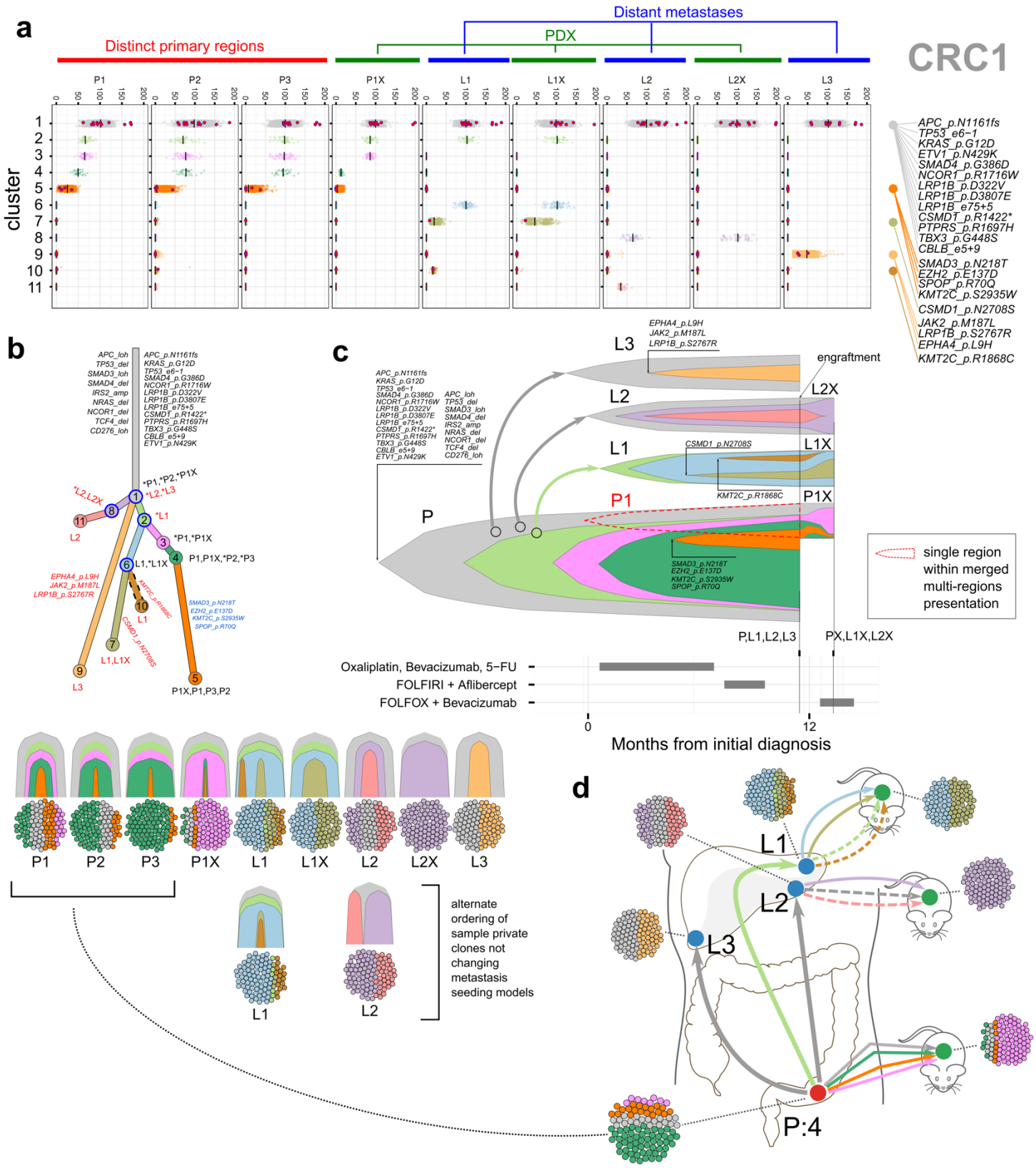

Fig. S9. Clonal evolution of patient CRC1. Panels a-d are organized similar to Fig. S2. In this patient, 11 clones were identified. A linear evolution was predicted for the primary tumor. Liver metastasis L1 was seeded by subclone 2 (light green). Liver metastases L2 and L3 were seeded by clone 1 (gray). Xenograft P1X derived from primary region P1 maintained all clones from its parental tumor. Xenografts L1X and L2X both lost clones from their parental tumors.

