Evolution of Cancer - Reconstructing the Past, Predicting the Future

742. WE-Heraeus-Seminar

21 - 25 March 2022

hybrid

at the Physikzentrum Bad Honnef, Germany

WILHELM UND ELSE
HERAEUS-STIFTUNG

Introduction

The Wilhelm und Else Heraeus-Stiftung is a private foundation that supports research and education in science with an emphasis on physics. It is recognized as Germany's most important private institution funding physics. Some of the activities of the foundation are carried out in close cooperation with the German Physical Society (Deutsche Physikalische Gesellschaft). For detailed information see https://www.we-heraeus-stiftung.de

Aims and scope of the 742. WE-Heraeus-Seminar:

Populations of cancer cells are not static, but keep evolving over the different stages of the disease. This complex evolutionary dynamics is shaped by inherently stochastic contributions (such as mutations and reproductive fluctuations) as well as deterministic components (such as selection for faster growth). For this reason, statistical models are key to modelling and analyzing the evolution of tumours.

Over the last decades, the sequencing of DNA from cancer biopsies has fundamentally changed our understanding of the evolution of cancer. Genomic data sampled across different patients and over time allows to address fundamental questions on how tumours develop, which genetic changes lead to rapid growth, or how a population of diverse tumour cells evolves under cancer therapy. Answering such questions on the basis of empirical data requires statistical models of cancer evolution, both to infer the past dynamics from current data and to derive predictions, for instance on the response to therapy.

This seminar will give an overview over the current state of cancer evolution research and the statistical models used to understand cancer genomic data. It will include tutorials on statistical physics and inference, cancer evolution, and cancer genomics, as well as lectures on the current state of the field by international specialists. The seminar is for MSc and PhD students and young researchers in physics (in particular statistical physics and biophysics), population genetics, and evolutionary biology.

Scientific Organizers:

Prof. Dr. Johannes Berg	Universität zu Köln, Germany E-mail: bergj@uni-koeln.de
Prof. Dr. Martin Peifer	Universität zu Köln, Germany E-mail: mpeifer@uni-koeln.de
Dr. Donate Weghorn	Centre for Genomic Regulation, Barcelona, Spain E-mail: dweghorn@crg.eu

Introduction

Administrative Organization:

Dr. Stefan Jorda
Elisabeth Nowotka

Wilhelm und Else Heraeus-Stiftung
Postfach 15 53
63405 Hanau, Germany

Phone +49 6181 92325-12
Fax +49 6181 92325-15
E-mail nowotka@we-heraeus-stiftung.de
Internet: www.we-heraeus-stiftung.de

Venue:

Physikzentrum
Hauptstrasse 5
53604 Bad Honnef, Germany

Conference Phone +49 2224 9010-120

Phone +49 2224 9010-113 or -114 or -117
Fax +49 2224 9010-130
E-mail gomer@pbh.de
Internetwww.pbh.de

Taxi Phone +49 2224 2222

Registration:

At the Physikzentrum, reception office
Sunday (17:00 h – 21:00 h) and Monday morning

Program

Sunday, 20 March 2022

17:00 – 20:00	Registration	
18:00	*BUFFET SUPPER and informal get-together*	

Monday, 21 March 2022

08:00	*BREAKFAST*	
09:00	Scientific organizers	**Welcome words**
09:10 – 10:10	Johannes Berg	Tutorial: **Population dynamics**
10:10 – 10:30	*COFFEE BREAK*	
10:30 – 11:30	Donate Weghorn	Tutorial: **Selection inference**
11:30 – 12:30	Martin Peifer	Tutorial: **Clonal inference**
12:30 – 12:40	**Conference Photo** (in the front of the lecture hall)	
12:40	*LUNCH*	

Program

Monday, 21 March 2022

14:00 – 15:00	Leonid Mirny	**Limited evidence of tumour mutational burden as a biomarker of response to immunotherapy**
15:00 – 15:30	*COFFEE BREAK*	
15:30 – 16:30?	Philipp Altrock	**Modeling cancer ecology to understand tumor evolution**
16:30 – 17:30	Eytan Domany	**tba**
17:30 – 17:45	Stefan Jorda	**About the Wilhelm and Else Heraeus-Foundation**
19:00	*DINNER*	

Program

Tuesday, 22 March 2022

08:00	*BREAKFAST*	
09:00 – 10:00	Bartlomiej Waclaw	**Darwinian evolution in spatial models of cancer**
10:00 – 10:30	*COFFEE BREAK*	
10:30 – 11:30	Peter Campbell	**tba**
11:30 – 12:30	Peter Van Loo	**Molecular archeology of cancer**
12:30	*LUNCH*	
14:00 – 15:00	Ville Mustonen	**The impact of cell turnover on cancer therapy and evolution**
15:00 – 15:30	*COFFEE BREAK*	
15:30 – 16:30	Noemi Andor	**Characterizing cytotoxic therapy induced shifts in the cost-to-benefit ratio of high ploidy**
16:30 – 17:30	Christina Curtis	**Charting genotype to phenotype maps of tumorigenesis**
17:30 – 18:30	**Online posters with 'elevator pitch'-talks**	
19:00	*DINNER*	

Program

Wednesday, 23 March 2022

08:00	*BREAKFAST*	
09:00 – 10:00	Julie George	**Deciphering tumor evolution in small cell lung cancer**
10:00 – 10:30	*COFFEE BREAK*	
10:30 – 11:30	Nicholas McGranahan	**Exploiting sequencing data to understand cancer evolution**
11:30 – 12:30	Marco Gerlinger	**Genetic and immune landscape co-evolution in colorectal cancers**
12:30	*LUNCH*	
14:00 – 17:30	**Excursion: Hike in the vicinity**	
19:00	*DINNER*	

Program

Thursday, 24 March 2022

08:00	*BREAKFAST*	
09:00 – 10:00	Trevor Graham	**Tracing evolution before cancer in the human colon**
10:00 – 10:30	*COFFEE BREAK*	
10:30 – 12:30	**On site posters, livestreamed via Zoom**	
12:30	*LUNCH*	
14:00 – 15:00	Simon Tavare	**Modeling and simulation of cancer evolution in single cells**
15:00 – 15:30	*COFFEE BREAK*	
15:30 – 16:30	Katerina Politi	**Uncovering mechanisms of drug resistance in lung cancer**
16:30 – 17:30	Quaid Morris	**Reconstructing complex clone trees from multiple bulk DNA samples using Pairtree**
18:00 – 19:00	**Fireside chat**	
19:00	*HERAEUS DINNER* *(social event with cold & warm buffet with complimentary drinks)*	

Program

Friday, 25 March 2022

08:00	*BREAKFAST*	
09:00 – 10:00	Mariam Jamal-Hanjani	**Cancer evolution and the development of metastases**
10:00 – 10:30	*COFFEE BREAK*	
10:30 – 11:30	Niko Beerenwinkel	**Inferring tumor evolution from single-cell data**
11:30 – 12:30	Shamil Sunyaev	**Mechanisms of cancer mutation and what they can tell us about cancer evolution**
12:30	*LUNCH*	
14:00 – 15:00 (Q%A at 14:45)	Charles Swanton	**Cancer evolution, immune evasion and metastases.**

End of the seminar and departure

NO DINNER for participants leaving on Saturday morning

Posters

Posters

Maria Andrianova *online*	**Paradox of inherited mutation in exonuclease domain of POLD – low somatic mutation rate but hypermutable cancers**
Arman Angaji	**Different modes of spatial tumor growth under neutral evolution**
Toby Baker	**Timing complex copy number gains in whole genome duplicated tumours**
Nico Borgsmueller	**Deviations from clock-like evolution found in cancerous and healthy tissue using single-cell DNA-seq data**
Mehdi Damaghi	**Evolution and ecology of metabolic phenotypes to stratify DCIS progression**
Nicola Elisabeth Dick	**Modeling drug resistance: Evolutionary dynamics that shape fitness landscapes of human cancers**
Pedro Ferreira	**Mapping single-cell transcriptomes to copy number evolutionary trees**
Claudia Grimal Bosch	**Modelling synonymous mutations in human cancer**
Carino Gurjao	**Discovery and Features of an Alkylating Signature in Colorectal Cancer**
Luuk Harbers	**High clonal diversity and spatial genetic admixture in early prostate cancer and surrounding normal tissue**
Elif Yaren Itak	**An evolutionary model of cell death in cancer**
Tom Kaufmann	**Inferring the evolution of cancer genomes from multi-sample copy number data including wholegenome doubling with MEDICC2**
Alvaro Köhn-Luque *online*	**Phenotypic deconvolution in heterogeneous cancer cell populations using drug screening data**

Posters

Verena Körber	**Disrupted stem cell hierarchy as a driver of leukemogenesis**
Teemu Kuosmanen	**Turnover flux and its impact on cancer evolution**
Hossameldin Loay Ali	**Simulations show the possible effect of recurrent mutations, out of equilibrium genomic sequence, and multiple mutation events, on the substitution rates and site frequency spectrum.**
Ulrich Michel	**Iterative exhaustion of resistance mechanisms against targeted therapy in an EGFR driven cell line**
Francesca Mignacco *online*	**Statistical physics insights on stochastic gradient descent**
Verónica Miró Pina	**TAFI (Tumor Allele Frequency Interpreter): a new deep learning tool to reveal the evolutionary history of tumores**
Hussein Naji	**Predicting patterns of responses to therapy and relapse of aggressive B-cell lymphomas**
Marcel Schmiel *online*	**Integrative single-cell tracking of genome evolution and tumor cell plasticity in Small Cell Lung Cancer (SCLC)**
Hugh Selway	**In silico testing of hypotheses for lung cancer development**
Vladimir Seplyarskiy *online*	**Mosaic mutations at transcription factor binding sites are enriched in schizophrenia**
Claudia Serrano Colome	**SigNet: three ANN-based tools for extracting and analyzing mutational processes in tumor genomes**
Saumil Shah	**Phenotypic plasticity can contribute to cancer metastasis**

Posters

Maximilian Stammnitz	**The evolution of two transmissible cancers in Tasmanian devils**
Kavan Thakkar *online*	**Biophysical and Computational study of leukemic tumors at Extramedullary sites**
Daniele Tramontano	**Graphical models for cancer evolution.**
Ignacio Vazquez-Garcia *online*	**Genomic instability as a determinant of tumor-immune co-evolution in ovarian cancer**
Haixi Yan	**Chromosomal instability drives spatial and temporal phenotypic diversity in Schwann cancer cells**

Abstracts of Talks

(in alphabetical order)

Modeling Cancer Ecology to Understand Tumor Evolution

Philipp M. Altrock[1]

[1]*Max Planck Institute for Evolutionary Biology, Plön, Germany*

The accumulation of genetic or epigenetic changes over a lifetime is called somatic evolution. This process inevitably occurs in multicellular organisms and can lead to disease development. The conventional dogma of tumor evolution postulates that the observed genomic changes are due to the acquisition of driver mutations that improve cell-intrinsic fitness independent of context. In contrast, increasing evidence suggests that cell-phenotypic diversity and spontaneous changes in the environment lead to context-dependent fitness differences. The resulting selection is driven by inflammation in the tumor microenvironment or the exchange of growth factors that act as public goods. This environment includes stroma cells, immune cells, and signaling molecules. An important question is how selection in this complex environment leads to disease progression. The tumor (micro)environment provides a powerful adaptive force that codetermines the fate of the disease and could inform treatment options. On can integrate biological and clinical data with mathematical modeling, to ask how stochastic effects and nonlinear selection dynamics operate in concert to drive disease onset, progression, and shape observed cellular diversity. The resulting models can improve our understanding of the roles of selection, temporal variation, and spatial variation during cancer evolution.

Characterizing cytotoxic therapy induced shifts in the cost-to-benefit ratio of high ploidy

Richard Beck, Gregory J. Kimmel, Andriy Marusyk, Daria Miroshnychenko, Andrew Schultz, Thomas Veith, Samuel Bakhoum, Xiaoqing Yu, Philipp M. Altrock, Ana Gomes, Noemi Andor

Analyses of intra-tumor heterogeneity across multiple cancer types suggest that tumor cell fitness declines once aneuploidy exceeds a certain limit[1–3]. A significant difference in outcome between tumors above and below the limit however is only evident among therapy-naïve patients, not among patients who subsequently underwent cytotoxic therapy[1]. *Our hypothesis is that the context-dependent ambivalence of high ploidy is what accounts for both of these observations*. On one hand, high ploidy ameliorates the deleterious effects of missegregation-induced genome-dosage imbalances, on the other hand a high ploidy cell has higher energetic demands as it has to replicate and express more genetic material. We performed a series of in-vitro and in-silico experiments to quantify both, (i) the costs and (ii) benefits of high ploidy. We developed and used mathematical models to predict differences in S phase duration between high and low ploidy cells (i), and to evaluate the possibility of mis-segregation induced population extinction (ii). Model predictions include critical curves that separate viable from non-viable populations as a function of their turnover- and mis-segregation rates. Missegregation- and turnover rates estimated for nine cancer types are then compared to these predictions for various biological assumptions.

