WELCOME!

On behalf of the ECCB10 Organizing and Steering Committees we would like to welcome you to beautiful Ghent, one of Europe's most exciting cities. We hope you will enjoy ECCB10 - four days of presentations, poster sessions, workshops, software demonstrations and networking with other leading bioinformatics researchers, computational biologists and bio-IT professionals from around the world. ECCB is the leading European meeting in our field and a primary international conference at which people present and discuss advances in computational biology.

We are proud to have five keynote lectures by distinguished speakers Peer Bork (European Molecular Biology Laboratory, Heidelberg, Germany), Elaine Mardis (Washington University Genome Center, St-Louis, MO, USA), Michael J. E. Sternberg (Imperial College, London, UK), Yves Van de Peer (Flemish Institute of Biotechnology and University of Ghent, Belgium) and Hans Westerhoff (University of Manchester, UK and VU University Amsterdam, The Netherlands).

In addition, there will be a number of interesting workshops and tutorials on Sunday, including the first European Student Council Symposium, a workshop on leveraging perturbation effects towards reverse engineering biological networks and one on extraction and reuse of genotype-phenotype knowledge from scientific literature and databases. There is no better way to kick off ECCB10 than to participate in one of these workshops and tutorials!

On Monday afternoon, we will have a special joint lecture by artist Koen Vanmechelen and geneticist Jean-Jacques Cassiman. They will present the Cosmopolitan Chicken Project – Koen's quest to create a universal, cosmopolitan chicken. Further, they will discuss how art and science influence each other.

We would like to acknowledge the help of all people who made ECCB10 happen, as well as our commercial and academic sponsors – thanks to them we were able to offer 62 fellowships for early-stage scientists. Finally, we would like to thank you for your participation. Enjoy the conference!

Yves Moreau & Jaap Heringa

TABLE OF CONTENTS

HOTELS

NH Gent Sint Pieters
Koning Albertlaan 121, 9000 Ghent, Belgium
+32 9 222 60 65

Holiday Inn Gent Expo
Maaltekouter 3, 9051 Ghent, Belgium
+32 9 220 24 24

NH Gent Belfort
Hoogpoort 63, 9000 Ghent, Belgium
+32 9 233 33 31

Ghent Marriott Hotel
Korenlei 10, 9000 Ghent, Belgium
+32 9 233 93 93

Hotel Gravensteen
Jan Breydelstraat 35, 9000 Ghent, Belgium
+32 9 225 11 50

Ghent River Hotel
Waaistraat 5, 9000 Ghent, Belgium
+32 9 266 10 10

Hotel de Flandre
Poel 1, 9000 Ghent, Belgium
+32 9 266 06 00

Hotel Ibis Gent Centrum Opera
Nederkouter 24-26, 9000 Ghent, Belgium
+32 9 225 07 07

Hotel Ibis Centrum St-Baafs Kathedraal
Limburgstraat 2, 9000 Ghent, Belgium
+32 9 233 00 00

Holiday Inn Gent Express
Akkerhage 2, 9000 Ghent, Belgium
+32 9 222 58 85

Campanile
Akkerhage 1, 9000 Ghent, Belgium
+32 9 220 02 22

Poortackere Monasterium
Oude Houtlei 58, 9000 Ghent, Belgium
+32 9 269 22 30

Hotel Novotel Gent Centrum
Goudenleeuwplein 5, 9000 Ghent, Belgium
+32 9 224 22 30

Carlton Hotel Gent
Koningin Astridlaan 138, 9000 Ghent, Belgium
+32 9 222 88 36

Hostel 47
Blekerijstraat 47, 9000 Ghent, Belgium
+32 478 71 28 27

HI De Draecke Gent
Sint-Widostraat 11, 9000 Ghent, Belgium
+32 9 233 70 50

Ecohostel Andromeda
Bargiekaai 35, 9000 Ghent, Belgium
+32 486 67 80 33

LOCAL PUBLIC TRANSPORT

Basics

'De Lijn' (http://www.delijn.be) provides trams and buses in Ghent. Using public transport is easy and travelers are offered several ticket options, but in any case buying your ticket in one of the distribution points (train station, kiosks, supermarkets…) is at least 25% cheaper than buying it from the bus or tram driver. If you are considering taking several bus/tram trips you might be interested in the "lijnkaart", which grants the owner 10 trips at a discount rate. People traveling in groups of 5 or more can buy a "lijnkaart %" which also provides 10 trips at an even lower rate. If you plan on using public transportation a lot, a day-pass or a 3-day-pass might be the most convenient way of traveling through Ghent. Once you have bought a ticket traveling is as simple as getting on the bus, saying "Goeiedag!" to the driver and inserting the ticket into the yellow validation device (don't forget to take it back). Transfers are allowed for one hour after validation.

Ticket Type	Price at distribution point	Price on the bus/tram
Single ticket	€ 1,20	€ 2
Lijnkaart (10 trips)	€ 8	€ 15
Lijnkaart % (10 trips)	€ 6	€ 11
Day-pass	€ 5	€ 6
3-day-pass	€ 10	€ 12
Museum Pass (3-day-pass, includes access to the city's museums): € 20		

From the conference site to the city centre

Bus 5 is a convenient way to go to the city centre. It leaves from the bus stop in front of the conference site. You will find a map of public transport in the conference bag.

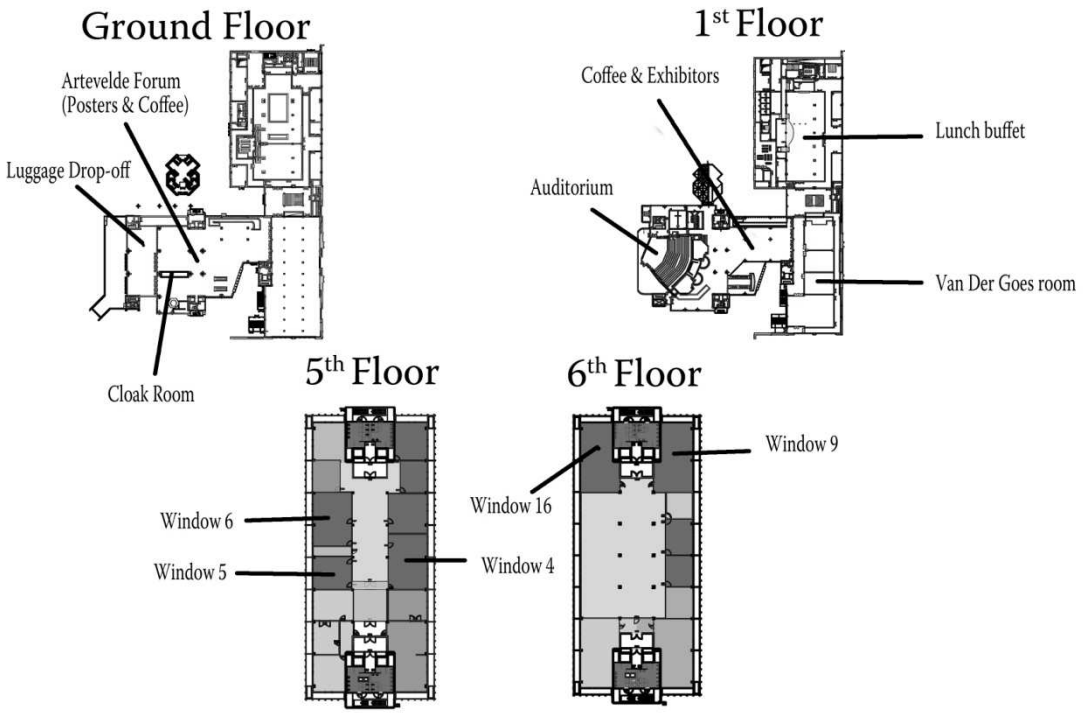
Ghent Museum Pass

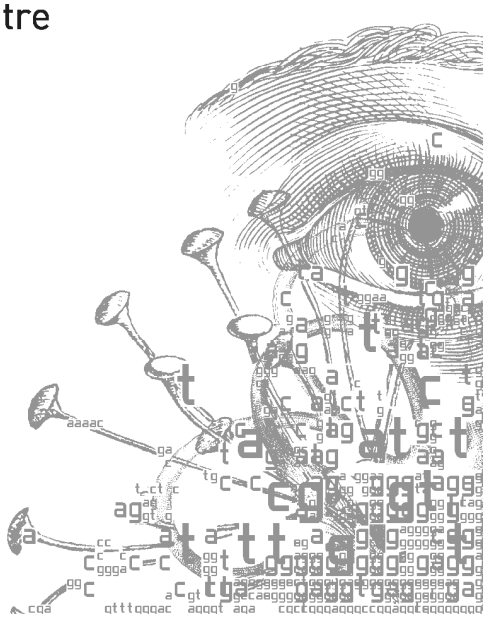
The Ghent Museum Pass costs € 20 and offers access to various museums and municipal buildings for 3 days. During those 3 days the Ghent Museum Pass is also valid as a ticket for all De Lijn vehicles in the Ghent city area. The validity period starts on the date that you write on the back of the pass. You can purchase a Ghent Museum Pass in advance and use it for your journey to the first museum. You can purchase the Ghent Museum Pass in the Lijnwinkels in East Flanders, in participating museums, from the tourism department of the city of Ghent, and in several Ghent hotels.

The participating museums and municipal buildings are: Belfry, Friary of the Carmelites, Design Museum Ghent, De wereld van Kina: the Garden, De wereld van Kina: the House, The Castle of the Counts, Alijn House, Art Centre St Peter's Abbey, MIAT (Museum for Industrial Archaeology and Textiles), Museum Dr. Guislain, Museum for the History of the Sciences, Museum of Fine Arts, St Bavo's Cathedral + The Adoration of the Mystic Lamb, SMAK (City Museum for Modern Art), STAM (City Museum Ghent).

More information available at De Lijn's website (http://www.delijn.be/en/vervoerbewijzen/types/uitstap/#5).

FLOOR PLAN





ORGANIZATION

Conference Chair

Yves Moreau, Katholieke Universiteit Leuven, Belgium

Proceedings Chair

Jaap Heringa, VU University Amsterdam, The Netherlands

Local Organizing Committee

Yves Moreau, Katholieke Universiteit Leuven, Belgium
Jaap Heringa, VU University Amsterdam, The Netherlands
Gert Vriend, Radboud University, Nijmegen, The Netherlands
Yves Van de Peer, University of Ghent and VIB, Belgium
Kathleen Marchal, Katholieke Universiteit Leuven, Belgium
Jacques van Helden, Université Libre de Bruxelles, Belgium
Louis Wehenkel, Université de Liège, Belgium
Antoine van Kampen, University of Amsterdam and Netherlands Bioinformatics Center (NBIC)
Peter van der Spek, Erasmus MC, Rotterdam, The Netherlands

Steering Committee

Michal Linial (Chair), Hebrew University, Jerusalem, Israel
Soren Brunak, Technical University of Denmark, Lyngby, Denmark
Philipp Bucher, University of Lausanne, Switzerland
David Gilbert, University of Glasgow, UK
Thomas Lengauer, Max-Planck Institute for Informatics, Saarbrücken, Germany
Yves Moreau, Katholieke Universiteit Leuven, Belgium
Dietrich Rebholz-Schuhmann, European Bioinformatics Institute, Hinxton, UK
Marie-France Sagot, INRIA and Université Claude Bernard, Lyon, France
Alfonso Valencia, Universidad Autonoma, Madrid, Spain
Jaak Vilo, University of Tartu, Estonia

Program Committee – Area Chairs

A—Sequence Analysis, Alignment, and Next-Generation Sequencing

Des Higgins, University College Dublin, Ireland
Geoff Barton, University of Dundee, Scotland, UK

B—Comparative Genomics, Phylogeny, and Evolution

Martijn Huynen, Radboud University Nijmegen Medical Centre, The Netherlands
Yves Van de Peer, Ghent University and VIB, Belgium

C—Protein and Nucleotide Structure

Anna Tramontano, University of Rome "La Sapienza", Italy
Jan Gorodkin, University of Copenhagen, Denmark

D—Annotation and Prediction of Molecular Function

Nir Ben-Tal, Tel-Aviv University, Israel
Fritz Roth, Harvard Medical School, USA

E—Gene Regulation and Transcriptomics

Jaak Vilo, University of Tartu, Estonia
Zohar Yakhini, Agilent Laboratories, Tel-Aviv and the Technion, Haifa, Israel

F—Text Mining, Ontologies, and Databases

Alfonso Valencia, National Center for Cancer Research, Madrid, Spain
Guy Cochrane, European Bioinformatics Institute, Hinxton, UK

G—Protein Interactions, Molecular Networks, and Systems Biology

Denis Thieffry, University of Marseille, France
Benno Schwikowski, Institut Pasteur, Paris, France

H—Genomic Medicine

Yves Moreau, Katholieke Universiteit Leuven, Belgium
Niko Beerenwinkel, Federal Institute of Technology, Zürich, Switzerland

I—Other Bioinformatics Applications

Martin Vingron, Max Planck Institute for Molecular Genetics, Berlin, Germany
Hans-Peter Lenhof, Saarland University, Saarbrücken, Germany

Program Committee – Members

Edoardo Airoldi
Tero Aittokallio
Yasmin Alam-Faruque
Mario Albrecht
Andre Altmann
Simon Andrews
Rolf Backofen
Gary Bader
Guy Baele
Timothy Bailey
Pierre Baldi
Geoff Barton
Niko Beerenwinkel
Tim Beissbarth
Asa Ben-Hur
Nir Ben-Tal
Panayiotis Benos
Mathieu Blanchette
Christian Blaschke
Christoph Bock
Alexander Bockmayr
Richard Bonneau
Erich Bornberg-Bauer
Sebastian Böcker
Sylvain Brohée
Christine Brun
Philipp Bucher
Jeremy Buhler
Joachim Buhmann
Mario Caccamo
Rita Casadio
Daniele Catanzaro
Claudine Chaouiya
J. Michael Cherry
Andrea Cilliberto
Kevin B. Cohen
Chris Cole
Tim Conrad
Thomas Dandekar
Susmita Datta
Peter Dawyndt
Hidde de Jong
Charlotte Deane
Rahul Deo
Francisco Domingues
Bas E. Dutilh
Arne Elofsson
Olof Emanuelsson
Andrea Foulkes
Jan Freudenberg
Toni Gabaldon
John S. Garavelli
Mikhail Gelfand
Robert Giegerich
Mark Girolami
Frank Oliver Gloeckner
Richard Goldstein
Didier Gonze
Jan Gorodkin
Julian Gough
Sam Griffiths-Jones
Michael Gromiha
Alain Guenoche
Roderic Guigo
Udo Hahn
Turkan Haliloglu
Thomas Hamelryck
Jackie Han
Monica Heiner
Jaap Heringa
Philippe Herve
Matt Hibbs
Des Higgins
Andreas Hildebrandt
Lynette Hirschman
Ivo Hofacker
Liisa Holm
Torsten Hothorn
Richard Hughey
Martijn Huynen
Johannes Jaeger
Inge Jonassen
Igor Jurisica
Daniel Kahn
Simon Kasif
Kazutaka Katoh
Paul Kersey

Daisuke Kihara
Philip Kim
Oliver King
Katja Kivinen
Stefan Klamt
Steven H. Kleinstein
Ina Koch
Rachel Kolodny
Peter Konings
Martin Krallinger
Michael Krauthammer
Martin Kuiper
Stefan Kurtz
Roman Laskowski
Sonia Leach
Florian Leitner
Hans-Peter Lenhof
Christina Leslie
Olivier Lichtarge
Michal Linial
Ole Lund
Florian Markowetz
David Martin
Aurélien Mazurie
Alice McHardy
Yves Moreau
Bernard Moret
Norman Morrison
Vincent Moulton
Arcady Mushegian
Luay Nakhleh
Goran Nenadic
See-Kiong Ng
Cedric Notredame
Guillaume Obozinski
Uwe Ohler
Ron Ophir
Sandra Orchard
Martin Oti
Christos Ouzounis
Kimmo Palin
Jong Park
Helen Parkinson
John Parkinson
Itsik Pe'er
Anders Gorm Pedersen
Matteo Pellegrini
Bengt Persson
Mihaela Pertea
Graziano Pesole
Gianluca Pollastri
Mihai Pop
Jiang Qian
Jörg Rahnenführer
Domenico Raimondo
Ben Raphael
Mathias Rarey
Magnus Rattray
Gunnar Rätsch
Dietrich Rebholz-Schumann
Chandan Reddy
Knut Reinert
Marylyn Ritchie
Stéphane Robin
David Rocke
Michal Rosen-Zvi
Volker Roth
Pierre Rouzé
Patrick Ruch
Marie-France Sagot
Jasmin Saric
Michael Schroeder
Ora Schueler-Furman
Stefan Schuster
Russell Schwartz
Benno Schwikowski
Joachim Selbig
Roded Sharan
Hagit Shatkay
Istvan Simon
Saurabh Sinha
Kimmen Sjolander
Ingolf Sommer
Christopher Southan
Rainer Spang
John Spouge
Peter Stadler
Jörg Stelling
Peter Sterk
Josh Stuart
Shamil Sunyaev
Hakim Tafer
Murat Tasan
Chris Taylor
Weidong Tian
Silvio Tosatto
Anna Tramontano
Achim Tresch
Sophia Tsoka
Alfonso Valencia
Jacques van Helden
Vera van Noort
Wessel van Wieringen
Yves Van de Peer
Klaas Vandepoele
Roel Verhaak
Jean-Philippe Vert
Allegra Via
Albert Vilella
Jaak Vilo
Martin Vingron
Gert Vriend
San Ming Wang
Bertram Weiss
Andreas Wilm
Shoshana Wodak
Haim Wolfson
Yu Xia
Zohar Yakhini
Ralf Zimmer
Andrei Zinovyev

Program Committee – Co-reviewers

Mohamed Abouelhoda
Aare Abroi
Pelin Akan
Sandro Andreotti
Inka Appel
Aaron Arvey
Eric Bonnet
Michael Beckstette
Christian Bender
Stephan Bernhart
Chris Bielow
Enrique Blanco
Regina Bohnert
Daniel Bolser
Anne-Laure Boulesteix
Quang Bao Anh Bui
Peter Butzhammer
Rainer Böckmann
Liran Carmel
Robert Castelo
Yi-Chien Chang
Pao-Yang Chen
Manfred Claassen
Diego Di Bernardo
Francisco Dominguez
Anne-Katrin Emde
Florian Erhard
Gaston Fiore
Xijin Ge
Jan Gertheiss
Ivan Gesteira Costa Filho
Giorgio Gonnella
Assaf Gottlieb
Philip Groth
Andreas Gruber
Mostafa Herajy
Silvia Heyde
Christian Hoener zu Siederdissen
Katharina Hoff
Steve Hoffmann
Franziska Hufsky
Rene Hussong
Bo Jiang
Marc Johannes
Klaus Jung
Crystal Kahn
Chen Keasar
Andreas Keller
Ilona Kifer
Hisanori Kiryu
Renzo Kottmann
Frank Kramer
Roland Krause
Sita Lange
Glenn Lawyer
Andreas Leha
Limor Leibovich
Xiaoli Li
Steve Lianoglou
Yao-Cheng Lin
Fei Liu
Ronny Lorenz
Claudio Lottaz
Claus Lundegaard
Fälth Maria
Alberto J. Martin
Daniel Maticzka
Andrey Mironov
Michael Molla
Simon Moxon
Arcady Mushegian
Mathias Möhl
Liat Perlman
Thomas N Petersen
Conrad Plake
Jaime Prilusky
Nan Qiao
Stephane Rombauts
Ovidiu Radulescu
Fidel Ramirez
Franck Rapaport
Florian Rasche
Elisabeth Remy
Anna Ritz
Christian Rohr
Patrick Ruch
Susanna Röblitz
Martin Schwarick
Michael Sammeth
Kerstin Scheubert
Ruben Schilling
Roland Schwarz
Donny Soh
Artem Sokolov
Eric Solis
Oliver Stegle
Israel Steinfeld
Christine Steinhoff
Gautier Stoll
Eric Tannier
Léon-Charles Tranchevent
Anke Truss
George Tsatsaronis
Charalampos Tsourakakis
Koji Tsuda
Andrei Turinsky
Fabio Vandin
Ian Walsh
Martin Welk
Christian Widmer
Sebastian Will
Rainer Winnenburg
Shanshan Zhu
Elena Zotenko

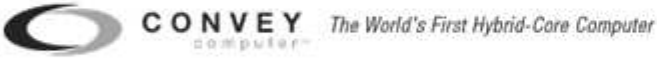
Sponsors

Platinum

Fellowships

Gold

Silver



Exhibitors

Other Sponsors

Platinum Sponsor	
IBM	http://www.ibm.com/be/en/

Fellowships	
ISCB	http://www.iscb.org/

Gold Sponsors	
NBIC	http://www.nbic.nl/

Silver Sponsors	
Convey Computer	http://www.conveycomputer.com/
myGrid	http://www.mygrid.org.uk/
CLC bio A/S	http://www.clcbio.com/
ENFIN	http://www.enfin.org/page.php
EBI	http://www.ebi.ac.uk/
ELIXIR	http://www.elixir-europe.org/page.php
Illumina	http://www.illumina.com/
Omixon	http://www.omixon.com/omixon/

Other Sponsors

FNRS	http://www2.frs-fnrs.be/
FWO	http://www.fwo.be/
Keygene	http://www.keygene.com/home/index.php
EMBO	http://www.embo.org/
Cancer Research	http://cancerres.aacrjournals.org/
BioMedCentral	http://www.biomedcentral.com/
Tibco	http://www.tibco.com
Microsoft	http://www.microsoft.com
BioSCENTer	http://www.kuleuven.be/bioscenter/
K.U.Leuven	http://www.kuleuven.be/
belspo IAP	http://www.belspo.com
BioMAGNet	http://www.kuleuven.be/biomagnet

Cambridge University Press
Cambridge University Press advances learning, knowledge and research worldwide. It is an integral part of the University of Cambridge and for centuries has extended its research and teaching activities by making available worldwide through its printing and publishing a remarkable range of academic and educational books, journals, and examination papers. For millions of people around the globe, the publications of the Press represent their only real link with the University of Cambridge. http://www.cambridge.org

Oxford University Press
Oxford University Press publishes some of the most prestigious books and journals in the world, including *Bioinformatics*, an official journal of the International Society for Computational Biology, *Nucleic Acids Research*, and our new fully open access journal, *Database: The Journal of Biological Databases and Curation*. Visit our stand to browse books and pick up journal sample copies. Visit us online: www.oxfordjournals.org

Chapman & Hall/CRC - Taylor & Francis group publishes books, journals and electronic databases. Check out new titles in our mathematical and computational biology series (including Golan Yona's "Introduction to Computational Proteomics") and key titles from our Garland list, including Understanding Bioinformatics and Molecular Biology of the Cell. Take advantage of 20% (50% off selected titles) convention discounts.

Applied Maths develops cutting-edge software for the biosciences. The software BioNumerics® is a software suite for integrated analysis and databasing of biological data in the broadest sense. When linked to Kodon®, a high level sequence analysis package or GeneMathsXT®, a highly sophisticated mathematical toolbox for the analysis of data from genechips and genearrays, it becomes a universal, powerful platform for bio-databasing and analysis. The software is licensed to more than 1.000 research sites worldwide and served for the preparation of >1.400 peer-reviewed papers.

The Springer Computer Science program serves research and academic communities around the globe. An impressive 6,000+ volumes published in the Lecture Notes in Computer Science (LNCS) contribute to 7,900+ eBooks, comprehensive reference works and 100+ journals in the Computer Science collection. Stop by the Springer booth to get acquainted with our multi-format publishing model: print – eBook - MyCopy.

Korilog is a bioinformatics company providing next generation of graphical softwares to integrate and explore pathways, genomes and proteomes data. The company provides academic and industry research labs with desktop and web-based softwares (koriblast.com; koriviewer.com) as well as expert software development services to support researchers in their R&D activities.

SciEngines provides innovative and efficient high-performance computers based on reconfigurable processors. For computational biology, data-center computing performance is achieved at a fraction of the cost - enabling new levels of analysis for modern science.

KEYNOTES

Systemic analysis of biological systems: lessons from a small single bacterium and a large complex community

Peer Bork

European Molecular Laboratory Lab, Heidelberg, Germany

Peer Bork, Ph.D., is senior group leader and joint head of the Structural and Computational Biology unit at EMBL, a European research organization with headquarters in Heidelberg. He also holds an appointment at the Max-Delbrueck-Center for Molecular Medicine in Berlin. Dr. Bork received his Ph.D. in Biochemistry (1990) and his habilitation in Theoretical Biophysics (1995). He works in various areas of computational biology and systems analysis with focus on function prediction, comparative analysis and data integration. He has published more than 400 research articles in international, peer-reviewed journals, among them more than 45 in Nature, Science and Cell. According to ISI (analyzing the last 10 years), Dr. Bork is currently the most cited European researcher in Molecular Biology and Genetics. He is on the editorial board of a number of journals including Science and PloS Biology, and functions as senior editor of the journal Molecular Systems Biology.
Dr. Bork co-founded four biotech companies, two of which went public. More than 25 of his former associates now hold professorships or other group leader positions in prominent institutions all over the world.
He received the "Nature award for creative mentoring" for his achievements in nurturing young scientists and was the recipient of the prestigious "Royal Society and Academie des Sciences Microsoft Award" for the advancement of science using computational methods.

Modeling protein structure, function, and interactions

Michael J. E. Sternberg

Imperial College, London, UK

Professor **Michael Sternberg** is the Director of the Centre for Integrative Systems Biology at Imperial College (CISBIC) and Centre for Bioinformatics (CfB) at Imperial College London and he holds the Chair of Structural Bioinformatics at Imperial. Michael Sternberg's research interest are protein bioinformatics and the development of logic-based chemoinformatics. His group have developed the Phyre server for protein structure prediction, 3D-Garden for protein-protein docking, and Confunc/ 3DLigand Site for protein function prediction. Recent work has developed methods to analyse the interactome and pathways. The chemoinformatics modelling employs a logic-based approach and is able to learn rules relating structure to activity and then use these rules to identify novel active molecules.

Analysis of whole-genome sequencing data from paired tumor and normal genomes

Elaine Mardis

Washington University Genome Center, St-Louis, MO, USA

Dr. **Elaine Mardis** graduated Phi Beta Kappa from the University of Oklahoma with a B.S. degree in zoology. She then went on to complete her Ph.D. in Chemistry and Biochemistry in 1989, also at Oklahoma. Following graduation, Dr. Mardis was a senior research scientist for four years at BioRad Laboratories in Hercules, CA.
In 1993, Dr. Mardis joined The Genome Center at Washington University School of Medicine. As Director of Technology Development, she helped create methods and automation pipelines for sequencing the Human Genome. She currently orchestrates the Center's efforts to explore next generation sequencing technologies and to transition them into production sequencing capabilities.

Dr. Mardis has research interests in the application of DNA sequencing to characterize cancer genomes. She also is interested in facilitating the translation of basic science discoveries about human disease into the clinical setting.
Dr. Mardis serves on several NIH study sections, is an editorial board member of Genome Research, and acts as a reviewer for Nature and Genome Research. She serves as chair of the Basic and Translational Sciences for the American College of Surgeons Oncology Group. Dr. Mardis recently received the Scripps Translational Research award for her work on cancer genomics.

Computing the flows of life

Hans Westerhoff

University of Manchester, UK and VU University of Amsterdam, The Netherlands

Hans Westerhoff is Professor of Systems Biology at Manchester University and also Professor of Microbial Physiology (Free University Amsterdam, VUA) and Professor of Mathematical Biochemistry (University of Amsterdam, UvA) at the BioCentrum Amsterdam. He heads a transnational research group on Systems Biology which spans the Manchester Centre for Integrative Systems Biology (MCISB) in the Manchester Interdisciplinary BioCentre (MIB) and the BioCentrum Amsterdam (see also http://www.bio.vu.nl/hwconf). His research interest focuses on how the interactions of macromolecules can lead to biological functioning, and integrates quantitative experimentation with mathematical analyses.

The evolutionary significance of ancient whole genome duplications and computational approaches to unveiling them

Yves Van de Peer

University of Ghent and Flemish Institute of Biotechnology, Belgium

Yves Van de Peer is professor in Bioinformatics and Genome biology in the Department of Plant Systems Biology at Ghent University, Belgium. He is leading a bioinformatics group of about 35 people. Yves Van de Peer has published about 250 papers in peer-reviewed journals and his group is a center of excellence in the fields of gene prediction and genome annotation, comparative genomics, and (top-down) systems biology.

SCHEDULE

Sunday, September 26 – Workshops and Tutorials							
8:30 AM	Workshop and tutorial registration & welcome coffee						
9:00 AM	**Tutorial 1 (Vd Goes)** Working with next-generation sequencing data	**Tutorial 2 (Window 9)** Use of semantic web resources in computational biology and bioinformatics	**Tutorial 3 (Window 16)** Current methods and applications for regulatory sequence analysis	**Tutorial 4 (Window5)** Protein structure validation	**Workshop 1 (Window 6)** Learning from perturbation effects	**Workshop 2 AIMM (Window 4)** Annotation, interpretation and management of mutations	**Workshop 3 ESCS1 (Bauwens)** Student symposium
10:30 AM	Coffee break						
11:00 AM	**Tutorial 1**	**Tutorial 2**	**Tutorial 3**	**Tutorial 4**	**Workshop 1**	**Workshop 2 AIMM**	**Workshop 3 ESCS1**
12:30 PM	Lunch						
1:30 PM	**Tutorial 1**	**Tutorial 2**	**Tutorial 3**	**Tutorial 4**	**Workshop 1**	**Workshop 2 AIMM**	**Workshop 3 ESCS1**
3:00 PM	Coffee break						
3:30 PM	**Tutorial 1**	**Tutorial 2**	**Tutorial 3**	**Tutorial 4**	**Workshop 1**	**Workshop 2 AIMM**	**Workshop 3 ESCS1**
5 PM	End of workshops and tutorials						

Sunday, September 26 – Conference Opening	
4:00 PM	Conference registration opens
5:30 PM	Conference opening and welcome
5:40 PM	**Keynote 1 - *Peer Bork.*** Systemic analysis of biological systems: lessons from a small single bacterium and a large complex community
6:30 PM	**PT1 - Utopia documents: linking scholarly literature with research data.** *Teresa K. Attwood, Douglas B. Kell, Philip McDermott, James Marsh, Steve Pettifer and David Thorne*
6:50 PM	**PT2 - Integrating genome assemblies with MAIA.** *Jurgen Nijkamp, Wynand Winterbach, Marcel van den Broek, Jean-Marc Daran, Marcel Reinders and Dick de Ridder*
7:10 PM	**PT3 - A graphical method for reducing and relating models in systems biology.** *Steven Gay, Sylvain Soliman and François Fages*
7:30 PM	Walk to the reception venue
8:00 PM	Welcome reception
9:30 PM	End of welcome reception

Monday, September 27		
8:30 AM	Registration and welcome breakfast	
9:00 AM	**Keynote 2 - *Michael J. E. Sternberg.*** Modeling protein structure, function, and interactions	
	PROCEEDINGS TRACK (Auditorium)	***TECH TRACK (Vd Goes)***
9:50 AM	**PT4 - Vorescore - Fold recognition improved by rescoring of protein structure models.** *Gergely Csaba and Ralf Zimmer*	**TT1 - Software for the data-driven researcher of the future**. *Paul Fisher et al.*
10:10 AM	**PT5 - Solenoid and non-solenoid protein recognition using stationary wavelet packet transform.** *An Vo, Nha Nguyen and Heng Huang*	
10:30 AM	**PT6 - Discriminatory power of RNA family models.** *Christian Hoener zu Siederdissen and Ivo L. Hofacker*	**TT2 - Clinical genomic analysis at IBM: from HIV positive to hypertension**. *Michael Rosen-Zvi et al.*
10:50 AM	**PT7 - RactIP: fast and accurate prediction of RNA-RNA interaction using integer programming.** *Yuki Kato, Kengo Sato, Michiaki Hamada, Yoshihide Watanabe, Kiyoshi Asai and Tatsuya Akutsu*	
11:10 AM	Coffee break	
	PROCEEDINGS TRACK (Auditorium)	***TECH TRACK (Vd Goes)***
11:40 AM	**PT8 - Semi-supervised multi-task learning for predicting interactions between HIV-1 and human proteins.** *Yanjun Qi, Oznur Tastan, Jaime G. Carbonell, Judith Klein-Seetharaman and Jason Weston*	**TT3 - The Microsoft Biology Foundation**. *Simon Mercer et al.*
12:00 PM	**PT9 - Candidate gene prioritization based on spatially mapped gene expression: an application to XLMR.** *Rosario M. Piro, Ivan Molineris, Ugo Ala, Paolo Provero and Ferdinando Di Cunto*	
12:20 PM	**PT10 - Prediction of a gene regulatory network linked to prostate cancer from gene expression, microRNA and clinical data.** *Eric Bonnet, Tom Michoel and Yves Van de Peer*	**TT4 - Study capturing: from research question to sample annotation.** *Kees van Bochove et al.*
12:40 PM	**PT11 - Inferring cancer subnetwork markers using density-constrained biclustering.** *Phuong Dao, Recep Colak, Raheleh Salari, Flavia Moser, Elai Davicioni, Alexander Schoenhuth and Martin Ester*	

1:00 PM	Lunch	
	PROCEEDINGS TRACK (Auditorium)	***TECH TRACK (Vd Goes)***
2:15 PM	**PT12 - Modeling associations between genetic markers using Bayesian networks.** *Edwin Villanueva and Carlos Dias Maciel*	**TT5 - NextGen biology with TIBCO Spotfire.** *Christof Gaenzler et al.*
2:35 PM	**PT13 - Mass spectrometry data processing using zero-crossing lines in multi-scale of Gaussian derivative wavelet.** *Nha Nguyen, Heng Huang, Soontorn Oraintara and An Vo*	
2:55 PM	**PT14 - Detecting host factors involved in virus infection by observing the clustering of infected cells in siRNA screening images.** *Apichat Suratanee, Ilka Wörz, Petr Matula, Anil Kumar, Lars Kaderali, Karl Rohr, Ralf Bartenschlager, Roland Eils and Rainer König*	**TT6 - Partitioning biological data with transitivity clustering.** *Jan Baumbach et al.*
3:15 PM	**Announcement of ISMB/ECCB 2011 and ECCB 2012**	
3:30 PM	Coffee break	
4:00 PM	**Keynote 3 – *Elaine Mardis***. Analysis of whole-genome sequencing data from paired tumor and normal genomes	
	PROCEEDINGS TRACK (Auditorium)	***TECH TRACK (Vd Goes)***
4:50 PM	**PT15 - Characteristics of 454 pyrosequencing data – enabling realistic simulation with Flowsim.** *Susanne Balzer, Ketil Malde, Anders Lanzén, Animesh Sharma and Inge Jonassen*	***ART MEETS SCIENCE*** **The Cosmopolitan Chicken Research Project.** *Koen Vanmechelen and Jean-Jacques Cassiman.*
5:10 PM	**PT16 - A Fast algorithm for exact sequence search in biological sequences using polyphase decomposition.** *Abhilash Srikantha, Ajit Bopardikar, Kalyan Kaipa, Venkataraman Parthasarathy, KyuSang Lee, TaeJin Ahn and Rangavittal Narayanan*	
5:30 PM	**PT17 - Classification of ncRNAs using position and size information in deep sequencing data.** *Florian Erhard and Ralf Zimmer*	
5:50 PM	Poster session (ODD numbers) & drinks	
8:00 PM	Visit city brewery Gruut and evening stroll through Ghent	

