

AUS DEM DEUTSCHEN KREBSFORSCHUNGSZENTRUM
UND DEM LEHRSTUHL FÜR ZELL- UND MOLEKULARBIOLOGIE
DER MEDIZINISCHEN FAKULTÄT MANNHEIM
LEITUNG: PROF. DR. MICHAEL BOUTROS

Multiparametric phenotyping of intestinal organoids to model disease initiation and treatment response in colorectal cancer

INAUGURALDISSERTATION

ZUR ERLANGUNG DES MEDIZINISCHEN DOKTORGRADES
AN DER
MEDIZINISCHEN FAKULTÄT HEIDELBERG DER
RUPRECHT-KARLS-UNIVERSITÄT

VORGELEGT VON NIKLAS TIMON RINDTORFF
AUS TÜBINGEN

2022

DEKAN: PROF. DR. ANDREAS DRAGUHN
DOKTORVATER: PROF. DR. MICHAEL BOUTROS

I DEDICATE THIS THESIS TO ANDREA AND KLAUS RINDTORFF AS WELL AS SOPHIE WEIL. YOU
SUPPORTED ME THROUGH THE DARKEST AND BRIGHTEST HOURS OF THIS JOURNEY.

FIGURES WITHIN THIS THESIS WERE CREATED USING BIORENDER.COM

Contents

1	INTRODUCTION	12
1.1	Disclosure	12
1.2	The Colon	13
1.3	Colorectal Cancer	16
1.4	Colorectal Cancer Therapy	20
1.5	Cancer Drug Discovery	21
1.6	Multi-View Representation Learning in Biology	22
1.7	Towards causal Representation Learning in biology	22
1.8	Aims of Thesis	22
2	MATERIALS AND METHODS	26
2.1	Patients	26
2.2	Organoid Culture	27
2.3	Biochemical assays	29
2.4	Organoid profiling	32
2.5	Image analysis	35
3	IMAGE-BASED PROFILING IDENTIFIES MOLECULAR DETERMINANTS OF CANCER ORGANOID ARCHITECTURE AND PLASTICITY	38
3.1	Disclosure	39
3.2	Establishing patient derived organoids for image-based profiling	39
3.3	Enabling methods for high-throughput image-based profiling of organoids	42
3.4	Image-based profiling captures the morphological diversity of patient-derived cancer organoids	44
3.5	Organoid phenotype-profiles capture organoid viability	49
3.6	Drug induced organoid phenotypes correspond to drug mechanism of action	52
3.7	Multi-omics factor analysis identifies shared factors linking morphology, genomic data and drug activity	54
3.8	An IGF1R signaling program is associated with increased organoid size, decreased EGFR inhibitor activity and can be induced by mTOR inhibition	55
3.9	An LGR5+ stemness program is associated with cystic organoid architecture and can be induced by inhibition of MEK	57
4	IMAGE-BASED PROFILING TO IDENTIFY REVERTANT THERAPEUTICS IN PRE-MALIGNANT MODELS OF COLON CANCER	68
4.1	Motivation	68
4.2	Generation of organoid colon adenoma models	70
4.3	Biochemical profiling of organoid models	73

4.4	Image-based profiling of organoid models	77
4.5	Quantifying small molecule induced phenotypes across organoid models	79
4.6	Multi-omics factor analysis identifies shared factors linking functional and structural biological views	81
4.7	A canonical Wnt signaling associated program caused by Apc loss	82
4.8	An oncogene-induced senescence program caused by isolated KrasG12D activation	88
4.9	oncogenic Kras leads to increased mTOR signaling conditioned on loss of Apc function	91
4.10	Ellagic acid is a candidate small molecule revertant of organoid adenoma states	93
5	DISCUSSION	94
	APPENDIX A SOME EXTRA STUFF	iv
	REFERENCES	vi

Listing of figures

1.1	Colon Cancer Progression	17
1.2	18
1.3	24
3.1	Central liquid handling methods a Organoid isolation procedure. Colorectal cancer tissue biopsies were collected via endoscopy, enzymatically removed from extracellular matrix proteins, washed and resuspended in basal membrane extract hydrogel. After solidification of hydrogel domes, organoids were overlayed with growth factor rich culture medium. b Organoid high-throughput experimentation. Colorectal cancer organoids were harvested, partially digested, seeded in hydrogel-coated 384-well plates	40
3.2	Organoid cohort overview a Tumor location (right/left/rectum) and AJCC/UICC stage of colorectal cancers that patient derived organoids were derived from. b Consensus molecular subtypes of organoids determined by RNA expression analysis. c Mutation status in PDOs, as analyzed by amplicon sequencing. Figure created with support from Johannes Betge (graphical presentation), Erica Valentini (sequencing data analysis) and Benedikt Rauscher (CMS type inference)	41
3.3	Overview of experiments a Organoids were isolated from endoscopic biopsies from patients with colorectal cancer. Organoids were dissociated and evenly seeded in 384-well plates before perturbation with an experimental (464 compounds) and a clinical compound library (63 compounds à 5 concentrations, 842 perturbations across both libraries). After treatment, high-throughput fluorescence microscopy was used to capture the morphology of organoids. The multi-channel (DNA, beta-actin, cell permeability) 3D imaging data was projected, segmented, and descriptive features were extracted to quantify potential drug-induced phenotypes. Untreated organoid morphology, organoid size and drug activity scores were integrated with mRNA expression and mutation data in a Multi-Omics Factor Analysis (MOFA) to increase interpretability of organoid variation. Figure created with support from Johannes Betge (graphical presentation)	42
3.4	Image-based profiling method a The image-processing pipeline illustrated with representative example images from 2 organoid lines: Organoids were imaged at multiple layers along the z-axis. Images were projected using a maximum contrast projection and segmented using a convolutional neural network, both designed and implemented by Jan Sauer. Descriptive features were extracted from all three channels to quantify phenotypes. Feature plots show the median phenotype of unperturbed organoids, six example features (Area, Phalloidin intensity, DAPI intensity, FITC intensity, FITC Haralick angular second moment (ASM) and FITC intensity 1-percentile) and their z-scores relative to all profiled organoid lines are shown. Figure created with support from Jan Sauer (data processing) and Johannes Betge (graphical presentation)	43

3.5	Dataset dimensions and compound library overview a Number of organoid models and number of perturbations in previous publications reporting high-throughput drug screenings with patient derived cancer organoids, b Graphical representation of the compound libraries used for drug screening in this project: A library targeting kinases and stem cell pathways (KiStem library, 464 compounds) and a clinical library with 63 drugs in 5 concentrations. Figure created with support from Johannes Betge (graphical presentation)	44
3.6	Image-based profiling captures the phenotype diversity of patient derived cancer organoids a Uniform Manifold Approximation and Projection (UMAP) of organoid-level features for a random 5% sample out of ca. 5.5 million organoids. The same sample is used for visualizations throughout the figure. Color corresponds to the log-scaled organoid area (dark blue: minimum size, yellow: maximum size). b organoid size distribution across lines. c UMAP representation of DMSO treated and drug treated organoids. Graph-based clustering of organoids by morphology. d UMAP embeddings of selected organoid lines (baseline state / 0.1% DMSO control-treated organoids) representing different morphological subsets, grey background consists of randomly sampled points. Depicted are representative example images for each line (right, cyan = DNA, magenta = actin, scale-bar: 200µm).	46
3.7	Basic image-based features and their role in organoid phenotype diversity. a-c Uniform Manifold Approximation and Projection (UMAP) of organoid-level features marked by DNA (DAPI) staining intensity (b), actin (Phalloid/FITC) staining intensity (c) and permeability (DeadGreen) staining intensity d Distribution of organoid size for all control (DMSO) treated organoids. e Replicate correlation of organoid size for control treated organoids. f UMAP representation of DMSO treated and drug treated organoids	47
3.8	Technical confounders and their impact on organoid phenotype. a UMAP of organoid level features stratified by organoid line and colored by experimental batch.	48
3.9	Organoid phenotype-profiles capture organoid viability. a Representative example images of negative-(0.1% DMSO) and positive control treated organoids (2.5µM bortezomib, cyan = DNA, magenta = actin, yellow = cell permeability; average images were selected and embedded in black background; scale bar: 50µm). b, Dose-dependent-trajectory of bortezomib drug effect. UMAP of organoid morphology at different bortezomib doses and (right panel) dose-dependent trajectory for three representative organoid lines. For visual purposes, trajectory inference was limited to partition 1, the left-hand set of measurements within the UMAP, representing ca. 95 % of all imaging data.	49
3.10	Organoid phenotype-profiles capture treatment specific changes in organoid viability. a Distribution of organoid size at different concentrations of paclitaxel. Shown is a random sample of 30% of all paclitaxel treated organoids for this and following figures. b Distribution of organoid number per well at different concentrations of paclitaxel. c Example images of Do22T organoids treated with paclitaxel. d Dose-response relationship of organoid size and paclitaxel dose. Do22T and Do46T are highlighted. e UMAP of organoid morphology highlighting Do22T organoids treated at different concentrations of paclitaxel. f UMAP of organoid morphology highlighting Do46T organoids treated at different concentrations of paclitaxel.	51

3.11	Organoid phenotype-profiles reflect ATP-dependent viability measurements. a Association of organoid size of selected example lines with organoid viability determined by luminescence-based, ATP-dependent viability profiling with CellTiter-Glo (CTG), which was performed in parallel with imaging on a subset of drug treatments. b UMAP visualization of dead organoids within our dataset, based on supervised machine learning of organoid viability using classifiers trained on positive- (high-dose bortezomib and SN-38) and negative (DMSO) controls (live-dead classifiers, LDC). c Association of organoid size of selected example lines with organoid viability determined by luminescence-based, ATP-dependent viability profiling with CellTiter-Glo (CTG), which was performed in parallel with imaging on a subset of drug treatments for benchmarking. d UMAP visualization of dead organoids within our dataset, based on supervised machine learning of organoid viability using classifiers trained on positive- (high-dose bortezomib and SN-38) and negative (DMSO) controls (live-dead classifiers, LDC). e Association of LDC and example organoid features (size, DAPI, actin and permeability dye intensities) with benchmark CTG viability read out. Figure created with support from Jan Sauer (LDC classifier training)	52
3.12	53
3.13	58
3.14	59
3.15	60
3.16	60
3.17	61
3.18	61
3.19	62
3.20	63
3.21	64
3.22	65
3.23	66
4.1	Visual abstract of adenoma model profiling project.	69
4.2	Establishing organoid models of colon adenoma, a Overview of organoid model establishment. Mouse colon organoids were isolated from a transgenic donor animal carrying an inactive conditional oncogenic KrasG12D allele. Homozygous truncation of Apc via CRISPR and activation of the heterozygous KrasG12D allele lead to four different genetically defined organoid models. b In vitro growth factor dependency of adenoma models. Organoids were cultured in complete or modified medium containing combinations of Wnt3A, R-Spondin1-Fc and EGF for 120h and subsequently imaged. Scalebar = 200um. c Oncogenic KrasG12D increases resistance to Egfr inhibition. Organoid ATP levels were measured 4 days after Gefitinib treatment and adjusted for organoid growth rate. Points represent mean of n=2 independent experiments. Error bars represent standard error of mean. d Erk phosphorylation is increased by oncogenic KrasG12D. Organoid models were cultured with or without Wnt3A and R-Spondin1-Fc for 72h and analyzed for protein levels. p, phospho. e Loss of Apc induces transcription of canonical Wnt-signaling target genes. qRT-PCR for Axin2 and Ccnd in the presence or absence of Wnt 3a and R-spondin1-Fc after 120h of culture. Expression levels are normalized to Sdha and Hprt transcript abundance. Bar graphs represent the mean of n=4 independent experiments.	71

4.7	Treatment and genotype dependent effects on organoid morphology distribution. a UMAP representation of DMSO treated (vehicle) and small molecule treated organoids. b, UMAP embeddings of four organoid genotypes (baseline state = 0.1% DMSO control-treated organoids), grey background consists of randomly sampled organoids.	77
4.8	Genotype dependent effects on organoid morphology. a Unperturbed organoid profiles from adenoma models were aggregated. Shown are representative individual organoids with selected features. Points show the mean phenotype for each independent biological replicate. Selected features and their z-scores relative to all single organoid profiles are shown (magenta = DNA, cyan = actin, cell permeability = yellow, scale-bar: 25 μ m)	78
4.9	Small molecule activity scoring. a A logistic regression classifier is trained to distinguish morphology profiles of individual treated and untreated organoids across all available replicates. Afterwards, the classifier is applied to a validation set of organoids and the classification performance is estimated using the area under the receiver operating characteristic curve (AUROC) metric. Method implemented by Jan Sauer. b Distribution of AUROC compound activity scores for all organoid lines, replicates and perturbations. AUROC scores were centered around 0.5 and treatments for this particular analysis were termed active when classification accuracy exceeded three times the Median Absolute Deviation (MAD) of the AUROC score distribution, an arbitrary threshold. A set of 184 compounds (16% of all screened small molecules) met the activity criteria. c Overlap of active compound treatments across organoid models. Shown is a Euler diagram of active compounds for each line. K (Kras G12D), AK (Apc loss and Kras G12D) and WT are color coded. Plot diagnostics: diagError: 0.011, stress: 0.002. d Identifying related treatment induced phenotypes. Normal vectors of treatment specific classifiers were compared by calculating the angular distance (related to cosine similarity, ranging from 0-180 degrees). Small angular distance between vectors correspond to a high similarity between the treatment-induced organoid phenotypes. Method implemented by Jan Sauer. e A map of compound induced phenotypes for Apc mutant and Apc wildtype organoids. Highlighted are clusters of compound induced phenotypes with related targets. Normal vectors for Apc mutant and Apc wildtype organoids were concatenated before angular distance calculation. Method implemented by Jan Sauer. f Treatment induced phenotypes by organoid genotype. Shown are significantly enriched treatment induced phenotypes for Apc mutant and wildtype organoid models. Clusters of similar phenotypes were tested for overrepresentation of known molecular targets using Fisher's exact test. Significantly enriched targets are shown. Method implemented by Jan Sauer.	80
4.10	Multi-omics factor analysis to identify shared factors linking morphology, size, gene expression, lipidomics, proteomics, genotype and drug activity. a Percent variance explained by the MOFA model for each factor. Untreated organoid morphology, organoid size and drug activity scores were integrated with genotype, proteomics, lipidomics and mRNA expression data. b Cumulative proportion of total variance explained by each experimental data modality within the MOFA model. c, Visualization of samples in factor space showing factors 1 and 2 as well as factor 1 and 3. Shown are independent replicates for each organoid line.	81
4.II	Multi-omics factor analysis input data and loadings. a measurement modalities, dimensionality and number of measurements. A third replicate of measurements were available for proteomics and lipidomics only. b Factor loadings for genotype information.	83

4.12	Factor 1, canonical Wnt signaling. a Gene-set enrichment network of factor 1 gene expression loadings. An edge connects Reactome pathways with more than 20% overlap. Central enriched processes include mitosis, DNA replication and DNA damage repair. b and c Gene set enrichment results of the "Lgr5 intestinal stem cell" and "proliferation" signature by Merlos-Suarez et al. over ranked factor 1 gene expression loadings (ranking from high factor 1 loading to low factor 1 loading, NES = normalized enrichment score). d Distributions of drug activity loadings grouped by drug target for factor 1. e Example images of compound treated organoids with WT or Apc-/- genotype. Representative images are displayed (magenta = DNA, cyan = actin, yellow = cell permeability, scale-bar: 200µm)	84
4.13	Small molecule Wnt signaling inhibitors. a AUROC activity score for three small molecule inhibitors of canonical Wnt signaling. b Target proteins for small molecules within the canonical Wnt signaling cascade with their relative position to the destruction complex (highlighted in blue). c Distributions of lipid abundance loadings grouped by lipid species for factor 1 (Abbreviations: HexCer - Hexosylceramide; PA - Phosphatidate; Cer - Ceramide; SM - Sphingomyelin; LPC - Lysophosphatidylcholine; PS - Phosphatidylserine; PE - Phosphatidylethanolamine; PI - Phosphatidylinositol; PG - Phosphatidylglycerol; DAG - Diacylglycerol; PC - Phosphatidylcholine; TAG - Triacylglycerol; Hex2Cer - Dihexosylceramide; CE - Cholersterolester)	85
4.14	Projection of factor 1 scores for treatment-induced phenotypes and viability changes. Highlighted are compounds leading to a significant change in projected factor scores across all organoid lines (ANOVA). Organoid viability is predicted using a random-forest based classifier (LDC) with scores from 0 (no toxicity) to 1 (complete toxicity)	86
4.15	GSK3 beta inhibition dependent morphology in colon organoid models. a Small molecule inhibition of GSK3 beta (CHIR98014) leads to phenocopying of Apc-/- genotype organoid models. b Shift of morphological features relative to scaled population mean for Apc wildtype organoid models treated with CHIR98014. c Excerpt of clustering from figure 4.9d, labeled with known binding activity of listed small molecules. Rucaparib is not member of the cluster and shown for comparison.	87
4.16	Factor 2, KrasG12D induced senescence. a Distributions of drug activity loadings grouped by drug target for factor 2. b Relationship of representative drugs' activity with factor 2 score. Shown are compounds from highlighted groups in panel (a). c Gene set enrichment results of a senescence signature by Fridman et al. over ranked factor 2 gene expression loadings (ranking from high factor 2 loading to low factor 2 loading, NES = normalized enrichment score). d Distributions of lipid abundance loadings grouped by lipid species for factor 2 (Abbreviations: HexCer - Hexosylceramide; PA - Phosphatidate; Cer - Ceramide; SM - Sphingomyelin; LPC - Lysophosphatidylcholine; PS - Phosphatidylserine; PE - Phosphatidylethanolamine; PI - Phosphatidylinositol; PG - Phosphatidylglycerol; DAG - Diacylglycerol; PC - Phosphatidylcholine; TAG - Triacylglycerol; Hex2Cer - Dihexosylceramide; CE - Cholersterolester). e Relationship of acyl chain length with factor 2 loading. Shown are lipids from highlighted species in panel (d)	89
4.17	Factor 3, KrasG12D effects in the context of Apc loss of function. a Distributions of drug activity loadings grouped by drug target for factor 3. b Relationship of representative mTOR inhibitor activity with factor 3 score. c Gene set enrichment results of a Reactome mTORC1 activation signature over ranked factor 3 gene expression loadings (ranking from high factor 3 loading to low factor 3 loading, NES = normalized enrichment score). d Visual summary of Myc gene set enrichment results for organoid state transitions. Myc signatures were significantly enriched in Apc-/- models and depleted in models with isolated KrasG12D+/- mutation.	91
4.18	93

Nothing in Biology Makes Sense Except in the Light of Evolution

Theodosius Dobzhansky

1

Introduction

1.1 DISCLOSURE

Parts of this introduction, especially the section on Wnt signaling, have been adapted from own publications, including *Wnt signaling in cancer*³

1.2 THE COLON

1.2.1 COLON FUNCTION AND VALUE AS MODEL ORGAN

evolutionary fundamental value representative model for complex organ well understood stem cell biology
enabled modeling in vitro devastating diseases emerge from colon

1.2.2 THE COLON STEM CELL NICHE

In order to understand the underlying principles of colorectal cancer development, or any cancer in general, it is advisable to turn to the stem cell biology governing the tissue's anatomy and function. The colon stem cell niche, or crypt, is the source of all epithelial cells lining the colon. Similar to the small intestine, Lgr5+ intestinal stemcells are located at the bottom of the crypt and continuously renew the epithelium by proliferating and pushing out new cells towards the colon's lumen.

This architecture serves multiple purposes, including protection of stem cells and the control of cell fate decisions across the epithelium. Multiple developmental pathways, especially Wnt signaling, Notch, BMP and ERK MAPK signaling, govern cell identity in the intestinal niche². The concentration of signaling cues for most of these pathways are organized in gradients along the crypt-lumen axis. For example, the concentration of stem cell property maintaining Wnt and EGF ligands, secreted by mesenchymal crypt cells and REG4+ deep secretory cells, decreases as cells are pushed outside of the crypt². In contrast, the effect of cell differentiating BMP ligands increases as the effect basal mesenchymal cell derived BMP inhibitors, such as Noggin, is reduced. In summary, as a cell is pushed outside the crypt by a continuous stream of fresh proliferating cells, developmental signaling cues vanish and subsequent gene expression changes lead to differentiation. Similarly, if cells were to move back into the crypt, the ambient signaling would lead to a reprogramming towards an intestinal stem cell fate.

Given the spacial confinement of proliferate signals, the crypt architecture leads to a protection against malignant transformation, too. At the bottom of the crypt, a neutral competition of proliferating intestinal stem cells leads to the rapid removal of cells that show reductions in their proliferation rate relative to wildtype stem cells, which is often the case in malignant neoplasms. Given this neutral competition and the dependence on external signaling, every dysfunctional or transformed cell is likely removed from the niche

and differentiates unless it acquires a set of molecular alterations that render it independent from niche signals. As mentioned above, the key signaling pathways that maintain stem cell properties in the crypt are canonical Wnt signaling and Notch signaling, while BMP signaling inhibits stemness and EGF dependent ERK MAPK signaling triggers cell proliferation. Given their exerted evolutionary pressure on malignant cells, it comes at no surprise that the majority of early driver mutations found in colorectal cancer, such as loss of APC (Wnt signaling), activation of β -catenin (Wnt signaling), activation of KRAS (ERK MAPK), BRAF (ERK MAPK) and loss of SMAD4 (TGF β /BMP signaling) are found in these exact signaling pathways. Of these mutations, especially mutations of APC and KRAS are frequently observed early in colorectal cancer development and are highly correlated with each other (50% of APC mutant tumors harbor mutations of KRAS), leading to both induced proliferative capacity and growth. In summary, the architecture of the colon stem cell niche and the signaling pathways required to regulate stemness and cell proliferation influence the evolutionary landscape of colorectal cancer development and thus account for the majority of early drive mutations, especially loss of APC and activation of KRAS, in this disease.

1.2.3 SIGNALING PATHWAYS CONTROLLING THE COLON STEM CELL NICHE

CANONICAL WNT SIGNALING

In 1973, the wingless gene was discovered in a screen for visual phenotypes, affecting patterning processes in *Drosophila melanogaster*, the fruit fly². Subsequently, further genetic screens identified components of the Wnt family as key regulators during embryonic development and later, cancer initiation as well as stem cell maintenance².

In canonical Wnt signaling, absence of Wnt ligands leads to phosphorylation of β -catenin by the destruction complex, which contains the scaffold protein Axin, the large protein APC (Adenomatous polyposis coli) and the kinases GSK3 β as well as casein kinase (CK1 α) (reviewed in Zhan, Rindtorff et al.³). In this state, β -catenin is phosphorylated by GSK3 β , ubiquitinated by β -TrCP and subsequently targeted for proteasomal degradation. In the absence of nuclear β -catenin, the TCF/LEF and transducin-like enhancer protein (TLE/Groucho) recruits Histone deacetylases to repress target genes.

The canonical pathway is activated upon binding of secreted Wnt ligands (for example, Wnt1 and Wnt

3a) to Fzd receptors and LRP co-receptors. Subsequently, LRP receptors are phosphorylated by CK α and GSK3 β , which then recruits Dishevelled (Dvl) proteins to the plasma membrane where they polymerize and are activated³. Next, the Dvl polymers inactivate the destruction complex by sequestration in multivesicular bodies. This results in stabilization and accumulation of β -catenin which then translocates into the nucleus. There, β -catenin forms an active complex with LEF (lymphoid enhancer factor) and TCF (T-cell factor) proteins by displacing TLE/Groucho complexes which leads to the recruitment of histone modifying co-activators such as CBP/p300, BRG1, BCL9 and Pygo³.

Next to Wnt ligands, members of the R-spondin ligand family are positive effectors of Wnt signaling^{3,4}. R-spondins bind to leucine-rich repeat containing G-protein-coupled receptors (Lgr) 4-6³. In the absence of R-spondin, the two E3 ubiquitin ligases Znrf3 and Rnf43 target the Frizzled (Fzd) receptor for lysosomal degradation³. The interaction of ubiquitin ligases and receptors is dependent on Dishevelled (Dsh)³. In the presence of external R-Spondins, binding of ligands to Lgr4-6 inhibits the activity of Znrf3/Rnf43 and leads to the accumulation of Fzd receptors on the cell surface³. Being transcriptional targets of Wnt signaling, Znrf3 and Rnf43 function as negative feedback regulators in Lgr5-positive cells³.

ERK MAPK SIGNALING

The extracellular-signal-regulated (ERK) mitogen-activated protein kinase family (MAPK) is one of three major MAPK families, together with the JNK (c-jun N-terminal kinase or stress-activated protein kinases) and MAPK14 group of protein kinases. These signaling cascades play a major role in (I) integrating external proliferative signals, (II) reacting to stress or ambient cytokines and (III) protecting cells from apoptosis, respectively³.

IGF AND mTOR SIGNALING

TGF BETA SIGNALING

TP53 SIGNALING

1.2.4 COLON ORGANOID MODELS

Intestinal organoids are three-dimensional cell culture models from primary adult tissue. Organoids develop from Lgr5+ adult stem cells and were first isolated from the small intestine of mice³. Subsequently, following the initial methodology, further organoid models across tissue-types and species have been developed. These include organoids from intact and cancerous human large intestine³, pancreas³, mammary epithelium²² and the hepatobiliary system²². Culturing these cells requires the addition of specific tissue-dependent growth factors and the embedding of cell in 3D hydrogels³. In the case of colon organoids, the necessary growth factors are inspired by signaling cues available in the intestinal stem cell niche: Wnt and R-spondin ligands secreted by PDGFR+ myofibroblasts activate and maintain canonical Wnt signaling; EGF ligands stimulate ERK MAPK signaling and Noggin ligands inhibit the differentiating effects of the BMP signaling cascade³. When combined with inhibitors of TGFβ and p38 mediated signaling, these growth factors can stimulate the formation and continuous proliferation of organoids ex-vivo.

Not only because of their high isolation efficacy of up to 80%, organoids are an increasingly popular model as they mimic their respective tissue of origin, including colorectal cancer (Pauli et al. 2017). Due to a high isolation efficiency and preserved tumor biology, patient derived colorectal cancer organoids have been used as personalized cancer models for precision oncology²². Moreover, organoids from healthy tissue can be cultured in-vitro as well and are amenable to genetic editing (Matano et al. 2015 Drost et al. (2015)). Therefore, these models also enable studying tumor development at a single mutation resolution.

1.3 COLORECTAL CANCER

1.3.1 COLORECTAL CANCER EPIDEMIOLOGY

Colorectal cancer is the third most common cancer worldwide and is associated with a Western lifestyle. Similar to other solid tumors, colorectal adenocarcinoma progression is classified into four stages by the UICC

(Union for International Cancer Control). These range from a *carcinoma-in-situ*, a malignant patch of cells that has not yet breached the basal lamina of the intestinal mucosa (stage 0), to metastatic disease (stage 4).

Figure 1.1: Colon Cancer Progression

1.3.2 COLORECTAL CANCER EMERGENCE AND EVOLUTION

Cancer in general, and colorectal cancer in particular, is a disease marked by the accumulation of genetic events in somatic cells, which over time lead to uncontrolled growth and spread beyond the tissue of origin. Genetic characterization efforts have identified recurring somatic genetic events in colorectal cancer that appear during its progression. Both sequencing and functional genomics experiments have helped distinguish a subset of genetic events, so-called functional events or drivers, that have a profound influence on cellular behavior and thus cancer development.

According to the adenoma-carcinoma sequence model, the majority of all colorectal adenocarcinomas arise from previously formed adenomas, benign neoplasms of the intestinal epithelium². As a consequence, one of the most effective medical interventions today to reduce death from colorectal cancer is the preventative removal of visible adenomas during lower endoscopy, such as colonoscopy².

