13th European Conference on Computational Biology

PROGRAM BOOK

ECCB 14

STRASBOURG · FRANCE

7-10 SEPTEMBER 2014

www.eccb14.org

Welcome

On behalf of the ECCB'14 Organizing and Steering Committees we are very happy to welcome you to Strasbourg, Heart of Europe. We hope that you will enjoy the many facets of the conference: keynote lectures, oral communications, workshops and tutorials, demos, posters, booths. This dense and attractive program is intended to be the substrate of fruitful discussions and networking among all of you.

We are proud to welcome seven distinguished Keynote Speakers: Nobel Prize laureate Jean-Marie Lehn (Strasbourg University, Nobel Prize in Chemistry in 1987), Patrick Aloy (Institute for Research in Biomedicine, Barcelona, Spain), Alice McHardy (Heinrich Heine University, Düsseldorf and Helmholtz Center for Infection Biology, Braunschweig, Germany), Nada Lavrač (Jožef Stefan Institute and University of Nova Gorica, Slovenia), Ewan Birney (European Bioinformatics Institute, Hinxton, United Kingdom), Doron Lancet (The Weizmann Institute of Science, Rehovot, Israel) and Eric Westhof (Strasbourg University, France).

A large variety of Workshops, Tutorials and Satellite Meetings will take place on Saturday, September 6 and Sunday, September 7 to kick off the conference!

We would like to stress the interest in the conference by the Council of Europe, whose headquarters are in Strasbourg. An invited talk by Laurence Lwoff, Head of the Bioethics Unit, is scheduled during the opening ceremony, in which she will share the Council of Europe's concerns about the bio-ethical challenges raised by the usage of biobanks and biomedical data in research and its applications. The Council of Europe will also host the Gala Evening on Tuesday in its nice reception hall and gardens along the Ill river.

Great thanks are due to all people who made ECCB'14 happen, including the ISCB and our commercial and academic sponsors who allowed us to distribute 67 travel fellowships for early-stage scientists.

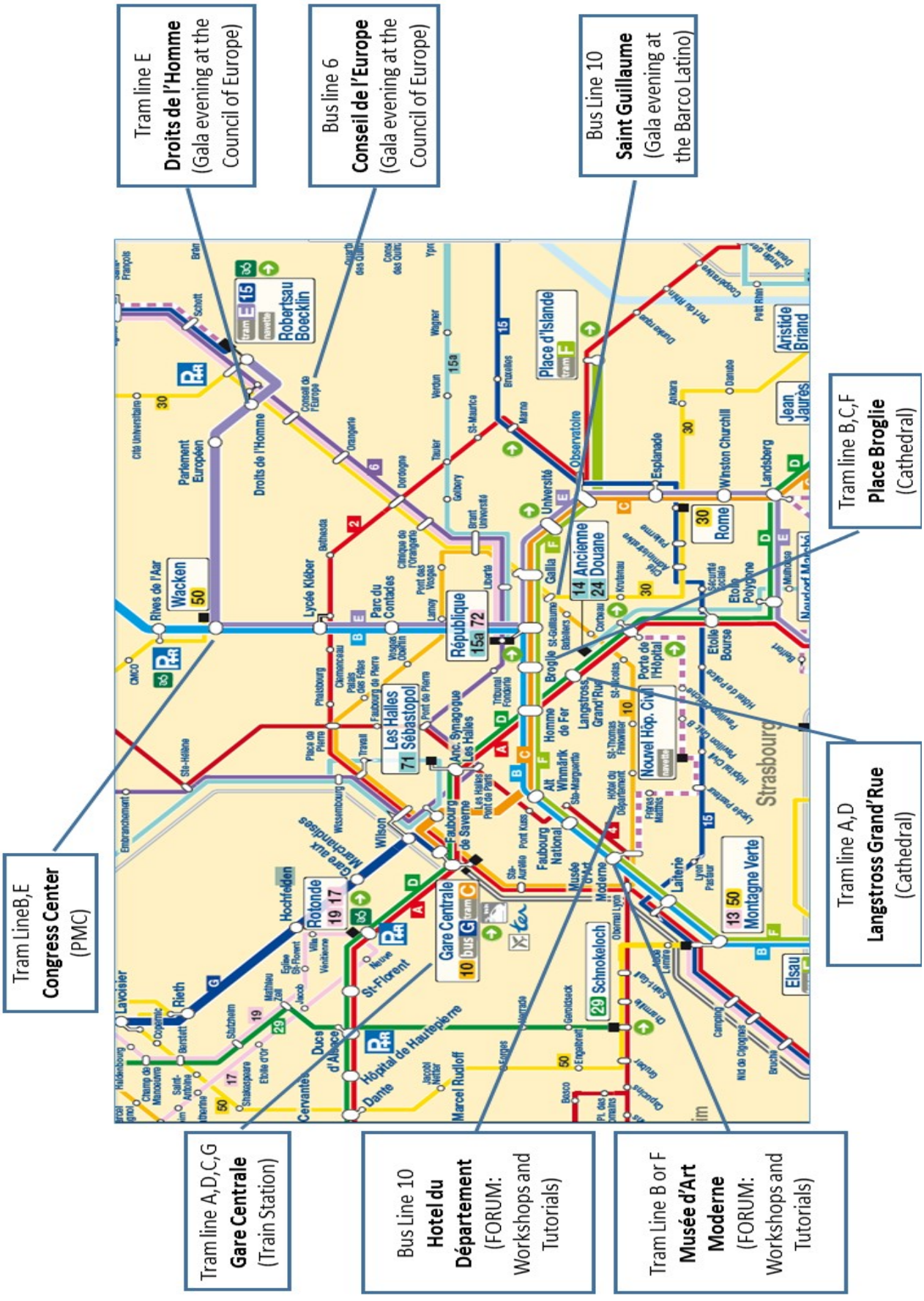
Thank you all for being there: each of you will bring something to the conference! We wish you an enjoyable ECCB'14 in Strasbourg!

Marie-Dominique Devignes and Yves Moreau

Conference chair and co-chair

Table of Contents

Map


Tram line E
Droits de l'Homme
(Gala evening at the Council of Europe)
Bus line 6
Conseil de l'Europe
(Gala evening at the Council of Europe)
Bus Line 10
Saint Guillaume
(Gala evening at the Barco Latino)
Tram line B,C,F
Place Broglie
(Cathedral)
Tram line A,D
Langstross Grand'Rue
(Cathedral)
Tram LineB,E
Congress Center
(PMC)
Tram line A,D,C,G
Gare Centrale
(Train Station)
Bus Line 10
Hotel du Département
(FORUM: Workshops and Tutorials)
Tram Line B or F
Musée d'Art Moderne
(FORUM: Workshops and Tutorials)
Strasbourg

Public Transport and Site Seeing

Pass 3-days for tramway and bus

Every participant has received a 3-day pass for Strasbourg public transport.

You can consult Strasbourg Public Transport maps on-line on http://www.cts-strasbourg.eu/en/

Itineraries

1. To reach the Forum at the Faculté de Médecine

- by bus : take line 10, stop at "Hotel du département".

- by tramway : stop at "Musée d'Art Moderne", take the bridge, cross the road and turn left.

Then follow the ECCB signs on the road

2. To reach the PMC (Palais de la Musique et des Congrès - Convention Centre)

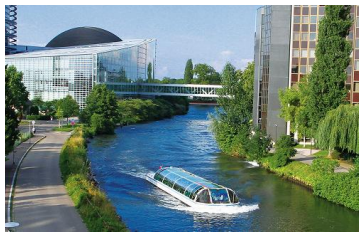
- by tramway line B or C, stop at "Wacken".

Then follow the ECCB signs on the road

3. Gala evening at the Council of Europe

The Gala at the Council of Europe includes a boat trip and the dinner at the Council.

In your congress bag, you will find a boat ticket with a departure time: PLEASE RESPECT THE DEPARTURE TIME WRITTEN ON YOUR TICKET.



The boat will take you to the Council of Europe after a small trip around Strasbourg.

The boat pier is located 150 m from the Cathedral: having the cathedral in front of you, go to the right (rue du Maroquin) down to the river. (See also map p10).

Information on batorama http://www.batorama.com/

In case you want to go to the Council of Europe by your own, take bus line 6 stop at "Conseil de l'Europe" or tramway line E stop at "Droits de l'Homme".

4. Gala evening at the Ancienne Douane

The Gala at the Ancienne Douane includes a mini-train trip and the dinner at a famous typical Alsacian restaurant. In the congress bag you received when registering, you will find a mini-train ticket with departure time that will take you for a small visit trip around Strasbourg: PLEASE RESPECT DEPARTURE TIME WRITTEN ON YOUR TICKET.

The mini-train station is located "place Gutemberg", in front of the Cathedrale (See also map p10).

Information on mini-train http://www.cts-strasbourg.eu/en/getting-around/tourism/

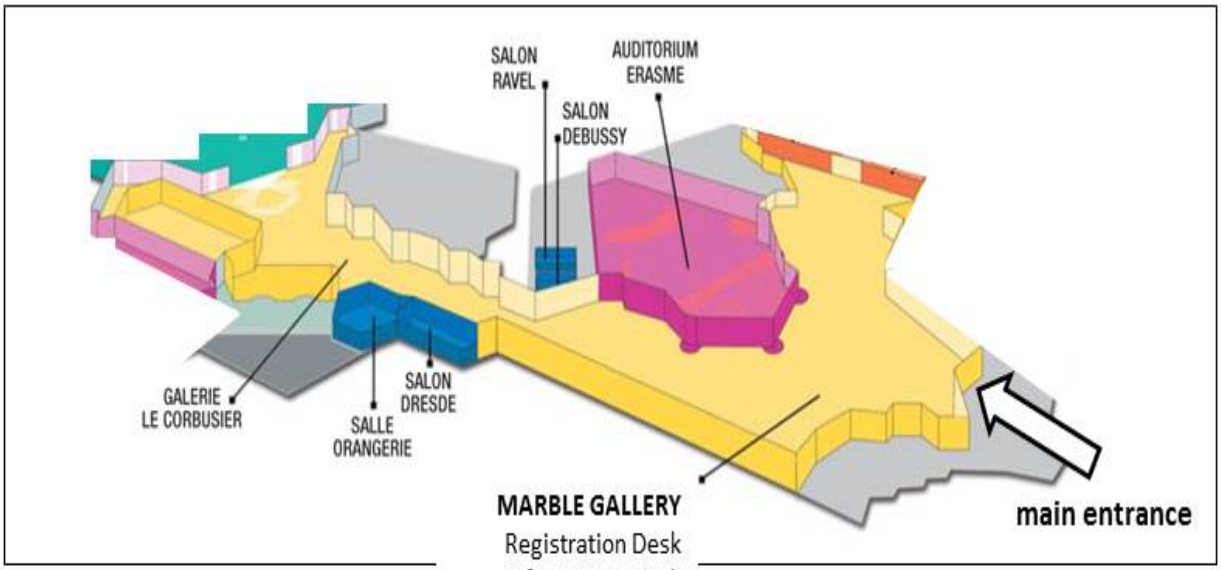
Site Seeing in Strasbourg and Nancy

1. Strasbourg Tourism Office: 17 place de la Cathédrale, Strasbourg ; Tél. +33 (0)3 88 52 28 28 http://www.otstrasbourg.fr/en/

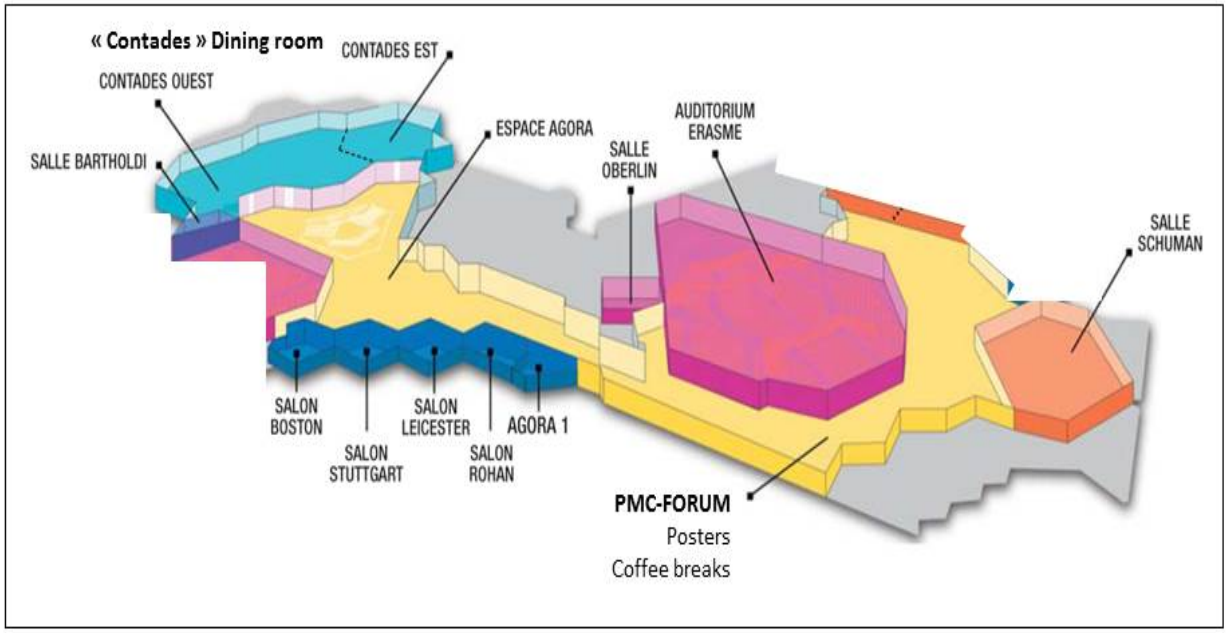
2. Nancy Tourism Office: Place Stanislas, Nancy ; Tel. +33 (0)3 83 35 22 41 http://en.nancy-tourisme.fr/
   Nancy is 90 min away from Strasbourg by train.

Floor Plan

Palais de la Musique et des Congrès (PMC) : Main Floor



Palais de la Musique et des Congrès (PMC) : 1st Floor



Organization

Conference Chairs

Marie-Dominique Devignes, CNRS-University of Lorraine, Nancy, France
Yves Moreau, University of Leuven, Belgium

Local Organizing committee

Mario Albrecht, Graz University of Technology, Austria
Francisco Azuaje, CRP-Santé, Luxembourg
Adrien Coulet, University of Lorraine, Nancy, France
Manuel Dauchez, University of Reims, France
Sophie Schbath, INRA, Jouy-en-Josas, France
Thomas Lengauer, Max Planck Institute for Informatics, Saarbrücken, Germany
Hans-Peter Lenhof, Center for Bioinformatics, Saarland University, Germany
Magali Michaut, Netherlands Cancer Institute, Amsterdam, The Netherlands
Emmanuelle Morin, INRA, Nancy – Champenoux, France
Luc Moulinier, CNRS, Strasbourg, France
Olivier Poch, CNRS, Strasbourg, France
Dave Ritchie, Inria Nancy-Grand Est, France
Marie-France Sagot, Inria Grenoble Rhône-Alpes, Lyon, France
Malika Smaïl-Tabbone, University of Lorraine, Nancy, France
Julie Thompson, CNRS, Strasbourg, France

ECCB Steering Committee

Ron Appel, Swiss Institute of Bioinformatics, Switzerland
Søren Brunak, Technical University of Denmark and University of Copenhagen, Denmark
Jaap Heringa, Vrije Universiteit, Amsterdam, The Netherlands
Oliver Kohlbacher, University of Tübingen, Germany
Michal Linial, The Hebrew University of Jerusalem, Israel
Rodrigo Lopez, European Bioinformatics Institute, Hinxton, United Kingdom
Yves Moreau, University of Leuven, Belgium
Marie-France Sagot, Inria Grenoble Rhône-Alpes, Lyon, France
Torsten Schwede, SIB and Biozentrum, University of Basel, Switzerland
Anna Tramontano (chair), University of Rome "La Sapienza", Italy
Alfonso Valencia, Centro Nacional de Investigaciones Oncologicas, Madrid, Spain
Jaak Vilo, University of Tartu, Estonia
Martin Vingron, Max Planck Institute for Molecular Genetics, Berlin, Germany
Janet Thornton, European Bioinformatics Institute, Hinxton, United Kingdom

Program Committee – Area Chairs

A. Sequencing and sequence analysis for genomics

Cedric Notredame, Centre for Genomic Regulation, Barcelona, Spain
Eric Rivals, Laboratory of Informatics, Robotics and Microelectronics, Montpellier, France

B. Gene expression

Jaak Vilo, University of Tartu, Estonia
Quaid Morris, University of Toronto, Canada

C. Pathways and molecular networks

Ralf Zimmer, Ludwig-Maximilians University Munich, Germany
Christian von Mering, University of Zürich and SIB, Switzerland

D. Computational systems biology

Thomas Sauter, University of Luxembourg, Luxembourg
Oliver Kohlbacher, University of Tübingen, Germany

E. Structural bioinformatics

Anna Tramontano, University of Rome "La Sapienza", Italy

Torsten Schwede, Swiss Institute of Bioinformatics and Biozentrum, University of Basel, Switzerland

F. Evolution and population genomics

Nicolas Galtier, CNRS, University Montpellier, France
Toni Gabaldon, Centre for Genomic Regulation, Barcelona, Spain

G. Bioinformatics of health and disease

Niko Beerenwinkel, Swiss Institute of Bioinformatics and ETH Zürich, Basel, Switzerland
Lodewyk Wessels, Netherlands Cancer Institute, The Netherlands

H. Biological knowledge discovery from data, texts and bio-images

Dietrich Rebholz-Schuhmann, University of Zürich, Switzerland
Michael Krauthammer, Yale University, USA

J. Methods and technologies for computational biology

Gert Vriend, Centre for Molecular and Biomolecular Informatics, Nijmeguen, The Netherlands
Christophe Blanchet, CNRS-University Lyon, France

Workshops and Tutorials

Mario Albrecht, Graz University of Technology, Austria
Olivier Poch, CNRS, Strasbourg, France

Industrial and Demo Track

Emmanuelle Morin, INRA, Nancy – Champenoux, France
Dave Ritchie, Inria Nancy-Grand Est, France

Posters

Francisco Azuaje, CRP-Santé, Luxembourg
Malika Smaïl-Tabbone, University of Lorraine, Nancy, France

Program Committee – Reviewers

Paste here the list of reviewers

Mohamed Abouelhoda
Tyler Alioto
Andre Altmann
Simon Anders
Kazuharu Arakawa
Saul Ares
Rolf Backofen
Pedro Ballester
Matteo Barberis
Geoff Barton
Thomas Bataillon
Michael A. Beer
Tim Beissbarth
Asa Ben-Hur
Nir Ben-Tal
Panayiotis Takis Benos
Bonnie Berger
Rafael Berlanga Llavori
Simon Bernèche
Inanc Birol
Judy Blake
Mathieu Blanchette
Brigitte Boeckmann
Paola Bonizzoni
Karsten Borgwardt
Céline Brochier-Armanet
Yana Bromberg
Andrew Brown
Javier Buceta
Alfonso Buil
Janusz Bujnicki
Eric Bullinger
Hauke Busch
Charlotte Deane
Alain Denise
Christophe Dessimoz
Eric W Deutsch
Diego Di Bernardo
Andreas Doncic
Finn Drablos
Ivana Drobnjak
Laurent Duret
Julien Dutheil
Martin Ebeling
Michael Ederer
Richard Edwards
Arne Elofsson
Laurent Falquet
Thilo Figge
Ronan Fleming
Samah Fodeh
Franca Fraternali
Caroline C. Friedel
Dmitrij Frishman
Julien Gagneur
Christine Gaspin
Tomas Gedeon
Olivier Gevaert
Jean-Francois Gibrat
Robert Giegerich
Alejandro Giorgetti
Georgios Gkoutos
Giorgio Gonnella
Liliana Greger
Andre Gruning
Raphael Guerois
Inge Jonassen
Klaus Jung
Senay Kafkas
Olga Kalinina
Michael Kane
Kazutaka Katoh
Lawrence Kelley
Patrick Kemmeren
Paul Kersey
Jin-Dong Kim
Jung-Jae Kim
Philip Kim
Martin Kircher
Gunnar W. Klau
Carole Knibbe
Raivo Kolde
Rachel Kolodny
Sergei Kosakovsky-Pond
Dima Kozakov
Roland Krause
Oscar Krijgsman
Andriy Kryshtafovych
Tobias Kuhn
Matthieu Legendre
Hans-Peter Lehnhof
Florian Leitner
Ulf Leser
Mingyao Li
Yao-Cheng Lin
Michal Linial
Rodrigo Lopez
Phillip Lord
Steven Ma
Robert Preissner
Jaime Prilusky
Miguel Angel Pujana
Tal Pupko
Sven Rahmann
Sanguthevar Rajasekaran
Gunnar Ratsch
Magnus Rattray
Knut Reinert
Bernhard Renard
Martin Renqiangmin
Paolo Ribeca
Fabio Rinaldi
Dave Ritchie
Mark Robinson
Peter Robinson
Marc Robinson-Rechavi
David Rocke
Stephane Rombauts
Jacques Rougemont
Julio Rozas
Ingo Ruczinski
Sten Rüdiger
Andrey Rzhetsky
Nicolas Salamin
Michael Sammeth
Matthias Samwald
Susanna-Assunta Sansone
Ugis Sarkans
Ivo Sbalzarini
Joerg Schaber
Alejandro Schäffer
Avner Schlessinger

Harmen Bussemaker
Carlos Camacho
Anne-Claude Camproux
Sander Canisius
Alessandra Carbone
Robert Castelo
Rui Chen
Hélène Chiapello
Lachlan Coin
Nigel Collier
Olivier Collin
Francisco Couto
Domenico Cozzetto
Forrest Crawford
Samuel Croset
Gergely Csaba
Attila Csikasz-Nagy
Florence D'Alché-Buc
Tal Dagan
Antoine Danchin
Susmita Datta
Vincent Daubin
Hélène Dauchel
Xavier De La Cruz
Dick De Ridder
Jeroen De Ridder
Jerzy Tiuryn
Jeffrey Townsend
Achim Tresch
Jerry Tsai
Pierre Tuffery
Alfonso Valencia
Pascal Vallotton
Roeland van Ham
Juergen Haas
Janna Hastings
Dieter Heermann
Manuela Helmer-Citterich
Jean-Karim Heriche
Henning Hermjakob
Carl Herrmann
Des Higgins
Sebastian Hiller
Asger Hobolth
Robert Hoehndorf
Ivo Hofacker
Daniel Hoffmann
Steve Hoffmann
Liisa Holm
Rob Hooft
Simon Hubbard
Wolfgang Huber
Philippe Hupé
Mark Ibberson
Dagmar Iber
Francesco Iorio
Christian Iseli
Samira Jaeger
Jeff Jensen
Antonio Jimeno Yepes
Jacques van Helden
David Vandijk
Česlovas Venclovas
Jean-Philippe Vert
Stéphane Vialette
Martin Vingron
Olga Vitek
Arndt Von Haeseler
Brian Marsden
Lennart Martens
Marc A. Marti-Renom
Liam Mcguffin
Aoife McLysaght
Denis Menshykau
George Michailidis
Stephen Montgomery
Burkhard Morgenstern
Sara Mostafavi
Matthieu Muffato
Mate Nagy
Markus Nebel
Claire Nédellec
Goran Nenadic
Zoran Nikoloski
Zemin Ning
Balazs Papp
Jong C. Park
Helen Parkinson
Leopold Parts
Sabine Peres
Kjell Petersen
Hervé Philippe
Yann Ponty
Natapol Pornputtapong
Jerome Waldispuhl
Xuefeng Wang
John Welch
Simon Whelan
Shoshana Wodak
Verena Wolf
Haim J. Wolfson
Ioannis Xenarios
Andreas Schlicker
Alexander Schliep
Birgit Schoeberl
Michael Schroeder
Joerg Schultz
Marcel Schulz
Russell Schwartz
Celine Scornavacca
Joachim Selbig
Nigam Shah
Noam Shomron
Barry Smith
Johannes Soeding
Rainer Spang
Henning Stahlberg
Alexandros Stamatakis
Alexander Statnikov
Oliver Stegle
Joerg Stelling
Jens Stoye
Wing-Kin Sung
Erdogan Taskesen
Denis Thieffry
Paul Thomas
Peter Tieleman
Jens Timmer
Elena Zaslavsky
Daniel Zerbino
Louxin Zhang
Xiuwei Zhang
Andrei Zinovyev
Scott Zuyderduyn

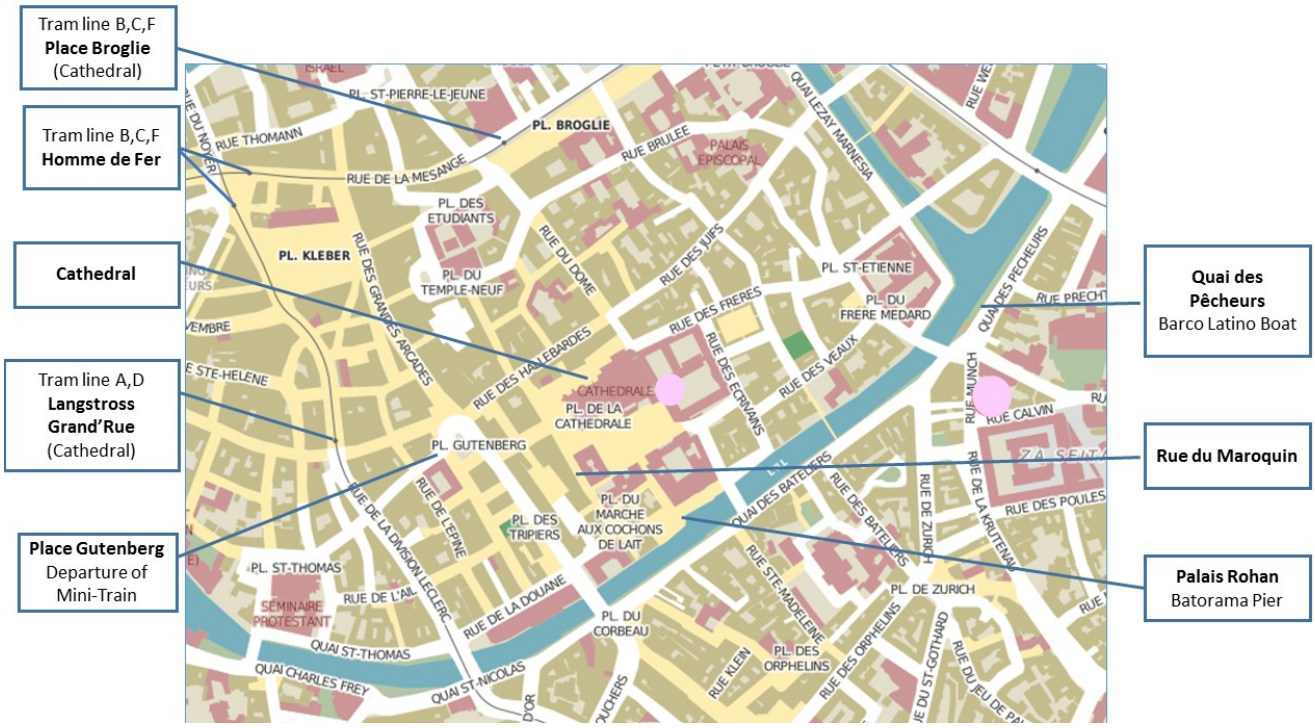
Program Committee – Co-reviewers

Nanne Aben
Amin Allahyar
Roland Arnold
Gabriele Ausiello
Sepideh Babaei
Mukesh Bansal
Pasha Baranov
Michaela Bayerlova
Vincenzo Belcastro
Evi Berchtold
Richard Bonneau
Andreas Bremges
Carlo Cannistraci
Annamaria Carissimo
Rita Casadio
Raffaele Cerulli
Jun Chen
Maria Chikina
Salim Chowdhury
Claire Colas
Fabrizio Costa
Luisa Cutillo
Neil Dalchau
Simone Daminelli
Arijit Das
Jérôme Dauvillier
Slavica Dimitrieva
Zhihao Ding
Norbert Dojer
Sebastian Dümcke
Ionas Erb
Florian Erhard
Corinna Ernst
Bernd Fischer
Patrick Fried
Martin Frith
Michael Garton
Ludwig Geistlinger
Vidhya Gnanasekaran
Luca Grieco
Alexey Gritsenko
Mario Guarracino
Stefan Haas
Kerstin Haase
Monika Heiner
Matthias Heinig
Shirley Hui
Dirk Husmeier
Mudassar Iqbal
Zerrin Isik
Jon Ison
Tobias Jakobi
Aleksander Jankowski
Jouhyun Jeon
Giuseppe Jurman
Darius Kazlauskas
Bernd Klaus
Sabrina Krakau
Frank Kramer
Gregory Kucherov
Ádám Kun
Anne Kupczok
Konstantina Kyriakouli
Elodie Laine
Henry Lam
Giddy Landan
Anastasia Lapytsko
Mario Lauria
Sangkyun Lee
Jingjing Li
Yu Lin
Philippe Lopez
Pier Luigi Martelli
Sjoerd M. H. Huisman
Sisi Ma
Cedrik Magis
Parag Mallick
Santiago Marco-Sola
Michael Menden
Laurence Menu-Bouaouiche
Sheinmann Misha
Andrzej Mizera
Arnau Montagud
Francesco Napolitano
Steffen Neumann
Marius Nicolae
Kliment Olechnovič
Yaara Oren
Christina Otto
Roberto Pagliarini
Luca Parca
Juliane Perner
Robert Pesch
Tobias Petri
Emma Pierson
Andrea Raiconi
Soheil Rastgou Talemi
Bisakha Ray
Hugues Richard
Philippe Rocca-Serra
Janine Roy
Daniel Runcie
Subrata Saha
Michael Schubert
Itamar Sela
Vassili Soumelis
David Spieler
Daniel Stöckel
Mark Sun
Zuojian Tang
Andrea Tanzer
Franziska Taruttis
Stefanie Tauber
Kęstutis Timinskas
Panuwat Trairatphisan
Timothy Travers
Armand Valsesia
Martijn van de Bunt
Gerard van Westen
Jean-Stéphane Varré
Jannik Vollmer
Astrid Wachter
Inken Wohlers
Feifei Xiao
Daniel Zerzion
Qing Zhao

Poster reviewers

Mario Albrecht
Theodore Alexandrov
Ayse Özlem Aykut
Francisco Azuaje
Nicolas Beaume
Nir Ben-Tal
Christophe Blanchet
Melanie Börries
Fiona Browne
Eric Bullinger
Anyela Camargo
Salvador Capella-Gutierrez
Hannah Carter
Frederic Cazals
Juan Miguel Cejuela
Sarah Cohen-Boulakia
Ana Conesa
Adrien Coulet
Olivier Dameron
Tjaart de Beer
Niek de Klein
Ana Dopazo
Rolf Findeisen
Anisah Ghoorah
Tatyana Goldberg
Hulda Haraldsdóttir
Michaël Heymann
Francesco Iorio
Edda Kloppmann
Dominique Lavenier
Natacha P. Leitão
Weizhong Li
Valentin Loux
Geoff Macintyre
Mahendra Mariadassou
Loredana Martignetti
Magali Michaut
Fouzia Moussouni-Marzolf
Isabel Nepomuceno-Chamorro
Cinta Pegueroles
Maude Pupin
Fabio Rinaldi
Anne Siegel
Amit Singh
Christine Sinoquet
Malika Smail-Tabbone
Nurgazy Sulaimanov
Helene Touzet
Federico Vaggi
Paul Walsh
Haiying Wang
Lucas Wetzel
Anke Wienecke-Baldacchino
Huiru Zheng

Strasbourrg Center: Departure for River Boat or Mini-Train Trips (Tuesday afternoon)



Sponsors and Exhibitors

Silver sponsors

BIOBASE — Booth 4

BIOBASE is the leading provider of expert-curated biological databases, software and services for the life sciences. Our products and services identify relations critical to drug and biomarker discovery as well as improve biomedical research by transforming data into scientific concepts.

www.biobase-international.com

Demo ID06
Monday 14:00-14:30
Tuesday 14:00-14:30
Room Schumann

GPB : Genomics, Proteomics and Bioinformatics — No booth

Genomics, Proteomics and Bioinformatics (GPB) is a peer-reviewed and fast-track open access journal that focuses on disseminating the newest discoveries in the -omics and bioinformatics studies.

www.journals.elsevier.com

Demo ID18
Tuesday 13:30-14:00
Tuesday 14:00-14:30
Salon Stuttgart

KoriScale — Booth 5

KoriScale is an Inria Innovation Lab that brings together Korilog SME and GenScale Inria project team. Korilog is a bioinformatics company specialized in genomic sequence comparison software tools. GenScale activities focus on methodological research targeting hight throughput genomic data processing. The Klast software (fast genomic databank comparison) is currently developed by KoriScale.

koriscale.inria.fr/

Demo ID08
Monday 14:00-14:30
Tuesday 13:30-14:00
Salon Boston

sbvImprover — Booth 14

sbv IMPROVER stands for systems biology verification and Industrial Methodology for Process Verification in Research. It is a robust methodology that verifies systems biology approaches using double-blind performance assessment and applies the wisdom of crowds to solve scientific challenges. The Network Verification Challenge (NVC) aims to verify and enhance existing biological network models using an innovative reputation-based crowd-sourcing approach for biological network curation and provides the opportunity to review, challenge and enhance network models of biological processes. The NVC is expected to increase the networks' value and promote their use in research applications such as drug discovery and personalized medicine.

sbvimprover.com

Demo ID07
Monday 13:30-14:00
Tuesday 14:00-14:30
Salon Boston

TotaLinuX — Booth 11

Created in 2004, TotaLinuX is a french compagny specialized in HPC and Big Data. Its core business is to design, install and distribute IT solutions to scientifical use.
Totalinux main skills are focused on designing and building IT infrastructure dedicated to computing, storage and graphical simulation.
TotaLinuX is also selling nearly 100 complex IT systems and the associated storage solutions a year. Regarding HPC, TotaLinuX is a specialist recognized for interfacing and managing computing clusters at very high speed (with InfiniBand Interconnect), and use of load balancing software. Our company also provides storage solutions to enable the preservation of data with new solutions such as BigData.

www.totalinux.fr

Demo ID13
Monday 13:30-14:00
Monday 14:00-14:30
Salon Stuttgart

Other Exhibitors

Active Motif — Booth 3

Active Motif is the industry leader in developing and delivering innovative tools to enable epigenetics and gene regulation research. We are committed to providing the highest quality products and superior service & support to serve the life science, clinical and pharmaceutical/drug discovery communities.
http://www.activemotif.com/

Cambridge University Press

Cambridge University Press is part of the University of Cambridge.
We further the University's mission by disseminating knowledge in the pursuit of education, learning and research at the highest international levels of excellence.
Find out more about us on our web site

www.cambridge.org/about-us/who-we-are

Booth 7

CRC Press

CRC Press, part of the Taylor and Francis Group, is the premier publisher of textbooks, reference books, and ebooks on computational biology and bioinformatics. Stop by our booth to view our latest titles and take advantage of our conference discount. If you are interested in writing a book please stop by the booth to speak with Sunil Nair about your idea.

www.crcpress.com/textbooks

Booth 8

EGI: European Grid Infrastructure

The European Grid Infrastructure (EGI) is a publicly funded e-infrastructure that gives scientists access to more than 370,000 logical CPUs and 170 PB of disk capacity to drive research and innovation in Europe. Its mission is to connect researchers from all disciplines with the reliable and innovative ICT services they need for their collaborative world-class research.
EGI offers computing capacity to speed up research, cloud capabilities and ready-to-use scientific software via the EGI Applications Database.
EGI's computing resources are free at point of use and allocation is not dependant on proposals.

www.egi.eu

Booth 1

EMBL-EBI : European Bioinformatics Institute

The European Bioinformatics Institute is part of EMBL, Europe's flagship laboratory for the life sciences. EMBL-EBI provides freely available data from life science experiments covering the spectrum of molecular biology. In addition to bioinformatics services, EMBL-EBI carries out bioinformatics research and has an extensive training programme. EMBL-EBI is a non-profit, intergovernmental organisation. The 500 staff represent 55 nationalities, and there are visiting scientists throughout the year. EMBL-EBI is located on the Wellcome Trust Genome Campus in Hinxton, Cambridge in the UK.

www.ebi.ac.uk/

Booth 9

ISCB Student Council

The ISCB Student Council (SC) is the student organization of the International Society for Computational Biology, whose members come from all around the world and share a passion for bioinformatics and computational biology. The mission of the SC is to promote the development of the next generation of computational biologists. This is achieved through the provision of scientific events, networking opportunities, soft-skills training, educational resources and career advice, while attempting to influence policy processes affecting science and education. The core activities carried out by the SC include the organization of several yearly symposia, the Regional Student Groups initiative and the Internships Program.

www.iscbsc.org/.

Booth 6

See also satellite meeting S01 - ESCS

ISCB : International Society for Computational Biology

The International Society for Computational Biology (ISCB) serves over 3,000 members from more than 70 countries by addressing scientific policies, providing access to high quality publications, organizing meetings, and serving as a portal to information about training, education, employment and news from related fields. The ISCB hosts annual meetings, including the ISMB, the world's longest running and largest bioinformatics conference (held jointly with the ECCB every other year in Europe). The ISCB also affiliates with several other significant meetings of our science, has two official journals of the highest impact factors in the Mathematical & Computational Biology category, and has affiliations in place with several other publications for the benefit of our members.
Now hosted at the San Diego Supercomputer Center at University of California, San Diego, the Society was officially formed in 1997 as an outgrowth of the International Conference on Intelligent Systems for Molecular Biology (ISMB). From humble beginnings, both the ISCB's membership and the ISMB's annual attendance have kept pace with the overall growth experienced in the field of bioinformatics/computational biology.

www.iscb.org/iscb-aboutus

Booth 10

Oxford University Press

Oxford University Press is a department of the University of Oxford. It furthers the University's objective of excellence in research, scholarship, and education by publishing worldwide.

http://global.oup.com/uk/about/

Booth 2

SIB : Swiss Institute of Bioinformatics

The SIB Swiss Institute of Bioinformatics is a leading player on the international bioinformatics scene. We have a long-standing tradition of producing state-of-the-art software and carefully annotated databases for the life science research community. UniProtKB/Swiss-Prot, SWISS-MODEL, Bgee, ViralZone and STRING are just a few of the more than 150 resources available on www.expasy.org.
Visit our booth and pre-register for our [BC]^2 Basel Computational Biology Conference! This SIB international conference will take place from 8 to 10 June 2015. Be among the first to discover its new format while savouring delicious Swiss specialities.

www.isb-sib.ch/

Booth 12

Springer

Springer is a leading global scientific, technical and medical publisher, providing researchers with quality content via innovative information products and services. Our eBook Collection contains more than 160,000 titles, adding 8,400 new titles per year. While we publish more than 2,500 journals in total, we are also home to the largest open access portfolio worldwide, with over 420 open access journals. And we are proud of our publishing partnership with more than 500 scientific societies.

www.springer.com

Booth 13

Sponsors without Booth

CGE	Canceropole Grand-Est	**www.canceropole-ge.org**
CNRS	Centre National de la Recherche Scientifique	**www.cnrs.fr**
Council of Europe	Human rights organisation	**hub.coe.int/en/**
GDR BIM	Group of Research in Molecular Bioinformatics	**www.gdr-bim.cnrs.fr/**
Harmonic Pharma	Start-up: 'Adding therapeutic value'	**harmonicpharma.com**
IFB	French Institute of Bioinformatics	**www.france-bioinformatique.fr/**
INRA	The French national institute for agricultural research	**www.inra.fr/en/**
Inria	A public science and technology institution dedicated to computational sciences	**www.inria.fr/en/**
Loria	Lorraine Laboratory of Research in Computer Science and its Applications	**www.loria.fr**
SFBI	French Society of Bioinformatics	**www.sfbi.fr**
University of Lorraine		**www.univ-lorraine.fr**
University of Strasbourg		**www.unistra.fr**

Fellowship Sponsors

ISCB	International Society for Computational Biology	**www.iscb.org**
SFBI	French Society of Bioinformatics	**www.sfbi.fr**
BITS	Italian Society of Bioinformatics	**www.bioinformatics.it**

Silver Sponsors

Other Exhibitors

Other sponsors

Keynotes

Perspectives in Chemistry: From Supramolecular Chemistry towards Adaptive Chemistry

Jean-Marie Lehn

University of Strasbourg, Strasbourg, France. Nobel Laureate in Chemistry (1987)

Abstract: Supramolecular chemistry explores systems undergoing *self-organization*, capable of generating well-defined functional supramolecular architectures by *molecular information controlled* self-assembly from their components, thus behaving as *programmed chemical systems*.

Supramolecular chemistry is intrinsically a *dynamic chemistry* in view of the lability of the interactions connecting the molecular components of a supramolecular entity and the resulting ability of supramolecular species to exchange their components. The same holds for molecular chemistry when the molecular entity contains covalent bonds that may form and break reversibility, so as to allow a continuous change in constitution by reorganization and exchange of building blocks. These features define a *Constitutional Dynamic Chemistry* (CDC) covering both the molecular and supramolecular levels.

CDC introduces a paradigm shift with respect to constitutionally static chemistry. It takes advantage of dynamic diversity to allow variation and selection and operates on dynamic constitutional diversity in response to either internal or external factors to achieve *adaptation*.

CDC generates networks of dynamically interconverting constituents, *constitutional dynamic networks*, presenting *agonistic* and *antagonistic* relationships between their constituents, that may respond to perturbations by physical stimuli or to chemical effectors.

The implementation of these concepts points to the emergence of *adaptive* and *evolutive chemistry*, towards systems of increasing complexity.

Bio: *Jean-Marie Lehn was born in Rosheim, France in 1939. In 1970 he became Professor of Chemistry at the Université Louis Pasteur in Strasbourg and from 1979 to 2010 he was Professor at the Collège de France in Paris. He is presently Professor Emeritus at the University of Strasbourg. He shared the Nobel Prize in Chemistry in 1987 for his studies on the chemical basis of "molecular recognition" (i.e. the way in which a receptor molecule recognizes and selectively binds a substrate), which also plays a fundamental role in biological processes. Over the years his work led him to the definition of a new field of chemistry, which he has proposed calling "supramolecular chemistry" as it deals with the complex entities formed by the association of two or more chemical species held together by non-covalent intermolecular forces, whereas molecular chemistry concerns the entities constructed from atoms linked by covalent bonds. Subsequently, the area developed into the chemistry of "self-organization" processes and more recently towards "adaptive chemistry", dynamic networks and complex systems.*

Author of more than 900 scientific publications, Lehn is a member of many academies and institutions. He has received numerous international honours and awards.

A network biology approach to novel therapeutic strategies

Patrick Aloy

Institute for Research in Biomedicine, Barcelona, Spain. ICREA Research Professor

Abstract: High-throughput interaction discovery initiatives are providing thousands of novel protein interactions, which are unveiling many unexpected links between apparently unrelated biological processes. In particular, analyses of the first draft human interactomes highlight a strong association between protein network connectivity and disease. Indeed, recent exciting studies have exploited the information contained within protein networks to disclose some of the molecular mechanisms underlying complex pathological processes. These findings suggest that both protein-protein interactions and the networks themselves could emerge as a new class of targetable entities, boosting the quest for novel therapeutic strategies. In this talk, I will summarize our work towards the characterization and modelling of the protein-interaction network underlying Alzheimer´s disease and breast cancer, together with our most recent attempts to decipher complex cell networks to the point of being able to predict how the perturbation of a node might affect the system as a whole.

Bio: *Born in Barcelona in 1972. He obtained his BSc in Biochemistry (1996), MSc in Biotechnology (1998) and PhD in Biochemistry & Molecular Biology (2000) from the Autonomous University of Barcelona. He then joined the Structural Bioinformatics Group at the European Molecular Biology Laboratory (2001) where he did his postdoctoral training. Since April 2006, Patrick Aloy is an ICREA Research Professor and Principal Investigator of the Structural Bioinformatics Lab in the Institute for Research in Biomedicine (IRB Barcelona). For fifteen years, Dr Aloy has been developing and implementing new technologies and algorithms, applying state-of-the-art methods to specific problems and bridging the gap between theoretical models and experiments in different disciplines. Nowadays, the main goal of the SB&NB lab is to combine computational and structural biology with interaction discovery experiments to unveil the basic wiring architecture of physio-pathological pathways. It is our believe that a deeper knowledge of the global topology of interactome networks related to human disease will have important bearings in the discovery of new drug targets and biomarkers, optimization of preclinical models and understanding how biological networks change from the healthy state to disease.*

Gaining Insight into the Uncultured Microbial World by Computational Metagenome Analysis

Alice Carolyn McHardy

HHU Düsseldorf and Helmholtz Center for Infection Biology in Braunschweig, Germany. Chair of the Department for Algorithmic Bioinformatics at HHU, Head of the Computational Biology for Infection Research Group at the Helmholtz Center.

