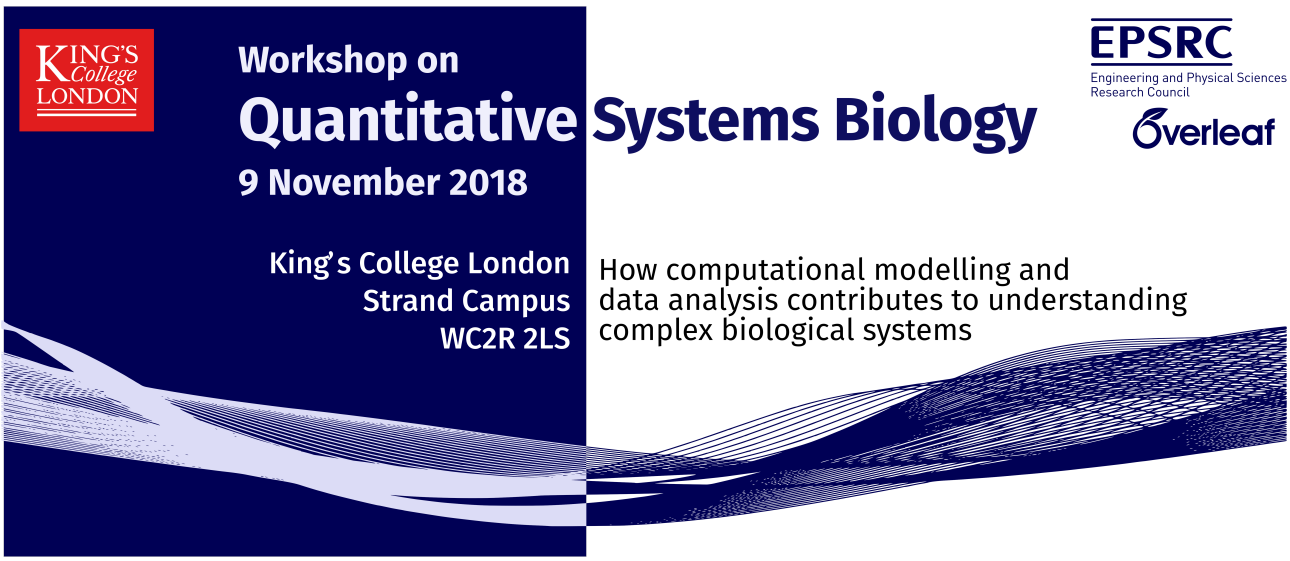

Workshop on

Quantitative Systems Biology

King's College London

November 9, 2018

EPSRC
Engineering and Physical Sciences
Research Council

KING'S College LONDON

Overleaf

Welcome Address

Following last year's success, it is with great pleasure that we welcome you to the second edition of the Quantitative Systems Biology (QSB) workshop. The aim of this workshop is to bring together experts working at the interface between biology/biomedicine and quantitative sciences, to discuss the latest developments in their research and novel approaches to existing problems.

Qualitative methods have been fundamental to natural science, as we make detailed observations on natural phenomena, analyse them and then put forward pertinent hypotheses to be tested. As the field grows quantitative tools have been indispensable in our understanding on the natural world: dynamical models are used to recapitulate the behaviour of genes, individuals, populations, enzymatic reactions, etc. Next-generation sequencing has become a staple of many experimental settings, only made possible by the use of bioinformatics and big data analytics. In this workshop we will explore the different ways in which quantitative science from fields such as physics, mathematics and computer science have fused with biology in order to further advance scientific developments. This workshop is organised by PhD students and wishes to encourage the interchange of ideas from people with different approaches and different questions, who nonetheless develop or apply quantitative tools in biological problems. This year we have invited a total of six speakers with a wide range of expertise spanning different areas of quantitative biology, and we are also delighted to feature two student talks in our programme. We hope these talks, together with two poster sessions, will provide much opportunity to learn about different methodologies and approaches in this emerging discipline.

We are fortunate to have the support of the EPSRC through the CANES (Cross-disciplinary Approaches to Non-Equilibrium Systems) Centre for Doctoral Training (CDT), based here at King's College London. We also owe many thanks to our sponsor Overleaf for helping to make this event possible. Without their support, QSB would be a much smaller affair.

