

DISS. ETH NO. 29540

**COMPUTATIONAL METHODS FOR DISSECTING  
CANCER HETEROGENEITY USING MULTI-MODAL  
DATA: FROM BULK TO SPATIAL RESOLUTION**

A thesis submitted to attain the degree of

**DOCTOR OF SCIENCES**

(Dr. sc. ETH Zurich)

presented by

**JOANNA FICEK-PASCUAL**

M. Sc. in Biostatistics, LMU Munich

born on 08.09.1992

accepted on the recommendation of

Prof. Dr. Gunnar Rätsch, examiner

Prof. Dr. Viktor H. Koelzer, co-examiner

Dr. Kjong-Van Lehmann, co-examiner

Dr. Ana Conesa, co-examiner

2023

# Abstract

Cancer is a disease characterized by a high degree of heterogeneity. The inter-patient variability at the histological, molecular, and functional levels hinders the development of one-fits-all treatments. Furthermore, the high intra-tumor heterogeneity, with cancer cells and the tumor microenvironment (TME) changing spatially and in the course of the disease, often leads to treatment resistance and cancer relapse. Given its complexity, the heterogeneity should be studied from various perspectives at multiple resolutions and biological levels. Recent technological advances enable molecular profiling of patients at bulk, single-cell, and even spatial resolution. The resulting data can provide new insights into tumorigenesis, cancer progression, TME composition, and cell-to-cell interactions. With the increasing number of features and cells profiled with multiple omic technologies, novel analytical algorithms are needed to integrate and extract meaningful information from the data. In this thesis, we propose new computational methods, adapt existing ones to new applications, and perform extensive analyses of the multi-omic profiles of cancer patients. To this end, we leverage the highly multi-modal dataset collected within the Swiss Tumor Profiler Consortium and study cancer heterogeneity in metastatic melanoma and high-grade serous ovarian cancer (HGSOC).

Studying the transcriptome using bulk RNA sequencing (bkRNA-seq) has previously elucidated many mechanisms underlying cancer. Given the high inter-patient heterogeneity, it is crucial first to identify and then obtain an in-depth understanding of the patient stratification that reflects different cancer subtypes and disease prognoses. Therefore, in this work, we employ bkRNA-seq to characterize differences between clinically and molecularly defined melanoma subtypes. We first perform a thorough data quality control and adjust the analyses for effects confounding the signal. We then perform

---

comparisons at gene and pathway levels, with the latter allowing for seamless integration with other data modalities. The analysis results align with the prior knowledge as well as provide further insights into the molecular landscape of metastatic melanoma disease.

The single-cell profiling of cancer samples can help dissect both the inter-patient and the intra-tumor heterogeneity by uncovering multiple clones and phenotypes of cancer cells or describing the cell-type composition and molecular characteristics of the TME. Hence, in this work, we employ single-cell proteomics to study HGSOC cancer cell heterogeneity before chemotherapy administration and at cancer relapse. Using the late-stage integration strategy, we then leverage matched genomic and transcriptomic information to identify features related to the intra-tumor heterogeneity observed in the HGSOC samples. However, to unlock the full potential of multi-omic single-cell datasets, methods allowing for an early-stage integration are needed. Even though procedures capable of measuring several modalities on the same cell are emerging, their scalability and widespread applicability are still limited. Therefore, we propose a computational method called Single-Cell Integration via Matching (SCIM) that enables a fusion of datasets stemming from single-cell profiling of a sample with multiple omics technologies.

Despite the information richness, single-cell data resulting from dissociated tissues cover only some aspects of tumor heterogeneity, as the progression of solid cancers is a spatial process. Thus, the growing interest in spatial biomarkers has sparked technological advances allowing multiplexed profiling while preserving the spatial context. In this work, we address two challenges of spatial profiling. First, we tackle the summarization of the fine-grained information on cellular neighborhoods at a sample level and propose an aggregation scheme that reduces information loss. Second, we develop a method for Multiplexed Virtual Staining of histopathology images (Multi-V-Stain) to extend the multiplex protein profiling of small regions of interest to the scale of the full histopathological slide. As the input and the desired output images come from subsequent tissue sections, we devise a comprehensive evaluation scheme accounting for the slice-to-slice discrepancy. Lastly, we provide an outlook on how analysis of spatial patterns of hallmark pathway activation can contribute to our understanding of cancer heterogeneity.

