

Aus dem Zentrum für Pathologie der Universität zu Köln
Institut für Allgemeine Pathologie und Pathologische Anatomie
Direktor: Universitätsprofessor Dr. med. R. Büttner

Über die Beschreibung und Modellierung einer Abhängigkeit des Zellwachstums von einer KRAS- Expression in KRAS Wildtyp Tumorzelllinien

Inaugural-Dissertation zur Erlangung der Doktorwürde
der Medizinischen Fakultät
der Universität zu Köln

vorgelegt von
Bastian Ulmer
aus Köln

promoviert am 06. April 2023

Gedruckt mit Genehmigung der Medizinischen Fakultät der Universität zu Köln

2023

Dekan: Universitätsprofessor Dr. med. G. R. Fink

1. Gutachterin: Professorin Dr. rer. nat. M. Odenthal
2. Gutachter: Privatdozent Dr. med. P. J. Bröckelmann
3. Gutachter: Privatdozent Dr. med. U. Schnetzke

Erklärung

Ich erkläre hiermit, dass ich die vorliegende Dissertationsschrift ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht.

Bei der Auswahl und Auswertung des Materials sowie bei der Herstellung des Manuskriptes habe ich Unterstützungsleistungen von folgenden Personen:

Herr Dr. Michael Kloth

Weitere Personen waren an der Erstellung der vorliegenden Arbeit nicht beteiligt. Insbesondere habe ich nicht die Hilfe einer Promotionsberaterin/eines Promotionsberaters in Anspruch genommen. Dritte haben von mir weder unmittelbar noch mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertationsschrift stehen.

Die Dissertationsschrift wurde von mir bisher weder im Inland noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

Die Forschungsarbeit habe ich mit beratender Unterstützung von Herrn Dr. Michael Kloth selbst konzipiert. Wie in der Publikation aufgeführt wurden die Daten der Krebszelllinien von der Internetseite des Dependency Map Consortium (Mutationen, RNA-Expression, CRISPR, RNAi, CCLE-Chemosensitivitätsdaten) heruntergeladen, sowie zusätzliche Chemosensitivitätsdaten von der Internetseite des Genomics of Drug Sensitivity in Cancer (GDSC) Projekts. Sowohl Datenanalysen als auch Implementierung der Modelle wurden von mir eigenständig in der R Programmierumgebung durchgeführt. Die Ergebnisse der Analysen wurden mit Herrn Dr. Michael Kloth, Frau Prof. Dr. Odenthal, Herrn Prof. Dr. Büttner und Herrn Prof. Dr. Roth diskutiert. Das Manuskript habe ich selbständig erstellt. Herr Dr. Michael Kloth und Frau Prof. Dr. Odenthal haben das Manuskript Korrektur gelesen.

Erklärung zur guten wissenschaftlichen Praxis:

Ich erkläre hiermit, dass ich die Ordnung zur Sicherung guter wissenschaftlicher Praxis und zum Umgang mit wissenschaftlichem Fehlverhalten (Amtliche Mitteilung der Universität zu Köln AM 132/2020) der Universität zu Köln gelesen habe und verpflichte mich hiermit, die dort genannten Vorgaben bei allen wissenschaftlichen Tätigkeiten zu beachten und umzusetzen.

Köln, den 09.10.2022

Unterschrift: 

Danksagung

An dieser Stelle möchte ich allen beteiligten Personen danken, die mich bei der Anfertigung meiner Dissertation unterstützt haben. Mein besonderer Dank gilt Herrn Dr. Michael Kloth für seine enorme Unterstützung bei der Umsetzung der gesamten Arbeit und die zahlreichen produktiven Gespräche. Außerdem möchte ich mich bei Frau Prof. Dr. Margarete Odenthal für ihren Rat und ihre Geduld im gesamten Verlauf meiner Promotion bedanken. Des Weiteren danke ich meiner Verlobten Frau Dr. Farina Grothey und meinen Eltern für ihr Verständnis, ihren Zuspruch und ihre Hilfe.

Meinem Großvater

Inhaltsverzeichnis

ABKÜRZUNGSVERZEICHNIS	6
1. ZUSAMMENFASSUNG	8
2. EINLEITUNG	9
2.1. RAS-Physiologie	9
2.2. RAS-Pathophysiologie	12
2.3. Zielgerichtete Therapien im RAS-Signalweg	14
2.4. Maschinelles Lernen	17
2.5. Fragestellungen und Ziel der Arbeit	21
3. PUBLIKATION	22
4. DISKUSSION	34
4.1. KRAS-Abhängigkeit in KRAS ^{wt} Zelllinien	34
4.2. Prädiktion der KRAS-Abhängigkeit	35
4.3. Diffusionskernel-basierte Prädiktion der KRAS-Abhängigkeit	37
5. LITERATURVERZEICHNIS	40
6. ANHANG	46
6.1. Abbildungsverzeichnis	46
7. VORABVERÖFFENTLICHUNGEN VON ERGEBNISSEN	47

Abkürzungsverzeichnis

AKT	AKT Serine/Threonine Kinase
ALK	Anaplastic Lymphoma Receptor Tyrosine Kinase
APT	Lysophospholipase
BAD	BCL2 Associated Agonist of Cell Death
BCL-X	B-cell lymphoma 2 Like 1
BRAF	B-Raf Proto-Oncogene, Serine/Threonine Kinase
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
DHHC9	Zinc Finger DHHC-Type Palmitoyltransferase 9
EGF	Epidermal Growth Factor
EGFR	Epidermal Growth Factor Receptor
ELK1	ETS transcription factor ELK1
EML4	Echinoderm Microtubule Associated Protein Like 4
ERBB	Human Epidermal Growth Factor Receptor
ERK	Extracellular signal-regulated kinase
ETS	E26 transformation-specific
FGFR	Fibroblast Growth Factor Receptor
FOS	FBJ Murine Osteosarcoma Viral Oncogene Homolog
GAP	GTPase-activating proteins
GCP16	Golgin A7
GDP	Guanosindiphosphat
GEF	Ras guanosine nucleotide exchange factors
GPCR	G-Protein-gekoppelte Rezeptoren
GRB2	Growth factor receptor-bound protein 2
GTP	Guanosintriphosphat
HDAC6	Histone Deacetylase 6
HRAS	Harvey Rat Sarcoma Viral Oncogene Homolog
JUN	Jun Proto-Oncogene, AP-1 Transcription Factor Subunit
KRAS	Kirsten Rat Sarcoma Viral Oncogene Homolog
LZTR1	Leucine Zipper Like Transcription Regulator 1
MAP2K1	Mitogen-Activated Protein Kinase Kinase 1
MAP2K2	Mitogen-Activated Protein Kinase Kinase 2
MAPK	Mitogen-activated protein Kinase
MEK	Mitogen-activated protein kinase kinase
MET	Hepatocyte Growth Factor Receptor
mTOR	Mechanistic Target of Rapamycin Kinase
MYC	Avian Myelocytomatosis Viral Oncogene Homolog

NF1	Neurofibromin 1
NFkB	Nuclear Factor Kappa B Subunit
NPM	Nucleophosmin
NRAS	Neuroblastoma Rat Sarcoma Viral Oncogene Homolog
NSCLC	Non-small cell lung cancer
PAK	P21 RAC1 Activated Kinase 1
PDGFR	Platelet-derived growth factor receptor
PI3K	Phosphoinositid-3-Kinase
PTB	Phosphotyrosin-bindende Domäne
RAE	RAS-activating Event
RAC	Ras-Related C3 Botulinum Toxin Substrate 1
RAF	Rapidly fibrosarcoma Proto-Oncogene Serine/Threonine-Protein Kinase
RAS	Rat Sarcoma
RasGRF	Ras guanosine nucleotide releasing factor
RasGRP	Ras guanosine nucleotide releasing protein
RET	Proto-Oncogene Tyrosine-Protein Kinase Receptor Ret
Rho	Rhodopsin
ROS1	ROS Proto-Oncogene 1, Receptor Tyrosine Kinase
RTK	Rezeptortyrosinkinase
RNAi	RNA-Interference
SH2	Src-homology 2
SHP2	SH2 Domain-Containing Protein Tyrosine Phosphatase 2
SIRT2	Sirtuin 2
SNV	Single-Nucleotide-Variants
SOS	Son-of-sevenless
STK19	Serine/Threonine Kinase 19
TIAM1	T-Lymphoma Invasion and Metastasis-Inducing Protein 1
TRK	Tropomyosin receptor kinase

1. Zusammenfassung

Mit der Entwicklung von mutationsspezifischen *Kirsten Rat Sarcoma Viral Oncogene Homolog* (KRAS) und panKRAS Inhibitoren konnten in der zielgerichteten Pharmakotherapie gegen die häufig mutierten *Rat Sarcoma* (RAS)-Onkogene zuletzt bedeutende Erfolge erreicht werden. Durch ihre zentrale Rolle innerhalb des Mitogen-aktivierten Signalwegs besitzen die RAS-GTPasen jedoch auch ohne Vorliegen von aktivierenden Mutationen eine wichtige Rolle in der Signaltransduktion von membranständigen Kinasen. Ziel dieser Arbeit war es daher, die Abhängigkeit des Zellwachstums von einer KRAS-Expression, auch bezeichnet als KRAS-Abhängigkeit, in Tumorzelllinien ohne aktivierende KRAS-Mutation zu charakterisieren. Zusätzlich sollten verschiedene Optionen für eine Modellierung dieser Abhängigkeit evaluiert werden.

Dafür wurden Daten von über 700 Zelllinien aus *Clustered Regularly Interspaced Short Palindromic Repeat* (CRISPR)-*knockout* und *RNA-Interference* (RNAi)-*knockdown* Experimenten des Achilles- und DRIVE-Projekts analysiert. Dabei konnte eine Subgruppe von Zelllinien ohne aktivierende KRAS-Mutationen identifiziert werden, deren Wachstum abhängig von einer KRAS-Expression ist. Durch den Nachweis einer erhöhten Chemosensitivität gegenüber zahlreichen *Mitogen-activated protein kinase kinase* (MEK)-Inhibitoren konnte zusätzlich Evidenz für die klinische Relevanz der Subgruppe gefunden werden.

Anschließend wurde die Abhängigkeit des Wachstums der Zelllinien von einer KRAS-Expression mit einem kombinierten Vorgehen unter Verwendung von RNA-Expressionsdaten und einer Diffusionskernel-basierten Variablenselektion erfolgreich vorhergesagt. Im Gegensatz zu diesem Ansatz konnten Modelle basierend auf Prädiktoren, bestehend aus ausgewählten *RAS-activating* Events (RAE) oder zuvor veröffentlichten RAS-Expressionssignaturen keine zufriedenstellenden Vorhersagen liefern. Diese Beobachtungen unterstreichen die Bedeutung eines standardisierten methodischen Vorgehens in der Modellierung mit Expressionsdaten und heben die heterogene Verteilung der RAEs in den Zelllinien mit KRAS Wildtyp Sequenz (KRAS^{wt}) hervor. Des Weiteren konnte gezeigt werden, dass Prädiktoren der KRAS^{wt} Modelle Informationssignale enthalten, die nicht durch andere Genexpressionen ersetzbar sind. Dies erhöht die Evidenz für einen Phänotyp in der KRAS^{wt} Subgruppe, dessen Wachstum abhängig von einer KRAS-Expression ist.

Zusammengefasst deuten die Ergebnisse der vorgelegten Studie daraufhin, dass Patienten mit Tumoren ohne aktivierende KRAS-Mutationen von einer Therapie mit den neu entwickelten panKRAS profitieren könnten. *Machine Learning*-Modelle basierend auf RNA-Expressionsdaten könnten zukünftig dabei helfen, klinische Entscheidungen bezüglich einer Therapieindikation zu treffen.

2. Einleitung

Der RAS Signalweg nimmt eine zentrale Stellung im Netzwerk verschiedener Signalkaskaden ein, die Wachstums- und Differenzierungsprozesse steuern¹⁻⁷. Aufgrund dieser zentralen Rolle ist der Signalweg einer der bedeutendsten onkogenen Treiber in der Entstehung zahlreicher Tumorentitäten⁸⁻¹¹. In der Vergangenheit wurden verschiedene Mechanismen beschrieben, die zu einer konstitutiven Aktivierung der Signaltransduktion führen. Insbesondere wurden für die RAS-GTPasen (KRAS, *Neuroblastoma Rat Sarcoma Viral Oncogene Homolog* (NRAS), *Harvey Rat Sarcoma Viral Oncogene Homolog* (HRAS)) selbst wichtige RAEs identifiziert¹². Durch die Entwicklung von panKRAS- und mutationsspezifischen Inhibitoren konnten hier zuletzt bedeutende Fortschritte in der zielgerichteten Therapie erreicht werden¹³⁻¹⁶.

2.1. RAS-Physiologie

Die Aktivierung des RAS-Signalwegs erfolgt an den RAS GTPasen KRAS, NRAS und HRAS durch den Austausch von Guanosindiphosphat (GDP) zu Guanosintriphosphat (GTP)^{12,17}. Wie von Gimple et al. zusammengefasst wurde, wird dieser Austausch durch sogenannte *Ras guanosine nucleotide exchange factors* (GEF) beschleunigt, wodurch die Signalweiterleitung verstärkt wird. Die entgegengesetzte Reaktion wird wiederum durch *GTPase-activating proteins* (GAPs) katalysiert, welche die Hydrolyse von GTP zu GDP beschleunigen, was zu einer Inaktivierung des Proteins führt¹². GEFs werden in die drei Gruppen, *Son-of-sevenless* (SOS), *Ras guanosine nucleotide releasing factor* (RasGRF) und *Ras guanosine nucleotide releasing protein* (RasGRP) unterteilt¹⁸. Bei Bindung von GEFs an die RAS-GTPasen dissoziiert gebundenes GDP. Anschließend wird dafür bevorzugt GTP gebunden, da es in zehnfach höherer Konzentration im Zytoplasma vorliegt¹⁸. Die Rekrutierung von GEFs kann durch eine Vielzahl unterschiedlicher membranständiger Rezeptoren initiiert werden. Dazu gehören Rezeptortyrosinkinasen (RTK), G-Protein-gekoppelte Rezeptoren (GPCR) und Integrine¹². Klinisch sind momentan insbesondere die RTKs von besonderem Interesse. Bislang wurden im menschlichen Genom 58 verschiedene RTKs identifiziert, die in 20 Klassen eingeteilt werden¹⁹. RTKs besitzen eine extrazelluläre Domäne für die Bindung von Liganden, eine transmembrane Helix sowie eine intrazelluläre Kinase-Domäne für die Signalweiterleitung²⁰. Zumeist liegen sie als monomere, seltener auch als multimere Komplexe in der Zellmembran vor. Im Fall der monomeren RTKs führt im Regelfall die Bindung von Wachstumsfaktoren auf der extrazellulären Seite zu einer Dimerisierung zweier Monomere mit anschließender Autophosphorylierung an zytoplasmatischen Tyrosinresten. Verschiedene Docking Proteine mit *Src-homology 2* (SH2)-Domänen und Phosphotyrosin-bindenden-Domänen (PTB) binden nun an die Phosphotyrosinreste der RTKs²⁰. Für die

Signalweiterleitung im RAS Signalweg ist hierbei insbesondere das *Growth factor receptor-bound protein 2* (GRB2) als Adapterprotein von Bedeutung^{21,22}. GRB2 bindet mit seiner SH2-Domäne an die phosphorylierten Tyrosinreste der aktivierten RTKs und bildet gleichzeitig über seine SH3-Domäne einen Komplex mit dem *guanine nucleotide exchange factor* SOS1. Die Rekrutierung des GRB2-SOS1 Komplexes zur Zellmembran führt nun zu einer Beschleunigung der durch SOS1 katalysierten Dissoziation von GDP an den RAS-GTPasen und somit zu einer gesteigerten Aktivität des Signalwegs^{21,22}.

Die Aktivierung der RAS-Proteine hat nun wiederum Einfluss auf diverse *Downstream* Signalkaskaden mit multiplen Zielgenen, die letztlich Prozesse wie Transkription, Zellzyklus, Zellwachstum und Zellüberleben steuern^{23,24}. Initiale Forschungsansätze fokussierten sich zunächst auf die Signalweiterleitung über den *Mitogen-activated protein Kinase* (MAPK)- sowie den *Phosphoinositid-3-Kinase* (PI3K)-Signalweg, welche intensiv untersucht wurden. Wie von Guo et al. zusammengefasst, erfolgt im MAPK-Signalweg durch RAS eine sequenzielle Signalweiterleitung über die folgenden Kinasen: *Rapidly fibrosarcoma Proto-Oncogene Serine/Threonine-Protein Kinase* (RAF), MEK und *Extracellular signal-regulated kinase* (ERK). Auf der Endstrecke aktiviert ERK Transkriptionsfaktoren wie ETS transcription factor ELK1 (ELK1) und *Avian Myelocytomatosis Viral Oncogene Homolog* (MYC), welche die Genexpression regulieren²⁵. Der PI3K-Aktivierung durch RAS folgt ebenfalls eine Signalweiterleitung über mehrere Stufen (bspw. PI3K – *AKT Serine/Threonine Kinase* (AKT) – *Nuclear Factor Kappa B Subunit* (NFkB) - etc.), wobei unter anderem Zellwachstum und anti-apoptotische Effekte positiv reguliert werden²⁶. Wie Abbildung 1 andeutet, ist die Signalweiterleitung jedoch äußerst komplex. Es kommen multiple zusätzliche Einflussfaktoren zusammen, sodass die aufgeführten linearen Prozesse nur eine starke Vereinfachung darstellen.

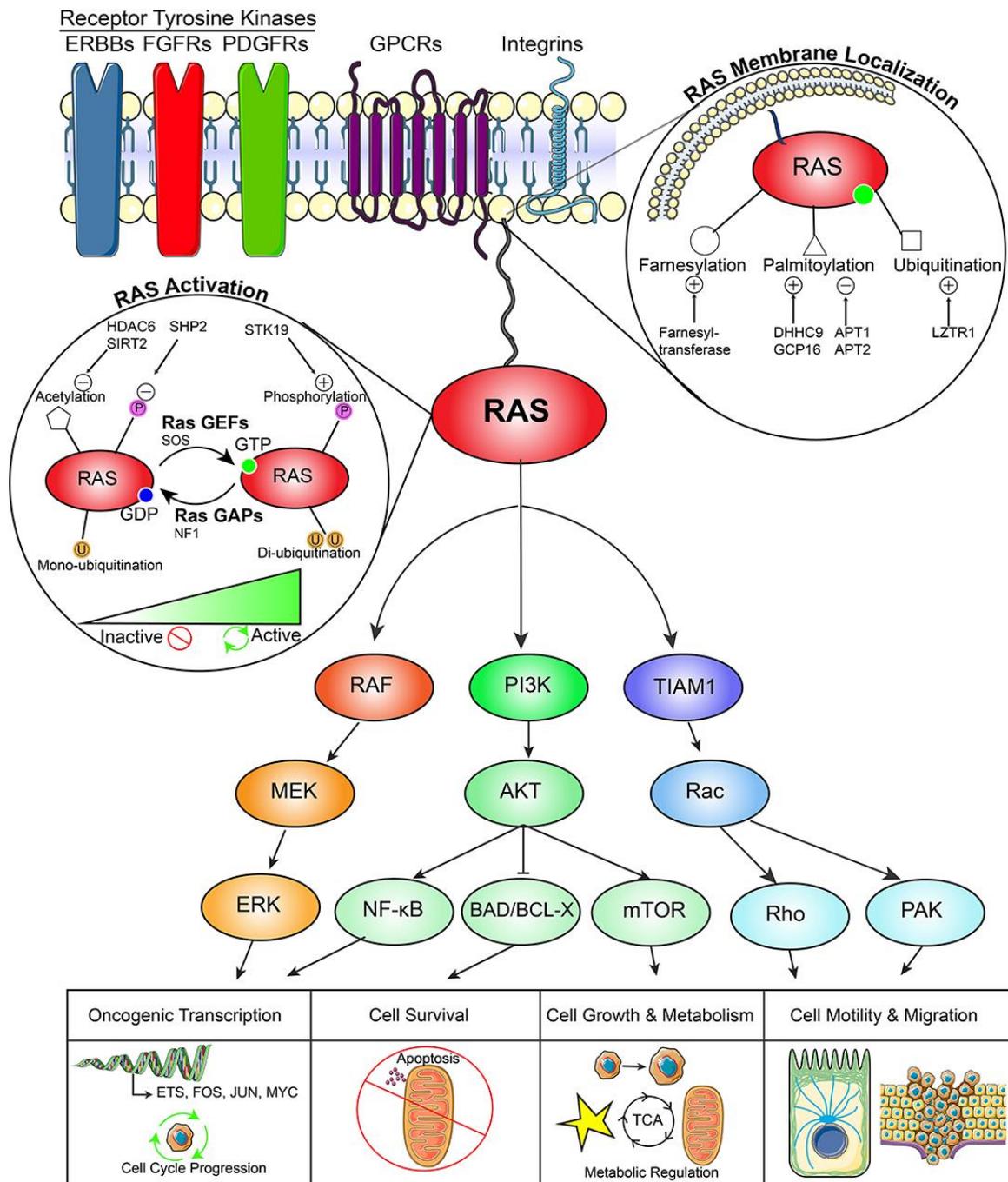


Abbildung 1. Vereinfachte schematische Darstellung des RAS-Signalwegs mit Fokus auf onkogene Effekte aus Gimble et al. (2019) ¹². Die Aktivierung des Signalwegs erfolgt *upstream* der RAS-Proteine an membranständigen Proteinen (z.B. RTKs). Anschließend werden die RAS-Proteine über mehrere Zwischenschritte aktiviert und das Signal von dort aus über verschiedene Effektorproteine (RAF, PI3K, etc.) weitergeleitet. *Downstream* nehmen diese Proteine Einfluss auf Prozesse die Zellwachstum, Differenzierung, Überleben und Migration beeinflussen.

