Inaugural dissertation

for

obtaining the doctoral degree

of the

Combined Faculty of Mathematics, Engineering and Natural Sciences

of the

Ruprecht - Karls - University

Heidelberg

Presented by

M.Sc. Mhd Haidar Kasem

born in Damascus

Oral examination: 19.04.2024

Intra-clonal gene expression heterogeneity in human colorectal cancer

Referees:

Prof. Dr. Jan Lohmann

Prof. Dr. Hanno Glimm

Abstract

Colorectal cancer (CRC) harbors heterogeneous subclonal cell populations that contribute to tumor progression, metastasis and therapy resistance. Subclonal heterogeneity arises not only from genetic and epigenetic differences, but also from functional heterogeneity of subclones sharing homogeneous genetic features. Genetic barcoding approaches have shown that subclones with long-term tumor-initiating cell (LT-TIC) activity are capable of fueling CRC expansion to metastasis and that CRC subclones can enter a drug tolerant persister state to tolerate chemotherapy. Nevertheless, subclonal transcriptional features which characterize functional states, such as treatment response and metastasis are underexplored. Characterizing how subclonal gene expression contributes to functional states of CRC enhances understanding the dynamics of tumor progression. Here, I describe subclonal dynamics in addition to associated single cell transcriptomics of treatment response and metastasis in barcoded CRC patient-derived organoids (PDOs) and xenografts, respectively.

To study subclonal gene expression heterogeneity in metastatic and therapy resistant populations, three CRC PDOs were transduced with a genetic RNA-expressed barcoding library for unique single cell marking. Barcoded PDOs were either orthotopically serially transplanted into NOD.Cg-Prkdc scid Il2rg tm1Wjl /SzJ (NSG) mice or treated with chemotherapy to study clonal and transcriptional dynamics in tumor progression to metastasis and in treatment response. Cells harvested from xenografts or treated PDOs were subjected to single cell RNA (scRNA) sequencing in addition to target barcode DNA sequencing to analyze clonal dynamics and subclonal gene expression programs. Combined scRNA data from more than 3.5x10E5 cells in xenografts demonstrated a drop in the number of clones by which 0.05-1% of detected clones formed the majority of cancer cells in serial transplantations of tumor, liver and lung metastasis, with an enrichment of few individual clones that drive tumor progression. The frequency of these enriched clones varied across serial transplantations of different tumor locations such as the tumor and metastases, suggesting a dynamic subclonal expansion in different microenvironments. Subclonal cell populations in both tumor and liver or lung metastases reflected distinct gene expression profiles of differentiated and undifferentiated intestinal cells. Differentiated epithelial profiles were reminiscent of epithelial intestinal cells such as absorptive enterocytes and secretory cells while undifferentiated cells shared transcriptional features of invasive and fetal/embryonic cell states. Individual metastatic subclones of liver metastasis showed higher expression of particular genes such as SOX4, VEGFA and PLCG2 which were enriched in the invasive fetal cell phenotype. Lung metastasis cells of the same subclones

demonstrated higher expression of S100A11, ANXA1 and KRT19, which were elevated in the differentiated epithelial cell phenotype. In addition, clones which represent more than 25% of the tumor or metastasis cell population did not show reproducible gene expression differences in comparison to smaller clones forming less than 5% of the cell population. Upon chemotherapy treatment and recovery of PDOs, the frequency of individual clones was marginally affected in comparison to control DMSO treated cells, without a strong selection of particular clones. In response to treatment, cells reflected abundant pre-existing gene expression profiles reminiscent to those of intestinal enterocytes, undifferentiated cycling and progenitor cells and a low frequency of transcriptional profiles related to proliferative and secretory cells. One week after drug washout, stem cell transcriptional clusters, in addition to treatment response abundant transcriptional clusters, were enriched. The functional relevance of 181 marker genes of treatment response-associated transcriptional clusters was assessed through a custom clustered regularly interspaced palindromic repeats (CRISPRi) screen, which demonstrated the depletion of 43 out of 52 significantly affected genes after chemotherapy treatment. Further, transposase-accessible chromatin single cell (scATAC) sequencing data from 2 PDOs treated with 5-Fluorouracil (5-FU) showed an open chromatin status in the region of genes detected to be differentially expressed in response-associated transcriptional clusters identified in scRNA data, thus linking gene transcription to epigenetic chromatin regulation.

Overall, this study links clonal dynamics to transcriptional states, thereby dissecting gene expression of subclonal functional states of tumor progression. This analysis reveals gene expression states and marker genes associated with metastasis and treatment response, that could be implemented in further validation and development of innovative treatment strategies in CRC.

Zusammenfassung

Darmkrebs (CRC) beherbergt heterogene subklonale Zellpopulationen, die zur Tumorprogression, Metastasierung und Therapieresistenz beitragen. Die subklonale Heterogenität ergibt sich nicht nur aus genetischen und epigenetischen Unterschieden, sondern auch aus der funktionellen Heterogenität von Subklonen mit homogenen genetischen Merkmalen. Genetische Barcoding-Ansätze haben gezeigt, dass Subklone mit langfristiger tumorinitiierender Zellaktivität (LT-TIC) in der Lage sind, die Ausbreitung von Darmkrebs bis hin zur Metastasierung voranzutreiben, und dass Darmkrebs-Subklone in einen arzneimitteltoleranten Persistenzzustand übergehen können, um eine Chemotherapie zu tolerieren. Dennoch sind subklonale transkriptionelle Merkmale, die funktionelle Zustände wie das Ansprechen die Behandlung und die Metastasierung charakterisieren, noch wenig erforscht. Die Charakterisierung des Beitrags der subklonalen Genexpression zu den funktionellen Zuständen des kolorektalen Karzinoms verbessert das Verständnis der Dynamik der Tumorprogression. Hier beschreibe ich die subklonale Dynamik sowie die damit verbundene Einzelzelltranskriptomik des Behandlungsansprechens und der Metastasierung in barcodierten patientenabgeleiteten CRC-Organoiden (PDOs) und Xenotransplantaten von Darmkrebs-Patienten.

Um die Heterogenität subklonaler Genexpression in metastatischen und therapieresistenten Populationen zu untersuchen, wurden drei CRC-PDOs mit einer genetischen RNA-exprimierten Barcoding-Bibliothek zur eindeutigen Einzelzellmarkierung transduziert. Barcodierte PDOs wurden entweder orthotopisch seriell in NOD.Cg-Prkdc scid Il2rg tm1Wjl /SzJ (NSG)-Mäuse transplantiert oder mit Chemotherapie behandelt, um die klonale und transkriptionelle Dynamik bei der Tumorprogression bis zur Metastasierung und beim Ansprechen auf die Behandlung zu untersuchen. Die aus Xenotransplantaten oder behandelten PDOs gewonnenen Zellen wurden zusätzlich zur Sequenzierung von Target-Barcode-DNA einer Einzelzell-RNA- (scRNA)-Sequenzierung unterzogen, um die klonale Dynamik und subklonale Genexpressionsprogramme zu analysieren. Kombinierte scRNA-Daten von mehr als 3,5x10E5 Zellen in Xenotransplantaten zeigten einen Rückgang der Anzahl der Klone, wobei 0,05-1% der entdeckten Klone die Mehrheit der Krebszellen in seriellen Transplantationen von Tumor-, Leber- und Lungenmetastasen bildeten, mit einer Anreicherung von wenigen einzelnen Klonen, die das Tumorwachstum vorantreiben. Die Häufigkeit dieser angereicherten Klone variierte bei seriellen Transplantationen an verschiedenen Tumororten wie dem Tumor und Metastasen, was auf eine dynamische subklonale Expansion in verschiedenen Mikroumgebungen hindeutet. Subklonale Zellpopulationen sowohl im Tumor als auch in Leber- oder Lungenmetastasen spiegelten

unterschiedliche Genexpressionsprofile von differenzierten und undifferenzierten Darmzellen wider. Differenzierte Epithelprofile erinnerten an epitheliale Darmzellen wie absorbierende Enterozyten und sekretorische Zellen, während undifferenzierte Zellen Transkriptionsmerkmale invasiver und fötaler/embryonaler Zellzustände aufwiesen. Einzelne metastatische Subklone der Lebermetastasen zeigten eine höhere Expression bestimmter Gene wie SOX4, VEGFA und PLCG2, die im Phänotyp der invasiven fötalen Zellen angereichert waren. Lungenmetastasenzellen derselben Subklone wiesen eine höhere Expression von S100A11, ANXA1 und KRT19 auf, die im differenzierten Epithelzell-Phänotyp erhöht waren. Darüber hinaus zeigten Klone, die mehr als 25% der Tumor- oder Metastasenzellpopulation ausmachen, keine reproduzierbaren Unterschiede in der Genexpression im Vergleich zu kleineren Klonen, die weniger als 5% der Zellpopulation ausmachen. Nach der Chemotherapie und der Wiederherstellung der PDOs wurde die Häufigkeit einzelner Klone im Vergleich zu den mit DMSO behandelten Kontrollzellen nur geringfügig beeinflusst, ohne dass eine starke Selektion bestimmter Klone stattfand. Als Reaktion auf die Behandlung wiesen die Zellen reichlich vorbestehende Genexpressionsprofile auf, die an die von intestinalen Enterozyten, undifferenzierten zyklischen Zellen und Vorläuferzellen erinnerten, sowie eine geringe Häufigkeit von Transkriptionsprofilen, die mit proliferativen und sekretorischen Zellen zusammenhängen. Eine Woche nach dem Absetzen des Medikaments waren stammzellassoziierte Transkriptionscluster angereichert, zusätzlich zu den Clustern, die bei der Reaktion auf die Behandlung reichlich vorhanden waren. Die funktionelle Bedeutung von 181 Markergenen der mit der Behandlungsreaktion verbundenen Transkriptionscluster wurde mit Hilfe eines benutzerdefinierten CRISPRi-Screens (Clustered Regular Interspaced Palindromic Repeats) bewertet, der die Verringerung von 43 der 52 signifikant betroffenen Gene nach einer Chemotherapiebehandlung zeigte. Darüber hinaus zeigten Transposase-accessible chromatin single cell (scATAC) Sequenzierungsdaten von 2 PDOs, die mit 5-Fluorouracil (5-FU) behandelt wurden, einen offenen Chromatin-Status in der Region von Genen, die nachweislich in den mit der Reaktion assoziierten transkriptionellen Clustern, die in den scRNA-Daten identifiziert wurden, unterschiedlich exprimiert werden, wodurch eine Verbindung zwischen der Gentranskription und der epigenetischen Chromatinregulierung hergestellt wurde.

Insgesamt verbindet diese Studie die klonale Dynamik mit transkriptionellen Zuständen, wodurch die Genexpression subklonaler Funktionszustände der Tumorprogression analysiert wird. Diese Analyse deckt Genexpressionszustände und Markergene auf, die mit Metastasierung und Behandlungsansprechen in Verbindung gebracht werden und die bei der weiteren Validierung und Entwicklung innovativer Behandlungsstrategien im CRC eingesetzt werden könnten.

Table of contents

Abbreviations

APC: Adenomatous Polyposis Coli

ARGLU1: Arginine and Glutamate Rich 1

ATOX1: Antioxidant 1 Copper Chaperone

BRAF: B-raf Proto-oncogene Serine/threonine Kinase

CASKIN2: CASK Interacting Protein 2

CCNB1: Cyclin B1

CEA: Carcinoembryonales Antigen

CCL20: C-C Motif Chemokine Ligand 20

CDH1: E-cadherin

CDH2: Osteogenic N-cadherin

CFTR: Cystic Fibrosis Transmembrane Conductance Regulator

CIN: Chromosomal Instability

COADREAD: Colorectal Cancer Metastatic Cohort

COL6A1: Collagen Type VI Alpha 1 Chain

CRC: Colorectal Cancer

CRISPR: Clustered Regularly Interspaced Short Palindromic Repeats

CS: Clonality Score

CXCL1/2/3: C-X-C Motif Chemokine Ligand 1/2/3

DAB2: DAB Adaptor Protein 2

DCS: Deep Crypt Secretory

DEG: Differential Gene Expression

DMSO: Dimethyl Sulfoxide

DNMT1: DNA Methyltransferase 1

DTPs: Drug-tolerant Persister Cells

EGFR: Epidermal Growth Factor Receptor

EMT: Epithelial-to-mesenchymal Transition

EPCAM: Epithelial Cell Adhesion Molecule

E2F: E2 Promoter Binding Transcription Factor

FACS: Fluorescence-activated Cell Sorting

FABP1: Fatty Acid Binding Protein 1

FOLFIRI: 5-FU, Folinic Acid and Irinotecan

FOLFOX: 5-florouracil 5-FU, folinic acid and oxaliplatin

FOSL1: FOS Like 1, AP-1 Transcription Factor Subunit

GSEA: Gene Set Enrichment Analysis

GPX1: Glutathione Peroxidase 1

GPX4: Glutathione Peroxidase 4

H3K27: Histone H3 Lysine 27

IGF-1R: Insulin Like Growth Factor 1 Receptor

ISC: Intestinal Stem Cells

ITH: Intra-tumor Heterogeneity

JUND: JunD Proto-Oncogene, AP-1 Transcription Factor Subunit

KI67: Marker of Proliferation Ki-67

KRAS: Kirsten Rat Sarcoma Virus Proto-Oncogene

LGR5: Quiescent Leucine-rich Repeat-containing G-protein Coupled Receptor 5

LINE-1: Type Transposase Domain Containing 1 Pseudogene 1

LRG1: Leucine Rich Alpha-2-Glycoprotein 1

LY6A: Lymphocyte Antigen 6 Family Member S

MALAT1: Metastasis Associated Lung Adenocarcinoma Transcript 1

MAPK: Mitogen-Activated Protein Kinase 1

MEK: Mitogen-Activated Protein Kinase Kinase 1

MET: Mesenchymal-to- Epithelial transition

MIC: Metastatic-initiating Cell

MIF: Macrophage Migration Inhibitory Factor

MMP9: Matrix Metallopeptidase 9

MSI: Microsatellite Instable

MSS: Microsatellite Stable

NCAM1: Neural Cell Adhesion Molecule 1

NES: Normalized Enrichment Score

NEAT1: Nuclear Paraspeckle Assembly Transcript 1

NFKb: Nuclear Factor 'Kappa-Light-Chain-Enhancer' of Activated B-cells

NGS: Next Generation Sequencing

NKD1: NKD Inhibitor of WNT Signaling Pathway 1

NRAS: NRAS Proto-Oncogene

NSCLC: Non-small Cell Lung Cancer

NSG: NOD.Cg-PrkdcscidIl2rgtm1Wjl/SzJ

PDOs: Patient-derived Organoids

PIK3CA: Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Alpha

PLCG2: Phospholipase C Gamma 2

PRAMEF12: PRAME Family Member 12

PRC2: Polycomb Repressive Complexes 2

PD-1: Programmed Cell Death 1

PD-L1: Programmed Death Ligand 1

PTEN: Phosphatase and Tensin Homolog

PTTG1: PTTG1 Regulator of Sister Chromatid Separation, Securin

RAB25: RAB25, Member RAS Oncogene Family

RAN: RAN, Member RAS Oncogene Family

REG4: Regenerating Family Member 4

RFP: Red Fluorescent Protein

RNF20: Ring Finger Protein 20

RUNX3: RUNX Family Transcription Factor 3

SCNA: Somatic Copy Number Alterations

scATAC-seq: Transposase-accessible Chromatin Single Cell Sequencing

scRNA: Single Cell RNA

SEMA3A: Semaphorin 3A

SMAD4: SMAD Family Member 4

SMARCC1: SWI/SNF Related, Matrix Associated, Actin Dependent Regulator of Chromatin Subfamily C Member 1

Slc38a1/2: Solute Carrier Family 38 Member 1/2

SMOC2: SPARC Related Modular Calcium Binding 2

SNAI1/2: Snail Family Transcriptional Repressor

SOX2/4/9: SRY-Box Transcription Factor 2/4/9

S100A: Calcium Binding Protein Family

SUMO2: Small Ubiquitin-like Modifier 2

SUZ12: SUZ12 Polycomb Repressive Complex 2 Subunit

T790M: Methionine for Threonine Substitution at Amino Acid Position 790 in EGFR

TACs: Tumor Amplifying Cells

TAZ: WW Domain Containing Transcription Regulator 1

TCGA: The Cancer Genome Atlas

TEAD: TEA Domain Transcription Factor 1

TF: Transcription factor

TFAP2A: Transcription Factor AP-2 Alpha

TFAP2C: Transcription Factor AP-2 Gamma

TGF-β: Transforming Growth Factor Beta

TGFBR2: Transforming Growth Factor Beta Receptor 2

TIC: Tumor Initiating Cell

TIMP: Tissue Inhibitor of Metalloproteinase

TLR4: Toll Like Receptor 4

TOP2A: DNA Topoisomerase II Alpha

TP53: Tumor Protein P53

TWIST1: Twist Family BHLH Transcription Factor 1

VEGF: Vascular Endothelial Growth Factor

WNT: Wingless-related Integration Site

YAP: Yes Associated Transcriptional Regulator

ZEB1: Zinc Finger E-Box Binding Homeobox 1

1. Introduction

1.1 Colorectal Cancer

CRC bears a significant global burden, ranking third in terms of incidence and second in terms of mortality among all cancer types (Morgan et al., 2023). An estimated incidence and mortality rate of 4.4 to 40.6 per 100,000 and 2.5 to 20.2 per 100,000, respectively is reported for the year of 2020, with a projection of 3.2x10E6 new cases in 2040 (Morgan et al., 2023). These statistics underscore the urgency in improving prevention, early detection, and treatment options to mitigate the impact of CRC on individuals and healthcare systems. Colorectal cancer CRC is a molecularly heterogeneous disease. It consists of cells with various genetic alterations and phenotypic features (Becht et al., 2016; Chan & Buczacki, 2021). The heterogeneity of CRC arises from the clonal origin of the disease by which a cell acquires genetic mutations that further diverge in disease progression giving rise to multiple subclones with different genetic and epigenetic features. The genetic, epigenetic, clinical, and pathological landscapes of CRC differ significantly among patients and within individual tumors (Becht et al., 2016). The latter, known as intra-tumor heterogeneity (ITH) adds an additional layer of complexity, signifying the presence of distinct cell populations even within a single tumor (Chan & Buczacki, 2021). ITH is observed in almost all types of cancer and contributes to drug resistance, cancer metastasis, and treatment failure (Stanta & Bonin, 2018). This diversity impacts patient's response to treatment, consequently affecting the patients' clinical prognosis.

1.2 The Intestinal Epithelium

Intestinal epithelium, from which CRC emerges, features finger-like projections known as villi, pivotal for nutrient absorption, and is encompassed by invaginations termed crypts (Clevers, 2013). Located at the crypt base, columnar base crypt (CBC) cells known as intestinal stem cells (ISCs) perpetually generate differentiated cell progeny ascending toward the villi, facilitating the continuous self-renewal of the intestinal epithelium approximately every 3–5 days (Barker, 2014). ISCs give rise to rapidly dividing transit-amplifying (TA) cells. These TA cells then differentiate into mature epithelial cells across various lineages (Figure 1A) (de Sousa & de Sauvage, 2019). At the crypt bottom alongside ISCs are located Paneth cells, which maintain homeostasis through secreting niche factors to ISCs. Upon damage, the intestinal epithelium exhibits remarkable regenerative capacity. Quiescent Leucine-rich repeat-containing G-protein coupled receptor 5 positive (LGR5+) stem cells at the +4 position from the crypt base could withstand damage due to their non-cycling nature and are proposed to contribute to repair upon injury (Barker et al.,

2007). A later report showed that, upon Lgr5+ cell ablation, crypt base columnar (CBC) cells can be efficiently replaced by cells higher up in the crypt, highlighting cell fate plasticity when the stem cell compartment is compromised (Figure 1B) (H. Tian et al., 2011).

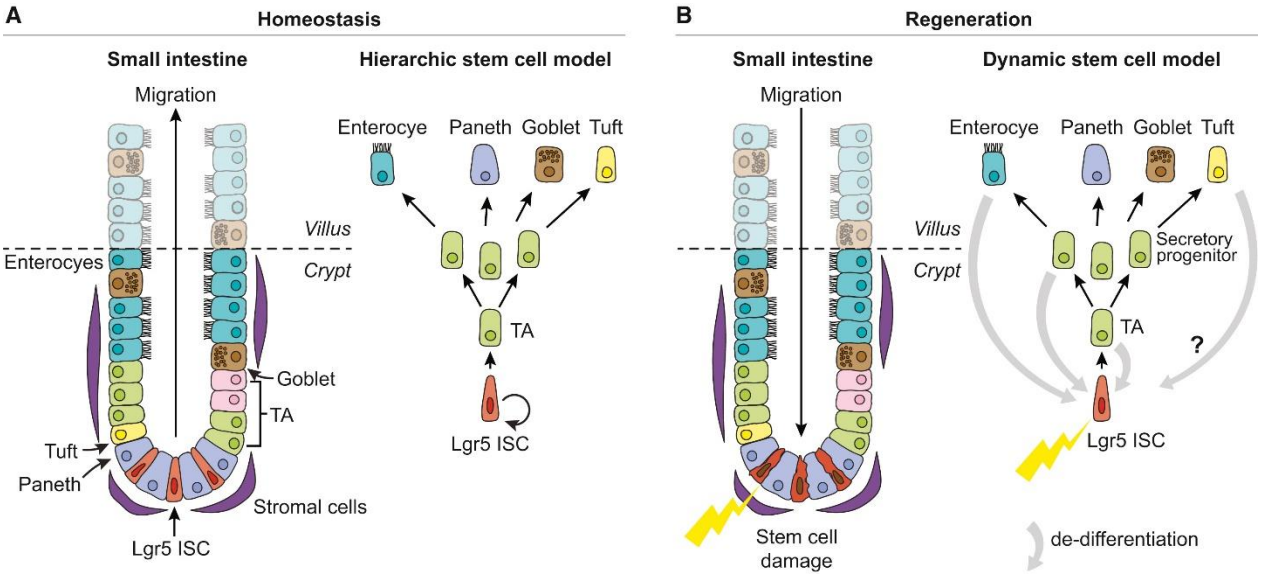


Figure 1. The intestinal crypt model. A. Representation of hierarchical organization of intestinal cells at the crypt base continuously generating transit-amplifying (TA) cells, which then differentiate into the various functional cells in the crypt and villi. Plasticity of LGR5- top crypt cells in regenerating LGR5+ stem cell compartment at the crypt bottom upon damage. Adapted by permission of ScienceDirect from Cellular plasticity in intestinal homeostasis and disease (de Sousa & de Sauvage, 2019).

In contrast to the small intestine, specializing in nutrient absorption, the colon, positioned at the terminal end of the intestinal tract, lacks villi and displays a flat epithelial surface arranged into densely packed crypts, emphasizing its specialization in water reuptake for waste compaction (Barker, 2014; Beumer & Clevers, 2021). While the colonic crypt shares cell types with the intestinal crypt, it stands out with a higher Goblet cell density and the absence of Paneth cells. In the colon, deep crypt secretory (DCS) cells, akin to Paneth cells in the small intestine, form the niche for LGR5-expressing stem cells, marked by regenerating family member 4 (REG4) expression (Altmann, 1983). Recent findings underscore that the removal of DCS cells leads to the loss of Lgr5+ stem cells both *in vitro* and *in vivo* (Sasaki et al., 2016).

1.3 Colorectal Cancer Tumorigenesis

The development of CRC is driven by a multistep process that encompasses genetic mutations, environmental factors, and epigenetic alterations (Fearon & Vogelstein, 1990; Markowitz & Bertagnolli, 2009). This intricate interplay drives the transformation of normal colon cells into malignant entities, molecularly through the accumulation of mutations in the colonic epithelial lining (Cross et al., 2018). Mutation in APC gene is a prevalent event that often leads to the formation of adenoma, which gradually evolves to an invasive adenocarcinoma by accumulation of additional mutations in oncogenes and tumor suppressor genes, often involving TP53 and KRAS (Figure 2) (Vogelstein et al., 1988). Subsequent alterations can drive tumor growth, local invasion, and ultimately, metastasis, which is the driving cause of lethality in CRC patients (Shin, Giancotti, & Rustgi, 2023).

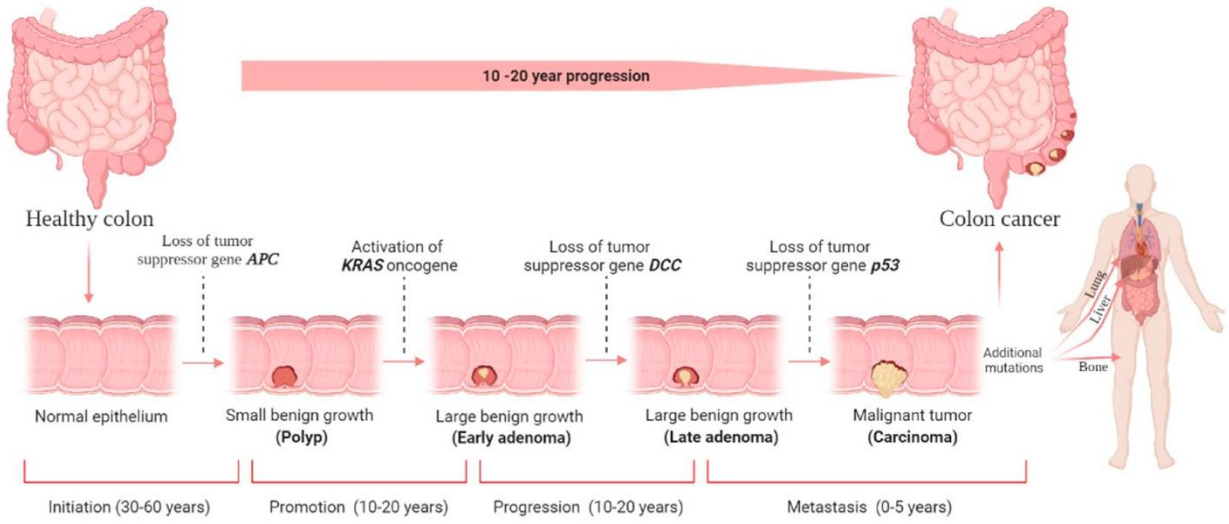


Figure 2. Colorectal cancer tumorigenesis. Decades of CRC development include initiation, promotion, progression, and metastasis. Liver, lung and bone are common metastatic sites of CRC. Adapted from (Hossain et al., 2022) with open access Creative Common CC BY license of MDPI.

1.4 Subtypes of Colorectal Cancer

Distinct molecular subtypes have been defined in CRC, each characterized by unique genetic and molecular features (Guinney et al., 2015). These molecular differences have practical implications for the treatment of CRC. The introduction of the consensus molecular subtypes

(CMS) system in CRC classification is based on gene expression profiles of various CRC patient cohorts (Guinney et al., 2015). This system defines four main CMSs, each associated with distinct molecular and clinical characteristics. CMS1 (microsatellite instable MSI immune) subtype accounts for approximately 14% of CRCs. Microsatellite instability is a molecular alteration caused by a defective DNA mismatch repair system, resulting in alterations of the short tandem repeats (STR) which are 1-6 base pairs repeating stretches of DNA. CMS1 is characterized by hypermutation, MSI-high status, and strong immune activation. These tumors are known to be highly immunogenic, making them responsive to immunotherapy approaches. On the other hand, microsatellite stable (MSS) tumors are typically treated with conventional chemotherapy or targeted therapies directed at molecules such as the epidermal growth factor receptor (EGFR) or vascular endothelial growth factor (VEGF) (Stintzing, Tejpar, Gibbs, Thiebach, & Lenz, 2017). CMS2 (canonical) is the largest subtype, representing around 37% of CRCs, is classified as the "epithelial" subtype and is characterized by the upregulation of WNT and MYC signaling pathways. About 13% of CRCs fall into the category of CMS3 (metabolic), which is characterized by metabolic dysregulation and epithelial features. CMS4 (mesenchymal) subtype accounts for roughly 23% of CRCs and exhibits prominent activation of TGF-β signaling and epithelial-to-mesenchymal transition (EMT), stromal invasion, and angiogenesis. CMS4 tumors are associated with a poorer prognosis, including lower relapse-free and overall survival rates. This CMS classification system provides valuable insights into the heterogeneity of CRC and can guide treatment approaches adequate to the molecular subtype of the patient, in a more personalized strategy (Guinney et al., 2015). Recognizing and classifying these heterogeneous subtypes enable targeted treatment strategies that may prevent disease progression. Therefore, the identification of molecular subtypes of colorectal cancer has been proposed as a way to improve personalized medicine (Gaiani et al., 2021).

In addition to molecular subtypes, colorectal cancers are further categorized according to their precise location within the colon, as either right-sided or left-sided tumors. Right-sided carcinomas are notably marked by impaired DNA mismatch repair mechanisms, resulting in a pronounced occurrence of microsatellite instability (MSI-high). Conversely, left-sided tumors tend to be predominantly linked with chromosomal instability (CIN-high), defined as persistently high rate of loss and gain of whole chromosomes, and can display discrete genetic and epigenetic modifications (Baran et al., 2018). The serrated pathway has been also identified in the context of CRC, apart from the traditional adenoma-carcinoma model. Serrated polyps, which are epithelial lesions of the colon and rectum displaying a “sawtooth” pattern on the polyp's surface and crypt epithelium, predominantly manifest mutations in the B-raf proto-oncogene,

serine/threonine kinase (BRAF), and further key genes such as catenin beta 1 (CTNNB1), KRAS, PIK3CA, and TGFBR2. APC mutations rarely occur as a driver event in this mode. Microsatellite instability and promoter hypermethylation often occur in these serrated polyps (Baran et al., 2018).

1.5 Intra-tumor Heterogeneity: An Additional Layer of Complexity

In addition to driving inter-tumor diversity between CRC patients, genomic instability, alterations in mismatch repair and cancer microenvironment can concurrently induce new heritable genetic and epigenetic variations in single cells within the same tumor (Figure 3) (Househam et al., 2022; Sobral et al., 2022). Throughout their expansion, these cells accumulate genetic and phenotypic distinctions from one another, giving rise to unique subclones within the same tumor (Sobral et al., 2022). Successively, more fit subclones characterized by enhanced self-renewal and growth potential replace less fit subclones (Greaves & Maley, 2012). This evolution of subclonal structure in disease progression, known as intra-tumor heterogeneity (ITH), refers to the coexistence of not only genetic or epigenetic, but also phenotypic distinctions of cancer cells within a tumor (Z. Zheng et al., 2020).

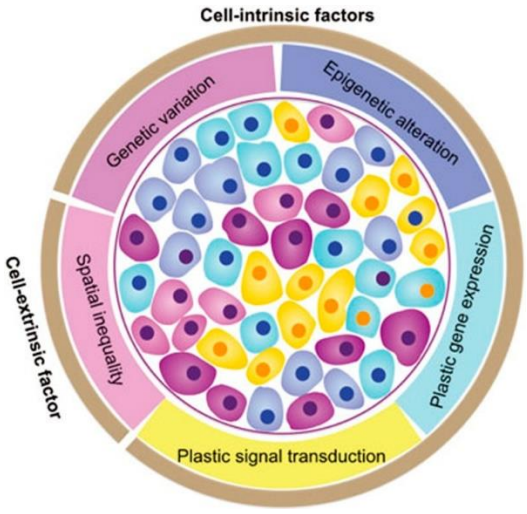


Figure 3. Diverse factors contribute to tumor heterogeneity. Genetic and epigenetic diversity within the same tumor or in different individuals are the drivers of cancer heterogeneity. Other players include spatial localization and microenvironment. Depicted from (Sun & Yu, 2015) under the Creative Commons Attribution 4.0 International License permission from Springer Nature.

1.6 Functional Intra-tumor Heterogeneity in Colorectal Cancer

Functional heterogeneity in CRC extends beyond genetic diversity. Phenotypic intra-tumor heterogeneity (ITH) manifests as variations in cellular properties within a tumor, including differences in proliferation, metabolism, invasiveness and differentiation. The coexistence of cells with distinct properties can endow the tumor with functional capacities to tolerate stress in different microenvironments and to resist therapy (Ramón et al., 2020). Indeed, cells derived from clones of different genetic background in the same tumor show different viability in response to different treatments (Roerink et al., 2018). In addition, cancer stem cells (CSCs) are a defined functional compartment, responsible for tumor progression and self-renewal, which shows advanced proliferative properties and less differentiation compared to other cancer cells (Batlle & Clevers, 2017; Driessens, Beck, Caauwe, Simons, & Blanpain, 2012). Our group has experimentally described how individual clones possess variable capacities in driving tumor progression and metastasis. Long-term tumor-initiating cells (LT-TICs) reminiscent to CSCs, characterized by significant self-renewal capabilities, predominantly steer tumor progression. Conversely, highly proliferative but short-lived tumor transient-amplifying cells (T-TACs) possess limited or no self-renewal potential and contribute to the tumor's mass (Dieter et al., 2011). Subsequently, our group has demonstrated that this functional diversity within the CRC TIC population remains stable through successive xenografts, even in the presence of substantial shifts in the genetic subclone composition (Giessler et al., 2017). These cancer stem cells are not an exclusively fixed population, but rather represent a dynamic cell feature that can be acquired by differentiated cells when TICs are damaged, as shown in separate publications (de Sousa e Melo et al., 2017; Dieter, Glimm, & Ball, 2017; Fumagalli et al., 2020; Shimokawa et al., 2017).

Additionally, ITH correlates with worse prognosis in CRC, by which genetic heterogeneity increases in worse pathological stage disease (Oh et al., 2019). While reports show that some driver mutations (mutations that provide a selective growth advantage and thus drive the cell lineage) such as KRAS, NRAS, BRAF and PIK3CA are concordant in both the primary tumor and metastatic lesions (Brannon et al., 2014; Petaccia de Macedo et al., 2017), other reports have shown higher genetic heterogeneity in metastasis compared to primary tumor, demonstrated by the appearance of additional mutations (Naxerova et al., 2017).

Genetic heterogeneity is widespread, but also epigenetic changes are responsible for phenotypic variation between cancer cells. For example, more than one CMS was shown to co-exist within the same tumor and even within colon cancer lines, indicating that individual tumors are a mixture of cells possessing different CMSs with heterogeneous gene expression profiles, a property

contributing to worse patient prognosis (Joanito et al., 2022; Marisa et al., 2021). Distinct landscapes of DNA methylation contribute to the heterogeneity of CRC, especially in distant metastatic lesions (Martínez-Cardús et al., 2016; Rodger et al., 2023; Ryser, Yu, Grady, Siegmund, & Shibata, 2018). In particular, demethylation of developmental gene transcription factor (TF)-binding sites in distinct sites of the tumor suggests the involvement of these genes in tumorigenesis (Heide et al., 2022).

In addition, high enrichment of transcription factor binding profiles of SUZ12 EZH2, both core components of the polycomb repressive complexes 2 (PRC2) catalyzing trimethylation of histone H3 lysine 27 (H3K27), is reported in metastatic lesions compared to tumor (Rodger et al., 2023). Histone modifications and chromatin accessibility are an additional layer of intra-tumor epigenetic heterogeneity that characterizes CRC. For instance, the chromatin accessible sites in the primary tumor of CRC are different from those in liver metastasis, demonstrating a heterogeneous landscape of chromatin remodeling in tumor progression (Li et al., 2023)

ITH is prevalent in numerous cancers (Stanta & Bonin, 2018) including colorectal cancer, and is associated with poor prognosis and outcome (Becht et al., 2016; Guinney et al., 2015), marking this phenomenon as a potential prognostic factor (Yu et al., 2021). Indeed, targeting intratumor heterogeneity has been proposed as a way to suppress colorectal cancer chemoresistance and metastasis (Chao et al., 2023). Understanding the genetic and functional diversity is essential, as it profoundly impacts the disease's response to therapeutic interventions. Recent advances in characterizing molecular heterogeneity in CRC are emerging (Marisa et al., 2021) and more investigation is needed to improve patient prognosis.

1.7 Metastasis

1.7.1 Metastasis Formation*: A Complex Cascade*

Metastasis, the leading cause of death in CRC patients, remains a significant challenge in CRC treatment, as it involves a multifaceted cascade encompassing invasion, intravasation, circulation, extravasation, and colonization at distant sites (Gupta & Massagué, 2006). This cascade initiates with tumor cells detaching from the primary tumor, either individually or as clusters, and entering the bloodstream in a process known as intravasation. The hepatic portal venous system, responsible for collecting most intestinal blood, makes the liver the most common site of CRC metastasis (W. Zhao et al., 2022). Subsequently, these tumor cells travel through the circulation, eventually exiting the bloodstream at secondary organ sites where they can instigate the development of overt metastases (Valastyan & Weinberg, 2011).This multifaceted process

relies on intricate interactions between tumor cells and their microenvironment, with numerous factors influencing the success of metastatic dissemination. Generally, metastasis formation is considered highly inefficient, as only a minuscule fraction of tumor cells leaving the primary site will ultimately give rise to metastases. One study reported that merely 0.01% of circulating tumor cells had the capacity to establish metastases when injected into the bloodstream (Fidler, 1970; Luzzi et al., 1998).

The initial phase of dissemination, during which cancer cells migrate away from the primary tumor, entails alterations in cell adhesion molecules (CAMs) like cadherins, changes in cell-matrix adhesion molecules, and remodeling of the extracellular matrix (ECM) (Kai, Drain, & Weaver, 2019). Tumor cells must lose their connections with neighboring cells and invade the surrounding tissue, facilitated by the secretion of matrix-degrading enzymes (Kessenbrock, Plaks, & Werb, 2010) and the expression of migration-associated proteins (Hamidi & Ivaska, 2018). Subsequently, tumor cells have various routes to take, such as the lymphatic or hematogenic route, where they enter lymph or blood vessels, respectively. Other routes include extravascular spread, observed in conditions like colorectal cancer or ovarian carcinomas, where tumor cells are carried by peritoneal fluid (Lengyel, 2010; Pretzsch et al., 2019). Additionally, tumor cells can employ perineural spread, as seen in head and neck squamous cell carcinomas (Liebig, Ayala, Wilks, Berger, & Albo, 2009), extravascular migratory metastatic spread, as observed in melanomas (Lugassy, Kleinman, Vermeulen, & Barnhill, 2020), or spread through air spaces, as occurs in lung carcinomas (Kadota et al., 2015).

During hematogenous metastasis formation, tumor cells enter the bloodstream by traversing the endothelial cell layer, a process facilitated by angiogenesis, which increases the number of blood vessels in the tumor. Angiogenesis is also essential for sustaining primary tumor growth (Hanahan & Folkman, 1996). Modulation of the endothelial cell barrier, for example, through the secretion of factors like Vascular Endothelial Growth Factor (VEGF), helps tumor cells enter blood vessels (Weis & Cheresh, 2005). Once in the bloodstream, tumor cells travel with the blood flow and eventually become trapped in capillary beds, where they can exit the bloodstream by penetrating the endothelial cell layer. After reaching secondary sites, tumor cells often enter a period of dormancy before initiating colonization of the organ, ultimately giving rise to secondary tumors (Giancotti, 2013).

1.7.2 Epithelial-to-Mesenchymal Transition in Metastasis Formation

Tumor cells must undergo phenotypic changes that enable them to migrate from the primary tumor in their journey of the invasive spreading to variable microenvironments in secondary locations. (Markowitz & Bertagnolli, 2009; Tsai & Yang, 2013; Vanharanta & Massagué, 2013). These phenotypic changes are often associated with a cellular process known as Epithelial-to-Mesenchymal Transition (EMT), during which cancer cells transition from an epithelial state to a mesenchymal state, akin to processes seen during embryonic development or wound healing (Fleischmajer, 1967). This property is hijacked by cancer cells to facilitate phenotypic transitions leading to intravasation into the circulation, extravasation, and colonizing distant organs. The hallmark of EMT is the loss of epithelial surface markers, most notably E-cadherin, and the acquisition of mesenchymal markers including vimentin and N-cadherin, enhancing cell motility (Serrano-Gomez, Maziveyi, & Alahari, 2016) (Figure 4). This transition is primarily induced by extracellular soluble factors, such as transforming growth factor-beta (TGF-β), and various components of the extracellular matrix (ECM) (Hao, Baker, & Ten Dijke, 2019; Potts & Runyan, 1989). E-cadherin repression involves the orchestrated activity of specific transcription factors, including SNAI1, SNAI2, TWIST1, and ZEB1, resulting in alterations in the expression of cell adhesion molecules (CAMs) (Batlle et al., 2000; Eger et al., 2005; Hajra & Fearon, 2002; Mittal, 2018). EMT leads to the loss of cell polarity and the upregulation of genes associated with cell migration. These changes collectively enhance the ability of cancer cells to disseminate from the primary tumor, facilitating metastatic spread (Yang et al., 2020). Furthermore, it has been observed that the acquisition of mesenchymal features during EMT is associated with the acquisition of stem cell-like properties, including self-renewal capacity and resistance to chemotherapy or radiotherapy (Dart, 2023; Debaugnies et al., 2023; Mani et al., 2008; Serrano-Gomez et al., 2016).

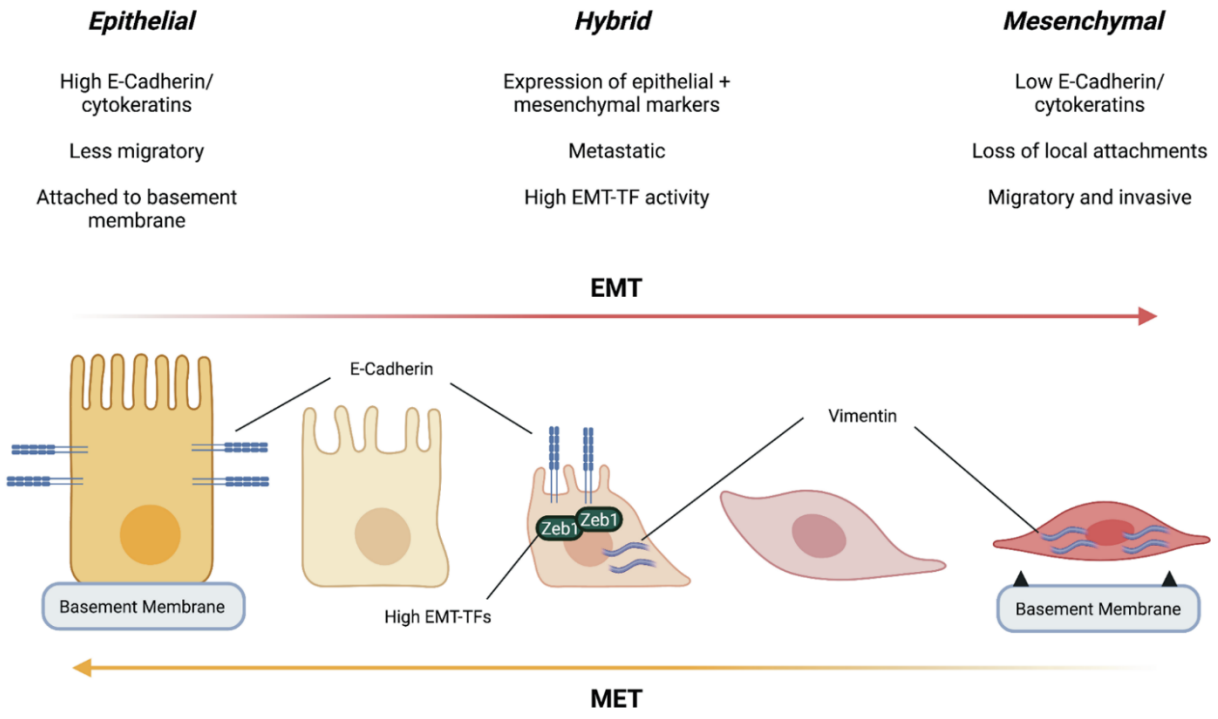


Figure 4. Epithelial to mesenchymal transition hybrid states. Loss of cadherins and cytokeratins characterizes loss of epithelial adhesive cell features and promotes an invasive mesenchymal phenotype. Adapted from (Evan, Wang, & Behrens, 2022) under the Creative Commons Attribution 4.0 International License permission (CCBY) of Springer Nature.

Notably, after cancer cells have exited the bloodstream at secondary sites, a reverse process known as mesenchymal-to-epithelial transition (MET) takes place (Hao et al., 2019). This supports cell to switch back to an epithelial phenotype in order to grow in the secondary site and become a detectable mass. The requirement for EMT and MET in the metastasis process remains a subject of discussion, as some studies have shown that metastasis can occur without EMT (Fischer et al., 2015; Revenco et al., 2019; X. Zheng et al., 2015). Moreover, the transition between epithelial and mesenchymal states is not always binary but can involve intermediate or hybrid states, leading to the concept of epithelial-mesenchymal plasticity (Evan et al., 2022; Yang et al., 2020). Hybrid EMT states have been recently reported to drive tumor progression and metastasis, but also as a plastic process functional in chemotherapy resistance (Dart, 2023; Simeonov et al., 2021).

1.7.3 Metastatic Cells: Trajectory and Destination

Metastatic organotropism, referring to the tendency of cancer cells to preferentially colonize specific secondary organs, presents a complex phenomenon. The 'seed and soil' theory, proposed by Stephen Paget in 1889, suggests that cancer cells necessitate specific

microenvironments akin to fertile soil for optimal growth, implying that the microenvironment at secondary sites significantly influences metastatic growth (Paget, 1989). Documented patterns of metastatic spread vary across organs; for instance, in colorectal cancer (CRC), liver metastases are frequently observed, while breast or prostate carcinomas often metastasize to bones (Hess et al., 2006).

Alternative theories, such as the anatomical concept proposed by Ewing in 1919, suggest that the vascular system's anatomy dictates the routes of metastasis, contributing to the non-random occurrence of metastasis (Ewing, 1919). Both the 'seed and soil' theory and vascular anatomy play roles in metastatic tropism. In CRC, the distinctive anatomy of venous blood flow guides blood through the hepatic-portal system before reaching the heart and lungs. Owing to their larger size compared to blood cells and the tendency of tumor cells to exit the primary tumor in clusters, tumor cells may become trapped in the initial capillaries they encounter, resulting in extravasation and the frequent observation of liver metastases in CRC patients (K. J. Cheung & Ewald, 2016). Certain studies propose that specific receptors on endothelial cells can trap tumor cells at particular organ sites, potentially leading to intravascular proliferation (Al-Mehdi et al., 2000; Ruoslahti, 2004). However, evidence also supports the idea that tumor cells can extravasate throughout the body without site-specific constraints (Chambers, Groom, & MacDonald, 2002). Nevertheless, some organs, such as the liver and lungs, appear more receptive to cancer cell extravasation, contributing to the concept of metastatic organotropism (Gao et al., 2019). The microenvironment at secondary sites plays a crucial role in determining whether disseminated tumor cells (DTCs) successfully reinitiate growth, with factors like growth factors, chemokines, metabolites, immune cell composition, and the stromal environment being key determinants (Lehuédé, Dupuy, Rabinovitch, Jones, & Siegel, 2016; Mohme, Riethdorf, & Pantel, 2017; Müller et al., 2001; Oskarsson, Batlle, & Massagué, 2014). To enhance conditions for metastatic outgrowth, tumor cells can influence the microenvironment, creating a premetastatic niche. This involves the secretion of factors such as the Vascular Endothelial Growth Factor (VEGF), S100 Calcium Binding Protein A8/9 (S100A8/9), or Matrix Metallopeptidase 9 (MMP9) in lung and melanoma (Hiratsuka, Watanabe, Aburatani, & Maru, 2006), or Tissue Inhibitor of Metalloproteinase (TIMP)-1 in colon and breast cancer or osteogenic N-cadherin (CDH2) in bone marrow (Y. Liu & Cao, 2016).

This long and complex process typically involves more than a decade to complete successive mutations of multiple genes in the primary tumor site, such as TGFBR2, SMAD4, PTEN, and RAS, and may spread early, long before the original tumor is detected (W. Zhao et al., 2022).

Numerous investigations have scrutinized mutational profiles in primary tumors versus metastases and have consistently found that metastases generally lack unique additional mutations common across various entities (Masoodi et al., 2020; Yachida et al., 2010). Instead, the majority of mutations were already present in the primary tumors. This suggests that epigenetic and transcriptional alterations may play a pivotal role in granting metastatic-initiating cells (MICs) the necessary adaptability (Hunter, Amin, Deasy, Ha, & Wakefield, 2018).

1.7.4 Clonality in Metastasis

In comparison to primary tumors, metastases generally composed of a smaller variety of genetic subclones (reduced subclonality) (Kerbel, Waghorne, Korczak, Lagarde, & Breitman, 1988; Siraj et al., 2020). Nevertheless, metastatic lesions display varying degrees of clonality. For instance, lymph node metastases tend to be polyclonal, implying that they originate from the primary tumor in multiple waves (Gerlinger, 2018; Naxerova et al., 2017). Conversely, ovarian or lung metastases in colorectal cancer (CRC) tend to be more monoclonal, suggesting a single-cell seeding mechanism (Siraj et al., 2020). Notably, the same study revealed a co-occurrence of both mono- and polyclonal metastases in CRC patients, with treatment-naïve metastases being predominantly polyclonal, while treated metastases exhibited a more monoclonal profile. While it is evident that multiple metastasis formation routes may occur simultaneously or sequentially from the same or independent subclones (Brown et al., 2018; Naxerova et al., 2017; Pereira et al., 2018; Quinn et al., 2021), additional research is required to gain a deeper understanding of the factors influencing metastatic tropism and even horizontal seeding from one metastasis into another.

1.8 Therapeutic Approaches in Colorectal Cancer

CRC treatment modalities have evolved significantly over the years. Standard treatments depend on the stage and location of the tumor and the overall health of the patient. It involves surgical intervention aiming at the complete resection of tumors and surrounding tissue, with subsequent management dictated by the stage of the disease. Adjuvant chemotherapy, often incorporating fluoropyrimidines and oxaliplatin (FOLFOX: 5-florouracil 5-FU, folinic acid and oxaliplatin) remains standard for reducing recurrence risk post-surgery (André et al., 2004). Targeted therapies that inhibit specific molecular pathways have revolutionized CRC management. Monoclonal antibodies, such as cetuximab and panitumumab, inhibit the epidermal growth factor receptor (EGFR), attenuating downstream signaling pathways. These agents demonstrate efficacy in RAS wild-type tumors, emphasizing the importance of molecular profiling in treatment decisions (Van

Cutsem et al., 2009). In advanced disease, anti-angiogenic agents like bevacizumab, targeting vascular endothelial growth factor (VEGF), have become integral components of chemotherapy regimens (Giantonio et al., 2007). The evolving understanding of CRC's molecular heterogeneity, including BRAF mutations, underscores the need for tailored approaches. Indeed a phase 2 clinical trial encompassing combined targeted therapy of PD-1, BRAF and MEK showed promising clinical outcomes in BRAF mutant patients (J. Tian et al., 2023). On the epigenetic level, inhibitors targeting pathways such as DNA methylation have been implemented in combination therapy and showed response in a subset of patients (10%) (Garrido-Laguna et al., 2013). For the subset of microsatellite instability-high (MSI-H) or mismatch repair-deficient (dMMR) tumors, immune checkpoint inhibitors, particularly pembrolizumab and nivolumab, have been implemented in the landscape of CRC treatment (André et al., 2020; Batlle & Clevers, 2017; Overman et al., 2017). This highlights immunotherapy as a promising avenue for CRC, leveraging the host immune system against malignant cells.

The contemporary therapeutic landscape for CRC embodies a dynamic interplay of surgery, traditional cytotoxic agents, targeted therapies, and immunomodulation, with ongoing research poised to refine and expand these approaches. Current studies focus on further unraveling the genomic landscape of colorectal cancer to identify novel therapeutic targets and refine personalized treatment strategies. The integration of these diverse therapeutic modalities marks a paradigm shift towards more combined, tailored and effective approaches in managing colorectal cancer, which despite these therapeutic advances is recurrent and still represents a global burden (Dienstmann et al., 2017; Morgan et al., 2023).

1.9 Drug Resistance in Cancer

Despite notable advancements in cancer therapy, the persistent challenge of drug resistance impedes the successful eradication of tumors in patients. CRC drug resistance can be attributed to different factors. Heterogeneous subclones with distinct genetic, epigenetic or phenotypic features within a tumor population may display variable responses to different lines of therapy and lead to tumor relapse despite an initial response (Marusyk, Janiszewska, & Polyak, 2020). In addition, adaptive resistance could emerge during prolonged selective pressure of drug exposure (Cayrefourcq et al., 2021). This could be attributed to the occurrence of additional mutations or to molecular adaptive mechanisms. Genetic alterations stand as well-established contributors to drug resistance mechanisms. For instance, cells treated with EGFR inhibitors may acquire MET amplification, perpetuating downstream signaling and proliferation (Engelman et al., 2007). Additionally, point mutations such as T790M in EGFR can hinder drug binding, exemplified in

non-small cell lung cancer (NSCLC) (Yun et al., 2008). The inactivation of tumor-suppressor genes also emerges as a driving force behind resistance (Vasan et al., 2019). Adaptive molecular mechanisms for therapy resistance include EMT, anti-apoptosis, detoxification from therapeutic drugs, expression of DNA damage checkpoints and cell dormancy (L. Chen, Yang, Chen, & Tai, 2022).

1.10 Drug Tolerant Persister Cells in Cancer

Drug-tolerant persister cells (DTPs) represent a unique subset of cancer cells capable of surviving drug treatments (Sharma et al., 2010). These cells constitute a small fraction (0.3–5%) of the parental tumor and exhibit a quiescent or slow-cycling state (Guler et al., 2017; Liau et al., 2017). DTPs can transiently tolerate high drug levels, similar to bacterial persisters (Kaushik, Sharma, Tiwari, & Tiwari, 2022), without underlying genetic mutations, thereby representing a reversible state that resumes growth upon treatment removal, yet remains chemotherapy-sensitive (Recasens & Munoz, 2019). This offers a potential therapeutic avenue before the onset of classical irreversible genetic mutation-driven drug resistance. Studies, including pioneering work by Sharma et al., have identified DTPs in cancer cell lines and demonstrated their selective ablation by targeting Insulin-like growth factor 1 receptor (IGF-1R) or employing chromatin-modifying agents (Sharma et al., 2010). Other investigations highlight the role of stress-induced long interspersed repeat element 1 (LINE-1) repression and dependency on lipid hydroperoxidase GPX4 in driving the DTP state (Hangauer et al., 2017). While initial studies focused on *in vitro*-generated DTPs, recent work in triple-negative breast cancer using patient-derived xenografts (PDXs) revealed a subpopulation capable of repopulating tumors after a DTP state, suggesting phenotypic properties may govern regrowth capacity without genomic selection (Echeverria et al., 2019). Another recent study carried out in CRC PDXs described a diapause-like state termed DTP (dormancy-like, therapy resistant) which enables cells to survive chemotherapy (Rehman et al., 2021). This state resembles an embryonic transcriptional stage that reflects stemness maintenance, transient cell cycles arrest, autophagy upregulation and metabolic reprogramming. Molecular pathways of MYC and mTOR signaling are reduced, while ERK pathway is activated in DTP cells (Dhimolea et al., 2021; Rehman et al., 2021). This mode of treatment resistance resembles a reversible form of non-genetic tumor heterogeneity that have been identified in various cancers in response to both chemotherapy and targeted agents. Understanding the molecular mechanisms underpinning DTP survival is essential to overcome drug resistance and enhance treatment efficacy.

