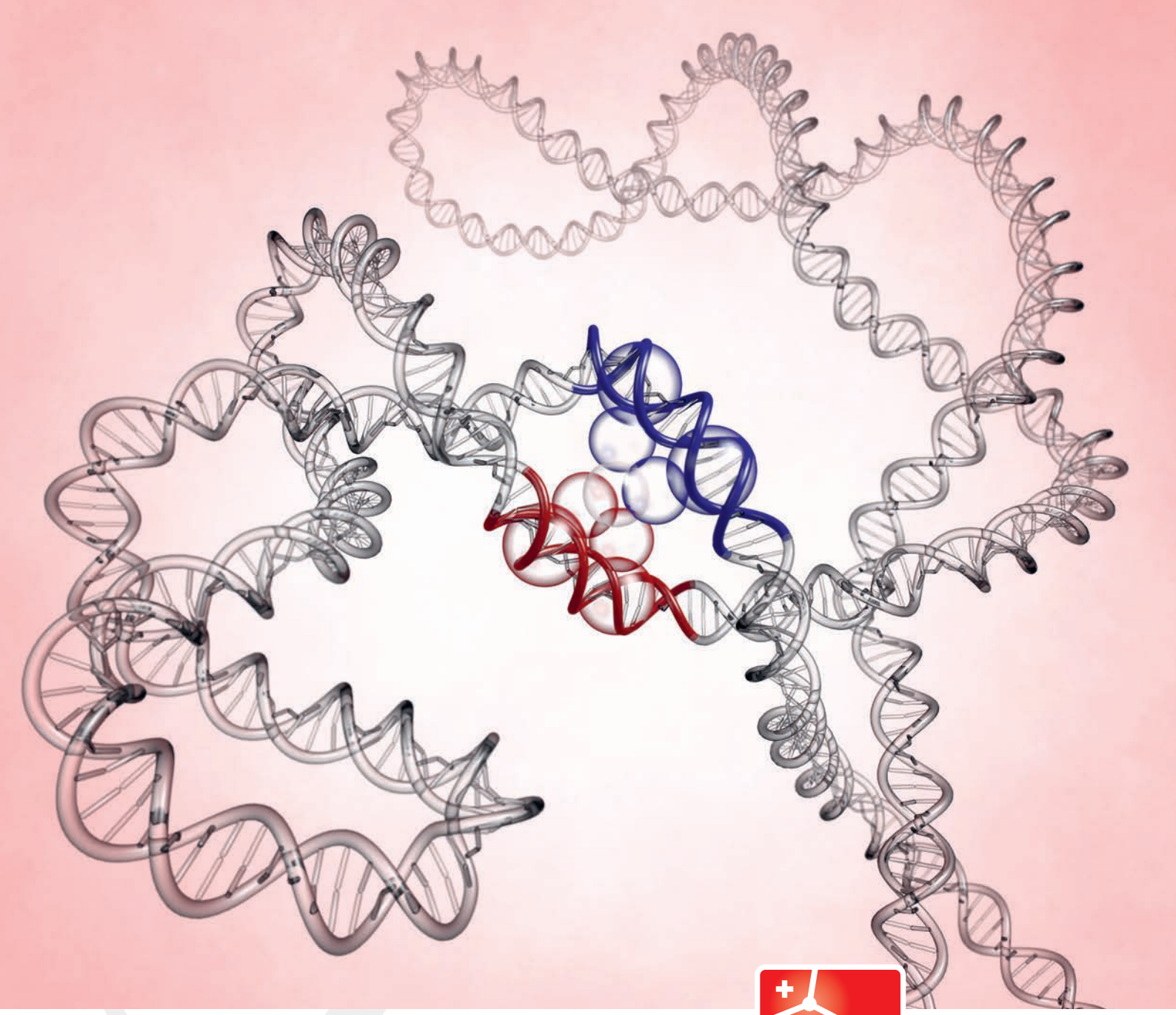


Swiss Institute of Bioinformatics

SIB Profile 2015

SIB, a crucial link in the life science chain

TABLE OF CONTENTS

Foreword 5

What is bioinformatics? 6

About SIB 8

Vision and missions 9

SIB organization: an efficient collaborative Swiss model 10

Governance 12

Bioinformatics resources for medicine and life sciences 14
- Databases and software platforms 15
- Core facilities and high-performance computing centres 18

Leading and coordinating bioinformatics in Switzerland 23
- Clinical bioinformatics: paving the way to a finer kind of health 24
- Legal and technology transfer office: enabling the community to benefit from SIB's innovations 26
- SIB technology: optimizing technology-related activities 27
- Training and outreach: training the next generation of bioinformaticians 28

SIB – the Swiss node of ELIXIR 30

In the spotlight in 2014 32

Finance 34

Staff 35

Research and service activities 36
- SIB collaborative network 36
- A wide variety of activity domains 38
  - Genes and genomes 40
  - Proteins and proteomes 47
  - Medicine and health 51
  - Evolution and phylogeny 56
  - Structural biology 63
  - Systems biology 66
  - Bioinformatics infrastructure 73

Acknowledgements 77

FOREWORD

> “Bioinformatics is at the heart of science and its long-term sustainability is crucial”

It has become difficult to imagine a life science project without bioinformatics. From understanding the 3D structure of macromolecules to designing drugs and mapping molecular pathways, bioinformatics continues to be at the forefront of many fields of research.

Today Switzerland has the highest concentration of bioinformaticians worldwide. SIB currently counts 56 research and service groups across Switzerland – a 25% increase over last year – and more than 650 members, demonstrating the interest in, and need for, such an infrastructure. Collaboration is another important measure of success, and as shown in pp. 36-37, SIB groups have established an outstanding collaborative network, which we are very proud of.

