



#Lungenforschung

## Die Vermessung der Lunge

Mit dem Human Lung Cell Atlas entsteht der erste aus vielen Datensätzen integrierte Atlas seiner Art für ein großes Organ. DZL-Wissenschaftler Malte Lücken, Gruppenleiter am Institut für Lungengesundheit und Immunität bei Helmholtz München, arbeitet federführend daran mit. Im Interview berichtet er, wie KI-Methoden die Pionierarbeit erst möglich machen.

DZG-Redaktion | Jörn Bullwinkel, Juliane Gringer

Malte Lücken



Sie arbeiten an nichts Geringerem als an der Erfassung eines menschlichen Organs auf zellulärer Ebene. Wie kam es dazu, dass der Human Lung Cell Atlas entwickelt wurde?

An einem menschlichen Zellatlas wird seit etwa 2017 gearbeitet. Die Vision war, dass er uns hilft zu verstehen, welche Gene in welchen Zellen des Körpers eines gesunden Menschen aktiv sind und wie diese sich zwischen verschiedenen Bevölkerungsgruppen unterscheiden. Zunächst fehlten uns ausreichend Daten, doch dann hat die COVID-Pandemie die internationale Zusammenarbeit vorangetrieben: Viele Daten wurden in einem Projekt

verschmolzen und wir konnten tatsächlich zeigen, welche Zellen tief in der Lunge die notwendigen Bauteile enthalten, um sich als Erstes mit Corona zu infizieren. Danach haben wir die Zusammenarbeit unabhängig von der Pandemie fortgesetzt und eine zelluläre Darstellung der menschlichen Lunge über anfangs 100 Individuen ausgearbeitet. Inzwischen haben wir das Modell auf 400 Individuen erweitert.

**DAS SIND RIESIGE DATENMENGEN, IN DENEN NUR METHODEN WIE DIE DER KÜNSTLICHEN INTELLIGENZ MUSTER ERKENNEN KÖNNEN**

### **Was macht künstliche Intelligenz für Ihre Forschung möglich?**

Als ich meinen ersten Einzelzell-Datensatz analysiert habe, hat es rund sechs Monate gedauert, bis ich sagen konnte, welche Zelltypen enthalten sind. Dank Human Lung Cell Atlas und KI-Methoden wissen wir jetzt innerhalb weniger Stunden, wo der Unterschied eines neuen Datensatzes zur gesunden Lunge liegt und auf welche Zelltypen wir uns fokussieren müssen. Wo wir vor einigen Jahren noch bei einer einzigen Messung pro Individuum waren, sind wir heute bei 20.000 Genen in je 5.000 Zellen pro Messung von je 400 bis 500 Individuen. Das sind riesige Datenmengen, in denen nur Methoden wie die der künstlichen Intelligenz Muster erkennen können. Wir nutzen sie, um die Daten zu erklären oder auch zu modellieren.

### **Vor welchen Herausforderungen stehen Sie dabei?**

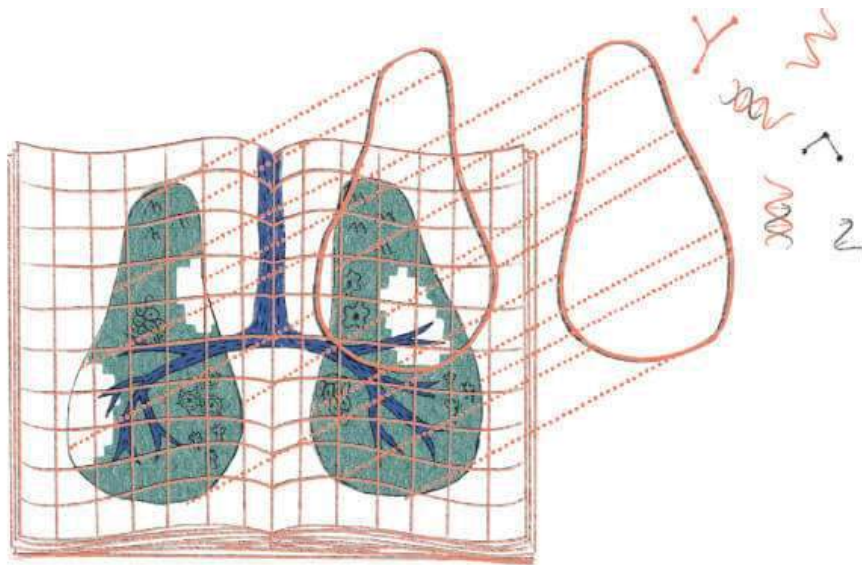
Anfangs wollten wir 14 Datensätze aus zwölf Studien nutzen, um eine Lunge abzubilden. Wir konnten diese aber nicht einfach zusammenbringen, denn jeder Datensatz wurde ja für einen bestimmten Zweck gewonnen, mit unterschiedlichen Technologien gemessen und verschiedenen Protokollen erfasst. Dadurch unterscheiden sich die Daten strukturell. KI-Methoden können diese Effekte herausrechnen – und genau daran arbeiten wir als Gruppe seit fünf Jahren: Wir vergleichen die dafür geeigneten Methoden und entwickeln sie weiter.

