

## Verwendungsnachweis SAGE

### Teil I: Kurzbericht

**Vorhaben:** Zellsegmentierungsfreie Identifizierung von mRNA-Lokalisationsmustern in Tumoren

**Förderkennzeichen:** 031L0265, **Zuwendung auf Ausgabenbasis**

### AKRONYM: SAGE

#### Aufgabenstellung und Ziele des Projektes:

Die transkriptionelle Heterogenität von Krebszell-Ökosystemen stellt eine therapeutische Herausforderung dar, da sie das Wachstum, die Immunflucht, die Metastasierung und die Reaktion auf Medikamente beeinflusst. Identische Zelltypen können gewebespezifische Expressionsmuster aufweisen. Die Aufklärung der Verbindung zwischen räumlicher Organisation und Expressionsprogrammen von Zellen ist besonders wichtig für das Verständnis der zellulären Heterogenität und der Mikroumgebung in soliden Tumoren.

Ziel des Projektes war es deshalb, eine effiziente, anwenderfreundliche und übergreifende Analyseumgebung für in-situ-Transkriptomik (IST)-Daten zu erstellen, welche eine zellsegmentierungsfreie Identifizierung von mRNA-Lokalisationsmustern in Tumoren ermöglicht.

Typischerweise beinhaltet die Analyse von IST-Daten einen Schritt, bei dem alle mRNA-Moleküle je einer der in der Gewebeprobe befindlichen Zellen zugewiesen werden. Dies ist jedoch problematisch, da gute Zellsegmentationen oft schwer zu erstellen sind. Deshalb besteht unmittelbar Bedarf an einer Methode, die ohne einen solchen Segmentationschritt auskommt.

Im Verlauf des Projektes wurde eine Reihe neuer Software-Instrumente geschaffen, die die oben beschriebene Fragestellung adressiert und deren erfolgreiche Anwendung unter Beweis stellt.

#### Durchführung des Projektes:

Die Durchführung des Projektes erfolgte im ersten Jahr strikt gemäß den Vorgaben des initialen Projektplans. Zu Beginn des Projektes wurden ein Doktorand und eine studentische Hilfskraft eingestellt. Zudem wurde aus Projektmitteln ein Laptop angeschafft. Das Projekt startete planmäßig und alle vorgesehenen Meilensteine konnten im ersten Jahr erreicht werden.

Im zweiten Jahr wurde der Projektplan aufgrund des Feedbacks aus der Omics-Community angepasst, um dem dringenden Bedarf an benutzerfreundlichen Webanwendungen für die Verarbeitung von räumlichen Transkriptomikdaten gerecht zu werden. Daher lag der Fokus im zweiten Jahr auf der Durchführung des Arbeitspakets 4 (Work Package WP4), das darauf abzielte, das Tool einer breiteren Öffentlichkeit ohne Zugang zu Hochleistungsrechnern zur Verfügung zu stellen. Der ursprüngliche Zeitplan wurde dahingehend geändert, dass die Aufgaben aus WP 4.1 (KDE-Algorithmus optimieren) und WP 4.3 (Online-App erstellen) in das zweite Jahr vorgezogen wurden. Der Algorithmus wurde in das Projekt „SSAM-lite“ integriert, sodass eine Analyse von SRT-Daten auf lokalen Rechnern ermöglicht wurde.

Das zweite Jahr war deshalb von der Durchführung des WP4 geprägt, welches es sich zur Aufgabe gemacht hatte, das Tool einer breiteren Öffentlichkeit ohne Zugang zu Hochleistungsrechnern zur Verfügung zu stellen. Der initiale Zeitplan wurde dahingehend verändert, dass zunächst die Aufgaben aus WP 4.1 (KDE-Algorithmus optimieren) und WP 4.3 (Online-App erstellen) in das zweite Jahr vorgezogen wurden. Der Algorithmus wurde in das Projekt „SSAM-lite“ integriert so dass eine Analyse von SRT-Daten von lokalen Rechnern ermöglicht wurde. Neben WP4 wurde auch WP3 schwerpunktmäßig bearbeitet. In diesem Arbeitspaket werden räumliche Zusammenhänge zwischen Zelltypen im Gewebekontext durch einen eigens dafür entwickelten Algorithmus erstellt. Die Bearbeitung dieses Arbeitspakets erstreckte sich über den größten Teil des dritten Berichtszeitraumes. Der Algorithmus ist nun sowohl in unserem SSAM-Tool als auch im SONAR-Tool verfügbar. Für die Berichtszeiträume des zweiten und dritten Jahres konnten die Meilensteine gemäß dem angepassten Projektplan planmäßig erreicht werden. Im Berichtszeitraum des vierten Jahres lag der Focus auf der Finalisierung des gesamten Projektes einschließlich des Abschlusses der Work Packages 2 und 5.

Dieser Berichtszeitraum war besonders produktiv, weil er die Abschlussphase des Projektes markierte und damit auch die meisten Veröffentlichungen hervorbrachte. Zusätzlich wurden Veranstaltungen wie ein Hackathon und eine Schulung zum Thema „Datenanalyse von räumlichen Omics-Daten“ organisiert und durchgeführt.

#### Ergebnisse und Verwertung des Projektes:

Die in diesem Projekt generierten SoftwareTools wurden auf gesunde, krankheits- und krebsbezogene Gewebedatensätze von Konsortien, Industriepartnern und lokalen Kollaborationspartnern angewendet. Dabei wurden bestehende und neu entstehende Speicherstrukturen und Datenstandards unterstützt, die Berechnungsanforderungen verringert sowie eine webbasierte App für unser Tool entwickelt. Weitere IST-Analyseplattformen wurden in unseren Algorithmus integriert. Veröffentlicht wurden der Quellcode, eine Cloud fähige Webanwendung und Publikationen, die unsere Werkzeuge und Ergebnisse beschreiben. Neben der Durchführung und Organisation von zwei Hackathons wurde ein zweitägiger Schulungskurs zur Nutzung unserer Software abgehalten. Alle diese Produkte werden aufgrund ihrer Veröffentlichung entsprechend den FAIR-Prinzipien bereits von der wissenschaftlichen Gemeinschaft genutzt.

Das Projekt hat eine Vielzahl von verwertbaren Ergebnissen in verschiedenster Form geliefert und ist insofern außerordentlich erfolgreich bezüglich der Menge und Qualität der erzeugten Ergebnisse. Es gibt insgesamt 10 wissenschaftliche Publikationen, die aufgrund der SAGE-Förderung ermöglicht wurden, zumeist in Fachzeitschriften mit hohem Impact-Faktor wie Nature Methods (Impact-Faktor: 36.1) <https://doi.org/10.1038/s41592-025-02617-2>.

Zusätzlich gibt es eine Vielzahl von Veröffentlichungen verschiedener Web-Tools und entsprechender Repositorien, deren Auflistung im Sachbericht im Detail vorliegen. Die erzeugten Webtools sind entsprechend den FAIR - Vorgaben öffentlich verfügbar und werden bereits von einer Vielzahl Forschender angewendet. Die Arbeit an web-basierten Tools wie den oben genannten wird insbesondere vorangetrieben durch die Zusammenarbeit während Hackathons, weil hier Experten der wissenschaftlichen Gemeinschaft auf internationaler Ebene im konzentrierten Austausch für einen begrenzten Zeitraum zusammenarbeiten. Im Projektzeitraum fanden 3 Hackathons statt, von denen 2 von Naveed Ishaque organisiert und durchgeführt wurden. Zusätzlich veranstaltete Naveed Ishaque einen Kursus zur Datenanalyse von räumlichen Omics Daten. Diese Hackathons und Kurse stellen eine großartige Möglichkeit dar, weitere Kontakte zu knüpfen und neue Kooperationen anzuregen, ähnlich wie die Präsentation von Ergebnissen auf Fachkonferenzen. Anwendung der erzeugten Ergebnisse erfolgt bereits im Open-Source-Vertrieb nach FAIR-Prinzipien, die Integration in große Forschungskonsortien und die Bereitstellung von Schulungen und Beratung wird bereits genutzt durch z. B. das Angebot hochwertiger cloudbasierter Analysedienste oder die Integration des Rahmens in diagnostische und pharmazeutische Pipelines besteht ein Potential der Kommerzialisierung

Naveed Ishaque hat die erzielten Ergebnisse in verschiedenen Stadien national und international auf einer Reihe von Fachtagungen und Konferenzen oder auch auf Veranlassung interessierter Arbeitsgruppen vorgestellt, so dass es insgesamt mehr als 20 Präsentationen gab, zu denen Ergebnisse dieses Projektes innerhalb der Forschungsgemeinschaft geteilt wurden. Zudem ist er seit 2023 Mitglied des zentralen Ausschusses der Global Alliance for Spatial Technologies (GESTALT), die den Zugang zu Spitzeninformationen über räumliche Omics-Technologien demokratisieren will.

Abschliessend lässt sich sagen, dass eine effiziente, anwenderfreundliche und übergreifende State-of-the-Art Analyseumgebung für in-situ-Transkriptomik (IST)-Daten erstellt wurde, die eine zellsegmentierungsfreie Identifizierung von mRNA-Lokalisationsmustern in Tumoren erlaubt und welche bereits erfolgreich von weiten Teilen der wissenschaftlichen Gemeinschaft angewendet wird.

**Teil II: Eingehende Darstellung**

**Vorhaben:** Zellsegmentierungsfreie Identifizierung von mRNA-Lokalisationsmustern in Tumoren

**Förderkennzeichen:** 031L0265, **Zuwendung auf Ausgabenbasis**

**AKRONYM:** SAGE (Spatial Analysis of Gene Expression)

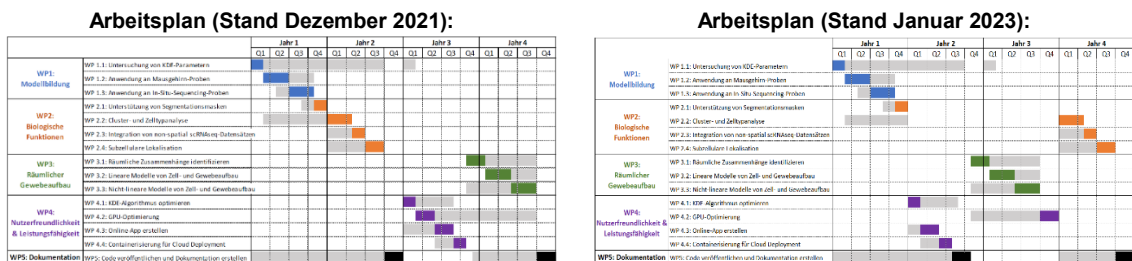
**Abgabefrist:** 31.07.2025

**Ausführliche Darstellung durchgeführter Arbeiten und Vergleich zur ursprünglichen Vorhabensbeschreibung.**

Das Projekt hat zu einer unerwarteten Fülle von Ergebnissen und Erkenntnissen geführt, die die räumliche Analyse von Omics Daten deutlich verbessert, präzisiert und vertieft haben. Wie im weiteren Verlauf dargestellt, gibt es aus diesem Grund zahlreiche Veröffentlichungen, zum Beispiel eine Reihe wissenschaftlicher Publikationen, aber auch zahlreiche Repositorien, die der Scientific Community zur Verfügung gestellt werden konnten.

