Phosphoproteomic Characterization of Colorectal Cancer for Precision Oncology

Luisa Schwarzmüller

Oral examination: 23.09.2025

Inaugural dissertation

for

obtaining the doctoral degree

of the

Combined faculty of Mathematics, Engineering and Natural Sciences

of the

Ruprecht – Karls – University

Heidelberg

Presented by

M.Sc. Luisa Schwarzmüller

Born in: Germersheim, Germany

Oral examination: 23.09.2025

Phosphoproteomic Characterization of Colorectal Cancer for Precision Oncology

Referees: Prof. Dr. Stefan Wiemann

Prof. Dr. Stefan Fröhling

Declaration of Authorship

I hereby declare that the work presented in my dissertation was carried out between October 2021 and July 2025 under the supervision of Prof. Dr. Stefan Wiemann in the group of Molecular Genome Analysis at the German Cancer Research Center (DKFZ, Heidelberg, Germany).

If not stated otherwise and referenced within the text, the data described in my dissertation are original, have been gathered by myself, and have not yet been presented as a part of a university examination. All main sources, as well as the work of cooperation, have been referenced appropriately.

I, as the author, hereby declare no potential conflict of interest.

Heidelberg,

(Luisa Schwarzmüller)

Table of Contents

List of Figures v
List of Tables vi
Summary 1
Zusammenfassung 2
List of Abbreviations 4
1. Introduction 6
1.1 Cancer 6
1.1.1 Definition and Biological foundations 6
1.1.2 Hallmarks of cancer 6
1.1.3 Cancer staging 7
1.2 Colorectal cancer 7
1.2.1 Epidemiology 7
1.2.2 Classification 8
1.2.3 Metastasis formation 9
1.3 Colorectal cancer treatment 10
1.3.1 Treatment approaches and therapeutic advances 10
1.3.2 Immunotherapy 11
1.3.2.1 Principle of immune checkpoint inhibition 11
1.3.2.2 Effectiveness of ICI treatment of primary and metastatic colorectal cancer 12
1.3.2.3 Tumor mutational burden and PD-L1 expression are insufficient biomarkers in CRC 13
1.3.2.4 Emerging biomarkers for ICI response in CRC 14
1.3.3 The role of KRAS mutations in colorectal cancer: therapeutic implications and treatment paradigms 15
1.4 Precision oncology programs: revolutionizing cancer treatment through personalized medicine 17
1.5 Mass spectrometry-based proteomics 19
1.5.1 Principle of MS-based proteomics measurement 19
1.5.2 Technical advancements of recent years 21
1.5.3 The advantages of phosphoproteomics 21
1.6 Patient-derived organoids for drug screenings 22
1.6 Aims of this thesis 23
2. Results 25
2.1 Technical optimization of proteomic measurements 25
2.1.1 Automated protein clean-up and digestion using SP3 on the Bravo liquid handling robot 25

2.1.2 Optimization of automated phospho-enrichment protocol 27
2.2.1 Optimization of Spectronaut search parameters reduces false discovery rate 31
2.2.2 Assessing the accuracy of relative protein quantification 35
2.2.3 False discovery rate and quantification accuracy of identified phosphopeptides 36
2.3 Optimization of tissue sample processing protocol 39
2.3.1 Establishment of tissue lysis procedure 39
2.3.2 Application of tissue lysis workflow to NCT-MASTER samples reveals high variance in sample quality 42
2.3.3 Variance in sample quality was not related to duration of storage, but rather appeared to be sample-intrinsic 43
2.3.4 Further lysis tests to identify the source of quality differences between samples 45
2.4 Description of colorectal cancer cohort samples received from DKFZ/NCT/DKTK-MASTER program 49
2.5 Quality control of kinase activity estimations 51
2.5.1 Patient samples with few identified phosphosites still gave reliable kinase activity estimates 51
2.5.2 Some kinases share a high proportion of target phosphosites, thus leading to indistinguishable kinase activities 54
2.5.3 Estimated MAPK1 activity based on phosphoproteome footprint agrees largely with the intensities of known activating phosphosites 57
2.5.4 Extending the kinase activity estimation by predicted kinase-substrate interactions 58
2.6 Comparison of the proteome with information from other molecular layers 62
2.6.1 Are gene copy number alterations reflected in the transcriptome and proteome? 62
2.6.2 Correlation of mRNA and protein expression 68
2.6.3 Does kinase activity correlate with protein expression? 71
2.7 Kinase activity-based patient clusters 74
2.7.1 Identification of potential kinase activity-based patient subgroups 74
2.7.2 Characterization of kinase activity-based patient subgroups 76
2.8 Per-patient combination of different molecular information and evaluation of MTB treatment recommendation in the light of proteomic data 80
2.9 Integration of possible biomarkers for immune checkpoint inhibition 84
2.10 Characterization of KRAS-mutated CRC patients to identify potential new vulnerabilities 88
2.10.1 Activating KRAS mutations do not lead to significantly higher activities of downstream signaling 88
2.10.2 Molecular characterization of KRAS-mutated patients reveals high activities of NF-κB, metabolic adaptations, and GSK3β activation 91

2.11 Integrating protein and pathway activities into an interventional basket-based, patient-specific score 95
2.11.1 Calculation of the patient-specific basket scores 95
2.11.2 Validation of basket-based scores using PDOs 99
3. Discussion 107
3.1 Technical developments 108
3.1.1 Workflow automation is crucial for the adaptation of proteomics in the clinical setting 108
3.1.2 False-discovery rates can be controlled with mixed-species experiments 109
3.2 The analyzed CRC cohort was highly heterogeneous, regarding both the sample quality and patient characteristics 109
3.3 Biological insights from the proteomic characterization of CRC patient samples 110
3.3.1 KRAS-mutant CRC tumors have distinct proteomic and signaling profiles that could be clinically relevant. 110
3.3.2 Traditional CRC patient classification does not apply to metastatic, pre-treated cohorts and three kinase activity-based patient subgroups were identified instead. 112
3.3.3 Integrated basket-based score did not predict drug response in patient-derived organoids 114
3.3.4 Protein-level data adds complementary information to genomics and transcriptomics 114
3.4 The lack of proteomics-based clinical studies 115
3.5 Limitations of my study 117
3.6 Outlook 118
4. Materials & Methods 119
4.1 Cell line lysates used for method development 120
4.2 Tissue lysis 120
Protein clean-up and tryptic digestion 121
4.3.1 Manual SP3 121
4.3.2 Automated SP3 122
4.3.3 In-gel 122
4.3.4 PreOmics iST kit 123
4.3.5 In-solution digestion 123
4.4 Enrichment of phosphorylated peptides 124
4.4.1 Manual phosphopeptide-enrichment 124
4.4.2 Automated phosphopeptide-enrichment 125
4.5 HYE species mixture 126
4.6 LC-MS/MS measurement 127

4.7 Mass Spectrometry measurement 127
4.8 Raw data analysis 130
4.9 Data filtering and normalization 131
4.10 Statistics for group comparisons 132
4.11 Transcription factor activities 132
4.12 Pathway activities 132
4.13 Kinase activities 132
4.14 Correlation analyses between CNV, mRNA and protein 132
4.15 Gene set enrichment analysis 133
4.16 Score computation 133
4.17 Use of large language models (LLMs) 134
Acknowledgements 135
Publication bibliography 137

List of Figures

Figure 1: Mechanism of immune checkpoint inhibition 12
Figure 2: Schematic illustration of NCT-MASTER workflow. 18
Figure 3 Principle of mass spectrometry-based protein analysis. 20
Figure 4: Optimization of autoSP3 26
Figure 5: Dilution curve of starting material (autoSP3) 27
Figure 6: Comparison of manual and automated phosphopeptide-enrichment protocols 29
Figure 7: Sensitivity and capacity of automated phospho-enrichment protocol 30
Figure 8: Optimization of phosphoproteome measurement strategy 31
Figure 9: Two-species proteome mixture to control false discovery rate 32
Figure 10: Spectronaut parameter optimization 34
Figure 11: Protein quantification accuracy 36
Figure 12: Controlling false-discovery rate of phosphopeptides 38
Figure 13: Optimization of tissue lysis procedure. 40
Figure 14: Final tissue lysis and MS sample preparation procedure. 41
Figure 15: Application of workflow to DKFZ/NCT-MASTER samples. 42
Figure 16: Identifying the source of sample quality variance 44
Figure 17: Additional tissue lysis tests to find the source of quality variance 46
Figure 18: Full proteome result of lysis test 47
Figure 19: Phosphoproteome result of lysis test. 49
Figure 20: Data availability and clinical patient metadata 51
Figure 21: Patient samples with low phosphoproteome coverage still deliver meaningful insights. 53
Figure 22: Correlating activities of different kinases 54
Figure 23: Numbers of unique substrates distinguishing each kinase from any other kinase 56
Figure 24: MAPK1 known activating sites compared to kinase activity based on footprint 57
Figure 25: Comparison of the Omnipath-based and the extended kinase-substrate-network (KSN) 59
Figure 26: Correlation of classical and extended kinase activities 61
Figure 27: Reflection of copy number alterations on gene expression and protein abundance . 65
Figure 28: Most frequently used copy number gain biomarkers 67
Figure 29: TROP2 expression 68
Figure 30: Correlation of mRNA expression and protein abundance 70
Figure 31: EGFR as an example 72
Figure 32: correlations of all layers 73
Figure 33: Full proteome PCA of CPTAC and my dataset. Color based on CMS subtype. 75

Figure 34: Identification of three patient subgroups with similar kinase activity landscapes. 76
Figure 35 Characterization of kinase activity-based patient subgroups 79
Figure 36: Example patient: mRNA level 82
Figure 37: Example patient: Full proteome level 83
Figure 38: Example patient: kinase activities 84
Figure 39: Potential biomarkers for immune checkpoint inhibition 87
Figure 40: Kinase activities based on classical Omnipath KSN: kinases downstream of KRAS grouped by KRAS mutation status. 89
Figure 41: Kinase activities based on extended KSN: kinases downstream of KRAS grouped by KRAS mutation status. 90
Figure 42: Transcription factor and pathway activities of KRAS-mutated patients in comparison to KRAS WT patients. 92
Figure 43: Full proteome comparison between KRAS-mutated patients and WT patients. 93
Figure 44: Kinase activities of KRAS-mutated patients and phosphosite substrates of GSK3B 94
Figure 45 Schematic illustration of per-patient basket score calculation 96
Figure 46 Final scores per basket per patient. 97
Figure 47: Distribution of basket scores among the kinase activity-based patient clusters and comparison to MTB recommendation. 98
Figure 48: Batch effect removal (full proteome) 100
Figure 49: Proteome-based basket score applied to CRC PDOs 101
Figure 50: Example PDO: basket scoring and drug response 102
Figure 51: Correlation of PDO basket scores with drug response. 103
Figure 52: Correlation of basket score with drugs targeting proteins in that basket 104
Figure 53: Response to Bortezomib grouped by mTOR activity 106

List of Tables

Table 1: Metadata of CRC patient-derived organoids 99
Table 2: List of used products 119
Table 3: Composition of the final lysis buffer 121
Table 4: Cost comparison of automated and manual phosphopeptide-enrichment 126
Table 5: LC gradient (numbers in brackets are for phosphoproteome measurements) 127
Table 6: MS settings for full (phospho) proteome data acquisition 128
Table 7: DIA windows for full proteome data acquisition 129
Table 8: DIA windows for phosphoproteome measurement 130
Table 9: Spectronaut search parameters 131
Table 10: List of used R packages 134

Summary

Colorectal cancer (CRC) is the third most common cancer worldwide and a leading cause of cancer-related mortality. Despite advances in molecular stratification and the integration of genomic and transcriptomic data into clinical decision-making, treatment options for patients with metastatic CRC remain limited, and therapeutic resistance is common. A key challenge in precision oncology is the inability of DNA and RNA profiling alone to capture the functional state of cancer cells, particularly the activity of signaling pathways that govern cell behavior. To address this gap, this thesis presents a phosphoproteomic characterization of CRC within the framework of the DKFZ/NCT/DKTK-MASTER precision oncology program.

I developed and optimized workflows for protein extraction, clean-up, digestion, and phosphopeptide enrichment from fresh-frozen clinical tissue samples using a high-throughput liquid handling robotic platform. These protocols enabled consistent and scalable preparation of patient samples with minimal hands-on time, while maintaining high proteome and phosphoproteome coverage. In parallel, I fine-tuned mass spectrometry acquisition and data analysis strategies to control false discovery rates and improve quantification accuracy, even in samples with low phosphosite abundance.

The resulting dataset revealed substantial heterogeneity in kinase signaling activity among CRC patients. By inferring kinase activation states from phosphosite profiles, I identified three kinase-based patient subgroups with distinct biological and clinical features, including differences in immune infiltration. These subtypes did not overlap with established transcriptomic classifications, highlighting the unique value of phosphoproteomic data in uncovering functional tumor phenotypes.

Importantly, I demonstrate that genomic alterations such as gene amplifications (e.g., *ERBB2*, *CDH17*) do not consistently translate to elevated protein expression or activity. This disconnect suggests that relying solely on DNA or RNA data for therapeutic stratification may lead to suboptimal treatment choices. Moreover, comparison of tumor tissues with patient-derived organoids (PDOs) revealed marked divergence in proteomic signatures, calling into question the fidelity of PDOs as preclinical models, particularly in the absence of the tumor microenvironment.

Overall, this work provides both technical and conceptual advances for integrating phosphoproteomics into precision oncology workflows. It underscores the need to consider protein-level information when making clinical decisions and sets the stage for future studies evaluating proteome-informed treatment strategies in CRC and beyond.

Zusammenfassung

Darmkrebs ist die dritthäufigste Krebsart weltweit und eine der Hauptursachen für krebsbedingte Sterblichkeit. Trotz Fortschritten bei der molekularen Stratifizierung und der Integration genomischer und transkriptomischer Daten in die klinische Entscheidungsfindung sind die Behandlungsmöglichkeiten für Patienten mit metastasiertem Darmkrebs nach wie vor begrenzt, und Therapieresistenz ist häufig. Eine zentrale Herausforderung in der Präzisionsonkologie besteht darin, dass DNA- und RNA-Profile allein nicht in der Lage sind, den Funktionszustand von Krebszellen zu erfassen, insbesondere die Aktivität von Signalwegen, die das Zellverhalten steuern. Um diese Lücke zu schließen, wird in dieser Arbeit eine phosphoproteomische Charakterisierung von Darmkrebs im Rahmen des DKFZ/NCT/DKTK MASTER-Präzisionsonkologieprogramms vorgestellt.

Ich entwickelte und optimierte Arbeitsabläufe für die Proteinextraktion, die Reinigung, den Verdau und die Anreicherung von Phosphopeptiden aus frischen, gefrorenen klinischen Gewebeproben unter Verwendung einer Hochdurchsatz-Roboterplattform für die Flüssigkeitsverarbeitung. Diese Protokolle ermöglichten eine konsistente und skalierbare Vorbereitung von Patientenproben mit minimalem Zeitaufwand, während gleichzeitig eine hohe Proteom- und Phosphoproteom-Abdeckung gewährleistet wurde. Parallel dazu habe ich die Strategien für die Massenspektrometrieerfassung und Datenanalyse feinabgestimmt, um die Falschentdeckungsraten zu kontrollieren und die Quantifizierungsgenauigkeit zu verbessern, selbst bei Proben mit geringer Phosphoproteom-Abdeckung.

Der resultierende Datensatz zeigte eine erhebliche Heterogenität der Kinase-Signalaktivität bei Darmkrebs-Patient*innen. Durch die Ableitung von Kinase-Aktivierungszuständen aus Phosphopeptid-Profilen habe ich drei Kinaseaktivitäts-basierte Patienten-Untergruppen mit unterschiedlichen biologischen und klinischen Merkmalen, einschließlich Unterschieden bei der Immuninfiltration, identifiziert. Diese Subtypen überschneiden sich nicht mit etablierten transkriptomischen Klassifizierungen, was den einzigartigen Wert phosphoproteomischer Daten bei der Aufdeckung funktioneller Tumorphänotypen unterstreicht.

Wichtig ist, dass ich zeigen konnte, dass genomische Veränderungen wie Genamplifikationen (z. B. *ERBB2*, *CDH17*) nicht durchgängig mit einer erhöhten Proteinexpression oder -aktivität einhergehen. Diese Diskrepanz deutet darauf hin, dass die ausschließliche Verwendung von DNA- oder RNA-Daten für die therapeutische Stratifizierung zu suboptimalen Behandlungsentscheidungen führen kann. Darüber hinaus ergab der Vergleich von Tumorgewebe mit von Patienten gewonnenen Organoiden (PDOs) deutliche Unterschiede in der Proteomsignatur, was die Zuverlässigkeit von PDOs als präklinische Modelle in Frage stellt, insbesondere wenn die Mikroumgebung des Tumors fehlt.

Insgesamt bietet diese Arbeit sowohl technische als auch konzeptionelle Fortschritte für die Integration der Phosphoproteomik in die Arbeitsabläufe der Präzisionsonkologie. Sie unterstreicht die Notwendigkeit, Informationen auf Proteinebene bei klinischen Entscheidungen zu berücksichtigen, und schafft die Voraussetzungen für künftige Studien, die proteomgestützte Behandlungsstrategien bei Darmkrebs und darüber hinaus untersuchen.

List of Abbreviations

CRC	Colorectal cancer
MSI	Microsatellite instability
MSS	Microsatellite stable
EMT	Epithelial-mesenchymal transition
CSC	Cancer stem cells
5-FU	5-fluorouracil
TME	Tumor microenvironment
ADC	Antibody-drug conjugate
MASTER	Molecularly aided stratification for tumor eradication
MTB	Molecular tumor board
TMB	Tumor mutational burden
SP3	Single pot solid phase enhanced sample preparation
DIA	Data-independent acquisition
PEP	Posterior error probability
HYE	Human-yeast-e.coli
FDR	False discovery rate
TEAB	Triethylammoniumbicarbonat
EDTA	Ethylendiamintetraacetat
SDS	Sodium dodecyl sulfate
PBS	Phosphate buffered saline
FBS	Fetal bovine serum
SPL	Sample processing lab
DNA	deoxyribonucleic acid
(m)RNA	(messenger) ribonucleic acid
SNV	Single nucleotide variant
CNA	Copy number alteration
CNV	Copy number variant
(T)CN	(total) copy number
WGS	Whole genome sequencing
WES	Whole exome sequencing
BH	Benjamini-Hochberg
KSN	Kinase-substrate network

CMS	Consensus molecular subtype
CC	Cell cycle
HR	Hormone receptor
RME	Ras-Mek-Erk
RTK	Receptor tyrosine kinase
DDR	DNA damage response
IS	Immune system
GSEA	Gene set enrichment analysis
ICI	Immune checkpoint inhibition
MHC	Major histocompatibility complex
WT	Wild type
TF	Transcription factor
PC(A)	Principal component (analysis)
PDO	Patient-derived organoid
CPTAC	Clinical proteome tumor analysis consortium
HRD	Homologous repair deficiency
FDA	Food and drug administration
AUC	Area under the curve

1. Introduction

1.1 Cancer

1.1.1 Definition and Biological foundations

Cancer represents one of the most significant challenges in modern medicine, affecting millions globally and imposing substantial burdens on healthcare systems worldwide. Despite remarkable advances in our understanding of cancer biology and the development of novel therapeutic approaches, cancer remains a leading cause of mortality, with about 20 million new cases diagnoses annually (Bray et al. 2024). Cancer is fundamentally a disease of cellular dysregulation characterized by uncontrolled proliferation, resistance to cell death signals, and the potential for invasion and metastasis (Hanahan and Weinberg 2011; Hanahan 2022). At its core, cancer arises from genetic alterations that progressively transform normal cells into malignant ones through a complex multistep process (Cairns 1975; Nowell 1976; Vogelstein et al. 2013; Stratton et al. 2009; Alexandrov and Stratton 2014). This transformation represents not merely a single disease but a diverse collection of disorders unified by the common feature of abnormal and unregulated cell growth. The completion of the Human Genome Project in 2004 marked a milestone in our understanding of cancer biology, enabling unprecedented insights into the genetic basis of malignancy (International Human Genome Sequencing Consortium 2004). The subsequent technological revolution has dramatically advanced our capacity to sequence individual cancer genomes rapidly and at low cost, revealing the extraordinary genetic complexity underlying different cancer types. Within a relatively short period following the human genome sequencing, whole-genome analysis has uncovered previously unknown cancer genes, facilitated novel drug development, illuminated unsuspected flexibility in human DNA, and provided alternative strategies for tumor classification (reviewed by Koboldt et al. (2013).

1.1.2 Hallmarks of cancer

Our contemporary understanding of cancer biology is organized around several fundamental capabilities that cancer cells must acquire to enable malignant growth. These "hallmarks of cancer" provide a conceptual framework for understanding the complexity of neoplastic disease (Hanahan and Weinberg 2011; Hanahan 2022). These hallmarks include:

- Sustained proliferative signaling
- Evasion of growth suppressors
- Resistance to cell death
- Enabling replicative immortality
- Induction of angiogenesis
- Activation of invasion and metastasis

More recent additions to these hallmarks include metabolic reprogramming, immune evasion, genomic instability, and tumor-promoting inflammation. These capabilities are not acquired simultaneously but rather develop progressively through various mechanisms including genetic mutations, epigenetic alterations, and interactions with the tumor microenvironment (Hanahan 2022).

1.1.3 Cancer staging

The Tumor-Node-Metastasis (TNM) staging system represents the cornerstone of cancer staging, providing a standardized approach to describing the anatomical extent of disease (Wittekind 2017). This system classifies cancers based on three key parameters:

T: Primary tumor characteristics (size and local invasion)

N: Regional lymph node involvement

M: Presence or absence of distant metastasis

Stage I tumors, representing the earliest detectable malignancies, typically manifest as small, localized neoplasms (T1-2) without evidence of regional lymph node involvement (N0) or distant metastatic spread (M0). Stage II tumors represent locally advanced malignancies that have grown beyond their initial site but remain confined to the region of origin without distant metastasis. A common example is T3N0M0 in colorectal cancer, where the tumor invades into peri-colorectal tissues (T3) but shows no lymph node involvement (N0) or distant spread (M0) (Zhang et al. 2021). Stage III cancer represents locally advanced disease (T1-4) characterized by regional lymph node metastasis (N1-2) without distant spread (M0). Stage IV tumors represent the most advanced cancer classification, characterized by distant metastatic spread beyond regional lymph nodes. The TNM pattern for Stage IV universally includes M1 designation, indicating detectable distant metastases, combined with any T (primary tumor) and N (regional lymph node) categories (Brierley et al. 2025).

1.2 Colorectal cancer

1.2.1 Epidemiology

Colorectal cancer represents a substantial and growing global health burden, ranking as the second most deadly and third most commonly diagnosed cancer worldwide. According to global cancer statistics (GLOBOCAN) 2022 data, 1.9 million new cases and nearly 1 million deaths were expected globally (Bray et al. 2024), with projections suggesting an increase by 2030 to more than 2.2 million new cases and 1.1 million deaths (Arnold et al. 2017). Colorectal cancer accounts for 10% of all cancer diagnoses worldwide, making it the third most common cancer after lung and female breast cancer. As the second leading cause of cancer mortality globally, colorectal cancer

was responsible for an estimated 9.3% of all cancer deaths in 2022, highlighting its significant impact on public health (Bray et al. 2024). While incidence rates remain among the highest in highly developed nations, these regions are experiencing stabilizing or decreasing trends due to advances in screening and prevention, whereas many low-income and middle-income countries continue to see rapidly rising incidence rates as they adopt more westernized lifestyles associated with increased colorectal cancer risk (Arnold et al. 2017; Bray et al. 2024).

A particularly concerning trend in colorectal cancer epidemiology is the increasing incidence among young adults, a phenomenon observed across multiple countries despite overall declining or stabilizing rates in older populations. In the United States, colon cancer incidence rates increased by 1.0% to 2.4% annually since the mid-1980s in adults aged 20 to 39 years and by 0.5% to 1.3% since the mid-1990s in adults aged 40 to 54 years. Even more dramatically, rectal cancer incidence rates have been increasing longer and faster, with an annual growth rate of 3.2% from 1974 to 2013 in adults aged 20-29 years. This shift in age distribution has been accompanied by a doubling in the proportion of rectal cancers diagnosed in adults younger than 55 years, from 14.6% to 29.2% between 1989-1990 and 2012-2013 (Siegel et al. 2017). Similar patterns of increasing early-onset colorectal cancer have been documented internationally, with a comprehensive global analysis revealing rising incidence in young adults in 19 countries, nine of which simultaneously demonstrated stable or declining trends in older adults, including Australia, Canada, Denmark, Germany, New Zealand, Slovenia, Sweden, the United Kingdom, and the United States. Particularly steep increases in young adult colorectal cancer were observed in Korea and New Zealand, with average annual percent changes of 4.2% and 4.0%, respectively (Siegel et al. 2019). These concerning trends in early-onset colorectal cancer highlight the need for heightened awareness, research into underlying causes, and consideration of revised screening guidelines for younger populations.

1.2.2 Classification

Colorectal cancer encompasses a heterogeneous group of malignancies that can be classified based on anatomical location, histological features, and molecular characteristics (Schneider and Langner 2014). The traditional anatomical distinction between colon and rectal cancers remains clinically relevant, with each demonstrating distinct epidemiological patterns, treatment approaches, and outcomes (Paschke et al. 2018). From a molecular perspective, colorectal cancers can be categorized based on specific genetic and epigenetic alterations that influence disease behavior and treatment response. Major molecular pathways in colorectal carcinogenesis include chromosomal instability (CIN), microsatellite instability (MSI), and the CpG island methylator phenotype (CIMP). MSI-high tumors, which result from deficiencies in DNA mismatch repair, typically demonstrate better prognosis and different treatment responses compared to

microsatellite stable tumors. The presence of specific oncogenic mutations, such as those in *KRAS*, *NRAS*, and *BRAF* genes, further refines tumor classification and has significant implications for targeted therapy selection. Histologically, the vast majority of colorectal cancers are adenocarcinomas, with less common variants including mucinous adenocarcinoma, signet ring cell carcinoma, and neuroendocrine tumors (International Agency for Research on Cancer 2010). Carcinomas are cancers that originate from epithelial cells, which line the surfaces of organs and tissues throughout the body. They are the most common type of cancer and include subtypes based on the specific epithelial tissue affected, e.g. adenocarcinomas arise from glandular epithelial cells (Hinck and Näthke 2014).

The commonly used Consensus Molecular Subtypes (CMS) classification system for colorectal cancer (CRC) is based on transcriptomic analysis of specific gene expression patterns that define four distinct molecular subtypes. The classification was developed through a collaborative effort by the Colorectal Cancer Subtyping Consortium, which analyzed gene expression data from over 4,000 CRC samples (Guinney et al. 2015). CMS1 is characterized by genes associated with immune activation and microsatellite instability, CMS2 (Canonical) displays epithelial characteristics with marked WNT and MYC signaling activation, CMS3 (Metabolic) is defined by genes related to metabolic dysregulation, and CMS4 (Mesenchymal) exhibits prominent TGFβ activation, angiogenesis, and stromal invasion.

1.2.3 Metastasis formation

At the time of diagnosis, 25-30% of CRC patients already present with distant metastases, while in an additional 15-20% of initially metastasis-free patients the disease will progress and metastases develop eventually (Rahbari et al. 2019). While the overall survival rate of patients with CRC is 64% (REF), this drops to 14% for patients that are diagnosed with metastatic CRC (Siegel et al. 2023) and treatment options remain limited.

Cancer metastasis involves a multi-step process where cells from the primary tumor invade distant tissues, as reviewed by Cañellas-Socias et al. (2024). Metastases most frequently arise in the liver because the portal vein carries circulating tumor cells in the blood from the gastrointestinal tract to the liver. Further frequent metastatic sites include the lungs and the peritoneal cavity (Riihimäki et al. 2016). Genetic alterations affecting WNT, EGF, TGFb and BMP signaling pathways enable the disseminated cancer cells to grow autonomously in a new host environment (Fujii et al. 2016). Also therapy-induced damage has been shown to stimulate a fetal-like CRC cell state that allows the tumor to recover and metastasize (Moorman et al. 2025; Ohta et al. 2022; Solé et al. 2022).

At host sites, metastatic niches are established through interactions with host cells. For example, TGFb-activated fibroblasts can produce ECM components, cytokines and survival factors that

facilitate tumor invasion (Calon et al. 2012) and have been shown to exclude T-cells (Tauriello et al. 2018). Different metastases (at different host organs) exhibit different immune environments (Angelova et al. 2018) and adaptations and Cañellas-Socias et al. (2024) even suggest that "each individual metastasis should be considered a disease on its own". In conclusion, metastatic CRC differs from the primary tumor with regards to prognosis, treatment response, and immune environment and although mCRC poses a major challenge, research has mainly focused on the primary disease. For instance, the Clinical Proteomic Tumor Analysis Consortium (CPTAC) currently provides the only available phosphoproteome dataset of CRC and was acquired from primary, treatment-naïve tumor samples (Vasaikar et al. 2019).

1.3 Colorectal cancer treatment

1.3.1 Treatment approaches and therapeutic advances

The management of colorectal cancer involves a multidisciplinary approach combining surgical intervention, radiation therapy, and systemic treatments tailored to disease stage, location, and molecular characteristics. For localized disease, surgical resection remains the cornerstone of curative treatment, with total mesorectal excision representing the standard approach for rectal cancer to minimize local recurrence risk (Heald and Ryall 1986). Neoadjuvant chemoradiation has become standard practice for locally advanced rectal cancer, enabling tumor downsizing and potentially allowing sphincter-preserving surgery in selected cases (Loria et al. 2024). Adjuvant chemotherapy, typically fluoropyrimidine-based regimens with or without oxaliplatin, is recommended for high-risk stage II and stage III colon cancer to reduce recurrence risk following surgical resection (Taieb and Gallois 2020).

For metastatic colorectal cancer, systemic therapy options have expanded significantly in recent decades, incorporating conventional cytotoxic agents (fluoropyrimidines, irinotecan, oxaliplatin) and targeted therapies, such as anti-EGFR antibodies for RAS wild-type tumors or anti-VEGF agents. Combining EGFR or VEGFR targeting therapies with chemotherapy has improved outcomes in some groups of mCRC patients. Almost 50% responded to VEGFR-targeting therapy (Hurwitz et al. 2004) and 10% responded to EGFR-targeting therapy (van Cutsem et al. 2007). This therapy is however not curative. The development of antibody-drug conjugates (ADCs), such as Sacituzumab govitecan, targeting trophoblast cell surface antigen 2 (TROP2) represents a promising therapeutic approach for metastatic CRC. TROP2 expression demonstrates significant associations with aggressive tumor characteristics and adverse clinical outcomes in colorectal cancer (Foersch et al. 2024) and treatment of metastatic CRC with Sacituzumab govitecan is currently being evaluated in a phase II/III clinical trial (Köhler et al. 2025).

1.3.2 Immunotherapy

1.3.2.1 Principle of immune checkpoint inhibition

The development of immune checkpoint inhibitors (ICIs) has revolutionized treatment paradigms for several cancer types. This type of immunotherapy is based on the mechanism that mutated proteins are presented on the cancer cell surface as neoantigens and can be recognized by the adaptive immune system (Schumacher 2015). Neoantigen recognition by T cells represents a critical mechanism in cancer immunosurveillance, enabling the immune system to distinguish malignant cells from healthy tissues. Neoantigens arise from somatic mutations in tumor cells, including single-nucleotide variants (SNVs), insertions/deletions (indels), and gene fusions. These mutations generate novel peptide sequences absent in normal cells, creating targets for T-cell recognition (Sim and Sun 2022). The ubiquitin-proteasome system degrades tumor proteins into peptides (van den Eshof et al. 2021) which are then transported to the endoplasmic reticulum (ER) where they are loaded onto MHC-I molecules (Blees et al. 2017). Loaded MHC I-neoantigen complexes traffic to the cell surface. Each MHC I molecule presents a single peptide (8–11 amino acids) for recognition by CD8+ T cells. Mutations that enhance peptide-MHC binding (e.g., KRASG12D) increase immunogenicity by improving complex stability. Cytotoxic CD8+ T-cells infiltrating the tumor can recognize presented neoantigens by highly specific interactions between the T-cell receptors (TCRs) and the peptide-MHC-I complex. TCR binding induces structural shifts in both the TCR and peptide-MHC, enabling discrimination between neoantigens and self-peptides (Mariuzza et al. 2023; Sim and Sun 2022). Upon neoantigen recognition, T cells kill cancer cells primarily by releasing perforin and granzymes to induce apoptosis, and by engaging death receptors such as Fas. This process not only eliminates tumor cells but can also stimulate broader immune responses against the tumor (Durgeau et al. 2018; Jaime-Sanchez et al. 2020; Haabeth et al. 2014). In response, many tumors upregulate PD-L1 on their surface, which binds to PD-1 receptors on T cells. This interaction sends an inhibitory signal to T cells, reducing their activity and leading to T-cell exhaustion or dysfunction (Mundhara and Sadhukhan 2024; Suresh and O'Donnell 2021; Manjili et al. 2012). Tumors may also exploit other checkpoint pathways, such as CTLA-4, LAG-3, and TIM-3, to suppress T-cell responses (Suresh and O'Donnell 2021; Lehtiö et al. 2021). Immune checkpoint inhibitors are monoclonal antibodies that target PD-1, PD-L1 or other immune checkpoints (CTLA4) and thereby re-active the cytotoxic T-cell response (Benelli et al. 2024; Medina and Adams 2016; Qiu et al. 2024).

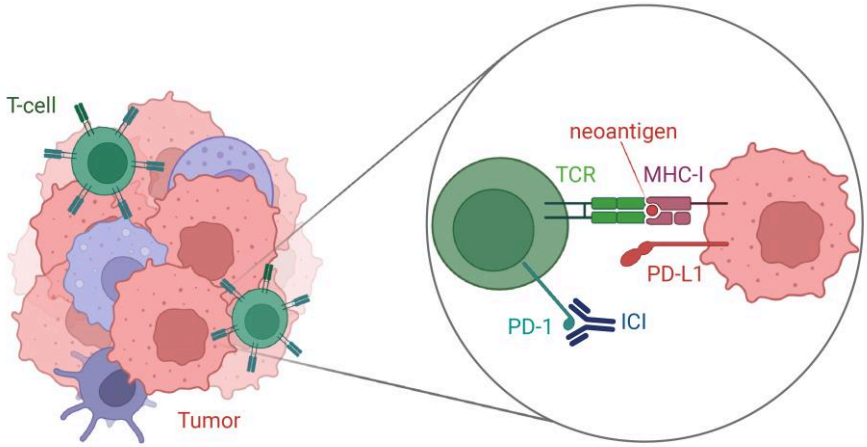


Figure 1: Mechanism of immune checkpoint inhibition
Neoantigen recognition by CD8+ cytotoxic T cells is performed by specific interactions between the T-cell receptors (TCRs) and the peptide-MHC-I complex. Many tumors upregulate PD-L1 on the cell surface, which binds to PD-1 receptors on T cells and sends an inhibitory signal, leading to T-cell exhaustion or dysfunction. Immune checkpoint inhibitors are monoclonal antibodies that target for example PD-1 and re-active the cytotoxic T-cell response.
Figure created with Biorender.

1.3.2.2 Effectiveness of ICI treatment of primary and metastatic colorectal cancer

Immune checkpoint blockade is highly effective in CRC patients that exhibit deficiencies in the mismatch repair machinery and high microsatellite instability (dMMR/MSI-H), which renders them responsive to immunotherapy (San-Román-Gil et al. 2023). These tumors are characterized by a dysfunctional DNA repair system that generates numerous neoantigens, creating a highly immunogenic tumor microenvironment more susceptible to immune checkpoint inhibitors. The effectiveness of ICIs in these 5-15% of CRC cases was demonstrated by several clinical trials such as the NICHE (Chalabi et al. 2020) and NICHE2 (Chalabi et al. 2022) studies that both reported 100% pathological response of early-stage, non-metastatic MSI tumors to neoadjuvant treatment with ipilimumab and nivolumab. Also single-agent treatment targeting PD-1 resulted in complete clinical response of all 12 evaluated early-stage MSI patients in a small phase II study (Cercek et al. 2022). However, the vast majority of CRC patients are microsatellite stable (MSS) and demonstrate primary resistance to immunotherapy, significantly limiting the broader applicability of this therapeutic approach (Ilerhunmwuwa et al. 2025; Chalabi et al. 2020). While a subset of MSS patients could benefit from ICI (Chalabi et al. 2020), response biomarkers are unclear (Guven et al. 2024).

While the benefit of ICI in early-stage MSI-H CRC has been established, the response rate in metastatic CRC (mCRC) is considerably lower. Metastases at different host organs exhibit different immune activities and persistent metastatic clones have the ability to escape the immune

system (Angelova et al. 2018) which could influence the responsiveness to ICI. Several studies showed that response to ICI is influenced by the specific organs affected by metastasis (Yu et al. 2021; Fakih et al. 2024). Furthermore, the proportion of microsatellite stable (MSS) cases is higher in mCRC than in primary CRC (Sinicrope and Sargent 2012) and thus 95% of mCRC patients rarely benefit from ICI treatment (Le et al. 2015). The KEYNOTE-177 study tested the effectiveness of pembrolizumab as a first-line therapy for dMMR/MSI-H mCRC. This phase III randomized trial demonstrated a response rate of 43.8% and a median progression-free survival of 16.5 months with pembrolizumab monotherapy, significantly outperforming conventional chemotherapy with a response rate of 33.1% and median progression-free survival of 8.2 months (André et al. 2020). The long-term benefits of pembrolizumab were further confirmed through 5-year follow-up where median overall survival with pembrolizumab was 77.5 months, compared to 36.7 months with chemotherapy, although 62% of patients in the chemotherapy arm crossed over to pembrolizumab treatment during the course of the study (André et al. 2025). In conclusion, while ICI treatment outperforms chemotherapy also in the metastatic setting, the response rate is lower than in primary CRC tumors and more than 50% of mCRC patients do not benefit from the therapy. Ongoing research aims to identify reliable response biomarkers.

1.3.2.3 Tumor mutational burden and PD-L1 expression are insufficient biomarkers in CRC

Tumor mutational burden (TMB), with the total number of coding mutations as a proxy for neoantigen load (Wen et al. 2022; Manca et al. 2023), is commonly used as a response biomarker for ICI treatment (Moeckel et al. 2023). Pembrolizumab (anti PD-1) was the first ever cancer therapy the FDA approved for any metastatic, solid tumor with high mutational burden independent of the cancer type (FDA 2020). While high TMB is associated with better responses to ICIs in several cancers, such as melanoma and lung cancer, its predictive value for CRC is not as strong or consistent (Li et al. 2021). The recognition of neoantigens by immune cells is a complex process that is regulated at many levels besides the mere presence of coding mutations. Current research efforts are intensively focused on characterizing these immunotherapy resistance mechanisms and identifying predictive biomarkers to guide therapeutic decision-making for optimal patient selection and treatment outcomes.

Measuring PD-L1 expression levels of tumor cells by immunohistochemistry (IHC) was the first FDA-approved response biomarker used in clinical trials to test the efficacy of PD-1-targeting antibodies (nivolumab, atezolizumab) (Topalian et al. 2012; Herbst et al. 2014). Although PD-L1 abundance undoubtedly helps identify patient populations that likely benefit from ICI therapy in many cancer types (Doroshow et al. 2021), in the case of colorectal cancer PD-L1 expression alone is not sufficient to reliably predict clinical benefit. A 2024 systematic review of 36 studies

(*n*=1,867 patients) found no significant association between tumor PD-L1 expression and ICI response in CRC (OR 1.01, 95% CI 0.48–2.14) (Yu et al. 2024).

1.3.2.4 Emerging biomarkers for ICI response in CRC

High expression of CD74 shows emerging potential as a predictive biomarker for ICI response in colorectal cancer, both in MSS (Acha-Sagredo et al. 2025) and MSI (Sun et al. 2025) patients. A 2024 clinical trial (NCT03261011) involving advanced solid tumors, including CRC, found that high CD74 protein expression (≥10% tumor area occupied by CD74+ immune cells) correlated with improved progression-free survival (HR = 0.21, 95% CI 0.12–1.08, p = 0.0615) in patients treated with the PD-1/CTLA-4 bispecific antibody AK104. Furthermore, high CD74 expression was associated with enhanced immune cell infiltration in the tumor microenvironment (TME), suggesting a more inflamed, ICI-responsive phenotype. Mechanistically, CD74 is a chaperone for MHC class II antigen presentation and interacts with macrophage migration inhibitory factor (MIF), influencing T-cell activation and tumor immunity (Wang et al. 2024).

A further known immune evasion mechanism consists of cancer cells decreasing the expression of MHC class I molecules on the cell surface (Gettinger et al. 2017; Sade-Feldman et al. 2017; Angell et al. 2014), which are needed to present tumor neoantigens to T cells. Without antigen presentation, T cells cannot recognize or target the cancer cells (Kallingal et al. 2023).

Additionally, the extent and types of immune cells infiltrating the solid tumor substantially influence the effectiveness of ICI treatment. Particularly the infiltration of CD8+ T cells is a predictive factor for response to ICIs across multiple cancer types (Kazama et al. 2022; Kümpers et al. 2019; Nojima et al. 2021). In microsatellite-stable (MSS) CRC, which make up the majority of cases, immune infiltration is generally lower. However, a phase II trial (AtezoTRIBE) showed that patients with higher immune cell infiltration had significantly longer progression-free survival when treated with a combination of chemotherapy, bevacizumab, and atezolizumab (an anti-PD-L1 ICI), compared to those with low immune infiltration (Moretto et al. 2023).

In the case of metastatic CRC, predicting ICI effectiveness is even more complex. The presence of liver metastases presents a particular challenge for immunotherapy efficacy in colorectal cancer patients. Liver metastasis represents the most common site of spread in CRC (Riihimäki et al. 2016), and research indicates that the presence of liver metastases significantly limits the efficacy of immunotherapy regardless of microsatellite instability status (Sahin et al. 2021). The exact mechanisms underlying this resistance phenomenon remain poorly understood, though several factors within the liver tumor microenvironment (TME) have been implicated. The immunosuppressive nature of the liver microenvironment, characterized by unique cellular populations and signaling networks, creates conditions unfavorable for effective anti-tumor

immune responses despite checkpoint inhibition (Shen et al. 2024). This observation underscores the complexity of immunotherapy resistance in colorectal cancer, particularly in the metastatic setting.

In conclusion, the interplay of multiple possible response biomarkers is complex, likely tumor type-specific, and experts suggest "the integration of multiple biomarkers as a path forward to most accurately predict response in patients" (Wang and Chan 2025).

1.3.3 The role of KRAS mutations in colorectal cancer: therapeutic implications and treatment paradigms

KRAS mutations represent one of the most clinically significant molecular alterations in colorectal cancer (CRC), profoundly influencing treatment decisions and patient outcomes. These mutations occur in 40-45% of patients with metastatic CRC (Ilhan et al. 2025; Osterlund et al. 2022; Radanova et al. 2023) and serve as crucial biomarkers for therapeutic selection, particularly regarding targeted therapies. *KRAS* mutations most commonly affect codons 12 and 13 in exon 2 (Ilhan et al. 2025), though clinically relevant mutations also include codon 61 (Q61) and codon 146 (A146) (Hunter et al. 2015). Among the specific *KRAS* variants, the G12D mutation is most common, occurring in approximately 36% of metastatic colorectal cancers (Neumann et al. 2009). Conversely, The *KRAS* G12C mutation, which has attracted particular interest due to recent therapeutic developments in lung cancer (Skoulidis et al. 2021; Langen et al. 2023), has a reported prevalence of only 3% in CRC patients (Potocki et al. 2023; Osterlund et al. 2022; Chida et al. 2021; Henry et al. 2021).

KRAS mutations define a subpopulation of colorectal cancer patients with inferior prognosis compared to those with wild-type tumors (Deng et al. 2015). The most established clinical application of *KRAS* mutation testing in colorectal cancer relates to its predictive value for response to anti-epidermal growth factor receptor (EGFR) therapy. Multiple studies have consistently demonstrated that patients with *KRAS*-mutant tumors derive minimal benefit from anti-EGFR monoclonal antibodies such as cetuximab and panitumumab (Chang et al. 2009). This therapeutic resistance occurs because *KRAS* mutations constitutively activate downstream signaling pathways independent of EGFR activation, rendering EGFR blockade ineffective. However, combining drugs against KRAS signaling and EGFR has shown promising results in early clinical trials (Miyashita and Hong 2024; Kataoka et al. 2023).

Similarly, resistance to directly targeted KRAS therapies also involves molecular adaptations. In preclinical studies of KRAS G12D inhibitors, continuous treatment led to the development of secondary KRAS mutations that conferred resistance to the targeted agent (Choi et al. 2024). This suggests that resistance to KRAS-directed therapies may follow patterns similar to those observed

with other targeted agents, involving adaptation through additional mutations that restore downstream pathway activation despite inhibitor presence.

The first-line systemic therapy for metastatic colorectal cancer typically consists of a combination of one targeted therapy agent and a chemotherapy doublet. For patients with KRAS wild-type tumors, particularly those with left-sided primary tumors, anti-EGFR antibodies combined with chemotherapy represent a standard first-line approach. However, for patients with *KRAS*-mutant disease, treatment options are more limited, with anti-VEGF therapy (bevacizumab) plus chemotherapy serving as the primary treatment approach (Liang et al. 2023). The selection of specific chemotherapy regimens is typically guided by factors such as patient performance status, comorbidities, and toxicity profiles rather than *KRAS* mutation status itself. Beyond first-line treatment, options for *KRAS*-mutant colorectal cancer have historically been limited to sequential chemotherapy regimens and continued anti-angiogenic therapy. The poor prognosis associated with *KRAS* mutations in metastatic colorectal cancer underscores the substantial unmet need for effective targeted approaches specifically designed for this molecular subtype (Li et al. 2020). This clinical challenge has driven extensive research efforts aimed at developing novel therapeutic strategies to overcome the inherent resistance of *KRAS*-mutant tumors to available targeted agents. Recent advances in medicinal chemistry and drug development have led to significant progress in directly targeting mutated KRAS proteins. KRAS G12C inhibitors have shown promising results in non-small cell lung cancer (Mok et al. 2024) and are now being evaluated in colorectal cancer patients with this specific mutation, however, show little effect (Piazza et al. 2024; Potocki et al. 2023). Additionally, KRAS G12D-specific inhibitors such as MRTX1133 have entered clinical trials for patients with advanced solid tumors harboring the common KRAS G12D alteration (NCT05737706). Furthermore, pan-KRAS inhibitors have been developed and currently undergo clinical trials (NCT06607185, NCT06625320, NCT06056024). These direct KRAS-targeting approaches represent a potential paradigm shift in treating KRAS-mutant colorectal cancer, offering the possibility of effective targeted therapy for patients previously limited to cytotoxic chemotherapy options.

Alternative strategies to address KRAS-mutant colorectal cancer include targeting downstream effectors of KRAS signaling. MEK inhibitors have been investigated, though resistance develops through various mechanisms, including upregulation of growth factor receptor-bound protein 7 (GRB7) as identified through CRISPR-Cas9 knockout screening (Yu et al. 2022). Innovative approaches being explored include combinations of EGFR and HER2-targeted therapies for specific molecular subtypes of KRAS wild-type disease, and small molecule approaches designed to degrade RAS proteins while repressing EGFR activity (Lee et al. 2018). These diverse therapeutic strategies reflect the complexity of KRAS-driven colorectal cancer and the multiple approaches being pursued to overcome treatment resistance.

Despite significant advances in understanding the role of KRAS mutations in colorectal cancer, several important challenges remain unresolved. The heterogeneity of KRAS mutations within tumors and across metastatic sites complicates treatment decisions and may contribute to mixed responses observed in some patients. The development of effective therapeutic strategies for KRAS-mutant colorectal cancer also faces significant hurdles. Unlike KRAS G12C mutations, which have proven amenable to direct targeting, other common KRAS mutations such as G12D and G13D have been more challenging to address with small molecule inhibitors. Additionally, the rapid development of resistance to targeted therapies through various mechanisms necessitates combination approaches that simultaneously target multiple pathways to achieve durable responses.

1.4 Precision oncology programs: revolutionizing cancer treatment through personalized medicine

Precision oncology represents a paradigm shift in cancer management, leveraging comprehensive molecular analyses to guide individualized treatment decisions. This approach has evolved from broad population-based therapies to highly personalized interventions targeting specific molecular alterations in each patient's tumor. The integration of advanced sequencing technologies, sophisticated bioinformatic pipelines, and multidisciplinary molecular tumor boards has facilitated this transition toward truly personalized cancer care. As precision oncology continues to transform clinical practice, numerous programs worldwide are implementing systematic approaches to molecular profiling and treatment selection.

The NCT-MASTER (Molecularly Aided Stratification for Tumor Eradication Research) program represents a pioneering effort in precision oncology, establishing a clinically applicable platform for prospective, biology-driven stratification of younger adults with advanced-stage cancer across all histologies and patients with rare tumors. Developed at the National Center for Tumor Diseases (NCT) Heidelberg and the German Cancer Consortium (DKTK), this comprehensive program employs a standardized workflow encompassing patient selection and consent, sample processing, whole-exome/genome and RNA sequencing, rigorous bioinformatic analysis, validation of potentially actionable findings, and data evaluation by a dedicated molecular tumor board (MTB) (Horak et al. 2017). This multidimensional approach enables the categorization of patients into different intervention baskets and formulation of evidence-based recommendations for clinical management, addressing a critical need for systematic molecular stratification in challenging cancer cases.

The NCT-MASTER program leverages comprehensive molecular profiling (whole-genome/exome sequencing, RNA sequencing, methylome analysis) to guide precision oncology. However, the sheer complexity and heterogeneity of the resulting data present significant

challenges for interpretation and clinical decision-making. Multilayered tumor profiles are difficult to analyze, and there is limited guidance on integrating these diverse data types into routine clinical practice (Mock et al. 2023). The NCT-MASTER program's molecular tumor board is able to recommend evidence-based, often experimental, therapies for more than 80% of patients with advanced rare cancers or early-onset. Despite the high rate of recommendations, less than one-third of patients (31.8%) actually receive the recommended therapies (Horak et al. 2021). Similar numbers have been reported from other precision oncology programs, such as 17% implementation rate in the CCCMunich^{LMU} program (Heinrich et al. 2023) and 31.7% at the Comprehensive Cancer Center Erlangen (Tögel et al. 2023). Barriers that prevent the administration of recommended drugs include lack of access to experimental or off-label therapies, reimbursement issues, and patient-specific factors such as disease progression or deteriorating patient condition. Implementation of off-label drug recommendations is especially difficult for common cancer types, like CRC, because many lines of standard therapy exist. For those patients who do receive the recommended therapies, there is a significant improvement in disease with a progression-free survival ratio compared to prior treatment >1.3 in 35.7% of patients (Horak et al. 2021). This suggests that while implementation is limited, the clinical benefit for those treated is substantial.

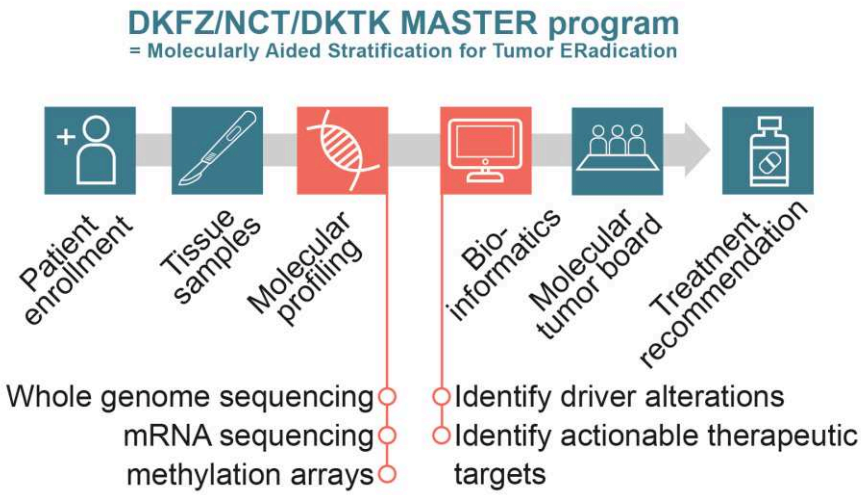


Figure 2: Schematic illustration of NCT-MASTER workflow.
Relatively young patients presenting with advanced-stage cancer for which standard therapy options have been exhausted can be enrolled in the DKFZ/NCT/DKTK MASTER program. Following surgery, molecular analyses of the tumor material are carried out and bioinformatically analyzed. Based on molecular markers, the molecular tumor board gives treatment recommendations. Illustration modified from https://www.nct-heidelberg.de/forschung/molecular-stratification/master.html.

Due to the epigenetic state and the regulation by transcription factors, gene expression is cell type- and context-specific. Therefore, sequencing information about mutations and other genetic alterations offer only limited information on the actual functional state of a cell. While transcriptomic measurements add the knowledge of which genes are expressed in a given context,

due to additional factors, such as mRNA stability, protein turnover and post-transcriptional modifications (Vogel and Marcotte 2012), the gene expression level is often only poorly correlated with the protein level (Sousa Abreu et al. 2009). Additionally, the mere expression level of a protein does not necessarily reflect its activity, which can be regulated by protein-protein interactions and post-translational modifications (PTMs), such as phosphorylation (Ardito et al. 2017). Thus, only proteomics and phosphoproteomics measurements offer direct information about protein abundance and activity, respectively.

1.5 Mass spectrometry-based proteomics

1.5.1 Principle of MS-based proteomics measurement

Different proteomics methods are available, including immunoblotting, protein arrays, enzyme-linked immunosorbent assay (ELISA) and mass spectrometry. While the first three techniques are based on antibody-protein interactions, mass spectrometry-based proteomics offers the unique possibility for unbiased protein detection and quantification, independent of antibody-specificity and -availability.

MS-based proteomics is based on the ability of the mass spectrometer to accurately determine the mass-to-charge ratio (m/z) of peptides, defined as the ratio of a peptide's mass (in Da) and the number of positive charges it carries. Peptides are generated from cellular protein extracts by protease digestion. The most commonly used protease is Trypsin which cleaves proteins C-terminally of lysine (K) or arginine (R). To reduce the complexity of simultaneously measured peptides, liquid chromatography (LC) is performed to separate the peptides according to their hydrophobicity. To this end, peptides are loaded onto a column filled with C18 material and are subsequently eluted with a gradient of acetonitrile. Hydrophilic peptides elute early, while hydrophobic peptides are retained on the column for a longer time. The LC is coupled to the MS system and at the interphase the peptides are protonated by electrospray ionization (ESI). During ESI, peptides that have been dissolved in a slightly acidic buffer are sprayed towards the MS in small droplets. It is hypothesized that during the flight, the solvent evaporates while the protons are retained by the peptides, which are thus positively charged and transferred to the gaseous phase (Bruins 1998). Each peptide can receive one proton each at its N-terminus and at basic amino acids (i.e., arginine, lysine and histidine). Due to the tryptic digestion, the C-terminus of the peptides is characterized by the presence of either lysine or arginine, both of which are basic amino acids, ensuring that each peptide has at least two protonation sites. Ionized peptides are called precursors and enter the MS where they are guided in the vacuum by strong electromagnetic fields.

Different types of mass spectrometers are available and the measurements are based on different mass filters and analyzers or combinations thereof. Since the experiments in this thesis were performed on a Thermo Orbitrap Exploris 480 hybrid mass spectrometer, the following

explanations will focus on this instrument. The most important parts of this mass spectrometer include the quadrupole and the orbitrap. The quadrupole acts as a mass filter and allows for isolation of molecules with specific m/z values within a certain range. This range can be as narrow as 0.4 m/z. The working mechanism of the quadrupole is based on direct and (radio frequency) alternating electric currents that guide the precursors in wave-like movements. Depending on the voltage of the alternating current, different precursor species can stably travel through the quadrupole and are thus isolated (Figure 3). Since only ions can be guided by electromagnetic fields, mass spectrometric analysis can be applied exclusively to charged molecules.

The orbitrap acts as a mass analyzer for the precursors that passed the mass filter. Ions rotate around the spindle-like electrode of the orbitrap and the frequencies of these motions are converted to a mass spectrum by Fourier transformation. Orbitrap mass analyzers allow for highly accurate (mass error < 5 ppm) and highly resolved (up to 480,000 at 200 m/z) determination of peptide masses (Figure 3). The resolution is calculated as the peptide's m/z divided by its peak width at half maximum intensity. For instance, when a precursor mass-to-charge ratio was determined to be 500 m/z and its peak width at half maximum intensity was 0.01 m/z, it was acquired with a resolution of 50,000, commonly abbreviated as 50K. The longer a precursor's rotation in the Orbitrap is acquired, the higher the resolution of the m/z determination.

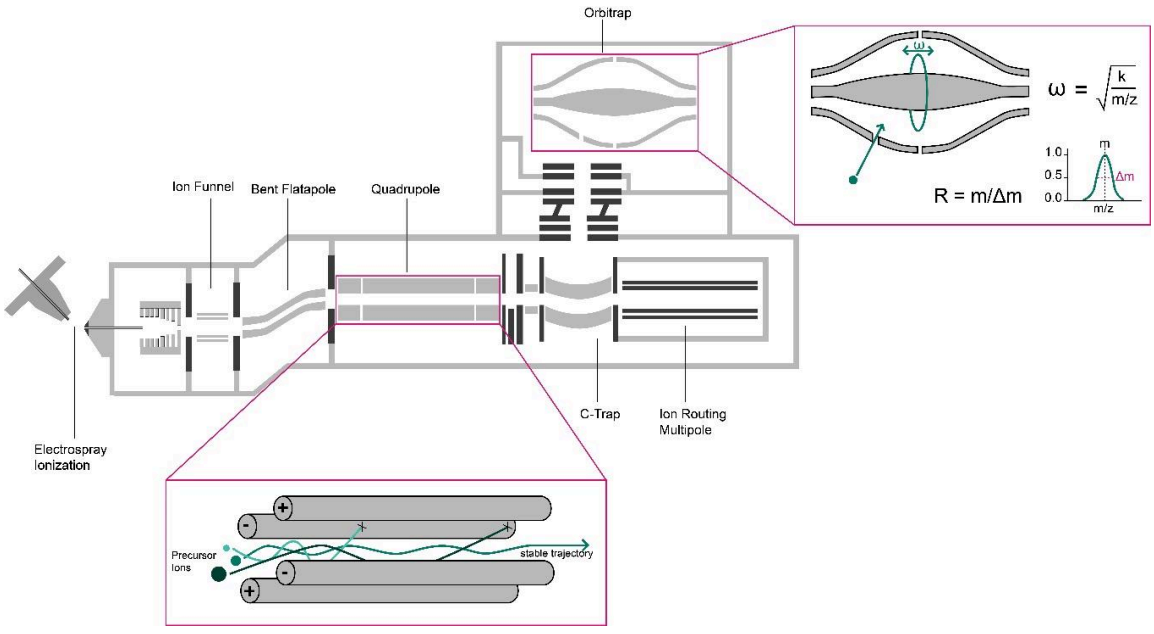


Figure 3 Principle of mass spectrometry-based protein analysis.
Chromatographically separated peptides are protonated by electrospray ionization (ESI) and enter the mass spectrometer guided by strong electromagnetic fields. The ion beam is focused in the ion funnel and uncharged molecules are filtered out by the bent flatapole. The quadrupole acts as a mass filter and isolates selected precursors or precursor ranges. The mass-to-charge ratio of isolated precursors is determined by the Orbitrap via the frequency of rotation along its spindle-shaped electrode. In the Ion Routing Multipole, precursor ions are collected and fragmented by gas collision. Fragments are then injected into the Orbitrap via the C-Trap. Figure adapted from Thermo Fisher Scientific.