On a molecular level, colorectal adenocarcinoma can be organized into tumors arising through (I) a chromosomal instability or (II) a DNA-mismatch repair deficiency associated route² (Figure 1.1). These two forms of tumor development have been associated with characteristic clinical, pathological and molecular findings. For example, tumors of the DNA-mismatch repair phenotype are more frequently located in the right colon, have a higher proportion of microsatellite instability, frequent BRAF mutations and a higher immune-cell infiltration². In contrast, tumors of the common chromosomal instability (CIN) phenotype are mostly microsatellite stable and have frequent APC and KRAS mutations.

The cascade of genetic events leading to the more frequent chromosomal instability associated form of colorectal cancer cause hyperactivation of a range of signaling pathways. Briefly, this cascade, known as the Vogelstein sequence², starts with the loss of the tumor suppressor APC in the intestinal epithelium. Loss of APC, which leads to adenoma formation, is followed by the activation of KRAS, PIK3CA, loss of SMAD4

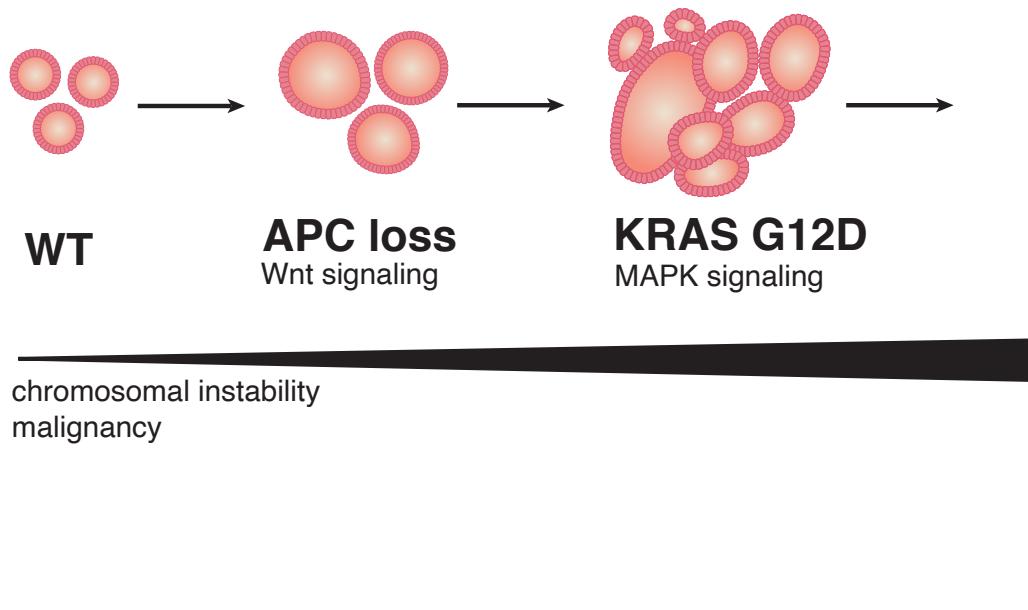


Figure 1.2

and TP53. Other forms of the disease, especially microsatellite-instable forms of colorectal cancer also harbor mutations of APC but show strong prevalence of BRAF mutations instead of KRAS mutations³.

Prior studies trying to further define colorectal cancer beyond these two developmental routes have proposed a set of different molecular subtypes³. In an attempt to unify these models, four consensus molecular subtypes (CMS) have been proposed³. Briefly, these subtypes organize colorectal cancer into classes defined by (I) a high fraction of microsatellite instable tumors, (II) APC mutations, (III) KRAS mutations and (IV) stromal infiltration, respectively. However, recent evidence has questioned the interpretability of these subtypes in multiple ways. First, the existence of non-malignant cells within the analyzed samples does not allow a direct interpretation of cancer-exclusive molecular processes which might govern treatment response and prognosis³. Second, the sampled intra-tumor region and its cellular composition influence the subtype classification result³. Third, validating studies of the consensus molecular subtype have shown that a

large fraction of cancer samples can not be confidently assigned to a single subtype and, more importantly, that subtypes, instead of being well separated, are rather on a high-dimensional continuum. This continuum can be defined by (I) markers of inflammation and T-cell activity and (II) markers of CDK-regulated DNA replication, leading to four quadrants in a continuous space³. For example, microsatellite instable tumors, which are most frequently found in the first CMS subgroup, were associated with a strong T-cell activity signature and a low CDK-regulated DNA replication signature. Along these lines, immune-cell infiltration has been established as an independent prognostic factor of overall survival and recurrence risk of colorectal cancer and is exploited in immunotherapy, which is especially active in MSI-high tumors^{??}. In contrast, CDK-dependent signaling has recently been recognized as a potential driver of immune-escape in multiple solid cancers³.

In summary, the molecular landscape of colorectal cancer is organized by two distinct forms of tumor development, chromosomal instability and microsatellite instability, that are linked to characteristic genetic changes resulting in a continuum of gene expression states which present themselves with varying degrees of cell proliferation and inflammation.

1.3.3 SIGNALING PATHWAYS CONTROLLING COLORECTAL CANCER

WNT SIGNALING IN COLORECTAL CANCER

The role of Wnt signaling during colorectal cancer development is well established³. While activating mutations of β -catenin do exist, loss of APC is the most frequent driver of Wnt signaling in colorectal cancer and can be found in about 80% of colorectal cancer patients³. In line with its role as a tumor suppressor, truncating of APC using the CRISPR/Cas9 technology, leads to colorectal cancer development, which can be modeled ex vivo in human intestinal organoids^{??}. Furthermore, by using a mouse model allowing the reversible knockdown of Apc via shRNA, it was demonstrated that adenomas could regress to normal tissue once APC function is restored, underlining the importance of continuous Wnt signaling for tumor maintenance³.

Although loss of APC in general is a driving event of colorectal cancer development and persistence, not every mutation of the APC gene leads to a similar phenotype. Studies of human colorectal cancer samples and tumors from mouse models revealed that different mutations of APC result in distinct levels of canonical

Wnt pathway activity and, in addition, are associated with characteristic tumor locations within the large intestine^{??}.

Besides APC, mutations in R-spondin and RNF43, regulators of Wnt receptor abundance at the cell surface level, were implicated as drivers of Wnt-dependent tumor growth. Deleterious RNF43 mutations have been described in 20% of colorectal cancer cases and are mutually exclusive to APC mutations. Also, amplified R-spondin3 fusion proteins have been described in 10% of CRC cases. While APC and β -catenin are generally considered independent of Wnt ligand availability, RNF43 mutant cancers are strongly dependent on Wnt secretion, rendering them highly susceptible to Wnt secretion targeted therapy.

ERK MAPK SIGNALING IN COLORECTAL CANCER

In colorectal cancer, the ERK MAPK signaling cascade and its members Ras/Raf/MEK and ERK are key regulators of cell proliferation in malignant cells. The Ras kinase members are mutated in about 36% of colorectal cancers[?]. According to the Vogelstein model of colorectal cancer initiation[?], activating mutations of KRAS takes place early during cancer development, more specifically, after loss of APC. Next to KRAS, BRAF mutations can also be found in around 10% of colorectal cancers[?]. Of note, mutations of KRAS and BRAF occur mostly in a mutually exclusive pattern with BRAF mutations being enriched in Microsatellite instable colorectal cancers^{??}.

IGF AND mTOR SIGNALING IN COLORECTAL CANCER

TGF BETA SIGNALING IN COLORECTAL CANCER

TP53 SIGNALING IN COLORECTAL CANCER

1.3.4 COLORECTAL CANCER ORGANOID MODELS

1.4 COLORECTAL CANCER THERAPY

The treatment of colorectal adenocarcinoma depends on disease stage. While surgical removal of the tumor is at the center of the treatment strategy, neoadjuvant and adjuvant chemotherapy are part of the recommended therapy from UICC stage 2 and 4 on, respectively. Today, for metastatic colorectal cancer the first line

treatment includes combination chemotherapy (FOLFOX or FOLFIRI) paired with Cetuximab (anti-EGFR) for KRAS wildtype disease, combination chemotherapy with Bevacizumab (anti-VEGFR) for KRAS mutant disease or triple chemotherapy (FOLFOXIRI) in combination with Bevacizumab for BRAF mutant disease². Following lines of therapy include different combinations of the aforementioned agents with the exception of Regorafenib and Triflouridin/Tipiracil as preferred third line agents for non-KRAS wildtype disease³. Consequentially, the only genetic tests currently recommended during therapy are the determination of KRAS and BRAF status³. Other genetic tests or targeted inhibitors have so far not found their way into clinical practice.

Both the important role of adenocarcinoma development and the limitations of personalized therapy in advanced stages of the disease motivated the research presented in this dissertation.

1.5 CANCER DRUG DISCOVERY

1.5.1 TYPES OF BIOLOGICAL MODELS

counterfactual reasoning - if x then y

1.5.2 GENETIC INTERACTIONS

1.5.3 RATIONAL AND FUNCTIONAL DRUG DISCOVERY

phenotype first

1.5.4 IMAGE-BASED PROFILING

A challenge for organoid research is the need for information-rich drug-testing methods. In the past, automated microscopy of 2D-cells has been used to measure the biological activity of compounds (Breinig et al. 2015). Prompted by this, we build a platform for high-throughput drug activity profiling in organoids. The platform uses confocal microscopy to collect fluorescent images of treated organoids in 3D. We developed SCOPE, a software package, to process these images and measure organoid phenotypes. We used this platform to test compounds on both patient derived organoids and genetically engineered organoid models.

1.6 MULTI-VIEW REPRESENTATION LEARNING IN BIOLOGY

PCA CCA MOFA Theis lab papers Schoelkopf causal representation learning and Fischer causal genetic interaction

1.7 TOWARDS CAUSAL REPRESENTATION LEARNING IN BIOLOGY

1.8 AIMS OF THESIS

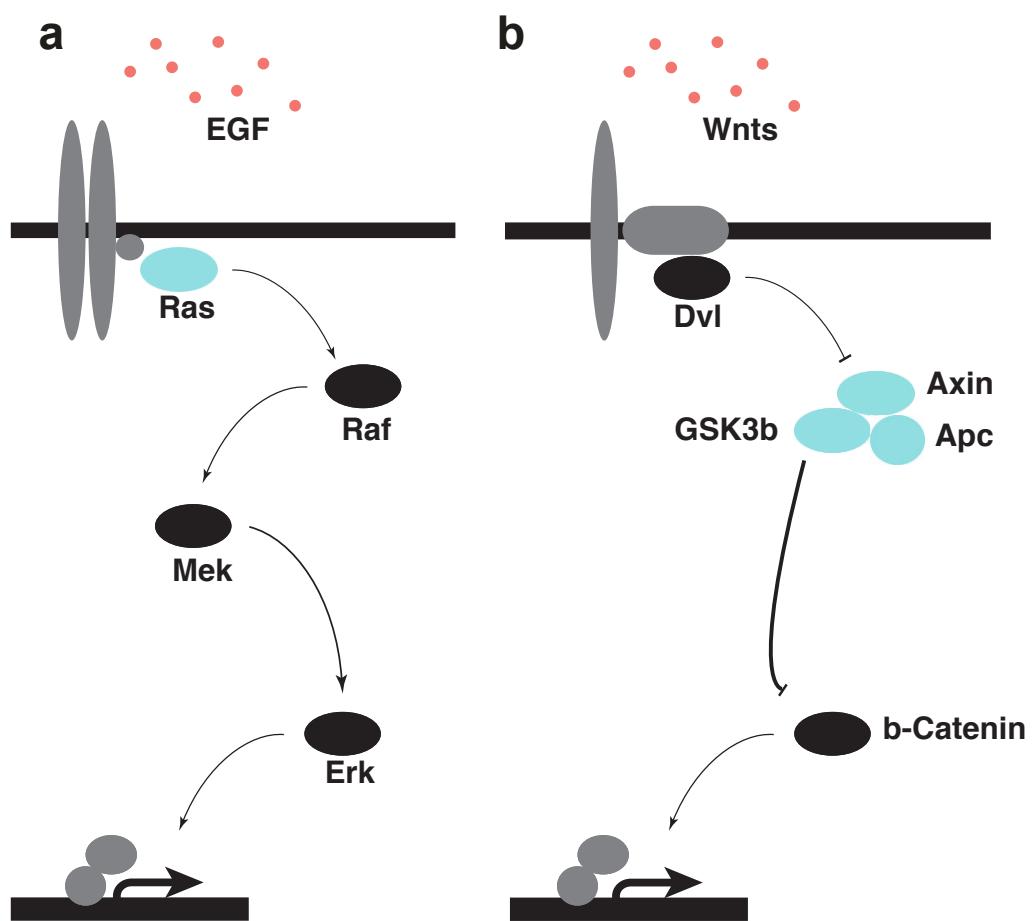
1.8.1 IMAGE-BASED PROFILING OF COLON ORGANOID MODELS

Patient derived organoids (PDOs) are physiological 3D tumor models that can be efficiently derived from cancer and normal tissues.^{1–3} Organoid isolation from human primary tumors and metastases^{1,4} has enabled the establishment of living biobanks.^{2,3,5–7} Notably, patient derived organoids have been shown to represent their origin's molecular features and morphology,^{2–4,8} enabling functional experiments such as drug testing ex vivo.^{3,5,9–14} Image-based profiling is a high-throughput microscopy-based methodology to systematically measure phenotypes of in vitro models. By learning a lower dimensional representation of biological images, biological states can be described, quantified and deciphered. When scaled and combined with chemical or genetic perturbations, this becomes a powerful approach to gain systematic insights into biological processes, making it a popular method in drug discovery and functional genomics research.^{15–17} Image-based assays have for instance been used to screen large libraries of small molecules to identify potential drug candidates, to analyze a drugs mode of action, or to classify drug-gene interactions by cell-morphology.^{18–21} Recently, image-based profiling of monogenic disease models has been particularly relevant for drug discovery. Here, primary cells from diseased tissue are profiled, perturbed and candidate drugs reversing the phenotype from a diseased to a healthy cell morphology are prioritized for drug development or used as diagnostics.^{22,23} Previously, high-throughput image-based profiling experiments have mostly been performed in a limited number of adherent cell lines and not been available to a growing number of disease models that require more sophisticated culture conditions, such as models from diverse cancers, rare genotypes or personalized disease models. Performing image-based profiling experiments of patient derived cancer organoids, a prime example for a complex 3D polygenic disease model, are, however, a technical and biological challenge. While

image-based drug testing in organoids have been performed, the morphological heterogeneity of patient-derived cancer organoids between and within patient donors as well as their diverging behavior upon molecular perturbations are not yet systematically understood.^{24–26} Here we report on a large-scale image based phenotyping study of patient derived cancer organoids. We generated patient derived organoids from colorectal cancer patients and treated 11 organoid models with more than 500 experimental and clinically used compounds at different concentrations. We systematically mapped the morphological heterogeneity of patient derived organoids and their response to compound perturbations from more than 3,700,000 confocal microscopy images. We identified a heterogeneous landscape of organoid phenotypes mainly driven by differences in organoid size, viability and cystic vs. solid organoid architecture. Using multi-omics factor analysis, we identified biological programs underlying these phenotypes and compounds that modulate them. For example, we linked cystic organoid architecture with a LGR5+ enriched, Wnt signaling dependent organoid state that could be induced via MEK inhibition and organoid size to an IGF-1R driven organoid state that was relatively insensitive to EGFR inhibition and could be induced via mTOR inhibition. A better understanding of organoid phenotypes and the ability to use multi-omics data to annotate organoid states and their plasticity have the potential to further accelerate image-based drug discovery for complex multigenic diseases.

1.8.2 PATIENT DERIVED ORGANOIDS IDENTIFIES COMPOUND-INDUCED PHENOTYPES

1.8.3 MULTI-OMICS PROFILING OF INTESTINAL ORGANOIDS IDENTIFIES AN EPISTATIC RELATIONSHIP OF APC LOSS AND KRAS ACTIVATION DURING COLORECTAL CANCER DEVELOPMENT



Details matter, it's worth waiting to get it right.

Steve Jobs

2

Materials and Methods

2.1 PATIENTS

All patients were identified at the University Hospital Mannheim, Mannheim, Germany. We included untreated patients with a new diagnosis of colon or rectal cancer in this study and obtained biopsies from their primary tumors and adjacent normal tissue via forceps based endoscopy. Exclusion criteria were active HIV, HBV or HCV infections. Biopsies were transported in phosphate buffered saline (PBS) on ice for subsequent organoid extraction. Clinical data, tumor characteristics and molecular tumor data were pseudonymized. The study was approved by the Medical Ethics Committee II of the Medical Faculty Mannheim, Heidelberg University (Reference no. 2014-633N-MA and 2016-607N-MA). All patients gave

written informed consent before tumor biopsy was performed. In this study, we extracted PDOs from 25 patients with colorectal cancer, 10 of them female, 15 male, with a mean age of 66 years (median 65). 16 patients had a rectum carcinoma, 9 a colon carcinoma.

2.2 ORGANOID CULTURE

2.2.1 PATIENT DERIVED ORGANOID CULTURE

Patient derived organoid (PDO) cultures were extracted from biopsies as reported by Sato et al.⁴ with slight modifications. Tissue fragments were washed in DPBS (Life technologies) and digested with Liberase TH (Roche) before embedding into BME R1 (Trevigen). The medium, termed ENA, contained Advanced DMEM/F12 (Life technologies) medium with 1% v/v penicillin/streptomycin (Life Technologies), Glutamax and HEPES (basal medium) supplemented with 100 ng/ml Noggin (Peprotech), B27 (Life technologies), 1,25 mM n-Acetyl Cysteine (Sigma), 10 mM Nicotinamide (Sigma), 50 ng/ml human EGF (Peprotech), 10 nM Gastrin (Peprotech), 500 nM A83-01 (Biocat), 10 nM Prostaglandin E2 (Santa Cruz Biotechnology), 10 µM Y-27632 (Selleck chemicals) and 100 mg/ml Primocin (Invivogen). After isolation, cells were kept in 2 conditions including medium as described (ENA), or supplemented with additional 3 µM SB202190 (Biomol) (ENAS) as described by Fujii et al.³. The tumor niche was determined after 14 days and organoids were subsequently cultured in the condition with best visible growth. Organoids were passaged every 7 days and medium was changed every 2-3 days.

2.2.2 MOUSE ORGANOID CULTURE

A heterozygous LSL-Kras G12D (B6.129S4-Krastm4Tyj/J) female mouse was crossed with a homozygous Rosa26-CreERT2 (B6.129-Gt(ROSA)26Sortm1(cre/ERT2)Tyj/J) male to generate offspring with a Tamoxifen activatable KRAS G12D allele. A single healthy LSL-Kras G12D CreERT2 mouse (male, 8 weeks) was sacrificed for organoid generation.

Mouse colon organoids were isolated based on work by Sato et al⁵. After cervical dislocation of the sacrificed mouse the colon was prepared and excised between caecum and rectum. The tissue was stored on ice in cooled DPBS (Life technologies), cut open lengthwise and washed three times with DPBS. After thorough

washing, colon fragments were cut into 2mm pieces and incubated in a 5mM EDTA/DPBS (Sigma) solution for 60 minutes on a rocking table. Digested fragments were allowed to settle and resuspended in DMEM/F12 (Life technologies) by repeated up- and down-pipetting with a serological pipette. The resulting crypt suspension was filtered with a 70ul filter (Falcon), crypts were counted and centrifuged at 150g, 10min, 4C. The resulting pellet was resuspended in 10mg/ml Matrigel (Corning) and plated on prewarmed 6-well suspension plates (Greiner). After 30-60 minutes of solidification, droplets were overlaid with complete organoid growth medium and incubated at 37C, 5% CO₂ in atmospheric air.

Complete colon organoid medium, termed WENRAS, contained 30% advanced DMEM/F12 (Life Technologies) supplemented with 1% v/v penicillin/streptomycin solution (Life Technologies), 1% v/v HEPES buffer (Life Technologies) and 1% v/v Glutamax (Life Technologies), 50% Wnt3A conditioned medium, and 20% R-spondin1-FC conditioned medium. The medium was further supplemented with recombinant Noggin (100 ng/ml), IX B27 (IX), n-Acetyl-cysteine (1.25 mM), Nicotinamide (10 mM), EGF (50 ng/ml), 500 nM A83-01 (Tocris), SB202190 (3 μM), Y-27632 (10 μM) and Primocin (100 μg/ml). All small molecule inhibitors were dissolved in DMSO.

After isolation, colon organoids were cultured in solidified BME R1 (10mg/ml) droplets and overlaid with genotype and experiment dependent growth medium. The medium was exchanged every 48-72 h. APC mutant colon organoid lines were cultured without Wnt and R-spondin conditioned medium, which was replaced by basal medium instead. Organoids were passaged weekly by digestion with TrypLE (Gibco) and resuspension in BME R1 (10mg/ml). Organoids were regularly tested for Mycoplasma contaminations.

2.2.3 GENETIC EDITING OF ORGANOIDS

An sgRNA targeting the murine ortholog of the APC mutation cluster region (MCR) was designed using E-CRISP(Heigwer et al., 2014). The Apc targeting sgRNA was cloned into the one-vector plasmid pSpCas9(BB)-2A-Puro (PX459) V2.0 according to Ran et al.³. Briefly, the vector was digested with BbsI-HF (Thermo) and the phosphorylated and annealed oligonucleotide sgApc1 (sgAPC1 F and -R) was ligated using T4-Ligase (Thermo). The construct was transformed into chemically competent bacterial cells (Stellar, Clontech) and plated on Carbenicillin agar. Individual colonies were isolated and sequencing of plasmid DNA from cultured colonies confirmed successful molecular cloning. Extracted organoids (termed

“wildtype”, “WT”) were cultured for multiple passages before transfection of the plasmid with Lipofectamine 2000 (Thermo). Here, grown organoids were digested with TrypLE (Gibco) and treated with Lipofectamine and plasmid DNA according to the manufacturer’s protocol. Transfected organoids were seeded in BME RI (10mg/ml) and Wnt₃A/R-Spondin1-Fc withdrawal was started 7 days after transfection. Surviving organoids were cultured continuously without Wnt₃A and R-Spondin1-Fc conditioned medium (termed “A”). To activate oncogenic Kras, Wildtype and APC mutant organoid lines were treated for 7 days with 0.5uM 4-Hydroxytamoxifen (Sigma) without EGF in the medium. 4-Hydroxytamoxifen was dissolved in Ethanol. After treatment, organoids were cultured with EGF containing media thereafter (termed “K” or “AK”, respectively).

2.3 BIOCHEMICAL ASSAYS

2.3.1 AMPLICON SEQUENCING OF PATIENT DERIVED ORGANOIDS

DNA was isolated from 19 organoid cultures with the DNA blood and tissue kit (Qiagen). Sequencing libraries were prepared with a custom panel (Tru-Seq custom library kit, Illumina) according to the manufacturers protocol and sequenced on a MiSeq (Illumina). Targeted regions included the most commonly mutated hot spots in colorectal cancer in 46 genes captured with 157 amplicons of approximately 250bp length. A list of targeted hot-spots that were sequenced can be found in supplementary table S5. After mapping of the reads to GRC38 reference genome using Burrows-Wheeler Aligner (BWA), data were analyzed using the Genome Analysis Toolkit (GATK) (McKenna et al., 2010; Van der Auwera et al., 2013). Base recalibration was performed and variants were called using MuTect2 pipeline. Variants with a variant frequency below 10%, with less than 10 reads, or with a high strand bias (FS<60) were filtered out. Variants were annotated with Ensemble variant effect predictor (McLaren et al., 2016) and manually checked and curated using integrative genomics viewer, if necessary³. Only non-synonymous variants present in COSMIC (Forbes et al., 2015) were considered true somatic cancer mutations. Also, all variants annotated “benign” according to PolyPhen database and “tolerated” in SIFT database were excluded, as well as variants with a high frequency in the general population as determined by a GnomAD (Lek et al., 2016) frequency of >0.001.

2.3.2 AMPLICON SEQUENCING OF MOUSE ORGANOIDS

Amplicon sequencing was performed to validate the genetic perturbation of Apc. DNA from Apc targeted and untargeted organoid lines was prepared using the DNA Blood and Tissue Kit (Qiagen), according to the manufacturer's tissue protocol including an RNase digestion. The targeted region was PCR amplified using primers F1 to R2, and sequencing libraries were prepared according to the manufacturer's protocol. Libraries were sequenced on a MySeq (Illumina) using 100bp single end reads.

2.3.3 GENOMIC PCR OF KRAS G12D ALLELE

To confirm activation of oncogenic Kras in 4-Hydroxytamoxifen treated lines, genomic DNA was isolated from all 4 organoid lines as described above. Presence or absence of the Lox-STOP-Lox cassette was evaluated by PCR according to the Kras G12D conditional PCR protocol by Tyler Jacks' group(Jackson et al., 2001). Briefly, primers 2 and 3 were used for genotyping on genomic DNA using the Q5 PCR protocol (NEB).

2.3.4 WESTERN BLOTTING

Organoids were cultured in Matrigel (Corning). Organoids were collected, and cells were isolated using Matrisperse (Corning) for 40 minutes on a rocking table. Isolated organoids were lysed in RIPA buffer (Sigma) with Protease inhibitor (Sigma) and Phosphatase inhibitor 3 (Sigma). Protein concentration was measured using the Pierce BCA kit (Thermo) according to the manufacturers protocol and samples were loaded onto NuPage gels (Thermo). Western Blotting was performed with following antibodies: anti-p(hospho)-Erk (1:2000), anti-Erk (1:1000), anti-SREBP1 (1:1000), anti-SCAP (1:5000) and anti-beta-actin-HRP (1:150,000).

2.3.5 ORGANOID GROWTH PATTERNS

Organoids were passaged and seeded in 4 different growth media with medium changes every 48h. Images were taken 120h after seeding.

2.3.6 RT-qPCR

Organoids were passaged and seeded in 4 different growth media with medium changes every 48h. After 120h, organoid RNA was isolated using the RNeasy Kit (Qiagen) with beta-Mercaptoethanol (Invitrogen) and a DNase digestion step. cDNA was synthesized using Oligo-dT primers (Thermo), Ribolock Ribonuclease inhibitor (Thermo) and Revert Aid H Minus reverse transcriptase (Thermo). RT-qPCR was performed using the ROCHE UPL kit (Roche), Sdha and Hprt expression levels were used as controls and averaged. Relative transcript abundance was measured using the ddCT method. RT-qPCR primers targeting Axin2, Ccnd, as well as the mouse reference genes Sdha and Hprt were used.

2.3.7 MASS SPECTROMETRY

Organoids were cultured according to the screening protocol (above). Organoids were isolated with Matrisperse (Corning) as described above. Isolated organoids were lysed in Ammonium Bicarbonate lysis buffer (50mM, pH 8.2) with 2.5% w/v SDC and 25U/ml Benzonase. Samples were applied to 1D-SDS-PAGE and fractionated. Gel pieces were extracted, cysteins residues reduced by DTT and carbamidomethylated using iodoacetamide. The samples were digested with Trypsin overnight. Resulting peptides were loaded on a cartridge trap column, packed with Acclaim PepMap300 C18, 5µm, 300Å wide pore (Thermo) and segregated in a 60 min gradient from 3% to 40% ACN on a nanoEase MZ Peptide analytical column (300Å, 1.7 µm, 75 µm x 200 mm, Waters). Eluted peptides were analyzed by an online coupled Q-Exactive-HF-X mass spectrometer.

