****

**Hard link to access the presentation slides**

[**https://drive.google.com/open?id=1\_yWGoP1488y\_wyW0NFIAn4zUYxwfd91u**](https://drive.google.com/open?id=1_yWGoP1488y_wyW0NFIAn4zUYxwfd91u)

**Altrock, Philip,** [**philipp.altrock@gmail.com**](mailto:philipp.altrock@gmail.com)

***Lecture***

***Breakouts***

***Projects***

**Bennett, Matthew,** [**bennett@rice.edu**](mailto:bennett@rice.edu)

***Lecture***

*Dynamics of synthetic gene circuits: from cells to consortia to organisms*

One challenge of synthetic biology is the creation of cooperative microbial systems that exhibit emergent multicellular behaviors. Such systems use cellular signaling mechanisms to regulate gene expression between cells. However, it is difficult to coordinate gene expression in spatially extended microbial systems because the range of signaling molecules is limited by diffusion. Here, we show that spatiotemporal coordination of gene expression can be achieved even when the spatial extent of the system is much greater than the diffusion distance of the signaling molecules. To do this, we used a combination of experimental perturbations and computational simulations to examine the dynamics of several two-strain synthetic microbial consortia. In small colonies, these consortia generate coherent transcriptional oscillations for a variety of regulatory topologies. In large colonies, we find that temporally coordinated oscillations across the population are possible, but rely on the presence of an intrinsic positive feedback loop that amplifies and propagates intercellular signals. These results demonstrate that spatially extended synthetic multi-cellular systems can be engineered to exhibit coordinated gene expression using only transient, short-range coupling between the constituent cells.

***Breakouts***

***Projects***

**Braun, Rosemary,** [**rbraun@northwestern.edu**](mailto:rbraun@northwestern.edu)

***Lecture***

*Lively Networks*

Many systems -- including living cells -- exhibit collective behaviors that emerge from complex networks of many interacting processes.  What can the "wiring diagram" of those interactions tell us about the dynamics of the system, and can we deduce the underlying network from the collective dynamics?  In this talk, I will discuss what we can learn about the dynamics of interacting systems from the topology of the underlying network of interactions.  I will introduce the formalisms of spectral graph theory and network filtration, and illustrate how these approaches can help us model how living systems respond and adapt to perturbations.

***Breakouts***

*Seeing the forest and the trees: multi-scale approaches for analyzing cancer omics data.*

Advances in high-throughput "\*omic" assays now make it possible to the molecular state of a sample in genome-wide detail, providing unprecedented opportunity to investigate disease mechanisms by simultaneously profiling thousands molecular markers per sample.  To date, however, most analyses of \*omic data consider each marker independently and treat regulatory pathways as a "sum of their parts." By neglecting the network of interactions, such approaches can miss crucial multi--gene effects associated with disease.  This talk will present some recent techniques to incorporate pathway information into the analysis of high--dimensional \*omic data.  By analyzing data at the systems level, the methods enable us to integrate disparate types of \*omic data, make inferences about disease mechanisms, and distinguish sets of cumulatively deleterious alterations from those that compensate one-another to preserve the overall function of a pathway.  We will show how these analyses can overcome the high variability of \*omics data to yield results that are more reproducible across studies, and demonstrate how these methods can be used to identify novel therapeutic and diagnostic targets.

***Projects***

**Ding, Khanh,** [**knd2127@columbia.edu**](mailto:knd2127@columbia.edu)

***Lecture***

***Breakouts***

***Projects***

**Faeder, Jim,** [**faeder@pitt.edu**](mailto:faeder@pitt.edu)

***Lecture***

*Introduction to modeling signal transduction using rules*

This lecture will describe the challenges faced in developing detailed models of biochemical networks, which encompass large numbers of interacting components. Although simpler coarse-grained models are often useful for gaining insight into biological mechanisms, such detailed models are necessary to understand how molecular components work in the network context and essential to developing the ability to manipulate such networks for practical benefits. The rule-based modeling (RBM) approach, in which biological molecules can be represented as structured objects whose interactions are governed by rules that describe their biochemical interactions, is the basis for addressing multiple scaling issues that arise in the development of large scale models. Currently available software tools for RBM, such as BioNetGen, Kappa, and Simmune, enable the specification and simulation of large scale models, and these tools are in widespread use by the modeling community. I will review some of the developments that gave rise to those capabilities, and then I will present several applications of the approach which yield new insights into the mechanisms that govern cell fate decisions.

