gefördert von:

Bundesministerium
für Forschung, Technologie
und Raumfahrt

Zuwendungsempfänger

Deutsches Krebsforschungszentrum

Thema der Förderung

Systemmedizinisches Forschungsnetz zur Früherkennung und Prävention von Leberkrebs LiSyM-Krebs

Konsortialprojekt:

Detaillierte Analyse der räumlichen Organisation der Entstehung des hepatozellulären Karzinoms (DEEP-HCC), Teilprojekt J

Verantwortliche

Prof. Dr. Benedikt Brors
Abt. Angewandte Bioinformatik
Deutsches Krebsforschungszentrum
Im Neuenheimer Feld 280, 69120 Heidelberg
Tel. 06221/42-3614 E-Mail: b.brors@dkfz.de

Förderkennzeichen

031L0258J

Projektlaufzeit:

01.07.2021 - 31.12.2024

Die Verantwortung für den Inhalt dieser Veröffentlichung liegt beim Autor.

II. Eingehende Darstellung

1. Verwendung der Zuwendung und erzielte Ergebnisse im Einzelnen

Das vorliegende Teilprojekt von DEEP-HCC war an den Arbeitspaketen WP-E1 (*Somatic mutations and full transcriptomes*) und WP-M1 (*Genomic mapping to identify the early HCC niche*) beteiligt. Zentrale Ziele waren dabei die Gesamtgenom- und RNA-Sequenzierung von Tumor-, Leber- und Blutproben, von denen extrahierte DNA und RNA über die Projektpartner in der AG Hampe (Uniklinik Dresden) bereitgestellt wurde, sowie die bioinformatische Analyse der resultierenden Daten. Diese umfasste die Qualitätskontrolle der initialen Sequenzierungsdaten, die Detektion somatischer Varianten einschließlich Einzelnukleotidvarianten (SNV), kurzer Insertionen/Deletionen (Indels), Kopienzahlaberrationen, struktureller Varianten und Fusionen. Zusätzlich wurden RNA-Sequenzierungsdaten aligniert, quantifiziert und normalisiert und die resultierenden Expressionsprofile analysiert.

WP-E1 (Somatic mutations and full transcriptomes)

Projektverlauf

Zur Erfüllung der regulatorischen Anforderungen wurde ein Data-Sharing-Abkommen sowie ein Joint-Controller-Vertrag mit dem zentralen Datenmanagement in LiSym-Krebs abgeschlossen. Es wurde ein Zugang zur zentralen RedCap-Datenbank mit den klinischen Daten von HCC-Patienten etabliert.

Zusammen mit Projektpartnern aus WP-E1 (AG Hampe/Brosch, Uniklinik Dresden) wurden Anforderungen an Gewebeproben formuliert, um daraus Gesamtgenom- und Proteomprofile erstellen zu können. Da der Fokus auf frühen Veränderungen lag, sollte auch benachbartes nicht-malignes Lebergewebe mit sequenziert werden. Um darin somatische Mutationen erkennen zu können, war Keimbahn-DNA als Kontrolle erforderlich (für Gesamtgenomsequenzierung), dafür mussten Blutproben desselben Patienten zur Verfügung stehen, aus denen DNA gewonnen werden konnte. Serum- bzw. Plasmaproben eignen sich dafür nicht, da daraus die zellulären Bestandteile des Bluts entfernt wurden, aus denen die DNA gewonnen wird. Zusammen mit der High-Throughput Sequencing Core Facility des DKFZ wurden Qualitätskriterien formuliert, die DNA und RNA schließlich bei unseren Projektpartnern in Dresden extrahiert und per Trockeneisversand an das DKFZ geliefert. Die Nukleinsäure-Extrakte wurden dann mit hoher Abdeckung (Gesamtgenom-Sequenzierung: mindestens 80x, RNA-Sequenzierung: mindestens 50 Millionen Reads, beides paired-end) auf Sequenzierern der Firma Illumina (HiSeq-X bzw. NovaSeq 6000) sequenziert.

