

Gefördert durch:



Bundesministerium
für Forschung, Technologie
und Raumfahrt

Heinrich-Heine-Universität Düsseldorf
Universität zu Köln
Philipps-Universität Marburg

Deep Insight - Integration von Keimbahn- und
somatischen genetischen Profilen durch
maschinelles Lernen zum Verständnis der Ätiologie
von Speiseröhrenkrebs

Prof. Dr. Dominik Heider, Prof. Dr. Axel Hillmer, Prof.
Dr. Johannes Schumacher

031L0267A, 031L0267B, 031L0267C,

Die Verantwortung für den Inhalt dieser Veröffentlichung liegt beim Autor.

I. Kurzbericht

Aufgabenstellung

Das Adenokarzinom des Ösophagus (EAC) stellt aufgrund seiner aggressiven Natur, schlechten Prognose und zunehmenden Inzidenz eine erhebliche medizinische Herausforderung dar. Trotz umfangreicher Forschung zu somatischen und Keimbahn-basierten molekularen Veränderungen fehlt es an einem integrativen Verständnis der Mechanismen, die zur Entstehung dieser Krebsart führen. Insbesondere fehlt es an prädiktiven Biomarkern für die maligne Transformation von Barrett-Ösophagus (BE) zu EAC. BE ist eine gutartige Metaplasie der unteren Speiseröhre aufgrund von Reflux des Magensafts. BE erhöht das Risiko für die Entstehung von EAC. Ziel des Projektes *Deep Insight* war es daher, mithilfe moderner Methoden des maschinellen Lernens – insbesondere Deep Learning – Keimbahnvarianten und somatische Tumorveränderungen zu integrieren, um neue Mechanismen der Krebsentstehung aufzudecken und multivariate prädiktive Risikoscores (AI-PRS) zu entwickeln. Dies sollte nicht nur die Präzisionsmedizin fördern, sondern auch neue Ansätze zur Früherkennung ermöglichen.

Ablauf des Vorhabens

Zu Beginn des Projekts wurden große genomische Datensätze aus internationalen EAC-Kohorten (insbesondere des *International Cancer Genome Consortium*, ICGC) sowie eine umfangreiche Biobank von über 2.500 BE-Patienten (bereitgestellt über das g4b-Konsortium) zusammengeführt. Die Patienten wurden basierend auf Polygenic Risk Scores stratifiziert, und Extremgruppen hinsichtlich des genetischen Risikos identifiziert.

Anschließend wurden neue Deep-Learning-Modelle auf Basis dieser Daten entwickelt und trainiert. Hierbei wurden Herausforderungen wie die hohe Dimensionalität (Millionen SNPs) und begrenzte Stichprobengrößen durch gezielte Speicheroptimierungen und modellbasierte Parallelisierung auf HPC-Clustern adressiert. Es wurden Modelle auf Chromosomenebene trainiert, die durch neuartige Speicherstrategien effizient ausgeführt werden konnten.

Darüber hinaus wurden die Keimbahnvarianten der ICGC Kohorte für EAC bestimmt. Dies erlaubte die Stratifizierung der Kohorte in Extremgruppen für das Keimbahnrisiko. Die Risikogruppen konnten hinsichtlich unterschiedlicher somatischer Mutationsprofile analysiert werden.

In einem weiteren Schritt wurden Formalin-fixierte Paraffin-eingebettete (FFPE) BE Gewebeproben aus den Extremgruppen analysiert. Insgesamt konnten von 58 BE Patienten genomweite Sequenzierungen durchgeführt werden, die somatische Varianten wie SNVs, indels, Kopienzahlveränderungen und Mutationssignaturen enthielten. Diese wurden anschließend mit Keimbahnvarianten in Beziehung gesetzt. Parallel dazu wurden die prädiktiven Modelle mithilfe von Multi-Label-Learning und Explainable-AI-Verfahren weiterentwickelt.

Die finale Implementierung der entwickelten Pipeline erfolgte über containerisierte Softwarelösungen (Docker).

Wesentliche Ergebnisse

Es wurden multivariate Polygenic Risk Scores (AI-PRS) entwickelt, die das individuelle Keimbahnrisiko besser abbilden als klassische lineare PRS-Modelle. Trotz begrenzter Datenmengen konnten Deep-Learning-Modelle auf Chromosomenebene valide Vorhersagen erzeugen. Die Kombination der Modelle führte jedoch nur zu marginalen Verbesserungen, was die Notwendigkeit größerer Datensätze unterstreicht. In BE-Patienten mit hohem AI-PRS konnten Mutationssignaturen identifiziert werden, die typischerweise in EAC beobachtet werden – ein Hinweis darauf, dass molekulare Transformation bereits früh nachweisbar ist. Die gesamte Analysepipeline wurde containerisiert, automatisiert und zur Nutzung auf HPC- sowie Cloud-Infrastrukturen vorbereitet. Dies erlaubt künftig eine breite Anwendung in Forschungsverbänden. Die Arbeiten führten zu mehreren internationalen Publikationen und trugen zu einem besseren Verständnis der genetischen Architektur und molekularen Evolution von EAC bei.

Das Projekt *Deep Insight* liefert damit einen wertvollen Beitrag zur integrativen Krebsforschung und zeigt, wie KI-gestützte Analysen neue Wege in der Onkologie eröffnen können.

II. Eingehende Darstellung

II.1 Ursprüngliche Aufgabenstellung sowie der wissenschaftliche und technische Stand, an den angeknüpft wurde

Das Adenokarzinom des Ösophagus (EAC) gehört zu den Krebserkrankungen mit der höchsten Zuwachsrate in der westlichen Welt. Trotz intensiver Forschung sind die Überlebensraten weiterhin sehr niedrig, da geeignete Marker für Früherkennung und effektive Therapiestratifizierung fehlen. Besonders kritisch ist die Transformation von Barrett-Ösophagus (BE) zu EAC, ein Prozess, der bislang nur unzureichend verstanden wird. Bisherige molekulare Studien fokussierten entweder auf somatische Mutationen oder auf Keimbahnveränderungen, eine umfassende Integration beider Ebenen erfolgte kaum.

Die Genetik von EAC ist komplex: Im somatischen Genom finden sich Mutationen in p53, amplifizierte Onkogene (z. B. *ERBB2*, *CCND1*) sowie extensive strukturelle Aberrationen. Auf der anderen Seite zeigen Keimbahnanalysen eine Beteiligung von Polygenic Risk Scores (PRS), die jedoch bisher nur geringe Effektstärken aufweisen. Zudem ist wenig darüber bekannt, wie Keimbahnrisiken die somatische Evolution des Tumors beeinflussen. Die Hypothese des Projekts war, dass durch Deep Learning eine integrative Modellierung möglich ist, die diese Interaktion aufklärt und neue prädiktive Risikowerte liefert.