***Breakouts***

*Introduction to modeling reaction-diffusion systems with MCell/CellBlender*

In this tutorial participants will learn how to construct reaction-diffusion models of cell systems using CellBlender, which is a powerful graphical interface to the popular MCell simulator. MCell performs particle-based stochastic simulations of biochemical reactions with arbitrarily complex boundaries. Participants will learn how to construct simple geometries and reaction networks and also how to import more complex geometries derived from images and more complex reaction networks developed in BioNetGen. The tutorial will also introduce participants to a number of relatively simple models where spatial effects produce large deviations from the well-mixed limit and even emergent phenomena that are otherwise not observed.

*Introduction to biochemical modeling with BioNetGen*

In this tutorial participant will learn how to construct, simulate, and analyze detailed models of biochemical systems using RuleBender, which provides a unified interface to the popular rule-based modeling software BioNetGen. Participants will learn the structure of a model input file, and how to run simulations using a variety of simulation engines, including ODE’s, SSA, and network-free stochastic simulation. Participants will work with a range of models that cover commonly-modeled features of signaling networks

***Projects***

**Igoshin, Oleg, igoshin@rice.edu**

***Lecture***

*Design principles of cellular differentiation control: lessons from Bacillus subtilis*

Successful execution of differentiation programs requires cells to assess multitudes of internal and external cues and respond with appropriate gene expression programs. In this talk, I will discuss results illustrating how *Bacillus subtilis* sporulation network deals with these tasks focusing on the lessons generalizable to other systems. With feedforward loops controlling both production and activation of downstream transcriptional regulators, cells achieve ultrasensitive threshold-like responses. The arrangement of sporulation network genes on the chromosome and transcriptional feedback loops allow coordination of sporulation decision with DNA-replication. Furthermore, to assess the starvation conditions without sensing specific metabolites, cells respond to changes in their growth rates with increased activity of sporulation master regulator. These design features of the sporulation network enable cells to robustly decide between vegetative growth and sporulation.

***Breakouts***

The speed at which catalytic enzymes function in complex biological systems and living cells is determined by the analysis of various reaction mechanisms and corresponding kinetic rates. The Michaelis-Menten (MM) mechanism is the most well-known example of a chemical kinetic network and, in principle, it is sufficient to determine the kinetic properties of any enzymatic process. The MM mechanism consists of only two steps: the free enzyme binds reversibly to the substrate to form an enzyme-substrate complex, and this complex subsequently dissociates to form the the product while also regenerating the free enzyme at the same time. Although the MM mechanism provides a very useful description of the enzyme kinetics, a more sophisticated and complex mechanism developed independently by Hopfield and Ninio, namely, kinetic proofreading (KPR)4,5 provides clearer and more accurate information about the dynamic properties of the enzymatic process. For example, the KPR mechanism explains the remarkably high substrate selectivity (i.e., low error rates ) observed experimentally for fundamental biological processes such as DNA replication by the DNA polymerase (), mRNA transcription by the RNA polymerase (), and protein translation by the ribosome (). The KPR mechanism enables the enzymes to choose the correct substrate from a pool of chemically similar substrates with high fidelity. The enhanced accuracy is due to the presence of an additional “proofreading” step such that the enzyme can correct the mistakes it made by removing incorrect substrates, thereby resetting the enzymatic system back to its original state without forming the product. However, the higher fidelity comes at the cost of reduced enzymatic speed and increased energy expenditure such that the ratio of the number of ATP molecules hydrolyzed to the number of product molecules formed is > 1. In this hands-on session of the q-Bio summer school course, I will guide the students through a lecture and tutorial designed to teach them how to use the Mathematica software to solve a set of backward master equations by Laplace transformation that govern the temporal evolution and the dynamics of biological systems. Specifically, I will instruct the students on how to apply this framework to the MM and KPR mechanisms (in this order) for a DNA polymerase enzyme from bacteriophage T7 in the presence of two substrates denoted as right R and wrong W. The main objective of my presentation and assignment is to introduce the students to the backward master equation formalism such that they can use the underlying mathematics to determine analytic expressions for two important first-passage properties of the T7 DNA polymerase, namely, the error rate η and the mean first-passage time (MFPT) τ . Moreover, using pre-determined kinetic parameters for the T7 DNA polymerase (i.e., experimental rate constants) 7–9 the students will determine specific values for the error rate η (the ratio of the splitting probabilities ΠR and ΠW to make the right R and wrong W products, respectively, η = ΠW /ΠR) and the conditional MFPT τR to make the right R product before making the wrong W one. The students will then compare their results on η and τR 1 obtained for the MM and KPR mechanisms to verify which mechanism gives the lower error rate and the slightly longer MFPT (slower speed).

