SIB
Swiss Institute of Bioinformatics

SIB Profile 2016

SIB, a crucial link in the life science chain

SIB | Swiss Institute of Bioinformatics
Quartier Sorge
Bâtiment Génopode
CH-1015 Lausanne
Switzerland
t +41 21 692 40 50
f +41 21 692 40 55

www.sib.swiss

elixir SWITZERLAND

Table of contents

Foreword

2015 was a particularly fruitful year for SIB. To name but a few, the Food and Agriculture Organization of the United Nations (FAO) appointed the institute as its Reference Centre for Bioinformatics. ELIXIR – the pan-European research infrastructure for biological information of which SIB is the largest national node – was awarded 19 million euros by the European Commission. SIB was invited to the 10th edition of the Swiss Innovation Forum, the national platform for promoting innovation, creativity and design which attracts leaders in their fields. SIB's computational solutions in two essential fields – medicine and drug development – met with great success.

"Bioinformatics continues to be at the heart of science."

In the field of research, SIB announced its collaboration with Ariana® Pharma to develop early detection tools for gastric cancer. The institute collaborated in discovering new possibilities for the treatment of breast cancer as well as treatment options for an incurable paediatric leukaemia, and contributed to the largest ever genome-wide study which strengthened the genetic link to obesity. On a lighter note, studies within SIB on the chameleon created a global buzz when a group discovered how the reptiles change colour and regulate their temperature.

As is the case every year, SIB and its services continue to expand. During the course of 2015, five new group leaders and three new partner institutions joined the institute. Currently, SIB counts some 60 research and service groups, 150 resources, 19 institutional members and 11 core facilities. And the trend is showing no signs of slowing down. However, it is less a question of quantity than quality: one of SIB's missions is to lead and coordinate the field of bioinformatics throughout Switzerland. The more groups there are, the more information will flow, thus serving research not only on the national level but also on the international level. The number of embedded bioinformaticians will also increase, as will the bioinformatics training programmes for life scientists, clinicians and bioinformaticians and the programmes for undergraduate and PhD students.

Bioinformatics continues to be at the heart of science. Year after year, SIB proves that its existence and organization are essential on the national and international levels. We would like to thank the Swiss government, the Federal Assembly, the State Secretariat for Education, Research and Innovation, the Swiss National Science Foundation and all those in funding roles, as well as our partner institutions, for their unwavering and invaluable support.

We would also like to express our heartfelt gratitude to all SIB members whose expertise and dedication have helped bring Swiss Bioinformatics to where it is today.

Felix Gutzwiller
President of
the Foundation Council

Manuel Peitsch
Chairman of the Board
of Directors

Ron Appel
Executive Director

What is bioinformatics?

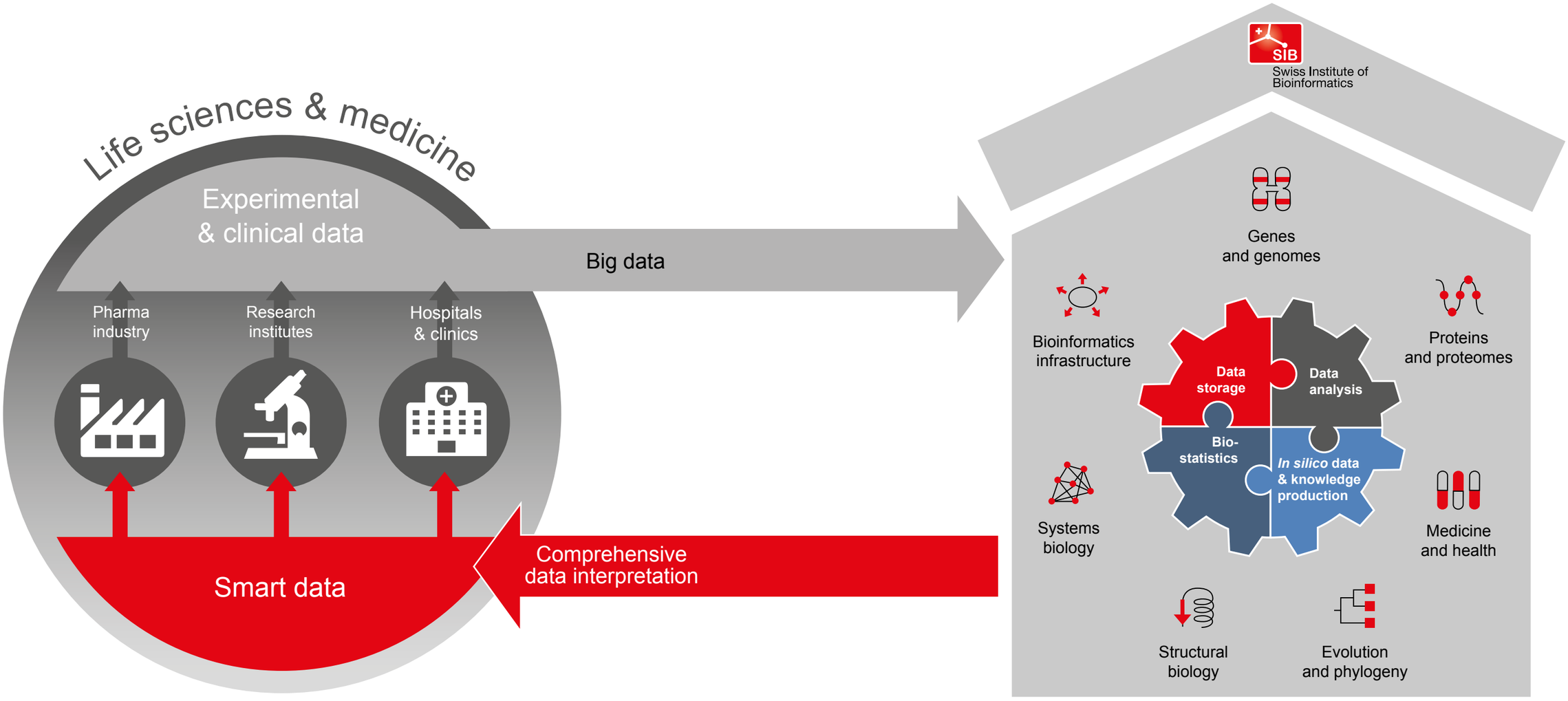
Bioinformatics is the application of computer technology to the understanding and effective use of biological and clinical data. It helps to convert "big data" into "smart data" or knowledge. Ensuring online access to bioinformatics resources as well as training and support from skilled bioinformaticians is essential for medicine and life sciences.

Computing has become a central component of modern life science projects performed by the pharma industry, research institutes, and hospitals and clinics: large volumes of experimental and clinical data ("big data") are generated by increasingly automated measuring devices. These data need to be stored, organized and analysed to extract new insights and knowledge. Extensive computational comparison of large datasets and computational simulation has become a third pillar of science – along with experimentation and theory – allowing researchers to advance their understanding of complex systems *in silico*.

Bioinformatics provides:

- **Databases and knowledgebases** for storing, retrieving and organizing biological data
- **Software** for modelling, visualizing, analysing, interpreting and comparing biological data
- **Computing and storage infrastructure** for "big data" processing
- **Biocuration and bioinformatics expertise** enabling life scientists to have accurate and comprehensive representation of biological knowledge and take full advantage of bioinformatics technologies.

Bioinformatics is thus an interdisciplinary field that brings major advances in many different life science and medical areas (for more details, please see corresponding sections on pp. 40-77).



About SIB

SIB, a crucial link in the life science chain

The SIB Swiss Institute of Bioinformatics is a unique success story at the frontier of life sciences and computer science. When the Institute was founded in 1998, bioinformatics was still in its infancy, both in Switzerland and abroad. Today SIB is an independent, non-profit foundation recognized of public utility that provides world-class bioinformatics to the national and international life science community.

By sharing its expertise in storage, analysis and dissemination of large biological datasets and through education and collaborations with research institutes and industrial partners, SIB has created a true bioinformatics culture in Switzerland, which counts today the highest concentration of bioinformaticians in the world.

The Institute continues to lead developments in the field of bioinformatics. Well aware that the wealth of data produced by modern technologies and the growing self-awareness of patients will change the way of considering medical data, SIB is now also embracing the challenge to strive for the excellence of bioinformatics in personalized health.

For more information: www.sib.swiss

19
institutional members
spread across Switzerland

Some
750
members

61
research
and service groups

11
core facilities

More than
1,300
peer-reviewed articles
published
over the past 15 years

More than
150
resources accessible
through the SIB portal
ExPASy

26
bioinformaticians
embedded in research
and clinical labs

Vision

The SIB Swiss Institute of Bioinformatics **fosters excellence in data science** to support progress in biological research and health.

Mission

SIB **leads and coordinates the field of bioinformatics in Switzerland**. Our data science experts **join forces** to advance biological and medical research and enhance health by:

1. Providing the national and international life science community with a state-of-the-art **bioinformatics infrastructure, including resources, expertise and services**
2. **Federating** world-class researchers and delivering **training** in bioinformatics

To achieve its mission, SIB is committed to:

Federate bioinformatics research groups from Swiss universities and research institutes

Foster collaboration and innovation at the highest level of scientific excellence

Create, maintain and disseminate **core bioinformatics databases, software and services** worldwide

Offer **key competencies and research support** in bioinformatics to the national life science community

Train first-rate researchers

Organization

SIB, an efficient collaborative Swiss model

The decentralized, federating organizational structure of the SIB Swiss Institute of Bioinformatics is an efficient collaborative model for countries setting up their own bioinformatics infrastructure, as well as for the European bioinformatics programme ELIXIR.

SIB is an independent, non-profit foundation with a unique organization modelled on Switzerland's federal structure. SIB consists of bioinformatics research and service groups from the major Swiss schools of higher education and research institutions (see Swiss map on the right). While SIB group leaders are senior academic staff of the partner institutions, a number of SIB scientists are paid directly by the Institute (see p. 13). Although each research group carries out its own research and teaching activities independently within its host institution, it can benefit from a wide range of resources and support provided by SIB. Moreover, SIB offers its partner institutions an efficient nationwide coordination of bioinformatics research, core resources and teaching activities. In return, the Swiss universities and research institutes provide SIB members with the infrastructure needed to perform their tasks.

SIB at ELIXIR

SIB acts as the Swiss node of ELIXIR, the pan-European bioinformatics programme. ELIXIR has recently received funding from the European Union to accelerate the implementation of Europe's life science data infrastructure.

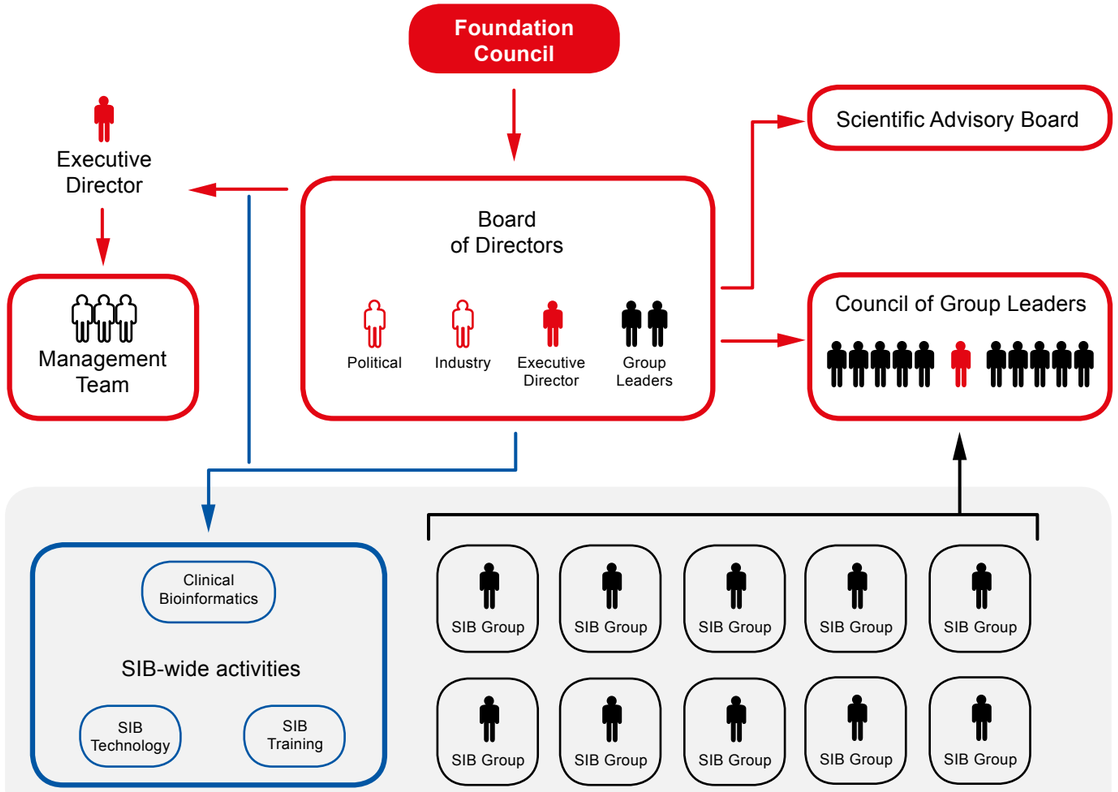
ELIXIR is a pan-European bioinformatics programme whose mission is to build a sustainable European infrastructure for biological information to support life science research and its translation to medicine and other fields. In 2015, ELIXIR was awarded €19 million from the European Horizon 2020 funding to accelerate the implementation of Europe's life science data infrastructure over the next four years. Started in September 2015, ELIXIR-EXCELERATE will facilitate the integration of Europe's bioinformatics resources, supporting all sectors of life science research and development. It will deliver excellence to ELIXIR's users by fast-tracking the development and deployment of essential data services.

SIB acts as the Swiss node of ELIXIR and is the biggest national node within the consortium. As a provider of various renowned bioinformatics resources, expertise and teaching activities to the international life science community, SIB plays an important role in ELIXIR. Torsten Schwede, leader of the SIB Computational Structural Biology Group and member of the SIB Board of Directors, has been Chair of the ELIXIR Board since 1 January 2015.

SIB organization. SIB consists of bioinformatics research and service groups from the major Swiss schools of higher education and research institutions.

Governance

SIB is governed by five bodies indicated in red in the organigram below.



The Foundation Council

It is the highest authority of the Institute, with supervisory powers. Its responsibilities include changes to the SIB statutes, the nomination of Group Leaders, and the approval of the annual budget and financial report. The SIB partner institutions are represented in this Council.

President

Prof. Felix Gutzwiller
Former Senator

Founding Members

Prof. Ron Appel
SIB Executive Director
and University of Geneva

Prof. Amos Bairoch
Group Leader, SIB
and University of Geneva

Dr Philipp Bucher
Group Leader, SIB and EPFL

Prof. Denis Hochstrasser
Vice-Rector, University of Geneva,
Head of Genetic and Laboratory
Medicine Department, Geneva
University Hospital (HUG)

Prof. C. Victor Jongeneel
Director, Bioinformatics
and Biomedical Informatics,
University of Illinois
at Urbana-Champaign

Prof. Manuel Peitsch
Chairman, SIB Board
of Directors and Vice-President
for Biological Systems Research
at Philip Morris International (PMI)

Ex officio Members

Prof. Karl Aberer
Vice-President
for Information Systems,
EPFL

Dr Claire Baribaud
Director,
Geneva School of Business
Administration (HEG), HES-SO

Prof. Henri Bounameaux
Dean, Faculty of Medicine,
University of Geneva

Prof. Edwin Constable
Vice-Rector of Research
and Talent Promotion,
University of Basel

Prof. Carlo Catapano
Director, IOR Institute
of Oncology Research

Prof. Nicolas Fasel
Vice-Dean, Faculty of Biology
and Medicine, University
of Lausanne

Mr Marc Filliettaz
General Manager,
GeneBio

Prof. Susan Gasser
Director,
Friedrich Miescher Institute
for Biomedical Research (FMI)

Prof. Detlef Günther
Vice President Research and
Corporate Relations, Swiss Federal
Institute of Technology Zurich

Ms Catherine Hirsch
Director, School of Business
and Engineering Vaud (HEIG-VD),
HES-SO

Prof. Denis Hochstrasser
Vice-Rector, University of Geneva,
Head of Genetic and Laboratory
Medicine Department, Geneva
University Hospital (HUG)

Prof. Christophe Hock
Vice-President for Medicine
and Science, University of Zurich

Prof. Rolf Ingold
Vice-Rector, University of Fribourg

Dr Caroline Kant
Executive Director, EspeRare
Foundation Switzerland

Prof. Jérôme Lacour
Dean, Faculty of Science, University
of Geneva

Prof. Piero Martinoli
President, Università della Svizzera
Italiana

Prof. Philippe Moreillon
Vice-Rector,
University of Lausanne

Prof. Jean-Marc Piveteau
President, Zurich University
of Applied Sciences (ZHAW)

Prof. Alexandre Reymond
CIG Director,
Faculty of Biology
and Medicine,
University of Lausanne

Dr Paul Steffen
Head of the Institute
for Sustainability Sciences
and Head of Corporate Research,
Agroscope

Prof. Marcel Tanner
former Director, Swiss Tropical
and Public Health Institute
(Swiss TPH)

Prof. Martin Täuber
Rector, University of Bern

Mr Richard Walker
Chief Financial Officer
and Secretary to the Board,
Ludwig Institute for Cancer
Research (LICR)

Co-opted Members

Prof. Manolo Gouy
CNRS Research Director,
Laboratory of Biometry
and Evolutionary Biology,
Claude Bernard-Lyon 1
University, France

The Board of Directors (BoD)

It takes all the decisions necessary to achieve the aims of the Institute e.g. defines the scientific strategy and internal procedures, and allocates federal funds to service and infrastructure activities. It consists of two Group Leaders elected jointly by the Council of Group Leaders and the BoD; two external members elected by the Foundation Council on the recommendation of the BoD; and the Executive Director. The BoD members are appointed for a renewable five-year period.

Prof. Manuel Peitsch
(Chairman)
Vice President for Biological
Systems Research
at Philip Morris International (PMI)

Ms Martine Brunschwig Graf
Former National Councillor

Prof. Ron Appel
SIB Executive Director
and University of Geneva

Prof. Christian von Mering
SIB Group Leader and University
of Zurich

Prof. Torsten Schwede
SIB Group Leader and University
of Basel

The Scientific Advisory Board

It acts as a consultative body providing recommendations to the BoD and the Council of Group Leaders. Its main tasks consist in monitoring the service and infrastructure activities, as well as the core bioinformatics resources. It is made up of at least five members who must be internationally renowned scientists from the Institute's fields of activities.

Prof. Alfonso Valencia
(Chairman)
Director of the Structural Biology
and Biocomputing Programme,
Spanish National Cancer
Research Centre, Madrid, Spain

Prof. Manolo Gouy
CNRS Research Director,
Laboratory of Biometry
and Evolutionary Biology,
Claude Bernard-Lyon 1
University, France

Dr David de Graaf
President and CEO
of Selventa,
Cambridge, MA, USA

Prof. Alexey I. Nesvizhskii
Department of Pathology
and Department of Computational
Medicine & Bioinformatics,
University of Michigan,
Ann Arbor, USA

Prof. Christine Orengo
Department of Structural
and Molecular Biology,
University College London,
United Kingdom

Prof. Ron Shamir
Computational Genomics
Group at the Blavatnik
School of Computer Science,
Tel Aviv University, Israel

Prof. Anna Tramontano
Computational Biology
Laboratory, La Sapienza
University, Rome, Italy

The Council of Group Leaders

It discusses all matters relating to the SIB groups as a whole, and proposes the nomination of new Group Leaders. It consists of the Group Leaders, the Affiliate Group Leaders and the Executive Director.

Group Leaders: The SIB Group Leaders are staff members at SIB partner institutions (see p. 37). In addition, OsiriX led by Prof. Osman Ratib is an Affiliate Group.

Honorary Members

Dr Johannes R. Randegger
Former National Councillor,
Honorary President
of the SIB Foundation Council

Mr Peter Malama (1960-2012)
Former National Councillor,
Honorary President
of the SIB Foundation Council

Prof. Ernest Feytmans
Honorary Director

Ms Christiane Langenberger (1941-2015)
Former Senator, Honorary President
of the SIB Foundation Council.

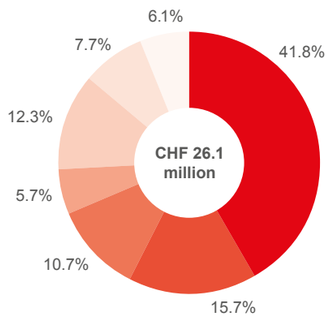
Finance

In a context of challenging times for research funding worldwide, SIB funds remained stable in 2015, thanks to the continued support of its funders.

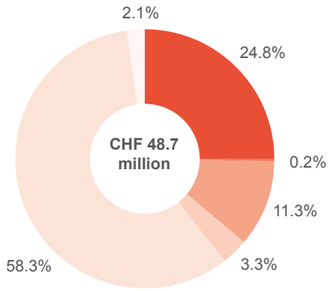
Sources of income

In 2015, the total income of the SIB groups reached CHF 74.8 million, of which CHF 26.1 million were managed by SIB, and CHF 48.7 million were managed by the groups' respective institutions, as indicated in the table below. The largest source of SIB funds is the Swiss government (CHF 11 million, 43.8%), followed by SNSF/European Funds and industry (CHF ~3.4 million each, 13.5%).

	Funds managed by SIB		Funds managed by the partner institutions		Total number of grants/ contracts
	CHF million	%	CHF million	%	
Swiss government (article 16)	10.9	41.8			
SNSF and European Funds	4.1	15.7	12.1	24.8	108
NIH	2.8	10.7	0.1	0.2	3
SystemsX.ch	1.5	5.7	5.5	11.3	58
Industry	3.2	12.3	1.6	3.3	25
Universities and hospitals	2.0	7.7	28.4	58.3	60
Other	1.6	6.1	1.0	2.1	12
	26.1		**48.7**		



Funds managed by SIB

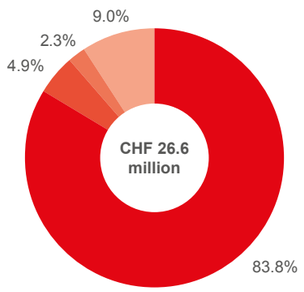


Funds managed by the partner institutions

Spending

Of the CHF 26.6 million spent by SIB in 2015, 84% were dedicated to salaries and the rest to equipment, scientific events and running costs.