2.2. RAS-Pathophysiologie

Durch die vielfältigen Einflüsse auf Prozesse, die Zellwachstum und Zellüberleben fördern, ist der RAS-Signalweg von hoher Bedeutung in der Onkogenese zahlreicher Tumorentitäten¹⁻¹¹. Wie von Santarpia et al. und Du et al. zusammengefasst, wurden diverse RAEs beschrieben, die auf verschiedenen Ebenen der Signalkaskade zu einer konstitutiven Aktivierung führen^{27,28}. Für die einzelnen Gene des Signalwegs sind teilweise auch mehrere unterschiedliche Aktivierungsmechanismen beschrieben worden. Zumeist lassen sich die Ursachen einer pathologisch gesteigerten Signaltransduktion auf strukturelle Veränderungen oder Überexpression der Proteine zurückführen^{27,28}. Wie zahlreiche große Studien belegen, kann die Inzidenz der RAEs dabei stark zwischen den Tumorentitäten variieren^{9,10}.

Zu den häufig von RAEs betroffenen Genen zählen RTKs der äußeren Zellmembran. RAEs, die zu strukturellen Veränderungen in den RTKs führen, werden oft durch einzelne Basenaustausche (*Single-Nucleotide-Variants* (SNV)), kurze Insertionen und Deletionen oder Translokationen hervorgerufen^{9,10,29}. Im Fall der SNV ist meist die Kinasedomäne direkt betroffen, wohingegen die Bruchpunkte der Translokationen C- oder N-Terminal der Kinase zu finden sind²⁸. Pathologische strukturelle Veränderungen mit unterschiedlicher klinischer Signifikanz lassen sich in zahlreichen RTKs finden. Im *Epidermal Growth Factor Receptor* (EGFR) wurden SNV und Deletionen aufgrund therapeutischer Implikationen ausführlich charakterisiert. Aktivierende Mutationen betreffen zumeist die Exone 18 bis 21, welche für die Kinasedomäne kodieren. Deletionen treten dabei mit hoher Frequenz in Exon 19 auf, wohingegen in Exon 21 durch Missense-Mutation häufig zur L858R Variante führen. Insgesamt machen beide Alterationen ungefähr 90% der Mutationen in der Kinase von EGFR aus²⁸. Zusätzlich ist die Variante T790M im Exon 20 als Resistenzmechanismus von hoher klinischer Bedeutung für eine Therapie mit EGFR-Inhibitoren (siehe unten)³⁰. Rekurrente Translokationen mit RAS-aktivierender Wirkung lassen sich wiederum beispielsweise in folgenden RTKs finden: *Anaplastic Lymphoma Receptor Tyrosine Kinase* (ALK), *Proto-Oncogene Tyrosine-Protein Kinase Receptor Ret* (RET) und *ROS Proto-Oncogene 1 Receptor Tyrosine Kinase* (ROS1)²⁸. Die Fusionspartner sind dabei oft unterschiedlich und variieren auch für das jeweilige Gen. Gemeinsames Merkmal ist jedoch, dass die Kinasedomäne bei gleichzeitiger Steigerung ihrer Aktivität stets erhalten bleibt. Wichtige Partnergene für ALK sind beispielsweise *Nucleophosmin* (NPM) und *Echinoderm Microtubule Associated Protein Like 4* (EML4)^{28,31}. Ursächlich für die onkogenen Effekte der beiden Translokationen sind zwei Faktoren. Die Transkription des Fusionsgens wird durch den Promotor des Fusionspartners von ALK gesteuert, was zu einer Überexpression der Kinase führt. Gleichzeitig vermittelt der Fusionspartner eine konstitutive Dimerisierung von zwei Monomeren des Fusionsproteins, die unabhängig von der Bindung eines Liganden ist und somit zur Aktivierung von ALK führt³¹. Die Proteinüberexpression ist der zweite übergeordnete Mechanismus, der zu einer

pathologischen RAS-Aktivierung durch RTKs führt. Ursächlich sind neben den bereits erwähnten Translokationen vor allem Genamplifikationen wie sie in ERBB2, im *Hepatocyte Growth Factor Receptor* (MET) und zahlreichen weiteren Genen auftreten³²⁻³⁵. Die Steigerung der Genkopienanzahl führt dabei zu einer erhöhten Expression der Onkogene, wodurch das Protein akkumuliert und die Signaltransduktion gesteigert wird²⁸.

Unterhalb einer Vielzahl von membranständigen Kinasen werden die RAS-GTPasen als zentrale Knotenpunkte der Signaltransduktion aktiviert. Aufgrund dieser zentralen Stellung in Signalwegen von Rezeptorkinasen gelten sie deshalb als starke onkogene Treiber. Wie zuvor können sowohl strukturelle Veränderungen des Proteins als auch Überexpression zu einer konstitutiven Aktivierung führen. Die strukturellen Veränderungen in Form von Missense-Mutationen überwiegen jedoch deutlich in ihrer Inzidenz¹¹. In Abhängigkeit von den Isoformen der GTPasen variiert ihre Häufigkeit je nach Entität. So ist KRAS häufig in Pankreaskarzinomen mutiert, NRAS in malignen Melanomen und HRAS in Urothelkarzinomen¹¹. Die Mutationen treten meist als Hotspots in den Kodons 12, 13, 61 und 146 auf, welche im Protein für die Nukleotidbindung verantwortlich sind (Abbildung 2)³⁶. Auch hier konnte eine Spezifität der Isoformen für bestimmte Kodons beobachtet werden. 80% der KRAS-Mutationen treten in Kodon 12 auf, wohingegen bei NRAS mit 60% Kodon 61 am häufigsten betroffen ist. Bei HRAS sind die häufigsten Mutationen wiederum gleichmäßiger verteilt. Mechanistisch führen die mutationsbedingten strukturellen Veränderungen zu einem Verlust der GTPase-Aktivität und zu einer erschwerten Bindung von GAPs. Die Hydrolyse von GTP zu GDP wird dadurch eingeschränkt und die RAS Proteine verbleiben in ihrer aktiven Form³⁶. Die Überexpression von RAS Proteinen ist im Vergleich dazu ein deutlich selteneres RAE¹¹. Für KRAS wurde sie in mehreren Entitäten wie Ösophagus-, Magen- und Ovarialkarzinomen beschrieben, wo sie den Hauptanteil der pathologischen Veränderungen in RAS ausmachen. Ursächlich sind auch hier Amplifikationen, die eine gesteigerte Anzahl von aktivierten RAS Proteinen mit gebundenem GTP zur Folge haben und somit auch die Aktivität des Signalwegs erhöhen¹¹.

Diese Abbildung kann aufgrund des Urheberrechts
nicht gezeigt werden.

Abbildung 2. Vergleichende Gegenüberstellung der Proteinsequenzen der RAS-Isoformen mit zugehörigen Sekundärstrukturelementen aus Prior et al. (2012) ³⁶. Sequenzabschnitte, die zuständig für Bindungsprozesse sind wurden farblich hervorgehoben (siehe Legende). Des Weiteren unterscheiden sich grau markierte Aminosäuren zwischen den Isoformen. Für rot markierte sind Mutationen in Tumoren bekannt, für schwarz unterstrichene in Entwicklungsstörungen. In malignen Tumoren lässt sich der Großteil der Mutationen in den Kodons 12, 13 und 61 finden ³⁶⁻³⁹.

Da die RAS-Proteine zahlreiche Effektoren besitzen, verzweigt sich der Signalweg im weiteren Verlauf, wodurch die Anzahl und die Diversität der RAEs zunimmt. In der klassischen RAS-RAF-MAPK-ERK Kaskade wurden für die Elemente zahlreiche weitere pathologische Ereignisse mit onkogener Wirkung identifiziert ⁴⁰⁻⁴³. Klinisch höchst relevant sind RAEs in der *B-Raf Proto-Oncogene Serine/Threonine Kinase* (BRAF). Zumeist führen Missense-Mutationen im Kodon V600 mit der Variante V600E zu einer Steigerung der Kinaseaktivität und damit zu einer konstitutiven Aktivierung der *Downstream* Effektoren *Mitogen-Activated Protein Kinase Kinase 1/2* (MAP2K1/MAP2K2) ⁴³. Darüber hinaus sind Mutationen in weiteren Aminosäure-Positionen beschrieben, die jedoch deutlich seltener auftreten ^{44,45}.

2.3. Zielgerichtete Therapien im RAS-Signalweg

In den vergangenen Jahren wurden erhebliche Anstrengungen unternommen, um zielgerichtete Therapien gegen onkogene Treiber des RAS-Signalwegs zu entwickeln. Es wurden sowohl Antikörper- als auch *Small Molecule*-basierte Therapeutika entwickelt, welche die Signaltransduktion an verschiedensten Zielstrukturen hemmen.

Fortschritte konnten insbesondere durch Therapieansätze erzielt werden, welche gegen die membranständigen RTKs gerichtet sind. Zielstrukturen aus dieser Gruppe sind unter anderem

die unter 1.2 beschriebenen *Epidermal Growth Factor* (EGF)-Rezeptoren, ALK, MET und ROS1⁴⁶⁻⁵⁰. In der Therapie RAS-aktivierter Tumore mit zugrundeliegenden pathologischen Veränderungen in EGFR wurden spezifische Antikörper wie Cetuximab entwickelt, welche an die extrazelluläre Domäne des Rezeptors binden⁵¹. Dadurch wird dieser für den Liganden EGF blockiert, was zu einer Reduktion des Tumorwachstums führt. Eingesetzt wird Cetuximab in der Therapie des metastasierten kolorektalen Karzinoms und in Plattenepithelkarzinomen des Kopf-Hals Bereichs⁵¹. Entscheidend für eine erfolgreiche Behandlung ist der RAS-Mutationsstatus^{52,53}. Oftmals bestehen RAS-Mutationen bereits vor einer anti-EGFR Therapie als unabhängiges RAEs, wodurch die Tumore eine intrinsische Resistenz aufweisen. Die aktivierenden RAS-Mutationen führen zu einer konstitutiv gesteigerten Signalweiterleitung *Downstream* des Rezeptors und verhindern somit eine antineoplastische Wirkung von Cetuximab^{52,53}. Teilweise treten sie allerdings auch als Resistenzmutation im Verlauf einer Behandlung auf⁵⁴. Des Weiteren wurden neben den Antikörpern verschiedene gegen EGFR gerichtete *Small Molecules* entwickelt, welche die Aktivität der Tyrosinkinase-Domäne hemmen (sog. Tyrosinkinase-Inhibitoren (TKI))⁵¹. Angewendet werden sie beispielsweise in der Therapie des nicht-kleinzelligen Lungenkarzinoms (NSCLC) bei Vorliegen von aktivierenden EGFR-Mutationen⁵⁵. Um Resistenzen zu umgehen, wurden die EGFR-TKI im Laufe der Jahre kontinuierlich weiterentwickelt. Daraus gingen Inhibitoren mit unterschiedlichen Wirkmechanismen und Therapieindikationen hervor, welche anhand ihrer Eigenschaften in drei Generationen unterteilt werden⁵⁶⁻⁶². Die erste Generation (Gefitinib, Erlotinib) wirkt durch eine reversible Blockierung der ATP-binding side von EGFR. Die zweite Generation (Afatinib, Dacomitinib) blockiert die Kinasedomäne irreversibel und verhindert dadurch die Transphosphorylierung. Die dritte Generation (Osimertinib) wird bei Vorliegen der Resistenzmutation T790M eingesetzt⁵⁶⁻⁶². Diese tritt häufig im Verlauf einer Therapie mit Inhibitoren der ersten und zweiten Generation auf und führt zu einer verminderten oder aufgehobenen Wirksamkeit der Medikamente^{63,64}. Inhibitoren der dritten Generation binden spezifisch an EGF-Rezeptoren mit der T790M Variante und ermöglichen dadurch eine Fortführung der Therapie⁶⁵. Gleichzeitig werden Nebenwirkungen durch die Spezifität für die T790M Variante verringert, da Wildtyp-Rezeptoren deutlich seltener blockiert werden⁶⁶. Diesbezüglich konnte im unbehandelten NSCLC gezeigt werden, dass Osimertinib (3. Generation) verglichen mit Erlotinib (1. Generation) bei deutlich längerer Therapiedauer und höherem Gesamtüberleben ein ähnliches Nebenwirkungsprofil besitzt⁵⁰.

Mit den ERBB2-Inhibitoren wurden auch für eine weitere Isoform der ERBB-Familie erfolgreich zielgerichtete Therapieansätze etabliert. Zunächst wurden rekombinant hergestellte Antikörper wie Trastuzumab entwickelt. Ihre antineoplastische Wirkung beruht auf der Blockierung der extrazellulären Domäne des Rezeptors, wodurch die Bindung von Liganden verhindert wird⁴⁶. Außerdem werden immunregulatorische Eigenschaften diskutiert⁶⁷. Die

Zulassung wurde für Tumoren mit immunhistochemisch oder durch Fluoreszenz-in-Situ-Hybridisierung nachgewiesener ERBB2-Amplifikationen erteilt ^{46,68}. Diese treten gehäuft in Mamma-, Magen- und Ösophaguskarzinomen auf ⁶⁹⁻⁷¹. Des Weiteren wurden *Small Molecules* wie Lapatinib entwickelt, welche die Aktivität der Tyrosinkinase von ERBB2 und zusätzlich von EGFR hemmen ^{57,72}. Angewendet werden sie in der Therapie des fortgeschrittenen metastasierten Mammakarzinoms ⁷³.

Für die weiteren genannten Beispiele der Rezeptortyrosinkinasen ALK, ROS1 und MET wurden sowohl selektive als auch Multikinase-Inhibitoren entwickelt ⁴⁷⁻⁴⁹. Zu den Multikinase-Inhibitoren gehören TKI wie Crizotinib und Entrectinib ^{74,75}. Crizotinib hemmt die Aktivität von ALK, ROS1 und MET, Entrectinib von ALK, ROS1 und der *Tropomyosin receptor kinase* (TRK) mit den Isoformen A - C. Eine Therapie kann beispielsweise bei Vorliegen einer ALK oder ROS1 Translokation im fortgeschrittenen nicht-kleinzelligen Lungenkarzinom erfolgen ^{74,75}. Darüber hinaus wird Entrectinib aufgrund der Wirksamkeit gegen TRK-Fusionen in zahlreichen weiteren Entitäten solider Tumore eingesetzt ⁷⁵.

Auch *Downstream* der Rezeptortyrosinkinasen wurden zielgerichtete Therapien gegen verschiedene Onkogene mit RAS aktivierender Wirkung entwickelt. Wie von Proietti et al. zusammengefasst, waren insbesondere Substanzen wie Dabrafenib, Vemurafenib und Encorafenib erfolgreich, welche zu der Gruppe der BRAF-Inhibitoren gehören ⁷⁶. Der antineoplastische Effekt wird durch die Hemmung der Kinaseaktivität von BRAF erreicht. Die Anwendung kann in der Therapie des malignen Melanoms erfolgen, in denen aktivierende V600E Mutationen gehäuft auftreten ⁷⁶. Jedoch sind sie auch für weitere Entitäten wie dem metastasierten kolorektalen Karzinom und dem nicht-kleinzelligen Lungenkarzinom zugelassen ^{77,78}. Im kolorektalen Karzinom ist allerdings der gleichzeitige Einsatz des EGFR-Antikörpers Cetuximab erforderlich, da die Monotherapie aufgrund eines Feedback Mechanismus über EGFR nicht wirksam ist ⁷⁸.

Des Weiteren stehen die RAS-GTPasen selbst seit längerer Zeit im Fokus intensiver Forschung. Aufgrund der hohen Inzidenz von aktivierenden Mutationen und Amplifikationen in zahlreichen Tumorentitäten sowie ihrer Rolle als Resistenzmechanismus in der anti-EGFR Behandlung, stellen sie ein attraktives Therapieziel dar ^{11,52,79}. Durch die hohe absolute Häufigkeit von KRAS-Mutationen konzentrierte sich die Forschung in den vergangenen Jahren insbesondere auf diese Isoform. Die Entwicklung war jedoch lange Zeit aufgrund fehlender *Drug binding Pockets* erschwert, weshalb nach anderen Möglichkeiten gesucht wurde, den Signalweg im weiteren Verlauf zu hemmen ⁸⁰. Hieraus erfolgte die Entwicklung von MEK-Inhibitoren, welche die Kinaseaktivität der Mitogen-aktivierten Protein Kinase Kinasen hemmen. Die klinische Zulassung ist momentan jedoch auf wenige Indikationen wie der Therapie von fortgeschrittenen Stadien des NRAS mutierten malignen Melanoms und des

nicht-kleinzelligen Lungenkarzinoms beschränkt ^{77,81-83}. Eine breite Anwendung war bislang nicht erfolgreich.

In jüngerer Vergangenheit konnten mit der Entwicklung von pan-KRAS- und mutationsspezifischen-Inhibitoren auch für die RAS-Proteine Erfolge erzielt werden. Ostrem et al. konnten als erste Moleküle entwerfen, welche sich bei Vorliegen der KRAS G12C Variante, an die Reste des falsch eingebauten Cysteins sowie an eine durch die Mutation neu entstandene Bindungstasche binden können. Dies führt unter anderem zu einer veränderten Bindungspräferenz von GDP und GTP, sodass GDP anschließend bevorzugt gebunden wird. KRAS verbleibt damit im inaktiven Zustand. Zusätzlich wird durch die Spezifität des Moleküls für die G12C Variante eine Bindung an KRAS-Proteine mit Wildtyp-Sequenz verhindert und somit mögliche Nebenwirkungen verringert ⁸⁴. Als erstes Medikament unter den G12C Inhibitoren erhielt Sotorasib, nach den Ergebnissen einer Phase II Studie von Skoulidis et al., eine vorläufige Zulassung für Patienten mit austherapiertem NSCLC. Die Studienergebnisse zeigen in 37,1 % der Patienten eine Reduktion der Tumormasse mit einem medianen Gesamtüberleben von 12,5 Monaten. Eine Tumorkontrolle konnte in 80,6 % erreicht werden ¹³. Momentan werden weitere Studien geplant, welche unter anderem Sotorasib mit dem bisherigen Therapiestandard (Docetaxel) des austherapierten NSCLC vergleichen ⁸⁵.