Finally, we would like to thank all participants, speakers, and the wider scientific community for the fantastic reception we have received when organising this workshop. We hope QSB will act as a platform for networking and future collaborations.

We hope you enjoy the exciting scientific programme that we have on offer!

Yours faithfully,

The Organising Committee:
Tereza Gerguri
Edgar Herrera Delgado
Irene Marzuoli
Joseph Ng

Contents

Welcome Address **1**

Schedule **1**

Invited Speakers **2**

Learning biological dynamics from static data: constructing temporal trajectories using machine learning (*Christopher Yau*) . . . 2

Inferring the true tree: what to do when phylogenetic analyses disagree (*Michelle Kendall*) . . . 2

Supercomputing in the Cellular Jungle (*Sarah Harris*) . . . 3

Hybrid frameworks for modelling reaction-diffusion processes (*Kit Yates*) . . . 4

Programming bimolecular systems (*Neil Dachau*) . . . 5

From transcriptomic data to plausible biology hypotheses via bioinformatics (*Irenia Nobeli*) . . . 5

Selected Talks **6**

Simulations and Markov state models of protein disorder (***Sarah-Beth Amos***) . . . 6

Beating cancer escape room: let's use mathematical modelling to unlock cells! (***Nuria Folguera Blasco***) . . . 6

Poster Presentations **7**

Joel Nulsen . . . 7

Johann Bauer . . . 7

Rowan Howell . . . 7

Michael John Casey (presenting on behalf of Joseph R. Egan) . . . 8

Gregory Parkes . . . 8

Christopher Rookyard . . . 8

Bethan Cornell . . . 9

Cristina Parigini . . . 9

Carolina Jarmillo Oquendo . . . 9

Daniel Koch . . . 10

Irepan Salvador-Martinez . . . 10

Xiuyun Jiang . . . 11

Schedule

Time	Speaker	Title
09:30-10:00		**Coffee and registration**
10:00-10:10		**Introduction & registration**
10:10-10:55	Christopher Yau (Birmingham)	Learning biological dynamics from static data: constructing temporal trajectories using machine learning
10:55-11:40	Michelle Kendall (Oxford)	Inferring the true tree: what to do when phylogenetic analyses disagree
11:40-11:45		**5-min short break**
11:45-12:25	Sarah Harris (Leeds)	Supercomputing in the Cellular Jungle
12:25-13:30		**Lunch & poster session**
13:30-14:10	Kit Yates (Bath)	Hybrid frameworks for modelling reaction-diffusion processes
14:10-14:50	Neil Dalchau (Microsoft Research)	Programming bimolecular systems
14:50-15:20	Sarah-Beth Amos (Oxford)	Simulations and Markov state models of protein disorder (Student talk)
15:20-15:50		**Coffee**
15:50-16:30	Irilenia Nobeli (Birkbeck)	From transcriptomic data to plausible biology hypotheses via bioinformatics
16:30-17:00	Nuria Folguera Blasco (The Francis Crick Institute)	Beating cancer's escape room: let's use mathematical modelling to unlock cells! (Student talk)
17:00-17:15		**Concluding remarks**

Invited Speakers

Learning biological dynamics from static data: constructing temporal trajectories using machine learning

10:10

Christopher Yau
University of Birmingham

Bio: Christopher Yau is a Reader in Computational Biology at the Institute of Cancer and Genomic Sciences where he is based at the Centre for Computational Biology and leads the Statistical Machine Learning BioHealth group. He is an expert in statistical methodologies for machine learning and data science and specialises in genomic science particularly cancer. His research ranges from mathematical and statistical algorithm development to collaborations with experimental scientists and clinicians involving modelling real world biomedical data sets. He leads a diverse group of researchers who specialise in both experimental and computational modelling and regularly gives talks and lectures around the world on data science. His goal is to conceive of a computational intelligence framework that will provide the foundation of learning health systems that will support novel health-related research from the molecular scale through to whole populations.