Im Grossen und Ganzen wurde das Projekt innerhalb der Laufzeit entsprechend der Planung durchgeführt. Allerdings gab es eine wesentliche Änderung im Projektplan, die die zeitliche Abarbeitung der Work Packages (WP, Arbeitspakete) betrifft. Ursprünglich sollte das Work Package 2, welches sich mit der Darstellung der biologischen Funktionen innerhalb einer Zelle beschäftigt, chronologisch nach dem Abschluss von Work Package 1 durchgeführt werden.

Aufgrund des Feedbacks aus der Scientific Community wurde aber deutlich, dass es grossen Bedarf gab, das Tool benutzerfreundlich und für eine breite Öffentlichkeit verfügbar zu machen, d.h. eben auch für Forscher ohne Zugang zu Hochleistungsrechnern und Programmierkenntnissen. Aus diesem Grund wurden zunächst die WPs 3 und 4 durchgeführt, damit das Tool reichweitenstark auf kleineren Rechnern zur Verfügung gestellt werden konnte, bevor WP 2 bearbeitet wurde. So wurden die beiden WPs 3 und 4 in den zweiten bzw. dritten Förderzeitraum vorgezogen.



**Abbildung 1:** Vergleich ursprünglicher Arbeitsplan 2021 (oben) zu Arbeitsplan 2023 (unten): WP2 ist verschoben nach 2024, WP 4 und WP3 sind vorgezogen in die Berichtszeiträume des 2. bzw. des 3. Jahres.

Die Bearbeitung der Work Packages wurde entsprechend dem Arbeitsplan vom Januar 2023 vorgenommen. Die folgende detaillierte Darstellung der Work Packages ist nicht streng chronologisch, sondern dem angepassten Projektplan entsprechend gegliedert mit Ausnahme des Work Package 5, welches am Ende dieses Textes beschrieben ist.

**WP 1: Modellbildung**

Unser Work Package 1 zielte darauf ab, das KDE-Modell so zu verbessern, dass es neuartige Methoden und Gewebe unterstützt, indem es gewebe- und technologiespezifische Verzerrungen identifiziert und beseitigt. Seit der Einreichung unseres Vorschlags haben wir unseren Rahmen für die Analyse von IST-Daten in der renommierten Zeitschrift Nature Communications (<https://doi.org/10.1038/s41467-021-23807-4>) veröffentlicht und verbessert und den Quellcode auf <https://github.com/HiDiHlabs/ssam> offengelegt.

### WP 1.1: Untersuchung von Kernel-Density-Estimation-Parametern

Work Package 1.1 zielte darauf ab, einen Rahmen für die Untersuchung von Kernel-Density-Estimation (KDE) Parametern zu implementieren. Dieser Rahmen ist in Python implementiert und ermöglicht die Anpassung einer großen Anzahl von KDE-Parametern: Kernel, Bandbreite, Schwellenwerte für die Expression und Filterung von Vektoren für das Clustering.

Wir haben festgestellt, dass ein sehr wichtiger Aspekt der Qualitätssicherung darin besteht, umliegende mRNA-Moleküle zu entfernen, die nicht aus Zellen stammen und zu falschen lokalen Maxima führen können. Wir implementierten erfolgreich eine K-Nächste-Nachbarn-Klassifikation zur Identifizierung ungewollter lokaler Maxima, welche sowohl global als auch lokal beobachtbar waren. Darüber hinaus untersuchten wir auch die Verwendung verschiedener Kernelformen im bestehenden Rahmen. In unserem Vorschlag geben wir an, dass wir in unserer Proof-of-Principle-Arbeit einen Gauß-Kernel verwendet haben. Es wäre jedoch interessant, Gauß-, Epanechnikov-, Quart-, Kosinus- und Halbkreis-Kerne zu untersuchen. In unserer ersten Veröffentlichung verwenden wir einen Gaußschen KDE, um die Interpretation der Ergebnisse zu vereinfachen. In einer späteren Studie über die menschliche Bauchspeicheldrüse (Tosti et al. 2021, <https://doi.org/10.1053/j.gastro.2020.11.010>) verwendeten wir jedoch einen halbkreisförmigen Kernel. Diese beiden Kernelformen (Gauß und Halbkreis) stellen die beiden Extreme der Kernelformen dar.

Unser anfängliches Interesse an der Untersuchung von Kernelformen war darauf zurückzuführen, dass der Epanechnikov-Kernel ein „theoretisch optimaler“ Kernel ist, da er den asymptotischen mittleren integrierten quadratischen Fehler (AMISE) minimiert, der eine Leistungskennzahl für die Effizienz eines Kernels ist. In Kombination mit einer variablen Bandbreite und einer minimalen Ausprägungsschwelle können diese Kernelformen jedoch einander angenähert werden. Da auch experimentelle Methoden sowohl systematische als auch stochastische Fehler bei der Datengenerierung aufweisen, gehen wir davon aus, dass die Form des Kernels im Vergleich zu den Bandbreiten- und Schwellenwertparametern eine weitaus geringere Rolle spielt.

Wir konnten feststellen, dass IST-Assays auf der Grundlage der In-situ-Hybridisierung (ISH) eine hohe Empfindlichkeit aufweisen (mehr als 10 Moleküle pro Zelle), so dass der Bereich der optimalen Bandbreite von 2,5, der in unserer Studie an drei Bereichen des Mäusehirns unter Verwendung von osmFISH, MERFISH und smFISH verwendet wurde, zu hervorragenden Ergebnissen führte.

Vor diesem Hintergrund haben wir unseren KDE-Modellierungsansatz an dem von MERFISH (Moffitt et al. 2018, <https://doi.org/10.1126/science.aau5324>) erstellten Datensatz des präoptischen Bereichs der Maus validiert. Dieser Datensatz ist außergewöhnlich, da er mehr als 100 räumlich aufgelöste Zelltypen identifiziert. Unser Ansatz rekapitulierte nicht nur die breiten Zellklassen (inhibitorische, exzitatorische, Oligodendrozyten, Astrozyten, Mikroglia, Ependymale, Endothelzellen und Muralzellen), sondern identifizierte auch die Mehrheit der einzelnen Zelltypen und -zustände. Wir waren sehr erfreut zu sehen, dass einige neuronale Zelltypen unterschiedliche räumliche Lokalisierungsmuster aufwiesen, die in der ursprünglichen Arbeit von Moffitt et al. identifiziert wurden. Diese Ergebnisse wurden in Park et al. 2021 veröffentlicht <https://doi.org/10.1038/s41467-021-23807-4>.

### WP 1.2: Anwendung an Mausgehirn-Proben

Bei der Anwendung dieses Frameworks stellte sich heraus, dass die Verwendung des Gauß-Kerns für eine einfache Interpretation optimal ist. Nach unserer ausführlichen Validierungsarbeit haben wir dieses Framework angewandt, um die Zelltypenorganisation im somatosensorischen Kortex der Maus mit osmFISH, im präoptischen Bereich der Maus mit MERFISH und im visuellen Kortex der Maus mit smFISH erfolgreich zu rekapitulieren. In allen Fällen können wir Zelltypen sehr detailliert abbilden und feinkörnige Gewebedomänen/Nischen identifizieren. Gewebedomänen-Signaturen wurden durch Clustering der lokalen Zelltyp-Zusammensetzung über gleitende kreisförmige 100-µm-Fenster identifiziert und auf die Zelltyp-Karte zurückprojiziert. Ein besonderer Erfolg war die Lokalisierung eines einzigen Zelltyps, Sst Chodl, im Gewebe, was die Empfindlichkeit unseres Ansatzes unterstreicht. Dieses Ergebnis wurde in Park et al. 2021 veröffentlicht. <https://doi.org/10.1038/s41467-021-23807-4>.

Darüber hinaus wurde das SSAM-Framework auf das Space-TX-Konsortium angewandt, wo wir eine sehr hohe Leistung über eine Reihe von IST-Technologien für das Mausgehirn nachweisen konnten. Unser segmentierungsfreier Ansatz ist sehr gut mit segmentierungsbasierten Ansätzen vergleichbar. Im Allgemeinen zeigt sich, dass unser SSAM-Tool empfindlicher ist als andere Methoden. Im Fall der BaristaSeq-Daten, die GMCS-Signaturen verwenden, schneidet unser Tool bemerkenswert gut ab, wenn man bedenkt, dass die Leistung bei segmentierungsbasierten Ansätzen sehr schlecht ist. Dieses Ergebnis wurde in Zhang et al. 2022 veröffentlicht, <https://doi.org/10.1101/2022.03.28.486139>.

### **WP 1.3: Anwendung an In-Situ-Sequencing-Proben**

Im Vergleich zu ISH-Technologien ist die In-Situ-Sequenzierung (ISS) ein viel weniger empfindlicher Assay, so dass wir durch den Einsatz von ISS viel weniger Moleküle pro Zelle detektieren konnten. Die generierten Daten waren wesentlich schwieriger zu analysieren. Diese Technologie ist jedoch von großer Bedeutung, da sie die Grundlage für ein neues Produkt bildet, das noch in diesem Jahr als „Xenium“-Plattform von 10x Genomics auf den Markt kommen soll: <https://www.10xgenomics.com/in-situ-technology>.

Wir haben 2 ISS-Datensätze analysiert: die CA1-Region des Mäusegehirns (Qian et al. 2020) und ein menschliches Pankreasläppchen (Tosti et al. 2021). Für die CA1-Region musste die Bandbreite auf 3 erhöht werden, aber aufgrund erheblicher Unstimmigkeiten bei der Lokalisierung von Cck1-Zellen im Vergleich zur ursprünglichen Veröffentlichung konnten wir die Ergebnisse nicht veröffentlichen.