***Projects***

**Iyer-Biswas, Srividya,** [**iyerbiswas@purdue.edu**](mailto:iyerbiswas@purdue.edu)

***Lecture***

***Breakouts***

***Projects***

**Jilkine, Aleksandra,** [**ajilkine@nd.edu**](mailto:ajilkine@nd.edu)

***Lecture***

*Control of Cell Fraction and Population Recovery during Tissue Regeneration in Stem Cell Lineages*

Multicellular tissues are continually turning over, and homeostasis is maintained through regulated proliferation and differentiation of stem cells and progenitors. Following tissue injury, a dramatic increase in cell proliferation is commonly observed, resulting in rapid restoration of tissue size. This regulation is thought to occur via multiple feedback loops acting on cell self-renewal or differentiation. Prior modeling studies of the cell lineage system have suggested that loss of homeostasis and initiation of tumorigenesis can be contributed to the loss of control of these processes, and the rate of symmetric versus asymmetric division of the stem cells may also be altered.

Here, we compare three variants of hierarchical stem cell lineage tissue models with different combinations of negative feedbacks and use sensitivity analysis to examine what are the possible strategies for the cells to achieve certain performance objectives. Our results suggest that multiple negative feedback loops must be present in the stem cell lineage the fractions of stem cells to differentiated cells in the total population as robust as possible to variations in cell division parameters, and to minimize the time for tissue recovery in a non-oscillatory manner. When one of these negative feedback loops on stem cell division been knocked out, most of the stem cell lineage population will be in the form of stem cells, suggestive of "precancerous" tissue. Furthermore, modeling suggests that positive feedback loops in stem cell homeostasis may also be required. We contrast and compare the differences between deterministic and stochastic versions of the models.

***Breakouts***

***Projects***

**Johnson, Maggie,** [**margaret.johnson@jhu.edu**](mailto:margaret.johnson@jhu.edu)

***Lecture***

*Single-particle reaction-diffusion applied to self-assembly in the cell*

Computational modeling excels at predicting the behavior of complex systems with many interacting components. By solving physics-based models, they help us to interpret, distill, and predict shared and simplified phenomena out of challenging experiments. However, current computational tools for studying self-assembly dynamics cannot manage the multi-component, non-equilibrium, minutes or longer processes that occur within cells. Molecular modeling approaches (e.g. Molecular Dynamics (MD)) provide detailed structures of assembly pathway but do not couple to most non-equilibrium events (such as phosphorylation). Conversely, reaction-diffusion (RD) approaches readily capture non-equilibrium events, but lack molecular structure. We will introduce here the single-particle reaction-diffusion approach, and show how integration of molecular details can enable unique and powerful studies of cell-scale self-assembly

***Breakouts***

*Computational modeling of clathrin-coat assembly with NERDSS*

In diverse cellular pathways including clathrin-mediated endocytosis and viral bud formation, cytosolic proteins must self-assemble at the plasma membrane. To understand the mechanisms whereby assembly is triggered and how perturbations can lead to dysfunction requires dynamics of not just assembly components, but their coupling to active, force-responsive, and ATP-consuming structures in cells. Current computational tools for studying self-assembly dynamics are not feasible for simulating cellular dynamics due to the slow time-scales and the dependence on energy-consuming events such as phosphorylation. We have developed reaction-diffusion algorithms and software that enable detailed computer simulations of nonequilibrium self-assembly over long time-scales. In this session, we will learn how to design models for reaction-diffusion simulations. We will go through how to set up models in the NERDSS software to run simulations over a range of parameters, in this case applied to understanding self-assembly of clathrin coats at the plasma membrane. The software is generalizable and applicable to a wide range of problems in protein self-assembly.