Ein zentrales Ziel war die Entwicklung von *Artificial Intelligence* (AI)-gestützten PRS-Modellen (AI-PRS), die durch den Einsatz nichtlinearer Deep-Learning-Architekturen (z. B. CNNs, Transformer) das Risikoprofil auf Basis großer *Single Nucleotide Polymorphism* (SNP)-Matrizen besser abbilden als klassische lineare Modelle. Der Ausgangspunkt war ein europäischer EAC-Datensatz (ICGC, >400 Patienten) sowie die große g4b-BE-Kohorte (>2.500 Patienten), die eine Stratifikation in Risikogruppen ermöglichte. Für eine systematische Untersuchung wurden Extremgruppen ausgewählt, bei denen eine genomische Charakterisierung durch WGS und phänotypische Daten kombiniert werden konnte.

Technisch bestand eine Herausforderung in der Handhabung extrem hochdimensionaler SNP-Daten (bis zu 30 Mio. Varianten pro Patient). Um diese effizient zu verarbeiten, wurden speicheroptimierte Kodierungen und modulare Trainingsstrategien auf HPC-Systemen vorgesehen. Darüber hinaus sollte mit Explainable-AI-Methoden nachvollzogen werden, welche genetischen Loci das Modell für die Risikoabschätzung heranzieht. So sollten neue, bislang unentdeckte genetische Determinanten identifiziert werden.

Schließlich war ein weiteres Ziel, somatische Profile (Mutationstypen, CNVs, Signaturen) in Bezug zu den AI-PRS-Modellen zu analysieren, um einen möglichen Einfluss der Keimbahn auf die Tumorevolution zu erfassen. Die Integration dieser multidimensionalen Informationen sollte neue Einblicke in die molekulare Ätiologie des EAC liefern und damit auch die Basis für personalisierte Diagnostik und Prävention legen.

II.2 Erzielte Ergebnisse

Im Folgenden möchten wir einen Überblick über unsere Projektergebnisse präsentieren. Eine vollständige Liste unserer Publikationen findet sich am Ende dieses Berichts.

Aufbau und Integration multizentrischer Daten

Im Rahmen des Projekts wurden umfangreiche genomische Daten aus internationalen Kohorten (u.a. ICGC) zusammengeführt. Parallel dazu wurde auf die g4b-Biobank zurückgegriffen, um Proben von über 2.500 BE-Patienten zu analysieren. Diese Proben wurden anhand klassischer und AI-basierter PRS-Modelle stratifiziert, wobei gezielt Extremgruppen für weiterführende molekulare Analysen ausgewählt wurden.

Entwicklung und Training von Deep-Learning-Modellen

Zunächst haben wir die bislang umfangreichste Metanalyse für genomweite Assoziation von 16.790 Patienten mit BE oder EAC durchgeführt und konnten 11 neue Risikoloci identifizieren (Schröder et al. 2023). Das Magenkarzinom zeigt eine ätiologische Überlappung mit EAC und die somatischen Mutationen weisen große Ähnlichkeit mit EAC auf. Wir haben daher in einer genomweiten Assoziationsstudie das Risikoprofil des Magenkarzinoms bestimmt und konnten Ähnlichkeiten zwischen dem Risikoprofil für Karzinome im oberen Magen (Kardia) und EAC zeigen (Hess et al. 2023). In Klau et al. (2023) haben wir untersucht, ob sich die Vorhersagekraft polygenetischer Risikoscores (PRS) durch die Kombination mehrerer PRS (multi-PRS) und den Einsatz moderner Machine-Learning-Verfahren wie Deep Neural Networks (DNNs) verbessern lässt. Wir konnten zeigen, dass Modelle, die mehrere PRS aus verwandten Krankheitsbildern integrieren, eine höhere Vorhersageleistung erzielen als Modelle, die nur einen einzelnen PRS verwenden. Diese Verbesserung beruht auf der Tatsache, dass viele Krankheiten gemeinsame genetische Grundlagen teilen.

Im Vergleich verschiedener Modellierungsansätze zeigte sich, dass Deep Neural Networks in fast allen Fällen bessere Ergebnisse lieferten als klassische Methoden wie Ridge Regression oder Random Forests. Obwohl die Verbesserungen im Bereich von ein bis zwei Prozent lagen, haben sie durch die Größe der betroffenen Bevölkerungsgruppen potenziell große Auswirkungen auf die Identifikation zusätzlicher Krankheitsfälle.

Der Einsatz von Deep Learning in Kombination mit multiplen PRS ist daher ein vielversprechender Weg, um die genetische Risikovorhersage zu verbessern. Diese Ansätze bieten ein großes Potenzial für die personalisierte Medizin und könnten in der klinischen Praxis zur besseren Risikostratifizierung genutzt werden.

Wir haben dann verschiedene Kodierungsstrategien und Deep-Learning-Architekturen getestet, darunter CNNs und Transformer-basierte Modelle, um diese chromosomenspezifisch zu trainieren und für die Fusion verschiedener genetischer Signalquellen optimiert. Speicheroptimierte Datenpipelines und eine effektive Parallelisierung auf HPC-Systemen erlaubten ein Training trotz großer Datenmengen.

Genomische und funktionelle Charakterisierung

Um die nach Keimbahnrisiko stratifizierten somatischen EAC Mutationen in den Kontext der generellen EAC Genomik und Transkriptomik zu setzen, haben wir in einer Übersichtsarbeit die bisher bekannten internationalen Studien zusammengefasst und interpretiert (Hoppe et al. 2021). Durch die Reanalyse von Daten des The Cancer Genome Atlas (TCGA) konnten wir zeigen, dass der Verlust des Y-Chromosoms besonders häufig beim papillären Nierenkarzinom und EAC auftreten (Müller et al. 2023). Im Weiteren haben wir die bioinformatische Pipeline für die integrative Analyse von genomischen und transkriptomischen Daten entwickelt (Plum et al. 2024). Darauf aufbauend haben wir die Rohdaten der ICGC Kohorte von 400 EAC Patienten reprozessiert und Keimbahnvarianten bestimmt. Hierfür wurde extensiv die HPC-Infrastruktur der Universität zu Köln und des Cologne Excellence Cluster on Aging-Associated Diseases (CECAD) genutzt. Basierend auf den Keimbahnvarianten wurden PRS bestimmt und die Kohorte in Patienten mit hohem bzw. niedrigem PRS aufgeteilt. Das erlaubte die vergleichende Analyse von Mutationsprofilen in Abhängigkeit der Keimbahnbelastung. Unter Verwendung des Medians des PRS zur Aufteilung der 400 EAC Patienten haben sich keine signifikanten Unterschiede zwischen Patienten mit hohem und niedrigem PRS für klassische quantifizierbare Mutationsphänomene wie Tumor Mutational Burden (TMB) und Mutationssignaturen (Alexandrov 2013 Nature; Alexandrov 2020 Nature) ergeben. Generell konnten wir Korrelationen von Mutationsphänomenen untereinander, wie TMB und Mutationen in CTCF Bindungstellen beobachten und zwischen Mutationssignaturen, wie z.B. SBS2 und SBS13 (Abb. 1).