1.5.2 Technical advancements of recent years

Over the past decade, mass spectrometry (MS) has undergone remarkable technical advances that are accelerating its integration into clinical applications. The advent of data-independent acquisition (DIA) has revolutionized proteomic profiling by enabling comprehensive, reproducible, and high-throughput quantification of thousands of proteins across complex clinical samples, providing robust diagnostic and prognostic signatures even in challenging scenarios such as cancer of unknown primary origin (Poulos et al. 2020; Boys et al. 2023; Fröhlich et al. 2024). Simultaneously, new generations of MS instruments have achieved unprecedented speed and sensitivity, allowing for rapid analysis with minimal sample input and enhanced detection of low-abundance biomarkers, thus supporting high-throughput workflows essential for routine clinical use (Bekker-Jensen et al. 2020; Heil et al. 2023; Meier et al. 2018). These improvements are complemented by the development of automated and user-friendly workflows, including streamlined sample preparation (Müller et al. 2020) and intuitive raw data analysis software platforms (Bekker-Jensen et al. 2019), all of which reduce operator variability and make MS-based assays more accessible for clinical laboratories. Collectively, these innovations are transforming MS from a research tool into a powerful, scalable technology for clinical diagnostics and personalized medicine.

1.5.3 The advantages of phosphoproteomics

Phosphoproteomics holds particular promise for clinical applications because it enables the direct and comprehensive analysis of protein phosphorylation, a key regulatory mechanism in virtually all cellular signaling pathways (Schmelzle and White 2006). Aberrant phosphorylation events are central to the development and progression of many diseases, especially cancer, where they drive oncogenic signaling, therapy resistance, and tumor heterogeneity (Ardito et al. 2017). By profiling the phosphoproteome in patient samples, clinicians and researchers can gain real-time insights into dynamic cellular processes, identify dysregulated pathways, and discover actionable biomarkers or therapeutic targets that are not evident from genomic or total proteome data alone (Harsha and Pandey 2010). Advances in sensitive and automated phosphopeptide enrichment methods now allow robust phosphoproteomic analysis from minimal clinical material, making it feasible to study disease-relevant signaling events even in small or precious patient samples (Murillo et al. 2018). Multiple studies have demonstrated the potential of phosphoproteomics to guide combination therapies to overcome resistance (Beekhof et al. 2023; Yip et al. 2024) or to reveal predictive markers of drug response (Lee et al. 2024; Yip et al. 2024).

Collectively, these capabilities position phosphoproteomics as a powerful tool for precision medicine, enabling the identification of clinically relevant biomarkers and informing personalized treatment strategies.

1.6 Patient-derived organoids for drug screenings

Patient-derived organoids (PDOs) have become a transformative model system in colorectal cancer (CRC) research, offering a highly relevant platform for both basic and translational studies. To establish CRC PDOs, tumor tissue surgically obtained from patients is then mechanically and enzymatically dissociated into single cells or small clusters, which are embedded in a three-dimensional extracellular matrix such as Matrigel and cultured in media supplemented with growth factors that support the self-organization and expansion of tumor cells into organoid structures (Sato et al. 2009). Organoids derived from colon cancer patients have been demonstrated to recapitulate the genetic landscape of the original tumor – including mutations, copy number alterations, and MSI status - (Codrich et al. 2021; van de Wetering et al. 2015; Weeber et al. 2015). One major challenge, however, is their incomplete recapitulation of the tumor microenvironment: PDOs typically lack the full spectrum of stromal cells, immune components, and vasculature present in vivo, which are critical for accurately modeling tumor behavior, angiogenesis, and drug responses. This limitation can affect the predictive value of PDOs, especially for therapies that target the tumor microenvironment or rely on immune interactions. Additionally, PDOs may not fully capture the intra-tumor heterogeneity of the original tumor, particularly if only a subset of cells successfully grows out during organoid establishment, potentially leading to clonal selection and loss of rare subpopulations (Qu et al. 2024). In the clinical context, CRC PDOs are increasingly used for high-throughput drug screening to predict individual patient responses to chemotherapy, targeted agents, and combination regimens, thereby informing personalized treatment strategies (He et al. 2023; Narasimhan et al. 2020; Smabers et al. 2024). Additionally, they serve as valuable tools to validate clinical finding and as living biobanks of patient tumors, PDOs are poised to play a central role in precision oncology for CRC.

1.6 Aims of this thesis

The overarching aim of this thesis was to explore the clinical potential of proteomic and phosphoproteomic profiling in the context of precision oncology for colorectal cancer (CRC). Specifically, the project was designed around three key objectives:

1. To establish a robust and scalable proteomics workflow suitable for clinical CRC tissue samples.

The goal was to develop and optimize protocols for protein extraction, automated peptide digestion, and automated phosphopeptide enrichment. These workflows were intended to be compatible with limited and variable-quality tissue samples typically obtained from clinical biopsies. Data-independent acquisition (DIA) mass spectrometry should be employed to ensure comprehensive and reproducible proteome and phosphoproteome coverage. Emphasis should be placed on quality control procedures and reproducibility across experimental batches. Finally, a cohort of CRC patient samples should be analyzed applying the optimized protocols to generate the data basis for subsequent analysis of total and phosphoproteomic data.

2. To integrate proteomic and phosphoproteomic data with existing genomic and transcriptomic datasets.

The aim was to design a computational framework capable of aligning protein-level measurements with previously acquired molecular data. This integration should involve cross-comparison of DNA copy number, RNA expression, and protein abundance, as well as the inference of kinase activity profiles using phosphosite information. This multi-layered approach was intended to identify concordant and discordant relationships between different molecular modalities.

3. To derive functional, clinically relevant insights in the form of an integrative scoring system.

The final objective was to translate the multi-omics data into actionable outputs that can support therapeutic decision-making. For this, a basket-based scoring system should be developed to summarize the activity of signaling pathways and potential drug targets in individual patient samples. Utilizing patient-derived organoids, this score should then be evaluated for stratifying treatment options and identifying promising therapeutic interventions.

Through these three aims of my thesis, I sought to lay a methodological and conceptual foundation for incorporating proteomics into the clinical precision oncology pipeline.

2. Results

2.1 Technical optimization of proteomic measurements

2.1.1 Automated protein clean-up and digestion using SP3 on the Bravo liquid handling robot

Automated sample preparation is a crucial prerequisite for high-throughput standardized processing of clinical samples. Single-pot solid-phase-enhanced sample preparation (SP3) had been previously established on the Agilent Bravo liquid handling robot and significantly reduces hands-on time and variability, while improving longitudinal reproducibility (Müller et al. 2020). This automated SP3 (autoSP3) workflow was further optimized by Dr. Barbara Helm and Yannik Dieter (AG Klingmüller) and with their support implemented at the Proteomics Core Facility of DKFZ. To test this automated workflow, I compared its performance to the manual SP3 method as published by Hughes et al. (2019). When using 20 µg protein from a cell line lysate, both sample preparation methods deliver comparable MS1 intensities (Figure 4A), however the manual method outperforms autoSP3 regarding the number of identified protein groups (Figure 4B). SP3 is known to perform better with higher protein concentrations during the bead binding step (Johnston et al. 2022). When using a higher protein starting material of 55 µg, the amount needed if subsequent phosphopeptide-enrichment is desired, autoSP3 performed equally well as manual SP3 regarding proteome coverage (Figure 4B). Reduced manual sample handling time could potentially reduce the risk of organic or inorganic contaminants entering the sample. Using the R package HowDirty (Gomez-Zepeda et al. 2024), I observed similar levels of inorganic contaminations with manual and automated SP3 (Figure 4C). Polyethylene glycol (PEG) was the highest abundant inorganic contamination. PEGs have a high ionization efficiency and are present in most laboratories since these polymers are included in skin creams and moisturizing products, detergents, pipette tips, and chemical wipes that contain different PEGs (Meston 2022). Using the MaxQuant (Cox and Mann 2008) contaminants.fasta database for common contaminants (provided with MaxQuant version 1.6.3.3), I identified similar levels of organic contamination with both manual and auto SP3, consisting mainly of trypsin, albumin, and keratins (Figure 4D). Porcine trypsin had been added in high amounts for protein digestion, while the bovine albumin most likely stemmed from the fetal bovine serum (FBS) used in cell culture. Overall, autoSP3 delivered results comparable with the manual workflow, but is scalable, standardized and greatly reduces hands-on sample preparation time.

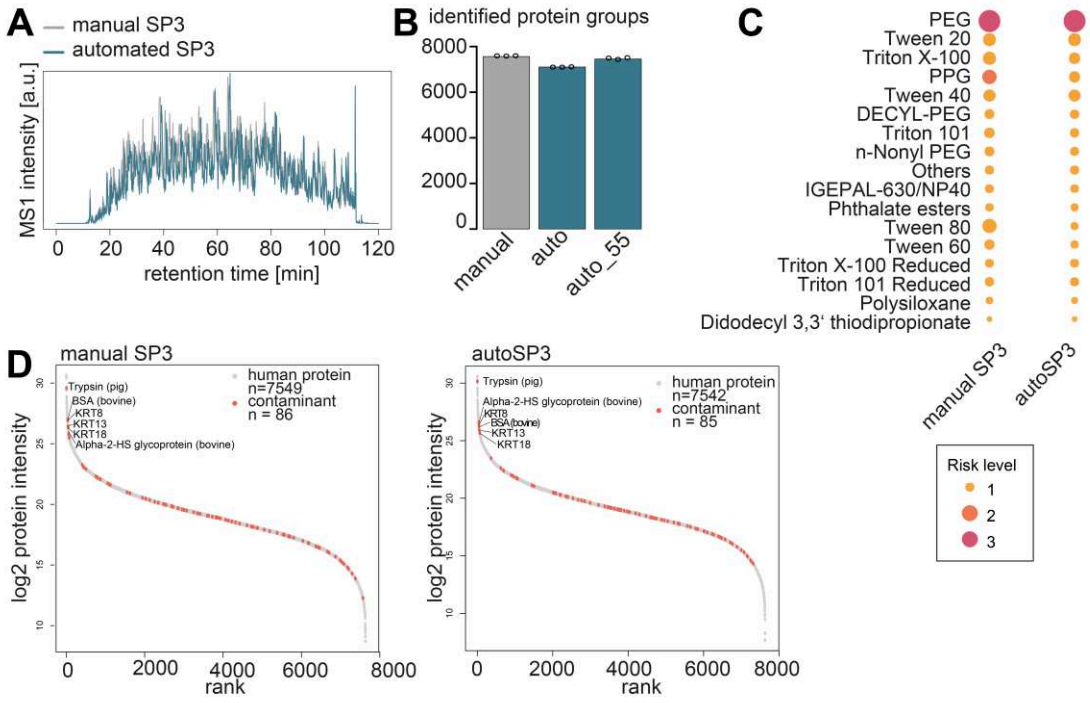


Figure 4: Optimization of autoSP3
MCF7 cells were lysed in RIPA buffer (Thermo Fisher Scientific), treated with DNase and Benzonase, and protein concentration was determined using Pierce BCA assay. Manual and automated SP3 was performed with 20 µg protein input in three technical replicates. Peptides from 1.5 µg protein were injected for LC-MS/MS measurement for 120 minutes and mass spectra were acquired with the DIA method (Table 6). Raw data were searched against a human protein database with isoforms (79,052 entries, downloaded from Uniprot 15.03.2022) with Spectronaut (version 17) in directDIA mode (Table 9). "Method development" setting was enabled to search the files independently from each other. (A) MS1 intensities over retention time. (B) Automated SP3 also additionally performed with 55 µg input (as needed for phosphopeptide-enrichment). Shown are numbers of identified and quantified protein groups. (C) Extent of inorganic contamination based on HowDirty package (Gomez-Zepeda et al. 2024). (D) DIA search with MaxQuant_contaminations.fasta file (provided in version 1.6.3.3), most abundant contaminations highlighted.

While autoSP3 works well with high amounts of protein input, such as 20 µg or 55 µg (Figure 4), processing low input samples might become necessary, for example when working with limited clinical samples. I assessed the sensitivity of autoSP3 by using different amounts of protein input (0.1 µg – 20 µg) from the cell line lysate and injecting the same theoretical amounts of peptides for measurement (0.5 µg or 0.1 µg). The samples were pre-diluted in RIPA lysis buffer so that the protein concentration differed, while the volume was kept constant.

Between 20 µg and 0.5 µg protein, I observed a continuous decline in the numbers of identified protein groups despite injecting peptides from the same amount of protein (Figure 5A), possibly due to insufficient protein precipitation to the magnetic beads at low concentrations (Johnston et al. 2022). From 0.5 µg protein lysate, I was still able to reproducibly identify 4,433 protein groups

– 68.4% of the number of protein groups identified from 20 µg lysate. According to the original publication, autoSP3 reproducibly quantifies 500–1,000 proteins from samples as small as 100–1,000 cells (in a range below 0.1 µg protein) while maintaining high precision and sensitivity (Müller et al. 2020). However, when processing very low sample amounts of 0.2 µg and 0.1 µg, I observed a drastic drop in proteome coverage with only 483 (±404) and 140 (±14) identified protein groups (Figure 5B). Processing of very low sample amounts (< 500 ng) is therefore not yet possible in the current set up and would need to be further optimized.

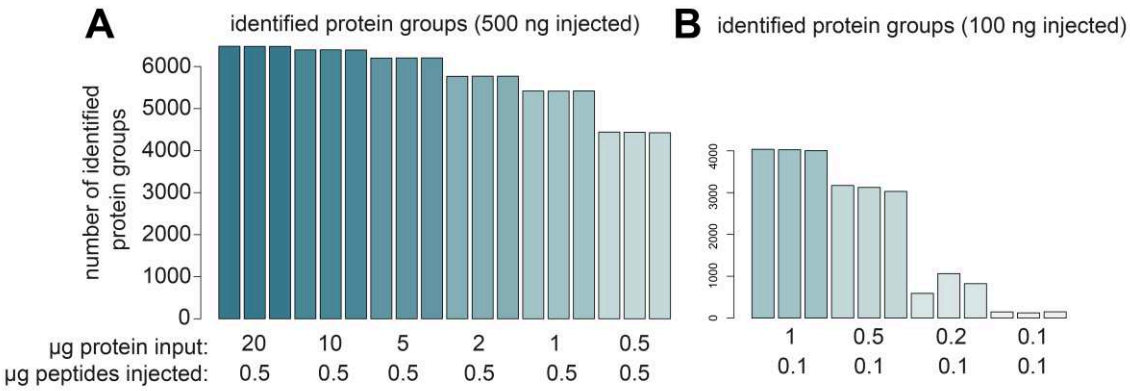


Figure 5: Dilution curve of starting material (autoSP3)
MCF7 cells were lysed in RIPA buffer (Thermo Fisher Scientific), treated with DNase and Benzonase, and protein concentration was determined using Pierce BCA assay. Automated SP3 was performed with 20,10,5,2,1,0.5,0.2,0.1 ug protein input in three technical replicates. The amount of trypsin was adjusted to always be 1:25. Peptides from 500 ng (A) or 100 ng (B) protein were injected for LC-MS/MS measurement for 120 minutes and mass spectra were acquired with the DIA method (Table 6). Raw data were searched against a human protein database with isoforms (79,052 entries, downloaded from Uniprot 15.03.2022) with Spectronaut (version 17) in directDIA mode (Table 9). "Method development" setting was enabled to search the files independently from each other. Shown are numbers of identified and quantified protein groups.

In conclusion, the automation of SP3 works and delivered very similar results to manual SP3, but required considerably less manual work and allows for higher throughput. Since I was not only interested in protein abundance, but also in protein activity, phosphopeptide-enrichment needed to be automated as well.

2.1.2 Optimization of automated phospho-enrichment protocol

Phosphorylated peptides, being inherently low-abundant species in cellular systems (Wu et al. 2011), traditionally require enrichment from large amounts of protein (0.5 – 5 mg) with 1 mg being the current standard for robust phosphoproteome coverage in most MS workflows (Ruprecht et al. 2015; Potel et al. 2018; Ruprecht et al. 2017).

Manual phosphopeptide-enrichment consists of multiple and labor-intensive steps. These include protein precipitation, in-solution digestion, Fe(III)-IMAC (immobilized metal affinity chromatography) column stripping and loading, peptide loading and elution of phosphorylated fraction, and repeated sample clean-up steps (see materials & methods 4.4.1). In total, this manual protocol requires four full working days to process a batch of maximally 20 samples. Its scalability limitations prompted the evaluation of an automated workflow in 96-well format established by Agilent on the Bravo liquid handling robot. This automated workflow has the additional advantage of using significantly reduced input (50 µg protein). I compared six technical replicates of phosphopeptide-enrichment from a RIPA lysate of a breast cancer cell line with the manual (500 µg) and the automated (50 µg) workflow, following the default protocol provided by Agilent (2016). It should be noted that manual protein digestion and phosphopeptide-enrichment was performed and measured a few months earlier because the manual setup was disassembled once the Bravo robot has been purchased. All samples were measured in DIA mode for 120 minutes, raw data search was performed using Spectronaut in direct-DIA mode, and phosphopeptides were site-collapsed to the gene level using Perseus with the plugin PeptideCollapse. A phosphosite localization probability cut-off of 0.95 was applied.

Though MS1 intensities were significantly lower in automated runs (Figure 6A) – consistent with 10-fold reduced input – reproducibility of phosphopeptide quantification across the six technical replicates proved comparable (manual: 3.4% CV; automated: 3.1% CV, Figure 6B). The manual workflow (500 µg) yielded 20,962 ± 366 identified, quantified and localized phosphosites versus 15,080 ± 201 in the automated protocol (50 µg input), with phosphosite localization filtered at >95% confidence (Figure 6C). Notably, 65.3% of phosphosites were consistently identified across all replicates of the automated workflow versus 56.6% in manual workflows, suggesting improved technical reproducibility (Figure 6D). The dramatic reduction in hands-on time (4 days → 3 hours per batch) and cost (60€ → 15€ per sample,Table 4) positions automation as essential for high-throughput clinical studies, particularly given the 4–5 week manual processing time for 96 samples versus 2 days when automated. Cost savings stem primarily from reduced trypsin and LysC usage due to the lower amount of protein required as input. Furthermore, the decrease in hands-on time leads to an enormous reduction of labor cost. The main cost factor of the automated protocol are the Fe(III) cartridges for the Bravo robot. Since it is possible to wash and re-load the cartridges with iron for multiple uses (SOPs: /LINCS/Overview Information 2025), this cost could potentially be further reduced.

In conclusion, although the manual workflow results in higher numbers of identified phosphosites, likely due to the ten-fold higher amount of input material, the fact that it is not scalable, more expensive, and requires at least ten times more manual work excludes it from being used in a high-throughput clinical proteomics setting.

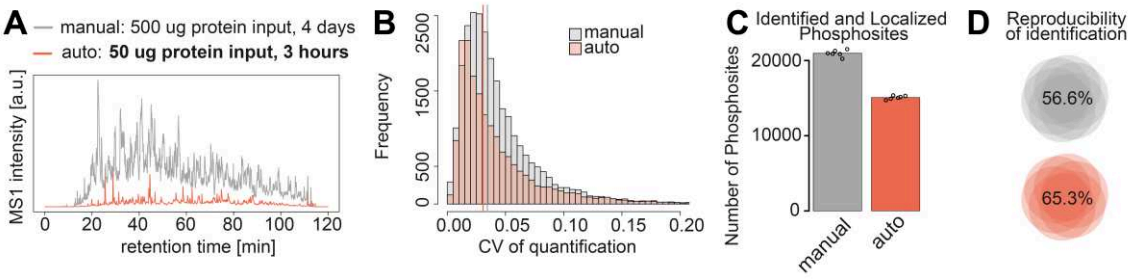


Figure 6: Comparison of manual and automated phosphopeptide-enrichment protocols
Manual phosphopeptide enrichment was performed with peptides from 500 µg protein as starting material. Automated phosphopeptide enrichment was performed on the Agilent Bravo liquid handling robot with peptides from 50 µg autoSP3 digested proteins. Phospho-enriched fractions were measured for 120 minutes with the phospho-DIA method. Raw data were searched against a human protein database with isoforms (79,052 entries, downloaded from Uniprot 15.03.2022) using Spectronaut (version 16) in directDIA mode. Identified phosphopeptides were site-collapsed to the gene level with Perseus (v 1.6.2.3) and plugin PeptideCollapse (v 1.4.3, cut-off 0.95). n = 6 technical replicates. (A) MS1 intensities of phosphopeptide-enriched fraction over retention time, estimated manual labor time is indicated. (B) Distribution of coefficient of phosphopeptide intensity variation over the six technical replicates. (C) Number of identified, quantified, and localized (> 0.95) phosphosites. (D) Proportion of phosphosites identified in all six replicates.

To reach higher numbers of identified phosphosites, I next tested whether the recommended 50 ug protein was really the capacity limit of the Fe(III)-cartridges or if the protein amount could be scaled up. To test this, I used a pool of autoSP3-digested peptides from MCF10A and MDA-MB-231 cell line lysates and performed automated phosphopeptide-enrichment with different amounts of peptides, ranging from 10 µg to 400 µg. Additionally, I tested if the measurement time could be reduced from 120 minutes to 60 minutes to further increase the throughput of the workflow. Enrichment of peptides from 10 µg to 100 µg protein resulted in a linear increase of phosphoproteome coverage (Figure 7A). Using as little as 10 µg protein as input and measuring for 60 minutes, allowed for the identification of around 5,000 phosphosites, highlighting the potential for also processing smaller biopsies with a fast turn-around. Beyond 100 µg, the number of identified phosphosites actually decreased, suggesting that 100 µg was the capacity limit of the cartridges (Figure 7A). Decreasing the measurement time from 120 minutes to 60 minutes, led to a 25% lower phosphoproteome coverage in the case of 50 µg input (120 minutes: 11,784 phosphosites, 60 minutes: 9,444 phosphosites). As long as high sample throughput was not critical, I therefore used the 120 minutes measurement method. With this longer measurement method, the difference between starting with 50 µg or 100 µg of protein was minimal, with only about 5% more sites identified from 100 µg.

Looking at the numbers of phosphosites per peptide, meaning single-, double-, or triple-phosphorylated - further confirmed that the capacity of the cartridges was exceeded with 200 ug and 400 ug (Figure 7B). When the capacity is exceeded, high affinity peptides (> 1 phosphate

group) will be enriched relative to single-phosphorylated peptides. At 50 µg and 100 µg only 3% of identified phosphopeptides were triple-phosphorylated, 22% were double-phosphorylated, and the vast majority of 75% carried only a single phosphate group. When using 200 µg or 400 µg protein as input, 40-44% of identified phosphopeptides were double-phosphorylated and 10% even carried three phosphate groups.

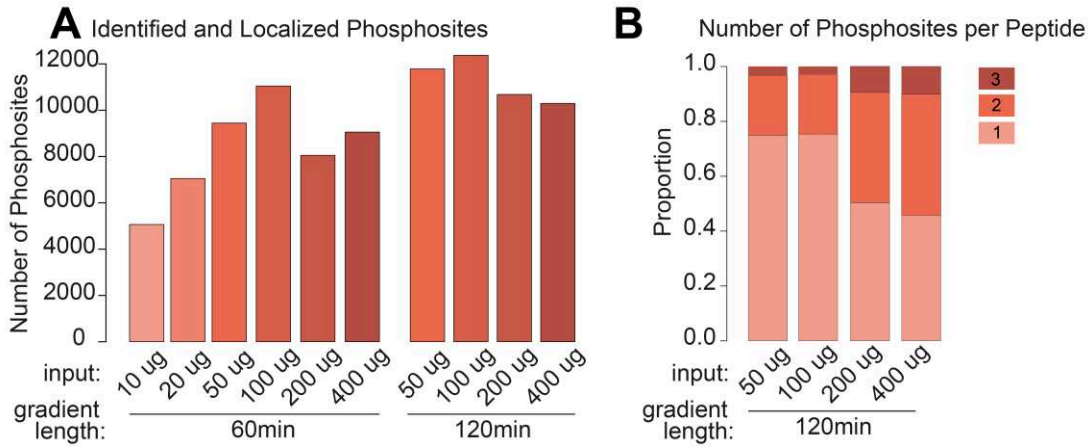


Figure 7: Sensitivity and capacity of automated phospho-enrichment protocol
Pool of digested 50 µg MCF10A and MDA-231 cell lines samples with autoSP3, pooled peptides and enriched different amounts of starting material. Measured for 60 min or 120 min with phospho-DIA method. Raw data were searched against a human protein database with isoforms (79,052 entries, downloaded from Uniprot 15.03.2022) using Spectronaut (version 16) in directDIA mode. Identified phosphopeptides were site-collapsed to the gene level with Perseus (v 1.6.2.3) and plugin PeptideCollapse (v 1.4.3, cut-off 0.95). n = 1 technical replicate. (A) Number of identified, quantified, and localized phosphosites from enrichment with different amounts of starting material. (B) Proportion of single-, double- or triple-phosphorylated peptides.

Since there was no large gain when using the cartridges at full capacity and clinical material might be limited, I decided to use 50 ug protein as input, as recommended by the Bravo manufacturer. Next, I tested different measurement strategies for optimal phosphoproteome coverage. I compared injecting the entire enriched fraction for LC-MS/MS measurement, or retaining half of the enriched sample as a back-up, or injecting two times half of the sample and measuring it with two different FAIMS setting (CVs) to achieve a sort of gas phase fractionation (Klaeger et al. 2021; Faivre et al. 2023).

Injecting the entire sample and measuring for 120 minutes delivered the highest numbers of identified and localized phosphosites and lowest quantification variance (Figure 8).

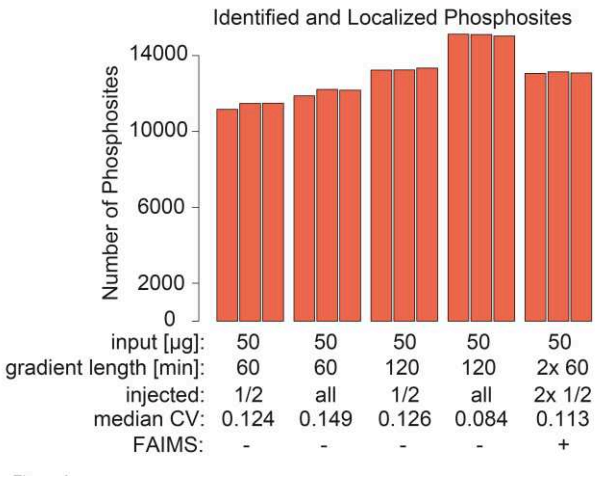


Figure 8: Optimization of phosphoproteome measurement strategy
Pool of digested 50 µg MCF10A and MDA-231 cell lines samples with autoSP3, pooled peptides and enriched phosphopeptides of peptides from 50 µg protein. Measured for 60 min or 120 min with phospho-DIA method. Either injected the entire phospho-enriched fraction, or retained half as a backup. FAIMS was used with CV settings of -45 and -70. Raw data were searched against a human protein database with isoforms (79,052 entries, downloaeed from Uniprot 15.03.2022) using Spectronaut (version 16) in directDIA mode. Identified phosphopeptides were site-collapsed to the gene level with Perseus (v 1.6.2.3) and plugin PeptideCollapse (v 1.4.3, cut-off 0.95). n = 3 technical replicates.

In conclusion, the automated phosphopeptide-enrichment workflow performed well, even with relatively low amounts of protein input. Since working with clinical samples, I wanted to make sure that what I was identifying was actually true. I was not only interested in measuring as many proteins or phosphopeptides as possible, but in a reliable method with as few as possible false identifications.

2.2.1 Optimization of Spectronaut search parameters reduces false discovery rate

Controlling the false discovery rate (FDR) in shotgun proteomics is critical to ensure the reliability and reproducibility of peptide and protein identifications. Shotgun proteomics involves analyzing complex mixtures of peptides, generating thousands of mass spectra that are matched to theoretical sequences of a protein database, e.g. the entire human proteome. FDR estimation methods, such as the target-decoy approach that generates a decoy database by shuffling the protein database**,** quantify these false matches by comparing target and decoy database hits (Levitsky et al. 2017). Shotgun proteomics generates vast datasets, where random matches between spectra and decoy sequences are inevitable and a threshold must be set to determine which rate of false discoveries are accepted in a given experiment. However, the theoretical FDR calculated based on the target-decoy approach is only a statistical estimation and the true proportion of falsely matched peptides is unknown.

Together with Martin Schneider, I designed an experiment with proteome mixtures of different species (Figure 9A). This allowed me to evaluate whether the software identified proteins that cannot truly be present.

The Spectronaut search engine (v. 17) includes multiple parameter thresholds for FDR control, such as precursor q-value (default: 0.01), precursor posterior error probability (PEP) (default: 0.2), protein q-value (default: 0.05), and protein PEP (default: 0.75). With the default search engine settings, 452 false positive human proteins were identified in a sample that only contained *E. coli* proteins (Figure 9B). Of the overall 7,446 identified human proteins across the experiment, that corresponded to a false discovery proportion of 6%. From the total 1,763 identified *E. coli* proteins across all samples, 285 proteins were identified in a sample containing only human proteins – a false discovery proportion of 16%. None of the identified proteins could be attributed to both species, eliminating peptide sequence homologies as a reason for false identifications.

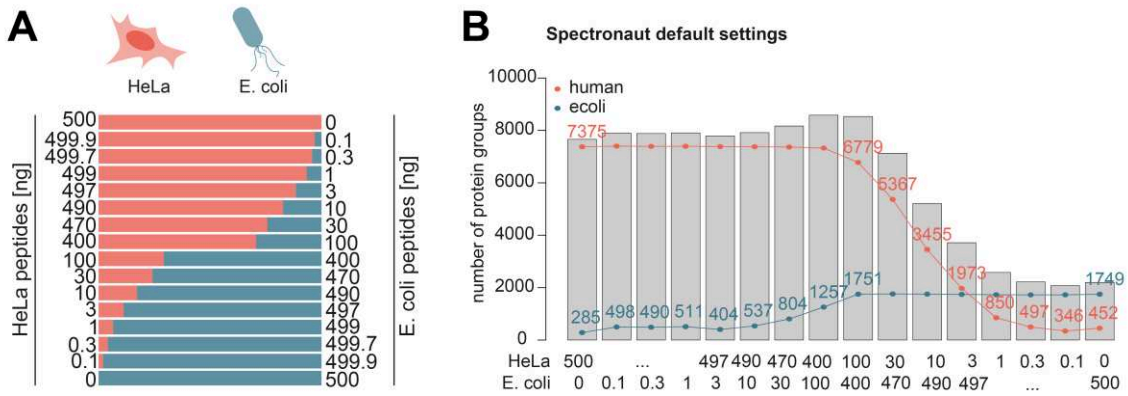


Figure 9: Two-species proteome mixture to control false discovery rate
Commercial HeLa (Thermo Fisher Scientific) and E. coli (Waters) protein digests were mixed in different proportions to a total of 500 ng. Each sample was measured for 120 minutes with the DIA method. Raw data were searched against the human (79,052 entries, downloaded from Uniprot 15.03.2022) and E. coli (5,062 entries, downloaded from Uniprot 15.03.2022) protein databases using Spectronaut (version 17) in directDIA mode with default settings. (A) Illustration of two-species proteome mixtures in different ratios. (B) Identified human and E.coli protein groups in each sample.

To minimize false-positive identifications without substantially compromising the detection of true-positive proteins, I systematically evaluated the impact of varying FDR-related parameter thresholds in Spectronaut. Specifically, I adjusted the precursor PEP, the protein q-value, and the protein PEP - three key parameters that influence the stringency of peptide and protein identification. For precursor PEP, I tested thresholds of 0.2 (default), 0.1, 0.01, and 0.001 (Figure 10A). Similarly, for the protein q-value, I explored values of 0.05 (default), 0.025, 0.01, and 0.001 (Figure 10B). Finally, for protein PEP, I assessed cut-offs of 0.75 (default), 0.25, 0.1, and 0.05 (Figure 10C). Through this optimization, I aimed to identify a combination of parameters that

significantly reduced the number of false protein identifications, namely cross-species identifications in the samples only containing the other species, while preserving the majority of (presumably) true positives. The final optimized combination consisted of:

- Precursor PEP: 0.05
- Protein q-value: 0.01
- Protein PEP: 0.05

This stricter set of FDR search parameters led to a remarkable 80% reduction in false positive protein identifications, while the total number of correctly identified proteins decreased by only 2%, highlighting a favorable trade-off between sensitivity and specificity (Figure 10D). Even with these refined parameters, 104 human proteins were still falsely identified in the sample composed entirely of *E. coli* proteins (Figure 10E). However, this accounts for only 1.4% of the total protein identifications, which is considered an acceptable level of false discovery in proteomics. Given the intrinsic uncertainty of mass spectrometry-based database matching, where imperfect fragmentation spectra and shared peptide motifs can occasionally lead to incorrect assignments, some degree of false positive identification is inevitable. Another strategy to eliminate potential false protein identifications is to filter out proteins that were only identified based on one single peptide ("single-hit"). Since we would assume that matching an acquired mass spectrum by chance to a peptide of a protein that is not actually present in the sample is unlikely to happen more than once to the same protein, this approach might eliminate false-positive identifications. However, when I exclude single-hit proteins in my data set, the (presumably) true proteins were affected to a greater extent than the false-positives. With single-hit proteins removed, 11.2% and 4.5% of true human and *E. coli* proteins were lost, respectively, while it reduced the numbers of false-positive identifications only by 2% and 3.7% for human and *E. coli* proteins, respectively (Figure 10F).

Ultimately, while statistical control methods like the target-decoy approach and software-based FDR parameters provide robust means to filter data, empirical validation, such as species-mixing experiments, remains a valuable tool to assess the real-world performance of these strategies.

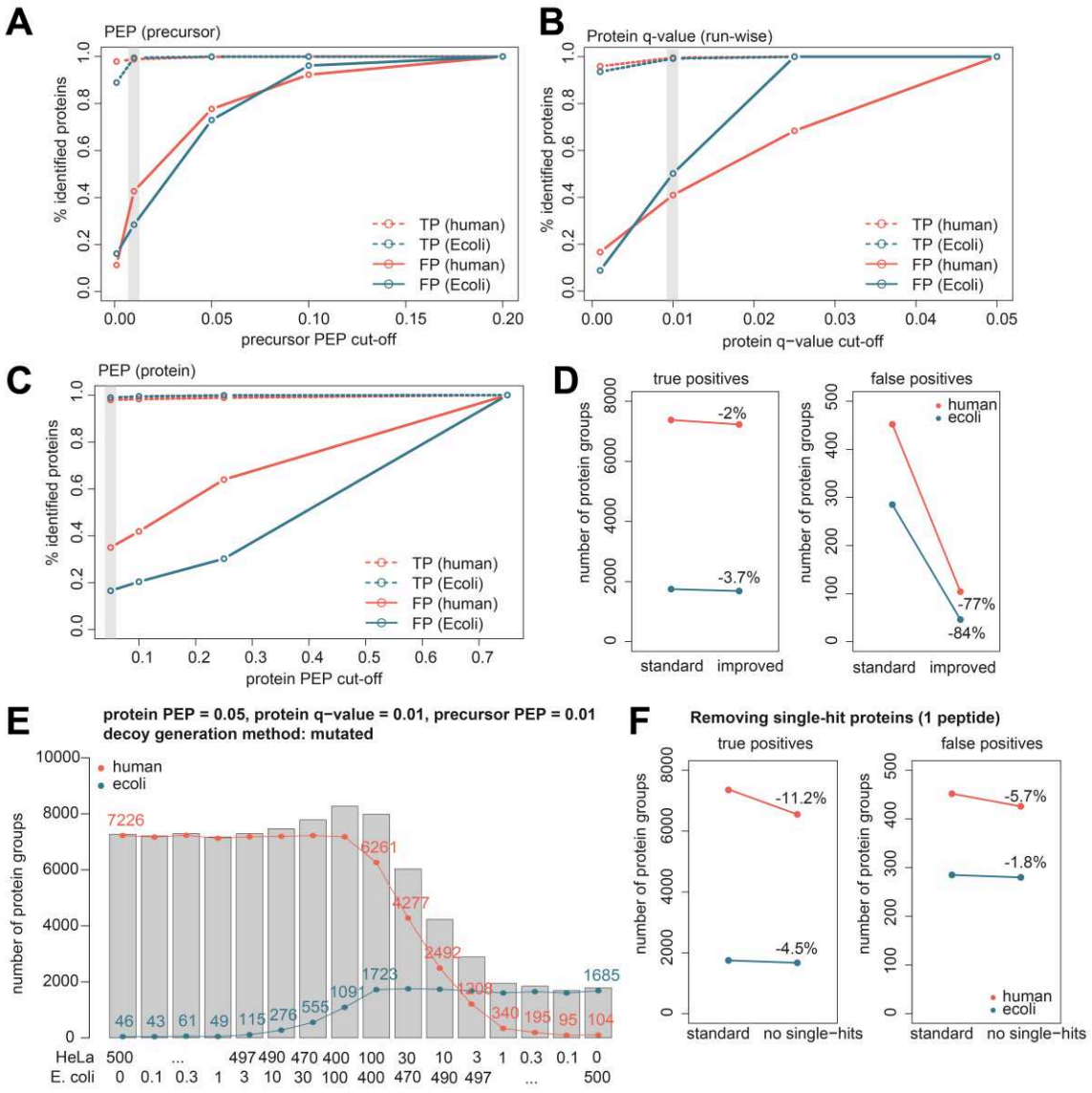


Figure 10: Spectronaut parameter optimization
Samples from Figure 9 analyzed with Spectronaut (version 17) with default settings. posterior error probability (PEP) of precursors (A), Protein q-value (B), and posterior error probability (PEP of proteins (C) were separately set to be increasingly more stringent. Shown are percentages of true positives (number of proteins identified in 100% sample) and false positives (number of proteins identified in 0% sample) compared to default settings. The chosen cut-offs for the final settings are highlighted in grey. (D) Reduction of false-positive and true-positive protein identifications when combining all stricter FDR cut-offs compared to default search settings. (E) Combined analysis of all files with new FDR cut-offs. Numbers of identified human and E. coli protein groups are highlighted. (F) Excluding proteins that were only identified by one peptide ("single hit") compared to default settings.

Especially in clinical settings, where molecular profiling of tumor tissue can directly influence therapy decisions, the demands on proteomics go beyond accurate protein identification. Reliable relative quantification becomes equally critical, particularly when comparing patients with each other using single analytical measurements without repetitions (n=1). In such contexts, even small deviations in quantification accuracy can affect clinical interpretation and decision-making.

2.2.2 Assessing the accuracy of relative protein quantification

To assess the quantification accuracy of my measurement method, I prepared a controlled mixture of proteomes from three different species in two defined ratios (Kuharev et al. 2015). Commercially available tryptic digests of HeLa (human), *E. coli*, and yeast proteins were used to create two distinct samples:

HYE-A: 65% HeLa, 15% *E. coli*, 20% yeast

HYE-B: 65% HeLa, 30% *E. coli*, 5% yeast

This experimental design ensured that the expected relative protein abundances were known *a priori*: human proteins should remain constant between the two samples (1:1), *E. coli* proteins should double in abundance (1:2 ratio in HYE-A vs. HYE-B), and yeast proteins should decrease four-fold (4:1 ratio in HYE-A vs. HYE-B). Each sample was analyzed using DIA, injecting 500 ng of total protein and acquiring data over a 120 minutes LC-MS method. The raw data were searched using Spectronaut (v17), matching spectra against all three species' sequence databases at once. Using the optimized search parameters (precursor PEP 0.05, protein q-value 0.01, protein PEP 0.05), I identified a total of 9,096 protein groups, distributed as follows: 5,970 human, 2,237 yeast, and 888 *E. coli*. One protein, which could be assigned to multiple species based on sequence overlap, was excluded from further analysis. To evaluate quantification performance, I calculated $\log_2$ intensity ratios of each protein between the HYE-A and HYE-B samples. The results demonstrated excellent clustering of proteins around their expected values, confirming high quantification precision. The mean squared error (MSE) - a measure of deviation from the expected ratios - was 0.31 for *E. coli*, 0.06 for yeast, and 0.14 for human proteins (Figure 11A).

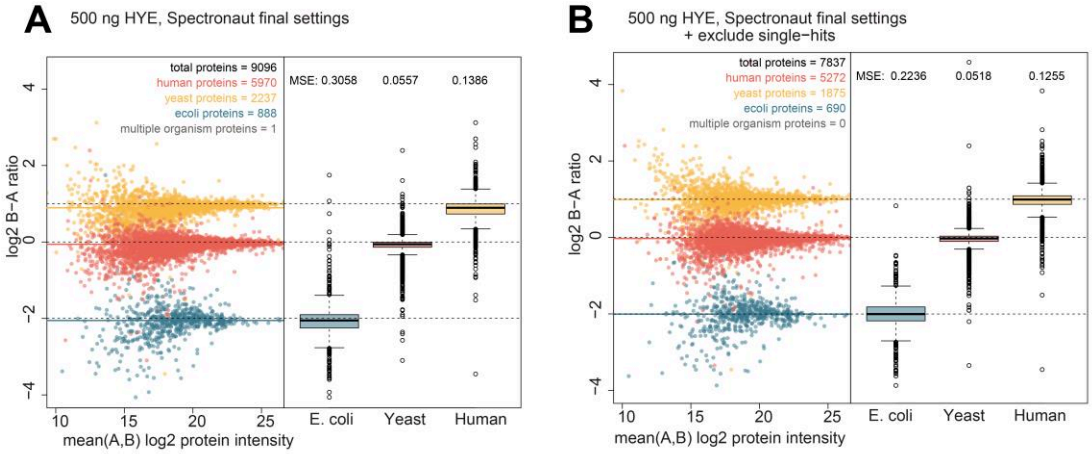


Figure 11: Protein quantification accuracy
Three species proteome mixture was prepared by combining HeLa, E. coli and yeast commercial digests to a total of 500 ng peptides in two different ratios: A = 65% human, 15% yeast, 20% E.coli, B = 65% human, 30% yeast, 5% E. coli. Both HYE-A and HYE-B were measured for 120 minutes with the DIA method. (A) final parameters, all proteins. Raw data were searched against the human (20,594 entries, downloaded from Uniprot 09.02.2023), yeast (6,060 entries, downloaded from Uniprot 06.03.2024), and E. coli (4,403 entries, downloaded from Uniprot 06.03.2024) protein databases without isoforms using Spectronaut (version 18) in directDIA mode with final settings (Table 9). (A) Log2 ratios between HYE-A and HYE-B were calculated for each identified protein and grouped by species. Mean squared errors were calculated by summing all squared deviations from the expected ratio and dividing by the number of identified proteins. (B) Proteins identified based on only one peptide ("single hits") were excluded.

When removing single-hit proteins, I observed a modest improvement in quantification accuracy. The MSE values decreased to 0.22 for *E. coli*, 0.05 for yeast, and 0.13 for human proteins, suggesting that protein quantifications based on more than one peptide are slightly more robust. However, excluding single-hit proteins led to a 14% decrease in overall proteome coverage (Figure 11B). Since the loss in protein numbers was much larger than the minute improvement of quantification accuracy, I decided to keep single-hit proteins in future analyses. The results of these mixed-proteome experiments highlight that mass spectrometry-based proteome analyses result in reliable, specific protein identification by careful FDR parameter optimization as well as accurate relative protein quantification, which is essential in both research and translational proteomics.

2.2.3 False discovery rate and quantification accuracy of identified phosphopeptides

Building on the results from the global proteome analysis, I next applied a similar experimental strategy to evaluate false discovery rates and quantification accuracy in phosphoproteomics. Post-translational modifications like phosphorylation add a layer of complexity to proteomic data analysis, and ensuring confidence in both site localization and quantification is particularly important for biological interpretation, especially in signaling-focused or clinical studies.

For this, I prepared a series of samples composed of phosphosite-enriched human peptides derived from the MCF7 breast cancer cell line, spiked into a constant background of 500 ng commercial yeast tryptic digest. By varying the amount of spiked-in human phosphopeptides, I could not only evaluate quantification performance but also assess the actual false discovery proportion, i.e., the number of human phosphosites falsely identified in the pure yeast background. Each sample was measured using the phospho-DIA acquisition method (Table 6, Table 8) over a 120 minutes LS-MS method, ensuring deep phosphoproteome coverage. Raw data were searched in Spectronaut (v17) using the directDIA+ workflow, with both a human and a yeast proteome database. Importantly, during the initial search, no threshold was applied for phosphosite localization confidence, to capture the full set of potential matches. Following identification, I applied post-processing using the PeptideCollapse (Bekker-Jensen et al. 2019) plugin in Perseus (Tyanova and Cox 2018), removing non-phosphorylated peptides and retaining only phosphopeptides with >95% localization probability, ensuring high confidence in site assignment.

Under default Spectronaut search parameters, I observed 1,303 human phosphosites falsely identified in the pure yeast background sample. This corresponds to a false discovery proportion of 5.4% (Figure 12A), indicating that phosphoproteomics, even with site localization filters, is susceptible to similarly high false positive rates as the full proteome. However, applying the stricter identification settings established in the earlier experiments (precursor PEP 0.01, protein q-value 0.01, protein PEP 0.05) drastically reduced the number of false-positive identifications. Only 68 human phosphosites were detected in the pure yeast sample, corresponding to a false discovery proportion of just 0.3% out of a total of 23,355 confidently localized phosphosites. This reinforces the effectiveness of stricter FDR control in minimizing erroneous identifications, even in highly complex phosphopeptide datasets.

To evaluate quantification accuracy, I calculated $\log_2$ intensity ratios between the sample with the highest amount of spiked-in human phosphopeptides (8 µL) and the other dilution points. The resulting distribution of phosphopeptide ratios clustered closely around the expected values, supporting the quantitative robustness of the workflow (Figure 12B). However, there was a slight overestimation of fold changes visible at all spike-in levels. This may be attributed to column saturation effects during LC-MS analysis, which could affect ionization efficiency and signal linearity at higher peptide loads (Alfaro et al. 2014).

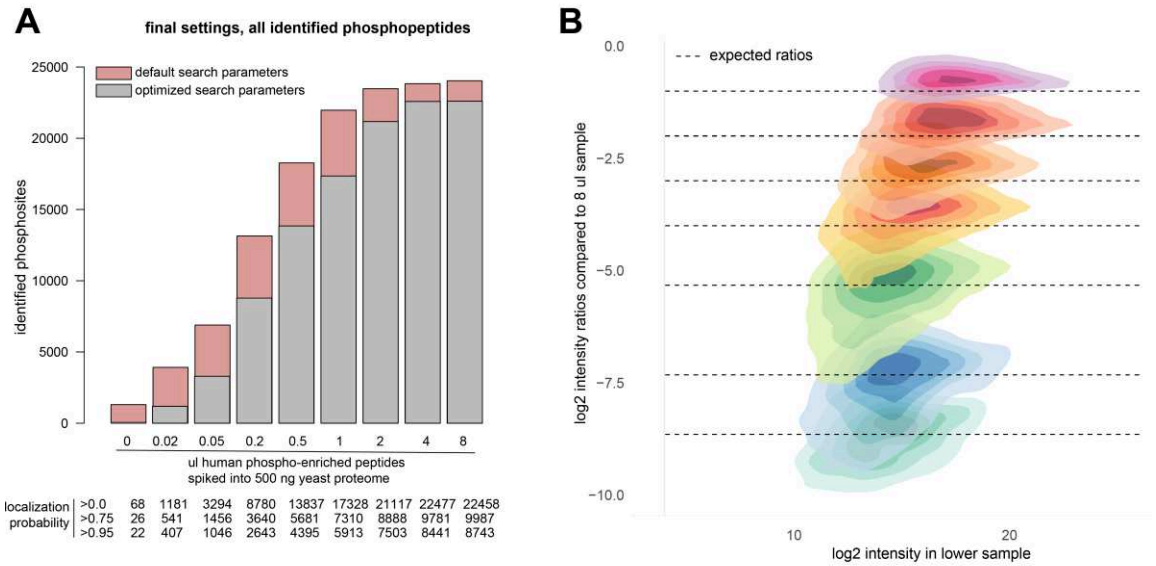


Figure 12: Controlling false-discovery rate of phosphopeptides
MCF7 cells were lysed in RIPA buffer, protein digestion and phosphopeptide enrichment were performed following the manual workflow (4.4.1). The phosphopeptide-fraction was vacuum centrifuged to dryness and subsequently dissolved in 16 µl MS buffer and pooled. Different amounts ranging from 0 µl to 8 µl were spiked into a constant background of 500 ng commercial yeast digest. Each sample was measured for 120 minutes with the phospho DIA method and raw data were searched against the human (20,594 entries, downloaded from Uniprot 09.02.2023) and yeast (6,060 entries, downloaded from Uniprot 06.03.2024) protein databases without isoforms using Spectronaut (version 17) in directDIA mode with either the default settings of the optimized, more stringent, settings (Table 9). (A) Numbers of identified and quantified human phosphopeptides in each sample. (B) Log2 ratios of phosphopeptide intensities between the highest spike-in (8 µl) and any other sample were computed, compared to the expected ratio and shown as 2D density plots (plotted using geom_density_2d_filled() function of ggplot2). The x-axis represents the log2 intensity of each peptide in the sample with lower amounts of spiked-in human phosphopeptides that was used to compute the respective ratios.

Overall, these results demonstrate that, even in phosphoproteomics, where complexity and variability are inherently higher compared to the full proteome, appropriate adjustment of search parameters can significantly improve the identification confidence and the identified phosphosites are trustworthy and quantitative. This is especially important when working with clinical datasets where both the presence and abundance of specific phosphosites are used to draw mechanistic or diagnostic conclusions.

Having established robust sample processing, measurement, and raw data search procedures, a tissue lysis protocol needed to be implemented to reach the final goal of measuring clinical samples.

2.3 Optimization of tissue sample processing protocol

2.3.1 Establishment of tissue lysis procedure

To ensure high-quality protein extraction from clinical tumor tissue specimens, I established a robust tissue lysis protocol optimized for both global proteomics and phosphoproteomics. Efficient lysis is particularly critical in tissue samples, where cell density, extracellular matrix content, and enzymatic activity can vary widely, often compromising protein recovery and post-translational modification stability if not properly controlled.

The final protocol was developed as a two-step process:

1. Mechanical disruption of the frozen tissue in lysis buffer using the TissueLyser II instrument (Qiagen)
2. Ultrasound sonication using the Bioruptor Pico (Diagenode) for protein extraction and concurrent DNA/RNA shearing

Unlike cell culture samples, tissue lysates are not sterile and may contain bacterial proteases or other enzymes that degrade proteins and post-translational modifications, in addition to the human cells' proteases and phosphatases (Ciernikova et al., 2022). To address this, I compared three different lysis buffer formulations:

Buffer #1: 8 M urea, 100 mM NaCl, 50 mM TEAB, cOmplete EDTA-free protease inhibitor (Merck), PhosSTOP phosphatase inhibitor (Merck).

Buffer #2: 8 M urea, 50 mM Tris pH 8.2, supplemented with a variety of protease and phosphatase inhibitors (Table 3)

Buffer #3: 4% SDS in PBS

Additionally, for buffer #2, I tested whether pulverizing frozen tissue on dry ice prior to lysis would enhance protein recovery. While tissue pulverization showed no significant benefit in protein yield or depth of coverage, adding extra inhibitors to the urea-based buffer (Buffer #2) had a marked positive effect (Figure 13A). With the optimized lysis buffer, I achieved identification of nearly 7,000 protein groups and around 9,000 phosphosites, indicating improved stabilization of labile targets during the lysis process. While the SDS-based buffer (Buffer #3) performed equally well for full proteome profiling, it proved less compatible with phosphoproteomics, possibly due to interference during phosphopeptide enrichment, where negatively charged SDS molecules might compete with phosphorylated peptides.

The next step in protocol development was to determine the optimal number of Bioruptor sonication cycles required to maximize protein yield. Each sonication cycle consisted of 30 seconds of sonication followed by 30 seconds of rest at 4°C. I mechanically disrupted multiple fresh-frozen tumor tissue samples (test tissues obtained from SPL: esophageal carcinoma #3TRFXU and ovarian carcinoma #5NHVNT) in a urea-based lysis buffer, pooled the lysates, and

then subjected aliquots to varying numbers of sonication cycles (0 and 4–7 cycles). Protein concentrations were quantified using a BCA assay. The results showed that 4 or more sonication cycles consistently resulted in high protein yields, with minimal gains observed beyond 5 cycles. Compared to no sonication, 5 cycles increased protein yield by 15% and 24% for the two tissue types, respectively. Further, the addition of 50 mg Diagenode sonication beads slightly improved protein extraction efficiency, resulting in an additional 5–7% increase in protein yield (Figure 13B).

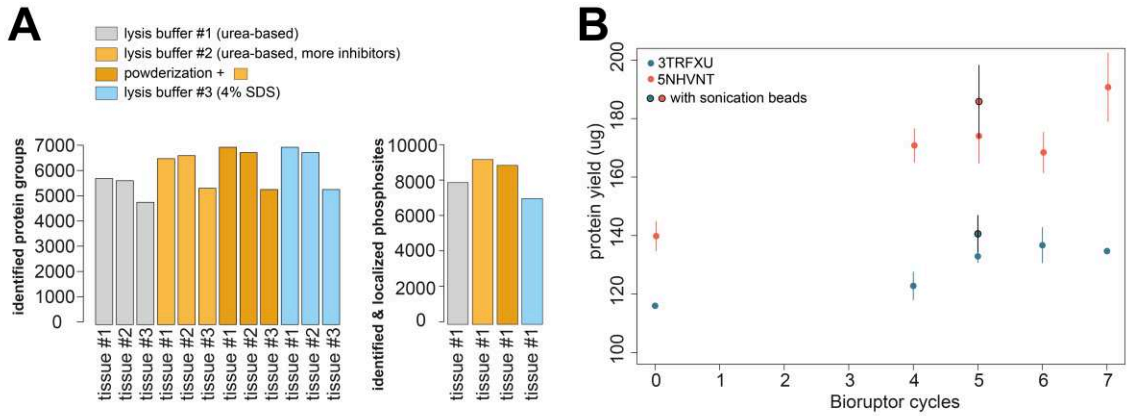


Figure 13: Optimization of tissue lysis procedure.
Two different fresh-frozen tissue samples (3TRFXU: esophageal carcinoma; 5NHVNT: ovarian carcinoma) were cut on dry ice into roughly 2 mm x 2mm x 2mm big pieces. (A) Different lysis buffers were tested on three different fresh-frozen tissue samples (all CRC). For lysis buffer #2, I additionally performed tissue powderization on dry ice prior to lysis. With the SDS buffer, the samples were incubated at 95°C for 10 minutes instead of TissueLyser disruption. Protein digestion was performed using the automated workflow, starting with 10 µg protein. From tissue #1, automated phosphopeptide-enrichment was performed starting with peptides from 55 µg protein. Samples were measured for 120 minutes with DIA and searched against the human database without isoforms (20,594 entries, downloaded from Uniprot 09.02.2023) using Spectronaut (version 17) in directDIA mode. Phosphosites were filtered for localization probability > 95%. (B) Steel bead and 8 M urea lysis buffer were added, followed by mechanical disruption with the TissueLyser II (2 x 2 minutes at 20 Hz). Lysates were pooled and aliquoted into 200 µl in Bioruptor sonication tubes. Sonication beads (50 mg) were added optionally. Lysates were sonicated for 4-7 cycles (30 seconds on, 30 seconds off). Protein yield was quantified using Pierce BCA assay.

Based on these optimizations, the final protocol is as follows (Figure 14):

1. Mechanical disruption
    a. Keep the fresh-frozen tissue piece on dry ice until processing
    b. Add one steel bead and 200 µL of optimized urea-based lysis buffer with inhibitors
    c. Mechanically disrupt using TissueLyser II for 2 × 2 minutes at 20 Hz
    d. Immediately place the lysate on ice
2. Ultrasound sonication
    a. Transfer the sample to Diagenode sonication tubes
    b. add 50 mg sonication beads
    c. sonicate for 5 cycles using the Bioruptor Pico

Protein extracts prepared using this protocol were then processed using the automated SP3 protocol and phospho-enrichment workflow described in the earlier sections.

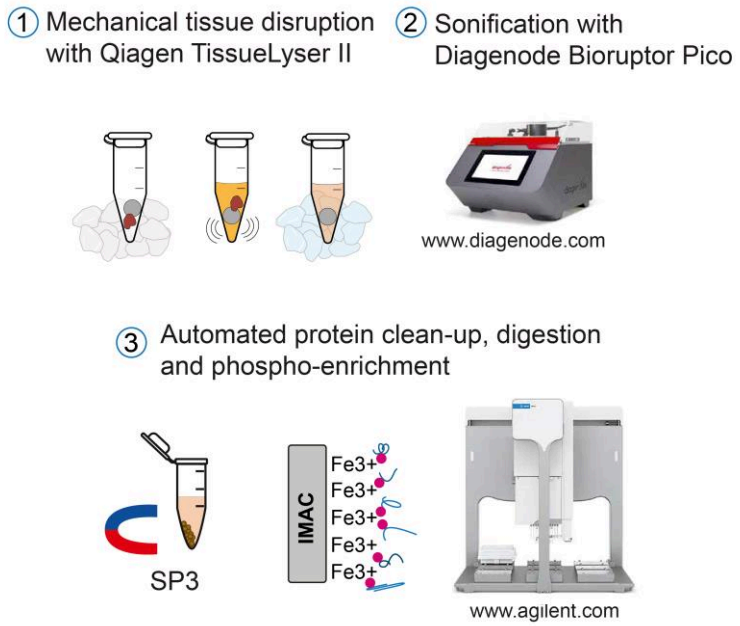


Figure 14: Final tissue lysis and MS sample preparation procedure.
Steel bead was added to frozen tissue on dry ice. Following the addition of the urea-based lysis buffer, mechanical disruption was performed using the TissueLyser II for 2 × 2 minutes at 20 Hz. Lysates were transferred to Bioruptor tubes and 50 mg sonication beads were added. Lysates were ultrasound sonicated for 5 cycles using the Bioruptor Pico. Following BCA assay, the protein extracts were then processed using the automated SP3 protocol and phospho-enrichment workflow.

2.3.2 Application of tissue lysis workflow to NCT-MASTER samples reveals high variance in sample quality

Now that tissue lysis, automated sample prep, and raw data analysis were established, the workflow was applied to clinical samples. I received 65 fresh-frozen colorectal cancer (CRC) tissue samples from the NCT sample processing lab (SPL) in two batches about eight months apart from each other. These tissue samples were collected from patients that had been enrolled in the DKFZ/NCT/DKTK-MASTER program and were analyzed retrospectively.

After applying the full established tissue proteome workflow to these samples, I observed a high variance regarding both protein and phosphosite coverage. The number of protein groups identified per patient ranged from 3,000 to more than 8,000 (Figure 15A) and the number of identified and localized phosphosites from the tissue pieces ranged from 3,000 to more than 25,000 (Figure 15B).

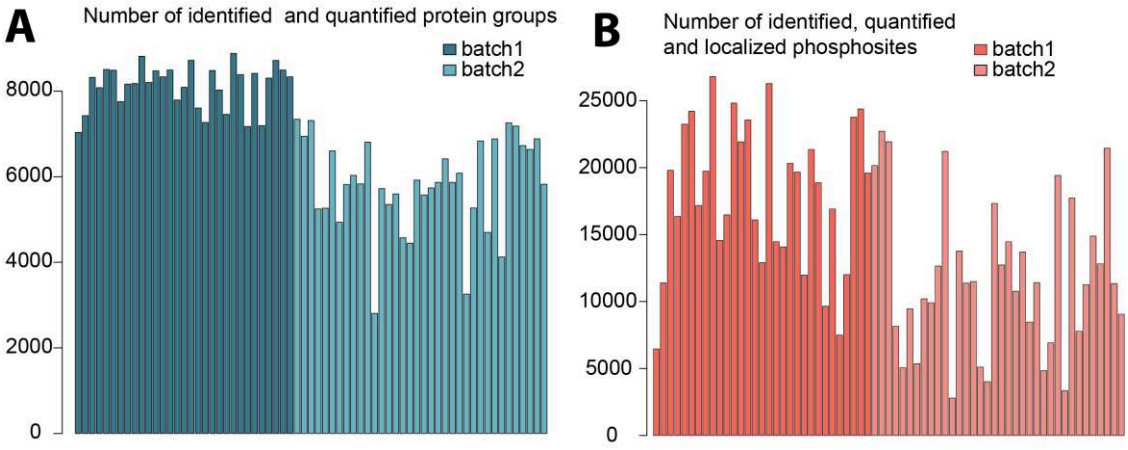


Figure 15: Application of workflow to DKFZ/NCT-MASTER samples.
A total of 65 fresh-frozen CRC tissue samples from the DKFZ/NCT/DKTK-MASTER program were received from the NCT sample processing lab in two batches (batch1: March 2023; batch2: November 2023). Tissue lysis was performed following the optimized workflow (Figure 13), protein concentration was determined with the Pierce BCA assay, and protein digestion and phosphopeptide-enrichment was performed using the automated workflow starting with 55 µg protein. Full and phosphoproteome fractions were measured for 120 minutes each using the (phospho) DIA method (

Table 6, Table 7, Table 8). Raw data were searched against the human protein database without isoforms (20,594 entries, downloaded from Uniprot 09.02.2023) using Spectronaut (version 18) in directDIA mode with optimized settings (Table 9). Identified phosphopeptides were site-collapsed to the gene level with Perseus (v 1.6.2.3) and plugin PeptideCollapse (v 1.4.3, cut-off 0.95). Shown are numbers of identified protein groups (A) and phosphosites (B) per patient sample.