	SIB spending	
	CHF million	%
Salaries	22.3	83.8
Equipment	1.3	4.9
Scientific events	0.6	2.3
Running costs	2.4	9.0
	26.6	



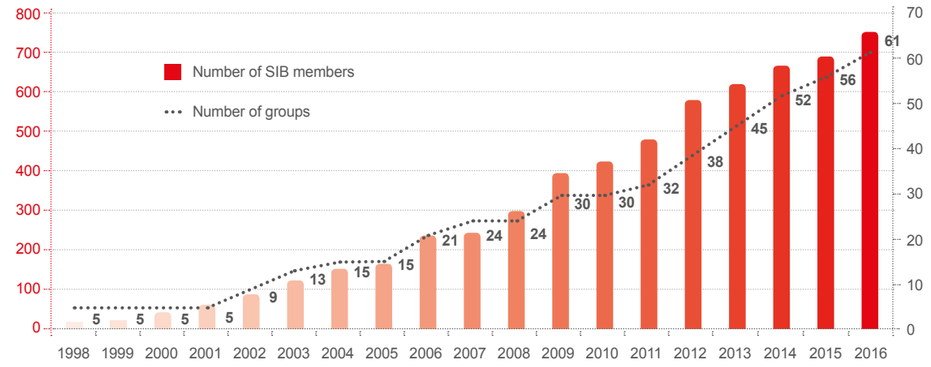
SIB spending

Membership & staff

SIB continues to grow in terms of membership and staff.

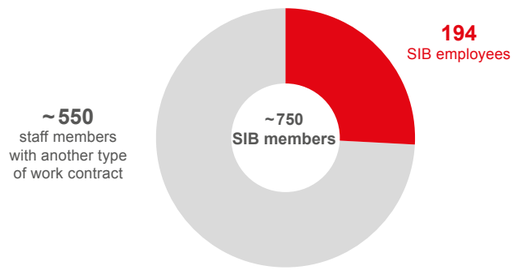
Evolution of SIB membership

Over the past 6 years, the number of SIB members has almost doubled as well as the number of groups.



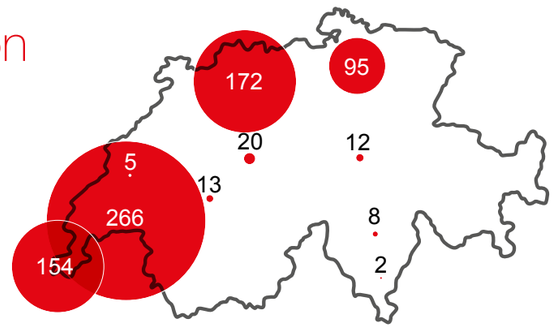
SIB members

As of 1 March 2016, SIB included approximately 750 members, of whom 194 are SIB employees (i.e. with an SIB work contract) and ~550 have another type of work contract (e.g. from universities).



Geographical representation of SIB members

About one third of SIB members is located in Lausanne, followed by Basel (23%), Geneva (20%) and Zurich (13%) (figures for March 2016).



In the spotlight 2015

SIB becomes FAO Reference Centre for Bioinformatics

Food and Agriculture Organization of the United Nations

The Food and Agriculture Organization of the United Nations (FAO) has appointed SIB as FAO Reference Centre for Bioinformatics, recognizing its expertise together with its state-of-the-art scientific services. In particular, SIB is collaborating with FAO on the screening, monitoring and follow-up of zoonotic diseases by providing open-access databases such as Viralzone (the virus knowledgebase maintained by the SIB Swiss-Prot Group), OpenFlu and OpenFMD (resources on influenza and foot-and-mouth disease, respectively, which are maintained by the SIB Vital-IT Group). These databases provide information on the pathogens' genomes, their epidemiology, evolution and parenthood and contribute to the fight against dangerous viral infections, including avian influenza and foot-and-mouth disease, in farm animals and wildlife.

© Wikicommons

SIB, key player at the European level

ELIXIR has been awarded €19 million by the EU to accelerate the implementation of Europe's life science data infrastructure over the next four years. SIB is actively involved in the ELIXIR-EXCELERATE project, which aims at facilitating the integration of Europe's bioinformatics resources, supporting all sectors of life science R&D. In particular, SIB focuses on the promotion of excellence in resource development and operation, the sustainability of core resources, as well as on training aspects. SIB hosted the ELIXIR Board meeting in Geneva on 13 and 14 April 2015. The ELIXIR Board meeting was attended by representatives from ELIXIR's then 11 member countries and the European Molecular Biology Laboratory (EMBL), four of ELIXIR's observer countries and guests from several other countries working towards membership - a total of 23 countries were represented at the meeting.

SIB continues its growth

Five new groups have joined the Institute as of 1 January 2016 (see pp. 38-39):

- Prof. Karsten Borgwardt, Head of Machine Learning & Computational Biology Lab, ETHZ
- Prof. Christophe Dessimoz, Head of Laboratory of Computational Evolutionary Biology and Genomics, University of Lausanne
- Prof. Carlos-Andrés Peña-Reyes, Head of Computational Biomed Research and Applications Group, Haute Ecole d'Ingénierie et Gestion de Vaud (HEIG-VD) – HES-SO
- Dr Hubert Rehrauer, Group Leader of Genome Informatics, Functional Genomics Center Zurich, ETHZ Zurich
- Dr Daniel Stekhoven, Head of Clinical Bioinformatics Unit (CBU), NEXUS Personalized Health Technologies, ETHZ Zurich

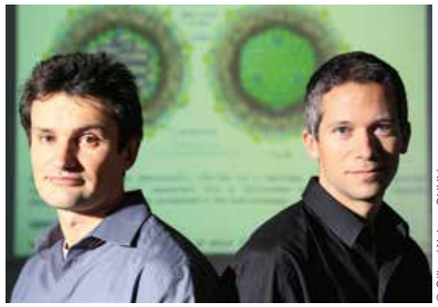
Three new institutional members have joined SIB:

- EspeRare Foundation, a non-profit organization specialized in drug development for rare diseases
- Geneva University Hospitals (HUG): SIB is collaborating with HUG for the development of a clinical-grade diagnostic pipeline based on NGS analysis
- Haute Ecole d'Ingénierie et Gestion de Vaud (HEIG-VD) – HES-SO

SIB extends a very warm welcome to all new members.

Leenaards Prize 2015 won by two SIB Group Leaders

On 26 March 2015, SIB Group Leaders E. Zdobnov (UNIGE) and J. Fellay (EPFL-CHUV) were among the Leenaards Prize laureates for translational medical research with the project "Genome against genome: the impact of human genetic variations on chronic viral infections". The study explores how genetic variations in humans can impact chronic viral infections and allows the design of novel diagnostic and therapeutic strategies for better patient management.



© Gilles Weber, CHUV

Launch of the new SIB website

SIB introduced its new website with a modern look, novel features, a more intuitive navigation, and specific sections dedicated to various audiences.
Please visit our new website: www.sib.swiss

Selected highlights 2015

Drug Design workshop at the Musée de la Main

During the “Rencontres avec les chercheurs” at the LabLife exhibition, 150 visitors attended SIB’s workshop about bioinformatics and drug design. Visitors could try their hand at modifying molecules *in silico* to improve the melanoma treatment properties.

Maths to serve breast cancer treatment

A collaborative work between SIB, the Ludwig Institute and CHUV combined clinical and experimental oncology with mathematical modelization to reprogram a flawed immune response into an anti-tumoural one, which could be used to treat breast cancer.

© Ernesto del Aguila, NHGRI

GTEx, linking gene activity and disease

Researchers co-led by E. Dermitzakis, Group Leader at SIB, have created a new data resource called Genotype-Tissue Expression (GTEx) to reveal how differences in an individual’s genetic make-up can affect gene activity and contribute to disease.

A great success at the Campus Biotech Open Day

SIB’s booth was very successful and crowded with attendees from 10 a.m. to 4 p.m. non-stop! During the day, over 1,000 visitors were introduced to bioinformatics.

Ariana® Pharma and SIB partnership

SIB officially announced the collaboration of F. Lisacek’s SIB Group with Ariana® Pharma for the discovery of novel biomarkers for gastric cancer. This partnership aims at developing early detection tools for this aggressive disease.

JANUARY | FEBRUARY | MARCH | APRIL | MAY | JUNE

The largest genome-wide study related to obesity

Two SIB Group Leaders, Z. Kutalik and J. Beckmann, contributed to the largest ever genome-wide study strengthening the genetic link to obesity. The study, tripling the number of known genetic locations related to obesity, has been published in *Nature*.

An SIB publication on the cover of *The Economist*

A collaborative work, involving R. Waterhouse from the SIB Computational Evolutionary Genomics Group, linked the mosquito’s sexual behaviour to its ability to transmit malaria. These findings were published in *Science* and covered in *The Economist*.

Photo: © James Gathany, Scientific Photographer, CDC.

How chameleons change colour finally unveiled

SIB Group Leader M. Milinkovitch and his team have discovered how chameleons accomplish their vivid colour changes and keep cool. The study, published in *Nature Communications*, created a bit of a global buzz.

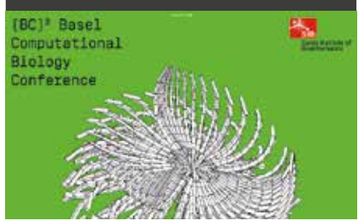
© Michel Milinkovitch

SIB associates genetic variants with cognitive impairment

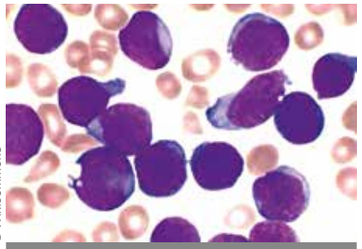
Z. Kutalik and J. Beckmann from SIB, together with a UNIL Group, published a study highlighting the negative association between individually rare but collectively common intermediate-size copy number variations and educational attainment.

12th [BC]² - Basel Computational Biology Conference and ELIXIR/SIB Innovation & SME Forum

The Forum organized by SIB and ELIXIR alongside the successful [BC]² conference in Basel under the headline “Data-driven innovation in the pharma and biotech industries” was a good opportunity for exchanges with industry.



JULY



© Wikicommons

New cure for a paediatric leukaemia

Researchers from the University Children's Hospital in Zurich together with M. Delorenzi's SIB Group have decoded the genome and transcriptome of a lethal sub-type of acute lymphoblastic leukaemia.

AUGUST

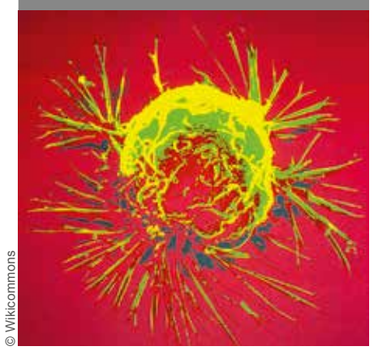
© Bernard Martinez

Recording of OPUS 23 - Music for a Gene

OPUS 23 - Music for a Gene is a string quartet by French composer O. Calmel and commissioned by SIB. The project sponsored by the Cogito Foundation and directed by SIB Group Leader L. Lane transposes the complexity of the human genome into musical emotions. The piece was recorded in 2015 thanks to a successful crowdfunding campaign.

New insights into breast cancer heterogeneity

The group of M. Stadler at SIB together with another FMI Group discovered how the most common breast cancer mutation can induce multipotency and tumour heterogeneity. These results have been published in *Nature*.



© Wikicommons

OCTOBER

SIB booth at Recherche Live – a great success

During the itinerant exhibition Recherche Live that was passing through Geneva for the Swiss Academy of Sciences' 200th anniversary, SIB's Drug Design booth met with huge success, welcoming approximately 400 students.

Molecular subtypes of bowel cancer uncovered

The SIB group of M. Delorenzi participated in a study consortium, which showed that bowel cancer can be classified into four distinct diseases associated with different diagnoses and specific treatments. This work has been published in *Nature Medicine*.

NOVEMBER

SIB Metagenomic Pizza at Expanding Your Horizons

During the 4th bi-annual event organized for young girls by Expanding Your Horizons at UNIGE, SIB presented its Metagenomic Pizza workshop about the human genome and various bioinformatics resources to an audience of 40 participants.

UNIGE Faculty Diabetes Centre Open House

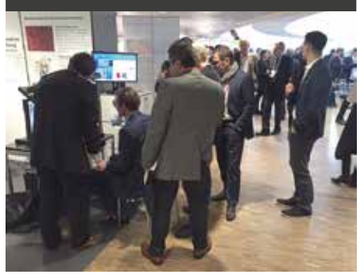
SIB held a workshop at the 13th annual "Journée Portes Ouvertes du Diabète et Obésité" of the UNIGE Medical Centre and newly formed Faculty Diabetes Centre. Visitors could discover how bioinformatics helps find genetic variations associated with diabetes.

SIB initiates young Israeli students to bioinformatics

Outstanding high school students from Israel visited SIB at the Campus Biotech Geneva. They were taught the basics of bioinformatics prior to their attendance at a Drug Design workshop which turned out to be a great success.

SIB presence at the Swiss Innovation Forum

SIB took part in the Swiss Innovation Forum and its exhibition "Future Expo" with a booth presenting how SIB's computational solutions contribute to innovation in medicine and drug development.



DECEMBER

Workshops for high school teachers

SIB organized a two-day workshop to introduce bioinformatics and drug design to high school teachers from the American School in Switzerland.

New publication about SIB's curated resources in NAR

An overview of SIB's resources and competence areas, with a strong focus on curated databases and SIB's most popular and widely used resources, was published in the *Nucleic Acids Research (NAR)* Database Issue.

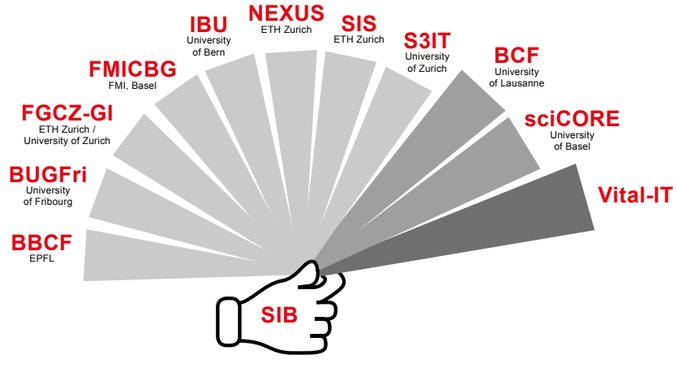
Services

SIB is instrumental to good science.

SIB provides world-class expertise and core bioinformatics resources to the national and international life science and medical community in academia and industry.

SIB provides the necessary bioinformatics services and resources for researchers thanks to:

- Personal expertise, key competencies and on-demand support in bioinformatics, biocuration, high-throughput data analysis, and statistics
- Over 150 internationally recognized and extensively used bioinformatics databases and software tools, which SIB continuously develops and maintains
- Eleven core facilities and high-performance computing (HPC) centres that provide state-of-the-art bioinformatics and statistical support
- Approximately 30 bioinformaticians embedded in labs at Swiss Universities and University hospitals who benefit from the SIB expert network to provide on-site customized support to researchers and clinical labs
- The SIB Legal and Technology Transfer Office (LTTO), whose mission is to showcase, manage and transfer SIB knowledge and know-how so that the scientific community can benefit from SIB's many innovations
- The SIB Technology Group, which is in charge of optimizing the coordination of all SIB technology-related activities and advising on technology-related matters.



Through eleven bioinformatics core facilities and HPC centres, SIB provides bioinformatics and statistical support as well as services, expertise and infrastructure to life scientists.

Databases & platforms

SIB develops, supplies and maintains more than 150 high-quality databases and software platforms for the global life science community.

Most of SIB resources are in open-access on the SIB bioinformatics resource portal ExPASy (www.expasy.org). Created in 1993, ExPASy was at that time the first website available in the biomedical field. The SIB resources cover different areas of life sciences, such as genomics, proteomics and evolution.

ExPASy Bioinformatics Resource Portal

CATEGORIES	SUB-CATEGORIES	EXAMPLES OF DATABASES	EXAMPLES OF SOFTWARE TOOLS
Genes and genomes	Sequence alignment		Codon Suite, LALIGN, Newick Utilities, T-Coffee
	Similarity search		LALIGN, Phylogibbs
	Characterization/annotation	CLIPZ, EPD, miROrtho, OMA, OpenFlu, OrthoDB, smirnaDB, SwissRegulon	CLIPZ, ChIP-Seq, EPD, ISA, OMA, smirnaDB
	Transcriptomics	Bgee, CleanEx, CLIPZ, smirnaDB, SwissRegulon	CLIPZ, ISMARA, MirZ, PPA, smirnaDB, TopAnat
Proteins and proteomes	Protein sequences and identification	neXtProt, UniProtKB, UniProtKB/Swiss-Prot, ViralZone	HAMAP, PeptideMass, Translate
	Mass spectrometry and 2-DE data	SWISS-2D PAGE, WORLD-2D PAGE Repository	FindPept, GlycoMod, MSight
	Protein characterization and function	neXtProt, UniProtKB, UniProtKB/Swiss-Prot	AACompSim, Biochemical Pathways, ProtScale
	Families, patterns and profiles	MyHits, PROSITE	MyDomains, MyHits, pftools, PRATT
	Post-translational modification	UniCarbKB, UniCarb-DB, SugarBind, UniProtKB/Swiss-Prot	FindMod, GlycanMass, ISMARA, UniCarbKB
	Protein-protein interaction	STRING, UniProtKB/Swiss-Prot	PredictProtein, ProtBud
	Similarity search/alignment	MyHits, UniProtKB	BLAST, ClustalW, MyHits
	Imaging		ImageMaster / Melanie, MSIght
Medicine and health		SwissSidechain	SwissDock, SwissParam, SwissBioisostere, SwissTargetPrediction
Evolution and phylogeny		Bgee, ImmunoDB, miROrtho, OMA, OrthoDB	Arlequin, CT-CBN, Newick utilities, OMA, TriFLe
Structural biology		SWISS-MODEL Repository, SwissSideChain	SwissDock, SWISS-MODEL Workspace, Swiss-PdbViewer
Systems biology		Progenetix, SwissRegulon	arrayMap, MetaNetX, The Systems Biology Research Tools
Bioinformatics infrastructure			nfswatch, Soaplab services, SPARQL-playground

This table shows, for each bioinformatics domain, examples of SIB databases and software programs that are available on ExPASy.

The list below presents SIB's core resources with a brief description:

UniProtKB/Swiss-Prot

Protein knowledgebase

UniProtKB/Swiss-Prot is the most widely used protein information resource in the world, with over 800,000 requests per month. It provides concise, but thorough, descriptions of a non-redundant set of hundreds of thousands of proteins including their function, domain structure, post-translational modifications and variants. Its high-quality annotation is the fruit of manual expert curation by biologists, who use information available in the scientific literature to provide an accurate description of each protein's features (see also p. 52).

neXtProt

Human protein knowledge platform

neXtProt is an innovative knowledge platform dedicated to human proteins. neXtProt includes curated information on various aspects of human protein biology such as function, mRNA/protein expression, protein/protein interactions, post-translational modifications and protein variations. Its aim is to help life science researchers in their quest to unravel the complexity of human life processes (see also p. 50).

SWISS-MODEL

Structure homology-modelling

SWISS-MODEL is an automated protein structure homology-modelling server for generating 3D models of a protein. Comparative approaches are currently the most accurate and reliable computational methods to derive 3D models of proteins, for which experimental structures are not available. SWISS-MODEL automates the complex process of model building on an easy to use web-based system, thereby making model information also available for non-specialists (see also p. 67).

STRING

Protein-protein interactions

STRING is a database of known and predicted protein-protein interactions, including direct (physical) and indirect (functional) associations. They are derived from different sources such as the genomic context, high-throughput experiments, (conserved) co-expression, and the literature. STRING quantitatively integrates the interaction data for a large number of organisms, and transfers information between these organisms where applicable. The database currently covers 9,643,763 proteins from 2,031 organisms (see also p. 51).

SwissRegulon

Annotations of regulatory sites

The SwissRegulon web portal provides information and tools for the analysis of genome-wide transcription regulatory networks in organisms ranging from *E. coli* to human. The database frontend offers an intuitive interface showing genomic information in a clear and comprehensible graphical form (see also p. 46).

SwissDrugDesign

Drug design

The SwissDrugDesign project is an ambitious initiative aiming to provide the first comprehensive, integrated and freely accessible web-based *in silico* drug design environment to the scientific community worldwide. It offers a large collection of tools covering all aspects of computer-aided drug design, from target prediction of small molecules (SwissTargetPrediction) to the provision of topology and parameters of drug-like molecules (SwissParam). Other tools include SwissDock, SwissSideChain and SwissBioisostere (see also p. 67).

EPD

Collection of eukaryotic promoters

Genes are first "transcribed" into messenger RNA, which is used as template to synthesize proteins. Transcription begins on "promoters". The Eukaryotic Promoter Database (EPD) provides quality-controlled information on experimentally defined promoters of higher organisms as well as web-based tools for promoter analysis (see also p. 44).

Bgee

Gene expression evolution

Bgee is a database of gene expression evolution, which integrates all types of transcriptome and expression information for animals – including human, model organisms such as mouse or *Drosophila*, and diverse species of evolutionary or agronomical relevance. Bgee is the only resource, which provides homologous expression between species (see also p. 62).

UniCarbKB

Glycan knowledgebase

UniCarbKB is a knowledgebase that offers public access to a curated database of information on glycoproteins, which are proteins to which carbohydrates (or glycans) are attached. UniCarbKB provides comprehensive information on glycan structures, and published glycoprotein information including global and site-specific attachment information (see also p. 51).

SugarBind

Pathogen sugar-binding

Host-pathogen communication is known to be mediated by carbohydrate-protein interactions, which ensure adhesion at the cell surface. The SugarBind database provides information on known mammalian carbohydrate sequences to which pathogenic organisms (bacteria, toxins and viruses) specifically adhere. Information in SugarBindDB is manually curated and supports the investigation of bacterial and viral infections (see also p. 51).

OrthoDB

Hierarchical catalogue of orthologs

OrthoDB is a catalogue of "equivalent" genes among species, called orthologs. Resolving gene ancestry is the most accurate way to predict putative gene functions by association with genes studied in model organisms. OrthoDB is hence critical both for evolutionary studies and for interpreting gene content from the newly sequenced genomes (see also p. 48).

OMA

Orthology prediction

The Orthology MAtrix (OMA) Browser provides orthology predictions among publicly available genomes. Started in 2004, it has undergone many releases and now elucidates orthology among millions of genes from thousands of species, making it one of the largest resources of its kind (see also p. 60).