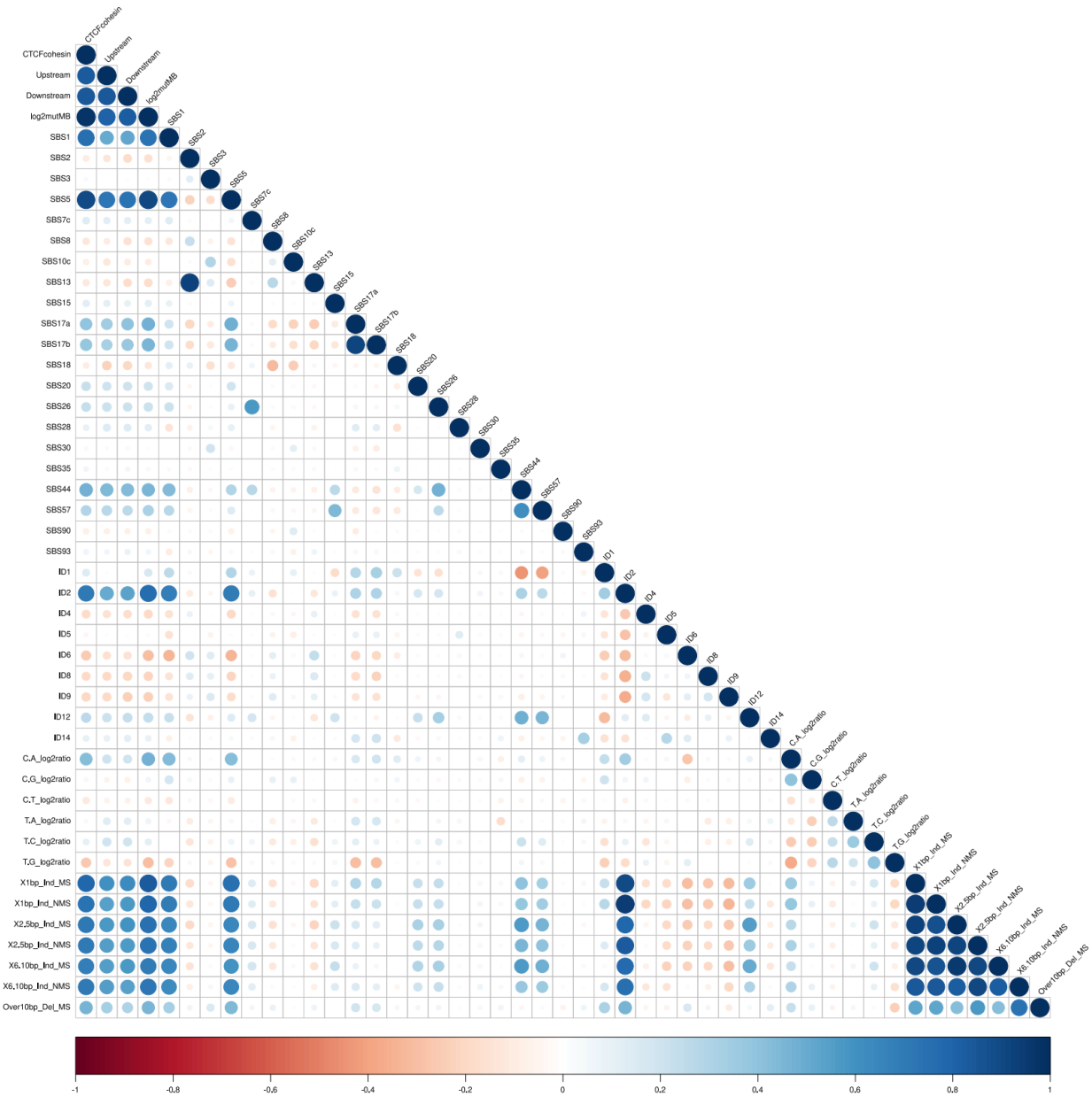


Abbildung 1. Korrelation zwischen verschiedenen somatischen Mutationsklassen der EAC Kohorte von 400 Patienten. Die verschiedenen Mutationsklassen sind vertikal und horizontal gleich aufgetragen. Die oberen drei Klassen beschreiben Häufigkeiten von Mutationen in CTCF/Cohesin Bindungsstellen bzw. upstream und downstream, gefolgt von TMB (log2mutMB). Darunter sind die Mutationssignaturen basierend auf Einzelbasenaustauschen (SBS1 bis SBS93), indel Signaturen (ID1 bis ID14), definierten Einzelbasenaustauschen (z.B. C.A_log2ratio für C>A Austausch) und unterschiedlich langen indels.

Es zeigte sich unter den Tumoren von Patienten mit extrem hohem PRS (>1,5; n=24) verglichen mit Patienten mit extrem niedrigem PRS (<1,5; n=30) eine Verschiebung der typischen Mutationssignaturen mit geringeren Anteilen von SBS17a, SBS17b und SBS40 sowie einem höheren Anteil von SBS5 in Tumoren von Patienten mit niedrigem PRS (Abb. 2).

