

# Computational Studies in Genomic and Transcriptomic Tumor Diversity

**Doctoral Thesis**

**Author(s):**

Moore, Ariane Leoni

**Publication date:**

2018

**Permanent link:**

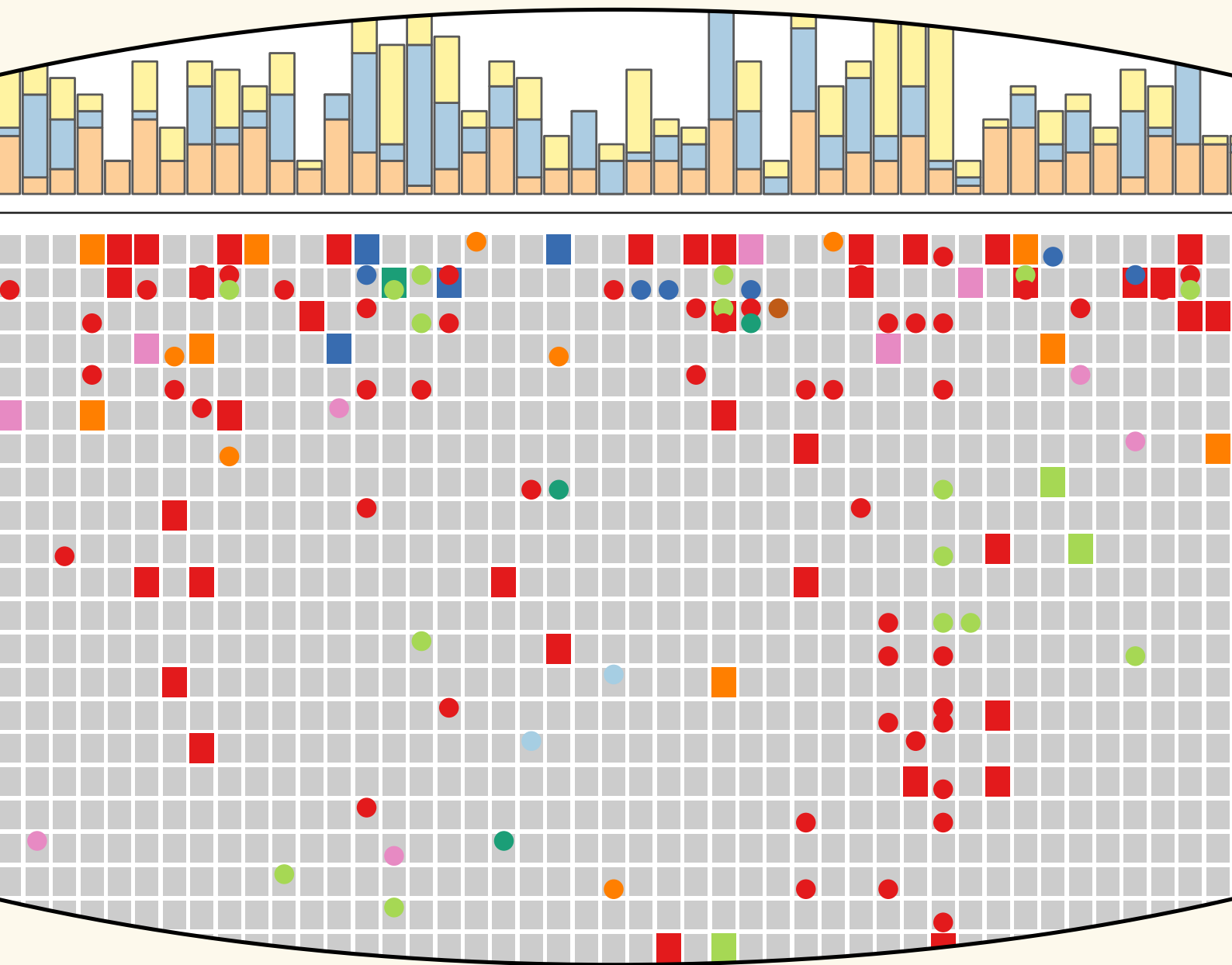
<https://doi.org/https://doi.org/10.3929/ethz-b-000319695>

**Rights / license:**

[In Copyright - Non-Commercial Use Permitted](#)

**COMPUTATIONAL STUDIES IN**

***GENOMIC AND TRANSCRIPTOMIC  
TUMOR DIVERSITY***



Diss. ETH N° 25442

**COMPUTATIONAL STUDIES IN  
GENOMIC AND TRANSCRIPTOMIC TUMOR DIVERSITY**

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES of ETH ZURICH  
(Dr. sc. ETH Zurich)

presented by

ARIANE LEONI MOORE

M.Sc., Major - Mathematics  
Virginia Polytechnic Institute and  
State University, USA

born on 20.05.1990

citizen of Germany

accepted on the recommendation of

Prof. Dr. Niko Beerenwinkel  
Dr. Christian Beisel  
Prof. Dr. Karsten Borgwardt  
Prof. Dr. Holger Moch

