

Zelldifferenzierung

Mit Einzelzell-Genomik die Entscheidungen von Zellen verfolgen

BO HU, JAN PHILIPP JUNKER

BERLIN INSTITUTE FOR MEDICAL SYSTEMS BIOLOGY, MAX-DELBRÜCK-CENTRUM FÜR MOLEKULARE MEDIZIN, BERLIN

During embryonic development, cells need to take a series of successive fate decisions in order to reach their final differentiated stage. Understanding the processes that give rise to the multitude of different cell types in an organism is a major question in developmental biology. New methods in single cell genomics enable researchers to decipher the transcriptional programs and gene regulatory mechanisms that underlie cell fate decisions during embryonic development.

DOI: 10.1007/s12268-021-1526-4
© Die Autoren 2021

■ In der embryonalen Entwicklung bildet sich aus einer einzigen Zelle, der befruchteten Eizelle, innerhalb erstaunlich kurzer Zeit ein kompletter Organismus aus vielen verschiedenen Zelltypen (Muskelzellen, Hautzellen, Nervenzellen usw.), die räumlich präzise angeordnet sind und spezialisierte Aufgaben wahrnehmen. Die verschiedenen Zellen im Körper tragen im Wesentlichen alle die gleiche genomische Information. Die Unterschiede zwischen verschiedenen Zelltypen gehen auf die Genomregulation zurück, beruhen also darauf, welche Gene in einem bestimmten Zelltyp „aktiv“ und exprimiert sind. Die Mechanismen zu verstehen, die zum Entstehen der verschiedenen Zelltypen im Organismus führen, ist eine der zentralen Fragen der Entwicklungsbiologie.

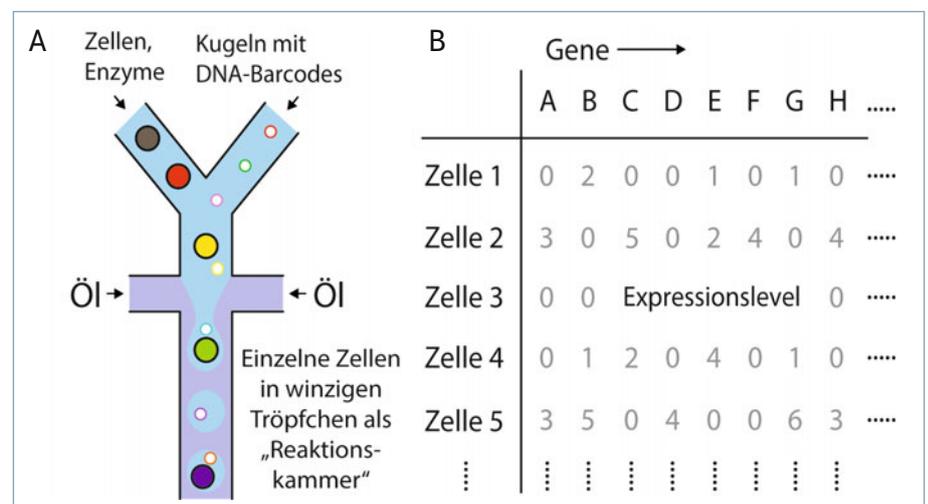
Identifizierung von Zelltypen mittels Einzelzellgenomik

Genetische Experimente, bei denen die Expression oder die Funktion einzelner Gene gezielt verändert wird, sind und bleiben unverzichtbar für die Analyse von Zelldifferenzierung. In den letzten Jahren haben sich aber zunehmend neue Methoden der Einzelzell-Genomik etabliert, die unser Verständnis für entwicklungsbiologische Differenzierungsvorgänge revolutionieren. Durch Sequenzierung der RNA einzelner Zellen können wir in tausenden Zellen messen, welche Gene exprimiert werden. Dazu

