

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

Abschlussbericht

LAMarCK - Longitudinal Multiomics Characterization of disease using prior Knowledge

**CompLS - Runde 1 - Verbundprojekt: LAMarCK - Charakterisierung von
Krankheiten mittels longitudinaler Multi-Omics Analyse**

Förderkennzeichen: 031L0181A

Förderperiode 01.04.2019 – 31.12.2023

Verbundkoordinator und Projektleiter für Teilprojekt A:

Dr. Georg Zeller
European Molecular Biology Laboratory
Meyerhofstr. 1
69117 Heidelberg
Tel.: +49 6221 387-0
E-mail: zeller@embl.de

Teilprojektpartner:

Prof. Julio Saez-Rodriguez
Director, Institute for Computational Biomedicine
Heidelberg University, Faculty of Medicine,
Im Neuenheimer Feld 1303, 10th floor 69120 Heidelberg

Prof. Dr. Matthias Schlesner
Augsburg University
Biomedical Informatics, Data Mining, Data Analytics
Alter Postweg 101 (Büro Center Messe)
86159 Augsburg

Prof. Dr. Peer Bork
European Molecular Biology Laboratory
Structural and Computational Biology Unit
Meyerhofstr. 1
69117 Heidelberg

Prof. Dr. Jan Korbel
European Molecular Biology Laboratory
Genome Biology Unit
Meyerhofstr. 1
69117 Heidelberg

Die Verantwortung für den Inhalt dieser Veröffentlichung liegt beim Autor.

Teil I. Kurzbericht

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

Das LAMarCK-Projekt hatte sich zum Ziel gesetzt, Software-Tools zu entwickeln, die eine integrative Analyse von Multi-omics-Daten ermöglichen. Auf diese Weise sollten biomedizinische Einsichten in (patho-)physiologische Zustände des menschlichen Organismus gewonnen werden. Die Differenzierung von Krankheits- und Gesundheitsstadien sowie die Unterscheidung verschiedener Altersgruppen sind Beispiele für solche Zustände mit translationaler Relevanz. Dazu sollte insbesondere die Nutzung zeitlich und räumlich aufgelöster Omics-Daten unter Einbeziehung biologischen Vorwissens optimiert werden. Im Rahmen des Projekts wurde nicht nur der Fokus auf molekulare Daten menschlicher Zellen gelegt, sondern auch das Mikrobiom berücksichtigt. Für beide Typen von Omics-Daten, sowohl für die des Mikrobioms als auch die des menschlichen Wirts, ergeben sich ähnliche Herausforderungen hinsichtlich integrativer Analysen. Die Integration dieser Daten verspricht neuartige biomedizinische Erkenntnisse, beispielsweise für die Entwicklung von Biomarkern zur Vorhersage von Frühstadien bestimmter Krebsarten.

2. Ablauf des Vorhabens

Im Rahmen des LAMarCK-Projekts kooperierten fünf Forschungsgruppen, die sich primär mit Bioinformatik / Computerbiologie befassen. Ihr Fokus liegt auf der Entwicklung, Integration und Bewertung von Methoden zur Analyse von Multi-Omics. Außerdem generieren alle Arbeitsgruppen selbst entsprechende Daten, wobei die Finanzierung aus anderen Quellen erfolgt. Die Koordination der gemeinsamen Arbeit im LAMarCK-Projekt erfolgte in halbjährlichen Treffen, unter Nutzung moderner Kommunikationssysteme wie Slack, Zoom und GitHub sowie in zusätzlichen Meetings (virtuell und persönlich), je nach Bedarf. Da die Projekt-Arbeiten vielfach auf früheren und laufenden Arbeiten, Datenbanken und Softwaretools aller beteiligten Gruppen basierten, wurden viele dieser Tools und Analysestrategien gemäß den Anforderungen von (longitudinalen und räumlichen) Multi-Omics-Analysen weiterentwickelt.

3. Ergebnisse sowie ggf. die Zusammenarbeit mit anderen Forschungseinrichtungen

Zum Zeitpunkt der Antragstellung lag der Fokus des Projekts auf der Analyse von longitudinalen Omics-Daten aus Zeitreihen. Dass räumlich aufgelöste Omics-Daten ähnliche Analyse-Anforderungen wie Zeitreihen stellen würden, war schon antizipiert worden – zeitliche (lineare) Abhängigkeiten entsprechen dabei räumlichen Abhängigkeiten in zwei oder drei Dimensionen. Bei Antragstellung waren allerdings Spatial-Omics Daten praktisch nicht verfügbar, wohingegen bereits einige publizierte longitudinale Multi-Omics-Datensätze existierten. Im Verlauf des Projekts veränderte sich dieses Verhältnis jedoch grundlegend: Während kaum weitere longitudinale Multi-Omics-Datensätze veröffentlicht wurden (insbesondere nicht von Studien, die Mikrobiom und menschlichen Organismus gleichzeitig untersuchten), entstanden in kurzer Folge viele neuartige räumliche Omics-Technologien, was zur Publikation vieler interessanter Spatial-Omics Datensätze führte (z.B. Sequenzierungs-basierte Spatial Transcriptomics, Mikroskopie-basierte Multiplex-FISH-Techniken und Antikörper-basierte Spatial Proteomics-Techniken).

Entsprechend hat das LAMarCK-Konsortium im Laufe des Projekts seine bioinformatischen Methodenentwicklungs-Aktivitäten refokussiert in Synergie mit weiteren Projekten der beteiligten Arbeitsgruppen (finanziert aus anderen Quellen), in denen auch Spatial Multi-Omics-Daten generiert wurden: So untersuchten wir beispielsweise die räumlichen Strukturen von Darmtumoren unter Berücksichtigung mikrobieller Besiedelung, Zelltyp-Verteilung, Tumorgenotypen und Expressionsmuster anhand von räumlich aufgelösten Multi-Omics- und Imaging-Daten. Alle Projektpartner sind in mehrere Forschungsprojekte zu Spatial-Omics Datengenerierung und Analyse eingebunden und kooperieren dazu auch vielfach mit weiteren Partnern. Im Rahmen von LAMarCK wurden entsprechende Analyse-Werkzeuge für longitudinale und räumliche Omics-Daten entwickelt und publiziert (**Türei et al., Mol. Syst. Biol., 2021; Dimitrov et al., Nat. Commun., 2022; Velten et al., Nat. Methods 2022; Dimitrov et al., bioRxiv, 2023**).

Neben dem Darmmikrobiom rückten im Verlauf des LAMarCK-Projekts Gewebe-residente Mikroben in den Fokus der Untersuchungen. Dies ist darauf zurückzuführen, dass durch eigene Studien und Arbeiten Dritter deutlich wurde, dass sich das Tumor-Gewebe-Mikrobiom direkt aus klassischen Omics-Daten von menschlichen Tumorproben charakterisieren lässt. Zu diesen Daten zählen beispielsweise Tumorgenome und -Transkriptome. Aufgrund der geringen mikrobiellen Biomasse sind die bioinformatische Analysen technisch anspruchsvoll, jedoch derzeit weltweit von großem Interesse, um molekulare Interaktionen zwischen mikrobiellen und Wirtszellen im Tumorgewebe untersuchen zu können. In diesem Zusammenhang hat das LAMarCK-Konsortium mehrere mikrobielle Profiling-Tools in Kombination mit verschiedenen Omics-Daten evaluiert, bioinformatische Workflows validiert und publiziert (Ruscheweyh, Milanese et al., *Microbiome*, 2022; Sambruni et al., *Genome Med.*, 2023).