Abstract: Next generation sequencing allows to extensively survey the genome-wide genetic diversity of microbial communities, as well as populations from all domains of life. A major challenge is the development of computational methods for hypothesis generation and basic computational analysis of these large-scale data sets. I will present our work on computational methods for metagenome analysis. We have developed fast and accurate methods for the taxonomic assignment of sequence fragments obtained by DNA shotgun sequencing of microbial communities to the species or higher-ranking taxa from which they originate. Furthermore, we are working on techniques for predicting and characterizing microbial phenotypes, as well as identifying the relevant protein repertoire for a given phenotype, focusing hereby on microbial plant biomass degradation and plant-associated microbial communities.

Bio: *Alice Carolyn McHardy's research interest is the design of computational methods to solve problems from the medical and biotechnological domains using next generation sequence data. A particular focus is the study of the evolution of influenza A viruses and characterizing the taxonomic composition, functions and evolution of microbial communities using metagenome sequence samples. She has a PhD in bioinformatics and a master's degree in biochemistry, both from Bielefeld University. From 2005 to 2007 she worked as a postdoctoral researcher, then as a permanent staff member in the Bioinformatics & Pattern Discovery Group at the IBM T.J. Watson Research Center in Yorktown Heights, NY. From 2007 to 2012 she led the Max-Planck research group for Computational Genomics & Epidemiology at the Max-Planck institute for Informatics in Saarbrücken, Germany.*

Advances in data mining for biomedical research

Nada Lavrač

Jožef Stefan Institute, Ljubljana and University of Nova Gorica, Nova Gorcia, Slovenia. Professor, Head of Department of Knowledge Technologies.

Abstract: The talk will first outline standard approaches to data mining, with the emphasis on subgroup discovery which proves to be an effective tool for data analysis in biomedical applications. The core of the talk will be devoted to inductive logic programming and relational data mining which also have a great potential for biomedical research, with a focus on recently developed approaches to semantic data mining, which enable the use of domain ontologies as background knowledge in data analysis. The use of described techniques and tools will be illustrated on selected biomedical applications. Moreover, the talk will present several data mining workflows, implemented in the recently developed platforms Orange4WS and ClowdFlows, which implement advanced methodologies which can be reused in biomedical applications.

Bio: *Nada Lavrač is Head of Department of Knowledge Technologies at Jozef Stefan Institute, Ljubljana, Slovenia. She is also Professor at the Jozef Stefan International Postgraduate School in Ljubljana and at the University of Nova Gorica. Her main research interests are in Knowledge Technologies, with particular interests in machine learning, data mining, text mining, knowledge management and computational creativity. She is author of several books, including the recently published Foundations of Rule Learning, Springer 2012. Her special interest is in supervised descriptive rule induction, where the research goal is to automatically induce rules from class labeled data, stored either in simple tabular format or in complex relational databases. Areas of her applied research include data mining applications in medicine, health care and bioinformatics.*

Big Data in Biology

Ewan Birney

European Bioinformatics Institute, Hinxton, United Kingdom. Associate Director.

Abstract: Molecular biology is now a leading example of a data intensive science, with both pragmatic and theoretical challenges being raised by data volumes and dimensionality of the data. These changes are present in both "large scale" consortia science and small scale science, and across now a broad range of applications – from human health, through to agriculture and ecosystems. All of molecular life science is feeling this effect.

This shift in modality is creating a wealth of new opportunities and has some accompanying challenges. In particular there is a continued need for a robust information infrastructure for molecular biology. This ranges from the physical aspects of dealing with data volume through to the more statistically challenging aspects of interpreting it. A particular problem is finding causal relationships in the high level of correlative data. Drawing on recent experience I will explore both the "blue collar" challenges of data volume and the "white collar" challenges of interpretation.

I will end with the serendipitous invention of using DNA for an entirely different reason – as a long-time horizon digital archiving material. I will describe this method and some of its benefits (as well as a few downsides) and explain how a future culture in 10,000 years time may still be able to read all of Shakespeare's sonnets – and perhaps much more.

Bio: *Ewan Birney, a British biochemist and bioinformatician born in 1972, is a leader in the computing and analysis of the genome.*

After leaving Eton College and before enrolling at the University of Oxford, Birney interned at Cold Spring Harbor Laboratory, under the supervision of James Watson - a discoverer of the structure of DNA. Later, as a PhD student at the Wellcome Trust Sanger

Institute in Cambridge, Birney was involved in the Human Genome Project, which was led by James Watson. Within the project, he helped to assemble the genome database. Shortly after receiving his PhD Birney was employed by the EMBL European Bioinformatics Institute in Cambridge, and soon started to coordinate large-scale collaborative projects. Among others, two big public databases resulted from these projects: Ensembl Genome Browser, with vertebrate genomes and gene annotations, and ENCODE (Encyclopedia of DNA Elements), a database of functional elements in the human genome. Ewan Birney and his team also developed a number of open-source and widely-used bioinformatics tools. For his outstanding achievements, Birney was awarded with various awards such as the Francis Crick Lecture from the Royal Society (2003), Chris Overton Prize from the international Society for Computational Biology (2005) and the Benjamin Franklin Award (2005) for contributions in Open Source Bioinformatics from Bioinformatics.org in association with BioIT. The current research of Birney's group focuses on sequence algorithms to store digital data in DNA molecules, and on associating natural genome variations with cellular phenotypes. Since 2012, Ewan Birney also serves as an Associate Director of EMBL-EBI. Ewan Birney maintains a blog, Bioinformatician at Large, and is active on Twitter with over 5000 followers.

Rational confederation of genes and diseases

Doron Lancet

The Weizmann Institute of Science, Rehovot, Israel. Professor in the Department of Molecular Genetics, Head of the Crown Human Genome Center.

Abstract: Human diseases are disposed at the heart of extensive research that encompasses genomics, bioinformatics, systems biology and systems medicine. Some of the challenges facing disease bioinformatics are disease nomenclature, standard symbols (as for genes), and integration of information from diverse sources. A most important issue to be tackled is generating a global view of gene-disease relationships. This is relatively straightforward for the monogenic diseases, natural human knockouts that constitute a rich source of biological insight. However, for complex diseases, a concerted effort is needed to sort out signal from noise. This necessitates the use of comprehensive disease and gene compendia with extensive cross-relations. Two relevant such tools will be presented. The first is the widely used GeneCards, (www.genecards.org), encompassing automatically mined information from >100 sources on ~120,000 gene entries, including the most comprehensive compilation of non protein-coding RNA genes. In the GeneCards pipeline we plan the incorporation of genomic enhancers, for which there is increasing evidence for involvement in disease. The second relevant tool is MalaCards (http://www.malacards.org), a most comprehensive resource of human diseases, with ~17,000 entries, mined from >60 sources. The development of MalaCards posed many algorithmic challenges, such as disease names unification, integrated classification and disease-gene scrutiny. In analogy to GeneCards, MalaCards displays a web card for each human disease, with 17 sections, including textual summaries, related diseases, genetic variations, genetic tests and relevant publications.

Next generation sequencing of malady-affected individuals has become a key technology for relating genes to diseases. We have developed VarElect for linking disease phenotypes to gene variants. It performs judicious prioritization among short-listed variations, leveraging the rich information and inter-links within GeneCards and MalaCards. VarElect's algorithm affords inferring direct as well as indirect links between genes and phenotype-related keywords. For indirect implication, gene-to-gene relations are formed via expanded paralogy relations, shared publications, interaction networks and shared biological pathways. A recent enhancement for the latter is PathCards, which unifies >3,000 pathways from 12 data sources into ~1000 SuperPaths with optimal informativeness and minimal redundancy. PathCards greatly enhances VarElect's capacity to portray unsuspected disease-gene relations. GeneCards, MalaCards and their affiliated tool VarElect and PathCards thus provides a facile and robust avenue for confederating genes with diseases.

Bio: *Prof. Lancet earned his BSc in chemistry from the Hebrew University of Jerusalem, and his PhD is in Chemical Immunology from the Weizmann Institute of Science. Following postdoctoral training at Harvard and Yale, he joined the Weizmann Institute faculty in 1981, where he is currently the incumbent of the Silver Professorial Chair in Human Genomics and Head of the Crown Human Genome Center. Prof. Lancet has played a key role in establishing genome research in Israel, and is presently involved in the establishment of a National Center for Personalized Medicine at Weizmann. A pioneer of research on the sense of smell, he discovered key protein components of this sensory mechanism. He now studies the population genetics of human olfaction, showing that each human individual has a "personal nose". In parallel, Lancet runs several programs for disease gene discovery, including an international collaborative study on genome sequencing for rare monogenic diseases. In the realm of bioinformatics, Lancet developed GeneCards, a world-renown automatically-mined web compendium of human genes. Recently, he developed the companion database MalaCards, a comprehensive web tool for human diseases. Lancet was awarded the first international Takasago Award in Japan on olfactory research (1986), the Wright Award in the USA (1998) and the Landau Prize in Human Genetics in Israel (2008). He is member of the European Molecular Biology Organization since 1996 and was Council member of the Human Genome Organization (2007-2012). Lancet wrote a science column in the major Israeli daily Haaretz, and often delivers public lectures and interviews on the human genome project.*

The Detection of Architectural Modules in RNA sequences and the RNA-Puzzles Modeling Contest

Eric Westhof

University of Strasbourg, Strasbourg, France. Professor, Director of the Institute of Molecular and Cellular Biology, CNRS.

Abstract: RNA molecules are characterized by the formation of hydrogen-bonded pairs between the bases along the polymer. All base-base interactions present in nucleic acids, with at least two "standard" H-bonds, can be classified in twelve families where each family is a 4x4 matrix of the usual bases. The common Watson-Crick pairs belong to one of these families and the other eleven families gather the non-Watson-Crick pairs. The Watson-Crick pairs form the

secondary structure and all the other families are critical for the tertiary structure. RNA architecture is thus viewed as the hierarchical assembly of preformed double-stranded helices defined by Watson-Crick base pairs and RNA modules maintained by non-Watson-Crick base pairs. RNA modules are recurrent ensembles of ordered non-Watson-Crick base pairs. Such RNA modules constitute a signal for detecting structured non-coding RNAs with specific biological functions. It is therefore important to be able to recognize such elements within genomes. Through systematic comparisons between homologous sequences and x-ray structures, followed by automatic clustering, the whole range of sequence diversity in recurrent RNA modules has been characterized. These data permitted the construction of a computational pipeline for identifying known 3D structural modules in single and multiple RNA sequences in the absence of any other information. Any module can in principle be searched, but four can be searched automatically: the G-bulged loop, the Kink-turn, the C-loop and the tandem GA loop. The present pipeline can be used for RNA 2D structure refinement, 3D model assembly, and for searching and annotating structured RNAs. RNA-Puzzles are collective and blind experiments in RNA three-dimensional structure prediction. The goals are to assess the leading edge of RNA structure prediction techniques, compare existing methods and tools, and evaluate their relative strengths, weaknesses, and limitations in terms of sequence length and structural complexity. The results should give potential users insight into the suitability of available methods for different applications and facilitate efforts in the RNA structure prediction community in their efforts to improve their tools. Generally, the less well predicted models always had worse non-Watson-Crick scores, demonstrating the importance of identifying non-Watson-Crick pairs and RNA modules.

Bio: *Eric Westhof received his Ph.D. in Biophysics in 1974 (Liège University, Belgium) after graduate work at Regensburg University, Germany. In 1977, as a* FULBRIGHT-HAYS *Research Fellow, he joined the Department of Biochemistry, University of Wisconsin (Madison, USA) to work with M. Sundaralingam in crystallography of nucleic acids. In 1981, with a EMBO post-doctoral fellowship, he moved to Strasbourg to work with Dino Moras on transfer RNA crystals. In 1988, he became Professor of Structural Biochemistry at the University of Strasbourg. Since 2006, he is Director of the* Institut de Biologie Moléculaire et Cellulaire *and head of the unit* Architecture et Réactivité de l'ARN *of the CNRS. He is an executive editor of* RNA Journal *and* Nucleic Acids Research *and a member of* EMBO, Deutsche Akademie der Naturforscher LEOPOLDINA, Academia Europaea*, the* Académie des Sciences*.*

His research activities are centered on the relationships between sequences, architectures, evolution and functions of RNA molecules, especially those with catalytic activity. With his collaborators, he develops and studies RNA sequence alignments in the light of RNA structures and architectures in order to identify RNA modules and to develop rules for predicting RNA folds and functions. The rules are transformed into algorithms for manipulating and assembling RNA architectures ab initio *or in density maps. The tools used are X-ray crystallography, bioinformatics, sequence comparisons, three-dimensional modeling and molecular dynamics simulations.*

Schedule

Workshops, Tutorials and Satellite Meetings

Venue: see tables below to check event location

FORUM: Faculty of Medecine of Strasbourg: 4 Rue Kirschleger, Strasbourg.

PMC: Palais de la Musique et des Congrès, Place de Bordeaux - Avenue Schutzenberger, Strasbourg.

Registration

Unless otherwise indicated, one hour before the beginning of the event

Saturday, September 6

FORUM	**Faculty of Medecine of Strasbourg: 4 Rue Kirschleger, Strasbourg.**			
W07	Workshop on Integrative Dynamic Analyses of Large Biomedical Network Data	**Room 3**	9:00	18:00
W09	Workshop on Machine Learning for Systems Biology (Day1)	**Amphi Forum**	9:00	18:00
W11	The 1st Unified Workshop on Proteome and Metabolome Informatics	**Room 2**	9:00	18:00
T01	Analysis of Cis-Regulatory Motifs from High-Throughput Sequence Sets	**Room 4**	9:00	18:00
T06	Reuse, Develop and Share Biological Visualisation with BioJS	**Room 1**	9:00	17:00
T07	Scientific Workflows for Analysing, Integrating and Scaling Bioinformatics Data: a Practical Introduction to Galaxy, Taverna and WS-PGRADE	**Room 5**	9:00	18:00
T09	TADbit: Automated Analysis and Three-Dimensional Modeling of Genomic Domains	**Room 6**	9:00	17:00
PMC	**Strasbourg Convention Centre, Place de Bordeaux - Avenue Schutzenberger, Strasbourg.**			
W06	Workshop on Informatics based Approaches for Circular Dichroism Data	**Salon Rohan**	9:00	18:00
W08	Workshop on Logical Modelling and Analysis of Cellular Networks (Day1)	**Salon Boston**	9:00	18:00
W16	Workshop on Tools and Techniques for Analysis and Design of Macromolecular Structures	**Salon Stuttgart**	9:00	17:00
S01	European Student Council Symposium	**Salon Leicester**	9:00	18:00

Sunday, September 7

FORUM	**Faculty of Medecine of Strasbourg: 4 Rue Kirschleger, Strasbourg.**			
W02	BioNetVisA Workshop: From biological network reconstruction to data visualization and analysis in molecular biology and medicine	**Room 3**	9:00	17:00
W04	Workshop on Computational and Systems Biology for Disease Comorbidities	**Room 2**	9:00	17:00
W09	Workshop on Machine Learning for Systems Biology (Day2)	**Amphi Forum**	9:00	17:00
T04	Multivariate projection methodologies for the exploration of large biological data sets. Application in R using the mixOmics package	**Room 5**	9:00	17:30
T05	Protein Evolution Analysis: on the Use of Phylogenetic Trees	**Room 4**	9:00	17:00

PMC	Strasbourg Convention Centre, Place de Bordeaux - Avenue Schutzenberger, Strasbourg.			
W01	RADIANT Workshop: Analysis of Differential Isoform Usage by RNA-seq: Statistical Methodologies and Open Software	**Salon Dresden**	9:00	18:00
W03	First Workshop on Computational Methods for Structural RNAs - CMRS'14	**Salon Bartholdi**	9:00	18:00
W05	Drug Development 2.0 - Computational integrative biology methods for drug repurposing, target discovery and translational research.	**Salon Boston**	9:00	18:00
W12	Workshop on Recent Computational Advances in Metagenomics	**Salon Orangerie**	9:00	17:30
W13	sbv IMPROVER Workshop	**Salon Leicester**	9:00	17:00
T02	Computational Tools to Define and Analyse Logical Models of Cellular Networks	**Salon Rohan**	13:30	18:00
T03	IMGT, the Global Reference in Immunogenetics and Immunoinformatics	**Salon Rohan**	9:00	13:00
T08	Statistics and Numerics for Dynamical Modeling	**Salon Stuttgart**	9:00	17:00

Main conference

Venue:

PMC: Palais de la Musique et des Congrès, Place de Bordeaux - Avenue Schutzenberger, Strasbourg.

Sunday, September 7 – Conference Opening

16:00	Conference registration opens	PMC – Main Floor
18:00	Conference opening and welcome	Auditorium ERASME
18:15	**Keynote 1: Jean-Marie LEHN**. ***Perspectives in Chemistry: From Supramolecular Chemistry towards Adaptive Chemistry***.	
19:05	**Invited Talk:** Laurence LWOFF. *Biomedical Research and Human Rights: the case of biobanking.*	
19:30	Welcome Cocktail	Dining Room Contades
21:00	End of welcome cocktail	

Monday, September 8

7:45	Registration opens	
8:45	**Keynote 2: Patrick ALOY.** ***A network biology approach to novel therapeutic strategies.***	Auditorium ERASME
9:35	Distribution in two parallel sessions	
	Auditorium ERASME	Room SCHUMANN
	Session Mon1: Pathways and Molecular Networks (1)	**Session Mon5: Evolution and Population Genetics (1)**
9:40	*PP01 - HubAlign: An accurate and efficient method for global alignment of protein-protein interaction networks.* Daniela Boernigen	*PP11 - Polytomy Refinement for the Correction of Dubious Duplications in Gene Trees.* Manuel Lafond
10:05	*PP02 - Alignment-free protein interaction network comparison.* Waqar Ali	*PP12 - RidgeRace: Ridge regression for continuous ancestral character estimation on phylogenetic trees.* Christina Kratsch
10:30	*PP03 - Fast randomisation of large genomic datasets while preserving alteration counts.* Francesco Iorio	*PP13 - Point estimates in phylogenetic reconstructions.* Philipp Benner

10:55	Coffee Break	Main Floor & 1st Floor
	Session Mon3: Sequencing and Sequence Analysis for Genomics (1)	**Session Mon6: Evolution and Population Genetics (2)**
11:15	*PP06 - Lambda: The local aligner for massive biological data.* Hannes Hauswedell	*PP14 - ASTRAL: Genome-scale coalescent-based species tree estimation*. Siavash Mirarab
11:40	*PP07 - Fiona: a parallel and automatic strategy for read error correction.* Marcel Schulz	**Highlight Talk:** *HP03 - Patterns of positive selection in seven ant genomes.* Julien Roux
12:05	**Highlight Talk**: *HP02 - Comparison of mapping algorithms used in high-throughput sequencing: application to Ion Torrent data.* Ségolène Caboche	
12:30	LUNCH	Dining Room Contades
13:30	**Industrial and Demo Track**	See corresponding pages
14:35	**Keynote 3: Alice McHardy. *Gaining Insight into the Uncultured Microbial World by Computational Metagenome Analysis.***	Auditorium ERASME
15:25	Distribution in two parallel sessions	
	Auditorium ERASME	Room SCHUMANN
	Session Mon4: Sequencing and Sequence Analysis for Genomics (2)	**Session Mon7: Structural Bioinformatics (1)**
15:30	*PP08 - FastHap: fast and accurate single individual haplotype reconstruction using fuzzy conflict graphs.* Sepideh Mazrouee	*PP15 - Assessing the local structural quality of transmembrane protein models using statistical potentials (QMEANBrane).* Gabriel Studer
15:55	*PP09 - Probabilistic single-individual haplotyping*. Volodymyr Kuleshov	*PP16 - A new statistical framework to assess structural alignment quality using information compression*. James Collier
16:20	*PP10 - cnvOffSeq: detecting intergenic copy number variation using off-target exome sequencing data*. Evangelos Bellos	*PP17 - Entropy driven partitioning of the hierarchical protein space*. Nadav Rappoport
16:45	Coffee Break	Main Floor & 1st Floor
	Session Mon2: Pathways and Molecular Networks (2)	**Session Mon8: Structural Bioinformatics (2)**
17:05	*PP04 - Identifying transcription factor complexes and their roles*. Thorsten Will	*PP18 - PconsFold: Improved contact predictions improve protein models*. Mirco Michel
17:30	*PP05 - Personalized identification of altered pathways in cancer*. Taejin Ahn	*PP19 - Microarray R-based analysis of complex lysate experiments with MIRACLE*. Markus List
17:55	**Highlight Talk:** *HP01 - Wiring miRNAs to pathways: a topological approach to integrate miRNA and mRNA expression profiles*. Enrica Calura	**Highlight Talk:** *HP04 - Comprehensive analysis of DNA polymerase III alpha subunits and their homologs in bacterial genomes.* Česlovas Venclovas
18:20	Break – Moving to Poster Session	
18:30	**Poster Session (Odd numbers)**	1st Floor
19:30	**Poster Session (Even numbers)**	1st Floor
20:30	Launching Ice-Breaking Event	Main Floor
20:45	Ice-breaking Event	

Tuesday, September 9

7:45	Registration opens	
8:45	**Keynote 4: Nada LAVRAČ. *Advances in data mining for biomedical research.***	Auditorium ERASME
9:35	Distribution in the two parallel sessions	

	Auditorium ERASME	Room SCHUMANN
	Session Tue1: Computational Systems Biology (1)	**Session Tue4: Biological Knowledge Discovery from Data**
9:40	*PP20 - Stronger findings for metabolomics through Bayesian modeling of multiple peaks and compound correlations*. Tommi Suvitaival	*PP28 - Unveiling new biological relationships using shared hits of chemical screening assay pairs.* Monica Campillos
10:05	*PP21 - Causal network inference using biochemical kinetics*. Chris Oates	*PP29 - Identification of structural features in chemicals associated with cancer drug response: A systematic data-driven analysis*. Suleiman Ali Khan
10:30	**Highlight Talk:** *HP05 - High-dimensional Bayesian parameter estimation: Case study for a model of JAK2/STAT5 signaling*. Sabine Hug	**Highlight Talk:** *HP06 - Shaping the interaction landscape of bioactive molecules*. David Gfeller
10:55	Coffee Break	Main Floor & 1st Floor
	Session Tue2: Computational Systems Biology (2)	**Session Tue5: Gene Expression (1)**
11:15	*PP22 - Effects of small particle numbers on long-term behaviour in discrete biochemical systems*. Peter Dittrich	*PP30 - Estimating the activity of transcription factors by the effect on their target genes*. Rainer Koenig
11:40	*PP23 - TEMPI: Probabilistic modeling time-evolving differential PPI networks with multiple information.* Yongsoo Kim	*PP31 - Modeling DNA methylation dynamics with approaches from phylogenetics*. Dennis Kostka
12:05	*PP24 - Experimental design schemes for learning Boolean network models.* Nir Atias	**Highlight Talk:** *HP07 - Key regulators control distinct transcriptional programmes in blood and progenitor and mast cells*. Felicia Ng
12:30	LUNCH	Dining Room Contades
13:30	**Industrial and Demo Track**	See corresponding pages
14:35	**Keynote 5: Ewan Birney. *Big Data in Biology.***	Auditorium ERASME
15:25	Distribution in two parallel sessions	
	Auditorium ERASMUS	Room SCHUMANN
	Session Tue3: Bioinformatics of Health and Disease (1)	**Session Tue6: Gene Expression (2)**
15:30	*PP25 - OncodriveROLE classifies cancer driver genes in Loss of Function and Activating mode of action.* Michael Philipp Schroeder	*PP32 - Two-dimensional segmentation for analyzing HiC data*. Celine Levy-Leduc
15:55	*PP26 - ContrastRank: a new method for ranking putative cancer driver genes and classification of tumor samples.* Emidio Capriotti	*PP33 - Broad-Enrich: Functional interpretation of large sets of broad genomic regions*. Raymond Cavalcante
16:20	*PP27 - Drug susceptibility prediction against a panel of drugs using kernelized Bayesian multitask learning.* Mehmet Gönen	**Highlight Talk:** *HP08 - Chromatin position effects quantified from thousands of reporters integrated in parallel*. Lodewyk Wessels
16:45	Coffee Break - Poster Session begins	Main Foor & 1st Floor
17:30	1st departure to river boat trip (then every ¼ hour)	Main Floor
18:45	End of Poster Session	1st Floor
19:15	Last departure to river boat trip	Main Floor
From 19:00	Gala Evening at the Council of Europe or Ancienne Douane	

Wednesday, September 10

8:00	Conference registration opens	Main Floor
9:00	**Keynote 6: Doron LANCET**. ***Rational confederation of genes and diseases***.	Auditorium ERASME

	Session Wed1: Databases and Ontologies	
9 :50	*PP34 - The impact of incomplete knowledge on the evaluation of protein function prediction: a structured-output learning perspective*. Yuxiang Jiang	
10:15	*PPE35 - Integration of molecular network data reconstructs Gene Ontology*. Vladimir Gligorijevic	
10 :40	Coffee Break	Main Floor
	Session Wed2: Bioinformatics of Health and Disease (2) and Bio-Imaging	Auditorium ERASME
11:00	*PP36 - Transcriptome-guided amyloid imaging genetic analysis via a novel structured sparse learning algorithm.* Jingwen Yan	
11:25	*PP37 - Large-scale automated identification of mouse brain cells in confocal light sheet microscopy images*. Paolo Frasconi	
11:50	**Highlight Talk:** *HP09 - Novel Developments in computational clinical breath analysis and biomarker detection*. Anne Christine Hauschild	
12:15	Moving to Dining Room	
12:30	LUNCH	Dining Room Contades
13:30	**Industrial and Demo Track**	See corresponding pages
	Session Wed3: Text Mining for Computational Biology	Auditorium Erasme
14:35	*PP38 - Extracting patterns of database and software usage from the bioinformatics literature*. Geraint Duck	
15:00	**Highlight Talk:** *HP10 - Text mining technologies for database curation*. Fabio Rinaldi	
	Session Wed4: RNA prediction	Auditorium Erasme
15:25	*PP39 - CRISPRstrand: Predicting repeat orientations to determine the crRNA-encoding strand at CRISPR loci*. Omer S. Alkhnbashi	
15:50	*PP40 - Towards a piRNA prediction using multiple kernel fusion and support vector machine.* Fariza Tahi.	
16:15	**Keynote 7: Eric WESTHOF**. ***The Detection of Architectural Modules in RNA sequences and the RNA-Puzzles Modeling Contest.***	
17:05	Prize Ceremony and Concluding remarks	
17:15	End of Conference – Farewell Coffee	Main Floor

Workshops

W01: Analysis of Differential Isoform Usage by RNA-seq: Statistical Methodologies and Open Software

Description

There are many computational and statistical challenges in the analysis of differential isoform usage by RNA-Seq. The workshop addresses users and method developers who are interested in better understanding the features of current tools and exploring approaches to method benchmarking. The workshop will cover the most important current statistical and computational challenges. The workshop is aimed at computational biologists, statisticians and genome biologists with an interest in data analysis and statistical methods and applications in transcriptomics.

Organizers

Simon Anders (EMBL) anders@embl.de, Wolfgang Huber (EMBL) whuber@embl.de, Magnus Rattray (University of Manchester) magnus.rattray@manchester.ac.uk, Mark Robinson (University of Zurich) mark.robinson@imls.uzh.ch

Program

Session 1 (9.00-9.05 Welcome)
9.05-9.40 Jernej Ule (Institute of Neurology, University College London, UK): *Regulation of cryptic splicing elements within long neuronal genes*
9.40-10.15 Mar Gonzales-Porta (EMBL-EBI, Cambridge, UK): *Identification, annotation and visualisation of extreme changes in splicing from RNA-seq experiments with SwitchSeq*
10.15-10.30 Axel Rasche (Max Planck Institute for Molecular Genetics, Berlin, Germany): *ARH/ARH-Seq – Discovery tool for differential splicing in High-throughput data*
10.30-10.45 **Coffee Break**
Session 2
10.45-11.20 Alejandro Reyes (EMBL, Heidelberg, Germany): *Studying differential usage of exons with the DEXSeq package*
11.20-11.35 Thomas Risch (Max Planck Institute for Molecular Genetics, Berlin, Germany): *Prediction of subtype-specific alternative exon usage in medulloblastoma*
11.35-11.50 Jose M. Garcia-Manteiga (OSR, Milan, Italy): *Alternative splicing in response to hypoxia in HUVEC cells by differential exon usage and differential expression of isoforms*
11.50-12.00 Benchmarking team activity
12.00-13.30 **Lunch**
Session 3
13.30-14.05 Alyssa Frazee (John Hopkins Bloomberg School of Public Health, Baltimore, USA): *Engineering annotation-agnostic tools for differential expression analysis*
14.05-14.20 Daniel Gautheret (Insitut de Génétique et Microbiologie, Paris, France): *Detecting Differential RNA-processing Events from RNA-seq data*
14.20-14.55 David Rossell (Dept. of Statistics, University of Warwick): *Casper for efficient summarization and design of RNA-seq studies*
14.55-15.30 Elsa Bernard (Institut Curie, Paris): *Fast isoform detection from RNA-Seq data with network flow techniques*
15.30-15.45 Malgorzata Nowicka (Institute of Molecular Sciences, University of Zurich, Switzerland): *Dirichlet-multinomial model for differential splicing detection in RNA-seq data*
15.45-16.00 Panos Papastamoulis (Faculty of Life Sciences, University of Manchester, UK): *Identifying differentially expressed transcripts via a Reversible Jump MCMC sampler*
16.00-16.20 **Coffee Break**
Session 4
16.20-16.55 David Kreil (Bioinformatics Research Group, Boku University Vienna, Austria) : *Benchmarking gene transcript expression profiling - challenges and platform specific results*
16.55-17.10 Azra Krek (Memorial Sloan-Kettering Cancer Center, New York, USA): *Comparison of isoform quantification methods for short-reads sequencing data using SEQC data*
17.10-18.00 Benchmarking team presentations and discussion

ECCB14 web site http://www.eccb14.org/program/workshops/rna-seq

W02: BioNetVisA

Description

BioNetVisA workshop will bring together different actors of network biology from database providers, networks creators, computational biologists, biotech companies involved in data analysis and modeling to experimental biologists, clinicians that use systems biology approaches. The participants will be exposed to the different paradigms of network biology and the latest achievements in the field.
The goal of BioNetVisA workshop is to build a discussion around various approaches for biological knowledge formalisation, data integration and analysis; compatibility between different methods and biological networks

resources available the field; applicability for concrete research and clinical projects depending on scientific question and type of high-throughput data.
The BioNetVisA workshop aims at identifying bottlenecks and proposing short- and long-term objectives for the community as discussing questions about accessibility of available tools for wide range of user in every-day standalone application in biological and clinical labs. In addition, the possibilities for collective efforts by academic researchers, clinicians, biotech companies and future development directions in the field will be discussed during the round table panel.

Organizers

Inna Kuperstein (Institut Curie, Paris, France) ; Emmanuel Barillot (Institut Curie, Paris, France) ; Andrei Zinovyev (Institut Curie, Paris, France) ; Hiroaki Kitano (Okinawa Institute of Science and Technology Graduate University, RIKEN Center for Integrative Medical Sciences) ; Nicolas Le Novère (Babraham Institute, Cambridge) ; Robin Haw (Ontario Institute for Cancer Research, Toronto) ; Mario Albrecht (University Medicine Greifswald, Greifswald, Germany) ; Benno Schwikowski (Institut Pasteur, France).

Program

9.00-9.40 Keynote talk: Tom Freeman (University of Edinburgh, Edinburgh, UK) *Network visualization and analysis of gene expression data.*
Session1 (part 1): Development, curation and maintenance of biological network databases Chair: Emmanuel Barillot (Institut Curie, Paris, France)
9.40-10.00 *Reactome: Linking Biological Pathways and Networks to Disease.* Robin Haw (Ontario Institute for Cancer Research, Toronto, Canada)
10.00-10.20 *IntAct - High Resolution Disease-centric Networks.* Henning Hermjakob (EMBL-European Bioinformatics Institute, Hinxton, UK)
10.20-10.40 *Coffee break and posters*
Session1 (part 2): Development, curation and maintenance of biological network databases Chair: Emmanuel Barillot (Institut Curie, Paris, France)
10.40-11.00 *Visualization and analysis of data using Atlas of Cancer Signalling Networks (ACSN) and NaviCell tools for integrative systems biology of cancer.* Inna Kuperstein (Institut Curie, Paris, France)
11.00-11.20 *Graphical data representation*. Pauline Gloaguen (CEA, Grenoble, France)
11.20-12.00 Keynote talk: Samik Ghosh (Systems Biology Institute, Tokyo, Japan) *Garuda – The way biology connects*
12.00-12.20 Posters flash presentations
12.20-13.30 *Lunch and poster session*
Session 2: Data visualisation and analysis in the context of biological networks_Chair: Mario Albrecht (Graz University of Technology, Austria)
13.30-13.50 *A Pathway-centric Approach to Multiomics Research Powered by GeneSpring Analytics.* Nigel Skinner (Agilent Technologies, London, UK)
13.50-14.10 *An integrative network analysis pipeline in Cytoscape.* Mohammed El-Kebir (Centrum Wiskunde & Informatica, Amsterdam, Netherlands)
14.10-14.30 *Using network analysis: HyperSet is a novel framework for functional interpretation of 'omics' data in global networks.* Andrey Alexeyenko (Stockholm Science for Life Laboratory, Solna, Sweden)
Session3: Network biology in research and medicine Chair: Robin Haw (Ontario Institute for Cancer Research, Toronto, Canada)
14.30-14.50 *Gene networks, tumor subtypes and patient prognosis signatures associated with ovarian cancer mutations.* Vladimir Kuznetsov (Bioinformatics Institute, Biopolis Singapore)
14.50-15.10 *Decoding Network Dynamics in Cancer.* Rune Linding (Technical University of Denmark, Lyngby, Denmark)
15.10-15.30 *Using Topological Analysis to Study Metabolism of Heterotrophic Plant Cell Network.* Nguyen V. N. Tung, (Université Bordeaux 1, Bordeaux, France)
15.30-15.50 *Conserved cross-species network modules elucidate Th17 T cell differentiation in human and mouse.* Hayssam Soueidan (NKI-AVL, Amsterdam, Netherlands)
15.50-16.10 *Meta expression analysis of regulatory T cell experiments for gene regulatory network reconstruction*. Stefan Kroeger (Humboldt-Universität zu Berlin, Berlin, Germany)
16.10-16.30 *Coffee break and posters*
16.30-17.10 Keynote talk***THE EMBO LECTURE: Yosef Yarden (The Weizmann Institut of Science, Rehovot, Israel) *HER2 and EGFR: at last, cancer therapy meets systems biology.*
17.10-17.50 Round table - discussion, awarding poster prizes and conclusions

ECCB14 web site: http://www.eccb14.org/program/workshops/bionetvisa

W03: First Workshop on Computational Methods for Structural RNAs - CMRS'14

Description

Ribonucleic acids (RNAs), the "genomic dark matter", have emerged as one of the most important of biomolecules. They play key roles in various aspects of the gene transcription and regulation processes, and are the focus of an ever-increasing interest in all subfields of biology. Deciphering the function of a non-protein coding RNA requires an intimate knowledge of its structure, motivating the development of structure-centric analyses. Underlying RNA structural workflows, one finds computational methods focusing on discrete abstractions for conformations, such as the secondary structure, which possibly include non-canonical interactions and pseudoknots. Such representations have led to efficient and accurate methods and algorithms, amenable to transcriptomic-scale studies. Recently, novel

experimental techniques, based on next-generation sequencing, have led to an unprecedented deluge of soft structural data (chemical/enzymatic footprinting, FRET...). It is therefore one of the exciting challenges of computational biology, and the subject of much work within the RNA bioinformatics community, to contribute new paradigms and methods addressing the challenge of scalability for RNA structure-centric pipelines in the big-data era. This challenge not only originates in the sheer magnitude of produced transcriptomic data, but also in the combinatorial explosion of sequences and structures associated with a given transcript. Successfully addressing this challenge will require a combination of highly efficient algorithms, mathematical models and versatile, yet compact, discrete representations and data structures.
The workshop aims at bringing together researchers and students, contributing and using algorithms and methods for structure-centric analyses of RNA. It strives at providing a forum for the dissemination of state-of-the-art methods and tools using discrete representations of folding landscapes in the context of big data. It will contribute to establish best-practices towards a better support for large-scale data within structure-centric algorithms.

Organizers

Fabrice Jossinet (Université de Strasbourg, France) ; Yann Ponty (CNRS – Ecole Polytechnique, France) ; Jérôme Waldispühl (University McGill, Canada).

Program

Session 1 – Comparative Structural RNAomics I

Jan Gorodkin: *Searching for SNPs disrupting RNA secondary structures*
Laetitia Bourgeade, Julien Allali and Cédric Chauve: *Chaining Sequence/Structure Seeds for Computing RNA Similarity*
Coffee break

Session 2 – Comparative Structural RNAomics II

Azadeh Saffarian, Mathieu Giraud and Helene Touzet: *Searching for alternate RNA structures in genomic sequences*
Shay Zakov, Nimrod Milo, Tamar Pinhas, Sivan Yogev, Erez Katzenelson, Eitan Bachmat, Yefim Dinitz, Dekel Tsur and Michal Ziv-Ukelson: *Recent Results on Three Problems in Comparative Structural RNAomics*
Lunch break

Session 3 – Genome-scale RNA bioinformatics

Mihaela Zavolan: *Deciphering the regulatory functions of miRNAs*
Svetlana Shabalina, Aleksey Ogurtsov, Anna Kashina and Nikolay Spiridonov: *The role of periodic mRNA secondary structure and RNA-RNA interactions in biological regulation and complexity*
Fabrizio Costa, Steffen Heyne, Dominic Rose and Rolf Backofen: *Scalable structural clustering of local RNA secondary structures*

Coffee break

Session 4 – Algorithmic foundations

Liang Ding, Xingran Xue, Sal Lamarca, Mohammad Mohebbi, Abdul Samad, Russell L. Malmberg and Liming Cai: *Ab initio Prediction of RNA Nucleotide Interactions with Backbone k-Tree Model*
Cedric Saule and Robert Giegerich: *Observations on the Feasibility of Exact Pareto Optimization*

ECCB14 web site http://www.eccb14.org/program/workshops/structural-rnas

Workshop web site http://cmsr2014.wordpress.com/

W04: Workshop on Computational and Systems Biology for Disease Comorbidities

Description

Disease comorbidity – or multimorbidity – exists if 2 or more disorders affect the same individuals more often than expected by chance. Such co-occurrence of diseases are the rule rather than the exception, in particular in ageing populations. Disease comorbidities hence represent a major burden in public health, and greatly affect disease costs and outcomes. But the recent expansion of large-scale biomedical databases and post-genomics disease-related datasets are now offering an unprecedented avenue for the investigations on comorbidities between human disorders. And Computational and Systems Biology are undeniable actors of such investigations.

Organizers

Anaïs Baudot, CNRS, Marseille, France. Contact: anais.baudot@univ-amu.fr ; Alfonso Valencia, CNIO, Madrid, Spain. Contact: avalencia@cnio.es

Program

8h-9h Welcome

Session 1 – Discovering Disease Comorbidities

9h-9h45 Keynote Søren Brunak (CBS, Copenhague, Denmark)
9h45-10h15 Laura I Furlong (IMIM, Universitat Pompeu Fabra, Barcelona, Spain): *PsyGeNET: a curated resource on associations between genes and psychiatric disorders. Application to the study of the comorbidities between alcohol and cocaine dependences and depression.*
10h15- 10h45h *Coffee break*

Session 2 – Interpreting Disease Comorbidities

10h45- 11h30 Keynote Ioannis Xenarios (SIB, Lausanne, Switzerland)
11h30-12h Jane A Driver (Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, USA): *Cellular Teleology and the Inverse Association between Cancer and Alzheimer's disease.*
12h-13h30 *Lunch*
13h30-14h15 Keynote Alfonso Valencia (CNIO, Madrid, Spain): *Molecular Evidence for the Inverse Comorbidities between*

CNS and Cancer detected by transcriptomic meta-analyses.
14h15- 14h45 Paulina Gomez (CNIO, Madrid, Spain): *Comorbidities and omics data integration towards a pancreas cancer risk prediction model.*
Session 3 – Using Disease Comorbidities
14h45-15h30 Keynote Patrick Aloy (IRB, Barcelona, Spain): *Blending chemistry and biology to understand human health and disease.*
15h30-16h Anaïs Baudot (CNRS, Marseille, France): *Functional Module approach for the study of Prostate Cancer and Disease Comorbidities.*
16h-16h20 *Coffee break*
16h20- 17h Round table
17h End of the workshop, departure to the ECCB opening lecture.

ECCB14 web site http://www.eccb14.org/program/workshops/csbdc

W05: Drug Development 2.0 -- Computational integrative biology methods for drug repurposing, target discovery and translational research

Description
The discovery of differentiated new medicines that provide clear benefits to patients remains challenging in spite of ever increasing investments. At the same time the quantity and diversity of patient related data continues to grow exponentially (pre-clinical data, clinical data, patient data, EHR, 'OMICs data, and information associated with medicines in general). New business models are also emerging with stronger collaborations between the industry and the academia with large public private partnerships (PPPs such as IMI in the EU) or the use of crowd sourcing. The scientific community is starting to leverage this wealth of information in ways that have the potential to disrupt the traditional drug discovery and development process as we know it. This workshop also invites submissions that leverage and translate similar techniques and data sets to new application domains including Health Outcomes Research, Biotech and Agriculture.
This workshop intends to bring together leading members from academia, biotech and pharmaceutical companies to discuss how to maximize the use of this data deluge to accelerate the discovery of new medicines that will ultimately benefit patients.

