GENOPOLE

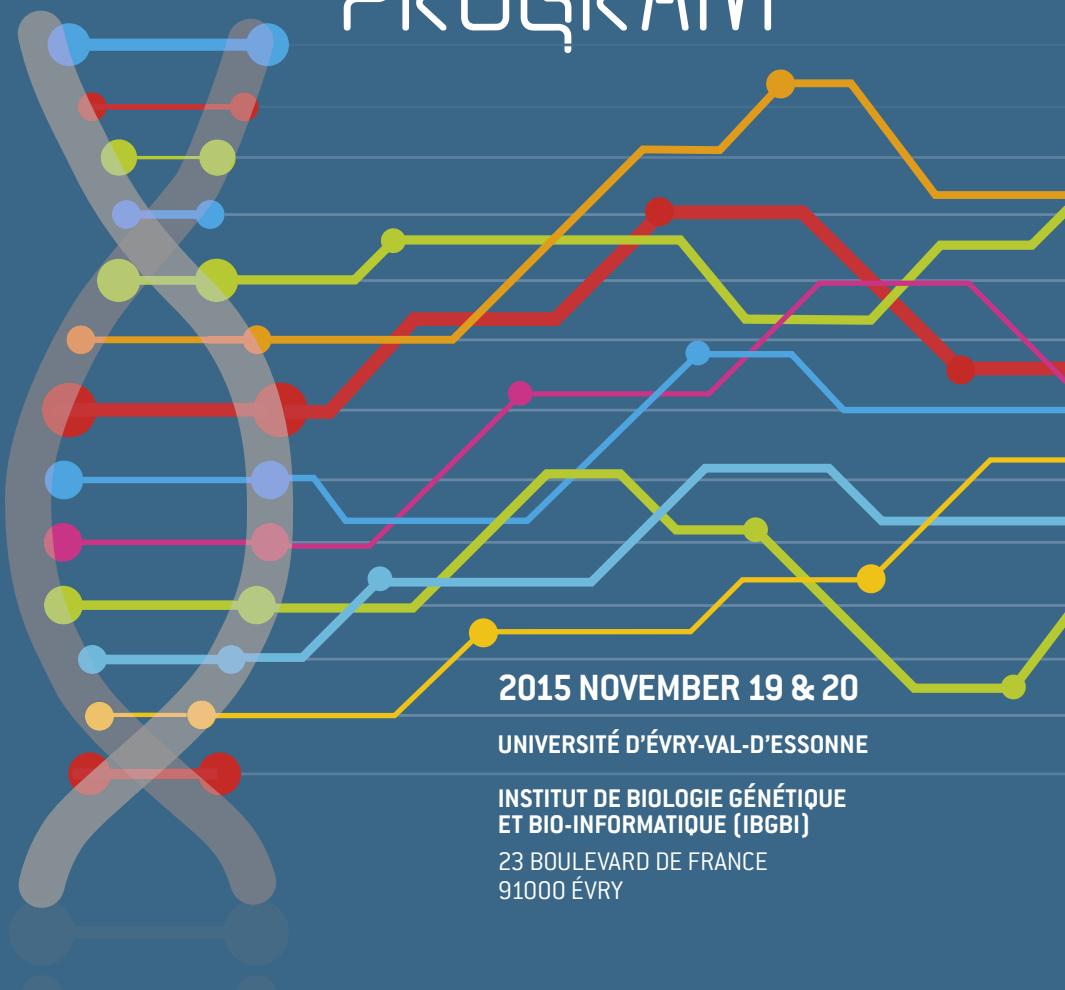
LaMME

Statistique et Génome

BIOMATHEMATICS CONFERENCE

STATISTICAL ANALYSIS OF MASSIVE GENOMIC DATA

PROGRAM



2015 NOVEMBER 19 & 20

UNIVERSITÉ D'ÉVRY-VAL-D'ESSONNE

INSTITUT DE BIOLOGIE GÉNÉTIQUE ET BIO-INFORMATIQUE (IBGBI)