Des Weiteren wurde, sofern sinnvoll und möglich, auch die integrative Analyse mit Einzelzell-Daten in einem größeren Umfang berücksichtigt als ursprünglich angenommen. Die Arbeitsgruppen Korbelt und Saez-Rodriguez sind in diesem Forschungsfeld sehr aktiv und kooperieren eng mit anderen führenden Forschungsgruppen (beispielsweise mit der Arbeitsgruppe Stegle, EMBL/DKFZ Heidelberg, und der Arbeitsgruppe Huber, EMBL Heidelberg). Des Weiteren entwickelten die AGs Zeller, Korbelt und Saez-Rodriguez Verfahren zur Detektion intrazellulärer oder adhäsiver Mikroben anhand von Einzelzell-Daten. Dies ist beispielsweise für intrazelluläre Pathogene wie *Fusobacterium nucleatum* im Kontext von Darmtumoren (und anderen Krebserkrankungen) von großem Interesse. Die AGs Korbelt und Zeller arbeiten derzeit an der Erstellung eines Manuskripts, in dem die Verfahren und biologischen Erkenntnisse aus Einzelzell-Daten und räumlich aufgelösten Multi-Omics-Daten von Darmtumoren beschrieben werden.

Mehr noch als im Projektantrag formuliert, standen Benchmarks und Validierung bestehender statistischer Analyseverfahren und Methoden im Fokus. Wir arbeiteten an mehreren Benchmarks von statistischen Analyseverfahren für Mikrobiom-Krankheits-Assoziationen, sowohl für Daten aus Querschnittsstudien, als auch aus Zeitreihen. Oft erwiesen sich dabei klassische statistische Techniken als geeignet (und neuartigen Methoden mehr als nur gelegentlich überlegen), weswegen sich das LAMarCK-Konsortium eher darauf konzentrierte, diese klassischen Techniken mit geeigneter Präprozessierung und Vorwissen zu koppeln (Wuyts, Alves, Zimmermann-Kogadeeva, Nishijima et al. *Mol. Syst. Biol.*, 2023; Lobentanzer et al., *Nat. Biotechnol.*, 2023; Wirbel, Essex et al., *bioRxiv* 2022).

Aufbauend auf die innerhalb des LAMarCK-Projekts durchgeführten Arbeiten und entwickelten Analyse-Werkzeuge ergaben sich weitere, in der Regel angewandte, Drittmittel-geförderte Projekte, an denen eine oder mehrere Arbeitsgruppen des Konsortiums beteiligt sind: So arbeiten die AGs Zeller und Korbelt im Rahmen des von der Baden-Württemberg-Stiftung geförderten Spatial-CRC-ECO Projekts an einer räumlichen Kartierung des Darmkrebs-Ökosystems unter Berücksichtigung heterogener Wirtszellpopulationen und Mikroben. Weitergehend arbeiten die AGs Korbelt und Zeller in einem Health + Life Science Alliance Heidelberg Mannheim Stimulator Projekt auch an der entsprechenden Charakterisierung hereditärer Darmkrebsproben. Die AGs Saez-Rodriguez und Zeller sind an dem ERC Synergy Projekt CartoHostBug beteiligt, das sich die funktionale Kartierung von molekularen und zellulären Mikrobiom-Wirts-Interaktionen in Darmkrebs und chronisch entzündlichen Darmerkrankungen in menschlichen Patienten und Mausmodellen zum Ziel gesetzt hat. Im Rahmen des BMBF-geförderten MiEOCRC Projekts, an dem die AG Zeller beteiligt ist erfolgt eine Untersuchung von Darmkrebs bei jungen Patienten mittels (prospektiver) longitudinaler Multi-Omics-Studien. Zusammenfassend demonstrieren diese Projekte, die auf LAMarCK aufbauen, eindrucksvoll, dass die Thematik hochaktuell ist und die entwickelten Analysestrategien das Potenzial besitzen, neue biologische Erkenntnisse zu gewinnen. Dadurch können wesentliche Impulse für die Erforschung menschlicher Krankheiten sowie für die Diagnostik und Therapie dieser gewonnen werden.

Teil II. Ausführlicher Bericht der im Rahmen des Vorhabens durchgeführten Arbeiten

1. Ausführliche Darstellung der durchgeführten Arbeiten im Vergleich zur ursprünglichen Vorhabenbeschreibung

Omics-Techniken, die Analysen von menschlichen Genomen, Transkriptomen, Epigenomen sowie des Darm-Mikrobioms ermöglichen, haben sich als Schlüsselwerkzeug erwiesen, um die Molekularbiologie von Krankheiten zu ergründen, Genotypen und Phänotypen mit Umweltfaktoren in Beziehung zu setzen und Anwendungen in der Präzisionsmedizin zu finden. Krankheitsprozesse sowie die Wechselwirkungen zwischen Mensch und Umwelt (die beispielsweise durch das Mikrobiom vermittelt werden können) sind hochdynamisch und erfordern die Erhebung von zeitlich oder räumlich aufgelösten Daten, deren Analyse weitere molekulare Erkenntnisse liefern kann.

Das LAMarCK-Projekt verfolgte daher vier Hauptziele, die ihrerseits durch wichtigen biomedizinischen Fragen motiviert sind:

1. Sammlung und Bereitstellung von biologischen Wissensgrundlagen, die dem Benchmarking, der Kontextualisierung und der Interpretation der Ergebnisse von Multi-Omics-Datenintegration dienen.
2. Implementierung von biologischem Wissen zur Unterstützung von Multi-Omics-Daten-Analyse, um die Gewinnung biologischer Erkenntnisse aus solchen Analysen zu ermöglichen und zu beschleunigen.
3. Entwicklung und Implementierung von Analysestrategien für zeitlich oder räumlich aufgelöste Multi-Omics-Daten, um auf molekularer Ebene Zell-Zell Interaktionen aufklären zu können, auch zwischen Mikrobiom und Wirt.
4. Validierung und Verbesserung der entwickelten Methodik anhand von Multi-Omics-Daten die aus Darmkrebsproben für Wirtszellen und Mikrobiom gewonnen wurden.

Die Arbeiten erfolgten in fünf Arbeitspaketen (WPs), deren Ergebnisse im folgenden detailliert dargestellt werden.