With the flourishing development of personalized medicine, our Institute is increasingly taking part in this field (pp. 24-25). Personalized medicine proposes to tailor medical decisions to the individual patient’s data, which rely on DNA sequencing and other “omics” technologies, as well as imaging. The huge amount of data generated by these methods is then processed with specific bioinformatics tools to transform “big data” into meaningful, “smart data”. This is the new field of clinical bioinformatics, whose importance was recognized by SIB very early on.

Science has never been confined to geographical boundaries. And the same goes for bioinformatics. Over the years, SIB has not only multiplied its international partnerships but thanks to its intercantonal and interinstitutional model of collaboration, the Institute was a key player in the creation of ELIXIR – the European Life Science Infrastructure for Biological Information – and is today its largest national node (pp. 30-31). In 2014, the European Council identified ELIXIR as one of Europe’s three priority research infrastructures, further highlighting the key role of bioinformatics in life sciences and medicine.

Bioinformatics is at the heart of science and its long-term sustainability is crucial. We would like to thank the Swiss government, the Federal Assembly, the State Secretariat for Education, Research and Innovation, the Swiss National Science Foundation and all those in funding roles, as well as our partner institutions for their unwavering and invaluable support.

Last but not least, our heartfelt gratitude goes to all SIB members without whose talent and dedication Swiss bioinformatics would not be where it is today.

Felix Gutzwiller
President of
the Foundation Council

Manuel Peitsch
Chairman of the Board
of Directors

Ron Appel
Executive Director

WHAT IS BIOINFORMATICS?

Bioinformatics is the application of computer technology to the understanding and effective use of biological data. In other words, it helps to convert “big data” into “smart data” or knowledge.

Computing has become a central component of modern scientific research: large volumes of data (**“big data”**) are generated by increasingly automated measuring devices. These data need to be stored, organized and analysed to extract new insights and knowledge. Because new discoveries often reveal their relevance when they are compared with the already available knowledge, extensive comparison with large datasets is a great advantage. In addition, **computational simulation has become a third pillar of science** – along with experimentation and theory – allowing researchers to advance their understanding of complex systems *in silico*.

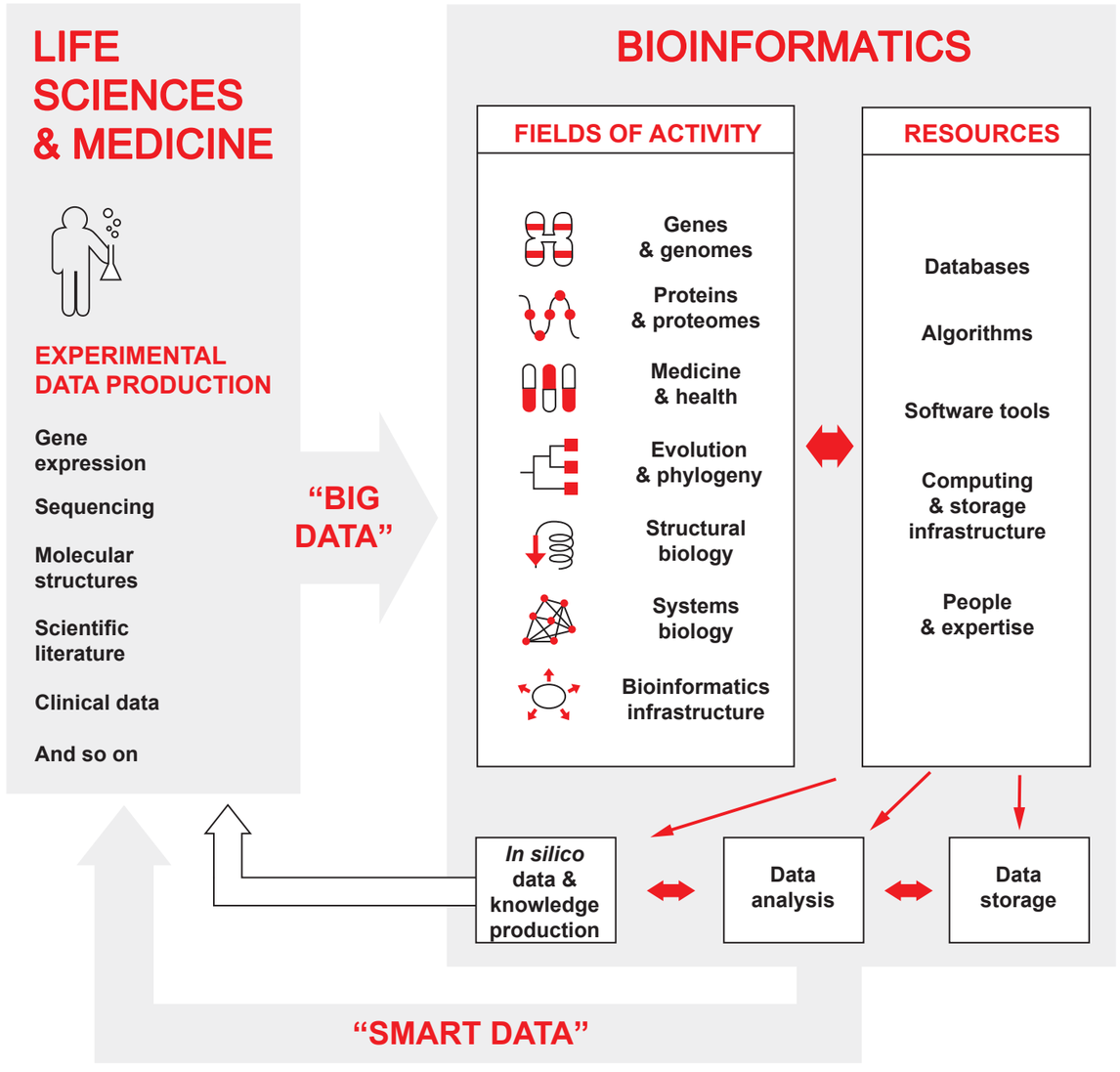
Bioinformatics provides:

- **Biocuration and bioinformatics expertise** enabling life scientists to have accurate and comprehensive representation of biological knowledge and take full advantage of bioinformatics technologies
- **Databases and knowledgebases** for giving life scientists access to curated biological data and information
- **Software** for analysing, visualizing, interpreting and comparing biological data and for modelling biological systems
- **Computing and storage infrastructures** for storing, analysing and processing biological data, including “big data”

Bioinformatics is thus an interdisciplinary field that brings major advances in many different life science and medical areas. Bioinformatics domains include (for more details, please see corresponding sections on p. 38 and following pages):

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

Ensuring online access to bioinformatics resources, such as databases and software platforms, as well as training and support from skilled bioinformaticians, is therefore essential for medicine and life sciences.



This chart shows the crucial role of bioinformatics in medicine and life sciences. Thanks to resources such as databases and software platforms, bioinformatics allows the storage and analysis of “big data” produced by experimentation. These data are then converted into “smart data” or knowledge that can be used by life scientists and clinicians. Bioinformatics can also generate data knowledge through *in silico* experiments.

ABOUT SIB

SIB, a crucial link in the life science chain

The SIB Swiss Institute of Bioinformatics is a unique success story at the frontier of life sciences and computer science. When the Institute was founded in 1998, bioinformatics was still in its infancy, both in Switzerland and abroad. Today SIB is an **independent, non-profit foundation** recognized of public utility that provides world-class bioinformatics. The **decentralized, federating organizational structure** of SIB is a model for countries setting up their own bioinformatics infrastructure, as well as for the European bioinformatics programme ELIXIR, which has adopted the hub and nodes model.

By sharing its expertise in storage, analysis and dissemination of large biological datasets and through education and **collaborations with research institutes and industrial partners**, SIB has created a true bioinformatics culture in Switzerland, which counts today the highest concentration of bioinformaticians in the world. The Institute continues to lead developments in the field of bioinformatics. Well aware that the wealth of data produced by modern technologies and the growing self-awareness of patients will change the way of considering medical data, SIB is also now embracing the challenge to strive for bioinformatics excellence in **personalized health**.

For more information: www.isb-sib.ch

SIB in figures

More than
650
members
who carry out outstanding science and services

More than
150
resources, including high-quality databases and software
accessible through the SIB portal ExPASy; created in 1993, ExPASy was at that time the first website available in the biomedical field

More than
1,300
peer-reviewed articles
published by SIB members over the last 15 years

More than
900,000
requests
per month on the UniProtKB/ Swiss-Prot website

16
institutional members
spread across Switzerland

56
research and service groups
from the major Swiss schools of higher education and research institutes

VISION AND MISSIONS

VISION

SIB helps **shape the future of life sciences through excellence in bioinformatics**

Mission 1

To **provide world-class core bioinformatics resources** to the national and international life science research community

Mission 2

To **lead and coordinate the field of bioinformatics** in Switzerland

In order to achieve these two missions, SIB has established

four strategic goals

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

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

Federate bioinformatics research groups from Swiss universities and research institutes and foster collaboration and innovation at the highest level of scientific excellence

Train first-rate researchers

SIB ORGANIZATION
AN EFFICIENT COLLABORATIVE SWISS MODEL

Modelled on Switzerland's federal structure, SIB is a federation of bioinformatics research and service groups from the major Swiss schools of higher education. The relationship between the academic world and SIB represents an efficient form of symbiosis: both SIB and the host institutions gain from intergroup synergies and from the national and international recognition of the high quality of its science and services.

SIB's institutional members are the Universities of Basel, Bern, Fribourg, Geneva, Lausanne, and Zurich, the Università della Svizzera italiana (USI), the Swiss Federal Institutes of Technology of Lausanne (EPFL) and Zurich (ETH Zurich), the Geneva School of Business Administration (HEG), the Zurich University of Applied Sciences (ZHAW); as well as research institutes: the Friedrich Miescher Institute for Biomedical Research (FMI), the Ludwig Institute for Cancer Research, the Swiss Tropical and Public Health Institute (Swiss TPH), the Agroscope, and the Institute of Oncology Research (IOR) (**see map on the facing page**). SIB also works closely with industrial partners.