Neben Sotorasib befinden sich aktuell weitere mutationsspezifische Inhibitoren gegen verschiedene KRAS-Varianten in unterschiedlichen Stadien der Entwicklung ^{80,86,87}. Für einen großen Teil der Varianten fehlen jedoch weiterhin angreifbare Bindungstaschen, weshalb Therapieoptionen mit mutationsspezifischen Inhibitoren nicht für alle Mutationen zur Verfügung stehen. Vor diesem Hintergrund wurde nach weiteren Möglichkeiten für eine Hemmung der KRAS-Aktivität gesucht, woraus die Entwicklung der panKRAS Inhibitoren BI-1701963 und BI-3406 erfolgte ¹⁴⁻¹⁶. Der antineoplastische Effekt wird durch die Blockierung der Interaktion zwischen SOS1 und KRAS erreicht. Nach Bindung an SOS1 verhindern die Moleküle die beschleunigte Dissoziation von GDP und damit den Austausch mit GTP an KRAS, welches dadurch im inaktiven Zustand verbleibt ¹⁴. BI-1701963 wird momentan innerhalb von klinischen Studien an Patienten mit fortgeschrittenen, KRAS mutierten, soliden Tumoren nach erfolgloser Standardtherapie getestet. Die erste Studie testet dabei die Kombination mit dem MEK Inhibitor Trametinib, die zweite die Kombination mit Irinotecan unter Beschränkung auf kolorektale Karzinome ^{15,16}.

2.4. Maschinelles Lernen (*Machine Learning*)

Die neu entwickelten Antikörper und *Small Molecules* ermöglichen eine zielgerichtete Therapie einzelner RAEs. Verschiedene Beobachtungen zeigen jedoch, dass die Einflüsse auf den Signalweg und das Therapieansprechen multifaktoriell sind. So konnte wie unter 1.3 beschrieben, in Studien zur Behandlung mit dem KRAS G12C Inhibitor Sotorasib trotz

Nachweis der spezifischen Mutation nur in 37% der Fälle eine Reduktion der Tumormasse erreicht werden¹³. Ein weiteres Beispiel lässt sich in der Prädiktion eines Therapieansprechens mit Cetuximab im metastasierten kolorektalen Karzinom finden. Hierbei zeigte sich, dass auch Patienten mit Tumoren ohne hohe EGFR-Expression von einer Therapie profitieren können⁸⁸. Ebenfalls muss keine aktivierende Mutation in EGFR für eine Behandlung mit Cetuximab vorliegen⁸⁹.

Trotz hoher Evidenz für das Vorliegen von komplexen Interaktionen im RAS-Signalweg werden die meisten Therapieindikationen unter Berücksichtigung aktueller klinischer Leitlinien im Regelfall anhand von einzelnen RAEs gestellt. Um eine Übertherapie zu verhindern und Patienten zu identifizieren, die besonders profitieren, steigt der Bedarf an präziseren Modellen, die der Komplexität des Signalwegs gerecht werden. In den vergangenen zwei Jahrzehnten konnten in der molekularbiologischen Charakterisierung von malignen Tumoren insbesondere durch biotechnologische Fortschritte in DNA- und RNA-Sequenzierung enorme Fortschritte erreicht werden. Mittlerweile können tausende Charakteristika eines Tumors durch eine einzelne Sequenzierung bestimmt werden^{9,10,90,91}. Guinney et al. war es dadurch möglich in einem integrativen Ansatz vier „*Consensus molecular Subtypes*“ im kolorektalen Karzinom mit komplexen Phänotypen abzugrenzen⁹¹. Außerdem wurden die Daten in verschiedenen Publikationen erfolgreich für die Modellierung von Chemosensitivität, Überlebenszeitanalysen und Gen-Abhängigkeiten verwendet⁹²⁻⁹⁶. So konnten Dempster et al. für zahlreiche Onkogene zeigen, dass transkriptomweite RNA-Expressionsdaten besser für die Prädiktion einer Gen-Abhängigkeit geeignet sind als der Mutationsstatus des einzelnen Onkogens⁹⁵.

Durch die Größe und Struktur der neuen Datensätze ergeben sich jedoch verschiedene Probleme in der statistischen Modellierung, da herkömmliche Methoden nicht dafür ausgelegt sind^{97,98}. Exemplarisch lässt sich dies an einfachen linearen Regressionsmodellen darstellen. Diese weisen nur eine geringe Verzerrung auf, können jedoch eine hohe Varianz entwickeln. Entsprechend können die Modelle in manchen Situationen zur Überanpassung (*Overfitting*) an ihre Trainingsdaten neigen. Dadurch werden sie anfällig für Schwankungen in den Testdaten, was zu fehlerhaften Vorhersagen führt. Dies ist mitunter der Fall, wenn die Anzahl der Prädiktoren hoch ist oder diese stark miteinander korrelieren (Multikollinearität)^{98,99}. Beides sind Eigenschaften der beschriebenen molekularbiologischen Daten^{9,10,90,91}.

Um dennoch Modelle mit robusten Vorhersagen erstellen zu können, müssen relevante Prädiktoren identifiziert werden und der Einfluss von nicht-relevanten Variablen minimiert werden. Allgemein lassen sich für diesen Schritt der sogenannten Variablenselektion zwei übergeordnete Lösungsansätze unterscheiden. Der erste Ansatz besteht darin die Prädiktoren anhand von zusätzlichem Hintergrundwissen auszuwählen^{100,101}. Die Grundannahme ist hierbei, dass durch die zusätzlichen Informationen Prädiktoren mit einer höheren prognostischen Validität selektioniert werden. Im einfachsten Fall kann dies durch

Berücksichtigung von Expertenwissen oder Literaturrecherche sowie kritischer Prüfung der Variablen auf Plausibilität im Kontext der Fragestellung erfolgen¹⁰⁰. Für eine systematischere wissensbasierte Variablenselektion können Inhalte aus umfassenden Daten- und Wissensdatenbanken verwendet werden^{101,102}. Die Inhalte müssen an der Fragestellung orientiert aufgearbeitet werden, sodass eine Interpretation und somit auch eine Variablenselektion durch den Untersucher ermöglicht wird. Beispielsweise können Gene assoziiert mit bestimmten biologischen Prozessen aus Genontologien extrahiert oder zentrale regulatorische Knotenpunkte in literaturbasierten Protein-Protein Interaktionsnetzwerken identifiziert werden^{101,103-105}.

Unter dem zweiten Ansatz zur Variablenselektion können verschiedene Methoden der Statistik und des maschinellen Lernens zusammengefasst werden, die sich wiederum subklassifizieren lassen. Die Gemeinsamkeit ist, dass sie innerhalb des zu modellierenden Datensatzes nach Mustern oder Klassifikationen suchen, um geeignete Prädiktoren zu identifizieren^{98,100,106-111}. Zunächst kann zwischen Methoden mit unüberwachtem und überwachtem Lernen unterschieden werden^{100,106}. Methoden des unüberwachten Lernens suchen in den Daten nach unbekanntem Mustern, ohne dass Zielwerte vorgegeben werden, wohingegen beim überwachten Lernen Zielwerte vordefiniert sind. Eine weitere Differenzierung lässt sich durch die Einteilung in Methoden mit „*Filter*“ oder „*Wrapper*“ Mechanismen erreichen. Im Fall von *Filter*-basierten Methoden werden die Prädiktoren vor der Modellierung anhand von festgelegten Kriterien selektioniert. Diese beziehen sich auf intrinsische Eigenschaften der Daten, welche oftmals durch statistische Tests wie Korrelations- und Varianzanalysen untersucht werden. Hingegen werden bei *Wrapper*-basierten Methoden zahlreiche unterschiedliche Modelle mit verschiedenen Kombinationen der Prädiktoren getestet. Anschließend werden die Ergebnisse der resultierenden Modelle anhand von festgelegten Kriterien bewertet und die optimale Prädiktorkombination ausgewählt^{100,106}.

Neben *Filter*- und *Wrapper*-basierten Methoden wird oftmals noch eine zusätzliche Gruppe der *Embedded Models* abgegrenzt, welche die Variablenselektion direkt in den Prozess des Modelltrainings integrieren¹⁰⁷. In den vergangenen Jahrzehnten wurden in dieser Gruppe zahlreiche Methoden und Algorithmen für Problemstellungen des überwachten Lernens entwickelt, die in vielfältigen Bereichen eingesetzt werden. Darunter fallen unter anderem *Shrinkage* Methoden, *Random Forests*, *Support Vector Machines*, *Gradient Boosting Machines* sowie neuronale Netzwerke, die jeweils spezifische Vor- und Nachteile aufweisen^{98,108-111}. Das für den zu modellierenden Datensatz bestgeeignete Verfahren muss dabei oft empirisch auf Basis der Qualität der Vorhersagen bestimmt werden. Exemplarisch für Verfahren aus der Gruppe der *Embedded Models* werden hier *Shrinkage* Methoden und *Random Forests* näher beschrieben.

Shrinkage Methoden sind im statistischen Lernen eine Gruppe von verschiedenen Verfahren, welche den Einfluss unbedeutender unabhängiger Variablen auf die Modellierung begrenzen und somit die Modellvarianz verringern ⁹⁸. Zu diesen Verfahren gehören unter anderem die *Ridge*- und die *Lasso*-Regression, welche sich durch die Einführung eines zusätzlichen Strafterms von linearen Regressionen unterscheiden. In einer herkömmlichen multiplen linearen Regression gilt, es die Summe der quadrierten Abweichungen nach der Gleichung in Abbildung 3 (1) über die Koeffizienten (β) zu minimieren. Wie in den Abbildungen 3 (2) und 3 (3) zu erkennen ist, werden für die *Ridge*- und *Lasso*-Regression die Gleichungen jeweils durch leicht unterschiedliche Strafterme ergänzt. Diese bewirken eine Verkleinerung (*shrinkage*) der Koeffizienten von weniger relevanten unabhängigen Variablen, wodurch sich die Varianz unter gleichzeitiger Erhöhung der Verzerrung verringert. Im Fall der *Ridge*-Regression werden die Koeffizienten gegen Null gehend verkleinert, bei der *Lasso*-Regression gleich Null gesetzt. Die *Lasso*-Regression stellt somit zusätzlich ein Selektionsverfahren für Variablen dar. Die Stärke des Strafterms wird über den Lambda Parameter reguliert, welcher durch Kreuzvalidierung optimiert werden muss ⁹⁸.

$$RSS = \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 \quad (1)$$

$$\beta_{Ridge} = \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^p \beta_j^2 \quad (2)$$

$$\beta_{Lasso} = \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^p |\beta_j| \quad (3)$$

Abbildung 3. (1) Herkömmliche lineare Regression, (2) *Ridge* Regression, (3) *Lasso* Regression; n ist die Anzahl der Beobachtungen, p die Anzahl der unterschiedlichen Prädiktoren, β_0 der y -Achsenabschnitt, β_j der Koeffizient des jeweiligen Prädiktors, y_i die Werte der unabhängigen Variable sowie x_{ij} die Werte der Prädiktoren ⁹⁸.

Der *Random Forest* Algorithmus als weiteres Verfahren mit Eignung für hochdimensionale Daten basiert auf *Decision Trees* ¹¹⁰. Einfache *Decision Trees* haben wie die *Ordinary Least Square* Regression das Problem des *Overfitting* an den Trainingsdaten ¹¹⁰. Die Verringerung der Varianz wird, wie zuvor auch bei den regularisierten Regressionsmethoden durch die Erhöhung der Verzerrung erreicht. *Random Forests* teilen dafür die Trainingsdaten wiederholt in neue *Training*- und *Test-Sets* auf (*Bagging*). Für die neuen einzelnen *Training Sets* werden nun separat *Decision Trees* erstellt. Dabei wird zusätzlich nur ein begrenzter Anteil der unabhängigen Variablen verwendet. Ziel dabei ist es, eine Korrelation zwischen den *Decision Trees* zu verringern, die durch Überfokussierung auf besonders starke Prädiktoren entstehen kann. Nach Erstellung der einzelnen *Decision Trees* werden deren Vorhersagen

zusammengeführt. Dieses Vorgehen des Algorithmus führt schließlich zu einer erhöhten Verzerrung und somit zur angestrebten verringerten Varianz der Modelle ¹¹⁰.

2.5. Fragestellungen und Ziel der Arbeit

Die KRAS GTPase ist ein zentraler Mediator in der Signaltransduktion von zahlreichen Rezeptortyrosinkinasen. Mit der jüngsten Entwicklung der panRAS-Inhibitoren (SOS1-Inhibitoren) wird eine Hemmung der KRAS-Aktivität unabhängig vom Mutationsstatus ermöglicht. Es stellt sich deshalb die Frage nach einer Anwendbarkeit dieser Inhibitoren in Tumoren ohne aktivierende KRAS-Mutation.

Ziel dieser Arbeit ist die Charakterisierung einer Abhängigkeit des Zellwachstums von einer KRAS-Expression in Tumorzelllinien ohne aktivierende KRAS-Mutation basierend auf Daten aus CRISPR-knockout und RNAi-knockdown Experimenten des Achilles- und DRIVE-Projekts. Zusätzlich sollen verschiedene Optionen für eine Modellierung dieser Abhängigkeit evaluiert werden.

3. Diffusion kernel-based predictive modeling of KRAS dependency in KRAS wild type cancer cell lines

Die Ergebnisse der vorgestellten Doktorarbeit wurden im Rahmen der folgenden Publikation beschrieben:

Ulmer B, Odenthal M, Buettner R, Roth W, Kloth M. Diffusion kernel-based predictive modeling of KRAS dependency in KRAS wild type cancer cell lines. *NPJ Syst Biol Appl* 2022; **8**(1): 2.

ARTICLE OPEN



Diffusion kernel-based predictive modeling of KRAS dependency in KRAS wild type cancer cell lines

Bastian Ulmer¹, Margarete Odenthal¹, Reinhard Buettner¹, Wilfried Roth² and Michael Kloth²

Recent progress in clinical development of KRAS inhibitors has raised interest in predicting the tumor dependency on frequently mutated RAS-pathway oncogenes. However, even without such activating mutations, RAS proteins represent core components in signal integration of several membrane-bound kinases. This raises the question of applications of specific inhibitors independent from the mutational status. Here, we examined CRISPR/RNAi data from over 700 cancer cell lines and identified a subset of cell lines without KRAS gain-of-function mutations (KRAS^{wt}) which are dependent on KRAS expression. Combining machine learning-based modeling and whole transcriptome data with prior variable selection through protein-protein interaction network analysis by a diffusion kernel successfully predicted KRAS dependency in the KRAS^{wt} subgroup and in all investigated cancer cell lines. In contrast, modeling by RAS activating events (RAE) or previously published RAS RNA-signatures did not provide reliable results, highlighting the heterogeneous distribution of RAE in KRAS^{wt} cell lines and the importance of methodological references for expression signature modeling. Furthermore, we show that predictors of KRAS^{wt} models contain non-substitutable information signals, indicating a KRAS dependency phenotype in the KRAS^{wt} subgroup. Our data suggest that KRAS dependent cancers harboring KRAS wild type status could be targeted by directed therapeutic approaches. RNA-based machine learning models could help in identifying responsive and non-responsive tumors.

npj Systems Biology and Applications (2022)8:2; <https://doi.org/10.1038/s41540-021-00211-8>

INTRODUCTION

The RAS signaling pathway is a key driver of carcinogenesis in many different tumor entities^{1–4}. Frequently, gain-of-function mutations or copy number alterations (CNA) at different levels of the signaling cascade lead to overactivity and thereby increased cell growth, migration, and invasion^{2,3,5}. Due to its high relevance in cancer per se, intensive research has been conducted aiming to develop targeted therapies. In recent years, several drugs entered clinical application focusing on inhibitors that are directed against membrane receptors preventing constitutively activated signal transduction^{6–8}. EGFR represents one of the most well-known examples⁹. However, therapy is limited by various resistance mechanisms in the receptor itself or other RAS/RAF pathway elements^{7,10–12}. Common mechanisms with high clinical relevance include mutations in downstream RAS GTPases^{2,3,10}, which represent important nodes in signal integration from cell membrane to nucleus¹³. Here, somatic point mutations lead to constitutive activation downstream of membrane receptors, thereby hindering therapeutic success^{5,10}. Despite intensive research, drug binding pockets could not be identified in these proteins for a long time, making the development of direct inhibitory pharmacotherapy difficult. MEK inhibitors have become a first option to overcome this mechanism of resistance by inhibiting downstream mitogen-activated protein kinase kinases. However, until now, clinical efficacy has only been demonstrated for specific applications such as NRAS/BRAF mutated melanoma^{14–16}. More recently, mutation-specific and panKRAS inhibitors have been developed that inhibit KRAS activity directly or indirectly^{17–20}. The mutation-specific inhibitors exploit structural changes in the KRAS protein that result from oncogenic point mutations so that cells expressing wild-type protein are less affected, which is expected to reduce toxicity of the therapy. However, so far specific inhibitors could only be designed for a few KRAS mutations such as G12C. This led to the development of the

panKRAS inhibitors, which downregulate KRAS activity by binding to SOS1 and thus enable an application independent of the mutation status. PanKRAS inhibitors such as BI1701963 are currently being tested in clinical trials including combinations with MEK inhibitors²¹. Furthermore, following the results of a phase II trial, Sotorasib became the first drug of mutation-specific inhibitors to receive preliminary approval for patients with therapy-refractory NSCLC²².

The emergence of new therapies with small molecule RAS inhibitors also increases the relevance of identifying responsive and resistant tumors as accurately as possible. Depending on the drug and tumor entity, different markers are currently used as predictors. Those include mutations, CNA and gene expression^{6,23}. The selection of predictors depends on the respective tumor entity and its characteristics. For example, activating EGFR and ERBB2 mutations are considered positive predictive markers for therapy with tyrosine kinase inhibitors against receptors of the EGFR-family in non-small cell lung cancer (EGFR) and colorectal cancer (ERBB2)^{23,24}. Response to ERBB2-antibody Trastuzumab correlates with ERBB2 expression and copy number status in breast cancer^{6,25}. Furthermore, EGFR resistance mutations such as T790M are important in anti-EGFR therapy in non-small-cell lung cancer as well as activating KRAS mutations in colorectal cancer^{11,26}. Besides activating mutations, expression-based RAS signatures may also improve therapy response prediction including treatment with KRAS inhibitors^{27,28}. This could be particularly important in tumors with more complex activation mechanisms by yet unknown RAS-activating events (RAE).

Recent progress in clinical development of KRAS inhibitors has raised interest in predicting the tumor dependency on frequently mutated RAS-pathway oncogenes. However, even without such activating mutations, RAS proteins represent core components in signal integration of several membrane-bound kinases. This raises the question of applications of the inhibitors independent from

¹Institute of Pathology, Cologne University Hospital, Cologne, Germany. ²Institute of Pathology, Mainz University Hospital, Mainz, Germany. [✉]email: bastian.ulmer@uk-koeln.de

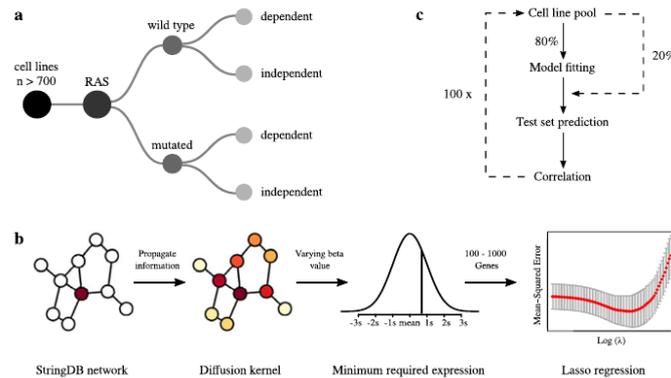


Fig. 1 Cancer cell line classification algorithm and gene dependency modeling strategies. **a** Strategy of cell line subgrouping leading to the investigated subgroups (HRAS^{wt}/HRAS^{mut}, KRAS^{wt}/KRAS^{mut}, NRAS^{wt}/NRAS^{mut}). **b** Variable selection workflow for whole transcriptome RNA-expression data consisting of the construction of a literature-based gene network followed by further selections steps with centrality quantification through a diffusion kernel and a minimum required expression level. Several different constellations of the hyperparameters were tested. Final modeling was performed using a Lasso, Elastic Net or Random Forest regression. **c** Workflow of iterative model fitting and performance evaluation for each gene dependency dataset.

the mutational status. In this work, we analyzed KRAS dependency by CRISPR/RNAi data from the Achilles- and DRIVE-Project^{29–32}. We identified a subgroup of KRAS dependent cell lines harboring wild type status in KRAS (Fig. 1a). KRAS dependency of this subgroup could not be predicted by RAE-based models or those of previously published expression signatures. Instead, our machine learning approach based on whole transcriptome data and diffusion kernel-based variable selection using protein-protein interaction network analysis significantly improved KRAS dependency prediction in KRAS^{wt} cancer cell lines (Fig. 1b/c, see methods).

