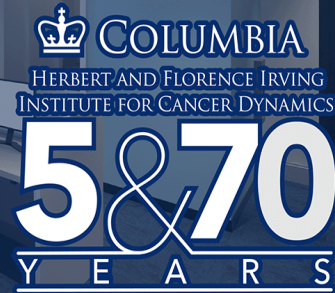


Mathematical and Computational Methods in Cancer and Biology

A celebration of the 5th anniversary
of the IICD and Simon Tavaré's 70th



16 - 18
March 2023



Mathematical and Computational Methods in Cancer and Biology Symposium

A Celebration of the 5th Anniversary of the Irving Institute
for Cancer Dynamics and Simon Tavaré's 70th

March 16th-18th, 2023

Table of Contents

Overview	3
Agenda	4
Symposium Speakers	6
Poster Session	39

Overview

The importance of the mathematical and computational sciences and technological developments in driving fundamental discoveries in cancer biology and many other biological fields is now widely acknowledged. To highlight the latest research in those areas, the Herbert and Florence Irving Institute for Cancer Dynamics is organizing the following symposium titled "Mathematical and Computational Methods in Cancer and Biology." The event will take place on March 16th-18th, 2023 at Columbia University. The conference will feature local, national, and international speakers describing their work on cancer 'omics, population genetics, computational biology, probability, and statistics. The symposium will also be an occasion to celebrate two milestones: Simon Tavaré's 70th birthday and the 5th anniversary of the Irving Institute for Cancer Dynamics.

Symposium Organizing Committee:

Simon Tavaré, Irving Institute for Cancer Dynamics Director, Professor of Statistics and Biological Sciences

Poly da Silva, Postdoctoral Research Scientist, Irving Institute for Cancer Dynamics and Department of Statistics

Arash Jamshidpey, Associate Research Scientist, Irving Institute for Cancer Dynamics and Department of Mathematics

Lorenza Favrot, Program Manager, Outreach and Communication, Irving Institute for Cancer Dynamics

Reed Black, Admin Coordinator, Irving Institute for Cancer Dynamics

Agenda

Thursday, March 16th		
8:00 AM-9:00 AM	Check-in & Breakfast, Faculty House, Presidential Ballroom	
9:00 AM-9:25 AM	Opening Remarks, Faculty House, Presidential Ballroom	
9:00 AM-9:25 AM	Jeannette Wing, <i>Executive Vice President for Research/The Fu Foundation School of Engineering and Applied Science, Professor of Computer Science - Columbia University</i> Michael Waterman, <i>University Professor Emeritus, Quantitative & Computational Biology, Mathematics, Computer Science - University of Southern California</i> Anil Rustgi, <i>Director of the Herbert Irving Comprehensive Cancer Center/Associate Dean of Oncology/Chief of Cancer Services/Professor of Medicine - Columbia University</i>	
9:25 AM-12:20 PM	Session 1: Cancer, Faculty House, Presidential Ballroom	
9:25 AM-9:50 AM	Sam Aparicio	University of British Columbia/BC Cancer Research Institute/IICD
9:50 AM-10:15 AM	Darryl Shibata	University of Southern California
10:15 AM-10:40 AM	Khanh N. Dinh	Columbia University/IICD
10:40 AM-11:05 AM	Coffee Break	
11:05 AM-11:30 AM	Trevor Graham	Institute of Cancer Research
11:30 AM-11:55 AM	Karol Nowicki-Osuch	Columbia University/IICD/New York Genome Center
11:55 AM-12:20 PM	Christina Curtis	Stanford University
12:20 PM-2:00 PM	Lunch Break, Faculty House, Presidential Ballroom	
2:00 PM-5:00 PM	Session 2: Population Genetics & Evolution, Faculty House, Presidential Ballroom	
2:00 PM-2:25 PM	Gesine Reinert	University of Oxford
2:25 PM-2:50 PM	Molly Przeworski	Columbia University
2:50 PM-3:15 PM	John Wakeley	Harvard University
3:15 PM-3:45 PM	Coffee Break	
3:45 PM-4:10 PM	Ignacio Vazquez-Garcia	Memorial Sloan Kettering/Columbia University/IICD
4:10 PM-4:35 PM	Magnus Nordburg	Gregor Mendel Institute of Molecular Plant Biology
4:35 PM-5:00 PM	Robert Griffiths	Monash University
7:00 PM	Speaker Dinner (Only for symposium speakers)	

Friday, March 17th		
8:30 AM-9:00 AM	Breakfast, Faculty House, Presidential Ballroom	
9:00 AM-12:20 PM	Session 3: Computational Biology, Faculty House, Presidential Ballroom	
9:00 AM-9:25 AM	Elham Azizi	Columbia University/IICD
9:25 AM-9:50 AM	Bianca Dumitrascu	Columbia University/IICD
9:50 AM-10:15 AM	John Marioni	Genentech
10:15 AM-10:40 AM	Sanja Vickovic	New York Genome Center/Columbia University/IICD
10:40 AM-11:05 AM	Coffee Break	
11:05 AM-11:30 AM	David Blei	Columbia University/IICD
11:30 AM-11:55 AM	Daniele Biasci	University of Oxford
11:55 AM-12:20 PM	Jellert Gaublomme	Columbia University/IICD
12:20 PM-2:00 PM	Lunch Break, Faculty House, Presidential Ballroom	
2:00 PM-5:00 PM	Session 4: Mathematics and Probability, Faculty House, Presidential Ballroom	
2:00 PM-2:25 PM	Ivan Corwin	Columbia University/IICD
2:25 PM-2:50 PM	Jim Pitman	University of California, Berkeley
2:50 PM-3:15 PM	Andrew Blumberg	Columbia University/IICD
3:15 PM-3:45 PM	Coffee Break	
3:45 PM-4:10 PM	Alison Etheridge	University of Oxford
4:10 PM-4:35 PM	Arash Jamshidpey	Columbia University/IICD
4:35 PM-5:00 PM	Andrew Barbour	University of Zurich
5:30 PM-7:30 PM	Reception and Poster Session, Faculty House, Skyline Dining Room	

Saturday, March 18th		
8:30 AM-9:00 AM	Breakfast, Faculty House, Presidential Ballroom	
9:00 AM-11:55 PM	Session 5: Statistics, Faculty House, Presidential Ballroom	
9:00 AM-9:25 AM	Richard Wilkinson	University of Nottingham (UK)
9:25 AM-9:50 AM	Poly Hannah da Silva	Columbia University/IICD
9:50 AM-10:15 AM	Russell Kunes	Columbia University/IICD
10:15 AM-10:40 AM	Coffee Break	
10:40 AM-11:05 AM	Sanket Rane	Columbia University/IICD
11:05 AM-11:30 AM	Richard Davis	Columbia University
11:30 AM-11:55 AM	Bin Yu	University of California, Berkeley
11:55 AM-12:15 PM	Closing Remarks, Faculty House, Presidential Ballroom	

Symposium Speakers

Talks will take place in the Faculty House, Presidential Ballroom

Opening Remarks



Jeannette Wing

Executive Vice President for Research
Professor of Computer Science, The Fu Foundation School of Engineering
and Applied Science, Columbia University



Michael Waterman

University Professor Emeritus, Quantitative & Computational Biology,
Mathematics, Computer Science, University of Southern California
Distinguished Research Professor, Biocomplexity Institute, University of
Virginia



Anil Rustgi

Director of the Herbert Irving Comprehensive Cancer Center
Associate Dean of Oncology
Chief of Cancer Services
Professor of Medicine, Columbia University

Samuel Aparicio, Canada Research Chair, University of British Columbia, and Affiliate Member of the Irving Institute for Cancer Dynamics, Columbia University

Decoding the origins of mutational single cell fitness in cancer drug resistance

Copy number-structural variant (CNA-SV) mutations have a pervasive effect on gene transcription because single mutations can directly affect hundreds of genes. Some 75% of triple negative breast cancer (TNBC, ~15% of early breast cancers) exhibit genome instability with poor prognosis; the instability features are shared by high grade serous ovarian cancer (HGSOC). Using novel methods for single genome sequencing (Laks *et al.*, Cell 2019) we have discovered copy number-structural variant (CNA-SV) “foreground” processes distributed over single cells and associated with background patterns of instability (Funnell *et al.*, Nature 2022). These mechanisms are associated with resistance to platinum salts. Our second discovery associates CNA-SV genotypes with clonal resistance to platinum in TNBC (Salehi *et al.*, Nature 2021), re-purposing Wright-Fisher models in time series samples of cancer to measure fitness. We now understand ongoing CNA-SV mutations contribute to evolution of fitness landscapes for DNA targeting therapeutics. A model for estimating gene fitness in large scale perturbation experiments of transplantable human tumours will be presented. Future prospects for multi-mode decoding of fitness will be discussed.



Professor Samuel Aparicio is the Nan & Lorraine Robertson Chair in Breast Cancer Research UBC, chair of the BC Cancer Research Department of Breast and Molecular Oncology in Vancouver, Canada and Fellow of the Royal Society of Canada. He has most recently conducted foundational work on methods for studying the evolution of human cancers using next-generation sequencing approaches and single cell sequencing methods. Dr Aparicio is also working to develop quantitative measures of clonal fitness and epigenetic re-wiring in patients, including methods for single cell genome sequencing and measuring fitness in patient derived xenograft models of human cancer. He is co-founder of Inflex, targeting DNA tertiary structures.

Elham Azizi, Herbert & Florence Irving Assistant Professor of Cancer Data Research (in the Herbert and Florence Irving Institute for Cancer Dynamics and in the Herbert Irving Comprehensive Cancer Center) and Assistant Professor of Biomedical Engineering, Columbia University

Machine learning dynamics in the tumor microenvironment

Cancer therapies succeed only in a subset of patients partly due to the heterogeneity of cells across and within tumors. Single-cell and spatial genomic technologies present exciting opportunities to characterize unknown cell types in complex tissues such as tumor microenvironments and elucidate their interactions, circuitry, and role in driving response to therapies. However, analyzing and integrating single-cell data across conditions, patients, time points, and data modalities involve significant statistical and computational challenges. I will present a set of probabilistic and deep generative models developed for addressing these problems and modeling temporal and spatial dynamics of key immune subsets defining cancer progression and response to immunotherapy. I will also present Starfysh for spatial mapping of heterogeneous cell states and crosstalk in complex tissues, from the integration of spatial transcriptomics and histology images.



