NHS
National Institute for Health Research

Joint BRC Biomedical Research Centre and Kings Centre for Bioinformatics

---

BRC Methodology Collaboration: Translational Bioinformatics seminar

Tuesday 21st February 2012, 12:30pm – 5:00pm

Lecture Theatre, Floor 30, Guys Tower Wing

Location map: http://www.kcl.ac.uk/campuslife/campuses/guys/Guys.aspx

Please register at comp-bio@kcl.ac.uk with subject "register"

Programme

12.30pm: Buffet lunch

1.00pm - Professor Ton Coolen: *What you see is not what you get: how sampling affects macroscopic features of biological networks.*

1.30pm - Dr Davide Bacciu: *Learning Bayesian Network skeletons with high-dimensional and large-sample size data.*

2.00pm - Professor Alfonso Valencia: *Computational Challenges in Personalize Medicine*

2.30 Tea break (30mins)

3pm - Dr Eric Yang: *Combining Disparate Clinical Datasets For Integrated Analysis*

3.30pm - Dr Paolo Vigneri: *The Untouchables: mapping critical residues in the BCR-ABL catalytic domain to understand TKI resistance in Chronic Myeloid Leukemia*

4pm - Professor Peter Coveney: *Computational Biomedicine: A Challenge for the Twenty-First Century*

4.30pm - Dr. Nour Shublaq: *Getting Personal through Translational Bioinformatics: The Future Now*

5.00pm Visit to local pub or move to Lecture Theatre 1, New Hunt's House (King's College London Guy's Campus, London SE1 1UL) for the next talk in the King's International Lecture Series (http://www.kingshealthpartners.org/info/kings-international)
Professor Mark Musen: *Becoming a meta-physician: better clinical information for the future of health care*

Seminar Location
King's College (Guy's) Tower
Floor 30, Lecture Theatre
Guys Hospital
Great Maze Pond
London
SE1 9RT

Detailed programme

12.30pm - Buffet lunch

1pm - Prof Ton Coolen
Kings College London (http://www.mth.kcl.ac.uk/~tcoolen/)
Title*:* What you see is not what you get: how sampling affects macroscopic features of biological networks.
Abstract*:* We use methods from the theory of tailored random graphs to study the effects of imperfect sampling on topological features of large biological signaling networks. Our aim is to increase our quantitative understanding of the relation between true biological networks and the imperfect and often biased samples of these networks that are reported in public data repositories and used by biomedical scientists. We derive exact explicit formulae for the macroscopic structural features of sampled networks, in terms if those of the underlying true network, for a broad family of sampling protocols that include (un-)biased node and/or link under-sampling as well as (un-)biased link over-sampling. These formulae should enable us in the near future to decontaminate e.g. protein interaction networks rigorously for the effects of method-specific experimental bias.

1.30pm - Dr Davide Bacciu
Universita' di Pisa
Title*:* Learning Bayesian Network skeletons with high-dimensional and large-sample size data.
Abstract*:* High throughput technologies make available biomedical data collections characterized both by high sample dimensionality as well as by a large number of observations. Constraint-based structure learning methods are theoretically and computationally well-suited for identifying networks of probabilistic interactions with high-dimensional data. However, the interplay between a high-dimensionality and a large sample-size introduces issues that

have, so far, received little attention. We discuss a fundamental issue concerning the reliability of association tests within such a large-scale scenario and we introduce a general, principled strategy for the ordering of association tests. Finally, we provide a review of available software packages, focusing on the trade-off between computational and reconstruction performance.

2pm - Prof. Alfonso Valencia
Spanish National Cancer Research Centre (CNIO)
(http://www.cnio.es/ing/grupos/plantillas/investigacion.asp?pag=728)
Title*:* Computational Challenges in Personalize Medicine
Abstract*:* Sequencing technology is making of the use of personal genomic information a pressing clinical reality. In this scenario Bioinformatics and Computational Biology play a central role in the organization and interpretation of the information.
The Computational workflows required for the analysis of the individual genomes involve complex Bioinformatics tasks and many hurdles related with data structure and integration. Less obvious, but perhaps more important than the engineering problems, are the scientific challenges underlying every steps of the analysis. Indeed, the interpretation of complex genomic data requires computational methods that in most cases belong to active research areas.
In this context my group participates both the implementation of systems in a pre-clinical setting and in the development of the science required at various levels. In particular, we focus in the analysis of the incidence of mutations in: a) structure and function of proteins, particularly in binding sites, b) splicing and splice sites, c) comparative analysis of affected pathways, and d) extraction of mutation-drug-disease relations from databases and text-sources. (Needless to say that many other important facets of genomic information are equally important in a comprehensive genome analysis study).
The initial experiences of the groups working in personalize-medicine initiatives shows that the many complex challenges in this emerging area will require a concerted and collaborative effort of the bioinformatics community.