RESULTS

Characterization of KRAS dependency in KRAS wild type cancer cell lines

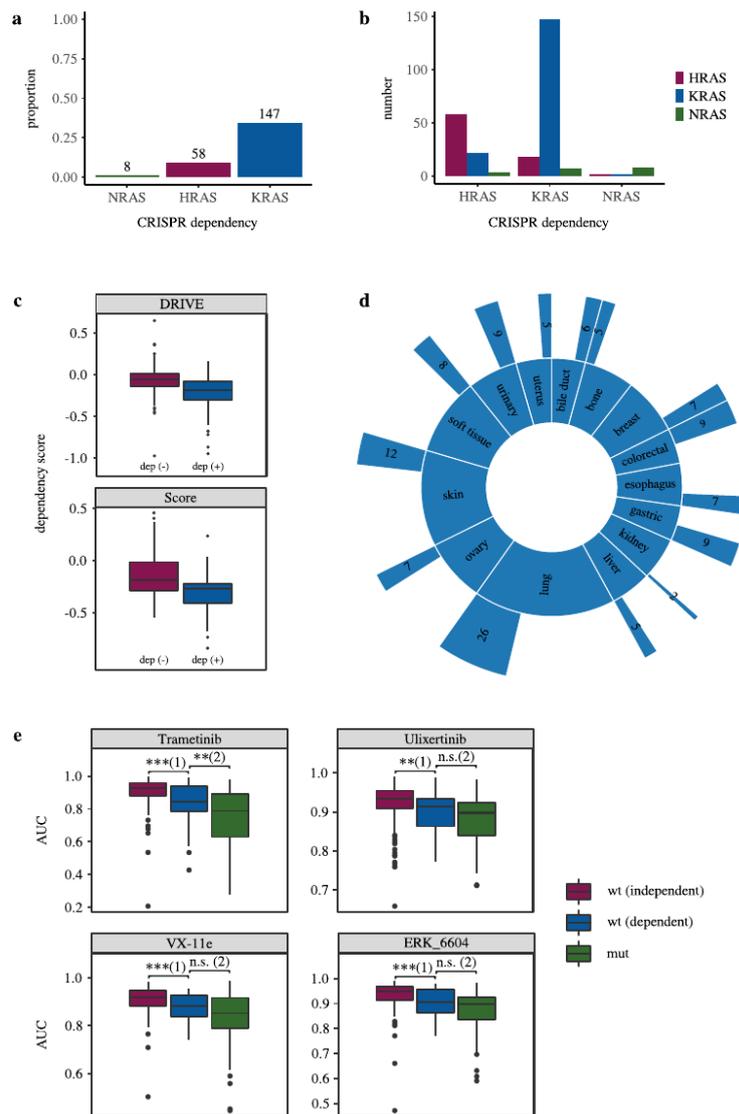
Currently, several KRAS inhibitors are being tested in clinical trials, first preliminary approvals have been granted and further are likely to follow^{20–22,33}. Primarily these inhibitors are developed with the aim to overcome gain-of-function mutations^{17–20}. However, there is evidence that patients with wild type status in RAS/RAS-oncogenes may also benefit from therapies targeting RAS genes²⁸. To investigate this phenomenon, we analyzed CRISPR knockout data of cancer cell lines from the Achilles Project focusing on a dependency characterization of KRAS, NRAS, and HRAS^{29–31,34}. These data provide a valid approximation of chemosensitivity to inhibitors targeting wild type RAS, which are not yet available in databases such as GDSC or CCLE. The KRAS^{wt} subgroup exhibited the largest fraction of dependent cell lines followed by HRAS^{wt} (Fig. 2a), whereas dependencies in NRAS^{wt} tended to be limited to a few cases. Co-dependencies of the individual wild type subgroups, i.e., the simultaneous presence of two dependencies, were most frequently observed for KRAS^{wt} and HRAS^{wt}, but rather rare overall (Fig. 2b). To validate the existence of a KRAS dependent subgroup in KRAS^{wt} cell lines we additionally analyzed two other dependency data sources of the DRIVE- (RNAi) and the Score-Project (CRISPR). Again, cell lines classified as KRAS dependent in the initial Achilles CRISPR screen exhibited a significantly higher dependency in both data (Fig. 2c). Next, we examined the data for an association between tissue origin and KRAS dependency to rule out any potential bias. For each entity, we performed a Fisher test with the respective binary

characteristics of belonging to the individual entity (yes/no) and KRAS dependency (yes/no). Although we observed a proportionally increased number of KRAS-dependent cases for colorectal and gastric cell lines within the entities (Fisher test; colorectal: $p = 0.03$; gastric: $p = 0.03$; $n = 573$), the overall composition of the dependent cell lines demonstrated a heterogeneous distribution (Fig. 2d). With 26 cases, the largest fraction of responsive cell lines originated from the lung, followed by skin tumors with 12 and tumors of the central nervous system (CNS) with 11 cases. However, these entities account for only 18% (lung), 9% (skin), and 8% (CNS) of the subgroup. We therefore assumed only a limited impact of tissue-specific effects on the results of our further studies.

Our results so far suggest that KRAS has an important survival function in signal integration in a specific subgroup of KRAS^{wt} cell lines. To further investigate a potential clinical relevance, we analyzed drug sensitivity to several downstream interacting MEK inhibitors. For this analysis, we additionally excluded cell lines with non-deleterious mutations in BRAF, HRAS, and NRAS from the KRAS^{wt} subgroup (RAS^{wt}/RAF^{wt}). We chose AUC (GDSC) or Active Area (CCLE) for drug sensitivity quantification because validity of extrapolated IC50 values is limited for cell lines that were only partially responsive or unresponsive within the experimentally tested inhibitor concentrations, as stated by GDSC³⁵. To this end, the AUC has been successfully used in a variety of other publications^{36–38}. As expected, we found consistent associations to MEK inhibitors across multiple compounds and datasets (Fig. 2e, Supplementary Fig. 1). Overall, the sensitivity of the KRAS^{wt} cell lines with KRAS dependency was significantly higher compared to the KRAS^{wt} independent subgroup, but slightly lower than the one of the KRAS^{mut} cell lines. Thus, we were able to provide further evidence for an increased RAS activity in the depicted subgroup. Furthermore, these analyses suggest that a combined therapy approach with panKRAS and MEK inhibitors, as it is currently being tested in KRAS mutated tumors²¹, may also be effective for patients with RAS^{wt}/RAF^{wt} tumors.

RAE possess limited sensitivity and specificity for KRAS dependency prediction in KRAS^{wt} subgroup

According to our results, KRAS^{wt} cancers with KRAS dependency potentially reflect a subgroup of cancers with therapeutic



relevance. Therefore, we aimed to specify this subgroup in more detail. There are numerous well characterized mechanisms of RAS activation that can lead to KRAS dependency (RAS activating events, RAE). These include mutations, overexpression and CNA in upstream genes as well as amplification of KRAS itself^{6,23,39–41}. Using differential expression analysis in KRAS^{wt} cancer cell lines between KRAS dependent and independent subsets we were able to detect several of already known pathway activating regulators. These included ERBB2 and KRAS (Fig. 3a), which were significantly overexpressed in the KRAS dependent KRAS^{wt} subgroup. In

addition, overrepresentation analysis revealed an association of the differentially expressed genes ($n_{genes} = 358$) to various receptor tyrosine kinase (RTK) signaling pathways such as EGFR, FGFR, MET, NTRK, and IGFR (Supplementary Data 1). However, RAE identification based on molecular genetic events or expression is not trivial and may lead to erroneous conclusions. Mutations provide an illustrative example, as their clinical implications are sometimes not known with certainty^{42–44}. Therefore, we used the genome wide CRISPR data from the Achilles Project for the detection of RAE, as they reflect direct

Fig. 2 RAS dependency characterization and associations to MEK inhibitor responsiveness. **a** Proportion of gene-dependent cell lines in the four subgroups of wild type cancer cell lines. The numbers above the columns indicate the absolute number of dependent cell lines. **b** Total number of co-dependencies for each gene. **c** Validation of the elaborated KRAS dependent and independent subgroups. After dividing the cell lines into KRAS dependent and independent using the Achilles Project CRISPR data, we verified the existence of the two subgroups in data of the DRIVE (RNAi) and the Score Project (CRISPR). Cell lines classified as KRAS dependent in the Achilles data exhibited a significantly higher dependency in both screens (Wilcox Test; DRIVE: $p = 4.6 \times 10^{-10}$, $n = 342$; Score: $p = 2.1 \times 10^{-4}$, $n = 124$). **d** Proportions of the different entities in the KRAS^{wt} group (inner circle) and the proportion of KRAS-dependent cell lines in each entity with indicated absolute numbers (outer circle). Only entities with at least ten cell lines were included in the figure. In absolute numbers, lung tumors were the most represented entity among KRAS dependent cell lines, followed by skin tumors. Overall, the group is very heterogeneous without one entity clearly dominating. **e** Characterization of MEK-inhibitors sensitivity in KRAS^{wt} cancer cell lines with dependent (blue, wt (dependent)) independent (purple, wt (independent)) status and as a reference KRAS^{mut} cases (green, mut). The symbols above the brackets refer to the following significance codes: *** < 0.001; ** < 0.01; * < 0.05; 'n.s.' > 0.05. In the overall comparison between the three groups, KRAS^{wt} cell lines with KRAS dependency are significantly more responsive to MEK inhibitors (lower AUC) than KRAS^{wt} independent group, but for some inhibitors less responsive than the KRAS^{mut} cell lines (Wilcox Test: Trametinib: (1) $p = 4.6 \times 10^{-4}$ ($n = 183$), (2) $p = 7.4 \times 10^{-3}$ ($n = 113$); Ulixertinib: (1) $p = 4.1 \times 10^{-3}$ ($n = 183$), (2) $p = 9.5 \times 10^{-2}$ ($n = 112$); VX-11e: (1) $p = 4.9 \times 10^{-4}$ ($n = 178$), (2) $p = 9.7 \times 10^{-2}$ ($n = 111$); ERK_6604: (1) $p = 1.8 \times 10^{-4}$ ($n = 179$), (2) $p = 1.3 \times 10^{-1}$ ($n = 111$)). For further compounds of CCLE, GDSC1, and GDSC2 see also Supplementary Fig. 1. Box plot annotation (**c**, **e**): 25th percentile (box bottom), 75th percentile (box top), median (box center), whiskers top/bottom $\pm 1.5 \times$ interquartile range, outliers are shown as dots.

information about the survival potential of the gene. Consistent with the expression data, genes of the pathways mentioned above were associated with higher dependency in the KRAS dependent subgroup. In addition, we found other RTK pathway associations such as those to KIT and PI3K (Fig. 3b). In the next step, we analyzed each cell line for potential RAE that could be causal for KRAS dependency. In total we used 43 binary markers (Supplementary Data 2) including 42 gene dependencies and KRAS amplification status, as this is associated with worse outcome and tumor progression in different cancer types^{40,45,46} (see methods). This simple approach allowed us to assign at least one RAE to 67% of the KRAS dependent cell lines either by co-dependency or amplification (Fig. 3c). For the remaining 33%, other monogenic activation mechanisms that were not considered or complex-genetic activation would be conceivable. In addition, our analysis showed that the distribution of markers is relatively heterogeneous and that multiple co-dependencies may also occur (Fig. 3d/e). However, a causal relationship is not assured by this approach. Eventually, independent simultaneous occurrences cannot be ruled out despite targeted and literature-based marker selection. Furthermore, it may not always be accurate to conclude that an RAE leads to KRAS dependency. Strikingly, this is illustrated by repeating the analysis on wild type cell lines without KRAS dependency, which assigned a RAE to 63% of the cell lines (Fig. 3c). To further reveal more complex interactions between RAE and KRAS dependency, we performed Lasso regression using quantitative RAE dependencies as predictors (see methods). This first approach based on RAE achieved no significant correlation between model predictions and experimental data (Pearson's $r = 0$, $p = 0.92$, $n = 529$), which underlines the need for improved variable selection and integrative modeling.

KRAS dependency prediction in KRAS^{wt} subgroup using previously published RNA signatures

Our initial approach indicates some challenges in the prediction of KRAS dependency in KRAS^{wt} cancers by known RAE. We therefore searched for suitable alternative modeling strategies. Regarding the identification of RAS- and KRAS-dependent cancers, several approaches have been taken including RNA-based expression signatures^{27,28}. To characterize whether these known signatures represent an improvement in prediction, we created machine learning models and examined their predictive performance (see methods). Overall, the resulting predictive performance was not satisfactory. Models based on the signature published by Loboda et al. achieved a correlation of 0.15 (Pearson's r , $n = 567$), those based on the signature by Singh et al. 0.18 (Pearson's r , $n = 567$). However, it should be noted that we were not able to assign all gene identifiers in the given signatures, leading to a loss of four

genes (2.7%) in the signature by Loboda et al. and 14 (2.7%) in the signatures from Singh et al. (Supplementary Data 3). Due to the small number of missing genes, we considered the influence on the results to be neglectable in both cases. Consequently, our analyses indicate a more complex situation in the prediction of KRAS dependency in KRAS^{wt} cancers even with previously published expression signatures.

Diffusion kernel-based protein-protein interaction network analysis improves predictive modeling of KRAS dependency in KRAS^{wt} cancer cell lines

Consistent with our results so far, recent literature showed that gene dependencies of KRAS, NRAS, and HRAS are difficult to model by whole transcriptome analysis and that mutation status is proposed to be more robust⁴⁷. To improve modeling in KRAS^{wt} cancers, we tested our own strategies including different machine learning algorithms as well as variable selection through gene centrality estimation in a protein-protein interaction network by a diffusion kernel (See methods, Fig. 1b/c). Initially, reference models with all predictors available in the RNA expression dataset were created (47768 genes) using Lasso regression. Compared to the previous signatures we achieved significantly higher correlations between test set predictions and the experimentally determined dependency demonstrating that the predictive performance of RNA-based models can be improved depending on the modeling strategy. For the CRISPR data Pearson's r was 0.23 and for RNAi data 0.25 (Fig. 4a). However, due to the high number of predictors, it is possible that the models are prone to overfitting. Dataset-specific artifacts such as random associations of predictors to the dependent variable, characteristics of the cell line model or effects caused by the entity distribution could affect external validity. To further improve the performance and to test strategies against possible overfitting we included a variable selection step before modeling. This strategy involved an initial restriction of predictors to genes derived from a literature-based protein-protein interaction network (StringDB), followed by further reduction steps through estimation of gene centrality in the network using a diffusion kernel (See methods, Fig. 1b/c). The hypothesis is that the selected genes are important regulatory elements in KRAS signaling and thus represent suitable predictors of KRAS dependency with biological significance. This approach combined with additional hyperparameter optimization (see methods) resulted in a total of 105 different predictor sets with a size ranging from 100 to 1000 genes. In the subsequent modeling by Lasso regression with optimized hyperparameters we reached a maximum correlation of 0.43 (Pearson's r) between observed and predicted KRAS dependency (Fig. 4a, Supplementary Data 4). This corresponds to a performance improvement of

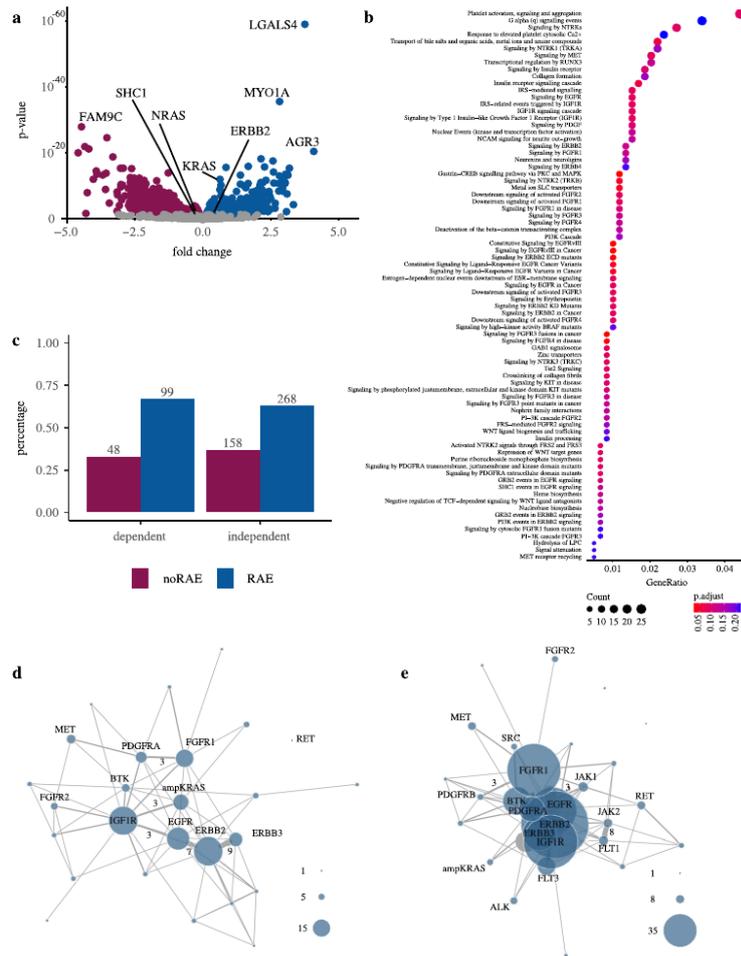


Fig. 3 Transcriptional characterization and dependency analysis of KRAS dependent KRAS^{wt} cell lines. **a** Differentially expressed genes in KRAS^{wt} dependent vs independent cell lines ($n = 567$). Positive values on the x-axis reflect higher expression in the dependent subgroup, correspondingly negative values reflect higher expression in the independent subgroup. **b** Overrepresentation analysis (Reactome) of genome wide CRISPR screen genes exhibiting a higher dependency in the KRAS^{wt} dependent subgroup (Wilcoxon–Mann–Whitney Test; $n_{genes} = 1038$). **c** Percentage of cell lines harboring at least one RAE (blue) or no RAE (purple) in KRAS^{wt} subgroup for KRAS dependent (left) and independent (right) cell lines. Absolute values are shown above each column. **d**, **e** Binary co-dependency network of RAE in KRAS^{wt} highlighting the heterogeneous distribution of RAE (Dependent cell lines **d**), independent **e**). Number of co-dependencies shared between two genes is shown if there were more than two co-dependencies. Node size refers to the number of cell lines classified as dependent on the respective gene. Cell lines without RAE were not included in the figures.

72% compared to the reference models with all available predictors of the RNA sequencing data. Best predictions could be achieved by using an initial variable set reduced to only 500 genes. Additional models based on the complete gene set of the protein network consistently performed weaker when compared to the diffusion kernel-based prediction (Fig. 4a).