Tuesday, September 28		
8:30 AM	Registration and coffee	
	PROCEEDINGS TRACK (Auditorium)	***TECH TRACK (Vd Goes)***
9:50 AM	**PT18 - Prototypes of elementary functional loops unravel evolutionary connections between protein functions.** *Alexander Goncearenco and Igor Berezovsky*	
10:10 AM	**PT19 - Improved sequence-based prediction of disordered regions with multilayer fusion of multiple information sources.** *Marcin J. Mizianty, Wojciech Stach, Ke Chen, Kanaka Durga Kedarisetti, Fatemeh Miri Disfani and Lukasz Kurgan*	**TT7 - CLC bio, a comprehensive platform for NGS data analysis.** *Roald Forsberg*
10:30 AM	**PT20 - A predictor for toxin-like proteins exposes cell modulator candidates within viral genomes.** *Guy Naamati, Manor Askenazi and Michal Linial*	
10:50 AM	Coffee break	
	PROCEEDINGS TRACK (Auditorium)	***TECH TRACK (Vd Goes)***
11:20 AM	**PT21 - Discriminative and informative features for biomolecular text mining with ensemble feature selection.** *Sofie Van Landeghem, Thomas Abeel, Yvan Saeys and Yves Van de Peer*	**TT8 - DNA sequencing with Illumina instruments and chemistry: current and future applications.** *Dirk Evers*
11:40 AM	**PT22 - BioXSD: the common data-exchange format for everyday bioinformatics web services.** *Matúš Kalaš, Pål Puntervoll, Alexandre Joseph, Edita Bartaševičiūtė, Armin Töpfer, Prabakar Venkataraman, Steve Pettifer, Jan Christian Bryne, Jon Ison, Christophe Blanchet, Kristoffer Rapacki and Inge Jonassen*	
12 PM	**PT23 - Improving disease gene prioritization using semantic similarity of Gene Ontology terms.** *Andreas Schlicker, Thomas Lengauer and Mario Albrecht*	**TT9 - Super-scale sequence data analysis with hybrid core computing.** *John Leidel*
12:20 PM	**PT24 - Discovering drug-drug interactions: a text-mining and reasoning approach based on properties of drug metabolism.** *Luis Tari, Saadat Anwar, Shanshan Liang, James Cai and Chitta Baral*	

12:40 PM	Lunch	
	PROCEEDINGS TRACK (Auditorium)	***TECH TRACK (Vd Goes)***
2:10 PM	**PT25 - Nonlinear dimension reduction and clustering by Minimum Curvilinearity unfold neuropathic pain and tissue embryological classes.** *Carlo Vittorio Cannistraci, Timothy Ravasi. Franco Maria Montevecchi, Trey Ideker and Massimo Alessio*	**TT10 - How robust are NGS whole-genome assemblies? A case study with plant genomes.** *Niina Haiminen, Laxmi Parida*
2:30 PM	**PT26 - A varying threshold method for ChIP peak-calling using multiple sources of information.** *KuanBei Chen and Yu Zhang*	
2:50 PM	**PT27 - is-rSNP: a novel technique for in silico regulatory SNP detection.** *Geoff Macintyre, James Bailey, Izhak Haviv and Adam Kowalczyk*	
3:10 PM	Poster session (EVEN numbers) and coffee	
5:10 PM	Bus departures and ride to Bruges. Free time in Bruges	
8 PM	Banquet dinner	
10:30 PM	Bruges to Ghent. **Last bus: 12 AM**	

Wednesday, September 29		
8:30 AM	Registration and coffee	
9 AM	***Keynote 4 – Hans Westerhoff.*** Computing the flow of life	
	PROCEEDINGS TRACK (Auditorium)	***TECH TRACK (Vd Goes)***
9:50 AM	**PT28 - Efficient parameter search for qualitative models of regulatory networks using symbolic model checking.** *Gregory Batt, Michel Page, Irene Cantone, Gregor Goessler, Pedro Monteiro and Hidde de Jong*	**TT11 - Data integration in proteomics through EnVsion and EnCore webservices.** *Rafael Jimenez et al.*
10:10 AM	**PT29 - Bayesian experts in exploring reaction kinetics of transcription circuits.** *Ryo Yoshida, Masaya Saito, Hiromichi Nagao and Tomoyuki Higuchi*	
10:30 AM	**PT30 - Dynamic deterministic effects propagation networks: learning signalling pathways from longitudinal protein array data.** *Christian Bender, Frauke Henjes, Holger Fröhlich, Stefan Wiemann, Ulrike Korf and Tim Beißbarth*	**TT12 - ELIXIR: a sustainable European infrastructure for biological information.** *Andrew Lyall*
10:50 AM	**PT31 - A novel approach for determining environment-specific protein costs: the case of Arabidopsis thaliana.** *Max Sajitz-Hermstein and Zoran Nikoloski*	
11:10 AM	Coffee break	
	PROCEEDINGS TRACK (Auditorium)	***TECH TRACK (Vd Goes)***
11:40 AM	**PT32 - De-correlating expression in gene-set analysis.** *Dougu Nam*	**TT13 - The Protein Data Bank in Europe (PDBe) - bringing structure to biology.** *Wim Vranken et al.*
12 PM	**PT33 - How threshold behaviour affects the use of subgraphs for network comparison.** *Tiago Rito, Zi Wang, Charlotte Deane and Gesine Reinert*	
12:20 PM	**PT34 - Discovering graphical Granger causality using the truncating lasso penalty.** *Ali Shojaie and George Michailidis*	**TT14 - The Universal Protein Resource (UniProt).** *Maria J. Martin*
12:40 PM	Lunch	

2:00 PM	**PT35 - Parsimony and likelihood reconstruction of human segmental duplications.** *Crystal Kahn, Benjamin Raphael and Borislav Hristov*	**TT15 - Accurate next gen sequencing data analysis on cloud computing.** *Attila Bérces et al.*
2:20 PM	**PT36 - Maximum likelihood estimation of locus-specific mutation rates in Y-chromosome short tandem repeats.** *Osnat Ravid-amir and Saharon Rosset*	
2:40 PM	Coffee break	
3:10 PM	***Keynote 5 – Yves Van de Peer.*** The evolutionary significance of ancient whole genome duplications and computational approaches to unveiling them	
4:00 PM	Closing remarks and awards	
4:30 PM	***END OF CONFERENCE***	

SOCIAL EVENTS

Sunday

The opening reception in the historic Bijloke abbey starts at 8 PM. The Bijloke, originally donated by Countess Johanna van Constantinopel, is home to the Bijloke hospital and the Bijloke abbey. The hospital, founded in 1228 A.D., is a magnificent example of medieval hospital architecture. The unique wooden roof of the first patient hall is constructed from a complete oak forest coming from the Ardennes. The second patient hall, the Craeckhuys, built in the 16th century, is a unique example of 16th century hospital architecture. These days the site serves as a music centre for the city of Ghent.

There will be live music by Swingalicious, a young band of 5 swinging musicians who studied at the Conservatory of Ghent, brought together by singer Katrien Van Opstal. They will play a fine set of standards, mostly jazz. Known or unknown pieces, swing or bossa, they will make sure that you learn one thing: jazz is fun!

Monday

Brewery Gruut

City brewery **Gruut** is located in the very heart of Ghent, Belgium. In the Middle Ages the river Lys divided the city of Ghent into two parts. On the right bank the brewers used typical beer ingredients like hops. The brewers on the left bank, which was ruled by the French, added a mixture of herbs or Gruut to their beers. Landlords collected a tax from the brewers based on the amount of spice ("Gruuts") used in the beer. Our brewery combines present day brewing techniques with centuries old tradition. The beer is unique and especially healthy because of the use of special spices instead of hop.

We will visit the brewery (and taste the beer!) on Monday, leaving the conference site at 8 PM. There will be a limited number of free tickets available at the registration desk on Monday morning.

Evening Stroll through Ghent

For those who want to explore the historic center of Ghent and try some of the local beverages, we offer a scenic walk through the city center along many of Ghent's impressive medieval buildings. Teams can scout through the city center, assembling bits of a map to guide them to the next location. In typical Belgian pubs along the route, participants have the opportunity to sample some exclusive Belgian beers. Register for free on Monday.

Tuesday

Conference Dinner

The banquet dinner will take place in the magnificent city of Bruges. Only a short bus ride away from Ghent, the "Venice of the North" is a prominent World Heritage Site of UNESCO. There will be time for a stroll through the city before the reception and evening dinner.

Busses will leave the conference site at 5:10 PM. The banquet dinner starts at 8 PM in Oud Sint Jan, a XIXth century former hospital. Last busses to Ghent will leave Bruges at midnight.

WORKSHOPS

Workshop 1: Learning from perturbation effects

Reverse engineering of biological networks is a key to the understanding of biological systems and to identifying new drug targets. Perturbation techniques, like target specific inhibitors and RNA interference, open tremendous possibilities to detect so far unknown interdependencies from (possibly multidimensional) effects. At the same time computational methods being able to derive network hypotheses from such complex data play a crucial role. This workshop aims to bring together computational scientists working with various approaches for this challenging task. The goal is to give an overview about different computational methods in the field and to strengthen and initiate new cooperations.

Organizer

Prof. Dr. Holger Fröhlich, Bonn-Aachen International Center for IT (B-IT), Bonn, Germany
e-mail: frohlich@bit.uni-bonn.
http://www.abi.bit.uni-bonn.de

Program

Session 1: Models for RNAi-screening data

Niko Beerenwinkel (ETH Zurich (D-BSSE), Basel, Switzerland): *Computational approaches to reconstruct signaling networks of pathogen entry from RNAi screens*

Mirko Birbaumer (ETH Zurich, Zurich, Switzerland): *From vesicle features to cellular phenotypes: statistical clustering in image-based high-throughput RNAi screens*

Lars Kaderali (University of Heidelberg, Bioquant, Germany): *Reconstructing signaling pathways with probabilistic boolean threshold networks*

Coffee Break

Session 2: Models for static perturbation effects I

Florian Markowetz (Cancer Research UK Cambridge Research Institute, Cambridge, UK): *The end of the screen is the beginning of the experiment: pathway integration of hits in genome-wide RNAi screens*

Holger Fröhlich (University of Bonn, Bonn-Aachen International Center for IT (BIT), Germany): *Demo session: nested effects models at work*

Charles Vaske (Princeton University, USA): *Predicting and expanding biological pathways using a factor graph nested effects model*

Lunch

Session 3: Models for static perturbation effects II

Lan Zagar (University of Ljubljana, Slovenia): *Inference of epistasis from yeast genome-scale genetic interaction map*

Achim Tresch (Gene Center Munich, Ludwig-Maximilians University Munich, Germany): *Modeling combinatorial interventions in transcriptional networks*

Nicole Radde (University of Stuttgart, Germany): *A statistical Bayesian framework for identification of biological networks from perturbation experiments*

Marloes Maathuis (ETH Zurich, Switzerland): *Predicting perturbation effects in large-scale systems from observational data*

Coffee Break

Session 4: Models for dynamic perturbation effects

Rainer Spang (University of Regensburg, Germany): *Dynamic nested effects models*

Christian Bender (German Cancer Research Center, Heidelberg, Germany): *Dynamic deterministic effects propagation networks*

Sven Nelander (Sahlgrenska-CMR/ Wallenberg Laboratory and Gothenburg Mathematical Modeling Center, Sweden): *Human transcriptional networks revealed by endogenous perturbations in cancer tumors*

Workshop 2: Annotation, Interpretation and Management of Mutations (AIMM)

This workshop presents the state of the art in extraction and reuse of genotype-phenotype information. Annotation of mutations with their impact on phenotypic expression is crucial to understanding genetic mechanisms involved in phenotypic processes and ultimately in complex diseases. This knowledge is key to generating novel hypotheses. Despite the existence of literature and databases describing impacts of mutations, association studies fail to deliver linkage to phenotypes which is the most important contemporary research interest. Extraction of such information from scientific literature is a promising research field and existing solutions are ready to be deployed as services.

Organizers

Christopher J.O. Baker
Ph.D., Associate Professor / Innovatia Research Chair, Department of Computer Science and Applied Statistics, University of New Brunswick, Saint John, Canada.
e-mail: bakerc@unb.ca
Dietrich Rebholz-Schuhmann
MD, Ph.D., Research Group Leader, European Bioinformatics Institute, Welcome Trust Genome Campus, Hinxton, Cambridge, United Kingdom.
e-mail: Rebholz@ebi.ac.uk
René Witte
Dr.-Ing., Assistant Professor, Group Leader, Semantic Software Lab, Concordia University, Department of Computer Science and Software Engineering, Montreal, Canada.
E-mail: rwitte@cse.concordia.ca

Keynote Speakers

Michael Schroeder (BIOTEC Technical University Dresden, Germany): *Extraction and reuse of mutations and annotations from literature using ontologies*

Joost Schymkowitz (VIB Switch Laboratory, Vrije Universiteit Brussel, Belgium): *A knowledgebase for phenotyping of human SNPs and disease mutations*

Accepted Papers

Toward a richer representation of sequence variation in the sequence ontology
Michael Bada and Karen Eilbeck

Deploying the mutation impact mining pipeline with SADI: an exploratory case study
Alexandre Riazanov, Jonas Bergman Laurila and Christopher JO Baker

GenoS: segment-based representation of genomics data. Application to genotyping data management
Jean-pierre Kocher, Hugues Sicotte, Yaxiong Lin and Eric Klee

A semantic assistant for mutation mentions in PubMed abstracts
Jonas Bergman Laurila, Alexandre Kouznetsov, and Christopher JO Baker

Frankenstein Junior: a relational learning approach toward protein engineering
Elisa Cilia and Andrea Passerini

Improving the prediction of disease-related variants using protein three-dimensional structure
Emidio Capriotti and Russ B Altman

Characterization of pathogenic germline mutations in human protein kinases
Jose Maria Gonzalez-Izarzugaza Martinez, Lisa Hopcroft, Anja Baresic, Christine Orengo, Andrew Martin and Alfonso Valencia

Workshop 3: ESCS1—European Student Council Symposium

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.

Chairs

Magali Michaut, ESCS1 chair (Terrence Donnelly CCBR, University of Toronto, Ontario, Canada)
Thomas Abeel, ESCS1 co-chair (Department of Plant Systems Biology, VIB, Ghent University, Gent, Belgium and Broad Institute of MIT and Harvard, Cambridge, MA, USA)

Program

9:00 AM	Ice breaker
9:30 AM	Keynote
	"Generating hypotheses for molecular biology through (top-down) systems biology"
	Yves Van de Peer, VIB and University of Ghent, Belgium
10:30 AM	Coffee break
11:00 AM	Students: contributed presentations based on abstract submission
12:30 PM	Lunch and poster session
1:30 PM	Soft skills workshop
	"Presenting Science using the theatre approach"
	Gijs Meeusen
3:00 PM	Coffee break
3:30 PM	Soft skills workshop
	presentation skills continues
5:00 PM	Posters

TUTORIALS

The purpose of the tutorial program is to provide participants with lectures and instruction covering topics relevant to the bioinformatics field. It offers participants an opportunity to learn about new areas of bioinformatics research, to get an introduction to important established topics, or to develop advanced skills in areas about which they are already familiar.

Tutorial 1: Working with next-generation sequencing data

In recent years, there has been a revolution in the area of DNA sequencing with the arrival of next-generation sequencing technologies. The type and volume of the data produced by next-generation sequencing machines presents many previously unseen informatics challenges. This tutorial will help people who are getting started on next-generation sequencing get an idea of the tools, flows, and procedures that they may need to set up to handle this data. In this short course, we will introduce the participants to the different next-generation sequencing technologies, show how to do some basic quality checking of the data, how to run the various next-generation alignment tools, create de novo sequence assemblies, and call variants (such as SNPs, short indels, and structural variants) from a reference sequence.
All slides and data from examples will be available online. Prerequisites for this tutorial are an interest in genome sequencing and basic UNIX skills.

Presenters

Thomas Keane completed his Ph.D. degree in the area of distributed computing and high-throughput phylogenomics from NUI Maynooth (Ireland) in 2006. He subsequently moved to the Pathogen Genomics group at the Wellcome Trust Sanger Institute to work on sequence assembly of several pathogens such as Plasmodium falciparum strains, Trypanosoma brucei, and several other pathogen genomes. In 2008, he co-founded the Vertebrate Resequencing Informatics group and manages the sequencing, informatics, and variation pipelines for large projects such as the 1000 genomes and mouse genomes project.
Jan Aerts received his Ph.D. at Wageningen University (Netherlands) in 2005 on the subject of chicken genome mapping and sequencing. After a post-doc position at the Roslin Institute near Edinburgh in Scotland - working on the cow genome assembly - he now works at the Wellcome Trust Sanger Institute near Cambridge (UK). His current work includes downstream analysis of next-generation sequencing data in order to identify putative SNPs and indels. He will start an assistant-professorship at the University of Leuven in October.

Schedule

Slot 1 - Thomas

Overview of next-generation sequencing technologies
Applications of next-generation sequencing
Quality control measures and metrics of libraries/lanes
Data storage (file formats) and metadata

Slot 2 - Thomas

Introduction to short read alignment algorithms and tools
Practical instructions and examples on use of short read aligners
Parallelising short read alignment
Introduction to sequence assembly methods and tools
Practical instructions on use of short read assembly tools to get the optimal sequence assembly

Slot 3 - Jan

Overview of variation calling from next-generation sequence data
SNP calling theory and tools

Short indel calling theory and tools
Practical examples of variation calling
File formats for variation calling and storage

Slot 4 - Jan

Introduction to structural variation
Summary of different types of structural variants (large insertions, deletions, inversions, translocations, copy number variants)
Overview of algorithms and tools for calling structural variants
Practical examples of calling structural variants
Visualisation of structural variants

Tutorial 2: Use of Semantic Web resources in computational biology and bioinformatics

The Semantic Web is a set of technologies, or a framework, which is designed to make data integration possible via the web, with the addition of a precise semantic characterization of entities and relations (ontologies). As data integration is a pre-requisite for systems biology and translational research, the Semantic Web can bring relevant benefits in these areas. The aim of this tutorial is to briefly introduce the key basic principles needed to understand what it means to represent information on the Semantic Web, and then to provide the attendees with basic hands on competences to start using biomedical information resources which are now available on this framework.
Prerequisites for this tutorial are knowledge of main biology databases; programming skills are a plus.
People with laptops and more proficient in programming will have the chance to explore a little bit further the exercises (we will leave some slightly more challenging exercises for the after course), however the course will be designed to be followed as a frontal presentation, at times very interactive.

Presenters

Paolo Romano, Ph.D., Bioengineer, is a Senior Scientist in bioinformatics at the National Cancer Research Institute of Genoa, Italy. His research interests include biomedical data management, network standards and tools, data integration, interoperability, ontologies, semantics methods and tools. He designed and contributed to the development of the Cell Line Database and its associated hypertext HyperCLDB, and he designed and developed Biowep, the Workflow Enactment Portal for Bioinformatics.
Andrea Splendiani, Ph.D., is a Senior Scientist in data integration at Rothamsted Research (BBSRC), Harpenden, UK. He has previous experience in microarray databases design and standards (University of Milano-Bicocca and Genopolis consortium, Italy), in Systems Biology (Institut Pasteur, France) and in Medical Informatics (University of Rennes1 – now ISERM U936, France). His research interests include standardization of biological data and in particular pathway information, biological data integration and the development of interactive systems to access and analyse biological information.

Schedule

Morning session

Introduction to basic principles of Semantic Web based representation of biological information (1h, theory). In this hour we will introduce the basics of RDF, and focus on the importance of URIs, shared relations and the implication of the open world assumption. We will invite the participants to think at RDF in terms of of a conceptual model, rather than XML, where is important to be precise and "what" we are referring to and on which predicates we use. We will show that the Semantic Web is easy.

Introduction to relevant biomedical resources, including UniProt, Bioportal, Pathwaycommons (1h, theory). In this hour we introduce briefly a few main biomedical resources which are available on the Semantic Web. The objective is to contextualise what we have presented before to the biomedical domain, and to provide some resources which will be used in the following examples and exercises. At the same time, given the widespread use of the resources above, it is likely that this introduction will

enable participants to quickly adopt them via the Semantic Web in their daily work. (Note: this lesson can easily accommodate a break as distinct resources are presented, which don't require a continuous flow of attention).

The remainder of the tutorial will introduce technologies through a very simple hands-on use case which will guide the participants through exporting data on the Semantic Web, integrating this data with other existing biomedical resources and querying the resulting integrated (distributed) knowledge-base.

Introduction to the simple examples and explanation of how several technologies fit together (15')
Introduction to the D2RQ relational mapping system (45'). We will show how to generate an automatic mapping between a relational database and RDF via the D2RQ tool. We will then show how to tune this mapping to better represent URIs and relations, and then we will show how to open a SPARQL endpoint via D2RQ.

Afternoon session

Practical example on how to use D2RQ (1h). Participants with laptops or more proficient in programming will be provided an extended example and invited to practice hands on, while we will present a subset of this example in detail to the audience, making this section very interactive by inviting participants to propose how they would conceptually map their information to RDF (which will help them to refine their understanding of this language).

Triplestores (30'). We will introduce other triplestores, which can complement D2RQ in a production environment. We will briefly introduce their pros and cons and provide tips on how to install and operate them.

Introduction to the SPARQL query language (45'). How to query D2RQ or any triplestore providing a SPARQL endpoint: examples of the most common and useful SPARQL constructs.

Practical example on performing queries with SPARQL. We will begin by presenting a list of queries of increasing complexity, which will involve the mapping realized in the previous example (via D2RQ) and the biomedical resources presented in the morning. We will explain how to express these queries in SPARQL, and we will show how the proper definition of URIs for entities and relations is the basis for data integration, which is now automatically realized at query time. This will refer to what we have presented in the introduction and the participants have practiced in the previous session.

Tutorial 3: Current methods and applications for regulatory sequence analysis

The annotation of the non-coding genome with gene regulatory function is lagging far behind the annotation of protein-coding genes and improved annotation will depend both on deeper biological insight into cis-regulatory logic and on more efficient computational prediction algorithms. Recent data obtained by high-throughput experiments accelerate the genome-wide identification of regulatory elements but also provide additional bioinformatics challenges. In the light of these developments, this tutorial will focus on bioinformatics methods to predict cis-regulatory elements and to aid the process of regulatory annotation. Participants will be provided with an overview of existing resources (databases, tools) and methods for detecting cis-regulatory elements in genome sequences, and generate testable hypotheses about the binding specificity of transcription factors (motifs discovery), their precise binding locations (binding site prediction), and their target genes (target identification), and go towards regulatory networks.

A list of software will be provided in advance, in order to allow participants to install it on their laptop. Some knowledge of the Unix interface is required.

Presenters

Jacques van Helden, Ph.D., is heading the Laboratory of Genome and network Bioinformatics (BiGRe - ULB). The BiGRe laboratory is specialized in the development, evaluation and application of bioinformatics approaches to analyze genome regulation, biomolecular networks, metabolic pathways and mobile genetic elements. Since 1997, Jacques van Helden has been developing a suite of specialized software tools for the analysis of genome regulation, the Regulatory Sequence Analysis Tools (RSAT), which integrates a variety of algorithms for motif discovery, sequence scanning for motifs, phylogenetic footprinting, analysis of ChIP-seq data, etc.

Stein Aerts, Ph.D., is heading the Laboratory of Computational Biology at the University of Leuven. The lab focuses on computational identification of cis-regulatory elements, on mapping transcriptional networks, on "omics" data integration, and on next-generation sequencing (NGS) data analysis. Stein Aerts is the developer of the TOUCAN software for regulatory sequence analysis, of several other motif and module discovery algorithms and text-mining applications, of the ENDEAVOUR application for gene prioritization, and of the cisTargetX method for transcriptional target identification in Drosophila.

Schedule

- Retrieving regulatory sequences from UCSC, EnsEMBL, RSAT, Toucan,....
  Upstream sequences, introns, repeat masking, multi-genome sequence retrieval,...
  Web-server, Web service or command-line access to resources
- From binding sites to binding motifs
  - Public databases about transcriptional regulation : motifs, sites and regulatory regions.
    JASPAR, PAZAR, ORegAnno, FlyReg
  - Regular expressions
  - Position-specific scoring matrices (PSSM)
    Concepts: counts, frequencies, pseudo-counts, weights, information content
  - PSSM formats and their inter-conversions
- Pattern matching: predicting binding sites
  - String-based pattern matching
  - Matrix-based pattern matching
    - matching scores
    - estimation of the risks (P-value distribution)
  - Cis-regulatory modules (clusters of motifs)
    - homotypic
    - heterotypic
- Pattern discovery: methods and applications
  - String-based pattern discovery
  - PSSM-based pattern discovery
  - PSSM enrichment detection
  - Combinations of PSSMs
- Phylogenetic footprinting
  - detecting conserved binding sites by comparative genomics
- Applications
  - Clusters of co-expressed genes
  - Motif discovery using high-throughput data (ChIP-seq, ...)
  - Genome-wide target prediction
  - Gene regulatory networks
- Evaluation of prediction performances
  - Benchmark datasets
  - Strength and pitfalls of evaluation: sensitivity, positive predictive value, accuracy, ROC curves,...
  - Method comparisons
- Browsing genomes for cis-regulatory elements
  - Visualization
  - Integration of predictions and annotations

Tutorial 4: Protein structure validation

Recent advances in homology modeling and drug design have made clear that the quality of protein structures is really important for good results. Today's software and CPU time availability on clusters, super computers, and the grid allow for easy improvement of old files. It is therefore important for all protein structure bioinformaticians to be able to evaluate the quality of the structures used in their studies and to know when it might be beneficial to ask an NMR spectroscopist or X-ray crystallographer to help bring an old file up to

today's standards. This workshop will teach the participants how to use a series of protein structure validation tools and how to interpret the results. The origins of problems are discussed as well as their importance for follow up studies.
In silico experience with protein structures is required. We will assume that the participants of this workshop have spent at least a hundred hours studying protein structures with molecular visualisation software. Extensive knowledge of any particular visualisation program, however, is not required. Candidates who cannot bring their own portable computer are requested to contact the organisers as only a limited number of PCs is available.

Presenters

Robbie Joosten recently completed his Ph.D. at the CMBI with as topic "Protein Structure Validation." His PhD project consisted of two major topics. The first topic was the detection of errors in PDB files. The second topic was rerefinement of all X-ray PDB files. After obtaining his Ph.D., Robbie moved to the X-ray department of the Dutch Cancer Institute (the NKI in Amsterdam) where he continues with the further improvement of (all) X-ray PDB files.
Jurgen Doreleijers got his Ph.D. in the group of Rob Kapteyn in Utrecht with as topic NMR protein structure quality. After his Ph.D. defense Jurgen moved to the BMRB in Wisconsin, USA. Here he worked on the remediation of the experimental NMR data that has been deposited with the NMR structure ensembles in the PDB. Jurgen recently moved to the CMBI where he works in the group of Geerten Vuister on the CING software that evaluates NMR structures.
Hanka Venselaar is a Ph.D. student at the CMBI. Her research project (automation of the prediction of the phenotypic effects of protein point mutations). This project includes extensive studies of protein structures and often involves a multitude of protein structure quality aspects.
Gert Vriend is one of the pioneers in the field of protein structure quality determination. His WHAT_CHECK software is for many years already the de facto standard in this field. He has published dozens of articles in this field, and he has been the Ph.D. supervisor of a series of Ph.D. students who worked on the structure validation of X-ray or NMR files.

Schedule

Seminars
Gert Vriend - Introduction
Robbie Joosten - X-ray structure evaluation
Jurgen Doreleijers - NMR structure evaluation

Practicals
Hanka Venselaar - Using the YASARA twinset for protein structure evaluation
Gert Vriend - Reading WHAT_CHECK reports

Parallel sessions
Jurgen Doreleijers - Using CING to evaluate protein structures
Robbie Joosten - Errors in X-ray structures

ART MEETS SCIENCE

ECCB has a tradition of organizing 'Art Meets Science' events. This year's event is a joint lecture by Koen Vanmechelen and Jean-Jacques Cassiman, an artist and a geneticist.

Belgian artist **Koen Vanmechelen** makes iconoclastic sculpture, paintings, glasswork and installations. His daring expeditions into contemporary science, philosophy and ethics have resulted into several internationally acclaimed projects. Best known is The Cosmopolitan Chicken Project (TCCP), a vast attempt to create and manipulate scores of chicken breeds from all over the world into a new species, a universal chicken or Superbastard. Four subprojects constitute the TCCP: Virtual crossing, Experimental crossing, The Walking Egg, and The Accident, Chronicles of The Cosmopolitan Chicken.
Prof. **Jean-Jacques Cassiman** is a human geneticist, renowned for pioneering genetic testing and forensic medicine in Belgium and for genetic research on cystic fibrosis. From 1993 to 1999, he was secretary general of the European Society of Human Genetics (ESHG). In 1998, he demonstrated through DNA testing that Karl Wilhelm Naundorff was not a descendant of the Bourbons and certainly not Louis XVII. In 2004, he demonstrated the heart that had been kept in Paris belonged to Louis XVII. For several decades, he has been a key figure in communicating genetic research to the general public and the media in Flanders.
Together they have initiated the *Cosmopolitan Chicken Research Project*, a genetic research project that ties art and science together by studying the genetic diversity of chicken breeds and Cosmopolitan Chicken hybrids.

Website

http://www.koenvanmechelen.be

TECHNOLOGY TRACK

TT1—Software for the data-driven researcher of the future

Alan Williams[2], Aleksandra Nenadic[2], Shoaib Sufi[2], Danius Michaelides[2], David Withers[2], Don Cruickshank[2], Franck Tanoh[2], Ian Dunlop[2], Jiten Bhagat[2], Katy Wolstencroft[2], Paolo Missier[2], Sergejs Aleksejevs[2], Stian Soiland-Reyes[2], Stuart Owen[2], Peter Li[2], Finn Bacell[2], Mannie Taggs[2], Rishi Ramgolam[2], Marco Roos[2], Eric Nzuobontane[2], Thomas Laurent[2], David De Roure[2], Robert Stevens[2], Steve Pettifer[2], Rodrigo Lopez[2],Carole Goble[2]

The myGrid project have developed a suite of open source software tools for the registration and discovery of Life Science Web Services (BioCatalogue), and the designing, execution, and sharing of scientific workflows (Taverna and myExperiment) for analytics and data processing.

The Taverna workflow workbench provides an environment in which scientists design and execute workflows, combining local and public distributed services and data resources into a single experimental protocol. The latest Taverna 2.2.0 release features: a command line tool for running workflows without the need to interact directly with a user-interface; access to the BioCatalogue of Life Science Web Services via a plug-in interface; and a Taverna Server for running workflows on a server as well as on the desktop.

myExperiment is a social networking site that provides a collaborative environment where scientists can safely publish their workflows, experiment plans, and standard operating procedures (SOPs). This allows researchers to share them with individuals, groups, and even discover those of others that can be subsequently re-used. myExperiment makes it easy for the next generation of scientists to contribute to a pool of scientific methods, build communities and form relationships – reducing time-to-experiment, sharing expertise and avoiding reinvention.

The BioCatalogue is a community driven registry of Life Science Web Services. It provides an open platform for Web Services registration, annotation and monitoring, with a comprehensive REST API. Moreover, BioCatalogue is a platform for Web Service providers to publish and advertise their services, providing a centralised and curated catalogue of Web Services, and to build a collaborative environment where the community can find, contact and meet the experts and maintainers of these services.

Taverna, BioCatalogue and myExperiment together create an integrated solution for bioinformatics, data analysis and analytics. We show how these tools together address the challenges of large scale data processing that comes with Next Generation Sequencing.

URL

http://www.mygrid.org.uk

Speaker

Dr. Paul Fisher[1]

Author affiliations

[1]University of Manchester [2]University of Manchester, University of Southampton, University of Oxford, EMBL-EBI, University of Leiden

TT2—Clinical genomic analysis at IBM: from HIV positive to hypertension

Ehud Aharoni[1], Hani Neuvirth[1], Noam Slonim[1], EuResist GEIE partners[2], Hypergenes partners[3]

The domain of personalized health and genome-based therapy has flourished in recent years due to the significant reduction of information storage devices, the reduction in care delivery costs, and improved clinical outcomes. Building on IBM's leadership in areas like systems integrations, cloud computing, massive scale

analytics and even emerging areas of science like nanomedicine, we focus on creating technologies and processes that will build an evidence-centric healthcare ecosystem.

IBM has defined four main areas of research: evidence generation, which uses scientific methods to turn raw health data into proof of effective treatment methods; the ability to deliver evidence in a context-dependant and personalized way at the point of care; improving service quality through simplifying the complex healthcare delivery process; and incentives and models to shift the healthcare industry to a system that rewards based on outcomes and healthier patients rather than only treatment and volume of care.

In recent years, IBM Research in Haifa, Israel has been involved in research related to evidence generation, with a focus on using the massive volumes of clinical and genomic data arriving from different hospitals to find new ways of optimizing patient treatment. Data mining and machine learning techniques were also applied to decipher the relationship between clinical status and genomic variations, with the goal of helping improve diagnostics and treatment.

From 2006 to 2008, IBM Research and the EuResist GEIE consortium developed a drug-interaction modeling tool that lets users predict the success rate of various drug combinations and their impact on virus evolution via an online portal. A set of prediction engines that leverage medical data (for example, viral gene sequences, patient histories) from seven sources are available in the portal and recommend which therapies are expected to be most efficient given viral, genomic, and other clinical and demographic measures. This engine predicts patient response to therapy with 78% accuracy, outperforming other common tools. Since then more data has been contributed and the algorithms were updated based on the new information. In 2010, IBM and the GEIE partners updated the recommendation to include new drugs.

IBM researchers are now working with the European HYPERGENES consortium to identify the genetic variations responsible for hypertension and associated organ damage. The team is working with clinical data accumulated from hypertensive and healthy people and genomic biomarkers provided by Illumina's 1Million SNP chip. The aim is to create a comprehensive genetic-epidemiological model that takes into account how genomics and other factors help improve diagnostic accuracy and introduce new strategies for early detection, prevention, and therapy. This effort will help create more personalized treatment plans for individuals that suffer from hypertension.

URL

http://engine.euresist.org/ and *http://srv-rimon.haifa.il.ibm.com:8080/rimon_web/snpWeights.jsp*

Speaker

Dr. Michal Rosen-Zvi, manager[1]

Author affiliations

[1] machine learning and data mining group, IBM Haifa Research Lab [2] EuResist GEIE [3] Hypergenes

TT3—The Microsoft Biology Foundation

Michael Zyskowski, Research Program Manager[1]; Bob Davidson, Principal Software Architect[1]

The aim of the Microsoft Biology Foundation (MBF) project is to produce a well-architected and comprehensively-documented library of common functionality related to bioinformatics and genomics, with the intention of making it easier to write life science applications on the Windows platform. Using C# and the .NET 4.0 framework provides additional levels of flexibility for the developer – over 70 .NET programming languages are compatible, from Visual Basic and Python to C++ and F#. It also leverages the power of .NET – over 15,000 pre-written functions - and takes advantage of .NET Parallel Extensions, a new feature which can parallelize algorithms across all cores and processors of the local machine.

This demonstration will include a brief tour of the MBF library, including details of its free, open source, community-curated and community-owned philosophy and how scientists and developers can participate in future development. We will also demonstrate the flexibility and usability of the library through a range of applications, including a DNA sequence assembler using the Windows Presentation Foundation, an add-in for

Microsoft Excel integrating bioinformatics functionality directly with the spreadsheet, access to webservices including demonstration of the Microsoft cloud computing solution Azure, and integration with HPC and scientific workflows.

