

Eine Suchmaschine für RNA-Moleküle

In Sekundenschnelle kann eine Google-ähnliche Suchmaschine die Daten von Millionen einzelner Zellen aus Tausenden Experimenten weltweit analysieren. Mit dem Tool lassen sich Fragen zur RNA-Biologie klären, die niemand zuvor beantworten konnte, berichtet ein Team vom Max Delbrück Center in „Nature“.

Stellen Sie sich vor, Ärzt*innen würden genau verstehen, welche Zellen zum Beispiel einen Tumor ihrer Patient*innen verursacht haben und ob dabei vielleicht auch Viren oder Bakterien im Spiel gewesen sind. Dann könnten sie die Informationen nutzen, um für jeden und jede der Erkrankten einen maßgeschneiderten Therapieplan zu erstellen. Um Fragen wie diese im Detail zu beantworten, müssen sie jedoch zunächst Tausende von Experimenten durchforsten, deren Ergebnisse in riesigen Datenbanken rund um den Globus gespeichert sind. Die Suche hätte bislang mehrere Tage in Anspruch genommen. Mindestens.

Jetzt ist ein solches Szenario ein gutes Stück aber näher gerückt. Denn Forschende des Berliner Instituts für Medizinische Systembiologie des Max Delbrück Center (MDC-BIMSB) haben eine Suchmaschine namens „Malva“ entwickelt, die solche Aufgaben radikal vereinfacht. Es sei die erste Plattform, die riesige Mengen an Einzelzelldaten allein anhand von Sequenzinformationen sekundenschnell durchforsten könne, sagt Daniel León-Periñán. León-Periñán ist ein Erstautor der Publikation in „Nature“ und Doktorand in der Arbeitsgruppe „Systembiologie von Gen-regulatorischen Elementen“ von Professor Nikolaus Rajewsky, dem Direktor des MDC-BIMSB.

„So ähnlich wie Google es vor 30 Jahren für das Internet getan hat, ermöglicht Malva es Wissenschaftler*innen und KI-Tools, innerhalb von Sekunden Informationen zu Millionen von Zellen zu durchsuchen – ohne riesige Dateien herunterladen zu müssen, ohne ein Referenzgenom zu benötigen und ohne dass tiefgreifende Fachkenntnisse in der Datenwissenschaft nötig wären“, fügt Rajewsky, der Letztautor der Studie, hinzu. „Malva verwandelt statische Transkriptom-Atlanten in dynamische Ressourcen, die unser Verständnis der RNA-Biologie vertiefen werden. Die Plattform könnte zudem entscheidend dazu beitragen, dass Forschende künftig besser verstehen, wie aus Gesundheit Krankheit wird oder wie und welche Zellen auf bestimmte medizinische Behandlungen reagieren.“

Unvorstellbar viele RNA-Daten

Mithilfe der Einzelzell-RNA-Sequenzierung bekommen Wissenschaftler*innen einen äußerst detaillierten Einblick in die Vorgänge im Inneren einzelner Zellen zu jedem beliebigen Zeitpunkt. In den vergangenen zehn Jahren haben sie so weltweit Terabytes an Daten gesammelt. Wer diese Daten auswerten möchte, um mehr über eine bestimmte DNA- oder RNA-Sequenz zu erfahren, steht allerdings vor zahlreichen Hürden.

Es müssten zum Beispiel unvorstellbare Mengen an Rohdaten heruntergeladen und erneut verarbeitet werden, wozu kein einzelnes Labor die Kapazitäten hat. Zudem müsste man einen Weg finden, Daten aus verschiedenen Quellen zu standardisieren. Darüber hinaus sind aufgrund der Art und Weise, wie die Daten indexiert sind, Informationen über RNA-Isoformen – also verschiedene RNA-Varianten, die von dem gleichen Gen kodiert werden – äußerst begrenzt.

Um die Aufgabe zu vereinfachen und die Bandbreite der Fragen zu erweitern, die sich anhand der Daten beantworten lassen, haben León-Periñán und Dr. Nikos Karaiskos, ebenfalls Erstautor aus der Gruppe von Rajewsky, öffentlich zugängliche Daten aus den Repositorien neu aufbereitet – sodass man sie nun nach bestimmten Nukleotidsequenzen durchsuchen kann. Malva indexiert zudem räumliche Daten und gibt dadurch auch Auskunft darüber, an welcher Stelle in einem Gewebeschnitt sich eine bestimmte RNA befindet.