For (i), we evaluated three key building blocks of dNTP synthesis –PO_4, O_2 and Glucose– as candidate cell-extrinsic resources that cap ploidy in Glioblastoma and stomach cancers. We predict that at limiting dNTP concentrations, high-ploidy cells will take longer to replicate their DNA than low-ploidy cells. In-vitro experiments support these predictions showing that PO_4 depletion imposes a higher fitness cost on near-tetraploid than on near-diploid breast cancer cells. For (ii), the majority of tumors across all nine cancer types had missegregation- and turnover rates that were within viable regions of the parameter space. When a dependency of mis-segregation rate on ploidy was introduced, ploidy states associated with low mis-segregation rates rendered MIE impossible at low turnover rates. Exposing a heterogeneous stomach cancer cell line to the microtubule-targeting drug Vinblastine confirmed that the high ploidy subpopulation had a fitness advantage.

If our hypothesis is true, the implications are broad. It would explain vast differences in the extent of inter-tumor karyotype heterogeneity. It may contribute to explain why agents that block dNTP production work well in combination with DNA damaging agents. It would also explain differences in ploidy across different primary and metastatic sites. As solid tumors progress, resources in the tumor microenvironment become scarcer than the resources available in normal surrounding tissues. These resource-poor environments may push high-ploidy cells to leave the primary tumor into circulation and thrive at locations with abundant access to nutrients. Understanding the resource cost of high ploidy can help uncover its therapeutic vulnerabilities across tissue sites with versatile energy supplies.

References

1. Andor, N. *et al.* Pan-cancer analysis of the extent and consequences of intratumor heterogeneity. *Nat. Med.* **22**, 105–113 (2016).
2. Birkbak, N. J. *et al.* Paradoxical relationship between chromosomal instability and survival outcome in cancer. *Cancer Res.* **71**, 3447–3452 (2011).
3. Roylance, R. *et al.* Relationship of extreme chromosomal instability with long-term survival in a retrospective analysis of primary breast cancer. *Cancer Epidemiol. Biomark. Prev. Publ. Am. Assoc. Cancer Res. Cosponsored Am. Soc. Prev. Oncol.* **20**, 2183–2194 (2011).

Inferring tumor evolution from single-cell data

N. Beerenwinkel

ETH Zurich, Department of Biosystems Science and Engineering, Basel, Switzerland

Cancer progression is an evolutionary process characterized by the accumulation of genetic alterations and responsible for tumor growth, clinical progression, and drug resistance development. We discuss how to reconstruct the evolutionary history of a tumor from single-cell sequencing data and present probabilistic models and efficient inference algorithms for mutation calling and learning tumor phylogenies from mutation and copy number data [1]. We present methods for integrating single-cell DNA and RNA data obtained from tumor biopsies [2] and for detecting common patterns of tumor evolution among patients, such as re-occurring evolutionary pathways and clonally exclusive mutations [3].

References

[1] https://doi.org/10.1101/2022.01.06.475205
[2] https://doi.org/10.1101/2021.11.04.467244
[3] https://doi.org/10.1101/2021.11.04.467347

Introduction to Population Genetics

Johannes Berg

Institute for Biological Physics, University of Cologne,
Zülpicher Str. 77a, 50937 Cologne, Germany

This tutorial gives a brief introduction to the basics of population genetics from the perspective of cancer evolution. We discuss the fundamental evolutionary forces of mutations, selection, and genetic drift as well as the signature of neutral evolution. We will discuss both the standard scenario of a population of constant size as well as the feature that defines the population genetics of cancer: a population that, for at least part of its lifetime, grows exponentially in time.

This tutorial has two companion tutorials on selection inference (D. Weghorn) and clonal inference (M. Peifer).

Deciphering Tumor Evolution in Small Cell Lung Cancer

Julie George[1,2]

[1]Department of Translational Genomics, Medical Faculty, University of Cologne, Weyertal 115b, 50931 Cologne, Germany;
[2]Department of Otorhinolaryngology, Head and Neck Surgery, Medical Faculty, University of Cologne, Germany
E-Mail: jgeorge@uni-koeln.de

Small cell lung cancer (SCLC) is a highly aggressive tumor that accounts for 15% of all lung cancers. SCLC patients are commonly diagnosed at a late stage, which complicates surgical resections. Therefore, the current standard of care is a platinum-based chemotherapy, and recently a combination with anti-PD-L1 therapy[1]. SCLC patients initially respond well to this treatment, but quickly relapse with a therapy-resistant tumor. Targeted therapies have not been identified, which leaves no other effective treatment options. The clinical performance quickly deteriorates at the time of relapse, resulting in a poor 5-year overall survival rate of less than 2%.

We have performed comprehensive genome sequencing of SCLC tumors to identify key genomic events for SCLC[2]. As a next step, we aimed to shed light on the dynamics of tumor progression in SCLC. We performed multi-regional and longitudinal genome sequencing of >160 tumor specimens obtained from >60 SCLC patients to study tumors at the time of first diagnosis and throughout therapy. We applied computational approaches to determine individual clones from tumor sequencing data. We reconstructed phylogenetic trees for all patient cases, and our data provides insights into patterns of tumor progression as a consequence of therapy.

[1] Leora Horn et al., "First-Line Atezolizumab plus Chemotherapy in Extensive-Stage Small-Cell Lung Cancer," *New England Journal of Medicine*, 2018, NEJMoa1809064, https://doi.org/10.1056/NEJMoa1809064.
[2] Julie George et al., "Comprehensive Genomic Profiles of Small Cell Lung Cancer," *Nature* 524, no. 7563 (2015): 47–53, https://doi.org/10.1038/nature14664.

Genetic and immune landscape co-evolution in colorectal cancers

Marco Gerlinger[1,2]

[1]*Barts Cancer Institute, Queen Mary University of London, UK*
[2] *St Bartholomew's Hospital GI Cancer Centre, London, UK*

Immune cell infiltrates in colorectal cancers (CRC) are important for immunotherapy responses and they have prognostic relevance. We investigated how Darwinian evolution and immune landscape evolution are linked in MMR proficient CRCs during treatment with anti-EGFR antibodies, and during cancer progression in hypermutated MMR deficient colorectal cancers. EGFR antibodies promoted the evolution of genetic resistance drivers in the RAS-RAF pathway as well as non-Darwinian resistance evolution through microenvironmental remodeling. Microenvironment changes encompassed an increase in T-cell infiltrates and expression of immune checkpoints and immunotherapy response signatures. This may open new opportunities for targeting MMR proficient colorectal cancers with immunotherapy which we are now testing in a clinical trial. The hypermutator phenotype in MMR deficient CRCs leads to very high mutation loads and immunogenicity. Using multi-region sequencing, we found a clear hierarchy of driver evolution despite pervasive intratumour heterogeneity. Immune evasion drivers were predominantly subclonal and showed parallel evolution. Phylogenetic analysis furthermore identified three patterns of immune evasion driver evolution (pan-tumor evolution, subclonal evolution, and evolutionary stasis) and these differed in their T-cell infiltrates. These distinct patterns of co-evolution demonstrate a dynamic interplay between immune cells and Darwinian evolution. Our results finally reinforce the importance of clonality and evolution subtype assessments for immunotherapy biomarker development.

Tracing evolution before cancer in the human colon

Trevor Graham

Centre for Evolution and Cancer, Institute of Cancer Research

Colon cancers arise from mutated stem cells, and so the dynamics of stem cell evolution in healthy colon likely influence cancer risk. In the first part of the talk, I will describe a novel method called “flip flop” that measures stochastic fluctuations in DNA methylation to quantitatively measure stem cell evolution[1]. Applying flip flop we find only slight differences in stem cell evolutionary dynamics between human colon and small intestine that are too small to explain the large difference in cancer risk between the two organs. In patients with germline APC mutations, the first genetic hit on the road to cancer, we observe increase stem cell numbers. In the second part of the talk, I will describe pre-cancer evolution in patients with inflammatory bowel disease. We observe widespread clonal expansion of aneuploid clones through inflamed tissue, and show that the detection of aneuploidy in inflamed colon strongly predicts the risk of future cancer development.

[1] doi: 10.1038/s41587-021-01109-w
[2] doi: 10.1136/gutjnl-2018-316191

Cancer evolution and the development of metastases

Mariam Jamal-Hanjani

University College London Cancer Institute, 72 Huntley Street, Paul O-Gorman Building, London, WC1 6DD, United Kingdom

Studies of cancer evolution to date have often relied on archival tissue surplus to diagnostic requirements from living patients obtained in early stage disease, and less commonly in relapsed or metastatic disease. Research autopsies provide unprecedented access to tissue samples that would otherwise often be inaccessible due to the site of disease and the invasive nature of sampling. PEACE is a pan cancer national autopsy programme that allows access to matched primary and metastatic tumours and aims to investigate the biological processes underpinning metastatic disease, including genomic drivers of tumour dissemination and failure of the adaptive immune system. Using phylogenetic analyses in PEACE, tumour evolutionary histories can be reconstructed and clonal relationships across anatomical sites inferred. Furthermore, clonal lineage and migration histories can be inferred to demonstrate the complex patterns of metastatic seeding that can exist. Whilst bulk tissue tumour sequencing shows widespread heterogeneity in somatic copy number alterations (SCNAs), single cell sequencing technologies have the ability to reveal with greater resolution allele-specific SCNAs in subclones at the single cell level, and shed light on the true extent of SCNA heterogeneity. Finally, detailed clinical annotation of patient histories and multi-region samples are crucial to map the evolution of primary to metastatic cancer, in particular, radiological imaging which can track the development and progression of tumours and focus genomic analyses on specific sites of metastases that may help identify the mechanisms of drug resistance and immune evasion.

Exploiting sequencing data to understand cancer evolution

N. McGranahan

Cancer development within an individual is an evolutionary process. This has important clinical implications for cancer prevention and therapy, as well as our understanding of cancer progression and metastatic spread.

In this talk, I will outline how we can exploit cancer genomic sequencing data to decipher cancer evolutionary histories and the extent of diversity within individual tumours. I will explore the importance of large-scale genomic events, including whole genome doubling, in shaping cancer evolution. I will consider the extent to which chromosomal instability is ongoing during cancer evolution.

Finally, I will explore how we can use novel bioinformatics tools to shed light on the interface between the cancer cell and the immune microenvironment, and mechanisms of immune escape. I will explore how DNA sequencing data can be harnessed to identify T cells in tumour samples, and the clinical relevance of T cell infiltrate in predicting response to immunotherapy.

Limited evidence of tumour mutational burden as a biomarker of response to immunotherapy

Carino Gurjao[2,1], and Leonid A. Mirny[1,2]

[1]Institute for Medical Engineering and Science, and Department of Physics, MIT, Cambridge, MA, USA
[2] Department of Medical Oncology, Dana-Farber Cancer Institute and Harvard Medical School, Boston, MA, USA

Cancer immunotherapy by immune checkpoint blockade (ICB) is effective for several cancer types, however, its clinical use is encumbered by a high variability in patient response. Several studies have suggested that Tumour Mutational Burden (TMB) correlates with patient response to ICB treatments, likely due to immunogenic neoantigens generated by novel mutations accumulated during cancer progression. Association of TMB and response to checkpoint inhibitors has become widespread in the oncoimmunology field, within and across cancer types, and has led to the development of commercial TMB-based biomarker platforms. Furthermore, patient prioritization for ICB based on individual TMB level was recently approved by the FDA. Here we revisit the association of mutational burden with response to checkpoint inhibitors by aggregating the largest pan-cancer dataset with more than 2500 ICB-treated patients with sequencing data and clinical annotation. Surprisingly, we find little evidence that TMB is predictive of patient response to immunotherapy. Our analysis suggests that previously reported associations arise from a combination of confounding disease subtypes and incorrect statistical testing. We show that using a TMB threshold for clinical decisions regarding immunotherapy could skew access to treatment for patients who may benefit from these therapies. Finally, we present a simple mathematical model that extends the neoantigen theory, is consistent with the lack of association between TMB and response to ICB and highlights the role of immunodominance. Our analysis calls for caution in the use of TMB as a biomarker and emphasizes the necessity of continuing the search for other genetic and non-genetic determinants of response to immunotherapy.