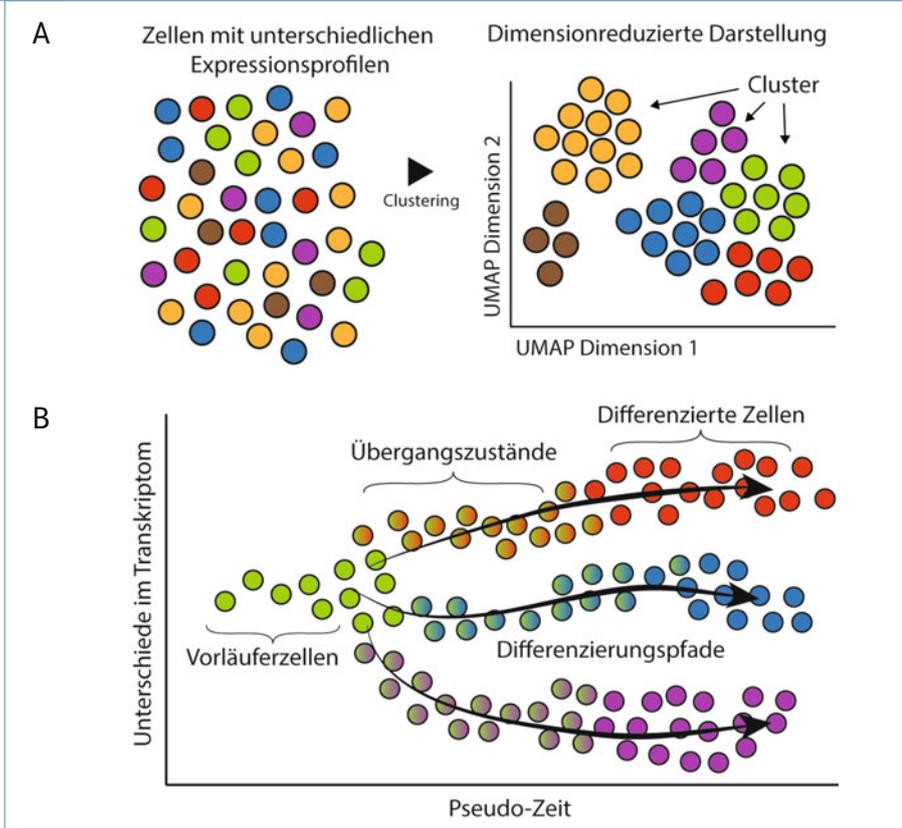
wird das zu untersuchende Gewebe in der Regel in eine Einzelzell-Suspension dissoziiert. Die RNA der einzelnen Zellen wird daraufhin mittels reverser Transkriptase in DNA umgeschrieben und per *next generation sequencing* analysiert. Dazu werden die einzelnen Zellen beispielsweise über Mikrofluidik-Verfahren unabhängig voneinander prozessiert (Abb. 1A). Die resultierenden Datensätze sind große Tabellen, bestehend

aus tausenden Genen in tausenden Zellen (Abb. 1B). Allerdings ist es mit momentanen molekularbiologischen Methoden nicht möglich, jedes einzelne RNA-Molekül in der Zelle zu detektieren. Je nach Technik werden üblicherweise nur zwischen einem und zehn Prozent der RNA-Moleküle erfasst. Dieser Informationsverlust führt zu Rauschen in den Daten und ist eine der wesentlichen Herausforderungen bei der Analyse von Einzelzell-Genomik-Daten. Neben RNA-Sequenzierung (Einzelzell-Transkriptomik) können auch andere genomische Parameter effizient in einzelnen Zellen gemessen werden (z. B. Chromatin-Zugänglichkeit oder DNA-Methylierung). RNA-Sequenzierung ist allerdings bei Weitem die etablierteste Technik der Einzelzell-Genomik und wird daher im Zentrum dieses Artikels stehen.