1.11 Embryonic/fetal Programs in Cancer

Activation of fetal or embryonic genetic programs has recently emerged as a mechanism of tumor progression and drug resistance (Fazilaty & Basler, 2023; Heide et al., 2022). These embryonic programs facilitate cell plasticity and re-activation of regenerative stem cell features that allow cancer cells to adapt to stress (Shivdasani, Clevers, & de Sauvage, 2021). In particular, DTPs have been linked to active embryonic signatures and diapause-like states (Dhimolea et al., 2021; Rehman et al., 2021). Diapause is an embryonic reversible condition that the embryo undergoes in its development, allowing for a postponement of the implantation process as an adaptive strategy to synchronize development with favorable environmental conditions (Fenelon, Banerjee, & Murphy, 2014). This state is characterized by the depletion of the expression of transcription factors c-Myc and Akt, in addition to Rictor, a component of mTORC2. The transcription factor and stress response gene EGR1, and glutamine transporters Slc38a1 and Slc38a2 are highly upregulated during diapause, and histone marks associated with active transcription H3K4me3, H3K36me2 and H4K5/8/12/16 acetylation are reduced while H3K27me3 is activated (Hussein, Balachandar, Mathieu, & Ruohola-Baker, 2022). This scenario is reminiscent in treated cancer cells, which enter embryonic-like states to tolerate therapy and re-multiply in less stressful conditions. Experimental reversal of this state by MYC activation or chemical inhibition of glutamine transporter Slc38a1/2 sensitized cancer cells to chemotherapy (Dhimolea et al., 2021; Hussein et al., 2020). Early embryonic programs include pluripotency states which confers cells with stemness features that can drive differentiation to variable fates. This property can be hijacked by cancer cells to drive tumor initiation and progression (Fazilaty, 2023). Cancer cells acquire variable features during tumorigenesis such as invasion, migration, adhesion and proliferation which can be reprogrammed through stemness or embryonic factors. For instance, the activity of the pluripotency transcription factor SOX2 influences both epithelial-mesenchymal transition (EMT) associated with invasion and the activation of developmental factor YAP linked to cell division (Pastushenko et al., 2021). This functional developmental plasticity in cancer progression can be reflected through a less differentiated phenotype in metastasis compared to the primary tumor (Laughney et al., 2020). Sox transcription factors, especially SOX2 and SOX9 expressed in embryogenesis and in stem cells, exert regulatory functions in tumor progression and metastasis (Jo et al., 2014). YAP proteins are transcriptional co-activators which direct cell migration in embryonic development (Sousa-Ortega et al., 2023), promote cancer metastasis by interacting with TEAD transcription factor family (Lamar et al., 2012). Among other embryonic genes active in cancer development are S100A6, LY6A, AREG and LRG1 (Fazilaty & Basler, 2023).

1.12 Plasticity: The Key to Survival

1.12.1 Plasticity of Crypt Cells

The intestinal crypt is characterized by phenotypic heterogeneity. Cells within the crypt exhibit diverse features along the differentiation axis of stem cells in the bottom of the crypt to fully differentiated cells on the top (Figure 1A) (de Sousa & de Sauvage, 2019). However, this differentiation can be reversible by which differentiated cells at the top of the crypt restore the stem cell compartment upon stress or tissue damage, possibly caused by different factors such as inflammation or irradiation (Figure 1B) (de Sousa & de Sauvage, 2019). This plastic ability of self-renewal and regeneration is suggested to be regulated by different mechanisms such as the fetal conversion caused by YAP/TAZ, the Notch signaling or through the open chromatin status (de Sousa & de Sauvage, 2019; Jadhav et al., 2017). Upon environmental stimuli, differentiated intestinal cells can be replaced by other cell types in a process named metaplasia (Giroux & Rustgi, 2017) through regenerative genes such as SOX9 (Willet et al., 2018). This metaplastic process is a precursor of dysplasia, which is linked to cancer. In the context of metaplasia, paligenosis has been described as an evolutionary conserved cellular program of regeneration in response to damage, mediated by DDIT4 (Q. Miao et al., 2019; Willet et al., 2018). Mechanisms governing such plasticity in normal intestinal epithelial cells, where CRC derives, can be possibly hijacked in a more robust fashion by cancer cells which are continuously challenged by hypoxic stress, varying spatial localization and anti-cancer treatment.

1.12.2 Cancer Cell Plasticity

Plasticity, the ability of cells to reprogram and transition between different functional phenotypes, is a characteristic of cancer cells that can be exploited for treatment resistance, tumor progression and metastasis (Z. Guo et al., 2023). In addition, phenotypic plasticity of cancer cells boosts intra-tumor heterogeneity on top of genetic heterogeneity (Househam et al., 2022). This adaptive capacity allows cancer cells to evade treatment and poses a significant challenge in cancer therapy. In melanoma, cells resist targeted and immune therapies by dedifferentiation to a neural crest state (Kim et al., 2021) or by transcriptional reprogramming (Shaffer et al., 2017). EMT for instance, is a plastic process related to tumor initiating cells (TICs), which are effective in driving metastasis and treatment resistance (Tsai & Yang, 2013). Reverting mesenchymal cells to an epithelial phenotype can reduce the tumor initiating capacity and progression of cancer (Pattabiraman et al., 2016). In addition, targeting EMT exposed tumor cells to immunotherapy treatment (Dongre et al., 2021). Microenvironmental changes related to treatment or spatial

localization of cancer cells can induce plastic gene expression programs (Lenos et al., 2018) and influence the proliferation of individual clones over others (van der Heijden et al., 2019). Such findings reflect the plasticity of cancer cells in responding to environmental factors and further highlights its importance in directing tumorigenesis (Burdziak et al., 2023).

In attempt to impair tumor progression, researchers targeted the stem cell compartment of cancer cells, which fuels tumor progression and self-renewal. Indeed, ablation of LGR5+ CRC stem cells can lead to tumor regression, but shortly after, LGR5+ cells appeared from previously LGR5- cell compartment and restored tumor growth, in a striking example of the significance of plasticity in tumorigenesis (Shimokawa et al., 2017). In a more recent work, LGR5- cells were shown to dominate disseminated CRC cells in circulation and form distant metastasis, where LGR5 expression was restored, in a demonstration of the plastic properties of CRC cells (Fumagalli et al., 2020). This plasticity poses a major challenge in cancer treatment and more research is needed to efficiently manage this phenomena.

1.13 Model Systems to Study Colorectal Cancer Tumors

1.13.1 *In Vitro* Cell Cultures

In vitro cell culture serves as an indispensable tool for biological research and for studying tumors in particular. More than 100 cell lines representing diverse tumor stages and genetic subtypes are extensively used for colorectal cancer research in both 2D and 3D settings (L. Liu et al., 2023). Cell lines are amenable to genetic modification, treatment and diverse experimental settings such as hypoxia and migratory assays. However, the inherent homogeneity of cell lines limits their ability to faithfully represent the intricate heterogeneity observed in patient tumors. For this reason, our group has used primary patient-derived tumor spheroid cultures (TSCs) from CRC tumors or metastases, grown under non-adherent, serum-free conditions with epidermal growth factor (EGF) and basic fibroblast growth factor (b-FGF). Patient-derived TSCs display functional heterogeneity, reflecting the diversity of tumor-initiating cells (TICs) (Dieter et al., 2011; Ricci-Vitiani et al., 2007).

3D organoid cultures represent a significant advancement in patient-derived tumor models (Sato et al., 2009). Organoid cultures derived from tumors can recapitulate some characteristics of the respective patient on the genetic, transcriptional and methylation levels (Roerink et al., 2018). In addition, normal CRC intestinal derived organoids mimic the structure of the intestinal crypt and generate villus-like domains containing differentiated cell types. Cells either with stemness

(LGR5+) or differentiated features are represented in patient-derived organoids (PDOs) (Sato et al., 2011). These cultures can be expanded for long-term and used for functional studies, drug screening and genetic manipulation. In particular, CRC derived PDOs provide a more representative model for pre-clinical studies than other *in vitro* cell culture models through relatively enhanced molecular resemblance to the heterogeneity of original tumors (Sasaki & Clevers, 2018).

1.13.2 *In vivo* Animal Models: Xenografts

Although *in vitro* cell lines and patient-derived cultures represent a fundamental tool in cancer research, they do not recapitulate the spatial structure of the tumor, nor the physiological microenvironment surrounding cancer cells. To overcome this limitation, CRC can be effectively modeled *in vivo* using (PDXs), which show superiority in mimicking structural and physiological conditions of cancer cells than *in vitro* systems (Rizzo, Bertotti, Leto, & Vetrano, 2021). PDXs are created by implanting patient-derived tumor samples into immunodeficient mice, such as NOD.Cg-PrkdcscidIl2rgtm1Wjl/SzJ (NSG) or other strains (Okada, Vaeteewoottacharn, & Kariya, 2019). Establishment of PDXs involves transplantation of small tumor pieces or cell suspensions of the tumor fragments, obtained through mechanical or enzymatic digestion. This maintains the genetic heterogeneity of the patient's tumor. In addition, PDXs provide suitable physiological microenvironment for cancer such as stromal cells and blood access. This environment maintains oxygen, nutrients and hormones' levels needed for tumor cell survival (Rizzo et al., 2021). Depending on the research goal, tumor cells are injected subcutaneously into the mouse's flank or orthotopically, mimicking the tissue of origin. While subcutaneous injections facilitate easy tumor size measurement, orthotopic models more closely emulate the tumor microenvironment and permits metastasis development through physiological routes (Alamo et al., 2014). Comparative studies between PDX tumors and original tumors revealed similarities in histology and genetic characteristics, even across serial passages (Hidalgo et al., 2014). PDX models can exhibit correlations between drug responses in the models and patients' responses, making them considerably valuable for translational research. However, the use of immunocompromised mice in PDX models restricts the study of the tumor microenvironment including bacterial and immune cells since they lack immune cell populations and live in a sterile environment (Radaelli, Santagostino, Sellers, & Brayton, 2018). In addition, PDX models face challenges such as low engraftment rates and extended tumor growth periods, (Bürtin, Mullins, & Linnebacher, 2020). Consequently, the advantages and the limitations of this model should be carefully considered in preclinical testing.

1.14 Aim of the study

The TIC activity in CRC is not dependent on individual genetic subclones (Dieter et al., 2011; Giessler et al., 2017; Rehman et al., 2021), and emerging evidence indicates that epigenetic factors, especially transcriptional states may play a pivotal role in clonal TIC activity (Household et al., 2022; Ryser et al., 2018; Zowada et al., 2021). Epigenetic changes in response to micro-environmental stress such as chemotherapeutics may serve as key determinants of the phenotypic plasticity observed in cancer (Flavahan, Gaskell, & Bernstein, 2017; Rehman et al., 2021). Such studies have utilized cellular marking combined with NGS sequencing, in addition to retrospective clonal mapping to describe clonal dynamics and associate epigenetic transcriptional events to CRC progression. These studies mainly relied on either cell lines, bulk sequencing, or subcutaneous xeno-transplants. Despite progress in characterizing functional heterogeneity of CRC cell populations and identifying relevant transcriptional states, exploration of transcriptional programs governing sub-clonal expansion and heterogeneity at the single-cell level remains limited. More investigation is required to understand the association of transcriptional states to clonal dynamics in metastatic spread and therapy response. Accordingly, this thesis aims at:

1- Deciphering clonal dynamics i.e. expansion or reduction in metastasis and treatment response to depict individual clonal contribution to tumor progression
2- Untangling gene expression states in subclones associated with metastasis and chemotherapy response on the single cell level
3- Linking clonal dynamics to transcriptional states identified in chemotherapy response and metastasis, to associate gene expression profiles to clonal expansion

Genetic barcoding and tracing of patient-derived cancer cells in combination with single cell RNA sequencing was utilized to tackle these aims. By dissecting the single cell diversity of gene expression in tumor and metastasis-forming subclones in real-time, this work aims at a better understanding of the transcriptional heterogeneity of CRC. Targeting transcriptional states associated with metastatic spread and treatment response can be a suitable approach to impede cancer progression and drug resistance. This work contributes to the understanding of transcriptional mechanisms characterizing distinct cell phenotypes observed in CRC subclones, in an attempt to facilitate their eradication.

2. Results

2.1 Contributions

All experiments, data analysis and interpretation were performed by myself unless indicated here.

Patient-derived culture models have been established in our group before. For characterization, I embedded PDOs or PDO xenografts in agarose or paraffin, while further immunohistochemistry procedures were performed by the Institute of Pathology at the University Hospital Heidelberg and the National Center for Tumor Diseases tissue bank (NCT Gewebebank). Further characterization of PDOs through FACS and 3D cell titer glow drug assay was performed by myself and our master intern, Pravin Velmurugan, under my supervision. Panel sequencing of organoids was done at the Core Unit for Molecular Tumor Diagnostics at TU Dresden and data were analyzed and visualized by Dr. Stephan Drukewitz or Dr. Attila Jady. Dr. Stella Arelaki performed immunofluorescence staining of paraffin embedded slides of xenografts. Single cell sequencing experiments were performed by myself at the single cell open lab (scOpenLab) of the DKFZ headed by Dr. Jan-Philipp Mallm, who supported the amplification of tracing barcodes in the single cell RNA (scRNA) library preparation, and sequencing was performed by the Genomics and Proteomics Core Facility (GPCF) and the Omics IT and Data Management Core Facility (ODCF) of the German Cancer Research Center (DKFZ). Bioinformatic analysis of sequencing data was performed by Dr. Attila Jady. Dr. Tim Holland-Letz from the Division of Biostatistics supported the statistical analysis of the clonality score of tracing barcodes and the cluster enrichment analysis in scRNA data. Dr. Alexander Wurm and Vida Kufrin performed and analyzed the custom CRISPRi library knock down screen of selected genes.

Disclaimer: I have drafted a part of the work presented in this thesis as a first author for a manuscript that is currently in preparation for submission to a scientific journal. I produced figures related to barcode analysis and single cell RNA data in this thesis and in the drafted manuscript with bioinformatics support from Attila Jady.

2.2. Patient-derived organoid xenotransplantation results in tumor formation and metastatic spread

To examine subclonal dynamics in CRC tumor progression and metastasis, I used patient-derived organoid (PDO) cultures from tumors (colon or rectum) or liver metastases in generating xeno-transplants as a study model (Supplementary table 1). To evaluate the functional capacity of

these PDOs in terms of tumor-forming potential and metastatic burden *in vivo*, I transplanted 500x10E3 freshly dissociated cells from eight organoid cultures (OC). Transplantations were performed orthotopically into the caecal wall of 2 to 6 immuno-deficient NOD.*Cg-Prkdc*^{scid}*Il2rg*^{tm1Wjl}/SzJ (NSG) mice for each culture by means of surgical procedure. This allows CRC cells to grow in their counterpart mouse microenvironment and to further disseminate into distant organs such as the liver (Alamo et al., 2014). To ensure the precise positioning of cells into the epithelial lining of the caecum and avoid surplus into the peritoneal cavity, the procedure involved precise injection between the serosal and sub-mucosal layers of the caecal wall, supported by stereomicroscopic magnification of 20X and application of a seprafilm adhesion barrier to ensure the integrity of the epithelial lining of the caecum after injection (Fumagalli et al., 2018).

Over a duration of 2 to 6 months, I observed that seven out of the eight cultures grew tumors in-vivo at varying frequencies (Figure 5). Specifically, OC3, OC7, OC13, OC17, OC19 exhibited primary tumor growth within a time frame of 2 to 3 months, while the remaining cultures grew primary tumors after a minimum of 5 months. OC15 and OC2 displayed the lowest growth frequency of 0 (0/3 mice) and 16.7% (1/6 mice), respectively, without metastatic occurrences in distant organs. Consequently, I excluded these two cultures from subsequent experiments. OC3, OC13, OC17 and OC19 showed a 100% rate of tumor growth (6/6, 3/3, 3/3, 2/2 mice) and corresponding liver metastasis formation rate of 100, 66 and 50% (6/6, 2/3, 2/3, 1/2 mice) respectively.

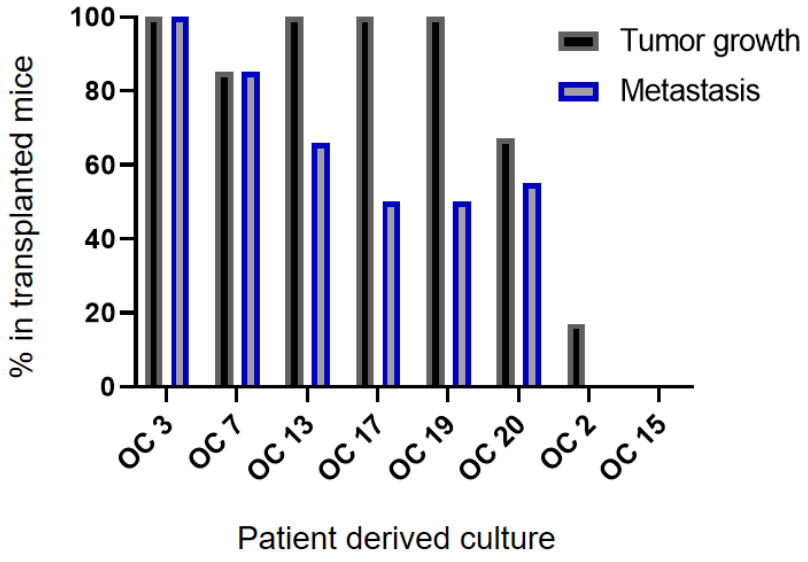


Figure 5. Tumor and metastasis formation capacity of PDOs in xenotransplantation. Barplot showing the % (y-axis) of tumor engraftment (in black) and metastasis formation (in gray) of each organoid culture OC (x-axis).

OC7 and OC20 showed a growth rate of 83% and 66.7 % (5/6 and 4/6 mice), respectively, along with liver metastasis rates of 83% and 50% (5/6 and 3/6 mice), respectively. This reflects the heterogeneity of this PDO small cohort in terms of xenograft engraftment capacity and duration, in addition to metastasis formation. This provided valuable information for the selection criteria of cultures included in subsequent experiments. Cultures that derived from treatment-naïve patients and demonstrated engraftment and metastasis formation in xenografts (OC7, OC13 and OC17), were selected for subsequent experiments.

2.3. Generation of a uniquely labeled cancer cell pool by lenti-viral barcoding of PDOs

To elucidate clonal dynamics and sub-clonal transcriptional changes characterizing intra-tumor heterogeneity in metastatic tumor progression and response to treatment, I employed low-passage organoid cultures derived from the three selected treatment-naïve CRC patients whose xenografts developed metastasis, in a genetic barcoding approach combined with single cell RNA sequencing (scRNA seq). These selected cultures were designated as PDO1, PDO2 and PDO3 (Supplementary table 1).

Pairing individual cell barcoding with scRNA seq enables tracking individual subclones and dissecting their transcriptome simultaneously. I utilized a commercially available clone-tracker barcode library (CloneTracker™ XP-RFP 10M Barcode-3'), which harbors up to 10x10E6 unique RNA-expressed barcode sequences, permitting their detection via scRNA seq. Barcodes which are lentivirally introduced into cells integrate into the genomic DNA, and therefore pass to daughter cells with cell division. This creates a uniquely labeled cell pool, which upon cell division generates cell progenies carrying the same barcode, designated as clones. The progression of such clones can be captured by NGS methods such as DNA amplicon or scRNA sequencing. In addition, the barcoding transfer vectors encode for a red fluorescent protein (RFP) used for identification and potential selection of labeled cells.

I transduced 1x10E6 cells of each of the three selected PDOs with the barcode library at a multiplicity of infection (MOI) < 0.25, to limit the viral integration to one barcode per cell, and thus generating uniquely barcoded cells. Two days after transduction, I sorted cells for RFP expression by FACS (Fluorescence-activated Cell Sorting) (300 x10E3, 110 x10E3 and 50 x10E3 cells from PDO1, PDO2 and PDO3 respectively) (Figure 6A, Supplementary figure 1). Sorted transduced cells were passaged for two to four weeks until a total number of 8x10E6 cells was achieved, thus

generating a stably barcoded pool of cells per patient, namely baseline cell pool (t0**;** Figure 6A) (PDO1: 2 weeks PDO2: 3 weeks, PDO3: 4 weeks). The barcode RFP signal was maintained at > 90% of the total PDO cell population 8 weeks after lentiviral transduction and RFP sorting, demonstrating the stability of barcode expression over the course of the experiment (Supplementary figure 2).

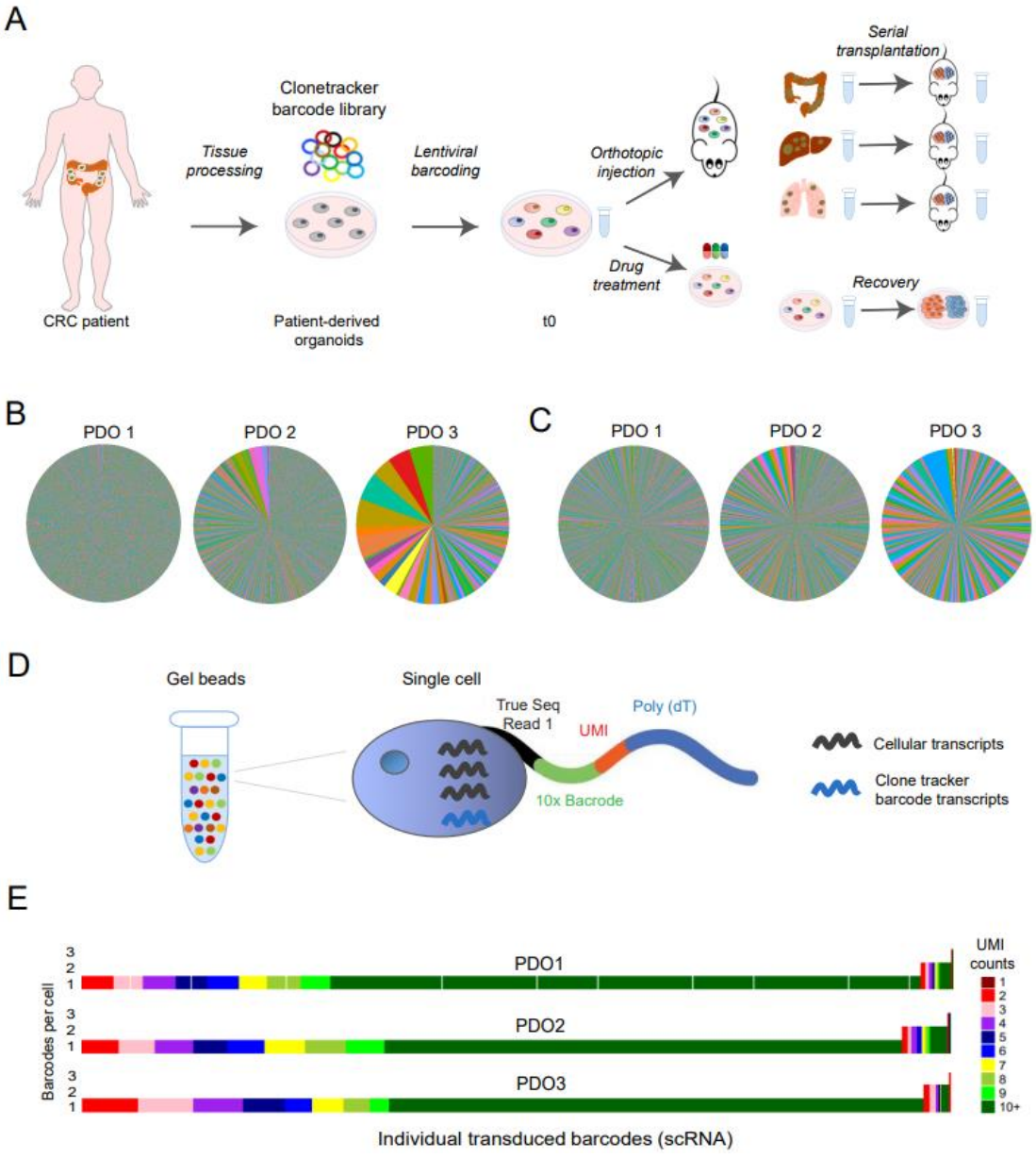


Figure 6. Tracing PDO clones through stable genetic barcoding. A. Experimental layout of PDO barcoding, xenotransplantation and *in vivo* drug treatment of barcoded PDO cells. B. Pie charts representing the clonal distribution (in %) of baseline cell population captured through targeted DNA (amplicon) sequencing. Barcodes representing less than 0.01% of all barcodes are collapsed into gray color, and barcodes with less than 10 reads were omitted for visualization purposes. C. Pie charts

representing the clonal distribution (in %) of baseline cell population captured through targeted scRNA sequencing. Individual clones detected in both approaches share the same color. D. Capturing endogenous cellular RNA transcripts in addition to clone tracker barcode transcripts in a single cell using the 10x scRNA library preparation protocol. Gel beads contain cell emulsions of single cells tagged to the 10x cell identifier barcode sequence. True Seq Read1: Adaptor sequence, UMI: unique molecular identifier, Poly (dT): polyadenylated tail. E. Number of barcodes and barcodes per cell detected through scRNA seq. I produced subfigures B, C and E with support from Dr. Attila Jady and adapted them within manuscript in preparation.

The baseline cell pool (t0**;** Figure 6A) was mixed and 1x10E6 cells were used for targeted barcode DNA sequencing. Additional 1x10E6 cells were used for 10x scRNA sequencing to assess the clonal composition and gene expression profile of the starting baseline cell population. I observed that targeted barcode DNA sequencing of baseline barcoded PDOs revealed a high number of unique barcodes (260 x10E3, 18 x10E3 and 4x10E3 in PDO1, PDO2 and PDO3, respectively) in each culture. Each barcode represented 3.3x10E-5 to 0.01% of the total cell population in PDO1, 7.1x10E-5 to 0.01% in PDO2, and 1.2x10E-4 to 5.9% in PDO3 (Figure 6B). This indicated a highly complex but homogeneous distribution of barcodes among cells of PDO1 and PDO2, while PDO3 barcode distribution was less homogeneous with a higher abundance of several clones, due to a lower starting barcoded cell number and longer passaging to reach an equal cell number to the other two cultures.

To monitor the distribution of barcodes on the single cell level, I amplified the barcode sequences out of the cDNA produced in scRNA library preparation, by means of targeted PCR. As all transcripts of a cell in 10x scRNA library preparation, transduced barcode transcripts were assigned to individual cells through unique 10x cell identifier barcodes, which are linked to all transcripts of a given cell (Figure 6D). 2183 to 6875 cells per culture were measured in the baseline population, whereby a unique barcode was detected in at least 95% of individual cells (Figure 6C). Each barcode in PDO1 and PDO2 represented a maximum of 0.6% of the total cell population and a maximum of 5% in PDO3 (Figure 6C). This observation is concordant with the barcode distribution detected by DNA barcode sequencing, whereby barcode detected by scRNA seq were also present among those detected by DNA seq (Figure 6 B, C). In addition, more than 85% of cells displayed one barcode integration per cell in baseline samples (5117/5929 in PDO1; 6666/6875 in PDO2; 1881/2183 in PDO3) (Figure 6E), indicating predominantly single vector integrations during transduction. Cells with multiple barcodes were removed from scRNA analysis, and cells sharing the same barcode were allocated as clones.

2.4.1 Clonal dynamics in tumor and metastasis formation *in vivo*

To elucidate the dynamics of clonal expansion and the heterogeneity of clonal gene expression in tumor progression and metastasis, I serially transplanted dissociated barcoded CRC PDOs orthotopically into NSG mice. In each mouse (n = 8 per patient culture), 250 x 10E3 barcoded PDO cells of t0 were injected into the caecal wall of the mice (Figure 6A). Tumors grew in the caecum within 8 to 16 weeks after transplantation, and at the timepoint of tumor harvesting, I observed metastases in the liver and lung (Figure 7). Cancer cells from primary tumors and metastatic sites, specifically the liver and lung, were harvested separately from two primary xenografts per patient for subsequent analysis of barcodes and gene expression. scRNA library preparation involved FACS sorting of RFP positive living cells to ensure capturing living barcoded cells (Supplementary figure 3). Cancer cells from organs of each of the two primary mice (caecum, liver and lung), were serially transplanted into the caecum of secondary mice (Figure 6A). One set of serially transplanted mice (one tumor-derived, one liver metastasis-derived, one lung metastasis-derived) were further analyzed for barcode composition and gene expression.

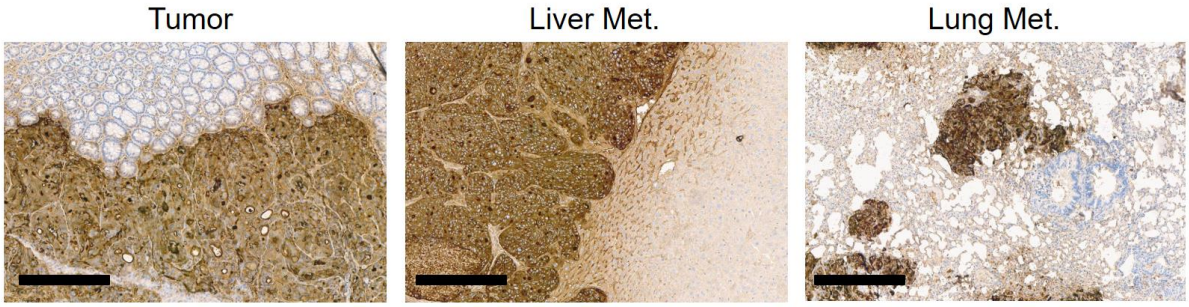


Figure 7. Tumor progression and metastasis in xenografts. Immunohistochemistry IHC staining of Carcinoembryonales Antigen CEA in primary tumor in addition to liver and lung metastasis of a PDO1-derived xenograft. Brown color corresponds to tumorous areas while lighter color depicts stromal cells. Scale bar: 300µM.

2.4.2 Clonal selection in serial xenografts and derived metastases

A. Barcode target DNA sequencing

Primary tumors

Barcode DNA sequencing of the baseline samples prior to transplantation at t0 revealed the presence of 4 x10E3 - 260 x10E3 barcodes per sample (260 x10E3 in PDO1, 180 x10E3 in PDO2, 4 x10E3 in PDO3). In contrast, primary xenografts harbored 2.2 x10E3 to 6.6 x10E3 barcodes

(4.7 x10E3 in PDO1, 6.6 x10E3 in PDO2, 2.2 x10E3 in PDO3), indicating a drop in the number of unique barcodes forming the total sampled cell population in tumors of primary xenografts, compared to the baseline samples of PDOs prior to transplantation (Supplementary table 2). While individual clones formed a maximum of 0.01% at t0 in PDO1 and PDO2 indicating a similar distribution of these clones (Figure 6B), ten individual major barcodes (barcodes representing more than 2.5% of total cell population i.e. > 25 x10E3 cells per sample) represented more than 50% of the total cell population in primary xenografts of PDO1 and PDO2 (Figure 8, Supplementary figure 4). In PDO3, 7 major clones at t0 formed 32% of sampled cells while 3 major barcodes formed more than 50% of tumor cells in the primary xenograft (Figure 8). This analysis I performed indicates a heterogeneous abundance of individual clones in xenografts, with an enrichment of some clones compared to others.

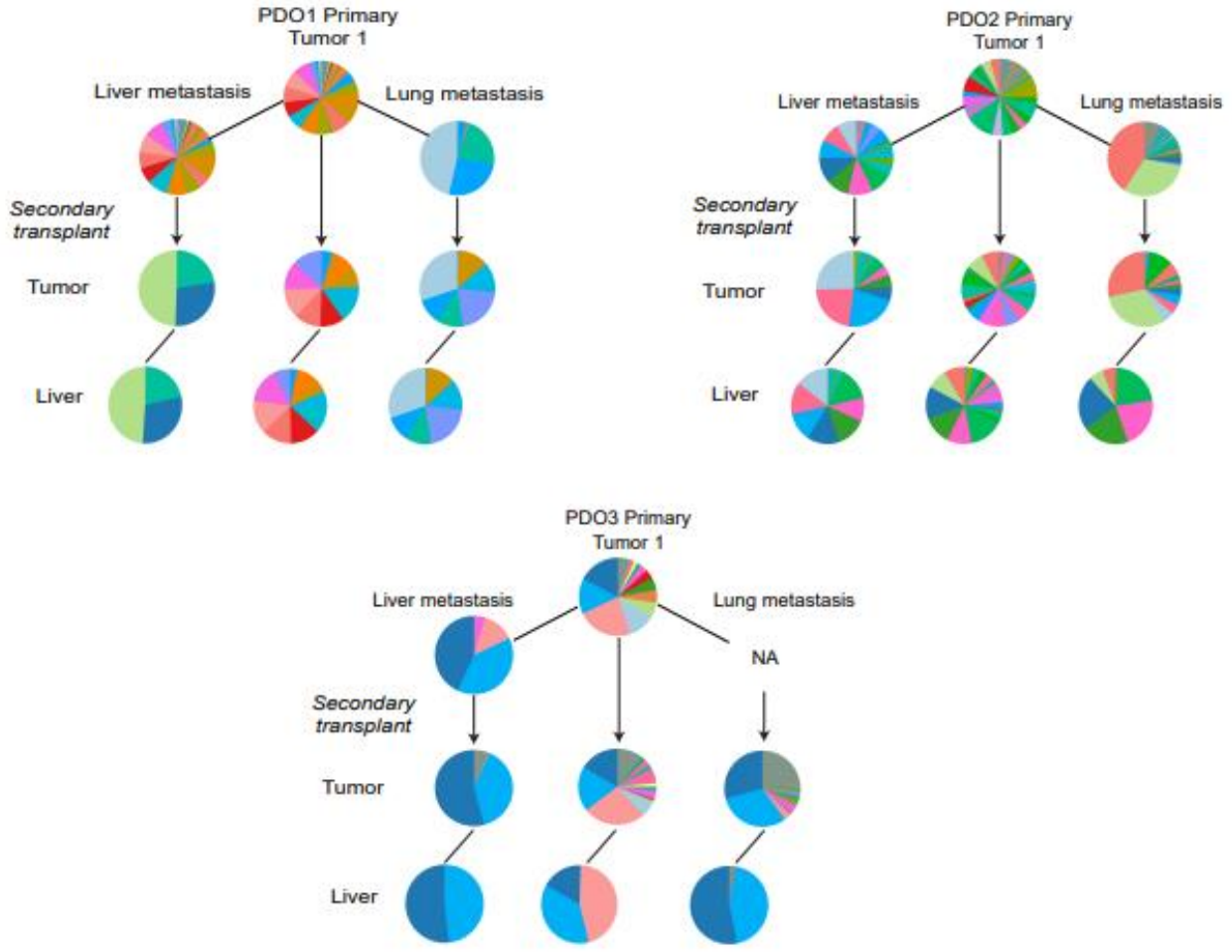


Figure 8. Clonal distribution in barcoded xenografts. Pie charts depicting the barcode distribution (in %) in PDXs captured through targeted DNA (amplicon) sequencing. One primary tumor per PDO is represented, in addition to metastases and serial transplantations. Each color portrays a unique clone. Gray color represents clones of low abundance (less than 0.1% of total barcode count). NA: non available. I

produced this figure with bioinformatics support from Dr. Attila Jady and adapted it within the draft of manuscript in preparation.

Metastasis

In harvested liver metastases spots of each of the two PDO1 xenografts, major clones (barcodes) demonstrated a frequency of 5 to 7% of the cell population, with a similar representation in the tumor (Figure 8, Supplementary figure 4). In liver metastasis of PDO2, 2 major clones represented in blue and pink color showed a frequency of 5-6% of the cell population compared to less than 1% in tumor. In liver metastasis of PDO3, each of the two major blue clones represent 37.5% of the cell population compared to less the 15 % in tumor. While red clones compose 2 to 3% in tumors of PDO2 and PDO3, they form less than 0.1% of cells in liver metastasis (Figure 8). This suggests a dynamic expansion of individual clones from tumor to liver metastasis, by which some clones maintain their contribution and others diminish, while some previously smaller clones expand. In lung metastasis of PDO1 or PDO2, major clones (3 clones in PDO1 and 2 in PDO2) formed up to 70% of all captured cells while they composed less than 7% of cells in tumor. This indicates that a few clones form the bulk of lung metastasis in each xenograft in comparison to the tumor which displays a more diverse clonal distribution.

Serial transplantation

In serial transplantation of tumor cells, I observed that major clones forming the tumor of secondary mice were found to be previously detected in the tumor of primary xenograft generation in the three PDOs (Figure 8). While major clones of PDO1 primary xenograft formed 5-7% of the cell population, those of the secondary tumor formed up to 12.5 % of cells. In PDO2 and PDO3, clones of serially transplanted tumor cells reflected a similar distribution compared to the primary xenograft tumor (Figure 8). In serial transplantation of liver or lung metastasis, only two to three clones collectively accounted for 66-98 % of the cell bulk in second generation xenografts. This suggests that clonal dominance increases with tumor progression in serial transplantation, especially in liver and lung metastasis, whereby major clones maintain their expansion during progression and others tend to diminish. A proportion of clones forming less than 1% of the total cell population was observed in all mice (collectively forming up to 25% of cells), indicating that clones of small size are maintained even in the presence of major clones.

B. Barcode scRNA sequencing

Primary tumors

To depict clonal dynamics and understand subclonal gene expression in tumors on the single cell level, scRNA barcode analysis was carried out in PDO xenografts. scRNA data of harvested cells shows that the number of detected unique barcodes (clones) derived from the tumor is 55, 92 in and 26 in 1,668 cells of PDO1, 5,082 cells of PDO2 and 4,539 cells of PDO3, respectively. This suggests that the tumor cell population comprises a relatively low number of clones compared to the total number of cells analyzed and that the tumor bulk is constituted of individual clones, whose number accounts for 0.5-3% of the total cell number (Figure 9).

To measure the clonal complexity of PDOs and xenotransplants with scRNA seq, I introduced a clonal diversity score, termed CS i.e. number of unique clones/total number of cells per sample. Downsampling was used to account for variations in cell numbers, by dividing cells into subgroups equal in size across different samples, and to ensure statistical measurement, which was not feasible in bulk barcode data. CS serves as an indicator of the clonal heterogeneity of the cell population, whereby a CS of 1 suggests a clonally diverse cell population where almost each individual cell belongs to a separate clone. In contrast to the baseline cell pool which had a CS of ~1, single cell sequencing of primary xenografts showed CSs below 0.25, indicating a significant decrease in the clonal complexity of engrafted tumor cells (Figure 10). In each of the two xenotransplanted primary mice per patient, one to three major clones were detected in more than 50% of the total cell population within tumors or metastases (Figure 9, Supplementary figure 4, Supplementary table 3), while each unique clone formed a maximum of 0.1% of the cells in baseline samples of PDO1 and PDO2, or 7% in PDO3 (Figure 10, figure 6C). This indicates that few clones can outcompete a large number of additional clones to engraft and form tumors and metastases.

Metastasis

In liver and lung metastasis of primary mice, CS showed lower values in comparison to the tumor (0.15 vs 0.2 and 0.14 vs 0.17 in primary mice of PDO1; 0.22 vs 0.32 in PDO2; 0.05 vs 0.11 in PDO3), indicating a lower clonal diversity in metastases compared to primary tumor (Figure 10). Some major/large clones of tumor i.e. clones depicted in light green color in PDO2 and in pink color in PDO3, demonstrated an abundance reaching 35% and 5% in tumor, respectively and an

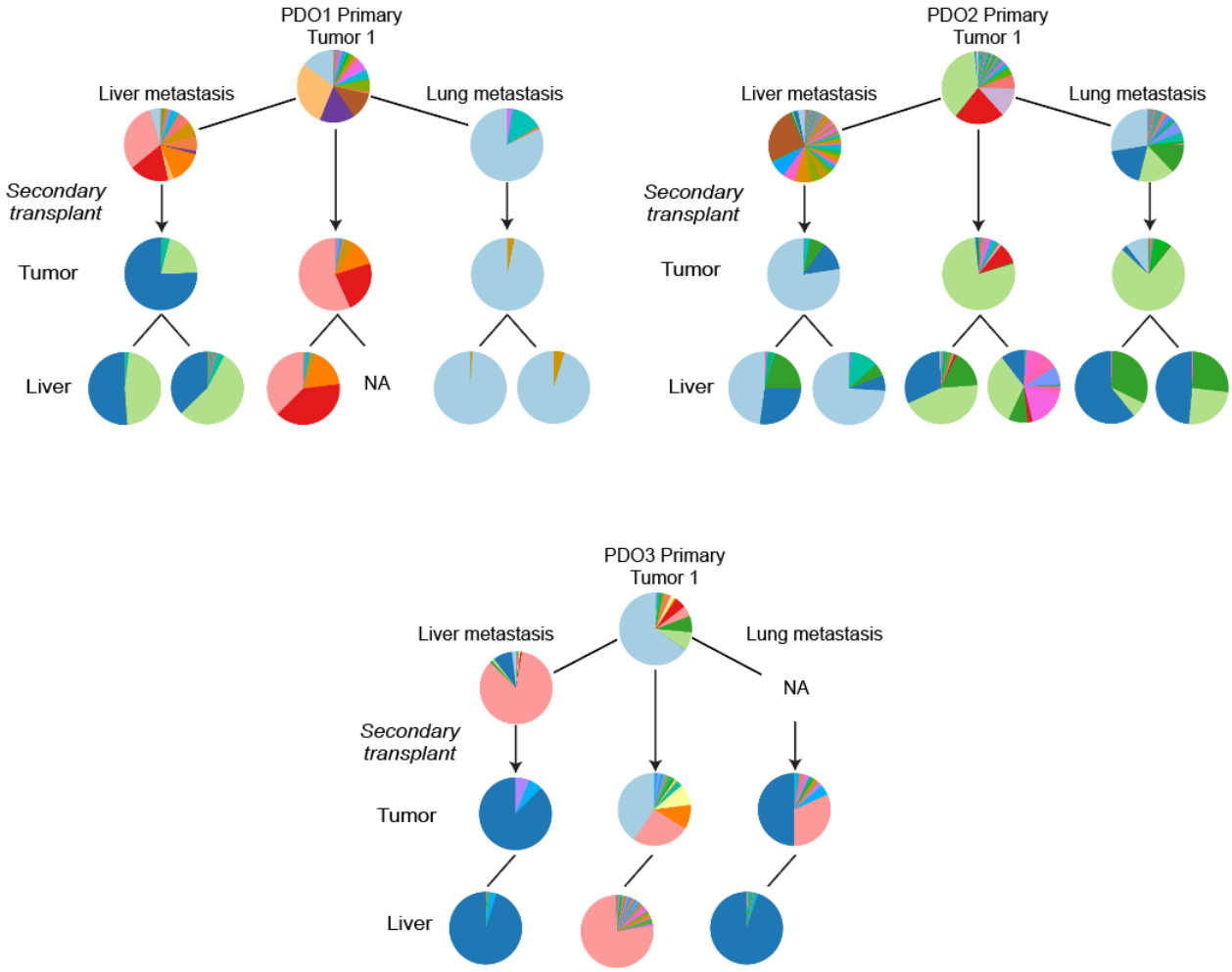


Figure 9. Pie charts depicting barcode distribution PDXs captured with scRNA sequencing. One primary tumor per PDO is represented, in addition to metastases and serial transplantations. Each color depicts a unique clone. Lines to the left direction correspond to liver metastasis while those directed to the right correspond to lung metastasis. NA: not available.

abundance of 75% and 25% in liver metastasis (Figure 9). Small clones of the tumor (pink and red color-depicted clones forming less than 1% of cells in PDO1; orange and light yellow colored clones forming less than 5% of cells in PDO3) reflected a higher frequency in liver metastasis (pink and red clones forming more than 75% of cells in PDO1; orange and light yellow clones forming up to 20% of cells in PDO3) (Figure 9). This indicates that both large and small clones in tumor can further expand in liver metastasis. Other abundant clones in tumor (violet, beige and light blue 12.5-20% in PDO1; light violet and orange clones 5-10% in PDO2; green clones 5-7% in PDO3) showed a lower representation in liver metastasis (violet, beige and light blue below 1% in PDO1; light violet and orange clones below 1% in PDO2; green clones below 2% in PDO3), indicating that large clones of tumor can diminish in liver metastasis.

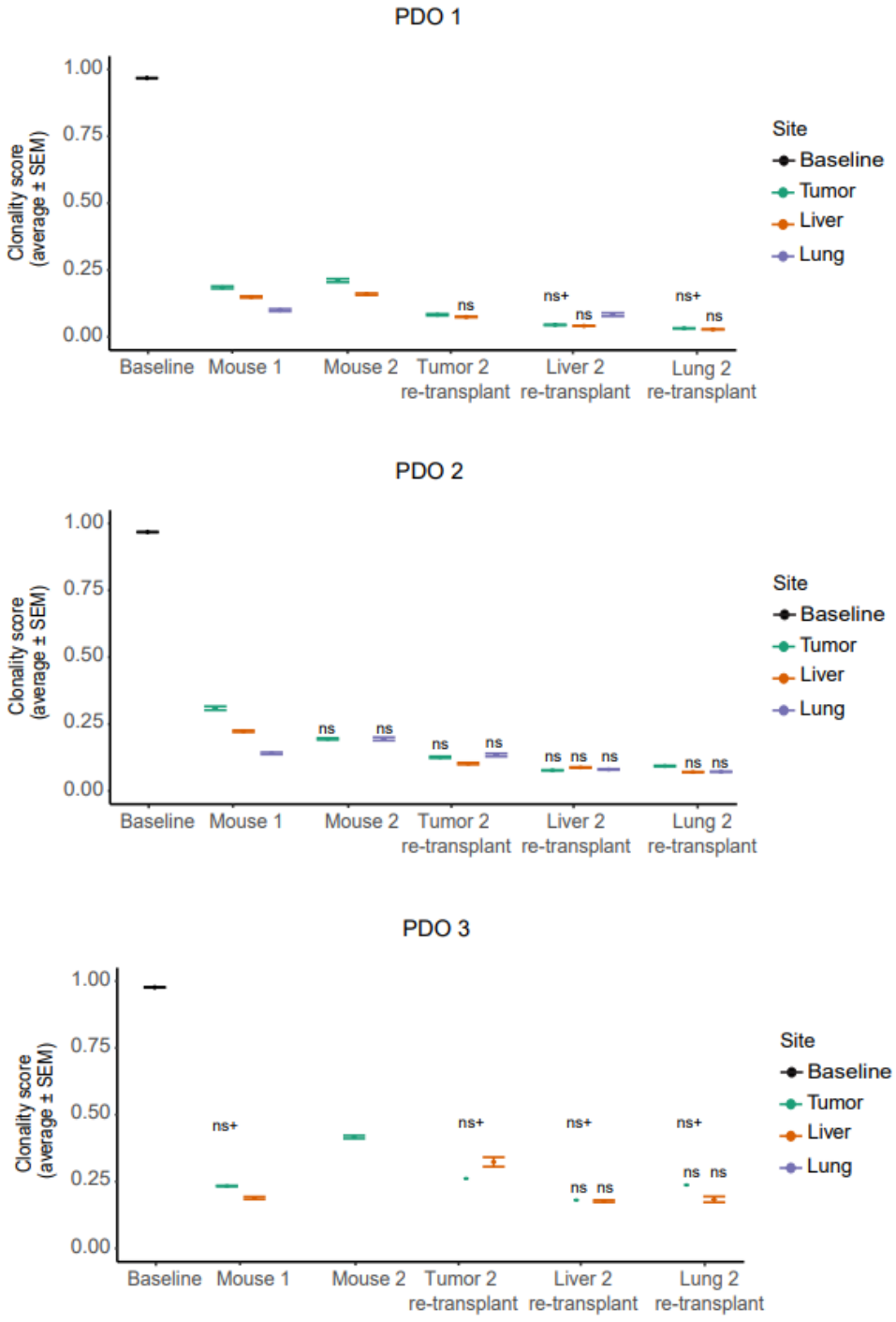


Figure 10. Clonal complexity of tumor and metastasis in barcoded xenografts. Plots illustrating the clonality score of baseline samples in each PDO and their respective xenograft. A CS of 1 means each cell has a unique barcode, while a low CS suggests a low clone number forming the cell population. Average SEM: average standard error of the mean derived from t-tests of downsampling (CS is calculated for blocks of 80 cells and combined for all cells of a sample). Non-included samples were not available or did not pass quality check. Non-significant comparisons (adjusted p-value>) of CS within the same mouse are indicated as ns, and those between different mice are indicated as ns+. CS comparisons were otherwise statistically significant. I produced this figure with bioinformatics support from Dr. Attila Jady and adapted it within the draft of manuscript in preparation.

Overall this analysis I performed demonstrates that major individual clones of the tumor appear in variable proportions in liver metastasis and suggests a dynamic contribution of individual clones, similar to that observed in DNA bulk sequencing, where clones can either persist, expand or diminish with time in liver metastasis. It further shows that not all clones in the primary tumor contribute equally to metastasis formation and that the relative distribution of clones can differ between primary tumor and metastasis.

Serial transplantation

In serial transplantation of tumor, liver and lung metastasis of primary mice, the CS showed lower values in comparison to the primary xenograft, indicating a lower clonal diversity in metastasis compared to primary tumor (Figure 10). In serial transplantations, only two to three clones collectively accounted for >75% of cells in second generation xenografts (Figure 9). This data suggests clonal exhaustion in tumor progression accompanied by the expansion of few individual clones that constitute the majority of the cancer cell population.

2.5 Tumor and metastatic cell populations demonstrate transcriptional diversity

Bioinformatic analysis was supported by Dr. Attila Jady.

2.5.1 Tumor and metastatic cells distribute in distinct transcriptional clusters

To understand the transcrptional heterogeneity of individual CRC subclones, I assessed the single cell gene expression profiles in barcoded xenografts. To address this, scRNA data of barcoded cancer cells from xenograft tumors, liver and lung metastases of each patient (n=15/patient) were integrated in one uniform manifold approximation and projection (UMAP). I observed that cells distributed in the projection among various clusters (PDO1:13, PDO2:15, PDO3: 12) (Figure 11.A) which are defined by distinct transcriptional programs. Cells of the 3 different cancerous sites (colon, liver and lung) distributed in all distinct transcriptional clusters (Figure 11.B), suggesting that they maintain an intrinsic gene expression heterogeneity despite the very distinct microenvironment the metastases develop in within the xenografted mouse.

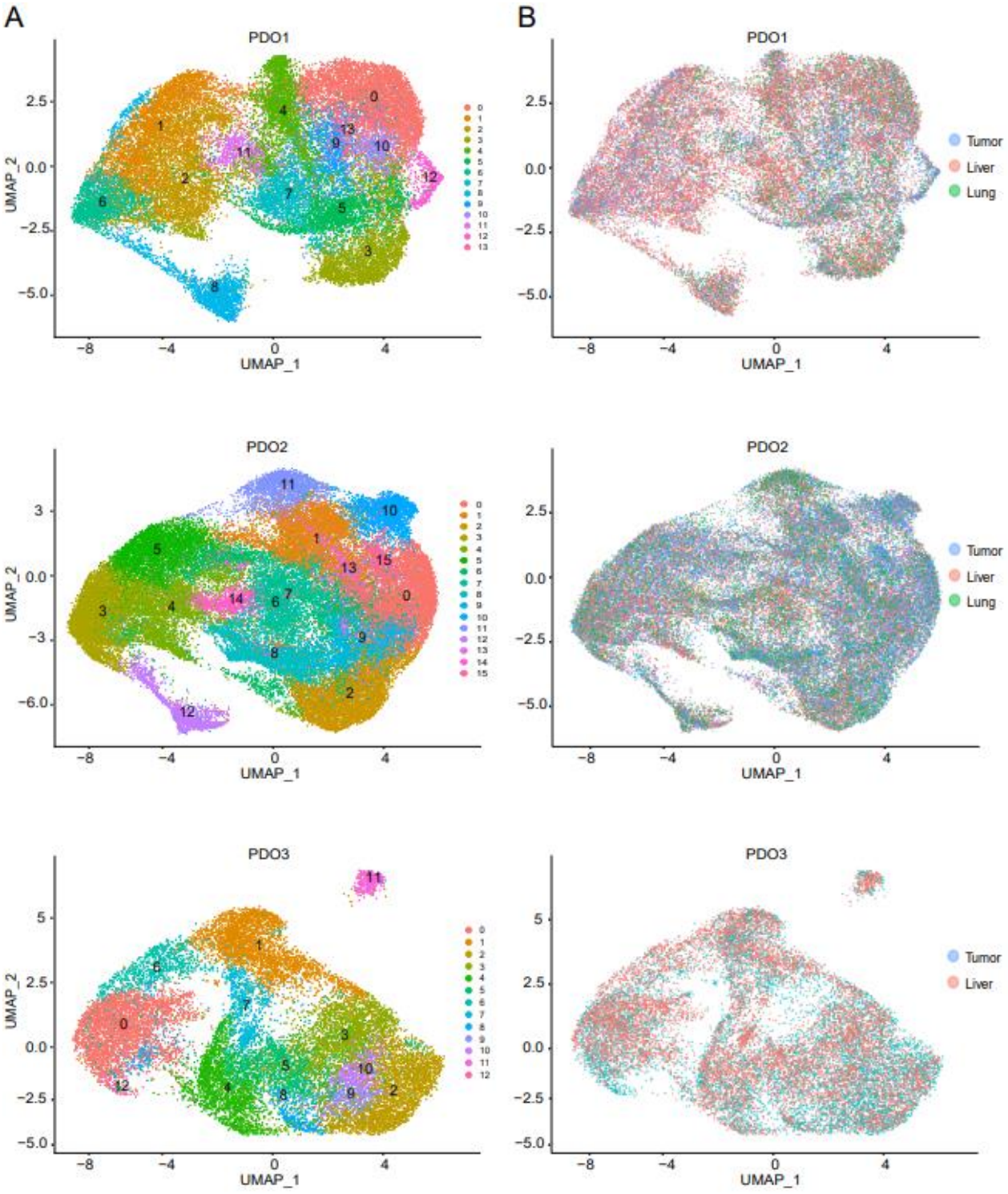


Figure 11. UMAPs of barcoded PDO xenografts. A. Distribution of transcriptional cell clusters of tumor, liver metastasis and lung metastasis in xenografts of each PDO. Each gene expression cluster is denoted with a unique color. B. Distribution of cells in tumor or metastases in the transcriptional space.

2.5.2 Tumor and metastatic cells express markers of different cell types of the intestinal epithelium

Since CRC originates as an abnormal growth on the inner lining of the colon or rectum, I assessed whether transcriptional clusters of barcoded xenografts share expression programs with known cell types of the normal intestinal epithelium. I examined this through the differential gene expression of previously described gene markers and signatures assigned to intestinal crypt cell types, in the single cell gene expression clusters of PDO xenografts (Figure 12) (Barker, van de Wetering, & Clevers, 2008; Uhlitz et al., 2021; Y. Wang et al., 2020).