Organizers
Prof. Dr. Jan Baumbach, University of Southern Denmark, Denmark ; Prof. Dr. Michael Schroeder, TU Dresden, Germany ; Dr. Philippe Sanseau, GlaxoSmithKline, UK ; Dr. Jacob Koehler, DOW AgroSciences, USA

Program
9 :00 Welcome - Dr Jacob Koehler
Session 1 - Chair: Dr Jacob Koehler
9:00-9:20 Dopazo, Joaquin (invited speaker)(Centro de Investigacion Principe Felipe,Valencia, Spain) tba
9:30-9:50 Dogan, Tunca (European Molecular Biology Laboratory, European Bioinformatics Institute, United Kingdom) *Comprehensive Drug-Target Predictions with the Combination of Domain Mapping and Ligand Similarity Detection.*
10:00-10:20 Tiwari, Sandeep; Almeida, Sintia; Hassan, Syed; Costa, Marcília Pinheiro da; Folador, Edson; Jamal, Syed; Junior, Alberto F. Oliveira Junior; Abreu, Vinícius Augusto Carvalho de Abreu; Silva, Artur; Barh, Debmalya; Azevedo, Vasco (Federal University of Minas Gerais, Brazil) *Targeting PhoPR System of Corynebacterium pseudotuberculosis via Potential Antibacterial Compounds from Natural Resources: An in silico Approach*.
10:30 - 11:00 *Break*
Session 2 - Chair: Dr Philippe Sanseau
11:00-11:20 Overington, John (invited speaker)(European Molecular Biology Laboratory, European Bioinformatics Institute, United Kingdom) *ChEMBL - a database of drug-like molecules, molecular targets and bioassays.*
11:30-11:50 Sun, Kai; Buchan, Natalie; Larminie, Chris; Przulj, Natasa (Imperial College, United Kingdom) *The integrated disease network.*
12:00-12:20 Borukhov, Itamar (Compugen Ltd., Israel) *Computational Discovery of Drug Targets and Therapeutic Biologics: Identifying Novel Natural Peptides and Functional Epitopes.*
12:30-14:00 *Lunch & posters*
Session 3 - Chair: Prof Dr Michael Schroeder
14:00-14:20 Przulj, Natasa (invited speaker) (Imperial College, United Kingdom) tba
14:30-14:50 Sinha, Shriprakash (Netherlands Bioinformatics Center, The Netherlands) *Integration Of Prior Biological Knowledge And Epigenetic Information Enhances Prediction Accuracy Of Bayesian Wnt Pathway.*
15:00-15:20 Jiménez, Francisco Martínez (Center for Genomic Regulation, Barcelona, Spain) *nAnnoLyze: ligand-target prediction by structural network biology.*
15:30-15:50 Folador, Edson; Hassan, Syed; Lemke, Ney; Barh, Debmalya; Silva, Artur; Ferreira, Rafaela; Azevedo, Vasco (Federal University of Minas Gerais, Brazil) *An improved interolog mapping-based computational prediction of protein-protein interactions with increased network coverage.*
16:00-16:30 *Break*
Session 4 - Chair: Prof Dr Jan Baumbach
16:30-16:50 Sarajlic, Anida; Gligorijevic, Vladimir; Radak, Djordje; Przulj, Natasa (Imperial College, United Kingdom) *Network wiring of pleiotropic kinases yields insight into dissociation of diabetes and aneurysm.*
17:00-17:20 Qurrat-ul-Ain, Ainee Modelling (University of Cambridge, United Kingdom) *Ligand Selectivity of Serine Proteases using Proteometric Approaches.*

17:30-17:50 Poongavanam, Vasanthanathan; Olsen, Jógvan Magnus; Kongsted, Jacob (University of Southern Denmark, Denmark) *Binding Free Energy Based Structural Dynamics Analysis of HIV-1 RT RNase H-Inhibitor Complexes*.
Closing remarks - Dr Jacob Koehler

ECCB14 web site http://www.eccb14.org/program/workshops/dd

W06: Workshop on Informatics based Approaches for Circular Dichroism Data

Description

Circular Dichroism (CD) is a very important technique used in structural biology for determining protein secondary structure, observing conformational changes brought about by protein:protein or protein:ligand interactions, and for monitoring protein folding. It is regularly employed and widely used for characterization in both academia and industry. With the advent of the enhanced technique of Synchrotron Radiation Circular Dichroism (SRCD) this has extended the wavelength range and therefore the amount of data able to be collected with a concomitant increase in the available information content. It has been realized that there is more than just the secondary structure content of a protein that can be derived from CD, and particularly SRCD, data. This spectroscopic technique is now ripe for informatics based approaches to enrich the amount of protein structural information that can be derived from such spectra as the entire structure makes pivotal contributions to the spectra produced. This workshop is designed to introduce and invite the computational biology world to exploit this area of science thereby enhancing the technique as a result through an increase in information available through the provision of new and novel analysis methods.

Historically it was proposed very early in the development of CD as a technique that a spectrum could be deconvoluted to provide the secondary structure content of the protein that produced the data. To a first approximation this is reasonably accurate, and many mathematical methods have been derived to gain this information from the spectral data. However, it is clear that this is only a first approximation as there are often inaccuracies apparent in the results, and it is obvious that more information content about the structures of the proteins is present in the data. We feel that the time is at hand for researchers in computational biology to become more involved in this field of research as there is more structural information available from CD/SRCD spectra This is particularly a timely opportunity as we now have the valuable resource that is the Protein Circular Dichroism Data Bank that contains a growing number of validated spectra of proteins. These can be utilized to develop more sophisticated and in-depth approaches to gaining structural information from CD/SRCD spectra. This workshop offers the opportunity for computational biologists to gain knowledge about the background of CD/SRCD spectroscopy, and of the analysis methods currently used from the experts in this field.

We aim to attract interest from the broader computational biology community by showing that the technique of Circular Dichroism has a lot to offer in regards to gaining yet more information from the data than is currently obtained. This workshop will provide essential background knowledge on this technique and will show the extent to which protein structural information is related to the CD spectrum obtained and on the various diverse approaches utilised in getting this information. It is therefore open to all who are interested in broadening and diversifying their computational repertoire towards obtaining as much useful protein structure information as is available from this spectroscopy area of science.

At this workshop we will be announcing the establishment of a new assessment competition specifically aimed towards the relationship between Circular Dichroism Spectra and Information Content. More details of the competition (which will run for a year) will be given in the workshop.

Organizers

Dr. Robert W. Janes ; Dr. Lazaros Mavridis. Queen Mary University of London, UK.

Program

PART I Data Resources for Circular Dichroism Spectroscopic Analyses

09:10 – 10:00 B.A. Wallace – *Introduction to CD Spectroscopy and Analyses* (Birkbeck, UK)
10:00 – 10:30 L. Whitmore – The PCDDB (Birkbeck, UK)
10:30 – 11:00 *Tea/Coffee*

PART II New Methods of Analysis - I

11:00 – 11:25 C. Perez-Iratxeta – *K2D2 and K2D3: Two Tools for Protein Secondary Structure Prediction from Circular Dichroism Spectra* (Ottawa,Canada)
11:25 – 11:50 P. Bellstedt – *CAPITO : A user-friendly web server-based analysis and plotting tool for circular dichroism data* (Jena, Germany)
11.50 - 13.00 *LUNCH & POSTERS*

PART III New Methods of Analysis - II

13:00 – 13:25 A. Rodger – *Trying to get the most out of your circular (and other) dichroism data* (Warwick, UK)
13:25 – 13:50 J. Kardos – *BeStSel: From secondary structure determination to fold recognition using CD spectroscopy* (Budapest, Hungary)
13:50 – 14:15 L. Cole – *Higher order structural (HOS) comparison of pharmaceutical protein drugs using CD spectroscopy. The opportunity and challenges for the adoption of CD as a true industrial tool*. (APP, Leatherhead, UK)
14:15 – 14:40 (To be confirmed)
14:40 – 15:10 Tea/Coffee

PART IV First Principles-Based Analyses

15:10 – 15:35 J. Hirst – *First Principles Calculations of Protein Circular Dichroism* (Nottingham, UK)
15:35 – 16:00 K. Thomasson – *The Dipole Interaction Model, a Classical Approach to Protein Circular Dichroism* (North Dakota, USA)

PART V **The CIDASE Challenge**
16:00 – 16:25 L. Mavridis (Queen Mary, UK)
16:25 – 16:30 Poster Prize
16:30 – 16:50 Poster Presentation Talk/Discussion
16:50 - 17:00 Closing Remarks

ECCB14 web site http://www.eccb14.org/program/workshops/cdd

W07: Workshop on Integrative Dynamic Analyses of Large Biomedical Network Data

Description
Networks are used to model many real-world phenomena. In the biological and medical domains, networks can model: physical or functional interactions between genes, proteins, metabolites, or other biomolecules; disease-gene or drug-target associations; relationships between diseases, between drugs, or between patients based on electronic healthcare record data; and so on. Analogous to revolutionary influence that genetic sequence research has had, research of biological network data is expected to transform our understanding of biological function, disease, therapeutics, and lead to personalized medicine. Significant efforts have already been made to extract novel biological and medical knowledge from network topology. Initially, the efforts have mostly focused on analyzing homogeneous and static network data. However, with an increasing availability of large amounts of high-throughput network data of many different types and of dynamic nature, the network research community has recognized the need for efficient network-based methods for fusing different data types into an integrated heterogeneous framework and for studying the dynamic properties of complex systems in biomedicine. These advances are needed to deepen biological understanding and improve healthcare. For these reasons, this workshop aims to bring together scientists at all stages of their career with interests in (but not limited to) large-scale analyses of dynamic and heterogeneous biological network data or related practical biomedical applications, such as understanding disease, drug-repositioning, pharmacogenomics, aging, and personalized medicine, or healthcare.

Organizers
Natasa Przulj, Imperial College London. Contact: n.przulj@imperial.ac.uk ; Tijana Milenkovic, University of Notre Dame. Contact: tmilenko@nd.edu

Program
(KT=keynote talk; IT=invited talk; CT=contributed talk)

Time	Event	
8-9am	*Registration*	
8:50-9am	Nataša Pržulj (Imperial College London) Tijana Milenković (University of Notre Dame)	Welcome and introduction
9-9:50am	KT1: Igor Jurisica (University of Toronto)	"Avoiding fusion of illusion and confusion - reducing bias in network-based analyses"
9:50-10:30	IT1: Ben Raphael (Brown University)	"Algorithms for Analyzing Mutated Networks and Pathways in Cancer"
10:30-10:45am	*Coffee break*	
10:45-11:15am	CT1: Omer Nebil Yaveroglu (Imperial College London)	"Topology-Function Conservation in Protein-Protein Interaction Networks"
11:15-11:55am	IT2: Michael Kramer (Trey Ideker's lab, University of California San Diego)	"Inferring gene ontologies from molecular networks"
12-1:30pm	*Lunch*	
1:30-2:20pm	KT2: Roded Sharan (Tel Aviv University)	"Protein networks: from topology to logic"
2:20-2:50pm	CT2: Rim Zaag (Plant Genomics Research, Evry, France)	"Global analysis of co-regulation for the identification of functional modules"
2:50-3:30pm	IT3: Teresa Przytycka (US National Institutes of Health)	"Towards systems level analysis of tumor heterogeneity"
3:30-4pm	CT3: Yuriy Hulovatyy (University of Notre Dame)	"Integrative, dynamic, and comparative biological network research of aging"
4-4:20pm	*Coffee break*	
4:20-5pm	IT4: Jan Baumbach (University of Southern Denmark)	"Integrating multiple omics data types with biological networks"
5-5:50pm	KT3: Alfonso Valencia (Spanish National Cancer Research Centre)	"Building an Epigenetic Network with Co-Evolutionary Methods"
5:50-6pm	Nataša Pržulj (Imperial College London) Tijana Milenković (University of Notre Dame)	Closing remarks

ECCB14 web site http://www.eccb14.org/program/workshops/idalbnd

W08: Workshop on Logical Modelling and Analysis of Cellular Networks

Description
Since the seminal publications of Stuart Kauffman and René Thomas about 40 years ago, logical modelling has been increasingly used to model the dynamical behaviour of biological regulatory networks. However, behind the common denomination "logical modelling", computational biologists are using different formalisations (Boolean, multilevel,

deterministic, stochastic, etc.) and have developed various methods and tools, which might be somewhat difficult to choose from to address specific biological questions.
This workshop proposal aims at providing an extensive and organised overview of recent developments that render possible the modelling and the analysis of large cellular networks, encompassing various signalling pathways, sophisticate transcriptional networks, as well as novel regulatory mechanisms (e.g. regulatory effects of miRNA, epigenetic regulations, etc.).
Speakers will refer to applications dealing with the control of cell proliferation, cell differentiation and reprogramming, embryonic development, immune response, and cancer drug resistance. They will emphasise the biological insights gained with logical modelling in general, as well as by using specific analysis methods.

Organizers

Denis Thieffry (Ecole Normale Supérieure, France) ; Ioannis Xenarios (Swiss Institute of Bioinformatics, Switzerland)

Scientific committee

Bockmayr Alexander ; Chaouiya Claudine ; Fages François ; Helikar Tomas ; Saez-Rodriguez Julio ; Thieffry Denis ; Xenarios Ioannis.

Program

9h00-10h30 **Chair: Ioannis Xenarios (Lausanne, Switzerland)**
- Denis Thieffry (Paris, France): *Introduction*
- Bornholdt Stefan (Bremen, Germany): *Boolean threshold networks as a modeling tool in systems biology.*
- Tomas Helikar (Lincoln, Nebraska, USA): *A collaborative and crowd-sourcing approach to the construction and analysis of logical models.*

10h30-11h00 - *Coffee break*
11h00-12h30 **Chair: Claudine Chaouiya (Oeiras, Portugal)**
- Reinhard Laubenbacher (Blacksburg, Virginia, USA): *Steady analysis of Boolean network models through model reduction and computational algebra.*
- Reka Albert (State College, Pennsylvania, USA): *Logic network allows the determination and control of cellular attractors.*
- Heike Siebert (Berlin, Germany): *Logical network analysis with symbolic steady states.*

12h30-14h00 *Lunch break*
14h00-15h30 **Chair: Tomas Helikar (Lincoln, Nebraska, USA)**
- Siegel Anne (Rennes, France): *Improving robustness in the study of logical model of signalling networks with answer set programming.*
- Julio Saez-Rodriguez (Hinxton, UK): *Logic models of signalling networks from phosphoproteomic data.*
- Gautier Stoll (Paris, France): *Continuous time Boolean modeling with MaBoSS software: algorithm and applications.*

15h30-16h - *Coffee break*
16h00-17h30 **Chair: Anne Siegel (Rennes, France)**
- Wassim Abou-Jaoude (Paris, France): *Logical modeling of T-helper cell differentiation and plasticity.*
- Louis Filipo Fitime (Nantes, France): *Integrating time-series data on large-scale cell-based models: application to skin differentiation.*
- Adrien Fauré (Yamaguchi, Japan): *A discrete model of Drosophila eggshell patterning reveals cell-autonomous and juxtacrine effects.*

17h30-18h - Chair: Denis Thieffry (Paris, France) - *General discussion and conclusions*

ECCB14 web site http://www.eccb14.org/program/workshops/lmacn

W09: Workshop on Machine Learning for Systems Biology

Description

Molecular biology and all the biomedical sciences are undergoing a true revolution as a result of the emergence and growing impact of a series of new disciplines/tools sharing the "-omics" suffix in their name. These include in particular genomics, transcriptomics, proteomics and metabolomics, devoted respectively to the examination of the entire systems of genes, transcripts, proteins and metabolites present in a given cell or tissue type.
The availability of these new, highly effective tools for biological exploration is dramatically changing the way one performs research in at least two respects. First, the amount of available experimental data is not a limiting factor any more; on the contrary, there is a plethora of it. Given the research question, the challenge has shifted towards identifying the relevant pieces of information and making sense out of it (a "data mining" issue). Second, rather than focus on components in isolation, we can now try to understand how biological systems behave as a result of the integration and interaction between the individual components that one can now monitor simultaneously (so called "systems biology").
Taking advantage of this wealth of "omics" information has become a condition sine qua non for whoever ambitions to remain competitive in molecular biology and in the biomedical sciences in general. Machine learning naturally appears as one of the main drivers of progress in this context, where most of the targets of interest deal with complex structured objects: sequences, 2D and 3D structures or interaction networks. At the same time bioinformatics and systems biology have already induced significant new developments of general interest in machine learning, for example in the context of learning with structured data, graph inference, semi-supervised learning, system identification, and novel combinations of optimization and learning algorithms.
MLSB14, the Eighth International Workshop on Machine Learning in Systems Biology, is a workshop of the ECCB 2014 conference. It aims to contribute to the cross-fertilization between the research in machine learning methods and their applications to systems biology by bringing together method developers and experimentalists. We are soliciting

submissions bringing forward methods for discovering complex structures (e.g. interaction networks, molecule structures) and methods supporting genome-wide data analysis
Please see the workshop website http://www.mlsb.cc for more details.

Organizers
Florence d'Alché-Buc (University of Evry, France) ; Pierre Geurts (University of Liege, Belgium).

Organizing committee
Florence d'Alché-Buc (University of Evry, France) ; Markus Heinonen (University of Evry, France) ; Pierre Geurts (University of Liege, Belgium) ; Vân Anh Huynh-Thu (University of Edinburgh, UK).

Program
Saturday, September 6
9am-10:30am Session 1
Introduction
Invited Talk: Pierre Baldi, UCI University, *Carbon-Based Computing Vs Silicon-Based Computing: A New Theory of Circadian Rhythms*.
Daniel Trejo-Banos: *Structural inference in oscillatory networks: a case study of the Arabidopsis Thaliana circadian clock*.
10:30am-10:45am *Coffee Break*
10:45am-12pm Session 2
Xin Liu: *Parameter Estimation in Computational Biology by Approximate Bayesian Computation coupled with Sensitivity Analysis.*
Van Anh Huynh-Thu: *A hybrid approach for the inference and modelling of gene regulatory networks*.
Karel Jalovec: *Classification of metagenomic samples using discriminative DNA superstrings (short talk)*.
12pm-1:30pm *Lunch*
1:30pm-3:30pm Session 3
Invited Talk: Nicola Segata, University of Trento, *Machine learning challenges in computational meta'omics*.
Eugen Bauer: *Metabolic Meta-Reconstruction and Community Modeling of Intestinal Microbes.*
Aalt van Dijk: *Interspecies Association Mapping: connecting phenotypes to sequence regions across species*.
3:30pm-4:30pm Poster Session
4:30pm-6:00pm Session 4
Invited Talk: Anne-Laure Boulesteix, LMU Munich, *Statistical testing and variability in real-data-based benchmark experiments for supervised learning methods.*
Adrien Dessy: *Computationally Efficient Test for Gene Set Dysregulation (short talk)*.
Sohan Seth: *Differential analysis of whole-genome shotgun sequences (short talk)*.

Sunday, September 7
9:00am-10:30am Session 5
Invited Talk: Jean-Loup Faulon, CNRS, *Using Machine Learning in Synthetic Biology: The Design-Build-Test and Learn cycle.*
Tom Mayo: *M3D: a kernel-based test for shape changes in methylation profiles.*
10:30am-10:45am *Coffee Break:*
10:45am-12pm Session 6
Pooya Zakeri: *Application of Geometric Kernel Data Fusion in Protein Fold Recognition and Protein Sub-nuclear Localization.*
Yawwani Gunawardana: *Outlier-Detecting Support Vector Regression for Modelling at the Transcriptome-Proteome Interface*.
Olivier Poirion: *Structuration of the bacterial replicon space (short talk)*.
12pm-1:30pm *Lunch*
1:30pm-3:30pm Session 7
Invited Talk: Karsten Borgwardt, ETH Zürich, *Machine Learning for Personalized Medicine.*
Anna Cichonska: *Meta-analysis of Genome-Wide Association Studies with Multivariate Traits*.
Roland Barriot: *Semi-automatic Validation of Genome-wide Reassembled Systems by Gene Prioritization through Weighted Data Fusion*.
3:30pm-4:20pm Poster Session with coffee
4:20pm-5:20pm Session 8
Invited Talk: Katheen Marchal, KU Leuven & U Ghent, TBA
Closing Remarks

ECCB14 web site http://www.eccb14.org/program/workshops/mlsb

Workshop website http://www.mlsb.cc

W11: The 1st Unified Workshop on Proteome and Metabolome Informatics

Description
At present there is little contact between metabolomics and proteomics fields at both wet and dry-lab levels, and so key synergies are being lost. A number of packages offer the complete quantification and identification workflow for discovery proteomics and metabolomics but there is little cross-fertilisation between metabolome and proteome informatics groups, despite numerous overlaps and similarities between disciplines. In some ways this is because metabolome and proteome informatics research has originated from different fields (broadly chemometrics and genome bioinformatics, respectively), yet their distinct perspectives have been applied to identical or similar problems. We therefore believe there is a timely opportunity to bring together the informatics communities in metabolomics and proteomics:

- To underpin existing and to drive new cross-disciplinary collaborations in order to facilitate better statistical integration and therefore better downstream modelling and ultimately biological understanding.
- To allow these disciplines to better 'borrow strength' from each other, improving quantitative workflows at all levels.
- To impact on researcher mobility due to clearer understanding of the commonality and subtle differences in workflows between disciplines.
- To support development of joint data exchange and reporting standards for optimal integration of omics data.

This workshop will be of interest to all researchers in proteomics or metabolomics, whether at the wet-lab or informatics level. It will also be of relevance to systems modellers and network biologists who integrate proteomics and metabolomics data and who wish to gain a greater understanding of how upstream processing of this data is performed.

Organizers

Dr Andrew Dowsey, CADET, University of Manchester ; Dr Simon Rogers, School of Computer Science, University of Glasgow ; Prof Rainer Breitling, MIB, University of Manchester ; Dr Richard Unwin, CADET, University of Manchester.

Program

8:00 Registration
9:00 Welcome - Andrew Dowsey
9:05 *Introduction to the new HUPO and ISCB Computational Mass Spectrometry Working Groups* - Oliver Kohlbacher, University of Tübingen, Germany.

Session 1

9:15 "*From spectra to biology - Joint analysis of proteomics and metabolomics data with OpenMS*", Oliver Kohlbacher, University of Tübingen, Germany.
9:40 "*mzMatch: software for metabolomics workflow design and data analysis*" - Andris Jankevics, University of Manchester, UK.
10:05 "*Galaxy: A viable platform for analysing mass spectrometry data?*" - Conrad Bessant, Queen Mary University of London, UK.
10:30 *Coffee break*

Session 2

10:45 "*Progress in PROCESS for proteomics and metabolomics data standards*" - Simon Perkins, University of Liverpool, UK.
11:10 "*World-wide data exchange in Metabolomics*" - Christoph Steinbeck, European Bioinformatics Institute, UK.
11:35 "*Revisiting hypothetical proteins: Combining 'omic' approaches for better functional prediction*" - Prashanth Suravajhala, Bioclues.org and Bioinformatics.Org.
12:00 *Lunch*
12:30 Poster presenters in attendance

Session 3

13:30 "*Proteomics Data, Functional Analysis and Dissemination*" - Henning Hermjakob, European Bioinformatics Institute, UK.
13:55 "*Advances in functional mixed modelling applied to LC-MS data*" - Jeffrey S Morris, The University of Texas MD Anderson Cancer Center, USA.
14:20 "*Spatial metabolomics using imaging mass spectrometry*" - Theodore Alexandrov, University of Bremen / University of California San Diego / SCiLS.
14:45 "*Absolute Murder: Informatics challenges for absolute quantitation in proteomics*" - Simon Hubbard, University of Manchester, UK.
15:10 "'*The importance of metabolite annotation – current software and next steps*" - Rick Dunn, University of Birmingham, UK.
15:35 "*Tsunami or alluvial plain? A primer on dealing with public data*", Lennart Martens, Universiteit Gent, Belgium.
16:00 Coffee break

Session 4

16:20 "*Stronger findings from mass spectral data through hierarchical Bayesian multi-peak modeling*" - Tommi Suvitaival, Helsinki Institute for Information Technology HIIT, Finland.
16:45 "*A Comparative MS/MS Based Workflow for the Detection of Stable Protein Adducts Induced by Small Molecules*" - Markus Müller, Swiss Institute of Bioinformatics, University of Geneva, Switzerland.
17:10 "*Critical assessment of the elemental isotope definition in mass-spectrometry-based proteomics*" - Jürgen Claesen, I-BioStat, UHasselt, Belgium.
17:35 "*Metabolite identification via multiple kernel learning*" - Huibin Shen, Department of Information and Computer Science, Aalto University, Espoo, Finland.
18:00 Close

Workshop web site http://www.cadetbioinformatics.org/workshops/eccb14/

W12: Recent Computational Advances in Metagenomics

Description

This workshop aims at promoting discussions and collaborations between biologists (modelers), computer scientists and applied-mathematicians involved in metagenomics and/or metatranscriptomics studies, either in the bioinformatics or statistical aspects of such analysis.
Metagenomics studies refer to analyses based on high-throughput sequencing of environmental samples and microbial ecosystems. Both marker-gene (16S, 18S, ITS, ...) and whole-genome strategies will be addressed to cover a wide array of question ranging from quantifying the microbial diversity in order to find structuring factors to assessing

the functional role of microbial communities.
The workshop will provide an overview of the state-of-the-art methods currently used in metaomics including comparative metagenomics and metatranscriptomics. At the other end of the spectrum, case-studies will illustrate how these methods produce insightful biological knowledge.

Organizers
Sophie Schbath (INRA/MIG, Jouy-en-Josas, France) ; Valentin Loux (INRA/MIG, Jouy-en-Josas, France) ; Mahendra Mariadassou (INRA/MIG, Jouy-en-Josas, France).

Program
09:00 - 09:10 Welcome
09:10 - 10:10 Keynote Steven Kembel (University of Québec at Montréal): "*Using metagenomics to model community assembly in the plant microbiome*"
10:10 - 10:45 Pierre Péricard: "*SortMeRNA 2: ribosomal RNA classification for taxonomic assignation*"
10:45 - 11:15 *Coffee break*
11:15 - 11:50 Simon Foucart: "*Quikr & WGSQuikr: Rapid Bacterial Community Reconstruction Via Compressive Sensing*"
11:50 - 12:25 Kévin Vervier: "*Towards Large-scale Machine Learning for Metagenomics Sequence Classification*"
12:25 - 13:00 Frédéric Mahé: "*Swarm: robust and fast clustering method for amplicon-based studies*"

13:00 - 14h15 *Lunch*
14:14 - 15:15 Keynote Aaron Darling (University of Technology, Sydney): "*Toward resolving the fine scale genetic structure of microbial populations: a metagenomic Hi-C approach*"
15:15 - 15:50 Edi Prifti: "*Quantitative metagenomics: form reads to biomarkers*"
15:50 - 16:20 *Coffee break*
16:20 - 16:55 Frédéric Plewniak: "*Metagenome-scale metabolic network reconstruction*"
16:55 - 17:30 Clovis Galiez: "*Identifying distant homologous viral sequences in metagenomes using protein structure information*"
17:30 Closing remarks

ECCB14 web site http://www.eccb14.org/program/workshops/rcam

W13: sbv IMPROVER Workshop

Description
During this workshop participants will be given the opportunity to learn more about using the crowd engagement methods to advance research. In the first part of the workshop we will focus on the design and results of our first two challenges (Case Study 1).
In the second part, we will introduce the ongoing Network Verification Challenge and give the workshop participants the opportunity to experience our network verification platform. This will lead to an open discussion about the platform and how methods to visualize and interact with biological data can be further developed to improve functionality for the user (Case Study 2).
In the third part of the workshop we will look into other analysis and visualization tools and their capability to deal with data sets from different fields of research, e.g. data sets from metabolomics, proteomics, transcriptomics or genomics. The key discussion here will focus on finding optimal ways, both to integrate different datasets into the analysis, and to be able to visualize results in a comprehensive and intuitive manner (Case Study 3).

Organizers
Manuel Peitsch, Ph.D. (Philip Morris International R&D, Vice President Biological Systems Research) ; Julia Hoeng, Ph.D. (Philip Morris International R&D, Manager Systems Toxicology) ; William Hayes, Ph.D. (Selventa, Chief Technical Officer) ; Jennifer Park, Ph.D. (Selventa, Associate Director Research).

Program
09:00 - 09:15 Introduction of Session Chair (Manuel Peitsch, Ph.D., Philip Morris International, Switzerland).
09:15 - 09:45 Keynote: *Diagnostic Signature Benchmarking* (Heinz Koeppl, Ph.D., TU Darmstadt, Germany).
09:45 - 10:15 *Coffee Break*
10:15 - 10:45 Keynote: *Network Verification Challenge* (Jennifer Park, Ph.D., Selventa, USA).
10:45 - 11:15 Keynote: *Quantification of miRNA-mRNA interactions* (Angel Rubio, Ph.D. University of Navarra, Spain).
11:15 - 11:45 Keynote: *A semiautomatic extraction process for causal network generation* (Juliane Fluck, Ph.D., Fraunhofer Institute SCAI, Germany).
11:45 - 12:30 Case study: Mini network modeling jamboree moderated by Selventa.
12:30 - 13:30 *Lunch Break*
13:30 - 14:00 Keynote: *Integrating public -omics data to leverage the analysis of your own experiments: challenges and opportunities* (Philip Zimmermann, Ph.D., NEBION, Switzerland).
14:00 - 14:30 Keynote: *Landscape of DNA sequences specifically binding transcription factors* (Vsevolod Makeev, Ph.D., Vavilov Institute of General Genetics, Russian Academy of Sciences, Russia).
14:30 - 15:00 *Coffee Break*
15:00 - 15:45 Keynote*: Integrative omics analysis and computational approaches* (Joaquin Dopazo, Ph.D., Principe Felipe Research Center, Spain).
15:45 - 16:30 Keynote: *GARUDA alliance and gateway* (Samik Ghosh, Ph.D., The Systems Biology Institute, Japan).
16:30 - 17:00 Keynote: *Systems pharmacology modeling of disease development and progression* (Oleg Demin, Ph.D., Institute for Systems Biology Moscow, Russia).
17:00 End of Workshop

ECCB14 web site http://www.eccb14.org/program/workshops/sbvi

W16: Tools and Techniques for Analysis and Design of Macromolecular Structures

Description

In this 1-day workshop, we will present techniques to analyze, design, and optimize models of proteins, DNA, RNA, and their complexes for scientific and industrial applications. The workshop will comprise two parts, a lecture session and hands-on exercises with participants laptops. Invited speakers will talk about new approaches to structure determination of macromolecules and macromolecular complexes. During the practical sessions participants will learn how to use methods for modelling protein interactions and 3D structures of RNA and protein-RNA complexes, e.g. from low-resolution density maps and biochemistry.

Why you should attend this workshop?

Several fast-growing fields in academic and industrial research are centered around nanotechnology and macromolecular analysis and design. Many protein and RNA complexes (e.g. the ribosome) are known structurally, but their dynamics are not well understood. RNA Origami structures to deliver drugs are under development. There is a boom in antibodies and alternative scaffolds, which are designed to bind proteins in human patients for therapy and diagnosis. Enzymes can be engineered for stability and specificity. In many of these applications, development based only on experiments is possible, but very laborious and often unreliable. On the other hand, the quality of computer methods has improved to the point that a non-specialist can get quantitative predictions from modeling which can save months of time in the laboratory. Participants in this workshop will learn to use these tools to become more effective modelers, or to complement their experimental work with modeling.

Organizers

Samuel C. Flores (Uppsala University, Sweden) ; Grzegorz Chojnowski (International Institute of Molecular and Cell Biology, Poland).

Program

9:00 Opening remarks
9:10 Eric Westhof, "*RNAPuzzles: Recent comparisons between Predictions and Structures*"
9:30 Mark Bathe, "*Computational Design Principles for Functional Nucleic Acid Nanostructures*"
9:50 Juan Fernández-Recio, "*Structural prediction of protein-protein complexes: are we ready for modeling the interactome?*"
10:10 Samuel Flores, "*Evaluating the effect of mutations in protein-protein interfaces with a multiscale approach*"
10:30 *Coffee*
11:00 Grzegorz Chojnowski (Bujnicki Lab) "*RNA Bricks - a database of RNA 3D motifs and their interactions*"
11:20 Francois Major, "*Structural dynamics control the microRNA maturation pathway*"
11:40 João Rodrigues (Bonvin Lab), "*Integrative Modeling of Protein Interactions*"
12:00 *Lunch*

Tutorials

1:00 *Using HADDOCK to model protein interactions* (João Rodrigues)
2:00 *Predict RNA 2D and 3D structure over the Internet using MC-Tools* (Francois Major)
3:00 *Modeling Large Macromolecular Complexes with PyRy3D* (Grzegorz Chojnowski)
4:00 *MMB (MacroMoleculeBuilder) tutorial* (Samuel Flores)

ECCB14 web site http://www.eccb14.org/program/workshops/ms

Tutorials

T01: Analysis of Cis-Regulatory Motifs from High-Throughput Sequence Sets

Description
Next Generation Sequencing led to the development of novel methods (ChIP-seq, FAIRE-seq, CLIP-seq) to acquire massive data about diverse signals involved in genome regulation and function (cis-regulation, chromatin conformation, RNA maturation, recombination, replication, …). Extracting relevant information from the raw data requires not only specialized software tools, but also a good understanding of their principles and parameters.
In this tutorial we will demonstrate **how to analyse ChIP-seq data using the Regulatory Sequence Analysis Tools (http://www.rsat.eu/)**, via their different ways of access: web-interface, command-line, and web-services.
This tutorial aims at introducing the theoretical principles and giving practical skills to use specialized software tools to **extract motifs** from full-scaled NGS datasets (typically covering tens of Mb).

Topics covered
Motif discovery & motif enrichment in peak sequences.
Impact of the background models.
Building control sets to estimate the rates of false positives.
Comparison between motifs.
Evaluating the quality of motifs extracted from NGS peaks.

Organizers
Jacques van Helden (Aix-Marseille Université, France), Morgane Thomas-Chollier (Ecole Normale Supérieure, Paris, France).

Co-teachers: Alejandra Medina-Rivera (SickKids Research Institute, Toronto, Canada), Carl Herrmann (IPMB University Heidelberg and DKFZ - B080, Heidelberg, Germany).

Program
08:00-09:00 *Registration and configuration of laptops (network access)*
09:00-10:30 Session 1: Analysis of cis-regulatory elements with the Regulatory Sequence Analysis Tools (RSAT): methods and website utilization (http://www.rsat.eu/).
10:30-10:45 *Coffee break*
10:45-12:00 Session 1 (continued)
12:00-13:30 *Lunch*
13:30-16:00 Session 2: Using RSAT in command-line (Unix terminal via ssh connection)
16:00-16:20 *Coffee break*
16:20-18:00 Session 3: Writing client scripts for RSAT web services.

Requirements
The course is addressed to bioinformaticians and biologists. The afternoon will require being familiar with the Unix shell (Linux) and basic programming skills (Perl or Python), in order to implement clients for web services.
To follow the practical sessions, participants are expected to bring a laptop with wireless capability. Any operating system is adequate (Linux, Mac OSX, Windows). The morning session will rely on classical use of Web site via Web browsers, and the afternoon session will require the capability to establish an ssh connection. A more detailed description of software requirement will be communicated to participants one week before the tutorial.

ECCB14 web site: http://www.eccb14.org/program/tutorials/cis-r

T02: Computational Tools to Define and Analyse Logical Models of Cellular Networks

Description
The logical framework is being increasingly used to model the dynamical behaviour of biological regulatory networks. In this context, a range of computational tools have been developed to support model definition and analysis. These tools reflect the diversity of formalisations covering “logical modelling” (Boolean, multilevel, deterministic, stochastic, etc.). This tutorial aims at providing an overview of this diversity, demonstrating the complementarity between existing computational tools made possible by a novel exchange format --Systems Biology Markup Language Qualitative Models (SBML qual). Instructors will provide an overview of existing technology powering efficient qualitative modelling of regulatory and signalling networks. Real biological examples will be used in the proposed tutorial session.

Organizer(s)
Claudine Chaouiya (Instituto Gulbenkian de Ciência, Oeiras Portugal), Tomas Helikar (University of Nebraska-Lincoln, USA), Julio Saez-Rodriguez (EMBL-EBI, UK).

Program
13:30-14:20 Introduction and overview
14:20-15:10 CellNOpt

15:10:16:00 GINsim
16:00-16:20 *Coffee break*
16:20-17:10 The Cell Collective
17:10-18:00 Conclusions - discussions

ECCB14 web site: http://www.eccb14.org/program/tutorials/lmcn

T03: IMGT, the Global Reference in Immunogenetics and Immunoinformatics

Description

IMGT®, the international ImMunoGeneTics information system® (http://www.imgt.org) was created in 1989 by the Laboratoire d'ImmunoGénétique Moléculaire (LIGM) of the Professors Marie-Paule Lefranc and Gérard Lefranc (University Montpellier 2 and CNRS) at Montpellier, France. IMGT® is the global reference in immunogenetics and immunoinformatics and is the first and, up to now, the only integrated information system in immunogenetics and immunoinformatics.

The aim of the tutorial is to familiarise biologists, bioinformaticians and computer scientists involved in immunogenetics and immunoinformatics research projects with the IMGT unique approach which bridges the gap from genes to 3D structures for immunoglobulin and T cell receptor repertoire analysis from fishes to humans, and for antibody engineering and humanization.

This tutorial will start with an overview of the IMGT® system built on the IMGT-ONTOLOGY axioms and concepts, and of its major databases. The main topics will be presented from sequences to structures and exemplified with IMGT/V-QUEST and IMGT/JunctionAnalysis for the analysis of nucleotide sequences and the high-throughput version IMGT/HighV-QUEST for next generation sequencing (NGS), IMGT/DomainGapAlign for amino acid sequences, IMGT/3Dstructure-DB for 3D structures, contact analysis and paratope/epitope interactions, and the IMGT/mAb-DB interface with access to therapeutical antibodies data.

Organizer

Marie-Paule Lefranc, IMGT founder and director, Institut de Génétique Humaine (IGH), UPR CNRS 1142, Université de Montpellier, e-mail : Marie-Paule.Lefranc@igh.cnrs.fr , http://www.imgt.org

Program

Session 1. IMGT immunoinformatics for repertoire analysis of the adaptive immune response

Chairperson : Véronique Giudicelli, PhD, IMGT Bioinformatics manager (Université de Montpellier, IGH CNRS, Montpellier, France)

- Marie-Paule Lefranc : IMGT®, the international ImMunoGeneTics information system® : IMGT-ONTOLOGY. Bridging the gap between genes (nomenclature), sequences (labels) and three-dimensional structures (IMGT unique numbering and IMGT Collier de Perles).
- Véronique Giudicelli : Analysis of immunoglobulin (IG) or antibody and T cell receptor (TR) rearranged nucleotide sequences : IMGT/V-QUEST, IMGT/JunctionAnalysis and IMGT/HighV-QUEST (presentation, demos, tests and discussion).

Coffee Break

Session 2. IMGT immunoinformatics for therapeutic antibodies

Chairperson : Patrice Duroux, PhD, IMGT Informatics manager (IGH CNRS, Montpellier, France)

- Marie-Paule Lefranc : Analysis of amino acid sequences and structures of antibodies : IMGT/DomainGapAlign and antibody humanization, IMGT/3Dstructure-DB and antigen receptor-antigen interactions (IG/Ag and TR/pMH), IMGT/mAb-DB for therapeutic antibodies.
- Patrice Duroux : Demos, tests and discussion.

ECCB14 web site: http://www.eccb14.org/program/tutorials/imgt

T04: Multivariate projection methodologies for the exploration of large biological data sets. Application in R using the mixOmics package

Description

The objective of this tutorial is to introduce the fundamental concepts behind projection-based approaches and illustrate their application on some exemplar studies using the R package mixOmics.

Multivariate projection approaches are useful exploratory tools to get a first understanding of large and complex data sets. These approaches are extremely efficient on large data sets, and can also answer complex questions. Such approaches include Principal Component Analysis (PCA, Joliffe 2002) and other variants, Partial Least Squares regression (PLS, Wold 2001), PLS-Discriminant Analysis, Canonical Correlation Analysis (CCA, Hotelling 1936). These approaches enable dimension reduction by projecting the data into a smaller subspace. Recent developments proposed the so-called `sparse' approaches, which include Lasso penalisations to allow variable selection (Tibshirani 2001).

PCA is the oldest and most popular multivariate technique but often, little is known about how this approach is solved and what are the limitations. More sophisticated approaches like PLS and CCA have recently been extended to deal with the large dimension (sparse PLS, or regularized CCA) and were proven to bring biologically meaningful results in many studies. Contrary to PCA, PLS and CCA enable the integration of two types of data sets.

Since 2009, we have implemented several multivariate approaches and their sparse variants in the R package mixOmics to be used by the statistical and bioinformatics community. Full tutorials are given on our website: http://perso.math.univ-toulouse.fr/mixomics/

In this tutorial, we will focus on the application of these approaches to medium and high throughput biological data

(transcriptomics, metabolomics, proteomics data) using PCA, CCA, PLS, PLS-DA and the variants that the mixOmics team and collaborators have developed.

Organizers and presenters

Dr Kim-Anh Lê Cao (The University of Queensland Dimantina Insitute, Brisbane, Australia), Dr Sébastien Déjean (Institut de Mathématiques de Toulouse, Université de Toulouse, France), Dr Ignacio González (Institut de Mathématiques de Toulouse, Université de Toulouse, Institut National de la Recherche Agronomique, France). Contact the mixOmics team: mixomics@math.univ-toulouse.fr

Program

9am – 5.30pm (registration starts from 8 am)

Morning

1. Principal Component Analysis

Concepts
Interpreting outputs
Application in mixOmics on some biological data sets

2. Canonical Correlation Analysis

Concepts and limitations
Application in mixOmics on some biological data sets

Afternoon

1. Partial Least Squares regression

Concepts about PLS and sparse PLS
Application in mixOmics on some biological data sets

2. PLS – Discriminant Analysis

Concepts about PLS-DA and sparse PLS-DA
Application in mixOmics on some biological data sets

ECCB14 web site: http://www.eccb14.org/program/tutorials/mixomics

T05: Protein Evolution Analysis: on the Use of Phylogenetic Trees

Description

Homologous proteins, that share a common ancestor, can be classified into families. These homologs can be orthologs, that were separated by a speciation event, or paralogs, that were separated by a duplication event. Within a protein family, all members are related by a phylogenetic tree, which consists of a root (the last common ancestor of the protein family), nodes (which are speciation/duplication events), branches (whose lengths correspond to the number of substitutions) and tips (which correspond to modern sequences). The tree is helpful for inferring the evolutionary history of the protein family. For example, we can reconstruct the ancestral sequences at each node of the tree. These ancestral sequences can be used for homology modelling, to reveal the ancestral 3D structures, or synthesised in vitro. Or we can compare trees to reveal similar evolutionary history between protein families (co-evolution). Manipulating tree topologies are complex operations that require tools to perform operations such as reading, pruning, collapsing, re-rooting. These operations can be done with programs with graphical user interfaces (GUI). However, in the area of large-scale data, in which hundreds or thousands of trees may be manipulated, it is impractical to use such programs. To this end, new software/libraries have been developed to deal with such large data sets in an automated manner.

This tutorial will present recent concepts regarding the evolution and adaptation of protein sequences. It will be divided into three sections, in which we will present methods relating to the use of phylogenetic trees to infer protein function. These sections will be 1) using scripts to manipulate trees, 2) using ancestral sequence reconstruction to infer history of a protein family and 3) the detection of coevolution between protein families. Each section will have an introduction explaining the concepts underlying any analysis methods, and a discussion of the power and limitations of different methods and tools used to explore these concepts and which participants will learn how to use during the practical for that section.

Organizer(s)

Brandon Invergo is a post-doctoral fellow at the European Bioinformatics Institute (EMBL-EBI) and the Sanger Institute, David Ochoa is a post-doctoral fellow in the laboratory of Pedro Beltrao at the European Bioinformatics Institute, Romain Studer is a senior post-doc research scientist in the laboratory of Dr. Pedro Beltrao, EMBL-EBI.

Program

Morning session: Performing phylogenetic analyses with Biopython.
9:00 Talk (45min): Performing phylogenetic analyses with Biopython (B. Invergo, EBI)
9:45 Practical (45min): Performing phylogenetic analyses with Biopython (B. Invergo, EBI)
10:30 *Coffee break*
10:45 Practical (1h15): Performing phylogenetic analyses with Biopython (B. Invergo, EBI)
12:00 *Lunch*
Afternoon session: ancestral sequence reconstuction and molecular co-evolution.
13:30 Talk (30min): Ancestral sequence reconstruction (R. Studer, EBI)
14:00 Practical (1h30): Ancestral sequence reconstruction (R. Studer, EBI)
15h30: Talk (30 min): Studying molecular co-evolution (D. Ochoa, EBI)
16h:00 *Coffee break*
16h20: Practical (50min): Studying molecular co-evolution (D. Ochoa, EBI)
17:30 *End*

ECCB14 web site: http://www.eccb14.org/program/tutorials/pea

T06: Reuse, Develop and Share Biological Visualisation with BioJS

Description
BioJS is a community project aiming to create a collection of JavaScript components to present biological information following a common guideline. This workshop aims to introduce the BioJS project and provide enough skills to use BioJS components. For people interested in contributing to BioJS we will also provide a hello world tutorial to quickly have a taste on how to create a BioJS component.

Organizers
Manuel Corpas, The Genome Analysis Centre, UK, Rafael Jimenez, ELIXIR Hub, Hinxton, UK.

Program
9:00 - 9:45 -- Introduction to the BioJS project.
9:45 - 10:15 -- *Coffee break*
10:15 - 10:30 -- Developer environment, download and install from BioJS GitHub repository, quick check.
10:30 - 11:00 -- How to use a BioJS component.
11:00 - 11:30 -- How to integrate multiple BioJS components.
11:30 - 11:45 -- BioJS registry, will show how to use the BioJS registry, how it is used and its benefits.
11:45 - 12:00 -- Creating a BioJS development environment.
12:00 - 12:30 -- A biological version of "Hello World" component.
12:30 - 1:30 -- *Lunch*
1:30 - 3:00 -- Design exercise. Here students work on an interesting
but simple example to be done by groups, with pen and paper. We would require big paper sheets for further discussion and creative input.
3:00 - 3:30 -- *Coffee break*
3:30 - 4:30 – Developing a BioJS component.
4:30 - 5:00 -- Wrap up. How to become involved in the BioJS community.