Einer der wichtigsten Vorzüge von Einzelzell-Transkriptomik ist die Möglichkeit, auf systematische Weise die verschiedenen Zelltypen in einem Gewebe zu identifizieren. Verschiedene Zelltypen sind durch unterschiedliche Genexpressionsprofile charakte-



▲ **Abb. 1:** Einzelzell-Transkriptomik. **A**, Einzelne Zellen werden zusammen mit kleinen Polymerkugeln, die DNA-Barcodes auf ihrer Oberfläche enthalten, in winzige Tröpfchen eingebettet. Diese dienen als Reaktionskammern für reverse Transkription. **B**, Die Daten lassen sich in Form von Tabellen darstellen, bei denen für jede Zelle die Expressionswerte sämtlicher Gene angegeben werden.



▲ **Abb. 2:** Zelltypen und Differenzierungspfade. **A,** Identifizierung von Zelltypen. Durch rechnergestützte Algorithmen werden Cluster von Zellen mit ähnlichen Expressionsprofilen identifiziert. Diese Cluster entsprechen den verschiedenen Zelltypen der Probe. **B,** Identifizierung von Differenzierungspfadern durch pseudo-zeitliches Ordnen von Zellen nach Ähnlichkeit ihrer Expressionsmuster.

risiert. Durch rechnergestütztes Clustering lassen sich die Zellen in verschiedene Gruppen sortieren, die den unterschiedlichen Zelltypen in der Probe entsprechen (**Abb. 2A**). Anhand der exprimierten Gene können wir die Identität des Clusters erschließen und dem Zelltyp einen Namen zuordnen. Eine Herausforderung besteht allerdings in der hohen Dimensionalität der Daten: Für jede Zelle haben wir tausende Messwerte, nämlich die Expressionwerte der einzelnen Gene. Um die Daten graphisch darzustellen, werden daher Methoden zur Reduktion der Dimensionalität gewählt.

Die Erstellung von „Atlanten“ von Zelltypen in verschiedenen Organsystemen ist eine wichtige Basis für weiterführende Arbeiten, z. B. für das Verständnis von Zelltyp-abhängigen Krankheiten. Konsortien wie der Human Cell Atlas arbeiten daran, alle Zelltypen im menschlichen Körper zu identifizieren. Die Anwendungen der Einzelzell-Transkriptomik gehen aber mittlerweile weit über die Identifizierung von Zelltypen hinaus. Insbesondere das Gewinnen von Informationen über die räumliche Anordnung von Zelltypen, über Mechanismen der Zelldifferenzierung und über den Ursprung von Zelltypen steht im Mittelpunkt aktueller Arbeiten [1].

Die Pfade der Zelldifferenzierung

Da die Zellen für die Sequenzierung zerstört werden, ist es nicht möglich, den Differenzierungsprozess einzelner Zellen direkt transkriptomweit zu beobachten. Allerdings kann diese Einschränkung durch „pseudo-zeitliches Ordnen“ von Einzelzell-Transkriptomen zu einem gewissen Grad überwunden werden. Wenn Zellen von Zelltyp A nach B differenzieren, können – sofern ausreichend Zellen sequenziert werden – auch extrem seltene und kurzlebige Übergangszustände detektiert werden. So lässt sich aus einzelnen Momentaufnahmen ein dynamisches Bild der Zelldifferenzierung rekonstruieren (**Abb. 2B**, [2]). Diese Methode ist insbesondere für die schnellen Differenzierungsprozesse der frühen Embryonalentwicklung gut geeignet und ermöglicht es zu beobachten, welche Gene entlang eines Differenzierungspfadens herauf- oder herunterreguliert werden.

