

# Schlussbericht

LiSyM

Verbundprojekt: Pillar I – Metabolischer Frñhschaden (EMI)

Teilprojekt 2

FKZ 031L0032

Universitätsmedizin Greifswald

Prof. Dr. Lars Kaderali

## I. Kurze Darstellung

### 1. Aufgabenstellung

Ziel des Gesamtverbundes LiSyM ist die Translation systembiologischer Verfahren auf klinische Fragestellungen am Beispiel der Leber und unter Nutzung der umfangreichen Vorarbeiten aus dem Netzwerk „Die virtuelle Leber“. Der Verbund „LiSyM - Verbundprojekt: EMI - Metabolischer Frñhschaden“ adressiert das klinische Problem der nichtalkoholischen Fettlebererkrankung (englisch NAFLD) – und hier spezifisch die Ereignisse, die die Transition von benigner Steatose zur Nekroinflammation als krankheitdefinierendes Ereignis für die nichtalkoholische Steatohepatitis (englisch: NASH) bedingen. Aufbauend auf den Steatose- und LIAM-Showcases und der klinischen systempharmakologischen Studie im Netzwerk „Die virtuelle Leber“ (VLN), wurden in LiSyM\_EMI diese frühen Krankheitsereignisse multidimensional erforscht und modelliert. Ziel war die Entwicklung von Modellen, die das molekulare Verständnis der Ereignisse dieses „metabolischen Frñhschaden“ erlauben. Dabei wurden Tiermodelle und humane Proben der im VLN aufgebauten Biobank genutzt. In der klinischen Studie wurde die systempharmakologische Modellierung auf Patienten mit NAFLD erweitert.

Aufgabe des Teilprojektes 2 (Universitätsmedizin Greifswald) war die integrative Bioinformatik-Analyse der im Verbundprojekt LiSyM\_EMI erhobenen Daten. Dabei wurden multi-OMICs-Methoden und maschinelle Lernverfahren eingesetzt, um prädiktive Signaturen zu entwickeln und zentrale molekulare Prozesse der Krankheitsprogression zu identifizieren. Mit dem Wechsel des Projektleiters des TP 2 von der Technischen Universität Dresden an die Universitätsmedizin Greifswald (nach Antragstellung) ergab sich zusätzlich die Möglichkeit, umfangreiche Daten der Epidemiologischen Study of Health in Pomerania (SHIP) der Universitätsmedizin Greifswald in die Analysen in LiSyM\_EMI zu integrieren.

## 2. Voraussetzungen, unter denen das Vorhaben durchgeführt wurde

Die Entwicklung des systembiologischen Methodenarsenals hatte in den letzten Jahren vor Projektbeginn erhebliche Fortschritte gemacht, sowohl auf der experimentellen als auch auf der methodischen Seite. Neben Verfahren für eine systematische Datengenerierung (genomische, epigenetische, lipidomische) haben auch Modellierungstechniken und Methoden der Quantitativen Biologie breiteren Einzug in die lebenswissenschaftliche Methodik genommen, maßgeblich auch getrieben durch Entwicklungen im Rahmen des VLN Verbundes, auf den mit LiSyM aufgebaut wurde. In der Summe und zusammen mit den Modellierungsanstrengungen des Lebermetabolismus (einschließlich VLN) lagen daher zu Projektstart die Voraussetzungen für eine modellhafte Analyse des Lebermetabolismus vor.

Die Pathogenese der Manifestationen metabolischer Erkrankungen (insbesondere des Typ 2 Diabetes) in Form einer Fettleberhepatitis an der Leber war (und ist) bisher nur unvollständig verstanden. Bisherige Studien nutzten selten systembiologische oder genomweite Ansätze und waren oft auf einzelne Gene und Pathomechanismen fokussiert. Die Vereinigung dieser beiden Ansätze – d.h. die Anwendung moderner systembiologischer Analysen – im Kontext der erwähnten Erkrankungen lässt auf neue Erkenntnisse zur Pathogenese und die Identifizierung neuer Biomarker und molekularer Zielmoleküle für die Therapie hoffen. Diese Ansätze sollten im LiSyM Verbundprojekt kombiniert werden und zum Verständnis der molekularen Prozesse bei Progression von benigner Steatose zur Nekroinflammation als krankheitdefinierendes Ereignis für die nichtalkoholische Steatohepatitis (englisch: NASH) angewendet werden, um wesentliche molekulare Prozesse aufzuklären.

## 3. Planung und Ablauf des Vorhabens

Aufgabe („Deliverable“)	Projektmonat										
	0-6	6-12	12-18	18-24	24-30	30-36	36-42	42-48	48-54	54-60	61-66
Del 1: Integrative OMICS-Datenanalyse	X	X	X	X							
Del 2: Kombinatorische, prädiktive Biomarker		X	X	X	X						
Del 3: Netzwerkinferenz			X	X	X	X	X	X			
Del 4: Identifikation von Schlüsselprozessen des Krankheitsfortschritts			X	X	X	X	X	X	X	X	X
Del 5: Validierter Biomarker-Score						X	X	X	X	X	X
Del 6: Adaption Deep Learning							X	X			
Del 7: SHIP / GANI_MED							X	X	X	X	X