Initial wurde ein Pilotexperiment mit insgesamt 18 Gewebeproben von 6 Patienten (6x Tumor, 6x nicht-malignes Lebergewebe, 6x Keimbahnkontrolle aus Blut) durchgeführt, um die Durchführbarkeit und Qualität der extrahierten Nukleinsäuren abzuschätzen. Nach erfolgreicher Durchführung des Pilotexperiments und Auswertung des initialen Datensets (siehe unten, WP-M1) wurde ein weiteres Set mit 30 Gewebeproben von 10 Patienten ausgewählt, die Nukleinsäuren bei den Projektpartnern in Dresden extrahiert, diese ans DKFZ verschickt und dort mit den gleichen Verfahren sequenziert. Ein drittes Set mit Proben von 10 Patienten wies leider nicht die notwendige DNA-Konzentration auf, so dass eine Sequenzierung innerhalb dieser Förderphase nicht mehr durchgeführt werden konnte.

Projektergebnisse

Insgesamt liegt ein Datensatz von Gesamtgenomsequenzierungen von 16 Patienten vor, mit Sequenzdaten von Tumor, benachbartem nicht-malignem Lebergewebe und Keimbahnkontrolle aus

Blutproben. Die Abdeckung beträgt 91 – 144x (Median: 110x). Elf der 16 Patienten wurden in Rostock behandelt, fünf in Dresden. Von diesen 16 Patienten liegt ein weiterer Datensatz mit RNA-Sequenzdaten von Tumor- und Lebergewebe vor (55M – 270M reads, Median: 195M). Die Sequenzdaten sind auf dem Genomics-Cluster des DKFZ im OTP-System gespeichert (https://otp.dkfz.de, aus Datenschutz-Gründen nur innerhalb der Firewall zugänglich) und werden nach Abschluss der bioinformatischen Analysen in das Deutsche Humane Genom-Phänom-Archiv GHGA übertragen (https://ghga.de), wo sie für die Forschergemeinschaft zugänglich sind, sofern ein passendes Ethikvotum vorhanden ist und ein Datennutzungsvertrag abgeschlossen wird.

WP-M1 (Genomic mapping to identify the early HCC niche)

Projektverlauf

Die im Arbeitspaket WP-E1 generierten Daten wurden mit bioinformatischen Algorithmen, die im DKFZ implementiert sind und auch für das PCAWG-Projekt sowie die MASTER- und N2M2-Studien genutzt werden, prozessiert. Das umfasst eine initiale Post-Sequenzierungs-Qualitätskontrolle hinsichtlich der Basenqualität, der Überrepräsentation von *k*-meren, um die Vollständigkeit des Adapter-Trimmings zu belegen, die Analyse des GC-Gehalts, die Duplikationsrate (Sequenzreads mit identischer Start-Position), die Verteilung der Read-Längen und der Read-Qualitätsscores. Adapter-Sequenzen aus Multiplex-Sequenzierungen wurden getrimmt und die Sequenzen mittels bwa-mem (DNA) bzw. STAR (RNA) gegen das humane Referenzgenom gemappt (Version: hg19). Zur Analyse somatischer Mutationen wurden in-house-Pipelines zur Detektion von Einzelnukleotidvarianten (SNV), Indels, Kopienzahlveränderungen und strukturellen Varianten angewendet. Diese wurden weiter nach Befunden hoher mit Hilfe intern entwickelter Qualitätsscores gefiltert. Diese Pipelines sind u.a. in Horak et al. (2021) beschrieben (P. Horak et al. Cancer Discov 11:2780-2795, 2021). Für die Analyse der RNA-Sequenzen wurden Reads pro Gen mit HTseq quantifiziert und zu transcripts-per-million (TPM) Werten normalisiert. Für Leber- und Tumorproben wurden separat die Gene mit der höchsten Varianz bestimmt und die Genexpressionswerte in Heatmaps dargestellt. Die am stärksten differentiell exprimierten Gene wurden berechnet und deren Expressionswerte ebenfalls in Heatmaps dargestellt. Heatmaps beinhalten eine inhärente Clusteranalyse mittels hierarchischer Clusterung (average linkage). Die Genlisten wurden einer Pathway-Enrichment-Analyse mit dem DAVID-Algorithmus unterworfen.