Allerdings benötigt die rechnergestützte Rekonstruktion von Differenzierungspfadern häufig zusätzliches Wissen über das biologische System. So lässt sich beispielsweise die Richtung eines Differenzierungsprozess nicht messen. Eine neue Methode namens *RNA velocity* [3] versucht diese Einschränkung zu beheben, indem die Richtung eines

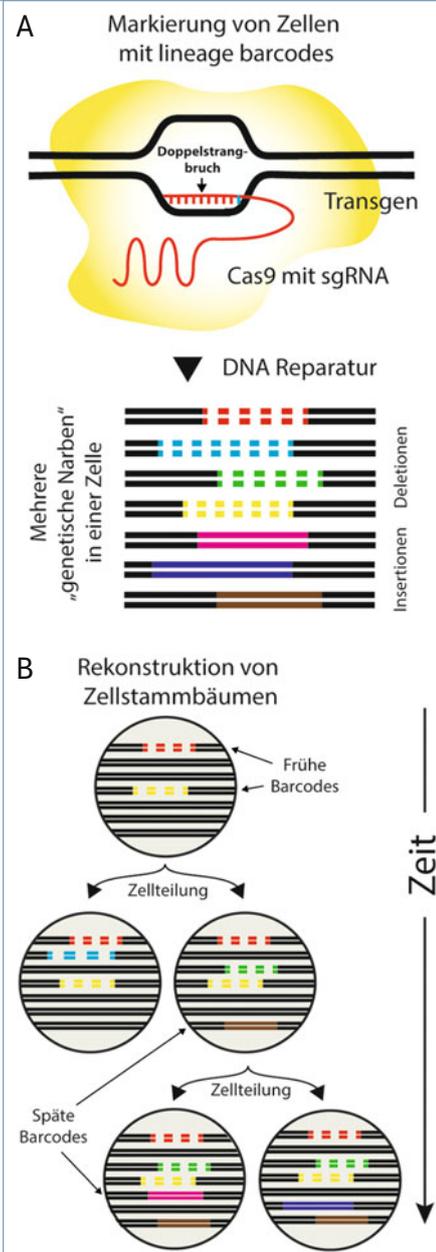
Differenzierungsprozess anhand des Verhältnisses von ungespleißten zu gespleißten Molekülen (d. h. „neue“ im Vergleich zu „alten“ Transkripten) bestimmt wird. Eine weitere vielversprechende Technik nutzt die chemische Markierung von RNA-Molekülen, um die unmittelbare Vergangenheit von zellulären Transkriptomen zu messen [4].

Barcodes in lebenden Organismen

Das Sortieren von Zellen nach Ähnlichkeit ihrer Transkriptome gibt uns zeitliche Information über relativ kurze Zeiträume. Für bestimmte Fragen ist es allerdings wichtig, die Verwandtschaftsbeziehungen von Zellen über lange Zeiträume zu bestimmen. Das beinhaltet Fragen nach dem entwicklungsbiologischen Ursprung adulter Zellen oder auch Krankheitsprozesse, wie z. B. die Bildung von Tumoren. Das Feld des *lineage tracing*, d. h. die Bestimmung von Zellstammbäumen, hat eine lange Tradition, die bis ins 19. Jahrhundert zurückreicht. Üblicherweise werden visuelle Markierungen (heutzutage meist fluoreszente Farbstoffe) als *lineage marker* verwendet, um Zellen zu markieren und zu verfolgen. In jüngerer Zeit wird verstärkt auch die enorme Speicherkapazität des Genoms verwendet, um die Verwandtschaftsbeziehungen von Zellen aufzuzeichnen. Es gibt dabei zwei konzeptionell unterschiedliche Ansätze: Methoden, die natürlich auftretende Mutationen ausnutzen, und solche, die das Genom aktiv modifizieren, um Informationen zu speichern.