ECCB14 web site: http://www.eccb14.org/program/tutorials/biojs

T07: Scientific Workflows for Analysing, Integrating and Scaling Bioinformatics Data: a Practical Introduction to Galaxy, Taverna and WS-PGRADE

Description
This tutorial will provide an introduction to the use of scientific workflows for bioinformatics data analysis, integration and scaling and will highlight the main differences and similarities between the myriad of systems available. Examples and comparisons will be illustrated based on WS-PGRADE, Taverna and Galaxy as representatives of grid workflows, distributed computing workflows and local tools workflows respectively. Case studies from large-scale transcriptome analysis projects will be used to motivate the technical comparison and demonstrate where the use of workflows can bring large advantages to researchers. Practical exercises will provide basic hands-on experience with those three systems.

Organizers
Katy Wolstencroft, Leiden Institute of Advanced Computer Science, University of Leiden, The Netherlands, k.j.wolstencroft@liacs.leidenuniv.nl ; Leon Mei, Sequencing Analysis Support Core, Leiden University Medical Center, The Netherlands, h.mei@lumc.nl ; Silvia Delgado Olabarriaga, Academic Medical Center, University of Amsterdam, The Netherlands, s.d.olabarriaga@amc.uva.nl .

Program
Session 1
- Round introduction
- Introduction to using scientific workflows in bioinformatics (Katy Wolstencroft)
- Introduction to Galaxy with demo (Leon Mei)

Coffee break
Session 2
- Hands-on practical Galaxy on RNAseq analysis (Leon Mei)
- Introduction to Taverna - distributed tools and services for downstream enrichment analysis (Katy Wolstencroft)

Lunch
Session 3
- Hands-on taverna (Katy Wolstencroft)
- Introduction to WS-PGRADE - tapping into Grid and Cloud resources for analyses requiring heavy computation (Silvia Delgado Olabarriaga)
- Hands-on ws-pgrade (Silvia Delgado Olabarriaga)

Coffee break
Session 4
- Discussion: comparison between workflow system properties (All)

recap of system properties for Galaxy, Taverna and WS-PGRADE;
brief discussions in groups

- Wrap-up and conclusions (All)
- Q&A: hands-on or bring your own workflow problems (All)

ECCB14 web site: http://www.eccb14.org/program/tutorials/workflows

T08: Statistics and Numerics for Dynamical Modeling

Description

A successful mathematical description of cell biological processes based on experimental data requires efficient and reliable numerical methods for parameter estimation as well as a suitable statistical methodology to reconstruct the underlying biochemical reaction networks. In this tutorial, statistical and numerical aspects for dynamic modeling in Systems Biology are discussed and the computational implementation is demonstrated. One major focus is the assessment of uncertainties of both, parameters and model predictions, which is efficiently and intuitively judged by the profile likelihood.

In summary, the following aspects are discussed:

- appropriate numerical algorithms for parameter estimation
- judging the quality of experimental data
- model discrimination by the likelihood ratio tests
- identifiability analysis and confidence intervals for estimated parameters
- observability analysis and confidence intervals for model predictions
- selection of informative new experimental conditions

In the tutorial, we use a basic system of ordinary differential equations (ODEs) to implement and illustrate the methods in MATLAB together with the participants. In addition, a published model for JAK-STAT signaling and a comprehensive software package for quantitative dynamic modelling is introduced. The illustrated methodology has been awarded three times as “Best Performer” in parameter estimation challenges within the Dialogue for Reverse Engineering Assessment and Methods (DREAM) competitions.

Organizers

Bernhard Steiert and Clemens Kreutz, university of Freiburg, Germany. Contact: clemens.kreutz@fdm.uni-freiburg.de

Program

9:00 – 10:00 Introduction and Motivation, some applications of dynamic modelling based on experimental data are presented.
10:00 – 10:30 *Coffee Break*
10:30 – 11:00 A basic example and computational implementation of parameter estimation and calculation of the profile likelihood to assess parameter uncertainties and detect non-identifiabilities.
11:00 – 12:30 Hands on session 1:
- Simulation of data for the basic example and a slightly changed setting
- Parameter estimation, Profile likelihood

12:30 – 13:30 *Lunch*
13:30 – 15:00 Presentation of the comprehensive methodology and its introduction of its implementation in the publicly available “Data 2 Dynamics Software”. https://bitbucket.org/d2d-development/d2d-software/wiki/Home
15:00 - 16:00 Hand on session 2: Analysis of real experimental data using the “Data 2 Dynamics Software”
16:00 – 17:00 Presentation of other methodological aspects and recently published approaches. Discussion and Summary.

ECCB14 web site: http://www.eccb14.org/program/tutorials/sndm

T09: TADbit: Automated Analysis and Three-Dimensional Modeling of Genomic Domains

Description

The sequence of a genome alone does not carry enough information to understand how genomic processes are carried out in the cell nucleus. To achieve this, the knowledge of its three-dimensional (3D) architecture is necessary. Advances in genomic technologies and the development of new analytical methods, such as Chromosome Conformation Capture (3C)-based methods, have allowed getting insights at unprecedented resolution into how the genome is organized. Recently, it has been shown that chromatin is organized in Topologically Associating Domains (TADs), large interaction domains that appear to be conserved among different cell types and even between species. In this tutorial, attendants will learn to use the TADbit library for the analysis and 3D modeling of TADs and genomes. TADbit is a python library that uses the Integrative Modeling Platform (IMP, http://www.integrativemodeling.org) to model 3C data for determining the 3D architecture of genomic domains and entire genomes at unprecedented resolutions.

Motivation: The genome is non-randomly organized within the cell nucleus, with chromosomes occupying precise nuclear regions, the so called "chromosome territories", separated by inter-chromatin compartments. It has been shown that chromosomes undergo additional levels of arrangements and organize themselves into Topologically Associating Domains (TADs), regions of the DNA that frequently interact with each other. Knowing how chromatin is arranged within these compartments is necessary for understanding how genes and their regulatory elements get spatially close to carry out their function.

Chromosome Conformation Capture (3C) derived methods have shown to be particularly efficient in revealing the three-dimensional (3D) spatial arrangement of genomic regions; in particular, the Hi-C method allows to identify genome-wide inter-loci interaction patterns. Although inter-loci interaction frequencies can be used as a proxy for their spatial distance, they do not give direct information on their 3D organization. Nonetheless, this information can be inferred with computational methods.

The attendants of this tutorial will learn, through TADbit, how to build up computational pipelines for the analysis and 3D

modeling of Hi-C data. TADbit takes as input Hi-C interaction matrices, which are then used to calculate TAD boundaries. Next, TADbit can build 3D models of selected TADs. This computational library allows building models to directly visualize and analyze looping interactions between distal regulatory elements and can be used for statistical or genomic analysis. TADbit will help in the characterization of global chromatin features and their relation to gene expression and other phenotypic variations genomic functions.
Overall Goals: This tutorial will provide basic knowledge of the 3D modeling of genomic domains. In particular, the attendants will learn:

- Fundamentals of structure prediction
- Interpretation and analysis of 3C-like data
- Topologically Associated Domains detection and analysis
- 3D structure determination and analysis of genomic domains

Organizers

Davide Baù: dbau@pcb.ub.cat, François Serra: fjserra@pcb.ub.cat, Guillaume Filion and Marc A. Marti-Renom: mmarti@pcb.ub.cat, National Center for Genomic Analysis (CNAG), Barcelona, Spain; Gene Regulation, Stem Cells and Cancer Program, Centre de Regulació Genòmica (CRG), Barcelona, Spain, Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain.

Program

09:00 – 10:30 Chromatin structure and Hi-C data: introduction and principles
10:30 - 11:00 Coffee Break
11:00 – 12:30 Introduction to TADbit
12:30 - 13:30 Lunch Break
13:30 - 15:00 3D Modeling of real Hi-C data with TADbit
15:00 - 15:30 Coffee Break
15:30 - 17:00 Analysis of the results

ECCB14 web site: http://www.eccb14.org/program/tutorials/tadbit

Satellite Meeting

S01: European Student Council Symposium 2014

Description

The ISCB Student Council and its European Regional Student Groups are organizing the third European Student Council Symposium to be held on Saturday, September 6th, during one of the Tutorials/Workshop Days of ECCB 2014 in Strasbourg, France.

The European Student Council Symposium is a forum for students and young researchers in the fields of Computational Biology and Bioinformatics. Participants will have the opportunity to present their work to an international audience, build a network within the computational biology community and develop important soft skills in an environment that fosters exchange of ideas and knowledge.

Organizers

Pieter Meysman, Chair ; Margherita Francescatto, Co-Chair.

Program

08:00-09:00	Registration & Ice Breaking
	Oral Presentations Session 1
09:00-09:05	Welcome and opening
09:05-09:50	**Keynote**: *Saprotrophics: a new natural habitat for bioinformaticians?* Prof. Lennart Martens, Ghent University
09:50-10:10	*Viral DNA replication: new insights and discoveries from large scale computational analysis.* Darius Kazlauskas, Institute of Biotechnology, Vilnius University, Lithuania
10:10-10:30	*Using the PDB to explore the conformational space of query proteins with at least one known conformation.* Aya Narunsky, Tel Aviv University, Israel
10:30-11:00	*Coffee Break*
	Oral Presentations Session 2
11:00-11:20	*Applications of Proteochemometrics – From Species Extrapolation to Cell Line Sensitivity Modelling.* Isidro Cortes, Institut Pasteur, Unité de Bioinformatique Structurale; CNRS, France
11:20-11:40	*Genotyping microsatellites in next- An exploration of the 3D chemical space has highlighted a specific shape profile for the compounds intended to inhibit protein-protein interactions.* Mélaine A. Kuenemann, University of Paris Diderot, France
11:40-12:00	*Mining the human proteome for conserved mechanisms.* Stefan Naulaerts, University of Antwerp, Belgium
12:00-14:15	*Lunch & Poster Session*
	Oral Presentations Session 3
14:15-15:00	**Keynote**: *Scale matters! - The importance of scale in the analysis of chromatin landscapes, mutation profiles and protein network architecture*. Dr. Jeroen de Ridder, Delft University of Technology
15:00-15:20	*Tetranucleotide usage in Identification and Analysis of Methylation Call Differences between Bisulfite Microarray and Bisulfite Sequencing Data with Statistical Learning Techniques.* Matthias Döring,Max Planck Institute for Informatics, Germany
15:20-15:40	*Hybrid approaches for the detection of networks of critical residues involved in functional motions in protein familie.* Dagoberto Armenta Medina,IBT/UNAM, Mexico
15:40-16:00	*Improving duplicated nodes position in vertebrate gene trees.* Amelie Peres, Ecole Normale Supérieure, Institut de Biologie de l'ENS, IBENS, France
16:00-16:30	*Coffee Break*
	Oral Presentations Session 4
16:30-17:00	Top Poster Flash Presentations
17:00-17:45	**Keynote**: *Network biology: large-scale data and text mining.* Prof. Lars Juhl Jensen, The Novo Nordisk Foundation Center for Protein Research
17:45	Symposium Final Words & Award Ceremony
19:00	Student Social Event

ECCB14 web site http://www.eccb14.org/program/satellite-meetings/escs

Symposium web site http://escs.iscbsc.org

Oral Presentations

Oral presentations include 40 proceeedings papers (PP01 to PP40) and 10 highlight papers (HP01 to HP10).

ECCB'14 proceedings are available on-line from the *Bioinformatics* (Oxford University Press) web site: http://bioinformatics.oxfordjournals.org/ from August 27, 2014.

You can also download the abstracts of both proceedings and highlight papers (pdf files) from the ECCB'14 web site.

Proceedings abstracts: http://www.eccb14.org/programme/proceedings_papers_abstracts.pdf

Highlight abstracts: http://www.eccb14.org/programme/higlight_papers_abstracts.pdf

Monday, September 8, 2014

Mon1 (Area C): Pathways and Molecular Networks (1)

Chairs: Ralf Zimmer, Anaïs Baudot

PP01 - HubAlign: An accurate and efficient method for global alignment of protein-protein interaction networks

Daniela Boernigen, Somaye Hashemifar and Jinbo Xu

Toyota Technological Institute at Chicago, IL 60637, USA.

ABSTRACT

Motivation: High-throughput experimental techniques have pro-duced a large amount of protein-protein interaction (PPI) data. The study of PPI networks, such as comparative analysis, shall benefit the understanding of life process and diseases at the molecular level. One way of comparative analysis is to align PPI networks to identify conserved or species-specific subnetwork motifs. A few methods have been developed for global PPI network alignment, but it still remains challenging in terms of both accuracy and efficiency.

Results: This paper presents a novel global network alignment algorithm, denoted as HubAlign, that makes use of both network topology and sequence homology information, based upon the observation that topologically important proteins in a PPI network usually are much more conserved and thus, more likely to be aligned. HubAlign uses a minimum-degree heuristic algorithm to estimate the topological and functional importance of a protein from the global network topology information. Then HubAlign aligns topologically important proteins first and gradually extends the alignment to the whole network. Extensive tests indicate that HubAlign greatly out-performs several popular methods in terms of both accuracy and efficiency, especially in detecting functionally similar proteins.

Availability: HubAlign is available freely for non-commercial purposes at
http://ttic.uchicago.edu/~hashemifar/software/HubAlign.zip

Contact: jinboxu@gmail.com

PP02 - Alignment-free protein interaction network comparison

Waqar Ali[1], Tiago Rito[1], Gesine Reinert[1], Fengzhu Sun[2] and Charlotte M. Deane[1]

[1]Department of Statistics, 1 South Parks Road, Oxford OX1 3TG, UK. [2]Molecular and Computational Biology Program, University of Southern California, California, USA.

ABSTRACT

Motivation: Biological network comparison software largely relies on the concept of alignment where close matches between the nodes of two or more networks are sought. These node matches are based on sequence similarity and/or interaction patterns. However due to the incomplete and error prone data sets currently available, such methods have had limited success. Moreover, the results of network alignment are in general not amenable for distance based evolutionary analysis of sets of networks. In this paper we describe Netdis, a topology based distance measure between networks, which offers the possibility of network phylogeny reconstruction.

Results: We first demonstrate that Netdis is able to correctly separate different random graph model types independent of network size and density. The biological applicability of the method is then shown by its ability to build the correct phylogenetic tree of species based solely on the topology of current protein interaction networks. Our results provide new evidence that the topology of protein interaction networks contains information about evolutionary processes, despite the lack of conservation of individual interactions. As Netdis is applicable to all networks due to its speed and simplicity we apply it to a large collection of biological and non-biological networks where it clusters diverse networks by type.

Availability: The source code of the program is freely available at

http://www.stats.ox.ac.uk/research/proteins/resources.
Contact: w.ali@stats.ox.ac.uk

From Area J: Methods and Technologies for Computational Biology

PP03 - Fast randomisation of large genomic datasets while preserving alteration counts

Andrea Gobbi[1*], Francesco Iorio[2, 3*], Kevin J. Dawson[3], David C. Wedge[3], David Tamborero[4], Ludmil B. Alexandrov[3], Nuria Lopez-Bigas[4], Mathew J. Garnett[3], Giuseppe Jurman[1] and Julio Saez-Rodriguez[2].

[1]Fondazione Bruno Kessler, Trento, Italy. [2]European Molecular Biology Laboratory, European Bioinformatics Institute, Cambridge, UK. [3]Wellcome Trust Sanger Institute, Cambridge, UK. [4]Universitat Pompeu Fabra, Barcelona, Spain. *Equally contributing authors.

ABSTRACT

Motivation: Studying combinatorial patterns in cancer genomic datasets has recently emerged as a tool for identifying novel cancer driver networks. Approaches have been devised to quantify, for example, the tendency of a set of genes to be mutated in a 'mutually exclusive' manner. The significance of the proposed metrics is usually evaluated by computing p-values under appropriate null models. To this end, a Monte Carlo method (the switching-algorithm) is used to sample simulated datasets under a null-model that preserves patient- and gene-wise mutation rates. In this method, a genomic dataset is represented as a bipartite network, to which Markov chain updates (switching-steps) are applied. These steps modify the network topology, and a minimal number of them must be executed in order to draw simulated datasets independently under the null model. This number has previously been deducted empirically to be a linear function of the total number of variants, making this process computationally expensive.

Results: We present a novel approximate lower bound for the number of switching-steps, derived analytically. Additionally we have developed the R package BiRewire, including new efficient implementations of the switching-algorithm. We illustrate the performances of BiRewire by applying it to large real cancer genomics datasets. We report vast reductions in time requirement, with respect to existing implementations/bounds and equivalent pvalue computations. Thus, we propose BiRewire to study statistical properties in genomic datasets, and other data that can be modeled as bipartite networks.

Availability: BiRewire is available on BioConductor at
http://www.bioconductor.org/packages/2.13/bioc/html/BiRewire.html

Supplementary information: Available on Bioinformatics online and at http://www.ebi.ac.uk/_iorio/BiRewire

Contact: iorio@ebi.ac.uk

Mon2 (Area C): Pathways and Molecular Networks (2)

Chairs: Ralf Zimmer, Anaïs Baudot

PP04 - Identifying transcription factor complexes and their roles

Thorsten Will and Volkhard Helms

Center for Bioinformatics, Campus Building E2.1, Saarland University, D-66123 Saarbrücken, Germany.

ABSTRACT

Motivation: Eukaryotic gene expression is controlled through molecular logic circuits that combine regulatory signals of many different factors. In particular, complexation of transcription factors and other regulatory proteins is a prevailing and highly conserved mechanism of signal integration within critical regulatory pathways and enables us to infer controlled genes as well as the exerted regulatory mechanism. Common approaches for protein complex prediction that only use protein interaction networks, however, are designed to detect self-contained functional complexes and have difficulties to reveal dynamic combinatorial assemblies of physically interacting proteins.

Results: We developed the novel algorithm DACO that combines protein-protein interaction networks and domain-domain interaction networks with the cluster-quality metric cohesiveness. The metric is locally maximized on the holistic level of protein interactions and connectivity constraints on the domain level are used to account for the exclusive and thus inherently combinatorial nature of the interactions within such assemblies. When applied to predicting transcription factor complexes in the yeast *S.cerevisiae*, the proposed approach outperformed popular complex prediction methods by far. Furthermore, we were able to assign many of the predictions to target genes, as well as to a potential regulatory effect in agreement with literature evidence.

Availability: A prototype implementation is freely available at https://sourceforge.net/projects/dacoalgorithm/.

Contact: volkhard.helms@bioinformatik.uni-saarland.de

PP05 - Personalized identification of altered pathways in cancer

Taejin Ahn[1, 2, 3], Eunjin Lee[1, 2], Nam Huh[1] and Taesung Park[3]

[1]Samsung Advanced Institute of Technology, [2]Samsung Genome Institute, Republic of Korea. [3]Interdisciplinary Program in Bioinformatics, Seoul National University, Republic of Korea.

ABSTRACT

Motivation: Identifying altered pathways in an individual is important for understanding disease mechanisms and for the future application of custom therapeutic decisions. Existing pathway analysis techniques are mainly focused on discovering altered pathways between normal and cancer groups and are not suitable for identifying the pathway aberrance that may occur in an individual sample. A simple way to identify individual's pathway aberrance is to compare normal and tumor data from the same individual. However, the matched normal data from the same individual is often unavailable in clinical situation. We therefore suggest a new approach for the personalized identification of altered pathways, making special use of accumulated normal data in cases when a patient's matched normal data is unavailable. The philosophy behind our method is to quantify the aberrance of an individual sample's pathway by comparing it to accumulated normal samples. We propose and examine personalized extensions of pathway statistics, Over-Representation Analysis (ORA) and Functional Class Scoring (FCS), to generate individualized pathway aberrance score (iPAS).

Results: Collected microarray data of normal tissue of lung and colon mucosa is served as reference to investigate a number of cancer individuals of lung adenocarcinoma and colon cancer, respectively. Our method concurrently captures known facts of cancer survival pathways and identifies the pathway aberrances that represent cancer differentiation status and survival. It also provides more improved validation rate of survival related pathways than when a single cancer sample is interpreted in the context of cancer-only cohort. In addition, our method is useful in classifying unknown samples into cancer or normal groups. Particularly, we identified 'amino acid synthesis and interconversion' pathway is a good indicator of lung adenocarcinoma (AUC 0.982 at independent validation). Clinical importance of the method is providing pathway interpretation of single cancer even though its matched normal data is unavailable.

Availability: The method was implemented using the R software, available at our website: http://bibs.snu.ac.kr/ipas.

Contact: tspark@stat.snu.ac.kr

Supplementary information: Available at *Bioinformatics* online.

Highlight Talk: *HP01 - Wiring miRNAs to pathways: a topological approach to integrate miRNA and mRNA expression profiles.*

Enrica Calura, Paolo Martini, Gabriele Sales and Chiara Romualdi.

Department of Biology, University of Padova, via U. Bassi 58/B, 35121 Padova, Italy.

ABSTRACT

The production rate of gene expression data is nothing less than astounding. However, with the benefit of hindsight we can assert that, since we completely ignored the non-coding part of the transcriptome, we spent the last decade to study cell mechanisms having few data in our hands. In this scenario, microRNAs, which are key post-trascriptional regulators, deserve special attention. Currently, miRNA and gene circuits are identified through the combination of binding prediction and expression correlation analyses, MAGIA, the web tool we developed, is an example to feel this aim (Sales et al NAR 2010, Bisognin et al NAR 2012). Although effective in many cases the simple correlation does not imply a causal relationship and a lot of false positive miRNA-mRNA interactions are still found. Moreover, miRNA and target genes are characterized by many-to-many relationships and they should be considered as part of a much more complex system of cellular interactions. Recently, to analyze the cellular circuits we developed a new web tool dedicated to topological pathway analyses called Graphite Web (Sales et al NAR 1013). Given the state of knowledge about the biogenesis of miRNAs, their mechanisms of action and the numerous experimentally validated target genes, miRNAs are also gradually appearing in the formal pathway representations such as KEGG and Reactome maps. However, the number of miRNAs annotated in pathway maps is very small and pathway analyses exploiting this new regulatory layer are still lacking. To fill these gaps, we developed micrographite a new pipeline to perform topological pathway analysis integrating gene and miRNA expression profiles. Micrographite analysis of gene and miRNA integrated transcriptome is used to study and dissect the epithelial ovarian cancer gene complexity and miRNA transcriptome defining and validating a new regulatory circuits.

Publications:

Calura E, Fruscio R, Paracchini L, Bignotti E, Ravaggi A, Martini P, Sales G, Beltrame L, Clivio L, Ceppi L, Di Marino M, Fuso Nerini I, Zanotti L, Cavalieri D, Cattoretti G, Perego P, Milani R, Katsaros D, Tognon G, Sartori E, Pecorelli S, Mangioni C, D'Incalci M, Romualdi C, Marchini S. MiRNA landscape in stage I epithelial ovarian cancer defines the histotype specificities. Clin Cancer Res. 2013 Aug 1;19(15):4114-23.

Calura E, Martini P, Sales G, Beltrame L, Chiorino G, D'Incalci M, Marchini S, Romualdi C. Wiring miRNAs to pathways: a topological approach to integrate miRNA and mRNA expression profiles. Nucleic Acids Res. 2014;42(11):e96.

Contact: enrica.calura@unipd.it

Mon3 (Area A): Sequencing and Sequence Analysis for Genomics (1)

Chairs: To be announced

PP06 - Lambda: The local aligner for massive biological data

Hannes Hauswedell, Jochen Singer and Knut Reinert

Department of Mathematics and Computer Science, Freie Universität Berlin, Takustr. 9, 14195 Berlin, Germany.

ABSTRACT

Motivation: Next-generation sequencing technologies produce unprecedented amounts of data, leading to completely new research fields. One of these is metagenomics, the study of large-size DNA samples containing a multitude of diverse organisms. A key problem in metagenomics is to functionally and taxonomically classify the sequenced DNA, to which end the well known BLAST program is usually used. But BLAST has dramatic resource requirements at metagenomic scales of data, imposing a high financial or technical burden on the researcher. Multiple attempts have been made to overcome these limitations and present a viable alternative to BLAST.

Results: In this work we present Lambda, our own alternative for BLAST in the context of sequence classification. In our tests Lambda often outperforms the best tools at reproducing BLAST's results and is the fastest compared to the current state-of-the art at comparable levels of sensitivity.

Availability: Lambda was implemented in the SeqAn open source C++ library for sequence analysis and is publicly available for download at http://www.seqan.de/projects/lambda .

Contact: hannes.hauswedell@fu-berlin.de or knut.reinert@fu-berlin.de

PP07 - Fiona: a parallel and automatic strategy for read error correction

Marcel Schulz[1, 2, §], David Weese[3, §], Manuel Holtgrewe[3, §], Viktoria Dimitrova[4, 5], Sijia Niu[4, 5], Knut Reinert[3] and Hugues Richard[4, 5, §]

[1]Cluster of Excellence "Multimodal Computing and Interaction", Saarland University & Max Planck Institute for Informatics, Saarbrücken, Germany. [2]Ray and Stephanie Lane Center for Computational Biology, Carnegie Mellon University, Pittsburgh, USA. [3]Department of Mathematics and Computer Science, Freie Universit ¨ at Berlin, Berlin, Germany. [4]Université Pierre et Marie Curie, UMR7238, CNRS-UPMC, Paris, France. [5]CNRS, UMR7238, Laboratory of Computational and Quantitative Biology, Paris, France. [§]These authors contributed equally to this work.

ABSTRACT

Motivation: Automatic error correction of high throughput sequencing data can have a dramatic impact on the amount of usable base pairs and their quality. It has been shown that the performance of tasks such as de novo genome assembly and SNP calling can be dramatically improved after read error correction. While a large number of methods specialized for correcting substitution errors as found in Illumina data exist, few methods for the correction of indel errors, common to technologies like 454 or Ion Torrent, have been proposed.

Results: We present Fiona, a new stand-alone read error correction method. Fiona provides a new statistical approach for sequencing error detection, optimal error correction and estimates its parameters automatically. Fiona is able to correct substitution, insertion, and deletion errors and can be applied to any sequencing technology. It uses an efficient implementation of the partial suffix array to detect read overlaps with different seed lengths in parallel. We tested Fiona on several real data sets from a variety of organisms with different read lengths and compared its performance to state-of-the-art methods. Fiona shows a constantly higher correction accuracy over a broad range of data sets from 454 and Ion Torrent sequencers, without compromise in speed.

Conclusion: Fiona is an accurate, parameter-free read error correction method that can be run on inexpensive hardware and can make use of multi-core parallelization whenever available. Fiona was implemented using the SeqAn library for sequence analysis and is publicly available for download at http://www.seqan.de/projects/_ona.

Contact: mschulz@mmci.uni-saarland.de , hugues.richard@upmc.fr

Highlight Talk: *HP02 - Comparison of mapping algorithms used in high-throughput sequencing: application to Ion Torrent data*

Ségolène Caboche[1], Christophe Audebert[2], Yves Lemoine[3] and David Hot[2,3]

[1]FRE 3642 Molecular and Cellular Medecine, CNRS, Institut Pasteur de Lille and University of Lille Nord de France. [2]Genes Diffusion, Douai, France. [3]Transcriptomics and Applied Genomics, Center for Infection and Immunity of Lille, Inserm U1019, Lille, France.

ABSTRACT

A fundamental step in High-throughput sequencing (HTS) data analysis is the mapping of reads onto reference sequences. Choosing a suitable mapper is a subtle task because of the difficulty of evaluating mapping algorithms. We present a benchmark procedure to compare mappers using both real and simulated datasets and considering computational resource and time requirements, robustness of mapping, ability to report

positions for reads in repetitive regions, and ability to retrieve true genetic variation positions. To measure robustness, a new definition for a correctly mapped read was introduced. We developed CuReSim, a read simulator, and CuReSimEval, a tool to evaluate the mapping quality of the simulated reads. The benchmark procedure was applied to evaluate mappers in the context of whole genome sequencing of small genomes with Ion Torrent data. These results were used to develop a pipeline to quickly and automatically characterize pathogens during an episode of infection.

Publication:

Caboche S, Audebert C, Lemoine Y, Hot D. Comparison of mapping algorithms used in high-throughput sequencing: application to Ion Torrent data. BMC Genomics. 2014 Apr 5;15:264.

Contact: segolene.caboche@pasteur-lille.fr

Mon4 (Area A): Sequencing and sequence analysis for genomics (2)

Chairs: To be announced

PP08 - FastHap: fast and accurate single individual haplotype reconstruction using fuzzy conflict graphs

Sepideh Mazrouee and Wei Wang.

Computer Science Department, University of California Los Angeles (UCLA), 3551 Boelter Hall, Los Angeles, CA 90095-1596, USA.

ABSTRACT

Motivation: Understanding exact structure of an individual's haplotype plays a significant role in various fields of human genetics. Despite tremendous research effort in recent years, fast and accurate haplotype reconstruction remains as an active research topic, mainly due to the computational challenges involved. Existing haplotype assembly algorithms focus primarily on improving accuracy of the assembly, making them computationally challenging for applications on large high-throughput sequence data. Therefore, there is a need to develop haplotype reconstruction algorithms that are not only
accurate but also highly scalable.

Results: In this paper, we introduce FastHap, a fast and accurate haplotype reconstruction approach, which is up to one order of magnitude faster than the state-of-the-art haplotype inference algorithms while also delivering higher accuracy than these algorithms. FastHap leverages a new similarity metric that allows us to precisely measure distances between pairs of fragments. The distance is then utilized in building the fuzzy conflict graphs of fragments. Given that optimal haplotype reconstruction based on minimum error correction (MEC) is known to be NP-hard, we use our fuzzy conflict graphs to develop a fast heuristic for fragment partitioning and haplotype reconstruction.

Availability: An implementation of FastHap is available for sharing upon request.

Contact: sepideh@cs.ucla.edu

PP09 - Probabilistic single-individual haplotyping

Volodymyr Kuleshov

Department of Computer Science, Stanford University, Stanford, CA, 94305, USA.

ABSTRACT

Motivation: Accurate haplotyping – determining from which parent particular portions of the genome were inherited – is still mostly an unresolved problem in genomics. Only recently have modern long read sequencing technologies begun to offer the promise of routine, cost-effective haplotyping. Here, we introduce ProbHap, a new haplotyping algorithm targeted at such technologies. ProbHap is based on a probabilistic graphical model; it is highly accurate and provides useful confidence scores at phased positions.

Results: On a standard benchmark dataset, ProbHap makes 11% fewer errors than current state-of-the-art methods. This accuracy can be further increased by excluding low-confidence positions, at the cost of a small drop in haplotype completeness.

Availability: Our source code is freely available at https://github.com/kuleshov/ProbHap .

Contact: kuleshov@stanford.edu

From Area J: Methods and Technologies for Computational Biology

PP10 - cnvOffSeq: detecting intergenic copy number variation using off-target exome sequencing data

Evangelos Bellos[1] and Lachlan Coin[1,2]

[1]Department of Genomics of Common Disease, Imperial College London, London W12 0NN, UK. [2]Institute for Molecular Bioscience, University of Queensland, St Lucia, QLD 4072, Australia.

ABSTRACT

Motivation: Exome sequencing technologies have transformed the field of Mendelian genetics and allowed for efficient detection of genomic variants in protein-coding regions. The target enrichment process that is intrinsic to exome sequencing is inherently imperfect, generating large amounts of unintended off-target sequence. Off-target data is characterized by very low and highly heterogeneous coverage and is usually discarded by exome analysis pipelines. We posit that off-target read depth is a rich but overlooked source of information that could be mined to detect intergenic copy number variation (CNV). We propose cnvOffseq, a novel normalization framework for off-target read depth that is based on local adaptive singular value decomposition (SVD). This method is designed to address the heterogeneity of the underlying data and allows for accurate and precise CNV detection and genotyping in off-target regions.
Results: cnvOffSeq was benchmarked on whole-exome sequencing samples from the 1000 Genomes Project. In a set of 104 gold standard intergenic deletions, our method achieved a sensitivity of 57.5% and a specificity of 99.2%, while maintaining a low FDR of 5%. For gold standard deletions longer than 5kb, cnvOffSeq achieves a sensitivity of 90.4% without increasing the FDR. cnvOff-Seq outperforms both whole-genome and whole-exome CNV detection methods considerably and is shown to offer a substantial im-provement over naïve local SVD.
Availability and Implementation: cnvOffSeq is available at http://sourceforge.net/p/cnvoffseq/
Contact: evangelos.bellos09@imperial.ac.uk ; l.coin@imb.uq.edu.au

Mon5 (Area F): Evolution and Population Genomics (1)

Chairs: To be announced

PP11 - Polytomy refinement for the correction of dubious duplications in gene trees

Manuel Lafond[1], Cedric Chauve[2,3], Riccardo Dondi[4] and Nadia El-Mabrouk[1]

[1]Department of Computer Science, Université de Montreal, Montreal (QC), Canada. [2]LaBRI, Université Bordeaux 1, Bordeaux, France. [3]Department of Mathematics, Simon Fraser University, Burnaby (BC), Canada. [4]Universitá degli Studi di Bergamo, Bergamo, Italy.

ABSTRACT
Motivation: Large scale methods for inferring gene trees are errorprone. Correcting gene trees for weakly supported features often results in non-binary trees, i.e., trees with polytomies, thus raising the natural question of refining such polytomies into binary trees. A feature pointing toward potential errors in gene trees are duplications that are not supported by the presence of multiple gene copies.
Results: We introduce the problem of refining polytomies in a gene tree while minimizing the number of created non-apparent duplications in the resulting tree. We show that this problem can be described as a graph-theoretical optimization problem. We provide a bounded heuristic with guaranteed optimality for well characterized instances. We apply our algorithm to a set of ray-finned fish gene trees from the Ensembl database to illustrate its ability to correct dubious duplications.
Availability: The C++ source code for the algorithms and simulations described in the paper are available at http://wwwetud.iro.umontreal.ca/lafonman/software.php .
Contact: lafonman@iro.umontreal.ca , mabrouk@iro.umontreal.ca

PP12 - RidgeRace: Ridge regression for continuous ancestral character estimation on phylogenetic trees.

Christina Kratsch and Alice McHardy.

Department for Algorithmic Bioinformatics, Heinrich Heine University, Universitätsstr. 1, 40225 Düsseldorf, Germany.

ABSTRACT
Motivation: Ancestral character state reconstruction describes a set of techniques for estimating phenotypic or genetic features of species or related individuals that are the predecessors of those present today. Such reconstructions can reach into the distant past and can provide insights into the history of a population or a set of species when fossil data are not available, or they can be used to test evolutionary hypotheses e.g. on the co-evolution of traits. Typical methods for ancestral character state reconstruction of continuous characters consider the phylogeny of the underlying data and estimate the ancestral process along the branches of the tree. They usually assume a Brownian motion model of character evolution or extensions thereof, requiring specific assumptions on the rate of phenotypic evolution.
Results: We suggest using ridge regression to infer rates for each branch of the tree and the ancestral values at each inner node. We performed extensive simulations to evaluate the performance of this method and have shown that the accuracy of its reconstructed ancestral values is competitive to reconstructions using other state-of- the-art software. Using a hierarchical clustering of gene mutation profiles from an ovarian cancer dataset, we demonstrate the use of the method as a feature selection tool.
Availability: The algorithm described here is implemented in C++ as a standalone program, and the source code is freely available at http://algbio.cs.uni-duesseldorf.de/software/RidgeRace.tar.gz .
Contact: mchardy@hhu.de

PP13 - Point estimates in phylogenetic reconstructions

Philipp Benner[1], Miroslav Bacak[1] and Pierre-Yves Bourguignon[1,2]

[1]Max-Planck Institute for Mathematics in the Sciences, Inselstr. 22, 04103 Leipzig, Germany. [2]Isthmus SARL, 81 rue Réaumur, 75002 Paris, France.

ABSTRACT

Motivation: The construction of statistics for summarizing posterior samples returned by a Bayesian phylogenetic study has so far been hindered by the poor geometric insights available into the space of phylogenetic trees, and adhoc methods such as the derivation of a consensus tree make up for the ill-definition of the usual concepts of posterior mean, while bootstrap methods mitigate the absence of a sound concept of variance. Yielding satisfactory results with sufficiently concentrated posterior distributions, such methods fall short of providing a faithful summary of posterior distributions if the data does not offer compelling evidence for a single topology.

Results: Building upon previous work of Billera et al. (2001), summary statistics such as sample mean, median, and variance are defined as the geometric median, Fréchet mean and variance respectively. Their computation is enabled by recently published works (Băcák, 2013; Miller et al., 2012), and embeds an algorithm for computing

shortest paths in the space of trees (Owen and Provan, 2011). Studying the phylogeny of a set of plants, where several tree topologies occur in the posterior sample, the posterior mean balances correctly the contributions from the different topologies, where a consensus tree would be biased. Comparisons of the posterior mean, median, and consensus trees with the ground truth using simulated data also reveals the benefits of a sound averaging method when reconstructing phylogenetic trees.

Availability: We provide two independent implementations of the algorithm for computing Fréchet means, geometric medians, and variances in the space of phylogenetic trees.

TFBayes: https://github.com/pbenner/tfbayes , TrAP: https://github.com/bacak/TrAP .

Contact: philipp.benner@mis.mpg.de

Mon6 (Area F): Evolution and Population Genomics (2)

Chairs: To be announced

PP14 - ASTRAL: Genome-scale coalescent-based species tree estimation

Siavash Mirarab[1], Rezwana Reaz Rimpi[1], Md. Shamsuzzoha Bayzid[1], Théo Zimmermann[1], Shel Swenson[2] and Tandy Warnow[1]

[1]Department of Computer Science, The University of Texas at Austin, Austin TX, USA. [2]Department of Electrical Engineering, The University of Southern California, Los Angeles CA, USA.

ABSTRACT

Motivation: Species trees provide insight into basic biology, including the mechanisms of evolution and how it modifies biomolecular function and structure, biodiversity, and co-evolution between genes and species. Yet gene trees often differ from species trees, creating challenges to species tree estimation. One of the most frequent causes for conflicting topologies between gene trees and species trees is incomplete lineage sorting (ILS), which is modelled by the multi-species coalescent. While many methods have been developed to estimate species trees from multiple genes, some which have statistical guarantees under the multi-species coalescent model, existing methods are too computationally intensive for use with genome-scale analyses or have been shown to have poor accuracy under some realistic conditions.

Results: We present ASTRAL, a fast method for estimating species trees from multiple genes. ASTRAL is statistically consistent, can run on datasets with thousands of genes, and has outstanding accuracy – improving upon MP-EST and the population tree from BUCKy, two statistically consistent leading coalescent-based methods. ASTRAL is often more accurate than concatenation using maximum likelihood, except when ILS levels are low or there are too few gene trees.

Availability: ASTRAL is available in open source form at https://github.com/smirarab/ASTRAL /. Datasets studied in this paper are available at http://www.cs.utexas.edu/users/phylo/datasets/astral .

Contact: warnow@illinois.edu

Highlight Talk: *HP03 - Patterns of positive selection in seven ant genomes.*

Julien Roux[1], Eyal Privman[2], Sébastien Moretti[1], Josephine Daub[3], Marc Robinson-Rechavi[1] and Laurent Keller[1]

[1]University of Lausanne, Switzerland. [2]University of Haifa, Israel.[3]University of Bern, Switzerland.

ABSTRACT

The evolution of ants is marked by remarkable adaptations that allowed the development of very complex

social systems. To identify how ant-specific adaptations are associated with patterns of molecular evolution, we searched for signs of positive selection on amino-acid changes in proteins. We identified 24 functional categories of genes which were enriched for positively selected genes in the ant lineage. We also reanalyzed genome-wide datasets in bees and flies with the same methodology, to check whether positive selection was specific to ants or also present in other insects. Notably, genes implicated in immunity were enriched for positively selected genes in the three lineages, ruling out the hypothesis that the evolution of hygienic behaviors in social insects caused a major relaxation of selective pressure on immune genes. Our scan also indicated that genes implicated in neurogenesis and olfaction started to undergo increased positive selection before the evolution of sociality in Hymenoptera. Finally, the comparison between these three lineages allowed us to pinpoint molecular evolution patterns that were specific to the ant lineage. In particular, there was ant-specific recurrent positive selection on genes with mitochondrial functions, suggesting that mitochondrial activity was improved during the evolution of this lineage. This might have been an important step toward the evolution of extreme lifespan that is a hallmark of ants.

Publication:

Roux J, Privman E, Moretti S, Daub JT, Robinson-Rechavi M, Keller L. Patterns of positive selection in seven ant genomes. Mol Biol Evol. 2014 Jul;31(7):1661-85.

Contact: julien.roux@unil.ch

Mon7 (Area E): Structural Bioinformatics (1)

Chairs: Torsten Schwede, Anna Tramontano

PP15 - Assessing the local structural quality of transmembrane protein models using statistical potentials (QMEANBrane)

Gabriel Studer[1,2], Marco Biasini[1,2] and Torsten Schwede[1,2]

[1]Biozentrum, University of Basel, Basel, 4056, Switzerland. [2]SIB Swiss Institute of Bioinformatics, Basel, 4056, Switzerland.

ABSTRACT

Motivation: Membrane proteins are an important class of biological macromolecules involved in many cellular key processes including signalling and transport. They account for one third of genes in the human genome and more than 50% of current drug targets. Despite their importance, experimental structural data is sparse, resulting in high expectations for computational modelling tools to help filling this gap. However, as many empirical methods have been trained on experimental structural data, which is biased towards soluble globular proteins, their accuracy for transmembrane proteins is often limited.

Results: We developed a local model quality estimation method for membrane proteins ("QMEANBrane") by combining statistical potentials trained on membrane protein structures with a per-residue weighting scheme. The increasing number of available experimental membrane protein structures allowed us to train membrane-specific statistical potentials that approach statistical saturation. We show that reliable local quality estimation of membrane protein models is possible, thereby extending local quality estimation to these biologically relevant molecules.

Availability: Source code and data sets are available on request.

Contact: torsten.schwede@unibas.ch

PP16 - A new statistical framework to assess structural alignment quality using information compression

James Collier[1], Lloyd Allison[1], Arthur Lesk[2], Maria Garcia de La Banda[1] and Arun Konagurthu[1]

[1] Clayton School of Information Technology, Monash University, Clayton, VIC 3800 Australia. [2] Department of Biochemistry and Molecular Biology, Pennsylvania State University, University Park, PA 16802 USA.

ABSTRACT

Motivation: Progress in protein biology depends on the reliability of results from a handful of computational techniques, structural alignments being one. Recent reviews have highlighted substantial inconsistencies and differences between alignment results generated by the ever-growing stock of structural alignment programs. The lack of consensus on how the quality of structural alignments must be assessed has been identified as the main cause for the observed differences. Current methods assess structural alignment quality by constructing a scoring function that attempts to balance conflicting criteria, mainly alignment coverage and fidelity of structures under superposition. This traditional approach to measuring alignment quality, the subject of considerable literature, has failed to solve the problem. Further development along the same lines is unlikely to rectify the current deficiencies in the field.

Results: This paper proposes a new statistical framework to assess structural alignment quality and significance based on lossless information compression. This is a radical departure from the traditional approach of formulating scoring functions. It links the structural alignment problem to the general class of statistical inductive inference problems, solved using the information-theoretic criterion of minimum message length. Based on this, we developed an efficient and reliable measure of structural alignment quality, I-value.

The performance of I-value is demonstrated in comparison with a number of popular scoring functions, on a large collection of competing alignments. Our analysis shows that I-value provides a rigorous and reliable quantification of structural alignment quality, addressing a major gap in the field.
Availability: http://lcb.infotech.monash.edu.au/I-value
Supplementary Information: http://lcb.infotech.monash.edu.au/I-value/suppl.html
Contact: arun.konagurthu@monash.edu

From Area J: Methods and Technologies for Computational Biology

PP17 - Entropy driven partitioning of the hierarchical protein space

Nadav Rappoport[1], Amos Stern[1], Nathan Linial[1] and Michal Linial[2]

[1]School of Computer Science and Engineering, The Hebrew University of Jerusalem, Israel. [2]Department of Biological Chemistry, Institute of Life Sciences, The Hebrew University of Jerusalem, Israel.