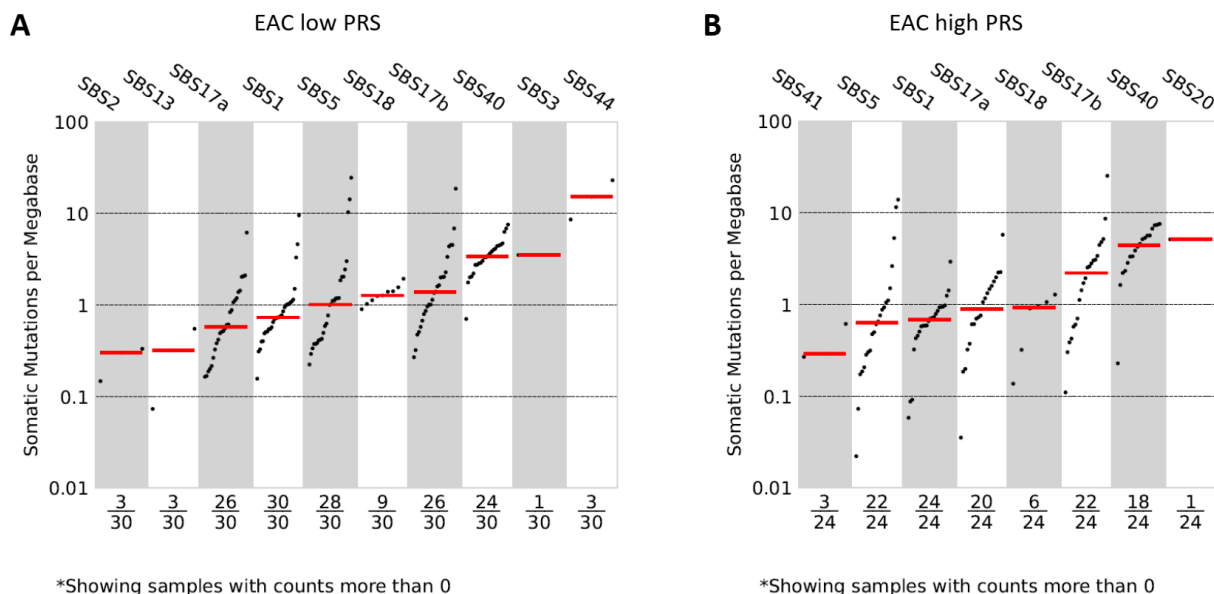


Abbildung 2. Frequenzen der Mutationssignaturen der ICGC Kohorte für EAC für Patienten mit extrem niedrigem (A) und extrem hohem (B) PRS. Die Mutationssignaturen wurden mittels der *Non-negative Matrix Factorization* (NMF) Methode für die EAC Patienten mit extrem niedrigem (n=30) und Patienten mit extrem hohem PRS (n=24) ermittelt (Alexandrov et al. 2020 Nature) und deren Häufigkeiten (Y-Achse) für die detektierten Signaturen (X-Achse) geplottet. Rote horizontale Linien markieren die Mediane.

Im Weiteren zeigten sich auf der Ebene individueller Gene Unterschiede. Während die am häufigsten mutierten Gene *TP53* und *TTN* vergleichbar häufig in den PRS-hoch bzw. PRS-niedrig Gruppen mutiert waren (in etwa 80% bzw. 60% der Tumoren) traten Mutationen in dem Tumorsuppressorgen *FAT4* deutlich häufiger in der PRS-niedrig Gruppe auf (33%) als in der PRS-hoch Gruppe (<10%; Abb. 3). Ähnlich verhält es sich mit Punktmutationen im Tumorsuppressorgen *CDKN2A*, welches in der Regel durch Deletionen des gesamten Gens inaktiviert ist. 23% der PRS-niedrig Tumoren wiesen *CDKN2A* Mutationen auf, wohingegen <10% der PRS-hoch Tumoren solche Punktmutationen zeigten. Das können erste Hinweise sein, dass das Keimbahnrisiko zur Mutationsarchitektur von EAC beitragen kann.

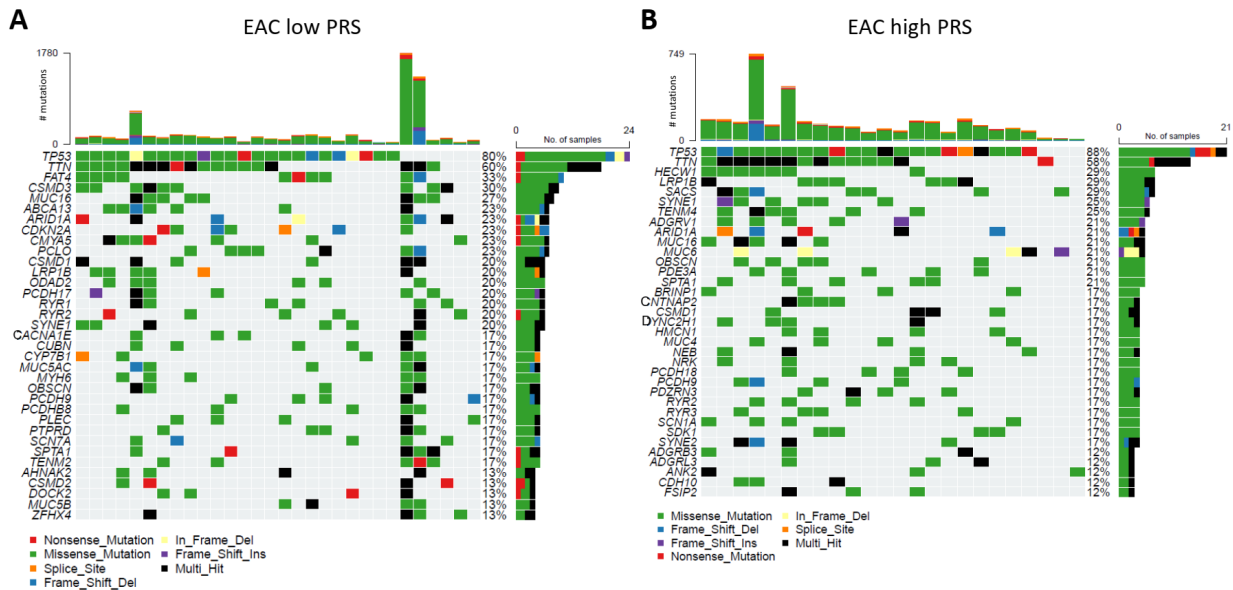


Abbildung 3. Oncoplots für EACs von Patienten der ICGC Kohorte stratifiziert nach PRS. Es sind jeweils Gene dargestellt, die in mindestens 10% der Patienten somatische Mutationen aufweisen. Es wurden die 30 Patienten mit dem niedrigsten PRS (A) verglichen mit den 24 Patienten mit dem höchsten PRS (B). Oben sind die Gesamtzahlen der somatischen Mutationen pro Tumor aufgeführt.