Core facilities & HPC

Through eleven core facilities and high-performance computing (HPC) centres, as well as embedded bioinformaticians, SIB groups provide expert data analysis services and computing power to life scientists in academia and industry, thus enabling them to perform world-class biomedical research.

Vital-IT

Vital-IT
University of Lausanne
Ioannis Xenarios

The Vital-IT multidisciplinary team of scientists and technical staff maintains a competency centre in bioinformatics and computational biology. Vital-IT's infrastructure currently spreads across six institutions that maintain biotechnological platforms: SIB, the universities of Geneva, Lausanne, Fribourg and Bern as well as EPFL. The core facility enables scientists to access state-of-the-art computational infrastructures (processing, storage and archiving) as well as expertise in data analysis and algorithmic development. Vital-IT partners with scientists to build computational solutions facilitating their research or to transform their ideas into production-quality software. It supports postgraduate education through training and workshops in coordination with SIB and institutional partners.

More than
8,000,000
CPU hours consumed per year

18
expert embedded bioinformaticians providing on-site data analysis services to life scientists and clinicians

More than
20,000,000
jobs run by users per year

Author of, or acknowledged in:
87
publications in 2015

22
in *Science* or *Nature* over the last 5 years

730
active users of Vital-IT in 2015

6
Petabytes of storage capacity

+

7,000
CPUs in 2015

sciCORE

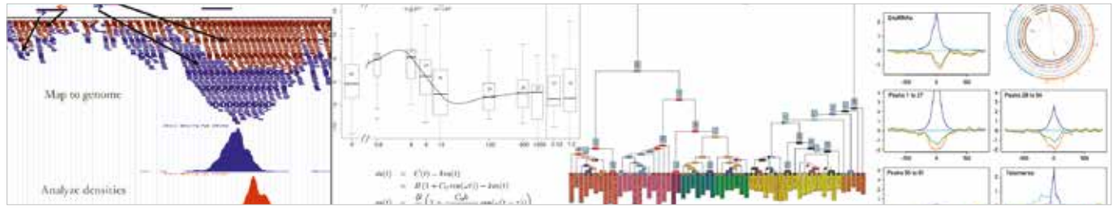
sciCORE
University of Basel
Torsten Schwede

sciCORE is a centre of competence in scientific computing – providing high-performance computing infrastructure, large-scale storage resources, scientific software and databases, server infrastructures and user support, as well as know-how and expertise to scientific research groups. sciCORE provides a professional environment for scientific applications: from bioinformatics, computational chemistry, physics, systems biology, to medicine and economics. In direct collaboration with scientific research groups, the centre of competence helps, develops, deploys, operates and extends the computational tools required for performing modern life science and biomedical research. It also operates the IT infrastructure for several SIB services, e.g. SWISS-MODEL and SwissRegulon.

BCF

Bioinformatics
Core Facility (BCF)
University of Lausanne
Mauro Delorenzi

The Bioinformatics Core Facility (BCF) is a centre of excellence that provides state-of-the-art know-how for data analysis for the life science community and supports biomedical groups in applying high-throughput laboratory approaches. It offers consulting, teaching and training, data analysis support and collaborations.



BBCF

Bioinformatics
and Biostatistics
Core Facility (BBCF)
EPFL, Lausanne
Jacques Rougemont

The EPFL Bioinformatics and Biostatistics Core Facility (BBCF) provides research labs with extensive support in bioinformatics and biostatistics. BBCF main competences are in management and analysis of genomic data, mathematical modelling and statistical analysis of quantitative biological data. BBCF provides support for the analysis of large or complex data sets, the development of data management pipelines for new high-throughput technologies (e.g. high-density arrays, high-throughput sequencing), and the statistical planning in complex experimental designs. The core facility also helps researchers in the areas of mining public data, designing and setting up local databases, building mathematical models from experimental data and running simulations to evaluate a model.

BUGFri

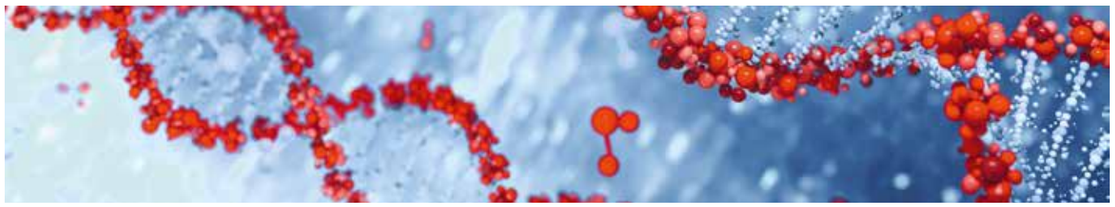
Bioinformatics Core Facility
of Fribourg (BUGFri)
University of Fribourg
Laurent Falquet

The Bioinformatics Core Facility of the University of Fribourg (BUGFri) supports life science researchers by providing expertise in data analysis of next generation sequencing (NGS) experiments, or any large-scale biological experiment requiring bioinformatics resources. BUGFri focuses on genome assembly, annotation and comparison as well as on mutant and structure variant identification by resequencing. The core facility also performs metagenomics, RNAseq and ChIPseq data analysis, proteome clustering and ortholog/ paralog classification, as well as pathway and gene set enrichment analysis.

FMICBG

FMI Computational Biology
Group (FMICBG)
Friedrich Miescher
Institute, Basel
Michael Stadler

The FMI Computational Biology Group (FMICBG) helps biologists of FMI in data analysis and visualization through collaborations, a bioinformatics helpdesk and relevant training depending on the needs. The core facility focuses on providing solutions based on free and open-source software, allowing the scientists to continue their own bioinformatics research even after leaving the FMI.



IBU

Interfaculty Bioinformatics
Unit (IBU)
University of Bern
Rémy Bruggmann

The Interfaculty Bioinformatics Unit of the University of Bern (IBU) provides services and expertise to assist researchers of the three “Life Science” Faculties (i.e. Sciences, Medicine, and VetSuisse) in data analysis and project planning for large-scale experiments (e.g. next generation sequencing, genome assembly). Furthermore, IBU has its own research programme and collaborates on large and complex projects. IBU develops methods to analyse high-throughput data. The core facility also has a high-performance compute cluster and a data storage system that are used for IBU’s own research, collaborations and service projects.

SIS

Scientific IT Services (SIS)
ETH Zurich
Bernd Rinn

The Scientific IT Services (SIS) is an interdisciplinary bioinformatics and scientific IT support group that builds up computational tools, ranging from lab databases to reusable framework components, enabling and supporting data management and analysis in life science research and beyond. SIS collaborates with Swiss and European research groups and industry in the life science sector such as SystemsX.ch, SyBIT, FAIR-DOM, HPC-CH and Swiss universities eSCT. SIS provides our partners with data management solutions and services, integrates and operates data analysis pipelines, and provides training and consulting in databases, scientific software development, high-performance and cloud computing.

S3IT

Service and Support
for Science IT (S3IT)
University of Zurich
Peter Kunszt
a.i. Marcel Riedi
(as of 2016)

The Service and Support for Science IT (S3IT) unit provides support for science in general, and life sciences and medicine in particular. S3IT serves as a partner for projects locally and nationally to enable competitive research with the advanced use of computational methods and resources. The S3IT team advises groups and projects about data management, data analysis and cooperates to optimise their specific workflow. S3IT also takes part in national projects and cooperates with similar technology-oriented groups to ensure that its expertise is always up-to-date.

New core facilities

In 2015, the SIB welcomed two new core facilities:

FGCZ-GI

The Functional Genomics Center Zurich Genome Informatics (FGCZ-GI) ETH Zurich / University of Zurich
Hubert Rehrauer

The Functional Genomics Center Zurich Genome Informatics (FGCZ-GI) of the ETH Zurich and the University of Zurich is a centre of competence for bioinformatics, especially in the area of microarray and next generation sequencing. The core facility brings the data generated by the latest “omics” technologies to life, by offering bioinformatics collaborations, services and education for single molecule, single cell as well as high-throughput data. The group covers entire analysis workflows from the processing of massive raw data up to statistical evaluation and interpretation.

NEXUS

NEXUS Personalized Health Technologies
Clinical bioinformatics Unit (CBU)
ETH Zurich
Daniel Stekhoven

NEXUS Personalized Health Technologies is an ETH Technology Platform created to enable and accelerate the execution of translational research projects by providing key technological resources, tools and collaboration opportunities for the personalized health research community. NEXUS encompasses two units - Clinical Bioinformatics (CBU) and Theragnostics Discovery (TDU). The TDU provides state-of-the-art robotic screening technologies and assists the research community in pursuit of innovative projects related to molecular circuit analyses, chemical hit identification and lead development. The CBU, now part of the SIB core facilities, provides computational expertise to process large-scale heterogeneous data and offers support in “omics” and statistical data analyses. The CBU collaborates with the research community to develop solutions for the integration of “omics” data and clinical information. This includes the development of decision support systems for oncologists, which are based on the analysis of genome sequencing data coming from patient tumor samples.

Embedded bioinformaticians

Bioinformatics skills are indispensable in today’s life science projects. SIB supports the Swiss universities and university hospitals not only through its bioinformatics resources and expertise, but also by embedding bioinformaticians in the various research and clinical labs.

There are more and more bioinformaticians physically collocated with scientists in wet labs of research institutes or sequencing departments in the hospitals. These bioinformaticians are called “embedded bioinformaticians”. Their presence in research and clinical groups is an advantage, as they can provide direct guidance on how to design experiments, how to manage and analyse data, and on the optimized use of the various bioinformatics tools.

Similarly, the physical collocation of clinicians and bioinformaticians represents a benefit for both disciplines. With the emergence of personalized medicine, this close collaboration allows the development of clinical bioinformatics tools especially designed and optimized for clinical research, patient data analysis, diagnosis, and precision medicine.

The 26 embedded bioinformaticians can take advantage of the SIB expert network, which they can consult to benefit from state-of-the-art expertise and support in the field.

Personalized health

SIB is active at two levels: clinical bioinformatics and the SPHN initiative

Medical practice is undergoing a revolution around personalized health. The clinical use of the gigantic quantities of data produced by high-throughput technologies poses novel technical, analytical, ethical and educational challenges to both clinicians and researchers. SIB plays a leading role at two levels in the national initiatives currently being set up, namely: the development of clinical bioinformatics to support hospitals and clinicians, and the data infrastructure for the Swiss Personalized Health Network (SPHN) initiative to share patient data for research.

In the near future, health practitioners will have access to a wide variety of data for a given patient, such as their genomic sequence, metabolic profile and lifestyle. Consequently, a patient's medical file will gradually become not only more and more detailed but also more personal. It will provide a basis on which practitioners will rely to diagnose a disease, adapt and fine-tune a patient's treatment and track down illnesses for better prevention. This so-called "personalized health" involves a multidisciplinary approach in which patients, practitioners and researchers work together to understand health at an individual level. In this context, several national initiatives have been launched in Switzerland to bridge the gap between medical practices, research and big data to convert them into clinically relevant knowledge. Bioinformatics and SIB have a crucial role to play in such translational projects, at **two different but complementary levels: clinical bioinformatics and data infrastructure for research**.

Clinical bioinformatics: SIB's support to clinicians

To respond to the needs of personalized medicine, SIB created in 2013 a Clinical Bioinformatics Group, led by Jacques Beckmann, to set up bridges between the world of bioinformatics and the medical community, for the benefit of patients (see figure on facing page). The missions of the Clinical Bioinformatics Group are:

- To identify sectors in which huge quantities of data are likely to enter the clinical arena
- To develop clinical bioinformatics tools specialized in storing, organizing, analysing and interpreting molecular data which are linked to an individual's health
- To offer assistance to clinicians in exploiting big data and adapting their methods, as well as in learning how to use bioinformatics tools, manage databanks, and interpret multiple types of data
- To collaborate with individual hospitals in order to establish bioinformatics pipelines to be used within their respective firewalls and on their own patient data.

Highlights 2015
During the course of 2015 and with the help of numerous health practitioners, SIB's Clinical Bioinformatics Group reached a national consensus on the diagnostic needs and expectations of hospitals in the multi "omics" areas. Strong interactive relationships were developed with Swiss hospitals with an initial focus on oncology and hemato-oncology. A workshop was organized for clinicians, bioinformaticians and researchers to review practices in clinical next generation sequencing (NGS), followed by the establishment of a working group dedicated to the topic. A first agreement was also signed with the Geneva University Hospitals (HUG), leading to the initial implementation of specific diagnostic bioinformatics tools at the hospitals.

Patient medical consultation | "Big data" generated by modern technologies, e.g. genetic sequencing | Data management, analysis & interpretation, integration of information from specialized databases | Consolidated report to support diagnosis and treatment decisions | Patient receives diagnosis, treatment and counseling

Clinical bioinformatics pipelines: from patient medical consultation, through "big data" generation, analysis and interpretation, to diagnosis and treatment.

SPHN initiative: SIB to provide infrastructure for research

In order to establish nationwide interoperability of clinical and "omics" data, allowing researchers in Switzerland to share patient data, the Swiss State Secretariat for Education, Research and Innovation (SERI) has launched a national research initiative entitled Swiss Personalized Health Network (SPHN). SPHN, which will officially start in January 2017 if subject to the approval by the Parliament in fall 2016, will integrate the university hospitals, schools of higher education, research institutes, organizations working in the area of personalized health, as well as currently ongoing projects across Switzerland. To achieve this goal, the medical informatics systems of these Swiss organizations and personalized health platforms will have to become mutually compatible ("interoperable").

SIB plays a leading role in the SPHN initiative and will set up two types of infrastructures:

- **Data Coordination Centre:** SIB is in charge of setting up and running the Data Coordination Centre (DCC) within the SPHN that will deal with data interoperability and organization nationwide to ensure that research projects will have access to data from the various Swiss hospitals. SIB is contributing to the establishment of standards for data production and storage.

- **BioMedIT:** In response to a joint call for applications for new research infrastructures of the Swiss National Science Foundation (SNSF) and SERI, SIB submitted the BioMedIT proposal, which aims to establish a coordinated nationwide network of secure infrastructures to support biomedical research in Switzerland. If the funding is accepted at the end of 2016, BioMedIT will be fully integrated into SIB's contribution to the SPHN initiative.

SIB technology

Optimizing technology-related activities

The SIB Technology Group is in charge of coordinating technology-related activities. Headed by Heinz Stockinger, the group works in close cooperation with technology and infrastructure providers, as well as competence centres such as Vital-IT, Swiss-Prot and sciCORE, to combine forces wherever necessary and practical. The group also advises SIB on technology-related matters.

Core competencies

- Design, development, testing and operation of scientific, technical and administrative software in bioinformatics, in cooperation with SIB groups, with a strong focus on web and internet technologies
- Technical coordination of topics that require an SIB-wide approach, i.e. web application deployment, security and related guidelines, code repositories, etc.
- Support and operation of SIB-wide services developed and/or deployed by the group, such as:
  - ExPASy SIB Bioinformatics Resource Portal and some scientific resources available on the portal
  - Requesting tracking operations for user support
  - Web applications for SIB course registration and administration
  - Web applications for SIB personnel management/administration
- Coordination of SIB technical activities within the ELIXIR project and operation

Highlights 2015

The SIB Technology Group was created and collaborates with SIB and its groups to develop and operate scientific services (incl. ExPASy.org), training (with SIB Training) and operations software (mainly web applications). The SIB Technology Group relies on the infrastructure support of the Vital-IT Group.

In 2015, the SIB Technology Group mainly collaborated on the following projects:

- *TopAnat/Bgee:* Tool to identify and visualize enriched anatomical terms from the expression patterns of a list of genes. Co-developed with the group of Marc Robinson-Rechavi.
- *ismara-client:* Software for Mac OS X and Linux to pre-process jobs submitted to the ISMARA server for final execution. Developed in cooperation with the group of Erik van Nimwegen.
- *Biocuration 2016:* Design and development of a website and registration system for the Biocuration 2016 conference. Co-developed with the group of Torsten Schwede.
- *GitLab:* Coordinated the installation and operation of the GitLab service. This is the first time SIB has had a common code version control management system for all SIB members.

Technology transfer

Enabling the community to benefit from SIB's innovations

With expertise that covers a broad spectrum of application fields, SIB occupies a pivotal hub position in bioinformatics innovation in Switzerland. Under the leadership of Marc Filliettaz, the SIB Legal and Technology Transfer Office (LTTO)'s mission is to showcase, manage and transfer SIB knowledge and know-how so that the scientific community can benefit from SIB's many innovations.

Main activities

- **Partnerships with industry:** The LTTO strives to enhance the scientific and industrial visibility of SIB innovation by assisting SIB members in their contacts with external partners. The LTTO works closely with SIB's Group Leaders in order to offer innovative products and services. Companies involved in medicine and life sciences can collaborate with SIB to complement their internal capacity.
- **Services to industry:**
  - Scientific support and data analysis thanks to in-house computational tools and in-depth expertise
  - Education and practical training in the use of software and analysis methods
  - Text mining and web monitoring in the life science and clinical fields (e.g. creation of a patient cohort, monitoring of social media platforms for drug safety surveillance).
- **Management of the company GeneBio:** the SIB's commercial arm GeneBio commercializes software tools and resources developed by SIB, e.g. Melanie, Prosite and SmileMS.
- **Legal Advice:** The LTTO is responsible for the legal matters of the Institute and advises SIB Management and Group Leaders on a broad range of evolving legal issues such as copyrights, personal data protection and research involving human beings.

Highlights 2015

- Increase in the service activities of SIB Groups for Swiss and international industry
- Definition of the legal framework for collaborations with academic and hospital institution partners in the context of innovative projects, such as Personalized Health
- Set-up of the SIB Licensing Advisory Committee to help Group Leaders choose appropriate models (open source, services for a fee, commercialization, etc.) for distributing their original work
- Launch of the Melanie software version 8 with a brand new dedicated website
- Implementation of a Contract Management System at SIB.

Training

One of SIB's missions is to train the next generation of bioinformaticians and ensure that life scientists make the best use of bioinformatics and SIB resources. Under the leadership of Patricia Palagi, SIB Training is in charge of promoting and coordinating training in bioinformatics, both in Switzerland and internationally.

20
SIB groups engaged in teaching activities

Nearly
1,000
trainees

Over
75
experts and trainers

41
short workshops organized

11
long block-courses coordinated

In 2015, SIB Training focused its efforts on four main axes:

- **Professional training**
  The SIB professional training portfolio, which continuously evolves to meet the scientific community's needs, proposes courses on several bioinformatics topics. In 2015, the "top ten" courses were related to next generation sequencing analysis, high performance computing, statistics and R, the statistical package. More information is available on the SIB Training portal: www.sib.swiss/training

- **SIB PhD Training Network**
  The SIB PhD Training Network specifically targets Swiss bioinformatics and computational biology PhD students. Special events in 2015 included the annual retreat, Bioinformatics in the Chalet, and two seasonal schools, among them the summer school on "Genomics and Evolution of Microbial Pathogens" jointly organized with Zurich University of Applied Sciences (ZHAW) and SystemsX.ch.

- **The SIB Fellowship programme**
  Through its Fellowship programme launched in 2012, SIB aims to create a pool of excellent young bioinformaticians by giving them the opportunity to carry out their PhD research in one of SIB's groups. Thanks to the generous support of committed partners, four outstanding laureates were selected in 2015: Monica Ticlla (Peru), Emma Ricart Altimiras (Spain), Luis Miguel de Oliveira Vilaca (Portugal) and Dennis Haake (Germany).

- **International collaborations**
  In order to reinforce connections with the international and European bioinformatics training community, SIB Training once again had the pleasure of co-organizing the Workshop in Education for Bioinformatics for the International Society for Computational Biology. SIB Training has also co-organized two workshops on e-learning methods and strategies in collaboration with GOBLET and the ELIXIR nodes in Slovenia, the United Kingdom and the Netherlands.

Outreach

Another SIB's missions is to bring bioinformatics to the layman, contributing to a better understanding of this science. SIB has created a broad range of workshops and organized many events to describe bioinformatics in an easy and playful way.

More than
3,000
participants in SIB activities for the layman

More than
1,000
children participating in SIB school activities

In 2015, more than 1,000 students (aged from 10 to 20) participated in various bioinformatics-related activities during events in or outside the classroom such as:

- **Workshops organized in collaboration with various schools or public laboratories**, BioScope (UNIGE), Chimiscope (UNIGE) and L'Eprouvette (UNIL).
- **The TecDays**, a SATW initiative that aims to promote technical understanding and stimulate curiosity about scientific and technical training.
- **(R)amène ta Science** (ramene-ta-science.unige.ch/), a concept developed by UNIGE. Academic experts train students who will then themselves conduct the workshop at their own school.
- **Training high school teachers**: SIB organized a two-day workshop to introduce bioinformatics and drug design (www.drug-design-workshop.ch/) to high school teachers from the American School in Switzerland.
- **Initiation of 30 young Israeli high school students** to basic bioinformatics, drug design and personalized health.

More than 3,000 people participated in various bioinformatics-related activities during public events for the layman, such as:

- **LabLife at the Musée de la Main, Lausanne**: SIB was invited to give several workshops to high school students and visitors of all ages during the event "Rencontres avec les chercheurs".
- **Campus Biotech Open Day, Geneva**: SIB's booth presented posters, videos and workshops about the role of bioinformatics and its contribution to medicine and health.
- **RECHERCHE LIVE**: SIB took part in an exhibition organized by UNIGE, which was paired with the Swiss Academy of Sciences' itinerant exhibition to mark its 200th anniversary. SIB offered activities related to its Drug Design workshop.
- **Expanding your Horizons, Geneva**: During the 4th bi-annual event for young girls organized by Expanding Your Horizons (www.eyhn.org), SIB presented its "Metagenomic Pizza" workshop, which allows the discovery of various bioinformatics resources, such as Blast and UniProtKB.
- **UNIGE Faculty Diabetes Centre Open House**: During the Open House event of the newly formed UNIGE Faculty Diabetes Centre, UNIGE and SIB held a workshop during which visitors were able to discover how bioinformatics helps find genetic variations associated with diabetes.

Research

As of today, SIB counts 61 bioinformatics research groups and some 750 scientists from the major Swiss schools of higher education and research institutes.

