

Sachbericht zum Verwendungsnachweis

SIMONA

(FKZ 031L0263B)

Gefördert vom Bundesministerium für Bildung und Forschung

Teil I: Kurze Darstellung

Ursprüngliche Aufgabenstellung sowie der wissenschaftliche und technische Stand an den angeknüpft wurde

Ziel des SIMONA-Projekts war die Entwicklung eines Machine-Learning-Frameworks zur Analyse multimodaler Einzelzell-Omics-Daten über verschiedene Krankheitszustände hinweg, mit besonderem Fokus auf die Tumorerogenität beim B-Zell-Non-Hodgkin-Lymphom (B-NHL).

Im Zentrum stand die Entwicklung eines Faktorenmodells, das es ermöglicht, komplexe patientenspezifische Heterogenität – wie etwa intratumorale Diversität und Therapieresistenz – abzubilden und mit klinischen Variablen zu verknüpfen.

Zum Projektbeginn standen keine geeigneten maschinellen Lernverfahren zur Verfügung, die in der Lage gewesen wären, multimodale Einzelzellendaten zuverlässig zu integrieren. Insbesondere fehlten Modelle, die sowohl nicht-gaußverteilte Datenstrukturen als auch komplexe biologische Zustände und Krankheitskontexte adäquat berücksichtigen konnten.

Ablauf des Vorhabens

Im ersten Projektjahr (2021–2022) untersuchten wir verschiedene Methoden zur Clusteranalyse von Lymphozyten-Subpopulationen basierend auf 3'-Einzelzell-RNA- und Epitopexpression (CITE-seq). Unsere Voranalysen identifizierten über 20 verschiedene T- und B-Zell-Subtypen, deren Häufigkeit sich stark zwischen den unterschiedlichen Tumorentitäten unterschied. Zudem konzentrierten wir uns auf die Analyse von Gewebeschnitten mittels multiplexer Fluoreszenzfärbung (CODEX/Akoya) und entwickelten eine automatisierte Pipeline zur Detektion und Quantifizierung der Proteinexpression in über 3 Millionen Einzelzellen. Auch die Integration von CITE-seq- und CODEX-Daten über gemeinsame Proteinmarker wurde in Erwägung gezogen. Ein entsprechendes statistisches Modell wurde konzipiert, dessen Entwicklung jedoch aufgrund technischer Herausforderungen bei der Angleichung der Expressionsprofile beider Plattformen nicht weiterverfolgt wurde.

Im zweiten Projektjahr (2022–2023) schlossen wir die detaillierte Analyse der T-Zellen in unseren CITE-seq- und CODEX-Datensätzen ab. Die daraus resultierenden Erkenntnisse zu molekularen Veränderungen in tumorassoziierten T-Zellen wurden auf BioRxiv veröffentlicht und zur Begutachtung in einem Fachjournal eingereicht. Parallel bereiteten wir eine vergleichbare Analyse für B-Zellen vor. Darüber hinaus adressierten wir logistische Herausforderungen wie Datenmanagement und -weitergabe an Kooperationspartner und entwickelten eine Software zur Handhabung und Online-Visualisierung von CODEX-Daten. Zusätzlich wurden probabilistische Faktormodelle zur Analyse von Einzelzell-RNA-Sequenzdaten unter verschiedenen experimentellen Bedingungen entwickelt. Diese Modelle unterscheiden sich von konventionellen Ansätzen, indem sie Gruppenzugehörigkeit auf Zellebene berücksichtigen und die Zähldaten adäquat modellieren.