These results demonstrate that variable selection by the network-based approach significantly improves the performance of models predicting KRAS dependency in KRAS wild-type cancer

cell lines. The predictions were clearly superior to RAE-based models or those of previously published expression signatures (Fig. 4b). In addition, it enables a substantial reduction of predictors prior modeling without impairing model performance. However, a closer analysis of the models revealed specific challenges. For all models, we observed that absolute errors increase significantly toward the outsides of the distribution, thereby also reflecting outliers which are difficult to predict (Fig. 4c). We assumed biological effects or attributes of dependency

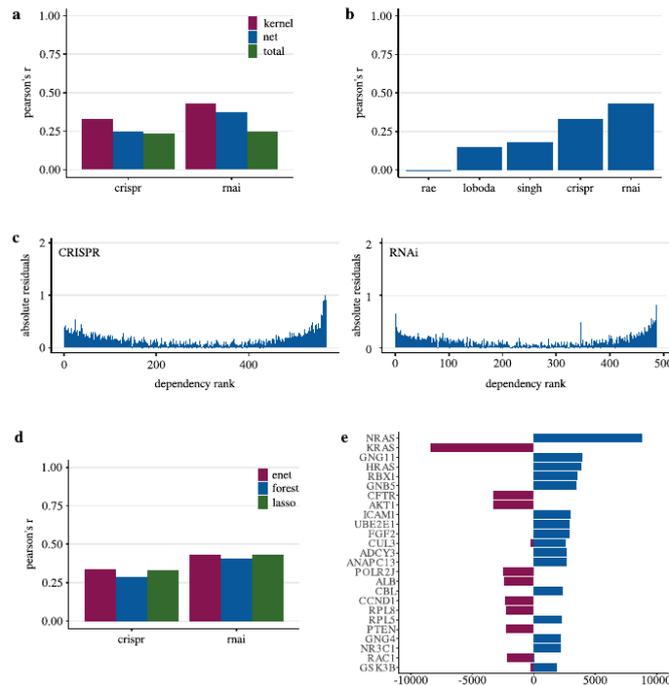


Fig. 4 Performance of different KRAS dependency modeling strategies and predictor analysis in KRAS^{wt} cell lines. **a** Correlation analysis (Pearson's r) in independent test sets between the experimentally determined KRAS cancer cell line dependency and our machine learning-based predictions for varying sets of predictors (see methods). Results are shown for models using all available predictors of the RNA sequencing data (total), all available predictors of the protein interaction network (net) and predictors selected by the diffusion kernel with hyperparameter optimization (kernel). Models were based on KRAS^{wt} cell lines of the different datasets (crispr - Achilles CRISPR effect data ($n = 567$), rna - DRIVE RNAi (DEMETER2) data ($n = 487$)). In case of the diffusion kernel variable selection workflow maximum correlation was reached with a hyperparameter constellation using 500 predictors. For complete results of hyperparameter tuning see Supplementary Data 4. **b** Performance (Pearson's r) of KRAS dependency models in KRAS^{wt} group compared between the different approaches (RAE - RAE-based models (CRISPR data), Loboda - Models using RNA expression of the gene selection by Loboda et al. (CRISPR data), Singh - Models using RNA expression of the gene selection by Singh et al. (CRISPR data), CRISPR - Best performing models using RNA expression of the gene selection by the diffusion kernel with optimized hyperparameters (CRISPR data), RNAi - Best performing models using RNA expression of the gene selection by the diffusion kernel with optimized hyperparameters (RNAi data)). For CRISPR/RNAi correlation analysis was performed similarly to (a). Correlation coefficients for RAE, Loboda and Singh were determined as described above. **c** Absolute error of CRISPR/RNAi models for each cell line using mutation- and best performing RNA-predictor set. Summarized results of 400 unique models are shown in the two waterfall plots. Cell lines were ordered by ascending observed KRAS dependency from left to right. The absolute error was estimated by summing the individual absolute differences of the predicted values from the observed values. **d** Correlation analysis (Pearson's r , $n_{\text{CRISPR}} = 567$, $n_{\text{RNAi}} = 487$) performed similarly to (a) this time comparing models using different algorithms (Elastic Net regression - enet, Random Forest regression - forest, Lasso regression - lasso). Neither Elastic net nor Random Forest Regression could improve the Lasso predictions of KRAS dependency. **e** Occurrence frequency of RNA-predictors in 12000 unique models of KRAS dependency (CRISPR/RNAi) in KRAS^{wt} cancer cell lines. Only models using the variable selection by the diffusion kernel were included. Negative values indicate the frequency of how often the predictor had a negative coefficient in the models (associated with higher KRAS dependency), positive values the frequency of how often the predictor had a positive coefficient (associated with lower KRAS dependency). The 25 most redundant genes are shown here.

distribution to be the cause for this observation. Therefore, we tested Elastic Net and Random Forest regression as two additional algorithms for modeling. However, these methods could not improve the prediction accuracies (Fig. 4d). For the predictor sets selected by the diffusion kernel maximum observed correlation coefficients (Pearson's r) with Elastic Net regression after hyperparameter optimization were 0.33 (CRISPR data) and 0.43 (RNAi data). Interestingly, despite the skewed distribution of the KRAS dependency with more independent cell lines the decision tree-based Random Forest regression performed worse than the linear regression models (Pearson's r ; CRISPR: 0.29; RNAi: 0.41).

Central components of KRAS interaction network contain non-substitutable information signals for dependency modeling

Next, we characterized the predictors of KRAS dependency in KRAS wild type models. Among the 12,000 models of CRISPR/RNAi data using gene sets selected by the diffusion kernel and hyperparameter optimization, the frequency of non-zero model coefficients was quantified for each predictor. In total, 1964 genes were used at least once in a model. Most frequently used predictors were NRAS and KRAS expression, followed by GNG11, HRAS, and RBX1 (Fig. 4e, Supplementary Data 5). KRAS expression

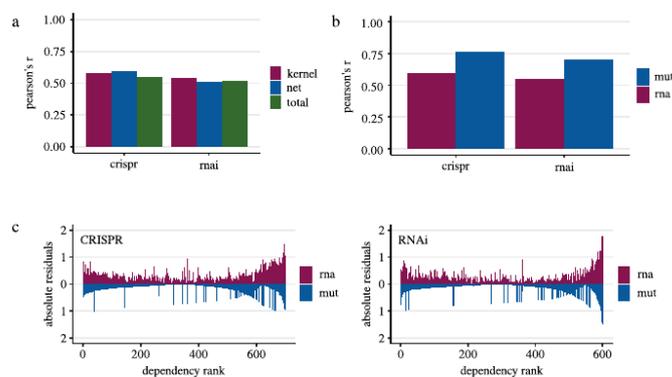


Fig. 5 Performance of mutation status-/RNA expression-based KRAS dependency models and analysis of error distributions in the complete cell line dataset. **a** Correlation analysis (Pearson's r) in independent test sets between the experimentally determined KRAS cancer cell line dependency and our machine learning-based predictions for varying sets of predictors (see methods). Results are shown for models using all available predictors of the RNA sequencing data (total), all available predictors of the protein interaction network (net) and predictors selected by the diffusion kernel with hyperparameter optimization (kernel). For both dependency datasets models were based on the entire cell line set (crispr - Achilles CRISPR effect data ($n = 698$), rnai - DRIVE RNAi (DEMETER2) data ($n = 601$)). In case of diffusion kernel variable selection workflow maximum correlation was reached with a hyperparameter constellation using 1000 predictors. For complete results of hyperparameter tuning see Supplementary Data 4. **b** Correlation analysis (Pearson's r) performed similarly to (a) this time comparing models either using KRAS mutation status (mut) or best performing predictors of the RNA sequencing data (rna) in the respective datasets (crispr - Achilles CRISPR effect data ($n_{rna} = 698$, $n_{mut} = 704$), rnai - DRIVE RNAi (DEMETER2) data ($n_{rna} = 601$, $n_{mut} = 613$)). Using RNA sequencing data as predictors, the best performance was achieved either with the complete protein interaction network (CRISPR) or a subset of the network consisting of 1000 genes selected by the diffusion kernel (RNAi). **c** Absolute error of CRISPR/RNAi models for each cell line using mutation- and best performing RNA-predictor set (CRISPR: complete protein interaction network; RNAi: diffusion kernel selection with 1000 genes). Summarized results of 400 unique models are shown in the two waterfall plots. Cell lines were ordered by ascending observed KRAS dependency from left to right. The absolute error was estimated by summing the individual absolute differences of the predicted values from the observed values (RNA expression-based (rna): purple bars; Mutation status-based (mut): blue bars). Predictions of models using mutation status show two local minima in the absolute error distributions indicating the binary prediction results. For both types of predictors residuals tend to increase at both sides of the distribution.

was associated with an increase, NRAS, GNG11, HRAS, and RBX1 with a decrease of KRAS dependency. The inverse regulation of KRAS, NRAS, and HRAS expression in context of KRAS dependency is consistent with results from Fig. 2b showing that co-dependency between these genes is a rare event in KRAS wild type cancer cell lines. We also performed an overrepresentation analysis of predictors that were used in at least 10% of all models, which corresponds to a list of 63 genes. As expected, numerous pathways related to signal transduction were overrepresented because of the variable selection based on the KRAS centralized network (Supplementary Data 6). Accordingly, the results show a profile comparable to Fig. 3b. Most significant regulated pathways were extra-nuclear estrogen signaling ($p = 3.1 \times 10^{-16}$), Diseases of signal transduction by growth factor receptors and second messengers ($p = 2.4 \times 10^{-15}$) as well as several pathways which involve tyrosine kinase signaling. Again, this highlights the heterogeneity of RAE and further potential approaches for combination therapies in the KRAS^{wt} subgroup.

The results indicate that KRAS dependent cancer cell lines in the KRAS^{wt} group exhibit a distinctive phenotype or activation state, which might be represented by expression data. However, RNA expression data are highly structured and intercorrelated^{31,48}, so that genes selected by our literature-based approach might be replaced without losing predictive power of the models. Consequently, a specific phenotype or gene regulation in the context of KRAS dependency would be less likely. On the other hand, a decrease in prediction power would mean a loss of information that could not be compensated by other predictors and context-specific regulation would be more likely. To investigate this, we selected the models with best performing hyperparameters and

extracted all predictors with non-zero coefficients, resulting in lists of 293 (CRISPR) and 184 (RNAi) genes. Then, separately for each gene list and dependency dataset we excluded the lists from the superset of all proteins in the interaction network and repeated modeling with these selections. In fact, we found a lower correlation for both datasets compared to the models using the total number of network genes. Performance in models of Achilles CRISPR dependency data dropped about 20% (Pearson's $r = 0.20$, $n = 567$) and for DRIVE RNAi data about 29% (Pearson's $r = 0.27$, $n = 487$). The loss of predictive power suggests that a non-compensable loss of information has taken place. This speaks for a regulation of genes selected by our interaction network-based in context of KRAS dependency, indicating a KRAS dependency phenotype in the KRAS^{wt} subgroup.

Comparison of RNA expression data and KRAS mutation status as predictors of generalized KRAS dependency

Finally, after predicting KRAS dependency in the KRAS^{wt} subgroup we also applied our selective modeling strategy to the entire set of cancer cell lines. Initial reference models using Lasso regression and all predictors available in the RNA expression dataset achieved correlations (Pearson's r) of 0.55 (CRISPR) and 0.52 (RNAi) between observed and predicted KRAS dependency (CRISPR) (Fig. 5a). Prior variable selection by the diffusion kernel yields slight improvements in performance (Pearson's r : 0.58 (CRISPR), 0.54 (RNAi)). Using all available predictors of the protein interaction network, a minimal improvement in prediction accuracy was observed for the CRISPR data again (Pearson's r : 0.59), but not for the RNAi data, where models performed weaker (Pearson's r : 0.51). Despite the good results, consistent with

recently published literature models based on KRAS mutation status outperformed RNA-based models (Fig. 5b; Pearson's r : 0.76 CRISPR, 0.7 RNAi)¹⁷. Again a closer analysis of the models revealed specific challenges. For both types of predictors, we observed that absolute errors increase significantly toward the outsides of the distribution (Fig. 5c). As expected, the mutation-based models provided near-binary predictions, as they were built on three binary predictors which cannot represent the continuous distribution of KRAS dependency. This is illustrated by the distribution of absolute residuals in Fig. 5c, which exhibits two local minima in the mutation-based models, each reflecting one of the two almost constant prediction values of the models. In contrast, RNA-based models did allow quantitative prediction and identified potential borderline cases. Nevertheless, as before, they struggled to adequately predict highly responsive cases. These results show distinct advantages and disadvantages of the two types of predictors. Accordingly, comprehensive models of KRAS dependency may benefit from a combination of KRAS mutation status and RNA expression data, as well as other predictor types.

DISCUSSION

The RAS signaling pathway possesses a central position in the oncogenesis of several tumor entities via numerous mechanisms. Recent progress in clinical development of KRAS inhibitors has raised interest in predicting the tumor dependency on frequently mutated RAS-pathway oncogenes. However, even without such activating mutations, RAS proteins represent core components in signal integration of several membrane-bound kinases. This raises the question of applications of the inhibitors independent from the mutational status.

In this work we identified a KRAS dependent subset of KRAS wild type cell lines. Consistently, across six different datasets from CRISPR, RNAi, and chemosensitivity experiments, we found evidence for increased RAS activity in the elaborated subgroup. Our results suggest that patients without activating KRAS mutations may also benefit from targeted therapies against KRAS. Here, compounds such as the SOS1-KRAS interaction inhibitors BI-3406 or BI-1701963 could be of particular interest. In contrast to mutation-specific inhibitors¹⁷, these pan-KRAS inhibitors do not require a specific mutation and could be used in the therapy of KRAS wild type malignancies^{19,20}. In addition, since the subgroup exhibits a significantly increased sensitivity to MEK inhibitors, a combined therapy, as currently being tested in clinical trials for KRAS mutated tumors, may also offer further advantages for some patients²¹.

Our results provide evidence for the presence of a KRAS-dependent subgroup in KRAS^{wt} tumors. The high number of KRAS-dependent cell lines within the KRAS^{wt} subgroup raises expectations for an equally high proportion of patients with increased response to therapy. In the future, further clinical research efforts are needed. In this regard our study should serve for deeper understanding of mechanisms in panKRAS inhibition and associated current clinical trials.

Identification of responsive and resistant tumors will be an essential task for an optimal therapy of the identified subgroup. We demonstrated that RAEs such as KRAS amplifications or EGFR co-dependencies are not reliable predictors of KRAS dependency in the KRAS^{wt} subgroup. Moreover, RAEs were heterogeneously distributed, and presence of RAE did not necessarily follow a KRAS dependency and vice versa. This observation is corroborated by results from clinical trial data, as the presence of an activating KRAS G12C mutation predicted response to Sotorasib in only 32% of patients¹⁷. We interpreted this as further evidence for the complexity of signal transduction as previously described^{9,49}. To address this complexity, we investigated modeling solutions using whole transcriptome RNA sequencing data as predictors. Recently,

difficulties in RNA-based modeling of gene dependencies of the tumor drivers KRAS, NRAS, and HRAS have been described⁴⁷.

In our analyses we tested modeling using Lasso regression with varying sets of predictors with a focus on variables selected by a literature-based protein-protein interaction network and subsequent gene centrality quantification through a diffusion kernel. This systematic approach with an initial variable reduction to 100–1000 genes significantly improved predictions of KRAS dependency in KRAS^{wt} cell lines when compared to models using all available RNA predictors, RAE-based models or those of previously published expression signatures. Although not the main focus of the depicted study, we also achieved improvements in performance by our approach in the complete cell line set but could not outperform KRAS mutation status.

The importance of variable selection for modeling is emphasized by inferior predictions of models based on genes of previous RAS expression signatures. In the respective initial publication these signatures could provide robust predictions for training and test data^{27,28}. However, in our study the reported performance dropped significantly. This suggests that predictions of expression-based models are only valid for applications with high similarity to the training data and tend to overfitting. Generalizations of specific gene signatures without methodological and algorithmic reference seem to be limited, which underlines the need for standardized procedures for the prediction of therapy response to KRAS inhibitors in tumor patients.

Even with high agreement between the training set data and those of the planned application, steps to reduce overfitting caused by artifacts in the training data are highly relevant to ensure external validity. In this regard, besides the use of statistical methods suitable for high-dimensional datasets collecting additional information about the predictors to select variables with context specific biological relevance can be important before modeling. Literature-based protein-protein interaction networks as used in this work represent one option for this purpose since they enable the integration of preexisting scientific knowledge to identify key regulatory genes in the context of KRAS signaling. These genes possibly represent more robust predictors, which reduce overfitting caused by dataset-specific artifacts. Due to the lack of additional data, we were not able to validate this hypothesis. Nevertheless, the significantly improved internal performance of KRAS dependency models in the KRAS^{wt} subgroup, using predictors selected by the diffusion kernel, indicate the advantages of this strategy. Future efforts with the objective to establish diagnostically applicable models may possibly benefit from similar approaches as well.

In summary, our results suggest that a subset of patients without oncogenic KRAS mutations may benefit from targeted therapy with KRAS inhibitors. In the long run, we assume that machine learning models based on high-dimensional RNA expression data could help with therapeutic decisions. As already mentioned before, a crucial factor for clinical applicability of the proposed models will be a highly standardized test methodology. This includes all steps of the analysis including sample preparation, sequencing and bioinformatic analysis as well as the choice of a suitable parameter for therapy response quantification. The realization of such solutions appears to be possible and reasonable as costs for quantifying gene expression continue to fall.

METHODS

Data

Cancer cell line data were obtained from the website of Dependency Map Consortium including Genomic, RNA-expression, CRISPR, RNAi and CCLL drug sensitivity data (Release 21Q1)^{29–32,34,50}. GDSC drug data (Release 8.2) were downloaded from the project's website⁵¹. For all analyses, only cell lines originating from solid tumors were used.

Cell line classification

Cell line classification was performed in two steps using mutation and CRISPR dependency data of the Achilles project. The investigated entire cell line set was divided into wild type and mutated according to their mutation status in the examined CRISPR knockout gene. All cell lines harboring non-deleterious mutations in KRAS, NRAS, and HRAS were regarded as mutated, reflecting typically activating hotspot mutations. This resulted in six different groups of cell lines (HRAS^{wt}/HRAS^{mut}, KRAS^{wt}/KRAS^{mut}, NRAS^{wt}/NRAS^{mut}). Initial analyses were performed in cell lines with wild type status and for each gene individually (Fig. 1a). In a second step, each group was divided into dependent and independent cell lines by the respective CRISPR gene dependency using the Achilles dependency format. The format indicates the probability of being part of a distribution of essential or non-essential genes for each cell line and each gene (Supplementary Fig. 2). At a threshold of over 50%, cell lines were classified as dependent.

Binary RAE classification

For RAE classification based on CRISPR dependency we selected oncogenes of the oncoKB database with a restriction to receptor kinases and non-RTKs annotated in HGNC^{52–54}. After that we searched for co-dependencies between the selected genes and KRAS using Achilles CRISPR dependency format. As before, cell lines were classified as dependent if dependency exceeded a threshold of 50%. RAE detection based on KRAS amplification status was performed using CCLE copy number data. Cell lines were classified as KRAS amplified if the relative copy number of KRAS exceeded a threshold of 3 compared to the mean copy number of the sample.

Overrepresentation analysis

Overrepresentation analysis was performed using R package ReactomePA⁵⁵. Following differential expression analysis, consideration was given to all statistically significant negative and positive associations with a minimum required level of gene expression higher than the 75th percentile. For CRISPR data, only those with significantly increased dependency in the KRAS-dependent group (Wilcoxon–Mann–Whitney test) were used.

Dependency modeling

CRISPR/RNAi data of Achilles and DRIVE projects (gene effect format [−∞; ∞]) were used as dependent variables in gene dependency modeling. Different predictor types (RAE, mutations, RNA expression) were processed as follows. RAE were represented by Achilles CRISPR gene effect format and KRAS gene copy number. We restricted gene dependency RAE to oncogenes annotated in oncoKB database and receptor kinases/non-RTKs from HGNC^{52–54}, resulting in 43 different markers for RAE-based predictions (Supplementary Data 2). Using a similar approach to Dempster et al., mutations were divided into three categories (deleterious, hotspot, other) based on annotations from DepMap data⁴⁷. The categories hotspot and other were restricted to non-silent, non-deleterious mutations with or without TCGA/COSMIC hotspot classification. Subsequently, a binary predictor variable was created from each of the three categories. In the presence of one or more mutations from one of the categories, the respective cell line was classified as mutated in the corresponding predictor variable, and as non-mutated in the absence of mutations in the category. Gene identifiers of previous expression signatures (Supplementary Data 3) were obtained from the respective publications and were used without further selection steps^{27,28}. To quantify gene expression levels, we used CCLE RNA sequencing data in TPM format.

For our workflow based on whole transcriptome RNA-expression data (TPM) we either used the total number of available RNA predictors or subsets selected by our variable selection approach. For this purpose, we constructed a literature-based protein-protein network to identify significant predictors by gene centrality (Fig. 1b). The hypothesis is that these genes are important regulatory elements in KRAS signaling and thus represent suitable predictors of KRAS dependency. First, protein-protein interactions were downloaded from the STRING database⁵⁶. Gene identifiers were assigned to each protein and duplicated interaction scores between the same genes were averaged. Interactions with a score lower than the 90th percentile were discarded. To focus the network on genes which may be involved in the context of KRAS signaling we restricted genes to those with a direct KRAS interaction (1st shell) and their

respective interaction partners (2nd shell). The final network consisted of 7070 genes. Centrality (closeness) of each gene was determined by a diffusion kernel which captures the information flow within the network as previously described^{57,58}. [Eq. 1]

$$K = e^{\beta H} = I + \beta H + \frac{\beta^2}{2!} H^2 + \frac{\beta^3}{3!} H^3 + \dots \quad (1)$$

K refers to the diffusion kernel, e to Euler's number, H to the negative Laplacian matrix, I to the identity matrix and β to a parameter which controls the degree of information flow. Similar to Lee et al.⁵⁸ we tested several values of the β parameter ($\beta = 0.75^n$, $n = 1, 5, 10, 20, 30$) for variable selection. This resulted in seven different closeness estimates of the genes in our network.