2.30 Tea break (30mins)

3pm - Dr Eric Yang
Johnson&Johnson
Title*:* Combining Disparate Clinical Datasets For Integrated Analysis
Abstract*:* As a company, Johnson and Johnson sees the integration of datasets generated by public groups such as ADNI, academic partners such as Kings College London, and our internal trial databases as a significant opportunity to enhance our underlying understanding of Alzheimer's Disease. However, the integration of these different datasets is a challenge, because each individual group will have a different underlying schemas, different underlying attribute labels, and finally different patient populations. The challenge is to transform such data so that we can utilize a single set of tools and analyses and not only duplicate analysis across datasets easily, but also combine datasets in the same analysis to increase statistical power. To do so, we have developed techniques for schema level integration, semantic data integration, and finally patient level data integration. Schema level integration is the process by which we transform

all of the different data sources, whether they be available in the form of tabular formats, SAS files, etc, into a common data schema which we store in a SQL database that can be queried through one set of tools and/or scripts. Semantic data integration attempts to take descriptors for a given attribute such as ADAS-Cog, and attempts to map them to a standardized vocabulary so that we can cross query for the same attribute across multiple datasets with a minimum amount of customization in our analysis pipeline. Finally in our patient level integration, we attempt to deal with the fact that in many cases trials are run for different lengths of time, with different entry criteria. To accomplish this, we have introduced a relatively simple concept of synchronizing all of the patients via the concept of disease duration prior to trial entry. Utilizing this concept, we can relate the progression of each patient individually to every other patient, thus allowing us to utilize them all in a single analysis. Taken as a whole, we believe that this work will be pivotal to our ability both understand the disease as well as derive the maximum utility from currently available datasets.

3.30pm – Dr. Paolo Vigneri
University of Catania
(http://www.linkedin.com/pub/paolo-vigneri/38/607/1a7)
***Title**:* The Untouchables: mapping critical residues in the BCR-ABL catalytic domain to understand TKI resistance in Chronic Myeloid Leukemia
Abstract: Suppression of BCR-ABL1 tyrosine kinase activity by Imatinib Mesylate (IM) has dramatically improved the natural history of Chronic Myeloid Leukemia (CML) ushering the era of molecular targeted therapy. Despite the unparalleled results achieved with IM, 30% of CML patients fail the drug, mostly because of point mutations in the BCR-ABL catalytic domain that interfere with IM binding.
These point mutations are not randomly distributed in the BCR-ABL kinase sequence. Indeed, only one (T315) of five residues critical for IM interaction is the object of amino acidic substitutions in IM-resistant patients. Combining computational and biological approaches, we demonstrate that the four remaining amino acids are pivotal to preserve both BCR-ABL kinase activity and oncogenic potential.
These findings contribute to explaining the efficacy of second and third generation tyrosine kinase inhibitors (TKIs) on CML patients failing IM. Furthermore, they may have far reaching implications for the design of further TKIs aimed against the ever-growing number of protein kinases involved in solid and haematologic malignancies.

4pm - Prof. Peter Coveney
Director, Computational Life & Medical Sciences Network; Centre for Computational Science, University College London
(http://www.ucl.ac.uk/chemistry/staff/academic_pages/peter_coveney)
***Title**:* Computational Biomedicine: A Challenge for the Twenty-First Century
***Abstract**:* With the relentless increase of computer power and the widespread availability of digital patient-specific medical data, we are now entering an era when it is becoming possible to develop predictive models of human disease and pathology which can be used to support and enhance clinical decision-making. The approach amounts to a grand challenge to computational science insofar as

we need to be able to provide seamless yet secure access to large scale heterogeneous personal healthcare data in a facile way, typically integrated into complex workflows some parts of which may need to be run on high performance computers, integrated into clinical decision support software. In this talk, I look at the state of the art in terms of case studies drawn from neurovascular pathologies and HIV/AIDS. These projects are representative of a large number currently being performed within the European Virtual Physiological Human and related e-Health initiatives. They make demands of information technology at many scales, from the desktop to national and international infrastructures for data storage and processing, linked by high performance networks, and set the agenda for much of twenty-first century biomedical research and clinical practice.

4.30pm - Dr Nour Shublaq
Scientific Research and Strategic coordinator for the UCL Computational Life& Medical Sciences Network, UCL
Title*:* Getting Personal through Translational Bioinformatics: The Future Now
Abstract*:* With the advent of whole human genome sequencing in minutes and the costs now counted in hundreds of pounds, the potential of the Human Genome Project can finally begin to be realised. At the same time, large initiatives such as the Virtual Physiological Human are addressing pathologies at a phenotypical level. The challenges in front of us to meld and convert these rapidly developing scientific capabilities into useful medical and clinical information remain very substantial -- from the secure management of the clinical derived data across hospital-university interfaces, via the development of large scale data integration warehouses, and back into clinical decision support systems. There are also fundamental issues associated with how we treat the genotype-phenotype relationship. We will survey the state of the art and highlight the problems that currently limit our ability to implement the vision for personalised medicine in its full form.