ABSTRACT
Motivation: Modern protein sequencing techniques have led to the determination of over 50 million protein sequences. *ProtoNet* is a clustering system that provides a continuous hierarchical agglomerative clustering tree for all proteins. While ProtoNet performs unsupervised classification of all included proteins, finding an optimal level of granularity for the purpose of focusing on protein functional groups remain elusive. Here, we ask whether knowledge-based annotations on protein families can support the automatic, unsupervised methods for identifying high quality protein families. We present a method that yields within the ProtoNet hierarchy an optimal partition of clusters, relative to manual annotation schemes. The methods principle is to minimize the entropy-derived distance between annotation-based partitions and all available hierarchical partitions. We describe the *best front* (BF) partition of 2,478,328 proteins from UniRef50. Out of 4,929,553 ProtoNet tree clusters, BF based on Pfam annotations contain 26,891 clusters. The high quality of the partition is validated by the close correspondence with the set of clusters that best describe thousands of keywords of Pfam. The BF is shown to be superior to naïve cut in the ProtoNet tree that yields a similar number of clusters. Finally, we used parameters intrinsic to the clustering process to enrich a-priori the BF's clusters. We present the entropy-based method's benefit in overcoming the unavoidable limitations of nested clusters in ProtoNet. We suggest that this automatic information-based cluster selection can be useful for other large-scale annotation schemes, as well as for systematically testing and comparing putative families derived from alternative clustering methods.
Availability: A catalogue of BF clusters for thousands of Pfam keywords is provided at: http://protonet.cs.huji.ac.il/bestFront/
Contact: michal.linial@huji.ac.il

Mon8 (Area E): Structural Bioinformatics (2)

Chairs: Torsten Schwede, Anna Tramontano

PP18 - PconsFold: Improved contact predictions improve protein models

Mirco Michel[1,2], Sikander Hayat[3], Marcin J. Skwark[4], Chris Sander[5], Debora S. Marks[3] and Arne Elofsson[1,2]

[1]Department of Biochemistry and Biophysics, Stockholm University, 10691 Stockholm, Sweden, [2]Science for Life Laboratory, Box 1031, 17121 Solna, Sweden, [3]Department of Systems Biology, Harvard Medical School, Boston, Massachusetts, USA, [4]Department of Information and Computer Science, Aalto University, PO Box 15400, FI-00076 Aalto, Finland, and [5]Computational Biology Center, Memorial Sloan-Kettering Cancer Center, New York, New York, USA.

ABSTRACT
Motivation: Recently is has been shown that the quality of protein contact prediction from evolutionary information can be improved significantly if direct and indirect information is separated. Given sufficiently large protein families the contact predictions contain sufficient information to predict the structure of many protein families. However, since the first studies contact prediction methods have improved. Here, we ask how much the final models are improved if improved contact predictions are used.
Results: In a small benchmark of 15 proteins we show that the TM-scores of top ranked models are improved by on average 33% using PconsFold compared to the original version of EVfold. In a larger benchmark we find that the quality is improved with 15-30% when using PconsC in comparison to earlier contact prediction methods.Further, using Rosetta instead of CNS does not significantly improve global model accuracy but the chemistry of models generated with Rosetta is improved.
Availability: PconsFold is a fully automated pipeline for ab-initio protein structure prediction based on evolutionary information. PconsFold is based on PconsC contact prediction and uses the Rosetta folding protocol. Due to its modularity, the contact prediction tool can be easily exchanged. The source code of PconsFold is available on GitHub at https://www.github.com/ElofssonLab/pcons-fold under the MIT license.

PconsC is available from http://c.pcons.net/ .
Contact: arne@bioinfo.se
Supplementary information: Supplementary data are available at Bioinformatics online.

From Area J: Methods and Technologies for Computational Biology

PP19 - Microarray R-based analysis of complex lysate experiments with MIRACLE

Markus List[1,2,3,§], Ines Block[1,2,§], Marlene Lemvig Pedersen[1,2], Helle Christiansen[1,2], Steffen Schmidt[1,2], Mads Thomassen[1,3], Qihua Tan[3,4], Jan Baumbach[5] and Jan Mollenhauer[1,2]

[1]Lundbeckfonden Center of Excellence in Nanomedicine NanoCAN, University of Southern Denmark, Odense, Denmark. [2]Molecular Oncology, Institute of Molecular Medicine, University of Southern Denmark, Odense, Denmark. [3]Institute of Clinical Research, University of Southern Denmark, Odense, Denmark. [4]Epidemiology, Biostatistics and Biodemography, Institute of Public Health, University of Southern Denmark, Odense, Denmark. [5]Department of Mathematics and Computer Science, University of Southern Denmark, Odense, Denmark. [§] joint first authorship.

ABSTRACT
Motivation: Reverse phase protein arrays (RPPAs) allow sensitive quantification of relative protein abundance in thousands of samples in parallel. Typical challenges involved in this technology are antibody selection, sample preparation and optimization of staining conditions. The issue of combining effective sample management and data analysis, however, has been widely neglected.
Results: This motivated us to develop MIRACLE, a comprehensive and user-friendly web application bridging the gap between spotting and array analysis by conveniently keeping track of sample information. Data processing includes correction of staining bias, estimation of protein concentration from response curves, normalization for total protein amount per sample and statistical evaluation. Established analysis methods have been integrated with MIRACLE, offering experimental scientists an end-to-end solution for sample management and for carrying out data analysis. In addition, experienced users have the possibility to export data to R for more complex analyses. MIRACLE thus has the potential to further spread utilization of RPPAs as an emerging technology for high-throughput protein analysis.
Availability: Project URL: http://www.nanocan.org/miracle/
Contact: mlist@health.sdu.dk

Highlight Talk: *HP04 - Comprehensive analysis of DNA polymerase III alpha subunits and their homologs in bacterial genomes*

Kęstutis Timinskas[1], Monika Balvočiūtė[2], Albertas Timinskas[1] and Česlovas Venclovas[1]

[1]Institute of Biotechnology, Vilnius University, Lithuania. [2]University of Otago, New Zealand.

ABSTRACT
Bacteria, unlike archaea and eukaryotes, use distinct C-family DNA polymerases for genome replication. Unfortunately, except for a few species, bacterial genome replication is poorly characterized. It is not known whether all bacteria use C-family DNA polymerases for DNA replication, how many distinct C-family groups are there, and how many different replication systems they form. In order to address these questions, we performed extensive computational analysis of C-family polymerases in nearly 2000 complete bacterial genomes. We found that all the genomes without exception encode at least one C-family polymerase implying the universal use of this polymerase family for bacterial DNA replication. Our analysis revealed four distinct groups of C-family polymerases. Based on their properties and distribution in genomes we discovered a novel, so far experimentally uncharacterized, replication system in Clostridia. Computational results also indicated that one of the C-family groups might be responsible for shaping genomic G+C content.
Publication:
Timinskas K, Balvočiūtė M, Timinskas A, Venclovas Č. Comprehensive analysis of DNA polymerase III α subunits and their homologs in bacterial genomes. Nucleic Acids Res. 2014 Feb;42(3):1393-413.
Contact: venclovas@ibt.lt

Tuesday, September 9, 2014

Tue1 (Area D) : Computational Systems Biology (1)

Chairs : Oliver Kohlbacher, Anne Siegel

PP20 - Stronger findings for metabolomics through Bayesian modeling of multiple peaks and compound correlations

Tommi Suvitaival[1], Simon Rogers[2] and Samuel Kaski[1,3]

[1]Helsinki Institute for Information Technology HIIT, Department of Information and Computer Science, Aalto University, FI-00076, Espoo, Finland. [2]School of Computing Science, University of Glasgow, Glasgow, G12 8QQ, UK. [3]Helsinki Institute for Information Technology HIIT, Department of Computer Science, University of Helsinki, Helsinki, Finland.

ABSTRACT

Motivation: Data analysis for metabolomics suffers from uncertainty due to the noisy measurement technology and the small sample-size of experiments. Noise and the small sample-size lead to a high probability of false findings. Further, individual compounds have natural variation between samples, which in many cases renders them unreliable as biomarkers. However, the levels of similar compounds are typically highly correlated, which is a phenomenon that we model in this work.

Results: We propose a hierarchical Bayesian model for inferring differences between groups of samples more accurately in metabolomic studies, where the observed compounds are collinear. We discover that the method decreases the error of weak and non-existent covariate effects, and thereby reduces false positive findings. To achieve this, the method makes use of the mass spectral peak data by clustering similar peaks into latent compounds, and by further clustering latent compounds into groups that respond in a coherent way to the experimental covariates. We demonstrate the method with three simulated studies and validate it with a metabolomic benchmark data set.

Availability and Implementation: An implementation in R is available at http://research.ics.aalto.fi/mi/software/peakANOVA/ .

Contact: tommi.suvitaival@aalto.fi , simon.rogers@glasgow.ac.uk , samuel.kaski@aalto.fi .

PP21 - Causal network inference using biochemical kinetics

Chris Oates[1], Frank Dondelinger[2], Nora Bayani[3], James Korkola[4], Joe Gray[4] and Sach Mukherjee[2,5]

[1]Department of Statistics, University of Warwick, Coventry, UK. [2]MRC Biostatistics Unit, Cambridge, UK. [3]Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, USA. [4]Knight Cancer Institute, Oregon Health and Science University, Portland, USA. [5]School of Clinical Medicine, University of Cambridge, Cambridge, UK.

ABSTRACT

Motivation: Networks are widely used as structural summaries of biochemical systems. Statistical estimation of networks is usually based on linear or discrete models. However, the dynamics of biochemical systems are generally nonlinear, suggesting that suitable nonlinear formulations may offer gains with respect to causal network inference and aid in associated prediction problems.

Results: We present a general framework for network inference and dynamical prediction using time-course data that is rooted in nonlinear biochemical kinetics. This is achieved by considering a dynamical system based on a chemical reaction graph with associated kinetic parameters. Both the graph and kinetic parameters are treated as unknown; inference is carried out within a Bayesian framework. This allows prediction of dynamical behavior even when the underlying reaction graph itself is unknown or uncertain. Results, based on (i) data simulated from a mechanistic model of mitogen activated protein kinase signaling and (ii) phosphoproteomic data
from cancer cell lines, demonstrate that nonlinear formulations can yield gains in causal network inference and permit dynamical prediction and uncertainty quantification in the challenging setting where the reaction graph is unknown.

Availability: MATLAB R2014a software is available to download from http://warwick.ac.uk/chrisoates .

Contact: c.oates@warwick.ac.uk ; sach@mrc-bsu.cam.ac.uk

Highlight Talk: *HP05 - High-dimensional Bayesian parameter estimation: Case study for a model of JAK2/STAT5 signaling*

Sabine Hug[1], Andreas Raue[2], Jan Hasenauer[1], Julie Bachmann[3], Ursula Klingmüller[3], Jens Timmer[4] and Fabian J. Theis[1]

[1]Institute of Computational Biology, Helmholtz Zentrum München, Germany. [2]Merrimack Pharmaceuticals, Inc., Cambridge, MA, USA. [3]Systems Biology of Signal Transduction, DKFZ-ZMBH Alliance, German Cancer Research Center, Heidelberg, Germany. [4]Institute for Physics, University of Freiburg, Germany.

ABSTRACT

Mechanistic dynamical models are nowadays commonly used for the analysis for complex datasets. Dynamical models depend however on many unknown parameters which have to be inferred from experimental data. The statistical inference in a high-dimensional parameter space is however conceptually and computationally challenging. In this paper we provide a proof of principle that the rigorous statistical analysis is also feasible in these demanding situations.

To ensure rigorous assessment of model and prediction uncertainties we take advantage of both a profile posterior approach and Markov chain Monte Carlo sampling.

We analyzed a dynamical model of the JAK2/STAT5 signaling pathway containing more than hundred parameters. The profile posterior reveals that the corresponding posterior distribution is bimodal. To

nevertheless guarantee efficient mixing we applied a multi-chain sampling approach. The Bayesian parameter estimation enables the assessment of prediction uncertainties and the design of additional experiments enhancing the explanatory power of the model.

Publication:
Hug S, Raue A, Hasenauer J, Bachmann J, Klingmüller U, Timmer J, Theis FJ. High-dimensional Bayesian parameter estimation: case study for a model of JAK2/STAT5 signaling. Math Biosci. 2013 Dec;246(2):293-304.

Contact: sabine.hug@helmholtz-muenchen.de

Tue2 (Area D): Computational Systems Biology (2)

Chairs: Oliver Kohlbacher, Anne Siegel

PP22 - Effects of small particle numbers on long-term behaviour in discrete biochemical systems

Peter Kreyssig[1], Christian Wozar[1], Stephan Peter[1], Tomas Veloz[2,3,4], Bashar Ibrahim[1,5,6] and Peter Dittrich[1]

[1]Bio Systems Analysis Group, Department of Mathematics and Computer Science and Jena Centre for Bioinformatics, Friedrich Schiller University Jena, 07743 Jena, Germany. [2]Mathematics Department, University of British Columbia, Kelowna, BC V1V 1V7, Canada. [3]Instituto de Filosofía y Ciencias de la Complejidad - IFICC, Los Alerces 3024 Ñuñoa, Santiago, Chile. [4]Center Leo Apostel, Vrije Universiteit Brussel, Krijgskundestraat 33, B-1160 Brussels, Belgium. [5]Umm Al-Qura University, 1109 Makkah Al-Mukarramah, Kingdom of Saudi Arabia. [6]Al-Qunfudah Center for Scientific Research (QCSR), 21912 Al-Qunfudah, Kingdom of Saudi Arabia.

ABSTRACT

Motivation: The functioning of many biological processes depends on the appearance of only a small number of a single molecular species. Additionally, the observation of molecular crowding leads to the insight that even a high number of copies of species does not guarantee their interaction. How single particles contribute to stabilising biological systems is not well understood yet. Hence we aim at determining the influence of single molecules on the long-term behaviour of biological systems, *i.e.* whether they can reach a steady state or not.

Results: We provide theoretical considerations and a tool to analyse SBML models for the possibility to stabilise due to the described effects. The theory is an extension of chemical organisation theory which we called discrete chemical organisation theory. Furthermore we scanned the BioModels Database for the occurrence of discrete chemical organisations. To exemplify our method we describe an application to the Template model of the mitotic spindle assembly checkpoint mechanism.

Availability: http://www.biosys.uni-jena.de/Services.html

Contact: bashar.ibrahim@uni-jena.de , dittrich@minet.uni-jena.de

Supplementary Information: Supplementary data are available at Bioinformatics online.

PP23 - TEMPI: Probabilistic modeling time-evolving differential PPI networks with multiple information

Yongsoo Kim[1], Jin-Hyeok Jang[1], Seungjin Choi[2] and Daehee Hwang[1,3]

[1]School of Interdisciplinary Bioscience and Bioengineering and [2]Department of Computer Science and Engineering, Pohang University of Science and Technology, Pohang 790-784, Korea, [3]Center for Systems Biology of Plant Senescence and Life History, Institute for Basic Science, Daegu Gyeongbuk Institute of Science and Technology, Daegu 711-873, Korea.

ABSTRACT

Motivation: Time-evolving differential protein-protein interaction (PPI) networks are essential to understand

serial activation of differentially regulated (up- or down-regulated) cellular processes (DRPs) and their interplays over time. Despite developments in the network inference, current methods are still limited in identifying temporal transition of structures of PPI networks, DRPs associated with the structural transition, and the interplays among the DRPs over time.
Results: Here, we present a probabilistic model for estimating Time-Evolving differential PPI networks with MultiPle Information (TEMPI). This model describes probabilistic relationships among network structures, time-course gene expression data, and Gene Ontology biological processes (GOBPs). By maximizing the likelihood of the probabilistic model, TEMPI estimates jointly the time-evolving differential PPI networks (TDNs) describing temporal transition of PPI network structures together with serial activation of DRPs associated with transiting networks. This joint estimation enables us to interpret the TDNs in terms of temporal transition of the DRPs. To demonstrate the utility of TEMPI, we applied it to two time-course datasets. TEMPI identified the TDNs that correctly delineated temporal transition of DRPs and time-dependent associations between the DRPs. These TDNs provide hypotheses for mechanisms underlying serial activation of key DRPs and their temporal associations.
Availability: Source code and sample data files are available at http://sbm.postech.ac.kr/tempi/sources.zip .
Contact: seungjin@postech.ac.kr or dhwnag@dgist.ac.kr

PP24 - Experimental design schemes for learning Boolean network models

Nir Atias, Michal Gershenzon, Katia Labazin and Roded Sharan

Blavatnik School of Computer Science, Tel Aviv University, Tel Aviv 69978, Israel.

ABSTRACT
Motivation: A holy grail of biological research is a working model of the cell. Current modeling frameworks, especially in the protein-protein interaction domain, are mostly topological in nature, calling for stronger and more expressive network models. One promising alternative is logic-based, or Boolean network modeling, which was successfully applied to model signaling regulatory circuits in human. Learning such models requires observing the system under a sufficient number of different conditions. To date, the amount of measured data is the main
bottleneck in learning informative Boolean models, underscoring the need for efficient experimental design strategies.
Results: We developed novel design approaches that greedily select an experiment to be performed so as to maximize the difference or the entropy in the results it induces with respect to current best-fit models. Unique to our maximum difference approach is the ability to account for all (possibly exponential number of) Boolean models displaying high fit to the available data. We applied both approaches to simulated and real data from the EFGR and IL1 signaling systems in human. We demonstrate the utility of the developed strategies in substantially improving on a random selection approach. Our design schemes highlight the redundancy in these data sets, leading up to 11-fold savings in the number of experiments to be performed.
Availability: Source code will be made available upon acceptance of the manuscript.
Contact: roded@post.tau.ac.il

Tue3 (Area G): Bioinformatics of Health and Disease (1)

Chairs: To be announced

PP25 - OncodriveROLE classifies cancer driver genes in loss of function and activating mode of action

Michael P Schroeder[1], Carlota Rubio-Perez[1], David Tamborero[1], Abel Gonzalez-Perez[1,*] and Nuria Lopez-Bigas [1,2]

[1] Research Unit on Biomedical Informatics, Department of Experimental and Health Sciences, Universitat Pompeu Fabra, Dr. Aiguader 88, Barcelona, Spain. [2] Institució Catalana de Recerca i Estudis Avançats (ICREA), Passeig Lluís Companys, 23, Barcelona, Spain.

ABSTRACT
Motivation: Several computational methods have been developed to identify cancer drivers genes – genes responsible for cancer development upon specific alterations. These alterations can cause the loss of function of the gene product, for instance in tumor suppressors, or increase or change its activity or function, if it is an oncogene. Distinguishing between these two classes is important to understand tumorigenesis in patients and has implications for therapy decision making. Here, we assess the capacity of multiple gene features related to the pattern of genomic alterations across tumors to distinguish between activating and loss of function cancer genes and we present an automated approach to aid the classification of novel cancer drivers according to their role.
Result: OncodriveROLE is a machine learning-based approach that classifies driver genes according to their role, using several properties related to the pattern of alterations across tumors. The method shows an accuracy of 0.93 and Matthew's Correlation Coefficient of 0.84 classifying genes in the Cancer Gene Census. The OncodriveROLE classifier, its results when applied to two list of predicted cancer drivers and TCGA-

derived mutation and copy number features used by the classifier are available at http://bg.upf.edu/oncodrive-role.
Contact: abel.gonzalez@upf.edu , nuria.lopez@upf.edu

PP26 - ContrastRank: a new method for ranking putative cancer driver genes and classification of tumor samples

Rui Tian[1], Malay Basu[1,2] and Emidio Capriotti[1,2,3]

[1]Division of Informatics, Department of Pathology, University of Alabama at Birmingham, 619 19th St. South, 35249 Birmingham, AL, USA. [2]Department of Clinical and Diagnostic Sciences, University of Alabama at Birmingham, 1705 University Boulevard, 35249 Birmingham, AL, USA. [3]Department of Biomedical Engineering, University of Alabama at Birmingham, 1075 13th Street South, 35249 Birmingham, AL, USA.

ABSTRACT
Motivation: The recent advance in high-throughput sequencing technologies is generating a huge amount of data that are becoming an important resource for deciphering the genotype underlying a given phenotype. Genome sequencing has been extensively applied to the study of the cancer genomes. Although a few methods have been already proposed for the detection of cancer-related genes, their automatic identification is still a challenging task. Using the genomic data made available by The Cancer Genome Atlas Consortium (TCGA), we propose a new prioritization approach based on the analysis of the distribution of putative deleterious variants in a large cohort of cancer samples.
Results: In this paper, we present ContastRank, a new method for the prioritization of putative impaired genes in cancer. The method is based on the comparison of the putative defective rate of each gene in tumor versus normal and 1000 genome samples. We show that the method is able to provide a ranked list of putative impaired genes for colon, lung and prostate adenocarcinomas. The list significantly overlaps with the list of known cancer driver genes previously published. More importantly, by using our scoring approach, we can successfully discriminate between TCGA normal and tumor samples. A binary classifier based on ContrastRank score reaches an overall accuracy higher than 90% and the Area Under the Curve (AUC) of Receiver Operating Characteristics (ROC) higher than 0.95 for all the three types of adenocarcinoma analysed in this paper. In addition, using ContrastRank score we are able to discriminate the three tumor types with a minimum overall accuracy of 77% and AUC of 0.83.
Conclusions: We describe ContrastRank, a method for prioritizing putative impaired genes in cancer. The method is based on the comparison of exome sequencing data from different cohorts and can detect putative cancer driver genes. ContrastRank can also be used to estimate a global score for an individual genome about the risk of adenocarcinoma based on the genetic variants information from a whole-exome VCF (Variant Calling Format) file. We believe that the application of ContrastRank can be an important step in genomic medicine to enable genome-based diagnosis.
Availability: The lists of ContrastRank scores of all genes in each tumor type are available as supplementary materials. A webserver for evaluating the risk of the three studied adenocarcinomas starting from whole-exome VCF file is under development.
Contact: emidio@uab.edu

PP27 - Drug susceptibility prediction against a panel of drugs using kernelized Bayesian multitask learning

Mehmet Gönen and Adam A. Margolin

Sage Bionetworks, 1100 Fairview Avenue North, Seattle, WA 98109, USA. Present address: Department of Biomedical Engineering, Oregon Health & Science University, 3303 SW Bond Avenue, Portland, OR 97239, USA.

ABSTRACT
Motivation: Human immunodeficiency virus (HIV) and cancer require personalized therapies due to their inherent heterogeneous nature. For both diseases, large-scale pharmacogenomic screens of molecularly characterized samples have been generated with the hope of identifying genetic predictors of drug susceptibility. Thus, computational algorithms capable of inferring robust predictors of drug responses from genomic information are of great practical importance. Most of the existing computational studies that consider drug susceptibility prediction against a panel of drugs formulate a separate learning problem for each drug, which cannot make use of commonalities between subsets of drugs.
Results: In this study, we propose to solve the problem of drug susceptibility prediction against a panel of drugs in a multi-task learning framework by formulating a novel Bayesian algorithm that combines kernel-based nonlinear dimensionality reduction and binary classification (or regression). The main novelty of our method is the joint Bayesian formulation of projecting data points into a shared subspace and learning predictive models for all drugs in this subspace, which helps us to eliminate off-target effects and drug-specific experimental noise. Another novelty of our method is the ability of handling missing phenotype values due to experimental conditions and quality control reasons. We demonstrate the performance of our algorithm via cross-validation experiments on two benchmark drug susceptibility datasets of HIV and cancer. Our method obtains statistically significantly better predictive performance on most of the drugs compared to baseline single-task algorithms

that learn drug-specific models. These results show that predicting drug susceptibility against a panel of drugs simultaneously within a multi-task learning framework improves overall predictive performance over single-task learning approaches.
Availability: Our Matlab implementations for binary classification and regression are available at https://github.com/mehmetgonen/kbmtl .
Contact: mehmet.gonen@sagebase.org

Tue4 (Area H): Biological Knowledge Discovery from data

Chairs: To be announced

PP28 - Unveiling new biological relationships using shared hits of chemical screening assay pairs

Xueping Liu[1,2] and Monica Campillos[1,2]

[1]Institute of Bioinformatics and Systems Biology and [2]German Center for Diabetes Research, Helmholtz Center Munich, 85764, Neuherberg, Germany.

ABSTRACT
Motivation: Although the integration and analysis of the activity of small molecules across multiple chemical screens is a common approach to determine the specificity and toxicity of hits, the suitability of these approaches to reveal novel biological information is less explored. Here, we test the hypothesis that assays sharing selective hits are biologically related.
Results: We annotated the biological activities (i.e. biological processes or molecular activities) measured in assays and constructed chemical hit profiles with sets of compounds differing on their selectivity level for 1,640 assays of ChemBank repository. We compared the similarity of chemical hit profiles of pairs of assays with their biological relationships and observed that assay pairs sharing non promiscuous chemical hits tend to be biologically related. A detailed analysis of a network containing assay pairs with the highest hit similarity confirmed biological meaningful relationships. Furthermore, the biological roles of predicted molecular targets of the shared hits reinforced the biological associations between assay pairs.
Contact: monica.campillos@helmholtz-muenchen.de

PP29 - Identification of structural features in chemicals associated with cancer drug response: A systematic data-driven analysis

Suleiman Ali Khan[1], Seppo Virtanen[1], Olli Kallioniemi[2], Krister Wennerberg[2], Antti Poso[2,3] and Samuel Kaski[1,4]

[1]Helsinki Institute for Information Technology HIIT, Department of Information and Computer Science, Aalto University, 00076 Espoo, Finland. [2]Institute for Molecular Medicine Finland FIMM, University of Helsinki, 00014 Helsinki, Finland. [3]School of Pharmacy, Faculty of Health Sciences, University of Eastern Finland, 70211 Kuopio, Finland. [4]Helsinki Institute for Information Technology HIIT, Department of Computer Science, University of Helsinki, 00014 Helsinki, Finland.

ABSTRACT
Motivation: Analysis of relationships of drug structure to biological response is key to understanding off-target and unexpected drug effects, and for developing hypotheses on how to tailor drug therapies. New methods are required for integrated analyses of a large number of chemical features of drugs against the corresponding genome-wide responses of multiple cell models.
Results: In this paper, we present the first comprehensive multi-set analysis on how the chemical structure of drugs impacts on genome-wide gene expression across several cancer cell lines (CMap database). The task is formulated as searching for drug response components across multiple cancers to reveal shared effects of drugs and the chemical features that may be responsible. The components can be computed with an extension of a very recent approach called Group Factor Analysis (GFA). We identify 11 components that link the structural descriptors of drugs with specific gene expression responses observed in the three cell lines, and identify structural groups that may be responsible for the responses. Our method quantitatively outperforms the limited earlier methods on CMap and identifies both the previously reported associations and several interesting novel findings, by taking into account multiple cell lines and advanced 3D structural descriptors. The novel observations
include: previously unknown similarities in the effects induced by 15-delta prostaglandin J2 and HSP90 inhibitors, which are linked to the 3D descriptors of the drugs; and the induction by simvastatin of leukemia-specific response, resembling the effects of corticosteroids.
Availability: Code http://research.ics.aalto.fi/mi/software/GFAsparse
Contact: suleiman.khan@aalto.fi, samuel.kaski@aalto.fi
Supplementary Information: Available at *Bioinformatics* online

Highlight Talk: *HP06 - Shaping the interaction landscape of bioactive molecules*

David Gfeller, Aurelien Grosdidier, Matthias Wirth, Antoine Daina, Olivier Michielin and Vincent Zoete

Swiss Institute of Bioinformatics, Lausanne, Switzerland.

ABSTRACT

Bioactive small molecules, such as drugs or metabolites, interact with proteins targets to modulate their activity, which in turn results in the observed phenotypic effects. However, for most bioactive compounds the list of targets is only partially known. Therefore computational predictions of bioactive molecule targets are powerful to narrow down the number of potential targets and to rationalize side effects of known molecules. Here, we introduce a new computational approach to accurately predict the targets of bioactive small molecules based on a combination of 2D and 3D similarity measures with known ligands. The method is trained on a large dataset of 280,381 small molecules interacting with 2686 targets from the ChEMBL database. Predictions can be carried out in five different organisms, and mapping predictions by homology within and between different species is enabled for close paralogs and orthologs. The method is accessible free of charge at http://www.swisstargetprediction.ch.

Publication:

Gfeller D, Michielin O, Zoete V. Shaping the interaction landscape of bioactive molecules. Bioinformatics. 2013 Dec 1;29(23):3073-9.

Contact: david.gfeller@gmail.com

Tue5 (Area B): Gene Expression (1)

Chair: Morgane Thomas-Chollier

PP30 - Estimating the activity of transcription factors by the effect on their target genes

Theresa Schacht[1,2,3], Marcus Oswald[1,2], Roland Eils[3,4], Stefan Eichmüller[5] and Rainer Koenig[1,2,3]

[1]Integrated Research and Treatment Center, Center for Sepsis Control and Care (CSCC), Jena University Hospital, D-07747 Jena, Erlanger Allee 101, Germany. [2]Network Modeling, Leibniz Institute for Natural Product Research and Infection Biology - Hans Knöll Institute Jena, Beutenbergstrasse 11a, 07745 Jena, Germany. [3]Theoretical Bioinformatics, German Cancer Research Center, INF 580, 69121 Heidelberg, Germany. [4]Department of Bioinformatics and Functional Genomics, Institute of Pharmacy and Molecular Biotechnology, and Bioquant, University of Heidelberg, Im Neuenheimer Feld 267, Germany. [5]Translational Immunology, German Cancer Research Center (DKFZ), INF 280, 69120 Heidelberg, Germany.

ABSTRACT

Motivation: Understanding regulation of transcription is central for elucidating cellular regulation. Several statistical and mechanistic models have come up the last couple of years explaining gene transcription levels using information of potential transcriptional regulators as transcription factors (TFs) and information from epigenetic modifications. The activity of TFs is often inferred by their transcription levels, promoter binding and epigenetic effects. However, in principle, these methods do not take hard-to-measure influences such as post-transcriptional modifications into account.

Results: For TFs, we present a novel concept circumventing this problem. We estimate the regulatory activity of TFs using their cumulative effects on their target genes. We established our model using expression data of 59 cell lines from the National Cancer Institute. The trained model was applied to an independent expression dataset of melanoma cells yielding excellent expression predictions and elucidated regulation of melanogenesis.

Implementation: Using mixed integer linear programming (MILP), we implemented a switch like optimization enabling a constrained but optimal selection of TFs and optimal model selection estimating their effects. The method is generic and can also be applied to further regulators of transcription.

Contact: Rainer.Koenig@uni-jena.de

PP31 - Modeling DNA methylation dynamics with approaches from phylogenetics

John A. Capra[1] and Dennis Kostka[2]

[1]Center for Human Genetics Research and Department of Biomedical Informatics, Vanderbilt University, Nashville, TN, 37232, USA. [2]Departments of Developmental Biology and Computational & Systems Biology, University of Pittsburgh, Pittsburgh, PA, 15201, USA.

ABSTRACT

Motivation: Methylation of CpG dinucleotides is a prevalent epigenetic modification that is required for proper development in vertebrates. Genome-wide DNA methylation assays have become increasingly common, and this has enabled characterization of DNA methylation in distinct stages across differentiating cellular lineages. Changes in CpG methylation are essential to cellular differentiation; however, current methods for modeling methylation dynamics do not account for the dependency structure between precursor and dependent cell

types.
Results: We developed a continuous-time Markov chain approach, based on the observation that changes in methylation state over tissue differentiation can be modeled similarly to DNA nucleotide changes over evolutionary time. This model explicitly takes precursor to descendant relationships into account and enables inference of CpG methylation dynamics. To illustrate our method, we analyzed a high-resolution methylation map of the differentiation of mouse stem cells into several blood cell types. Our model can successfully infer unobserved CpG methylation states from observations at the same sites in related cell types (90% correct), and this approach more accurately reconstructs missing data than imputation based on neighboring CpGs (84% correct). Additionally, the single CpG resolution of our methylation dynamics estimates enabled us to show that DNA sequence context of CpG sites is informative about methylation dynamics across tissue differentiation. Finally, we identified genomic regions with clusters of highly dynamic CpGs and present a likely functional example. Our work establishes a framework for inference and modeling that is well-suited to DNA methylation data, and our success suggests that other methods for analyzing DNA nucleotide substitutions will also translate to the modeling of epigenetic phenomena.
Availability: Source code is available at www.kostkalab.net/software .
Contact: tony.capra@vanderbilt.edu , kostka@pitt.edu

Highlight Talk: *HP07 - Key regulators control distinct transcriptional programmes in blood and progenitor and mast cells*

Felicia Ng, Fernando Calero-Nieto, Nicola Wilson, Rebecca Hannah, Victoria Moignard, Ana Leal-Cervantes, Isabel Jimenez-Madrid, Evangelia Diamanti, Lorenz Wernisch and Berthold Gottgens

Cambridge Institute for Medical Research, United Kingdom.

ABSTRACT
Despite major advances in the generation of genome-wide binding maps, the mechanisms by which transcription factors (TFs) regulate cell type identity have remained largely obscure. Through comparative analysis of 10 key haematopoietic TFs in both mast cells and blood progenitors, we demonstrate that the largely cell type-specific binding profiles are not opportunistic, but instead contribute to cell type-specific transcriptional control, because (i) mathematical modelling of differential binding of shared TFs can explain differential gene expression, (ii) consensus binding sites are important for cell type-specific binding and (iii) knock-down of blood stem cell regulators in mast cells reveals mast cell-specific genes as direct targets. Finally, we show that the known mast cell regulators Mitf and c-fos likely contribute to the global reorganisation of TF binding profiles. Taken together therefore, our study elucidates how key regulatory TFs contribute to transcriptional programmes in several distinct mammalian cell types.
Publication:
Calero-Nieto FJ, Ng FS, Wilson NK, Hannah R, Moignard V, Leal-Cervantes AI, Jimenez-Madrid I, Diamanti E, Wernisch L, Göttgens B. Key regulators control distinct transcriptional programmes in blood progenitor and mast cells. EMBO J. 2014 Jun 2;33(11):1212-26.

Contact: fn231@cam.ac.uk

Tue6 (Area F): Gene expression (2)

Chairs: To be announced

PP32 - Two-dimensional segmentation for analyzing HiC data

Celine Levy-Leduc[1], Maud Delattre[1], Tristan Mary-Huard[1,2] and Stephane Robin[1]

[1]AgroParisTech/INRA MIA 518, 16 rue Claude Bernard, 75005 Paris, France. [2]UMR de Génétique Végétale, INRA/Université Paris-Sud/CNRS, 91190 Gif-sur-Yvette, France.

ABSTRACT
Motivation: The spatial conformation of the chromosome has a deep influence on gene regulation and expression. HiC technology allows the evaluation of the spatial proximity between any pair of loci along the genome. It results in a data matrix where blocks corresponding to (self-)interacting regions appear. The delimitation of such blocks is critical to better understand the spatial organization of the chromatin. From a computational point of view, it results in a 2D-segmentation problem.
Results: We focus on the detection of cis-interacting regions, which appear to be prominent in observed data. We define a block-wise segmentation model for the detection of such regions. We prove that the maximization of the likelihood with respect to the block boundaries can be rephrased in terms of a 1D-segmentation problem, for which the standard dynamic programming applies. The performance of the proposed methods are assessed by a simulation study on both synthetic and re-sampled data. A comparative study on public data shows good concordance with biologically confirmed regions.
Availability: The HiCseg R package is available from the Comprehensive R Archive Network (CRAN) and from the web page of the corresponding author.
Contact: celine.levy-leduc@agroparistech.fr

PP33 - Broad-Enrich: Functional interpretation of large sets of broad genomic regions

Raymond Cavalcante[1], Chee Lee[1], Ryan Welch[1,2], Snehal Patil[3], Terry Weymouth[3], Laura Scott[2] and Maureen Sartor[1,2,3]

[1]Department of Computational Medicine and Bioinformatics, [2]Department of Biostatistics, and [3]Center of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI 48109, USA.

ABSTRACT

Motivation: Functional enrichment testing facilitates the interpretation of ChIP-seq data in terms of pathways and other biological contexts. Previous methods developed and used to test for key gene sets affected in ChIP-seq experiments treat peaks as points, and are based on the number of peaks associated with a gene or a binary score for each gene. These approaches work well for transcription factors, but histone modifications often occur over broad domains, and across multiple genes.

Results: To incorporate the unique properties of broad domains into functional enrichment testing, we developed Broad-Enrich, a method that uses the proportion of each gene's locus covered by a peak. We show that our method has a well-calibrated false positive rate, performing well with ChIP-seq data having broad domains compared to alternative approaches. We illustrate Broad-Enrich with 55 ENCODE ChIP-seq datasets using different methods to define gene loci. Broad-Enrich can also be applied to other datasets consisting of broad genomic domains such as copy number variations.

Availability: http://broad-enrich.med.umich.edu for web version and R package.

Contact: sartorma@umich.edu

Supplementary Information: Supplementary data are available at *Bioinformatics* online.

Highlight Talk: *HP08 - Chromatin position effects quantified from thousands of reporters integrated in parallel*

Lodewyk Wessels[1], Waseem Akhtar[1], Johann de Jong[1], Alex Pindyurin[2], Ludo Pagie[1], Wouter Meuleman[3], Jeroen de Ridder[4], Anton Berns[1], Maarten van Lohuizen[1] and Bas van Steensel[1]

[1]Netherlands Cancer Institute, Amsterdam, The Netherlands. [2]Institute of Molecular and Cellular Biology, Novosibirsk, Russian Federation. [3]Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [4]Delft University of Technology, The Netherlands.

ABSTRACT

Reporter genes integrated into the genome are a powerful tool to reveal effects of regulatory elements and local chromatin context on gene expression. However, such assays have been low throughput. Here, we describe an approach to monitor transcriptional activity of thousands of randomly integrated reporters. Computational analyses of more than 27,000 distinct reporter integrations in mouse embryonic stem cells reveal the following. First, lamina associated domains act as attenuators of transcription, likely by reducing access of transcription factors to binding sites. Second, chromatin compaction as derived from HiC data is predictive of reporter activity. Third, we find evidence of cross-talk between neighbouring genes and estimate that enhancers can influence gene expression on average over ~20 kb. Most importantly, the richness and size of the datasets opens up the opportunity for additional extensive and robust computational analyses. We will showcase the utility with recent analyses shedding new light on gene regulation.

Publication:

Akhtar W, de Jong J, Pindyurin AV, Pagie L, Meuleman W, de Ridder J, Berns A, Wessels LF, van Lohuizen M, van Steensel B. Chromatin position effects assayed by thousands of reporters integrated in parallel. Cell. 2013 Aug 15;154(4):914-27.

Contact: l.wessels@nki.nl

Wednesday, September 10, 2014

Wed1 (Area H): Biological Ontologies

Chairs: To be announced

PP34 - The impact of incomplete knowledge on the evaluation of protein function prediction: a structured-output learning perspective

Yuxiang Jiang[1], Wyatt Clark[1], Iddo Friedberg[2,3] and Predrag Radivojac[1]

[1]Department of Computer Science and Informatics, Indiana University, Bloomington, Indiana, USA. [2]Department of Microbiology, Miami University, Oxford, Ohio, USA. [3]Department of Computer Science and Software Engineering, Miami University, Oxford, Ohio, USA.

ABSTRACT
Motivation: The automated functional annotation of biological macro-molecules is a problem of computational assignment of biological concepts or ontological terms to genes and gene products. A number of methods have been developed to computationally annotate genes using standardized nomenclature such as Gene Ontology (GO). However, questions remain about the possibility for development of accurate methods that can integrate disparate molecular data as well as about an unbiased evaluation of these methods. One important concern is that experimental annotations of proteins are incomplete. This raises questions as to whether and to what degree currently available data can be reliably used to train computational models and estimate their performance accuracy.
Results: We study the effect of incomplete experimental annotations on the reliability of performance evaluation in protein function prediction. Using the structured-output learning framework, we provide theoretical analyses and carry out simulations to characterize the effect of growing experimental annotations on the correctness and stability of performance estimates corresponding to different types of methods. We then analyze real biological data by simulating the prediction, evaluation, and subsequent re-evaluation (after additional experimental annotations become available) of GO term predictions. Our results agree with previous observations that incomplete and accumulating experimental annotations have the potential to significantly impact accuracy assessments. We find that their influence reflects a complex interplay between the prediction algorithm, performance metric, and underlying ontology. However, using the available experimental data and under realistic assumptions, our results also suggest that current large-scale evaluations are meaningful and almost surprisingly reliable.
Contact: predrag@indiana.edu
Supplementary information: Supplementary data are available at Bioinformatics online.

PP35 - Integration of molecular network data reconstructs Gene Ontology

Vladimir Gligorijevic, Vuk Janjic and Natasa Przulj

Department of Computing, Imperial College London, SW7 2AZ, UK.

ABSTRACT
Motivation: Recently, a shift was made from using Gene Ontology (GO) to evaluate molecular network data to using these data to construct and evaluate GO: Dutkowski et al. [2013] provide the first evidence that a large part of GO can be reconstructed solely from topologies of molecular networks. Motivated by this work, we develop
a novel data integration framework that integrates multiple types of molecular network data to reconstruct and update GO. We ask how much of GO can be recovered by integrating various molecular interaction data.
Results: We introduce a computational framework for integration of various biological networks using Penalized Non-negative Matrix Tri-Factorization (PNMTF). It takes all network data in a matrix form and performs simultaneous clustering of genes and GO terms, inducing new relations between genes and GO terms (annotations) and between GO terms themselves. To improve the accuracy of our predicted relations, we extend the integration methodology to include additional topological information represented as the similarity in wiring around non-interacting genes. Surprisingly, by integrating topologies of bakers yeasts protein-protein interaction, genetic interaction and co-expression networks, our method reports as related 96% of GO terms that are directly
related in GO. The inclusion of the wiring similarity of non-interacting genes contributes 6% to this large GO-term association capture. Furthermore, we use our method to infer new relationships between GO terms solely from the topologies of these networks and validate 44% of our predictions in the literature. In addition, our integration
method reproduces 48% of cellular component, 41% of molecular function and 41% of biological process GO terms, outperforming the previous method in the former two domains of GO. Finally, we predict new GO annotations of yeast genes and validate our predictions through genetic interactions profiling.
Supplementary information: Supplementary Tables of new GO term associations and predicted gene annotations are available at: http://bio-nets.doc.ic.ac.uk/GO-Reconstruction/ .
Contact: natasha@imperial.ac.uk

Wed2 (Area G): Bioinformatics of Health and Disease (2) and Bio-imaging

Chair: Lodewyk Wessels

PP36 - Transcriptome-guided amyloid imaging genetic analysis via a novel structured sparse learning algorithm

Jingwen Yan[1,2], Lei Du[2], Sungeun Kim[2], Shannon Risacher[2], Heng Huang[3], Jason Moore[4], Andrew Saykin[2] and Li Shen[2], and the Alzheimer's Disease Neuroimaging Initiative[§]

[1]BioHealth, Indiana University School of Informatics & Computing, Indianapolis, IN, 46202, USA.

[2]Radiology & Imaging Sciences, Indiana University Sch. of Medicine, Indianapolis, IN, 46202, USA. [3]Computer Science & Engineering, The University of Texas at Arlington, TX, 76019, USA. [4]Genetics, Community & Family Medicine, Dartmouth Medical School, Lebanon, NH, 03756, USA. [§]A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

ABSTRACT

Motivation: Imaging genetics is an emerging field that studies the influence of genetic variation on brain structure and function. The major task is to examine the association between genetic markers such as single nucleotide polymorphisms (SNPs) and quantitative traits (QTs) extracted from neuroimaging data. The complexity of these data sets have presented critical bioinformatics challenges that require new enabling tools. Sparse canonical correlation analysis (SCCA) is a bi-multivariate technique used in imaging genetics to identify complex multi-SNP-multi-QT associations. However, most of the existing SCCA algorithms are designed using the soft thresholding method, which assumes that the input features are independent from one another. This assumption clearly does not hold for the imaging genetic data. In this paper, we propose a new knowledge-guided SCCA algorithm (KG-SCCA) to overcome this limitation as well as improve learning results by incorporating valuable prior knowledge.

Results: The proposed KG-SCCA method is able to model two types of prior knowledge: one as a group structure (e.g., linkage disequilibrium blocks among SNPs) and the other as a network structure (e.g., gene co-expression network among brain regions). The new model incorporates these prior structures by introducing new regularization terms to encourage weight similarity between grouped or connected features. A new algorithm is designed to solve the KG-SCCA model without imposing the independence constraint on the input features. We demonstrate the effectiveness of our algorithm with both synthetic and real data. For real data, using an Alzheimer's disease (AD) cohort, we examine the imaging genetic associations between all SNPs in the *APOE* gene (i.e., top AD gene) and amyloid deposition measures among cortical regions (i.e., a major AD hallmark). In comparison with a widely used SCCA implementation, our KG-SCCA algorithm produces not only improved cross-validation performances but also biologically meaningful results.

Availability: Software is freely available upon request.