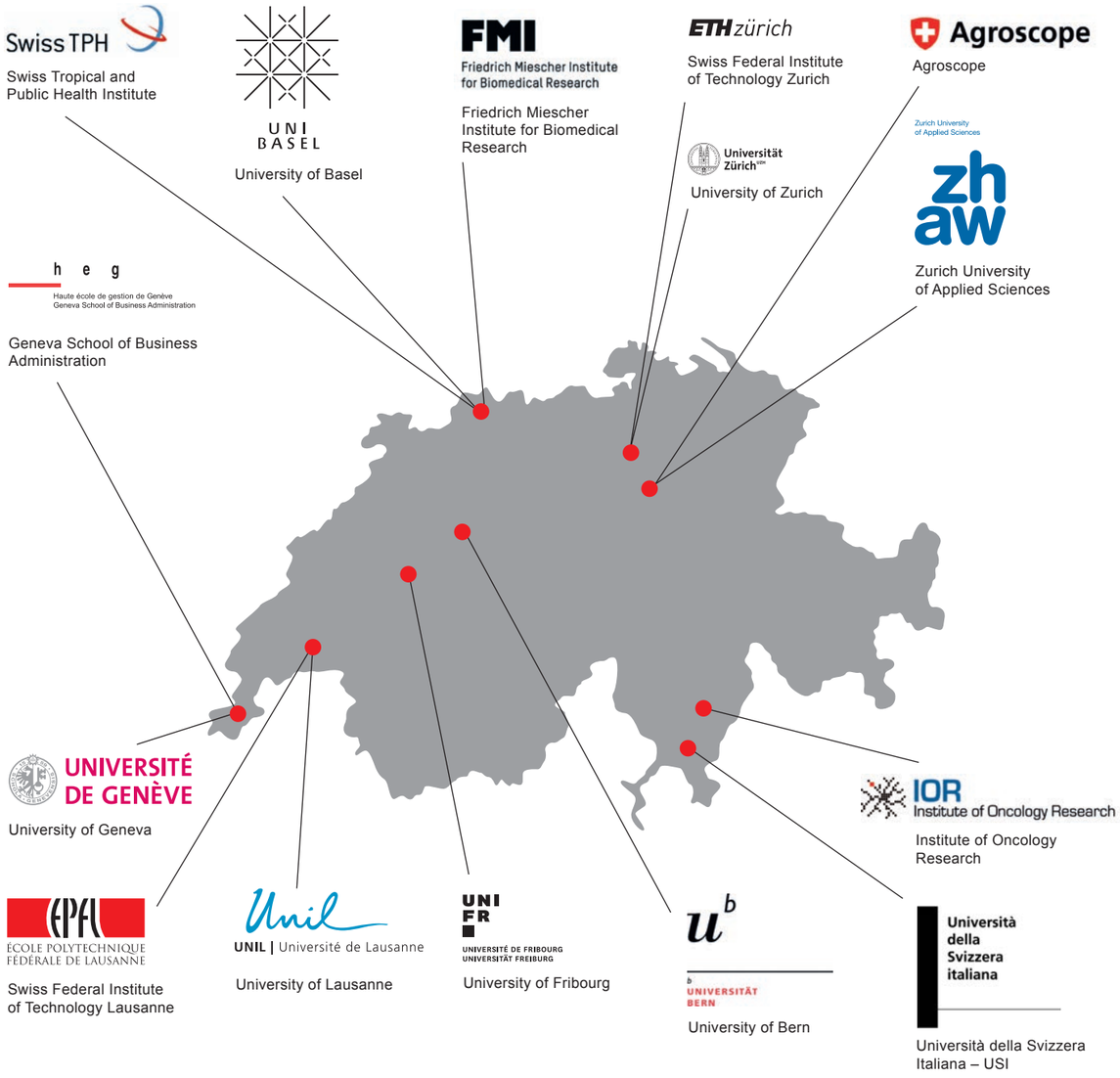
SIB has a unique organization: the group leaders are senior academic staff of partner institutions, while a number of the scientists are paid directly by SIB. Although each research group carries out its own research and teaching activities independently within its host institution, it can benefit from a wide range of resources and support provided by SIB. Moreover, SIB offers its partner institutions an efficient nationwide coordination of bioinformatics research, core resources and teaching activities. In return, the Swiss universities and research institutes provide SIB members with the infrastructure needed to perform their tasks.

campus biotech

SIB sets up offices at Campus Biotech

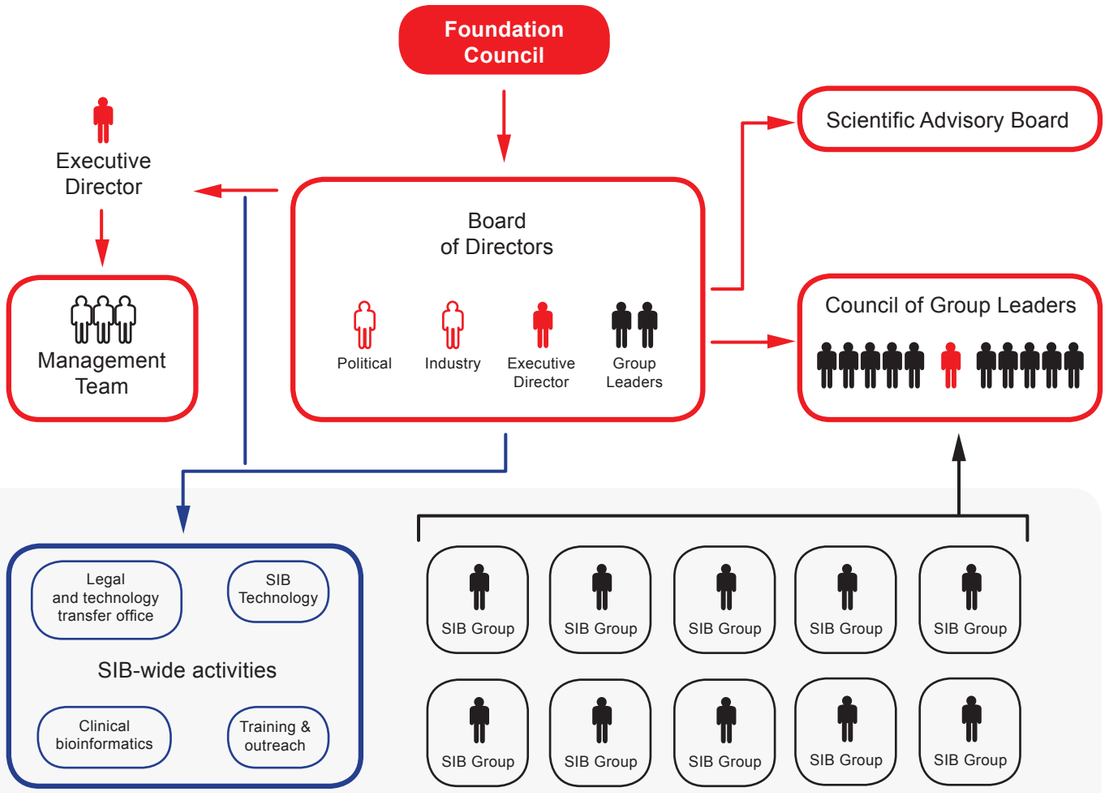
Members of the SIB groups Swiss-Prot and Vital-IT entered Campus Biotech to contribute their expertise to the new centre of excellence in biotechnology and life science research. Campus Biotech is located in the former Merck Serono headquarters in Geneva and hosts academic groups from the University of Geneva and EPFL, a new Wyss Centre for Bio- and Neuro-Engineering and related organizations and businesses. SIB is closely associated with this endeavour, in particular through its involvement in personalized health research. www.campusbiotech.ch