Im dritten Projektjahr (2023–2024) finalisierten wir das Manuskript zu den Reifungsstadien von B-Zellen bei B-NHL, veröffentlichten es auf BioRxiv und reichten es zur Veröffentlichung ein. Unser probabilistisches Faktormodell zur Modellierung von Einzelzell-Daten über mehrere Bedingungen hinweg wurde zunächst auf simulierten In-silico-Datensätzen intensiv getestet und anschließend auf reale multi-condition-Datensätze erfolgreich angewendet. Zudem entwickelten wir Softwaretools zur Vorverarbeitung und Analyse von CODEX-Daten. Für beide Projekte wurden Manuskripte verfasst, die sich zum Zeitpunkt dieses Berichts in der finalen Einreichungsphase bei BioRxiv befinden. Darüber hinaus initiierten wir ein Anschlussprojekt, das auf den entwickelten Prinzipien aufbaut, um hochdimensionale Einzelzell-Omics-Daten mit mehreren Modalitäten und über unterschiedliche experimentelle Bedingungen hinweg simultan zu modellieren.

Wesentliche Ergebnisse sowie Zusammenarbeit mit anderen Stellen

Die wichtigsten biologischen Ergebnisse unserer Studien umfassen eine umfassende Analyse von T- und B-Zell-Subtypen bei B-Zell-Non-Hodgkin-Lymphomen (B-NHL), die Entwicklung fundierter Ansätze zur linearen Modellierung hochdimensionaler, multi-konditioneller Datensätze, sowie die Entwicklung von Software zur Verarbeitung und Analyse hochgradig multiplexierter Bildgebungsdaten.

Analyse von T-Zell-Subtypen bei B-NHL (Projekt B führend): In unserer Untersuchung analysierten wir diffuses großzelliges B-Zell-Lymphom, Mantelzelllymphom, folliculäres und marginalzonenassoziiertes Lymphom. Mithilfe eines multimodalen Ansatzes – bestehend aus Transkriptom- und Epitope-Profilierung, T-Zellrezeptor-Sequenzierung, Durchflusszytometrie und multiplexierter Immunfluoreszenz – wurden 101 Lymphknotenproben von B-NHL-Patienten und gesunden Kontrollen untersucht. Dadurch konnten wir signifikante quantitative und räumliche Unterschiede im T-Zell-Mikromilieu zwischen den Subtypen und innerhalb einzelner Lymphome identifizieren. Die T-Zell-Phänotypisierung zeigte insbesondere Unterschiede in der Häufigkeit spezifischer T-Zell-Populationen, wie PD1⁺ TCF7⁻ zytotoxische T-Zellen, T-Follikel-Helferzellen und IKZF3⁺ regulatorische T-Zellen. Die Analyse der T-Zellrezeptor-Sequenzen wies auf eine klonale Expansion als zugrunde liegenden Mechanismus hin. Zudem war eine erhöhte Präsenz von PD1⁺ TCF7⁻ T-Zellen mit einer ungünstigeren Prognose assoziiert. Insgesamt liefert diese Studie eine detaillierte Referenzkarte infiltrierender T-Zellen bei B-NHL und stellt damit eine wertvolle Ressource für die weitere Forschung zur Lymphombiologie und Prognose dar.

Analyse von B-Zell-Reifungsstadien im B-NHL (Projekt B führend): Wir untersuchten die Reifungsprozesse von B-Zellen im Kontext des B-Zell-Non-Hodgkin-Lymphoms (B-NHL) und konnten zeigen, dass die intratumorale Heterogenität nicht ausschließlich durch onkogenvermittelte Mechanismen entsteht. Vielmehr trägt auch der fortschreitende Differenzierungsprozess der B-Lymphozyten wesentlich zur intratumoralen Diversität bei. Auf Basis von Einzelzell-Multiomik-Daten sowie räumlichen Gewebeanalysen von Lymphomen aus 51 Patient:innen analysierten wir verschiedene B-NHL-Subtypen sowie reaktive Lymphknoten. Unsere Ergebnisse zeigen, dass innerhalb eines Tumors unterschiedliche Reifungsstadien von B-Zellen koexistieren, die auf einen gemeinsamen Zellklon zurückzuführen sind. Dies weist auf kontinuierlich ablaufende Differenzierungsprozesse innerhalb des malignen Zellpools hin. Die Zusammensetzung dieser Reifungsstadien variierte sowohl zwischen den Subtypen als auch zwischen einzelnen Patient:innen. Besonders bemerkenswert war, dass DLBCL-Tumoren (diffus großzelliges B-Zell-Lymphom) gemischte klinische Subtypen der Zellherkunft aufwiesen.