***Projects***

**Josic, Kresimir**

***Lecture***

How to infer parameters from experimental observations of biochemical reaction processes is a key problem in systems biology. However, biochemical data are intrinsically stochastic, and observations may be noisy. I will describe different Bayesian approaches that can be used to infer rate parameters in such situations. While conceptually straightforward, the main difficulty in implementing such methods is the need for the numerical computation of high-dimensional integrals to find the estimates of interest. I will describe how Markov Chain Monte Carlo (MCMC) methods can be used in this situation. I will also briefly describe more recent Hamiltonian Monte Carlo techniques that are rapidly gaining in popularity.

Experimental data is frequently obtained using discrete-time sampling systems. This means that we can at best observe the state of the system at discrete intervals, and the exact timing of reactions is unknown. This provides a challenge to a direct implementation of classical methods. I will discuss algorithms that can be applied in such situations.

I will provide Python code to both generate synthetic time series of some simple biochemical reaction processes, as well as implement the different inference techniques described in the lecture.

***Breakouts***

***Projects***

**Kimmel, Marek,** [**kimmel@rice.edu**](mailto:kimmel@rice.edu)

***Lecture***

*Neutral Cancer Evolution with Selective Sweep(s) Analysis of Site Frequency Spectra of Tumor Genomes*

The aim of this talk is to present mathematical models that can be used to extract information regarding cancer evolution from the genome sequences of human cancers. This includes the history of growth and mutation and effects such as genetic drift and selective sweeps. Our aim is to point out how mathematical and statistical modeling may help in elucidating problems that frequently have been tackled using intuitive approaches.

Biological cells undergo mutations as they proliferate and such mutations can be neutral, advantageous, or deleterious. The rate of mutation depends on the environment and DNA repair mechanisms. Progress in genome sequencing has allowed cataloguing not only reference genomes of many biological species but also of variants characteristic of human, animal and plant diseases. In particular, initiatives such as the The Cancer Genome Atlas program and the International Cancer Genome Consortium have allowed determination of sets of genomic variants characteristic of many human tumors, with several hundred specimens of each, thus detailing their common mutational features.

One difficulty that arises is that most of the genome sequences available result from so-called bulk sequencing, in which DNA from a sample of cells obtained from the tumor and its environment is cut into fragments, amplified and sequenced, resulting in reads that are aligned with the human reference genome. The resulting genome sequence includes variants that are characteristic of different but not easily identifiable sub-populations of tumor cells. Short of sequencing a representative subset of genomes of individual cells, this difficulty cannot at present be radically improved. Nevertheless, bulk-sequencing data constitute most of the material currently available and it seems important to try to understand the message they carry regarding tumor origin and natural course, perhaps distorted by treatment. This might be called ``the genomic archaeology of tumors".

***Breakouts***

I will discuss 2 modeling studies carried out by my group, which illustrate our interest in evolution and population genetics of cell populations. Corresponding publications are:

Wojdyla T, Mehta H, Glaubach T, Bertolusso R, Iwanaszko M, Braun R, Corey SJ, Kimmel M. Mutation, drift and selection in single-driver hematologic malignancy: Example of secondary myelodysplastic syndrome following treatment of inherited neutropenia. PLoS computational biology. 2019 Jan 7;15(1):e1006664.

Mura M, Feillet C, Bertolusso R, Delaunay F, Kimmel M. Mathematical modelling reveals unexpected inheritance and variability patterns of cell cycle parameters in mammalian cells. PLoS computational biology. 2019 Jun 3;15(6):e1007054.

***Projects***

*Mathematical model of clonal hematopoiesis* As humans age, their hemopoietic stem cells slow down, in the sense that interdivision times become progressively longer. This is probably compensated by acceleration of proliferation of intermediate maturation stages (committed cells). Data for this are available. The project will involve setting up a ODE system modeling overall proliferation kinetics and coupling it with a stochastic model of clonality, to investigate possible scenarios of evolution of clonal hematopoiesis and its transition to malignancy.