SIB'S INSTITUTIONAL MEMBERS



GOVERNANCE

SIB ORGANIGRAM



The Foundation Council

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

President

Prof. Felix Gutzwiller
Senator

Founding Members

Prof. Ron Appel
Executive Director,
SIB and University of Geneva

Prof. Amos Bairoch
Director, Department
of Human Protein Sciences,
University of Geneva;
and SIB Group leader

Dr Philipp Bucher
Group leader,
SIB and EPFL

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

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

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

Ex officio Members

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

Dr Claire Baribaud
Director,
Geneva School of Business
Administration

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

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

Prof. Carlo Catapano
Director, Institute
of Oncology Research (IOR)

Mr Marc Filliettaz
General Manager,
GeneBio

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

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

Prof. Piero Martinoli
President,
Università della Svizzera italiana

Prof. Andreas Mayer
Vice-Dean,
Faculty of Biology and Medicine,
University of Lausanne

Prof. Philippe Moreillon
Vice-Rector,
University of Lausanne

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

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

Prof. Roland Y. Siegwart
Vice-President Research
and Corporate Relations,
ETH Zurich

Prof. Pierre Spierer
Professor emeritus,
Faculty of Science,
University of Geneva

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

Dr Robert L. Strausberg
Executive Director
of Collaborative Sciences
Ludwig Institute
for Cancer Research (LICR)

Mrs Turkan Tahtasakal
Hewlett Packard, Geneva

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

Prof. Martin Täuber
Rector, University of Bern

Prof. Guido Vergauwen
Rector, University of Fribourg

Prof. Daniel Wyler
Vice-Rector, Medicine
and Natural Sciences,
University of Zurich

Co-opted Members

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

The Board of Directors (BoD)

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

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

Ms Martine Brunschwig Graf
Former National Councillor

Prof. Ron Appel
SIB Executive Director
and University of Geneva

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

Prof. Torsten Schwede
SIB Group Leader and University
of Basel

The Scientific Advisory Board

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

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

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

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

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

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

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

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

The Council of Group Leaders

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.

The SIB Group Leaders are located in SIB partner institutions (see pp. 36-37). In addition, OsiriX led by Prof. Osman Ratib is an Affiliate Group.

Honorary Members

Prof. Ernest Feytmans
Honorary Director

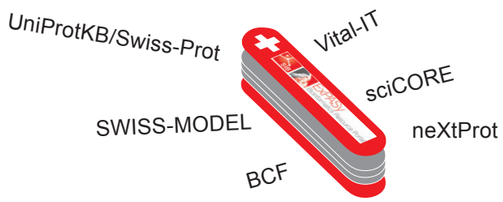
Ms Christiane Langenberger
Former Senator, Honorary President
of the SIB Foundation Council

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

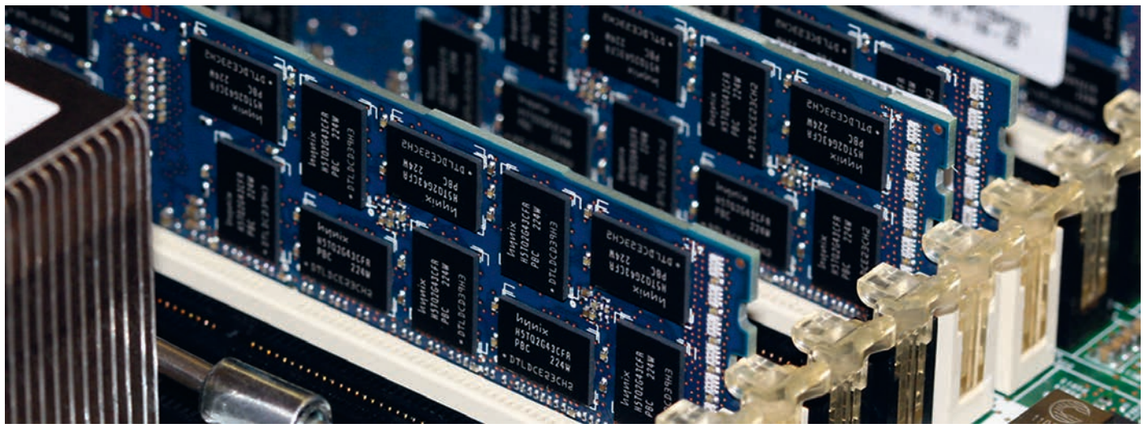
BIOINFORMATICS RESOURCES
FOR MEDICINE AND LIFE SCIENCES

SIB is instrumental to good science in Switzerland and worldwide. The institute provides the necessary bioinformatics services and resources for life scientists.