Unsere räumlichen Analysen zeigten zudem, dass die unterschiedlichen B-Zell-Reifungsstadien distinkte Immunmikromilieus innerhalb des Tumors besiedeln und dort spezifische zelluläre Interaktionen aufrechterhalten. Diese Befunde unterstreichen die hohe Plastizität der zellulären Differenzierung im Tumorkontext und ihre zentrale Rolle in der Tumorevolution. Unsere Studie stellt damit die konventionelle Annahme infrage, dass Differenzierung und intratumorale Heterogenität getrennte Phänomene seien. Vielmehr zeigen sich beide als eng miteinander verknüpfte Prozesse im Verlauf der Lymphomentwicklung.

Methodische Entwicklungen (Projekt A führend): Auf methodischer Ebene entwickelten wir neue Ansätze zur integrativen Analyse hochdimensionaler, multi-konditioneller Einzelzell-Transkriptomdaten. Verfahren zur Dimensionsreduktion sind essenziell, um komplexe Genexpressionsprofile in handhabbare, niedrigdimensionale Repräsentationen zu überführen. Die Anwendung linearer Faktorisierungsverfahren wird jedoch durch Variablen wie Batch-Effekte oder experimentelle Störungen erschwert, die zusätzliche Varianzquellen darstellen. Um dies zu adressieren, entwickelten wir single-cell PCA (scPCA), ein Faktorisierungsframework zur gemeinsamen Modellierung zellulärer Heterogenität und konditioneller Variablen in scRNA-seq-Daten. Die Anwendung von scPCA auf stimulierte und unstimulierte PBMCs sowie Einzelzell-Zellliniendaten zeigte, dass scPCA interpretierbare, PCA-ähnliche latente Repräsentationen erzeugt und Batch-Effekte effektiv ausgleicht. Die Analyse alternder Lungenzellen bei Nagern offenbarte eine bislang unbekannte Hochregulation von *Ccl5* in T-Zellen. Darüber hinaus identifizierte scPCA koordinierte Transkriptionsänderungen in Neuronen über verschiedene Zeitpunkte hinweg sowie genregulatorische Verschiebungen im Zebrafisch-Modell nach *chordin*-Knockout, selbst bei stark unterschiedlichen Zellhäufigkeiten. Zusammenfassend stellt scPCA ein leistungsstarkes Werkzeug zur Integration und Interpretation multi-konditioneller Einzelzellendaten dar und liefert wichtige Einblicke in komplexe biologische Prozesse.

Zusammenarbeit mit anderen

Während der SIMONA-Studie arbeiteten die Arbeitsgruppen von Prof. Wolfgang Huber und Prof. Sascha Dietrich eng mit mehreren renommierten Forschungsteams zusammen:

1. Nolan-Gruppe (Stanford University): Generierung und Analyse von CODEX-Bilddaten.
2. Seifert-Gruppe (Universitätsklinikum Essen): Bereitstellung und Analyse von FACS-Daten zur Validierung der T- und B-Zell-Studien.
3. Weigert-Gruppe (Universität München): Durchführung von DNA-Sequenzierungsexperimenten.
4. Zaugg-Gruppe (EMBL Heidelberg): Analyse der Transkriptionsfaktoraktivitäten.

Präsentationen und Konferenzen

1. Die Ergebnisse des Projekts wurden auf mehreren internationalen Konferenzen vorgestellt:
2. Donnacha Fitzgerald präsentierte die B-Zell-Studie auf: Jahrestagung der American Society of Hematology (ASH) 2022), dem EHA-Kongress 2023, der Cold Spring Harbor Laboratory Biology of Genomes Conference 2024

3. Tobias Roeder stellte die Ergebnisse der T-Zell-Studie beim EHA-Kongress 2020 vor.
4. Harald Vöhringer präsentierte die methodischen Entwicklungen zur scPCA auf der Cold Spring Harbor Biological Data Science Conference 2022.