In the following, I aimed to identify the source of these apparent differences in sample quality. Were there intrinsic variations to the clinical samples, like protein degradation, or was the sample preparation process unreproducible or faulty, such as insufficient DNA and RNA removal?

2.3.3 Variance in sample quality was not related to duration of storage, but rather appeared to be sample-intrinsic

Eight of the 65 patient samples had no mRNA sequencing data available, which was – according to the SPL - due to tissue degradation and resulting low quality of the extracted mRNA. I hypothesized that tissue samples with low quality mRNA might also present with degraded protein content. To test this hypothesis, I compared the numbers of identified protein groups between the samples with (n = 57) and without (n = 8) available mRNA sequencing data and found that the samples with insufficient mRNA quality indeed showed significantly lower numbers of identified protein groups (Figure 16A) as well as phosphosites (Figure 16B). However, all eight samples with low-quality mRNA stemmed from batch number 2. This could either mean that the comparison is biased or that something changed in the handling or storage of the samples between batch 1 and batch 2 causing lower quality tissues. Several of the poorly performing samples did have mRNA sequencing data available, suggesting that poor mRNA and protein quality do not necessarily coincide. Further, I estimated the duration of sample storage in the SPL biobank by subtracting the MTB date of each patient from the date that I processed the samples. The sample storage time ranged between 0 and 99 months with a median storage time of 23.5 months. There is a slight positive correlation between storage duration and number of identified proteins (Figure 16C) or phosphosites (Figure 16D), however with very low R^2 values. Interestingly, there is a high correlation between the proteome and phosphoproteome coverage in both batches (Figure 16E), suggesting general quality differences of the samples. Most samples of batch 1 (30 out of 31) were obtained at the NCT Heidelberg, while an increasing number of samples from NCT partner sites in Berlin, Dresden, Mainz and Munich were included in batch 2 (Figure 16F). Compared to Heidelberg, the proteome coverage was significantly lower in samples from Berlin and Munich, suggesting that differences in tissue handling or the longer shipping time could influence the sample quality.

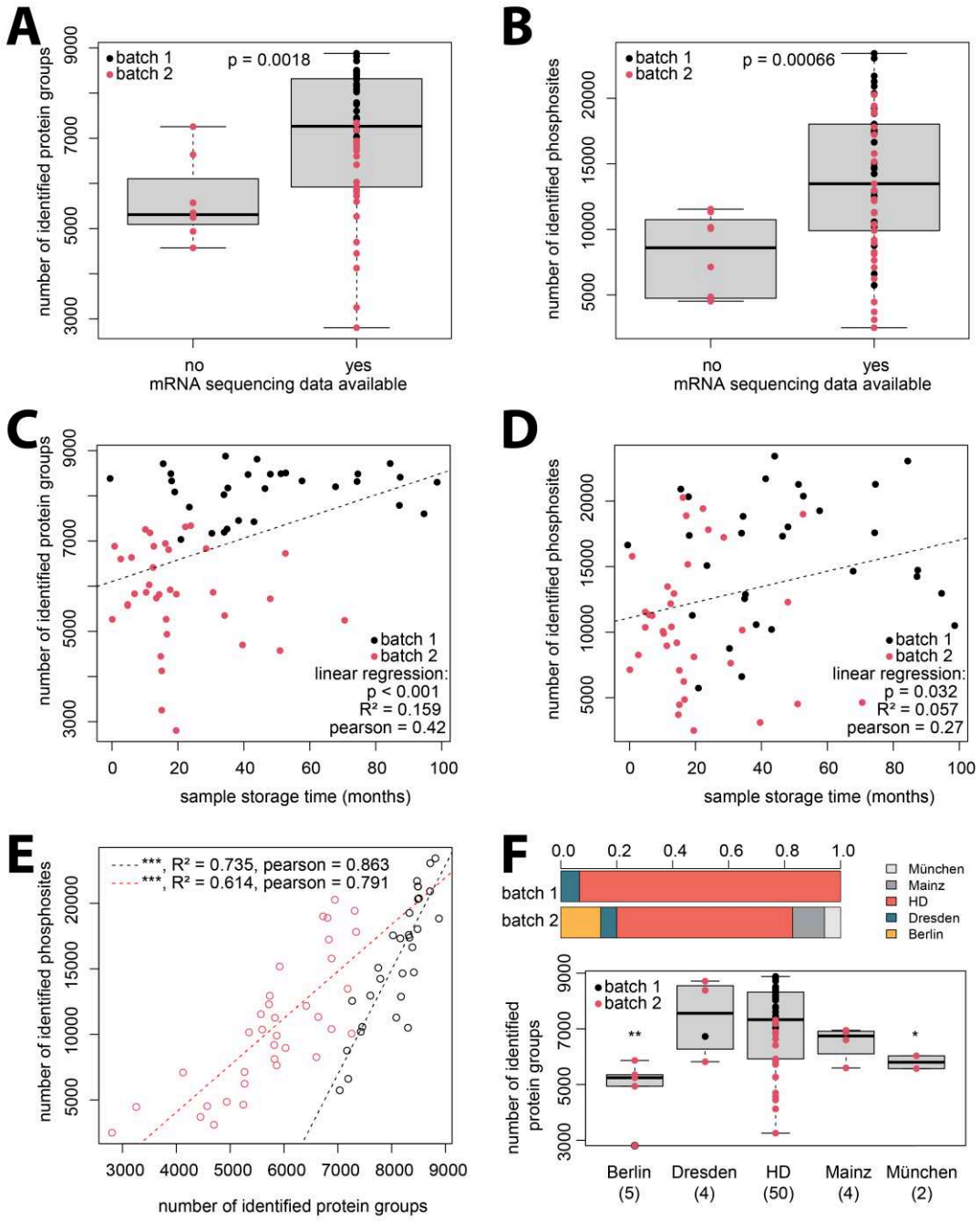


Figure 16: Identifying the source of sample quality variance

(A) Number of identified protein groups in each patient sample was grouped by availability of RNA sequencing data. Significance was assessed using a two-sided Student's t-test. The sample batch is indicated by color. (B) Number of identified and localized phosphosites in each patient sample was grouped by availability of RNA sequencing data. Significance was assessed using a two-sided Student's t-test. The sample batch is indicated by color. (C, D) The sample storage time was estimated by using the date of the molecular tumor board (minus two weeks) and subtracting it from the date I processed the samples. Correlation between the numbers of identified protein groups (C) or phosphosites (D) and the sample storage time was assessed using Pearson's correlation coefficient (cor) and linear regression (lm). (E) Pearson correlation and linear regression analyses between the numbers of identified protein groups and phosphosites, separated by batch. (F) Distribution of NCT partner clinics where samples were obtained. * = p-value < 0.05, ** = p-value < 0.01, *** = p-value < 0.001.

In conclusion, there appeared to be some connection between poor mRNA quality and poor proteome quality of a tissue sample, but the duration of sample storage did not play an obvious role. To rule out low reproducibility of the sample processing and measurement steps, as well as insufficient DNA/RNA shearing and sample clean-up, I next performed several additional tissue lysis tests.

2.3.4 Further lysis tests to identify the source of quality differences between samples

To further understand the sources of variability in sample quality, I conducted a series of additional tissue lysis and processing tests using selected samples from batch 1**,** batch 2, and three independent test tissues. The aim was to pinpoint which steps in the workflow might be responsible for the differences observed in proteome and phosphoproteome coverage, particularly in a subset of underperforming samples from batch 2.

From batch 2, I selected two well-performing and four poorly-performing samples based on their proteomic depth in the initial measurement (Figure 15). The tests were designed to rule out specific steps in the workflow and included the following (Figure 17):

i. Repetition of the full TissueLyser and Bioruptor protocol with fresh tissue pieces to assess reproducibility of the entire lysis procedure. This test was only performed with samples from batch 1, since there was no tissue left for batch 2.

ii. Repetition of autoSP3 and automated phosphopeptide enrichment from the same lysates, to evaluate reproducibility of automated sample processing

iii. Additional sonication cycles with the Bioruptor to assess whether shearing potential residual long strands of DNA/RNA would improve results

iv. Protein precipitation of the original lysates to remove possible contaminants, such as fatty acids, salts, or nucleic acids, that might interfere with protein digestion or phosphopeptide enrichment

v. Testing long-term storage of the lysis buffer by comparing freshly prepared buffer to aliquots that had been stored at –80 °C for one year

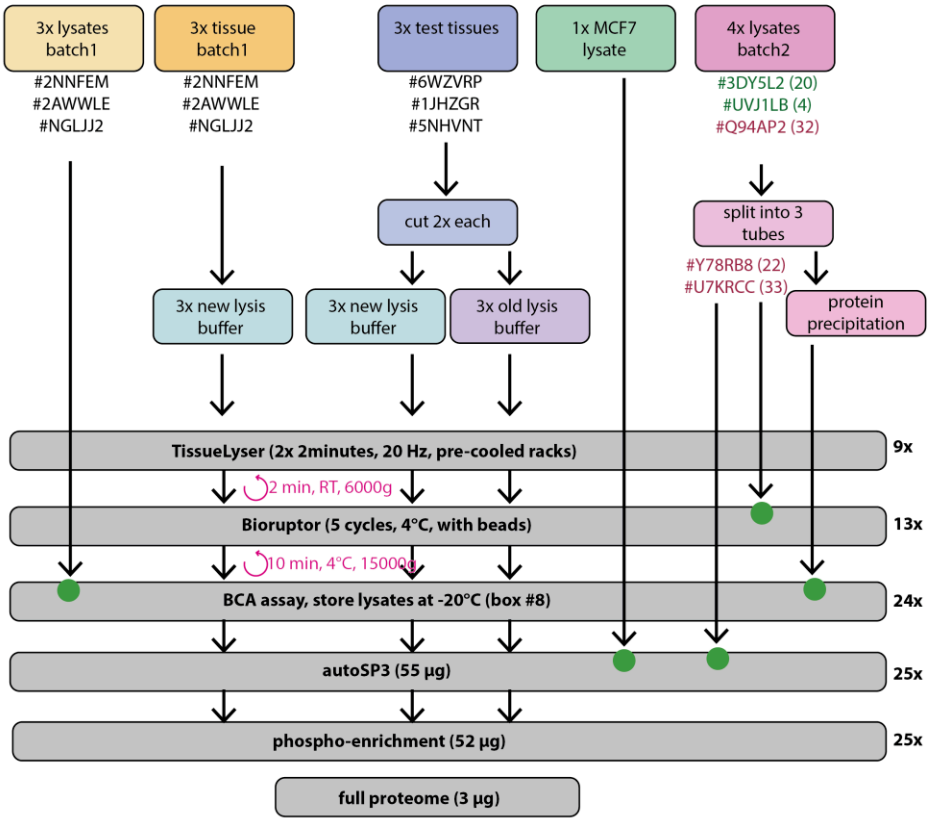


Figure 17: Additional tissue lysis tests to find the source of quality variance
To assess the reproducibility of different tissue sample preparation steps, I repeated different steps of the process with samples from batch 1, independent test tissues, and well performing as well as poorly performing samples from batch 2. I compared repetition of the entire process, additional Bioruptor sonication cycles or protein precipitation, and repetition of the automated digestion and phosphopeptide-enrichment steps. Furthermore, I tested lysis buffer prepared fresh that day against lysis buffer stored at -80°C for one year.

The repeat lysis and processing of samples from batch 1 demonstrated excellent reproducibility across all tested steps, including mechanical disruption, digestion, enrichment, and MS measurement. Protein and phosphosite coverage from these repeats (Figure 18, yellow bars) matched or even exceeded that of the MCF7 cell line lysate, confirming that when all conditions are optimal, the protocol yields robust and consistent results.

For the test tissues, the influence of lysis buffer storage showed sample-dependent effects. In one case (TT_1), there was no observable difference between fresh and stored buffer. However, for the other two test tissues, the freshly prepared buffer resulted in higher proteome coverage (+39% and +11%, respectively; Figure 18, blue/purple bars). This suggests that preparing the lysis buffer fresh for each experiment might be preferable, since aliquoting and storing at –80 °C does not guarantee consistent performance across all samples.

Repeating autoSP3 and phosphopeptide enrichment for the lysates from the two well-performing batch 2 samples led to slight increases in protein identifications (+15% and +6%, respectively), indicating some degree of variability in these steps. Since additional sonication and protein precipitation did not result in any further gains (Figure 18, green bars), it appears that autoSP3 or phospho-enrichment steps, and potentially instrument-related variance, could be the more likely sources of fluctuation.

In contrast, the four poorly-performing samples from batch 2 showed clear improvement when autoSP3 and phospho-enrichment were repeated from the same lysates (Figure 18, red bars). Protein identifications increased 2.3-fold in one case, and by 32–47% in the other three samples, strongly suggesting inconsistencies in automated sample processing rather than issues with lysis or lysate quality. For the future, I propose to always include a standard – like the MCF7 lysate – on any 96-well plate of samples that is processed with the automated Bravo workflows and to conduct a longitudinal evaluation of the reproducibility. Additional sonication or precipitation did not have a consistent impact, with proteome coverage changes ranging from –19% to +7%, ruling out significant interference from remaining DNA/RNA or other contaminants.

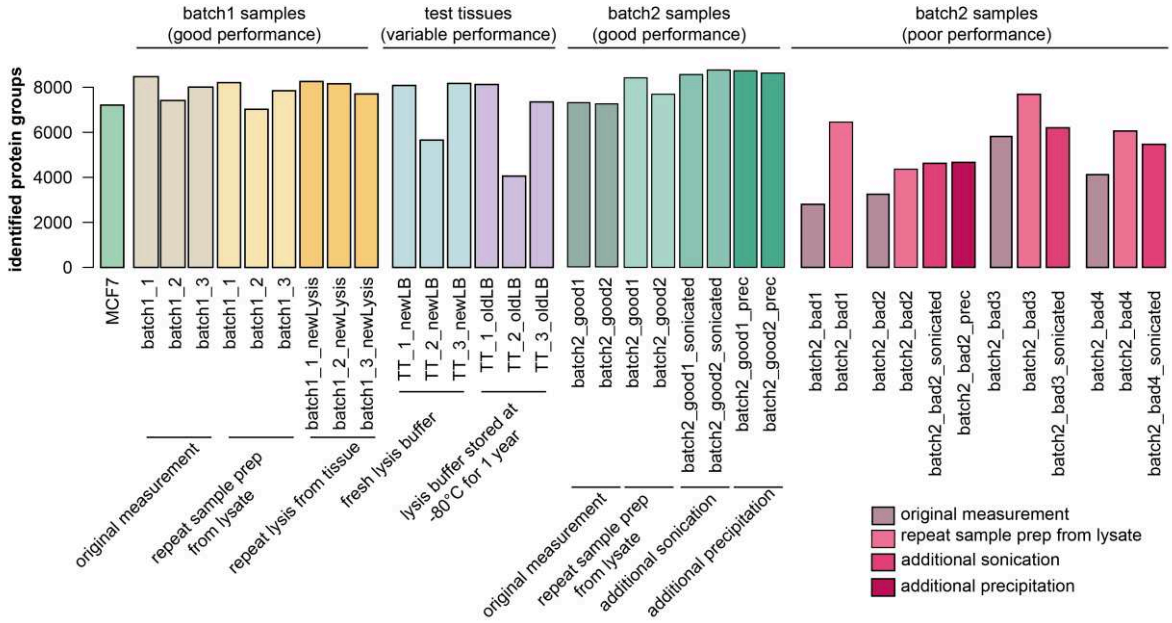


Figure 18: Full proteome result of lysis test
Result of samples from Figure 17. Protein concentration was determined using BCA assay. Starting with 55 µg, proteins were digested and peptides were phospho-enriched following the automated workflow. Peptides from 1.5 µg protein were injected for measurement of the full proteome and data were acquired for 120 minutes in DIA mode. Raw data were searched against the human proteome database without isoforms (20,594 entries, downloaded from Uniprot 09.02.2023) using Spectronaut (version 18) with the optimized settings (Table 9). Shown are numbers of identified and quantified protein groups in each sample. The MCF7 sample serves as a reference for how a standard cell culture lysate performs.

Despite the improvements resulting from the repetition of the automated protocols, these batch 2 samples continued to underperform relative to others, indicating intrinsic sample-related limitations or unresolved processing variability that cannot be fully corrected through re-processing. These findings underscore the importance of routine quality control and, where possible, repeat measurements to mitigate outlier behavior and maximize reproducibility in clinical proteomics workflows.

Evaluating the phosphoproteome measurements of these additional tests, revealed a very similar picture: For well performing samples, independent of batch 1 or 2, repeated protein digestion and phosphopeptide-enrichment as well as additional sonication or protein precipitation, did not alter the phosphoproteome coverage (Figure 19, yellow and green bars). Also, for the long-term storage of the lysis buffer, the phosphoproteome agreed with the full proteome and showed somewhat lower phosphoproteome coverage with the stored compared to the freshly prepared lysis buffer (Figure 19, blue/purple bars). This difference was moderate for TT_1 and TT_3 with an advantage of the fresh buffer of 5% and 7% respectively, whereas the fresh lysis buffer had an immense impact on TT_2 and increased the phosphoproteome coverage 2.7-fold. Since the lysis with either the stored or the fresh buffer can never be performed on exactly the same tissue piece, I cannot exclude the possibility that the observed trends could be explained by quality differences at different sites of the tissue. However, there seems to be a trend that fresh preparation is preferable over long-term storage of the lysis buffer.

Repeating the entire lysis and sample processing step from fresh tissue pieces of samples from batch 1 did indeed lead to some differences (Figure 19, yellow bars): Compared to the initial measurement, the first sample lost 27% of phosphoproteome coverage, the second sample gained 69%, and the third sample lost 17%. This could suggest that in contrast to the full proteome, the phosphoproteome is quite affected by either slight differences in the lysis conditions or – more likely – by quality differences at different sites of the tumor.

As for the poorly performing samples of batch 2, repetition of the protein digestion and phosphopeptide-enrichment steps improved the phosphoproteome coverage in only one sample, while it even slightly decreased the numbers in the other three samples (Figure 19, red bars). This suggests that there could be some noticeable variation in the automated sample processing steps, however they do not explain the overall poor performance of these samples. Again, additional sonication steps or protein precipitation did not improve the sample quality.

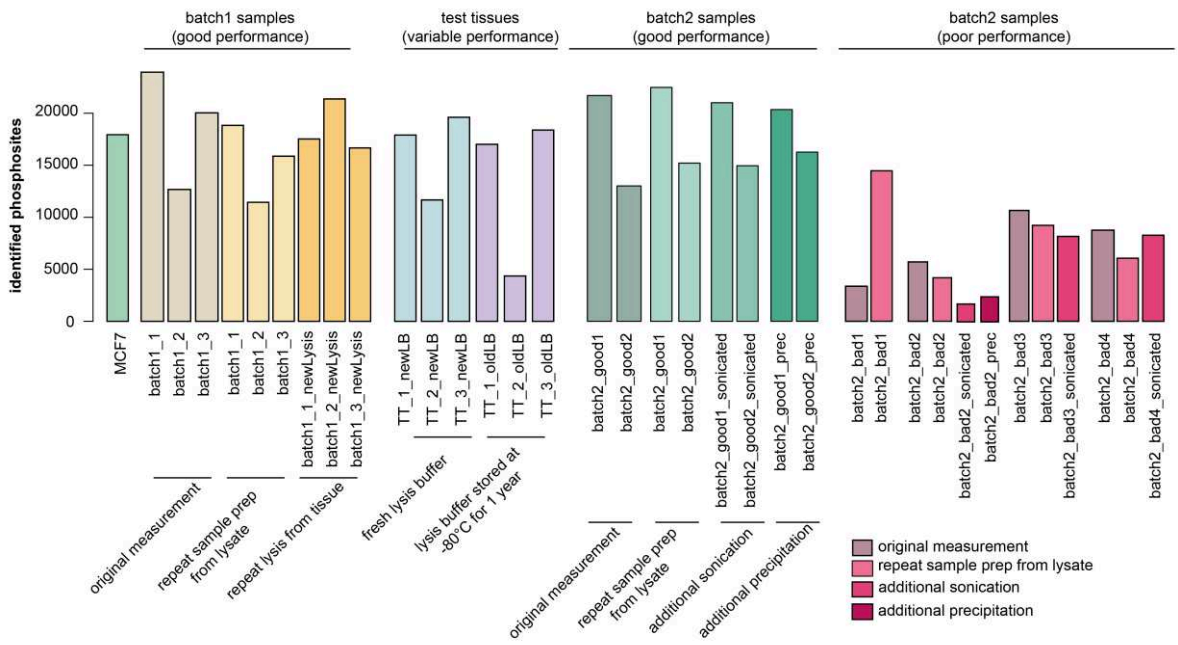


Figure 19: Phosphoproteome result of lysis test.
Result of samples from Figure 17. Protein concentration was determined using BCA assay. Starting with 55 µg, proteins were digested and peptides were phospho-enriched following the automated workflow. The entire phospho-enriched fraction was injected for measurement and data were acquired for 120 minutes in phospho-DIA mode. Raw data were searched against the human proteome database without isoforms (20,594 entries, downloaded from Uniprot 09.02.2023) using Spectronaut (version 18) with the optimized settings (Table 9). Shown are numbers of identified, quantified, and localized (>0.95) phosphosites in each sample. The MCF7 sample serves as a reference for how a standard cell culture lysate performs.

In conclusion, the tissue lysis procedure appeared to be very reproducible, the number of Bioruptor sonication cycles proved sufficient, and protein precipitation was found to not be advantageous. Automated SP3 and phospho-enrichment showed some variance but overall it seems more likely that the differences are intrinsic to the samples and have nothing – or not much - to do with the lysis and MS sample preparation steps. My best guess is that differences in ischemia time during sample collection after surgery, differences in sample transportation, and storage might influence the sample quality.

2.4 Description of colorectal cancer cohort samples received from DKFZ/NCT/DKTK-MASTER program

For all 65 patients included in this study, comprehensive molecular profiles were available, including full and phosphoproteome acquired by myself as well as single nucleotide variant (SNV) information previously acquired and shared with me by the NCT-MASTER program. Additionally, mRNA sequencing data were available for all but eight patients, and copy number variant (CNV) data were available for all patients who underwent whole genome sequencing (WGS), which

included 49 out of 65 patients. The cohort had a balanced sex distribution, with 32 male and 33 female patients (Figure 20).

What made this dataset particularly unique is the age profile of the patients: the cohort consisted of unusually young individuals for colorectal cancer (American Cancer Society), with ages at the time of the molecular tumor board ranging from 20 to 80 years, and an average age of just 43 years. Another important characteristic is that most tissue samples were not derived from the primary tumor (n = 11), but rather from metastatic sites (n = 54). Unfortunately, the exact anatomical location of the metastases was not consistently documented. In some instances, metadata indicated that the biopsy was taken for example from the liver or lung, but in the majority of cases, this information was missing.

Only five patients (7.7%) showed MSI sensor values higher than 3.5 which is recommended as the cut-off (Niu et al. 2014), indicating that microsatellite instability (MSI) was not a common feature in this cohort, but comparable to published MSI frequencies (5%) in metastatic CRC (Sinicrope and Sargent 2012).

The cohort was also marked by extensive prior treatment. As part of the NCT-MASTER program, patients were not treatment-naïve, instead, they typically presented after years of heavy pre-treatment, having received up to 11 different therapeutic agents per patient. Importantly, no two patients shared the exact same pre-treatment regimen, highlighting the heterogeneity of this cohort. Among the most frequently administered therapies were 5-fluorouracil (5-FU), received by 90% of patients (mean duration: 10.2 months), oxaliplatin (administered to 81% of patients for a mean of 4.6 months), and irinotecan (given to 71%, average 5.1 months). About one third of the patients also underwent one to four rounds of radiation therapy (Figure 20).

In addition to standard chemotherapeutics, patients were treated with a variety of targeted agents, including anti-VEGFR therapies (e.g., Bevacizumab, Ramucirumab, Aflibercept, Zaltrap), immune checkpoint inhibitors (ICIs) (e.g., Ipilimumab, Nivolumab, Pembrolizumab, Atezolizumab), and anti-EGFR antibodies (e.g., Cetuximab, Panitumumab). Further treatments included additional chemotherapeutic agents such as Gemcitabine, Vinorelbine, and Docetaxel.

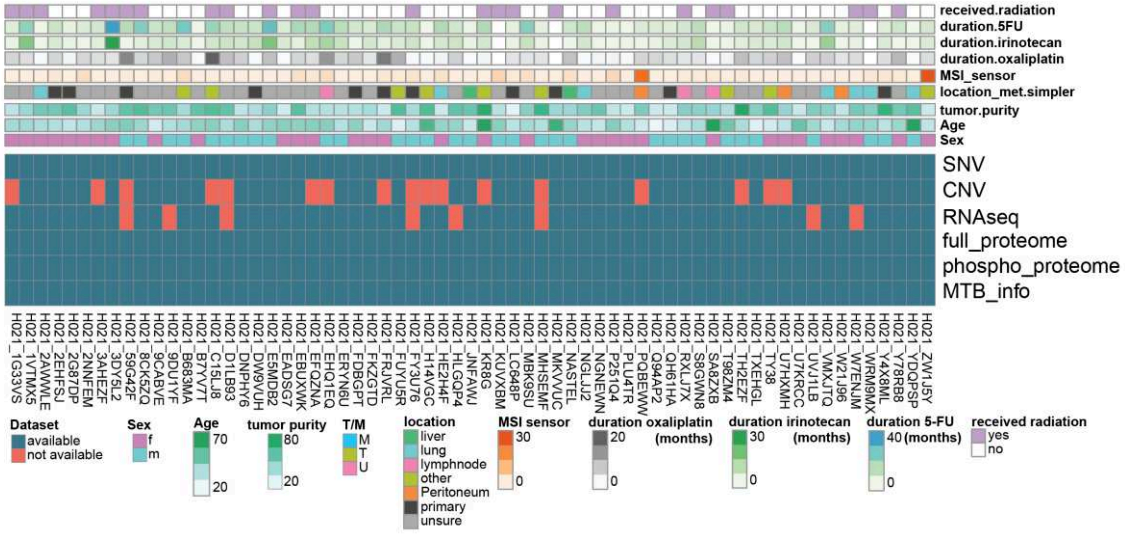


Figure 20: Data availability and clinical patient metadata
Full and phosphoproteome data were acquired by myself in this present study. Single nucleotide variants (SNV), copy number variants (CNV), RNA sequencing (RNAseq) and metadata were provided by the DKFZ/NCT/DKTK-MASTER program. CNV data were available for 48 patients, RNAseq for 57 of the 65 patients. Clinical metadata included, amongst others, sex, age, estimated tumor purity, whether the sample was obtained from the primary tumor (T) or a metastatic site (M) or unknown (U), and in some cases the organ of the metastatic site. Furthermore, detailed documentation of received treatments, such as radiatio nor chemotherapy, were available.

In conclusion, the present CRC patient cohort represented a unique and novel data set characterized by relatively young and heavily pre-treated patients, presenting with end-stage, metastatic disease. Since the tissue samples had been obtained from different (and often unknown) metastatic sites and no two patients have received the same pre-treatment regimen, comparing different subgroups within this cohort will be highly influenced by all these confounders. Instead, I attempted a personalized analysis comparing each patient to the rest of the cohort in the upcoming chapters.

2.5 Quality control of kinase activity estimations

2.5.1 Patient samples with few identified phosphosites still gave reliable kinase activity estimates

Although phosphoproteome data were technically available for all 65 patients, the coverage varied substantially, with some samples containing as few as 3,000 quantified phosphosites. This raised an important question: Can meaningful conclusions still be drawn from samples with such limited phosphoproteome depth? To address this, I performed a series of quality control steps aimed at assessing the reliability of relative kinase activity estimates even in poorly performing samples.

I began by selecting the top 10 samples with the highest numbers of identified phosphosites (Figure 21A) and calculated relative kinase activities for each patient, using z-scored phosphosite intensities and requiring at least five known substrate sites per kinase. To get robust z-scores, I had

filtered out phosphosites that were identified in less than 20 samples. I then repeated the analysis, but restricted it to only the phosphosites that were also present in each of the 10 samples with the lowest phosphosite coverage. For example, the 14,788 identified phosphosites in the tumor sample from patient H14VGC were filtered to only retain those sites that were also identified in the low-performing sample 1G33VS, resulting in the deletion of 63% of phosphosites. The filtered H14VGC sample now mimics the performance of the 1G33VS sample and by correlating the kinase activities based on the complete H14VGC data set compared to the filtered data set – mimicking 1G33VS – I was able to evaluate whether reliable kinase activity estimates can be drawn even from low-performing samples.

As expected, the number of kinases for which activities could be calculated dropped from over 150 to between 70 and 110, depending on the sample pairing, due to fewer identified known target sites (Figure 21B).

Importantly, for the kinases that could still be evaluated, the activity estimates correlated strongly with those from the full dataset with an average Pearson correlation coefficient (r) of 0.835 (Figure 21C). Even in the worst-case pairing, deleting 85% of the identified phosphosites, the agreement remained high ($r = 0.652$), suggesting that kinase activity profiles are robust to reduced phosphosite coverage — at least when a minimum threshold of relevant target sites is still met. I had a closer look at the pair with the lowest Pearson correlation coefficient, namely top sample #2NNFEM filtered to only retain phosphosites identified in sample #D1LB93 which left only 3,913 of the original 14,127 sites. Most of the kinases that were originally (based on the full data set) found to be more active compared to all other patients – with a score higher than 2 - are still captured after deleting 72% of phosphosites to mimic a low-performing sample (Figure 21D). These include GSK3b, MAPK8, AKT1, RPS6KA3, and others. A few kinases, namely PAK1, MAPK3, and RPS6KA1 were initially found to be more active but this was lost in the filtered data set, suggesting that some kinase activities might be underestimated in the low-performing patient samples. There is not a single kinase that was newly found to be more active after filtering the data set, suggesting that the low performing samples are not at risk of falsely overestimating kinase activities.

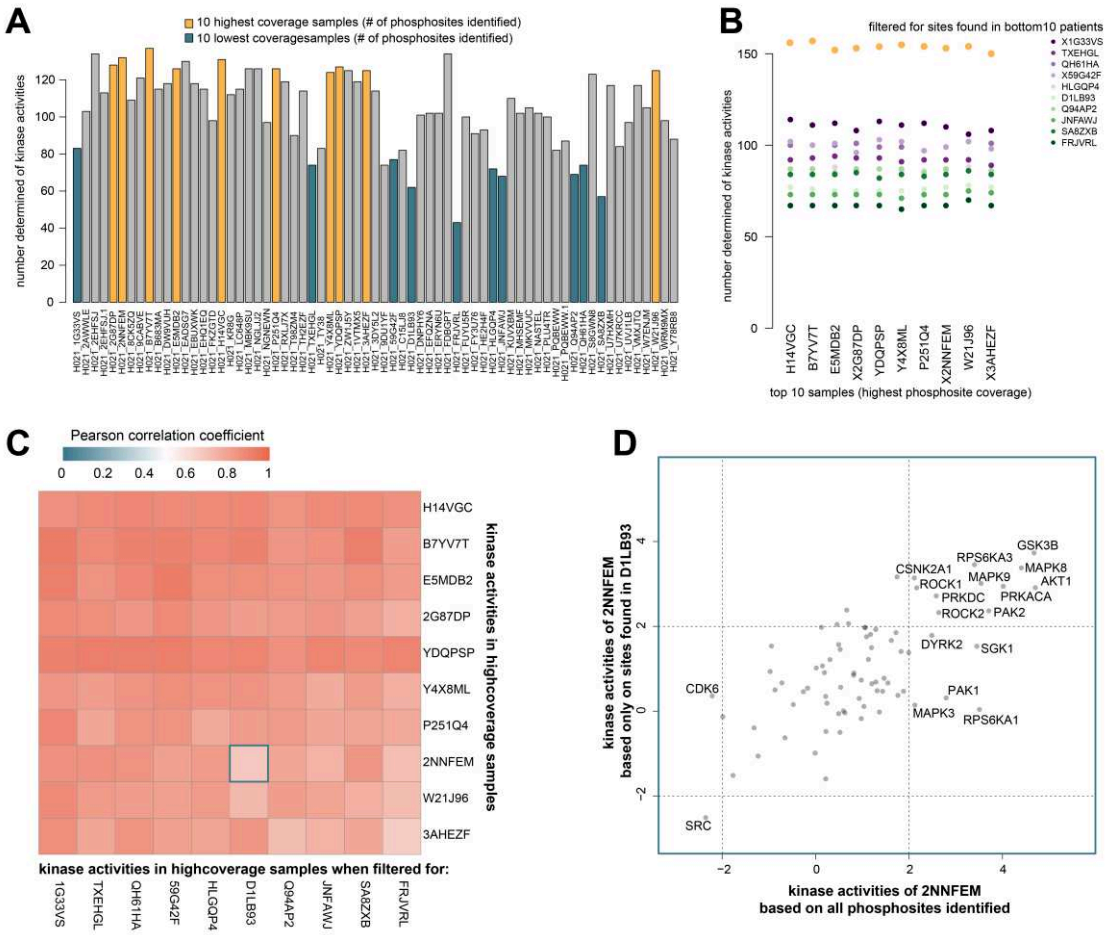


Figure 21: Patient samples with low phosphoproteome coverage still deliver meaningful insights.
The phosphoproteome data were filtered to remove phosphosites that were identified in fewer than 20 patients. Subsequently, the phosphosite intensities were z-scored across all patients. Using the decoupleR package and the Omnipath database, I calculated relative kinase activities by applying the run_wmean function to the z-scored phosphosites. Only kinases with at least 5 identified substrates were included in the analysis. (A) Number of kinases that a relative activity score could be calculated per patient. Highlighted are the 10 highest and the 10 lowest performing samples, i.e. the samples with the highest or lowest number of identified phosphosites. (B) The phosphosites of the top 10 patient samples were subsequently filtered to only retain the phosphosites identified in each of the bottom 10 samples. This reduced the number of kinases that an activity score could be calculated for. (C) The kinase activities based on the original, complete data set were correlated to the activities resulting from each of the downsampling. Shown are Pearson correlation coefficients and the lowest correlating pairing is highlighted. (D) Scatter plot of kinase activities from the lowest correlating pairing (2NNFEM down-sampled to only retain phosphosite found in D1LB93).

In conclusion, while fewer phosphosites limit the number of kinases that can be assessed, the quality and consistency of kinase activity estimation remains intact for those that do pass the threshold. This gives me confidence that even low-coverage phosphoproteome samples can be included in downstream analyses without introducing major bias.

2.5.2 Some kinases share a high proportion of target phosphosites, thus leading to indistinguishable kinase activities

Having confirmed the reliability of individual kinase activity estimations, I next asked whether activities of different kinases correlate across the patient cohort. To explore this, I calculated Pearson correlation coefficients between every pair of kinases using their relative activity scores across all 65 patients (Figure 22).

Interestingly, several kinase pairs showed high correlations, suggesting potential co-regulation. For instance, different isoforms of protein kinase C exhibited strong correlations (r between 0.82 and 0.92), as did MAPK1 and MAPK3 (r = 0.91) and various cyclin-dependent kinases (CDKs), with correlations ranging from 0.701 to 0.999.

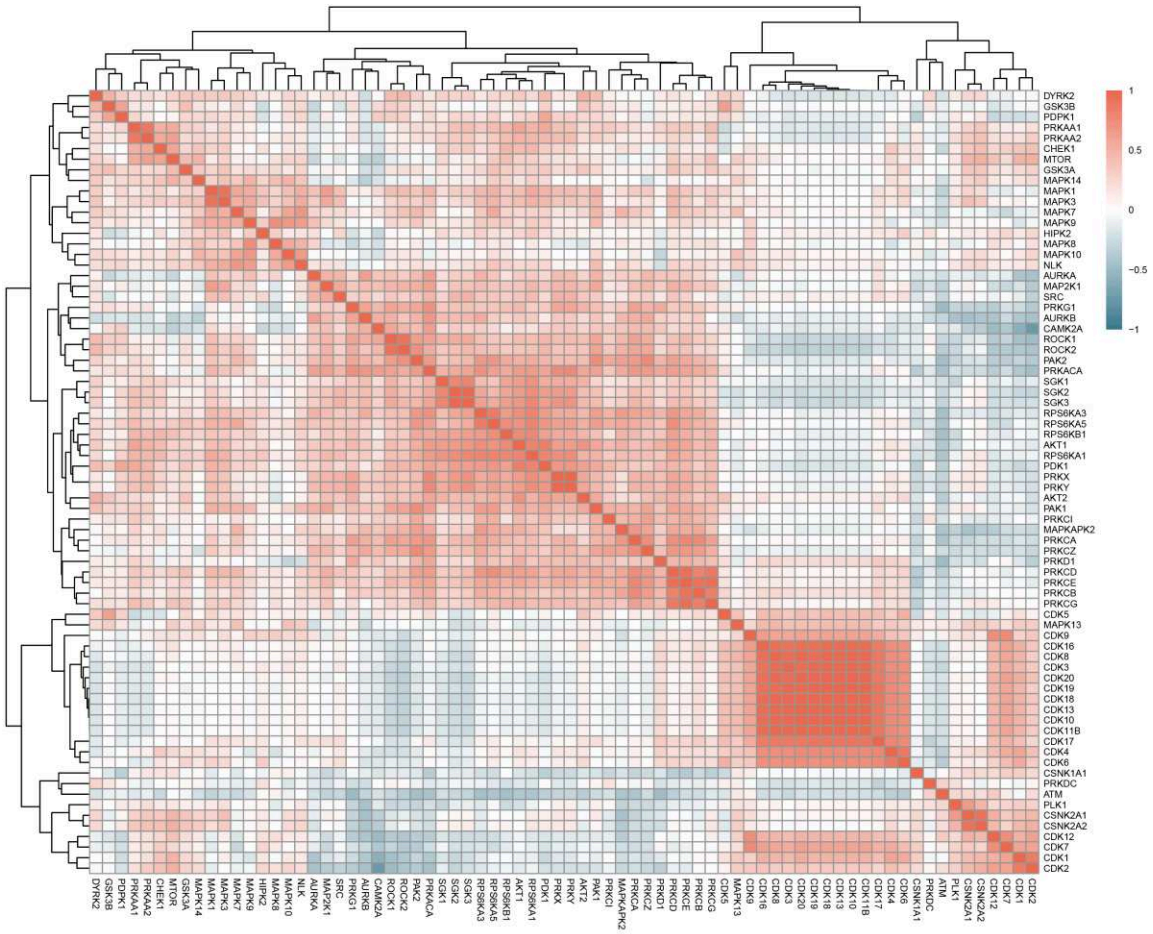


Figure 22: Correlating activities of different kinases
The activity scores of each kinase over the 65 patients were correlated with any other kinase. Shown are pearson correlation coefficients with unsupervised hierarchical clustering.

However, this raised a critical question: Are these high correlations due to genuine biological co-regulation, or are they merely artifacts arising from shared target phosphosites used in the

calculation of kinase activities? To address this, I examined the degree of substrate overlap between every possible kinase pair in the dataset (Figure 23).

As an example, 79% of AKT2's substrates (based on the phosphosites identified in my data) were also assigned to AKT1. In contrast, only 19% of AKT1's substrates were shared with AKT2, highlighting an asymmetry likely driven by the uneven depth of prior research — AKT1 has been more extensively studied and thus has more annotated phosphosites. This asymmetry implies that AKT2 activity estimates may not be trustworthy, as they are largely influenced by sites also attributed to AKT1.

To better quantify this effect, I calculated the number of exclusive phosphosites — those assigned to only one kinase and not shared with any other — for each kinase pair (Figure 23). For example, AKT1 had 65 exclusive substrates that distinguish it from AKT2, while AKT2 had only 4 unique sites, making it difficult to confidently separate its activity from AKT1. A similar pattern was observed for other kinase families: while CDK1/2/4/5/6 had numerous unique substrates, the less well-characterized CDKs (e.g., CDK16–CDK20) had fewer than five unique phosphosites, rendering their activity estimations unreliable.

On the other hand, the activities of MAPK1 (ERK2) and MAPK3 (ERK1) were also highly correlated (Figure 22), however in this care the correlation cannot be explained by a large overlap of known substrate phosphosites (Figure 23). Therefore, the correlation likely stems from actual co-regulation of these related kinases and does not reflect an artefact of the kinase activity estimation tool.

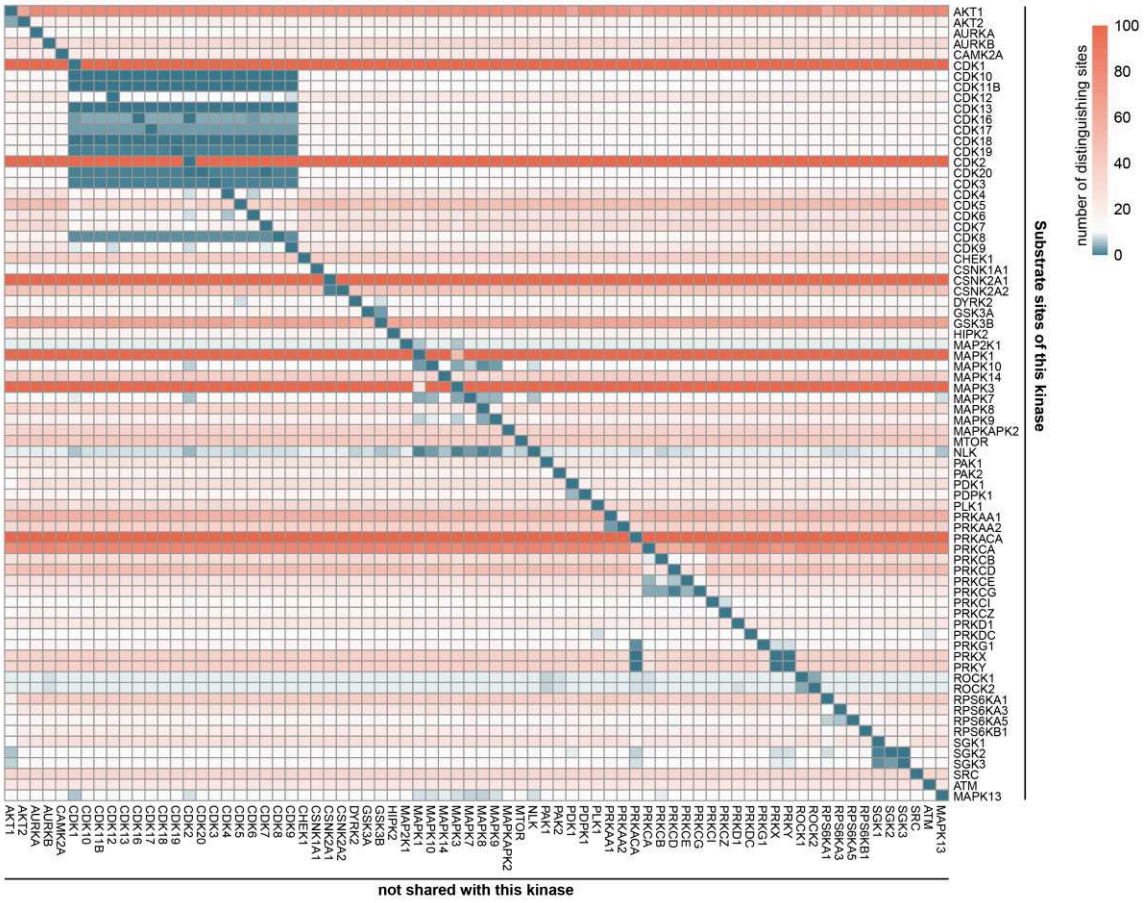


Figure 23: Numbers of unique substrates distinguishing each kinase from any other kinase
Numbers of substrates from the Omnipath database assigned to a kinase on the right that are not shared with the respective kinase on the bottom. Kinases with fewer than 5 unique substrates are highlighted in blue.

Based on this analysis, I concluded that the 28 kinases lacking a sufficient number of uniquely assigned substrates cannot be confidently assessed and should be excluded from further analyses. Therefore, all downstream figures and interpretations are based only on kinases with enough distinguishing phosphosites to ensure meaningful and specific activity profiles.

This observation likely reflects the broader issue of incomplete phosphosite annotation in the field. Many phosphosites identified in proteomics experiments are of unknown function, and their regulating kinases remain unidentified, creating a bias toward well-studied kinases like CDK2 or AKT1. This highlights the need for continued expansion and refinement of kinase-substrate databases to improve interpretability of phosphoproteomic data.

2.5.3 Estimated MAPK1 activity based on phosphoproteome footprint agrees largely with the intensities of known activating phosphosites

Some kinases have well-characterized activating phosphosites that are commonly used as markers of kinase activity in experimental settings. For example, T185 and Y187 of MAPK1 (ERK2) are widely accepted indicators of its activation and are frequently assessed via western blotting (Canagarajah et al. 1997; Shaul and Seger 2005). To further validate the reliability of my footprint-based kinase activity estimates, I compared the inferred MAPK1 activities with the intensities of these two activating sites across patients (Figure 24). The results showed moderate agreement (Spearman's $\rho = 0.439$ and 0.484), supporting the validity of the approach. However, relying solely on one or two phosphorylation sites may not capture the full complexity of kinase regulation. Not all activating sites are known, and focusing only on a few well-studied ones can provide a narrow and potentially biased view. In contrast, the footprint-based method leverages multiple downstream substrates, offering a more robust and comprehensive estimate of kinase activity.

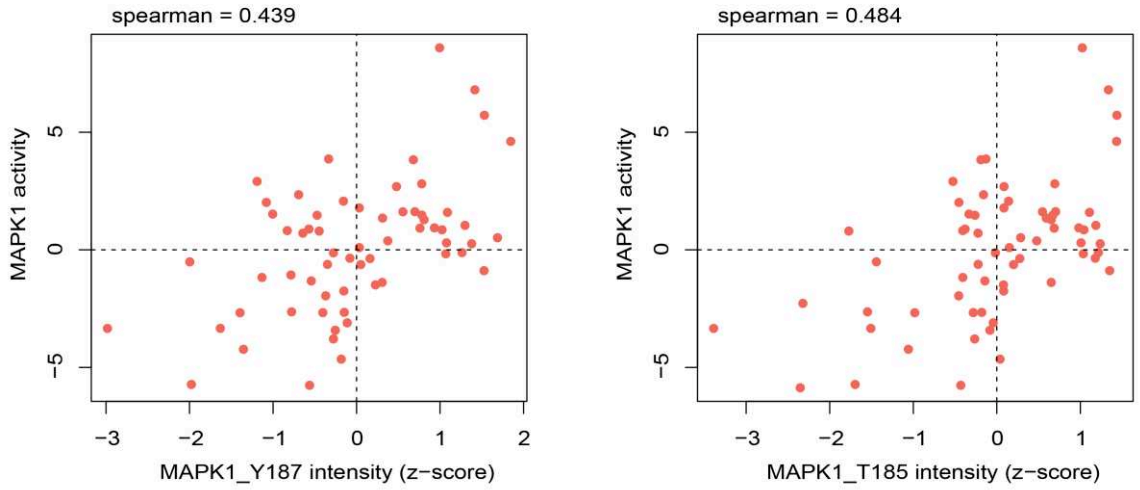


Figure 24: MAPK1 known activating sites compared to kinase activity based on footprint
Correlation of z-scored intensities of activating MAPK1 phosphosites (Y187 and T185) with the relative kinase activity score based on the entire Omnipath database. Each point represents one patient.

2.5.4 Extending the kinase activity estimation by predicted kinase-substrate interactions

One of the major challenges in interpreting phosphoproteomic data is that the biological function of the vast majority of phosphosites remains unknown - estimates suggest that only around 5% of the phosphosites identified by mass spectrometry have an annotated functional role (Hornbeck et al. 2015; Needham et al. 2019). As a result, large portions of phosphoproteomic datasets are often neglected during analysis.

In some cases, even when the function of a specific phosphosite is unknown, the upstream kinase responsible for its phosphorylation may be known. This allows us to use phosphoproteomic data as an indirect readout or "footprint" of kinase activity. However, this strategy comes with a major limitation: it is biased toward well-characterized kinases, because it relies on reference databases such as Omnipath (Türei et al. 2021), which exclusively cover experimentally validated kinase-substrate relationships.

This means that kinases like CDK1 or MAPK1, which have been extensively studied, are overrepresented, whereas less characterized kinases are largely absent or only have very few annotated phosphosite substrates. Moreover, the actual substrate landscape of each kinase may be far broader and more diverse than what has been captured in targeted experiments. Additionally, many target phosphosites are likely shared between multiple kinases and the usage might be context-specific, making strict one-to-one kinase-substrate assignments overly simplistic.

To address these limitations, several efforts have been undertaken to predict kinase-substrate relationships at scale, using either high-throughput kinase screening technologies, such as peptide arrays (Johnston et al. 2022; Yaron-Barir et al. 2024; Johnson et al. 2023), or AI-based predictions that consider the sequence motifs surrounding phosphorylation sites (Zhou et al. 2023). One particularly comprehensive integration of such efforts comes from Garrido-Rodriguez et al. (2024), who compiled multiple sources of predicted interactions into a unified framework.

To benchmark its usefulness, I compared the classical KSN, which is based solely on the Omnipath database, to the “extended KSN”, a merged resource combining Omnipath with predicted kinase-substrate interactions.

The comparison revealed a dramatic expansion: In Omnipath, each kinase has a median of 14 substrates, with a few well-known kinases like CDK1 or MAPK1 having over 500. In contrast, the extended KSN lists a median of 520 substrates per kinase, and many kinases now have more than 1,000 predicted phosphosite targets (Figure 25A). The number of kinases covered with at least 5 substrates also increases from 310 in Omnipath to 375 in the extended KSN (Figure 25B). The number of total kinase-substrate relationships jumps from 22,457 (Omnipath) to 1,049,954 (extended KSN; Figure 25C).

When applying this to my patient data, the effect is substantial: using only the Omnipath-based network, I could estimate on average 104 kinase activities per patient, but with the extended KSN, this increased to an average of 298 - almost a three-fold increase in coverage (Figure 25D). This demonstrates how much more of the phosphoproteome could potentially be utilized by incorporating predicted relationships.

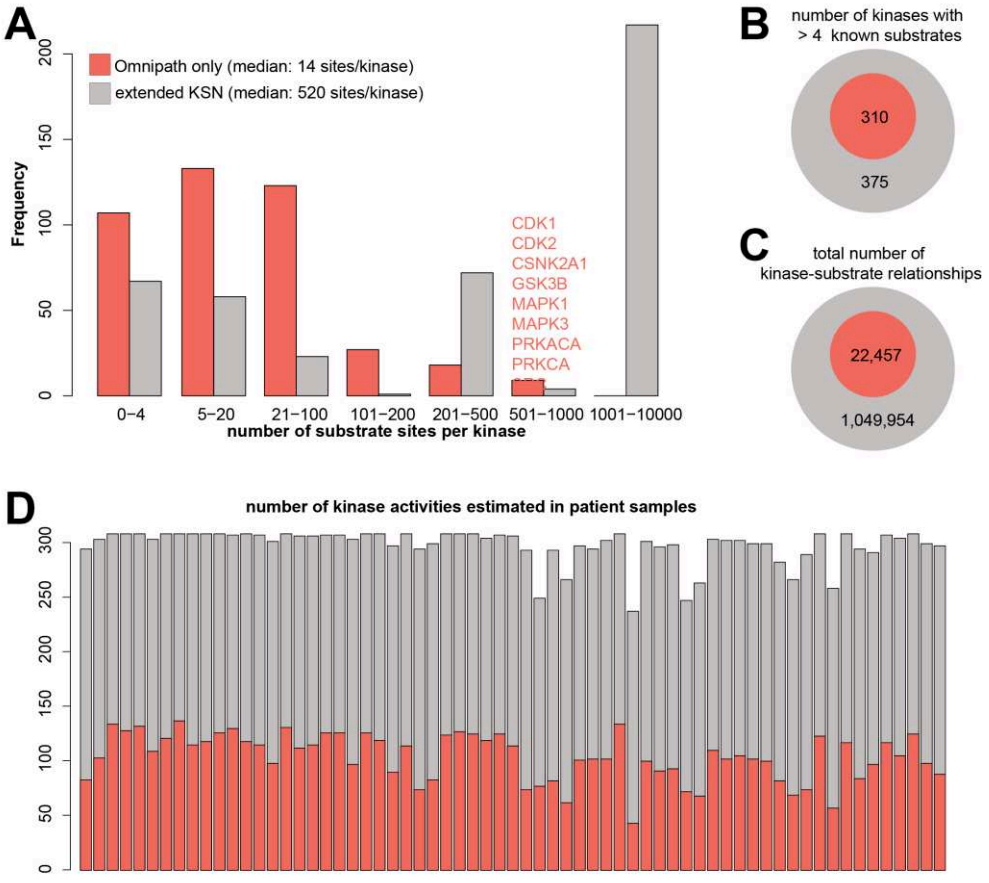


Figure 25: Comparison of the Omnipath-based and the extended kinase-substrate-network (KSN)
The Omnipath database of experimentally known kinase-substrate interactions was assessed via the OmnipathR package. I excluded substrates that were only known by ProtMapper, KEA, or PhosphoNetworks as the only source. The extended KSN was obtained from Garrido-Rodriguez et al. (2024) and was merged with any additional Omnipath entries. (A) Number of known phosphosite substrates per kinase. (B) Number of kinases with 5 or more annotated target phosphosites. (C) Total number of kinase-substrate interactions entailed in the databases. (D) Using the decoupleR package, I computed relative kinase activity scores for my 65 patient samples either based on the Omnipath KSN or the extended KSN. Shown are the numbers of kinases that an activity score could be computed for based on 5 or more substrates.

Of course, this extension comes with a caveat: there is no definitive way to validate whether kinase activities estimated from predicted interactions are truly accurate - but the same holds true for activities inferred from experimentally validated networks. All such estimations remain indirect,

and in both cases, the kinase activity represents a proxy, based on the behavior of known or presumed substrates.

As a next step, I sought to assess whether kinase activities derived from the extended KSN agree with those estimated using the classical Omnipath network, to better understand the reliability of the expanded predictions. To evaluate the agreement between the classical Omnipath-based kinase activity estimates and those derived from the extended KSN, I calculated Pearson correlation coefficients for both sets of activities for each individual patient. The results showed a positive correlation in all but one patient, with an average Pearson's r of 0.42, ranging from –0.07 to 0.72 (Figure 26A, left panel). While this indicates a moderate overall concordance, I wanted to explore whether certain types of kinases might contribute more strongly to this agreement.

To that end, I divided the kinases into two categories:

i. "Robust" kinases: those with at least 20 experimentally validated substrate sites (according to Omnipath) found in my dataset

ii. Less robust kinases: those with fewer than 10 Omnipath-based substrates found in my data set

This filtering revealed a clear pattern: The correlation between Omnipath-based and extended KSN-based activities was much higher for robust kinases, with a mean r of 0.54, compared to a mean r of only 0.18 for less robust kinases (Figure 26A, right panel). Assuming that better Omnipath coverage leads to a more correct activity estimate for a given kinase, this shows that the extended KSN-based activities more closely resembles these presumably correct activities.

To illustrate this further, I took a closer look at a single patient with a correlation close to the cohort average ($r = 0.42$). In this patient, I visualized the kinase activities estimated by both methods, with each dot representing a kinase and the color encoding the number of Omnipath substrates identified for that kinase in my dataset (Figure 26B). The plot revealed that most of the "robust" kinases (i.e., those with higher substrate coverage) clustered around the identity line, indicating good agreement between methods. In contrast, kinases with poor Omnipath coverage tended to deviate more strongly. This observation suggests that when Omnipath coverage is low, the kinase activity derived from the extended KSN may actually be more reliable, as it includes predicted substrates that expand the otherwise limited footprint. Conversely, when Omnipath coverage is high, both methods seem to agree well, likely reflecting the increased confidence in the activity estimation.

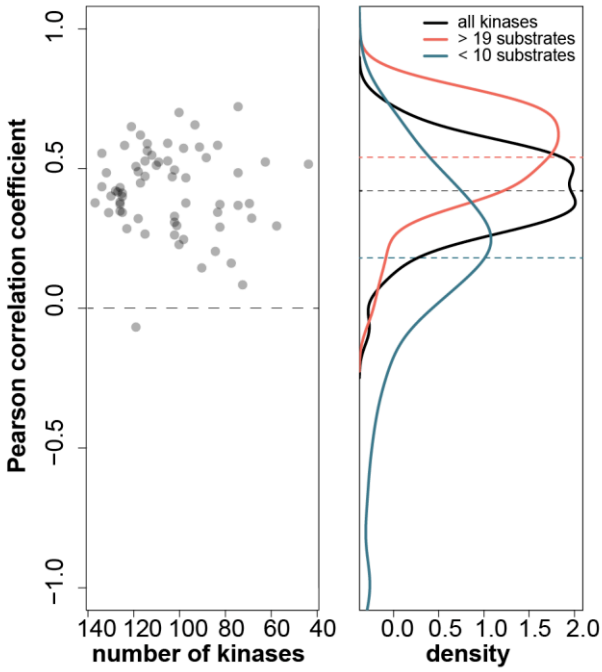


Figure 26: Correlation of classical and extended kinase activities
The Omnipath database of experimentally known kinase-substrate interactions was assessed via the OmnipathR package. I excluded substrates that were only known by ProtMapper, KEA, or PhosphoNetworks as the only source. The extended KSN was obtained from Garrido-Rodriguez et al. (2024) and was merged with any additional Omnipath entries. Relative kinase activities were computed by mapping the identified phosphosites to kinases using either the Omnipath database or the extended KSN and applying the run_wmean function of decoupleR to the z-scored phosphosite intensities of each patient. The resulting kinase activities for each patient based on Omnipath or the extended KSN, respectively, were correlated with each other. Shown are the Pearson correlation coefficients and the number of kinases for which an activity score was computed, each dot represents one patient. The distribution of Pearson correlation coefficients are shown for all kinases, kinases with at least 20 identified substrates, and kinases with less than 10 identified substrates.

In conclusion, although there is no definite way to determine which prior knowledge network best reflects the true kinase activities - and the answer may be context-specific - this analysis shows that the extended KSN can complement and potentially improve upon classical methods, particularly for understudied kinases. For this reason, I decided to include kinase activity estimates from both the classical Omnipath-based and the extended KSN-based networks in all downstream analyses of my patient cohort, while treating predictions based on the extended KSN with appropriate caution.

2.6 Comparison of the proteome with information from other molecular layers

Having established a robust phosphoproteome and proteome data set the next step was to explore how this layer of information relates to the other molecular profiles available for the same patients. Specifically, I aimed to compare the proteomic and phosphoproteomic data to genomic alterations (such as copy number variants and single nucleotide variants) and transcriptomic profiles (mRNA expression), with the goal of determining whether the proteome adds distinct or complementary insights. In other words**:** Does the proteome simply reflect upstream molecular events, or does it provide unique, additional biological information that might otherwise be missed? This question is especially relevant given the increasing appreciation of post-transcriptional and post-translational regulation in cancer biology, which may not be fully captured by DNA- or RNA-based approaches alone.