Hyperparameter optimization of the variable selection by the diffusion kernel was performed for the following variables. The minimum level of gene expression was tested for the percentiles P25, P50, P75, and P95. β parameter was tuned for the values mentioned above. Subsequently, the number of selected genes by centrality was tested for 100, 500, and 1000 genes. This approach resulted in a total of 60 different predictor sets for modeling. Independent of the hyperparameter optimization, we also tested modeling either with all unfiltered genes of the network (6967 genes) or with the total number of RNA predictors available (47,768 genes). For all predictor sets gene expressions with missing values were discarded before the final selection step.

Despite our variable selection workflow, the number of predictors for RNA-based models was still high. To prevent overfitting, we consistently used Lasso regression for all models and predictor types. In addition, we also tested Elastic Net and Random Forest regression in KRAS dependency models of the KRAS^{wt} subgroup after variable selection through the diffusion kernel. Lasso and Elastic Net regression minimize the following quantities [Eqs. 2–3]:

$$\beta_{\text{Lasso}} = \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^p |\beta_j| \quad (2)$$

$$\beta_{\text{Elastic net}} = \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^p \beta_j^2 + \lambda \sum_{j=1}^p |\beta_j| \quad (3)$$

n is the number of observations, p the number of different predictors, β_0 the y -intercept, β_j the coefficient of the respective predictor, y_i the values of the independent variable, and x_{ij} the values of the predictors.

For every set of independent variables 100 models were trained, each using 80% of the cell lines randomly selected by R base sample() function. The remaining 20% served as an independent test set. Test set predictions were averaged for each predictor set and correlated with the experimental data (Fig. 1c).

Statistical analyses

All statistical analyses were performed in R programming language⁵⁹. The package glmnet was used for Lasso regression and Elastic Net⁶⁰. Random Forest regression was performed with randomForest package⁶¹. Differential expression analyses were carried out with DESeq2⁶² and network analysis was generated with igraph⁶³. RAWGraphs 2.0 was used for additional graphics⁶⁴. Standard statistical methods are mentioned in figures and legends.

Reporting summary

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

DATA AVAILABILITY

The datasets analyzed in the study are available on the websites of the Dependency Map Consortium (<https://depmap.org/portal/>) and the Genomics of Drug Sensitivity in Cancer Project (<https://www.cancerrxgene.org/>).

CODE AVAILABILITY

All computational analyses were performed in R programming language⁵⁹. The codes used for the other computational analyses are available upon request to the authors.

Received: 9 June 2021; Accepted: 16 December 2021;
Published online: 19 January 2022

REFERENCES

- Winters, I. P. et al. Multiplexed in vivo homology-directed repair and tumor barcoding enables parallel quantification of Kras variant oncogenicity. *Nat. Commun.* **8**, 2053 (2017).
- Kandath, C. et al. Mutational landscape and significance across 12 major cancer types. *Nature* **502**, 333–339 (2013).
- Zehir, A. et al. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat. Med.* **23**, 703–713 (2017).
- Prior, I. A., Hood, F. E. & Hartley, J. L. The Frequency of Ras Mutations in Cancer. *Cancer Res.* **80**, 2969–2974 (2020).
- Santarpia, L., Lippman, S. M. & El-Naggar, A. K. Targeting the MAPK-RAS-RAF signaling pathway in cancer therapy. *Expert Opin. Ther. Targets* **16**, 103–119 (2012).
- Piccari-Gebhart, M. J. et al. Trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer. *N. Engl. J. Med.* **353**, 1659–1672 (2005).
- Ramalingam, S. S. et al. Overall Survival with Osimertinib in Untreated, EGFR-Mutated Advanced NSCLC. *N. Engl. J. Med.* **382**, 41–50 (2020).
- Wolf, J. et al. Capmatinib in MET Exon 14-Mutated or MET-Amplified Non-Small-Cell Lung Cancer. *N. Engl. J. Med.* **383**, 944–957 (2020).
- Wee, P. & Wang, Z. Epidermal Growth Factor Receptor Cell Proliferation Signaling Pathways. *Cancers (Basel)* **9**, 52 (2017).
- De Roock, W. et al. Effects of KRAS, BRAF, NRAS, and PIK3CA mutations on the efficacy of cetuximab plus chemotherapy in chemotherapy-refractory metastatic colorectal cancer: a retrospective consortium analysis. *Lancet Oncol.* **11**, 753–762 (2010).
- Karapetis, C. S. et al. K-ras mutations and benefit from cetuximab in advanced colorectal cancer. *N. Engl. J. Med.* **359**, 1757–1765 (2008).
- Le, X. et al. Landscape of EGFR-Dependent and -Independent Resistance Mechanisms to Osimertinib and Continuation Therapy Beyond Progression in EGFR-Mutant NSCLC. *Clin. Cancer Res.* **24**, 6195–6203 (2018).
- Vojtek, A. B. & Der, C. J. Increasing complexity of the Ras signaling pathway. *J. Biol. Chem.* **273**, 19925–19928 (1998).
- Robert, C. et al. Five-Year Outcomes with Dabrafenib plus Trametinib in Metastatic Melanoma. *N. Engl. J. Med.* **381**, 626–636 (2019).
- Dummer, R. et al. Overall survival in patients with BRAF-mutant melanoma receiving encorafenib plus binimetinib versus vemurafenib or encorafenib (COLUMBUS): a multicentre, open-label, randomised, phase 3 trial. *Lancet Oncol.* **19**, 1315–1327 (2018).
- Dummer, R. et al. Binimetinib versus dacarbazine in patients with advanced NRAS-mutant melanoma (NEMO): a multicentre, open-label, randomised, phase 3 trial. *Lancet Oncol.* **18**, 435–445 (2017).
- Hong, D. S. et al. KRAS(G12C) Inhibition with Sotorasib in Advanced Solid Tumors. *N. Engl. J. Med.* **383**, 1207–1217 (2020).
- Kessler, D. et al. Drugging an undruggable pocket on KRAS. *Proc. Natl. Acad. Sci. USA* **116**, 15823–15829 (2019).
- Hofmann, M. H. et al. BI-3406, a Potent and Selective SOS1-KRAS Interaction Inhibitor, Is Effective in KRAS-Driven Cancers through Combined MEK Inhibition. *Cancer Discov.* **11**, 142–157 (2021).
- Gort, E. et al. A phase I, open-label, dose-escalation trial of BI 1701963 as monotherapy and in combination with trametinib in patients with KRAS mutated advanced or metastatic solid tumors. *J. Clin. Oncol.* **38**, TPS3651–TPS3651 (2020).
- Ingelheim, B. A Study to Test Different Doses of BI 1701963 Alone and Combined With Trametinib in Patients With Different Types of Advanced Cancer (Solid Tumours With KRAS Mutation). <https://clinicaltrials.gov/ct2/show/NCT04111458> (2019).
- Skoulidis, F. et al. Sotorasib for Lung Cancers with KRAS p.G12C Mutation. *N. Engl. J. Med.* **384**, 2371–2381 (2021).
- Paez, J. G. et al. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science* **304**, 1497–1500 (2004).
- Kloth, M. et al. Activating ERBB2/HER2 mutations indicate susceptibility to pan-HER inhibitors in Lynch and Lynch-like colorectal cancer. *Gut* **65**, 1296–1305 (2016).
- Yardley, D. A. et al. Quantitative measurement of HER2 expression in breast cancers: comparison with 'real-world' routine HER2 testing in a multicenter Collaborative Biomarker Study and correlation with overall survival. *Breast Cancer Res.* **17**, 41 (2015).
- Pao, W. et al. Acquired resistance of lung adenocarcinomas to gefitinib or erlotinib is associated with a second mutation in the EGFR kinase domain. *PLoS Med.* **2**, e73 (2005).
- Singh, A. et al. A gene expression signature associated with "K-Ras addiction" reveals regulators of EMT and tumor cell survival. *Cancer Cell* **15**, 489–500 (2009).
- Loboda, A. et al. A gene expression signature of RAS pathway dependence predicts response to PI3K and RAS pathway inhibitors and expands the population of RAS pathway activated tumors. *BMC Med. Genom.* **3**, 26 (2010).
- DepMap, Broad. DepMap 21Q1 Public. figshare <https://doi.org/10.6084/m9.figshare.13681534.v1> (2020).
- Meyers, R. M. et al. Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. *Nat. Genet.* **49**, 1779–1784 (2017).
- Dempster, J. M. et al. Extracting Biological Insights from the Project Achilles Genome-Scale CRISPR Screens in Cancer Cell Lines. Preprint at <https://www.biorxiv.org/content/10.1101/720243v1> (2019).
- McFarland, J. M. et al. Improved estimation of cancer dependencies from large-scale RNAi screens using model-based normalization and data integration. *Nat. Commun.* **9**, 4610 (2018).
- Ingelheim, B. A Study to Test Different Doses of BI 1701963 in Combination With Irinotecan in People With Advanced Bowel Cancer With Kirsten Rat Sarcoma Viral Oncogene Homologue (KRAS) Mutation. <https://clinicaltrials.gov/ct2/show/NCT04627142> (2020).
- Behan, F. M. et al. Prioritization of cancer therapeutic targets using CRISPR-Cas9 screens. *Nature* **568**, 511–516 (2019).
- Wellcome Sanger Institute & Center for Molecular Therapeutics. The Genomics of Drug Sensitivity in Cancer Project. <https://www.cancernxgene.org> (2020).
- Iorio, F. et al. A Landscape of Pharmacogenomic Interactions in Cancer. *Cell* **166**, 740–754 (2016).
- Basu, A. et al. An interactive resource to identify cancer genetic and lineage dependencies targeted by small molecules. *Cell* **154**, 1151–1161 (2013).
- Rees, M. G. et al. Correlating chemical sensitivity and basal gene expression reveals mechanism of action. *Nat. Chem. Biol.* **12**, 109–116 (2016).
- Pirker, R. et al. EGFR expression as a predictor of survival for first-line chemotherapy plus cetuximab in patients with advanced non-small-cell lung cancer: analysis of data from the phase 3 FLEX study. *Lancet Oncol.* **13**, 33–42 (2012).
- Essakly, A. et al. PIK3CA and KRAS Amplification in Esophageal Adenocarcinoma and their Impact on the Inflammatory Tumor Microenvironment and Prognosis. *Transl. Oncol.* **13**, 157–164 (2020).
- Valtorta, E. et al. KRAS gene amplification in colorectal cancer and impact on response to EGFR-targeted therapy. *Int. J. Cancer* **133**, 1259–1265 (2013).
- Heigener, D. F. et al. Afatinib in Non-Small Cell Lung Cancer Harboring Uncommon EGFR Mutations Pretreated With Reversible EGFR Inhibitors. *Oncologist* **20**, 1167–1174 (2015).
- Wu, J. Y. et al. Effectiveness of tyrosine kinase inhibitors on "uncommon" epidermal growth factor receptor mutations of unknown clinical significance in non-small cell lung cancer. *Clin. Cancer Res.* **17**, 3812–3821 (2011).
- Tu, H. Y. et al. A comprehensive review of uncommon EGFR mutations in patients with non-small cell lung cancer. *Lung Cancer* **114**, 96–102 (2017).
- Birkeland, E. et al. KRAS gene amplification and overexpression but not mutation associates with aggressive and metastatic endometrial cancer. *Br. J. Cancer* **107**, 1997–2004 (2012).
- Wagner, P. L. et al. Frequency and clinicopathologic correlates of KRAS amplification in non-small cell lung carcinoma. *Lung Cancer* **74**, 118–123 (2011).
- Dempster, J. M. et al. Gene expression has more power for predicting in vitro cancer cell vulnerabilities than genomics. Preprint at <https://www.biorxiv.org/content/10.1101/2020.02.21.959627v3> (2020).
- Klebanov, L. & Yakovlev, A. Diverse correlation structures in gene expression data and their utility in improving statistical inference. *Ann. Appl. Stat.* **1**, 538–559 (2007).
- Coyle, S. M. & Lim, W. A. Mapping the functional versatility and fragility of Ras GTPase signaling circuits through in vitro network reconstitution. *Elife* **5**, e12435 (2016).
- Ghandi, M. et al. Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature* **569**, 503–508 (2019).
- Yang, W. et al. Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res.* **41**, D955–D961 (2013).
- HUGO Gene Nomenclature Committee (HGNC), European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus. HGNC Database. www.genenames.org (2020).
- Braschi, B. et al. Genenames.org: the HGNC and VGNC resources in 2019. *Nucleic Acids Res.* **47**, D786–D792 (2019).
- Chakravarty, D. et al. OncoKB: A Precision Oncology Knowledge Base. *JCO Precis Oncol.* **1**, 1–16 (2017).
- Yu, G. & He, Q. Y. ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. *Mol. Biosyst.* **12**, 477–479 (2016).
- Szklarczyk, D. et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–D613 (2019).

57. Kondor, R. I. & Lafferty, J. Diffusion kernels on graphs and other discrete structures. *In Proc. 19th International Conference on Machine Learning*, 315–322 (2002).
58. Lee, J. K. et al. Pharmacogenomic landscape of patient-derived tumor cells informs precision oncology therapy. *Nat. Genet.* **50**, 1399–1411 (2018).
59. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/> (2020).
60. Friedman, J., Hastie, T. & Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J. Stat. Softw.* **33**, 1–22 (2010).
61. Breiman, L. Random Forests. *Mach. Learn.* **45**, 5–32 (2001).
62. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
63. Csardi, G. & Nepusz, T. The igraph software package for complex network research. *Inter Journal Complex Syst.* **1695**, 1–9 (2006).
64. Mauri, M., Elli, T., Caviglia, G., Ubaldi, G. & Azzi, M. RAWGraphs: A Visualisation Platform to Create Open Outputs. *In Proceedings of the 12th Biannual Conference on Italian SIGCHI Chapter*. 28:1–28:5 (Association for Computing Machinery, New York, NY, USA; 2017).

ACKNOWLEDGEMENTS

M.K. received grant support from the Else-Kröner-Forschungskolleg.

AUTHOR CONTRIBUTIONS

B.U. and M.K. conceived the project, performed the data analyses and implemented the models. B.U., M.K., M.O., R.B., and W.R. were involved in the result interpretation, drafting the content and structure of the paper. All authors approved the final version.

FUNDING

Open Access funding enabled and organized by Projekt DEAL.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41540-021-00211-8>.

Correspondence and requests for materials should be addressed to Bastian Ulmer.

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

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



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

© The Author(s) 2022

4. Diskussion

Der RAS-Signalweg besitzt über verschiedene Mechanismen eine zentrale Rolle in der Onkogenese zahlreicher Tumorentitäten⁸⁻¹¹. Insbesondere die RAS-GTPasen selbst sind dabei von entscheidender Bedeutung in der Signaltransduktion von der Zellmembran zum Nukleus¹². Zuletzt wurden mit den panKRAS- und mutationsspezifischen KRAS-Inhibitoren bedeutende Fortschritte in der Entwicklung zielgerichteter Therapien erreicht¹³⁻¹⁶. In der vorliegenden Arbeit wurden Zelllinien solider Tumoren mit KRAS Wildtyp Sequenz auf die Abhängigkeit des Zellwachstums von einer KRAS-Expression, basierend auf Daten aus CRISPR-knockout und RNAi-*knockdown* Experimenten, untersucht. Innerhalb dieser KRAS^{wt} Zelllinien konnte eine KRAS-abhängige Subgruppe identifiziert werden. Die beschriebene KRAS-Abhängigkeit konnte anschließend durch den kombinierten Einsatz von RNA-Expressionsdaten, einer Netzwerk-basierten Variablenselektion sowie von Algorithmen des maschinellen Lernens erfolgreich vorhergesagt werden. Die vorgestellten Ergebnisse deuten darauf hin, dass eine therapeutische Anwendung der neu entwickelten panKRAS-Inhibitoren auch für Tumore ohne aktivierende KRAS-Mutationen geeignet sein könnte. Unsere Daten sprechen somit für eine verbesserte Prädiktion des Therapieansprechens durch RNA-basierte Modelle unter Nutzung von maschinellem Lernen.

4.1. KRAS-Abhängigkeit in KRAS^{wt} Zelllinien

Innerhalb des RAS-Signalwegs lassen sich zahlreiche Onkogene identifizieren, die durch RAEs aktiviert werden können und daher zu einer gesteigerten Signaltransduktion beitragen^{27,28}. Die RAS-GTPasen stellen dabei zentrale Knotenpunkte in der Signalintegration vieler membranständiger RTKs wie EGFR, ERBB2 und ALK dar¹². Bei einer pathologischen Überaktivierung der RTKs durch ein oder mehrere RAEs kann die Signaltransduktion zum Zellkern entsprechend auch über die RAS-Proteine stattfinden. Klinisch sind diese Prozesse insbesondere bei Vorliegen von aktivierenden RAS-Mutationen relevant, da die Tumoren resistent gegen eine Therapie mit EGFR-spezifischen Antikörpern sind⁵². Aufgrund ihrer zentralen Rolle in der Signalkaskade stellen die RAS-Proteine jedoch auch ohne aktivierende Mutation ein attraktives Therapieziel dar.

Basierend auf der biologischen Rationalen einer RAS-Abhängigkeit in RAS-Wildtyp-Tumoren haben wir für jede RAS-Isoform die Abhängigkeit von einer Expression in Tumor-Zelllinien analysiert. Hierfür haben wir Daten aus genomweiten CRISPR-*knockout* Experimenten des Achilles Projects von über 700 Zelllinien unterschiedlicher Entität verwendet¹¹²⁻¹¹⁷. Separat für jede der drei Isoformen haben wir zunächst die Zelllinien mit Wildtyp Sequenz selektioniert und anschließend anhand der CRISPR Daten in abhängig und nicht-abhängig klassifiziert. Dabei konnten wir für KRAS mit 34% sowohl relativ als auch absolut die höchste Anzahl

abhängiger Zelllinien feststellen. Für NRAS und HRAS ergaben sich deutlich kleinere Anteile. Eine Co-Abhängigkeit, also das gleichzeitige Vorliegen einer Abhängigkeit in mehreren RAS-Isoformen, wurde nur in wenigen Fällen vorgefunden. Die Ergebnisse zeigen somit, dass innerhalb der Wildtyp Zelllinien KRAS die größte Bedeutung für die Signaltransduktion besitzt. Dieses Verhalten deckt sich mit den Inzidenzen RAS mutierter Tumoren, in denen KRAS Mutationen ebenfalls am häufigsten auftreten ¹¹. Die geringe Anzahl von Co-Abhängigkeiten deutet außerdem an, dass es sich um distinkte Aktivierungsmechanismen der Isoformen handelt.

Des Weiteren konnten wir durch die Analysen von Chemosensitivitätsdaten der Zelllinien eine Evidenz für eine therapeutische Relevanz der KRAS-abhängigen Subgruppe ableiten. Hierbei zeigten die KRAS abhängigen Zelllinien im Vergleich zu den unabhängigen eine höhere Sensitivität gegenüber zahlreichen MEK-Inhibitoren. Folglich wäre eine Therapie mit MEK-Inhibitoren, die sich bereits in klinischer Anwendung befindet, auch in Tumoren ohne aktivierende KRAS-Mutationen denkbar. Außerdem stehen mit den kürzlich entwickelten panKRAS/SOS1 Inhibitoren BI-1701963 und BI-3406 in Zukunft möglicherweise weitere Therapieoptionen mit gezielter KRAS-Inhibierung zur Verfügung, die in Kombination mit einer MEK-Inhibition genutzt werden können ¹⁴⁻¹⁶. Im Gegensatz zu den mutationsspezifischen KRAS-Inhibitoren hemmen diese auch die Aktivität des KRAS-Wildtyp Proteins und sind dementsprechend in der von uns ausgearbeiteten Subgruppe der KRAS-abhängigen Wildtyp Tumoren potenziell anwendbar. Des Weiteren ist eine Kombination mit MEK-Inhibitoren, wie sie zurzeit für BI-1701963 in KRAS mutierten Tumoren getestet wird, denkbar ¹⁵. Bei den aus unserer Arbeit abgeleiteten Therapieansätzen ist zu berücksichtigen, dass es sich um Ergebnisse aus präklinischen Experimenten handelt. Für eine perspektivische Etablierung einer antineoplastischen Therapie bedarf es weiterer translationaler Forschung an den panRAS Inhibitoren und klinische Studien.