*ABC estimation of evolution of mutant receptor based on Moran Model* In the paper by Wojdyla et al. (2019) it has been argued that transition to a premalignant disorder MDS in neutropenia patients treated with G-CSF can be explained using a simple Moran model with recurrent mutation and selection. Recently, new data were published, which may allow a direct verification by fitting a Moran Model and finding if resulting estimates of parameters make biological sense.

*Population genetics of proliferating cells.* Classical population genetics models are frequently used to describe evolution of cell populations. However, modeling studies show that there exist properties that are inherited by cells in a non-classical way, which may accelerate the action of genetic drift. These effects may be meaningful in the context of evolution of cancer but also proliferating normal cells, for example in bone marrow. Project involves mathematical analysis or a simulation study of a model based on bifurcating auto-regression, using literature data.

**King, Katherine, kyk@bcm.edu**

***Lecture***

*The biology of hematopoietic stem and progenitor cells and the process of primitive hematopoiesis*

We will discuss the process of primitive hematopoiesis, including the biology of hematopoietic stem and progenitor cells (HSPCs). We will review experimental methods to define HSPCs and their properties including quiescence, proliferation and differentiation rates. We will explore the current state of knowledge regarding how environmental stresses such as infection and inflammation perturb these properties. We will describe lineage tracing and bar-coding methodologies to study such perturbations. Finally, we will discuss how changes in HSPC populations contribute to clonal hematopoiesis and the potential contributions of infection and inflammation to leukemic transformation. This will include a basic discussion of leukemia, leukemic stem cells, and genetic basis of cancer.

***Breakouts***

*Hematopoietic stem cells and their quantification*

*Working session on environmental drivers of clonal hematopoiesis*

***Projects***

**Massey, Susan,** [**Massey.Susan@mayo.edu**](mailto:Massey.Susan@mayo.edu)

***Lecture***

*Multi-scale modeling of paracrine PDGF-driven glioma growth and invasion*

The most common primary brain tumor in adults, glioblastoma claims thousands of lives each year. It is a notoriously aggressive and invasive disease, and despite efforts to improve survival rates, the standard of care has remained unchanged for more than a decade. Platelet¬–derived growth factor (PDGF) is often over–expressed in gliomas, where it can drive tumor growth via autocrine and paracrine stimulation of PDGF receptor (PDGFR)–expressing glioma cells, as well as paracrine recruitment of non–neoplastic oligodendroglial progenitor cells (OPCs), which also express PDGFRs. Constructing multi-scale mechanistic mathematical models allows us to better examine the effects of paracrine PDGF signaling on glioma growth. Model simulations show that increased PDGF signaling can increase growth rates and alter the distribution of glioma cells infiltrating adjacent normal brain tissue. While the use of PDGF inhibitors has remained largely unsuccessful at improving patient outcomes in glioblastoma, this may be due to inadequate targeting of these agents to the best candidates. By incorporating different treatment simulations in our model, we show that PDGF inhibition results in decreased OPC recruitment, which leads to slower growing, but more diffusely infiltrating tumors.

***Breakouts***

*Model parameter sensitivity analysis using Latin hypercube sampling and partial rank correlation coefficients*

Description: This will include a brief lecture providing a motivation and application for parameter sensitivity analysis in model development, an overview of the LHS and PRCC procedure, and an overview of some code files I have written (matlab and python) that will allow them to get experience using the technique. Hopefully this will be something accessible that the students can use in their projects if appropriate.

***Projects***

1. Sensitivity analysis of a model for tumor growth incorporating hypoxia, angiogenesis, and edema
2. Analysis of Ivy GAP data (<http://glioblastoma.alleninstitute.org/> )

**Mitra, Eshan,** [**emitra@lanl.gov**](mailto:emitra@lanl.gov)

***Lecture***

*Parameter Estimation and Uncertainty Quantification for Systems Biology Models*Mathematical models of cell signaling processes often contain unknown parameters. Parameter estimation and uncertainty quantification are often required before making reliable predictions. I will discuss gradient-based parameter estimation methods (first and second order methods, and methods for calculation of gradients) as well as metaheuristic parameter estimation methods. The discussion will include recent approaches of using these methods in combination with non-numerical, qualitative data. On the topic of uncertainty quantification, I will cover profile likelihood, bootstrapping, and Bayesian uncertainty quantification by MCMC. Finally, I will discuss  current software tools available that implement these methods.