Projektergebnisse

Die Anzahl somatischer SNV-Mutationen mit Konfidenzscore 8-10 betrug bei HCC-Proben zwischen 24 und 161 (median 85,5). Dabei wurden rekurrent Mutationen in typischen Onkogenen gefunden, u.a. in TP53, FAT2, FAT3, ARID2, CTNNB1 oder ERBB3. Auch Mutationen in ALB und APOB wurden detektiert, diese kommen sowohl in HCC als auch im nicht-malignen Lebergewebe vor. Ein auffälliger Befund war die Detektion einer Mutation in FOXO1 (hotspot, S22W), das als rekurrent vorkommend in Lebergewebe, nicht jedoch in HCC beschrieben ist (Ng et al. Nature 598:473-478, 2021). Diese Mutation wird als Reaktion auf metabolischen Stress angesehen (Ng et al. 2021). Interessanterweise konnte diese Mutation nicht im umgebenden nicht-malignen Lebergewebe nachgewiesen werden, auch nicht mit niedriger Allelfrequenz. Möglicherweise handelt es sich hier um eine frühe Mutation in HCC. Eine Übersicht über die am häufigsten detektierten Mutationen bei HCC ist in Abb. 1 in Form eines Onkoprints gegeben.

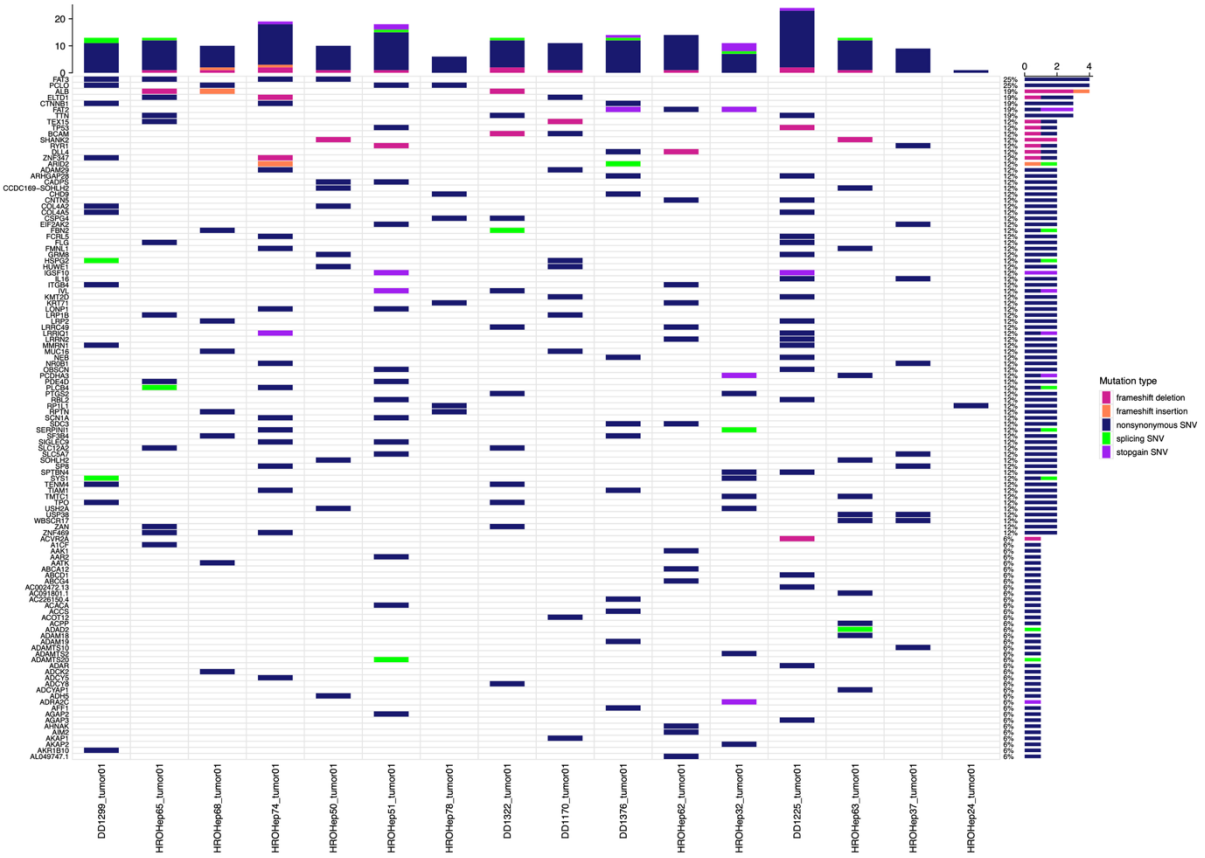

Abbildung 1: Onkoprint-Darstellung der am häufigsten detektierten SNV in HCC-Proben.