BE ist eine gutartige Metaplasie des distalen (unteren) Ösophagus. BE erhöht deutlich das Risiko, EAC zu entwickeln. Es ist bekannt, dass Mutationen, die typischerweise beim EAC auftreten, in BE bereits vorliegen können. Es konnten bisher jedoch keine spezifischen Mutationen oder andere Biomarker identifiziert werden, die verlässlich vorhersagen lassen, welche BE Patienten später EAC entwickeln. Wir haben daher im nächsten Schritt getestet, ob bereits in der BE Mukosa Mutationssignaturen auftreten, die typisch für EAC sind, und ob diese mit dem Keimbahnrisiko assoziiert werden können. Wir haben daher im nächsten Schritt von unserer Kohorte von BE Patienten, die in der genomweiten Assoziationsuntersuchung analysiert wurden, archiviertes BE Gewebe von Patienten mit sehr hohem und sehr niedrigem PRS entnommen. Wir haben als Analysemethode genomweite Sequenzierung ausgewählt, da diese Daten am besten mit den Daten des ICGC verglichen werden können und am zuverlässigsten Mutationssignaturen darstellen können. Aus 58 selektierten FFPE-BE-Gewebeproben und gepaarten Blutproben als Keimbahnkontrolle wurden genomweite Sequenzdaten erzeugt. Hierauf wurde der wesentliche Teil der Verbrauchsmittel (Vergabe von Aufträgen) verwendet. Dabei wurden somatische Varianten (SNVs, Indels, CNVs) in Beziehung zu Keimbahn-PRS gesetzt. Generell sind weniger klassische Onko- und Tumorsuppressorgene in BE mutiert als in EAC (Abb. 4). Insbesondere fanden wir Mutationen in *TP53* in weniger als 10% von BE.

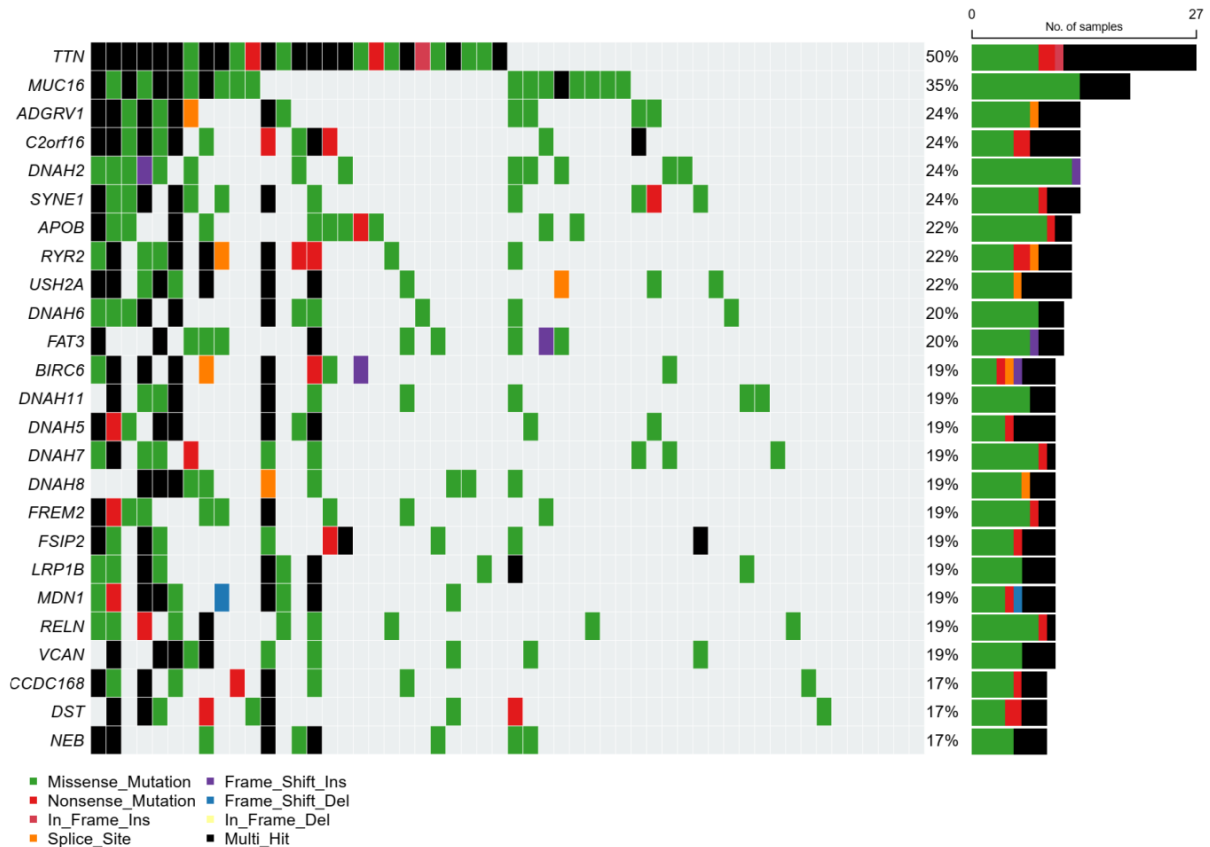


Abbildung 4. **Oncoplot für BEs von Patienten der DeepInsight Kohorte.** Es sind jeweils Gene dargestellt, die in mindestens 10% der Patienten somatische Mutationen aufweisen.

Die Mutationssignaturen SBS1, die durch spontane Deaminierung des Cytosins hervorgerufen wird und daher als "Alterungssignatur" bezeichnet wird, sowie SBS5, die als Uhr-ähnliche (clock-like) Signatur beschrieben wird, sind deutlich in BE ungeachtet des PRS zu sehen (Abb. 5). SBS4 (Signatur für Tabakrauchen) und SBS10b (Signatur von Polymerase Epsilon Mutationen) waren stärker in BE von Patienten mit niedrigem PRS ausgeprägt wohingegen SBS2, SBS32, SBS41, SBS96A und SBS18 stärker in der PRS-hoch Gruppe ausgeprägt waren. Das mag zum Einen bedeuten, dass es zusätzliche Mutationsprozesse (Tabakrauchen, Polymerase Epsilon Alteration) in PRS-niedrig Patienten benötigt, um BE zu entwickeln. Zum Anderen könnte das Auftreten der für EAC typischen Signatur SBS18 in der PRS-hoch Gruppe auf eine höhere Wahrscheinlichkeit hindeuten, zu EAC zu entarten.