Contact: shenli@iu.edu

PP37 - Large-scale automated identification of mouse brain cells in confocal light sheet microscopy images

Paolo Frasconi[1], Ludovico Silvestri[2], Paolo Soda[3], Roberto Cortini[1], Francesco Pavone[2] and Giulio Iannello[3]

[1]Department of Information Engineering (DINFO), Università di Firenze, Italy. [2]European Laboratory for Nonlinear Spectroscopy (LENS), Università di Firenze, Italy. [3]Integrated Research Centre, Università Campus Bio-Medico di Roma, Italy.

ABSTRACT

Motivation: Recently, confocal light sheet microscopy has enabled high-throughput acquisition of whole mouse brain 3D images at the micron scale resolution. This poses the unprecedented challenge of creating accurate digital maps of the whole set of cells in a brain.

Results: We introduce a fast and scalable algorithm for fully automated cell identification. We obtained the whole digital map of Purkinje cells in mouse cerebellum consisting of a set of 3D cell center coordinates. The method is very accurate and we estimated an F_1 measure of 0.96 using 56 representative volumes, totaling 1.09 GVoxel and containing 4,138 manually annotated soma centers.

Availability and implementation: Source code and its documentation are available at http://bcfind.dinfo.unifi.it/. The whole pipeline of methods is implemented in Python and makes use of ylearn2 (Goodfellow et al., 2013) and modified parts of Scikitlearn (Pedregosa et al., 2011). Brain images are available on request.

Contact: paolo.frasconi@unifi.it

Supplementary information: Coordinates of predicted soma centers of a whole mouse cerebellum and additional figures.

Highlight Talk: *HP09 - Novel developments in computational clinical breath analysis and biomarker detection.*

Anne-Christin Hauschild[1], Jörg Ingo Baumbach[2] and Jan Baumbach[3]

[1]Max Planck Institute for Informatics, Saarbrücken, Germany. [2]Faculty Applied Chemistry, Reutlingen University, Germany. [3]University of Southern Denmark, Denmark.

ABSTRACT

The volatolom is the sum of volatile organic compounds that are emitted by all living cells and tissues. We seek to non-invasively "sniff" biomarker molecules that are predictive for the biomedical fate of individual patients. This promises great hope to move the therapeutic windows to earlier stages of disease progression. While portable devices for breathomics measurement exist, we face the traditional biomarker research barrier: a lack of robustness hinders translation to the world outside laboratories. To move from biomarker discovery to

validation, from separability to predictability, we have developed several bioinformatics methods for computational breath analysis, which have the potential to redefine non-invasive biomedical decision making by rapid and cheap matching of decisive medical patterns in exhaled air. We aim to provide a supplementary diagnostic tool complementing classic urine, blood and tissue samples. The presentation will review the state of the art, highlight existing challenges and introduce new data mining methods for identifying breathomics biomarkers.

Publications:

Hauschild AC, Kopczynski D, D'Addario M, Baumbach JI, Rahmann S, Baumbach J. Peak detection method evaluation for ion mobility spectrometry by using machine learning approaches. Metabolites. 2013 Apr 16;3(2):277-93.

Eckel SP, Baumbach J, Hauschild AC. On the importance of statistics in breath analysis--hope or curse? J Breath Res. 2014 Mar;8(1):012001.

Maurer F, Hauschild AC, Eisinger K, Baumbach J, Mayor A and Baumbach JI. MIMA - a software for analyte identification in MCC/IMS chromatograms by mapping accompanying GC/MS measurements. Int. J. Ion Mobil. Spec. 2014 Apr:17:95–101.

Smolinska A, Hauschild AC, Fijten RR, Dallinga JW, Baumbach J, van Schooten FJ. Current breathomics-a review on data pre-processing techniques and machine learning in metabolomics breath analysis. J Breath Res. 2014 Jun;8(2):027105.

Contact: jan.baumbach@imada.sdu.dk

Wed3 (Area H): Text Mining for Computational Biology

Chairs: To be announced

PP38 - Extracting patterns of database and software usage from the bioinformatics literature

Geraint Duck[1], Goran Nenadic[1,2], Andy Brass[1,3], David Robertson[3] and Robert Stevens[1]

[1]School of Computer Science, University of Manchester, UK. [2]Manchester Institute of Biotechnology, University of Manchester, UK. [3]Computational and Evolutionary Biology, Faculty of Life Sciences, University of Manchester, UK.

ABSTRACT

Motivation: As a natural consequence of being a computer-based discipline, bioinformatics has a strong focus on database and software development, but the volume and variety of resources are growing at unprecedented rates. An audit of database and software usage patterns could help provide an overview of developments in bioinformatics and community common practice, and comparing the links between resources through time could demonstrate both the persistence of existing software and the emergence of new tools.

Results: We study the connections between bioinformatics resources and construct networks of database and software usage patterns, based on resource co-occurrence, that correspond to snapshots of common practice in the bioinformatics community. We apply our approach to pairings of phylogenetics software reported in the literature, and argue that these could provide a stepping-stone into the identification of scientific best practice.

Availability: The extracted resource data, the scripts used for network generation and the resulting networks are available at: http://bionerds.sourceforge.net/networks/

Contact: robert.stevens@manchester.ac.uk

Highlight Talk: *HP10 - Text mining technologies for database curation*

Fabio Rinaldi[1], Simon Clematide[1], Simon Hafner[1], Gerold Schneider[1], Gintare Grigonyte[2], Martin Romacker[3] and Therese Vachon[4]

[1]University of Zurich, Switzerland. [2]University of Stockholm, Sweden. [3]F. Hoffmann-LaRoche, Switzerland. [4]Novartis, Switzerland.

ABSTRACT

Although human curation for life science databases offers the best guarantee of high quality annotations, it suffers from severe bottlenecks which have long been recognized in the curation community. The most pressing problem is that of efficiency of the process: it is impossible for human curators to keep up with the growing pace of publication. Text mining technologies, coupled with advanced user interfaces, offer the potential to partially alleviate this bottleneck. We survey the results of several recent competitive evaluations of text mining technologies, discuss how text mining systems can be integrated in a curation workflow, and illustrate our approach to assisted curation, which has been tested in collaboration with major databases.

Publication :

Rinaldi F, Clematide S, Hafner S, Schneider G, Grigonyte G, Romacker M, Vachon T. Using the OntoGene pipeline for the triage task of BioCreative 2012. Database (Oxford). 2013 Feb 9;2013:bas053.

Contact: fabio@ontogene.org

Wed4 (Areas A and E): RNA prediction

Chair: Yann Ponty

PP39 - CRISPRstrand: Predicting repeat orientations to determine the crRNA-encoding strand at CRISPR loci

Omer S. Alkhnbashi[1], Fabrizio Costa[1], Shiraz A. Shah[2], Roger A. Garrett[2], Sita J. Saunders[1] and Rolf Backofen[1,3]

[1]Bioinformatics Group, Department of Computer Science, University of Freiburg, Georges-Köhler-Allee 106, 79110 Freiburg, Germany, [2]Archaea Centre, Department of Biology, University of Copenhagen, Ole Maaloes Vej 5, DK2200 Copenhagen, Denmark, [3]BIOSS Centre for Biological Signalling Studies, Cluster of Excellence, University of Freiburg, Germany.

ABSTRACT

Motivation: The discovery of CRISPR-Cas systems almost 20 years ago rapidly changed our perception of the bacterial and archaeal immune systems. CRISPR loci consist of several repetitive DNA sequences called repeats, inter-spaced by stretches of variable length sequences called spacers. This CRISPR array is transcribed and processed into multiple mature RNA species (crRNAs). A single crRNA is integrated into an interference complex, together with CRISPR-associated (Cas) proteins, to bind and degrade invading nucleic acids. Although existing bioinformatics tools can recognize CRISPR loci by their characteristic repeat-spacer architecture, they generally output CRISPR arrays of ambiguous orientation and thus do not determine the strand from which crRNAs are processed. Knowledge of the correct orientation is crucial for many tasks, including the classification of CRISPR conservation, the detection of leader regions, the identification of target sites (protospacers) on invading genetic elements, and the characterization of protospacer-adjacent motifs (PAMs).

Results: We present a fast and accurate tool to determine the crRNA-encoding strand at CRISPR loci by predicting the correct orientation of repeats based on an advanced machine learning approach. Both the repeat sequence and mutation information were encoded and processed by an efficient graph kernel to learn higher order correlations. The model was trained and tested on curated data comprising more than 4,500 CRISPRs and yielded a remarkable performance of 0.95 AUC ROC (area under the curve of the receiver operator characteristic). In addition, we show that accurate orientation information greatly improved detection of conserved repeat sequence families and structure motifs. We integrated CRISPRstrand predictions into our CRISPRmap web server of CRISPR conservation and updated the latter to version 2.0.

Availability: CRISPRmap and CRISPRstrand are available at http://rna.informatik.uni-freiburg.de/CRISPRmap

Contact: backofen@informatik.uni-freiburg.de

PP40 - Towards a piRNA prediction using multiple kernel fusion and support vector machine

Jocelyn Brayet[1,2], Farida Zehraoui[1], Laurence Jeanson-Leh[2], David Israeli[2] and Fariza Tahi[1]

[1]IBISC, UEVE/Genopole, IBGBI, 23 bv. de France, 91000 Evry, France. [2]Genethon, 1, bis rue de l'Internationale, 91002 Evry Cedex, France.

ABSTRACT

Motivation: Piwi interacting RNA (piRNA) is the most recently discovered and the least investigated class of AGO/Piwi protein interacting small non-coding RNAs. PiRNAs are mostly known to be involved in protecting the genome from invasive transposable elements. But recent discoveries suggest their involvement in the pathophysiology of diseases, such as cancer. Their identification is therefore an important task, and computational methods are needed. However, the lack of conserved piRNA sequences and structural elements makes this identification very challenging and difficult.

Results: In the present study, we propose a new modular and extensible machine learning method based on multiple kernels and a support vector machine (SVM) classifier for piRNA identification. Very few piRNA features are known to date. The use of a multiple kernels approach allows editing, adding or removing piRNA features that can be heterogeneous in a modular manner according to their relevance in a given species. Our algorithm is based on a combination of the previously identified features (sequence features (k-mer motifs and a uridine at the first position) and piRNAs cluster feature) and a new telomere/centromere vicinity feature. These features are heterogeneous and the kernels allow to unify their representation. The proposed algorithm, named piRPred, gives very promising results on Drosophila and Human data and outscores previously published piRNA identification algorithms.

Availability: piRPred is freely available to non-commercial users on our Web server EvryRNA: http://EvryRNA.ibisc.univ-evry.fr

Contact: tahi@ibisc.univ-evry.fr

Industrial and Academic Demos

The purpose of the Industrial and Demo Track is to provide commercial organizations and academic institutions the opportunity to present their products or services in a scheduled series of "demo" presentations. Demonstrators will be given a one-hour time-slot in which they are strongly encouraged to give two demos/presentations each of up to 30 minutes, including discussion and exchanges with the audience. Alternatively the demo presenters can choose to present their two demos on different days. The aim of giving two demos in the one-hour slot is to reach as large a number of conference attendees as possible. Presenters must therefore finish their first half-hour slot on time to allow attendees to visit another demo/presentation.

Room Assignment and Schedule

Monday, September 8

ID01	Bringing the Tools to the Data – Providing Scientists with Personalized and On-Demand Bioinformatics Services on the Cloud of the French Institute of Bioinformatics - IFB	Oberlin	13:30	14:00
ID02	A Panel of European Public Galaxy Instances - Galaxy	Oberlin	14:00	14:30
ID03	Building and Testing Executable Biological Models in the BioModelAnalyzer – Microsoft Research	Leicester	13:30	14:00
ID04	EMC Isilon – the Foundation of NGS Data Analysis	Leicester	14:00	14:30
ID05	UniProt: New Website and Latest Developments	Schumann	13:30	14:00
ID06	Enhanced Human NGS Variant and Gene Regulation Analysis – BIOBASE (A)	Schumann	14:00	14:30
ID07	A Reputation-Based Web Application (sbv IMPROVER Network Verification Challenge) that Facilitates Collaboration and Applications on Biological Network Models (Part1)	Boston	13:30	14:00
ID08	KLAST: Fast, Accurate and NGS Scalable Bank-To-Bank Sequence Similarity Search Tool - KoriScale	Boston	14:00	14:30
ID09	Protein Model Portal and SWISS-MODEL Workspace: Giving the Proteome a Third Dimension - SIB	Bartholdi	13:30	14:00
1D10	BioMercator 4.0: A Complete Framework to Integrate QTL, Meta-QTL and Genome Annotation - INRA	Bartholdi	14:00	14:30
ID12	Garuda : Fly to the Future of Biology – SBI and RIKEN	Rohan	13:30	14:00
ID12	Garuda : Fly to the Future of Biology – SBI and RIKEN	Rohan	14:00	14:30
ID13	TotaLinux	Stuttgart	13:30	14:00
ID13	TotaLinux	Stuttgart	14:00	14:30

Tuesday, September 9

ID01	Bringing the Tools to the Data – Providing Scientists with Personalized and On-Demand Bioinformatics Services on the Cloud of the French Institute of Bioinformatics - IFB	Oberlin	13:30	14:00
ID02	A Panel of European Public Galaxy Instances - Galaxy	Oberlin	14:00	14:30
ID03	Building and Testing Executable Biological Models in the BioModelAnalyzer – Microsoft Research	Leicester	13:30	14:00
ID04	EMC Isilon – the Foundation of NGS Data Analysis	Leicester	14:00	14:30

ID05	UniProt: New Website and Latest Developments	Schumann	13:30	14:00
ID06	Enhanced Human NGS Variant and Gene Regulation Analysis – BIOBASE (B)	Schumann	14:00	14:30
ID08	KLAST: Fast, Accurate and NGS Scalable Bank-To-Bank Sequence Similarity Search Tool - KoriScale	Boston	13:30	14:00
ID07	A Reputation-Based Web Application (sbv IMPROVER Network Verification Challenge) that Facilitates Collaboration and Applications on Biological Network Models (Part 2)	Boston	14:00	14:30
ID09	Protein Model Portal and SWISS-MODEL Workspace: Giving the Proteome a Third Dimension - SIB	Bartholdi	13:30	14:00
1D10	BioMercator 4.0: A Complete Framework to Integrate QTL, Meta-QTL and Genome Annotation - INRA	Bartholdi	14:00	14:30
ID11	iPlant Collaborative™ A Scalable Cyberinfrastructure for Life Science – Cold Spring Harbor	Rohan	13:30	14:00
ID11	iPlant Collaborative™ A Scalable Cyberinfrastructure for Life Science – Cold Spring Harbor	Rohan	14:00	14:30
ID18	Goes Open: publish with Genomics, Proteomics & Bioinformatics	Stuttgart	13:30	14:00
ID18	Goes Open: publish with Genomics, Proteomics & Bioinformatics	Stuttgart	14:00	14:30

Wednesday, September 10

Wednesday presentations are given only once as they correspond to papers submitted to the "Methods and technologies for computational biology" track that were selected as deserving demo.

ID14	Sushi: An Exquisite Recipe for Fully Documented, Reproducible and Reusable NGS Data Analysis	Boston	13:30	14:00
ID15	JAMM: A Peak Finder for Joint Analysis of NGS Replicates	Boston	14:00	14:30
ID16	POPS: Predicting and Enhancing Protein Solubility	Leicester	13:30	14:00
ID17	Computational Tools for the Taxonomic Analysis of Shotgun Metagenome Samples	Leicester	14:00	14:30

ID01: Bringing the Tools to the Data – Providing Scientists with Personalized and On-Demand Bioinformatics Services on the Cloud of the French Institute of Bioinformatics

Life science researchers, thanks to the continuous improvement of experimental technologies, face a deluge of data whose exploitation requires large computing resources and appropriate software tools. To tackle these issues, the French Institute of Bioinformatics (IFB) developed predefined, personalized bioinformatics cloud services and turnkey appliances encapsulating common bioinformatics tools, workflows and gateways. The size of these appliances, at most a few gigabytes, allows users to move them easily to the data, rather than moving the data (whose size can be three order of magnitude larger - terabytes) to the computing resources. IFB is currently running an academic cloud infrastructure with the appropriate biological data and bioinformatics tools to meet the needs of the life science community.

Presentation Flier: www.eccb14.org/programme/id_track/ID01-summary.pdf

Presenters

Christophe Blanchet [Christophe.Blanchet@france-bioinformatique.fr]

Jean-Francois Gibrat [Jean-Francois.Gibrat@france-bioinformatique.fr]

Institut Français de Bioinformatique, IFB-core, CNRS UMS3601, France

External Link: http://www.france-bioinformatique.fr

ID02: A Panel of European Public Galaxy Instances

Galaxy, a web-based platform for data intensive biomedical research, is available as a web server (http://UseGalaxy.org), and as customisable open source software. The first session will introduce the Galaxy project and then the deepTools webserver. The second session will feature the public Galaxy instances "Nebula" and "IFB's cloud".

Presentation Flier: www.eccb14.org/programme/id_track/ID02-summary.pdf

Presenters

Hans-Rudolf Hotz[1] [hansrudolf.hotz@fmi.ch] ; Björn Grüning[2] [bjoern.gruening@gmail.com] ; Alban Lermine[3] [alban.lermine@curie.fr] ; Valentina Boeva[3] [Valentina.Boeva@curie.fr] ; Olivier Inizan[4] [olivier.inizan@versailles.inra.fr] ; Christophe Blanchet[5] [Christophe.Blanchet@france-bioinformatique.fr].

[1] Friedrich Miescher Institute for Biomedical Research, Maulbeerstrasse 66, 4058 Basel, Switzerland

[2] Albert-Ludwig-Universität Freiburg, Georges-Köhler-Allee 106, 79110 Freiburg, Germany

[3] Institut Curie, INSERM U900, Mines ParisTech, 26 rue d'Ulm, 75248 Paris, France

[4] INRA, Unité de recherche en genomique info. RD 10 Rte de Saint Cyr, 78026 Versailles, France

[5] IFB, CNRS UMS 3601, Avenue de la Terrasse - Bât 21, 91190 Gif-sur-Yvette, France

External Link: http://UseGalaxy.org

ID03: Building and Testing Executable Biological Models in the BioModelAnalyzer

Executable models of biological phenomena offer a new set of techniques to address problems arising from biological complexity. Here we present the BioModelAnalyzer, a web-based tool for building and testing executable models, and showcase how it can be used to identify missing protein-protein interactions and understand complex signaling networks.

Presentation Flier: www.eccb14.org/programme/id_track/ID03-summary.pdf

Presenter

Ben Hall [benhall@microsoft.com]

Microsoft Research, Cambridge, CB1 2FB, UK

External Link: http://biomodelanalyzer.research.microsoft.com/

ID04: EMC Isilon – the Foundation of NGS Data Analysis

NGS usually means PBytes of data. But not only that, the workflow involves different steps, different files with different sizes and bandwidth requirements and different algorithms are used to do the data analysis. And to add another angle different algorithms may even use different protocols to access the data. EMC Isilon takes care of all that in one system. It can be flexibly deployed to your data storage needs and handles all possible requirements. So you end up doing your analysis not solving the data storage problem.

Presentation Flier: www.eccb14.org/programme/id_track/ID04-summary.pdf

Presenter

Wolfgang Mertz [wolfgang.mertz@emc.com]

Isilon Storage Division, EMC International S.à.r.l., Osterfeldstr. 84, 85737 Ismaning, Germany

ID05: UniProt: New Website and Latest Developments

The demonstration will cover:
1. A description of UniProt
2. New UniProt beta site, functionalities and design
3. Accessing UniProt using simple query syntax
4. Proteomes querying and retrieval
5. Sequence similarity searches, alignments and ID mapping tools provided

Presentation Flier: www.eccb14.org/programme/id_track/ID05-summary.pdf

Presenter

Diego Poggioli [diego@ebi.ac.uk]

EMBL-European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, CB10 1SD, UK

External Link: http://www.ebi.ac.uk/

ID06: Enhanced Human NGS Variant and Gene Regulation Analysis

BIOBASE is the leading provider of expert-curated biological databases, software and services for the life sciences. Our products and services identify relations critical to drug and biomarker discovery as well as improve biomedical research by transforming data into scientific concepts.
In the first session (A), you will learn more about the TRANSFAC® database of eukaryotic transcription factors and in the second session (B) more about the NGS analysis tool Genome Trax™.

Presentation Flier: www.eccb14.org/programme/id_track/ID06-summary.pdf

Presenter

Jasmin Dröge [Jasmin.Droege@biobase-international.com]

BIOBASE GmbH, Halchtersche Str. 33, 38304 Wolfenbüttel, Germany

External Link: http://www.biobase-international.com/

BIOBASE Exhibition Booth N°4

ID07: A Reputation-Based Web Application (sbv IMPROVER Network Verification Challenge) that Facilitates Collaboration and Applications on Biological Network Models

Part 1. Collaborative crowd-verification approach allows domain experts from various fields of biology to gather robust peer-reviewed information from which relationships are identified and evaluated. The approach is used to investigate smarter solutions to complement classical peer review and expand biological network models suitable for research in respiratory diseases.
Part 2. Using transcriptomics data, we assess the activation of causal network models from sbv IMPROVER BioNet in a quantitative, statistical and interpretable manner. The Biological Impact Factor (BIF) quantifies the response of a cell system to an insult by aggregating the overall network responses.

Presentation Flier: www.eccb14.org/programme/id_track/ID07-summary.pdf

Presenters

Jennifer Park [jpark@selventa.com] ; Florian Martin [florian.martin@pmi.com]

Philip Morris International R&D, Quai Jeanrenaud 5, Neuchatel, 2000, Switzerland

External Link: www.pmi.com

sbvIMPROVER Exhibition Booth N°14

ID08: KLAST: Fast, Accurate and NGS Scalable Bank-To-Bank Sequence Similarity Search Tool

KLAST is a professional sequence similarity search tool for processing high volumes of genomic sequences. The demo will present the use of the KLAST software on the following application domains: genome annotation, genome comparison and analysis of metagenomic data.

Presentation Flier: www.eccb14.org/programme/id_track/ID08-summary.pdf

Presenters

Dominique Lavenier [dominique.lavenier@irisa.fr] ; Patrick Durand [pdurand@korilog.com]

Inria/IRISA, Campus de Beaulieu, Avenue du General Leclerc, 35042 Rennes, France

External Link: https://koriscale.inria.fr

KoriScale Exhibition Booth N°5

ID09: Protein Model Portal and SWISS-MODEL Workspace: Giving the Proteome a Third Dimension

The three-dimensional structure of a protein provides valuable information for understanding its molecular function and guides the rational design of experiments. In this tutorial, we will illustrate how protein structure homology modelling can be used to study proteins still lacking experimental characterization.

Presentation Flier: www.eccb14.org/programme/id_track/ID09-summary.pdf

Presenters

Juergen Haas[1,2] [juergen.haas@unibas.ch] ; Torsten Schwede[1,2] [torsten.schwede@unibas.ch]

[1]Biozentrum Unversität Basel, Klingelbergstr 50-70, 4056 Basel, Switzerland

[2]SIB - Swiss Institute of Bioinformatics, Klingelbergstr 50-70, 4056 Basel, Switzerland

External Link(s): www.proteinmodelportal.org | www.swissmodel.expasy.org

SIB Exhibition Booth N°12

ID10: BioMercator 4.0: A Complete Framework to Integrate QTL, Meta-QTL and Genome Annotation

Compilation of genetic maps combined to QTL meta-analysis has proven to be a powerful approach for identification of candidate genes underlying quantitative traits. BioMercator is the first software covering all steps required to perform QTL meta-analysis and mine genome functional annotation related to QTL and meta-QTL. BioMercator V4 is freely available: http://moulon.inra.fr/biomercator

Presentation Flier: www.eccb14.org/programme/id_track/ID10-summary.pdf

Presenters

Yannick de Oliveira ; Olivier Sosnowski ; Alain Charcosset ; Johann Joets [joets@moulon.inra.fr]

INRA, UMR 0320 / UMR 8120 Génétique Végétale, F-91190 Gif-sur-Yvette, France

External Link: http://moulon.inra.fr/

ID11: iPlant Collaborative™ A Scalable Cyberinfrastructure for Life Science

The iPlant Collaborative develops cyberinfrastructure to solve data-intensive problems of biology - those involving genome, phenotype, and environmental data. iPlant provides free access to cyberinfrastructure through web-based platforms and services, many of which are accessible to biologists without extensive computational backgrounds. This demo surveys these tools and their applications.

Presentation Flier: www.eccb14.org/programme/id_track/ID11-summary.pdf

Presenter

Jason Williams [williams@cshl.edu]

Cold Spring Harbor Laboratory, 1 Bungtown Rd. Cold Spring Harbor, NY 11724, USA

External Link: http://www.iplantcollaborative.org

ID12: Garuda: Fly to the Future of Biology

With the ever-increasing diversity of omics-scale experimental data, a key challenge is the ability to discover the right tools for a specific analysis and navigate through their specific formats. Garuda is an open, community-driven, platform that provides a framework to discover, connect & navigate through different applications in bio-medical research.

Presentation Flier: http://www.eccb14.org/programme/id_track/ID12-summary.pdf

Presenter(s)

Samik Ghosh [ghosh@sbi.jp] ; Yukiko Matsuoka ; Hiroaki Kitano

The Systems Biology Institute, Tokyo, Japan

External Link: http://www.garuda-alliance.org

ID13: TotaLinuX

The description of the demo was not received before printing the program book. Please check the ECCB web site.

Presentation Flier: http://www.eccb14.org/programme/id_track/ID13-summary.pdf

Presenter

Jacques Pasteau

TotaLinuX, 2/4 rue Jean Baptiste Huet, 78350 Jouy-en-Josas, France

External Link: http://www.totalinux.fr

TotaLinuX Exhibition Booth N°11

ID14: Sushi: An Exquisite Recipe for Fully Documented, Reproducible and Reusable NGS Data Analysis

We present Sushi, an agile framework for web- and commandline-based data analysis that lets users build fully reproducible analysis workflows. All results and associated meta-information are fully defined on the file system in tabular clear text format that is independent of Sushi and ready for sharing and distribution.

Presentation Flier: www.eccb14.org/programme/id_track/ID14-summary.pdf

Presenters

Masomi Hatakeyama [masaomi.hatakeyama@ieu.uzh.ch] ; Hubert Rehrauer [Hubert.Rehrauer@fgcz.ethz.ch]

Functional Genomics Center, University of Zurich, Switzerland

External Link: http://www.fgcz.ch/

ID15: JAMM: A Peak Finder for Joint Analysis of NGS Replicates

We introduce JAMM: a peak finder that can integrate biological replicates and determine enrichment site widths accurately. JAMM is a universal peak finder that is applicable to different types of datasets. It is available for free and can run on Linux machines through the command line: http://code.google.com/p/jammpeak-finder

Presentation Flier: www.eccb14.org/programme/id_track/ID15-summary.pdf

Presenters

Mahmoud Ibrahim [mahmoud.ibrahim@mdc-berlin.de] ; Scott A. Lacadie ; Uwe Ohhler [uwe.ohler@mdc-berlin.de]

Max-Delbrück Center for Molecular Medicine Berlin-Buch, Robert-Rössle-str. 10, Berlin 13125, Germany

External Link: https://www.mdc-berlin.de/

ID16: POPS: Predicting and Enhancing Protein Solubility

Protein solubility is a prerequisite for many biophysical and biochemical applications. However, it remains challenging to produce soluble proteins by using heterologous expression systems. In this regard, computational approaches to correctly predict the solubility level of a protein and enhancing the solubility level by modifying the protein sequences are highly valuable.

Presentation Flier: www.eccb14.org/programme/id_track/ID16-summary.pdf

Presenters

Yi-An Tung [callsobing@gmail.com] ; Chien-Yu Chen [cychen@mars.csie.ntu.edu.tw]

National Taiwan University and Academia Sinica, Taipei 106, Taiwan

External Link: http://c4lab.bime.ntu.edu.tw/pops/service.html

ID17: Computational Tools for the Taxonomic Analysis of Shotgun Metagenome Samples

Metagenomics characterizes microbial communities by shotgun sequencing of environmental DNA. We present software (taxator-tk, PhyloPythiaS+) to facilitate the taxonomic analysis of large NGS datasets. The taxonomic assignment of individual (sub)sequences allows estimating the abundances of community members and to reconstruct taxonomic bins with sequence data for the individual taxa.

Presentation Flier: www.eccb14.org/programme/id_track/ID17-summary.pdf

Presenter

Johannes Dröge [Johannes.Droege@uni-duesseldorf.de]

Heinrich-Heine-Universität Düsseldorf, Germany

External Link: http://algbio.cs.uni-duesseldorf.de/software/

ID18: Goes Open: publish with Genomics, Proteomics & Bioinformatics

Genomics, Proteomics and Bioinformatics (GPB) is a peer-reviewed and fast-track open access journal that focuses on disseminating the newest discoveries in the -omics and bioinformatics studies. A thorough introduction on the GPB would facilitate your publishing with this fast-processing journal and help increasing the impacts of your study.

Presentation Flier: www.eccb14.org/programme/id_track/ID18-summary.pdf

Presenter

Andreas Keller [andreas.keller@ccb.uni-saarland.de]

Chair for Clinical Bioinformatics, Saarland University, Saarbrücken, Germany

External Link: http://www.journals.elsevier.com/genomics-proteomics-and-bioinformatics/

Posters

A. Sequencing and sequence analysis for genomics

A01: Pedro Oliveira, Marie Touchon and Eduardo Rocha. Genetic Mobility and the Distribution of Restriction Modification Systems in Prokaryotes

A02: Vivek Srinivas, Swati Patankar and Santosh Noronha. Translatability can predict functional small open reading frames in intergenic regions of the Plasmodium falciparum genome.

A03: Amin Ardeshirdavani, Erika Souche, Luc Dehasbe, Jeroen Van Houdt, Joris Vermeesch and Yves Moreau. NGS logistics: federated analysis of ngs sequence variants across multiple locations

A04: Susan Jones, Linda Milne and Glenn Bryan. Genetic marker discovery in potato: using RNA-Seq for single nucleotide polymorphism discovery in wild Solanum species

A05: Javier Perez-Florido, F.Javier Lopez-Domingo, Antonio Rueda, Joaquin Dopazo and Javier Santoyo-Lopez. ngsCAT: an easy-to-use tool to evaluate the efficiency of targeted enrichment sequencing

A06: Arnaud Kerhornou, Dan Bolser and Paul Kersey. The polyploid bread wheat genome in Ensembl Plants

A07: Jasmin Schlotthauer, Agnes Hotz-Wagenblatt, Karl-Heinz Glatting, Sabine Weiss and Annette Altmann. Two pipelines for the identification of tumor specific peptides from Next Generation Sequencing data originating from display analysis

A08: Laurent Jourdren, Sandrine Perrin, Sophie Lemoine and Stéphane Le Crom. Aozan: an automated post sequencing data processing pipeline

A09: Lin Wang, Bingding Huang, Agnes Hotz-Wagenblatt, Karl-Heinz Glatting and Christopher Previti. Theoretical and practical coverage achieved by targeted exome enrichment

A10: Kévin Vervier, Jean-Philippe Vert, Maud Tournoud, Jean-Baptiste Veyrieras and Pierre Mahé. Towards Large-scale Machine Learning for Metagenomics Sequence Classification

A11: Aurelia Caputo, Gregory Dubourg and Olivier Croce. A different approach to assemble a whole genome directly from human stool

A12: Michał Kierzynka, Wojciech Frohmberg, Piotr Żurkowski and Jacek Błażewicz. A novel approach to the whole genome assembly problem

A13: Dominik Forster, Lucie Bittner, Slim Karkar, Micah Dunthorn, Sarah Romac, Stéphane Audic, Philippe Lopez, Eric Bapteste and Thorsten Stoeck. Testing ecological theories with sequence similarity networks

A14: Mathieu Labernardiere, Patrice Baa-Puyoulet, Jean-Pierre Gauthier, Gérard Febvay, Federica Calevro, Yvan Rahbé, Hubert Charles, Jean-Christophe Simon and Stefano Colella. SymbAphidBase: a new database dedicated to aphid symbionts to store novel sequenced genomes and standardize their annotations.

A15: Nik Shazana Nik Mohd Sanusi, Rozana Rosli, Chan Kuang Lim, Low Eng-Ti Leslie, Meilina Ong Abdullah, Rajinder Singh and Ravigadevi Sambanthamurthi. SNPs Discovery from RNA-seq Data of dura, pisifera and tenera Fruit Forms of Oil Palm

A16: Rozana Rosli, Chan Kuang Lim, Low Eng Ti Leslie, Meilina Ong-Abdullah, Rajinder Singh and Ravigadevi Sambanthamurthi. Comparative genomics in oil palm

A17: Electra Tapanari, Julien Lagarde, Javier Santoyo, Laurens Wilming, Jose Manuel Gonzalez, Barbara Uszczynska, Anne-Maud Ferreira, Alexandre Reymond, Roderic Guigo and Jennifer Harrow. Examining tissue specific RACEseq extension of lncRNAs in the Human GENCODE gene set

A18: Jérôme Compain, Renaud Jullien, Sivasangari Nandy, Olivier Collin, Jean-Francois Gibrat, Valentin Loux, Véronique Martin and Sophie Schbath. Mapdecode : inventory and benchmark of read mapping tools

A19: Aurélien Bernard, Nicolas Guilhot, Fréderic Choulet, Etienne Paux and Philippe Leroy. The TriAnnot pipeline and its application to the wheat chromosome 3B annotation

A20: Chloé Cabot, Mélissa Mary, Chadi Saad, Alexandre Renaux, Alexis Bertrand, Amandine Velt, Arnaud Lefebvre, Caroline Bérard, Nicolas Vergne and Hélène Dauchel. GC- VC/DGE: a user-friendly web application for Going over Concordance across results from NGS bioinformatics analytic pipelines

A21: Samia Benamar, Morgan Gaia and Olivier Croce. Genome analysis of a new virophage to highlight the specificity with its host

A22: Paul Bailey, Sarah Ayling, Cristobal Uauy, Ksenia Krasileva, Hans Vasquez-Gross and Jorge Dubcovsky. Development Of An Exome Capture Resource For Functional Genomics In Bread Wheat

A23: Vincent Walter, Julie Thompson, Olivier Poch and Hoan Nguyen. NeoPipe : A workflow for protein family analysis

A24: Koen Illeghems, Luc De Vuyst and Stefan Weckx. Unravelling ecosystem composition and functional potential of the microbial metagenome involved in the cocoa bean fermentation process

A25: Gabriel Renaud, Udo Stenzel, Tomislav Maricic, Victor Wiebe and Janet Kelso. deML: Likelihood-based approach for robust demultiplexing of next-generation sequencing data

A26: Anna Ershova, Ivan Rusinov, Anna Karyagina, Sergei Spirin and Andrei Alexeevski. Underrepresented Words in Prokaryotic Genome Shed Light on Lifespan of Restriction-Modification Systems in Genome

A27: Rian Pierneef, Oliver Bezuidt and Oleg Reva. Getting insight into ontological relations between mobile genetic elements in bacterial genomes using SeqWord LingvoCom tools

A28: Julien Pelé, Matthieu Moreau, Hervé Abdi, Patrice Rodien, Hélène Castel and Marie Chabbert. Evolutionary hubs in protein families: A comparative analysis of sequence covariation methods

A29: Anne-Laure Abraham, Mathieu Almeida, Nicolas Pons, Charlie Pauvert, Sophie Schbath and Pierre Renault. A shotgun metagenomic method to characterise low abundant species and assign precisely taxonomy in complex microbial ecosystems

A30: Heiner Klingenberg and Peter Meinicke. Tools for fast and accurate metatranscriptome analysis

A31: Rita Pancsa, Mauricio Macossay-Castillo, Simone Kosol and Peter Tompa. Structural and functional implications of stop codon readthrough in an evolutionary context

A32: Valentina Boeva, Tatiana Popova, Maxime Lienard, Sebastien Toffoli, Maud Kamal, Christophe Le Tourneau, David Gentien, Nicolas Servant, Pierre Gestraud, Thomas Rio Frio, Philippe Hupé, Emmanuel Barillot and Jean-François Laes. OncoCNV: a multi-factor data normalization method for the detection of copy number aberrations in amplicon sequencing data

A34: Sebastien Tempel, David Servillo and Emmanuel Talla. Survey of insertion sequence domestication in prokaryotic genomes

A35: Anaïs Gouin, Anthony Bretaudeau, Claire Lemaitre and Fabrice Legeai. Identification and correction of genome mis-assemblies due to heterozygosity

A36: Amin Zarif Saffari, Marie-Theres Gansauge, Matthias Meyer, Svante Pääbo, Janet Kelso and Kay Prüfer. DNA damage analysis of ancient DNA

A37: Kirsley Chennen, Corinne Stoetzel, Jean Muller, Julie Thompson, Hélène Dollfus and Olivier Poch. VarScrut: a tool for deciphering new genes involved in rare diseases in the post-genomic era

A38: Jose de Vega-Bartol, Leif Skot, Matthew Hegarty, Wayne Powell, Mario Caccamo and Sarah Ayling. Annotation of the the red clover (Trifolium pratense) genome

A39: Felicia Ng, David Ruau, Evangelia Diamanti, Rebecca Hannah, Berthold Gottgens and Judith Schütte. Features of motif co-occurrence in blood cell development

A40: Aslihan Gerhold-Ay, Johanna Mazur and Harald Binder. Optimal Mapping of Methylation and RNA-Seq Data with Prediction Performance as a Measure for Optimality

A41: Patrick Durand, Erwan Drezen, Sébastien Brillet and Dominique Lavenier. KLAST: a new high-performance sequence similarity search tool

A42: Ruslan Soldatov, Svetlana Vinogradova and Andrey Mironov. Translation causes global unfolding of mRNA structures in vivo

A43: Stephen Newhouse, Amos Folarin, Hamel Patel and Richard Dobson. eaNGS (Easy Analysis of Next Generation Sequencing): a flexible, easy to use automated NGS pipeline for research and clinical laboratories

A44: Léa Siegwald, David Hot, Yves Lemoine, Hélène Touzet and Ségolène Caboche. How can you trust your metagenomic analysis pipeline ?

A45: Alexandre Loywick, Gael Even, Sophie Merlin, Renaud Blervaque and Christophe Audebert. A targeted metagenomic analysis pipeline dedicated to Ion Torrent PGM Data

A46: Martina Visnovska, Petra Hlouskova, Terezie Mandakova and Martin Lysak. Fragility of genomic block I among species of the mustard family (Brassicaceae)

A47: Boris Nagaev and Andrei Alexeevski. NPG-explorer: a new tool for nucleotide pangenome construction and analysis of closely related prokaryotic genomes

A48: Gaetan Benoit, Dominique Lavenier, Claire Lemaitre and Guillaume Rizk. Bloocoo, a memory efficient read corrector

A49: Leo Colmet Daage, Nadia Bessoltane, Virginie Bernard, Eve Lapouble, Nathalie Clement, Angela Bellini, Gaëlle Pierron, Valérie Combaret, Jean Michon, Isabelle Janoueix-Lerosey, Olivier Delattre and Gudrun Schleiermacher. Whole-Genome Sequencing Analysis of Neuroblastoma's Clonal Evolution

A50: Guillaume Rizk, Anaïs Gouin, Rayan Chikhi and Claire Lemaitre. MindTheGap : integrated detection and assembly of short and long insertions

A51: Nadia Bessoltane Bentahar, Virginie Bernard and Olivier Delattre. Pubmed Based Gene Annotation Mining, tool helps clinicians and biologists to highlight candidate genes of interest

A52: Sven Schuierer and Guglielmo Roma. The exon quantification pipeline (EQP)

A53: Claire Kuchly, Gerald Salin, Gaelle Vilchez, Jerome Mariette and Frederic Escudie. NGS Goes Automatic : From library preparation to quality control of data

A54: Kevin Lebrigand. De Novo Genome Assembly of Drechmeria Coniospora using ION Torrent, SOLiD and Optical Mapping data.

A55: Katarina Matthes and Mark D. Robinson. A comparison of count-based and assembly-based methods for differential splice detection

A56: Susete Alves Carvalho, Raluca Uricaru, Jorge Duarte, Claire Lemaitre, Nathalie Rivière, Gilles Boutet, Alain Baranger and Pierre Peterlongo. Reference-free high-throughput SNP detection in pea: an example of discoSnp usage for a non-model complex genome.

A57: Bettina Halwachs, Henrik R. Nilsson, Kessy Abarenkov and Gerhard G. Thallinger. Integration and validation of resources for high-throughput classification of fungal communities

A58: Sergei Lebedev and Oleg Shpynov. A switching hidden markov model for bisulfite sequencing

A59: Pierre-Marie Chiaroni, Denis Thieffry and Morgane Thomas-Chollier. Prediction of transcription factor motifs and binding sites from multiple histone mark ChIP-seq datasets

A60: Svetlana Vinogradova and Andrew Mironov. New structure-based RNA alignment method

A61: Adam Clooney, Marcus Claesson, Roy Sleator and Aisling O'Driscoll. High-resolution of microbiota, inflammation and diet in Inflammatory Bowel Disease using parallelised big data processing and analytics

A62: Evgenii Kurbatckii. Detecting differential histone modification sites from ChIP-seq data via ranking

A63: Imene Boudellioua and Victor Soloviev. Investigation of Prokaryotic Gene Regulation in Distant Phylogenetic Groups

A64: Takeru Nakazato, Tazro Ohta and Hidemasa Bono. DBCLS SRA: Functional mining and characterization of public NGS data

A65: Chaehwa Seo, Sangok Kim, Jieun Kim, Wan Kyu Kim and Sanghyuk Lee. Exome and Transcriptome Profiling of Lung Adenocarcinoma in Female Never-Smokers

A66: Ari Ugarte, Juliana Bernardes and Alessandra Carbone. CASH : a tool to identify domains and functionally annotate metagenomic and metatranscriptomic sequences

A67: Sabine Van Dillen, Anne-Claire Coûté-Monvoisin and Philippe Horvath. Development of a pan-genome tool and its application to comparative genomics analysis in Lactobacillus plantarum

A68: Ronald Schuyler and Simon Heath. Nucleosome-influenced DNA methylation gain and loss during differentiation

A69: Francesca Nadalin and Alberto Policriti. A new approach to RNA-Seq data analysis based on local paired reads assembly

A70: Dimitrios Zisis, Iris Hovel, Rurika Oka, Blaise Weber, Maike Stam, Jan-Jaap Wesselink and Pawel Krajewski. 4C-seq data processing, normalization and differential analysis of chromosomal contact profiles in Arabidopsis thaliana

A71: Renaud Vanhoutrève, Julie Thompson, Pierre Collet and Olivier Poch. YAMSA (Yet Another Multiple Sequence Alignment) method using Parisian Evolution on GPGPU

A72: Mi Ni Huang, John R. McPherson, Bin Tean Teh, Patrick Tan and Steven G. Rozen. Assessing Microsatellite Instability in Tumor Exome Sequences

A73: Jaime Castro-Mondragon and Jacques van Helden. Comparing, clustering and aligning transcription factor binding motifs with RSAT matrix-clustering

A74: Ching Chang, Mei-Ju May Chen, Tony Kuo, Jian-Long Huang, David S. Haymer, Ju-Chun Hsu and Chien-Yu Chen. Improving completeness of de novo transcriptome assembly and gene annotation by multi-species transcriptome sequencing in fruit fly genus Bactrocera

A75: Jocelyn De Goer De Herve, Myoung-Ah Kang, Xavier Bailly and Engelbert Mephu Nguifo. A perceptual hash algorithm for indexing and similarity search in a database of DNA sequences

A76: Joanna Sasin-Kurowska, Piotr Borsuk, Jan Gawor, Robert Gromadka, Jakub Grzesiak and Marek Zdanowski. Supraglacial community responces to increasing global temperatures as revealed by comparative metagenomics

A77: Mathias Vandenbogaert, Anne-Sophie Delannoy-Vieillard, Laure Diancourt, Aurélia Kwasiborski, Jean-Michel Thiberge and Valérie Caro. Alignment-free sequence composition analysis of High-Throughput Sequencing runs, for pre-assembly read clustering

A78: Ivan Kel, Christoph Dieterich, Zisong Chang, Luciano Milanesi and Ivan Merelli. Applying eQTL analysis to RNA-Seq data to study genetic regulation of miRNA expression in C.elegans

A79: Sangok Kim, Yukyung Jun, Charny Park, Pora Kim, Jieun Kim, Chaehwa Seo, Kyoohyoung Rho, Jong-Eun Lee, Wan Kyu Kim, Harkyun Kim and Sanghyuk Lee. Multi-dimensional Genomic Study of Early-Onset Gastric Cancer

A80: Margus Lukk. Disk-free computing framework for NGS Big Data – how fast can we compute?

