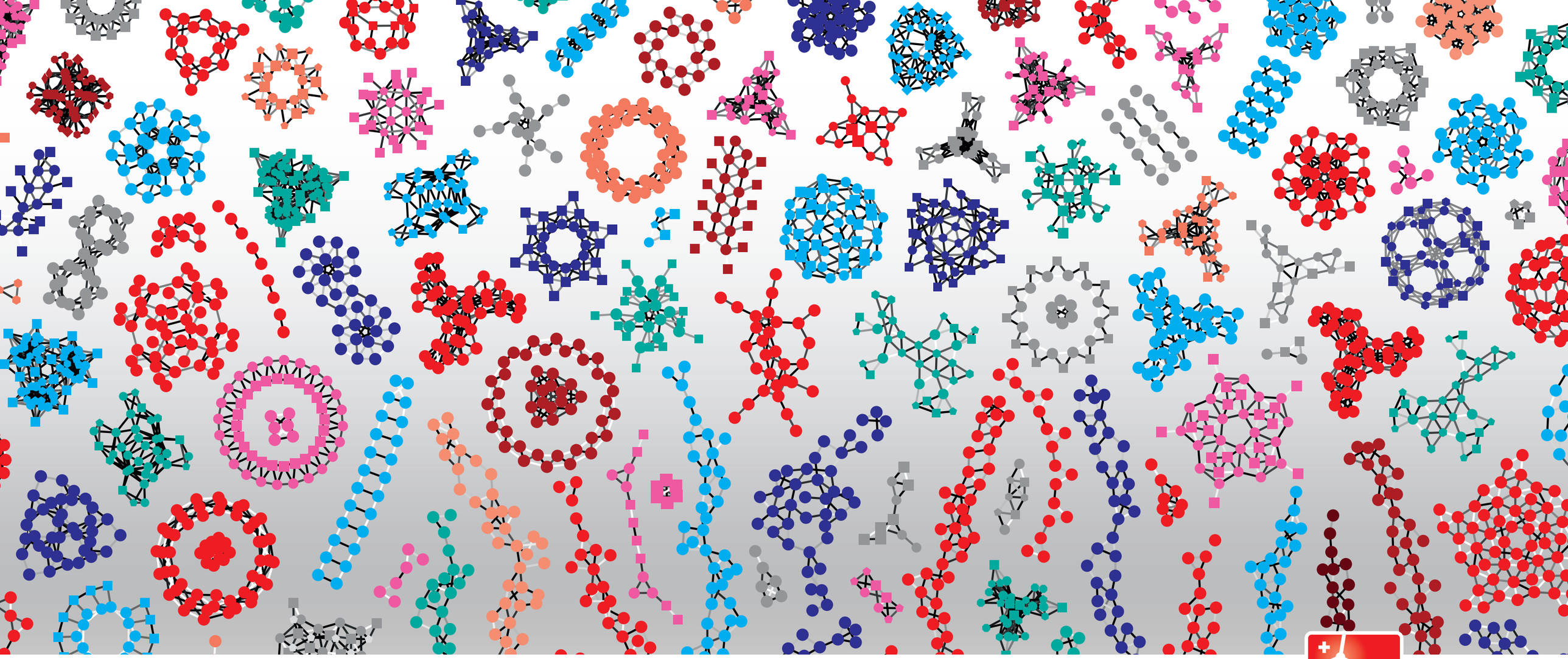


Swiss Institute of Bioinformatics

SIB Profile 2017

Empowering advances in life sciences and health

Swiss Institute of Bioinformatics

Quartier Sorge
Bâtiment Génopode
CH-1015 Lausanne
Switzerland
t +41 21 692 40 50

www.sib.swiss

elixir SWITZERLAND

Table of contents

Foreword

2016 proved to be yet another eventful year for SIB. The Institute continued to strengthen its position as an important representative of bioinformatics on the national and international scale, and to offer its unswerving support to research, services and training while pursuing important endeavours in the field of personalized health.

"Year after year, SIB continues to demonstrate its support for all fields of the life sciences, with a more recent accent on personalized health."

A large part of 2016 was spent contributing to the official start of the Swiss Personalized Health Network (SPHN) that began in January 2017, for which SIB is leading the Data Expert group and is in charge of the Data Coordination Centre. The Institute was also asked to coordinate the integration of clinical data for the pan-European project – known as RHAPSODY – for precision therapy and the prevention of diabetes, which was launched in September 2016. This project brings together researchers and experts from 26 partner institutions from both the public and private sectors. Finally, in October 2016, thanks to the joint collaboration between SIB's Vital-IT and Clinical Bioinformatics groups, and staff at the Geneva University Hospitals (HUG), SIB's first clinical bioinformatics tool – OncoBench™ – was inaugurated at HUG. It is used for the routine molecular diagnosis of cancer patients, thus opening a new and exciting chapter in the Institute's history.

On the data resource front, SIB co-led the development of the governance and processes that support the identification and evaluation of ELIXIR Core Data Resources. These resources, which are of crucial importance for the life sciences, will form the focal point of technical and science policy actions to ensure their long-term sustainability and promote excellence in resource development.

In April 2016, SIB was proud to host the 9th International Biocuration Conference in Geneva. The event coincided with the 30th anniversary of one of SIB's core resources used worldwide: the Swiss-Prot database created in 1986, which then became UniprotKB/Swiss-Prot, the knowledgebase of protein sequences and functional information produced by SIB in collaboration with PIR and EMBL-EBI. The challenges created by the exponential growth of biomedical data and how to make them accessible to scientists and computers were the main topics addressed. The event lasted five days and was a great success. In September, another of SIB's core resources reached the mature age of 30: the Eukaryotic Promoter Database (EPD). A well-attended symposium took place in Lausanne on this occasion.

Year after year, SIB continues to demonstrate its support for all fields of the life sciences, with a more recent accent on personalized health. It has become a leading organization on the national front and its influence continues to grow beyond its borders. With several new groups joining SIB in 2016, the total number of member groups has now reached 65, representing about 800 scientists. None of this, however, could be achieved without the backing of many bodies. We would like to thank the Swiss government and in particular the State Secretariat for Education, Research and Innovation SERI, the Federal Assembly, 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 to raise Swiss bioinformatics to the influential position it holds 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.

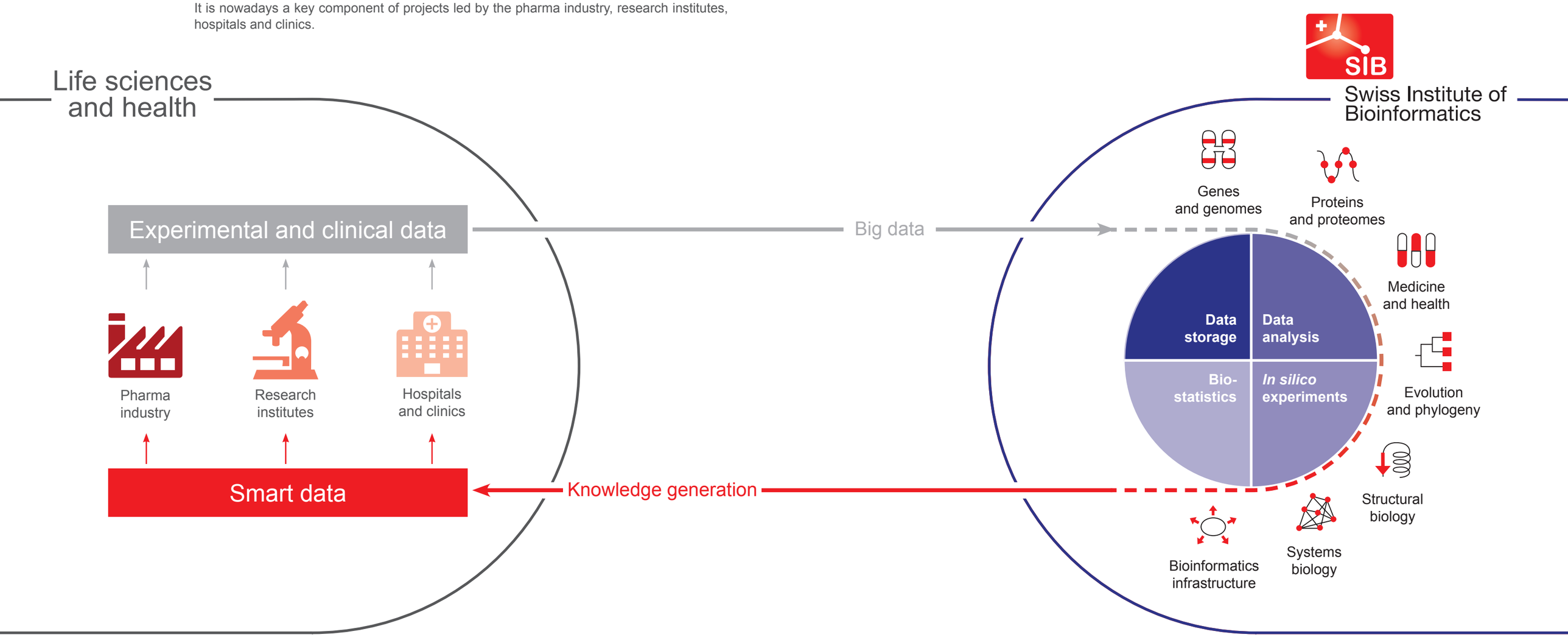
Bioinformatics helps convert “big data” into “smart data”

With the advent of new technologies and automated measuring devices, ever-larger volumes of experimental and clinical data (“big data”) are generated. These data need to be stored, organized and analysed in order to extract new insights and knowledge (“smart data”). Computer-based approaches have therefore become a third pillar of science, by allowing researchers to advance their understanding of complex systems. The same applies to life sciences, where little research could be performed today without the help of bioinformatics. It is nowadays a key component of projects led by the pharma industry, research institutes, hospitals and clinics.

Bioinformatics encompasses:

- **Data- and knowledgebases** for storing, retrieving and organizing biological information
- **Software** for modelling, visualizing, analysing, interpreting and comparing biological data
- **Computing and storage infrastructure** for “big data” processing
- **Biocuration and expertise** providing life scientists with an accurate and comprehensive representation of biological knowledge and enabling them to take full advantage of bioinformatics technologies.

Bioinformatics is therefore an exciting interdisciplinary field, which is driving major advances in many different life science and health-related areas.



About SIB

Empowering advances in life sciences and health

The SIB Swiss Institute of Bioinformatics is a unique success story at the frontier of life 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 as being of public utility, which 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 contributes significantly to creating a true bioinformatics culture in Switzerland.

The Institute is leading developments in the field of bioinformatics, including in the rapidly developing area of personalized health. Having anticipated the challenges raised by the advent of the 'omics' era and growing self-awareness among patients, SIB is now at the forefront of personalized health endeavours and has been selected by the Swiss government as its bioinformatics partner in national personalized health projects.

19
institutional members across Switzerland

Some **800**
members

65
research and service groups

12
core facilities

More than **2,000**
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. Its 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. **Bringing together** world-class researchers and delivering **training** in bioinformatics.

To achieve its mission, SIB is committed to:

Creating, maintaining and disseminating worldwide a large portfolio of reliable, sustainable and top-quality core bioinformatics services and resources, such as databases, software and competence centres

Serving the national life science community by offering easily accessible world-class competencies, expertise and support in bioinformatics

Supporting hospitals and clinicians with know-how, resources and infrastructure dedicated to personalized health

Bringing together bioinformatics research groups from Swiss universities and research institutes

Fostering collaboration and innovation at the highest level of scientific excellence

Providing life scientists, clinicians, and PhD students with a large portfolio of courses and workshops

Fostering exchanges among bioinformatics and computational biology PhD students and training them in the most up-to-date methods necessary for their doctoral research

Representing and promoting bioinformatics locally and internationally

Organization

SIB, an efficient collaborative Swiss model...

The decentralized, federating organizational structure of the SIB Swiss Institute of Bioinformatics serves as a collaborative model for countries setting up their own bioinformatics infrastructure.

SIB's unique organization is modelled on Switzerland's federal structure. It consists of bioinformatics research and service groups from the major Swiss schools of higher education and research institutions (see opposite page). While most SIB Group Leaders are senior academic staff of the partner institutions, a number of SIB scientists are employed directly by the Institute (see p. 15). Although each research group carries out its own research and teaching activities independently within its host institution, it benefits from a wide range of resources and support provided by SIB. In return, the Swiss universities and research institutes provide SIB members with the infrastructure needed to perform their mission.

...and the largest national bioinformatics network in Europe

SIB acts as the Swiss node of ELIXIR, a pan-European organization that is building a sustainable European infrastructure to support life science research.

SIB co-leads the ELIXIR Data and Training Platforms and is involved in several other work streams, such as the benchmarking of software tools.

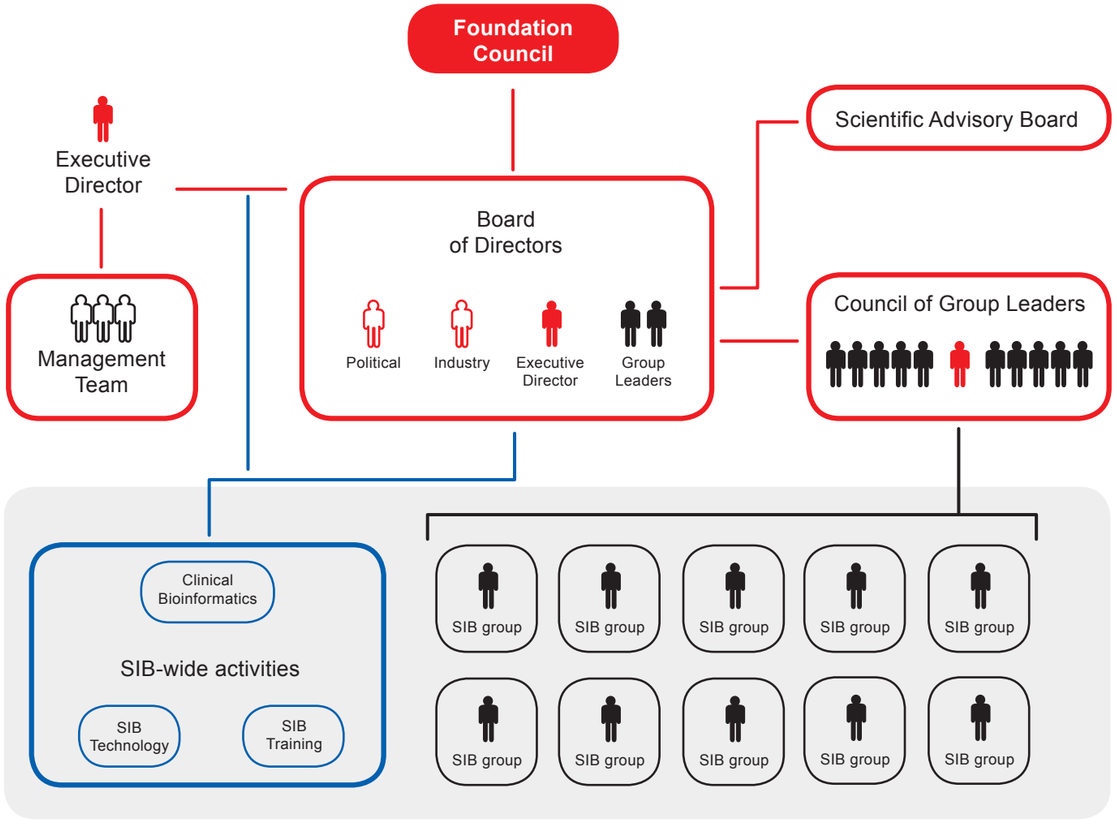
Highlights 2016

- SIB co-led the development of the governance and processes supporting the identification and evaluation of ELIXIR Core Data Resources. These resources, which are of crucial importance for the life sciences, will form the focal point of technical and science policy actions to drive their long-term sustainability and promote excellence in resource development (Durinx C *et al., F1000Research* 2016, 5:2422).
- The ELIXIR Training Platform aims to establish a training community across all member states that coherently delivers ELIXIR-related training. During the past year, the SIB Training group was involved in ELIXIR activities on metrics related to the quality and impact of training and eLearning, as well as the identification of training needs across the ELIXIR community.
- After chairing the ELIXIR Board in 2015 and 2016, Torsten Schwede, Director of the SIB Personalized Health Informatics group and member of the SIB Board of Directors, has handed over as chairman to Prof. Rein Aasland from Norway.

SIB partner institutions

Governance

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



The Foundation Council

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

President

Prof. Felix Gutzwiller
Former Senator

Founding Members

Prof. Ron Appel
SIB Executive Director, Professor at the 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 Hospitals (HUG)

Prof. C. Victor Jongeneel
Carl R. Woese Institute for Genomic Biology, University of Illinois, USA

Prof. Manuel Peitsch
Chairman, SIB Board of Directors and Chief Scientific Officer Research at Philip Morris International

Ex officio Members

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

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

Prof. Edouard Bugnion
Vice-President for Information Systems, EPFL

Prof. François Bussy
Vice-Rector for Research, International Relations and Continuing Education, University of Lausanne

Prof. Carlo Catapano
Director, IOR Institute of Oncology Research

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

Prof. Boas Erez
Rector, Università della Svizzera Italiana

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

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

Prof. Detlef Günther
Vice-President Research and Corporate Relations, ETH Zurich

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

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

Prof. Rolf Ingold
Vice-Rector for Research and Information Technology, University of Fribourg

Dr Corinne Jud
Head of the Competence Division Method Development and Analytics, Agroscope

Dr Caroline Kant
Executive Director, EspeRare Foundation Switzerland

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

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

Prof. Alexandre Reymond
Director, Centre for Integrative Genomics (CIG), Faculty of Biology and Medicine, University of Lausanne

Prof. Eduardo Sanchez
Dean, School of Management and Engineering Vaud (HEIG-VD), HES-SO

Prof. Christian Leumann
Rector, University of Bern

Prof. Juerg Utzinger
Director, Swiss Tropical and Public Health Institute (Swiss TPH)

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

The Board of Directors (BoD)

The BoD takes all the decisions necessary to achieve the aims of the Institute, such as defining the scientific strategy and internal procedures, and allocates federal funds to service and infrastructure activities. The BoD 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. Members of the BoD are appointed for a renewable five-year period.

Prof. Manuel Peitsch
Chairman
Chief Scientific Officer Research at Philip Morris International

Ms Martine Brunschwig Graf
Former National Councillor

Prof. Ron Appel
SIB Executive Director, Professor at the University of Geneva

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

Prof. Torsten Schwede
Group Leader, SIB and University of Basel

The Scientific Advisory Board

This board 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 activity.

Prof. Alfonso Valencia
Chairman
Life Sciences Department Director, Barcelona Supercomputing Centre, Spain

Prof. Søren Brunak
Founder of the Centre for Biological Sequence Analysis, Technical University of Denmark, Denmark

Dr Laurent Duret
CNRS Research Director, Laboratory of Biometry and Evolutionary Biology, Claude Bernard-Lyon 1 University, France

Dr David de Graaf (until June 2017)
President and CEO of Selventa, Cambridge, MA, USA

Prof. Melissa Haendel
Director of the Ontology Development Group, Oregon Health & Science University, Portland, 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 (passed away in March 2017)
Computational Biology Laboratory, La Sapienza University, Rome, Italy

The Council of Group Leaders

The council 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: SIB Group Leaders are staff members at SIB partner institutions (see p. 39). In addition, OsiriX, led by Prof. Osman Ratib, is an Affiliate Group.

Honorary Members

Prof. Ernest Feytmans
Honorary Director

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

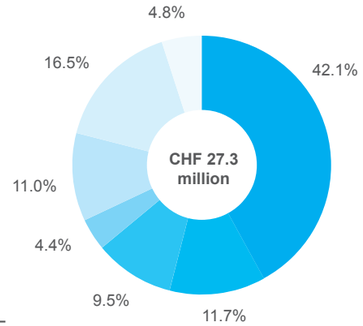
Finance

SIB funds remained stable in 2016, thanks to the continued support of its funders.

Sources of income

In 2016, the total amount of funds managed by SIB reached CHF 27.3 million, as indicated in the table below. The largest source of SIB funds is the Swiss government (CHF 11.5 million, 42.1%).

Altogether, the total 2016 budget for bioinformatics in Switzerland (including both funds managed by SIB and funds managed by SIB's partner institutions) amounted to CHF 82.7 million.

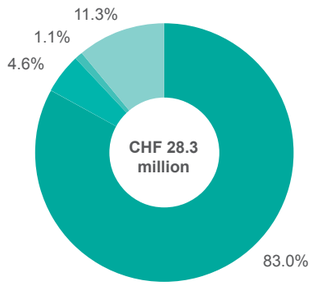


Funds managed by SIB

	Funds managed by SIB		Total number of grants/ contracts managed by SIB
	CHF million	%	
Swiss government (SERI)	11.5	42.1	1
SNSF and European funds	3.2	11.7	21
NIH	2.6	9.5	2
SystemsX.ch	1.2	4.4	9
Industry	3.0	11.0	
Universities and hospitals	4.5	16.5	18
Other	1.3	4.8	
	27.3		**51**

Spending

Of the CHF 28.3 million spent by SIB in 2016, 83% was allocated to salaries and the rest to equipment, scientific events and running costs.



SIB spending

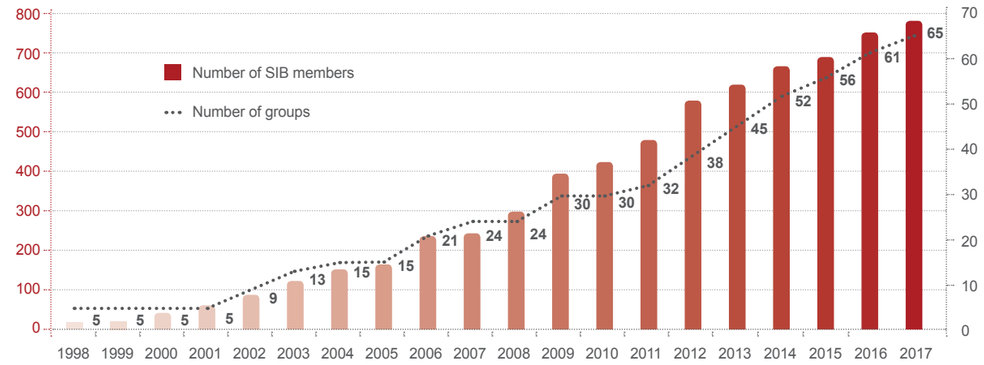
	SIB spending	
	CHF million	%
Salaries	23.5	83.0
Equipment	1.3	4.6
Scientific events	0.3	1.1
Running costs	3.2	11.3
	28.3	

Membership and staff

SIB continues to grow in terms of membership and staff.