I observed that differentiated enterocytes and secretory cell marker genes FABP1, TFF1/3, MUC13 and LCN2 were highest differentially expressed (average expression >+1; p-value <0.05) in closely related clusters of the transcriptional space UMAP within each patient (clusters 3, 5 and 12 in P1; clusters 2, 6, 8 and 9 in P2; clusters 4, 5 and 12 in P3) in comparison to other clusters (Figures 11, 12, 13). Cycling cell markers CCNB1, PTTG1 and RAN (Ishitsuka et al., 2012; X. Wang et al., 2020) were highly differentially expressed (average expression >+1; p-value <0.05) in distinct clusters (clusters 0, 10 and in P1; clusters 0, 1 and 7 in P2; clusters 2 and 9 in P3) whereas stem cell markers LGR5, SMOC2, NOTUM (Barker et al., 2007; Flanagan et al., 2021; J. Liu et al., 2021; Y. Liu et al., 2023; Q. Miao et al., 2019; Roukens et al., 2021; Y. Wang et al., 2020) were all positively expressed in other defined clusters (average expression >+1; p-value <0.05) (cluster 4 in P1; clusters 11 in P2; cluster 3 in P3). Markers of embryonic /fetal (VEGFA and SOX2/4) (Fazilaty & Basler, 2023; Q. Miao et al., 2019; Roukens et al., 2021; Weis & Cheresh, 2005) and mesenchymal invasive (NEAT1, MALAT1 and ARGLU1) (Y. Chen et al., 2021; Koedoot et al., 2019; Zhou et al., 2015) phenotype were highest expressed (average expression >+1; p-value <0.05) in a group of adjacent clusters, opposite to differentiated cell clusters in the transcriptional space of each patient (clusters 1, 2 and 6 in P1; clusters 3, 4 and 5 in P2; clusters 1, 6 and 7 in P3) (Figures 11, 12, 13). These latter clusters negatively expressed markers of differentiated intestinal crypt cells (FABP1, TFF3, KRT20) and positively expressed SOX4, a driver of embryonic gene program (J. Liu et al., 2021) (Figure 11, 12) and were assigned as invasive-fetal transcriptional clusters. Further transcriptional cell clusters were identified i.e. inflammatory-chemokine (CXCL1, CCL20) (Sawant et al., 2016), progenitor-anti-oxidant (GPX1, SLPI, ATOX1) (Y. Zhao, Wang, Zhou, & Shao, 2022) among the proximity of differentiated epithelial cell clusters, in addition to metaplastic and regenerative programs (SOX9, DDIT4) (Z. F. Miao et al., 2020) (Figure 13, Supplementary figure 5). This analysis I performed shows that the expression of cell markers related to distinct cell types of the intestinal epithelium is distributed in different transcriptional clusters of xenografted PDOs.

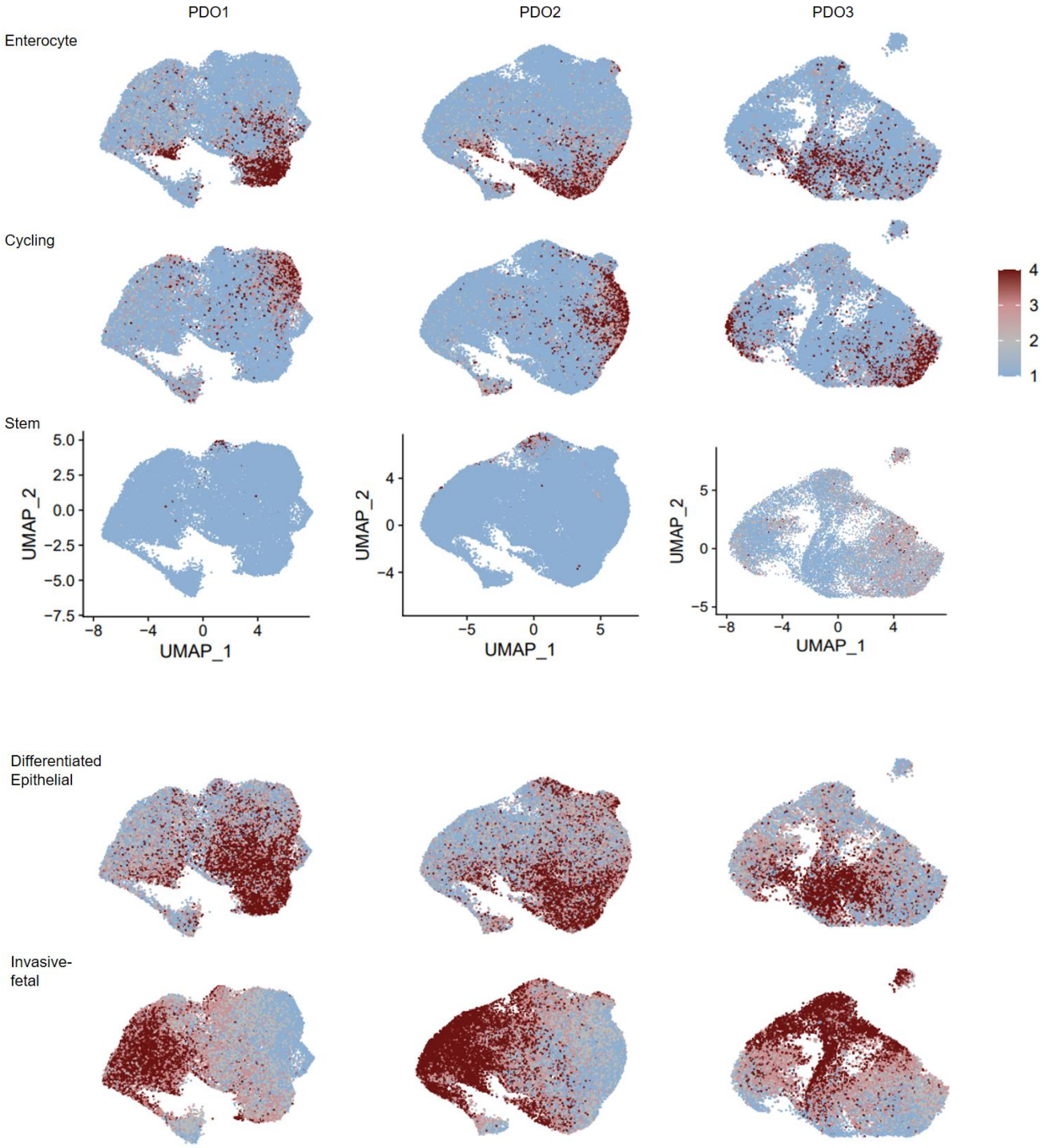


Figure 12. Expression of cell type marker genes in transcriptional clusters of barcoded PDO xenografts. Enterocyte, cycling and stem cell markers are depicted in the first panel. Gene expression level in cells is represented by color intensity depicted in the scale. UMAP dimensions are indicated in the third row. The second panel shows differentiated epithelial and fetal invasive transcriptional states.

2.5.3 Distinct biological processes and canonical pathways characterize the transcriptional clusters of barcoded xenografts

To detect key biological processes and canonical pathways in identified transcriptional clusters of tumor and metastasis, I analyzed gene signature enrichment analysis GSEA (Subramanian et al., 2005) in these clusters.

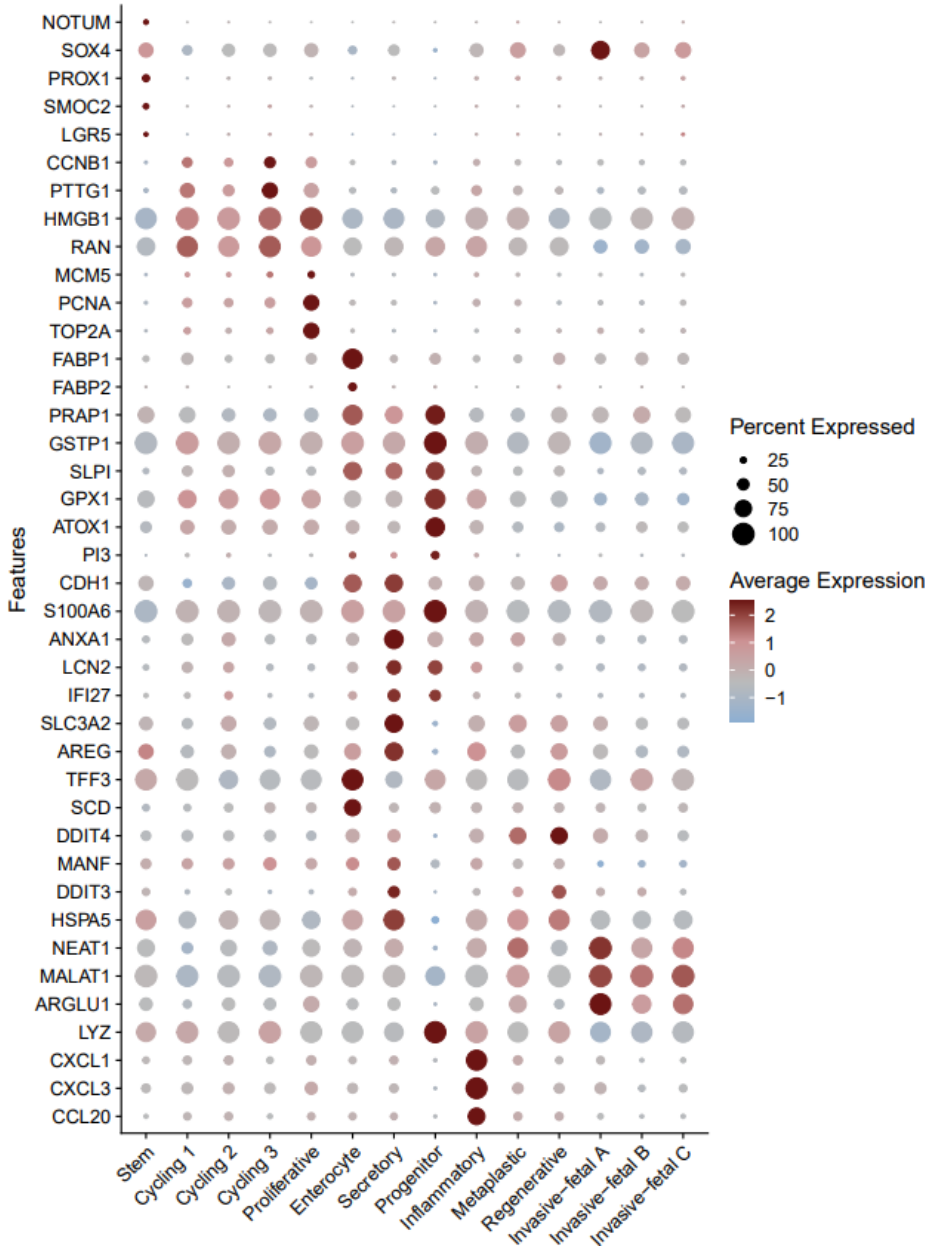


Figure 13. Cell type marker genes of transcriptional clusters of xenografts. Plot depicting the gene expression level of selected genes in assigned transcriptional clusters of xenografts derived from PDO1. All presented points are statistically significant.

GSEA demonstrated that cancer cells of the differentiated epithelial cell phenotype (enterocyte-secretory) enriched biological processes of absorption (endocytosis), secretion, adhesion and calcium/iron binding (NES of 2-3; p-value adj. <0.05), while downregulating cell division (G2M) (NES < -2; p-value adj. <0.05) (Figure 14, Supplementary figures 6, 7). Canonical pathways in these cells included elevated MTORC and KRAS signaling and down regulated E2F, DNMT1 and MYC targets (NES <-1; p-value adj. <0.05). Enriched hallmark processes of the identified undifferentiated cycling cell clusters included cell cycle, oxidative phosphorylation (cellular respiration), and apoptosis repression, while major active pathways included enriched MYC, E2F and DNMT1 targets, along with downregulated p53, KRAS and TNFA signaling (Figure 14, Supplementary figures 6, 7). In the stem cell-like cluster, hallmarks of G2M, elongation factor E2F and oxidative phosphorylation OXPHOS were downregulated (NES <-1; p-value adj. <0.05), while ribosomal activity was enriched. Of the canonical pathways in this cell cluster, TNFa-NFKb signaling was activated while MTORC and MYC pathways were downregulated. Undifferentiated invasive cells were characterized by GSEA hallmarks of negative regulation of apoptosis, cell differentiation, adhesion and OXPHOS, in addition to downregulated mTOR and MYC pathways. Cells within this transcriptional program upregulated biological processes of vasculature development and tissue migration. Histone 3 lysine 4 H3K4 methylation, related to chromatin regulation in embryonic cells (Bernstein et al., 2006), and gene sets of embryonic organ development were upregulated in these clusters (NES +1.78; p value.adj <0.05; NES +1.8; p value.adj <0.05, respectively). I observed that Progeny analysis (Pathway RespOnsive GENes for activity inference), which can infer pathway activities in scRNA data, identified PI3K and VEGF as highest expressed pathways (score of >+1; p-value<0.05) in these cell clusters (Supplementary figure 8) (Schubert et al., 2018). Promotion of PI3K pathway, reported in invasive cancer cells, has been previously reported to be promoted through the activation of SOX4 (Mehta et al., 2017), which is highly expressed in identified invasive clusters (average expression >2 in 75% of cells) (Figure 13).

2.5.4 Tumor cells express genes of epithelial to mesenchymal transcriptional trajectory

Since clusters located at the opposite extremities of the UMAPs enrich the expression of either differentiated epithelial (KRT20, CDH1) or invasive-fetal and mesenchymal (VEGFA, SOX4) genes and signatures (Cadherin binding in epithelial clusters and PI3K pathway activation, in invasive-fetal mesenchymal-like clusters) (Figures 12, 13), I assigned these clusters as epithelial differentiated and fetal invasive. To re-draw how cells transverse through different transcriptional states of differentiation, a trajectory analysis that enables the visualization of dynamic gene expression changes in pseudo-time was carried out. Trajectory analysis using RNA velocity

Figure 14 **Figure 14. GSEA markers of transcriptional clusters.** First panel: plot showing the normalized enrichment score of GSEA terms biological processes and hallmarks in cell type-assigned transcriptional clusters of PDO1 xenograft. BP: biological process; CC: cellular component; H: hallmark.

Second panel represents molecular pathways. CGN: cancer gene neighborhoods; MF: molecular function; K: canonical pathway. Statistical significance is indicated by the adjusted p-value (adj. p).

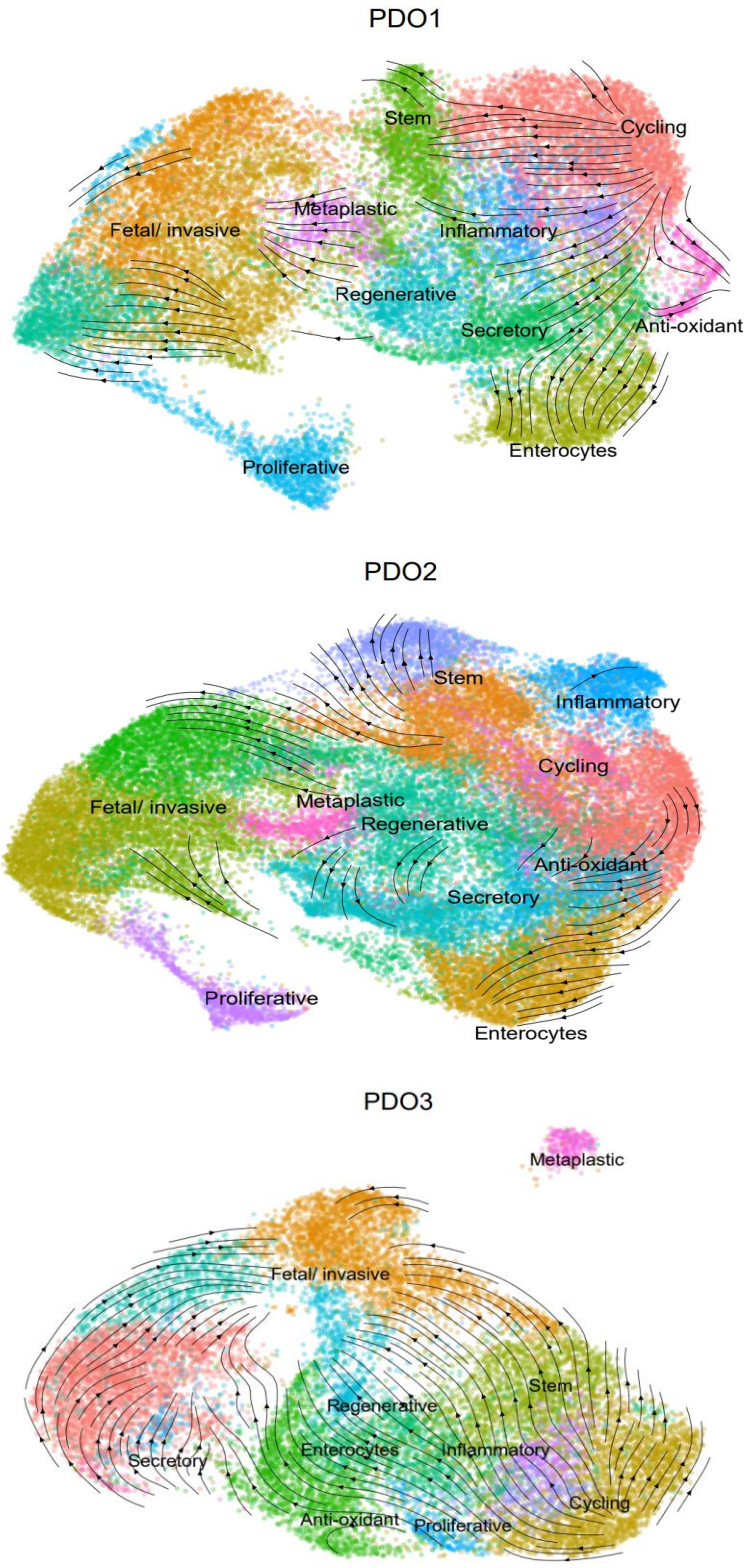


Figure 15. Differentiation trajectory in tumor and metastasis of CRC xenograft. Velocity plot derived

from trajectory analysis illustrating the initial and final differentiation states of cells. Trajectory direction is depicted with arrows.

(Weiler, Van den Berge, Street, & Tiberi, 2023) identified an initial transcriptional cluster in the vicinity of differentiated epithelial-like cells (enterocytes) with the direction of transcriptional change pointed toward a final state among mesenchymal-like invasive/fetal cell clusters (Figure 15).

An increasing gradient expression of genes previously described to be associated with embryonic stages (SOX2/4, VEGFA, CTNND1) and metastasis formation (MALAT1, NEAT1, PNN) (Y. Chen et al., 2021; Koedoot et al., 2019; Wei et al., 2016; Zhou et al., 2015) was observed across the trajectory of differentiated epithelial-like to invasive mesenchymal-like cell clusters (Figure 12, 13, Supplementary figure 5). Canonical epithelial markers (CDH1, KRT20, EPCAM) and S100 gene family, on the other hand, were gradually downregulated through the epithelial to invasive mesenchymal trajectory (Figures 12, 13, Supplementary figure 5). Intermediate clusters lying between the differentiated epithelial and invasive fetal clusters showed highest expression of regenerative genes DDIT4 and SOX9 among other clusters (log2FC 1 and 0.3 respectively). The activity of these genes was described in paligenosis, a regenerative program which fuels differentiated cells re-entry into cell cycle (Figure 13, Supplementary figure 5) (Z. F. Miao et al., 2020; Roukens et al., 2021; Willet et al., 2018). In particular, SOX9 has been reported for its neoplastic functions (Matheu et al., 2012). YAP1, involved in cellular plasticity and regeneration (Yui et al., 2018), is highest expressed in these clusters (log2FC 0.15, average expression of 2).

Overall, this analysis indicates that xenografted patient-derived CRC cells are heterogeneous in their gene expression, and individual cells express two main distinct transcriptional programs namely epithelial differentiated and invasive-fetal among others (cycling, stem) in both tumor and metastatic cells. These transcriptional phenotypes span an epithelial-differentiated to an invasive fetal/embryonic state, with an intermediate state resembling metaplasia (Giroux & Rustgi, 2017).

To illustrate the transcriptional heterogeneity of differentiated/epithelial and fetal/invasive clusters (Figure 12) on the spatial level, the expression of one marker of each of these two gene expression phenotypes (S100A4 and VEGFA, in epithelial differentiated and invasive-fetal, respectively) was evaluated in paraffin embedded xenograft material by means of fluorescent microscopy. The expression of these markers reveals a heterogeneous spatial distribution within tumor and metastasis, which varied among neighboring cells (Figure 16). In addition, the expression of these markers (S100A4 and VEGFA) did not seem to be mutually exclusive in

xenografts' cancer cells. These observations, in addition to scRNA data indicate that cells of both differentiated epithelial and invasive-fetal phenotypes co-exist in tumor and metastasis.

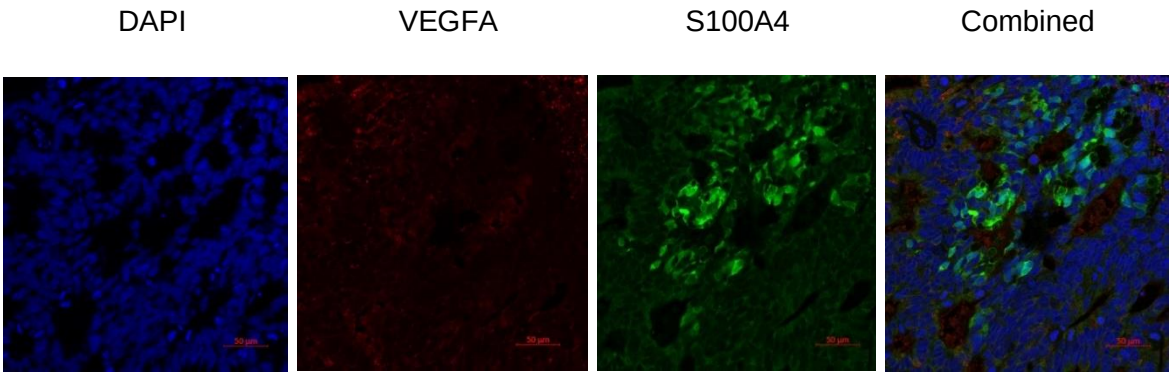


Figure 16. Immunofluorescence IF of epithelial and invasive mesenchymal-like cluster markers in liver metastasis of PDO1 xenograft. Scale bar: 50μM. IF performed by Dr. Stella Arelaki.

2.5.5 Liver metastasis cells enrich invasive cell clusters while lung metastasis enrich a differentiated epithelial cell phenotype.

To understand the contribution of particular transcriptional cell clusters in tumor progression, I quantified their abundance in tumors compared to metastases. Cells of liver metastases exhibited a higher frequency of invasive-fetal clusters compared to tumor and lung metastases cells demonstrated by an average increase of 5.3-15.6% of invasive-fetal B in PDO1 (adj. p-value <0.05) and 3.4-9.8% of invasive fetal A in PDO2 (adj. p-value <0.05) among all clusters. Tumor and lung metastases were more abundant in epithelial differentiated (enterocyte-secretory) transcriptional clusters (3.7-9.9% higher in PDO1; adj. p-value<0.05 and up to 8.3% higher in secretory 2 and 3 of PDO2; adj. p-value <0.05) in comparison to liver metastases (Figure 17). Cycling cell cluster abundance was high in both tumor and metastases of all 3 PDO PDXs (15-28% of all clusters).

2.6 Subclonal gene expression heterogeneity characterizes tumor and metastasis of CRC xenografts

To understand the intra-clonal gene expression of CRC in barcoded xenografts, I analyzed the distribution of subclonal cells within the identified transcriptional clusters of each patient. This supports determining whether individual clones lie within defined transcriptional cell clusters. Single cell gene expression data shows that cells of large individual clones (cells sharing the same barcode) subsisted among all identified transcriptional programs, corresponding to differentiated epithelial, stem, invasive-fetal, amplifying cell states and others (Figure 18). Intra-

clonal cells were not restricted to particular gene expression clusters. This suggests that distinct clones are not defined by specific transcriptional profiles but are rather distributed in heterogeneous gene expression programs. Clones maintained this heterogeneous distribution between identified states of epithelial to invasive transcriptional clusters in tumor, liver and lung metastasis (Figure 18).

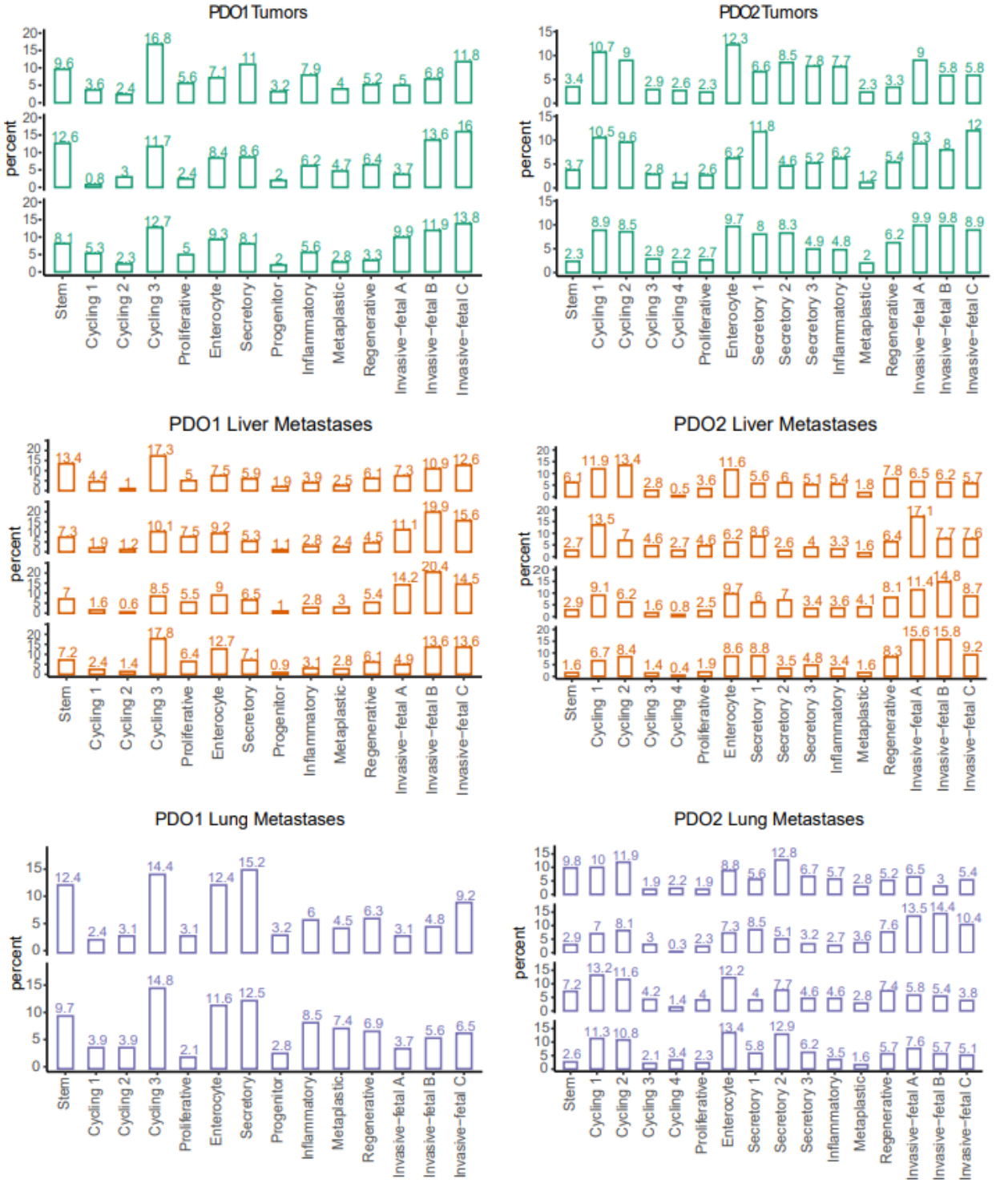


Figure 17. Cluster abundance in tumors, liver metastases and lung metastases. Barplots illustrating the frequency of gene expression cell clusters in %. Green color corresponds to tumor cells, red to liver metastases and blue to lung metastases.

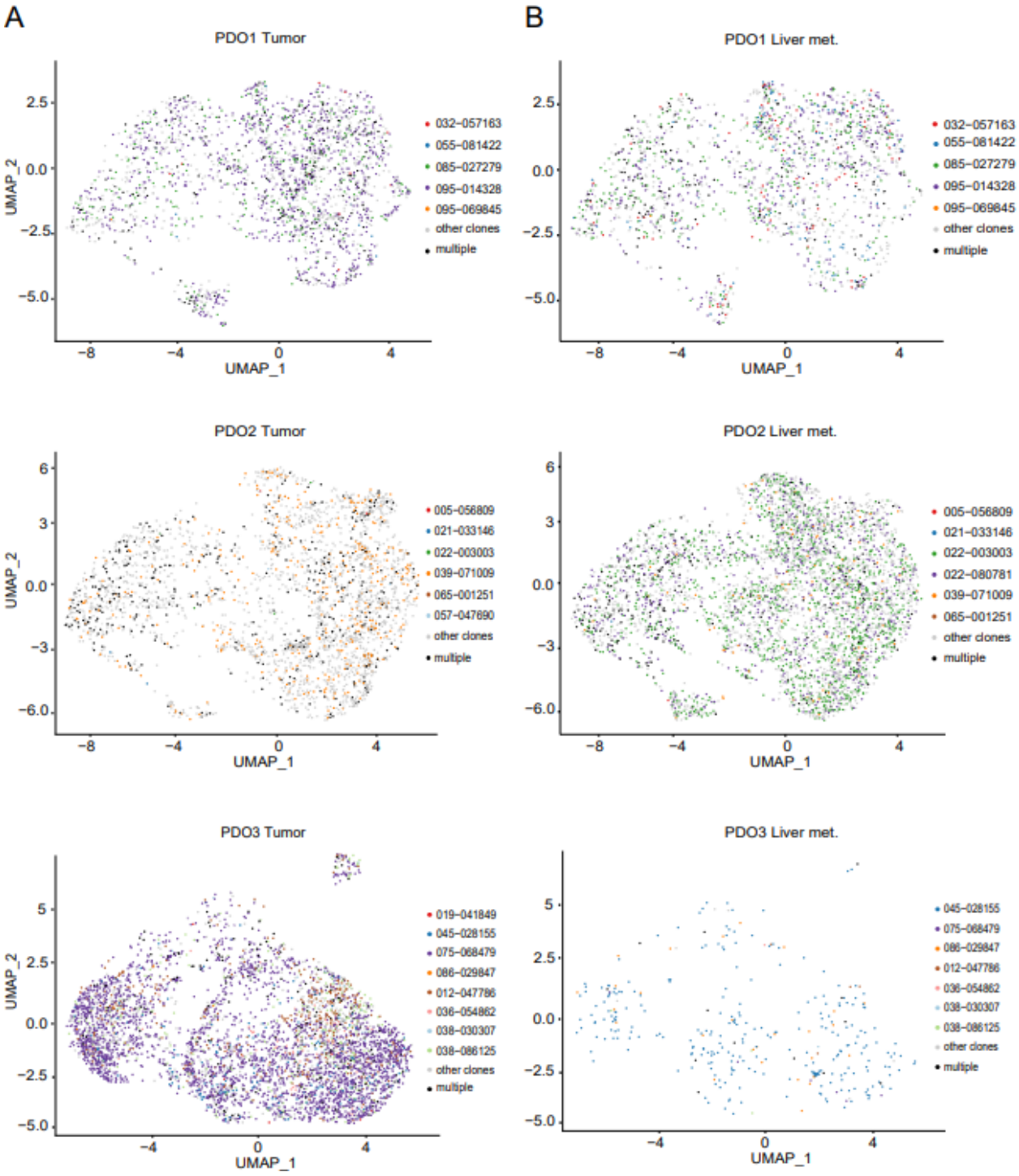


Figure 18. Distribution of individual clones in transcriptional programs. UMAPs of 3 patients in tumor and metastasis, each patient in a row. Cells of each clone are colored uniquely. Cells with multiple transductions are colored in black and clones forming less than 1% of the cell population are colored in gray (other clones). I produced this figure with bioinformatic support from Dr. Attila Jady.

2.7 The frequency of transcriptional clusters in individual clones

To dissect gene expression changes among cells of individual clones in liver metastasis compared to tumor, I quantified the frequency of transcriptional clusters in sub-clonal cells (cells corresponding to a given clone) of tumor and liver metastasis. The frequency of stem cell and enterocyte transcriptional programs in single clones remained unchanged (maximum change of 1.6% and 2.5%, respectively) in both tumor and liver metastasis, and the cycling cell phenotype remained highest among other programs (up to 30%) (Figure 19). In liver metastasis, individual clones enriched the fetal/invasive (mesenchymal-like) cell clusters (25% in liver vs 18% in tumor for PDO1) (Figure 19) and reflected a reduced frequency of the secretory phenotype (2.9 % in liver vs 10% in tumor of PDO1; 2.9% vs 7.7 % in PDO2) (Figure 19) in a pattern comparable to the overall gene expression in liver metastases (Figure 17).

To confirm this observation on the gene level, the expression of invasive fetal identified genes (PCLG2, MALAT1, VEGFA and SOX4) and marker genes of differentiated epithelial features (KRT20, ANXA1, S100A11) was quantified in subclones. The expression of invasive markers was higher in the liver metastasis subclonal compartment compared to the tumor compartment (average expression > 0.75 in liver met. vs < 0.5 in tumor), while the contrary is observed for differentiated cell markers (>1 in tumor and <-0.5 in liver metastasis) (Figure 20).

Subclones in lung metastasis demonstrate an expanded differentiated secretory and enterocyte phenotype transcription in comparison to liver metastasis (PDO1: enterocyte 13% in lung vs 6.1% in liver; secretory 13% vs 2.9% in lung met and liver met., respectively; PDO2: 24.2% vs 15.7% secretory in lung met. and liver met., respectively while enterocyte-like cell clusters were comparable), and a reduced invasive cell program (PDO1: 25.5% vs 14.5% in liver met. vs lung met.; PDO2: comparable in this clone) (Figure 19), in a pattern similar to overall gene expression program in lung metastasis (Figure 17). The secretory phenotype cluster positively expressed gene sets of adhesive, endocytic, and secretory biological processes (Figure 14). In addition, among top genes in epithelial differentiated transcriptional programs (enterocyte and secretory) are Annexin ANXA1, which is required for survival of metastatic cells in combination with calcium-binding protein encoding gene S100A, expressed in the anti-oxidant progenitor cluster (Figure 13) (Jaiswal & Nylandsted, 2015). This indicates that individual clones maintain gene expression heterogeneity in both tumor and metastases but enrich a fetal invasive phenotype in liver metastasis and a differentiated epithelial phenotype in primary tumor and lung metastasis.

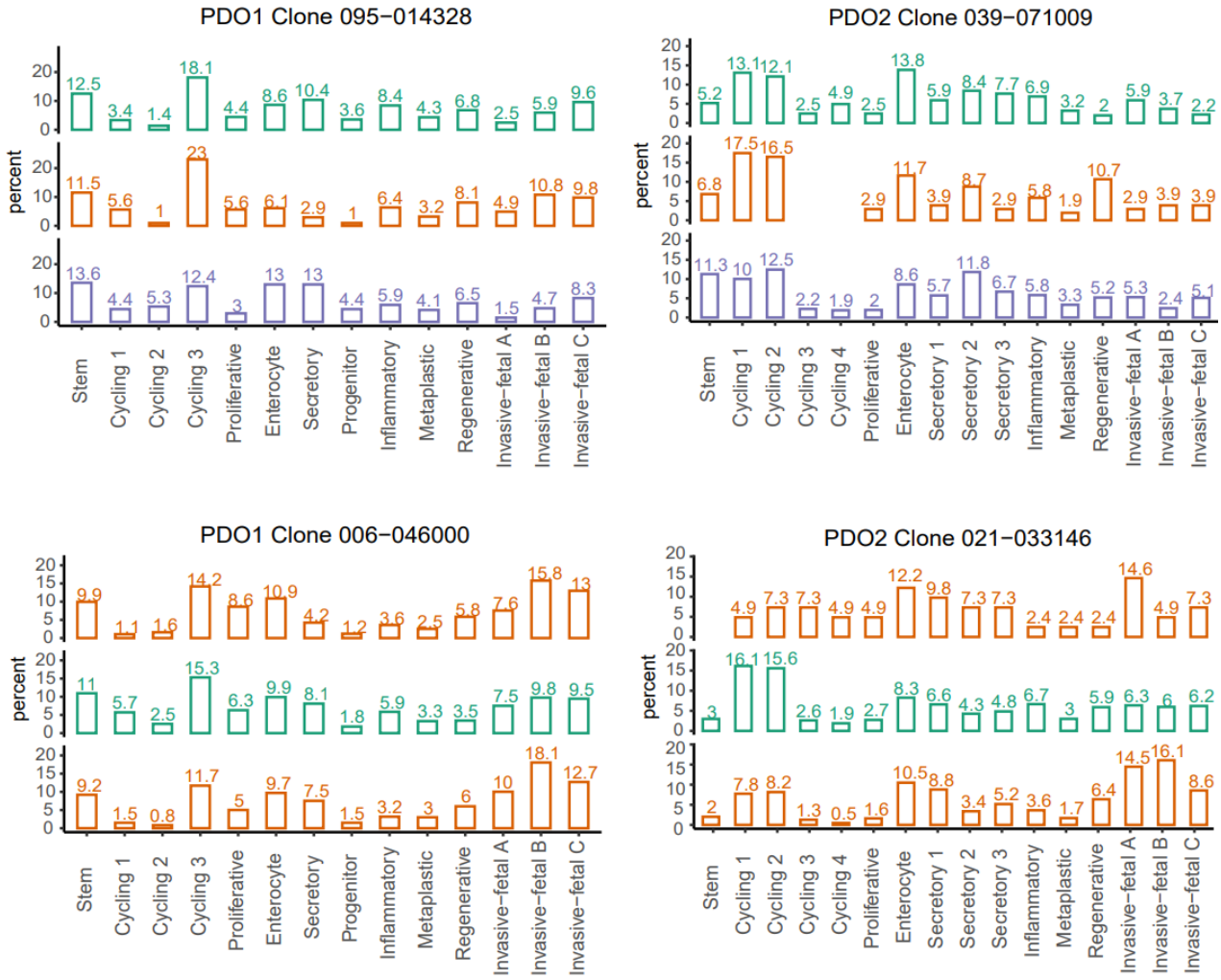


Figure 19. Distribution of transcriptional clusters in individual clones. Barplot illustrating the frequency of transcriptional cell clusters in individual clones in %. Green color corresponds to tumor cells, red to liver metastasis and blue to lung metastasis. The top panel represents data of primary mice, while the panel below shows data derived from the serial transplantation of liver metastasis into the caecum of secondary mice.

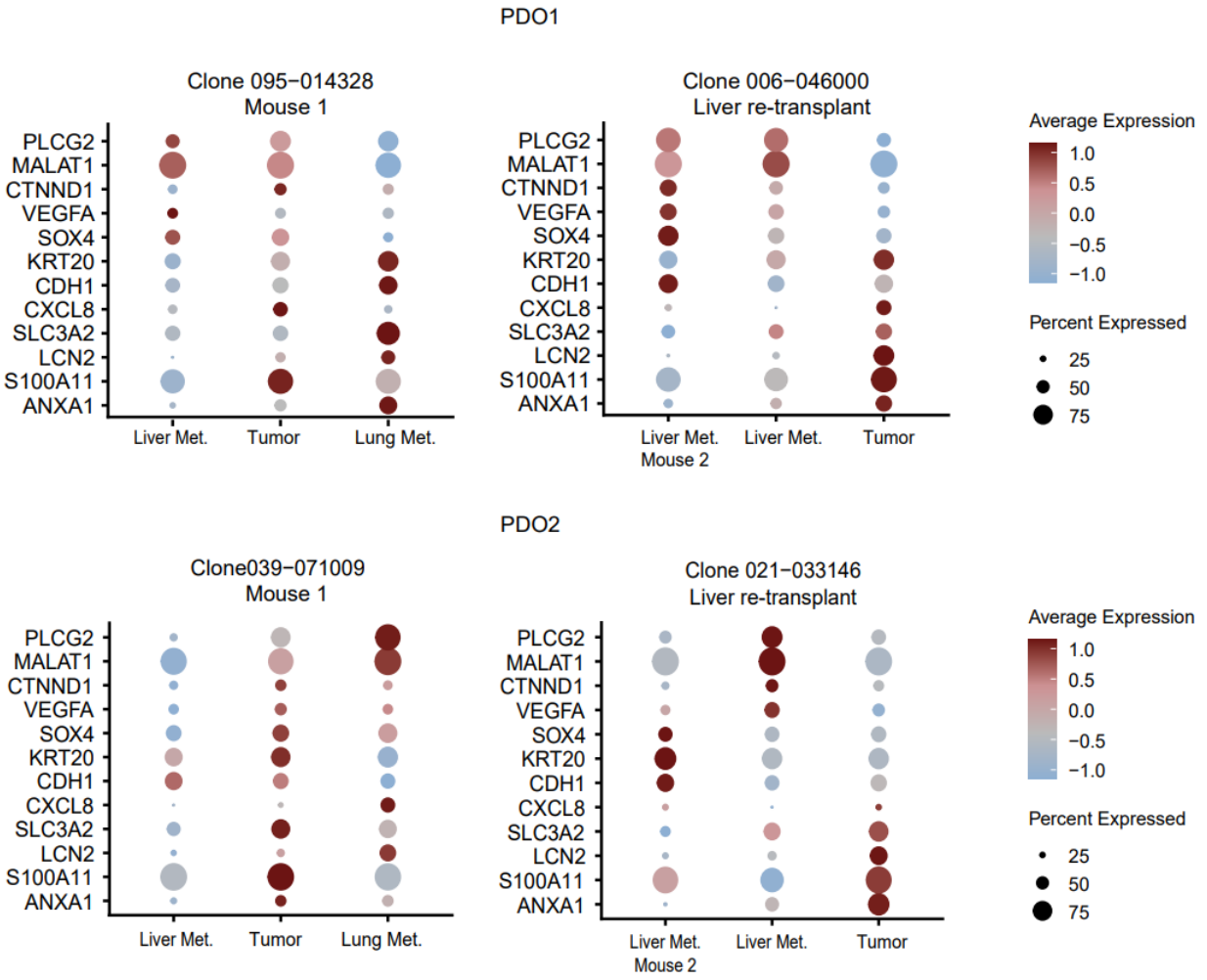


Figure 20. Differentiated and invasive cell marker expression in individual clones. Plot illustrating the selected genes expression in individual clones in tumor and metastases of liver and lung. In the left panel, data derived from primary mouse 1 is shown, while the left panel reflects data from serial transplantation of liver metastasis of primary mouse 2. A representative clone form each mouse is shown. Statistically significant results are presented (adjusted p-value<0.05).

In serial transplantation of liver metastasis clones into the caecum of secondary mice, the tumor transcription profile (lower invasive and higher epithelial cluster frequency) (Figure 17) was restored (Figure 19). In liver metastasis of secondary mice, subclones showed the transcriptional program related to liver metastasis of primary mice (higher invasive and lower epithelial cluster frequency in liver metastasis, vice versa in lung metastasis) (Figure 19), indicating a plastic potential of CRC clones to adapt their gene expression in different microenvironments.

2.8 Gene expression of big clones is similar to that of minor clones

Since clones exhibited varying contribution with big and small sizes (Figures 8, 9), I hypothesized that major (dominant) clones of bigger size (clones forming more than 10% of all cell population) could have high expression of genes associated with cell proliferation and survival, in comparison to smaller clones forming less than 5% of the cell population in a given sample. To assess this, differential gene expression analysis (DEG) was carried out between these big and small clones. Surprisingly, only small (log2FC<0.4, adj p-value <0.05) changes in the expression level of a few genes (CALB1, IFI27, S100A4, LYZ, DNAJC15 and MTRNR2L12) was observed in 11% of all clonal comparisons (Figure 21.A). This indicates a lack of major gene expression variations between big and small clones (inter-clonal level) and suggests that differential gene expression may not explain higher growth of specific clones.

To understand whether cells of particular subclones reflect gene expression variations in different microenvironments (primary tumor and metastasis), I compared the differentially expressed genes (DEG) of particular abundant (forming more than 10% of the cell population) clones (n=21 in PDO1; n=18 in PDO2) in different locations i.e. colon, liver and lung (intra-clonal comparison). Specific genes within the same clone showed enriched expression in lung or liver metastases compared to the primary tumor (51 genes in PDO1 and 72 genes in PDO2; log2FC >1; adj p-value <0.05) in more than 90% of the observed cases (Figure 21.B). Among the top genes (29 in PDO1 and 35 in PDO2; log2FC> 1; adj p-value <0.05) observed in liver metastasis versus tumor were MIF and PLCG2, which are upregulated in the fetal-invasive (mesenchymal) transcriptional program, associated with liver metastasis. LCN2 and ANXA1, which are top genes of the enterocyte and secretory epithelial programs, were positively differentially expressed in single clones residing in lung metastasis compared to tumor and liver metastasis in the same mouse (Figure 21). The differential expression of some of these (12 of 22 genes in PDO1 and 37 genes in PDO2) was observed in a previous analysis (Figure 20). This indicates that the metastatic site influences the gene expression of individual clones in CRC xenografts and that transcriptional heterogeneity of individual clones can support tumor growth and progression.

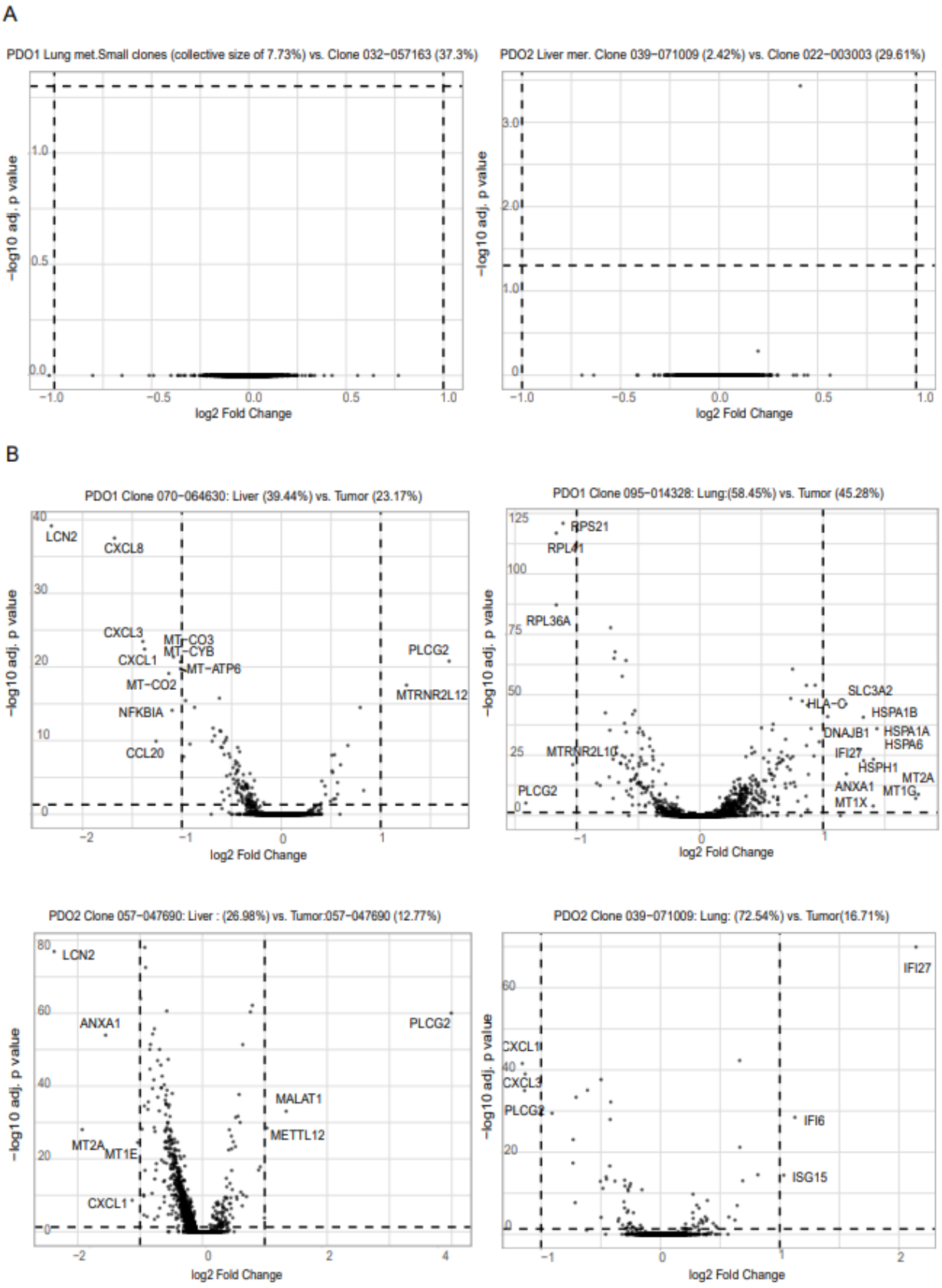


Figure 21. Inter-clonal and subclonal differential gene expression. A. Volcano plot depicting differential

gene expression analysis in representative small and big clones of liver and lung metastasis of PDO1 and PDO2, respectively. B. Volcano plot depicting differential gene expression analysis of subclonal cells in liver or lung metastasis versus subclonal cells of the primary tumor; PDO1 (upper panel) and PDO2 (lower panel) are shown. Statistical significance is indicated in the y-axis, all represented gene names are statistically significant (adjusted p-value <0.05)

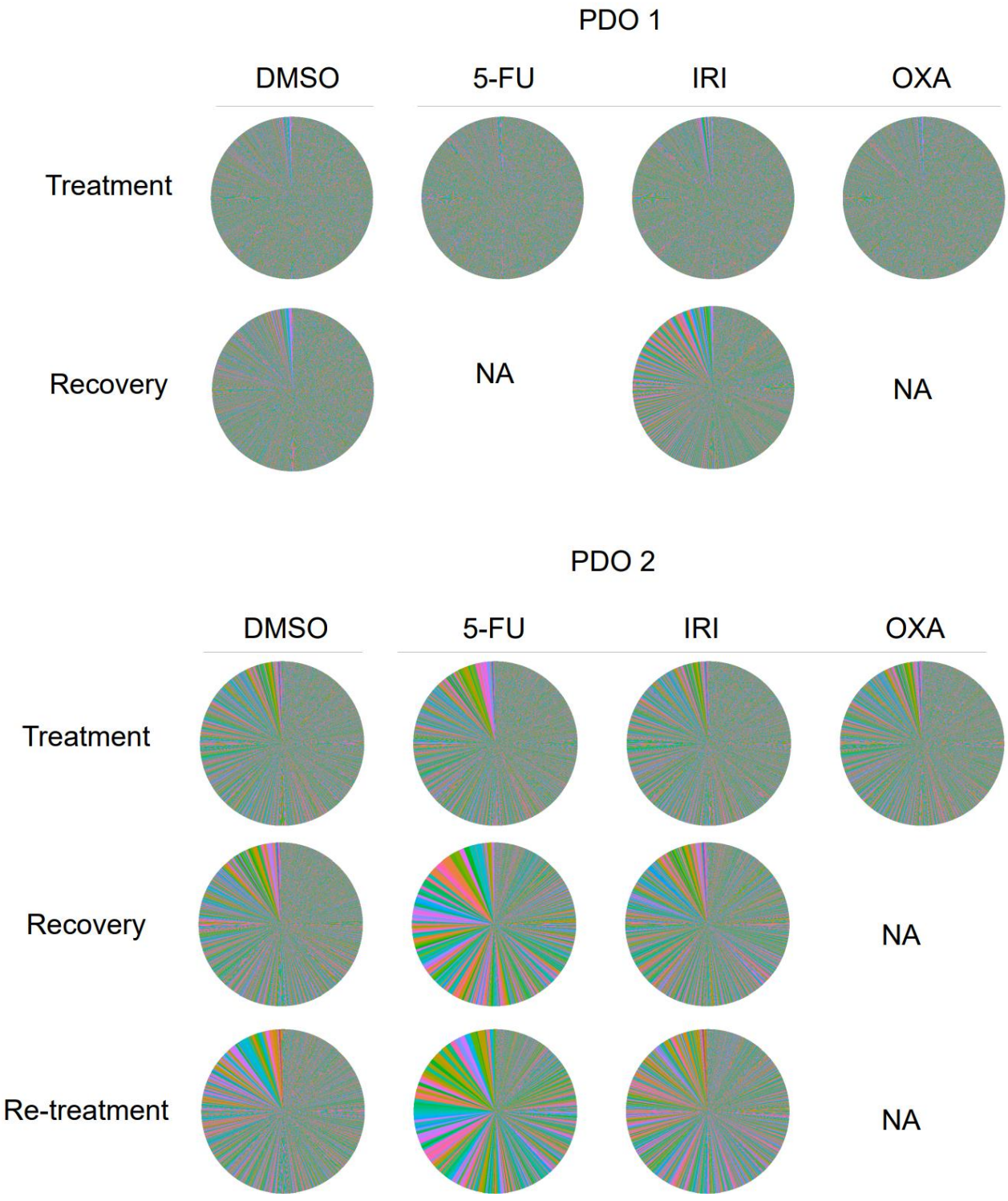
2.9 Clonal gene expression response to chemotherapy

2.9.1 Chemotherapy exerts a minor effect on barcode distribution of CRC PDOs

Since the transitions in pre-existing transcriptional states such as differentiated epithelial and fetal-invasive (mesenchymal) phenotypes were shared by and not restricted to individual clones, and that big clones did not reflect high differential gene expression in comparison to small clones, I aimed at understanding how challenging CRC clones with chemotherapeutic stress would affect the clonal distribution and gene expression within individual clones. To test this, I treated one million barcoded PDOs of t0 (Figure 6.A) with an estimated 70% of maximal inhibitory concentration (IC70) of chemotherapeutic agents conventionally used in CRC treatment (André et al., 2004) i.e. 5-fluorouracil, Oxaliplatin and Irinotecan separately, for five consecutive days. I used the drug solvent compound dimethyl sulfoxide DMSO as control. Cells from DMSO and chemotherapy PDOs were collected for DNA barcode sequencing and scRNA sequencing whenever available, and splitted for further passaging. Cells were further cultured for an additional week in drug-free media for recovery and were collected for DNA barcode sequencing and scRNA sequencing.