Elham Azizi joined Columbia in 2020 as the Herbert and Florence Irving Assistant Professor of Cancer Data Research in the Irving Institute for Cancer Dynamics and Assistant Professor of Biomedical Engineering. She is also affiliated with the Department of Computer Science, Data Science Institute, and the Herbert Irving Comprehensive Cancer Center. Elham holds a BSc in Electrical Engineering from Sharif University of Technology, an MSc in Electrical Engineering and a PhD in Bioinformatics from Boston University. She was a postdoctoral fellow in the Dana Pe'er Lab at Columbia University and Memorial Sloan Kettering Cancer Center. Her multidisciplinary research utilizes novel machine learning techniques and single-cell genomic and imaging technologies to study the dynamics and circuitry of interacting cells in the tumor microenvironment. She is a recipient of the CZI Science Diversity Leadership Award, NSF CAREER Award, Tri-Institutional Breakout Prize for Junior Investigators, NIH NCI Pathway to Independence Award, an American Cancer Society Postdoctoral Fellowship, and an IBM Best Paper Award at the New England Statistics Symposium.

Andrew Barbour, Emeritus Professor of Mathematics, University of Zürich

Estimating the correlation in network disturbance models

The Network Disturbance Model of Doreian (1989) expresses the dependency between observations taken at the vertices of a network by modelling the correlation between neighbouring vertices, using a single correlation parameter r . It has been observed that estimation of r in dense networks, using the method of Maximum Likelihood, leads to results that can be both biased and very unstable. In this talk, we sketch why this is the case, showing that the variability cannot be avoided, no matter how large the network. We also propose a more intuitive estimator of r , which shows little bias. (Joint work with Gesine Reinert, University of Oxford)



Andrew Barbour: My research has been largely directed towards aspects of probability theory, and its uses in understanding problems in biology. More concretely, my primary research areas are: Stein's method for distributional approximation; Mathematical epidemiology, modelling and data analysis; Combinatorial and geometrical probability; Branching and (meta-)population processes; and Stochastic networks. I retired from my chair at the University of Zürich in January 2011, and hold an honorary professorial fellowship at the University of Melbourne.

Daniele Biasci, Principal Investigator (Innovator Track), The Kennedy Institute of Rheumatology, University of Oxford

The landscape of high-affinity human antibodies against intratumoral antigens

Intratumoral expression of antibody transcripts can be detected across different cancer types and is correlated with positive clinical outcomes. However, the antigens recognized by these antibodies remain largely unknown. Here we inferred the paired sequence of thousands of clonally expanded antibodies using bulk RNA sequencing data from solid tumors in The Cancer Genome Atlas (TCGA). After expressing 283 candidate antibodies in mammalian cells, we individually tested them against a library of twenty thousand full-length human proteins and a library of five thousands membrane-only proteins. Using surface plasmon resonance we confirmed high-affinity antibodies that bind to their targets with KD in the low nanomolar range. Our work provides insights into the antigens that drive B cell responses in human cancers while also directly identifying fully human, high-affinity antibodies against them.

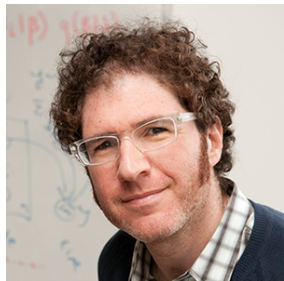


Daniele Biasci is a Principal Investigator at the University of Oxford, Kennedy Institute of Rheumatology, focusing on genetics of tertiary lymphoid structures formation in cancer and autoimmunity. Daniele received his PhD in Immunology from the University of Cambridge and completed his postdoc at the Cancer Research UK Cambridge Institute (CRUK CI) in the Computational Biology Group.

David Blei, Professor of Statistics and Computer Science, Associate Member of the Irving Institute for Cancer Dynamics, Columbia University

Scaling and generalizing approximate Bayesian inference

A core problem in statistics and machine learning is to approximate difficult-to-compute probability distributions, a problem to which Simon Tavaré has made seminal contributions. Such approximations are especially important in Bayesian statistics, which frames all inference about unknown quantities as a calculation about a conditional distribution. In this talk I review and discuss innovations in variational inference (VI), a method that approximates probability distributions through optimization. VI has been used in myriad applications in machine learning and Bayesian statistics. In this talk, I will discuss how to scale VI to large datasets and how to apply VI to models for which other methods are difficult to implement.



David Blei is a Professor of Statistics and Computer Science at Columbia University, and a member of the Columbia Data Science Institute. He studies probabilistic machine learning and Bayesian statistics, including theory, algorithms, and application. David has received several awards for his research. He received a Sloan Fellowship (2010), Office of Naval Research Young Investigator Award (2011), Presidential Early Career Award for Scientists and Engineers (2011), Blavatnik Faculty Award (2013), ACM-Infosys Foundation Award (2013), a Guggenheim fellowship (2017), and a Simons Investigator Award (2019). He is the co-editor-in-chief of the *Journal of Machine Learning Research*. He is a fellow of the Association for Computing Machinery (ACM) and the Institute of Mathematical Statistics (IMS).

Andrew Blumberg, Herbert and Florence Irving Professor of Cancer Data Research (in the Herbert and Florence Irving Institute of Cancer Dynamics and in the Herbert Irving Comprehensive Cancer Center) and Professor of Mathematics and Computer Science, Columbia University

Statistical foundations for time series in topological data analysis

Topological data analysis studies qualitative multiscale descriptors of the shape of data. This talk will describe joint work with Anne van Delft that integrates ideas from functional time series and topological data analysis to provide statistical foundations for studying temporally varying shape information.



Andrew Blumberg is currently the Herbert and Florence Irving Professor of Cancer Data Research (in the Herbert and Florence Irving Institute of Cancer Dynamics and in the Herbert Irving Comprehensive Cancer Center) and Professor of Mathematics and Computer Science. Prior to joining Columbia University, Andrew was Professor of Mathematics at the University of Texas. Prior to arriving in Austin, he was an NSF postdoctoral fellow from 2005-2009 at Stanford, with a year's stint as a member at the Institute for Advanced Study in 2007-2008. Andrew received his PhD in 2005 from University of Chicago. Andrew has broad research interests in mathematics and computer science. His research includes work in algebraic topology, topological data analysis, and computer security and privacy. He is particularly interested in the applications of geometry and topology to the analysis of genomic data.

Ivan Corwin, Professor of Mathematics, Associate Member of the Irving Institute for Cancer Dynamics, Columbia University

Extreme Diffusion -- or was Einstein wrong about diffusion

Diffusion is pervasive in the natural world. Over one hundred years ago Einstein created a remarkably simple and powerful theory describing the behavior of a single diffusing particle. That theory has since been applied countless times to successfully model widely disparate systems. In this talk, I will explain a failure of this theory when applied to systems with many particles diffusing in the same environment. In particular, in such systems, the particles that move the furthest (the extremes of the diffusion) are governed by behaviors much different than would follow from Einstein's theory. I will demonstrate this through analysis of a mathematical model for random walks in a random environment, and discuss ongoing numerical and experimental works to confirm the conclusion that we draw from this model. I will also discuss why studying extreme diffusion is important in some physical, biological, epidemiological, and social applications.



Ivan Corwin is a professor of mathematics at Columbia University. He studies aspects of probability and mathematical physics including random interface growth, interacting particle systems, random matrix theory and stochastic partial differential equations. He received his PhD from the Courant Institute in 2011 and has since held positions at Microsoft Research, MIT, Institut Henri Poincare (at the Poincare Chair), U.C. Berkeley (currently as a Visiting Miller Professor), and Columbia. He has held a Clay Research Fellowship, a Packard Fellowship, and Simons Fellowship and Investigator Award, a Schramm Fellowship and is a Fellow of the American Mathematical Society and of the Institute of Mathematical Statistics. He was the recipient of the 2021 Loeve prize in probability, 2018 Alexanderson Award, 2014 Rollo Davidson Prize, 2012 Young Scientist Prize of the IUPAP, and gave an invited lecture at the 2014 International Congress of Mathematicians.

Christina Curtis, Professor of Medicine, Genetics, Biomedical Data Science, Director of AI and Cancer Genomics, Stanford University, and Investigator, Chan Zuckerberg Biohub

Deterministic evolution and stringent selection during pre-neoplasia

The earliest events during human tumor initiation, while poorly characterized, may hold clues to malignancy detection and prevention. Here we model occult pre-neoplasia by bi-allelically inactivating TP53, a common early event in gastric cancer, in human gastric organoids. Causal relationships between this initiating genetic lesion and resulting phenotypes were established using experimental evolution in multiple clonally derived cultures over two years. TP53 loss elicited progressive aneuploidy, including copy number alterations and structural variants prevalent in gastric cancers, with evident preferred orders. Longitudinal single cell sequencing of TP53 deficient gastric organoids similarly indicates progression towards malignant transcriptional programs. Moreover, high-throughput lineage tracing with expressed cellular barcodes demonstrates reproducible dynamics whereby initially rare subclones with shared transcriptional programs repeatedly attain clonal dominance. This powerful platform for experimental evolution exposes stringent selection, clonal interference and a striking degree of phenotypic convergence in pre-malignant epithelial organoids. These data imply predictability in the earliest stages of tumorigenesis and reveal evolutionary constraints and barriers to malignant transformation.



Christina Curtis is a Professor of Medicine, Genetics, and Biomedical Data Science and Director of AI and Cancer Genomics at Stanford University. Her research has led to new paradigms in understanding how human tumors evolve and metastasize and has redefined the molecular map of breast cancer. Dr. Curtis has been the recipient of numerous awards, including the NIH Director's Pioneer Award and the AACR Award for Outstanding Achievement in Basic Science. She is a Kavli Fellow of the National Academy of Sciences, a Susan G. Komen Scholar, and a Chan Zuckerberg Biohub Investigator. Dr. Curtis serves on the editorial boards of numerous journals, including *Science* and *Cancer Discovery*, as an advisor to biotech, and on the AACR Board of Directors.