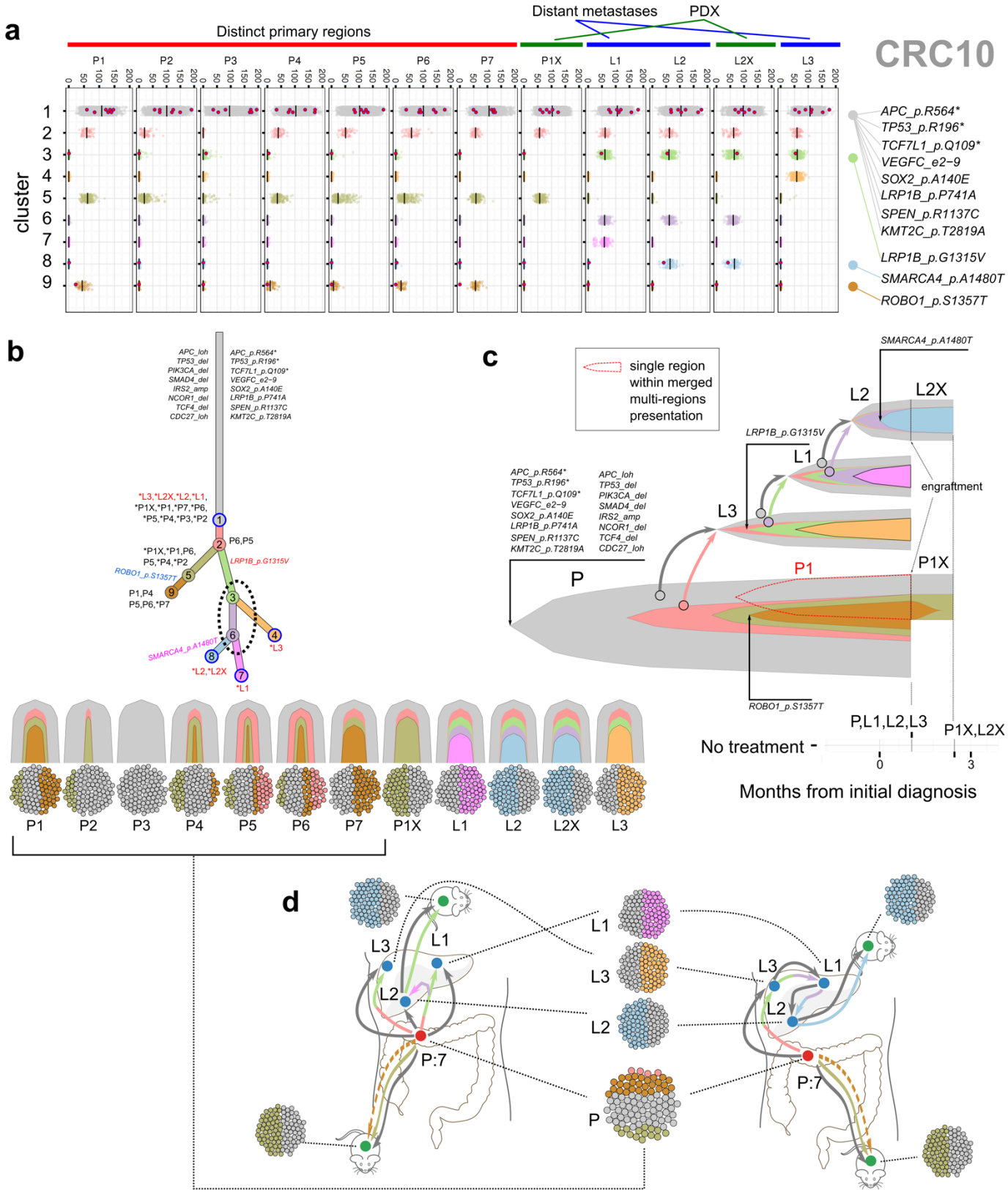

Fig. S10. Clonal evolution of patient CRC10. Panels a-d are organized similar to Fig. S2. In this patient, 9 clones were identified. A linear evolution was predicted for the primary tumor. Liver metastasis L3 was seeded by two clones (1/gray, 2/salmon) from the primary. Subclone 3 (light green) clonal marker variants were present at subclonal levels in both L1 and L3 but absent from the primary tumor. This raises a distinct possibility of a polyclonal model of metastasis seeding between L1 and L3. Similarly, subclone 6 (purple) clonal marker variants were found in only L2 and L2. Together, this raises a possibility of a metastasis cascade where L3 seeded L1, and L1 later seeded L2 (c, d - right). A model where an unknown site evolved from the primary and seeded L1 and L2 was also presented (d - left). Lastly, both xenografts P1X and L1X lost one clone compared with their parental tumor.

REFERENCES AND NOTES

1. E. R. Fearon, B. Vogelstein, A genetic model for colorectal tumorigenesis. *Cell* **61**, 759–767 (1990).

2. Cancer Genome Atlas Network, Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330–337 (2012).

3. A. R. Brannon, E. Vakiani, B. E. Sylvester, S. N. Scott, G. McDermott, R. H. Shah, K. Kania, A. Viale, D. M. Oschwald, V. Vacic, A.-K. Emde, A. Cercek, R. Yaeger, N. E. Kemeny, L. B. Saltz, J. Shia, M. I. D'Angelica, M. R. Weiser, D. B. Solit, M. F. Berger, Comparative sequencing analysis reveals high genomic concordance between matched primary and metastatic colorectal cancer lesions. *Genome Biol.* **15**, 454 (2014).