Eine Herausforderung war auch, dass nicht alle weltweit Forschenden einen bestimmten Zelltyp gleich benennen. Wir mussten also erst mal sehen, welche Zelltypen in unseren Datensätzen enthalten sind, und sie zusammenfassen – beispielsweise Goblet-Zellen, die in der Nasenschleimhaut beheimatet sind. So konnten wir herausfinden, wo die größten Unterschiede in der Zellbenennung sind, um uns darauf bei der Datenintegration zu fokussieren. Durch das Integrieren der 14 Datensätze konnten wir diese Zellen dann letztendlich miteinander vergleichen, um einen Konsens herzustellen.

### **Welche Methoden nutzen Sie noch und wie setzen Sie sie ein?**

Für den ersten Human Lung Cell Atlas setzen wir beispielsweise einen sogenannten Conditional Variational Autoencoder (CVAE) ein. Dessen Funktionsweise ähnelt der des MP3-Formats: Wenn ich ein komplexes Musikstück als MP3-Datei sichere, wird es so komprimiert, dass nur die essenziellen Informationen gespeichert werden, die dann wieder dekomprimiert werden müssen, um es als große Sounddatei abzuspielen. Genauso benutzen wir einen CVAE, um unsere Daten von 20.000 Genen in nur wenigen Dimensionen – beispielsweise in zweidimensionalen Grafiken – darzustellen.

Lange Zeit brauchte man große Mengen an Zellmaterial, um Technologien wie die PCR anzuwenden. Dadurch erhielt man aber nur den Mittelwert aus der Summe aller Zellen. Dass sich diese Zellen individuell sehr unterscheiden, konnte erst etwa 2012 mit der Einzelzell-Analyse aufgelöst werden. Nun ging es darum, die Zahl der parallel zu messenden Zellen zu steigern. Daraus entwickelte sich die Idee, den ganzen menschlichen Körper hochauflösend zu vermessen und besser zu verstehen, wie sich Zellen im Krankheitsfall im Vergleich zum gesunden Patienten verhalten.

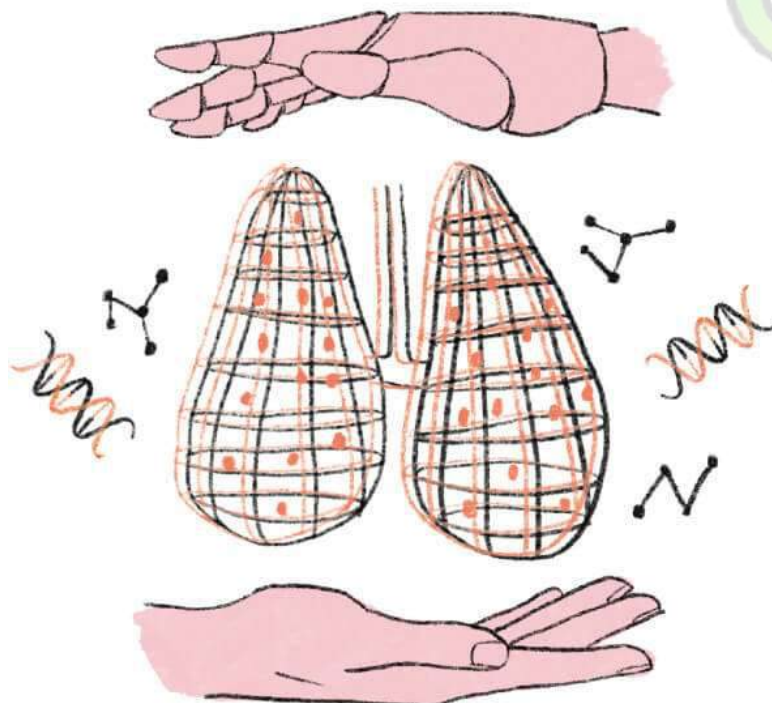


### Was kann uns der Atlas in seiner jetzigen Form bereits über die Lunge erzählen?

Wir haben ganz neue Zelltypen gefunden, die noch niemand zuvor beschrieben hat. Außerdem haben wir uns international darauf geeinigt, wie wir alle in der menschlichen Lunge vorkommenden Zelltypen charakterisieren und einheitlich benennen. Auf molekularbiologischer Ebene wissen wir nun beispielsweise, welche Gene in einer Zelle aktiv sein oder welche Proteine vorhanden sein müssen, damit ein Virus andocken kann. Dann schauen wir, welche Zelltypen vermutlich die ersten sind, die infiziert werden – was für die therapeutische Entwicklung sehr relevant ist und auch zum besseren Verständnis beiträgt, wie diese Krankheit sich verbreitet.