23 BOULEVARD DE FRANCE
91000 ÉVRY

PROGRAM

BIOMATHEMATICS CONFERENCE

STATISTICAL ANALYSIS OF MASSIVE GENOMIC DATA

NOVEMBER 19

9:00 / 9:30
Registration of participants

• – 9:30 / 9:40
Opening remarks and introduction
Patrick Curmi, *President of the Université d'Évry-Val-d'Essonne*

• – 9:40 / 10:40
François Cambien *(Inserm, UPMC, Paris, France)*

10:40 / 11:10
Coffee break – **Poster session**

• – 11:10 / 12:10
Anne-Laure Boulesteix *(University of Munich, Germany)*

12:10 / 14:00
Lunch – **Poster session**

• – 14:00 / 15:00
Stéphane Robin *(INRA-AgroParisTech, Paris, France)*

15:00 / 15:30
Coffee break – **Poster session**

• – 15:30 / 16:30
Mark van de Wiel *(Vrije Universiteit Amsterdam, the Netherlands)*

• – 16:30 / 17:30
Andras Paldi *(Généthon-EPHE, Évry, France)*

NOVEMBER 20

9:00 / 9:30
Registration of participants

• – 9:30 / 10:30
Peter Visscher *(University of Queensland, Australia)*

10:30 / 11:00
Coffee break – **Poster session**

• – 11:00 / 12:00
Françoise Clerget *(Inserm U781, Hopital Necker Enfants Malades, Paris-Descartes University, Paris, France)*

12:00 / 14:00
Lunch – **Poster session**

• – 14:00 / 15:00
Sandrine Dudoit *(University of California, Berkeley, USA)*

15:00 / 15:30
Coffee break – **Poster session**

• – 15:30 / 16:30
Lodewyk Wessels *(The Netherlands Cancer Institute, Amsterdam, the Netherlands)*

• – 16:30 / 17:30
Jocelyn Laporte *(IGBMC, Illkirch, France)*

• – 17:30 / 17:40
Closing and conclusion
Emmanuel Dequier, *Director, Genopole Research and Global Infrastructure*

This two-day cross-disciplinary conference will bring together biologists, geneticists, clinicians, bioinformaticians and statisticians in order to discuss emerging challenges raised by the analysis of high-throughput genomic data, and present dedicated innovative approaches.

It will first and foremost focus on applications to personalized and predictive medicine, but also to various issues related to genetics, epigenetics, molecular phenotypes and metagenomics.

This conference is part of Genopole's strategic objective to host new research teams or researchers within the next few years in the field of biomathematics and biostatistics.

Patrick Curmi

President of the Université d'Évry-Val-d'Essonne, Director of Research at Inserm, Université d'Évry-Val-d'Essonne, Laboratory SABNP (Structure-Activité des Biomolécules Normales et Pathologiques)

presidence@univ-evry.fr

After his medical doctorate, Patrick Curmi turned to fundamental research, and completed a PhD in Biochemistry. As a specialist in molecular and structural biology and in nanotechnologies, he has worked in various Inserm laboratories since 1985. Convinced of the importance of multidisciplinarity, he coordinated and steered a European consortium to build fluorescent diamond nanoparticles, with the idea of using them as multipurpose and permanent markers in biology/medicine. In 2007, he founded the laboratory SABNP (Structure-Activité des Biomolécules Normales et Pathologiques), a joint research unit (UMR) Inserm-Université d'Évry. He taught optical and electronical microscopy at the Université d'Évry, and nanomaterials at the École supérieure des mines of Paris. Member of the Scientific Council of the Université d'Évry from 2006 to 2011, he became its Dean in 2012. He was elected President of the Université d'Évry-Val-d'Essonne in January 2015.

NOVEMBER 19

9h30

Opening remarks and introduction

Welcoming address by Patrick CURMI to scientists from around the globe, gathered to discuss, over two days, the latest advancements in biology, genetics, clinics, bioinformatics and statistics in the analysis of massive genomic data. Deeply involved in building Tomorrow's Medicine, the Université d'Évry-Val-d'Essonne is particularly proud to host this event.

