

# Berichtsblatt

1. ISBN oder ISSN	2. Berichtsart (Schlussbericht oder Veröffentlichung) Schlussbericht
3. Titel SyMBoD: Systemmedizinische Analysen zur personalisierten Behandlung von Knochendefekten in Patienten mit Diabetes - Tiermodell, Biobanken und Modellierung	
4. Autor(en) [Name(n), Vorname(n)] Dr. Tanja Laske Klaudia Magdalena Adamowicz	5. Abschlussdatum des Vorhabens 30.11.2024
	6. Veröffentlichungsdatum 25.Juli 2025
	7. Form der Publikation
8. Durchführende Institution(en) (Name, Adresse) Universität Hamburg Mittelweg 177, 20148 Hamburg	9. Ber. Nr. Durchführende Institution
	10. Förderkennzeichen 01ZX2210D
	11. Seitenzahl
12. Fördernde Institution (Name, Adresse) Bundesministerium für Bildung und Forschung (BMBF) 53170 Bonn	13. Literaturangaben 0
	14. Tabellen 0
	15. Abbildungen 0
16. Zusätzliche Angaben	
17. Vorgelegt bei (Titel, Ort, Datum) BMBF	
18. Kurzfassung SyMBoD hat das Ziel, die körpereigene Regeneration bei bisher unzureichend oder nicht gelösten medizinischen Herausforderungen zu fördern, indem es innovative digitale Technologien aus der Systemmedizin einsetzt. Als Modell dienen Patienten mit nicht verheilenden Knochenbrüchen sowie Typ-2-Diabetes mellitus (T2DM). Das interdisziplinäre SyMBoD-Konsortium verfolgt das Ziel, die Mechanismen hinter der fehlenden Knochenheilung bei T2DM zu entschlüsseln. Durch die Analyse von Omics- und Bildgebungsdaten bei Patienten mit eingeschränkter Knochenregeneration wird SyMBoD personalisierte, intelligente 3D-gedruckte Implantate entwickeln, die eine Heilung schwerwiegender Knochendefekte ermöglichen. Dazu werden klinische sowie molekulare Daten sowohl von Patienten als auch aus präklinischen Modellen mit computergestützten Methoden und mechano-biologischen Optimierungsansätzen kombiniert, um patientenspezifische, intelligente Implantate für die Knochenheilung zu entwickeln. Das Institut für computergestützte Systembiologie an der Universität Hamburg bringt seine Expertise in der Identifikation von Krankheitsmechanismen sowie im Bereich der Medikamenten-Umwidmung unter Einsatz künstlicher Intelligenz, Netzwerkbiologie und systemmedizinischer Methoden ein. Es ist insbesondere in den Teilprojekten (TP) 3 (Identifikation von Biomarkern) und TP5 (Entwicklung der systemmedizinischen SyMBoD-Plattform) beteiligt.	
19. Schlagwörter Systems medicine, bone regeneration, omics data, personalized therapy, network analysis	
20. Verlag n.a.	21. Preis n.a.

# Titelblatt zum Schlussbericht

(Gemäß Nrn. 5.4 NABF)

Thema: SyMBoD: Systemmedizinische Analysen zur personalisierten Behandlung von Knochendefekten in Patienten mit Diabetes - Tiermodell, Biobanken und Modellierung

*Themen des Teilvorhabens:*

TP3 - SyMBoD: Identifizierung von Biomarkern für die frühzeitige Diagnose und Prognose des Heilungsverlaufs von Knochendefekten bei Type 2 Diabetes Mellitus (T2DM)

TP5 - SyMBoD: Integrative Systemmedizin-Plattform

Förderkennzeichen	Autor/innen <i>(hier bitte alle Autorinnen und Autoren nennen, um Plagiatsvorwürfe zu vermeiden)</i>	Zuwendungsempfänger	Laufzeit <i>(Beginn und Ende)</i>	Hinweis auf Vertraulichkeit
01ZX2210D	Dr. Tanja Laske, Klaudia Magdalena Adamowicz	Universität Hamburg	01.12.2022 - 30.11.2024	x

Kontaktperson:

Name: Dr. Nina Wenke

Straße: Albert-Einstein-Ring 8-10

PLZ Ort: D-22761 Hamburg

E-Mail: nina.wenke@uni-hamburg.de

Telefon: +49 40 42838-7643

## Teil I (Kurzbericht)

1. Erläuterung zur ursprüngliche Aufgabenstellung sowie den wissenschaftlichen und technischen Stand, an den angeknüpft wurde

SyMBoD hat das Ziel, die körpereigene Regeneration bei bisher unzureichend oder nicht gelösten medizinischen Herausforderungen zu fördern, indem es innovative digitale Technologien aus der Systemmedizin einsetzt. Als Modell dienen Patienten mit nicht verheilenden Knochenbrüchen sowie Typ-2-Diabetes mellitus (T2DM). Das interdisziplinäre SyMBoD-Konsortium verfolgt das Ziel, die Mechanismen hinter der fehlenden Knochenheilung bei T2DM zu entschlüsseln. Durch die Analyse von Omics- und Bildgebungsdaten bei Patienten mit eingeschränkter Knochenregeneration wird SyMBoD personalisierte, intelligente 3D-gedruckte Implantate entwickeln, die eine Heilung schwerwiegender Knochendefekte ermöglichen. Dazu werden klinische sowie molekulare Daten sowohl von Patienten als auch aus präklinischen Modellen mit computergestützten Methoden und mechano-biologischen Optimierungsansätzen kombiniert, um patientenspezifische, intelligente Implantate für die Knochenheilung zu entwickeln. Das Institut für computergestützte Systembiologie an der Universität Hamburg bringt seine Expertise in der Identifikation von Krankheitsmechanismen sowie im Bereich der Medikamenten-Umwidmung unter Einsatz künstlicher Intelligenz, Netzwerkbiologie und systemmedizinischer Methoden ein. Es ist insbesondere in den Teilprojekten (TP) 3 (Identifikation von Biomarkern) und TP5 (Entwicklung der systemmedizinischen SyMBoD-Plattform) beteiligt.