2.6.1 Are gene copy number alterations reflected in the transcriptome and proteome?

Copy number alterations (CNAs), including gains of oncogenes or deletions of tumor suppressor genes, are frequently used as biomarkers to guide treatment decisions, particularly in precision oncology programs. In my cohort, treatment recommendations based on CNAs were made for 23 out of 65 patients (35%). The most frequently considered alterations included *CDH17* gains (in six patients), which led to recommendations for a bispecific antibody therapy (BI 905711, within the clinical trial NCT04137289); *ERBB2* (HER2) gains (three patients), supporting the use of HER2-targeted therapies such as Trastuzumab; and *KRAS* gains (four patients), used to justify consideration of pathway-targeting agents like MEK inhibitors (e.g. Trametinib) or pan-RAF inhibitors (e.g. Sorafenib). On the deletion side, *CDKN2A* losses were used to recommend CDK4/6 inhibitors in two patients.

However, the assumption that gene copy number directly translates into functional protein output is not always valid. Between DNA and active protein lie many layers of regulation - transcriptional, post-transcriptional, translational, and post-translational - that can decouple genomic alterations from proteomic phenotypes (Gitton et al. 2002). Given how often CNAs are used as the basis for treatment decisions, I set out to investigate whether these alterations were consistently reflected at the mRNA and protein levels in my cohort. If a treatment recommendation hinges on a copy number gain, it is important to verify that this gain actually results in higher protein expression and, ideally, increased activity.

To extract the total copy number (TCN) levels of individual genes, I mapped genes to chromosome positions and for genes that were not in any of the chromosome regions with copy number alterations, I assumed that the copy number equaled 2. To assess whether there is any correlation between the gene copy number and the mRNA or protein expression, I used relative expression values by computing the log2 ratio of the protein intensity, or the mRNA TPM value, compared to

the median expression of all patients. Since for most genes the majority of patients have the normal two copies, I excluded these patients' values from the correlation analysis of a respective gene. Then, I computed the Spearman correlation coefficient (ρ) for any gene with at least 10 pairs of TCN and mRNA or protein expression, meaning that these genes had copy number alterations in at least 10 patients and mRNA or protein expression values were available.

The average TCN to mRNA correlation was 0.137 which was significantly higher than the average TCN to protein correlation ($\rho = 0.076$). This shows that neither mRNA nor protein expression appear to be affected much by the copy number status of the respective gene in my cohort (Figure 27A, B).

To better understand how individual gene copy number gains translate into actual changes in expression, I took a closer look at one patient with an exceptionally high number of affected genes—around 500 genes showed copy numbers between 4 and 6. I first analyzed how many of these copy number gains actually resulted in increased mRNA or protein levels. To define a meaningful expression change, I considered expression to "agree" with the copy number gain if the relative $\log_2$ mRNA or protein, respectively, expression ratio was greater than 1 (i.e., at least a two-fold increase compared to the cohort median), and to "disagree" if it was below -1. For transcripts or proteins that did not show any strong regulation in either direction ($-1 < \text{ratio} < 1$) I decided that I could not draw a definite conclusion regarding the agreement with the gene copy number level. At the mRNA level, 92 of the amplified genes showed increased expression, while 25 showed lower expression than expected (Figure 27C). At the protein level, 59 genes had elevated expression consistent with the copy number gain, and 16 showed lower-than-expected levels (Figure 27D). One notable is *SLC2A3*, which had a total of 5.19 gene copies and exhibited both very high mRNA expression ($\log_2$ ratio 3.81) and also markedly elevated protein levels ($\log_2$ ratio 6.44). On the other hand, 4.75 gene copies of *ECHDC3* were associated with a more than 2-fold higher mRNA expression (log2 ratio 1.24) but unchanged protein expression (log2 ratio -0.03) compared to the median of the cohort. Interestingly, only 21 genes overlapped between the mRNA and protein "agreeing" sets, suggesting that agreement at the transcript level does not always translate to the protein level. These findings underscore the importance of directly measuring protein expression: most of the copy number–protein expression agreements or disagreements would not have been detected based on mRNA alone. Thus, relying solely on transcriptomics may overlook important discrepancies that could influence treatment decisions.

Next, I investigated whether there might be a threshold effect, that is a minimum level of copy number increase required before changes in mRNA or protein expression become detectable. To do this, I grouped all genes across all patients into categories based on their total copy number: losses (<2), normal (2), moderate gains (2–4), high gains (4–6), and very high gains (>6). I then examined the distribution of relative expression values ($\log_2$ ratios) for each category. Both mRNA

and protein expression levels showed a gradual, though modest, increase with rising copy number. For protein expression, the median log_2 ratios progressed from –0.11 (copy number <2), –0.02 (copy number = 2), to 0.10, 0.17, and 0.23 for the moderate, high, and very high copy number groups, respectively (Figure 27E). A similar trend was observed at the mRNA level, with medians of –0.05, –0.03, 0.16, 0.21, and 0.37 (Figure 27F). While these increases are not dramatic, they suggest a cumulative effect, where higher copy number gains tend to result in proportionally higher expression, especially at the transcript level.

Together, these results suggest that while CNAs can contribute to changes in gene expression, the effect is often subtle and variable, and additional regulatory layers may modulate the final protein abundance. This reinforces the need to integrate multiple data layers - genomic, transcriptomic, and proteomic - to comprehensively understand the molecular state of a tumor and to guide rational treatment decisions.

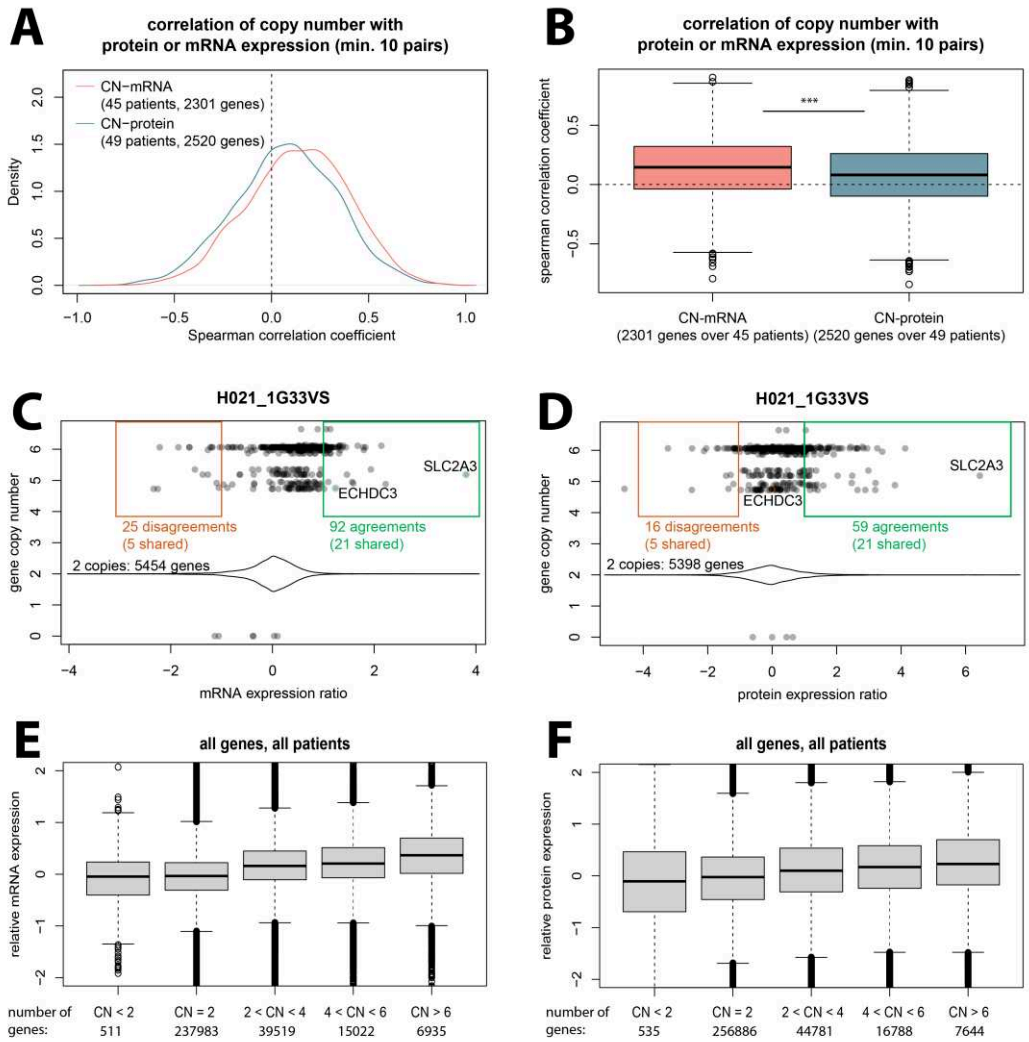


Figure 27: Reflection of copy number alterations on gene expression and protein abundance

CNV data were available for 49 patients. Genes were mapped to chromosome positions using GenomicRanges R package (v 1.54.1) and Homo.sapiens R package (v 1.3.1) and all TCN (total copy number) values were extracted. For genes that were not in any chromosome region with altered copy number, I assumed CN = 2. (A) Spearman correlation coefficients were calculated for each gene where at least 10 pairs of copy number and mRNA TPM resp. protein log2 expression values were available. (B) Distribution of Spearman correlation coefficients. Significance between CN~mRNA and CN~protein correlation coefficients was assessed using two-sides, unpaired Student's t-test. *** = p-value < 0.001. (C) Relative mRNA expression of example patient 1G33VS and copy number levels of all genes. Highlighted are genes for which the mRNA expression agrees with the increased copy number, i.e. is higher than 1, and disagrees, i.e. is lower than -1. (D) Relative protein intensities of example patient 1G33VS and copy number levels of all genes. Highlighted are genes for which the protein abundance agrees with the increased copy number, i.e. is higher than 1, and disagrees, i.e. is lower than -1. (E, F) Genes were grouped into copy number losses (<2), normal two copies, moderate copy number gain (2<CN<4), high copy number gain (4<CN<6) and very high copy numbers (>6). The distribution of relative mRNA (E) and protein (F) expression of all genes and all patients are illustrated by boxplots.

To further explore how copy number alterations translate into expression changes at the mRNA and protein level, I examined three genes whose copy number gains were most frequently used by the molecular tumor board to guide treatment decisions: *CDH17*, *KRAS*, and *ERBB2*. For each gene, I plotted the total copy number against the relative mRNA and protein expression for patients with the respective data available (45 patients for total copy number and mRNA, 49 for total copy number and protein). Patients who received a treatment recommendation based on copy number gain of that specific gene were highlighted in green (Figure 28).

For all three genes, Spearman correlation coefficients were calculated using only patients whose copy number differed from the normal value of 2, to better assess whether deviations in gene dosage were associated with changes in expression. Overall, mRNA expression showed a stronger correlation with copy number than protein expression, reinforcing earlier observations that genomic alterations tend to be more immediately reflected at the transcript level, while additional regulatory layers may obscure the signal at the protein level.

Looking more closely at *CDH17*, a gene frequently considered for treatment with a bispecific antibody (BI 905711), the pattern was particularly interesting (Figure 28A). Despite several patients showing elevated copy numbers, treatment recommendations were not always aligned with the highest gene dosage. Among the five patients who received CDH17-targeted treatment, four had elevated mRNA levels, but only two showed elevated protein expression. One striking example is patient Q94AP2, who had six copies of *CDH17* and slightly elevated mRNA levels, yet displayed downregulated protein expression compared to the cohort. This discrepancy raises questions about the actual utility of CDH17 as a biomarker in this case, and highlights the importance of protein-level data when evaluating targetable alterations.

In contrast, KRAS showed relatively high agreement between mRNA and protein levels (Figure 28B). Among the four patients who received treatment recommendations based on *KRAS* amplification, two (NGNEWN and 2EHFSJ) had strongly elevated expression at both the mRNA and protein level, while the other two (FKZGTD and VMXJTQ) showed neither. Notably, only the patients with extreme copy number gains - 26 and 50 gene copies - had clearly elevated expression, suggesting that for KRAS, a copy number threshold must be surpassed before any increase in expression is detectable.

Finally, ERBB2 presented yet another layer of complexity (Figure 28C). While mRNA levels were elevated in two of the three patients who were recommended HER2-targeted therapy (Q94AP2 and FDBGTP), the corresponding protein abundance did not confirm these findings. In fact, HER2 protein was not even detected in Q94AP2, the patient with both the highest copy number (13.3) and highest mRNA expression. While absence of detection does not confirm absence of the protein, it does indicate that HER2 levels were relatively low compared to the rest of the cohort. Similarly, FDBGTP, who had the second highest *ERBB2* copy number (7.3), showed only a modest protein

expression ratio (0.13). These findings emphasize the disconnect that can occur between DNA copy number, mRNA expression, and final protein abundance, and they serve as a clear reminder that copy number alterations alone are not always sufficient indicators of functional target presence.

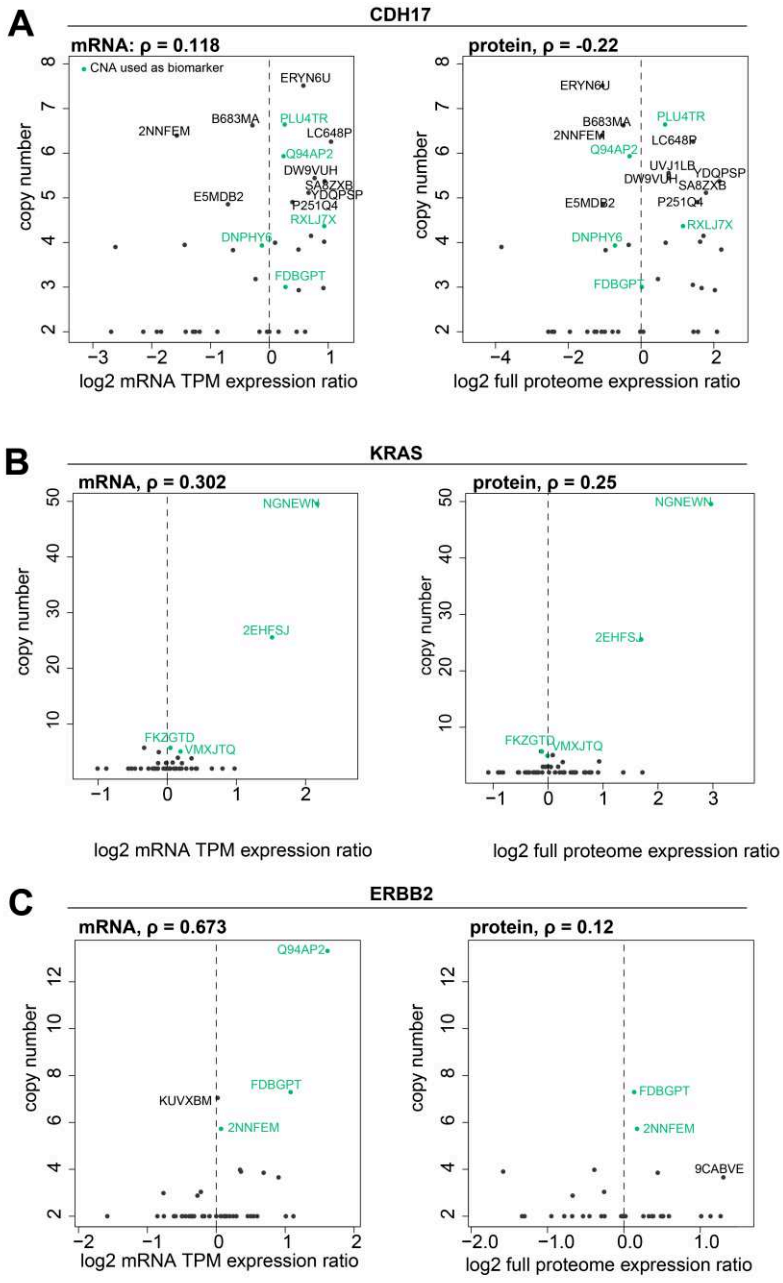


Figure 28: Most frequently used copy number gain biomarkers
CNV data were available for 49 patients. Genes were mapped to chromosome positions using GenomicRanges R package (v 1.54.1) and Homo.sapiens R package (v 1.3.1) and all TCN (total copy number) values were extracted. For genes that were not in any chromosome region with altered copy number, I assumed CN = 2. Expression ratios were calculated by dividing the mRNA TPM or protein intensity value of each patient by the median value over all patients, follower bx log2 transformation. Copy numbers compared to relative mRNA and protein expression of the three genes most commonly used as a biomarker when copy number gains are detected: CDH17 (A), KRAS (B), and ERBB2 (C). Spearman correlation coefficients (ρ) are based only on patients in which the gene copy number was not 2.

Overall, these examples underline the critical value of integrating protein-level data - especially in clinical contexts where treatment recommendations are based on assumptions about gene dosage effects. Having established that protein-level measurements provide valuable insights beyond what is captured by DNA copy number alone, I next examined the relationship between mRNA and protein expression in order to determine whether proteomics adds additional layers of biological information that are not accessible through transcriptomic data.

2.6.2 Correlation of mRNA and protein expression

To compare mRNA with protein expression levels, I first standardized the expression data by z-scoring the $\log_2$-transformed protein intensities and mRNA TPM values, respectively. I focused on genes that were detected at both the mRNA and protein level in at least ten patients - resulting in a final set of 7,820 genes across 57 patients.

High mRNA expression was frequently used by the molecular tumor board to recommend treatment with antibody-drug conjugates targeting the respective protein. For example, high Trophoblast cell surface antigen-2 (TROP2) mRNA levels were used as a biomarker to recommend enrolment of the patient in an ongoing clinical trial testing treatment with Sacituzumab Govitecan, an anti-TROP2 ADC (NCT06243393; Figure 29). While proteomic measurement confirmed the high TROP2 expression in these two patients, MKVVUC and DNPHY6, it also revealed two more patients, JNFAWJ and QH61HA, with relatively high TROP2 protein expression that was not detected on mRNA level. Potentially, these two patients could have been recruited for the TROP2 ADC clinical trial.

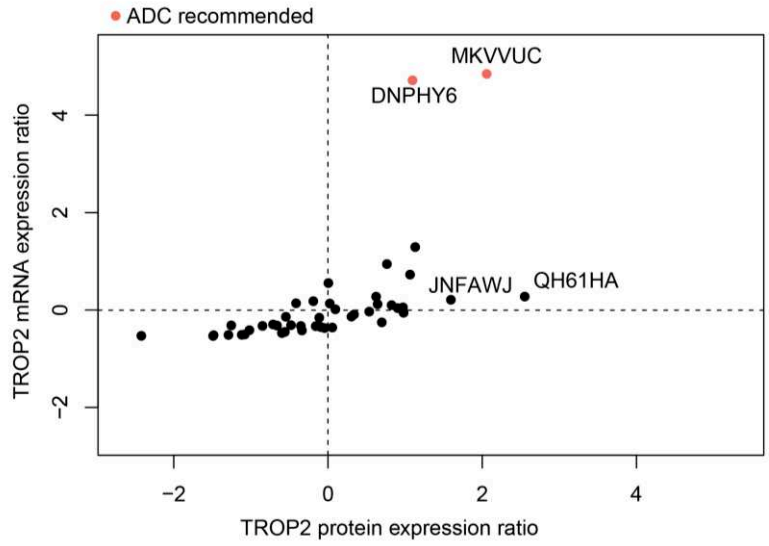


Figure 29: TROP2 expression
Protein log2 intensities and mRNA TPM values were each z-scaled over all patients and TROP2 values were plotted.

To assess the global mRNA-protein-correlation in my data set, I next computed the Spearman correlation coefficient for each gene across patients to quantify the strength of association between its mRNA and protein expression levels. The correlations ranged widely, from –0.57 to 0.95, with a mean correlation of 0.45 (Figure 30A). This variability reflects the complex and gene-specific nature of the relationship between transcription and translation, suggesting that for many genes, protein levels are not solely dictated by mRNA abundance, but instead are influenced by post-transcriptional and post-translational regulatory mechanisms.

To further explore the biological processes associated with different levels of mRNA-protein concordance, I grouped genes into four categories based on their correlation values: those with negative correlation, those with low correlation (ρ between 0 and 0.2), moderate correlation ($0.2 < \rho < 0.7$), and those with high correlation ($\rho > 0.7$). I then performed gene set overrepresentation analysis using the ReactomePA package, applying Benjamini–Hochberg correction for multiple testing and an adjusted p-value cut-off of 0.05 (Figure 30A).

This analysis revealed distinct biological themes within each correlation group. Genes with high mRNA-protein correlation were enriched in pathways related to extracellular matrix organization and fatty acid metabolism (Figure 30B), suggesting that for these functions, protein expression may be largely driven by transcriptional regulation. In contrast, genes with low correlation were overrepresented in pathways such as cellular response to starvation and various translation-related processes (Figure 30C), where post-transcriptional mechanisms are likely to play a more dominant regulatory role. Intriguingly, some gene sets were significantly enriched even among genes with negative correlation between mRNA and protein levels. Notably, this included respiratory electron transport and the TCA cycle (Figure 30D), which may reflect tightly regulated metabolic processes where protein stability, localization, or activity are finely tuned and not directly linked to transcriptional output.

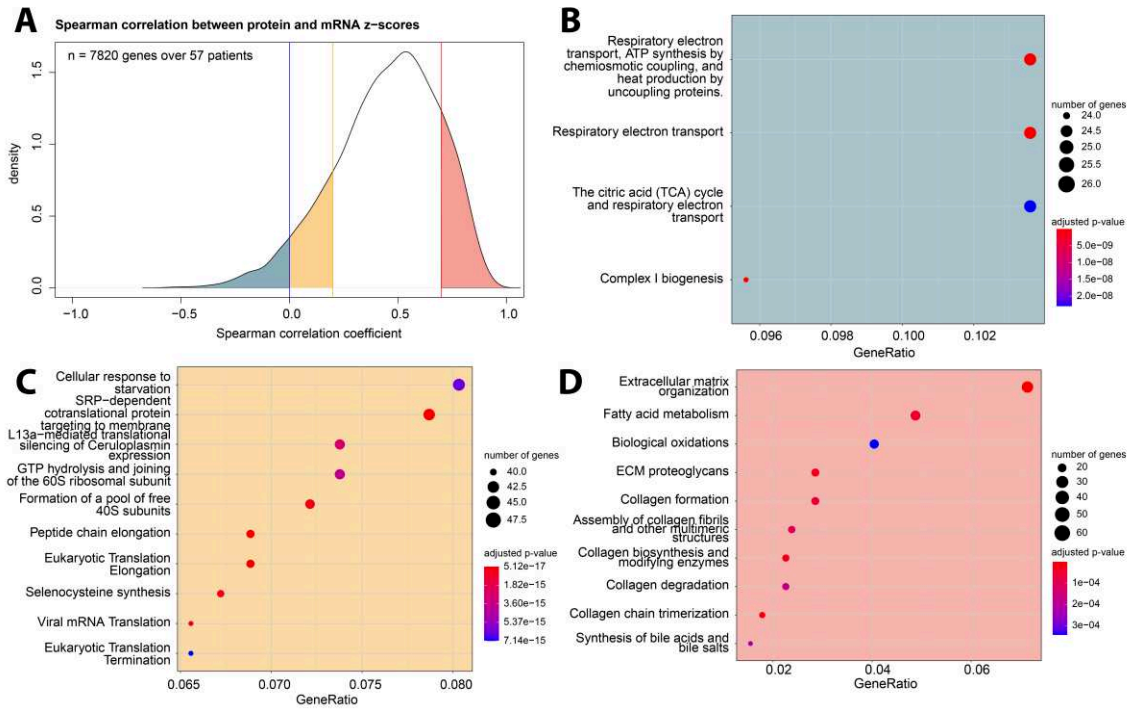


Figure 30: Correlation of mRNA expression and protein abundance
Protein log2 intensities and mRNA TPM values were each z-scaled over all patients. For 57 patients, both mRNA sequencing and proteome data were available. Only genes were kept for which both mRNA and protein values were present in at least 10 patients and Spearman correlation coefficients (ρ) were computed for each transcript/protein. (A) Distribution of Spearman correlation coefficients of 7820 genes for which both mRNA and protein values were present in at least 10 patients. The cut-offs for highly correlating genes ($\rho > 0.7$), lowly correlating genes ($0 < \rho < 0.2$) and negatively correlating genes ($\rho < 0$) are highlighted. (B, C, D) Overrepresentation analysis of Reactome pathways was performed using ReactomePA package, applying Benjamini–Hochberg correaction and a p-value cut-off of 0.05. The top 10 Reactome pathways overrepresented in the negatively correlating (B), lowly correlating (C), and highly correlating (D) genes are shown.

These findings underscore the importance of integrating proteomic data into multi-omics analyses: while mRNA levels offer a proxy for gene activity, protein expression captures a different and complementary dimension of cellular regulation - one that is particularly critical in contexts where post-transcriptional control mechanisms are at play.

While understanding protein abundance is valuable, what we are ultimately most interested in is not just how much protein is present, but how active it is. This is particularly relevant for kinases, which are among the most frequent drug targets in oncology (Latham et al. 2024). Since kinase activity cannot be measured directly in tissue samples, I relied on the phosphoproteome as a functional footprint, as described earlier, to infer protein activity for kinases.

2.6.3 Does kinase activity correlate with protein expression?

To explore the relationship between protein expression and protein activity, I examined the epidermal growth factor receptor (EGFR), a receptor tyrosine kinase frequently altered in cancer and a prominent drug target. For each patient, I compared EGFR gene copy number, mRNA expression, protein abundance, and inferred kinase activity, plotting each pairwise combination to assess how these layers of information relate to each other (Figure 31).

The results show a high correlation between mRNA and protein expression (Spearman's $\rho = 0.71$), suggesting that transcription is a key determinant of EGFR protein levels in this cohort. However, correlations between copy number and mRNA ($\rho = 0.175$) or copy number and protein ($\rho = –0.13$) were weak or even negative, consistent with my earlier findings (2.6.1) that copy number alterations do not reliably predict downstream expression.

Most strikingly, there was almost no correlation between EGFR protein expression and its inferred activity ($\rho = 0.025$). This illustrates that protein abundance alone is not a good proxy for activity - a critical insight when considering treatment decisions. EGFR activity depends on multiple additional layers of regulation, including post-translational modifications (e.g., phosphorylation), subcellular localization, and interaction with binding partners or ligands. The classical model of EGFR activation begins with ligand binding, most notably epidermal growth factor (EGF), which triggers a cascade of conformational changes leading to receptor dimerization and subsequent kinase activation (Lemmon et al. 2014). This activation is transient and regulated by several mechanisms, such feedback inhibition (Gotoh 2009; Avraham and Yarden 2011), and receptor internalization (Gorden et al. 1978) followed by either lysosomal degradation (Miller et al. 1986) or recycling of membrane-standing EGFR (Sorkin et al. 1991).

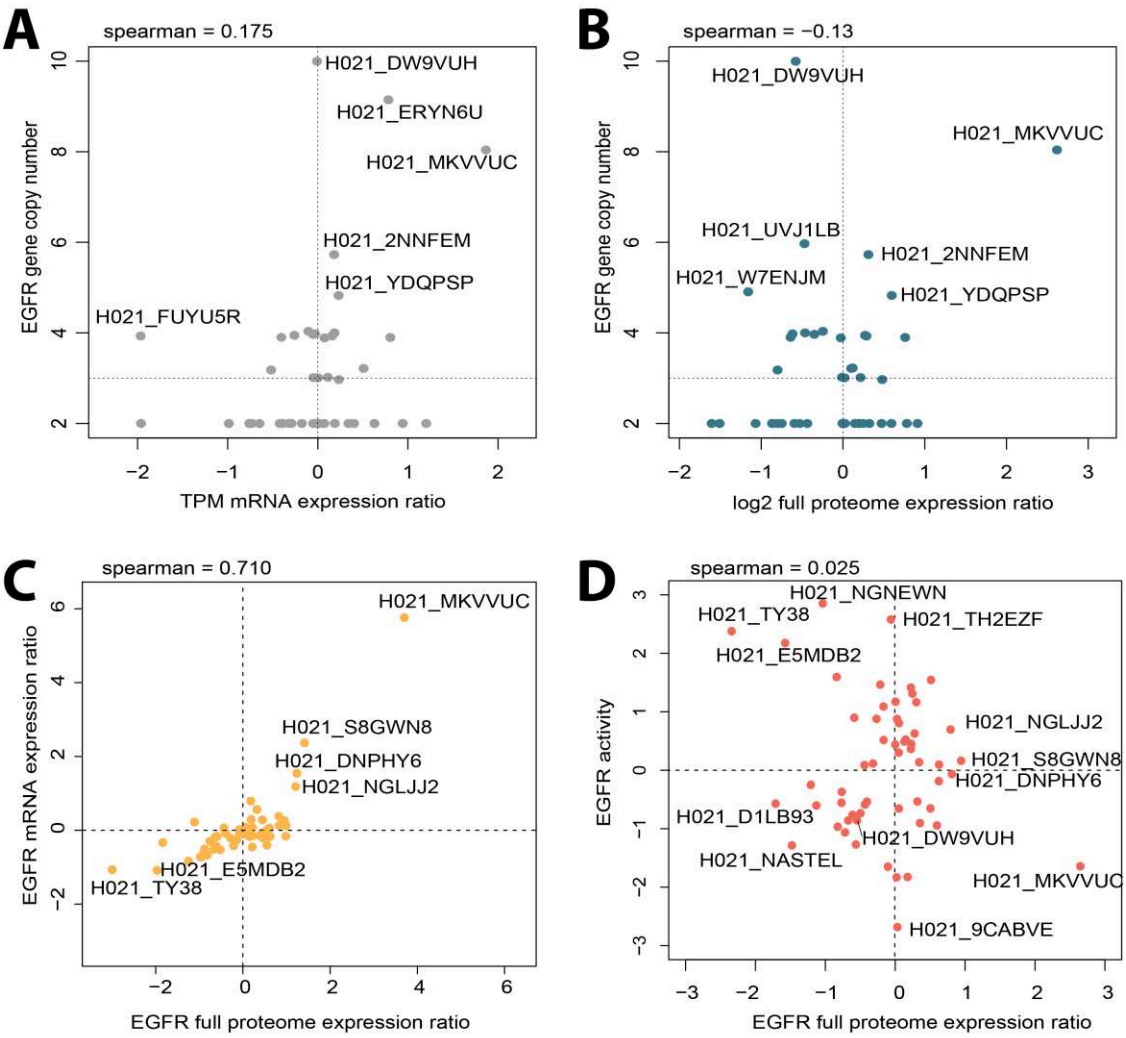


Figure 31: EGFR as an example
CNV data were available for 49 patients. Genes were mapped to chromosome positions using GenomicRanges R package (v 1.54.1) and Homo.sapiens R package (v 1.3.1) and all TCN (total copy number) values were extracted. For genes that were not in any chromosome region with altered copy number, I assumed CN = 2. Expression ratios were calculated by dividing the mRNA TPM or protein intensity value of each patient by the median value over all patients, follower bx log2 transformation. The phosphoproteome data were filtered to remove phosphosites that were identified in fewer than 20 patients. Subsequently, the phosphosite intensities were z-scored across all patients. Protein (kinase) activities were computed based on the extended KSN (see 2.5.3) and applying the run_wmean function of decoupleR to the z-scored phosphosite intensities of each patient. Only kinases with at least 5 identified substrates were included in the analysis. (A) EGFR mRNA expression compared to EGFR gene copy numbers, (B) EGFR protein expression compared to EGFR gene copy numbers, (C) EGFR protein expression compared to EGFR mRNA expression, (D) EGFR protein expression compared to EGFR kinase activity.

This case underscores the value of including functional readouts such as kinase activity in clinical proteomics, as they can reveal regulatory dynamics and signaling states that are invisible to transcriptomics or proteomics alone.

Finally, I illustrated the correlations between all investigated molecular levels (Figure 32) demonstrating that only mRNA and protein levels show relatively high correlations, while all other levels appear to be rather disconnected.

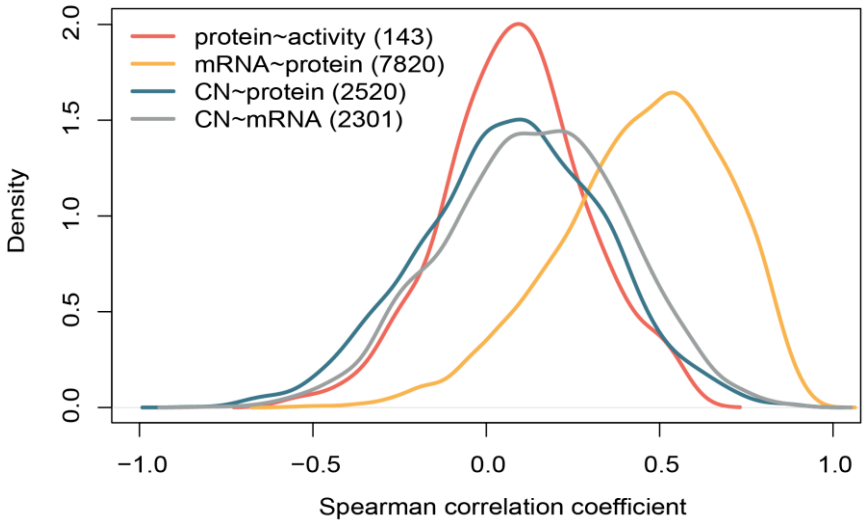


Figure 32: correlations of all layers
CNV data were available for 49 patients. Genes were mapped to chromosome positions using GenomicRanges R package (v 1.54.1) and Homo.sapiens R package (v 1.3.1) and all TCN (total copy number) values were extracted. For genes that were not in any chromosome region with altered copy number, I assumed CN = 2. Expression ratios were calculated by dividing the mRNA TPM or protein intensity value of each patient by the median value over all patients, follower bx log2 transformation. Protein (kinase) activities were computed based on the extended KSN and applying the run_wmean function of decoupleR to the z-scored phosphosite intensities of each patient. Density of correlations between protein and protein activity (65 patients), mRNA and protein (57 patients), copy number and protein (49 patients), and copy number and mRNA (45 patients). The numbers of genes/transcripts/proteins for which at least 10 values were found in both compared molecular levels are noted in brackets.

Having established that the molecular layers of protein abundance and activity does add additional information that is complementary to genetic and transcriptomic data, I next characterized the proteome of this patient cohort.

2.7 Kinase activity-based patient clusters

2.7.1 Identification of potential kinase activity-based patient subgroups

As mentioned previously, the patient cohort used in my study is clinically and biologically unique - composed of younger individuals with end-stage disease, many of whom had undergone extensive treatment and whose samples were primarily obtained from metastatic lesions (Figure 20). In contrast, publicly available CRC molecular data sets have been obtained mainly from treatment-naïve primary tumors.

For example, the Cancer Genome Atlas (TCGA) datasets for colon adenocarcinoma (COAD) and rectal adenocarcinoma (READ) provide genomic and transcriptomic sequencing data from 459 and 167 patient samples, respectively. The vast majority of these samples were obtained from primary tumors. These tissues are collected at the time of initial surgical resection or biopsy, prior to any systemic therapy or radiation, to ensure that the molecular profiles reflect the untreated disease state (https://portal.gdc.cancer.gov/projects/TCGA-COAD; https://portal.gdc.cancer.gov/projects/TCGA-READ; accessed 17.06.2025).

Similarly, the Clinical Proteomic Tumor Analysis Consortium (CPTAC) colorectal cancer cohorts are composed of samples exclusively from treatment-naïve patients and do not include samples from metastatic sites (Vasaikar et al. 2019).

These differences in the cohorts raised the question: Does my patient cohort represent a fundamentally different class of colorectal cancer (CRC) than what is typically captured in more classical, treatment-naïve datasets?

To begin addressing this, I evaluated whether my cohort aligned with the Consensus Molecular Subtypes (CMS) - a widely used classification system for CRC based on RNA-seq data from treatment-naïve, primary tumor samples. The CMS framework, developed using such classical cohorts, defines four transcriptional subtypes with prognostic and potential therapeutic relevance (Guinney et al. 2015). I obtained full proteome data and clinical metadata, including CMS classification, from the CPTAC CRC proteogenomic study (Vasaikar et al., 2019), which includes 97 primary, untreated tumors. As expected, when visualized using principal component analysis, the CPTAC samples clustered according to their CMS assignment, confirming the stability of these subtypes on the proteome level in classical settings (Figure 33A). Next, I applied the CMScaller R package to assign CMS labels to my patient samples using their RNA-seq profiles. However, unlike the CPTAC cohort, my samples did not cluster according to CMS subtype (Figure 33B). This suggests that the transcriptional landscape of advanced, heavily treated, and metastatic CRC tumors may no longer reflect the canonical CMS structure.

The CMS classification faces significant challenges in heavily pre-treated, metastatic CRC cases due to substantial alterations in the tumor microenvironment (TME) induced by chemotherapy.

Single-cell analysis reveals that preoperative chemotherapy causes profound reprogramming of the TME, including activation of B cells, lower diversity of tumor-associated macrophages with immature phenotypes, and accumulation of myofibroblasts (Che et al. 2021). The reliability of CMS classification is further compromised by organ-specific changes in metastatic sites. Research demonstrates that only 64% of CRC metastases exhibit concordant CMS groups with matched primary tumors. Liver metastases show predominant CMS2 classification, while lung and peritoneal metastases are mainly CMS4, suggesting that the "seed and soil" hypothesis influences CMS distribution in metastatic sites (Luo et al. 2024).

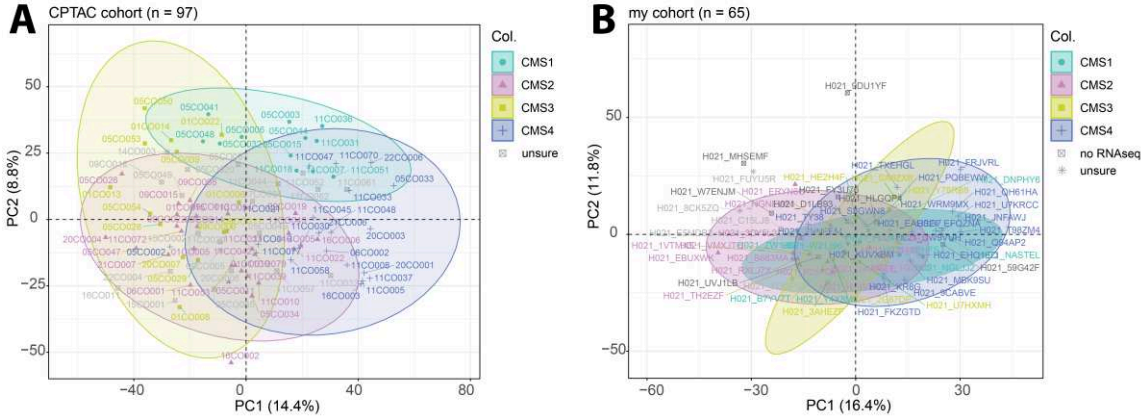


Figure 33: Full proteome PCA of CPTAC and my dataset. Color based on CMS subtype. (A) CPTAC full proteome data set was obtained from LinkedOmics (https://www.linkedomics.org/cptac-colon/) and CMS assignment was obtained from the supplementary table 1 of Vasaikar et al. (2019). (B) CMScaller R package was used to predict CMS from bulk RNA sequencing data (1000 permutations, FDR cut-off 0.05). For eight patients, RNA sequencing data were not available. PCA was performed on the proteome data using the prcomp R function and visualized using the factoextra R package, color-coded by CMS.

Taken together, these results support the hypothesis that the classical CMS framework may not apply to end-stage, pre-treated CRC, and that such patients may form distinct molecular subgroups requiring different approaches to classification and potentially therapy. While PDOs likely lose tissue characteristics during cultivation, the unique cohort captured in my CRC patient data set might in part explain the separate PCA clustering (Figure 48). This observation led me to perform a more data-driven exploration of patient subtypes, independent of existing classification systems.

When correlating the relative kinase activities of single patients within my cohort, I noticed the formation of three clusters of patients with highly similar kinase activity landscapes (Figure 34A). The first patient group (cluster 1) was characterized by high activities of cyclin-dependent kinases (CDKs), casein kinase 2 (CSNK2), and GSK3b. Cluster 2 was characterized by high activities of, among others, PRKC, PRKA, AURKB and ROCK1/2. The third patient group presents with high activities of several MAPK members, AKT1 and RPS6KA (Figure 34B). At first glance, the

clusters cannot be explained by the location of the primary tumor (ICD10), CMS, or sex of the patient. However, all available categorical and numerical metadata will be tested systematically in the following paragraph.

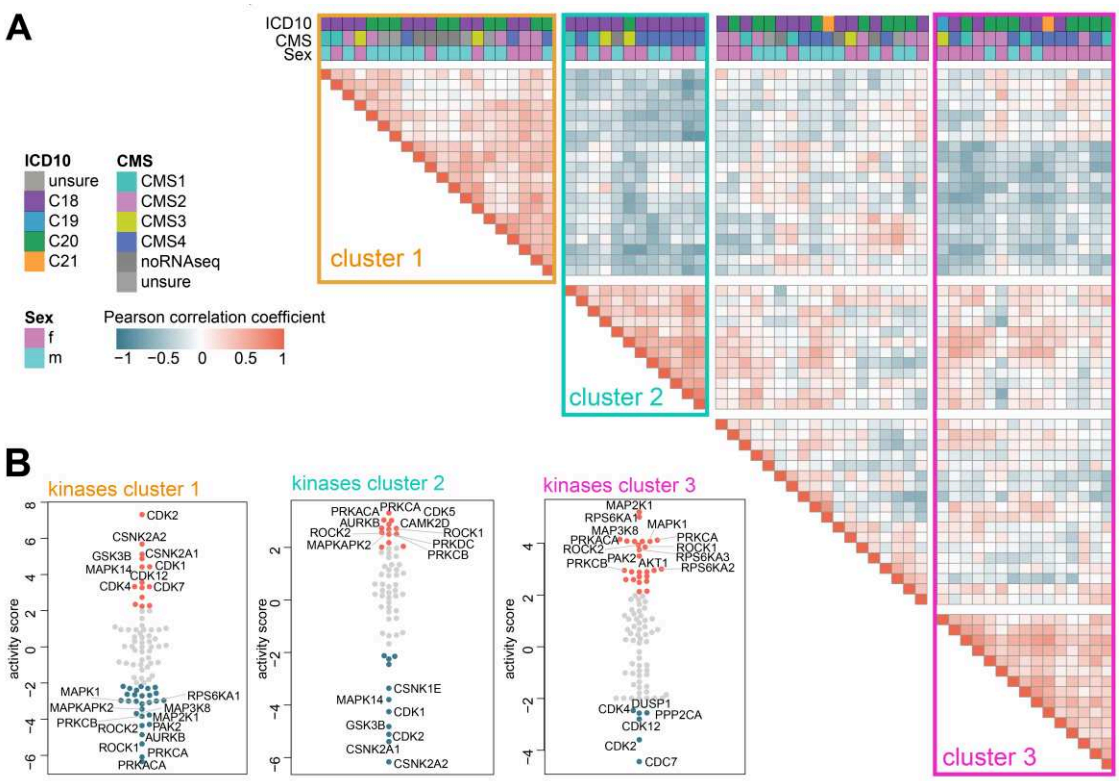


Figure 34: Identification of three patient subgroups with similar kinase activity landscapes. (A) Pearson correlation of patients' kinase activities with each other. Pearson correlation coefficients were visualized using pheatmap R package with unsupervised hierarchical clustering (k = 4). Three highly correlating clusters of patients are highlighted. (B) Kinase activities of patients within one the respective three clusters compared against all other patients were calculated using the Omnipath database and the t-statistics resulting from eBayes statistics of the limma R package (Ritchie et al. 2015), comparing the phosphopeptide intensities of the two respective patient groups (e.g. cluster 1 vs. all other). Kinases with an absolute activity score > 2 are highlighted.

2.7.2 Characterization of kinase activity-based patient subgroups

To characterize the three kinase activity-based patient subgroups on the molecular level, I determined transcription factor and pathway activities based on the mRNA sequencing data. Briefly, I compared the gene expression between each of the three clusters with all other patients by applying limma statistics (Ritchie et al. 2015). The resulting t-statistics were then used by the decoupleR package (Badia-I-Mompel et al. 2022) to compute relative transcription factor and pathway activities based on the CollecTRI (Müller-Dott et al. 2023) and PROGENy (Schubert et al. 2018) databases, respectively. I similarly used limma to assess differential protein and phosphosite abundance between any of the respective clusters and all other patients. For the full

proteome, I then performed GSEA of Reactome pathways based on the sorted log2 fold-changes. Relative kinase activities were computed based on the t-statistics using decoupleR and the Omnipath database (Türei et al. 2021).

Cluster 1 displayed a proliferation-driven molecular phenotype. Patients in this group exhibited high transcription factor activities of the E2F family (E2F1–4) and MYC, all of which are known to regulate genes involved in cell cycle progression and replication. Concordantly, Reactome pathway enrichment based on differentially abundant proteins highlighted cell cycle, transcriptional, and translational machinery as significantly upregulated. Kinase activity estimates further supported this phenotype, with elevated activities of CDK1, CDK2, and CDK4—central regulators of cell cycle transitions (Figure 35A). This cluster could potentially represent tumors with more aggressive proliferative behavior, though clinical correlation remains limited due to incomplete follow-up data.

Cluster 2 was characterized by high TGFβ signaling activity, particularly through SMAD2/3/4 transcription factors, as well as increased activity of TP53, suggesting a possibly more differentiated or stress-responsive tumor state. Interestingly, this group also had fewer copy number alterations, which may point to a more genomically stable profile (Figure 35C). Although not statistically significant, patients in this cluster were also less likely to have received radiotherapy (Figure 35C), possibly reflecting differences in clinical presentation or disease progression (Fisher's exact $p = 0.22$). Furthermore, there was a higher proportion of colon cancer (C18) samples – rather than rectal cancer (C20) – in cluster 2 patients (Figure 35E). Since colon cancer is less frequently treated with radiation (American Cancer Society), this could explain why fewer patients in this group received radiation, which in turn could explain the higher genomic stability.

Cluster 3 exhibited a clearly immune-infiltrated and inflammation-associated phenotype. Transcription factors such as IRF1, IRF4, NFκB, and REL were more active, alongside the JAK-STAT, TNFα and NF-κB pathways. These signatures were supported by protein-level enrichment for immune-related Reactome pathways (adaptive immunity, cytokine signaling, interleukin and interferon signaling). Notably, this group also showed the highest average ESTIMATE immune score (Yoshihara et al. 2013), indicating higher inferred immune cell infiltration across the tumor samples (Figure 35B).

To investigate which other factors characterize these three patient groups, aside from their kinase activity landscapes, I had a look at numerical metadata, including age, estimated tumor purity, total duration of treatment with oxaliplatin, 5-FU and irinotecan, survival duration since MTB, and MSI sensor value. Differences between groups were tested using Kruskal-Wallis test. There was no significant difference in the total duration of chemotherapy treatment (Figure 35D), as well as the age of patients (Figure 35F). There was also no difference in the number of days the patients

survived after the MTB took place. However, for more than half of the patients this follow-up information was not available. Especially for cluster 1 that contains patients with potentially more aggressive tumors, a complete follow-up would have been interesting. The only significant numerical information was tumor purity with higher values in tissue samples from cluster 1 (Figure 35F). This suggests that there are fewer other cell types, such as immune or stroma cells, present in these tissue samples which agrees with the immune infiltration estimate (Figure 35B).

Available categorical metadata included sex, whether the tissue sample was obtained from the primary tumor (T) or a metastatic site (M), location of the primary tumor (ICD10), batch, CMS, MSI status, whether the patient received radiation, anti-VEGF, ICI, or anti-FGFR therapy, KRAS mutation status, and classification into high tumor mutational burden (TMB), BRCAness and homologous repair deficiency (HRD). Differences between groups were tested using Chi-Square test or Fisher's exact test for variables with less than 5 values in any group. Categorical metadata that were found to differ significantly between the kinase activity-based clusters included batch, BRCAness, and ICD10 location (Figure 35E). Since all cluster 2 patient samples were received in batch 2, this cluster could potentially represent a batch effect artefact rather than a biologically distinct group of patients. However, colon cancer was overall more frequent in batch 2 (71% of cases) compared to batch 1 (35% of cases) which could explain the apparent batch effect.

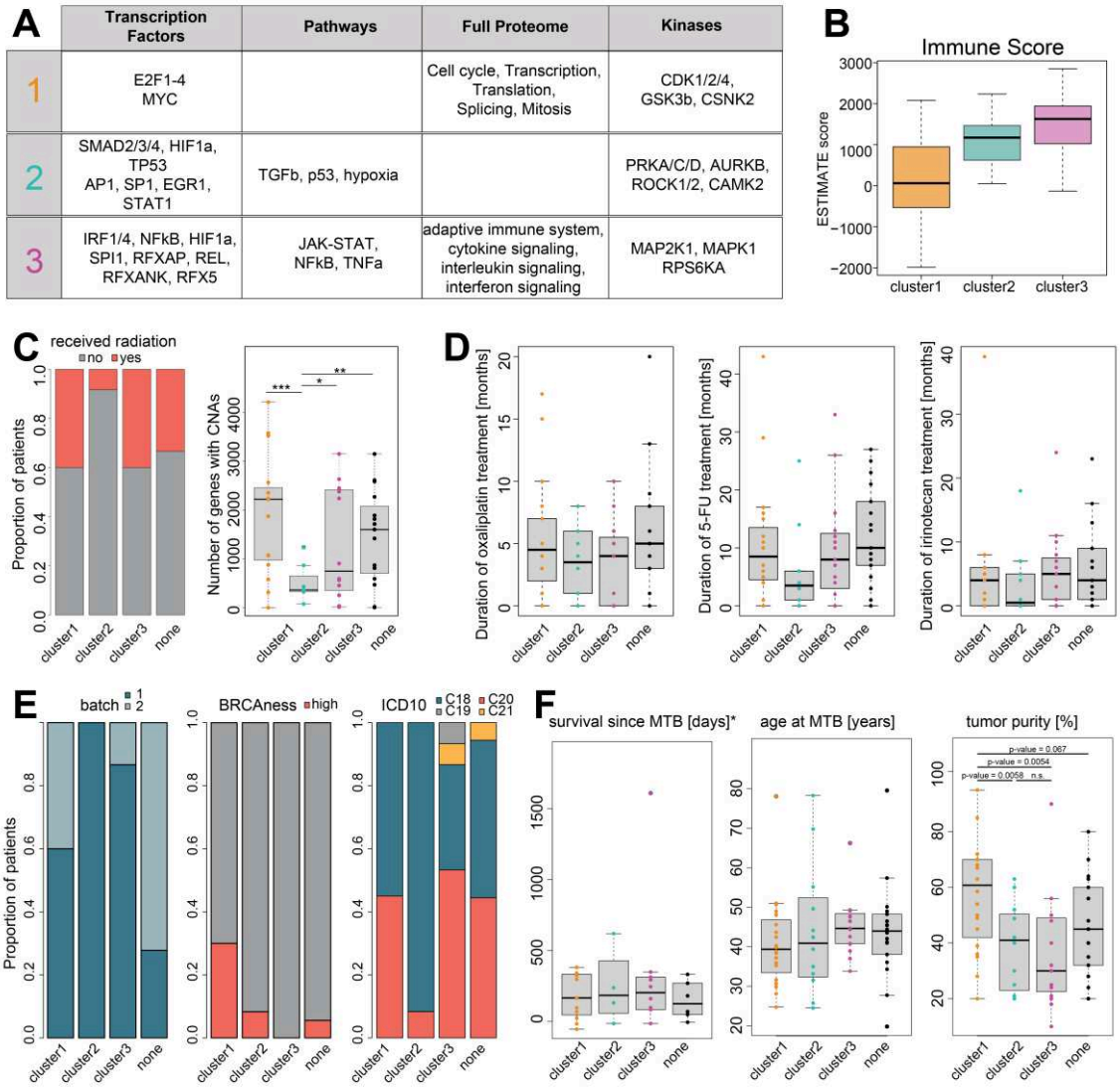


Figure 35 Characterization of kinase activity-based patient subgroups
(A) Summary of findings from TF and pathway activities, Reactome GSEA, and kinase activities (each cluster against all other patients). (B) ESTIMATE immune score obtained from immunedeconv R package applied to bulk RNA sequencing data (C) Proportion of patients that received radtioation. Number of genes with copy number alterations (duplication or deletions) per patient. (D) Normality of continuous metadata was assessed using Shapiro-Wilk test. Duration of chemotherapy pre-treatments showed non-normal distribution and differences between groups were therefore tested using Kruskal-Wallis test. No significant difference was found between any groups. (E) Statistical differences in the proportion of categorical metadata was asessed using Chi-Square test. If the expected frequency in any group was too low for Chi-Square test (<5), Fisher's Exact Test was used istead. Distribution of Batch, BRCAness, and ICD10 were found to be signficantly different between the clusters ($p < 0.05$). (F) Normality of continuous metadata was assessed using Shapiro-Wilk test. Most of the metadata showed non-normal distribution and differences between groups were therefore tested using Kruskal-Wallis test. Distribution of number of days survived since MTB (*Out of the 20, 12, 15, and 18 patients in each cluster 9, 8, 5, and 12 patients have no follow-up), age at MTB, and estimated tumor purity are shown. Significance of tumor purity was followed up with two-sided Student's t-tests between the groups.

The present data set represents the first proteome and phosphoproteome analysis of young, end-stage, metastatic, heavily pre-treated colorectal cancer patients. I revealed three potential subgroups based on kinase activities that might have implications for therapy recommendations. Next, I aimed to integrate the different data on a per-patient basis and evaluate the treatment recommendations given by the NCT-MASTER molecular tumor board.

2.8 Per-patient combination of different molecular information and evaluation of MTB treatment recommendation in the light of proteomic data

To illustrate the benefit proteomic data can likely have in clinical decision-making, I selected one representative example patient from the cohort and examined how the proteomic data aligned with the treatment recommendations proposed by the NCT-MASTER molecular tumor board (MTB). I selected this specific patient because they received several different treatment recommendations based on different potentially actionable molecular biomarkers, including a KRAS mutation, high mutational burden, and homologous recombination deficiency.

The patient was a 49-year-old female diagnosed with metastatic colorectal cancer. RNA-seq-based consensus molecular subtyping using the CMSCaller R package (Eide et al. 2017) classified the tumor as CMS1. The tumor sample was taken from a metastatic site (specific organ unknown), with an estimated tumor purity of 63%, and the tumor was classified as microsatellite stable (MSS). Genomic profiling revealed a KRAS G13D mutation, a frequent driver mutation in colorectal cancer. The patient had undergone extensive treatment over 14 years, including four rounds of radiation, a combined 36 months of irinotecan + 5-FU, 5 months of oxaliplatin + 5-FU, and 3 months of irinotecan combined with bevacizumab, an anti-VEGF antibody. The patient presented with progressive disease at the time of enrolment into the DKFZ/NCT/DKTK-MASTER program and molecular analyses were applied to uncover potential treatment options.

Following genome sequencing and mRNA sequencing, several treatment options were indeed proposed by the MTB, although none were ultimately implemented, for reasons unknown. The recommendations were prioritized as follows:

- Priority 1: MEK1 inhibitor, due to *KRAS* G13D mutation
- Priority 2: Immune checkpoint inhibitor (ICI) in combination with MEK inhibitor, due to high tumor mutational burden and *KRAS* G13D mutation
- Priority 3: pan-RAF inhibitor, due to *KRAS* G13D mutation
- Priority 4: PARP inhibitor, due to a BRCAness profile, meaning homologous recombination deficiency (HRD)

To explore whether the proteomic layer supported any of these recommendations, I integrated multiple levels of molecular information, including estimates of transcription factor activities,

pathway activities, protein expression, and kinase activities, and categorized findings into 11 therapeutic baskets, inspired by the NCT-MASTER interventional classification. These included: apoptosis, cell cycle (CC), hormone receptor signaling (HR), RAS-MEK-ERK signaling (RME), signaling by receptor tyrosine kinases (RTK), Wnt signaling, DNA damage response (DDR), immune system (IS), metabolism, and TGF-β signaling. This basket-based approach aimed to reveal whether multiple molecular alterations in a given patient converge on specific targetable biological processes.

Using RNA-seq as a footprint, I estimated transcription factor (Garcia-Alonso et al. 2019) and pathway activities (Schubert et al. 2018) relative to the rest of the cohort. Although these analyses are technically transcriptome-based, they provide a functional readout that estimates protein-level activity. In this particular patient, transcription factors related to the immune system were among the most active, including CIITA, NF-KB2, REL, and IRF4 (Figure 36A). CIITA is a master regulator of the expression of major histocompatibility complex (MHC) class II genes, which are essential for antigen presentation and the proper functioning of the adaptive immune system (Steimle et al. 1993; Steimle et al. 1994; Chang et al. 1994). CIITA drives MHC class II expression on antigen-presenting cells like dendritic cells or B cells, enabling them to present tumor antigens to CD4+ T cells. This primes anti-tumor T-cell responses, which are essential for immune checkpoint inhibitors to function effectively (Wang et al. 2023; Amrane et al. 2025; Axelrod et al. 2019). NF-KB2 has been shown to be upregulated in advanced CRC with hepatic metastasis and to promote immune escape by inducing CD8+ T-cell exhaustion. It directly interacts with STAT2, enhancing STAT2 phosphorylation and upregulating PD-L1 expression on cancer cells. This suppresses anti-tumor immunity by inhibiting cytotoxic T-cell activity. Inhibition of NFκB2 with the small molecule Rg5 reduced PD-L1 expression and synergized with anti-PD-1 immunotherapy, improving tumor control in preclinical models (Zhang et al. 2024). REL, a further member of the NF-κB family, regulates CD4+ T cell function, suggesting that REL may influence anti-tumor immune responses (Lalle et al. 2024). IRF4 is predominantly expressed in tumor-infiltrating immune cells (e.g., T cells, macrophages) and associates with immunosuppressive pathways. High IRF4 levels may drive immune evasion by modulating cytokine networks and T-cell exhaustion (Tan et al. 2025).

Pathway activities also highlighted immune system-related signaling, including high activities of JAK-STAT, NF-κB, and TNF-α pathways (Figure 36B). Additionally, the PI3K pathway was found to be more active relative to the rest of the cohort.

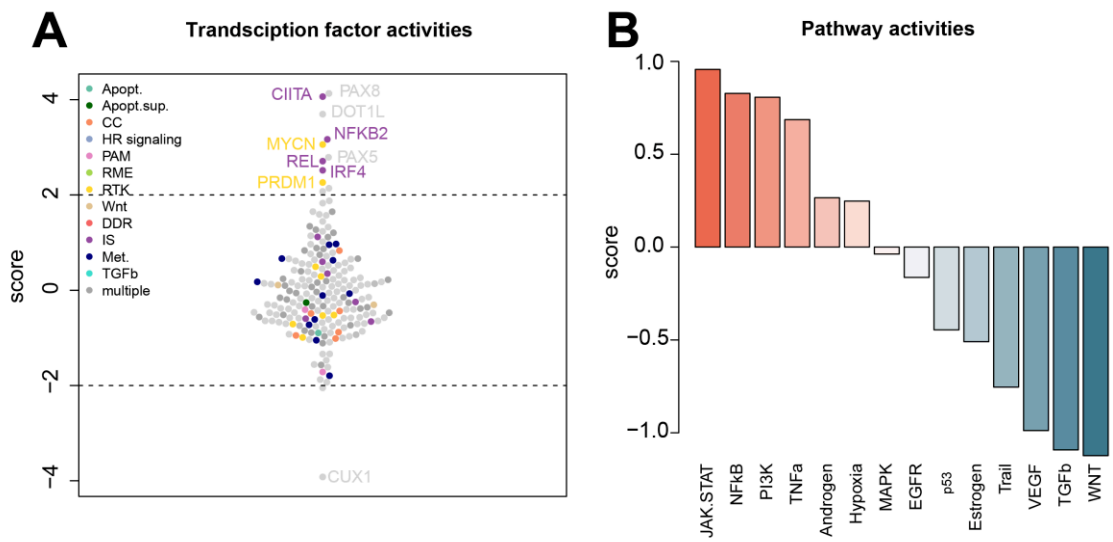


Figure 36: Example patient: mRNA level
Relative mRNA expression for each patient compared to the entire cohort was computed as the log2 RNA TPM value ratio to the median TPM of all patients. (A) Using deceoupleR package, I estimated transcription factor (Garcia-Alonso et al. 2019) activities based on the CollecTRI database for this example patient relative to the rest of the cohort. The transcription factors were assigned to one, or multiple, baskets based on Reactome pathways and highlighted in diferent colors. (B) Using deceoupleR package, I estimated pathway activities (Schubert et al. 2018) based on the PROGENy database for this example patient relative to the rest of the cohort.