Poly H. da Silva, Postdoctoral Research Scientist, Irving Institute for Cancer Dynamics and Department of Statistics, Columbia University

Modeling the arrival of COVID-19 samples in a database with a birth-immigration process

This talk investigates the use of some non-mechanistic continuous stochastic models and their discrete counterparts to predict the appearance of certain biological data, such as Covid DNA sequences, in a database. In particular, by using a generalized version of the Birth-Immigration process, we model the accumulation of Covid DNA sequences at the GISAID database, classifying each arriving sequence as either an existing or a novel type. The goal is to estimate the weekly arrival rates of different variants, providing valuable insights into the spread of the virus in different locations. We analyze the correlation of the number of sub-variants found in samples drawn sequentially from non-overlapping time intervals. The methods developed in this work can be applied to model the entrance of other types of data in databases. This is joint work with Arash Jamshidpey and Simon Tavaré.



Poly H. da Silva is a Postdoctoral Research Scientist in the Department of Statistics and the Irving Institute for Cancer Dynamics working under supervision of Prof. Simon Tavaré. She is interested in discrete mathematics, probabilistic combinatorics, graph theory and algorithms with their applications in bioinformatics, computational biology, genomics, cancer evolution and population dynamics.

Richard A. Davis, Howard Levene Professor of Statistics, Columbia University

Kernel PCA for multivariate extremes

We propose kernel PCA as a method for analyzing the dependence structure of multivariate extremes and demonstrate that it can be a powerful tool for clustering and dimension reduction. Our work provides some theoretical insight into the preimages obtained by kernel PCA, demonstrating that under certain conditions they can effectively identify clusters in the data. We build on these new insights to characterize rigorously the performance of kernel PCA based on an extremal sample, i.e., the angular part of random vectors for which the radius exceeds a large threshold. More specifically, we focus on the asymptotic dependence of multivariate extremes characterized by the angular or spectral measure in extreme value theory and provide a careful analysis in the case where the extremes are generated from a linear factor model. We give theoretical guarantees on the performance of kernel PCA preimages of such extremes by leveraging their asymptotic distribution together with Davis-Kahan perturbation bounds. Our theoretical findings are complemented with numerical experiments illustrating the finite sample performance of our methods. This is joint work with Marco Avella Medina and Gennady Samorodnitsky.



Richard Davis is the Howard Levene Professor of Statistics at Columbia University and former chair of the Statistics Department (2013-19). He has held academic positions at MIT, Colorado State University, and visiting appointments at numerous other universities. He was Hans Fischer Senior Fellow at the Technical University of Munich (2009-12), Villum Kan Rasmussen Visiting Professor (2011-13) at the University of Copenhagen, and Jubilee Professor at Chalmers University (2019). Davis is a fellow of the Institute of Mathematical Statistics and the American Statistical Association, and is an elected member of the International Statistical Institute. He was president of IMS in 2016 and Editor-in-Chief of *Bernoulli Journal* 2010-12. He is co-author (with Peter Brockwell) of the books, *Time Series: Theory and Methods*, and *Introduction to Time Series and Forecasting*.

Bianca Dumitrascu, Herbert & Florence Irving Assistant Professor of Cancer Data Research (in the Herbert and Florence Irving Institute for Cancer Dynamics and in the Herbert Irving Comprehensive Cancer Center) and Assistant Professor of Statistics, Columbia University

Learning statistical representations of embryonic development

During embryonic development, single cells read in local information from their environments and use this information to move, divide and specialize. As a result, the environments themselves change. However, it remains unclear how gene expression programs interact with cell morphology and mechanical forces to orchestrate organogenesis in early embryos. Recent advances in single cell techniques and in toto imaging enable unique venues in exploring this link between genomics and biophysics, which dynamically maps cells to organisms. In this talk, I will describe statistical machine learning frameworks aimed at understanding how tissue level mechanical and morphometric information impact gene expression patterns in spatio-temporal contexts. We use these tools to understand boundary formation in the early development of mouse embryos and to align data from light sheet recordings of pre-gastrulation development.

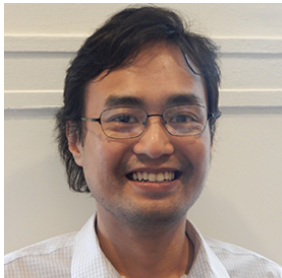


Bianca Dumitrascu is an Assistant Professor in the Department of Statistics and the Irving Institute for Cancer Dynamics at Columbia University, NYC. Bianca works at the intersection of statistical machine learning and genomics. Her current focus is quantifying the relationship between gene regulatory networks, morphology and mechanical forces in wound healing and early development, with tools from explainable machine learning and representation learning. Previously, Bianca was an Affiliated Lecturer in the Department of Computer Science at the University of Cambridge and a Member in the School of Mathematics at the Institute for Advanced Studies. She received a PhD and MS in Computational Biology from Princeton University and a BS in Mathematics from MIT.

Khanh N. Dinh, Postdoctoral Research Scientist, Irving Institute for Cancer Dynamics and Department of Statistics, Columbia University

Modeling and simulation of cancer evolution in single cells

It has been apparent during the DNA sequencing era that cancer is characterized both by small point mutations and indels affecting driver genes and large-scale copy number aberrations (CNAs) and structural variants. We have constructed a mathematical framework for modeling cancer evolution as driven by creation and selection of both of these mechanisms. The framework allows for adaptive cancer-specific population dynamics through negative feedback, and implementation of distinct selection models. We also developed an algorithm that can simulate realistically large cell populations, yet retain statistical accuracy in the sampled cells' phylogeny, while maintaining runtime to the level that is practical for simulation-based parameter inference. The model has been fitted to PCAWG/TCGA data and provided a mechanical link between frequencies of gains and losses of chromosome arms and their distributions of Tumor Suppressor Genes (TSGs) and oncogenes, which has been validated with experimental pan-cancer measures. The model and algorithm can be employed to study effects of CNAs and point mutations on the fitness landscape, or be used to infer relevant biological rates from single-cell DNA data.



Khanh N. Dinh: My background is in theoretical mathematics and numerical analysis. My work as a PhD student focused on developing numerical solvers for the Chemical Master Equation. As a Postdoctoral Research Scientist, my foray into cancer research was through modeling the evolution of Acute Myeloid Leukemia and its progression from Severe Congenital Neutropenia. I developed a stochastic model that linked heterogeneity in AML patient outcomes to random fluctuations affecting the Minimal Residual Disease. Another target of my postdoctoral research was in development of an algorithm to decompose the Site Frequency Spectra from bulk DNA-seq data into the neutral component and selective humps. This algorithm was further extended to account for spatial information. My current project involves development of a mathematical model, simulation algorithm and inference method for cancer evolution in single cells.

Alison Etheridge, Professor of Probability, Mathematical Institute and Department of Statistics, University of Oxford

Forwards and backwards in spatially heterogeneous populations

We introduce a broad class of mechanistic models that might describe how spatially heterogeneous populations live, die, and reproduce. Questions we (start to) address include: how does population density change across space and time? And how does genetic ancestry spread across geography when looking back through time in these populations? A novelty is that by explicitly splitting reproduction into two phases (production of juveniles and their maturation) we produce a framework that not only captures models that when suitably scaled converge to classical reaction diffusion equations, but also ones with nonlinear diffusion that exhibit quite different behaviour. This is joint work with Tom Kurtz (Madison), Peter Ralph (Oregon) and Ian Letter and Terence Tsui (Oxford).



Alison Etheridge is a mathematician with a background in probability and mathematical analysis. After student life in Oxford and McGill, she worked in Berkeley, Cambridge, Edinburgh, and London before returning to Oxford. Much of her work focuses on infinite dimensional stochastic processes and their applications. Such processes are essential if one is interested in modelling populations evolving in spatial continua. In recent years her central interest has been a collection of mathematical problems arising in population genetics.

Jellert Gaublomme, Assistant Professor in Biological Sciences, Associate Member of the Irving Institute for Cancer Dynamics, Columbia University

Optical pooled multi-omic single-cell screening

Pooled CRISPR screens enable perturbation studies of pathways at scale, and with single-cell resolution. Until recently the phenotypic readout associated with pooled screens was typically limited to fitness, or the expression of a tagged marker protein. Single-cell RNA-sequencing vastly expanded our readout to the whole transcriptome, but often relies on lysis and is thus oblivious to cell morphology and subcellular protein localization patterns as a function of the incurred CRISPR perturbation. Here, we show our development of a multi-omic optical pooled screening approach that combines in situ RNA, protein and barcode readout at the single-cell level.



Jellert Gaublomme is an Assistant Professor in the department of Biological Sciences at Columbia University. His lab develops and applies multi-omic technologies to study biomolecular and cellular interactions. The application areas include immunology and cancer. He received his PhD in 2014 from Harvard University with Professor Hongkun Park, and subsequently did a postdoctoral fellowship at the Broad Institute with Professor Aviv Regev.

Trevor Graham, Director of the Centre for Evolution and Cancer, Institute of Cancer Research

Inferring the dynamics of cancer evolution using stochastic branching processes

Cancers evolve. But rarely can we directly observe this dynamic process, because our data derives from single samples that capture only snapshots in time. I will discuss how we can construct simple mathematical models of tumour evolution and fit them to these "snapshots" (cancer genome sequencing data derived from a single point in time) to learn the unobservable evolutionary dynamics of cancer development. This productive line of work was first inspired by a visit to Simon Tavaré's lab in 2005, when I was an early PhD student, and Simon opened my eyes to stochastic processes and Bayesian inference.



Prof. Trevor Graham is Director of the Centre for Evolution and Cancer at the Institute of Cancer Research, London. His background is in mathematics and his interdisciplinary research group combines theory and experiment to understand how cancers evolve. For his contributions to the understanding of cancer genome evolution he was elected a Fellow of the Academy of Medical Sciences in 2022. Trevor's love for stochastic processes and cancer evolution was cemented by a visit, as a young PhD student, to Simon Tavaré's and Darryl Shibata's groups at USC in 2005, where Trevor's lack of surfing skills were also bitterly exposed.