2.3.8 EXPRESSION PROFILING

Organoid RNA was isolated from 19 PDO lines with the RNeasy mini kit after snap freezing organoids on dry ice. Samples were hybridized on Affymetrix U133 plus 2.0 arrays. Raw microarray data were normalized using the robust multi-array average (RMA) method (Irizarry et al., 2003) followed by quantile normalization as implemented in the ‘affy’ (Gautier et al., 2004) R/Bioconductor (Huber et al., 2015) package. In order to exclude the presence of batch effects in the data, principal component analysis and hierarchical clustering were applied. Consensus molecular subtypes were determined as described previously (Guinney et al., 2015) using

the single sample CMS classification algorithm with default parameters as implemented in the R package ‘CMSclassifier’. In all cases, differential gene expression analyses were performed using a moderated t-test as implemented in the R/Bioconductor package ‘limma’ (Ritchie et al., 2015). Gene set enrichment analyses were performed using ConsensusPathDB (Kamburov et al., 2013) for discrete gene sets or GSEA as implemented in the ‘fgsea’ (Sergushichev, 2016; Subramanian et al., 2005) R/Bioconductor package for ranked gene lists. Mouse organoids were cultured according to a standardized screening protocol. Briefly, organoid models were seeded and cultured for 72h in WENRAS+Y and additional 96h in ENR-Y. Samples were harvested after 7 days and RNA was isolated using the RNAEasy Kit (Qiagen) as described above. Transcript expression levels were measured using MoGene-2_o-st chips (Affymetrix).

2.4 ORGANOID PROFILING

2.4.1 CELL SEEDING DURING COMPOUND TESTING

PDO drug profiling followed a standardized protocol with comprehensive documentation of all procedures. Organoids were collected and digested in TrypLE Express (Life technologies). Fragments were collected in basal medium with 300 U/ml DNase and strained through a 40µm filter to achieve a homogeneous cell suspension with single cells and small clusters of cells, but without large organoid fragments. 384 well µclear assay plates (Greiner) were coated with 10µL BME V2 (Trevigen) at a concentration of 6.3 mg/ml in basal medium, centrifuged and incubated for >20 min at 300G and 37° C to allow solidification of the gel. PDO cell clusters together with culture medium (ENA) and 0,8 mg/ml BME V2 were added in a volume of 50µl per well using a Multidrop dispenser (Thermo Fisher Scientific). Plates were sealed with a plate-loc (Agilent) and centrifuged for additional 20 min allowing cells to settle on the pre-dispensed gel. Cell number was normalized before seeding by measuring ATP levels in a 1:2 dilution series of digested organoids with CellTiter-Glo (Promega). The number of cells matching 10,000 photons was seeded in each well. After seeding of organoid fragments, plates were incubated for three days at 37°C to allow organoid formation before addition of compounds. Two biological replicates of each PDO line from different passages were profiled at different time points.

Mouse organoid screening was performed as described above with slight modifications. Clotting of organoid

fragments was avoided by adding 10 U/ml of bovine DNase1 to the medium during filtration. The cell viability of digested fragment suspensions was estimated using Cell-Titer-Glo (Promega). 40ul of cell suspension was mixed with 40ul of undiluted reagent and measured after 30 minutes on a Mithras plate reader (Berthold). Cell fragments with a viability corresponding to 5000 photons were seeded per well on pre-coated 384 well plates using a Multidrop peristaltic pump robot (Thermo). After 72 hours of organoid expansion in WENRAS+Y, the medium was changed to ENR-Y and compound libraries were added using a BiomekFX (Beckmann). Screening plates were further incubated for 96h before a cell permeability dye Image-IT DeadGreen was added for 4 hours and cells were fixed in a 3% PFA buffer supplemented with 1% BSA. After fixation, plates were stored for up to 4 days or directly stained with DAPI (Sigma) and Phalloidin-TRITC. Processed plates were imaged using an Incell6000 automated line-scanning confocal fluorescence microscope.

2.4.2 COMPOUND LIBRARIES

Two compound libraries were used for screening: A library containing 63 clinically relevant drugs (clinical cancer library) and a large library of 464 compounds targeting kinases and stem cell or developmental pathways associated genes (KiStem library). The clinical cancer library was manually curated by relevance for current (colorectal) cancer therapy, mechanism of action and potential clinical applicability. Compounds of this library are in clinical use or at least in phase I/II clinical trials. Five concentrations per compound were screened (five-fold dilutions). The concentrations were determined by analysis of literature data from previous 3D and 2D drug screens and own experiments. A list of compounds included in this library and maximum concentrations used can be found in S3. The KiStem library includes 464 compounds targeting a diverse set of kinases and stem cell relevant pathways S4. All compounds in this library were used in a concentration of 7.5 μ M. All compounds were obtained from Selleck chemicals. Compounds of both libraries were arranged in an optimized random layout. We stored compound libraries in DMSO at -80 C.

2.4.3 COMPOUND TREATMENT

Medium was aspirated from all screening plates and replaced with fresh ENA medium devoid of Y-27632, resulting in 45 μ l volume per well. Drug libraries were diluted in basal medium and subsequently 5 μ l of each

compound was distributed to screening plates. All liquid handling steps were performed using a Biomek FX robotic system (Beckmann Coulter). Plates were sealed and incubated with the compounds for four days. All PDO lines underwent profiling with the clinical cancer library, while the KiStem library was used with 13 PDO lines.

2.4.4 LUMINESCENCE VIABILITY READ OUT

Plates undergoing viability screening were treated with 30 μ l CellTiter-Glo reagent after medium aspiration with a Biomek FX. After incubation for 30 minutes, luminescence levels were measured with a Mithras reader (Berthold technologies).

For mouse experiments, before compound addition, organoid viability was measured using Cell-Titer-Glo (Promega) and WENR-Y was added to the remaining plates. Cell viability was measured after compound exposure as described above. The pre-treatment viability of organoids was used to estimate growth-rate controlled dose-response curves according to Clark et al.(Hafner et al., 2016) Measurements of GR metrics for was not robust for slow proliferating WT lines. Therefore, we omitted plotting dose-response curves for WT lines when a GRfit was conducted. Dose response curves of relative viability, including the WT line can be found in the supplementary figures.

2.4.5 IMAGE-BASED PHENOTYPING

Image-IT DeadGreen (Thermo Fisher) was added to the cultures with a Multidrop dispenser (Thermo Fisher) in 100nM final concentration and incubated for four hours. Afterwards, medium was removed and organoid cultures were fixed with 3% PFA in PBS with 1% BSA. Fixed plates were stored at 4° C for up to three days before permeabilization and staining. On the day of imaging, organoids were permeabilized with 0.3% Triton-X-100 and 0.05% Tween in PBS with 1% BSA and stained with 0.1 μ g/ml TRITC-Phalloidin (Sigma) and 2 μ g/ml DAPI (Sigma). All liquid handling steps were performed with a BiomekFX. Screening plates were imaged with an Incell Analyzer 6000 (GE Healthcare) line-scanning confocal fluorescent microscope. We acquired 4 fields per well with z-stacks of 16 slices at 10x magnification. The z-steps between the 16 slices had a distance of 5 μ m, the depth of field of each slice was 3.9 μ m.

2.4.6 IMMUNOHISTOCHEMISTRY

PDOs were fixed for 20 min in 4% (v/v) Roti Histofix (Carl Roth) followed by embedding into MicroTissues 3D Petri Dish micromolds (Sigma Aldrich) using 2% (w/v) Agarose LE (Sigma) in PBS supplemented with 0.5 mM DTT. Thereafter, PDOs were subjected to dehydration steps and embedding in paraffin. Formalin-Fixed Agarose/Paraffin-Embedded sections (3- 5 μ m) were manually cut from blocks with a microtome (Leica RM 2145) and transferred to glass slides (Superfrost, ThermoFisher Scientific) before HE staining using automated staining devices.

2.4.7 PROLIFERATION ASSAY

Organoids were collected, digested, strained, normalized and seeded as in the drug profiling protocol described above. Medium change was performed on days 2 and 4. Viability of plated organoids was measured 2, 4 and 8 days after seeding using CellTiter-Glo as described above.

2.5 IMAGE ANALYSIS

2.5.1 IMAGE PROCESSING

Microscopic image z-stacks were compressed to HDF5 format for archival and underwent maximum contrast projection using the R/Bioconductor package MaxContrastProjection for further processing of the images. Segmentation of the projections based on intensity did not sufficiently identify organoids. Instead, we trained a deep convolutional neural network (DCNN) on the partially correct intensity segmentation, leveraging the robustness of DCNNs with regards to mislabeled training data and eliminating the need for expensive manual annotations. Standard image features, including shape, moment, intensity, and Haralick texture features on multiple scales, were extracted using the R/Bioconductor package EBImage (Pau et al., 2010). Of note, the strong diversity of unperturbed organoid phenotypes between PDO lines did not allow the definition of a core set of individual reproducible descriptive features across all screened organoids. Therefore, no correlation-based filtering of features was done, allowing comparisons between different PDO lines. Out-of-focus objects were programmatically removed from the dataset using a feature based random forest

classifier. Two PDO lines (Do15T01 and Do21T01) had to be excluded from analysis because of too many of out-of-focus elements.

2.5.2 DRUG-INDUCED PHENOTYPES

A principal component analysis (PCA) was performed on the entire dataset to reduce the dimensionality. 25 principal components were selected, explaining approx. 81% of the total variance within the dataset. A linear support vector machine (SVM) was trained per line and treatment (and per concentration where applicable) to differentiate treated organoids from negative controls based on the PCA-transformed features (Loo et al., 2007). To allow comparison between various PDO lines and drug perturbations, the distributions of features describing organoids from different batches were adjusted. Drugs were categorized as either active or inactive based on the accuracy of the SVM. The histogram of accuracies made a threshold of 85% the most intuitive. The direction of the vector perpendicular to the SVM hyperplane was interpreted as the drug-induced effect. Drugs were clustered with regard to the angles of their corresponding effect vectors in PCA-feature space.

2.5.3 LIVE-DEAD CLASSIFICATION

A random forest classifier was trained on the original single organoid features to differentiate living from dead organoids. Organoids treated with DMSO were used as negative (i.e. living) controls while organoids treated with Bortezomib and SN-38 at the two highest concentrations were used as positive (i.e. dead) controls. Visual inspection of the projected images confirmed our choice of positive controls. Binary classification results were averaged within wells to obtain viability scores ranging from 0 to 1, indicating how lethal a treatment was. A separate classifier was trained for each individual line to ensure inter-line independence.

2.5.4 ANALYSIS OF LUMINESCENCE DATA AND DOSE-RESPONSE RELATIONSHIPS

Raw luminescence data of each plate were first normalized using the Loess-fit method (Mpindi et al., 2015) in order to correct for edge effects where increased luminescence intensity was observed along the edges of each plate. Subsequently, each plate was normalized by division with the median luminescence intensity of the DMSO controls. Drug response Hill curves (DRC) were fitted and area under the curve values were calculated for each DRC using the ‘PharmacoGx’ (Smirnov et al., 2016) R/Bioconductor package.

The beginning is the most important part of the work.

Plato, The Republic

3

Image-based profiling identifies molecular
determinants of cancer organoid architecture and
plasticity

3.1 DISCLOSURE

Significant parts of this chapter have been adapted from own manuscripts, including *Image-based profiling identifies molecular determinants of cancer organoid architecture and plasticity*³ and its public pre-print versions. The maximum contrast projection method, organoid segmentation method, feature extraction procedure and organoid viability classification (LDC) was previously developed by Jan Sauer as part of his dissertation. Image-based profiling experiments were supported by Johannes Betge and Clara Dingert.

3.2 ESTABLISHING PATIENT DERIVED ORGANOIDS FOR IMAGE-BASED PROFILING

Patient derived organoids can be established from diverse healthy or malignant tissues and have been shown to represent their tissue of origin with respect to morphological and molecular features including gene expression and somatic mutations^{2,8,7,3}. To generate personalized cancer models for image-based profiling, I designed and implemented a standardized laboratory workflow to generate patient derived organoids from colorectal cancer samples via endoscopic biopsy (Figure 1a). Briefly, fresh patient samples were washed, digested and embedded in a basal membrane extract, a proprietary mixture of extracellular matrix proteins especially rich in Laminin and Collagen 4. The isolated tumor cells were then overlaid with a growth factor rich medium, containing Epidermal Growth Factor (EGF), the BMP-signaling antagonist Noggin and the small-molecule inhibitor A83-01, which inhibits TGF β -signaling by targeting the Activin receptor-like kinase family.

Following this protocol, patient derived organoids from 13 patients with colorectal cancer were prospectively developed. Donors to the biobank represented different UICC stages (Figure ??). Gene expression profiling and amplicon sequencing of frequently altered genes in colorectal cancer showed molecular profiles characteristic for the disease . Similar to sequencing studies of primary tumors, patient derived organoids harbored a high frequency of APC (6/13), KRAS (8/13) and TP53 (5/13) mutations³. On a gene expression level, patient derived organoids mainly represented the canonical consensus molecular subtype CMS 2 of colorectal cancer². No patient derived organoid line with a MSI-high phenotype and the associated CMS 1 molecular subtype was established. Also, no organoid line matched the molecular subtype CMS 4, which is associated with stromal infiltration and TGF β -signaling. These results are in line with previous

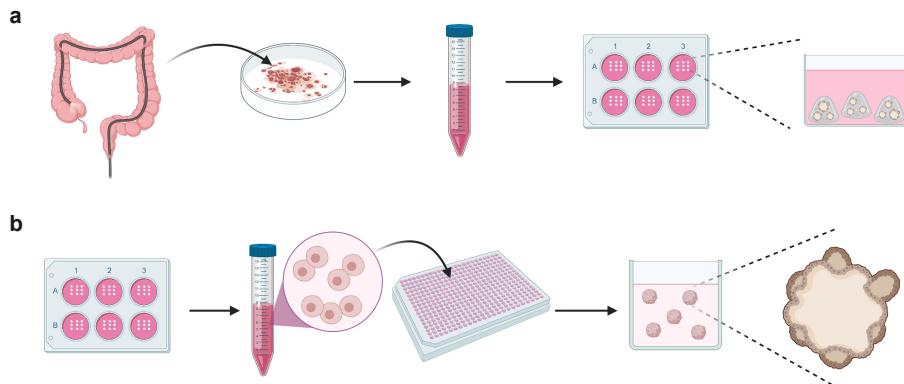


Figure 3.1: Central liquid handling methods **a** Organoid isolation procedure. Colorectal cancer tissue biopsies were collected via endoscopy, enzymatically removed from extracellular matrix proteins, washed and resuspended in basal membrane extract hydrogel. After solidification of hydrogel domes, organoids were overlayed with growth factor rich culture medium. **b** Organoid high-throughput experimentation. Colorectal cancer organoids were harvested, partially digested, seeded in hydrogel-coated 384-well plates

observations^{7,6} and the limitations of the organoid culture system, which selects for growth of epithelial cells (favored by canonical Wnt signaling high, BMP4 signaling low) over mesenchymal cells (favored by canonical Wnt signaling low, BMP4 signaling high) *ex vivo*⁴.

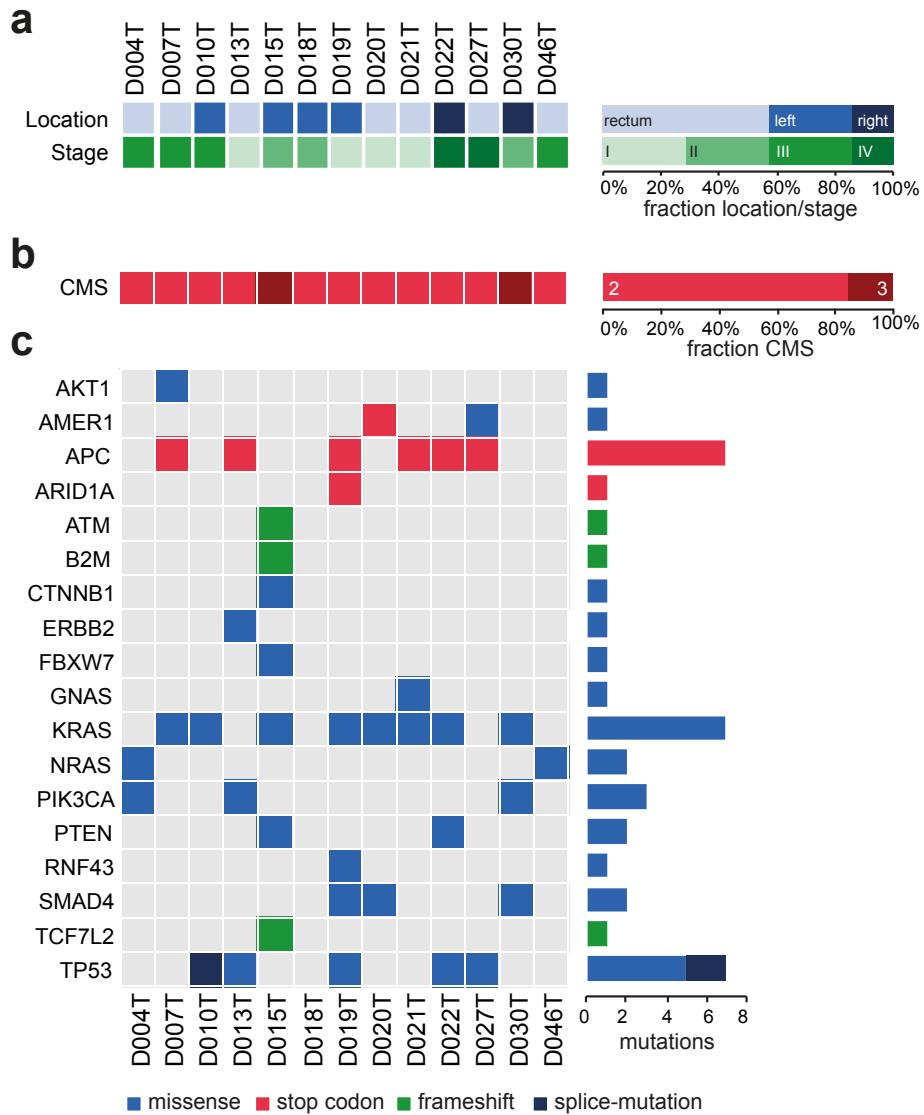


Figure 3.2: Organoid cohort overview a Tumor location (right/left/rectum) and AJCC/UICC stage of colorectal cancers that patient derived organoids were derived from. b Consensus molecular subtypes of organoids determined by RNA expression analysis. c Mutation status in PDOs, as analyzed by amplicon sequencing. Figure created with support from Johannes Betge (graphical presentation), Erica Valentini (sequencing data analysis) and Benedikt Rauscher (CMS type inference)

3.3 ENABLING METHODS FOR HIGH-THROUGHPUT IMAGE-BASED PROFILING OF ORGANOIDS

To systematically measure organoid morphology, I established a platform for high-throughput image-based profiling experiments. The three engineering problems that had to be solved were (1) control of organoid size and density in a 384-well plate format, (2) control of organoid location within the hydrogel for efficient microscopy and (3) maintenance of organoid integrity during automatic fixation and permeabilization.

Figure 3.3: Overview of experiments a Organoids were isolated from endoscopic biopsies from patients with colorectal cancer. Organoids were dissociated and evenly seeded in 384-well plates before perturbation with an experimental (464 compounds) and a clinical compound library (63 compounds à 5 concentrations, 842 perturbations across both libraries). After treatment, high-throughput fluorescence microscopy was used to capture the morphology of organoids. The multi-channel (DNA, beta-actin, cell permeability) 3D imaging data was projected, segmented, and descriptive features were extracted to quantify potential drug-induced phenotypes. Untreated organoid morphology, organoid size and drug activity scores were integrated with mRNA expression and mutation data in a Multi-Omics Factor Analysis (MOFA) to increase interpretability of organoid variation. Figure created with support from Johannes Betge (graphical presentation)

A standard protocol for cell-based assays, including image-based profiling, is seeding cells into microwell plates at a fixed cell number. In order to determine cell number, adherent cells are dissociated and counted using optical methods. Patient derived organoids, however, demonstrated a low rate of organoid outgrowth when passaged by complete organoid dissociation down to the single cell level. To improve organoid outgrowth, the dissociation protocol was stopped early, yielding cell clusters of ca. 1-10 cells. These organoid fragments showed an increased outgrowth rate, which could be further improved by treating cells with 10 µM of Rho-Kinase inhibitor Y-27632 (data not shown). To control organoid size and density, organoids were digested with a modified trypsin derivative, and filtered through a 40 µm cell strainer to ensure an upper limit of organoid fragment size. To effectively estimate the cell number while maintaining organoid fragments, organoid fragments were titrated based on their ATP concentration, instead of cell count. The ATP concentration of the organoid fragment suspension was determined using an ATP-dependent luminescence readout. After controlling for ATP concentration, the organoid fragment suspension was seeded onto basal membrane extract covered 384 well plates.

Conventional image-based profiling of adherent cells is based on automatic microscopy of one 2D plane per field of view. Given the 3D growth patterns of organoids, more data has to be acquired to fully capture organoid phenotype. Acquiring multiple planes of imaging data per field of view, however, creates a technical

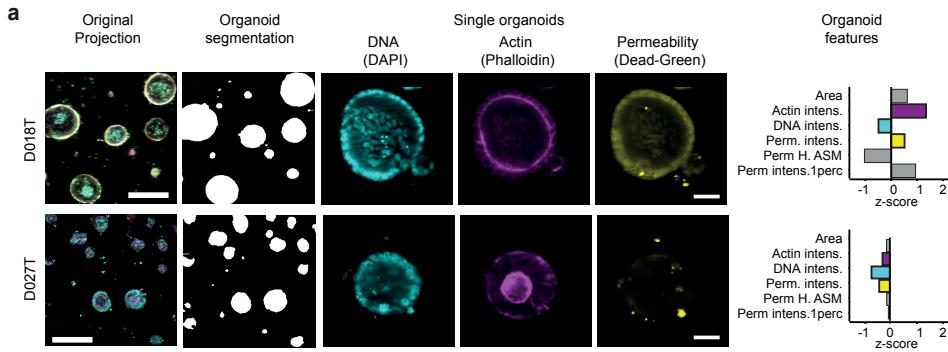


Figure 3.4: Image-based profiling method a The image-processing pipeline illustrated with representative example images from 2 organoid lines: Organoids were imaged at multiple layers along the z-axis. Images were projected using a maximum contrast projection and segmented using a convolutional neural network, both designed and implemented by Jan Sauer. Descriptive features were extracted from all three channels to quantify phenotypes. Feature plots show the median phenotype of unperturbed organoids, six example features (Area, Phalloidin intensity, DAPI intensity, FITC intensity, FITC Haralick angular second moment (ASM) and FITC intensity 1-percentile) and their z-scores relative to all profiled organoid lines are shown. Figure created with support from Jan Sauer (data processing) and Johannes Betge (graphical presentation)

data storage and processing burden. For a fixed 3D volume, the dimensions of the collected data increase linearly with the number of acquired planes and quadratically with the target z-axis resolution. To reduce the observed 3D volume and thus the number of required imaging planes, the vertical distribution of organoid fragments within the basal membrane extract layer was controlled by centrifuging organoid fragments post-seeding at 500G for 20 minutes at 37 degrees Celsius. The centrifugal force led to an accelerated sedimentation of organoid fragments onto the same optical plane before the polymerization of the hydrogel was complete.

After three days of culture and four days of compound treatment, organoids were fixed and stained for actin (Phalloidin/TRITC), DNA (DAPI), and cell permeability (DeadGreen/FITC). Subsequently, plates were imaged at multiple z-positions by automated confocal microscopy. During treatment with the hyperosmolar fixative (3% para-formaldehyde in phosphate buffered saline) the protein rich hydrogel underwent an irreversible volume contraction (data not shown). To reduce this artefact, the fixative was supplemented with bovine serum albumin to a final concentration of 1% weight/volume. After image acquisition, 3D data was

projected into a 2D plane by applying a maximum contrast projection followed by segmentation with a weakly-supervised convolutional neural network and single-organoid-level feature extraction 3.4. In summary, seeding well-quantifiable organoid fragments instead of single cells, centrifuging organoid fragments to reduce the imaged 3D volume, and modifying liquid handling buffers to avoid hydrogel-driven artefacts technically enabled high-throughput image-based profiling of organoid models.

3.4 IMAGE-BASED PROFILING CAPTURES THE MORPHOLOGICAL DIVERSITY OF PATIENT-DERIVED CANCER ORGANOID

To better understand the diversity of organoid phenotypes and how morphology links to molecular processes, I performed image-based profiling at single organoid resolution with 11 organoid models using compounds targeting developmental pathways (464 compounds), as well as compounds in clinical use (63 compounds in 5 concentrations) 3.5. The resulting data comprised morphological profiles for each organoid with 528 phenotypic features that were subsequently reduced into 25 principle components representing 81% of morphological variance.

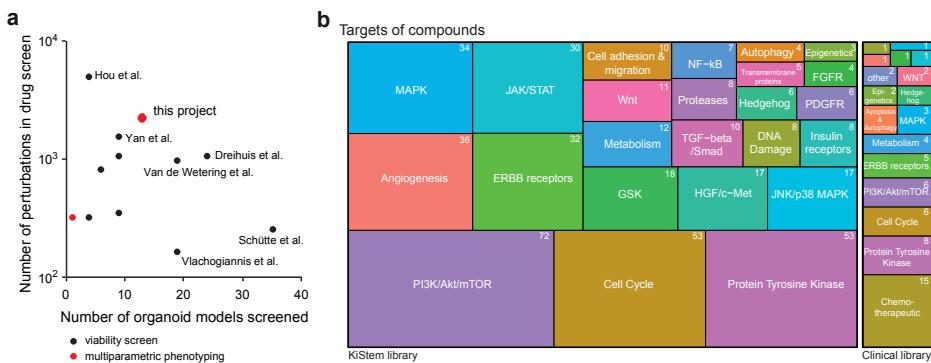


Figure 3.5: Dataset dimensions and compound library overview a Number of organoid models and number of perturbations in previous publications reporting high-throughput drug screenings with patient derived cancer organoids, b Graphical representation of the compound libraries used for drug screening in this project: A library targeting kinases and stem cell pathways (KiStem library, 464 compounds) and a clinical library with 63 drugs in 5 concentrations. Figure created with support from Johannes Betge (graphical presentation)

To visualize the heterogeneity of colorectal cancer organoids and drug induced changes across and within cancer organoid lines, the features of ca. 5.5 million profiled organoids were embedded using uniform manifold approximation and projection (UMAP) (figure 4.7 a and 3.7 a-c). Most organoid lines showed characteristic bimodal log-normal distributions of organoid size with one component containing small organoids and another component made up of larger organoids with varying, line specific, average size (figure 4.7 b, and 3.7 d-e). The log-normal-like size distribution likely resulted from intrinsic differences in cellular size and growth rate compounding over time in multicellular organoids. While DNA and actin staining intensity were positively correlated with organoid size, cell permeability was negatively correlated and enriched in regions with relatively smaller organoids (3.7 a-c). Graph-based clustering of this identified 12 regions within the embedding (figure 4.7 c). When comparing drug-treated organoids to organoids treated with the negative control (DMSO), no clear separation of these two groups, except an increased presence of drug-treated organoids in region 3, was seen. This finding suggested that organoid morphology was distributed on a continuum of phenotypes spanning perturbed and unperturbed conditions of the experiment (figure 3.7 f).