URL

http://mbf.codeplex.com

Speaker

Simon Mercer, Ph.D. Director of Health and Wellbeing[1]

Author affiliations

[1] Microsoft Research

TT4—Study capturing: from research question to sample annotation

Machiel Jansen [2], Jeroen Wesbeek[1], Tjeerd Abma[3], Jahn-Takeshi Saito[4], Adem Bilican[4], Michael van Vliet[5], Prasad Gajula[6], Vincent Ludden[1], Robert Horlings[1], Siemen Sikkema[7], Margriet Hendriks[3], Chris Evelo[4], Ben van Ommen[1], Jildau Bouwman[1]

The demonstration of the software tool will focus mainly on the part about capturing biological studies. During the demonstration, it will become clear how we achieve the following goals:

- Facilitate standardization of the description of biological research studies
- Leverage existing standards and ontologies and integrate them into one user interface
- Cover complex study designs, such as double-blind crossover designs
- Provide an overview of the different studies done in e.g. a consortium, to encourage cooperation and data exchange
- Track samples that are used for omics assays back to the source, the circumstances under which and the subject from which they were taken
- Provide the first cornerstone needed to do 'multi-omics' data analysis over multipe different studies and multiple omics platforms
- Focus on user experience, through extensive testing with biologists

The demonstration will have the following outline:

- Walk the attendants through the process of adding a study to the study capture tool
- Demonstration of the connection of samples in the study capture tool to omics data
- Give an example of a query involving both study design data and omics data

URL

http://www.dbnp.org

Speaker

Kees van Bochove [1,2]

Author affiliations

[1] TNO Quality of Life, Zeist [2] NBIC BioAssist Engineering Team [3] UMCU Metabolomics Centre, UMC Utrecht [4] BiGCaT Department of Bioinformatics, Maastricht University [5] Leiden/Amsterdam Centre for Drug Research [6] Plant Research International, Wageningen University [7] Biosystems Data Analysis Group, Swammerdam Institute for Life Sciences, University of Amsterdam

TT5—NextGen biology with TIBCO Spotfire

Prof. Peter van der Spek[2], Dr. Andreas Kremer[2]

Spotfire helps you to integrate and analyse Microarray and other –omics data and align it with your LIMS and screening data. The complexity of such an analysis is rising with every new data format you are adding. The analyst needs powerful and flexible software tools to get results out of the combined data.

The TIBCO Spotfire software platform fulfils all important business needs of data analysis in modern bioinformatics. Every bioinformatician knows that collections of .CEL and .csv files are not sufficient for analyses as well as simple relational data bases are not. Even Data Warehouses and associated reporting tools are not sufficient to turn data into knowledge. Spotfire DecisionSite is well known for high quality ad-hoc analysis, but also this is not sufficient anymore. We will show how the new TIBCO Spotfire brings all pieces together, reduces the complexity of the analysis and is compliant with your company IT infrastructure. We will demonstrate how Spotfire and the bioinformatics department of the Erasmus Medical Center in Rotterdam built an interactive human brain atlas. This project shows the combination of genomics, proteomics and cytogenetic data with medical imaging. The aim is to identify genes associated with neurological symptoms and diseases (see ECCB2010 poster H13, A. Kremer).

URL

http://spotfire.tibco.com

Speaker

Dr. Christof Gaenzler, Senior Solution Consultant [1]

Author Affiliations

[1] TIBCO Spotfire, Germany [2] Erasmus Medical Center, Rotterdam

TT6—Partitioning biological data with Transitivity Clustering

Tobias Wittkop[1], Sita Lange[2], Dorothea Emig[4], Sven Rahmann[3], Jan Baumbach[4,5]

Partitioning biomedical data objects into groups, such that the objects in each group share common traits, is a long-standing challenge in computational biology. Here we present an integrated data clustering framework based on weighted transitive graph projection: Transitivity Clustering. We illustrate a typical, biomedical clustering task that starts with a list of amino acid sequences, investigates similarity functions and parameter estimation problems, and finally deals with an integrated result interpretation; all of which can be done easily with Transitivity Clustering, but with no other clustering software. We will further demonstrate how to use Transitivity Clustering to identify protein complexes in protein-protein interaction data. With Transitivity Clustering, we provide the user with easy-to-use interfaces that significantly ease and improve each step of the typical data clustering workflow: (1) A web interface for a quick analysis of medium-sized data sets, (2) a powerful stand-alone Java implementation for large-scale data clustering, and (3) a collection of Cytoscape plugins that also provide further methods to answer typical follow-up questions.

Reference: Wittkop T, Emig D, Lange S, Morris JH, Boecker S, Rahmann S, Albrecht M, Stoye J, Baumbach J (2010) *Partitioning biological data with Transitivity Clustering.* Nature Methods. 2010 Jun;7(6):419-20.

URL

http://transclust.cebitec.uni-bielefeld.de

Speaker

Dr. Jan Baumbach[1,2,3,4,5]

Author affiliations

[1] Buck Institute for Age Research, USA [2] Albert-Ludwigs-University, Freiburg, Germany [3] Technische Universität Dortmund, Germany [4] Max Planck Institute for Informatics, Germany [5] International Computer Science Institute, University of California at Berkeley, USA

TT7—CLC bio, a comprehensive platform for NGS data analysis

Not long ago high-throughput DNA sequencing (HTS) was reserved for a few dedicated genome centers that have dedicated and specialized bioinformatics staff and hardware for the analysis of sequencing data. But over the course of only a few years HTS has become broadly accessible to many non-specialized researchers and groups and the responsibility for analyzing very large HTS data sets may now rest upon researchers with a background in molecular biology or medicine and no formal computer training. For these

groups the establishment of an appropriate software and hardware platform for bioinformatics analysis presents a very hard and often insurmountable challenge.

Most bioinformatics software for analysis of HTS data exist as stand-alone command line tools that are inaccessible to non-specialists. This creates the need for a dedicated bioinformatician in the data analysis workflow to perform simple and tedious tasks such as simply establishing the analysis software infrastructure, executing commands and parsing textual or binary output. This is unfortunate since it makes the access to bioinformatics personnel a serious and poorly scalable bottleneck in the data analysis work flow and directs valuable bioinformatics brain resources away from creative and proactive tasks and towards tedious and non value creating routine tasks. Furthermore, it disempowers the biomedical researchers that requested the data and formulated the biological hypotheses by making them unable to directly inspect, handle and analyze the data themselves.

Another serious bottleneck for data analysis is the access to appropriate hardware. The output of a single run of a sequencing instrument is now at around 50 giga bases and still increasing. This places very tough demands on the hardware used in the analysis regarding CPU power and accessible memory. Again, the recent broad dissemination of HTS data means that very large data sets are now accessible to research group that have insufficient financial or personnel resources to establish a dedicated and powerful computational infrastructure. In order to give these groups access to data analysis, software must be designed that can operate on and maximize the utility of standard, economic and omnipresent computer solutions. In organizations where powerful central computing facilities can be installed and dedicated to HTS data analysis these can only be accessed through technically challenging interfaces that are inaccessible to most employees and require specialist intervention and a drain on the bioinformatics personnel resources.

To resolve these issues, we have developed and present the CLC bio integrated solution for analysis and data handling of high-throughput sequencing (HTS) data.

URL

http://www.clcbio.com

Speaker

Roald Forsberg, PhD[1]

Author affiliations

[1]Director of Scientific Software Solutions

TT8—DNA sequencing with Illumina instruments and chemistry: current and future applications

Klaus Maisinger[1], Anthony Cox[1], Lisa Murray[1], Come Raczy[1]

The session will be separated into three parts, starting with the current range of Illumina sequencers and the latest performance figures achieved in the production sequencing facilities at Illumina Cambridge in Chesterford Research Park, UK.

The data will serve as a basis to describe the Illumina software pipeline from RTA real time analysis on the instrument PC to CASAVA for mapping and genome variation detection via grid computing.

This will be followed by a characterization of the current state of the art in high-throughput sequencing workflows. Starting from typical computing and storage volumes we will extrapolate application parameters for future research scenarios in de novo assembly, whole genome and whole exome re-sequencing as well as metagenomics. Finally, we will describe bioinformatics scientist positions currently open at our R&D site in Chesterford near Cambridge, UK.

Speaker
Dirk J. Evers, Director Computational Biology[1]

Author affiliations
[1]Illumina Inc.

TT9—Super-scale sequence data analysis with Hybrid Core computing

It is clear that the latest generation of sequencing devices are a vast improvement over previous generations. However, the volume of data created by these new devices has significantly outpaced the ability to sufficiently analyze the data in a timely manner with traditional processors. As such, the only feasible methodology is the adoption of accelerator technologies. The current general purpose graphics processing methodologies provide interesting platforms for development, yet fail to deliver the sufficient order of speedup required to keep pace with the data explosion. Field programmable gate arrays (FPGAs) have typically performed quite well for bioinformatics applications. However, the difficulty in programming the soft cores and the difficulty in data management have historically failed to allow for widespread adoption.

Convey's Hybrid Core computing platform is an ideal platform for sequencing tasks in that it combines the traditional x86 economy of scale with an abstracted set of FPGAs for direct algorithmic synthesis coupled with a cache-coherent memory subsystem. The platform includes a significant set of compilers, tools and libraries such that the historical pain in programming FPGAs is fundamentally eliminated. In this way, Convey has the ability to implement very discretely managed algorithmic kernels (called *personalities*), which serve as logic and instruction set extensions to the x86 environment. In doing so, the Convey platform allows users and applications the ability to significantly parallelize and outperform traditional Von Neumann architectures.

Speaker
John Leidel[1]

Author affiliations
[1]Software Architect, Convey Computer Corp

TT10—How robust are NGS whole-genome assemblies? A case study with plant genomes

Laxmi Parida[1]

A decade after the human genome was sequenced, a large section of the computational/bioinformatics community believes the general assembly problem to be "solved", in spite of the recent changes in the underlying technologies. However, the "quality" of a draft genome of an organism continues to dog the community that studies the genomes closely with a fine-toothed comb for biological insights such as phenotypic associations with specific loci, strain-specific polymorphism detection and so on.

As genome assemblers come of age and their numbers grow, is it possible to consolidate multiple assemblies so that the quality of the resulting draft is of a higher quality than the individual constituents? We are developing a system to investigate these questions and their related issues. One such burning issue is to understand, model and detect erroneously assembled segments. Clearly the misassemblies require more attention than gaps in the assembly (the former is the usual problem of we-don't-know-what-we-don't-know) and is also perhaps more amenable to computational/algorithmic remedies. We are focusing on two directions to study these: one is on intrinsic single assembly statistics and the other on comparing assemblies.

Our work is motivated by the need for the generation of a high quality draft for Theobroma cacao, a plant with estimated 440 Mb genome size that produces cocoa beans, the basic ingredient in chocolate. The case study with the plant genomes originated within the USDA/Mars/IBM Consortium for sequencing the T. cacao genome. In this talk I will describe our assembly evaluation environment and present some preliminary results.

Speaker

Niina Haiminen[1]

Author affiliations

[1] IBM T. J. Watson Research Center, Computational Genomics, Yorktown Heights, NY, USA

TT11—Data integration in proteomics through EnVsion and EnCore webservices

Pascal Kahlem[1], Henning Hermjakob[1]

This demo will introduce the EnCore infrastructure as collection of webservices to query proteomics data. We will describe how to use individual EnCore webservices and we will explain how to create workflow connecting different services. We will explain how the EnCore platform is technically structured and how it simplifies the way to use webservice using a common format named "EnXML". We will list and briefly describe webservices available in EnCore. Among these services we will explore how to query molecular interactions, protein identifications, biological pathways, protein sequence information, biological models, protein localization and ontology distribution. We will learn how to use EnVision, a web graphical user interface to query EnCore webservices and display elaborate information from its results. We will see examples to query EnCore using EnVision and we will describe in detail the results obtained by EnVsion.

URL

http://www.enfin.org

Speaker

Rafael Jimenez[1]

Author affiliations

[1] European Bioinformatics Institute

TT12—ELIXIR: a sustainable European infrastructure for biological information

Dr Andrew Lyall[1]

ELIXIR has recently completed a two year, European wide consultation involving academic and industrial users, data providers and international collaborators, including three stakeholders meetings, two surveys, and fourteen work packages.

ELIXIR will be a distributed infrastructure arranged as a hub and nodes, with the hub at the European Molecular Biology Laboratory's European Bioinformatics Institute (EMBL-EBI) in Hinxton, UK.

The ELIXIR Steering Committee has recently issued a Request for Suggestions for ELIXIR nodes, seeking input from organisations that are interested in hosting one of its 'nodes', in order to help shape ELIXIRs construction. Organisations wishing to contribute their input have been requested to consider how they might contribute: data resources; bio-computing capacity; infrastructure for data integration; and services for the research community, including training and standards development.

Responses from this round of stakeholder input will enable ELIXIR to: define the landscape of potential nodes; provide the coordination necessary to ensure interoperability; and identify any duplications and missing capabilities.

URL

http://www.elixir-europe.org

Author affiliations

[1] ELIXIR Project Manager, EMBL-EBI

TT13—The Protein Data Bank in Europe (PDBe) - bringing structure to biology

Sameer Velankar[1], Gerard Kleywegt[1]

In 2011, the Protein Data Bank (PDB) celebrates its 40th anniversary. To date, the PDB has essentially been a historic archive capturing structures (and the underpinning experimental data) of biomacromolecules published in the primary literature. As the use of structural data by non-experts becomes commonplace, the demands on the archive (by both these users and funding agencies) and the way it is made accessible will inevitably change. It is therefore necessary to transform sites that serve the archive into resources that are directly relevant for scientists who work in biomedicine and related disciplines, while simultaneously taking care not to alienate the communities that produce the structures. EMBL-EBI's Protein Data Bank in Europe (PDBe), one of the founders of the Worldwide Protein Data Bank (wwPDB), is committed to becoming such a resource.

The first step towards this goal is the recent make-over of our webpages. During the ECCB session, we will demonstrate how to use the new PDBe pages to navigate to the five areas ("ALIVE") where we are traditionally strong: Advanced services, Ligands, Integration, Validation and Experimental data. We will also relate this to our core position within the European Bioinformatics Institute (EBI).

During the session we will demonstrate:

- New front pages and functionality
- Wizard for structure newbies
- PDBprints (visual overview of PDB entry)
- PDBeXplore (exploring biological data)
- Easy navigation to the five areas ("ALIVE") on which the PDBe focuses
- Advanced services
- SSM (Secondary Structure Matching)
- Pisa (Quaternary Structure)
- PDBeMotif (Motifs and Sites)
- Ligands
- PDBeChem (Ligand information)
- Integration with other databases and resources
- SIFTS project (in collaboration with UniProt)
- Validation (implementing new tools and resources)
- Experimental data
- The Electron Density Server (which will be ported from Uppsala to PDBe
- The Electron Microscopy Database (EMDB)
- Nuclear Magnetic Resonance (NMR) data

URL

www.pdbe.org

Speaker

Dr Wim Vranken[1]

Author affiliations

[1] Protein Data Bank in Europe (PDBe), EMBL-EBI

TT14—The Universal Protein Resource (UniProt)

The UniProt Consortium[1,2,3]

UniProt is the central resource for storing and interconnecting information from large and disparate sources, and the most comprehensive catalog of protein sequence and functional annotation. UniProt is built upon the

extensive bioinformatics infrastructure and scientific expertise at European Bioinformatics Institute (EBI), Protein Information Resource (PIR) and Swiss Institute of Bioinformatics (SIB). It has four components optimized for different uses. The UniProt Knowledgebase (UniProtKB) is an expertly curated database, a central access point for integrated protein information with cross-references to multiple sources. The UniProt Archive (UniParc) is a comprehensive sequence repository, reflecting the history of all protein sequences. UniProt Reference Clusters (UniRef) merge closely related sequences based on sequence identity to speed up searches. The UniProt Metagenomic and Environmental Sequences (UniMES) database is a repository specifically developed for the expanding area of metagenomic and environmental data. Other developments include the ID mapping service which allows users to map between UniProtKB and more than 30 other data sources; and UniSave, a comprehensive history service of UniProtKB entries.

The demonstration will cover:

- A brief description of the UniProt databases.
- Accessing UniProt using simple query syntax. The user will be presented with helpful suggestions and hints.
- Exploration of sequence similarity searches, alignments and ID mapping tools provided.
- Accessing UniProt data programmatically.

This demonstration will also encourage user interaction and feedback.

URL

http://www.uniprot.org/

Speaker

Maria J. Martin, Ph.D.[1]

Author affiliations

[1] Team Leader, UniProt (Development), EMBL-European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK [2] Protein Information Resource, Georgetown University Medical Center, 3300 Whitehaven St. NW, Suite 1200 Washington, DC 20007, USA [3] Swiss Institute of Bioinformatics, Centre Medical Universitaire 1 rue Michel Servet 1211 Geneva 4, Switzerland.

TT15—Accurate next gen sequencing data analysis on cloud computing

Miklós Csűrös[2], Szilveszter Juhos[1]

We present an internet based, automated, highly accurate genome variant analysis toolkit for next generation sequencing data called Omixon Variant Toolkit. This method is applicable for exome analysis or for analyzing highly variable genomes.

Accurate, reliable and highly sensitive genome variant analysis based on next-generation sequencing short-read data poses a significant computational challenge. There is always a trade-off between computational speed and guaranteed high accuracy. (P. Flicek & E. Birney, S6 vol. 6 no. 11s 2009 Nature Methods) A small increase in the desired accuracy can exponentially increase computational demand. Since many genomics applications do not require high precision, the most popular methods are fast but less accurate especially for identifying small insertions and deletions. These variants are important for variable genomes, in cancer research or biomarker discovery. We developed an accurate and computationally efficient method implemented in the Omixon Variant Toolkit which we offer as an internet-based service. In this presentation we demonstrate how to use our automated analysis tool for genome variant discovery. We also demonstrate the accuracy of the method through comparing results obtained with the most popular and the best open source tools available. We show that the Omixon Variant Toolkit is able to map short reads in the genome in areas of high density of insertions, deletions and SNP and longer (10-12 bp) deletions.

We carried out comparative analysis of 14 strains of a bacteria causing inflammatory disorder in humans including some pathogenic, non-pathogenic, drug-resistant, non-resistant strains. Our method was able to identify correctly variants and many of them were later confirmed by Sanger sequencing. We shall present additional example applications in human exome analysis.

Additional benefits of our internet-based method is that users do not need expensive computational facilities

since the calculation is carried out on a cloud computing platform, where the user pays only for the particular calculation. It is significantly more cost efficient than owning your own infrastructure. In addition, our software is very easy to use and guarantees high accuracy without having to know large number of different parameter settings. Users can control the maximum number of mismatches and indels and can choose from three levels of sensitivity: fast, sensitive and ultra sensitive.

Accuracy is achieved by a spaced seeds mapping, followed by greedy extension and a precise statistical alignment based on a pair-hidden Markov model, combining DNA sequence evolution models and sequencing errors (from read quality files). The method was published in Csuros, Juhos, Berces "Fast Mapping and Precise Alignment of AB SOLiD Color Reads to Reference DNA" Springer Lecture Notes in Bioinformatics 6293:176-188, 2010.

URL

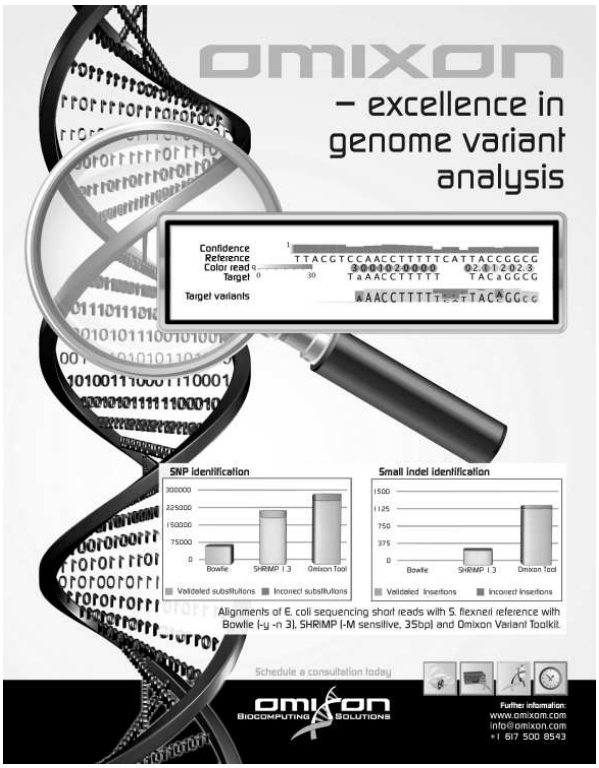
http://www.omixon.com

Speaker

Attila Bérces[1]

Author affiliations

[1]Omixon, Budapest, Hungary [2]University of Montreal, Canada



ORAL PRESENTATIONS

Sequence Analysis, Alignment and Next Generation Sequencing

PT2—Integrating genome assemblies with MAIA

Jurgen Nijkamp[1,2,3,*], Wynand Winterbach[1,4], Marcel van den Broek[2,3], Jean-Marc Daran[2,3], Marcel Reinders[1,3,5] and Dick de Ridder[1,3,5]

[1]The Delft Bioinformatics Lab, Department of Mediamatics, Delft University of Technology, Mekelweg 4, 2628 CD Delft, [2]Industrial Microbiology Group, Department of Biotechnology, Delft University of Technology, Julianalaan 67, 2628 BC Delft, [3]Kluyver Centre for Genomics of Industrial Fermentation, P.O. Box 5057, 2600 GA Delft, [4]Network Architectures and Services, Department of Telecommunications, Delft University of Technology, Mekelweg 4, 2628 CD Delft and [5]Netherlands Bioinformatics Center, 260 NBIC, P.O. Box 9101, 6500 HB Nijmegen, The Netherlands

Motivation: De novo assembly of a eukaryotic genome with next-generation sequencing data is still a challenging task. Over the past few years several assemblers have been developed, often suitable for one specific type of sequencing data. The number of known genomes is expanding rapidly, therefore it becomes possible to use multiple reference genomes for assembly projects. We introduce an assembly integrator that makes use of all available data, i.e. multiple de novo assemblies and mappings against multiple related genomes, by optimizing a weighted combination of criteria.

Results: The developed algorithm was applied on the de novo sequencing of the Saccharomyces cerevisiae CEN.PK 113-7D strain. Using Solexa and 454 read data, two de novo and three comparative assemblies were constructed and subsequently integrated, yielding 29 contigs, covering more than 12 Mbp; a drastic improvement compared with the single assemblies.

Availability: MAIA is available as a Matlab package and can be downloaded from http://bioinformatics.tudelft.nl

Contact: j.f.nijkamp@tudelft.nl

PT15—Characteristics of 454 pyrosequencing data—enabling realistic simulation with flowsim

Susanne Balzer[1,2,*], Ketil Malde[1,*], Anders Lanzén[3,4], Animesh Sharma[2] and Inge Jonassen[2,3]

[1]Institute of Marine Research, PO Box 1870, N-5817, [2]Department of Informatics, University of Bergen, PO Box 7803, N-5020, [3]Computational Biology Unit, Bergen Center for Computational Science, Thormøhlensgate 55, N-5008 and [4]Department of Biology, University of Bergen, PO Box 7803, N-5020, Bergen

Motivation: The commercial launch of 454 pyrosequencing in 2005 was a milestone in genome sequencing in terms of performance and cost. Throughout the three available releases, average read lengths have increased to ~500 base pairs and are thus approaching read lengths obtained from traditional Sanger sequencing. Study design of sequencing projects would benefit from being able to simulate experiments.

Results: We explore 454 raw data to investigate its characteristics and derive empirical distributions for the flow values generated by pyrosequencing. Based on our findings, we implement Flowsim, a simulator that generates realistic pyrosequencing data files of arbitrary size from a given set of input DNA sequences. We finally use our simulator to examine the impact of sequence lengths on the results of concrete whole-genome assemblies, and we suggest its use in planning of sequencing projects, benchmarking of assembly methods and other fields.

Availability: Flowsim is freely available under the General Public License from http://blog.malde.org/index.php/flowsim/

Contact: susanne.balzer@imr.no; ketil.malde@imr.no

PT16—A fast algorithm for exact sequence search in biological sequences using polyphase decomposition

Abhilash Srikantha[1,*], Ajit S. Bopardikar[1,*], Kalyan Kumar Kaipa[1], Parthasarathy Venkataraman[1], Kyusang Lee[2], TaeJin Ahn[2] and Rangavittal Narayanan[1]

1Samsung Advanced Institute of Technology India Lab, Bangalore, Karnataka, India and 2Genetic Analysis Group, Emerging Center, Samsung Advanced Institute of Technology, Suwon, South Korea

Motivation: Exact sequence search allows a user to search for a specific DNA subsequence in a larger DNA sequence or database. It serves as a vital block in many areas such as Pharmacogenetics, Phylogenetics and Personal Genomics. As sequencing of genomic data becomes increasingly affordable, the amount of sequence data that must be processed will also increase exponentially. In this context, fast sequence search algorithms will play an important role in exploiting the information contained in the newly sequenced data. Many existing algorithms do not scale up well for large sequences or databases because of their high-computational costs. This article describes an efficient algorithm for performing fast searches on large DNA sequences. It makes use of hash tables of Q-grams that are constructed after downsampling the database, to enable efficient search and memory use. Time complexity for pattern search is reduced using beam pruning techniques. Theoretical complexity calculations and performance figures are presented to indicate the potential of the proposed algorithm.
Contact: s.abhilash@samsung.com; ajit.b@samsung.com

PT17—Classification of ncRNAs using position and size information in deep sequencing data

Florian Erhard* and Ralf Zimmer
Institut für Informatik, Ludwig-Maximilians-Universität München, Amalienstraße 17, 80333 München, Germany

Motivation: Small non-coding RNAs (ncRNAs) play important roles in various cellular functions in all clades of life. With next-generation sequencing techniques, it has become possible to study ncRNAs in a high-throughput manner and by using specialized algorithms ncRNA classes such as miRNAs can be detected in deep sequencing data. Typically, such methods are targeted to a certain class of ncRNA. Many methods rely on RNA secondary structure prediction, which is not always accurate and not all ncRNA classes are characterized by a common secondary structure. Unbiased classification methods for ncRNAs could be important to improve accuracy and to detect new ncRNA classes in sequencing data.
Results: Here, we present a scoring system called ALPS (alignment of pattern matrices score) that only uses primary information from a deep sequencing experiment, i.e. the relative positions and lengths of reads, to classify ncRNAs. ALPS makes no further assumptions, e.g. about common structural properties in the ncRNA class and is nevertheless able to identify ncRNA classes with high accuracy. Since ALPS is not designed to recognize a certain class of ncRNA, it can be used to detect novel ncRNA classes, as long as these unknown ncRNAs have a characteristic pattern of deep sequencing read lengths and positions. We evaluate our scoring system on publicly available deep sequencing data and show that it is able to classify known ncRNAs with high sensitivity and specificity.
Availability: Calculated pattern matrices of the datasets hESC and EB are available at the project web site http://www.bio.ifi.lmu.de/ALPS. An implementation of the described method is available upon request from the authors.
Contact: florian.erhard@bio.ifi.lmu.de

Comparative Genomics, Phylogeny and Evolution

PT35—Parsimony and likelihood reconstruction of human segmental duplications

Crystal L. Kahn[1,*], Borislav H. Hristov[1] and Benjamin J. Raphael[1,2,*]
[1]Department of Computer Science and [2]Center for Computational Molecular Biology, Brown University, Providence, RI, 02912, USA

Motivation: Segmental duplications >1 kb in length with ≥90% sequence identity between copies comprise nearly 5% of the human genome. They are frequently found in large, contiguous regions known as duplication blocks that can contain mosaic patterns of thousands of segmental duplications. Reconstructing the evolutionary history of these complex genomic regions is a non-trivial, but important task.
Results: We introduce parsimony and likelihood techniques to analyze the evolutionary relationships between duplication blocks. Both techniques rely on a generic model of duplication in which long, contiguous substrings are copied and reinserted over large physical distances, allowing for a duplication block to be constructed by aggregating substrings of other blocks. For the likelihood method, we give an efficient dynamic programming algorithm to compute the weighted ensemble of all duplication scenarios that account for the construction of a duplication block. Using this ensemble, we derive the probabilities of various duplication scenarios.We

formalize the task of reconstructing the evolutionary history of segmental duplications as an optimization problem on the space of directed acyclic graphs. We use a simulated annealing heuristic to solve the problem for a set of segmental duplications in the human genome in both parsimony and likelihood settings.
Availability: Supplementary information is available at http://www.cs.brown.edu/people/braphael/supplements.
Contact: clkahn@cs.brown.edu; braphael@cs.brown.edu.

PT36—Maximum likelihood estimation of locus-specific mutation rates in Y-chromosome short tandem repeats

Osnat Ravid-Amir and Saharon Rosset*
Department of Statistics and Operation Research, Tel Aviv, Israel

Motivation: Y-chromosome short tandem repeats (Y-STRs) are widely used for population studies, forensic purposes and, potentially, the study of disease, therefore knowledge of their mutation rate is valuable. Here we show a novel method for estimation of site-specific Y-STR mutation rates from partial phylogenetic information, via the maximum likelihood framework.
Results: Given Y-STR data classified into haplogroups, we describe the likelihood of observed data, and develop optimization strategies for deriving maximum likelihood estimates of mutation rates. We apply our method to Y-STR data from two recent papers. We show that our estimates are comparable, often more accurate than those obtained in familial studies, although our data sample is much smaller, and was not collected specifically for our study. Furthermore, we obtain mutation rate estimates for DYS388, DYS426, DYS457, three STRs for which there were no mutation rate measures until now.
Contact: saharon@post.tau.ac.il

Protein and Nucleotide Structure

PT4—Vorescore—fold recognition improved by rescoring of protein structure models

Gergely Csaba* and Ralf Zimmer*
Practical Informatics and Bioinformatics Group, Department of Informatics, Ludwig-Maximilians-Universität München, Amalienstr. 17, D-80333 München, Germany

Summary: The identification of good protein structure models and their appropriate ranking is a crucial problem in structure prediction and fold recognition. For many alignment methods, rescoring of alignment-induced models using structural information can improve the separation of useful and less useful models as compared with the alignment score. Vorescore, a template-based protein structure model rescoring system is introduced. The method scores the model structure against the template used for the modeling using Vorolign. The method works on models from different alignment methods and incorporates both knowledge from the prediction method and the rescoring.
Results: The performance of Vorescore is evaluated in a largescale and difficult protein structure prediction context. We use different threading methods to create models for 410 targets, in three scenarios: (i) family members are contained in the template set; (ii) superfamily members (but no family members); and (iii) only fold members (but no family or superfamily members). In all cases Vorescore improves significantly (e.g. 40% on both Gotoh and HHalign at the fold level) on the model quality, and clearly outperforms the state-of-the-art physics-based model scoring system Rosetta. Moreover, Vorescore improves on other successful rescoring approaches such as Pcons and ProQ. In an additional experiment we add high-quality models based on structural alignments to the set, which allows Vorescore to improve the fold recognition rate by another 50%.
Availability: All models of the test set (about 2 million, 44GB gzipped) are available upon request.
Contact: csaba@bio.ifi.lmu.de; ralf.zimmer@ifi.lmu.de

PT5—Solenoid and non-solenoid protein recognition using stationary wavelet packet transform

An Vo[1,*,†], Nha Nguyen[2,†] and Heng Huang[3]
[1]The Feinstein Institute for Medical Research, North Shore LIJ Health System, NY, [2]Department of Electrical Engineering and [3]Department of Computer Science and Engineering, University of Texas at Arlington, TX, USA

Motivation: Solenoid proteins are emerging as a protein class with properties intermediate between structured and intrinsically unstructured proteins. Containing repeating structural units, solenoid proteins are expected to

share sequence similarities. However, in many cases, the sequence similarities are weak and non-detectable. Moreover, solenoids can be degenerated and widely vary in the number of units. So that it is difficult to detect them. Recently, several solenoid repeats detection methods have been proposed, such as self-alignment of the sequence, spectral analysis and discrete Fourier transform of sequence. Although these methods have shown good performance on certain data sets, they often fail to detect repeats with weak similarities. In this article, we propose a new approach to recognize solenoid repeats and non-solenoid proteins using stationary wavelet packet transform (SWPT). Our method associates with three advantages: (i) naturally representing five main factors of protein structure and properties by wavelet analysis technique; (ii) extracting novel wavelet features that can capture hidden components from solenoid sequence similarities and distinguish them from global proteins; (iii) obtaining statistics features that capture repeating motifs of solenoid proteins.

Results: Our method analyzes the characteristics of amino acid sequence in both spectral and temporal domains using SWPT. Both global and local information of proteins are captured by SWPT coefficients. We obtain and integrate wavelet-based features and statistics-based features of amino acid sequence to improve the classification task. Our proposed method is evaluated by comparing to state-of-the-art methods such as HHrepID and REPETITA. The experimental results show that our algorithm consistently outperforms them in areas under ROC curve. At the same false positive rate, the sensitivity of our WAVELET method is higher than other methods.

Availability: http://www.naaan.org/anvo/Software/Software.htm

Contact: anphuocnhu.vo@mavs.uta.edu

PT6—Discriminatory power of RNA family models

Christian Höner zu Siederdissen* and Ivo L. Hofacker

Institute for Theoretical Chemistry, University of Vienna, Währinger Strasse 17, A-1090 Wien, Austria

Motivation: RNA family models group nucleotide sequences that share a common biological function. These models can be used to find new sequences belonging to the same family. To succeed in this task, a model needs to exhibit high sensitivity as well as high specificity. As model construction is guided by a manual process, a number of problems can occur, such as the introduction of more than one model for the same family or poorly constructed models. We explore the Rfam database to discover such problems.

Results: Our main contribution is in the definition of the discriminatory power of RNA family models, together with a first algorithm for its computation. In addition, we present calculations across the whole Rfam database that show several families lacking high specificity when compared to other families. We give a list of these clusters of families and provide a tentative explanation. Our program can be used to: (i) make sure that new models are not equivalent to any model already present in the database; and (ii) new models are not simply submodels of existing families.

Availability: www.tbi.univie.ac.at/software/cmcompare/. The code is licensed under the GPLv3. Results for the whole Rfam database and supporting scripts are available together with the software.

Contact: choener@tbi.univie.ac.at

PT7—RactIP: fast and accurate prediction of RNA-RNA interaction using integer programming

Yuki Kato[1,*,†], Kengo Sato[2,*,†], Michiaki Hamada[3,4], Yoshihide Watanabe[5], Kiyoshi Asai[2,4] and Tatsuya Akutsu[1]

[1]Bioinformatics Center, Institute for Chemical Research, Kyoto University, Gokasho, Uji, Kyoto 611-0011, [2]Graduate School of Frontier Sciences, University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8561, [3]Mizuho Information & Research Institute, Inc, 2-3 Kanda-Nishikicho, Chiyoda-ku, Tokyo 101-8443, [4]Computational Biology Research Center (CBRC), National Institute of Advanced Industrial Science and Technology (AIST), 2-41-6, Aomi, Koto-ku, Tokyo 135-0064 and [5]Department of Mathematical Sciences, Faculty of Science and Engineering, Doshisha University, 1-3 Tataramiyakodani, Kyotanabe, Kyoto 610-0321, Japan

Motivation: Considerable attention has been focused on predicting RNA–RNA interaction since it is a key to identifying possible targets of non-coding small RNAs that regulate gene expression posttranscriptionally. A number of computational studies have so far been devoted to predicting joint secondary structures or binding sites under a specific class of interactions. In general, there is a tradeoff between range of interaction type and efficiency of a prediction algorithm, and thus efficient computational methods for predicting comprehensive type of interaction are still awaited.