Inferring the true tree: what to do when phylogenetic analyses disagree

10:55

Michelle Kendall
University of Oxford

Bio: I am a senior researcher in statistical genetics and pathogen dynamics at the Oxford Big Data Institute working on the BEEHIVE project which is a cross-European study of HIV genomics and virulence. I am particularly interested in developing tree-based methods for answering statistical questions about HIV genotypic variation within and between patients.

Supercomputing in the Cellular Jungle

11:45

Sarah Harris
University of Leeds

Abstract: Computational models have huge potential to provide insight into molecular biology by providing detailed animations of biomolecules and their interactions. In principle, these simulations act as a computational microscope, so long as the results that are obtained can be validated against experimental data. Molecular simulation can show how the shapes of biomolecules change due to their thermal motion, how the structure of individual biomolecules is affected by subjecting them to mechanical stress and the possible biological consequences conformational diversity. However, the computational expense of atomically detailed calculations, which require high performance supercomputer facilities, places serious limitations on the length and time-scales that can be accessed. I shall describe the successes and the challenges of simulations of biomacromolecules using examples from our own research on the importance of dynamics in protein recognition and DNA supercoiling.

Now that techniques such as cryo electron tomography are revealing highly organised supermacromolecular architectures at the length-scale directly above that of single molecules, which was previously invisibile, there is a need for new computational tools to intepret these experiments. I will present a new algorithm we are developing called Fluctuating Finite Element Analysis (FFEA) that uses continuum mechanics to model biomolecules that are too large to be simulated at the atomisic level. I will use our computer simulations of the molecular motor dynein in the axoneme to demonstrate that mesoscale super-macromolecular organisation cannot be ignored in mechanistic biology and conclude by commenting on future prospects for computer simulation in molecular biology.

Bio: My group uses high performance computing to model biomolecules and aim to cover all of the time and length-scales of relevance to molecular biology. This requires computational methods ranging from quantum mechanical, through atomistic to mesoscale modelling.

Hybrid frameworks for modelling reaction-diffusion processes

13:30

Kit Yates

University of Bath

Abstract: Spatial reaction-diffusion models have been employed to describe many emergent phenomena in biological systems. The modelling technique for reaction-diffusion systems that has predominated due to its analytical tractability and ease of simulation has been the use of partial differential equations (PDEs). However, due to recent advances in computational power, the simulation, and therefore postulation, of computationally intensive individual-based models has become a popular way to investigate the effects of noise in reaction-diffusion systems.

In a wide variety of biological situations, computationally-intensive, high-resolution models are relevant only in particular regions of the spatial domain. In other regions, coarser representations may suffice to capture the important dynamics. Such conditions necessitate the development of hybrid models in which some areas of the domain are modelled using a coarse-grained representation and others using a more fine-grained representation.

In this talk I will discuss recent work from my group on connecting coarse and fine representations of reaction-diffusion phenomena. The models to be coupled will include both on and off-lattice individual-based representations of diffusion with and without volume exclusion as well as macroscopic partial differential equations. In each scenario we will demonstrate good agreement between our hybrid models and the full individual-based representation whilst achieving significant computational savings.

Bio: Kit Yates is a Senior Lecturer in the Department of Mathematical Sciences and co-director of the Centre for Mathematical Biology at the University of Bath. He completed his PhD in Mathematics at the University of Oxford in 2011. His research demonstrates that mathematics can be used to describe all sorts of real-world phenomena: from embryo formation to locust swarming and from sleeping sickness to egg-shell patterning. He is particularly interested in the role that randomness plays in Biology. Some specific areas of work include cell migration, spatial hybrid simulation methods for reaction-diffusion systems and methodologies in stochastic simulations.

Programming bimolecular systems

14:10

Neil Dachau

Microsoft Research

Bio: I am a Scientist in the Biological Computation research group at Microsoft Research Cambridge. I am interested in how to program computation and decision-making in biological systems. The applications of programmed biology are numerous, including the synthesis of medicines and industrial chemicals, through to the direct treatment of disease. I studied Mathematics at the University of Oxford, UK (2001-2005), before becoming a biologist at the University of Cambridge. My PhD project was a collaboration between Alex Webbs group at the Department of Plant Sciences and Jorge Goncalves in the Control Group at the Department of Engineering. Following my PhD, I briefly held a research associate position in the Control Group, working with Glenn Vinnicombe on applications of stochastic control theory to gene networks. I came to Microsoft Research as a postdoc in 2009, during which I worked with Andrew Phillips on modelling immune systems and synthetic gene networks.