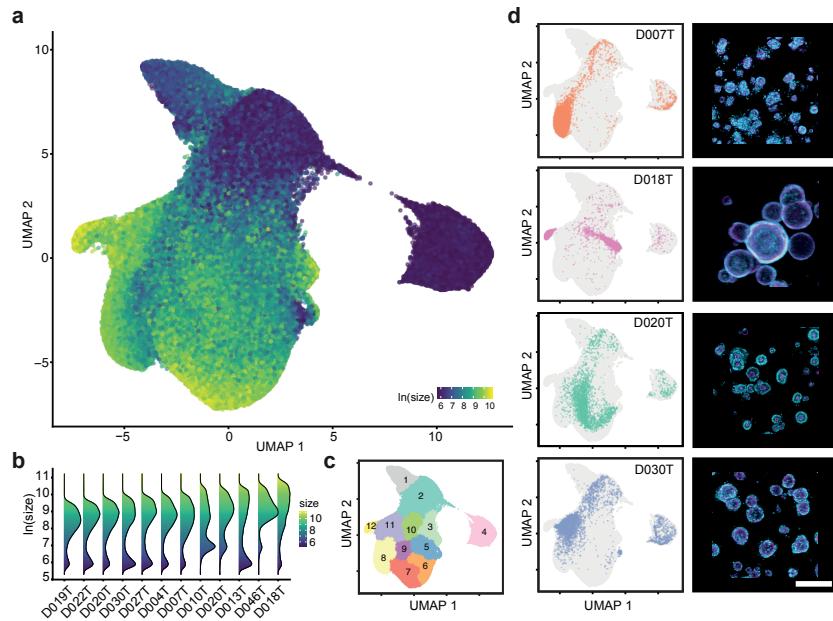


Figure 3.6: Image-based profiling captures the phenotype diversity of patient derived cancer organoids **a** Uniform Manifold Approximation and Projection (UMAP) of organoid-level features for a random 5% sample out of ca. 5.5 million organoids. The same sample is used for visualizations throughout the figure. Color corresponds to the log-scaled organoid area (dark blue: minimum size, yellow: maximum size). **b** organoid size distribution across lines. **c** UMAP representation of DMSO treated and drug treated organoids. Graph-based clustering of organoids by morphology. **d** UMAP embeddings of selected organoid lines (baseline state / 0.1% DMSO control-treated organoids) representing different morphological subsets, grey background consists of randomly sampled points. Depicted are representative example images for each line (right, cyan = DNA, magenta = actin, scale-bar: 200 μ m).

Different organoid lines within the embedding were located in characteristic regions, with organoid size and organoid architecture as primary organizing factors (figure 4.7 b and d). For example, organoid line Do18T had the largest median organoid size within the dataset and a cystic organoid architecture, while Do20T organoids had a solid architecture and smaller median size. In most cases, organoid lines had two areas of main density, with one of them in regions 2, 3 or 4, reflecting the previously mentioned bimodal size distribution. In summary, image-based profiling of patient derived colorectal cancer organoids showed strong morphological heterogeneity with line dependent differences in size and organoid architecture.

Exploratory data analysis of the relationship between organoid morphology and batch showed overall

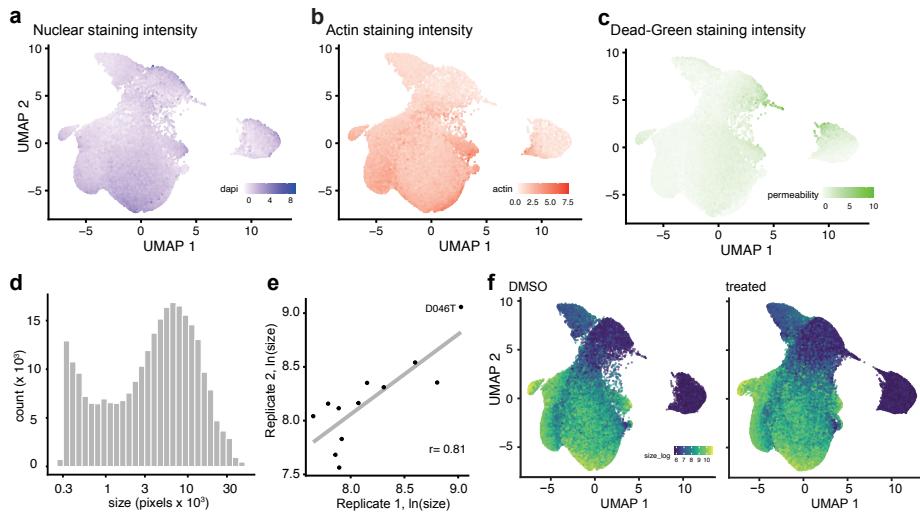


Figure 3.7: Basic image-based features and their role in organoid phenotype diversity. a-c Uniform Manifold Approximation and Projection (UMAP) of organoid-level features marked by DNA (DAPI) staining intensity (b), actin (Phalloidin/FITC) staining intensity (c) and permeability (DeadGreen) staining intensity d Distribution of organoid size for all control (DMSO) treated organoids. e Replicate correlation of organoid size for control treated organoids. f UMAP representation of DMSO treated and drug treated organoids

reproducible measurements of organoid profiles across experiments. While objects with a log-area of 8 pixels and larger showed reproducible phenotypes across contexts, smaller objects (mostly dead organoids) showed batch-dependent differences in phenotype. For example, region 1 within the UMAP embedding was exclusively occupied by observations from batch HC1092-09 and HC1092-10, while region 4 was relatively underoccupied. Given the confounding of line differences by experimental batches (experimental batches and tested organoid lines were not independent) and the stronger prevalence of batch effects for small objects, no procedure to remove these batch-dependent differences in organoid phenotype were performed.

a

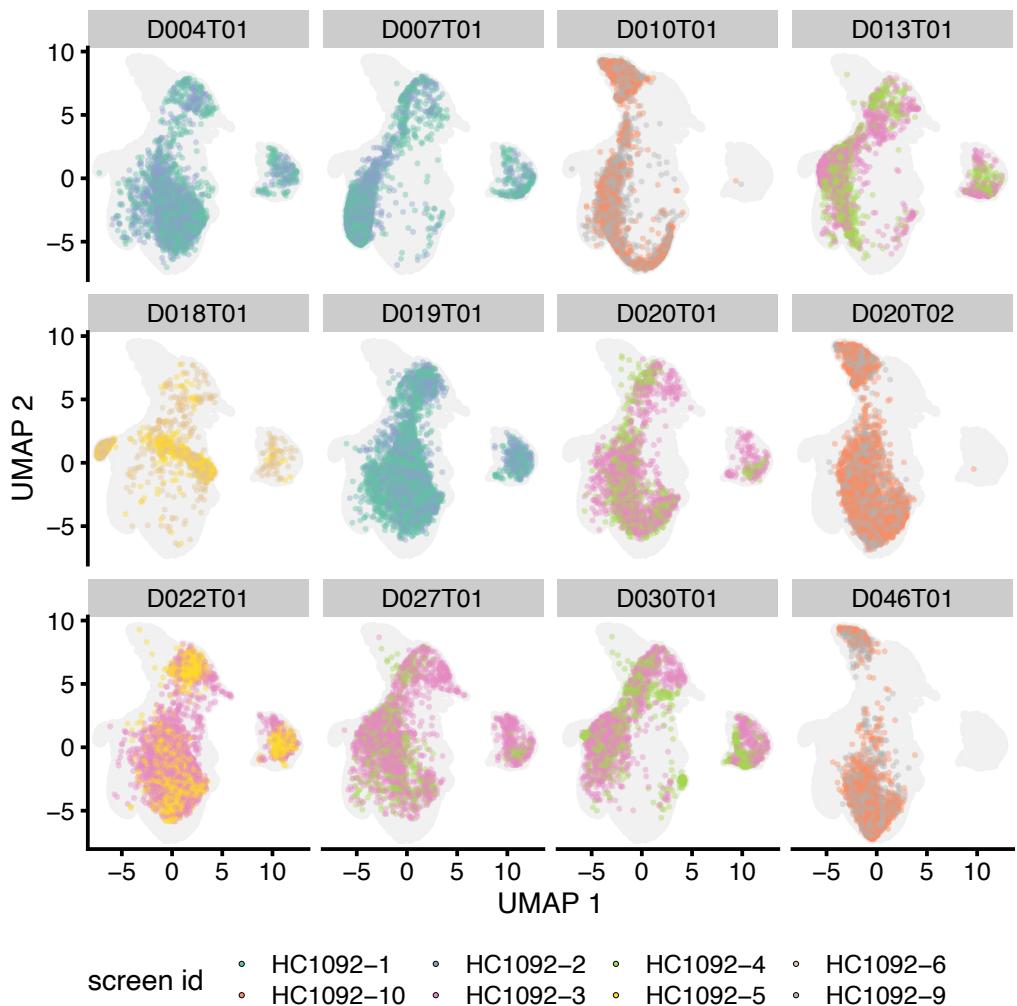


Figure 3.8: Technical confounders and their impact on organoid phenotype. a UMAP of organoid level features stratified by organoid line and colored by experimental batch.

3.5 ORGANOID PHENOTYPE-PROFILES CAPTURE ORGANOID VIABILITY

Drug induced changes in cell viability are a basic readout in oncology drug discovery. Prompted by the observation that organoid size was a major factor determining the structure of the phenotype embedding (UMAP and factor 1 in MOFA analysis, see below), I hypothesized that low organoid size was at least partially the result of cell death within the organoid and, more broadly, that phenotype data could be used to estimate organoid viability. Bortezomib, a small molecule proteasome inhibitor with high in vitro toxicity led to dose dependent organoid death in all organoid lines, thus representing suitable positive controls (Fig. 2a). Analogous to pseudotime in single-cell gene expression analysis, dose-dependent trajectories of bortezomib drug response could be fitted (Fig. 2b) using the non-parametric principle curve method. Starting from diverse baseline morphologies, increasing doses of this compound led to a step-wise convergence on a final death-related phenotype, which corresponded to the areas with enrichment of small objects (regions 2, 3 and 4).

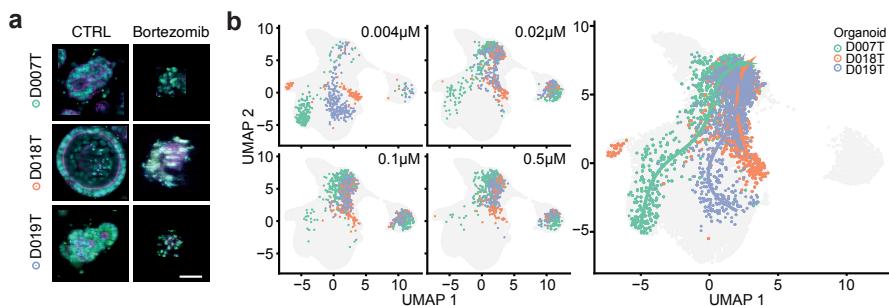


Figure 3.9: Organoid phenotype-profiles capture organoid viability. a Representative example images of negative- (0.1% DMSO) and positive control treated organoids (2.5 μ M bortezomib, cyan = DNA, magenta = actin, yellow = cell permeability; average images were selected and embedded in black background; scale bar: 50 μ m). b, Dose-dependent-trajectory of bortezomib drug effect. UMAP of organoid morphology at different bortezomib doses and (right panel) dose-dependent trajectory for three representative organoid lines. For visual purposes, trajectory inference was limited to partition 1, the left-hand set of measurements within the UMAP, representing ca. 95 % of all imaging data.

Similarly, paclitaxel, a microtubule disassembly inhibitor, shifted the bimodal size distribution of organoids in a dose-dependent fashion (figure 4.4a), while organoid count remained largely unchanged (figure 4.4b). This

effect, however, was organoid line-specific, as median organoid size in paclitaxel sensitive lines (e.g. Do22T) decreased, while the size of other organoids remained unaffected (e.g. Do46T, Fig. 2c-f). These observations suggested a link between organoid morphology, especially organoid size, with a loss of cell viability. To further test the link between organoid morphology and cell viability, I performed a luminescence-based, ATP dependent, cell viability assay (CTG) in parallel with imaging as benchmark. A strong association of CTG viability with organoid size (Fig. 2g) was visible.

To test whether a more accurate prediction of organoid viability was achievable by using all available imaging data, I used a previously trained set of random forest classifiers (live/dead classifiers, LDC). These classifiers were trained on individual organoid phenotype profiles to distinguish between negative and positive control treatments (DMSO, bortezomib and sn-38, Supplemental Fig. S3d-e). When applying the classifier to the whole imaging dataset and visualizing predictions via UMAP, organoids within previously identified regions 2, 3 and 4 had the highest probabilities for death (Fig. 2h). LDC had the highest correlation with CTG based viability data (Fig. 2i), however, the association with organoid size was almost as strong in the majority of organoid lines (Fig. 2i, Supplemental Fig. S3g), while other simple features, such as DAPI actin, and permeability (DeadGreen) intensity were less suitable to predict viability of organoids (Fig 2i). In conclusion, organoid size is an informative metric to describe organoid viability, but is biased by line-specific differences in untreated organoid size. Models consuming more comprehensive morphological information can achieve even higher predictive performance of organoid viability.

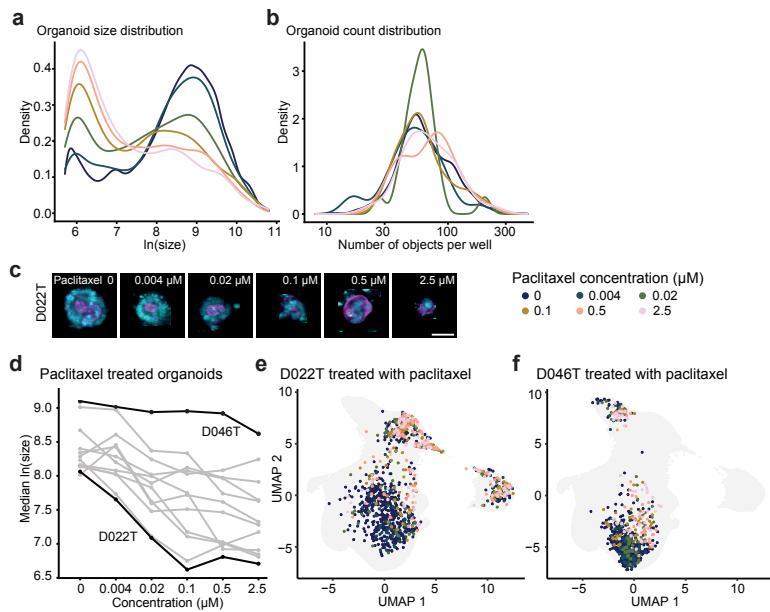


Figure 3.10: Organoid phenotype-profiles capture treatment specific changes in organoid viability. **a** Distribution of organoid size at different concentrations of paclitaxel. Shown is a random sample of 30% of all paclitaxel treated organoids for this and following figures. **b** Distribution of organoid number per well at different concentrations of paclitaxel. **c** Example images of D022T organoids treated with paclitaxel. **d** Dose-response relationship of organoid size and paclitaxel dose. D022T and D046T are highlighted. **e** UMAP of organoid morphology highlighting D022T organoids treated at different concentrations of paclitaxel. **f** UMAP of organoid morphology highlighting D046T organoids treated at different concentrations of paclitaxel.

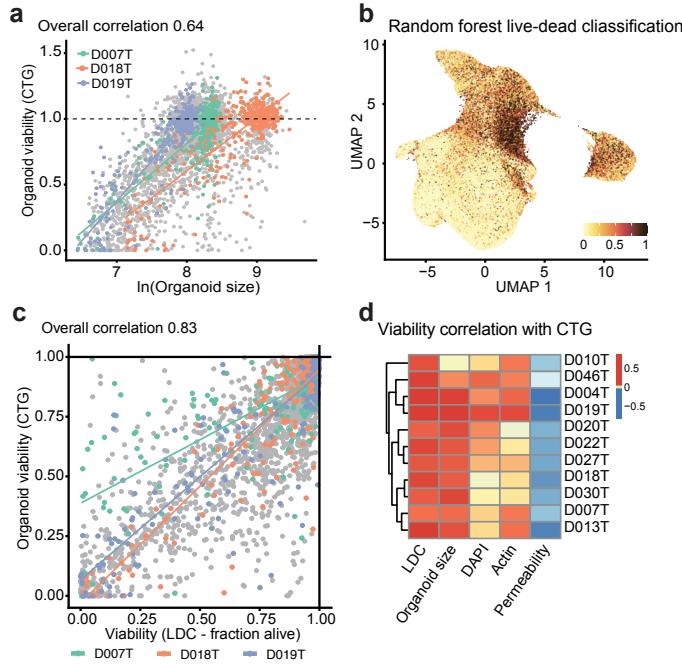


Figure 3.11: Organoid phenotype-profiles reflect ATP-dependent viability measurements. **a** Association of organoid size of selected example lines with organoid viability determined by luminescence-based, ATP-dependent viability profiling with CellTiter-Glo (CTG), which was performed in parallel with imaging on a subset of drug treatments. **b** UMAP visualization of dead organoids within our dataset, based on supervised machine learning of organoid viability using classifiers trained on positive- (high-dose bortezomib and SN-38) and negative (DMSO) controls (live-dead classifiers, LDC). **c** Association of organoid size of selected example lines with organoid viability determined by luminescence-based, ATP-dependent viability profiling with CellTiter-Glo (CTG), which was performed in parallel with imaging on a subset of drug treatments for benchmarking. **d** UMAP visualization of dead organoids within our dataset, based on supervised machine learning of organoid viability using classifiers trained on positive- (high-dose bortezomib and SN-38) and negative (DMSO) controls (live-dead classifiers, LDC). **e** Association of LDC and example organoid features (size, DAPI, actin and permeability dye intensities) with benchmark CTG viability read out. Figure created with support from Jan Sauer (LDC classifier training)

3.6 DRUG INDUCED ORGANOID PHENOTYPES CORRESPOND TO DRUG MECHANISM OF ACTION

An advantage of image-based phenotyping over cell viability measurements in drug discovery is the ability to use the high dimensional drug-induced phenotype-profiles to identify active but not necessarily lethal drugs and estimate their mechanism of action by unsupervised clustering. To test whether this approach could be used in cancer organoids, we used a weakly supervised learning approach to identify drug effect vectors and group them by similarity. First, we trained logistic regression models to separate individual compound-treated

organoids from unperturbed controls and used the resulting normal vector between control- and treated organoid profiles as the drug effect vector. Next we scored every model's ability to separate treated and untreated organoids (AUROC, ranging from 0.5 to 1) to identify active treatments that induce a robust change in organoid morphology (Fig. 3a, 3b). Based on our observations, drug activity was necessary but not sufficient for a viability effect (Fig. 3c). A fraction of drugs led to identifiable changes in organoid morphology but were not classified as lethal by our LDC model.

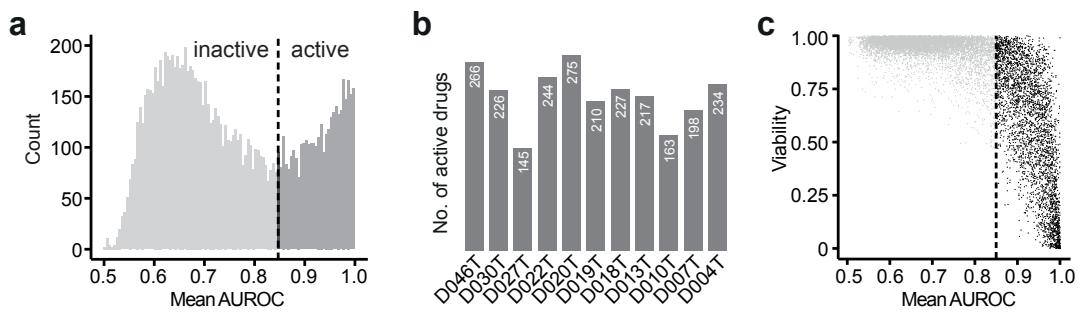


Figure 3.12

To test whether active drugs systematically induce organoid phenotypes that are informative of mechanism of action, we calculated the cosine distance between concatenated drug effect vectors. We identified a clustering of specific mode-of-actions, including inhibitors of MEK, Aurora kinase, CDK, mTOR, AKT, EGFR or GSK₃ which we also identified with an alternative computational approach, phenotype profile averaging followed by pearson correlation (Fig. 3d-h, Supplemental Fig. S4a-c, S5a-c). Furthermore, compounds with related targets also induced similar phenotypes, for example MEK inhibitors clustered with specific RAF- and ERK inhibitors (Fig. 3e) or AKT and PI₃K inhibitors were part of a cluster mainly containing mTOR targeting compounds (Fig. 3f, Supplemental Fig. 5a). The clustering also suggested additional mode-of-actions or off-target effects for well-described compounds (Fig. 3g-h). For example, the PKC inhibitor enzastaurin was related to GSK₃, substantiating a previously described interaction with the alpha

and beta subunits of GSK3 β (Fig. 3h). To assess whether morphological profiles of active drug treatments were primarily driven by differences in organoid viability, we compared LDC predictions with the phenotypic clustering (Fig. 3i). We observed a larger cluster of lethal treatments (including molecules targeting ATM, JAK, PLK, CDK). However, the majority of clusters were caused by non-lethal phenotypes, including those induced by inhibitors of AKT, mTOR, EGFR or GSK3. Visual inspection of several phenotypes (Fig. 3j) revealed recurring drug target dependent phenotypes. Most notably, MEK inhibitors led to reorganization towards more cystic organoid architecture. These drug target dependent phenotypes were observable across organoid lines and drugs.

3.7 MULTI-OMICS FACTOR ANALYSIS IDENTIFIES SHARED FACTORS LINKING MORPHOLOGY, GENOMIC DATA AND DRUG ACTIVITY

A limitation of image-based profiling experiments is that both unperturbed and drug induced phenotypes are challenging to interpret in terms of their underlying biology. Theoretically, in the presence of multiple *in vitro* models with both phenotype and genomic measurements, links between the two data modalities can be learned. Based on the observation that organoid morphology was distributed in a continuous space, I hypothesized that variation in organoid baseline morphology could be associated with differences in gene expression, mutations, as well as drug activity for the 11 cancer organoid lines in our sample. To learn a joint representation of unperturbed organoid morphology, unperturbed organoid size, gene expression, somatic mutations, and drug activity, multi-omics factor analysis (MOFA) was performed. MOFA is a matrix factorization method that decomposes a set of different measurements into a shared table of factors scoring each observed sample and a set of corresponding tables linking each factor to features in the set of original measurements. When trained with a low number of $k = 3$ factors, MOFA recovered factors explaining ca.

41-24% of variance across the different data modalities, with the first two factors accounting for ca. 29-17% in aggregate (Supplemental Fig S6a-b). While gene expression, mutations and drug activity profiles for organoid lines contributed to all factors, factor 1 captured an exceptional amount of variation in median organoid size (ca. 39%). In contrast, factor 2 was primarily capturing variation within baseline organoid morphology (ca. 16%) (Fig 4a). Overall, MOFA factors explained up to 40% of variance in median organoid size, drug activity and gene expression, while less than 30% of variance in baseline organoid morphology was explained by the model (Fig 4b). Organoid lines Do46T and Doo4T stood out as lines with the strongest score for factor 1, while lines Do18T and Do13T had the strongest score in factor 2. Visual inspection of organoids revealed that organoid lines with a higher factor 1 score tended to be larger in size and organoids with high factor 2 score tended to have a more cystic organoid architecture based on manual classification. No interpretable morphological differences between factor 3 low and high organoids was identifiable, so the subsequent analysis was focused on the first two interpretable factors generated by MOFA (Supplemental Figures S6c). Visualising factor scores for untreated organoid lines within the UMAP embedding showed characteristic regions within the representation space linked to factors (Fig 4d). Factor 1 high lines were located mostly in region 7 (the bottom part of the embedding) while factor 2 and 3 positive lines were located in region 3, 10, 11, 12 (the middle part of the embedding) and region 9 (a central part of the embedding), respectively. To summarize, MOFA identified factors within the dataset that explained variation between organoid lines across different data modalities, including organoid morphology and median organoid size.

3.8 AN IGF1R SIGNALING PROGRAM IS ASSOCIATED WITH INCREASED ORGANOID SIZE, DECREASED EGFR INHIBITOR ACTIVITY AND CAN BE INDUCED BY mTOR INHIBITION

Differences in organoid size are an obvious contributing factor to intra- and inter-organoid line heterogeneity. Organoid size was influenced by both organoid line and drug treatments and was associated with factor 1 scores (Fig 6a). An unsupervised gene set enrichment analysis (GSEA) for reactome pathways across factor 1 loadings showed an enrichment for IGF1R signaling and mitogen-activated protein kinase signaling related genes. In fact, the IGF signaling related transcripts H19 (rank 1) and IGF2 (rank 13) were among the strongest contributors to factor 1. This increase in proliferative signaling was confirmed by GSEA of a previously identified intestinal proliferation signature.³¹ To better understand clinical correlates to the identified gene

expression patterns, we tested for molecular subtypes stemming from an analysis of cancer-cell intrinsic gene expression profiles.³² Factor 1 showed an enrichment for CRIS D, a molecular subtype linked to IGF2 overexpressing tumors with resistance to EGFR inhibitor therapy (Fig 6c), and a depletion for CRIS C, which has been linked to EGFR dependency (Supplemental Figure S8a). In fact, activity of EGFR inhibitors was the strongest contributor to a negative factor 1 score while IGF1R and MEK inhibitor activity contributed to a positive factor 1 score (Fig 6d-e, Supplemental Figure 8b-d). Prompted by the observation that mitogenic signaling, including IGF1R signaling, was underlying factor 1, we hypothesized that other compound treatments could influence the plasticity between the observed organoid states by modulating signaling pathway activity within organoids. To test whether drug treatments shifted organoid phenotype profiles in factor space, I took advantage of the previous observation that unperturbed and certain perturbed organoids shared similar phenotypic profiles. To this end, the previously estimated factor loading matrix for unperturbed organoid morphology, which was generated during MOFA training, was used as a starting point. By generating the pseudoinverse of the loading matrix and multiplying with average phenotypic profiles of drug-treated organoids, the influence various drug treatments had on biological programs previously identified in unperturbed organoids was approximated. A group of cell cycle related kinase inhibitors targeting polo like kinases, Aurora kinases and cyclin dependent kinases shifted organoids to a low factor 1 score. In contrast, mTOR inhibitor treatment increased factor 1 scores in cancer organoids (Fig 6f and Supplemental Figure S8e). Given the observation that factor 1 was associated with IGF-1R signaling and mTOR inhibitor treatment led to an increase in factor 1 scores, I hypothesized that mTOR inhibition leads to a reactive upregulation of IGF1R signaling in cancer organoids. In fact, inhibition of mTOR signaling had previously been linked to transcriptional disinhibition of IRS-1 in a negative feedback loop³³ and reactive induction of IGF1R signaling had previously been described as a resistance mechanism to small molecule mTOR inhibitors in cancer.³⁴ When testing this hypothesis in patient derived organoids, a dose-dependent increase of IRS-1 protein abundance in organoids treated with the ATP competitive mTOR inhibitor WYE-132 was observable (Fig 6g). To summarize, factor 1 described an organoid state with relatively large organoid size, elevated IGF1R dependent mitogenic signaling and relative inactivity of EGFR inhibitor treatment that could be induced by inhibiting an mTOR dependent negative feedback loop in patient derived cancer organoids.

3.9 AN LGR₅+ STEMNESS PROGRAM IS ASSOCIATED WITH CYSTIC ORGANOID ARCHITECTURE AND CAN BE INDUCED BY INHIBITION OF MEK

A particularly strong recurring organoid phenotype was the presence of a cystic organoid architecture, seen in untreated Doi8T organoids and organoids treated with MEK inhibitors (Fig 1e, 3f, 5a). In the cystic state, which was observed in factor 2 high organoid lines, organoids consisted of a monolayer of uniform cells lining a central spherical lumen with a distinct apico-basally oriented actin cytoskeleton (Fig 5b). This phenotype was reminiscent of organoid morphologies previously seen in APC-/ or Wnt ligand treated human intestinal organoids. To test if factor 2 comprised Wnt signaling and intestinal stem cell identity related gene expression programs, gene set enrichment analyses (GSEA) was performed for cell identity signatures previously identified in intestinal crypts and colorectal cancer. GSEA revealed an enrichment of Lgr₅+ stem cell signature-related genes for the factor 2 loadings (FDR=0.002, NES=1.74) (Fig 5c and Supplemental Figure S7a).³¹ Next, we wondered whether factor 2 was associated with particular drug activity or inactivity patterns. Activity of Wnt signaling inhibitors and EGFR inhibitors were the strongest average contributors to a positive factor 2 score (t statistic = 3.02, FDR = 0.046 and t statistic = 3.08, FDR = 0.046, respectively), while activity of ERK and MEK inhibitors were associated with a low factor 2 score (Fig 5d), albeit not significantly. As expected from these results, factor 2 high organoid lines showed a stronger morphological response to the Wnt pathway inhibitor PRI-724. (Fig 5e and Supplemental Figure S7b). Next, we again used phenotype profiles of drug treated organoids and approximated how drug treatment shifted organoids along the factor 2 program. We observed MEK and focal adhesion kinase inhibitors significantly shifting all tested organoid lines towards higher factor 2 scores (Fig 5f and Supplemental Figure S7c). This change in factor 2 scores was concentration dependent for MEK inhibitors (Fig 5g and Supplemental Figure S7d-e) and coincided with a visual shift in organoid morphology (Fig 5h). Given the observation that factor 2 was enriched for an LGR₅+ stem cell signature, we measured the expression of LGR₅ transcripts at different concentrations of MEK inhibitor treatment and observed analogous dose-dependent increases in transcript abundance. In summary, factor 2 represents an organoid state with cystic architecture, increased expression of LGR₅+ stem cell related genes and increased sensitivity to Wnt signaling inhibitors that could be induced by MEK inhibition.