Robert Griffiths, Adjunct Professor, School of Mathematics, Monash University

Coalescence in Feller branching diffusions

A Feller branching diffusion $X(t)$ has infinitesimal mean λx and variance λx . The process is subcritical if $\lambda < 0$, critical if $\lambda = 0$ and supercritical if $\lambda > 0$. It can be thought of as a limit from a Bienayme-Galton-Watson (BGW) process. Coalescence is described by $A_n(s,t)$, the number of ancestors of a sample of n individuals taken at time t , at an earlier time $t-s$. Ancestral lineages “come down from infinity” and the number of ancestors of the population $A_\infty(s,t)$ has a proper distribution which is compound Poisson-Geometric. The limit coalescent distribution as t tends to infinity, conditional on non-extinction at t , in the subcritical case, has the property that the mean time between coalescent events is the same as in the Kingman coalescent, but the distribution is different. Another limit is to take s tending to zero in a supercritical model. The asymptotics of the coalescent process, looking at s back to zero, are related to the reduced tree of a non-homogeneous birth and death process. Applications are made to finding approximate distributions for models with d types in the Feller branching process, where individuals have a general mutation scheme between types and the overall mutation rate is small. This is joint research with Conrad Burden from the Australian National University.



Robert Griffiths is an Adjunct Professor of Mathematics at Monash University. He was an academic at Monash University for 25 years from 1973, then at the University of Oxford from 1998 for 20 years, an Emeritus Professor since 2012, returning to Australia in 2018. His main research interest is in Stochastic Processes in Mathematical Population Genetics. He was elected as a Fellow of the Royal Society in 2010 for his contributions to Mathematical Genetics.

Arash Jamshidpey, Associate Research Scientist, Irving Institute for Cancer Dynamics and Department of Mathematics, Columbia University

Asymptotic behavior of multiple samples drawn from a birth process with immigration

We study the joint distribution of the number of families observed in multiple samples drawn sequentially from disjoint time intervals of a birth process with immigration. We investigate the asymptotic behavior of the joint distribution and find conditions under which the system is exchangeable. We provide some limit theorems under certain conditions, as the number of samples and/or their average sizes tend to infinity. We extend the model to one with time-varying mutation rates and relate it to the sampling theory from the equilibrium of more complicated population models.



Arash Jamshidpey is an Associate Research Scientist in the Irving Institute for Cancer Dynamics and the Department of Mathematics. Before this, he was a postdoc at Columbia University, working under the supervision of Prof. Simon Tavaré. He received his PhD in mathematics from University of Ottawa, under the supervision of Prof. Donald Dawson and Prof. David Sankoff. Arash is interested in probability theory and its applications in population genetics, cancer evolution, genomics and phylogenetics. Over the past years, he has been working on the effects of environmental changes on various continuous and discrete population models in which mutation and selection interact and fluctuate in time. His current research includes measure-valued processes, branching processes, birth-death processes, interacting particle systems in random environments, mathematical genomics and phylogenetics.

Russell Kunes, PhD Student, Department of Statistics and Irving Institute for Cancer Dynamics, Columbia University

Gradient estimation for binary latent Variables via gradient variance clipping

Gradient estimation is often necessary for fitting generative models with discrete latent variables, in contexts such as reinforcement learning and variational autoencoder (VAE) training. The DisARM estimator (Yin et al. 2020; Dong, Mnih, and Tucker 2020) achieves state of the art gradient variance for Bernoulli latent variable models in many contexts. However, DisARM and other estimators have potentially exploding variance near the boundary of the parameter space, where solutions tend to lie. To ameliorate this issue, we propose a new gradient estimator bitflip-1 that has lower variance at the boundaries of the parameter space. As bitflip-1 has complementary properties to existing estimators, we introduce an aggregated estimator, $\text{unbiased gradient variance clipping}$ (UGC) that uses either a bitflip-1 or a DisARM gradient update for each coordinate. We theoretically prove that UGC has uniformly lower variance than DisARM. Empirically, we observe that UGC achieves the optimal value of the optimization objectives in toy experiments, discrete VAE training, and in a best subset selection problem.



Russell Kunes is a PhD student in the Columbia Department of Statistics (Simon Tavaré lab). He is interested in developing interpretable machine learning tools for studying gene regulation in cancer. Previously, he has worked on spectral algorithms for network inference and variable selection methods for supervised topic models. He graduated from Harvard University with a B.A. in Statistics and Mathematics.

John Marioni, Senior Vice President and Head of Computation, Genentech Research and Early Development, Genentech, CA, USA

Beyond reference atlases - computational approaches for single-cell genomics

Over the past 10 years novel experimental approaches, powered by new computational methods, have enabled the generation of molecular profiles from millions of individual cells, resulting in atlases of multiple tissues and organs. In this presentation, I will discuss some of the computational approaches that have underpinned this analysis, before turning to challenges that are faced in effectively using these atlases to answer questions in the context of both normal development and disease.



John Marioni is the Senior Vice President and Head of Computation at Genentech Research and Early Development (gRED) in South San Francisco. Previously, he was Head of Research at the EMBL-EBI and a Senior Group Leader at the CRUK Cambridge Institute within the University of Cambridge. John read for his PhD under the supervision of Simon Tavaré before working as a postdoc under the supervision of Matthew Stephens. He is a member of EMBO and a Fellow of the United Kingdom's Academy of Medical Sciences.

Magnus Nordborg, Scientific Director of the Gregor Mendel Institute of Molecular Plant Biology

The genetics of epigenetics in Arabidopsis

I will describe our efforts to carry out a 1001 Genomes Project in the model plant *Arabidopsis thaliana*, and what we have learned about the genetic architecture of various traits, in particular molecular ones, like DNA methylation.



Dr. Magnus Nordborg, born in Stehag, Sweden, completed his PhD on theoretical population genetics at Stanford in 1995. Thereafter he made significant contributions to coalescence theory at the University of Chicago, and later became well-known for his pioneering work at USC on genome-wide association studies (GWAS) in non-human organisms, working both independently and with leading gene researchers such as Joe Ecker and Detlef Weigel. He has over 110 research publications, including two in *Nature*, four in *Science* and ten in *Nature Genetics*. He has been Scientific Director of the Gregor Mendel Institute of Molecular Plant Biology (Vienna) since 2009, is Principal Investigator on an ERC Advanced Grant, a Member of the American Association for the Advancement of Science (AAAS), a member of EMBO and a Corresponding Member of the Austrian Academy of Sciences.

Karol Nowicki-Osuch, Associate Research Scientist, Irving Institute for Cancer Dynamics, Columbia University

Precancerous lesions of gastric and esophageal cancers show profound phenotypic similarity

Esophageal adenocarcinomas (EAC) and intestinal gastric cancers (IGC) are jointly the third leading cause of cancer deaths worldwide. EAC and IGC share histological and molecular features and are associated with Intestinal Metaplasia (IM) precancerous lesions – Barrett’s Esophagus (BE) and Gastric Intestinal Metaplasia (GIM), respectively. The cells-of-origin of BE and GIM have been debated and our recent data suggest a gastric origin for BE. Identical origin (gastric cells) of GIM and BE and molecularly similar final destination – adenocarcinomas – suggest that the precancerous lesion might also be molecularly similar entities. A comparative RNA-seq analysis of GIM and BE at the single-cell level showed a profound similarity between all three phenotypes. Computational deconvolution using single-cell methods characterized non-goblet cell types of GIM and BE as single-cell chimeras composed of gastric and small intestinal phenotypes. Furthermore, in line with the histological assessment, we observed an absence of parietal and chief cells and a lack of goblet cells in these samples. The chimeric state was driven by gastric foveolar and mucous neck cells. Sub-clustering of mucous neck cells (marked by MUC6 expression) showed that pathological changes are caused by the acquisition of intestinal stem-like phenotype by these cells. Taken together these data suggest that mucous neck cells are the cells of origin for gastric and esophageal IM. Our data suggest that although triggered by different pathological stimuli, gastric and esophageal IM are profoundly similar diseases likely originating from gastric mucous neck cells and their convergence indicates shared development pathways of EAC and IGC.



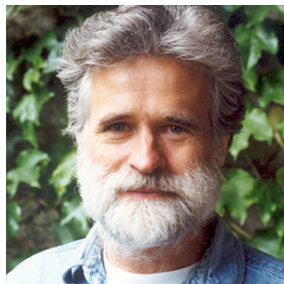
Karol Nowicki-Osuch is an Associate Research Scientist in IICD. Before joining the Institute, Karol was a Postdoctoral Research Associate in Prof Rebecca Fitzgerald's laboratory at the University of Cambridge UK since 2016. Karol’s experience and interests lie at the interface between wet- and dry-lab cancer biology research. Over recent years, his work has been focused on questions associated with the early stages of esophageal cancer development including the transition from normal to cancer tissues. As a Research Associate Scientist, Karol will spearhead

IICD's major experimental focus on the development of experimental and computational methods for the analysis of solid tumors in three dimensions in partnership with the New York Genome Center and Cambridge University.

Jim Pitman, Emeritus Professor, Statistics and Mathematics, University of California, Berkeley

The range of a self-similar additive gamma process is a scale invariant Poisson point process

For a positive self-similar stochastic process T with independent increments, the range of T forms a Poisson point process with sigma-finite intensity if and only if the one-dimensional distribution of $T(1)$ is of the gamma type. The "if" part of this result is implicit in the proof by Arratia, Barbour and Tavaré that the spacings between consecutive points of a scale invariant Poisson point process on the positive half line are the points of another scale invariant Poisson point process with the same intensity. The result is also related to the theory of records and to Gneden's Poisson corner process. This is joint work with Zhiyi You. A preliminary version is available at <https://arxiv.org/abs/2111.09409>



Jim Pitman is an Emeritus Professor of Statistics and Mathematics at the University of California, Berkeley. The main theme of his research career has been the study of various kinds of stochastic processes, principally Markov processes, and the analysis of how the distributions of such processes are affected by various operations such as path transformations and conditionings. In recent years, he has become interested in interfaces between this core of ideas and ideas in other areas of mathematics, especially combinatorics, special functions, and analytic number theory.