Die somatischen Mutationen im nicht-malignen Lebergewebe waren zum einen deutlich seltener, und zeigten zum anderen nur wenig Überlappung mit den Mutationen in HCC-Tumoren. Die Liste der gemeinsam veränderten Gene enthielt u.a. ALB, APOB, AKAP1, NOTCH2, FOXG1, IGFR1, PCLO, ARID1A, DNAH11, BAT1 und FLT4. Eine Enrichment-Analyse mittels DAVID ergab als häufigste betroffene biologische Prozesse u.a. Angiogenese, Entwicklung des Nervensystems, der *insulin-like growth factor* Rezeptor Signalweg, sowie positive Regulation der MAPK-Kaskade. Die Analyse ist aktuell aber noch nicht abgeschlossen, da die Mutationen in den jeweiligen Genen noch hinsichtlich ihrer möglichen Auswirkungen auf die Proteinfunktion oder -struktur bewertet werden und Allelfrequenzen verglichen werden müssen. Die häufigsten Veränderungen im Lebergewebe sind in Abb. 2 gezeigt.

Kopienzahlveränderungen waren hinsichtlich der Fragestellung nicht sehr aufschlussreich. Abb. 3 zeigt exemplarisch das Kopienzahlprofil einer Leber- und zugehörigen HCC-Probe. Alle Leberproben zeigten ein diploides Profil, unter Berücksichtigung der Vertauschung von Leber- und Tumorproben bei zwei Patienten. Die HCC-Proben zeigten Tumor-typische Veränderungen mit Amplifikation von Onkogenen und Deletionen von Tumorsuppressorgenen. Eine Analyse der rekurrenten Veränderungen wurde nicht durchgeführt, da sie bei der gegebenen Fragestellung nur hilfreich gewesen wäre, wenn auch Leberproben Aneuploidie gezeigt hätten.