WP 1 Etablierung einer Datenbasis von strukturiertem biologischem Vorwissen zur Validierung

WP 1.1 Im Rahmen dieses Arbeitspakets wurde biologisches Vorwissen in Form von Expressionssignaturen sowohl aus (Bulk-) Transkriptomdaten als auch aus Einzelzelldaten gesammelt. Da der technologische Fortschritt während der Projektlaufzeit zu einer großen Fülle zelltyp- und physiologiespezifischer Signaturen führte, die präziser und leichter zu extrahieren oder zu identifizieren sind als Signaturen, die aus der Massentranskriptomik abgeleitet werden, wurde der Fokus auf Signaturen aus Einzelzelldaten gelegt. Außerdem wurde Dekonvolutionstechniken angewandt, um die weithin verfügbaren Transkriptomik-Massendaten, wie z. B. Tumorproben in den TCGA-Repositoryn, besser nutzen zu können. Im Rahmen des LAMarCK-Projekts haben wir mehrere Signaturressourcen in einem einheitlichen, vollständig integrierten Software-Framework zur Verfügung gestellt. Das Framework umfasst in erster Linie den OmniPath Prior Knowledge Processing Stack (Türei 2021) und das DecoupleR Activity Inference Framework (Badia-i-Mompel 2022), die beide in der Saez-Rodriguez Group entwickelt wurden. Außerdem wurden weitere Ressourcen, wie DoRothEA und PROGENy (beide aus der AG Saez-Rodriguez), sowie PanglaoDB und CytoSig (aus anderen Forschungsgruppen) integriert. Unsere Integration dieser Signaturen machte diese für eine Vielzahl von Einzelzell- als auch in Bulk-Transkriptomik-Analysen und Analyse-Workflows leicht verfügbar. Sämtliche Signaturen und alles andere Vorwissen stehen nicht nur für den Menschen, sondern auch für die Maus und andere Modellorganismen zur Verfügung.

Eine wichtige Anwendung dieser Signaturen ist die Abschätzung des Anteils verschiedener Zelltypen (Stroma- und Immunzellen) in großen Krebs-Bulk-Transkriptomik-Datensätzen mit

Hilfe sogenannter Dekonvolutionsverfahren. Zu diesem Zweck haben wir an der Organisation einer DREAM (Dialog for Reverse Engineering Assessments and Methods) Challenge zum Thema Tumordekonvolution teilgenommen (<https://www.synapse.org/#!/Synapse:syn15589870/wiki/582446>). Die Teilnahme an der Challenge ermöglichte es uns, die verfügbaren Methoden eingehend zu bewerten (Arbeiten der AG Saez-Rodriguez, Heidelberg University).

WP 1.2 Da die in WP 1.1 integrierten Signaturen nicht spezifisch für einen physiologischen Zustand, einen Zelltyp oder eine Krebsart sind, wurde in WP1.2 dieser Kontext inferiert. Dazu stützen wir uns auf die mechanistischen Modellierungsmethoden, die in der Saez-Rodriguez-Gruppe gut etabliert sind um Signalwege in unserer kuratierten Datenbank OmniPath zu kontextualisieren und Spezifität für physiologische Zustände abzuleiten. Die Methodik der Wahl, kausale Inferenz, ist im Softwarepaket CARNIVAL implementiert (Liu 2019), welches mit einer API versehen wurde und in Analyse-Pipelines integriert wurde.

Eine weitere Methode zur Ableitung zustandsspezifischer Signalzustände ist NicheNet, ein Tool zur Abschätzung von Ligandenaktivitäten aus Transkriptomikdaten, das in einer anderen Gruppe entwickelt wurde. Wir haben NicheNet anschließend an verschiedenen Problemen evaluiert, seinen Input an Vorwissen in unsere Datenbank OmniPath integriert und die Methode in das OmniPath R Client-Paket implementiert.

Schließlich haben wir auch ein Framework für die Inferenz von Liganden-Rezeptor-Aktivität entwickelt. Dieses LIANA-Framework umfasst mehr als 10 Methoden, verwendet Daten aus OmniPath und ist über Standarddatenstrukturen mit anderen Pipeline-Komponenten verbunden, was eine weitere Möglichkeit zur Ableitung von zustandsspezifischen Aktivitäten bietet (Dimitrov 2022). Um die Methoden in LIANA mit umfassendem Vorwissen zu unterstützen, haben wir mehr als 10 Liganden-Rezeptor-Datenbanken in OmniPath integriert (Türei 2021) (Arbeiten der AG Saez-Rodriguez, Heidelberg University).

WP 1.3 Die Anpassung und Erweiterung von Annotations-Ressourcen für Mikrobiomdaten bildete den Schwerpunkt von WP1.3. So wurden beispielsweise das taxonomische und phylogenetische Annotations- und Visualisierungstool iTOL (Letunic 2021), das taxonomische Profiling-Tool mOTUs (Ruscheweyh 2022) und die prokaryotische Genomdatenbank proGenomes (Fullam 2022) überarbeitet. Weiterhin haben ein neues Tool (GUNC) zur Qualitätskontrolle prokaryotischer Genome entwickelt (Orakov 2021). Mit einer neu entwickelten Analyse-Pipeline annotierten wir mobile genetische Elemente (MGEs) in solchen Genomen und fanden wir in der ersten größeren Vergleichsstudie ihrer Art heraus, dass diese ca. 13% eines typischen prokaryotischen Genoms ausmachen (Khedkar 2022). Darüberhinaus haben wir Tools integriert für die Annotation des mikrobiellen Sekundärstoffwechsels, um diese für funktionale Analysen von Mikrobiom-Multi-Omics-Daten nutzen zu können. Insbesondere haben wir ein sehr präzises Software-Tool (GECCO) zur Annotation und Analyse von biosynthetischen Gen-Clustern entwickelt (Carroll 2021). Schließlich haben wir eine großangelegte Re-Annotation des komplexen Kohlenhydrat-Stoffwechsels in bakteriellen Genomen durchgeführt (Ducarmon 2024). Um die funktionale Gen-Annotation durch verbesserte Parallelisierung generell zu beschleunigen, entwickelten wir außerdem das PyHMMER Softwaretool (Larralde 2023) (Arbeiten der AG Bork und AG Zeller, EMBL Heidelberg).

Publikationen des LAMarCK-Konsortiums im Zusammenhang mit WP1 (teilweise existieren Überschneidungen mit den in anderen WPs genannten Publikationen):

1. Liu et al., *NPJ Sys. Biol. Appl.*, 2019 (doi: 10.1038/s41540-019-0118-z)
2. Türei et al., *Mol. Syst. Biol.*, 2021 (doi: 10.15252/msb.20209923)
3. Letunic and Bork, *Nucl. Acids Res.*, 2021 (doi: 10.1093/nar/gkab301)
4. Carroll, Larralde, Fleck et al., *bioRxiv*, 2021 (doi: 10.1101/2021.05.03.442509)
5. Orakov, Fulham et al., *Genome Biol.*, 2021 (doi: 10.1186/s13059-021-02393-0)
6. Cantalapiedra et al., *Mol. Biol. Evol.*, 2021 (doi: 10.1093/molbev/msab293)
7. van Rossum et al., *Bioinform.*, 2022 (doi: 10.1093/bioinformatics/btab789)
8. Khedkar et al., *Nucl. Acids Res.*, 2022 (doi: 10.1093/nar/gkac163)
9. Ruscheweyh, Milanese et al., *Microbiome*, 2022 (doi: 10.1186/s40168-022-01410-z)
10. Fullam et al., *Nucl. Acids Res.*, 2022 (doi: 10.1093/nar/gkac1078)