From transcriptomic data to plausible biology hypotheses via bioinformatics

15:50

Irenia Nobeli

Birkbeck, University of London

Bio: Irilenia Nobeli's lab has expertise in computational biology and chemoinformatics. Current interests include regulatory RNAs, transcriptomics in health and disease, disorders of the brain and the role of small molecules in Biology.

Selected Talks

Simulations and Markov state models of protein disorder

14:50

Sarah-Beth Amos
University of Oxford

Abstract: Peripheral membrane proteins bind to the cell membrane and modify its physical properties to convey intracellular messages. Many of these proteins contain regions of disorder that are important for their specific function, but their dynamic behaviour is difficult to capture on experimental timescales. Here we present work combining multiscale molecular dynamics (MD) simulations and Markov State models (MSM) to investigate how peripheral membrane proteins recognise and bind to the cell membrane. Type I phosphatidylinositol phosphate kinase A (PIP5K1A) is one such example and catalyses lipid phosphorylation in the cell membrane. PIP5K1A contains a disordered activation loop which confers lipid specificity. Our work shows the process of membrane recognition and binding by the activation loop. MSM analysis shows that the loop adopts specific conformations in the membrane to mediate lipid phosphorylation.

Beating cancer escape room: let's use mathematical modelling to unlock cells!

16:30

Nuria Folguera Blasco
The Francis Crick Institute

Abstract: The capacity of somatic cells to switch their phenotype in response to damage stimuli in vivo might have a pivotal role in ageing and cancer. However, how the mechanisms of phenotype reprogramming are established remains poorly understood. In order to elucidate such mechanisms, we present a stochastic model of combined epigenetic regulation(ER)-gene regulatory network to study the plastic phenotypic behaviours driven by ER heterogeneity. Our analysis of the coupled system reveals the existence of pluripotent stem-like and differentiated steady-states. Crucially, ER heterogeneity of differentiation genes is responsible for conferring abnormal robustness to pluripotent stem-like states. We formulate epigenetic heterogeneity-based strategies capable of unlocking and facilitating the transit from differentiation-refractory (pluripotent stem-like) to differentiation-primed epistates. Our results suggest that epigenetic heterogeneity regulates phenotypic robustness of cell fate reprogramming.

Poster Presentations

Joel Nulsen
King's College London

Abstract: Cancer progression is shaped by evolution. During tumour growth, somatic alterations are acquired by cells. Some of these alterations confer selective advantages on cells, and thereby promote cancer progression; these are termed 'driver alterations'. However, the selective pressures acting in a tumour are themselves influenced by various factors. We investigate the hypothesis that, at the level of biological processes, germline (inherited) genetic perturbations can lead to an 'inherited vulnerability' to certain somatic driver alterations in oesophageal adenocarcinoma patients.

Johann Bauer
City, University of London

Abstract: One of the simplest game theoretical models for host-parasite systems is the Matching Pennies game. If the co-evolution of host and parasite populations playing Matching Pennies is modelled by a replicator dynamics, one obtains non-convergent, periodic behaviour. We examine an extension of this case by explicitly considering the possibility of mutations in both the host and the parasite populations, obtaining a replicator-mutator dynamics. We show that the inclusion of arbitrarily small mutation probabilities rules out the usual periodic behaviour. Instead, small mutations cause the system to converge to a unique equilibrium arbitrarily close to the equilibrium of the system without mutation. Thus, if mutation probabilities decrease over time, the system converges to the equilibrium of the pure replicator dynamics. This highlights that under certain conditions arbitrarily low mutation probabilities can alter the quality of the co-evolution of such systems.