Die Arbeitsplanung für Teilprojekt 2 erfolgt in mehreren Arbeitsschritten, die im Balkenplan oben dargestellt sind. Für den ersten Förderabschnitt (M1-M36) waren 5 Arbeitsschritte vorgesehen, die sich mit der integrativen Analyse der im LiSyM-Verbund erhobenen Daten, der Ableitung von kombinatorischen Biomarker-Signaturen sowie der Netzwerkinferenz zur Identifikation von

Schlüsselprozessen des Krankheitsfortschritts befassen. In der zweiten Projekthälfte wurde dann vermehrt auf die Anwendung fortgeschrittenerer maschineller Lernverfahren fokussiert, und durch den Wechsel des Hauptantragstellers an die Universitätsmedizin Greifswald konnte zusätzlich auf Daten der SHIP-Studie zugegriffen werden, die für weitergehende Analysen genutzt und verstärkt in die Arbeiten in LiSyM eingebunden wurden.

Die Arbeiten konnten im Großen und Ganzen wie geplant durchgeführt werden. Durch die Coronavirus-Pandemie kam es durch Lockdown-Maßnahmen und Umstellung auf Homeoffice zu Verzögerungen in den Meilensteinen 4, 5 und 7, die durch die Verlängerung des Projektes bis 31.6.2021 (M61-66) aufgefangen werden konnten. Insgesamt konnte das Projekt damit erfolgreich zu Ende geführt werden.

#### **4. Wissenschaftlicher und technischer Stand, an den angeknüpft wurde**

LiSyM\_EMI Teilprojekt 2 fokussierte auf die integrative Bioinformatik-Analyse der in LiSyM erhobenen Daten, und ergänzte diese mit Daten der Greifswalder Study of Health in Pomerania (SHIP), die außerhalb des LiSyM-Verbundes erhoben wurden, aber für ergänzende Analysen im Rahmen von LiSyM zur Verfügung standen.

Im Bereich der OMICs-Datenanalyse, der integrierten Analyse von klinischen und klinik-nahen Daten und auch im Bereich der Anwendung von Methoden des Data Mining und der künstlichen Intelligenz auf biomedizinische Daten sind in den letzten Jahren vor Projektstart erhebliche methodische Fortschritte erreicht worden. Netzwerkbasierte Verfahren wurden unter anderem vom Antragsteller entwickelt um die integrative Auswertung von Hochdurchsatz-/OMICs-Daten zu ermöglichen (z.B. Smoot 2011, Wang 2014, Alcaraz 2014, Amberkar 2015). Maschinelle Lernverfahren und Netzwerkinferenz-Methoden ermöglichen nun die automatisierte Rekonstruktion von regulatorischen und Signaltransduktionsnetzwerken, und können Hypothesen über mechanistische Zusammenhänge generieren (Prill 2011, Marbach 2010, Mator 2015, Kiani 2014, Böck 2012). Diese Verfahren wurden im vorliegenden Projekt für die Untersuchung von NAFLD angepasst und zur Analyse der Projektdaten verwendet, mit dem Ziel prädiktive Signaturen zu entwickeln und zentrale molekulare Prozesse der Krankheitsprogression zu identifizieren und deren Zusammenhang zu verstehen.

Eine regelmäßige Literaturrecherche erfolgte über die Datenbanken „PubMed“ und „Scopus“, darüber hinaus wurden neue methodische Entwicklungen im Bereich der Bioinformatik-Datenanalyse regelmäßig über Fachliteratur und wissenschaftliche Tagungen verfolgt und in den Entwicklungen in LiSyM berücksichtigt.

#### **5. Zusammenarbeit mit anderen Stellen**

Im Rahmen des Konsortiums „LiSyM“ erfolgte eine enge Zusammenarbeit mit den übrigen Projektpartnern, insbesondere innerhalb des „Pillar I“ Verbundes LiSyM\_EMI. Eine sehr enge Kooperation erfolgte mit der experimentellen / klinischen Arbeitsgruppe von Jochen Hampe (TU Dresden), deren Daten maßgeblich in den Analysen in diesem Teilprojekt ausgewertet wurden (Deliverables 1-5).

Außerhalb des LiSyM-Verbundes wurde eng mit dem Institut für Community Medicine in Greifswald (Prof. Völzke) im Zusammenhang mit der Auswertung der SHIP-Studie zusammengearbeitet.

Während der Projektlaufzeit erfolgte eine intensive Zusammenarbeit zwischen allen Verbundpartnern. Während regelmäßiger Projekttreffen wurden die erzielten Ergebnisse und der weitere Fortgang des Projektes diskutiert und abgestimmt. In mehreren persönlichen Treffen in Dresden und regelmäßigen Video- und Telefonkonferenzen erfolgte eine enge Abstimmung mit den unmittelbaren Partnern in einzelnen Arbeitspaketen.