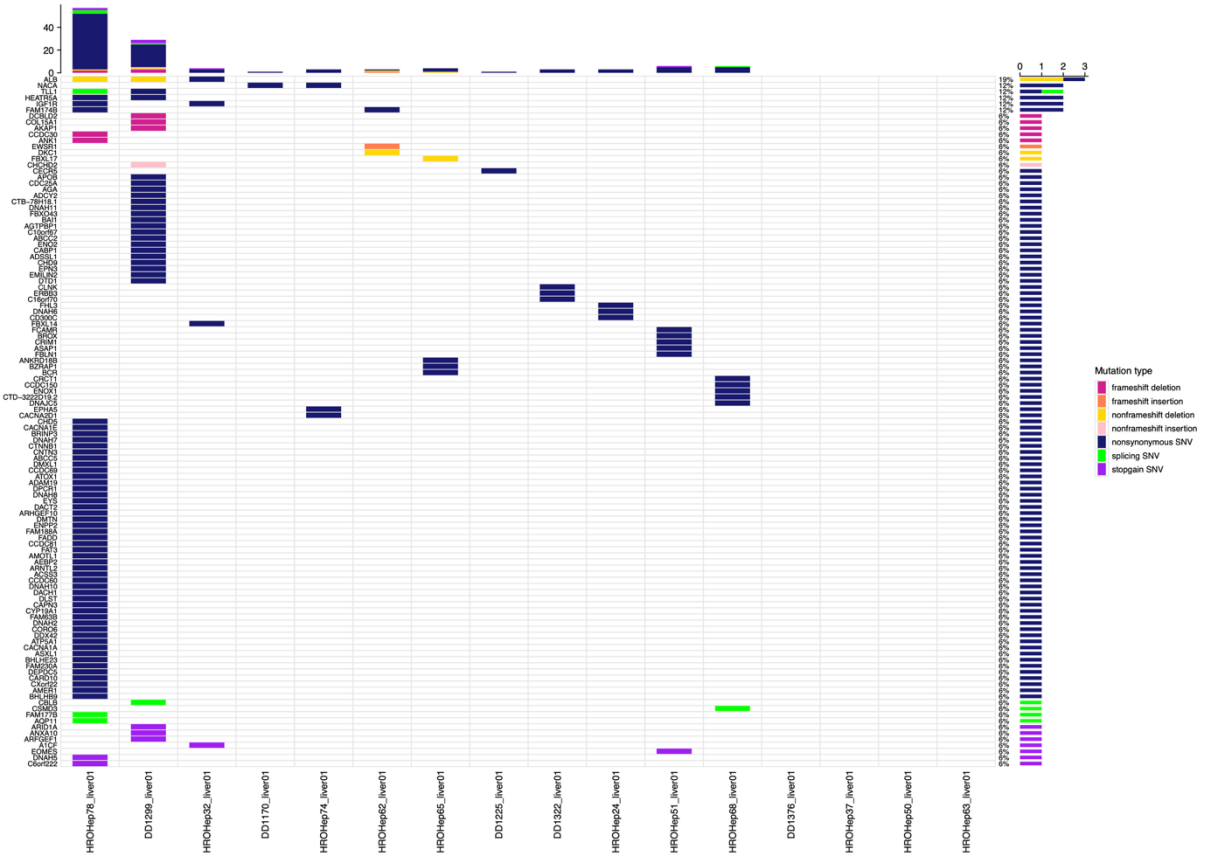

Abbildung 2: Onkoprint-Darstellung der am häufigsten detektierten SNV in nicht-maligenen Leber-Proben.

Die Analyse der RNA-Expressionswerte ergab, wie erwartet, deutliche Unterschiede zwischen nicht-malignem Lebergewebe und Tumorproben (Abb. 4). Die Gene mit der höchsten Varianz innerhalb von nicht-malignem Lebergewebe und innerhalb von HCC-Proben sind in Abb. 5 gezeigt. Die Analyse dieser Daten ist noch nicht abgeschlossen, aber erste Hinweise deuten auf unterschiedliche Aktivitäten von Genen, die an Entzündungs- und Infektionswegen beteiligt sind, innerhalb von Leber- und HCC-Gewebe. Es lässt sich auch erkennen, dass die HCC-Proben in mindestens 2, vielleicht auch 3 Cluster geteilt sind, die unterschiedliche Expressionsprofile aufweisen (Abb. 4 und 5). In Abb. 4 sind die beiden sehr unterschiedlichen Clustern durch die unterschiedliche Aktivität von Genen gekennzeichnet, die maßgelbliche Kontrolle über die Zellproliferation ausüben (CCNA2, CCNB1, CCNB2, CDKN3, MKI67, TOP2A). Der weitere Abgleich dieser Cluster mit beschriebenen Subgruppen (z.B. aus TCGA) ist in Arbeit.

Die Analyse hinsichtlich des Timings der Mutationen anhand von Allelfrequenzen mit Hinsicht auf Kopienzahlveränderungen konnte bislang noch nicht abgeschlossen werden, ebenso steht die Analyse des Einflusses verschiedener Mutationssignaturen, die auf Aktivitäten von mutations-generierenden Prozessen hindeuten können, noch aus. Grund dafür war, dass diese Analysen aufgrund der begrenzten statistischen Power stark von einer erhöhten Stichprobengröße profitiert hätten. Bis zum Abschluss der Förderperiode stand im Raum, dass ggf. ein weiteres Set an Tumor- und nicht-malignen Gewebeproben hätte sequenziert werden können, was sich dann aufgrund geringer DNA-Konzentrationen bzw. nachfolgendem Ablauf der Föderperiode nicht mehr realisieren ließ. Diese

Analysen werden aktuell außerhalb von LiSym-Krebs fortgesetzt und werden in ein Manuskript einfließen, das anschließend publiziert werden soll.