11. **Dimitrov et al., Nat. Commun., 2022** (doi: 10.1038/s41467-022-30755-0)
12. **Badia-i-Mompel et al., Bioinform. Adv., 2022** (doi: 10.1093/bioadv/vbac016)
13. **Larralde and Zeller, Bioinformatics, 2023** (doi: 10.1093/bioinformatics/btad214)
14. **Ducarmon, Karcher et al., bioRxiv, 2024** (doi: 10.1101/2024.01.08.574624)

WP 2 Bereitstellung und Vorverarbeitung von longitudinalen Multi-Omics-Daten

WP 2.1 Entgegen der bei Antragstellung gemachten Annahme, wurden im Verlauf des LAMarCK-Projekts kaum longitudinalen Multi-Omics-Daten publiziert, insbesondere nicht im Feld der Krebsgenomik. Während der COVID-19-Pandemie haben wir jedoch auch entsprechende Datensätzen zu dieser Krankheit gesammelt und aufbereitet (u.a. Bouhaddou 2020). Als Resultat haben wir eine COVID-19 Disease-Map veröffentlicht, eine umfassende, kuratierte computergestützte Wissensdatenbank für SARS-CoV-2-Virus-Wirt-Interaktionen und Krankheitsmechanismen (Ostaszewski 2021). In diesem Zusammenhang haben wir auch Interaktionswege zwischen SARS-CoV-2-Virus und Wirt in in-vivo Darmgewebe und Darmorganoiden mit scRNA-Seq-Daten analysiert (Poletti 2021). Damit war diese Studie eine erste exemplarische Analyse von Mikroben-Wirt-Interaktionen anhand von menschlichen Transkriptomdaten (Arbeiten der AG Saez-Rodriguez, Heidelberg University).

In einer Studie, die durch die Entwicklung von ArbiGent, einem Tool für die Genotypisierung von Strukturellen Genomvarianten anhand von DNA Einzelstrang-Sequenzierung (Strand-Seq), ermöglicht wurde, katalogisierten wir umfassend polymorphe Inversionen. Dieser neuartige und umfassende Datensatz erlaubte uns erstmals das Phänomen der "Rekurrenz" von Inversionen in menschlichen Genomen umfassend zu beschreiben, sowie neue Verbindungen zu mehreren Erbkrankheiten (Williams-Beuren Syndrome, 3q29 Microdeletion, 15q13.3 Microdeletion) aufzuzeigen (Porubsky 2022) (Arbeiten der AG Korbel, EMBL Heidelberg).

WP 2.2 konzentrierte sich auf Mikrobiom-Zeitreihendaten. Unsere umfassende Sammlung (>5400 Metagenome von >1400 Individuen) ermöglichte umfassende Analysen des Genpools des menschlichen Mikrobioms, aber auch anderer prokaryotischer Gemeinschaften (Hildebrand 2021 und Coelho 2021). Multi-Omics-Datensätzen unter Einbeziehung des Mikrobioms wurden jedoch während der Laufzeit des LAMarCK-Projekts kaum publiziert (teilweise wird die Veröffentlichung solcher Daten durch DSGVO in Europa stark behindert, was zum Zeitpunkt der Antragstellung nicht absehbar war). Die vom US-amerikanischen HMP2 Konsortium publizierten longitudinalen Multi-Omics-Datensätze zum Darmmikrobiom bei (Prä-)Diabetes und Chronisch-entzündlichen Darmerkrankungen, sowie zum vaginalen Mikrobiom während der Schwangerschaft und Frühgeburt inklusive aller verfügbarer Metadaten wurden ein wichtiger Bestandteil unserer oben genannten Mikrobiom-Ressource.

Außerdem haben wir einen Referenzdatensatz bestehend aus einer künstlichen Bakteriengemeinschaft zu Benchmark-Zwecken selbst generiert. In dieser definierten *in vitro* Bakteriengemeinschaft sahen wir einerseits gute Übereinstimmungen zwischen metabolischen, metagenomischen meta-proteomischen und meta-transkriptomische Profilen (Wuyts 2023) (Arbeiten der AG Bork, EMBL Heidelberg).

Publikationen des LAMarCK-Konsortiums im Zusammenhang mit WP2 (teilweise existieren Überschneidungen mit den in anderen WPs genannten Publikationen):

1. **Bouhaddou, Memon, Meyer, White et al., Cell, 2020** (doi: 10.1016/j.cell.2020.06.034)
2. **Ostaszewski, Niarakis, Mazein, Kuperstein, Phair et al., Mol. Syst. Biol., 2021** (doi: 10.15252/msb.202110387)
3. **Poletti et al., npj Syst. Biol. Appl., 2021** (doi: 10.1038/s41540-022-00224-x)
4. **Hildebrand et al., Cell Host Microbe, 2021** (doi: 10.1016/j.chom.2021.05.008)
5. **Coelho et al., Nature, 2021** (doi: 10.1038/s41586-021-04233-4).
6. **Porubsky, Höps, Ashraf et al., Cell, 2022** (doi: 10.1016/j.cell.2022.04.017)
7. **Wuyts, Alves, Zimmermann-Kogadeeva, Nishijima et al. Mol. Syst. Biol., 2023** (doi: 10.15252/msb.202311525)

WP 3 Entwicklung und Validierung eines Multi-Omics Analyse-Frameworks

WP 3.1 Aufgrund der eingeschränkten Verfügbarkeit von longitudinalen Multi-Omics-Daten wurden schwerpunktmäßig Tools für die Analyse von Einzelzell- und Spatial-Omics Daten entwickelt. Resultate aus der AG Saez-Rodriguez waren eine end-to-end Multi-Omics-Analyse-Pipeline, COSMOS (Dugourd 2021), LIANA, ein umfassendes Software-Framework für Zell-Zell-Kommunikation (CCC) (Dimitrov 2022), und die Einrichtung einer umfassenden Datenbank für interzelluläre Signalübertragung in OmniPath (Türei 2021). LIANA wurde anschließend erweitert, um seine CCC-Inferenzmethoden auf Mikroben-Wirt-Interaktionen anwenden zu können. Eine Möglichkeit der Interaktion zwischen Mikroben und Wirt ist durch die Produktion bakterieller Metabolite. Auch menschliche Zellen produzieren verschiedene Metabolite für CCC. Um Vorwissen für alle Schritte der Inferenz mit entsprechenden Metaboliten bereitzustellen, haben wir eine umfassende Datenbankunterstützung in OmniPath entwickelt, einschließlich Ligand-Rezeptor-Interaktionen, metabolische Netzwerke und Metabolit-Struktur, Ontologie und funktionale Annotationsdatensätze. Letztere sind für die Vorverarbeitung von experimentellen Metabolomics-Daten, die funktionelle Charakterisierung und die Interpretation der Ergebnisse von wesentlicher Bedeutung (Dimitrov 2023). Das erweiterte LIANA+ Framework ist auch in der Lage, longitudinale oder multiregionale Daten mittels des statistischen MOFA-Ansatzes zu analysieren, um Zell-Zell-Kommunikation zu inferieren.