4. K. Naxerova, J. G. Reiter, E. Brachtel, J. K. Lennerz, M. van de Wetering, A. Rowan, T. Cai, H. Clevers, C. Swanton, M. A. Nowak, S. J. Elledge, R. K. Jain, Origins of lymphatic and distant metastases in human colorectal cancer. *Science* **357**, 55–60 (2017).

5. Q. Wei, Z. Ye, X. Zhong, L. Li, C. Wang, R. E. Myers, J. P. Palazzo, D. Fortuna, A. Yan, S. A. Waldman, X. Chen, J. A. Posey, A. Basu-Mallick, B. H. Jiang, L. Hou, J. Shu, Y. Sun, J. Xing, B. Li, H. Yang, 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).

6. T.-M. Kim, S.-H. Jung, C. H. An, S. H. Lee, I.-P. Baek, M. S. Kim, S.-W. Park, J.-K. Rhee, S.-H. Lee, Y.-J. Chung, Subclonal genomic architectures of primary and metastatic colorectal cancer based on intratumoral genetic heterogeneity. *Clin. Cancer Res.* **21**, 4461–4472 (2015).

7. M. L. Leung, A. Davis, R. Gao, A. Casasent, Y. Wang, E. Sei, E. Vilar, D. Maru, S. Kopetz, N. E. Navin, Single-cell DNA sequencing reveals a late-dissemination model in metastatic colorectal cancer. *Genome Res.* **27**, 1287–1299 (2017).

8. Z. Hu, J. Ding, Z. Ma, R. Sun, J. A. Seoane, J. Scott Shaffer, C. J. Suarez, A. S. Berghoff, C. Cremolini, A. Falcone, F. Loupakis, P. Birner, M. Preusser, H.-J. Lenz, C. Curtis, Quantitative evidence for early metastatic seeding in colorectal cancer. *Nat. Genet.* **51**, 1113–1122 (2019).

9. E. Stewart, S. M. Federico, X. Chen, A. A. Shelat, C. Bradley, B. Gordon, A. Karlstrom, N. R. Twarog, M. R. Clay, A. Bahrami, B. B. Freeman III, B. Xu, X. Zhou, J. Wu, V. Honnell, M. Ocarz, K. Blankenship, J. Dapper, E. R. Mardis, R. K. Wilson, J. Downing, J. Zhang, J. Easton, A. Pappo, M. A. Dyer, Orthotopic patient-derived xenografts of paediatric solid tumours. *Nature* **549**, 96–100 (2017).