***Breakouts***

*Parameter Estimation and Uncertainty Quantification in PyBNF.*PyBNF is a recently developed software tool for parameter estimation and uncertainty quantification of models defined in BNGL or SBML. We will work through configuring and running some example fitting and uncertainty quantification jobs in PyBNF.

Before the session, students should install PyBNF from <https://github.com/lanl/PyBNF> as well as BioNetGen from [http://bionetgen.org](http://bionetgen.org/).

***Projects***

**Munsky, Brian,** [**Brian.Munsky@colostate.edu**](mailto:Brian.Munsky@colostate.edu)

***Lecture***

***Breakouts***

***Projects***

**Navin, Nicolas,** [**NNavin@mdanderson.org**](mailto:NNavin@mdanderson.org)

***Lecture***

***Breakouts***

***Projects***

**Palmer, Adam**

***Lecture***

*Understanding the origin of benefit of combination cancer therapies*Developing optimal drug combinations is one of the central challenges of cancer treatment research: drug combinations are used to treat most types of cancer, and are almost exclusively responsible for cures of advanced cancers. However, historically successful combination therapies were developed empirically, and the mechanistic basis for their efficacy has been largely speculative. I will present studies of clinically successful combination therapies that identify the control of between-tumor and within-tumor heterogeneity by independently active drugs as critical contributors to the efficacy of drug combinations in human patients. Mathematical descriptions of heterogeneity in cellular or patient populations, and experimental measurements of how drug combinations address heterogeneity, lead to accurate predictions of clinical trial results across many types of cancer and types of therapies, including curative chemotherapy regimens and recent immunotherapies. These results have broad significance for the treatment of cancers, for the interpretation of clinical trials, and point to new opportunities to use combination therapies with greater precision.

***Breakouts***

*Analysis and simulation of tumor heterogeneity in clinical trials*

In this break-out session we will begin with classical models of tumor kinetics, and tumor response to therapy, and expand these models to include inter-tumor and intra-tumor heterogeneity, to investigate the patient-to-patient variability observed in clinical trials of cancer therapies. Finally, we will apply these concepts to the analysis of human clinical trial data for combinations of cancer therapies, to understand the mechanistic basis for the clinical efficacy of combination therapies.

***Projects***

*Tumor-specific and drug-specific simulation of clinical trials of cancer therapies*

There exist many approaches to model tumor growth and response to therapy, some of which we will learn in the break-out session. Rarely are such models expanded to consider the immense inter-patient heterogeneity in response to anti-cancer therapies. However, there exists vast cohorts of human clinical trials data against which such models could be calibrated. In these projects students can calibrate these models to clinical data for specific tumor types (which exhibit different kinetic parameters) and specific therapies (which have different effects on rate of proliferation and rate of death of cancer cells). These models can close the gap between generic mechanistic understanding of how cancers respond to treatment, and specific outcomes in human clinical trials of cancer therapies, which can support more rational design of treatment regimens and clinical trials thereof.

**Sumazin, Pavel,** [**Pavel.Sumazin@bcm.edu**](mailto:Pavel.Sumazin@bcm.edu)

***Lecture***

*Strategies for tumor phylogeny inference from bulk tumor profiles*

Cancers are composed of multiple types of malignant and non-malignant cells. Understanding their composition may inform on the interactions between cancer cells and their environment and identify molecular features of aggressive subclones, including subclones that may be resistant to therapy or form metastases.  Characterizing the composition of cancers can help infer their evolution, and may help predict patient outcomes and identify effective therapeutic strategies.