Im Rahmen der Maßnahmen zur Verbreitung der Projektergebnisse organisierten die Gruppen von Huber und Stegle den EMBO-Workshop „Integrative Analyse von Multi-Omics-Daten“ im September 2022 und Februar 2024. Der Workshop bot einen umfassenden Überblick über fortgeschrittene Anwendungen von Multi-Omics-Methoden, vermittelte Kriterien zur Auswahl geeigneter Analyseverfahren und ermöglichte den Teilnehmenden durch praxisnahe Übungen einen direkten Zugang zu methodischem Know-how.

Teil II: Eingehende Darstellung

Eingehende Darstellung der im Rahmen des Vorhabens durchgeführten Arbeiten, insbesondere im Vergleich zur ursprünglichen Vorhabenbeschreibung

Multimodales und räumlich aufgelöstes Profiling identifiziert unterschiedliche Muster der T-Zell-Infiltration bei nodalen B-Zell-Lymphomen

Wir nutzten die Einzelzellsequenzierungstechnologie CITE-seq, um 14 verschiedene T-Zell-Subtypen innerhalb unseres großen B-NHL-Datensatzes zu identifizieren. Dazu gehörten verschiedene Arten von Helfer-T-Zellen, regulatorischen T-Zellen (Tregs) und zytotoxischen T-Zellen (Ttox). Unter diesen zeigte sich eine signifikante klonale Expansion von PD1+ TIM3+ zytotoxischen T-Zellen (Ttox EM3) bei diffus großzelligem B-Zell-Lymphom (DLBCL), follikulärem Lymphom (FL) und Marginalzonenlymphom (MZL). Zudem stellten wir fest, dass T-follikuläre Helferzellen (TFH) und IKZF3+ Tregs speziell in FL und MZL klonal expandiert waren, was auf ihre Mitwirkung am Krankheitsgeschehen hinweist.

Unsere Studie untersuchte auch die Erschöpfung von T-Zellen und deren Auswirkungen auf die Prognose durch Trajektorien- und Pseudotime-Analysen. Wir fanden heraus, dass terminal erschöpfte T-Zellen, gekennzeichnet durch niedrige TCF7-Aktivität und hohe Expression inhibitorischer Rezeptoren wie TIM3, in DLBCL und FL am häufigsten vorkamen. Ein höherer Anteil dieser erschöpften T-Zellen korrelierte mit einer schlechteren progressionsfreien Überlebensrate, was darauf hindeutet, dass die Erschöpfung von T-Zellen ein kritischer Faktor für den Krankheitsverlauf ist.

Mittels CODEX-Bildgebung kartierten wir die räumliche Verteilung von Immunzellen innerhalb des Tumormikromilieus und identifizierten spezifische „Nachbarschaften“ in den Lymphknoten, die mit unterschiedlichen Immunzelltypen angereichert waren. Beispielsweise zeigten DLBCL-Proben eine diffuse Präsenz von Bereichen mit erschöpften T-Zellen und Makrophagen, die eng mit Tumorzellen assoziiert waren. Im Gegensatz dazu wiesen FL-Proben ausgeweitete Keimzentren mit klonalen TFH-Zellen und Tregs auf, umgeben von Bereichen mit Gedächtnis- und erschöpften T-Zellen.

Relevanz für die Immuntherapie

Unsere Ergebnisse haben wichtige Implikationen für die Immuntherapie bei B-NHL. Die hohe Präsenz terminal erschöpfter T-Zellen in bestimmten Subtypen könnte darauf hinweisen, dass hier immuntherapeutische Ansätze (z.b.: bi-spezifische Antikörper) nur schlecht wirksam sind. Patienten mit einem höheren Anteil an PD1+ TIM3- Ttox-Zellen, die eine gewisse

Selbsterneuerungskapazität beibehalten, könnten dagegen stärker von kombinierten Immuntherapien, einschließlich T-Zell-aktivierenden Ansätzen, profitieren.