Molly Przeworski, Professor, Departments of Biological Sciences and Systems Biology, Columbia University

Relating pathogenic loss-of-function mutations in humans to their evolutionary fitness costs

Causal loss-of-function (LOF) variants for Mendelian and severe complex diseases are enriched in 'mutation intolerant' genes. We show how such observations can be interpreted in light of a model of mutation-selection balance, and use the model to relate the pathogenic consequences of LOF mutations at present-day to their evolutionary fitness effects. To this end, we first infer posterior distributions for the fitness costs of LOF mutations in 17,318 autosomal and 679 X-linked genes from exome sequences in 56,855 individuals. Estimated fitness costs for the loss of a gene copy are typically above 1%; they tend to be largest for X-linked genes, whether or not they have a Y homolog, followed by autosomal genes and genes in the pseudoautosomal region. We then compare inferred fitness effects for all possible de novo LOF mutations to those of de novo mutations identified in individuals diagnosed with one of six severe, complex diseases or developmental disorders. Proband carry an excess of mutations with estimated fitness effects above 10%; as we show by simulation, when sampled in the population, such highly deleterious mutations are typically only a couple of generations old. Moreover, the proportion of highly deleterious mutations carried by probands reflects the typical age of onset of the disease. The study design also has a discernible influence: a greater proportion of highly deleterious mutations is detected in pedigree than case-control studies, and for autism, in simplex than multiplex families and in female versus male probands. Thus, anchoring observations in human genetics to a population genetic model allows us to learn about the fitness effects of mutations identified by different mapping strategies and for different traits.



Molly Przeworski is a population geneticist working at the interface of human genetics and evolutionary biology. She received a B.A. in Mathematics from Princeton University and a Ph.D. from the Committee on Evolutionary Biology at the University of Chicago. Her postdoc was in the Statistics Dept. of the University of Oxford, and was followed by a two-year stint as a research scientist at the Max Planck Institute for Evolutionary Anthropology. Before moving to Columbia University in 2014, she was a faculty member at the University of Chicago (where she was also a Howard Hughes Medical Institute Early Career Scientist) as well as, briefly, at Brown University.

Sanket Rane, Associate Research Scientist, Irving Institute for Cancer Dynamics, Columbia University

Build, break & restore: decoding the rules that define and reshape immune equilibria

Our T and B lymphocyte populations are multilayered, organized ecosystems of functionally diverse subsets that work together to elicit stunningly rich and dynamic responses against infections and tumors. The sizes and clonal architecture of lymphocyte subpopulations are determined by an intricate interplay of production, proliferative renewal, death, and differentiation. During an immune response, the relative contributions of these processes vary dynamically, and after the resolution of the response, new equilibria are established. The mechanisms that determine this reconfiguration of B and T cell ecology after antigenic perturbations, such as infections or malignancies, remain unclear. This talk will focus on our efforts to combine mathematical modeling with dedicated experiments to quantify, compare, and connect lymphocyte dynamics at steady state and during immune responses in order to understand how their sizes and clonal structures evolve over a lifetime.



Sanket Rane is a theoretical immunologist trained in deterministic and probabilistic mathematical modeling, inference-driven Bayesian statistical methods, and cell biology techniques. Sanket earned his PhD in immunology in 2015 at the National Institute of Immunology, India, for investigating molecular and cellular factors that govern T cell numbers and function during healthy aging. He transitioned into using computational approaches to study the lymphocyte ecology during his postdoctoral years. The overarching theme of his research is to synthesize tightly-knit, iterative connections between math and biology to build a quantitative map of our immune competence in healthy aging, infections, and cancer pathologies.

Gesine Reinert, Professor, Department of Statistics, University of Oxford

A multilayer network approach for a COVID-19 multimodal dataset

Developments in experimental biology have enabled the collection of multiple molecular modalities per patient for large cohorts, increasing the importance of developing methods to combine these datasets. First, we construct single-modality patient-to-patient similarity networks, ensuring their optimality via our resampling approach, COGENT. Then we combine them into a multilayer network, coupling the layers according to the level of information shared between them, before carrying out community detection. Here we concentrate on data from the COVID-19 Multi-omic Blood Atlas, which also includes data from healthy volunteers. Communities identified on this multilayer network show a strong association with the major clinical diagnosis while providing a more fine-grained view of the disease. The communities can be linked to differential activities of biological pathways as well as to clinical features, thus aiding a better understanding of this multimodal molecular health dataset. This is based on joint works with Mariano Beguerisse-Diaz, Lyuba Bozhilova, Charlotte Deane, Heather Harrington, Julian Knight, Javier Pardo-Diaz, Philip Poole, and Piotr Sliwa.

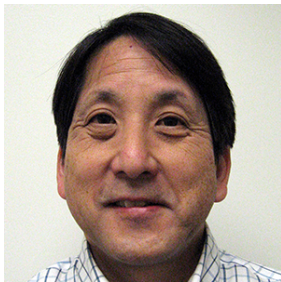


Gesine Reinert is a Research Professor in the Department of Statistics at the University of Oxford, and a Fellow of the Alan Turing Institute. She is also a Fellow of the Institute of Mathematical Statistics. Her research includes probabilistic and statistical methods for network analysis, as well as applied probability and computational biology, and connections with machine learning methods.

Darryl Shibata, Professor of Pathology, University of Southern California Keck School of Medicine

Contemporary somatic cell lineage tracing with a rapidly fluctuating tick tock molecular clock

The pasts of human tissues can be reconstructed by sequencing their genomes. A shortfall is the ability to infer more recent events because most “molecular clocks” tick slowly. A new type of tick tock clock based on “rapid” (~1 per 1,000 divisions) fluctuations between methylated and unmethylated states at hundreds of fCpG sites facilitates more contemporary lineage tracing. In bulk polyclonal populations, fCpGs states are unsynchronized, with average methylation ~50%. However, recent bottlenecks leading to clonal populations synchronize fCpGs to the pattern in the progenitor, resulting in “W”-shaped distributions, with trimodal peaks at 0, 50 and 100% methylation. W-shaped fCpG distributions are seen in “clonal” intestinal crypts, and quantitative analysis can infer stem cell numbers and their turnover. fCpGs can record the dynamics of neoplasia, with W-shapes in acute leukemia and more broad distributions in chronic leukemias. Multiregional sampling of fCpGs in colorectal cancer tissue sections are consistent with single expansions. Integration of fCpG clocks with other multiomic and spatial approaches can help reconstruct normal and neoplastic tissue dynamics with higher resolution of clinically relevant (months) time frames.



Darryl Shibata is a Professor of Pathology at the University of Southern California Keck School of Medicine. He started working with Dr. Simon Tavaré around the turn of the century, and early DNA methylation clock studies inferred that human colon crypts were maintained by multiple, randomly surviving, mitotic stem cells. He is interested in human tissue reconstructions with an emphasis on intestinal crypts and colorectal cancers.

Ignacio Vazquez-Garcia, Postdoctoral Research Scientist in the Irving Institute for Cancer Dynamics, Columbia University, and a Research Fellow in Computational Oncology, Memorial Sloan Kettering

Chromosomal instability and its role in cancer genome evolution

Chromosomal instability (CIN) is a major driver of tumor progression and treatment resistance in cancer. CIN results from ongoing errors in chromosome segregation during cell division, generating copy number heterogeneity that provides a substrate for natural selection. However, the dynamic impact of CIN on the number and structure of chromosomes remains underexplored, particularly in the context of human cancer. Newly developed single-cell assays now allow for the measurement of genomes and transcriptomes of single cells obtained from patient tissues, providing a direct view into the evolution of tumor genomes undergoing CIN and into their surrounding microenvironment. To study the role of CIN in cancer evolutionary dynamics, we designed a prospective multi-modal study of high-grade serous ovarian cancer (HGSOC), an archetypal tumor with high CIN. We performed multi-region analyses of 128 samples from 45 HGSOC patients using single-cell whole genome sequencing (scWGS), single-cell RNA sequencing (scRNA-seq), and in situ immunofluorescence (IF) microscopy. These single-cell and spatial measurements enable us to estimate rates of chromosome missegregation, genome doubling and micronucleus formation. We are able to trace likely paths from euploid founder cells to a heterogeneous tumor cell population, identifying divergent evolutionary trajectories related to loss of DNA repair mechanisms. Furthermore, we establish a mechanistic link between CIN rates and immune evasion, revealing an interplay between cytosolic DNA sensing and inflammatory signaling in the natural history of the disease. Together, our findings reveal new insights into the evolution of ovarian cancer, beginning with loss of heterozygosity of tumor suppressors and subsequent tolerance of WGD. This is followed by a transient state of CIN, with elevated rates of chromosome missegregation, micronucleus rupture, and potent activation of immune signaling in the tumor and its microenvironment. These scaled measurements of CIN and the tumor microenvironment opens new avenues for quantitative modeling of cellular fitness in human cancer.



Ignacio Vazquez-Garcia is a Research Fellow in Computational Oncology at Memorial Sloan Kettering and Columbia University, working with Dr. Sohrab Shah and Dr. Simon Tavaré. His research focuses on fundamental principles of somatic evolution in human cancer. He currently focuses on the role of genome instability and tumor-immune co-evolution in driving the emergence of malignancy from normal tissues towards lethal and resistant disease. He integrates innovations in computational modeling with technology development, particularly in single-cell technologies which are beginning to highlight intra-tumoral heterogeneity in cell genomes, cell states, spatial dynamics and interactions with the tumor microenvironment.

Sanja Vickovic, Director of Technology Innovation and Core Faculty, New York Genome Center/Assistant Professor of Biomedical Engineering (in the Herbert and Florence Irving Institute for Cancer Dynamics), Columbia University/Wallenberg Academy Fellow at Science for Life Laboratory and Uppsala University

Connections in cells and tissues

Mucosal and barrier tissues such as the gut, lung or skin, are composed of a complex network of cells and microbes forming a tight niche that prevents pathogen colonization and supports host-microbiome symbiosis. Characterizing these networks at high molecular and cellular resolution is crucial for our understanding of homeostasis and disease. Spatial transcriptomics has recently emerged as a key technique to capture and positionally barcode RNAs directly in tissues. Today, I will introduce two of the recent advances in the application of spatial transcriptomics at scale, by presenting Spatial Multi-Omics (SM-Omics); as a fully automated, high-throughput platform for combined and spatially resolved transcriptomics and antibody-based protein measurements, and Spatial Host-Microbiome sequencing (SHM-seq); an all-sequencing based approach that captures tissue histology, polyadenylated RNAs and bacterial 16S sequences directly from tissues on spatially barcoded glass surfaces.