François Cambien

Director of Research Emeritus
at INSERM UMRS 1166, former director of UMRS 937,
co-coordinator of LABEX GENMED
francois.cambien@upmc.fr

NOVEMBER 19

François Cambien received his MD degree from Faculty Cochin, Paris. He later specialized in cardiovascular epidemiology and statistics then in human genetics and genetic epidemiology. He joined the French institute of health and medical research (INSERM), and, in 1988, was appointed director of the INSERM 'DNA bank for cardiovascular research', one of the first large-scale human DNA resources in this area of research. The department was involved in many national and international studies and was in charge of managing and analysing over 40,000 DNA samples from patients and healthy controls covering most common cardiovascular disorders. Over the past 5 years, the laboratory (INSERM UMRS 937 and then UMRS 1166) has been involved in genome-wide association studies (GWAS) of various cardiovascular diseases. Most recently, Dr Cambien has been particularly implicated in GWAS of dilated cardiomyopathy. He has also investigated the contribution of genetic and non-genetic factors to genome wide expression in circulating cells, in large healthy populations.

9h40

Genome-wide transcriptomics and eQTL studies in human populations

Using microarray or sequencing technologies, it is possible to simultaneously quantify all RNA transcripts in a biological sample. The transcriptome is considerably more complex and difficult to investigate than the genome. It differs in different tissues and cell-types and is affected by genetic variability (eQTLs) but also by numerous other factors. The covariation of RNA transcripts reflects a modular, hierarchical and dynamic architecture which has pathophysiological implications. Thousands of eQTLs and biologically relevant networks of genes have been identified. However, studies are still limited in size, tissue/cell type coverage, availability of non-genetic data, and by the disconnection of data sets (e.g. different tissues in different studies). In addition, it is generally impossible to investigate disease pathways or to integrate different levels of high-dimensional data. When the appropriate resources become available, data analysis and integration (across diseases, tissues/cells, species) will constitute an interesting statistical challenge.

Anne-Laure Boulesteix

Prof. Dr. HDR Associate professor,
Université Louis Maximilien,
Munich, Germany
boulesteix@ibe.med.uni-muenchen.de

NOVEMBER 19

Anne-Laure Boulesteix is an associate professor at the Department of Medical Informatics, Biometry and Epidemiology of the University of Munich. She received her PhD in statistics in 2005 from the same university and her accreditation to supervise research (HDR) in 2011 from the Université d'Évry-Val-d'Essonne. Her research focuses on the statistical analysis of massive omics data, with emphasis on prediction modelling, non-parametric supervised classification, validation and issues related to scientific practice. She also works as a statistical consultant for medical doctors and teaches biostatistics to students in statistics, epidemiology and medicine.

11h10

Prediction models with low-dimensional clinical and high-dimensional omics data

Anne-Laure Boulesteix, Riccardo De Bin, Willi Sauerbrei

Over the last 15 years, literature has suggested numerous signatures derived from high-dimensional omics data (e.g. gene expression data) could predict patient outcomes such as survival time or response to therapy. A bitter disillusion followed the enthusiasm of early years, as researchers realized that the predictive ability of many signatures failed to be validated when evaluated based on independent datasets or did not show any added predictive value compared to classical clinical predictors which are much easier and cheaper to collect. In this talk, I will present results on different aspects of the combination of low-dimensional clinical and high-dimensional omics data in the context of prediction modelling. Firstly, it is not trivial to build a prediction model which fully exploits both types of predictors. I will present different strategies, illustrate them through the application to two case studies and show results of an extensive simulation. Secondly, I will address the problem of validating the added predictive value of high-dimensional omics data compared to classical clinical predictors. Thirdly, I will discuss general issues related to the benchmarking of prediction methods and the stability of model selection with particular focus on the combination of low and high-dimensional data.