A81: Andre Kahles, Cheng Soon Ong and Gunnar Rätsch. SplAdder: Integrated Quantification, Visualization and Differential Analysis of Alternative Splicing

A82: Aleksander Jankowski, Jerzy Tiuryn and Shyam Prabhakar. MOCCA: accurate identification of transcription factor footprints by modeling DNase I cut profiles

A83: Tahila Andrighetti, Gunther Johannes Lewczuk Gerhardt, Agnes Alessandra Sekijima Takeda, Ney Lemke and Jose Luiz Rybarczyk-Filho. Identification of metagenomic data by genome signatures and n-entropy analysis

A84: Séverine Gagnot, Mireille Ansaldi and Emmanuel Talla. Identification and characterization of small viral proteins in bacteriophage and prokaryotic genomes.

A85: Thomas Abeel, Bruce Walker, Alex Salazar, Terrance Shea, Chris Desjardins, Jennifer Wortman, Sarah Young and Ashlee Earl. Identifying large sequence variants in collections of bacterial genomes with Pilon and Emu

A86: Jakob Hull Havgaard, Kortine Kleinheinz, Sachin Pundhir and Jan Gorodkin. Combining read profile alignment with structural alignment to predict structured RNAs in transcriptomic data

A87: Sepideh Mazrouee and Wei Wang. Single Individual Haplotyping - third generation sequencing

A88: Anaïs Vittu, John Randy Clayton and Stéphanie Blandin. A pipeline for the identification of contaminant microorganisms in high-throughput RNA/DNA sequencing data

A89: Aleksandra Pfeifer, Barbara Jarzab and Joanna Polanska. Algorithms for fusion transcripts detection in RNA-seq data - comparison and improvement

A90: Gift Nuka, Matthew Fraser and Maxim Scheremetjew. InterProScan 5: Large-scale protein sequence analysis

A91: Konstantin Okonechnikov, Aki Imai-Matsushima, Lukas Paul, Alexander Seitz, Thomas F. Meyer and Fernando Garcia-Alcalde. InFusion: advancing discovery of fusion genes and chimeric transcripts from deep RNA-sequencing data

A92: Andigoni Malousi, Justine Guégan, Vincent Guillemot, Vincent Perlbarg, Arthur Tenenhaus and Ivan Moszer. Scaling Sequencing Pipelines for Whole Genomes

A93: Jacques Lagnel, Khalid Belkhir, Tereza Manousaki, Erick Desmarais, Anastasia Tsagkarakou and Alban Mancheron. New tools to optimise the analysis of a large RNA-Seq dataset from non model species: development of a hybrid assembly strategy and assessment of library complexity from raw sequencing output

A94: Martin Kircher, Daniela M. Witten, Preti Jain, Brian J. O'Roak, Gregory M. Cooper and Jay Shendure. A general framework for estimating the relative pathogenicity of human genetic variants

B. Gene expression

B01: Huy Tran, Samuel Oliveira, Olli Yli-Harja and Andre Ribeiro. Inducer intake kinetics under high extracellular concentrations in live Escherichia coli cells

B02: Olivier Rué, Philippe Bardou, Jérôme Mariette, Sarah Maman, Matthias Zytnicki and Christine Gaspin. sRNA-PlAn : a workflow for the sRNAseq data analysis

B03: Marion Verdenaud, Michel Degueldre, Sheila Zuniga, Juan Carlos Triviño, Loïc Quinton, Edwin De Pauw, Pierre Escoubas and Frédéric Ducancel. VENOMICS : High-throughput peptidomics and transcriptomics of animal venoms for discovery of novel therapeutic peptides

B04: Yoichi Takenaka, Shigeto Seno and Hideo Matsuda. Chronological analysis of regulatory strength on gene regulatory networks

B05: Liberata Mwita and Oleg Reva. Comparison of gene expression profiles of two plant growth promoting strains Bacillus atrophaeus UCMB-5137 and Bacillus amyloliquefaciens FZB42 stimulated by maize root exudate

B06: Kiyohiko Sakamoto and Y-H. Taguchi.Subtype specific promoter methylation in glioblastoma

B07: Mireya Plass, Simon H. Rasmussen, Lykke Pedersen and Anders Krogh. Computational analysis of RNA binding proteins in miRNA-mediated downregulation

B08: Julien Roux, Irene Hernando-Herraez, Claudia Chavarria, Amy Mitrano, Jonathan Pritchard, Tomas Marques-Bonet and Yoav Gilad. A genomic study of the contribution of DNA methylation to regulatory evolution in primates

B09: Rim Zaag, Guillem Rigaill, Jean-Philippe Tamby, Véronique Brunaud, Zakia Tariq, Sébastien Aubourg, Etienne Delannoy and Marie-Laure Martin-Magniette. Global Analysis of coRegulation for the identification of functional modules

B10: Valentin Voillet, Magali San Cristobal, Pascal G.P. Martin, Yannick Lippi, Louis Lefaucheur and Laurence Liaubet. Integrative approach to define biomarkers of piglet maturity

B11: Nicola Voyle, Aoife Keohane, Stephen Newhouse, Katie Lunnon, Andy Simmons, Eric Westman, Hilkka Soininen, Iwona Kloszewska, Patrizia Meccoci, Magda Tsolaki, Bruno Vellas, Simon Lovestone, Angela Hodges,

Richard Dobson and Steven Kiddle. Blood based gene expression markers of Alzheimer's Disease diagnosis: a pathway based approach.

B12: Bart Cuypers, Manu Vanaerschot, Maya Berg, Pieter Meysman and Kris Laukens. Understanding Leishmania development and adaptation by using a systems biology approach

B13: Pieter Meysman, Ehsan Sabaghian, Riet De Smet, Kristof Engelen, Yves Van de Peer, Yvan Saeys and Kris Laukens. A detailed comparison of seven prokaryotic transcriptional regulatory networks from an evolutionary perspective

B14: Emmanuel Chaplais, Alice Talpin, Félicie Costantino, Clémence Desjardin, Nelly Bonilla, Ariane Leboime, Roula Said-Nahal, Franck Letourneur, Jacques Sébastien, Gilles Chiocchia, Maxime Breban and Henri-Jean Garchon. Multiway transcriptome analysis discriminates effects of disease and of HLA-B27 in Spondyloarthritis

B15: Alexandra Popa and Rainer Waldmann. Genome-wide characterization of translation by ribosome profiling – bioinformatics challenges

B16: Axel Rasche, Matthias Lienhard and Ralf Herwig. ARH/ARH-Seq: Discovery tool for differential splicing in High-throughput data

B17: Claudia Coronnello, Giovanni Perconti, Patrizia Rubino, Flavia Contino, Serena Bivona, Salvatore Feo and Agata Giallongo. Statistical Validation of a Comprehensive Gene/miRNA Expression Profile Dataset for miRNA:mRNA Interaction Analysis

B18: Tom Lesluyes, Gaëlle Pérot, Marine R. Largeau, Céline Brulard, Pauline Lagarde, Jean-Michel Coindre, Agnès Neuville, Carlo Lucchesi and Frédéric Chibon. From laboratory bench to patient, micro-arrays to NGS: clinical transfer of a gene expression signature.

B19: Metsada Pasmanik-Chor, Shay Ben Shachar, Henit Yanai, Liran Baram, Hofit Elad, Amos Ofer, Eli Brazowski, Noam Shomron, Hagit Tulchinsky and Iris Dotan. Gene and microRNA expression as tools to infer spectrums of Inflammatory Bowel Diseases

B20: Laurence Josset, Lisa Gralinski, Amie Eisfeld, Ralph Baric, Yoshihiro Kawaoka and Michael Katze. Analysis of a cellular gene response network to SARS-CoV and influenza A virus infection identifies specific virus-host dynamics.

B21: Mitra Barzine and Alvis Brazma. Integration of independent RNAseq datasets

B22: Emilie Chautard, Clara Benoit-Pilven, Vincent Lacroix and Didier Auboeuf. Development of a new bioinformatics pipeline to annotate, quantify and visualize alternative splicing events from human and mouse RNA-Seq data

B23: Liliana Greger and Alvis Brazma. Characterizing transcriptome diversity by RNA chimeras across human populations

B24: Julie Aubert, Christelle Hennequet-Antier, Cyprien Guérin, Delphine Labourdette, Anne de La Foye, Nathalie Marsaud, Fabrice Legeai, Frédérique Hilliou and Brigitte Schaeffer. How to design a good RNA-Seq experiment in an interdisciplinary context ?

B25: Martina Sattlecker, Hamel Patel, Susie Humby, Richard Dobson and Stephen Newhouse. Cross-disorder assessment of a diagnostic Alzheimer's disease gene expression signature

B26: Bernard Fongang and Andrzej Kudlicki. Conserved transcriptional regulatiory modules in mouse, chicken and zebrafish somitogenesis networks.

B27: Xiaobei Zhou, Helen Lindsay and Mark Robinson. Robustly detecting differential expression in RNA sequencing data using observation weights

B28: Chakravarthi Kanduri, Minna Ahvenainen, Anju K Philips, Tuire Kuusi, Harri Lähdesmäki and Irma Järvelä. Transcriptional modulation of neurotransmission by music performance

B29: Rafael Takahiro Takahiro, Pedro Rafael Costa and Ney Lemke. Transcription of the ribosomal DNA in density traffic of RNA polymerases in Escherichia coli using a stochastic model sequence-dependent

B30: Feargal Ryan, Ian Jeffrey and Marcus Claesson. Metatranscriptomics of colonic biopsies in Inflammatory Bowel Disease

B31: Francisco J Altimiras, Barbara Uszczynska, David E Loyola, Emilio Palumbo, Anna Vlasova, Robert Mj Deacon, Rodrigo A Vasquez, Roderic Guigó and Patricia Cogram. Whole Transcriptome Analysis by RNA-sequencing Reveals Novel Alzheimer's Disease Biomarkers in Natural Population of the Rodent Octodon degus

B32: Henry Han. Feature Selection for RNA-Seq Data Analysis

B34: Anita Lerch, Cristian Koepfli, Zbynek Bozdech, Ivo Mueller, Liam O'Connor and Ingrid Felger. Inferring gametocyte stage specific transcriptomes from mixed life stages of Plasmodium vivax field samples

B35: Arthur Chun-Chieh Shih, Ling Li, Ya-Ting Chang and Chien-Chang Chen. Co-expressed Regulatory Functional Modules in Pressure overload-induced Cardiac Hypertrophy

B37: Enrica Calura, Gabriele Sales, Paolo Martini and Chiara Romualdi. Wiring miRNAs to pathways: a topological approach to integrate miRNA and mRNA expression profiles

B38: Owen Dando, Peter Kind and Ian Simpson. Piquant: a pipeline for assessing the performance of transcriptome quantification tools

B39: Mei-Ju May Chen, You-Yu Lin, Wen-Hsiung Li and Chien-Yu Chen. Transcriptional Regulation of Long Non-coding Gene Expression in Drosophila melanogaster: A Genome-wide study using RNA-seq

B40: Audrey Bihouée, Erwan Delage, Sébastien Charneau, Abdelhalim Larlhimi, Audrey Donnart, Damien Eveillard, Jérémie Bourdon, Pierre Lindenbaum, Gilles Toumaniantz, Flavien Charpentier, Richard Redon and Géraldine Jean. A combined approach to identify therapeutic targets involved in Progressive Cardiac Conduction Defect

B41: Sepideh Babaei, Ahmed Mahfouz, Boudewijn P.F. Lelieveldt, Marcel Reinders and Jeroen De Ridder. Multi-scale chromatin interactions are predictive for spatial co-expression patterns in the mouse cortex

B42: Frédéric Fer, Julie Aubert and Jean-Marie Beckerich. Combining kinetic modeling and transcriptomic analysis to study the resilience on a model of microbial ecosystem

B43: Philipp Senger and Shweta Bagewadi. Automatic Quality Assessment Of Microarray Datasets Using Ensemble Methods

B44: Marcelo Segura, Hector Keun and Tim Ebbels. Pathway based models have similar predictivity and robustness to models based on random collections of genes

B45: Candida Vaz, Choon Wei Wee, Gek Ping Serene Lee, Vivek M Tanavde and Sinnakarupan Mathavan. Next generation sequencing reveals tissue and sex specific known and novel miRNAs in zebrafish

B46: Luiz Augusto Bovolenta, Danillo Pinhal, Simon Moxon, Arthur Casulli Oliveira, Pedro Gabriel Nachtigall, Marcio Luis Acencio, Cesar Martins and Ney Lemke. Nile tilapia miRNAs: characterization and target prediction

B47: Sandra Koser, Jan-Philipp Mallm, Sabrina Schumacher, Stephan Wolf, Stephan Stilgenbauer, Karsten Rippe, Daniel Mertens and Benedikt Brors. Differential expression analysis of microRNA in CLL and their influence on mRNA expression

B48: Chee Lee, Yanxiao Zhang and Maureen A. Sartor. RNA-Enrich: A cut-off free gene set enrichment testing method for RNA-seq that adjusts for gene read count

B49: Marta Rosikiewicz and Marc Robinson-Rechavi. Benchmarking of quality control parameters for RNA-seq data

B50: José Luis Gaete and Marta Fernandez. RNA-Seq applications for plant transcriptome analysis in response to cold acclimation by Ion Torrent technology

C. Pathways and molecular networks

C01: Inna Kuperstein, Simon Fourquet, Jean-Marie Ravel, Emmanuel Barillot and Andrei Zinovyev. A comprehensive map of programmed cell death signalling network: an analytical tool for studying regulation of different modes of cell death in human disorders

C02: Yvonne Mayer, Karin Zimmermann, Berit Haldemann, Dido Lenze, Michael Hummels and Ulf Leser. Inclusion of miRNAs improves differential network analysis

C03: Antonio Fabregat, Joel Weiser, Steven Jupe, Phani Garapati, Oscar Forner, Pablo Porras, Robin Haw, Peter D'Eustachio, Lincoln Stein and Henning Hermjakob. Reactome: Pathway Analysis Tool Suite

C04: Jeanne Cambefort, Guillaume Collet, Sylvain Prigent, Simon Dittami, Olivier Dameron, Thierry Tonon and Anne Siegel. AuReMe: an integrative method for Reconstruction of Metabolic networks including Automatic Gap-Filling

C05: Michaela Bayerlova, Florian Klemm, Annalen Bleckmann, Frank Kramer, Tobias Pukrop and Tim Beissbarth. Integration of breast cancer RNA-Seq data into newly constructed WNT signaling networks

C07: Sylvain Bournais, Pauline Gloaguen, Gilles Curien, Christophe Bruley, Florence Combes, Marianne Tardif, Yves Vandenbrouck, Giovanni Finazzi, Myriam Ferro and Norbert Roland. Towards the virtual chloroplast

C08: Patrice Baa-Puyoulet, Augusto F. Vellozo, Jaime Huerta-Cepas, Gérard Gérard Febvay, Federica Federica Calevro, Marie-France Sagot, Hubert Charles, Toni Gabaldon and Stefano Colella. Annotating arthropods genome to study and compare their metabolism: the ArthropodaCyc collection of Cyc databases powered by CycADS

C09: Fazle Faisal and Tijana Milenkovic. Dynamic networks reveal key players in aging

C10: Kaveh Pouran Yousef, Adam Streck, Heike Siebert and Max von Kleist. Analysing the c-di-GMP-dependent bistable regulation of curli fibers in Escherichia coli by combining boolean, continuous and stochastic modelling

C11: Sun Sook Chung, Alessandro Pandini, Alessia Annibale, Anthony C. C. Coolen, Nicolas Shaun B. Thomas and Franca Fraternali. Topological analysis of protein interaction networks: the importance of loop motifs and their biological implications

C13: Berit Haldemann, Daniel Heinze, Michael Hinz, Claus Scheidereit and Ulf Leser. A Comprehensive Approach to the Characterization of Transcriptional Regulation of DNA Damage Responses

C14: Rafael S. Costa, Nguyen Hoang Son and Susana Vinga. Comparison of cellular objectives in flux balance constraint-based models

C15: David Hill, Tanya Berardini, Peter D'Eustachio, Harold Drabkin, Chris Mungall and Judith A. Blake. Modeling Glycolysis in The Gene Ontology: All Roads Lead to Pyruvate

C16: Eva Strakova, Jan Bobek, Alice Zikova, Klara Novotna and Jiri Vohradsky. Inference of sigma factor controlled networks in germinating prokaryote

C17: Stefan Kroeger, Melanie Venzke, Ria Baumgrass and Ulf Leser. Meta expression analysis of regulatory T cell experiments for gene regulatory network reconstruction

C18: Azzurra Carlon, Barbara Di Camillo, Federica Eduati and Gianna Toffolo. A rule based modeling approach to insulin signaling pathway analysis: signal amplification and robustness

C19: Dmitry Ravcheev and Ines Thiele. Genomics analysis of the respiratory capacities of the human intestinal microbiota

C20: Patrick Trampert, Tim Kehl, Daniel Stöckel, Hans-Peter Lenhof, Andreas Keller and Christina Backes. GeneTrail2 - A statistical analysis tool for molecular signatures

C21: Jeanne Marie Onana Eloundou-Mbebi, Sabrina Kleesseen, Michaël Méret, Thomas Degenkolbe, Lothar Willmitzer and Zoran Nikoloski. Reconstruction of substrate complexes in biochemical networks from time-resolved relative compound levels

C22: Claire Rioualen, Quentin Da Costa, Guillaume Pinna, Annick Harel-Bellan, Emmanuelle Charafe-Jauffret, Christophe Ginestier and Ghislain Bidaut. Interactome–regulome integrative approach for genome-wide screening data analysis in a breast cancer stem cells study

C23: Yang Xiang, Florian Martin and Joe Whittaker. A Stochastic Penalty for Incorporating Prior Information to Improve Reconstruction of Biological Networks

C24: Noel Malod-Dognin and Natasa Przulj. L-GRAAL: Lagrangian Graphlet-based Network Aligner

C26: Cedric Simillion, Heidi Lischer, Robin Liechti and Rémy Bruggmann. Avoiding the pitfalls of gene set enrichment analysis with SetRank

C27: Melanie Boerries, Hauke Busch, Jie Bao, Juliana M. Nascimento, Margareta Müller, Sofia Depner and Dennis Dauscher. Global Mean First Passage Time Analysis of Secretome to Transcriptome Signaling Reveals Endothelial-Derived TNFa and CXCL12 as Enhancers of Lung-Tumor Cell Migration

C28: Ashutosh Malhotra and Martin Hofmann-Apitius. Exploring novelty in mechanistic models for Alzheimer's disease by assessing reliability of protein interactions.

C29: Darren Davis, Omer Nebil Yaveroglu, Noel Malod-Dognin, Aleksandar Stojmirovic and Natasa Przulj. Topology-Function Conservation in Protein-Protein Interaction Networks

C31: Aristidis Vrahatis, Konstantina Dimitrakopoulou, Athanasios Tsakalidis and Anastasios Bezerianos. A time-varying method for subpathway enrichment analysis

C32: Léonard Jaillet and Stephane Redon. Characterizing reaction pathways with an energy-driven motion planning method.

C33: David Henriques, Miguel Rocha, Julio Saez-Rodriguez and Julio Banga. Reverse engineering of logic-based models using a mixed-integer dynamic optimization approach

C34: Hulda Haraldsdóttir and Ronan Fleming. Graph theoretical analysis of atom transitions in human dopamine metabolism

C35: Shriprakash Sinha. Prior Biological Knowledge And Epigenetic Information Enhances Prediction Accuracy Of Bayesian Wnt Pathway

C36: Juris Viksna, Alvis Brazma, Karlis Cerans, Dace Ruklisa and Thomas Schlitt. Hybrid systems for modeling and analysis of qualitative behaviour of gene regulatory networks

C37: Ricardo de Matos Simoes, Kate E Williamson and Frank Emmert-Streib. Comparative analysis of gene regulatory networks inferred from large-scale urothelial cancer RNAseq, Bead and Oligo gene expression data sets

D. Computational systems biology

D03: Alicia Amadoz, Patricia Sebastián-León, Francisco Salavert and Joaquín Dopazo. PATHiPRED: prediction models using the activation status of stimulus-response signaling circuits.

D04: Liliana Ironi and Diana X Tran. Model-based design of synthetic networks

D05: Isaac Crespo, Nicolas Guex, Sylvian Bron, Assia Ifticene-Treboux, Eveline Faes-Van'T Hull, Solange Kharoubi, Robin Liechti, Patricia Werffeli, Mark Ibberson, Francois Majo, Michäel Nicolas, Julien Laurent, Abhisheck Garg, Khalil Zaman, Hans-Anton Lehr, Brian J. Stevenson, Curzio Rüegg, Jean-François Delaloye, Ioannis Xenarios, George Coukos and Marie-Agnès Doucey. Angiogenic activity of breast cancer patients' monocytes reverted by combined use of systems modeling and experimental approaches

D06: Gabor Beke, Matej Stano and Lubos Klucar. Modelling the interaction between bacteriophages and bacteria

D07: Mathias Weyder, Marc Prudhomme, Patrice Polard and Gwennaele Fichant. Modeling competence regulation during bacterial transformation in S. pneumoniae

D08: Rafael Björk, Patrik Rydén and Tatjana Pavlenko. Structure learning for improved classification accuracy for high-dimensional omics data

D09: Guillaume Brysbaert, Mélany Tanchon, Ralf Blossey, Marc Aumercier and Marc Lensink. Targeting the interactions of the Ets-1 oncoprotein

D10: Ganna Androsova, Sophie Rodius, Petr Nazarov, Arnaud Muller, François Bernardin, Céline Jeanty, Simone Niclou, Laurent Vallar and Francisco Azuaje. A comprehensive integrative analysis of the transcriptional network underlying the zebrafish heart regeneration

D11: Wout Bittremieux, Dirk Valkenborg, Aida Mrzic, Hanny Willems, Bart Goethals and Kris Laukens. Pattern mining of mass spectrometry quality control data

D12: Valérie Sautron, Elena Terenina, Élodie Merlot, Pascal Martin, Yannick Lippi, Laurence Liaubet, Armelle Prunier, Pierre Mormede and Nathalie Villa-Vialaneix. Longitudinal CCA to analyze stress responses in pigs

D13: Teppo Annila, Anantha-Barathi Muthukrishnan, Abhishekh Gupta, Ramakanth Neeli Venkata and Andre Ribeiro. Properties of the spatial organization of Tsr protein clusters in live Escherichia coli cells

D14: Ralph Patrick, Kim-Anh Le Cao, Bostjan Kobe and Mikael Boden. PhosphoPICK: Probabilistic Modelling of Cellular Context for Predicting Kinase-Substrate Phosphorylation Events

D15: Alejandro F. Villaverde, Federico Morán and Julio R. Banga. Computationally efficient network inference using information theory: fMIDER

D16: Elson Tomás, Alexandra M. Carvalho, Paulo Mateus and Susana Vinga. Unsupervised classification of pharmacokinetic responses using non-linear mixed effects models

D17: Lujia Chen, Chunhui Cai, Vicky Chen and Xinghua Lu. Trans-species learning of cellular signaling systems with bimodal deep belief networks

D18: Monica Golumbeanu, Pejman Mohammadi, Celine Hernandez, Manfredo Quadroni, Amalio Telenti, Angela Ciuffi and Niko Beerenwinkel. Characterizing the dynamics of cellular response to HIV-1 infection through clustering of time-series proteomics data

D19: Arnau Montagud, Andrei Zinovyev and Emmanuel Barillot. Multiscale mathematical modelling of breast cancer invasion

D20: Joe Wandy, Rónán Daly and Simon Rogers. Incorporating peak grouping information for alignment of multiple liquid chromatography-mass spectrometry datasets

D21: Mahsa Ghanbari, Julia Lasserre and Martin Vingron. Reconstruction of gene networks using prior knowledge

D22: Samuel Collombet, Morgane Thomas-Chollier, Touati Benoukraf, Annouck Luyten, Chris Van Oevelen, Daniel G. Tenen, Thomas Graf and Denis Thieffry. Logical modelling of immune cell specification and reprogramming

D23: Aristotelis Kittas, Amelie Barozet, Jekaterina Sereshti, Niels Grabe and Sophia Tsoka. CytoASP: A Cytoscape plug-in for logical modelling of signalling networks using BioASP

D24: Sabeur Aridhi, Haitham Sghaier, Mondher Maddouri and Engelbert Mephu Nguifo. Domain knowledge-based model for phenotype prediction of ionizing-radiation-resistance in bacteria

D25: Konstantin Kozlov and Alexander Samsonov. Differential Evolution Entirely Parallel Method for Sequence-based Modeling of Gene Expression

D26: Nicole Radde, Karsten Kuritz, Caterina Thomaseth and Frank Allgöwer. The circuit-breaking algorithm for systems with order preserving flow

D27: Alain Sewer and Florian Martin. Using data-biased random walks on signed graphs to quantify perturbations in causal biological network models

D28: David Cohen, Loredana Martignetti, Emmanuel Barillot, Andrei Zinovyev and Laurence Calzone. Modelling the intracellular molecular network of tumoural invasion

D29: Andreas Troll. A new Approximation Approach for the Chemical Master Equation

D30: Anida Sarajlic, Vladimir Gligorijevic, Djordje Radak and Natasa Przulj. Network wiring of pleiotropic kinases yields insight into dissociation of diabetes and aneurysm

D32: Isa Kirk, Søren Brunak and Kirstine Belling. The correlation between the human protein interactome and conserved mammalian synteny blocks

D33: Lingjian Yang, Aristotelis Kittas, Johnathan Watkins, Anita Grigoriadis, Sophia Tsoka and Lazaros Papageorgiou. An optimisation framework inferring module activity for breast cancer classification

D34: Otoniel Rodríguez Jorge, Linda Kempis Calanis, Denis Thieffry and Angélica Santana Calderón. Logical modelling of TLR5 signals helps unravel the mechanism of neonatal CD4 T cell activation by flagellin.

D35: Djordje Djordjevic, Andrian Yang, Armella Zadoorian, Kevin Rungrugeecharoen and Joshua Ho. How difficult is inference of mammalian causal gene regulatory networks?

D36: Adrien Fauré, Barbara Vreede, Élio Sucena and Claudine Chaouiya. A Discrete Model of Drosophila Eggshell Patterning Reveals Cell-Autonomous and Juxtacrine Effects

D37: Vijayabaskar Ms, Nadine Obier, Monika Lichtinger, Debbie Goode, Michael Lie-A-Ling, Elli Marinopoulou, Josh Lilly, Constanze Bonifer, Valarie Kouskoff, Georges Lacaud, Berthold Göttgens and David Westhead. Understanding the mechanism of in vitro cellular differentiation in mouse through integrative analysis of genome-wide chromatin accessibility, chromatin modifications, transcription factor binding and gene expression data

D38: Wassim Abou-Jaoudé, Maximilien Grandclaudon, Pedro T. Monteiro, Aurélien Naldi, Claudine Chaouiya, Vassili Soumelis and Denis Thieffry. Logical modeling of T-helper cell differentiation and plasticity

D39: Emilia Wysocka, James Snowden, Matthew Page and Ian Simpson. Towards a semi-automated framework of rule-base model creation for neuropsychiatric disease.

D40: Azim Dehghani Amirabad and Marcel H. Schulz. Exploiting RNA-Seq data to the fullest: Models for miRNA-transcript target interactions

D41: Léon-Charles Tranchevent, François-Olivier Desmet, Hussein Mortada, Emilie Chautard, Marion Dubarry, Clara Benoit-Pilven and Didier Auboeuf. A computational platform to predict the functional consequences of alternative splicing variations

D42: Priscila Da Silva Figueiredo Celestino Gomes, Isaure Chauvot de Beauchene, Nicolas Panel, Pedro Geraldo Pascutti, Eric Solary and Luba Tchertanov. Impact Of Oncogenic Mutations On Allosteric Regulation Of Receptor Tyrosine Kinases: Application To The Drugs Design

D43: Georgij Arapidi, Igor Fesenko, Konstantin Babalyan, Emile Zakiev, Anna Seredina, Regina Chazigaleeva, Elena Kostrukova, Sergey Kovalchuk, Nikolay Anikanov, Tatiana Semashko, Vadim Govorun and Vadim Ivanov. Identification of small open reading frames with high coding potential in moss Physcomitrella patens

D44: Amhed Vargas-Velazquez, Pierre-Marie Chiaroni, Morgane Thomas-Chollier and Denis Thieffry. Modelling the interplay between transcriptional regulation and chromatin remodeling during cell differentiation in response to Retinoic Acid

D45: Pauline Traynard, Adrien Fauré, François Fages and Denis Thieffry. Logical modeling of the mammalian cell cycle

D46: Pauline Traynard and François Fages. A bi-directional coupled model of the cell cycle and the circadian clock

D47: Johannes Barth and Christian Fufezan. Applying novel computational tools to dissect the interwoven light and oxygen effects in the ROS stress response network of Chlamydomonas reinhardtii by enhanced quantitative mass spectrometry

D48: Shelly Mahlab and Michal Linial. miRNA-mRNA interactions: Probabilistic and dynamic perspectives

D49: Monika Kurpas, Katarzyna Jonak and Krzysztof Puszynski. The novel mathematical model of ATR-p53-Wip1 signaling pathway: studies on prediction of cellular response to DNA damages

D50: Costas Bouyioukos, Ivan Junier and François Képès. Genome REgulatory Architecture Tools (GREAT). The SCAN suite for the detection of regular patterns along genomes. GREAT:SCAN

D51: Wojciech Bensz and Krzysztof Puszynski. A stochastic model of the p53 ubiquitination system.

D52: Karsten Kuritz and Frank Allgöwer. Determining cell-cycle induced variations from snap-shot data sets

D53: Gabriella Sferra, Federica Fratini, Marta Ponzi and Elisabetta Pizzi. Dynamics of P. falciparum protein-protein interaction network: the membrane microdomain interactome.

E. Structural bioinformatics

E01: Ruben Acuna, Zoe Lacroix and Jacques Chomilier. SPROUTS 2.0: a database and workflow to predict protein stability upon point mutation

E02: Babak Sokouti, Farshad Rezvan and Siavoush Dastmalchi. Improving the “per residue” prediction accuracy of helical transmembrane segments of GPCRs using GPCRTOP v.1.0 web server

E03: Ruben Acuna, Zoe Lacroix, Jacques Chomilier and Nikolaos Papandreou. SMIR: a method to predict the residues involved in the core of a protein

E05: Mauricio Macossay Castillo, Simone Kosol, Peter Tompa and Rita Pancsa. Protein structural aspects of multifunctional gene regions

E06: Pierrick Craveur, Agnel Praveen Joseph, Pierre Poulain, Joseph Rebehmed, Sylvain Léonard, Floriane Noël, Yassine Ghouzam, Romain Deniau, Amine Ghozlane, Jérémy Esque, Guilhem Faure, Aurélie Bornot, Ramachandra Moorthy Bhaskara, Lakshmiparum S. Swapna, Swapnil Mahajan, Garima Agarwal, Vincent Jallu, Jiří Černý, Bohdan Schneider, Catherine Etchebest, Jean-Christophe Gelly, Narayanaswamy Srinivasan and Alexandre G. de Brevern. A short journey inside the protein structures at the light of a structural alphabet

E07: Ludis Morales, Janneth González, George Barreto and David Diaz. Structural and functional predictions of the hypothetical protein PA2481 in Pseudomonas Aeruginosa PAO1

E08: Stanislav Engel and Yosef Kuttner. Misfolded SOD1 noxious "gain-of-interaction" - studying the molecular mechanism of amyotrophic lateral sclerosis (ALS) pathogenesis

E09: Michelle Mukonyora. The in silico prediction of foot-and-mouth disease virus (FMDV) epitopes on the South African Territories (SAT)1, SAT2 and SAT3 serotypes

E10: Joseph Rebehmed, Patrick Revy, Guilhem Faure, Jean-Pierre de Villartay and Isabelle Callebaut. The SRI domain family: a common scaffold for RNA polymerase II CTD binding

E11: Yassine Ghouzam, Guillaume Postic, Alexandre G. de Brevern and Jean-Christophe Gelly. Improving remote protein homology detection using a structural alphabet

E12: Olga Kalinina, Jennifer Herrmann and Rolf Mueller. Novel mechanism of S. aureus RNA polymerase inhibition by disciformycins from P. fallax, discovered through structural modeling

E13: Nina M. Fischer, Marcelo D. Polêto, Anders Gärdenäs, Daniel S. D. Larsson and David van der Spoel. Analyzing RNA structures and molecular dynamics simulations

E14: Marcos Tadeu Geraldo, Agnes Alessandra Sekijima Takeda, Antônio Sérgio Kimus Braz and Ney Lemke. The basis of nuclear import by importin-alpha based on molecular dynamics simulations and normal modes analysis

E15: Edda Kloppmann, Burkhard Rost, Jonas Reeb and Michael Bernhofer. Target selection and data analysis for the New York Consortium of Membrane Protein Structure (NYCOMPS)

E16: Eugenia Polverini and Valeria Gherardi. Inside the mechanism of SMN-SmD1 protein complex formation: effects of the Spinal Muscular Atrophy - causing E134K mutation. A molecular dynamics simulation study.

E17: Nicholas Furnham, Natalie Dawson, Christine Orengo and Janet Thornton. Coupling Similarities In Enzyme Reactions With The Evolution Of Their Function Across 375 Domain Superfamilies

E18: Olga S. Voitenko and Olga V. Kalinina. Tight clusters of extremely conserved and non-conserved positions co-localize with protein-protein interaction interfaces of HIV-1 intra-virus and virus-host interactions

E19: Gulin Ozcan, Zeynep Kutlu Kabas, Onur Sercinoglu and Pemra Ozbek. Binding Behavior of HLA-B Alleles Related to Ankylosing Spondylitis (AS) Disease: A Comparative Study by Computational Methods

E20: Grzegorz Chojnowski, Tomasz Waleń, Pawel Piatkowski, Wojciech Potrzebowski and Janusz M. Bujnicki. BrickworX builds models of low resolution nucleic acid crystal structures from recurrent motifs

E21: Gwénaëlle André-Leroux, Stéphanie Petrella and Claudine Mayer. Peptidomimetics Based Inhibitor Design for Tuberculosis

E22: Gabriele Marchler, Farideh Chitsaz, Myra Derbyshire, Noreen Gonzales, Marc Gwadz, Fu Lu, James Song, Narmada Thanki, Josie Wang, Roxanne Yamashita, Chanjuan Zheng, Steve Bryant and Aron Marchler-Bauer. Annotation of Structural Motifs in the Conserved Domain Database

E23: Zeynep Kutlu Kabas, Gulin Ozcan, Onur Sercinoglu and Pemra Ozbek. Computational Study on the Effect of pH on the Binding Behaviour of HLA-B Alleles

E24: Karolis Uziela, Nanjiang Shu, Björn Wallner and Arne Elofsson. How to select the best protein model using ProQ2?

E25: Mirco Michel, Sikander Hayat, Marcin J. Skwark, Chris Sander, Debora S. Marks and Arne Elofsson. PconsFold: Improved contact predictions improve protein models

E27: Adva Yeheskel, Rony Seger and Malka Cohen-Armon. Protein Complexes Assembly and Function Revealed by Structural Motion Prediction

E28: Mateusz Banach, Elodie Duprat, Mathilde Carpentier, Barbara Kalinowska, Irena Roterman and Jacques Chomilier. Identification of the folding nucleus of globular protein: application to immunoglobulin-like and flavodoxin fold domains

E29: Marc Lensink and Shoshana Wodak. Score_set: A CAPRI Benchmark for Scoring Protein Complexes

E30: Yannick Spill and Michael Nilges. Variance and Information Content in SAS profiles

E31: Emilio Potenza, Tomas Di Domenico, Ian Walsh and Silvio Tosatto. MobiDB 2.0: an improved database of intrinsically disordered and mobile proteins.

E32: Olga Zanegina, Anna Karyagina, Andrei Alexeevski and Sergei Spirin. Structural classification of protein-DNA complexes and their families based on interacting elements.

E33: Maciej Antczak, Tomasz Zok, Martin Riedel, David Nebel, Piotr Lukasiak, Marta Szachniuk, Thomas Villmann and Jacek Blazewicz. Accurate approach for nucleotide conformation prediction of RNAs

E34: Sayoni Das, David Lee, Natalie Dawson and Christine Orengo. FunFHMMer : Exploiting CATH-Gene3D functional families to predict functions and functional sites of uncharacterised sequences

E35: Wim Vranken, Daniele Raimondi and Elisa Cilia. Applying dynamics-based interaction potentials in a residue network

E36: Elisa Cilia, Rita Pancsa, Peter Tompa, Tom Lenaerts and Wim F. Vranken. DynaMine: a web-server for predicting protein dynamics from sequence

E37: Amrita Roy Choudhury, Marjana Novič and Igor Zhukov. The transmembrane regions of Bilitranslocase

E38: Amine Ghozlane, Etienne Ruppé, Julien Tap, Nicolas Pons, Alexandre De Brevern, Joseph Rebehmed, Sean Kennedy and Stanislav Ehrlich. Pairwise Comparative Modeling For Identification Of Class A Beta-lactamases In the Human Intestinal Microbiota

E39: Tunca Dogan, Alex Bateman and Maria Martin. UniProt Domain Architecture Alignment: A New Approach for Protein Similarity Search using InterPro Domain Annotation

E40: Benjamin Bardiaux, Barth van Rossum, Olivera Francetic, Nadia Izadi-Pruneyre, Christiane Ritter, Hartmut Oschkinat and Michael Nilges. Structural modelling of symmetric protein assemblies from distance constraints.

E41: Yuezhou Zhang, Alexandre Borrel, Leslie Regad, Anne-Claude Camproux, Gustav Boije Af Gennäs, Jari Yli-Kauhaluoma and Henri Xhaard. Phosphate and ribose structural isosteric replacement in the Protein Data Bank

E42: Bjoern-O. Gohlke, Robert Preissner, Tim Overkamp, Antje Richter, Bernd Gillissen and Peter Daniel.Target landscapes identifies Vatalanib as PARP inhibitor

E43: Alexandre Borrel, Leslie Regad, Henri Xhaard, Michel Petitjean and Anne-Claude Camproux. Druggability prediction performances related to different pocket estimations

E44: Jairo Rocha, Ricardo Alberich and Emidio Capriotti. DRFLEX: An RNA Structural Classification Database with RNAFlex

E45: Pietro Lovato, Alejandro Giorgetti and Manuele Bicego. A multimodal approach to protein remote homology detection

E46: Deepti Jaiswal, Radka Svobodova Varekova, David Sehnal, Crina-Maria Ionescu, Stanislav Geidl, Lukas Pravda, Vladimır Horsky, Michaela Wimmerova and Jaroslav Koca. Consistency of sugar structures and their annotation in the PDB

E47: Nicolas Denis and David Ritchie. Fine-Grained Structure-Function Clustering of Pfam Protein Domain Families: A Case Study Using CYP450

E48: Jad Abbass and Jean-Christophe Nebel. Customised fragment libraries for ab initio protein structure prediction: usage of functional and structural annotations

E49: Irena Roterman, Mateusz Banach, Barbara Kalinowska and Leszek Konieczny. Similar Structure – Different Stabilization – Analysis Of Immunoglobulin-Like Domains

E50: Aram Gyulkhandanyan. Binding of cationic porphyrins to heme proteins

E51: Bedrat Amina, Amrane Samir, Guédin Aurore and Mergny Jean-Louis. Algorithm to predict G-quadruplex folding through score computing

E52: Luigi D'Ascenzo and Pascal Auffinger. Electrostatic potential dissimilarities between aromatic amino acids and nucleobases lead to anion-π stacking events in nucleic acids

E54: Lukáš Pravda, Radka Svobodová Vařeková, David Sehnal, Crina-Maria Ionescu, Karel Berka, Michal Otyepka and Jaroslav Koča. Anatomy of enzymatic channels and algorithm for its detection

E55: Evgeniy Aksianov and Andrey Alexeevsky. ProtOn: a tool for automatic annotation of beta-structures

E56: Yasaman Karami, Serge Amselem, Elodie Laine and Alessandra Carbone. Disease-related mutations in proteins: a study of dynamically correlated networks and coevolved residue clusters

E57: Nidhi Tyagi, Edward Farnell, Colin Fitzsimmons, Stephanie Ryan, Rick Maizels, David Dunne, Janet Thornton and Nicholas Furnham. Allergenic Proteins Are Targets For Mammalian IgE Mediated Immune Response Against Metazoan Parasites: Linking allergy to immunity against metazoan parasites

E58: Géraldine Caumes, Hiba Abi Hussein, Jean-Baptiste Chéron, Anne-Claude Camproux and Leslie Regad. Effects of the pocket estimation algorithms in one pocket-ligand complex classification

E59: Inès Rasolohery, Imen Daoud, Pierre Tufféry, Gautier Moroy and Frédéric Guyon. PatchSearch: a new method for surface patch comparison in proteins

E60: Sergei Grudinin and Georgy Derevyanko. HermiteFit: Fast fitting atomic structures into a low-resolution density map using 3D orthogonal Hermite functions

E61: Iain H. Moal and Juan Fernández-Recio. Training energy functions from changes in protein-protein binding affinity upon mutation

E62: Isidro Cortes, Guillaume Bouvier, Michael Nilges, Luca Maragliano and Therese Malliavin. Enhanced conformational sampling of the catalytic domain of the adenyl cyclase CyaA from Bordetella pertussis

E63: Thomas Coudrat, John Simms, Denise Wootten, Arthur Christopoulos and Patrick Sexton. Computer-aided drug discovery: development of a method for G protein-coupled receptor binding pocket refinement

E64: Kliment Olechnovic and Ceslovas Venclovas. The CAD-score webserver: contact area-based comparison of structures and interfaces of proteins, nucleic acids and their complexes

E65: Nadia Znassi and Andrey V. Kajava. Mapping amyloidogenicity to the known 3D structures of proteins

E66: Claudia Caudai, Emanuele Salerno, Monica Zoppè and Anna Tonazzini. A multiscale model for 3D chromatin structure estimation using quaternions

E67: Justas Dapkunas, Albertas Timinskas, Kliment Olechnovic, Mindaugas Margelevicius, Rytis Diciunas and Ceslovas Venclovas. The PPI3D web server for searching, analyzing and modeling protein-protein interactions in the context of 3D structures

E68: Martino Bertoni, Marco Biasini, Florian Kiefer and Torsten Schwede. Comparative protein quaternary structure modelling using evolutionary interaction fingerprints

E69: Nolan Chatron, Florent Langenfeld, Virginie Lattard and Luba Tchertanov. In silico study of Vitamin K epoxide reductase and its mutants

E70: Jessica Andreani and Johannes Soeding. Prediction of beta-strand interactions from direct coupling patterns

E71: Kęstutis Timinskas and Česlovas Venclovas. Computational Analysis of DNA Polymerases and their Homologs in Bacterial Genomes

E72: You-Yu Lin, Mei-Ju May Chen and Chien-Yu Chen. A study of inter- and intra-protein correlated mutations on highly similar protein sequences

E73: Sucharita Dey and Bin Tean Teh. Molecular dynamic simulation reveals altered function of a chromatin regulatory protein involved in myeloid malignance due to single missense mutation

E74: Alexander Monzon, Emidio Capriotti and Gustavo Parisi. Conformational diversity of protein functional regions improves the characterization of deleterious mutations

E75: Tomas Bastys, Nadezhda T. Doncheva, Hauke Walter, Rolf Kaiser, Mario Albrecht and Olga V. Kalinina. Molecular dynamics simulations reveal a potential reason for selecting of distinct mutation combinations in HIV-1 protease clinical isolates: a study of several clones with and without L76V resistance mutation

E76: Dominic Simm, Klas Hatje and Martin Kollmar. Waggawagga: a Web Service for the Comparative Visualization of Coiled-Coil Predictions and the Detection of Charged Single-Alpha-Helices (CSAHs)

E77: Akito Taneda. A multi-objective genetic algorithm for multi-target RNA design

E78: Lukasz P. Kozlowski and Janusz M. Bujnicki. Identification of potential protelomerases and their target sites in publicly available genomes

E79: Marco Pietrosanto, Eugenio Mattei, Manuela Helmer-Citterich and Fabrizio Ferré. BEAM: A new method for RNA secondary structure motifs discovery

E80: Arumay Pal and Chandra S. Verma. Investigating molecular mechanism and dynamics of ErbB family ligand binding by molecular dynamics simulation

E81: Nadezhda T. Doncheva, John H. Morris, Olga Voitenko, Tomas Bastys, Karsten Klein, Eric F. Pettersen, Dina Schneidman, Andrej Sali, Thomas E. Ferrin, Olga V. Kalinina and Mario Albrecht. Analyzing the dynamic nature of proteins using residue interaction networks

E82: Peter Cimermancic, Patrick Weinkam, Justin Retternmaier, Daniel A. Keedy, Rahel Woldeyes, James A. Wells, James S. Fraser and Andrej Sali. Expanding the druggable proteome by characterization and prediction of cryptic binding sites

E83: Pedro Rafael Costa and Ney Lemke. A heuristic approach to study the influence of transcription pausing on RNA folding

E84: Aya Narunsky, Haim Ashkenazy, Rachel Kolodny and Nir Ben-Tal. Using PDB to explore conformational space of query proteins with at least one known conformation

E85: Katerina Taškova, Marco Carnini, Sonika Rao and Andreas Hildebrandt. Exploring the space of co-optimal alignments for template-based protein model quality assessment

E86: Yves Dehouck and Alexander S. Mikhailov. A sequence-specific elastic network model for coarse-grained studies of protein dynamics

E87: Marco Pasi, John Maddocks, Richard Lavery and The Ascona B-Dna Consortium. Microsecond-scale sequence-dependence of B-DNA mechanics and cation binding.