Wir haben auch modelliert, wie sich die Genaktivität von der Nase bis tief in die Lunge im gleichen Zelltyp verändert. Außerdem können wir modellieren, wie sich Genaktivitäten von Rauchern oder Menschen mit höherem BMI verändern. Wenn wir solche Abläufe besser verstehen, können wir daraus ableiten, wieso Krankheitsbilder mit gewissen Prozessen assoziiert sind, und sagen, in welchen Zelltypen die Prozesse gerade stattfinden. Außerdem können wir verfolgen, wie Zellen im Laufe ihrer individuellen Entwicklung bestimmte Gene an- und andere ausschalten. Wie viele Zellen sich in einem bestimmten Entwicklungsstadium befinden, kann ebenfalls für Gesundheit oder Krankheit ausschlaggebend sein.

**KI-METHODEN KÖNNTEN HIER BALD GENAUSO WICHTIG SEIN WIE EIN MRT-GERÄT.**



Es stehen immer größere Datensätze zur Verfügung, mit denen die Maschinen lernen können, um Krankheiten zu klassifizieren. Das ist der erste Schritt, vor dem Diagnostizieren. In der Bildgebung erkennt KI Veränderungen bereits sehr gut. KI-Methoden könnten hier bald genauso wichtig sein wie ein MRT-Gerät.

### **Was macht der Human Lung Cell Atlas möglich?**

Er eröffnet vor allem neue Möglichkeiten in der Früherkennung von Krankheitsprozessen. Wenn wir verstehen, was die Zellprofile gesunder Menschen ausmacht, können wir erkennen, wann sich ein Krankheitsprozess von der natürlichen Variation abhebt. Wir haben dazu weiterführende Methoden entwickelt, mit denen wir neue Datensätze mit so einem Atlas vergleichen oder sie integrieren können. Ich denke, dass es bis zur klinischen Umsetzung noch ein weiter Weg ist, aber die Möglichkeit, den automatischen Vergleich zu nutzen, gibt es jetzt. Persönlich finde ich es besonders spannend, dass wir dadurch in den Bereich der personalisierten Medizin kommen.

ALS BIOINFORMATIKER MALE ICH MIR DIE ZUKUNFT IMMER SO AUS, WIE ICH SIE GERNE HÄTTE. DOCH WAS REALISTISCH IST, HÄNGT VON VIELEN FAKTOREN AB, DIE ICH NICHT BEEINFLUSSEN KANN.

### **Was kann das alles für Patientinnen und Patienten bedeuten?**

Als Bioinformatiker male ich mir die Zukunft immer so aus, wie ich sie gerne hätte. Doch was realistisch ist, hängt von vielen Faktoren ab, die ich nicht beeinflussen kann. Dennoch motiviert es mich unglaublich, wenn ich mir vorstelle, dass wir Screening-Methoden entwickeln können, die sehr genau und empfindlich in Bezug auf die Erkennung von Erkrankungen im Frühstadium sind. Oder dass wir vielleicht irgendwann sagen können, wo im Körper ein Medikament besonders wirksam ist. Weiterhin könnten die Informationen von Medizinerinnen und Medizinern einfacher genutzt werden – ähnlich wie der Genom-Atlas ist der Lungenatlas offen gestaltet. Das Feld der Einzelzell-Genomik lebt schließlich von der Idee, Daten zu teilen. Dadurch ist die Entwicklung der Einzelzell-Technologie auch mit dem Fortschritt der KI so schnell gewachsen. Und genau das finde ich sehr spannend daran: dass wir sehr offen miteinander kommunizieren können.



Der Human Lung Cell Atlas wird stetig weiterentwickelt: Er soll bald eine diversere Bevölkerungsstruktur abbilden. Und stärker in klinische Projekte einfließen, die beispielsweise erforschen, wie sich die zelluläre Landschaft über die Schweregrade einer COPD hinweg verändert. Ziel ist es, aus den 20.000 menschlichen Genen diejenigen herauszufiltern, die für die Diagnose einer bestimmten Erkrankung relevant sind. Damit ließe sich der Atlas zukünftig mit weniger Datenaufwand zur Diagnose nutzen.

Dieser Artikel erschien im Oktober 2024 in der SYNERGIE-Ausgabe #2 | 2024