DNA sequencing

To measure clonal dynamics in response to treatment, I quantified the proportion of unique barcodes (clones) within treatment and control settings by depicting the percentage of reads of the same barcode in the total read count. Targeted barcode DNA sequencing of both DMSO and chemotherapy treated barcoded PDOs revealed that single barcodes represented 7.7x10E-5 to 0.1% of the total cell population in PDO1 and PDO2, and 2.6x10E-4 to 6.07% in PDO3 (Figure 22, Supplementary table 4). This barcode representation, comparable to the baseline sample (3.3x10E-5 to 0.01% in PDO1; 7.1x10E-5 to 0.01% in PDO2; 1.2x10E-4 to 5.9% in PDO3) (Figure 6B), and without major outgrowth of individual pre-existing clones in chemotherapy treated and DMSO control cells, indicates that chemotherapy treatment *in vitro* has a minor impact on clonal complexity on the short term of one week. After one week of recovery in drug-free culture media,


PDO 1
DMSO
5-FU
IRI
OXA
Treatment
Recovery
NA
NA
PDO 2
DMSO
5-FU
IRI
OXA
Treatment
Recovery
NA
Re-treatment
NA

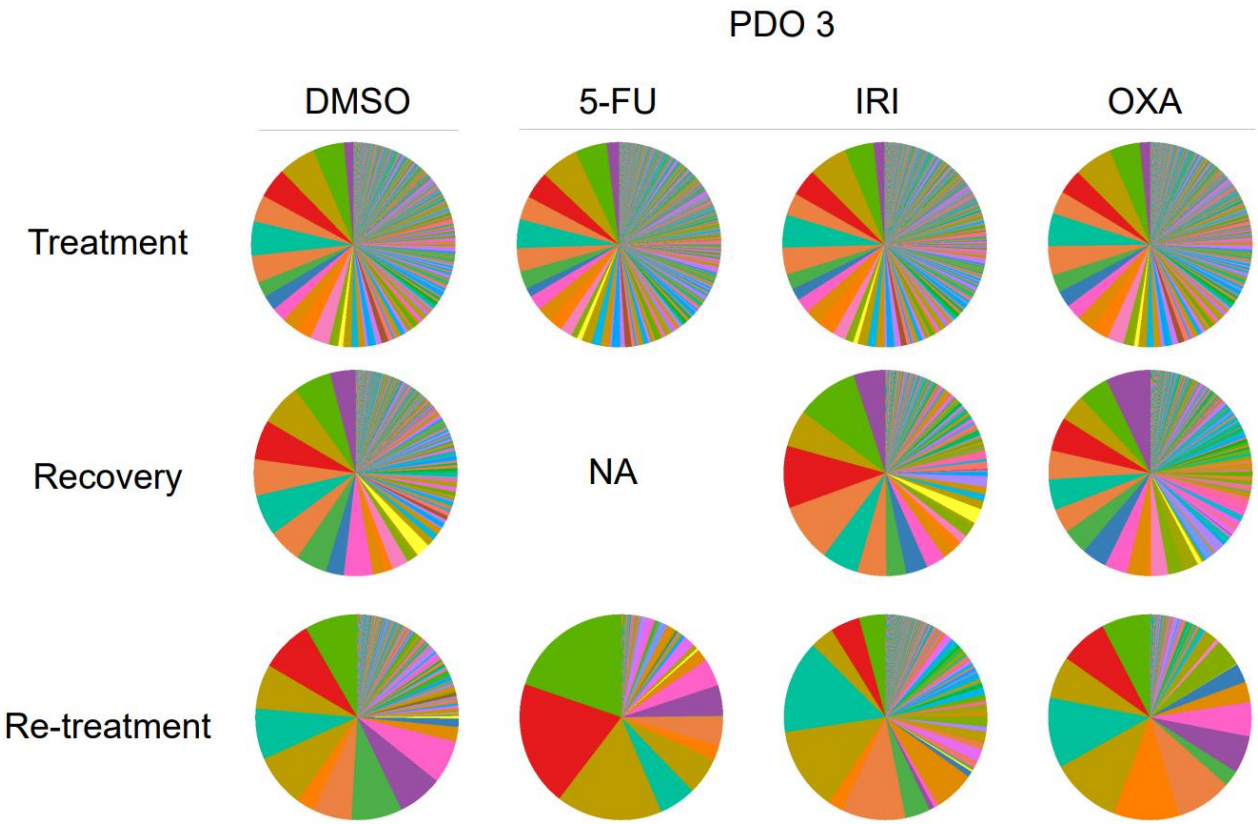


Figure 22. Clonal distribution in treated PDOs. Pie charts depicting the barcode distribution (in %) in PDOs, captured through targeted DNA (amplicon) sequencing. DMSO and chemotherapy treated PDOs are illustrated, in addition to recovered and re-treated timepoints. Each color portrays a unique clone. Gray color represents clones of low representation (less than 0.1% of total barcode count). NA: not available. I produced this figure with bioinformatics support from Dr. Attila Jady.

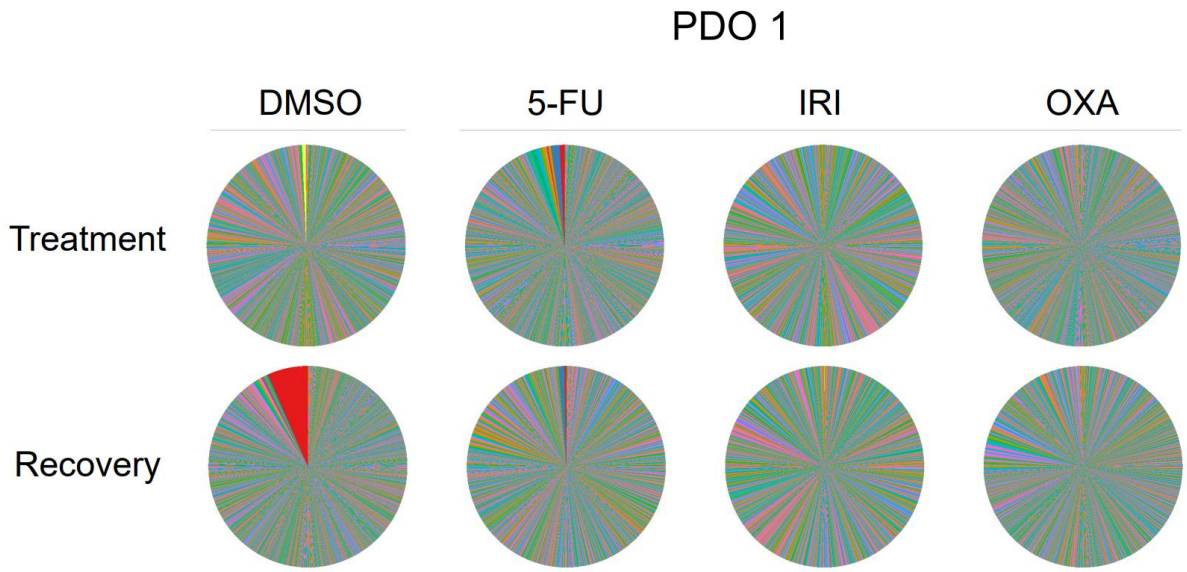
single barcodes showed frequency ranging between 6.5x10E-5 to 0.2% in DMSO vs 7.8x10E-5 to 2.3x10E-3 % in treated PDO1, 4x10E-5 to 0.04 % in DMSO vs 2.6x10E-4 to 0.43% in treated PDO2 and 2.6x10E-4 to 6.4% in DMSO vs 2.6x10E-4 to 9.8% in treated PDO3 (Figure 22). This suggests a slow enrichment of individual clones with passaging for an additional weel, observed in both treated and untreated cells, with a slightly enhanced enrichment in treated cells.

To further evaluate whether a second treatment cycle affects the clonal complexity of barcoded PDOs, I treated recovered cells of 2 cultures (PDO2 and PDO3) with an additional dose of chemotherapy for 5 days. Each barcode represented 4.9x10E-4 to 0.07% of the total cell population in DMSO treated cells of PDO2 in comparison to 6.2x10E-5 to 0.43% in chemotherapy treated cells, and a range of 3.7x10E-4 to 8.3% in DMSO treated cells of PDO3, in comparison to 3.7x10E-4 to 19.7% in chemotherapy treated cells (red color-depicted clone in 5-FU re-treatment) (Figure 22). This suggests that pre-existing individual clones, in comparison to others, may display

higher growth potential which is independent of, but subtly enhanced by chemotherapy treatment after 3-4 weeks.

scRNA barcode sequencing

To quantify clonal dynamics in response to treatment on the single cell level, the percentage of cells sharing the same barcode transcript in the total cell population of a sample was depicted. Targeted scRNA barcode sequencing of both DMSO and chemotherapy treated barcoded PDOs demonstrated that each barcode represented a range of 0.05 to 1.41 % of the total cell population in PDO1 and 0.01 to 0.1% inPDO2, and a range of 7 to 8.5 % in PDO3 (Figure 23). After one week of recovery in drug-free culture media, individual barcodes represented 0.02 to 6.59% (red depicted clone) in DMSO and treatment recovery cells of PDO1. In PDO2, this ranged between 0.1 and 6.4 %. The most frequent barcodes were detected in the DMSO controls (red depicted clone 6.59% in PDO1; blue and red depicted clones in PDO2, 6.4 and 2.3% respectively). In PDO3, individual barcode frequency ranged between 0.37% and 18.2 %. The most abundant barcode (violet color) was detected in 5-FU recovery sample, while it was among the most abundant barcodes in the DMSO sample (7.9%). In re-treatment, PDO2 most abundant barcode was detected in the DMSO control (6.9%) while the most frequent barcode in PDO3 was detected in 5-FU re-treatment (34%). This analysis I performed indicates a slow enrichment of individual clones with passaging, observed in both treated and untreated cells, as observed in DNA sequencing (Figure 22), without substantial outgrowth of individual pre-existing clones under treatment for 5 days.



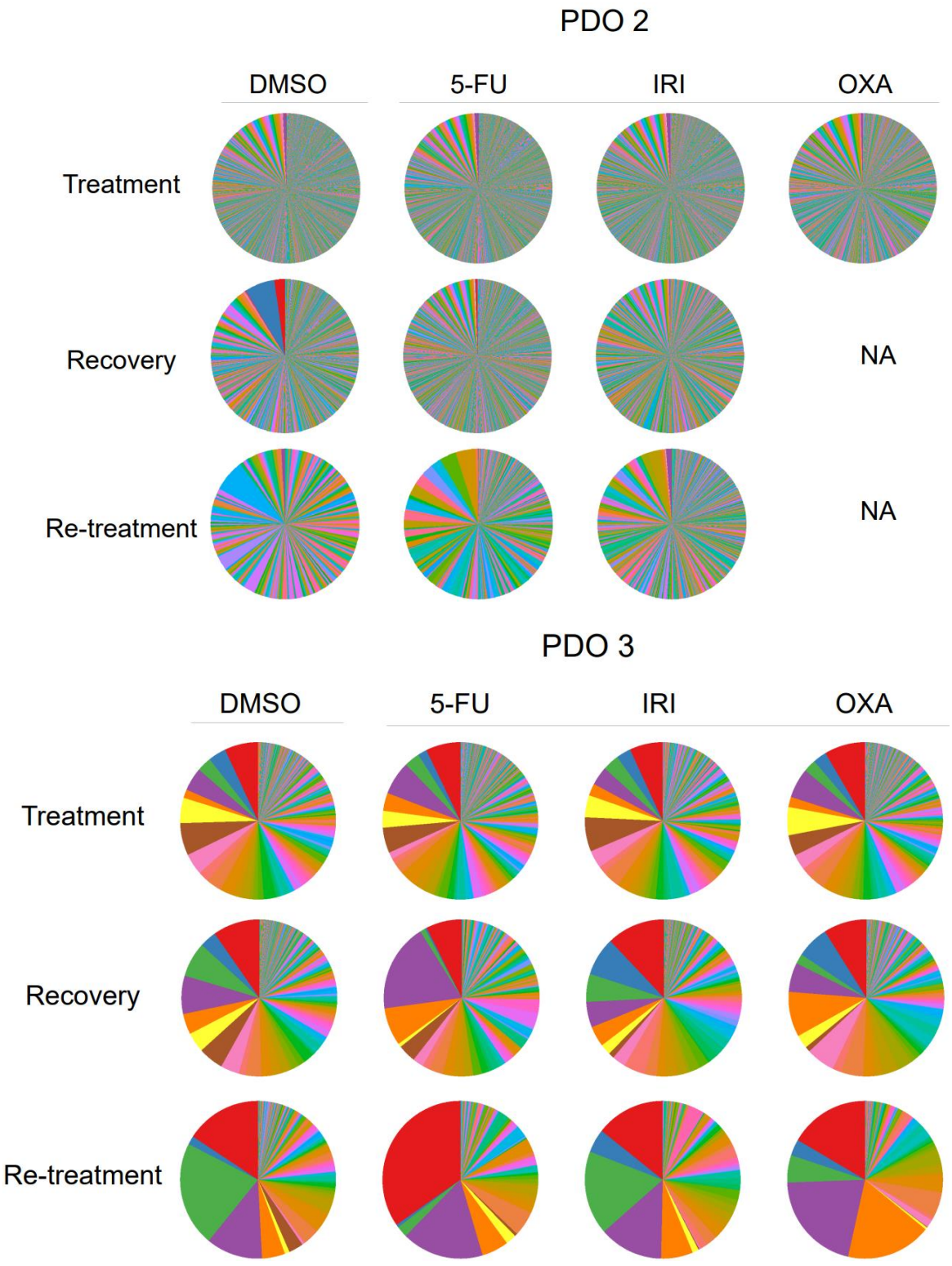


Figure 23. Clonal distribution in treated PDOs –scRNA seq. Pie charts depicting barcode distribution (in %) in treated PDOs, captured with scRNA sequencing. DMSO and chemotherapy treated PDOs are illustrated, in addition to recovered and re-treated timepoints whenever a sample was available. Each color

represents a unique clone. Gray color represents clones of low representation (less than 0.1% of total barcode count). NA: not available

To examine the clonal complexity through the total number of barcodes in each sample, the clonality score was calculated. I found out that clonality score was comparable in treated samples and untreated control at each time point (Figure 24), indicating that chemotherapy treatment has an undetectable impact on clonal complexity or individual clonal outgrowth in the total period of four weeks and suggesting an enrichment of individual clones with passaging, regardless of treatment (Figure 23).

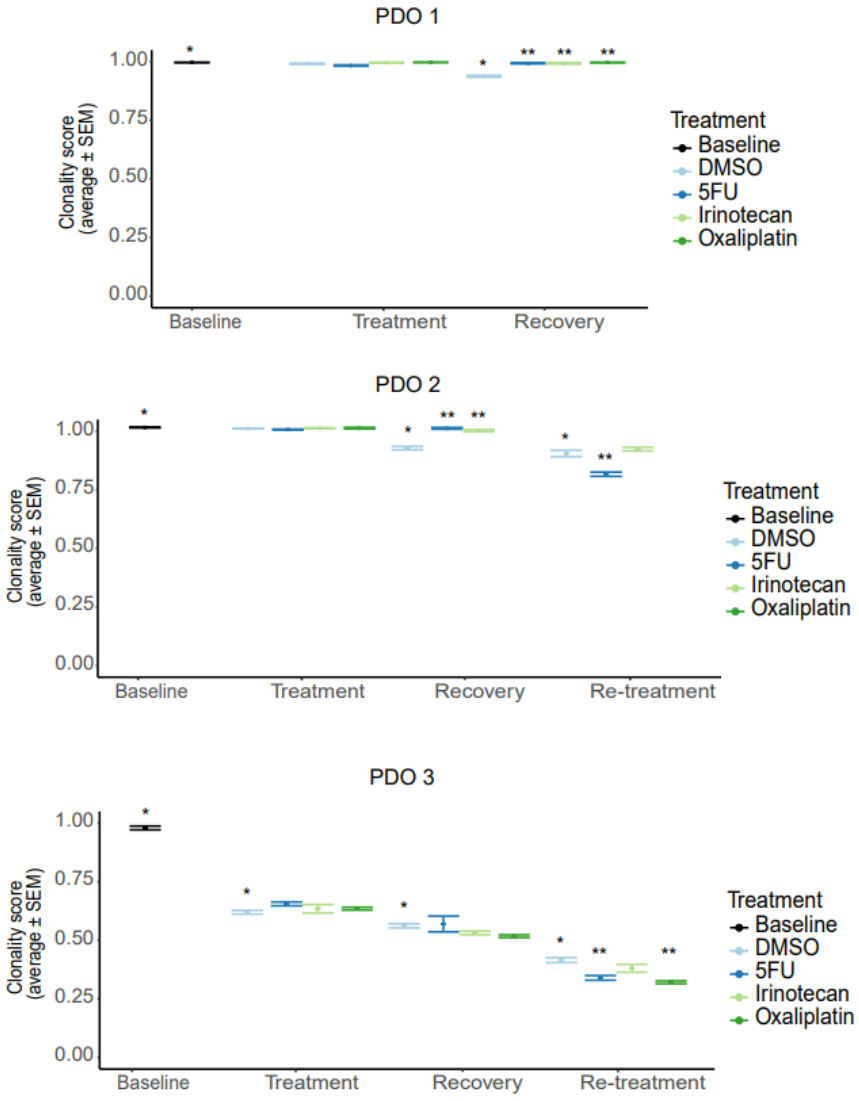


Figure 24. Clonal diversity in chemotherapy treatment of PDOs. Plots illustrating the clonality score of baseline, DMSO and chemotherapy treated samples in each PDO. Statistical significance is indicated by an adjusted p-value <0.05. Significant CS changes are indicated with * between DMSO samples of different timepoints, while significant changes within samples of the same timepoint are indicated with **. Samples without asterisks are not statistically significant.

2.10 Treatment tolerant cells enrich particular pre-existing gene expression clusters

2.10.1 Gene expression heterogeneity in chemotherapy treated PDOs

To understand the gene expression response to chemotherapy treatment, scRNA seq data from samples of all time points (pre-treatment, treatment, recovery and DMSO control samples) were integrated in one projection of unsupervised clustering for each patient. Untreated, treated and recovered cells distributed among the regions of transcriptional space projection in various shared clusters (clusters: PDO1:14, PDO2:14, PDO3: 12) (Figure 25), indicating that CRC PDO cells maintain an intrinsic transcriptional heterogeneity despite external stress factors. To characterize individual transcriptional cell clusters in each patient, I denoted them using their gene expression markers and gene set enrichment analysis GSEA terms.

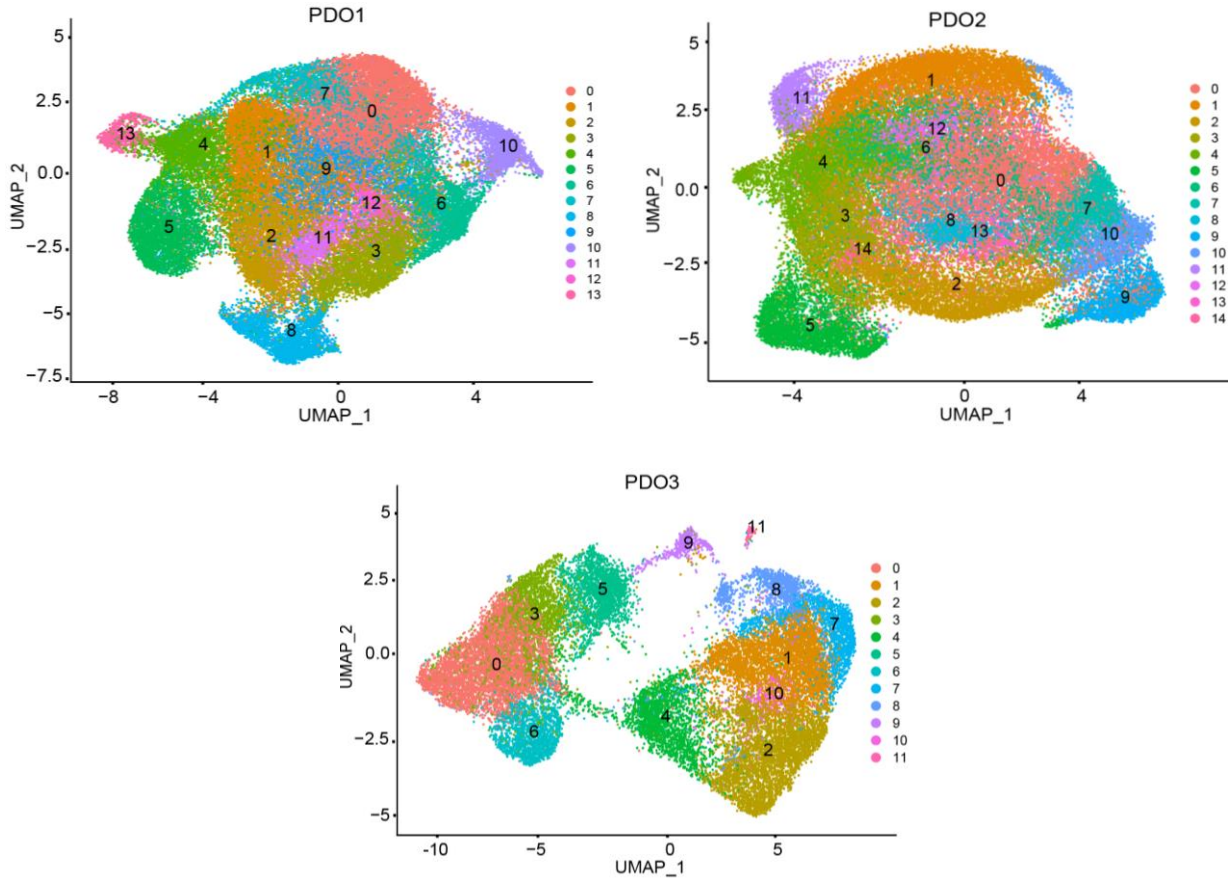


Figure 25. Gene expression of treated PDOs. UMAPs of 3 treated PDOs. Each color depicts an individual cell cluster. Cell cluster in the transcriptional space according to similarities in their gene expression.

Differential gene expression analysis of scRNA integrated data (DMSO control + chemotherapy treated i.e. 5-FU, OXA and IRI + recovered cells of each sample, n=9) from each patient showed highest expression of genes related to highly mitotic proliferative cells (KI67, MCM5, PCNA, TOP2A), normally expressed on the top of colon crypts (Y. Wang et al., 2020), in particular clusters (clusters 5 and 7 of PDO1; clusters 11 and 12 of PDO2; clusters 0, 1 and 4 of PDO3; avg log2FC>1; p-value <0.01) (Figure 26, Supplementary figure 9). Other clusters (cluster 0 of PDO1; clusters 1 of PDO2; clusters 5 and 6 of PDO3) displayed positive differential expression of cell cycle markers CCNB1 and HMGB1 (log2FC 0.46 and 0.36, p-value adj. <0.05) (Hirata et al., 2013; Tang, Kang, Zeh, & Lotze, 2010) among top 10 genes, but negative differential expression of proliferation markers (KI67, MCM5, PCNA, TOP2A; log2FC< -0.3, p-value <0.01). In addition, transcripts of SOX9 and CTNNB1, expressed at the lower non-differentiated side of the crypts (Hirata et al., 2013; Ramalingam, Daughtridge, Johnston, Gracz, & Magness, 2012), were positively expressed in this cluster (Log2FC >0.3; p-value <0.01). The expression of respiratory genes MT-ATP6 and MT-CO3 ranked highest within cells of this transcriptional program (Log2FC >1; p value <0.01). GSEA analysis of both transcriptional programs (TA proliferative and non-differentiated cycling cells) showed an enriched expression of cell cycle progression factors and hallmarks (gene sets representing well-defined biological processes) (G2M, E2F, MYC and DNMT1 NES>2; p value <0.01) (Figure 27, Supplementary figures 10, 11). Clusters 2 in PDO1, 3 in PDO2 and 8 in PDO3 differentially expressed secretion factors TFF3 and SCD (avg Log2FC>1), typically expressed by goblet secretory cells (Rios-Esteves & Resh, 2013; Y. Wang et al., 2020). Clusters 3 in PDO1 and PDO2 and 2 in PDO3 were annotated with high expression of enterocyte FABP1 (Log2FC > 2; p-value<0.05) and epithelial (KRT19, CDH17 avg log2FC 0.84; p-value <0.01) marker genes. Endocytic and absorption processes typical to enterocytes were enriched in this cluster (NES>2) (Figure 27, Supplementary figures 10, 11). Clusters 6, 8 and 6 in PDO1, PDO2 and PDO3, respectively, demonstrated highest expression of the enterocyte and enterocyte progenitor markers FABP1 and FABP5 among other clusters (log2FC > 0.5; p-value <0.05), and thus was annotated as progenitor cluster. Clusters 1 in all three PDOs, in addition to cluster 11 in PDO3 harbored highest expression of stem cell fetal-markers and fate determinants NOTUM, SOX4, PROX1, NKD1 and SEMA3A (Y. Wang et al., 2020), and thus were annotated as the stem cell cluster. This transcriptional program reflected high expression of MAPK and NOTCH signaling in PDO1 (NES >2; p value <0.01) (Figure 27, Supplementary figures 10, 11) and downregulation of MYC and MTOR pathways in all three PDOs (NES<-2; p value <0.01). Accordingly, I annotated each cluster with their highest expressed intestinal cell type marker, supported by previous publications (Figure 26) (Uhlitz et al., 2021; Y. Wang et al., 2020). This suggests that CRC PDOs reflect intrinsic gene expression phenotypes, reminiscent of intestinal

cell features which can perform different functions of absorption, secretion, proliferation, stemness and invasion. In addition, this shows that transcriptional clusters of PDOs resemble those detected in xenografts.

Figure 26. Cell type marker genes of transcriptional clusters of treated PDO1. Plot depicting the gene expression level of selected genes (vertical axis) in assigned transcriptional clusters (horizontal axis) of treated PDO1.

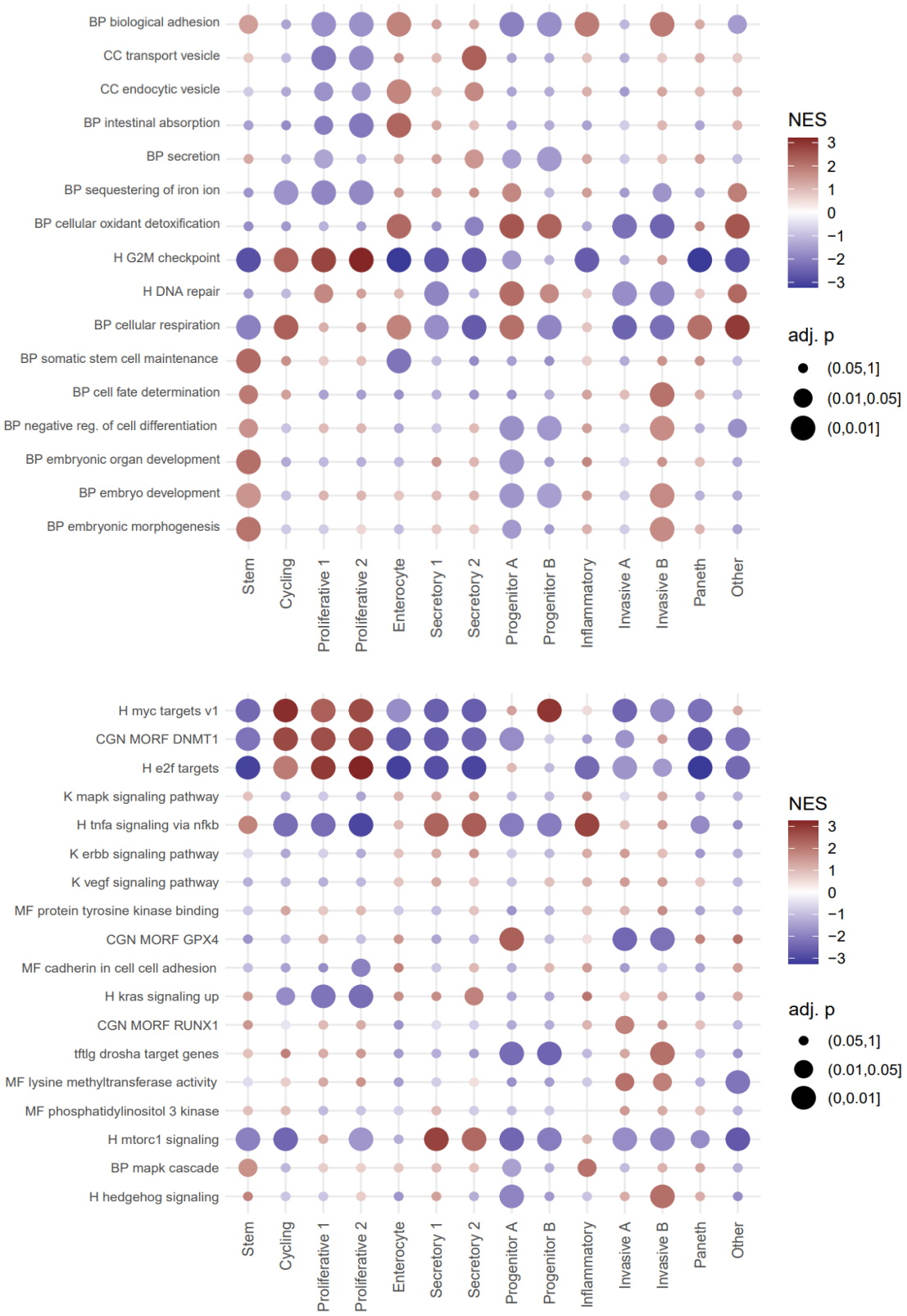


Figure 27. GSEA markers of transcriptional clusters. First panel: plot showing the normalized enrichment score of GSEA terms biological processes and hallmarks in cell type-assigned transcriptional clusters of PDO1 integrated scRNA data. BP: biological process; CC: cellular component; H: hallmark.

Second panel represents molecular pathways. CGN: cancer gene neighborhoods; MF: molecular function; K: canonical pathway.

2.10.2 Transcriptional treatment response

To depict gene expression of PDO cells in treatment, I quantified the frequency of transcriptional clusters in baseline, DMSO control, treated and recovered samples. Undifferentiated cycling cells were highest in all PDOs, representing 15 to 22% of all baseline cells. Stem cell clusters were second highest in PDO2 and PDO3 (15 to 17%) while the secretory cell cluster ranked second in PDO1 (14.8%) (Figure 28, Supplementary figure 12). The enterocyte cell cluster ranked third in PDO2 and PDO3 (9.5 and 19.5%, respectively) and the proliferating cell cluster was the third highest in PDO1 (10.8%) (Figure 28). This indicates that PDO1 has higher proliferative and secretory features than PDO2 and PDO3 which demonstrated higher stem and enterocytic features. In addition, this analysis shows that particular transcriptional clusters detected in xenografts (cycling, proliferative, enterocyte, secretory, progenitor) are also detected in cultured PDOs. However, the invasive clusters were less abundant in PDOs compared to xenografts (2.4-8.4% in PDO1, 2.5% in PDO3 and 2-20% in PDO1 xenograft, 0.5-27% in PDO3 xenograft).

Proliferative (TA) cell program, characterized with high cell cycle and DNA activity (Figure 27, Supplementary figures 10, 11), was reduced in response to 5-FU and Irinotecan treatment compared to the DMSO control (avgFC 2.5) (Figure 28). Indeed, Irinotecan is known for cell killing by poisoning topo-isomerase (Mathijssen, Loos, Verweij, & Sparreboom, 2002), which is highly expressed in proliferative cells cluster (Figure 26). This cluster remained low in treatment recovery samples (2.3 to 4.6%). This cluster was lowly represented in PDO2 and PDO3 baseline samples (2.6% and 6.3%, respectively) and remained so after treatment (1.86-3%). Secretory goblet-like cell cluster was also reduced upon treatment in PDO1 (FC 2.56) and remained underrepresented in treatment recovery at t2 in comparison to DMSO control, after drug washout (FC 1.9). This cluster was lowly represented in PDO2 and PDO3 baseline samples (6 to 8%) and remained so, or further declines after treatment (2.72 to 9% in PDO2 and 6.1-10 % in PDO3) (Figure 28, Supplementary figure 12).

Cells of the differentiated enterocyte phenotype were enriched in treatment response of PDO1 (8.4 to 14%) in comparison to baseline and DMSO controls (5.6-8.4%) and remained around this value in recovery samples (Figure 28). In PDO2 and PDO3, the frequency of this cluster was among highest in DMSO and treatment samples (9 to 12% in PDO2 and 13 to 19.5% in PDO3). This transcriptional state resembling the enterocyte phenotype is characterized by negative

expression of gene set enrichment (GSEA) pathways of DNA replication and MYC targets (Figure 27, Supplementary figures 10, 11). I found out that among top 15 genes in this transcriptional cluster are Annexin ANXA1 (log2FC 1) and calcium-binding protein encoding gene S100A11 (log2FC 0.85), which are required for wounded plasma membrane repair and survival (Figure 26) (Jaiswal & Nylandsted, 2015). In addition, secretory leukocyte protease inhibitor SLPI, necessary for wound healing (Ashcroft et al., 2000), peptidase inhibitor 3 functional in tissue injury prevention (Tejera et al., 2014), and LGALS3 mediating treatment resistance in cancer (Wang et al., 2019), were among top 15 highly expressed genes in these cells (log2FC>1) (Figure 26).

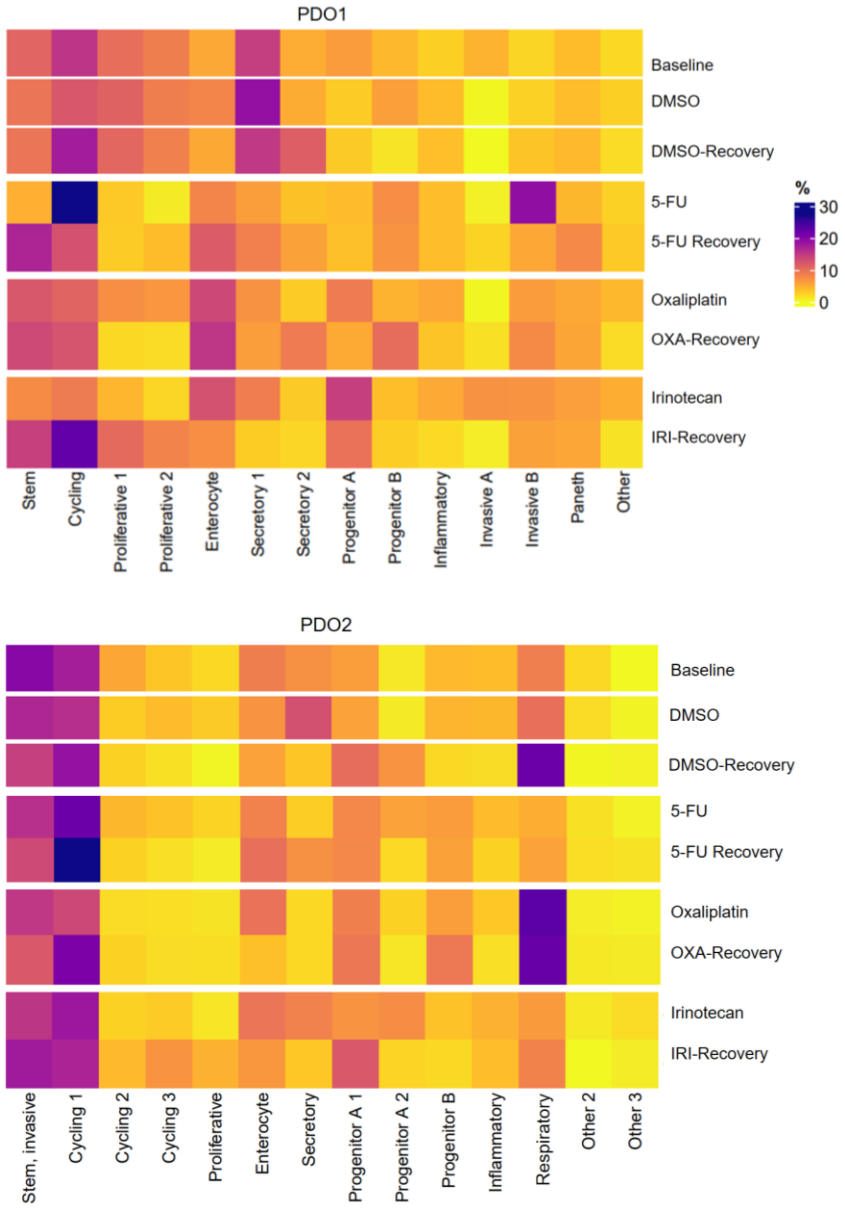


Figure 28. Cluster enrichment in treated PDOs. Heatmap illustrating cluster abundance (in %) in baseline, treated and recovered cells of PDO1 and PDO2.

This transcriptional program remained enriched in treatment recovery at t2, after drug washout (7.3%-10.3% in 5-FU rec and IRI rec of PDO2 and 10-17% in recovery samples of PDO3) (Figure 28, Supplementary figure 12).

The enterocyte progenitors program (progenitor A), was elevated in Irinotecan response in PDO1 and PDO2 (3 to 14% and 1.6 to 8%, respectively) (Figure 28). I observed that cells of this transcriptional phenotype are characterized by detoxification and anti-oxidant activity (Figure 27). Top 15 expressed genes include anti-oxidant and detoxification markers GPX1, ATOX1, and GSTP1 (Jin et al., 2022; Schnekenburger, Karius, & Diederich, 2014; Y. Zhao et al., 2022), in addition to SLPI, PI3 and S100 of differentiated enterocytes. The abundance of enterocyte progenitor clusters were maintained in Irinotecan recovered cells in comparison to DMSO control (10.7% in PDO1 and12.8% in PDO2) (Figure 28).

The undifferentiated cycling cell program was among the three highest abundant clusters in all conditions (control, treatment, recovery) of the 3 PDOs (9-28% in PDO1; 14-28.5% in PDO2; 10.5-20% in PDO3), suggesting this transcriptional profile to be a treatment resistance related phenotype. Among active GSEA motifs in these LGR5 positive cells are proliferation associated factors HOXA4, MYC and DNMT1 in addition to the biological process of cellular respiration (NES >2; p value <0.01) (Figure 27, Supplementary figures 10, 11).

Treatment recovery

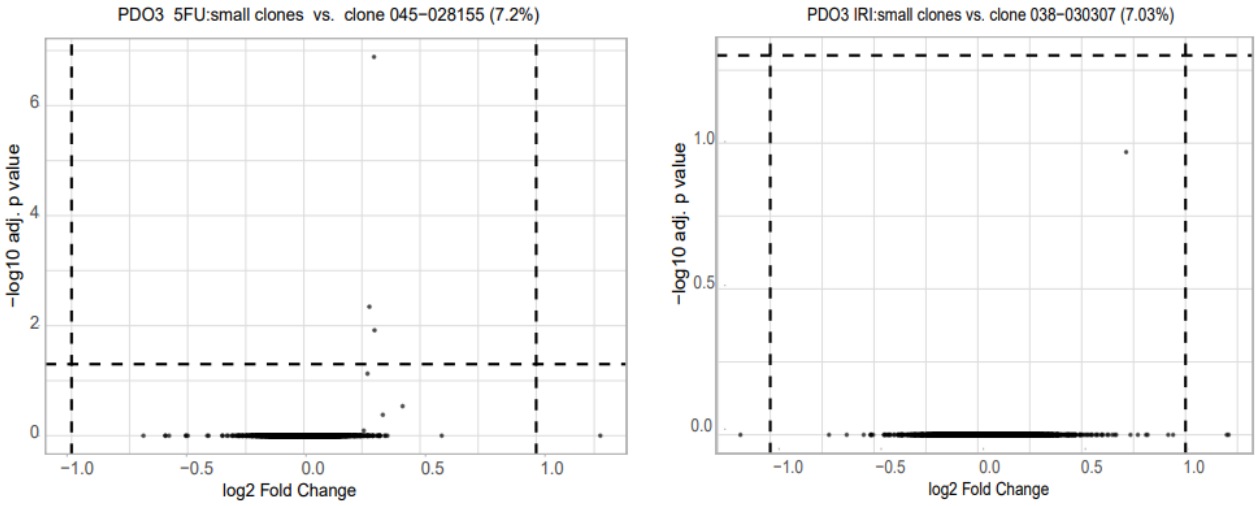
After culturing CRC PDOs for one week with drug-free media, the frequency of LGR5+ undifferentiated cycling and stem cell phenotype clusters was maintained among highest in all three PDOs (7 to 17% in PDO1, 12 to 28% in PDO2 and 15 to 22.4% in PDO3) (Figure 28, Supplementary figure 12). This indicates an association of both programs to treatment response and recovery and suggests that CRC cells can adapt to chemotherapy stress through particular pre-existing transcriptional programs.

2.10.3 Clonal gene expression dynamics in response to therapy

Since abundant clones did not show differential gene expression in comparison to small clones in barcoded PDO xenografts, I further compared the differential gene expression of small versus major clones of treated PDOs *in vitro*, to understand whether major clones (forming >10% of the cells of a given sample) have transcriptional features differentially expressed in comparison to small clones (forming less than 1% of a sample cell population). Differential gene expression analysis (DEG) was carried out between these major clones captured by scRNA seq either in treatment or DMSO control, in comparison to small clones, whenever cell numbers permit (>50 cells per clone). I observed no significant particular inter-clonal differences in gene expression were observed (Figure 33.A).

To detect differences of gene expression in the same clone in DMSO and chemotherapy treatment, DEG was performed between cells of the same clone in two conditions (DMSO and chemotherapy). When carrying this intra-clonal (cells within the same clone) comparison, I observed that differential expression of gene sets related to particular treatment response clusters (enterocyte, progenitor) appears (Figure 33.B). These genes include S100A11, ANXA1 and KRT20 associated with the enterocyte and enterocyte progenitor (progenitor A) transcriptional program (Figure 26). This again suggests that clones with different growth dynamics express similar gene sets and that clones respond to chemotherapy by modulating their gene expression into pre-existing transcriptional phenotypes.

A



B

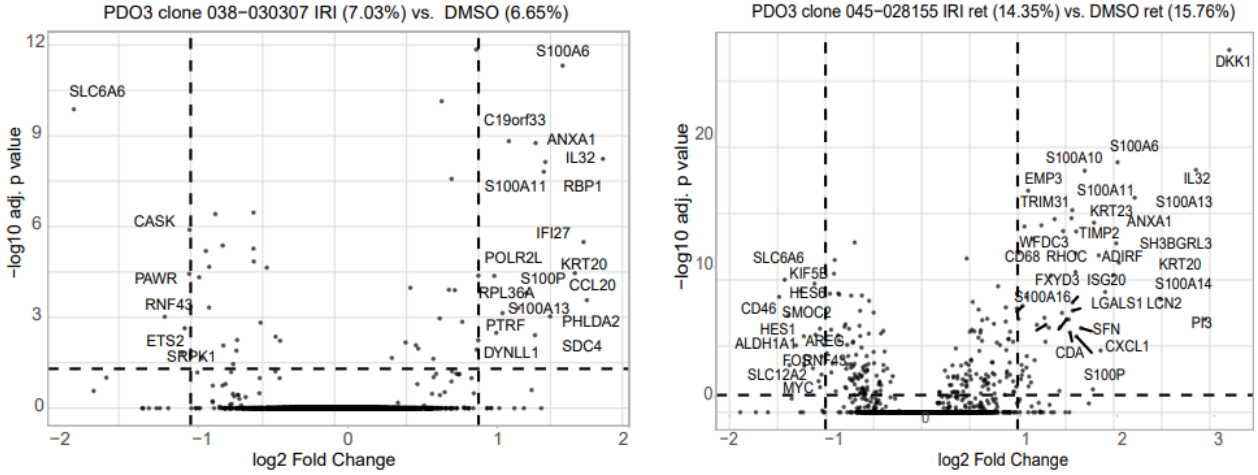


Figure 29. Intra and inter-sample clonal gene expression. A. Volcano from PDO3 showing differentially expressed gene in abundant vs small clones. B. Intra-clonal violin plot of individual clones in treatment versus DMSO control. OXA: Oxaliplatin. Small clones: collective clones individually forming less than 1% of all clones. Statistical significance is indicated in the y-axis, all represented gene names are statistically significant (adjusted p-value <0.05)

2.10.4 CRISPRi supports the significance of top differentially expressed genes in treatment response and recovery-associated clusters

This experiment was performed and analyzed by Vida Kufrin and Alexander Wurm, NCT/UCC Dresden.

Since chemotherapy treated and recovered cells enrich specific transcriptional clusters (cycling, stem, enterocytes), I hypothesized that genes expressed in these clusters might be relevant in chemotherapy treatment response. To assess their functional relevance, these genes were targeted by means of a customized knock-in clustered regularly interspaced short palindromic

repeats (CRISPRi) screen, in 2 separate CRC models (3D CRC sphere culture and 2D HT29 cell line), in combination with chemotherapy treatment (Figure 29). I selected a set of common upregulated genes in treatment response clusters shared among at least two BC-PDOs (181 genes) (Supplementary table 5).

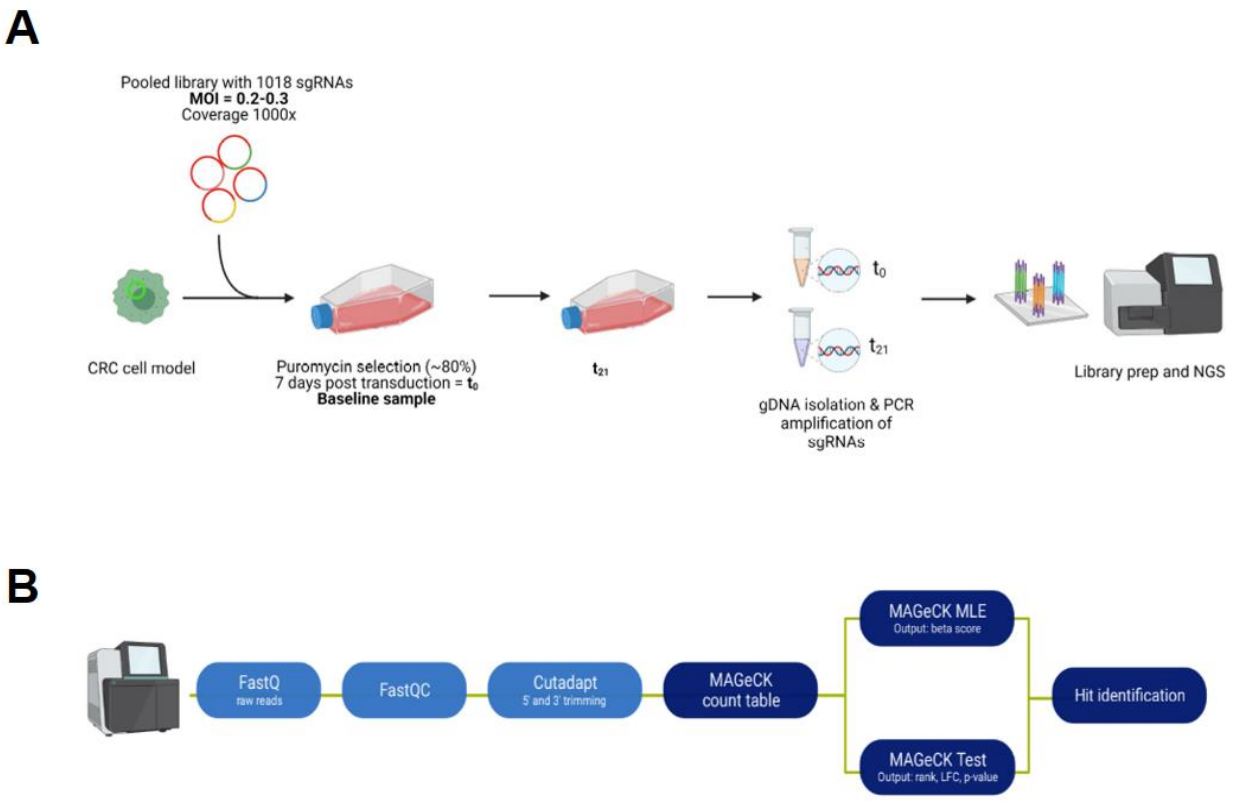


Figure 30. Experimental settings of CRISPRi screen. A. Experimental procedure: Cells of 2 CRC models were transduced with sgRNAs, selected in puromycin and treated with chemotherapy at t0. Cells pellets were collected at day 21 and genomic DNA was isolated for sgRNA sequencing. B. Analysis pipeline: MAGeCK software was used for analyzing gRNA abundance and for hit calling. This figure was created by Vida Kufrin.

After confirming the expression of RNA guides of selected genes in these two models, cells were either treated with chemotherapy (5-FU or OXA) or DMSO in 3 technical triplicates for 48 hours. I reasoned that downregulating (knocking-down) genes upregulated in response to treatment in PDOs would enhance the sensitivity of cells to treatment, leading to an increased decay of these cells compared to those with downregulation of non-relevant genes. After drug washout (Figure 29), guide RNA amplicon sequencing of treated cells (day 21) revealed depletion of 23 genes in

oxaliplatin samples (-7.5 <Log2FC<0; log score >2) and enrichment of 3 genes (0 <Log2FC<2.5; log score >2) in comparison to the DMSO control. In 5-FU treatment, 20 genes were depleted (-15 <Log2FC<0; log score >2) and 6 were enriched (0 <Log2FC<2.5; log score >2) in 3D CRC culture, while HT29 showed a similar trend (Figure 30). Control core essential genes MYC and RPS27 were depleted in control cells (-12.5 <Log2FC<0; log score >2).

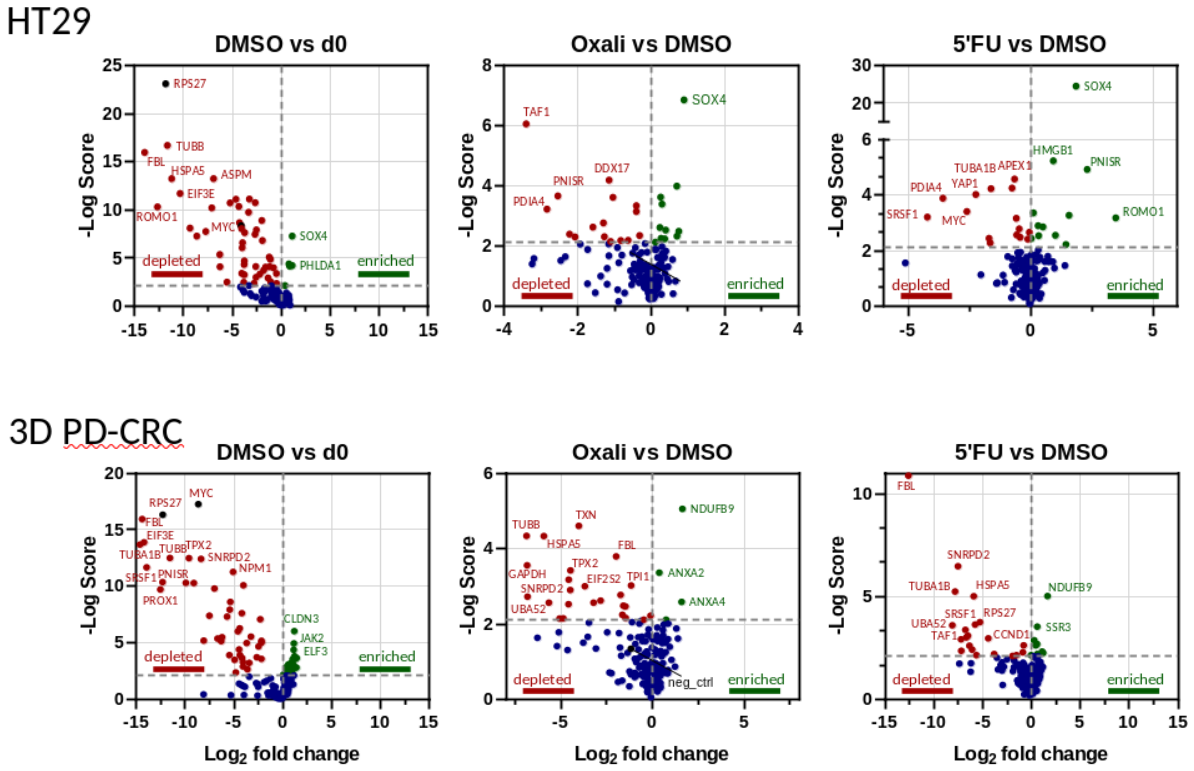


Figure 31. Depletion of cells with knockdown of genes elevated in treatment response transcriptional programs. Cell abundance is determined through NGS read count of guide gRNA. Depleted and enriched genes in treatment were corrected to genes of the DMSO control (left panel). Statistical significance is indicated in the y-axis, all represented gene names are statistically significant (logscore>2). This figure is created by Vida Kufrin and Alexander Wurm.

This indicates that a set of selected top expressed genes in identified treatment response transcriptional programs demonstrate correlation with reduced cell survival when downregulated under chemotherapy treatment. SRSF1 was identified among the common factors in both treatments and models (Figure 31). SRSF1 is an oncogenic splicing factor and a critical transcriptional target of MYC, which acts as a key mediator of tumor cell plasticity (Das, Anczuków, Akerman, & Krainer, 2012). MYC, being considered a positive control gene in this screen, was depleted (Figure 30, 31; log score >2). Mutual positive expression of canonical MYC and SRSF1 is solely maintained in undifferentiated LGR5 positive cycling cell clusters (PD1:

cluster 0, PDO3: cluster 0, 1; PDO3: cluster 1), which are sustained in chemotherapy treatment and recovery (Figure 28).

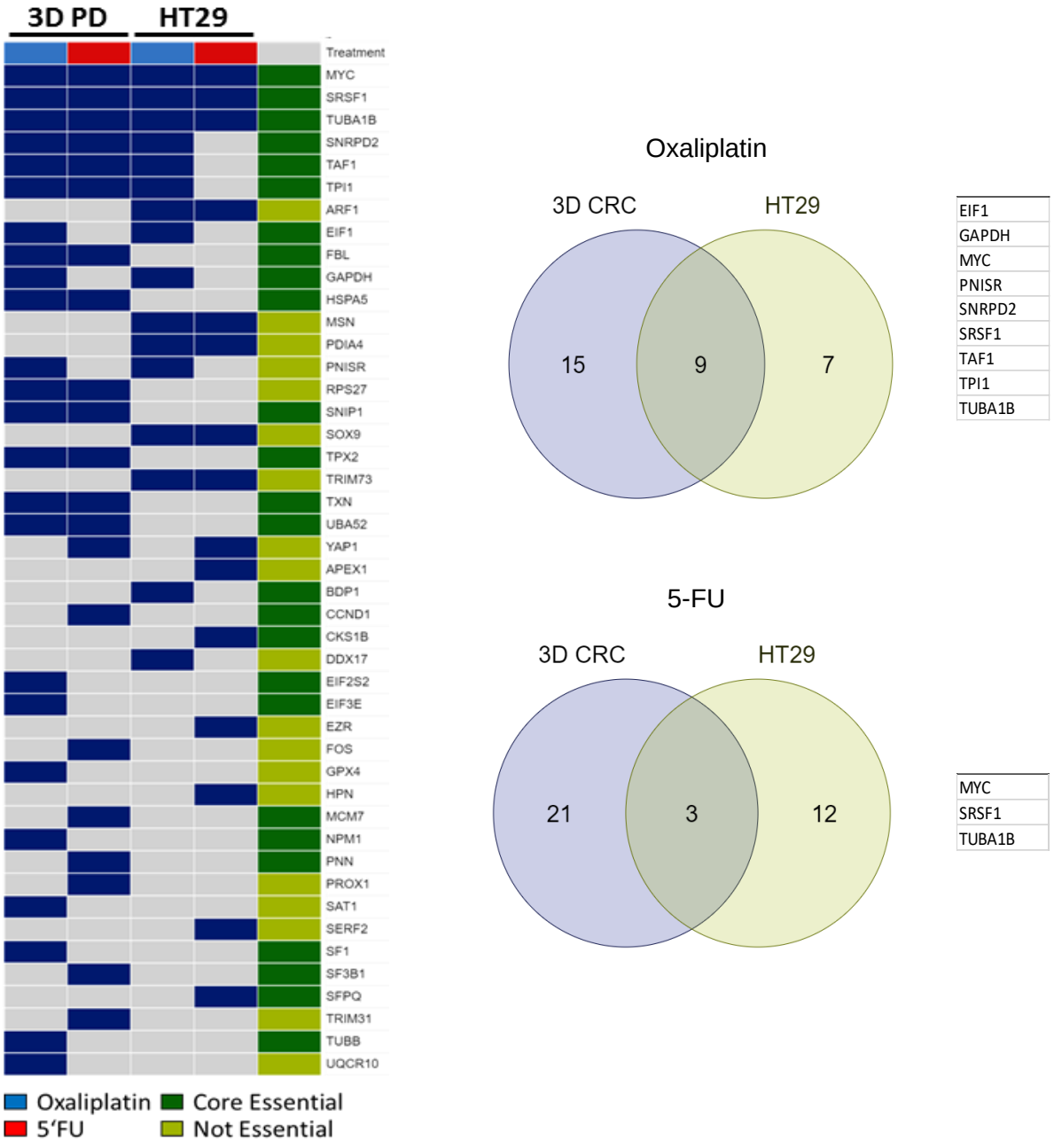


Figure 32. Depleted genes of CRISPRi custom screen in treated CRC models. Plot showing common genes depleted after chemotherapy in CRISPRi cells (left panel). Venn diagram displaying shared depleted genes and their names in 2 CRC models of the CRISPRi custom screen. Core essential: genes coding for machineries of all eukaryotic cells such as the ribosome, RNA polymerases and central metabolic enzymes.