Ferner haben wir unserer Software mit der Softwarefamilie 'sc verse' eng verzahnt, indem wir deren Standarddatenstruktur verwenden, was Pipeline-Entwicklung und -Optimierung beschleunigt. Eine solche, flexibel anpassbare Pipeline, die Qualitätskontrolle, Probenintegration, Clustering, automatische oder manuelle Zelltyp-Annotation, funktionelle Analyse und Aktivitätsinferenz (TF, Pathway, Zytokin) umfasst, haben wir zur Verfügung gestellt (Arbeiten der AG Saez-Rodriguez, Heidelberg University).

WP 3.2 Auf der Mikrobiomseite lag ein Fokus des LAMarCK-Projekts auf realistischen Simulationen von krossektionalen und longitudinalen Mikrobiom-Daten als Basis für das Benchmarking von entsprechenden Analysemethoden (die Arbeiten zu krossektionalen Mikrobiomsimulationen und Benchmarking sind publiziert (Wirbel 2022, für longitudinale Daten ist ein Manuskript in Vorbereitung) (Arbeiten der AG Zeller, EMBL Heidelberg).

In Zusammenarbeit mit der Arbeitsgruppe von Oliver Stegle, EMBL/DKFZ, waren Mitglieder des LAMarCK-Konsortiums daran beteiligt, eine zuvor entwickelte Methode zur Multi-Omics Integration, MOFA, auf ihre Eignung für Mikrobiomdatenanalyse zu untersuchen. Insbesondere wurde evaluiert, wie gut sich MOFA und der Nachfolger MOFA+ für die integrative Analyse von Mikrobiomdaten zusammen mit Daten zu Ernährungsgewohnheiten, die molekularen Omics-Daten in ihrer Struktur gleichen, eignen. Dass diese Faktoranalysetools durchaus in der Lage sind, relevante Trends zu erkennen und interessante und plausible Bezüge zwischen Omics-Layers aufzuzeigen wurde anhand vergleichender Evaluationen mit anderen multivariaten Techniken (wie z.B. sparse canonical correlation analysis) klar (Arbeiten der AG Zeller, EMBL Heidelberg).

Außerdem wurde im Rahmen des LAMarCK-Projekts Software entwickelt um möglichst genaue taxonomische Mikrobiom-Profile aus Tumor-Genomik- und -Transkriptomik-Daten (die u.a. vom TCGA Konsortium erhoben wurden) zu extrahieren. Dazu wurde ein NextFlow Workflow (vknicht genannt) erstellt, der die Anwendung verschiedener existierender Tools für taxonomisches Profiling (PathSeq, Kraken2, MAPSeq und mOTUs3) ermöglicht und für die flexible Anwendung in verschiedenen HPC und Cloud Computing Environments validiert und optimiert. Eine erste Version wurde im Rahmen von NFDI4Microbiota publiziert. Eine erste Anwendung wurde ebenfalls publiziert (Sambruni 2023) (Arbeiten der AG Zeller und AG Bork, EMBL Heidelberg).

Schließlich haben wir die Transkriptionsfaktor-Regulone in OmniPath durch eine neue Ressource namens CollecTRI (Müller-Dott 2023) erweitert. Dieser neue Datensatz deckt wesentlich mehr Transkriptionsfaktoren ab und enthält vollständigere Regulone, was die funktionelle und mechanistische Analyse von Transkriptomikdaten verbessert. Außerdem wurde OmniPath weiterentwickelt, um die funktionelle Analyse von Metabolomics-Daten zu unterstützen (Arbeiten der AG Saez-Rodriguez, Heidelberg University).

WP 3.3 Im Rahmen von WP3.3 wurde ein Workflow zur Analyse von RNA-Seq-Datensätzen mit komplizierten experimentellen Designs entwickelt, der es erlaubt, entweder mehrere experimentelle Zustände oder Variablen zu vergleichen. Dieser Snakemake Workflow wurde auf GitHub veröffentlicht (<https://github.com/schlesnerlab/multicondition-deseq2-enrichment>) und ist per Conda oder Docker standardisiert einsetzbar. Im Rahmen dieses Workflows werden standardisiert eine Vielzahl von Downstream-Analysen durchgeführt, u.a. DoRoThEA und PROGENy mithilfe des decoupleR Frameworks (Badia-i-Mompel 2022). Dieser Workflow ermöglicht es insbesondere, auch für longitudinale Daten, standardisiert, Vorwissen für andere Anwendungen zur Verfügung zu stellen (Arbeiten der AG Schlesner, Augsburg University).

Ferner wurde BioCypher (Lobentanzer 2023) entwickelt, ein breit anwendbares, flexibles und hochleistungsfähiges Datenbankmanagement-Framework für biomedizinisches Vorwissen. Das BioCypher-Framework stützt sich auf biomedizinische Standard-Ontologien, so dass es praktisch alles im Projektkontext relevante Vorwissen aufnehmen kann. BioCypher basiert auf der leistungsstarken Graphen-Datenbanktechnologie Neo4j, die die Komplexität und Anzahl der Abfragen in den Analyse-Pipelines erheblich erweitert. BioCypher vereinfacht so nicht nur die Anbindung von Vorwissen an unsere Datenanalyse-Pipelines, sondern erleichtert auch die Kombination unabhängiger Datenbanken (Arbeiten der AG Saez-Rodriguez, Heidelberg University).

WP 3.4 Die in WP 3.2 vorgestellten neu entwickelten Tools für Mikrobiom-Profilierung anhand von Tumor-Omics-Daten ermöglichen die Integration dieser mikrobiellen Profile mit humanen Expressionsprofilen (aus denselben Daten). Dadurch werden Analysen der Mikroben-Wirts-Interaktion möglich, wie sonst nur anhand von Multi-Omics-Daten. Zu den Ergebnissen der integrativen Analysen des Wirtstranskriptoms mit Gewebe-residentem Mikrobiom in Darmtumoren ist ein Manuskript in Vorbereitung. Außerdem wurden diese Analysestrategien auf Pankreaskrebs- und Leberkrebs-Transkriptomdaten angewendet (in Kollaboration zwischen LAMarCK-Partnern und der Forschungsgruppe von Mathias Heikenwälder, DKFZ Heidelberg und M3 Institut, Uniklinikum Tübingen). In allen Anwendungsfällen wurden sowohl unüberwachte Faktor-Analysemethoden als auch univariate gemischte lineare Modelle (siehe WP 3.2) verwendet und vergleichend evaluiert (für die Resultate der integrativen Mikrobiom-Wirts-Analysen bei Leberkrebs befindet sich ein Manuskript unter Begutachtung) (Arbeiten der AG Zeller und AG Korbel, EMBL Heidelberg).