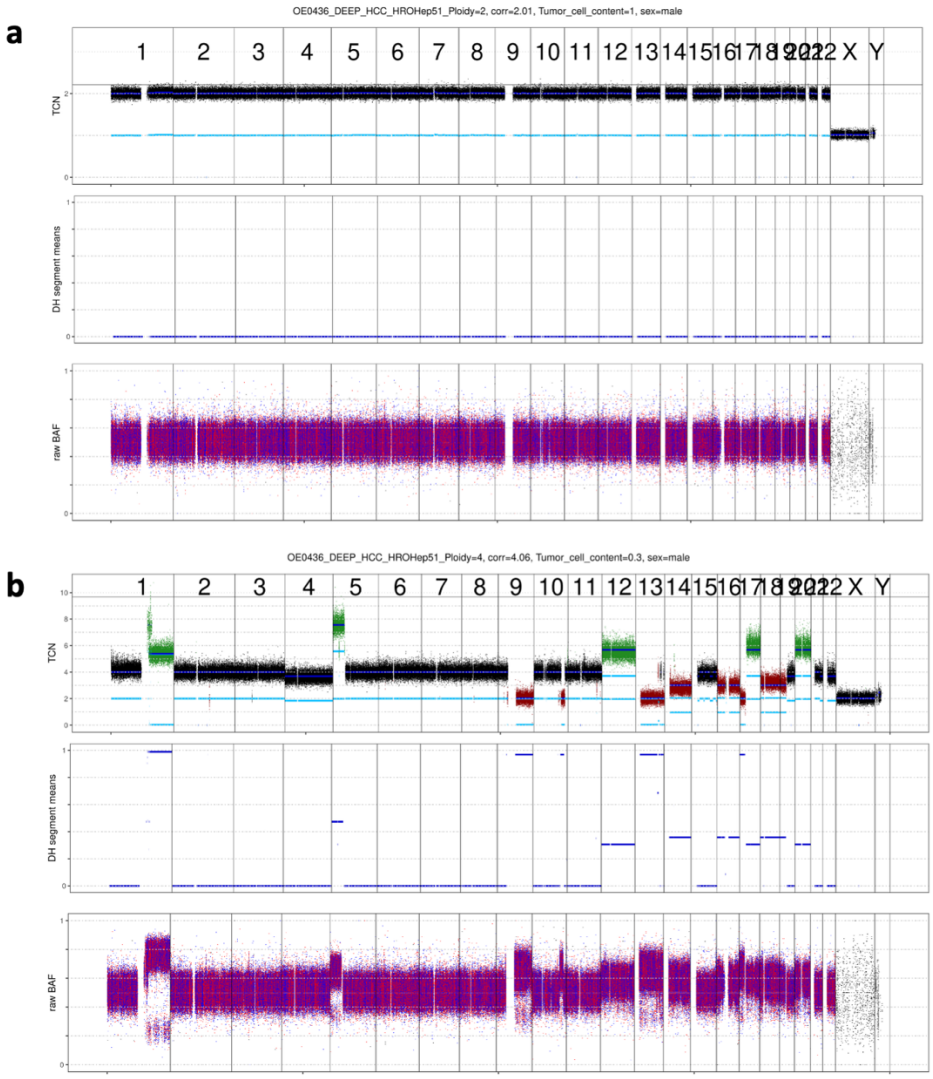

Abbildung 3: Kopienzahlprofil (Tumor-Kopienzahl, TCN; Segmentierung; B-Allel-Frequenz, BAF) für nicht-malignes Lebergewebe (a) und HCC-Gewebe (b) eines exemplarischen Patienten (HRO51).