10. L. Ding, M. J. Ellis, S. Li, D. E. Larson, K. Chen, J. W. Wallis, C. C. Harris, M. D. McLellan, R. S. Fulton, L. L. Fulton, R. M. Abbott, J. Hoog, D. J. Dooling, D. C. Koboldt, H. Schmidt, J. Kalicki, Q. Zhang, L. Chen, L. Lin, M. C. Wendl, J. F. McMichael, V. J. Magrini, L. Cook, S. D. McGrath, T. L. Vickery, E. Appelbaum, K. Deschryver, S. Davies, T. Guintoli, L. Lin, R. Crowder, Y. Tao, J. E. Snider, S. M. Smith, A. F. Dukes, G. E. Sanderson, C. S. Pohl, K. D. Delehaunty, C. C. Fronick, K. A. Pape, J. S. Reed, J. S. Robinson, J. S. Hodges, W. Schierding, N. D. Dees, D. Shen, D. P. Locke, M. E. Wiechert, J. M. Eldred, J. B. Peck, B. J. Oberkfell, J. T. Lolofie, F. Du, A. E. Hawkins, M. D. O'Laughlin, K. E. Bernard, M. Cunningham, G. Elliott, M. D. Mason, D. M. Thompson, J. L. Ivanovich, P. J. Goodfellow, C. M. Perou, G. M. Weinstock, R. Aft, M. Watson, T. J. Ley, R. K. Wilson, E. R. Mardis, Genome remodelling in a basal-like breast cancer metastasis and xenograft. *Nature* **464**, 999–1005 (2010).

11. G. Vlachogiannis, S. Hedayat, A. Vatsiou, Y. Jamin, J. Fernández-Mateos, K. Khan, A. Lampis, K. Eason, I. Huntingford, R. Burke, M. Rata, D.-M. Koh, N. Tunariu, D. Collins, S. Hulkki-Wilson, C. Ragulan, I. Spiteri, S. Y. Moorcraft, I. Chau, S. Rao, D. Watkins, N. Fotiadis, M. Bali, M. Darvish-Damavandi, H. Lote, Z. Eltahir, E. C. Smyth, R. Begum, P. A. Clarke, J. C. Hahne, M. Dowsett, J. de Bono, P. Workman, A. Sadanandam, M. Fassan, O. J. Sansom, S. Eccles, N. Starling, C. Braconi, A. Sottoriva, S. P. Robinson, D. Cunningham, N. Valeri, Patient-derived organoids model treatment response of metastatic gastrointestinal cancers. *Science* **359**, 920–926 (2018).

12. S. F. Roerink, N. Sasaki, H. Lee-Six, M. D. Young, L. B. Alexandrov, S. Behjati, T. J. Mitchell, S. Grossmann, H. Lightfoot, D. A. Egan, A. Pronk, N. Smakman, J. van Gorp, E. Anderson, S. J. Gamble, C. Alder, M. van de Wetering, P. J. Campbell, M. R. Stratton, H. Clevers, Intra-tumour diversification in colorectal cancer at the single-cell level. *Nature* **556**, 457–462 (2018).

13. S. Misale, R. Yaeger, S. Hobor, E. Scala, M. Janakiraman, D. Liska, E. Valtorta, R. Schiavo, M. Buscarino, G. Siravegna, K. Bencardino, A. Cercek, C.-T. Chen, S. Veronese, C. Zanon, A. Sartore-Bianchi, M. Gambacorta, M. Gallicchio, E. Vakiani, V. Boscaro, E. Medico, M. Weiser, S. Siena, F. Di Nicolantonio, D. Solit, A. Bardelli, Emergence of KRAS mutations and acquired resistance to anti-EGFR therapy in colorectal cancer. *Nature* **486**, 532–536 (2012).

14. A. T. Boutin, W.-T. Liao, M. Wang, S. S. Hwang, T. V. Karpinets, H. Cheung, G. C. Chu, S. Jiang, J. Hu, K. Chang, E. Vilar, X. Song, J. Zhang, S. Kopetz, A. Futreal, Y. A. Wang, L. N. Kwong, R. A. DePinho, Oncogenic Kras drives invasion and maintains metastases in colorectal cancer. *Genes Dev.* **31**, 370–382 (2017).