Für den Datensatz der menschlichen Bauchspeicheldrüse setzten wir den Bandbreiten-Parameter auf 4 und verwendeten ein halbkreisförmiges Kernel, um gute Ergebnisse für die Zelltyp-Lokalisierung zu erzielen, und die Zelltyp-Häufigkeit stimmte mit den Einzelkern-Seq-Experimenten überein. Dieser Datensatz der menschlichen Bauchspeicheldrüse war angesichts der sehr geringen Molekülzahl pro Zelle (3 bis 5 Moleküle pro Zelle) einer der schwierigsten Datensätze, mit denen wir arbeiten konnten. Unsere positiven Ergebnisse deuten also stark darauf hin, dass unsere Methoden tatsächlich für eine breite Palette von IST-Technologien geeignet sind. Dieses Ergebnis wurde in Tosti et al. 2021 veröffentlicht, <https://doi.org/10.1053/j.gastro.2020.11.010>.

Insgesamt zeigt unsere Analyse, dass unser methodischer Ansatz tatsächlich für ISS geeignet ist.

## **WP 4: Nutzerfreundlichkeit & Leistungsfähigkeit**

Work Package 4 hatte den Hintergedanken unser Tool benutzerfreundlich und für eine breite Öffentlichkeit verfügbar zu machen, einschließlich Forschern ohne Zugang zu Hochleistungsrechnern und Programmierkenntnissen. Durch Feedback von wichtigen Kooperationspartnern mit denen tumorbefallene-SRT-Samples analysiert wurden, haben wir beschlossen, dieses Work Package in den zweiten Förderzeitraum vorzuziehen. Das Vorgehen sollte den Zweck erfüllen, unser Tool und Analysekonzept möglichst frühzeitig und reichweitenstark zu publizieren. Der Algorithmus wurde in das webbasierte Projekt "SSAM-lite" integriert, auf das Benutzer über das Internet zugreifen und mit dessen Hilfe sie auf ihren lokalen Rechnern SRT-daten analysieren können. Damit die Software auf der Hardware von Endnutzern effizient ausgeführt werden kann, wurden Teile der Algorithmen und Implementierungen angepasst und optimiert.

### **WP 4.1: KDE-Algorithmus optimieren**

Aufgabe 4.1 hatte zum Ziel, die Implementierung des KDE-Algorithmus zu verbessern, welcher eine Schwachstelle in der ursprünglichen SSAM-Implementierung aus dem ersten Berichtszeitraum darstellte.

Die Optimierung stützt sich größtenteils auf die ursprüngliche Implementierung der KDE in SSAM. Eine Heuristik wurde angewendet, um SSAMs internes Modell der Moleküldichte nur an Positionen zu berechnen, an denen ein deutliches Signal existiert, wobei die marginalen Randbereiche des Gaußschen Kernelmodells ignoriert werden. Diese Implementierung erhielt den Namen "SSAM-fast\_kde" und wurde sowohl der ursprünglichen SSAM-Implementierung als auch dem Web-Tool hinzugefügt. In einem Benchmarking-Experiment verbesserte fast\_kde die Laufzeit des Algorithmus um einen Faktor von ~100. Nach diesem erheblichen Leistungszuwachs haben wir weiter an der Optimierung des Algorithmus gearbeitet, um eine reibungslose Bedienerfahrung für die breite

Nutzerschaft zu ermöglichen. Eine zweite Heuristik wurde umgesetzt, indem die Molekülkoordinaten leicht auf die exakten Pixelkoordinaten des Vektorfeldes des SSAM-Gewebemodells verschoben wurden. Auf diese Weise konnte ein universeller Gauß-Kernel vorberechnet und an den beobachteten Molekülkoordinaten in das Vektorfeld kopiert werden. Diese zweite Heuristik konnte die KDE-Laufzeit um einen Faktor von ~10 verbessern (als "SSAM-lite" bezeichnet). Dies führte zu einer 1000-fachen Steigerung der Geschwindigkeit gegenüber der ursprünglichen Implementierung. Beide Heuristiken wurden in unserer neuesten Publikation über SSAM-lite von 2022 getestet, beschrieben und veröffentlicht.

#### **WP 4.2: GPU-Optimierung**

Der Sonar-Algorithmus nutzt zur Ermittlung von Konkurrenzkurven pixel-diskretisierte Repräsentationen von Zelltypkarten. Während der Berechnung werden mehrere Matrixfaltungen, Multiplikationen und Reduktionen durchgeführt, welche allesamt großes Parallelisierungspotential bieten. Durch die Nutzung moderner, im großen Stil parallelisierter GPU-Architekturen kann Sonar entsprechend große räumliche Datensätze effizienter verarbeiten und die Berechnungszeit im Vergleich zu CPU-basierten Ansätzen reduzieren.

Um eine effiziente Konvertierung auf GPU-basierte Architekturen zu gewährleisten, wurde Sonar größtenteils in PyTorch implementiert. PyTorch ist ein Framework für numerische Berechnungen in Python, das eine CUDA-Bindung für seine GPU-beschleunigten Berechnungen verwendet. Das Framework ist weit verbreitet in modernen Anwendungen des maschinellen Lernens und bietet als solches eine flexible Umgebung für dynamische Berechnungsgraphen und eine umfangreiche Bibliothek von voroptimierten Operationen, die es ermöglichen, komplexe Algorithmen effizient auf GPUs auszuführen. Die Entscheidung für PyTorch erleichtert somit nicht nur die Nutzung von GPUs, sondern unterstützt auch die Skalierbarkeit und Anpassungsfähigkeit des Sonar-Algorithmus im Hinblick auf zukünftige Entwicklungen im Bereich der räumlichen Analyse mit erheblich größeren und tieferen Datensätzen.

Die Matrixfaltung als ressourcenintensivste Operation im Berechnungsgraphen des Sonar-Algorithmus profitiert besonders von der Umstellung zu PyTorch. Matrixfaltungen sind kennzeichnend für moderne Bildanalysealgorithmen im maschinellen Lernen und werden als solche intensiv erforscht und optimiert. Sonar verwendet eine in PyTorch implementierte `fft-convolution`, welche, dem Faltungstheorem folgend, die Faltung als punktweise Multiplikation im Frequenzraum durchführt. Um Daten nicht redundant in den Frequenzraum konvertieren zu müssen, werden konvertierte Dichteverteilungen von Zelltypkarten zwischengespeichert und wiederverwendet.

#### **WP 4.3: Online-App erstellen**

Wie oben beschrieben wurde der grundlegende SSAM-Algorithmus in ein leicht zugängliches, browserbasiertes Analysewerkzeug namens "SSAM-lite" übersetzt. SSAM-lite verfügt über eine benutzerfreundliche Oberfläche, die aus einem "Datenzentrum" (in dem Benutzer ihre Daten vor der Analyse überprüfen können), einem Analysebereich (der zur Anpassung von Modellparametern und schnellen Anwendung der verschiedenen Verarbeitungsfunktionen der SSAM-Analyse-Pipeline genutzt werden kann) und einem "Ergebnisbereich" (in dem der Benutzer das Ergebnis überprüfen, das Layout der Zelltypkarte für eine optimale Ergebnisausgabe vorbereiten und die Ergebnisse herunterladen kann) besteht. Die Online-Bereitstellung erfolgt über Netlify - ein Bereitstellungssystem, das stets die neueste Version des Tools über ein CI/CD (continued integration/continued deployment) - System bereitstellt. Die Online-App ist unter <https://ssam-lite.netlify.app/> erreichbar.

Unsere Implementierung ist geräteunabhängig, und in Verbindung mit den Leistungssteigerungen im KDE-Schritt, die in WP4.1 beschrieben wurden, konnten wir die Online-App erfolgreich auf einen Datensatz von 250.000 Datenpunkten auf einem Mobiltelefon anwenden.

Das Tool mündete in einer Publikation und wird aktiv vom Entwicklerteam unterstützt und gewartet. Es wurde für die explorative Datenanalyse in mehreren Projekten zur räumlich aufgelösten Transkriptomik verwendet (z.B. Dr. Olivia Nonn, Charité Berlin, ECRC; Dr. Amin El-Heliebi, Medizinische Universität Graz; Prof. Elena Braicu, Charité – Universitätsmedizin Berlin). Darüber hinaus wurde das Tool auf verschiedenen Treffen vorgestellt, einschließlich des EASI-Genomics-Konsortiums (EU H2020-Projekt 824110).

#### **WP 4.4: Containerisierung für Cloud Deployment**

In enger Zusammenarbeit mit dem de.NBI-Cloud-Team haben wir erfolgreich eine Cloud-Version von SSAM-lite erstellt, die 4 VMs (56 Kerne und 116 GB RAM) nutzt und für sehr große Datensätze geeignet ist. Diese Instanz kann über <https://ssam-lite.bihealth.org/> mit der gleichen Benutzeroberfläche wie die Netlify-App (WP4.3) aufgerufen werden.

In der ursprünglichen Projektbeschreibung hatten wir das Ziel, ein Docker-Image der Anwendung über DockerHub zu veröffentlichen. Allerdings hat DockerHub sich von seinem kostenlosen und offenen Anwendungsmodell entfernt, indem sie den kostenlosen Team-Plan eingestellt haben. Obwohl wir ein Docker-Image verfügbar haben, haben wir uns entschieden, dieses Work Package durch direkten Zugriff auf die Cloud-Instanz von SSAM-lite zu implementieren. Das Docker-Image können wir noch immer bereitstellen, sollte ein neues, geeigneteres Hosting-Repository verfügbar werden.

#### **WP 3: Räumlicher Gewebeaufbau**

WP 3 wurde konzipiert, um den räumlichen Aufbau von Tumorgewebe zu untersuchen. Wir haben einen Algorithmus zur Analyse räumlicher Zusammenhänge zwischen Zelltypen im Gewebekontext erstellt. Die entsprechende Anwendung ist jetzt in unserem SSAM-Tool

(<https://github.com/eilslabs/ssam>) und/oder Sonar-Tool (<https://github.com/eilslabs/sonar>) verfügbar.

##### **WP 3.1: Räumliche Zusammenhänge identifizieren**

Zusätzlich zu der bis zum Ende des Jahres 2022 erfolgten teilweisen Aktualisierung von Sonar haben wir ein weiteres statistisches Modell implementiert. Wir haben eine Möglichkeit gefunden, den erwarteten Hintergrund korrekt zu modellieren, indem wir die Größe des radialen Kerns und die Gesamtanteile der über das gesamte Gewebe beobachteten Zelltypen berücksichtigen. Diese Normalisierung ermöglichte es uns, intuitive Anreicherungsgrade zu erhalten und auch ein statistisches Modell zu implementieren. Angesichts der großen Anzahl von Datenpunkten gingen wir davon aus, dass der zentrale Grenzwertsatz die Annahme der Normalität rechtfertigt, die wir durch Anwendung des Kolmogorov-Smirnov-Tests oder des Shapiro-Wilk-Tests nicht verwerfen konnten. Somit ist die Ausgabe der Daten des Sonar-Tools jetzt wesentlich genauer und besser.