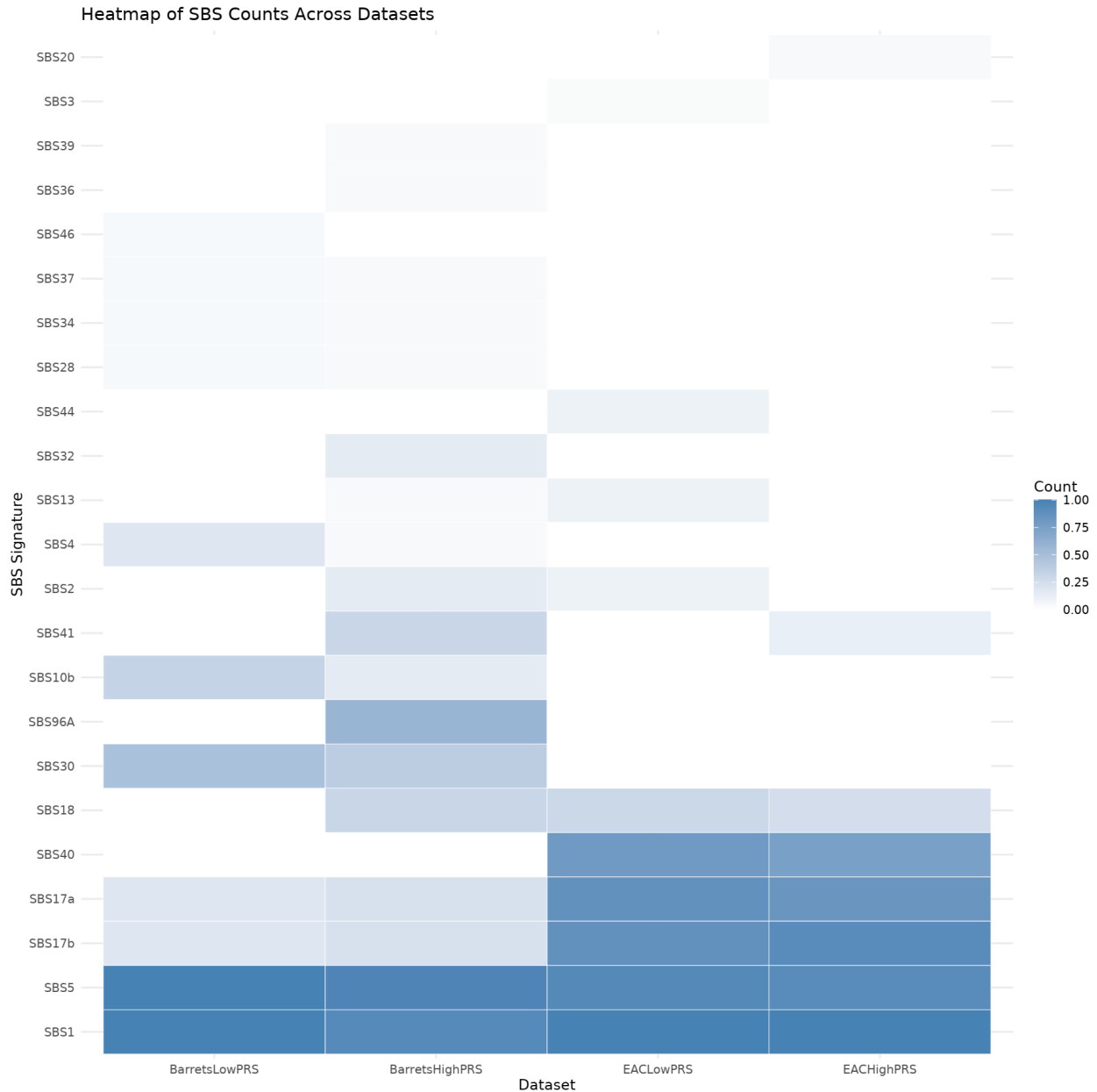


Abbildung 5. Mutationssignaturen für BE und EAC stratifiziert nach PRS. BE (unsere Sequenzierungen) und EAC Patienten (ICGC Daten) mit extrem niedrigem bzw. extrem hohem PRS (X-Achse) wurden hinsichtlich des prozentualen Anteils der Mutationen, die einzelnen Signaturen zugeordnet wurden (Y-Achse) geplottet.

Softwareentwicklung und Infrastruktur

Es wurde eine containerisierte Analysepipeline entwickelt, die alle Schritte von der Datenverarbeitung bis zur Modellbewertung des PRS umfasst. Die Nutzung über de.NBI und Docker garantiert eine einfache Nachnutzung.

Abweichungen vom Projektplan

Die im Projektplan avisierten Meilensteine konnten größtenteils ohne Verzögerungen umgesetzt werden. Aufgrund von technischen Problemen kam es zu erheblichen Verzögerungen bei der Bearbeitung der Daten. Diese Verzögerungen wurden jedoch bis zum Ende des Projekts kompensiert. Weitere Verzögerungen entstanden durch die verspätete Bereitstellung von Proben aufgrund der noch laufenden Corona-Pandemie zu Beginn des Projektes. Die verzögerten Analysen konnten jedoch im Laufe des Projektes nachgeholt werden, so dass es zu keinen Abweichungen im Projektplan gekommen ist.

II.3 Wichtigste Positionen des zahlenmäßigen Nachweises

Die bewilligten Personalmittel wurden eingesetzt zur Finanzierung (1) eines Wissenschaftlers zur Erstellung, Training und Evaluation komplexer Machine-Learning- und Deep-Learning-Modelle, (2) eines Wissenschaftlers zur bioinformatischen Analyse der somatischen Mutationen, (3) eines Wissenschaftlers für die GWAS Analysen in den Keimbahndaten. Die bereitgestellten Sachmittel wurden überwiegend zur Sequenzierung und Charakterisierung der Proben verwendet. Die Reisemittel wurden primär zur Teilnahme an nationalen und internationalen Konferenzen verwendet, soweit dies im Pandemie-Zeitraum möglich war.

II.4 Notwendigkeit und Angemessenheit der geleisteten Arbeit

Die Verbindung somatischer und Keimbahn-basierter genetischer Informationen stellt eine zukunftsweisende Methode für die Krebsrisikoforschung dar. Deep-Learning-Verfahren konnten hier erfolgreich eingesetzt werden, um sowohl prädiktive als auch mechanistische Modelle zu entwickeln. Die geleisteten Arbeiten waren daher nicht nur notwendig, sondern auch methodisch und technologisch wegweisend.