## II. Eingehende Darstellung

### 1. Verwendung der Zuwendung und des erzielten Ergebnisses im Einzelnen

Im Rahmen des LiSyM-Verbundprojektes sind in Teilprojekt 2 vor allem Personalausgaben angefallen / gefördert worden. Über die gesamte Projektlaufzeit wurde ein wissenschaftlicher Mitarbeiter (Doktorand) mit einer 50% Stelle nach TV-L E13 finanziert. Dieser wissenschaftliche Mitarbeiter hat die im Antrag skizzierten Aufgaben wie in der folgenden Darstellung zusammengefasst bearbeitet:

#### Meilenstein 1: Integrative OMICS-Datenanalyse (M0-M24)

**Arbeitsbeschreibung gemäß Antrag:** *Wir werden unsere etablierten Verfahren zur integrativen, netzwerkbasierten Datenanalyse für die im Projekt erhobenen Daten geeignet anpassen (Amberkar 2013 & 2015) um die verschiedenen, im Projekt erhobenen OMICS-Daten zu integrieren und gemeinsam auszuwerten. Anschließend werden die im Verbundprojekt erhobenen experimentellen Daten ausgewertet um zentrale molekulare Prozesse auf regulatorischer und metabolischer Ebene zu identifizieren, die mit der Progression der Erkrankung assoziiert sind.*

**Durchgeführte Arbeiten:** Die Anpassung der Datenanalysepipeline wurde mit bereits publizierten Daten zur Fettleber mit Proben aus Lebergewebe (Ahrens et al., 2013), sowie auf Daten der Greifswalder SHIP Studie in Vollblut (Study of Health in Pomerania, Völzke et al., 2011) bereits ab Projektstart durchgeführt. Eine Datenanalysepipeline wurde in R aufgesetzt, und zur Analyse aller in der Folge im Projekt angefallenen Daten angewendet. Alle von den Projektpartnern erhalten Daten wurden ausgewertet. Daten der Arbeitsgruppe Hampe (TU Dresden) zur Zonierung der Leber wurden ausgewertet; diese Daten umfassen Transkriptom-Daten aus drei Zonen der Leber (Perizentral, Periportal, sowie mittlerer Bereich dazwischen) aus drei Patientengruppen (Frühe Steatohepatitis, Steatose, Gesund übergewichtige) plus normale Kontrollen. Ergebnisse der Analysen (gerankte Liste signifikanter Gene) wurden an die Arbeitsgruppe Hampe weitergegeben, sowie in Meilenstein 2 weiterverwendet. **Die Arbeiten zum Meilenstein 1 wurden entsprechend den Planungen im Projektzeitraum abgeschlossen.**

#### Meilenstein 2: Kombinatorische, prädiktive Biomarker (M6-M30)

**Arbeitsbeschreibung gemäß Antrag:** *Basierend auf den Ergebnissen von Arbeitsschritt 1 werden wir kombinatorische, prädiktive Biomarker auf Netzwerkebene identifizieren, die die Krankheitsprogression vorhersagen können. Diese werden an die Projektpartner zur Validierung weitergegeben.*

**Durchgeführte Arbeiten:** Basierend auf der gerankten Genliste aus Meilenstein 1 wurden Netzwerkanalysen durchgeführt. Diese deuten auf eine wichtige Rolle der Prozesse „Bile secretion“ sowie des „Hedgehog Signalling Pathways“ hin. Während eine Analyse einzelner Biomarker (univariate Analyse) keine Probleme bereitete und signifikante Ergebnisse ergab, zeigten sich bei der integrativen, netzwerkbasierten Datenanalyse Grenzen durch die geringe Anzahl an Leberbiopsien, die zur Verfügung standen. Zwar ist ein Ranking der Pathways möglich, jedoch mit nur eingeschränkter statistischer Signifikanz der Ergebnisse. Da größere Zahlen an Leberbiopsien nur schwer zu generieren sind, wurde in der zweiten Projektphase zur Validierung und weiteren

Datenanalyse verstärkt auf Daten der SHIP-Studie zurückgegriffen. Dabei sollte untersucht werden, ob sich in Blutproben der SHIP-Probanden ähnliche Signaturen wie in den Leberproben finden lassen, die zumindest als diagnostisches Proxy geeignet sind ohne die Erfordernis einer Leberbiopsie (siehe Meilenstein 7). Darüber hinaus wurden Analysen mit fortgeschrittenen maschinellen Lernverfahren durchgeführt, um Ergebnisse der Pathway-Analysen mit anderen Methoden zu validieren (siehe Meilenstein 3). **Die Arbeiten zum Meilenstein 2 wurden entsprechend den Planungen abgeschlossen.**

#### Meilenstein 3: Netzwerkinferenz (M12-M36)

**Arbeitsbeschreibung gemäß Antrag:** *Um die Dynamik der molekularen Interaktionen im systembiologischen Sinne bei fortschreitender Erkrankung zu untersuchen werden wir in diesem Arbeitsschritt Netzwerkinferenz-Methoden anwenden, die das Zusammenspiel zwischen verschiedenen Biomarkern und in Arbeitsschritt 1 identifizierten Prozessen untersuchen (Matos 2015, Kiani 2014, Knapp 2013, Böck 2012). Die Verfahren müssen hierbei angepasst werden, um metabolische Prozesse mit regulatorischen / Signaltransduktionsprozessen integrieren zu können.*

**Durchgeführte Arbeiten:** Aufbauend auf den Ergebnissen der Pathway-Analyse in Meilenstein 2 wurden weitere Analysen mit regularisierten Bayes'schen Netzen durchgeführt. Hierbei konnten die Ergebnisse aus Meilenstein 2 zur Bedeutung des „Bile secretion“ und „Hedgehog Signalling“ für die Progression der Fettleber bestätigt werden. Allerdings zeigten sich auch hier Probleme durch Overfitting der Modelle durch die geringe Fallzahl und daraus resultierend die Notwendigkeit, die Methoden weiter zu verfeinern, insbesondere auch im Hinblick auf die Anwendung der Methoden auf Daten der SHIP Studie (Meilenstein 7).