Reconstructing complex clone trees from multiple bulk DNA samples using Pairtree

J.A. Wintersinger[1,2,3], S.M. Dobson[1,4], E. Kulman[5], L.D. Stein[1,2], J.E. Dick[1,4], and Q. Morris[1,3,5]

[1]*University of Toronto, Toronto, ON, Canada*
[2]*Ontario Institute for Cancer Research, Toronto, ON, Canada*
[3]*Vector Institute for Artificial Intelligence, Toronto, ON, Canada*
[4]*Princess Margaret Cancer Centre, Toronto, ON, Canada*
[5]*Memorial Sloan Kettering Cancer Center, New York City, NY, USA*

Cancers are composed of genetically distinct subpopulations of malignant cells. DNA sequencing data can be used to determine the somatic point mutations specific to each population and build *clone trees* describing the evolutionary relationships between them. These clone trees can reveal critical points in disease development and inform treatment. Pairtree[1] is a new method that constructs more accurate and detailed clone trees than previously possible using variant allele frequency data from one or more bulk cancer samples. It does so by first building a *Pairs Tensor* that captures the evolutionary relationships between pairs of subpopulations, and then it uses these relations to constrain clone trees and infer violations of the infinite sites assumption. Pairtree can accurately build clone trees using up to 100 samples per cancer that contain 30 or more subclonal populations. On 14 B-progenitor acute lymphoblastic leukemias, Pairtree replicates or improves upon expert-derived clone tree reconstructions.

Clone trees illustrate the evolutionary history of a cancer and can provide insights into how the disease changed through time (e.g., between diagnosis and relapse). Pairtree uses DNA sequencing data from many samples of the same cancer to build more detailed and accurate clone trees than previously possible.

References

[1] **J.A. Wintersinger, et al.** Blood Cancer Discovery, accepted (2022)

The impact of cell turnover on cancer therapy and evolution

V. Mustonen[1]

[1] *Organismal and Evolutionary Biology Research Programme, Department of Computer Science, Institute of Biotechnology, University of Helsinki, 00014 Helsinki, Finland*

A tumour grows when the total birth rate of its cells exceeds their total death rate. We study mathematically how to best impact such evolutionary dynamics by cytotoxic (increasing death rate) or cytostatic (decreasing birth rate) therapy. Comparing the treatments we derive conditions for choosing optimal therapy. In particular, we quantify how the choice and the efficacy gain of optimal therapy depends on driver mutation, metastasis, intrinsic cell birth and death rates, and the details of cell competition [1]. We then consider optimal growth strategies from the perspective of the target cell population using methods from optimal control theory [2]. Finally, we consider how growth strategies themselves are under evolutionary pressure that depends on the tumour micro-environment [3]. In summary, we show that detailed understanding of the cell population dynamics could be exploited in choosing the right mode of treatment with substantial therapy gains. The talk is based on collaborative work [1-3].

References

[1] J. V. Anttila, M. Shubin, J. Cairns, F. Borse, Q. Guo, T. Mononen, I. Vazquez-Garcia, O. Pulkkinen, V. Mustonen PLOS Comput Biol **15(11)**:1–18 (2019)

[2] T. Mononen, T. Kuosmanen, J. Cairns, V. Mustonen, *submitted* (2022)

[3] T. Kuosmanen, S. Särkkä, V. Mustonen, *in preparation* (2022)

Clonal Inference

Martin Peifer

Translational Genomics, University of Cologne, Cologne

Tumors form distinct clonal (sub)populations during their evolutionary time. Studying the development of these populations can provide a better understanding of therapy response and failure. Current DNA sequencing techniques are powerful tools to infer clonal populations even from bulk tumor data. In this tutorial, I will discuss how this inference is performed on bulk tumor whole exome and genome sequencing data. I will further discuss common obstacles and pitfalls that can lead to spurious results, which can finally disturb or lead to unresolvable phylogenetic relationships of the inferred clones. Finally, I will give a short introduction into clonal inference from current single-cell sequencing approaches.

Uncovering Mechanisms of Drug Resistance in Lung Cancer

Katerina Politi

Departments of Pathology and Medicine (Section of Medical Oncology)
Yale School of Medicine & Yale Cancer Center
New Haven, CT USA 06510

Targeted therapies and immunotherapies have transformed the landscape for the diagnosis and treatment of metastatic lung cancer. These tumors are now routinely tested for the presence of mutations or rearrangements in specific oncogenic drivers that, if present, predict sensitivity to targeted therapies directed to the genomic alterations present. Genotype-directed therapies have improved outcomes in specific subsets of patients with metastatic lung cancer. In the absence of specific actionable alterations, many metastatic lung cancers are treated with immune checkpoint inhibitors alone or in combination with chemotherapy. Despite the success of these agents, targeted therapies and immunotherapies are rarely curative and acquired resistance is a major impediment to cures for patients treated with these therapies. Moreover, there is heterogeneity in the durability and depth of responses between patients and at different metastatic sites within individual patients. Here we describe studies to address these issues using mouse models and studies of patient specimens.

To investigate mechanisms of acquired resistance to therapy we established a tissue collection program that allows us to compare tumors at the time of acquired resistance to pre-treatment specimens. Analysis of a cohort of cases of patients treated with immune checkpoint inhibitors who developed acquired resistance to these agents revealed the presence of defects in β2-microglobulin an essential component of the MHC I antigen presentation machinery. Additional alterations in genes in this pathway were also identified for which the functional consequences are unknown. Integration of genomic, transcriptomic and functional studies to understand the impact of alteration in MHC I antigen presentation on response to immunotherapies will be discussed.

A paradigm for the success of targeted therapies in lung cancer, comes from *E*pidermal *G*rowth *F*actor *R*eceptor (*EGFR)* mutant lung cancer. Mutations in exons encoding the tyrosine kinase domain of EGFR confer sensitivity to tyrosine kinase inhibitors (TKIs) and four TKIs (erlotinib, gefitinib, afatinib and, most recently, osimertinib) are currently approved for the first-line treatment of *EGFR* mutant lung cancer. Acquired drug resistance, however, is a major challenge with all of these TKIs and especially for osimertinib we have very limited knowledge of the mechanisms of resistance given its recent adoption in the clinic. Without knowledge about resistance mechanisms, optimal post-osimertinib treatment strategies remain to be defined. We have developed a novel mouse model to study the consequences of co-occurring tumor suppressor gene alterations on the progression and TKI sensitivity of *EGFR* mutant tumors. Data from these models can be integrated with studies from patient specimens to study tumor progression and response to TKIs.

Collectively, our studies provide new insights and integrative approaches to study drug resistance.

Modeling and simulation of cancer evolution in single cells

Khanh Dinh[1] and Ignacio Vazquez-Garcia[1,2] and Simon Tavaré[1]

[1]Irving Institute for Cancer Dynamics, Columbia University, New York, USA
[2] Memorial Sloan Kettering Cancer Center, New York, USA

Following Nowell's [1] formalization of tumor evolution as a series of clonal expansions, many authors have used bulk sequencing data to infer the nature of these expansions; for example [2-4].

Recent advances in single-cell whole genome sequencing enable profiling of copy number aberrations at high resolution in thousands of cells [5]. Single-cell genomics data generated by such technologies enables quantitative measurements of tumor dynamics in space and time [6], and measurements of the rate of chromosomal aneuploidy, whole-genome duplications and replication errors in tumors.

We have developed a simulation algorithm for studying single-cell dynamics in a population of cells, incorporating somatic copy number changes, clonal selection of driver mutations and accumulation of neutral passenger mutations. The simulator follows population dynamics as input by the user, generates the clonal evolution forward in time, where clones are defined by their copy number and driver mutation profiles. The phylogeny of a sample is then computed backward in time. The algorithm is designed to be efficient for large cell populations while maintaining statistical accuracy.

We present two examples from the simulator package. The first follows the neutral evolution of copy number events in the population of epithelial cells in the fallopian tube. The second investigates the evolution of high-grade serous ovarian cancer (HGSOC) driven by genomic instability. The simulator may also be used to calibrate clonal reconstruction algorithms used on single-cell DNA sequencing data.

References

[1] Nowell, P. Science **194**, 23-28 (1976)
[2] Sottoriva, A. et al. Nature Genetics **47**, 209-216 (2015)
[3] Williams, M. Nature Genetics **48**, 238-244 (2016)
[4] Dinh, K. et al. Statistical Science **34**, 129-144 (2020)
[5] Laks, E. et al. Cell **179**, 1207–21.e22 (2019)
[6] Salehi, S. et al. Nature **595**, 585-590 (2021)

Molecular archeology of cancer

Peter Van Loo[1,2]

[1]*The University of Texas MD Anderson Cancer Center, Houston, USA*
[2]*The Francis Crick Institute, London, UK*

Tumor development is driven by changes to the genome and epigenome leading to fitness advantages underlying successive clonal expansions. As somatic genetic and epigenetic changes occur across most or all cell cycles, the cancer (epi)genome carries an archeological record of its past. Over the past years, we have developed several approaches to mine that archeological record from the cancer genome, which we collectively call 'molecular archeology of cancer'. Using these approaches, we are able to infer the subclonal architecture of tumors, and gain key insights into the order and timing of the genomic changes that occurred over their evolutionary history. We have applied these approach in a large-scale pan-cancer setting, showing that intra-tumor heterogeneity is pervasive across cancers, and that the timelines of tumor evolution span multiple years to decades, with typically similar key driver events occurring early. We have recently also started to develop methods that leverage the cancer epigenome to reconstruct the subclonal architecture of cancer, which will open up new avenues to study tumor evolution.

References

[1] Gerstung et al., Nature 578, 122-128 (2020)
[2] Dentro et al., Cell 184:2239-2254 (2021)
[3] Demeulemeester et al., Nat. Genet. 54, 128-133 (2022)
[4] Larose Cadieux et al., bioRxiv doi:https://doi.org/10.1101/2020.11.03.366252 (2020)

Darwinian evolution in spatial models of cancer

B. Waclaw[1,2]

[1]Dioscuri Centre for Physics and Chemistry of Bacteria, Institute for Physical Chemistry PAS, Warsaw, Poland
[2] School of Physics and Astronomy, The University of Edinburgh, Edinburgh, United Kingdom

Mathematical modelling of cancer has a long history but it traditionally focused on replicating growth laws observed for different tumour types, the role of angiogenesis, or predicting the outcome of chemotherapy. Recently, advances in genomics have made it possible to investigate Darwinian evolution in populations of cancer cells. This has opened up many interesting questions. In particular, as the cancerous tumour grows, cells accumulate further mutations. Are these mutations neutral “passengers” (i.e. they do not change the net growth rate) or are some of them “driver mutations” that increase the growth rate? Is there evidence of selection acting on certain traits of cancer cells? How genetically diverse a typical tumour is? How is evolution affected by the spatial structure of the tumour? In this talk I will discuss how computer models can be used to shed light on these questions. I will show how different processes: replication, death, migration, and mechanical interactions between cells in a tumour affect its structure and genetic composition. I will also discuss how different models compare to experimental data, and their implications for cancer therapy.

Selection inference

Donate Weghorn

Centre for Genomic Regulation (CRG), Barcelona, Spain

Cancer is a genetic disease, elicited by so-called driver mutations. These mutations evolve under positive selection pressure, as they confer a growth advantage to the cell they arise in. On the other hand, we may expect to find negative selection acting on loci that the tumor cannot afford to lose to functionally impactful mutations. Many methods have been developed to estimate selection pressures in cancer tumors, both for protein-coding genes as well as non-coding genomic elements. One crucial aspect of all of these methods is to account for the high variability of mutation probability as a function of genomic position and sequence context. In this primer, I will discuss the sources of this variability of mutation probability, different approaches to modeling it and, finally, how these approaches are used in selection inference algorithms.

Abstracts of Posters

(in alphabetical order)

Paradox of inherited mutation in exonuclease domain of POLD – low somatic mutation rate but hypermutable cancers

M. Andrianova[1] and V. Seplyarskiy[2]

[1]The Skolkovo Institute of Science and Technology, Moscow,Russia
[2] Harvard Medical School, Boston, USA

Two main DNA polymerases (POLE and POLD) replicate DNA in eukaryotic cells. Extensive sequencing of cancer samples revealed highly mutable cancers with mutations in exonuclease domains (ED) of POLE or POLD. Interestingly samples with somatic mutation in POLD always have deficiency in mismatch repair system. Thus, direct influence of mutation in ED of POLD on mutation rate and cancer development was unclear. Recently it was shown that germline mutations in POLE drastically increase mutation rate (~7-fold) when the effect of POLD mutations is quite moderate (up to 3 fold) [1]. The germline variant L474P in POLD showed the lowest influence on mutation rate (~1.2 fold) however, it is strongly associated with colorectal cancer [2]. To explore this contradiction, we sequenced extended family with L474P variant and familial cancer. From our data, we showed that despite negligible effect on overall number of mutations L474P variant strongly alters mutational spectra. In tumor from the same family, we observed very high mutation rate and mutational spectrum different from spectrum for somatic mutations of the same individual. Similar pattern was previously observed in hypermutable adenomas from individual with another germline POLD mutation [1]. We suggested that probably heterozygous mutation in ED of POLD is not sufficient for cancer development but increases the probability of additional event accelerating mutation rate. To our surprise, we found that sequenced tumor has copy-neutral loss of heterozygosity (LOH) leading to homozygous L474P mutation. Indeed, LOH was previously observed in other tumors in families with germline POLD mutations [3] and recent experimental paper also observed recessive effect of mutation in ED of POLD in yeasts (Zhou et al., 2021).