One of SIB's missions is to lead and coordinate the field of bioinformatics in Switzerland:

1) by fostering collaborations:

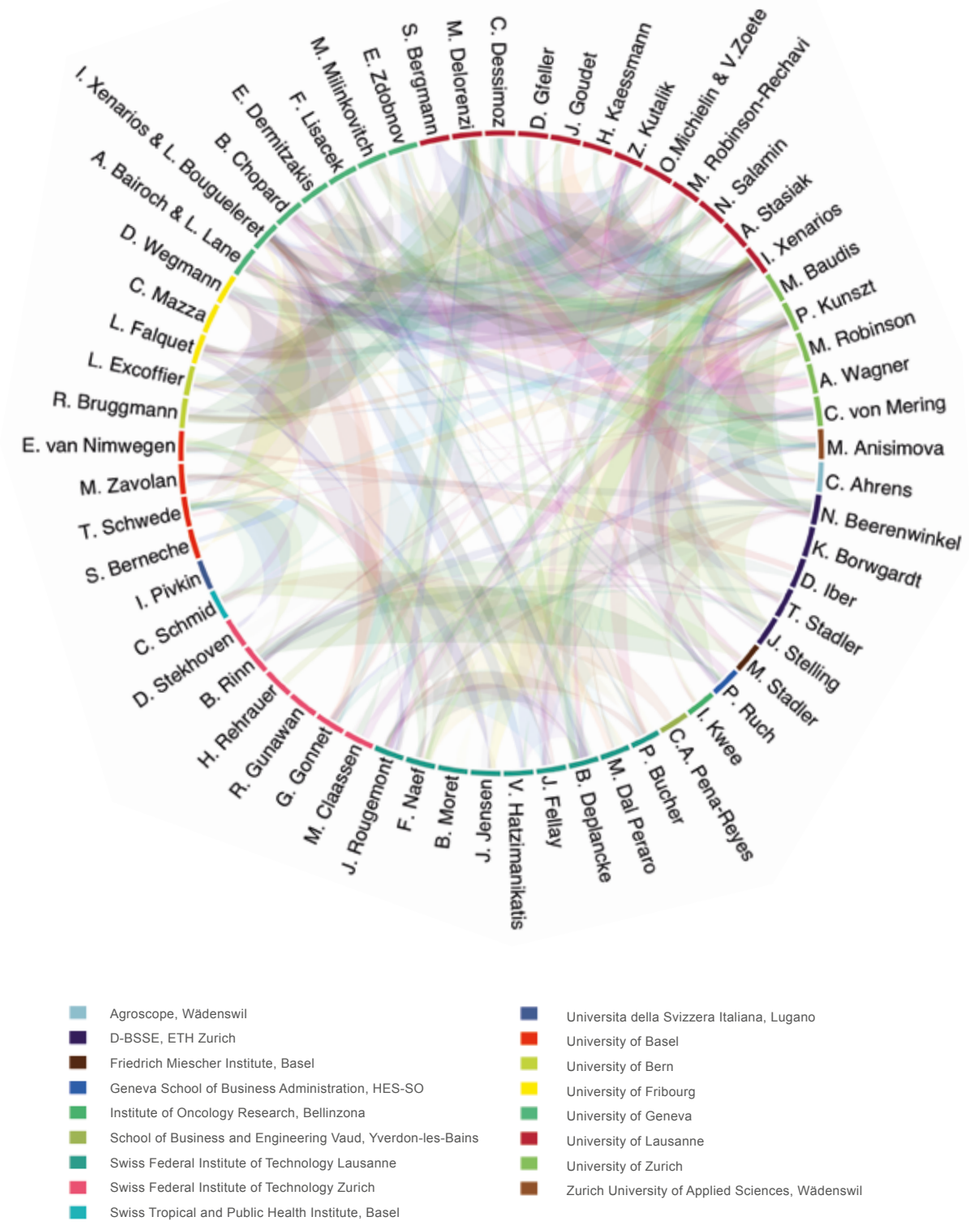
The federal structure of SIB allows its data science experts to join forces in order to advance biological and medical research. Over time, a dense collaborative network has been established between the SIB groups from Swiss universities and research institutes located in the cantons of Basel, Bern, Geneva, Fribourg, Ticino, Vaud and Zurich.

SIB collaborates at the international level with many renowned institutions, for instance:

- In Europe: the European Molecular Biology Laboratory-European Bioinformatics Institute (EMBL-EBI), the Bioinformatics Services to Swedish Life Science (BILS), the Spanish National Bioinformatics Institute (INB), the Dutch Techcentre for Life Sciences (DTL)
- In the US: the National Institutes of Health (NIH), the National Center for Biotechnology Information (NCBI), the Protein Information Resource (PIR).
- Elsewhere: SOKA University (Japan), Macquarie University (Australia), the University of Cape Town (South Africa), the Weizmann Institute of Science (Israel).

2) by promoting and fostering excellence and innovation in the field of bioinformatics:

Since 2008, SIB has been honoring young researchers and ground-breaking resources on the national and international level through the SIB awards (SIB International Young Bioinformatician Award, SIB International Bioinformatics Resource Award, SIB Best Swiss Bioinformatics Graduate Paper Award).



SIB collaboration network. The collaboration network was generated from a programme developed by Michael Baudis, SIB Group Leader: progenetix.org/collabplots/

New groups

Karsten Borgwardt

Machine Learning & Computational Biology Lab
ETH Zurich

What do they do?
Karsten Borgwardt's lab develops novel data mining algorithms to detect patterns and statistical dependencies in large datasets from biology and medicine. Their hope is to reach two goals: 1) Enable the automatic generation of new knowledge from big data through machine learning, and 2) gain an understanding of the relationship between biological systems and their molecular properties. This understanding is of fundamental importance for personalized medicine, which tailors medical treatment to the molecular properties of a person.

Highlights 2015
Variant annotation is important in genomics as it serves to quantify a patient's disease risk based on sequence variants. In 2015, the Borgwardt lab published the most comprehensive empirical comparison of various genome variant annotation tools to date. They discovered that, 1) training and test datasets for the prediction tools often overlap, leading to overly optimistic results and 2) naïve strategies that predict the same annotation for all variants from the same gene or protein outperform all state-of-the-art tools due to the way current variant databases are populated. These findings will be essential when designing new genome variant annotation tools.

Many branches of Computational Biology use high-dimensional feature selection as their key algorithmic instrument to discover relationships between genomic properties and phenotypes of interest. When looking for multifactorial combinations of these properties, one faces a huge multiple testing problem due to the large number of candidate combinations. Recent work by Borgwardt's lab has shown that this problem can be solved in important settings in practice.

Publications 2015

Grimm D *et al.* The evaluation of tools used to predict the impact of missense variants is hindered by two types of circularity. Human Mutation 2015; 36(5):513-523.

Llinares-López F *et al.* Fast and Memory-Efficient Significant Pattern Mining via Permutation Testing. Proceedings of the 21st ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD2015) 2015: 725-734.

Llinares-López F *et al.* Genome-wide detection of intervals of genetic heterogeneity associated with complex traits. ISMB 2015, Bioinformatics 2015; 31 (12): i240-i249.

Christophe Dessimoz

Laboratory of Computational Evolutionary Biology and Genomics
University of Lausanne

What do they do?
Christophe Dessimoz' laboratory studies evolutionary and functional relationships between genes, genomes and species. Two key questions are: How can we extrapolate to the rest of life and in the best way possible our current knowledge in molecular biology while concentrating on just a handful of model organisms? How can we exploit the wealth and diversity of life to get a better grasp on specific organisms or systems of interest?

The group's activities are divided between bioinformatics methodology and resource development, and the application of these – most typically in collaboration with experimentalists.

Highlights 2015
Christophe Dessimoz and his team published a large empirical study which shows that, on average, alignment filtering makes phylogenetic trees overwhelmingly worse (Tan et al. Syst Biol 2015). The group also demonstrated that contrary to a recent claim in the literature – with regard to performing multiple sequence alignments – the evolutionarily correct guide tree gives the best results (Tan et al. PNAS 2015). A review was also published on lateral gene transfer both as an article in a peer-reviewed journal (Ravenhall et al. PLOS Comp Biol 2015) and on Wikipedia – using an exciting new format called the "Topic Page". This brought them to reflect on this new experience as well as on the challenges and opportunities of publishing scholarly work on Wikipedia. The Dessimoz lab was involved in co-organizing the Quest for Orthologs 4 conference in Barcelona, which was combined with the lab retreat. An updated paper on the OMA orthology database (Altenhoff et al. NAR 2015) was published, and coincided with the 10th anniversary of the resource and a new web interface.

Comments and thoughts can be found on the lab's blog: lab.dessimoz.org/blog/

Publications 2015

Altenhoff AM *et al.* The OMA orthology database in 2015: function predictions, better plant support, synteny view and other improvements. Nucleic Acids Res 2015;43:D240–D249.

Tan G *et al.* Current Methods for Automated Filtering of Multiple Sequence Alignments Frequently Worsen Single-Gene Phylogenetic Inference Syst Biol 2015;64:778–791.

Tan G *et al.* Simple chained guide trees give poorer multiple sequence alignments than inferred trees in simulation and phylogenetic benchmarks. Proc Natl Acad Sci USA 2015;112:E99–E100.

Carlos-Andrés Peña-Reyes

Computational Biomed Research and Applications Group, Haute Ecole d'Ingénierie et Gestion de Vaud (HEIG-VD) – HES-SO

What do they do?
With the advent of high-throughput technologies and clinical information systems, life and clinical sciences now produce very large amounts of data (big data). Peña-Reyes and his team's goal is to uncover hidden patterns in these data and to build data-driven models as a tool to discover biomarkers and assist clinicians in their decisions. Their projects encompass the fields of transcriptomics, systems biology, and clinical bioinformatics and analytics.

Highlights 2015
Carlos Andrés Peña and his team were strongly involved in the Eurostars-2 FISHGUARD consortium, which brings together the SMEs Bioscientia and Biotem, and their group, CI4CB, at HEIG-VD. The aim of this project is to develop a rapid in-field screening test against two fish viruses, which cause high losses for the European aquaculture industry. Developing this assay requires generating highly sensitive and specific antibodies against virus antigens. As the main bioinformatics partner, the group developed an immunoinformatics pipeline to predict linear epitopes. For this, they first identified conserved regions in the antigens using all the publicly-available data on the virus isolates. They then predicted epitopes in these antigens using nine competitive classifiers. In order to combine their predictions, the team defined a novel ensemble combination rule taking into account classifier diversity. The method is currently being benchmarked and their partners have just started generating the antibodies based on the group's predictions.

Hubert Rehrauer

Genome Informatics
Functional Genomics Center Zurich, ETH Zurich

What do they do?
Hubert Rehrauer and his team are dedicated to the processing, analysis and interpretation of next generation sequencing data. They interact closely with research groups and provide tailored comprehensive bioinformatics solutions. Additionally, they provide standard analysis pipelines for the more frequent research questions. They also train researchers and bioinformaticians on various aspects of data analysis, as well as provide access to their computing infrastructure for running analyses.

Highlights 2015
During 2015 and for the scientific user community, Hubert Rehrauer and his team made web-based analyses of genomic and other sequencing data available (fgcz-sushi.uzh.ch). With their approach, researchers are able to access a high-performance computing infrastructure with their web-browser and apply state-of-the-art statistical analyses to their data. Besides the intuitive interface provided by the system, researchers benefit from a fully documented, portable and shareable result. This is a major step towards reproducible and collaborative research.

The group also participated in the CAMDA (Critical Assessment of Massive Data) challenge and assessed the value of molecular profiling for the prognosis of cancer survival. In this regard, they analysed, in detail, data from the TCGA consortium (The Cancer Genome Atlas) and evaluated the prognostic performance when different data sources are integrated to form a joined prognosis.

Publications 2015

Müller-Edenborn K *et al.* Hypoxia attenuates the proinflammatory response in colon cancer cells by regulating IκB. Oncotarget 2015;6(24):20288-301.

Daniel Stekhoven

NEXUS Personalized HealthTechnologies (NEXUS) Clinical Bioinformatics Unit (CBU)
ETH Zurich

What do they do?
The Clinical Bioinformatics Unit of NEXUS Personalized Health Technologies, a technology platform of ETH Zurich, offers customized bioinformatics and statistics services for analyses and projects in the field of biomedical research. Daniel Stekhoven and his group maintain close collaborations with hospitals in Zurich and Basel, introducing state-of-the-art data analyses. Based on a fee-for-service model, their aim is to meet individual project goals and establish close collaborations with their customers. Their ambition is to help the end user understand what the results mean. However, their services are not limited to analysis alone. They also offer support in writing and reviewing manuscript texts, as well as delivery of manuscript-ready tables and figures.

Highlights 2015
Daniel Stekhoven and his team have been working on their somatic mutation calling pipeline for clinical diagnostics, which enables the use of whole exome sequencing for clinical decision-making. With the Clinic of Dermatology at the University Hospital of Zurich, the team expects to start giving therapy suggestions in the first quarter of 2016.

Enabling biomedical researchers to access public repositories that contain genomic data and offer an easy way to make exploratory analyses is an excellent way of accelerating research. The group is also developing a resource called www.tcgabrowser.com, which provides differential expression and survival analysis on all cancer types available on TCGA.

In 2015, the Zurich Metastasis Project (Z-Met) was announced by the Competence Center Personalized Medicine, and covers several aspects of their five flagship projects. Daniel Stekhoven and his team are the main bioinformatics/statistics provider for these flagship projects, together with their sister unit Theragnostics Discovery inside NEXUS Personalized Health Technologies.

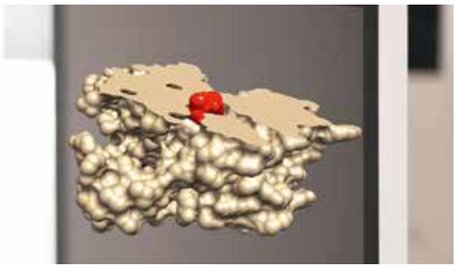
Publications 2015

Moisan A *et al.* White-to-brown metabolic conversion of human adipocytes by JAK inhibition. Nature Cell Biology 2015;17(1):57-67

Singer F *et al.* Supporting Personalized Medicine with State-of-the-art Technologies. EMBL | Stanford Conference – Personalised Health 2015

Zickmann F *et al.* MSProGene: integrative proteogenomics beyond six-frames and single nucleotide polymorphisms. Bioinformatics 2015;31(12):106-115

A wide variety of activity domains



Bioinformatics is the application of computer technology to the understanding and effective use of biological data. It is thus an interdisciplinary field, targeting different areas of medicine and life sciences. The vast majority of SIB groups are therefore involved in numerous domains.

SIB's research activities focus on seven main domains:

- Genes and genomes
- Proteins and proteomes
- Medicine and health
- Evolution and phylogeny
- Structural biology
- Systems biology
- Bioinformatics infrastructure

The table on the facing page presents SIB's main fields of activity. The SIB groups (as of 1 January 2016) are classified according to their main research area (indicated by a full black square) and then in alphabetical order. The other domains, in which the groups are working, are indicated by empty squares.

	GENES AND GENOMES pp. 43-48	PROTEINS AND PROTEOMES pp. 49-52	MEDICINE AND HEALTH pp. 53-57	EVOLUTION AND PHYLOGENY pp. 58-64	STRUCTURAL BIOLOGY pp. 65-67	SYSTEMS BIOLOGY pp. 68-74	BIOINFORMATICS INFRASTRUCTURE pp. 75-77
Sven Bergmann	■		□				□
Rémy Bruggmann	■	□	□			□	□
Philipp Bucher	■		□				□
Bart Deplancke	■		□	□		□	□
Emmanouil Dermitzakis	■		□			□	
Laurent Falquet	■	□	□				□
Zoltán Kutalik	■		□	□			□
Erik van Nimwegen	■			□		□	□
Mark D. Robinson	■		□				
Michael Stadler	■		□			□	□
Andrzej Stasiak	■	□	□		□		
Evgeny Zdobnov	■	□	□	□			□
Christian Ahrens	□	■				□	□
Amos Bairoch & Lydie Lane		■	□				□
Frédérique Lisacek		■	□			□	□
Christian von Mering	□	■		□		□	□
Ioannis Xenarios & Lydie Bougueleret	□	■	□	□	□	□	□
Michael Baudis	□		■			□	□
Niko Beerenwinkel	□		■	□		□	□
Mauro Delorenzi	□		■			□	□
Jacques Fellay	□		■			□	□
David Gfeller	□	□	■		□		
Ivo Kwee	□		■			□	
Patrick Ruch		□	■				□
Christoph Schmid	□		■				
Maria Anisimova	□	□		■			□
Laurent Excoffier	□			■			□
Gaston Gonnet	□	□		■			□
Jérôme Goudet	□			■			□
Jeffrey D. Jensen	□			■			□
Bernard Moret	□			■			
Marc Robinson-Rechavi	□			■			□
Nicolas Salamin	□			■			□
Tanja Stadler	□		□	■			□
Andreas Wagner	□			■			
Daniel Wegmann	□		□	■			□
Simon Bernèche		□			■		
Matteo Dal Peraro		□	□		■		□
Olivier Michielin & Vincent Zoete		□	□		■	□	□
Torsten Schwede		□			■		□
Bastien Chopard			□			■	□
Manfred Claassen			□			■	
Rudiyanto Gunawan	□		□			■	□
Vassily Hatzimanikatis	□	□	□			■	□
Dagmar Iber			□			■	□
Christian Mazza	□					■	
Michel Milinkovitch	□			□		■	□
Félix Naef	□	□	□			■	
Igor V. Pivkin			□			■	
Jörg Stelling						■	□
Mihaela Zavolan	□	□	□			■	
Peter Kunszt	□	□	□	□	□	□	■
Bernd Rinn							■
Jacques Rougemont	□			□		□	■
Ioannis Xenarios	□	□	□	□	□	□	■

Genes and genomes

Genome is the word used by life scientists to describe the sum of genetic material, including genes, inherited by a living being. A genome is like an open book on the processes of life, if you know how to read it.

Bioinformatics develops tools not only to read the genetic information, but also to store the resulting data, analyse and interpret them. Aberrations in genetic material can be at the heart of diseases such as cancer or Down syndrome.

Sven Bergmann
Computational Biology Group
University of Lausanne

What do they do?
At the Computational Biology Group, Sven Bergman and his team develop concepts and algorithmic tools for the analysis of large-scale biological and clinical data. They focus on the integration of genotypic and phenotypic datasets from mammalian cells or clinical studies. A key approach is the reduction of complexity through modular and network analysis. A complementary direction of their research pertains to relatively small genetic networks whose components are well known.

Highlights 2015
During the course of 2015, Sven Bergmann and his team made substantial progress in the fast and accurate computation of gene and pathway scores [PLOS Genet. 2016; 12(1):e1005616]. A direct application of their analysis tool – called PASCAL – resulted from the analysis of data from the FANTOM5 project and revealed that genetic variants associated with different diseases can be used to identify the relevant tissues enriched for genetic networks that are perturbed by the variants [*Nature Methods*, in press].

Main publications 2015
Lamparter D *et al.* Fast and Rigorous Computation of Gene and Pathway Scores from SNP-Based Summary Statistics. PLOS Comput Biol 2016;12(1):e1004714.
Marbach D *et al.* Tissue-specific regulatory circuits reveal variable modular perturbations across complex diseases. Nat Methods: in press.
Vonesch SC *et al.* Genome-Wide Analysis Reveals Novel Regulators of Growth in Drosophila melanogaster. PLOS Genet 2016;12(1):e1005616.

Rémy Bruggmann
Interfaculty Bioinformatics Unit – IBU
University of Bern

What do they do?
At the Interfaculty Bioinformatics Unit of the University of Bern (IBU), Rémy Bruggmann and his team provide services and expertise to assist researchers of the three “Life-Science” Faculties (i.e. Sciences, Medicine, and VetSuisse) in data analysis and project planning for large-scale experiments (e.g. next generation sequencing, genome assembly). Furthermore, the group has its own research programme and collaborates on large and complex projects. They develop methods to analyse high-throughput data. They have a high-performance compute cluster and a data storage system that they use for their own research, collaborations and service projects.

Highlights 2015
Rémy Bruggmann and his team participated in a project to assess whole exome sequencing (WES) methods. WES is used more and more in both research and diagnostics to search for potentially deleterious mutations in a patient. However, several protocols and versions exist, and which one to choose is a difficult task for a researcher when faced with a specific question. To guide researchers in an appropriate selection, the team carried out an in-depth analysis of current WES enrichment protocols. They found that the performance varies significantly between different protocols. As an illustration, the captured regions can be quite different from the theoretical ones, and the robustness of one and the same method between different service providers varies for some protocols more than it does for others. Besides WES, the team also suggests that enrichment-free whole genome sequencing (WGS) can overcome the limitations of WES particularly in GC-rich exonic regions.

In 2015, the group set up a comprehensive bioinformatics teaching module to promote bioinformatics. They successfully trained many biologists at all stages of their career, i.e. from postdocs to full professors.

Main publications 2015
Li H *et al.* The outer mucus layer hosts a distinct intestinal microbial niche. Nat Commun 2015;6:8292.
Meienberg J *et al.* New insights into the performance of human whole-exome capture platforms. Nucleic Acids Res 2015;43:e76–e76.

Philipp Bucher
Computational Cancer Genomics Group
EPFL, Lausanne

What do they do?
At the Computational Cancer Genomics Group, Philipp Bucher and his team are interested in gene regulation both in healthy and diseased cells. Breakthroughs in genomics technologies have led to the production of large volumes of data that could potentially tell us something about how gene regulatory instructions are encoded in our DNA. Bucher's group develops new algorithms, computer programs, web services and databases that will help them and others to extract knowledge and understanding from such data.

Highlights 2015
During the course of 2015, Philipp Bucher and his team extended EPDNew to two new model organisms, i.e. *S. cerevisiae* (baker's yeast) and *S. pombe* (fission yeast) for which they released comprehensive promoter collections.
The group also gave an SIB course entitled "Chip-seq data analysis: from quality check to motif discovery and more - An introduction to the tools and databases of the EPD team". The goal was to teach new and prospective users how to use their public resources in an efficient and effective manner.

A workshop entitled "Beyond Position weight matrices – towards next generation tools for predicting protein-DNA interactions" was also organized at the [BC]² Basel Computational Biology Conference. Experts in the field discussed the opportunities and challenges of novel high-throughput technologies for profiling transcription factor binding specificity both *in vitro* and *in vivo*.

With regard to ChIP-seq tools on the Amazon cloud, the group created a public AMI (Amazon Machine Image) for EC2 under the name ChIP-Seq-Tools_SIB (ID: ami-a9c3dc99).

Main publications 2015

Ambrosini G *et al.* Principles of ChIP-seq data analysis illustrated with examples. Genomics Comp Biol 2015;1(1):e22.
Dreos R *et al.* The Eukaryotic Promoter Database: expansion of EPDnew and new promoter analysis tools. Nucleic Acids Res 2015;43(Database issue):D92-6.