Dieses Problem wurde durch explizit an die spezifischen Eigenschaften der Netzwerke angepasste Regularisierungsmethoden (Einbinden eines Strafterms für komplexe Modelle in der Zielfunktion) sowie die Integration biologischen Vorwissens aus Pathway-Datenbanken (z.B. zur Topologie einzelner Signaltransduktions oder genregulatorischer Netzwerke) behoben. Die Methoden wurden angepasst, um mit den in Greifswald zusätzlich verfügbaren Daten aus der SHIP Studie zu arbeiten, und insbesondere klinische Parameter einzubinden. Darüber wurde an der Einbindung von Verfahren zur Imputation fehlender Werte sowie zur Integration von sehr unterschiedlich verteilten Daten (z.B. Genexpressiondaten, sozioökonomische Daten, Laborparameter) gearbeitet. Diese Probleme treten vor allem bei Daten der SHIP und Gani\_Med Studien auf, die in die Analysen miteinbezogen wurden (Meilenstein 7). Multi-OMICs-Daten wurden in die Analysen einbezogen und dafür erforderliche methodische Anpassungen durchgeführt. Die Methoden wurden in Meilenstein 7 verwendet, um SHIP/GANI\_MED Daten zu analysieren. Ergebnisse der Methodenentwicklung sowie Resultate auf SHIP-Daten zur Fettleber konnten in einer wissenschaftlichen Publikation in der Fachzeitschrift PLOS Computational Biology veröffentlicht werden (Becker AK, Dörr M, Felix SB, Frost F, Grabe HJ, Lerch MM, Nauck M, Völker U, Völzke H, **Kaderali L**. *From heterogeneous healthcare data to disease-specific biomarker networks: A hierarchical Bayesian network approach*. **PLoS Comput Biol.** 2021 Feb 12;17(2):e1008735).

**Die Arbeiten zum Meilenstein 3 wurden entsprechend den Planungen abgeschlossen.**

#### Meilenstein 4: Identifikation von Schlüsselprozessen des Krankheitsfortschritts (M12-M36)

**Arbeitsbeschreibung gemäß Antrag:** *Basierend auf den Ergebnissen der Arbeitsschritte 1-3 und in enger Kooperation mit den biologischen und klinischen Arbeitspartnern werden wir zentrale Schlüsselprozesse des Krankheitsfortschritts identifizieren. Diese werden an die experimentellen Partner weitergegeben zur Validierung, sowie an die Modellierungspartner zum Einbau in die systembiologischen Modelle.*

**Durchgeführte Arbeiten:** Die Ergebnisse aus den Analysen mit regularisierten Bayes'schen Netzen aus Meilenstein 3 stützen die Analyseergebnisse aus den Meilensteinen 1 und 2, dass „Bile Secretion“ und „Hedgehog Signalling“ eine wichtige Rolle bei der Progression der nichtalkoholischen Fettlebererkrankung spielen. Die Ergebnisse wurden an die Projektpartner übermittelt.

Weitere Analysen und methodische Verfeinerungen wurden durchgeführt, um zusätzliche OMICS-Daten in den Analysen verarbeiten zu können. Dabei wurden als weiterer Datensatz Plasma-Proteom und Urin-Metabolomdaten aus der SHIP-Studie integriert. Ein besonderes Augenmerk musste dabei auf die Vorprozessierung der SHIP-Metabolom-Daten gelegt werden, da hier in nicht unerheblichem Umfang fehlende Werte in den Messdaten auftreten; zudem erfordert die Anwendung der Bayes'schen Netze eine Diskretisierung der kontinuierlichen Daten, um die Netzwerkstruktur lernen zu können. Hierfür wurden verschiedene Verfahren getestet, die besten Ergebnisse konnten mit k-nearest-Neighbor Imputation zur Behandlung fehlender Werte erzielt werden; die Diskretisierung der Daten hingegen erforderte aufwändige manuelle Analysen und konnte nicht automatisiert erfolgen. Die entwickelten Methoden und prozessierten Daten wurden in Meilenstein 7 (Analyse der SHIP-Daten) weiter prozessiert.

In Zusammenarbeit mit Partnern außerhalb des LySiM-Netzwerkes konnten zudem weitere Patientenproben von Patienten mit unterschiedlichen Lebererkrankungen, darunter die Nichtalkoholische Fettleber in verschiedenen Stadien, gewonnen werden, um den Einfluss von microRNAs in diesen Erkrankungen zu untersuchen. Diese Daten konnten ergänzend zu den Daten der SHIP Studie untersucht werden, und zeigen einen Cluster von microRNAs einschließlich der mir195, mir182 und mir23a, die sich als nützlich zur Diagnose der Steatose erwiesen. Ergebnisse dieser Arbeiten konnten in einer wissenschaftlichen Fachzeitschrift veröffentlicht werden (Gholizadeh, M.; Szlag-Pieniek, S.; Post, M.; Kurzawski, M.; Prieto, J.; Argemi, J.; Drozdziak, M.; Kaderali, L. Identifying Differentially Expressed MicroRNAs, Target Genes, and Key Pathways Deregulated in Patients with Liver Diseases. *Int. J. Mol. Sci.* 2020, 21, 7368).