References

[1] Philip S. Robinson et al., J Nature Genetics **53**, 1434–1442 (2021)
[2] Laura Valle et al., Human Molecular Genetics **23**, 3506–3512 (2014)
[3] Claire Palles et al., Nature Genetics **45,** 136–144 (2013)
[4] Zhi-Xiong Zhou et al., Nature Structural & Molecular Biology **28**, 1020–1028 (2021)

Different modes of spatial tumor growth under neutral evolution

A. Angaji[1], M. Owusu[3], C. Velling[2], N. Dick[3], D. Weghorn[3], J. Berg[1]

[1]Institute for Biological Physics, University of Cologne, 50825 Cologne, Germany
[2] Kishony Lab, Faculty of Biology, Technion City, 32000 Haifa, Israel
[3]Cenctre for Genomic Regulation, 08003 Barcelona, Spain

Cancer development is an evolutionary process of somatic mutation, selection, and genetic drift which shapes the genetic profile of a tumor. Next-generation sequencing technology has revealed high intra-tumor heterogeneity at the genomic level. This genetic heterogeneity also extends to the spatial dimension of clonal architecture, but only recently has high-depth multi-region sequencing data become available [1]. We use spatially resolved genetic data to study a tumor's mode of evolution and growth as well as population genetic theory of mutant type extinction to infer cancer cell turnover rates and mutation rates under the null model of neutral evolution [2].

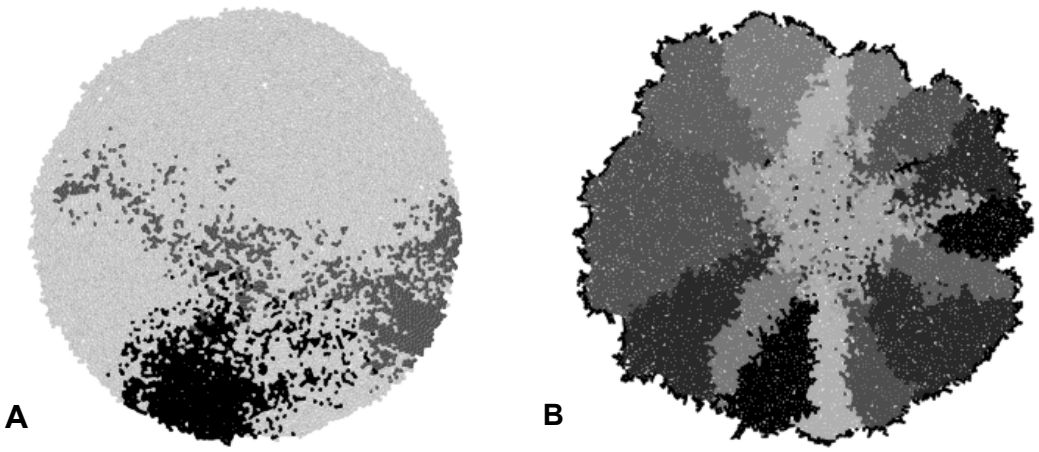


Fig: A) Mutations are highly dispersed without need of active migration under bulk driven growth. **B)** Clones show directed, outward bias under surface driven growth.

References

[1] Ling S *et al* 2015, Extremely high genetic diversity in a single tumor points to prevalence of non-Darwinian cell evolution, Proc. Natl Acad. Sci. USA 112 E6496–505

[2] Angaji A *et al* 2021, Stochastic clonal dynamics and genetic turnover in exponentially growing populations, J. Stat. Mech. 103502

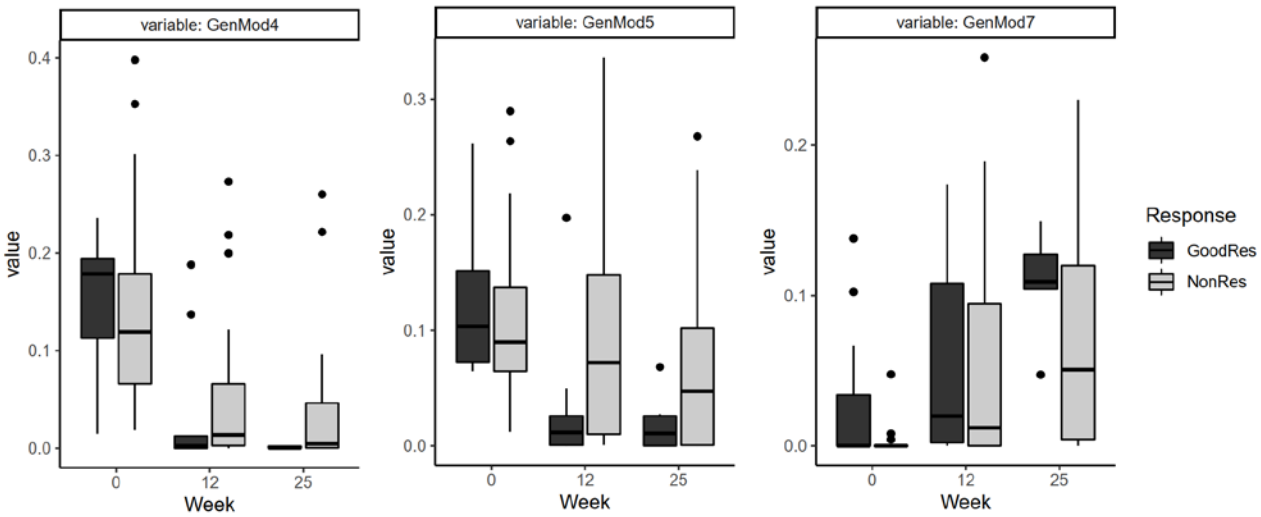
The role of phenotypic diversity in resistance and their response to therapy

Youness Azimzade, Alvaro Köhn-Luque and Arnoldo Frigessi

Institute of Basic Medical Sciences, University of Oslo, Oslo, Norway

As diversity of tumors becomes more and more clear, the quest to understand how subpopulations inside tumor participate into response to drugs becomes more and more.

Using bulk gene expression profiles for breast tumor cells before and during therapy, we have been able to identify the frequency of subpopulations and their dynamics of during therapy. Interestingly, the population composition and their evolution is different between good responders and non-responders.



Figur 1. Fraction of three subpopulations (GenMod4, GenMod5 and GenMod7) vs time for Luminal B breast tumors.

We are trying to develop a simple mathematical model that can regenerate the observed data. Using this model, we will explore alternative therapeutic schedules to see if with changes the combination of the drugs and their timing/dosing; we can regulate dynamics of these subpopulations and patient outcome.

Timing Complex Copy Number Gains in Whole Genome Duplicated Tumours

Toby Baker[1], Stefan Dentro[2], Paul Spellman[3], Maxime Tarabichi[1,4] and Peter Van Loo[1,5]

[1]Francis Crick Institute, London, UK
[2] DKFZ, Heidelberg, Germany
[3] OHSU, Portland OR, USA
[4] IRIBHM Université Libre de Bruxelles, Brussels, Belgium
[5] MD Anderson Cancer Center, Houston, TX, USA
Email toby.baker@crick.ac.uk

The accumulation of genetic copy number aberrations is a process common to the development of many cancers. Tumours that have undergone a whole genome duplication are known to have high degrees of genomic instability and often contain genomic regions that have undergone a series of genomic gains resulting in multiple copies of both alleles.

The relative timing of copy number gains in whole genome duplicated tumours is complicated by the fact that there are multiple plausible evolutionary histories for the tumour that could give rise to them. Typically, the most parsimonious history, the one with the fewest number of events is assumed to take place[1]. Here we detail a method, GRITIC, that identifies the route taken and the timing of such gains in whole genome duplicated samples by leveraging the simultaneous occurrence of the whole genome duplication in the sample. On a representative simulated cohort, it accurately assigns route history and timing to complex copy number gains.

By applying the method to 3,262 whole genome duplicated metastatic and primary tumour samples from the Pan-Cancer Analysis of Whole Genomes[2] and Hartwig Medical Foundation[3] datasets we find that the principle of maximum parsimony is frequently violated with large numbers of copy number gains occurring earlier than thought under this assumption.

GRITIC allows for a more accurate inference of evolutionary histories in different cancer types and better insights into the early genomic events that occur in tumours that go on to undergo a whole genome duplication.

References

[1] M. Gerstung, Nature **578**, 122-128 (2020)
[2] The ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium, Nature **578**, 82-93 (2020)
[3] P. Priestly, Nature **575**, 210-216 (2019)

Deviations from clock-like evolution found in cancerous and healthy tissue using single-cell DNA-seq data

Nico Borgsmüller[1,2,†], Monica Valecha[3,4,5,†], Jack Kuipers[1,2], Niko Beerenwinkel[1,2] and David Posada[3,4,5,‡]

[1]Department of Biosystems Science and Engineering, ETH Zürich, 4058 Basel,Switzerland

[2]SIB, Swiss Institute of Bioinformatics, 4058 Basel, Switzerland
[3]CINBIO, Universidade de Vigo, 36310 Vigo, Spain
[4]Department of Biochemistry, Genetics, and Immunology, Universidade de Vigo, 36310 Vigo, Spain

[5]Galicia Sur Health Research Institute (IIS Galicia Sur), SERGAS-UVIGO.

[†]These authors contributed equally to this work.
[‡]To whom correspondence should be addressed.

The mode of tumor evolution impacts cancer progression, therapy response, and relapse. To what extent mutations with a selective advantage or random sampling of neutral mutations drive tumor evolution is, however, still under debate. One explanation for neutral evolution is a molecular clock, implying that mutations occur with a fixed rate, figuratively with every tick of the clock. Here we test the existence of a molecular clock for the first time with single-cell DNA sequencing data. We developed and benchmarked a novel statistical test and applied it to 26data sets from cancerous and healthy tissue. In 18 cases, we rejected a molecular clock, with no rejection in two normal, two polyp, and three cancerous samples. Our findings show that somatic evolution in cancerous and, notably, also non-cancerous tissues generally does not follow clock-like patterns, with selection being one possible explanation.

Evolution and ecology of metabolic phenotypes to stratify DCIS progression

M. Damaghi[1,2,3]

[1]Stony Brook Cancer Center, Stony Brook, New York, USA
[2] Department of Pathology, Stony Brook University, Stony Brook, New York, USA
[2] Department of Applied Mathematics & Statistics, Stony Brook University, Stony Brook, New York, USA

At the moment, it is not possible to distinguish at diagnosis ductal carcinoma in situ (DCIS) patients whose lesions will develop into invasive cancer from those stay indolent. Despite variability in outcome, DCIS patients are all treated the same, which results in significant over-treatment for individuals who are at lower risk. Therefore, it would be valuable to have new biomarkers and techniques that, at screening, identify patients under or over-treatment. Such research strategies and techniques may spare treatment to some and morbidity to others.

Tumors are evolving ecosystems that undergo many steps of selection. In early breast cancer microenvironmental selection factors include acidosis, hypoxia, and nutrient deprivation, as well as other cells in the regional niche (other cancer cells, stromal cells, or immune cells) define the ecosystem. This highly variable ecosystem selects for highly adaptive cancer cells that are more aggressive. One good example of these metabolic phenotypes is Warburg Effect (WE) accompanied by acidic microenvironment. We have shown that WE and acid phenotype is correlated to aggressive phenotype and metastasis in mouse and human. In this proposal we will use these phenotypes at early stage of breast cancer to predict its evolution to upstaged and metastasis. Evolutionary theory predicts that micro domains or habitats will have differential evolutionary trajectories. To define habitats, we will quantify the frequency and distribution of WE phenotype of cancer and stromal cells using advanced multiplex immunohistochemistry along with novel neighborhood analyses of landscape ecology using machine learning and pathomics to identify DCIS lesions that are likely to progress to invasive disease. We hypothesize that WE habitats select for the emergence of aggressive DCIS lesions and hence would define evolutionary path of the tumor that will also provide important decision support. Our proposal will also. develop approaches to study the evolutionary process of breast cancer in both animal models of DCIS and patients that will not only illuminate novel biology but can be used for personalized therapy.

Modeling drug resistance: Evolutionary dynamics that shape fitness landscapes of human cancers

D. Weghorn[1] and N. Dick[1,2]

[1]Centre for Genomic Regulation, Barcelona, Spain
[2] Universitat Pompeu Fabra, Barcelona, Spain
Email: ndick@crg.es

A common problem in cancer treatment is the development of drug resistance.

From a population genetics perspective the therapeutic pressure on a cancer cell population can be seen as a shift in its fitness landscape. With this, the application of Fishers fundamental theorem leads to the prediction that tumors with higher variablity in cellular fitness will recover from the attack more quickly and/or with a higher probability[1]. By modelling tumor growth and drug response in numerical simulations we aim to study this hypothesis and deepen our understanding of the interplay between different modes of selection that lead to therapy resistance.