II.5 Voraussichtlicher Nutzen und Verwertbarkeit

Die entwickelten AI-PRS-Modelle ermöglichen eine individualisierte Risikobewertung für EAC. Die Softwareinfrastruktur erlaubt eine einfache Adaption auf andere Krebsarten. Die Erkenntnisse über somatische Signaturen und deren Assoziation mit Keimbahn-basierten Risiken könnten neue Möglichkeiten für die Prävention und Früherkennung bieten. Die Ergebnisse sollen zukünftig in klinische Studien einfließen und könnten langfristig in der onkologischen Diagnostik Anwendung finden.

Die im Rahmen des Projekts erforschten Deep Learning-Verfahren repräsentieren bedeutsame Fortschritte im methodischen Bereich der Künstlichen Intelligenz, insbesondere im Kontext medizinischer und onkologischer Anwendungen. Von besonderer Relevanz sind dabei zwei Schlüsselaspekte: Erstens die verbesserte Analyse von Chromosomen über hierarchische Deep Learning Architekturen sowie die Einbeziehung weiterer PRS in den AI-PRS. Letzteres ermöglicht die Übertragung von gelerntem Wissen, das in einem Kontext erworben wurde, auf einen anderen.

Sämtliche während des Projekts entwickelten und notwendigen Software-Skripte, die für die Erstellung, das Training und die Evaluation von KI-Modellen verwendet wurden, sind öffentlich zugänglich und stehen der Fachgemeinschaft über entsprechende Plattformen, wie beispielsweise GitHub, zur Verfügung. Damit ist auch für andere Arbeitsgruppen eine schnelle Entwicklung weiterer Vorhersagemodelle möglich.

Zudem wurden verschiedene Studien durchgeführt, um ein besseres Verständnis der Mutationen im EAC zu erlangen. In der Arbeit von Schröder et al. (2022) wurde die bisher größte GWAS-Meta-Analyse zu BE und EAC durchgeführt, basierend auf Daten von 16.790 Patienten und 32.476 Kontrollen. Die Studie identifizierte 27 Risikoloci (darunter 11 neu), die mit BE und/oder EAC assoziiert sind. Ergänzt wurde die GWAS durch eine Transkriptom-weite Assoziationsstudie (TWAS), die zusätzliche krankheitsrelevante Gene identifizierte. Eine genetische Korrelationsanalyse ergab starke Überschneidungen mit Risikofaktoren wie Refluxkrankheit (GERD) und Fettleibigkeit, wobei GERD genetisch stärker mit BE als mit EAC korreliert war. Schließlich konnte durch Kombination genetischer Risikoprofile (Polygenic Risk Scores, PRS) mit klinischen Faktoren die Vorhersagekraft für BE/EAC verbessert werden – ein vielversprechender Ansatz für künftige Risikomodelle und Präventionsstrategien. In der Studie von Hess et al. (2023) wurde dann die genetische Heterogenität verschiedener Krebsarten, insbesondere Magenkrebs (GC) umfassend analysiert und verglichen. Mithilfe einer Meta-Analyse von zehn europäischen genomweiten Assoziationsstudien (GWAS) mit insgesamt 5.816 Patienten und 10.999 Kontrollen konnte gezeigt werden, dass es eine starke genetische Heterogenität gibt. Zwei neue genetische Risikoloci wurden identifiziert und fünf bekannte bestätigt, wobei alle eine Subtyp-spezifische Assoziation zeigten. Zusätzlich lieferte eine Transkriptom-weite Assoziationsstudie (TWAS) Hinweise auf pathophysiologische Relevanz der Gene *MUC1*, *ANKRD50*, *PTGER4* und *PSCA*. Auch wurde gezeigt, dass die Blutgruppe einen Einfluss auf das Erkrankungsrisiko von GC hat (Blutgruppe A erhöht, Blutgruppe 0 schützt). Schließlich ergab ein Vergleich mit Daten zu BE und EAC, dass GC und EAC genetische Gemeinsamkeiten aufweisen. Die Arbeit von Müller et al. (2023) beschäftigte sich mit dem Verlust des Y-Chromosoms (Loss of Y, LoY) in Tumorgeweben bei männlichen Krebspatienten. Anhand genomischer und transkriptomischer Daten von 2.375 Patienten aus 13 Krebsentitäten zeigte die Studie, dass LoY mit erhöhter genomischer Instabilität, vermehrter Aneuploidie und erhöhter Mutationslast einhergeht. Zudem fanden sich bei LoY häufiger Mutationen im Tumorsuppressorgen *TP53* und Amplifikationen von Onkogenen wie *MET*, *KRAS*, *EGFR* und *CDK6*. Auch auf transkriptomischer Ebene wurden relevante Veränderungen beobachtet, darunter eine Hochregulation des Invasionsmarkers *MMP13* und eine Runterregulation des Tumorsuppressorgens *GPC5*. LoY korrelierte stark mit Krebsarten, die eine höhere Inzidenz bei Männern zeigen, was auf eine geschlechtsspezifische Rolle von LoY in der Krebsentstehung hinweist. Wir haben Einzelzell-RNA-Sequenzierungen durchgeführt und die zelltypspezifischen Expressionsprofile mit Keimbahnrisikoprofilen in Linkage Disequilibrium Regression Score (LDSR) Analysen auf Anreicherung getestet (Wenzel et al. Cell Genomics, *accepted*). Hier konnten wir zeigen, dass die Entwicklung von EAC stärker durch Keimbahnrisiko über lokale zelluläre Prozesse beeinflusst wird als die Entwicklung von BE. Intestinale Metaplasiezellen und die Differenzierung in Säulenepithelzellen sind besonders relevant für die EAC Entstehung. Darüber hinaus sind Fibroblasten und Endothelzellen an der Entstehung von EAC und BE beteiligt.

II.6 Veröffentlichungen

Die im Projekt durchgeführten Arbeiten und Ergebnisse wurden in insgesamt sechs Veröffentlichungen in international anerkannten peer-reviewed Journals publiziert. Ein weiteres Manuskript ist *accepted* und die Publikation zu unserer BE Kohorte und der Stratifizierung der somatischen Mutationssignaturen hinsichtlich PRS ist in Vorbereitung.