SIB develops and maintains **more than 150** internationally recognized **databases and software**, which life scientists worldwide use extensively.



Through eight bioinformatics core facilities and high-performance computing centres and more than 30 **embedded bioinformaticians**, SIB provides bioinformatics and statistical support as well as services, expertise and infrastructure to life scientists.



DATABASES AND SOFTWARE PLATFORMS

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

Most of SIB resources are in open-access on ExPASy, the SIB bioinformatics resource portal. The SIB resources cover different areas of life sciences, such as genomics, proteomics and evolution.
www.expasy.org

ExPASy Bioinformatics Resource Portal

CATEGORIES	SUB-CATEGORIES	EXAMPLES OF DATABASES	EXAMPLES OF SOFTWARE
Genes and genomes	Sequence alignment		Codon Suite, LALIG, Newick Utilities
	Similarity search		LALIGN, Phylogibbs
	Characterization/annotation	CLIPZ, EPD, miROrtho, OMA, OpenFlu, OrthoDB, smirnaDB, SwissRegulon	CLIPZ, EPD, ISA, OMA, smirnaDB
	Transcriptomics	Bgee, CleanEx, CLIPZ, smirnaDB, SwissRegulon	CLIPZ, ISMARA, MirZ, PPA, smirnaDB
Proteins and proteomes	Protein sequence and identification	neXtProt, UniProtKB, UniProtKB/Swiss-Prot, ViralZone	LALIGN, 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, PRATT
	Post-translational modification	UniCarbKB, SugarBind, UniProtKB/Swiss-Prot	FindMod, GlycanMass, ISMARA, UniCarbKB
	Protein-protein interaction	STRING, UniProtKB/Swiss-Prot	PredictProtein, ProtBud
	Similarity search/alignment	MyHits, UniProtKB	BLAST, ClustalW, MyHits
	Imaging		ImageMaster / Melanie, MSIght
Medicine and health		SwissSidechain	SwissDock, SwissParam, SwissBioisostere, SwissTargetPrediction
Evolution and phylogeny		Bgee, ImmunoDB, miROrtho, OMA, OrthoDB	Arlequin, CT-CBN, Newick utilities, OMA, TriFLe
Structural biology		SWISS-MODEL Repository, SwissSideChain	SwissDock, SWISS-MODEL Workspace, Swiss-PdbViewer
Systems biology		Progenetix, SwissRegulon	arrayMap, MetaNetX, The Systems Biology Research Tools
Bioinformatics infrastructure			nfswatch, Soaplab services

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

SOME OF SIB'S CORE RESOURCES

UniProtKB/Swiss-Prot

Protein sequence database

UniProtKB/Swiss-Prot is the most widely used protein information resource in the world, with over 900,000 requests per month. It provides concise, but thorough, descriptions of a non-redundant set of over 547,000 proteins including their function, domain structure, post-translational modifications and variants. Its high-quality annotation is the fruit of expert curation by biologists, who use information available in the scientific literature to provide an accurate description of each protein's features (see also p. 50).

neXtProt

Human protein knowledge platform

neXtProt is an innovative knowledge platform dedicated to human proteins. neXtProt includes curated information on various aspects of human protein biology such as function, mRNA/protein expression, protein/protein interactions, post-translational modifications and protein variations. Its aim is to help life science researchers in their quest to unravel the complexity of human life processes (see also p. 48).

SWISS-MODEL

Structure homology-modelling

SWISS-MODEL is an automated protein structure homology-modelling server for generating 3D models of a protein. Comparative approaches are currently the most accurate and reliable computational methods to derive 3D models of proteins, for which experimental structures are not available. SWISS-MODEL automates the complex process of model building on an easy to use web-based system, thereby making model information also available for non-specialists (see also p. 65).

STRING

Protein-protein interactions

STRING is a database of known and predicted protein-protein interactions, including direct (physical) and indirect (functional) associations. They are derived from different sources such as the genomic context, high-throughput experiments, (conserved) co-expression, and the literature. STRING quantitatively integrates the interaction data for a large number of organisms, and transfers information between these organisms where applicable. The database currently covers 5,214,234 proteins from 1,133 organisms (see also p. 49).

SwissRegulon

Annotations of regulatory sites

The SwissRegulon web portal provides information and tools for the analysis of genome-wide transcription regulatory networks in organisms ranging from *E.coli* to human. The database frontend offers an intuitive interface showing genomic information in a clear and comprehensible graphical form (see also p. 44).

SwissDrugDesign

Drug design

The SwissDrugDesign project is an ambitious initiative aiming to provide the first comprehensive, integrated and freely accessible web based *in silico* drug design environment to the scientific community worldwide. It offers a large collection of tools covering all aspects of computer-aided drug design, from target prediction of small molecules (SwissTargetPrediction) to the provision of topology and parameters of drug-like molecules (SwissParam). Other tools include SwissDock, SwissSideChain and SwissBioisostere (see also p. 65).

EPD

Collection of eukaryotic promoters

Genes are first "transcribed" into messenger RNA used as templates to synthesize proteins. Transcription begins on "promoters". The Eukaryotic Promoter Database (EPD) provides quality-controlled information on experimentally defined promoters of higher organisms as well as web-based tools for promoter analysis (see also p. 42).

Bgee

Gene expression evolution

Bgee is a database of gene expression evolution, which integrates all types of transcriptome and expression information for animals – including human, model organisms such as mouse or *Drosophila*, and diverse species of evolutionary or agronomical relevance. Bgee is the only resource which provides homologous expression between species (see also p. 60).

UniCarbKB

Glycan database

UniCarbKB is a knowledgebase that offers public access to a curated database of information on glycoproteins, which are proteins to which carbohydrates (or glycans) are attached. UniCarbKB provides comprehensive information on glycan structures, and published glycoprotein information including global and site-specific attachment information (see also p. 49).