Using mixture models to analyse genome-wide screens

Rowan Howell
King's College London, The Francis Crick Institute

Abstract: Statistical analysis of genome-wide screens often assumes the data is distributed normally and determines a cutoff based on a z-score for outliers, however as the number of hits in a screen increases, this assumption breaks down. We reasoned that genome-wide data could be divided into two categories: genes that are unaffected by the screening conditions, and those that exhibit a response. Therefore, we applied bimodal normal mixture models to analyse a dataset of 23 genome-wide synthetic physical interaction screens. We found that the distribution of true interactions was accurately predicted by the models and cutoffs were less dependent on the number of hits. Mixture models offer a more rigorous foundation for analysis of genome-wide datasets as well as metrics for between-screen comparison. The tools developed also provide flexible methods to adapt cutoffs to optimise true and false positive rates.

Mathematical modelling of cancer immunology: Receptor-ligand interactions and T cell signaling

Michael John Casey (presenting on behalf of Joseph R. Egan)
University of Southampton

Abstract: T cell receptors can bind with antigens presented on the surface of tumour cells allowing an activating signal to be sent to the T cell. However, a tumour cell can avoid such an immune response by sending an inhibitory signal to the T cell via a second binding event. Let R represent a receptor, L represent its associated ligand, and B represent the complex that is formed when the ligand and receptor bind together. This interaction can be described as a reversible hetero-dimerization reaction where the ratio of the unbinding and binding rates is known as the dissociation constant, K_d. In a stochastic framework we have showed that if K_d is approximately equal to the larger of the total number of receptors, R_T, or the total number of ligands, L_T, then the number of bound complexes, B, is in a regime of high variability. We speculate that these high fluctuations in the number of bound complexes may have an impact on the immune response of T cells against tumour cells.

Is it possible to reconstruct an accurate cell lineage using CRISPR recorders?

Gregory Parkes
University of Southampton

Abstract: In this work we are interested in quantitative modelling of correlation across such gene products. Building on recent work, we develop computational models spanning transcript, translation and protein levels at different stages of the H. sapiens cell cycle. We enhance this analysis by incorporating 25+ sequence-derived features which are likely determinants of cellular protein concentration and quantitatively select for relevant features, producing a vast dataset with thousands of genes. We reveal the complex modelling interplay between expression levels across time as a domain with many small, cumulative factors using machine learning methods to highlight outliers with respect to such models as proteins associated with post-translationally regulated modes of action.

Christopher Rookyard
King's College London

Abstract: Early in the development of the nervous system of the zebrafish, the brain and spinal cord undergo a transition from a solid, rod-like structure to a hollow tube. To be able to properly make this hollow space along the midline of the tissue, the neuroepithelial cells on opposing sides must adjust their positions to make a straight interface. We are interested in the mechanics of this process; utilising laser incisions, we have seen that the tissue is under tension in two orthogonal directions, aligning with two of the principal axes of the tissue, the anterior-posterior and the medial-lateral. We are investigating the relative importance of these forces in making a straight midline interface through computational modelling, and have found that the straightening process is fundamentally different for the two tension orientations. Thus we return to experiments to look for signatures of either of these distinct straightening modes.

Bethan Cornell
King's College London

Abstract: Fluorescent molecular rotors are molecules whose fluorescence quantum yield and lifetime are functions of the viscosity and polarity of their local environment. These properties can be exploited to obtain real time imaging of processes involving viscosity changes, which may be signatures of specific diseases. Rotors based on BODIPY have long fluorescent lifetimes and, hence, make excellent probes. However, a detailed description of their working mechanisms has not yet been obtained.

Here, we focus on a prototypical BODIPY-based molecular rotor. We investigate computationally its ground and excited state properties as a function of the rotation of the rotatable portion, in order to assess and compare different methods and levels of theory. Moreover, we also measure experimentally its spectroscopical features. These are the first steps towards developing a computational/experimental protocol to unravel how BODIPY-based molecular rotors function and sense viscosity.