At the protein level, the expression of immune-related markers was markedly elevated compared to the cohort median, including CD74, STAT1, HLA-A, and ICAM1 (Figure 37A). CD74, a key regulator of MHC class II antigen presentation, facilitates the assembly and surface expression of HLA-DR dimers on MSI-H CRC cells, enhancing tumor antigen presentation to CD4+ T cells. Improved HLA-DR dimerization correlates with better responses to anti-PD-1 therapy in preclinical models and patients (Sun et al. 2025). While Sun et al. showed this mechanism only in MSI-H cell lines and patients, CD74 might have a similar role in MSS patients like the present example. In another study Acha-Sagredo et al. (2025) showed that CRC patients can be classified into interferon (IFN)-high or -low groups. In the IFN-high patients, T cells induce expression of CD74 in tumor cells and tumor-associated macrophages (TAM) which was predictive of response to anti-PD-1 ICI in both MSI and MSS patients.

HLA-A, a key component of the MHC class I which is responsible for tumor neoantigen presentation. MSS CRC often exhibits low HLA-A expression, contributing to an "immune-cold" phenotype resistant to ICIs (Angelova et al. 2015). Reduced HLA-A limits neoantigen visibility, even in tumors with moderate TMB (Montesion et al. 2021). Since the tumor sample from this patient expressed high levels of HLA-A, tumor-specific neoantigens might indeed be presented on the cell surface and accessible for T cells.

Gene set enrichment analysis of selected Reactome pathways further confirmed significant upregulation of proteins involved in cytokine signaling, adaptive and innate immunity, and hormone receptor signaling (Figure 37B).

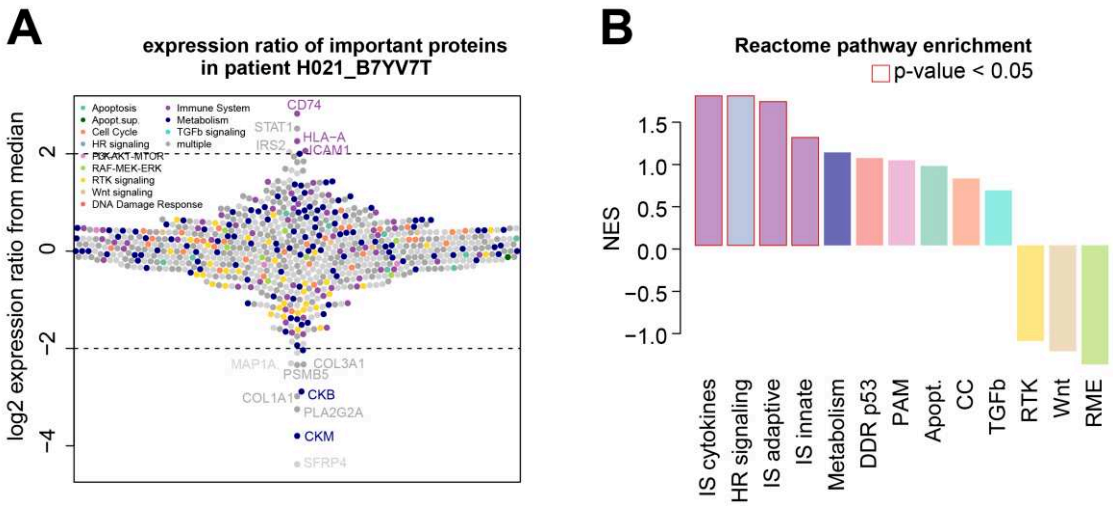


Figure 37: Example patient: Full proteome level
Relative protein expression for each patient compared to the entire cohort was computed as the log2 protein intensity ratio to the median intensity of all patients. (A) Proteins were assigned to different baskets based on Reactome pathways and extreme outliers are highlighted. (B) Gene set enrichment analysis of selected Reactome pathways was performed on the expression ratios using the GSEA function of the clusterProfiler R package.

In terms of kinase activity, several kinases showed relatively high activity, including CSNK2, CDKs 1/2/4/6, MAPK1/3, and the PI3K-AKT-mTOR axis (AKT1, AKT3, mTOR), indicating potential vulnerabilities in cell cycle regulation and growth factor signaling (Figure 38).

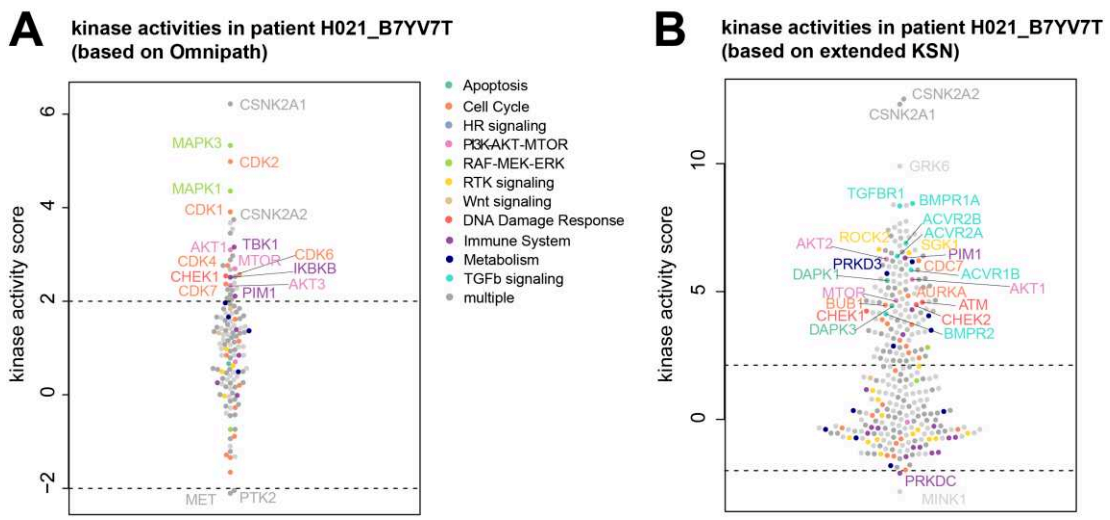


Figure 38: Example patient: kinase activities
Relative kinase activities were computed based on the Omnipath database (A) or the extended KSN (B) by applying the run_wmean function of the decoupleR package to the z-scored phosphosite intensities of each patient. Kinases were assigned to one, or multiple, baskets based on Reactome pathways and are highlighted in different colors.

Taken together, these findings suggest that the immune system-related signals observed across multiple molecular levels could have supported the prioritization of immunotherapy, aligning with Priority 2 on the MTB list. This case exemplifies how proteomics could have helped refine treatment priorities, especially in a patient where genomic biomarkers alone do not fully capture the underlying tumor biology.

2.9 Integration of possible biomarkers for immune checkpoint inhibition

Immune checkpoint inhibition (ICI) is an emerging and promising therapeutic strategy that has gained traction across various tumor types, particularly due to its pan-cancer approval for tumors with high tumor mutational burden (Gandara et al. 2025). Reflecting this trend, ICI-based therapies were the most frequently recommended treatment option by the molecular tumor board in this cohort, proposed for 17 patients as either a first or second priority. However, despite the theoretical promise, ICI was only implemented in one case, where high TMB served as the qualifying biomarker. Unfortunately, this patient showed progressive disease upon follow-up, suggesting a non-response.

This limited success aligns with broader clinical observations in colorectal cancer (CRC), where ICI responses are generally modest, especially in metastatic (André et al. 2023; Shiu et al. 2023) or microsatellite stable (MSS) tumors (San-Román-Gil et al. 2023; Chalabi et al. 2020). While microsatellite instability (MSI-high) tumors tend to show better responses to ICI, the vast majority

of CRC cases are MSS and lack robust predictive biomarkers for immune checkpoint therapy response (Yamaguchi et al. 2025; Chalabi et al. 2020; Guven et al. 2024).

Unlike certain cancer types where a clear biomarker, such as PD-L1 expression in non-small cell lung cancer (Reck et al. 2019), can help guide treatment decisions, the biomarker landscape for ICI response in metastatic CRC is more fragmented and complex. The predictive value of PD-L1 expression was found to be insignificant (Yu et al. 2024).

A wide range of potential response indicators have been proposed, including:

- Total number of coding mutations, serving as a proxy for neoantigen load (Wen et al. 2022; Manca et al. 2023)
- High expression of CD74 (Sun et al. 2025; Acha-Sagredo et al. 2025)
- MSI status, with high scores correlating with better ICI response rate (Le et al. 2015; Chalabi et al. 2020; Chalabi et al. 2022), although the response in the metastatic setting is limited (André et al. 2023).

I further hypothesized that

- Elevated TNF-α and NF-κB pathway activities, indicating a pro-inflammatory tumor microenvironment,
- Presence of infiltrating CD8+ T-cells, reflecting an immunologically active tumor microenvironment,
- And abundant expression of MHC class I components, which is crucial for effective antigen presentation

Might be beneficial for ICI treatment.

To better explore which patients in my cohort might have been candidates for immune checkpoint inhibition, I compiled a composite panel of potential response biomarkers, integrating transcriptomic and proteomic information. The goal was to capture the multifaceted nature of immune infiltration and immune responsiveness—factors that are critical in identifying patients who might benefit from ICI, but which are often poorly reflected in single-layer biomarkers.

To this end, I included the following features for each patient:

- CD8+ T-cell infiltration estimates, derived from tumor bulk RNA-seq data using the EIPC method (Racle et al., 2017) as implemented in the immunedeconv R package (Sturm et al., 2019),
- General immune infiltration scores, estimated using ESTIMATE (Yoshihara et al., 2013) from the same package,

- Number of exonic single nucleotide variants (SNVs) as a proxy for tumor mutational burden,
- MSI sensor scores (Niu et al. 2014) with a recommended threshold of 3.5, provided by the NCT-MASTER program,
- Pathway activity scores for TNF-α and NFκB signaling, calculated using PROGENy (Schubert et al., 2018),
- Mean protein abundance of MHC class I complex components (HLA-A, HLA-B, HLA-C, and B2M),
- And CD74 protein abundance, a potential immune activation marker.

To ensure comparability across the different data types and measurement scales, I z-scored all values across patients, transforming them into standardized scores that reflect how much a patient's value deviates from the cohort mean. This normalization allowed me to compare the relative "immune-readiness" of each patient using a shared scale.

By aggregating the z-scored values for all these potential biomarkers, I derived a composite immune infiltration score for each patient and ranked them accordingly—from those with the lowest immune signatures to those with strong immune-related signals (Figure 39).

This integrative visualization revealed substantial variation in immune infiltration and potential ICI responsiveness across the cohort. Notably, several patients who had not received ICI recommendations by the tumor board showed strong immune-related signatures, suggesting that a multi-layered molecular perspective might help identify additional patients who could benefit from checkpoint inhibition—and, conversely, help avoid recommending ICI in cases where molecular markers suggest non-responsiveness, despite high TMB alone. The two patients with very high MSI sensor values (Y78RB8 and HE2H4F) were found to have the highest and fourth-highest aggregated biomarkers score, while the other three patients with a sensor value higher than 3 (59G42D, EBUXWK, and FUYU5R) were found more towards the lower end of the score.

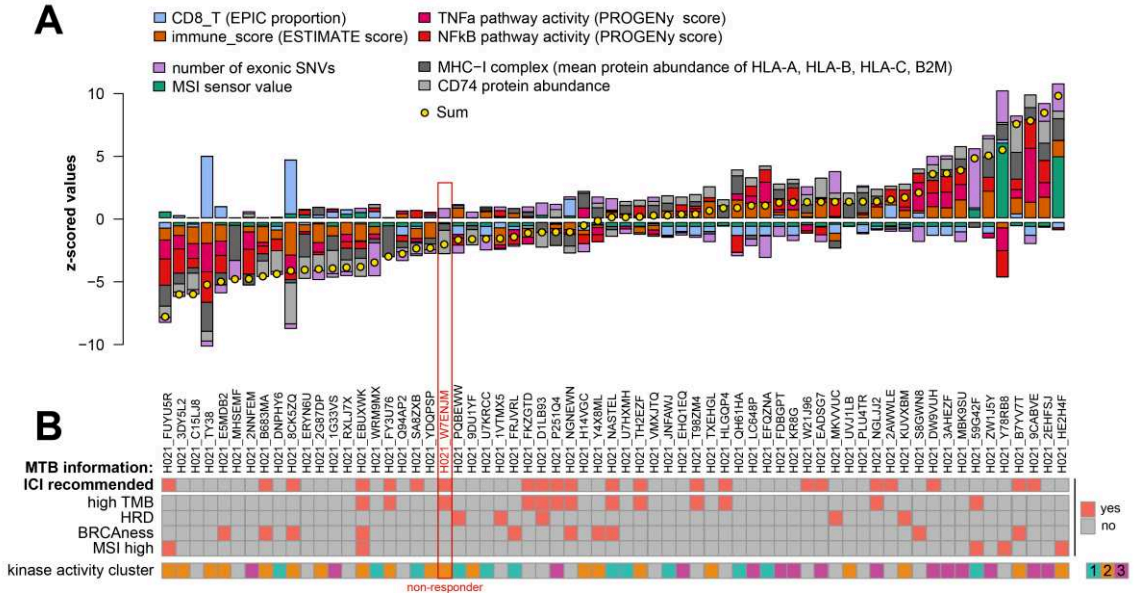


Figure 39: Potential biomarkers for immune checkpoint inhibition
(A) Starting from the bulk RNA sequencing data, I used the immunedeconv R package with the EPIC method to estimate immune cell proportions and the ESTIMATE method to obtain a general score for the extent of immune cell infiltration. Furthermore, I calculated the number of exonic single nucleotide variants (SNVs) from the SNV table provided by DKFZ/NCT/DKTK-MASTER. The MSI sensor values were also provided by MASTER. Relative pathway activity scores for TNFα and NF-κB signaling were computed based on the PROGENy database (Schubert et al. 2018) using the decoupleR package function run_wmean on the z-scaled TPM values. For the abundance of the MHC-I proteins, I calculated the mean expression ratio of the HLA-A, HLA-B, HLA-C and B2M proteins. All final values were z-scaled over the patients so that they have equal influence on the final score. To obtain the final score, all values were summed up for each patient (yellow dots). (B) From the clinical metadata received from the DKFZ/NCT/DKTK-MASTER program, I extracted the information on whether immune checkpoint inhibition was recommended, whether the tumor had a high mutational burden (TMB), whether the tumor showed homologous recombination deficiency (HRD), BRCAness or a high MSI sensor value (> 3.5). The one patient that was treated with ICI according to the MTB recommendation (W7ENJM), but did not respond to the treatment, is highlighted.

The only patient that received ICI in accordance with the MTB recommendation - patient W7ENJM - did not respond to the treatment and showed progressive disease. This patient had a relatively high mutational burden, but low MHC-I complex and CD74 expression. This might indicate that neoantigens might not be sufficiently presented on the cell surface to be recognized by CD8+ T cells and could potentially explain why this patient did not benefit from ICI. While one single patient is definitely not enough to conclude that this kind of aggregation of potential immune biomarkers reflects the true responsiveness, it might help clinicians to evaluate a broader range of factors that affect ICI response and give a better-informed recommendation.

2.10 Characterization of KRAS-mutated CRC patients to identify potential new vulnerabilities

2.10.1 Activating KRAS mutations do not lead to significantly higher activities of downstream signaling

Returning to the example patient described in chapter 2.7, the top-priority treatment recommendation from the molecular tumor board was a RAF/MEK inhibitor (Regorafenib), based on the presence of a KRAS G13D mutation in the tumor. Since this variant is known to result in constitutively active KRAS (Johnson et al. 2019), one could expect that the downstream signaling cascade, particularly involving kinases like RAF, MEK, and ERK, should also exhibit elevated activity in comparison to KRAS wild-type patients. As shown in Figure 33A, this specific patient indeed displayed higher activity of ERK1 (MAPK3) and ERK2 (MAPK1), raising the question: Could this increased downstream activity be attributed to the KRAS G13D mutation?

To explore this more systematically, I compared the relative kinase activities of RAF1, MAP2K1 (MEK1), MAPK1 (ERK2), and MAPK3 (ERK1) across patients grouped by KRAS mutation status: KRAS WT (n=38), codon 12 mutations (n=17), codon 13 mutations (n=5), and four other KRAS mutations, including rarer activating variants such as A146T and Q61L. One patient with an NRAS Q61K alteration, also known to be activating (Ahmadian et al. 1999), was included in the “other” group.

Surprisingly, this analysis did not reveal consistently higher activity of downstream kinases in any of the KRAS-mutant groups compared to WT patients (Figure 40). The only statistically significant difference was observed in the “other” group, where MAP2K1 activity was significantly elevated relative to the WT group (Figure 40B). This appeared to be largely driven by the patient with the A146T mutation, who showed high activity of RAF, MEK, and ERK. The KRAS A146T mutation has been previously shown to have specific phenotypic effects that differ from the classical G12D alteration (Poulin et al. 2019). Furthermore, the KRAS A156T alteration has been described to represent a distinct subgroup of metastatic CRC patients with high tumor burden and poor prognosis (van 't Erve et al. 2021) and to be more sensitive to MEK inhibition (Puliga et al. 2024).

Interestingly, some KRAS WT patients also exhibited high activity in this signaling axis, such as patients 2EHFSJ and P251Q4, emphasizing that elevated RAS-MAPK pathway activity is not exclusively linked to KRAS mutation status. In conclusion, this suggests that while mutations like KRAS G13D can be informative, they do not universally predict downstream kinase activity. Instead, the proteomics-based activity estimates may reveal functionally active signaling in patients regardless of KRAS mutational status, and could therefore serve as a more direct and individualized indicator for MAPK pathway-targeted therapies.

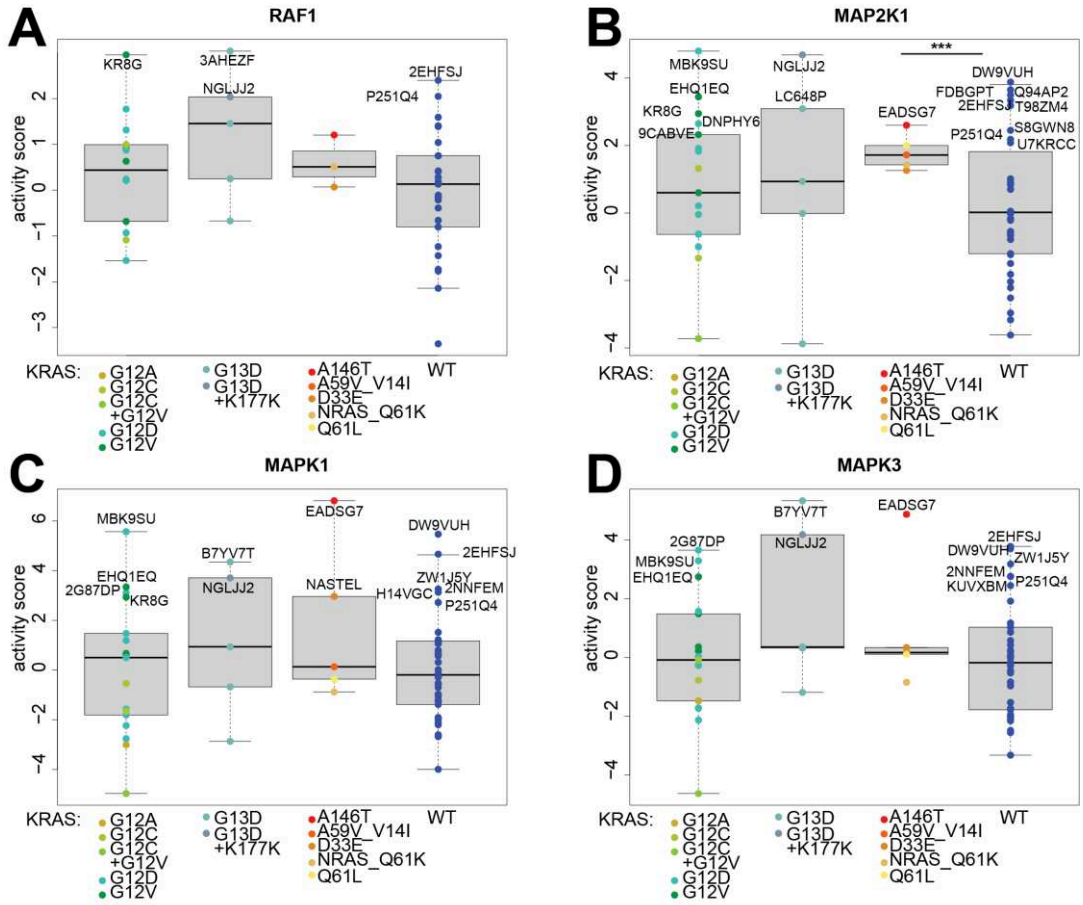


Figure 40: Kinase activities based on classical Omnipath KSN: kinases downstream of KRAS grouped by KRAS mutation status.
Kinase activities were calculated by mapping identified phosphosites to their responsible kinases, as annotated in the Omnipath database. For kinases with at least 5 identified substrate phosphosites, the relative activity was then computed by applying the run_wmean function of the decoupleR package to the z-scaled phosphosite intensities of each patient. Activities of kinases downstream of KRAS, namely RAF1 (A), MAP2K1 (B), MAPK1 (C), and MAPK3 (D) were grouped by KRAS mutation status and visualized as boxplots. Significance between KRAS-mutated groups and the KRAS WT patients was assessed using two-sided, unpaired Student's t-tests. *** = p-value < 0.001.

Using the extended kinase-substrate network (see 2.5.3), I could additionally evaluate the per-patient activities of BRAF (Figure 41A) and MAP2K2 (Figure 41B). These KRAS-downstream kinases showed a very similar picture with neither of them having significantly higher activities in the KRAS-mutated patient groups compared to the WT patients. The same KRAS WT patients (P251Q4 and 2EHFSJ) stand out with very high BRAF and MAP2K2 activities.

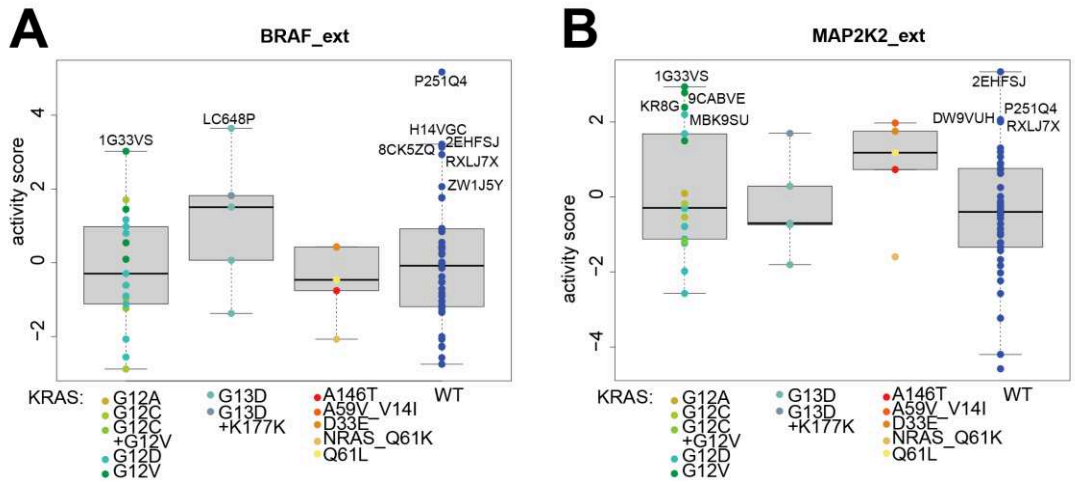


Figure 41: Kinase activities based on extended KSN: kinases downstream of KRAS grouped by KRAS mutation status.
Kinase activities were calculated by mapping identified phosphosites to their responsible kinases, as annotated in the extended KSN. For kinases with at least 5 identified substrate phosphosites, the relative activity was then computed by applying the run_wmean function of the decoupleR package to the z-scaled phosphosite intensities of each patient. Activities of additional kinases downstream of KRAS, namely BRAF (A) and MAP2K2 (B), were grouped by KRAS mutation status and visualized as boxplots. Significance between KRAS-mutated groups and the KRAS WT patients was assessed using two-sided, unpaired Student's t-tests. No comparison was found to be significant.

Ultimately, it is not known whether high activities of KRAS-downstream kinases are indeed indicative of better response to drugs targeting the RAF-MEK-ERK pathway. However, having a closer look at the individual kinase activities might still help clinicians to select those patients in which the KRAS mutation actually shows a downstream effect.

KRAS, long considered undruggable due to its lack of suitable binding pockets and high affinity for GTP, is now recognized as a druggable target, at least for certain mutations, especially for the KRAS G12C mutation. Drugs like sotorasib (AMG510) and adagrasib (MRTX849) have shown significant clinical efficacy, leading to regulatory approvals for non-small cell lung cancer (Skoulidis et al. 2021; Langen et al. 2023). While these advances have transformed the landscape for KRAS-targeted therapy, challenges remain. Not all KRAS mutations are currently druggable - most progress has been made with inhibitors of KRAS G12C, while inhibitors for other KRAS variants (such as G12D, G13D, and A146T), or pan-RAS, are still under clinical development (Herzberg and Manji 2023). Additionally, other pathways can compensate for KRAS inhibition and resistance to KRAS inhibitors can develop, necessitating research into combination therapies and next-generation inhibitors (Zhu et al. 2022; Saleh et al. 2024; Isermann et al. 2025).

This led me to the next question: What else characterizes the KRAS-mutated patients? Can I identify any other potential vulnerabilities and (co-)treatment options?

2.10.2 Molecular characterization of KRAS-mutated patients reveals high activities of NF-κB, metabolic adaptations, and GSK3β activation

To identify potential additional vulnerabilities in KRAS-mutated CRC, I began by analyzing the RNA-seq data, focusing on transcriptional changes associated with KRAS hotspot mutations. For this, patients were grouped into those carrying a KRAS G12 or G13 mutation (n=22) and KRAS wild-type (WT) patients (n=38), while patients with other, less common KRAS variants were excluded due to limited representation. I performed differential gene expression analysis on the RNA-seq TPM values using the eBayes function of the limma R package. To move beyond gene-level changes and explore functional alterations, I employed the decoupleR package (Badia-I-Mompel et al. 2022) to estimate transcription factor (TF) and pathway activity scores. For TF activity inference, I used t-statistics from the differential expression analysis in combination with the CollecTRI database (Müller-Dott et al. 2023), while pathway activities were derived using the PROGENy database (Schubert et al. 2018).

This analysis revealed that KRAS-mutant tumors exhibited elevated activities of several subunits of the NF-κB transcription factor, including NF-KB1, RELA, REL, and NF-KB (Figure 42A). NF-κB is a master regulator of cancer-related processes, promoting tumor growth, survival, metastasis, inflammation, and therapy resistance. Its constitutive activation is a hallmark of many cancers, making it a critical target for therapeutic intervention, as reviewed by Taniguchi, Karin (2018). The activation of KF-κB by mutated KRAS has been shown in lung cancer (Bassères et al. 2010; Meylan et al. 2009; Xia et al. 2012). All three studies demonstrated the requirement of NF-κB activation for enhanced proliferation and decreased apoptosis in KRAS-driven lung cancer. My data suggest that a similar connection between mutated KRAS and NF-κB activation exists in colorectal cancer.

The MAPK pathway, which acts downstream of KRAS, emerged as the most active signaling pathway in the mutant group (Figure 42B), suggesting that there is increased expression of MAPK signaling target genes although the single kinases, such as MAP2K1 or MAPK1, did not show significantly higher activity. Additional pathways showing increased activity in KRAS-mutant tumors included JAK-STAT, TNFα, and NF-κB, further supporting a pro-inflammatory, growth-promoting environment. Interestingly, despite the well-established resistance of KRAS-mutant CRC tumors to EGFR inhibition (Lièvre et al. 2006; Benvenuti et al. 2007; Freeman et al. 2008), I identified EGFR signaling as one of the most active pathways compared to KRAS WT patients (Figure 42B). Combining KRAS and EGFR inhibition has shown promising results in colon cancer cell lines (Kataoka et al. 2023; Feng et al. 2023) as well as in early clinical trials (Shergill et al. 2022). Overall, these findings highlight distinct transcriptional and signaling profiles in KRAS-mutated tumors that could be exploited for targeted or combinatorial treatment strategies.

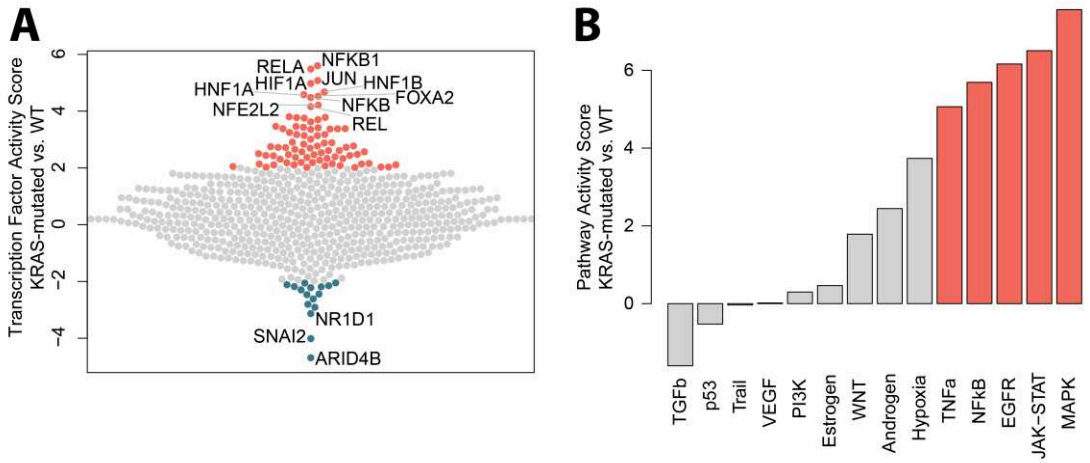


Figure 42: Transcription factor and pathway activities of KRAS-mutated patients in comparison to KRAS WT patients.
The eBayes function of the limma R package was used to calculate differential gene expression from RNA-seq TPM values for KRAS mutated (n = 19) vs. WT (n = 34) patients. (A) Using the decoupleR package, I obtained the collecTRI database of transcription factor target genes. To compute relative transcription factor activity scores between the two groups, I applied the run_wmean function to the t-statistics of the differential gene expression analysis. The top 10 most active transcription factors are highlighted. (B) Using the decoupleR package, I obtained the PROGENy database containing the top100 most responsive genes of 14 pathways. To compute relative pathway activity scores between the two groups, I applied the run_wmean function to the t-statistics of the differential gene expression analysis. Significant pathway activities are highlighted in red.

To identify potential KRAS-mutated patients-specific vulnerabilities, I filtered the protein intensities for at least 9 values in both mutated (n=22) and WT (n=38) patient groups. I performed differential gene expression analysis on log2 protein intensity values using the eBayes function of the limma R package. After Benjamini-Hochberg adjustment of p-values only seven proteins were found to be significantly differentially expressed in the mutated patients (Figure 43A).

Among the most upregulated proteins was pyruvate dehydrogenase kinase 3 (PDK3) which plays a crucial role in metabolic reprogramming of cancer. PDK3 inhibits the pyruvate dehydrogenase complex, thus blocking oxidative metabolism of pyruvate and shifting cells toward glycolysis - a hallmark of cancer metabolism (Woolbright et al. 2019). High PDK3 expression correlates with advanced tumor stage, metastasis, and reduced survival in colorectal cancer (Lu et al. 2011). PDK3 overexpression is furthermore linked to chemotherapy resistance across multiple cancer types (Lu et al. 2011; Lu et al. 2008) and drives immune evasion (Liu et al. 2024). In a large drug repurposing study, PDK3 was simulated to be druggable by FDA-approved compounds Bagrosin and Dehydrocholic acid (Khan et al. 2024) which makes it a potentially druggable vulnerability of KRAS-mutated colorectal cancer.

Gene set enrichment analysis (GSEA) of Reactome pathways revealed further metabolic reprogramming with upregulated expression of proteins involved in the TCA cycle, metabolism of lipids, vitamins, amino acids, and cofactors (Figure 43B).

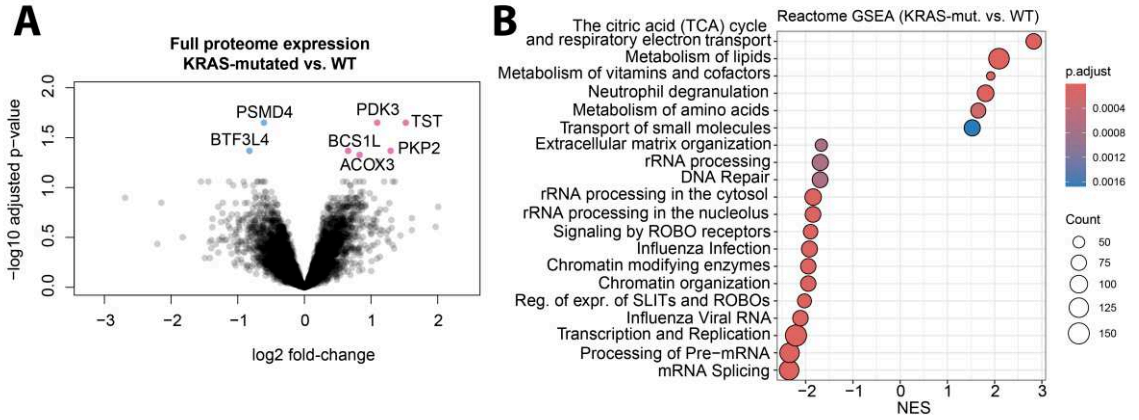


Figure 43: Full proteome comparison between KRAS-mutated patients and WT patients. The eBayes function of the limma R package was used to perform differential protein abundance analysis between the KRAS mutated (n = 22) and the WT (n = 39) patients. Proteins with less than 9 values present in both groups were excluded. P-values were corrected for multiple testing using the Benjamini-Hochberg method. (A) Log2 fold-changes and -log10 of adjusted p-values are visualized as a scatter plot. Significantly differentially abundant proteins are highlighted. (B) The org.Hs.eg.db package was used to map gene names to entrez IDs. Then, the gsePathway function of the ReactomePA package was applied to the sorted log2 fold-changes to perform GSEA of Reactome pathways (min gene set size 120, BH adjustment, p-value cut.off = 0.05). The 20 most enriched pathways are visualized.

To further explore functional differences between KRAS-mutated and wild-type (WT) CRC tumors, I extended the analysis to kinase activity inference using phosphoproteomics data. Kinase–substrate relationships were obtained from the OmnipathR database and I applied several filtering steps to ensure robust results. First, phosphosites were excluded if ProtMapper was their only annotation source, and only those with at least nine data points in both patient groups were retained. I then performed differential phosphorylation analysis using the limma framework with empirical Bayes moderation and BH-adjusted p-values. Next, I filtered the Omnipath kinase–substrate network (KSN) to include only kinases with at least five matched substrates in my dataset. Kinase activities were estimated using run_wmean, based on the t-statistics from the differential analysis. Additionally, to avoid misleading results due to substrate redundancy, I removed 28 kinases previously flagged as non-trustworthy due to high substrate overlap (see 2.5.2).

Among the most striking findings was GSK3β, which showed the highest inferred activity in KRAS-mutated tumors (Figure 44A). This could be clinically relevant, since GSK3β has been shown to be essential for the survival of KRAS-mutated pancreatic cancer cells (Kazi et al. 2018),

and its role in CRC is also supported by literature. A review by Vidri and Fitzgerald (2020) highlighted GSK3β as a key modulator linking KRAS signaling to the Wnt pathway, a concept that was initially proposed by Li et al. (2005). The activation of GSK3β is supported by increased phosphorylation intensities of a large number of its known substrates in the mutant group, with phosphorylated LMO7 (S988 and S991) being the most differentially abundant phosphopeptides (Figure 44B). LMO7 has been implicated in metastasis (Dai et al. 2025) and immune evasion (Xie et al. 2019), reinforcing the potential clinical relevance of this kinase in KRAS-mutated CRC. These results point to GSK3β as a potential vulnerability and an actionable target worth exploring further.

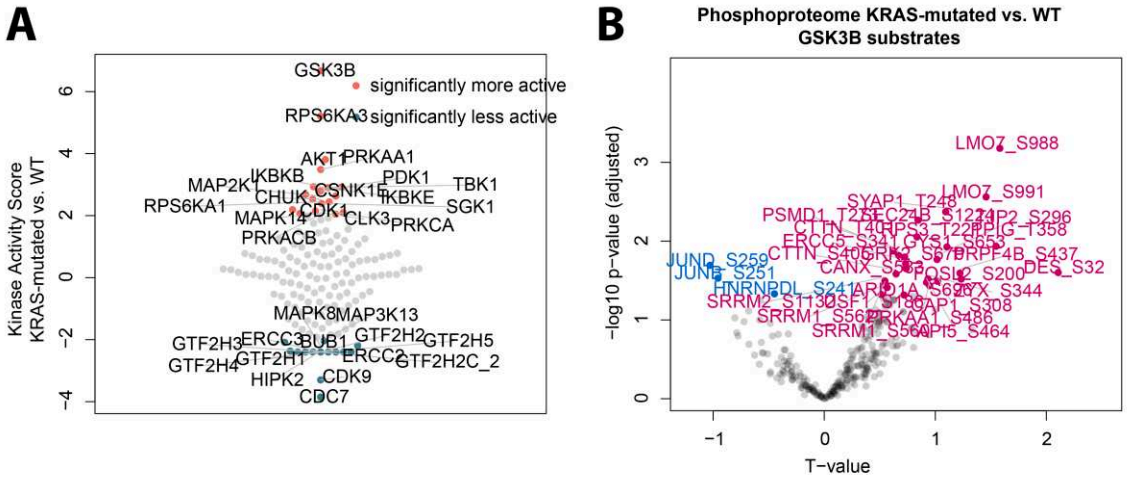


Figure 44: Kinase activities of KRAS-mutated patients and phosphosite substrates of GSK3B.
The eBayes function of the limma R package was used to assess differential abundance of phosphosites between the KRAS mutated (n = 22) and the WT (n = 39) patients. Phosphosites with less than 9 values present in both groups were excluded. P-values were corrected for multiple testing using the Benjamini-Hochberg method. Kinase activities were then computed based on the Omnipath database using the decoupleR package. (A) Kinase activities in the KRAS-mutated patients comapred to WT patients, visualized as a beeswarm plot. Kinases with an absolute score > 2 are highlighted. (B) T-statistic values and -log10 adjusted p-values of GSK3b substrate phosphosites are visualized as a scatter plot. Significantly differentially abundant phosphosites are highlighted.

Having demonstrated the variety and complexity of proteome-level information that can be extracted from the molecular patient data, including transcription factor and pathway activities, relative protein expression, and kinase activities, I next aimed to integrate the data into a patient-specific score.

2.11 Integrating protein and pathway activities into an interventional basket-based, patient-specific score

I developed a composite molecular activity score designed to integrate information across multiple biological layers and prioritize oncogenic pathway dependencies for each patient. The overarching goal was to translate complex molecular signatures into actionable insights, inspired by the concept of interventional baskets as used in the NCT-MASTER program. To approximate pathway-level vulnerabilities, I selected a curated list of 16 Reactome pathways corresponding to major oncogenic processes such as cell cycle regulation, apoptosis, PI3K/AKT and MAPK signaling, WNT and TGFβ pathways, DNA damage response, metabolism, and both innate and adaptive immunity.

Selected Reactome pathways included: cell cycle ("R-HSA-1640170"), apoptosis ("R-HSA-109581"), suppression of apoptosis ("R-HSA-9635465"), growth hormone receptor signaling ("R-HSA-982772"), PI3K/AKT signaling in cancer ("R-HSA-2219528"), PI3K/AKT activation ("R-HSA-198203"), oncogenic MAPK signaling ("R-HSA-6802957"), signaling by receptor tyrosine kinase ("R-HSA-9006934"), signaling by WNT in cancer ("R-HSA-4791275"), p53-dependent G1 DNA damage response ("R-HSA-69563"), p53-independent DNA damage response ("R-HSA-69610"), innate immune system ("R-HSA-168249"), adaptive immune system ("R-HSA-1280218"), cytokine signaling in immune system ("R-HSA-1280215"), signaling by TGFb family members ("R-HSA-9006936"), and metabolism ("R-HSA-1430728").

2.11.1 Calculation of the patient-specific basket scores

For each patient, I aggregated data from six distinct molecular layers:

- Transcription factors (TFs) with relative activity > 2
- All PROGENy-derived pathway activity scores
- Proteins with relative expression ratio > 2
- GSEA enrichment scores of the selected Reactome pathways
- Kinases with relative activity > 2 based on Omnipath
- Kinases with relative activity > 2 based on the extended KSN

Each selected feature (e.g., a protein or TF) was assigned to one - or multiple - of the predefined baskets. Within each layer and basket, the sum of activity or enrichment values was normalized by the basket size, to prevent large baskets (e.g., cell cycle) from dominating the score. These per-basket, per-layer values were then z-score transformed across patients, ensuring that no single molecular layer (e.g., transcription factors or kinases) disproportionately influenced the overall result. Finally, the composite score was calculated by summing the six z-scores - giving a half-

weight to the extended kinase activities - to produce a patient-specific activity score for each interventional basket (Figure 45).

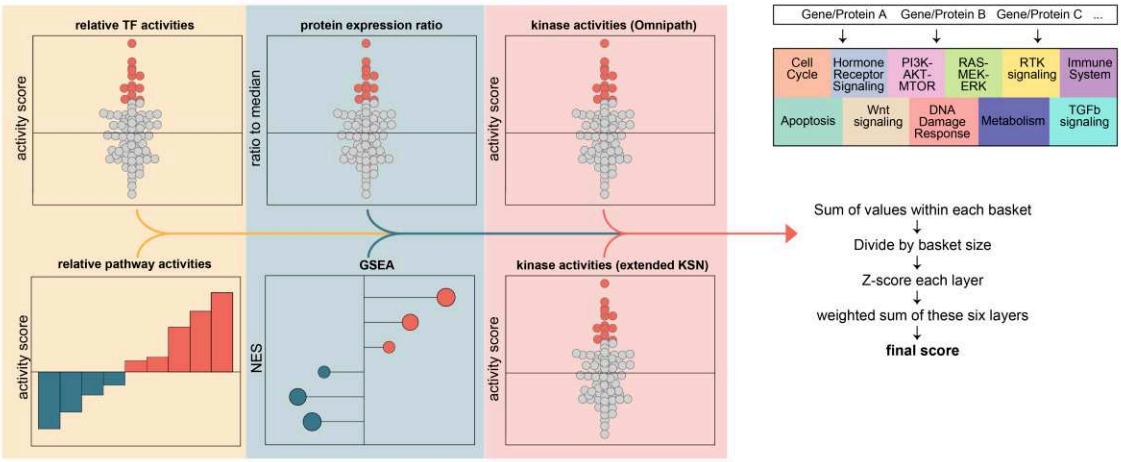


Figure 45 Schematic illustration of per-patient basket score calculation
For each patient, the following values were included into the score: transcription factors with a score > 2, all 14 pathway activity scores, proteins with expression ratios > 2, normalized enrichment scores (NES) of selected Reactome pathways, kinases with an activity score > 2. Subsequently, all included genes/proteins were assigned to one, or multiple, baskets based on selected Reactome pathways. For each basket, a sum of all containing values is computed, followed by division of this sum by the total basket size. Each layer (TF, pathway, protein, kinase) is then z-scored across all patients to ensure the same contribution of each layer. Subsequently, the six layers are summed up while applying a weight of 0.5 to the layer of extended kinase activities. The result is a final score for each basket in each patient that contains information on how active each basket is (potentially) compared to the other patients. All molecular layers contributed equally, except for the extended kinases that contribute with half weight.

The basket with the highest positive score was then selected as the final recommendation for each patient. Additionally, I recorded the highest-scoring basket per individual molecular layer to explore consistency and potential co-dependencies. Only one patient received no recommendation, having no positive score across any basket. Several patients showed multiple high-scoring baskets, potentially reflecting co-activation of pathways or mixed phenotypes (Figure 46). The agreement between different molecular layers (e.g. transcription factor activity and protein expression) was limited: Only 5 patients (7.7%) displayed agreement between the top scoring basket and at least 3 different molecular layers. This suggests that information based on gene expression is not necessarily reflected in protein abundance or activity – or vice versa. It remains to be explored which molecular information is most informative about the true tumor dependencies and vulnerabilities.

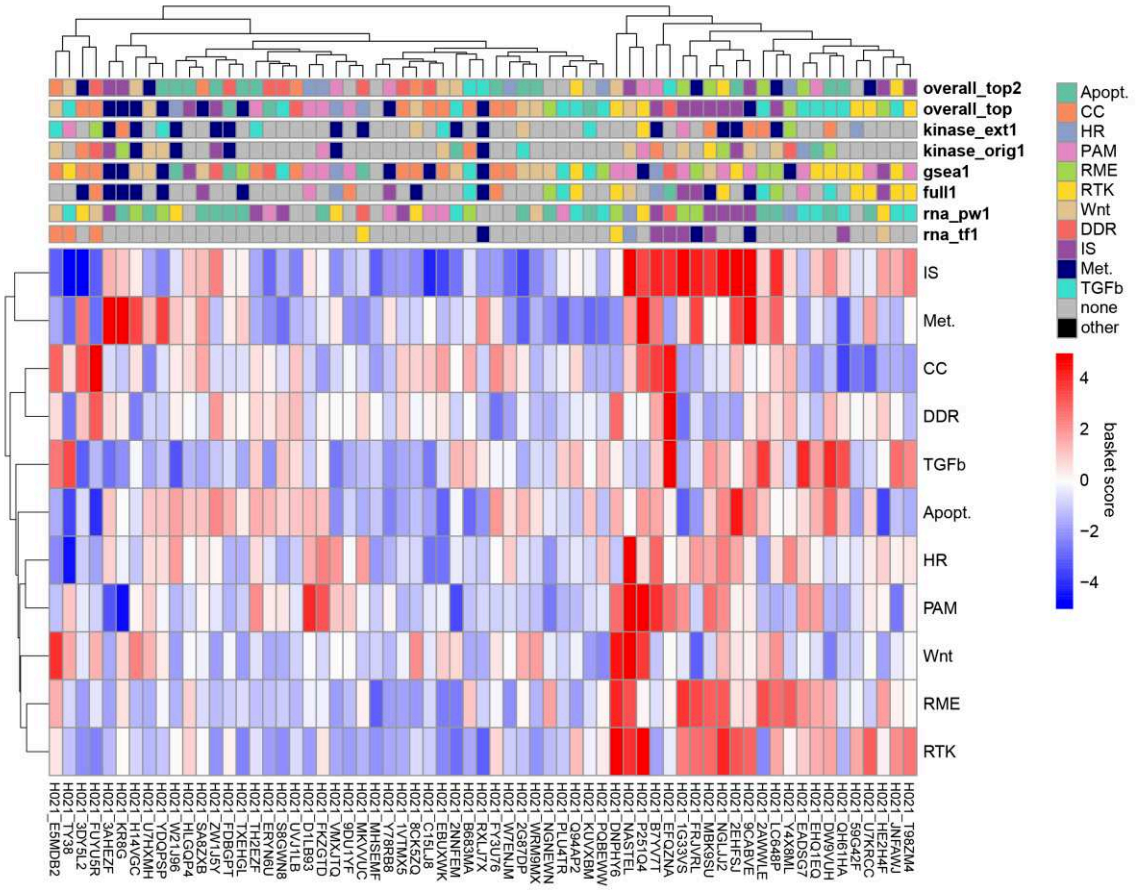


Figure 46 Final scores per basket per patient.
The final score values for each basket in every patient is visualized as a heatmap. The highest basket score was selected for each patient, if that highest score was positive ("overall top"). Additionally, the second highest scoring basket ("overall top2") was selected, if it was positive. Also, the highest scoring baskets based only on transcription factor, pathway, protein abundance, or kinase activity information, respectively, is shown.

Overall, each basket was recommended with similar frequencies, however when grouped by the previously defined kinase activity-based clusters, distinct trends emerged: Cluster 1 patients were more frequently scored into the Cell Cycle (CC) basket (Figure 47A), consistent with their highly proliferative molecular phenotype (Figure 35A). Cluster 3 patients more frequently showed the highest score in the immune system-related (IS) basket (Figure 47A), again mirroring their inflammation-heavy, immune-infiltrated signature (Figure 35A, B). The distribution of the actual basket scores over all patients reinforced these associations: Cluster 1 had significantly higher CC scores, while Cluster 3 displayed significantly elevated IS and metabolism scores. Conversely, RTK pathway scores were significantly lower in Cluster 1, potentially indicating a reduced reliance on receptor tyrosine kinase signaling in this highly proliferative group (Figure 47B).

I then compared these pathway-derived basket assignments to the actual treatment recommendations made by the Molecular Tumor Board (MTB). MTB recommendations, when

given, typically focused on specific alterations (e.g., a mutation or copy number gain of a specific gene) rather than a holistic pathway activation signature. Of the 65 patients 22 patients received no MTB recommendation, due to the absence of an actionable biomarker. For patients who did receive recommendations, I manually assigned them to baskets based on the drug's mechanism of action (e.g., CDK inhibitor → CC basket). The overall agreement between MTB and pathway-based recommendations was limited, likely reflecting their differing conceptual approaches—specific biomarker targeting vs. systems-level pathway activity (Figure 47C). Notably, only one patient (W7ENJM) actually received treatment based on the MTB recommendation (immune checkpoint inhibitor), however did not respond to the immunotherapy. In my approach this patient scored very low for the IS basket and instead scored high in cell cycle and hormone receptor signaling baskets (Figure 46). While my scoring approach is in no way validated, this suggests that the additional molecular context provided by the basket score might have prompted a different, possibly more suitable, therapeutic direction.

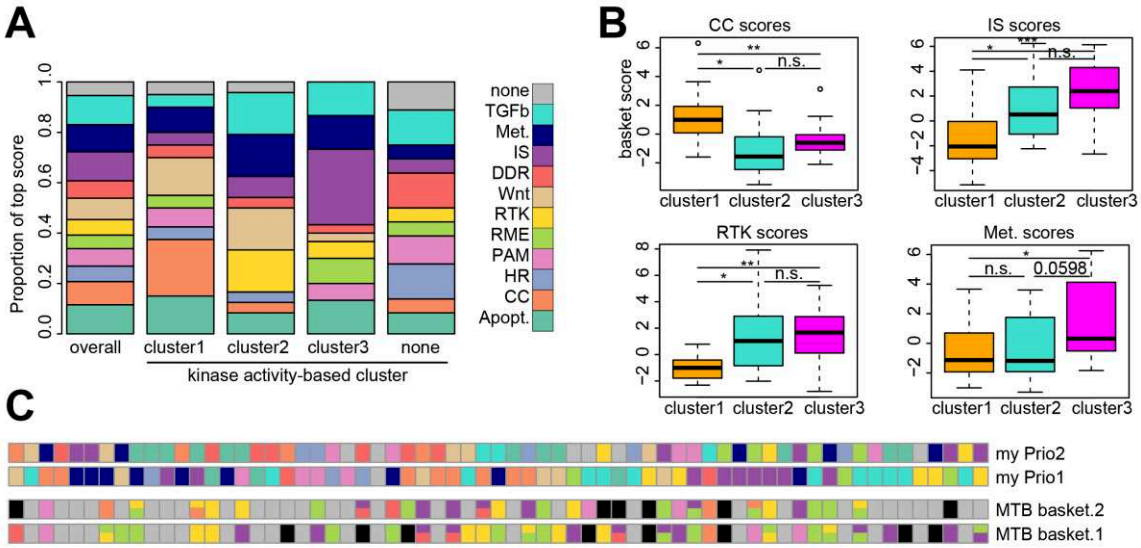


Figure 47: Distribution of basket scores among the kinase activity-based patient clusters and comparison to MTB recommendation.
(A) Distribution of baskets among the top two scores, if they were > 0, of each patient grouped by the kinase clusters of Figure 34. (B) Distribution of CC, IS, RTK, and Metabolism basket score values within the three kinase activity-based patient clusters. Significance was assessed using a two-sided Student's t-test. (C) Treatment recommendations given by the molecular tumor board were manually assigned to baskets and compared to the top scoring baskets of my analysis. In 22 cases no recommendation was given by the MTB, for the lack of actionable biomarkers.

In summary, this scoring approach offers a multilayered, systems-based perspective on oncogenic pathway activity, aiming to translate complex omics data into actionable treatment strategies. While direct validation is difficult due to limited treatment data and the MTB's single-alteration focus, this method holds potential as a complementary tool for patient stratification. To evaluate its utility, I next tested these pathway-specific vulnerability scores in patient-derived organoid

(PDO) models, leveraging available drug screening data to assess whether high basket scores predict therapeutic response.

2.11.2 Validation of basket-based scores using PDOs

RIPA lysates were generated from 24 PDO lines kindly provided by the group of Dr. Johannes Betge, with sample preparation carried out by Dr. Theresa Mulholland. The samples were processed and analyzed in two batches: initially 8 PDOs, followed by an additional 16 PDOs (Table 1) and five healthy colon biopsies.

Table 1: Metadata of CRC patient-derived organoids
PDOs were established and cultivated by Dr. Theresa Mulholland who provided me with the lysates and relevant metadata. Cells were either derived from the primary tumor (T) or a metastatic lesion (M). The PDOs were screened against drug libraries – either containing FDA-approved used drugs (clinical) or, additionally, kinase inhibitors (kinase).

PDO	batch	age	stage	T/M	KRAS	library
D104	1	68	4	T	WT	Kinase
D174	1	82	1	T	G12C	Clinical
D172	1	47	4	T	Q61P	Kinase
D080	1	64	3	T	G12D	Kinase
D235	1	51	3	T	G12A	Kinase
D086	1	72	3	T	G12D	Kinase
D098	1	49	4	T	G12D	Kinase
D021	1	86	1	T	G12D	Kinase
D007	2	33	3	T	G12D	Clinical
D013	2	84	2	T	WT	Clinical
D030	2	71	2	T	A146T	Clinical
D046	2	51	3	T	WT	Clinical
D089	2	49	4	T	G12D	Clinical
D105	2	45	4	M	G13D	Clinical
D131	2	77	3	T	WT	Clinical
D147	2	60	4	M	G12S	Clinical
D192	2	61	4	T	WT	Clinical
D208	2	34	3	T	G12D	none
D223	2	69	1	T	G12A	Clinical
D236	2	63	2	T	G13D	Clinical
D284	2	40	4	M	G12S	none
D293	2	72	4	T	Q61L	none
D318	2	48	3	T	WT	none
D330	2	69	1	T	n.d.	none

To enable cross-batch comparison, I included two repeated PDO lysates I had analyzed already in the first batch (D172 and D086) and four CRC tissue lysates (2G87DP, FDBGPT, P251Q4,

U7HXMH) in the sample preparation, measurement and data analysis of the second batch. Protein clean-up, digestion, and phosphopeptide-enrichment were performed following the automated workflow starting with 55 µg protein of each sample. Full proteome and phosphorylated fractions were measured for 120 minutes each and analyzed with Spectronaut (version 18). Protein groups were filtered for identification in at least 50% of the samples.

Initially, I observed a strong batch effect in the proteome measurements: samples clustered clearly by batch along the first principal component (PC1), which accounted for 51.5% of the total variance (Figure 48A). Notably, the replicated lysates did not cluster together, confirming the presence of batch-related artifacts.

To correct for these batch effects, I applied the removeBatchEffect() function from the limma R package, using the batch identifiers as the condition and the sample origin (patient, PDO, healthy) as the design factor. After correction, the replicate samples clustered closely with their original measurements, indicating successful batch correction. However, the PDOs and patient tissue samples still clustered apart along PC1, which now explained 27.2% of the variance (Figure 48B). Interestingly, the healthy colon biopsies clustered closely with the patient samples, suggesting that CRC tumor samples share greater proteomic similarity with healthy tissue than with PDOs.

This separation could reflect true biological differences: for example, the PDOs were derived from the primary tumor of treatment-naïve, mixed-stage patients (

Table 1), while the tissue samples came from young, advanced, heavily pre-treated patients. Alternatively, it may result from the culturing process itself, which enriches for stem-like populations and lacks the tumor microenvironment (Le Tong et al. 2024), making PDOs inherently distinct from in vivo tumor tissue.

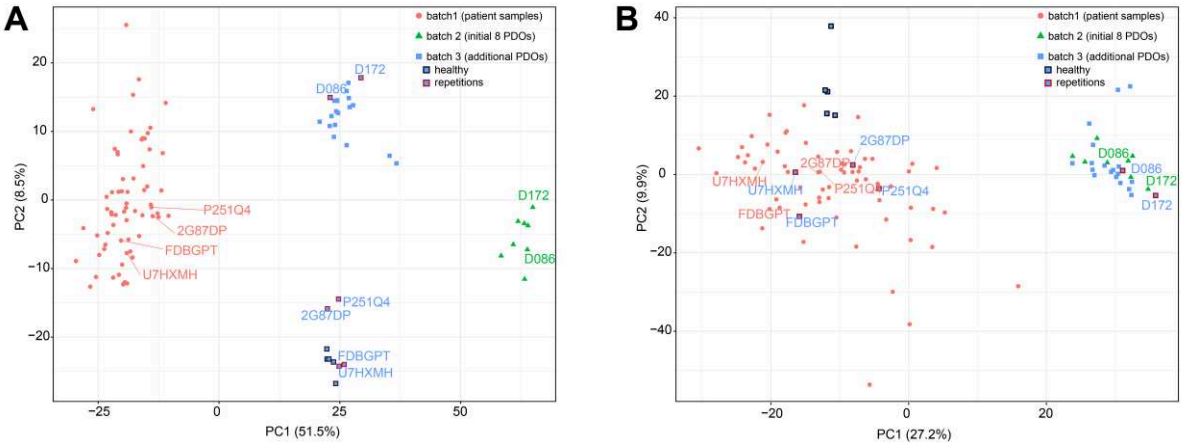


Figure 48: Batch effect removal (full proteome)
Full proteome, merged from the three batches, filtered for proteins identified in at least 50% of the samples. PCA. (A) after merge, (B) after batch effects removed.

Batch effect correction was also applied to the phosphoproteome data using the same strategy as for the total proteome, allowing for a more accurate comparison between samples and better integration with the findings from the CRC patient tissues. Since no RNA sequencing data were available, I computed the basket scores based on protein abundance and activity only, as described in Figure 45. For each PDO line the two highest positive scoring baskets were selected (Figure 49).

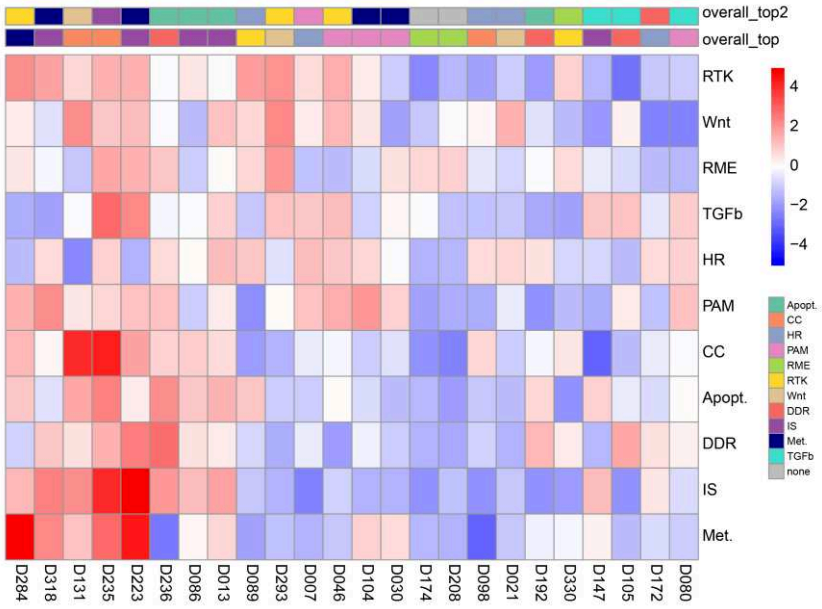


Figure 49: Proteome-based basket score applied to CRC PDOs
Full proteome and Phosphoproteome data were acquired for 24 CRC patient-derived organoids (PDOs) from the group of Johannes Betge. The basket scores were calculated as described previously (Figure 45) with the difference that no RNA sequencing data were available for the PDOs and the score was thus based on proteome level information only (full proteome ratios, ssGSEA, kinase activities, extended kinase activities). The final score values are illustrated as a heatmap and the overall highest and second highest scoring baskets are highlighted.