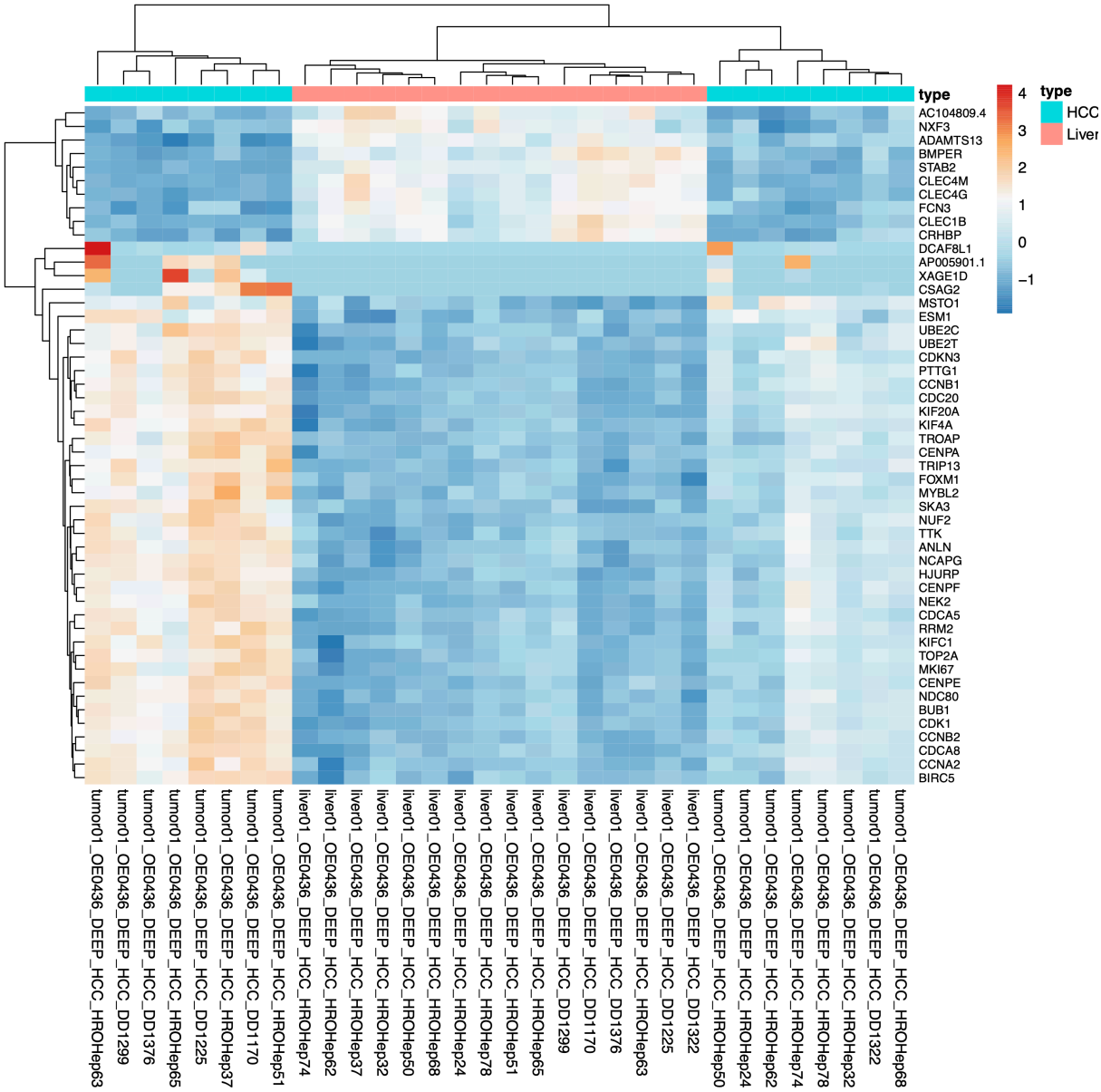

Abbildung 4: Heatmap der Genexpressionswerte für die am meisten differentiell exprimierten Gene zwischen Tumor- (türkis) und nicht-malignem Lebergewebe (hellrot). Dargestellt sind Z-Scores.

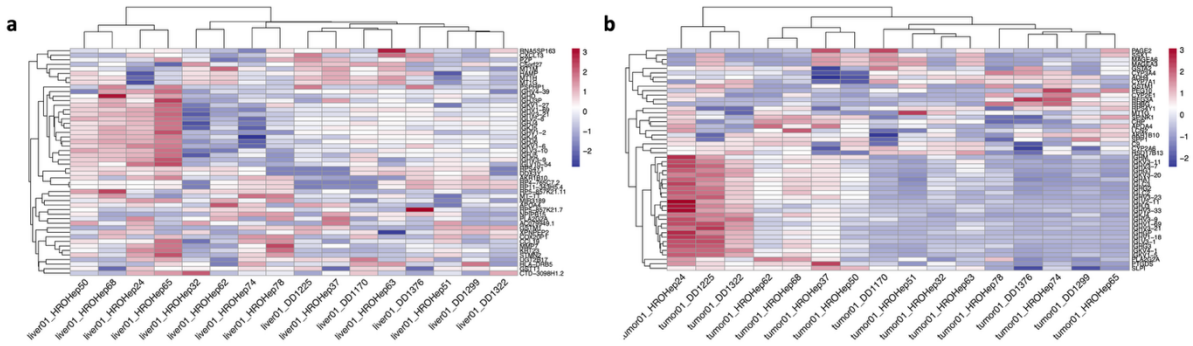

Abbildung 5: Heatmap der Genexpressionswerte für die Gene mit der höchsten Varianz innerhalb von nicht-malignem Lebergewebe (a) oder einer HCC-Probe (b). Dargestellt sind Z-Scores.