##### **WP 3.2: Lineare Modelle von Zell- und Gewebeaufbau**

Die zelluläre Organisation spielt eine herausragende Rolle für die Definition der Gewebefunktion, daher ist die Analyse der Gewebearchitektur wichtig. Um diese lineare Beziehung zwischen den Zellen innerhalb des Gewebes zu modellieren, haben wir einen Ansatz mit gleitenden Fenstern („sliding window approach“) implementiert, um Domänensignaturen auf der Grundlage der Anzahl der Zelltypkennzeichnungen im Fenster zu berechnen. Die Gewebedomänen werden durch Clustering dieser Signaturen mithilfe von agglomerativem Clustering definiert.

Dieser Algorithmus wurde auf eine bekannte Probe des somatosensorischen Kortex von Mäusen angewandt, die mittels osmFISH profiliert wurde (Codeluppi et al. 2018). Die identifizierten Domänen korrelierten gut mit den bekannten Schichten der Großhirnrinde, was gut mit zuvor publizierten Ergebnissen übereinstimmt. Zusätzlich zur Identifizierung der korrekten Verteilung von pyramidalen Neuronen in den erwarteten Neokortexschichten zeigte die Zelltypzusammensetzung der Domänen, dass Mfge8 exprimierende Astrozyten 7-14 % jeder der Gewebeschichten ausmachten, im Gegensatz zu der deutlich geringeren Anzahl von Astrozyten-Mfge8- Zellen, die in der vorherigen Studie genannt wurden.

##### **WP 3.3: Nicht-lineare Modelle von Zell- und Gewebeaufbau**

Ein Hidden- Markov- Random- Field (hMRF)-Modell lässt sich auf natürliche Weise auf das Problem der Zuweisung von Gewebebereichen übertragen. Jeder Zelle sind zwei Graphen zugeordnet. Der verborgene Knoten modelliert die Wahrscheinlichkeit der Zugehörigkeit einer Zelle zu einer Gewebedomäne. Der sichtbare Knoten stellt die beobachtete Genexpression dar. In unserer Implementierung haben wir ein Potts-Modell für die Interaktionsenergie gewählt, d. h. sie ist gleich Null, wenn alle Nachbarn derselben Gewebedomäne zugeordnet sind, und erhöht sich bei jeder Abweichung

um eins. Die unäre Energie wird als Kosinusabstand zwischen der Genexpression einer Zelle und dem Mittelwert aller Zellen, die demselben Gewebebereich angehören, berechnet.

Die Durchführung einer hMRF-Analyse für einen Datensatz der räumlichen Transkriptomik erfolgt in folgenden Schritten: Zunächst wird die Datei mit den Rohdaten vom Programm gelesen und in eine konsistente interne Datendarstellung umgewandelt. Dies wird derzeit für den Loom-Dateistandard, Baysor-Ausgabedateien und ein Standard-Komma-separiertes Dateiformat unterstützt. Aus dieser Darstellung wird durch Delaunay-Triangulation ein Umgebungsdiagramm erstellt. Für nicht-konvexe Gewebeformen wird ein Modell von 'fehlerbehafteten Umgebungen' eingeführt. Die entsprechenden Ausreißer werden durch einen Abstandsschwellenwert herausgefiltert, der vom Benutzer als Vielfaches des Medians der globalen Abstandsverteilung angegeben werden kann. Die Clusterzuordnungen werden mit Hilfe eines k-Means Clustering des Genexpressionsraums initialisiert. Die Cluster-Mittelwerte werden aus dieser anfänglichen Zuordnung berechnet. Der **Expectation-Maximization-Algorithmus** alteriert daraufhin zwischen einem Gibbs-Sampling-Schritt für die Zuordnung der Gewebebereiche und der anschließenden Berechnung der Mittelwerte der Gewebebereiche.

Das hMRF-Modell wurde auf den somatosensorischen Kortex mit zwei verschiedenen Zellsegmentierungen angewendet. Alle Gewebedomänen (Ventrikel, Hippocampus, weiße Substanz, kombinierte innere Kapsel (IC) / Caudoputamen (CP) und Schichten 1, 2/3, 4, 5,6) konnten zugeordnet werden, obwohl die hMRF-Datenausgabe Domänen aufbrechen oder unterschiedliche Grenzen aufweisen kann. Diese Ergebnisse stellen eine wesentliche Verbesserung gegenüber dem linearen Modell (unter WP 3.2 geschildert) dar.

Das allgemeine Interesse der Community an der Modellierung der zellulären Organisation in Gewebe war so groß, dass wir einen weiteren Hackathon organisierten (siehe Sonstiges für weitere Einzelheiten).

## WP2: Biologische Funktionen

### WP 2.1: Unterstützung von Segmentierungsmasken

Wir haben Unterstützung für Segmentierungsmasken hinzugefügt, indem wir eine „Eingabemaske“-Option hinzugefügt haben, die die Analyse auf bestimmte Regionen des Bildes beschränkt. Dies kann eine Segmentierungsmaske über die Python-Bibliothek `matplotlib.patches.Polygon` sein. Wir haben diese Funktion bei der Analyse des Maus-VISP-Datensatzes verwendet.

### WP2.2 Clustering und Zelltypsignaturanalyse

Bei der Untersuchung alternativer Clustering-Methoden sind uns zwei wichtige Punkte aufgefallen: (i) die Detektion nach Leiden ist nach wie vor die am häufigsten verwendete Methode in diesem Bereich, und (ii) die meisten Clustering-Methoden sind nicht raumbezogen. „Leiden Clustering“ ist eine deutliche Verbesserung gegenüber der Methodik „Leiden für Einzelzell-RNAseq-Daten“ (<https://www.nature.com/articles/s41598-019-41695-z>). Die Implementierung unterstützt jedoch keine räumlichen Einbettungen. Daher haben wir eine raumbezogene Version des Leiden-Algorithmus implementiert, die wir SpatialLeiden nennen und die kürzlich in **Genome Biology** zur Veröffentlichung angenommen wurde (<https://genomebiology.biomedcentral.com/articles/10.1186/s13059-025-03489-7>). Wir zeigen, dass durch die Integration räumlicher Informationen in verschiedenen Schritten (Merkmalsauswahl, Dimensionalitätsreduktion und/oder Graphenkonstruktion) das Leiden-Clustering zu einer rechnerisch sehr leistungsfähigen, bewusst räumlichen Clustering-Methode wird, die sich gut mit anderen modernen räumlichen Clustering-Algorithmen messen kann. Bei der Anwendung auf eine Mausgehirnprobe zeigen wir eine deutliche Verbesserung gegenüber dem normalen Leiden-Clustering, das auf der Auswahl hochvariabler Gene (HVGs) und PCA basiert.

Außerdem haben wir festgestellt, dass das derzeitige Paradigma zu einer Klassifizierung von Zelltypen ohne Berücksichtigung der Unsicherheit führt. Um dies zu beheben, haben wir den „Assignment Score“ in unser neues Tool `Sainsc` implementiert, das in *Small Methods* (<https://onlinelibrary.wiley.com/doi/full/10.1002/smt.202401123>) veröffentlicht wurde. Dabei

berücksichtigten wir die Differenz zwischen der besten und der zweitbesten möglichen Zelltypzuordnung und definierten anhand der Kosinusähnlichkeit einen Vertrauenswert für die Zelltypzuordnung. Dies hat erhebliche Auswirkungen auf die Interpretation der Zelltypen im Gewebe, da der Benutzer nun beurteilen kann, ob neuartige Beobachtungen auf vertrauenswürdige Zelltypzuweisungen zurückzuführen sind oder ob es sich bei den neuartigen Beobachtungen möglicherweise um ein Artefakt handelt, das auf unsichere Zelltypzuweisungen zurückzuführen ist.

Um dies zu demonstrieren, haben wir in unserer Analyse einen bestimmten Zelltyp in der Region des paraventriculären Nukleus (PVT) im Gehirn der Maus identifiziert. Die PVT-Region enthält glutamaterge Neuronen. Die ursprüngliche Analyse von Salas et al. bezeichnete die Neuronen im PVT als „Thal-Glut-Neuronen“, während Sainsc diese als „MEA-COA-BMA Slc17a6 Glut“ erkannte. Die Genexpressionssignaturen von „Thal-Glut-Neuronen“ und „MEA-COA-BMA Slc17a6 Glut“ sind sehr ähnlich (Kosinusähnlichkeit 0,79). In der Sainsc-Analyse war der Zuordnungsscore der PVT-Region viel niedriger als der des umgebenden Thalamus, was uns zu der Vermutung veranlasste, dass die Kombination aus der Zelltyp-Signaturmatrix und dem Gen-Panel für die Annotation der PVT-Region nicht optimal war. Während wir das Gen-Panel nicht ändern können, untersuchten wir die Auswirkungen einer Änderung der Zelltyp-Signaturmatrix. Wir entschieden uns, die Probe anhand der Zelltyp-Signaturen aus der Einzelzell-RNA-seq-Studie von Yao et al. neu zu annotieren, die die Signatur einer PVT-spezifischen neuronalen Klasse, „PVT-PT Ntrk1 Glut“, enthielt. Unter Verwendung dieser Signaturmatrix hat Sainsc die PVT-Region korrekt annotiert, so dass sie die neuronale Klasse „PVT-PT Ntrk1 Glut“ mit einem hohen Zuordnungswert enthält.

Geplant war hier als Meilenstein D2.2, welcher als Minor github release von SSAM mit Unterstützung für Clusteringmethoden und vorher bekannten Zuweisungen von Zelltypen (M16) definiert war.

Stattdessen wurde der Meilenstein D2.2 dadurch erreicht, dass ein „Assignment Score“ in unser neues Tool Sainsc implementiert wurde: <https://onlinelibrary.wiley.com/doi/full/10.1002/smt.d.202401123>

### WP2.3 Integration von nicht-räumlichen scRNAseq-Daten

Entgegen unseren Erwartungen haben wir festgestellt, dass anspruchsvollere maschinelle Lernmethoden für die Integration von Einzelzell- und räumlichen Transkriptomikdaten nicht unbedingt zu einer besseren Qualität der Annotation von Zelltypen für bildgebungsbasierte räumliche Transkriptomikdaten führen. Hier konzentrierten wir uns auf die Daten des somatosensorischen Kortex des Mäusegehirns, die mittels osmFISH (Codeluppi *et al.*) erstellt wurden. Wir haben versucht, verschiedene traditionelle maschinelle Lernverfahren (Random Forest, Naive Bayes, XGBoost) und eine Reihe von Deep-Learning-Methoden zu verwenden. Allerdings waren adversarische Autoencoder (AAE) die einzigen Modelle, die ein vernünftiges Ergebnis lieferten, das die ursprünglichen Ergebnisse der Studie von Codeluppi *et al.* nachahmte, die von unserem SSAM de novo-Modus bereits sehr gut dargestellt wurden. Der AAE-Ansatz war viel zu empfindlich und relativ langsam im Vergleich zu dem ursprünglich in SSAM implementierten statistischen Ansatz.