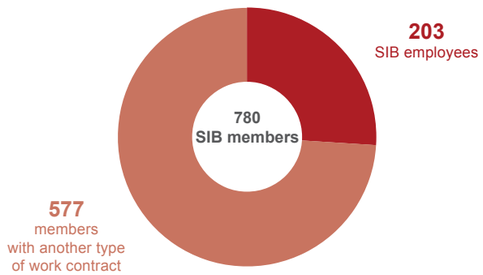
Evolution of SIB membership

Over the past eight years, the number of SIB members has doubled, along with the number of groups.



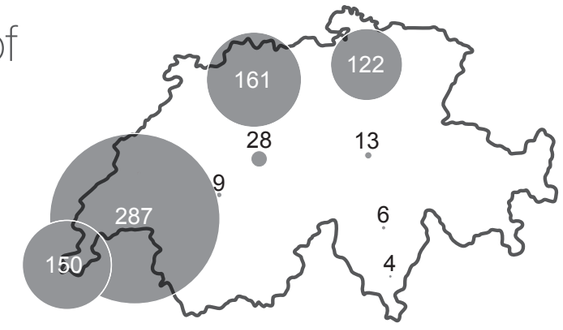
SIB members

As of 1 February 2017, SIB had 780 members, of whom 203 are SIB employees (i.e. with an SIB work contract). Among SIB employees, 28 nationalities are represented.



Geographical distribution of SIB members

As of 1 February 2017, more than one third of SIB members were located in Lausanne, followed by Basel (21%), Geneva (19%) and Zurich (16%).



In the spotlight

Focus on personalized health

The past year was particularly rich in developments for SIB in the field of personalized health. From oncology to diabetes, at the national and international level, here are a few examples of the endeavours in which SIB is playing a key role to improve health.

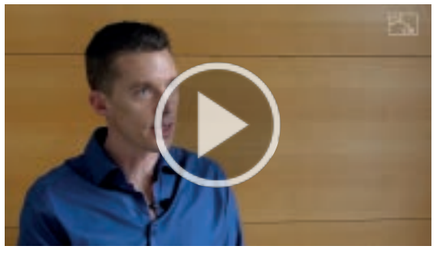
Interviews

Bridging the gap between data science and medicine

Valérie Barbié, Head of the SIB Clinical Bioinformatics group, and Jacques Beckmann, Head of the group until his retirement on 31 December 2016, present the initiative that led in September 2016 to the rollout of an oncology diagnostic pipeline at the Geneva University Hospitals (HUG), in collaboration with SIB's Vital-IT group.

Building the national research infrastructure for personalized health

Torsten Schwede, Head of the Personalized Health Informatics group (PHI), explains the pivotal role of SIB in the Swiss Personalized Health Network at two levels: as Data Coordination Centre and via the secure nationwide network BioMedIT.



RHAPSODY for personalized health of diabetes

Mark Ibberson, senior scientist at SIB's Vital-IT group, presents a pan-European public-private partnership established to find new ways to predict and fight diabetes, with SIB acting as Data Coordination Centre.

In addition to providing the infrastructure and competence centres that are essential to support the growing field of personalized health, SIB is also actively fostering research for patients' benefit, with many of the Institute's scientists tackling issues such as cancer and immunotherapy, obesity or HIV. Read more on these topics in our Research pages (see pp. 42-83).

Scan to watch these videos on our SIB YouTube channel.

A new adventure: Digital Humanities

© watchara / Shutterstock

Today, the humanities can benefit from the accumulated experience of bioinformatics solutions in terms of infrastructure, technical know-how, data visualization and computational approaches. This is the realm of Digital Humanities (DH), and SIB's Vital-IT group is supporting a number of projects in the field, led by Claire Clivaz under the label "Vital-DH projects@Vital-IT".

One joint project involves the transcription and interpretation of ancient manuscripts. An illustration is the design of a viewer for an Arabic manuscript dating back to the ninth century, for which Claire and her team make use of SIB's expertise in developing tools for curation. In this way, manuscripts can be made available to all via a virtual research environment while giving room to comments, discussions and annotations made by third parties. A manuscript then becomes a dynamic and interactive entity as it is continuously enhanced with information that can also be immediately verified.

Yet another project focuses on the enhancement of a pre-existing concept known as "eTalks". eTalks are speeches that have been recorded and are subsequently edited, but in a very special way: each speech is split into units enriched with text and image, and can be explicitly referred to via a URL. Each unit can thus be cited, as an article would be. To date, 25 eTalks have been produced, ranging from human enhancement technologies to personalized health.

Autophagy: an important step in biocuration

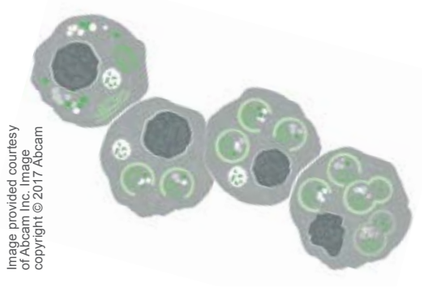


Image provided courtesy of Abcam Inc. Image copyright © 2017 Abcam

Autophagy – literally meaning "self-digestion" – was first observed in the 1960s. Why would a cell digest parts of its insides in the first place, and what are the underlying molecular mechanisms? When under stress, such as starvation for example, a cell cannot rely on the outside world for sustenance anymore, so it resorts to food it is already carrying. This form of self-recycling is done in cell organelles known as lysosomes. The genes involved in autophagy have now been discovered and reveal not only the mechanisms at work but also diseases that can occur when a gene is deficient, namely certain forms of cancer and neurological diseases such as Parkinson's disease.

The 2016 Nobel Prize in Physiology or Medicine paid tribute to the fascinating cellular process of autophagy. This coincided, and quite by chance, with a major endeavour on the subject carried out by SIB's Swiss-Prot group.

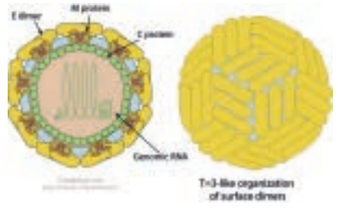
The Swiss-Prot group develops and maintains the UniProtKB/Swiss-Prot protein knowledgebase and, over the years, special attention has been paid to autophagy. In 2016, Marc Feuermann of the Swiss-Prot group completed the curation of the process within the UniProtKB/Swiss-Prot knowledgebase. Autophagy is used throughout the eukaryotic kingdom – from yeast to humans – so the task was not a small one, and the information was disseminated across thousands of entries describing homologous protein sequences.

A second achievement soon followed. There is an ongoing collaboration between the UniProtKB/Swiss-Prot database and a project known as Gene Ontology, or GO. GO is a controlled vocabulary that provides a unified description of gene function across all species. GO is used by many bioinformatics resources including UniProtKB/Swiss-Prot. Using the new and updated representation of autophagy in UniProtKB/Swiss-Prot as a base, Marc, along with Pascale Gaudet of SIB's CALIPHO group, developed a complete representation of autophagy in GO, and used it to annotate UniProtKB. The work was described in a publication in the journal *Database* in December 2016.

2016 at a glance

Build-up of harmful mutations during early human migrations

SIB Group Leader L. Excoffier and his team contributed to an international study published in *PNAS* on early migrations out of Africa, which showed that the further away a population moves from its place of origin, the more harmful mutations it will carry.



© ViralZone 2016 - SIB

Bioinformatics *vs* Zika

SIB's Swiss-Prot group developed a Zika virus page on the virology resource ViralZone, giving access to a wealth of biological data and publications about this insect-borne virus.

© Valdo Peixoto and Daniel Marbach. valdopeixoto.com

"Social networks" of genes disrupted in complex diseases

Accurate mapping of gene networks in human cell and tissue types sheds light on disease mechanisms and targeted treatments. The study, led by researchers from SIB and the University of Lausanne, was published in *Nature Methods* and *PLoS Computational Biology*.

Orthology benchmarking made easy

To identify the best methods for finding orthologs (genes that are directly related in different species), a team led by SIB Group Leader C. Dessimoz developed an innovative web-based service, called "Orthology Benchmarking", and published their work in *Nature Methods*.

International Biocuration Conference 2016

The conference, hosted by SIB at Campus Biotech Geneva, gave the opportunity to curators and developers of biological databases to meet and showcase their work in this fast-expanding field.

SIB Days 2016

Organized in Biel, this internal biyearly scientific event was a great opportunity for members to network and attend lectures on bioinformatics and other scientific fields, poster sessions, and workshops.

JANUARY | FEBRUARY | MARCH | APRIL // JUNE

Software and Data Carpentry instructors' training

SIB hosted a workshop for ELIXIR trainers wishing to become Software and Data Carpentry instructors.

The tick genome brings hope against Lyme disease

The genome of the deer tick was sequenced by an international team including SIB researchers, and the results published in *Nature Communications*. Understanding how these disease-carrying arachnids function will help to develop novel tick-control programmes.

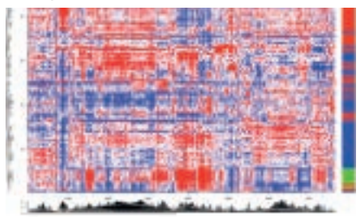
© Dr Andrew Nuss

"Metagenomic Pizza" in *EMBnet Journal*

The Metagenomic Pizza workshop was published in *EMBnet Journal*. Developed by SIB, the workshop aims to explain to a lay audience how DNA and bioinformatics tools available on the internet can be used to identify food ingredients.

Swiss-Prot 30th anniversary

SIB's Swiss-Prot knowledgebase celebrated its 30th anniversary. Today, UniProtKB/Swiss-Prot is the most widely used protein information resource in the world.



Genetic test to reduce chemotherapy use

SIB's Bioinformatics Core Facility led by M. Delorenzi played a key role in the MINDACT study, which showed that nearly 50% of women with breast cancer, to whom chemotherapy would traditionally be given, do not require it.

© Michel Milinkovitch

A lot in common for hairs, feathers and scales

The potential evolutionary link between hairs in mammals, feathers in birds and scales in reptiles has been debated for decades. Researchers at UNIGE and M. Milinkovitch's SIB group demonstrate that all these skin appendages share a common ancestry.

SIB at “Nuit de la Science”

SIB took part in this public event, organized by the Science History Museum in Geneva. With its Drug Design workshop, the Outreach team demonstrated to a wide audience how bioinformatics is used to design a drug today.

A pan-European project to fight diabetes

SIB, via its competence centre Vital-IT, became the data coordination centre of a pan-European public-private partnership on Risk Assessment and Progression of Diabetes (RHAPSODY).

© Stephen Gowland (Kiaco)

First genomic history of Australia’s peopling

SIB Group Leaders A-S. Malaspinas, L. Excoffier and four of their groups’ members participated in the first comprehensive genomic study on Australia’s peopling history, published in *Nature*.

Launch of OncoBench™

A new genomic analysis platform for a faster cancer diagnosis was launched thanks to an exemplary collaboration between the SIB Clinical Bioinformatics group, Vital-IT experts and the Geneva University Hospitals.

The Gene Ontology Handbook

This new book on Gene Ontology by SIB Group Leader C. Dessimoz and other present and past SIB members has been published by Springer and is freely available online.

© Jody McIntyre

Mapping the spread of avian flu

In collaboration with scientists from the Université Libre de Bruxelles and the Food and Agriculture Organization of the United Nations (FAO), the SIB groups Swiss-Prot and Vital-IT published the first global model to predict the geographic spread of avian flu.

JULY // SEPTEMBER OCTOBER NOVEMBER DECEMBER

SIB’s Virtual Computational Biology seminars online

The seminar series, allowing life scientists and clinicians to learn more about SIB’s research, expertise and resources, is now available to everyone, everywhere, via a live webcast and a dedicated playlist on the SIB YouTube channel.

SIB clinical metagenomics workshop in Bern

SIB Clinical Bioinformatics organized a workshop of clinical metagenomics to understand the challenges and solutions linked to pathogen identification. The event brought together more than 50 participants from various organizations from across Switzerland.

EPD 30th anniversary

The specialized Eukaryotic Promoter Database EPD developed by SIB Group Leader P. Bucher reached the mature age of 30, with a symposium organized to mark the occasion.

SIB at “Planète Santé Live”

SIB took part in the “Planète Santé Live” fair, one of the largest health-related public events in Switzerland, and welcomed a large number of students, families and clinicians who all participated with enthusiasm in the drug design or clinical bioinformatics activities.

SIB Training’s offer tailored for companies

SIB Training team has tailored its offer to the needs of companies. Three groups (Roche, Debiopharm and a leading Swiss consumer goods company) benefited from these customized on-site, private courses.

Services and resources

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 industries.

SIB provides the necessary bioinformatics services and research infrastructure for scientists thanks to:

- **Over 150 internationally recognized and extensively used bioinformatics databases and software tools,** which SIB continuously develops and maintains
- **26 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 protect SIB's knowledge and know-how and ensure its transfer to research institutes and the industry.

Databases and tools

SIB groups develop, supply and maintain more than 150 high-quality databases and tools for the global life science community.

Most of SIB's resources are available via open access on the SIB bioinformatics resource portal ExPASy (www.expasy.org). Created in 1993, ExPASy was at that time the first website in the biomedical field. SIB's 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, SwissADME, Swiss-Similarity
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 tools that are available on ExPASy.

SIB's core resources

UniProtKB/Swiss-Prot
Protein knowledgebase

Type - Knowledgebase with manual expert curation

Description - Hundreds of thousands of protein descriptions, including function, domain structure, subcellular location, post-translational modifications and functionally characterized variants

▸ **Highlight** - Most widely used protein information resource in the world, with over 900,000 user requests per month. The Swiss-Prot section of the UniProt resource celebrated its 30th birthday in 2016

See also p. 56

neXtProt
Human protein knowledgebase

Type - Knowledgebase with manual expert curation

Description - Information on various aspects of human protein biology such as function, involvement in diseases, mRNA/protein expression, protein/protein interactions, post-translational modifications, protein variations and their phenotypic effects

▸ **Highlight** - Advanced search option enabling the user to make very precise queries, e.g. "proteins highly expressed in brain but not expressed in testis" or "proteins that bind a metal and are secreted"

See also p. 54

SWISS-MODEL
Protein structure homology-modelling service and repository

Type - Software tool and knowledgebase

Description - Automated protein structure homology-modelling platform for generating 3D models of a protein using a comparative approach, and database of annotated models for key reference proteomes based on UniProtKB

▸ **Highlight** - Reliable and easy-to-use web-based platform making model information available also for non-specialists

See also p. 72

STRING
Protein-protein interactions

Type - Knowledgebase

Description - Resource of known and predicted protein-protein interactions, including direct (physical) and indirect (functional) associations derived from various sources, such as genomic context, high-throughput experiments, (conserved) co-expression and the literature

▸ **Highlight** - Currently includes 9,643,763 proteins from 2,031 organisms and is the most used resource of its kind

See also p. 55

SwissRegulon Portal
Tools and data for regulatory genomics

Type - Software tools and knowledgebases

Description - Web portal for regulatory genomics, including genome-wide annotations of regulatory sites and motifs, ISMARA webserver for automated inference of regulatory networks, and CRUNCH for automated analysis of ChIP-seq data

▸ **Highlight** - User can upload raw micro-array, RNA-seq or ChIP-seq data to the ISMARA webserver to automatically infer the core regulatory networks acting in the system of interest

See also p. 49

SwissDrugDesign
Drug design

Type - Software tools

Description - Web-based computer-aided drug design tools, from molecular docking (SwissDock) to pharmacokinetics and druglikeness (SwissADME), through virtual screening (SwissSimilarity) and target prediction of small molecules (SwissTargetPrediction)

▸ **Highlight** - Comprehensive and integrated web-based drug design environment

See also p. 72

EPD
Eukaryotic Promoter Database

Type - Knowledgebase with manual expert curation and software tools

Description - Quality-controlled information on experimentally defined promoters of higher organisms, as well as web-based tools for promoter analysis

▸ **Highlight** - Celebrated its 30th birthday in 2016

See also p. 47

Bgee
Gene expression evolution

Type - Knowledgebase with manual expert curation

Description - Information on gene expression evolution (incl. all types of transcriptomes), allowing to retrieve and compare expression patterns between animals including human, model organisms and diverse species of evolutionary or agronomical relevance

▸ **Highlight** - Only resource to provide homologous gene expression between species

See also p. 67

UniCarbKB
Glycan knowledgebase

Type - Knowledgebase with manual expert curation

Description - Comprehensive information on glycan structures and published glycoprotein information including global and site-specific attachment information

▸ **Highlight** - International collaboration between institutions across six countries

See also p. 55

SugarBind
Pathogen-sugar binding knowledgebase

Type - Knowledgebase with manual expert curation

Description - Information on known mammalian carbohydrate sequences to which pathogenic organisms (bacteria, toxins and viruses) specifically adhere, and supporting the investigation of bacterial and viral infections

▸ **Highlight** - Only searchable resource describing interactions between pathogen proteins and mammalian carbohydrates

See also p. 55

OrthoDB
The hierarchical catalog of orthologs

Type - Automated phylogenomic database and software tool

Description - Comprehensive online catalogue of animal, fungal, plant, archaeal, bacterial and viral orthologs, including functional and evolutionary gene annotations and enabling the inference of putative gene functions

▸ **Highlight** - Largest orthology resource. Users can generate publication-quality comparative genomics charts, as well as upload, analyse and interactively explore their own pre-publication data

See also p. 52

OMA
Orthology MAtrix browser

Type - Automated phylogenomic database and software tool

Description - High-quality orthology predictions among complete genomes

▸ **Highlight** - Broad scope and size, feature-rich web interface, availability in a wide range of formats and interfaces, frequent update schedule

See also p. 65

Embedded bioinformaticians

Bringing data science experts to research labs

Bioinformatics skills are essential 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 supporting embedded bioinformaticians in various research and clinical labs.

An increasing number of bioinformaticians are physically co-located with scientists in wet labs in research institutes or sequencing departments in hospitals. 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 optimal use of various bioinformatics tools.

Similarly, the physical co-location 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 specifically designed and optimized for clinical research, patient data analysis or diagnosis.

The 26 embedded bioinformaticians from the different institutions can take advantage of the SIB expert network via an SIB host group, providing them with state-of-the-art expertise and support in the field.

Hiring a bioinformatician – available next door to the lab space – has probably been the single most impactful decision I have taken since starting my group. It has transformed our research in many ways, including the nature of the projects we design and the experimental approaches we can actually take."
David Gatfield, Associate Professor, Centre for Integrative Genomics, University of Lausanne

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. The SIB Legal and Technology Transfer Office (LTTO)'s mission is to protect and transfer SIB's knowledge and know-how to research institutes and the industry so that the public can benefit from the SIB groups' many innovations.

Main activities

- **Partnerships with the industry:** the LTTO strives to enhance the scientific and industrial visibility of the SIB groups' innovation by assisting SIB members in their contacts with external partners. Companies involved in medicine and life sciences can collaborate with SIB to complement their internal capacity.

  Examples of such collaborations include:

  - Scientific support and data analysis thanks to on-site computational tools and in-depth expertise
  - 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 from health records, monitoring of social media platforms for drug safety surveillance).

- **Management of the company GeneBio:** SIB's commercial arm GeneBio commercializes software tools and resources developed in-house, e.g. Melanie, Prosite and SmileMS.

- **Legal advice:** The LTTO is responsible for the Institute's legal affairs and advises the SIB Management and Group Leaders on a broad range of legal topics such as copyright, personal data protection and research involving human beings.

In the current scientific context, where scientific innovation is led by academic groups, legal aspects such as the establishment of research or service contracts and the management of intellectual property have become essential. The SIB LTTO is of major importance for SIB groups in this regard."
Vincent Zoete, Associate Group Leader, Molecular Modelling group, SIB and Assistant Professor at the University of Lausanne

Competence centres

Because the whole is greater than the sum of its parts, gathering expertise, infrastructure and computing power under the same roof unleashes interdisciplinary research and excellence. The competence centres at SIB include its core facilities and high-performance computing (HPC) centres, as well as technology coordination.

Core facilities

Through 12 core facilities and HPC centres, SIB groups provide expert data analysis services and computing power to life scientists in academia and industry, thus enabling them to perform world-class research.

Vital-IT

Vital-IT
University of Lausanne
Ioannis Xenarios

The Vital-IT multidisciplinary team of scientists and technical staff maintains a competence centre in bioinformatics and computational biology, which also serves as a reproducible science platform and life cycle management centre.
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 infrastructure (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 the SIB Training group and institutional partners.