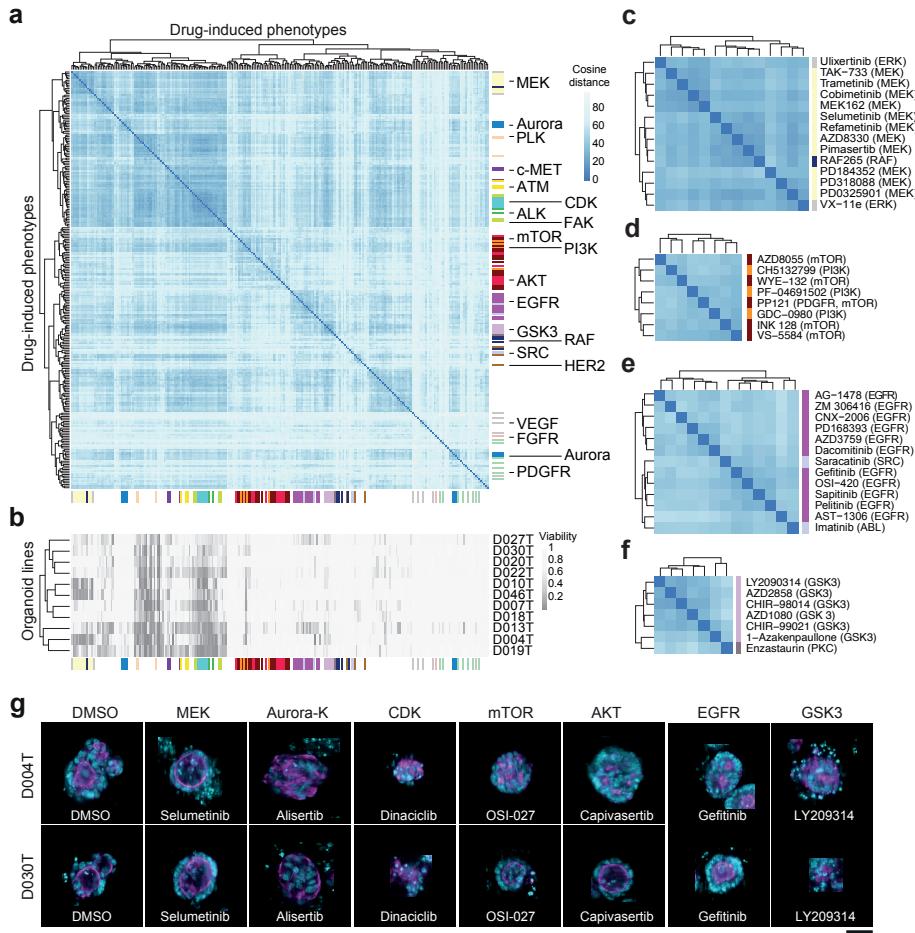


Figure 3.13

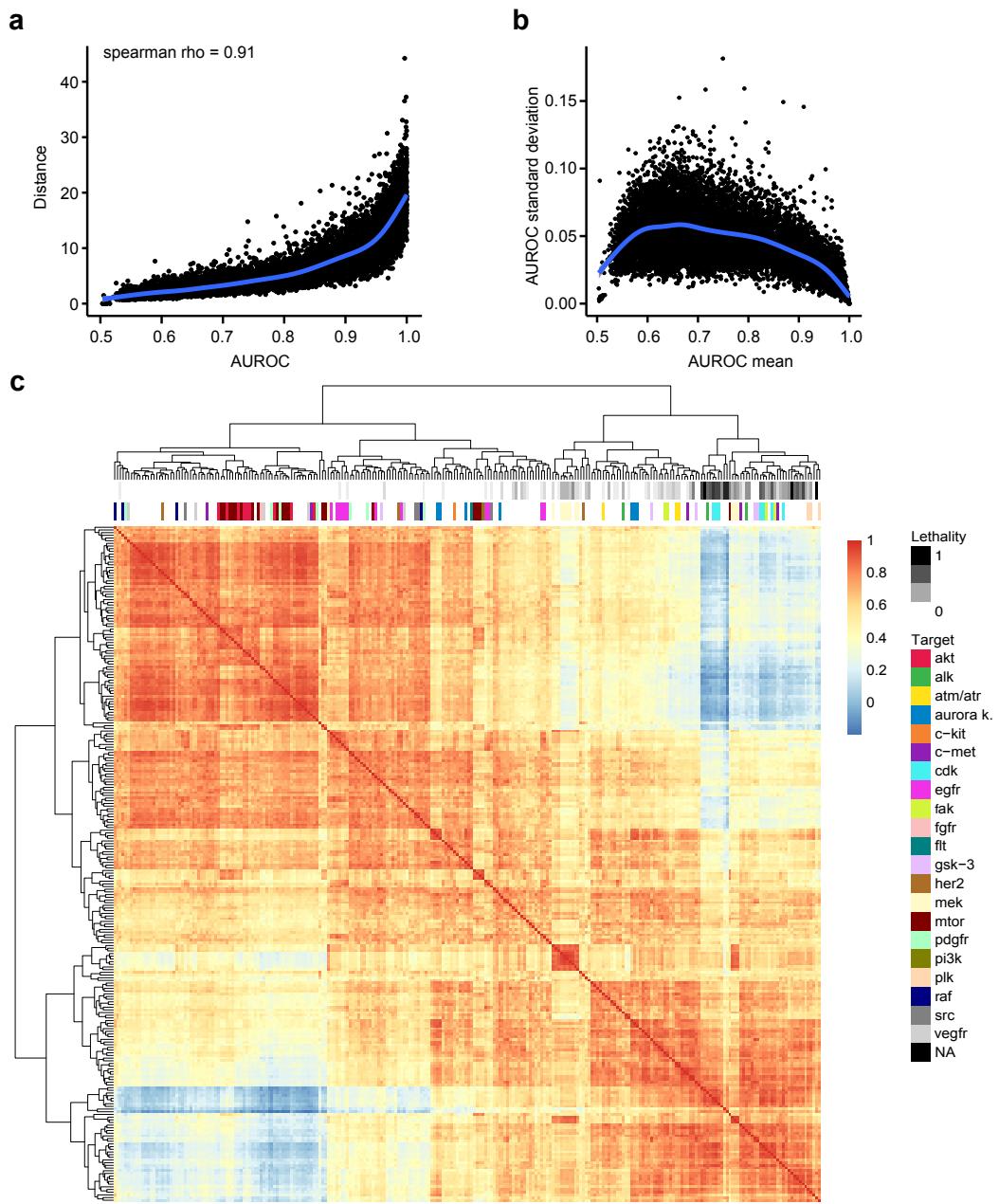


Figure 3.14

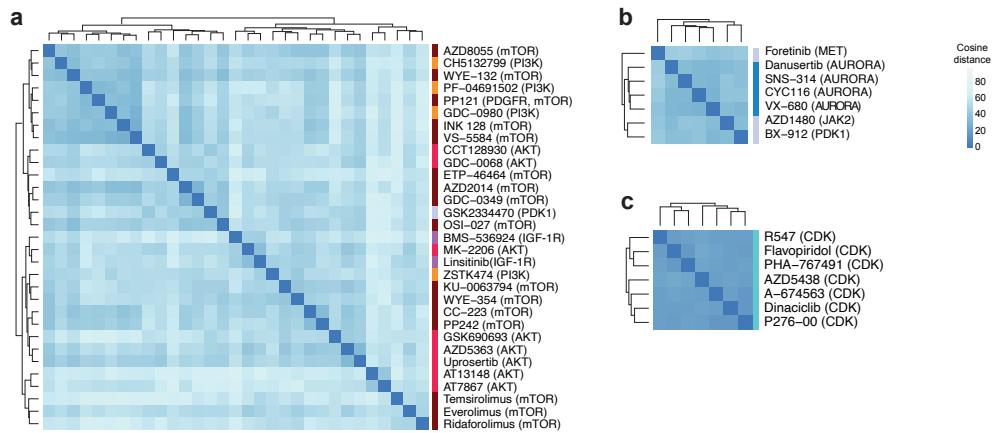


Figure 3.15

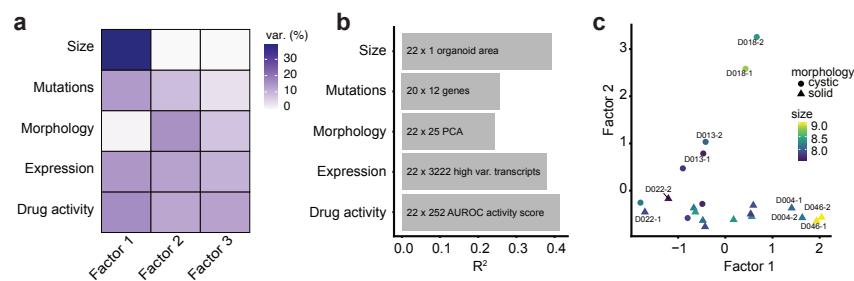


Figure 3.16

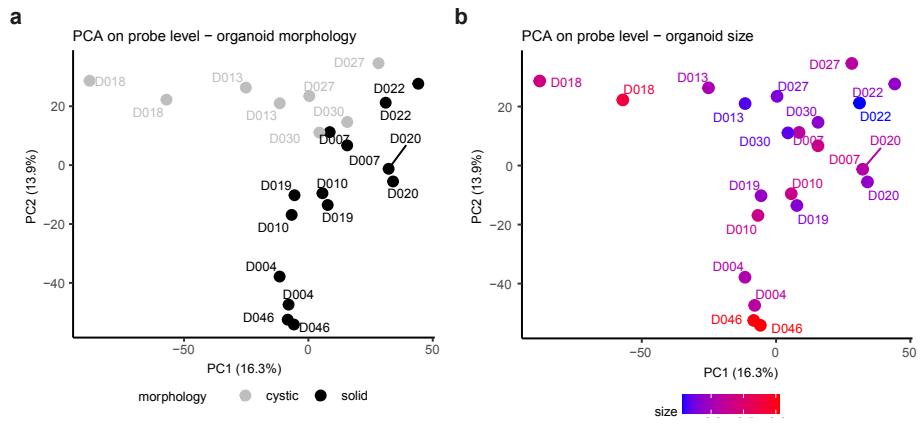


Figure 3.17

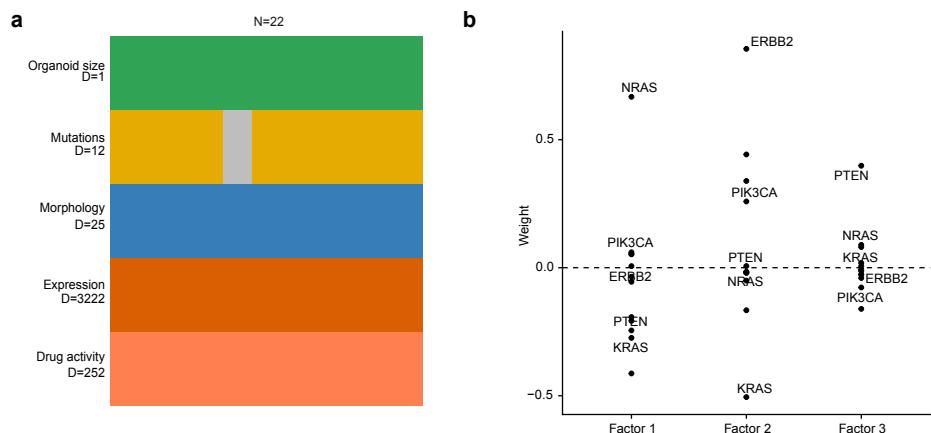


Figure 3.18

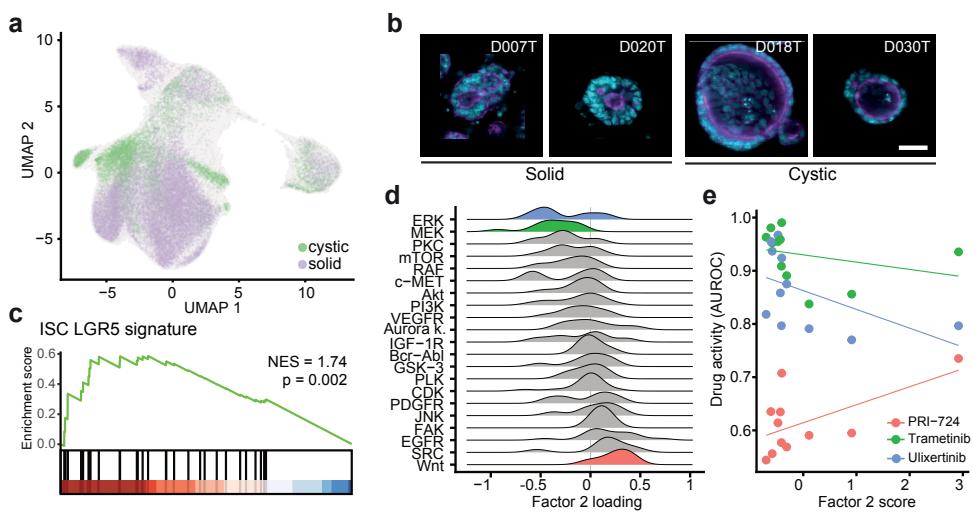


Figure 3.19

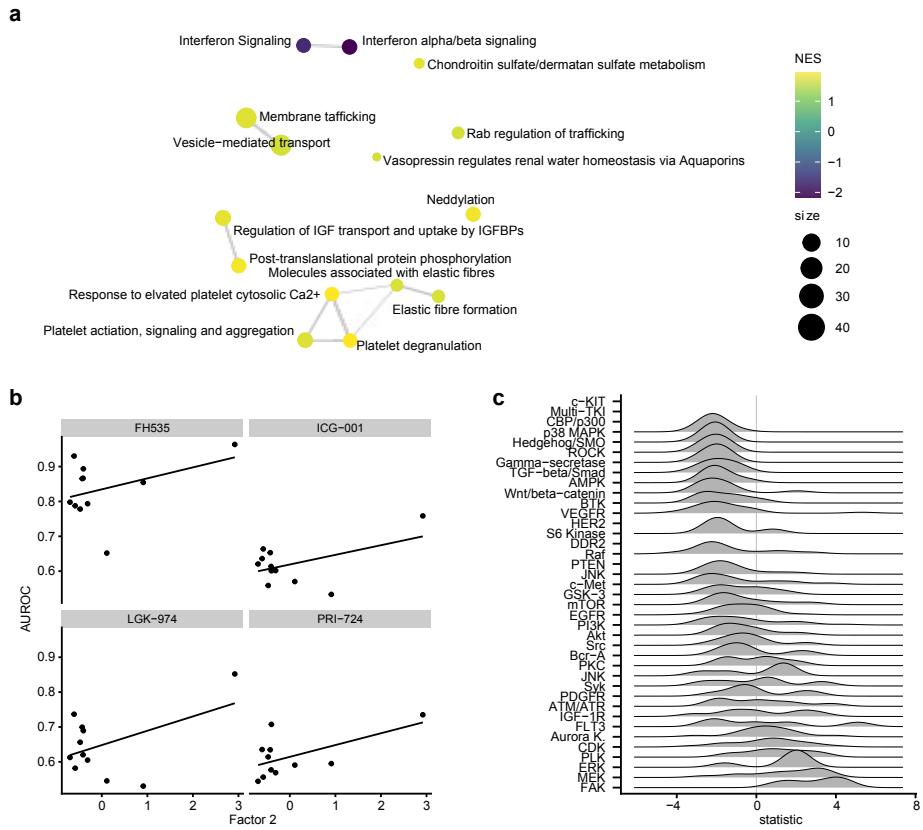


Figure 3.20

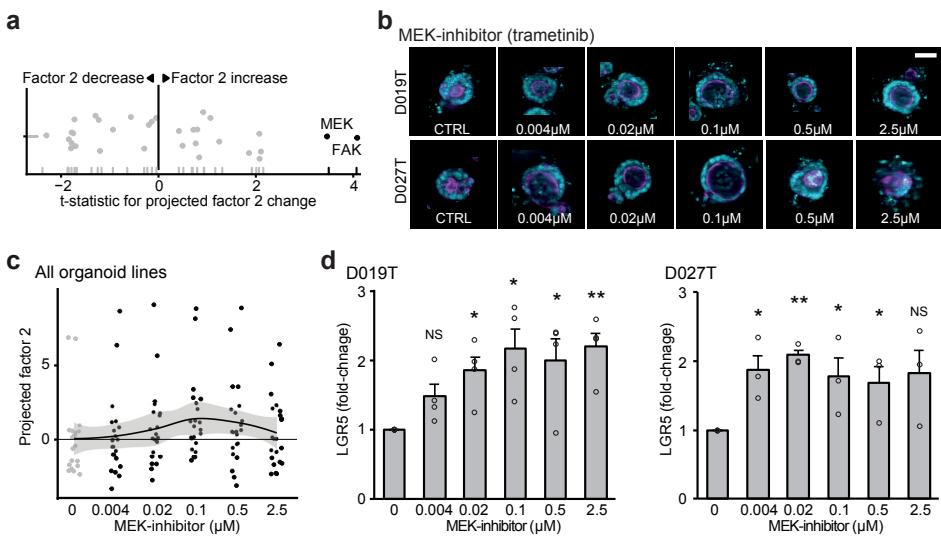


Figure 3.21

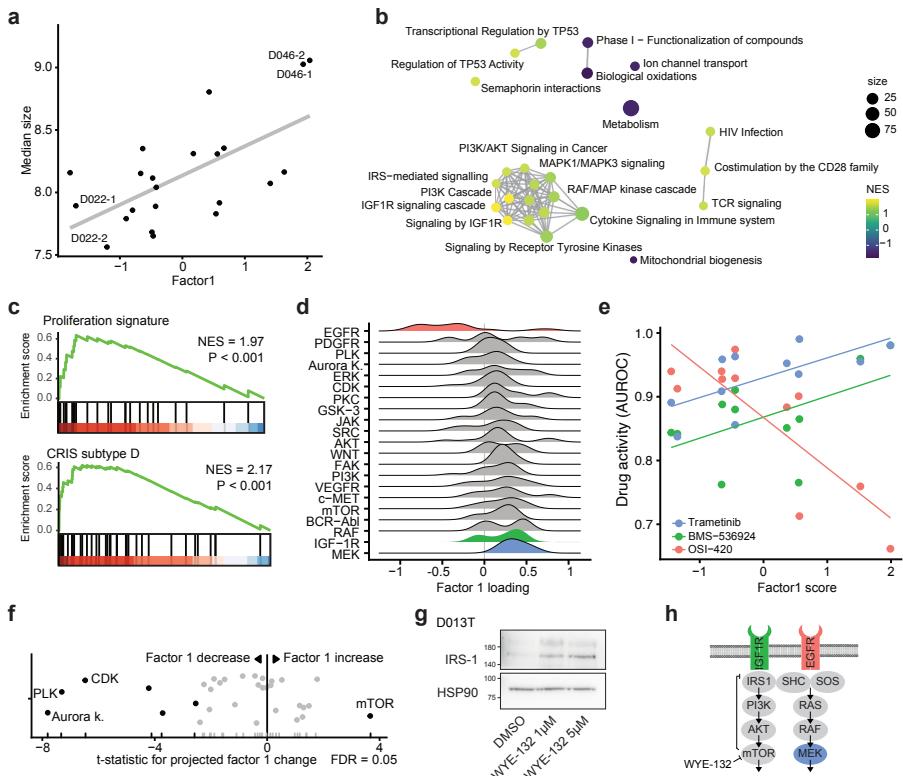


Figure 3.22

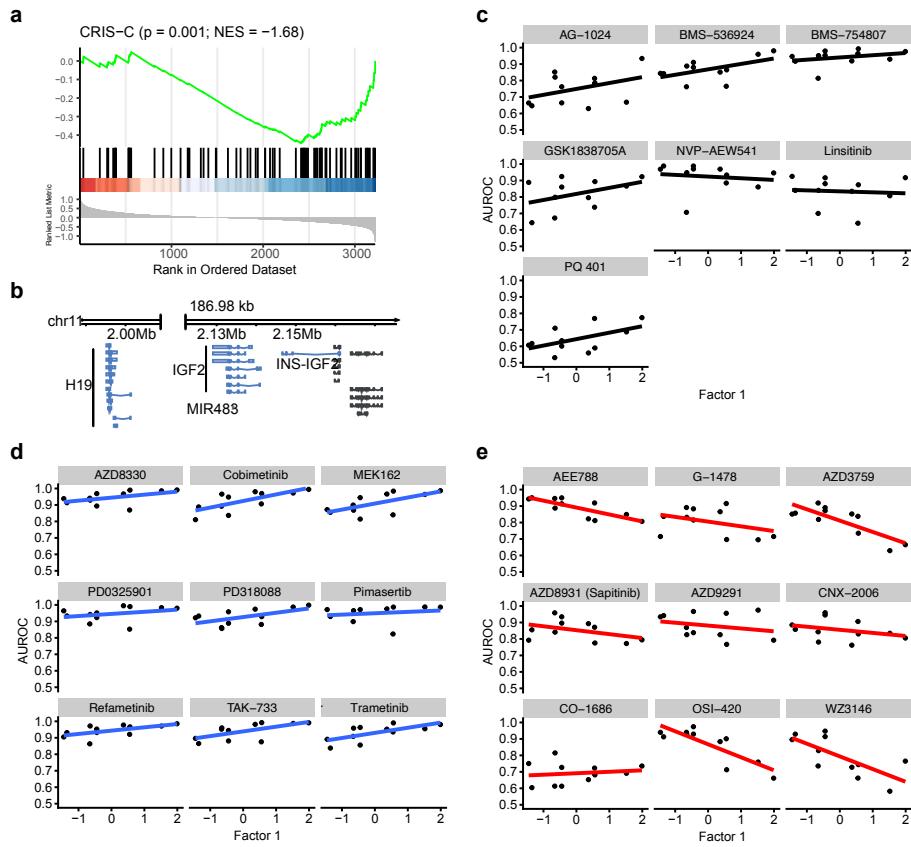


Figure 3.23

What I cannot create, I do not understand.

Richard Feynman

4

Image-based profiling to identify revertant therapeutics in pre-malignant models of colon cancer

4.1 MOTIVATION

Based on the observation that (1) joint representations of organoid morphology and biochemical state can be learned, and (2) small molecule perturbations can shift organoids in representation space, the following

hypothesis can be formulated: If well-annotated small molecules can help identify a causal representation of unknown organoid states, unknown small molecules with desirable properties can be identified by their ability to shift the state of well-annotated organoid models. Put differently, instead of using small-molecule perturbations to learn about the biology of non-characterized patient derived organoids, we should be able to use genetically engineered organoid models to learn about desirable properties of non-characterized small molecules.

The emergence of colorectal cancer via the chromosomal instability process is a well understood sequence of genetic events that start with hyperactivation of canonical Wnt signaling, often via loss of APC, followed by the hyperactivation of RAS-MAPK signaling, often via oncogenic mutations of KRAS. These two mutations are frequent and significantly co-occurring in colorectal cancer patients, suggesting an interplay between these two acquired genetic functional events that lead to a selfish evolutionary advantage of premalignant cells. While the genetic events within this process are well understood, no therapeutics targeting the loss of APC or the hyperactivation of KRAS via the frequent G12D activation have yet been developed.

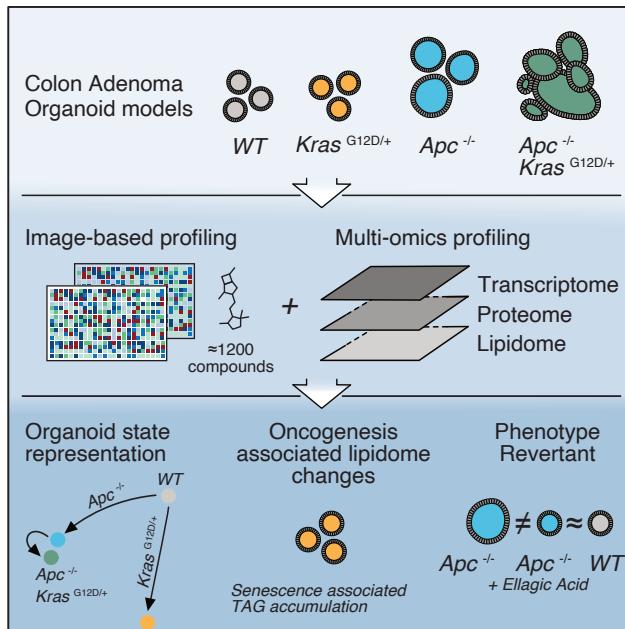


Figure 4.1: Visual abstract of adenoma model profiling project.

By using image-based profiling of genetically engineered mouse colon organoid models carrying Apc truncating mutations and/or a Kras G12D allele, I modelled the first set of genetic events within the Adenoma-Carcinoma sequence. Using a similar strategy than in the first chapter of this thesis, organoid models were characterized in-depth and subjected to a high-throughput small molecule screen of ca. 1700 FDA-approved substances and experimental compounds. The goal of this project was to (1) understand the biological state changes caused by individual transforming genetic changes in Apc and Kras within colon epithelial cells and (2) identify putative candidates that shift organoid states within the learned representation away from a pre-malignant state.

4.2 GENERATION OF ORGANOID COLON ADENOMA MODELS

APC and KRAS mutant lesions are considered intermediate colon adenomas (Fearon and Vogelstein, 1989). To model the formation of colon adenomas in vitro, I used a transgenic mouse to derive organoid cultures. The transgenic animal carried a conditional tamoxifen inducible KrasG12D/+ allele (Jackson et al., 2001) (Figure 4.2a). After isolation, I confirmed that extracted colon organoids did not express an activated form of KrasG12D (Figure 4.3a) and defined these organoids as wildtype (WT). To model loss-of-function mutations of the tumor suppressor Apc, the frequently mutated mutation-cluster-region on the APC gene was targeted by CRISPR (Figure 4.2a). Generated organoids harbored biallelic loss-of-function mutations in Apc (Figure 4.3a). Subsequent activation of oncogenic KrasG12D by treatment with 4-Hydroxytamoxifen led to four distinct organoid adenoma models (Figure 4.2a and Figure 4.3a-b); wildtype (WT), Apc-/- (A), KrasG12D/+ (K), and Apc-/- / KrasG12D/+ (AK).