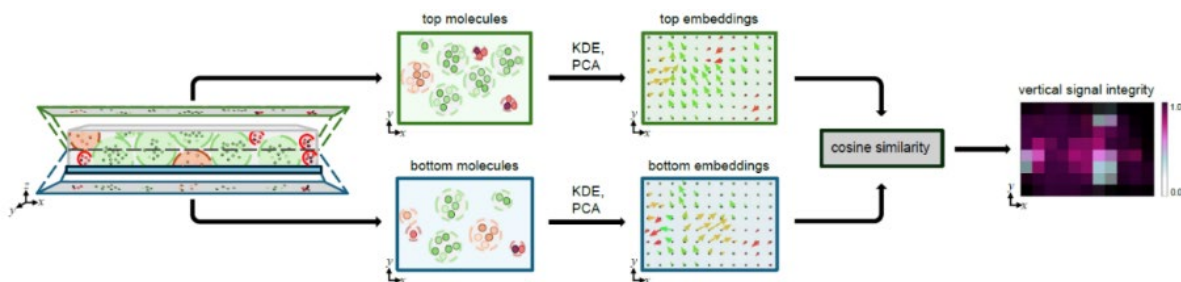
Darüber hinaus haben wir festgestellt, dass eine weitere Vereinfachung der Integration von scRNAseq-Daten zu akzeptablen Ergebnissen führt. Bei einer Analyse von Bauchspeicheldrüsendaten (menschliche Bauchspeicheldrüsenläppchen, die mittels In-situ-Sequenzierung mit einem Genpanel von 120 Markergenen profiliert wurden) konnten wir tatsächlich beobachten, dass die Integration mit scRNAseq-Daten, bei der nur die binäre Zuordnung von Markergenen zu Zelltypen verwendet wurde, entgegen unseren Erwartungen den räumlichen Zelltyp tatsächlich verbesserte. Diese Ergebnisse wurden in der Zeitschrift **Gastroenterology** veröffentlicht ([https://www.gastrojournal.org/article/S0016-5085\(20\)35399-3/fulltext](https://www.gastrojournal.org/article/S0016-5085(20)35399-3/fulltext)). Wir glauben, dass dies auf die vereinfachte Natur der Gen-Panels zurückzuführen ist, die für die bildgebungsbasierte räumliche Transkriptomik verwendet werden - diese Gene werden für 1:1-Beziehungen mit Zelltypen ausgewählt, was vielleicht die Verwendung von binären Klassifizierungsmethoden im Vergleich zur Bewertung der qualitativen Genexpression rechtfertigt. Wir gehen jedoch davon aus, dass sich dies bei der Verwendung größerer Gen-Panels ändern kann, bei denen die Zuordnung von Genen zu Zelltypen möglicherweise nicht so einfach ist. Wir betrachten dies als einen Bereich, der weitere Untersuchungen rechtfertigt, sobald neuere Modelle verfügbar sind.

Geplant war hier als Meilenstein D2.3, welcher als Minor github release von SSAM mit Integrierung von scRNAseq (M18) angelegt war. Stattdessen wurde dieses Feature in Sainsc integriert. Die Integration der scRNAseq Daten ist beschrieben unter: <https://sainsc.readthedocs.io/stable/usage.html>

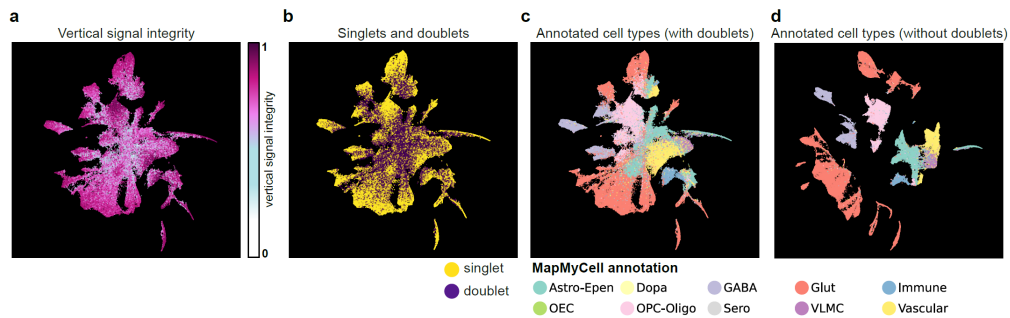
## WP2.4 Subzelluläre Lokalisierung

Bei der Bearbeitung dieser Aufgabe stießen wir auf eine konkurrierende Publikation, in der das Berechnungsprogramm BENTO beschrieben wird, welches diese sehr spezifische Aufgabe effektiv angeht: <https://genomebiology.biomedcentral.com/articles/10.1186/s13059-024-03217-7>. Während diese Publikation in diesem Jahr veröffentlicht wurde, nahmen wir persönlichen Kontakt mit dem Erstautor, Clarence Mah, auf und diskutierten seine Preprint-Ergebnisse (<https://www.biorxiv.org/content/10.1101/2022.06.10.495510v1>) im Zusammenhang mit dem von uns vorgeschlagenen Plan für SAGE. Die Diskussion führte schließlich zu dem Gedanken, dass die Analyse der subzellulären Lokalisierung mit der genauen Abgrenzung der Zellgrenzen (d. h. der Zellsegmentierung) zusammenhängt, ein Problem, das wir aus unserer früheren SSAM-Veröffentlichung kennen, die als Teil von SAGE (<https://www.nature.com/articles/s41467-021-23807-4>) veröffentlicht wurde. Daher haben wir den Ansatz verfolgt, subzelluläre Genexpressionssignale auszuwerten, die von sich überlappenden Zellen stammen könnten. Kurz gesagt, räumliche Transkriptomikdaten werden aus Gewebeschnitten von ~10  $\mu\text{m}$  erzeugt, und bisher haben Forscher angenommen, dass diese Gewebeschnitte zweidimensional sind. Wir zeigen jedoch systematisch, dass dies eine zu einfache Annahme ist und die dreidimensionale Organisation von Zellen in Gewebeschnitten zu einer großen Anzahl von Überlappungen (d. h. räumlichen Dubletten) führt, wie in unserem Preprint auf biorxiv mit dem Titel „2D, or not 2D? Investigating Vertical Signal Integrity of Tissue Slices“ (<https://www.biorxiv.org/content/10.1101/2025.01.13.632601v1>) dargestellt ist.

In der Publikation wird beschrieben, dass bildgebungs-basierte, orts aufgelöste Transkriptomik Transkripte innerhalb von Zellen in 3D lokalisieren kann. Die Zellsegmentierung, ein wichtiger Schritt, der der Zuordnung von Transkripten zu Zellen und der Annotation von Zellfunktionen vorausgeht, wird jedoch in der Regel in 2D durchgeführt und ist daher nicht in der Lage, mit räumlichen Dubletten umzugehen, die sich aus überlappenden Zellen ergeben, was dazu führt, dass segmentierte Zellen Transkripte enthalten, die von mehreren Zelltypen stammen. In unserer Studie entwickeln wir ein Proof-of-Principle-Modell, stellen ein computergestütztes Tool namens `ovrlpy` vor, das überlappende Zellen, Gewebefalten und ungenaue Zellsegmentierung identifiziert (**Abbildung 2**), das benutzerfreundliche und dennoch detaillierte Ausgabebilder erzeugt, um Zellüberlappungen und Gewebefalten zu identifizieren. Wir zeigen auch, dass die Anwendung des `ovrlpy`-Tools auf die herkömmliche, auf Zellsegmentierung basierende Analyse die Qualität der Clusterbildung verbessern kann, was zu einer besseren nachgeschalteten Analyse führt (**Abbildung 3**).



**Abbildung 2:** eine Darstellung der ovrlpy-Signal-Kohärenz-Pipeline. Die Werkzeuge erstellen zunächst „virtuelle Teilscheiben“ von räumlichen Transkriptomikdaten, um den Vergleich von Molekülen am oberen und unteren Rand des Gewebes zu ermöglichen, die dann vor dem Vergleich in niedrigdimensionale Einbettungen projiziert werden. Der quantitative Vergleich der Ähnlichkeit der Genexpression im oberen und unteren Teil des Gewebes bildet die Grundlage für den „vertikalen Signalintegritäts-Score“ (VSI), der anzeigt, ob sich eine Zelle überlappt oder nicht.



**Abbildung 3:** Die unüberwachte Dublettenerkennung von Ovrly verbessert die Zelltypentrennung im Genexpressionsraum. UMAP-Einbettungen von Genexpressionsdaten aus Xenium-Mausgehirn-Zellsegmenten. (A) Segmente eingefärbt nach flächengemittelter vertikaler Signalintegrität (VSI). (B) Segmente, die durch die Dublettenerkennung von ovrly gefiltert wurden. (C) Segmente, die durch MapMyCells-Zelltyp-Klassifizierungen kommentiert wurden. (D) UMAP-Einbettung nach Ausschluss der von ovrly identifizierten Doubletten, eingefärbt durch MapMyCells-Zelltyp-Anmerkungen.

Als Meilenstein D2.4 war das Minor github release von SSAM mit Unterstützung für die Segmentierung und Identifizierung subzellulärer Strukturen (M21) geplant. Stattdessen wurde dieses Feature implementiert in Sainsc anstelle von SSAM und wurde veröffentlicht unter: [https://sainsc.readthedocs.io/stable/tutorials/VisiumHD\\_CRC.html](https://sainsc.readthedocs.io/stable/tutorials/VisiumHD_CRC.html)

#### WP5: Code veröffentlichen und Dokumentation erstellen

Der Code für SSAM und SSAM-lite wurde jeweils auf einem kuratierten GitHub-Repository veröffentlicht und wird aktiv gewartet (<https://github.com/HiDiHlabs/ssam>, <https://github.com/HiDiHlabs/ssam-lite>). Es existieren weiterführende Entwicklungsaktivitäten, und Beiträge und Feedback von der Scientific Community werden aktiv ermutigt und integriert, wie z.B. eine Optimierung der Performance des Tools in Bezug auf die Nutzung des Arbeitsspeichers. Beide Projekte erhielten eine umfangreiche Dokumentation (<https://ssam.readthedocs.io/stable/tutorials/VisiumHDCRC.htm/>, <https://spatialeiden.readthedocs.io/>), wurden geprüft und veröffentlicht ([doi.org/10.1101/2021.09.29.462194](https://doi.org/10.1101/2021.09.29.462194), [doi.org/10.1038/s41467-021-23807-4](https://doi.org/10.1038/s41467-021-23807-4)). Somit war der Meilenstein D5.2 erfüllt, der das Major Release des Tools zur räumlichen Analyse mit allen neuen Features zur Aufgabe hatte (M48).