Stéphane Robin

Research Director, French National Institute for Agronomic Research (INRA), AgroParisTech, Paris, France

robin@agroparistech.fr

NOVEMBER 19

Stéphane Robin is a senior researcher (DR) at the French Institute for Agronomic Research (INRA). His research focuses mainly on the development of original statistical methods for the analysis of biological data, more specifically genomic data of various types. From a methodological point of view, Stéphane Robin is interested in latent variable models, change-point detection and network analysis. In terms of teaching, he is associated with the Master 2 "Mathématiques pour les sciences du vivant" at the University Paris-Saclay.

14h00

Detection of recurrent alterations: Adapting statistics to data dimension

Modern genomic technologies enable the detection of alterations (loss or gain of chromosomal regions) that may occur in a patient's genome. We then look at alterations that are frequently observed in a cohort of patients with the same disease, e.g. the same cancer type. Such alterations are named recurrent.
Under simple but reasonable assumptions, the significance of such regions can be rephrased in terms of excursions of some Markovian process. The calculation of the corresponding p-value raises combinatorial issues, the complexity of which increases with both the number of patients n and the number of loci p. Due to advances of molecular biology technologies, both numbers have increased in the last decade.
Three approaches will be presented, corresponding to three historical periods. Each approach relies on a different probabilistic tool. CGH array analysis (small p, small n) can be achieved using finite embedded Markov chain. Continuous time Markov processes (birth and death process) can be used to deal with SNP arrays (small p, large n).
As for deep sequencing technologies (NGS: large p, large n), a limit Ornstein-Uhlenbek process will be introduced.

Mark van de Wiel

Full Professor in Statistics for Genomics, Dep. Epidemiology & Biostatistics and Dep. Mathematics, VU University (medical center), Amsterdam, the Netherlands

mark.vdwiel@vumc.nl

NOVEMBER 19

After obtaining his PhD in 2000 on 'Nonparametric Inference', Mark van de Wiel decided to start working in Biostatistics and specialized in developing methodology for, in those times emerging, genomics data like microarrays. He has worked on topics like FDR-based multiple testing, but also on more data-centered methodology in particular for DNA copy number microarrays. Since 2010 his focus has shifted somewhat towards (Bayesian) methodology for network estimation, biomarker detection and clinical prediction. 'Penalized regression' and 'empirical Bayes' are central notions in his recent work. He teaches two master-courses on 'High-dimensional data analysis': one for mathematics students and one for applied statistics students. Since October 2013, he has chaired the Statistics for Genomics unit, which consists of 10 researchers. This unit aims at developing statistical methodology for the analysis of genomics data, with a focus on Molecular Networks (Wessel van Wieringen), Data Integration (Renee de Menezes) and Prediction. In addition, the unit helps biomedical researchers with the analysis of their genomics data.

15h30

How to learn from a lot: Empirical Bayes in Genomics

The high-dimensional character of genomics data generally forces statistical inference methods to apply some form of penalization, e.g. multiple testing, penalized regression or sparse gene networks. The other side of the coin, however, is that the dimension of the variable space may also be used to learn across variables (like genes, tags, methylation probes, etc). Empirical Bayes is a powerful principle to do so. In both the Bayesian and frequentist paradigms it comes down to estimation of the a priori distribution of parameter(s) from the data. We briefly review some well-known statistical methods that use empirical Bayes to analyze genomic data. We believe, however, that the principle is often not used at its full strength. We illustrate the flexibility and versatility of the principle in three settings: 1) Bayesian inference for differential expression from count data (e.g. RNAseq), 2) prediction of binary response, and 3) network reconstruction. For 1) we develop a novel algorithm, ShrinkBayes, for the efficient simultaneous estimation of multiple priors, allowing joint shrinkage of multiple parameters in differential gene expression models. This can be attractive when sample sizes are small or when many nuisance variables like batch effects are present. For 2) we demonstrate how auxiliary information in the form of 'co-data', e.g. p-values from an external study or genomic annotation, can be used to improve prediction of binary response, like tumor recurrence. We derive empirical Bayes estimates of penalties of groups of variables in a classical logistic ridge regression setting, and show that multiple sources of co-data may be used. Finally, for 3) we combine empirical Bayes with computationally efficient variational Bayes approximations of posteriors for the purpose of gene network reconstruction through the use of structural equation models. These models regress each gene on all others, and hence this setting can be regarded as a combination of 1) and 2).