WP 3.5 Für eine künstlich zusammengestellte Bakteriengemeinschaft wurde unter kontrollierten Bedingungen und externen chemischen Stimuli ein longitudinaler Multi-Omics Datensatz generiert. Im einzelnen wurden 16S rRNA Amplikon-, Metagenomik-, Metatranskriptomik-, Metaproteomik- und Metabolomik-Daten erhoben. In einer integrativen Analyse wurden mehrere Multi-Omics-Analyse-Tools, wie z.B. MOFA (s.o.), mit verschiedenen Knowledge-basierten Ansätzen verglichen. Dieses Benchmarking ergab u.a., dass alle Omics-Techniken mit Spezies-Auflösung die relative Häufigkeiten der Bakterienspezies sehr konsistent schätzen. Die verschiedenen Omics-Techniken lieferten in diesem kontrollierten Experiment komplementäre Informationen über funktionelle Veränderungen der Bakteriengemeinschaft. Während zum Beispiel negative Spezies-Spezies-Interaktionen von fast allen Omics-Datentypen erfasst wurde, gaben Metatranskriptom-Daten Hinweise auf den molekularen Mechanismus und Metabolomik-Daten zeigten eine Veränderung der Zuckeraufnahme aus dem Medium durch die veränderte Bakteriengemeinschaft an (Wuyts 2023) (Arbeiten der AG Bork, EMBL Heidelberg).

Publikationen des LAMarCK-Konsortiums im Zusammenhang mit WP3 (teilweise existieren Überschneidungen mit den in anderen WPs genannten Publikationen):

1. Dugourd et al., *Mol. Syst. Biol.*, **2021** (10.15252/msb.20209730)
2. Dimitrov et al., *Nat. Commun.*, **2022** (doi: 10.1038/s41467-022-30755-0)
3. Türei et al., *Mol. Syst. Biol.*, **2021** (doi: 10.15252/msb.20209923)
- a. Wirbel, Essex, et al., *bioRxiv* **2022** (doi: 10.1101/2022.05.09.491139), bei **Genome Biol.** akzeptiert
4. Müller-Dott et al., *Nucl. Acids Res.*, **2023** (doi: 10.1093/nar/gkad841)
5. Badia-i-Mompel et al., *Bioinform Adv.*, **2022** (doi: 10.1093/bioadv/vbac016)

6. **Lobentanzer et al., Nat. Biotechnol., 2023** (doi: 10.1038/s41587-023-01848-y)
7. **Wuyts, Alves, Zimmermann-Kogadeeva, Nishijima et al. Mol. Syst. Biol., 2023** (doi: 10.15252/msb.202311525)
8. **Sambruni et al., Genome Med., 2023** (doi: 10.1186/s13073-023-01180-9)
9. **Dimitrov et al., bioRxiv, 2023** (doi: 10.1101/2023.08.19.553863) bei **Nat. Methods** akzeptiert

WP 4. Erweiterung des Analyse-Frameworks zur longitudinalen Datenanalyse

WP 4.1 Die Erweiterung des Mikrobiomdaten-Simulations-Frameworks für Zeitreihen (ausgehend von den Simulationen für krossektionale Daten siehe WP 3.2) ermöglichte die vergleichende Evaluation statistischer Analysemethoden für solche longitudinalen Omics-Daten. Das zeigte die hervorragende Eignung von linearen gemischten Modellen (LMMs) für die Analyse solcher Zeitreihen.

Diese Auswertungen von univariaten statistischen Methoden (um Assoziationen einzelner mikrobieller Taxa zu identifizieren) wurden durch Arbeiten an multivariablen Methoden ergänzt. In Zusammenarbeit mit der Arbeitsgruppe von Oliver Stegle, EMBL/DFKZ, wurde die Eignung des Faktoranalyse-Frameworks MOFA und seiner longitudinalen Erweiterung MEPHISTO auf Eignung für Mikrobiomdatenanalyse untersucht und publiziert (Velten 2022). Außerdem wurde anhand verschiedener Datensätze die Anwendung dieser Faktoranalysetools dafür optimiert, Mikrobiomdaten zusammen mit anderen Multi-Omics-Daten (z.B. Wirtstranskriptom) zu analysieren, um plausible Bezüge zwischen diesen Omics-Layers aufzuzeigen. Dabei zeigte sich, dass diese Faktormodelle wichtige Trends über mehrere Omics-Layers hinweg identifizieren können – auch als Vorverarbeitungsschritt für überwachte Klassifikationsverfahren. Aufgrund dieser Ergebnisse erübrigte sich die Entwicklung überwachter Verfahren für longitudinale Daten (wie den Group-LASSO-Algorithmus), wie das zur Antragsstellung noch geplant war (Arbeiten der AG Zeller und AG Bork, EMBL Heidelberg).

Schließlich wurde das statistische MOFA-Multi-Omics-Framework für die Analyse der Signalübertragung zwischen mehreren Zelltypen und Proben mit LIANA erweitert. Bei diesem Verfahren wird MOFA anstelle von Omics-Modalitäten auf die Kommunikation zwischen Paaren von Zelltypen angewendet, die aus Einzelzell-Transkriptomikdaten abgeleitet werden. Dieser Ansatz ermöglicht die Analyse der Zell-Zell-Kommunikation über mehrere Bedingungen und Zelltypen hinweg. Um LIANA mit MOFA zu verbinden, haben wir erstere in die Muon-Multiomics-Datenstruktur integriert, die Teil des Scverse Single-Cell Computational Framework ist (Arbeiten der AG Saez-Rodriguez, Heidelberg University).

WP 4.2 Auch auf Wirtsseite wurde die Nutzung von Linearen gemischten Modelle (LMMs) zur Analyse anhand eines großen RNAseq Datensatzes mit alternden Mäusen evaluiert. Das ergab, dass sich LMMs besonders für die Analyse von longitudinalen Daten eignen, da sie bei wiederholten Messungen, wie sie bei longitudinalen Daten vorkommen, die zufälligen Effekte pro Individuum modellieren können (Arbeiten der AG Schlesner, Augsburg University).

Ferner wurde ein mechanistisches Multi-omics-Framework COSMOS (Dugourd 2021) entwickelt, eine Methode, die kausale Schlussfolgerungen über Transkriptomik, Phosphoproteomik und Metabolomik hinweg durchführt. Die parallele Anwendung von COSMOS und MOFA lieferte ergänzende Erkenntnisse, so dass wir beide Methoden in unseren Analyse-Workflows verwenden werden. Wir haben auch eine Integration zwischen LIANA und AnnData entwickelt, die eine einfache Inferenz der Zell-Zell-Kommunikation in scanpy Einzelzell-Pipelines ermöglicht (Arbeiten der AG Saez-Rodriguez, Heidelberg University).