More than
12,000,000
CPU hours consumed per year

20
expert embedded bioinformaticians providing on-site data analysis services to life scientists and clinicians

More than
24,000,000
jobs run by users per year

Author of, or acknowledged in:
112
publications in 2016

39
publications in *Science* or *Nature* over the last 6 years

820
active users of Vital-IT in 2016

15
Petabytes of storage capacity

8,600
CPUs in 2016

sciCORE

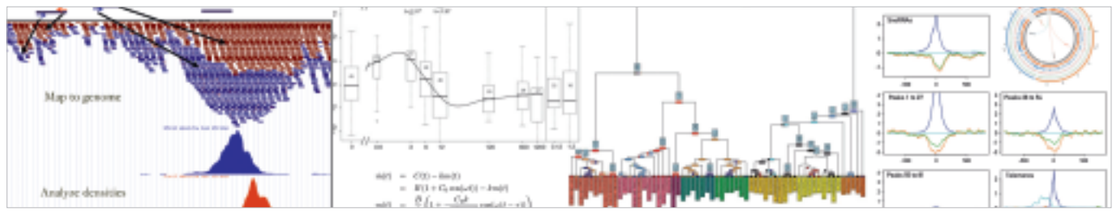
sciCORE
University of Basel
Torsten Schwede
Thierry Sengstag

sciCORE is a competence centre in scientific computing – providing high-performance computing infrastructure, large-scale storage resources, scientific software and databases, server infrastructure 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 and systems biology to medicine and economics. Working in direct collaboration with scientific research groups, the competence centre 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

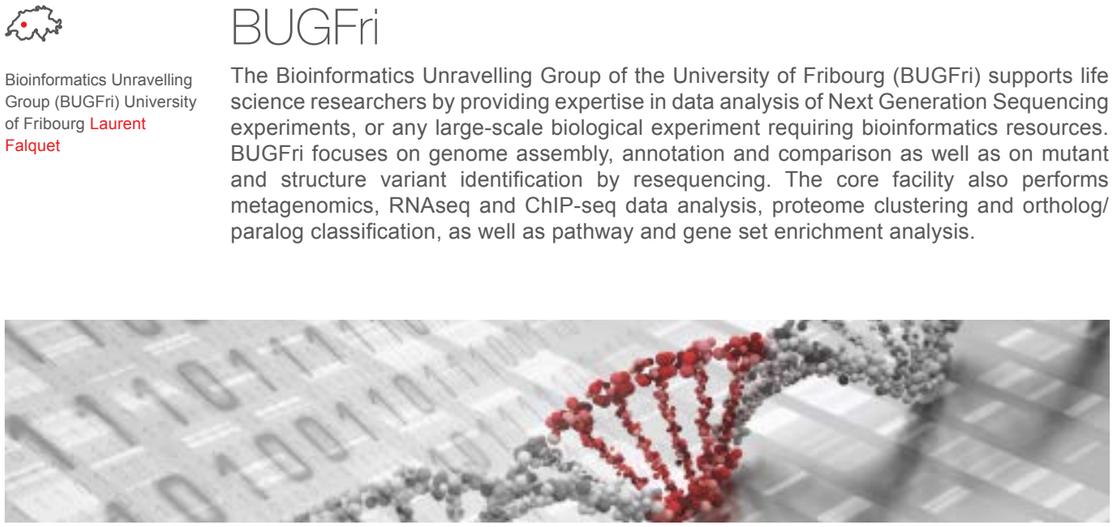
The Bioinformatics Core Facility (BCF) is a centre of excellence that provides state-of-the-art know-how for data analysis and discovery science. Specializing in biostatistical methods, clinical statistics, design of experiments, DNA/RNA Next Generation Sequencing data analysis and bioinformatics of high-throughput technologies, the core facility offers data analysis services and collaboration, statistical consulting, teaching and training aimed at supporting the Swiss and international life science community.



BBCF

Bioinformatics
and Biostatistics
Core Facility (BBCF)
EPFL, Lausanne
BBCF group

The Bioinformatics and Biostatistics Core Facility (BBCF) provides research labs with extensive support in bioinformatics and biostatistics. BBCF's 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 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 Unravelling Group (BUGFri) University of Fribourg Laurent Falquet

The Bioinformatics Unravelling Group of the University of Fribourg (BUGFri) supports life science researchers by providing expertise in data analysis of Next Generation Sequencing 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 ChIP-seq data analysis, proteome clustering and ortholog/paralog classification, as well as pathway and gene set enrichment analysis.

FGCZ-GI

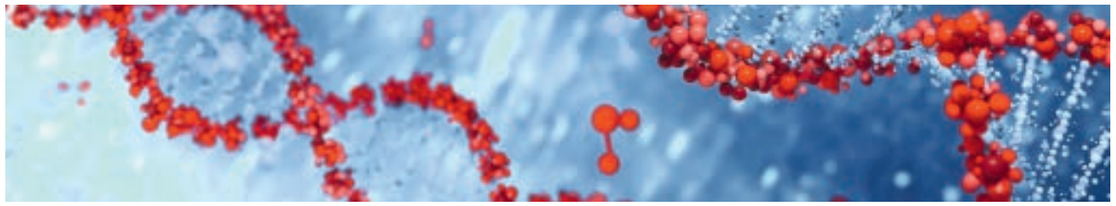
The Functional Genomics Centre Zurich Genome Informatics (FGCZ-GI) ETHZ / University of Zurich Hubert Rehrauer

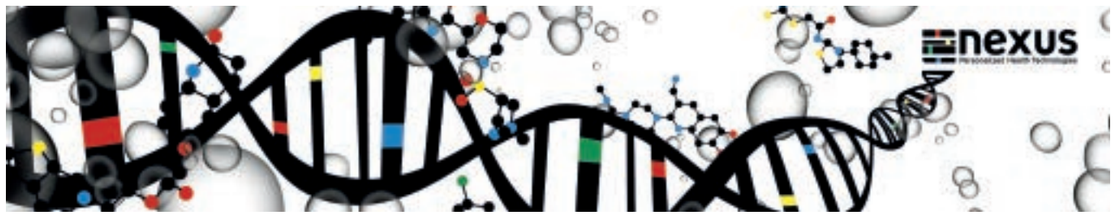
The Functional Genomics Centre Zurich Genome Informatics (FGCZ-GI) is dedicated to the processing, analysis and interpretation of Next Generation Sequencing data. FGCZ-GI interacts closely with research groups, and provides tailored comprehensive bioinformatics solutions. It also provides standard analysis pipelines for the more frequent research questions. The core facility team provides training to researchers and bioinformaticians on various aspects of data analysis, and access to its computing infrastructure for running analyses.

FMICBG

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

The FMI Computational Biology Group (FMICBG) helps biologists of the FMI in data analysis and visualization through collaboration, a bioinformatics helpdesk and training. 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 and genome assembly). Furthermore, IBU has its own research programme and collaborates on large and complex projects. It develops methods to analyse high-throughput data. The core facility also has a high-performance computing cluster and a data storage system that are used for IBU’s own research, collaboration and service projects.

NEXUS

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

NEXUS Personalized Health Technologies is an ETH Zurich Technology Platform that was 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 is built around two interdisciplinary technology units that are staffed by professional scientists. The Clinical Bioinformatics Unit (CBU) provides computational expertise to process large-scale heterogeneous data. The Theragnostics Discovery Unit (TDU) provides state-of-the-art robotic screening and analytic technologies in conjunction with chemical libraries, genome-scale gene manipulation tools and high-content cell imaging devices.

SIS

Scientific IT Services (SIS) ETH Zurich Bernd Rinn

The Scientific IT Services (SIS) is an interdisciplinary bioinformatics and scientific IT support group which builds computational tools. These tools range from lab databases to reusable framework components that enable and support both data analysis and data management 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, FAIRDOM, HPC-CH and swissuniversities’ eSCT/EnhanceR community. The core facility improves and ports scientific software, develops data management solutions and provides associated services. It also integrates and operates data analysis pipelines, and provides training and consulting in databases, scientific software development, high-performance and cloud computing.

Service and Support
for Science IT (S3IT)
University of Zurich
Marcel Riedi

S3IT

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 and 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.

In 2016, SIB welcomed a new core facility:

DBM Bioinformatics Core
Facility, Department of
Biomedicine (DBM)
University Hospital Basel
Robert Ivanek

DBM Bioinformatics

The DBM Bioinformatics Core Facility provides a centralized resource of expertise in computational biology and statistics to all researchers at the Department of Biomedicine. It helps scientists with the analysis, interpretation and visualization of expression, epigenetic and genomic data, mainly derived from Next Generation Sequencing experiments.

Technology coordination

Optimizing technology-related activities

The SIB Technology group supports the Institute and its groups by enabling knowledge and technology exchange and improved resource operation as well as by providing software engineering expertise. The group works in close cooperation with infrastructure providers and competence centres such as Vital-IT, Swiss-Prot and sciCORE.

Core competencies

- Design, development, testing and operation of scientific, technical and administrative software, 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, SIB internet domain, etc.
- Knowledge and technology exchange within SIB groups
- Support and operation of SIB-wide services developed and/or deployed by the group:
  - **ExPASy.org** (SIB's bioinformatics resource portal) as well as several resources available on the portal
  - Requesting tracking operations for user support
  - Applications for SIB Training activity, etc.
- Coordination of SIB's technical activities within **ELIXIR**

Highlights 2016

A specific focus was placed on web application testing and monitoring, which was put in place for several SIB resources. Additionally, the group supported major SIB-organized events such as the International Biocuration Conference, the SIB Days and the EPD Symposium.

Furthermore, to stimulate **technical knowledge exchange** between SIB groups, several internal SIB technical events took place in various cities in Switzerland. The group contributed to the following projects:

- **Beacon** interface extension and implementation on top of **arrayMap.org**, in collaboration with the Global Alliance for Genomics and Health (GA4GH) and the group led by Michael Baudis (Zurich)
- **MetaPIGA**: development of a web application on top of the MetaPIGA command line tool with the group of Michel Milinkovitch (Geneva) running on Vital-IT resources.

The SIB Technology group has been instrumental in implementing our GA4GH "Beacon" pilot study, and supported us in practical issues related to our server setup and monitoring."
Michael Baudis, Professor at the University of Zurich and Computational Oncogenomics Group Leader, SIB

Personalized health

Making the most out of health-related data

SIB is today playing a leading role in Switzerland's personalized health landscape, by sharing its expertise in bioinformatics with hospitals and by building the infrastructure that will enable patients' data to be used for research and to foster novel developments in personalized health.

Clinical bioinformatics: SIB's support for medical practice

In addition to data pertaining to a patient's lifestyle, eating habits and vital signs, medical practitioners are increasingly faced with complex molecular data such as their patients' genomic sequence, proteomic profile, or even metabolic profile.

The use of these data poses novel technical, analytical, ethical and educational challenges to both clinicians and scientists: clinicians have to learn how to handle and interpret this new type of data, and society must define where the boundaries of privacy lie.

In this context, there is a critical need **to bridge the gap between current medical practices and technology outputs.** To convert these outputs into clinically relevant knowledge is the aim of clinical bioinformatics. This particular application of bioinformatics is dedicated to the organization, analysis, interpretation and storage of data pertaining to an individual's state of health, which can be utilized by medical professionals.

In 2013, SIB created a Clinical Bioinformatics group, whose missions are:

- To establish consensus and common good practices for high-throughput omics data analysis in diagnostics across Swiss hospitals
- To establish trusted partnerships with Swiss public clinical institutions to develop, implement and sustain state-of-the-art approaches and tools for upcoming technologies and needs
- To provide harmonized clinical bioinformatics training across Switzerland, in collaboration with Swiss hospitals and universities
- To facilitate the interactions between SIB's research groups and the medical realm for clinical research projects.

Highlights 2016

An optimized Next Generation Sequencing (NGS) diagnostic pipeline for oncology was developed with the Molecular Pathology laboratory of the Geneva University Hospitals (HUG). OncoBench™ was specifically designed to improve sample tracking, somatic mutations identification, annotation and reporting for diagnostic purposes, in compliance with international data standards. It is currently being used for routine diagnostics in Molecular Pathology.

In addition, SIB Clinical Bioinformatics continued building close relationships with Swiss hospitals to foster harmonization of bioinformatics pipelines at the national level: the oncology and haemato-oncology working group is currently comparing hospitals' bioinformatics methods for clinical NGS. A second working group was launched to review hospitals' NGS practices in clinical microbiology. Both working groups include more than 40 participants from all over Switzerland.

> *Collaborating with SIB has been a clear positive for our laboratory. The OncoBench™ programme that we developed together has greatly simplified our workflows and increased our efficiency, while preserving the security of our patients' personal data."*
> **Dr Thomas A. McKee, Associate Physician and Unit Manager, Clinical Pathology Service, Geneva University Hospitals (HUG)**

Patient medical consultation

"Big data" generated by modern technologies, e.g. genetic sequencing

Data management, analysis and 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 lead research data infrastructure activities

In order to bring Switzerland to the forefront of research in personalized health, the Swiss State Secretariat for Education, Research and Innovation (SERI) has launched a national research initiative called the Swiss Personalized Health Network (SPHN). SPHN will establish nationwide interoperability of clinical, "omics" and other health-related data, allowing researchers in Switzerland to share information and collaborate efficiently. Since January 2017, SPHN brings together university hospitals, schools of higher education, research institutes and organizations working in the area of personalized health, as well as other health-related research activities across Switzerland. To achieve its goals, information generated at the various organizations and personalized health platforms will have to become mutually compatible ("semantically interoperable").

SIB plays a leading role in the SPHN initiative and the **Personalized Health Informatics (PHI) group** in particular is in charge of setting up and running two types of data infrastructure:

- **Data Coordination Centre (DCC):** The DCC will deal with data interoperability and data management nationwide to ensure that research projects can efficiently collaborate and share data across the various Swiss hospitals and research institutions. It will be coordinating the establishment of standards for data generation and annotation, data quality indicators, semantic interoperability and exchange formats;

- **BioMedIT:** This research infrastructure aims to establish a coordinated nationwide network of secure IT infrastructure at Swiss universities to support biomedical research in Switzerland. BioMedIT will form an integral part of SIB's contribution to the SPHN initiative – providing expertise, software workflows, and high-performance storage and computing resources for analysing and interpreting large volumes of clinical and omics data within SPHN.

Training

One of SIB's missions is to train the next generation of bioinformaticians and ensure that life and health scientists make the best use of bioinformatics resources, many of which are developed by SIB groups. SIB's Training group is in charge of promoting and coordinating training in bioinformatics, both in Switzerland and internationally.

30
SIB groups engaged in teaching activities

About
1,350
trainees

Over
100
experts and trainers

57
courses and workshops

112
teaching days

Highlights 2016

- **SIB PhD Training Network**
  The SIB PhD Training Network is dedicated to Swiss bioinformatics and computational biology PhD students. Special events in 2016 included the autumn school in "Bioinformatics and Population Genomics", jointly organized with Staromics and Ecology & Evolution CUSO doctoral programmes, and the course on "Machine Learning" jointly organized with SystemsX.ch.

- **Reaching out to companies**
  In 2016, the SIB Training team widely advertised its course offer to companies. Tailored to the needs of biotech and pharma companies and offered as on-site private courses, this customized training offer has already benefited companies such as Roche and Debiopharm. Several scientists from companies also attended the regular SIB courses.

- **Software Carpentry and Data Carpentry courses**
  An ELIXIR Software Carpentry and Data Carpentry instructors' training was hosted by SIB in 2016. Four SIB members obtained their certification as instructors and three user workshops were co-organized by SIB. These courses teach basic computing skills and tools enabling scientists to work more efficiently with data as well as with ELIXIR and SIB's infrastructure.

  Professional training : The SIB Training portfolio is constantly evolving to meet the scientific community's needs. Most popular courses are related to Next Generation Sequencing analysis, statistics and R, the statistical package. Find the full list of courses at www.sib.swiss/training.

- **International collaboration**
  In order to strengthen connections with the international and European bioinformatics training community, the SIB Training team once again had the pleasure of co-organizing the "Workshop in Education for Bioinformatics" for the International Society for Computational Biology. The SIB Training group also co-organized a workshop on "Training needs for Biocurators" with GOBLET and EMBL-EBI.

> *As a beginner in computational biology, SIB's training courses helped me greatly to conceptualize and develop custom algorithms for the analysis of my deep RNA sequencing data."*
> **Dr Oriane Mauger, Postdoctoral fellow, Biozentrum, University of Basel**

Outreach

Another of 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 activities to explain, in a playful and intuitive way, the key role it plays in life sciences and medicine today.

Over
2,500
participants in SIB activities for the layman

Over
1,300
participants in SIB activities in schools

Highlights 2016

Over 2,500 people of all ages took part in bioinformatics-related outreach activities and events.

- **Events around the classroom**
  Outreach activities at schools focused on SIB's Drug Design workshops. Thanks to the professional web-based tools developed by SIB, over 1,300 pupils and students, aged 12 - 19, discovered how bioinformatics is used to design a drug.

> *Now I understand how a drug works, the difficulties that can be encountered when designing a new drug and how bioinformatics can help."*
> **Dafine, 17 years old**

- **Training high school teachers**
  SIB was invited by the Swiss Chemical Society to conduct a Drug Design workshop for high school teachers in Zurich, during the "Future of Chemical Education" symposium. Two similar events were organized in Geneva and Lausanne for biology, chemistry and mathematics high school teachers.

- **Career guidance**
  The technologies involved in life science research are evolving rapidly, creating new professional opportunities, including in bioinformatics. SIB informed several groups of young people within the framework of career guidance programmes on these new prospects.

- **Popular events**
  SIB took part in the 2nd edition of "Planète Santé Live" at the SwissTech Convention Centre (EPFL) – a major health-related public event in Switzerland that hosted nearly 29,000 visitors this year. Participating in this event enabled SIB to introduce concepts such as drug design and clinical bioinformatics to a large audience.
  SIB was present at "La Nuit de la Science", a public event organized by the Science History Museum in Geneva, and at the "Mystères de l'UNIL", a public event organized by the University of Lausanne.

- **Collaboration with public outreach laboratories**
  Several of the workshops were conducted in collaboration with the Chimiscope, the Bioscope and "(R)amène ta Science" (University of Geneva), l'Eprouvette (University of Lausanne), the Swiss Chemical Society and SATW (TecDays).

Research

SIB has 65 bioinformatics groups and some 800 scientists from the major Swiss schools of higher education and research institutes.

It is SIB's mission to lead and coordinate the field of bioinformatics in Switzerland, and to bring world-class researchers together.

SIB fosters collaboration and innovation at the highest level of scientific excellence

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

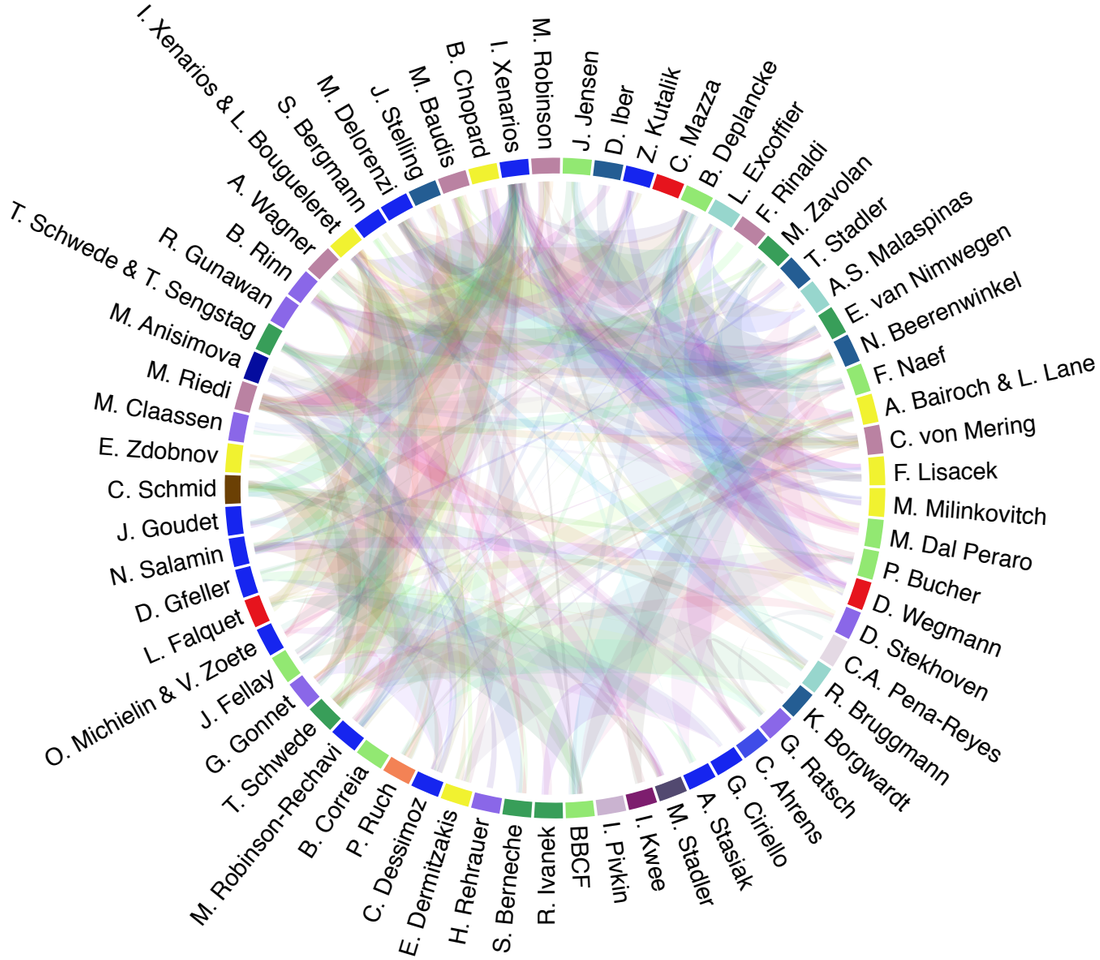
- **In Europe**: the European Bioinformatics Institute (EMBL-EBI), the Bioinformatics Services to Swedish Life Science (BILS), the Spanish National Bioinformatics Institute (INB) and the Dutch Techcentre for Life Sciences (DTL)
- **In the US**: the National Institutes of Health (NIH), the National Center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

SIB Awards:

Since 2008, SIB has been honouring young researchers and ground-breaking resources on the national and international level through the SIB awards. Every two years, SIB selects the best submissions in the following categories:
- International Young Bioinformatician Award
- Bioinformatics Resource Innovation Award
- Best Swiss Bioinformatics Graduate Paper Award.

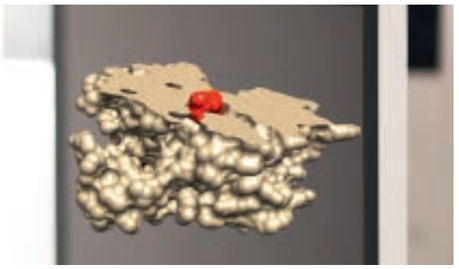
The SIB Awards 2017 will be presented in September 2017 during the 13th [BC]² Basel Computational Biology Conference.