2.10.5 Prognostic value of identified treatment response and recovery genes

To examine whether top genes identified in abundant cell clusters of drug response have a prognostic value, I analyzed their expression and correlation with patient prognosis in publicly available RNASeq dataset. This dataset corresponds to a colorectal cancer metastatic cohort colon and rectum adenocarcinoma (COADREAD) included in the cancer genome atlas (TCGA) ("Comprehensive molecular characterization of human colon and rectal cancer," 2012). I assessed 27 marker genes of treatment response (enterocyte, cycling), recovery (stem cell phenotype) and invasive transcriptional clusters of PDO1 (Figure 26). The majority of the genes that showed significance (p-value <0.05) (15 out of 16 genes) correlated with poor/infavorable prognosis indicated by a reduced patient overall survival, disease specific survival (time from the start of treatment until the reappearance of cancer-related symptoms) or disease-free interval (time from the start of treatment until cancer progresses into metastasis) (Figure 32, Supplementary table 6). This observation further supports the relevance of identified transcriptional programs pertinent to treatment response in scRNA data of treated PDOs.

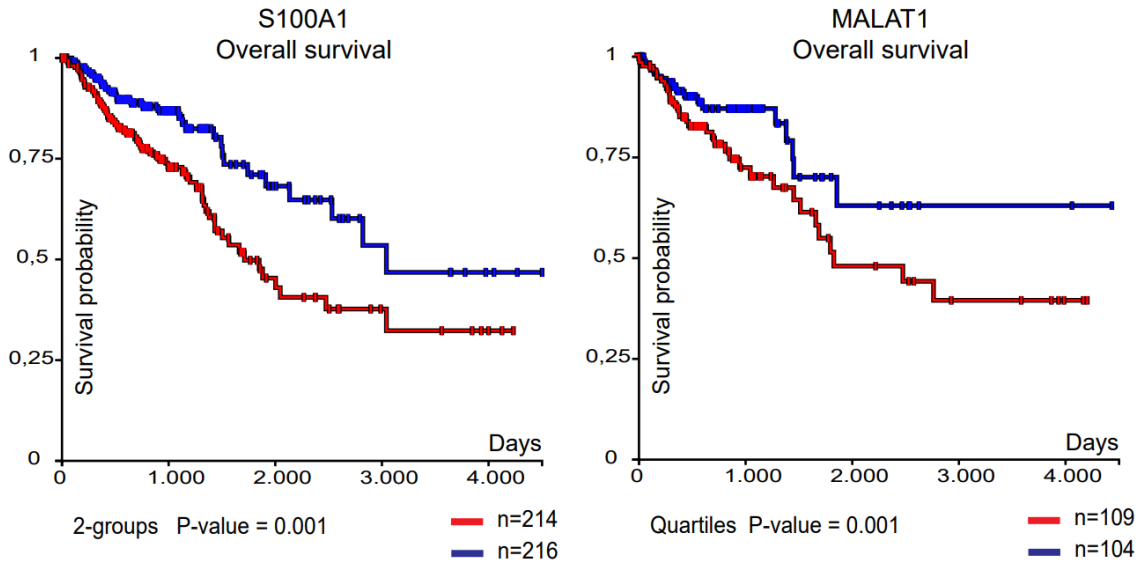


Figure 33. Prognostic value of genes identified in our approach. Kaplar Meier plots showing the survival probability of patients with high gene expression (RNA seq) of individual genes. Results of 2 differentially expressed genes identified in treatment response transcriptional programs are displayed. Statistical significance is indicated by p-value. 2-groups: samples divided on the median. Quartiles: upper vs lower quartile of the samples.

2.11. Chromatin regulation in chemotherapy treatment of CRC PDOs

Treated PDOs enriched particular transcriptional clusters in response to treatment (Figure 28). To gain further insights into mechanisms facilitating the expression of genes in transcriptional clusters enriched in chemoherapy treatment response, I analyzed the open chromatin status of treated cells was analyzed. Exploring the open chromatin status of treated cells could help understand epigenetic mechanisms regulating the expression of treatment associated programs, and depict chromatin remodeling in treated in comparison to untreated cells. Transcription takes place in DNA regions with an open chromatin state, where nucleosomes at gene promoters and enhancers are modified to allow transcription initiation (Tsompana & Buck, 2014). Accordingly, regions of highly expressed genes in treatment response are expected to be accessible to the transcription machinery. I treated 1x10E6 cells of two PDOs (PDO2 and PDO3) derived from the baseline cells with either 5-FU or DMSO as performed in the scRNA experiment (5 days of treatment at IC70), and performed transposase-accessible chromatin single cell sequencing (scATAC-seq). For each treated PDO, merged cluster visualization of cell populations with distinct chromatin accessibility patterns was produced. 3 major cell clusters per patient were identified (Figure 34). DMSO control cells grouped in one cluster which I assigned as ‘untreated’, since it harbors control cells not treated with chemotherapy, whereas treated cells clustered in another profile that was termed as treated cluster.

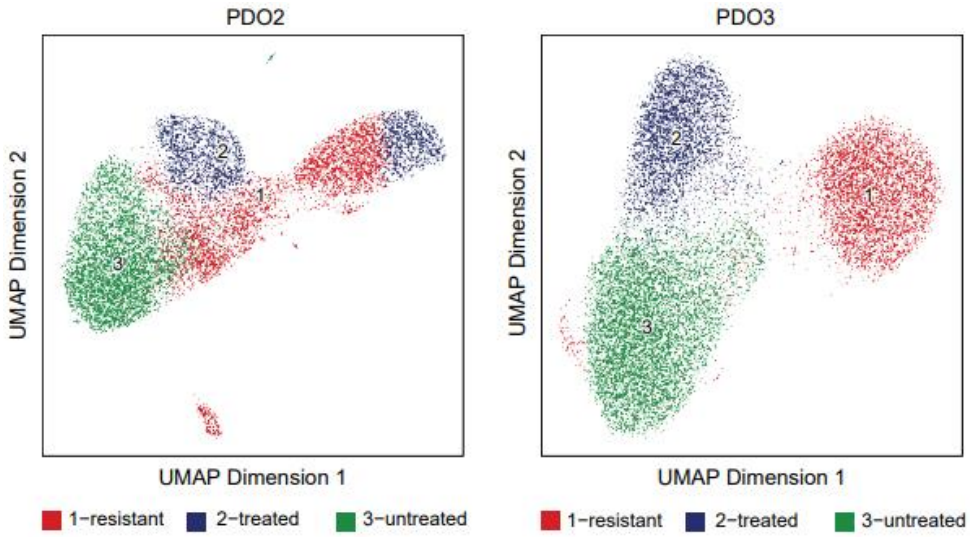


Figure 34. Cells of distinct open chromatin regions in chemotherapy. UMAPs depicting clusters of single cells with distinct open chromatin status. Merged data from DMSO and 5-FU treated PDOs is shown. I produced this figure with bioinformatics support from Dr. Attila Jady.

A third cluster contained cells from both conditions i.e. untreated and responsive, suggesting that cells within this profile are pre-existing treatment tolerant cells whose chromatin status is not largely affected by treatment.

Untreated clusters of both cultures showed an open chromatin status in the regions of tumor suppressor genes RNF20, NAV3 (transcriptional target of p73), and DAB2 (negative regulator of WNT/MAPK) (Figure 35). Treated clusters in both patients shared open chromatin regions of common genes DBH-AS1 and COL6A1, related to tumor survival and progression (P. Chen, Cescon, & Bonaldo, 2013; Song, Gao, Peng, & Yang, 2020), suggesting these as changes in chromatin structure in treatment response (Figure 35).

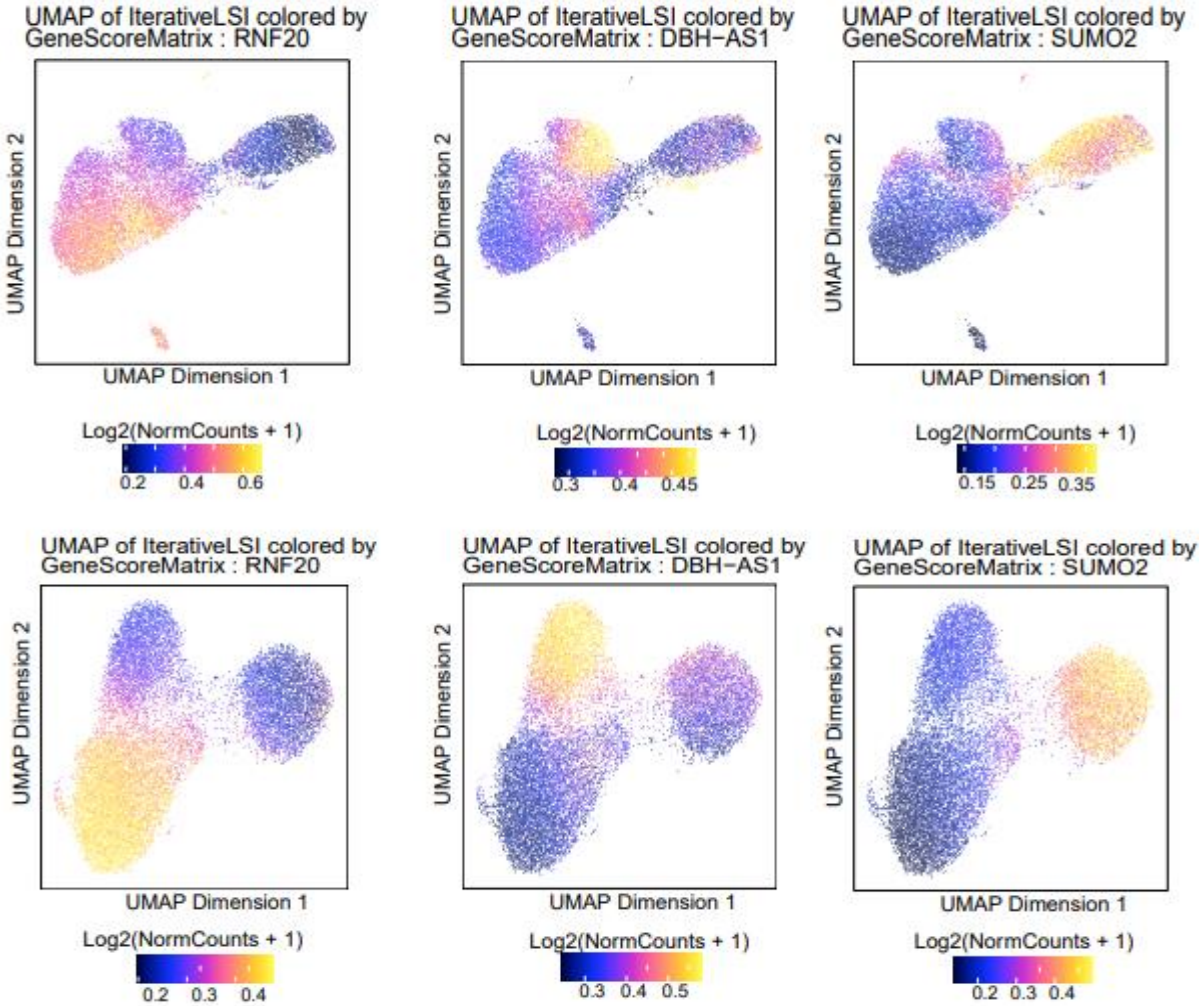


Figure 35. Chromatin cluster marker genes. Open chromatin status of representative marker genes in identified cell clusters. Higher counts indicate a more open chromatin status. PDO2 (upper panel) and PDO3 (lower panel) are depicted.

Moreover, I observed that this cluster showed enriched binding sites of apoptosis and EMT regulator transcription factors of JUND and FOSL2 (Figure 36), in comparison to other clusters. JUND is known to dimer with FOSL1 to form the activator protein complex AP1, which regulates gene expression and biological processes of differentiation and senescence (Sobolev et al., 2022). This shows that identified clusters are characterized with an open chromatic region of distinct genes.

Resistance related cluster comprised cells with distinct open chromatin regions in the proximity of SUMO2, CASKIN2, and PRAMEF12 genes (Figure 35) indicating a potential role of these genes in treatment resistance. SUMOylation (involving SUMO2) was shown to be a critical process in maintaining cell cycle progression and genome integrity (Eifler & Vertegaal, 2015). Transcription factor motif analysis showed open chromatin sites in targets of CTBP1-DT, DNMT1 and SMAD5 in resistance associated cells compared to other cells (Figure 36), suggesting their association to treatment resistance. These motifs are active regulators of tumor progression and DNA damage repair (Chau et al., 2012; Nishiyama & Nakanishi, 2021).

To link open chromatin status to gene expression in chemotherapy response, I assigned open chromatin sites detected by scATAC seqyencing to highly expressed genes and gene sets of 5-FU enriched clusters, detected in scRNA data. Marker genes of treatment response clusters (enterocyte, cycling, stem) with an open chromatin region in treated cells were summarized (Supplementary table 7). JUN, FOS, SUMO2, SMAD5 and SMARCC1, motifs detected in response and resistant clusters of scATAC, were differentially expressed in the invasive and the cycling cell clusters (Figures 26, 27) (Figure 36).

Transcription factor targets of DNMT1 enriched in scATAC resistant clusters, were upregulated in scRNA clusters abundant in treatment and recovery conditions i.e. undifferentiated cycling and stem cell transcriptional programs (Figure 27). This indicates that chromatin modulation in these sites could contribute to gene expression response in treatment and suggests an impact of chromatin modulation on gene expression in stress conditions. Such genome-wide chromatin accessibility approach I performed, with and without treatment, suggests that identified gene expression heterogeneity in chemotherapy treated cell populations can be influenced by chromatin remodeling.

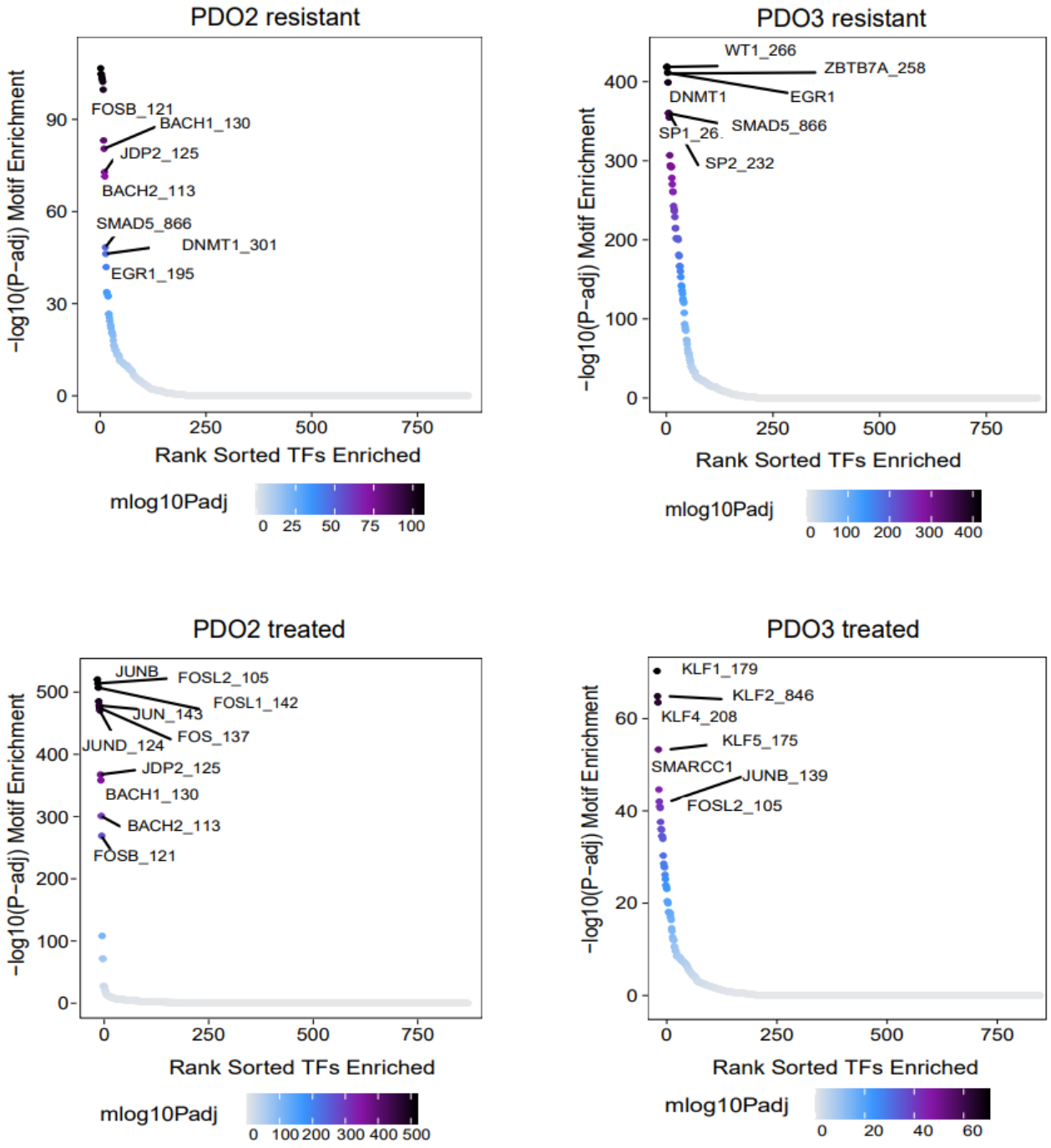


Figure 36. Enriched chromatin motifs of resistant and treatment response cell clusters. Plot displaying enriched chromatin motifs in differential abundance of open chromatin regions in treatment-resistance and treatment-response cell clusters of scATAC data.

3. Discussion

In this study, I utilized colorectal cancer derived organoids to understand the gene expression heterogeneity within individual clones of CRC. I provided novel insights into different gene expression profiles, detected in individual clones, which characterize tumor progression and metastasis in addition to chemotherapy response.

3.1 Orthotopic transplantation of PDOs as a study model

I found that xenotransplantation results showed heterogeneous capacity of CRC PDOs in terms of tumor formation and metastasis development *in vivo* (Figure 5). Disparities in tumor engraftment between different PDOs can be attributed to several factors such as the tumor content and the genetic background of a PDO, but also to the quality of dissociated cells and the precision of tumor cell injection into the mouse caecum (C. Chen et al., 2021). Other possible factors which are not explored here are the mutational burden of PDOs, the tumor stage and the cancer subtype, of which the cultures derive from. This aspect is beyond the purpose of this study, but can be considered when establishing PDO biobanks. This highlights that pre-clinical models such as organoid cultures should be carefully assessed for their tumor cell content for downstream applications, especially in tumor formation and drug response predictions. Discrepancy in metastasis formation rates in xenografts can derive from biological or even experimental factors. The humane endpoint of mice was determined either by palpable tumor detection or the overall health of the mice, which might not permit a sufficient time window for metastasis to form. Overall, orthotopic xenotransplantation provided a valuable demonstration of the functional aspect of PDOs in tumor and metastasis formation and supported their selection criteria for downstream experiments.

3.2 Genetic barcoding of PDOs

The barcoding approach I utilized here served as a valuable tool for tracking the growth dynamics of individual clones in tumor progression and treatment response. I noticed that baseline populations demonstrated an infrequent homogeneous distribution of transduced barcodes both by bulk DNA and scRNA sequencing methods (Figure 6 B, C). PDO3 exhibited a less homogenous baseline population, an observation attributed to the lower initial number of RFP

positive sorted cells in comparison to PDO 1 and 2 (50 x10E3 vs 300 x10E3 and 110 x10E3). More passaging time to reach an adequate cell number similar to the other two cultures, in addition, might have influenced this representation. Similar observations in PDO3 were achieved in downstream clonal dynamics in metastasis and under treatment in comparison to the other two PDOs, which supported the inclusion of this culture in the analysis. Even though a low multiplicity of infection (<0.25) was used, multiple transductions were still observed in less than 10% of the cell population (Figure 6 D). Although implementing a lower MOI in this scenario seems intuitive, it could have compromised the quality of the data, since specific easy-to-transduce cell populations might have been barcoded, limiting the inclusion criteria of heterogeneous cell populations in downstream analysis (Hines & Hines, 2022). This is an important aspect since this study aims at understanding the heterogeneity of CRC, which involves including diverse cell phenotypes existing in PDOs into the analysis. This might be less important in a cell line settings, which are known to comprise more harmonious cell content in comparison to organoid cultures (Corrò, Novellasdemunt, & Li, 2020). There is a tradeoff between avoiding multiple transductions and including additional potential cell populations into the analysis, which needs to be carefully considered.

3.3 Capturing genetic barcodes by NGS

To address the main question of understanding subclonal gene expression in treatment response and metastasis of CRC cells, I implemented an *in vivo* and an *in vitro* experimental arms. The clonal distribution was captured by means of target barcode DNA and scRNA sequencing. DNA sequencing has the advantage of higher coverage in comparison to scRNA sequencing, but it can display data from necrotic tumor cells, meaning a wider representation of the history of clonal distribution including nonfunctional dead cells. In contrast, scRNA data derives from a small proportion of tumor viable functional cells that are representative but not inclusive of the overall cell population. This can explain the difference observed in the abundance of individual clones using these two approaches, but can also suggest that scRNA offers an updated status of the viable functional clonal distribution. Although the contrast in technologies reflects some variations in clonal representation, their combination provides a more powerful approach enhancing the aspect of coverage and accuracy. In addition, the data I obtained suggests similar findings (Figures 8, 9, 22 and 23).

3.4 Clonal dynamics in tumor progression and metastasis of CRC

Upon transplantation of barcoded PDOs into NSG mice, tumors and metastases developed to liver and lung. I observed that barcode sequencing data, both from bulk DNA and scRNA, showed that the high number of diverse homogeneous barcodes in the initial barcoded cells was reduced after tumor engraftment and serial transplantation, whereby only few clones dominated the tumor and metastases (Figure 8, 9). The representation of individual clones fluctuated in tumor, liver and lung metastasis, with either over-or underrepresentation, indicating a dynamic clonal expansion in tumor progression. Indeed, the tumor microenvironment can affect the growth dynamics of individual clones (Lenos et al., 2018). In addition, different abundant clones were observed in the two primary transplants deriving from the same baseline cell population, especially in PDO1 and PDO2 (Supplementary figure 4), suggesting an extrinsic factor that influences clonal expansion. This goes in line with previous reports illustrating dissimilar clonal distribution in primary tumor compared to metastatic lesions (Lenos et al., 2018; Naxerova et al., 2017). Dissimilar distribution of individual clones was previously reported by our group by which different set of clones contributed to disease progression in different stages of xenograft serial transplantation (Dieter et al., 2011; Giessler et al., 2017). In serial transplantation, I noticed that further clonal exhaustion occurred, especially in liver metastasis-derived cell populations (Figure 8, 9) where dominant clones expanded further, indicating a high clonal dominance in later stages of the disease progression. Such phenomenon was reported in earlier studies (Kerbel et al., 1988). Tumor, liver and lung-derived serial transplantations were dominated by different sets of clones suggesting either a clonal selection specific for the microenvironment or a stochastic clonal selection (Figure 8, 9). The difference in the abundance of clones in tumors of primary mice supports a stochastic clonal selection (Supplementary figure 4).

Although some clones grew robustly in specific microenvironments such as the liver and lung or in serial transplantations, another set of clones co-existed in a lower frequency (Figure 8, 9). This suggests that the tumor cell composition of metastasis can be dominated by, but not restricted to specific clones. Such an observation I reported can be related to reported sequential dissemination of clones into metastatic sites (Naxerova et al., 2017; Quinn et al., 2021). This implies that particular clones can surpass the spatial bottleneck early on, and disseminate to highly expand in permissive microenvironments such as the liver and lung. Further clones can disseminate later on and thus become represented in a smaller cell pool in metastasis.

3.5 Gene expression heterogeneity in tumor progression and metastasis

Integrated data from tumor cells derived from the colon and the liver/lung metastases of xenografts showed that cells distributed across different gene expression phenotypes, resembling normal cell types of the intestinal epithelium (Y. Wang et al., 2020), but also cells of invasive-fetal phenotype characterized with low expression of epithelial markers (Sacchetti et al., 2021). I observed that trajectory analysis showed that the direction of transcriptional change spans the spectrum of differentiated epithelial to the invasive-fetal and mesenchymal cell clusters, with hybrid cell clusters in between (Figure 15). The significance of EMT in tumor progression, treatment resistance and metastasis has been extensively described (Debaugnies et al., 2023; Fleischmajer, 1967; Vanharanta & Massagué, 2013), including hybrid EMT states (Pastushenko et al., 2021; Simeonov et al., 2021). Here, I described these states with marker genes, biological processes and molecular pathways. I observed that transcriptional clusters lying between these two phenotypes showed an elevated expression of genes described in regulating metaplastic transformation of differentiated cells into less differentiated cycling cells (SOX9, DDIT4, YAP1) (Bala et al., 2023; Z. F. Miao et al., 2020; Willet et al., 2018). This transcriptional cell cluster namely metaplastic, which harbors high differential expression of these gene markers, can represent an intriguing phase in the regulation of cell de-differentiation and invasiveness.

These observations I made underline the variable gene expression phenotypes and biological features that can be adopted by CRC cells. These diverse biological features i.e. absorption, secretion, proliferation, stemness and invasion hijacked from cells of normal intestinal cells, could support tumor growth. This is consistent with a previous report suggesting that several functions can be carried out by distinct cells of the tumor (Bloch & Harel, 2016). The existence of different functional cell phenotypes within tumor cells might suggest that simultaneous targeting of these phenotypes can be tested for efficacy in eradicating tumor cells. This might be relevant since sequential targeting of molecular pathways could be possibly surpassed by tumors through plasticity (da Silva-Diz et al., 2018), by which active targeted genes or pathways can be compensated by cells initially expressing these pathways or genes at a low level (Shimokawa et al., 2017). Other approaches of immune cell therapy such as Chimeric Antigen Receptors CAR T-cell therapy (Cappell & Kochenderfer, 2023) can be also hindered by cancer cell plasticity (Dongre et al., 2021).

The relevance of genes (or their upstream regulators) identified in the epithelial (S100 family, ANXA1, LCN2 etc) and the invasive (SOX4, ARGLU1, PI3) phenotype, or genes related to metaplasia (SOX9, YAP1, DDIT4) can be further examined through knock out or overexpression

experiments through *in vitro* approaches such as migration/invasion trans-well assays (Justus, Leffler, Ruiz-Echevarria, & Yang, 2014) or through xeno-transplantation and metastasis assessment of engineered cells over or under expressing these genes. In addition, targeting such identified genes with available inhibitors can be also considered as an alternative approach for their functional validation.

3.6 Subclonal gene expression heterogeneity in tumor and metastasis of CRC

I observed that cells sharing the same barcode (clones) distributed among the regions of transcriptional space of tumor and metastases, and did not cluster uniquely in particular areas, reflecting an unexclusive heterogeneous gene expression on the subclonal level. Such transcriptional heterogeneity which is shared by individual clones, suggests that effective clonal outgrowth is not linked to a unique transcriptional signature that confers a growth advantage to a given clone over another. Indeed, direct comparison of differentially expressed genes in big versus small (clones representing more than 10 % of cells compared to clones of less than 5% of cells), or in monoclonal versus polyclonal cells, within the same compartment (tumor or metastasis) showed an absence of or subtle transcriptional changes of a few genes (including CALB1, S100A4, IFI27), that were not reproducible among different xenografts or patients (Figure 21). This observation argues that individual clones share similar transcriptional features and could exhibit an equipotency in expansion. The observed intra-clonal transcriptional heterogeneity, on the other hand, can confer these clones with plasticity in the challenging tumor microenvironment during tumor progression (Figure 21) (Lenos et al., 2018). The transcriptional similarity of distinct clones in xenografts suggests that clonal selection could be an early stochastic event influenced by the tumor microenvironment, especially given that the tumor initiating cells TICs in two primary mice of the same barcoded PDO were not redundant (Figure 18, 19, 20). This hypothesis can be supported by the observation that the functional heterogeneity of individual clones was previously shown to be independent of genetic mutations (Giessler et al., 2017) in xenografts. In serial transplantations of labeled CRC cells, a subset of previously underrepresented clones denoted as delayed contributing tumor initiating cells DC-TICs, was active in tumor progression (Dieter et al., 2011). This also props the multi-clonal potential of driving tumor progression in particular CRC stages. In combination with the data presented here, such reports might further support a clonal equipotency in driving CRC.

The data presented here showed that intraclonal gene expression heterogeneity along the differentiated epithelial (enterocyte-secretory) to fetal-invasive clusters was maintained in different microenvironments (tumor, liver and lung metastasis), with differences in the number of cells subsisting in specific programs in each site. In liver metastasis, I showed that individual clones showed downregulation of transcriptional programs with canonical epithelial markers (CDH1, KRT20, EPCAM) and an enrichment of transcriptional clusters related to a fetal-invasive phenotype (SOX4, NEAT1, MALAT1, VEGFA, PLCG2), while the same clones reflected the opposite scenario in lung metastasis. Differential gene expression analysis of the same clones in two different sites (tumor vs liver, liver vs lung) further supported this observation (Figure 20, 21). This suggests a clonal intrinsic transcriptional plasticity, which can be implemented in response to different microenvironments or stress. Although this observation is in accordance with previous reports demonstrating a plastic conversion of cancer cells from differentiated to stem cell-like phenotypes in tumor progression and response to different microenvironments (Bala et al., 2023; Fumagalli et al., 2020), data describing intra-clonal gene expression in the tumor or in metastasis were not previously described in this detail.

To dissect the spatial distribution of the barcodes in tumor and metastasis, it would be possible to design fluorescent probes tailored to specifically bind the barcode sequences. This helps understanding whether individual clones form isogenic islands or whether multiple clones co-distribute spatially. Further, the gene expression profiles of individual clones can be examined by combination of spatial transcriptomics with spike-in of barcode probes. This can reveal inter-clonal variations in different spots of tumor expansion and confirm particular transcriptional patterns of individual clones on the spatial level. For example, such experiment can support the identification of the spatial localization of identified transcriptional programs (invasive, secretory, enterocyte) to specific regions such as the tumor edge or center, or their spread and interaction along different tumor regions.

Here, it is worth mentioning that upon xenograft purification and library preparation, a few thousand cells were captured with scRNA, and therefore a partial proportion of the whole tumor cell population was included for analysis but not the entire clonal transcriptional diversity was covered. Accordingly, potential clonal transcriptional events could have been missed. On the other hand, with this resolution used here, I managed to examine the clonal dynamics of thousands of representative cells and their respective gene expression.

Another important aspect not covered in this work is the effect of the immune system on the progression of cancer and its gene expression. Since PDOs are transplanted into NSG immunodeficient mice, cancer cells do not interact with functional immune cells which normally

exert their cytotoxic innate and adaptive immune functions on cancer cells (Hiam-Galvez, Allen, & Spitzer, 2021). Nevertheless, xenografted cells demonstrated the presence of a small transcriptional cluster (2-8%) expressing immune-related genes, located in the vicinity of differentiated cell clusters in the transcriptional space. This cell cluster is characterized with high expression of chemokine and inflammatory associated genes (CXCL1, CCL20) and signalling (TNFa-NFKb) (Figures 13, 14, 26, 27, Supplementary figures 5-7 and 9-11). The expression of such immune-related markers can be associated to survival and progression of cancer cells despite the absence of a functional immune system in NSG mice (Hoesel & Schmid, 2013).

3.7 Clonal dynamics in chemotherapy

I reasoned that if subclones differ in their sensitivity to chemotherapy, it is anticipated that chemotherapy would perturb the clonal composition of a cancer cell population, through eliminating more sensitive clones and marginally impacting resistant ones. In the scenario of clonal equipotency, where each subclone shares similar features with others, chemotherapy is expected to eliminate cells across all individual subclones. When I treated 3 PDOs with chemotherapeutic agents 5-Flurouracil, Oxaliplatin and Irinotecan, the abundance of individual clones of PDOs was not largely affected compared to DMSO controls, and the clonal diversity was only slightly disturbed, an observation reported previously in CRC xenotransplants (Rehman et al., 2021) and patient material (Househam et al., 2022), however not in CRC PDOs. I did not detect major outgrowth of particular clones was detected in direct response to chemotherapy or in recovery, although a slight enrichment of some clones occurred in recovered cells. These results argue toward an equipotency of individual clones in response to chemotherapy, as in xenotransplantation, although long-term effects of chemotherapy on the clonal composition cannot be excluded. This scenario could differ when using targeted therapy, where cells of specific genetic mutations or amplifications are aimed for (A. H. Cheung et al., 2018). Using targeted therapy in combination with the utilized barcoding approach I used in this thesis can be an intriguing aspect, whereby the effect of such drugs on individual clones can be further investigated (Szebényi et al., 2023; Turdo et al., 2022; Zhang et al., 2022).

3.8 Gene expression heterogeneity in chemotherapy treatment response

I observed that PDOs' cells maintained gene expression heterogeneity by subsisting in different transcriptional programs reminiscent of the intestinal epithelium (enterocyte, secretory, stem) under chemotherapy treatment and upon drug washout. The abundance of such transcriptional programs varied under treatment and recovery, demonstrating a decrease in the secretory and proliferative cell phenotypes, and maintenance of enriched enterocyte, stem cell and cycling cell phenotypes. This goes in accordance with a recent report describing the decay of a secretory goblet-like phenotype and the enrichment of a fetal cell phenotype in CRC drug tolerant persister cells DTPs (Álvarez-Varela et al., 2022), but provides additional insights into relevant cell phenotypes and transcriptional programs in chemotherapy response of CRC. The maintenance of these clusters indicates their potential relevance in treatment resistance and proposes a concurrent effect of pre-existing transcriptional programs of different cell phenotypes to support tumor growth after treatment. In individual treatments (5-FU, Oxaliplatin and Irinotecan), both common (enterocyte, cycling, stem) and distinct (enterocyte progenitors in 5-FU treatment) transcriptional programs were enriched or remained abundant. In treatment recovery, the stem cell-like cluster was enriched or remained abundant (Figure 28, Supplementary figure 12). Stem cells are responsible for fast renewal of intestinal epithelium upon damage (Humphries & Wright, 2008; Zeuner et al., 2014), which can explain the enriched abundance of this transcriptional cluster in treatment recovery samples. On the other hand, the decrease of the proliferating cell cluster or its low abundance is an intuitive outcome of using chemotherapy drugs which mainly target cell division activity. Xenografts could offer a better model for studying the effect of treatment on gene expression since they mimic tumor physiological conditions better than *in vitro* cultures. Nevertheless, PDOs have shown a considerable gene expression heterogeneity in this experimental setup (Figure 25), offering a potential alternative to the demanding *in vivo* approaches.

On the clonal level, abundant clones forming more than 5% of the cell population did not show abundant or recurrent transcriptional variations in comparison to small clones forming less than 1% of cells. These comparisons I performed were mainly possible in PDO3, where a less complex starting cell population was implemented. This observation is similar to the same clonal comparison in PDXs, where clones of big size did not reflect major transcriptional variations in comparison to small clones, further supporting this finding. Such results suggests that both, abundant or small clones reflect similarities in gene expression and that clonal expansion of particular clones is independent from their intrinsic transcription. The enrichment of particular clones could be possibly pre-defined in the baseline sample. For instance, clonal outgrowth could

be a stochastic event defined by microenvironment such as cell located at the organoid or tumor edge (Lenos et al., 2018). Another possibility is that at the time of treatment, sub-clonal cell populations could have been previously residing in a treatment response state (enterocyte, stem, cycling) that favors their growth and expansion, in a stochastic fashion. Such cells could be less sensitive to chemotherapeutic killing due to their gene expression profile and therefore could expand further than other clones. For example, MEX3A expressing cells have been reported to protect CRC cells from chemotherapy (Álvarez-Varela et al., 2022). Since subclones, however, harbored cells spanning distinct transcriptional states, their representation would thus not be largely affected by chemotherapy, especially due to the expression of treatment response phenotype (enterocyte, stem, cycling) in a set of cells within subclones. Cells of the same subclones which express genes of the sensitive phenotype (amplifying, secretory) would be either sensitive to treatment and undergo apoptosis, or activate plastic features switching their gene expression profile into treatment resistance related phenotype and resume cell division. This scenario of intra-clonal gene expression heterogeneity might explain the unperturbed clonal representation even with treatment.

3.9 CRISPRi custom screen demonstrates the relevance of selected genes in identified transcriptional clusters of treatment response

Downregulating marker genes of transcriptional phenotypes maintained or enriched in treatment response and recovery, by means of a custom CRISPRi screen, in combination with chemotherapy in two CRC cell models, resulted in a significant reduction of chemotherapy treated cells of these models. This highlights the functional role of such genes in treatment response and endorses the observations depicted by the scRNA analysis. CRISPR activation screen (CRISPRa) can be an additional option to further validate relevance of these selected genes and others. Further functional validation can reveal additional insights about the gene expression response to chemotherapy treatment. For instance, double knock-in experiments can validate the collaborative effect of such genes (ANXA1 and S100; SRSF1 and MYC), through downregulating their co-expression. In addition, targeting such genes with specific inhibitors in combination with chemotherapy can reveal whether their downregulation impacts treatment response in CRC cells. This also suggests testing additional targets in combinatorial treatment strategies in pre-clinical models. Further, microscopy can be an invaluable method for confirming the upregulation of proteins which were identified in our approach in response to treatment, especially marker genes

of the adhesive enterocyte cluster (CDH1), the invasive feature (PI3) or the stem cell (NOTUM, SOX4) phenotype.

Technically, CRC cell lines and 3D sphere cultures were previously tested in our group for CRISPR gene manipulation, and therefore were used in this small size screen. Organoid cultures can also be a relevant model for this approach, that needs further testing. In addition, since targeted cells are assumed to be sensitive to treatment, a small chemotherapy dose was used (IC15 and IC20) so that recovered cells can further expand after treatment with the possibility of their detection through NGS, instead of wiping out the major part of the cells and therefore losing the screen value. This is important taking into account the small size of this screen.

I showed that the predictive clinical relevance of 27 genes identified in the clusters with highest frequency (stem, enterocyte, progenitor) in treatment response, in addition to the invasive metastatic cluster in xenografts, was demonstrated through their correlation with worse prognosis in COADREAD TCGA patient cohort. Many of these genes were previously characterized as potential therapy targets or implemented in functional validation (Jaiswal & Nylandsted, 2015; Mehta et al., 2017; G. Zhao et al., 2023; Zhou et al., 2015). This endorses the relevance of identified transcriptional programs in our data set and suggests them as possible targets for further validation and targeting for cancer therapy in the future. Further exploration of these findings could potentially lead to the identification of novel regulatory genes in colorectal cancer, with implications in therapeutic strategies. Moreover, the identified transcriptional signatures can be further tested for prognostic prediction.

3.10 Open chromatin status of chemotherapy treated cells

I carried out single cell assay for transposase-accessible chromatin sc-ATAC in 5-FU treated and untreated PDOs to link transcriptional changes in gene expression to DNA regulation. Through data analysis, I identified three main cell clusters of open chromatin regions, encompassing untreated and treated cells, with a third cluster harboring cells from both conditions. This latter cell cluster was assigned as the treatment resistant cluster due to its abundance in both conditions, which indicated that cells within this cluster were not perturbed by treatment. Open chromatin regions of genes and transcriptional motifs in resistant and treatment clusters were differentially expressed in treatment response and recovery clusters of scRNA data, in comparison to other clusters. These included MYC and DNMT1 which were enriched in the cycling cell cluster of scRNA data, highlighting the possible relevance of this transcriptional

program in treatment response. Indeed, DNMT1 and MYC are implicated in drug resistance (L. Guo, Lee, Zhou, & Huang, 2022; Si et al., 2016). This highlights the relevance of the analyzed data through identifying similar drug response gene sets in two separate approaches, and links DNA regulation through the chromatin status to gene expression. Such link identifies possible regulatory mechanisms that could be exploited to target drug resistance in CRC cells.

4. Conclusions

This approach in tracking single clones in tumor progression, metastasis and treatment response revealed findings that can be further pursued to better understand and tackle the heterogeneity of colorectal cancer.

The results of this experimental framework suggest that few CRC subclones maintain tumor progression in xenografts while other subclones become underrepresented. In addition, it reveals that single clones are highly heterogeneous in their gene expression, through resembling cellular states of the intestinal epithelium with differentiated cell functions (adhesion, secretion, absorption), but also invasive-fetal transcriptional cell features. These subclonal transcriptional features undergo a selective pressure in different microenvironments. Both transcriptional phenotypes (differentiated and invasive fetal) characterize abundant subclones, with an enhancement of the fetal invasive phenotype in liver metastasis and the differentiated enterocyte and secretory phenotype in lung metastasis. The intermediate transcriptional state I identified between these two main phenotypes namely the metaplastic state can represent a transcriptional state enabling the switch between the differentiated and invasive phenotypes. Additional transcriptional profiles such as the cycling and the stem cell phenotype maintain a high representation through tumor progression. In chemotherapy treatment response, these latter cellular phenotypes in addition to the enterocyte transcriptional program are maintained abundant, while secretory and proliferative cell clusters decline. Combinatorial targeting of these phenotypes can be a possible alternative treatment strategy. Inter-clonal transcriptional differences are scarce, even between highly abundant and small clones both in xenografts and in cultured CRC organoids, where individual clones propagate similarly with or without treatment in the course of the experiment. These observations suggest a characteristic equipotency of single CRC subclones, which can undergo selection through extrinsic microenvironmental opportunities. Targeting transcriptional phenotypes, rather than individual clones, can therefore be more efficient in epigenetic treatment approaches.

Finally, this approach unravels previously underexplored gene expression in functional states of tumor progression into metastasis and under treatment in individual CRC subclones. This highlights the significance of cell tracing approaches in understanding clonal dynamics of cancer progression.

5. Materials and Methods

Human Tissue Sample

Colorectal cancer tissue samples were collected from the surgery department of Heidelberg University hospital. Experiments on human material were carried out in accordance with the declaration of Helsinki and approved by the ethics committee of the medical faculty at the University Heidelberg with ethics vote number 323/2004. Informed consent on tissue collection was received from each patient.

Organoid Culture Generation

Tumor sample purification and cell culture generation was performed as previously described (Giessler et al., 2017; Miura & Suzuki, 2017). Single cell suspensions were obtained by cutting and digesting tumor tissue pieces in 0.08 u/ml Dispase (BD) and 50 µm magnesium chloride, supplemented with 1% antibiotic–antimycotic (Thermo fisher scientific) in a rotating incubator for 1.5 h at 37°C. Cells were washed with PBS (Thermo Fisher scientific) and strained through 100 µM filter (Corning). For generating organoids, a previously described protocol was used with minor modifications (Miura & Suzuki, 2017). Up to 10x10E3 cells were seeded in ice-cold basement membrane extract BME drops (R&D Systems) in one well of 6-well non treated dishes (Corning). Advanced Dulbecco's Modified Eagle Medium DMEM/F12 with additives (10 mm Hepes, 1× B27, 10 mM nicotinamide, 1.25 mM n-acetylcysteine, 1 µm SB202190 , 500 nm A83-01, 10 nm Gastrin, 10 nm PGE2 (Sigma-Aldrich) and 100 µg/ml primocin (Invivogen) was used as CRC organoid culture media. Epidermal growth factor EGF (20 ng/ml R&D Systems,) was freshly added 3 times per week. Rho associated kinase (ROCK) inhibitor Y-27632 (10 µM, STEM CELL technologies) was added when seeding dissociated cells. Tumor organoids were dissociated with TrypLE Express (Gibco) for 10 minutes at 37 °C and replated in six-well plate. To dissociate tumor organoids, cells were collected in TryplE and incubated for 10–20 min at 37 °C and mechanically disrupted by pipetting. The reaction was stopped by adding double volume of PBS. Cells were washed with PBS before reseeding. PDOs were authenticated by Multiplexion (Heidelberg, Germany) as described previously (Castro et al., 2013) and were tested for contaminations by Multiplexion (Schmitt & Pawlita, 2009) and regularly screened for mycoplasma using the Venor GeM Classic kit (Minerva Biolabs).

CloneTracker Barcode Library Lentiviral Production

CloneTracker XP™ (#BCXP10M3RP-P) library plasmids (Cellecta) were packaged using 293T cells (Dull et al., 1998). 1×107 cells per dish were plated on 15 cm tissue culture plates (Corning) in 15 ml Iscove's Modified Dulbecco's Medium IMDM. Transfection was performed 24hrs later, media was exchanged with IMDM containing 10% FBS and 1% L-glutamine. 6 µg of library plasmid was co-transfected with packaging plasmids p101 (12.5 µg), p102 (6.25 µg) and p103 (9 µg) in addition to lipofectamine and PLUS transfection reagent (Thermo Fisher). After 18hrs, media was replaced with DNase I (1 U/ml, Epicentre Illumina), 10% Fetal bovine serum FBS, magnesium chloride MgCl2 (5 mM), and 20mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), pH 7.4 (Sigma-Aldrich) containing fresh media. Virus supernatant was collected 48 hours after transfection and was filtered through 0.22 µm filters (Techno Plastic Products). Viral particles were pelleted through ultracentrifugation 2hrs at 20,000 rpm. Viral pellets were dried, resuspended in DPBS, aliquoted and stored at -80°C. Lentiviral titre was determined by serial dilutions and measurements of red fluorescence protein RFP in Hela cells using flow cytometry (LSR II and FACS Diva).

Generating Barcoded Patient-derived Organoids (BC-PDOs)

CRC patient-derived organoids were dissociated and transduced with the barcode lentivirus with a multiplicity of infection MOI of 0.25 for 16hrs. Cells were harvested and washed twice before seeding in BME drops. Four days later, barcoded cells were sorted via fluorescence-activated cell sorting (FACS) for RFP using a BD FACSAria II (BD Biosciences) and seeded in BME for one week till organoid growth was observed. To obtain replicates of barcoded cells (clones) in baseline population t0 (Figure 6A), BC-PDOs were dissociated and seeded again for one/two weeks. These baseline cells were resuspended, mixed and splitted in 6 parts. Part 1 were collected for scRNA library prep, part 2 were pelleted for DNA seq, part 3 were transplanted into mice, part 4 were seeded for drug treatment, part 5 were seeded for passaging and part 6 were vitally frozen.

BC-PDOs Chemotherapy Treatment

1x10E6 single cells of each baseline BC PDO were seeded into each of four wells in a 6-well plate, with CRC culture media for three days for organoids to form. In each well, culture media

was refreshed with either chemotherapy drugs 5-Fluorouracil 10µM, SN-38 12nM (active form of Irinotecan) and Oxaliplatin 10 µM (Selleck Chemicals) or dimethyl sulfoxide (DMSO; Sigma-Aldrich) 72h post seeding. Drugs were replenished 2-3 days after first dose. 5 days post treatment, PDOs in each of the four conditions were dissociated into single cell suspensions, washed and collected in culture media in 3 parts. Part 1 was used for scRNA library prep, part 2 were pelleted for DNA seq, part 3 were seeded again in drug-free media to let cells recover the treatment for one week. One week after recovery, PDOs were dissociated again and collected in 3 parts. Part 1 was used for scRNA library prep, part 2 were pelleted for DNA seq, part 3 were vitally frozen or re-treated (PDO2 and PDO3).

Laboratory Animals (Mice)

Immunodeficient NOD.Cg-PrkdcscidIl2rgtm1Wjl/SzJ (NSG) mice (The Jackson Laboratory, Bar Harbor, ME, USA) were inbred by the DKFZ Centralized Laboratory Animal Facilities under pathogen-free conditions, according to the German Animal Protection laws and regulations approved by the ethical committee headed by the responsible animal welfare officer. Animal experiments were approved by the Regional Authority of Karlsruhe, Germany under the approval numbers G233/15 and G11-20. Mice were kept in ventilated cages under controlled temperature in groups of five mice with wood chip embedding and nesting material and had free access to water and food. Light/dark cycle was 14 h and 10 h, respectively. Both male and female mice (8-12 week old) were randomly assigned for xeno-transplantation.

BC PDOs Orthotopic Surgical Transplantation

Barcoded (BC)-organoids were dissociated with TryplE (Thermo Fisher #12604013) into cell clumps 3-5 days post seeding. 1x10E6 cells were quenched in 30-50 ul Matrigel (BD Biosceonce) and organoid media in 30-gauge syringe (VWR) on ice. Mice were sedated with 1.75% (vol/vol) isoflurane (Dräger, Lübeck, Germany) on a heating pad and treated subcutaneously with a single dose of 200 mg/kg Metapyrin (Metamizol) (Serumwerk Bernburg). Eye ointment (Vidisic) was used during the surgical procedure. A 1-cm abdominal incision was applied to expose the caecum where BC-PDOs were injected in the caecal submucosa aided by M60 stereomicroscope (Leica). Seprafilm (Genzyme #6380-01) was added on the injection site and absorbable sutures (Resorba) were applied to close the incision. Mice were observed daily post procedure and tumor growth

was monitored by palpation. 8 mice were transplanted at least for the first generation of each barcoded PDO. Tumor growth and metastasis developed in a period of 8-18 weeks. For serial transplantation, tumor or metastatic pieces from each organ (caecum, liver and lung) were chopped, mixed and digested as described above. Up to 5x10E5 freshly dissociated cells, either from tumor, liver or lung metastasis were transplanted into the caecum of secondary mice (Figure 6A). A maximum of 3 mice were serially transplanted whenever a sufficient cell number was obtained from primary mice. Another dissociated cell portion (5x10E5) cells underwent RFP FACS sorting (Supplementary figure 3) and 10x single-cell RNA sequencing, another part (1x10E6 cells) was subjected to barcode bulk DNA sequencing.

Single-cell RNA Library Preparation and Sequencing

Single-cell RNA sequencing (scRNA-seq) was performed following the manufacturer's user guide for the Chromium Single Cell 3' Library (10× Genomics, V2). After tumor and metastasis dissociation, single barcoded cells (RFP+) were sorted (Aria II cell sorter BD Biosciences) at 4ºC to remove mouse cells and debris, pelleted and resuspended in 100 µL cold phosphate-buffered saline with 0.04% bovine serum albumin BSA. Cultured PDOs were dissociated in in TryplE (Thermo Fisher #12604013) for 10 minutes at 37 ºC. Cells were then washed with DPBS and filtered through cell strainers (40 µm - Corning) into cold DPBS - 0.04% BSA to obtain single cells. The concentration of cells was determined on a Luna cell counter (Logos biosystems) and 20-25 x 10E3 cells were loaded into the Chromium chip. Separate Chromium lanes were used for each sample. After cDNA amplification, 10ul of the amplified cDNA was used for the downstream fragmentation, adapter ligation, and sample index PCR. PCR products were purified and size-selected using SPRI magnetic beads (Beckman) and quantified by BioAnalyzer (Agilent). Sequencing libraries from each tissue sample were pooled to yield approximately equal coverage per cell per sample. Libraries were sequenced using an Illumina NovaSeq 6000 S2/4 at an average coverage of 50,000 raw reads per cell (paired-end; read1: 28 cycles; i7 index: 8 cycles; read 2: 98 cycles).

Transduced Barcode Amplification From scRNA cDNA Product

10 ng of the scRNA amplified cDNA was used to PCR amplify molecules that contain both transduced genetic tracking barcodes and the Chromium cellular barcodes.

Target Site-specific primers containing Illumina-compatible adapters (forward: 5′-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCGACCACCGAACGCAACGCACGCA; reverse: 5′- AATGATACGGCGACCACCGAGATCT-ACACTCTTTCCCTACACGACGCTC)

were used in Kapa Taq ReadyMix™ (Sigma) to amplify barcodes, cellular barcodes and UMI. In a second PCR step, Illumina indeces were added for each sample using UDI forward primer

CAAGCAGAAGACGGCATACGAGAT-NNNNNNNN-GTGACTGGAGTTCAGACGTGT GC; in combination with IS reverse primers AATGATACGGCGACCACCGAGATCT-NNNNNNNN-ACACTCTTTCCCTACACGACGCTC;

where “N” denotes sample indices. PCR products were purified, size-selected and quantified as mentioned above. These molecules were then pooled in equal molarity in addition to 10% PhiX genomic DNA library for quality-control. Libraries were sequenced using the Novaseq SP sequencing platform (Illumina) toread and map genetic tracking barcodes to Chromium cellular barcodes, thereby connecting clonal tracking data and gene expression data from the same cell.

Reads of dual indices (8 cycles each) identified sample identities while 10X cell barcode and unique molecular identifier (UMI) sequences were read first (Read1: 28 cycles). The target site sequence was read next (Read2: 94 cycles). Sequencing clusters which passed Illumina QC filters and were processed as described below.

Genomic Barcode Amplification and Sequencing

Barcoded PDOs were harvested and pelleted from indicated samples (baseline, treatment recovery, primary xenografts, serial transplantation) (Figure 6A). Genomic DNA was isolated from 0.5-1x10E6 cells depending on available cell number using DNeasy Blood and Tissue DNA extraction kit (Qiagen) according to manufacturer’s recommendations. Qubit (Life Technologies) was used to quantify isolated DNA. Genomic barcode amplification and indexing was performed in a 2-PCR step using Cellecta NGS Prep Kit for barcode libraries in (pScribe) (LNGS-300) with a maximum 10-50ng of DNA per reaction. The first PCR reaction was performed for amplifying barcode-flanking sequences using specific primers for 14 cycles. The second 10-12 cycle PCR indexed the product of each sample for demultiplexing upon sequencing, whereby primers contained a 6bp index sequence and the Illumina sequencing adapters.