Dr. Sanja Vicković is a Core Faculty Member and the Director of the Technology Innovation Lab, at the New York Genome Center. She holds joint appointments as an Assistant Professor at the Fu Foundation School of Engineering and Applied Science and the Herbert and Florence Irving Institute for Cancer Dynamics at Columbia University, and as a Wallenberg Academy Fellow of the Royal Swedish Academy of Sciences and the Royal Swedish Academy of Engineering Sciences at Uppsala University. Dr. Vicković is an experienced and accomplished engineer and an inventor of the spatial transcriptomic technology called “Visium” and now commercialized by 10x Genomics.

John Wakeley, Professor, Department of Organismic and Evolutionary Biology, Harvard University

Multiple mutations in the ancestry of rare SNP variants

Mutation rates vary over two orders of magnitude among sites in the human genome. A sampling theory for multiple latent mutations in the ancestry of a rare variant is presented. Results for constant-size populations are related to Ewens sampling formula. Results for growing or otherwise fluctuating populations follow an independent-Poissons model. These results can explain the widely divergent patterns of rare SNP frequencies among sites in recent surveys of human genomic variation.



John Wakeley is Professor of Organismic and Evolutionary Biology at Harvard University. His main area of research is the theory of gene genealogies. Current projects include studies of human genomic variation, the roles of organismal genealogies in structuring genetic variation, and the evolutionary effects of extreme variation in offspring numbers. He is the author of a textbook on the mathematics of gene genealogies: *Coalescent Theory, An Introduction*.

Richard Wilkinson, Professor of Statistics, School of Mathematical Sciences, University of Nottingham

Adjoint aided inference of linear systems

Linear systems occur throughout engineering and the sciences, most notably as differential equations. In many cases the forcing function for the system is unknown, and interest lies in using noisy observations of the system to infer the forcing, as well as other unknown parameters. In this talk, I will show how the adjoint of a linear system can be used to efficiently infer forcing functions modelled as Gaussian processes (GPs), using a truncated basis expansion of the GP kernel. In this case, exact conjugate Bayesian inference for the truncated GP can be achieved, often with substantially lower computation than would be required using MCMC methods.



Richard Wilkinson is Professor of Statistics at the University of Nottingham. He works on uncertainty quantification problems, primarily on how to fit complex models to data. He did his PhD with Simon Tavaré, finishing in 2008, on approximate Bayesian computation methods to estimate primate divergence times.

Bin Yu, Chancellor's Distinguished Professor and Class of 1936 Second Chair in Statistics, EECS and Computational Biology, University of California, Berkeley

Fast interpretable greedy-tree sums with applications to clinical decision instruments

Modern machine learning has achieved impressive prediction performance, but often sacrifices interpretability, a critical consideration in high-stakes domains such as clinical-decision-instrument modeling. In such settings, practitioners often turn to highly interpretable shallow decision tree models, but these suffer from inductive bias against additive structure, which limits their prediction performance. To extend decision trees to capture additive structure in data, we propose Fast Interpretable Greedy-Tree Sums (FIGS), which generalizes the CART algorithm to simultaneously grow a flexible number of trees in a summation. FIGS combines logical rules with addition, and so remains highly interpretable. In particular, applying FIGS to 3 clinical-decision datasets results in models that reflect domain knowledge and enjoy substantially improved specificity (by up to 20%) without sacrificing sensitivity or interpretability. All code and models are released in a full-fledged package on GitHub.



Bin Yu is Chancellor's Distinguished Professor and Class of 1936 Second Chair in Statistics, EECS, and Computational Biology at UC Berkeley. Her research has focused on the practice and theory of ML and solving interdisciplinary data problems in neuroscience, genomics, and precision medicine, including developing algorithms such as iterative random forests (iRF), stability-driven NMF, and adaptive wavelet distillation (AWD) from deep learning models. She is a member of the National Academy of Sciences and of the American Academy of Arts and Sciences. She was a Guggenheim Fellow, and holds an Honorary Doctorate from The University of Lausanne. She served on the inaugural scientific advisory board of the UK Turing Institute of Data Science and AI and is serving on the editorial board of PNAS.

Poster Session

Friday, March 17th (5:30-7:30 PM)

Faculty House, Skyline Dining Room

Poster presenters are highlighted in bold.

1. Fast and interpretable genomic data analysis using multiple approximate kernel learning

Ayyüce Begüm Bektaş, Çiğdem Ak, Mehmet Gönen

Memorial Sloan Kettering Cancer Center

Dataset sizes in computational biology have been increased drastically with the help of improved data collection tools and increasing size of patient cohorts. Previous kernel-based machine learning algorithms proposed for increased interpretability started to fail with large sample sizes, owing to their lack of scalability. To overcome this problem, we proposed a fast and efficient multiple kernel learning (MKL) algorithm to be particularly used with large-scale data that integrates kernel approximation and group Lasso formulations into a conjoint model. Our method extracts significant and meaningful information from the genomic data while conjointly learning a model for out-of-sample prediction. It is scalable with increasing sample size by approximating instead of calculating distinct kernel matrices.

To test our computational framework, namely, Multiple Approximate Kernel Learning (MAKL), we demonstrated our experiments on three cancer datasets and showed that MAKL is capable to outperform the baseline algorithm while using only a small fraction of the input features. We also reported selection frequencies of approximated kernel matrices associated with feature subsets (i.e. gene sets/pathways), which helps to see their relevance for the given classification task. Our fast and interpretable MKL algorithm producing sparse solutions is promising for computational biology applications considering its scalability and highly correlated structure of genomic datasets, and it can be used to discover new biomarkers and new therapeutic guidelines.

MAKL is available at <https://github.com/begumbektas/makl> together with the scripts that replicate the reported experiments. MAKL is also available as an R package at <https://cran.r-project.org/web/packages/MAKL>.

2. Evolutionary Dynamics of Oligodendrogliomas

Sarah Benedetto, Verena Korber, Yonghe Wu, Jörg Felsberg, Guido Reifenberger, Bernhard Radlwimmer, Peter Lichter, Thomas Hofer

DKFZ - German Cancer Research Center

The development and progression of cancer is an evolutionary process. Deep genome sequencing data now allow us to time the origin and study the progression of individual tumors. We developed a model based on the birth-death process to infer tumor evolution from whole-genome sequencing data and applied it to six samples of an adult brain tumor, grade II oligodendrogliomas. Specifically, we used statistical and population-genetic approaches to infer the clonal evolution of primary tumors. We found that the tumors originated in early childhood and thereafter evolved slowly. Within the primary tumor, we found selected subclones in all samples with widely different selective advantages. In weakly selected subclones, we did not detect known tumor drivers whereas subclones harboring an activating mutation in the TERT promoter, which supports telomere maintenance, were strongly selected. These findings suggest that mutations in the TERT

promoter may be acquired during oligodendroglioma evolution and confer a large selective advantage by which they eventually reach fixation in the majority of cases. Thus, TERT promoter mutations supporting telomere maintenance may be a key determinant of the aggressiveness of the disease.

3. Computational Multiscale Model of Cancer Cell Migration and Invasion Phenotypes

Temitope O. Benson, Ashlee N. Ford Versypt

University of Buffalo

The spread of cancer cells from a localized tumor mass in one part of the body to another is known as metastasis. It plays a crucial role in cancer-related death in cancer patients and reducing the efficacy of cancer treatment. Cancer cells in a tumor mass interact with one another as well as their local tumor microenvironment, particularly the extracellular matrix (ECM), during metastasis. The ECM undergoes structural remodeling of biochemical, physical, and mechanical characteristics because of this interaction. Cancer cells exhibit different mode of migration and invasion properties which includes single and collective migration modes.

Single and collective cancer cell migration from the tumor is influenced by this structural remodeling. The processes and techniques that produce these cancer cell migratory characteristics are still unknown. Our group used the free open-source software CompuCell3D to create a computer model that mimicked in vivo cancer cell movement across the ECM during structural remodeling. Here, we discuss how we used in vitro migration tests for varied ECM collagen fiber concentrations and pore diameters to evaluate phenotypic changes from single cell to collective cell migration to validate this model. We also investigate the impact of cell adhesion. Chemotaxis-induced cancer cell motility is also investigated and quantified. The cancer cells are represented as discrete agents in our model, and the ECM components, including collagen fibers and remodeling enzyme(s), are modeled using a system of partial differential equations. The goal is to be able to validate the in vitro model and then go on to in vivo investigations to provide cancer metastasis prediction capacity.

4. Bayesian imputation of unmeasured metabolites

Casey Bradshaw, Amy Xie, Ed Reznik, Wesley Tansey

Columbia University

Metabolomics data is a rich source of insight into the cellular processes occurring in a biological specimen, which in turn has implications for our understanding of disease states. Modern metabolomics experiments are typically designed to measure a small subset of the metabolites present in a tissue sample, and the particular subset chosen varies across experiments. We present a method for leveraging data from multiple studies to impute metabolites which were unmeasured in some, but not all, of those studies. This method represents the metabolite abundance rankings via a Bayesian Plackett-Luce model, and performs approximate inference to generate the imputed metabolite ranks. Extending this analysis to allow imputation of metabolites from transcriptomic data is a key area of interest, as the latter is much more abundant and readily available.

5. Mathematical modeling identifies optimum palbociclib-fulvestrant dose administration schedules for the treatment of estrogen receptor-positive breast cancer patients

Yu-Chen Cheng, Shayna Stein, Agostina Nardone, Weihan Liu, Wen Ma, Gabriella Cohen, Cristina Guarducci, Thomas McDonald, Rinath Jeselsohn, Franziska Michor

Dana Farber Cancer Institute

Cyclin-dependent kinases 4/6 (CDK4/6) inhibitors such as palbociclib are approved for the treatment of metastatic estrogen receptor-positive (ER+) breast cancer in combination with endocrine therapies, and significantly improve outcomes in patients with this disease. However, given the large number of possible pairwise drug combinations and administration schedules, it remains unclear which clinical strategy would lead to best survival. Here, we developed a computational, cell cycle-explicit model to characterize the pharmacodynamic (PD) response to palbociclib-fulvestrant combination therapy. This PD model was parameterized, in a Bayesian statistical inference approach, using in vitro data from cells with wild-type estrogen receptor (WT-ER) and cells expressing the activating missense ER mutation, Y537S, which confers resistance to fulvestrant. We then incorporated pharmacokinetic (PK) models derived from clinical data into our computational modeling platform. To systematically compare dose administration schedules, we performed in silico clinical trials based on integrating our PD and PK models as well as considering clinical toxicity constraints. We found that continuous dosing of palbociclib is more effective for lowering overall tumor burden than the standard, pulsed-dose palbociclib treatment. Our mathematical modeling and statistical analysis platform provide a rational method for comparing treatment strategies in search of optimal combination dosing strategies with cell cycle inhibitors in ER+ breast cancer.