Similar to genetically modified human colon organoids (Drost et al., 2015a; Matano et al., 2015), adenoma models showed characteristic niche requirements. Both Apc mutant organoid lines grew independent of the Wnt-signaling activating factors Wnt 3a and R-Spondin1 (Figure 4.2b). In fact, Apc mutant lines showed an increased growth in a Wnt3a and R-Spondin1 free environment when compared to the complete medium. Organoid models with an activated KrasG12D allele were less sensitive to removal of EGF from the media. However, as observed before (Drost et al., 2015a), the mutant KrasG12D allele was insufficient to compensate completely for the loss of EGF from the medium. Nevertheless, KrasG12D mutant organoid lines were more

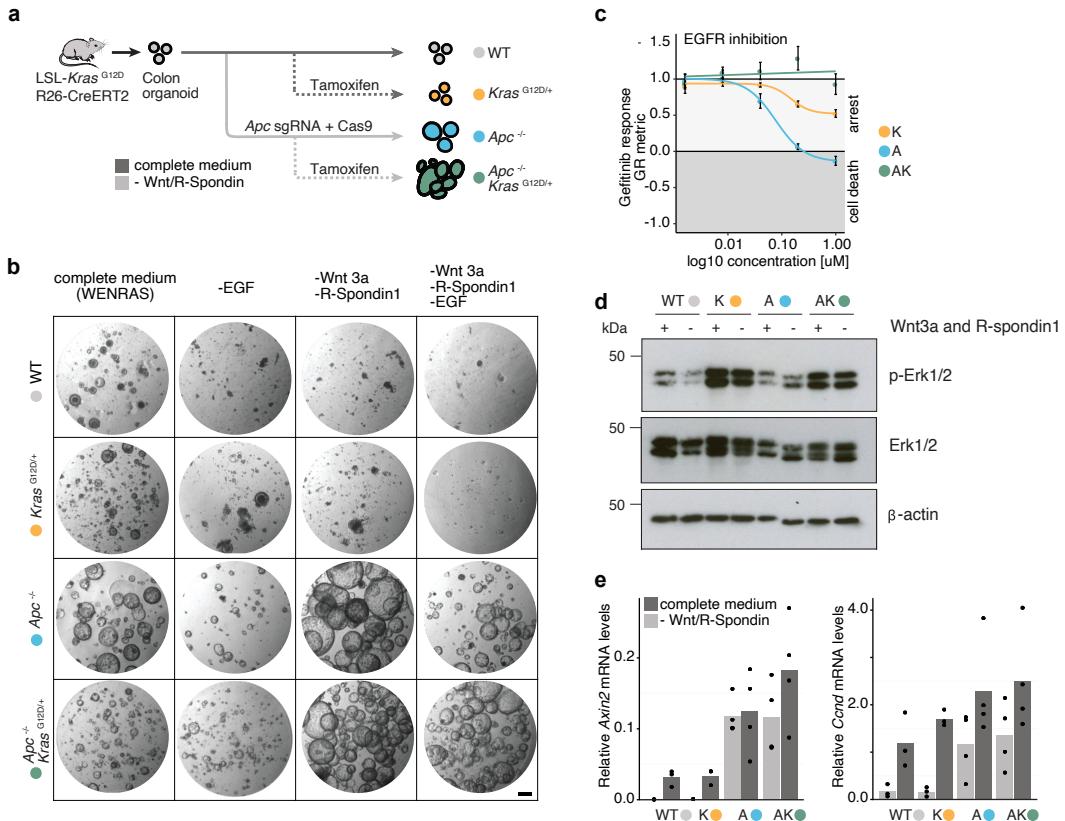


Figure 4.2: Establishing organoid models of colon adenoma. **a** Overview of organoid model establishment. Mouse colon organoids were isolated from a transgenic donor animal carrying an inactive conditional oncogenic KrasG12D allele. Homozygous truncation of Apc via CRISPR and activation of the heterozygous KrasG12D allele lead to four different genetically defined organoid models. **b** In vitro growth factor dependency of adenoma models. Organoids were cultured in complete or modified medium containing combinations of Wnt3A, R-Spondin1-Fc and EGF for 120h and subsequently imaged. Scalebar = 200um. **c** Oncogenic KrasG12D increases resistance to EGFR inhibition. Organoid ATP levels were measured 4 days after Gefitinib treatment and adjusted for organoid growth rate. Points represent mean of n=2 independent experiments. Error bars represent standard error of mean. **d** ERK phosphorylation is increased by oncogenic KrasG12D. Organoid models were cultured with or without Wnt3A and R-Spondin1-Fc for 72h and analyzed for protein levels. p, phospho. **e** Loss of Apc induces transcription of canonical Wnt-signaling target genes. qRT-PCR for Axin2 and Ccnd in the presence or absence of Wnt 3a and R-spondin1-Fc after 120h of culture. Expression levels are normalized to Sdha and Hprt transcript abundance. Bar graphs represent the mean of n=4 independent experiments.

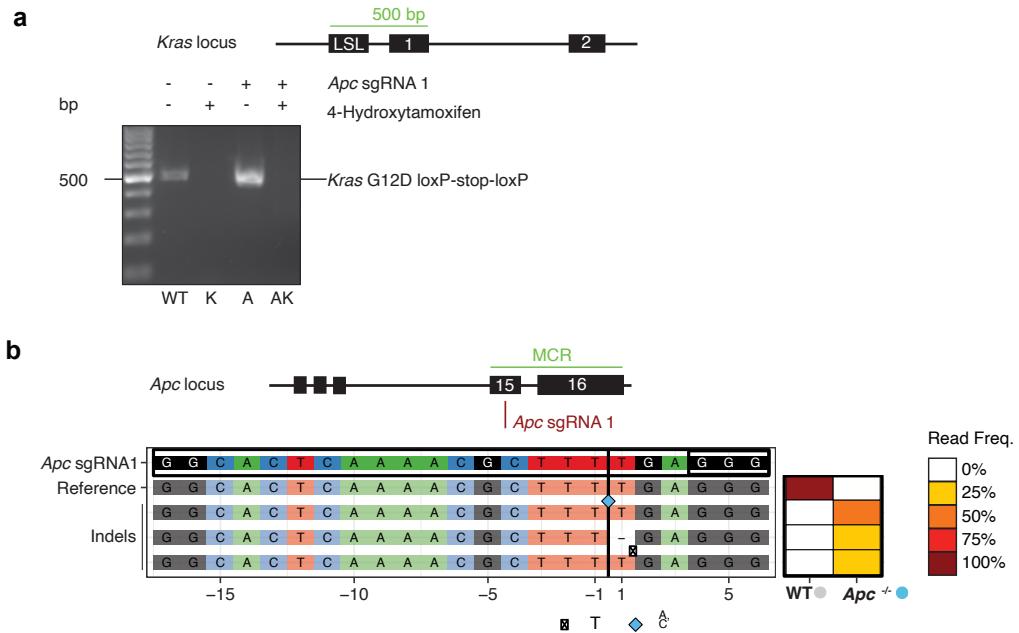


Figure 4.3: Structural validation of organoid colon adenoma models. **a** Allele-specific PCR products of colon organoid models isolated from a transgenic mouse with a conditional tamoxifen inducible KrasG12D/+ allele. **b** Amplicon sequencing result of the murine mutation cluster region ortholog for organoids transfected with an Apc targeting sgRNA and Cas9 carrying plasmid. The sequencing results show the presence of 3 different insertion/deletions within the pool of sgRNA treated organoid models. Wildtype sequences are absent within the CRISPR targeted pool, while mutant sequences are absent in the untreated organoid pool.

resistant to pharmacological inhibition of Egfr signaling (Figure 4.2c). In conclusion, organoid model genotypes were reflected in characteristic growth factor dependencies.

Next, I investigated the effects of mutations in Apc and Kras on both canonical Wnt- and Erk dependent signaling. While the presence of the KrasG12D/+ allele led to an increase in Erk-phosphorylation across models, Apc-/- / KrasG12D/+ organoids showed no marked additional increase in Erk-phosphorylation when compared to KrasG12D/+ organoids (Figure 4.2d). Moreover, Apc-/- / KrasG12D/+ adenoma models showed no significant differences in expression of the Wnt target genes Axin2 and Ccnd when compared to Apc-/- single-mutant models (A) ($p > 0.34$ for all conditions, Wilcoxon rank sum test) (Figure 4.2e). These results indicate that organoid adenoma models show genotype-dependent activity of characteristic signaling pathways, while there is no extensive crosstalk between the Apc-/- and KrasG12D/+ allele in mouse colon

organoids that is directly reflected in canonical Wnt- and Erk dependent signaling.

4.3 BIOCHEMICAL PROFILING OF ORGANOID MODELS

To explore comprehensive molecular differences between organoid models, I next performed transcriptome, proteome and lipidome profiling of all four organoid models (Figure 3.9). Transcriptome profiling of organoid models showed an increased expression of the stem-cell marker Lgr5 and Wnt-signaling regulators such as Nkdr1, Notum, Wif1 and Znrf3 in Apc mutant organoid lines (Figure 3.9b). To the contrary, Apc wildtype organoid lines showed an increased expression of epithelial differentiation markers, such as Krt20, Alpp and Abcb1 (P-glycoprotein). Overall, the number of genes with significant expression changes after Apc loss was 2.5 times greater compared to isolated KrasG12D activation (FDR = 0.1, Apc-/-: 44.5%, KrasG12D/+: 18.3% of assessed genes).

A related observation was made during the analysis of protein abundance. Proteome profiling and imputation identified 3906 gene products in all measured organoid lines. Again, Wnt signaling regulators (Axin2, Notum) were enriched in Apc mutant organoid lines and the number of significantly regulated proteins after Apc loss was 2.5 times greater compared to an isolated KrasG12D activation (FDR = 0.1, Apc-/-: 260, KrasG12D/+: 105 assessed proteins).

Principal component analysis of both transcriptome, proteome and lipidome data showed a related organization of variance across measurements. Across modalities, the first principal component captured differences between Apc wildtype and Apc mutant organoid models, while the second (in case of proteomics measurements the third) principal component captured differences between wildtype and KrasG12D/+ single-mutant models (Figure 4.4b, 4.4c and 4.4d). In every modality, a high degree of similarity was observed among Apc-/- and Apc-/- / KrasG12D/+ organoid lines. While activation of oncogenic KrasG12D in wildtype organoids led to global changes in transcript, protein and lipid expression, these changes were not as pronounced in organoids without functional Apc. In fact, only the mRNA expression of 91 genes was significantly altered between Apc-/- and Apc-/- / KrasG12D/+ organoids (FDR = 0.1).

To explore activated biological processes, gene set enrichment analysis on organoid mRNA expression data was performed. The strongest changes in gene expression after loss of Apc were linked to an increased

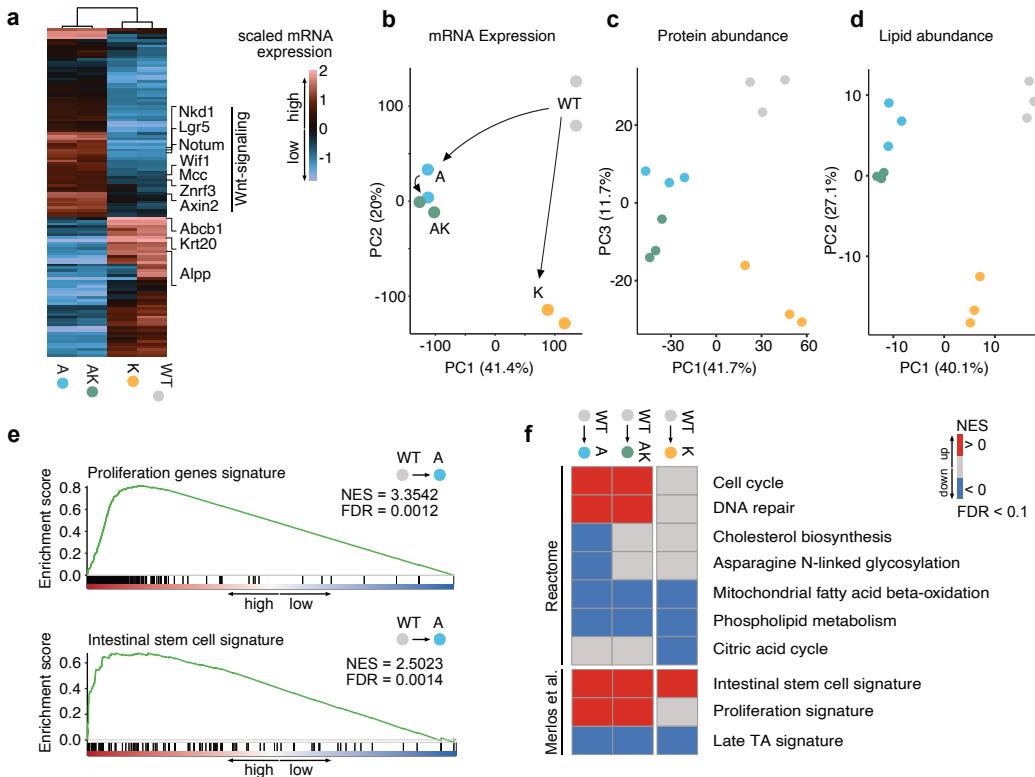


Figure 4.4: Molecular profiling of organoid adenoma models. **a** Differential gene expression of adenoma models. Shown are scaled expression values for the top 125 differentially expressed genes for every organoid line. Selected genes are highlighted. All organoids were cultured for 3 days in WENRAS before exposure to ENR for 4 days. Cell number was controlled between experiments. Whole organoid lysates were analyzed. **b** Transcript abundance data. Shown are the first two principal components of scaled gene expression data. The proportion of variance of each principal component is listed in parenthesis. **c** Protein abundance data. Shown are the first and third principal component of scaled protein expression data. The proportion of variance of each principal component is listed in parenthesis. **d** Lipid species abundance data. Shown are the first two principal components of scaled lipid abundance data. The proportion of variance of each principal component is listed in parenthesis. **e** Loss of Apc leads to increased expression of proliferation and intestinal stem cell associated genes. Shown is a gene set enrichment analysis of differentially expressed genes between Apc mutant and WT organoids. Intestinal gene expression signatures were used according to Merlos et al. NES, normalized enrichment score. **f** Overview of cellular processes in organoid adenoma models. Shown are selected enriched differential gene expression signatures from reactome and Merlos et al. NES, normalized enrichment score. NES > 0 suggests an enriched/ activated biological process. FDR < 0.1.

proliferative activity (Figure 3.9e). Gene set enrichment analysis of published intestinal cell-proliferation and stem cell signatures showed an enrichment of both signatures in Apc^{-/-} organoids (Figure 3.9e) (Merlos-Suárez et al., 2011). In contrast, a signature for differentiating transit-amplifying cells was depleted. Gene set enrichment analysis of Apc^{-/-} / KrasG12D/+ double-mutant organoids showed the same results.

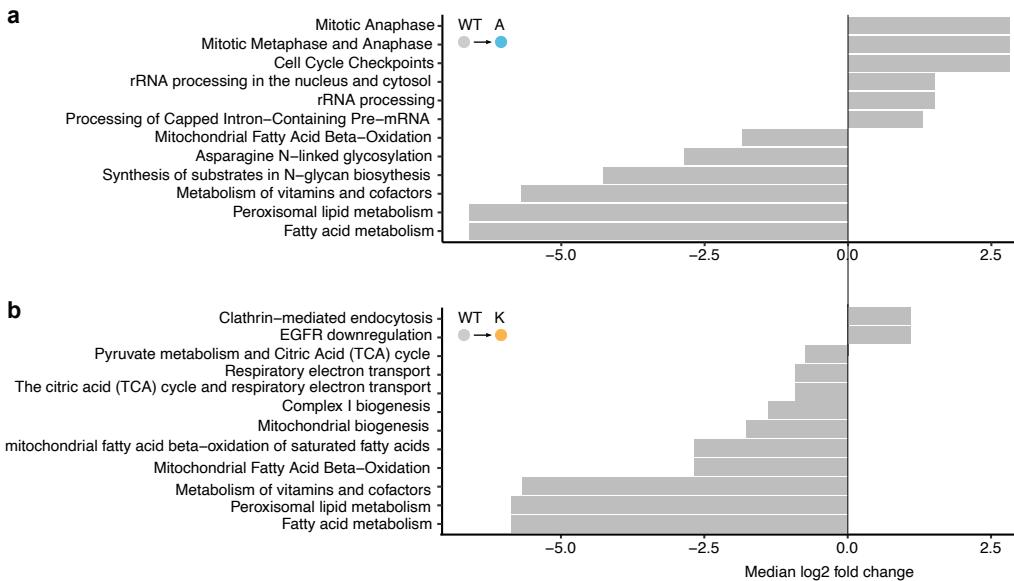


Figure 4.5: **a** Representative up and down-regulated transcriptional processes after loss of Apc. Expression signatures were sourced from reactome and average log2 fold changes for included transcripts are illustrated. FDR < 0.1. **b** Representative up and down-regulated transcriptional processes after activation of oncogenic Kras G12D. Expression signatures were sourced from reactome and average log2 fold changes for included transcripts are shown. FDR < 0.1.

Next to these published signatures, I explored the enrichment of curated gene sets from the Reactome database (Fabregat et al., 2018). Here, both Apc^{-/-} and Apc^{-/-} / KrasG12D/+ double-mutant lines showed a positive enrichment of cell cycle and DNA repair related genes when compared to wildtype organoids (Figure 4.5a). Unique to the KrasG12D/+ organoid line was a decreased expression of citric acid cycle and respiratory chain related genes (Figure 4.5b). This effect, was not observed in Apc^{-/-} / KrasG12D/+ double mutant organoids (Figure 2.4f). In addition, organoid models with an KrasG12D/+ genotype showed a downregulation of the EGFR receptor, in line with a potential negative feedback response to hyperactivated

RAS-MAPK signaling (Figure 4.5b). Both Apc-/- and KrasG12D/+ organoid models showed a strong reduction of lipid metabolism and beta-oxidation (Figure 4.5a,b).

In summary, loss of Apc leads to a global shift in transcript, protein, and lipid abundance in colon organoids, including a strong increase in cell proliferation associated genes. Activation of isolated oncogenic KrasG12D leads to pronounced reduction in citric acid cycle related gene expression while this phenotype was not seen in organoid models with a loss of Apc.

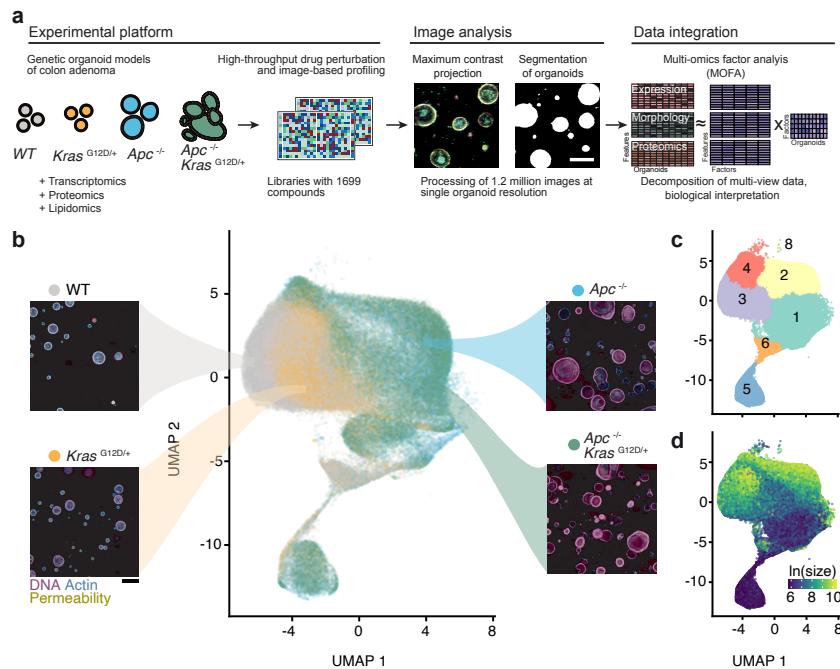


Figure 4.6: Image-based profiling of organoid adenoma models. **a** Overview of experiments. Organoids were isolated from a transgenic mouse model and genetically edited. Organoids were dissociated and evenly seeded in 384-well plates before perturbation with an experimental small molecule library. After treatment, high-throughput fluorescence microscopy was used to capture the morphology of organoids in 16 selected z-layers and 3 channels. 3D imaging data were projected on a 2D plane using a maximum contrast projection. Here, only pixel areas with the largest contrast among the z-axis were retained. Morphological features were computed based on the projection. Untreated organoid morphology, organoid size and drug activity scores were integrated with transcript expression, protein abundance, lipid abundance and genotype data in a Multi-Omics Factor Analysis (MOFA) model. Figure created with support from Johannes Betge (graphical presentation). **b** Uniform Manifold Approximation and Projection (UMAP) of organoid-level features for a random 5% sample out of imaged organoids. The identical sample is used for visualizations throughout the figure. Organoid genotype is colorcoded and representative images are displayed (magenta = DNA, cyan = actin, cell permeability = yellow, scale-bar: 200µm). **c** Graph-based clustering of organoids by morphology with 8 resulting clusters. **d** Organoid size distribution. Color corresponds to the log-scaled organoid area (dark blue: minimum size, yellow: maximum size).

4.4 IMAGE-BASED PROFILING OF ORGANOID MODELS

To measure how organoids change their biological state as a response to small molecule perturbation, I used the previously developed image-based profiling method to observe organoid morphology. Organoid models of four different genotypes were perturbed with a library of ca. 1700 compounds and morphological profiles were computed (Figure 4.6a). A UMAP projection of 25 principal components representing single-organoid morphology showed distinct genotype-dependent morphological states for viable organoids (Figure 4.6b). Graph based clustering of organoid morphology profiles resulted in 8 different clusters (Figure 4.6c). While developed organoids within cluster 4 and 3 were enriched for *Apc*^{+/+} organoid models, cluster 2 and 1 were populated by *Apc*^{-/-} models. Analogous to gene expression, lipidomics and proteomics representation space, *Apc* mutant organoid models were less distinct from each other than organoids with a WT and isolated *KrasG12D/+* genotype (Figure 4.7b). While developed organoids that present with a larger organoid area showed distinct genotype-specific morphologies, small and dead organoids clustered together across genotypes within cluster 5 (Figure 4.6c and d).

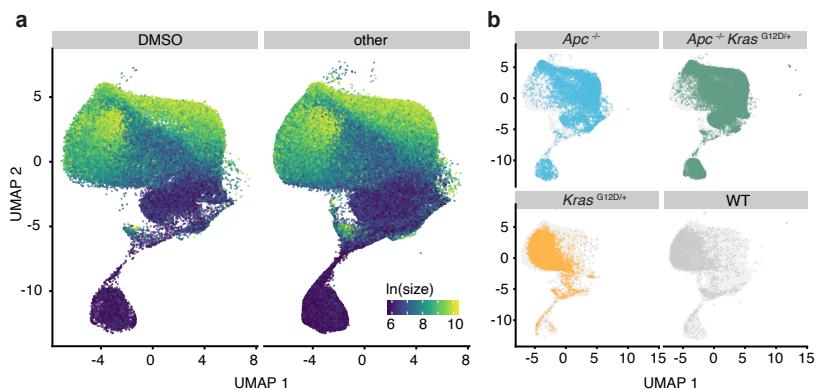


Figure 4.7: Treatment and genotype dependent effects on organoid morphology distribution. a UMAP representation of DMSO treated (vehicle) and small molecule treated organoids. b, UMAP embeddings of four organoid genotypes (baseline state = 0.1% DMSO control-treated organoids), grey background consists of randomly sampled organoids.

The distribution of DMSO-treated organoids and small molecule perturbed organoids in morphological space overlapped strongly (Figure 4.7a), possibly because of large amount of treatments that did not alter organoid morphology.

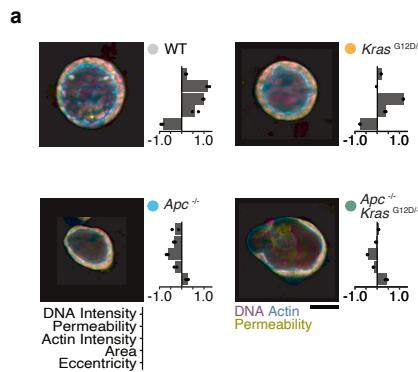


Figure 4.8: Genotype dependent effects on organoid morphology. a Unperturbed organoid profiles from adenoma models were aggregated. Shown are representative individual organoids with selected features. Points show the mean phenotype for each independent biological replicate. Selected features and their z-scores relative to all single organoid profiles are shown (magenta = DNA, cyan = actin, cell permeability = yellow, scale-bar: 25μm)

When comparing the morphologies of different organoid models in detail, characteristic differences were identifiable (Figure 4.8a). DMSO-treated *Apc*^{+/+} organoids showed a strong, regular apical actin cytoskeleton (high average actin intensity) that organized the multicellular formation into a regular-patterned spherical morphology (low average eccentricity). In contrast, *Apc*^{-/-} organoids showed a relative lack of a regular actin cytoskeleton (low average actin intensity) and a irregular, non-spherical morphology (high average eccentricity).

In summary, developed organoids showed genotype-dependent differences in morphology. Analogous to differences in biochemical state, a primary source of variation was the loss of the tumor suppressor gene *Apc*. Organoids with truncated *Apc* presented with a loss of the spherical, potentially structure-conferring, and cell-spanning apical actin cytoskeleton that was observed in *Apc*^{+/+} organoid models.

4.5 QUANTIFYING SMALL MOLECULE INDUCED PHENOTYPES ACROSS ORGANOID MODELS

To study the effect that small molecule perturbations had on organoid models, the classification based approach developed during the study of human cancer organoid phenotypes was used. Briefly, for every treatment and genotype, a linear classifier was trained to distinguish DMSO-treated organoids from treated organoids. The classification performance, expressed as the AUROC obtained on a hold-out dataset, was used to determine the activity of a compound. A high AUROC (approaching 1) is observed for compounds that lead to a treatment-induced organoid morphology that is very distinct from DMSO treated organoids. In contrast a low AUROC (minimum of 0.5) is observed for compounds where the classification performance approaches random guessing (Figure 4.9a).

Related to the approach chosen in the previous chapter, active treatments were identified based on the AUROC score that a classifier reached. Given differences in the distribution of AUROC scores between lines, with KRAS G12D/+ organoid lines being shifted towards higher AUROC values, I centered the distribution of AUROC scores around 0.5 and defined an arbitrary activity threshold at 3 times the Median Adjusted Deviation for all tested models (Figure 4.9b). A primary source of variation in the genotype-dependent identity of active compounds (16% of all tested small molecules) was the functional state of Apc, as most active compounds were shared among Apc+/+ and Apc/- models, while little relative overlap existed between these two alleles (Figure 4.9c). As described above, the higher classification performance seen in organoids with the Kras G12D/+ allele, led to a larger number of active drugs identified for these models. A possible reason for this systematic difference in the number of active treatments might be a larger number of developed organoids seen in the images of these models. This difference might be linked to the previously described increased colony-forming capacity seen in Kras G12D/+ colon cancer models.