Results: We present RactIP, a fast and accurate prediction method for RNA–RNA interaction of general type using integer programming. RactIP can integrate approximate information on an ensemble of equilibrium joint structures into the objective function of integer programming using posterior internal and external base-paring probabilities. Experimental results on real interaction data show that prediction accuracy of RactIP is at least comparable to that of several state-of-the-art methods for RNA–RNA interaction prediction. Moreover, we demonstrate that RactIP can run incomparably faster than competitive methods for predicting joint secondary structures.
Availability: RactIP is implemented in C++, and the source code is available at
http://www.ncrna.org/software/ractip/
Contact: ykato@kuicr.kyoto-u.ac.jp; satoken@k.u-tokyo.ac.jp

Prediction and Annotation of Molecular Function

PT18—Prototypes of elementary functional loops unravel evolutionary connections between protein functions

Alexander Goncearenco[1,2] and Igor N. Berezovsky[1,*]
[1]Computational Biology Unit, Bergen Center for Computational Science and [2]Department of Informatics, University of Bergen, N-5008 Norway

Motivation: Earlier studies of protein structure revealed closed loops with a characteristic size 25–30 residues and ring-like shape as a basic universal structural element of globular proteins. Elementary functional loops (EFLs) have specific signatures and provide functional residues important for binding/activation and principal chemical transformation steps of the enzymatic reaction. The goal of this work is to show how these functional loops evolved from pre-domain peptides and to find a set of prototypes from which the EFLs of contemporary proteins originated.
Results: This article describes a computational method for deriving prototypes of EFLs based on the sequences of complete genomes. The procedure comprises the iterative derivation of sequence profiles followed by their hierarchical clustering. The scoring function takes into account information content on profile positions, thus preserving the signature. The statistical significance of scores is evaluated from the empirical distribution of scores of the background model. A set of prototypes of EFLs from archaeal proteomes is derived. This set delineates evolutionary connections between major functions and illuminates how folds and functions emerged in pre-domain evolution as a combination of prototypes.
Contact: Igor.Berezovsky@uni.no

PT19—Improved sequence-based prediction of disordered regions withmultilayer fusion of multiple information sources

Marcin J. Mizianty, Wojciech Stach, Ke Chen, Kanaka Durga Kedarisetti, Fatemeh Miri Disfani and Lukasz Kurgan*
Department of Electrical and Computer Engineering, University of Alberta, Edmonton, Canada T6G 2V4

Motivation: Intrinsically disordered proteins play a crucial role in numerous regulatory processes. Their abundance and ubiquity combined with a relatively low quantity of their annotations motivate research toward the development of computational models that predict disordered regions from protein sequences. Although the prediction quality of these methods continues to rise, novel and improved predictors are urgently needed.
Results: We propose a novel method, named MFDp (Multilayered Fusion-based Disorder predictor), that aims to improve over the current disorder predictors. MFDp is as an ensemble of 3 Support Vector Machines specialized for the prediction of short, long and generic disordered regions. It combines three complementary disorder predictors, sequence, sequence profiles, predicted secondary structure, solvent accessibility, backbone dihedral torsion angles, residue flexibility and B-factors. Our method utilizes a custom-designed set of features that are based on raw predictions and aggregated raw values and recognizes various types of disorder. The MFDp is compared at the residue level on two datasets against eight recent disorder predictors and top-performing methods from the most recent CASP8 experiment. In spite of using training chains with ≤25% similarity to the test sequences, our method consistently and significantly outperforms the other methods based on the MCC index. The MFDp outperforms modern disorder predictors for the binary disorder assignment and

provides competitive real-valued predictions. The MFDp's outputs are also shown to outperform the other methods in the identification of proteins with long disordered regions.
Availability: http://biomine.ece.ualberta.ca/MFDp.html
Contact: lkurgan@ece.ualberta.ca

PT20—A predictor for toxin-like proteins exposes cell modulator candidates within viral genomes

Guy Naamati[1], Manor Askenazi[2,3] and Michal Linial[2,*]
[1]School of Computer Science and Engineering, [2]Department of Biological Chemistry, Sudarsky Center for Computational Biology, Hebrew University of Jerusalem, Israel and [3]Blais Proteomics Center, Dana-Farber Cancer Institute, Boston, MA, USA

Motivation: Animal toxins operate by binding to receptors and ion channels. These proteins are short and vary in sequence, structure and function. Sporadic discoveries have also revealed endogenous toxin-like proteins in non-venomous organisms. Viral proteins are the largest group of quickly evolving proteomes. We tested the hypothesis that toxin-like proteins exist in viruses and that they act to modulate functions of their hosts.
Results: We updated and improved a classifier for compact proteins resembling short animal toxins that is based on a machine-learning method. We applied it in a large-scale setting to identify toxin-like proteins among short viral proteins. Among the ~ 26 000 representatives of such short proteins, 510 sequences were positively identified. We focused on the 19 highest scoring proteins. Among them, we identified conotoxin-like proteins, growth factors receptor-like proteins and anti-bacterial peptides. Our predictor was shown to enhance annotation inference for many 'uncharacterized' proteins. We conclude that our protocol can expose toxin-like proteins in unexplored niches including metagenomics data and enhance the systematic discovery of novel cell modulators for drug development.
Availability: ClanTox is available at http://www.clantox.cs.huji.ac.il
Contact: michall@cc.huji.ac.il

Gene Regulation and Transcriptomics

PT25—Nonlinear dimension reduction and clustering by Minimum Curvilinearity unfold neuropathic pain and tissue embryological classes

Carlo Vittorio Cannistraci[1,2,3,4,5,*], Timothy Ravasi[1,5], Franco Maria Montevecchi[3], Trey Ideker[5] and Massimo Alessio[2,*]
[1]Red Sea Integrative Systems Biology Lab, Computational Bioscience Research Center, Division of Chemical and Life Sciences and Engineering, King Abdullah University for Science and Technology (KAUST), Jeddah, Kingdom of Saudi Arabia, [2]Proteome Biochemistry, San Raffaele Scientific Institute, via Olgettina 58, 20132 Milan, [3]Department of Mechanics, [4]CMP Group, Microsoft Research, Politecnico di Torino, c/so Duca degli Abruzzi 24, 10129 Turin, Italy, [5]Department of Bioengineering and Department of Medicine, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093 USA

Motivation: Nonlinear small datasets, which are characterized by low numbers of samples and very high numbers of measures, occur frequently in computational biology, and pose problems in their investigation. Unsupervised hybrid-two-phase (H2P) procedures—specifically dimension reduction (DR), coupled with clustering—provide valuable assistance, not only for unsupervised data classification, but also for visualization of the patterns hidden in high-dimensional feature space.
Methods: 'Minimum Curvilinearity' (MC) is a principle that—for small datasets—suggests the approximation of curvilinear sample distances in the feature space by pair-wise distances over their minimum spanning tree (MST), and thus avoids the introduction of any tuning parameter. MC is used to design two novel forms of nonlinear machine learning (NML): Minimum Curvilinear embedding (MCE) for DR, and Minimum Curvilinear affinity propagation (MCAP) for clustering.
Results: Compared with several other unsupervised and supervised algorithms, MCE and MCAP, whether individually or combined in H2P, overcome the limits of classical approaches. High performance was attained in the visualization and classification of: (i) pain patients (proteomic measurements) in peripheral neuropathy; (ii) human organ tissues (genomic transcription factor measurements) on the basis of their embryological origin.
Conclusion: MC provides a valuable framework to estimate nonlinear distances in small datasets. Its extension to large datasets is prefigured for novel NMLs. Classification of neuropathic pain by proteomic profiles offers

new insights for future molecular and systems biology characterization of pain. Improvements in tissue embryological classification refine results obtained in an earlier study, and suggest a possible reinterpretation of skin attribution as mesodermal.
Availability: https://sites.google.com/site/carlovittoriocannistraci/home
Contact: kalokagathos.agon@gmail.com; massimo.alessio@hsr.it

PT26—A varying threshold method for ChIP peak-calling using multiple sources of information

Kuan-Bei Chen[1] and Yu Zhang[2,*]
[1]Department of Computer Science and Engineering, The Pennsylvania State University, University Park and [2]Department of Statistics, The Pennsylvania State University, 422A Thomas, University Park, PA 16802, USA

Motivation: Gene regulation commonly involves interaction among DNA, proteins and biochemical conditions. Using chromatin immunoprecipitation (ChIP) technologies, protein–DNA interactions are routinely detected in the genome scale. Computational methods that detect weak protein-binding signals and simultaneously maintain a high specificity yet remain to be challenging. An attractive approach is to incorporate biologically relevant data, such as protein cooccupancy, to improve the power of protein-binding detection. We call the additional data related with the target protein binding as supporting tracks.
Results: We propose a novel but rigorous statistical method to identify protein occupancy in ChIP data using multiple supporting tracks (PASS2). We demonstrate that utilizing biologically related information can significantly increase the discovery of true proteinbinding sites, while still maintaining a desired level of false positive calls. Applying the method to GATA1 restoration in mouse erythroid cell line, we detected many new GATA1-binding sites using GATA1 co-occupancy data.
Availability: http://stat.psu.edu/~yuzhang/pass2.tar
Contact: yuzhang@stat.psu.edu

PT27—is-rSNP: a novel technique for in silico regulatory SNP detection

Geoff Macintyre[1,2,*], James Bailey[1,2], Izhak Haviv[3,4,5] and Adam Kowalczyk[1,2,*]
[1]Department of Computer Science and Software Engineering, [2]NICTA, Victoria Research Lab, The University of Melbourne, Victoria 3010, [3]Bioinformatics and Systems Integration, The Blood and DNA Profiling Facility, Baker IDI Heart and Diabetes Institute, 75 Commercial Rd, Prahran, Victoria 3004, [4]Metastasis Research Lab, Peter MacCallum Cancer Centre, St Andrews Place, East Melbourne, Victoria 3002 and [5]Department of Biochemistry and Molecular Biology. The University of Melbourne, Victoria 3010, Australia

Motivation: Determining the functional impact of non-coding disease-associated single nucleotide polymorphisms (SNPs) identified by genome-wide association studies (GWAS) is challenging. Many of these SNPs are likely to be regulatory SNPs (rSNPs): variations which affect the ability of a transcription factor (TF) to bind to DNA. However, experimental procedures for identifying rSNPs are expensive and labour intensive. Therefore, in silico methods are required for rSNP prediction. By scoring two alleles with a TF position weight matrix (PWM), it can be determined which SNPs are likely rSNPs. However, predictions in this manner are noisy and no method exists that determines the statistical significance of a nucleotide variation on a PWM score.
Results: We have designed an algorithm for in silico rSNP detection called is-rSNP. We employ novel convolution methods to determine the complete distributions of PWM scores and ratios between allele scores, facilitating assignment of statistical significance to rSNP effects. We have tested our method on 41 experimentally verified rSNPs, correctly predicting the disrupted TF in 28 cases. We also analysed 146 disease-associated SNPs with no known functional impact in an attempt to identify candidate rSNPs. Of the 11 significantly predicted disrupted TFs, 9 had previous evidence of being associated with the disease in the literature. These results demonstrate that is-rSNP is suitable for high-throughput screening of SNPs for potential regulatory function. This is a useful and important tool in the interpretation of GWAS.
Availability: is-rSNP software is available for use at: http://www.genomics.csse.unimelb.edu.au/is-rSNP
Contact: gmaci@csse.unimelb.edu.au; adam.kowalczyk@nicta.com.au

PT32—De-correlating expression in gene-set analysis

Dougu Nam
School of Nano-Biotechnology and Chemical Engineering, Ulsan National Institute of Science and Technology, Republic of Korea

Motivation: Group-wise pattern analysis of genes, known as geneset analysis (GSA), addresses the differential expression pattern of biologically pre-defined gene sets. GSA exhibits high statistical power and has revealed many novel biological processes associated with specific phenotypes. In most cases, however, GSA relies on the invalid assumption that the members of each gene set are sampled independently, which increases false predictions.
Results: We propose an algorithm, termed DECO, to remove (or alleviate) the bias caused by the correlation of the expression data in GSAs. This is accomplished through the eigenvalue-decomposition of covariance matrixes and a series of linear transformations of data. In particular, moderate de-correlation methods that truncate or rescale eigenvalues were proposed for a more reliable analysis. Tests of simulated and real experimental data show that DECO effectively corrects the correlation structure of gene expression and improves the prediction accuracy (specificity and sensitivity) for both gene and sample-randomizing GSA methods.
Availability: The MATLAB codes and the tested data sets are available at ftp://deco.nims.re.kr/pub or from the author.
Contact: dougnam@unist.ac.kr

PT34—Discovering graphical Granger causality using the truncating lasso penalty

Ali Shojaie* and George Michailidis
Department of Statistics, University of Michigan, Ann Arbor Michigan 48109, USA

Motivation: Components of biological systems interact with each other in order to carry out vital cell functions. Such information can be used to improve estimation and inference, and to obtain better insights into the underlying cellular mechanisms. Discovering regulatory interactions among genes is therefore an important problem in systems biology. Whole-genome expression data over time provides an opportunity to determine how the expression levels of genes are affected by changes in transcription levels of other genes, and can therefore be used to discover regulatory interactions among genes.
Results: In this article, we propose a novel penalization method, called truncating lasso, for estimation of causal relationships from time-course gene expression data. The proposed penalty can correctly determine the order of the underlying time series, and improves the performance of the lasso-type estimators. Moreover, the resulting estimate provides information on the time lag between activation of transcription factors and their effects on regulated genes. We provide an efficient algorithm for estimation of model parameters, and show that the proposed method can consistently discover causal relationships in the large p, small n setting. The performance of the proposed model is evaluated favorably in simulated, as well as real, data examples.
Availability: The proposed truncating lasso method is implemented in the R-package 'grangerTlasso' and is freely available at http://www.stat.lsa.umich.edu/~shojaie/
Contact: shojaie@umich.edu

Text Mining, Ontologies and Databases

PT1—Utopia documents: linking scholarly literature with research data

T. K. Attwood[1,2,*], D. B. Kell[3,4], P. McDermott[1,2], J. Marsh[3], S. R. Pettifer[2] and D. Thorne[3]
[1]School of Computer Science, [2]Faculty of Life Sciences, [3]School of Chemistry, University of Manchester, Oxford Road, Manchester M13 9PL and [4]Manchester Interdisciplinary Biocentre, 131 Princess Street, Manchester M1 7DN, UK

Motivation: In recent years, the gulf between the mass of accumulating-research data and the massive literature describing and analyzing those data has widened. The need for intelligent tools to bridge this gap, to rescue the knowledge being systematically isolated in literature and data silos, is now widely acknowledged.
Results: To this end, we have developed Utopia Documents, a novel PDF reader that semantically integrates visualization and data-analysis tools with published research articles. In a successful pilot with editors of the Biochemical Journal (BJ), the system has been used to transform static document features into objects that can

be linked, annotated, visualized and analyzed interactively (http://www.biochemj.org/bj/424/3/). Utopia Documents is now used routinely by BJ editors to mark up article content prior to publication. Recent additions include integration of various textmining and biodatabase plugins, demonstrating the system's ability to seamlessly integrate on-line content with PDF articles.
Availability: http://getutopia.com
Contact: teresa.k.attwood@manchester.ac.uk

PT21—Discriminative and informative features for biomolecular text mining with ensemble feature selection

Sofie Van Landeghem[1,2,†], Thomas Abeel[1,2,†], Yvan Saeys[1,2] and Yves Van de Peer[1,2,*]
[1]Department of Plant Systems Biology, VIB and [2]Department of Plant Biotechnology and Genetics, Ghent University, Gent, Belgium

Motivation: In the field of biomolecular text mining, black box behavior of machine learning systems currently limits understanding of the true nature of the predictions. However, feature selection (FS) is capable of identifying the most relevant features in any supervised learning setting, providing insight into the specific properties of the classification algorithm. This allows us to build more accurate classifiers while at the same time bridging the gap between the black box behavior and the end-user who has to interpret the results.
Results: We show that our FS methodology successfully discards a large fraction of machine-generated features, improving classification performance of state-of-the-art text mining algorithms. Furthermore, we illustrate how FS can be applied to gain understanding in the predictions of a framework for biomolecular event extraction from text.We include numerous examples of highly discriminative features that model either biological reality or common linguistic constructs. Finally, we discuss a number of insights from our FS analyses that will provide the opportunity to considerably improve upon current text mining tools.
Availability: The FS algorithms and classifiers are available in Java- ML (http://java-ml.sf.net). The datasets are publicly available from the BioNLP'09 Shared Task web site
(http://www-tsujii.is.s.u-tokyo.ac .jp/GENIA/SharedTask/).
Contact: yves.vandepeer@psb.ugent.be

PT22— BioXSD: the common data-exchange format for everyday bioinformatics web services

Matúš Kalaš[1,2,*], Pål Puntervoll[1], Alexandre Joseph[3], Edita Bartaševičiūtė[4], Armin Töpfer[1,5], Prabakar Venkataraman[1], Steve Pettifer[6], Jan Christian Bryne[1,2], Jon Ison[7], Christophe Blanchet[3], Kristoffer Rapacki[4] and Inge Jonassen[1,2]
[1]Computational Biology Unit, Bergen Center for Computational Science, Uni Research, 5008 Bergen, Norway, [2]Department of Informatics, University of Bergen, 5008 Bergen, Norway, [3]Université Lyon 1; CNRS, UMR 5086; IBCP, Institut de Biologie et Chimie des Protéines, 69367 Lyon Cedex 07, France, [4]Center for Biological Sequence Analysis, Department of Systems Biology, Technical University of Denmark, 2800 Kongens Lyngby, Denmark, [5]Institute for Bioinformatics, Center for Biotechnology, Bielefeld University, 33594 Bielefeld, Germany, [6]School of Computer Science, The University of Manchester, Manchester, M13 9PL, UK and [7]European Bioinformatics Institute, EMBL, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SD, UK

Motivation: The world-wide community of life scientists has access to a large number of public bioinformatics databases and tools, which are developed and deployed using diverse technologies and designs. More and more of the resources offer programmatic web-service interface. However, efficient use of the resources is hampered by the lack of widely used, standard data-exchange formats for the basic, everyday bioinformatics data types.
Results: BioXSD has been developed as a candidate for standard, canonical exchange format for basic bioinformatics data. BioXSD is represented by a dedicated XML Schema and defines syntax for biological sequences, sequence annotations, alignments and references to resources. We have adapted a set of web services to use BioXSD as the input and output format, and implemented a testcase workflow. This demonstrates that the approach is feasible and provides smooth interoperability. Semantics for BioXSD is provided by annotation with the EDAM ontology. We discuss in a separate section how BioXSD relates to other initiatives and approaches, including existing standards and the Semantic Web.
Availability: The BioXSD 1.0 XML Schema is freely available at http://www.bioxsd.org/BioXSD-1.0.xsd under the Creative Commons BY-ND 3.0 license. The http://bioxsd.org web page offers documentation, examples of data in BioXSD format, example workflows with source codes in common programming

languages, an updated list of compatible web services and tools and a repository of feature requests from the community.
Contact: matus.kalas@bccs.uib.no; developers@bioxsd.org; support@bioxsd.org

PT23—Improving disease gene prioritization using the semantic similarity of Gene Ontology terms

Andreas Schlicker†, Thomas Lengauer and Mario Albrecht*
Max Planck Institute for Informatics, Department of Computational Biology and Applied Algorithmics, Campus E1.4, 66123 Saarbrücken, Germany

Motivation: Many hereditary human diseases are polygenic, resulting from sequence alterations in multiple genes. Genomic linkage and association studies are commonly performed for identifying disease-related genes. Such studies often yield lists of up to several hundred candidate genes, which have to be prioritized and validated further. Recent studies discovered that genes involved in phenotypically similar diseases are often functionally related on the molecular level.
Results: Here, we introduce MedSim, a novel approach for ranking candidate genes for a particular disease based on functional comparisons involving the Gene Ontology. MedSim uses functional annotations of known disease genes for assessing the similarity of diseases as well as the disease relevance of candidate genes. We benchmarked our approach with genes known to be involved in 99 diseases taken from the OMIM database. Using artificial quantitative trait loci, MedSim achieved excellent performance with an area under the ROC curve of up to 0.90 and a sensitivity of over 70% at 90% specificity when classifying gene products according to their disease relatedness. This performance is comparable or even superior to related methods in the field, albeit using less and thus more easily accessible information.
Availability: MedSim is offered as part of our FunSimMat web service (http://www.funsimmat.de).
Contact: mario.albrecht@mpi-inf.mpg.de

PT24—Discovering drug–drug interactions: a text-mining and reasoning approach based on properties of drug metabolism

Luis Tari[1,*], Saadat Anwar[2], Shanshan Liang[2], James Cai[1] and Chitta Baral[2]
[1]Disease and Translational Informatics, Hoffmann-La Roche, Nutley, NJ 07110 and [2]Department of Computer Science and Engineering, Arizona State University, Tempe, AZ 85287, USA

Motivation: Identifying drug–drug interactions (DDIs) is a critical process in drug administration and drug development. Clinical support tools often provide comprehensive lists of DDIs, but they usually lack the supporting scientific evidences and different tools can return inconsistent results. In this article, we propose a novel approach that integrates text mining and automated reasoning to derive DDIs. Through the extraction of various facts of drug metabolism, not only the DDIs that are explicitly mentioned in text can be extracted but also the potential interactions that can be inferred by reasoning.
Results: Our approach was able to find several potential DDIs that are not present in DrugBank. We manually evaluated these interactions based on their supporting evidences, and our analysis revealed that 81.3% of these interactions are determined to be correct. This suggests that our approach can uncover potential DDIs with scientific evidences explaining the mechanism of the interactions.
Contact: luis.tari@roche.com

Protein Interactions, Molecular Networks and Systems Biology

PT3—A graphical method for reducing and relating models in systems biology

Steven Gay, Sylvain Soliman and François Fages*
EPI Contraintes, Institut National de Recherche en Informatique et Automatique, INRIA Paris-Rocquencourt, France

Motivation: In Systems Biology, an increasing collection of models of various biological processes is currently developed and made available in publicly accessible repositories, such as biomodels.net for instance, through common exchange formats such as SBML. To date, however, there is no general method to relate different models to each other by abstraction or reduction relationships, and this task is left to the modeler for re-using and coupling models. In mathematical biology, model reduction techniques have been studied for a long time, mainly in the case where a model exhibits different time scales, or different spatial phases, which can be

analyzed separately. These techniques are however far too restrictive to be applied on a large scale in systems biology, and do not take into account abstractions other than time or phase decompositions. Our purpose here is to propose a general computational method for relating models together, by considering primarily the structure of the interactions and abstracting from their dynamics in a first step.

Results: We present a graph-theoretic formalism with node merge and delete operations, in which model reductions can be studied as graph matching problems. From this setting, we derive an algorithm for deciding whether there exists a reduction from one model to another, and evaluate it on the computation of the reduction relations between all SBML models of the biomodels.net repository. In particular, in the case of the numerous models of MAPK signalling, and of the circadian clock, biologically meaningful mappings between models of each class are automatically inferred from the structure of the interactions. We conclude on the generality of our graphical method, on its limits with respect to the representation of the structure of the interactions in SBML, and on some perspectives for dealing with the dynamics.

Availability: The algorithms described in this article are implemented in the open-source software modeling platform BIOCHAM available at http://contraintes.inria.fr/biocham. The models used in the experiments are available from http://www.biomodels.net/.

Contact: francois.fages@inria.fr

PT28—Efficient parameter search for qualitative models of regulatory networks using symbolic model checking

Gregory Batt[1,*], Michel Page[2,3], Irene Cantone[4], Gregor Goessler[2], Pedro Monteiro[2,5] and Hidde de Jong[2]

[1]INRIA Paris - Rocquencourt, Le Chesnay, [2]INRIA Grenoble - Rhône-Alpes, Montbonnot, [3]IAE, Université Pierre Mendès France, Grenoble, France, [4]Clinical Sciences Center, Imperial College, London, UK and [5]INESC/Instituto Superior Técnico, Lisbon, Portugal

Motivation: Investigating the relation between the structure and behavior of complex biological networks often involves posing the question if the hypothesized structure of a regulatory network is consistent with the observed behavior, or if a proposed structure can generate a desired behavior.

Results: The above questions can be cast into a parameter search problem for qualitative models of regulatory networks. We develop a method based on symbolic model checking that avoids enumerating all possible parametrizations, and show that this method performs well on real biological problems, using the IRMA synthetic network and benchmark datasets. We test the consistency between IRMA and time-series expression profiles, and search for parameter modifications that would make the external control of the system behavior more robust.

Availability: GNA and the IRMA model are available at http://ibis.inrialpes.fr/

Contact: gregory.batt@inria.fr

PT29—Bayesian experts in exploring reaction kinetics of transcription circuits

Ryo Yoshida*, Masaya M. Saito, Hiromichi Nagao and Tomoyuki Higuchi

The Institute of Statistical Mathematics, Research Organization of Information and Systems, 10-3 Midori-cho, Tachikawa, Tokyo, 190-8562, Japan

Motivation: Biochemical reactions in cells are made of several types of biological circuits. In current systems biology, making differential equation (DE) models simulatable *in silico* has been an appealing, general approach to uncover a complex world of biochemical reaction dynamics. Despite of a need for simulation-aided studies, our research field has yet provided no clear answers: how to specify kinetic values in models that are difficult to measure from experimental/theoretical analyses on biochemical kinetics.

Results: We present a novel non-parametric Bayesian approach to this problem. The key idea lies in the development of a Dirichlet process (DP) prior distribution, called *Bayesian exper*ts, which reflects substantive knowledge on reaction mechanisms inherent in given models and experimentally observable kinetic evidences to the subsequent parameter search. The DP prior identifies significant local regions of unknown parameter space before proceeding to the posterior analyses. This article reports that a Bayesian expert-inducing stochastic search can effectively explore unknown parameters of *in silico* transcription circuits such that solutions of DEs reproduce transcriptomic time course profiles.

Availability: A sample source code is available at the URL http://daweb.ism.ac.jp/~yoshidar/lisdas/
Contact: yoshidar@ism.ac.jp

PT30—Dynamic deterministic effects propagation networks: learning signalling pathways from longitudinal protein array data

Christian Bender[1,*], Frauke Henjes[1], Holger Fröhlich[2], Stefan Wiemann[1], Ulrike Korf[1] and Tim Beißbarth[3]
[1]Department of Molecular Genome Analysis, German Cancer Research Center, 69120 Heidelberg, [2]Department of Algorithmic Bioinformatics, Bonn-Aachen International Center for IT, 53113 Bonn and [3]Department of Medical Statistics, University of Göttingen, 37099 Göttingen, Germany

Motivation: Network modelling in systems biology has become an important tool to study molecular interactions in cancer research, because understanding the interplay of proteins is necessary for developing novel drugs and therapies. De novo reconstruction of signalling pathways from data allows to unravel interactions between proteins and make qualitative statements on possible aberrations of the cellular regulatory program. We present a new method for reconstructing signalling networks from time course experiments after external perturbation and show an application of the method to data measuring abundance of phosphorylated proteins in a human breast cancer cell line, generated on reverse phase protein arrays.
Results: Signalling dynamics is modelled using active and passive states for each protein at each timepoint. A fixed signal propagation scheme generates a set of possible state transitions on a discrete timescale for a given network hypothesis, reducing the number of theoretically reachable states. A likelihood score is proposed, describing the probability of measurements given the states of the proteins over time. The optimal sequence of state transitions is found via a hidden Markov model and network structure search is performed using a genetic algorithm that optimizes the overall likelihood of a population of candidate networks. Our method shows increased performance compared with two different dynamical Bayesian network approaches. For our real data, we were able to find several known signalling cascades from the ERBB signaling pathway.
Availability: Dynamic deterministic effects propagation networks is implemented in the R programming language and available at http://www.dkfz.de/mga2/ddepn/
Contact: c.bender@dkfz.de

PT31—A novel approach for determining environment-specific protein costs: the case of **Arabidopsis thaliana**

Max Sajitz-Hermstein[1] and Zoran Nikoloski[1,2,*]
[1]Max-Planck Institute of Molecular Plant Physiology and [2]Institute of Biochemistry and Biology, University of Potsdam, Potsdam, D-14476 Germany

Motivation: Comprehensive understanding of cellular processes requires development of approaches which consider the energetic balances in the cell. The existing approaches that address this problem are based on defining energy-equivalent costs which do not include the effects of a changing environment. By incorporating these effects, one could provide a framework for integrating 'omics' data from various levels of the system in order to provide interpretations with respect to the energy state and to elicit conclusions about putative global energy-related response mechanisms in the cell.
Results: Here we define a cost measure for amino acid synthesis based on flux balance analysis of a genome-scale metabolic network, and develop methods for its integration with proteomics and metabolomics data. This is a first measure which accounts for the effect of different environmental conditions. We applied this approach to a genome-scale network of *Arabidopsis thaliana* and calculated the costs for all amino acids and proteins present in the network under light and dark conditions. Integration of function and process ontology terms in the analysis of protein abundances and their costs indicates that, during the night, the cell favors cheaper proteins compared with the light environment. However, this does not imply that there is squandering of resources during the day. The results from the association analysis between the costs, levels and well-defined expenses of amino acid synthesis, indicate that our approach not only captures the adjustment made at the switch of conditions, but also could explain the anticipation of resource usage via a global energy-related regulatory mechanism of amino acid and protein synthesis.
Contact: nikoloski@mpimp-golm.mpg.de

PT33—How threshold behaviour affects the use of subgraphs for network comparison

Tiago Rito[1,2,*], Zi Wang[1], Charlotte M. Deane[1,2] and Gesine Reinert[1,2]
[1]Department of Statistics, University of Oxford, OX1 3TG and [2]OCISB/DTC Systems Biology, Department of Biochemistry, South Parks Road, OX1 3QU, UK

Motivation: A wealth of protein–protein interaction (PPI) data has recently become available. These data are organized as PPI networks and an efficient and biologically meaningful method to compare such PPI networks is needed. As a first step, we would like to compare observed networks to established network models, under the aspect of small subgraph counts, as these are conjectured to relate to functional modules in the PPI network. We employ the software tool GraphCrunch with the Graphlet Degree Distribution Agreement (GDDA) score to examine the use of such counts for network comparison.
Results: Our results show that the GDDA score has a pronounced dependency on the number of edges and vertices of the networks being considered. This should be taken into account when testing the fit of models. We provide a method for assessing the statistical significance of the fit between random graph models and biological networks based on non-parametric tests. Using this method we examine the fit of Erdös–Rényi (ER), ER with fixed degree distribution and geometric (3D) models to PPI networks. Under these rigorous tests none of these models fit to the PPI networks. The GDDA score is not stable in the region of graph density relevant to currentPPI networks. We hypothesize that this score instability is due to the networks under consideration having a graph density in the threshold region for the appearance of small subgraphs. This is true for both geometric (3D) and ER random graph models. Such threshold behaviour may be linked to the robustness and efficiency properties of the PPI networks.
Contact: tiago@stats.ox.ac.uk

Genomic Medicine

PT8—Semi-supervised multi-task learning for predicting interactions between HIV-1 and human proteins

Yanjun Qi[1,*], Oznur Tastan[2], Jaime G. Carbonell[2], Judith Klein-Seetharaman[2] and Jason Weston[3]
[1]NEC Labs America, 4 Independence Way, Princeton, NJ 08540, [2]School of Computer Science, Carnegie Mellon University, PA 15213 and [3]Google Research NY, 75 Ninth Avenue, New York, NY 10011, USA

Motivation: Protein–protein interactions (PPIs) are critical for virtually every biological function. Recently, researchers suggested to use supervised learning for the task of classifying pairs of proteins as interacting or not. However, its performance is largely restricted by the availability of truly interacting proteins (*labeled*). Meanwhile, there exists a considerable amount of protein pairs where an association appears between two partners, but not enough experimental evidence to support it as a direct interaction (*partially labeled*).
Results: We propose a semi-supervised multi-task framework for predicting PPIs from not only *labeled*, but also *partially labeled* reference sets. The basic idea is to perform multi-task learning on a supervised classification task and a semi-supervised auxiliary task. The supervised classifier trains a multi-layer perceptron network for PPI predictions from *labeled* examples. The semi-supervised auxiliary task shares network layers of the supervised classifier and trains with *partially labeled* examples. Semi-supervision could be utilized in multiple ways. We tried three approaches in this article, (i) classification (to distinguish partial positives with negatives); (ii) ranking (to rate partial positive more likely than negatives); (iii) embedding (to make data clusters get similar labels). We applied this framework to improve the identification of interacting pairs between HIV-1 and human proteins. Our method improved upon the state-of-the-art method for this task indicating the benefits of semi-supervised multi-task learning using auxiliary information.
Availability: http://www.cs.cmu.edu/~qyj/HIVsemi
Contact: qyj@cs.cmu.edu

PT9—Candidate gene prioritization based on spatially mapped gene expression: an application to XLMR

Rosario M. Piro*, Ivan Molineris, Ugo Ala, Paolo Provero and Ferdinando Di Cunto
Molecular Biotechnology Center and Department of Genetics, Biology and Biochemistry, University of Torino, Via Nizza 52, 10126 Torino, Italy

Motivation: The identification of genes involved in specific phenotypes, such as human hereditary diseases, often requires the time-consuming and expensive examination of a large number of positional candidates

selected by genome-wide techniques such as linkage analysis and association studies. Even considering the positive impact of next-generation sequencing technologies, the prioritization of these positional candidates may be an important step for disease-gene identification.
Results: Here, we report a large-scale analysis of spatial, *i.e.* 3D, gene expression data from an entire organ (the mouse brain) for the purpose of evaluating and ranking positional candidate genes, showing that the spatial gene-expression patterns can be successfully exploited for the prediction of gene–phenotype associations not only for mouse phenotypes, but also for human central nervous system-related Mendelian disorders. We apply our method to the case of X-linked mental retardation, compare the predictions to the results obtained from a previous large-scale resequencing study of chromosome X and discuss some promising novel candidates.
Contact: rosario.piro@unito.it

PT10—Prediction of a gene regulatory network linked to prostate cancer from gene expression, microRNA and clinical data

Eric Bonnet[1,2], Tom Michoel[1,2] and Yves Van de Peer[1,2,*]
[1]Department of Plant Systems Biology, VIB and [2]Department of Molecular Genetics, Ghent University, Technologiepark 927, B-9052 Gent, Belgium

Motivation: Cancer is a complex disease, triggered by mutations in multiple genes and pathways. There is a growing interest in the application of systems biology approaches to analyze various types of cancer-related data to understand the overwhelming complexity of changes induced by the disease.
Results: We reconstructed a regulatory module network using gene expression, microRNA expression and a clinical parameter, all measured in lymphoblastoid cell lines derived from patients having aggressive or non-aggressive forms of prostate cancer. Our analysis identified several modules enriched in cell cycle-related genes as well as novel functional categories that might be linked to prostate cancer. Almost one-third of the regulators predicted to control the expression levels of the modules are microRNAs. Several of them have already been characterized as causal in various diseases, including cancer.We also predicted novel microRNAs that have never been associated to this type of tumor. Furthermore, the condition-dependent expression of several modules could be linked to the value of a clinical parameter characterizing the aggressiveness of the prostate cancer. Taken together, our results help to shed light on the consequences of aggressive and non-aggressive forms of prostate cancer.
Availability: The complete regulatory network is available as an interactive supplementary web site at the following URL: http://bioinformatics.psb.ugent.be/webtools/pronet/
Contact: yves.vandepeer@psb.vib-ugent.be

PT11—Inferring cancer subnetwork markers using density-constrained biclustering

Phuong Dao[1,†], Recep Colak[1,†], Raheleh Salari[1], Flavia Moser[2], Elai Davicioni[3], Alexander Schönhuth[4,*,‡] and Martin Ester[1,*,‡]
[1]School of Computing Science, Simon Fraser University, Burnaby, [2]University of British Columbia Centre for Disease Control, Vancouver, [3]GenomeDX Biosciences Inc., Vancouver, BC, Canada and [4]Department of Mathematics, University of California, Berkeley, CA, USA

Motivation: Recent genomic studies have confirmed that cancer is of utmost phenotypical complexity, varying greatly in terms of subtypes and evolutionary stages. When classifying cancer tissue samples, subnetwork marker approaches have proven to be superior over single gene marker approaches, most importantly in cross-platform evaluation schemes. However, prior subnetwork-based approaches do not explicitly address the great phenotypical complexity of cancer.
Results: We explicitly address this and employ *density-constrained biclustering* to compute subnetwork markers, which reflect pathways being dysregulated in many, but not necessarily all samples under consideration. In breast cancer we achieve substantial improvements over all cross-platform applicable approaches when predicting TP53 mutation status in a well-established non-cross-platform setting. In colon cancer, we raise prediction accuracy in the most difficult instances from 87% to 93% for cancer versus non-cancer and from 83% to (astonishing) 92%, for with versus without liver metastasis, in well-established cross-platform evaluation schemes.