15. W. De Roock, V. De Vriendt, N. Normanno, F. Ciardiello, S. Tejpar, KRAS, BRAF, PIK3CA, and PTEN mutations: Implications for targeted therapies in metastatic colorectal cancer. *Lancet Oncol.* **12**, 594–603 (2011).

16. S. Semba, N. Itoh, M. Ito, E. M. Youssef, M. Harada, T. Moriya, W. Kimura, M. Yamakawa, Down-regulation of PIK3CG, a catalytic subunit of phosphatidylinositol 3-OH kinase, by CpG hypermethylation in human colorectal carcinoma. *Clin. Cancer Res.* **8**, 3824–3831 (2002).

17. M. Janakiraman, E. Vakiani, Z. Zeng, C. A. Pratilas, B. S. Taylor, D. Chitale, E. Halilovic, M. Wilson, K. Huberman, J. C. Ricarte Filho, Y. Persaud, D. A. Levine, J. A. Fagin, S. C. Jhanwar, J. M. Mariadason, A. Lash, M. Ladanyi, L. B. Saltz, A. Heguy, P. B. Paty, D. B. Solit, Genomic and biological characterization of exon 4 KRAS mutations in human cancer. *Cancer Res.* **70**, 5901–5911 (2010).

18. J. E. Talmadge, I. J. Fidler, AACR centennial series: The biology of cancer metastasis: Historical perspective. *Cancer Res.* **70**, 5649–5669 (2010).

19. N. Aceto, A. Bardia, D. T. Miyamoto, M. C. Donaldson, B. S. Wittner, J. A. Spencer, M. Yu, A. Pely, A. Engstrom, H. Zhu, B. W. Brannigan, R. Kapur, S. L. Stott, T. Shioda, S. Ramaswamy, D. T. Ting, C. P. Lin, M. Toner, D. A. Haber, S. Maheswaran, Circulating tumor cell clusters are oligoclonal precursors of breast cancer metastasis. *Cell* **158**, 1110–1122 (2014).

20. K. J. Cheung, V. Padmanaban, V. Silvestri, K. Schipper, J. D. Cohen, A. N. Fairchild, M. A. Gorin, J. E. Verdone, K. J. Pienta, J. S. Bader, A. J. Ewald, Polyclonal breast cancer metastases arise from collective dissemination of keratin 14-expressing tumor cell clusters. *Proc. Natl. Acad. Sci. U.S.A.* **113**, E854–E863 (2016).

21. I. J. Fidler, The pathogenesis of cancer metastasis: The "seed and soil" hypothesis revisited. *Nat. Rev. Cancer* **3**, 453–458 (2003).

22. D. S. Micalizzi, S. Maheswaran, D. A. Haber, A conduit to metastasis: Circulating tumor cell biology. *Genes Dev.* **31**, 1827–1840 (2017).

23. D. Hanahan, R. A. Weinberg, Hallmarks of cancer: The next generation. *Cell* **144**, 646–674 (2011).

24. P. J. Ulintz, J. K. Greenson, R. Wu, E. R. Fearon, K. M. Hardiman, Lymph node metastases in colon cancer are polyclonal. *Clin. Cancer Res.* **24**, 2214–2224 (2018).

25. R. S. Goswami, K. P. Patel, R. R. Singh, F. Meric-Bernstam, S. Kopetz, V. Subbiah, R. H. Alvarez, M. A. Davies, K. J. Jabbar, S. Roy Chowdhuri, A. J. Lazar, L. J. Medeiros, R. R. Broaddus, R. Luthra, M. J. Routbort, Hotspot mutation panel testing reveals clonal evolution in a study of 265 paired primary and metastatic tumors. *Clin. Cancer Res.* **21**, 2644–2651 (2015).

26. K. Pantel, C. Alix-Panabières, Liquid biopsy and minimal residual disease - latest advances and implications for cure. *Nat. Rev. Clin. Oncol.* **16**, 409–424 (2019).