B-Zell-Differenzierung bleibt trotz maligner Transformation in Lymphomen erhalten: Erkenntnisse aus einem Einzelzell-Multiomic-Atlas

Unsere integrative Einzelzell-Analyse (Multiomics- und räumlich aufgelöst) verschiedener B-Zell-Lymphomentitäten belegt, dass Differenzierungsprozesse der B-Zellen eine wesentliche Quelle intratumoraler Heterogenität darstellen. Wir verwendeten CITE-seq zur Untersuchung der B-Zell-Heterogenität bei B-NHL. Unsere Ergebnisse zeigen, dass die B-Zell-Reifung innerhalb von Tumoren kein linearer Prozess ist. Stattdessen koexistieren innerhalb eines Tumors mehrere Reifungszustände, die vom gleichen klonalen Ursprung abstammen. Dies legt nahe, dass B-Zellen auch nach ihrer malignen Transformation flexibel bleiben können.

Unsere Bilddatenanalyse zeigte, dass diese Reifungszustände keine zufällige räumliche Verteilung aufweisen, sondern spezifische Nischen innerhalb des Tumors besetzen. Diese ähneln normalen Reifungsumgebungen wie Keimzentren, sind jedoch bei malignem B-NHL häufig gestört oder umorganisiert.

Zudem fanden wir heraus, dass bestimmte genetische Varianten, wie Kopienzahlveränderungen (CNVs), mit spezifischen Reifungszuständen assoziiert sind. Das bedeutet, dass diese genetischen Veränderungen entweder bestimmte Differenzierungspfade fördern oder deren Folge sein können; ein zusätzlicher Komplexitätsfaktor in der Tumorbilogie von B-NHL.

Diese Erkenntnisse widersprechen der herkömmlichen Vorstellung, dass der Ursprung einer Tumorzelle eine fixe Reifungsstufe vorgibt. Stattdessen ist die Differenzierung ein dynamischer Prozess, der zur Tumorevolution und Resistenzentwicklung beiträgt.

Gemeinsames Modellieren zellulärer Heterogenität und Bedingungeffekten mit scPCA bei Einzelzell-RNA-Seq

Unser flexibler Faktorisierungsansatz scPCA erlaubt die gleichzeitige Modellierung zellulärer Heterogenität und experimenteller Bedingungen. Aufbauend auf PCA berücksichtigt scPCA Vorinformationen über jede Zelle und ermöglicht so eine präzisere Analyse mehrbedingter scRNA-seq-Daten.

Im Gegensatz zu komplexen, schwer interpretierbaren nicht-linearen Modellen trennt scPCA diese Effekte auf lineare Weise – mit hoher Interpretierbarkeit. Wir validierten scPCA mit verschiedenen Datensätzen (z. B. PBMCs, Lungenzellen alternder Mäuse, Zebrafisch-Embryonen) und zeigten, dass scPCA zuverlässig Bedingungeffekte von zellulärer

Variabilität trennt.

Eingehende Darstellung der wichtigsten Positionen des zahlenmäßigen Nachweises.

Der Zahlungsnachweis wurde separat eingereicht. Die bereitgestellten Mittel für Personal, Verbrauchsmaterial und Reisen wurden ausgeschöpft.