#### Liste der erreichten Meilensteine (alle Meilensteine wurden erreicht):

- D1.1: Unterstützung für unterschiedliche KDE Kernel für SSAM auf GitHub veröffentlichen (M3)
- D1.2: Zellarten für Maus-VISp and MTG für unterschiedliche IST-assays kartieren(M6)
- D1.3: Minor github release von SSAM mit ISS-Unterstützung (M10)
- D2.1: Minor github release von SSAM mit Unterstützung von Segmentationsmasken (M12)
- D2.2: Minor github release von SSAM mit Unterstützung für Clusteringmethoden und vorher bekannten Zuweisungen von Zelltypen (M16)
- D2.3: Minor github release von SSAM mit Integrierung con scRNAseq (M18)
- D2.4: Minor github release von SSAM mit Unterstützung für die Segmentierung und Identifizierung subzellulärer Strukturen (M21)
- D3.1: Erster Release eines neuen Tools zur Analyse räumlicher Strukturen (M37)
- D3.2: Update des Tools zur räumlichen Analyse mit linearem Modell des Gewebeaufbaus (M41)
- D3.3: Update des Tools zur räumlichen Analyse mit nicht-linearem Modell des Gewebeaufbaus (M41)
- D4.1: Veröffentlichung einer vorläufigen GitHub-Version von SSAM mit optimierter KDE (M26).
- D4.2: Minor github release von SSAM mit GPU-Optimierung (M29)
- D4.3: Aktiver IST-Tool-Webserver (M32).
- D4.4: Angepasst: Reproduzierbare Cloud-Bereitstellung von IST-Tools mit leistungsstarkem Hardware-Backend
- D5.1: Hauptversion von SSAM, mit allen neu erstellten Funktionen, als SSAM v2 veröffentlicht (M24).
- D5.2: Major release des Tools zur räumlichen Analyse mit allen neuen Features (M48)

### Wichtigste Positionen des zahlenmäßigen Nachweises und weitere Erläuterungen zu den Ausgaben:

Das Projekt konnte neben den erzielten fachlichen Ergebnissen auch in Bezug auf die Finanzen insgesamt positiv beendet werden, so dass abschließend 682,82€ (ohne PP) des bewilligten Budgets nicht in Anspruch genommen werden müssen. Allerdings wurden im Verlauf des Projektes die Personalmittel etwas überzogen, da sich besonders in der Endphase herausstellte, dass die Vielzahl publizierbarer Ergebnisse mehr personelle Unterstützung benötigte als initial geplant. Hier fand eine Umwidmung von Mitteln hauptsächlich aus dem Bereich der sächlichen Verwaltungsausgaben in die Personalmittel statt.

Beschreibung	Plan (Gesamt)	IST (esamt)	Differenz Plan-IST
Summe Personalausgaben	219.511,08 €	226.123,27 €	-6.612,19 €
Summe sächlicher Verwaltungsausgaben	18.300,00 €	12.529,90 €	5.770,10 €
Gegenstände und andere Inv. >800€	3.500,00 €	1.975,09 €	1.524,91 €
Gesamtausgaben	241.311,08 €	240.628,26 €	682,82 €
Gesamtausgaben (inkl.PP)	289.573,30 €	288.753,91 €	819,38 €

**Tabelle 1:** Übersicht Plan (AZAP)/Ist der Personalausgaben, Summe sächlicher Verwaltungsausgaben und Gegenstände und andere Investitionen im Projektverlauf sowie der Differenz in den Gesamtausgaben (€).

**Personalausgaben (F0824):** Mit dem Start des Projektes im Juli 2021 wurde ein Doktorand beschäftigt. Ein Doktorand konnte neben studentischen Hilfskräften über die volle Laufzeit des Projektes gefördert werden, so dass dieser jetzt erfolgreich seine Dissertation abschließen kann. Ein weiterer Doktorand unterstützte speziell in den späten Phasen des Projektes mit dem Ziel den Output an wissenschaftlichen Artikeln zu unterstützen. Des Weiteren wurden studentische Hilfskräfte beschäftigt um gezielt bei bestimmten Projektaktivitäten unterstützend tätig zu werden.

Es fand zum Ende des Projektes noch eine Umwidmung von Sachmitteln/Investitionsmitteln im Wert von insgesamt 6.612,19€ in Personalmittel statt zur Gewährleistung des Outputs publizierbarer Ergebnisse.

#### **Sächliche Verwaltungsausgaben (F0847):**

Aus Tabelle 1 lässt sich entnehmen, dass das Budget für sächliche Verwaltungsausgaben nicht im vollen Umfang entsprechend dem Plan genutzt wurde.

Position F0847 unterteilt sich in verschiedene Finanzpositionen, zu denen wir Ausgaben geplant hatten (siehe *Tabelle 2*). Auch in dieser Tabelle sind Plan- und Ist -Ausgaben verglichen worden, wobei in allen Positionen das geplante Budget nicht vollständig ausgeschöpft wurde. Unter den weiteren Sachausgaben 2 (F0842) waren Kosten für Publikationen und Workshops geplant. Das Budget wurde verwendet zur Finanzierung zweier der zahlreichen Publikationen sowie der Finanzierung eines Hackathons, der im Dezember 2024 stattfand. Reisekosten waren in diesem Projekt in großem Maß eingeplant und konnten dann aufgrund der Pandemie nur vermindert ausgeschöpft werden. Hier fallen insgesamt nur 298,15€ für nationale und 2.969,13€ für internationale Reisen an.

Beschreibung	Plan (Gesamt)	IST gesamt	Delta
Weitere Sachausgaben 2	10.300,00 €	9.262,62 €	1.037,38 €
Dienstreisen Inland	3.000,00 €	298,15 €	1.701,85 €
Dienstreisen Ausland	5.000,00 €	2.969,13 €	2.030,87 €
Summe sächlicher Verwaltungsausgaben	18.300,00 €	12.529,90 €	4.770,10 €
Summe sächlicher Verwaltungsausgaben (inkl. PP)	21.960,00 €	15.035,88 €	5.724,12 €

**Tabelle 2:** Übersicht Plan (AZAP) /Ist der sächlichen Verwaltungsausgaben (F0847) im Detail sowie der Differenz im Verhältnis Plan – Ist.

### **Gegenstände und andere Investitionen >800€ (F0850):**

Für die Position 0850 waren ursprünglich 3500,00€ Budget geplant, die sich zusammensetzten aus den Kosten für eine externe GPU und einen Laptop. Die Ausgaben beliefen sich hier jedoch nur auf 1.975,09€ (ohne PP). Dieses ist begründet darin, dass der Kauf einer externen GPU nicht notwendig erschien, weil diese aufgrund der vorhandenen Anzahl an GPU-Cluster-Knoten nicht mehr benötigt wurde.

### **Notwendigkeit und Angemessenheit der geleisteten Projektarbeiten:**

Das Projekt hat wesentlich dazu beigetragen eine bestehende Lücke in der Analyse von Spatial Omics-Daten zu schließen. Es ist seit Beendigung des Projektes möglich wesentlich präzisere Aussagen im Bereich der zellsegmentierungsfreien Frameworks für die Analyse räumlicher Omics Daten vorzunehmen. Dadurch entsteht ein tieferes Verständnis für die Vorgänge in kleineren Zellverbänden und in Mikroumwelten von Tumoren, welches dabei helfen wird, Tumore und das Verhalten von Zellverbänden weiterhin zu ergründen räumliche Heterogenität das Fortschreiten der Krankheit, das Ansprechen auf die Therapie und die Prognose beeinflusst.

Die im Projekt erarbeiteten Tools sind veröffentlicht und werden bereits in der Scientific Community benutzt.

### **Voraussichtlicher Nutzen, die angestrebte Verwertbarkeit des Ergebnisses:**

Der Nutzen und die Verwertbarkeit der Ergebnisse liegen in deren Veröffentlichung und Verbreitung, um einer möglichst breiten Community die neu erzeugten Tools zur Benutzung zur Verfügung zu stellen.

Das SAGE-System ermöglicht eine robuste, hochauflösende räumliche Kartierung der Genexpression und ermöglicht so einen Durchbruch beim Verständnis der Mikroumgebung von Tumoren, der Zell-Zell-Interaktionen und der räumlichen Organisation von Geweben. Dies hat unmittelbare Auswirkungen auf die Krebsforschung, wo es die räumliche Heterogenität das Fortschreiten der Krankheit, das Ansprechen auf die Therapie und die Prognose beeinflusst.

Die Werkzeuge erleichtern auch die Integration der räumlichen Transkriptomik mit anderen Omics-Daten, und tragen so dazu bei, dass mit höherer Sicherheit und Zuverlässigkeit neue biologische Erkenntnisse gewonnen werden können.

Aus wirtschaftlicher Sicht werden die Projektergebnisse durch Open-Source-Vertrieb nach FAIR-Prinzipien, Integration in große Forschungskonsortien und die Bereitstellung von Schulungen und Beratung genutzt. Es besteht Potenzial für die Kommerzialisierung, z. B. durch das Angebot hochwertiger cloudbasierter Analysedienste oder die Integration des Rahmens in diagnostische und pharmazeutische Pipelines.

### **Konkrete Planungen im Sinne des fortgeschriebenen Verwertungsplanes**

Die Verwertung der Ergebnisse fand primär zunächst auf der Ebene der Veröffentlichungen von

Publikationen und Repositorien nach den FAIR -Prinzipien statt. Details dazu sind unter **Erfolgte und geplante Veröffentlichungen des Ergebnisses nach Nr. 5 der NABF** auf Seite 12 aufgeführt.