2. Ablauf des Vorhabens

Das Hauptziel in TP3 ist die Identifizierung und Validierung von Biomarkern, welche die Diagnose, Prognose und Therapie von Knochendefekten verbessern, die bei Patienten mit T2DM unzureichend verheilen. Zusätzlich sollen bioinformatisch gewonnene Parameter mathematische Modelle zur Simulation der Knochendefektheilung informieren und Vorschläge für die Funktionalisierung von 3D-gedruckten Implantaten ermöglichen, z. B. mittels Medikamenten-Umwidmung. Um diese Forschungsziele zu erreichen, werden Plasma- und Gewebeprobe aus einem T2DM-Rattenmodell mit segmentalen Femurdefekt untersucht. Um die Beschreibung und das Verständnis des Heilungsprozesses von Knochendefekten zu verbessern, werden longitudinale multi-OMICS-Analysen von Patienten und Tiermodellen durchgeführt und dafür maßgeschneiderte Computerprogramme für die Vorhersage von Teilnetzmarkern entwickelt. Dabei verknüpft das TP3 mehrere TPs zur Probengewinnung von Patienten und Tiermodellen (TP1 und TP2), die Datenerfassung und -analyse (TP3), Modellentwicklung und -parametrisierung durch Datenintegration (TP4) und die Implementierung der system-medizinischen SyMBoD-Plattform (TP5).

Während der Projektlaufzeit etablierte sich eine agile Forschungspraxis im SyMBoD-Konsortium, um auf die Herausforderungen verbunden mit der COVID-19 Pandemie sowie personellen Veränderungen (Personalabgang und Ausbildung neuer Mitarbeitender) und logistischen Herausforderungen (Wechsel zum ZDF Rattenmodell für T2DM) angemessen reagieren zu können. Dafür ermöglichten regelmäßige online-Meetings in Themengruppen, z. B. auch mit dem Sys\_CARE-Konsortium, und/oder persönliche Treffen mit zuständigen Partnern nach Bedarf, die Entwicklung von Lösungsstrategien, um Zwischenziele trotz zeitlicher Verzögerungen in der Probengewinnung zu erreichen und die Durchführung des

Arbeitsplans sicherzustellen.

3. Wesentlichen Ergebnisse sowie ggf. die Zusammenarbeit mit anderen Forschungseinrichtungen

In den Teilprojekten 3 und 5 (TP3, TP5) wurden computergestützte Methoden entwickelt und diese auf einer systemmedizinischen Plattform implementiert, um u. a. Protein- und Metabolitprofile aus Plasma- und Gewebeproben zu analysieren.

In TP3 wurden zunächst eine Vielzahl öffentlich verfügbarer multi-Spezies Proteomics-Daten für eine Meta-Analyse zusammengetragen, um solche Proteine zu identifizieren, die bei der Knochenentwicklung und/oder Knochenregeneration reguliert sind und als Biomarker oder zur Funktionalisierung von Scaffolds dienen könnten. Mittels Biostatistik sowie durch systembiologische Analysen der Protein-Protein-Netzwerke (Modul-Mining) identifizierten wir potenzielle Zielmoleküle für eine mögliche Medikamenten-Umwidmung bereits verfügbarer Pharmazeutika, die für die zukünftige Experimentplanung sowie Validierungsversuche herangezogen werden können (Schmidt et al. 2024, Journal of Tissue Engineering).

Im Rahmen der Meta-Analyse sowie der Analyse der OMICs-Daten entwickelten wir Ansätze für eine verbesserte Datenvorbereitung, um eine qualitativ hochwertige und statistisch sichere Auswertung zu gewährleisten. Dazu gehören u. a. die Entwicklung des Computerprogramms ProHarMed (Adamowicz et al. 2023, npj Syst Biol Appl, <https://apps.cosy.bio/proharmed/>) zur verbesserten Integration und Schnittmengenanalyse von Datensätzen verschiedener Spezies, sowie die Entwicklung der Software PRONE zur Evaluierung des Einflusses von Normalisierungsmethoden auf biostatistischen Auswertungen wie die differentielle Expressionsanalyse (Arend et al., Brief Bioinform, akzeptiertes Manuskript, <https://exbio.wzw.tum.de/prone/>). Zusammen mit Partnern aus dem Sys\_CARE-Projekt entwickeln wir eine Methode zum Clustering für die Zeitreihenanalyse von Plasmapdaten. Die damit identifizierten Biomarker zeigen Übereinstimmungen mit den Ergebnissen aus der Meta-Analyse und sind deshalb vielversprechende Kandidaten für Validierungsexperimente (Schmidt et al., in Vorbereitung). Darüber hinaus entwickelten wir das Computerprogramm DRaCOoN (Delgado-Chaves et al., BMC Bioinformatics, in Revision, <https://pypi.org/project/draconopy/>) für die *de novo* Rekonstruktion von differentiell regulierten Netzwerken und analysierten damit einen bereits publizierten Transkriptom-Datensatz zu Knochenheilung im Tiermodell. Die Analyse der Knochen-OMICs Daten zeigte distinkte molekulare Veränderungen in der kontralateralen Kontrolle im Vergleich zum Femurdefekt. Bei der weiteren Datenanalyse wurde deshalb die Nutzung der kontralateralen Kontrolle für ein molekulares Referenzprofil sowie bei der Planung von Validierungsexperimenten berücksichtigt, sodass neue zusätzliche Kontrollen eingeführt wurden (Sham-operated) und damit ein neues experimentelles Design motiviert wurde. Darüber hinaus führte die bioinformatische Auswertung der Knochen-Metabolomics Daten zur Identifizierung von Molekülen, deren heilungsunterstützende Wirkung in Validierungsexperimenten bestätigt werden konnte (Dias et al., Acta Biomaterialia, in Revision). In Zusammenarbeit mit Genevention (TP5) integrieren wir die in TP3 entwickelten Programme und Analysemethoden in der SyMBoD-Plattform. Insbesondere richteten wir die Möglichkeit ein, Analysen und Workflows agil anzupassen, indem wir zusätzliche öffentlich einsehbare Software-Verzeichnisse (git) und Docker-Container für robuste Kontrolle der Programmierumgebung nutzen.