2. Positionen des zahlenmäßigen Nachweises

Die im Rahmen des Projektes bewilligten Personalmittel wurden für die Finanzierung der Wissenschaftlerstelle verwendet. Die Reisemittel wurden für die Teilnahme an LiSym-Projekttreffen und die Zwischenevaluation von LiSym eingesetzt. Weiterhin wurden Mittel im Rahmen der internen Leistungsverrechnung mit der DKFZ High-Throughput-Sequencing Core Facility verwendet, um die DNA und RNA aus Tumor-, Leber- und Blutproben zu sequenzieren.

3. Notwendigkeit und Angemessenheit der geleisteten Arbeit

Die Arbeiten wurden wie geplant durchgeführt. Wie im Antrag dargestellt waren sie für die Durchführung des Projektes notwendig und angemessen.

4. Voraussichtlicher Nutzen und Verwertbarkeit der Ergebnisse

Die Ergebnisse werden nach Abschluss der bioinformatischen Analysen über die interne Datenmanagement-Plattform https://seek.lisym.org bereitgestellt und können für weiterführende Analysen und Information von Modellierungsansätzen in der nächsten Förderphase von LiSym-Krebs eingesetzt werden. Die Rohdaten werden, wie oben genannt, über GHGA archiviert und stehen unter einem Modell des kontrollierten Zugangs zur Verfügung, bei dem die besondere Natur der Daten, die immer Personenbezog im Sinne der Datenschutzgrundverordnung haben, berücksichtigt wird.

5. Bekannt gewordener Fortschritte auf dem Gebiet des Vorhabens

Durch die AG von Peter Campbell (Sanger Centre, Hinxton, UK) wurde in 2021 eine Untersuchung veröffentlicht, bei der insgesamt 1590 mikrodissezierte Leberproben von 34 Patienten mit Lebererkrankungen mittels Gesamtgenomsequenzierung charakterisiert wurden (Ng et al. Nature 598:473-478, 2021). Dabei konnten verschiedene somatische Mutationen gefunden werden, die offenbar unter positiver Selektion standen und als Anpassungen an metabolischen oder inflammatorischen Stress in der Leber gedeutet wurden. Einige dieser Mutationen treten nie oder selten bei HCC auf, unter anderem die S22W Mutation in FOXO1, die in dem Artikel genauer funktionell charakterisiert wird. In dieser Studie waren allerdings keine korrespondierenden HCC-Tumore aus denselben Patienten verfügbar, so dass die Rolle der beschriebenen Veränderungen bei der Genese von HCC ungeklärt blieb. In einer Arbeit von Nguyen et al. (Nguyen et al. Nat Commun 13:441, 2022) wurde die Einzelzellzusammensetzung von HCC, nicht-malignem Lebergewebe und peripherem Blut mittels CyTOF-Massenspektrometrie und Einzelzell-RNA-Sequenzierung untersucht. Dabei konnte eine graduelle Abnahme antitumorigener Zelltypen und eine Zunahme nicht-funktionaler oder immunsuppresiver Immunzellen im Tumorverlauf beobachtet werden, wobei eine Re-Immunisierung bei den höchstgradigen Tumoren beobachtet wurde. Hier standen allerdings keine longitudinalen Tumorproben zur Verfügung, sondern Proben verschiedener Patienten wurden – recht grob – anhand der auf dem TNM-System basierenden Stadieneinteilung dem vermuteten Tumorverlauf zugeordnet. Trotz regelmäßiger Literaturrecherchen in der Datenbank PubMed konnten keine weiteren Artikel mit erheblicher Bedeutung für das aktuelle Teilprojekt in diesem Vorhaben identifiziert werden. Die oben genannten Erkenntnisse werden zur Interpretation der erhaltenen Ergebnisse mit herangezogen.

6. Veröffentlichungen

Ein Manuskript zu den Ergebnissen der Sequenzierung und der bioinformatischen Analyse ist in Vorbereitung.