First PCR primers:

Forward: 5'-ACCGAACGCAACGCACGCA

Reverse: 5'-AGCACCAAGCCCAGCCAGCACCAGCA

Second PCR (Nested) primers:

Forward: 5'-TGCGTGCGTTGCGTTCGGTGCTATCTCGTATGCCGTCTTCTGCT

Reverse: 5'-AGATACGGCGACCACCGAGATCTACACGCACGACGAGACGCAGACGAANNNN

NNACGACGACCGACCCGAACCACGA where "N" denotes sample indices

TapeStation 2200 (Agilent) was used to detect ~230-bp PCR amplicon including the 48- bp barcode sequence. After purification with SPRIselect beads (Beckman Coulter), NGS libraries were mixed and diluted according to equimolarity for sequencing. NGS was performed using NextSeq550 platform (Ilumina) (single read 75 bp, covering sample index and 48-bp barcode sequence).

scATAC Library Preparation of Treated PDOs

Two PDO cultures were treated with either 5-FU or DMSO control for five days, and dissociated as mentioned above. Single cell suspension was treated with DNase (Epicentre-Illumina) and lysed in 1 mL prepared lysis buffer (10 nM Tris-HCl, 10 mM NaCl, 3 mM MgCl2, 0.1% Tween-20, 0.01% Digitonin, 1% BSA, 0.1% NonidetTM P40 Substitute (Sigma-Aldrich)) and incubated on ice for 3 min. The culture was mixed with 1 mL washing buffer (10 nM Tris-HCl, 10 mM NaCl, 3 mM MgCl2, 0.1% Tween-20, 1% BSA) and centrifuged at 400 rcf for 7 min at 4°C. The homogenate was suspended in nuclei buffer and filtered through a 40-µm cell strainer (Corning) and the nuclear concentration was determined using LUNA Automated Cell Counter (Logos biosystems). 10,000 nuclei were loaded on the chips of Chromium Next GEM Single Cell ATAC Library & Gel Bead Kit v.1.1 (10x Genomics, 1000175). Samples were sequenced on Illumina NovaSeq 6000 using paired-end 150 bp protocol.

Immunofluorescence Staining

Staining with S100A10 and VEGF antibodies was performed on FFPE embedded xenograft material for both tumor and metastasis (liver and lung). The primary antibodies (S100A4 1:100

AB196380 rabbit conjugated to Alexa Fluor® 488 Abcam, and VEGF MA5-13182 1:50, respectively) were incubated overnight at 4°C. Incubation with secondary Alexa Fluor plus 555-conjugated goat anti-rabbit IgG antibody (1:1000, Life Technologies) was carried out for 1h at room temperature. DAPI (1:1000, ThermoFisher Scientific, Gent, Belgium) was used for nuclear counterstaining. Cell observer fluorescence microscope (Leica) was used for detection of immunofluorescence and XEN software was used for image analysis.

Immunohistochemistry Staining

Formalin-fixed and paraffin-embedded pieces of xenograft material (primary tumor in the colon, liver and lung metastases) were processed at the Institute of Pathology, Heidelberg University. Tissue sections were cut and treated with an antigen retrieval buffer. Staining for CEA was performed using an automatic staining device (Ventana Benchmark Ultra, Roche, Basel, Switzerland) as described before (Kriegsmann et al., 2019).

CRISPRi screen and analysis

Cell Culture

Three-dimensional patient-derived (Female, CRC liver metastasis, TP53, APC and KRAS mutant; MSS) spheroid cells SC1 were cultured as previously described (Zowada et al., 2021). Briefly, SC1 cells were maintained as suspension cells in ultra-low attachment plates in Advanced DMEM/F-12 medium supplemented with 2 mM L-glutamine, 5 mM HEPES, 0.6% glucose, 4 µg/mL heparin and 4 mg/mL BSA. Growth factors EGF (20 ng/mL) and FGF (10 ng/mL) were added to the medium twice per week. HT29 cells (German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany) were cultured in McCoy's 5A medium (Thermo Fisher Scientific, Waltham, MA, USA) supplemented with 10% FBS and 1% penicillin/streptomycin. HEK293T cells (ATCC) were cultured in DMEM supplemented with 10% FBS. All cell models were routinely tested for mycoplasma contamination and authenticated by Multiplex Cell Authentication (Multiplexion).

Drug Sensitivity Assay

For generation of dose response curves, ATPlite 1step Luminescence Assay System (Perkin Elmer) was used according to manufacturer`s instructions. Briefly, 1x104 cells (SC1dCas9-KRAB)

and 5x103 cells (HT29dCas9-KRAB) were seeded into a 96-well plate. After 24-hour recovery phase, cells were treated with 1:1 diluted drugs (5-fluorouracil and Oxaliplatin) in 11 descending concentrations. DMSO served as a negative control, whereas blank wells served as a positive control. Following 72-hour incubation, ATP content as a surrogate for cell viability was determined by measuring luminescence using Infinite 200 Pro plate reader (Tecan). The IC50 values were determined using GraphPad Prism version 8.1.2.

Custom sgRNA Library Synthesis

Target selection criteria

A custom sgRNA library was designed to target 181 genes of interest with 5 independent sgRNAs per gene using the CRISPick algorithm (Doench et al., 2016). Additionally, 97 non-targeting sgRNAs as negative (C. Liu, Zhang, Liu, & Cheng, 2017), and 18 sgRNAs targeting core essential genes RPS27 and MYC as positive controls were included, accounting for a total of 1018 unique sgRNAs. The library was cloned in pCRISPRia-v2 sgRNA lentiviral expression vector (Gift from Jonathan Weissman, Addgene plasmid # 84832 ; http://n2t.net/addgene:84832 ; RRID:Addgene_84832) as previously described (Gilbert et al., 2014; Horlbeck et al., 2016). Briefly, sgRNA library was PCR amplified from a customized oligonucleotide pool (Twist Bioscience). The PCR products and pCRISPRia-v2 vector were digested with BstXI and BlpI, gel purified and ligated in 1:1 ratio. Subsequently, custom sgRNA library was electroporated in MegaX DH10B T1R Electrocomp™ cells and directly transferred in liquid overnight culture. Custom sgRNA library was later purified by ZymoPURE II Plasmid Maxiprep Kit (Zymo Research) and equal representation of individual sgRNAs was verified by amplicon NGS, confirming that all designed sgRNAs were cloned in the pCRISPRia-v2 vector.

CRISPRi Screening

SC1dCas9-KRAB and HT29dCas9-KRAB cells constitutively expressing dCas9-KRAB were produced by lentiviral transduction of parental SC1 spheroids and HT29 cell line with pLV hUbC-dCas9 KRAB-T2A-GFP (Gift from Charles Gersbach, Addgene plasmid # 67620; http://n2t.net/addgene:67620; RRID:Addgene_67620) followed by sorting to achieve >95% GFP purity. Expression of dCas9-KRAB was verified by Western blot and dCas9-KRAB activity was confirmed by transcriptional repression of RPS27 gene (pCRISPRia-v2,

GCCTGGACCAAAGCACTCAC). For CRISPRi negative selection screening, SC1dCas9-KRAB and HT29dCas9-KRAB cell models were transduced with custom sgRNA lentiviral library at multiplicity of infection of 0.3 to ensure a single sgRNA integration per cell. Transduced cells were selected by 1 µg/mL puromycin for 48 hours. After 5 days, cells were split in four parts: one part was harvested as a baseline sample (t0), whereas the rest of the cells were used for screening by either vehicle (DMSO), oxaliplatin (OXA) or 5-fluorouracil (5-FU). For OXA screening, cells were cultured in the medium containing a concentration that reflects the IC15 for each line (135 nM for SC1dCas9-KRAB, 850 nM for HT29dCas9-KRAB). For 5-FU screening, cells were cultured in the medium containing a concentration that reflects the IC20 for each line (1600 nM for SC1dCas9-KRAB, 1000 nM for HT29dCas9-KRAB). Cells were incubated in aforementioned conditions for 48 hours, washed and cultured for another 19 days to ensure at least 10 population doublings. Cells from all three arms (DMSO control, OXA treated, 5-FU treated) were harvested at a final timepoint (t21). Genomic DNA from t0 and t21 was extracted (Qiagen DNeasy Blood and Tissue kit). The sgRNA cassettes were PCR amplified using an adapted protocol from a previous study (Yau and Rana, 2018) to ensure appropriate amplicon heterogeneity. Amplicons were purified and subjected to sequencing using a the NextSeq 550 platform (Illumina) together with a 20% PhiX DNA spike-in. Sequencing data were preprocessed using in-house script. sgRNA abundance and hit calling was analyzed using MAGeCK v0.5.9.2 algorithm as previously published (Li et al., 2014). Gene abundance with logscore >2 were considered statistically significant. All the steps in the CRISPRi screen were conducted ensuring a minimum of 1000x sgRNA coverage. CRISPRi screens were performed in biological triplicates and all replicates were included in the final analysis.

Transcriptomics bioinformatics analysis

Alignment and quality check QC

Single cell transcriptome sequencing (scRNA) data was aligned and processed using 10x Cell Ranger Count v3.1.0 with the hg-GRCh38 reference genome. Seurat v4.1 pipeline along with R 4.1.0 was used to analyze the data with default parameters based on the Seurat's manuals, if not mentioned otherwise. Low quality cells (≤ 1000 genes or ≥ 10 percent of mitochondrial gene fraction) were eliminated. In case of the xenograft samples, an additional run of the Cell Ranger Count was performed with a concatenated hg-GRCh38 and hg-GRChm38 reference genome to

determine the human mapped read percentage. Cells below 75% were considered as mouse cells and were subsequently eliminated from the analysis.

SCTransform v2 was used for normalization and variance stabilization of scRNA-seq data. Cell cycle scores and phases were determined by the CellCycleScoring function (with nbin=22). Furthermore, SCTransform v2 was used to regress out the cell cycle scores. DoubletFinder was used to determine cell doublets (two different cells sharing the same 10x barcode). Detected doublets were eliminated. Remaining cell clusters with less than 25 cells were eliminated.

Data integration and analysis

Samples were integrated separately in six groups corresponding to each patient (PDO1, PDO2, PDO3) and experimental setup (PDOs *in vitro*, xenotransplants) using the IntegrateData function including all genes. Clustering was performed with a resolution of 0.5. Differential gene expression analyses were performed between each cluster and the rest of the cells (Bonferroni multiple error correction was performed additionally for the clusters, adjusted p value ≤ 0.05 was considered significant). Gene set enrichment analyses were performed on the differentially expressed genes with fgsea algorithm (msigdbr 7.4.1 geneset collections: H, oncogenic, TFT, GO-BP, GO-MF, MIR, CGN, GO-CC). Gene sets bigger than 500 genes were excluded and additional manually curated gene sets were included. Benjamini Hochberg multiple error correction was performed for gene sets and clusters and the adjusted p-value/FDR ≤ 0.05 was considered significant. For comparisons of the abundance of selected clusters (enterocytes, secretory, invasive) two-way ANOVA and post hoc tests were used for determining the statistical significance.

Trajectory analysis

Velocyto version 0.17.17 was run on each individual sample with the hg-GRCh38 reference genome to carry out the trajectory analysis. Integrated Seurat objects were converted to Anndata with SeuratDisk package. Trajectory analyses were performed with CellRank 1.5.1 pipeline based on the usage guide. Integrated Seurat and Velocyto data were merged. Data were processed with scVelo. Cell rank was used to identify terminal and initial states and compute fate maps. Directed PAGA was used to aggregate the individual fate maps into a cluster-level fate map. Lineage drivers were determined and gene expression trends were computed using GAM model with Gaussian distribution and identity link.

Transduced barcode analysis

Transduced barcode reads were aligned to the reference provided by company with BWA-MEM 0.7.17 and processed using python 3.7.0. The transduced barcodes were determined with accepting 1 mismatch followed by extracting and merging cell barcodes and UMIs. R 4.1.0 was used for further processing of the barcodes to eliminate low quality reads (min. quality: 10, no N, no homopolymer UMI, PCR duplicates carrying the same UMI for the transduced and cell barcode). Barcode copies were counted based on the different UMIs and non-valid cells (Seurat pipeline) were eliminated. The most frequent barcode in a cell was considered. Remaining cells containing multiple barcodes (Figure 6E) with similar read numbers were eliminated from the clonal analyses.

Clonality analysis

Clonality score was calculated by dividing the detected clone number with the cell numbers. Downsampling i.e. subgrouping samples to groups equal in size and calculating their mean was carried out to overcome the challenge of the different cell numbers in different samples. Groups of 87 cells were randomly selected from each sample and clonality scores were determined per group. Clonality scores of cell groups were averaged and SEM calculated for each sample. ANOVA and TukeyHSD posthoc test was used to compare the clonality scores (adjusted p value ≤ 0.05 was considered significant).

Clone composition

To assess clonal composition, most abundant clones were displayed for each group and relative abundance (%) of each clone was determined. Top unique clones were shown in different colors and remaining smaller clones were collapsed in grey color.

Clone differential gene expression analyses

To determine the differential gene expression between tumor sites or treatments, intersample comparisons which consider significant gene changes with a minimum 1 log2 fold change were displayed. The difference between major and small clones was determined only within each sample to avoid batch effects.

Differential gene expression analyses were performed between each of the most abundant clones (8 *in vitro* and 12 *in vivo*) considering the other clones (small) group as a separate clone. A clone was considered for analysis if it contained at least 30 cells. Seurat's method was used in addition to Bonferroni multiple error correction for the comparisons, where adjusted p value ≤ 0.05 was considered significant.

Bulk transduced barcode analyses

Transduced barcode reads were aligned to the reference provided by company using BWA-MEM 0.7.17. The alignment results were processed with python 3.7.0. and the transduced barcodes were determined (1 mismatch was accepted). Further, R 4.1.0 was used to process the barcodes. Low quality reads (min. quality: 10, no N, no homopolymer (UMI)) were eliminated and barcode copies were counted (based on transduced barcodes). Clone composition was shown as described earlier.

scATACseq Analyses

Single cell ATAC sequencing data was aligned and processed using 10x Cell Ranger ATAC Count v2.1.0 with the hg-GRCh38 reference genome. ArchR v1.0.2 pipeline (along with R 4.1.0) was used to analyze the data with the recommended parameters of the manual, if not mentioned otherwise. Arrow files were created, doublet score were determined and doublets were filtered out. Iterative Latent Semantic Indexing was used for dimension reduction, followed by clustering. UMAP was created and gene scores were calculated. Additional grouping was defined based on the relative cell abundance of clusters, which were categorized as sensitive group if the treatment/control ratio was > 10, response if the treatment/control ratio < 0.1 or otherwise resistant. Marker features were extracted.

Pseudo-bulk replicates were created and peaks were identified with MACS2. Motif (cisbp), EncodeTFBS, ATAC and Codex peak annotation was used. ChromVAR deviaton enrichment was used to retrieve information on single cell level. Motif Footprinting was performed and co-accessibility was obtained. Positive TF-Regulators were identified between gene scores and motifs.

Statistical analysis

Statistical analysis was performed in R 4.1.0 using testing methods indicated in different experiments or otherwise indicated. Data is displayed as mean ± SEM. Experiments were performed with three biological replicates unless noted otherwise. Statistical significance was defined as *, $p \leq 0.05$.

6. Supplementary material

6.1 Tables

Culture Name	Origin	Tumor type	MS Status	Patient sex	KRAS status	Treatment
OC2	Patient tissue	Colon ascendens	MSI	Male	N/A	FOLFOX+
OC3	Xenograft	Liver metastasis	MSS	Female	p.G12V	FOLFOX+
OC7 (PDO1)	Patient tissue	Rectum	MSS	Male	p.G12D	None
OC13 (PDO3)	Xenograft	Colon	MSS	Male	p.G12D	None
OC15	Xenograft	Rectum	MSS	Male	WT	FOLFOX+
OC17 (PDO2)	Patient tissue	Colon	MSS	Female	WT	None
OC19	Patient tissue	Colon	MSS	Male	WT	FOLFOX+
OC20	Xenograft	Liver metastasis	MSS	Female	WT	FOLFOX+

Supplementary table 1. Organoid cultures background. Table showing the tumor type, microsatellite status (MSS: stable, MSI: instable) and KRAS gene mutations (WT: wild type). Patient sex, previous treatment and organoid generation material (origin) is shown. FOLFOX: chemotherapy combination containing the drugs leucovorin calcium (folinic acid), fluorouracil, and oxaliplatin. + stands for Bevacizumab or Avastin treatment in addition to FOLFOX.

	PDO1	PDO2	PDO3
Baseline	260 x 10E3	18 x 10E3	4 x 10E3
Tumor 1	6.6 x 10E3	5.1 x 10E3	2 x 10E3
Tumor 2	2.8 x 10E3	8.1 x 10E3	2.3 x10E3
Liver Met. 1	1.5 x 10E3	2 x 10E3	N/A
Liver Met. 2	0.87x 10E3	1.2 x 10E3	4 x 10E3

Supplementary table 2. Total clone number in PDOs and xenografts. Table exhibiting the total number of unique barcodes detected through barcode amplicon sequencing. Data corresponding to baseline cells in addition to two primary mice tumors and liver metastases of each PDO is shown. N/A: not available.

Clone ID	Tumor 1	Liver Met.	Lung Met.	Tumor serial	Liver Met.	Lung Met.	Liver serial	Liver Met.	Lung Met.	Lung serial	Liver Met.	Lung Met.
095-069845	246	113	270	//	//	//	//	//	//	1880	4863	587
040-060666	486	61	//	//	//	//	//	//	//	//	//	//
061-074113	10	139	//	11	6	190	//	//	//	//	//	//
006-046000	//	754	//	666	598	1	//	//	//	//	//	3
037-006036	//	358	//	136	274	//	//	//	//	//	//	//
070-064630	//	423	//	272	629	3	//	//	//	//	//	//
095-014328	//	//	//	//	//	147	//	//	//	//	//	//
032-057163	//	//	//	//	//	53	634	1625	545	//	//	//
055-081422	//	//	//	//	//	//	172	1495	803	//	//	//
076−082004	//	//	//	//	//	//	//	//	//	//	//	39
090−004937	//	//	63	//	//	//	//	//	//	//	//	//
Others	926	565	8	89	88	46	33	58	113	59	55	2
Clone number	55	27	5	13	12	11	5	8	68	3	4	2
Cell number	1668	2413	341	1174	1568	440	839	3178	1461	1939	4918	628

Supplementary table 3. Clone cell number in barcoded xenografts captured by scRNA seq. Table exhibiting the number of unique barcodes detected through barcode amplification from scRNA libraries. Data corresponding to primary xenograft cancer sites (tumor, liver and lung) and their serial transplantations into the caecum of secondary mice, in addition to metastases in secondary mice are displayed. Clone identification number (ID) and its corresponding cell number is indicated in the first column. A representative summary of data from PDO1 is shown. Liver serial: tumor formed in the caecum of serially transplanted secondary mice, by liver metastasis cells of primary xenograft. Others: clones that have less than 3 cells in all samples. //: below detection limit.

Clone ID	Baseline	DMSO	5-FU	OXA	IRI	DMSO rec	IRI rec
002-065605	376	169	156	406	146	559	368
003-067634	1917	1889	781	1248	1158	3028	46
004-011230	2250	1386	800	1547	1022	2338	148
022-051677	500	260	345	333	613	176	31
033-034589	439	161	175	438	162	604	314
050-078683	2303	1791	770	1629	979	2968	121
052-045772	822	254	408	246	565	217	33
058-074148	1570	1469	816	1462	886	2378	115
069-036038	3284	2218	1226	2357	1715	4427	219
097-057464	2279	1572	753	1677	998	2728	131
other clones	29432654	12835516	8078699	15256802	13995425	15372139	12818236
Clone number	29448394	12846685	8084929	15268145	14003669	15391562	12819762

Supplementary table 4. Clone read number in barcoded PDO captured by DNA seq. Table displaying the number of reads for the top 8 individual barcodes detected through DNA barcode amplification. Data corresponding to 1 x 10E6 cells of baseline sample, DMSO and chemotherapy treated PDOs is displayed, in addition to cells recovered from Irinotecan treatment (Oxaliplatin and 5-fluorouracil recovery samples are not available). A representative summary of data from PDO1 is shown. Clone identification number (ID) and its corresponding read number is indicated in the first column. Others: clones that have less than 3 cells in all samples.

Gene	PDO1 log2FC	PDO2 log2FC	PDO3 log2FC	significance
SOX4	1.031	0.3	0.8	*
NOTUM	1.222	0.87	NA	*
SOX9	0.49	0.48	0.6	*
PROX1	0.98	0.4	0.6	*
MALAT1	3	0.6	2.2	*
NEAT1	3	0.7	2.47	*
ARGLU1	2.14	NA	0.9	*
FABP1	2.3	3.3	NA	*
LGALS4	1.17	0.9	0.9	*
TFF1	1	1.8	1.4	*
SLPI	1.2	1.5	0.7	*
ANXA1	1.06	0.6	1.2	*
S100A	1.24	1	1.4	*
SLC3A2	1.6	1.3	0.95	*
SRSF1	0.37	0.26	0.35	*
KRT18	0.87	0.87	1.7	*
KRT20	0.87	0.87	1.7	*
PDIA4	1.94	1.76	0.98	*
DDIT4	1.5	1.74	2	*
MANF	1.5	1.3	0.67	*
PI3	0.9	0.9	0.9	*
ATOX	0.9	2.4	2.4	*
GSTP1	0.9	0.9	0.6	*
GPX1	0.9	2.4	3	*
ROMO	0.4	0.5	NA	*
HMGB1	0.8	0.3	0.2	*
MYC	0.15	0.17	0.34	*
LGALS4	0.7	0.9	0.24	*
SLPI	1.1	1.1	0.7	*

SLC52A	0.36	0.25	0.24	*
BCL2	0.3	0.5	0.9	*
RAN/BP1	1	2.8	2.3	*
DDX5	1.7	2.7	2	*
TPT1	0.7	4.7	4.5	*
NME2	4.5	2.3	NA	*
SAT1	0.5	0.6	1.3	*
SF1	0.9	1.8	2	*
PNN	1.74	0.5	0.8	*
GOLGB1	1.24	0.8	0.9	*
HSPA5	2.93	2.4	1.5	*
ADIRF	0.9	0.5	0.9	*
FXYD3	0.88	0.99	0.84	*
DNAJB/C	1.5	1.4	0.7	*
ZFAS1	1.5	1.3	1	*
VIMP	1.4	1	NA	*
TXN	1.34	0.8	0.9	*
ARF	1.25	0.13	NA	*
EIF1	1.15	2.5	2	*
TAF1D	1.3	1.4	0.5	*
ATF4	1	0.9	0.8	*
NDRG1	0.9	1	1.7	*
P4HB	0.9	1	0.7	*
ERO1	0.9	0.7	1	*
TRIM	1.55	0.5	0.7	*
ATP5D/O	0.9	0.8	0.9	*
UQCR	0.8	0.9	NA	*
COX5	0.8	0.8	0.8	*
SNRPD2	0.8	0.9	0.9	*
POMP	0.6	0.75	NA	*
PHLDA1	0.6	0.6	2.4	*

PRAP	1.8	0.9	NA	*
EZR	0.6	0.2	1	*
APEX1	2.85	0.8	0.6	*
FOS	0.9	1.9	NA	*
CDK2	1.2	1	2	*
CD46	0.5	1	0.22	*
KCNQ1OT1	2.2	0.58	1.24	*
TPI1	0.37	0.48	0.57	*
TUBA1B	1.1	1.01	0.8	*
GAPDH	0.6	0.31	0.78	*

Supplementary table 5. Selected list of genes from CRISPRi customized screen. A table showing a list of genes selected for CRISPRi custom screen in two CRC cell models. 70 gene are depicted. The level of relative expression of these genes is indicated by the Log2 fold change (FC) in integrated scRNA data of each culture. The significance is indicated by an asterisk representing an adjusted p-value <0.05. NA: not available.

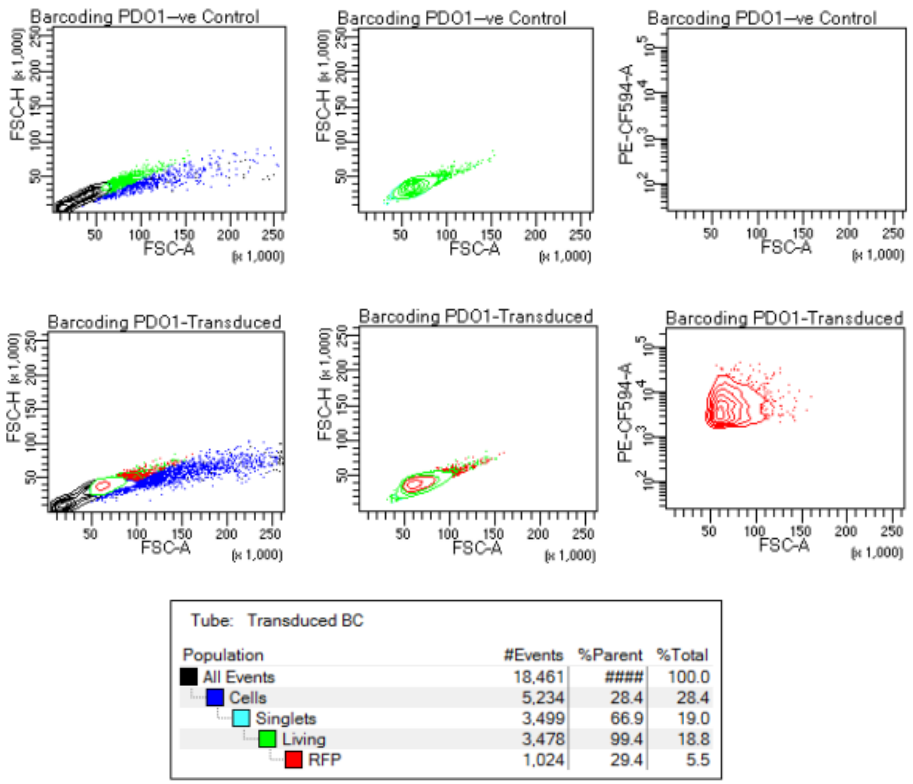
Gene	Survival	Prognosis	p-value	Groups	Cluster
NEAT1	Disease free-interval	Worse	0.0014	2 groups	Stem
SOX4	Disease free-interval	Worse	0.068	2 groups	Stem
NOTUM	Disease free-interval	Worse	0.008	2 groups	Stem
SOX9	Overall	Worse	0.05	quartiles	Stem
PROX1			NS		Stem
LGR5			NS		Stem
SMOC2			NS		Stem
SEMA3A	Disease specific-survival	Worse	0.014	2 groups	Stem
CCNB1	Overall	Worse	0.009	quartiles	Cycling
PTTG1	Overall				Cycling
HMGB1	Overall		NS		Cycling
FABP1	Overall		NS		Enterocyte
PRAP1	Disease free-interval	Worse	0.05	2 groups	Enterocyte
S1001	Overall	Worse	0.001	2 groups	Enterocyte
GSTP1	Overall		NS		Progenitor
SLPI	Overall		NS		Progenitor
GPX1	Disease specific-survival	Better	0.02	3 groups	Progenitor
ATOX1	Overall		NS		Progenitor
PI3	Overall		NS		Progenitor
ANXA1	Overall		NS		Progenitor
MALAT1	Overall	Worse	0.047	quartiles	Invasive
NEAT1	Disease free-interval	Worse	0.001	2 groups	Invasive
ARGLU1	Disease free-interval	Worse	0.004	2 groups	Invasive
CTNNB1	Overall	Worse	0.032	2 groups	Invasive
VEGFA	Disease free-interval	Worse	0.0001	2 groups	Invasive
DDX17	Disease free-interval	Worse	0.011	2 groups	Invasive
PLCG2	Disease specific-survival	Worse	0.009	2 groups	Invasive

Supplementary table 6. Prognostic value of marker genes of treatment response transcriptional clusters. Table showing the prognostic value of genes corresponding to transcriptional clusters in treatment response in COAD-READ patient cohort. Statistical significance is indicated by p-value. 2-groups: samples divided on the median. Quartiles: upper vs lower quartile of the samples. NS: not significant.

Cluster	Response (scATAC)	Resistance (scATAC)	Untreated (scATAC)
Stem (scRNA)	PRAP1, SOX9	NOTUM, SMOC2, PRAP1	
Undifferentiated cycling (scRNA)	PTTG1, CCNB1, CTNNB1		PTTG1, CCNB1, CTNNB1
Enterocyte (scRNA)	PRAP1, ANXA1, LCN2, GPX1, S100A11, CDH1, FABP1	PRAP1	GPX1, S100A11, CDH1, FABP1
Enterocyte progenitor (scRNA)	GPX4, ANXA1, SLPI, PI3, LCN2	GPX4	
Invasive (scRNA)	SRSF1, MALAT1, NEAT, VEGFA	SRSF1	MALAT1, NEAT, VEGFA, TAF1

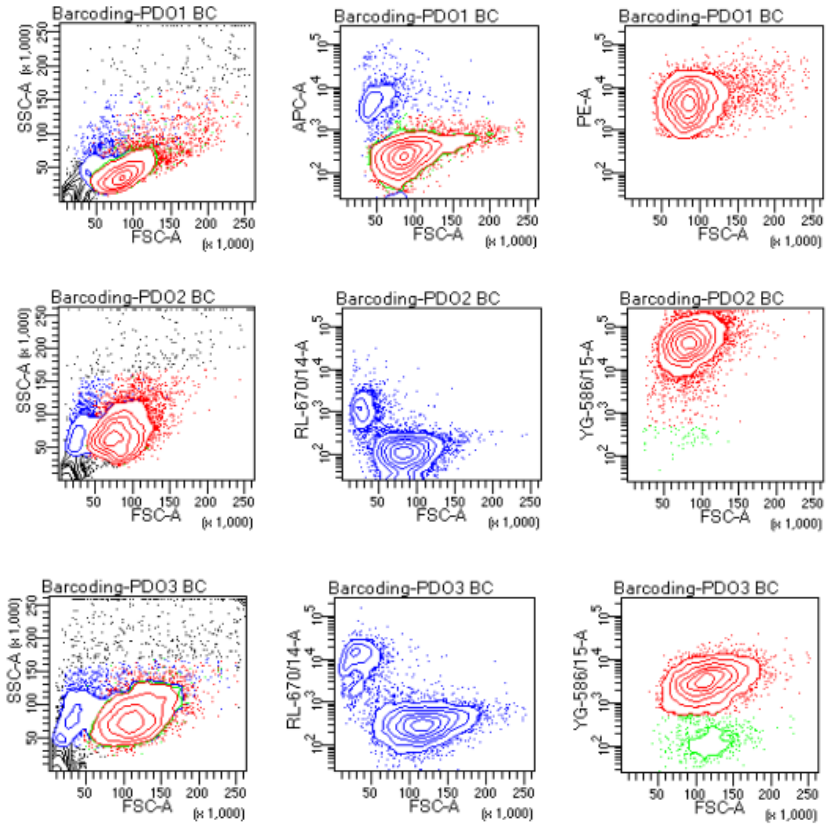
Supplementary table 7. Open chromatin status of marker genes of transcriptional clusters. List of genes corresponding to transcriptional clusters in treatment response (vertical left panel) and their cluster allocation in scATAC treatment experiment.

6.2 Figures



Supplementary figure 1. Sorting RFP+ cells in a representative PDO. Untransduced cells were used as a negative control for RFP cell sorting. RFP+ single cells represent 29.4% of living cells. Dead cells were excluded using toto-3 cell membrane staining. Single dots and contour represent cell populations.

A



B

Tube: PDO1 BC

Population	#Events	%Parent	%Total
All Events	10,000	####	100.0
Cells	6,247	62.5	62.5
Living	5,257	84.2	52.6
RFP	4,586	87.2	45.9

Tube: PDO2 BC

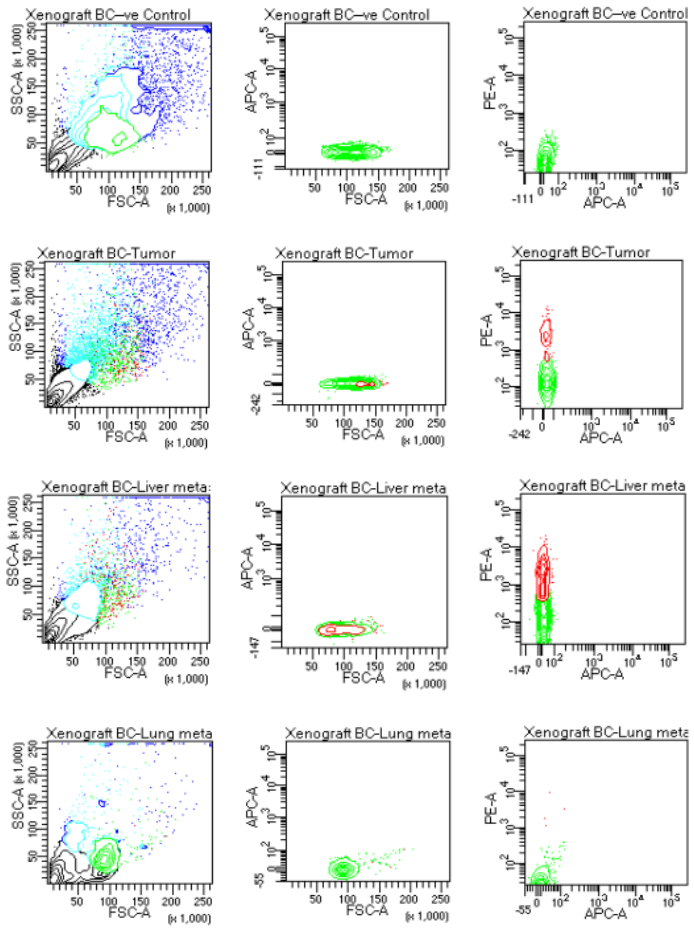
Population	#Events	%Parent	%Total
All Events	10,000	####	100.0
Cells	7,143	71.4	71.4
Living	6,004	84.1	60.0
RFP	5,947	99.1	59.5

Tube: PDO3 BC

Population	#Events	%Parent	%Total
All Events	10,000	####	100.0
Cells	5,925	59.2	59.2
Living	4,492	75.8	44.9
RFP	3,971	88.4	39.7

Supplementary figure 2. RFP+ cells of barcoded PDOs 8 weeks after transduction. A. Plots of RFP+ cell quantification by flow cytometry. Untransduced cells were used as a negative control for RFP living cells. RFP+ cells are measured living cells. Dead cells were excluded using toto-3 cell membrane staining. Single dots and contour depict cell populations. B. Percentage of RFP+ cells in FACS measurement events.

A



B

Tube: -ve Control

Population	#Events	%Parent	%Total
All Events	10,000	####	100.0
Cells	6,311	63.1	63.1
Singlets	4,220	66.9	42.2
Living	1,070	25.4	10.7
RFP	0	0.0	0.0

Tube: Tumor

Population	#Events	%Parent	%Total
All Events	30,000	####	100.0
Cells	5,498	18.3	18.3
Singlets	3,768	68.5	12.6
Living	610	16.2	2.0
RFP	117	19.2	0.4

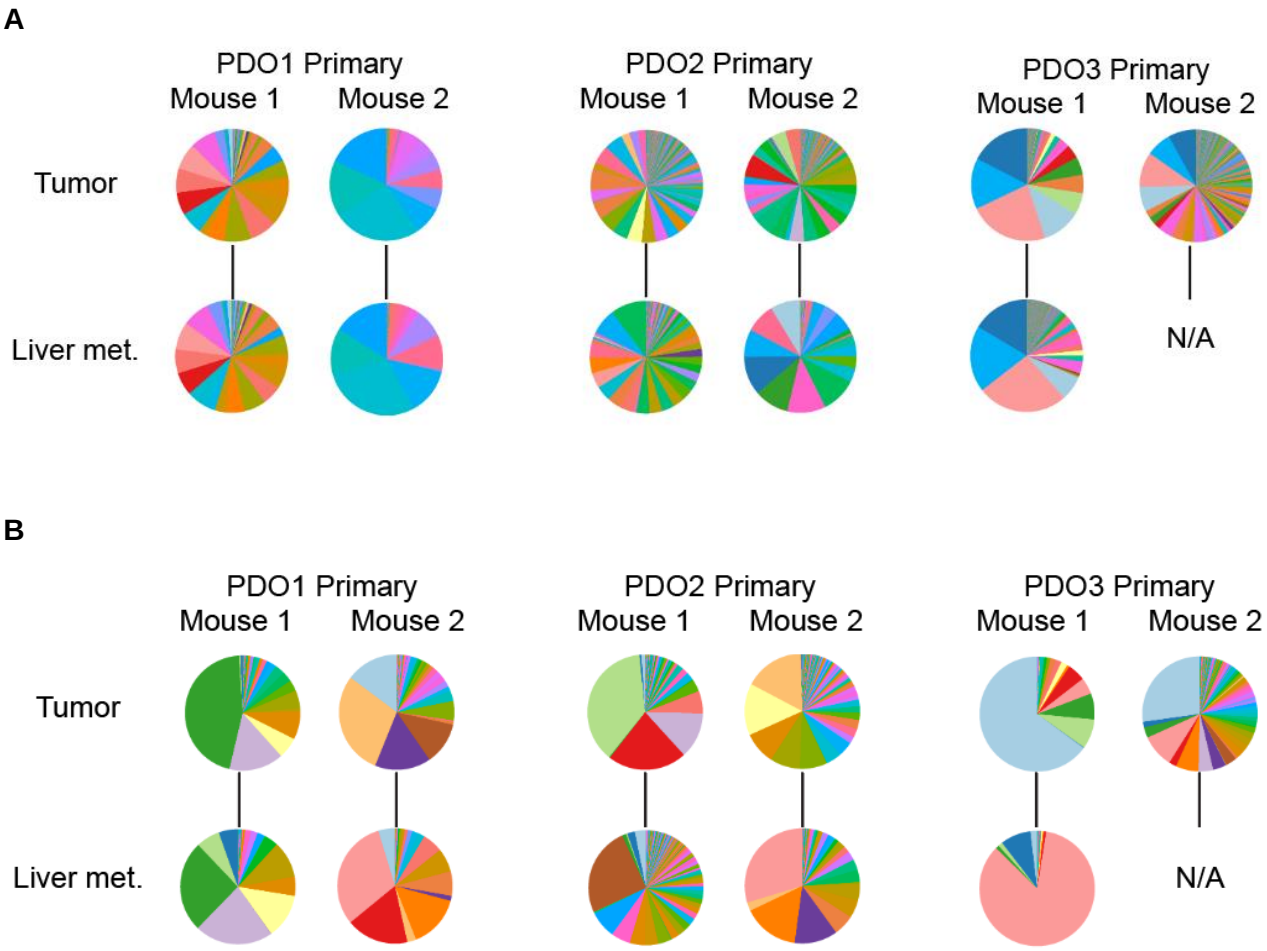
Tube: Liver metastasis

Population	#Events	%Parent	%Total
All Events	10,000	####	100.0
Cells	2,493	24.9	24.9
Singlets	1,801	72.2	18.0
Living	648	36.0	6.5
RFP	250	38.6	2.5

Tube: Lung metastasis

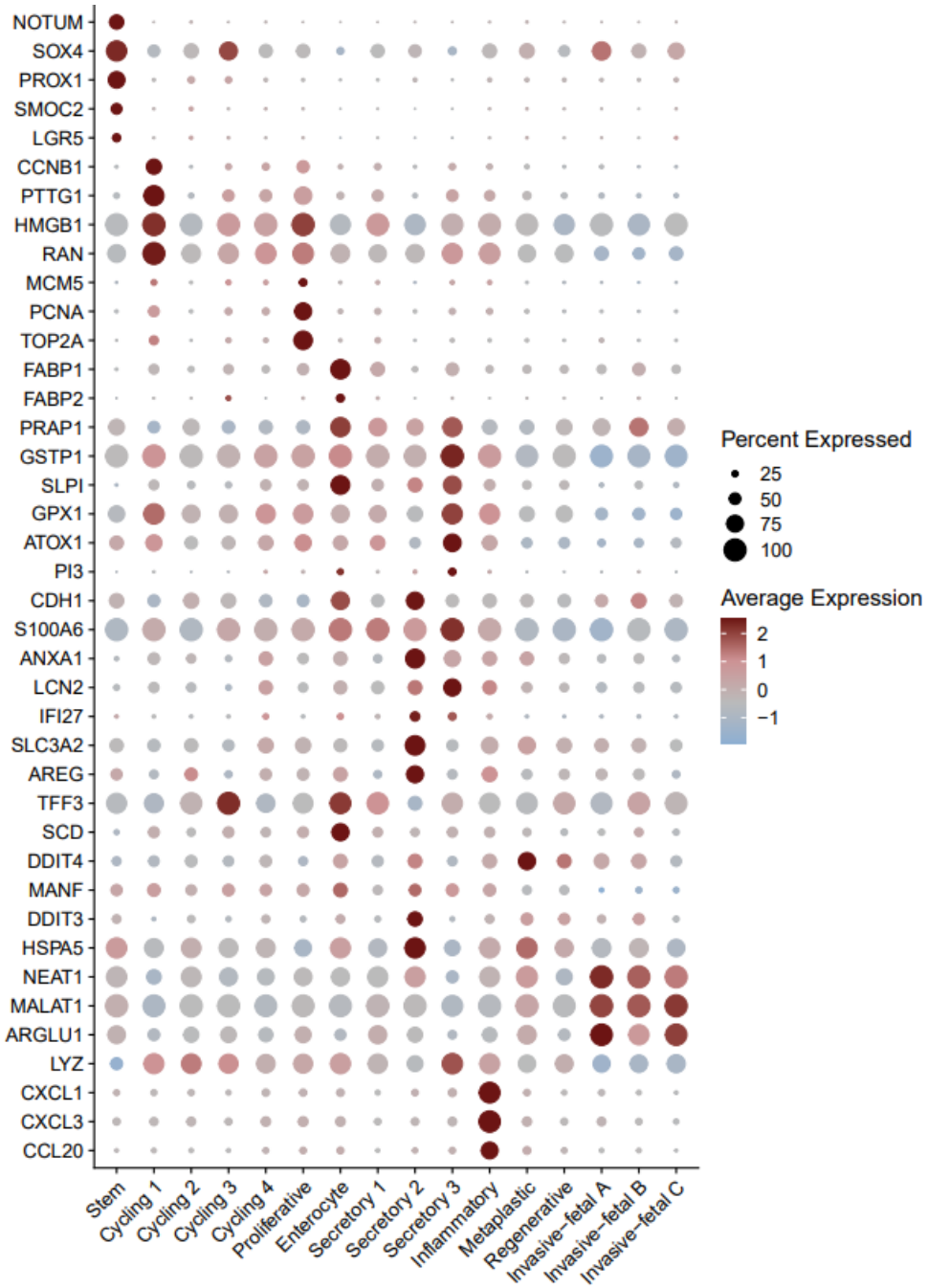
Population	#Events	%Parent	%Total
All Events	1,992	####	100.0
Cells	872	43.8	43.8
Singlets	749	85.9	37.6
Living	513	68.5	25.8
RFP	4	0.8	0.2

Supplementary figure 3. Sorting RFP+ cells from xenografts. A. Plots of single cell RFP+ cell sorting by flow cytometry. Untransduced cells were used as a negative control for RFP living cell. RFP+ cells are sorted from living cells. Dead cells were excluded using toto-3 cell membrane staining. Single dots and contour depict cell populations. B. Percentage of RFP+ cells in FACS sorting events.

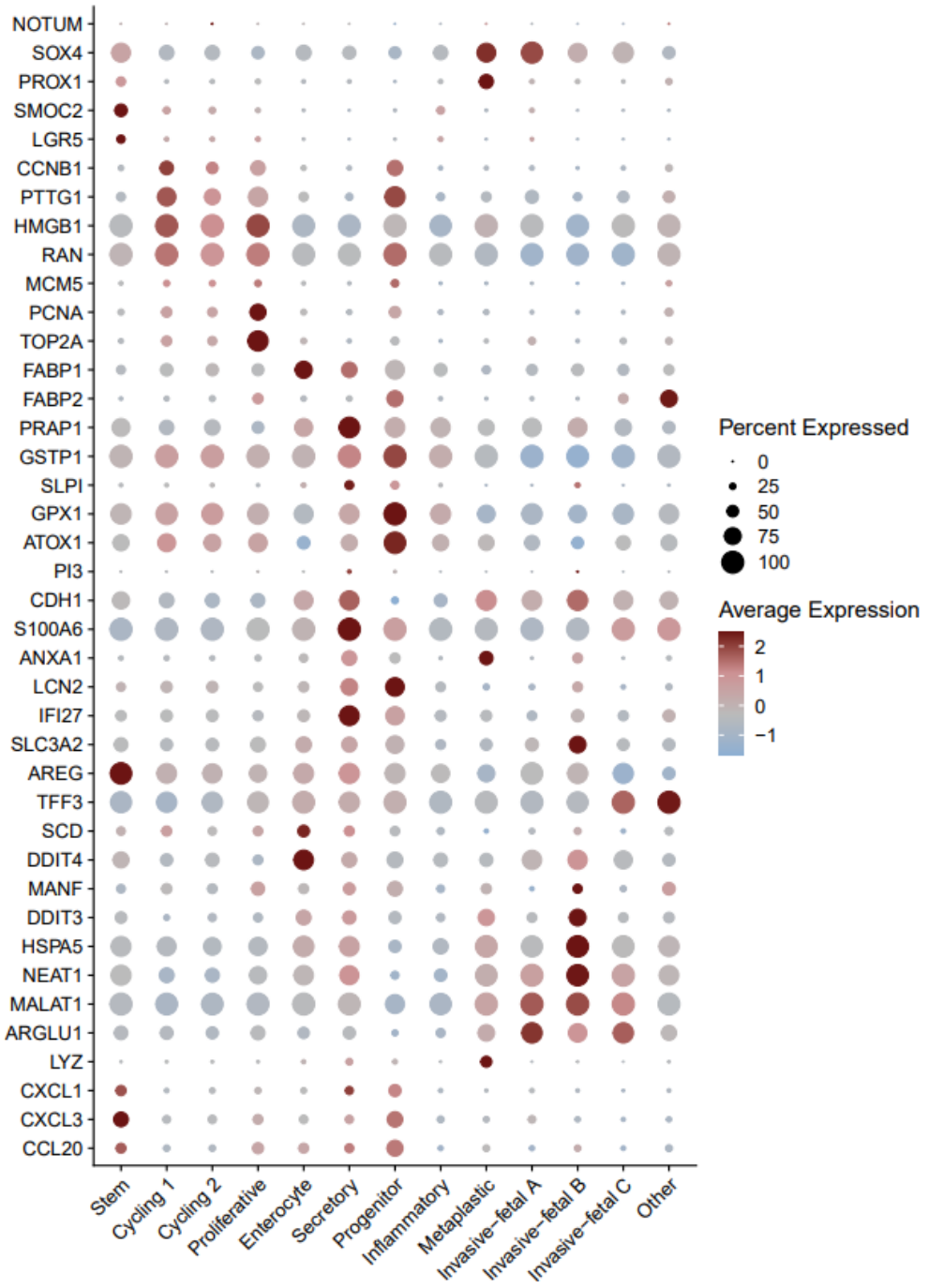


Supplementary figure 4. Clonal distribution in primary barcoded xenografts. Pie charts depicting the barcode distribution (in %) in PDXs captured through targeted DNA (amplicon) sequencing (A) or scRNA seq (B). Each color portrays a unique clone. Gray color represents clones of low representation (less than 0.1% of total barcode count). Liver met.: liver metastasis. NA: non available.

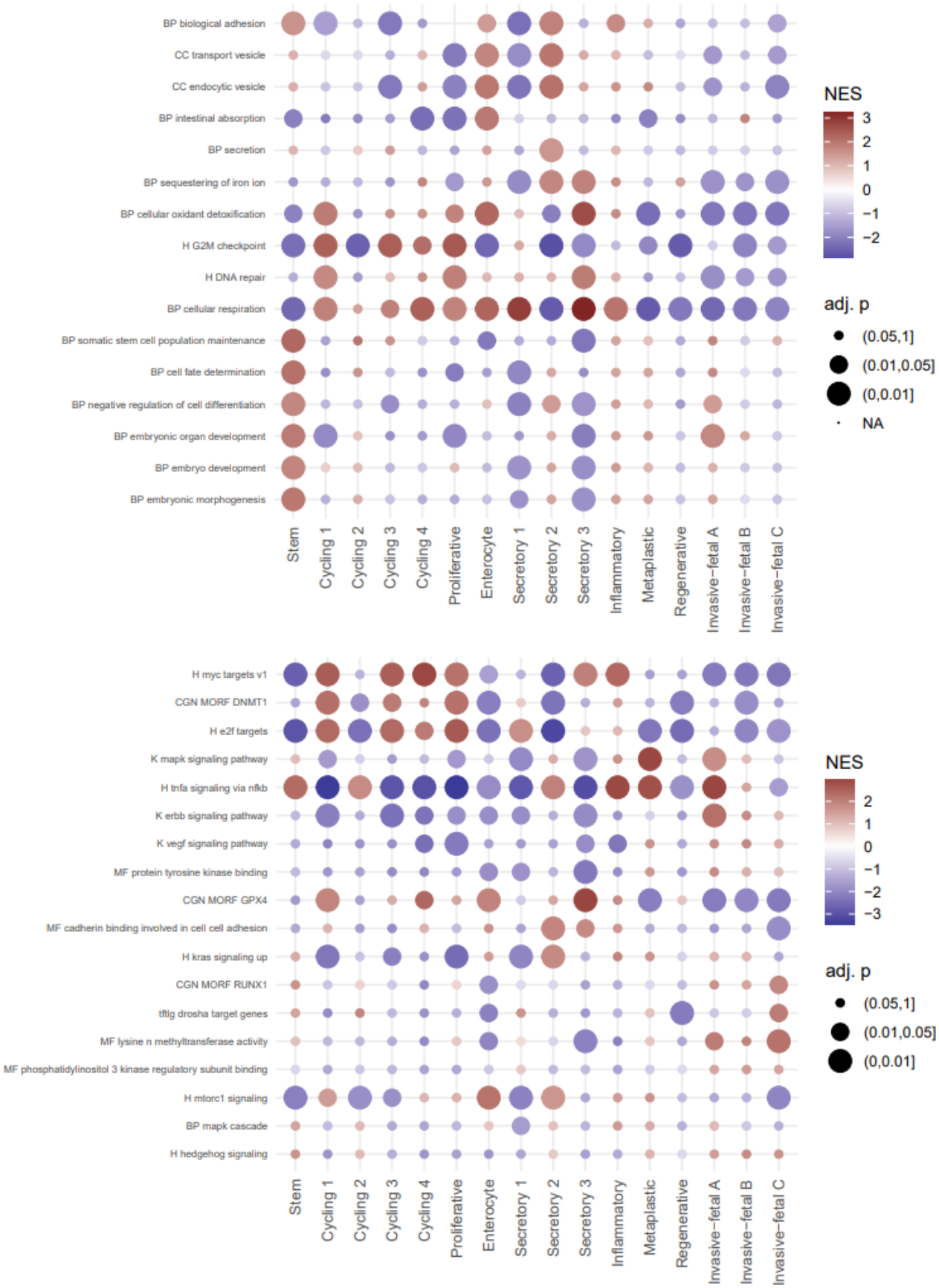
A


NOTUM
SOX4
PROX1
SMOC2
LGR5
CCNB1
PTTG1
HMGB1
RAN
MCM5
PCNA
TOP2A
FABP1
FABP2
PRAP1
GSTP1
SLPI
GPX1
ATOX1
PI3
CDH1
S100A6
ANXA1
LCN2
IFI27
SLC3A2
AREG
TFF3
SCD
DDIT4
MANF
DDIT3
HSPA5
NEAT1
MALAT1
ARGLU1
LYZ
CXCL1
CXCL3
CCL20
Stem
Cycling 1
Cycling 2
Cycling 3
Cycling 4
Proliferative
Enterocyte
Secretory 1
Secretory 2
Secretory 3
Inflammatory
Metaplastic
Regenerative
Invasive−fetal A
Invasive−fetal B
Invasive−fetal C
Percent Expressed
25
50
75
100
Average Expression
2
1
0
−1

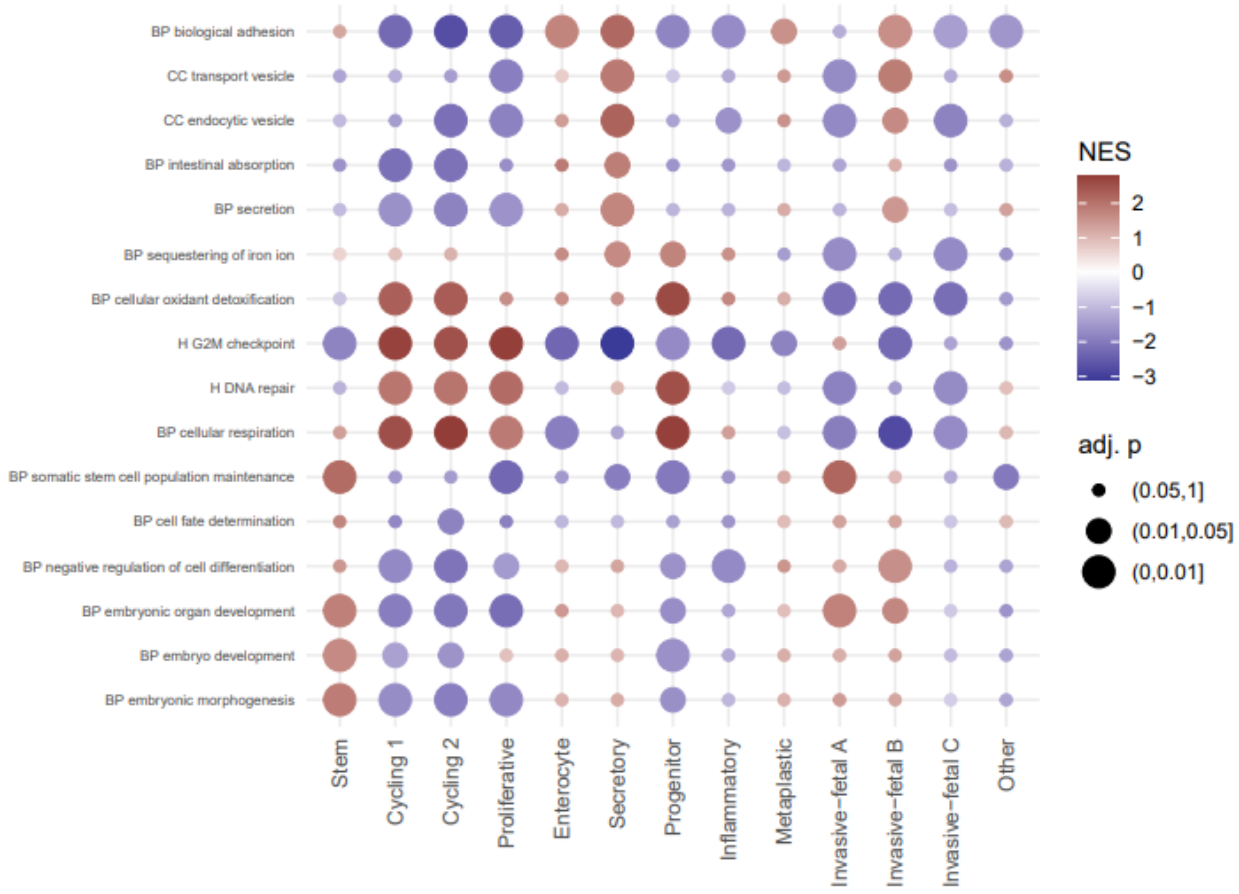
B

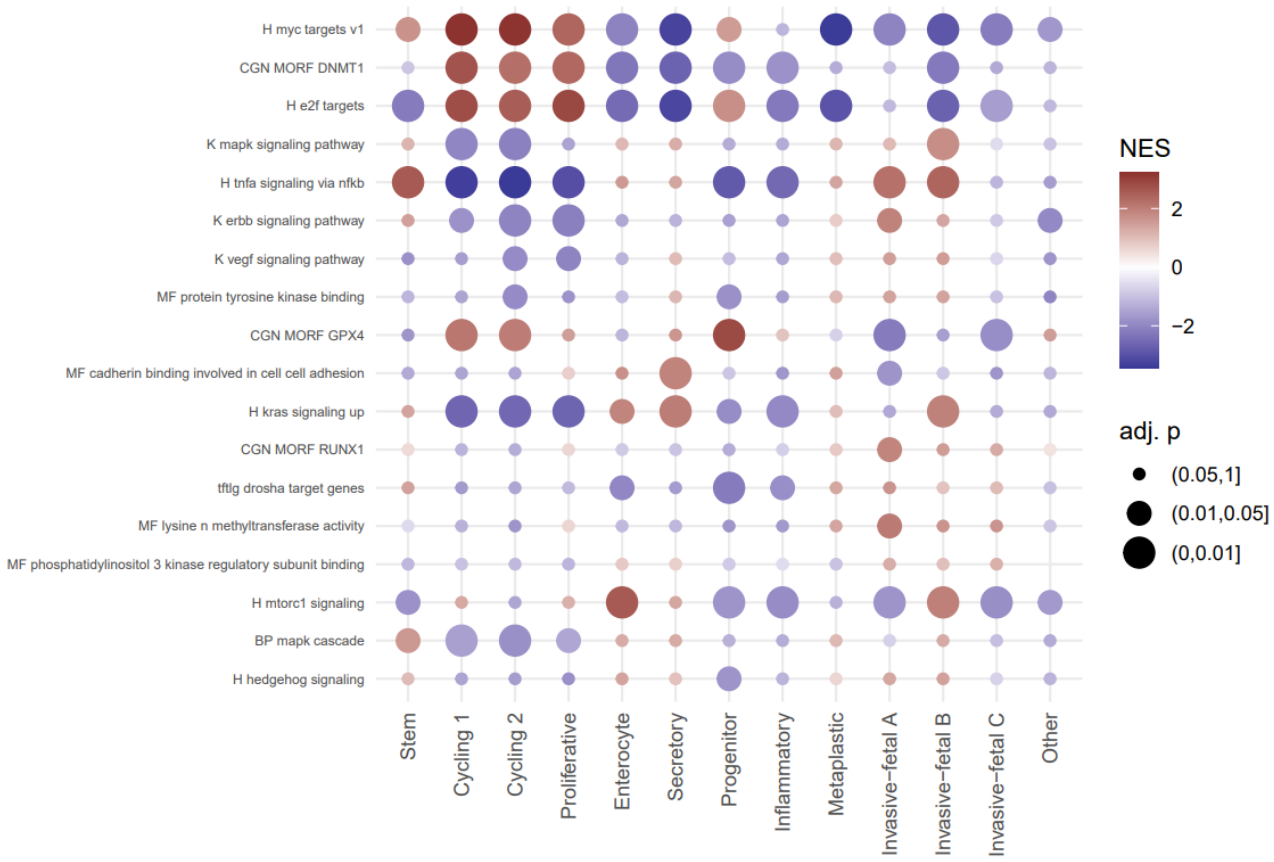


Supplementary figure 5. Cell type marker genes of transcriptional clusters of xenografts. Plots depicting the gene expression level of selected genes (y-axis) in assigned transcriptional clusters (x-axis) of xenografts derived from PDO2 and PDO3 (A and B respectively).