To test if this theoretical treatment recommendation is associated with response to drugs targeting the respective basket, I compared the final score with drug response data provided by Dr. Theresa Mulholland. Since the score represents an accumulation of pathway activity relative to the other PDOs, I also computed a relative drug response for easier interpretability. The 1-AUC drug response values were divided by the median response of all PDOs to that respective drug, followed by log2 transformation. PDO line D105 showed the highest relative score in the DNA damage response (DDR) basket (Figure 50A), did however respond poorly to nutlin3a treatment – a P53 inhibitor (Figure 50B). On the other hand, the lowest scoring basket for this PDO line was RTK signaling, but the cells responded exceptionally well to drugs targeting RTKs (VEGFR and EGFR

inhibitors). This suggests that the aggregated basket scores do not predict general response to drugs targeting proteins of that basket.

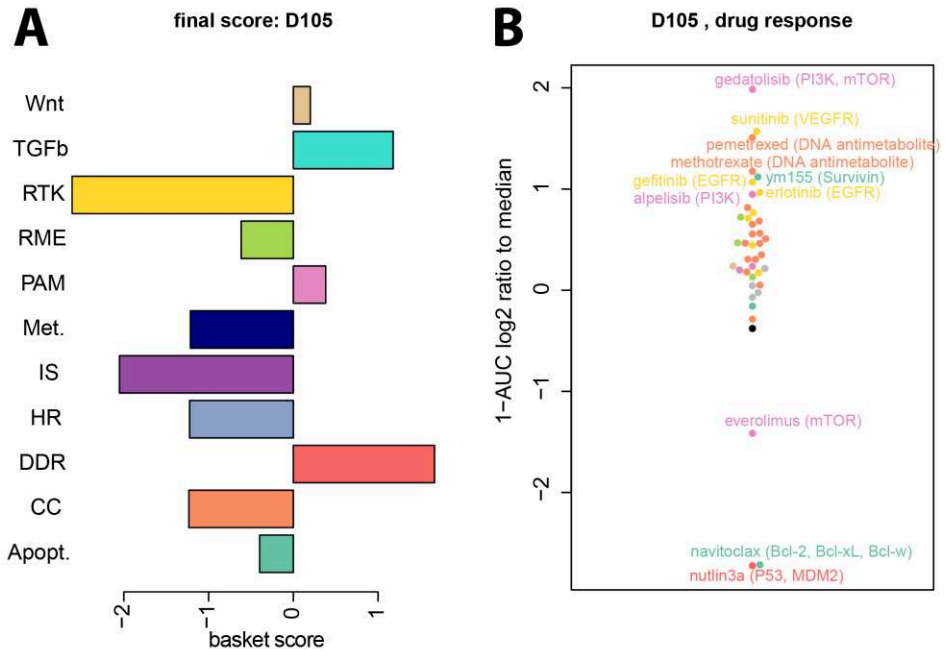


Figure 50: Example PDO: basket scoring and drug response
(A) Score value for each basket in the example PDO D105. (B) Response to a clinical drug library was measured by Dr. Theresa Mulholland with a SYBR green-based proliferation assay at different drug concentrations. The resulting area under the curve (AUC) of proliferation upon treatment with increasing drug concentrations demonstrates the sensitivity or resistance, respectively, of each PDO to the respective drug and was normalized to range between 0 and 1. I calculated 1-AUC values for every drug response of each PDO and filtered for drugs that showed some response (1-AUC > 0.3) in at least one PDO. To estimate the relative sensitivity of a given PDO, I calculated the log2 ratio of 1-AUC compared to the median response. Drugs with an absolute relative response ratio of larger than 1 are highlighted and the target protein or mode of action is indicated, color-coded based on the basket they target.

To provide a global correlation analysis of basket scores and drug response, I assigned each tested drug to one of the interventional baskets based on the drug target or mode of action. Drugs were excluded from this analysis when they were screened in less than 7 PDOs or if no PDO responded to it with a 1-AUC value > 0.3. I then performed Spearman correlation analyses between the relative response values (log2 ratio compared to the median response) and the relative basket scores. Interestingly, some basket scores, such as apoptosis, cell cycle and DNA damage response, showed high correlation with response to almost all screened drugs, independent of which basket is targeted by them (Figure 51). In contrast, high HR and PAM scores correlated negatively with response to almost any drug. This suggests that activities of certain baskets could predict whether a PDO line is generally a good responder or not. However, with an average of 13 data points the

numbers of screened PDO lines per drug are relatively small which means that a few outliers could bias the correlation between basket score and drug response.

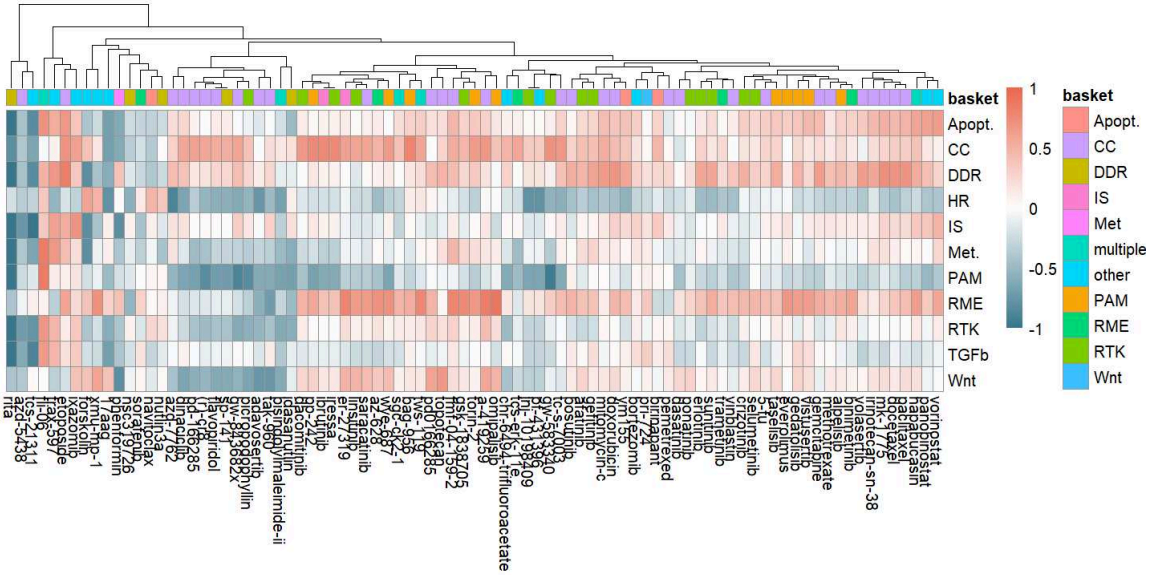


Figure 51: Correlation of PDO basket scores with drug response.
Drug response data and basket scores were filtered for drugs that were screened in at least 7 PDOs. Drugs were further filtered for response (1-AUC > 0.3) in at least one PDO, leaving a total of 75 drugs. To estimate the relative sensitivity of a given PDO, I calculated the log2 ratio of 1-AUC compared to the median response. Based on the target protein or mode of action, I manually assigned each drug to one of the baskets. Spearman correlation coefficients were calculated for every relative drug response and basket score.

To quantify whether basket scores hold any predictive value, I focused on drugs targeting the PI3K/AKT pathway (PAM) (Figure 52A, n = 10), RTK signaling (Figure 52B, n = 12), any protein involved in the cell cycle (Figure 52C, n = 28), or RAS-MEK-ERK signaling (Figure 52D, n = 5). I compared the Spearman correlation coefficients between the respective target score and the relative response to drugs targeting that same basket compared to all other drugs (Figure 52). None of the basket scores showed higher correlation with self-targeting drugs compared to the general drug response. This suggests once more that this sort of aggregated basket score holds no general predictive value for drugs targeting any protein of the respective pathway.

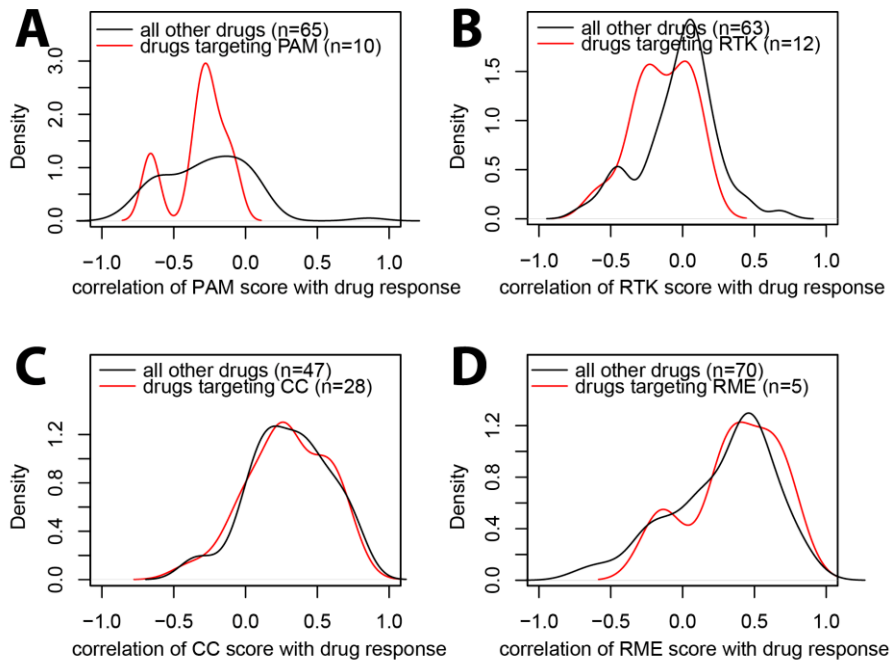


Figure 52: Correlation of basket score with drugs targeting proteins in that basket
Distribution of Spearman correlation coefficients between the relative basket score value and drugs targeting that same basket. For comparison, the distribution of Spearman correlation coefficients between the relative basket score value and any other drug, not targeting that same basket, is shown. (A) PAM basket, (B) RTK basket, (C) CC basket, (D) RME basket.

While an overall basket score might not predict response to any possible drug intervention in the same pathway, specific protein activities might: I hypothesized that, for example, the relative activity of a given kinase should represent some kind of reliance of the PDO line on that kinase and should thus predict a better response to drugs targeting that specific kinase. To test this hypothesis, I compared kinase activities (63 kinases with at least 16 activity values) and relative drug responses (87 drugs tested on at least 7 PDOs with at least 1 PDO responding 1-AUC > 0.3) across 19 PDO lines. For each pairing of drug and kinase, the PDOs were divided into three equally sized groups based on kinase activity (low, medium, high) and the drug response values (1-AUC) were visualized as boxplots. Significantly different drug responses were tested using two-sided Student's t-tests. One particularly interesting association was found between MTOR activity and the response to the proteasome-targeting drug Bortezomib (Figure 53A). PDO lines with high MTOR activity responded significantly better to Bortezomib treatment. I was further able to reproduce this result using a phosphoproteome dataset of CRC cell lines (Figure 53B).

The connection between mTOR activity and effectiveness of Bortezomib treatment has been previously shown in multiple myeloma (MM). Decaux et al. demonstrated that overexpression of REDD1, a negative regulator of mTOR, rendered MM cells resistant to Bortezomib treatment

(2010). A further mechanism through which mTOR activity influences bortezomib sensitivity appears to be through regulation of protein synthesis and cellular stress. Research on the YWHAE/14-3-3ε protein demonstrated that 14-3-3ε promotes translation initiation and protein synthesis in multiple myeloma cells by binding and inhibiting the TSC1/TSC2 complex and directly promoting phosphorylation of mTORC1. Higher 14-3-3ε expression correlates with increased proteasome inhibitor sensitivity, as it increases the protein load and consequent stress when proteasome function is inhibited (Xu et al. 2020).

Given the interaction between mTOR signaling and bortezomib response, combined targeting of both pathways presents a promising therapeutic strategy. A compound library screening study identified a synergistic effect between the mTOR inhibitor everolimus and bortezomib against osteosarcoma cells. This combination showed enhanced inhibition of cell proliferation compared to monotherapy with either drug alone, effectively induced apoptosis and interfered with cell cycle progression, and significantly reduced tumor growth in vivo (Asanuma et al. 2022).

My analysis suggests that elevated mTOR activity might play a similar role in colorectal cancer, increasing the sensitivity to proteasome inhibition and could serve as a therapeutic biomarker or combination treatment target.

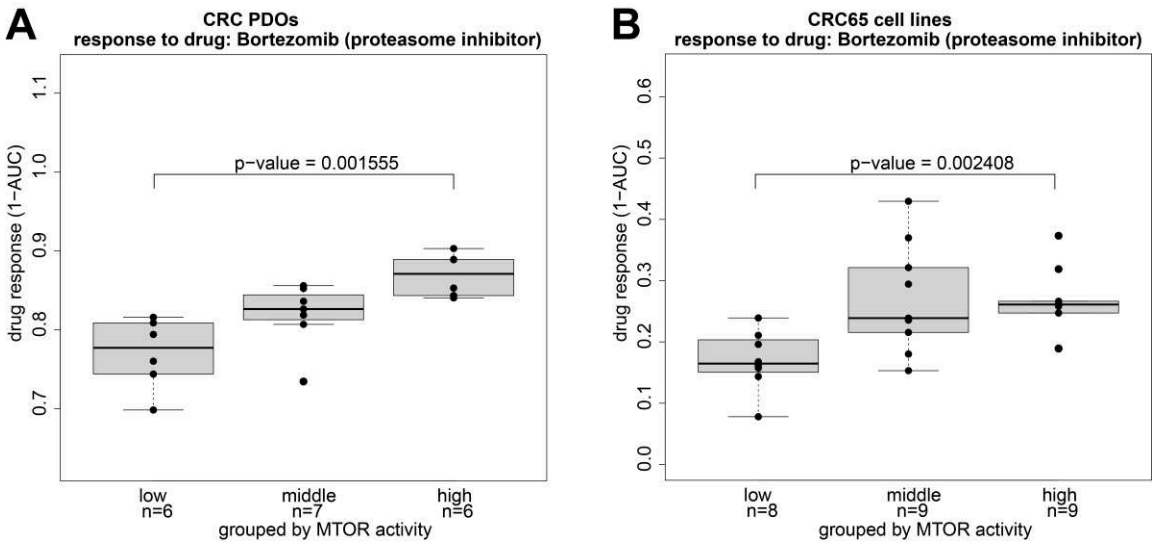


Figure 53: Response to Bortezomib grouped by mTOR activity
(A) Phosphoproteome data of 24 CRC PDOs, provided by the group of Johannes Betge, were acquired by DIA and filtered for phosphosites localized with a probability > 95% and identified in at least 50% of PDO samples. Since the PDOs were measured in two batches, batch effects were removed and variance stability normalization (VSN) was performed using the removeBatchEffect and normalizeVSN functions, respectively, of the limma R package. Kinase activities were calculated based on the Omnipath database and z-scored phosphopeptide intensities. Non-trustworthy kinases (Figure 23) were removed. Kinases with at least 16 activity values were kept, leaving 63 kinases. For each kinase the 24 PDOs were grouped into equally sized groups of low, intermediate and high kinase activity. Drug response data were filtered for drugs that were screened in at least 7 PDOs. Drugs were further filtered for response (1-AUC > 0.3) in at least one PDO, leaving a total of 75 drugs. The differential response to Borzetomib between the low and high activity patients was assessed using a two-sided Student's t-test. (B) Phosphoproteome and drug screening data of 65 CRC cell lines were obtained from Frejno et al. (2020). Phosphoproteome data were filtered for phosphosites identified in at least 50% of the cell lines. Relative kinase activities were computed based on the Omnipath database and z-scored phosphopeptide intensities. Non-trustworthy kinases (Figure 23) were removed. Drug response data and drug target annotations from CCLE (obtained via Frejno et al. (2020)) were available for 32 of these cell lines. I filtered for drugs that were tested on at least 20 cell lines, showed response in at least one cell line (1-AUC > 0.3), and were also screened in the Betge lab drug library, leaving 80 drugs. For each kinase the 32 cell lines were grouped into equally sized groups of low, intermediate and high kinase activity. The differential response to Borzetomib between the low and high activity patients was assessed using a two-sided Student's t-test.

3. Discussion

In the era of next generation sequencing, genomic and transcriptomic analyses of patient material to guide cancer treatment has been implemented and trusted for several years (D'Adamo et al. 2021; Kandoth et al. 2013). Proteomics on the other hand has so far mainly been used in the form of immunohistochemistry to validate transcriptomic biomarkers, such as the expression of HER2 in breast cancer (Duraiyan et al. 2012). The use of global mass spectrometry-based proteomics in the clinical setting has, however, only started to emerge. One reason for the hesitant implementation is that the technical advancements of proteomics measurements, with regards to sample turn-over time, cost, and measurement depth, were lagging behind genomics and transcriptomics for a long time (Mundt et al. 2023). Mass spectrometry-based proteomics has advanced significantly in recent years and in terms of speed, sensitivity and coverage has almost caught up to RNA-sequencing. New generations of mass spectrometers have improved speed and sensitivity (Bekker-Jensen et al. 2020). Automation of sample processing steps, like autoSP3 (Müller et al. 2020), has enabled faster turn-around times – a critical prerequisite when working with real-time clinical samples. Automated enrichment protocols for protein post translational modifications (PTMs) such as phosphorylation have extended the possibility of proteomics beyond measuring mere protein abundance. The development of data-independent acquisition (DIA) has revolutionized the way proteomics data are acquired, resulting in almost complete data sets and high proteome coverage (Ludwig et al. 2018). Finally, the development of AI-based software to analyze DIA raw data such as Spectronaut (Bruderer et al. 2015) has enabled the reliable identification and quantification of proteins.

With this present thesis I aimed to demonstrate the feasibility of proteomic and phosphoproteomic data acquisition from fresh-frozen cancer samples and how the data could be utilized to inform treatment decisions in a precision oncology setting. I established and optimized a protein extraction process, evaluated automated protein digestion and phosphopeptide enrichment workflows, and performed rigorous quality control experiments to ensure the reliability of the acquired data. Utilizing various bioinformatic tools, I gained biological insights into the characteristics of late-stage, pre-treated colorectal cancer and integrated multi-omics data on a per-patient basis to identify potential vulnerabilities.

3.1 Technical developments

3.1.1 Workflow automation is crucial for the adaptation of proteomics in the clinical setting

A major focus of my project was to evaluate the automated SP3 and phospho-enrichment protocols on the Agilent Bravo liquid-handling platform to allow for efficient and consistent protein clean-up, digestion, and enrichment. I demonstrated that autoSP3 results comparable to the manual version, with regards to the number of identified protein groups and the extent of contamination (Figure 4). The automated phosphopeptide-enrichment workflow allowed me to decrease the required sample input and the amount of hands-on labor ten-fold each (Figure 6), while reducing the cost per sample by 75% (Table 4). With this set-up I was able to identify and quantify up to 8,879 protein groups and 23,429 phosphosites (Figure 15). The number of proteins and phosphopeptides typically identified from fresh-frozen tissue samples varies significantly based on analytical platforms, tissue type, sample preparation methods, and analysis depth. Zhou et al. (2020) performed a proteome analysis of 246 fresh-frozen tissue samples across 16 cancer types, using a lysis buffer, as well as MS and DIA method settings that were highly comparable to my set up, and identified up to 7,505 protein groups per cancer type. Mertins et al. (2018) employed deep offline fractionation and TMT labelling and were able to identify >7.700 protein groups and >31,000 phosphosites from 300 µg peptides. Hence, the workflow I established was comparable with published protocols in terms of numbers of identified protein groups and phosphosites.

However, applying mass spectrometric analysis of samples with a measurement time of 120 minutes each for the full proteome and the phosphoproteome fraction, I was still limited to analyzing 6 patients per day. More advanced high-throughput LC-MS setups specifically designed for clinical applications, such as the Evosep One system, can analyze more than 200 samples per day while maintaining deep proteome coverage (Bache et al. 2018). The most advanced systems can process up to 1,000 samples per day. These capillary-flow LC/MS/MS systems maintain high sensitivity while achieving unprecedented speed. For example, a recent development enables analysis of 500+ samples per day while identifying more than 3,000 protein groups from less than 100 ng of human cultured cells (Tomioka et al. 2023). Orbitrap Astral instruments, the newest generation, enable analysis of up to 96 samples with short 24-minute gradients while identifying up to 10,000 proteins and 11,000 phosphopeptides from clinical FFPE samples (Haines et al. 2024). Micro-flow LC-MS systems using Orbitrap technology can achieve 11 proteomes in 16 hours using tandem mass tag multiplexing, identifying around 9,000 proteins and 30,000 phosphopeptides (Bian et al. 2020).

In conclusion, while the numbers of identified protein groups and phosphosites with my method are in the same range as recently published studies with a comparable set-up and the automated

sample preparation workflows allow for a high throughput, the measurement turn-around time could be improved in the future.

3.1.2 False-discovery rates can be controlled with mixed-species experiments

The probability-based identification (Perkins et al. 1999) of proteins by mass spectrometry, make it difficult to determine how trustworthy the acquired data are. Especially in the clinical setting, where protein measurements should be used to infer treatment decisions for cancer patients, it is crucial to know how likely it is that a certain protein was falsely identified.

To estimate and control the reliability of protein and phosphosite identification, I conducted mixed-species experiments by combining protein digests of human, yeast, and *E. coli* origin in different ratios. Searching the raw data against the databases of all included species, allowed me to quantify how many proteins were falsely identified in a sample that did not contain this species. Initially, I observed a 6% false discovery proportion in human protein identifications (Figure 9) but was able to reduce these by 80% by adjusting the search parameters of the Spectronaut software (Figure 10). Testing these optimized search settings on phosphoproteome data, I observed a false discovery proportion of only 0.3% (Figure 12). The developers of Spectronaut have in the meantime published a very similar strategy of using mixed-species samples to assess the false discovery rate and recommend stricter search settings that are very similar to my final settings (Table 9) (Baker et al. 2024).

In summary, I have established a workflow that could then be applied to clinical samples, allowing for highly sensitive detection of proteins and of phosphosites, and achieving high reliability.

3.2 The analyzed CRC cohort was highly heterogeneous, regarding both the sample quality and patient characteristics

I received, prepared, and measured the CRC tissue samples in two batches, eight months apart, and observed highly variable sample qualities with regards to proteome and phosphoproteome coverage (Figure 15). Samples of batch 2 performed worse in general, but I was unable to identify the source of these apparent quality differences. I was able to rule out sample storage time (Figure 16 C; D) as well as irregularities with the autoSP3, phospho-enrichment, and LC-MS/MS measurement processes (Figure 18; Figure 19) as determinants of sample quality. RNA sequencing data were not available for eight of the patient samples, due to RNA degradation, and these eight samples also showed significantly lower (phospho)proteome coverage (Figure 16 A; B). Therefore, my best guess was that the variance in proteome coverage represents sample-intrinsic quality differences, potentially due to sample handling, shipping, or ischemia time. It remains unknown why the samples from batch 2 generally performed worse than the ones from batch 1, but the same

appeared to be true for the mRNA-sequencing data that were missing exclusively from patients of batch 2. It would be interesting to investigate whether any changes in the tissue sampling process were made in the eight months between batch 1 and batch 2. I observed increased inclusion of samples from different NCT partner sites, such as Berlin or Munich, in batch 2 (Figure 16F) and I suggest to compare the tissue sample handling protocols between the NCT partner sites to ensure high quality standards. While I was able to show that useful information could still be drawn from low-coverage samples (Figure 21), I suggest to include a number of reference samples as quality controls in future cohort measurements. Besides the heterogeneity in sample quality, this CRC patient cohort was also characterized by highly heterogeneous disease states and clinical histories. Samples were obtained either from the primary tumor (11 cases) or from different metastatic sites (54 cases) with the metastasized organ largely undocumented (Figure 20). Additionally, each patient had undergone extensive prior treatments, including radiation, chemotherapy, and targeted therapies. With so many confounding factors potentially influencing the tumor biology, comparisons of patient subgroups should be treated with caution. Instead, most of my analyses were performed on a per-patient level, based on relative values across the entire cohort.

3.3 Biological insights from the proteomic characterization of CRC patient samples

3.3.1 KRAS-mutant CRC tumors have distinct proteomic and signaling profiles that could be clinically relevant.

KRAS mutations are found in approximately 40-50% of colorectal cancer cases (Ilhan et al. 2025; Osterlund et al. 2022; Radanova et al. 2023) and are associated with poor prognosis (Deng et al. 2015) and resistance to anti-EGFR therapies (Chang et al. 2009). Interestingly, my findings revealed that *KRAS* mutation status does not correlate directly with the activity levels of downstream RAF, MEK, or ERK proteins (Figure 40; Figure 41), suggesting that additional regulatory or compensatory mechanisms are involved in the activation of these pathways. My study demonstrated that *KRAS*-mutated CRC tumor cells exhibited heightened NF-κB, EGFR pathway, and GSK3β activities (Figure 42; Figure 44) as well as increased expression of TCA cycle proteins, such as PDK3 (Figure 43). These findings suggest new potential vulnerabilities of *KRAS*-mutated CRC tumors that might be exploited by targeted therapies.

NF-κB activation represents a promising therapeutic target in *KRAS*-mutant CRC, since NF-κB plays a crucial role in regulating the inflammatory tumor microenvironment, which is particularly relevant in *KRAS*-mutant CRC that typically shows reduced immune cell infiltration compared to BRAF-mutant tumors (Edin et al. 2024). Targeting NFκB could potentially reverse the immunosuppressive phenotype associated with *KRAS* mutations. Furthermore, NFκB activation has been implicated as a mechanism of resistance to KRAS-targeted therapies, suggesting that combination approaches including NFκB inhibitors might enhance the efficacy of KRAS inhibitors

(Xia et al. 2014). However, the clinical development of specific NFκB inhibitors has been challenging due to the pleiotropic effects of NFκB in normal tissues and potential toxicity concerns (Prescott and Cook 2018).

Inhibition of GSK3β could be a further vulnerability of *KRAS*-mutated CRC cells, since it has been shown to be essential for the survival of *KRAS*-mutated pancreatic cancer cells (Kazi et al. 2018), and is a key modulator linking KRAS signaling to the Wnt pathway (Li et al. 2005). The development of GSK3β inhibitors has however faced significant challenges and despite extensive research efforts by pharmaceutical companies and academic institutions, attempts to develop potent and selective GSK3β inhibitors for clinical use have been largely unsuccessful (Bhat et al. 2018). GSK3β plays crucial roles in normal cellular physiology, raising concerns about potential toxicity from systemic inhibition. The enzyme is involved in glycogen metabolism, insulin signaling, and Wnt pathway regulation in healthy tissues (Wadhwa et al. 2020). Disrupting these essential functions could lead to significant adverse effects in normal cells and indeed, the challenges in clinical GSK3β inhibitor development have been partly attributed to toxicity issues and safety margin considerations (Bhat et al. 2018).

The combination of sotorasib (a KRAS G12C inhibitor) with panitumumab (an EGFR inhibitor) has shown significant clinical benefit in patients with *KRAS* G12C-mutant metastatic CRC, with improved progression-free survival compared to standard care (Fakih et al. 2023). This approach addresses the feedback activation of EGFR signaling that occurs with KRAS inhibition alone. Similarly, *BRAF*-mutant (V600E) CRC patients respond better to BRAF inhibition when combined with EGFR inhibition, as there is a feedback activation from ERK1/2 on EGFR (Prahallad et al. 2012; Di Nicolantonio et al. 2021). In conclusion, there is increasing evidence that inhibiting the RAS-RAF-MEK-ERK axis in colorectal cancer is only effective when blocking EGFR signaling in addition. This strategy could be further tested in *KRAS*-mutated patients by, for example, combining

Combinations targeting both KRAS signaling and metabolic vulnerabilities, such as glucose transport (GLUT1) (Erber et al. 2019) or glutamine metabolism (Hu et al. 2023), have shown promise in preclinical models of KRAS-driven cancers. TCA cycle modulation could potentially exploit the metabolic vulnerabilities created by KRAS-driven metabolic reprogramming (Zhang et al. 2025).

The complex signaling and metabolic alterations in *KRAS*-mutant CRC suggest that multi-targeted approaches will likely be necessary for effective treatment. NFκB, EGFR, TCA cycle, and GSK3β all represent promising co-treatment targets that could address different aspects of KRAS-driven oncogenesis.

3.3.2 Traditional CRC patient classification does not apply to metastatic, pre-treated cohorts and three kinase activity-based patient subgroups were identified instead.

The consensus molecular subtype (CMS) classification system, while valuable for treatment-naïve primary colorectal cancer, faces significant limitations when applied to heavily pre-treated, metastatic disease. Recent studies demonstrate that only 64% of CRC metastases exhibit concordant CMS groups with matched primary tumors, with liver metastases showing predominant CMS2 classification while peritoneal metastases were enriched for the CMS4 type (Luo et al. 2024). This organ-specific alteration of CMS distribution suggests that the "seed and soil" hypothesis fundamentally influences molecular classification in metastatic disease. Furthermore, CRC has been shown to gain higher plasticity and lose colonic cell lineage identity during cancer progression, which causes profound changes compared to the primary tumor (Moorman et al. 2025; Cammareri et al. 2025).

The effects of systemic therapy on the tumor biology further complicate patient classification. Preoperative chemotherapy causes profound reprogramming of the tumor microenvironment, including activation of B cells, reduced diversity of tumor-associated macrophages with immature phenotypes, and accumulation of myofibroblasts (Harryvan et al. 2024). These treatment-induced changes create a molecular landscape that may no longer reflect the canonical CMS structure, necessitating alternative classification approaches for advanced disease (Rejali et al. 2023; Back et al. 2024).

Using kinase activity correlation analysis, I identified three subgroups of pre-treated, metastatic CRC patients that portrayed different molecular characteristics (Figure 34; Figure 35). Functional proteomics, particularly kinase activity profiling, provides direct readouts of cellular signaling states rather than static gene expression measurements, making them particularly valuable for understanding dynamic tumor biology in the context of treatment pressure. This approach has demonstrated success in other cancer types, with phosphoproteomics-based classification systems showing prognostic value that complements or exceeds traditional genomic classifications (Zagorac et al. 2018; Gámez-Pozo et al. 2017). Studies in triple-negative breast cancer have successfully used phosphoproteomics to build functional tumor taxonomies, identifying kinase activity patterns that predict treatment outcomes with greater accuracy than mutation-based classifications (Zagorac et al. 2018). Similarly, kinase activity profiling of peripheral blood mononuclear cells has shown predictive value for immune checkpoint inhibitor response in lung cancer patients (Hurkmans et al. 2020).

The three subgroups identified in my CRC cohort possibly represent distinct CRC subtypes not captured by traditional RNA-based CMS classification. Cluster 1 represents a proliferation-driven molecular phenotype characterized by high activities of cyclin-dependent kinases (CDK1, CDK2,

CDK4), casein kinase 2 (CSNK2), and GSK3β. This kinase activity profile corresponds with elevated transcription factor activities of the E2F family (E2F1-4) and MYC, master regulators of cell cycle progression and replication. The concordant upregulation of cell cycle, transcriptional, and translational machinery identified through Reactome pathway enrichment supports the aggressive proliferative nature of this subgroup.

Cluster 2 exhibits a distinct molecular profile characterized by high TGFβ signaling activity through SMAD2/3/4 transcription factors and increased TP53 activity. The reduced copy number alterations in Cluster 2 patients point to a more genomically stable profile, which may be linked to the observed enrichment for colon cancer (C18) samples rather than rectal cancer (C20). This anatomical distinction has clinical relevance, as colon cancer is less frequently treated with radiation therapy compared to rectal cancer (American Cancer Society), potentially explaining the lower radiotherapy exposure in this cluster. However, it remains unclear whether cluster 2 is a batch effect artefact (Figure 35) or the biological consequence of more frequent colon cancer patient enrolment, rather than rectal cancer in batch 2.

Cluster 3 demonstrates the most pronounced immune-infiltrated and inflammation-associated phenotype, with active transcription factors including IRF1, IRF4, NFκB, and REL, alongside elevated JAK-STAT, TNFα, and NF-κB pathway activities. This molecular signature is supported by protein-level enrichment for immune-related pathways including adaptive immunity, cytokine signaling, and interferon responses. The elevated immune infiltration in Cluster 3 may indicate active anti-tumor immune responses that could be enhanced through checkpoint inhibition or other immunomodulatory strategies.

Kinase activity estimation has demonstrated remarkable robustness even with limited sample quality or coverage. Systematic down-sampling studies show that kinase activity scores maintain high reproducibility (Pearson $r > 0.8$) even with 80% of phosphosites removed, indicating that meaningful functional insights can be obtained even from degraded clinical samples (Figure 21). This robustness is crucial for clinical implementation, where sample quality can be highly variable. The direct connection between kinase activities and drug targets makes this classification system particularly relevant for precision oncology. More than 95% of FDA-approved drugs target proteins, and kinase inhibitors represent one of the largest classes of targeted therapies (Santos et al. 2017). By classifying patients based on the functional activity of druggable targets, this approach provides a more direct path to therapeutic intervention compared to gene expression-based classifications.

3.3.3 Integrated basket-based score did not predict drug response in patient-derived organoids

The idea of developing a basket-based score that integrates different molecular information such as transcription factor and kinase activities, was to identify enriched pathway activations in certain interventional baskets that might hint at a vulnerability of the tumor (Figure 45). However, when applied to CRC PDO lines (Figure 49) the basket score did not correlate with increased response to drugs targeting that respective basket (Figure 52). This likely underscores the complexity of pathway rewiring and compensatory mechanisms in cancer. However, PCA of the full proteome showed that patient tissue samples clustered more closely with the healthy colon biopsies than the CRC PDO lines (Figure 48), suggesting that CRC tumor samples share greater proteomic similarity with healthy tissue than with PDOs, challenging the utility of the latter as a 1:1 tumor model for drug testing. While patient-derived organoids offer an accessible and scalable platform for ex vivo drug screening, PDOs diverge significantly from the originating tumors (Shi et al. 2024; Bozym et al. 2025; Chan et al. 2023). This is likely due to the absence of tumor microenvironment components, including immune and stromal cells, which are critical for both tumor behavior and drug response (Habič et al. 2022; Shi et al. 2024; Polak et al. 2024). Moreover, culture conditions may induce metabolic and signaling adaptations not present in vivo (Zeng et al. 2023). These limitations suggest that PDOs, in their current form, are insufficient proxies for intact tumors and should be interpreted with caution in therapeutic screening efforts. Recent advances in organoid technology aim to address these limitations through co-culture systems where PDOs are cultured with immune cells and fibroblasts to reconstitute the TME (Polak et al. 2024; Chan et al. 2023). Additionally, organoids-on-a-chip based on microfluidic technology and assembloid approaches that fuse tumor organoids with other organoid types are showing promise for more accurately recapitulating the in vivo environment (Zeng et al. 2023; ha et al. 2024).

In conclusion, it is currently difficult to assess whether the non-correlation of basket scores with drug responses represents true biology – meaning the complexity of pathway regulations and compensations is too high to capture vulnerabilities with this method – or stems from the inadequate representation of the in vivo CRC tissue by organoids.

3.3.4 Protein-level data adds complementary information to genomics and transcriptomics

My study showed that relying solely on DNA copy number or mRNA levels can mislead therapeutic decisions, as highlighted by the discrepancies in CDH17 and ERBB2 expression at the protein level despite high gene copy numbers (Figure 28A, C). Both these genes were frequently used as biomarkers to recommend treatment with drugs targeting these proteins, but the gene copy

numbers showed little to no correlation with the protein levels (Pearson's r = 0.12 and -0.22 respectively).

Furthermore, in the case of KRAS, one of the most important biomarkers that characterizes colorectal cancer, measuring gene copy number levels was not a reliable predictor of increased protein abundance, as at least 25 gene copies were needed to see the effect on the protein level (Figure 28B). These examples demonstrate that relying on genomic information is insufficient and had the protein levels been known, the treatment recommendations given by the molecular tumor board would have likely been different.

Even mRNA levels, although demonstrating higher correlation, does not necessarily directly predict protein levels. For some genes, for example EGFR, the correlation between mRNA and protein levels is strong (Pearson's r = 0.71, Figure 31C) suggesting that transcriptomic measurements might be sufficient, at least in this cohort of CRC patients, to reliably predict protein expression. However, for many other genes, for example genes involved in respiratory electron transport or TCA cycle, the mRNA and protein abundances appear to be not correlated at all (Figure 30).

Values from literature that studied mRNA-protein correlation support these findings. Across large cancer datasets, the Pearson correlation coefficient between mRNA and protein abundance for matched genes typically ranges from 0.3 to 0.6. This means that, on average, only about 10–36% of the variance in protein levels can be explained by mRNA levels alone (Wang et al. 2013; Maier et al. 2009; Fortelny et al. 2017).

Many reasons explain why mRNA and protein levels are not directly correlated, including post-transcriptional regulation, protein stability differences, and translational control mechanisms (Liu et al. 2016). Depending on the gene(s) of interest, and likely depending on the disease context, mRNA sequencing might suffice or proteomics measurements are necessary to reveal the true protein abundance, respectively.

3.4 The lack of proteomics-based clinical studies

This study underscores the potential of proteomics to enhance clinical decision-making in precision oncology. While genomic and transcriptomic data are routinely used to guide treatment recommendations, protein activity, modulated by post-translational modifications, offers a more direct readout of cellular state. Integrating proteomic and phosphoproteomic information into molecular tumor board (MTB) discussions could help refine therapy stratification, particularly in complex or ambiguous cases. By identifying protein-level discordances with DNA and RNA data, our findings argue for a paradigm shift toward multi-layered molecular diagnostics in routine oncology practice.

The field of mass spectrometry-based proteomics faces a significant translational challenge in clinical medicine, primarily due to a stark paucity of dedicated clinical trials that would enable rigorous evaluation of its diagnostic and therapeutic value (He 2019; Whiteaker 2010). Despite the tremendous potential of proteomic biomarkers to provide crucial insights into disease mechanisms and treatment responses, the application of proteomics in early-phase clinical development remains minimal (Nandal and Burt 2017). This scarcity of mass spectrometry proteomics-based clinical trials creates a fundamental barrier to establishing the clinical utility and economic value of these sophisticated analytical approaches (Migneault and Hunter 2009; Brown et al. 2020). The challenges inherent in implementing proteomics in clinical settings include high biological variability in patient populations, diverse disease stages, nonuniform responses to therapy, and special requirements for handling samples from clinical trials (Migneault and Hunter 2009; Maes et al. 2019). Additionally, the lack of efficient technologies for verification and validation of proteomic biomarkers has significantly hindered their translation from discovery to clinical application (Whiteaker 2010; Hristova and Chan 2019). The complex nature of proteomic data interpretation, requiring specialized bioinformatics platforms and computational expertise, further compounds these implementation challenges (Guingab-Cagmat et al. 2013; Brown et al. 2020). In stark contrast, genomic and transcriptomic biomarkers have achieved widespread adoption in clinical trials, with comprehensive genomic profiling becoming routine practice in oncology (Dienstmann et al. 2015; Cristescu et al. 2018). The success of genomic biomarkers is exemplified by their extensive use in immunotherapy trials, where tumor mutational burden and gene expression profiles serve as established predictive markers for treatment response. Clinical trial designs targeting patient subgroups based on genetic characteristics have become commonplace, with molecular prescreening programs processing thousands of patients annually to match genomic alterations with targeted therapies. The regulatory pathway for genomic biomarkers is well-established, with standardized reporting frameworks and validated diagnostic platforms that facilitate clinical implementation. This disparity between proteomics and genomics creates a critical knowledge gap, as the limited number of proteomic clinical trials makes it extremely difficult to assess the true clinical value, cost-effectiveness, and patient outcomes associated with mass spectrometry-based protein biomarkers. The lack of robust clinical evidence for proteomic approaches ultimately impedes their integration into precision medicine strategies, despite their potential to capture functional information that genomic alterations alone cannot provide (Vinaiphat et al. 2021). Consequently, the field urgently needs increased investment in proteomic clinical trials and standardized validation frameworks to unlock the full potential of this powerful analytical technology for improving patient care.

3.5 Limitations of my study

One key limitation of this study lies in the high heterogeneity of the patient cohort. While some tissue samples were derived from the primary tumor (n = 11), most were obtained from different metastatic sites (n = 54) and the location of the metastases was in most cases not documented. The cohort was furthermore marked by extensive and heterogeneous prior treatment. As part of the NCT-MASTER program, patients presented after years of heavy pre-treatment, having received up to 11 different therapeutic agents per patient. Treatments included chemotherapies like 5-FU, irinotecan, and oxaliplatin, radiation therapy, and targeted agents, such as anti-VEGFR, -EGFR treatments, or immune checkpoint inhibitors. No two patients shared the exact same pre-treatment regimen, which likely led to specific treatment-induced signaling adaptations. This heterogeneity may obscure more direct genotype-to-phenotype relationships, particularly in kinase activity and pathway activation.

Furthermore, technical limitations obscure the comparability of proteome data with the genomic and transcriptomic data acquired by the NCT MASTER program. Since the proteome was not acquired from the exact same tissue piece as genome sequencing and RNA-seq, there could be potential spatial differences between the samples that cannot be accounted for in the analysis. Another technical limitation is that the proteome data, similarly to the transcriptome data, was not adjusted according to the estimated tumor purity. Therefore, it remains unclear whether the expression of certain proteins of interest actually comes from the cancer cells themselves or possibly from the tumor microenvironment.

Protein activity estimation is currently only possible for transcription factors and kinases, both of which are not directly measured but estimated from footprints (Türei et al. 2021; Badia-I-Mompel et al. 2022; Müller-Dott et al. 2023). This indirect measurement approach introduces uncertainty that must be considered when interpreting functional insights from proteomic datasets. Future developments should focus on expanding the range of proteins for which activity can be reliably inferred and improving the accuracy of these predictions.

3.6 Outlook

Future work should focus on validating proteomic biomarkers and pathway activities in prospective clinical studies, ideally integrating proteomics into the NCT-MASTER pipeline for real-time clinical decision support. The development of fast, standardized, clinically certified phosphoproteomic workflows will be essential to ensure reproducibility and scalability. Current bottlenecks lie in the time-intensive MS measurement and downstream bioinformatics analysis – future efforts should focus on making these steps more efficient.

To improve the translational relevance of patient-derived organoids for predicting drug responses, co-culture or microfluidic models that preserve immune and stromal interactions could be implemented.

Longitudinal sampling may also reveal dynamic changes in signaling activity that could inform adaptive treatment strategies. However, repeated sampling of tissue biopsies will likely be difficult to realize in clinical practice, and might therefore be limited to measurement of liquid biopsies.

The clinical implementation of kinase activity-based patient classification will require extensive validation in larger, independent cohorts. Key areas for validation include the reproducibility of cluster assignments across different analytical platforms, the prognostic significance of each cluster, and the predictive value for specific therapeutic interventions. Furthermore, extending the knowledge database of kinase-substrate relationships – either experimentally or computationally – would make kinase activity estimates more robust and less biased for extensively researched proteins. Estimating protein activities based on the downstream “footprint” is currently only possible for transcription factors - by looking at the transcription of known target genes - and for kinases – by looking at the phosphorylation of known substrate sites. Extending this type of activity measurement to other protein classes would improve the functional readout of proteomics data.

Future developments should focus on integrating kinase activity-based classification with existing molecular biomarkers and clinical parameters. This multi-dimensional approach may provide more robust patient stratification than any single molecular feature alone. My data suggest that with stringent quality controls, proteomics is ready to be routinely applied in the clinical setting and the combination of functional proteomics data with genomic, transcriptomic, and clinical variables may represent the future of precision oncology.

4. Materials & Methods

Table 2: List of used products

Name	Order number	Vendor
Urea	1084880250	MILLIPORE
Trizma Base	T1503	Sigma-Aldrich
Sodium vanadate (NaVO4)	SIAL590088-25G	Sigma-Aldrich
Sodium pyrophosphate (H2Na2P2O7)	33385.22	ALFA
PMSF	PMSF16-S-50	Alpha Diagnostic International
Sodium fluoride (NaF)	2618.1	Carl ROTH
Beta-glycerophosphate	L03425.14	ALFA
Halt protease inhibitor cocktail	PIER87786	THERMO SCIENTIFIC PIERCE
EDTA	PIER87786	THERMO SCIENTIFIC PIERCE
Trypsin	V5280	Promega
LC/MS-grade H2O	10728098	Thermo Fisher Scientific
Acetonitrile	10001334	Thermo Fisher Scientific
TFA	00202341A8BS	Biosolve BV
FA	00069141A8BS	Biosolve BV
Ethanol	1009831000	Merck Milipore
TEAB	T7408-100ml	Sigma-Aldrich
Magnetic beads A	45152105050250	Ge Healthcare
Magnetic beads B	65152105050250	Ge Healthcare
CAA	C0267-100g	Sigma-Aldrich
TCEP	C4706-2G	Sigma-Aldrich
Lo-bind reaction tubes	04-212-3500	Nerbe Plus
96-well plates	017219	Kisker Biotech
Fe(III) cartridges	G5496-60085	Agilent
Commercial yeast digest	A47951	Life Technologies
Commercial E. coli digest	186003196	Waters
Commercial HeLa digest	15245733	Thermo Fisher Scientific

4.1 Cell line lysates used for method development

MCF7 cells were cultivated in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% FCS, 50 units/mL penicillin, 50 µg/mL streptomycin sulfate (Invitrogen AG, Carlsbad, CA, USA), and 10−8 M 17-ß-estradiol (E2, Sigma-Aldrich, Saint-Louis, MI, USA) at 37°C with 5% CO2 in a humidified incubator. were lysed in RIPA buffer (Thermo Fisher Scientific) that was supplemented with 10 mM NaF, 1 mM Na3VO4, cOmplete EDTA-free protease inhibitor (Merck), PhosSTOP phosphatase inhibitor (Merck), 250 U/mL Benzonase (Merck), and 10 U/mL DNase (Qiagen). Lysates were kept on ice for 30 minutes and then centrifuged for 30 minutes at 15,000xg and 4°C. Supernatants were then transferred to fresh tubes. Protein concentrations were determined using the Pierce BCA assay (Thermo Fisher Scientific) and lysates were stored at -20°C.

4.2 Tissue lysis

Tumor tissue samples were collected within the DKFZ/NCT/DKTK MASTER program after written informed consent for biobanking and molecular analysis, approved by the Ethics Committee of Heidelberg University (protocol number S-206/2011). Fresh-frozen tissue samples were cut into 30 µm thick slices by the NCT Sample Processing Lab (SPL) and stored at -80°C. For protein extraction, the tissue slices were kept on dry ice, 200 µl of lysis buffer (Following sonication, lysates were centrifuged for 5 minutes at 4°C and maximum speed (14,000 g). The supernatants were transferred to new tubes and protein concentration was determined using Pierce Protein BCA assay (Thermo Fisher Scientific). Lysates were stored at -20°C until further use.

Table 3) was added, and the samples were subjected to mechanical disruption as fast as possible. Mechanical disruption of the tissue was performed using the TissueLyser II (Qiagen) at 20 Hz for two times two minutes. Subsequently, to shear DNA and RNA and to enhance protein extraction, 50 mg sonication beads were added and the lysates were sonicated using the Bioruptor Pico (Diagenode) device at 4°C for 5 cycles (if not stated otherwise) of 30 seconds on, 30 seconds off.

Additional lysis buffers that were used in this study include:

- RIPA lysis buffer for cell line and PDO lysis (commercial RIPA buffer (Thermo Fisher Scientific) supplemented with 10 mM NaF, 1 mM Na3VO4, cOmplete EDTA-free protease inhibitor (Merck), PhosSTOP phosphatase inhibitor (Merck))
- Urea-based lysis buffer for tissue lysis (8 M urea, 100 mM NaCl, 50 mM TEAB, cOmplete EDTA-free protease inhibitor (Merck) and PhosSTOP phosphatase inhibitor (Merck))
- SDS-based lysis buffer for tissue lysis (4% SDS in PBS)

Following sonication, lysates were centrifuged for 5 minutes at 4°C and maximum speed (14,000 g). The supernatants were transferred to new tubes and protein concentration was determined using Pierce Protein BCA assay (Thermo Fisher Scientific). Lysates were stored at -20°C until further use.

Table 3: Composition of the final lysis buffer

Recipe kindly provided by Dr. Ali Azmal Syed.

Compound	Stock concentration	End concentration
Urea	60.06 g/mol	8 M
Sodium chloride (NaCl)	58.44 g/mol	75 mM
Tris pH 8.2	1 M	50 mM
Sodium vanadate (NaVO4)	1 M	1 mM
Sodium pyrophosphate (H2Na2P2O7)	221.94 g/mol	1 mM
PMSF	100X	1 mM
Sodium fluoride (NaF)	1 M	1 mM
Beta-glycerophosphate	306.12 g/mol	1 mM
Halt protease inhibitor cocktail	100X	1X
EDTA	100X	1X

Protein clean-up and tryptic digestion

4.3.1 Manual SP3

Magnetic beads mixture was prepared by combining 20 µl beads A with 20 µl beads B, washing three times with 200 µl H2O, and finally resuspending the bead slurry in 20 µl H2O. From the cell or tissue lysate, 20 µg protein were used as input and the volume was filled up to 60 µl with lysis buffer. Reduction and carbamidomethylation of disulfide bridges was performed by adding 40 mM CAA and 10 mM TCEP, followed by incubation at 95°C for 5 minutes (37°C for 1 hour in case of urea-based lysis buffers). After returning to room temperature, 2 µl beads preparation were added, followed by ethanol to a final concentration of 50%. Protein precipitation on the magnetic beads was performed at room temperature, while shaking at 800 rpm, for 15 minutes. Subsequently, the samples were placed on a strong magnet, and the supernatant was removed. The proteins precipitated on the beads were washed twice with 200 µl 80% ethanol and once with 200 µl acetonitrile. Beads were resuspended in 75 µl digestion buffer (100 mM TEAB and 0.4 µg/µl Trypsin in H2O) and sonicated briefly in a water bath. Digestion was performed at a trypsin-to-protein ratio of 1:25, at 37°C, while shaking at 650 rpm, for 16 hours. The following day, the supernatant was transferred to a new

tube and peptides were vacuum centrifuged to dryness (30°C, 1200 rcf). Peptides were stored at -20°C until further use.

4.3.2 Automated SP3

Automated SP3-based protein clean-up and digestion was performed on the Agilent Bravo liquid handling robot, based on the protocol established by Müller et al. (2020) and further optimized by Yannik Dieter and Dr. Barbara Helm. Briefly, magnetic beads, CAA and TCEP solutions were prepared similarly to the manual workflow (4.3.1). From each tissue lysate 20 µg protein were placed into the Bravo system in a 96-well plate. The reduction, carbamidomethylation, protein binding, and washing steps were performed by the liquid handling robot. In the last step, the protein-bound beads were dissolved in 75 µl 100 mM TEAB and I manually added 0.8 µg Trypsin to each sample. Digestion was performed at 37°C, while shaking at 650 rpm, for 16 hours. The following day, the supernatant was transferred to a new tube and peptides were vacuum centrifuged to dryness (30°C, 1200 rcf). Peptides were stored at -20°C until further use.

4.3.3 In-gel

For in-gel clean-up and digestion, 10 ug of protein lysate was used, unless noted otherwise for a respective experiment. 4x LSD running buffer (Thermo Fisher, #NP0007) and 10x reducing agent (Thermo Fisher, #NP0009) were added and the samples were heated to 70°C for 10 minutes. Subsequently, samples were loaded onto NuPAGE Bis-Tris (4-12%) gels (Life Technologies, #NP0321BOX) and placed in MES buffer (Life Technologies, #NP0002). An electric field of 100 V was applied for 9 minutes letting proteins migrate 0.5 cm into the gel. Gels were stained in Coomassie Brilliant blue R-250 for 3 hours and the 0.5 cm long protein band was cut out. The band was cut into 6 smaller pieces for further processing.

100 µl water was added to the gel pieces and incubated for 5 minutes at 37°C and 600 rpm. The supernatant was discarded and the gel pieces were washed in 100 µl 50% acetonitrile for 5 minutes at 37°C and 600 rpm. For reduction of disulfide-bridges, 100 µl 10 mM DTT was added and incubated for 60 minutes at 56°C and 600 rpm. Subsequently, gel pieces were again washed with 100 µl water. For alkylation, 100 µl 55 mM IAA was added and incubated for 30 minutes in the dark at 25°C and 600 rpm. Gel pieces were washed three times each with water and 50% acetonitrile as described previously. To de-hydrate the gel pieces, 100% acetonitrile

was added, incubated for 1 minute at room temperature and the supernatant was discarded. This step was repeated once. Gel pieces were air-dried for 15 minutes. Subsequently, 15 µl (0.25 µg) of Trypsin (dissolved in 50 mM TEAB) was added and soaked up by the gel pieces. Once the Trypsin was soaked into the gel pieces, they were covered with further 50 µl of 50 mM TEAB. Digestion was performed for 16 hours at 37°C and 500 rpm.

For extraction of peptides from the gel pieces, 50 µl of 50% acetonitrile with 0.1% TFA was added and incubated in the sonication bath for 5 minutes. The supernatant was collected in a fresh tube. This procedure was repeated with 100% acetonitrile, 0.1% TFA again 100 % acetonitrile. The collected supernatants were vacuum centrifuged to dryness and the dried peptides were stored at -20°C until measurement.

4.3.4 PreOmics iST kit

Used iST kit from PreOmics (P.O.00001) and followed the application note for lysis of tissue. Briefly, I placed a roughly 3 mm x 3 mm x 3 mm tissue piece into a 2 ml tube on dry ice, added a steel bead and, at room temperature, added 200 ul LYSE buffer from iST kit. Mechanical tissue disruption was performed using the TissueLyer II (Qiagen) at 20 Hz for two times two minutes. Lysates were centrifuged briefly and transferred to Diagenode sonication tubes together with 50 mg sonication beads. Five sonication cycles with the Bioruptor Pico were applied to the lysates at 4°C. Subsequently, lysates were centrifuged (5 minutes, 8°C, 15,000 rpm) and supernatants were transferred to new low-bind tubes. Samples were heated to 95°C for 10 minutes, followed by brief centrifugation. After returning to room temperature, 50 µl of the lysate were loaded onto the iST cartridge. For digestion, 50 µl of the DIGEST buffer was added to the cartridge, followed by incubation at 37°C for three hours. To stop the digestion, 100 µl of STOP buffer were added and the cartridges were centrifuged at 3,800 g for 90 seconds. Samples were washed with 200 µl WASH1, followed by 200 µl WASH2 and eluted with 100 µl ELUTE. The elution step was repeated once. Eluted peptides were vacuum centrifuged to dryness (45°C) and stored at -20°C until further use.

4.3.5 In-solution digestion

Ten µg of protein were precipitated according to Wessel and Flügge (1984), protein pellets were dried for 15 minutes at room temperature. For tryptic digestion, protein pellets were dissolved in 8 M urea, supplemented with 100 mM NaCl, 50 mM TEAB, cOmplete EDTA-free protease inhibitor (Merck) and PhosSTOP phosphatase inhibitor (Merck). Disulfide

bridges were reduced with 10 mM DTT for 1 hour at 27°C, then alkylated with 30 mM IAA for 30 minutes at room temperature. Reactions were quenched with 10 mM DTT and incubation for 15 minutes at room temperature.

Lysyl endopeptidase (LysC, Fujifilm) was added in a 1:100 enzyme-to-protein ratio, and proteins were digested for 4 hours at 30°C. Then, the urea concentration was diluted to 1.6 M with 50 mM TEAB, Trypsin was added in a 1:50 ratio and the digestion continued for 16 hours at 37°C. Formic acid (FA) was added to a final concentration of 2% (v/v) to terminate digestion.

4.4 Enrichment of phosphorylated peptides

4.4.1 Manual phosphopeptide-enrichment

Five hundred µg of protein were precipitated (Wessel and Flügge 1984) and the protein pellets were dried for 15 minutes at room temperature. For tryptic digestion, protein pellets were dissolved in 8 M urea, supplemented with 100 mM NaCl, 50 mM TEAB, cOmplete EDTA-free protease inhibitor (Merck) and PhosSTOP phosphatase inhibitor (Merck). Disulfide bridges were reduced by adding 10 mM DTT and incubating for 1 hour at 27°C, then alkylated with 30 mM IAA for 30 minutes at room temperature. Reactions were quenched with 10 mM DTT for 15 minutes at room temperature.

Lysyl endopeptidase (LysC, Fujifilm) was added in a 1:100 enzyme-to-protein ratio, and proteins were digested for 4 hours at 30°C. Then the urea concentration was diluted to 1.6 M by adding 50 mM TEAB, then trypsin was added in a enzyme-to-protein ratio of 1:50 and the digestion continued for 16 hours at 37°C. To terminate digestion, formic acid (FA) was added to a final concentration of 2% (v/v).

Subsequently, peptides were desalted using Sep-Pak C18 cartridges (Waters). Briefly, the cartridges were conditioned with 100% acetonitrile, washed with 80% acetonitrile containing 0.6% acetic acid and equilibrated twice with 2.5% formic acid. Peptide solutions were loaded, and the flow-through was collected and loaded a second time. Bound peptides were washed four times with 2.5% formic acid and then eluted twice with 80% acetonitrile containing 0.6% acetic acid. The eluate was vacuum centrifuged to dryness and stored at -20°C.

For enrichment of phosphorylated peptides, an Immobilized Metal Affinity Chromatography (IMAC) column (Thermo Fisher Scientific) was charged with 25 mM iron chloride ($FeCl_3$) in 100 mM acetic acid at a flow rate of 0.2 mL/min for 30 minutes. Then, the column was rinsed with 0.1 % formic acid at a flow rate of 4 ml/min for 4 hours. Dried peptides were reconstituted

in 30% acetonitrile with 0.07 % TFA, and were loaded onto the IMAC column at a flow rate of 0.2 mL/min. Elution of phosphorylated peptides was achieved with a gradient of ammonium (NH_4OH). The phosphorylated peptide fraction was collected and vacuum centrifuged to dryness.

Finally, the enriched phosphopeptides were desalted by using Stop and Go Extraction (STAGE) tips. Briefly, three layers of C18 material (Supelco) were placed in a pipette tip, pre-wetted with methanol, washed with 80% acetonitrile in 0.1% trifluoroacetic acid (TFA) and equilibrated with 0.1% TFA. The dried peptides were reconstituted in 0.1% TFA and loaded onto the STAGE tip. Subsequently, the peptides were washed with 0.1% TFA, followed by two elution steps with 60% and 80% acetonitrile, respectively. The eluate was vacuum centrifuged to dryness and stored at -20°C until further analysis.