In addition to the in silico approach we aim to identify cancer drug resistance genes. Clones with resistance mutations are expected to be at a fitness disadvantage in the untreated cancer cell population relative to the wild-type tumor cells and experience positive selection only at the onset of a therapy. By analyzing inferred selection patterns on DNA sequencing data of cancer patients we want to detect genes that play a role in cancer drug resistance. The data sets we are using include whole exome and whole genome sequencing of primary tumors and metastasis of more than 36 different cancer types and are containing detailed information on the patients therapies and survival.

References

[1] R. Fisher: "The Genetical Theory of Natural Selection", Dover, **2nd edition** (1958)

Mapping single-cell transcriptomes to copy number evolutionary trees

Pedro F. Ferreira[1,2], Jack Kuipers[1,2] and Niko Beerenwinkel[1,2]

[1]Department of Biosystems Science and Engineering, ETH Zürich, Basel, Switzerland
[2] Swiss Institute of Bioinformatics, Basel, Switzerland

Cancer arises and evolves by the accumulation of somatic mutations that provide a selective advantage. The interplay of mutations and their functional consequences shape the evolutionary dynamics of tumors and contribute to different clinical outcomes. In the absence of scalable methods to jointly assay genomic and transcriptomic profiles of the same individual cell, the two data modalities are usually measured separately and need to be integrated computationally. Here, we introduce SCATrEx, a statistical model to map single-cell gene expression data onto the evolutionary history of copy number alterations of the tumor. SCATrEx jointly assigns cancer cells assayed with scRNA-seq to copy number profiles arranged in a copy number aberration tree and augments the tree with clone-specific clusters. Our simulations show that SCATrEx improves over both state-of-the-art unsupervised clustering methods and cell-to-clone assignment methods. In an application to real data, we observe that SCATrEx finds inter-clone and intra-clone gene expression heterogeneity not detectable using other integration methods. SCATrEx will allow for a better understanding of tumor evolution by jointly analysing the genomic and transcriptomic changes that drive it.

Modelling synonymous mutations in human cancer

C. Grimal-Bosch[1] and D. Weghorn[1]

[1]Centre for Genomic Regulation, Barcelona, Spain
E-mail: claudia.grimal@crg.eu

Synonymous mutations have been traditionally thought to be functionally silent and therefore, evolutionarily neutral. With time, however, several mechanisms involving the relative usage and distribution of synonymous codons in coding sequences have been proven to affect gene expression. The decoding rates of codons can vary depending on the availability of their cognate tRNAs, which has promoted the classification of codons into slow and fast, or even optimal and non-optimal. The location and abundance of these so-called non-optimal codons can improve the translation process by allowing for the correct folding of the protein, or can promote mRNA degradation via extended ribosome pausing. These are just some examples of the different roles that synonymous codons can play in regulating the efficiency of protein synthesis, so the question is whether the impact that synonymous mutations have when altering the pattern of codon usage is significant enough to be under selection in certain biological systems. In this case, we are modelling synonymous mutations in human cancer to determine if their observed distribution is consistent with neutral evolution. For that particular purpose, a model has been developed based on the abundance of the different codons and the sequence-context-dependent mutational probabilities. The results obtained have been correlated against the corresponding change in different measures of codon optimality, looking for a way to explain the tensions between the model and the observations.

Discovery and Features of an Alkylating Signature in Colorectal Cancer

Carino Gurjao 1,2+, Rong Zhong 3,4+, Koichiro Haruki 3+, Yvonne Li 1,2, Liam F. Spurr 1,2,5, Henry Lee-Six 6, Brendan Reardon 1,2, Tomotaka Ugai 3,7, Xuehong Zhang 8,9, Andrew D. Cherniack 1,2, Mingyang Song 7,8,9,10,11, Eliezer M. Van Allen 1,2, Jeffrey A. Meyerhardt 1, Jonathan A. Nowak 12, Edward L. Giovannucci 7,8,9, Charles S. Fuchs 13, Kana Wu 9, Shuji Ogino 2,3,7,12, Marios Giannakis 1,2

1. Department of Medical Oncology, Dana-Farber Cancer Institute and Harvard Medical School, Boston, MA, USA
2. Broad Institute of MIT and Harvard, Cambridge, MA, USA
3. Program in MPE, Department of Pathology, BWH and Harvard Medical School, Boston, MA, USA
4. Department of Epidemiology and Biostatistics and Ministry of Education Key Lab of Environment and Health, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, China;
5. Pritzker School of Medicine, Biological Sciences Division, The University of Chicago, Chicago, IL, USA
6. Wellcome Sanger Institute, Hinxton, UK
7. Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA
8. Channing Division of Network Medicine, Department of Medicine, BWH and Harvard Medical School, Boston, MA, USA
9. Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, MA
10. Clinical and Translational Epidemiology Unit, MGH and Harvard Medical School, Boston, MA, USA
11. Division of Gastroenterology, Massachusetts General Hospital, Boston, MA; USA
12. Department of Pathology, Brigham and Women's Hospital and Harvard Medical School, Boston, Boston, MA, USA
13. Yale Cancer Center, Yale School of Medicine, Smilow Cancer Hospital, New Haven, CT, USA

Abstract

Several risk factors have been established for colorectal carcinoma (CRC), yet their direct mutagenic effects in patients' tumours remain to be elucidated. Here, we leveraged whole-exome sequencing data from 900 CRC cases that had occurred in three US-wide prospective studies with extensive dietary and lifestyle information. We found an alkylating signature which was previously undescribed in CRC, and then showed the existence of a similar mutational process in normal colonic crypts. This alkylating signature is associated with high intakes of processed and unprocessed red meat prior to diagnosis. Additionally, this signature was more abundant in the distal colorectum, predicted to target cancer driver mutations KRAS p.G12D, KRAS p.G13D and PIK3CA p.E5454K, and associated with poor survival. Together, these results link for the first time a colorectal mutational signature to a component of diet, and further implicate the role of red meat in CRC initiation and progression.

References

[1] Gurjao, C. et al. Cancer Discov. 11, 2446–2455 (2021).

High clonal diversity and spatial genetic admixture in early prostate cancer and surrounding normal tissue

L. Harbers*[1,2], N. Zhang*[1,2,3], M. Simonetti*[1,2], G. Longo[4], F. Tarish[5], N. Schultz[2,6], V. Roukos[4], T. Helleday[2,5,6], C. Marchiò[7,8], M. Bienko[1,2] and N. Crosetto[1,2]

[1]Department of Microbiology, Tumor and Cell Biology, Karolinska Institutet, Stockholm, Sweden
[2]Science for Life Laboratory, Solna, Sweden
[3]Department of Breast Surgery, Qilu hospital of Shandong University, Ji'nan, China
[4]Institute of Molecular Biology (IMB), Mainz, Germany
[5]Södersjukhuset, Stockholm, Sweden
[6]Department of Oncology and Pathology, Karolinska Institutet, Stockholm, Sweden
[7]Candiolo Cancer Institute, FPO – IRCCS, Candiolo, Turin, Italy
[8]Department of Medical Sciences, University of Turin, Turin, Italy
** Equally contributing authors*

Copy number alterations (CNAs) are pervasive in advanced human cancers, but their prevalence in early-stage, localized tumors and their surrounding normal tissues is poorly characterized. To investigate this phenomenon, here we developed a method for spatially resolved single-cell CNA profiling and applied it to characterize the CNA landscape in 10,007 nuclei extracted from 70 tumor and normal tissue regions (~125 mm3 tissue cubes) from prostatectomies performed in six patients with localized prostate cancer. We identified two distinct groups of cells with abnormal karyotype, one mainly consisting of sparse alterations ('pseudo-diploid' cells) and the other characterized by genome-wide karyotypic changes ('monster' cells). Pseudo-diploid cells displayed high clonal diversity and formed numerous small sized clones ranging from highly spatially localized to broadly spread clones, whereas monster cells were singular events detected throughout the prostate. We observed a remarkable correlation between the fraction of the genome affected by CNAs and the number of tissue regions in which pseudo-diploid cells were found. Highly localized pseudo-diploid clones were enriched in tumor regions and carried deletions of known or putative tumor suppressors, including APC, CDKN1B, FOXO1, FOXP1, and RB1. Strikingly, in two regions in which targeted deep sequencing detected a point mutation affecting the DNA-binding activity of the FOXA1 transcription factor, we also found a co-deletion of FOXO1 and FOXO3 genes in cells from two different pseudo-diploid clones, implicating combinatorial perturbations of Forkhead transcription factors as an early driver of prostate carcinogenesis. Our study reveals that CNAs and mutations are widespread across normal and tumor regions in the prostate glands of patients with localized prostate cancer and suggests that a subset of alterations—most likely small deletions causing the loss of key tumor suppressors—confer a fitness advantage and channel cells towards tumorigenesis.

An evolutionary model of cell death in cancer

Elif Yaren Itak and Johannes Berg

Institute for Biological Physics, University of Cologne, Zülpicher Straße 77, 50937 Cologne, Germany
Email: eitak1@uni-koeln.de and bergj@uni-koeln.de

The loss of cell death mechanisms has long been considered as a hallmark of cancer [1]. Nevertheless, recent studies are also showing evidence of the oncogenic role of stress-induced cell death in early tumour progression. In this context, the evolution of tumours is characterised by a complex and dynamical balance of proliferation and cell death [2]. To model the resulting evolutionary dynamics, we use a stochastic fitness model which links the rate of controlled cell death of a single cell to two quantities: the amount of damage suffered by an individual cell (e.g. DNA damage), and the activation level of a cell death pathway. This model is based on stochastic increases and decreases in these levels per cell at rates that depend on externally controlled parameters and the individual levels. By performing simulations of logistically growing populations of cells, we show how tumour dynamics depend on adjustable parameters, such as the stochastic rates at which cell damage is accumulated or repaired. Using a quasi-species approach, we provide a mathematical framework to describe these underlying stochastic dynamics of a tumour population. This allows us to investigate what degree of intra-tumour heterogeneity is to be expected over the long term for different choices of parameters. Within this approach we can predict the fitness of a clone depending on the pathway activity and the DNA damage. The aim of this work is to study the selection pressure that cell death exerts on a population.

References

[1] Hanahan, D., and Weinberg, R. A. The hallmarks of cancer. Cell vol. 100,1 (2000): 57-70.

[2] Labi, V., Erlacher, M. How cell death shapes cancer. Cell Death Dis 6, e1675 (2015).

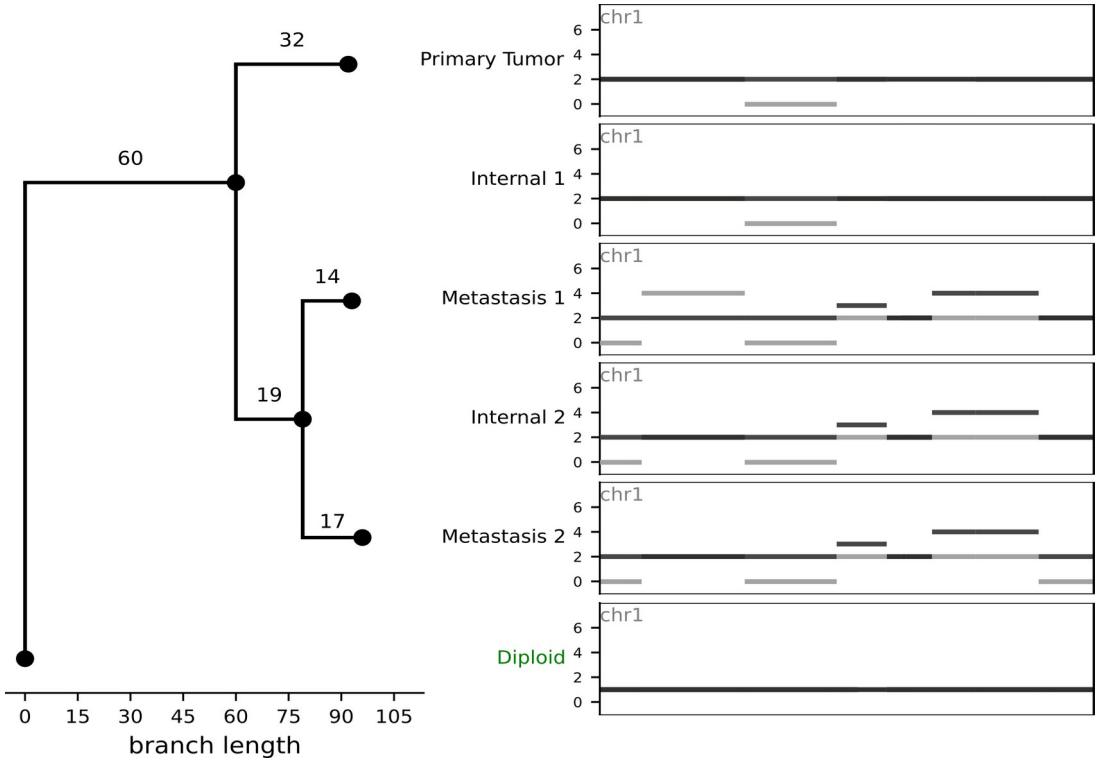
Inferring the evolution of cancer genomes from multi-sample copy number data including whole-genome doubling with MEDICC2

T.L. Kaufmann[1,2], M. Petkovic[1], T.B.K. Watkins[3] and R.F. Schwarz[1,2]

[1]Max Delbrück Center for Molecular Medicine, Berlin, Germany
[2]Berlin Institute for the Foundations of Learning and Data, Berlin, Germany
[3]The Francis Crick Institute, London, UK

Somatic copy number alterations (SCNA) include large-scale events, such as chromosome arm-level gains and losses as well as focal amplifications and deletions and play a key role in the evolutionary processes that shape cancer genomes. SCNAs often appear together with whole genome doubling (WGD) which generates near-tetraploid cells and is associated with poor patient outcome. While the importance of SCNAs and WGD events for tumor evolution is widely accepted, there are currently no methods for phylogenetic inference from SCNAs that include WGD events.