## Teil II

Durchgeführten Arbeiten (ausführliche Darstellung):

Im Folgenden werden die wichtigsten Ergebnisse zu den zugehörigen Zielstellungen der von der Universität Hamburg zu bearbeitenden Teilprojekte (TPs) dargestellt.

### Teilprojekt 3 - Biomarker Identifizierung

AP1.4 - Datenintegration, Biostatistik und Bioinformatik

**Zielstellung:** Identifizierung von diagnostischen und prognostischen Biomarkern. Analyse molekularer Profile sowie Entwicklung von verbesserten Bioinformatik-Workflows zur Integration von Multi-OMICs-Daten zum Vergleich von Literaturdaten und Daten aus (prä)-klinischen Studien an Tieren und Menschen.

#### Biomarker aus der Literatur:

Die Literatursuche aus Phase 1 wurde ausgeweitet, sodass zu den 24 Proteomik-Datensätzen, sieben weitere Studien integriert wurden um gemäß der PRISMA-Richtlinien relevante Arbeiten zu den Themen "liquid biopsy" und Implantat-gestützter Heilung von Knochendefekten zu ergänzen (Schmidt et al. 2024, Journal of Tissue Engineering). Diese Vielzahl von multi-Spezies Datensätzen, die vom SyMBoD Partner zusammengetragen und re-analysiert wurden (PI Kalkhof, Fraunhofer IZI), wurde mittels unseres neu entwickeltes Programm ProHarMed (Adamowicz et al. 2023, npj Syst Biol Appl, <https://apps.cosy.bio/proharmed/>) für eine integrative systembiologische Analyse vorbereitet. ProHarMed erleichtert die Meta- und Schnittmengenanalyse von Datensätzen verschiedener Spezies (Mensch, Ratte, Maus, Kaninchen) mit unterschiedlicher Detailtiefe, d. h. mit unterschiedlicher Verfügbarkeit von Gen-/Protein-IDs und liefert eine verbesserte Abdeckung relevanter biologischer Netzwerke, da es die Identifikation einer höheren Anzahl an Biomarkern und damit Startknoten für die Netzwerkanalyse ermöglicht. Die Relevanz der so gewonnenen Netzwerke konnten wir sowohl biostatistisch, mittels der Methode DIGEST (Adamowicz et al. 2022, Brief Bioinform, <https://digest-validation.net/>), als auch qualitativ bestätigen. Die qualitative Validierung beruht dabei auf dem Abgleich der ProHarMed-Module mit bereits bekannten Krankheitsmechanismen und Medikamenten, die wir mittels Modul-Mining und Netzwerk-basierter Medikamenten-Umwidmung mittels drugst.One (<https://drugst.one/>) identifizieren konnten.

Bei der Auswahl der Startknoten berücksichtigen wir ebenfalls die verschiedenen Experimente (*in vivo*, *ex vivo*, *in vitro*) sowie der analysierten Gewebe (extrazelluläre Vesikeln, Plasma, Knochen, Zellkultur) durch eine Gewebe-spezifische Gewichtung mit der Anzahl der verfügbaren Studien und der darin gefundenen Biomarker. Insbesondere galt es Gewebe-übergreifend Biomarker (Kernbiomarker) zu identifizieren, die z. B. nach der Messung im Plasma Rückschlüsse auf andere Gewebe ermöglichen. Bei der Identifizierung von Kernbiomarkern berücksichtigen wir neben der DIGEST-Validierung weitere Kriterien: Zugänglichkeit im Plasma, Netzwerk-Zentralität, Funktion in biologischen Prozessen des Netzwerks, Einfluss auf die Knochenheilung, Anzahl bekannter Medikamenten, welche diesen Zielproteinen zugeordnet werden. Dieser umfassende Workflow ermöglicht auch zukünftig, neu ermittelte Biomarker aus Literaturdaten und Daten aus (prä)-klinischen Studien einzubeziehen und zu bewerten, z. B. für potenzielle Behandlungsmöglichkeiten.

Zusammenfassend ermöglichte unsere Analyse die Identifikation von 51 zentralen Proteinen in der Knochenregeneration, wovon 30 bereits über bekannte Medikamente adressierbar sind. Insgesamt wurden 133 zugelassene Wirkstoffe identifiziert, die gezielt in Knochenheilungsprozesse eingreifen könnten. Besonders hervorzuheben sind die Substanzen Trifluoperazin, Quercetin, Phenethyl-Isothiocyanat und Arteminol, die mehrere dieser Zielproteine modulieren und Potenzial für eine gezielte therapeutische Steuerung der Knochenheilung besitzen. Neben klassischen ECM-Komponenten zeigen sich auch 12 weniger untersuchte Proteine wie CCN1, HSPG2, TGFBI oder ANXA1 als attraktive Kandidaten für die bioaktive Funktionalisierung von Knochenimplantaten.

### **Verbesserte Analyse von Proteomik-Daten durch optimale Datenvorbereitung, Biostatistik und Netzwerkanalyse:**

Um eine qualitativ hochwertige und statistisch abgesicherte Datenanalyse zu gewährleisten, evaluierten wir systematisch den Einfluss von Normalisierungsmethoden auf nachgeschaltete Analysen, wie z. B. der differentiellen Expressionsanalyse. Dafür bieten wir das neue Programm PRONE (Arend et al., Brief Bioinform, in Revision, <https://exbio.wzw.tum.de/prone/>) an, welches es dem Nutzer erlaubt 17 Normalisierungsmethoden vergleichend zu evaluieren. Wir haben einen Methoden- sowie Datensatz-abhängigen Einfluss auf die Analysen festgestellt und empfehlen deshalb die Nutzung von PRONE für jeden neuen Datensatz. Zur Qualitätskontrolle nutzen wir standardmäßig, sowohl in PRONE als auch in anderen Analyse-Workflows für Proteomik- als auch Metabolomik-Daten, das Programm POMA (Castellano-Escuder et al. 2021, Plos Comp Biol) zur Identifizierung von Ausreißern.

Auf Grundlage verschiedener Analyse-Strategien, um Knochen-OMICs Daten auszuwerten, wurde die Nutzung der kontralateralen Kontrolle, d. h. des Knochens ohne Defekt, für ein molekulares Referenzprofil sowie bei der Planung von Validierungsexperimenten berücksichtigt, sodass neue zusätzliche Kontrollen eingeführt wurden (Sham-operated) und damit ein neues experimentelles Design motiviert wurde.