E88: Juergen Haas, Alessandro Barbato, Steven Roth, Tobias Schmidt, Konstantin Arnold, Khaled Mostaguir, Lorenza Bordoli and Torsten Schwede. The ProteinModelPortal - How good is my modeling ? First Results From The Continuous Automated Model EvaluatiOn (CAMEO)

E89: Qingzhen Hou, Jaap Heringa and K. Anton Feenstra. Differential Conservation Between Interacting and Non-interacting Homologs Identifies Interface Residues

E90: Qingzhen Hou, Kamil Krystian Belau, Marc F. Lensink and K. Anton Feenstra. Mapping the Protein-protein Interaction Free Energy Landscape

E91: K. Anton Feenstra, Tom L.G.M. van den Kerkhof and Esther F. Gijsbers. Application Showcases for Sequence Harmony: Specificity Detection in HIV Protein Sequences

E92: Irene Farabella, Daven Vasishtan, Agnel-Praveen Joseph, Arun Prasad Pandurangan and Maya Topf. Validation of 3D Electron Microscopy Density Fits assembly models using TEMPY

F. Evolution and population genomics

F01: Xiaoyu Yu and Oleg Reva. Mathematical modeling of the genetic amelioration of horizontally transferred genomic islands in bacterial genomes

F02: Kay Prüfer, Janet Kelso and Svante Pääbo. Searching for Regions of Superarchaic Introgression in the Denisovan Genome

F03: Corinne Rancurel, Martine Da Rocha and Etienne G J Danchin. Alienness : Rapid detection of horizontal gene transfers in metazoan genomes

F04: Anne Friedrich, Cyrielle Reisser, Paul Jung, Gilles Fischer and Joseph Schacherer. Population genomics reveals the evolutionary fate of a large-scale introgression in a protoploid yeast species

F05: David Pflieger, Anastasie Sigwalt, Jing Hou and Joseph Schacherer. An automated R pipeline to analyze the genetic complexity of stress tolerance in Saccharomyces cerevisiae

F06: Stefanie Mühlhausen and Martin Kollmar. Analysing the taxonomic distributions of conserved introns with GenePainter

F07: Sophie Siguenza, Hélène Badouin, Stéphane De Mita, Jérôme Gouzy and Ludovic Cottret. WeggLib, a web interface for population genomics

F08: Hélène Badouin, Jérôme Gouzy, Alodie Snirc, Sophie Siguenza, Antoine Branca and Tatiana Giraud. Population genomics of the phytopathogenic fungi Microbotryum violaceum

F09: Gabriel V Markov, Praveen Baskaran and Ralf J Sommer. The same or not the same: Lineage-specific gene expansions and homology relationships in multigene families in nematodes

F10: Dong Seon Kim, Hye Ji Oh, Dongjin Choi and Yoonsoo Hahn. MOXD2 Gene Inactivation in Apes and Whales

F11: Hye Ji Oh, Dongjin Choi and Yoonsoo Hahn. Evolution of Intronless Genes in Ciona genus

F12: Dong Seon Kim, Hye Ji Oh, Dongjin Choi and Yoonsoo Hahn. Gains and Losses of N-glycosylation Sites during Human Evolution

F13: Jackson Peter, Anne Friedrich, Agnès Llored, Anders Bergstrom, Anastasie Sigwalt, Kelle Freel, Gianni Liti and Joseph Schacherer. The 1002 yeast genomes project: a framework for genome-wide association studies

F14: Francesc Peris-Bondia and Laurence Van Melderen. Comprehensive analysis of the genomic localization of bacterial toxin-antitoxin systems

F15: Julien Fumey, Céline Noirot, Hélène Hinaux, Sylvie Rétaux and Didier Casane. Evo Devo of Astyanax mexicanus cavefish: A new time frame and its consequence on the underlying evolutionary mechanisms.

F16: Fanny Pouyet, Marc Bailly-Bechet and Laurent Guéguen. Evolution of Codon Usage in E. coli

F17: Darius Kazlauskas and Česlovas Venclovas. Viral DNA replication: new insights and discoveries from large scale computational analysis

F18: Heloise Philippon, Céline Brochier-Armanet and Guy Perrière. Origin and Evolution of Cellular Signaling Pathways: PI3K as a case study

F19: Annalisa Fierro, Sergio Cocozza, Antonella Monticelli, Giovanni Scala and Gennaro Miele. Continuos and Discontinuos Phase Transitions in Quantitative Genetics: the role of stabilizing selective pressure

F20: Giovanni Scala, Ornella Affinito, Gennaro Miele, Antonella Monticelli and Sergio Cocozza. Evidence for Evolutionary and non Evolutionary phenomena shaping the genetic variant distribution near Transcription Start Sites

F21: Cécile Pereira, Alain Denise and Olivier Lespinet. A new method for improving the prediction and the functional annotation of ortholog groups

F23: Alexandra Vatsiou, Eric Bazin and Oscar Gaggiotti. Pathways enriched for selection

F24: Rafael Piergiorge, Ana Carolina Ramos Guimarães and Marcos Catanho. Evolution and functional genomics of analogous enzymes in the human genome

F25: Fabricia Nascimento and Allen Rodrigo. The evolutionary dynamics of endogenous retroviruses: testing the strict master and transposon models by computer simulations

F26: Nadav Rappoport and Michal Linial. Comparative Analysis of Insects' Complete Proteomes

F27: Nicolas Rodrigue and Nicolas Lartillot. Phylogenetic measurements of departures from the mutation-selection equilibrium

F28: Thies Gehrmann and Marcel Reinders. Proteny Discovering and visualizing statistically significant syntenic clusters at the proteome level between divergent genomes

G. Bioinformatics of health and disease

G01: Emile Rugamika Chimusa, Jacquiline Wangui Mugo and Nicola Mulder. Leveraging ancestry along the genome of admixed individuals to resolve missing heritability in disease scoring statistics

G02: Sylvain Mareschal, Pierre-Julien Viailly, Philippe Bertrand, Fabienne Desmots-Loyer, Elodie Bohers, Catherine Maingonnat, Karen Leroy, Thierry Fest and Fabrice Jardin. Next-Generation Sequencing applied to tailor targeted therapies in lymphoma: the RELYSE project

G03: Vivien Deshaies, Alban Lermine and Elodie Girard. Galaxydx – a web-server dedicated to ngs diagnosis data analyses

G04: Pravinkumar Patchaiappan. In silico molecular docking studies of biological active compounds from aegle marmelos against p53

G05: Marcos Avila, Daniel Torrente, Ludis Morales, Francisco Capani, Janneth Gonzalez and George E. Barreto. Structural insights from GRP78- NF-κB binding interactions: A computational approach to understand a possible neuroprotective pathway in brain injuries

G06: Robin Haw. Reactome Knowledgebase - Linking biological pathways, networks and disease.

G07: Niek de Klein, Sophie Rodius, Peter Nazarov, Arnaud Muller, François Bernardin, Céline Jeanty, Simone Niclou, Laurent Vallar and Francisco Azuaje. Connecting multiple gene expression signatures with candidate drugs for boosting heart regeneration potential.

G08: Amelie Desvars, Linda Vidman, Chinmay Dwibedi, Maria Furberg, Pär Larsson, Anders Sjöstedt, Anders Johansson and Patrik Rydén. Modeling the spatiotemporal distribution of tularemia in Sweden

G09: Nora Speicher and Nico Pfeifer. Integrative cancer subtype discovery using multiple kernel learning

G10: Andrew Nightingale, Tunca Dogan, Diego Poggioli, Guoying Qi, Jie Luo and Maria-Jesus Martin. The Role of UniProt's Protein Sequence Databases in Biomedical Research

G11: Cristina Menni, Steven Kiddle, Massimo Mangino, Ana Vinuela, Maria Psatha, Claire Steves, Martina Sattlecker, Alfonso Buil, Stephen Newhouse, Sally Nelson, Stephen Williams, Nicola Voyle, Hilkka Soininen, Iwona Kloszewska, Patrizia Meccoci, Magda Tsolaki, Bruno Vellas, Simon Lovestone, Tim Spector, Richard Dobson and Ana Valdes. Circulating proteomic signatures of chronological age

G12: Johnathan Watkins, Kayleigh Ougham, Markus Mayrhofer, Anders Isaksson, Andrew Tutt and Anita Grigoriadis. Exploring large-scale somatic variation in cancer for biomarker discovery

G13: Tom Petty, S Cordey, I Padioleau, M Docquier, L Turin, O Preynat-Seauve, E Zdobnov and L Kaiser. ezVIR: a user-friendly bioinformatics tool for human virus diagnostics from high-throughput sequencing of clinical specimens

G14: Pierre-Julien Viaillly, Arnaud Lefebvre and Hélène Dauchel. FunEVA: a user-friendly web application for Functional Effect Variation Analysis of human coding variations and their prioritization

G15: Arnaud Lefebvre, Alexandra Martins, Karim Labrèche, Vivien Deshaies, Alan Lahure, Pascaline Gaildrat and Hélène Dauchel. HExoSplice: a new software based on overlapping hexamer scores for prediction and stratification of exonic variants altering splicing regulation of human genes

G16: Yong Li, Ekkehart Lausch, Anika Salfelder, Karl Otfried Schwab, Natascha van der Werf-Grohmann, Tanja Velten, Dieter Lütjohann, Pablo Villavicencio Lorini, Uta Matysiak-Scholze, Bernhard Zabel and Anna Köttgen. Whole exome sequencing identifies variants causing different monogenic diseases in one nuclear family

G17: Anthony Mathelier, Calvin Lefebvre, Jiarui Ding, David J. Arenillas, Wyeth W. Wasserman and Sohrab P. Shah. Cis-Regulatory Somatic Mutations and Gene-Expression Alteration in B-cell Lymphomas

G18: Johanna Mazur, Isabella Zwiener and Harald Binder. Combining gene expression measurements from different platforms with a stratified boosting approach

G19: Medi Kori and Kazım Yalcın Arga. Uncovering the Interconnectivity Between Infertility-Associated Woman Diseases

G20: Pedro Brazão-Faria, Alexandra M. Carvalho, Susana Vinga and Nuno L. Barbosa-Morais. Analyses of alternative splicing landscapes in clear cell renal cell carcinomas reveal putative novel prognosis factors

G21: Anita Schuerch, Debby Schipper, Maarten A. Bijl, Jim Dau, Kimberlee B. Beckmen, Claudia M.E. Schapendonk, V. Stalin Raj, Albert D. M. E. Osterhaus, Bart L. Haagmans, Morten Tryland and Saskia L. Smits. Metagenomic survey for viruses through iterative assembly of taxonomic units

G22: Jelmar Quist, Johnathan A. Watkins, Pierfrancesco Marra, Andrew Nj Tutt and Anita Grigoriadis. Uncovering driver mechanisms of cancer using a gene module based approach

G23: Mathilde Daures, Anne Degavre, Cécile Julier and Anne Philippi. GAMES: Genetic Analysis and Mining of Exome Sequencing

G24: Therese Kellgren and Patrik Rydén. Experimental designs for finding disease-causing mutations in rare diseases

G25: Elizabeth Baker, Hamel Patel, Mizanur Khondoker, Stephen Newhouse and Richard Dobson. The use of Polygenic Risk Scores for predicting rate of cognitive decline in Alzheimer's disease

G26: Owen Lancaster, Tim Beck, Raymond Dalgleish and Anthony Brookes. Cafe Variome: enabling the federated discovery of genetic variant and phenotype data

G27: Clara Benoit-Pilven, Amandine Rey, Léon-Charles Tranchevent, Marie-Pierre Lambert, Hussein Mortada, Emilie Chautard, Laura Corbo, Béatrice Eymin and Didier Auboeuf. Alternative splicing and resistance to cancer targeted therapies

G28: Anna Feldmann and Nico Pfeifer. Predicting and Understanding HIV-1 Susceptibility to Broadly Neutralizing Antibodies

G29: Noémie Robil, Benoit Grellier, Fabien Petel, Ronald Rooke and Jacques Haiech. A new gene expression-based tool for selecting putative membrane cancer antigens named KANT

G30: Abhishek Dixit and Richard Dobson. BBGRE: brain and body genetic resource exchange

G31: Yupeng Cun and Holger Froehlich. Network and Data Integration for Biomarker Signature Discovery via Network Smoothed T-Statistics

G32: Kerstin Haase, Silke Raffegerst, Dolores Schendel and Dmitrij Frishman. Expitope: Web server for Epitope Expression

G33: Claudia Rincon, Isabel Brito and Philippe Hupé. Evaluation of different algorithms to stratify cancer tumors based on gene interaction networks and somatic mutations data

G34: Adrin Jalali and Nico Pfeifer. Interpretable per Case Weighted Ensemble Method for Cancer Associations

G35: Alejandra Medina-Rivera, Lina Antounians, Jessica Dennis, France Gagnon and Michael Wilson. Defining the cJun regulatory network in vascular endothelial cells from multiple species.

G36: Mahmoud Elhefnawi, Asmaa Ezzat and Mohd Noor Isa. Metagenomic and Metatranscriptomic analyses of the hepatocellular carcinoma-associated microbial communities and the potential role of microbial communities in liver cancer

G37: Maxim Ivanov, Kamil Khafizov and Sergey Kovalenko. OrphaPRED – functional effect prediction tool for mutations associated with rare diseases

G38: Laura Buzdugan and Peter Bühlmann. High-dimensional, predictive GWAS

G39: Francesco Iorio, Hayley Francies, Jayeta Saxena, Graham Bignell, Cyril Benes, Ultan McDermott, Simon Cook, Mathew Garnett and Julio Saez-Rodriguez. Systematic Prediction of Transcription Factors modulating Drug Response in Human Cancer Cell Lines

G40: Wolfgang Raffelsberger, Hélène Polveche, Amélie Weiss, Mickael Renaud, Anne Maglott-Roth, Johann Foloppe, Benoit Grellier, Etienne Weiss, Laurent Brino, Olivier Poch, Philippe Ancian and Philippe Erbs. Improving oncolytic viruses : Integrated data treatment pipeline

G41: Sumaiya Nazeen and Bonnie Berger. Integrative Analysis of Multiple Gene Expression Studies Reveals Genes and Pathways involved in Alteration of Steroidogenesis in Polycystic Ovary Syndrome (PCOS)

G42: Tim Beck, Robert Hastings and Anthony Brookes. GWAS phenotype data: standardised in the GWAS Central resource and harmonised in the GWAS PhenoMap browser

G43: Khalid Abnaof, Joao Dinis and Holger Fröhlich. Using Consensus Clustering to Explore Biological Effect Similarities of Drug Treatments based on Integrated Biological Knowledge from Multiple Sources – An Example Study on HIV and Cancer

G44: Jennifer E. Mollon, Steven J. Kiddle, Claire Steves, Kerrin Small, Martina Sattlecker, Katie Lunnon, Petra Proitsi, John Powell, Angela Hodges, Steven Williams, Tim Spector, Iwona Kloszewska, Patrizia Mecocci, Hilkka Soininen, Magda Tsolaki, Bruno Vellas, Simon Lovestone, Richard J. B. Dobson and Stephen Newhouse. Identification and Replication of Cis and Trans Effect Protein Quantitative Trait Loci in Ageing Adults

G45: David Källberg, Mattias Landfors, Yuri Belyaev and Patrik Rydén. Identifying subgroups of cancer – the blind two-sample test

G47: Laurence Pearl, Amanda Schierz, Simon Ward, Bissan Al-Lazikani and Frances Pearl. Drugging the DNA Damage Response

G48: Fiona Browne, Haiying Wang and Huiru Zheng. Network Driven Analysis for Biomarker Discovery in Alzheimer's Disease

G49: Trevor Clancy and Eivind Hovig. Mining immune cell activity from tumor transcriptomes

G51: Mamunur Rashid, Alistair Rust, Jeroen Ridder and David Adams. Identification of Novel Non-Coding Driver Mutations in Cancer using Pattern Recognition

G52: Russel Sutherland, Salvador Diaz-Cano, Jane Moorhead and Richard Dobson. Predicting tumour grade across multiple adenocarcinomas using exome sequence data.

G54: Haeseung Lee and Wan Kyu Kim. An integrative analysis of multi-dimensional compound-disease signatures for in silico drug repositioning in human cancers

G56: Bartosz Wojtas and B. Kaminska. TCGA-based analysis of gliomas uncovers a putative role of miRNA 155 in regulation of gene expression

H. Biological knowledge discovery from data, texts and bio-images

H01: Louis Cronje, Rian Pierneef and Oleg Reva. Ribosomal RNA operons in horizontally transferred genomic islands – facts or artifacts?

H02: Sylvain Demey, Evelyne Begaud, Nahla Chaïbi, Jean-Mary Gallais, Elodie Raynal-Melchy, Loïc Talignani, Emmanuelle Helloin, Serge Casaregola, Anne Favel, Martial Briand, Florence Valence-Bertel, Marie-Laure Dardé, Isabelle Villena, Rémy Guyoneaud, Michaël Pressigout and Chantal Bizet. BRC-LIMS : application for the management of French Biological Resource Centres of Microorganisms

H03: Eero Lihavainen, Mikhail Kislin, Leonard Khirug and Andre Ribeiro. Automatic quantification of mitochondrial fragmentation from two-photon microscope images of mouse brain tissue

H04: Melissa Mary and Gansel Xavier. LOINC & SNOMED-CT: usability and challenges to code identification tests and results for automated in vitro diagnostics systems

H05: Matej Stano, Gabor Beke and Lubos Klucar. viruSITE - database of viral genomes

H06: Esther Schmidt, Oliver Pelz, Svetlana Buhlmann, Maximilian Koch and Michael Boutros. GenomeRNAi: A Phenotype Database for Large-scale RNAi Screens

H07: Michael Lenz, Daniela Malan, Joana Frobel, Wolfgang Wagner, Philipp Sasse and Andreas Schuppert. A two-scale gene expression landscape for the characterization of in vitro differentiated cells

H08: Anne Mai Wassermann. Integrating historical biological data for small molecule-target predictions

H09: Catherine Kirsanova, Ugis Sarkans and Gabriella Rustici. Cellular Phenotype Database with Ontology assisted data browsing

H10: Lisa M Breckels, Sean Holden, Kathryn S Lilley and Laurent Gatto. A Transfer Learning Framework for Organelle Proteomics Data

H11: Tsubasa Ogawa, Kenji Etchuya and Yuri Mukai. Studies on factors and prediction of protein palmitoylation using a back propagation artificial neural network (BP-ANN)

H12: Afaf Benhouda, Mouloud Yahia and Hachani Khadraoui. Gastroprotective activity of Umbilicus rupestris leaf extract in experimental animal

H13: Jaroslav Budis, Rastislav Hekel, Gabriel Minarik and Tomas Szemes. Application and web service for functional annotation of variants

H14: Nikolay Samusik, Brice Gaudilliere, Gabriela Fragiadakis, Robert Bruggner, Martin Angst and Garry Nolan. Fast and accurate mapping of phenotypic space in single-cell data with X-shift

H15: Peter Ebert, Christoph Bock and Thomas Lengauer. A computational approach towards cross-species epigenomics

H16: Kota Hamada, Kenji Etchuya and Yuri Mukai. Influence of signal-peptide sequences on subcellular location of mature proteins

H17: Kenji Etchuya and Yuri Mukai. Structural Characteristics of Fuc Modification Sites in Mammalian Proteins

H18: Emad Elsebakhi, Rashid Al-Ali, Mohamed-Ramzi Temanni, Abdou Kadri, Radja Badji and Rawan Alsaad. Decision support and outcome prediction within SIDRiTrip translational research informatics platform

H19: Nikolaos Papanikolaou, Georgios Pavlopoulos, Evangelos Pafilis, Theodosios Theodosiou, Reinhard Schneider, Venkata Pardhasaradhi Satagopam, Christos Ouzounis, Aristides Eliopoulos, Vasilis Promponas and Ioannis Iliopoulos. BioTextQuest+: A knowledge integration platform for literature mining and concept discovery

H20: Yan Zhang, Isabel Riba-Garcia, Richard Unwin, Henning Hermjakob and Andrew Dowsey. Streaming Visualisation for Raw Mass Spectrometry Data and Results Based on a Novel Compression Algorithm

H21: Gaston Mazandu and Nicola Mulder. Information Content-based Gene Ontology Functional Similarity Measures: How good are these measures?

H22: Arthur Tenenhaus, Vincent Guillemot, Vincent Perlbarg, Andigoni Malousi, Justine Guégan and Ivan Moszer. RGCCA: a versatile tool for the analysis of multiblock and multigroup datasets

H23: Abhishek Dixit and Richard Dobson. CohortExplorer: A generic application programming interface (API) for entity attribute value database schemas

H24: Alexandre Angers-Loustau, Mauro Petrillo, Alex Patak and Joachim Kreysa. GMO-Scan: a tool for fast identification of transgenic elements in DNA sequences.

H25: Amos Folarin, Caroline Johnston, Zina Ibrahim, Mark Begale, David Mohr and Richard Dobson. Exploiting the Quantified Self for clinical care: a framework for integrating mobile and sensor data into the electronic health record

H26: Gabin Personeni, Simon Daget, Céline Bonnet, Philippe Jonveaux, Marie-Dominique Devignes, Malika Smail-Tabbone and Adrien Coulet. Mining Linked Open Data : a Case Study with Genes Responsible for Intellectual Disability

H27: Christophe Becavin, Nina Sesto, Jeffrey Mellin, Francis Impens and Pascale Cossart. Listeriomics: Systems biology of the model pathogen Listeria

H29: Vladimir Gligorijevic, Vuk Janjic and Natasa Przulj. Integration of molecular network data reconstructs Gene Ontology

H30: Matthias Ziehm, Aditi Bhat, Dobril K. Ivanov, Matthew D. Piper, Linda Partridge and Janet M. Thornton. Computational Biology of Ageing and Longevity in Model Organisms – Meta-Analyses of Survival Data and the SurvCurv Online Resource

H31: Ranjeet Bhamber, Yan Zhang, Isabel Riba-Garcia, Julian Selley, Richard Unwin and Andrew Dowsey. The seaMass Framework: Peptide feature extraction from raw mass spectrometry data with non-negative sparse Poisson regression and learnt predictors

H32: Hanqing Liao, Isabel Riba-Garcia, Richard Unwin, Jeffrey Morris and Andrew Dowsey. Group-wise Image Registration-nOrmalisation (GIRO): Retention time alignment and abundance normalisation for LC-MS data

H33: Nina Verstraete, Ignacio Sanchez and Diego Ferreiro. Spatial organization and distribution of linear motifs in the Ankyrin repeat protein family and its binding partners

H34: Yasuhito Inoue and Yasutaka Nakata. Statistical Analysis of Adrenergic Receptors

H35: Stefan Naulaerts, Pieter Meysman, Wim Vanden Berghe, Bart Goethals and Kris Laukens. Mining the human proteome for conserved mechanisms

H36: Pooya Zakeri, Leon-Charles Tranchevent and Yves Moreau. Kernel-Based Gene Prioritization Using Late Integration versus Geometric Kernel Fusion

H37: Xin He, Ernest Walzel, Douglas Armstrong and Ian Simpson. DisEnT, a unified ontology based gene set enrichment analysis (GSEA) framework for gene disease and gene phenotype studies.

H38: Lucila Aimo, Robin Liechti, Anne Niknejad, Nevila Nouspikel, Anne Gleizes, Dmitry Kuznetsov, Fabrice David, Vassily Hatzimanikatis, Howard Riezman, F. Gisou van der Goot, Lydie Bougueleret, Ioannis Xenarios and Alan Bridge. SwissLipids – a knowledge resource for lipid biology

H39: Nisar Shar, Vijayabaskar Ms and David Westhead. Predicting transcription factor mutual interactions from ENCODE data

H40: Jian-Long Huang, Ming-Yi Hong and Chien-Yu Chen. SeqHouse: a web platform for biological data integration and management

H41: Mauro Petrillo, Alexandre Angers, Alex Patak and Joachim Kreysa. Screening of public nucleotide databases with PCR simulation to generate a secondary database containing sequences related to Genetically Modified Organisms.

H42: Aida Mrzic, Trung Nghia Vu, Dirk Valkenborg, Evelyne Maes, Filip Lemière, Bart Goethals and Kris Laukens. Pattern detection in associated artifact peaks in mass spectra with frequent itemset mining

H44: Vidya Oruganti, Martin C. Simon and Marcel H. Schulz. Small RNA Analysis and Visualization in P. tetraurelia

H45: Albert Pallejà Caro, Sune Frankild, David Westergaard, Pope Moseley and Søren Brunak. Linking ICD10/SNOMED CT concepts to human genes

J. Methods and technologies for computational biology

J01: Francesco Musacchia, Swaraj Basu, Marco Salvemini and Remo Sanges. Annocript: a flexible pipeline for transcriptome annotation that can also identify putative long non-coding RNAs

J02: Jerome Mariette, Frederic Escudie, Philippe Bardou and Christophe Klopp. Jflow: A fully scalable Javascript workflow management system

J03: Jean Philippe Tamby, Rim Zaag, Jean-Paul Bouchet, Cecile Guichard, Philippe Grevet, Marie-Laure Martin-Magniette, Sébastien Aubourg and Véronique Brunaud. Evolution of the FLAGdb++ integrated environment for exploring plant genomes

J04: Yuriy Vaskin, Francesco Venco and Heiko Muller. SMITH: managing NGS data and workflows in a sequencing facility

J05: E'Krame Jacoby, François Le Fèvre, Ludovic Fleury, Sandrine Lalami, Audrey Lemaçon, David Vallenet, Guillaume Albini, Claudine Médigue and Claude Scarpelli. BIRDS, a rule based framework for automating generation and management of bioinformatics treatments

J06: Oliver Horlacher, Frederic Nikitin, Davide Alocci, Julien Mariethoz, Markus Mueller and Frederique Lisacek. MzJava: an open source mass spectrometry library

J07: Yannick de Oliveira, Olivier Sosnowski, Alain Charcosset and Johann Joets. BioMercator 4.0: A complete framework to integrate QTL, meta-QTL and genome annotation

J08: Yannick De Oliveira, Guy-Ross Assoumou-Ella, Johann Joets and Alain Charcosset. Thalia : A database dedicated to association genetics in plants

J09: Jeongsu Oh, Chi-Hwan Choi, Soon Gyu Hong, Wan-Sup Cho and Kyung Mo Kim. CLUSTOM-CLOUD: In-Memory Data Grid-based software for clustering large scale 16S rRNA sequence data in the cloud environment

J10: Katarina Truvé, Martin Norling and Erik Bongcam-Rudloff. SEQscoring: a web-based tool for interpretation and visualization of case control data from massive parallel sequencing (MPS) projects

J11: Daniele Pierpaolo Colobraro and Paolo Romano. A new implementation of CABRI Web Services

J12: Rim Zaag, Jean-Philippe Tamby, Cécile Guichard, Zakia Tariq, Guillem Rigaill, Etienne Delannoy, Jean-Pierre Renou, Sébastien Aubourg, Marie-Laure Martin-Magniette and Véronique Brunaud. GEM2NET: From gene expression modeling to -omics network to discover Arabidopsis thaliana genes involved in stress response

J13: Prachi Mehrotra, Vimlakany G Ami and Narayanaswamy Srinivasan. Classification of multi-domain proteins using CLAP server: case studies on proteins containing tyrosine phosphatases and SH3 domains

J14: Jason Williams. iPlant Collaborative: A Unified Cyberinfrastructure for the Life Sciences

J16: David Brown, Rowan Hatherley and Özlem Tastan Bishop. HUMA: A web server and database for the analysis of genetic variations in humans

J17: Burçak Otlu, Sunduz Keles and Oznur Tastan. GLANET: Genomic Loci Annotation and Enrichment Tool

J18: Dominik Lutter. A generalized additive model approach for high throughput screening approaches

J19: Roland Barriot, Petra Langendijk-Genevaux, Yves Quentin and Gwennaele Fichant. Semi-automatic Validation of Genome-wide Reassembled Systems by Gene Prioritization through Weighted Data Fusion

J20: Sebastian Seitz and Tatyana Goldberg. Sorting the nuclear proteome using machine learning

J22: Aurélien Naldi, Pedro T. Monteiro, Denis Thieffry and Claudine Chaouiya. GINsim: a software tool for the modelling and analysis of logical regulatory networks

J23: Pedro L. Varela, Pedro T. Monteiro, Nuno Mendes, Adrien Fauré and Claudine Chaouiya. EpiLog, a computational tool for the logical modelling of epithelial pattern formation

J24: Stephanie Le Gras, Serge Uge, Matthieu Jung, Ludovic Roy, Valerie Cognat, Frederic Plewniak, Irwin Davidson and Julien Seiler. GalaxEast: an open and powerful Galaxy instance for integrative Omics data analysis

J25: Y-H. Taguchi. Heuristic principal component analysis based unsupervised feature extraction and its application to bioinformatics

J26: Xavier Prudent and Michael Hiller. Linking Phenotypes and Genomic regions: the Forward Genomics Approach

J27: Fotis Psomopoulos and Christos Ouzounis. Computation and visualization of ancestral pathway reconstruction and inference

J28: Jörgen Brandt, Marc Bux and Ulf Leser. Cuneiform - Parallel Execution of NGS Workflows

J29: Jasmin Straube, Alain Dominique Gorse, Bevan Emma Huang and Kim-Anh Lê Cao. A linear mixed model spline framework for analyzing time course 'omics' data

J30: Svetlana Artemova, Mael Bosson, Jocelyn Gate, Sergei Grudinin, Leonard Jaillet, Marc Piuzzi, Petr Popov and Stephane Redon. SAMSON: Software for Adaptive Modeling and Simulation Of Nanosystems

J31: Nicolas Sapay, Ghita Rahal and Artem Khlebnikov. Multi-omics data integration platform in public private partnership

J32: Jonathan Mercier, Alexandre Renaux, David Vallenet, Adrien Josso, François Lefèvre, E'Krame Jacoby Ayari, Guillaume Albini, Aurélie Genin-Lajus, Claude Scarpelli and Claudine Médigue. The MicroScope platform: from data integration to a rule-based system for massive and high-quality microbial genome annotation

J33: Óscar Torreño Tirado and Oswaldo Trelles. Easily registering bioinformatics services metadata

J34: Ryohei Suzuki, Daisuke Komura, Kazuki Yamamoto and Shumpei Ishikawa. MOLding: Gesture-based Interactive Molecular Dynamics for Protein Structure Manipulation

J35: Sandie Arnoux, Yvon Jégou, Gaël Beaunée and Pauline Ezanno. A generic framework to model infection dynamics in a metapopulation of cattle herds

J36: Tor Johan Mikael Karlsson, Óscar Torreño Tirado and Oswaldo Trelles. jORCA: Jumping to the Cloud

J37: Alexis Allot, Laetitia Poidevin, Kirsley Chennen, Raymond Ripp, Julie Thompson, Olivier Poch and Odile Lecompte. GeneBook: a social network linking genes, diseases and researchers

J38: Caroline Siegenthaler and Rudiyanto Gunawan. Assessing Inference Methods in the Absence of Gold Standard Networks: Can Crowdsourcing Help?

J39: Felipe Albrecht, Christoph Bock and Thomas Lengauer. DeepBlue: Epigenomic Data Server

J40: Sarvesh Nikumbh and Nico Pfeifer. On the Hardness of Computationally Predicting Long-Range Chromatin Interactions

J41: Kazuki Kishi, Daisuke Komura, Takayuki Isagawa and Shumpei Ishikawa. Visualizing whole cancer-stromal interactome

J42: Nicolas Tchitchek and Christophe Becavin. MDS-Reference Maps and MDS-Voronoi Representations for Visualization and Analysis of High Dimensional –Omics Profiles

J43: Frédéric Mahé, Torbjørn Rognes, Christopher Quince, Colomban de Vargas and Micah Dunthorn. Swarm: robust and fast clustering method for amplicon-based studies

J44: Christofer Bäcklin and Mats Gustafsson. Developer friendly and computationally efficient predictive modeling without information leakage: The emil package for R

J46: Guoxian Yu, Hailong Zhu and C. Domeniconi. Predicting Protein Functions using Incomplete Hierarchical Labels

J47: Jan Grau, Jens Boch and Stefan Posch. Genome-wide TALEN off-target prediction and its utility for TALEN design

J48: Iris Leitner and Oswaldo Trelles. Intuitive library for efficient access of compressed genome sequences

J49: Jose Arjona-Medina, Óscar Torreño Tirado and Oswaldo Trelles. Software for featuring genome evolution

J50: Hanna Ćwiek, Augustyn Markiewicz and Paweł Krajewski. Phenalyser: a web-based ISA-TAB-compliant tool for analysis of phenotyping experiments

J51: Helen Lindsay, Alexa Burger, Jonas Zaugg, Christian Mosimann and Mark Robinson. An R toolkit for studying the CRISPR-Cas9 mutation spectrum

J52: Maciej Pajak, Clive Bramham and Ian Simpson. Computational approaches to improving miRNA-mRNA interaction predictions

J53: Chen Meng, Bernhard Küster, Aedín Culhane and Amin Moghaddas Gholami. Integration of multiple omics data for detecting cluster specific perturbed gene sets

J54: Jorge Álvarez-Jarreta and Gregorio de Miguel Casado. PhyloViewer: A Viewer for Large Phylogenies

J55: Maarja Lepamets, Lauris Kaplinski, Reidar Andreson and Maido Remm. GenomeTester 4.0 – a k-mer analyzer package for sequencing reads and genomes

J56: Kerstin Johnsson, Jonas Wallin and Magnus Fontes. Model-based mutual clustering of flow cytometry data through Bayesian hierarchical modeling

J57: Damien Correia, Olivia Doppelt-Azeroual, Jean-Baptiste Denis, Mathias Vandenbogaert and Valérie Caro. MetaGenSense: A Web application for analysis and visualization of High throughput Sequencing metagenomic data

J58: Nadia El-Mabrouk, Laurent Gueguen, Manuel Lafond, Emmanuel Noutahi, Jonathan Séguin, Magali Semeria and Eric Tannier. Genome-wide gene tree correction

J59: Eugenio Mattei, Fabrizio Ferrè and Manuela Helmer-Citterich. BEAR-Suite: a collection of tools for RNA structural comparison

J60: Bryan Kowal, Akram Mohammed and Tomas Helikar. Building, simulating, and analyzing large-scale logical models of complex biological systems in a collaborative fashion with the Cell Collective

J61: Atefeh Lafzi, Saber Hafezqorani, Yesim Aydin Son and Hilal Kazan. Post-transcriptional regulation mediated by the interplay between RNA-binding proteins and miRNAs

J62: Johannes Köster and Sven Rahmann. Massively parallel read mapping on GPUs with PEANUT

J63: Matúš Kalaš, Sveinung Gundersen, László Kaján, Jon Ison, Steve Pettifer, Christophe Blanchet, Rodrigo Lopez, Kristoffer Rapacki and Inge Jonassen. BioXSD — A data model for sequences, alignments, features and measurements

J64: Jorge Alvarez-Jarreta, Gregorio de Miguel Casado and Elvira Mayordomo. PhyloFlow: A Fully Customizable and Automatic Workflow for Phylogeny Estimation

J65: Michał Woźniak, Limsoon Wong and Jerzy Tiuryn. A new statistical test for detection of drug resistance-associated mutations in bacteria

J66: Gonzalo Garcia Accinelli, Bernardo Clavijo, Sarah Ayling and Mario Caccamo. A systematic approach to plant genome assembly.

J67: Christophe Blanchet and Jean-François Gibrat. Bioinformatics cloud services of the French Institute of Bioinformatics

J68: Foteini Pappa, Varvara Karagkiozaki, Paraskeui Kavatzikidou and Stergios Logothetidis. Development of conductive fiber-based scaffolds for tissue engineering and cellular uptake

Social events

Sunday: Welcome cocktail

The ECCB'14 welcome cocktail will take place right after the opening lectures, at the Congress Center, Palais de la Musique et des Congrès (PMC), in the dining room Contades, on Sunday, September 7 from 19:30 to 21:00.

Monday: Ice-breaking event

An informal ice-breaking event will be organized on Monday, September 8 from 20:30 to 22:30.

Meeting point is at the main floor right after the poster session.

Tuesday: Gala evening

Due to the unexpected number of participants, we have been obliged to arrange two propositions for the gala evening. The first initial proposition takes place at the Council of Europe, in a smart historical ambience, preceded by a guided river boat tour (about 50 min). Capacity: 800 persons. The second proposition takes place at a typical and famous Alsacian restaurant:the Ancienne Douane, and is preceded by a guided mini-train tour of Strasbourg (about 1 hour). Capacity is about 150 persons. People distribution between the two events will be determined in advance because the Council of Europe requires an exact list of participants' names for security reasons. Sorry for this constraint. Possible exchange modalities will be announced at the conference.

Proposition 1. A river boat trip followed by a dinatory cocktail at the Council of Europe

Departures of river boat will be at 'Palais Rohan' Batorama pier, close to the Cathedral (see itinerary p.5 and map p.10) by groups of about 130 persons at following embarking times:

18:00 ; 18:15 ; 18:45 ; 19:00 ; 19:15 ; 19:30 ; 19:45

The boat will drop you at the pier close to the Council of Europe.

Venue: Council of Europe, Avenue de l'Europe, Strasbourg (see Transport Map p.5).

Standing buffet will begin around 21:00. Drinks and music will be proposed during the waiting time.

Return to the hotels will be with the bus or tram (until 0:30).

Proposition 2. A mini-train round trip followed by Alsacian food, drinks and music at the Ancienne Douane

Departures of mini-train will be Place Gutenberg, close to the Cathedral (see itinerary p.5 and map p.10) by groups of about 50 persons at following times:

18:00 ; 19:00 ; 20:00

Venue: The Ancienne Douane is located Rue de la Douane in Strasbourg center.

Standing buffet will begin around 21:00. Drinks and music will be proposed before.

Return to the hotels will be with the bus or tram (until 0:30) or by foot (no time limit!).

Science Web Art: Banner Contest

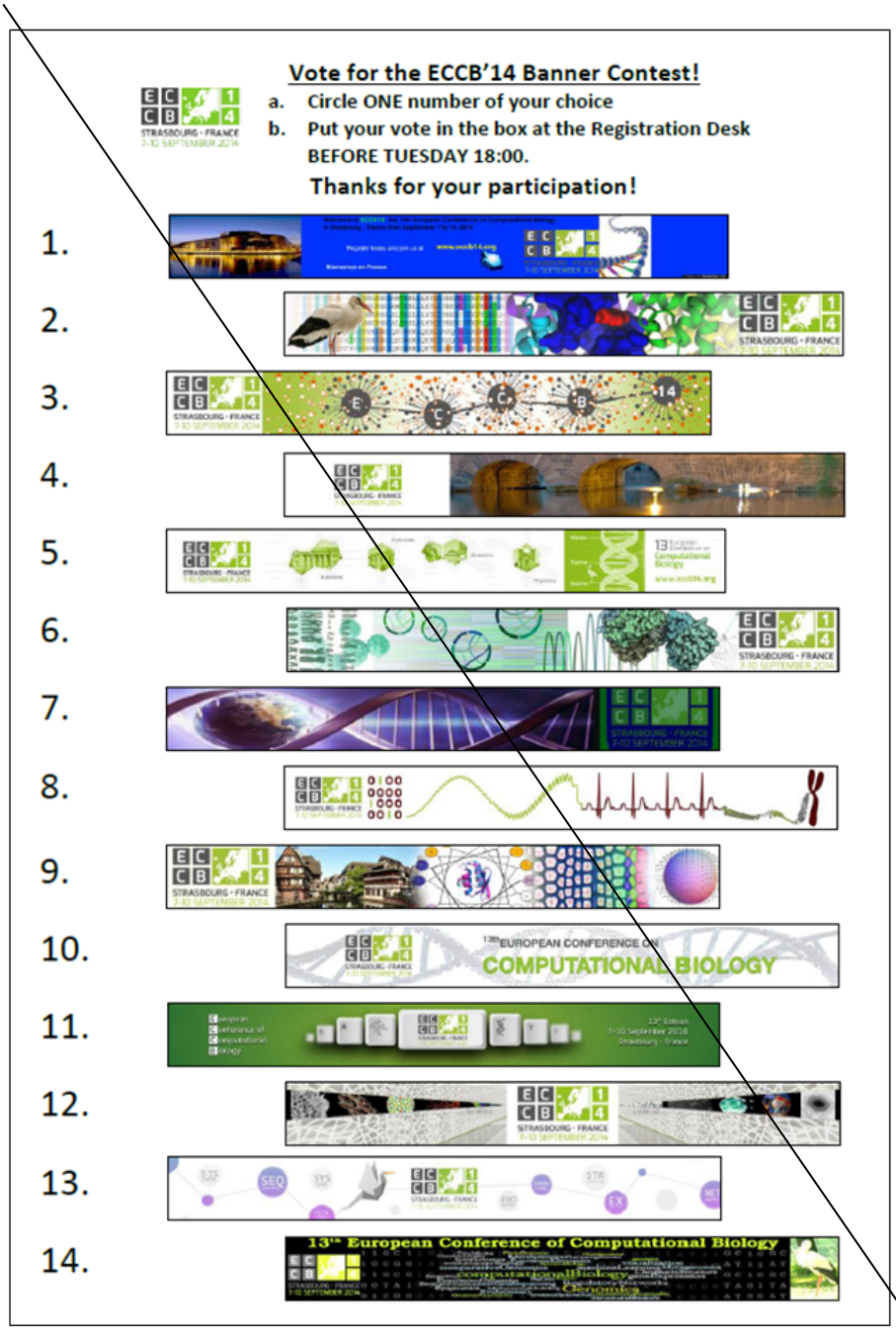
For its 2014 edition, the ECCB organizing committee has organized a competition for the best web banner dedicated to the ECCB 2014 web site. Details of the call can be found at http://www.eccb14.org/guidelines/banner-contest. This contest is sponsored by the LORIA: Lorraine Laboratory of Computer Science and its Applications, Nancy, France.

Fourteen original banners have been submitted and are visible at http://www.eccb14.org/home/banner-gallery?page=1#category. A new banner is loaded randomly at each new connection to the eccb14.org site

Every ECCB'14 participant will receive a ballot to vote for his favorite banner. A ballot box will be available at the Registration Desk.

The winner will have the possibility to choose between an iPad and a free (refunded) registration.

The vote will be closed Tuesday Sept. 9th at 18:00 and the winner will be announced during the evening gala.



13th European Conference on Computational Biology

ECCB 14

STRASBOURG · FRANCE

7-10 SEPTEMBER 2014

Organizing committee : Strasbourg : Luc Moulinier, Anne Ney , Olivier Poch, Julie Thompson · **Nancy :** Adrien Coulet, Emmanuelle Morin, Dave Ritchie, Malika Smaïl-Tabbone · **ISCB :** Marie-France Sagot · **Jebif :** Magali Michaut · **SFBI :** Sophie Schbath · **Luxembourg :** Francisco Azuaje · **Germany :** Mario Albrecht, Hans-Peter Lenhof, Thomas Lengauer.

ISCB · SFBI · cnrs · Inria · Université de Lorraine · Université de Strasbourg · Loria

www.eccb14.org