Here, we present MEDICC2, a new phylogenetic algorithm for multi-sample haplotype-specific SCNA data based on a minimum-evolution criterion that infers phylogenetic trees, reconstructs ancestral genomes and reliably detects WGD events. MEDICC2 accurately locates clonal and subclonal copy number events, including WGDs, timing them relative to each other. It explicitely recreates ancestral copy number states which are present as internal nodes in the tree. Efficient parallel implementations enable the application to single-cell experiments with thousands of samples.



Phenotypic deconvolution in heterogeneous cancer cell populations using drug screening data

A. Köhn-Luque[1]

[1]*Oslo Centre for Biostatistics and Epidemiology, Faculty of Medicine, Oslo, Norway*
E-mail: alvaro.kohn-luque@medisin.uio.no

Tumor heterogeneity is an important driver of tumor recurrence, as treatments that initially elicit clinical responses can select for drug-tolerant tumor subpopulations, leading to the outgrowth of resistant clones and cancer treatment failure. Profiling the drug-response heterogeneity of tumor samples using traditional genomic deconvolution methods has yielded limited results, due in part to the imperfect mapping between genomic variation and functional characteristics. Here, by introducing an underlying population dynamic model of tumor subclonal response to therapy, we enable the phenotypic deconvolution of bulk drug response data into component subpopulations, and the estimation of their differential drug sensitivities and population frequencies. We validate this method, called DECIPHER, using simulated tumor drug screening data as well as measurements of drug response in known mixture experiments of cancer cell lines. We then use this method to profile the population drug response heterogeneity in multiple myeloma patient samples, and we demonstrate how these results can be used to produce personalized predictions of tumor response to therapy. This methodology can be applied across cancer types and therapies. Our study (available as preprint [1]) demonstrates how mechanistic population modeling can be leveraged to develop statistical frameworks for profiling phenotypic heterogeneity from bulk tumor samples and to perform individualized patient treatment predictions.

References

[1] A. Köhn-Luque et al, bioRxiv 2022.01.17.476604 (2022)

Disrupted stem cell hierarchy as a driver of leukemogenesis

V. Körber[1], C. Kongsaysak-Lengyel[2], T. Feyerabend[2], N. Claudino[1], H.-R. Rodewald[2] and T. Höfer[1]

[1]Theoretical Systems Biology, German Cancer Research Center, Heidelberg, Germany

[2]Cellular Immunology, German Cancer Research Center, Heidelberg, Germany

Oncogenic mutations perturb the balance between proliferation and differentiation, and can ultimately cause cancer. How often, in which order and in which cell types they arise are fundamental open questions. Here, we study the genetic evolution of T-cell acute lymphoblastic leukemia, spontaneously developing in mice if thymi are decoupled from hematopoietic stem cell input and become self-sustaining. Sequencing whole genomes of T cell precursors from these autonomous thymi, we find that leukemia onset is driven by at least two mutations: typically, trisomy 15 and a mutation in *Notch1*. Developing a population-genetic model of leukemogenesis, we estimate that each driver mutation occurs at least once in a million cell divisions. This rate predicts that several pre-leukemic clones arise, which we confirmed using single-cell RNAseq. Moreover, these data corroborate that the first mutation favors clonal expansion of mutant cells and show that the second hit may then occur in a downstream stage of T cell development. In sum, our findings show that disrupting the normal stem cell hierarchy can cause the rapid development of leukemia at normal rates of mutagenesis.

Turnover Flux and Its Impact on Cancer Evolution

Teemu Kuosmanen[1]

[1]*Department of Computer Science, University of Helsinki, 00014 Helsinki, Finland*
E-mail: teemu.kuosmanen@helsinki.fi

Population genetics provides the cornerstone of understanding the interplay of different evolutionary forces and its rich theory lies at the heart of quantitative descriptions of evolution. However, the standard models fail to accurately describe the complex population dynamics seen in real populations and thus obscure the crucial link between natural selection and the underlying birth and death events. Here we rederive the key results related to fixation and establishment of new mutants starting from the intrinsic birth and death rates and show how also their sum, the turnover rate, plays an important but widely unrecognized role in evolution. We show how both the absolute and relative turnover rates influence evolution and describe a deterministic turnover selection, the turnover flux, which operates in finite populations. Furthermore, by simulating tumor progression via multiple growth epochs, we show how the evolutionary trajectories and waiting times to cancer significantly depend on the tumor microenvironment which modulates the effective birth and death rates. The obtained results further deepen our knowledge on how different life-history strategies, demographic stochasticity, ecological feedback mechanisms, and evolution are inseparably intertwined. This contribution is based on the collaborative work [1].

References

[1] T. Kuosmanen, S. Särkkä, V. Mustonen, in preparation (2022)

Simulations show the possible effect of recurrent mutations, out of equilibrium genomic sequence, and multiple mutation events, on the substitution rates and site frequency spectrum.

H. Loay[1,2] and D. Weghorn[1]

[1]Center of Genomic regulation (CRG), Barcelona, Spain
[2] Universitat Pompeu Fabra (UPF), Barcelona, Spain

Genomic substitutions are often used to estimate mutation rates in studies of human and cancer evolution; however, in many of these studies certain assumptions are usually made. These assumptions include a linear relationship between the number of substitutions and mutation rates, a genomic sequence in detailed balance, and a single mutation event to be the origin of any substitution event (infinite sites model).

Based on computer simulations of realistic human sequence and germline mutation rates, we present evidence of nonlinearity between the number of substitutions and mutation rates. This may cause an underestimation of mutation rates of higher values or for bigger population sizes, two conditions also encountered in many cancer tumors.

We also show a possible skew in the human site frequency spectrum, previously considered as an outcome of the bias in gene conversion phenomena, to happen as a result of the current out of detailed-balance human sequence. Our simulations show a limited effect of the detailed balance status on the substitution rates of smaller populations, compared to the effect of multiple mutation events, where a single site can undergo more than one substitution event.

Breaking the infinite sites model by correcting for the effect of recurrent mutations on the larger population sizes and mutation rates of cancer, may increase the accuracy of the inferred mutational signatures from observed cancer tumor mutations.

Iterative exhaustion of resistance mechanisms against targeted therapy in an EGFR driven cell line

N. Müller[1], J. Brägelmann[2,3,4], C. Lorenz[2,3,4], U. P. Michel[1], D. Plenker[3], S. Ortiz-Cuaran[3], J. Weiss[5], R. Büttner[6], M. Pfeifer[3,4], R. K. Thomas[3,6,7], M. L. Sos[2,3,4] and J. Berg[1]

[1]Institute for Biological Physics, University of Cologne, Germany
[2]Molecular Pathology, Institute of Pathology, University Hospital of Cologne, Germany
[3]Department of Translational Genomics, Center of Integrated Oncology Cologne-Bonn, University of Cologne, Germany
[4]Center for Molecular Medicine Cologne, University of Cologne, Germany
[5]Department I of Internal Medicine, Center for Integrated Oncology Aachen Bonn Cologne Düsseldorf, Germany
[6]Institute for Pathology, University Hospital Cologne, Germany
[7]German Cancer Research Center, German Cancer Consortium (DKTK), Heidelberg, Germany

In solid tumors, the response to targeted therapy is typically short-lived, as therapy-resistant mutants can quickly expand during therapy. We analyze the spectrum of such resistance mutations coexisting in a large population of cancer cells. We use an iterative scheme of artificial evolution to amplify and isolate different resistance mechanisms. As a proof of concept, we apply our scheme to PC-9 cells, a human non-small cell lung cancer cell line with an activating EGFR mutation. The mechanisms we find comprise the well-known gatekeeper-mutation T790M in EGFR, a mutation in NRAS, the amplification of MET-ligand HGF, as well as induction of AKT-mTOR signaling. In this experiment, a combination of four drugs targeting these mechanisms prevents not only the expansion of resistant cells, but also inhibits the growth of drug-tolerant cells, which can otherwise act as a reservoir for further resistance mutations. These results suggest that a finite number of drugs specifically acting on individual resistant clones may be able to control resistance in oncogenically driven lung cancer.

References

[1]N. Müller, J. Brägelmann, C. Lorenz, U. P. Michel, D. Plenker, S. Ortiz-Cuaran, J. Weiss, R. Büttner, M. Pfeifer, R. K. Thomas, M. L. Sos, J. Berg, bioRxiv 756593 (2019)

TAFI (Tumor Allele Frequency Interpreter): a new deep learning tool to reveal the evolutionary history of tumores

Verónica Miró Pina[1] ,David Castellano[1], and Donate Weghorn[1,3]

[1] Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Dr. Aiguader 88, Barcelona 08003, Spain
[2] University of Arizona
[3] Universitat Pompeu Fabra (UPF), Barcelona, Spain

Tumor progression is a somatic evolutionary process in which population expansion is driven by the accumulation of mutations that promote cell proliferation (cancer drivers). Other mutations, called passengers, with an indifferent fitness effect, subsequently accumulate. Passenger mutations are footprints of the tumor's evolutionary history. The distribution of both driver and passenger variant allele frequencies (VAF) can be used to infer relevant biological parameters, such as mutation rates, or to distinguish between tumor growth models under the theoretical framework of population genetics. However, some technical issues arise when trying to compute these estimators from available data. In large scale sequencing projects (e.g., PCAWG [1]), the technique used is bulk sequencing, with low read depth. This curtails our ability to call low frequency variants and yields a general underestimation of the amount of genetic diversity in the tumor. Strong biases arise when comparing data from tumors that have been sequenced using different protocols.
Our goal is to develop an algorithm that can estimate mutation rate and demographic history while accounting for the systematic errors and biases generated by sequencing techniques, mutation calling tools and subsequent filters, that are applied to generate current public datasets. Our method combines Deep Learning with Bayesian statistics [2], to estimate the true amount of genetic diversity and provide reliable estimates of the tumor's mutation rate, demography and age. Our algorithm produces estimates of the mutation load that are comparable across samples with different read depths. This algorithm could be applied to create a comprehensive study of how different evolutionary parameters vary across tumor types and across individual tumors. We present some preliminary results of this study.

References

[1] Pan-cancer analysis of whole genomes. The ICGC/TCGA Pan-Cancer Analysis of Whole Genomes. 2020, Nature, Vol. 578, pp. 81-93.

[2] Approximate Bayesian computation with deep learning supports a third archaic introgression in Asia and Oceania. Mondal, M., Bertranpetit, J. & Lao, O. 246, 2019, Nat Commun, Vol. 10.

Predicting Patterns of Responses to Therapy and Relapse of Aggressive B-cell Lymphomas

Hussein Naji, Katarzyna Bozek

Center for Molecular Medicine Cologne, Cologne, Germany

Diffuse large B-cell lymphoma (DLBCL) is a tumor of B lymphocytes (a subgroup of white blood cells) and is the most common non-Hodgkin's lymphoma in adults. Approx. 37 % of all B cell lymphomas are DLBCL. It is a morphologically, biologically, and clinically extraordinarily heterogeneous disease. Although there are treatment approaches based on high-dose chemotherapy with autologous stem cell support, a long-term disease control is not always possible as up to 50 % of patients suffer from relapses after a first-line treatment. Thus, there is a pressing need for a reliable upfront detection of those patients that will not benefit from standard frontline regimens. The reason for this lack of efficient treatment is that the causes of the DLBCL and its relapse are still not well understood.

The aim of this project is to find a solution for this problem by combining biological expertise with computational approaches. So far, the research on characterizing this tumor was mainly based on genome sequencing techniques. We will expand on the existing studies by implementing data science and machine learning algorithms on microscopy images of the DLBCL. The underlying project objective is to determine characteristics that are predictive of the occurrence of relapse and non-relapse. These features can be used not only to predict treatment outcomes but also to find patterns in relapsing and non-relapsing patients which might provide new information regarding the causes of DLBCL and its relapse. The insights gained from this project might contribute to the development of new, personalized approaches for diagnosis and treatment of DLBCL.