### **Proteomik:**

Wir führten verschiedene bioinformatische Analysen der SyMBoD-Daten von diabetischen und nicht-diabetischen Ratten mit einem Femurdefekt durch, die u. a. im letzten Jahr des ersten Förderzeitraums gewonnen wurden. Es galt, Veränderungen im Plasma- und Knochenproteom, sowie zusätzlich Daten der kontralateralen Kontrollen zu analysieren.

Im Anschluss an die Datenvorbereitung (siehe oben: Normalisierung, Filterung, Ausreißeranalyse) erfolgte die differentielle Expressions-Analyse mittels Limma sowie Netzwerkanalysen mit Methoden des Modul-Mining und der Medikamenten-Umwidmung, z. B. mittels des neuen KPM R-Paketes (KeyPathwayMineR, Mechteridis et al. Front Genet, 2022) sowie ProHarMed (Adamowicz et al. 2023, npj Syst Biol Appl, <https://apps.cosy.bio/proharmed/>). Wir verfolgten den kontinuierlichen und statischen Ansatz, d. h. entlang der Zeit innerhalb jeder Gruppe (diabetisch und nicht-diabetisch) und zwischen den Gruppen zu jedem gemessenen Zeitpunkt. Der mit den Sys\_CARE-Partnern (Förderkennzeichen 01ZX1908/01ZX2208) neu entwickelte Ansatz für das Clustering zeitlich aufgelöster Plasma-Datensätze wurde weiter optimiert, um diejenigen Plasmaproteine zu finden, die sowohl in diabetische als auch nicht-diabetische Ratten reguliert sind, aber unterschiedliche Cluster-Prototypen aufweisen. Dazu werteten wir systematisch den Einfluss des linkage modes auf die erhaltenen Cluster-Prototypen aus. Die mittels Clustering

identifizierten Biomarker zeigen Übereinstimmungen mit den Ergebnissen aus der Meta-Analyse (siehe oben) und sind deshalb vielversprechende Kandidaten für Validierungsexperimente (Schmidt et al., in Vorbereitung). Von insgesamt 432 über das Clustering identifizierten Biomarkern überschneiden sich 53 Kandidaten mit der Meta-Analyse. Unter diesen befinden sich 13 sogenannte Kernbiomarker, die in mehreren biologischen Kontexten der Meta-Analyse vorkommen (z. B. Bone; EVs oder Bone; Liquid-Biopsy). Zusätzlich wurden 189 differentiell exprimierte Proteine identifiziert, entsprechend 114 einzigartigen Genen. Davon finden sich 22 Gene auch in der Metaanalyse wieder, darunter vier Kernbiomarker, die mehreren funktionellen Kategorien zugeordnet sind.

Die molekulare Subtypisierung, d. h. Stratifizierung, der Ratten mittels Grand Forest (Netzwerk-basiertes Clustering, Larsen et al. 2020, Systems Medicine) und MoSBI (Ensemble Biclustering, Rose et al. 2022, PNAS, <https://exbio.wzw.tum.de/mosbi/>) wird durch die geringe Stichprobengröße (n = 4-8 Tiere pro Zeitpunkt und Bedingung) erschwert und war nicht unmittelbar evident. Die neu entwickelte Methode UnPaSt (nicht-überwachtes Clustering, <https://apps.cosy.bio/unpast/>, Pre-Print: Hartung et al., <https://doi.org/10.48550/arXiv.2408.00200>) wurde für die SyMBoD Daten geprüft. Vorläufige Ergebnisse möchten wir mittels Analyse der neuen Ratten-Kohorten (ZDF Rattenmodell, siehe unten) nach der zweiten Förderperiode weiterverfolgen. Die Verwendung von Grand Forest soll ebenfalls wiederholend geprüft werden, um molekulare Profile mit dem Erfolg der Knochenheilung zu assoziieren (mechanistische Endophänotypisierung).

Bei der differentiellen Expressions-Analyse des Knochenproteoms war unerwartet, dass in den Proben der nicht-diabetischen kontralateralen Kontrolle im Vergleich zum Femurdefekt doppelt so viele differentiell exprimierte Proteine ermittelt wurden. Wir haben Grund zur Annahme, dass in der kontralateralen Kontrolle distinkte molekulare Veränderungen als Antwort auf den Femurdefekt induziert werden. Dies motivierte neue Analysestrategien sowie die Einführung neuer zusätzlicher Kontrollen (Sham-operated) in den Validierungsexperimenten (ZDF Rattenmodell).

Im Rahmen der Evaluierung eines neuen Extraktionsprotokolls (Wiltzsch et al., Journal of Proteome Research, in Revision) zur kombinierten Metabolomics- und Proteomics-Analyse unterstützten wir die bioinformatische Auswertung der DIA-NN-basierten Proteomdaten (Fraunhofer IZI). Drei Methoden wurden verglichen: eine Standard-Ein-Schritt-Extraktion (mit und ohne kollagenspezifischen Puffer), ein aufwendig etabliertes Vier-Schritt-Protokoll sowie das neu entwickelte Zwei-Schritt-Protokoll. Letzteres quantifizierte im Median 4.518 Proteine mit hoher Reproduzierbarkeit und einem dynamischen Bereich über vier Größenordnungen – nahezu vergleichbar zur Vier-Schritt-Methode mit 4.934 Proteinen, jedoch bei halbiertem Extraktions- und Messaufwand. Der Kollagenanteil wurde effektiv reduziert, ohne kollagenasespezifische Enzyme. Zur Prüfung der Multi-Omics-Tauglichkeit wurde das Protokoll mit einer vorgeschalteten Metabolitextraktion (2-Step+M) kombiniert: Die Abdeckung sank nur minimal auf 4.393 Proteine. Die funktionelle Analyse zeigte eine hohe Übereinstimmung in der Abdeckung biologischer Prozesse, insbesondere von Signalweg-bezogenen und ECM-assoziierten Proteinen, was die Eignung für integrative OMICs-Studien bestätigt.