Bart Deplancke
Laboratory of Systems Biology and Genetics
EPFL, Lausanne

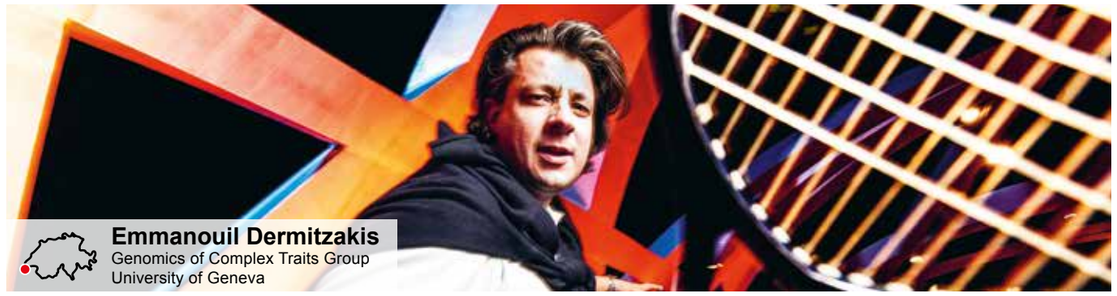
What do they do?
At the Laboratory of Systems Biology and Genetics (LSBG), Bart Deplancke and his team use high-throughput sequencing, single cell genomics, microfluidics, large-scale yeast screens, and computational approaches 1) to decipher the regulatory code in Drosophila and Mammals with a specific focus on mesenchymal stem cell function and gut immunity and 2) to examine how variations in this code affect molecular and organismal diversity. In addition to their research interests, they are actively pursuing the development of new research tools and pipelines that enable a better characterization of gene regulatory networks.

Highlights 2015
The fundamental discovery that most complex trait-associated variants are located in non-coding, putatively regulatory regions of the genome has brought into the spotlight transcription factor (TF)-DNA interactions as key mediators of phenotypic variation. Over the years, a canonical model emerged pointing to nucleotide variation in TF motifs as the principle source of variable TF binding. Recent findings are, however, challenging this model and revealing that the majority of variable TF binding events are in fact driven by variants that are far away (several hundred to several thousand bp away) from the cognate motifs.

During the course of 2015, Bart Deplancke's lab, with the Laboratories of Professors Dermitzakis (University of Geneva), Hernandez and Reymond (University of Lausanne) has now introduced the novel concept of variable chromatin modules (VCMs) which provides a molecular rationale for these long-distance genetic effects on TF DNA binding, as detailed in Waszak, Delaneau et al. Cell 2015.

Main publications 2015

Bou Sleiman MS *et al.* Genetic, molecular and physiological basis of variation in Drosophila gut immunocompetence. Nat Commun 2015;6:7829.
Schertel C *et al.* A large-scale, in vivo transcription factor screen defines bivalent chromatin as a key property of regulatory factors mediating Drosophila wing development. Genome Res 2015;25(4):514-23.
Waszak SM *et al.* Population Variation and Genetic Control of Modular Chromatin Architecture in Humans. Cell 2015;162(5):1039-50.



Emmanouil Dermitzakis
Genomics of Complex Traits Group
University of Geneva

What do they do?
At the Genomics of Complex Traits Group, Emmanouil Dermitzakis and his team have a strong interest in population genomics and genetics of complex traits. They use various methodologies to understand the role of genetic variation in phenotypic variation. They also aim to understand what fraction of genetic variation is harboured within functional elements of the human genome. The group's main focus is on genome-wide analysis of gene expression and cellular phenotypes and their association with nucleotide variation. They attempt to detect functional genetic variation in regulatory elements and subsequently use functional variation and accurately measured gene expression variation to bridge the genotype with disease phenotypes in association studies.

Highlights 2015
During the course of 2015, Emmanouil Dermitzakis and his team had a paper published on their analysis of RNA sequencing data generated as part of the pilot phase of the Genotype-Tissue Expression (GTEx) project. Their findings help to understand the cellular and biological consequences of human genetic variation, and of the heterogeneity of such effects among a diverse set of human tissues.
The team also developed a molecular quantitative trait loci (QTL) browser, which is currently in its alpha stage. A software package was also developed to become QTL tools.

Main publications 2015

GTEx Consortium (authors). The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science 2015 May 8;348(6235):648-60.
Ongen H *et al.* Fast and efficient QTL mapper for thousands of molecular phenotypes. Bioinformatics 2015 Dec 26; pii: btv722.
Waszak SM *et al.* Population Variation and Genetic Control of Modular Chromatin Architecture in Humans. Cell 2015;162(5):1039-50.

Laurent Falquet
Bioinformatics Unravelling Group – BUGFri
University of Fribourg

What do they do?
At the Bioinformatics Unraveling Group of the University of Fribourg (BUGFri), Laurent Falquet and his team support life science researchers by providing expertise in data analysis of next generation sequencing (NGS) experiments, or any large-scale biological experiment requiring bioinformatics resources. They focus on genome assembly, annotation and comparison as well as on mutant and structure variant identification by resequencing. They also perform metagenomics, RNAseq and ChIPseq data analysis, proteome clustering and ortholog/paralog classification, as well as pathway and gene set enrichment analysis.

Highlights 2015
During the course of 2015, the web server PACMAN (PACific biosciences Methylation Analyzer) was developed by a master's degree student in Laurent Falquet's group. This web site allows a user to upload a full, or draft, bacterial genome, together with the motifs.gff file of a PacBio sequencing analysis.
The PACMAN web server uses Circos to generate a graphical view of the most important methylation motifs. The user can preselect among several possible views and filters. The output is a publication-ready PDF or PNG. In addition, the detailed page can be used to identify genes near hyper- and hypo- methylated regions. PACMAN is hosted by Vital-IT.
Link: www.unifr.ch/bugfri/pacman

Main publications 2015

Liljander A *et al.* (2015). A field-applicable recombinase polymerase amplification assay for rapid detection of Mycoplasma capricolum subsp. capripneumoniae. J Clin Microbiol 2015 Jun 17; pii: JCM.00623-15.
Pillet B *et al.* The Dedicated Chaperone Acl4 Escorts Ribosomal Protein Rpl4 to Its Nuclear Pre-60S Assembly Site. PLOS Genet 2015;11:e1005565.
Rich MK *et al.* A petunia GRAS transcription factor controls symbiotic gene expression and fungal morphogenesis in arbuscular mycorrhiza. Plant Physiol 2015;168:788-797.

Zoltán Kutalik
Statistical Genetics Group
University of Lausanne

What do they do?
At the Statistical Genetics Group, Zoltan Kutalik and his team are interested in the development of statistical methodologies in order to decipher the genetic architecture of complex human traits related to obesity. In order to do this, the group efficiently combines genome-wide association studies (GWAS) with different -omics data to enhance their understanding of the genetic network of the human genome. Furthermore, they are heavily involved in the activities of the GIANT consortium as well as in various clinical genetic analyses.

Highlights 2015
During 2015, Kutalik and his team conducted a systematic genome-wide search, involving 320,000 individuals, to identify genetic loci with age- and/or sex-dependent effects on body size and shape. They identified 15 loci whose effect on BMI (Body Mass Index) was different in older adults compared to younger ones. The team also identified 44 loci whose effect on waist-to-hip ratio differed between men and women.
Furthermore, the group identified 831 carriers (10.5%) of rare intermediate size copy number variants (CNVs) among 7,877 genotyped Estonian Biobank individuals. This group of carriers had increased prevalence of intellectual disability and decreased education attainment. 5.1% of these deletion carriers had an intellectual disability compared with 1.7 % in the Estonian cohort without detected CNVs. Among the deletion carriers, 33.5% did not graduate from high school, while this ratio was only 25.3% in the Estonian general population.
In 2015, Zoltan Kutalik was appointed an honorary senior lecturer at the University of Exeter.

Main publications 2015
Männik K *et al.* Copy Number Variations and Cognitive Phenotypes in Unselected Populations. JAMA 2015 May 26;313(20):2044-2054.
Rüeger S *et al.* Impact of common risk factors of fibrosis progression in chronic hepatitis C. Gut 2015 Oct;64(10):1605-15.
Winkler TW *et al.* The Influence of Age and Sex on Genetic Associations with Adult Body Size and Shape: A Large-Scale Genome-Wide Interaction Study. PLOS Genet 2015 Oct 1;11(10).

Erik van Nimwegen
Genome Systems Biology Group
University of Basel

What do they do?
The main research interest of the Genome Systems Biology (GSB) Group is the study of genome-wide regulatory systems, to reconstruct them from high-throughput molecular data, understand and model how they have evolved, and search for design principles in their construction. In particular, the group is developing and applying new algorithmic tools for the automated reconstruction of genome-wide regulatory networks from comparative genomic, deep sequencing, and other high-throughput data. In addition, methods are being developed for studying genome evolution and the evolution of regulatory networks in particular.

Highlights 2015
The main highlight in 2015 was the publication of the first major work from the group's wet lab in which they uncovered a general mechanism for the de novo evolution of gene regulation, and found that gene expression noise plays a crucial role in facilitating its evolution.

A second major highlight involves Crunch – a completely automated webserver for ChIP-seq analysis – which went online at crunch.unibas.ch as part of the group's SwissRegulon portal. Crunch performs all steps of the ChIP-seq analysis, from raw data quality control, read mapping, fragment-size estimation, peak detection, and peak annotation to novel DNA sequence motif analysis. In particular, Crunch finds a complementary set of motifs that can explain the ChIP-seq data, and then comprehensively characterizes the occurrence of these motifs across the ChIP binding peaks.

Main publications 2015
Pemberton-Ross *PJ et al.* ARMADA: Using motif activity dynamics to infer gene regulatory networks from gene expression data. Methods 2015;85:62-74.
Schertel C *et al.* A large-scale, in vivo transcription factor screen defines bivalent chromatin as a key property of regulatory factors mediating Drosophila wing development. Genome Res 2015;25(4):514-23.
Wolf L *et al.* Expression noise facilitates the evolution of gene regulation. eLife 2015;4:e05856.

Mark D. Robinson
Statistical Bioinformatics Group
University of Zurich

What do they do?
At the Statistical Bioinformatics Group, Mark Robinson and his group develops robust data analysis solutions, including new or improved methods, for the analysis of genome-scale data. They develop statistical methods for interpreting data from high-throughput sequencing and other technologies in the context of genome sequencing, gene expression and regulation and analysis of epigenomes. They are largely data- and problem-driven, and ultimately the methods they develop are catered to the characteristics of the technology platform generating the data. The group develops publicly-available open-source software tools, generally through the Bioconductor project. The majority of their time is spent on collaborative projects and development of statistical methods with accompanying software. Where needed, they design experiments and collect data to compare the performance of competing methods and platforms.

Main publications 2015
Soneson C *et al.* Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000 Research 2015 Dec 30 [revised 2016 Feb 29];4:1521.
Zhou X *et al.* Do count-based differential expression methods perform poorly when genes are expressed in only one condition? Genome Biol 2015;16:222.
Bilgin T *et al.* Tandem repeat variation in human and great ape populations and its impact on gene expression divergence. Genome Res 2015;25(11):1591-9.

Michael Stadler
FMI Computational Biology Group
Friedrich Miescher Institute, Basel

What do they do?
The FMI Computational Biology Group is located at the Friedrich Miescher Institute in Basel. Michael Stadler and his group study gene regulation through the analysis and modelling of genome-wide datasets. They collaborate closely with experimental researchers on various biological topics including cancer progression and cellular differentiation. Using statistical approaches, they aim to gain a better understanding on how the different layers of epigenetic, transcriptional and post-transcriptional regulation interact and contribute to the control of gene expression. The great majority of their projects measure various aspects of gene expression including DNA methylation, single cell transcription, protein-binding to DNA, and translation using high-throughput sequencing.

Highlights 2015
In 2015, Michael Stadler and his team developed a novel computational approach to analyse gene expression data. The approach compares sequenced RNA fragments in introns and exons, and can thereby discern the contribution of transcriptional and post-transcriptional regulation to gene expression. The description of this new method and the results of the analysis have been published in *Nature Biotechnology*.

For more information see "The hidden treasure in RNA-seq":
www.fmi.ch/news/releases/articles/stadler.150713.html

Main publications 2015
Gaidatzis D *et al.* Analysis of intronic and exonic reads in RNA-seq data characterizes transcriptional and post-transcriptional regulation. Nat Biotechnol 2015;33(7):722-9.
Gaidatzis D *et al.* QuasR: quantification and annotation of short reads in R. Bioinformatics 2015;31(7):1130-2.
Royo H *et al.* Silencing of X-Linked MicroRNAs by Meiotic Sex Chromosome Inactivation. PLOS Genet 2015;11(10):e1005461.

Andrzej Stasiak
DNA and Chromosome Modelling Group
University of Lausanne

What do they do?
At the DNA and Chromosome Modelling Group, Andrzej Stasiak and his group apply Metropolis Monte-Carlo and Brownian dynamics simulations to elucidate how DNA molecules and chromatin fibres behave in living cells. Their group is especially interested in understanding chromosome structure and organization during interphase. They investigate effects of high crowding such as those known to occur in cell nuclei. They study consequences of transcription-induced supercoiling and topological consequences of DNA replication. The group also builds relatively simple models of interphase chromosomes that recapitulate the results of Chromosome Conformation Capture (3C) experiments.

Highlights 2015
1) In a collaborative project involving researchers in Poland and the USA, Stasiak's group has been involved in setting up a database of proteins that form knots and slipknots: knotprot.cent.uw.edu.pl/
2) In living cells, DNA molecules are highly crowded, which has an important effect on their overall shape and other properties. This crowding effect is difficult to assess and measure experimentally. The group therefore performed numerical simulation studies on the effects of physiological self-crowding on supercoiled DNA molecules, such as bacterial plasmids. It was observed that crowding does effectively enhance the known effects of DNA supercoiling.
3) It has been puzzling to understand how postreplicative catenanes that arise during the replication of circular DNA molecules are so efficiently decatenated in bacteria. Using numerical simulations, Stasiak and his team revealed a new form of long-distance cooperation between DNA gyrase and topoisomerase III that results in very efficient DNA decatenation.

Main publications 2015

Benedetti F *et al.* Effects of physiological self-crowding of DNA on shape and biological properties of DNA molecules with various levels of supercoiling. Nucleic Acids Res 2015;43(4):2390-9.
Jamroz M *et al.* KnotProt: a database of proteins with knots and slipknots. Nucleic Acids Res 2015;43(D1):D306-14.
Racko D *et al.* Generation of supercoils in nicked and gapped DNA drives DNA unknotting and postreplicative decatenation. Nucleic Acids Res 2015;43(15):7229-36.

Evgeny Zdobnov
Computational Evolutionary Genomics Group
University of Geneva

What do they do?
At the Computational Evolutionary Genomics Group, Evgeny Zdobnov and his team are active in the fields of comparative genomics and shotgun metagenomics. They study molecular evolution, develop approaches to genomics data analyses, and implement computational pipelines. The group also applies evolutionary models to digest sequencing data, and revises these models using the novel data. They study functional genomic elements on the basis of sequence variability among different species and within populations. Their interests range from arthropod genomics – including invertebrate vectors of human pathogens – to the evolution of viruses and clinical microbiology.

Highlights 2015
Evgeny Zdobnov and his team's 2015 highlights are two new bioinformatics resources: BUSCO and CEGA.
BUSCO (busco.ezlab.org) is an open-source software package for quantitative assessment of genome assembly and annotation completeness based on evolutionarily informed expectations of gene content, i.e. sets of Benchmarking Universal Single-Copy Orthologs derived from OrthoDB. Genomics has revolutionized biological research, but quality assessment of the resulting assembled sequences remains mostly limited to technical measures like N50, and now BUSCO addresses this important need from the biological perspective.
CEGA (cega.ezlab.org) is a catalogue of conserved elements (CNCs/UCEs) derived from multiple genomic alignments. The team identified CNCs independently for five vertebrate clades, each referring to a different last common ancestor and therefore to an overlapping but varying set of CNCs.

Main publications 2015

Dousse A *et al.* CEGA-a catalogue of conserved elements from genomic alignments. Nucleic Acids Res. Epub 2015 Nov 2.
Kriventseva EV *et al.* OrthoDB v8: update of the hierarchical catalogue of orthologs and the underlying free software. Nucleic Acids Res 2015;43:D250-6.
Simão FA *et al.* BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 2015;31(19):3210-2.

Proteins and proteomes

Proteome describes the entire set of proteins expressed by a cell, a tissue or an organism at a given time. Proteins are the products of genes and are involved in nearly every task in the body – from shaping cells to defending the body against pathogens.

An altered protein, produced by a mutation in its gene, can be at the heart of diseases such as cystic fibrosis or Creutzfeldt-Jakob disease. Bioinformatics develops tools to understand how proteins exercise their role.

Christian Ahrens

Bioinformatics and Proteogenomics Group
Agroscope, Wädenswil

What do they do?

Christian Ahrens and his team carry out research at the Bioinformatics and Proteogenomics Group centres around the bioinformatic integration and analysis of datasets from state-of-the-art -omics technologies, which they obtain through close collaboration with experimental biologists. These datasets include genome sequence, gene and protein expression, as well as metabolomics data. One particular focus is to exploit the unique advantages of proteomics data, including strategies to achieve complete proteome coverage (including the membrane proteome) and to identify all proteins encoded in a genome (proteogenomics).
Recently, they also started to study the role of microbiomes, e.g. for plant protection applying metagenomics, genomics and transcriptomics approaches.

Highlights 2015

Christian Ahrens and his team provided their proteomics data analysis and integration expertise to help uncover the fact that a new peptidomimetic (a small protein-like chain designed to mimic a peptide) which is active against Gram-negative pathogens, exhibits a novel mode of action (MoA), i.e. it selectively ruptures the outer membrane of Gram-negative bacteria.

Such novel peptidomimetics hold great potential for clinical use and will be developed further by Polyphor, a firm which is part of the research consortium led by Professor J.A. Robinson (Chemistry Dept., UZH). The consortium also includes the group of Professor L. Eberl (Microbiology Dept., UZH) and our bioinformatics group (Agroscope/SIB). This is the second peptidomimetic with a novel MoA to have been discovered by this consortium, and has raised great interest in the research community.

Main publications 2015

De Vrieze M *et al.* Volatile organic compounds from native potato-associated pseudomonas as potential anti-oomycete agents. Front Microbiol 2015;6:1295.
Lardi M *et al.* σ54-Dependent response to nitrogen limitation and virulence in *Burkholderia cenocepacia* strain H111. Appl Environ Microbiol 2015;81(12):4077-89.
Urfer M *et al*. A peptidomimetic antibiotic targets outer membrane proteins and disrupts selectively the outer membrane in *Escherichia coli*. J Biol Chem 2016;291(4):1921-32.

Amos Bairoch & Lydie Lane

Computer and Laboratory Investigation of Proteins
of Human Origin – CALIPHO / University of Geneva

What do they do?

At the CALIPHO Group, which stands for “Computer and Laboratory Investigation of Proteins of Human Origin”, Amos Bairoch, Lydie Lane and their team aim to use a combination of bioinformatics and experimental methodologies to increase the knowledge about the function of the 20,000 or so protein-coding genes that exist in the human genome. Their main mission is the development of neXtProt, a human protein knowledge resource. Recently, they focused on annotating the effects of human protein variations in the context of cancers and genetic diseases, and analysing results of high-throughput experiments to shed light on the function of selected sets of uncharacterized human proteins.

Highlights 2015

In 2015, Amos Bairoch, Lydie Lane and their team were particularly active in the continued development of their new advanced search: search.nextprot.org/. Based on SPARQL/RDF technology it allows the user to make very precise queries across the wealth of knowledge stored in neXtProt as well as other resources that also use SPARQL.

The team also continued its efforts in the annotation of the effect of protein variations in cancers and diseases, with an emphasis on sodium channel proteins, a family of proteins involved in many types of epilepsies as well as other genetic diseases.

Main publications 2015

Desmurs M *et al.* C11orf83, a mitochondrial cardiolipin-binding protein involved in bc1 complex assembly and supercomplex stabilization. Mol Cell Biol 2015;35:1139-1156.
Gaudet P *et al.* The neXtProt knowledgebase on human proteins: current status. Nucleic Acids Res 2015;42:D764-D770.
Holliday GL *et al.* Key challenges for the creation and maintenance of specialist protein resources. Proteins 2015;83:1005-1013.

Frédérique Lisacek

Proteome Informatics Group – PIG
University of Geneva

What do they do?

At the Proteome Informatics Group (PIG), Frédérique Lisacek and her team are involved in software and database development for the benefit of the proteomics and the glycomics communities. These resources are made available through the ExPASy server. Software tools support experimental mass spectrometry data analysis, mainly for the detection of posttranslational modifications. Databases store knowledge of carbohydrates attached to proteins as well as protein-carbohydrate interactions.

Highlights 2015

In living organisms, the size of glycan (or sugar, or carbohydrate) molecules varies significantly according, for instance, to the availability of enzymes that synthesize them. Irrespective of their size, the recognition of glycans by other molecules – mainly proteins – is usually limited to a substructure. One of the challenges of glycobiology is to establish which part of a full glycan binds to other molecules, for example surface proteins. Exploring specific substructures can be useful to predict this binding potential. GlyS3 is a new tool that matches a glycan molecule fragment to large collections of full glycan structures as, for instance, those contained in UniCarbKB (www.unicarbkb.org) or GlycomeDB (www.glycome-db.org). The implementation of GlyS3 takes advantage of the latest developments in web semantics, and the tool was integrated into SugarBindDB – a database that collects information on glycan recognition by pathogens.

Over the past years, the group’s production has reached a critical mass warranting the creation of a dedicated tab on the ExPASy server where databases and tools useful for glycomics and glycoproteomics studies are now gathered. These resources are now centraliszed to improve usability as well as to enhance their interconnections. Besides SIB resources, the ExPASy glycomics tab will preferentially host tools that reveal the links between glycans and proteins, or the effect glycosylation has on cellular processes.

Main publications 2015

Alocci D *et al.* Graph Database vs RDF Triple Store: A Comparison on Glycan Substructure Search, PLOS ONE 2015;10(12):e0144578.
Bilbao A *et al.* Ranking fragment ions based on outlier detection for improved label-free quantification in Data-Independent Acquisition LC-MS/MS, J Prot Research 2015;14(11):4581-93.
Horlacher O *et al.* MzJava: an open source mass spectrometry library, J Proteomics 2015;129:63-70.

Christian von Mering

Bioinformatics / Systems Biology Group
University of Zurich

What do they do?