4.4.2 Automated phosphopeptide-enrichment

For automated phosphopeptide-enrichment, protein clean-up and digestion was performed with 55 µg protein material as input following the automated SP3 workflow adapted from Müller et al. (2020) on the Agilent Bravo liquid handling robot. Proteins were digested with Trypsin in a protease-to-protein ratio of 1:25 for 16 hours at 37°C. Peptides were vacuum centrifuged to dryness and subsequently enriched for phosphorylation on the Agilent Bravo liquid handling robot following the Agilent application note (https://www.agilent.com/cs/library/applications/5991-6073EN.pdf). Briefly, peptides were dissolved in 20 µl 0.1% TFA in H2O, followed by addition of 80 µl 0.1 TFA in acetonitrile. Peptides were transferred into a 96-well plate and placed into the Bravo robot. Fe(III) cartridges were primed with 100 µl Priming buffer (50 % ACN/0.1 % TFA) at a flow rate of 300 µl/min. Subsequently, the cartridges were equilibrated with 50 µl Equilibration and internal cartridge wash buffer (80 % ACN/0.1 % TFA) at a flow rate of 10 µl/min. Then, the peptides were loaded onto the cartridges at a flow rate of 3.5 µl/min. The cartridges were washed with 75 µl Equilibration and internal cartridge wash buffer (80 % ACN/0.1 % TFA) at a flow rate of 10 µl/min and the flow-through fraction was collected in a 96-well plate. Lastly, the phosphopeptides were eluted with 20 µl Elution buffer (1 % ammonium hydroxide) at a flow rate of 5 µl/min and the fraction was collected in a second 96-well plate. Both fractions were vacuum centrifuged to dryness and stored at -20°C.

Table 4: Cost comparison of automated and manual phosphopeptide-enrichment

autoPhospho (96 samples a 50 µg)			**Manual phosphopeptide-enrichment (96 samples a 500 µg)**		
Amount	Item	Cost per sample (Euros)	Amount	Item	Cost per sample (Euros)
5	96-well plates	0.23	800	eppis	0.732
30 mg	TCEP	0.00051	170 ml	methanol	0.019
30 mg	CAA	0.00006	20 ml	Chloroform	0.031
500 µl	Magnetic beads	0.23	20.2 liters	H2O	1.41
50 ml	ethanol	0.02	10 g	urea	0.0088
50 ml	Acetonitrile	0.012	120 mg	NaCl	0.00000276
1 ml	TEAB	0.0194	4.75 ml	TEAB	0.092
200 µg	Trypsin	1.5	2 pills	Protease inh.	0.167
2	Packages Bravo tips	0.124	2 pills	Phosphatase inh.	0.2195
100	loBind tubes	0.0916	60 mg	DTT	0.01344
1 pack	Fe-cartridges	12.15	110 mg	IAA	0.0154
20 ml	H2O	0.0014	500 µg	LysC	41.25
1 ml	ammonia	0.00124	1000 µg	Trypsin	7.5
100	PCR tubes	0.073	100	15 ml Falcons	0.246
			1.6 liters	acetonitrile	0.38
			8 ml	Acetic acid	0.00244
			100	SepPak cartridges	3.11
			100	SepPak liners	0.385
			1/20	Fe-column	3.295
			8 g	FeCl	0.0069
			6 g	EDTA	0.0094
			12 ml	ammonia	0.01488
			1	C18 disk	0.0965
TOTAL		**14.46 Euros**			**59.21 Euros**
Manual labor		**1X**			**10X**

4.5 HYE species mixture

The Human-Yeast-Ecoli (HYE) mixed proteome samples were prepared using commercial digests of HeLa cells (Thermo Fisher, #88329, 20 µg), Yeast cells (Thermo Fisher, #A47951, 20 µg) and E. coli cells (Waters, #1191967, 100 µg). Each were dissolved to 100 ng/µl in MS buffer (2.5% HFIP, 0.1% TFA). Next, the three species were mixed in two different ratios. A: 65% HeLa, 15%

Yeast, 20% E. coli; B: 65% HeLa, 30% Yeast, 5 % E. coli. Therefore, they contain proteins in the known ratios of 1:1 for HeLa,1:2 for yeast, and 4:1 for E. coli. Aliquots of 1 µg total protein were prepared, vacuum centrifuged to dryness and stored at -20°C.

4.6 LC-MS/MS measurement

For LC-MS/MS analysis, peptides were dissolved in ULC/MS grade water containing 0.1% trifluoracetic acid (TFA) and 2.5% 1,1,1,3,3,3-Hexafluoro-2-propanol (HFIP), followed by sonication for 5 minutes. For phospho-enriched samples, the reconstitution buffer additionally contained 50 mM citric acid. The samples were transferred to autosampler vials and placed in the autosampler module of the Ultimate 3000 (Thermo Fisher Scientific) liquid chromatography (LC) system. The LC was operated at a flow of 300 nL/min and the column was heated to 35°C. Peptides were first loaded onto a trapping column (Thermo Fisher Scientific, 160454) in the presence of 98% loading buffer A (0.1% TFA in water) and 2% loading buffer B (0.1% TFA in acetonitrile). Peptides were then transferred to the analytical column (Waters, 186008795, BEH C18 130Å 1.7 µm 75x250 mm), from which they were separated by hydrophobicity with a linear gradient of 4-30% acetonitrile. For phosphopeptide-enriched samples, a gradient of 2-28% acetonitrile was applied (Table 5).

Table 5: LC gradient (numbers in brackets are for phosphoproteome measurements)

Time (minutes)	% Buffer A (0.1% FA in H2O)	% Buffer B (0.1% FA in ACN)	
0	98	2	
3	98	2	
4	96	4 (2)	Linear gradient
106	70	30 (28)	
107	24	76	
109	24	76	
110	98	2	
120	98	2	

4.7 Mass Spectrometry measurement

Separated peptides were ionized by electrospray ionization (ESI) with an applied voltage of 2200 V. Raw data acquisition was performed with the Orbitrap Exploris 480 mass spectrometer (Thermo Fisher Scientific) which was operated in data-independent acquisition (DIA) mode. MS1 scans were acquired with an orbitrap resolution of 120K and covered the range from 350 - 1400 m/z.

Precursor ions were collected for a maximum injection time of 45 ms and the automated gain control (AGC) target was set to 3e6. MS2 scans were acquired in 47 precursor isolation windows of variable width and 1 m/z overlap that covered the range from 400 – 1000 m/z (up to 1200 m/z for phospho-enriched samples). See Table 7 and Table 8 for the detailed window distribution. For MS2 scans a normalized collision energy of 28% (26% for phospho-enriched samples) was applied and the orbitrap was operated at a resolution of 30K. Maximum injection time was 54 ms and the automated gain control (AGC) target was set to 1e6 (Table 6). The total cycle time of the method equaled 3.6 seconds.

Table 6: MS settings for full (phospho) proteome data acquisition

Setting	Value	
Ion Source Type	NSI	
Positive Ion (V)	2200	
Ion Transfer Tube Temp (°C)	275	General
Expected LC Peak Width (s)	30	
Lock Mass (m/z)	445.12003	
Orbitrap Resolution	120,000	
Scan Range (m/z)	350-1400	
AGC Target (%)	300	MS1
Max Injection Time (ms)	45	
Data Type	Centroid	
HCD Collision Energy (%)	28 (26)	
Orbitrap Resolution	30,000	
AGC Target (%)	1000	MS2
Max Injection Time (ms)	54	
Data Type	Centroid	

Table 7: DIA windows for full proteome data acquisition

Window number	Start m/z	End m/z	Width (m/z)
1	399.9	413.9	14
2	412.9	426.9	14
3	425.9	438.9	13
4	438	451	13
5	450	462	12
6	461	473	12
7	472	484	12
8	483	494	11
9	493	504	11
10	503	515	12
11	514	525	11
12	524	535	11
13	534	544	10
14	543	554	11
15	553	564	11
16	563	574	11
17	573	583	10
18	582	593	11
19	592	603	11
20	602	613	11
21	612	623	11
22	622	633	11
23	632	643	11
24	642	653	11

Window number	Start m/z	End m/z	Width (m/z)
25	652	663	11
26	662.1	674.1	12
27	673.1	684.1	11
28	683.1	695.1	12
29	694.1	706.1	12
30	705.1	717.1	12
31	716.1	728.1	12
32	727.1	741.1	14
33	740.1	753.1	13
34	752.1	766.1	14
35	765.1	778.1	13
36	777.1	792.1	15
37	791.1	806.1	15
38	805.1	820.1	15
39	819.1	835.1	16
40	834.1	851.1	17
41	850.1	868.1	18
42	867.1	886.1	19
43	885.2	906.2	21
44	905.2	926.2	21
45	925.2	949.2	24
46	948.2	973.2	25
47	972.2	1001.2	29

Table 8: DIA windows for phosphoproteome measurement

Window number	Start m/z	End m/z	Width (m/z)
1	399.9	442.9	
2	441.9	472.9	
3	471.9	495.9	
4	494.9	515.9	
5	514.9	533.9	
6	532.9	550.9	
7	549.9	565.9	
8	564.9	580.9	
9	579.9	593.9	
10	593	608	
11	607	621	
12	620	633	
13	632	645	
14	644	657	
15	656	669	
16	669	681	
17	680	692	
18	691	704	
19	703	715	
20	714	727	
21	726	738	
22	737	750	
23	749	761	
24	760	773	

Window number	Start m/z	End m/z	Width (m/z)
25	772	785	
26	784	797	
27	796	809	
28	808.1	822.1	
29	821.1	834.1	
30	833.1	848.1	
31	847.1	862.1	
32	861.1	875.1	
33	874.1	890.1	
34	889.1	904.1	
35	903.1	920.1	
36	919.1	936.1	
37	935.1	952.1	
38	951.1	968.1	
39	967.1	985.1	
40	984.1	1004.1	
41	1003.1	1025.1	
42	1024.2	1047.2	
43	1046.2	1071.2	
44	1070.2	1099.2	
45	1098.2	1130.2	
46	1129.2	1164.2	
47	1163.2	1201.2	
48			

4.8 Raw data analysis

MS raw data search was performed using Spectronaut software (Biognosys) in directDIA (library-free) mode. A human proteome database either with protein isoforms (79,052 entries, downloaeded from Uniprot 15.03.2022) or without isoforms (20,594 entries, downloaded from Uniprot 09.02.2023) was used, as indicated in the figure legends. If not indicated otherwise, the search parameters listed in Table 9 were used. Phosphoproteomic data sets were further processed using the PeptideCollapse (Bekker-Jensen et al. 2019) plug-in of Perseus (Tyanova and Cox 2018), applying a localization probability cut-off at 95%.

Table 9: Spectronaut search parameters

Parameter	Value
Version	16.3 and 17.1 for method development, 18.6 for final analyses
Analysis Type	directDIA+
Precursor Qvalue Cutoff	0.01
Precursor PEP Cutoff	0.01
Protein Qvalue Cutoff (Experiment)	0.01
Protein Qvalue Cutoff (Run)	0.01
Protein PEP Cutoff	0.05
Exclude Single Hit Proteins	False
Decoy Generation Method	Mutated
Imputation Strategy	None
Quantity MS Level	MS2
Protein LFQ Method	Automatic
Quantity Type	Area
Cross-Run Normalization	True (Automatic)
Major Group Quantity	Sum peptide quantity
Minor Group Quantity	Mean precursor quantity
Fixed Modifications	Carbamidomethyl (C)
Variable Modifications	Acetyl (Protein N-term), Oxidation (M), (Phosphorylation (S/T/Y))
Max Peptide Length	35
Min Peptide Length	7
Missed Cleavages	2
Fragment Ions m/z	200 - 3000

4.9 Data filtering and normalization

I received RNA sequencing data in the form of TPM-normalized values from the DKFZ/NCT/DKTK MASTER program. I summed up duplicated gene names, filtered out any transcripts with TPM < 1, and filtered for genes identified in at least 80% of the samples. I then performed median normalization using the limma R package.

I performed variance stability normalization (VSN) on both the global proteome and phosphoproteome data using the limma R package. Duplicated phosphosites were summed up. Subsequently, I filtered for proteins identified in at least 50% of patients and phosphosites identified in at least 20% of patients. To compare relative expression of proteins, I calculated log2 ratios compared to the median intensity of each protein over all patients. To compare relative intensities of phosphosites, I scaled the values over all patients by z-scoring.

4.10 Statistics for group comparisons

Whenever comparing two groups with each other, e.g. KRAS mutated vs. wild-type, I filtered the data for a minimum of 75% values present per condition. Statistical comparison of groups was performed using the eBayes() function of the limma R package, followed by Benjamini-Hochberg adjustment of p-values.

4.11 Transcription factor activities

Using the decoupleR package (Badia-I-Mompel et al. 2022), I imported the CollecTRI (Müller-Dott et al. 2023) database (get_collectri function). To compute relative transcription factor activity scores between two groups, I applied the run_wmean() function to the t-statistics values of the respective comparison. For per-patient or per-organoid specific TF activities, relative to all patients or organoids, I applied the run_wmean() function to the z-scored TPM values.

4.12 Pathway activities

Using the decoupleR package (Badia-I-Mompel et al. 2022), I imported the PROGENy (Schubert et al. 2018) database (get_progeny function). To compute relative pathway activity scores between two groups, I applied the run_wmean() function to the t-statistics values of the respective comparison. For per-patient or per-organoid specific pathway activities, relative to all patients or organoids, I applied the run_wmean() function to the z-scored TPM values.

4.13 Kinase activities

Using the OmnipathR package (Türei et al. 2021) I imported the kinase substrate network (KSN) with the get_signed_ptms() function. I removed kinase-substrate relationships if they were based on ProtMapper as the only source and refined the KSN to include only kinases with at least 5 target phosphosites identified in my dataset. To compute relative kinase activities between two groups, I applied the run_wmean function to the t-statistics values of the respective comparison. For per-patient or per-organoid specific kinase activities, relative to all patients or organoids, I applied the run_wmean() function to the z-scored log2 phosphosite intensities.

4.14 Correlation analyses between CNV, mRNA and protein

I received total copy number values of genomic regions for 49 of the 65 patients from the DKFZ/NCT/DKTK MASTER program. Using the GenomicRanges and Homo.sapiens R packages, I mapped gene names to the chromosome regions that showed any copy number

variation. For correlation analysis I included the patients with both CNV and RNA seq data available (n = 45) or patients with both CNV and proteome data available (n = 49), respectively. Both mRNA TPM values and protein log2 intensities were transformed into log2 ratios relative to the median expression of each gene. Spearman correlation coefficients were computed between the copy number values and the mRNA or protein expression ratios, respectively, requiring at least 10 value pairs.

4.15 Gene set enrichment analysis

For gene set enrichment analysis (GSEA) of Reactome pathways, I used the mapIds() function of the AnnotationDbi R package to map gene names to Entrez IDs. Next, I selected the log2 fold-changes from the limma statistics and ordered them decreasingly. GSEA was then performed using the gsePathway() function of the ReactomePA R package, applying it to the sorted log2 fold-change values, requiring a minimum gene set size of 120 and BH-adjustment von p-values. Results were visualized using the dotplot() function of the enrichplot R package.

4.16 Score computation

Selected Reactome pathways included: cell cycle ("R-HSA-1640170"), apoptosis ("R-HSA-109581"), suppression of apoptosis ("R-HSA-9635465"), growth hormone receptor signaling ("R-HSA-982772"), PI3K/AKT signaling in cancer ("R-HSA-2219528"), PI3K/AKT activation ("R-HSA-198203"), oncogenic MAPK signaling ("R-HSA-6802957"), signaling by receptor tyrosine kinase ("R-HSA-9006934"), signaling by WNT in cancer ("R-HSA-4791275"), p53-dependent G1 DNA damage response ("R-HSA-69563"), p53-independent DNA damage response ("R-HSA-69610"), innate immune system ("R-HSA-168249"), adaptive immune system ("R-HSA-1280218"), cytokine signaling in immune system ("R-HSA-1280215"), signaling by TGFb family members ("R-HSA-9006936"), and metabolism ("R-HSA-1430728").

For each patient, I included transcription factors with relative activity > 2, all PROGENy-derived relative pathway activity scores, proteins with relative expression ratio > 2, GSEA enrichment scores of the selected Reactome pathways, kinases with relative activity > 2 based on Omnipath KSN, and kinases with relative activity > 2 based on the extended KSN. For the PROGENy pathway scores, EGFR and VEGF were assigned to the RKT basket, JAK-STAT, NF-κB and TNFα to the immune system basket, Trail to apoptosis, and estrogen and androgen signaling to the hormone receptor signaling basket.

Each selected feature (e.g., a protein or TF) was assigned to one - or multiple - of the Reactome baskets. Within each layer and basket, the sum of activity or enrichment values was normalized by the basket size, to prevent large baskets (e.g., cell cycle) from dominating the score. These per-basket, per-layer values were then z-score transformed across patients, ensuring that no single molecular layer (e.g., transcription factors or kinases) disproportionately influenced the overall result. Finally, the composite score was calculated by summing the six z-scores - giving a half-weight to the extended kinase layer - to produce a patient-specific activity score for each interventional basket.

Table 10: List of used R packages

R package	Version
limma	3.58.1
dplyr	1.1.4
stringr	1.5.1
readxl	1.4.3
OmnipathR	3.10.1
Purrr	1.0.2
beeswarm	0.4.0
tools	4.3.1
BiocManager	1.30.25
biomaRt	2.58.2
clusterProfiler	4.10.1
CMScaller	2.0.1
decoupleR	2.8.0
DOSE	3.28.2
enrichplot	1.22.0
EPIC	1.1.7
Homo.sapiens	1.3.1
ReactomePA	1.46.0
immunedeconv	2.1.0
GenomicRanges	1.54.1
pheatmap	1.0.12
Org.Hs.eg.db	3.18.0
factoextra	1.0.7
AnnotationDbi	1.64.1
Homo.sapiens	1.3.1

4.17 Use of large language models (LLMs)

Perplexity.ai was used to help with literature review, but everything was always manually verified by myself. ChatGPT was used to proofread text and to re-phrase it for improved readability. No text was accepted without editing by myself.

Acknowledgements

My time as a PhD student was challenging, but also very enriching for my personal development and it would not have been possible without the support of many people.

First of all, I would like to thank Prof. Stefan Wiemann for taking me in as a doctoral researcher and guiding me through these four years. After my master's I thought I was done with research, but you convinced me to stay and pursue a PhD, which I am very grateful for now! Thank you for believing in me, for always being super enthusiastic, and for treating the people around you with kindness! If more group leaders were like you, I believe fewer people would leave academia.

I am also very grateful for having had Dominic Helm as my co-supervisor. Doing my master's thesis with you ignited my passion for mass spectrometry and ultimately convinced me to stay in research for a little longer. Thank you for teaching me everything I know about mass spectrometry, trusting me with maintaining the instruments, and your continued guidance and advice!

Another irreplaceable source of support was Cindy Körner. Discussing my project's challenges with you always brought out the best ideas! Thank you also for being there for us students whenever we were struggling and needed mental support to continue the PhD – or sometimes just to vent about everything.

I would also like to thank my TAC members, Prof. Stefan Fröhling and Prof. Benedikt Brors, for taking the time out of their busy days to listen to my project's progress and to offer encouraging words, new ideas, and continued support.

I highly appreciate the fruitful collaboration with the NCT MASTER program and everyone involved: Prof. Stefan Fröhling, Dr. Barbara Hutter, Annika Baude, Katrin Pfütze, and Katja Beck. Providing me with clinical samples is what made my project possible in the first place, and I would like to thank you for sharing all your data with me and for your reliable, quick responses to all my questions.

I am very grateful for the entire MGA group, especially Janina Müller, Sabine Karolus, Angelika Wörner, and Sara Burmester, for making my PhD time enjoyable!

I would like to thank Syed Azmal Ali for lending his expertise in tissue sample processing, as well as Martin Schneider and Efsthathios Vlachavas for their input to bioinformatics analysis. Furthermore, I acknowledge and appreciate Prof. Ursula Klingmüller and Prof. Jeroen Krijgsveld for providing measurement time on the MSCoreSys-funded mass spectrometer.

While working on my PhD project was fun – at least sometimes – what I took away most from my time at DKFZ were the countless opportunities for my personal development. Engaging in other projects than my PhD was very fulfilling and was ultimately what made me push through these past four years. On that note, I would like to thank Marion Gürth and Barbara Janssens from the

Career Service for their support and for letting us organize the R&D Career Day! Furthermore, I am very thankful for the experience of organizing the CONTACT job fair and would like to thank the entire BioContact team for the great work! Last but not least, working for the DKFZ sustainability group was very fulfilling and I would like to thank Hollyn Hartlep for everything she does to make DKFZ more sustainable and for letting me help out in her projects!

On a private note, I would like to thank my parents, Andrea and Steffen, who have enabled me to study and pursue a PhD by convincing me to finish my Abitur more than ten years ago. This was one of the few times that they ever interfered with my decisions and in retrospect I am glad they did. Also, I couldn't have pursued my interests and studied for the past 10 years without their support! Finally, I would like to thank Nick for his constant mental support and for teaching me to take better care of myself and to sometimes say no.

Publication bibliography

Acha-Sagredo, Amelia; Andrei, Pietro; Clayton, Kalum; Taggart, Emma; Antoniotti, Carlotta; Woodman, Chloé A. et al. (2025): A constitutive interferon-high immunophenotype defines response to immunotherapy in colorectal cancer. In *Cancer cell* 43 (2), 292-307.e7. DOI: 10.1016/j.ccell.2024.12.008.

Agilent AssayMAP Bravo Technology Enables Reproducible Automated Phosphopeptide Enrichment from Complex Mixtures Using High-Capacity Fe (III) … (2016). Available online at https://scholar.google.com/citations?user=yh_yj6iaaaaj&hl=de&oi=sra.

Ahmadian, M. R.; Zor, T.; Vogt, D.; Kabsch, W.; Selinger, Z.; Wittinghofer, A.; Scheffzek, K. (1999): Guanosine triphosphatase stimulation of oncogenic Ras mutants. In *Proceedings of the National Academy of Sciences of the United States of America* 96 (12), pp. 7065–7070. DOI: 10.1073/pnas.96.12.7065.

Alexandrov, Ludmil B.; Stratton, Michael R. (2014): Mutational signatures: the patterns of somatic mutations hidden in cancer genomes. In *Current opinion in genetics & development* 24 (100), pp. 52–60. DOI: 10.1016/j.gde.2013.11.014.

Alfaro, Clint M.; Uwakweh, Agbo-Oma; Todd, Daniel A.; Ehrmann, Brandie M.; Cech, Nadja B. (2014): Investigations of analyte-specific response saturation and dynamic range limitations in atmospheric pressure ionization mass spectrometry. In *Analytical chemistry* 86 (21), pp. 10639–10645. DOI: 10.1021/ac502984a.

American Cancer Society: Colorectal Cancer Facts & Figures 2020-2022.

American Cancer Society: Radiation Therapy for Colorectal Cancer. Available online at https://www.cancer.org/cancer/types/colon-rectal-cancer/treating/radiation-therapy.html.

Amrane, Karim; Le Noac'h, Pierre; Hemon, Patrice; Abgral, Ronan; Le Meur, Coline; Pradier, Olivier et al. (2025): MHC class II: a predictor of outcome in melanoma treated with immune checkpoint inhibitors. In *Melanoma research*. DOI: 10.1097/CMR.0000000000001022.

André, T.; Shiu, K-K; Kim, T. W.; Jensen, B. V.; Jensen, L. H.; Punt, C. J. A. et al. (2025): Pembrolizumab versus chemotherapy in microsatellite instability-high or mismatch repair-deficient metastatic colorectal cancer: 5-year follow-up from the randomized phase III KEYNOTE-177 study. In *Annals of oncology : official journal of the European Society for Medical Oncology* 36 (3), pp. 277–284. DOI: 10.1016/j.annonc.2024.11.012.

André, Thierry; Pietrantonio, Filippo; Avallone, Antonio; Gumus, Mahmut; Wyrwicz, Lucjan; Kim, Jong Gwang et al. (2023): KEYSTEP-008: phase II trial of pembrolizumab-based combination in MSI-H/dMMR metastatic colorectal cancer. In *Future oncology (London, England)* 19 (37), pp. 2445–2452. DOI: 10.2217/fon-2022-1105.

André, Thierry; Shiu, Kai-Keen; Kim, Tae Won; Jensen, Benny Vittrup; Jensen, Lars Henrik; Punt, Cornelis et al. (2020): Pembrolizumab in Microsatellite-Instability-High Advanced Colorectal Cancer. In *The New England journal of medicine* 383 (23), pp. 2207–2218. DOI: 10.1056/NEJMoa2017699.

Angell, Trevor E.; Lechner, Melissa G.; Jang, Julie K.; LoPresti, Jonathan S.; Epstein, Alan L. (2014): MHC class I loss is a frequent mechanism of immune escape in papillary thyroid cancer that is reversed by interferon and selumetinib treatment in vitro. In *Clinical cancer research : an official journal of the American Association for Cancer Research* 20 (23), pp. 6034–6044. DOI: 10.1158/1078-0432.CCR-14-0879.

Angelova, Mihaela; Charoentong, Pornpimol; Hackl, Hubert; Fischer, Maria L.; Snajder, Rene; Krogsdam, Anne M. et al. (2015): Characterization of the immunophenotypes and antigenomes of colorectal cancers reveals distinct tumor escape mechanisms and novel targets for immunotherapy. In *Genome biology* 16 (1), p. 64. DOI: 10.1186/s13059-015-0620-6.

Angelova, Mihaela; Mlecnik, Bernhard; Vasaturo, Angela; Bindea, Gabriela; Fredriksen, Tessa; Lafontaine, Lucie et al. (2018): Evolution of Metastases in Space and Time under Immune Selection. In *Cell* 175 (3), 751-765.e16. DOI: 10.1016/j.cell.2018.09.018.

Ardito, Fatima; Giuliani, Michele; Perrone, Donatella; Troiano, Giuseppe; Lo Muzio, Lorenzo (2017): The crucial role of protein phosphorylation in cell signaling and its use as targeted therapy (Review). In *International journal of molecular medicine* 40 (2), pp. 271–280. DOI: 10.3892/ijmm.2017.3036.

Arnold, Melina; Sierra, Mónica S.; Laversanne, Mathieu; Soerjomataram, Isabelle; Jemal, Ahmedin; Bray, Freddie (2017): Global patterns and trends in colorectal cancer incidence and mortality. In *Gut* 66 (4), pp. 683–691. DOI: 10.1136/gutjnl-2015-310912.

Asanuma, Kunihiro; Nakamura, Tomoki; Nakamura, Koichi; Hagi, Tomohito; Okamoto, Takayuki; Kita, Kouji et al. (2022): Compound Library Screening for Synergistic Drug Combinations: mTOR Inhibitor and Proteasome Inhibitor Effective Against Osteosarcoma Cells. In *Anticancer research* 42 (9), pp. 4319–4328. DOI: 10.21873/anticanres.15932.

Avraham, Roi; Yarden, Yosef (2011): Feedback regulation of EGFR signalling: decision making by early and delayed loops. In *Nat Rev Mol Cell Biol* 12 (2), pp. 104–117. DOI: 10.1038/nrm3048.

Axelrod, Margaret L.; Cook, Rebecca S.; Johnson, Douglas B.; Balko, Justin M. (2019): Biological Consequences of MHC-II Expression by Tumor Cells in Cancer. In *Clinical cancer research : an official journal of the American Association for Cancer Research* 25 (8), pp. 2392–2402. DOI: 10.1158/1078-0432.CCR-18-3200.

Bache, Nicolai; Geyer, Philipp E.; Bekker-Jensen, Dorte B.; Hoerning, Ole; Falkenby, Lasse; Treit, Peter V. et al. (2018): A Novel LC System Embeds Analytes in Pre-formed Gradients for Rapid, Ultra-robust Proteomics. In *Molecular & cellular proteomics : MCP* 17 (11), pp. 2284–2296. DOI: 10.1074/mcp.TIR118.000853.

Back, Tim R. de; Wu, Tan; Schafrat, Pascale Jm; Hoorn, Sanne ten; Tan, Miaomiao; He, Lingli et al. (2024): A consensus molecular subtypes classification strategy for clinical colorectal cancer tissues. In *Life science alliance* 7 (8). DOI: 10.26508/lsa.202402730.

Badia-I-Mompel, Pau; Vélez Santiago, Jesús; Braunger, Jana; Geiss, Celina; Dimitrov, Daniel; Müller-Dott, Sophia et al. (2022): decoupleR: ensemble of computational methods to infer biological activities from omics data. In *Bioinformatics advances* 2 (1), vbac016. DOI: 10.1093/bioadv/vbac016.

Baker, Christa P.; Bruderer, Roland; Abbott, James; Arthur, J. Simon C.; Brenes, Alejandro J. (2024): Optimizing Spectronaut Search Parameters to Improve Data Quality with Minimal Proteome Coverage Reductions in DIA Analyses of Heterogeneous Samples. In *Journal of proteome research* 23 (6), pp. 1926–1936. DOI: 10.1021/acs.jproteome.3c00671.

Bassères, Daniela S.; Ebbs, Aaron; Levantini, Elena; Baldwin, Albert S. (2010): Requirement of the NF-kappaB subunit p65/RelA for K-Ras-induced lung tumorigenesis. In *Cancer research* 70 (9), pp. 3537–3546. DOI: 10.1158/0008-5472.CAN-09-4290.

Beekhof, Robin; Bertotti, Andrea; Böttger, Franziska; Vurchio, Valentina; Cottino, Francesca; Zanella, Eugenia R. et al. (2023): Phosphoproteomics of patient-derived xenografts identifies targets and markers associated with sensitivity and resistance to EGFR blockade in colorectal cancer. In *Science translational medicine* 15 (709), eabm3687. DOI: 10.1126/scitranslmed.abm3687.

Bekker-Jensen, Dorte B.; Bernhardt, Oliver M.; Hogrebe, Alexander; del Val, Ana Martinez; Verbeke, Lynn; Gandhi, Tejas et al. (2019): Rapid and site-specific deep phosphoproteome profiling by data-independent acquisition (DIA) without the need for spectral libraries.

Bekker-Jensen, Dorte B.; Martínez-Val, Ana; Steigerwald, Sophia; Rüther, Patrick; Fort, Kyle L.; Arrey, Tabiwang N. et al. (2020): A Compact Quadrupole-Orbitrap Mass Spectrometer with FAIMS Interface Improves Proteome Coverage in Short LC Gradients. In *Molecular & cellular proteomics : MCP* 19 (4), pp. 716–729. DOI: 10.1074/mcp.TIR119.001906.

Benelli, Nicolas D.; Brandon, Ian; Hew, Karina E. (2024): Immune Checkpoint Inhibitors: A Narrative Review on PD-1/PD-L1 Blockade Mechanism, Efficacy, and Safety Profile in Treating Malignancy. In *Cureus* 16 (4), e58138. DOI: 10.7759/cureus.58138.

Benvenuti, Silvia; Sartore-Bianchi, Andrea; Di Nicolantonio, Federica; Zanon, Carlo; Moroni, Mauro; Veronese, Silvio et al. (2007): Oncogenic activation of the RAS/RAF signaling pathway impairs the response of metastatic colorectal cancers to anti-epidermal growth factor receptor antibody therapies. In *Cancer research* 67 (6), pp. 2643–2648. DOI: 10.1158/0008-5472.CAN-06-4158.

Bhat, Ratan V.; Andersson, Ulf; Andersson, Shalini; Knerr, Laurent; Bauer, Udo; Sundgren-Andersson, Anna K. (2018): The Conundrum of GSK3 Inhibitors: Is it the Dawn of a New Beginning? In *Journal of Alzheimer's disease : JAD* 64 (s1), S547-S554. DOI: 10.3233/JAD-179934.

Bian, Yangyang; Zheng, Runsheng; Bayer, Florian P.; Wong, Cassandra; Chang, Yun-Chien; Meng, Chen et al. (2020): Robust, reproducible and quantitative analysis of thousands of proteomes by micro-flow LC-MS/MS. In *Nature communications* 11 (1), p. 157. DOI: 10.1038/s41467-019-13973-x.

Blees, Andreas; Januliene, Dovile; Hofmann, Tommy; Koller, Nicole; Schmidt, Carla; Trowitzsch, Simon et al. (2017): Structure of the human MHC-I peptide-loading complex. In *Nature* 551 (7681), pp. 525–528. DOI: 10.1038/nature24627.

Boys, Emma L.; Liu, Jia; Robinson, Phillip J.; Reddel, Roger R. (2023): Clinical applications of mass spectrometry-based proteomics in cancer: Where are we? In *Proteomics* 23 (7-8), e2200238. DOI: 10.1002/pmic.202200238.

Bozym, David J.; Zheng, David X.; Revach, Or-Yam; Aref, Amir; Jenkins, Russell W. (2025): Patient-derived organotypic tumor spheroids, tumoroids, and organoids: advancing immunotherapy using state-of-the-art 3D tumor model systems. In *Lab on a chip*. DOI: 10.1039/D5LC00062A.

Bray, Freddie; Laversanne, Mathieu; Sung, Hyuna; Ferlay, Jacques; Siegel, Rebecca L.; Soerjomataram, Isabelle; Jemal, Ahmedin (2024): Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. In *CA: a cancer journal for clinicians* 74 (3), pp. 229–263. DOI: 10.3322/caac.21834.

Brierley, James D.; van Eycken, Elizabeth; Rous, Brian; Giuliani, Meredith (2025): TNM Classification of Malignant Tumours: John Wiley & Sons.

Brown, Kyle A.; Melby, Jake A.; Roberts, David S.; Ge, Ying (2020): Top-down proteomics: challenges, innovations, and applications in basic and clinical research. In *Expert review of proteomics* 17 (10), pp. 719–733. DOI: 10.1080/14789450.2020.1855982.

Bruderer, Roland; Bernhardt, Oliver M.; Gandhi, Tejas; Miladinović, Saša M.; Cheng, Lin-Yang; Messner, Simon et al. (2015): Extending the limits of quantitative proteome profiling with data-independent acquisition and application to acetaminophen-treated three-dimensional liver microtissues. In *Molecular & cellular proteomics : MCP* 14 (5), pp. 1400–1410. DOI: 10.1074/mcp.M114.044305.

Bruins, Andries P. (1998): Mechanistic aspects of electrospray ionization. In *Journal of Chromatography A* 794 (1-2), pp. 345–357. DOI: 10.1016/S0021-9673(97)01110-2.

Cairns, J. (1975): Mutation selection and the natural history of cancer. In *Nature* 255 (5505), pp. 197–200. DOI: 10.1038/255197a0.

Calon, Alexandre; Espinet, Elisa; Palomo-Ponce, Sergio; Tauriello, Daniele V. F.; Iglesias, Mar; Céspedes, María Virtudes et al. (2012): Dependency of colorectal cancer on a TGF-β-driven program in stromal cells for metastasis initiation. In *Cancer cell* 22 (5), pp. 571–584. DOI: 10.1016/j.ccr.2012.08.013.

Cammareri, Patrizia; Raponi, Michela; Hong, Yourae; Billard, Caroline V.; Peckett, Nat; Zhu, Yujia et al. (2025): Loss of colonic fidelity enables multilineage plasticity and metastasis. In *Nature*. DOI: 10.1038/s41586-025-09125-5.

Canagarajah, B. J.; Khokhlatchev, A.; Cobb, M. H.; Goldsmith, E. J. (1997): Activation mechanism of the MAP kinase ERK2 by dual phosphorylation. In *Cell* 90 (5), pp. 859–869. DOI: 10.1016/s0092-8674(00)80351-7.

Cañellas-Socias, Adrià; Sancho, Elena; Batlle, Eduard (2024): Mechanisms of metastatic colorectal cancer. In *Nature reviews. Gastroenterology & hepatology* 21 (9), pp. 609–625. DOI: 10.1038/s41575-024-00934-z.

Cercek, Andrea; Lumish, Melissa; Sinopoli, Jenna; Weiss, Jill; Shia, Jinru; Lamendola-Essel, Michelle et al. (2022): PD-1 Blockade in Mismatch Repair-Deficient, Locally Advanced Rectal Cancer. In *The New England journal of medicine* 386 (25), pp. 2363–2376. DOI: 10.1056/NEJMoa2201445.

Chalabi, M.; Verschoor, Y. L.; van den Berg, J.; Sikorska, K.; Beets, G.; Lent, A. V. et al. (2022): LBA7 Neoadjuvant immune checkpoint inhibition in locally advanced MMR-deficient colon cancer: The NICHE-2 study. In *Annals of Oncology* 33, S1389. DOI: 10.1016/j.annonc.2022.08.016.

Chalabi, Myriam; Fanchi, Lorenzo F.; Dijkstra, Krijn K.; van den Berg, José G.; Aalbers, Arend G.; Sikorska, Karolina et al. (2020): Neoadjuvant immunotherapy leads to pathological responses in MMR-proficient and MMR-deficient early-stage colon cancers. In *Nature medicine* 26 (4), pp. 566–576. DOI: 10.1038/s41591-020-0805-8.

Chan, Wai Sun; Mo, Xuetang; Ip, Philip Pun Ching; Tse, Ka Yu (2023): Patient-derived organoid culture in epithelial ovarian cancers-Techniques, applications, and future perspectives. In *Cancer medicine* 12 (19), pp. 19714–19731. DOI: 10.1002/cam4.6521.

Chang, C. H.; Fontes, J. D.; Peterlin, M.; Flavell, R. A. (1994): Class II transactivator (CIITA) is sufficient for the inducible expression of major histocompatibility complex class II genes. In *J Exp Med* 180 (4), pp. 1367–1374. DOI: 10.1084/jem.180.4.1367.

Chang, David Z.; Kumar, Vikas; Ma, Ying; Li, Kuiyuan; Kopetz, Scott (2009): Individualized therapies in colorectal cancer: KRAS as a marker for response to EGFR-targeted therapy. In *Journal of hematology & oncology* 2, p. 18. DOI: 10.1186/1756-8722-2-18.

Che, Li-Heng; Liu, Jing-Wen; Huo, Jian-Ping; Luo, Rong; Xu, Rui-Ming; He, Cai et al. (2021): A single-cell atlas of liver metastases of colorectal cancer reveals reprogramming of the tumor microenvironment in response to preoperative chemotherapy. In *Cell discovery* 7 (1), p. 80. DOI: 10.1038/s41421-021-00312-y.

Chida, Keigo; Kotani, Daisuke; Masuishi, Toshiki; Kawakami, Takeshi; Kawamoto, Yasuyuki; Kato, Kyoko et al. (2021): The Prognostic Impact of KRAS G12C Mutation in Patients with Metastatic Colorectal Cancer: A Multicenter Retrospective Observational Study. In *The Oncologist* 26 (10), pp. 845–853. DOI: 10.1002/onco.13870.

Choi, Jeesoo; Shin, Ju-Young; Kim, Taeyul K.; Kim, Kiwook; Kim, Jiyun; Jeon, Eunhye et al. (2024): Site-specific mutagenesis screening in KRASG12D mutant library to uncover resistance mechanisms to KRASG12D inhibitors. In *Cancer letters* 591, p. 216904. DOI: 10.1016/j.canlet.2024.216904.

Codrich, Marta; Dalla, Emiliano; Mio, Catia; Antoniali, Giulia; Malfatti, Matilde Clarissa; Marzinotto, Stefania et al. (2021): Integrated multi-omics analyses on patient-derived CRC organoids highlight altered molecular pathways in colorectal cancer progression involving PTEN. In *Journal of experimental & clinical cancer research : CR* 40 (1), p. 198. DOI: 10.1186/s13046-021-01986-8.

Cox, Jürgen; Mann, Matthias (2008): MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. In *Nature biotechnology* 26 (12), pp. 1367–1372. DOI: 10.1038/nbt.1511.

Cristescu, Razvan; Mogg, Robin; Ayers, Mark; Albright, Andrew; Murphy, Erin; Yearley, Jennifer et al. (2018): Pan-tumor genomic biomarkers for PD-1 checkpoint blockade-based immunotherapy. In *Science (New York, N.Y.)* 362 (6411). DOI: 10.1126/science.aar3593.

D'Adamo, Gemma L.; Widdop, James T.; Giles, Edward M. (2021): The future is now? Clinical and translational aspects of "Omics" technologies. In *Immunology and cell biology* 99 (2), pp. 168–176. DOI: 10.1111/imcb.12404.

Dai, Shangnan; Peng, Yunpeng; Wang, Guangfu; Chen, Chongfa; Chen, Qiuyang; Yin, Lingdi et al. (2025): LIM domain only 7: a novel driver of immune evasion through regulatory T cell differentiation and chemotaxis in pancreatic ductal adenocarcinoma. In *Cell death and differentiation* 32 (2), pp. 271–290. DOI: 10.1038/s41418-024-01358-7.

Decaux, Olivier; Clément, Monique; Magrangeas, Florence; Gouraud, Wilfried; Charbonnel, Catherine; Campion, Loïc et al. (2010): Inhibition of mTORC1 activity by REDD1 induction in myeloma cells resistant to bortezomib cytotoxicity. In *Cancer science* 101 (4), pp. 889–897. DOI: 10.1111/j.1349-7006.2009.01467.x.

Deng, Yanhong; Wang, Li; Tan, Shuyun; Kim, George P.; Dou, Ruoxu; Chen, Dianke et al. (2015): KRAS as a predictor of poor prognosis and benefit from postoperative FOLFOX chemotherapy in patients with stage II and III colorectal cancer. In *Molecular oncology* 9 (7), pp. 1341–1347. DOI: 10.1016/j.molonc.2015.03.006.

Di Nicolantonio, Federica; Vitiello, Pietro Paolo; Marsoni, Silvia; Siena, Salvatore; Tabernero, Josep; Trusolino, Livio et al. (2021): Precision oncology in metastatic colorectal cancer - from biology to medicine. In *Nature reviews. Clinical oncology* 18 (8), pp. 506–525. DOI: 10.1038/s41571-021-00495-z.

Dienstmann, Rodrigo; Jang, In Sock; Bot, Brian; Friend, Stephen; Guinney, Justin (2015): Database of genomic biomarkers for cancer drugs and clinical targetability in solid tumors. In *Cancer discovery* 5 (2), pp. 118–123. DOI: 10.1158/2159-8290.CD-14-1118.

Doroshow, Deborah Blythe; Bhalla, Sheena; Beasley, Mary Beth; Sholl, Lynette M.; Kerr, Keith M.; Gnjatic, Sacha et al. (2021): PD-L1 as a biomarker of response to immune-checkpoint inhibitors. In *Nature reviews. Clinical oncology* 18 (6), pp. 345–362. DOI: 10.1038/s41571-021-00473-5.

Duraiyan, Jeyapradha; Govindarajan, Rajeshwar; Kaliyappan, Karunakaran; Palanisamy, Murugesan (2012): Applications of immunohistochemistry. In *Journal of pharmacy & bioallied sciences* 4 (Suppl 2), S307-9. DOI: 10.4103/0975-7406.100281.

Durgeau, Aurélie; Virk, Yasemin; Corgnac, Stéphanie; Mami-Chouaib, Fathia (2018): Recent Advances in Targeting CD8 T-Cell Immunity for More Effective Cancer Immunotherapy. In *Frontiers in immunology* 9, p. 14. DOI: 10.3389/fimmu.2018.00014.

Edin, Sofia; Gylling, Björn; Li, Xingru; Stenberg, Åsa; Löfgren-Burström, Anna; Zingmark, Carl et al. (2024): Opposing roles by KRAS and BRAF mutation on immune cell infiltration in colorectal cancer - possible implications for immunotherapy. In *British journal of cancer* 130 (1), pp. 143–150. DOI: 10.1038/s41416-023-02483-9.

Eide, Peter W.; Bruun, Jarle; Lothe, Ragnhild A.; Sveen, Anita (2017): CMScaller: an R package for consensus molecular subtyping of colorectal cancer pre-clinical models. In *Scientific reports* 7 (1), p. 16618. DOI: 10.1038/s41598-017-16747-x.

Erber, Johanna; Steiner, Joachim D.; Isensee, Jörg; Lobbes, Leonard A.; Toschka, André; Beleggia, Filippo et al. (2019): Dual Inhibition of GLUT1 and the ATR/CHK1 Kinase Axis Displays Synergistic Cytotoxicity in KRAS-Mutant Cancer Cells. In *Cancer research* 79 (19), pp. 4855–4868. DOI: 10.1158/0008-5472.CAN-18-3959.

Faivre, Danielle A.; McGann, Christopher D.; Merrihew, Gennifer E.; Schweppe, Devin K.; MacCoss, Michael J. (2023): Comparing peptide identifications by FAIMS versus quadrupole gas-phase fractionation.

Fakih, Marwan; Wang, Chongkai; Sandhu, Jaideep; Ye, Jian; Egelston, Colt; Li, Xiaochen (2024): Immunotherapy response in microsatellite stable metastatic colorectal cancer is influenced by site of metastases. In *European journal of cancer (Oxford, England : 1990)* 196, p. 113437. DOI: 10.1016/j.ejca.2023.113437.

Fakih, Marwan G.; Salvatore, Lisa; Esaki, Taito; Modest, Dominik P.; Lopez-Bravo, David P.; Taieb, Julien et al. (2023): Sotorasib plus Panitumumab in Refractory Colorectal Cancer with Mutated KRAS G12C. In *The New England journal of medicine* 389 (23), pp. 2125–2139. DOI: 10.1056/NEJMoa2308795.

FDA (2020): FDA approves pembrolizumab for adults and children with TMB-H solid tumors.

Feng, Juanjuan; Hu, Zhongwei; Xia, Xinting; Liu, Xiaogu; Lian, Zhengke; Wang, Hui et al. (2023): Feedback activation of EGFR/wild-type RAS signaling axis limits KRASG12D inhibitor efficacy in KRASG12D-mutated colorectal cancer. In *Oncogene* 42 (20), pp. 1620–1633. DOI: 10.1038/s41388-023-02676-9.

Foersch, Sebastian; Schmitt, Maxime; Litmeyer, Anne-Sophie; Tschurtschenthaler, Markus; Gress, Thomas; Bartsch, Detlef K. et al. (2024): TROP2 in colorectal carcinoma: associations with histopathology, molecular phenotype, and patient prognosis. In *The journal of pathology. Clinical research* 10 (5), e12394. DOI: 10.1002/2056-4538.12394.

Fortelny, Nikolaus; Overall, Christopher M.; Pavlidis, Paul; Freue, Gabriela V. Cohen (2017): Can we predict protein from mRNA levels? In *Nature* 547 (7664), E19-E20. DOI: 10.1038/nature22293.

Freeman, Daniel J.; Juan, Todd; Reiner, Maureen; Hecht, J. Randolph; Meropol, Neal J.; Berlin, Jordan et al. (2008): Association of K-ras mutational status and clinical outcomes in patients with metastatic colorectal cancer receiving panitumumab alone. In *Clinical colorectal cancer* 7 (3), pp. 184–190. DOI: 10.3816/CCC.2008.n.024.

Frejno, Martin; Meng, Chen; Ruprecht, Benjamin; Oellerich, Thomas; Scheich, Sebastian; Kleigrewe, Karin et al. (2020): Proteome activity landscapes of tumor cell lines determine drug responses. In *Nature communications* 11 (1), p. 3639. DOI: 10.1038/s41467-020-17336-9.

Fröhlich, Klemens; Fahrner, Matthias; Brombacher, Eva; Seredynska, Adrianna; Maldacker, Maximilian; Kreutz, Clemens et al. (2024): Data-Independent Acquisition: A Milestone and Prospect in Clinical Mass Spectrometry-Based Proteomics. In *Molecular & cellular proteomics : MCP* 23 (8), p. 100800. DOI: 10.1016/j.mcpro.2024.100800.

Fujii, Masayuki; Shimokawa, Mariko; Date, Shoichi; Takano, Ai; Matano, Mami; Nanki, Kosaku et al. (2016): A Colorectal Tumor Organoid Library Demonstrates Progressive Loss of Niche Factor Requirements during Tumorigenesis. In *Cell stem cell* 18 (6), pp. 827–838. DOI: 10.1016/j.stem.2016.04.003.

Gámez-Pozo, Angelo; Trilla-Fuertes, Lucía; Berges-Soria, Julia; Selevsek, Nathalie; López-Vacas, Rocío; Díaz-Almirón, Mariana et al. (2017): Functional proteomics outlines the complexity of breast cancer molecular subtypes. In *Scientific reports* 7 (1), p. 10100. DOI: 10.1038/s41598-017-10493-w.

Gandara, David R.; Agarwal, Neeraj; Gupta, Shilpa; Klempner, Samuel J.; Andrews, Miles C.; Mahipal, Amit et al. (2025): Tumor mutational burden and survival on immune checkpoint inhibition in 8000 patients across 24 cancer types. In *Journal for Immunotherapy of Cancer* 13 (2). DOI: 10.1136/jitc-2024-010311.

Garcia-Alonso, Luz; Holland, Christian H.; Ibrahim, Mahmoud M.; Turei, Denes; Saez-Rodriguez, Julio (2019): Benchmark and integration of resources for the estimation of human transcription factor activities. In *Genome research* 29 (8), pp. 1363–1375. DOI: 10.1101/gr.240663.118.

Garrido-Rodriguez, Martin; Potel, Clement; Burtscher, Mira Lea; Becher, Isabelle; Rodriguez-Mier, Pablo; Müller-Dott, Sophia et al. (2024): Evaluating signaling pathway inference from kinase-substrate interactions and phosphoproteomics data.

Gettinger, Scott; Choi, Jungmin; Hastings, Katherine; Truini, Anna; Datar, Ila; Sowell, Ryan et al. (2017): Impaired HLA Class I Antigen Processing and Presentation as a Mechanism of Acquired Resistance to Immune Checkpoint Inhibitors in Lung Cancer. In *Cancer discovery* 7 (12), pp. 1420–1435. DOI: 10.1158/2159-8290.CD-17-0593.

Gitton, Yorick; Dahmane, Nadia; Baik, Sonya; Ruiz i Altaba, Ariel; Neidhardt, Lorenz; Scholze, Manuela et al. (2002): A gene expression map of human chromosome 21 orthologues in the mouse. In *Nature* 420 (6915), pp. 586–590. DOI: 10.1038/nature01270.

Gomez-Zepeda, David; Michna, Thomas; Ziesmann, Tanja; Distler, Ute; Tenzer, Stefan (2024): HowDirty: An R package to evaluate molecular contaminants in LC-MS experiments. In *Proteomics* 24 (8), e2300134. DOI: 10.1002/pmic.202300134.

Gorden, P.; Carpentier, J. L.; Cohen, S.; Orci, L. (1978): Epidermal growth factor: morphological demonstration of binding, internalization, and lysosomal association in human fibroblasts. In *Proceedings of the National Academy of Sciences of the United States of America* 75 (10), pp. 5025–5029. DOI: 10.1073/pnas.75.10.5025.

Gotoh, Noriko (2009): Feedback inhibitors of the epidermal growth factor receptor signaling pathways. In *The international journal of biochemistry & cell biology* 41 (3), pp. 511–515. DOI: 10.1016/j.biocel.2008.06.019.

Guingab-Cagmat, J. D.; Cagmat, E. B.; Hayes, R. L.; Anagli, J. (2013): Integration of proteomics, bioinformatics, and systems biology in traumatic brain injury biomarker discovery. In *Frontiers in neurology* 4, p. 61. DOI: 10.3389/fneur.2013.00061.

Guinney, Justin; Dienstmann, Rodrigo; Wang, Xin; Reyniès, Aurélien de; Schlicker, Andreas; Soneson, Charlotte et al. (2015): The consensus molecular subtypes of colorectal cancer. In *Nature medicine* 21 (11), pp. 1350–1356. DOI: 10.1038/nm.3967.

Guven, Deniz Can; Kavgaci, Gozde; Erul, Enes; Syed, Masood Pasha; Magge, Tara; Saeed, Anwaar et al. (2024): The Efficacy of Immune Checkpoint Inhibitors in Microsatellite Stable Colorectal Cancer: A Systematic Review. In *The Oncologist* 29 (5), e580-e600. DOI: 10.1093/oncolo/oyae013.

ha, Taehyeong; Hwang, Kihwan; Nam, Kyung Mi; Ji, So Young; Kang, Ho; Han, Jung Ho; Kim, Chae-Yong (2024): TMOD-34. GENERATION OF BRAIN-VESSEL-GBM ASSEMBLOIDS USING PATIENT-DERIVED GLIOBLASTOMA TISSUE. In *Neuro-Oncology* 26 (Supplement_8), viii327-viii327. DOI: 10.1093/neuonc/noae165.1298.

Haabeth, Ole Audun Werner; Tveita, Anders Aune; Fauskanger, Marte; Schjesvold, Fredrik; Lorvik, Kristina Berg; Hofgaard, Peter O. et al. (2014): How Do CD4(+) T Cells Detect and Eliminate Tumor Cells That Either Lack or Express MHC Class II Molecules? In *Frontiers in immunology* 5, p. 174. DOI: 10.3389/fimmu.2014.00174.

Habič, A.; Majc, B.; Porčnik, A.; Bošnjak, R.; Mlakar, J.; Lah Turnšek, T. et al. (2022): P02.07.B Patient-derived glioblastoma organoids: Elucidating the mechanisms of glioblastoma therapeutic resistance in the context of tumor microenvironment. In *Neuro-Oncology* 24 (Supplement_2), ii30-ii31. DOI: 10.1093/neuonc/noac174.100.

Haines, Moe; Thorup, John R.; Gohsman, Simone; Ctortecka, Claudia; Newton, Chelsea; Rohrer, Dan C. et al. (2024): High-throughput proteomic and phosphoproteomic analysis of formalin-fixed paraffin-embedded tissue. In *bioRxiv : the preprint server for biology*. DOI: 10.1101/2024.11.17.624038.

Hanahan, Douglas (2022): Hallmarks of Cancer: New Dimensions. In *Cancer discovery* 12 (1), pp. 31–46. DOI: 10.1158/2159-8290.CD-21-1059.

Hanahan, Douglas; Weinberg, Robert A. (2011): Hallmarks of cancer: the next generation. In *Cell* 144 (5), pp. 646–674. DOI: 10.1016/j.cell.2011.02.013.

Harryvan, T. J.; Abudukelimu, S.; Stouten, I.; van der Wel, E. J.; Janson, S. G.T.; Lenos, K. J. et al. (2024): TGFβ signaling in cancer-associated fibroblasts drives a hepatic gp130-dependent pro-metastatic inflammatory program in CMS4 colorectal cancer subtype.

Harsha, H. C.; Pandey, Akhilesh (2010): Phosphoproteomics in cancer. In *Molecular oncology* 4 (6), pp. 482–495. DOI: 10.1016/j.molonc.2010.09.004.

He, Tianlin (2019): Implementation of Proteomics in Clinical Trials. In *Proteomics. Clinical applications* 13 (2), e1800198. DOI: 10.1002/prca.201800198.

He, Xingfeng; Jiang, Yan; Zhang, Long; Li, Yaqi; Hu, Xiang; Hua, Guoqiang et al. (2023): Patient-derived organoids as a platform for drug screening in metastatic colorectal cancer. In *Frontiers in bioengineering and biotechnology* 11, p. 1190637. DOI: 10.3389/fbioe.2023.1190637.

Heald, R. J.; Ryall, R. D. (1986): Recurrence and survival after total mesorectal excision for rectal cancer. In *Lancet (London, England)* 1 (8496), pp. 1479–1482. DOI: 10.1016/s0140-6736(86)91510-2.

Heil, Lilian R.; Damoc, Eugen; Arrey, Tabiwang N.; Pashkova, Anna; Denisov, Eduard; Petzoldt, Johannes et al. (2023): Evaluating the Performance of the Astral Mass Analyzer for Quantitative Proteomics Using Data-Independent Acquisition. In *Journal of proteome research* 22 (10), pp. 3290–3300. DOI: 10.1021/acs.jproteome.3c00357.

Heinrich, Kathrin; Miller-Phillips, Lisa; Ziemann, Frank; Hasselmann, Korbinian; Rühlmann, Katharina; Flach, Madeleine et al. (2023): Lessons learned: the first consecutive 1000 patients of the CCCMunichLMU Molecular Tumor Board. In *Journal of cancer research and clinical oncology* 149 (5), pp. 1905–1915. DOI: 10.1007/s00432-022-04165-0.

Henry, Jason T.; Coker, Oluwadara; Chowdhury, Saikat; Shen, John Paul; van Morris, K.; Dasari, Arvind et al. (2021): Comprehensive Clinical and Molecular Characterization of KRASG12C-Mutant Colorectal Cancer. In *JCO precision oncology* 5. DOI: 10.1200/PO.20.00256.

Herbst, Roy S.; Soria, Jean-Charles; Kowanetz, Marcin; Fine, Gregg D.; Hamid, Omid; Gordon, Michael S. et al. (2014): Predictive correlates of response to the anti-PD-L1 antibody MPDL3280A in cancer patients. In *Nature* 515 (7528), pp. 563–567. DOI: 10.1038/nature14011.

Herzberg, Benjamin O.; Manji, Gulam A. (2023): KRAS: Druggable at Last. In *The Oncologist* 28 (4), pp. 283–286. DOI: 10.1093/oncolo/oyad014.

Hinck, Lindsay; Näthke, Inke (2014): Changes in cell and tissue organization in cancer of the breast and colon. In *Current opinion in cell biology* 26, pp. 87–95. DOI: 10.1016/j.ceb.2013.11.003.

Horak, Peter; Heining, Christoph; Kreutzfeldt, Simon; Hutter, Barbara; Mock, Andreas; Hüllein, Jennifer et al. (2021): Comprehensive Genomic and Transcriptomic Analysis for Guiding Therapeutic Decisions in Patients with Rare Cancers. In *Cancer discovery* 11 (11), pp. 2780–2795. DOI: 10.1158/2159-8290.CD-21-0126.

Horak, Peter; Klink, Barbara; Heining, Christoph; Gröschel, Stefan; Hutter, Barbara; Fröhlich, Martina et al. (2017): Precision oncology based on omics data: The NCT Heidelberg experience. In *International journal of cancer* 141 (5), pp. 877–886. DOI: 10.1002/ijc.30828.

Hornbeck, Peter V.; Zhang, Bin; Murray, Beth; Kornhauser, Jon M.; Latham, Vaughan; Skrzypek, Elzbieta (2015): PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. In *Nucleic acids research* 43 (Database issue), D512-20. DOI: 10.1093/nar/gku1267.

Hristova, Ventzislava A.; Chan, Daniel W. (2019): Cancer biomarker discovery and translation: proteomics and beyond. In *Expert review of proteomics* 16 (2), pp. 93–103. DOI: 10.1080/14789450.2019.1559062.

Hu, Sha-Sha; Han, Yue; Tan, Tian-Yuan; Chen, Hui; Gao, Jia-Wen; Wang, Lan et al. (2023): SLC25A21 downregulation promotes KRAS-mutant colorectal cancer progression by increasing glutamine anaplerosis. In *JCI insight* 8 (21). DOI: 10.1172/jci.insight.167874.

Hughes, Christopher S.; Moggridge, Sophie; Müller, Torsten; Sorensen, Poul H.; Morin, Gregg B.; Krijgsveld, Jeroen (2019): Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. In *Nature protocols* 14 (1), pp. 68–85. DOI: 10.1038/s41596-018-0082-x.

Hunter, John C.; Manandhar, Anuj; Carrasco, Martin A.; Gurbani, Deepak; Gondi, Sudershan; Westover, Kenneth D. (2015): Biochemical and Structural Analysis of Common Cancer-Associated KRAS Mutations. In *Molecular cancer research : MCR* 13 (9), pp. 1325–1335. DOI: 10.1158/1541-7786.MCR-15-0203.

Hurkmans, Daan P.; Verdegaal, Els M. E.; Hogan, Sabrina A.; Wijn, Rik de; Hovestad, Lies; van den Heuvel, Dianne M. A. et al. (2020): Blood-based kinase activity profiling: a potential predictor of response to immune checkpoint inhibition in metastatic cancer. In *Journal for Immunotherapy of Cancer* 8 (2). DOI: 10.1136/jitc-2020-001607.

Hurwitz, Herbert; Fehrenbacher, Louis; Novotny, William; Cartwright, Thomas; Hainsworth, John; Heim, William et al. (2004): Bevacizumab plus irinotecan, fluorouracil, and leucovorin for metastatic colorectal cancer. In *The New England journal of medicine* 350 (23), pp. 2335–2342. DOI: 10.1056/NEJMoa032691.

Ilerhunmwuwa, Nosakhare Paul; Sahin, Ibrahim Halil; Saeed, Anwaar (2025): Immunotherapy resistance in colorectal cancer with liver metastases: challenges & therapeutic advances. In *Chinese clinical oncology* 14 (1), p. 7. DOI: 10.21037/cco-24-93.

Ilhan, Nurullah; Dane, Faysal; Goker, Erdem; Uygun, Kazım; Orhan, Bülent; Okutur, Kerem et al. (2025): Regional and Gender-Based Distribution of KRAS Mutations in Metastatic Colorectal Cancer Patients in Turkey: An Observational Study. In *Medicina (Kaunas, Lithuania)* 61 (4). DOI: 10.3390/medicina61040694.