Die Suchmaschine ist so konzipiert, dass man sie kontinuierlich erweitern kann: Sobald neue Einzelzell-RNA-Daten in der Fachliteratur veröffentlicht sind, werden diese auf einen Server am Max Delbrück Center heruntergeladen, verarbeitet und den vorhandenen Daten hinzugefügt.

Ein breites Anwendungsspektrum

Es gebe unzählige Anwendungsszenarien, sagt Karaiskos. „Sie reichen von ganz simplen Fragen wie ‚In welchem Zelltyp wird dieses spezielle Gen exprimiert?‘ bis hin zu weitaus komplexeren.“ Solch einfache Fragen könne man zwar auch mit anderen Plattformen beantworten, fügt Karaiskos hinzu. „Mit ihnen lassen sich jedoch zum Beispiel keine Fragen zur RNA-Biologie klären, da sie RNA-Isoformen als Einheit behandeln und nicht einzeln einem Referenzgen zugeordnen.“ So könne man nicht herausfinden, ob eine bestimmte RNA-Isoform in einem bestimmten Zelltyp exprimiert werde, erklärt der Forscher. „Die Flexibilität, in Malva nach RNA-Sequenzen zu suchen, ermöglicht es uns, Fragen anhand dieser Daten zu beantworten, die zuvor unmöglich schienen.“

Da andere Plattformen ihre Daten zudem auf mithilfe eines Referenzgenom abbilden, das aus wenigen menschlichen DNA-Proben zusammengesetzt ist, finden Forschende, die nach Informationen über RNA suchen, die aus seltenen oder einzigartigen Genvarianten entsteht, möglicherweise nicht viel. Malva erlaubt den Zugriff auf einen wesentlich vielfältigeren RNA-Datensatz. Die neue Plattform umfasst außerdem Sequenzinformationen verschiedener Spezies, darunter Bakterien, Viren und Pilze. So können Wissenschaftler*innen untersuchen, wie diese Mikroorganismen menschliche Zellen beeinflussen und Krankheiten verursachen.

Derzeit können alle Forschenden Malva kostenlos nutzen. Rajewsky und sein Team bereiten allerdings die Gründung eines Start-ups vor, um Malva kommerziell verfügbar zu machen. Ein Patent auf die Technologie ist bereits angemeldet.

Max Delbrück Center

Das Max-Delbrück-Centrum für Molekulare Medizin in der Helmholtz-Gemeinschaft legt mit seinen Entdeckungen von heute den Grundstein für die Medizin von morgen. An den Standorten in Berlin-Buch, Berlin-Mitte, Heidelberg und Mannheim arbeiten unsere Forschenden interdisziplinär zusammen, um die Komplexität unterschiedlicher Krankheiten auf Systemebene zu entschlüsseln – von Molekülen und Zellen über Organe bis hin zum gesamten Organismus. In wissenschaftlichen, klinischen und industriellen Partnerschaften sowie in globalen Netzwerken arbeiten wir gemeinsam daran, biologische Erkenntnisse in praxisnahe Anwendungen zu überführen – mit dem Ziel, Frühindikatoren für Krankheiten zu identifizieren, personalisierte Behandlungen zu entwickeln und letztlich Krankheiten vorzubeugen. Das Max Delbrück Center wurde 1992 gegründet und vereint heute eine vielfältige Belegschaft mit rund 1.800 Menschen aus mehr als 70 Ländern. Wir werden zu 90 Prozent durch den Bund und zu 10 Prozent durch das Land Berlin finanziert.

Originalpublikation:

Daniel León-Periñán, Nikos Karaiskos, Nikolaus Rajewsky (2026): „Ultrafast and reference-free sequence discovery in single-cell data“. Nature, DOI: 10.1038/s41586-026-10975-w

Weitere Informationen:

<https://www.mdc-berlin.de/de/n-rajewsky> – AG N. Rajewsky

<https://malva.mdc-berlin.de/> – Malva Website