**Die Arbeiten zum Meilenstein 4 wurden entsprechend den Planungen im Berichtszeitraum abgeschlossen.**

#### Meilenstein 5: Validierter Biomarker-Score (M30-M36)

**Arbeitsbeschreibung gemäß Antrag:** *In diesem Arbeitsschritt wird ein finaler, prädiktiver Biomarker-Score (Blut und Gewebe) zusammengestellt und anhand der verfügbaren Daten statistisch validiert.*

**Durchgeführte Arbeiten:** Basierend auf den Ergebnissen des Meilenstein 3 und 4 wurde ein kombinierter Biomarker-Score basierend auf den SHIP-Daten berechnet und mit anderen Leber-Risikoscores verglichen (Fatty Liver Index, Hepatic Steatosis Index, NAFLD Ridge Score), die alle zwischen 3 und 6 anthropometrische Parameter und biochemische Tests kombinieren. Auf den SHIP-Daten konnten diese publizierten Scores AUROC (Area under the Receiver Operator Characteristic

Curve)-Werte zwischen 0.67 und 0.78 und AUPR (Area under the Precision-Recall Curve)-Werte zwischen 0.24 und 0.34 erreichen. Mit dem in diesem Projekt entwickelten Score konnte ein AUROC-Wert von 0.82 sowie ein AUPR-Wert von 0.42 erzielt werden, und damit eine deutliche Verbesserung über etablierte Scores. Diese Ergebnisse konnten in PLOS Computational Biology publiziert werden (Becker AK, Dörr M, Felix SB, Frost F, Grabe HJ, Lerch MM, Nauck M, Völker U, Völzke H, **Kaderali L**. *From heterogeneous healthcare data to disease-specific biomarker networks: A hierarchical Bayesian network approach*. **PLoS Comput Biol**. 2021 Feb 12;17(2):e1008735).

**Die Arbeiten zum Meilenstein 5 wurden entsprechend den Planungen abgeschlossen.**

#### Meilenstein 6: Adaption Deep Learning (M36-48)

**Arbeitsbeschreibung gemäß Antrag:** *In diesem Arbeitsschritt sollen Deep Learning Verfahren angepasst werden, um eine Analyse von Daten der SHIP und Gani\_Med Studien zu ermöglichen. Diese Studien sind charakterisiert durch eine sehr breite molekulare und klinische Untersuchung der Studienteilnehmer, und ermöglichen daher eine systemische Untersuchung (organ-to-organ crosstalk). Deep-Learning bezeichnet „tiefe“, modulare neuronale Netze; hierbei können Teilbereiche der neuronalen Netze verwendet werden, um Teilprobleme eines Klassifikationsproblems zu lösen. In diesem Meilenstein sollen solche Module verwendet werden, um Parameter zu verschiedenen Organen und/oder molekularen Pathways innerhalb eines Moduls zu gruppieren, und deren Einfluss auf die Progression der Fettleber zu untersuchen – das neuronale Netz lernt dabei diese Zusammenhänge automatisch.*

**Durchgeführte Arbeiten:** Im Rahmen einer am Institut für Bioinformatik der Universitätsmedizin Greifswald unter Betreuung des Projektleiters und im Kontext von LiSyM durchgeführten Masterarbeit wurden Deep-Learning-Verfahren zur Analyse der SHIP-Daten eingesetzt. Ein entsprechendes tiefes neuronales Netz wurde aufgesetzt, und auf Daten der SHIP Studie zur Fettleber angewendet. Dabei lag der Fokus der Arbeit auf der Optimierung der Topologie des Neuronalen Netzes, um Pathwayinformationen aus der KEGG-Datenbank in die Analyse einbinden zu können. Ein wesentliches Ergebnis der Arbeit ist, dass die Einbindung von Pathway-Informationen in die Topologie des neuronalen Netzes zu Verbesserungen der Vorhersagegüte führt, allerdings geht im Vergleich zu „klassischen“ Nicht-Neuronale-Netzwerk-basierten Methoden (Meilensteine 1-4) die Interpretierbarkeit der Modelle dabei verloren, d.h. ein einfacher Rückschluss auf zugrundeliegende Prozesse ist nicht mehr direkt möglich. Im Rahmen der Arbeit wurden entsprechend verschiedene Methoden evaluiert, wie ein Rückschluss auf relevante Prozesse möglich ist, dabei kamen insbesondere Krebsassoziierte Pathways sowie RAS Signalling als wichtige Prozesse mit hohem prädiktiven Wert zur Vorhersage einer Fettleber heraus. Die Masterarbeit wurde erfolgreich abgeschlossen (Jan Oldenburg, „Optimierung eines künstlichen neuronalen Netzes zur Vorhersage einer nichtalkoholischen Fettleber mit Hilfe von KEGG-Pathways“, Universität Greifswald, Studiengang Biomathematik, 2019). Herr Oldenburg hat sich im Anschluss an die Masterarbeit entschieden, weiter an der Universitätsmedizin Greifswald am Institut für Bioinformatik tätig zu sein, und setzt die Arbeit außerhalb des LiSyM-Projektes im Rahmen einer Folge-Dissertation fort.