Availability: Software is available on request.
Contact: alexsch@math.berkeley.edu; ester@cs.sfu.ca

PT12—Modeling associations between genetic markers using Bayesian networks

Edwin Villanueva and Carlos Dias Maciel*
Electrical Engineering Department, Sao Carlos School of Engineering, University of Sao Paulo, Sao Carlos, Sao Paulo, Brazil

Motivation: Understanding the patterns of association between polymorphisms at different loci in a population (linkage disequilibrium, LD) is of fundamental importance in various genetic studies. Many coefficients were proposed for measuring the degree of LD, but they provide only a static view of the current LD structure. Generative models (GMs) were proposed to go beyond these measures, giving not only a description of the actual LD structure but also a tool to help understanding the process that generated such structure. GMs based in coalescent theory have been the most appealing because they link LD to evolutionary factors. Nevertheless, the inference and parameter estimation of such models is still computationally challenging.
Results: We present a more practical method to build GM that describe LD. The method is based on learning weighted Bayesian network structures from haplotype data, extracting equivalence structure classes and using them to model LD. The results obtained in public data from the HapMap database showed that the method is a promising tool for modeling LD. The associations represented by the learned models are correlated with the traditional measure of LD *D′*. The method was able to represent LD blocks found by standard tools. The granularity of the association blocks and the readability of the models can be controlled in the method. The results suggest that the causality information gained by our method can be useful to tell about the conservability of the genetic markers and to guide the selection of subset of representative markers.
Availability: The implementation of the method is available upon request by email.
Contact: maciel@sc.usp.br

Other Bioinformatics Applications

PT13—Mass spectrometry data processing using zero-crossing lines in multi-scale of Gaussian derivative wavelet

Nha Nguyen[1,2], Heng Huang[1,*], Soontorn Oraintara[2] and An Vo[3]
[1]Department of Computer Science and Engineering, [2]Department of Electrical Engineering, University of Texas at Arlington, TX and [3]The Feinstein Institute for Medical Research, North Shore LIJ Health System, New York, USA

Motivation: Peaks are the key information in mass spectrometry (MS) which has been increasingly used to discover diseases-related proteomic patterns. Peak detection is an essential step for MS-based proteomic data analysis. Recently, several peak detection algorithms have been proposed. However, in these algorithms, there are three major deficiencies: (i) because the noise is often removed, the true signal could also be removed; (ii) baseline removal step may get rid of true peaks and create new false peaks; (iii) in peak quantification step, a threshold of signal-to-noise ratio (SNR) is usually used to remove false peaks; however, noise estimations in SNR calculation are often inaccurate in either time or wavelet domain. In this article, we propose new algorithms to solve these problems. First, we use bivariate shrinkage estimator in stationary wavelet domain to avoid removing true peaks in denoising step. Second, without baseline removal, zero-crossing lines in multi-scale of derivative Gaussian wavelets are investigated with mixture of Gaussian to estimate discriminative parameters of peaks. Third, in quantification step, the frequency, SD, height and rank of peaks are used to detect both high and small energy peaks with robustness to noise.
Results: We propose a novel Gaussian Derivative Wavelet (GDWavelet) method to more accurately detect true peaks with a lower false discovery rate than existing methods. The proposed GDWavelet method has been performed on the real Surface-Enhanced Laser Desorption/Ionization Time-Of-Flight (SELDI-TOF) spectrum with known polypeptide positions and on two synthetic data with Gaussian and real noise. All experimental results demonstrate that our method outperforms other commonly used methods. The standard receiver operating characteristic (ROC) curves are used to evaluate the experimental results.
Availability: http://ranger.uta.edu/~heng/MS/GDWavelet.html or http://www.naaan.org/nhanguyen/archive.htm
Contact: heng@uta.edu

PT14—Detecting host factors involved in virus infection by observing the clustering of infected cells in siRNA screening images

Apichat Suratanee[1,2,†], Ilka Rebhan[3,†], Petr Matula[1,2], Anil Kumar[3], Lars Kaderali[4], Karl Rohr[1,2], Ralf Bartenschlager[3], Roland Eils[1,2,*] and Rainer König[1,2,*]

[1]Department of Bioinformatics and Functional Genomics, Institute of Pharmacy and Molecular Biotechnology, Bioquant, University of Heidelberg, INF 267, [2]Division of Theoretical Bioinformatics, German Cancer Research Center (DKFZ), INF 280, [3]The Department for Infectious Diseases, Molecular Virology, University of Heidelberg, INF 345 and [4]Viroquant Research Group Modeling, Bioquant, University of Heidelberg, INF 267, 69120 Heidelberg, Germany

Motivation: Detecting human proteins that are involved in virus entry and replication is facilitated by modern high-throughput RNAi screening technology. However, hit lists from different laboratories have shown only little consistency. This may be caused by not only experimental discrepancies, but also not fully explored possibilities of the data analysis.We wanted to improve reliability of such screens by combining a population analysis of infected cells with an established dye intensity readout.

Results: Viral infection is mainly spread by cell–cell contacts and clustering of infected cells can be observed during spreading of the infection *in situ* and *in vivo*. We employed this clustering feature to define knockdowns which harm viral infection efficiency of human Hepatitis C Virus. Images of knocked down cells for 719 human kinase genes were analyzed with an established point pattern analysis method (Ripley's *K*-function) to detect knockdowns in which virally infected cells did not show any clustering and therefore were hindered to spread their infection to their neighboring cells. The results were compared with a statistical analysis using a common intensity readout of the GFP-expressing viruses and a luciferase-based secondary screen yielding five promising host factors which may suit as potential targets for drug therapy.

Conclusion: We report of an alternative method for high-throughput imaging methods to detect host factors being relevant for the infection efficiency of viruses. The method is generic and has the potential to be used for a large variety of different viruses and treatments being screened by imaging techniques.

Contact: r.eils@dkfz.de; r.koenig@dkfz.de

POSTERS

Sequence Analysis, Alignment and Next Generation Sequencing

A-1 Resolving the genomic breakpoints of a microdeletion using next generation sequencing in a family

Approaches to finding the exact breakpoints of the deletion causing a recurrent mental retardation syndrome. Combining microarray, mate pair sequencing and familial data.

—de Ligt J, Vissers LELM, Kloosterman W, Hehir-Kwa JY, Bruijn E, Gilissen C, Guryev V, Cuppen E , Brunner HG, Veltman JA

A-2 Identification of unknown environmental sequences by biased distribution of signature oligonucleotides

We present a database of signature oligonucleotides and software tools that provides a graphical user interface to the database. These tools allow one to manage the database and infer the phylogeny of unknown DNA sequences.

—Labuschange P, Emmett W, Davenport CF, **Reva ON**

A-3 Identification of conserved repeats with advanced background models

In a survey of single amino acid repeats in over 100 eukaryotes, leucine repeats are unexpectedly frequent in signal peptides according to an application specific background model. The relatively stronger conservation of the repeats suggests a functional role.

—Łabaj P-P, Sykacek P, Kreil D-P

A-4 Searching for structural homologues of LeuT-like secondary transporters beyond sequence similarity

We present AlignMe, a program that can use various types of information for alignment of two membrane protein sequences. We use AlignMe to detect relationships between transporters that share a structural fold but lack detectable sequence similarity.

—Khafizov K, Staritzbichler R, Stamm M, Forrest LR

A-5 Why inverse proteins are relatively abundant

Using a new artificial set of peptide sequences, we suggest the relative abundance of inverse proteins can be explained by the fact they display the same repeat structures and amino acid propensity of existing proteins.

—**Nebel J-C**

A-6 HERD: the highest expected reward decoding for HMMs with application to recombination detection

We present a new approach for decoding hidden Markov models that can be tailored to a particular application, and therefore can achieve better prediction accuracy than traditional decoding methods, such as the Viterbi algorithm.

—Nánási M, Vinař T, Brejová B

A-7 Visualizing the next generation of sequencing data with GenomeView

GenomeView can interactively browse high-throughput sequencing data and genome alignments of dozens of genomes with dynamic navigation and zooming. It can handle dozens of aligned genomes, thousands of annotation features and millions of short reads

—Abeel T, Van Parys T, Galagan J, Van de Peer Y

A-8 IsUnstruct: a method based on Izing's model for prediction of disordered residues from protein sequence alone.

A new method based on the dynamic programming was developed for searching disordered residues in protein chain. Testing of the method on two databases has shown that our method is one of the best.

—Lobanov MY, **Galzitskaya OV**

A-9 MASiVE: Mapping and analysis of SireVirus elements in plant genome sequences

We present MASiVE, an algorithm for the large-scale, yet sensitive and highly accurate, discovery of intact Sirevirus LTR-retrotransposons in plant genomic sequences. MASiVE was validated on the annotated large Zea mays chromosome one.

—Darzentas N, Bousios A, Apostolidou V, Tsaftaris AS

A-10 Signals in human promoter sequences

We have located GC rich 6-nts in human promoters and identified sigmoidal behavior in their distribution.These hexamers are also observed in miRNAs and TF binding sites.We aim to establish

correlation among promoters, TFs and miRNAs,based on 6-nts.
—**Putta P**

A-11 Iterative read mapping and assembly allows the use of more distant reference in metagenome assembly

We assemble a quasispecies consensus genome from 32nt metagenomic reads with 91.1% identity to the original reference. Iterating the mapping and assembly procedure allows us to approach the consensus of the sequenced community.
—Dutilh BE, Huynen MA, Gloerich J, Strous M

A-12 APPRIS, a database of annotated principal splice variants for the human genome

We have developed a database that deploys a range of computational methods to provide value to the annotations of the human genome. The database selects one of the CDS for each gene as the principal functional variants based on these algorithms.
—Rodríguez J-M, Ezkurdia I, Lopez G, Pietrelli A, Maietta P, Wesselink J-J, Valencia A, Tress M

A-13 Accurate long read mapping using enhanced suffix arrays

We present an algorithm to map long reads to reference genomes, based on chaining maximal exact matches and using heuristics and the Needleman-Wunsch algorithm to bridge the gaps. It was tested on simulated BRCA1 data with positive results.
—Vyverman M, Fack V, De Schrijver J, Dawyndt P

A-14 Detection and architecture of small heat shock protein monomers

We studied the architecture of small Heat Shock Proteins monomers, based on a new approach that detects and delineates the alpha-crystallin domain . We annotated ACD structural components and quantified sHSPs properties.
—Poulain P, Gelly J-C, Flatters D

A-15 Assembly of mitochondrial genomes using next-generation sequencing data

We describe the identification and analysis of complete mitochondrial genomes using next-generation data derived from whole-genome shotgun sequencing of two isolates of the plant pathogen Rhynchosporium secalis.
—Felder M, Taudien S, Groth M, Münsterkötter M, Navarro-Quezada A, Knogge W, Platzer M

A-16 AnsNGS: Annotation system of structural variation for next generation sequencing data

We introduce a new data integration and visualization tool AnsNGS to visualize short read mapping, identify and annotate genetic variation based on the reference genome.
—Na YI, Park CH, Cho Y, Kim, JH

A-17 Assessing the accuracy and completeness of the Bonobo genome sequence

We present the detailed analysis of the quality of the Bonobo genome sequence, assembled from 25x coverage of 454 sequencing data, by comparison to the human and chimpanzee genomes.
—Prüfer K, Ptak SE, Fischer A, Good JM, Mullikin JC, Miller J, Kodira CD, Knight JR, The Bonobo Genome Consortium, Kelso J, Pääbo S

A-18 Estimating effective DNA database size via compression

Biological databases often contain identical or very similar sequences. This is not taken into account in estimation of statistical significance of an alignment. We propose to use the compressed size of database as an estimate of its effective size.
—Visnovska M, Nanasi M, Vinar T, Brejova B

A-19 ChIP-seq analysis: from peaks to motifs

We present ChIP-motifs, a specialized pipe-line combining various algorithms for discovering cis-regulatory motifs in massive ChIP-seq datasets.
—Thomas-Chollier M, Defrance M, Sand O, Herrmann C, Thieffry D, van Helden J

A-20 Applying mass spectrometry data to improve annotation of the reference mouse genome

We have constructed a novel genome annotation pipeline for tandem mass spectrometry data and have applied this to the reference mouse genome, enabling both the validation of current annotation and the identification of novel protein-coding loci.
—Saunders GI, Brosch M, Frankish A, Collins MO, Yu L, Harrow J, Choudhary JS, Hubbard T

A-21 Prediction of splice-modifying SNPs in human genes using a new analysis pipeline called AASsites

Most tools for the annotation of single nucleotide polymorphisms analyze possible effects on the transcriptome and proteome levels. In contrast, our AASsites pipeline indicates if a change in the splice pattern due to a SNP is likely to occur.
—Hotz-Wagenblatt A, Faber K, Risch A, Glatting K-H

A-22 Structator: a tool for fast index-based matching of RNA structural motifs

Structator is a tool for ultra-fast matching of RNA structural motifs. It employs affix arrays, which is an index data structure that allows for bidirectional search of RNA structural motifs while taking advantage of their symmetric nature.

—Meyer F, Will S, Beckstette M

A-23 Project HOPE: Providing the last piece of the puzzle...

We present Project HOPE; an automatic analysis method for point mutations. HOPE collects and combines protein information from a series of sources and produces a conclusion while focusing on the structural effects.

—**Venselaar H**

A-24 Simple not-PWM-based approach for transcription factor binding site prediction

We suggest a novel simple not-PWM-based approach to the detection of transcription factor binding sites. It outperforms other tools in both sensitivity and specificity and allows to significantly reduce the number of false positives

—Fazius E, Shelest V, Shelest E

A-25 PICMI: mapping point variations on genomes

PICMI is a web server to simplify the mapping of nucleotide variations on a genome, to retrieve information about their effect on all isoforms of a gene and, to rapidly add value to existing amino acidic variation data.

—Le Pera L, Marcatili P, Tramontano A

A-26 Detection of alternative splice isoforms in the human genome

We have carried out a comprehensive re-identification of proteomics spectra from several data repositories and confirmed that many human genes express multiple alternative isoforms in sufficient quantities to be detected in proteomics experiments.

—Ezcurdia I, del Pozo A, Rodriguez JM, Ashman K, Valencia A, Tress ML

A-27 Boundaries for short-read, low-coverage deNovo assembly and resequencing

A set of formulas are presented to calculate the expected distribution of contig length for a short-read sequencing experiment before performing it. Boundaries can be provided by these formulas to reduce the resources needed.

—Schatz F, Schimmler M

A-28 TAPIR, a web server for the prediction of plant microRNA targets, including target mimics

We present a new web server called TAPIR, designed for the prediction of plant microRNA targets, including target mimics. The TAPIR web server can be accessed at:

http://bioinformatics.psb.ugent.be/webtools/tapir

—Bonnet E, He Y, Billiau K, Van de Peer Y

A-29 GRID/CPCA analysis of the SULTs superfamily: multivariate characterization of the most relevant structural and physicochemical differences

We used computational methods to provide a multivariate characterization of human and mouse cytosolic sulfotransferases,revealing the most relevant structural and physicochemical differences within the SULT family towards to substrate specificity.

—**Rebollido-Rios R**, Zamora-Rico I

A-30 Theoretical and empirical quality assessment of transcription factor binding motifs

We propose a method, implemented in the program matrix-quality, that combines theoretical and empirical score distributions to assess the reliability of position-specific matrices for predicting transcription factor binding sites.

—Medina-Rivera A, Abreu-Goodger C, Thomas-Chollier M, Salgado-Osorio H, Collado-Vides J, van Helden J

A-31 Characterization and prediction of protein nucleolar localization sequences

By extensive curation of the literature, we have uncovered 46 human nucleolar localization sequences . These sequences were used to create an artificial neural network predictor which accurately predicts NoLSs and distinguishes them from NLSs.

—Scott MS, Boisvert F-M, McDowall MD, Lamond AI, Barton GJ

A-32 Jalview: past, present, future.

Jalview is a widely used open-source program for sequence alignment visualization. In this poster, we describe its new annotation and web service capabilities, and our plans for its future development.

—Procter JB, Troshin PV, Martin MAM, Barton GJ

A-33 Enrichment of eukaryotic linear motifs in viral proteins.

We have found that there is an evidence for enrichment of PCNA binding motif in the ordered region and PDZ ligand binding motifs in the

disorder region of 33,269 proteins from 1,646 viral genomes.

—Pushker R, Edwards R-J, Shields D-C

A-34 Analyzing ChIP-Seq data using qips

We propose a new tool called 'qips' especially suited for the analysis of ChIP-Seq data also containing broader enriched regions which occur for example when targetting polymerase or histone marks.

—**Gogol-Döring A**, Chen W

A-35 COMA server for protein distant homology search

Detection of distant homology is a widely used computational approach for studying proteins. Here we report a homology search web server based on sequence profile comparisons. The server capabilities are illustrated with some non-trivial examples.

—Margelevičius M, Laganeckas M, Venclovas Č

A-36 Protein disorder and short motifs in disordered regions are enriched near the cytoplasmic side of single-pass transmembrane proteins

Human single-pass transmembrane proteins were analyzed to predict intrinsic disorder and potentially functional short linear motifs.

—Stavropoulos I, Khaldi N, Davey NE, Finian M, Shields DC

A-37 Variation in positional preferences by 20 amino acids in alpha helices with different size.

We have dissected alpha helixes with different number of amino acids . We then compared the AA compositions of each position for each class of helices. We have also compared the AA contents of the same position in different helices.

—Fallahi H, Dokanehiifard S-A, Hosseini-Nasab S-M-E

A-38 Comparison of mapping and variant calling accuracies for next-generation sequence data in relation to coverage depth: a case study in leukemia cell lines.

We have assembled nine analysis pipelines to analyze targeted sequencing data and assessed their performances under high and low coverage thresholds separately and showed that it is possible to identify correct variations even in low coverage regions

—Kalender Z, Geerdens E, Cuppens H, Cools J, Aerts S

A-39 Monitoring strain diversity in metagenomics data using meta-MLST SNP profiling

Monitoring bacterial strain diversity in metagenomics sequence data, is typically done using 16S rRNA sequences supporting resolution up to the genus level. We describe a method that allows following diversity in metagenomes at the strain level.

—De Jager VCL, Van der Sijde MR, Hazelwood LA, Kutahya O, Kleerebezem M, Smid EJ, Siezen RJ, Van Hijum SAFT

A-40 A variance stabilizing parametrization to detect differentially expressed genes in RNA-Seq data

Processing of RNA-Seq data is currently affected by a transcript length bias in the calling of differentially expressed genes. Here we propose a novel approach based on a variance stabilizing parametrization and on an unbiased Wald type test.

—Sales G, Risso D, Chiogna M, Romualdi C

A-41 DNA annotation induction: from RefGene on human chr. 22 to genome-wide CAGE for human and mouse

We have shown that the knowledge of functional annotation such as RefGene locations of transcription start sites on the smallest autosomal human chromosome is sufficient for genome-wide prediction of CAGE tags, for human and also mouse.

—Bedo J, Steininger A, Haviv I, Kowalczyk A

A-42 A new resampling method

We introduced a new, biologically inspired resampling method to use in molecular sequences. The method is crucial especially for the non alignment based methods.

—Bakis Y, Sezerman OU

A-43 Clustering massive sequence databases

We developed a tool for clustering massive number of DNA sequences. This tool has been implemented in BioloMICS, a software solution for biological data. Experiments show promising results.

—Vu D, Robert V

A-44 WordCluster: detecting clusters of DNA words and genomic elements

WordCluster finds statistical significant clusters of DNA words or any other genomic elements. The co-localization with gene annotations and the enrichment/depletion analysis in GO terms may facilitate relating them to biological function.

—**Barturen G**, Oliver JL, Hackenberg M

A-45 Analysis of copy loss and gain variations in Holstein cattle autosomes using BeadChip SNPs

Using the BovineSNP50 BeadChip for 912 bulls, we analyzed copy losses and gains based either on deviation from the expected Hardy-Weinberg equilibrium or on signal intensity using PennCNV. This approach was more effective than previous methodologies.

—Seroussi E, Glick G, Shirak A, Yakobson E, Weller JI, Ezra E, Zeron Y

A-46 The role of the Occludin MARVEL domain in intracellular targeting and clustering

Occludin, a MARVEL domain protein is an essential component of tight junctions. Bioinformatic sequence analysis was applied in conjunction with live cell imaging to study the functions of this protein-lipid interacting domain.

—Yeheskel A, Yaffe Y, Hirschberg K, Pasmanik-Chor M

A-47 Multi-Harmony: detecting functional specificity from sequence alignment

Specialized functions in protein sub-families generally are conferred by a small number of amino acids. Our improved multi-Harmony web server reliably detects these sub-type-specific sites from an input multiple alignment and sub-grouping.

—**Feenstra KA**, Brandt BW, Heringa J

A-48 Fast approximate statistical ranking for sequence motif discovery under higher order Markov background model

A new method of motif p-value estimation, based on a polynomial approximation and a MCMC optimization of the coefficients, is proposed. Higher order background correlation can be handled with minimum sacrifice of precision and complexity.

—**Shida K**

A-49 Networking interacting residues, evolutionary conservation and energetics in protein structures

Residue interaction networks consider single amino acids in the protein structure as nodes and connections as physico-chemical interactions. Here, we present RING as a novel tool to generate RINs for use in Cytoscape.

—Vidotto M, Martin AJM, Boscariol F, Walsh I, Tosatto SCE

A-50 An iterative scheme for making long structural alignments of ncRNAs using FOLDALIGN.

Pairwise structural alignment of ncRNAs is a computational very demanding task. To lower these requirements, we present an iterative approach which recursively extends shorter alignments. The approach applies to both local and global alignments.

—**Havgaard JH**, Gorodkin J

A-51 A multi-objective approach to the prediction of sRNAs in Bacteria

We present a methodology, which uses a multi-objective approach to extract the best methods' aggregations to predict sRNAs in Bacteria by maximizing the specificity and sensitivity of the program´s individual predictions.

—Arnedo J, Romero-Zaliz R, del Val C

A-52 Scalability of large-scale protein domain family inference

We present MPI_MkDom2, a distributed algorithm for protein domain family inference enabling the construction of regular new releases of ProDom, while preserving the structure of domain families built by the now inapropriate sequential method MkDom2.

—Rezvoy C, Vivien F, Kahn D

A-53 Evolutionary study of eye-developmental gene expression across multiple Drosophila species by high-throughput tag sequencing.

We have used Illumina's digital gene expression in imaginal discs of three Drosophila species. We show that DGE produces accurate measures of gene expression and we identify evolutionary conserved and divergent expression during eye development

—**Naval-Sanchez M**, Aerts S

A-54 Higher order repeat with largest primary repeat unit and located within a gene in human genome

Presentation of our novel computational method Global Repeat Map, which is particularly convenient for identification and analysis of repeats and higher order repeats characterized by very long repeat units on a case study of Build 37.1 genomic ensemble of Y chromosome.

—Glunčić M, Paar V, Paar P, Rosandić M, Vlahović I

A-55 Fast and accurate digging for binding motifs in ChIP-Seq data using ChIPMunk software

We present the ChIPMunk tool for motif discovery in ChIP-Seq data using base coverage information. Comparison with other motif discovery tools shows that ChIPMunk identifies the correct motifs of the same or better quality and works much faster.

—Kulakovskiy I V, Boeva VA, Favorov AV, Makeev V J

A-56 Mebitoo - A sequence analysis toolbox framework

We present Mebitoo, a bioinformatics software tool to embrace disequence analysis methods within a uniform graphical user interface. The framework simplifies both the development of methods and the workflow when assessing large sets of sequences.

—**Spaniol C**, Helms V

A-57 Practical aspects of RNA-Seq data analysis that you should know and they don't tell you

We present a guide of practical problems for the analysis of RNA-Seq data. We analysed a variety of published data by means of state-of-the-art software and provide valuable hints that should be taken into consideration on these kind of analysis

—Garcia F, Tarazona S, Zuñiga S, Santoyo J, Dopazo J, Conesa A

A-58 MutaBase: a framework and web-interface for the archival, management and interpretation of single nucleotide variants obtained by next generation sequencing.

MutaBase is an easy-to-use, flexibly queryable framework which assists in the discovery of interesting single-nucleotide variants called by next-generation sequencing platforms. We here present its application in full-exome sequencing of 4 patients

—Sifrim A, Van Houdt J, Vermeesch J, Moreau Y

A-59 Mining cleavage sites of the mouse peptidome

We present the application of constraint-based itemset mining techniques on peptide cleavage sites. Mining experimentally verified cleavage sites allows discovery of recognition motifs for endopeptidases.

—De Grave K, Guns T, Vandekerckhove TTM, Landuyt B, Menschaert G, Van Criekinge W, Schoofs L, Luyten W

A-60 Properties of plant microRNAs affect miRDeep statistical scoring as well as accuracy

Given the differences in properties of plant miRNAs from that of animal miRNAs, we examined their effect on key prediction measures used by programs like miRDeep, and estimated the relevant parameters.

—Thakur V, Wanchana S, Xu M, Mosig A, Zhu X

A-61 Using the Amazon Elastic Cloud Computing Resource for the analysis of next-generation sequencing data in Ensembl

The Ensembl project has ported parts of its Next-Generation Sequence analysis pipeline and web services to the Amazon Cloud. With this infrastructure in place, the analysis and visualisation of large data sets is attractive for smaller labs.

—Vogel J, Aken B, Chiang G-T Clapham P, Coates G, Fairley S, Hourlier T, Ruffier M, Tang A, White S, Zadissa A, Searle S, Hubbard T

A-62 Automated sequence extraction of relevant genomic features for targeted resequencing

We present an easy to use webtool for genomic target selection based on the most complete and uniform annotation sources currently available. The tool focuses on usability and should allow to easily prepare next generation sequencing experiments.

—De Wilde B, D'Hont B, Lefever S, Hellemans J, Speleman F, Pattyn F, Vandesompele J

A-63 An Ensembl-based pipeline for microRNA prediction and expression profiling using next generation sequencing data

We present a workflow appliance for predicting miRNA loci and profiling miRNA expression based on short read sequences generated from small RNA libraries. Is is based on Ensembl eHive and incorporates a number of open-source miRNA analysis software.

—**James N**, Donepudi M, Spooner W, Watson M

A-64 The Goby file formats: towards scalable next-generation sequencing data analysis

The poster will present an overview of the Goby framework and provide the results of a storage benchmark comparing Goby formats to widely used next-gen file formats.

—Chambwe N, Dorff KC, Srdanovic M, Deng M, Andrews SJD, **Campagne F**

A-65 Quantitative transcriptome analysis using de novo assembly of RNA-seq from the non-model C4 species Cleome gynandra

Here we describe the de novo assembly of Illumina-based RNA sequencing data to establish the Cleome gynandra transcriptome, and to assess global changes in gene expression during leaf development.

—**Salmon-Divon M**, Aubry S, Rutherford K-M, Kelly K-A, Hibberd J-M, Bertone P

A-66 iCount – comprehensive analysis of iCLIP data

iCount is a computational pipeline for the analysis of individual-nucleotide resolution UV-crosslinking and immunoprecipitation (iCLIP) data: mapping, quantification of bound RNA, peak finding, binding sequence motif enrichment and visualization.

—Rot G, König J, Gorup C, Zupan B, Ule J, **Curk T**

A-67 Relating genomic variation to drug response in childhood acute lymphoblastic leukemia by multiplexed targeted sequencing.

We present a reproducible, cost-effective, multiple-sample pooling customization of a genome enrichment technique and demonstrate its applicability for custom SNP genotyping experiments using childhood acute lymphoblastic leukemia as an example.

—**Wesolowska A**, Dalgaard MD, Borst L, Gautier L, Bak M, Weinhold N, Nielsen BF, Nersting J, Tommerup N, Brunak S, Sicheritz-Ponten T, Leffers H, Schmiegelow K, Gupta R

A-68 Exploring 3D pooled next-generation sequencing using SNP-Cub^3

We developed a simulation and analysis pipeline - called SNP-Cub^3 - for 3D pooled illumina resequencing that allows optimization and scaling of the multidimensional pools, optimization of the sequencing setup and analysis of the sequencing data.

—**De Schrijver J**, De Wilde B, Trooskens G, Vandesompele J, Van Criekinge W

Comparative Genomics, Phylogeny and Evolution

B-1 The role of transposable elements in the evolution of non-mammalian vertebrates and invertebrates

We present a comparative analysis of transposable elements exonization in five different species. our analysis shed light on the differences between vertbrates and invertbrates regarding the birth of new exons.

—Sela N, Kim E, Ast G

B-2 A tool for comparison of complete proteomes between pairs of organisms

We present a tool that implements an algorithm that permits an easy comparison of the full proteomes between two organisms. This tool is implemented in PERL and will be made available both on a website and as a set of downloadable scripts.

—**Karathia HM**, Alves R

B-3 Toward an unified measure of intraspecific selective pressure

We present an attempt to unify some of the most used measures of intraspecific selective pressure. By their exhaustive analysis, we highlighted their pros and cons, suggesting a way for their integration.

—Amato R, Miele G, Pinelli M, Cocozza S

B-4 Assessment of genetic structure within and among Normandie populations of wild beet

We assess the level of genetic differentiation in B. vulgaris sps. maritima form coastal northern France at three different geographical scales.

— Tran HT, Cuguen J, Touzet P, Saumitou-Laprade P

B-5 Trees inside trees for genome wide association studies

We consider extensions of Random Forests candidate splits for taking into account linkage desequilibrium between SNP markers in genome-wide association studies. We obtain improved predictive power and interpretability on artificial and real datasets.

—Botta V, Geurts P, Wehenkel L

B-6 Validation of single cell arrayCGH on a 60-mer oligo microarray platform for preimplantation genetics diagnosis

We present a novel research to detect copy number variations of single cell DNA derived from blastomeres. This research is helpful for the future study to improve the selection of high quality human embryos before implantation.

—Cheng J, Vanneste E, Konings P, Van Eyndhoven W, Voet T, Ampe M, Verbeke G, Vermeesch J, Moreau Y

B-7 Polymorphisms associated with mtDNA of elite Kenyan endurance athletes

We analyzed the complete mtDNA sequences of elite Kenyan endurance athletes to look for SNPs correlated with excellence in international competition. We have identified 4 polymorphisms in members of the L0 clade which are significantly enriched.

—Seiler M, Fuku N, Diao L, Solovyov A, Seiler S, Tanaka M, Bhanot G

B-8 Bayesian inference of community average gene copy numbers and its application for the metagenomic characterization of bacterioplankton community types in the Sargasso Sea

We present a Bayesian approach for estimating community average gene copy numbers for comparative metagenomic analyses, and illustrate its use in a comparative metagenomic study of microbial communities from the Sargasso Sea.

—Beszteri B, Frickenhaus S, Giovannoni SJ

B-9 Phylogenetic mapping of non-model organism RNA-seq reads using a graph algorithm

We present a new methodology to map RNA-seq datasets for non-model organisms that don't have a reference genome to reference gene family alignments. The resulting fragments can then be assembled into transcripts, complementing existing de novo tools.

—Vilella A, Massingham T, Loytynoja A

B-10 GEI-DB: A database of atypical genomic elements in bacteria

We present GEI-DB, a database of genomic islands that were identified in bacteria by the analysis of genome-wide distribution of tetra-nucleotide patterns using oliginucleotide usage statistics.

—Bezuidt O, Lima-Mendez G, Reva O

B-11 The CNS twilight zone: limitations in comparing upstream regions in plants

The limits of successful genomic comparisons to understand upstream regions of plants are dependent on the twilight zone shaped by divergence time, relative positioning, differences between plant groups, and duplication events.

—Reineke AR , Bornberg-Bauer E, Gu J

B-12 TurboOrtho – a high performance alternative to OrthhoMCL

We have re-implemented in C++ the popular OrthoMCL algorithm for protein clustering. This resulted in 14-216 times increase in the performance.

—Ekseth O, Lindi B, Kuiper M, Mironov V

B-13 i-ADHoRe 3.0 Detection of collinearity in large scale datasets.

We present a new version of i-ADHoRe capable of detecting collinearity in large datasets. The combination of algorithmical improvements and support for MPI and multithreading allows the efficient analysis the full Ensembl database .

—Proost S, Fostier J, Dhoedt B, Demeester P, Van de Peer Y, Vandepoele K

B-14 Comparative expression analysis of orthologous genes between Arabidopsis and rice

In this study we developed a general framework to measure expression context conservation for orthologs between Arabidopsis and rice and show that for several gene functions sequence and expression evolution are only weakly correlated.

—Movahedi S, Van de Peer Y, Vandepoele K

B-15 Phylogenomics and robust construction of prokaryotic evolutionary trees

We show that our method of construction of prokaryotic evolutionary trees based on sets of selected genes produces very robust results. The sets are constructed from as many as possible clusters of orthologous groups of genes.

—Korenblat K, Volkovich Z, Bolshoy A

B-16 Mapping the human membrane proteome: a majority of the human membrane proteins can be classified according to function and evolutionary origin.

Membrane proteins are essential in all cells and are also the targets for the majority of our drugs. In this study we present a detailed overview of the members of the human membrane proteome.

—Almén MS, Nordström KJ, Fredriksson R, Schiöth HB

B-17 Towards molecular trait-based ecology through integration of biogeochemical, geographical and metagenomic data

Metagenomics is a powerful tool in environmental and medical microbiology . Here we show that an ecosystems 'parts list' can be lead to ecological knowledge through integration of the Venter GOS study with external data.

—Raes J, Letunic I, Yamada T, Jensen LJ, Bork P

B-18 Orthologous proteins associated with yeast telomeric complex identified by synteny and sequence similarity

We attempt to identify orthologs in several yeast species for proteins known to have function in maintenance of telomeric complex. By combination of sequence similarity and synteny information we evaluate confidence for putative orthologs.

—Macko M, Tomáška L, Vinař T

B-19 Estimating average heterozygosity using multilocus dominant marker data: a maximum likelihood approach via the EM algorithm.

We present an Expectation-Maximization algorithm for obtaining the maximum likelihood estimate of average heterozygosity under presence of ascertainment bias, using multilocus dominant marker data.

—**Khang TF**, Yap VB

B-20 PLAZA 2.0 : a resource for plant comparative genomics

We present an update of the PLAZA plaform for comparative genomics. Besides a substantial increase in number of available species, various tools have been updated and new tools have been added.

—Van Bel M, Proost S, Wischnitzki E, Van de Peer Y, Vandepoele K

B-21 Branch testing in phylogenetic trees: comparison of computationally efficient methods

We compare several methods to test branches in phylogenetic trees in terms of false positive and false negative rate, computational efficiency and robustness to the underparametization of the evolutionary model.

—Czarna A, Wróbel B

B-22 Parsimonious higher-order hidden Markov models for enhanced comparative genomics

An efficient extension of the standard Hidden Markov Model is developed for the analysis of Array-CGH data. The benefits of this method over others are demonstrated for the genomes of two important Arabidopsis ecotypes.

—Seifert M, Banaei A, Gohr A, Strickert M, Grosse I

B-23 Correction of bootstrap confidence levels using an iterated bootstrap procedure with computational efficient methods.

We apply iterated bootstrap proposed by Rodrigo to correct bootstrap confidence levels using two computationally efficient methods for the reconstruction of pseudotrees.

—Czarna A, Wróbel B

B-24 Molecular evolution, structure and functional divergence of Lipoxygenase gene family in vertebrate.

Lipoxygenase comprise a multi-gene family of non-heme, iron containing dioxygenases. This study describes the phylogenetic relationships, structural and functional divergence of vertebrate Lipoxygenase gene family with special emphasis on mammalia.

—Padmanabhan R, Kuhn H, Reddanna P

B-25 Cassis: detection of genomic rearrangement breakpoints

Cassis is a software developed for the precise detection of genomic rearrangement breakpoints by the comparison of the genomes of two related species. It detects and narrows breakpoints and statistically assesses the obtained results.

—Baudet C, Lemaitre C, Dias Z, Gautier C, Tannier E, Sagot M-F

B-26 Methylated cytosines are less likely to mutate within CpG islands

5mCpGs are over 1.5-times less likely to mutate into TpG within CpG islands. Possible causes of this effect include a higher percentage of deleterious C>T mutations in CpG islands, or interactions between 5mCpG mutation rates and local GC content.

—Medvedeva Y, Panchin A, Mitrofanov S, Makeev V

B-27 Entropy approach reveals new features of genes and genomes

We present an investigation of entropy characteristics for 1031 bacterial, 89 archaeal and 9 yeast genomes. Genome regions differing considerably in their entropy characteristics from the average over the genome, were studied.