SugarBind

Pathogen sugar-binding

Host-pathogen communication is known to be mediated by carbohydrate-protein interactions, which ensure adhesion at cell surface. The SugarBind database covers knowledge of interactions between pathogens and carbohydrate ligands of mammalian hosts. Information in SugarBindDB is manually curated and supports the investigation of microbial or viral infections (see also p. 49).

OrthoDB

Hierarchical catalogue of orthologs

OrthoDB is a catalogue of "equivalent" genes among species, called orthologs. Resolving gene ancestry is the most accurate way to predict putative gene functions by association with genes studied in model organisms. OrthoDB is hence critical both for evolutionary studies and for interpreting gene content from newly sequences genomes (see also p. 46).

OMA

Orthology prediction

The Orthology MAtrix (OMA) Browser provides orthology predictions among publicly available genomes. Started in 2004, it has undergone 17 releases and now elucidates orthology among 8.82 million genes from 1,706 species, making it one of the largest resources of its kind (see also p. 58).

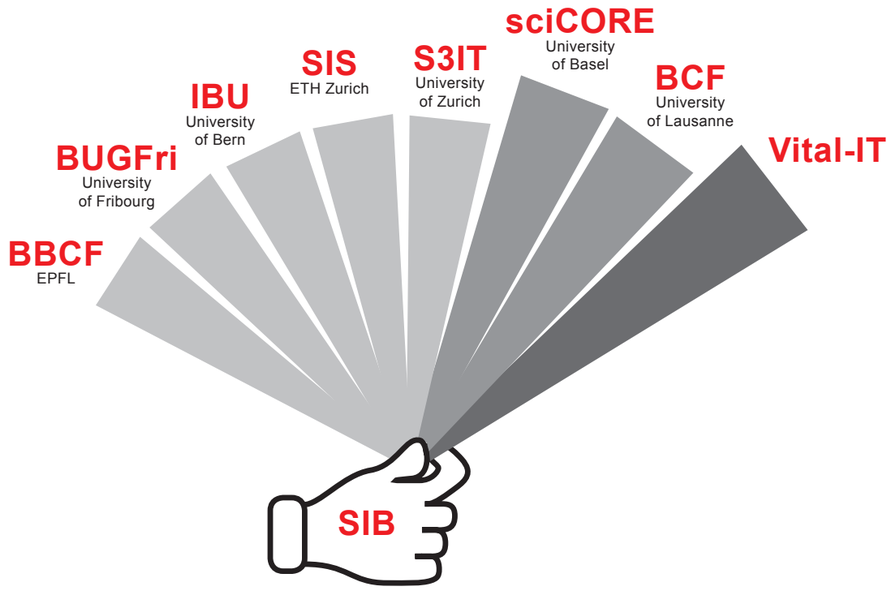
CORE FACILITIES AND HIGH-PERFORMANCE COMPUTING CENTRES

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

The services provided include analysis of high-throughput data (genome/exome sequence, RNA sequencing, proteomics), scientific support of (bio)medical projects, development of algorithms, biostatistics training, as well as access to computational space, helpdesk and support.

The SIB **Vital-IT** group provides computational infrastructure, development support and bioinformatics expertise to the life science community. SIB also co-manages the Center for Scientific Computing (**sciCORE**) and the Bioinformatics Core Facility (**BCF**), and collaborates with the Service and Support for Science IT (**S3IT**) facility, the Scientific Information Services (**SIS**), the Bioinformatics and Biostatistics Core Facility (**BBCF**), the Bioinformatics Unravelling Group (**BUGFri**) and the Interfaculty Bioinformatics Unit (**IBU**) (see figure below).

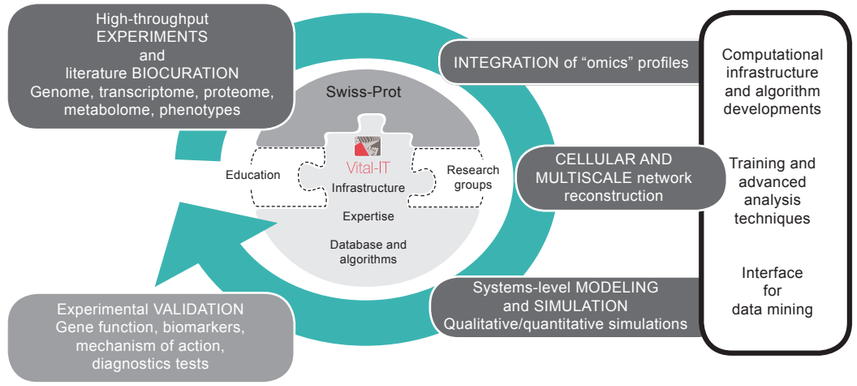
In addition to its bioinformatics core facilities and HPC centres, SIB provides assistance to the Swiss universities and institutes by embedding bioinformaticians in laboratories. The presence of these **embedded bioinformaticians** in the research groups is a real advantage, as they can provide direct guidance on how to manage and analyse data for optimized use of the various bioinformatics tools.



Vital-IT

Ioannis Xenarios' group

Vital-IT is a bioinformatics competence centre that supports and collaborates with life scientists in Switzerland and beyond. The multidisciplinary team provides expertise, training and support in high-performance computing (HPC) and storage resources and infrastructure, so as to help develop, maintain and extend life science and medical research. Additionally, Vital-IT performs research within the group and in collaboration with groups worldwide and serves as an interface between academia and industry.
See also p. 75 and www.vital-it.ch.