## **Metabolomik:**

Wir führten verschiedene bioinformatische Analysen der SyMBoD-Daten von diabetischen und nicht-diabetischen Ratten mit einem Femurdefekt durch, die u. a. im letzten Jahr des ersten Förderzeitraums gewonnen wurden. Es galt, Veränderungen im Plasma- und Knochen-Metabolom zu analysieren. Dafür standen uns Datensätze sowohl von GC-MS- als auch Biocrates-Messungen vom SyMBoD-Partner aus TP1 zur Verfügung.

Bei den Analysen der GC-MS Knochen-Metabolom-Daten konzentrierten wir uns auf die Analyse der Daten des Knochendefektes. Dazu gehören eine explorative Datenanalyse, Metabolite Set Enrichment Analysis sowie eine Biomarker-Suche, mit Hilfe des Programms MetaboAnalyst (<https://www.metaboanalyst.ca/>). Die Analyse zur Identifizierung differentiell abundanter Metabolite lieferte dabei vergleichsweise wenig Kandidaten, was vermutlich mit der geringe Datensatz- und Stichprobengröße zusammenhängt (70 Metabolite ohne fehlende Werte, n = 4 Tiere pro Zeitpunkt und Bedingung). Mittels AUC-ROC (Area under the receiver operator curve) ermittelten wir eine Rangliste von Metaboliten, die als therapeutischer Biomarker dienen könnten. Die Analyse gibt dabei für jeden Metabolit an, ob es mit einem Schwellenwert als Biomarker zur Differenzierung zwischen diabetischen und nicht-diabetischen Tieren geeignet ist. Die heilungsunterstützende Wirkung einige der so erhaltenen Metabolite konnten in Validierungsstudien der TP1-Partner bestätigt werden (Dias et al., Acta Biomaterialia, in Revision). Die molekulare Subtypisierung, d. h. Stratifizierung, der Ratten anhand der Metabolom-Daten war nur eingeschränkt möglich und wird wie bei den Proteom-Daten zukünftig weiter verfolgt (siehe oben).

Im Rahmen der Biocrates-Analyse lagen Metabolomdaten sowohl zu Knochen- als auch zu Blutplasma-Proben vor. Der Datensatz der Knochenmetabolom-Analysen umfasste 455 Metabolite und 32 Proben, wovon zwei Proben, die mehr als 90 % fehlende Werte aufwiesen, von den weiteren Analyseschritten ausgeschlossen wurden. Der gesamte Datensatz wurde so gefiltert, dass kein Metabolit mehr als 20 % fehlende Werte aufwies. Somit umfasste der gefilterte Datensatz 344 Metabolite, 30 Proben und insgesamt 3,8 % fehlende Werte. Diese 30 Proben stammten zu gleichen Teilen aus dem kontralateralen und dem operierten Bein, jeweils zusammengesetzt aus 7 Proben von diabetischen und 8 Proben von nicht-diabetischen Tieren.

Für die Blutplasma-Metabolom-Daten lagen initial 72 Proben und 106 Metabolite vor. Nachdem drei Proben als Ausreißer identifiziert und der Datensatz entsprechend gefiltert worden war, umfasste dieser 69 Proben mit je 99 Metaboliten und insgesamt 1,4 % fehlenden Werten. Die Gesamtzahl der 69 Proben setzt sich aus 35 Proben von diabetischen Tieren und 34 Proben von nicht-diabetischen Tieren zusammen. Von den diabetischen Tieren waren 27 Proben von Knochendefekten, die mit einem Scaffold versorgt wurden, während 8 Proben ohne Scaffold vorlagen. Von den Proben der nicht-diabetischen Tiere wiesen 30 ein Scaffold auf, während in 4 Proben kein Scaffold vorhanden war. Die PLS-DA-Analyse ergab, dass sich Knochenproben von diabetischen und nicht-diabetischen Tieren mit einer Genauigkeit von 82 % unterscheiden ließen. Die Differenzierung zwischen operierten und kontralateralen Gliedmaßen erreichte eine Genauigkeit von 90 %. Bei den Blutplasma-Daten konnten Proben von diabetischen und nicht-diabetischen Tiere mittels PLS-DA mit 95 % Genauigkeit voneinander abgegrenzt werden, während die Differenzierung zwischen Proben mit und ohne Scaffold lediglich 60 % Genauigkeit erreichte.

Im weiteren Verlauf wurden die Plasma-Metabolom-Proben, die über mehrere Zeitpunkte vor und nach der Operation erhoben wurden, hinsichtlich signifikanter Abundanzänderungen der Metaboliten untersucht ( $\log_{2}FC = 0.5$ ,  $p\text{-value} = 0.05$ ). Zur Erfassung der zeitlichen Dynamik erfolgte eine differentielle Expressionsanalyse an benachbarten Zeitpunkten, die getrennt für nicht-diabetische und diabetische Proben durchgeführt wurde.

Bei den Proben nicht-diabetischer Tiere zeigte sich im kurzfristigen Verlauf – von drei Tagen vor bis einen Tag nach der Operation – eine Herabregulation von 18 Metaboliten. Mit zunehmendem zeitlichem Abstand zur Operation nahm die Zahl der herabregulierten Metaboliten kontinuierlich ab, während die Anzahl der hochregulierten Metaboliten auf bis zu 17 anstieg.

Ein ähnlicher Ansatz wurde bei den Proben der diabetischen Tiere verfolgt. Im Zeitraum von postoperativem Tag 1 bis Tag 3 erreichte die Zahl der herabregulierten Metaboliten mit 19 ihren Höchstwert, während im Intervall zwischen Tag 3 und Tag 21 der größte Wert an hochregulierten Metaboliten mit 12 beobachtet werden konnte. Anders als bei den Proben der nicht-diabetischen Tiere ließ sich hier jedoch keine kontinuierliche Abnahme der herabregulierten beziehungsweise ein kontinuierlicher Anstieg der hochregulierten Metaboliten über die Zeit erkennen.