—**Putintseva Y**, Sadovsky M

B-28 The hidden duplication past of the plant pathogen Phytophthora and its consequences for infection

We present a newly developed strategy to uncover ancient polyploidy events, in particular for heavily rearranged genomes. Furthermore, we have applied our new approach on the Phytophthora genomes to unveil their duplication history.

—Martens C, Van de Peer Y

B-29 Proteins linkages between African Trypanosoma and 8 pathogenetic eukaryotic organisms

Trypanosoma brucei differs from higher eukaryotes in many aspects.In the present study, we use the Domain Fusion Analysis for the prediction of functional functional linkages and associations between Trypanosoma brucei and 8 pathogenic organisms.

—Dimitriadis D, Trimpalis P, Choli-Papadopoulou T, Anagnou NP, Kossida S

B-30 Reconstructing phylogenetic trees from clustering trees

Some top-down methods for phylogenetic tree construction can be viewed not just as constructing trees, but as identifying constraints that the phylogenetic tree must satisfy. We show that this viewpoint can lead to improved phylogenetic trees.

—Costa E, Vens C, Blockeel H

B-31 Predictive modeling of psychrophilic adaptation on the proteome sequence-level

A numerical approach to characterize psychrophilic adaptation on the sequence level is presented. For some protein families discriminating models are derived, allowing for classification of adapted vs. non-adapted sequences.

—**Frickenhaus S**, Beszteri B

B-32 Microbial phenotype prediction based on protein domain profiles

We present a discriminative machine learning approach for prediction of microbial phenotypes from the organism-specific domain content. Our method achieves high prediction accuracy and allows biological interpretation of the learned genomic features.

—Lingner T, Mühlhausen S, Meinicke P

B-33 New insights into the metazoan evolution of cadherins: from basal to modern

Exhaustive identification and comparative analysis of the cadherin repertoires in key organisms such as amphioxus, sea anemone and the placozoan Trichoplax enabled us to reconstruct the complex metazoan evolution of the cadherin superfamily.

—Hulpiau P, van Roy F

B-34 Comparative mapping of transcription factor binding sites in plant genomes

Comparative mapping of transcription factor binding sites results in enrichment for functional binding sites and reduces the number of non-functional and therefore false positive predictions which frequently occur in genome wide motif studies.

—Wischnitzki E, Van de Peer Y, Vandepoele K

B-35 Protein model selection with Prottest increases phylogenetic performance

We carried out computer simulations to assess the role of model selection on the inference of maximum likelihood phylogenetic trees from protein alignment, using best-fit empirical models of amino acid replacement identified by ProtTest.

—Patricio M, Abascal F, Zardoya R

B-36 cn.FARMS: a probabilistic latent variable model to detect copy number variations

High-density oligonucleotide genotyping microarrays, especially Affymetrix SNP6 chips, are widely used for copy number analysis. In order to identify CNVs more reliable, we have developed a maximum a posteriori factor analysis model called cn.FARMS.

—Clevert D-A, Mitterecker A, Mayr A, Klambauer G, Tuefferd M, De Bondt A, Talloen W , Goehlmann, and Hochreiter S

B-37 Genome-wide heterogeneity of the substitution process

We present a genomic characterization of variation in best-fitting nucleotide substitution models. Different genes are best explained by different models, suggesting that there is a large amount of phylogenetically relevant genomic heterogeneity.

—Arbiza L, Dopazo H, Posada D

B-38 Comparative microarray analysis to elucidate TF-networks in activated T-cells

T-cell activation is crucial for the adaptive immune response against pathogens. We assembled a large set of publicly accessible gene expression data sets concerned with T-cell activation to study and re-engineer transcription factor networks.

—Kröger S, Scheel T, Leser U , Baumgrass R

B-39 Gene-trait matching analysis of Lactobacillus plantarum strains

We present a gene-trait matching method that allows integration of large data sets. We test our method with genotype and phenotype data of 38 Lactobacillus plantarum strains. It allowed the identification of phenotype-specific gene clusters.

—Bayjanov J R, Siezen RJ, Van Hijum SAFT

B-40 ProtTest-HPC: fast selection of best-fit models of protein evolution.

We have developed and evaluated a parallel version of ProtTest for the fast selection of best-fit models of protein evolution in multi-core desktops and high performance computing environments.

—Darriba D, Taboada G-L, Doallo R, Posada D

B-41 NTRFinder: an algorithm to find nested tandem repeats

We introduce the algorithm NTRFinder to find complex repetitive structures in DNA that we call nested tandem repeats . We have tested our algorithm on both real and simulated data, and to date have found 3 NTRs of interest in real sequences.

—Matroud A, Hendy M, Tuffley C

B-42 Computational epigenomics of plant SNF2 genes

Chromatin remodeling is at the core of epigenetic gene regulation, in plants responsible for proper development and genotype*environment interactions. Comparative analysis of plant Snf2 genes helps elucidate mechanisms of chromatin remodeling.

—Bargsten JW, Mlynárová L, Nap JPH

B-43 Comprehensive analysis of splice site evolution in primates using whole genome alignments and RNASeq data

We present the comparative analysis of changes in splicing sites between several primate species. Based on human annotated transcripts, we identify non-functional splice sites in other primates and verify our results using RNASeq data.

—Ongyerth M, Prüfer K, Kelso J

B-44 An ancient motif signature defines hundreds of vertebrate developmental enhancers

We uncover a striking enrichment within vertebrate CNEs for binding-site motifs of the Pbx-Hox hetero-dimer. We anticipate our findings will have a significant impact on understanding the gene-regulatory elements that underlie vertebrate development.

—Parker H, **Piccinelli P**, Elgar G

B-45 Diversity of the chromosomal beta-lactamase in 181 clinical Klebsiella oxytoca isolates. Comparison of the relation between beta-lactamase and gyrase-A sequences

Phylogenetic analysis of beta-lactamase genes from 181 Danish Klebsiella oxytoca isolates shows a clear separation into five known types. More than half is type Oxy2. GyraseA sequences support the grouping with one exception Oxy1 and Oxy5 are fused.

—**Worning P**, Boye K, Hansen DS

B-46 Origin and evolution of the organellar release factors

Organellar gene expression is far from understood, with the human mitochondrion genetic code being elucidated in 2010. We uncovered the remarkable ribosomal release factor functional differentiation and co-evolution with the organellar genetic code.

—**Duarte I**, Huynen M .

B-47 Identification and characterisation of novel "atypical" odorant binding proteins in the mosquito genomes

We present the identification of a new class of atypical odorant binding proteins and the classification of the family in the three mosquito genomes Anopheles gambiae, Aedes aegypti and Culex quinquifasciatus based on comparative genome analysis.

—**Manoharan M**, Ng Fuk Chong M, Vaitinadapoulle A, Frumence E, Sowdhamini R, Offmann B

Protein and Nucleotide Structure

C-1 Developing a methodology for protein-ligand docking based on genetic algorithm and normal modes

In this poster we describe the methodology and discuss the results and the performance obtained with the two first version of methodology that combine genetic algorithm, rotamer library and normal modes anlaysis to simluate docking protein-ligand.

—Lima AN, Philot EA, Scott LPB, Perahia D

C-2 A statistical potential for modelling of protein-RNA complexes.

We provide two knowledge-based potentials for discrimination of near native decoys, with RMS

—Tuszynska I, Rother K, Bujnicki JM

C-3 ProBiS: A web server for detection of structurally similar protein binding sites

A web server, ProBiS, freely available at http://probis.cmm.ki.si, is presented. This provides access to the program ProBiS, which detects protein binding sites based on local structural alignments.

—Konc J, Janežič D

C-4 Toward the prediction of the absolute quality of single protein structure models

Here we present a new method for the prediction of the absolute quality of a model in which quality is expressed based on how similar the model's "energy" is compared to values calculated for experimental structures solved by X-ray crystallography.

—Benkert P, Biasini M, Schwede T

C-5 NMR data and protein structure

The PDBe works on relating NMR data to protein structures, and on presenting NMR structures and data to the user. This poster gives an overview of the current status of NMR at the PDBe.

—**Vranken W**

C-6 Predicting the effect of mutations on the thermal stability of proteins with statistical potentials and artificial neural networks

We present a new method to predict changes in melting temperature induced by mutations in proteins. It relies on a set of statistical potentials and a neural network, and outperforms predictions based on estimations of the thermodynamic stability.

—Folch B, Rooman M, Dehouck Y

C-7 BioShell - a universal utility library for structural bioinformatics

BioShell is a bioinformatical toolkit written in Java, providing programs and a scripting library . It can also be used for development of Java programs. Applications include sequence and structure alignments and structural analysis.

—**Gront D**

C-8 Molecular dynamic simulation studies of membrane bound fully solvated β3 adrenergic receptor

A 5.0 ns molecular dynamics simulation was performed on the predicted 3D model of β3AR. The trajectories obtained during production phase were analyzed. The average coordinate file generated during production phase of was taken as the final model.

—Tewatia P, Malik BK, Sahi S

C-9 The significance of the ProtDeform score

We show that as TM-scores, PD-scores are protein length independent and follow approximately an Extreme Value Distr. On fold and homology levels, PD-scores are more discrimiant that TM-scores. We calculate 3 different PD-values for different tests.

—Rocha J, Alberich R

C-10 Exploiting synergy between computational biology and X-ray crystallography for solving challenging macromolecular structures

We present a novel method for completing fragmented models automatically obtained from low-resolution electron density maps, exploiting the presence of intrinsic information - the non-crystallographic symmetry .

—**Wiegels T**, Lamzin V

C-11 Pattern recognition with moment invariants for interpretation of very low resolution macromolecular density maps

We present a novel method for the interpretation of low-resolution maps, which does not rely on any map segmentation or knowledge about the position of the individual structural fragments. The structures of the fragments should be known in advance.

—**Heuser P**, Lamzin VS

C-12 PTools: an open source molecular docking library

PTools is an open source docking library that handles biomolecular structures at atomic and reduced resolutions. The ATTRACT program based on the PTools library can perform protein-protein, protein-DNA and 3-body docking simulations.

—Poulain P, Saladin A., Fiorucci S, Zacharias M, Prévost C

C-13 Development of a structural alignment tool for protein local surfaces

We present a structural alignment tool that produces similarity scores, superposed structures and atom alignments of detected similar surface regions of the input pair of proteins. It detected similar active sites of serine proteases in test case.

—Minai R, Horiike T

C-14 Blender game engine as a tool to navigate the conformational space of proteins

Using Blender, a software of computer graphics and video games, we developed a system to elaborate transitions between different models present in NMR collections. While providing a intermediate conformations, the system also builds a navigation map

—Zini M-F, Andrei R, Loni T, Zoppè M

C-15 Using data mining to identify structural rules in proteins

The progress in bioinformatics and biotechnology area has generated a huge amount of data sequences that requires a profoundly analysis. In this work, we study formation rules of the secondary structures of proteins using data mining.

—Stelle D, Scott LPB, Barioni MC

C-16 Detecting protein domains of biological assemblies using TopDomain-web

TopDomain-web is a tool that automatically assigns domains to different structural entities of proteins. We demonstrate that comprehensive structural information as provided by biological

assemblies is essential to assign compact structural domains.
—Senn S, Sippl MJ

C-17 Metals in protein structures: classification and functional prediction

We present a new method to compare metal sites in protein structures in a systematic and automated fashion. This provides the basis for their rigorous classification, the development of prediction tools, and the design of a metalloprotein database.
—Andreini C, Cavallaro G

C18. Intuitive visualization of surface properties of moving proteins

Using Blender, a free, open-source, 3D creation software, we developed a method for simultaneous visualization of surface properties of macromolecules, such as Electrostatic Potential and hydropathy.
—Andrei R, Callieri M, Zini M-F, Loni T, Zoppè M

C-19 Hierarchical classification of helical distortions related to proline

The presence of proline in a helix can be accommodated by several types of distortions. We present here a hierarchical method to classify them into five canonical structures. This method can be implemented on a large scale and will help to develop predictive tools for molecular modeling.
—Rey J, Devillé J, **Chabbert M**

C-20 A "smooth" knowledge based potential for protein structure refinement

We construct a geometrical knowledge-based potential for protein structures by optimizing 'nice smooth' energy minima around native structures. We use only the carbon alpha atoms to allow for faster sampling methods and the potential is spline based.
—Røgen P, Koehl P

C-21 PROTEIN PEELING in 2010: recent developments and applications

We present an update of our Protein Peeling web server for protein structure analysis. New improvements of server are described: mobile extremity identification, Protein Units energy computation and structural domain delineation.
—Gelly J-C, de Brevern AG

C-22 Local moves for efficient sampling of protein conformational space

We present a novel method to perform local deformations of protein backbone that includes an analytical solution for the loop closure problem. We show that this kinetic enables reliable sampling of protein equilibrium.
—Bottaro S, Boomsma W, Enøe Johansson K, Andreetta C, Hamelryck T, Ferkinghoff-Borg J

C-23 Self Organising Maps applied to protein structure classification

This work presents how a rather simple yet very powerful algorithm, called Self Organising Map, is able to cluster similar structures given a numerical representation of the proteins. Application of SOMs to structure prediction is analysed.
—**Alanis-Lobato G**

C-24 Construction of a sequence- and backbone-dependent rotamer library by hidden Markov model

Sidechain prediction is an important subproblem of protein design and structure prediction. Construction of rotamer library is the basis for protein sidechain prediction because it provides the basic searching space for prediction.
—Wen W, Lv Q, Yang P, Yang L-Y, Wu J-Z, Huang X

C-25 Codon usage and protein structure in prokaryotes: old stories, new findings and puzzling questions

We add a new chapter to the long standing debate on the potential relationship between codon usage and protein structure. We show that some early results do not hold true in light of an appropriate statistical framework and large scale analyses.
—Cozzetto D, Ward S, Jones DT

C-26 Analysis of FTICR mass spectra with H/D exchange of deprotonated dinucleotides

We present a statistical method for the interpretation of H/D exchange mass spectra. It allows estimating the exchange rate of hydrogens and drawing conclusions about the conformation of biomolecules.
—Claesen J, Valkenborg D, Burzykowski T

C-27 Voroprot: an interactive tool for the analysis of geometric features of protein structure

Voroprot is an advanced interactive tool that provides a unique set of possibilities for the visualization and analysis of geometric features of protein structure using additively weighted Voronoi diagram, also known as Apollonius diagram.
—Olechnovič K, Margelevičius M, Venclovas Č

C-28 BBP - Beta-Barrel Predictor: a web server for the prediction of the super-secondary structure of transmembrane β-barrel proteins

We present a web server for the prediction of the super-secondary structure of transmembrane β-barrel proteins whose motifs are based on permutations, such as Greek key or Jelly roll. It is compared to existing works and shows positive efficiency.

—Tran VD, Chassignet P, Steyaert J-M

C-29 Prediction of three dimensional structure of protein complexes

We present new benchmark of currently available docking software. Seven program were evaluated on extensive set of 1300 protein-ligand pairs. Ability to predict correct structure of complexes as well as scoring capabilities were tested.

—Plewczyński D, Łaźniewski M, Augustyniak R, Ginalski K

C-30 Investigating differences in the structural environment of parallel and anti parallel beta sheets

By inspecting the structural environment of beta sheets, we find an unexplained difference between the solvent exposed area of parallel versus anti-parallel beta sheets. This may have important consequences for the stability of beta sheets.

—**Bawono P**, Abeln S

C-31 Protein model quality assessment based on structural and functional similarities

We present GOBA - a model quality assessment programme that allows for the evaluation of the quality of protein model-structures based on the knowledge of the protein function.

—Konopka BM, Nebel J-C, Kotulska M

C-32 Data driven structure prediction: calculating accurate small angle X-ray scattering curves from coarse-grained protein models

We present a method for calculating accurate SAXS curves from coarse-grained protein structures. It was tested on a set of diverse structures and in protein decoy recognition and shows promise for use in inference of protein structures from SAXS data

—Stovgaard K, Andreetta C, Ferkinghoff-Borg J, Hamelryck T

C-33 OpenStructure: A flexible software framework for computational structural biology

We have developed OpenStructure, a flexible software framework for structural bioninformatics which integrates core computational methods with powerful visualization tools. It is an ideal working environment for the development of new algorithms.

—Biasini M, Mariani V, Haas J, Scheuber S, Schenk A-D, Schwede T, Philippsen A

C-34 Modeling structure of ionic channels from rigid network of contact sites

We present graph approach to characterize topology of protein channels and its characteristic features. The network of channel contact sites proved distinct from other proteins and indicated optimal definition of the contact site.

—Kotulska M, Gerstein M

C-35 Multi-task sequence labeling for protein annotation

We introduce an original multi-task approach for sequence labeling applied to protein annotation problems. We show successful results for the prediction of secondary structure, solvent accessibility and disorder regions on a set of 500 PDB proteins.

—Maes F, Becker J, Wehenkel L

C-36 A convex programming model for protein structure prediction

Convex optimization is applied to protein structure prediction. Globally minimum conformations may be found quickly and energy function parameters may be exactly determined. Applications for the framework are to NMR data and homology modeling.

—**Kauffman C**, Karypis G

C-37 Predicting protein function with the relative backbone position kernel

We propose a kernel method for predicting the function of proteins that makes use of 3D structural information. Using the kernel in support vector machines, we obtain a state-of-the-art accuracy on two datasets of protein structures.

—Schietgat L, Aryal S, Ramon J

C-38 The Torsional Network Model: improvements in modelling of protein flexibility

We present a benchmark of our new Torsional Network Model in terms of the robustness of the derived normal modes and its application to study protein conformational changes, both as compared to the closely related Anisotropic Network Model.

—**Méndez Giráldez R**, Bastolla U

C-39 MEDELLER: coordinate generation for membrane proteins

MEDELLER is a homology-based coordinate generation method for membrane proteins. Using membrane protein-specific information and accurate loop modelling it outperform the popular method Modeller on all test sets by an average of 0.48 Angstroms RMSD.

—Kelm S, Shi J, Deane CM

C-40 Structure-based predictor of HIV coreceptor tropism

We propose a structural predictor of the HIV coreceptor tropism offering a high accuracy and additional insight into the structural and physicochemical properties determining the viral phenotype.

—Bozek K, Lengauer T, Domingues FS

C-41 Different structures/same interaction partners: what can we learn from promiscuous protein-protein interactions?

In this work, we analyzed the binding sites of proteins that are able to bind similar partners, at equivalent binding sites, despite their structural dissimilarity. We found no clear evidence of convergent evolution.

—**Martin J**

C-42 3DM: A system to automatically build structure-based super-family systems

We present a system that automatically builds a structure-based super-family information system that integrates data from many different sources. The system has been tested in a series of protein engineering projects.

—Joosten HJ, Kuipers RKP, Vd Bergh T, Schaap PJ, **Vriend G**

C-43 Quantifying the relationship between RNA sequence and three-dimensional structure conservation for homology detection

We present an analysis of a selected set composed by 589 high similarity RNA structural alignment to quantify the relationship between RNA sequence and structure conservation and define the “twilight zone” for RNA homology detection.

—**Capriotti E,** Marti-Renom MA

C-44 IPknot: fast and accurate prediction of RNA secondary structures with pseudoknots using integer programming

We present IPknot for prediction of RNA secondary structures with pseudoknots using integer programming. IPknot is at least comparable in accuracy to several methods for RNA pseudoknot prediction and can run fast.

—**Sato K**, Kato Y, Akutsu T, Asai K

C-45 The Protein Model Portal, a resource of the Nature PSI Structural Biology Knowledgebase

The Protein Model Portal is a resource for 3-D protein structure models and their annotation. The current release allows searching millions of model structures and provides access to interactive services for model building and quality estimation.

—Bordoli L, Haas J, Benkert P, Mostaguir K, Kiefer F, Arnold K, **Schwede T**

C-46 PB-PENTAdb : a web resource for the analysis and prediction of local backbone structure and flexibility using pentapeptide protein blocks

Taking advantage of protein blocks, a 16-state description of the backbone of proteins, we have investigated the sequence to structure relationship at the pentapeptide level and are providing for the first time, an online tool.

—**Offmann B**, Tyagi M, Joseph A, Drula M, Grondin M, Frumence E, Vaitinadapoulle A, Cadet F, Srinivasan N, de Brevern A

C-47 RactIP: fast and accurate prediction of RNA-RNA interaction using integer programming

We present RactIP for RNA-RNA interaction prediction using integer programming. RactIP is at least comparable in accuracy to several methods for RNA-RNA interaction prediction and can run incomparably faster.

—**Kato Y**, Sato K, Hamada M, Watanabe Y, Asai K, Akutsu T

Prediction and Annotation of Molecular Function

D-1 Validating the structural information exchange within the domain Fyn SH2

We show in this poster the validation through NMR of the predicted information processing within the protein domain Fyn SH2 caused by peptide binding.

—Huculeci R, Buts L, Rousseau F, Schymkowitz J, van Nuland N, Lenaerts T

D-2 Discriminant functions for classification and prediction of T1SS, T3SS, T4SS and T6SS secreted proteins from proteobacteria

The two sets of predictor variables were obtained for multiple discrimination among T1SS, T3SS, T4SS & T6SS secreted proteins. The classifiers are considered as either independent or complementary tools to specify the leaderless secretion substrates.

—Kampenusa I, Zikmanis P

D-3 Computing fragmentation trees from tandem mass spectrometry data

We automatically annotate tandem mass spectra of metabolites with a fragmentation tree, which can reveal information about a completely unknown compound. The method does not require any information besides the spectra to work.

—Rasche F, Svatos A, Böcker S

D-4 Computational identification of new miRNAs in genomic regions associated with psychiatric diseases

We used an in-house developed miRNA prediction pipeline to predict new miRNA genes in genomic regions associated with psychiatric diseases. Using online available data from RNA-Seq experiments, we were able to find experimental evidence for 21 new miRNAs.

—Aelterman B, De Rijk P, Del-Favero J

D-5 CatANalyst: a web server for predicting catalytic residues

CatANalyst is a web server for predicting catalytic residues. It incorporates both a sequence-and a structure-based predictor and relies on an underlying SVM classifier that was shown to improve over previous state-of-the-art approaches.

—Cilia E, Passerini A

D-6 Functional characterization of Parkinson by high-throughput data analysis with l1l2 regularization

We present the results of an in-silico analysis on microarray Parkinson data, where the reliability of the employed variable selection technique is confirmed by the functional characterization of the identified signature.

—Squillario M, Masecchia S, Barla A

D-7 Feature elimination for one-class MicroRNA target pridection and gene identification

The usage of Zero norm feature elimination for one class machine learning to predict microRNA targets and gene identification cause an improving of the one-class results, where in some cases reaches the performance of the two-class approach.

—**Khalifa W**, Yousef M

D-8 Computational identification of synonymous SNPs in the human genome and their potential role in disease

We present a computational approach and analysis tool for the identification of synonymous SNPs within the human genome which may have a potential functional impact on gene expression and/or the gene product and may be associated with disease.

—Wood L, Ramsay M

D-9 The potential functional role of chimeric transcripts in higher eukaryotes

We studied the potential functional role of chimeric proteins in eukaryotes. We found that chimeric RNAs resulting from trans-splicing combine full functional and transmembrane domains, and use signal peptides to change localization of a chimera.

—**Morgenstern F**, Valencia A

D-10 A 'Functional Signature' for analysing annotation space

We have performed a comprehensive analysis of the current state of the functional annotations in protein databases. We describe homologous clusters by their 'Functional Signature' and indicate where functional transfer by homology should be possible.

—del Pozo A, Tress M, Valencia A

D-11 Predicting small molecule ligand binding sites and catalytic residues with firestar/FireDB

Here we present the new developments of firestar and FireDB, an expert system for predicting functional residues in protein sequences based on information extracted from structures and the Catalytic Site Atlas.

—López G, Maietta P, del Pozo A, Rodríguez JM, Valencia A, Tress ML

D-12 Classifying substrate specificities of membrane transporters from Arabidopsis thaliana

Based on a ranking method that measures the similarity of membrane transporters by their amino acid frequency and higher sequence order information, we annotated unknown transporters from Arabidopsis thaliana to four substrate classes.

—Schaadt NS, Christoph J, Helms V

D-13 Molecular modeling and comparison by docking and molecular dynamics of the interaction between fatty acids and proteins and EgFABP2 EgFABP1

In order to understand why similar fatty acid binding proteins are expressed in the flatworm parasite Echinococcus granulosus, we studied ligand affinities of EgFABP1 and EgFABP2 to different fatty acids: arachidonic, palmitic, oleic and linoleic.

—**Paulino Z-M**, Esteves A

D-14 GreenPhylDB version 2: web resources for comparative and functional genomics in plants

We present an update of the GreenPhyl website that contains a catalogue of gene families for 16 genomes of the plantae kingdom, ranging from algae to angiosperms. Gene families are first manually checked and then phylogenetically analysed.

—Rouard M, Guignon V, Aluome C, Laporte MA, Droc G, Walde C, Zmasek CM, Perin C, Conte MG

D-15 Collaboration-based function prediction in protein-protein interaction networks

We present a new approach for protein function prediction in protein-protein interaction networks. The main idea behind our approach is that topologically close proteins tend to have collaborative functions, not necessarily the same functions.

—Rahmani H, Blockeel H, Bender A

D-16 SVM feature selection applied to lignocellulose degradation

We study cellulose degradation by means of feature selection with an L1-regularized linear SVM to a microbal dataset.

—**Trukhina Y**, McHardy A

D-17 Substrate-specific HMMs for the classification of adenylation and acyltransferase domains of NRPS/PKS systems

We have developed an efficient strategy to classify the substrate specificity of the adenylation and acyltransferase domains of the non-ribosomal peptide synthetases and polyketide synthases .

—Khayatt B I, Siezen RJ, Francke C

D-18 Stochastic modeling of proteolytic activity from LC-MS data

We propose the stochastic model of serum proteolytic activity. Parameters of the model estimated from LC-MS/MS samples correspond to the cleavage activities of specific enzymes differentiating colorectal cancer patients from healthy subjects.

—Dittwald P, Gambin A, Ostrowski J, Karczmarski J

D-19 Mining posttranscriptional regulatory features by comparing total and polysomal mRNA profilings

The aim of this study is to relate the sensitivity of mRNAs to posttranscriptional control to UTRs contextual properties. We used machine learning techniques to extract informative features characterizing posttranscriptionally controlled mRNAs.

—Re A, Tebaldi T, Segata N, Passerini A, Blanzieri E, Quattrone A

D-20 Prototypes of elementary functional loops unravel evolutionary connections between protein functions

We derive prototypes of elementary functional loops having specific signatures and characteristic size 25-30 residues from sequences of complete proteomes. We show how folds and functions emerged in pre-domain evolution as a combination of prototypes

—Goncearenco A, Berezovsky IN

D-21 Utilization of proteins and nucleic acids in the study of gene function: a comparative review

This review examines protein amenability to prediction of gene function and the potential of proteomics in biological research. It is extracted from a wide array of literature on proteomics and gene function studies in view of current advancements in molecular biology and or bioinformatics

—Mwololo JK, Karaya HG, Munyua JK, Muturi PW, Munyiri SW

D-22 ClanTox: A predictor tool for toxin-like proteins reveals 500 such proteins within viral genomes

Animal toxins are short, highly stable proteins that operate by binding to receptors or ion channels. We developed a classifier that identifies toxin like proteins and tested it on viral proteins. We discovered 510 toxin like proteins in viruses.

—Naamati G, Askenazi M, Linial M

D-23 I-Patch web service: inter-protein contact prediction using local network information

Presentation of the I-Patch Web Service for prediction of inter-protein contact sites. I-Patch gives 59% precision with 20% recall on a blind test

set of 31 proteins - better than other existing methods on the same data set.

—Hamer R, Luo Q, Armitage J, Reinert G, Deane C, Krawczyk K

D-24 Predicting candidate genes for biomass traits

We present an Ondex workflow for automated functional annotation of newly sequenced genomes, as well as web interface for systematic analysis of QTL. It was used to annotate the poplar genome and to predict candidate genes for biomass traits.

—Hassani-Pak K, Kuo SC, Hanley S, Rawlings C

D-25 Exploring residue interaction networks to understand protein structure-function relationships

The new Cytoscape plugin RINalyzer enables the interactive, visual analysis of residue interaction networks derived from 3D protein structures. It affords exploring the molecular role of residues critical for protein structure and function.

—Doncheva N-T, Domingues F-S, Albrecht M

D-26 BioXSD: the canonical XML-Schema data model for bioinformatics Web services

BioXSD defines standard exchange formats of the basic bioinformatics data types: sequences, sequence annotations, alignments, and references to data resources. It aims to enable smooth interoperability of the world-wide bioinformatics Web services.

—Kalaš M, Puntervoll P, Joseph A, Bartaševičiūtė E, Töpfer A, Ison J, Blanchet C, Rapacki K, Jonassen I

D-27 Functional inference for experimental proteomics: the PANDORA (annotations graph) and ProtoNet (family tree) resources

From experimentalist to knowledge - A shortcut through PANDORA BOX. ProtoNet & PANDORA web tools for understanding proteins function by integration of sequence and annotations.

—**Rappoport N**, Linial M

D-28 Computational approaches to study glycosaminglycans recognition by IL-8 for the design of biomaterials for tissue regeneration

Glycosaminglycan-protein interactions in extracellular matrix play a critical role for a number of important biological processes. We characterize interactions of IL-8 with natural and modified glycosaminoglycans using computational approaches.

—**Samsonov SA** , Pisabarro MT

D-29 Expression profile based substrate specificity prediction in complete Saccharomyces cerevisiae methyltransferome

We identified complete S. cerevisiae methyltransferome (all methyltransferases encoded in genome) with structural and substrate specificity characterization, supplemented with development of new methodology for substrate specificity prediction.

—**Wlodarski T**, Rowicka M, Ginalski K

D-30 Discovery and annotation strategies of nonribosomal peptide synthetases from bacterial genomes

We propose a strategy to annotate nonribosomal peptide synthetases, huge enzymatic complexes producing peptides from free amino acids, not via central dogma. The classical automatic annotation process can't be applied on those enzymes.

—Saci Z, **Pupin M**, Deravel J, Krier F, Caboche S, Jacques P, Leclère V

D-31 Predicting N-glycosylation sites in human proteins

N-glycosylation site prediction in human proteins using artificial neural networks on sequence features. Distinguishes glycosylated and non-glycosylated canonical consensus sequons. Impact on human proteome and disease mutations is investigated.

—Gupta R, **Frederiksen J**, Jung E, Brunak S

D-32 AIGO: Toward a unified framework for the Analysis and the Inter-comparison of GO functional annotations

This work defines an ensemble of measures suitable to provide both a quantitative and a qualitative description of GO functional annotations and also to allow a comprehensive inter-comparison between them.

—Hindle M, Powers S, Habash D, Saqi M, Defoin-Platel M

Gene Regulation and Transcriptomics

E-1 TFactS: a tool to predict transcription factor regulation from gene expression data

We present TFactS, a transcription factor target gene database, used to predict the regulation of transcription factors from gene expression data. The tool was updated and applied on different biological systems .

—Essaghir A, Pachikian BD, Delzenne NM, Toffalini F, van Helden J, Demoulin JB

E-2 Time- and dose-dependent effects of cigarette smoke on monocytic cell transcriptome

In the context of atherosclerosis, our transcriptomics study investigates the molecular mechanisms underlying the effect of cigarette smoke exposure on monocytic cells over time. Hundreds of genes were found to be regulated.

—Poussin C, Lietz M, Schlage W, Lebrun S, Hoeng J, Peitsch M

E-3 Combined Monte Carlo and dynamical modelling of gene regulation enables multiple parameter estimations and hypothesis generation

Parameter estimation has revealed a bimodality of a genetic regulatory circuit and an additional interpretation of its outcome has been compared with sensitivity analyses. Moreover, an experimental verification of the model has been suggested.

—Herman D, Thomas CM, Stekel DJ

E-4 Module-based comparative gene expression analysis: evolutionary conserved coexpression in Bacillus subtilis and Escherichia coli

We developed a method that uses microarray data and homology mapping among genes as input and provides pairs of conserved modules as output. We applied our methodology to study coexpression conservation between Escherichia coli and Bacillus subtilis.

—Zarrineh P, Fierro C, Sánchez-Rodríguez A, De Moor B, Marchal K

E-5 Algorithm-driven artifacts in median polish summarization of microarray data

Over-similarity artifacts are introduced by the summarization step of Affymetrix microarrays preprocessing methods RMA and GCRMA . We present a modified RMA procedure and show its performance in microarray clustering tasks.

—Giorgi F-M, Bolger A, Lohse M, Usadel B

E-6 In silico study of the regulation of sRNAs in Escherichia coli

In this work we aimed at updating the known Escherichia coli transcriptional network with the sRNA regulatory network. This includes predicting the regulation of the sRNAs and extending sRNA regulons of with novel targets.

—Ishchukov I, Ryan D, Zarrineh P, Cloots L, Thijs I, Engelen K, Marchal K

E-7 Systems biology analysis at VIB

We have developed powerful software tools to infer gene regulation networks from transcriptional data that can help identifying molecular markers and deciphering the wiring of networks of interacting components.

—**Ren X-Y**, Bonnet E., Joshi A, Maere S, Michoel T, Van Parys T, Vermeirssen V, Van de Peer Y

E-8 A computational framework for automated analysis of nonlinear dynamics of gene regulatory networks

We present a computational tool for qualitative simulation of gene regulatory network dynamics modelled by a specific class of ODEs. It is appropriate to derive all different types of dynamical behaviors from given network structures.

—Ironi L, Panzeri L

E-9 Exploiting ESTs to infer plant alternative splicing

We present findAS, a novel java-based utility to find alternative splicing events from genomic alignments of mRNA sequences. FindAs extracts the putative AS from alignments of expressed sequences by means of a cluster analysis.

—Potenza E, Cestaro A, Fontana P, Bianco L, Velasco R

E-10 Inferring gene regulatory networks to identify conserved regulation between Arabidopsis and Populus

We infer the gene regulatory networks of Populus and Arabidopsis from microarray expression data using several methods. The networks from the two species are aligned to identify conserved functional and topological features characteristic to trees.

—Netotea S, Hvidsten TR

E-11 BayesPI - a new model to study protein-DNA interactions: a case study of condition-specific protein binding parameters for Yeast transcription factors

We have incorporated Bayesian model regularization with biophysical modeling of protein-DNA interactions, and of genome-wide nucleosome positioning to study protein-DNA interactions, using a high-throughput dataset.

—**Wang J**, Morigen

E-12 Use of structural DNA properties for the prediction of regulator binding sites with conditional random fields

We present a framework for integrating local structural DNA properties in a motif screening application and create a set of novel target site predictions for various E. coli transcription factors that could be validated using independent data sources

—Meysman P, Engelen K, Laukens K, Dang TH, Marchal K

E-13 Gene expression analysis in Scots pine in response to red and far-red light

We studied gene regulation in Pinus sylvestris from northern Sweden grown under red and far-red light using Pine cDNA microarrays. Computational annotation of the genes was done with Blast2GO, which reveals existence of differentially expressed genes

—Ranade SS, Abrahamsson S, Niemi J, García-Gil MR

E-14 Transcription regulatory regions and motifs database – the genomic data layer of the Nencki Regulatory Genomics Portal

Understanding transcriptional co-regulation in vertebrates remains a challenge. To address it, we are developing a database of several kinds of putative regulatory regions in human, mouse and rat, scored with three major TFBS motifs libraries.

—**Dabrowski M**, Lenart J, Mieczkowski J, Kaminska B

E-15 The microRNA body map: dissecting microRNA function through integrative genomics

We present the microRNA body map, an interactive online compendium and mining tool of high-dimensional newly generated and published microRNA expression profiles along with functional annotation inferred through integrative transcriptomics.