# Zusammenfassung

Krebs ist eine Krankheit, die durch eine hohe Heterogenität gekennzeichnet ist. Die Variabilität zwischen Patienten auf histologischer, molekularer und funktioneller Ebene behindert die Entwicklung einheitlicher Behandlungen. Darüber hinaus führt die hohe intratumorale Heterogenität, bei der sich Krebszellen und die Tumormikroumgebung (TME) räumlich und im Krankheitsverlauf verändern, häufig zu Behandlungsresistenz und Krebsrückfällen. Angesichts ihrer Komplexität sollte die Heterogenität aus verschiedenen Perspektiven mit mehreren Auflösungen und auf unterschiedlichen biologischen Ebenen untersucht werden. Jüngste technologische Fortschritte ermöglichen die molekulare Profilierung von Patienten mit Massen-, Einzelzell- und sogar räumlicher Auflösung. Die daraus resultierenden Daten können neue Einblicke in die Tumorentstehung, das Fortschreiten des Krebses, die TME-Zusammensetzung und Zell-Zell-Interaktionen liefern. Mit der zunehmenden Anzahl von Merkmalen und Zellen, die mit unterschiedlichen Omic-Technologien profiliert werden, sind neuartige Analysealgorithmen erforderlich, um die Daten zu integrieren und um die aussagekräftige Informationen aus ihnen zu extrahieren. In dieser Arbeit schlagen wir neue analytische Methoden vor, passen bestehende an neue Anwendungen an und führen umfangreiche Analysen der multiomischen Profile von Krebspatienten durch. Zu diesem Zweck nutzen wir den hochgradig multimodalen Datensatz, der im Rahmen des Swiss Tumor Profiler Consortium gesammelt wurde, und untersuchen die Krebsheterogenität bei metastasiertem Melanom und hochgradigem serösem Ovarialkarzinom (HGSOC).

Die Untersuchung des Transkriptoms mittels Massen-RNA-Sequenzierung (bkRNA-seq) hat bereits viele Mechanismen aufgeklärt, die Krebs zugrunde liegen. In Anbetracht der hohen Heterogenität zwischen Patienten ist es von

---

entscheidender Bedeutung, zunächst die Patientenstratifizierung zu identifizieren, die unterschiedliche Krebssubtypen und Krankheitsprognosen widerspiegelt, und anschließend ein tiefgreifendes Verständnis für diese zu erlangen. Daher verwenden wir in dieser Arbeit bkRNA-seq, um Unterschiede zwischen klinisch und molekular definierten Melanom-Subtypen zu charakterisieren. Wir führen zunächst eine gründliche Datenqualitätskontrolle durch und passen die Analysen auf signalverfälschende Effekte an. Anschließend führen wir Vergleiche auf Gen- und Signalwegebene durch, wobei letzteres eine nahtlose Integration mit anderen Datenmodalitäten ermöglicht. Die Analyseergebnisse stimmen mit dem bisherigen Wissen überein und geben weitere Einblicke in die molekulare Landschaft der metastasierten Melanomerkrankung.

Die Einzelzellprofilierung von Krebsproben kann dabei helfen, sowohl die Heterogenität zwischen Patienten als auch innerhalb des Tumors zu analysieren, indem mehrere Klone und Phänotypen von Krebszellen aufgedeckt oder die Zelltypzusammensetzung und die molekularen Eigenschaften des TME beschrieben werden. Daher verwenden wir in dieser Arbeit Einzelzell-Proteomik, um die Heterogenität von HGSOE-Krebszellen vor der Verabreichung einer Chemotherapie und bei einem Krebsrückfall zu untersuchen. Mithilfe der Spätphasen-Integrationsstrategie nutzen wir dann abgestimmte genomische und transkriptomische Informationen, um Merkmale zu identifizieren, die mit der in den HGSOE-Proben beobachteten intratumoralen Heterogenität zusammenhängen. Um jedoch das volle Potenzial multiomischer Einzelzell-datensätze auszuschöpfen, sind Methoden erforderlich, die eine frühzeitige Integration ermöglichen. Auch wenn immer mehr Verfahren aufkommen, mit denen mehrere Modalitäten an derselben Zelle gemessen werden können, sind ihre Skalierbarkeit und breite Anwendbarkeit noch begrenzt. Daher schlagen wir eine Berechnungsmethode namens Single-Cell Integration via Matching (SCIM) vor, die eine Fusion von Datensätzen ermöglicht, die aus der Einzelzellprofilierung einer Probe mit mehreren Omics-Technologien stammen.

Trotz des Informationsreichtums decken Einzelzelldaten aus dissoziierten Geweben nur einige Aspekte der Tumorerheterogenität ab, da das Fortschreiten solider Krebserkrankungen ein räumlicher Prozess ist. Das wachsende Interesse an räumlichen Biomarkern hat daher zu technologischen Fortschritten geführt, die eine Multiplex-Profilierung unter Beibehaltung des räumlichen

---

Kontexts ermöglichen. In dieser Arbeit befassen wir uns mit zwei Herausforderungen der räumlichen Profilierung. Zunächst nehmen wir die Zusammenfassung der feinkörnigen Informationen über zelluläre Nachbarschaften auf Stichprobenebene in Angriff und schlagen ein Aggregationsschema vor, das den Informationsverlust reduziert. Zweitens entwickeln wir eine Methode zur Multiplex-Virtual-Färbung histopathologischer Bilder (Multi-V-Stain), um die Multiplex-Proteinprofilierung kleiner ausgewählter Regionen auf den Maßstab des gesamten histopathologischen Bildes zu erweitern. Da die Eingabe- und die gewünschten Ausgabebilder aus aufeinanderfolgenden Gewebeschnitten stammen, entwickeln wir ein umfassendes Auswertungsschema, das die Diskrepanz zwischen den einzelnen Schichten berücksichtigt. Abschließend geben wir einen Ausblick darauf, wie die Analyse räumlicher Muster der Aktivierung von Signalwegen zu unserem Verständnis von Krebsheterogenität beitragen kann.