Stiftung Tierärztliche Hochschule Hannover

**Institut für Tierzucht und Vererbungsforschung**

**Bachelorarbeit**

**Evaluation von Resamplingverfahren zur Analyse von Single-cell RNA-seq Daten**

Sergej Ruff

Matrikelnummer: 10024140

Studiengang: Biologie (B.Sc)

Startdatum: 01.06.2022

Abgabedatum: 02.09.2022

Erstprüfer: Prof. Dr. Klaus Jung

Zweitprüfer: Prof. Dr. Bernd Schierwater

Inhaltsverzeichnis ( Aus Inhaltsverzeichnis entfernen)

[Inhaltsverzeichnis ( Aus Inhaltsverzeichnis entfernen) 2](#_Toc923735405)

[Zusammenfassung 3](#_Toc336791368)

[1. Einleitung 3](#_Toc2137218125)

[1.1. Single-cell RNA sequencing (scRNA-seq) im Vergleich zu Bulk-Ansätzen 4](#_Toc1378068543)

[1.2. Single-cell RNA sequencing: Durchführung 6](#_Toc1157975797)

[1.2.1 Single-cell RNA sequencing: Zellpräparation 6](#_Toc1691066332)

[1.2.2 Single-cell RNA sequencing: Datenanalyse 7](#_Toc1564904119)

[2. Methoden und Daten 7](#_Toc182052165)

[3. Ergebnisse 7](#_Toc1725178871)

[4. Diskussion 7](#_Toc351822704)

[5. Literaturverzeichnis 8](#_Toc731322696)

[Anhang 8](#_Toc302398698)

[R-Code 8](#_Toc128016496)

# Zusammenfassung

Eine Seite Zusammenfassung

# 1. Einleitung

Single-cell RNA sequencing (scRNA-seq) erlaubt die Detektion und quantitative Analyse von Transkripten auf der Ebene einzelner Zellen. Damit ist eine Analyse der Genexpression nicht mehr auf Zellpopulationen beschränkt und man ist in der Lage Genexpressionsunterschiede zwischen einzelnen Zellen zu detektieren, was mit den bisherigen Bulk-RNA-Ansätzen nicht möglich gewesen ist **(Macaulay und Voet, 2014)**. Diese Arbeit untersucht die Stabilität des scRNA-seq Verfahrens, indem man sich anschaut, wie oft ein Marker-Gen nach Bootstrapping einem bestimmten Cluster zugeordnet wird.

In den folgenden Teilkapiteln werden zunächst die Grundlagen für scRNA-seq gelegt. Zum einem werden hier die biologischen Aspekte betrachtet und geklärt, welche Vorteile ScRNA-seq im Vergleich zu Bulk-RNA-Verfahren hat; und zum andrem werden die technischen Fragen geklärt: Wie gewinnt man einzelne Zellen aus einem Gewebe und wie analysiert man die ScRNA-seq Daten in Studio R? Anschließend wird die Theorie hinter Bootstrapping beschrieben.

Meyer, J. S., Ingersoll, C. G., McDonald, L. L., & Boyce, M. S. (1986). Estimating uncertainty in population growth rates: jackknife vs. bootstrap techniques. *Ecology*, *67*(5), 1156-1166.

Saremi, B., Kohls, M., Liebig, P., Siebert, U., & Jung, K. (2021). Measuring reproducibility of virus metagenomics analyses using bootstrap samples from FASTQ-files. *Bioinformatics*, *37*(8), 1068-1075.

Saremi, B., Gusmag, F., Distl, O., Schaarschmidt, F., Metzger, J., Becker, S., & Jung, K. (2022). A comparison of strategies for generating artificial replicates in RNA-seq experiments. *Scientific reports*, *12*(1), 1-13.

Luecken, M. D., & Theis, F. J. (2019). Current best practices in single‐cell RNA‐seq analysis: a tutorial. *Molecular systems biology*, *15*(6), e8746.

Kulkarni, A., Anderson, A. G., Merullo, D. P., & Konopka, G. (2019). Beyond bulk: a review of single cell transcriptomics methodologies and applications. *Current opinion in biotechnology*, *58*, 129-136.