NOVEMBER 19

Andras Paldi

École Pratique des Hautes Études - UMR951 INSERM/GENETHON INTEGRARE - 1bis, rue de l'Internationale - 91002 ÉVRY - France

Andras.Paldi@ephe.sorbonne.fr

Andras Paldi received his PhD in 1987 from the Eötvös University in Budapest. He has been teaching at the École Pratique des Hautes Études since 1995 and is a researcher at Généthon. Andras Paldi's research is focused on two highly debated and challenging concepts and their application in biology: the role of epigenetic mechanisms in cell differentiation and the stochasticity of gene expression. Both concepts raise a number of questions regarding their impact on cell differentiation. His lab is investigating (i) how the clonal stability of cell phenotypes and the underlying stability of gene expression levels can be reconciled with the stochastic fluctuations within the cells and in its microenvironment; (ii) how cell fate decisions are taken in spite of the fluctuations; (iii) how experimental manipulations such as transduction using lentiviral vectors generate epigenetic changes in the host cell genome.

16h30

Stochasticity as a challenge to genetic determinism in the cell

Proteins, RNA-s and other biologically important molecules are present in low numbers in the cell leading to substantial fluctuation of their abundance. These fluctuations generate heterogeneity, even between genetically identical cells, and pose a challenge to our understanding of fundamental biological phenomena as differentiation or phenotypic stability of the cell.
There are two opposing views on this variation. One possibility is that the variability is harmful and simply represents a "noise" and the cells have to cope with it as best as they can.
The alternative is that molecular fluctuations are not only unavoidable but important, and cells exploit them to generate coherent behaviour at the cellular population or tissue level. It is not known what the relative contribution of these two possibilities is to normal and pathological functioning. The best way to address this question is to acquire a sufficient amount of single cell data and construct models that allow inferring on the mechanisms and understanding how reproducible population-level kinetics emerge from the highly variable individual cell behaviours.

NOVEMBER 20

Peter Visscher

Professor of Quantitative Genetics, University of Queensland, Australia

peter.visscher@uq.edu.au

Professor Visscher is an internationally recognized researcher with an outstanding track record working at the interface of quantitative genetics, statistical genetics, population genetics, human genetics, animal genetics, bioinformatics and genetic epidemiology. His expertise is reflected in a track record of publishing papers and obtaining grants across disciplines in life sciences and medicine. The fact that his expertise spans a variety of fields which are relevant to this application is also demonstrated by his publication record (Nature, Science, Nature Genetics, Nature Reviews Genetics, Genome Research, Genome Biology, PLoS Genetics, American Journal of Human Genetics, Nucleic Acids Research, Genetics, Behavior Genetics, Genetics and Human Molecular Genetics). The topics of his research output vary from quantitative and population genetics to genetic epidemiology, genomics, human genetics and bioinformatics. The focus of his current research activities is in the detection and fine-mapping of loci underlying complex traits, based upon theoretical studies and application of methods to real data, in population genetic studies using theoretical approaches and high-density genetic marker data, and in systems genomic studies.

9h30

Combining whole genome genetic and epigenetic data to improve phenotypic prediction of complex traits

Genome-wide association studies have succeeded in identifying important common genetic variants that underlie complex disease. There is also growing evidence that environmental exposures and risk factors for disease leave an epigenetic imprint that can be dissected using genomics technology. We quantify the accuracy of prediction of past or current disease-related environmental exposures from multi-locus epigenetic models and the precision of predicting body-mass-index and other complex trait phenotypes by combining a genetic and methylation predictor.