Wissenschaftliche Publikationen:

1. Plum PS, Hess T, Bertrand D, Morgenstern I, Velazquez Camacho O, Jonas C, Alidousty C, Wagner B, Roessler S, Albrecht T, Becker J, Richartz V, Holz B, Hoppe S, Poh HM, Chia BKH, Chan CX, Pathiraja T, Teo AS, Marquardt JU, Khng A, Heise M, Fei Y, Thieme R, Klein S, Hong JH, Dima SO, Popescu I, Hoppe-Lotichius M, Buettner R, Lautem A, Otto G, Quaas A, Nagarajan N, Rozen S, Teh BT, Goepfert B, Drebber U, Lang H, Tan P, Gockel I, Schumacher J, Hillmer AM. Integrative genomic analyses of European intrahepatic cholangiocarcinoma: Novel ROS1 fusion gene and PBX1 as prognostic marker. *Clin Transl Med.* 2024 Jun;14(6):e1723. doi: 10.1002/ctm2.1723. PMID: 38877653
2. Schröder J, Chegwiddden L, Maj C, Gehlen J, Speller J, Böhmer AC, Borisov O, Hess T, Kreuser N, Venerito M, Alakus H, May A, Gerges C, Schmidt T, Thieme R, Heider D, Hillmer AM, Reingruber J, Lyros O, Dietrich A, Hoffmeister A, Mehdorn M, Lordick F, Stocker G, Hohaus M, Reim D, Kandler J, Müller M, Ebigbo A, Fuchs C, Bruns CJ, Hölscher AH, Lang H, Grimminger PP, Dakkak D, Vashist Y, May S, Görg S, Franke A, Ellinghaus D, Galavotti S, Veits L, Weismüller J, Dommermuth J, Benner U, Rösch T, Messmann H, Schumacher B, Neuhaus H, Schmidt C, Wissinowski TT, Nöthen MM; Wellcome Trust Case Control Consortium 2 (WTCCC2); Esophageal Adenocarcinoma Genetics Consortium (EAGLE); Barrett's and Esophageal Adenocarcinoma Consortium (BEACON); Dong J, Ong JS, Buas MF, Thrift AP, Vaughan TL, Tomlinson I, Whiteman DC, Fitzgerald RC, Jankowski J, Vieth M, Mayr A, Gharahkhani P, MacGregor S, Gockel I, Palles C, Schumacher J. GWAS meta-analysis of 16 790 patients with Barrett's oesophagus and oesophageal adenocarcinoma identifies 16 novel genetic risk loci and provides insights into disease aetiology beyond the single marker level. *Gut.* 2023 Apr;72(4):612-623. doi: 10.1136/gutjnl-2021-326698. Epub 2022 Jul 26. PMID: 35882562.
3. Klau JH, Maj C, Klinkhammer H, Krawitz PM, Mayr A, Hillmer AM, Schumacher J, Heider D. AI-based multi-PRS models outperform classical single-PRS models. *Front Genet.* 2023 Jun 27;14:1217860. doi: 10.3389/fgene.2023.1217860.
4. Hess T, Maj C, Gehlen J, Borisov O, Haas SL, Gockel I, Vieth M, Piessen G, Alakus H, Vashist Y, Pereira C, Knapp M, Schüller V, Quaas A, Grabsch HI, Trautmann J, Malecka-Wojcieszko E, Mokrowiecka A, Speller J, Mayr A, Schröder J, Hillmer AM, Heider D, Lordick F, Pérez-Aísa Á, Campo R, Espinel J, Geijo F, Thomson C, Bujanda L, Sopena F, Lanás Á, Pellisé M, Pauligk C, Goetze TO, Zelck C, Reingruber J, Hassanin E, Elbe P, Alsabeah S, Lindblad M, Nilsson M, Kreuser N, Thieme R, Tavano F, Pastorino R, Arzani D, Persiani R, Jung JO, Nienhüser H, Ott K, Schumann RR, Kumpf O, Burock S, Arndt V, Jakubowska A, Ławniczak M, Moreno V, Martín V, Kogevinas M, Pollán M, Dąbrowska J, Salas A, Cussenot O, Boland-Auge A, Daian D, Deleuze JF, Salvi E, Teder-Laving M, Tomasello G, Ratti M, Senti C, De Re V, Steffan A, Hölscher AH, Messerle K, Bruns CJ, Siviņš A, Bogdanova I, Skieceviciene J, Arstikyte J, Moehler M, Lang H, Grimminger PP, Kruschewski M, Vassos N, Schildberg C, Lingohr P, Ridwelski K, Lippert H, Fricker N, Krawitz P, Hoffmann P, Nöthen MM,

Veits L, Izbicki JR, Mostowska A, Martínón-Torres F, Cusi D, Adolfsson R, Cancel-Tassin G, Höblinger A, Rodermann E, Ludwig M, Keller G, Metspalu A, Brenner H, Heller J, Neef M, Schepke M, Dumoulin FL, Hamann L, Cannizzaro R, Ghidini M, Plaßmann D, Geppert M, Malfertheiner P, Gehlen O, Skoczytas T, Majewski M, Lubiński J, Palmieri O, Boccia S, Latiano A, Aragones N, Schmidt T, Dinis-Ribeiro M, Medeiros R, Al-Batran SE, Leja M, Kupcinskis J, García-González MA, Venerito M, Schumacher J. Dissecting the genetic heterogeneity of gastric cancer. *EBioMedicine*. 2023 Jun;92:104616. doi: 10.1016/j.ebiom.2023.104616. Epub 2023 May 18. Erratum in: *EBioMedicine*. 2023 Aug;94:104709. doi: 10.1016/j.ebiom.2023.104709. PMID: 37209533; PMCID: PMC10212786.

5. Müller P, Velazquez Camacho O, Yazbeck AM, Wölwer C, Zhai W, Schumacher J, Heider D, Buettner R, Quaas A, Hillmer AM. Why loss of Y? A pan-cancer genome analysis of tumors with loss of Y chromosome. *Comput Struct Biotechnol J*. 2023 Feb 15;21:1573-1583. doi: 10.1016/j.csbj.2023.02.024. PMID: 36874157; PMCID: PMC9978323.
6. Hoppe S, Jonas C, Wenzel MC, Velazquez Camacho O, Arolt C, Zhao Y, Büttner R, Quaas A, Plum PS, Hillmer AM. Genomic and Transcriptomic Characteristics of Esophageal Adenocarcinoma. *Cancers (Basel)*. 2021 Aug 26;13(17):4300. doi: 10.3390/cancers13174300.

Accepted:

Wenzel MC, Giel A-S, Plum PS, Hoppe S, Franitza M, Jonas C, Dasmeh P, Thieme R, Zhao Y, Heider D, Palles C, Fitzgerald RC, Bruns CJ, Buettner R, Quaas A, Gockel I, Ma Cj, Chon S-H, Schumacher J, Hillmer AM. Single cell analysis of Barrett's esophagus and carcinoma reveals cell types conferring risk via genetic predisposition. *Cell Genomics*