BP biological adhesion
CC transport vesicle
CC endocytic vesicle
BP intestinal absorption
BP secretion
BP sequestering of iron ion
BP cellular oxidant detoxification
H G2M checkpoint
H DNA repair
BP cellular respiration
BP somatic stem cell population maintenance
BP cell fate determination
BP negative regulation of cell differentiation
BP embryonic organ development
BP embryo development
BP embryonic morphogenesis
Stem
Cycling 1
Cycling 2
Cycling 3
Cycling 4
Proliferative
Enterocyte
Secretory 1
Secretory 2
Secretory 3
Inflammatory
Metaplastic
Regenerative
Invasive−fetal A
Invasive−fetal B
Invasive−fetal C
NES
3
2
1
0
−1
−2
adj. p
(0.05,1]
(0.01,0.05]
(0,0.01]
NA
H myc targets v1
CGN MORF DNMT1
H e2f targets
K mapk signaling pathway
H tnfa signaling via nfkb
K erbb signaling pathway
K vegf signaling pathway
MF protein tyrosine kinase binding
CGN MORF GPX4
MF cadherin binding involved in cell cell adhesion
H kras signaling up
CGN MORF RUNX1
tflg drosha target genes
MF lysine n methyltransferase activity
MF phosphatidylinositol 3 kinase regulatory subunit binding
H mtorc1 signaling
BP mapk cascade
H hedgehog signaling
Stem
Cycling 1
Cycling 2
Cycling 3
Cycling 4
Proliferative
Enterocyte
Secretory 1
Secretory 2
Secretory 3
Inflammatory
Metaplastic
Regenerative
Invasive−fetal A
Invasive−fetal B
Invasive−fetal C
NES
2
1
0
−1
−2
−3
adj. p
(0.05,1]
(0.01,0.05]
(0,0.01]

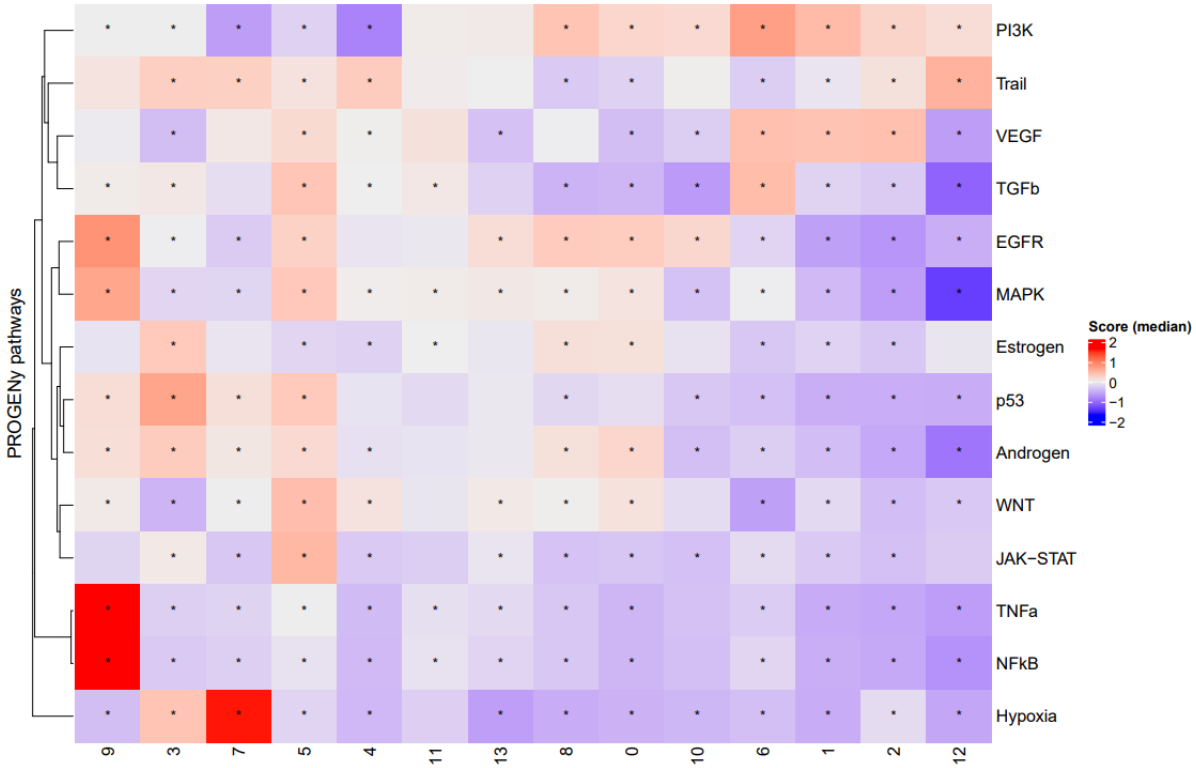
Supplementary figure 6. **GSEA markers of transcriptional clusters in PDO2.** First panel: plot showing the normalized enrichment score of GSEA terms biological processes and hallmarks in cell type-assigned transcriptional clusters of PDO2 xenograft. BP: biological process; CC: cellular component; H: hallmark. Second panel represents molecular pathways. CGN: cancer gene neighborhoods; MF: molecular function; K: canonical pathway.



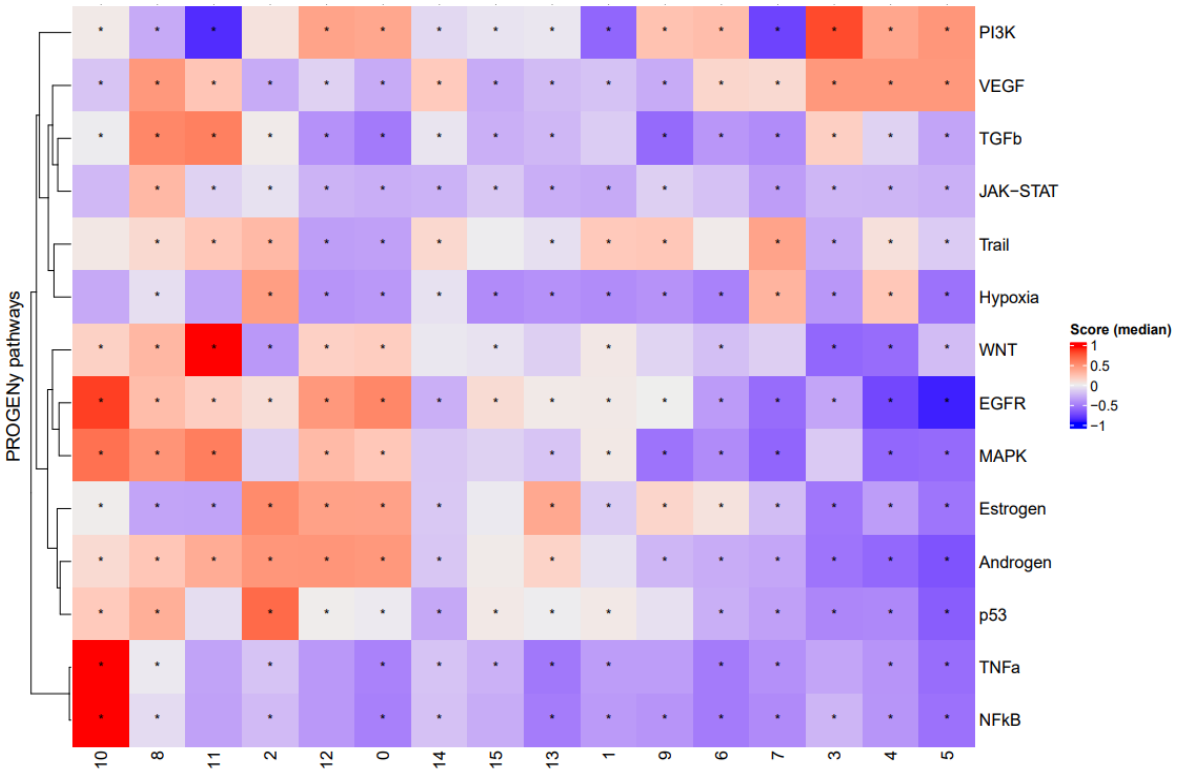


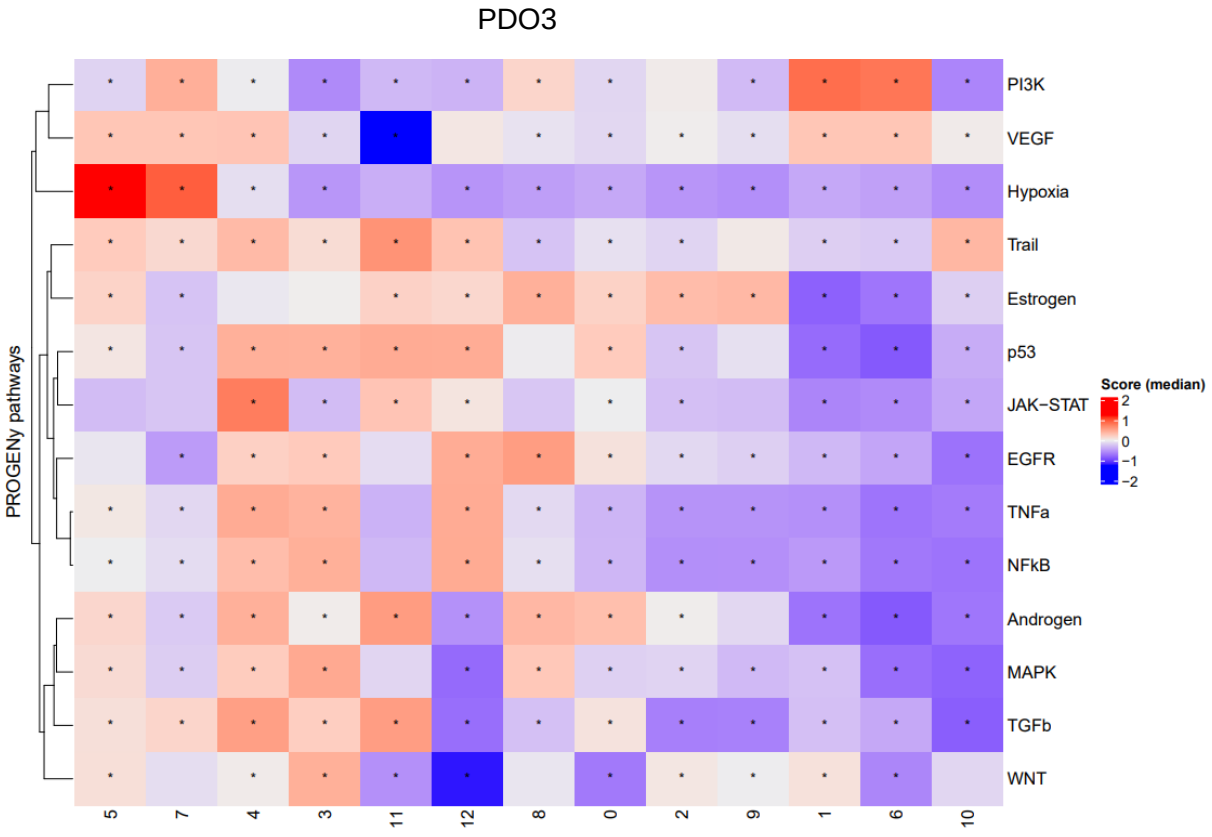
Supplementary figure 7. GSEA markers of transcriptional clusters in PDO3. First panel: plot showing the normalized enrichment score of GSEA terms biological processes and hallmarks in cell type-assigned transcriptional clusters of PDO3 xenograft. BP: biological process; CC: cellular component; H: hallmark. Second panel represents molecular pathways. CGN: cancer gene neighborhoods; MF: molecular function; K: canonical pathway.

PDO1


PROGENy pathways
PI3K
Trail
VEGF
TGFb
EGFR
MAPK
Estrogen
p53
Androgen
WNT
JAK−STAT
TNFa
NFkB
Hypoxia
9
3
7
5
4
11
13
8
0
10
6
1
2
12
Score (median)
2
1
0
−1
−2

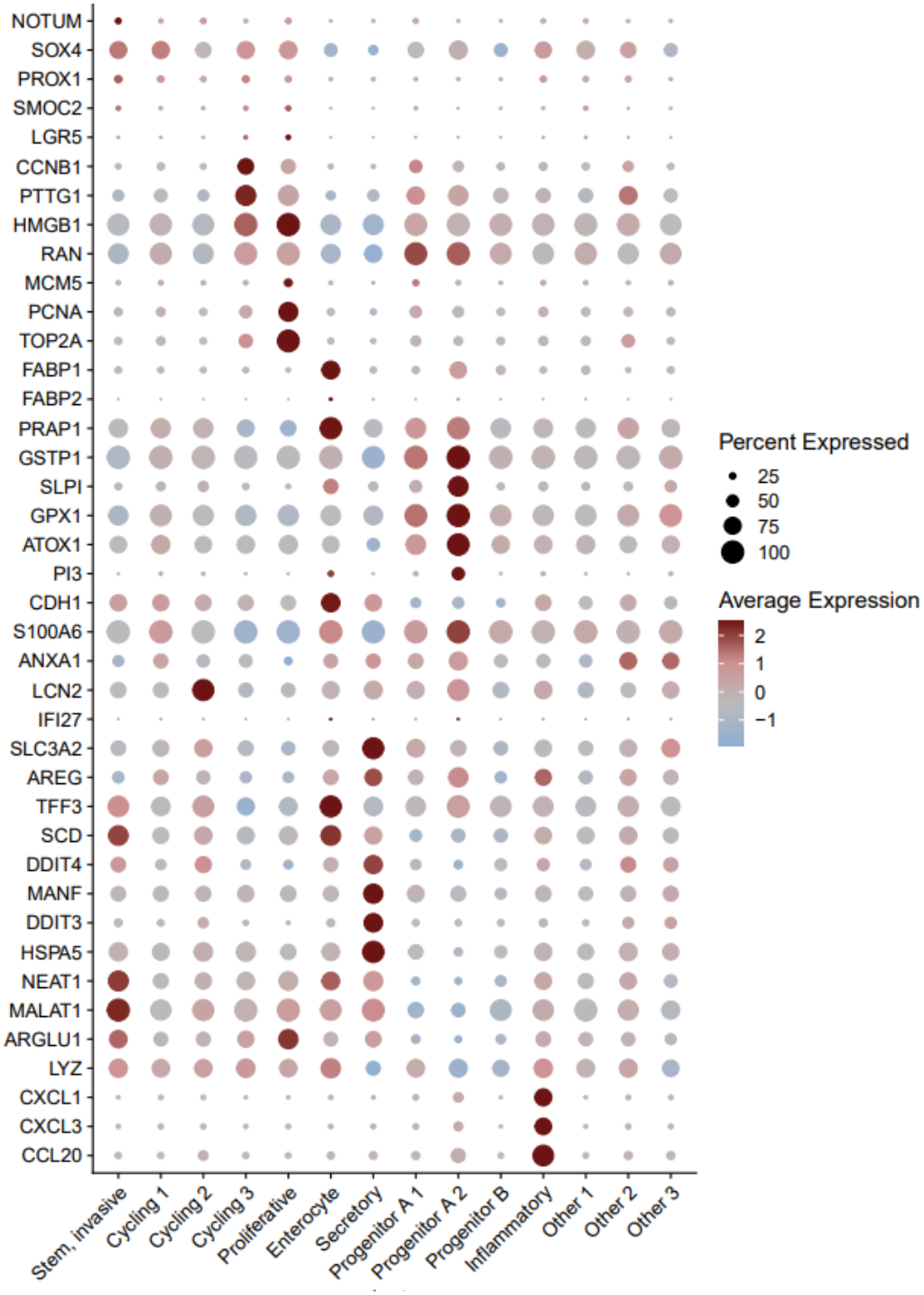
PDO2


PROGENy pathways
PI3K
VEGF
TGFb
JAK−STAT
Trail
Hypoxia
WNT
EGFR
MAPK
Estrogen
Androgen
p53
TNFa
NFkB
10
8
11
2
12
0
14
15
13
1
9
6
7
3
4
5
Score (median)
1
0.5
0
−0.5
−1

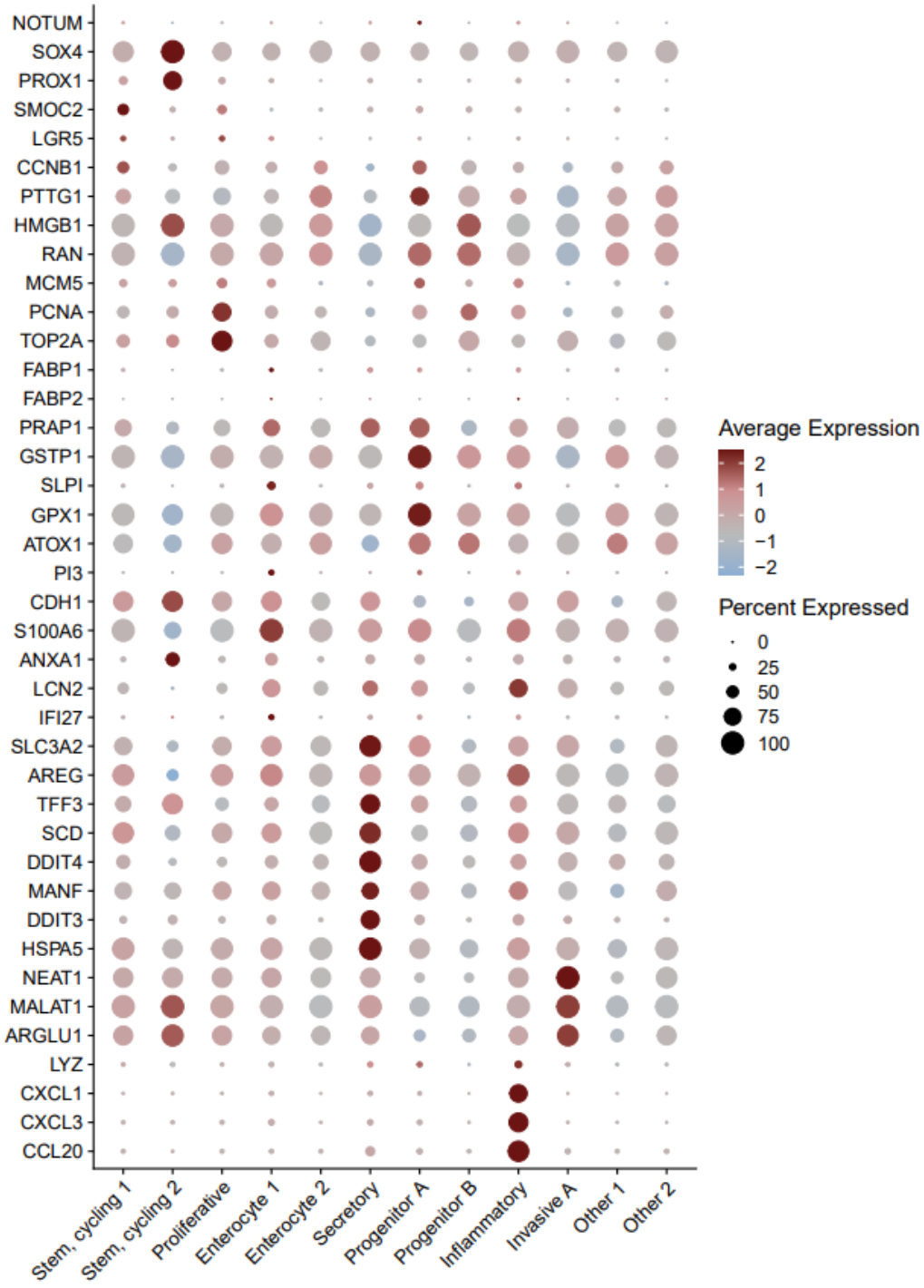


Supplementary figure 8: Progeny analysis in barcoded xenografts. Analysis showing the activation of canonical pathways in the transcriptional clusters of xenograft integrated data of tumors and metastases of each PDO. Numbers depict transcriptional clusters. Clusters (1, 2 and 6 of PDO1), (3, 4 and 5 of PDO2) and (1, 6 and 7 of PDO3) denote invasive fetal cell phenotypes. Clusters (3 and 5 of PDO1), (2 and 6, 8, 9 of PDO2) and (5 and 4 of PDO3) denote enterocyte and secretory cell phenotypes, respectively. Cluster description and annotation is detailed in the main text.

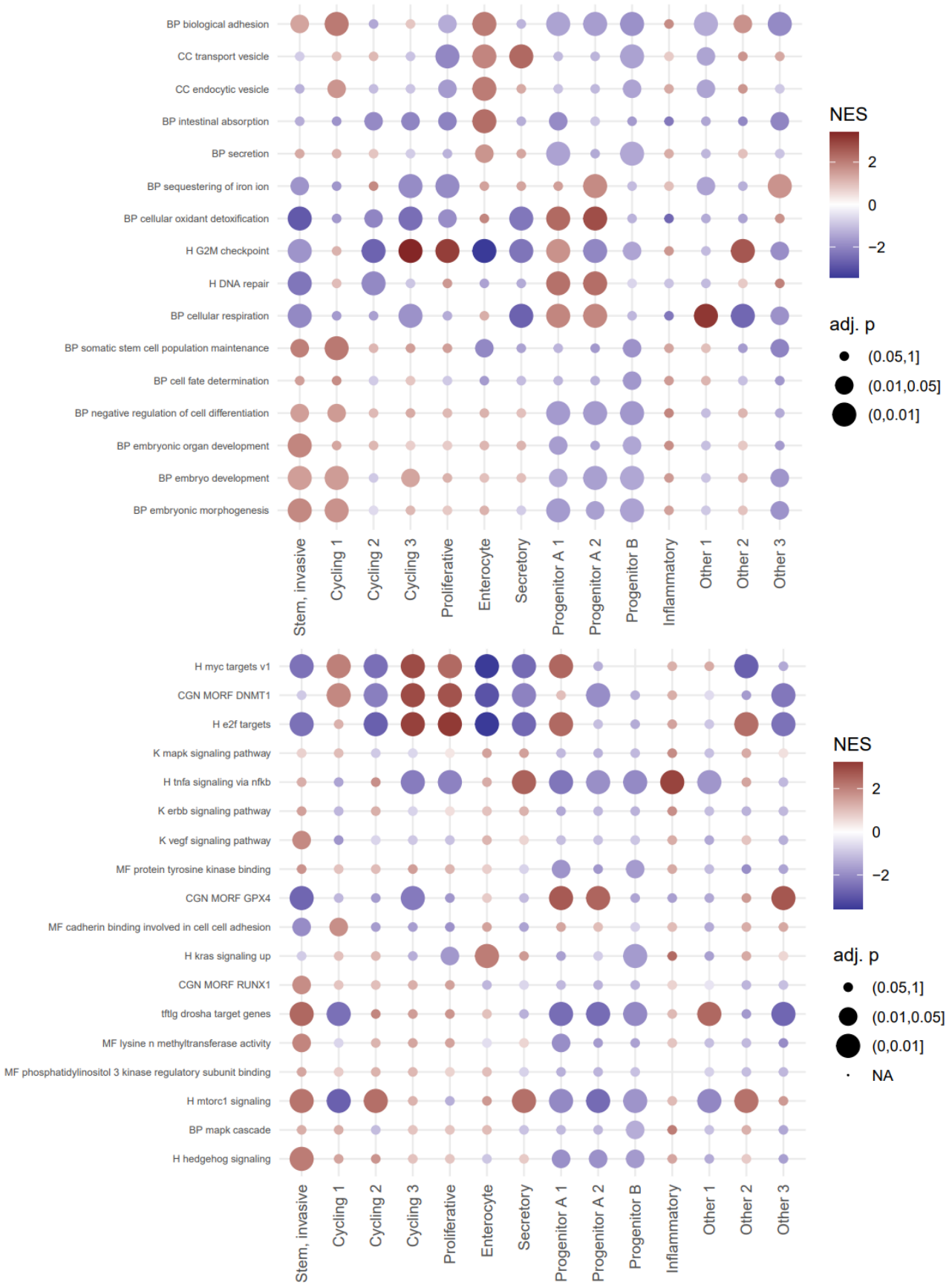
A


NOTUM
SOX4
PROX1
SMOC2
LGR5
CCNB1
PTTG1
HMGB1
RAN
MCM5
PCNA
TOP2A
FABP1
FABP2
PRAP1
GSTP1
SLPI
GPX1
ATOX1
PI3
CDH1
S100A6
ANXA1
LCN2
IFI27
SLC3A2
AREG
TFF3
SCD
DDIT4
MANF
DDIT3
HSPA5
NEAT1
MALAT1
ARGLU1
LYZ
CXCL1
CXCL3
CCL20
Stem, invasive
Cycling 1
Cycling 2
Cycling 3
Proliferative
Enterocyte
Secretory
Progenitor A 1
Progenitor A 2
Progenitor B
Inflammatory
Other 1
Other 2
Other 3
Percent Expressed
25
50
75
100
Average Expression
2
1
0
−1

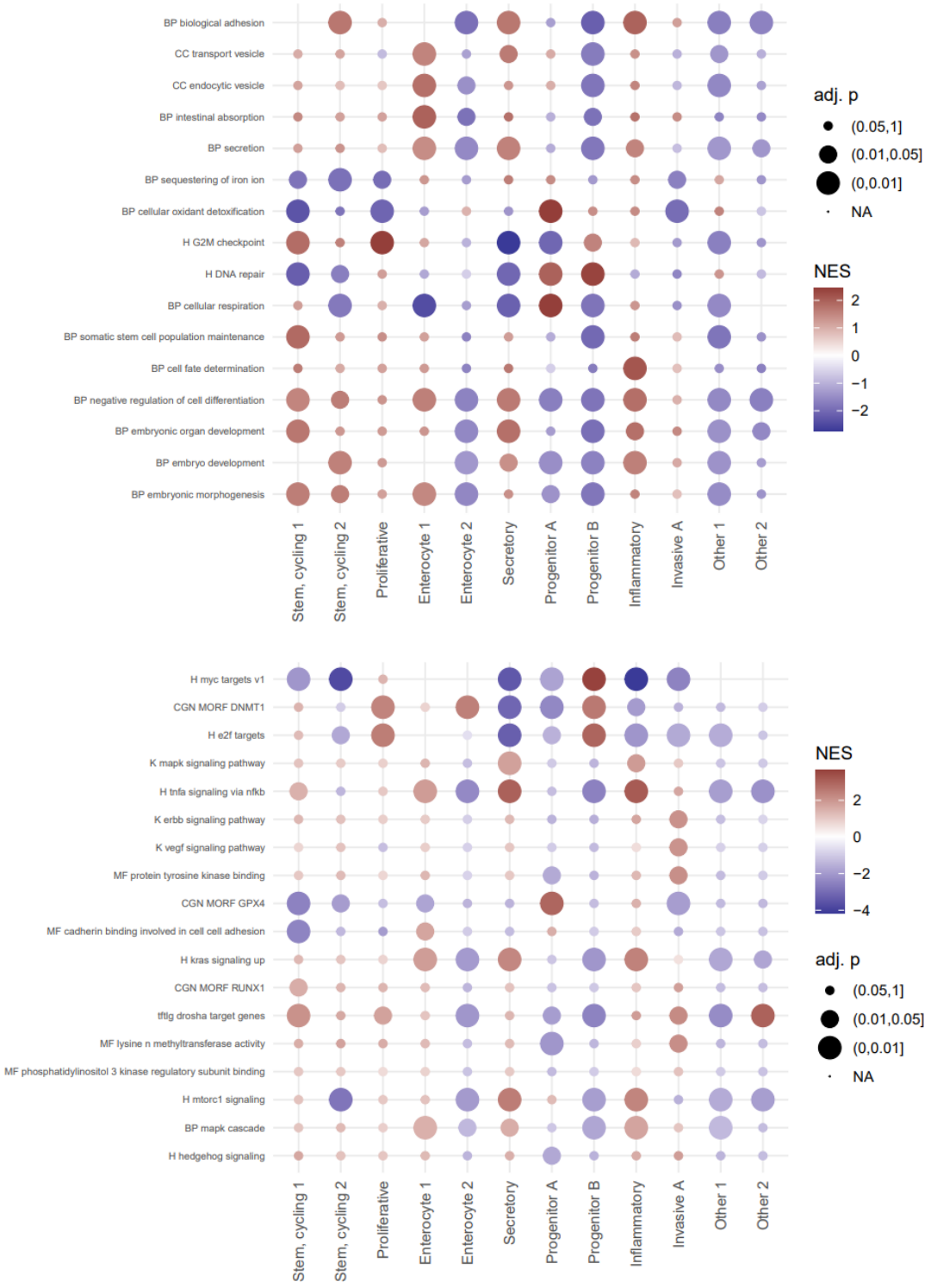
B



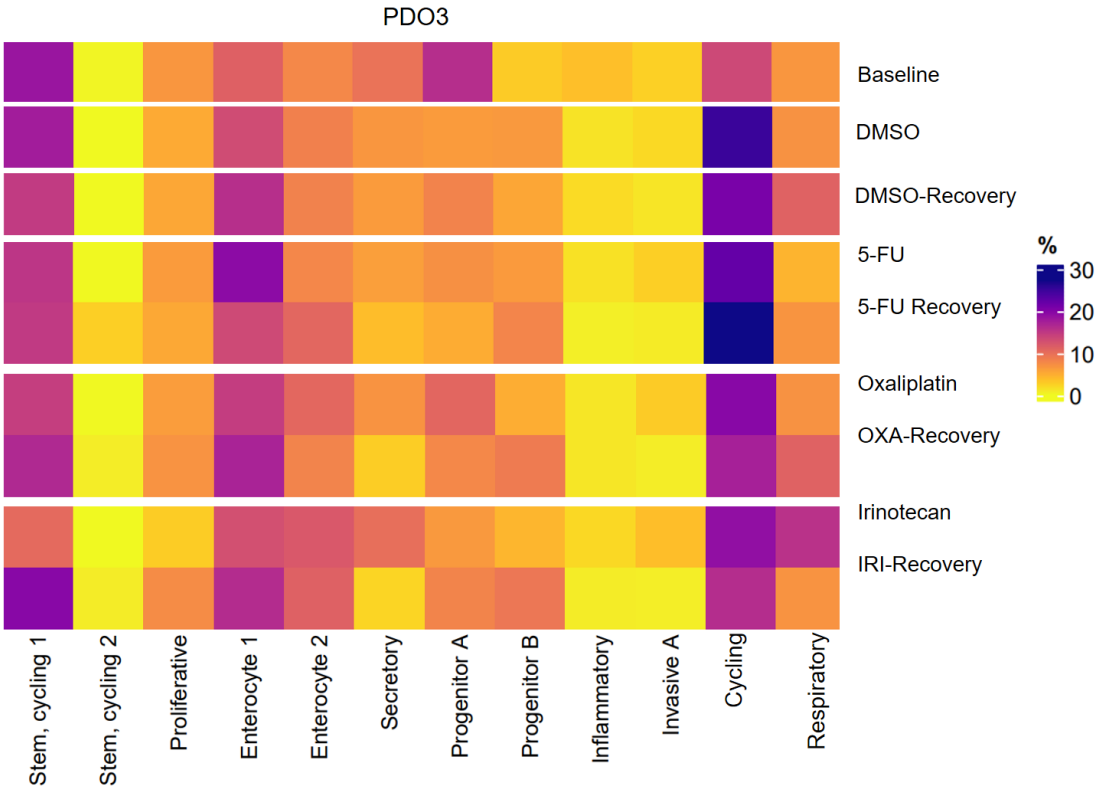
Supplementary figure 9. Cell type marker genes of transcriptional clusters of treated PDOs. Plots depicting the gene expression level of selected genes (vertical axis) in assigned transcriptional clusters (horizontal axis) of treated cells from PDO2 and PDO3 (A and B respectively).


BP biological adhesion
CC transport vesicle
CC endocytic vesicle
BP intestinal absorption
BP secretion
BP sequestering of iron ion
BP cellular oxidant detoxification
H G2M checkpoint
H DNA repair
BP cellular respiration
BP somatic stem cell population maintenance
BP cell fate determination
BP negative regulation of cell differentiation
BP embryonic organ development
BP embryo development
BP embryonic morphogenesis
Stem, invasive
Cycling 1
Cycling 2
Cycling 3
Proliferative
Enterocyte
Secretory
Progenitor A 1
Progenitor A 2
Progenitor B
Inflammatory
Other 1
Other 2
Other 3
NES
2
0
−2
adj. p
(0.05,1]
(0.01,0.05]
(0,0.01]
H myc targets v1
CGN MORF DNMT1
H e2f targets
K mapk signaling pathway
H tnfa signaling via nfkb
K erbb signaling pathway
K vegf signaling pathway
MF protein tyrosine kinase binding
CGN MORF GPX4
MF cadherin binding involved in cell cell adhesion
H kras signaling up
CGN MORF RUNX1
tftlg drosha target genes
MF lysine n methyltransferase activity
MF phosphatidylinositol 3 kinase regulatory subunit binding
H mtorc1 signaling
BP mapk cascade
H hedgehog signaling
Stem, invasive
Cycling 1
Cycling 2
Cycling 3
Proliferative
Enterocyte
Secretory
Progenitor A 1
Progenitor A 2
Progenitor B
Inflammatory
Other 1
Other 2
Other 3
NES
2
0
−2
adj. p
(0.05,1]
(0.01,0.05]
(0,0.01]
NA

Supplementary figure 10. GSEA markers of transcriptional clusters in treated PDO2. First panel: plot showing the normalized enrichment score of GSEA terms biological processes and hallmarks in cell type-assigned transcriptional clusters of PDO3 treated cells. BP: biological process; CC: cellular component; H: hallmark. Second panel represents molecular pathways. CGN: cancer gene neighborhoods; MF: molecular function; K: canonical pathway.



Supplementary figure 11. GSEA markers of transcriptional clusters in treated PDO3. First panel: plot showing the normalized enrichment score of GSEA terms biological processes and hallmarks in cell type-assigned transcriptional clusters of PDO3 treated cells. BP: biological process; CC: cellular component; H: hallmark. Second panel represents molecular pathways. CGN: cancer gene neighborhoods; MF: molecular function; K: canonical pathway.



Supplementary figure 12. Cluster enrichment in treated PDOs. Heatmap illustrating cluster abundance in baseline, treated and recovered cells of PDO1 and PDO2.

7. References

Al-Mehdi, A. B., Tozawa, K., Fisher, A. B., Shientag, L., Lee, A., & Muschel, R. J. (2000). Intravascular origin of metastasis from the proliferation of endothelium-attached tumor cells: a new model for metastasis. *Nat Med, 6*(1), 100-102. doi:10.1038/71429

Alamo, P., Gallardo, A., Pavón, M. A., Casanova, I., Trias, M., Mangues, M. A., . . . Céspedes, M. V. (2014). Subcutaneous preconditioning increases invasion and metastatic dissemination in mouse colorectal cancer models. *Dis Model Mech, 7*(3), 387-396. doi:10.1242/dmm.013995

Altmann, G. G. (1983). Morphological observations on mucus-secreting nongoblet cells in the deep crypts of the rat ascending colon. *Am J Anat, 167*(1), 95-117. doi:10.1002/aja.1001670109

Álvarez-Varela, A., Novellasdemunt, L., Barriga, F. M., Hernando-Momblona, X., Cañellas-Socias, A., Cano-Crespo, S., . . . Batlle, E. (2022). Mex3a marks drug-tolerant persister colorectal cancer cells that mediate relapse after chemotherapy. *Nat Cancer, 3*(9), 1052-1070. doi:10.1038/s43018-022-00402-0

André, T., Boni, C., Mounedji-Boudiaf, L., Navarro, M., Tabernero, J., Hickish, T., . . . de Gramont, A. (2004). Oxaliplatin, fluorouracil, and leucovorin as adjuvant treatment for colon cancer. *N Engl J Med, 350*(23), 2343-2351. doi:10.1056/NEJMoa032709

André, T., Shiu, K. K., Kim, T. W., Jensen, B. V., Jensen, L. H., Punt, C., . . . Diaz, L. A., Jr. (2020). Pembrolizumab in Microsatellite-Instability-High Advanced Colorectal Cancer. *N Engl J Med, 383*(23), 2207-2218. doi:10.1056/NEJMoa2017699

Ashcroft, G. S., Lei, K., Jin, W., Longenecker, G., Kulkarni, A. B., Greenwell-Wild, T., . . . Wahl, S. M. (2000). Secretory leukocyte protease inhibitor mediates non-redundant functions necessary for normal wound healing. *Nat Med, 6*(10), 1147-1153. doi:10.1038/80489

Bala, P., Rennhack, J. P., Aitymbayev, D., Morris, C., Moyer, S. M., Duronio, G. N., . . . Sethi, N. S. (2023). Aberrant cell state plasticity mediated by developmental reprogramming precedes colorectal cancer initiation. *Sci Adv, 9*(13), eadf0927. doi:10.1126/sciadv.adf0927

Baran, B., Mert Ozupek, N., Yerli Tetik, N., Acar, E., Bekcioglu, O., & Baskin, Y. (2018). Difference Between Left-Sided and Right-Sided Colorectal Cancer: A Focused Review of Literature. *Gastroenterology Res, 11*(4), 264-273. doi:10.14740/gr1062w

Barker, N. (2014). Adult intestinal stem cells: critical drivers of epithelial homeostasis and regeneration. *Nat Rev Mol Cell Biol, 15*(1), 19-33. doi:10.1038/nrm3721

Barker, N., van de Wetering, M., & Clevers, H. (2008). The intestinal stem cell. *Genes Dev, 22*(14), 1856-1864. doi:10.1101/gad.1674008

Barker, N., van Es, J. H., Kuipers, J., Kujala, P., van den Born, M., Cozijnsen, M., . . . Clevers, H. (2007). Identification of stem cells in small intestine and colon by marker gene Lgr5. *Nature, 449*(7165), 1003-1007. doi:10.1038/nature06196

Batlle, E., & Clevers, H. (2017). Cancer stem cells revisited. *Nat Med, 23*(10), 1124-1134. doi:10.1038/nm.4409

Batlle, E., Sancho, E., Francí, C., Domínguez, D., Monfar, M., Baulida, J., & García De Herreros, A. (2000). The transcription factor snail is a repressor of E-cadherin gene expression in epithelial tumour cells. *Nat Cell Biol, 2*(2), 84-89. doi:10.1038/35000034

Becht, E., de Reyniès, A., Giraldo, N. A., Pilati, C., Buttard, B., Lacroix, L., . . . Fridman, W. H. (2016). Immune and Stromal Classification of Colorectal Cancer Is Associated with Molecular Subtypes and Relevant for Precision Immunotherapy. *Clin Cancer Res, 22*(16), 4057-4066. doi:10.1158/1078-0432.Ccr-15-2879

Bernstein, B. E., Mikkelsen, T. S., Xie, X., Kamal, M., Huebert, D. J., Cuff, J., . . . Lander, E. S. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell, 125*(2), 315-326. doi:10.1016/j.cell.2006.02.041

Beumer, J., & Clevers, H. (2021). Cell fate specification and differentiation in the adult mammalian intestine. *Nat Rev Mol Cell Biol, 22*(1), 39-53. doi:10.1038/s41580-020-0278-0

Bloch, N., & Harel, D. (2016). The tumor as an organ: comprehensive spatial and temporal modeling of the tumor and its microenvironment. *BMC Bioinformatics, 17*(1), 317. doi:10.1186/s12859-016-1168-5

Brannon, A. R., Vakiani, E., Sylvester, B. E., Scott, S. N., McDermott, G., Shah, R. H., . . . Berger, M. F. (2014). Comparative sequencing analysis reveals high genomic concordance between matched primary and metastatic colorectal cancer lesions. *Genome Biol, 15*(8), 454. doi:10.1186/s13059-014-0454-7

Brown, M., Assen, F. P., Leithner, A., Abe, J., Schachner, H., Asfour, G., . . . Kerjaschki, D. (2018). Lymph node blood vessels provide exit routes for metastatic tumor cell dissemination in mice. *Science, 359*(6382), 1408-1411. doi:10.1126/science.aal3662

Burdziak, C., Alonso-Curbelo, D., Walle, T., Reyes, J., Barriga, F. M., Haviv, D., . . . Pe'er, D. (2023). Epigenetic plasticity cooperates with cell-cell interactions to direct pancreatic tumorigenesis. *Science, 380*(6645), eadd5327. doi:10.1126/science.add5327

Bürtin, F., Mullins, C. S., & Linnebacher, M. (2020). Mouse models of colorectal cancer: Past, present and future perspectives. *World J Gastroenterol, 26*(13), 1394-1426. doi:10.3748/wjg.v26.i13.1394

Cappell, K. M., & Kochenderfer, J. N. (2023). Long-term outcomes following CAR T cell therapy: what we know so far. *Nat Rev Clin Oncol, 20*(6), 359-371. doi:10.1038/s41571-023-00754-1

Castro, F., Dirks, W. G., Fähnrich, S., Hotz-Wagenblatt, A., Pawlita, M., & Schmitt, M. (2013). High-throughput SNP-based authentication of human cell lines. *Int J Cancer, 132*(2), 308-314. doi:10.1002/ijc.27675

Cayrefourcq, L., Thomas, F., Mazard, T., Assenat, E., Assou, S., & Alix-Panabières, C. (2021). Selective treatment pressure in colon cancer drives the molecular profile of resistant circulating tumor cell clones. *Mol Cancer, 20*(1), 30. doi:10.1186/s12943-021-01326-6

Chambers, A. F., Groom, A. C., & MacDonald, I. C. (2002). Dissemination and growth of cancer cells in metastatic sites. *Nat Rev Cancer, 2*(8), 563-572. doi:10.1038/nrc865

Chan, D. K. H., & Buczacki, S. J. A. (2021). Tumour heterogeneity and evolutionary dynamics in colorectal cancer. *Oncogenesis, 10*(7), 53. doi:10.1038/s41389-021-00342-x

Chao, S., Zhang, F., Yan, H., Wang, L., Zhang, L., Wang, Z., . . . Bu, P. (2023). Targeting intratumor heterogeneity suppresses colorectal cancer chemoresistance and metastasis. *EMBO Rep, 24*(8), e56416. doi:10.15252/embr.202256416

Chau, J. F., Jia, D., Wang, Z., Liu, Z., Hu, Y., Zhang, X., . . . Li, B. (2012). A crucial role for bone morphogenetic protein-Smad1 signalling in the DNA damage response. *Nat Commun, 3*, 836. doi:10.1038/ncomms1832

Chen, C., Lin, W., Huang, Y., Chen, X., Wang, H., & Teng, L. (2021). The Essential Factors of Establishing Patient-derived Tumor Model. *J Cancer, 12*(1), 28-37. doi:10.7150/jca.51749

Chen, L., Yang, F., Chen, S., & Tai, J. (2022). Mechanisms on chemotherapy resistance of colorectal cancer stem cells and research progress of reverse transformation: A mini-review. *Front Med (Lausanne), 9*, 995882. doi:10.3389/fmed.2022.995882

Chen, P., Cescon, M., & Bonaldo, P. (2013). Collagen VI in cancer and its biological mechanisms. *Trends Mol Med, 19*(7), 410-417. doi:10.1016/j.molmed.2013.04.001

Chen, Y., Li, J., Xiao, J. K., Xiao, L., Xu, B. W., & Li, C. (2021). The lncRNA NEAT1 promotes the epithelial-mesenchymal transition and metastasis of osteosarcoma cells by sponging miR-483 to upregulate STAT3 expression. *Cancer Cell Int, 21*(1), 90. doi:10.1186/s12935-021-01780-8

Cheung, A. H., Chow, C., Zhang, J., Zhou, Y., Huang, T., Ng, K. C., . . . To, K. F. (2018). Specific targeting of point mutations in EGFR L858R-positive lung cancer by CRISPR/Cas9. *Lab Invest, 98*(7), 968-976. doi:10.1038/s41374-018-0056-1

Cheung, K. J., & Ewald, A. J. (2016). A collective route to metastasis: Seeding by tumor cell clusters. *Science, 352*(6282), 167-169. doi:10.1126/science.aaf6546

Clevers, H. (2013). The intestinal crypt, a prototype stem cell compartment. *Cell, 154*(2), 274-284. doi:10.1016/j.cell.2013.07.004

Comprehensive molecular characterization of human colon and rectal cancer. (2012). *Nature, 487*(7407), 330-337. doi:10.1038/nature11252

Corrò, C., Novellasdemunt, L., & Li, V. S. W. (2020). A brief history of organoids. *Am J Physiol Cell Physiol, 319*(1), C151-c165. doi:10.1152/ajpcell.00120.2020

Cross, W., Kovac, M., Mustonen, V., Temko, D., Davis, H., Baker, A. M., . . . Tomlinson, I. P. M. (2018). The evolutionary landscape of colorectal tumorigenesis. *Nat Ecol Evol, 2*(10), 1661-1672. doi:10.1038/s41559-018-0642-z

da Silva-Diz, V., Lorenzo-Sanz, L., Bernat-Peguera, A., Lopez-Cerda, M., & Muñoz, P. (2018). Cancer cell plasticity: Impact on tumor progression and therapy response. *Semin Cancer Biol, 53*, 48-58. doi:10.1016/j.semcancer.2018.08.009

Dart, A. (2023). EMT in chemoresistance. *Nat Rev Cancer, 23*(6), 349. doi:10.1038/s41568-023-00581-7

Das, S., Anczuków, O., Akerman, M., & Krainer, A. R. (2012). Oncogenic splicing factor SRSF1 is a critical transcriptional target of MYC. *Cell Rep, 1*(2), 110-117. doi:10.1016/j.celrep.2011.12.001

de Sousa e Melo, F., Kurtova, A. V., Harnoss, J. M., Kljavin, N., Hoeck, J. D., Hung, J., . . . de Sauvage, F. J. (2017). A distinct role for Lgr5(+) stem cells in primary and metastatic colon cancer. *Nature, 543*(7647), 676-680. doi:10.1038/nature21713

de Sousa, E. M. F., & de Sauvage, F. J. (2019). Cellular Plasticity in Intestinal Homeostasis and Disease. *Cell Stem Cell, 24*(1), 54-64. doi:10.1016/j.stem.2018.11.019

Debaugnies, M., Rodríguez-Acebes, S., Blondeau, J., Parent, M. A., Zocco, M., Song, Y., . . . Blanpain, C. (2023). RHOJ controls EMT-associated resistance to chemotherapy. *Nature, 616*(7955), 168-175. doi:10.1038/s41586-023-05838-7

Dhimolea, E., de Matos Simoes, R., Kansara, D., Al'Khafaji, A., Bouyssou, J., Weng, X., . . . Mitsiades, C. S. (2021). An Embryonic Diapause-like Adaptation with Suppressed Myc Activity Enables Tumor Treatment Persistence. *Cancer Cell, 39*(2), 240-256.e211. doi:10.1016/j.ccell.2020.12.002

Dienstmann, R., Vermeulen, L., Guinney, J., Kopetz, S., Tejpar, S., & Tabernero, J. (2017). Consensus molecular subtypes and the evolution of precision medicine in colorectal cancer. *Nat Rev Cancer, 17*(2), 79-92. doi:10.1038/nrc.2016.126

Dieter, S. M., Ball, C. R., Hoffmann, C. M., Nowrouzi, A., Herbst, F., Zavidij, O., . . . Glimm, H. (2011). Distinct types of tumor-initiating cells form human colon cancer tumors and metastases. *Cell Stem Cell, 9*(4), 357-365. doi:10.1016/j.stem.2011.08.010

Dieter, S. M., Glimm, H., & Ball, C. R. (2017). Colorectal cancer-initiating cells caught in the act. *EMBO Mol Med, 9*(7), 856-858. doi:10.15252/emmm.201707858

Dongre, A., Rashidian, M., Eaton, E. N., Reinhardt, F., Thiru, P., Zagorulya, M., . . . Weinberg, R. A. (2021). Direct and Indirect Regulators of Epithelial-Mesenchymal Transition-Mediated Immunosuppression in Breast Carcinomas. *Cancer Discov, 11*(5), 1286-1305. doi:10.1158/2159-8290.Cd-20-0603

Driessens, G., Beck, B., Caauwe, A., Simons, B. D., & Blanpain, C. (2012). Defining the mode of tumour growth by clonal analysis. *Nature, 488*(7412), 527-530. doi:10.1038/nature11344

Dull, T., Zufferey, R., Kelly, M., Mandel, R. J., Nguyen, M., Trono, D., & Naldini, L. (1998). A third-generation lentivirus vector with a conditional packaging system. *J Virol, 72*(11), 8463-8471. doi:10.1128/jvi.72.11.8463-8471.1998

Echeverria, G. V., Ge, Z., Seth, S., Zhang, X., Jeter-Jones, S., Zhou, X., . . . Piwnica-Worms, H. (2019). Resistance to neoadjuvant chemotherapy in triple-negative breast cancer mediated by a reversible drug-tolerant state. *Sci Transl Med, 11*(488). doi:10.1126/scitranslmed.aav0936

Eger, A., Aigner, K., Sonderegger, S., Dampier, B., Oehler, S., Schreiber, M., . . . Foisner, R. (2005). DeltaEF1 is a transcriptional repressor of E-cadherin and regulates epithelial plasticity in breast cancer cells. *Oncogene, 24*(14), 2375-2385. doi:10.1038/sj.onc.1208429

Eifler, K., & Vertegaal, A. C. O. (2015). SUMOylation-Mediated Regulation of Cell Cycle Progression and Cancer. *Trends Biochem Sci, 40*(12), 779-793. doi:10.1016/j.tibs.2015.09.006

Engelman, J. A., Zejnullahu, K., Mitsudomi, T., Song, Y., Hyland, C., Park, J. O., . . . Jänne, P. A. (2007). MET amplification leads to gefitinib resistance in lung cancer by activating ERBB3 signaling. *Science, 316*(5827), 1039-1043. doi:10.1126/science.1141478

Evan, T., Wang, V. M., & Behrens, A. (2022). The roles of intratumour heterogeneity in the biology and treatment of pancreatic ductal adenocarcinoma. *Oncogene, 41*(42), 4686-4695. doi:10.1038/s41388-022-02448-x

Ewing, J. (1919). *Neoplastic disease. Philadelphia: W.B. Saunders;* .

Fazilaty, H. (2023). Restoration of embryonic gene expression patterns in tissue regeneration and disease. *Nat Rev Mol Cell Biol, 24*(6), 375-376. doi:10.1038/s41580-023-00586-y

Fazilaty, H., & Basler, K. (2023). Reactivation of embryonic genetic programs in tissue regeneration and disease. *Nat Genet, 55*(11), 1792-1806. doi:10.1038/s41588-023-01526-4

Fearon, E. R., & Vogelstein, B. (1990). A genetic model for colorectal tumorigenesis. *Cell, 61*(5), 759-767. doi:10.1016/0092-8674(90)90186-i

Fenelon, J. C., Banerjee, A., & Murphy, B. D. (2014). Embryonic diapause: development on hold. *Int J Dev Biol, 58*(2-4), 163-174. doi:10.1387/ijdb.140074bm

Fidler, I. J. (1970). Metastasis: quantitative analysis of distribution and fate of tumor emboli labeled with 125 I-5-iodo-2'-deoxyuridine. *J Natl Cancer Inst, 45*(4), 773-782.