NOVEMBER 20

Françoise Clerget-Darpoux

Research Director (DRCE) at INSERM.
Presently retired with an emeritus status at the Imagine Institute (INSERM unit781), Paris, France.
francoise.clerget@inserm.fr

After a double training in Mathematics and Genetics, Françoise Clerget-Darpoux has focused her research work on statistical genetics. For ten years, she has led the head of a department devoted to multiple aspects of statistical genetics, population genetics and genetic epidemiology encompassing both methodology developments and their application to the study of human diseases (in particular, auto-immune and neuro-psychiatric diseases). For 15 years, she has been in charge of a course on Statistical Genetics (master level) at University Paris-Sud.

11h00

Interpretation of GWAS data: beware of the GIGO syndrome

Thanks to molecular biology, progress on monogenic diseases over the last two decades of the 20th century has been incredible.
In this new century, with the sequencing of the genome, we may also expect a huge advance in the understanding of human diseases, in particular with the identification of many monogenic sub-entities.
It seems however that, in their enthusiasm, geneticists have forgotten that most human diseases are very heterogeneous and complex in their etiologies.
As a result, we observe recurrent GIGO (Garbage In Garbage Out) epidemics. A first outbreak took place in the late 1980s, when geneticists, building on the success of model-based linkage analysis in monogenic diseases, started to use monogenic models to study multifactorial diseases.
A second outbreak is under way, with the computation of heritability or of individual risk estimates for multifactorial diseases from Genome-Wide Association Study (GWAS) data. The assumed model is no longer a monogenic model but a polygenic additive.
It took a long time to escape from the monogenic paradigm; it is now urgent to escape from the polygenic one! GIGO syndrome prevention requires a move away from simplistic models and the development of novel strategies combining different sources of information to progress in the understanding of disease aetiology.

NOVEMBER 20

Sandrine Dudoit

Professor of Biostatistics and Statistics, Chair and Head Graduate Advisor, Graduate Group in Biostatistics, University of California, Berkeley, USA
sandrine@stat.berkeley.edu

Sandrine Dudoit is Professor of Biostatistics and Statistics and Chair of the Graduate Group in Biostatistics at the University of California, Berkeley. Professor Dudoit's methodological research interests regard high-dimensional inference and include exploratory data analysis (EDA), visualization, loss-based estimation with cross-validation (e.g. density estimation, regression, model selection), and multiple hypothesis testing. Much of her methodological work is motivated by statistical inference questions arising in biological research and, in particular, the design and analysis of high-throughput microarray and sequencing gene expression experiments, e.g. mRNA-Seq for transcriptome analysis and genome annotation and ChIP-Seq for DNA-protein interaction profiling (e.g. transcription factor binding). Her contributions include: exploratory data analysis, normalization and expression quantitation, differential expression analysis, class discovery, prediction, integration of biological annotation metadata (e.g. Gene Ontology (GO) annotation). She is also interested in statistical computing and, in particular, reproducible research. She is a founding core developer of the Bioconductor Project (www.bioconductor.org), an open-source and open development software project for the analysis of biomedical and genomic data.

14h00

Identification of Novel Cell Types Using Single-Cell Transcriptome Sequencing

Single-cell transcriptome sequencing (scRNA-Seq), which combines high-throughput single-cell extraction and sequencing capabilities, enables the transcriptome of large numbers of individual cells to be assayed efficiently.
Profiling of gene expression at the single-cell level for a large sample of cells is crucial for addressing many biologically relevant questions, such as the investigation of rare cell types or primary cells (e.g. early development, where each of a small number of cells may have a distinct function) and the examination of subpopulations of cells from a larger heterogeneous population (e.g. discovering cell types in brain tissues).
Sandrine Dudoit will discuss some of the statistical analysis issues that have arisen in the context of a collaboration funded by the Brain Research through Advancing Innovative Neurotechnologies (BRAIN) Initiative, with the aim of classifying neuronal cells in the mouse somatosensory cortex.
These issues, ranging from so-called low-level to high-level analyses, include: exploratory data analysis (EDA) for quality assessment/control (QA/QC) of scRNA-Seq reads, normalization to account for nuisance technical effects, cluster analysis to identify novel cell types, and differential expression analysis to derive gene expression signatures for the cell types.