Eingehende Darstellung der Notwendigkeit und Angemessenheit der geleisteten Projektarbeit

Die im Projekt SIMONA geleistete Arbeit adressierte eine zentrale methodische und inhaltliche Lücke in der modernen Einzelzellforschung: die fehlende Verfügbarkeit geeigneter Machine-Learning-Ansätze zur Analyse hochdimensionaler, multimodaler Einzelzell-Omics-Daten über verschiedene Krankheitszustände hinweg. Insbesondere im Kontext hämatologischer Neoplasien, wie dem B-Zell-Non-Hodgkin-Lymphom (B-NHL), besteht ein dringender Bedarf an computergestützten Methoden, die sowohl genetische als auch phänotypische Heterogenität erfassen und klinisch interpretierbar machen.

Die im Rahmen von SIMONA entwickelte Faktormodell-basierte Analyseplattform erfüllt genau diese Anforderung. Sie ermöglicht es, patientenspezifische Heterogenität – einschließlich intratumoraler Differenzierung und Therapieresistenz – differenziert zu modellieren und mit klinischen Parametern in Beziehung zu setzen. Die entwickelte Analyse-Pipeline stellt dabei eine integrative bioinformatische Infrastruktur dar, die auf reale klinische Fragestellungen anwendbar ist und den Weg für translationale Anwendungen in der Präzisionsmedizin ebnet.

Die Projektarbeit erwies sich nicht nur als methodisch innovativ, sondern auch als inhaltlich hochrelevant: Durch die Kombination von Einzelzell-Multiomics-Daten mit räumlicher Information konnte demonstriert werden, dass B-Zell-Differenzierung ein wesentlicher Treiber intratumoraler Heterogenität ist, eine Erkenntnis, die direkte Implikationen für Diagnostik, Prognose und Therapieentwicklung im Bereich der Lymphomforschung hat.

Angesichts der initialen methodischen Lücke, der hohen klinischen Relevanz der Fragestellung sowie der erfolgreichen Umsetzung innovativer Analyseverfahren ist die geleistete Projektarbeit als angemessen zu betrachten.

Eingehende Darstellung des voraussichtlichen Nutzens, insbesondere der Verwertbarkeit des Ergebnisses

Wir identifizierten prognostische Marker wie PD1⁺ TIM3⁺ T-Zellen, die mit einem ungünstigen Überleben assoziiert sind, und kartierten verschiedene B-Zell-Reifungszustände in Verbindung mit spezifischen genetischen Veränderungen.

Diese molekularen und immunologischen Signaturen liefern wertvolle Ansatzpunkte für die zukünftige Entwicklung personalisierter Therapieansätze, bei denen die intratumorale Heterogenität gezielt berücksichtigt werden kann.

Unsere Erkenntnisse könnten auch die Immuntherapie verbessern: Die hohe Zahl erschöpfter T-Zellen in Subtypen wie DLBCL und FL erklärt deren Therapieresistenz. Zudem identifizierten wir therapeutisch relevante Zellpopulationen (z. B. klonale TFH-Zellen und IKZF3⁺ Tregs), die als Zielstrukturen für neue Behandlungsansätze dienen könnten.

Die scPCA-Methode wiederum ist ein bedeutender Fortschritt für die Einzelzellbiologie, da sie interpretierbare Dimensionsreduktion über mehrere Bedingungen hinweg ermöglicht – besonders relevant für Krebs- und Immunforschung.

Für die Zukunft planen wir:

Die Validierung unserer Erkenntnisse zu T-Zell- und B-Zell-Zuständen mittels groß angelegter Proteomikdaten.

Die Weiterentwicklung von scPCA zur Analyse multimodaler Einzelzelldaten.

Eingehende Darstellung des während der Durchführung dem ZE bekannt gewordenen Fortschritts bei anderen Stellen

Nicht bekannt.

Eingehende Darstellung der erfolgten oder geplanten Veröffentlichungen

Die Ergebnisse wurden veröffentlicht in:

Nature Cell Biology: Roider et al. (2024)

BioRxiv: Fitzgerald et al. (2022)

Außerdem ist das scPCA-Paket öffentlich zugänglich auf:

PyPI & GitHub: <https://github.com/sagar87/scPCA>

Dokumentation: <https://sagar87.github.io/scPCA>