International Agency for Research on Cancer (2010): WHO classification of tumours of the digestive system. Reflects the views of a working group that convened for an editorial and consensus conference at the International Agency for Research on Cancer (IARC), Lyon, December 10 - 12, 2009 ; third volume of the 4th edition of the WHO series on histological and genetic typing of human tumours. 4. ed., 1. print run. Lyon: IARC (World Health Organization classification of tumours, 3 (der 4. ed.)).

International Human Genome Sequencing Consortium (2004): Finishing the euchromatic sequence of the human genome. In *Nature* 431 (7011), pp. 931–945. DOI: 10.1038/nature03001.

Isermann, Tamara; Sers, Christine; Der, Channing J.; Papke, Bjoern (2025): KRAS inhibitors: resistance drivers and combinatorial strategies. In *Trends in cancer* 11 (2), pp. 91–116. DOI: 10.1016/j.trecan.2024.11.009.

Jaime-Sanchez, Paula; Uranga-Murillo, Iratxe; Aguilo, Nacho; Khouili, Sofia C.; Arias, Maykel A.; Sancho, David; Pardo, Julian (2020): Cell death induced by cytotoxic CD8+ T cells is immunogenic and primes caspase-3-dependent spread immunity against endogenous tumor antigens. In *Journal for Immunotherapy of Cancer* 8 (1). DOI: 10.1136/jitc-2020-000528.

Johnson, Christian W.; Lin, Yi-Jang; Reid, Derion; Parker, Jillian; Pavlopoulos, Spiro; Dischinger, Patrick et al. (2019): Isoform-Specific Destabilization of the Active Site Reveals a Molecular Mechanism of Intrinsic Activation of KRas G13D. In *Cell reports* 28 (6), 1538-1550.e7. DOI: 10.1016/j.celrep.2019.07.026.

Johnson, Jared L.; Yaron, Tomer M.; Huntsman, Emily M.; Kerelsky, Alexander; Song, Junho; Regev, Amit et al. (2023): An atlas of substrate specificities for the human serine/threonine kinome. In *Nature* 613 (7945), pp. 759–766. DOI: 10.1038/s41586-022-05575-3.

Johnston, Harvey E.; Yadav, Kranthikumar; Kirkpatrick, Joanna M.; Biggs, George S.; Oxley, David; Kramer, Holger B.; Samant, Rahul S. (2022): Solvent Precipitation SP3 (SP4) Enhances Recovery for Proteomics Sample Preparation without Magnetic Beads. In *Analytical chemistry* 94 (29), pp. 10320–10328. DOI: 10.1021/acs.analchem.1c04200.

Kallingal, Anoop; Olszewski, Mateusz; Maciejewska, Natalia; Brankiewicz, Wioletta; Baginski, Maciej (2023): Cancer immune escape: the role of antigen presentation machinery. In *Journal of cancer research and clinical oncology* 149 (10), pp. 8131–8141. DOI: 10.1007/s00432-023-04737-8.

Kandoth, Cyriac; McLellan, Michael D.; Vandin, Fabio; Ye, Kai; Niu, Beifang; Lu, Charles et al. (2013): Mutational landscape and significance across 12 major cancer types. In *Nature* 502 (7471), pp. 333–339. DOI: 10.1038/nature12634.

Kataoka, Masahiro; Kitazawa, Masato; Nakamura, Satoshi; Koyama, Makoto; Yamamoto, Yuta; Miyazaki, Satoru et al. (2023): Cetuximab Enhances the Efficacy of MRTX1133, a Novel KRASG12D Inhibitor, in Colorectal Cancer Treatment. In *Anticancer research* 43 (10), pp. 4341–4348. DOI: 10.21873/anticanres.16629.

Kazama, Akira; Bilim, Vladimir; Tasaki, Masayuki; Anraku, Tsutomu; Kuroki, Hiroo; Shirono, Yuko et al. (2022): Tumor-infiltrating immune cell status predicts successful response to immune checkpoint inhibitors in renal cell carcinoma. In *Scientific reports* 12 (1), p. 20386. DOI: 10.1038/s41598-022-24437-6.

Kazi, Aslamuzzaman; Xiang, Shengyan; Yang, Hua; Delitto, Daniel; Trevino, José; Jiang, Rays H. Y. et al. (2018): GSK3 suppression upregulates β-catenin and c-Myc to abrogate KRas-dependent tumors. In *Nature communications* 9 (1), p. 5154. DOI: 10.1038/s41467-018-07644-6.

Khan, Zeba Firdos; Rathi, Aanchal; Khan, Afreen; Anjum, Farah; Chaudhury, Arunabh; Taiyab, Aaliya et al. (2024): Exploring PDK3 inhibition in lung cancer through drug repurposing for potential therapeutic interventions. In *Scientific reports* 14 (1), p. 29672. DOI: 10.1038/s41598-024-78022-0.

Klaeger, Susan; Apffel, Annie; Clauser, Karl R.; Sarkizova, Siranush; Oliveira, Giacomo; Rachimi, Suzanna et al. (2021): Optimized Liquid and Gas Phase Fractionation Increases HLA-Peptidome Coverage for Primary Cell and Tissue Samples. In *Molecular & cellular proteomics : MCP* 20, p. 100133. DOI: 10.1016/j.mcpro.2021.100133.

Koboldt, Daniel C.; Steinberg, Karyn Meltz; Larson, David E.; Wilson, Richard K.; Mardis, Elaine R. (2013): The next-generation sequencing revolution and its impact on genomics. In *Cell* 155 (1), pp. 27–38. DOI: 10.1016/j.cell.2013.09.006.

Köhler, B. C.; Haag, G. M.; Le Cornet, L.; Hoffmeister-Wittmann, P.; Schmidt, M.; Manjunath, A. et al. (2025): TROPHIT1—a randomized, open-label, multicenter, phase II/III trial of sacituzumab govitecan compared to standard of care in metastatic colorectal cancer patients. In *ESMO Gastrointestinal Oncology* 7, p. 100118. DOI: 10.1016/j.esmogo.2024.100118.

Kuharev, Jörg; Navarro, Pedro; Distler, Ute; Jahn, Olaf; Tenzer, Stefan (2015): In-depth evaluation of software tools for data-independent acquisition based label-free quantification. In *Proteomics* 15 (18), pp. 3140–3151. DOI: 10.1002/pmic.201400396.

Kümpers, Christiane; Jokic, Mladen; Haase, Ozan; Offermann, Anne; Vogel, Wenzel; Grätz, Victoria et al. (2019): Immune Cell Infiltration of the Primary Tumor, Not PD-L1 Status, Is Associated With Improved Response to Checkpoint Inhibition in Metastatic Melanoma. In *Frontiers in medicine* 6, p. 27. DOI: 10.3389/fmed.2019.00027.

Lalle, Guilhem; Lautraite, Raphaëlle; Bouherrou, Khaled; Plaschka, Maud; Pignata, Aurora; Voisin, Allison et al. (2024): NF-κB subunits RelA and c-Rel selectively control CD4+ T cell function in multiple sclerosis and cancer. In *J Exp Med* 221 (6). DOI: 10.1084/jem.20231348.

Langen, Adrianus Johannes de; Johnson, Melissa L.; Mazieres, Julien; Dingemans, Anne-Marie C.; Mountzios, Giannis; Pless, Miklos et al. (2023): Sotorasib versus docetaxel for previously treated non-small-cell lung cancer with KRASG12C mutation: a randomised, open-label, phase 3 trial. In *Lancet (London, England)* 401 (10378), pp. 733–746. DOI: 10.1016/S0140-6736(23)00221-0.

Latham, Bethany D.; Geffert, Raeanne M.; Jackson, Klarissa D. (2024): Kinase Inhibitors FDA Approved 2018-2023: Drug Targets, Metabolic Pathways, and Drug-Induced Toxicities. In *Drug metabolism and disposition: the biological fate of chemicals* 52 (6), pp. 479–492. DOI: 10.1124/dmd.123.001430.

Le, Dung T.; Uram, Jennifer N.; Wang, Hao; Bartlett, Bjarne R.; Kemberling, Holly; Eyring, Aleksandra D. et al. (2015): PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. In *The New England journal of medicine* 372 (26), pp. 2509–2520. DOI: 10.1056/NEJMoa1500596.

Le Tong; Cui, Weiyingqi; Zhang, Boya; Fonseca, Pedro; Zhao, Qian; Zhang, Ping et al. (2024): Patient-derived organoids in precision cancer medicine. In *Med (New York, N.Y.)* 5 (11), pp. 1351–1377. DOI: 10.1016/j.medj.2024.08.010.

Lee, Chien-Yun; The, Matthew; Meng, Chen; Bayer, Florian P.; Putzker, Kerstin; Müller, Julian et al. (2024): Illuminating phenotypic drug responses of sarcoma cells to kinase inhibitors by phosphoproteomics. In *Molecular systems biology* 20 (1), pp. 28–55. DOI: 10.1038/s44320-023-00004-7.

Lee, Sang-Kyu; Cho, Yong-Hee; Cha, Pu-Hyeon; Yoon, Jeong-Soo; Ro, Eun Ji; Jeong, Woo-Jeong et al. (2018): A small molecule approach to degrade RAS with EGFR repression is a potential therapy for KRAS mutation-driven colorectal cancer resistance to cetuximab. In *Experimental & molecular medicine* 50 (11), pp. 1–12. DOI: 10.1038/s12276-018-0182-2.

Lehtiö, Janne; Arslan, Taner; Siavelis, Ioannis; Pan, Yanbo; Socciarelli, Fabio; Berkovska, Olena et al. (2021): Proteogenomics of non-small cell lung cancer reveals molecular subtypes associated with specific therapeutic targets and immune evasion mechanisms. In *Nature cancer* 2 (11), pp. 1224–1242. DOI: 10.1038/s43018-021-00259-9.

Lemmon, Mark A.; Schlessinger, Joseph; Ferguson, Kathryn M. (2014): The EGFR family: not so prototypical receptor tyrosine kinases. In *Cold Spring Harb Perspect Biol* 6 (4), a020768. DOI: 10.1101/cshperspect.a020768.

Levitsky, Lev I.; Ivanov, Mark V.; Lobas, Anna A.; Gorshkov, Mikhail V. (2017): Unbiased False Discovery Rate Estimation for Shotgun Proteomics Based on the Target-Decoy Approach. In *Journal of proteome research* 16 (2), pp. 393–397. DOI: 10.1021/acs.jproteome.6b00144.

Li, Jingnan; Mizukami, Yusuke; Zhang, Xiaobo; Jo, Won-Seok; Chung, Daniel C. (2005): Oncogenic K-ras stimulates Wnt signaling in colon cancer through inhibition of GSK-3beta. In *Gastroenterology* 128 (7), pp. 1907–1918. DOI: 10.1053/j.gastro.2005.02.067.

Li, Yan; Ma, Yiqi; Wu, Zijun; Zeng, Fanxin; Song, Bin; Zhang, Yanrong et al. (2021): Tumor Mutational Burden Predicting the Efficacy of Immune Checkpoint Inhibitors in Colorectal Cancer: A Systematic Review and Meta-Analysis. In *Frontiers in immunology* 12, p. 751407. DOI: 10.3389/fimmu.2021.751407.

Li, Zi-Nan; Zhao, Lin; Yu, Li-Feng; Wei, Min-Jie (2020): BRAF and KRAS mutations in metastatic colorectal cancer: future perspectives for personalized therapy. In *Gastroenterology report* 8 (3), pp. 192–205. DOI: 10.1093/gastro/goaa022.

Liang, Yi-Hsin; Chen, Kuo-Hsing; Shao, Yu-Yun (2023): Treatment outcome comparisons of first-line targeted therapy in patients with KRAS wild-type metastatic colorectal cancer: A nationwide database study. In *Cancer medicine* 12 (14), pp. 15176–15186. DOI: 10.1002/cam4.6196.

Lièvre, Astrid; Bachet, Jean-Baptiste; Le Corre, Delphine; Boige, Valérie; Landi, Bruno; Emile, Jean-François et al. (2006): KRAS mutation status is predictive of response to cetuximab therapy in colorectal cancer. In *Cancer research* 66 (8), pp. 3992–3995. DOI: 10.1158/0008-5472.CAN-06-0191.

Liu, Yansheng; Beyer, Andreas; Aebersold, Ruedi (2016): On the Dependency of Cellular Protein Levels on mRNA Abundance. In *Cell* 165 (3), pp. 535–550. DOI: 10.1016/j.cell.2016.03.014.

Liu, Zhiqiang; Li, Liang; Liu, Lei; Zhu, Zhengming; Yu, Yi; Zhan, Shaochao et al. (2024): PDK3 drives colorectal carcinogenesis and immune evasion and is a therapeutic target for boosting immunotherapy. In *American journal of cancer research* 14 (6), pp. 3117–3129. DOI: 10.62347/QWKZ3078.

Loria, Anthony; Ramsdale, Erika E.; Aquina, Christopher T.; Cupertino, Paula; Mohile, Supriya G.; Fleming, Fergal J. (2024): From Clinical Trials to Practice: Anticipating and Overcoming Challenges in Implementing Watch-and-Wait for Rectal Cancer. In *JCO* 42 (8), pp. 876–880. DOI: 10.1200/JCO.23.01369.

Lu, Chun-Wun; Lin, Shau-Chieh; Chien, Chun-Wei; Lin, Shih-Chieh; Lee, Chung-Ta; Lin, Bo-Wen et al. (2011): Overexpression of pyruvate dehydrogenase kinase 3 increases drug resistance and early recurrence in colon cancer. In *The American journal of pathology* 179 (3), pp. 1405–1414. DOI: 10.1016/j.ajpath.2011.05.050.

Lu, Chun-Wun; Lin, Shih-Chieh; Chen, Ko-Fan; Lai, Yen-Yu; Tsai, Shaw-Jenq (2008): Induction of pyruvate dehydrogenase kinase-3 by hypoxia-inducible factor-1 promotes metabolic switch and drug resistance. In *The Journal of biological chemistry* 283 (42), pp. 28106–28114. DOI: 10.1074/jbc.M803508200.

Ludwig, Christina; Gillet, Ludovic; Rosenberger, George; Amon, Sabine; Collins, Ben C.; Aebersold, Ruedi (2018): Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial. In *Molecular systems biology* 14 (8), e8126. DOI: 10.15252/msb.20178126.

Luo, Qingqing; Quan, Yibo; Liu, Wei; Wu, Zixin; Qiu, Wenjing; Liang, Wenlong et al. (2024): Seed and Soil: Consensus Molecular Subgroups (CMS) and Tumor Microenvironment Features Between Primary Lesions and Metastases of Different Organ Sites in Colorectal Cancer. In *Cancer management and research* 16, pp. 225–243. DOI: 10.2147/CMAR.S441675.

Maes, Evelyne; Oeyen, Eline; Boonen, Kurt; Schildermans, Karin; Mertens, Inge; Pauwels, Patrick et al. (2019): The challenges of peptidomics in complementing proteomics in a clinical context. In *Mass spectrometry reviews* 38 (3), pp. 253–264. DOI: 10.1002/mas.21581.

Maier, Tobias; Güell, Marc; Serrano, Luis (2009): Correlation of mRNA and protein in complex biological samples. In *FEBS letters* 583 (24), pp. 3966–3973. DOI: 10.1016/j.febslet.2009.10.036.

Manca, Paolo; Corti, Francesca; Intini, Rossana; Mazzoli, Giacomo; Miceli, Rosalba; Germani, Marco Maria et al. (2023): Tumour mutational burden as a biomarker in patients with mismatch repair deficient/microsatellite instability-high metastatic colorectal cancer treated with immune checkpoint inhibitors. In *European journal of cancer (Oxford, England : 1990)* 187, pp. 15–24. DOI: 10.1016/j.ejca.2023.03.029.

Manjili, Masoud H.; Egilmez, Nejat; Knutson, Keith L.; Selvan, Senthamil R.; Ostberg, Julie R. (2012): Tumor escape and progression under immune pressure. In *Clinical & developmental immunology* 2012, p. 641079. DOI: 10.1155/2012/641079.

Mariuzza, Roy A.; Wu, Daichao; Pierce, Brian G. (2023): Structural basis for T cell recognition of cancer neoantigens and implications for predicting neoepitope immunogenicity. In *Frontiers in immunology* 14, p. 1303304. DOI: 10.3389/fimmu.2023.1303304.

Medina, Patrick J.; Adams, Val R. (2016): PD-1 Pathway Inhibitors: Immuno-Oncology Agents for Restoring Antitumor Immune Responses. In *Pharmacotherapy* 36 (3), pp. 317–334. DOI: 10.1002/phar.1714.

Meier, Florian; Brunner, Andreas-David; Koch, Scarlet; Koch, Heiner; Lubeck, Markus; Krause, Michael et al. (2018): Online Parallel Accumulation-Serial Fragmentation (PASEF) with a Novel Trapped Ion Mobility Mass Spectrometer. In *Molecular & cellular proteomics : MCP* 17 (12), pp. 2534–2545. DOI: 10.1074/mcp.TIR118.000900.

Mertins, Philipp; Tang, Lauren C.; Krug, Karsten; Clark, David J.; Gritsenko, Marina A.; Chen, Lijun et al. (2018): Reproducible workflow for multiplexed deep-scale proteome and phosphoproteome analysis of tumor tissues by liquid chromatography-mass spectrometry. In *Nature protocols* 13 (7), pp. 1632–1661. DOI: 10.1038/s41596-018-0006-9.

Meston, Daniel (2022): Pitfalls in Proteomics: Avoiding Problems That Can Occur Before Data Acquisition Begins. In *LCGC N. Am.*, pp. 524–528. DOI: 10.56530/lcgc.na.vp4480b1.

Meylan, Etienne; Dooley, Alison L.; Feldser, David M.; Shen, Lynn; Turk, Erin; Ouyang, Chensi; Jacks, Tyler (2009): Requirement for NF-kappaB signalling in a mouse model of lung adenocarcinoma. In *Nature* 462 (7269), pp. 104–107. DOI: 10.1038/nature08462.

Migneault, Isabelle; Hunter, Joanna M. (2009): Industrialized MS-based proteomics in the search for circulating biomarkers. In *Bioanalysis* 1 (6), pp. 1149–1163. DOI: 10.4155/bio.09.105.

Miller, K.; Beardmore, J.; Kanety, H.; Schlessinger, J.; Hopkins, C. R. (1986): Localization of the epidermal growth factor (EGF) receptor within the endosome of EGF-stimulated epidermoid carcinoma (A431) cells. In *The Journal of cell biology* 102 (2), pp. 500–509. DOI: 10.1083/jcb.102.2.500.

Miyashita, Hirotaka; Hong, David S. (2024): Combining EGFR and KRAS G12C Inhibitors for KRAS G12C Mutated Advanced Colorectal Cancer. In *Journal of cancer immunology* 6 (2), pp. 62–69. DOI: 10.33696/cancerimmunol.6.086.

Mock, Andreas; Teleanu, Maria-Veronica; Kreutzfeldt, Simon; Heilig, Christoph E.; Hüllein, Jennifer; Möhrmann, Lino et al. (2023): NCT/DKFZ MASTER handbook of interpreting whole-genome, transcriptome, and methylome data for precision oncology. In *NPJ precision oncology* 7 (1), p. 109. DOI: 10.1038/s41698-023-00458-w.

Moeckel, Camille; Bakhl, Katrina; Georgakopoulos-Soares, Ilias; Zaravinos, Apostolos (2023): The Efficacy of Tumor Mutation Burden as a Biomarker of Response to Immune Checkpoint Inhibitors. In *International journal of molecular sciences* 24 (7). DOI: 10.3390/ijms24076710.

Mok, Tony S. K.; Yao, Wenxiu; Duruisseaux, Michaël; Doucet, Ludovic; Azkárate Martínez, Aitor; Gregorc, Vanesa et al. (2024): KRYSTAL-12: Phase 3 study of adagrasib versus docetaxel in patients with previously treated advanced/metastatic non-small cell lung cancer (NSCLC) harboring a KRAS G12C mutation. In *JCO* 42 (17_suppl), LBA8509-LBA8509. DOI: 10.1200/JCO.2024.42.17_suppl.LBA8509.

Montesion, Meagan; Murugesan, Karthikeyan; Jin, Dexter X.; Sharaf, Radwa; Sanchez, Nora; Guria, Ameet et al. (2021): Somatic HLA Class I Loss Is a Widespread Mechanism of Immune Evasion Which Refines the Use of Tumor Mutational Burden as a Biomarker of Checkpoint Inhibitor Response. In *Cancer discovery* 11 (2), pp. 282–292. DOI: 10.1158/2159-8290.CD-20-0672.

Moorman, Andrew; Benitez, Elizabeth K.; Cambulli, Francesco; Jiang, Qingwen; Mahmoud, Ahmed; Lumish, Melissa et al. (2025): Progressive plasticity during colorectal cancer metastasis. In *Nature* 637 (8047), pp. 947–954. DOI: 10.1038/s41586-024-08150-0.

Moretto, Roberto; Rossini, Daniele; Catteau, Aurélie; Antoniotti, Carlotta; Giordano, Mirella; Boccaccino, Alessandra et al. (2023): Dissecting tumor lymphocyte infiltration to predict benefit from immune-checkpoint inhibitors in metastatic colorectal cancer: lessons from the AtezoT RIBE study. In *Journal for Immunotherapy of Cancer* 11 (4). DOI: 10.1136/jitc-2022-006633.

Müller, Torsten; Kalxdorf, Mathias; Longuespée, Rémi; Kazdal, Daniel N.; Stenzinger, Albrecht; Krijgsveld, Jeroen (2020): Automated sample preparation with SP3 for low-input clinical proteomics. In *Molecular systems biology* 16 (1), e9111. DOI: 10.15252/msb.20199111.

Müller-Dott, Sophia; Tsirvouli, Eirini; Vazquez, Miguel; Ramirez Flores, Ricardo O.; Badia-I-Mompel, Pau; Fallegger, Robin et al. (2023): Expanding the coverage of regulons from high-confidence prior knowledge for accurate estimation of transcription factor activities. In *Nucleic acids research* 51 (20), pp. 10934–10949. DOI: 10.1093/nar/gkad841.

Mundhara, Nikita; Sadhukhan, Pritam (2024): Cracking the Codes behind Cancer Cells' Immune Evasion. In *International journal of molecular sciences* 25 (16). DOI: 10.3390/ijms25168899.

Mundt, Filip; Albrechtsen, Nicolai J. Wewer; Mann, Sebastian Porsdam; Treit, Peter; Ghodgaonkar-Steger, Medini; O'Flaherty, Martina et al. (2023): Foresight in clinical proteomics: current status, ethical considerations, and future perspectives. In *Open Res Europe* 3, p. 59. DOI: 10.12688/openreseurope.15810.2.

Murillo, Jimmy Rodriguez; Kuras, Magdalena; Rezeli, Melinda; Miliotis, Tasso; Betancourt, Lazaro; Marko-Varga, Gyorgy (2018): Automated phosphopeptide enrichment from minute quantities of frozen malignant melanoma tissue. In *PloS one* 13 (12), e0208562. DOI: 10.1371/journal.pone.0208562.

Nandal, Savita; Burt, Tal (2017): Integrating Pharmacoproteomics into Early-Phase Clinical Development: State-of-the-Art, Challenges, and Recommendations. In *International journal of molecular sciences* 18 (2). DOI: 10.3390/ijms18020448.

Narasimhan, Vignesh; Wright, Josephine A.; Churchill, Michael; Wang, Tongtong; Rosati, Rachele; Lannagan, Tamsin R. M. et al. (2020): Medium-throughput Drug Screening of Patient-derived Organoids from Colorectal Peritoneal Metastases to Direct Personalized Therapy. In *Clinical cancer research : an official journal of the American Association for Cancer Research* 26 (14), pp. 3662–3670. DOI: 10.1158/1078-0432.CCR-20-0073.

Needham, Elise J.; Parker, Benjamin L.; Burykin, Timur; James, David E.; Humphrey, Sean J. (2019): Illuminating the dark phosphoproteome. In *Sci. Signal.* 12 (565). DOI: 10.1126/scisignal.aau8645.

Neumann, Jens; Zeindl-Eberhart, Evelyn; Kirchner, Thomas; Jung, Andreas (2009): Frequency and type of KRAS mutations in routine diagnostic analysis of metastatic colorectal cancer. In *Pathology, research and practice* 205 (12), pp. 858–862. DOI: 10.1016/j.prp.2009.07.010.

Niu, Beifang; Ye, Kai; Zhang, Qunyuan; Lu, Charles; Xie, Mingchao; McLellan, Michael D. et al. (2014): MSIsensor: microsatellite instability detection using paired tumor-normal sequence data. In *Bioinformatics* 30 (7), pp. 1015–1016. DOI: 10.1093/bioinformatics/btt755.

Nojima, Yuji; Shimizu, Katsuhiko; Saisho, Shinsuke; Maeda, A. I.; Kurosaki, Takeshi; Kurose, Koji et al. (2021): Tumor PD-L1 and VEGF Expression, and CD8 T Cell Infiltration Predict Clinical Response to Immune Checkpoint Inhibitors in Non-small Cell Lung Cancer. In *Anticancer research* 41 (11), pp. 5469–5475. DOI: 10.21873/anticanres.15359.

Nowell, P. C. (1976): The clonal evolution of tumor cell populations. In *Science (New York, N.Y.)* 194 (4260), pp. 23–28. DOI: 10.1126/science.959840.

Ohta, Yuki; Fujii, Masayuki; Takahashi, Sirirat; Takano, Ai; Nanki, Kosaku; Matano, Mami et al. (2022): Cell-matrix interface regulates dormancy in human colon cancer stem cells. In *Nature* 608 (7924), pp. 784–794. DOI: 10.1038/s41586-022-05043-y.

Osterlund, Emerik; Ristimäki, Ari; Kytölä, Soili; Kuopio, Teijo; Heervä, Eetu; Muhonen, Timo et al. (2022): KRAS-G12C Mutation in One Real-Life and Three Population-Based Nordic Cohorts of Metastatic Colorectal Cancer. In *Frontiers in oncology* 12, p. 826073. DOI: 10.3389/fonc.2022.826073.

Paschke, Stephan; Jafarov, Sakhavat; Staib, Ludger; Kreuser, Ernst-Dietrich; Maulbecker-Armstrong, Catharina; Roitman, Marc et al. (2018): Are Colon and Rectal Cancer Two Different Tumor Entities? A Proposal to Abandon the Term Colorectal Cancer. In *International journal of molecular sciences* 19 (9), p. 2577. DOI: 10.3390/ijms19092577.

Perkins, David N.; Pappin, Darryl J. C.; Creasy, David M.; Cottrell, John S. (1999): Probability-based protein identification by searching sequence databases using mass spectrometry data. In *Electrophoresis* 20 (18), pp. 3551–3567. DOI: 10.1002/(SICI)1522-2683(19991201)20:18<3551::AID-ELPS3551>3.0.CO;2-2.

Piazza, Gary A.; Chandrasekaran, Preethi; Maxuitenko, Yulia Y.; Budhwani, Karim I. (2024): Assessment of KRASG12C inhibitors for colorectal cancer. In *Frontiers in oncology* 14, p. 1412435. DOI: 10.3389/fonc.2024.1412435.

Polak, Roel; Zhang, Elisa T.; Kuo, Calvin J. (2024): Cancer organoids 2.0: modelling the complexity of the tumour immune microenvironment. In *Nature reviews. Cancer* 24 (8), pp. 523–539. DOI: 10.1038/s41568-024-00706-6.

Potel, Clement M.; Lin, Miao-Hsia; Heck, Albert J. R.; Lemeer, Simone (2018): Defeating Major Contaminants in Fe3+- Immobilized Metal Ion Affinity Chromatography (IMAC) Phosphopeptide Enrichment. In *Molecular & cellular proteomics : MCP* 17 (5), pp. 1028–1034. DOI: 10.1074/mcp.TIR117.000518.

Potocki, Paweł M.; Wójcik, Piotr; Chmura, Łukasz; Goc, Bartłomiej; Fedewicz, Marcin; Bielańska, Zofia et al. (2023): Clinical Characterization of Targetable Mutations (BRAF V600E and KRAS G12C) in Advanced Colorectal Cancer-A Nation-Wide Study. In *International journal of molecular sciences* 24 (10). DOI: 10.3390/ijms24109073.

Poulin, Emily J.; Bera, Asim K.; Lu, Jia; Lin, Yi-Jang; Strasser, Samantha Dale; Paulo, Joao A. et al. (2019): Tissue-Specific Oncogenic Activity of KRASA146T. In *Cancer discovery* 9 (6), pp. 738–755. DOI: 10.1158/2159-8290.CD-18-1220.

Poulos, Rebecca C.; Hains, Peter G.; Shah, Rohan; Lucas, Natasha; Xavier, Dylan; Manda, Srikanth S. et al. (2020): Strategies to enable large-scale proteomics for reproducible research. In *Nature communications* 11 (1), p. 3793. DOI: 10.1038/s41467-020-17641-3.

Prahallad, Anirudh; Sun, Chong; Huang, Sidong; Di Nicolantonio, Federica; Salazar, Ramon; Zecchin, Davide et al. (2012): Unresponsiveness of colon cancer to BRAF(V600E) inhibition through feedback activation of EGFR. In *Nature* 483 (7387), pp. 100–103. DOI: 10.1038/nature10868.

Prescott, Jack A.; Cook, Simon J. (2018): Targeting IKKβ in Cancer: Challenges and Opportunities for the Therapeutic Utilisation of IKKβ Inhibitors. In *Cells* 7 (9). DOI: 10.3390/cells7090115.

Puliga, Elisabetta; Bellis, Chiara de; Vietti Michelina, Sandra; Capeloa, Tania; Migliore, Cristina; Orrù, Claudia et al. (2024): Biological and targeting differences between the rare KRAS A146T and canonical KRAS mutants in gastric cancer models. In *Gastric cancer : official journal of the International Gastric Cancer Association and the Japanese Gastric Cancer Association* 27 (3), pp. 473–483. DOI: 10.1007/s10120-024-01468-8.

Qiu, Junyu; Cheng, Zilin; Jiang, Zheng; Gan, Luhan; Zhang, Zixuan; Xie, Zhenzhen (2024): Immunomodulatory Precision: A Narrative Review Exploring the Critical Role of Immune Checkpoint Inhibitors in Cancer Treatment. In *International journal of molecular sciences* 25 (10). DOI: 10.3390/ijms25105490.

Qu, Shanqiang; Xu, Rongyang; Yi, Guozhong; Li, Zhiyong; Zhang, Huayang; Qi, Songtao; Huang, Guanglong (2024): Patient-derived organoids in human cancer: a platform for fundamental research and precision medicine. In *Molecular biomedicine* 5 (1), p. 6. DOI: 10.1186/s43556-023-00165-9.

Radanova, Maria; Mihaylova, Galya; Stoyanov, George St; Draganova, Vyara; Zlatarov, Aleksandar; Kolev, Nikola et al. (2023): KRAS Mutation Status in Bulgarian Patients with Advanced and Metastatic Colorectal Cancer. In *International journal of molecular sciences* 24 (16). DOI: 10.3390/ijms241612753.

Rahbari, Nuh N.; Carr, Prudence R.; Jansen, Lina; Chang-Claude, Jenny; Weitz, Jürgen; Hoffmeister, Michael; Brenner, Hermann (2019): Time of Metastasis and Outcome in Colorectal Cancer. In *Annals of surgery* 269 (3), pp. 494–502. DOI: 10.1097/SLA.0000000000002564.

Reck, Martin; Rodríguez-Abreu, Delvys; Robinson, Andrew G.; Hui, Rina; Csőszi, Tibor; Fülöp, Andrea et al. (2019): Updated Analysis of KEYNOTE-024: Pembrolizumab Versus Platinum-Based Chemotherapy for Advanced Non-Small-Cell Lung Cancer With PD-L1 Tumor Proportion Score of 50% or Greater. In *JCO* 37 (7), pp. 537–546. DOI: 10.1200/JCO.18.00149.

Rejali, Leili; Seifollahi Asl, Romina; Sanjabi, Fatemeh; Fatemi, Nayeralsadat; Asadzadeh Aghdaei, Hamid; Saeedi Niasar, Mahsa et al. (2023): Principles of Molecular Utility for CMS Classification in Colorectal Cancer Management. In *Cancers* 15 (10). DOI: 10.3390/cancers15102746.

Riihimäki, Matias; Hemminki, Akseli; Sundquist, Jan; Hemminki, Kari (2016): Patterns of metastasis in colon and rectal cancer. In *Scientific reports* 6, p. 29765. DOI: 10.1038/srep29765.

Ritchie, Matthew E.; Phipson, Belinda; Di Wu; Hu, Yifang; Law, Charity W.; Shi, Wei; Smyth, Gordon K. (2015): limma powers differential expression analyses for RNA-sequencing and microarray studies. In *Nucleic acids research* 43 (7), e47. DOI: 10.1093/nar/gkv007.

Ruprecht, Benjamin; Koch, Heiner; Domasinska, Petra; Frejno, Martin; Kuster, Bernhard; Lemeer, Simone (2017): Optimized Enrichment of Phosphoproteomes by Fe-IMAC Column Chromatography. In *Methods in molecular biology (Clifton, N.J.)* 1550, pp. 47–60. DOI: 10.1007/978-1-4939-6747-6_5.

Ruprecht, Benjamin; Koch, Heiner; Medard, Guillaume; Mundt, Max; Kuster, Bernhard; Lemeer, Simone (2015): Comprehensive and reproducible phosphopeptide enrichment using iron immobilized metal ion affinity chromatography (Fe-IMAC) columns. In *Molecular & cellular proteomics : MCP* 14 (1), pp. 205–215. DOI: 10.1074/mcp.M114.043109.

Sade-Feldman, Moshe; Jiao, Yunxin J.; Chen, Jonathan H.; Rooney, Michael S.; Barzily-Rokni, Michal; Eliane, Jean-Pierre et al. (2017): Resistance to checkpoint blockade therapy through inactivation of antigen presentation. In *Nature communications* 8 (1), p. 1136. DOI: 10.1038/s41467-017-01062-w.

Sahin, Ibrahim Halil; Goyal, Subir; Pumpalova, Yoanna; Sonbol, Mohamad B.; Das, Satya; Haraldsdottir, Sigurdis et al. (2021): Mismatch Repair (MMR) Gene Alteration and BRAF V600E Mutation Are Potential Predictive Biomarkers of Immune Checkpoint Inhibitors in MMR-Deficient Colorectal Cancer. In *The Oncologist* 26 (8), pp. 668–675. DOI: 10.1002/onco.13741.

Saleh, Khalil; Ibrahim, Rebecca; Khoury, Rita; Tikriti, Zamzam; Khalife, Nadine (2024): KRAS and EGFR inhibitors: a new step in the management of colorectal cancer. In *Translational gastroenterology and hepatology* 9, p. 56. DOI: 10.21037/tgh-24-73.

San-Román-Gil, María; Torres-Jiménez, Javier; Pozas, Javier; Esteban-Villarrubia, Jorge; Albarrán-Fernández, Víctor; Álvarez-Ballesteros, Pablo et al. (2023): Current Landscape and Potential Challenges of Immune Checkpoint Inhibitors in Microsatellite Stable Metastatic Colorectal Carcinoma. In *Cancers* 15 (3). DOI: 10.3390/cancers15030863.

Santos, Rita; Ursu, Oleg; Gaulton, Anna; Bento, A. Patrícia; Donadi, Ramesh S.; Bologa, Cristian G. et al. (2017): A comprehensive map of molecular drug targets. In *Nature reviews. Drug discovery* 16 (1), pp. 19–34. DOI: 10.1038/nrd.2016.230.

Sato, Toshiro; Vries, Robert G.; Snippert, Hugo J.; van de Wetering, Marc; Barker, Nick; Stange, Daniel E. et al. (2009): Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche. In *Nature* 459 (7244), pp. 262–265. DOI: 10.1038/nature07935.

Schmelzle, Katrin; White, Forest M. (2006): Phosphoproteomic approaches to elucidate cellular signaling networks. In *Current opinion in biotechnology* 17 (4), pp. 406–414. DOI: 10.1016/j.copbio.2006.06.004.

Schneider, Nora I.; Langner, Cord (2014): Prognostic stratification of colorectal cancer patients: current perspectives. In *Cancer management and research* 6, pp. 291–300. DOI: 10.2147/CMAR.S38827.

Schubert, Michael; Klinger, Bertram; Klünemann, Martina; Sieber, Anja; Uhlitz, Florian; Sauer, Sascha et al. (2018): Perturbation-response genes reveal signaling footprints in cancer gene expression. In *Nature communications* 9 (1), p. 20. DOI: 10.1038/s41467-017-02391-6.

Shaul, Yoav; Seger, Rony (2005): The detection of MAPK signaling. In *Current protocols in cell biology* Chapter 14, Unit 14.3. DOI: 10.1002/0471143030.cb1403s28.

Shen, Ke-Yu; Zhu, Ying; Xie, Sun-Zhe; Qin, Lun-Xiu (2024): Immunosuppressive tumor microenvironment and immunotherapy of hepatocellular carcinoma: current status and prospectives. In *Journal of hematology & oncology* 17 (1), p. 25. DOI: 10.1186/s13045-024-01549-2.

Shergill, Ardaman; Liao, Chih-Yi; Kindler, Hedy L.; Polite, Blase N.; Catenacci, Daniel V.T. (2022): A phase 1b/2 study of VS-6766 in combination cetuximab in patients (pts) with advanced KRAS mt colorectal cancer (CRC). In *JCO* 40 (4_suppl), TPS219-TPS219. DOI: 10.1200/JCO.2022.40.4_suppl.TPS219.

Shi, Yixin; Guan, Zhanwen; Cai, Gengxi; Nie, Yichu; Zhang, Chuling; Luo, Wei; Liu, Jia (2024): Patient-derived organoids: a promising tool for breast cancer research. In *Frontiers in oncology* 14, p. 1350935. DOI: 10.3389/fonc.2024.1350935.

Shiu, K-K.; André, T.; Kim, T. W.; Vittrup Jensen, B.; Jensen, L. H.; Punt, C.J.A. et al. (2023): LBA32 Pembrolizumab versus chemotherapy in microsatellite instability-high (MSI-H)/mismatch repair-deficient (dMMR) metastatic colorectal cancer (mCRC): 5-year follow-up of the randomized phase III KEYNOTE-177 study. In *Annals of Oncology* 34, S1271-S1272. DOI: 10.1016/j.annonc.2023.10.024.

Siegel, Rebecca L.; Fedewa, Stacey A.; Anderson, William F.; Miller, Kimberly D.; Ma, Jiemin; Rosenberg, Philip S.; Jemal, Ahmedin (2017): Colorectal Cancer Incidence Patterns in the United States, 1974-2013. In *Journal of the National Cancer Institute* 109 (8). DOI: 10.1093/jnci/djw322.

Siegel, Rebecca L.; Torre, Lindsey A.; Soerjomataram, Isabelle; Hayes, Richard B.; Bray, Freddie; Weber, Thomas K.; Jemal, Ahmedin (2019): Global patterns and trends in colorectal cancer incidence in young adults. In *Gut* 68 (12), pp. 2179–2185. DOI: 10.1136/gutjnl-2019-319511.

Siegel, Rebecca L.; Wagle, Nikita Sandeep; Cercek, Andrea; Smith, Robert A.; Jemal, Ahmedin (2023): Colorectal cancer statistics, 2023. In *CA: a cancer journal for clinicians* 73 (3), pp. 233–254. DOI: 10.3322/caac.21772.

Sim, Malcolm J. W.; Sun, Peter D. (2022): T Cell Recognition of Tumor Neoantigens and Insights Into T Cell Immunotherapy. In *Frontiers in immunology* 13, p. 833017. DOI: 10.3389/fimmu.2022.833017.

Sinicrope, Frank A.; Sargent, Daniel J. (2012): Molecular pathways: microsatellite instability in colorectal cancer: prognostic, predictive, and therapeutic implications. In *Clinical cancer research : an official journal of the American Association for Cancer Research* 18 (6), pp. 1506–1512. DOI: 10.1158/1078-0432.CCR-11-1469.

Skoulidis, Ferdinandos; Li, Bob T.; Dy, Grace K.; Price, Timothy J.; Falchook, Gerald S.; Wolf, Jürgen et al. (2021): Sotorasib for Lung Cancers with KRAS p.G12C Mutation. In *The New England journal of medicine* 384 (25), pp. 2371–2381. DOI: 10.1056/NEJMoa2103695.

Smabers, Lidwien P.; Wensink, Emerens; Verissimo, Carla S.; Koedoot, Esmee; Pitsa, Katerina-Chara; Huismans, Maarten A. et al. (2024): Organoids as a biomarker for personalized treatment in metastatic colorectal cancer: drug screen optimization and correlation with patient response. In *Journal of experimental & clinical cancer research : CR* 43 (1), p. 61. DOI: 10.1186/s13046-024-02980-6.

Solé, Laura; Lobo-Jarne, Teresa; Álvarez-Villanueva, Daniel; Alonso-Marañón, Josune; Guillén, Yolanda; Guix, Marta et al. (2022): p53 wild-type colorectal cancer cells that express a fetal gene signature are associated with metastasis and poor prognosis. In *Nature communications* 13 (1), p. 2866. DOI: 10.1038/s41467-022-30382-9.

SOPs: /LINCS/Overview Information (2025). Available online at https://panoramaweb.org/wiki/LINCS/Overview%20Information/page.view?name=sops, updated on 4/14/2025, checked on 4/14/2025.

Sorkin, A.; Krolenko, S.; Kudrjavtceva, N.; Lazebnik, J.; Teslenko, L.; Soderquist, A. M.; Nikolsky, N. (1991): Recycling of epidermal growth factor-receptor complexes in A431 cells: identification of dual pathways. In *The Journal of cell biology* 112 (1), pp. 55–63. DOI: 10.1083/jcb.112.1.55.

Sousa Abreu, Raquel de; Penalva, Luiz O.; Marcotte, Edward M.; Vogel, Christine (2009): Global signatures of protein and mRNA expression levels. In *Molecular bioSystems* 5 (12), pp. 1512–1526. DOI: 10.1039/b908315d.

Steimle, V.; Siegrist, C. A.; Mottet, A.; Lisowska-Grospierre, B.; Mach, B. (1994): Regulation of MHC class II expression by interferon-gamma mediated by the transactivator gene CIITA. In *Science (New York, N.Y.)* 265 (5168), pp. 106–109. DOI: 10.1126/science.8016643.

Steimle, Viktor; Otten, Luc A.; Zufferey, Madeleine; Mach, Bernard (1993): Complementation cloning of an MHC class II transactivator mutated in hereditary MHC class II deficiency (or bare lymphocyte syndrome). In *Cell* 75 (1), pp. 135–146. DOI: 10.1016/s0092-8674(05)80090-x.

Stratton, Michael R.; Campbell, Peter J.; Futreal, P. Andrew (2009): The cancer genome. In *Nature* 458 (7239), pp. 719–724. DOI: 10.1038/nature07943.

Sun, Yiting; Yang, Bowen; Wen, Ti; Guo, Xiaoyu; Li, Danni; Shi, Ruichuan et al. (2025): ANXA10 sensitizes microsatellite instability-high colorectal cancer to anti-PD-1 immunotherapy via assembly of HLA-DR dimers by regulating CD74. In *Cell biology and toxicology* 41 (1), p. 25. DOI: 10.1007/s10565-024-09982-2.

Suresh, Shruthy; O'Donnell, Kathryn A. (2021): Translational Control of Immune Evasion in Cancer. In *Trends in cancer* 7 (7), pp. 580–582. DOI: 10.1016/j.trecan.2021.04.002.

Taieb, Julien; Gallois, Claire (2020): Adjuvant Chemotherapy for Stage III Colon Cancer. In *Cancers* 12 (9). DOI: 10.3390/cancers12092679.

Tan, Yiqing; Yang, Yiping; Zhang, Mingjun; Li, Ni; Hu, Lei; Deng, Mingyou et al. (2025): IRF4 as a molecular biomarker in pan-cancer through multiple omics integrative analysis. In *Analytical methods : advancing methods and applications* 17 (16), pp. 3183–3201. DOI: 10.1039/d4ay02269f.

Taniguchi, Koji; Karin, Michael (2018): NF-κB, inflammation, immunity and cancer: coming of age. In *Nature reviews. Immunology* 18 (5), pp. 309–324. DOI: 10.1038/nri.2017.142.

Tauriello, Daniele V. F.; Palomo-Ponce, Sergio; Stork, Diana; Berenguer-Llergo, Antonio; Badia-Ramentol, Jordi; Iglesias, Mar et al. (2018): TGFβ drives immune evasion in genetically reconstituted colon cancer metastasis. In *Nature* 554 (7693), pp. 538–543. DOI: 10.1038/nature25492.

Tögel, Lars; Schubart, Christoph; Lettmaier, Sebastian; Neufert, Clemens; Hoyer, Juliane; Wolff, Kerstin et al. (2023): Determinants Affecting the Clinical Implementation of a Molecularly Informed Molecular Tumor Board Recommendation: Experience from a Tertiary Cancer Center. In *Cancers* 15 (24). DOI: 10.3390/cancers15245892.

Tomioka, Ayana; Tomioka, Ryota; Mori, Issei; Isobe, Yosuke; Arita, Makoto; Imami, Koshi et al. (2023): One thousand samples per day capillary-flow LC/MS/MS for high-speed, high-sensitivity and in-depth proteomics.

Topalian, Suzanne L.; Hodi, F. Stephen; Brahmer, Julie R.; Gettinger, Scott N.; Smith, David C.; McDermott, David F. et al. (2012): Safety, activity, and immune correlates of anti-PD-1 antibody in cancer. In *The New England journal of medicine* 366 (26), pp. 2443–2454. DOI: 10.1056/NEJMoa1200690.

Türei, Dénes; Valdeolivas, Alberto; Gul, Lejla; Palacio-Escat, Nicolàs; Klein, Michal; Ivanova, Olga et al. (2021): Integrated intra- and intercellular signaling knowledge for multicellular omics analysis. In *Molecular systems biology* 17 (3), e9923. DOI: 10.15252/msb.20209923.

Tyanova, Stefka; Cox, Juergen (2018): Perseus: A Bioinformatics Platform for Integrative Analysis of Proteomics Data in Cancer Research. In *Methods in molecular biology (Clifton, N.J.)* 1711, pp. 133–148. DOI: 10.1007/978-1-4939-7493-1_7.

van Cutsem, Eric; Peeters, Marc; Siena, Salvatore; Humblet, Yves; Hendlisz, Alain; Neyns, Bart et al. (2007): Open-label phase III trial of panitumumab plus best supportive care compared with best supportive care alone in patients with chemotherapy-refractory metastatic colorectal cancer. In *JCO* 25 (13), pp. 1658–1664. DOI: 10.1200/JCO.2006.08.1620.

van de Wetering, Marc; Francies, Hayley E.; Francis, Joshua M.; Bounova, Gergana; Iorio, Francesco; Pronk, Apollo et al. (2015): Prospective derivation of a living organoid biobank of colorectal cancer patients. In *Cell* 161 (4), pp. 933–945. DOI: 10.1016/j.cell.2015.03.053.

van den Eshof, Bart L.; Medfai, Lobna; Nolfi, Emanuele; Wawrzyniuk, Magdalena; Sijts, Alice J. A. M. (2021): The Function of Immunoproteasomes-An Immunologists' Perspective. In *Cells* 10 (12). DOI: 10.3390/cells10123360.

van 't Erve, Iris; Wesdorp, Nina J.; Medina, Jamie E.; Ferreira, Leonardo; Leal, Alessandro; Huiskens, Joost et al. (2021): KRAS A146 Mutations Are Associated With Distinct Clinical Behavior in Patients With Colorectal Liver Metastases. In *JCO precision oncology* 5. DOI: 10.1200/PO.21.00223.

Vasaikar, Suhas; Huang, Chen; Wang, Xiaojing; Petyuk, Vladislav A.; Savage, Sara R.; Wen, Bo et al. (2019): Proteogenomic Analysis of Human Colon Cancer Reveals New Therapeutic Opportunities. In *Cell* 177 (4), 1035-1049.e19. DOI: 10.1016/j.cell.2019.03.030.

Vidri, Roberto J.; Fitzgerald, Timothy L. (2020): GSK-3: An important kinase in colon and pancreatic cancers. In *Biochimica et biophysica acta. Molecular cell research* 1867 (4), p. 118626. DOI: 10.1016/j.bbamcr.2019.118626.

Vinaiphat, Arada; Low, Jee Keem; Yeoh, Kheng Wei; Chng, Wee Joo; Sze, Siu Kwan (2021): Application of Advanced Mass Spectrometry-Based Proteomics to Study Hypoxia Driven Cancer Progression. In *Frontiers in oncology* 11, p. 559822. DOI: 10.3389/fonc.2021.559822.

Vogel, Christine; Marcotte, Edward M. (2012): Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. In *Nature reviews. Genetics* 13 (4), pp. 227–232. DOI: 10.1038/nrg3185.

Vogelstein, Bert; Papadopoulos, Nickolas; Velculescu, Victor E.; Zhou, Shibin; Diaz, Luis A.; Kinzler, Kenneth W. (2013): Cancer genome landscapes. In *Science (New York, N.Y.)* 339 (6127), pp. 1546–1558. DOI: 10.1126/science.1235122.

Wadhwa, Pankaj; Jain, Priti; Jadhav, Hemant R. (2020): Glycogen Synthase Kinase 3 (GSK3): Its Role and Inhibitors. In *Current topics in medicinal chemistry* 20 (17), pp. 1522–1534. DOI: 10.2174/1568026620666200516153136.

Wang, Jianghua; Li, Xiaoting; Xiao, Guanxi; Desai, Jayesh; Frentzas, Sophia; Wang, Zhongmin Maxwell et al. (2024): CD74 is associated with inflamed tumor immune microenvironment and predicts responsiveness to PD-1/CTLA-4 bispecific antibody in patients with solid tumors. In *Cancer immunology, immunotherapy : CII* 73 (2), p. 36. DOI: 10.1007/s00262-023-03604-2.

Wang, Mona Meng; Coupland, Sarah E.; Aittokallio, Tero; Figueiredo, Carlos R. (2023): Resistance to immune checkpoint therapies by tumour-induced T-cell desertification and exclusion: key mechanisms, prognostication and new therapeutic opportunities. In *British journal of cancer* 129 (8), pp. 1212–1224. DOI: 10.1038/s41416-023-02361-4.

Wang, Quanhui; Wen, Bo; Yan, Guangrong; Wei, Junying; Xie, Liqi; Xu, Shaohang et al. (2013): Qualitative and quantitative expression status of the human chromosome 20 genes in cancer tissues and the representative cell lines. In *Journal of proteome research* 12 (1), pp. 151–161. DOI: 10.1021/pr3008336.

Wang, Stephen L.; Chan, Timothy A. (2025): Navigating established and emerging biomarkers for immune checkpoint inhibitor therapy. In *Cancer cell* 43 (4), pp. 641–664. DOI: 10.1016/j.ccell.2025.03.006.

Weeber, Fleur; van de Wetering, Marc; Hoogstraat, Marlous; Dijkstra, Krijn K.; Krijgsman, Oscar; Kuilman, Thomas et al. (2015): Preserved genetic diversity in organoids cultured from biopsies of human colorectal cancer metastases. In *Proceedings of the National Academy of Sciences of the United States of America* 112 (43), pp. 13308–13311. DOI: 10.1073/pnas.1516689112.

Wen, Lei; Chen, Zhigang; Ji, Xiaomeng; Fong, William Pat; Shao, Qiong; Ren, Chao et al. (2022): Pathological complete response to immune checkpoint inhibitor in patients with colorectal cancer liver metastases harboring POLE exonuclease domain mutation. In *Journal for Immunotherapy of Cancer* 10 (7). DOI: 10.1136/jitc-2022-004487.

Wessel, D.; Flügge, U. I. (1984): A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. In *Analytical biochemistry* 138 (1), pp. 141–143. DOI: 10.1016/0003-2697(84)90782-6.

Whiteaker, Jeffrey R. (2010): The increasing role of mass spectrometry in quantitative clinical proteomics. In *Clinical chemistry* 56 (9), pp. 1373–1374. DOI: 10.1373/clinchem.2010.150383.

Wittekind, Christian (2017): Tnm. Klassifikation Maligner Tumoren. 8th ed. Weinheim: John Wiley & Sons Incorporated. Available online at https://livivo.idm.oclc.org/login?url=https://ebookcentral.proquest.com/lib/zbmed-ebooks/detail.action?docID=4770952.

Woolbright, Benjamin L.; Rajendran, Ganeshkumar; Harris, Robert A.; Taylor, John A. (2019): Metabolic Flexibility in Cancer: Targeting the Pyruvate Dehydrogenase Kinase:Pyruvate Dehydrogenase Axis. In *Molecular cancer therapeutics* 18 (10), pp. 1673–1681. DOI: 10.1158/1535-7163.MCT-19-0079.

Wu, Ronghu; Haas, Wilhelm; Dephoure, Noah; Huttlin, Edward L.; Zhai, Bo; Sowa, Mathew E.; Gygi, Steven P. (2011): A large-scale method to measure absolute protein phosphorylation stoichiometries. In *Nature methods* 8 (8), pp. 677–683. DOI: 10.1038/nmeth.1636.

Xia, Yifeng; Shen, Shen; Verma, Inder M. (2014): NF-κB, an active player in human cancers. In *Cancer immunology research* 2 (9), pp. 823–830. DOI: 10.1158/2326-6066.CIR-14-0112.

Xia, Yifeng; Yeddula, Narayana; Leblanc, Mathias; Ke, Eugene; Zhang, Yonghui; Oldfield, Eric et al. (2012): Reduced cell proliferation by IKK2 depletion in a mouse lung-cancer model. In *Nature cell biology* 14 (3), pp. 257–265. DOI: 10.1038/ncb2428.

Xie, Yi; Ostriker, Allison C.; Jin, Yu; Hu, Haidi; Sizer, Ashley J.; Peng, Gang et al. (2019): LMO7 Is a Negative Feedback Regulator of Transforming Growth Factor β Signaling and Fibrosis. In *Circulation* 139 (5), pp. 679–693. DOI: 10.1161/CIRCULATIONAHA.118.034615.

Xu, Yan; Fulciniti, Mariateresa; Samur, Mehmet K.; Ho, Matthew; Deng, Shuhui; Liu, Lanting et al. (2020): YWHAE/14-3-3ε expression impacts the protein load, contributing to proteasome inhibitor sensitivity in multiple myeloma. In *Blood* 136 (4), pp. 468–479. DOI: 10.1182/blood.2019004147.

Yamaguchi, K.; Tsuchihashi, K.; Ueno, S.; Uehara, K.; Taguchi, R.; Ito, M. et al. (2025): Efficacy of pembrolizumab in microsatellite-stable, tumor mutational burden-high metastatic colorectal cancer: genomic signatures and clinical outcomes. In *ESMO open* 10 (1), p. 104108. DOI: 10.1016/j.esmoop.2024.104108.

Yaron-Barir, Tomer M.; Joughin, Brian A.; Huntsman, Emily M.; Kerelsky, Alexander; Cizin, Daniel M.; Cohen, Benjamin M. et al. (2024): The intrinsic substrate specificity of the human tyrosine kinome. In *Nature* 629 (8014), pp. 1174–1181. DOI: 10.1038/s41586-024-07407-y.

Yip, Hon Yan Kelvin; Shin, Sung-Young; Chee, Annabel; Ang, Ching-Seng; Rossello, Fernando J.; Wong, Lee Hwa et al. (2024): Integrative modeling uncovers p21-driven drug resistance and prioritizes therapies for PIK3CA-mutant breast cancer. In *NPJ precision oncology* 8 (1), p. 20. DOI: 10.1038/s41698-024-00496-y.

Yoshihara, Kosuke; Shahmoradgoli, Maria; Martínez, Emmanuel; Vegesna, Rahulsimham; Kim, Hoon; Torres-Garcia, Wandaliz et al. (2013): Inferring tumour purity and stromal and immune cell admixture from expression data. In *Nature communications* 4, p. 2612. DOI: 10.1038/ncomms3612.

Yu, Chune; Luo, Dan; Yu, Jing; Zhang, Min; Zheng, Xiaobo; Xu, Guangchao et al. (2022): Genome-wide CRISPR-cas9 knockout screening identifies GRB7 as a driver for MEK inhibitor resistance in KRAS mutant colon cancer. In *Oncogene* 41 (2), pp. 191–203. DOI: 10.1038/s41388-021-02077-w.

Yu, Hang; Liu, Qingquan; Wu, Keting; Tang, Shuang (2024): Biomarkers to predict efficacy of immune checkpoint inhibitors in colorectal cancer patients: a systematic review and meta-analysis. In *Clinical and Experimental Medicine* 24 (1), p. 143. DOI: 10.1007/s10238-024-01408-x.

Yu, Jiali; Green, Michael D.; Li, Shasha; Sun, Yilun; Journey, Sara N.; Choi, Jae Eun et al. (2021): Liver metastasis restrains immunotherapy efficacy via macrophage-mediated T cell elimination. In *Nature medicine* 27 (1), pp. 152–164. DOI: 10.1038/s41591-020-1131-x.

Zagorac, Ivana; Fernandez-Gaitero, Sara; Penning, Renske; Post, Harm; Bueno, Maria J.; Mouron, Silvana et al. (2018): In vivo phosphoproteomics reveals kinase activity profiles that predict treatment outcome in triple-negative breast cancer. In *Nature communications* 9 (1), p. 3501. DOI: 10.1038/s41467-018-05742-z.

Zeng, Xiao; Ma, Qiong; Li, Xue-Ke; You, Li-Ting; Li, Jia; Fu, Xi et al. (2023): Patient-derived organoids of lung cancer based on organoids-on-a-chip: enhancing clinical and translational applications. In *Frontiers in bioengineering and biotechnology* 11, p. 1205157. DOI: 10.3389/fbioe.2023.1205157.

Zhang, Chundong; Mei, Zubing; Pei, Junpeng; Abe, Masanobu; Zeng, Xiantao; Huang, Qiao et al. (2021): A Modified Tumor-Node-Metastasis Classification for Primary Operable Colorectal Cancer. In *JNCI cancer spectrum* 5 (1). DOI: 10.1093/jncics/pkaa093.

Zhang, Jiaxin; Liu, Zuojia; Zhao, Wenjing; Li, Chang; Liu, Fei; Wang, Jin (2025): Targeting KRAS Sensitizes Ferroptosis by Coordinately Regulating the TCA Cycle and Nrf2-SLC7A11-GPX4 Signaling in Hepatocellular Carcinoma. In *Smart medicine* 4 (2), e70005. DOI: 10.1002/smmd.70005.

Zhang, Jiwei; Ma, Fen; Li, Zhe; Li, Yuan; Sun, Xun; Song, Mingxu et al. (2024): NFKB2 mediates colorectal cancer cell immune escape and metastasis in a STAT2/PD-L1-dependent manner. In *MedComm* 5 (5), e521. DOI: 10.1002/mco2.521.

Zhou, Yangying; Lih, T. Mamie; Pan, Jianbo; Höti, Naseruddin; Dong, Mingming; Cao, Liwei et al. (2020): Proteomic signatures of 16 major types of human cancer reveal universal and cancer-type-specific proteins for the identification of potential therapeutic targets. In *Journal of hematology & oncology* 13 (1), p. 170. DOI: 10.1186/s13045-020-01013-x.

Zhou, Zhongliang; Yeung, Wayland; Gravel, Nathan; Salcedo, Mariah; Soleymani, Saber; Li, Sheng; Kannan, Natarajan (2023): Phosformer: an explainable transformer model for protein kinase-specific phosphorylation predictions. In *Bioinformatics* 39 (2). DOI: 10.1093/bioinformatics/btad046.

Zhu, Chunxiao; Guan, Xiaoqing; Zhang, Xinuo; Luan, Xin; Song, Zhengbo; Cheng, Xiangdong et al. (2022): Targeting KRAS mutant cancers: from druggable therapy to drug resistance. In *Molecular cancer* 21 (1), p. 159. DOI: 10.1186/s12943-022-01629-2.