## 1.1. Single-cell RNA sequencing (scRNA-seq) im Vergleich zu Bulk-Ansätzen

Das Transkriptom beschreibt die Gesamtheit aller RNA-Moleküle, die eine Zelle herstellt, und anders als das Genom ist das Transkriptom ein dynamisches System, welches sich je nach Zelltyp und Zellzustand unterscheidet. Das bedeutet, dass sich die Genexpression von Zelle zur Zelle unterscheiden wird und selbst Zellpopulationen, die phänotypisch homogen zu seien scheinen, können Visibilitäten aufweisen **(Huang, 2009)**.

Huang S. (2009). Non-genetic heterogeneity of cells in development: more than just noise. *Development (Cambridge, England)*, *136*(23), 3853–3862. https://doi.org/10.1242/dev.035139

Die herkömmlichen Verfahren, die sich mit dem Transkriptom befasst haben, haben sich aber bei ihrer Analyse auf ganze Zellpopulationen bezogen. Unser jetziges Wissen zum Transkriptom bezieht sich also auf Arbeiten, die Tausende Zellen und ihre Transkriptome gleichzeitig analysiert haben. Diese Bulk-Ansätze, die ganze Zellpopulationen analysieren, können die Heterogenität der Transkriptome nicht einfangen.

Erst nachdem das erste Paper, welches sich mit dem Transkriptom auf Einzel-Zell-Ebene befasst hat, erschien ist **(Tang et al al., 2009)**, stieg die Anzahl an Arbeiten, die sich mit scRNA-seq befassen, rasant an.

Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B. B., Siddiqui, A., Lao, K., & Surani, M. A. (2009). mRNA-Seq whole-transcriptome analysis of a single cell. *Nature methods*, *6*(5), 377–382. <https://doi.org/10.1038/nmeth.1315>

Single-cell RNA sequencing ermöglicht die Analyse des Transkriptoms auf Einzel-Zell-Ebene. Bulk-Ansätze können nur die Expression von Genen über die ganze Zellpopulationen detektieren, während scRNA-seq die Analyse der Genexpression einzelner Zellen ermöglicht. Anders als Bulk-Ansätze kann scRNA-seq auch Co-Expressionsmuster zwischen zwei Genen kenntlich machen **(Macaulay und Voet, 2014)**. Das periphere Blut enthält mehrere verschiedene Zelltypen. Dazu gehören die Lymphozyten, die Monozyten, Die Neutrophilen Zellen und weitere Zellen. Das periphere Blut und viele andere Gewebe des Körpers bestehen also auch heterogenen Zellpopulationen. Die Identifikation und Charakterisierung der einzelnen Zellen innerhalb einer heterogenen Zellpopulation wird durch scRNA-seq ermöglicht. Genauso kann man auch seltene Zellpopulationen innerhalb eines Gewebes entdecken und die Genexpressionsunterschiede zwischen den Zellen innerhalb einer heterogenen Zellpopulation untersuchen. Die Untersuchung der Genexpressionsunterschiede auf Einzel-Zell-Ebene ist wichtig für die Aufdeckung vieler Krankheiten und ihrer Mechanismen, weil Krankheiten wie Malaria **(Rivera-Correa et al., 2017)** und COVID-19 **(De Biasi et al., 2020)** gezielt die Genexpression bestimmter Zelltypen verändern, um sich vor dem Immunsystem des Wirtes zu schützen.