Lodewyk Wessels

Professor of Computational Cancer Biology,
The Netherlands Cancer Institute,
Amsterdam, the Netherlands

l.wessels@nki.nl

NOVEMBER 20

Lodewyk Wessels received his M.Sc. (1990) and Ph.D.(1997) both from the Department of Electronic and Computer Engineering, Universtity of Pretoria, South Africa. From 1993 to 1997 he was a member of the Center for Spoken Language Understanding at the Oregon Graduate School of Science and Technology, initially as a graduate student and later as a post-doctoral fellow. In 1997, he joined the Faculty of Electrical Engineering, Mathematics and Computer Science at the Delft University of Technology, initially as postdoc and later as assistant professor. In 1997, he became a faculty member and head of the Bioinformatics and Statistics group at the Netherlands Cancer Institute in Amsterdam, The Netherlands. He was appointed chair of Computational Cancer Biology at the Technical University in Delft in April 2012. The computational Cancer Biology group focuses on three research themes. In Theme 1 we develop computational approaches to map the molecular landscape of cancers by, for instance, identifying cancer driver genes. In Theme 2 we study how these drivers are regulated in pathways by building in silico models of signaling and metabolic pathways. Finally, in Theme 3, we develop approaches that exploit the driver landscape and knowledge of its regulatory landscape to design personalized treatments.

15h30

From the cancer landscape to personalized treatment

The exact mechanisms involved in tumor development and therapy response are still largely unclear. Here we report on two computational approaches we developed to systematically unravel these genetic interactions, based on high throughput datasets, and map these to drug response. In the first approach, we developed a kernel-based, scale space approach to detect molecular interactions (co-occurrences and mutually exclusivities) from somatic variant and copy number data.
We show that the number of aberrations per gene and per sample have a major influence on results and propose a null-model to correct these effects. We demonstrate the approach on the TCGA and METABRIC breast cancer cohorts. We identify no co-occurring aberrations but multiple mutually exclusive interactions that represent a global breast cancer interaction network, shedding new light on breast cancer development and subtyping.
We also perform a pan-cancer interaction analysis which reveals cancer-type independent interactions. Secondly, we propose a novel computational approach based on integer programming that infers logic combinations of discrete genetic events that predict the observed phenotype. The use of a logic formalism enables the formulation of intelligible models, facilitating speedy hypotheses generation. Here, we report on the application of this approach to the Wellcome Trust Sanger Institute 1000 cancer cell line panel. On this panel, genome-wide mutation status, copy number, gene expression and methylation profiles have been recorded. The response to 560 anti-cancer therapeutics have also been recorded. Our models show that for most drugs, combinations of mutations explain the drug response better than single mutations. For example, of the eight BRAF inhibitors in the panel, the drug sensitivity to three of them is better explained using a logic combination of a BRAF mutation and one or more mutations in other genes. These results immediately suggest putative drug combination therapies.

Jocelyn Laporte

Research director, French National Institute of Health and Medical Research (Inserm), U964, UMR7104, Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), Illkirch, France

jocelyn@igbmc.fr

NOVEMBER 20

Jocelyn Laporte is a research director at INSERM, in the department of Translational Medicine at IGBMC, Illkirch. His team focuses on understanding and treating rare and severe neuromuscular disorders, mainly congenital myopathies, caused by mutations in proteins regulating organelles trafficking and intracellular organization. While focusing on these genetic diseases, their approaches are multidisciplinary and encompass the identification of the implicated genes by high-throughput sequencing, the study of the molecular and cellular functions of these proteins in cells, the validation of faithful mammalian disease models, and preclinical proof-of-concept of potential therapies.