Zusätzlich wurden direkte Vergleiche der Metabolitenabundanz zwischen diabetischen und nicht-diabetischen Proben zum gleichen Zeitpunkt vorgenommen. Einen Tag nach der Operation wurde der größte Unterschied mit insgesamt 30 differentiell regulierten Metaboliten festgestellt. Im weiteren zeitlichen Verlauf nahmen diese Differenzen ab: So wurden 42 Tage postoperativ lediglich 10 differentiell regulierte Metaboliten detektiert, was annähernd dem Niveau von 11 differentiell regulierten Metaboliten drei Tage vor der Operation entspricht.

Bei den Knochen-Daten standen die Zeitpunkte 21 und 42 Tage nach der Operation für die differentielle Abundanzanalyse ( $\log_{2}FC = 1$ ,  $p\text{-value} = 0.05$ ) zur Verfügung. In den nicht-diabetischen Proben wurden im kontralateralen Bein zwischen Tag 21 und Tag 42 insgesamt 41 hochregulierte Metaboliten detektiert, während im operierten Bein nur 25 hochregulierte Metaboliten identifiziert wurden. Die Zahl der herabregulierten Metaboliten war mit jeweils 1 im kontralateralen und 4 im operierten Bein vergleichsweise gering. Bei den diabetischen Proben wurde im kontralateralen Bein lediglich ein hochregulierter Metabolit festgestellt, wohingegen im operierten Bein 63 hochregulierte Metaboliten identifiziert wurden. Herabregulierte Metaboliten traten ausschließlich im kontralateralen Bein auf, wo 8 entsprechende Metaboliten detektiert wurden, während im operierten Bein keine herabregulierten Metaboliten beobachtet wurden.

Zudem wurden die diabetischen und nicht-diabetischen Proben zum selben Zeitpunkt verglichen. Hier zeigte sich, dass im diabetischen operierten Bein 89 hochregulierte Metaboliten vorlagen, verglichen mit dem nicht-diabetischen operierten Bein. Zum Zeitpunkt von 42 Tagen betrug die Anzahl der hochregulierten Metaboliten im diabetischen operierten Bein 129. Die Anzahl an herabregulierten Metaboliten war mit jeweils 2 zum Zeitpunkt von 21 Tagen und 1 zum Zeitpunkt von 42 Tagen vergleichsweise gering.

Bezüglich der kontralateralen Proben wurden zum Zeitpunkt von 21 Tagen 6 hochregulierte und 3 herabregulierte Metaboliten gemessen. Zum Zeitpunkt von 42 Tagen wurden in diesen Proben 21 herabregulierte Metaboliten gemessen, jedoch keine hochregulierten Metaboliten.

Im weiteren Verlauf unserer Arbeit werden wir diese Ergebnisse nutzen, um potenzielle Biomarker zur Differenzierung zwischen diabetischen und nicht-diabetischen Knochenheilungsprozessen zu identifizieren. Dazu planen wir bezüglich der Metabolomdaten den Einsatz weiterführender Analysemethoden, wie beispielsweise der Metabolite Set Enrichment Analysis und der ROC-Kurvenanalyse zur Biomarker-Identifizierung. Darüber hinaus sollen auch multi-OMICs Ansätze angewandt werden, um den Informationsgehalt aus Proteomik- und Metabolomik-Daten miteinander zu verknüpfen und in einen funktionellen Kontext zu setzen.

### **Transkriptomik:**

In Vorbereitung der Analyse von avisierten Transkriptomik-Daten entwickelten wir das Computerprogramm DRaCOoN (Delgado-Chaves et al., BMC Bioinformatics, in Revision, <https://pypi.org/project/draconopy/>) für die de novo Rekonstruktion von differentiell regulierten biologischen Netzwerken anhand eines bereits publizierten Transkriptom-Datensatz zur Knochenheilung im Tiermodell.

Am Ende der zweiten SyMBoD-Förderperiode wurden sowohl neue Proteomik- als auch Metabolomik-Datensätze der neuen ZDF-Ratten-Kohorte gewonnen. Mit denen während der Projektlaufzeit entwickelten Methoden und Workflows wird derzeit von Mitarbeitenden der Uni Hamburg die Analyse dieser Daten weiterverfolgt.

Für Erklärungen zu Abläufen bei der Gewinnung und Auswertung von Patientendaten verweisen wir auf die Projektberichte der Partner Fraunhofer IZI und Charité Berlin.

## **Teilprojekt 5 - Plattform für integrative Systemmedizin**

AP1.3 - Integration von Analysemodulen aus TP3 in die SyMBoD Plattform

**Zielstellung:** Entwicklung einer Plattform zur Integration von Forschungs- und klinischen Daten und deren Verknüpfung mit Bioinformatik, maschinellem Lernen und Simulationsmodellen.

Bereitstellung und Integration von Analysemodulen in die SyMBoD Plattform:

In Zusammenarbeit mit der Genevention GmbH werden R-Skripte für Analyse-Workflows aus TP3, wie die differenzielle Expressionsanalyse (limma) in der Proteomik sowie das Network Enrichment (KPM), angepasst und in die SyMBoD Plattform integriert. Für die molekulare Subtypisierung auf der Grundlage von Proteomik- und Metabolomikdaten aus Tiermodellen wurde MoSBi (Rose et al. 2022, PNAS, <https://exbio.wzw.tum.de/mosbi/>) ebenfalls auf der SyMBoD-Plattform zur Verfügung gestellt. Es wurden ebenfalls unsere neuen Programme ProHarMed (<https://apps.cosy.bio/proharmed/>) und drugst.One (<https://drugst.one/>) zum Modul-Mining und Netzwerk-basierter Medikamenten-Umwidmung erfolgreich integriert.

Während der stetigen Verbesserung unserer bioinformatischen Workflows und der fortschreitenden Datenanalyse der Plasma- und Knochen-OMICs Daten war es immer wieder notwendig, Anpassungen der bisherigen Analyseprogramme vorzunehmen. Dies führte zum wiederholten Aktualisierung und Kompilierung der bereits auf der Plattform

integrierten Analysemodule. Um dieses Vorgehen weitestgehend zu automatisieren, etablierten wir in der zweiten Förderperiode zusätzliche öffentlich einsehbare Software-Verzeichnisse (git) für die Versionierung und einen erleichterten Zugang. Des Weiteren wurde die Kontrolle der Programmierumgebung, und damit die Verwendung der korrekten Software-Paketversionen, durch Docker-Container bewerkstelligt. Dies erlaubt der SyMBoD-Plattform die aktuellste Version der Analysemodule abzurufen und somit dem Nutzer zur Verfügung zu stellen.