At the Bioinformatics / Systems Biology Group, Christian von Mering and his team study the dynamics of entire biological systems, both at evolutionary time scales and at shorter time scales down to a few minutes. They often work in close collaboration with laboratory scientists, focusing on the computational aspects of studying such systems, in fields ranging from genetics to genomics to proteomics. In addition, they produce and maintain several online resources for the life science community, including STRING-db (protein networks), EGGNOG-db (gene orthology relations), and PAX-db (protein abundances).

Highlights 2015

In 2015, the Bioinformatics/Systems Biology group released a major update of its protein-protein interaction database: STRING. This big update brings the coverage of STRING to over 2,000 model organisms, and more than 900 million interactions are now available. The greatest conceptual advance in this update concerned the scaling of the database – in past versions it was necessary to conduct all-against-all homology searches over the involved proteins in order to execute some of the interaction prediction algorithms in STRING. With the current update the group made to version 10, these homology searches have been replaced by group-orthology data that scales much better and allows the STRING pipeline to transfer interaction knowledge from one model organism to the other. This and other improvements have been very well received by the users, with more than 3,000 distinct users now working with the database on a daily basis.

Main publications 2015

Huerta-Cepas J *et al.* eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. Nucleic Acids Res 2015;44(D1):D286-93.
Szklarczyk D *et al.* STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res 2015;43:D447-52.
Wang MC *et al.* Version 4.0 of PaxDb: protein abundance data, integrated across model organisms, tissues and cell-lines. Proteomics 2015;15(18):3163-8.

Ioannis Xenarios & Lydie Bougueleret
Swiss-Prot Group
University of Geneva

What do they do?
The Swiss-Prot Group develops, annotates and maintains the UniProtKB/Swiss-Prot protein sequence database, the most widely used protein information resource in the world. It also develops and maintains other widely-used specialized databases such as PROSITE, ENZYME, HAMAP, RHEA, and SwissLipids. The group also offers to the virologists' community the ViralZone portal, as well as co-heading the development and maintenance of the ExPASy proteomics website. The Swiss-Prot group is one of the largest groups in the Swiss Institute of Bioinformatics.

Highlights 2015
The group continues to produce and maintain the Swiss-Prot section of the UniProt knowledgebase. During the course of 2015, specific emphasis was given to the curation of human protein variants and association with genetic diseases. Approximately 74,000 human SAPs (Single amino-acid polymorphism) are described in UniProtKB/Swiss-Prot, more than 38,000 of which are enriched by annotations describing involvement in disease and functional characteristics of the variants.
The team continued to develop the SwissLipids knowledgebase for lipid chemistry and biology adding over 100 new lipid classes and 60,000 new lipid structures as well as information and extensive curated links to proteins (UniProtKB) and metabolism (Rhea) sourced from over 400 publications. The group also released a new version of the Rhea website which will form the basis for the development of new tools and interfaces serving the systems modelling communities.
The Swiss-Prot group continues its SIB outreach and educational activities. The web-team plays an active and extensive role in the development of the SIB bioinformatics resource portal ExPASy, while our knowledge resource for virus biology "ViralZone" forms the basis for the development of specialized e-learning courses in partnership with FAO and the Vital-IT team.

Main publications 2015
Aimo L *et al.* The SwissLipids knowledgebase for lipid biology. Bioinformatics 2015;31(17):2860-6.
UniProt Consortium. UniProt: a hub for protein information. Nucleic Acids Res 2015;43(Database issue):D204-12
Hulo C *et al.* A structured annotation frame for the transposable phages: a new proposed family "Saltoviridae" within the Caudovirales. Virology 2015;477:155-63.

Medicine and health

Bioinformatics provides ever-growing support to the field of medicine and health by offering its expertise in many different ways. Drawing on patients' data, bioinformaticians develop tools that help medical practitioners in their decision making.

As an illustration, SIB has developed the algorithm for a non-invasive prenatal test that can detect the most frequent trisomies and chromosomal rearrangements. It also developed models to predict the evolution of brain aneurysms and estimate the dynamics of the Ebola virus during the 2014 outbreak in West Africa.

Michael Baudis

Computational Oncogenomics Group
University of Zurich

What do they do?

Michael Baudis and his team at the Computational Oncogenomics Group focus on the analysis of structural variations in cancer genomes using computational genomics, including bioinformatics and systems biology methods. Their work centres around their collections of molecular tumour data, assembled from genomic screening experiments in cancer e.g. through comparative cytogenetic hybridization (CGH) and genome sequencing studies. Specific projects deal with the development of computational methods for structural data analysis, genomic analyses in selected tumour entities as well as with the large-scale exploration of genomic patterns across malignancies.

Highlights 2015

In 2015, much of the group's activity was focused on advancing projects for the Global Alliance for Genomics and Health (GA4GH). In particular, it contributed to GA4GH schema elements for the exchange of data describing biological and clinical features. Additionally, Michael Baudis and his team designed a GA4GH data implementation project which has been accepted as one of the ELIXIR human data pilot studies and will be executed in 2016. Towards the end of 2015, the group started a new, major data acquisition round for the next version of the arrayMap repository.

Main publications 2015

Muff R *et al.* Genomic instability of osteosarcoma cell lines in culture: impact on the prediction of metastasis relevant genes. PLOS ONE 2015;10(5):e0125611.

Niko Beerenwinkel

Computational Biology Group
D-BSSE, ETH Zurich, Basel

What do they do?

The Computational Biology Group is located in Basel and part of the Department of Biosystems Science and Engineering (D-BSSE) of the ETH Zurich. Niko Beerenwinkel and his team's research and teaching activities are in the areas of computational biology, biostatistics, and systems biology. Their activities include the development of mathematical and statistical models, their implementation in computer programs, and application to biomedical problems. The group is conducting active research projects on HIV drug resistance, the somatic evolution of cancer, haplotype inference from ultra-deep sequencing data, and reconstruction of signalling pathways from RNAi screens.

Highlights 2015

Niko Beerenwinkel and his lab's 2015 highlights include the development and release of: (1) gespeR, a computational method for deconvoluting off-target-confounded RNA interference screens, (2) QuasiFit, a Bayesian MCMC sampler for inferring fitness landscapes from dynamic population equilibrium distributions, and (3) BitPhylogeny, a probabilistic framework for reconstructing intra-tumour phylogenies.

Main publications 2015

Schmich F *et al.* Deconvoluting off-target-confounded RNA interference screens. Genome Biol 2015; 16:220.

Seifert D *et al.* A framework for inferring fitness landscapes of patient-derived viruses using quasispecies theory. Genetics 2015;199(1):191-203.

Yuan K *et al.* BitPhylogeny: a probabilistic framework for reconstructing intra-tumor phylogenies. Genome Biol 2015;16:36.

Mauro Delorenzi

Bioinformatics Core Facility – BCF
University of Lausanne

What do they do?

In the Bioinformatics Core Facility (BCF), Mauro Delorenzi and his team promote trans-disciplinary collaborations between research teams in medicine, molecular biology, genetics, genomics, statistics and bioinformatics. In particular, they perform analysis of biomedical-genomics data with a focus on biomarker studies in cancer research, building on their specific expertise in statistical methods for genomics data analysis. Recently, they concentrated on molecular heterogeneity and pathway activation patterns in cancer subtypes, but they are open to any kind of research directions.

Highlights 2015

In 2015, Mauro Delorenzi and his team investigated the molecular heterogeneity of colon cancer (CC) with the aim of finding information that is useful to assess the expected risk of metastasis and the best way to treat the disease after surgical removal. Useful information consists in predicting the benefit of chemotherapy, if it outweighs its side effects, and in predicting which drug would be more effective.
A first approach consists in a direct statistical analysis of the relationships between one tumour feature and a variable of clinical interest (such as the risk of metastasis for example). In a second approach, the group begins by subdividing the tumours into several groups which differ more clearly from one another by the characteristics of their gene expression patterns (so called tumour subtypes). They then test the usefulness of these groups with respect to the clinical interest. In this way, the team completed a collaborative investigation (Guinney et al. 2015) designed to consolidate previously proposed gene expression subtype systems including their own (Budinska et al. 2013). This investigation classifies primary CC into four major "consensus molecular subtypes" (CMS 1-4). The CMS system will be useful to assess which treatments have high efficacy in which patient groups; to this end, Delorenzi's lab is working with teams that have relevant clinical trial data.

Main publications 2015

Fischer U *et al.* Genomics and drug profiling of fatal TCF3-HLF-positive acute lymphoblastic leukaemia identifies recurrent mutation patterns and therapeutic options. Nat Genet 2015;47(9):1020-9.

Guinney J *et al.* The consensus molecular subtypes of colorectal cancer. Nat Med 2015;21(11):1350-6.

Klingbiel D *et al.* Prognosis of stage II and III colon carcinoma treated with adjuvant 5-FU or FOLFIRI in relation to microsatellite status, results of the PETACC-3 trial. Ann Oncol 2015;26(1):126-32.

Jacques Fellay

Host-Pathogen Genomics Group
EPFL, Lausanne

What do they do?

At the Host-Pathogen Genomics Group, Jacques Fellay and his team explore the genetic roots of inter-individual differences in response to infections, with a particular focus on the genomic interactions between pathogens and their human hosts. At the crossroads between basic science and the clinical world, the group is committed to translational genomic research, aiming at identifying, validating and bringing to clinical use genetic markers of susceptibility to infectious diseases. Host genomics of HIV infection, joint analyses of interactions between human and viral genomes, and exome sequencing on patients with extreme infectious disease phenotypes are some of their important research directions.

Highlights 2015

In 2015, the Host-Pathogen Genomics Group finalized the largest study ever on human genetic factors involved in spontaneous control of HIV infection. The group obtained genome-wide genotyping data from 11,000 HIV infected individuals (25 cohorts), which gave rise to an unprecedented description of the kind of impact common genetic variants can have on HIV disease.
Using a combination of exome and transcriptome sequencing in carefully selected patients, Fellay's team identified rare genetic variants conferring unusual susceptibility to several pediatric infections, including bronchiolitis/ pneumonia caused by respiratory syncytial virus and severe sepsis by *Pseudomonas aeruginosa*.

With colleagues from the EPFL School of Computer Sciences, the group is also developing innovative solutions for genomic privacy – an essential trust-building component on the road towards genomic-based medicine

Main publications 2015

Bartha I *et al.* The Characteristics of Heterozygous Protein Truncating Variants in the Human Genome. PLOS Comp Biol 2015;11(12):e1004647.

Hammer C *et al.* Amino acid variation in HLA class II proteins is a major determinant of humoral response to common viruses. Am J Hum Genet 2015;97(5):738-43.

McLaren PJ *et al.* Polymorphisms of large effect explain the majority of the host genetic contribution to variation of HIV-1 virus load. PNAS 2015;112(47):14658-63.

David Gfeller
Computational Cancer Biology
University of Lausanne

What do they do?
At the Computational Cancer Biology Group, David Gfeller and his group develop algorithmic and modelling tools to analyse large-scale genomics and proteomics data from cancer samples. By integrating different kinds of experimental data, they aim at studying how tumours and tumour microenvironments are organized, with a special focus on infiltrating immune cells. They are also interested in predicting the functional impact of somatic mutations found in cancer cells and, in particular, whether some of these mutations could elicit recognition by the immune system and how this recognition can be exploited towards clinical benefits. To reach this goal, the team develops computational tools from statistics, machine learning and modelling.

Highlights 2015
During 2015, the group developed novel bioinformatics tools to model the interactions between cancer cells and immune cells. One of our areas of research is to predict how well somatic mutations are displayed on cancer cells. A second is to predict how tumours are infiltrated by immune cells and what the prognosis value is of immune infiltration across different tumour types.

Main publications 2015

Gfeller D *et al.* Protein homology reveals new targets for bioactive small molecules. Bioinformatics 2016;31(16):2721-7.

Ivo Kwee
Bioinformatics Core Unit
Institute of Oncology Research, Bellinzona

What do they do?
Our main task at the Bioinformatics Core Unit (BCU) is to support the research groups at the Institute of Oncology Research (IOR) with computational and statistical services. Ivo Kwee and his team's research interests are focused on the genetics and biology of cancer with a major emphasis on lymphomas and epithelial cancers, such as prostate, breast and ovarian cancer. Importantly, more than as just a supporting role, they proactively identify and develop novel bioinformatics projects that can complement and in many cases drive our biologic research. In collaboration with the SIB Swiss Institute of Bioinformatics the group develops innovative data analysis tools, visualization software and database resources for genomics research.

Main publications 2015

Akhmedov M *et al.* A Matheuristic Algorithm for the Prize-collecting Steiner Tree Problem. Proc. of IEEE ICOICT 2015:408–12.
Boi M *et al.* The BET Bromodomain Inhibitor OTX015 Affects Pathogenetic Pathways in Preclinical B-cell Tumor Models and Synergizes with Targeted Drugs. Clin Cancer Res 2015;21(7):1628-38.
Mensah AA *et al.* Novel HDAC inhibitors exhibit pre-clinical efficacy in lymphoma models and point to the importance of CDKN1A expression levels in mediating their anti-tumor response. Oncotarget 2015;6(7):5059-71.

Patrick Ruch
Text Mining Group
Geneva School of Business Administration (HEG)

What do they do?
The SIB Text Mining Group carries carry out activities in semantic and text analytics applied to health and life sciences. Previously hosted by the Radiology and Medical Informatics Department of the University Hospitals of Geneva, the group moved to the University of Applied Sciences Geneva (HES-SO – HEG Geneva) in 2008. Patrick Ruch and his team develop text mining solutions to support both the annotation of SIB databases and the work of a wide range of biomedical professionals, from drug designers to clinicians. The group thus designs, develops and maintains data and web analytic instruments, such as custom search engines, automatic text classifiers and information extraction systems, to help domain experts "make sense" of their biomedical data.

Highlights 2015
Over the course of 2015, Patrick Ruch and his team developed a new resource known as DeepQA4PA: Deep Question-Answering used for Protein Annotation. What they discovered is that Standard Question-Answering cannot answer omics questions because conceptual complexity is too high. As a result, a new type of Question-Answering is needed; i.e. one able to leverage curated molecular biology databases.

In the field of education, the group provides SIB/CUSO Training in Text Mining.

Main publications 2015

Gobeill J *et al.* Deep Question Answering for protein annotation. Database (Oxford) 2015 Sep 16;2015.pii:bav081.
Gobeill J *et al.* Instance-Based Learning for Tweet Monitoring and Categorization. In: Experimental IR Meets Multilinguality, Multimodality, and Interaction. Experimental IR Meets Multilinguality, Multimodality, and Interaction. Springer 2015:235–240.

Christoph Schmid
Computational PathoGenOmics Group
Swiss TPH, Basel

What do they do?
At the Computational PathoGenOmics Group at the Swiss Tropical and Public Health (Swiss TPH) Institute, Christoph Schmid and his team focus their activities on the analysis of data derived from recent high-throughput assays. In collaboration with groups at the Swiss TPH Institute and with external groups, they develop computational methods and apply them to research questions in infection biology and public health. They are involved in projects assessing genome sequences of a variety of pathogenic organisms, epigenetic profiles in prokaryotes and eukaryotes, and gene expression levels in a set of disease models.

Highlights 2015
During the course of 2015, the work on DNA methylation in bacteria was completed in a PhD thesis (Mohamed Sater). New activities on gut metaenomics started in a PhD project (Monica Ticlla) with the support of an SIB PhD fellowship. An institutional agreement was made between the Swiss Tropical and Public Health Institute and the Center for Scientific Computing Computing (sciCORE). The group was also involved in the teaching of DNA methylation and methylome analysis via next generation sequencing (StarOmics & SIB course, 2015).

Main publications 2015

Elkon R *et al.* RFX transcription factors are essential for hearing in mice. Nat Commun 2015 15;6:8549.
Sater MRA *et al.* DNA Methylation Assessed by SMRT Sequencing is Linked to Mutations in Neisseria meningitides Isolates. PLOS ONE 2015;10(12):e0144612.

Evolution and phylogeny

A genome can inform life scientists on how a species has evolved over time. Phylogeny studies the phenotypic and genetic closeness of species, and is illustrated by phylogenetic trees.

Bioinformatics develops tools that are able to read a species' genome, tell the story and build trees. In this way, life scientists have, for example, acquired a better understanding of human migration in the past, the formation of crocodile scales and the evolutionary history of grass.

Maria Anisimova
Applied Computational Genomics Team
Zurich University of Applied Sciences, Wädenswil

What do they do?
At the Applied Computational Genomics Team, Maria Anisimova and her team focus on theoretical and computational aspects of modelling the process of genome evolution and adaptive change. With the growing size and complexity of molecular data, they strive to keep pace, providing accurate, scalable and practical computational solutions that enable a wide range of scientists to analyse patterns of evolution and natural selection in large genomic and -omics data. Their goal is to bring new bioinformatics methods to real applications ranging from biotechnology to biomedical research, ecology and agriculture. They embrace an interdisciplinary approach by integrating different data sources and combining methods.

Highlights 2015
During the course of 2015, Maria Anisimova and her team updated their popular software CodonPhyML, which allows for fast maximum likelihood phylogeny inference from protein-coding genes under codon substitution models. An update of CodonPhyML allows for multi-loci inference and is currently distributed separately as CodonPhyML_multi: https://sourceforge.net/projects/codonphyml/

The team continued to work on statistical methods for detecting tandem repeats in genomic sequences. A library implementing these methods and known as TRAL was released. TRAL, or Tandem Repeat Annotation Library, is a highly modularized, flexible sequence tandem repeat annotation Python2/3 library: http://elkeschaper.github.io/tral/
Anisimova's group is also involved in teaching and organized the Zurich University of Applied Sciences ZHAW-SIB postgraduate summer course "Life in numbers: Genomics and Evolution of Microbial Pathogens" from 24 August to 4 September 2015 in Wädenswil: http://gemp2015.lifeinnumbers.ch. This intensive course gathered students from Switzerland and abroad, and was warmly received.

Main publications 2015
Anisimova M *et al.* Statistical approaches to detecting and analysing tandem repeats in genomic sequences. Frontiers Bioeng Biotech 2015;3:31.
Gil M *et al.* Methodologies for Phylogenetic Inference. Encyclopedia of Life Sciences eLS 2015:1–5.
Zoller S *et al.* Maximum-Likelihood Tree Estimation Using Codon Substitution Models with Multiple Partitions. Mol Biol Evol 2015;32(8):2208-16.

Laurent Excoffier
Computational Population Genetics Group
University of Bern

What do they do?
At the Computational and Molecular Population Genetics (CMPG) Group, Laurent Excoffier and his team are developing new methodologies for the simulation and analysis of molecular polymorphisms within species, with a particular focus on humans. They also develop and maintain the Arlequin software, a popular package for the analysis of multi-locus genetic diversity within and between populations as well as statistical methods to reconstruct and infer evolutionary processes from genomic data. The team focuses on the effect of range expansions on genomic and functional diversity, and the detection of signatures of adaptation and selection at the molecular level.

Highlights 2015
In 2015, Laurent Excoffier and his group continued their study on the effect of spatial range expansions on the functional diversity of a species. Using a mixture of computer simulations and analytical modelling, they demonstrated that populations which expanded their range should present an excess of deleterious mutations due to inefficient purifying selection. This mutation load should then correlate positively with the distance from their place of origin.

The group showed that such a mutation load was indeed present in humans by contrasting the genomic diversity of African and non-African populations. A clear gradient of mutation load from Africa to the Americas was observed, and probably occurred during the migration of modern humans out of Africa about 50,000 years ago.

Main publications 2015
Henn BM *et al.* Distance from Sub-Saharan Africa Predicts Mutational Load in Diverse Human Genomes. PNAS 2015;113:E440-9.
Peischl S *et al.* Expansion load and the evolutionary dynamics of a species range. Am Nat 2015;185:E81-E93.
Peischl S *et al.* Expansion load: recessive mutations and the role of standing genetic variation. Mol Ecol 2015;24:2084-2094.

Gaston Gonnet
Computational Biochemistry Research Group
ETH Zurich

What do they do?
At the Computational Biochemistry Research Group, Gaston Gonnet and his team are interested in the modelling and analysis of biological problems at the molecular level. Their expertise lies in particular in searching algorithms, optimization algorithms, mathematical modelling, and computational systems. Most of their research efforts are concentrated on the Orthologous MAtrix (OMA) project. This particular project aims to produce, automatically, reliable orthologous groups of proteins that are derived from entire genomes. The group offers the results and general services through the internet and through the distribution of the Darwin system for bioinformatics computations.

Highlights 2015
The OMA Browser is a SIB-funded, publicly available resource that provides orthology predictions among publicly available proteomes from all domains of life. During 2015, two new releases of the OMA database were made public, in which the number of covered genomes has steadily increased. The database now covers 1970 (+284) genomes.
The group is very engaged in teaching and supervising the advanced Master's Degree in Computational Biology and Bioinformatics at ETH Zurich.

Main publications 2015

Altenhoff AM *et al.* The OMA orthology database in 2015: function predictions, better plant support, synteny view and other improvements. Nucl Acids Res 2015;43(D1):D240-D249.
Ravenhall M *et al.* Inferring horizontal gene transfer. PLOS Comput Biol 2015;11(5):e1004095.
Škunca N *et al.* Phylogenetic profiling: how much input data is enough? PLUS ONE 2015;10 (2):e0114701.

Jérôme Goudet
Population Genetics and Genomics Group
University of Lausanne

What do they do?
At the Population Genetics and Genomics Group, Jérôme Goudet and his team's interest is focused on understanding how the interplay of population structure, trait architecture and selection can be disentangled. To this end, they use different approaches, from theory and the development of statistical tools to field observations. The main biological models currently used are the barn owl and Miniopterus bats. On the theoretical side, the group investigates the dynamics of multilocus genetic systems under the influence of selection, migration and drift, develop comprehensive individual-based models as well as statistical methods to infer selection, mating systems and population structure.

Jeffrey D. Jensen
Population Genetics Group
EPFL, Lausanne

What do they do?
Jeffrey Jensen and his team's primary research theme at the Population Genetics Group is centred around drawing statistical inference from DNA polymorphism data – specifically, describing the processes that determine the amount and distribution of genetic variation within and between natural populations, and between species. The group works on both applied and theoretical problems in fields ranging from population genomics to medical genetics. They focus on developing statistical methodology to infer the parameters of positive selection for specific sites in the genome, as well as on characterizing the full distribution of fitness effects of all new, segregating and fixed mutations in the genome.

Highlights 2015
In 2015, the Jensen Lab focused on the development of statistical inference methods for time-sampled population data, with primary applications to the evolution of drug resistance and the demographic history of infection characterizing human cytomegalovirus (HCMV) and influenza virus (IAV).