2018

# Abstract

Tumor development is an evolutionary process involving mutation and selection that generates a heterogeneous collection of clones, which can also interact with each other. Smaller clones in the tumor may have a critical impact on treatment response and relapse. Therefore, it is essential to have robust and reliable bioinformatics pipelines and variant callers that are also able to detect low-frequency mutations. We developed a computational pipeline for cancer whole-exome, panel, and RNA sequencing data. In order to improve the crucial step of variant calling, we performed a comparison of different tools based on simulated data. This study revealed many differences in the tools' performance and also inspired the development of a novel consensus variant detection approach named `rank-combination`. It outperformed all individual tools, especially in calling low-frequency variants. We next applied our bioinformatics pipeline including the `rank-combination` to a large cohort comprising 178 biopsies from 89 kidney cancer patients. This analysis revealed intra-tumor heterogeneity on the genomic and transcriptomic level. Furthermore, we introduced the term "clonal exclusivity", which describes the observation of mutual exclusivity of mutations in clones within a tumor. We developed an efficient statistical framework to detect clonally exclusive gene pairs in large patient cohorts. Such pairs can shed light on selection pressures favoring co-existent clones with synergistic effects. Our approach is implemented as an R software package called `GeneAccord`, and we applied it to the cohort of kidney cancer patients. This analysis revealed MUC16 and TP53 as significant candidate pairs for synergistic co-existent clones, and adds evidence that MUC16 is a cancer driver. Finally, metastases are oftentimes a death sentence for cancer patients. In order to learn more about how cancer progresses and metastasizes, we also applied our pipeline to analyze melanoma samples which stem from a mouse model that includes timed samples from the primary tumor and metastases. We identified upregulated expression of genes that pave the way for metastases, and showed that their overexpression is stronger in later time points of tumor progression. Furthermore, we tracked genomic diversity over time, and performed network-based integration to detect important signaling hubs during progression and in metastases. To sum up, our results contribute towards an improved understanding of oncogenesis. The provided tools and insight will be useful in future cancer studies and may eventually lead to better treatment strategies.



# Kurzfassung

Tumore entwickeln sich in einem evolutionären Prozess, der Mutation und Selektion beinhaltet, und welcher eine heterogene Sammlung von Klonen erzeugt, die auch miteinander interagieren können. Kleinere Klone im Tumor können eine kritische Auswirkung auf den Behandlungserfolg einer Krebstherapie haben und zu Rückfällen führen. Daher ist es wichtig, robuste und zuverlässige Bioinformatikanalysesysteme, sowie Methoden zur Detektion von Varianten zu haben, die auch niederfrequente Mutationen aus kleineren Klonen erkennen können. Wir entwickelten ein rechnergestütztes Analysesystem für Krebsdaten, welche von der Sequenzierung des gesamten Exoms, bestimmten Regionen des Exoms, oder des Transkriptoms stammen. Um den entscheidenden Schritt des Variantendetektierens zu verbessern, führten wir einen Vergleich durch, bei dem wir verschiedene Methoden basierend auf simulierten Daten evaluierten. Diese Studie zeigte viele Unterschiede in der Leistung der einzelnen Methoden und inspirierte auch die Entwicklung eines neuen Konsensansatzes namens *rank-combination*, bei dem mehrere Methoden zur Variantendetektion kombiniert werden. Es übertraf alle individuellen Methoden in seiner Leistung, und war besonders gut beim Aufspüren von niederfrequenten Varianten.

Als nächstes verwendeten wir unser rechnergestütztes Analysesystem einschließlich der *rank-combination*, um einen großen Datensatz mit 178 Biopsien von 89 Nierenkrebspatienten zu analysieren. Diese Analyse zeigte eine Intra-Tumorheterogenität auf der genomischen und transkriptomischen Ebene. Außerdem führten wir einen neuen Begriff ein, nämlich "klonale Exklusivität", welcher die Beobachtung der gegenseitigen Exklusivität von Mutationen in Klonen innerhalb eines Tumors beschreibt. Wir entwickelten eine innovative statistische Methode, um klonal exklusive Genpaare in großen Patientenkohorten zu finden. Solche Paare können Aufschluss über einen Selektionsdruck geben, welcher koexistente Klone mit synergistischen Effekten begünstigt. Unsere Methode ist als ein R-Softwarepaket namens *GeneAccord* implementiert, und wir wendeten es auf den großen Kohort von Nierenkrebspatienten an. Diese Analyse ergab, dass MUC16 und TP53 ein signifikantes Kandidatenpaar für synergistisch koexistente Klone ist, und bedeutet, dass MUC16 auch ein wichtiges Krebsgen sein könnte.

Metastasen sind oft ein Todesurteil für Krebspatienten. Um weitere Erkenntnisse darüber zu erlangen, wie Krebs fortschreitet und metastasiert, setzten wir unser Bioinformatikanalysesystem auch zur Analyse von Melanomproben von Mäusen ein. Die Proben von Primärtumor und Metastasen stammen von unterschiedlichen Zeitpunkten in der Tumorentwicklung. Wir identifizierten eine hochregulierte Expression von Genen, die den Weg für Metastasen ebnen, und zeigten, dass ihre Überexpression in späteren Zeitpunkten der Tumorprogression stärker ist. Darüber hinaus verfolgten wir die genomische Diversität zwischen Metastasen und Primärtumor über die Zeit, und führten netzwerkbasierende Datenintegration durch, um wichtige Signalknotenpunkte während der Progression und in Metastasen zu erkennen.

Zusammenfassend tragen unsere Ergebnisse zu einem verbesserten Verständnis der Onkogenese bei. Die zur Verfügung gestellten Werkzeuge und Erkenntnisse werden in zukünftigen Krebsstudien nützlich sein und können zur Entwicklung von verbesserten Behandlungsstrategien beitragen.