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



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

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 the SIB groups are therefore involved in several 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

Research areas:
■ Main area
□ Secondary area(s)

	GENES AND GENOMES pp. 45-52	PROTEINS AND PROTEOMES pp. 53-56	MEDICINE AND HEALTH pp. 57-62	EVOLUTION AND PHYLOGENY pp. 63-69	STRUCTURAL BIOLOGY pp. 70-72	SYSTEMS BIOLOGY pp. 73-79	BIOINFORMATICS INFRASTRUCTURE pp.80-83
Sven Bergmann	■		□				□
Rémy Bruggmann	■	□	□			□	□
Philipp Bucher	■		□				□
Bart Deplancke	■		□	□		□	□
Emmanouil Dermitzakis	■		□			□	
Laurent Falquet	■	□	□				□
Robert Ivanek	■		□			□	□
Zoltán Kutalik	■		□	□			□
Anna-Sapfo Malaspinas	■		□	□			
Erik van Nimwegen	■			□		□	□
Hubert Rehrauer	■		□			□	□
Mark D. Robinson	■		□				
Michael Stadler	■		□			□	□
Andrzej Stasiak	■	□	□		□		
Evgeny Zdobnov	■	□	□	□			□
Christian Ahrens	□	■				□	□
Amos Bairoch and Lydie Lane		■	□				□
Bruno Correia		■			□		
Frédérique Lisacek		■	□			□	□
Christian von Mering	□	■		□		□	□
Ioannis Xenarios and Lydie Bougueleret	□	■	□	□	□	□	□
Michael Baudis	□		■			□	□
Niko Beerenwinkel	□		■	□		□	□
Mauro Delorenzi	□		■			□	□
Jacques Fellay	□		■			□	□
David Gfeller	□	□	■		□		
Ivo Kwee	□		■			□	
Carlos-Andrés Peña-Reyes	□	□	■			□	□
Gunnar Rätsch	□		■				□
Patrick Ruch		□	■				□
Christoph Schmid	□		■				
Maria Anisimova	□	□		■			□
Christophe Dessimoz	□	□		■			□
Laurent Excoffier	□			■			□
Gaston Gonnet	□	□		■			□
Jérôme Goudet	□			■			□
Jeffrey D. Jensen	□			■			□
Marc Robinson-Rechavi	□			■			□
Nicolas Salamin	□			■			□
Tanja Stadler	□		□	■			□
Andreas Wagner	□			■			
Daniel Wegmann	□		□	■			□
Simon Bernèche		□			■		
Matteo Dal Peraro		□	□		■		□
Olivier Michielin and Vincent Zoete		□	□		■	□	□
Torsten Schwede		□			■		□
Karsten Borgwardt	□		□			■	
Bastien Chopard			□			■	□
Giovanni Ciriello			□			■	
Manfred Claassen			□			■	
Rudiyanto Gunawan	□		□			■	□
Dagmar Iber			□			■	□
Christian Mazza	□					■	
Michel Milinkovitch	□			□		■	□
Félix Naef	□	□	□			■	
Igor V. Pivkin			□			■	
Jörg Stelling						■	□
Mihaela Zavolan	□	□	□			■	
BBCF	□			□		□	■
Marcel Riedi	□	□	□	□	□	□	■
Fabio Rinaldi			□			□	■
Bernd Rinn							■
Torsten Schwede and Thierry Sengstag	□	□	□	□	□	□	■
Daniel Stekhoven			□				■
Ioannis Xenarios	□	□	□	□	□	□	■

New groups

What do we do?
The Computational Systems Oncology lab combines algorithmic design, numerical modelling, and molecular biology approaches to address relevant questions in cancer biology and therapeutics.

Highlights 2016
During 2016 our group designed new tools to investigate cancer genetic and epigenetic determinants, to understand how distinct evolutionary dynamics give rise to intratumoral heterogeneity and how these can reveal functional and clinically relevant dependencies between oncogenic alterations.

What do we do?
Our group is driven by the passion of expanding nature's repertoire by designing novel functional proteins to be used for practical purposes such as therapeutics, vaccines and biosensors. The broader vision of our research is to characterize our newly designed proteins at several levels: *in silico*, *in vitro* and, if applicable, *in vivo*.

Highlights 2016
Our group was awarded an ERC starting grant to support our algorithmic developments in the computational design of *de novo* proteins.
We made important advances on a number of computational tools, including an updated version of the Rosetta Fold From Loops protocol and web-based tool to enable the design of *de novo* proteins for non-expert users.
Using our new methodologies, we designed several novel proteins to serve as immunogens for the development of an RSV vaccine. We characterized these proteins experimentally and started animal studies. These novel technologies will soon be used for the design of immunogens for other important pathogens.

Main publications 2016

Blackburn MC *et al.* Integrating gene synthesis and microfluidic protein analysis for rapid protein engineering. Nucleic Acids Res. 2016; 44 (7): e68.
Silva DA *et al.* Motif-Driven Design of Protein-Protein Interfaces. Methods Mol Biol. 2016; 1414:285-304.
Backus K *et al.* Proteome-wide covalent ligand discovery in native biological systems. Nature 2016; 534, 570-4.
Mousa JJ *et al.* Structural basis for nonneutralizing antibody competition at antigenic site II of the respiratory syncytial virus fusion protein. Proc Natl Acad Sci U S A 2016; 113, E6849-E6858.
Matthews ML *et al.* Chemoproteomic profiling and discovery of protein electrophiles in human cells. Nat Chem. 2017; 9, 234–243.
Sarti E *et al.* Protein-protein structure prediction by scoring molecular dynamics trajectories of putative poses. Proteins 2016; 84(9):1312-20.

What do we do?
Our group is located in the Department of Biomedicine (DBM) at the University of Basel. We are collaborating with scientists from DBM on projects covering a broad spectrum of research topics, from cellular differentiation and evolutionary biology to the identification of a molecular basis for various human diseases. To do this, we focus on the analysis, interpretation and visualization of expression, epigenetic and genomic data, which are mainly derived from Next Generation Sequencing experiments.

What do we do?
Ancient and modern DNA research have both entered the genomics era. At the Evolutionary Genomics Group (EGG)/Computational Paleogenomics Group (CPG), we aim to characterize evolutionary processes (genetic drift, natural selection, migration and mutation) using genomics data from both modern and ancient samples. We develop analytical and computational methods to analyse and interpret time sampled data and we apply those methods to novel ancient DNA datasets via collaborative projects. Our work should allow us to quantify and time adaptive and migration events – notably related to the human colonization of the world – while generating unique datasets.

Main publications 2016

Malaspinas AS *et al.* A genomic history of Aboriginal Australia. Nature 2016; 538(7624):207-214.
Malaspinas AS. Methods to characterize selective sweeps using time serial samples: an ancient DNA perspective. Mol Ecol. 2016; 25(1):24-41.

Gunnar Rätsch
Biomedical Informatics
ETH Zurich

What do we do?
We are interested in modern machine learning techniques suitable for the analysis of problems that arise in medicine and biology. In particular, we develop new learning techniques that are capable of dealing with large amounts of genomic data, allow very accurate predictions of the phenomenon at hand and are able to comprehensibly provide reasons for their prognoses, and thereby assist in gaining new biomedical insights.

Highlights 2016
In May 2016, our group moved from MSKCC in New York to ETH Zurich. We are excited about the new environment and the newly formed collaborative relationships with SIB, the Max Planck-ETH Center for Learning Systems and the University Hospital Zurich. In 2016 we started a new project on distributed storage and computing on very large graph genomes, an effort that will be funded by the National Research Programme 75 “Big Data”. We are also heavily involved in the International Cancer Genome Consortium (ICGC) Pan Cancer Analysis Working Group and The Cancer Genome Atlas (TCGA) PanCanAtlas project. The lab is actively involved in the BRCA Challenge, a demonstration project of the Global Alliance for Genomics and Health (GA4GH). The BRCA Challenge aims to advance understanding of the genetic basis of breast, ovarian, and other cancers that are driven by germline variants in BRCA1 and BRCA2 (BRCAexchange.org). In this context, we presented the project at a session of the Human Variome Project at UNESCO in Paris and at the ASHG session in Vancouver. Gunnar Rätsch participates in and leads multiple efforts related to the Swiss Personalized Health Network.

Main publications 2016
Hyland SL *et al.* Knowledge transfer with medical language embeddings. In: Proc. Data Mining for Medicine and Healthcare, May 5-7, 2016, Miami.
Zhong Y *et al.* RiboDiff: detecting changes of mRNA translation efficiency from ribosome footprints. Bioinformatics 2017; 33(1):139-141.
Hartmann L *et al.* Alternative splicing substantially diversifies the transcriptome during early photomorphogenesis and correlates with the energy availability in *Arabidopsis*. Plant Cell 2016; 28(11):2715-2734.

Fabio Rinaldi
BioMeXT: Biomedical Information Extraction
University of Zurich

What do we do?
The BioMeXT group specializes in Information Extraction (IE) from the biomedical literature, as well as other textual sources. Information extraction involves automatically extracting structured information from textual documents, and is an important component of Text Mining systems. We specialize in the extraction of domain-specific entities (such as genes, proteins, drugs, diseases) and their semantic relationships (e.g. protein-protein interactions, gene-disease associations). Our tools are often evaluated through participation in community-run evaluation challenges (e.g. BioCreAtIvE: Critical Assessment of Information Extraction systems in Biology). Additionally, we provide an environment for Assisted Curation (ODIN), which is currently being used in the curation pipeline of the RegulonDB database in a project funded by the US-NIH.

Highlights 2016
The SNSF-funded MelanoBase project began in March 2016. The goal of the project is a large-scale extraction of information from the biomedical literature in order to build a disease-centric knowledgebase of information relevant for biological and clinical purposes. Later in the year, a collaboration was established with the FBK research institute in Trento (Italy) in order to pursue advanced natural language processing techniques relevant for the MelanoBase project. In December 2016 the collaborative project “Automated detection of adverse drug events from older inpatients’ medical records using structured data mining and natural language processing”, submitted within the “Smarter Health Care” National Research Programme (NRP74) was approved. We will participate in this project, contributing natural language processing technologies for the automated analysis of medical records.

Main publications 2016
Wang Q *et al.* Overview of the interactive task in BioCreative V. Database (Oxford) 2016: baw119.
Mazzocut M *et al.* Insight of Italian Web conversations in Complementary and Alternative Medicines and Cancer. J Med Internet Res. 2016; 18(6): e120.
Rinaldi F *et al.* BioCreative V Track 4: A shared task for the extraction of causal network information in biological expression language. Database (Oxford) 2016: baw067.

Genes and genomes

The genome is the sum of genetic material, including genes, inherited by an organism. It contains all the information needed to build and maintain this living being.

Aberrations in genetic material can be at the heart of diseases such as cancer or Down’s syndrome.

Bioinformatics develops tools to read the genetic information, store the resulting data, analyse and interpret them.

Sven Bergmann
Computational Biology Group
University of Lausanne

What do we do?
At the Computational Biology Group we develop concepts and algorithmic tools for the analysis of large-scale biological and clinical data. We participate in many genome-wide association studies (GWAS) for human traits and have a particular interest in the integration of genotypic and complex phenotypic datasets (such as gene expression or metabolomics). A key approach is the reduction of complexity through modular and network analysis. A complementary line of our research pertains to relatively small genetic networks whose components are well known.

Highlights 2016
In 2016, we studied the gene expression response to drugs affecting heart rate and blood pressure across a panel of genetically different mice [*BMC Genomics* 17:717]. Using our computational software PASCAL for fast and accurate computation of gene and pathway scores from SNP-wise association statistics [*PLoS Genet*. 12(1):e1005616], we analysed data from the FANTOM5 project, revealing that genetic variants associated with different diseases can be used to identify the relevant tissues enriched for genetic networks that are perturbed by variants [*Nature Methods* 13:366]. We also used PASCAL in combination with a large panel of GWAS data for the evaluation of modules that were identified from a set of gene networks by participants of a DREAM challenge we organized.

Main publications 2016

Prunotto A *et al.* RNAseq analysis of heart tissue from mice treated with atenolol and isoproterenol reveals a reciprocal transcriptional response. BMC Genomics 2016; 17:717.

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 2016; 13:366.

Rémy Bruggmann
Interfaculty Bioinformatics Unit (IBU)
University of Bern

What do we do?
At the Interfaculty Bioinformatics Unit of the University of Bern (IBU), we 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 and genome assembly). Furthermore, we have our own research programme and collaborate on large and complex projects. We develop methods to analyse high-throughput data. We have a high-performance computing cluster and a data storage system that we use for our own research, collaboration and service projects.

Highlights 2016
In 2016, we participated in several whole genome sequencing projects. In one project (published in *Nature Plants*), we sequenced the two parental wild species of *Petunia hybridae* (i.e. *P. axilaris* and *P. inflata*), a popular bedding plant with a long history as a genetic model system. The factors responsible for the shift from bee to moth pollination reside in very dynamic regions of the genome, which may have been essential to the diversity of floral colour patterns and pollination systems. Another important project (published in *Nature Communications*) is about the gustatory receptor neurons (GRNs) of *Drosophila* larvae and included the generation and analysis of low input amount RNA-seq data. Interestingly, GRNs employ a remarkably different mode of gustatory information coding. We identified a multimodal GRN that responds to chemicals of different taste modalities with opposing valence. This multimodal neuron is essential for bitter compound avoidance, and its artificial activation is sufficient to mediate aversion. Our findings support a model for taste coding in larvae, in which distinct receptor proteins mediate different responses within the same multimodal GRN.

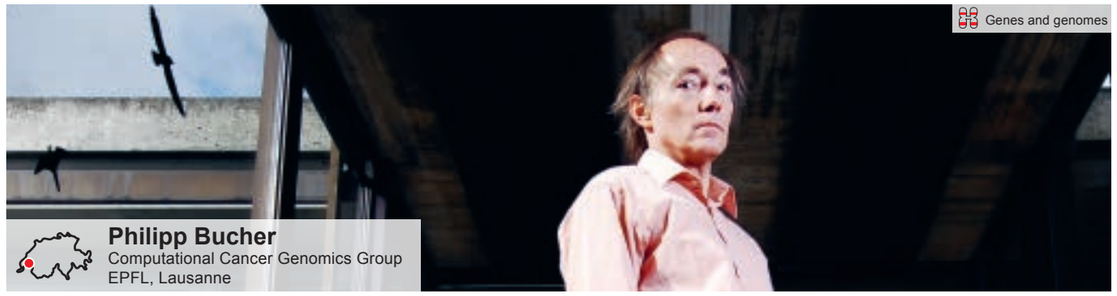
Main publications 2016

Bombarely A *et al.* Insight into the evolution of the *Solanaceae* from the parental genomes of *Petunia hybrida*. Nat Plants 2016; 2(6), 16074.

van Giesen L *et al.* Multimodal stimulus coding by a gustatory sensory neuron in *Drosophila* larvae. Nat Commun. 2016; 7, 10687.

Miclaus M *et al.* Maize cytolines unmask key nuclear genes that are under the control of retrograde signaling pathways in plants. Genome Biol Evol. 2016; 8(11), 3256–3270.

Schlegel M *et al.* Globally distributed root endophyte *Phialocephala subalpina* links pathogenic and saprophytic lifestyles. BMC Genomics 2016; 17(1):1015.



Philipp Bucher
Computational Cancer Genomics Group
EPFL, Lausanne

What do we do?
At the Computational Cancer Genomics Group, we are interested in gene regulation in both 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. Our group develops new algorithms, computer programmes, web services and databases that will help us and others to extract knowledge and understanding from such data.

Highlights 2016
EPD news:
- Promoter collections for two new model organisms were added: honey bee and maize. The human promoter collection now has more than 25,000 entries after the processing of new CAGE data from FANTOM5.
- A joint paper with Bernard Moret's group, entitled "A Maximum-likelihood approach for building cell-type trees by lifting", received the best paper award at the 2016 Asia Pacific Bioinformatics Conference in San Francisco.
- To mark the 30th anniversary of the Eukaryotic Promoter Database EPD, we organized a symposium entitled "Promoter Research: Past, Present and Future" at the Starling hotel on the EPFL campus. We had outstanding talks by speakers from three continents, and lively discussions among all participants.
- At the SIB Days 2016, our workshop entitled "Facilitating Reproducibility of Computational Research in Bioinformatics" attracted over 50 participants.
- An article presenting our ChIP-Seq tools was published in *BMC Genomics*

Main publications 2016

Kumar S and Bucher P. Predicting transcription factor site occupancy using DNA sequence intrinsic and cell-type specific chromatin features. BMC Bioinformatics 2016;17 Suppl 1:4.

Dreos R *et al.* Influence of rotational nucleosome positioning on transcription start site selection in animal promoters. PLoS Comput Biol. 2016; 12(10):e1005144.

Ambrosini G *et al.* The ChIP-Seq tools and web server: a resource for analyzing ChIP-seq and other types of genomic data. BMC Genomics 2016; 17(1):938.

Bart Deplancke
Laboratory of Systems Biology and Genetics
EPFL, Lausanne

What do we do?
At the Laboratory of Systems Biology and Genetics (LSBG), we are using high-throughput sequencing, single cell genomics, microfluidics, and computational approaches 1) to decipher the regulatory code in *Drosophila* and mammals with a specific focus on mesenchymal stem cell function, adipose biology and gut immunity, and 2) to examine how variations in this code affect molecular and organismal diversity. In addition to our research interests, we are actively pursuing the development of new research tools and computational pipelines that enable better characterization of gene regulatory networks.

Highlights 2016
Understanding the DNA binding behaviour of transcription factors (TFs) is critical for elucidating the transcriptional logic in a cell and for uncovering how genomic variation affects gene regulatory processes. Despite tremendous efforts to define the DNA binding specificities of TFs, less than half of all human TFs have so far been experimentally characterized, and this situation is even worse when considering obligate or facultative heterodimers. To address this data lacuna, we developed a novel digital microtechnology (SMiLE-seq) aimed at deriving quantitative DNA binding models of single and dimeric TFs that belong to different structural families. Furthermore, using a comparable microfluidic platform and in collaboration with the Hatzimanikatis Lab (EPFL), we performed a comprehensive study to quantify cooperativity between TFs that form heterodimers. The resulting data allowed us to build a mechanistic model that accounted for all possible intermediate and final complexes that can occur between two TFs and DNA. One of the main findings that emerged using this model is that the nucleotide composition of the heterodimer binding site has an important impact on the extent of DNA binding cooperativity.

Main publications 2016

Deplancke B *et al.* The genetics of transcription factor DNA binding variation. Cell 2016; 166:538.

Isakova A *et al.* Quantification of cooperativity in heterodimer-DNA binding improves the accuracy of binding specificity models. J Biol Chem. 2016; 291:10293.

Isakova A *et al.* SMiLE-seq identifies binding motifs of single and dimeric transcription factors. Nat Methods 2017; 14(3):316-322.

Emmanouil (Manolis) Dermitzakis

Genomics of Complex Traits Group
University of Geneva

What do we do?

At the Genomics of Complex Traits Group we have a strong interest in population genomics and genetics of complex traits. We are using various methodologies to understand the role of genetic variation in phenotypic variation. We also aim to understand what fraction of genetic variation is harboured within functional elements of the human genome. Our main focus is on genome-wide analysis of gene expression and cellular phenotypes and their association with nucleotide variation. We 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 2016

During this past year we have engaged in large-scale analysis of GTEx data, multi-omics cohort data and system genetics analysis, and produced a number of manuscripts that are on their way for publication. Our group has also moved to cancer genomics more intensively, with a focus on the contribution of non-coding regulatory DNA to cancer predisposition and progression.

Main publications 2016

Ongen H *et al.* Fast and efficient QTL mapper for thousands of molecular phenotypes. Bioinformatics 2016; 32(10):1479-85.

Glastonbury CA *et al.* Adiposity-dependent regulatory effects on multi-tissue transcriptomes. Am J Hum Genet. 2016; 99(3):567-79.

Lotta LA *et al.* Integrative genomic analysis implicates limited peripheral adipose storage capacity in the pathogenesis of human insulin resistance. Nat Genet. 2017; 49(1):17-26.

Laurent Falquet

Bioinformatics Unravelling Group (BUGFri)
University of Fribourg

What do we do?

At the Bioinformatics Unravelling Group of the University of Fribourg (BUGFri), we 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. We focus on genome assembly, annotation and comparison as well as on mutant and structure variant identification by resequencing. We also perform metagenomics, RNAseq and ChIP-seq data analysis, proteome clustering and ortholog/paralog classification, as well as pathway and gene set enrichment analysis.

Highlights 2016

At the beginning of 2016, the Sinergia project “Pangenomic and comprehensive analysis of the relationship between bacterial toxin-antitoxin systems and antibiotic phenotype” began. We are now working towards improving the analysis and description of toxin-antitoxin systems (TAs) using both NGS data analysis to identify cut sites of some toxins and public Tn-Seq data to identify new TAs. The web server PACMAN (PACific biosciences Methylation Analyzer) was also improved. 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 by making a choice from 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.

Main publications 2016

Ardissone S *et al.* Cell cycle constraints and environmental control of local DNA hypomethylation in α-proteobacteria. PLoS Genet. 2016; 12: e1006499.

Leake SL *et al.* The salivary microbiome for differentiating individuals: proof of principle. Microbes Infect. 2016; 18: 399-405.

Ritpitakphong U *et al.* The microbiome of the leaf surface of *Arabidopsis* protects against a fungal pathogen. New Phytol. 210: 1033-1043.

Zoltán Kutalik

Statistical Genetics Group
CHUV / University of Lausanne

What do we do?

At the Statistical Genetics Group we are interested in the development of statistical methodologies in order to decipher the genetic architecture of complex human traits related to obesity. To do this, we efficiently combine genome-wide association studies (GWAS) with different -omics data to enhance our understanding of the genetic network of the human genome. We are also very involved in the activities of the GIANT consortium as well as in various clinical genetic analyses.