**Die Arbeiten zum Meilenstein 6 wurden entsprechend den Planungen abgeschlossen.**

## Meilenstein 7: Analyse der SHIP/GANI\_MED Daten (M36-60)

**Arbeitsbeschreibung gemäß Antrag:** Die Methoden aus Meilenstein 3 und 6 werden auf SHIP/GANI\_MED Daten angewendet, um Zusammenhänge der Krankheitsprogression zu untersuchen. Darüber hinaus werden in diesem Arbeitsschritt Schlüsselprozesse des Krankheitsfortschritts (Meilenstein 4) sowie Biomarker-Scores (Meilenstein 5) aus Lebergewebe mit systemischen Daten und Signaturen auf SHIP/GANI\_MED Probanden aus Blut verglichen.

**Durchgeführte Arbeiten:** Daten der SHIP Studie wurden vorprozessiert und Analysen mit Bayes'schen Netzen wurden erfolgreich durchgeführt, unter Integration von phänotypischen Daten und Laborparametern. Es zeigte sich dabei, dass Signaturen in Lebergewebe (Meilensteine 1-3) nicht direkt mit Signaturen im Blut (SHIP) vergleichbar sind, sondern hier andere Prozesse im Vordergrund stehen. Interessanterweise zeigten die Daten der SHIP-Studie, dass durch Integration von molekularen Parametern (Leberfunktionstests, Serum Glukose, Cholesteroll) mit klinischen und phänotypischen Variablen (Körperfett, Geschlecht, Impedanzanalyse, Alter) gute diagnostische Vorhersagen getroffen werden können, die die Ergebnisse von etablierten Biomarker-Scores für die Fettleber übertreffen, allerdings nicht an die Qualität einer Leberbiopsie herankommen (siehe hierzu auch Meilenstein 5 sowie Abbildung).

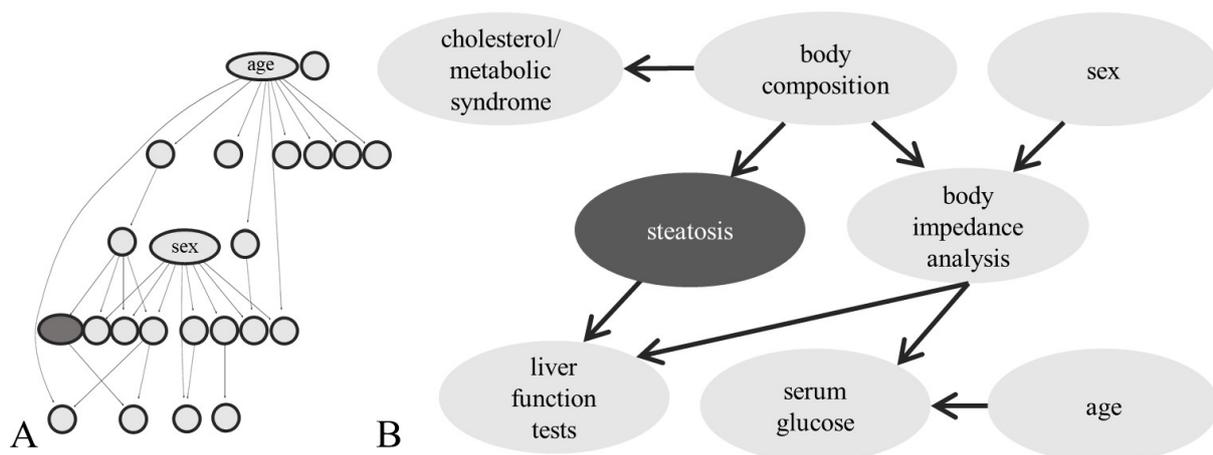


Abbildung: Hierarchisches Bayes'sches Netzwerk zur Vorhersage der Fettleber. Links gezeigt das vollständige Biomarker-Netzwerk (A), rechts ein Auszug der gruppierten relevanten Variablen (B). Jeder Knoten fasst dabei mehrere Messwerte / Variablen der SHIP Studie zusammen, z.B. umfasst „body composition“ Variablen wie Gewicht, BMI, Waist-to-Height Ratio und weitere Messungen, die gemeinsam in den finalen Score eingehen.

Ergebnisse wurden in PLOS Computational Biology publiziert (Becker AK, Dörr M, Felix SB, Frost F, Grabe HJ, Lerch MM, Nauck M, Völker U, Völzke H, **Kaderali L**. *From heterogeneous healthcare data to disease-specific biomarker networks: A hierarchical Bayesian network approach*. **PLoS Comput Biol**. 2021 Feb 12;17(2):e1008735).