Single-cell RNA profiling by scRNA-Seq and single-cell protein expression profiling by CyTOF are fast becoming the methods of choice for estimating the expression of genes in individual cancer cells. Assays to profiles DNA alterations in single cells are also being developed. These single-cell profiling assays help identify cell-type-specific gene expression programs that can be used to aggregate and identify cell types, thus allowing researchers to estimate tumor composition, which, in turn, can be used to reverse engineer tumor evolution.

However, these assays have limitations. scRNA-Seq can only work on fresh or frozen biopsies, and, consequently, 95% of tumor biopsies cannot be profiled by single-cell RNA-Seq with today’s technology. single-cell RNA-Seq also remains expensive: over $5,000 per sample. Scientists only recently have begun exploring the potential applications of protein-expression profiling by CyTOF, but this costly technology produces estimates for fewer than 50 genes per sample and many cancer genes cannot be profiled because of various biological and technological limitations. Single-cell DNA sequencing remains technologically challenging as well, with many labs reporting low accuracy. We asked whether similar analysis could be accomplished by deconvolving DNA, RNA, and protein expression estimates from profiles of cell populations across multiple tumor regions by next-generation DNA sequencing, RNA-Seq, mass spectrometry, or other bulk-tumor profiling technologies. These technologies are relatively mature, can be used to profile most tumor samples with adequate size, and can be implemented at a fraction of the cost of single-cell technologies.

We will present the problem and discuss available methods developed by us and by other groups.

***Breakouts***

Manual inference of tumor-subclone specific profiles and phylogenies from estimates of tumor subclones profiles

***Projects***

* Generate simulated tumor profiles from cell line and single-cell expression profiles
* Propose methods to evaluate solutions for simulated data
* Implement strategies to resolve profiles and construct tumor phylogenies

**Tavaré, Simon, st3193@columbia.edu**

***Lecture***

*Computational cancer genomics*

The mathematical sciences have contributed substantially to our understanding of the way cancer evolves. Cancer is thought of as a disease of the genome, so my focus will be on mutations in DNA and what they tell us about tumour evolution. Following a brief introduction to the history of cancer, I will discuss “tumour heterogeneity”, the DNA sequence variation observed between tumours and within them, and how we might use this to learn about progression, treatment and relapse. Along the way I will illustrate some of the underlying mathematical and statistical tools that have helped in this endeavor.

This talk does not assume any familiarity with the cancer field.

***Breakouts***

*Some statistical problems in cancer genomics*

The starting point for this talk comes from population genetics: how should we estimate evolutionarily relevant parameters from DNA sequence data taken from samples of individuals? I will give a brief overview of what we learned, starting from the Ewens Sampling Formula and touching on Approximate Bayesian Computation as an inference method when likelihoods are intractable. To illustrate ABC, I will give an example concerning inference of the number of distinct DNA sequences in a sample, given only information about the frequency of point mutations in the samples. This example provides an introduction to inference from typical cancer sequencing data, in which individuals are replaced by cells and in which typically we do not know which mutations occur in which cells. I will give a brief overview of what cancer evolution is about, the sort of statistical and computational problems it poses, and where we are in addressing some of them.

*The combinatorics of spaghetti hoops*

Starting with n cooked spaghetti strands, tie randomly chosen ends together to produce a collection of spaghetti loops. What is the expected number of loops? What can be said about the distribution of the number of loops of length 1, 2, …? What is the behavior of the longest loops when n is large? What is the probability that all the loops have different lengths? Questions like this appear in many guises in many areas of mathematics, the connection being their relation to the Ewens Sampling Formula (ESF). I will describe a number of related examples, including prime factorization, random mappings and random permutations, illustrating the central role played by the ESF. I will also discuss methods for simulating decomposable combinatorial structures by exploiting another wonder of the ESF world, namely the Feller Coupling. Analysis of a children’s playground game shows that apparently small departures from the Feller model can open up a number of unsolved problems.