6. Genealogies in expanding populations and stochastic reaction-diffusion waves on metric graphs

Wai-Tong (Louis) Fan, Rick Durrett, Wenqing Hu, Greg Terlov, Johnny Yang, John Yin

Indiana University

In spatial population genetics, it is important to understand the probability of extinction in multi-species interactions such as growing bacterial colonies, cancer tumor evolution and human migration. This is because extinction probabilities are instrumental in determining the probability of coexistence and the genealogies of populations. A key challenge is the complication due to spatial effect and different sources of stochasticity. In this poster, I will discuss about methods to compute the genealogies, the probability of extinction and other long-time behaviors for stochastic reaction-diffusion equations on metric graphs that flexibly parametrizes the underlying space.

7. Optimal Dosing Strategies in the Presence of Persister-Cells: Combination of EGFR inhibitors and MET inhibitor in first-line treatment of EGFR-mutant NSCLC

Christopher Graser, Kamrine Poels, Aya Shiba, Aaron Hata, and Franziska Michor

Dana Farber Cancer Institute

To delay the emergence of resistance in EGFRm non-small cell lung cancer (NSCLC) patients, a three-drug combination of two EGFR inhibitors and one MET inhibitor has been proposed. However, due to joint toxicity, maximally two out of these three drugs can be administered simultaneously. Here, we developed a mathematical model of tumor evolution and drug-sensitivity in this system, which allows us to optimize treatment schedules in a patient-specific manner. We parameterized this model using drug-assays and measuring division and apoptosis dynamics in real time. Generally, we show that treatment regimens that cycle quickly through all drug-combinations are most potent, once resistant cells have emerged in large proportions. For more rare clones of resistant cells, on the other hand, regimens with longer treatment cycles perform best. Our model, moreover, not only considers drug-resistance, but also explicitly models persister cells, which transiently enter a less proliferative, but also less drug-sensitive state. This allows us to explore treatment regimens that include treatment pauses during which the persister sub-clones become re-sensitized. Overall, our results suggest that optimizing treatment schedules can significantly delay the outgrowth of resistant clones.

8. Altered circadian rhythms in Luminal A breast tumors modulate chemotherapeutic targets, metastatic potential, and tumor prognosis

Jan A. Hammarlund, S-Y Li, G Wu, JB Hogenesch, Q-J Meng, RC Anafi

University of Pennsylvania

The molecular circadian clock regulates thousands of genes in a tissue-specific manner. The influence of circadian time on cell division – and by extension cancer - is particularly strong. Data from both shift workers and model organisms suggest that circadian disruption can increase the risk of breast cancer. Individual tumors likely use distinct machinery to disrupt the normal circadian regulation of cell division. While chronomedicine promises to improve therapy, our inability to describe rhythms in specific healthy human tissues and tumors has been a key barrier to clinical translation.

We modified CYCLOPS (CYClic-Ordering-by-Periodic-Structure), an established method for circadian data ordering to better account for confounding variables. Then, we combined RNA-seq data from 26 time-stamped, clinical breast biopsy pairs with data from the Genotype-Tissue Expression (GTEx) project and the Tissue Cancer Genome Atlas (TCGA). For time-stamped, non-cancerous samples, the CYCLOPS ordering was well correlated with collection time. The acrophases of core circadian genes inferred from the CYCLOPS ordering were in good accord with known physiology. Cycling was observed in pathways related to inflammation, hormone responsiveness, and DNA repair.

Both co-expression analysis and experimental measures from patient derived organoids suggested continued, albeit reduced, core clock rhythms in Luminal-A breast cancer samples. Application of CYCLOPS to Luminal-A data revealed disrupted rhythms with some output pathways gaining and others losing rhythmicity. Specifically, epithelial mesenchymal transition (EMT), a pathway critical to metastasis, demonstrated increased cycling. Among Luminal-A samples, there was marked variability in CYCLOPS magnitude, a composite measure of global circadian rhythm strength. When compared to tumors with lower rhythm strength, tumors with higher circadian magnitude demonstrated increased cycling of EMT pathway genes and a higher risk of metastasis (Relative Risk 1.8). Experiments with three-dimensional Luminal-A breast cancer cultures showed

that circadian disruption following the knockdown of the core clock gene ARNTL resulted in increased cell division but reduced matrix invasion and cellular spread.

These findings demonstrate the clinical importance of determining subtype specific circadian rhythms. We find rhythms in normal breast tissue that can guide chronotherapy to minimize local toxicity. Strikingly, high magnitude molecular rhythms in Luminal-A tumors may predict distant metastasis while, in-vitro, circadian disruption reduces tumor spread and migration.

9. Using stochastic processes to characterize the metastatic progression of breast cancer

Brianna Han, Carlos Mendonça, Andrada Nicolae, Jane Zhang

Columbia University

Due to the biochemical, physiological, and genetic complexity of cancer metastasis, a deterministic model which accurately predicts metastatic progression of breast cancer is not currently feasible. Therefore, we utilized a probabilistic approach, modeling the presence of cancer at sites within the body as states in a Markov chain. We used a transition matrix to represent the probabilities of metastasis from one site to another. The algorithm used to obtain such a transition matrix is inspired by an iterative adjustment method used in the analysis of lung cancer metastasis. Additionally, we sought to improve current models by further developing Monte Carlo simulations to more adequately account for a patient's metastatic history. After running 1000 simulations, we recorded the average time that each site was first reached. We were able to deduce other clinically relevant information from our simulations, such as the most likely metastatic pathways given metastases at specific sites.

10. Meta-analysis of the human gut microbiome

Dong-Min Jin, James T. Morton, Richard Bonneau

New York University

Advances in high-throughput sequencing have transformed our capacity to investigate the microbiome and its associations with diverse conditions. Shotgun metagenomic sequencing provides a better taxonomic resolution and genomic information compared to gene amplicon sequencing. The human gut microbiota includes trillions of microbes living in the gut. Studies have demonstrated associations between the human gut microbiome and multiple diseases, ranging from gastrointestinal disorders to neurodegenerative diseases. Getting to know the human gut microbiome is important for us to understand human biology and diseases. In this study, we processed datasets investigating different diseases with consistent methods, and by comparing the overlap of microbes more abundant or less abundant in different diseases pairs, we'll be able to figure out the microbes and functional similarities between diseases.

11. Traversome: A maximum likelihood framework for variable genome resolution and frequency estimation

Jian-Jun Jin, Deren A. R. Eaton

Columbia University

Genomic variants, including single-nucleotide variants (SNVs) and structural variants (SVs), are abundant in bulk sequencing of a multicellular individual or a population of unicellular or acellular individuals. The primary reported product of a de novo genome assembly is commonly a FASTA-formatted file containing a single haploid representation of each scaffold or chromosome sequence, which loses the information of alternative variants and potentially hampers the contiguity of the assembly. The utilization of assembly graphs for quantified resolution of genomic variants is underappreciated in available tools. Here we developed a new method for inferring one or more genome resolutions of an assembly graph, and the relative frequencies of each genome structural conformation, based on long-read mapping to an existing genome assembly graph. We implemented it in Traversome, which can generate accurate assemblies and genome structure frequency estimations according to our tests on both simulated and real datasets. Traversome is open source (GPLv3) and available at <https://github.com/Kinggerm/Traversome>.

12. Single cell master-regulator inference pipeline identifies the core regulators of liver fibrosis

Ajay Nair, Robert Schwabe, Andrea Califano, Ramon Bataller Alberola, Aveline Filliol, Yoshinobu Saito, Silvia Affo, Jorge Matias Caviglia, Oscar M. Rodriguez Fiallos

Columbia University

Background: Liver fibrosis affects 3.6-13% of the global population and shows an increasing trend. Fibrosis usually can progress to cirrhosis (excessive scarring affecting liver function), loss of liver function, and tumor development. Liver fibrosis is the excessive scarring of liver during chronic injury. It can be caused by fatty liver, chronic hepatic viral infection, excessive alcohol consumption, or cholestasis (impaired bile flow). Hepatic stellate cells (HepSC) are the main source of fibroblasts in the liver and activated HepSC represent the primary contributor of liver scarring. The key regulators of HepSC activation are still unknown and there are no currently approved drugs for liver fibrosis.

Hypothesis: Identifying the master-regulators of HepSC activation can help to understand the biology of liver fibrosis and identify effective drugs.

Research Design: We generated large single-nucleus/single-cell RNA-sequencing (sn/scRNA-seq) datasets of mouse (normal, NASH, CC14, and BDL) and human (normal, NASH, alcoholic hepatitis, and alcoholic cirrhosis) liver fibrosis. We also developed a single-cell regulator activity inference pipeline (scRegacti) to predict master-regulators of disease states from sn/scRNA-seq data. scRegacti has single-cell specific pre- and post-processing steps and analysis parameters for master-regulator inference. The sn/scRNA-seq datasets are noisy and have drop-outs (most genes have zero counts) making it impossible to directly infer master-regulators. Therefore, we developed procedures to de-noise and filter the data using statistical and biological principles. ARACNe and VIPER are well-known tools for regulatory-network and master-regulator inference from bulk RNA-seq data. Here, we optimized both the tools to work with sn/scRNA-seq data for identification of master-regulators of disease-states and their visualization. This pipeline integrates easily with the popular single-cell analysis package Seurat.

Results: scRegacti inferred master-regulators from both human and mouse datasets. A set of regulators was seen to be conserved across the different datasets, thus representing a core set of regulators of liver fibrosis. This core set of regulators contained many well-known and novel regulators. Knockdown of these core regulators showed reduction in fibrosis. Further, drug screening experiments identified drugs that reduce fibrosis and drugs that revert an activated HepSC to a quiescent state. These drugs also showed a reduction and reversion of core regulator activity, respectively.