Neben der Verwertung der Ergebnisse durch Publikationen findet die Verbreitung auf Ebene des Austauschs von Ergebnissen und Erkenntnissen statt. Dieses ist der Fall bei der Teilnahme an Konferenzen und Kongressen, sowie der Organisation oder Teilnahme an Hackathons, wo eine sehr gute Möglichkeit des Kontakts nicht nur mit wissenschaftlichen sondern auch kommerziellen Stakeholdern besteht. Diese sind unter **Sonstige Aktivitäten** auf Seite 13 im Detail aufgeführt. Die Projektergebnisse werden durch Open-Source-Vertrieb nach FAIR-Prinzipien, Integration in große Forschungskonsortien und die Bereitstellung von Schulungen und Beratung bereits genutzt. Durch z. B. das Angebot hochwertiger cloudbasierter Analysedienste oder die Integration des Rahmens in diagnostische und pharmazeutische Pipelines besteht ein Potential der Kommerzialisierung. Zu den nächsten Schritten in dieser Richtung gehören die Erweiterung der Nutzerbasis, die Entwicklung von Premium-Diensten, die Anpassung der Tools für neue räumliche Omics-Plattformen und die Suche nach weiteren Finanzierungsmöglichkeiten und Partnerschaften, um die Nachhaltigkeit zu gewährleisten. Wir haben bereits Interesse von mehreren kommerziellen Unternehmen (CellZome, RemedyBio) erhalten.

#### **Während der Durchführung des Vorhabens bekanntgewordene Fortschritte des Vorhabens bei anderen**

Es sind keine vergleichbaren Ergebnisse anderer Gruppen auf diesem Gebiet bekannt. Die Gemeinschaft ist sehr klein und es kann davon ausgegangen werden, dass Fortschritte innerhalb der Scientific Community geteilt werden.

#### **Erfolgte und geplante Veröffentlichungen des Ergebnisses nach Nr. 5 der NABF**

##### **Publikationen in wissenschaftlichen Zeitschriften:**

###### **2021:**

**Park J**, Choi W, **Tiesmeyer S**, Long B, Borm LE, Garren E, Nguyen TN, Tasic B, Codeluppi S, **Graf T**, Schlesner M, Stegle O, Eils R, **Ishaque N**. Cell segmentation-free inference of cell types from in situ transcriptomics data. Nat Commun. 2021 Jun 10;12(1):3545. Erratum in: Nat Commun. 2021 Jun 28;12(1):4103. PMID: 34112806;  
<https://doi.org/10.1038/s41467-021-23807-4>

Tosti L, Hang Y, Debnath O, **Tiesmeyer S**, Trefzer T, Steiger K, Ten FW, Lukassen S, Ballke S, Kühl AA, Spieckermann S, Bottino R, **Ishaque N**, Weichert W, Kim SK, Eils R, Conrad C. Single-Nucleus and In Situ RNA-Sequencing Reveal Cell Topographies in the Human Pancreas. Gastroenterology. 2021 Mar;160(4):1330-1344.e11. doi: 10.1053/j.gastro.2020.11.010. Epub 2020 Nov 16. PMID: 33212097.  
<https://doi.org/10.1053/j.gastro.2020.11.010>

###### **2022:**

**Yun Zhang**, Jeremy A. Miller, **Jeongbin Park**, Boudewijn P. Lelieveldt, Brian Long, Tamim Abdelaal, Brian D. Aevermann, Tommaso Biancalani, Charles Comiter, Oleh Dzyubachyk, Jeroen Eggermont, Christoffer Mattsson Langseth, Viktor Petukhov, Gabriele Scalia, Eeshit Dhaval Vaishnav, Yilin Zhao, Ed S. Lein, Richard H. Scheuermann, Reference-based cell type matching of spatial transcriptomics data. **Nature scientific reports 13, Article No. 9567 (2023)**  
<https://doi.org/10.1038/s41598-023-36638-8>

**Tiesmeyer, S., Sahay, S., Müller-Böttcher, N., Eils, R., Mackowiak, S. D., & Ishaque, N.** SSAM-lite: a light-weight web app for rapid analysis of spatially resolved transcriptomics data. **Front. Genet., 2022, Feb 28;12:78587713** <https://doi.org/10.3389/fgene.2022.785877>

Nonn, O., **Debnath, O.**, Valdes, D.S., Sallinger, K., Secener, A.K., Haider, S., Fischer, C.,

**Tiesmeyer, S.**, Nimo, J., Kuenzer, T. and Maxian, T., Knöfler, M., Karau, P., Bartholomaeus, H., Kroneis, T., Frolova, A., Neuper, L., Haase, N., Kräker, K., Kedziora, S., Forstner, D., Verlohren, S., Stern, C., Coscia, F., Tilaganathan, B., Eils, R., Huppertz, B., El-Heliebi, A., Staff, A.C., Müller, D., Dechend, R., Gauster, M., Ishaque, N., Herse, F., 2022. Disturbed trophoblast transition links preeclampsia progression from placenta to the maternal syndrome. **bioRxiv**, pp.2022-10. <https://doi.org/10.1101/2022.10.10.511539>

Salas, S.M., Kuemmerle, L.B., Mattsson Langseth, C., **Tiesmeyer, S.**, Avenel, C., Hu, T., Rehman, H., Grillo, M., Czarnewski, P., Helgadottir, S., Tiklova, K., Andersson, A., Rafati, N., Chatzinikolaou, M.P., Theis, F., Luecken, M., Wählby, C., **Ishaque, N.** und Nilsson, M.  
Optimizing Xenium In Situ data utility by quality assessment and best practice analysis workflows. **Nature Methods** **22**, 813-823 (2025) <https://doi.org/10.1038/s41592-025-02617-2>

2024:

**Mueller-Boetticher, N., Tiesmeyer, S., Eils, R., Ishaque, N. et al (2024).** Sainsc: A Computational Tool for Segmentation-Free Analysis of In Situ Capture Data. **Small Methods**, Vol.9, Issue 5, 2401123, <https://doi.org/10.1002/smt.202401123>

**Mueller-Boetticher, N. Sahay, S., Eils, R., Ishaque, N.,** SpatialLeiden: spatially aware Leiden clustering. **Genome Biology** **26**, Art. No. 24 (2025) <https://doi.org/10.1186/s13059-025-03489-7>.

Gavriildis, G., Vasileiou, V., Orfanou, A., **Ishaque, N.**, Psomopoulos, F.. A mini-review on perturbation modelling across single-cell omic modalities. **Comput. Struct. Biotechnol J.** **2024**, Apr **25**:23:1886-1896 <https://doi.org/10.1016/j.csbj.2024.04.058>

**Tiesmeyer, S., Mueller-Boetticher, N., Malt, A.,** Long, B., Salas, S.M., Kiessling, P., Horn, P., Guillot, A., Kuemmerle, L.B., Ma, L., Tacke, F., Theis, F., Kuppe, C., Nilsson, M., **Eils, R., Ishaque, N.** 2D, or not 2D? Investigating Vertical Signal Integrity of Tissue Slices. **bioRxiv**. <https://doi.org/10.1101/2025.01.13.632601>

Olivia Nonn et al. Olivia Debnath, Daniela S. Valdes, Katja Sallinger, Ali Kerim Secener, Cornelius Fischer, **Sebastian Tiesmeyer**, Jose Nimo, Thomas Kuenzer, Juliane Ulrich, Theresa Maxian, Martin Knöfler, Philipp Karau, Thomas Kroneis, Alina Frolova, Lena Neuper, Nadine Haase, Alexander Malt, **Niklas Müller- Bötticher**, Kristin Kräker, Sarah Kedziora, Désirée Forstner, **Roland Eils**, Ruth Schmidt-Ullrich, Sandra Haider, Stefan Verlohren, Christina Stern, Meryam Sugulle, Stuart Jones, Basky Thila Ganathan, Tu'uhevaha J. Kaitu'u-Lino, Stephen Tong, Berthold Huppertz, Amin El-Heliebi, Anne Cathrine Staff, Fabian Coscia, Dominik M. Gauster, **Naveed Ishaque**, Florian Herse. Senescent Syncytiotrophoblast Secretion During Early Onset Preeclampsia. **Hypertension**, **2025 May**; **82(5)**: 787-799 <https://doi.org/10.1161/HYPERTENSIONAHA.124.23362>

*Veröffentlichungen von Source Code, unterstützender Dokumentation des Source Code, und neuer oder revidierter Software-Repositoryn in GitHub entsprechend den FAIR -Prinzipien:*

<https://ssam-lite.bihealth.org/>

<https://ssam-lite.netlify.app/>

Code repository: <https://github.com/HiDiHlabs/ssam>

Dokumentation "Read the Docs": <https://ssam.readthedocs.io/en/release/>

[https://ssam.readthedocs.io/en/release/userguide/07-input\\_mask.html](https://ssam.readthedocs.io/en/release/userguide/07-input_mask.html)

Code repository: <https://github.com/HiDiHlabs/ssam-lite>

Documentation on "Read the Docs": <https://ssam-lite.readthedocs.io/en/release/>

Validierungsdaten: <https://zenodo.org/record/5517606>

Beispiel Jupyter notebooks: [https://github.com/HiDiHlabs/ssam\\_example](https://github.com/HiDiHlabs/ssam_example)

Pypi: <https://pypi.org/project/ssam/>

Data: [https://zenodo.org/record/3478502/files/supplemental\\_data\\_ssam\\_2019.zip](https://zenodo.org/record/3478502/files/supplemental_data_ssam_2019.zip)

[https://ssam.readthedocs.io/en/release/userguide/07-input\\_mask.html](https://ssam.readthedocs.io/en/release/userguide/07-input_mask.html)

<https://github.com/HiDiHlabs/sainsc>

<https://github.com/HiDiHlabs/spatialleiden>  
<https://github.com/HiDiHlabs/ovrl.py>  
<https://github.com/HiDiHlabs/sonar>

### Sonstige Aktivitäten

#### 2021:

In diesem Jahr wurden verschiedene ELIXIR-Aktivitäten ausgebaut und erweitert. Basierend auf dem erfolgreichen ELIXIR-EXCELERATE-Kurs „Single cell RNA-seq data analysis in R“ (Jeongbin Park, Naveed Ishaque) am CSC-IT Center for Science in Espoo, Finnland (<https://www.csc.fi/web/training/-/scrnaseq>) wurde Naveed Ishaque kürzlich als de.NBI-Vertreter für die neu gegründete „ELIXIR Single Cell Omics Emerging Community“ (unter der Leitung von Eija Korpelainen, CSC, Finnland) nominiert.

