

Umfangreiche medizinische Datensätze verstehen: ScPoli ermöglicht Multiskalenrepräsentation von Zellen und Proben

Nur die korrekte Analyse und Interpretation der stetig wachsenden Menge an medizinischen Daten kann zu wissenschaftlichen Durchbrüchen und essenziellen Therapien für Patient:innen verhelfen. Computerwissenschaftler:innen von Helmholtz Munich haben ein generatives Modell namens scPoli (single-cell population level integration) entwickelt, um die Datenintegration von hochauflösenden Einzelzell-Datensätzen durchzuführen. So können Referenzkarten des menschlichen Körpers auf Einzelzellebene, sogenannte Einzelzell-Atlanten, für die medizinische Forschung erstellt werden.

Mit Hilfe dieser Atlanten werden Daten schneller und detailreicher interpretiert, was schließlich zu neuen biologischen Erkenntnissen und einem besseren Verständnis von Krankheiten führt. Das Modell wurde jetzt in Nature Methods vorgestellt.

Seit einigen Jahren erleben wir einen immensen Anstieg sowohl in der Menge als auch in der Komplexität von aufgenommenen Daten – insbesondere im Bereich der medizinischen Forschung. Wissenschaftler:innen sind nun in der Lage, Gewebe und Organe in faszinierender Präzision zu erfassen, und das bis hin zur Ebene einzelner Zellen. Die Zusammenlegung der resultierenden Datensätze hat zur Erstellung sogenannter Einzelzell-Atlanten geführt, also Referenzkarten jeder Zelle in einem bestimmten Organ. In Zukunft soll auf Basis dieser Atlanten eine zusammenhängende Karte des gesamten menschlichen Körpers entstehen. Die hochauflösenden und umfangreichen Datensätze ermöglichen nicht nur neue biologische Erkenntnisse über die zelluläre Vielfalt bestimmter Gewebe, sondern beschleunigen auch zeitaufwändige Schritte während der Analyse. Mit Hilfe der Atlanten können Forschende beispielsweise Organe von gesunden Menschen mit denen von erkrankten vergleichen und dadurch wertvolle Erkenntnisse über die Entwicklung und den Verlauf von Krankheiten erhalten.

Auf Grund der zunehmenden Größe dieser Atlanten werden Methoden des maschinellen Lernens und Computeralgorithmen benötigt, um die Daten zu analysieren und zu integrieren. Prof. Fabian Theis, von Helmholtz Munich und Professor an der Technischen Universität München, sowie Dr. Mohammad Lotfollahi und Carlo De Donno vom Computational Health Center von Helmholtz Munich und haben ein neues generatives Modell namens scPoli (Abkürzung für single-cell population level integration) mit einem sogenannten Multiskalenrepräsentationsansatz für sowohl Zellen als auch Proben entwickelt, welcher Patient:innen in großen Studien entspricht. Es ist das erste Datenintegrationsmodell, das Repräsentationen sowohl für Zellen als auch für Proben erstellen kann. Mit diesem neuen generativen Modell können medizinische Datensätze einfacher analysiert werden, um die Hauptfaktoren für Variabilität zu identifizieren, während gleichzeitig die natürlichen Unterschiede berücksichtigt werden.

Das Team hat scPoli mit dem Ziel entwickelt, die Interpretierbarkeit von Einzelzell-Studien zu verbessern. Im Gegensatz zu anderen Modellen, die nur Zellrepräsentationen liefern, bietet scPoli Forschenden eine neuartige Perspektive, um Muster auf der Probenebene zu untersuchen und zu verknüpfen, wodurch Integrationsabläufe und Interpretationen verbessert werden.

Funktionalität bestätigt: Einzelzell-Atlanten lassen sich schneller erstellen

Die Forschenden von Helmholtz Munich haben bereits die Funktionalität ihres Modells unter Beweis gestellt, indem sie Daten aus zwei relevanten Einzelzell-Atlanten integriert haben. Die Integration des Human Lung Cell Atlas, einer Referenzkarte der Lunge, die kürzlich vom Team um Prof. Fabian Theis veröffentlicht wurde, zeigte sowohl eine Verbesserung der Leistung als auch neue Einblicke auf der Probenebene. Darüber hinaus wurde scPoli verwendet, um einen umfangreichen PBMC-Atlas (ein Einzelzell-Atlas der peripheren Blutmonozyten), bestehend aus 7,8 Millionen Zellen, zu integrieren. Dies zeigte die Skalierbarkeit des Modells, welche für Integrationsstudien unabdingbar ist.

Einzigartig: Das Modell zeigt Variationen zwischen Proben und Zellen

ScPoli ist im Vergleich zu früheren Bemühungen zur Datenintegration einzigartig. Es bietet verschiedene Anwendungsfälle für die Multiskalenanalyse, da Wissenschaftler:innen zum ersten Mal in der Lage sind, gleichzeitig Zell-, Proben- und Patient:innenrepräsentationen zu analysieren und die Eigenschaften und Merkmale einzelner Zellen viel detaillierter als je zuvor zu erfassen. Es wurde gezeigt, dass scPoli eine Multiskalenklassifizierung von Zellen und Proben sowie kontrollierte Datenintegrationsabläufe ermöglichen kann. Das Modell kann die Erstellung und Nutzung von Atlanten beschleunigen, was wiederum das Verständnis von Krankheiten und die Entwicklung neuartiger Therapien vorantreibt.

Originalpublikation:

De Donno et al. (2023): Population-level integration of single-cell datasets enables multi-scale analysis across samples. Nature Methods. Doi: [10.1038/s41592-023-02035-2](https://doi.org/10.1038/s41592-023-02035-2)