WP 4.3 Um Tools für die statistische Auswertung von Zeitreihen des Mikrobioms weiter in der Praxis zu erproben untersuchten wir mehrere Datensätze, die zu Veränderungen in der Zusammensetzung des Darmmikrobiom unter Ernährungsinterventionen erhoben worden waren. Auch diese Studien bestätigten deutlich die Eignung von linearen gemischten Modellen (LMMs) (siehe auch WP 4.1). Die Ergebnisse einer solchen Studie befinden sich derzeit in Begutachtung (Ducarmon 2024). Die technischen Herausforderungen bei der Analyse longitudinaler Mikrobiomdaten in Zusammenhang mit Ernährungsänderungen und unter kalorienreduzierten Diäten wurden (neben anderen Aspekten des Fastens auf den

menschlichen Organismus und das Darmmikrobiom) außerdem in einem Übersichtsartikel publiziert (Ducarmon 2023) (Arbeiten der AG Zeller, EMBL Heidelberg).

Publikationen des LAMarCK-Konsortiums im Zusammenhang mit WP4 (teilweise existieren Überschneidungen mit den in anderen WPs genannten Publikationen):

1. **Velten et al., Nat. Methods 2022** (doi: 10.1038/s41592-021-01343-9)
2. **Dugourd et al., Mol. Syst. Biol. 2021** (doi: 10.15252/msb.20209730)
3. **Ducarmon et al., Trends Microbiol., 2023** (doi: 10.1016/j.tim.2023.02.009)
4. **Ducarmon et al., bioRxiv, 2024** (doi: 10.1101/2024.04.19.590209)

WP5. Anwendung des Analyse-Frameworks zur integrierten Analyse von Mikrobiom und Wirtsorganismus

WP 5.1 Anwendungen der oben beschriebenen mikrobiellen Profiling-Strategie (s.o. WP 3.2) auf Omics-Datensätze verschiedener Tumor-Gewebe war auf Tumoren des Darms, der Leber und der Bauchspeicheldrüse fokussiert. Unsere Validierungsexperimente zeigten deutlich, dass die mikrobiellen Profile aus Whole-Exome-, Whole-Genome- und RNA-Seq-Daten Muster zeigen, die sehr genau mit denen aus metagenomischen Sequenzierungsdaten (basierend auf 16S Amplikonsequenzierung von Gewebeproben oder Metagenomsequenzierung von Stuhlproben) übereinstimmen. Darauf aufbauend wurde eine integrative Analyse der Whole-Exome- und RNA-Seq-Daten aller oben genannten Tumorentitäten durchgeführt. Unter Verwendung von univariaten LMMs und multivariablen Faktormodellen ließen sich interessante Bezüge zwischen intratumoralen Bakterien und dem Tumortranskriptom herstellen, wobei sich insbesondere bei Darmkrebs klare Assoziationen zwischen inflammatorischen Signalen, Mutationslast und verstärkter mikrobieller Präsenz im Tumor zeigten (Manuskript in Vorbereitung). Bei Leberkrebs zeigten sich Unterschiede im Tumormikrobiom in Assoziation mit viraler Tumor-Aetiologie (Manuskript in Vorbereitung). Diese Analysen stützen sich auf eine Vielzahl der oben genannten, im Rahmen des LAMarCK-Projekts entwickelten oder validierten Tools (DoRothea, COSMOS, LIANA, mOTUs, LMMs, MOFA). wurden diese Analysen außerdem auf räumlich aufgelöste Multi-Omics-Daten ausgeweitet (kollaborative Arbeiten der AGs Zeller, Korbelt und Saez-Rodriguez, EMBL und Heidelberg University).

WP 5.2 Im Falle von Darmkrebs wurden diese Analysen auch auf räumlich aufgelöste Multi-Omics-Analyse von Tumorgewebeproben ausgeweitet. Die Daten dafür stammen größtenteils aus unseren Laboren am EMBL (finanziert durch interne und andere Drittmittel). Dabei wurden bioinformatische Workflows zur Analyse von Multiplex-FISH Daten etabliert, um die Identifikation und Lokalisation von intratumoralen Bakterien möglich zu machen. Diese Workflows umfassen die Registrierung verschiedener Fluoreszenz-Mikroskopie-Bilder, deren Alignierung mit anderen Bilddaten (von anderen Spatial-Omics-Modalitäten, wie z.B. Spatial Transcriptomics), und die automatisierte Erkennung von Bakterien und deren genaue taxonomische Bestimmung. Auf der Wirtsseite zielten unsere Analyse-Workflows darauf ab, die Zelltyp-Verteilung und räumlich aufgelöste Genexpression im Tumor abzubilden (dabei wurden Einzelzell- und räumliche Transkriptomdaten, die mittels 10X Genomics' Spatial Transcriptomics-Technologien Visium und Xenium generiert wurden, integriert). Diese integrativen Analysen zeigten hochinteressante Kollokalisationsmuster zwischen spezifischen intratumoralen Bakterien, Tumor- und Immunzellen und deren Genexpression. Ein Manuskript, das diese Ergebnisse beschreibt, befindet sich derzeit in Vorbereitung kollaborative Arbeiten der AG Korbelt und AG Zeller, EMBL Heidelberg).

Publikationen des LAMarCK-Konsortiums im Zusammenhang mit WP5 befinden sich derzeit noch in Vorbereitung

2. Wichtigsten Positionen des zahlenmäßigen Nachweises

EMBL

Die Verwendungsnachweise wurden separat vorgelegt. Es wurden Personal-, Reisemittel und Verbrauchskosten verausgabt. Insgesamt wurden weniger Mittel als geplant verausgabt (**538,327€ vs. 544,299€**).

Personalmittel. Durch das LAMarCK Projekt wurden Personalmittel zur Verfügung gestellt, die vor allem für bioinformatischen Analysen durch ein Team aus Wissenschaftlern (Postdoktorand:innen), und Doktorand:innen am EMBL genutzt wurden, die zu allen 5 WPs beigetragen haben, insgesamt **529,284€**.

Reisemittel: Aufgrund der Corona-Pandemie konnte ein Großteil der geplanten Reisen nicht angetreten werden. Die Mittel wurden im Projektverlauf teilweise in Personalmittel umgewidmet. Es wurden **2,573 €** für Reisen verausgabt.

Gegenstände über 410 / 400€. Mit diesen wurde ein Laptop für die alleinige Nutzung für das Projekt angeschafft (**2,418.49€**).

Verbrauchsmittel. Insgesamt wurden **1,040€** für Verbrauchsmittel verausgabt.