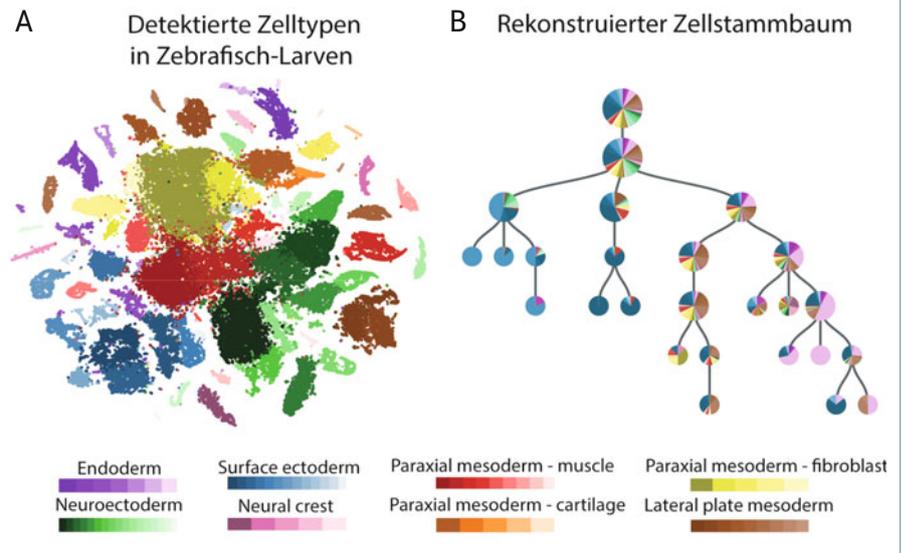
Natürlich auftretende somatische Mutationen eignen sich hervorragend als *lineage marker*, da sie direkt per DNA-Sequenzierung ausgelesen werden können. Da keine Eingriffe in den Organismus nötig sind, können die Verwandtschaftsbeziehungen von Zellen auch im Menschen ausgelesen werden. In den letzten Jahren wurde dieser Ansatz in einigen Arbeiten verwendet, um frühe embryologische Zellentscheidungen zu verfolgen. Die Kosten für das Sequenzieren ganzer Genome in vielen einzelnen Zellen sind bislang jedoch sehr hoch. *Lineage tracing* basierend auf den (wesentlich häufigeren) Mutationen in Mitochondrien ist daher eine interessante Alternative, um Zellstammbäume in Menschen zu rekonstruieren [5].

Für Modellorganismen sind jedoch meist Techniken besser geeignet, bei denen das Genom aktiv modifiziert wird. Wir und andere haben Methoden zur Bestimmung von Zellstammbäumen in tausenden einzelnen Zellen mittels Einzelzell-Transkriptomik entwickelt



▲ **Abb. 3:** Rekonstruktion von Zellstambäumen. **A**, Cas9 erzeugt einen Doppelstrangbruch in einem Transgen. Nach der DNA-Reparatur entsteht eine „genetische Narbe“, die als *lineage barcode* benutzt werden kann. **B**, Rekonstruktion von Zellstambäumen anhand der genetischen Narben, die während der embryonalen Entwicklung geschaffen wurden.

(**Abb. 3**). Diese Techniken verbinden das Erstellen von *lineage barcodes* durch CRISPR/Cas9 mit Zelltyp-Identifizierung via Einzelzell-Transkriptomik [6–8]. Auch wenn es einige technische Unterschiede im Detail gibt, liegt allen diesen Methoden das gleiche Prinzip zugrunde: Cas9 erzeugt Doppelstrangbrüche in exprimierten Transgenen. Während des Reparaturprozesses entstehen „genetische Narben“, kleine Reparaturfehler, die als *lineage barcodes* verwendet werden können. Diese genetischen Narben sind ideal



▲ **Abb. 4:** Beispiel: Zellstammbaum in der Zebrafischlarve. **A**, Clustering von Einzelzell-Transkriptomen in Zebrafischlarven. In der Darstellung sind die detektierten Zelltypen farblich codiert gemäß ihrer Zugehörigkeit zu verschiedenen entwicklungsbiologischen Kategorien. **B**, Mithilfe von *lineage barcodes* kann ein Zellstammbaum rekonstruiert werden. Die Tortendiagramme zeigen den Anteil verschiedener Zelltypen an den jeweiligen Knotenpunkten des Stammbaums. Jeder Knotenpunkt entspricht einem spezifischen *lineage barcode* (modifiziert nach [6]).