27. A. Woolston, K. Khan, G. Spain, L. J. Barber, B. Griffiths, R. Gonzalez-Exposito, L. Hornsteiner, M. Punta, Y. Patil, A. Newey, S. Mansukhani, M. N. Davies, A. Furness, F. Sclafani, C. Peckitt, M. Jiménez, K. Kouvelakis, R. Ranftl, R. Begum, I. Rana, J. Thomas, A. Bryant, S. Quezada, A. Wotherspoon, N. Khan, N. Fotiadis, T. Marafioti, T. Powles, S. Lise, F. Calvo, S. Guettler, K. von Loga, S. Rao, D. Watkins, N. Starling, I. Chau, A. Sadanandam, D. Cunningham, M. Gerlinger, Genomic and transcriptomic determinants of therapy resistance and immune landscape evolution during anti-EGFR treatment in colorectal cancer. *Cancer Cell* **36**, 35–50.e9 (2019).

28. P. A. Ott, Y.-J. Bang, S. A. Piha-Paul, A. R. A. Razak, J. Bennouna, J.-C. Soria, H. S. Rugo, R. B. Cohen, B. H. O'Neil, J. M. Mehnert, J. Lopez, T. Doi, E. M. J. van Brummelen, R. Cristescu, P.

Yang, K. Emancipator, K. Stein, M. Ayers, A. K. Joe, J. K. Lunceford, T-cell-inflamed gene-expression profile, programmed death ligand 1 expression, and tumor mutational burden predict efficacy in patients treated with pembrolizumab across 20 cancers: KEYNOTE-028. *J. Clin. Oncol.* **37**, 318–327 (2018).

29. C. Bettegowda, M. Sausen, R. J. Leary, I. Kinde, Y. Wang, N. Agrawal, B. R. Bartlett, H. Wang, B. Luber, R. M. Alani, E. S. Antonarakis, N. S. Azad, A. Bardelli, H. Brem, J. L. Cameron, C. C. Lee, L. A. Fecher, G. L. Gallia, P. Gibbs, D. Le, R. L. Giuntoli, M. Goggins, M. D. Hogarty, M. Holdhoff, S. M. Hong, Y. Jiao, H. H. Juhl, J. J. Kim, G. Siravegna, D. A. Laheru, C. Lauricella, M. Lim, E. J. Lipson, S. K. Marie, G. J. Netto, K. S. Oliner, A. Olivi, L. Olsson, G. J. Riggins, A. Sartore-Bianchi, K. Schmidt, l.-M. Shih, S. M. Oba-Shinjo, S. Siena, D. Theodorescu, J. Tie, T. T. Harkins, S. Veronese, T. L. Wang, J. D. Weingart, C. L. Wolfgang, L. D. Wood, D. Xing, R. H. Hruban, J. Wu, P. J. Allen, C. M. Schmidt, M. A. Choti, V. E. Velculescu, K. W. Kinzler, B. Vogelstein, N. Papadopoulos, L. A. Diaz Jr., Detection of circulating tumor DNA in early- and late-stage human malignancies. *Sci. Transl. Med.* **6**, 224ra24 (2014).

30. J. Tie, Y. Wang, C. Tomasetti, L. Li, S. Springer, I. Kinde, N. Silliman, M. Tacey, H.-L. Wong, M. Christie, S. Kosmider, I. Skinner, R. Wong, M. Steel, B. Tran, J. Desai, I. Jones, A. Haydon, T. Hayes, T. J. Price, R. L. Strausberg, L. A. Diaz Jr., N. Papadopoulos, K. W. Kinzler, B. Vogelstein, P. Gibbs, Circulating tumor DNA analysis detects minimal residual disease and predicts recurrence in patients with stage II colon cancer. *Sci. Transl. Med.* **8**, 346ra92 (2016).

31. J. G. Grossman, T. M. Nywening, B. A. Belt, R. Z. Panni, B. A. Krasnick, D. G. DeNardo, W. G. Hawkins, S. P. Goedegebuure, D. C. Linehan, R. C. Fields, Recruitment of CCR2+ tumor associated macrophage to sites of liver metastasis confers a poor prognosis in human colorectal cancer. *Oncoimmunology* **7**, e1470729 (2018).