—Mestdagh P, Lefever S, Pattyn F, Ridzon D, Fredlund, Fieuw A, Vermeulen J, De Paepe A, Wong L, Speleman F, Chen C, Vandesompele J

E-16 D-Light: Transcription factor binding site management and visualization

We present D-Light, a new client-server system for compiling, managing, querying and displaying annotation data for transcription factor binding sites. A GUI allows combinatorial queries within or between genomes.

—Laimer J, Zuzan C, Ehrenberger T, Freudenberger M, Gschwandtner S, Lebherz C, Lirk G, Lackner P

E-17 Expression and chromosomal clustering of tissue restricted antigens

Tissue restricted antigens are crucial in the negative selection of autoreactive T cells in the thymus to prevent autoimmune dieseases. We defined TRAs from microarray data and could show that they are significantly clustered on chromosomes.

—Dinkelacker M, Pinto S, Derbinski J, Eils R,Kyewski B, Brors B

E-18 Community dynamics and metatranscriptome analysis of lactic acid bacteria ecosystems through functional gene microarray analysis

A lactic acid bacteria functional gene microarray was successfully used to unravel community dynamics and to analyze the metatranscriptome of a ten-day backslopped sourdough fermentation in a temporal way.

—Weckx S, Allemeersch J, Van der Meulen R, Vrancken G, Huys G, Vandamme P, Van Hummelen P, De Vuyst L

E-19 A functional gene microarray for lactic acid bacteria exceeding the species level: design, validation, and data analysis

A functional gene microarray for lactic acid bacteria was developed to investigate whole-ecosystem gene expression in fermented food ecosystems.

—Allemeersch J, Van der Meulen R, Vrancken G, Huys G, Vandamme P, Van Hummelen P, De Vuyst L, Weckx S

E-20 Combinatorial analysis of transcription regulation of response to oxidative stresses in E. coli

To study mechanisms involved in stress response in E. coli from gene expression data, an original approach based on Answer Set Programming was

designed. With this method, new targets for predicted transcription factors can be discovered.

—Thiele S, Klie S, Jozefczuk S, Selbig J, Blachon S

E-21 Time delays in gene activation: sensitivity to external noise

We present a study of a genetic switch capable of delaying its transition from low to high expression state after being excited by an external input, and discuss the conditions under which these delays are most resistant to parameter perturbations.

—Albert J, Rooman M

E-22 Discovering regulatory mechanisms that govern zygotic genome activation in Drosophila

From clusters of genes co-expressed at different stages of early embryogenesis, we detected over-represented motifs potentially involved in the zygotic genome activation in Drosophila.

—Darbo E, Thieffry D, Lecuit T, van Helden J

E-23 Detection of extended UTRs with 3' expression microarrays

Affymetrix 3' expression arrays were used to predict transcript extension beyond currently known boundaries.

—Thorrez L, Tranchevent L-C, Chang H-J, Moreau Y, Schuit F

E-24 A dynamic regulatory network model reveals hard-wired heterogeneity in blood stem cells

We present the most advanced mammalian regulatory network model based on known cis-regulatory interactions to date, from which we made novel, experimentally testable hypotheses about the control of differentiation in haematopoietic stem cells.

—Bonzanni N, Garg A, Foster SD, Wilson NK, Kinston S, Miranda-Saavedra D, Heringa J, Feenstra A, Xenarios I, Göttgens B

E-25 Alternative splicing as a marker of disease

In this work we try to identify aberrant alternative splicing isoforms that are cause or consecuence of human disease with a focus on neurodegenerative disease. We will classify them according to tissue and age.

—**Muro EM**, Andrade-Navarro MA

E-26 TilSeg: an automated pipeline to process tiling arrays expression data

TilSeg is a tool dedicated to tiling arrays automated analysis. This software uses standard quantification files, performs data analysis and provides standardized outputs files which can be easily imported to genome browsers or other analysis tools.

—Barreau D, Gao Y, Jaffrezic F, Hugo K, Rogel-Gaillard C, Marthey S

E-27 Evaluation of computational miRNA target predictions in human

MicroRNA target prediction is a popular research topic. Many computational prediction tools have been developed that predict different targets for the same microRNA. This poster presents an objective comparison between 7 often used tools.

—**Gevaert O**

E-28 Model-based multifactor dimensionality reduction to detect epistasis for quantitative traits in the presence of error-free and noisy data

We present the power of Model-Based Multifactor Dimensionality Reduction in identifying significant epistasis models for quantitative traits in the presence of noisy data.

—Mahachie John JM, Van Lishout F , Van Steen K

E-29 Time coherent three-dimensional clustering: unraveling local transcriptional patterns in multiple gene expression time series

We present an efficient algorithm, relying on string matching using suffix trees and itemset mining, to unravel gene modules exhibiting a coherent transcriptional response in specific time frames across multiple gene expression time series.

—Gonçalves JP, Moreau Y, Madeira SC

E-30 The role of incoherent microRNA-mediated feedforward loops in noise buffering

Incoherent microRNA mediated feedforward loop are remarkably overrepresenetd in the human regulatory network. We show analytically and through simulations that their role in the network is to control the fluctuations in the target protein level

—Osella M, Bosia C, Corà D, Caselle M

E-31 Dynamic transcriptomics of helper T cell differentiation

We performed transcription profiling of helper T cells. Using our polar score method, we were able to identify genes specifically associated with Th1/Th2 differentiation in the large set of – non polarizing – differentially expressed genes.

—van den Ham HJ, de Waal L, Zaaraoui F, Bijl M, van IJcken WF, Osterhaus ADME, de Boer RJ, Andeweg AC

E-32 Effect of size and heterogeneity of samples on biomarker discovery: synthetic and real data assessment

We analyzed the effect of sample size and heterogeneity on a number of biomarker discovery methods on real and synthetic data. Classification based approaches improve selection; using external cross-validation loops improves results reproducibility.

—Di Camillo B, Martini M, Sanavia T, Jurman G, Sambo F, Barla A, Squillario M, Furlanello C,Toffolo G, Cobelli C

E-33 Analysis of DNA sequence dependent structural properties in prokaryotic genomes

Sequence dependent DNA structural property analysis in E. coli, B. subtilis and M. tuberculosis promoter sequences as well as from other prokaryotic genomes reveal that lower stability and bendability are more common than higher intrinsic curvature.

—**Rangannan V**, Bansal M

E-34 Computational analysis of transcription factors' binding co-localization

We studied gene expression and biological pathways data to investigate DNA Binding Targets co-localization in human and mouse. Specific nucleotide distances between DBTs in promoter regions suggest patterns of transcription factor cooperation.

—Martinez P, Blanc E, Coolen A, Holzwarth J, Fraternali F

E-35 Cross-checking experimental results with publicly available gene expression data: a query-driven approach

We present a computational tool that allows for cross-checking a list of experimentally derived genes with a gene expression compendium comprising publicly available gene expression data.

—De Smet R, Hermans K, De Keersmaecker S, Marchal K

E-36 Using tiling microarrays to predict transcription start sites: application to Lactobacillus plantarum WCFS1

We present a method for detecting transcription start sites based on whole-genome tiling microarray expression data to validate genome-wide predicted promoter locations.

—Todt T, Wels M, Siezen RJ, Van Hijum SAFT, Kleerebezem M

E-37 Predicting regulatory networks from large scale time-course microarray data under several stimulation conditions on MCF-7 cells

We analyzed large scale time-course microarray data under several stimulation conditions on MCF-7 cells. Our analysis presented a putative regulatory network leading to biological functions initiated by distinct stimulations.

—Shiraishi Y, Saeki Y, Nagashima T, Yumoto N, Takahashi K, Okada M

E-38 Inferring regulatory networks from expression data using tree-based methods

We present a new algorithm for the inference of genetic regulatory networks from expression data that exploits tree-based ensemble methods. This algorithm was best performer in the DREAM4 In Silico Multifactorial challenge.

—Huynh-Thu VA, Irrthum A, Wehenkel L, Geurts P

E-39 Dynamic profile of transcription factors inferred from mRNA expression time courses in Gefitinib-treated lung cancer cells

We present a statistical test to reverse-engineer dynamic activities of transcription factors from gene expression time courses. Key regulatory pathways involving new clinical biomarkers are found in analyses of Gefitinib-treated lung cancer cells.

—Nagao H, Yoshida R, Saito M-M, Imoto S, Yamaguchi R, Yamauchi M, Goto N, Miyano S, Higuchi T

E-40 Deciphering the genomic plasticity of bacterial species

We developed a computational procedure and compared the predicted regulatory systems from four closely related bacterial species in the genus Bordetellae, of which three are pathogenic with different host specificity.

—Janky R, Harvill ET, Madan Babu M

E-41 Expression arrays & outlying processes: building a database of outliers

We present the development of a database of probes for Affymetrix expression arrays. A php interface allow to summarize probe features and empirical measures, in order to detect both theoric and empirical outliers.

—Berger F, Kroll KM, van Dorp M, Carlon E

E-42 Detecting significantly correlated gene clusters and their interactions in a continuous scale-space

We apply scale space theory to detect genomic regions showing significant clusters of genes with highly correlated gene expression profiles. We introduce a novel method to control the FWE when detecting clusters with non-homogeneous gene densities.

—Van Dyk E, Reinders M, Wessels L

E-43 MAMMOD: Multi-Agent Motif and MOdule Discovery

We present Mammod, a novel tool for simultaneous discovery of cis-regulatory motifs and modules in eukaryotic promoters. We demonstrate Mammod's performance on benchmark data sets and show that it can compete with several state-of-the-art methods.

—Van Delm W, Ayoubi T, De Moor B, Moreau Y

E-44 Modulon identification from bacterial gene expression array data using improved supervised approach

We present and improvement of correlation matrix-based method of data mining for microarray data. We have analyzed different data source and several basic regulons in E. coli, coming up with a software for supervised prediction of new modulon members

—Hedge S, Permina E, Medvedeva Y, Mande SC, Makeev V

E-45 GCQM: quality assessment of microarray studies and samples using gene correlations

The quality of gene expression data is variable, and difficult to measure. We propose a quality measure based on the consistency of gene-gene correlations. We show that it allows the flagging of bad quality studies and samples across platforms.

—Venet D, Detours V, Bersini H

E-46 Integration of correlation structure improves performance of pathway-based classification

By integrating the correlation structure of gene set expression profiling, we imporved the performance in the gene set-based classification analysis. Our method was validated with simulation data and four breast cancer datasets.

—Cho SB, Kim JH

E-47 Multi Experiment Matrix – web tool for mining co-expressed genes over hundreds of datasets

MEM is a query engine atop of the microarray experiments from ArrayExpress that performs search for co-expressed genes over hundreds of datasets at a time. Given the query gene MEM finds a list of genes that have similar expression in many datasets.

—Adler P, Kolde R, Kull M, Peterson H, Reimand J, Tkatšenko A, Vilo J

E-48 Linking parallel measurements of high-throughput miRNA, gene and protein expression data

We presented a graph based workflow for linking different kinds of omics data using pathway information and component wise boosting. It was tested on a colorectal cancer data set with miRNA and gene expression measurements of 97 samples.

—Gade S, Binder H, Beißbarth T

E-49 An extended translational network involving transcription factors and RNA binding proteins derived from the identification of phylogentic footprings in 5' and 3' untranslated regions

We present a computational pipeline and a first benchmarking experiment aimed at the prediction of a network coming from the analysis hyperconserved elements of untranslated regions, applied to a 46 vertebrates species alignment.

—Dassi E, Tebaldi T, Zuccotti P, Riva P, Quattrone A

E-50 Unvealing heterogeneity of stage II and stage III colorectal cancer by defining molecular subtypes

We present the results of a meta-analytical approach for identifying robust subgroups of patients with colorectal cancer. Unsupervised analysis of gene expression profiles revealed major molecular subgroups, stable across experiments.

—Budinska E, Popovici V, Delorenzi M

E-51 A method for improved prediction of in vivo transcription factor binding sites by using biophysical DNA features

We present a sequence-based biophysical approach for the identification of transcription factor binding sites. This method was used on widely available ChIP-Seq data.

—Hooghe B, Broos S ,Van Roy F, De Bleser P

E-52 Nonlinear dimension reduction and clustering by Minimum Curvilinearity unfold neuropathic pain and tissue embryological classes

'Minimum Curvilinearity' is a principle which inspires two novel nonlinear machine learning: Minimum Curvilinear Embedding for dimensionality reduction; and Minimum Curvilinear Affinity Propagation for clustering.

—Cannistraci CV, Ravasi T, Montevecchi FM, Ideker T, Alessio M

E-53 Seeded Cis-Regulatory Module discovery using rank based motif scoring

We have development an in silico cis-regulatory module detection algorithm for aid in interpretation of gene regulatory networks in humans. Our biologically motived design results in improved performance over existing CRM predictors.

—Macintyre G, Bailey J, Kowalczyk A, Haviv I

E-54 Proximity-based cis-regulatory module detection using constraint programming for itemset mining

We present an extendible method for the detection of cis-regulatory modules, based on constraint programming for itemset mining. The method finds all combinations of transcription factors that frequently have binding sites in each others proximity.

—Guns T, Sun H, Nijssen S, Sanchez-Rodriguez A, De Raedt L, Marchal K

E-55 High-throughput detection of epistasis in studies of the genetics of complex traits

We present a novel tool that has been developed to allow routine high throughput analysis of epistasis in complex traits in populations genotyped with high density single nucleotide polymorphism markers.

—Gyenesei A, Laiho A, Semple C, Haley C, Wei W

E-56 Heavy metal resistance in Cupriavidus metallidurans: a complex evolutionary and transcriptional process

The metal-resistant soil bacterium Cupriavidus metallidurans CH34 displays myriads of gene expression patterns when exposed to a wide range of heavy metals at non-lethal concentrations pointing towards a complex regulatory network.

—Monsieurs P, Moors H, Van Houdt R, Janssen P, Mergeay M, Leys N

E-57 Identification of co-regulated mRNA/miRNA pairs by the CoExpress software tool

Here we present the software tool CoExpress and apply it for the analysis of co-regulation between miRNA and mRNA in 14 cell lines. Co-regulation effects were validated using numerical methods, bioinformatics and qPCR experiments.

—Nazarov P V, Khutko V, Schmitz S, Muller A, Kreis S, Vallar L

E-58 Identification of genes with preferential expression in the egg cell

Gametes, like egg and central cell, are crucial cell types in the plant life cycle. To understand gamete differentiation and function, we isolate and characterize egg cell specific genes of wheat and the model plant Arabidopsis.

—Köszegi D, Czhial A, Kumlehn J, Altschmied L, Baumlein H

E-59 Towards real-time control of gene expression: controlling the HOG signalling cascade

We present preliminary results on the development of a platform for the real-time control of gene expression. The platform integrates a microfluidic device, an epi-fluorescence microscope, and software implementing control approaches.

—Uhlendorf J, Bottani S, Fages F, Hersen P, **Batt G**

E-60 Sequencing the transcriptome of a deep-sea hydrothermal vent mussel: new possibilities for the discovery of immune genes in an unconventional model organism

We have established the first tissue transcriptional analysis of a deep-sea hydrothermal vent animal and generated a searchable catalog of genes which can be applied in gene expression profiling experiments from a non-conventional model organism.

—**Bettencourt R**, Stefanni S, Pinheiro M, Egas C Serrão Santos R

E-61 Towards a computational tool to uncover genes involved in signaling crosstalk in Arabidopsis thaliana

Crosstalk between signaling pathways is of critical importance in organisms. We present a computational method to identify genes involved in signaling crosstalk in Arabidopsis thaliana, with the goal of providing a tool for experimental scientists.

—**Omranian N**, Arvidsson S, Riaño-Pachón DM, Mueller-Roeber B

E-62 Prediction of the stage of embryonic stem cells differentiation from genome-wide expression data

We evaluated several machine learning approaches to predict the stage of embryonic stem cells differentiation from genome-wide expression profiles. Results are very encouraging, even when learning and prediction use samples from different cell lines.

—**Zagar L**, Mulas F, Sacchi L, Garagna S, Zuccotti M, Bellazzi R, Zupan B

E-63 High throughput sequencing of the Anopheles transcriptome through sexual development

Despite the importance of Anopheline mosquito species as vectors of human malaria, the mechanisms leading to sexual differentiation in these insects is yet to be elucidated. 454 pyrosequencing was used to identify sex-specific splicing in A. gambiae

—**Koscielny G**, Nolan T, Severgnini M, Rizzi E, Lawson D, Kersey P, De Bellis G, Crisanti A

E-64 The complexity of gene expression dynamics revealed by permutation entropy

The analysis of gene expression data has so far focused primarily on the identification of differentially expressed genes. We aim at studying gene expression time series data from the viewpoint complexity.

—Sun X, Zou Y, Nikiforova V, Kurths J, **Walther D**

E-65 Efficient query-based biclustering of gene expression data using probabilistic relational models

We present ProBic, a query-based biclustering method. ProBic identifies biologically sound biclusters and, bicluster identification is robust as the set of query genes can contain genes that are not part of the bicluster of interest.

—Cloots L, Zhao H, Van den Bulcke T, **Wu Y**, De Smet R, Storms V, Meysman P, Engelen K, Marchal K

Text Mining, Ontologies and Databases

F-1 Resource of Asian Primary Immunodeficiency Diseases update: an open web-based integrated molecular database on primary immunodeficiencies

RAPID hosts information on PID gene-specific molecular alterations, expression profiles, interaction networks and mouse studies. By scanning whole genome, predicted 1442 candidate PID genes using SVM, out of which 15 genes have been so far confirmed.

—Ramadoss SK, Keerthikumar S, Raju R, Kandasamy K, Balakrishnan L, Dhevi L, Selvan N, Sekhar NR, Mohan S, Bhattacharjee M, Hijikata A, Koh

F-2 Measuring the functional similarity of drugs

We presents a novel strategy to assessing the functional similarity of annotation profiles based on the Gene Ontology topology. We applied this strategy to an extensive dataset of drug profiles derived from genome-wide expression data. The dominant b

—Götz S, Behrens S, Tarazona S, Marba M, Dopazo J, Conesa A

F-3 G-language Bookmarklet: a gateway for semantic web, linked data, and web services

G-language Bookmarklet is a visual gateway to numerous biological resources that works on any web page viewed with any browser, implemented as an animated interactive ring-shaped menu, freely available at:

http://www.g-language.org/wiki/bookmarklet

—Arakawa K, Kido N, Oshita K, Tomita M

F-4 A graphical view of the world of protein-protein interaction data-bases

The graphical representantion of data exchange between various protein-protein databases is visualized using Cytoscape and presented at http://www.pathguide.org/interactions.php

—Klingström T, Plewczynski D

F-5 StrainInfo: your microbiological stepping stone

StrainInfo is a global catalog of microbial material integrating the catalogs of more than 60 Biological Resource Centers . StrainInfo integrates and bundles information on its strain, sequence, taxon and publication passports.

—Verslyppe B, De Smet W, Gillis W, De Vos P, De Baets B, Dawyndt P

F-6 Semantic integration of isolation habitat and location in StrainInfo

StrainInfo integrates multiple redundant textual descriptions of isolation habitat and location by means of an ontology-based integration algorithm.

—Verslyppe B, De Smet W, De Vos P, De Baets B, Dawyndt P

F-7 Towards an ontology for farm animal reproduction traits

A framework of an ontology of reproduction traits in farm animals was developed. A major goal was to integrate as much knowledge as possible about the domain reproduction traits in farm animals and that can be used by computer-based analyses.

—**Hulsegge B,** Smits MA, te Pas MFW, Woelders H

F-8 Human gene symbol validation at the HGNC

High throughput technologies has created an increasing demand to resolve large lists of human gene symbols to their HGNC ID. We have developed a simple web-based 'list search' tool which allows a researchers to quickly validate such datasets.

—**Lush MJ**, Seal RL, Gordon SM, Wright MW, Bruford EB

F-9 miRSel: automated extraction of associations between microRNAs and genes from the biomedical literature

We propose to use text mining of publication abstracts for extracting microRNA-gene associations including microRNA-target relations to complement current repositories.

—Naeem H, Küffner R, Csaba G, Zimmer R

F-10 Metarel, a relation metagraph for inferences in biomedical ontologies

We enabled queries on the Open Biomedical Ontologies, by automated reasoning techniques. All the ontologies were translated to RDF, a Semantic Web standard. Reasoning modules were added to the Metarel metagraph, which allowed SPARUL inferences.

—Blondé W, Antezana E, Venkatesan A, De Baets B, Kuiper M, Mironov V

F-11 Predicting disease causing genes using the information content surrounding the disease and gene in literature

We present a text-mining approach for predicting gene-disease relationships, even when the gene and the disease never have been co-mentioned together in an abstract before. The prediction is based on indirect links.

—van Haagen H, Aten E,'t Hoen P-B, Roos M, Messemaker T, Mons B, van Ommen G-J, Schuemie M

F-12 Revealing heterogeneities and inconsistencies in protein functional annotations

We propose a method able to highlight potential inconsistencies on functional annotations in the Gene Ontology database of a pool of proteins thanks to the simultaneous clustering of both GO terms and protein annotations.

—Sanavia T, Facchinetti A, Di Camillo B, Toffolo G, Lavezzo E, Toppo S, Fontana P

F-13 A formal framework for evaluating the significance of peptide-spectrum matches

We present a framework for assessing the significance of peptide-spectrum-matches in computational mass-spectrometry. The use of enumeration techniques for counting peptides that match a specific spectrum allows to increase the sensitivity of PSMs.

—**Vandenbogaert M**

F-14 Improving the execution of bioinformatical workflows

We present a novel approach called Data Flow Delegation to improve the execution of bioinformatical workflows, which optimizes the integration and coordination of different Biological Databases and tools for solving various biological problems.

—Subramanian S, Sztromwasser P, Puntervoll P

F-15 BridgeDb: standardized access to gene, protein and metabolite identifier mapping services

We present BridgeDb, a developer library to connect bioinformatics tools to online and offline identifier mapping services. BridgeDb already works with several mapping services such as BioMart, Synergizer, Cronos, PICR and HMS

—Van Iersel MP, Pico AR, Kelder T, Gao J, Ho I, Hanspers K, Conklin BR, Evelo CT

F-16 HGVbaseG2P: an advanced database for the integration and interrogation of genetic association datasets

The Human Genome Variation Genotype to Phenotype database provides a comprehensive publication medium for summary-level genetic association data. Powerful graphical presentation methods enable study visualisation & co-examination.

—Free R, Hastings R, Thorisson GA, Gollapudi VLS, **Beck T**, Lancaster O, Brookes AJ

F-17 Extracting phytogeography information from species distribution data

Availability of species distributions in digital format motivated us to analyze this data. Pair wise similarities among grid squares were calculated. Our results were highly accurate when we compare with the phytogeographical regions.

—Bakis Y, Sezerman OU, Babaç MT

F-18 D2K - data to knowledge - data integration for biological reasoning

We present data mining system by integration of quality controlled published, in silico generated and experimentally gathered data. The system is

designed to generate biological hypotheses a priori to wet-lab experiments or high-throughput studies.
—**Wienecke-Baldacchino AK**, Heinäniemi M, Carlberg C

F-19 BRENDA text-mining: new developments for obtaining enzyme-related disease information from scientific literature

BRENDA contains manual annotated and text mining derived information from scientific literature. The disease-related text mining component has been revised and extended to the further classification of mining results.
—Soehngen C, Scheer M, Schomburg I, Chang A, Grote A, Schomburg D

F-20 Beegle - a new search engine for discovering novel genes

We present a new method based on text mining and association network analysis which, from any free text query, returns known and potential novel genes linked to it.
—Brohée S, Gonçalves JP, Nitsch D, Moreau Y

F-21 An overview of the pathogen-host interactions database, PHI-base

The Pathogen-Host Interaction database is a phenomics database cataloguing experimentally verified pathogenicity genes. An orthologous gene cluster analysis is used to identify host specificity genes.
—Janowska-Sejda E, Defoin-Platel M, Hammond-Kosack K, Urban M, Tsoka S, Saqi M

F-22 Finding disease related genes by GeneRank algorithm using co-occurrence based network structures

We present a method to decide the ranks of genes for the disease related genes findings by using text mining for GeneRank algorithm. Our experiment shows that this approach outperforms the functional annotation approach with Gene Ontology.
—Lee H-M, Shin M-Y, Hong M-P

F-23 Identifying variability of human splicing forms in their interactions by using literature mining

We present a new protein-protein interaction database generated through a comprehensive text mining pipeline for human splicing variants. We analyze the PPI DB content to assess the variation in interactions of proteins involving isoforms.
—Kafkas S, Varoglu E, Rebholz-Schuhmann D, Taneri B

F-24 The meaning behind the choice of words: a case in cancer metastasis

We analyzed the language used by cancer experts while talking about several aspects of metastasis. Unlike the rigid ideas often found in the dry and strictly-written biomedical literature, our data revealed a world full of uncertainty and speculation
—Divoli A, Mendonca EA, Rzhetsky A

F-25 Linked open experimentation

We investigate the application of information technologies that can help explore knowledge beyond a single experiment and a single researcher. We link biological models, Taverna results, myExperiment, BioCatalogue, and ConceptWiki.
—Roos M, Van Haagen HHHBM, Singh B, Marshall MS, Mons BM

F-26 First CALBC Challenge – first results

CALBC aims at creating a large annotated corpus with about 5 different semantic types whose annotation is carried out automatically. The CALBC challenge I has not been terminated and first results are available.
—Rebholz-Schuhmann D, Jimeno Yepes A, Li C, van Mulligen E, Kors J, Milward D, Hahn U

F-27 EuroPhenome: a repository for high-throughput mouse phenotyping data

The Europhenome project is a comprehensive resource for raw and annotated high throughput phenotyping data arising from projects such as EUMODIC. These provide data on inbred and mutant mouse lines.
—**Morgan H**, Hassan A, Blake A, Hancock JM, Mallon AM

F-28 InSilico DB: an efficient starting point for the analysis of curated human Affymetrix gene expression microarray datasets in GenePattern

There is a large amount of Affymetrix gene expression datasets in the public domain, but reusing them requires tedious and error-prone retrieval and compilation. InSilico DB allows exports of uniform curated datasets into popular analysis platform.
—**Venet D**, Coletta A, Taminau J, Steenhoff D, Bentabet L, Walker N, Meganck S, Delgado Blanco J, de Schaetzen V, Savagner F, Rousseau F, Schymkowitz J, Detours V, Nowé A, **Weiss Solís DY**, **Bersini H**

Protein Interactions, Molecular Networks and Systems Biology

G-1 Phenotypic effects of network rewiring in regulatory hierarchies

We study the phenotypic effects of network rewiring events in regulatory hierarchies. By allowing the hierarchies to change upon deletions and insertions of nodes/edges, we find that location of changes more accurately reflects the phenotypic effect.

—Bhardwaj N, Kim PM, Gerstein M B

G-2 Bringing order to disorder: comparative genomics and genetic interactions uncover three biologically distinct forms of protein disorder

Intrinsically disordered regions are very common in many proteins. Using comparative genomics and genetic interactions we found that disorder can be split into three phenomena: flexible disorder, constrained disorder and non-conserved disorder.

—Bellay J, Han S, Michaut M, Constanzo M, Andrews BJ, Boone C, Bader GD, Myers CL, Kim PM

G-3 Prediction of genetic interactions in yeast using machine learning

We propose to use machine learning techniques to infer genetic interactions in yeast by integrating various feature sets defined on genes. The approach is validated using four available genetic interactions maps as training networks.

—Schrynemackers M, Geurts P, Wehenkel L, Madan Babu M

G-4 Prioritizing candidate genes by network analysis of differential expression using machine learning approaches

We present a method that can identify promising candidate genes using network-based machine learning approaches even if no knowledge is available about the disease or phenotype. It was benchmarked on 40 KO experiments in mice with positive results.

—Nitsch D, Gonçalves JP, Ojeda F, Tranchevent L-C, Moreau Y

G-5 The human E3 ubiquitin ligase enzyme protein interaction network

Using structural data,efficient structural comparison algorithms and appropriate filters,we construct human E3 ubiquitin ligase enzyme protein interaction network.Our analysis reveals important functional features and a priori unknown E3 interactions

—Kar G, Keskin O, Nussinov R, Gursoy A

G-6 Fast motif enumeration and clustering in integrated networks

We present novel algorithms and a Cytoscape user interface for identifying clusters of overlapping network motifs, which form functional modules in cellular networks composed of multiple types of interactions.

—Audenaert P, Van Parys T, Pickavet M, Demeester P, Van de Peer Y, Michoel T

G-7 Dynamic deterministic effects propagation networks: learning signalling pathways from longitudinal protein array data

We present a method for reconstruction of signalling networks from Reverse-Phase-Protein-Array time course data. Boolean signal propagation is combined with a Hidden Markov Model and used in a Genetic Algorithm to infer signalling networks from data.

—Bender C, Henjes F, Fröhlich H, Wiemann S, Korf U, Beißbarth T

G-8 Systematic mapping of multiple specificity in peptide recognition modules reveals new binding modes of protein domains

We develop a new computational model of binding specificity in peptide recognition modules using large experimental datasets of interacting peptides. Our method predicts unexpected protein interactions and reveals new binding modes of PDZ domains.

—Gfeller D, Ernst A, Verschueren E, Vanhee P, Dar N, Serrano L, Sidhu SS, Bader GD, Kim PM

G-9 Reverse-engineering gene regulatory networks for the abiotic stress response in Arabidopsis thaliana

We infer stress gene regulatory networks for Arabidopsis thaliana, consisting of functionally coherent coexpression modules and biologically relevant regulators. We also predict new functions for uncharacterized genes in the abiotic stress response.

—**Vermeirssen V**, De Clercq I, Van Parys T, Van Breusegem F, Van de Peer Y

G-10 Predicting protein-protein interaction using mirror tree

We introduce the Mutual Information measure as an alternative metric to evaluate the co-evolutionary information shared by a protein pair. We provide a comparative study of Mirror Tree

and enhanced versions using a large dataset of E.coli proteins.
—**Esmaielbeiki R**, Nebel J-C

G-11 Comparison of Bayesian networks and its extensions applied to the inference of regulatory networks

In this work a comparison among tree different Bayesian networks approaches is presented. BNs without any modification, with interventions and with addition of extra knowledge. To compare the methods synthetic and cytometry data are used.
—**Werhli AV**

G-12 Insight into the mechanism of specific regulation of STAT proteins by atypical dual-specificity phosphatases

We present our results based on an approach combining docking and molecular dynamic simulations to understand at a molecular level the mechanism responsible for the specific regulation of STAT proteins by atypical dual-specificity phosphatases.
—Jardin C, Sticht H

G-13 VoteDock: the consensus docking method for prediction of protein-ligand interactions

Our consensus-based docking method is able to dock properly about 20% of protein-ligand pairs more than docking methods in average, and more than 10% of pairs more than single best program. Also drop in RMSD of top conformation is observed.
—Plewczynski D, Łaźniewski M, von Grotthuss M, Rychlewski G

G-14 Inferring genetic regulatory networks with an hierarchical Bayesian model and a parallel sampling algorithm

We propose the use of a Metropolis Coupled Markov Chain Monte Carlo in order to improve a Bayesian coupling scheme for learning genetic regulatory networks from a combination of related data sets, obtained under different experimental conditions.
—Mendoza MR, **Werhli AV**

G-15 Prediction of protein-protein interactions in the apoptosis pathway

We focus on predicting protein-protein interactions in the apoptosis pathway, by structurally comparing the genes/proteins in that pathway with the template interfaces of non-obligate complexes obtained from the PDB. PRISM server is used as the prediction tool.
—Acuner Ozbabacan SE, Gursoy A, Keskin O, Nussinov R

G-16 Systematic network analysis

We present a tool for large-scale network analysis and demonstrate its relevance to the study of biological networks, in particular the inference of network features constrained by evolution. Our approach helps motivate generative models of networks.
—Agarwal S, Villar G, Jones N

G-17 Model of trisporic acid synthesis in Mucorales shows bistable behaviour

We present an ODE model for the synthesis pathway of trisporic acid in Mucorales. This pathway has some very interesting characteristics, like exchange of intermediates as well as a positive feedback loop. Our model shows bistable behaviour.
—Werner S, Vlaic S, Schroeter A, Schuster S

G-18 Simulation of immune response to Mycobacterium tuberculosis using an agent-based model

We have developed a three-dimensional agent-based model for the immune response to M. tuberculosis. This model captures the cellular migration, activation, phagocytosis and death. Also, it includes the bacillus replication and necrosis formation.
—Galvão V, Miranda JGV, Andrade RFS

G-19 Design, optimization and predictions of a coupled model of the cell cycle, circadian clock, DNA repair system, irinotecan metabolism and exposure control under temporal logic constraints

We propose a coupled model of the mammalian cell cycle, the circadian clock, the p53/Mdm2 DNA-damage repair system, the metabolism of irinotecan and the control of cell exposure to it. We discuss the use of this model for cancer chronotherapies.
—De Maria E, Fages F, Rizk A, Soliman S

G-20 Adding structural information to the von Hippel-Lindau tumor suppressor interaction network

The von Hippel-Lindau tumor suppressor gene is a protein interaction hub, controlling numerous genes implicated in tumor progression. Using structural information and computational analysis we have located three distinct interaction interfaces.
—Leonardi E, Murgia A, Tosatto S

G-21 Computational multiscale modeling of brain tumor growth

We present a novel computational multiscale model for the spatio-temporal dynamics of brain cancer. This model combines a molecular interaction network with its emerging cellular activities to describe avascular tumor progression.

—Schütz T A, Toma A, Becker S, Mang A, Buzug TM

G-22 SLIDER: a generic metaheuristic for the discovery of correlated motifs in protein-protein interaction networks

We present the metaheuristic SLIDER for finding correlated motifs in sequences of interacting proteins. We show that it outperforms existing motif-driven CMM methods and scales to large protein-protein interaction networks.

—Boyen P, Van Dyck D, Neven F, van Ham RCHJ, van Dijk ADJ

G-23 The effective interactions between proteins via coarse-grained modeling

We propose a method that enables the fast calculation of the potential of mean force between interacting protein pairs. Such information can be used for predicting protein-protein interactions.

—Pool R, Feenstra KA, Heringa J

G-24 Predicting cellular perturbation effects and identifying causal perturbations using a message-passing based machine learning method

A novel network reconstruction is proposed that is based on a message-passing scheme in which the strength of the links are learned. Its effectiveness is shown as it is able to explain effects of cellular perturbations.

—Hulsman M, Reinders MJT

G-25 Towards better receptor-ligand prediction

Receptor-ligand interactions mediate virtually every aspect of our multicellular life. Here we discuss the use of kernels to correctly pair these important receptors and ligands.

—Iacucci E, Ojeda F, De Moor B, Moreau Y

G-26 Extracting and visualising protein-protein interfaces for data-driven docking

Protein-protein interactions play a key role in many cellular processes but they are difficult to predict at the molecular level. We are developing an approach to exploit knowledge of existing PPIs to help guide protein docking calculations.

—Ghoorah AW, Devignes M-D, Smaïl-Tabbone M, Ritchie DW

G-27 Visualizing spatiotemporal information in heterogeneous biological networks with Arena3D

We present Arena3D, a visualization and analysis platform that enables the understanding of connections between biological networks and the identification of patterns in gene dynamics. It was tested on the gene regulatory network of the cell cycle.

—Secrier M, Pavlopoulos GA, Schneider R

G-28 Efficient learning of signaling networks from perturbation time series via dynamic nested effects models

We propose a novel, computationally efficient approach for reconstructing signaling cascades from measured temporal downstream effects of targeted interventions; via an extension of Nested Effects Models .

—Praveen P, Tresch A, Froehlich H

G-29 Discovering patterns of differentially regulated enzymes in metabolic pathways of tumors

We present a pattern recognition method on networks thatcomplements normal enrichment tests to detect such functionally related regulation patterns

—Schramm G,Wiesberg S, Diessl N, Kranz A-J, Sagulenko V, Oswald M, Reinelt G, Westermann F, Eils R, Konig R

G-30 Extending the yeast metabolic network using a data integration approach

We present a novel methodology, using the data integration and visualisation capabilities of Ondex, with Taverna's ability to combine bioinformatics Web Services, to create a powerful platform for analysing data to support systems biology research.