Fischer, K. R., Durrans, A., Lee, S., Sheng, J., Li, F., Wong, S. T., . . . Gao, D. (2015). Epithelial-to-mesenchymal transition is not required for lung metastasis but contributes to chemoresistance. *Nature, 527*(7579), 472-476. doi:10.1038/nature15748

Flanagan, D. J., Pentinmikko, N., Luopajärvi, K., Willis, N. J., Gilroy, K., Raven, A. P., . . . Sansom, O. J. (2021). NOTUM from Apc-mutant cells biases clonal competition to initiate cancer. *Nature, 594*(7863), 430-435. doi:10.1038/s41586-021-03525-z

Flavahan, W. A., Gaskell, E., & Bernstein, B. E. (2017). Epigenetic plasticity and the hallmarks of cancer. *Science, 357*(6348). doi:10.1126/science.aal2380

Fleischmajer, R. (1967). Epithelial-mesenchymal interactions. *Science, 157*(3795), 1472-1482. doi:10.1126/science.157.3795.1472

Fumagalli, A., Oost, K. C., Kester, L., Morgner, J., Bornes, L., Bruens, L., . . . van Rheenen, J. (2020). Plasticity of Lgr5-Negative Cancer Cells Drives Metastasis in Colorectal Cancer. *Cell Stem Cell, 26*(4), 569-578.e567. doi:10.1016/j.stem.2020.02.008

Fumagalli, A., Suijkerbuijk, S. J. E., Begthel, H., Beerling, E., Oost, K. C., Snippert, H. J., . . . Drost, J. (2018). A surgical orthotopic organoid transplantation approach in mice to visualize and study colorectal cancer progression. *Nat Protoc, 13*(2), 235-247. doi:10.1038/nprot.2017.137

Gaiani, F., Marchesi, F., Negri, F., Greco, L., Malesci, A., de'Angelis, G. L., & Laghi, L. (2021). Heterogeneity of Colorectal Cancer Progression: Molecular Gas and Brakes. *Int J Mol Sci, 22*(10). doi:10.3390/ijms22105246

Gao, Y., Bado, I., Wang, H., Zhang, W., Rosen, J. M., & Zhang, X. H. (2019). Metastasis Organotropism: Redefining the Congenial Soil. *Dev Cell, 49*(3), 375-391. doi:10.1016/j.devcel.2019.04.012

Garrido-Laguna, I., McGregor, K. A., Wade, M., Weis, J., Gilcrease, W., Burr, L., . . . Sharma, S. (2013). A phase I/II study of decitabine in combination with panitumumab in patients with wild-type (wt)

KRAS metastatic colorectal cancer. *Invest New Drugs, 31*(5), 1257-1264. doi:10.1007/s10637-013-9947-6

Gerlinger, M. (2018). Metastasis Seeding Cells: Lone Invaders or Mass Migrators? *Clin Cancer Res, 24*(9), 2032-2034. doi:10.1158/1078-0432.Ccr-17-3644

Giancotti, F. G. (2013). Mechanisms governing metastatic dormancy and reactivation. *Cell, 155*(4), 750-764. doi:10.1016/j.cell.2013.10.029

Giantonio, B. J., Catalano, P. J., Meropol, N. J., O'Dwyer, P. J., Mitchell, E. P., Alberts, S. R., . . . Benson, A. B., 3rd. (2007). Bevacizumab in combination with oxaliplatin, fluorouracil, and leucovorin (FOLFOX4) for previously treated metastatic colorectal cancer: results from the Eastern Cooperative Oncology Group Study E3200. *J Clin Oncol, 25*(12), 1539-1544. doi:10.1200/jco.2006.09.6305

Giessler, K. M., Kleinheinz, K., Huebschmann, D., Balasubramanian, G. P., Dubash, T. D., Dieter, S. M., . . . Glimm, H. (2017). Genetic subclone architecture of tumor clone-initiating cells in colorectal cancer. *J Exp Med, 214*(7), 2073-2088. doi:10.1084/jem.20162017

Gilbert, L. A., Horlbeck, M. A., Adamson, B., Villalta, J. E., Chen, Y., Whitehead, E. H., . . . Weissman, J. S. (2014). Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell, 159*(3), 647-661. doi:10.1016/j.cell.2014.09.029

Giroux, V., & Rustgi, A. K. (2017). Metaplasia: tissue injury adaptation and a precursor to the dysplasia-cancer sequence. *Nat Rev Cancer, 17*(10), 594-604. doi:10.1038/nrc.2017.68

Greaves, M., & Maley, C. C. (2012). Clonal evolution in cancer. *Nature, 481*(7381), 306-313. doi:10.1038/nature10762

Guinney, J., Dienstmann, R., Wang, X., de Reyniès, A., Schlicker, A., Soneson, C., . . . Tejpar, S. (2015). The consensus molecular subtypes of colorectal cancer. *Nat Med, 21*(11), 1350-1356. doi:10.1038/nm.3967

Guler, G. D., Tindell, C. A., Pitti, R., Wilson, C., Nichols, K., KaiWai Cheung, T., . . . Classon, M. (2017). Repression of Stress-Induced LINE-1 Expression Protects Cancer Cell Subpopulations from Lethal Drug Exposure. *Cancer Cell, 32*(2), 221-237.e213. doi:10.1016/j.ccell.2017.07.002

Guo, L., Lee, Y. T., Zhou, Y., & Huang, Y. (2022). Targeting epigenetic regulatory machinery to overcome cancer therapy resistance. *Semin Cancer Biol, 83*, 487-502. doi:10.1016/j.semcancer.2020.12.022

Guo, Z., Ashrafizadeh, M., Zhang, W., Zou, R., Sethi, G., & Zhang, X. (2023). Molecular profile of metastasis, cell plasticity and EMT in pancreatic cancer: a pre-clinical connection to aggressiveness and drug resistance. *Cancer Metastasis Rev*. doi:10.1007/s10555-023-10125-y

Gupta, G. P., & Massagué, J. (2006). Cancer metastasis: building a framework. *Cell, 127*(4), 679-695. doi:10.1016/j.cell.2006.11.001

Hajra, K. M., & Fearon, E. R. (2002). Cadherin and catenin alterations in human cancer. *Genes Chromosomes Cancer, 34*(3), 255-268. doi:10.1002/gcc.10083

Hamidi, H., & Ivaska, J. (2018). Every step of the way: integrins in cancer progression and metastasis. *Nat Rev Cancer, 18*(9), 533-548. doi:10.1038/s41568-018-0038-z

Hanahan, D., & Folkman, J. (1996). Patterns and emerging mechanisms of the angiogenic switch during tumorigenesis. *Cell, 86*(3), 353-364. doi:10.1016/s0092-8674(00)80108-7

Hangauer, M. J., Viswanathan, V. S., Ryan, M. J., Bole, D., Eaton, J. K., Matov, A., . . . McManus, M. T. (2017). Drug-tolerant persister cancer cells are vulnerable to GPX4 inhibition. *Nature, 551*(7679), 247-250. doi:10.1038/nature24297

Hao, Y., Baker, D., & Ten Dijke, P. (2019). TGF-β-Mediated Epithelial-Mesenchymal Transition and Cancer Metastasis. *Int J Mol Sci, 20*(11). doi:10.3390/ijms20112767

Heide, T., Househam, J., Cresswell, G. D., Spiteri, I., Lynn, C., Mossner, M., . . . Sottoriva, A. (2022). The co-evolution of the genome and epigenome in colorectal cancer. *Nature, 611*(7937), 733-743. doi:10.1038/s41586-022-05202-1

Hess, K. R., Varadhachary, G. R., Taylor, S. H., Wei, W., Raber, M. N., Lenzi, R., & Abbruzzese, J. L. (2006). Metastatic patterns in adenocarcinoma. *Cancer, 106*(7), 1624-1633. doi:10.1002/cncr.21778

Hiam-Galvez, K. J., Allen, B. M., & Spitzer, M. H. (2021). Systemic immunity in cancer. *Nat Rev Cancer, 21*(6), 345-359. doi:10.1038/s41568-021-00347-z

Hidalgo, M., Amant, F., Biankin, A. V., Budinská, E., Byrne, A. T., Caldas, C., . . . Villanueva, A. (2014). Patient-derived xenograft models: an emerging platform for translational cancer research. *Cancer Discov, 4*(9), 998-1013. doi:10.1158/2159-8290.Cd-14-0001

Hines, W. C., Jr., & Hines, W. C., 3rd. (2022). Lost in transduction: Critical considerations when using viral vectors. *Front Cell Dev Biol, 10*, 1080265. doi:10.3389/fcell.2022.1080265

Hirata, A., Utikal, J., Yamashita, S., Aoki, H., Watanabe, A., Yamamoto, T., . . . Yamada, Y. (2013). Dose-dependent roles for canonical Wnt signalling in de novo crypt formation and cell cycle properties of the colonic epithelium. *Development, 140*(1), 66-75. doi:10.1242/dev.084103

Hiratsuka, S., Watanabe, A., Aburatani, H., & Maru, Y. (2006). Tumour-mediated upregulation of chemoattractants and recruitment of myeloid cells predetermines lung metastasis. *Nat Cell Biol, 8*(12), 1369-1375. doi:10.1038/ncb1507

Hoesel, B., & Schmid, J. A. (2013). The complexity of NF-κB signaling in inflammation and cancer. *Mol Cancer, 12*, 86. doi:10.1186/1476-4598-12-86

Horlbeck, M. A., Gilbert, L. A., Villalta, J. E., Adamson, B., Pak, R. A., Chen, Y., . . . Weissman, J. S. (2016). Compact and highly active next-generation libraries for CRISPR-mediated gene repression and activation. *Elife, 5*. doi:10.7554/eLife.19760

Hossain, M. S., Karuniawati, H., Jairoun, A. A., Urbi, Z., Ooi, J., John, A., . . . Hadi, M. A. (2022). Colorectal Cancer: A Review of Carcinogenesis, Global Epidemiology, Current Challenges, Risk Factors, Preventive and Treatment Strategies. *Cancers (Basel), 14*(7). doi:10.3390/cancers14071732

Househam, J., Heide, T., Cresswell, G. D., Spiteri, I., Kimberley, C., Zapata, L., . . . Graham, T. A. (2022). Phenotypic plasticity and genetic control in colorectal cancer evolution. *Nature, 611*(7937), 744-753. doi:10.1038/s41586-022-05311-x

Humphries, A., & Wright, N. A. (2008). Colonic crypt organization and tumorigenesis. *Nat Rev Cancer, 8*(6), 415-424. doi:10.1038/nrc2392

Hunter, K. W., Amin, R., Deasy, S., Ha, N. H., & Wakefield, L. (2018). Genetic insights into the morass of metastatic heterogeneity. *Nat Rev Cancer, 18*(4), 211-223. doi:10.1038/nrc.2017.126

Hussein, A. M., Balachandar, N., Mathieu, J., & Ruohola-Baker, H. (2022). Molecular Regulators of Embryonic Diapause and Cancer Diapause-like State. *Cells, 11*(19). doi:10.3390/cells11192929

Hussein, A. M., Wang, Y., Mathieu, J., Margaretha, L., Song, C., Jones, D. C., . . . Ruohola-Baker, H. (2020). Metabolic Control over mTOR-Dependent Diapause-like State. *Dev Cell, 52*(2), 236-250.e237. doi:10.1016/j.devcel.2019.12.018

Ishitsuka, Y., Kawachi, Y., Taguchi, S., Maruyama, H., Fujisawa, Y., Furuta, J., . . . Otsuka, F. (2012). Pituitary tumor-transforming gene 1 enhances proliferation and suppresses early differentiation of keratinocytes. *J Invest Dermatol, 132*(7), 1775-1784. doi:10.1038/jid.2012.74

Jadhav, U., Saxena, M., O'Neill, N. K., Saadatpour, A., Yuan, G. C., Herbert, Z., . . . Shivdasani, R. A. (2017). Dynamic Reorganization of Chromatin Accessibility Signatures during Dedifferentiation of Secretory Precursors into Lgr5+ Intestinal Stem Cells. *Cell Stem Cell, 21*(1), 65-77.e65. doi:10.1016/j.stem.2017.05.001

Jaiswal, J. K., & Nylandsted, J. (2015). S100 and annexin proteins identify cell membrane damage as the Achilles heel of metastatic cancer cells. *Cell Cycle, 14*(4), 502-509. doi:10.1080/15384101.2014.995495

Jin, J., Ma, M., Shi, S., Wang, J., Xiao, P., Yu, H. F., . . . Teng, C. B. (2022). Copper enhances genotoxic drug resistance via ATOX1 activated DNA damage repair. *Cancer Lett, 536*, 215651. doi:10.1016/j.canlet.2022.215651

Jo, A., Denduluri, S., Zhang, B., Wang, Z., Yin, L., Yan, Z., . . . Haydon, R. C. (2014). The versatile functions of Sox9 in development, stem cells, and human diseases. *Genes Dis, 1*(2), 149-161. doi:10.1016/j.gendis.2014.09.004

Joanito, I., Wirapati, P., Zhao, N., Nawaz, Z., Yeo, G., Lee, F., . . . Tan, I. B. (2022). Single-cell and bulk transcriptome sequencing identifies two epithelial tumor cell states and refines the consensus molecular classification of colorectal cancer. *Nat Genet, 54*(7), 963-975. doi:10.1038/s41588-022-01100-4

Justus, C. R., Leffler, N., Ruiz-Echevarria, M., & Yang, L. V. (2014). In vitro cell migration and invasion assays. *J Vis Exp*(88). doi:10.3791/51046

Kadota, K., Nitadori, J. I., Sima, C. S., Ujiie, H., Rizk, N. P., Jones, D. R., . . . Travis, W. D. (2015). Tumor Spread through Air Spaces is an Important Pattern of Invasion and Impacts the Frequency and Location of Recurrences after Limited Resection for Small Stage I Lung Adenocarcinomas. *J Thorac Oncol, 10*(5), 806-814. doi:10.1097/jto.0000000000000486

Kai, F., Drain, A. P., & Weaver, V. M. (2019). The Extracellular Matrix Modulates the Metastatic Journey. *Dev Cell, 49*(3), 332-346. doi:10.1016/j.devcel.2019.03.026

Kaushik, V., Sharma, S., Tiwari, M., & Tiwari, V. (2022). Antipersister strategies against stress induced bacterial persistence. *Microb Pathog, 164*, 105423. doi:10.1016/j.micpath.2022.105423

Kerbel, R. S., Waghorne, C., Korczak, B., Lagarde, A., & Breitman, M. L. (1988). Clonal dominance of primary tumours by metastatic cells: genetic analysis and biological implications. *Cancer Surv, 7*(4), 597-629.

Kessenbrock, K., Plaks, V., & Werb, Z. (2010). Matrix metalloproteinases: regulators of the tumor microenvironment. *Cell, 141*(1), 52-67. doi:10.1016/j.cell.2010.03.015

Kim, Y. J., Sheu, K. M., Tsoi, J., Abril-Rodriguez, G., Medina, E., Grasso, C. S., . . . Ribas, A. (2021). Melanoma dedifferentiation induced by IFN-γ epigenetic remodeling in response to anti-PD-1 therapy. *J Clin Invest, 131*(12). doi:10.1172/jci145859

Koedoot, E., Smid, M., Foekens, J. A., Martens, J. W. M., Le Dévédec, S. E., & van de Water, B. (2019). Co-regulated gene expression of splicing factors as drivers of cancer progression. *Sci Rep, 9*(1), 5484. doi:10.1038/s41598-019-40759-4

Kriegsmann, K., Cremer, M., Zgorzelski, C., Harms, A., Muley, T., Winter, H., . . . Kriegsmann, M. (2019). Agreement of CK5/6, p40, and p63 immunoreactivity in non-small cell lung cancer. *Pathology, 51*(3), 240-245. doi:10.1016/j.pathol.2018.11.009

Lamar, J. M., Stern, P., Liu, H., Schindler, J. W., Jiang, Z. G., & Hynes, R. O. (2012). The Hippo pathway target, YAP, promotes metastasis through its TEAD-interaction domain. *Proc Natl Acad Sci U S A, 109*(37), E2441-2450. doi:10.1073/pnas.1212021109

Laughney, A. M., Hu, J., Campbell, N. R., Bakhoum, S. F., Setty, M., Lavallée, V. P., . . . Massagué, J. (2020). Regenerative lineages and immune-mediated pruning in lung cancer metastasis. *Nat Med, 26*(2), 259-269. doi:10.1038/s41591-019-0750-6

Lehuédé, C., Dupuy, F., Rabinovitch, R., Jones, R. G., & Siegel, P. M. (2016). Metabolic Plasticity as a Determinant of Tumor Growth and Metastasis. *Cancer Res, 76*(18), 5201-5208. doi:10.1158/0008-5472.Can-16-0266

Lengyel, E. (2010). Ovarian cancer development and metastasis. *Am J Pathol, 177*(3), 1053-1064. doi:10.2353/ajpath.2010.100105

Lenos, K. J., Miedema, D. M., Lodestijn, S. C., Nijman, L. E., van den Bosch, T., Romero Ros, X., . . . Vermeulen, L. (2018). Stem cell functionality is microenvironmentally defined during tumour

expansion and therapy response in colon cancer. *Nat Cell Biol, 20*(10), 1193-1202. doi:10.1038/s41556-018-0179-z

Li, S., Yang, M., Teng, S., Lin, K., Wang, Y., Zhang, Y., . . . Wang, D. (2023). Chromatin accessibility dynamics in colorectal cancer liver metastasis: Uncovering the liver tropism at single cell resolution. *Pharmacol Res, 195*, 106896. doi:10.1016/j.phrs.2023.106896

Liau, B. B., Sievers, C., Donohue, L. K., Gillespie, S. M., Flavahan, W. A., Miller, T. E., . . . Bernstein, B. E. (2017). Adaptive Chromatin Remodeling Drives Glioblastoma Stem Cell Plasticity and Drug Tolerance. *Cell Stem Cell, 20*(2), 233-246.e237. doi:10.1016/j.stem.2016.11.003

Liebig, C., Ayala, G., Wilks, J. A., Berger, D. H., & Albo, D. (2009). Perineural invasion in cancer: a review of the literature. *Cancer, 115*(15), 3379-3391. doi:10.1002/cncr.24396

Liu, C., Zhang, L., Liu, H., & Cheng, K. (2017). Delivery strategies of the CRISPR-Cas9 gene-editing system for therapeutic applications. *J Control Release, 266*, 17-26. doi:10.1016/j.jconrel.2017.09.012

Liu, J., Qiu, J., Zhang, Z., Zhou, L., Li, Y., Ding, D., . . . Lang, T. (2021). SOX4 maintains the stemness of cancer cells via transcriptionally enhancing HDAC1 revealed by comparative proteomics study. *Cell Biosci, 11*(1), 23. doi:10.1186/s13578-021-00539-y

Liu, L., Yan, Q., Chen, Z., Wei, X., Li, L., Tang, D., . . . Sun, D. (2023). Overview of research progress and application of experimental models of colorectal cancer. *Front Pharmacol, 14*, 1193213. doi:10.3389/fphar.2023.1193213

Liu, Y., & Cao, X. (2016). Characteristics and Significance of the Pre-metastatic Niche. *Cancer Cell, 30*(5), 668-681. doi:10.1016/j.ccell.2016.09.011

Liu, Y., Chen, H., Xiao, L., Dong, P., Ma, Y., Zhou, Y., . . . Shen, L. (2023). Notum enhances gastric cancer stem-like cell properties through upregulation of Sox2 by PI3K/AKT signaling pathway. *Cell Oncol (Dordr)*. doi:10.1007/s13402-023-00875-w

Lugassy, C., Kleinman, H. K., Vermeulen, P. B., & Barnhill, R. L. (2020). Angiotropism, pericytic mimicry and extravascular migratory metastasis: an embryogenesis-derived program of tumor spread. *Angiogenesis, 23*(1), 27-41. doi:10.1007/s10456-019-09695-9

Luzzi, K. J., MacDonald, I. C., Schmidt, E. E., Kerkvliet, N., Morris, V. L., Chambers, A. F., & Groom, A. C. (1998). Multistep nature of metastatic inefficiency: dormancy of solitary cells after successful extravasation and limited survival of early micrometastases. *Am J Pathol, 153*(3), 865-873. doi:10.1016/s0002-9440(10)65628-3

Mani, S. A., Guo, W., Liao, M. J., Eaton, E. N., Ayyanan, A., Zhou, A. Y., . . . Weinberg, R. A. (2008). The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell, 133*(4), 704-715. doi:10.1016/j.cell.2008.03.027

Marisa, L., Blum, Y., Taieb, J., Ayadi, M., Pilati, C., Le Malicot, K., . . . Laurent-Puig, P. (2021). Intratumor CMS Heterogeneity Impacts Patient Prognosis in Localized Colon Cancer. *Clin Cancer Res, 27*(17), 4768-4780. doi:10.1158/1078-0432.Ccr-21-0529

Markowitz, S. D., & Bertagnolli, M. M. (2009). Molecular origins of cancer: Molecular basis of colorectal cancer. *N Engl J Med, 361*(25), 2449-2460. doi:10.1056/NEJMra0804588

Martínez-Cardús, A., Moran, S., Musulen, E., Moutinho, C., Manzano, J. L., Martinez-Balibrea, E., . . . Esteller, M. (2016). Epigenetic Homogeneity Within Colorectal Tumors Predicts Shorter Relapse-Free and Overall Survival Times for Patients With Locoregional Cancer. *Gastroenterology, 151*(5), 961-972. doi:10.1053/j.gastro.2016.08.001

Marusyk, A., Janiszewska, M., & Polyak, K. (2020). Intratumor Heterogeneity: The Rosetta Stone of Therapy Resistance. *Cancer Cell, 37*(4), 471-484. doi:10.1016/j.ccell.2020.03.007

Masoodi, T., Siraj, S., Siraj, A. K., Azam, S., Qadri, Z., Parvathareddy, S. K., . . . Al-Kuraya, K. S. (2020). Genetic heterogeneity and evolutionary history of high-grade ovarian carcinoma and matched distant metastases. *Br J Cancer, 122*(8), 1219-1230. doi:10.1038/s41416-020-0763-4

Matheu, A., Collado, M., Wise, C., Manterola, L., Cekaite, L., Tye, A. J., . . . Lovell-Badge, R. (2012). Oncogenicity of the developmental transcription factor Sox9. *Cancer Res, 72*(5), 1301-1315. doi:10.1158/0008-5472.Can-11-3660

Mathijssen, R. H., Loos, W. J., Verweij, J., & Sparreboom, A. (2002). Pharmacology of topoisomerase I inhibitors irinotecan (CPT-11) and topotecan. *Curr Cancer Drug Targets, 2*(2), 103-123. doi:10.2174/1568009023333890

Mehta, G. A., Parker, J. S., Silva, G. O., Hoadley, K. A., Perou, C. M., & Gatza, M. L. (2017). Amplification of SOX4 promotes PI3K/Akt signaling in human breast cancer. *Breast Cancer Res Treat, 162*(3), 439-450. doi:10.1007/s10549-017-4139-2

Miao, Q., Hill, M. C., Chen, F., Mo, Q., Ku, A. T., Ramos, C., . . . Nguyen, H. (2019). SOX11 and SOX4 drive the reactivation of an embryonic gene program during murine wound repair. *Nat Commun, 10*(1), 4042. doi:10.1038/s41467-019-11880-9

Miao, Z. F., Lewis, M. A., Cho, C. J., Adkins-Threats, M., Park, D., Brown, J. W., . . . Mills, J. C. (2020). A Dedicated Evolutionarily Conserved Molecular Network Licenses Differentiated Cells to Return to the Cell Cycle. *Dev Cell, 55*(2), 178-194.e177. doi:10.1016/j.devcel.2020.07.005

Mittal, V. (2018). Epithelial Mesenchymal Transition in Tumor Metastasis. *Annu Rev Pathol, 13*, 395-412. doi:10.1146/annurev-pathol-020117-043854

Miura, S., & Suzuki, A. (2017). Generation of Mouse and Human Organoid-Forming Intestinal Progenitor Cells by Direct Lineage Reprogramming. *Cell Stem Cell, 21*(4), 456-471.e455. doi:10.1016/j.stem.2017.08.020

Mohme, M., Riethdorf, S., & Pantel, K. (2017). Circulating and disseminated tumour cells - mechanisms of immune surveillance and escape. *Nat Rev Clin Oncol, 14*(3), 155-167. doi:10.1038/nrclinonc.2016.144

Morgan, E., Arnold, M., Gini, A., Lorenzoni, V., Cabasag, C. J., Laversanne, M., . . . Bray, F. (2023). Global burden of colorectal cancer in 2020 and 2040: incidence and mortality estimates from GLOBOCAN. *Gut, 72*(2), 338-344. doi:10.1136/gutjnl-2022-327736

Müller, A., Homey, B., Soto, H., Ge, N., Catron, D., Buchanan, M. E., . . . Zlotnik, A. (2001). Involvement of chemokine receptors in breast cancer metastasis. *Nature, 410*(6824), 50-56. doi:10.1038/35065016

Naxerova, K., Reiter, J. G., Brachtel, E., Lennerz, J. K., van de Wetering, M., Rowan, A., . . . Jain, R. K. (2017). Origins of lymphatic and distant metastases in human colorectal cancer. *Science, 357*(6346), 55-60. doi:10.1126/science.aai8515

Nishiyama, A., & Nakanishi, M. (2021). Navigating the DNA methylation landscape of cancer. *Trends Genet, 37*(11), 1012-1027. doi:10.1016/j.tig.2021.05.002

Oh, B. Y., Shin, H. T., Yun, J. W., Kim, K. T., Kim, J., Bae, J. S., . . . Park, W. Y. (2019). Intratumor heterogeneity inferred from targeted deep sequencing as a prognostic indicator. *Sci Rep, 9*(1), 4542. doi:10.1038/s41598-019-41098-0

Okada, S., Vaeteewoottacharn, K., & Kariya, R. (2019). Application of Highly Immunocompromised Mice for the Establishment of Patient-Derived Xenograft (PDX) Models. *Cells, 8*(8). doi:10.3390/cells8080889

Oskarsson, T., Batlle, E., & Massagué, J. (2014). Metastatic stem cells: sources, niches, and vital pathways. *Cell Stem Cell, 14*(3), 306-321. doi:10.1016/j.stem.2014.02.002

Overman, M. J., McDermott, R., Leach, J. L., Lonardi, S., Lenz, H. J., Morse, M. A., . . . André, T. (2017). Nivolumab in patients with metastatic DNA mismatch repair-deficient or microsatellite instability-high colorectal cancer (CheckMate 142): an open-label, multicentre, phase 2 study. *Lancet Oncol, 18*(9), 1182-1191. doi:10.1016/s1470-2045(17)30422-9

Paget, S. (1989). The distribution of secondary growths in cancer of the breast. 1889. *Cancer Metastasis Rev, 8*(2), 98-101.

Pastushenko, I., Mauri, F., Song, Y., de Cock, F., Meeusen, B., Swedlund, B., . . . Blanpain, C. (2021). Fat1 deletion promotes hybrid EMT state, tumour stemness and metastasis. *Nature, 589*(7842), 448-455. doi:10.1038/s41586-020-03046-1

Pattabiraman, D. R., Bierie, B., Kober, K. I., Thiru, P., Krall, J. A., Zill, C., . . . Weinberg, R. A. (2016). Activation of PKA leads to mesenchymal-to-epithelial transition and loss of tumor-initiating ability. *Science, 351*(6277), aad3680. doi:10.1126/science.aad3680

Pereira, E. R., Kedrin, D., Seano, G., Gautier, O., Meijer, E. F. J., Jones, D., . . . Padera, T. P. (2018). Lymph node metastases can invade local blood vessels, exit the node, and colonize distant organs in mice. *Science, 359*(6382), 1403-1407. doi:10.1126/science.aal3622

Petaccia de Macedo, M., Melo, F. M., Ribeiro, H. S. C., Marques, M. C., Kagohara, L. T., Begnami, M. D., . . . Cunha, I. W. (2017). KRAS mutation status is highly homogeneous between areas of the primary tumor and the corresponding metastasis of colorectal adenocarcinomas: one less problem in patient care. *Am J Cancer Res, 7*(9), 1978-1989.

Potts, J. D., & Runyan, R. B. (1989). Epithelial-mesenchymal cell transformation in the embryonic heart can be mediated, in part, by transforming growth factor beta. *Dev Biol, 134*(2), 392-401. doi:10.1016/0012-1606(89)90111-5

Pretzsch, E., Bösch, F., Neumann, J., Ganschow, P., Bazhin, A., Guba, M., . . . Angele, M. (2019). Mechanisms of Metastasis in Colorectal Cancer and Metastatic Organotropism: Hematogenous versus Peritoneal Spread. *J Oncol, 2019*, 7407190. doi:10.1155/2019/7407190

Quinn, J. J., Jones, M. G., Okimoto, R. A., Nanjo, S., Chan, M. M., Yosef, N., . . . Weissman, J. S. (2021). Single-cell lineages reveal the rates, routes, and drivers of metastasis in cancer xenografts. *Science, 371*(6532). doi:10.1126/science.abc1944

Radaelli, E., Santagostino, S. F., Sellers, R. S., & Brayton, C. F. (2018). Immune Relevant and Immune Deficient Mice: Options and Opportunities in Translational Research. *Ilar j, 59*(3), 211-246. doi:10.1093/ilar/ily026

Ramalingam, S., Daughtridge, G. W., Johnston, M. J., Gracz, A. D., & Magness, S. T. (2012). Distinct levels of Sox9 expression mark colon epithelial stem cells that form colonoids in culture. *Am J Physiol Gastrointest Liver Physiol, 302*(1), G10-20. doi:10.1152/ajpgi.00277.2011

Ramón, Y. C. S., Sesé, M., Capdevila, C., Aasen, T., De Mattos-Arruda, L., Diaz-Cano, S. J., . . . Castellví, J. (2020). Clinical implications of intratumor heterogeneity: challenges and opportunities. *J Mol Med (Berl), 98*(2), 161-177. doi:10.1007/s00109-020-01874-2

Recasens, A., & Munoz, L. (2019). Targeting Cancer Cell Dormancy. *Trends Pharmacol Sci, 40*(2), 128-141. doi:10.1016/j.tips.2018.12.004

Rehman, S. K., Haynes, J., Collignon, E., Brown, K. R., Wang, Y., Nixon, A. M. L., . . . O'Brien, C. A. (2021). Colorectal Cancer Cells Enter a Diapause-like DTP State to Survive Chemotherapy. *Cell, 184*(1), 226-242.e221. doi:10.1016/j.cell.2020.11.018

Revenco, T., Nicodème, A., Pastushenko, I., Sznurkowska, M. K., Latil, M., Sotiropoulou, P. A., . . . Blanpain, C. (2019). Context Dependency of Epithelial-to-Mesenchymal Transition for Metastasis. *Cell Rep, 29*(6), 1458-1468.e1453. doi:10.1016/j.celrep.2019.09.081

Ricci-Vitiani, L., Lombardi, D. G., Pilozzi, E., Biffoni, M., Todaro, M., Peschle, C., & De Maria, R. (2007). Identification and expansion of human colon-cancer-initiating cells. *Nature, 445*(7123), 111-115. doi:10.1038/nature05384

Rios-Esteves, J., & Resh, M. D. (2013). Stearoyl CoA desaturase is required to produce active, lipid-modified Wnt proteins. *Cell Rep, 4*(6), 1072-1081. doi:10.1016/j.celrep.2013.08.027

Rizzo, G., Bertotti, A., Leto, S. M., & Vetrano, S. (2021). Patient-derived tumor models: a more suitable tool for pre-clinical studies in colorectal cancer. *J Exp Clin Cancer Res, 40*(1), 178. doi:10.1186/s13046-021-01970-2

Rodger, E. J., Gimenez, G., Ajithkumar, P., Stockwell, P. A., Almomani, S., Bowden, S. A., . . . Chatterjee, A. (2023). An epigenetic signature of advanced colorectal cancer metastasis. *iScience, 26*(6), 106986. doi:10.1016/j.isci.2023.106986

Roerink, S. F., Sasaki, N., Lee-Six, H., Young, M. D., Alexandrov, L. B., Behjati, S., . . . Clevers, H. (2018). Intra-tumour diversification in colorectal cancer at the single-cell level. *Nature, 556*(7702), 457-462. doi:10.1038/s41586-018-0024-3

Roukens, M. G., Frederiks, C. L., Seinstra, D., Braccioli, L., Khalil, A. A., Pals, C., . . . Coffer, P. J. (2021). Regulation of a progenitor gene program by SOX4 is essential for mammary tumor proliferation. *Oncogene, 40*(45), 6343-6353. doi:10.1038/s41388-021-02004-z

Ruoslahti, E. (2004). Vascular zip codes in angiogenesis and metastasis. *Biochem Soc Trans, 32*(Pt3), 397-402. doi:10.1042/bst0320397

Ryser, M. D., Yu, M., Grady, W., Siegmund, K., & Shibata, D. (2018). Epigenetic Heterogeneity in Human Colorectal Tumors Reveals Preferential Conservation And Evidence of Immune Surveillance. *Sci Rep, 8*(1), 17292. doi:10.1038/s41598-018-35621-y

Sacchetti, A., Teeuwssen, M., Verhagen, M., Joosten, R., Xu, T., Stabile, R., . . . Fodde, R. (2021). Phenotypic plasticity underlies local invasion and distant metastasis in colon cancer. *Elife, 10*. doi:10.7554/eLife.61461

Sasaki, N., & Clevers, H. (2018). Studying cellular heterogeneity and drug sensitivity in colorectal cancer using organoid technology. *Curr Opin Genet Dev, 52*, 117-122. doi:10.1016/j.gde.2018.09.001

Sasaki, N., Sachs, N., Wiebrands, K., Ellenbroek, S. I., Fumagalli, A., Lyubimova, A., . . . Clevers, H. (2016). Reg4+ deep crypt secretory cells function as epithelial niche for Lgr5+ stem cells in colon. *Proc Natl Acad Sci U S A, 113*(37), E5399-5407. doi:10.1073/pnas.1607327113

Sato, T., Stange, D. E., Ferrante, M., Vries, R. G., Van Es, J. H., Van den Brink, S., . . . Clevers, H. (2011). Long-term expansion of epithelial organoids from human colon, adenoma, adenocarcinoma, and Barrett's epithelium. *Gastroenterology, 141*(5), 1762-1772. doi:10.1053/j.gastro.2011.07.050

Sato, T., Vries, R. G., Snippert, H. J., van de Wetering, M., Barker, N., Stange, D. E., . . . Clevers, H. (2009). Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche. *Nature, 459*(7244), 262-265. doi:10.1038/nature07935

Sawant, K. V., Poluri, K. M., Dutta, A. K., Sepuru, K. M., Troshkina, A., Garofalo, R. P., & Rajarathnam, K. (2016). Chemokine CXCL1 mediated neutrophil recruitment: Role of glycosaminoglycan interactions. *Sci Rep, 6*, 33123. doi:10.1038/srep33123

Schmitt, M., & Pawlita, M. (2009). High-throughput detection and multiplex identification of cell contaminations. *Nucleic Acids Res, 37*(18), e119. doi:10.1093/nar/gkp581

Schnekenburger, M., Karius, T., & Diederich, M. (2014). Regulation of epigenetic traits of the glutathione S-transferase P1 gene: from detoxification toward cancer prevention and diagnosis. *Front Pharmacol, 5*, 170. doi:10.3389/fphar.2014.00170

Schubert, M., Klinger, B., Klünemann, M., Sieber, A., Uhlitz, F., Sauer, S., . . . Saez-Rodriguez, J. (2018). Perturbation-response genes reveal signaling footprints in cancer gene expression. *Nat Commun, 9*(1), 20. doi:10.1038/s41467-017-02391-6

Serrano-Gomez, S. J., Maziveyi, M., & Alahari, S. K. (2016). Regulation of epithelial-mesenchymal transition through epigenetic and post-translational modifications. *Mol Cancer, 15*, 18. doi:10.1186/s12943-016-0502-x

Shaffer, S. M., Dunagin, M. C., Torborg, S. R., Torre, E. A., Emert, B., Krepler, C., . . . Raj, A. (2017). Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance. *Nature, 546*(7658), 431-435. doi:10.1038/nature22794

Sharma, S. V., Lee, D. Y., Li, B., Quinlan, M. P., Takahashi, F., Maheswaran, S., . . . Settleman, J. (2010). A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. *Cell, 141*(1), 69-80. doi:10.1016/j.cell.2010.02.027

Shimokawa, M., Ohta, Y., Nishikori, S., Matano, M., Takano, A., Fujii, M., . . . Sato, T. (2017). Visualization and targeting of LGR5(+) human colon cancer stem cells. *Nature, 545*(7653), 187-192. doi:10.1038/nature22081

Shin, A. E., Giancotti, F. G., & Rustgi, A. K. (2023). Metastatic colorectal cancer: mechanisms and emerging therapeutics. *Trends Pharmacol Sci, 44*(4), 222-236. doi:10.1016/j.tips.2023.01.003

Shivdasani, R. A., Clevers, H., & de Sauvage, F. J. (2021). Tissue regeneration: Reserve or reverse? *Science, 371*(6531), 784-786. doi:10.1126/science.abb6848

Si, X., Liu, Y., Lv, J., Ding, H., Zhang, X. A., Shao, L., . . . Sun, Y. (2016). ERα propelled aberrant global DNA hypermethylation by activating the DNMT1 gene to enhance anticancer drug resistance in human breast cancer cells. *Oncotarget, 7*(15), 20966-20980. doi:10.18632/oncotarget.8038

Simeonov, K. P., Byrns, C. N., Clark, M. L., Norgard, R. J., Martin, B., Stanger, B. Z., . . . Lengner, C. J. (2021). Single-cell lineage tracing of metastatic cancer reveals selection of hybrid EMT states. *Cancer Cell, 39*(8), 1150-1162.e1159. doi:10.1016/j.ccell.2021.05.005

Siraj, S., Masoodi, T., Siraj, A. K., Azam, S., Qadri, Z., Ahmed, S. O., . . . Al-Kuraya, K. S. (2020). Clonal Evolution and Timing of Metastatic Colorectal Cancer. *Cancers (Basel), 12*(10). doi:10.3390/cancers12102938

Sobolev, V. V., Khashukoeva, A. Z., Evina, O. E., Geppe, N. A., Chebysheva, S. N., Korsunskaya, I. M., . . . Mezentsev, A. (2022). Role of the Transcription Factor FOSL1 in Organ Development and Tumorigenesis. *Int J Mol Sci, 23*(3). doi:10.3390/ijms23031521

Sobral, D., Martins, M., Kaplan, S., Golkaram, M., Salmans, M., Khan, N., . . . Costa, L. (2022). Genetic and microenvironmental intra-tumor heterogeneity impacts colorectal cancer evolution and metastatic development. *Commun Biol, 5*(1), 937. doi:10.1038/s42003-022-03884-x

Song, Y., Gao, F., Peng, Y., & Yang, X. (2020). Long non-coding RNA DBH-AS1 promotes cancer progression in diffuse large B-cell lymphoma by targeting FN1 via RNA-binding protein BUD13. *Cell Biol Int, 44*(6), 1331-1340. doi:10.1002/cbin.11327

Sousa-Ortega, A., Vázquez-Marín, J., Sanabria-Reinoso, E., Corbacho, J., Polvillo, R., Campoy-López, A., . . . Martínez-Morales, J. R. (2023). A Yap-dependent mechanoregulatory program sustains cell migration for embryo axis assembly. *Nat Commun, 14*(1), 2804. doi:10.1038/s41467-023-38482-w

Stanta, G., & Bonin, S. (2018). Overview on Clinical Relevance of Intra-Tumor Heterogeneity. *Front Med (Lausanne), 5*, 85. doi:10.3389/fmed.2018.00085

Stintzing, S., Tejpar, S., Gibbs, P., Thiebach, L., & Lenz, H. J. (2017). Understanding the role of primary tumour localisation in colorectal cancer treatment and outcomes. *Eur J Cancer, 84*, 69-80. doi:10.1016/j.ejca.2017.07.016

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., . . . Mesirov, J. P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A, 102*(43), 15545-15550. doi:10.1073/pnas.0506580102

Sun, X. X., & Yu, Q. (2015). Intra-tumor heterogeneity of cancer cells and its implications for cancer treatment. *Acta Pharmacol Sin, 36*(10), 1219-1227. doi:10.1038/aps.2015.92

Szebényi, K., Füredi, A., Bajtai, E., Sama, S. N., Csiszar, A., Gombos, B., . . . Szakács, G. (2023). Effective targeting of breast cancer by the inhibition of P-glycoprotein mediated removal of toxic lipid peroxidation byproducts from drug tolerant persister cells. *Drug Resist Updat, 71*, 101007. doi:10.1016/j.drup.2023.101007

Tang, D., Kang, R., Zeh, H. J., 3rd, & Lotze, M. T. (2010). High-mobility group box 1 and cancer. *Biochim Biophys Acta, 1799*(1-2), 131-140. doi:10.1016/j.bbagrm.2009.11.014

Tejera, P., O'Mahony, D. S., Owen, C. A., Wei, Y., Wang, Z., Gupta, K., . . . Christiani, D. C. (2014). Functional characterization of polymorphisms in the peptidase inhibitor 3 (elafin) gene and

validation of their contribution to risk of acute respiratory distress syndrome. *Am J Respir Cell Mol Biol, 51*(2), 262-272. doi:10.1165/rcmb.2013-0238OC

Tian, H., Biehs, B., Warming, S., Leong, K. G., Rangell, L., Klein, O. D., & de Sauvage, F. J. (2011). A reserve stem cell population in small intestine renders Lgr5-positive cells dispensable. *Nature, 478*(7368), 255-259. doi:10.1038/nature10408

Tian, J., Chen, J. H., Chao, S. X., Pelka, K., Giannakis, M., Hess, J., . . . Corcoran, R. B. (2023). Combined PD-1, BRAF and MEK inhibition in BRAF(V600E) colorectal cancer: a phase 2 trial. *Nat Med, 29*(2), 458-466. doi:10.1038/s41591-022-02181-8

Tsai, J. H., & Yang, J. (2013). Epithelial-mesenchymal plasticity in carcinoma metastasis. *Genes Dev, 27*(20), 2192-2206. doi:10.1101/gad.225334.113

Tsompana, M., & Buck, M. J. (2014). Chromatin accessibility: a window into the genome. *Epigenetics Chromatin, 7*(1), 33. doi:10.1186/1756-8935-7-33

Turdo, A., Gaggianesi, M., Di Franco, S., Veschi, V., D'Accardo, C., Porcelli, G., . . . Todaro, M. (2022). Effective targeting of breast cancer stem cells by combined inhibition of Sam68 and Rad51. *Oncogene, 41*(15), 2196-2209. doi:10.1038/s41388-022-02239-4

Uhlitz, F., Bischoff, P., Peidli, S., Sieber, A., Trinks, A., Lüthen, M., . . . Morkel, M. (2021). Mitogen-activated protein kinase activity drives cell trajectories in colorectal cancer. *EMBO Mol Med, 13*(10), e14123. doi:10.15252/emmm.202114123

Valastyan, S., & Weinberg, R. A. (2011). Tumor metastasis: molecular insights and evolving paradigms. *Cell, 147*(2), 275-292. doi:10.1016/j.cell.2011.09.024

Van Cutsem, E., Köhne, C. H., Hitre, E., Zaluski, J., Chang Chien, C. R., Makhson, A., . . . Rougier, P. (2009). Cetuximab and chemotherapy as initial treatment for metastatic colorectal cancer. *N Engl J Med, 360*(14), 1408-1417. doi:10.1056/NEJMoa0805019

van der Heijden, M., Miedema, D. M., Waclaw, B., Veenstra, V. L., Lecca, M. C., Nijman, L. E., . . . Vermeulen, L. (2019). Spatiotemporal regulation of clonogenicity in colorectal cancer xenografts. *Proc Natl Acad Sci U S A, 116*(13), 6140-6145. doi:10.1073/pnas.1813417116

Vanharanta, S., & Massagué, J. (2013). Origins of metastatic traits. *Cancer Cell, 24*(4), 410-421. doi:10.1016/j.ccr.2013.09.007

Vasan, N., Razavi, P., Johnson, J. L., Shao, H., Shah, H., Antoine, A., . . . Baselga, J. (2019). Double PIK3CA mutations in cis increase oncogenicity and sensitivity to PI3Kα inhibitors. *Science, 366*(6466), 714-723. doi:10.1126/science.aaw9032

Vogelstein, B., Fearon, E. R., Hamilton, S. R., Kern, S. E., Preisinger, A. C., Leppert, M., . . . Bos, J. L. (1988). Genetic alterations during colorectal-tumor development. *N Engl J Med, 319*(9), 525-532. doi:10.1056/nejm198809013190901

Wang, H., Song, X., Huang, Q., Xu, T., Yun, D., Wang, Y., . . . Chen, J. (2019). LGALS3 Promotes Treatment Resistance in Glioblastoma and Is Associated with Tumor Risk and Prognosis. *Cancer Epidemiol Biomarkers Prev, 28*(4), 760-769. doi:10.1158/1055-9965.Epi-18-0638

Wang, X., Li, D., Sun, L., Shen, G., Liu, H., Guo, H., . . . Lu, Y. (2020). Regulation of the small GTPase Ran by miR-802 modulates proliferation and metastasis in colorectal cancer cells. *Br J Cancer, 122*(11), 1695-1706. doi:10.1038/s41416-020-0809-7

Wang, Y., Song, W., Wang, J., Wang, T., Xiong, X., Qi, Z., . . . Chen, Y. G. (2020). Single-cell transcriptome analysis reveals differential nutrient absorption functions in human intestine. *J Exp Med, 217*(2). doi:10.1084/jem.20191130

Wei, Z., Ma, W., Qi, X., Zhu, X., Wang, Y., Xu, Z., . . . Li, G. (2016). Pinin facilitated proliferation and metastasis of colorectal cancer through activating EGFR/ERK signaling pathway. *Oncotarget, 7*(20), 29429-29439. doi:10.18632/oncotarget.8738

Weiler, P., Van den Berge, K., Street, K., & Tiberi, S. (2023). A Guide to Trajectory Inference and RNA Velocity. *Methods Mol Biol, 2584*, 269-292. doi:10.1007/978-1-0716-2756-3_14

Weis, S. M., & Cheresh, D. A. (2005). Pathophysiological consequences of VEGF-induced vascular permeability. *Nature, 437*(7058), 497-504. doi:10.1038/nature03987

Willet, S. G., Lewis, M. A., Miao, Z. F., Liu, D., Radyk, M. D., Cunningham, R. L., . . . Mills, J. C. (2018). Regenerative proliferation of differentiated cells by mTORC1-dependent paligenosis. *Embo j, 37*(7). doi:10.15252/embj.201798311

Yachida, S., Jones, S., Bozic, I., Antal, T., Leary, R., Fu, B., . . . Iacobuzio-Donahue, C. A. (2010). Distant metastasis occurs late during the genetic evolution of pancreatic cancer. *Nature, 467*(7319), 1114-1117. doi:10.1038/nature09515

Yang, J., Antin, P., Berx, G., Blanpain, C., Brabletz, T., Bronner, M., . . . Sheng, G. (2020). Guidelines and definitions for research on epithelial-mesenchymal transition. *Nat Rev Mol Cell Biol, 21*(6), 341-352. doi:10.1038/s41580-020-0237-9

Yu, T., Gao, X., Zheng, Z., Zhao, X., Zhang, S., Li, C., & Liu, G. (2021). Intratumor Heterogeneity as a Prognostic Factor in Solid Tumors: A Systematic Review and Meta-Analysis. *Front Oncol, 11*, 744064. doi:10.3389/fonc.2021.744064

Yui, S., Azzolin, L., Maimets, M., Pedersen, M. T., Fordham, R. P., Hansen, S. L., . . . Jensen, K. B. (2018). YAP/TAZ-Dependent Reprogramming of Colonic Epithelium Links ECM Remodeling to Tissue Regeneration. *Cell Stem Cell, 22*(1), 35-49.e37. doi:10.1016/j.stem.2017.11.001

Yun, C. H., Mengwasser, K. E., Toms, A. V., Woo, M. S., Greulich, H., Wong, K. K., . . . Eck, M. J. (2008). The T790M mutation in EGFR kinase causes drug resistance by increasing the affinity for ATP. *Proc Natl Acad Sci U S A, 105*(6), 2070-2075. doi:10.1073/pnas.0709662105

Zeuner, A., Todaro, M., Stassi, G., & De Maria, R. (2014). Colorectal cancer stem cells: from the crypt to the clinic. *Cell Stem Cell, 15*(6), 692-705. doi:10.1016/j.stem.2014.11.012

Zhang, Z., Qin, S., Chen, Y., Zhou, L., Yang, M., Tang, Y., . . . Wei, X. (2022). Inhibition of NPC1L1 disrupts adaptive responses of drug-tolerant persister cells to chemotherapy. *EMBO Mol Med, 14*(2), e14903. doi:10.15252/emmm.202114903

Zhao, G., Wang, Q., Zhang, Y., Gu, R., Liu, M., Li, Q., . . . Lin, P. (2023). DDX17 induces epithelial-mesenchymal transition and metastasis through the miR-149-3p/CYBRD1 pathway in colorectal cancer. *Cell Death Dis, 14*(1), 1. doi:10.1038/s41419-022-05508-y

Zhao, W., Dai, S., Yue, L., Xu, F., Gu, J., Dai, X., & Qian, X. (2022). Emerging mechanisms progress of colorectal cancer liver metastasis. *Front Endocrinol (Lausanne), 13*, 1081585. doi:10.3389/fendo.2022.1081585

Zhao, Y., Wang, H., Zhou, J., & Shao, Q. (2022). Glutathione Peroxidase GPX1 and Its Dichotomous Roles in Cancer. *Cancers (Basel), 14*(10). doi:10.3390/cancers14102560

Zheng, X., Carstens, J. L., Kim, J., Scheible, M., Kaye, J., Sugimoto, H., . . . Kalluri, R. (2015). Epithelial-to-mesenchymal transition is dispensable for metastasis but induces chemoresistance in pancreatic cancer. *Nature, 527*(7579), 525-530. doi:10.1038/nature16064

Zheng, Z., Yu, T., Zhao, X., Gao, X., Zhao, Y., & Liu, G. (2020). Intratumor heterogeneity: A new perspective on colorectal cancer research. *Cancer Med, 9*(20), 7637-7645. doi:10.1002/cam4.3323

Zhou, X., Liu, S., Cai, G., Kong, L., Zhang, T., Ren, Y., . . . Wang, X. (2015). Long Non Coding RNA MALAT1 Promotes Tumor Growth and Metastasis by inducing Epithelial-Mesenchymal Transition in Oral Squamous Cell Carcinoma. *Sci Rep, 5*, 15972. doi:10.1038/srep15972

Zowada, M. K., Tirier, S. M., Dieter, S. M., Krieger, T. G., Oberlack, A., Chua, R. L., . . . Ball, C. R. (2021). Functional States in Tumor-Initiating Cell Differentiation in Human Colorectal Cancer. *Cancers (Basel), 13*(5). doi:10.3390/cancers13051097

8. List of main figures

9. Acknowledgements

First of all, I would like to thank Prof. Dr. Hanno Glimm for giving me the opportunity to work on this project in his group at DKFZ/NCT Heidelberg. Thank you for your guidance and advice throughout the 5 years. I highly appreciate and admire your approach, your thoughtful comments and suggestions. I learned a lot from you, and I hope to collaborate with you in the future.

A big thank you to Dr. Claudia R. Ball who also gave me the opportunity to work in this group and who supervised me in this project. Thank you for the continuous support and for always being available for discussions and advice. I really appreciate all the effort you put into driving this project forward until the thesis submission. I wish it was more in person.

I would like to thank Dr. Friederike Herbst for joining in on this project and for providing her scientific expertise, for the everyday support and for the valuable scientific discussions, in addition to offering another perspective even from abroad. Thank you also for always having an open door and for your kindness.

Next, I would like to thank Dr. Sebastian Dieter who supported in supervising the project in the last two years.

I would like to thank Prof. Dr. Stefan Wiemann and Prof. Dr. Jan Lohmann for being members of my thesis advisory committee, in addition to Prof. Peter Angel for taking part in my thesis examination committee.

Special thanks for Dr. Attila Jády, who supported the bioinformatic analysis of the data produced in this project, for our long discussions and for being a great colleague. Thanks for bearing with me even when I asked for more.

I was lucky to have great colleagues, starting with Martina K. Zowada. I am so grateful that we could work together during the last years. Thank you for always being there to support in the lab and in scientific discussions. Thanks for the fun we had together in and outside the lab as friends. I wish I could make it more to the volleyball training.

Stella, it was and still is a great pleasure to have you as a colleague, thanks for your support and for being there in good and also challenging times, and of course in experimental lab support.

Thank you Tim Kindinger for being available for assistance since the first day, and especially in mice experiments where I learned a lot. Thanks Karin for your support in organoid cultures and the fun we had together.

Same goes for Sylvia Martin, who is a nice person that was always available for technical support but also for having a talk nice talk with. I appreciate a lot your help and getting to know you.

A special thank you goes to Pravin Velmurugan who supported additional aspects of this project during his master internship. It was a pleasure working with you and I appreciate your help a lot.

Thanks for taking the time to support my project. It was great to get to know you Michael, thanks for being a great colleague and office buddy.

Thank you Dr. Mario Huerta for supporting the early planning of this project in the very early steps and for always bringing fun with you. Thanks to my good colleagues Maximillian, Constantin and my previous colleagues Na, Tabea, Nina, Ceren and Mariam. Anna-Maria, thanks for the nice time together.

I also want to thank all my colleagues from Dresden, especially Sara Hrabovska and Dr. Zunamys Carrero for being nice colleagues. I greatly acknowledge colleagues who supported this project and were a pleasure to work with Dr. Alexander Wurm and Vida Kufrin.

Thanks to our Azubis Hannah Walzer, Celine Reifenberg, Leon Wambach, Ina Siebig, and Dylan Knopf, but especially Florian Heeg for their technical assistance and for the great time we had. Thank you for always being available and offering your help whenever needed.

I would also like to acknowledge the administrative support especially from Julia Dorok in addition to Yasmin Führer, Julia Poenisch and Dr. Daniela Richter.

I would like to thank the Core Facility of DKFZ for their technical support and excellent services.

A big thanks goes especially to Dr. Jan-Philipp Mallm for his support in the project and for all the scOpen lab team especially Katharina, Denise, Pooja and Michelle. I have spent a lot of time at your lab and it felt like my second lab group.

Thanks for the cancer patients who provided their tumor samples, allowing this research to happen.

Special thanks for the animal laboratory technicians and Vets who helped me in maintaining the mice. I hope experiments would be possible to do without animals one day.

The graduate office are taking good care of PhD students, thank you, especially Lindsay for your support.

I would like to thank the members of the group of Prof. Stefan Frohling and prof. Claudia Scholl for welcoming us in their group and for the great feedback to my project and for the scientific discussions.

Dr. Daniel Mertens, thanks for your great advice great mentor, for giving a part of your time to support me to advance with my career.

Julian, it was great knowing you, I learned a lot from and it was nice to speak to you or have a coffee together to talk about many of our outlooks. It helped me think much clearer in some moments where I needed it.

For my friends of the pumpkin group, Enrico, Menna, Julia, Simon, Carlo and Eirinni, we had a lot of fun together and Heidelberg was much more fun and easy with you. Thomas, I wish we had more lunches together but it was great to see you every now and then, especially in basketball matches. Thank you Isabella for sharing a good time together and thank you Federica for the nice coffee breaks.

Lukas, you have been a great friend and I had a lot of fun in your company, Thursday lunches were the highlight of many of my weeks.

Dani, we have been a team since the beginning until the end and your friendship was one of the greatest things I got out of my PhD journey. I am happy to know you and your family, which is now mine too. Thanks for the support, the trips and all the fun we had together.

Paola, I cannot thank you enough for your support throughout these last five years and especially through writing my thesis. Your company has made everything look better, thank you for being there, you are fundamental to me.

At the end, I want to thank all my family, especially my father, mother and brothers for their endless support. I cannot thank you enough for everything that you have done for me. I would not be who I am without you.