Highlights 2016

During 2016, Aurélien Macé and Jing Cui successfully defended their PhD theses, and Sina Rüeger (PhD student) received the Young Investigator Award for the best talk in statistical genetics at the European Society of Human Genetics (ESHG) conference. We developed a new software (PASCAL) for fast and rigorous computation of gene and pathway scores from SNP-based summary statistics. Another tool to reliably call and associate Copy Number Variants (CNVs) was also published. Finally, we participated in major collaborative efforts (published in *Nature, Nature Genetics* and *Nature Communications*) to unravel the genetic basis of birth weight, educational attainment, blood pressure and leptin levels.

Main publications 2016

Macé A *et al.* New quality measure for SNP array based CNV detection. Bioinformatics 2016; 32(21):3298-3305.

Marbach D *et al.* Tissue-specific regulatory circuits reveal variable modular perturbations across complex diseases. Nat Methods 2016;13(4):366-70.

Tafti M *et al.* Narcolepsy-associated HLA class I alleles implicate cell-mediated cytotoxicity. Sleep 2016; 39(3):581-7.

Erik van Nimwegen

Genome Systems Biology Group
University of Basel

What do we do?

Our main research interest at the Genome Systems Biology (GSB) Group is the study of genome-wide regulatory systems in order 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, we are 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 2016

Our first main highlight for 2016 was the completion of our integrated microfluidic and image-analysis setup for studying gene regulation *in vivo* at the single-cell level. Our setup, consisting of a dual-input microfluidic chip and accompanying analysis software, allows automated and highly accurate tracking of the growth and gene expression of lineages of single cells as they respond to continuously changing external conditions. The analysis software, which we developed in collaboration with the group led by Gene Myers, jointly optimizes both segmentation and tracking, and includes a highly novel curation procedure – called ‘leveraged editing’ – in which a single input directive can fix up to a dozen errors. Applying this methodology to the founding system of studies in gene regulation – i.e. induction of the lac operon in response to a switch of carbon source from glucose to lactose – we discovered that single-cell lag times have a multi-modal distribution and that lag times are controlled by a (as yet unknown) heritable factor.
A second highlight was the publication of our manuscript on our CRUNCH pipeline for completely automated analysis of ChIP-seq data. CRUNCH performs all ChIP-seq analysis steps, from quality analysis and mapping of the raw reads to the comprehensive *de novo* motif finding and annotation of binding sites in all ChIP peaks. Finally, this year we also completed our development of a new general motif model, called Dinucleotide Weight Tensor (DWT), which incorporates arbitrary dependencies between positions within regulatory sites. In our recently submitted manuscript we show that DWTs, which have no tunable parameters whatsoever, always perform at least as well as position-specific weight matrices, and strongly outperform them for a substantial fraction of transcription factors.

Main publications 2016

Berger S *et al.* Crunch: Completely automated analysis of ChIP-seq data. bioRxiv 2016; 042903.

Kaiser M *et al.* Tracking single-cell gene regulation in dynamically controlled environments using an integrated microfluidic and computational setup. bioRxiv 2016; 076224.

Omidi S and van Nimwegen E. Automated incorporation of pairwise dependency in transcription factor binding site prediction using dinucleotide weight tensors. bioRxiv 2016; 078212.

Hubert Rehrauer

Functional Genomics Center Zurich
ETH Zurich, University of Zurich

What do we do?

We are dedicated to the processing, analysis and interpretation of Next Generation Sequencing (NGS) data. We interact closely with research groups, and provide tailored comprehensive bioinformatics solutions. We also provide standard analysis pipelines for the more frequent research questions. We train researchers and bioinformaticians on various aspects of data analysis, and provide access to our computing infrastructure for running analyses.

Highlights 2016

In 2016, we made our web-based NGS data analysis platform fully available for direct self-managed data exploration and analysis. With this convenient web access, researchers can leverage a comprehensive set of NGS analysis tools and an associated high-performance computing backend for their omics-related research. Additionally, thanks to our bioinformatics analysis we made a few key contributions towards a better understanding of disease mechanisms and potential therapeutic options. Finally, with the use of long reads generated by FGCZ's PacBio platform, we were able to contribute to various genome assemblies.

Main publications 2016

Hatakeyama *et al.* SUSHI: an exquisite recipe for fully documented, reproducible and reusable NGS data analysis. BMC Bioinformatics 2016; 17(1):228.

Kulig *et al.* IL-12 protects from psoriasiform skin inflammation. Nat Commun. 2016; 7:13466.

Izuno *et al.* Genome sequencing of *Metrosideros polymorpha* (Myrtaceae), a dominant species in various habitats in the Hawaiian Islands with remarkable phenotypic variations. J Plant Res. 2016; 129(4):727-36.

Mark D. Robinson

Statistical Bioinformatics Group
University of Zurich

What do we do?

At the Statistical Bioinformatics Group we develop robust data analysis solutions, including new or improved methods, for the analysis of genome-scale data. We 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. We are largely data- and problem-driven, and ultimately the methods we develop cater to the characteristics of the technology platform generating the data. We develop publicly available open-source software tools, generally through the Bioconductor project. The majority of our time is spent on collaborative projects and development of statistical methods with accompanying software. Where needed, we design experiments and collect data to compare the performance of competing methods and platforms.

Main publications 2016

Weber LM and Robinson MD. Comparison of clustering methods for high-dimensional single-cell flow and mass cytometry data. Cytometry A 2016; 89 (12), 1084-1096.

H Lindsay *et al.* CrispRVariants charts the mutation spectrum of genome engineering experiments. Nat Biotechnol. 2016; 34 (7), 701-702.

Soneson C and MD Robinson. iCOBRA: open, reproducible, standardized and live method benchmarking. Nat Methods 2016; 13 (4), 283.

Michael Stadler

FMI Computational Biology Group
Friedrich Miescher Institute, Basel

What do we do?

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

Highlights 2016

In most cells, the DNA is packaged around histone proteins. However, mammalian sperm cells have a highly compacted genome, and during maturation they evict the large majority of their histone proteins from the DNA. Whether the histones that are left carry epigenetic marks and could therefore serve as a carrier for epigenetic information across generations remains an open question. In a re-analysis of published datasets, we were able to clarify former contradicting reports on the subject, which associated retained histones with repeat elements, although open questions still remain in this intriguing research field.

Main publications 2016

Habacher C *et al.* Ribonuclease-mediated control of body fat. Dev Cell 2016; 39(3):359-369.

Miki TS *et al.* XRN2 Autoregulation and control of polycistronic gene expression in *Caenorhabditis elegans*. PLoS Genet. 2016; 12(9):e1006313.

Eymery A *et al.* The methyltransferase Setdb1 is essential for meiosis and mitosis in mouse oocytes and early embryos. Development. 2016; 143(15):2767-79.

Royo H *et al.* Alternative computational analysis shows no evidence for nucleosome enrichment at repetitive sequences in mammalian spermatozoa. Dev Cell. 2016; 37(1):98-104.

.

Andrzej Stasiak

DNA and Chromosome Modelling Group
University of Lausanne

What do we do?

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

Highlights 2016

- Using the Metropolis Monte-Carlo approach, we simulated supercoiled DNA molecules that were also knotted or catenated. The analysis of the configurations obtained suggested a geometric selection mechanism permitting bacterial DNA topoisomerases to efficiently decatenate freshly replicated DNA. Our study may help in the design of antibiotics that target bacterial topoisomerases.
- In a collaboration with researchers in Poland, we used bioinformatics tools to analyse proteins that form deep, tight knots. We observed that knotted cores in these proteins have somewhat unusual properties. These regions show an increased number of inter-residue contacts, have high thermal stability and low solvent accessibility.
- Due to our numerous publications in which we simulated DNA molecules, chromatin fibres and topoisomerases, we were invited by *Methods in Molecular Biology* to formally present various simulation methods used by our group in the form of protocols.

Main publications 2016

Rawdon EJ *et al.* How topoisomerase IV can efficiently unknot and decatenate negatively supercoiled DNA molecules without causing their torsional relaxation. Nucleic Acids Res. 2016; 44(10): 4528-38.

Dabrowski-Tumanski P *et al.* In search of functional advantages of knots in proteins. PLoS One 2016; 11, e0165986.

Racko D *et al.* Molecular dynamics simulations of supercoiled, knotted and catenated DNA molecules including modelling of action of DNA gyrase. Methods Mol Biol. 2016. In press.

Evgeny Zdobnov
Computational Evolutionary Genomics Group
University of Geneva

What do we do?
At the Computational Evolutionary Genomics Group we are active in the fields of comparative genomics and shotgun metagenomics. We study molecular evolution, develop approaches to genomics data analyses, and implement computational pipelines. We apply evolutionary models to digest sequencing data, and revise these models using novel data. We study functional genomic elements on the basis of sequence variability among different species and within populations. Our interests range from arthropod genomics, including invertebrate vectors of human pathogens, to the evolution of viruses and clinical microbiology.

Highlights 2016
In 2016, we completed the v9.1 update of our database of orthologs, OrthoDB, which provides evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral genes. As well as a further increase in the coverage of organisms and the depth of collected annotations, the hierarchical catalogue of orthologs now features a comparative chart generator, as well as online BUSCO analysis and OrthoDB mapping for user uploaded sequences. The remarkable rate of adoption of our BUSCO tool for quantitative assessment of completeness of genome assemblies, gene sets, or transcriptomes (cited 162 times in one and a half years!) prompted us to release version 2 in 2016. The update includes significant expansion of the underlying data sets, now providing 44 clade-specific BUSCOs, and a major revision of the software implementing the procedure. The software is now distributed through GitLab. It is also available as an Ubuntu virtual machine, as well as being integrated as an online service for the OrthoDB logged-in users.
In 2016 we also finished the sequencing of several insects in our lab. The first manuscript, in print, is on the genomic features of the damselfly *Calopteryx splendens*, which represents a sister clade to most insect orders.

Main publications 2016
Zdobnov EM *et al.* OrthoDB v9.1: cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. Nucleic Acids Res. 2016; 5(D1):D744-D749.
Cordey S *et al.* Astrovirus MLB2, a new gastroenteric virus associated with meningitis and disseminated infection. Emerg Infect Dis. 2016; 22(5):846-53.
Dousse A *et al.* CEGA--a catalog of conserved elements from genomic alignments. Nucleic Acids Res. 2016; 44(D1):D96-100.

Proteins and proteomes

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.
The proteome describes the entire set of proteins expressed by a cell, a tissue or an organism at a given time.

A mutation in its gene can alter a protein's function, thereby causing diseases such as cystic fibrosis or Creutzfeldt-Jakob.

Bioinformatics develops tools to understand how proteins exercise their role.

Christian Ahrens

Bioinformatics and Proteogenomics Group
Agroscope, Wädenswil

What do we do?

Our research at the Bioinformatics and Proteogenomics Group revolves around the bioinformatic integration and analysis of datasets from state-of-the-art omics technologies, which we obtain through close collaboration with experimental biologists. These datasets include genome sequences, 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, we started to study the role of microbiomes – e.g. for plant protection – by applying metagenomics, genomics and transcriptomics approaches.

Highlights 2016

As part of a research consortium led by Professor J.A. Robinson (Chemistry Dept., UZH) and Professor L. Eberl (Microbiology Dept., UZH), our team provided its 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) active against Gram-negative pathogens, exhibits a novel mode of action (MoA). In this case, it selectively ruptures the outer membrane of Gram-negative bacteria.

Furthermore, our proteogenomics expertise helped to uncover protein expression evidence for novel, unannotated short coding sequences in the rhizobial model organism *Bradyrhizobium japonicum*. A current focus of the group is to develop a general approach to identify such missed protein-coding genes in prokaryotes.

Main publications 2016

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.

Turnbull L *et al.* Explosive cell lysis as a mechanism for the biogenesis of bacterial membrane vesicles and biofilms. Nat Commun. 2016; 7:11220.

Čuklina J *et al.* Genome-wide transcription start site mapping of *Bradyrhizobium japonicum* grown free-living or in symbiosis - a rich resource to identify new transcripts, proteins and to study gene regulation. BMC Genomics 2016; 17:302.

Amos Bairoch and Lydie Lane

Computer and Laboratory Investigation of Proteins of Human Origin (CALIPHO) / Geneva, Centre Médical Universitaire (CMU)

What do we do?

At the CALIPHO Group, we aim to use a combination of bioinformatics and experimental methodologies to increase knowledge about the function of the 20,000 or so protein-coding genes that exist in the human genome. Our main mission is the development of neXtProt, a human protein knowledge resource.

Recently, we focused on annotating the effects of human protein variations in the context of cancers and genetic diseases, and analysing the results of high-throughput experiments to shed light on the function of selected sets of uncharacterized human proteins.

We are part of the HUPO Human Protein Project, which aims to validate the existence of all predicted human proteins in biological samples by mass spectrometry.

We are also active in the development of ontologies/standardization resources, such as Cellosaurus for cell lines and ICEPO for ion channel electrophysiology.

Highlights 2016

In 2016, we continued to develop the new generation of the neXtProt platform. The new neXtProt site allows power users and software developers to make use of all of the integrated data in software tools (with our new API) or in powerful cross-resources queries (thanks to our SPARQL endpoint).

We also started to integrate the results of our effort to annotate the phenotypic effect of protein variants in cancers and genetic diseases in a new phenotype view, and through specific portals such as the one for ion channels.

Main publications 2016

Gaudet P *et al.* The neXtProt knowledgebase on human proteins: 2017 update. Nucleic Acids Res 2016; pii: gkw1062.

Hinard V *et al.* ICEPO: the ion channel electrophysiology ontology. Database (Oxford) 2016; pii: baw017.

Omenn GS *et al.* Metrics for the Human Proteome Project 2016: progress on identifying and characterizing the human proteome, including post-translational modifications. J Proteome Res. 2016; 15(11):3951-3960.

Duek P *et al.* Missing protein landscape of human chromosomes 2 and 14: progress and current status. J Proteome Res. 2016; 15(11):3971-3978.

Frédérique Lisacek

Proteome Informatics Group
University of Geneva

What do we do?

At the Proteome Informatics Group (PIG) we are involved in software and database development for the benefit of the proteomics and 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 2016

In recent years, we have set out to collect and integrate information on glycans, whose role is increasingly described as key in many normal and pathological cellular processes. With the creation of a dedicated tab on the ExPASy server in 2015, we now centralize databases and tools that are useful for glycomics and glycoproteomics and gradually expand our range. In 2016, we introduced three new tools:

- GlycoSiteAlign, which selectively aligns amino acid sequences surrounding glycosylation sites depending on structural properties of the glycan attached to the site. The tool previews and/or downloads alignments which may reveal amino acid patterns corresponding to selected features (e.g. fucosylated vs. non-fucosylated).
- PepSweetener, which facilitates the manual annotation of intact glycopeptide spectra through the interactive inspection of theoretical glycopeptides matching queried molecular masses. It is currently focused on human glycoproteins.
- Glydin', which compiles knowledge on glyco-epitopes (glycan ligands) and displays their interrelationships based on monosaccharide composition in an interactive network.

This tool collection is mainly designed to support scientists in shaping assumptions on glycan-binding properties and glycan attachment on proteins.

Main publications 2016

Gastaldello A *et al.* GlycoSiteAlign: glycosite alignment based on glycan structure. J Proteome Res. 2016; 15(10):3916-3928.

Horlacher O *et al.* Mining large scale tandem mass spectrometry data for protein modifications using spectral libraries. J Proteome Res. 2016; 15(3):721-31.

Mariethoz J *et al.* SugarBindDB, a resource of glycan-mediated host-pathogen interactions. Nucleic Acids Res. 2016; 44(D1):D1243-50.

Christian von Mering

Bioinformatics / Systems Biology Group
University of Zurich

What do we do?

In the Bioinformatics / Systems Biology Group, we study the dynamics of entire biological systems, on both evolutionary and shorter time scales – down to a few minutes. We often work in close collaboration with laboratory scientists, focusing on the computational aspects of studying such systems, in fields ranging from genetics to genomics and proteomics. In addition, we 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 2016

In 2016, our group completely re-designed the web interface of our protein-protein interaction database, STRING. This removed several bottlenecks in usability and throughput, and provided much better value to expert and non-expert users alike. We also took the opportunity to replace several outdated web-technologies, such as Adobe Flash. 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.

We also developed a high-throughput pipeline for recognizing microbial "species" from high-throughput sequencing data. We applied this pipeline to a truly global dataset, describing the quantitative occurrence patterns of many microbial species for the first time. We then used this dataset to derive novel measures for community similarity and for clustering quality.

Main publications 2016

Schmidt TS *et al.* A family of interaction-adjusted indices of community similarity. ISME J. 2016; 11(3):791-807.

Szklarczyk D *et al.* The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. Nucleic Acids Res. 2017; 45:D362-D368.

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. 2016; 44:D286-D293.

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

What do we do?

At the Swiss-Prot Group we develop, annotate and maintain the UniProtKB/Swiss-Prot protein sequence database, the most widely used protein information resource in the world. We also develop and maintain other databases including PROSITE, a database of protein families and domains, ENZYME, a database of enzyme nomenclature, HAMAP, a collection of manually curated family profiles for protein classification and associated, manually created annotation rules, Rhea, a curated database of chemical reactions and SwissLipids, an expert-curated resource that integrates lipidomics data with biological knowledge. We also offer the virologists' community the ViralZone portal. The group also co-heads the development and maintenance of the ExPASy proteomics website. We are one of the largest groups at SIB.

Highlights 2016

In 2016, our group continued to produce and maintain the Swiss-Prot section of the UniProt knowledgebase, adding expert-curated knowledge from over 7,000 new publications describing over 3,300 new proteins and functions. Expert curation is essential to the development of Swiss-Prot and remains the surest means of providing life science researchers with rapid access to comprehensive knowledge on protein function. In 2016, we conducted a detailed and wide-ranging investigation of published and curated literature in Swiss-Prot. This investigation revealed that the expert curation of UniProtKB/Swiss-Prot provides high coverage of available knowledge and that this activity is both scalable and sustainable. While the majority of our efforts are directed at UniProtKB/Swiss-Prot, our group continues to develop a range of complementary and specialized knowledge resources for life scientists. These include the Rhea knowledgebase of expert curated biochemical reactions, which has applications in enzyme annotation (Rhea will be used in UniProt from 2017 onwards) and the description of genome-scale metabolic networks (Rhea is a reference resource for MetaNetX at Vital-IT). We introduced a new reaction classification, which complements and extends the widely-used enzyme classification of the IUBMB. Rhea currently describes over 9,500 unique reactions curated from a similar number of publications. Our ViralZone team worked with collaborators from the University of KwaZulu-Natal to develop an improved resource for the life scientists working on HIV-1, the BioAfrica HIV-1 Proteome Resource. We collaborated with a number of SIB groups and other world-renowned institutes on the development of standardized benchmarks for the phylogenetic inference, a cornerstone for both evolutionary studies and functional annotation efforts.

Our 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.

Main publications 2016

Altenhoff AM *et al.* Standardized benchmarking in the Quest for orthologs. Nat Methods 2016; 13, 425-430.

Breuza L *et al.* The UniProtKB guide to the human proteome. Database (Oxford) 2016; pii:bav120.

Druce M *et al.* Improving HIV proteome annotation: new features of BioAfrica HIV Proteomics Resource. Database (Oxford) 2016; pii:baw045.

Morgat A *et al.* Updates in Rhea - an expert curated resource of biochemical reactions. Nucleic Acids Res. 2016; pii:gkw990.

Poux S *et al.* On expert curation and sustainability: UniProtKB/Swiss-Prot as a case study. bioRxiv 2016; 094011.

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 clinicians in their decision making.

What do we do?
Our focus at the Computational Oncogenomics Group is the analysis of structural variations in cancer genomes using computational genomics, including bioinformatics and systems biology methods. Our work centres around our collections of molecular tumour data, assembled from genomic screening experiments in cancer e.g. through molecular-cytogenetic 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.
Other aspects of the group's work are the development of computational methods for genome profiling datasets as well as the design and implementation of standards for genome and metadata annotation and sharing. Recently, we have become increasingly interested in questions of genome data epistemology, e.g. the identification of analysis biases related to geographical provenance and disease-type in cancer.

Highlights 2016
In 2016, much of the group's activity was focused on advancing projects for the Global Alliance for Genomics and Health (GA4GH). In particular, we contributed to GA4GH schema elements for the exchange of data describing biological and clinical features. Additionally, our team designed a GA4GH data implementation project, which has been accepted as one of the ELIXIR human data implementation studies. Also, together with the SIB technology group, we developed an implementation of a GA4GH Beacon, based on our arraymap data resource, to facilitate a forward looking development of the Beacon protocol for the incorporation of structural genomic variants (beacon.arraymap.org).

Main publications 2016
Ai N *et al.* CNARA: reliability assessment for genomic copy number profiles. BMC Genomics. 2016; 17(1):799.
Andersson A *et al.* PKCα and HMGB1 antagonistically control hydrogen peroxide-induced poly-ADP-ribose formation. Nucleic Acids Res. 2016; 44(16):7630-45.

What do we do?
The Computational Biology Group is located in Basel and part of the Department of Biosystems Science and Engineering (D-BSSE) of ETH Zurich. Our research and teaching activities are in the areas of computational biology, biostatistics, and systems biology. Our activities include the development of mathematical and statistical models, their implementation in computer programmes, and their application to biomedical problems. We are 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 2016
Our 2016 highlights include the development and release of (i) TiMEx and pathTiMEx, waiting time models for pathways of mutually exclusive cancer alterations and their progression dynamics, (ii) SCITE, a method for full Bayesian phylogenetic tree inference from single-cell data, and (iii) novel algorithms for the efficient learning of large-scale mutational networks.

Main publications 2016
Cristea S *et al.* pathTiMEx: Joint inference of mutually exclusive cancer pathways and their progression dynamics. J Comput Biol. 2017. In press.
Jahn K *et al*. Tree inference for single-cell data. Genome Biol. 2016; 17:86.
Montazeri H *et al.* Large-scale inference of conjunctive Bayesian networks. Bioinformatics 2016; 32(17):i727-i735.