3. Notwendigkeit und Angemessenheit der geleisteten Projektarbeiten

Die Entwicklung von Analysestrategien und Softwaretools für Multi-Omics Integration ist ein aktives Forschungsfeld. Insbesondere für die Analyse von Zeitreihen und räumlich aufgelösten Omics-Daten existierten zum Zeitpunkt der Antragstellung kaum bioinformatische Werkzeuge. Zusätzlich zur Entwicklung solcher bioinformatischer Tools bestand auch ein Bedarf an vergleichenden Evaluationen dieser Analyse-Werkzeuge, um der Forschungsgemeinschaft Anhaltspunkte und Empfehlungen geben zu können zu deren Eignung, Stärken und Schwächen. Ferner berücksichtigen die wenigsten dieser Analyse-Tools Vorwissen in zufriedenstellender Weise, was es erheblich erschwerte, mit den Ergebnissen dieser Tools auch neue biologische Erkenntnisse zu gewinnen. Daraus begründete sich die klare Notwendigkeit für das LAMarCK-Konsortium, sich an der Entwicklung und Validierung solcher Tools zu beteiligen. Im Laufe des Vorhabens hat das LAMarCK-Konsortium Datenbanken und Softwaretools etabliert, um die Analyse verschiedener Arten von Omics-Daten mit modernsten Methoden zu unterstützen. Die darauf aufbauende Entwicklung sowie die Evaluierung von Analyse-Workflows (die sich aus einzelnen solcher Tools zusammensetzen) an verschiedenen Datensätzen waren notwendig, um das Potential dieser Methoden zur Gewinnung neuer Erkenntnisse von translationaler Relevanz, zum Beispiel anhand von konkreten Darmkrebs-Multi-Omics-Daten, zu demonstrieren. Diese Aktivitäten waren alle angemessen und notwendig, um das letztendliche Ziel des Projekts zu erreichen, d. h. die Bereitstellung von Werkzeugen für die Multi-Omics-Analyse von zeitlich und räumlich aufgelösten Daten, soweit möglich auch unter Einbeziehung von Einzelzelldaten.

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

Das wissenschaftliche Ergebnis von Omics-Projekten hängt weitgehend von der Fähigkeit der Methoden ab, relevante biologische Zusammenhänge zu erfassen. Die im Rahmen des LAMarCK-Projekts entwickelten Werkzeuge und Pipelines ermöglichen die Gewinnung biologischer Erkenntnisse aus zeitlich und räumlich aufgelösten Multi-Omics-Daten, die derzeit viele biomedizinischen Forschungsprojekte prägen. Wir haben modernste Berechnungs- und Softwareentwicklungstechnologien eingesetzt, um Werkzeuge bereitzustellen, die gut konzipiert, flexibel, integriert und leistungsstark sind. Der Nutzen dieser Arbeit zeigt sich bereits in der Nutzung unserer Werkzeuge in der Forscher Community und in den Ergebnissen, die sie sowohl in unseren als auch in anderen Projekten liefern. Die gute Qualität und das Design dieser Software erleichtern die langfristige Wartbarkeit und Erweiterbarkeit, so dass sich die Methoden immer leicht an die sich ständig ändernden Anforderungen anpassen lassen. Insgesamt haben wir eine solide Infrastruktur für die wissenschaftliche Multi-Omics-Analyse geschaffen, einschließlich der Analyse der Zell-Zell-Kommunikation und der Mikroben-Wirt-Interaktion, die sich in vielen

Anwendungen als nützlich erwiesen hat und deren Bedeutung in der Zukunft voraussichtlich noch wachsen wird. Wir erwarten, dass diese Werkzeuge mindestens in den nächsten 5 Jahren von der akademischen Gemeinschaft intensiv genutzt werden. Eine Implementierung im pharmazeutischen Kontext im präklinischen Bereich ist ebenfalls in einem ähnlichen Zeitrahmen zu erwarten. Ein Einsatz in einem eher klinischen industriellen Kontext wird vermutlich innerhalb von 10 Jahren erfolgen, wenn die hier analysierten Omics-Technologien in diesen Umgebungen Einzug halten.

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

Nicht bekannt.

6. Erfolgten oder geplanten Veröffentlichungen des Ergebnisses

Veröffentlichungen in Peer-Review Fachzeitschriften:

1. Liu et al., *NPJ Sys. Biol. Appl.*, 2019 (doi: 10.1038/s41540-019-0118-z)
2. Bouhaddou, Memon, Meyer, White et al., *Cell*, 2020 (doi: 10.1016/j.cell.2020.06.034)
3. Ostaszewski, Niarakis, Mazein, Kuperstein, Phair et al., *Mol. Syst. Biol.*, 2021 (doi: 10.15252/msb.202110387)
4. Türei et al., *Mol. Syst. Biol.*, 2021 (doi: 10.15252/msb.20209923)
5. Letunic and Bork, *Nucl. Acids Res.*, 2021 (doi: 10.1093/nar/gkab301)
6. Carroll, Larralde, Fleck et al., *bioRxiv*, 2021 (doi: 10.1101/2021.05.03.442509)
7. Orakov, Fulham et al., *Genome Biol.*, 2021 (doi: 10.1186/s13059-021-02393-0)
8. Cantalapiedra et al., *Mol. Biol. Evol.*, 2021 (doi: 10.1093/molbev/msab293)
9. Dugourd et al., *Mol. Syst. Biol.*, 2021 (10.15252/msb.20209730)
10. Poletti et al., *npj Syst. Biol. Appl.*, 2021 (doi: 10.1038/s41540-022-00224-x)
11. Hildebrand et al., *Cell Host Microbe*, 2021 (doi: 10.1016/j.chom.2021.05.008)
12. Coelho et al., *Nature*, 2021 (doi: 10.1038/s41586-021-04233-4).
13. van Rossum et al., *Bioinform.*, 2022 (doi: 10.1093/bioinformatics/btab789)
14. Khedkar et al., *Nucl. Acids Res.*, 2022 (doi: 10.1093/nar/gkac163)
15. Ruscheweyh, Milanese et al., *Microbiome*, 2022 (doi: 10.1186/s40168-022-01410-z)
16. Fullam et al., *Nucl. Acids Res.*, 2022 (doi: 10.1093/nar/gkac1078)
17. Dimitrov et al., *Nat. Commun.*, 2022 (doi: 10.1038/s41467-022-30755-0)
18. Badia-i-Mompel et al., *Bioinform. Adv.*, 2022 (doi: 10.1093/bioadv/vbac016)
19. Porubsky, Höps, Ashraf et al., *Cell*, 2022 (doi: 10.1016/j.cell.2022.04.017)
20. Velten et al., *Nat. Methods* 2022 (doi: 10.1038/s41592-021-01343-9)
21. Wirbel, Essex, et al., *bioRxiv* 2022 (doi: 10.1101/2022.05.09.491139), bei *Genome Biol.* akzeptiert
22. Larralde and Zeller, *Bioinformatics*, 2023 (doi: 10.1093/bioinformatics/btad214)
23. Müller-Dott et al., *Nucl. Acids Res.*, 2023 (doi: 10.1093/nar/gkad841)
24. Lobentanzer et al., *Nat. Biotechnol.*, 2023 (doi: 10.1038/s41587-023-01848-y)
25. Wuyts, Alves, Zimmermann-Kogadeeva, Nishijima et al. *Mol. Syst. Biol.*, 2023 (doi: 10.15252/msb.202311525)
26. Ducarmon et al., *Trends Microbiol*, 2023 (doi: 10.1016/j.tim.2023.02.009)
27. Sambruni et al., *Genome Med.*, 2023 (doi: 10.1186/s13073-023-01180-9)
28. Dimitrov et al., *bioRxiv*, 2023 (doi: 10.1101/2023.08.19.553863) bei *Nat. Methods* akzeptiert
29. Ducarmon, Karcher et al., *bioRxiv*, 2024 (doi: 10.1101/2024.01.08.574624)
30. Ducarmon et al., *bioRxiv*, 2024 (doi: 10.1101/2024.04.19.590209)