Rivera-Correa, J., Guthmiller, J. J., Vijay, R., Fernandez-Arias, C., Pardo-Ruge, M. A., Gonzalez, S., Butler, N. S., & Rodriguez, A. (2017). Plasmodium DNA-mediated TLR9 activation of T-bet+ B cells contributes to autoimmune anaemia during malaria. *Nature communications*, *8*(1), 1282. <https://doi.org/10.1038/s41467-017-01476-6>

De Biasi, S., Lo Tartaro, D., Meschiari, M., Gibellini, L., Bellinazzi, C., Borella, R., Fidanza, L., Mattioli, M., Paolini, A., Gozzi, L., Jaacoub, D., Faltoni, M., Volpi, S., Milić, J., Sita, M., Sarti, M., Pucillo, C., Girardis, M., Guaraldi, G., Mussini, C., … Cossarizza, A. (2020). Expansion of plasmablasts and loss of memory B cells in peripheral blood from COVID-19 patients with pneumonia. *European journal of immunology*, *50*(9), 1283–1294. https://doi.org/10.1002/eji.202048838

## 1.2. Single-cell RNA sequencing: Durchführung

### 1.2.1 Single-cell RNA sequencing: Zellpräparation

Die Isolation von Zellen für scRNA-seq erfolgt aus festen Geweben wie zum Beispiel Darmzotten oder aus Zellsuspensionen wie dem peripheren Blut. Die Isolation erfolgt über Droplet basierte Technologien. Oft genutzte Technologien sind das 10X Genomics Chromium, Dropseq und inDrop (Kulkarni et al., 2019). Sie alle funktionieren nach demselben Grundprinzip und können deshalb in folgende Schritte eingeteilt werden: Vorbereitung der Zellsuspension, eine optionale Zellsortierung, Einkapseln einzelner Zellen in Droplets, cDNA-Synthese und Amplifikation, Erstellen einer Bibliothek und Sequenzierung sowie Vorprozessierung der gewonnen Daten **(Abbildung \_, Salomon et al., 2019)**.

|  |
| --- |
|  |
| ***Abbildung \_:*** *Schematische Darstellung einer Droplet basierten Zellpräparation für scRNA-seq. Bestehend aus Vorbereitung der Zellsuspension, eine optionale Zellsortierung, Einkapseln einzelner Zellen in Droplets, cDNA-Synthese und Amplifikation, Erstellen einer Bibliothek und Sequenzierung sowie Vorprozessierung der gewonnen Daten* ***(Salomon et al., 2019)****.* |

Zu Beginn der Zellpräparation müssen feste Gewebe in eine Zellsuspension dissoziiert werden. Ein FACS-Gerät kann genutzt werden, um mögliche tote Zellen zu entfernen, wodurch die gewünschten Zellen angereichert werden. Ziel der Droplet basierten Verfahren ist es, einen abgeschlossenen Reaktionsraum zu erschaffen. Ein Bead und einzelne Zellen aus der Suspension treffen hier in einem Mikrofluid aufeinander. Ein Öl trennt die Flüssigkeit in einzelne Droplets auf. Im Droplet befindet sich ein Lysispuffer. Sollte sich im Droplet eine Zelle befinden, dann wird diese Zelle durch den Lysispuffer lysiert und die RNA aus der Zelle wird im Droplet freigesetzt. Die eingeführten Beads besitzen einen Oligo-d(T)-Primer auf ihrer Oberfläche, wodurch sie komplementär an den Poly-A-Schwanz der RNA binden können. Des Weiteren besitzen die Beads Cell Barcodes und Unique Molecular Identifiers (UMI). Cell Barcodes sind identische, mehrmals auftauchende Sequenzen auf den Beads und Sie werden genutzt für die Identifikation einzelner Zellen. UMI sind einzigartige Sequenzen auf den Beads, die als Tags genutzt werden, um die einzelnen Transkripte zu markieren **(Salomon et al., 2019)**. So ist es möglich die einzelnen Transkripte einer bestimmten Zelle zuzuordnen. Nachdem die einzelnen RNAs an den Bead gebunden haben, werden über RT-PCR cDNAs mit Barcode und UMIs generiert und Bibliotheken erstellt. Diese cDNAs werden über Next-Generation Sequenzer sequenziert und vor-prozessiert.

### 1.2.2 Single-cell RNA sequencing: Datenanalyse

# 2. Methoden und Daten

Methoden

# 3. Ergebnisse

Ergebnisse

# 4. Diskussion

Diskussion

# 5. Literaturverzeichnis

# Anhang

## R-Code