What do we do?
In the Bioinformatics Core Facility (BCF) we promote trans-disciplinary collaboration between research teams in medicine, molecular biology, genetics, genomics, statistics and bioinformatics. In particular, we perform analysis of biomedical-genomics data with a focus on biomarker studies in cancer research, building on our specific expertise in statistical methods for genomics data analysis. Recently, we concentrated on molecular heterogeneity and pathway activation patterns in cancer subtypes, but we are open to any kind of research direction.

Highlights 2016
Our team continues to investigate the molecular heterogeneity of colon cancer (CC) with the aim of finding information that is useful to assess 1) the expected risk of metastasis and 2) the best way to treat the disease after surgical removal. Useful information consists in predicting the benefit of chemotherapy – with respect to its toxicity – 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). We then test the usefulness of these groups with respect to clinical interest.
Our team completed a collaborative investigation (Guinney *et al.* 2015) designed to consolidate previously proposed gene expression subtype systems including our own (Budinska *et al.* 2013), which we recently expanded (Barras *et al.* 2016).
After many years, 2016 also saw the completion of the first analysis of the MINDACT trial data, providing a confirmation of the utility of gene-expression-based assessment regarding the risk of metastasis in chemotherapy decisions for breast cancer patients (Cardoso *et al.* 2016).

Main publications 2016
Cardoso F *et al.* 70-Gene Signature as an aid to treatment decisions in early-stage breast cancer. N Engl J Med. 2016; 25;375(8):717-29.
Barras D *et al*. BRAF V600E mutant colorectal cancer subtypes based on gene expression. Clin Cancer Res. 2016; 23(1):104-115.
Bady P *et al.* Sensitivity analysis of the MGMT-STP27 model and impact of genetic and epigenetic context to predict the MGMT methylation status in gliomas and other tumors. J Mol Diagn. 2016; 18(3):350-61.

What do we do?
At the Host-Pathogen Genomics Group we 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, we are 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 our important research directions.

Highlights 2016
During the course of 2016, we used a combination of exome/genome and RNA sequencing approaches to search for human genetic determinants of unusual responses to infectious agents. Last year, our team identified rare genetic variants conferring extreme susceptibility to several paediatric infections, including bronchiolitis/pneumonia caused by common respiratory viruses, and severe sepsis due to *Pseudomonas aeruginosa*. Joint analyses of host and pathogen genomes are now ongoing on the same samples.
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 2016
Asgari S *et al.* Exome sequencing reveals primary immunodeficiencies in children with community-acquired *Pseudomonas aeruginosa* sepsis. Front Immunol. 2016; 7:357.
Rusert P *et al.* Determinants of HIV-1 broadly neutralizing antibody induction. Nat Med. 2016; 22(11):1260-1267.
McLaren PJ *et al.* Privacy-preserving genomic testing in the clinic – a model using HIV treatment. Genet Med. 2016; 18(8):814-22.

David Gfeller
Computational Cancer Biology
University of Lausanne

What do we do?
At the Computational Cancer Biology Group, our aim is to study the interactions between cancer and immune cells. To this end, we develop machine-learning algorithms to analyse large-scale genomics and proteomics data. In particular, we are focusing on molecular and cellular aspects of cancer immune cell interactions. At the molecular level, we develop tools to predict (neo-)antigen presentation by integrating large HLA peptidomics datasets. At the cellular level, we are developing novel approaches to characterize immune infiltrations and the different states of immune cells from gene expression profiles of tumours and immune cells.

Highlights 2016
During 2016, the group developed a novel bioinformatics tool to analyse large HLA peptidomics datasets, and gain a better understanding of the properties of HLA peptide ligands.

Main publications 2016
Bassani-Sternberg M and Gfeller D. Unsupervised HLA peptidome deconvolution improves ligand prediction accuracy and predicts cooperative effects in peptide-HLA interactions. J Immunol. 2016; 197, 2492.
Gfeller D *et al.* Current tools for predicting cancer-specific T cell immunity. OncoImmunology 2016; 5(7):e1177691.

Ivo Kwee
Bioinformatics Core Unit
Institute of Oncology Research, Bellinzona

What do we 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. Our 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 just playing a supporting role, we proactively identify and develop novel bioinformatics projects that can complement and in many cases drive our biological research. In collaboration with SIB, we develop innovative data analysis tools, visualization software and database resources for genomics research.

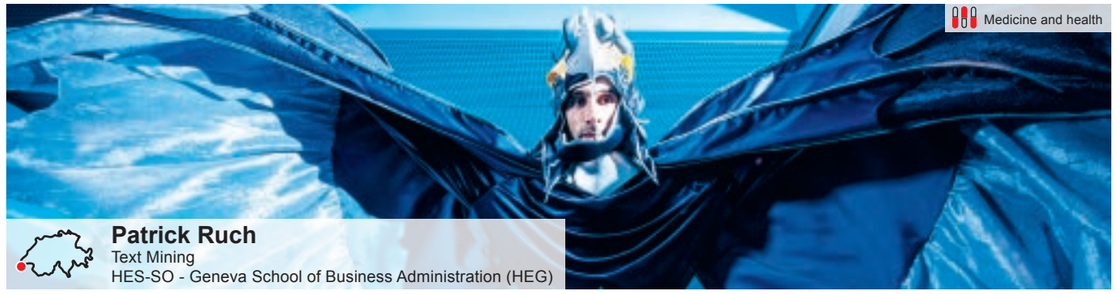
Highlights 2016
In 2016, the canton of Ticino approved the integration of IOR into the new Faculty of Biomedical Sciences of the University of Lugano from 2017. The new Faculty will offer a Master's degree in Medicine (three years) starting in 2020, in close collaboration with, on the academic side, ETH Zurich, the University of Basel and the University of Zurich, and with the Ente Ospedaliero Cantonale and the private clinics in Ticino for bedside teaching.

Main publications 2016
Akhmedov M *et al.* A divide and conquer matheuristic algorithm for the Prize-collecting Steiner Tree Problem. Computers & Oper Res. 2016; 70, 18-25.
Vasquez R *et al.* The bromodomain inhibitor OTX015 (MK-8628) exerts anti-tumor activity in triple-negative breast cancer models as single agent and in combination with everolimus. Oncotarget. 2017; 8(5):7598-7613.
Gaudio E *et al.* Combination of the MEK inhibitor pimasertib with BTK or PI3K-delta inhibitors is active in preclinical models of aggressive lymphomas. Ann Oncol. 2016; 27(6):1123-8.

Carlos-Andrés Peña-Reyes
Computational Intelligence for Computational Biology (CI4CB)
HEIG-VD, Yverdon-Les-Bains

What do we do?
With the advent of high-throughput technologies and clinical information systems, the life sciences and clinical sciences now produce very large amounts of data (Big Data). Our goal is to uncover hidden patterns in these data, as well as building data-driven models as tools to discover biomarkers and assist clinicians in their decisions. Our projects encompass the fields of transcriptomics, systems biology and clinical bioinformatics and analytics.

Highlights 2016
A few ongoing selected projects:
- INPHINITY, SNSF project with UNIL and Inselspital: We develop predictive models of phage-bacteria interactions to build dynamic and comprehensive infection networks. These models help biologists and clinicians select potential therapeutic phages against pathogenic bacteria.
- FISHGUARD, a Eurostars-2 project in partnership with European SMEs Bioscientia and Biotem: Our bioinformatics and machine-learning expertise contributes to the discovery and characterization of biomarkers that are to be embedded in a novel screening test against two viral infections in fish.
- BOSS, CTI project with the startup SimplicityBio: We develop new feature selection and visualization methods to improve their biomarker discovery process.
- D-Rex, Hasler Stiftung-funded project: We explore and develop methods to understand, prioritize and explain the knowledge obtained through a deep neural network by representing it in the form of hierarchical and logical rules.



Patrick Ruch
Text Mining
HES-SO - Geneva School of Business Administration (HEG)

What do we do?
In the SIB Text Mining Group, we carry out activities in semantic and text analytics applied to the health and life sciences. Previously hosted by the Radiology and Medical Informatics Department of the Geneva University Hospitals, our group moved to the University of Applied Sciences Geneva (HES-SO – HEG Geneva) in 2008. We 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. We are thus designing, developing and maintaining data and web analytic instruments, such as custom search engines, automatic text classifiers and information extraction systems, to help domain experts "make sense" of biomedical data.

Highlights 2016
Over the course of 2016, our team developed a new curation service, nextA5, which prioritizes the literature for specific curation requirements. This service has been shown to significantly improve the search effectiveness of curators along three important curation axes: diseases (+231%), molecular functions (+236%), and biological processes (+3,153%). In parallel, user-friendly interfaces powered with a set of JSON web services are currently being implemented into the neXtProt annotation pipeline.
Regarding scientific communication and publication, the group organized a workshop during the Biocuration 2016 conference in Geneva (10-14 April). The workshop featured different presentations related to advances and challenges in the field of text-mining applied to Biocuration.

Main publications 2016
Mottin L *et al.* neXtA5: accelerating annotation of articles via automated approaches in neXtProt. Database (Oxford) 2016; pii: baw098.
Mottin L *et al.* BiTeM at CLEF eHealth Evaluation Lab 2016 Task 2: Multilingual Information Extraction. CLEF (Working Notes) 2016: 94-102.

Christoph Schmid
Computational PathoGenOmics Group
Swiss Tropical and Public Health Institute (TPH), Basel

What do we do?
At the Computational PathoGenOmics Group at the Swiss Tropical and Public Health (Swiss TPH) Institute we focus our 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, we develop computational methods and apply them to research questions in infection biology and public health. We 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.

Evolution and phylogeny

In addition to being an open book of how an organism functions, genomes can inform life scientists on how a species has evolved over time. Phylogeny studies how species are related to each other.

Bioinformatics develops tools to read a species' genome, compare genetic information between organisms, develop computing methods to reconstruct their past and build their 'family' trees.

Maria Anisimova

Applied Computational Genomics Team
Zurich University of Applied Sciences, Wädenswil

What do we do?

At the Applied Computational Genomics Team, we 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, we strive to keep pace by 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. Our goal is to bring new bioinformatics methods to real applications, ranging from biotechnology to biomedical research, ecology and agriculture. We embrace an interdisciplinary approach by integrating different data sources and combining methods.

Highlights 2016

During the course of 2016, our team worked on the SNSF grant implementing new algorithms in the prototype JATI software, in order to allow for simultaneous fast alignment and phylogeny reconstruction from genomics sequences. This included the formulation of a dynamic programming algorithm explicitly including an indel model.

Interesting work has been completed in collaboration with Roche to analyse flu seasonal dynamics, based on genomic sequences collected worldwide and over several years.

The first author of this study, PhD student Lorenzo Gatti, presented this work at the SIB Days in Biel/Bienne and received an award for the best lightning talk.

This year our group organized and hosted for the first time the PhyloSIB workshop “Big data and phylogenomics” in ZHAW Wädenswil. The workshop attracted about 40 participants, thus representing the largest number of SIB groups in the history of PhyloSIB. As a highlight, Prof. Ziheng Yang opened the workshop with an invited talk.

Finally, at the end of 2016 we received two new research grants. As a result, two more postdoctoral fellows and one research assistant will join our group in 2017.

Christophe Dessimoz

Computational Evolutionary Biology and Genomics Group
University of Lausanne

What do we do?

Working at the interface of biology and computer science, our laboratory seeks to better understand evolutionary and functional relationships between genes, genomes and species. A few key underlying 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?
- Conversely, how can we exploit the wealth and diversity of life to get a better grasp on specific organisms or systems of interest?
- Can we summarize meaningfully the evolutionary history of species by arranging them into a small number of tree topologies that capture both vertical inheritance and the most important events of non-vertical inheritance?

Our activities are divided between bioinformatics methods and resource development, and their application – typically with experimentalists.

Highlights 2016

We published a paper reporting the key achievements of the Quest for Orthologs consortium benchmarking working group, which we have been conducting for the past four years. The work established minimum standards in orthology benchmarking and reported the outcome of a community experiment including 14 leading orthology methods. For these, a battery of 20 tests was carried out on a standard set of 66 genomes crossing all kingdoms of life. This will facilitate future orthology benchmarking by offering a web-based benchmarking service.

We also established a clear and tractable definition for the concept of “homoeology”, i.e. evolutionary relationships which arise via hybridization–allopolyploidization of the genome.

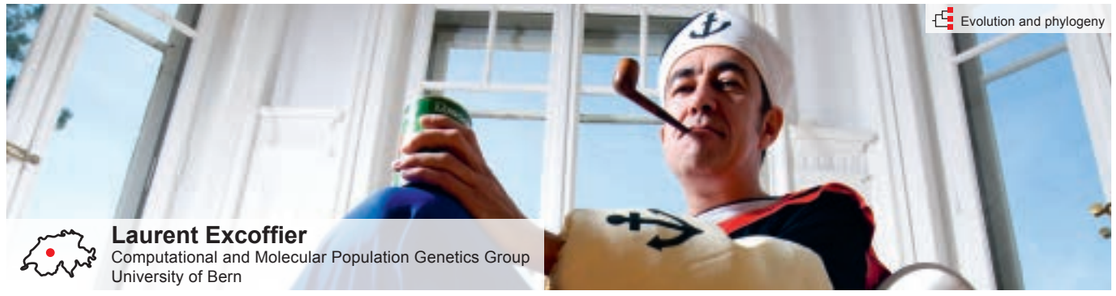
We also edited *The Gene Ontology Handbook*, an open-access book, which was published by Springer. The book provides a practical and self-contained overview of Gene Ontology – an essential resource in any bioinformatician’s toolbox – with several chapters contributed by SIB members.

Main publications 2016

Altenhoff A *et al.* Standardized benchmarking in the quest for orthologs. Nat Methods 2016; 13, 425–430.

Glover N *et al.* Homoeologs: what are they and how do we infer them? Trends Plant Sci. 2016; 21, 609–621.

Dessimoz C and Škunca N (Editors). *The Gene Ontology Handbook*. Meth Mol Biol. 2017, Springer (New York), Vol. 1446.



Laurent Excoffier

Computational and Molecular Population Genetics Group
University of Bern

What do we do?

At the Computational and Molecular Population Genetics (CMPG) Group, we develop new methodologies for the simulation and analysis of molecular polymorphisms within species, with a particular focus on humans. We 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 2016

In 2016, our group was heavily involved in reconstructing past demographic scenarios of human and chimpanzee evolution based on full genomes. In humans, we showed that all non-African populations originated from a single wave of migration out of Africa some 70 Kya, and that Australia had a complex and long history since its founding more than 50 Kya. In chimpanzees, we found evidence that bonobo hybridized with the ancestors of the Central and Eastern chimpanzee between 200 and 500 Kya, suggesting that admixture might have been widespread in hominids. We also used spatially explicit simulations to better examine the settlement of Eurasia by modern humans. We found that long-distance dispersal had been important in maintaining high levels of genetic diversity during these times, and that Eurasians retracted to southern refugia during the last glacial maximum.

Main publications 2016

Alves I *et al.* Long distance dispersal shaped patterns of human genetic diversity in Eurasia. Mol Biol Evol. 2016; 33: 946–958.

de Manuel M *et al.* Chimpanzee genomic diversity reveals ancient admixture with bonobos. Science 2016; 354: 477-481.

Henn BM *et al.* Distance from sub-Saharan Africa predicts mutational load in diverse human genomes. Proc Natl Acad Sci U S A 2016;113:E440-9.

Malaspinas AS *et al.* A genomic history of Aboriginal Australia. Nature 2016; 538: 207-214.

Peischl S *et al.* Genetic surfing in human populations: from genes to genomes. Curr Opin Genet Dev. 2016; 41: 53-61.

Gaston Gonnet

Computational Biochemistry Research Group
ETH Zurich

What do we do?

At the Computational Biochemistry Research Group, we are interested in the modelling and analysis of biological problems at the molecular level. In particular, our expertise lies in searching algorithms, optimizing algorithms, mathematical modelling, and computational systems. Most of our 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. We offer the results and general services via the internet and through the distribution of the Darwin system for bioinformatics computations.

Highlights 2016

The OMA Browser is a SIB-funded, publicly available resource that provides orthology predictions among publicly available proteomes from all domains of life. The most recent release of the OMA Browser covers now 2,024 complete genomes. The OMA algorithm is available as a stand-alone programme to analyse custom genome dataset. The group is actively participating in several community benchmarking efforts focused on orthology and protein function prediction.

The group is engaged in teaching and supervising the advanced Master’s degree in Computational Biology and Bioinformatics at ETH Zurich.

Main publications 2016

Altenhoff AM *et al.* Standardized benchmarking in the quest for orthologs. Nat Methods 2016; 13(5):425-430.

Yuxiang J *et al.* An expanded evaluation of protein function prediction methods shows an improvement in accuracy. Genome Biol. 2016; 17(1):184.

What do we do?
At the Population Genetics and Genomics Group, our interest is focused on understanding how the interplay of population structure, trait architecture and selection can be disentangled. To this end, we 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, we investigate the dynamics of multilocus genetic systems under the influence of selection, migration and drift, and develop comprehensive individual-based models as well as statistical methods to infer selection, mating systems and population structure.

Highlights 2016
In 2016, with the arrival of two new PhD students, the group continued to investigate the genome of the barn owl, which will help buttress the hypothesis put forward in our Evolution paper suggesting a ring-like colonization of this bird of prey around the Mediterranean. Meanwhile, our collaboration with Prof. Bruce Weir led to the largest survey of human genetic polymorphism using forensic markers.
The Hierfstat R package, developed by the group, has seen several new features added and it is now well connected with other population genomics packages such as Adegenet, Ape and Apex. PhD students and postdocs from SIB and Swiss universities had the chance to discover the features of these packages through a doctoral course organized by our group.

Main publications 2016

Kapun M *et al.* Genomic evidence for adaptive inversion clines in *Drosophila melanogaster*. Mol Biol Evol. 2016; 33(5):1317-36.
Buckleton J *et al.* Population-specific FST values for forensic STR markers: A worldwide survey. Forensic Sci Int Genet. 2016; 23:91-100.
Burri R *et al.* The genetic basis of color-related local adaptation in a ring-like colonization around the Mediterranean. Evolution 2016;70(1):140-53.
Ducret V *et al.* Sex-specific allelic transmission bias suggests sexual conflict at MC1R. Mol Ecol. 2016; 25(18):4551-63.
Rougemont Q *et al.* Reconstructing the demographic history of divergence between European river and brook lampreys using approximate Bayesian computations. PeerJ. 2016; 4:e1910.

What do we do?
Our 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. We work on both applied and theoretical problems in fields ranging from population genomics to medical genetics. We 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 2016
In 2016, we continued our focus on statistical inference in population genetics – in both natural and experimental settings.

Main publications 2016

Ewing G *et al.* The consequences of not accounting for background selection in demographic inference. Mol Ecol. 2016; 25(1): 135-141.
Bank C *et al.* An experimental evaluation of drug induced mutational meltdown as an antiviral treatment strategy. Evolution 2016; 70: 2470-84.
Bank C *et al.* On the (un)predictability of a large intragenic fitness landscape. Proc Natl Acad Sci U S A 2016; 113: 14085-90.

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

Highlights 2016
We added a new gene page to our Bgee database of gene expression evolution. Thanks to a new algorithm that ranks expression information of different types – from RNA-seq to *in situ* hybridization – we are now able to present the most relevant expression patterns for a gene, highlighting them among hundreds of expression patterns in different conditions.
We also distributed an R package allowing to access all Bgee data and to perform TopAnat computations. A manuscript describing it is in preprint at F1000research Bioconductor channel.
The group participated in a new paper resulting from the Master's degree course "Sequence a genome", during which Master's degree students sequence, assemble and annotate new bacterial genomes, using PacBio and RNA-seq for the first time. The paper is based directly on the students' work during the class.

Main publications 2016

Kryuchkova N and Robinson-Rechavi M. A benchmark of gene expression tissue-specificity metrics. Brief Bioinform. 2016; pii: bbw008.
Davydov II *et al.* State aggregation for fast likelihood computations in molecular evolution. Bioinformatics 2016; 33(3):354-362.
Kryuchkova N and Robinson-Rechavi M. Tissue-specificity of gene expression diverges slowly between orthologs, and rapidly between paralogs. PLoS Comput Biol. 2016; 12(12):e1005274.

What do we do?
At the Computational Phylogenetics Group, we develop software to better understand the evolutionary history between organisms and to test macroevolutionary hypotheses. We are looking at the ecological, genomic and morphological factors that limit and constrain speciation and adaptation. We focus on phylogenetic reconstruction methods, clownfish and plant genomics, the estimation of positive selection on genes, modelling the evolution of DNA sequences and phenotypes, the mode and tempo of species evolution and the spatially explicit evolution of diversity. Our aim is to develop better models to analyse sequence data and quantitative models to estimate macroevolutionary patterns and processes.

Highlights 2016
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 as well as the evolution of their phenotypic traits. The method was implemented into an R software called Jive. The models implemented in Jive were extended by developing a novel Bayesian approach that can estimate the rate of evolution of a quantitative trait and its variance along a phylogenetic tree. 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 in the rates of trait evolution, to assess the phylogenetic effects on the evolution of those traits.
The models fully complement existing approaches and are currently used to estimate the evolution of several groups of animals (e.g. mammals, birds or clownfish) or the evolution of floral morphologies in key groups of angiosperms.

Main publications 2016

Kostikova A *et al.* Bridging inter- and intraspecific trait evolution with a hierarchical Bayesian approach. Syst Biol. 2016; 65: 417-431.
Roland J and Salamin N. Niche width impacts vertebrate diversification. Global Ecol Biogeo. 2016; 10: 1252-1263.
Roland J *et al.* Molecular evolutionary rates are not correlated with temperature and latitude in *Squamata*: an exception to the metabolic theory of ecology? BMC Evol Biol 2016; 16:95.