Für Erklärungen zu Abläufen bei der Integration und Auswertung von Patientendaten verweisen wir auf den Projektbericht der Partner Genevention GmbH.

### 1. Wichtigsten Positionen des zahlenmäßigen Nachweises

Die Mittel wurden frist- und projektbezogen eingesetzt. Alle Stellen wurden ohne Verzögerung besetzt. Die Arbeitspakete wurden entsprechend der Projektbeschreibung abgearbeitet und somit die Personalkosten zur Erfüllung der Aufgaben eingesetzt.

### 2. Notwendigkeit und Angemessenheit der geleisteten Projektarbeiten

Alle Stellen wurden ohne Verzögerung besetzt. Die Arbeitspakete wurden entsprechend der Projektbeschreibung abgearbeitet und somit die Personalkosten zur Erfüllung der Aufgaben eingesetzt. Die gesamtheitliche Auswertung der Ratten multi-OMICs Daten von Plasma und Knochengewebe und die Akquise vergleichbarer Daten zur Modellierung der Patientenkohorte in der ersten Phase verzögerten sich aufgrund der COVID-19-Pandemie. Die entstandenen Verzögerungen bei der Datenproduktion und -übertragung wurden weitestgehend in dieser Phase aufgeholt und die bioinformatischen Arbeiten laut Planung abgeschlossen. Für Erklärungen zur Probengewinnung verweisen wir auf die Projektberichte der Partner Fraunhofer IZI und Charité Berlin.

### 3. Voraussichtliche Nutzen, insbesondere die Verwertbarkeit des Ergebnisses - auch konkrete Planungen für die nähere Zukunft - im Sinne des fortgeschriebenen Verwertungsplans

Die vom SyMBoD-Konsortium entwickelten Programme und Workflows, wie z. B. zur systematischen Evaluierung von Normalisierungsmethoden auf bioinformatische Downstream-Analysen; zur Meta-Analyse von interdisziplinären, multi-Spezies OMICs-Studien sowie die Netzwerk-basierte Medikamenten-Umwidmung und das automatisierte Clustern von zeitaufgelösten OMICs-Daten, und deren Bereitstellung auf einer integrativen Systemmedizin-Plattform, sind ein wichtiger Beitrag zur Analyse von molekularen Veränderungen in Zusammenhang mit komplexen Krankheitsbildern. Die Veröffentlichung einer Vielzahl der erreichten Ergebnisse wird derzeit vorbereitet bzw. befindet sich bereits kurz vor Veröffentlichung (in Revision). Vorhaben für die nähere Zukunft sind die Weiterverfolgung der Validierung der Ergebnisse durch Analysen der in der zweiten Förderperiode gewonnenen Daten der ZDF Ratten-Kohorten. Hierfür können die bereits

entwickelten Workflows weiterhin zur Auswertung der ZDF Ratten-Kohorten und Validierungsstudien verwendet werden.

Anhand der Meta-Analyse und unseren analogen Analysen von Plasma- und Knochen-OMICs-Daten zur Identifizierung möglicher theragnostischer Biomarker, die anhand der Veränderungen im Blut die Heilungsprognose des Gewebes vorhersagen können, eröffnen die Möglichkeit einer minimal-invasiven Vorhersage des Krankheitsverlaufs und könnten somit einen wichtigen Beitrag zur Verbesserung der Diagnostik und Therapie von T2DM Patienten leisten. Insbesondere die Bereitstellung dieser Analysen in der Patientenversorgung durch innovative digitale Technologien aus der Systemmedizin ist zukunftsweisend, sodass unsere Arbeiten einen wichtigen Anknüpfungspunkt für weiterführende Arbeiten bieten.

4. Während der Durchführung des Vorhabens dem Zuwendungsempfänger bekannt gewordenen Fortschritt auf dem Gebiet des Vorhabens bei anderen Stellen

Es ist nichts bekannt.

5. Erfolgten oder geplanten Veröffentlichungen des Ergebnisses nach Nr. 5 der NKBF/NABF

Im Rahmen des SyMBoD Projektes wurden folgende Publikationen vorbereitet und veröffentlicht:

**Geplante:**

Arend, Adamowicz, Schmidt, Burankova, Zolotareva, Tsoy, Pauling, Kalkhof, Baumbach, List, Laske. Systematic Evaluation of Normalization Approaches in Tandem Mass Tag and Label-Free Protein Quantification Data Using PRONE. Akzeptiertes Manuskript: Briefings in Bioinformatics. Pre-Print: <https://doi.org/10.1101/2025.01.27.634993>

Delgado-Chaves, Spurny, Laske, Oubounyt, Baumbach. DRaCOoN: A Novel Algorithm for Pathway-level Differential Co-expression Analysis in Transcriptomics. In Revision: BMC Bioinformatics. <https://pypi.org/project/draconpy/>

Dias, Chan, Ellinghaus, Fritsche-Guenther, Bembenek, Laske, Baumbach, Duda, Kirwan, Poh. Endogenous dysregulated energy and amino acid metabolism delays scaffold-guided large volume bone regeneration in a diabetic rat model with Leptin receptor deficiency. In Revision: Acta Biomaterialia.

Wiltzsch, Schmidt, Adamowicz, Lauterbach, Lehmann, Baumbach, Laske, Kalkhof. Systematic Comparison of Bone Proteome Extraction Methods to Allow for Integrated Proteomics-Metabolomics-Correlation. In Revision: Journal of Proteome Research.