Zur Integration von weiteren OMICs-Daten der SHIP-Studie in die Analysen wurden weitere Anpassungen der Methode vorgenommen, insbesondere was die Datenprozessierung, Regularisierung und Einbindung von Multi-OMICs-Daten in die Netzwerkanalysen angeht. Insbesondere wurde durch Kombination von Entscheidungsbäumen als Vorprozessierungsschritt vor der Bayes'schen Netzwerk-Analyse eine Methode vorgeschlagen, mit der eine Merkmalsauswahl für relevante Prädiktoren für die Netzwerkanalyse vorgenommen werden kann – die

Entscheidungsbäume dienen der Prädiktion und der Auswahl relevanter Messwerte, während das folgende Bayes'sche Netz diese interpretierbar macht und den Zusammenhang zwischen verschiedenen Messwerten und dem Auftreten der Erkrankung visualisiert.

Eine methodische Publikation zu diesem Ansatz wurde zur Veröffentlichung eingereicht und ist aktuell im Begutachtungsprozess (Becker et al., submitted / under review). **Die Arbeiten zum Meilenstein 7 wurden entsprechend den Planungen abgeschlossen.**

## **2. Die wichtigsten Positionen des zahlenmäßigen Nachweises**

### **Personal:**

Im Rahmen des Projektes wurde eine Doktorandenstelle über die gesamte Projektlaufzeit bewilligt (1 Doktorand, E13, 50% VZÄ). Die Stelle konnte zum Projektstart mit Frau Kleimeier besetzt werden, die durch Elternzeit während der Projektlaufzeit pausierte, währenddessen wurde Frau Becker auf dem Projekt beschäftigt:

Dana Kleimeier (Wissenschaftliche Mitarbeiterin / Doktorandin): 1.1.2017 bis 30.11.2017

Ann-Kristin Becker (Wissenschaftliche Mitarbeiterin / Doktorandin): 1.12.2017 bis 31.12.2020

Dana Kleimeier (Wissenschaftliche Mitarbeiterin / Doktorandin): 1.1.2021 bis 30.06.2021

Frau Kleimeier setzt ihr Dissertationsvorhaben auf Haushaltsmitteln des Instituts für Bioinformatik fort.

### **Verbrauchsmaterial:**

Im Rahmen des Projektes erfolgten keine Abweichungen von den ursprünglichen Planungen.

### **Vergabe von Aufträgen:**

Im Rahmen des Projektes erfolgte keine Vergabe von Aufträgen.

### **Investitionen:**

Im Rahmen des Projektes erfolgten keine Investitionen.

Insgesamt wurden alle bewilligten Mittel für die Durchführung des Projektes und die Erreichung der Projektziele verwendet.

## **3. Notwendigkeit und Angemessenheit der geleisteten Arbeit**

Die durchgeführten Arbeiten haben unmittelbar zu einem besseren Verständnis der molekularen Prozesse bei der Progression der Fettleber beigetragen, wobei systembiologische Methoden und mathematische Modelle eine maßgebliche Rolle zur Untersuchung der Vorgänge gespielt haben. Durch die Analysen insbesondere auf den SHIP-Daten konnte ein neuer Risikoscore für die Fettleber entwickelt werden, der nicht-invasiv (ohne Leberbiopsie) arbeitet und im Vergleich zu etablierten Scores zumindest auf den SHIP-Daten eine bessere Vorhersagegüte erreicht. Ergebnisse wurden

publiziert und erfordern weitere Validierung durch die wissenschaftliche Community auf anderen Datensätzen im Rahmen von klinischen Studien.

Damit hat das Teilprojekt wesentlich zu den Zielen des Gesamtverbundes beigetragen, und insbesondere auch wichtige Beiträge zur Erreichung der förderpolitischen Ziele der gesamten Ausschreibung geliefert.

Die angeforderten Personal- und Verbrauchsmittel wurden angemessen eingesetzt und haben einen wesentlichen wissenschaftlichen Erkenntnisgewinn sichergestellt (wird durch bereits erfolgte und weitere anstehende Publikation dokumentiert). Dabei sind in diesem Teilprojekt vor allem Personalmittel eingesetzt worden, die unmittelbar für die Entwicklung und Untersuchung der mathematischen Modelle und Datenanalysen benötigt wurden.

#### **4. Voraussichtlicher Nutzen, insbesondere Verwertbarkeit der Ergebnisse im Sinne des fortgeschriebenen Verwertungsplans**

Im Rahmen der Antragstellung wurden keine wirtschaftlichen Erfolgsaussichten für das vorliegende Projekt in Aussicht gestellt. In Bezug auf den fortgeschriebenen Verwertungsplan haben sich keine unmittelbaren wirtschaftlichen Erfolgsaussichten nach Projektende ergeben. Während der Projektlaufzeit gab es keine Erfindungen/Schutzrechtsanmeldungen.

Eine unmittelbare wissenschaftliche Anschlussfähigkeit ist gegeben. Methodische Entwicklungen zur Integration von OMICs-Daten mit klinischen Daten (hierarchische Bayes'sche Netze, Kombination von Entscheidungsbäumen zur Vorhersage mit Bayes'schen Netzen zur Modellinterpretation) sind bereits in weiteren Anschlussprojekten eingesetzt worden, und sind aktuell Grundlage für einen eingereichten DFG-Antrag, in dem diese Methoden im Rahmen einer Forschergruppe zur Analyse von neurologischen Erkrankungen eingesetzt werden sollen. Auch die im Rahmen von LiSyM\_EMI entwickelten Deep-Learning-Verfahren werden im Rahmen einer Dissertation am Institut für Bioinformatik weiter entwickelt.