Given the observation that the primary source of variation for drug activity, similar to other biochemical assays, was the state of the Apc allele, I focused the analysis of treatment-induced organoid phenotypes along this axes. Similar to the approach taken in the previous chapter, normal vectors of the logistic regression classifiers were compared by estimating the enclosed angle (Figure 4.9d). Normal vectors for organoids with the same Apc allele were concatenated. The resulting clustering of normal vectors by their similarity showed an enrichment for small molecules with related mechanism of action (Figure 4.9e and f). For example, EGFR

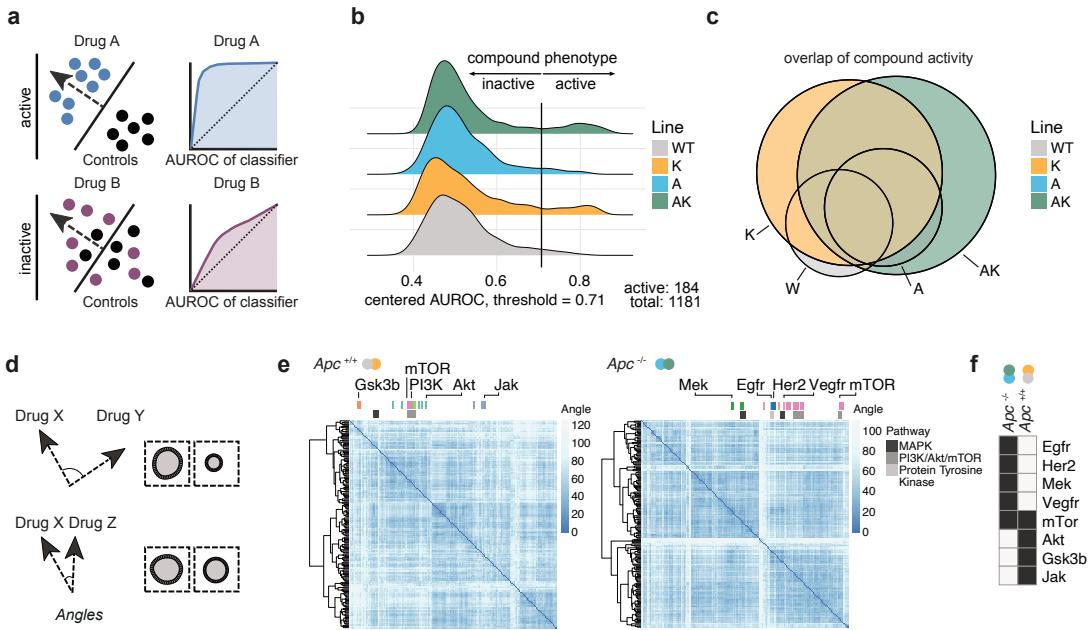


Figure 4.9: Small molecule activity scoring. **a** A logistic regression classifier is trained to distinguish morphology profiles of individual treated and untreated organoids across all available replicates. Afterwards, the classifier is applied to a validation set of organoids and the classification performance is estimated using the area under the receiver operating characteristic curve (AUROC) metric. Method implemented by Jan Sauer. **b** Distribution of AUROC compound activity scores for all organoid lines, replicates and perturbations. AUROC scores were centered around 0.5 and treatments for this particular analysis were termed active when classification accuracy exceeded three times the Median Absolute Deviation (MAD) of the AUROC score distribution, an arbitrary threshold. A set of 184 compounds (16% of all screened small molecules) met the activity criteria. **c** Overlap of active compound treatments across organoid models. Shown is a Euler diagram of active compounds for each line. K (Kras G12D), AK (Apc loss and Kras G12D) and WT are color coded. Plot diagnostics: diagError: 0.011, stress: 0.002. **d** Identifying related treatment induced phenotypes. Normal vectors of treatment specific classifiers were compared by calculating the angular distance (related to cosine similarity, ranging from 0-180 degrees). Small angular distance between vectors correspond to a high similarity between the treatment-induced organoid phenotypes. Method implemented by Jan Sauer. **e** A map of compound induced phenotypes for *Apc* mutant and *Apc* wildtype organoids. High-lighted are clusters of compound induced phenotypes with related targets. Normal vectors for *Apc* mutant and *Apc* wildtype organoids were concatenated before angular distance calculation. Method implemented by Jan Sauer. **f** Treatment induced phenotypes by organoid genotype. Shown are significantly enriched treatment induced phenotypes for *Apc* mutant and wildtype organoid models. Clusters of similar phenotypes were tested for overrepresentation of known molecular targets using Fisher's exact test. Significantly enriched targets are shown. Method implemented by Jan Sauer.

inhibitors were significantly enriched in Apc-mutant organoid lines, while GSK₃B-inhibitors, which lead to a stimulation of canonical Wnt signaling, were enriched in Apc-wildtype organoid models, only.

4.6 MULTI-OMICS FACTOR ANALYSIS IDENTIFIES SHARED FACTORS LINKING FUNCTIONAL AND STRUCTURAL BIOLOGICAL VIEWS

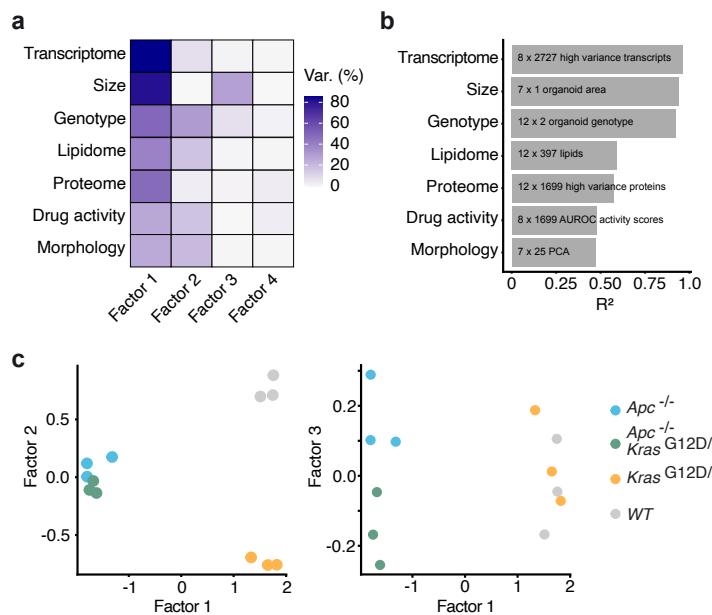


Figure 4.10: Multi-omics factor analysis to identify shared factors linking morphology, size, gene expression, lipidomics, proteomics, genotype and drug activity. **a** Percent variance explained by the MOFA model for each factor. Untreated organoid morphology, organoid size and drug activity scores were integrated with genotype, proteomics, lipidomics and mRNA expression data. **b** Cumulative proportion of total variance explained by each experimental data modality within the MOFA model. **c**, Visualization of samples in factor space showing factors 1 and 2 as well as factor 1 and 3. Shown are independent replicates for each organoid line.

To comprehensively model the biological state of organoid models and explore potential interventions that

move organoids in state-space, I performed multi-omics factor analysis (MOFA). Analogous to the process described in chapter 1, feature matrices from different sources were processed and factorized using $k=4$ factors (Figure 4.10a and 4.18a). The learned model was based on both functional (e.g. drug activity) and structural (e.g. genotype, proteomics, lipidomics and mRNA expression) information. To reduce the dimensionality of input data, only high variance features from gene expression and proteomics analysis were used. The resulting factorization explained an overall XX percent of variance across the analyzed views, of which the first three factors captured the majority (XX percent of variance, Figure 4.10a). The learned model explained most variance within the mRNA expression and genotype data, while measurements within the organoid morphology data had the lowest explained variance (Figure 4.10b). Visual inspection of factors as well as exploration of factor loadings within the genotype view showed that factor 1 explained state differences caused by Apc loss of function, while factor 2 explained state differences caused by the activation of KrasG12D in an Apc^{+/+} genotype (Figure 4.10c and 4.18b). In contrast to factor 2, factor 3 captured differences between Kras^{+/+} and KrasG12D^{+/+} organoids with Apc loss of function, albeit with low overall explained variance. While the initial number of factors is a user-defined feature within MOFA, the method automatically drops excess factors if they are not considered effective based on an applied automatic relevance determination (ARD) prior. Increasing the number of factors above $k=4$ in this analysis, did not lead to an increased number of interpretable factors. In fact, factor 4 already did not capture differences between organoid genotypes and was not interpretable from a biological point of view (Figure 4.18b). This observation and the fact that this study explored the biological effects of two oncogenic genetic events both in isolation and in concert reveals, I speculate, a potential conceptual intersection of representation learning methods, such as MOFA, and the theory of genetic interactions which will be explored in the last chapter of this thesis.

4.7 A CANONICAL WNT SIGNALING ASSOCIATED PROGRAM CAUSED BY APC LOSS

To understand the molecular changes associated with factor 1, factor loadings for mRNA expression data were analyzed using Reactome gene-set enrichment analysis (Figure 4.12a). Three clusters of biological processes were significantly associated with a negative factor loading, caused by Apc loss-of-function: 1) Mitotic Anaphase related processes, including spindle checkpoints; 2) Mitotic S-phase, including DNA replication

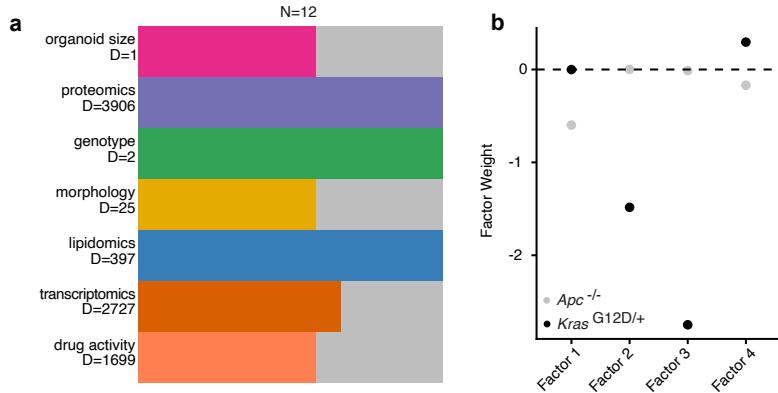


Figure 4.11: Multi-omics factor analysis input data and loadings. **a** measurement modalities, dimensionality and number of measurements. A third replicate of measurements were available for proteomics and lipidomics only. **b** Factor loadings for genotype information.

and 3) DNA repair mechanisms, including homology directed repair.

In line with the enrichment of processes seen in cell proliferation, factor 1 loadings were associated with an enrichment of a previously described intestinal proliferation signature (Figure 4.12c) and an LGR5+ intestinal stem cell identity signature (Figure 4.12b). These findings are in line with the long-standing evidence that loss of *Apc* leads to a hyperactivation of canonical Wnt signaling, which in turn leads to increased intestinal cell proliferation and Myc-dependent changes towards a stem-like cell state.

When focusing on factor loadings of compound activity measurements, a low factor 1 score was significantly linked to increased sensitivity towards small molecules targeting microtubuli and focal adhesion kinase (FAK, Figure 4.12d). This morphological sensitivity was presented itself primarily as reduced organoid size and number relative to the DMSO vehicle control (Figure 4.12e).

In contrast to microtubuli and FAK inhibitors, the average activity scores of small molecules targeting Wnt signaling was associated with increased factor 1 scores (Figure 4.12d). Further exploration of the association between the AUROC activity score and *Apc* genotype showed that small molecule inhibitors of the canonical Wnt secretion pathway protein Porcupine (Porcn), IWP-L6 and LGK-974, were more active in *Apc* WT

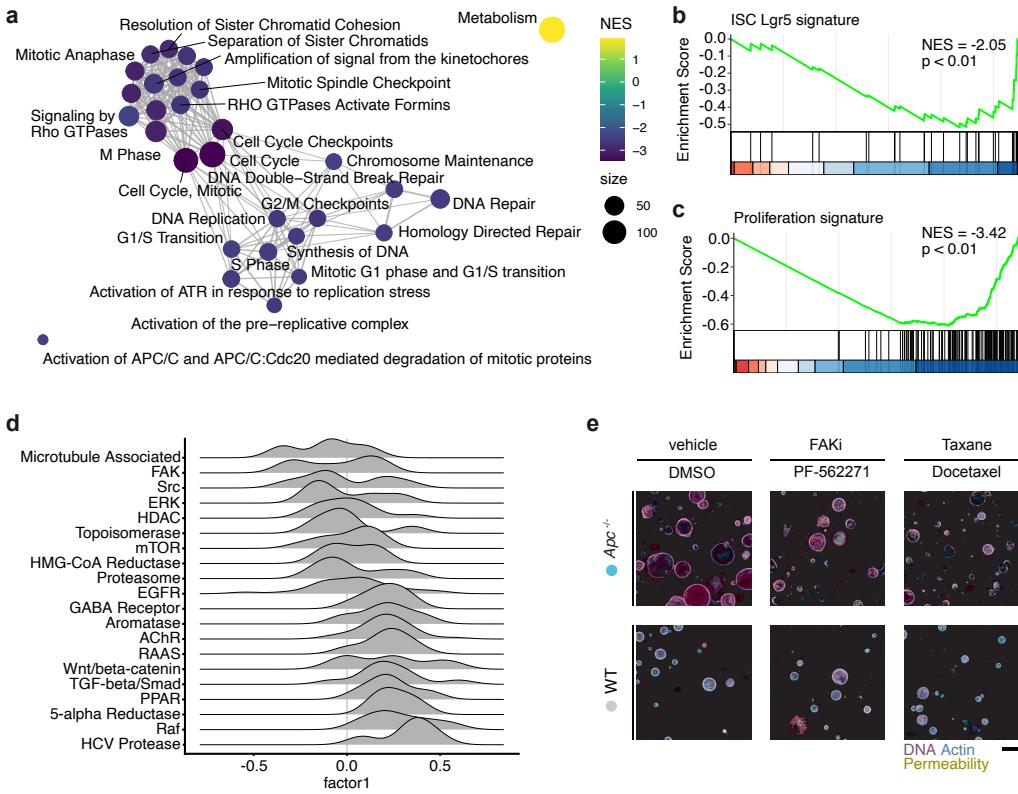


Figure 4.12: Factor 1, canonical Wnt signaling. a Gene-set enrichment network of factor 1 gene expression loadings. An edge connects Reactome pathways with more than 20% overlap. Central enriched processes include mitosis, DNA replication and DNA damage repair. b and c Gene set enrichment results of the "Lgr5 intestinal stem cell" and "proliferation" signature by Merlos-Suarez et al. over ranked factor 1 gene expression loadings (ranking from high factor 1 loading to low factor 1 loading, NES = normalized enrichment score). d Distributions of drug activity loadings grouped by drug target for factor 1. e Example images of compound treated organoids with WT or Apc-/- genotype. Representative images are displayed (magenta = DNA, cyan = actin, yellow = cell permeability, scale-bar: 200 μ m).

organoids relative to their Apc-/- counterparts (Figure 4.13a). In contrast, this effect was not observable for PRI-724, a small molecule inhibitor targeting the interaction of beta-catenin and CREB-binding-protein (4.13a). These differences in drug activity scores among these small molecule inhibitors are most likely related to their different targets' relative location to Apc in the canonical Wnt signaling cascade. While Porcn-dependent Wnt secretion is generally upstream of the destruction complex, the interaction of beta-catenin and CREB-binding-protein is located downstream of it. As a consequence, inhibition of destruction complex function by loss of Apc is expected to render cells less sensitive to perturbations of the Wnt secretion cascade than perturbations of transcription factor binding properties (Figure 4.13b).

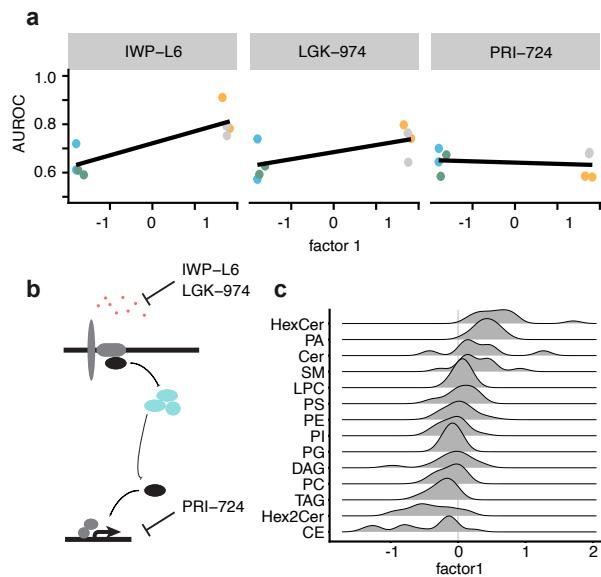


Figure 4.13: Small molecule Wnt signaling inhibitors. **a** AUROC activity score for three small molecule inhibitors of canonical Wnt signaling. **b** Target proteins for small molecules within the canonical Wnt signaling cascade with their relative position to the destruction complex (highlighted in blue). **c** Distributions of lipid abundance loadings grouped by lipid species for factor 1 (Abbreviations: HexCer - Hexosylceramide; PA - Phosphatidate; Cer - Ceramide; SM - Sphingomyelin; LPC - Lysophosphatidylcholine; PS - Phosphatidylserine; PE - Phosphatidylethanolamine; PI - Phosphatidylinositol; PG - Phosphatidylglycerol; DAG - Diacylglycerol; PC - Phosphatidylcholine; TAG - Triacylglycerol; Hex2Cer - Dihexosylceramide; CE - Cholesterol)

Next to factor related differences in morphological compound sensitivity, I analyzed the association of factor 1 weights with lipid species. High concentrations of cholesterol esters were associated with a low factor 1 score (seen in Apc^{-/-} organoids), while elevated concentrations of phosphatidic acid species were linked to a high factor 1 score (figure 4.13c).

After linking factor 1 to Apc loss and identifying features caused by this molecular change, I was interested in identifying small molecule treatments that -based on the morphology they induced- shifted organoid state along the factor 1 axis. As described in the previous chapter, small molecules that led to a predicted factor 1 change were identified using ANOVA. To identify treatments that led to a drop in organoid viability, a classifier (LDC) trained on ground-truth lethal treatments within the stem-cell library was applied. Two groups of compounds were identified that induced a shift towards lower factor 1 scores (group A, figure 4.18) and higher factor 1 scores (group B, 4.18).

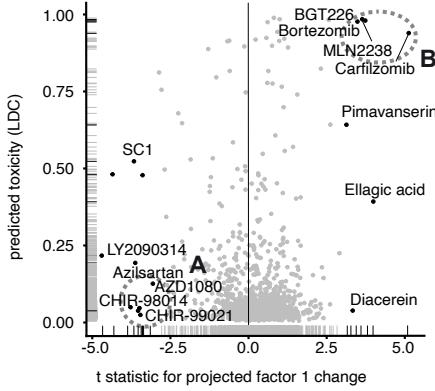


Figure 4.14: Projection of factor 1 scores for treatment-induced phenotypes and viability changes. Highlighted are compounds leading to a significant change in projected factor scores across all organoid lines (ANOVA). Organoid viability is predicted using a random-forest based classifier (LDC) with scores from 0 (no toxicity) to 1 (complete toxicity)

Small molecules within group A induced a morphology associated with Apc loss while maintaining organoid viability. In contrast, members of group B primarily led to a loss of viability and a shift towards a morphological state associated with Apc wildtype organoids. Of note, small molecules within these groups had related target proteins! Compounds within group A (incl. CHIR-98014, CHIR-99021, LY2090314) targeted GSK₃ beta - a kinase with central function within the canonical Wnt signaling destruction complex. Inhibition of GSK₃ beta leads to hyperactivation of canonical Wnt signaling. Members of group B primarily targeted the Proteasome (incl. Bortezomib, Carfilzomib, MLN2238).

Further validation of group A showed how treatment with the GSK₃ beta inhibitor CHIR-98014 led to treatment-induced phenotypes in WT and KrasG12D^{+/−} organoids that phenocopied the unperturbed morphology of Apc^{−/−} and double mutant organoid models (Figure 4.15a). On the feature level, treatment led to an increase in organoid size and DNA intensity (Figure 4.15b) in Apc wildtype models. This change in morphology was likely due to an increased proliferation rate of mutant cells, leading to rising organoid size and a higher density of nuclei per analyzed object. Guided by the identification of a strong GSK₃ beta

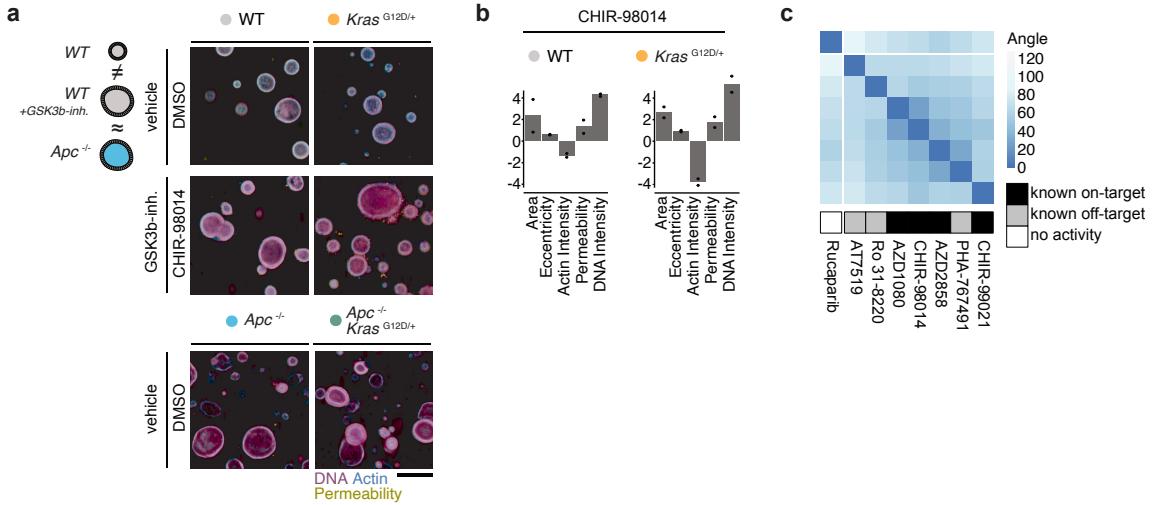


Figure 4.15: GSK3 beta inhibition dependent morphology in colon organoid models. a Small molecule inhibition of GSK3 beta (CHIR98014) leads to phenocopying of Apc^{-/-} genotype organoid models. b Shift of morphological features relative to scaled population mean for Apc wildtype organoid models treated with CHIR98014. c Excerpt of clustering from figure 4.9d, labeled with known binding activity of listed small molecules. Rucaparib is not member of the cluster and shown for comparison.

inhibition phenotype in Apc wildtype organoids, I analyzed small molecules that clustered with known inhibitors of this kinase based on the angular distance of their drug effect vectors within these models (Figure 4.15c) and 4.9d). All small molecules clustering with well-described inhibitors of GSK3 beta had previously described off-target binding activity against this kinase.

In conclusion of this section, loss of Apc function is the primary source of variation across the four analyzed organoid genotypes. Increased cell proliferation rate, canonical Wnt signaling and genomic stress caused by a loss of Apc is observable across profiling modalities and leads to characteristic molecular dependencies (microtubuli, FAK signaling). Small molecules inhibiting the function of the destruction complex member GSK3 beta can phenocopy loss of Apc among colon organoid models.

4.8 AN ONCOGENE-INDUCED SENESCENCE PROGRAM CAUSED BY ISOLATED KRAS_{G12D} ACTIVATION

While the Apc^{-/-} genotype contributed primarily to factor 1 (Figure 4.18b), the Kras_{G12D} allele showed a strong loading for both factor 2 and factor 3. This observation paired with the fact that only organoid models without Apc loss of function were separated by factor 2 (Figure 4.10c), led me to conclude that factor 2 described a Kras_{G12D} dependent change in cell state that was conditioned on intact Apc function.

To understand the molecular mechanisms represented by factor 2, I again analyzed features with large absolute loadings for this factor. Plotting of factor loadings for drug activity by small molecule target, showed that ERK and MEK inhibitors were more active in factor 2 low models (Kras_{G12D}^{+/+}) while EGFR/HER2 inhibitors were more active in factor 2 high organoids (WT, figure 4.16a and b). This juxtaposition in small molecule activity against RAS-MAPK pathway members was reminiscent of the previous observations made around canonical Wnt signaling inhibitors (Figure 4.13). With oncogenic Kras localized between the receptor-layer (including Egfr and Her2) and downstream kinases (for example Erk), hyperactive Kras signaling likely leads to a cell state with relative resistance to EGFR inhibitors and increased dependency on Erk signaling. The previously observed transcriptional process of Egfr-downregulation as a response to Kras_{G12D}^{+/+} is in line with these observations (Figure 4.5b).

In addition, HMG-CoA reductase inhibitors (incl. Rosuvastatin) showed an increased activity in Kras_{G12D}^{+/+} organoids (Figure 4.16a and b). This vulnerability towards cholesterol biosynthesis inhibitors in cells with oncogenic Kras signaling has previously been described.

Oncogene induced senescence is a cell state marked by an arrest of the cell cycle and expression of pro-inflammatory mediators as a response to an oncogenic perturbation. An activated Kras_{G12D}^{+/+} genotype leads to oncogene induced senescence of colon epithelial cells *in vivo*. Prompted by previous reports on the effect of oncogenic Kras, I identified an enrichment of a senescence related gene expression signature within the loadings of factor 2 (Figure 4.16c). Transcripts linked to cell senescence, including Igfbp3 (factor 2 loading ca. -0.999) and Hmgaz2 (factor 2 loading ca. -1.050) ranked among the strongest contributors to the factor. Further exploration of loadings for factor 2 revealed systematic shifts in the composition of the lipidome. Activation of oncogenic Kras led to an accumulation of sphingomyelin (SM) and triacylglycerol (TAG) species in colon cells. In contrast, I observed a relative depletion of cholesterol esters (CE) and

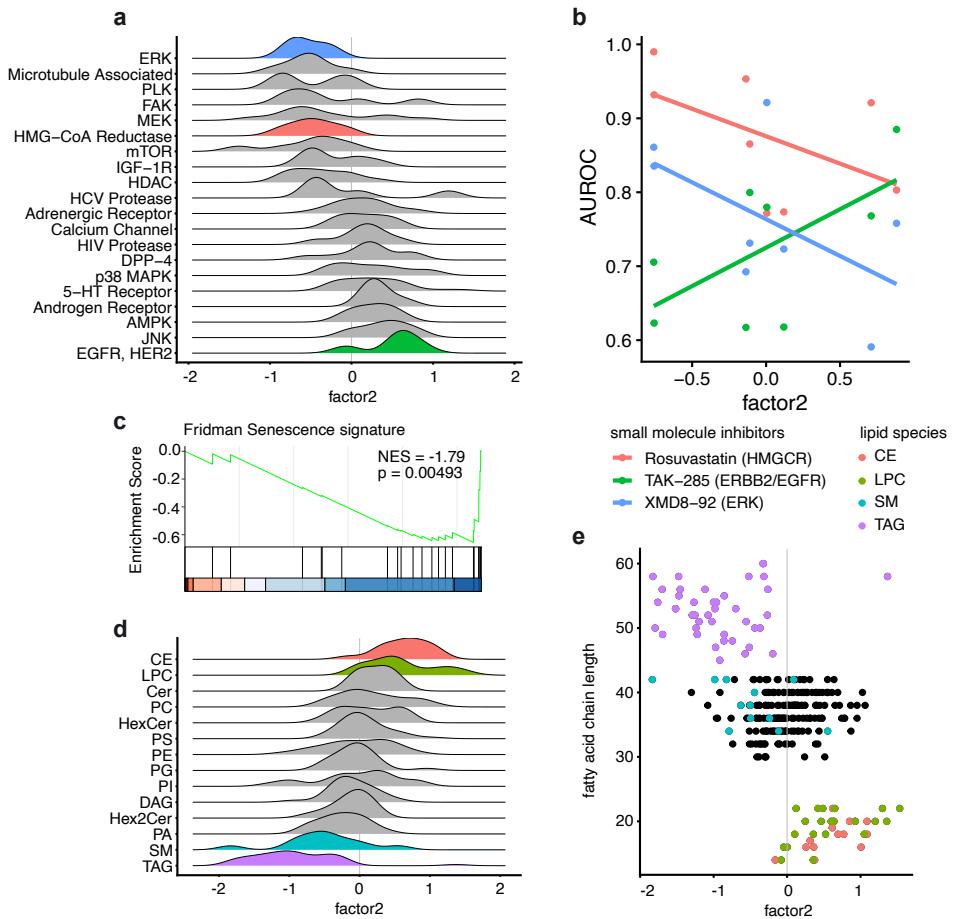


Figure 4.16: Factor 2, KrasG12D induced senescence. **a** Distributions of drug activity loadings grouped by drug target for factor 2. **b** Relationship of representative drugs' activity with factor 2 score. Shown are compounds from highlighted groups in panel (a). **c** Gene set enrichment results of a senescence signature by Fridman et al. over ranked factor 2 gene expression loadings (ranking from high factor 2 loading to low factor 2 loading, NES = normalized enrichment score). **d** Distributions of lipid abundance loadings grouped by lipid species for factor 2 (Abbreviations: HexCer - Hexosylceramide; PA - Phosphatidate; Cer - Ceramide; SM - Sphingomyelin; LPC - Lysophosphatidylcholine; PS - Phosphatidylserine; PE - Phosphatidylethanolamine; PI - Phosphatidylinositol; PG - Phosphatidylglycerol; DAG - Diacylglycerol; PC - Phosphatidylcholine; TAG - Triacylglycerol; Hex2Cer - Dihexosylceramide; CE - Cholesterolester). **e** Relationship of acyl chain length with factor 2 loading. Shown are lipids from highlighted species in panel (d)

lysophosphatidylcholine (LPC) (Figure 4.16d). Next to shifts in the lipid composition, I also observed a correlation of fatty acid residue chain length and factor 2 loadings, where a low factor 2 score was correlated with higher residue chain length (Figure 4.16e) and a higher degree of lipid desaturation (data not shown). Both of these association was primarily driven by differences in the abundance of lipid species, however.