Schmidt, Adamowicz, Wiltzsch, Dias, Arend, Lio, Lehmann, Lingner, Tsoy, Baumbach, Laske, Poh, Kalkhof. Longitudinal plasma proteomics and reveals differential systemic molecular responses in scaffold-guided bone regeneration at diabetic conditions (in Vorbereitung)

## Erfolgte:

Moral-Morales A, Sámano C, Ocampo-Cervantes JA, Topf M, Baumbach J, González-Barrios R, Soto-Reyes E (2024) Key Proteins for Regeneration in *A. mexicanum*: Transcriptomic Insights from Aged and Juvenile Limbs. *Scientifica* 2024 Nov 14;2024:5460694. <https://doi.org/10.1155/2024/5460694>

Schmidt, Adamowicz, Arend, Lehmann, List, Poh, Baumbach, Kalkhof, Laske. Meta-analysis of proteomics data from osteoblasts, bone and blood: Insights into druggable targets, active factors and potential biomarkers for bone biomaterial design. *J Tissue Eng* 15 (2024). <https://doi.org/10.1177/20417314241295332>

Kersting J, Lazareva O, Louadi Z, Baumbach J, Blumenthal D, List M (2024) DysRegNet: Patient-specific and Confounder-aware Dysregulated Network Inference towards Precision Therapeutics. *Br. J. Pharmacol.* 2024 Dec 4. <https://doi.org/10.1111/bph.17395>

Adamowicz, Arend, Maier, Schmidt, Kuster, Tsoy, Zolotareva, Baumbach, Laske. Proteomic Meta-Study Harmonization, Mechanotyping and Drug Repurposing Candidate Prediction with ProHarMeD. *npj Syst Biol Appl* 9, 49 (2023). <https://doi.org/10.1038/s41540-023-00311-7>

Del Moral-Morales A, Salgado-Albarrán M, Sánchez-Pérez Y, Wenke NK, Baumbach J, Soto-Reyes E (2023) CTCF And Its Multi-partner Network For Chromatin Regulation. *Cells* 2023 May 10;12(10),1357. <https://doi.org/10.3390/cells12101357>

Oubounyt M, Elkjaer ML, Laske T, Grønning AGB, Moeller MJ, Baumbach J (2023) De-novo reconstruction and identification of transcriptional gene regulatory network modules differentiating single-cell clusters. *NAR Genome Bioinform.* 2023 Mar 3;5(1):lqad018. <https://doi.org/10.1093/nargab/lqad018>

Adamowicz K, Maier A, Baumbach J, Blumenthal DB (2022) Online in silico validation of disease and gene sets, clusterings, or subnetworks with DIGEST. *Brief Bioinform.* 2022 Jul 18;23(4):bbac247. <https://doi.org/10.1093/bib/bbac247>

Hartung M, Anastasi E, Mamdouh Z, Nogales CC, Schmidt HHH, Baumbach J, Zolotareva O, List M (2022) Cancer Driver Drug Interaction Explorer. *Nucleic Acids Res.* 2022 May 17;gkac384. <https://doi.org/10.1093/nar/gkac384>

Rose TD, Bechtler T, Ciora OA, Le KAL, Molnar F, Koehler N, Baumbach J, Röttger R, Pauling JK (2022) MoSBI: Automated signature mining for molecular stratification and subtyping. *Proc Natl Acad Sci U.S.A.* 2022 Apr 11; 119 (16) e2118210119. <https://doi.org/10.1073/pnas.2118210119>

Bernett J, Krupke D, Sadegh S, Baumbach J, Fekete SP, Kacprowski T, List M, Blumenthal DB (2022) Robust disease module mining via enumeration of diverse prize-collecting Steiner trees. *Bioinformatics* 2022 Jan 4;btab876. <https://doi.org/10.1093/bioinformatics/btab876>

Sadegh S, Skelton J, Anastasi E, Bernett J, Blumenthal DB, Galindez G, Salgado-Albarrán M, Lazareva O, Flanagan K, Cockell S, Nogales C, Casas AI, Schmidt HHHW, Wipat A,

Baumbach J, Kacprowski T (2021) Network medicine for disease module identification and drug repurposing - The NeDRex platform. *Nat Commun.* 2021 Nov 25;12(1):6848.  
<https://doi.org/10.1038/s41467-021-27138-2>

Matschinske J, Alcaraz N, Benis A, Golebiewski M, Grimm DG, Heumos L, Kacprowski T, Lazareva O, List M, Louadi Z, Pauling JK, Pfeifer N, Röttger R, Schwämmle V, Sturm G, Traverso A, van Steen K, Vaz de Freitas M, Silva GCV, Wee L, Wenke NK, Zanin M, Zolotareva O, Baumbach J, Blumenthal DB (2021) The AIME registry for artificial intelligence in biomedical research. *Nat Methods* 18:1128-1131, Aug 25 2021.  
<https://doi.org/10.1038/s41592-021-01241-0>

Sarno F, Benincasa G, List M, Barabasi AL, Baumbach J, Ciardiello F, Filetti S, Glass K, Loscalzo J, Marchese C, Maron BA, Paci P, Parini P, Petrillo E, Silverman EK, Verrienti A, Altucci L, Napoli C (2021) Clinical Epigenetics Settings for Cancer and Cardiovascular Diseases: Real-life Applications of Network Medicine at the Bedside. *Clinical Epigenetics* 2021 Mar 30;13(1):66 <https://doi.org/10.1186/s13148-021-01047-z>

Manz Q, Tsoy O, Fenn A, Baumbach J, Völker U, List M, Kacprowski T (2021) ASimulatoR: splice-aware RNA-Seq data simulation. *Bioinformatics* 2021 Feb 27;btab142.  
<https://doi.org/10.1093/bioinformatics/btab142>

Lazareva O, Baumbach J, List M, Blumenthal DB (2021) On the limits of active module identification. *Brief Bioinform.* 2021 Mar 29;bbab066. <https://doi.org/10.1093/bib/bbab066>

Weber P, Pauling JK, List M, Baumbach J (2020) BALSAM - An Interactive Online Platform for Breath Analysis, Visualization and Classification. *Metabolites* 2020 Oct 2;10(10):E393.  
<https://doi.org/10.3390/metabo10100393>

Sadegh S, Matschinske J, Blumenthal DB, Galindez G, Kacprowski T, List M, Nasirigerdeh R, Oubounyt M, Pichlmair A, Rose TD, Salgado-Albarrán M, Späth J, Stukalov A, Wenke NK, Yuan K, Pauling JK, Baumbach J (2020) Exploring the SARS-CoV-2 virus-host-drug interactome for drug repurposing. *Nat Commun.* 2020 Jul 14;11(1):3518.  
<https://doi.org/10.1038/s41467-020-17189-2>