32. C. A. Miller, B. S. White, N. D. Dees, M. Griffith, J. S. Welch, O. L. Griffith, R. Vij, M. H. Tomasson, T. A. Graubert, M. J. Walter, M. J. Ellis, W. Schierding, J. F. DiPersio, T. J. Ley, E. R. Mardis, R. K. Wilson, L. Ding, SciClone: Inferring clonal architecture and tracking the spatial and temporal patterns of tumor evolution. *PLOS Comput. Biol.* **10**, e1003665 (2014).

33. H. X. Dang, B. S. White, S. M. Foltz, C. A. Miller, J. Luo, R. C. Fields, C. A. Maher, ClonEvol: Clonal ordering and visualization in cancer sequencing. *Ann. Oncol.* **28**, 3076–3082 (2017).

34. C. Kandoth, M. D. McLellan, F. Vandin, K. Ye, B. Niu, C. Lu, M. Xie, Q. Zhang, J. F. McMichael, M. A. Wyczalkowski, M. D. M. Leiserson, C. A. Miller, J. S. Welch, M. J. Walter, M. C. Wendl, T. J. Ley, R. K. Wilson, B. J. Raphael, L. Ding, Mutational landscape and significance across 12 major cancer types. *Nature* **502**, 333–339 (2013).

35. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

36. M. Griffith, O. L. Griffith, S. M. Smith, A. Ramu, M. B. Callaway, A. M. Brummett, M. J. Kiwala, A. C. Coffman, A. A. Regier, B. J. Oberkfell, G. E. Sanderson, T. P. Mooney, N. G. Nutter, E. A. Belter, F. Du, R. L. Long, T. E. Abbott, I. T. Ferguson, D. L. Morton, M. M. Burnett, J. V. Weible, J. B. Peck, A. Dukes, J. F. McMichael, J. T. Lolofie, B. R. Derickson, J. Hundal, Z. L. Skidmore, B. J. Ainscough, N. D. Dees, W. S. Schierding, C. Kandoth, K. H. Kim, C. Lu, C. C. Harris, N. Maher, C. A. Maher, V. J. Magrini, B. S. Abbott, K. Chen, E. Clark, I. Das, X. Fan, A. E. Hawkins, T. G. Hepler, T. N. Wylie, S. M. Leonard, W. E. Schroeder, X. Shi, L. K. Carmichael, M. R. Weil, R. W. Wohlstadter, G. Stiehr, M. D. McLellan, C. S. Pohl, C. A. Miller, D. C. Koboldt, J. R. Walker, J. M. Eldred, D. E. Larson, D. J. Dooling, L. Ding, E. R. Mardis, R. K. Wilson, Genome modeling system: A knowledge management platform for genomics. *PLOS Comput. Biol.* **11**, e1004274 (2015).

37. T. Kuilman, A. Velds, K. Kemper, M. Ranzani, L. Bombardelli, M. Hoogstraat, E. Nevedomskaya, G. Xu, J. de Ruiter, M. P. Lolkema, B. Ylstra, J. Jonkers, S. Rottenberg, L. F. Wessels, D. J. Adams, D. S. Peeper, O. Krijgsman, CopywriteR: DNA copy number detection from off-target sequence data. *Genome Biol.* **16**, 49 (2015).

38. M. Griffith, C. A. Miller, O. L. Griffith, K. Krysiak, Z. L. Skidmore, A. Ramu, J. R. Walker, H. X. Dang, L. Trani, D. E. Larson, R. T. Demeter, M. C. Wendl, J. F. McMichael, R. E. Austin, V. Magrini, S. D. McGrath, A. Ly, S. Kulkarni, M. G. Cordes, C. C. Fronick, R. S. Fulton, C. A. Maher, L. Ding, J. M. Klco, E. R. Mardis, T. J. Ley, R. K. Wilson, Optimizing cancer genome sequencing and analysis. *Cell Syst.* **1**, 210–223 (2015).

39. C. A. Miller, J. McMichael, H. X. Dang, C. A. Maher, L. Ding, T. J. Ley, E. R. Mardis, R. K. Wilson, Visualizing tumor evolution with the fishplot package for R. *BMC Genomics* **17**, 880 (2016).