Main publications 2015

Foll M *et al.* A Wright-Fisher ABC-based approach for inferring per-site effective population sizes and selection coefficients from time-sampled data. Mol Ecol Resour 2015;15:87-98.
Montano V *et al.* Worldwide population structure, long term demography, and local adaptation of Helicobacter pylori. Genetics 2015;200:947-63.
Renzette N *et al.* Limits and patterns of cytomegalovirus genomic diversity in humans. Proc Natl Acad Sci USA 2015;112:E4120-28.

Bernard Moret
Laboratory for Computational Biology and Bioinformatics
EPFL, Lausanne

What do they do?
At the Laboratory for Computational Biology and Bioinformatics, Bernard Moret and his team develop, implement and assess models and algorithms for genome and network evolution. They conduct foundational research in optimization as well as in stochastic models and associated algorithms. Areas of particular interest include reticulate evolution, genomic rearrangements, gene families, phylogenetic multiple sequence alignments, as well as large-scale, high-performance phylogenetic reconstruction. Software produced in the course of their research is available to the community in source form on their website.

Highlights 2015
During the course of 2015, the group developed exact, yet scalable, algorithms to compute optimal solutions to genomic rearrangement problems in the event of duplication or loss of entire regions.

Main publications 2015

Shao M *et al.* A fast and exact algorithm for the exemplar breakpoint distance. Proc. 19th Int'l Conf. on Research in Comput. Molecular Bio. RECOMB'15, in Lecture Notes in Computer Science, Springer Verlag 2015;9029:309-322.
Shao M *et al.* Comparing genomes with rearrangements and segmental duplications. Proc. 23rd Symp. on Intelligent Systems for Mol. Bio. ISMB'15, in Bioinformatics 2015;31(12):i329-i338.

Marc Robinson-Rechavi
Evolutionary Bioinformatics Group
University of Lausanne

What do they do?
At the Evolutionary Bioinformatics Group, Marc Robinson-Rechavi and his team are mainly concerned with determining the role of evolutionary innovation and constraint in animals. For this, they develop methods and databases to extract reliable information from genome and transcriptome data. These databases include Bgee, a database for gene expression evolution, Selectome, a database of positive selection. In developing these resources, they also conduct research on ontologies, biocuration, and high-performance computing. Their biological focus is to link Evo-Devo with phylogenomics. They notably study the role of gene duplication in divergence between genes and between species.

Highlights 2015
During the course of 2015, Marc Robinson-Rechavi and his team released new tools to analyse gene expression data in animals: 1) TopAnat, a GO-like enrichment of anatomical terms mapped to genes by expression patterns (bgee.org/?page=top_anat#/), and 2) Processed RNA-seq and microarray data for 17 species in Bgee. TopAnat is the first tool to provide a way of characterizing expression patterns of gene lists automatically.

The group also participated in writing two papers resulting from the master's degree course "Sequence a genome", a course during which master's degree students sequence, assemble and annotate new bacterial genomes. This course is a highlight of the Molecular Life Sciences master's degree at UNIL, and the papers are based directly on the students' work during the class: dx.doi.org/10.3389/fmicb.2015.00101 and dx.doi.org/10.1111/1462-2920.12498.

Main publications 2015

Bastian F.B *et al.* The Confidence Information Ontology: a step towards a standard for asserting confidence in annotations. Database (Oxford) 2015;2015:bav043.

Daub JT *et al.* Inference of Evolutionary Forces Acting on Human Biological Pathways. Genome Biol Evol 2015;7(6):1546-1558.

Kryuchkova-Mostacci N *et al.* Tissue-Specific Evolution of Protein Coding Genes in Human and Mouse. PLOS ONE 2015;10(6):e0131673.

Nicolas Salamin
Computational Phylogenetics Group
University of Lausanne

What do they do?
The Computational Phylogenetics Group develops software to understand the evolutionary history between organisms, and test macroevolutionary hypotheses. They study the ecological, genomic and morphological factors that limit and constrain speciation and adaptation. Amongst other subjects, the group also focuses on phylogenetic reconstruction methods, clownfish and plant genomics, the mode and tempo of species evolution and the spatially explicit evolution of diversity. The group's aim is to improve models to analyse sequence data, and quantitative models to estimate macroevolutionary patterns and processes.

Highlights 2015
The group is developing new ways to estimate the rate of species evolution by using complex Bayesian approaches. These developments are important to understand the factors that influence the emergence and extinction of species over time. The method was implemented into a python software called pyRate. The models implemented in pyRate were extended by developing a novel Bayesian approach that can estimate the standard heterogeneous birth-death process on fossil data. The advantage of such an approach is that the birth-death process can be directly compared with estimates done on phylogenetic trees, and can provide information on the divergence times of the nodes in a phylogenetic tree estimated by molecular data.

Such an approach is very flexible and the group incorporated several models to fully account for the heterogeneity in the tempo of species evolution. This allows for shifts of speciation and extinction rates during species evolution, but also the association with external factors such as climate variation, or competition with other clades.

The models fully complement existing approaches and are currently used to estimate the evolution of large groups of mammals such as canids, or the origins of flowering plants.

Main publications 2015

Dib L *et al.* Coev-web: a web platform designed to simulate and evaluate coevolving positions along a phylogenetic tree. BMC Bioinformatics 2015;16:394.

Laurent S *et al.* Detecting patterns of species diversification in the presence of both rate shifts and mass extinctions. BMC Evol Biol 2015;15:157.

Silvestro D *et al.* Measurement errors should always be incorporated in phylogenetic comparative analysis. Methods in Ecology and Evolution 2015;6(3):340-346.

Tanja Stadler
Computational Evolution Group
D-BSSE, ETH Zurich, Basel

What do they do?
At the Computational Evolution Group, Tanja Stadler and her team develop phylogenetic tools in order to understand evolutionary processes. Using their phylogenetic methods, they aim to improve their understanding of past evolutionary and population dynamic processes on different scales. They address questions in a number of fields, focusing on epidemiology, public health and medicine, ecology and evolution, and language evolution. In their daily work, they define and analyse stochastic models, implement computational methods, analyse empirical data, and discuss their new insights with clinicians, public health policy makers, as well as ecologists and palaeontologists.

Highlights 2015
In 2015, the group took a big step in bridging part of the gap between molecular evolution and palaeontology. It developed tools to integrate data sources from both fields in order to reconstruct the tree of life and assess the macroevolutionary processes which give rise to the present day species. These advances were discussed during a two-day meeting at the Royal Society in London in November. As this document goes into print, a paper is in press and two manuscripts are under revision on the topic. Furthermore, the *Neue Zürcher Zeitung* also featured our work on the evolution of penguins last December. To develop the area further, the group will be welcoming a new postdoc on an ETH fellowship grant.

Main publications 2015

Du Plessis L *et al.* Getting to the root of epidemic spread with phylodynamic analysis of genomic data. Trends Microbiol 2015;23(7):383-6.

Hagen O *et al.* Age-dependent speciation explains empirical tree shape distribution. Syst Biol 2015;64(3):432-40.

Stadler T *et al.* How well can the exponential-growth coalescent approximate constant rate birth-death population dynamics? Proc Roy Soc B 2015;282:20150420.

Andreas Wagner
Evolutionary Systems Biology Group
University of Zurich

What do they do?
At the Evolutionary Systems Biology Group, Andreas Wagner and his team study the evolution and evolvability of biological systems at all levels of biological organization, from genes and genomes to biological networks and whole organisms. They develop bioinformatics tools to integrate data from a variety of sources, including comparative whole-genome sequence data, microarray expression data, and high-throughput protein interaction data. Their work uses comparative analysis of genomic data, laboratory evolution experiments and mathematical modelling. They also develop a variety of bioinformatics tools to help them deal with the torrent of data in genomics and structural biology.

Main publications 2015

Bilgin Sonay T *et al.* Tandem repeat variation in human and great ape populations, and its impact on gene expression divergence. Genome Res 2015;25(11):1591-9.

Bratulic S *et al.* Mistranslation drives the evolution of robustness but not translational accuracy in TEM-1 β-lactamase. Proc Natl Acad Sci 2015;112(41):12758-63.

Hayden E *et al.* Intramolecular phenotypic capacitance in a modular RNA molecule. Proc Natl Acad Sci 2015;112(40):12444-9

Hosseini S-R *et al.* Exhaustive analysis of a genotype space comprising 1015 central carbon metabolisms reveals an organization conducive to metabolic innovation. PLOS Comput Biol 2015;11(8):e1004329.

Daniel Wegmann
Statistical and Computational Evolutionary Biology Group
University of Fribourg

What do they do?
When observing nature, one is easily impressed by the huge diversity seen at any biological scale. Daniel Wegmann and his team's primary aim at the Statistical and Computational Evolutionary Biology Group is to better understand the underlying evolutionary and ecological processes that have been shaping this diversity over the course of evolution on our planet. To achieve this, they design and evaluate new statistical and computational approaches to infer complex evolutionary histories. For this they develop and apply machine learning algorithms, with a particular focus on likelihood-free methods. They then apply these approaches to the wealth of data currently being generated, mostly in collaboration with experimental groups. They are further committed to making all our developments available to the scientific communities by releasing easy-to-use software packages.

Highlights 2015
Through methodological advances, the retrieval of DNA sequences from ancient bones has become an invaluable tool to study the prehistory of humans and other organisms. However, DNA obtained from very old samples show peculiar characteristics referred to as Post Mortem Damage (PMD).

Daniel Wegmann and his group have been particularly interested in understanding how to incorporate PMD into population genetic analysis. For instance, they developed a novel variant caller to infer accurately the genotypes of ancient samples, and found ways to infer accurately the level of genetic diversity from such data – even when the total amount of data is very low.
With Professor Burger in Mainz, the group applied these methods to learn more about how farming spread across prehistoric Europe. They found that the early farmers from the Aegean region are direct ancestors of the early farmers in western Europe thus prompting that farming entered Europe through Anatolia and spread predominantly as farmers colonized Europe.

Main publications 2015

Bruford MW *et al.* Prospects and challenges for the conservation of farm animal genomic resources, 2015-2025. Front Genet 2015;6:314.
Ferrer-Admetlla *A et al.* An approximate Markov model for the Wright-Fisher diffusion. Genetics 2015;(in press).

Structural biology

Biological macromolecules such as DNA and proteins acquire a specific architecture in space. The 3D conformation they adopt is a direct consequence of their nucleic acid or amino acid sequence, respectively. A protein's function is defined by its 3D structure.

Bioinformatics develops software that is able to model and predict a protein's 3D structure, and hence deduce its probable function. Such tools are of great assistance in the field of drug design, for instance.

Simon Bernèche
Computational Biophysics Group
University of Basel

What do they do?
At the Computational Biophysics Group, Simon Bernèche and his team are interested in the structure-function relationship of membrane proteins. Using molecular mechanics simulations and statistical approaches, the group aims at understanding the microscopic mechanisms underlying the functions of proteins involved in the membrane transport of various substrates to discover how the functions of proteins emerge from their 3D structure. A central topic of study concerns the elucidation of gating mechanisms regulating ion permeation and the activity of potassium channels in excitable cells, and the resulting impact on neuron signalling. Other subjects of interest involve transport mechanisms that are ATP-dependent or proton-coupled, and the mechanisms of protein folding.

Highlights 2015
Thanks to the latest generation of molecular mechanics force field, Simon Bernèche and his team were able to elucidate molecular mechanisms on a scale that is barely accessible by experimental approaches. For example, based on crystallographic and functional data, the proteins of the Rh family were described as ammonia channels which allow NH_3 to diffuse down its gradient. The group's free energy simulations revealed that Rh proteins actually recruit NH_4^+, which can then diffuse as NH_3 once a proton has been transferred to a conserved histidine. Recruitment of NH_4^+ is more efficient than that of NH_3 by orders of magnitude because of its higher abundance and binding affinity, and thus explains better the measured transport rates. Their simulations revealed mechanisms which provide a new perspective on the structural data and allow for a more rigorous interpretation of the functional data.

Main publications 2015

Baday S *et al.* Mechanism of NH_4^+ Recruitment and NH3 Transport in Rh Proteins. Structure 2015;23:1550-7.
Bignucolo O *et al.* Backbone Hydration Determines the Folding Signature of Amino Acid Residues. J Am Chem Soc 2015;137(13):4300-3.
Xu Y *et al.* Allocrite sensing and binding by the breast cancer resistance protein (ABCG2) and P-glycoprotein (ABCB1). Biochemistry 2015;54(40):6195-206.

Matteo Dal Peraro
Laboratory for Biomolecular Modelling
EPFL, Lausanne

What do they do?
Matteo Dal Peraro and his team's main goal at the Laboratory for Biomolecular Modeling is to understand the physical and chemical properties of complex biological systems, in particular their function with regard to structure and dynamics. To this end, they use and develop a broad spectrum of computational tools fully integrated with experimental data. Multiscale simulations and dynamic integrative modelling are used to investigate the function of molecular assemblies, mimicking conditions of the native cellular environment.

Highlights 2015
The physical and chemical characterization of biological membranes is of fundamental importance for understanding the functional role of lipid bilayers in shaping cells and organelles, steering vesicle trafficking and promoting membrane-protein signalling. Molecular dynamics simulations stand as a powerful tool to probe the properties of membranes at the atomistic level. However, the biological membrane is highly complex, and closely mimicking its physiological constitution *in silico* is not a straightforward task.

In 2015, Matteo Dal Peraro and his team introduced *LipidBuilder*, a framework for creating and storing models of biologically relevant phospholipid species with acyl tails of heterogeneous composition. *LipidBuilder* also enables the assembly of these database-stored lipids into realistic bilayers featuring asymmetric distribution on layer leaflets and concentration of given membrane constituents as defined, for example, by lipidomics experiments. *LipidBuilder* is a powerful tool to model biological membranes of near-biological complexity, and a robust complement to the current efforts to characterize the biophysical properties of biological membranes using molecular simulation.

Main publications 2015

Abriata LA *et al.* Assessing the potential of atomistic molecular dynamics simulations to probe reversible protein-protein recognition and binding. Sci Rep 2015;5:10549.
Bovigny C *et al.* LipidBuilder: A Framework To Build Realistic Models for Biological Membranes. J Chem Inf Model 2015;55(12):2491-9.
Pezeshgi Modarres H *et al.* Understanding and Engineering Thermostability in DNA Ligase from Thermococcus sp. 1519. Biochemistry 2015;54(19):3076-3085.

Olivier Michielin & Vincent Zoete
Molecular Modelling Group
University of Lausanne

What do they do?
At the Molecular Modelling Group (MMG), Olivier Michielin, Vincent Zoete and their team study mechanisms of molecular recognition in particular protein-protein or protein-small ligand interactions, using molecular modelling techniques such as homology modelling, molecular dynamics, docking and free energy simulations. Their main activity consists in the development and application of state-of-the-art methods in computer-aided protein engineering and drug design. Most efforts are concentrated on the development of new small molecule inhibitors of important targets for cancer therapy, as well as the design of optimized proteins like T cell receptor (TCR), for cancer immunotherapy. The group develops and maintains several web tools for drug design (see below). They also act as the Protein Modeling Facility (PMF) of the University of Lausanne.

Highlights 2015
Molecular docking predicts the position of ligands in the binding sites of macromolecules, and constitutes the cornerstone of structure-based drug design. In 2015, the MMG developed a new algorithm for docking – *Attracting Cavities* – which transiently replaces the rough potential energy of the protein by a smooth attracting potential that drives the ligands into protein cavities. The approach achieved a success rate of 80% in reproducing experimental binding modes. It will join the team's SwissDock web interface for docking.

During the course of 2015, the group developed SwissSimilarity, a new web tool for rapid ligand-based virtual screening. Screenable compounds include drugs, bioactive and commercial compounds, as well as an unprecedented ultra-large library of 205 million virtual compounds that are readily synthesizable. Predictions can be carried out using five different approaches. SwissSimilarity is part of a large SIB initiative to provide online tools for drug design – such as SwissDock, SwissBioisostere and SwissTargetPrediction with which SwissSimilarity can interoperate.

Main publications 2015

Gfeller D *et al.* Protein homology reveals new targets for bioactive small molecules. Bioinformatics 2015;31(16):2721-7.
Röhrig UF *et al.* Challenges in the Discovery of Indoleamine 2,3-Dioxygenase 1 (IDO1) Inhibitors. J Med Chem 2015;58(24):9421-37.
Zoete V *et al.* Attracting cavities for docking. Replacing the rough energy landscape of the protein by a smooth attracting landscape. J Comput Chem 2016;37(4):437-47.

Torsten Schwede
Computational Structural Biology Group
University of Basel

What do they do?
At the Computational Structural Biology (CSB) Group, Torsten Schwede and his team focus on the development of methods and algorithms to model, simulate and analyse 3D protein structures and their molecular properties in order to apply these techniques to understanding biological processes at a molecular level. Their main emphasis is on homology modelling approaches – using evolutionary information to model protein tertiary and quaternary structures. Applications in biomedical research include the study of protein-ligand interactions from different perspectives, such as the identification of small antiviral molecules to support drug development, the structure-guided engineering of enzymes or the interpretation of disease causing mutations in proteins.

Highlights 2015
In a collaboration with the Swiss Federal Laboratories for Materials Science and Technology, and GlycoVaxyn AG (Schlieren, Switzerland), the CSB group successfully engineered the specificity of a recombinant bacterial enzyme for the production of novel conjugate vaccines in *E.coli*. Computational modelling of the protein structure in complex with potential substrates provided the basis for the rational engineering of the enzyme's properties.

In June, the group organized the 12th edition of the Basel Computational Biology Conference: The meeting brought together over 500 scientists from a broad range of disciplines. The next $[BC]^2$ Conference is planned for September 2017.

In November, the team presented a novel user interface experience for exploring molecular interactions at the Swiss Innovation Forum in Basel – the national platform for promoting innovation, creativity and design. Transforming the molecular forces resulting from such simulations into a physical force, which can be perceived using a haptic feedback device, provides a new and innovative user interface experience.

Main publications 2015

Kryshtafovych A *et al.* Methods of model accuracy estimation can help selecting the best models from decoy sets: Assessment of model accuracy estimations in CASP11. Proteins 2015;doi:10.1002/prot.24919.
Sali A *et al.* Outcome of the First wwPDB Hybrid/Integrative Methods Task Force Workshop. Structure: with folding and design 2015;23(7):1156-67.
Ihssen J *et al.* Increased efficiency of Campylobacter jejuni N-oligosaccharyltransferase PglB by structure-guided engineering. Open Biol 2015;5(4):140227.

Systems biology

No biological macromolecule – nor living being – works on its own but interacts with many others which, in turn, interact with others. This is the field of systems biology.

Bioinformatics develops mathematical models that can illustrate such systems and even address their evolution in time. Such tools can help to delineate metabolic pathways, for instance, or predict what could happen if a given species is introduced into a pre-existing ecological system.

Bastien Chopard
Scientific and Parallel Computing Group
University of Geneva

What do they do?
At the Scientific and Parallel Computing (SPC) Group, Bastien Chopard and his team develop new algorithms and methods to better understand and/or predict various phenomena in Biology. They focus on multiscale modelling and computing, high-performance computing, cellular automata, lattice Botzmann methods, multi-agent systems, and optimizing techniques and machine learning. A core activity of their group is the modelling and simulation of complex systems. In bioinformatics, the objective is to develop advanced numerical methodology to model biological processes.

Highlights 2015
In 2015, Bastien Chopard and his team carried out research activities that focused on several issues: the modelling of epithelium, new MCMC methods for phylogeny, experimenting and modelling platelet adhesion and aggregation where new bio-physical phenomena have been evidenced.
The group also developed activities outside the field of bioinformatics within the European project Sophocles, in which information processing in complex systems was studied. Within the PACS national project, the group also developed multiscale HPC algorithms and software to describe the transport and sedimentation of volcanic ashes close to and far from a volcano.

Main publications 2015
Ribeiro de Sousa D *et al.* Determination of a wall shear rate threshold for thrombus formation in intracranial aneurysms. J Neuro Interv Surg 2015;doi: 10.1136/neurintsurg-2015-011737.
Zouaoui Boudejltia K *et al.* Quantitative analysis of platelets aggregates in 3D by Digital Holographic Microscopy, Biomed Opt Express 2015;6(9):3556-63.

Manfred Claassen
Computational Single Cell Biology Group
ETH Zurich

What do they do?
Manfred Claassen and his team's research at the Computational Single Cell Biology Group aims at elucidating the composition of heterogeneous cell populations and how these implement function in the context of cancer and immune biology. To accomplish this task the group builds on concepts from statistics, machine learning and mathematical optimization to develop probabilistic approaches to describe biological systems, learn these descriptions from data, and design experiments to validate hypotheses following from computational analyses. Their research can be used to pinpoint therapeutic targets with a perspective to designing drugs.

Highlights 2015
During the course of 2015, Manfred Claassen and his team developed CellCnn – a sensitive means to detect rare disease-associated cell subsets via representation learning. The team also developed Reactionet Lasso: structure learning for stochastic reaction networks and STILT, a particle filter based Bayesian model selection approach for single cell time-lapse imaging experiments. Claassen and his group also participate in an SIB Computational Biology block-course in Lausanne.

Main publications 2015
De Vargas Roditi L *et al.* Computational and experimental single cell biology techniques for the definition of cell type heterogeneity, interplay and intracellular dynamics. Curr Opin Biotechnol 2015;34:9-15.

Rudiyanto Gunawan

Chemical and Biological Systems Engineering
Group / ETH Zurich

What do they do?

At the Chemical and Biological Systems Engineering Laboratory, Rudiyanto Gunawan and his team develop tools for systems modelling and analysis of chemical and biological networks. Their mission is to create enabling theories and computational methods for generating systems insights, understanding and knowledge in chemical, biological and medical applications. Their research spans multiple length and time scales of cell biology, from gene/signalling/metabolic networks in single cells to the ageing process in human and cell culture bioreactors in the pharmaceutical industry.

Highlights 2015

During the course of 2015, Rudiyanto Gunawan and his team released two tools, both of which were described in two separate publications in the journal *Bioinformatics*:

1. REDEMPTION (Reduced Dimension Ensemble Modeling and Parameter Estimation). REDEMPTION provides an integrated MATLAB platform for ensemble modelling and parameter estimation using ordinary differential equations;
2. REDUCE (Reduction of Uncertain Edges). REDUCE is an optimal design of gene knock-out strategy for gene regulatory network inference.