16h30

Massively parallel sequencing for gene identification and diagnosis in rare genetic diseases: the case of myopathies

Muscle represents about half of dried body weight and is essential for physiological homeostasis, metabolism and movement. More than 200 different myopathies have been described but about half of patients are still devoid of a genetic diagnosis, suggesting the implication of yet unlinked genes and inadequate routine molecular diagnosis strategies. To tackle these rare genetic diseases, we launched a national consortium ("Myocapture") and performed exome sequencing together with the Centre National de Genotypage on 1,000 patients and relatives to identify causative mutations. Variant filtering and ranking, and applying different genetic scenarios highlighted several cases: 1) some patients were linked to mutations in known myopathy genes, mainly in very large genes that are not routinely tested in diagnostic laboratories; 2) some cases were due to mutations in genes previously implicated in other muscle diseases, thus enlarging the phenotypic spectrum (allelic diseases); 3) several novel myopathy genes were identified. Indeed, a significant proportion of patients were not resolved and are still being scrutinized. We are also building an integrated knowledge base dedicated to myopathies at first, in order to provide analytical tools for sequencing analysis and better gene prioritization, and to provide an integrated view of pathways implicated in muscle homeostasis and function under normal and pathological conditions.

ORGANISERS

- **Genopole** is France's leading biotech- and biotherapy-dedicated science and business park. It brings together twenty research labs, university teaching facilities (at the Université d'Évry-Val-d'Essonne) and 81 biotech companies. Genopole's objective is to fund research in genomics, post-genomics and related sciences, transfer technology to industry, develop high-level teaching in these fields and promote the creation and development of biotech companies.
  www.genopole.fr

- **The Université d'Évry-Val-d'Essonne (UEVE)** is a major research hub in the southern Greater Paris area, with no less than 18 research laboratories and 10,000 students. UEVE is a multidisciplinary, career-oriented university. It is renowned for the diversity of its training offer, the quality of its research and particularly its excellence in the fields of genomics and post-genomics in healthcare and industry, an expertise exercised in close collaboration with Genopole, the CNRS, the CEA and others.
  www.univ-evry.fr

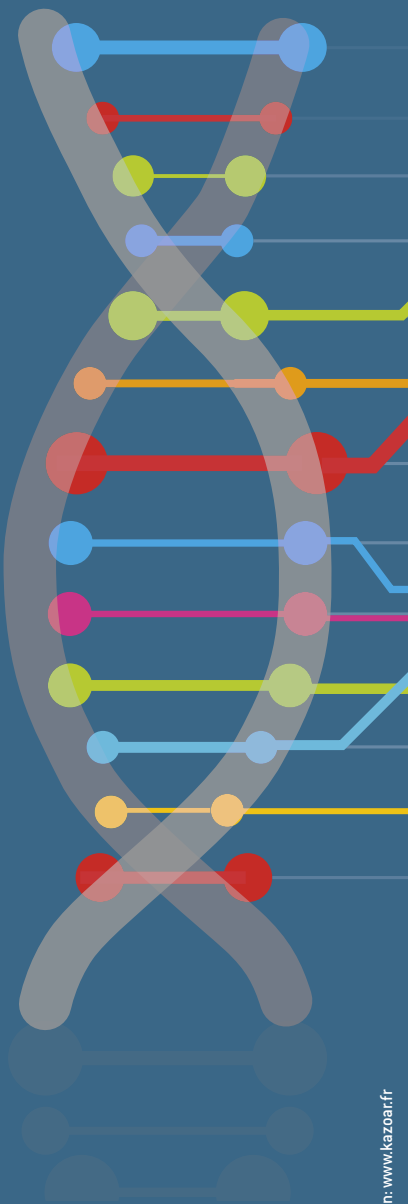
- **LaMME (Laboratory for Mathematics and Modeling)** contains 3 research teams: Analysis and partial derivative equations, Probabilities and financial mathematics, Statistics and genome.

  The **Statistics and Genome Team** focuses on the conception, development and mathematical analysis of original statistical methods for the analysis of biological data. These methods are made available to the scientific community via dedicated software or web interfaces. The team maintains strong interactions with biologists to ensure the relevance of these developments. Conversely, these interactions trigger the study of new, challenging mathematical questions.
  www.math-evry.cnrs.fr

SUPERVISING ORGANIZATIONS

www.cnrs.fr
www.ensiie.fr
www.inra.fr

PARTNERS



http://biomaths.genopole.fr

Design: www.kazoar.fr