References

[1] Chapuy B., Stewart C., Dunford A.J., Kim J., Kamburov A., et al. (2018). Molecular subtypes of diffuse large B cell lymphoma are associated with distinct pathogenic mechanisms and outcomes. Nat Med 24(5): 679-690.

[2] Klein S., Quaas A., Quantius J., Loser H., Meinel J., et al. (2021). Deep Learning Predicts HPV Association in Oropharyngeal Squamous Cell Carcinomas and Identifies Patients with a Favorable Prognosis Using Regular H&E Stains. Clin Cancer Res 27(4): 1131-1138.

[3] Schmitz R., Wright G.W., Huang D.W., Johnson C.A., Phelan J.D., et al. (2018). Genetics and Pathogenesis of Diffuse Large B-Cell Lymphoma. N Engl J Med 378(15): 1396-1407.

Integrative single-cell tracking of genome evolution and tumor cell plasticity in Small Cell Lung Cancer (SCLC)

Marcel Schmiel[1,2], Roman K[1]. Thomas, Julie George[1]

[1]Department of Translational Genomics, University Cologne, Faculty of Medicine and University Hospital Cologne, Cologne, Germany

[2] Institute of Pathology, University Cologne, Faculty and University Hospital Cologne, Cologne, Germany

SCLC is the deadliest type of lung cancer, which is mostly diagnosed at late stages and accounts for 15% of pulmonary tumors. Even though chemotherapy is initially effective, SCLC patients relapse quickly with therapy resistant tumors. Genome studies revealed universal loss of TP53 and RB1; recent transcriptional studies categorized SCLC tumors based on the expression of 4 lineage transcription factors (TFs). We aimed to dissect how genomic cues impact transcriptional phenotypes, tumor cell plasticity and the dynamics of tumor and immune cell interactions, to thus decipher molecular mechanisms of phenotypic divergence and therapy resistance in SCLC.

We performed genome sequencing and single cell transcriptome profiling on > 60 SCLC tumors, including primary tumors, metastases and paired relapse tumors acquired throughout therapy. Single cell RNA-seq was performed for patient tumors to study cancer cells and the tumor micro-environment (TME)(n=17), and on xenograft models to study tumor cell intrinsic heterogeneity (n=49). Single cell and genome sequencing data were processed with an in-house pipeline. To further validate our findings on the transcriptional level, we performed immunohistochemistry and imaging mass cytometry.

Single cell transcriptome profiles were determined for > 200 000 cells. Transcriptome profiling at single cell level revealed co-expression of at least two of the lineage transcription factors ASCL1, NEUROD1 or POU2F3 for each SCLC tumor. Instead of distinct subgroups, our data points to a concerted regulation of a multitude of lineage factors and transcriptional programs in each SCLC tumor. To specifically investigate how underlying genetic alterations affect molecular phenotypes, integrative studies across all patients were performed. This revealed shared transcriptional programs in tumors with MYC family member amplifications; trajectory inference for cases with low and high-level copy gains of MYCN pointed to distinct transcriptional states. We performed multi-regional studies of matched patient cases. Additionally, we reconstructed the clonal evolution and projected it to transcriptional trajectories, thus allowing to determine the effect of genome diversity on the transcriptional landscape. All together our data provides a first comprehensive framework for the study on how underlying genetic alterations shape the transcriptional landscape and phenotypic plasticity in SCLC.

In silico testing of hypotheses for lung cancer development

H. Selway[1] and A. Pennycuick[2]

[1]UCL, London, United Kingdom

Single-cell genomic analysis of airway basal cells [1] show two surprising findings:
1) The mutational burden of basal lung cells has a bimodal distribution, with some cells near normal while others are highly mutated
2) The proportion of near-normal cells is higher in ex-smokers than in current smokers, suggesting that the healthy population of cells expands after cessation of smoking

Different hypotheses for driving underlying mechanisms could include:
1 a) A quiescent subpopulation of basal stem cells which divide much slower and hence mutate less
b) A protected niche in airways where tobacco carcinogens cannot penetrate
c) Some variation in cells' inherent resistance to tobacco-induced mutagenesis
2 a) Loss of mutational advantage conferred by driver mutations in the absence of smoke
b) Accumulated high mutational burden leading to negative selection through either deleterious passenger mutations or immune surveillance, bolstered by the lack of immunosuppressant tobacco smoke

Here we present a modular computational model which builds on an experimentally verified homeostatic model [2] to account for mutations on division and relative cell fitness. The model simulates each patient's basal lung cell population based on their smoking record. Each hypothesis about underlying mechanisms is included as an independently excludable module.
The resultant models are fitted to experimental data via Bayesian Optimisation, minimising the Wasserstein distance between the mutational distributions of the simulated and real cell populations. By comparing how well the different models fit, we infer the relative validity of the hypotheses.
First-principles modelling of underlying mechanisms to compare hypotheses provides insight into the way cancer develops in the lungs. This could allow for more targeted analyses of datasets to improve risk modelling and streamline the search for novel preventative treatments.

References

[1] K. Yoshida et al., Nature **578**, 266-272 (2020)
[2] VH. Teixeira et al., eLife 2013;2:e00966 (2013)

Mosaic mutations at transcription factor binding sites are enriched in schizophrenia

Eduardo A. Maury[1,2,3,*], Attila Jones[4,5,*], Vladimir Seplyarskiy[6,7,*], Chaggai Rosenbluh[4], Yifan Wang,[8,9,10], Taejong Bae[8], Alexej Abyzov[8], Sattar Khoshkoo[1,3,11], Yasmine Chahine[1,3], Brain Somatic Mosaicism Network[12], Peter J. Park[6], Schahram Akbarian[13], Eunjung Alice Lee[1,3], Shamil R. Sunyaev[6,7], Christopher A. Walsh[1,3,14,15], Andrew Chess[4,5,11]

Schizophrenia (SCZ) is a complex neuropsychiatric disorder in which both germline genetic mutations[1–4] and maternal factors, such as infection and immune activation[5–7], have been implicated, but how these two strikingly different mechanisms might converge on the same phenotype is unknown. During development, cells accumulate somatic, mosaic mutations in ways that can be shaped by the cellular environment or endogenous processes[8–10], but these early developmental mutational patterns have not been studied in SCZ. Here we analyzed deep (267x) whole-genome sequencing (WGS) of DNA from cerebral cortical neurons isolated from 61 SCZ and 25 control postmortem brains to capture mutations occurring before or during fetal neurogenesis. SCZ cases overall showed a >15% increase in genome-wide sSNV compared to controls ($p < 2e-10$). Remarkably, mosaic T>G mutations and CpG transversions were 79- and 39-fold enriched, respectively, at transcription factor binding sites (TFBS) in SCZ, but not in controls. The pattern of T>G mutations resembles mutational processes in cancer attributed to oxidative damage that is sterically blocked from DNA repair by transcription factors (TFs) bound to damaged DNA[11–13]. The CpG transversions similarly suggest unfinished DNA demethylation resulting in abasic sites[14] that can also be blocked from repair by bound TFs[11,12,15]. Allele frequency analysis suggests that both localized mutational spikes occur in the first trimester[10]. We call this prenatal mutational process "*skiagenesis*" (from the Greek *skia*, meaning shadow), as these mutations occur in the shadow of bound TFs. Skiagenesis reflects as-yet unidentified prenatal factors and is associated with SCZ risk in a subset of cases. In turn, mutational disruption of key TFBS active in fetal brain is well positioned to create SCZ-specific gene dysregulation in concert with germline risk genes. *Skiagenesis* provides a fingerprint for exploring how epigenomic regulation and prenatal factors such as maternal infection or immune activation may shape the developmental mutational landscape of human brain.

SigNet: three ANN-based tools for extracting and analyzing mutational processes in tumor genomes

Claudia Serrano Colome[1,2], Oleguer Canal Anton[3], Donate Weghorn[1,2]

[1]Centre for Genomic Regulation (CRG), Barcelona, Spain
[2] Universitat Pompeu Fabra (UPF), Barcelona, Spain
[3]KTH Royal Institute of Technology in Stockholm, Stockholm, Sweden

Mutations that occur during cancer development can be classified into different mutational processes. This is done using the mutations' sequence contexts and statistical tools that decompose the observed mutation spectra into statistically independent mutational components, or "signatures". Many so-called refitting algorithms have been developed that then use a linear combination of these independent signatures to assign mutations from a given tumor to the underlying mutational processes. However, these algorithms were developed years ago, when a set of 30 signatures was originally identified, and they have not been tested for the newest catalogs that contain more than 70 different signatures. Furthermore, estimated per-tumor signature weights are noisy, yet their errors are rarely quantified in existing algorithms. Here I present **SigNet**: a collection of data-based algorithms useful in the study of mutational processes. First of all we present **SigNet Refitter**, a method that uses Deep Neural Networks to do signature refitting. This algorithm outperforms the existing methods, is easily adaptable to any new catalog of mutational signatures and provides very accurate prediction intervals for the contribution of each signature in a single tumor. This showcases how, using Artificial Neural Networks, we can take advantage of the correlations between the different mutational processes that occur during carcinogenesis to obtain accurate signature decompositions, even when the number of mutations in the sampled tumor is very low. However, one of the main challenges of taking a data-based approach is the need of a labeled dataset. In this regard, we designed a method called **SigNet Generator**, to synthetically create realistic-looking data based on the known correlations between signatures. Besides, since these known correlations are susceptible to change and more signatures are left to be identified, we also implemented a method capable of identifying samples that do not preserve such correlations. We refer to it as **SigNet Detector**. SigNet, with its three modules, provides useful tools for the field of signature decomposition analysis.

Phenotypic plasticity can contribute to cancer metastasis

S. Shah[1], M. Raatz[1], and A. Traulsen[1]

[1]Department of Evolutionary Theory, Max Planck Institute for Evolutionary Biology, Plön, Germany

Preventing metastasis, the systemic spread of cancer, remains a challenge even though it is the primary cause of cancer-related fatalities. Successful metastases require tumor cells to be invasive and proliferative; however, a cell can only have one of the two phenotypes. Thus, tumor cells change their phenotype during metastasis to make up for this trade-off. It is believed that this experimentally observed phenotypic change results from different environmental cues and phenotypic programs at the primary site of cancer and the secondary site of metastasis formation. Instead, we provide an alternative explanation for the observed phenotype changes by accounting for intrinsically phenotypically-plastic cell populations that maintain a phenotypic heterogeneity that is resilient against perturbations. Our model admits an equilibrium - corresponding to a stable distribution of tumor phenotypes. At the primary site of cancer with initially low invasive capacity, this equilibrium is approached by a relative increase of invasive cells. At the secondary site of metastasis with initially low proliferative capacity, this equilibrium is approached by a relative increase of proliferative cells. Thus, the experimentally observed phenotype shift during metastasis can result from intrinsic plasticity and different initial conditions reconstituting the phenotype distribution, thus extending the current understanding of the metastatic process. We further investigate the effect of the rate of phenotype transitions on the heterogeneity of the tumor population. We show that by slowing down the phenotype transitions or adjusting the tendency of the cells to switch phenotypes, the tumor is forced into a less heterogeneous state, leading to favorable treatment outcomes. Accounting for alternative mechanisms to the current understanding could identify new clinical targets to contain cancer metastasis.

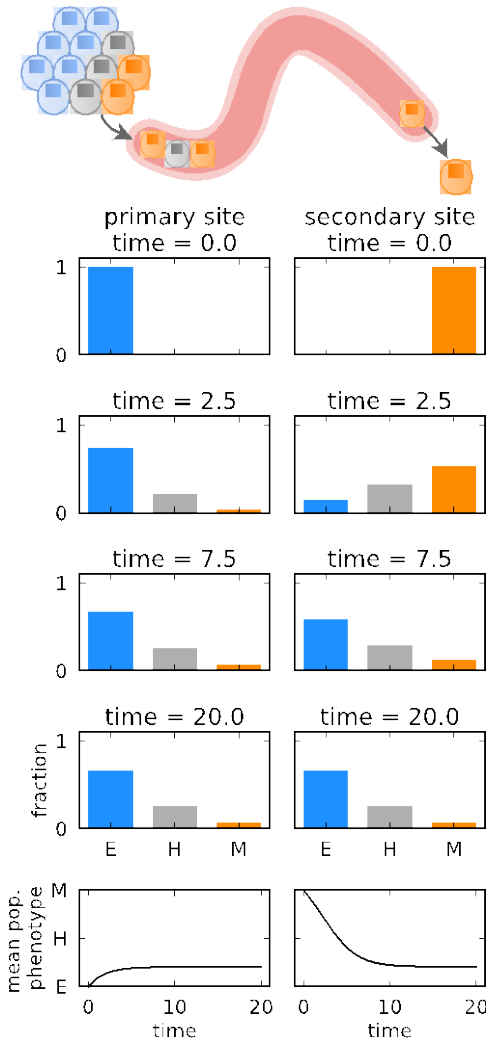


Figure 1: The proliferative (E), hybrid (H), and invasive (M) cells are represented by blue, grey, and orange colors.