4.2. Prädiktion der KRAS-Abhängigkeit

Nach erfolgreicher Identifizierung einer pharmakologisch angreifbaren KRAS-abhängigen Subgruppe in KRAS^{wt} Zelllinien, stellt sich die Frage nach einer geeigneten Prädiktion der responsiven Tumoren.

In ersten Ansätzen sind wir der Hypothese gefolgt, dass die KRAS Abhängigkeit der Wildtyp-Subgruppe durch RAEs in membranständigen Rezeptorkinasen entsteht, wie es für zahlreiche Beispiele in der Literatur beschrieben wurde ¹². Grundlage dafür ist allerdings die Annahme einer linearen Signaltransduktion, welche jedoch hinsichtlich einer Modellierung nicht immer zulässig ist. Hierbei ist insbesondere zu berücksichtigen, dass die Signalwege über ein komplexes Netzwerk aus unzähligen Interaktionspartnern und Feedback Mechanismen funktionieren ¹¹⁸. Ein Beispiel für die direkte klinische Relevanz lässt sich in der Behandlung

des BRAF mutierten kolorektalen Karzinoms finden. Hier ist eine Therapie mit BRAF-Inhibitoren wie Encorafenib deutlich effektiver bei gleichzeitiger Gabe von Cetuximab, da ansonsten Resistenzen über einen EGFR vermittelten Feedback-Mechanismus entstehen ⁷⁸. Des Weiteren ist bei vielen genomischen Ereignissen, wie im Fall von EGFR-Mutationen, das Potential einer RAS-aktivierenden Wirkung nicht immer eindeutig geklärt ¹¹⁹. Zur Vorhersage der Responsivität haben wir uns für den Versuch einer *Machine Learning*-basierten Modellierung entschieden, da RAEs der ausgewählten Onkogene teils als klinischer Standard in der Beurteilung einer RAS-Aktivierung verwendet werden und somit eine klinische Referenz besteht. In den Modellen haben wir neben genomischen Ereignissen die Genabhängigkeiten der Kinasen aus den CRISPR-Daten als Indikator für RAEs verwendet. Hierbei besteht die Annahme, dass die berücksichtigten Daten ein geeigneteres Maß für das RAS-aktivierende Potential der Onkogene sein könnten und somit auch zu besseren Vorhersagen beitragen. Durch dieses Vorgehen konnten wir insgesamt 67% der KRAS abhängigen Ziellinien ein RAE zuordnen. Die fehlenden 33% zeigen, dass in dieser klinisch orientierten Auswahl an Onkogenen offenbar nicht alle relevanten RAEs berücksichtigt wurden. Grundsätzlich ist anzunehmen, dass zusätzliche Prädiktoren die Sensitivität verbessern könnten. Insbesondere wurden keine Tumorsuppressor-Gene berücksichtigt, da diese nicht in eine auf CRISPR Daten basierte Modellierung integriert werden können. Ein knockout von Tumorsuppressor-Genen würde im Vergleich zum antiproliferativen Effekt bei Onkogenen eher einen proliferativen oder keinen Effekt zur Folge haben. Des Weiteren fiel nach zusätzlicher Untersuchung der KRAS unabhängigen Zelllinien auf, dass mit der bisherigen Prädiktorauswahl auch in dieser Gruppe 63% der Zelllinien ein RAE aufweisen. Zusätzliche Prädiktoren würden dieses Problem nicht lösen und es womöglich sogar negativ beeinflussen. Um komplexere Mechanismen der Aktivierung nicht zu vernachlässigen, haben wir als Versuch einer verhältnismäßig einfachen RAE-basierten Prädiktion eine quantitative Modellierung durch Regressionsverfahren getestet. Auch hier konnten wir keine zufriedenstellenden Ergebnisse erreichen. Eine intuitiv erscheinende Prädiktion der KRAS-Abhängigkeit auf Basis von *Upstream*-gelegenen RAEs ist demnach nicht einfach möglich. Die initialen Versuche zeigen, dass der zu vorhersagende Phänotyp komplexer zu sein scheint, als dass man ihn auf einige wenige Onkogene mit heterogen verteilten RAEs reduzieren kann. Unterstützt werden diese Beobachtungen durch Studienergebnisse des KRAS G12C Inhibitors Sotorasib. Trotz des Vorliegens der spezifischen Mutation wurde lediglich bei 37% der Patienten eine Reduktion der Tumormasse erreicht ¹³. Entsprechend sind die Einflüsse auf den RAS-Signalweg selbst bei Vorliegen eines starken onkogenen Treibers so vielfältig, dass selbst eine spezifische Therapie nicht sicher durch eine einfache Berücksichtigung von RAEs vorhergesagt werden kann. Durch unsere bisherigen Ergebnisse konnten wir die vorbeschriebene Komplexität des Signalwegs weiter objektivieren und zeigen, dass RAEs in bekannten Onkogenen nicht für

eine Prädiktion der KRAS-Abhängigkeit in Zelllinien mit KRAS Wildtyp Gensequenz geeignet sind. Insbesondere lässt sich ableiten, dass die KRAS-Abhängigkeit womöglich besser durch Daten erfasst werden könnte, die den Phänotyp des Tumors umfassender wiedergeben. Um die Qualität der Modellierung zu verbessern, haben wir deshalb verschiedene Ansätze basierend auf Expressionsdaten getestet. Dempster et al. konnten diesbezüglich nachweisen, dass die Abhängigkeit der Tumorzelllinien von zahlreichen Onkogenen basierend auf RNA-Expressionsdaten erfolgreich modelliert werden kann⁹⁵. Im Fall der RAS-Proteine zeigten sich jedoch Schwierigkeiten. Für unsere Modelle haben wir die RNA-Prädiktoren zunächst aus zuvor veröffentlichten Expressionssignaturen ausgewählt, die einen RAS-Aktivitätsstatus wiedergeben sollen. Anschließend haben wir basierend auf diesen Prädiktoren mittels *Lasso* Regressionsmodelle der KRAS-Abhängigkeit erstellt. Obwohl beide Signaturen in den initialen Publikationen gute Ergebnisse lieferten, konnten wir dies für unsere Modelle nicht wiederholen^{120,121}. Eine mögliche Ursache könnte darin liegen, dass die Signaturen sowohl anhand einer Kombination aus KRAS-Wildtyp als auch mutierten Zelllinien erstellt worden sind und somit nicht derselbe Phänotyp modelliert wurde. Des Weiteren könnte eine zu geringe externe Validität der Modelle vorliegen. Entsprechend würden die verwendeten Variablen zwar gute Vorhersagen für die Daten der initialen Publikationen ergeben, jedoch ohne für eine Verallgemeinerung auf weitere unabhängige Datensätze geeignet zu sein. Zusammengefasst sahen wir methodische Unterschiede in der Quantifizierung von RAS-Aktivität und Prädiktoren sowie eine nicht geeignete Variablenselektion als mögliche Ursachen für die schlechten Vorhersagen der in der Literatur beschriebenen Modellierung an.

4.3. Diffusionskernel-basierte Prädiktion der KRAS-Abhängigkeit

Die initialen Modellierungsansätze basierend auf RNA-Expressionssignaturen konnten ebenfalls keine zufriedenstellende Vorhersage der KRAS-Abhängigkeit liefern. In unseren weiteren Versuchen erfolgte eine fokussierte Variablenselektion, die wiederum auf eine Verbesserung der Prädiktion durch geeignetere Prädiktoren abzielt.

Deshalb haben wir ein an Lee et al. orientiertes eigenes Vorgehen, bestehend aus einer Protein-Interaktionsnetzwerk-basierten Variablenselektion und anschließender *Lasso*-Regression, implementiert¹⁰¹. Die Variablenselektion besteht hierbei aus einer Quantifizierung der Zentralität eines literatur-basierten Protein-Protein Interaktionsnetzwerks durch einen Diffusionskernel. Dadurch war es uns möglich, zusätzliche Informationen über den Signalweg aus bestehendem Wissen in die Modellierung zu integrieren und somit Gene mit zentralen regulatorischen Eigenschaften in der Umgebung von KRAS zu identifizieren. Zwar wäre eine Variablenselektion aufgrund der Verwendung einer *Lasso* Regression nicht zwingend notwendig gewesen, jedoch nahmen wir an, dass dieser Schritt die Störsignale in den Daten reduzieren würde. Die Anzahl an Variablen mit zufälligen Assoziationen im *Training Set* aber

ohne wirkliche Relevanz wäre somit verringert, wodurch die Vorhersagen der internen *Test Sets* verbessert werden sollten. Gleichzeitig bestand das Ziel durch eine partielle Abkopplung der Variablenselektion von den Trainingsdaten eine Möglichkeit zur Verbesserung der externen Validität zu ermöglichen.

Basierend auf diesem Vorgehen konnten wir für die KRAS^{wt} Gruppe erfolgreich Modelle der KRAS-Abhängigkeit erstellen. Verglichen mit den vorherigen RAS-Expressionssignaturen zeigte sich eine deutliche Verbesserung der Vorhersagen, was die allgemeine Bedeutung der Variablenselektion betont. Mithilfe des von uns verwendeten funktionellen Protein-Interaktionsnetzwerks und anschließender Zentralitätsanalyse durch einen Diffusionskernel konnten wir Gene identifizieren, die sowohl für CRISPR als auch für RNAi Daten als robuste Prädiktoren verwendet werden können. Die Ergebnisse deuten darauf hin, dass dieses literaturbasierte Vorgehen in der Variablenselektion potenziell geeignet ist um Gene zu identifizieren, die im Kontext einer RAS-Aktivierung reguliert werden. Somit könnten insbesondere in $p \gg n$ Situationen (deutlich mehr Prädiktoren als Untersuchungsobjekte) Prädiktoren mit zufälligen Assoziationen zur abhängigen Variablen, wie sie möglicherweise in den Expressionssignaturen der früheren Publikationen zu finden sind, verworfen werden. Unterstützt wird diese Vermutung durch die Ergebnisse einer Wiederholung der Modellierung ohne die durch Diffusionskernel und *Lasso*-Regression selektionierten Gene. Hierbei konnten wir eine Abnahme der Qualität der Vorhersagen von bis zu 29% feststellen, was auf einen nicht kompensierbaren Informationsverlust hindeutet und somit für die Spezifität der Prädiktoren spricht. Damit stellen unsere Ergebnisse weitere Hinweise auf das Vorliegen eines KRAS-abhängigen Phänotyps in der KRAS^{wt} Gruppe dar.

Trotz der deutlichen Verbesserungen konnten wir bei genauerer Untersuchung der Modelle eine größere Abweichung der Vorhersagen für stark KRAS-abhängige Zelllinien feststellen. Wir nahmen an, dass dies möglicherweise durch ein unausgeglichenes Verhältnis zwischen KRAS-abhängigen und nicht-abhängigen Zelllinien bedingt ist, wobei die nicht-abhängigen Zelllinien überwiegen. Um die Vorhersagen für die stark KRAS-abhängigen Zelllinien zu verbessern, haben wir deshalb eine Modellierung mittels *Elastic Net*- sowie der *Random Forest*-Regression getestet, die bereits zuvor erfolgreich in der Modellierung zahlreicher verwandter Fragestellungen eingesetzt wurden^{95,96,101}. Hier konnten wir jedoch mit beiden Verfahren keine weitere Verbesserung der Vorhersagen erreichen. Bei vergleichbarer Qualität der Vorhersagen von *Lasso*- und *Elastic Net*-Regression waren die Modelle der *Random Forest*-Regression interessanterweise denen der anderen Algorithmen sogar deutlich unterlegen. Dieses Verhalten muss als unerwartet angesehen werden, da, wie von James et al. erörtert, die *Random Forest*-Regression als nicht lineares Verfahren potenziell besser für die Modellierung von Daten mit schiefer Datenverteilung geeignet sein sollte⁹⁸. Folglich zeigen die Ergebnisse, dass die Wahl eines geeigneten Algorithmus für die Modellierung unserer

Fragestellung empirisch durch eine Testung verschiedener Optionen kontrolliert werden muss. Auch in unserer Studie ist es nicht auszuschließen, dass andere Algorithmen wie *Gradient Boosting Machines* besser für die Vorhersage der Abhängigkeit von einer KRAS-Expression geeignet sind. Unsere Experimente waren jedoch primär darauf ausgerichtet, ob eine Modellierung der KRAS-Abhängigkeit mit Prädiktoren aus molekularbiologischen Daten in der KRAS^{wt} Subgruppe möglich ist und eine solche Modellierung von einer Diffusionskernel-basierten Variablenselektion profitiert. Die Optimierung der Qualität der Vorhersagen (*Hyperparameter-Tuning*) war eher zweitrangig, da die in vitro basierten Modelle keine validen Vorhersagen über ein Therapieansprechen von panKRAS Inhibitoren bei Patienten zulassen. Wie zuvor gilt auch hier, dass die Methoden im Rahmen von klinischen Studien weiter evaluiert werden müssen. Die hier dargestellte Vorgehensweise kann entsprechend in diagnostisch anwendbare Modelle standardisiert integriert und unter Nutzung von Patienten-abhängigen Daten trainiert werden.

Neben der zuvor beschriebenen Vorhersage der KRAS-Abhängigkeit in Tumor-Zelllinien mit KRAS Wildtyp Sequenz, haben wir ergänzend unser Vorgehen erfolgreich auf KRAS-mutierte Zelllinien übertragen können. Durch die Variablenselektion mittels Diffusionskernel war hierbei ebenfalls eine Verbesserung der RNA-basierten Vorhersagen möglich. Gleichzeitig konnten wir die Anzahl der initialen Prädiktoren auf wenige hundert reduzieren. Die Modelle basierend auf KRAS-Mutationsdaten wurden jedoch nicht übertroffen. Allerdings besitzen beide Prädiktortypen unterschiedliche Vor- und Nachteile. Entsprechend lässt sich folgern, dass umfassende Modelle einer KRAS-Abhängigkeit möglicherweise von einer Kombination von Mutations- und Expressionsdaten sowie weiterer Prädiktoren profitieren können.

Zusammengefasst konnten wir in dieser Arbeit zeigen, dass Patienten ohne aktivierende KRAS-Mutationen von einer Therapie mit den neu entwickelten panKRAS und MEK Inhibitoren profitieren könnten. RNA-expressionsbasierte *Machine Learning*-Modelle könnten dabei zukünftig die Grundlage für eine verbesserte Identifizierung der responsiven Tumoren sein. Für die Entwicklung von klinisch anwendbaren Modellen bedarf es allerdings eines hoch standardisierten methodischen Vorgehens. Dies beinhaltet sämtliche Schritte der Analysen von der molekularbiologischen Aufarbeitung der Proben bis hin zu der bioinformatischen Prozessierung und Interpretation der Daten. Dabei ist bei der Modellierung von Beginn an auch auf die Auswahl einer geeigneten Outcome Variable und robusten Prädiktoren zu achten, die unter "*Real World*" Bedingungen adäquat erfasst werden können. Das Potential der *Machine Learning*-Modelle für eine Verbesserung der Versorgung von onkologischen Patienten ist vielversprechend. Eine Etablierung als klinisches Testverfahren erscheint nicht zuletzt aufgrund der fallenden Sequenzierungskosten zukünftig realisierbar zu sein.

5. Literaturverzeichnis

1. Mulcahy LS, Smith MR, Stacey DW. Requirement for ras proto-oncogene function during serum-stimulated growth of NIH 3T3 cells. *Nature* 1985; **313**(5999): 241-3.
2. Smith MR, DeGudicibus SJ, Stacey DW. Requirement for c-ras proteins during viral oncogene transformation. *Nature* 1986; **320**(6062): 540-3.
3. Hagag N, Halegoua S, Viola M. Inhibition of growth factor-induced differentiation of PC12 cells by microinjection of antibody to ras p21. *Nature* 1986; **319**(6055): 680-2.
4. Thomas SM, Demarco M, Darcangelo G, Halegoua S, Brugge JS. Ras Is Essential for Nerve Growth Factor-Induced and Phorbol Ester-Induced Tyrosine Phosphorylation of Map Kinases. *Cell* 1992; **68**(6): 1031-40.
5. Wood KW, Sarnecki C, Roberts TM, Blenis J. ras mediates nerve growth factor receptor modulation of three signal-transducing protein kinases: MAP kinase, Raf-1, and RSK. *Cell* 1992; **68**(6): 1041-50.
6. de Vries-Smiths AM, Burgering BM, Leervers SJ, Marshall CJ, Bos JL. Involvement of p21ras in activation of extracellular signal-regulated kinase 2. *Nature* 1992; **357**(6379): 602-4.
7. Li N, Batzer A, Daly R, et al. Guanine-nucleotide-releasing factor hSos1 binds to Grb2 and links receptor tyrosine kinases to Ras signalling. *Nature* 1993; **363**(6424): 85-8.
8. Winters IP, Chiou SH, Paulk NK, et al. Multiplexed in vivo homology-directed repair and tumor barcoding enables parallel quantification of Kras variant oncogenicity. *Nat Commun* 2017; **8**(1): 2053.
9. Kandoth C, McLellan MD, Vandin F, et al. Mutational landscape and significance across 12 major cancer types. *Nature* 2013; **502**(7471): 333-9.
10. Zehir A, Benayed R, Shah RH, et al. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat Med* 2017; **23**(6): 703-13.
11. Prior IA, Hood FE, Hartley JL. The Frequency of Ras Mutations in Cancer. *Cancer Res* 2020; **80**(14): 2969-74.
12. Gimble RC, Wang X. RAS: Striking at the Core of the Oncogenic Circuitry. *Front Oncol* 2019; **9**: 965.
13. Skoulidis F, Li BT, Dy GK, et al. Sotorasib for Lung Cancers with KRAS p.G12C Mutation. *N Engl J Med* 2021; **384**(25): 2371-81.
14. Hofmann MH, Gmachl M, Ramharter J, et al. BI-3406, a Potent and Selective SOS1-KRAS Interaction Inhibitor, Is Effective in KRAS-Driven Cancers through Combined MEK Inhibition. *Cancer Discov* 2021; **11**(1): 142-57.
15. Boehringer Ingelheim. A Study to Test Different Doses of BI 1701963 Alone and Combined With Trametinib in Patients With Different Types of Advanced Cancer (Solid Tumours With KRAS Mutation). <https://clinicaltrials.gov/ct2/show/NCT04111458> (Zuletzt abgerufen am 08.08.2021).
16. Boehringer Ingelheim. A Study to Test Different Doses of BI 1701963 in Combination With Irinotecan in People With Advanced Bowel Cancer With Kirsten Rat Sarcoma Viral Oncogene Homologue (KRAS) Mutation. <https://clinicaltrials.gov/ct2/show/NCT04627142> (Zuletzt abgerufen am 08.08.2021).
17. Gibbs JB, Sigal IS, Poe M, Scolnick EM. Intrinsic GTPase activity distinguishes normal and oncogenic ras p21 molecules. *Proc Natl Acad Sci U S A* 1984; **81**(18): 5704-8.
18. Rajalingam K, Schreck R, Rapp UR, Albert S. Ras oncogenes and their downstream targets. *Biochim Biophys Acta* 2007; **1773**(8): 1177-95.
19. Lemmon MA, Schlessinger J. Cell signaling by receptor tyrosine kinases. *Cell* 2010; **141**(7): 1117-34.
20. Schlessinger J. Cell signaling by receptor tyrosine kinases. *Cell* 2000; **103**(2): 211-25.
21. Lowenstein EJ, Daly RJ, Batzer AG, et al. The SH2 and SH3 domain-containing protein GRB2 links receptor tyrosine kinases to ras signaling. *Cell* 1992; **70**(3): 431-42.
22. McCormick F. Signal transduction. How receptors turn Ras on. *Nature* 1993; **363**(6424): 15-6.
23. Moodie SA, Willumsen BM, Weber MJ, Wolfman A. Complexes of Ras.GTP with Raf-1 and mitogen-activated protein kinase kinase. *Science* 1993; **260**(5114): 1658-61.