Furthermore, the team's analysis of lifespans and mitochondrial DNA (mtDNA) sequences of mammalian and avian species showed that in contrast to a long-held hypothesis, the number of direct repeats (DRs) is not a specific, evolutionarily selected feature of mtDNA. Instead, DRs form as an indirect consequence of bias in nucleotide composition and in synonymous codon usage.

Main publications 2015

Lakshmanan LN *et al.* Are mutagenic non D-loop direct repeat motifs in mitochondrial DNA under a negative selection pressure? Nucleic Acids Res 2015;43(8):4098-208.

Liu Y *et al.* REDEMPTION: Reduced Dimension Ensemble Modeling and Parameter Estimation. Bioinformatics 2015;31(20):3387-9.

Ud-Dean SMM *et al.* Optimal design of gene knock-out experiments for gene regulatory network inference. Bioinformatics 2015;doi:10.1093/bioinformatics/btv672.

Vassily Hatzimanikatis

Laboratory of Computational Systems Biotechnology
EPFL, Lausanne

What do they do?

At the Laboratory of Computational Systems Biotechnology, Vassily Hatzimanikatis and his team work at the interface of biology, process systems engineering, and reaction engineering. They are developing expertise in the mechanistic modelling of chemical and biophysical cellular processes and in the application of systems engineering methods to the analysis of these models. Their objectives are to develop approaches that will provide guidance in the study of problems in basic and applied biology and medicine, and that will accelerate discovery and development in pharmaceutical and industrial biotechnology. Their aim is to provide experimentally testable hypotheses and targets for purposeful redesign and manipulation of these processes.

Highlights 2015

The group continued to develop a computational framework for the atom-level reconstruction of metabolic networks from *in silico* labelled substrates, which allows tracking the atoms' fate through the reconstructed metabolic network. The method was applied for the reconstruction of an atom-level representation of a core metabolic network of *E. coli*.

Even for well-characterized model organisms like *E. coli*, new reactions remain to be discovered. Starting from the known biochemistry of *E. coli* metabolism, the group used BNICE.ch (Biochemical Network Integrated Computational Explorer) – a computational framework for the identification of novel metabolic reactions and pathways – to search for novel reactions.

Indeed, BNICE.ch applies known biotransformation rules to generate a "super" network which captures all possible reactions, given a set of *E. coli* core metabolites and known biotransformation rules present in *E. coli*. This super network captures all the known *E. coli* reactions as well as novel pathways that can serve as potential novel biosynthesis pathways for valuable chemicals.

Main publications 2015

Ataman M *et al.* Heading in the right direction: thermodynamics-based network analysis and pathway engineering. Curr Opin Chem Biol 2015;36:176-182.

Hadadi N *et al.* Design of computational retrobiosynthesis tools for the design of de novo synthetic pathways. Curr Opin Chem Biol 2015;28:99–104.

Tymoshenko S *et al.* Metabolic Needs and Capabilities of Toxoplasma gondii through Combined Computational and Experimental Analysis. PLOS Comput Biol 2015;11(5):e1004261.

Dagmar Iber

Computational Biology Group
D-BSSE, ETH Zurich, Basel

What do they do?

At the Computational Biology Group, Dagmar Iber and her team develop quantitative, predictive models of biological signalling networks with a view to gaining a comprehensive understanding of the dynamics and evolution of cellular signalling. Higher forms of life emerge from a more sophisticated use of often conserved signalling pathways to regulate biological function. The complex behaviour of the resulting cellular signalling networks is impossible to grasp by verbal models alone. Quantitative, computational models are required to integrate biological knowledge into a framework that permits the efficient generation of testable hypotheses and that enables an integrative understanding of biological networks.

Highlights 2015

In 2015, Dagmar Iber's group published the LBIBCell software (Tanaka et al., Bioinformatics, 2015). LBIBCell enables the detailed simulation of tissue development at cellular resolution; the tissue is described by a Navier-Stokes fluid and the biophysical details of the tissue's dynamics can be coupled with the PDE models for the regulatory networks. Iber's team also collaborated with the SBML developers on MOCCASIN to automate the conversion of MATLAB ODE models into SBML (Gomez et al., Bioinformatics, in press).

On the modelling side, the Iber group defined a mechanism that enables the definition of relative patterns on growing embryonic domains, and provided evidence for an evolutionary highly conserved mechanism that defines the dorsal-ventral axis.

Main publications 2015

Fried P *et al.* Read-Out of Dynamic Morphogen Gradients on Growing Domains. PLOS ONE 2015;e0143226.

Genikhovich G *et al.* Technau U. Axis patterning by BMPs: cnidarian network reveals evolutionary constraints. Cell Rep 2015 Mar 11;pii:S2211-1247(15)00181-3.

Tanaka S *et al.* LBIBCell: A Cell-based Simulation Environment for Morphogenetic Problems. Bioinformatics 2015;31:2340-7.

Christian Mazza

Biomathematics and Computational Biology Group
University of Fribourg

What do they do?

At the Biomathematics and Computational Biology Group, Christian Mazza and his team are a small team located at the Department of Mathematics of the University of Fribourg. The field of mathematics can provide models to the life science community to achieve a greater understanding of how a given biological system evolves in time with respect to the many interactions of a different nature which exist within an organism. The group studies biological networks, complex ecosystems and mathematical models of plant growth by focusing on both their geometrical structure (graphs, patterns) and their underlying dynamics (deterministic and stochastic). Typical examples are Lotka-Volterra dynamics on complex ecological networks and cellular processing systems.

Main publications 2015

Chrystel Feller *et al.* Self-Organization of Plant Vascular Systems: Claims and Counter-Claims about the Flux-Based Auxin Transport Model. PLOS ONE 2015;10(3):e0118238.

Michaël Dougoud *et al.* Ultrasensitivity and Sharp Threshold Theorems for Multisite Systems. ArXiv:1511.03150[q-bio.SC].

Michel Milinkovitch

Artificial & Natural Evolutionary Development of Complexity Group / University of Geneva

What do they do?

For more than 10 years, Michel Milinkovitch and his team's core activities at the Artificial & Natural Evolutionary Development of Complexity Group revolved around the production of experimental data and the development of tools and algorithms in Evolutionary Genetics. Since 2008, they have additionally combined evolutionary developmental biology (EvoDevo) and the study of physical processes to understand the mechanisms generating complexity and diversity in the living world. The group specializes in non-classical model species in reptiles and mammals and integrates data and analyses from comparative genomics, molecular developmental genetics, as well as computer modelling and numerical simulations.

Highlights 2015

During the course of 2015, Michel Milinkovitch and his lab:

- sequenced the corn snake genome;
- established the Reptilian transcriptome database v2.0;
- discovered the physical process which underlies colour change in chameleons;
- developed *R2OBBIE-3D* (a Fast Robotic High-Resolution System for Quantitative Phenotyping of Surface Geometry and Colour-Texture);
- and completed the first linkage mapping and characterization of a colour mutation in the corn snake.

Several of these highlights are illustrated on videos: goo.gl/RPXfaH

Main publications 2015

Saenko S.V *et al.* Amelanism in the corn snake is associated with the insertion of an LTR-retrotransposon in the OCA2 gene. Sci Rep 2015;5:17118.

Teyssier J *et al.* Photonic Crystals Cause Active Colour Change in Chameleons. Nat Commun 2015;6:6368.

Tzika AC *et al.* Reptilian Transcriptomes v2.0: An Extensive Resource for Sauropsida Genomics and Transcriptomics Genome Biol Evol 2015;7(6):1827-41.

Félix Naef

Computational Systems Biology EPFL, Lausanne

What do they do?

The aim of systems biology is to achieve a quantitative and dynamic understanding of cellular networks by combining experimental data with theoretical and computational methodologies. At the Computational Systems Biology Group, Félix Naef and his team's interest lies in the regulatory and cellular networks involved in oncogenic signalling, cell-cycle regulation, and molecular oscillators. Data obtained from technologies such as microarrays, chromatin-immunoprecipitation (ChIP) and genome sequencing are brought together to discover regulatory dependencies between genes and regulatory proteins involved in cell proliferation. One thematic focus is the study of biomolecular oscillators, in particular the circadian clock.

Highlights 2015

The group's major highlights in 2015 included 1) the analysis of translation initiation rates as a function of time in the mouse liver, using ribosome footprinting and 2) the analysis of transcriptional fluctuations in individual mammalian cells, which allowed the group to estimate the structure and number of rate limiting steps in promoter cycles.

Main publications 2015

Atger F *et al.* Circadian and feeding rhythms differentially affect rhythmic mRNA transcription and translation in mouse liver. Proc Natl Acad Sci 2015;112(47).

Blanchoud S *et al.* CAST: An automated segmentation and tracking tool for the analysis of transcriptional kinetics from single-cell time-lapse recordings. Methods 2015;85:3-11.

Zoller B *et al.* Structure of silent transcription intervals and noise characteristics of mammalian genes.Mol Syst Biol 2015;11(7):823.

Igor V. Pivkin

Scientific Computing Group
Università della Svizzera italiana, Lugano

What do they do?

Igor Pivkin and his team's research interests at the Scientific Computing Group lie in the area of multiscale/multiphysics modelling and parallel large-scale simulations of biological systems. They focus on the development of new computational models and corresponding numerical methods suitable for the next generation of super computers. The group is working on stochastic multiscale modelling of motion, the interaction, deformation and aggregation of cells under physiological flow conditions, biofilm growth, coarse grained molecular dynamics simulations, as well as the modelling of transport processes in healthy and tumour-induced microcirculation.

Highlights 2015

Computer simulations on an atomistic scale are suitable for studying biomolecular systems and can be used to complement experiments for exploring free energy landscapes and binding free energies. However, until now, all-atom approaches have been limited in their timescale and can be computationally expensive. Igor Pivkin and his group developed a new coarse-grained force field for describing proteins on the basis of the Dissipative Particle Dynamics (DPD) method, thus further extending the scope of applications of DPD to simulations of biomolecular systems. Their new model is based on the electrostatic polarization of the protein backbone and a detailed representation of the sidechains in combination with a polarizable water model. The model is suitable for the coarse-grained description of proteins and can improve the sampling of native states in coarse-grained protein simulations.

Main publications 2015

Christel S *et al.* Systems Biology of Acidophile Biofilms for Efficient Metal Extraction. Advanced Materials Research 2015;1130:312-315.

Lykov K *et al.* Inflow/outflow boundary conditions for particle-based blood flow simulations: application to arterial bifurcations and trees. PLOS Comput Biol 2015;11(8):e1004410

Peter E *et al.* A polarizable coarse-grained protein model for dissipative particle dynamics. Phys Chem Chem Phys 2015;17(37):24452-24461.

Jörg Stelling

Computational Systems Biology Group
D-BSSE, ETH Zurich, Basel

What do they do?

At the CSB Group at ETH Zurich in Basel, Joerg Stelling and his team develop and apply computational and – most recently – experimental methods to analyse and design complex cellular networks, with a focus on large-scale mechanistic approaches. The group comprises biologists, computer scientists, engineers, and mathematicians who perform interdisciplinary research in systems and synthetic biology. They focus on developing and applying computational methods and mechanistic mathematical models to study complex cellular networks, to elucidate their operating principles, and to enable their rational re-design. Their biological applications rely on the group's experimental biology section that uses budding yeast as a model organism, and on various external collaborations.

Highlights 2015

Highlights in 2015 for the Joerg Stelling lab concern research on the development of computational methods for model-based systems analysis as well as new software helping to ensure reproducible science in academic laboratories. Novel methods address key challenges in systems biology, namely uncertainty quantification (efficiently assessing how perturbations affect a system's behaviour using advanced numerical methods), and the integration of heterogeneous data types (where the group proposed a novel approach based on network motifs and validated it in a case study on nutrient signalling). Collaborations with Bernd Rinn's group (SIB) led to the development of an electronic laboratory notebook and a laboratory information management system (ELN-LIMS) for the open-source platform openBIS.

Main publications 2015

Barillari C *et al.* openBIS ELN-LIMS: an open-source database for academic laboratories. Bioinformatics 2016;32(4):638-40.

Oliveira AP *et al.* Inferring causal metabolic signals that regulate the dynamic TORC1-dependent transcriptome. Mol Syst Biol 2015;11(4).

Schillings C *et al.* Efficient characterization of parametric uncertainty of complex (bio) chemical networks. PLOS Comput Biol 2015;11:e1004457.

Mihaela Zavolan
RNA Regulatory Networks Group
University of Basel

What do they do?
Individual cells of a body exhibit a stunning diversity of phenotypes, despite carrying a largely identical genetic makeup. The differences between say, a neuron and a muscle cell are thus determined by the distinct ways in which the same genetic information can be read, interpreted and translated into function. This multifaceted process has been a major focus of study over the last decade, during which scientists have unveiled several additional layers of complexity. At the RNA Regulatory Networks (RRN) Group at the Biozentrum in Basel, Mihaela Zavolan and her team use both experimental and computational methods to discover and understand the regulatory networks governing the interpretation of genetic information at the level of tissues and single cells.

Highlights 2015
During the course of 2015, in a series of studies Mihaela Zavolan and her team developed a general approach to infer a biophysical model of miRNA-target interaction that considerably improves the principled prediction of miRNA-target interactions. The model can also be used to predict siRNA off-targets and thereby help in the interpretation of siRNA screening data.

Main publications 2015
Breda J *et al.* Quantifying the strength of miRNA-target interactions. Methods 2015;85:90-9.
Gumienny R *et al.* Accurate transcriptome-wide prediction of microRNA targets and small interfering RNA off-targets with MIRZA-G. Nucleic Acids Res 2015;43(3):1380-91.
Kanitz A *et al.* Comparative assessment of methods for the computational inference of transcript isoform abundance from RNA-seq data. Genome Biol 2015;16:150.

Bioinformatics infrastructure

As technology develops, the quantity of data generated by researchers grows and needs to be not only stored but processed. Life scientists need the help of bioinformaticians to do this.

Consequently, academic institutions and research centres are gradually developing their own infrastructures that provide computational facilities for their researchers and develop software and databases, besides providing a link with industry and offering training.

Peter Kunszt a.i. **Marcel Riedi**
Service and Support for Science IT – S3IT
University of Zurich

What do they do?
At the Service and Support for Science IT (S3IT) unit, Peter Kunzst and his team provide support for science in general, and life science and medicine in particular. S3IT serves as a partner for projects locally and nationally to enable competitive research with the advanced use of computational methods and resources. Kunzst's team advises groups and projects about data management, data analysis and cooperates to optimise their specific workflow. S3IT also takes part in national projects and cooperates with similar technology-oriented groups to ensure that its expertise is always up-to-date.

Highlights 2015
Over the course of the year, the team was able to establish successful collaborations with over 130 end-users in 45 research groups from 22 different departments at the University of Zurich. In particular, S3IT was part of two projects that are highly specialized in 3D imaging, making use of new Light Sheet Microscopy technology, and developed 3D segmentation and genealogy tracing algorithms in these projects.

With regards to infrastructure, in 2014 Peter Kunszt and his team finished the public procurement process for their local ScienceCloud infrastructure. The new infrastructure was installed during the course of 2015 and has over 3000CPU cores and 1.5PB of usable storage.

Main publications 2015

Malmström E *et al.* Large-scale inference of protein tissue origin in gram-positive sepsis plasma using quantitative targeted proteomics. Nat Commun 2016;7:10261.
Tykesson E *et al.* Deciphering the mode of action of the processive polysaccharide modifying enzyme dermatan sulfate epimerase 1 by hydrogen/deuterium exchange mass spectrometry. Chem. Sci 2016;7:1447-1456.
Walzthoeni T *et al.* xTract: software for characterizing conformational changes of protein complexes by quantitative cross-linking mass spectrometry. Nat Methods 2015;0:Epub ahead of print.

Bernd Rinn
Scientific IT Services – SIS
ETH Zurich

What do they do?
Bernd Rinn and his group at the Scientific IT Services (SIS) form an interdisciplinary bioinformatics and scientific IT support group building up computational tools ranging from lab databases to reusable framework components enabling and supporting data management and analysis in life science research and beyond. They collaborate with Swiss and European research groups and industry in the life science sector such as SystemsX, SyBIT FAIR-DOM, HPC-CH and Swiss universities eSCT. They provide their partners with data management solutions and services, integrate and operate data analysis pipelines, and provide training and consulting in databases, scientific software development, high-performance and cloud computing.

Highlights 2015
During the course of 2015 and with several biology wet labs at ETH Zurich, Bernd Rinn and his team developed a combined data management, inventory and electronic lab notebook system which is easy to use for biologists while being at the same time powerful.
In a pilot project of CRUS, the group developed a pilot of a Swiss metadata hub for research datasets from several research areas, including the life sciences.

They set up an Hadoop/Spark big data cluster to create an environment for everyone at ETH, so that they can learn about, test and use this technology. During the year, the team held several workshops on Spark, which turned out to be in high demand.
The group also started a project with a brain research lab and colleagues from S3IT to bring core brain research tools to the Spark big data platform and enable the scaling up of these methods beyond their current limitations. With colleagues from the UK, Germany and Switzerland Rinn and his group have been developing the OpenSEEK platform and the FAIRDOM hub for systems biology research data management in Europe.

Main publications 2015

Barillari C *et al.* openBIS ELN-LIMS: an open-source database for academic laboratories. Bioinformatics 2015;32(4):638-40.
Kiefer P *et al.* DynaMet: A Fully Automated Pipeline for Dynamic LC-MS Data. Anal Chem 2015;87(19):9679-9686.
Wicker J *et al.* enviPath – The environmental contaminant biotransformation pathway resource. Nucl Acids Res 2016;44(D1):D502-8.

Jacques Rougemont
Bioinformatics and Biostatistics Core Facility – BBCF
EPFL, Lausanne

What do they do?
At the EPFL Bioinformatics and Biostatistics Core Facility (BBCF), Jacques Rougemont and his team provide research labs with extensive support in bioinformatics and biostatistics. Their main competences are in management and analysis of genomic data, mathematical modelling and statistical analysis of quantitative biological data. They provide support for the analysis of large or complex data sets, the development of data management pipelines for new high-throughput technologies (e.g. high-density arrays, high-throughput sequencing), and the statistical planning in complex experimental designs. They also help researchers in the areas of mining public data, designing and setting up local databases, building mathematical models from experimental data and running simulations to evaluate a model.

Highlights 2015
During the course of 2015, with Professor D. Shore (UNIGE), Jacques Rougemont and his group analysed the presence of "fragile" nucleosomes – i.e. nucleosomes that are sensitive to MNase digestion – in the yeast genome. They found that these nucleosomes delineate a particular class of gene promoters with sequence-specific features affecting their regulatory architecture.

With Professor B. Deplancke (SIB/EPFL), the group made a major upgrade to the GetPrime system (http://bbcftools.epfl.ch/getprime) which is a database of rtPCR primer pairs for transcript-specific expression analysis. Primers are now available in 13 different species, and the group's pipeline is easily adaptable to others. The web site also links primers to other genomic data of potential interest such as known polymorphisms.

Main publications 2015

Droz-Georget Lathion, S. *et al.* A single epidermal stem cell strategy for safe ex vivo gene therapy. EMBO Mol Med 2015;7:380–393.
Kubik, S. *et al.* Nucleosome Stability Distinguishes Two Different Promoter Types at All Protein-Coding Genes in Yeast. Mol Cell 2015;60:422–434.
Zaiss MM *et al.* The Intestinal Microbiota Contributes to the Ability of Helminths to Modulate Allergic Inflammation. Immunity 2015;43:998–1010.

Ioannis Xenarios
Vital-IT Group

What do they do?
With a multidisciplinary team of scientists and technical staff, at Vital-IT Ioannis Xenarios and his team maintain a competency centre in bioinformatics and computational biology. Their infrastructure currently spreads across six institutions that maintain bio-technological platforms: SIB, the Universities of Geneva, Lausanne, Fribourg and Bern as well as EPFL. At Vital-IT, the team enables scientists to access state-of-the-art computational infrastructures (processing, storage and archiving) as well as expertise in data analysis and algorithmic development. The team partners with scientists to build computational solutions facilitating their research or to transform their ideas into production-quality software. They support postgraduate education through training and workshops in coordination with SIB and institutional partners.

Highlights 2015
During 2015, Vital-IT's infrastructure grew to 6PetaBytes of storage and 7,000 computational cores, all fully dedicated to life science and medical applications. It embeds over 2,400 bioinformatics-supporting software packages that enable reproducible science and data life cycle management.

Vital-IT's infrastructure supports several hundred research projects and maintains widely-used SIB resources, i.e. Swiss-Prot MetaNetX, NeXtProt, OpenFlu, LipidX, SwissDock, Swiss-Model and the ExPaSy portal. It also maintains more than 140 websites and services for its partner groups with thousands of daily visitors. The demand for supporting research projects grows constantly; there were over 730 active users in 2015. Vital-IT collaborates on research activities as part of SystemsX.ch, and technology and development projects funded by the European Commission. It also provides training to biomedical and life scientists on how to develop and use bioinformatics software in a high-performance computing environment.

Main publications 2015

Chasapi A *et al.* An extended, Boolean model of the septation initiation network in S.pombe provides insights into its regulation. PLOS ONE 2015;10(8):e0134214
Doucey MA *et al.* Toward a rational design of combination therapy in cancer. Oncoimmunology 2015;4(11):e1046674
Guex N *et al.* Angiogenic activity of breast cancer patients' monocytes reverted by combined use of systems modeling and experimental approaches. PLOS Comput Biol 2015;11(3):e1004050

Aknowledgements

We gratefully acknowledge the following funders, sponsors and partners for their financial support and encouragement in helping us fulfil our mission in 2015.

- The Swiss government and in particular:
  - The State Secretariat for Education, Research and Innovation (SERI)
  - The Swiss National Science Foundation (SNSF)
  - The Commission for Technology and Innovation (CTI)
- Our institutional members (see list p. 9)
- The European Commission
- The Cogito Foundation
- The Leenaards Foundation
- The Medic Foundation
- The National Institutes of Health (NIH)
- Oncosuisse
- The Research for Life Foundation
- The Solidar Immun Foundation
- SystemsX.ch

The SIB Fellowship programme continues, thanks to the generous support of:

- The Leenaards Foundation
- The R. Geigy Foundation
- The Swiss Foundation for Excellence and Talent in Biomedical Research
- SystemsX.ch
- The University of Geneva
- The University of Lausanne
- The University of Zurich

We also thank all industrial and academic partners who trust SIB's expertise.

SIB profile 2016 written by the SIB Communications Department and Vivactis (Suisse) SA, with contributions by SIB members.

SIB Group Leaders' pictures: Nicolas Righetti, rezo.ch
Cover image: DNA background, Fotolia

Design and layout: Atelier Valenthier, valenthier.ch