als Informationsträger für Abstammungsbeziehungen geeignet, da sie nach Erstellung nicht mehr verändert werden können, über die gesamte Lebensdauer des Organismus stabil sind, und bei der Zellteilung an beide Tochterzellen weitergegeben werden. Ein Beispiel für einen auf diese Weise rekonstruierten Zellstammbaum in einer Zebrafisch-Larve ist in **Abbildung 4** gezeigt. Auch wenn viele Herausforderungen bestehen, hat *lineage tracing* mittels CRISPR-Cas9 großes Potenzial, um den Ursprung sämtlicher Zelltypen zu verstehen und die Mechanismen Zelltyp-abhängiger Krankheiten zu entschlüsseln.

Literatur

- [1] Griffiths JA, Scialdone A, Marioni JC (2018) Using single-cell genomics to understand developmental processes and cell fate decisions. *Mol Syst Biol* 14: e8046
- [2] Kester L, van Oudenaarden A (2018) Single-cell transcriptomics meets lineage tracing. *Cell Stem Cell* 23: 166–179
- [3] La Manno G, Soldatov R, Zeisel A et al. (2018) RNA velocity of single cells. *Nature* 560: 494–498
- [4] Erhard F, Baptista MAP, Krammer T et al. (2019) scSLAM-seq reveals core features of transcription dynamics in single cells. *Nature* 571: 419–423
- [5] Ludwig LS, Lareau CA, Ulirsch JC et al. (2019) Lineage tracing in humans enabled by mitochondrial mutations and single-cell genomics. *Cell* 176: 1325–1339
- [6] Spanjaard B, Hu B, Mitic N et al. (2018) Simultaneous lineage tracing and cell-type identification using CRISPR-Cas9-induced genetic scars. *Nat Biotechnol* 36: 469–473
- [7] Raj B, Wagner DE, McKenna A et al. (2018) Simultaneous single-cell profiling of lineages and cell types in the vertebrate brain. *Nat Biotechnol* 36:442–450
- [8] Alemany A, Florescu M, Baron CS et al. (2018) Whole-organism clone tracing using single-cell sequencing. *Nature* 556: 108–112

Funding note: Open Access funding enabled and organized by Projekt DEAL.
Open Access: Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und

angeben, ob Änderungen vorgenommen wurden. Die in diesem Artikel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen. Weitere Details zur Lizenz entnehmen Sie bitte der Lizenzinformation auf <http://creativecommons.org/licenses/by/4.0/deed.de>.

Korrespondenzadresse:

Prof. Dr. Jan Philipp Junker
Berlin Institute for Medical Systems Biology
Max-Delbrück-Centrum für Molekulare Medizin
Hannoversche Straße 28
D-10115 Berlin
janphilipp.junker@mdc-berlin.de

AUTOREN



Bo Hu
2009–2015 Bachelor- und Masterstudium Molekulare Medizin an der Universität Freiburg. 2016–2020 Doktorand in der Arbeitsgruppe von Prof. Dr. J. P. Junker am Berlin Institute for Medical Systems Biology des Max-Delbrück-Centrums für Molekulare Medizin.



Jan Philipp Junker
2000–2009 Physikstudium an der TU München mit anschließender Promotion in der Gruppe von Prof. Dr. M. Rief. 2010–2015 Postdoc in der Gruppe von Prof. Dr. A. van Oudenaarden am Massachusetts Institute of Technology, USA und am Hubrecht Institut, Utrecht. Seit 2016 Gruppenleiter am Berlin Institute for Medical Systems Biology des Max-Delbrück-Centrums für Molekulare Medizin.