To summarize, factor 2 represents an axis of variation within the observed organoid models that accounts for

x% of variance across data modalities. Factor 2 explains differences between wildtype and KrasG12D+/- organoid models and captures the activation of RAS-MAPK signaling and subsequent oncogene induced senescence. KrasG12D+/- transformed colon cells showed a hyperactivation of RAS-MAPK signaling with characteristic changes in small molecule dependencies (ERK and EGFR inhibitors, respectively). Of note, KrasG12D+/- induced senescence coincided with the accumulation of triacylglycerol (TAG) and sphingomyelin (SM) species within colon organoids. Lipid accumulation in senescent cells has been observed in multiple senescent cells, including human fibroblasts (TAG accumulation) and, more specifically, has recently been described in non-mammalian model systems as a consequence of Kras mediated oncogene induced senescence.

4.9 ONCOGENIC KRAS LEADS TO INCREASED mTOR SIGNALING CONDITIONED ON LOSS OF APC FUNCTION

While factor 2 captured the effect of oncogenic Kras in an Apc wildtype state, factor 3 scores separated organoid models by Kras genotype in the context of an Apc loss of function (Figure 4.10c). Overall, factor 3 only accounted for x% of variance within the MOFA model, supporting the overall similarity of Apc-/- and Apc-/- KrasG12D+/- organoids observed on the transcriptome, proteome, lipidome and phenotype level (Figure 4.4a-d and 4.7b).

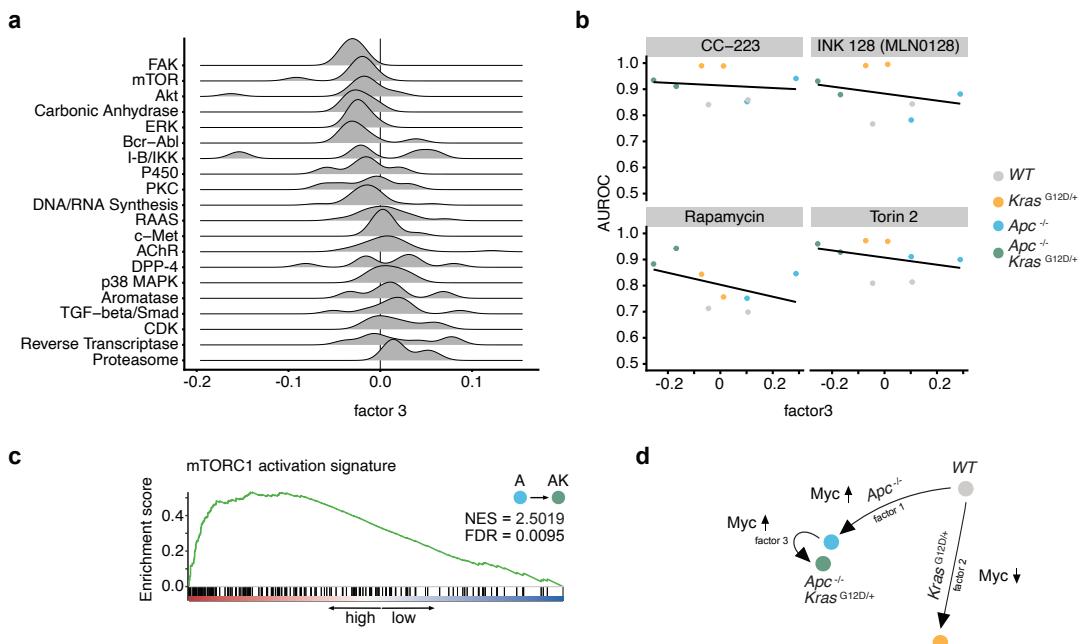


Figure 4.17: Factor 3, KrasG12D effects in the context of Apc loss of function. a Distributions of drug activity loadings grouped by drug target for factor 3. b Relationship of representative mTOR inhibitor activity with factor 3 score. c Gene set enrichment results of a Reactome mTORC1 activation signature over ranked factor 3 gene expression loadings (ranking from high factor 3 loading to low factor 3 loading, NES = normalized enrichment score). d Visual summary of Myc gene set enrichment results for organoid state transitions. Myc signatures were significantly enriched in Apc-/- models and depleted in models with isolated KrasG12D+/- mutation.

Exploration of factor 3 loadings identified mTOR and FAK inhibitors to systematically contribute to a negative factor score. Double-mutant organoid models (Apc-/- KrasG12D+/-) were more sensitive to these small molecule inhibitors than their Apc-/- counterparts (Figure 4.17a). This difference in drug activity was observable for both ATP-competitive (e.g. INK-128, Sapanisertib) and non-ATP-competitive (e.g.

Rapamycin) inhibitors (Figure 4.17b). In line with the increased sensitivity to mTOR inhibitors, differential gene expression analysis comparing single (*Apc* -/-, abbreviated A) and double mutant (*Apc* -/- *KrasG12D* +/+, abbreviated AK) organoids identified a significant increase in mTORC1 activation according to a Reactome signature (Figure 4.17c).

kras oncogene induced senescence next to mice *p16* or *TP53* deletion avoid cell state in zebrafish myc signaling avoids cell state Myc is a master regulator after loss of *Apc* Myc went up after loss of *Apc* even further up after combination of *Apc* *Kras* but down after *Kras*

gene expression only

fig,61cholesterolbiosynthesisgoesuprelativetoApclossmonorespiratoryelectrontransportgoeslowinKrasmono

In summary the effect of *Kras* conditioned on the loss of *Apc* function. This factor captures little variance relative to factor 1 and 2.

4.10 ELLAGIC ACID IS A CANDIDATE SMALL MOLECULE REVERTANT OF ORGANOID ADENOMA STATES

Related to factor 1 and previous chapter, I searched for treatments reverting phenotype for both factor 1 and factor 2 towards wildtype. The majority of compounds that led to a shift towards a wildtype organoid state along both axes were associated with a high predicted toxicity (color coding, predictions based on LDC viability classifier).

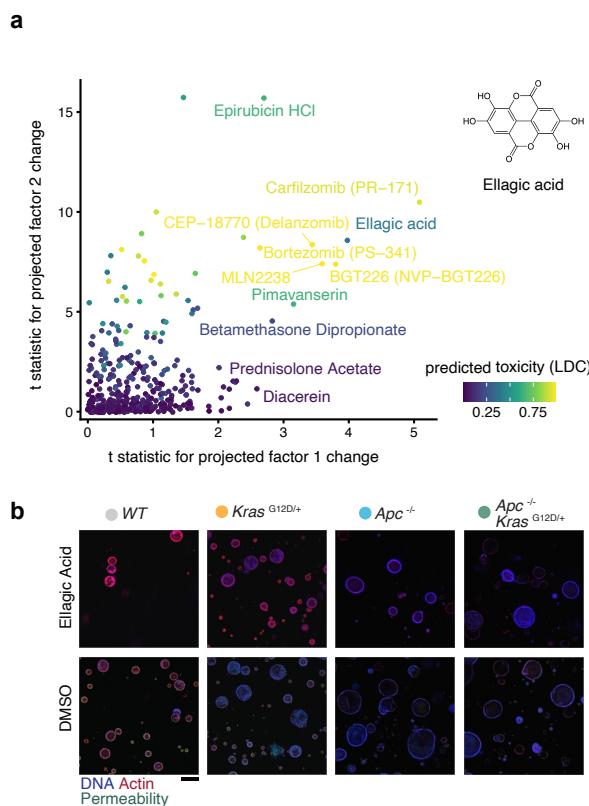


Figure 4.18

I learned that if you work hard and creatively, you can have just about anything you want, but not everything you want. Maturity is the ability to reject good alternatives in order to pursue even better ones.

Ray Dalio

5

Discussion

factor 1 Wnt-STOP

DMSO PDO

CTG LDC

Oncogene induced senescence TP53 dependent

switch from cholesterol esters to TAG

* in fact downregulation of [[beta oxidation]] seen in all 3 mutant organoid lines * in [[KRAS]] also reduction of [[citric acid cycle]] and [[respiratory electron transport]] -> [[warburg effect]] * accumulation of [[triacylglycerol TG]] * [[long chain fatty acids]] - also desaturated as in lizardo * [[cholesterol ester CE]]

and [[phosphatidyl choline PC]] depletion - previously described in adherent cell lines [[lizardo2017.pdf]] *
similar bias toward [[triacylglycerol TG]] has also been observed in [[zebrafish]] models with Kras solo
mutation

* bias away from [[glycerophospholipid]] also observed in [[zebrafish]]

Myc just like reprogramming high canonical Wnt signaling can rejuvenate organoid models interferon
response as mediator of oncogene induced senescence
oncogene induced senescence

why no KRAS alone? - [[oncogene induced senescence]] * why no APC alone? - proliferative advantage? cell
cycle? * weakness of this method is presence of [[EGF]] ligand

Sakai 2017 small intestinal organoids more polyps in AK vs. A! (tumorigenicity in-vitro - other papers) and
in-vivo) no invasion in AK/AKF K necessary but not sufficient for invasion/metastasis no differences in
proliferation A vs. AK they found them to be similar - confirmed with our data AKT program AK is the
most aggressive combo - mouse in-vivo studies - it is the most active form

outlook mouse genetics - coarse phenotypes - Sakai organoid - high genetic control - Matano, PI₃KCA + APC
paper image based profiling coupled with organoids - Rindtorff identify phase space of cancer development in
vitro and ways to therapeutically target it

limitations media condition overemphasize effect of Apc and potentially mask effect of oncogenic Kras
tradeoff between maximizing the phenotype state space while maintaining viable cells

Multiparametric phenotyping of intestinal organoids to model disease
initiation and treatment response in colorectal cancer

ZUSAMMENFASSUNG

Dies ist meine deutsche Zusammenfassung der Promotion

A

Some extra stuff

References

- [1] Broutier, L., Mastrogiovanni, G., Verstegen, M. M. A., Francies, H. E., Gavarró, L. M., Bradshaw, C. R., Allen, G. E., Arnes-Benito, R., Sidorova, O., Gasparsz, M. P., Georgakopoulos, N., Koo, B.-K., Dietmann, S., Davies, S. E., Praseedom, R. K., Lieshout, R., IJzermans, J. N. M., Wigmore, S. J., Saeb-Parsy, K., Garnett, M. J., van der Laan, L. J. W., & Huch, M. (2017). Human primary liver cancer-derived organoid cultures for disease modeling and drug screening. *Nat. Med.*, 23(12), 1424–1435.
- [2] Guinney, J., Dienstmann, R., Wang, X., de Reyniès, A., Schlicker, A., Soneson, C., Marisa, L., Roepman, P., Nyamundanda, G., Angelino, P., Bot, B. M., Morris, J. S., Simon, I. M., Gerster, S., Fessler, E., De Sousa E Melo, F., Missaglia, E., Ramay, H., Barras, D., Homicsko, K., Maru, D., Manyam, G. C., Broom, B., Boige, V., Perez-Villamil, B., Laderas, T., Salazar, R., Gray, J. W., Hanahan, D., Tabernero, J., Bernard, R., Friend, S. H., Laurent-Puig, P., Medema, J. P., Sadanandam, A., Wessels, L., Delorenzi, M., Kopetz, S., Vermeulen, L., & Tejpar, S. (2015). The consensus molecular subtypes of colorectal cancer. *Nat. Med.*, 21(11), 1350–1356.
- [3] Muzny, D. M., Bainbridge, M. N., Chang, K., Dinh, H. H., Drummond, J. A., Fowler, G., Kovar, C. L., Lewis, L. R., Morgan, M. B., Newsham, I. F., Reid, J. G., Santibanez, J., Shinbrot, E., Trevino, L. R., Wu, Y.-Q., Wang, M., Gunaratne, P., Donehower, L. A., Creighton, C. J., Wheeler, D. A., Gibbs, R. A., Lawrence, M. S., Voet, D., Jing, R., Cibulskis, K., Sivachenko, A., Stojanov, P., McKenna, A., Lander, E. S., Gabriel, S., Getz, G., Ding, L., Fulton, R. S., Koboldt, D. C., Wylie, T., Walker, J., Dooling, D. J., Fulton, L., Delehaunty, K. D., Fronick, C. C., Demeter, R., Mardis, E. R., Wilson, R. K., Chu, A., Chun, H.-J. E., Mungall, A. J., Pleasance, E., Gordon Robertson, A., Stoll, D., Balasundaram, M., Birol, I., Butterfield, Y. S. N., Chuah, E., Cope, R. J. N., Dhalla, N., Guin, R., Hirst, C., Hirst, M., Holt, R. A., Lee, D., Li, H. I., Mayo, M., Moore, R. A., Schein, J. E., Slobodan, J. R., Tam, A., Thiessen, N., Varhol, R., Zeng, T., Zhao, Y., Jones, S. J. M., Marra, M. A., Bass, A. J., Ramos, A. H., Saksena, G., Cherniack, A. D., Schumacher, S. E., Tabak, B., Carter, S. L., Pho, N. H., Nguyen, H., Onofrio, R. C., Crenshaw, A., Ardlie, K., Beroukhim, R., Winckler, W., Getz, G., Meyerson, M., Protopopov, A., Zhang, J., Hadjipanayis, A., Lee, E., Xi, R., Yang, L., Ren, X., Zhang, H., Sathiamoorthy, N., Shukla, S., Chen, P.-C., Haseley, P., Xiao, Y., Lee, S., Seidman, J., Chin, L., Park, P. J., Kucherlapati, R., Todd Auman, J., Hoadley, K. A., Du, Y., Wilkerson, M. D., Shi, Y., Liquori, C., Meng, S., Li, L., Turman, Y. J., Topal, M. D., Tan, D., Waring, S., Buda, E., Walsh, J., Jones, C. D., Mieczkowski, P. A., Singh, D., Wu, J., Gulabani, A., Dolina, P., Bodenheimer, T., Hoyle, A. P., Simons, J. V., Soloway, M., Mose, L. E., Jefferys, S. R., Balu, S., O'Connor, B. D., Prins, J. F., Chiang, D. Y., Neil Hayes, D., Perou, C. M., Hinoue, T., Weisenberger, D. J., Maglante, D. T., Pan, F., Berman, B. P., Van Den Berg, D. J., Shen, H., Triche, Jr, T., Baylin, S. B., Laird, P. W., Getz, G., Noble, M., Voet, D., Saksena, G., Gehlenborg, N., DiCara, D., Zhang, J., Zhang, H., Wu, C.-J., Yingchun Liu, S., Shukla, S., Lawrence, M. S., Zhou, L., Sivachenko, A., Lin, P., Stojanov, P., Jing, R., Park, R. W., Nazaire, M.-D., Robinson, J., Thorvaldsdottir, H., Mesirov, J., Park, P. J., Chin, L., Thorsson, V., Reynolds, S. M., Bernard, B., Kreisberg, R., Lin, J., Iype, L., Bressler, R., Erkkilä, T., Gundapuneni,

- M., Liu, Y., Norberg, A., Robinson, T., Yang, D., Zhang, W., Shmulevich, I., de Ronde, J. J., Schultz, N., Cerami, E., Ciriello, G., Goldberg, A. P., Gross, B., Jacobsen, A., Gao, J., Kaczkowski, B., Sinha, R., Arman Aksoy, B., Antipin, Y., Reva, B., Shen, R., Taylor, B. S., Chan, T. A., Ladanyi, M., Sander, C., Akbani, R., Zhang, N., Broom, B. M., Casasent, T., Unruh, A., Wakefield, C., Hamilton, S. R., Craig Cason, R., Baggerly, K. A., Weinstein, J. N., Haussler, D., Benz, C. C., Stuart, J. M., Benz, S. C., Zachary Sanborn, J., Vaske, C. J., Zhu, J., Szeto, C., Scott, G. K., Yau, C., Ng, S., Goldstein, T., Ellrott, K., Collisson, E., Cozen, A. E., Zerbino, D., Wilks, C., Craft, B., Spellman, P., Penny, R., Shelton, T., Hatfield, M., Morris, S., Yena, P., Shelton, C., Sherman, M., Paulauskis, J., Gastier-Foster, J. M., Bowen, J., Ramirez, N. C., Black, A., Pyatt, R., Wise, L., White, P., Bertagnolli, M., Brown, J., Chan, T. A., Chu, G. C., Czerwinski, C., Denstman, F., Dhir, R., Dörner, A., Fuchs, C. S., Guillem, J. G., Iacoboca, M., Juhl, H., Kaufman, A., Kohl, III, B., Van Le, X., Mariano, M. C., Medina, E. N., Meyers, M., Nash, G. M., Paty, P. B., Petrelli, N., Rabeno, B., Richards, W. G., Solit, D., Swanson, P., Temple, L., Tepper, J. E., Thorp, R., Vakiani, E., Weiser, M. R., Willis, J. E., Witkin, G., Zeng, Z., Zinner, M. J., Zornig, C., Jensen, M. A., Sfeir, R., Kahn, A. B., Chu, A. L., Kothiyal, P., Wang, Z., Snyder, E. E., Pontius, J., Pihl, T. D., Ayala, B., Backus, M., Walton, J., Whitmore, J., Baboud, J., Berton, D. L., Nicholls, M. C., Srinivasan, D., Raman, R., Girshik, S., Kigonya, P. A., Alonso, S., Sanbhadt, R. N., Barletta, S. P., Greene, J. M., Pot, D. A., Mills Shaw, K. R., Dillon, L. A. L., Buetow, K., Davidsen, T., Demchok, J. A., Eley, G., Ferguson, M., Fielding, P., Schaefer, C., Sheth, M., Yang, L., Guyer, M. S., Ozenberger, B. A., Palchik, J. D., Peterson, J., Sofia, H. J., & Thomson, E. (2012). Comprehensive molecular characterization of human colon and rectal cancer. *Nature*, 487(7407), 330–337.
- [4] Sato, T., Stange, D. E., Ferrante, M., Vries, R. G. J., Van Es, J. H., Van den Brink, S., Van Houdt, W. J., Pronk, A., Van Gorp, J., Siersema, P. D., & Clevers, H. (2011). Long-term expansion of epithelial organoids from human colon, adenoma, adenocarcinoma, and barrett's epithelium. *Gastroenterology*, 141(5), 1762–1772.
 - [5] Sato, T., Vries, R. G., Snippert, H. J., van de Wetering, M., Barker, N., Stange, D. E., van Es, J. H., Abo, A., Kujala, P., Peters, P. J., & Clevers, H. (2009). Single lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche. *Nature*, 459(7244), 262–265.
 - [6] Schutte, M. (2017). Molecular dissection of colorectal cancer in pre-clinical models identifies biomarkers predicting sensitivity to EGFR inhibitors. *Nat. Commun.*
 - [7] Van De Wetering, M., Francies, H. E., Francis, J. M., Bounova, G., Iorio, F., Pronk, A., Van Houdt, W., Van Gorp, J., Taylor-Weiner, A., Kester, L., McLaren-Douglas, A., Blokker, J., Jaksani, S., Bartfeld, S., Volckman, R., Van Sluis, P., Li, V. S. W., Seepo, S., Sekhar Pedamallu, C., Cibulskis, K., Carter, S. L., McKenna, A., Lawrence, M. S., Lichtenstein, L., Stewart, C., Koster, J., Versteeg, R., Van Oudenaarden, A., Saez-Rodriguez, J., Vries, R. G. J., Getz, G., Wessels, L., Stratton, M. R., McDermott, U., Meyerson, M., Garnett, M. J., & Clevers, H. (2015). Prospective derivation of a living organoid biobank of colorectal cancer patients. *Cell*, 161(4), 933–945.
 - [8] Weeber, F., van de Wetering, M., Hoogstraat, M., Dijkstra, K. K., Krijgsman, O., Kuilman, T., Gadella-van Hooijdonk, C. G. M., van der Velden, D. L., Peper, D. S., Cuppen, E. P. J. G., Vries, R. G., Clevers, H., & Voest, E. E. (2015). Preserved genetic diversity in organoids cultured from biopsies of human colorectal cancer metastases. *Proceedings of the National Academy of Sciences*, 112(43), 201516689.

EIGENANTEIL AN DATENERHEBUNG UND -AUSWERTUNG SOWIE EIGENE
VERÖFFENTLICHUNGEN

MANUSKRIPTE, WELCHE IN DIESER DISSERTATION BEHANDELT WERDEN:

1. BETGE J*, Rindtorff N*, SAUER J*, RAUSCHER B, DINGERT C, GAITANTZI H, HERWECK F, MIER SCH T, VALENTINI E, HAUBER V, GUTTING T, FRANK L, BELLE S, GAISER T, BUCHHOLZ I, JESENOFSKY R, HAERTEL N, ZHAN T, FISCHER B, BREITKOPF-HEINLEIN K, BURGERMEISTER E, EBERT MP AND BOUTROS M. AUTOMATED IMAGE-BASED PROFILING IDENTIFIES DISTINCT COMPOUND INDUCED PHENOTYPES IN PATIENT-DERIVED ORGANOID S. MANUSCRIPT IN PREPARATION.
2. Rindtorff N, SAUER J, BETGE J, ZHAN, T, BOHMANN N, FISCHER B., SCHOLL C, BOUTROS M.
MANUSCRIPT IN PREPARATION

IM RAHMEN DER ERSTGENANNTEN ARBEIT ETABLIERTE ICH DIE METHODIK ZUR KULTIVIERUNG UND HOCHDURCHSATZ-MEDIKAMENTENTESTUNG VON HUMANEN DARMKREBSORGANOÏDEN IN DER ARBEITSGRUPPE VON MICHAEL BOUTROS. VON BEGINN DES PROJEKTES AN WAR ICH MIT JOHANNES BETGE FÜR DIE ZIELSETZUNG UND PLANUNG VERANTWORTLICH. ZUSAMMEN MIT HERRN BETGE, INITIIERTE UND BEGLEITETEN WIR DIE ANALYSE DER MESSERGEBNISSE DURCH JAN SAUER UND, KURZ VOR EINREICHEN DER ERSTEN FASSUNG DES MANUSKRIPTES, AUCH DURCH BENEDIKT RAUSCHER. INSGESAMT BELÄUFT SICH MEIN BEITRAG ZU DIESEM MANUSKRIFT AUF 30-40% DER GESAMTLEISTUNG.

IM RAHMEN DER ZWEITGENANNTEN ARBEIT FÜHRTE ICH DIE KONSEPTION, ETABLIERUNG, EXEKUTION UND DATENANALYSE DES PROJEKTES KOMPLETT SELBSTSTÄNDIG DURCH. FÜR TEILE DER BILDANALYSE NUTZE ICH VON JAN SAUER BEREITES ENTWICKELTE METHODEN. JOHANNES BETGE UND TIANZUO ZHAN DIENTEN ALS BERATER WÄHREND DER INITIALEN PHASE DES PROJEKTES. NICK BOHMANN ARBEITETE ALS EIN BESUCHENDER STUDENT FÜR 6 WOCHE N AUF DEM PROJEKT MIT. BERND FISCHER WAR VOR SEINEM TOD DER ARBEITSGRUPPENLEITER VON JAN SAUER. CLAUDIA SCHOLLS ARBEITSGRUPPE STELLTE DAS MAUSMODELL ZUR VERFÜGUNG. MICHAEL BOUTROS WAR BETREUENDER DOKTORVATER.

WEITERE KOLLABORATIONEN:

1. HEIGWER F, SCHEEDER C, LAUFER C, RAUSCHER B, Rindtorff N, MIERSCH T, SCHMITT B, BLASS C, BOUTROS M. A HIGH-RESOLUTION MAP OF SIGNALING ACROSS THE METAZOAN CELL. SUBMITTED.
2. ZHAN T*, AMBROSI G*, WANDMACHER MA*, RAUSCHER B, BETGE J, Rindtorff N, HAEUSSLER R, HINSENKAMP I, BAMBERG L, HESSLING B, ERDMANN G, BURGERMEISTER E, EBERT MP AND BOUTROS M. MEK INHIBITION ACTIVATES WNT SIGNALLING AND INDUCES STEM CELL PLASTICITY IN COLORECTAL CANCER ORGANOIDS. NATURE COMMUNICATIONS - 2018
3. ZHAN T*, Rindtorff N*, BETGE J, EBERT M, BOUTROS M. CRISPR/CAS9 FOR CANCER RESEARCH AND THERAPY. SEMINARS IN CANCER BIOLOGY - 2017
4. ZHAN T*, RINDTORFF N*, BOUTROS M. WNT SIGNALING IN CANCER. ONCOGENE. - 2016

* EQUAL CONTRIBUTION

Danksagung

ZUM ENDE DIESER DISSERTATION MÖCHTE ICH MICH BEI EINER REIHE VON MENSCHEN BEDANKEN, die mich während der letzten Jahre begleitet haben. Nicht alle Gelisteten sind inzwischen noch unter uns und zu manchen habe ich leider nur noch wenig Kontakt.

EIDESSTATTLICHE VERSICHERUNG

1. BEI DER EINGEREICHTEN DISSERTATION ZU DEM THEMA "MULTIPARAMETRIC PHENOTYPING OF INTESTINAL ORGANOIDS TO MODEL DISEASE INITIATION AND TREATMENT REONSE IN COLORECTAL CANCER" HANDELT ES SICH UM MEINE EIGENSTÄNDIG ERBRACHTE LEISTUNG.
2. ICH HABE NUR DIE ANGEgebenEN QUELLEN UND HILFSMITTEL BENUTZT UND MICH KEINER UNZULÄSSIGEN HILFE DRITTER BEDIENT. INSbesondere HABE ICH WÖRTLICH ODER SINNGEMÄSS AUS ANDEREN WERKEN ÜBERNOMMENE INHALTE ALS SOLCHE KENNTLICH GEMACHT.
3. DIE ARBEIT ODER TEILE DAVON HABE ICH BISLANG NICHT AN EINER HOCHSCHULE DES IN- ODER AUSLANDS ALS BESTANDTEIL EINER PRÜFUNGS- ODER QUALIFIKATIONSLEISTUNG VORGELEGT.
4. DIE RICHTIGKEIT DER VORSTEHENDEN ERKLÄRUNGEN BESTÄTIGE ICH.
5. DIE BEDEUTUNG DER EIDESSTATTLICHEN VERSICHERUNG UND DIE STRAFRECHTLICHEN FOLGEN EINER UNRICHTIGEN ODER UNVOLLSTÄNDIGEN EIDESSTATTLICHEN VERSICHERUNG SIND MIR BEKANNT. ICH VERSICHERE AN EIDES STATT, DASS ICH NACH BESTEM WISSEN DIE REINE WAHRHEIT ERKLÄRT UND NICHTS VERSCHWIEGEN HABE.

HEIDELBERG, DEN