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

Highlights 2016
In 2016, we focused on developing tools for investigating ongoing epidemics such as Zika in South America or seasonal influenza in our hometown, Basel. We validated these software tools using data from the Ebola outbreak 2013-16 in West Africa. For the ongoing Zika outbreak we are using free available data, while for influenza we are collecting data ourselves (funded by an SNSF interdisciplinary grant). In particular, we conducted a city-wide survey on influenza (for media coverage see https://www.bsse.ethz.ch/cevo/cevo-press/2016/04/influenza-survey-in-progress.html), and we will be collecting blood samples from patients during the upcoming winter season.
Furthermore, the group took a big step in bridging part of the gap between molecular evolution and palaeontology. We 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 present-day species. To develop the area further, the group welcomed a new postdoc on an ETH fellowship grant.

Main publications 2016
Kuehnert D *et al.* Phylodynamics with migration: A computational framework to quantify population structure from genomic data. Mol Biol Evol. 2016; 33 (8): 2102-2116.
Drummond AJ *et al.* Bayesian phylogenetic estimation of fossil ages. Philos Trans R Soc Lond B Biol Sci. 2016; 371: 20150129.
Stadler T and Smrckova J. Estimating shifts in diversification rates based on higher-level phylogenies. Biol Lett. 2016; 12:20160273.

What do we do?
At the Evolutionary Systems Biology Group, we study the evolution and evolvability of biological systems at all levels of biological organization: from genes and genomes to biological networks and whole organisms. We 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. Our work uses comparative analysis of genomic data, laboratory evolution experiments and mathematical modelling. We also develop a variety of bioinformatics tools to help us take advantage of the torrent of data in genomics and structural biology.

Highlights 2016
In 2016, we developed the Genonets Server, a computational tool that allows the construction of genotype networks, which play an important role in understanding the evolutionary dynamics of evolving molecules. The server can:
1. construct genotype networks for categorical and univariate phenotypes from DNA, RNA, amino acid or binary sequences;
2. analyse genotype network topography and how it relates to the navigability of a genotype network via mutation and natural selection;
3. provide multiple interactive visualizations that facilitate exploratory research and education.

We also developed Growthcurver, an R package for obtaining interpretable metrics from microbial growth curves. Plate readers can measure the growth curves of many microbial strains in a high-throughput fashion. The hundreds of absorbance readings collected simultaneously for hundreds of samples create technical hurdles for data analysis. Growthcurver summarizes the growth characteristics of microbial growth curve experiments conducted in a plate reader. The data are fitted to a standard form of the logistic equation, and the parameters have clear interpretations on population-level characteristics, such as doubling time, carrying capacity and growth rate.

Main publications 2016
Khalid F *et al.* Genonets server – A web server for the construction, analysis, and visualization of genotype networks. Nucleic Acids Res. 2016; 44: W70-W76.
Sprouffske K *et al.* Growthcurver: An R package for obtaining interpretable metrics from microbial growth curves. BMC Bioinformatics 2016; 17:172.
Wagner A *et al.* From the primordial soup to self-driving cars: standards and their role in natural and technological innovation. J R Soc Interface. 2016; 13(115):20151086.

What do we do?
When observing nature, one is easily impressed by the huge diversity seen on any biological scale. Our 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, we design and evaluate new statistical and computational approaches to infer complex evolutionary histories. For this we develop and apply machine-learning algorithms, with a particular focus on likelihood-free methods. We then apply these approaches to the wealth of data currently being generated, mostly in collaboration with experimental groups. We are further committed to making all our developments available to the scientific communities by releasing easy-to-use software packages.

Highlights 2016
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). Our group has been particularly interested in understanding how to incorporate PMD into population genetic analysis. For instance, we 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.
We applied these methods to learn more about how farming spread across prehistoric Europe, and found that the early farmers from the Aegean region are direct ancestors of the early farmers in Western Europe, thus prompting the fact that farming spread, predominantly, as farmers colonized Europe. Interestingly, however, the first farmers of the Aegean region are genetically distinct from the first farmers in the fertile crescent, the presumed origin of farming, suggesting that farming initially spread as a cultural idea.

Main publications 2016
Kousathanas A *et al.* Inferring heterozygosity from ancient and low coverage genomes. Genetics 2017; 205: 317-332.
Broushaki F *et al.* Early neolithic genomes from the eastern Fertile Crescent. Science 2016; 353: 499-503.
Hofmanová Z *et al.* Early farmers from across Europe directly descended from Neolithic Aegeans. Proc Natl Acad Sci U S A 2016; 113: 6886-6891.
Kousathanas A *et al.* Likelihood-free inference in high-dimensional models. Genetics 2015; 203(2): 893-904.
Ferrer-Admetlla A *et al.* An approximate Markov model for the Wright-Fisher diffusion and its application to time-series data. Genetics 2016; 203(2):831-846.

Structural biology

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

Bioinformatics develops software to model and predict a protein's 3D structure, and hence deduce its probable function, or study its interaction with other molecules. Such tools are of great assistance in the field of drug design, for instance.

Simon Bernèche
Computational Biophysics Group
University of Basel

What do we do?
At the Computational Biophysics Group, we are interested in the structure-function relationship of membrane proteins. Using molecular mechanics simulations and statistical approaches, our group aims to understand 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 revolves around the elucidation of gating mechanisms, which regulate 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 2016
In the context of the Human Brain Project, one of the EU's flagship projects, we defined a new generation of kinetic models of K channels that will eventually be used in neuron models at the core of the brain simulator. These new kinetic models summarize all of our knowledge of K channels accumulated over decades through structural, functional, and simulation studies. This provides an ideal framework to better understand the role played by each of the channels in the modulation and propagation of action potentials within neurons. On this basis, we elaborated new hypotheses on the mechanisms that sustain working (or short) memory in the brain. We are currently developing neuron network models to test these hypotheses.

Main publications 2016
Sauer MM *et al.* Catch-bond mechanism of the bacterial adhesin FimH. Nat Commun. 2016; 7(7), 10738.
Leung HTA *et al.* A rigorous and efficient method to reweight very large conformational ensembles using average experimental data and to determine their relative information content. J Chem Theory Comput. 2016; 12(1):383-94.
.

Matteo Dal Peraro
Laboratory for Biomolecular Modelling
EPFL, Lausanne

What do we do?
Our main goal at the Laboratory for Biomolecular Modelling 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, we 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 2016
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.
Using *LipidBuilder*, a framework that we previously introduced for creating models of biologically relevant phospholipid species with acyl tails of heterogeneous composition, we used multiscale molecular dynamics simulations to investigate the stability of the amyloid precursor protein (APP) dimer in realistic models of the synaptic plasma membrane (SPM). The proteolytic cleavage of the transmembrane domain of APP releases amyloid-β (Aβ) peptides, whose accumulation in the brain tissue is an early indicator of Alzheimer's disease. We discovered that the specific composition of the SPM and, in particular, the abundance of highly unsaturated lipids were fundamental for selecting one of the two possible APP dimerization states so far proposed (Audagnotto M *et al.* 2016).

Main publications 2016
Audagnotto M *et al.* Effect of the synaptic plasma membrane on the stability of the amyloid precursor protein homodimer. J Phys Chem Lett. 2016;7:3572.
Abriata LA *et al.* Molecular effects of concentrated solutes on protein hydration, dynamics, and electrostatics. Biophys J. 2016;111:743.
Iacovache I *et al.* Cryo-EM structure of aerolysin variants reveals a novel protein fold and the pore-formation process. Nat Commun. 2016; 7:12062.
Song AS *et al.* Immobilization of the N-terminal helix stabilizes prefusion paramyxovirus fusion proteins. Proc Natl Acad Sci U S A 2016; 113:E3844.

Olivier Michielin and Vincent Zoete
Molecular Modelling Group
University of Lausanne

What do we do?
At the Molecular Modelling Group (MMG) we 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. Our main activity consists of 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 such as T cell receptors (TCR), for cancer immunotherapy. We develop and maintain several web tools for drug design, such as SwissDock, SwissBioisostere and SwissTargetPrediction. We also act as the Protein Modelling Facility (PMF) of the University of Lausanne.

Highlights 2016
In 2016, the group officially released SwissSimilarity, a new web tool for rapid ligand-based virtual screening. SwissSimilarity is part of a large SIB initiative to provide online tools for drug design – such as SwissDock, SwissBioisostere and SwissTargetPrediction, with which it can interoperate.
In addition, the group designed the BOILED-Egg classification model, which provides an intuitive graph prediction of passive intestinal absorption and brain penetration. The group also developed a new on-the-fly QM/MM docking algorithm within the in-house docking tool Attracting Cavities. The group finished the development of SwissADME, a new web tool to compute the physicochemistry and estimate the pharmacokinetics, druglikeness and medicinal chemistry friendliness of small molecules. Among others, SwissADME provides an exclusive access to our in-house models iLogP and BOILED-Egg. The SwissADME interface is designed to allow both experts and non-experts to use it. Official release is planned for 2017.

Main publications 2016

Daina A and Zoete V. A BOILED-Egg to predict gastrointestinal absorption and brain penetration of small molecules. ChemMedChem, 2016; 11(11):1117–1121.
Zoete V *et al.* SwissSimilarity: A web tool for low to ultra high throughput ligand-based virtual screening. J Chem Inf Model. 2016; 56(8):1399–1404.
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):437–447.
Chaskar P *et al.* On-the-fly QM/MM docking with attracting cavities. J Chem Inf Model. 2017; 57(1):73–84.
Daina A *et al.* SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. Sci Rep. 2017. In press.

Torsten Schwede
Computational Structural Biology Group
University of Basel

What do we do?
At the Computational Structural Biology (CSB) Group, we are focusing on the development of methods and algorithms to model, simulate and analyse three-dimensional protein structures and their molecular properties in order to apply these techniques to the understanding of biological processes at a molecular level. Our 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 2016
In 2016, we released a new re-engineered SWISS-MODEL Repository, which integrates the latest developments in the SWISS-MODEL pipeline and features a newly designed graphical interface. Proteins of model organisms are modelled on a weekly basis. Users of the Repository can also trigger an almost instantaneous update of a given entry if no models are available for a specific UniProt sequence.
According to the magazine “Horizons” (March 2016) of the SNSF Swiss National Science Foundation, our 2014 publication describing the developments of the SWISS-MODEL expert system was ranked as the sixth highest-impact scientific publication published in Switzerland over the period 2014/2015.
Our group participated in the TecDays at the Gymnasium Stadelhofen in Zurich and at the Liceo Cantonale in Bellinzona with a module about Structural Bioinformatics. This event is organized by the Swiss Academy of Engineering Sciences to promote the study of technical and natural sciences among students in Swiss schools.

Main publications 2016

Bienert S *et al.* The SWISS-MODEL Repository-new features and functionality. Nucleic Acids Res. 2017; 45(D1):D313-D319.
Moult J *et al.* Critical assessment of methods of protein structure prediction: progress and new directions in round XI. Proteins 2016; 84 Suppl 1:4-14.
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 2016; 84 Suppl 1:349-69.

Systems biology

Biological macromolecules do not work on their own. Instead they interact with others which, in turn, interact with others, thus creating complex systems. The interdisciplinary field of systems biology aims at studying these systems in a holistic way.

Bioinformatics develops computational and mathematical models that can illustrate such systems and even address their evolution over 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.

What do we do?
Our lab builds the bridge between Big Data Analysis and Biomedical Research. We develop novel Data Mining Algorithms to detect patterns and statistical dependencies in large datasets from the fields of biology and medicine. Our major goals are twofold: 1) to enable the automatic generation of new knowledge from Big Data through Machine Learning, and 2) to gain an understanding of the relationship between Biological Systems and their molecular properties. Such an understanding is of fundamental importance for personalized medicine, which tailors medical treatment to the molecular properties of a person.

Highlights 2016
Recent research exploiting the tremendous progress in sequencing technologies has generated huge data sets of genetic information that enable large-scale analyses, such as genome-wide association studies (GWAS) to explore genotype–phenotype relationships. An effort in this direction, to which the MLCB lab contributed, is the sequencing of the genomes of 1,135 naturally inbred lines of the model plant *Arabidopsis thaliana*, and the subsequent establishment of a high quality reference genome panel (The 1001 Genomes Consortium, 2016). Our lab was also at the forefront of establishing AraPheno, a public database that allows people to easily submit, download and visualize phenotypic data for *Arabidopsis thaliana* (Seren *et al.*, 2016). Our current work aims at bringing both genetic and phenotypic data together in one advanced online platform for performing genome-wide association studies: easyGWAS.
In biological and healthcare data, researchers are facing extremely high-dimensional representations of samples, from patients to bacteria. When linking these high-dimensional representations to phenotypes, multiple testing correction is of the utmost importance for practitioners. Due to the large number of dimensions, however, multiple testing correction is computationally challenging and prone to losing all detection power. We present a first approach for finding significant feature combinations, which properly corrects for multiple testing and at the same time makes it possible to account for categorical covariates such as age or gender of individuals (Papaxanthos *et al.*, 2016).

Main publications 2016
Papaxanthos L *et al.* Finding significant combinations of features in the presence of categorical covariates. NIPS 2016. In press.
Seren Ü *et al.* AraPheno: a public database for *Arabidopsis thaliana* phenotypes. Nucleic Acids Res. 2017; 45(D1):D1054-D1059.
The 1001 Genomes Consortium. 1,135 genomes reveal the global pattern of polymorphism in *Arabidopsis thaliana*. Cell 2016;166(2):481-491.

What do we do?
At the Scientific and Parallel Computing (SPC) Group we develop new algorithms and methods to better understand and/or predict various phenomena in biology. We 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 our group is the modelling and simulation of complex systems. In bioinformatics, the objective is to develop advanced numerical methodology to model biological processes.

Highlights 2016
In 2016, our team took part in an H2020 CompBioMed project, “A Centre of Excellence in Computational Biomedicine”. In collaboration with Dagmar Iber, a Sinergia project “A 3D Cell-Based Simulation Framework for Morphogenetic Problems” was also accepted, and will run for a period of four years.

Main publications 2016
Malaspinas O *et al.* A spatio-temporal model for spontaneous thrombus formation in cerebral aneurysms. J Theor Biol. 2016 Apr 7;394:68-76.
Meyer X *et al.* Accelerating Bayesian inference for evolutionary biology models. Bioinformatics 2016; pii: btw712.
Merzouki A *et al.* The properties of a cell-based numerical model of epithelium under stretching constraints. Soft Matter. 2016; 12(21):4745-54.

What do we do?
Our research at the Computational Single Cell Biology Group aims to elucidate the composition of heterogeneous cell populations and how these implement function in the context of cancer and immune biology. To accomplish this task, we build 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 computational analyses. Our research can be used to pinpoint therapeutic targets with a view to designing drugs.

Highlights 2016
During the course of 2016, our team overhauled 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.

Main publications 2016
Klimovskaia A *et al.* Sparse regression based structure learning of stochastic reaction networks from single cell snapshot time series. PLoS Comput Biol. 2016; 12(12):e1005234.
Feigelman JS *et al.* Analysis of cell lineage trees by exact Bayesian inference identifies negative autoregulation of Nanog in mouse embryonic stem cells. Cell Syst. 2016; 3(5):480-490.e13.

What do we do?
At the Chemical and Biological Systems Engineering Laboratory, we develop tools for systems modelling and the analysis of chemical and biological networks. Our mission is to create enabling theories and computational methods for the generation of systems insights, as well as for understanding and acquiring knowledge in chemical, biological and medical applications. Our 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 2016
During the course of 2016, our team released two tools which were described in separate publications in the journals of *Bioinformatics* and *BMC Bioinformatics*:
1. DeltaNet: bioinformatics tool that is used to infer the gene targets of drug compounds from gene transcriptional profiles;
2. TRaCE+ (Transitive Reduction and Closure Ensemble+): tool for gene regulatory network inference from gene transcriptional profiles. It extends the capability of our TRaCE algorithm by inferring the regulatory modes (activation/repression).

Furthermore, the team’s analysis of single cell gene transcriptional profiles of chicken erythrocytic cell differentiation revealed a transition period marked by a peak in the cell-to-cell variability of the gene expression and the connectivity of the gene co-expression network. Importantly, this peak preceded an irreversible cellular commitment to differentiation.

Main publications 2016
Noh H and Gunawan R. Inferring gene targets of drugs and chemical compounds from gene expression profiles. Bioinformatics 2016; 32(4):2120-7.
Ud-Dean SMM *et al.* TRaCE+: Ensemble inference of gene regulatory networks from transcriptional expression profiles of gene knock-out experiments. BMC Bioinformatics 2016; 17:252.
Richard A *et al.* Single-cell-based analysis highlights a surge in cell-to-cell molecular variability preceding irreversible commitment in a differentiation process. PLoS Biol. 2016; 14(12):e1002585.

Dagmar Iber

Computational Biology Group
ETH Zurich, D-BSSE, Basel

What do we do?

The Computational Biology Group (CoBi) develops computational models of developmental processes. We place a particular focus on mechanistic 4D image-based *in silico* models of organogenesis (mouse lung, kidney, pancreas, limb, brain, *Drosophila* wing and eye) and on the delineation of fundamental mechanisms such as those that restrict the size of organs and those that maintain the proportions of structures in different-sized embryos. The group collaborates with tissue engineers to build spatially-organized tissue from stem cells, and with clinicians to apply its techniques to disease models.

Highlights 2016

One of the key open problems in developmental biology concerns the mechanism of size and growth control. The organ growth rate declines continuously during embryonic development, but the underlying mechanism is elusive. Jannik Vollmer, a recipient of an SIB fellowship, has now shown that the growth rate in the eye disc declines in a manner that is inversely proportional to the increase in the eye disc area (Vollmer *et al.*, *Development* 2016; see commentary). This observation is consistent with growth control by dilution of a cytokine. In a separate line of work, Patrick Fried and his co-workers developed a quantitative model of eye disc development that is consistent with the experimental data and which explains the patterning dynamics in the eye disc, e.g. the linear progression of the morphogenetic furrow that separates proliferating and differentiated tissue (Fried *et al.*, *PLoS Comp Biol*. 2016). Both works were highlighted by their respective journals.

Together with the SBML development team, Harold Gomez published MOCCASIN, a software tool to automate the conversion of MATLAB ODE models into SBML (Gomez *et al.*, *Bioinformatics* 2016).

Main publications 2016

Fried P *et al.* A model of spatio-temporal dynamics of *Drosophila* eye disc development. PLoS Comput Biol. 2016 Sep 14;12(9):e1005052.

Vollmer J *et al.* A quantitative analysis of *Drosophila* eye disc growth. Development 2016; 143(9):1482-90.

Gomez H *et al.* MOCCASIN: converting MATLAB ODE models to SBML, Bioinformatics 2016; pii: btw056.

Christian Mazza

Biomathematics and Computational Biology Group
University of Fribourg

What do we do?

At the Biomathematics and Computational Biology Group, we are a small team based in the Department of Mathematics at 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 over time with respect to the many interactions of a different nature that exist within an organism. We study 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 2016

Dougoud M *et al.* Ultrasensitivity and sharp threshold theorems for multisite systems. J. Phys. A: Math. Theor. 2017; 50, 075601.

Dougoud M *et al.* The effect of gap junctional coupling on the spatiotemporal patterns of Ca^{2+} signals and the harmonization of Ca^{2+}-related cellular responses. PLoS Comput Biol. 2016; 12(12):e1005295.

Dougoud M *et al.* The feasibility of equilibria in large ecosystems: a primary but neglected concept in the complexity-stability debate. 2016 arXiv:1612.06735.

Michel Milinkovitch

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

What do we do?

We combine Evolutionary and Developmental Biology with the study of Physical Processes to understand the mechanisms generating life's complexity and diversity. We specialize in non-classical model species in reptiles and mammals and we integrate data and methods from physics, comparative genomics and molecular developmental genetics, as well as computer modelling and numerical simulations.

What has Physics got to do with it? Remarkably, many questions in development are conceptually similar to those investigated in soft-matter physics, statistical physics and mechanics. For example, self-organizational capabilities of cells and tissues and the role of geometry and form are pertinent to EvoDevo on multiple scales and levels of analysis. Our objectives are to understand the interactions between physical (e.g., mechanics or reaction diffusion) and biological (e.g., cell signalling, proliferation, migration) parameters, which generate patterns and shapes during development.

These studies are integrated into an evolutionary and molecular genetic perspective. Hence, we also perform comparative genomics/transcriptomics and develop phylogeny inference tools. Our projects are, by essence, highly multidisciplinary and integrative, and our team includes evolutionary and developmental biologists, computer scientists, engineers and physicists.

Highlights 2016

During the course of 2016, our lab:

- produced a high-quality / high-coverage sequencing and annotation of the corn snake genome;
- performed linkage mapping and characterization of several mutations in the corn snake that affect colour and colour patterns;
- identified new reaction-diffusion mechanisms that affect skin colour patterns;
- demonstrated the evolutionary homology of skin appendages (hairs, feathers and scales) across amniote vertebrates, a question that has been debated for decades.

Main publications 2016

Di-Poï N and Milinkovitch MC. The anatomical placode in reptile scale morphogenesis indicates shared ancestry among skin appendages in amniotes. Science Advances 2.6 2016; e1600708.

Dhillon DSJ *et al.* Bifurcation analysis of reaction diffusion systems on arbitrary surfaces. 2016 arXiv:1605.01583 [cs.CE]

Ullate-Agote A *et al.* A step-by-step guide to assemble a reptilian genome. In Press.

Félix Naef

Computational Systems Biology Group
EPFL, Lausanne

What do we 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 our 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 2016

The group's major highlights in 2016 included the analysis of the nuclear proteome as a function of time in the mouse liver, and the analysis of temperature-dependent RNA processing, notably for the Cirbp gene.

Main publications 2016

Wang J *et al.* Nuclear proteomics uncovers diurnal regulatory landscapes in mouse liver. Cell Metab. 2017; 25(1):102-117.

Gotic I *et al.* Temperature regulates splicing efficiency of the cold-inducible RNA-binding protein gene. Cirbp. Genes Dev. 2016; 30(17):2005-17.

Bischofberger M *et al.* Revealing assembly of a pore-forming complex using single-cell kinetic analysis and modeling. Biophys J. 2016; 110(7):1574-81.

What do we do?
Our research interests at the Scientific Computing Group lie in the area of multiscale/multiphysics modelling and parallel large-scale simulations of biological systems. We focus on the development of new computational models and corresponding numerical methods suitable for the next generation of super computers. We are working on stochastic multiscale modelling of motion, the interaction, deformation and aggregation of cells under physiological flow conditions, biofilm growth and coarse-grained molecular dynamics simulations, as well as the modelling of transport processes in healthy and tumour-induced microcirculation.