#### **5. Während der Durchführung des Vorhabens bekannt gewordener Fortschritt auf dem Gebiet des Vorhabens bei anderen Stellen**

Es sind keine Ergebnisse bekannt geworden, die die Projektziele oder den Projektablauf verändert haben.

#### **6. Erfolgte oder geplante Veröffentlichungen der Ergebnisse**

*Erfolgte Veröffentlichungen:*

1. C. Bruckmann, *Eine Analyse von Genexpressionsdaten mittels maschinellen Lernens zur Untersuchung der Krankheit Fettleber unter Ausschluss von alkoholbedingten Ursachen*. Bachelor-Arbeit, Universität Greifswald, Studiengang Biomathematik, 2017.
2. Becker AK, Dörr M, Felix SB, Frost F, Grabe HJ, Lerch MM, Nauck M, Völker U, Völzke H, **Kaderali L**. *From heterogeneous healthcare data to disease-specific biomarker networks: A hierarchical Bayesian network approach*. **PLoS Comput Biol**. 2021 Feb 12;17(2):e1008735

3. Oldenburg, J. *Optimierung eines künstlichen neuronalen Netzes zur Vorhersage einer nichtalkoholischen Fettleber mit Hilfe von KEGG-Pathways*, Master-Arbeit, Universität Greifswald, Studiengang Biomathematik, 2019
4. Gholizadeh, M.; Szlag-Pieniek, S.; Post, M.; Kurzawski, M.; Prieto, J.; Argemi, J.; Drozdik, M.; Kaderali, L. Identifying Differentially Expressed MicroRNAs, Target Genes, and Key Pathways Deregulated in Patients with Liver Diseases. *Int. J. Mol. Sci.* 2020, 21, 7368
5. Becker AK, Ittermann T, Dörr M, Felix S, Nauck M, Teumer A, Völker U, Völzke H, Kaderali L, Nath N (2021). Discovering association patterns of individual serum Thyrotropin concentrations using machine learning: An example from the Study of Health in Pomerania (SHIP). Submitted.
6. A.-K. Becker, *Discovering Latent Structures in High-Dimensional Healthcare Data: Towards Improved Interpretability*. Dissertation, Dr. rer. Nat., Universität Greifswald

## Berichtsblatt

1. ISBN oder ISSN	2. Berichtsart (Schlussbericht oder Veröffentlichung) Schlussbericht
3. Titel  LiSyM Verbundprojekt: Pillar I – Metabolischer Frühschaden (EMI), Teilprojekt 2	
4. Autor(en) [Name(n), Vorname(n)]  Kaderali, Lars	5. Abschlussdatum des Vorhabens 31.6.2021
	6. Veröffentlichungsdatum
	7. Form der Publikation
8. Durchführende Institution(en) (Name, Adresse)  Universitätsmedizin Greifswald, Institut für Bioinformatik, Felix-Hausdorff-Str. 8, 17475 Greifswald	9. Ber. Nr. Durchführende Institution
	10. Förderkennzeichen 031L0032
	11. Seitenzahl
12. Fördernde Institution (Name, Adresse)  Bundesministerium für Bildung und Forschung (BMBF) 53170 Bonn	13. Literaturangaben
	14. Tabellen
	15. Abbildungen
16. Zusätzliche Angaben	
17. Vorgelegt bei (Titel, Ort, Datum)	
18. Kurzfassung Die Pathogenese der Manifestationen metabolischer Erkrankungen in Form einer Fettleberhepatitis an der Leber war (und ist) bisher nur unvollständig verstanden. Bisherige Studien nutzten selten systembiologische oder genomweite Ansätze und waren oft auf einzelne Gene und Pathomechanismen fokussiert. Die Anwendung moderner systembiologischer Analysen lässt auf neue Erkenntnisse zur Pathogenese und die Identifizierung neuer Biomarker und molekularer Zielmoleküle für die Therapie hoffen. Diese Ansätze sollten im LiSyM Verbundprojekt kombiniert werden und zum Verständnis der molekularen Prozesse bei Progression von benigner Steatose zur Nekroinflammation als krankheitdefinierendes Ereignis für die nichtalkoholische Steatohepatitis (englisch: NASH) angewendet werden, um wesentliche molekulare Prozesse aufzuklären. Das vorliegende Teilprojekt befasst sich dabei mit der integrativen Bioinformatik-Analyse der in LiSyM_EMI erhobenen Daten. Es wurden Bioinformatik-Analyseverfahren zur Inferenz von Risikofaktor-Netzwerken entwickelt und zur Erstellung von prädiktiven, diagnostischen Modellen verwendet. Dabei konnte ein neuer, nicht-invasiver Score zur Prognose der Fettleber entwickelt und analysiert werden, und es wurden methodische Entwicklungen im Gebiet der hierarchischen Bayesschen Netze publiziert.	
19. Schlagwörter Fettleber, NASH, Biomarker, Bayes'sche Netze	
20. Verlag	21. Preis