Zusätzlich hat sich Naveed Ishaque auch aktiv am wissenschaftlichen Austausch beteiligt und auf verschiedenen Konferenzen und Workshops Vorträge gehalten:

- Making single molecule spatial transcriptomics data analysis easy, Oxford Global Next Gen Omics: UK, 2021, London, UK
- Spatial heterogeneity of gene expression, Jahrestagung der Society for Neuroscience (SfN), 2021, Chicago, USA
- Zellsegmentierungsfreie Modellierung der räumlichen Genexpression, Oxford Global Spatial Biology US Congress, 2021, Virtuell
- Beschleunigung der Forschung mit FAIR Einzelzellendaten: räumliche Transkriptomik, ELIXIR Virtual All Hands Meeting, 2021, Virtuell
- Identifizierung von Zelltypen, ohne Zellen identifizieren zu müssen, MuDS Graduate School Seminar Series, 2020, Virtuell
- Maschinelles Lernen für die räumliche Genexpressionsanalyse, FU-Berlin, 2020, Virtuell

#### 2022:

In diesem Jahr hat Naveed Ishaque als Hauptorganisator des Spatial Biology-Tracks zum de.NBI Biohackathon im Dezember 2022 beigetragen. Diese Fachveranstaltung hatte insgesamt 63 Teilnehmer, von denen 25 Wissenschaftler\*innen und Entwickler persönlich im Leucorea-Konferenzzentrum in Lutherstadt-Wittenberg anwesend waren. Bei dieser Konferenz handelte es sich um ein internationales Event, deren Teilnehmer in den wichtigsten Forschungszentren im Bereich der räumlichen Biologie zuhause sind (Allen Brain Institute, Universität Stockholm, Helmholtz-Zentrum München, BIH).

Die Veranstaltung wurde technisch über das de.NBI-Cloud-Setup unter <https://spatialhackathon.github.io/> realisiert. Tägliche Videokonferenzen fanden statt, um die Arbeit von Vor-Ort- und Remote-Teilnehmern über verschiedene Zeitzonen hinweg zu synchronisieren. Das Programm der Veranstaltung kann auf der Hauptwebsite eingesehen werden. Die Ergebnisse sind im Github-Projekt der Veranstaltung unter <https://github.com/SpatialHackathon> verfügbar. Der Hackathon produzierte Vignetten zu neuen räumlichen Metriken, ein Programm für die weitere Untersuchung von 3D-Modellierung und initiierte eine neue Zusammenarbeit mit dem Helmholtz-Zentrum München, in deren Rahmen momentan eine Publikation erarbeitet wird.

In diesem Jahr haben sowohl der Projektleiter als auch über das Projekt finanzierte Doktoranden aktiv an verschiedenen Konferenzen und Workshops teilgenommen und Vorträge gehalten:

- Analyzing spatial transcriptomics data on a mobile phone, EASI-Genomics Annual General Meeting, 2022, Barcelona, Spain
- Cell segmentation free analysis of single molecule spatial transcriptomics data, Finnish Single Cell Club, Helsinki, Finland
- Spatial transcriptomics data analysis and its challenges, Spatial Biology EU Conference, Berlin, Germany
- Spatial transcriptomics data analysis and its challenges, Invited talk, Leiden University Medical Centre, Netherlands

- Cell segmentation free analysis of single molecule spatial transcriptomics data, EMBO Spatial Omics Workshop, Online
- Spatial Technologies, ELIXIR Single Cell Omics Annual Meeting, Uppsala, Sweden
- Scientific web programming, CORE UNIT BIOINFORMATICS von BIH/Charité Berlin, Online
- Spatial transcriptomics data analysis and its challenges, BIMSB single cell symposium 2022, Berlin

### 2023:

Im Jahr 2023 hat Naveed Ishaque als Hauptorganisator des Spatial Biology-Tracks zum de.NBI Biohackathon im Dezember 2023 beigetragen. Insgesamt nahmen 57 Wissenschaftler\*innen und Entwickler teil, von denen 25 persönlich anwesend waren. Die Teilnehmer hatten einen breiten, internationalen Hintergrund, wobei viele der wichtigsten Forschungszentren im Bereich der räumlichen Biologie vertreten waren (Allen Brain Institute, Universität Stockholm, Helmholtz-Zentrum München, BIH, A\*STAR Singapore, ETH Zurich, EMBL).

Die Veranstaltung wurde technisch über die de.NBI-Cloud-Setup unter <https://spatialhackathon.github.io/> realisiert. Tägliche Videokonferenzen fanden statt, um die Arbeit von Vor-Ort- und Remote-Teilnehmern über verschiedene Zeitzonen hinweg zu synchronisieren. Das Programm der Veranstaltung kann auf der Hauptwebsite eingesehen werden. Die Ergebnisse sind im Github-Projekt der Veranstaltung unter <https://github.com/SpatialHackathon/SpaceHack2023> verfügbar. Der Hackathon produzierte Vignetten zu neuen räumlichen Metriken, ein Programm für die weitere Untersuchung von 3D-Modellierung und initiierte eine neue Zusammenarbeit mit dem Leiden University Medical Centre und UZH, in deren Rahmen momentan eine Publikation erarbeitet wird. Artikel, die im Rahmen des BMBF SAGE-Projekts entstanden sind, wurden auf dem Open-Access-Preprint-Server bioRxiv eingereicht.

In diesem Jahr haben Projektleiter und Doktoranden aktiv an verschiedenen Konferenzen und Workshops teilgenommen und Vorträge gehalten:

- Eingeladener Sprecher am DKFZ, November 2023 in Heidelberg
- Eingeladener Sprecher auf dem Meeting Spatial Omics UK, im Oktober 2023 in London
- Eingeladener Sprecher auf dem Kongress BioTechX Europe, Oktober 2023 in Basel
- Eingeladener Sprecher beim de.NBI Cloud User Meeting, November 2023 in Bielefeld
- Eingeladener Sprecher beim EASI Genomics Advanced Genomics Meeting, Mai 2023 in Berlin
- Eingeladener Sprecher beim ELIXIER All Hands Meeting, März 2023 in Dublin

### 2024/ Januar 2025:

In diesem Jahr haben sowohl der Projektleiter als auch der Doktorand aktiv an verschiedenen Konferenzen und Workshops teilgenommen und Vorträge gehalten:

#### Projektleiter:

- Eingeladener Sprecher bei der Arbeitsgruppe Schlesner, Januar 2024 in Augsburg
- Eingeladener Sprecher bei der Arbeitsgruppe Enders, Januar 2024 in Zürich
- Eingeladener Sprecher im Limerick Digital Cancer Research Center im Mai 2024 in Limerick
- Eingeladener Sprecher bei Remedy Bioscience im Dezember 2024, ONLINE
- Teilnahme ELIXIER All Hands Meeting im Juni 2024 in Uppsala
- Teilnahme BMBF CLS Meeting im September 2024 in Bonn
- Teilnahme Drug Development Innovation Forum im November 2024, Teilnahme ONLINE
- Teilnahme BioTechX im Oktober 2024, in Basel
- Teilnahme NextGenOmics im Oktober 2024 in London
- Teilnahme deNBI IndustrieForum im November 2024, Teilnahme ONLINE

### **Doktorand:**

- Vorstellung des Ovrply-Packages auf dem ComLS-Statusseminar in Bonn
- Teilnahme ELIXIER All Hands Meeting im Juni 2024 in Uppsala
- Teilnahme BMBF CLS Meeting im September 2024 in Bonn
- Teilnahme Drug Development Innovation Forum im November 2024, Teilnahme ONLINE
- Teilnahme BioTechX im Oktober 2024, in Basel
- Teilnahme NextGenOmics im Oktober 204 in London
- Teilnahme deNBI IndustrieForum im November 2024, Teilnahme ONLINE

Folgende Veranstaltungen wurden im Jahr 2024 und Januar 2025 organisiert im Rahmen von SAGE:

#### **1. Germany Elixier Space Hack 3.0 vom 9-12.12.2024 in Berlin**

Dieser Hackathon war -wie bereits die Vorgängerveranstaltungen- ein großer Erfolg, was sich auch daran gezeigt hat, dass es bereits nach 3 Wochen des Freischaltens der Website Überbuchungen gab, so dass es auch Interessenten gab, die nicht in Präsenz vor Ort, sondern nur online teilnehmen konnten. Insgesamt nahmen bei dieser Hybrid-Veranstaltung 30 Teilnehmer in Präsenz teil; es gab weitere 20 Teilnehmer, die nur online zugeschaltet waren.

Tägliche Videokonferenzen fanden statt, um die Arbeit von Vor-Ort- und Remote-Teilnehmern über verschiedene Zeitzonen hinweg zu synchronisieren. Das Programm der Veranstaltung kann auf der Hauptwebsite eingesehen werden. Die Ergebnisse sind im Github-Projekt der Veranstaltung unter <https://github.com/SpatialHackathon> verfügbar.

Zu den wichtigsten Ergebnissen des SpaceHack 3.0-Hackathons gehören (i) die internationale Vernetzung zwischen wichtigen Akteuren im Bereich der räumlichen Transkriptomik, (ii) wichtige Diskussionen über die Qualitätssicherung von Daten, die insbesondere bei der Charakterisierung von Tumorzellen von Bedeutung ist, (iii) große Fortschritte bei mehreren Projekten, darunter Qualitätskontrollwerkzeuge, Datendepots, Benutzerhandbücher und Deep-Learning-Analysen, und (iv) die Einbeziehung von Industriepartnern, die auf wichtige Probleme der Gemeinschaft und neue Projekte aufmerksam gemacht wurden.

#### **2. Fortbildungskurs „EU ELIXIER Spatial Omics Data Analysis“ vom 21.-24.01.2025 in Lausanne, Schweiz ([https://www.sib.swiss/training/course/20250121\\_SPODA](https://www.sib.swiss/training/course/20250121_SPODA)).**

Der Kurs „SpatialOmicsDataAnalysis“ war der zweite große Workshop, in dem Nutzer aus ganz Europa in der Analyse räumlicher Transkriptomikdaten geschult wurden, wobei ein wesentlicher Teil den segmentierungsfreien Analyseansätzen gewidmet war, die vom SAGE-Projekt angeführt werden. Dazu gehörte ein Vortrag des Projektleiters, der eine Einführung in bildgebungsbasierte räumliche Transkriptomikdaten und die Probleme der Zellsegmentierung gab und erläuterte, wie zellsegmentierungsfreie Ansätze diese Probleme lösen können. Es folgte ein praktischer Teil, bei dem die Teilnehmer unser Sainsc-Tool zur Analyse eines Krebsdatensatzes verwendeten. Dies führte dazu, dass die Gesamtleistung des SAGE-Projekts in ganz Europa bekannter wurde und unser ursprüngliches Ziel, das Tool in Deutschland bekannt zu machen, übertroffen wurde.