24. Rodriguez-Viciano P, Warne PH, Dhand R, et al. Phosphatidylinositol-3-OH kinase as a direct target of Ras. *Nature* 1994; **370**(6490): 527-32.
25. Guo YJ, Pan WW, Liu SB, Shen ZF, Xu Y, Hu LL. ERK/MAPK signalling pathway and tumorigenesis. *Exp Ther Med* 2020; **19**(3): 1997-2007.
26. Castellano E, Downward J. RAS Interaction with PI3K: More Than Just Another Effector Pathway. *Genes Cancer* 2011; **2**(3): 261-74.
27. Santarpia L, Lippman SM, El-Naggar AK. Targeting the MAPK-RAS-RAF signaling pathway in cancer therapy. *Expert Opin Ther Targets* 2012; **16**(1): 103-19.
28. Du Z, Lovly CM. Mechanisms of receptor tyrosine kinase activation in cancer. *Mol Cancer* 2018; **17**(1): 58.
29. Stransky N, Cerami E, Schalm S, Kim JL, Lengauer C. The landscape of kinase fusions in cancer. *Nat Commun* 2014; **5**: 4846.
30. Pao W, Miller VA, Politi KA, et al. Acquired resistance of lung adenocarcinomas to gefitinib or erlotinib is associated with a second mutation in the EGFR kinase domain. *PLoS Med* 2005; **2**(3): e73.
31. Roskoski R, Jr. Anaplastic lymphoma kinase (ALK): structure, oncogenic activation, and pharmacological inhibition. *Pharmacol Res* 2013; **68**(1): 68-94.
32. Slamon DJ, Godolphin W, Jones LA, et al. Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer. *Science* 1989; **244**(4905): 707-12.
33. Miller CT, Lin L, Casper AM, et al. Genomic amplification of MET with boundaries within fragile site FRA7G and upregulation of MET pathways in esophageal adenocarcinoma. *Oncogene* 2006; **25**(3): 409-18.
34. Hara T, Ooi A, Kobayashi M, Mai M, Yanagihara K, Nakanishi I. Amplification of c-myc, K-sam, and c-met in gastric cancers: detection by fluorescence in situ hybridization. *Lab Invest* 1998; **78**(9): 1143-53.
35. Fiala O, Pesek M, Finek J, et al. Epidermal Growth Factor Receptor Gene Amplification in Patients with Advanced-stage NSCLC. *Anticancer Res* 2016; **36**(1): 455-60.
36. Prior IA, Lewis PD, Mattos C. A comprehensive survey of Ras mutations in cancer. *Cancer Res* 2012; **72**(10): 2457-67.
37. Vetter IR, Wittinghofer A. The guanine nucleotide-binding switch in three dimensions. *Science* 2001; **294**(5545): 1299-304.
38. Cox AD, Der CJ. Ras history: The saga continues. *Small GTPases* 2010; **1**(1): 2-27.
39. Goody RS, Pai EF, Schlichting I, et al. Studies on the structure and mechanism of H-ras p21. *Philos Trans R Soc Lond B Biol Sci* 1992; **336**(1276): 3-10; discussion -1.
40. Yuan J, Ng WH, Tian Z, et al. Activating mutations in MEK1 enhance homodimerization and promote tumorigenesis. *Sci Signal* 2018; **11**(554).
41. Nikolaev SI, Rimoldi D, Iseli C, et al. Exome sequencing identifies recurrent somatic MAP2K1 and MAP2K2 mutations in melanoma. *Nat Genet* 2011; **44**(2): 133-9.
42. Jaiswal BS, Durinck S, Stawiski EW, et al. ERK Mutations and Amplification Confer Resistance to ERK-Inhibitor Therapy. *Clin Cancer Res* 2018; **24**(16): 4044-55.
43. Davies H, Bignell GR, Cox C, et al. Mutations of the BRAF gene in human cancer. *Nature* 2002; **417**(6892): 949-54.
44. Heinzerling L, Kuhnappel S, Meckbach D, et al. Rare BRAF mutations in melanoma patients: implications for molecular testing in clinical practice. *Br J Cancer* 2013; **108**(10): 2164-71.
45. Menzer C, Menzies AM, Carlino MS, et al. Targeted Therapy in Advanced Melanoma With Rare BRAF Mutations. *J Clin Oncol* 2019; **37**(33): 3142-51.
46. Piccart-Gebhart MJ, Procter M, Leyland-Jones B, et al. Trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer. *N Engl J Med* 2005; **353**(16): 1659-72.
47. Wolf J, Seto T, Han JY, et al. Capmatinib in MET Exon 14-Mutated or MET-Amplified Non-Small-Cell Lung Cancer. *N Engl J Med* 2020; **383**(10): 944-57.
48. Shaw AT, Kim DW, Nakagawa K, et al. Crizotinib versus chemotherapy in advanced ALK-positive lung cancer. *N Engl J Med* 2013; **368**(25): 2385-94.
49. Shaw AT, Ou SH, Bang YJ, et al. Crizotinib in ROS1-rearranged non-small-cell lung cancer. *N Engl J Med* 2014; **371**(21): 1963-71.

50. Ramalingam SS, Vansteenkiste J, Planchard D, et al. Overall Survival with Osimertinib in Untreated, EGFR-Mutated Advanced NSCLC. *N Engl J Med* 2020; **382**(1): 41-50.
51. Wee P, Wang Z. Epidermal Growth Factor Receptor Cell Proliferation Signaling Pathways. *Cancers (Basel)* 2017; **9**(5).
52. De Roock W, Claes B, Bernasconi D, et al. Effects of KRAS, BRAF, NRAS, and PIK3CA mutations on the efficacy of cetuximab plus chemotherapy in chemotherapy-refractory metastatic colorectal cancer: a retrospective consortium analysis. *Lancet Oncol* 2010; **11**(8): 753-62.
53. Rampias T, Giagini A, Siolos S, et al. RAS/PI3K crosstalk and cetuximab resistance in head and neck squamous cell carcinoma. *Clin Cancer Res* 2014; **20**(11): 2933-46.
54. Misale S, Yaeger R, Hobor S, et al. Emergence of KRAS mutations and acquired resistance to anti-EGFR therapy in colorectal cancer. *Nature* 2012; **486**(7404): 532-6.
55. Maemondo M, Inoue A, Kobayashi K, et al. Gefitinib or chemotherapy for non-small-cell lung cancer with mutated EGFR. *N Engl J Med* 2010; **362**(25): 2380-8.
56. Le T, Gerber DE. Newer-Generation EGFR Inhibitors in Lung Cancer: How Are They Best Used? *Cancers (Basel)* 2019; **11**(3).
57. Wishart DS, Feunang YD, Guo AC, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res* 2018; **46**(D1): D1074-D82.
58. DRUGBANK ONLINE. Gefitinib. <https://go.drugbank.com/drugs/DB00317> (Zuletzt abgerufen am 26.07.2022).
59. DRUGBANK ONLINE. Erlotinib. <https://go.drugbank.com/drugs/DB00530> (Zuletzt abgerufen am 26.07.2022).
60. DRUGBANK ONLINE. Afatinib. <https://go.drugbank.com/drugs/DB08916> (Zuletzt abgerufen am 26.07.2022).
61. DRUGBANK ONLINE. Dacomitinib. <https://go.drugbank.com/drugs/DB11963> (Zuletzt abgerufen am 26.07.2022).
62. DRUGBANK ONLINE. Osimertinib. <https://go.drugbank.com/drugs/DB09330> (Zuletzt abgerufen am 26.07.2022).
63. Hochmair MJ, Buder A, Schwab S, et al. Liquid-Biopsy-Based Identification of EGFR T790M Mutation-Mediated Resistance to Afatinib Treatment in Patients with Advanced EGFR Mutation-Positive NSCLC, and Subsequent Response to Osimertinib. *Target Oncol* 2019; **14**(1): 75-83.
64. Yu HA, Arcila ME, Rekhtman N, et al. Analysis of tumor specimens at the time of acquired resistance to EGFR-TKI therapy in 155 patients with EGFR-mutant lung cancers. *Clin Cancer Res* 2013; **19**(8): 2240-7.
65. Janne PA, Yang JC, Kim DW, et al. AZD9291 in EGFR inhibitor-resistant non-small-cell lung cancer. *N Engl J Med* 2015; **372**(18): 1689-99.
66. Cross DA, Ashton SE, Ghiorghiu S, et al. AZD9291, an irreversible EGFR TKI, overcomes T790M-mediated resistance to EGFR inhibitors in lung cancer. *Cancer Discov* 2014; **4**(9): 1046-61.
67. Triulzi T, Regondi V, De Cecco L, et al. Early immune modulation by single-agent trastuzumab as a marker of trastuzumab benefit. *Br J Cancer* 2018; **119**(12): 1487-94.
68. Yardley DA, Kaufman PA, Huang W, et al. Quantitative measurement of HER2 expression in breast cancers: comparison with 'real-world' routine HER2 testing in a multicenter Collaborative Biomarker Study and correlation with overall survival. *Breast Cancer Res* 2015; **17**: 41.
69. Abrahao-Machado LF, Scapulatempo-Neto C. HER2 testing in gastric cancer: An update. *World J Gastroenterol* 2016; **22**(19): 4619-25.
70. Cancer Genome Atlas Research N, Analysis Working Group: Asan U, Agency BCC, et al. Integrated genomic characterization of oesophageal carcinoma. *Nature* 2017; **541**(7636): 169-75.
71. Owens MA, Horten BC, Da Silva MM. HER2 amplification ratios by fluorescence in situ hybridization and correlation with immunohistochemistry in a cohort of 6556 breast cancer tissues. *Clin Breast Cancer* 2004; **5**(1): 63-9.
72. DRUGBANK ONLINE. Lapatinib. <https://go.drugbank.com/drugs/DB01259> (Zuletzt abgerufen am 26.07.2022).

73. European Medicines Agency. EMEA/H/C/000795 - IB/0067 - Tyverb : EPAR - Product Information. https://www.ema.europa.eu/en/documents/product-information/tyverb-epar-product-information_en.pdf (Zuletzt abgerufen am 26.07.2022).
74. European Medicines Agency. EMEA/H/C/002489 - R/0071 - Xalkori : EPAR - Product Information. https://www.ema.europa.eu/en/documents/product-information/xalkori-epar-product-information_en.pdf (Zuletzt abgerufen am 26.07.2022)
75. European Medicines Agency. EMEA/H/C/004936 - IB/0001 - Rozlytrek: EPAR - Product information. https://www.ema.europa.eu/en/documents/product-information/rozlytrek-epar-product-information_en.pdf (Zuletzt abgerufen am 26.07.2022).
76. Proietti I, Skroza N, Michelini S, et al. BRAF Inhibitors: Molecular Targeting and Immunomodulatory Actions. *Cancers (Basel)* 2020; **12**(7).
77. Planchard D, Besse B, Groen HJM, et al. Dabrafenib plus trametinib in patients with previously treated BRAF(V600E)-mutant metastatic non-small cell lung cancer: an open-label, multicentre phase 2 trial. *Lancet Oncol* 2016; **17**(7): 984-93.
78. Kopetz S, Grothey A, Yaeger R, et al. Encorafenib, Binimetinib, and Cetuximab in BRAF V600E-Mutated Colorectal Cancer. *N Engl J Med* 2019; **381**(17): 1632-43.
79. Pao W, Wang TY, Riely GJ, et al. KRAS mutations and primary resistance of lung adenocarcinomas to gefitinib or erlotinib. *PLoS Med* 2005; **2**(1): e17.
80. Kessler D, Gmachl M, Mantoulidis A, et al. Drugging an undruggable pocket on KRAS. *Proc Natl Acad Sci U S A* 2019; **116**(32): 15823-9.
81. Robert C, Grob JJ, Stroyakovskiy D, et al. Five-Year Outcomes with Dabrafenib plus Trametinib in Metastatic Melanoma. *N Engl J Med* 2019; **381**(7): 626-36.
82. Dummer R, Ascierto PA, Gogas HJ, et al. Overall survival in patients with BRAF-mutant melanoma receiving encorafenib plus binimetinib versus vemurafenib or encorafenib (COLUMBUS): a multicentre, open-label, randomised, phase 3 trial. *Lancet Oncol* 2018; **19**(10): 1315-27.
83. Dummer R, Schadendorf D, Ascierto PA, et al. Binimetinib versus dacarbazine in patients with advanced NRAS-mutant melanoma (NEMO): a multicentre, open-label, randomised, phase 3 trial. *Lancet Oncol* 2017; **18**(4): 435-45.
84. Ostrem JM, Peters U, Sos ML, Wells JA, Shokat KM. K-Ras(G12C) inhibitors allosterically control GTP affinity and effector interactions. *Nature* 2013; **503**(7477): 548-51.
85. Amgen. Study to Compare AMG 510 "Proposed INN Sotorasib" With Docetaxel in Non Small Cell Lung Cancer (NSCLC) (CodeBreak 200). <https://www.clinicaltrials.gov/ct2/show/NCT04303780> (Zuletzt abgerufen am 08.08.2021).
86. Sakamoto K, Masutani T, Hirokawa T. Generation of KS-58 as the first K-Ras(G12D)-inhibitory peptide presenting anti-cancer activity in vivo. *Sci Rep-Uk* 2020; **10**(1).
87. Mirati Therapeutics Inc. Phase 2 Trial of MRTX849 Plus Pembrolizumab for NSCLC With KRAS G12C Mutation KRYSTAL-7. <https://www.clinicaltrials.gov/ct2/show/NCT04613596> (Zuletzt abgerufen am 08.08.2021).
88. Qin S, Xu J, Wang L, et al. Impact of tumor epidermal growth factor receptor (EGFR) status on the outcomes of first-line FOLFOX-4 +/- cetuximab in patients (pts) with RAS-wild-type (wt) metastatic colorectal cancer (mCRC) in the randomized phase 3 TAILOR trial. *Annals of Oncology* 2016; **27**.
89. European Medicines Agency. EMEA/H/C/000558 - N/0087 - Erbitux : EPAR - Product Information. https://www.ema.europa.eu/en/documents/product-information/erbitux-epar-product-information_en.pdf (Zuletzt abgerufen am 26.07.2022)
90. Consortium ITP-CAoWG. Pan-cancer analysis of whole genomes. *Nature* 2020; **578**(7793): 82-93.
91. Guinney J, Dienstmann R, Wang X, et al. The consensus molecular subtypes of colorectal cancer. *Nat Med* 2015; **21**(11): 1350-6.
92. Schutte M, Risch T, Abdavi-Azar N, et al. Molecular dissection of colorectal cancer in pre-clinical models identifies biomarkers predicting sensitivity to EGFR inhibitors. *Nat Commun* 2017; **8**: 14262.
93. Mishra NK, Southekal S, Guda C. Survival Analysis of Multi-Omics Data Identifies Potential Prognostic Markers of Pancreatic Ductal Adenocarcinoma. *Front Genet* 2019; **10**: 624.

94. Gonzalez Bosquet J, Devor EJ, Newton AM, et al. Creation and validation of models to predict response to primary treatment in serous ovarian cancer. *Sci Rep* 2021; **11**(1): 5957.
95. Dempster JM, Krill-Burger J, Warren A, McFarland JM, Golub TR, Tsherniak A. Gene expression has more power for predicting in vitro cancer cell vulnerabilities than genomics. *bioRxiv* 2020: 2020.02.21.959627.
96. Cancer Cell Line Encyclopedia C, Genomics of Drug Sensitivity in Cancer C. Pharmacogenomic agreement between two cancer cell line data sets. *Nature* 2015; **528**(7580): 84-7.
97. Klebanov L, Yakovlev A. Diverse correlation structures in gene expression data and their utility in improving statistical inference. *Ann Appl Stat* 2007; **1**(2): 538-59.
98. James G, Witten D, Hastie T, Tibshirani R. An Introduction to Statistical Learning: with Applications in R. 2nd ed. New York: Springer, 2021.
99. Farrar DE, Glauber RR. Multicollinearity in Regression Analysis - Problem Revisited. *Rev Econ Stat* 1967; **49**(1): 92-107.
100. Heinze G, Wallisch C, Dunkler D. Variable selection - A review and recommendations for the practicing statistician. *Biom J* 2018; **60**(3): 431-49.
101. Lee JK, Liu Z, Sa JK, et al. Pharmacogenomic landscape of patient-derived tumor cells informs precision oncology therapy. *Nat Genet* 2018; **50**(10): 1399-411.
102. Xiao J, Wang R, Cai X, Ye Z. Coupling of Co-expression Network Analysis and Machine Learning Validation Unearthed Potential Key Genes Involved in Rheumatoid Arthritis. *Front Genet* 2021; **12**: 604714.
103. Stingo FC, Chen YA, Tadesse MG, Vannucci M. Incorporating Biological Information into Linear Models: A Bayesian Approach to the Selection of Pathways and Genes. *Ann Appl Stat* 2011; **5**(3): 1978-2002.
104. Tian S, Wang C, Wang B. Incorporating Pathway Information into Feature Selection towards Better Performed Gene Signatures. *Biomed Res Int* 2019; **2019**: 2497509.
105. Cun Y, Frohlich HF. Prognostic gene signatures for patient stratification in breast cancer: accuracy, stability and interpretability of gene selection approaches using prior knowledge on protein-protein interactions. *BMC Bioinformatics* 2012; **13**: 69.
106. Kuhn M., Johnson K. Applied Predictive Modeling. 1st ed. New York: Springer, 2013.
107. Remeseiro B, Bolon-Canedo V. A review of feature selection methods in medical applications. *Comput Biol Med* 2019; **112**: 103375.
108. Blagus R, Lusa L. Boosting for high-dimensional two-class prediction. *BMC Bioinformatics* 2015; **16**: 300.
109. Tibshirani R. Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society Series B (Methodological)* 1996; **58**(1): 267-88.
110. Breiman L. Random Forests. *Machine Learning* 2001; **45**(1): 5-32.
111. Cortes C, Vapnik V. Support-vector networks. *Machine Learning* 1995; **20**(3): 273-97.
112. DepMap, Broad (2020): DepMap 21Q1 Public. figshare. https://figshare.com/articles/dataset/public_21q1/13681534 (Zuletzt abgerufen am 26.06.2021).
113. Meyers RM, Bryan JG, McFarland JM, et al. Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. *Nat Genet* 2017; **49**(12): 1779-84.
114. Dempster JM, Rossen J, Kazachkova M, et al. Extracting Biological Insights from the Project Achilles Genome-Scale CRISPR Screens in Cancer Cell Lines. *bioRxiv* 2019: 720243.
115. Dempster JM, Boyle I, Vazquez F, et al. Chronos: a CRISPR cell population dynamics model. *bioRxiv* 2021: 2021.02.25.432728.
116. Pacini C, Dempster JM, Boyle I, et al. Integrated cross-study datasets of genetic dependencies in cancer. *Nat Commun* 2021; **12**(1): 1661.
117. Ghandi M, Huang FW, Jane-Valbuena J, et al. Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature* 2019; **569**(7757): 503-8.
118. Campbell SL, Khosravi-Far R, Rossman KL, Clark GJ, Der CJ. Increasing complexity of Ras signaling. *Oncogene* 1998; **17**(11 Reviews): 1395-413.
119. Chakravarty D, Gao J, Phillips SM, et al. OncoKB: A Precision Oncology Knowledge Base. *JCO Precis Oncol* 2017; **2017**.

120. Loboda A, Nebozhyn M, Klinghoffer R, et al. A gene expression signature of RAS pathway dependence predicts response to PI3K and RAS pathway inhibitors and expands the population of RAS pathway activated tumors. *BMC Med Genomics* 2010; **3**: 26.
121. Singh A, Greninger P, Rhodes D, et al. A gene expression signature associated with "K-Ras addiction" reveals regulators of EMT and tumor cell survival. *Cancer Cell* 2009; **15**(6): 489-500.

6. Anhang

6.1. Abbildungsverzeichnis

Abbildung 1: Schematische Darstellung des RAS-Signalwegs	11
Abbildung 2: Vergleichende Darstellung der Proteinsequenzen der RAS-Isoformen	14
Abbildung 3: Minimierungsprobleme verschiedener linearer Regressionsverfahren	20

7. Vorabveröffentlichungen von Ergebnissen

Die Resultate dieser Promotionsarbeit wurden in der Publikation Ulmer B, Odenthal M, Buettner R, Roth W, Kloth M. Diffusion kernel-based predictive modeling of KRAS dependency in KRAS wild type cancer cell lines. *NPJ Syst Biol Appl* 2022; **8**(1): 2. bereits veröffentlicht.