Conclusions: Liver fibrosis has high prevalence but no approved drugs. Activated HepSC are the primary cause of liver fibrosis. Using the large sn/scRNA-seq datasets and scRegacti we inferred core regulators of HepSC activation. We also identified drugs that reduce or revert HepSC activation, a central event in the development of liver fibrosis.

13. Predicting compound toxicity by leveraging large-scale rat toxicogenomic data

Heeju Noh, Nicholas Tatonetti, Andrea Califano

Columbia University

Predicting potential toxicity of small molecules, early in the development process, is still an elusive challenge, resulting in a significant number of compound failures in phase 1 clinical trials. In recent years, several computational approaches for compound toxicity prediction have been proposed. Yet, most of them have focused on cellular rather than organ or whole-organism effects. In this study, we present a network- and machine learning-based framework for identifying compounds that may be toxic to major organs, including liver, kidney, and heart. For this purpose, we leveraged transcriptomic data from rodents exposed to a variety of compounds, which were obtained from large toxicogenomics projects like DrugMatrix, SEquencing Quality Control (SEQC), and TG-GATE. Specifically, we used the activity of transcription factors (TFs) - measured based on the enrichment of their transcriptional targets in differentially expressed genes using the VIPER algorithm - as an accurate, highly multiplexed reporter of compound toxicity. Our analysis identified sets of drugs causing distinct toxicities in the liver, kidney, and heart. We trained our models using data from DrugMatrix and tested their accuracy using independent data from SEQC and TG-GATE. Our predictor was highly accurate and reproducible with AUROC scores ranging from 0.87 to 0.90 for liver toxicity and 0.82 to 0.85 for kidney toxicity. Critically, using TF activity significantly outperformed equivalent predictors based on gene expression. The toxicity signatures identified by our algorithm provided critical insight into the specific mechanisms underlying compound toxicity in each organ, including genotoxicity and cytotoxicity in the liver, calcium balance and epigenetic changes in the heart, and cell death and hyaline droplet formation in the kidney. We propose that this novel methodology can critically benefit drug development and research by reducing the experimental burden required to establish drug safety.

14. Development of trajectory analysis of a transcript's evolvability under APOBEC3 activity

Joon-Hyun Song, Jinyu Li, Thomas MacCarthy, Mehdi Damaghi

Stony Brook University

Recent studies have shown that the mutational signature by APOBEC activity is ubiquitously observed in various cancers including ovarian cancer. Three of the APOBEC3 sub-family, APOBEC3A, APOBEC3B, and APOBEC3G have been shown to play a role in cancer. APOBEC3A and B target a cytidine in TC motifs and result in C-to-T, C-to-G, or rarely C-to-A mutation. Similarly, mutations could arise by APOBEC3G which targets the last cytidine in CCC motifs. These mutations may be the source of the increased heterogeneity of cells in the early stage of cancer formation by targeting mutations to driver genes such as BRCA1/2, furthermore in later stages leading to resistance to treatments such as PARP inhibitors. Here, we introduce a novel technique to analyze transcripts' evolvability under APOBEC3 activity. The evolvability of a biological system is the potential to create phenotypic variation advantageous for the context the biological system finds itself. Transcriptional evolvability in cancer is, therefore, an ability to produce variants, transcripts with mutation, that might lead to higher fitness. For instance, the original transcript may acquire several synonymous mutations before a mutation that is non-synonymous and brings functional advantages. Our technique generates a series of sequences, a trajectory, by mutating APOBEC3 motifs randomly one by one until a non-synonymous mutation occurs. Then, we compute APOBEC3 motifs representation statistics of all sequences and measure the reference sequences evolvability by analyzing the trajectory pattern, the coverage area, the length of trajectories, and the general motile trend. We used the technique to compare the evolvability of BRCA1/2 and PARP reference sequences and mutated sequences from ovarian cell lines and patient samples adapted to PARP inhibitor.

15. Intra-tumour heterogeneity during evolutionary therapies

Alexander Stein, Katerina Stankova, Benjamin Werner

Barts Cancer Institute, Queen Mary University of London

Intra-tumour heterogeneity is a natural outcome of the cancer evolutionary process. It is a leading contributor to treatment resistance and disease progression. In the past decade, novel personalized therapy schedules have been developed that explicitly target cancer as a speciation process to improve patient outcomes. These evolutionary therapies are based on mathematical models describing the dynamics of typically two or three phenotypes of interest. However, in reality, the resistant subpopulation of a tumour consists of many subclones. A high genetic diversity of subclones is accompanied by a higher risk of multi-drug resistance and the potential of additional driver mutations. In my poster, I introduce a stochastic branching process that includes both interactions between cancer cells and mutation events. Based on this model, I present the analysis of the clonal diversity at the time of diagnosis and during treatment. We find that the clonal diversity remains at a relatively high level after the application of a single treatment. For evolutionary therapies, we quantify the trade-off between (1) keeping the tumour burden high such that the resistant subpopulation grows slower and (2) keeping the tumour burden low such that the chance of acquiring additional resistance or driver mutations is low. I suggest adaptations of the existing evolutionary therapies leading to lower genetic diversity, and, therefore, also lower risk of multi-drug resistance.

16. Integrating mathematical modelling and wet-lab experiments to examine the scope for adaptive treatment scheduling of PARP inhibitors in ovarian cancer

Maximilian Strobl, Mehdi Damaghi, Alexandra Martin, Samantha Byrne, Mark Robertson-Tessi, Robert Gatenby, Robert Wenham, Philip Maini, Alexander Anderson

Cleveland Clinic

PARP inhibitors (PARPis) represent a great advancement in the treatment of ovarian cancer, yet these drugs often fail after a few months due to emerging drug resistance. A recent clinical trial in prostate cancer showed that evolutionary-inspired, adaptive drug scheduling significantly delayed time to progression. This approach adaptively skipped treatment to maintain a pool of drug-sensitive cells that suppressed resistant cells through competition. Here, we present results from a combined modelling and experimental study in which we investigated whether adaptive therapy could delay resistance to the PARPi olaparib in ovarian cancer.

We performed a series of in vitro experiments in which we used Incucyte Zoom time-lapse microscopy to characterise the cell population dynamics under different PARPi schedules. Leveraging these data we developed an ordinary differential equation mathematical model of treatment response, and used this model to test different plausible adaptive treatment schedules. Our model can accurately predict the in vitro treatment dynamics, even to new schedules, and suggests that treatment modifications need to be carefully timed, or one risks losing control over tumour growth, even in the absence of any resistance. This is because multiple rounds of cell division are required for cells to acquire sufficient DNA damage to induce apoptosis. As a result, adaptive therapy algorithms that modulate treatment but never completely withdraw it are predicted to perform better in this setting than strategies based on treatment interruptions. Subsequent experiments confirm this prediction in vivo. Overall, this study contributes to a better understanding of the impact of scheduling on treatment outcome for PARPis, and showcases some of challenges involved in developing adaptive therapies for new treatment settings.

17. A Unified Modular Framework to Incorporate Structural Dependency in Spatial Omics Data

Jiayu Su, Jean-Baptiste Reynier, Xi Fu, David A Knowles, Raul Rabadan

Columbia University

Spatial omics technologies, such as spatial transcriptomics, allow the identification of spatially organized biological processes, while presenting computational challenges for existing analysis approaches that ignore spatial dependencies. Here we introduce Smoother, a unified and modular framework that integrates positional information into non-spatial models via spatial priors and losses. In simulated and real datasets, we show that Smoother enables spatially aware data imputation, cell-type deconvolution, and dimensionality reduction with high accuracy. (See our biorxiv preprint: <https://www.biorxiv.org/content/10.1101/2022.10.25.513785v1.full>).

18. Probabilistic modeling of single-cell DNA replication dynamics in genomically unstable cancer cells

Adam C. Weiner, Marc J. Williams, Hongyu Shi, Andrew McPherson, Sohrab P. Shah

Memorial Sloan Kettering Center

Emerging single-cell whole genome sequencing (scWGS) assays hold potential for studying replication dynamics of cancer cells; however, computational methods for identifying S-phase cells and inferring their single-cell replication timing (scRT) profiles remain immature for samples with heterogeneous copy number profiles. Here we report a new method, PERT, which jointly infers replication and copy number states of S-phase cells. This method enabled us to analyze the replication dynamics of >10,000 S-phase single-cell whole genomes across various genetically engineered cell lines, high grade serous ovarian cancers, and triple negative breast cancers. We show that this method enables robust cell cycle phase predictions, quantifies cell-to-cell replication timing variability, and approximates relative proliferation rates between tumor subclones. Our results illuminate how aberrant DNA replication processes can both drive and result from the evolution of structural variations in human tumors.

19. Detection and Reconstruction of Validated Extrachromosomal Circular DNA from Single Cell DNA Sequencing Data

Benjamin Wesley, Alvin Ng, Pankaj Pradeep, Rebecca Fitzgerald, Karol Nowicki-Osuch

Columbia University

Esophageal Adenocarcinoma (EAC) incidence is increasing, while 5-year survival rates stubbornly remain below 15%. A large contribution to this low survival rate is the relatively late diagnosis, with most EAC presenting at an advanced stage with numerous large scale genomic copy number alterations (CNA) and acquisition of recently identified, frequent (>50% of cases) extrachromosomal DNAs (ecDNA). Direct Library Preparation (DLP+) is a whole genome amplification (WGA)-free, single cell DNA sequencing method that provides even sequencing coverage across thousands of cells. DLP+ allows for an unbiased representation of the global CNA states throughout the genome. Here, we employed DLP+ to track CNA states in EAC organoids.

The DLP+-based analysis of patient derived EAC organoids (CAM277 and CAM401) demonstrated dynamic changes in CNA profiles over time. In both organoids, we observed selection of individual clones during their long-term maintenance. Next, we took advantage of traditional whole genome sequencing (WGS) and Oxford Nanopore (ONT) Long Read Sequencing to identify and reconstruct an ecDNA composed of fragments of chromosome 12. We were not only able to detect and quantify per-cell ecDNA states using our short-read, low coverage DLP+ data, but we also reconstructed the ecDNA structure independently of the ONT analysis. When combined with genome-wide CNA states, we showed that ecDNA are inherited independently of clonal composition of organoids. Taken together, this data shows that, for the first time, ecDNA states can directly tracked in individual cells and their effects on clonal dynamics can be investigated in cancer models.