Highlights 2016
The spleen plays multiple roles in the human body. Among them is the removal of old and altered red blood cells (RBCs), which is achieved by filtering cells through endothelial slits – i.e. small micron-sized openings. It was previously observed that people without a spleen have less deformable RBCs, indicating that the spleen may play a role in defining RBC size and shape. We used a detailed RBC model implemented within the Dissipative Particle Dynamics (DPD) simulation framework to study the filter function of the human spleen. Our results demonstrate that the spleen does indeed play a major role in defining the size and shape of healthy human RBCs, thus indicating a new function for a well-known organ. These results offer a better understanding of how the circulatory bottleneck for RBCs in the spleen could affect a variety of acute and chronic disease states arising from hereditary disorders, human cancers and infectious diseases, with implications for therapeutic interventions and drug efficacy assays.

Main publications 2016
Pivkin IV *et al.* Biomechanics of red blood cells in human spleen and consequences for physiology and disease. Proc Natl Acad Sci U S A 2016; 113(28):7804-7809.
Peter EK *et al.* Coarse kMC-based replica exchange algorithms for the accelerated simulation of protein folding in explicit solvent. Phys Chem Chem Phys. 2016; 18(18):13052-13065.
Peter EK *et al.* A canonical replica exchange molecular dynamics implementation with normal pressure in each replica. J Chem Phys. 2016; 145(4):044903.

What do we do?
At the CSB Group at ETH Zurich in Basel we 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. We 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. Our biological applications rely on the group's experimental biology section that uses budding yeast as a model organism, and on various external collaborations.

What do we 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 we 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.

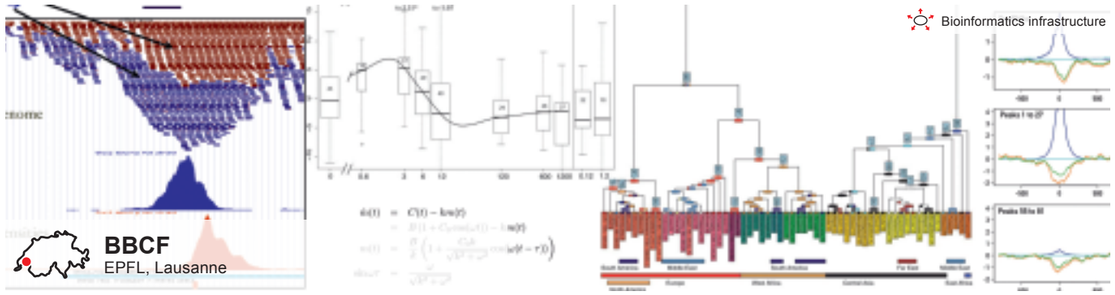
Highlights 2016
A first highlight of 2016 was the discovery of a novel regulator of pre-mRNA 3' end processing. Through the computational analysis of many 3' end sequencing data sets we found that the HNRNPC, a protein so far known as a splicing regulator, affects the choice of 3' end processing sites. These in turn determine the sequence of the 3' untranslated regions of transcripts and, further, their localization and translation. A second highlight was the quantification of ribosomal protein expression heterogeneity across human cell types. In this study we found that hematopoietic cells exhibit a striking lineage-specific expression of certain ribosomal proteins, as do malignant cells. These patterns of ribosomal protein expression can be explained through both transcription regulator and copy number variation, and have prognostic value in cancers.

Main publications 2016
Gruber AJ *et al.* A comprehensive analysis of 3' end sequencing data sets reveals novel polyadenylation signals and the repressive role of heterogeneous ribonucleoprotein C on cleavage and polyadenylation. Genome Res. 2016; 26(8):1145-59.
Jorjani H *et al.* An updated human snoRNAome. Nucleic Acids Res. 2016; 44(11):5068-82.
Guimaraes JC and Zavolan M. Patterns of ribosomal protein expression specify normal and malignant human cells. Genome Biol. 2016; 17(1):236.

Bioinformatics infrastructure

With the advent of new technologies, the quantity of data generated by researchers has grown exponentially and needs to be stored as well as processed.

This is where bioinformatics infrastructure comes into play. Academic institutions and research centres are gradually developing their own infrastructure that provides computational facilities, software and databases, in addition to providing a link with industry and offering training.



What do we do?
At the EPFL Bioinformatics and Biostatistics Core Facility (BBCF) we provide research labs with extensive support in bioinformatics and biostatistics. Our main competences are in management and analysis of genomic data, mathematical modelling and statistical analysis of quantitative biological data. We 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 statistical planning in complex experimental designs. We 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 2016
In 2016, the BBCF had the opportunity to participate in a pilot analysis with Pierre Fabre, PhD – from Prof. Duboule's group (EPFL) – whose aim was to highlight specific expression patterns of the Hox genes into different mouse tissues using Single-cell RNA-seq (scRNA-seq). This study should now be extended, and most probably completed with chromatin accessibility (ATAC-Seq or scATAC-seq) and chromatin folding (FISH and 4C-seq) data in order to understand specific cis-regulation at the single cell level.

With Prof. Deplancke's group (EPFL, SIB) we started the development of ASAP (Automated Single-cell Analysis Pipeline), a fully integrated, web-based platform aimed at the complete analysis of single-cell and bulk RNA-seq data post genome alignment: from the parsing, filtering and normalization of the input count data files to the visual representation of the data, identification of cell clusters, differentially expressed genes (including cluster-specific marker genes), and functional gene set enrichment.
In collaboration with Prof. van der Goot's group (EPFL) we continued the development of SwissPalm, and developed a tool to compare inter-species palmitoyl-proteomes. This new tool allows to identify – with higher confidence – truly palmitoylated proteins from proteomic screens.

Main publications 2016
Loviglio MN *et al.* Chromosomal contacts connect loci associated with autism, BMI and head circumference phenotypes. Mol Psychiatry. 2016
Deluz C *et al.* A role for mitotic bookmarking of SOX2 in pluripotency and differentiation. Genes Dev. 2016; 30(22):2538-2550.
Zeng YX *et al.* Extension of least squares spectral resolution algorithm to high-resolution lipidomics data. Anal Chim Acta. 2016; 914:35-46.

What do we do?
At the Service and Support for Science IT (S3IT) unit we provide support for science in general, and for life science and medicine in particular. S3IT serves as a partner for both local and national projects to enable competitive research with the advanced use of computational methods and resources. Our team advises groups and projects about data management and data analysis, and cooperates to optimize 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 2016
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 launched a platform for integrated data management, workflow management, and data analytics. Such a platform enables the detailed tracking of measured data, as well of the results derived from this data. It also provides users with an easy way to analyse their data through an interactive web portal based on Jupyter.

With regard to infrastructure, in 2016 our team extended the local ScienceCloud infrastructure, which now consists of over 6,000 CPU cores and 2PB of usable storage. Our Hydra system targets use-cases requiring up to 3TB of RAM. In 2016 it was renewed and more than doubled in size. Furthermore, a new GPU cluster system Vesta with a total of 80 GPUs was procured and put into production.

Main publications 2016
Röst HL *et al.* TRIC: an automated alignment strategy for reproducible protein quantification in targeted proteomics. Nat Methods 2016; 13, 777-783.
Röst HL *et al.* OpenMS: A flexible open-source software platform for computational mass spectrometry. Nat Methods 2016; 13, 741-748.
Stanford NJ *et al.* FAIRDOM: Data and model management for all. Nucl Acids Res. In press.

Bernd Rinn
Scientific IT Services (SIS)
ETH Zurich, D-BSSE, Basel

What do we do?
Our group at the Scientific IT Services (SIS) is an interdisciplinary bioinformatics and scientific IT support group, which develops computational tools. These tools range from lab databases to reusable framework components that enable and support both data analysis and data management in life science research and beyond. We collaborate with Swiss and European research groups and industry in the life science sector – such as SystemsX.ch, SyBIT FAIRDOM, HPC-CH and swissuniversities' eSCT / EnhanceR community. We improve and port scientific software, develop data management solutions and provide associated services. We also integrate and operate data analysis pipelines, and provide training and consulting in databases, scientific software development, high-performance and cloud computing.

Highlights 2016
This year, we had many discussions with ETH researchers from the personalized health domain on their computing and data privacy needs. In response to these discussions, we started to develop the new dedicated data and computing infrastructure *"Leonhard Med"* at ETH, along with associated data services. While parts of the concept are still in flux as the Swiss Personalized Health Network is shaping up, *Leonhard Med* is already in use today.
As in recent years, we have been busy porting data analysis platforms to Euler, either from desktops or from older computer clusters. We sparked the network analysis of antibody repertoires to allow the first-ever network analysis of a full antibody repertoire with samples containing up to 500k+ CDR3 sequences. We also ported a big software platform this year, the iPortal Proteomics data analysis system, which runs workflows like OpenSWATH. The new EulerPortal performs about twice as fast as the old system.
As the demand for computing power by ETH users has been constantly growing, the HPC cluster Euler has been upgraded with 452 new computing nodes, thus increasing the performance of the cluster from 570 TFlops to 1 PFlop. At the same time, Euler's project storage was extended by 1PB to adapt to ever-growing research data. The next substantial upgrade will be installed early 2017.
We have released a new major version 16.05 of the openBIS data management platform. Among other improvements, the new release features a new and more flexible application programming interface (v3), improved search capabilities and better performance for a set of advanced use cases. The new API incorporates feedback from many users, and in particular from the OpenSEEK software development in the FAIRDOM consortium.
In order to sustain the efforts of FAIRDOM beyond the lifetime of the project consortium and support data management that is Findable, Accessible, Interoperable and Reusable, the FAIRDOM association was founded and is now open to new members.
In 2016, the openBIS ELN-LIMS user base increased steadily and now contains more than 15 groups from three different departments. In the CRUS P2 project DLCM ("Data Lifecycle Management"), we worked on improving further the system based on biologists' feedback. Some of the improvements are web-based project summaries, downloadable reports on projects, simplified data upload and direct access to data in the system via a built-in fileserver.
As the complexity of data analysis approaches that are used in the life sciences is growing, ensuring the reproducibility of scientific findings is focusing. We have been working with life science wet labs to integrate the open scientific computing platform Jupyter with the openBIS ELN-LIMS data management system. This is ongoing work, and the first results are very encouraging.
Among the training activities, the four-day intensive workshop on *parallel programming with MPI* and *OpenMPI* in the summer was a highlight. It was conducted by Rolf Rabenseifner from HLRS Stuttgart, a well-known expert in the field.

Main publications 2016

Wolstencroft K *et al.* FAIRDOMHub: a repository and collaboration environment for sharing systems biology research. Nucleic Acids Res. 2017; 45(D1):D404-D407.
Khan TA *et al.* Accurate and predictive antibody repertoire profiling by molecular amplification fingerprinting. Sci Adv. 2016; 2(3):e1501371.
Fusco L *et al.* Computer vision profiling of neurite outgrowth dynamics reveals spatiotemporal modularity of Rho GTPase signaling. J Cell Biol. 2016; 212:91-111

Torsten Schwede and Thierry Sengstag
sciCORE
University of Basel

What do we do?
sciCORE is a centre of competence in scientific computing. It provides 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. At sciCORE we provide a professional environment for scientific applications: from bioinformatics, computational chemistry, physics and systems biology to medicine and economics. In direct collaboration with scientific research groups, we help develop, deploy, operate and extend the computational tools required for performing modern life science and biomedical research. We operate the IT infrastructure for several SIB services, e.g. SWISS-MODEL and SwissRegulon.

Highlights 2016
In 2016, the sciCORE team consolidated its support for the Swiss life-science community: notably, we expanded our support to web resources developed at Agroscope (ZH) and the University Hospital Basel, and deployed a new storage service available for groups using instruments that produce large amounts of data, such as high-resolution microscopes. Our data-management consulting to research groups at the University of Basel and SwissTPH was also extended.
By the end of 2016, sciCORE support had grown to about 115 groups at the University of Basel, SwissTPH and University Hospitals. Our team collaborates actively with the SIB Training group by contributing to computing and data analysis courses for life scientists (about 500 students in 21 courses covering an introduction to Linux, R and Python programming, statistics, genomics data mining, etc.). On the national scale, sciCORE remains active in infrastructure and research projects such as the Data-Life Cycle Management project, the eScience team and SystemsX.

Main publications 2016

Badri ND *et al.* Cyclic di-GMP mediates a histidine kinase/phosphatase switch by noncovalent domain cross-linking. Sci Adv. 2016; 2(9):e1600823.
Auffray C *et al.* Making sense of big data in health research: Towards an EU action plan. Genome Med. 2016; 8(1):71.
García-Senz D *et al.* Type Ia Supernovae: can Coriolis force break the symmetry of the gravitational confined detonation explosion mechanism? Astrophys J. 2016; 819:132.

Daniel Stekhoven
Clinical Bioinformatics Unit
ETH Zurich

What do we 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. We maintain close collaboration with hospitals in Zurich and Basel, introducing and enabling state-of-the-art multi-omics data analyses. Based on a fee-for-service model, our aim is to meet individual project goals and establish close interaction with our customers. Our ambition is to help the end user understand what the results mean. However, our services are not limited to analysis alone. We also offer support in writing and reviewing manuscript texts, as well as delivery of manuscript-ready tables and figures.

Highlights 2016
Building on its previous successes over the course of 2015, CBU's second year went well. The newly launched Molecular Tumor Board Zurich shaped the opinion of molecular cancer diagnostics in clinics both in Zurich and in Basel towards comprehensive analysis – with, on the one hand, comprehensive sequencing using WES and WGS for profiling tumours, and on the other hand, the comprehensive mining of public databases, clinical trial registries and research literature. While comprehensive sequencing and variant calling still have a more fundamental research characteristic, comprehensive mining is directly relevant for patients in the clinics today.
Together with the University of Zurich and the University Hospital of Zurich we were involved in a project which established the first *in vivo* mouse model for ccRCC, accurately capitulating the cellular and molecular features of clear cell renal cell carcinoma (ccRCC). This project combined RNA-Seq, WES and public data from TCGA, and we showed how well the mouse model represented human data. The project was finished and submitted at the end of 2016.
CBU planned and ran a workshop on the topic of Clinical Bioinformatics as a Service as part of the annual ECCB 2016 in The Hague. The workshop was conceived together with Niko Beerenwinkel (ETHZ), Wolfgang Huber (EMBL) and Simon Tavaré (CRUK). There was great interest in the topics and a large common ground between European core facilities providing bioinformatics in a clinical setting was identified. Given this success we are intending to organize another workshop covering the same topic soon.

Main publications 2016

Thurnherr T *et al.* Genomic variant annotation workflow for clinical applications. F1000Research 2016; 5:1963.
Lukamowicz-Rajska M *et al.* MiR-99b-5p expression and response to tyrosine kinase inhibitor treatment in clear cell renal cell carcinoma patients. Oncotarget. 2016; 7(48):78433-78447.

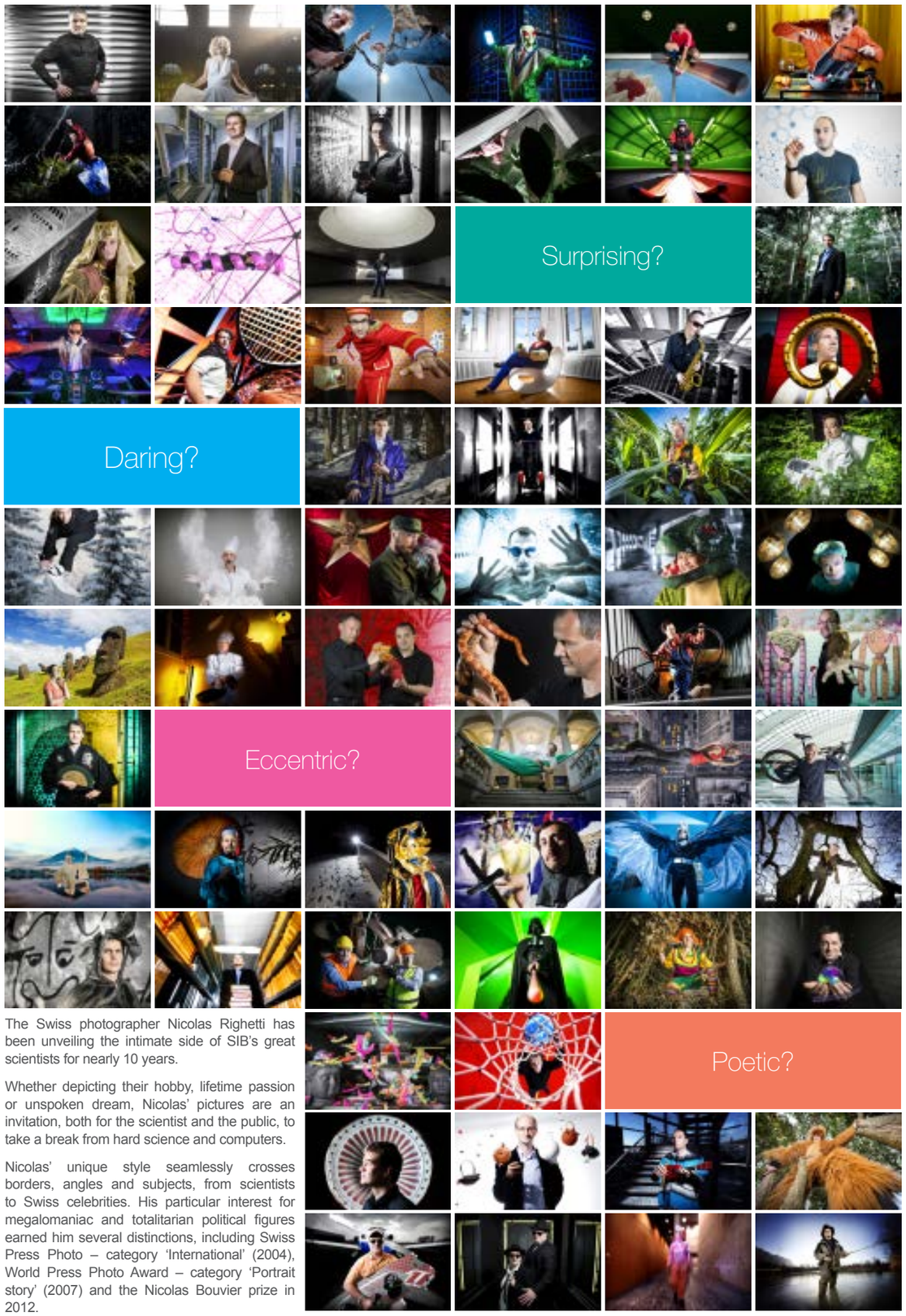
Ioannis Xenarios
Vital-IT
University of Lausanne

What do we do?
With a multidisciplinary team of scientists and technical staff, at Vital-IT we maintain a competence centre in bioinformatics and computational biology. Our infrastructure currently spreads across six institutions that maintain bio-technological platforms: SIB, the Universities of Geneva, Lausanne, Fribourg and Bern, and EPFL. At Vital-IT we enable scientists to access state-of-the-art computational infrastructure (for processing, storage and archiving) as well as expertise in data analysis and algorithmic development. We partner with scientists to build computational solutions facilitating their research or to transform their ideas into production-quality software. We support postgraduate education through training and workshops in coordination with SIB and institutional partners.
The Vital-IT Group in Lausanne and the Swiss-Prot Group in Geneva provide unique and complementary resources to the community. Vital-IT provides computational infrastructure (computing and storage), software and bioinformatics analysis, whereas the Swiss-Prot Group provides biological knowledge (UniProtKB/Swiss-Prot, PROSITE, etc.).

Highlights 2016
During 2016, Vital-IT's infrastructure grew to 15 petabytes of storage and 8,600 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, SwissLipids, 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 1,000 users in 2016. Vital-IT collaborates on research activities as part of SystemsX.ch, and technology and development projects funded by the European Commission or CTI/KTI. 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 2016

Moretti S *et al.* MetaNetX/MNXref - reconciliation of metabolites and biochemical reactions to bring together genome-scale metabolic networks. Nucleic Acids Res. 2016; 44(D1):D523-6.
Roduit S *et al.* Analysis of the dynamic co-expression network of heart regeneration in the zebrafish. Sci Rep. 2016; 6:26822.
Dorier J *et al.* Boolean regulatory network reconstruction using literature based knowledge with genetic algorithm optimization method. BMC bioinformatics 2016;17(1):410.



The Swiss photographer Nicolas Righetti has been unveiling the intimate side of SIB's great scientists for nearly 10 years.

Whether depicting their hobby, lifetime passion or unspoken dream, Nicolas' pictures are an invitation, both for the scientist and the public, to take a break from hard science and computers.

Nicolas' unique style seamlessly crosses borders, angles and subjects, from scientists to Swiss celebrities. His particular interest for megalomaniac and totalitarian political figures earned him several distinctions, including Swiss Press Photo – category 'International' (2004), World Press Photo Award – category 'Portrait story' (2007) and the Nicolas Bouvier prize in 2012.

Acknowledgements

We gratefully acknowledge the following funders, sponsors and partners for their financial support and encouragement in helping us fulfil our mission in 2016.

- 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 p. 11)
- BioAlps
- The European Commission
- Health Valley - Fondation Inartis
- The Leenaards Foundation
- The Medic Foundation
- The National Institutes of Health (NIH)
- The Research for Life Foundation
- The Solidar Immun Foundation
- The Swiss Foundation for Excellence and Talent in Biomedical Research
- SystemsX.ch

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

- The Leenaards Foundation
- The R. Geigy Foundation
- 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 Group Leaders' pictures: Nicolas Righetti, www.lundi13.ch

Design and layout: Vivactis (Suisse) SA

Cover image:
"Most proteins interact with others to carry out their function. This image is a schematic representation of some of the known 3D protein complexes in SWISS-MODEL, an SIB core resource maintained by the Computational Structural Biology group. Each complex is represented by a graph, where proteins are illustrated by nodes and their interactions by edges."