Cristina Parigini
University of Southampton

Abstract: Homeostasis in cycling tissues is related to a balance between cell proliferation and differentiation, for which the underlying dynamics are complex and not well understood. Although experimental data are of undoubted importance, mathematical modelling of cell dynamics are essential to understand this mechanism. The focus here is on the theoretical modelling of a generic system of cells, based on a continuous multi-type branching process, its deterministic approximation and graph-theory representation. Two biologically relevant types of process are identified: invariant and mixed cell fate. In one case, homeostasis is consistent to the asymmetric division model while in the other it is achieved at population level. The difference involves also the homeostasis robustness, the regulation via feedback and the clone size distribution. Clone size distribution is fundamental when dealing with the analysis of lineage tracing data, based on which the applicable cell fate model can be revealed.

Carolina Jarmillo Oquendo
University of Southampton

Abstract: A heterogeneous diagnosis combined with a lack of studies make it difficult to establish appropriate treatment and prognosis in splenic marginal zone lymphoma (SMZL). The purpose of this study is to optimise a bioinformatic pipeline to call somatic variants with tumour only samples to expand the catalogue of mutations in SMZL. Prior to analysis, a database of previously identified variants was constructed for comparison. HaloPlex HS was used for the amplicon library preparation and NextSeq for sequencing. Reads were aligned to the hg38 assembly. LocatIt merged duplicate reads and GATK was used for variant calling. On average, there were 194 coding variants called per individual sample and this number was reduced to 133 when duplicates were merged. This change partially accounted for a batch effect observed in raw data. Further optimisation of the pipeline will require optimising sensitivity for lower variant allele frequencies to detect subclonal mutations and exclude false positives.

Bayesian Network Methods for Mechanistic Modelling in Systems Biology

Daniel Koch

King's College London

Abstract: Bayesian Networks (BNs) are probabilistic graphical models that are able to simplify reasoning under uncertainty on the basis of directed acyclic graphs representing statistical independencies. The formalism also allows for learning models including network structure from data, making it attractive for discovering the structure of gene regulatory or signalling networks (e.g. from omics-type data). In the philosophy of science the theory of BNs facilitated major progress in understanding the concept of causality in a rigorous and axiomatic manner, rendering them attractive for predictive and mechanistic modelling, too. In an interdisciplinary collaboration, we have successfully applied causally interpreted Bayesian networks for modelling disease progression mechanisms in chronic myeloid leukemia. Our results indicate that multiple secondary changes are necessary for the progression to its final stage, some of which potentially form positive feedback loops. In our simulations, the success of tyrosine kinase inhibitors such as imatinib for treating this disease relies on their ability to interfere with these feedback loops. In another project, we currently explore the potentials to hybridise BNs with agent based methods such as cellular automata. In a first toy-model study we have shown that BNs, often inapt to represent rich spatial information, can be supplemented by cellular automata to capture spatially more detailed mechanisms.

Is it possible to reconstruct an accurate cell lineage using CRISPR recorders?

Irepan Salvador-Martinez

University College London

Abstract: Cell lineages provide the framework for understanding how organisms are built and cell fates are decided during development. Recently, the idea of using CRISPR to induce mutations during development and use them as heritable markers for lineage reconstruction has been proposed. While an attractive idea, its practical value depends on the accuracy of the lineages that can be generated by this method. Here, we use computer simulations, informed by empirical data on CRISPR mutational frequencies in Drosophila, to estimate the performance of this approach under different conditions. We show significant impacts from multiple biological and technical parameters - variable cell division rates, skewed mutational outcomes, target dropouts and sequencing strategies. Our approach reveals the limitations of existing recorders, and indicates how future implementations can be optimised to improve lineage reconstruction.

Xiuyun Jiang
University College London

Abstract: Multi-heme proteins have attracted much attention recently due to their prominent role in mediating extracellular electron transport (ET), but one of their key fundamental properties, the rate constants for ET between the constituent heme groups, have so far evaded experimental determination. For the modelling of the electron transfer reactions between the hemes, we need to compute the diabatic state coupling matrix element, which is an important parameter in calculating heme-heme electron transfer(ET) rates.Several quantum models are tested by systematically including more side chains of the heme porphyrin. Finally the calculation is done in a reasonably big model and averaged over 25 snapshots from a molecular dynamics trajectory. Rather surprisingly, we find that electronic coupling between the hemes is significantly enhanced by the cysteine linkages for two terminal hemes.