—**Fisher P**, Dobson P, Brenninkmeijer C, Canevet C, Taubert J, Rawlings C, Stevens R

G-31 Predicting genetic interactions by quantifying redundancy in biochemical pathways

We propose a novel unsupervised method to predict genetic interactions based on the information contained in protein interaction networks. Our method, combined with experiments, could shed light on the quantitative nature of biological robustness.

—Delgado-Eckert E, Gill S, Merdes G, Beerenwinkel N

G-32 Hub protein interfaces and hot region organization

Structural properties of hub proteins give clues about hot region organization in the interface.

Diversity between hot regions in hub proteins contribute to predict whether a given hub interface is constructed by two date hubs or two party hubs.
—Cukuroglu E, Gursoy A, Keskin O

G-33 A functional-genomics fermentation platform to identify and optimize industrial-relevant properties of Lactobacillus plantarum

We present a functional genomics platform that matches gene expression to industrially-relevant phenotypic traits in Lactobacillus plantarum WCFS1.
—Wels M, Bron PA, Wiersma A, Marco M, Kleerebezem M

G-34 DisGeNET: a Cytoscape plugin to visualize, integrate, search and analyze gene-disease networks

DisGeNET provides a user-friendly framework to explore human gene-disease associations by means of network analysis tools.
—Bauer-Mehren A, Rautschka M, Sanz F, Furlong LI

G-35 Combination of network topology and pathway analysis to reveal functional modules in human disease

A comprehensive database comprising human gene-disease associations for monogenic, complex and environmental diseases was subjected to topological and functional network analysis to identify and characterize disease-related biological processes.
—Bauer-Mehren A, Bundschus M, Rautschka M, Mayer MA, Sanz F, Furlong LI

G-36 Attacking interface & interaction networks

"Interface & Interaction Networks" result from integration of binding site information into PPI networks. Here proteins are depicted as nodes, interactions as edges and interfaces as a different kind of node.
—Engin B., Gursoy A., Keskin O

G-37 Pathway Projector: web-based zoomable pathway browser using KEGG Atlas and Google Maps API

Pathway Projector is a web-based pathway browser with zoomable user interface. This application allows multiple search, experimental data mapping, and editing. Pathway Projector is available at: http://www.g-language.org/ PathwayProjector/.
—Kono N, Arakawa K, Ogawa R, Kido N, Oshita K, Ikegami K, Tamaki S, Tomita M

G-38 Mapping the 'Farnesylome' - structure-based prediction of farnesyl transferase targets

We present a structure-based approach for the detection of farnesylated proteins. Using this approach we identify new putative targets for Farnesyl transferase not identified by traditional sequence-based approaches, providing new functional insight.
—London N, Schueler-Furman N

G-39 Integrative analysis of gene expression and copy number variation data to elucidate a reference human gene network

We developed algorithms and database to elucidate directed Gene Networks from gene expression and CNV. From human datasets we compiled a healthy reference network, which can be used in disease studies to identify differential disease networks.
—Orsini M, Pinna A, De Leo V, de la Fuente A

G-40 Unknown player in modelling of signal transduction pathway

We present a probabilistic models base on Nested effects Models to infer unknown players in the signalling pathway topologies from RNAi gene silencing data and the applicability of this method in controlled simulation study.
—Sadeh M, Anchang B, Spang R

G-41 Assembly of protein complexes by integrating graph clusterings

We propose a framework for integrating clustering methods for predicting protein functional modules. Besides, we provide a web service which compute six clustering results and their integration. It is interactive and powerful network analyzer.
—Chin C-H, Chen S-H, Chen C-Y, Hsiung C, Ho C-W, Ko M-T, Lin C-Y

G-42 How natural host species avoid CD4+ T cell depletion during SIV infection

Our study reveals one key mechanism by which natural host monkeys have adapted to simian immunodeficiency virus infection and therefore maintain their immune system and remain free from AIDS.
—Chan ML, Petravic J, Ortiz AM, Engram J, Paiardini M, Cromer D, Silvestri G, Davenport MP

G-43 Kinase-specific phosphorylation site prediction in Arabidopsis thaliana

We present a method for predicting kinase-specific phosphorylation sites in Arabidopsis thaliana based on finding kinases orthologous to kinases in other well-studied systems. Results obtained are subject to further computational characterisation.

—Dang T-H, Verschoren A, Laukens K

G-44 BiologicalNetworks: enabling systems-level studies of host-pathogen interactions

The approach at cross-scale data integration to study host-pathogen interactions is proposed and demonstrated on a study of the Influenza infections. The methods and data are available through the Java application .

—Kozhenkov S, Sedova M, Dubinina Y, Ponomarenko J, Baitaluk M

G-45 Genome-wide identification of disease-related genes

Elucidating the underlying disease mechanisms is crucial for understanding the onset of diseases, the development of specific diagnostic and therapeutic approaches. We present a linkage interval-independent method to identify disease-related genes.

—**Jaeger S**, Leser U

G-46 Bioinformatics tools for the investigation of proteomics aspects involved in cardiovascular diseases: from biomarker discovery to protein-protein interaction networks

The study of complex samples, by "high-throughput" proteomics technologies, requires the development of computational tools to handle the vast amount of data produced. Their employment will be shown in relation to myocardial infarction investigation.

—Di Silvestre D, Brambilla F, Brunetti P, Lionetti V, Agostini S, Cavallini C, Mauri PL

G-47 ABC database - the analysing biomolecular contacts database

We present un update of our ABC database that contais detailed information about protein-protein and protein-small molecule interfaces. The database is freely available via web interface for the statistical analysis of biomolecular contacts

—Walter P, Metzger J, Helms V

G-48 The core and pan metabolism in the Escherichia coli species

In this work, we analyze the metabolic diversity of the Escherichia coli species. Using an optimized reconstruction strategy, we built and investigated the metabolic networks of 29 E. coli strains, which showed high correlation with their phylogeny.

—Vieira G, Sabarly V, Bourguignon P-Y, Durot M, Le Fèvre F, Mornico D, Vallenet D, Bouvet O, Denamur E, Schachter V & Médigue C

G-49 BioGraph: discovering biomedical relations by unsupervised hypothesis generation

The BioGraph service integrates heterogeneous knowledge bases and allows for the successful automated formulation and ranking of comprehensible functional hypotheses relating biomedical contexts to candidate targets, e.g., for disease-gene relations.

—Liekens A, De Knijf J, Daelemans W, Goethals B, De Rijk P, Del-Favero J

G-50 Tuning noise propagation in a two-step series enzymatic cascade

We characterize noise propagation in a two-step series MAPK enzymatic cascade, which is ubiquitously conserved in eukaryotes. We identify the parameters that may be used to tune noise propagation in the enzymatic cascade.

—Dhananjaneyulu V, Sagar PVN, Kumar G, Viswanathan GA

G-51 The effect of interactome evolution on network alignment

We investigate the effect of network evolution and error on protein network alignment results. Using a distance metric for network alignments enables us to measure agreement between alignments of real and simulated networks and propose likely error rates.

—Ali W, Deane CM

G-52 Analyzing compounds' mode of action by biological relatedness of proteins.

Mode of action is primarily analyzed by gene expression. As not all regulation happens on the transcriptional level, we enrich protein interactions by comprehensive data to detect the biological context leading to/caused by expression changes.

—Schmid R, Baum P, Ittrich C, Fundel-Clemens K, Lämmle B, Birzele F, Weith A, Brors B, Eils R, Mennerich D, Quast K

G-53 Logical modelling of the regulatory network controlling the formation of the egg appendages in Drosophila

We present a map and a dynamic logical model of the intracellular regulatory network controlling dorsal appendage patterning in drosophila oogenesis. Our preliminary results support the hypothesis that Broad inhibits Fas3 in the roof cells.

—Fauré A, Vreede B, Sucena É, Chaouiya C

G-54 A curated database of microRNA mediated feed-forward loops involving MYC as master regulator

We present and discuss a database of mixed microRNA / Transcription Factor Feed-Forward regulatory Loops having MYC as master regulator and characterized completely by experimentally supported regulatory interactions, in human.

—El Baroudi M, Corà D, Bosia C, Osella M, Caselle M

G-55 The Biological Connection Markup Language : a SBGN compliant data format for visualization, filtering and analysis of biological pathways

BCML allows a SBGN compliant representation of pathways in a visual and machine-readable format. BCML permits a selection of elements basing on specific biological evidence, creating pathways usable for both the bioinformatician and the biologist.

—Calura E, Beltrame L,, Popovici R, Rizzetto L, Rivero Guedez D, Donato M, Romualdi C, Draghici S, Cavalieri D

G-56 Estimating the size of the S.cerevisiae interactome

Our method for estimating an interactome size effectively combines high-throughput and literature-curated data. We found a higher estimate for the S.cerevisiae interactome, showing that completing the yeast interactome map require extensive effort.

—Sambourg L, Thierry-Mieg N

G-57 Laplacian eigenmaps, penalized principal component regression on graphs, and analysis of biological pathways

We propose a network-based approach for analyzing the significance of biological pathways using Laplacian eingenmaps with Neumann boundary conditions, which provides a dimension reduction method that directly incorporates the network information.

—Shojaie A, Michailidis G

G-58 Predicting metabolic pathways from bacterial operons and regulons

We use the pathway extraction tool to predict metabolic pathways from bacterial genomes, by extracting from a metabolic network the subgraphs that connect at best a set of seed genes that belong to a same operon or regulon.

—Faust K, Croes D, Dupont P, van Helden J

G-59 Inferring translationally active RNA binding proteins - mRNA interactions from polysomal profiling data with a Bayesian inference approach

We present a Bayesian inference approach to predict direct or indirect interactions between RNA binding proteins and their targets based on the comparison between translatome and transcriptome profiling experiments.

—Tebaldi T, Sanguinetti G, Niranjan M, Quattrone A

G-60 DASS-GUI – a new data mining framework

We have developed the new data mining framework DASS-GUI. It allows the pattern identification itself and additional analyses of the identified patterns .

—Hollunder J, Van de Peer Y, Wilhelm T

G-61 EnrichNet: network-based gene set enrichment analysis

EnrichNet extends classical gene set enrichment analysis by mapping gene and protein sets of interest onto interaction networks to investigate their associations using network distance calculations and sub-network visualizations.

—Glaab E, Baudot A, Krasnogor N, Valencia A

G-62 Interacting copy number alterations in breast cancer

We present a framework for the detection of co-occurring copy number alterations from high-density array data. The method is applied to 216 breast tumours, revealing several deregulated gene interactions driven by interdependent genomic alterations.

—Canisius S, Klijn C, Smid M, Martens J, Foekens J, Wessels L

G-63 Compression of mass spectral imaging data using discrete wavelet transform guided by spatial information.

The presented method reduces the size of MSI data sets considerably while still achieving excellent reconstruction of the original mass spectra. This is done by retaining only a limited number of wavelet coefficients that express spatial structure.

—Verbeeck N, Van de Plas R, De Moor B, Waelkens E

G-64 Learning ancestral polytrees for HIV-1 mutation pathways against nelfinavir

This retrospective study shows that even in the presence of sampling biases and confounding effects ancestral polytrees can depict HIV-1 resistance mutation pathways under drug selective pressure from nelfinavir, a protease inhibitor .

—Li GD, Beheydt G, Bielza C, Larrañaga P, Camacho RJ, Grossman Z, Torti C, Zazzi M, Prosperi M, Kaiser R, Van Laethem K, De Maeyer M, Jansen M, Vandamme AM

G-65 Ranking of genes from RNAi perturbation dat

We present a comparative study in which we test methods for ranking genes from RNAi perturbation data. We identify the most robust techniques and validate them on first and secondary RNAi screening data.

—Siebourg J, Beerenwinkel N

G-66 A network centered approach for genome wide data integration applied to Alzheimers disease prediction

Network centered analysis of SNP and expression data yields predictive and biologically relevant signatures for Late Onset Alzheimers Disease.

—van den Akker E, Heijmans B, Kok J, Slagboom P, Reinders M

G-67 How to efficiently hunt the disease causing gene

We present Endeavour, a tool that identifies the most promising candidate genes from large gene lists by integration of multiple genomic data sources. It has been extensively benchmarked and also experimentally validated.

—Tranchevent L-C, Aerts S, Van Loo P, Hassan BA, Moreau Y

G-68 Study of the ultrasensitivity of the BCL-2 apoptotic switch

In our work we compare different hypothesis about functioning of the Bcl-2 apoptotic switch. We used mathematical modelling and measuring of the steady-state stimulus-response sensitivity as the judging criterion of plausibility of these hypotheses.

—**Tokár T**, Uličný J

G-69 Improved genome-wide protein-protein interaction prediction and analysis of biological process coordination in Escherichia coli

We build an improved and high-quality genome-wide interaction map for Escherichia coli, using different machine learning methods and genomic feature datasets from several sources. These predictions are used to propose various pathway interactions

—Muley VY, Ranjan A

G-70 A generic gene regulatory network reconstruction method: application to Lactococcus lactis MG1363

We present a novel method for the whole-genome reconstruction of regulatory networks. A classification model is trained using known interactions for E. coli and B. subtilis. This model is used to classify gene interactions in any prokaryote.

—Theunissen D, Brouwer R, Kuipers O, Hugenholtz J, Siezen R, Van Hijum S

G-71 Dissecting the specificity of E2-E3 interaction in the ubiquitination pathway by a molecular docking approach

Ubiquitin enzymes are key regulating units for many fundamental processes required for cell viability. In this work, we used our molecular docking software HADDOCK to get insights at an atomic level on the specificity of E2/E3 interactions.

—Melquiond ASJ, Rodrigues J, Bonvin AMJJ

G-72 PopCover – selecting peptides with optimal population and pathogen coverage

PopCover integrates pan-specific MHC binding predictions with MHC allelic prevalence and select pools of peptides that in concert will provide broad coverage of both pathogen genomic and the population MHC diversity.

—**Lundegaard C**, Buggert M, Karlsson AC, Lund O, Perez C, Nielsen M

G-73 A framework for functional selection of biomarkers

Molecular biomarkers can be used for evaluating the health state of an individual. Here, a framework for biomarker selection is presented, and the ability of the selected markers to cover different diseases is evaluated.

—**Kivinen V**, Nykter M, Yli-Harja O, Shmulevich I

G-74 iPath: interactive pathways explorer

iPath is a web-based tool for the visualization and analysis of the metabolic pathways. An interactive viewer provides straightforward navigation through various pathways and enables easy access to the underlying chemicals and enzymes.

—**Yamada T**, Letunic I, Okuda S, Bork P

G-75 In silico comparative modeling of MTHFR A1298C polymorphism in acute leukemia

Alterations in folate levels are related with the risk factor for the cancer. The C677T and A1298C mutations of MTHFR were associated with

susceptibility in leukemia. Simulations in silico are proposed to clarify a paradoxal comportment to A1298C.
—Ramos F, Lima J, **Melo J**

G-76 Feasibility space as a tool to understand regulation of metabolic networks

Hierarchical regulation analysis shows that metabolic networks are regulated both hierarchically and metabolically, but it is unclear why and when. We study whether basic physiological constraints can help explain the observed regulatory mechanisms.
—Nikerel IE, Hu F, Berkhout J, Teusink B, Reinders MJT, **De Ridder D**

Genomic Medicine

H-1 DISCOVERY: A resource for the rational selection of drug target proteins and leads for the malaria parasite, Plasmodium falciparum

The DISCOVERY project provides a resource for the rational in silico selection of putative drug target proteins and lead compounds in malaria, using data mining on automated protein annotations and the creation of relations to ligand molecules.
—Odendaal CJ, Harrison CM, Szolkiewicz MS, Joubert F

H-2 Variance estimators for t-test ranking influence the stability and predictive performance of microarray gene signatures

Identification of differentially expressed genes in cancer prognosis or diagnosis is commonly performed with a t-Test. Specific variance estimators are shown to affect the classification performance and the stability of small gene signatures.
—Touleimat N, **Hernández-Lobato D**, Dupont P

H-3 Selecting small subsets of genes for predicting the outcome of chemotherapy treatments : a dynamic programming approach

We adress the issue of selecting small sets of genes involved in the response to a preoperative chemotherapy treatment in breast cancer. Our approach is to select genes that maximize the `responders - non responders' interclass distance.
—Natowicz R, Moraes Pataro C-D, Incitti R, Costa M-A, Cela A, Souza T, Braga A-P, Rouzier R

H-4 Networks of generalized top scoring pairs for robust phenotype classification

We present a network analysis of the top scoring pairs classifiers and we show how one can build more powerful, but still simple, decision rules by exploiting the graph topologies induced by TSP. Also, we introduce a generalized version of TSP and its parallelized implementation.
—Popovici V, Budinska E, Delorenzi M

H-5 Comparing network and pathway based classification for breast cancer. Network and pathway based classifiers do not outperform single gene classifiers

We compared the accuracy of three methods to predict outcome in breast cancer based on network or pathways data in combination with gene expression data with the performance of a single gene classifier. We found that single gene markers perform best.
—Staiger C, Kooter R, Dittrich M, Müller T, Klau GW, Wessels L

H-6 Enterotypes of the human gut microbiome

In the gut metagenome of 39 individuals from 6 countries we identified several robust clusters that suggest that the intestinal microbiota variation is stratified, not continuous. We study the influencing factors of this phenomenon.
—Raes J, Arumugam M, The Metahit Consortium, Dore J, Weissenbach J, Ehrlich, SD, Bork P

H-7 Identifying chemotherapy resistance genes using outlier detection

We have developed a novel algorithm that allows the detection of genes that show class specific differential expression in a subset of samples. We use this algorithm to find genes related to chemotherapy resistance.
—de Ronde J, Mulder L, Lips E, Rodenhuis S, Wessels L

H-8 Genome wide association study of non-synonymous single nucleotide polymorphisms for seven common diseases

We have conducted a genome wide scan for identifying association of non-synonymous SNPs with seven common diseases. In our study we identified 18 new associations with diseases studied in additional to the 24 associations publised by WTCCC study.
—Surendran P, Shields D, Stanton A

H-9 Comparative bioinformatics approach to multiple tumors reveals novel prognostic markers in breast cancer

We provide a comparative analysis of pathways enriched in multiple tumors and disclosed new prognostic markers by applying gen set enrichment analysis to pathway groups highly conserved among the tumors.

—Krupp M, Marquardt J, Maas T, Galle P, Tresch A, Teufel A

H-10 Genomic and epigenomic molecular signatures reveals network mechanisms associated with ovarian cancer prognosis

The Cancer Genome Atlas has recently made available the molecular characteristics of more than 200 patients. Using computational network models, we were able to find sub-networks that significantly stratify survival rates.These interactions are assoc

—**Ben-Hamo R**, Efroni S

H-11 Inferring distributions of trait-associated SNPs with application to genetic association studies

Published SNP associations were used to learn SNPs' a-priory potential of being trait-associated. The predicted potential is integrated with classic association methods. Results are improved for 16/19 simulated diseases, and for Type-2 diabetes data.

—Neuvirth H, Aharoni E, Azencott C-A, Farkash A, Landau D, Rosen-Zvi M, Geiger D

H-12 Optimizing exon CGH array designs for robust rearrangements detection.

We propose new approach to measure the quality of array CGH designs focusing on robustness of rearrangements detection to the noise. We implemented the efficient Monte Carlo method for testing noise robustness within DNAcopy procedure.

—Gambin T, Sykulski M, Gambin A

H-13 Interactive human brain atlas for a better understanding of diseases

The ever increasing complexity & amount of research data require new approaches to data analysis & visualization. We describe an application that facilitates the integration & visualization of gene expression data in the context of the brain anatomy.

—Kremer A, Gaenzler C, van der Spek P

H-14 Clinical prognosis with transcriptional association networks and regression trees

We present a new supervised prediction method for prognostic problems based on the discovery of clinically-relevant association networks. The method integrates regression trees and clinical class-specific networks.

—Nepomuceno I, Azuaje F, Devaux Y, Nazarov PV, Muller A, Aguilar-Ruiz JS, Wagner DR

H-15 Prediction of clinical outcome after cardiac arrest and induced therapeutic hypothermia

We analyzed gene expression biosignatures relevant to the prediction of clinical outcome after cardiac arrest . In order to identify mechanisms potentially driving a clinical outcome, we developed a method to infer class-specific networks.

—Nepomuceno I, Azuaje F, Devaux Y, Stammet P, Aguilar-Ruiz JS, Wagner DR

H-16 Allele-specific copy number analysis of breast carcinomas

Using a novel algorithm, we present the first allele-specific copy number analysis of the in vivo breast cancer genome, and identify specific recurring signatures of aberrations.

—Van Loo P, Nordgard SH, Lingjærde OC, Russnes HG, Rye IH, Sun W, Weigman VJ, Marynen P, Zetterberg A, Naume B, Perou CM, Børresen-Dale A-L, Kristensen VN

H-17 Gene dosage balance and the evolution of protein interaction networks

We present evidence of different mechanisms to control gene dosage balance between hubs in vertebrates and lower eukaryotes. Such mechanisms reflect different properties in terms of protein-protein interactions, gene origin and gene duplicability.

—**D'Antonio M**, Ciccarelli FD

Other Bioinformatics Applications

I-1 Universal virtual screening

We optimized scoring functions of molecular docking programs for virtual screenings of universal targets. The optimization improves enrichment rates by 77.2%, and approximately 70% of ligands for target proteins were predictable.

—**Onodera K**, Kamijo S

I-2 Accuracy and computational efficiency of a graphical modeling approach to linkage disequilibrium estimation

We develop recent work on using graphical models for linkage disequilibrium to provide efficient programs for model fitting, phasing and imputation of missing data in large data sets.

—Abel H J, Thomas A

I-3 Stepwise classifier for heterogeneous genomic data

We introduce stepwise classification method which use different data types independently, so that intrinsic classification power of each data types play their role without overcome by others.Experiment result showed it's distinct classification power

—Wubulikasimu A, van de Wiel MA

I-4 ALADYN: a web server for aligning proteins by matching their large-scale motion

We present the ALADYN web server which aligns proteins by matching their large-scale internal motion. The method allows to efficiently detect meaningful relationships between proteins that are not necessarily well-alignable in sequence or structure.

—Potestio R, Aleksiev T, Pontiggia F, Cozzini S, Micheletti C

I-5 Fitness differences as an explanation for difference between chronic myeloid leukemia therapies

Using a computational models of the hematopoietic dynamics, we show here why Nilotinib, a second line treatment for chronic myeloid leukemia, has a faster and deeper response than Imatinib.

—Lenaerts T, Castagnetti F, Traulsen A, Pacheco JM, Rosti G, Dingli D

I-6 Stochastic extinction of leukemic stem cells provides a road to cure chronic myeloid leukemia

We show through a computational model of the hematopoietic system that tyrosine kinase inhibitors like Imatinib can cure CML as a result of stochastic effects in the system.

—Lenaerts T, Traulsen A, Pacheco JM, Dingli D

I-7 sscMap perturbation: unbiased candidate therapeutic ranking in connectivity mapping

Connectivity mapping is a method for discovering connections between different biological states based on gene-expression similarities. A perturbation approach aids in the selection of potential therapeutics among discovered drug-disease connections.

—McArt D, Zhang S-D

I-8 In silico study of expression profile correlation between microRNAs and cancerous genes

We investigate the possibility that microRNA can act as an oncogene or tumor suppressor gene. Experimentally verified microRNA target genes information (TarBase) are integrated with microRNA and mRNA expression data (NCI-60) to study this hypothesis, in which the Pearson correlation and Spearman rank coefficients are used to quantify these relations for nine cancer tissues.

—Ng K-L, Weng C-W, Huang C-H

I-9 Harry Plotter: a user friendly program to visualize genome and genetic map features

We present Harry Plotter a simple program that visualizes genomic scaffolds anchored to a genetic map and chromosome features distribution, like QTL significance, gene density, repeated elements and GpC abundance, using multi-colored gradients.

—Moretto M, Cestaro A, Troggio M, Costa F, Velasco R

I-10 Inhibitory activity of Thiadiazoles on Protein Kinase PKnB from Mycobacterium Tuberculosis: a virtual screening and molecular docking study

Tuberculosis causes more than two million deaths per year. We did the study of 1800 different Thiadiazoles to discover new derivatives that can be drugs with higher potency and specificity.

—Raj U, Singh V-B, Swati, Srivastava A, Naqvi S-A-H

I-11 A microarray format standardization study: meaningful structures

Microarray data is too big and disparate. Repositories cannot exchange data. The records are not visible as required. We extended the structure and semantics of MINiML file on GEO within a metadata framework with a case study to prove the performance

—Kocabas F, Can T, Baykal N

I-12 Normalization of proteomic ratio data.

In order to correct proteomic ratio data for variable loading we commonly centre the logarithm of the ratios. Here we set this “normalization” in a larger context, explain why the common practice is a

favoured, and show that sometimes we can do even better.

—Sherman J, Molloy M, Descallar J, Lance B, Uitto P, Wood G

I-13 BioCatalogue: the curated catalogue of life science web services

We present the BioCatalogue, a registry of curated biological web services. A place where web service providers, users and curators can register, annotate, monitor and search for web services.

—Tanoh F, Bhagat J, Eric Nzuobontane E, Laurent T, Wolstencroft K, Stevens R, Pettifer S, Lopez R, Goble C

I-14 Regression system for prediction of errors in the data on gene expression in situ obtained from confocal images

We present a regression system for the prediction and correction of errors in gene expression data extracted from confocal images. The system is based on the method based on the censoring technique that we have published elsewhere.

—Myasnikova E, Surkova S Samsonova M

I-15 Impact of genetic variations on phosphorylation sites

We analyzed the cross-talking between mutations, alternative splicing and phosphorylation sites in proteins and transcripts, describing how evolution shapes genes creating different contexts in which a given site is immersed.

—Via A, Le Pera L, Ferré F, Tramontano A

I-16 FuSiGroups: grouping gene products by functional similarity

We present FuSiGroups, a novel algorithm to group gene products based on functional similarity. Unlike existing approaches, FuSiGroups allows each gene product to be in multiple groups, representing its association with different biological concepts.

—Welter DN, Gray WA, Kille P

I-17 A file system strategy for high-performance sequence interval queries of very large datasets

We present results of performance tests of a file system solution for fast sequence interval queries of large volumes of genomics feature data. GF2S efficiently provides scalability, speed and flexibility for visualization of huge datasets.

—**Karcz SR**, Links MG, Parkin IAP

I-18 Evaluation of multivariate data analysis strategies for high-content screening

Multivariate data, as generated by high-content screening, is analyzed using a modular analysis pipeline. Dimension reduction methods and approaches to summarize single cell results are such easily evaluated to determine the best analysis strategy.

—Kümmel A, Parker CN, Gabriel D

I-19 Studies of the binding capacity of cyclooxygenase II and biodistribution of phenols contained in natural products

A set of phenols present in propolis and grape were studied in their interaction with an inflammatory processes mediator, the ciclooxygenase II enzyme. Additionally, their biodistribution properties were studied using in silico techniques.

—Paulino Z-M, Aguilera M-S, García Jenifer García, Iribarne F

I-20 Finding logic networks using the Dutch Life Science Grid: handling 15 million jobs

We present a method that enables running a huge amount of jobs on grid infrastructure. The solution involves a pilot job framework for scheduling jobs and an output receiver for storing the outputs. It scales beyond 5k cores and can handle 15M jobs.

—Bot J, de Ridder J, Reinders M

I-21 A detailed view on several model-based multifactor dimensionality reduction methods for detecting gene-gene interactions in case-control data in the absence and presence of noise

Model-Based Multifactor Dimensionality Reduction is a promising method for detecting epistasis that flexibly deals with different outcome types. Its power and type I error are evaluated in a variety of simulation settings.

—Cattaert T, Van Lishout F, Mahachie John JM, Van Steen K

I-22 Quality ranking of 16S sequences: an approach based on poset theory

We present a new approach to rank 16S rRNA sequences according to several predefined quality criteria. We devised applications and compared the results with those of a generally available manually curated data set.

—De Smet W, Verslyppe B, De Loof K, De Vos P, De Baets B, Dawyndt P

I-23 Clustering and kernel methods using R

Two different kernel based clustering methods, k K-Means and spectral clustering, for unsupervised learning on chemoinformatics data are presented in this poster.

—**Adefioye A**, De Moor B

I-24 Ensuring data integrity in large modern integrative genomic studies

We present an algorithm for the detection and correction of sample mis-identification in studies combining genotypes with microarray expression data. We illustrate its worth with simulated errors in a breast cancer dataset.

—Lynch A, Dunning M, Chin S-F, Curtis C

I-25 Testing branches in dendrograms representing species abundance data

We apply four approaches for dendrograms representing species abundance data, and analyse which method enables to identify biologically relevant clusters.

—Calkiewicz J, Wlodarska-Kowalczuk M, Wrobel B

I-26 Reliable identification of hundreds of proteins without peptide fragmentation

We present evidence that, combining high precision mass determination and modern prediction of the retention times of tryptic peptides with a robust scoring system, we can detect hundreds of proteins from HPLC-MS experiments with a low FDR.

—Bochet P, Rügheimer F, Guina T, Brooks P, Goodlett D, Clote P, Schwikowski B

I-27 Exploring unassigned peaks in protein fragment mass spectra with frequent itemset mining techniques

We introduce an approach to exploit information from collections of unassigned spectral peaks lists by using frequent itemset mining techniques, and evaluate it based on a set of over 20.000 validated peptide mass fingerprint spectra.

—Vu T-N, Valkenborg D, Eeckhout D, De Jaeger G, Goethals B, Witters E, Lemière F, Laukens K

I-28 Benchmarking a new semantic similarity measure using reference sets and clustering: evaluation and interpretation for the discovery of missing information

Several semantic similarity measures exist using Gene Ontology controlled vocabulary. Comparing performances requires evaluation criteria. We have defined a new measure and set up an evaluation methodology based on reference sets.

—Benabderrahmane S, Smaïl-Tabbone M, Napoli A, Poch O, Devignes MD

I-29 Utopia:GPCRDB: a domain-specific PDF reader.

We present a GPCR-specific PDF reader that dynamically links concepts in scientific articles to relevant entities in the GPCRDB, providing readers with text-specific annotations containing data and information on i.e. GPCRs, residues and mutations.

—Vroling B, Thorne D, McDermott P, Pettifer S and Vriend G

I-30 Locally optimal structures of an RNA

RNA locally optimal secondary structures give a rich picture of the folding space of an RNA. We propose an efficient algorithm computing all such structures for a given RNA sequence. We implemented this new method in a software called REGLISS.

—Saffarian A, Giraud M, Touzet H

I-31 Anesthetics and the role of the biological membrane in the molecular regulation of K2P potassium channels

The TREK-1, a two-pore domain potassium channel, is of paramount importance for anesthesia. We have investigated its structure in a membrane environment and shown that the channels opening must be an indirect reaction of the membrane swelling.

—Bernardi RC, Treptow W, Klein ML

I-32 Comparison of validation methods for merging cancer microarray data sets

Gene expression information abounds in public databases. However, this information is fragmented in small datasets and thus of limited use. We benchmark datasets merging methods designed to achieve greater statistical power.

—Taminau J, Meganck S, Weiss-Solis DY, Van Staveren WCG, Dom G, Venet D, Bersini H, Detours V, Nowé A

I-33 Mapping insertions to putative target genes

We investigate the association of insertions with gene expression in MuLV and Sleeping Beauty data sets. We present a knowledge-based approach to map insertions to target genes and demonstrate its superiority over nearest gene target selection.

—De Jong J, De Ridder J, Sun N, Van Uitert M, Adams, DJ, Wessels LFA

I-34 The zebrafish embryo in toxicology: A combination of toxicogenomics and other screening techniques

We present here our system for studying toxicological effects in zebrafish embryos. In our approach, we combine gene expression systems like microarrays and automated high throughput imaging techniques.

—Legradi J, Yang L, Alshut R., Ho N, Klüver N, Scholz S, Mikut R, Reischl M, Liebel U, Strähle U

I-35 Disease gene prediction based on tissue-specific conserved coexpression

We show tissue-specific and multi-tissue conserved coexpression to be highly complementary and exploit tissue-specific relationships between disease and candidate genes to provide novel high-confidence candidates for several genetic disorders.

—Piro RM, Ala U, Molineris I, Grassi E, Damasco C, Bracco C, Provero P, Di Cunto F

I-36 Improved classification by integrating multiple patient data sets with literature information using co-inertia analysis

We describe a mathematical frame work to integrate two genome scale expression datasets of patients and its corresponding literature information from PUBMED as prior data sets. We illustrate this frame work improve the classification ac

—Thomas M, Daemen A, De Moor B

I-37 Time-resolved signatures from weighted visibility graph representation of time-series data

We provide a representation of time series via directed weighted visibility graphs. From this a measure is employed in statistical tests that account for the dependencies between not only the time points but also the entities from which time-series data were obtained.

—Nikoloski Z, Grimbs S, Schäfer R, Sers C, Selbig J

I-38 Genome Sequence of the Edible Cyanobacterium Arthrospira sp. PCC 8005

The cyanobacterium Arthrospira sp. PCC 8005 is of great interest to the European Space Agency for its nutritive value and oyxgenic properties in the MELiSSA biological life support system for long-term manned missions into space.

—Janssen PJ, Morin N, Monsieurs P, Mergeay M, Leroy B, Wattiez, Vallaeys T, Waleron K, Waleron M, Wilmotte A, Quillardet P, Tandeau de Marsac N, Talla E, Zhang C-C, Médigue C, Barbe V, Ley N

I-39 Moa: managing command line bioinformatics

Bioinformatics projects are often command line driven, while this allows flexibility it can make maintenance difficult. Moa is designed to address this issue; a tool to organise and maintain command line projects whilst retaining flexibility.

—**Fiers M**

I-40 SAAPdb: structural effects of single amino acid polymorphisms

We present an update of SAAPdb, a database of single amino acid polymorphisms and the effects they are likely to have on protein structure. This update includes new mutations, a new structural effect and analyses of several datasets.

—Baresic A, Alnumair N, Martin ACR

I-41 Detecting human proteins involved in virus infection by observing the clustering of infected cells in siRNA screening images

We present a clustering method analyzing siRNA screening images of HCV/DV infected cells to detect host factors needed for the replication of the virus. The results were compared with common intensity readouts and promising host factors were found.

—**Suratanee A**, Rebhan I, Matula P, Kumar A, Kaderali L, Rohr K , Bartenschlager R, Eils R, König R

I-42 A benchmark on cancer classification using LS-SVMs and microarray data

We compare several cancer classification models based on the Least Squares Support Vector Machine (LS-SVM)s and micrroarray data. In total, we test 120 combinations of preprocessing methods, feature selection methods, kernel functions and kernel parameter

—**Popovic D**, Daemen A, De Moor B

I-43 MiRPara: a SVM-based software tool for prediction of mature microRNAs

We have developed an SVM based system for microRNA prediction based on results from experimental studies. Our software can be run on full length genome sequences from any species. In random tests, the software outperformed existing prediction tools

—Wu Y, Liu H, **Rayner S**

I-44 A Platform for identifying prostate-cancer-related microRNA and mRNA using the empirical Bayes method in analysing microarray data

A platform has been set up to study the regulatory role of miRNAs in tumorigenesis. Certain putative pairs of miRNA-mRNA are confirmed to be cancer related, hence, the effectiveness of the present approach is demonstrated.

—Chen S-T, **Ng K-L**

I-45 PRIDE Inspector: a new tool to browse, visualize and review proteomics data

PRIDE (http://www.ebi.ac.uk/pride) is one of the main public repositories of MS proteomics data. PRIDE Inspector (http://code.google.com/p/pride-inspector) is an open source application to visualize and perform a first quality assessment of MS data.

—**Wang R**, Ríos D, Reisinger F, Vizcaíno JA, Hermjakob H

I-46 The scientist/staff, project collaboration and content management system (PCCMS)

We present a scientist and project collaboration and relationship management platform, designed to streamline and integrate Omics experiment data and project management for various technologies in genomics, proteomics and bioinformatics.

—**Kumuthini J**, Dominy J

AUTHOR INDEX