The evolution of two transmissible cancers in Tasmanian devils

M.R. Stammnitz[1,#], K. Gori[1], Y.M. Kwon[1], E. Harry[2], F.J. Martin[3], K. Billis[3], Y. Cheng[4], W. Chow[2], S. Comte[5], H.P. Eggertsson[6], S. Fox[7], R. Hamede[5], B. Lazenby[7], S.A. McCarthy[2], S. Peck[7], M.A. Quail[2], K. Swift[8], J. Wang[1], J. Wood[2], K. Howe[2], M.R. Stratton[2], Z. Ning[2], E.P. Murchison[1]

[1]Department of Veterinary Medicine, University of Cambridge, United Kingdom
[2]Wellcome Sanger Institute, Hinxton, United Kingdom
[3]European Bioinformatics Institute, Hinxton, United Kingdom
[4]University of Sydney, Sydney, Australia
[5]University of Tasmania, Hobart, Australia
[6]deCODE genetics Inc, Reykjavik, Iceland
[7]Save the Tasmanian Devil Program, Hobart, Australia
[8]Mount Pleasant Laboratories, Prospect, Australia
[#]Current address: Centre for Genomic Regulation, Barcelona, Spain,
E-mail: maximilian.stammnitz@crg.eu

Transmissible cancers are clonal lineages that spread through populations via transfer of contagious cancer cells. Such diseases have rarely been observed in mammals, yet two distinct clones, known as devil facial tumour 1 (DFT1) and devil facial tumour 2 (DFT2), affect the same marsupial carnivore: the Tasmanian devil. Both cancers are spread between devils by biting, emerged within the past 25 years – DFT1 was first observed in 1996, DFT2 in 2014 – and have caused substantial population declines in the species. In addition to their importance as unusual pathogens, DFT1 and DFT2 permit repeated sampling of individual cancer lineages over decades, thus providing a unique system with which to study *in vivo* mutational processes over long timespans.

By deep sequencing of whole genomes (>60X) from large tumour cohorts of DFT1 (N=78) and DFT2 (N=41) from infected devils sampled between 2003 and 2018, we have reconstructed detailed mutational trajectories of both lineages. Time-resolved phylogenetic trees suggest that DFT1 first emerged in 1986 (1982-1988), and DFT2 in 2011 (2009-2012). Subclone analysis documents transmission of heterogeneous cell populations. DFT2 has faster mutation rates than DFT1 across all variant classes, including substitutions, indels, rearrangements, LINE-1 transposable element insertions and copy number alterations, and we identify a hypermutated DFT1 lineage with defective DNA mismatch repair. Several loci show plausible evidence of positive selection in DFT1 or DFT2, including loss of chromosome Y and inactivation of MGA, but none are common to both cancers. This study illuminates the parallel long-term evolution of two transmissible cancers, and that mutations of several classes accumulate linearly with time within individual cancer clones.

Biophysical and Computational study of leukemic tumors at Extramedullary sites

Kavan Thakkar[1]

[1]*Indian Institute of Technology*

Every cancer except some specific type of brain cancer is linked by the inherent property of metastasis. The metastasis usually occurs in the five steps: invasion and migration, intravasation, circulation, extravasation, and finally, the combination of colonization, proliferation, and angiogenesis. Due to the circulatory nature of leukocytes, the first two steps are generally eliminated in the spread of leukemia at the non-marrow sites, so the rate of metastasis will be comparatively high in the liquid tumors [1].

The proto-oncogene c-myb is a nuclear transcription factor that shares significant sequence homology with two other myb family members, A-myb and B-myb. The c-myb gene encodes a transcription factor that regulates the cell cycle via regulating cell proliferation, differentiation, and apoptosis through various protein-protein interactions, DNA binding, and signaling pathways. It is found in breast epithelia, hematopoietic cells, colon, brain, and kidneys, where it plays essential functions via various routes [2,3].

Due to the general location and molecular functions of c-myb, dysregulation of c-myb is associated with various types of human cancers, such as Acute myeloid leukemia (AML), Glioblastoma, Breast cancers, Ovarian cancers, Colon cancers, Hepatic carcinomas, etc. In humans, c-myb is located on the 6q. c-myb shows its oncogenic activity by mainly translocation with several genes that have a poor prognosis profile in patients with leukemia and other solid cancers [4,5].

Patients with extramedullary leukemia have poor survival rates because we lack information on how leukemic cells adopt such metastatic behavior. With the help of single-cell mechanics and transcriptomic sequencing, I want to determine the path by which the leukemic cells adopt the more malignant behavior to develop therapeutic strategies that can aid in the betterment of prognosis.

References

[1] Trendowski M. The inherent metastasis of leukaemia and its exploitation by sonodynamic therapy. Crit Rev Oncol Hematol. 2015 May;94(2):149-63. doi: 10.1016/j.critrevonc.2014.12.013. Epub 2015 Jan 3. PMID: 25604499

[2] Liu X, Xu Y, Han L, Yi Y. Reassessing the Potential of Myb-targeted Anti-cancer Therapy. J Cancer. 2018 Mar 15;9(7):1259-1266. doi: 10.7150/jca.23992. PMID: 29675107; PMCID: PMC5907674.

[3] Extramedullary leukemia behaving as solid cancer: clinical, histologic, and genetic clues to chemoresistance in organ sites. **https://doi.org/10.1002/ajh.25594**.

[4] Oh IH, Reddy EP. The myb gene family in cell growth, differentiation and apoptosis. Oncogene. 1999 May 13;18(19):3017-33. doi: 10.1038/sj.onc.1202839. PMID: 10378697.

[5] Extramedullary leukemia behaving as solid cancer: clinical, histologic, and genetic clues to chemoresistance in organ sites. **https://doi.org/10.1002/ajh.25594**.

Graphical models for cancer evolution.

D. Tramontano[1], R. Talbut[2], A. Monod[2] and M. Drton[1]

[1]Technical University of Munich, Munich, Germany
[2] Imperial College London, London, UK

We propose a novel algorithm to learn the causal behavior of graphs from observational data. With the aim of learning cancer evolution patterns in mind, and in order to leverage powerful theory from algebraic geometry which additionally bypasses identifiability issues as well as computational limitations, we focus on polytrees (i.e., directed graphs whose skeletons are trees) in the linear non Gaussian casual model context. Our algorithm functions in two steps, which reconstructs the tree skeleton using the classical Chow-Liu algorithm in a first stage, and subsequently orients the edges of the graph according to algebraic conditions on moment variables. Furthermore, we prove a consistency theorem for our algorithm in a high-dimensional setting where the size of the tree grows at a faster rate than the sample size subject to log-concavity of the variables. These features make our contribution a promising tool to reconstruct and study the behavior of cancer evolution from the algebraic statistical perspective of graphical models.

Genomic instability as a determinant of tumor-immune co-evolution in ovarian cancer

I. Vázquez-García[1,*], F. Uhlitz[1,*], N. Ceglia[1], J. Lim[1], M. Wu[1], N. Mohibullah[1], A. Ruiz[1], K. Boehm[1], V. Bojilova[1], C. Fong[1], T. Funnell[1], D. Grewal[1], E. Havasov[1], S. Leung[1], A. Pasha[1], D. Patel[1], M. Pourmaleki[1], N. Rusk[1], H. Shi[1], R. Vanguri[1], M. Williams[1], A. Zhang[1], V. Broach[1], D. Chi[1], A. Da Cruz Paula[1], G. Gardner[1], S. Kim[1], M. Lennon[1], K. Long Roche[1], Y. Sonoda[1], O. Zivanovic[1], R. Kundra[1], A. Viale[1], F. Derakhshan[1], L. Geneslaw[1], A. Maroldi[1], R. Nunez[1], F. Pareja[1], A. Stylianou[1], M. Vahdatinia[1], Y. Bykov[1], R. Grisham[1,2], Y. Liu[1], Y. Lakhman[1], I. Nikolovski[1], D. Kelly[1], J. Gao[1], A. Schietinger[1], T. Hollmann[1], S. Bakhoum[1], R. Soslow[1], L. Ellenson[1], N. Abu-Rustum[1,2], C. Aghajanian[1], C. FFriedman[1,2], A. McPherson[1], B. Weigelt[1], D. Zamarin[1,2,†], S. Shah[1,†]

[1]Memorial Sloan Kettering Cancer Center, New York, NY, USA
[2] Weill Cornell Medical Center, New York, NY, USA

*Equal contribution; †Senior author

Genomic instability is a hallmark of human cancer, with fundamental relevance to cancer etiology and evolution, anti-tumor immunity and therapeutic response. High-grade serous ovarian cancer (HGSOC) is an archetypal cancer of genomic instability defined by distinct mutational processes, intraperitoneal spread and tumor heterogeneity. As immunotherapies have thus far proven ineffective in this disease, we sought to establish the determinants of immune recognition and evasion and to quantify their impact in the natural evolutionary history of HGSOC [1]. Accordingly we studied the impact of mutational processes on cellular phenotypes in the tumor microenvironment (TME) using genome-based stratification of homologous recombination proficient (HRP) and deficient (HRD) disease subtypes, and profiling single cell phenotypes from ~1 million cells including cancer cells, T cells, myeloid cells and fibroblasts derived from single cell RNA sequencing, and in situ spatial profiling of cancer cell, T cell and macrophage states of 160 tumor sites obtained from 42 treatment-naive patients.

Mutational processes in HRD-Dup (*BRCA1* mutant-like) tumors were associated with a high neoantigen burden, cancer cell-intrinsic JAK/STAT signaling and predominance of highly-differentiated dysfunctional $CD8^+$ T cells in the TME; HRD-Del (*BRCA2* mutant-like) tumors were linked with cancer cell-intrinsic NF-κB and TNFα signaling; and foldback inversion (FBI, HRP) tumors were associated with cancer cell-intrinsic TGFβ signaling and immune exclusion, with predominantly naive/central memory-like T cells. HLA loss of heterozygosity (LOH) was a common mechanism of immune escape in the HRD, but not FBI tumors, connecting evolutionary selection with immune states. Our findings yield mechanistic insights for how distinct mutational processes in HGSOC lead to diverse patterns of intra- and inter- patient variation in immune resistance, which can be leveraged to optimize future immuno-therapeutic strategies.

References

[1]I. Vázquez-García, F. Uhlitz et al, bioRxiv **2021.08.24.454519** (2021)

Chromosomal instability drives spatial and temporal phenotypic diversity in Schwann cancer cells

H. Yan[1], J. Demeulemeester[1,2], A. Verfaillie[1], C. Cotobal Martin[1], T. Kaufmann[3], R. Schwarz[3], T. Voet[4], S. Zaccaria[5], A. Flanagan[5], M. Tarabichi[1,6], P. Van Loo[1]

[1]Francis Crick Institute, London, United Kingdom
[2] KU Leuven, Leuven, Belgium
[3]MDC, Berlin, Germany
[4] Wellcome Trust Sanger Institute, Hinxton, United Kingdom
[5]University College London, London, United Kingdom
[6]Université Libre de Bruxelles, Brussels, Belgium

Malignant peripheral nerve sheath tumours (MPNSTs) are rare, aggressive soft tissue sarcomas derived from Schwann cells with poor prognosis[1]. Previous studies have shown that these tumours are typically genomically complex, with a moderate burden of single nucleotide variants but extensive copy number aberrations (CNAs)[2].
We applied a multi-omics approach to scrutinize the evolution and heterogeneity of a primary MPNST and five recurrence regions from one patient.
We find significant heterogeneity in copy number profiles, suggestive of ongoing chromosomal instability and evolution. We perform in-depth tumour phylogenetic reconstruction from bulk whole genome sequencing data. Single-cell DNA sequencing (scDNA-seq) revealed further CNA heterogeneity across and within regions, allowing us to refine the tree down to single-cell resolution.
We also profile different populations of tumour and non-tumour cells and confirmed this with a genotyping approach. CNA profiles inferred from single-cell transcriptomes reflect the within-region heterogeneity seen in scDNA-seq. We find tumour cells with similar CNA profiles typically also cluster together by their gene expression profiles, implying that the majority of gene expression heterogeneity in this tumour is underpinned by copy number changes.
Finally, we explore the spatial relationships between tumour subclones using laser capture microdissection. We detect heterogeneity within tissue sections evidencing local clonal expansions and CNA events that follow spatial distributions. Using spatial transcriptomics, we show tumour cells are homogeneously admixed with tumour microenvironment populations.
Our work demonstrates the power of combining spatial multi-omics at the single-cell and bulk levels to study cancer evolution. This enables a spatio-temporal representation of a tumour's development with annotation of genetic and transcriptomic events.

References

[1] G. Dunn, *J. Neurosurg.* **118**, 142-148 (2013)
[2] A. Pemov, *Neuro-Oncol. Adv.* **2**, i50-i61 (2020)