***Projects***

**Wang, Wenyi,** [**wangwenyi6@gmail.com**](mailto:wangwenyi6@gmail.com)

***Lecture***

*Mathematical deconvolution to understand tumor heterogeneity*

Tumor tissue samples are comprised of a mixture of cancerous and surrounding stromal cells. Understanding tumor heterogeneity is crucial to analyzing gene signatures associated with cancer prognosis and treatment decisions. Compared with the experimental approach of laser micro-dissection to isolate different tissue components, in silico dissection of mixed cell samples is faster and cheaper. Numerous computational approaches previously developed have their limitations to deconvolute heterogeneous tumor samples. In this talk, I will present our new models for transcriptomic deconvolution DeMixT, which accounts for the immune cell compartment explicitly and is able to address the challenging problem when the observed signals are assumed to come from a mixture of three cell compartments, infiltrated immune cells, tumor microenvironment and cancerous tissues. I will also present our models for genomics deconvolution, CliP/CSR.

***Breakouts***

*Multi-omic deconvolution analyses of heterogeneous tumor samples reveal cell-type specific molecular interactions underlying clinical outcomes*

Most tumors consist of a variable proportion of malignant and nonmalignant cells including epithelial cells, fibroblasts, and infiltrating immune cells, which confounds biomarker studies of response to treatment. Deconvolution approaches have been developed for transcriptomes to address this heterogeneity in tumor samples. We applied the DeMixT (Bioconductor R package) in two component mode (tumor and stroma) to 16 solid tumor types from The Cancer Genome Atlas (TCGA) data: BLCA, BRCA, COAD, HNSC, KICH, KIRC, KIRP, LIHC, LUAD, LUSC, PAAD, PRAD, READ, STAD, THCA, UCEC. We compared results of pathway and survival analyses from the original mixed expression data with those from the deconvolved expression data. Additionally, using cancer-specific biological information, we performed three-component deconvolution (tumor stroma and immune) for 5 cancer types (BLCA, COAD, PRAD, HNSC, KIRC) without requiring data from the immune profiles. For these 16 cancer types, we obtained tumor purities, immune proportions (if available) and deconvolved individual-level gene expression of each components. Downstream analyses in these cancer types were performed and biological findings from previous literature were validated using deconvolved expressions but not with original mixed expressions. Novel cell-type specific pathway activities were identified. Immune proportions were shown to be highly predictive of good prognosis in multiple cancer types. In summary, we provide comprehensive transcriptome deconvolution outputs for the TCGA datasets. These datasets will enable investigation of the tumor-stroma-immune microenvironment, which will elucidate new biological mechanisms for cancer.

***Projects***

Datasets: Download data from the TCGA data portal. Choose your favorite cancer types: Thyroid cancer, Breast cancer, or Lung cancer.

1. RNAseq deconvolution (short-term): Please perform 2-component deconvolution, and then 3-component deconvolution using DeMixT. Following the pipeline described here:

<https://www.biorxiv.org/content/10.1101/566075v1>

1. Matching RNAseq deconvolution metrics with H&E slide images (bonus): Define your biological hypothesis with the metric to be used, then use the matching H&E slides to accept or reject your hypothesis. The H&E slide images of TCGA samples are available here: <http://cancer.digitalslidearchive.net>/

DNAseq deconvolution (long-term): Here we aim to understand what is subclonal reconstruction. Step 1. Use a bam file simulator (a popular one is BamSurgeon) to simulate three bam files for heterogeneous tumor samples. Step 2. Run mutation calling (MuSE) and CNV calling (ascatNGS) on the simulated bam files. Step 3, choose three subclonal reconstruction methods to run on your data. E.g., Pyclone, PhyloWGS, and CliP. Step 4. Run concensus subclonal reconstruction using CSR.

**Wheeler, David, wheeler@bcm.edu**

***Lecture***

*The Genomic Landscape of DNA Damage Repair Deficiency in Cancer Patients*

The DNA of our genomes is under constant assault by mutagenic processes both endogenous and exogenous. There are over 270 genes known to play a role in the maintenance and repair of our DNA. Since cancer is a disease of the genome, damage to various components of the repair system contribute to the etiology of the disease, sometimes leading to surprising observations into genome structure and function. This lecture will discuss insights into the role of the DNA Damage Response system in shaping the cancer genome, and how we are using this information to uncover new avenues toward effective cancer therapy.

***Breakouts***

***Projects***

**Xiao, Jie,** [**xiao@jhmi.edu**](mailto:xiao@jhmi.edu)

***Lecture***

***Breakouts***

***Projects***