Services

- Providing an HPC environment to support the research work of its partners in areas ranging from sequence analysis, over molecular modelling, to image processing
- Developing new computer algorithms to meet the users' needs and remain up-to-date with the new generation of computers
- Maintaining and developing specialist software engineering techniques for parallelization, optimization and validation of complex algorithms
- Development activities to turn concepts derived from research into robust software solutions
- Consulting and educational activities geared towards the computational needs of life science companies
- Acting as an agent for new collaborations with industry, with potential for spinning off new companies in the field of life science informatics

Vital-IT in figures

More than
5 million
CPU hours
consumed every year

More than
10 million
jobs
run by users
every year

5
Petabytes
of storage capacity

+

5,000
CPUs
in 2014

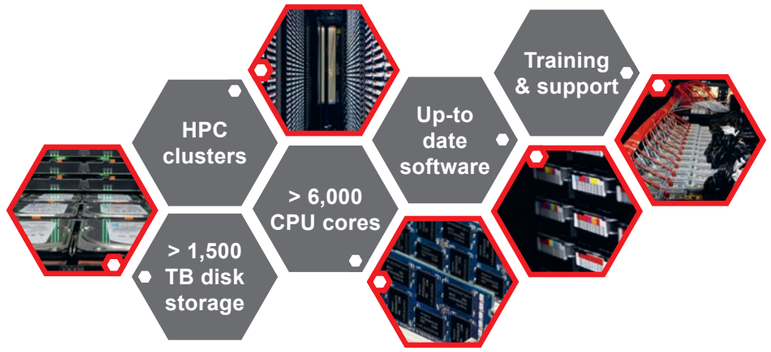
Author of,
or acknowledged in,
more than
90
publications
in 2014 including
28
in *Science* or *Nature*
over the last 5 years

760
users
in 2014 *versus* 671 in 2013

sciCORE

Center for Scientific Computing
Torsten Schwede's group, University of Basel

sciCORE is a centre of competence in scientific computing – providing high-performance computing (HPC) infrastructure, large-scale storage resources, scientific software and databases, server infrastructures, user support, training and teaching, as well as know-how and expertise to scientific research groups. sciCORE provides a professional environment for a broad spectrum of scientific applications, ranging from bioinformatics, computational chemistry, physics, systems biology, to medicine and economics. In direct collaboration with scientific research groups, sciCORE helps develop, deploy, operate and extend the computational tools required for performing modern life science and biomedical research. sciCORE operates the IT infrastructure for several of SIB's services, such as SWISS-MODEL and SwissRegulon, which are widely used by the international research community.
See also p. 65 and http://scicore.unibas.ch.



Services

- Operation of a professional storage infrastructure for scientific data, which is required for managing the vast amounts of data produced in modern life sciences
- Development and operation of a state-of-the-art HPC infrastructure environment to support large-scale data analysis, modelling and simulation applications
- Support to research groups and projects on complex data management questions, from requirement analysis to development, deployment and operation of productive software solutions
- Consulting for researchers on bioinformatics and computational biology aspects of their projects
- Support to users in computational aspects of project definition for applications to national and international grant programs
- Organization of training and teaching courses in scientific HPC, bioinformatics, biostatistics and data visualization
- Operation of computational scientific applications as professional and robust services for the life science community

BCF

Bioinformatics Core Facility
Mauro Delorenzi's group, University of Lausanne

The core competence and activities of BCF reside in the interface between biomedical sciences, statistics and computation, particularly in the application of high-throughput "omics" technologies, such as all sequence- and array-based methods, to problems of clinical importance including development of cancer biomarkers. BCF offers consulting, teaching and training, data analysis support and research collaborations for both academic and industrial partners. In particular, BCF provides a consulting service on biostatistics matters, with a focus on high-throughput technologies in genomics or proteomics. This service is aimed at all people active in life sciences in Switzerland.
See also p. 53 and http://bcf.isb-sib.ch.

Services

- Statistical consulting
- Education and practical training
- Data analysis
- Personnel embedding
- Scientific collaboration

S3IT

Service and Support for Science IT
Peter Kunszt's group, University of Zurich

S3IT is proactively participating and partnering in collaborative research projects with a clear focus on support and enabling IT technologies for science. S3IT assists in resolving current and future challenges around data collection, data validation and analysis, as well as data publication and reproducibility.
See also p. 74 and www.s3it.uzh.ch/index.en.html.

Services

- Partnership with research groups and projects in computational and data intensive research
- Access to competitive infrastructure: HPC supercomputing resources, capacity cloud computing resources as well as optimized server infrastructures for research with access to standardized tools and software
- Data management, automation and data analysis capabilities
- Consultancy and development of software and methods for usage in scientific projects
- Re-engineering of existing software and tools, for contribution to community resources
- Training and education in the usage of the infrastructure and software

SIS

Scientific IT Services
Bernd Rinn's group, ETH Zurich

SIS is an interdisciplinary service group enabling and supporting computing- and data-intensive research. To this end, SIS collaborates with researchers at ETH Zurich and beyond on solving scientific computing problems, and is involved in national and international scientific IT efforts, such as SystemsX SyBIT (http://sybit.net), FAIR-DOM (http://fair-dom.org), HPC-CH (http://www.hpc-ch.org) and Swiss universities eSCT.
See also p. 74 and http://sis.id.ethz.ch.

Services

- Big data and HPC infrastructure, and related support
- Flexible and robust data management and integration solutions for life science data
- Open research data portal and data publication services
- Custom data analysis pipelines
- Scientific software engineering services
- Consulting and training on resolution of scientific computing and software problems

BBCF

Bioinformatics and Biostatistics Core Facility
Jacques Rougemont's group, EPFL
BBCF supports research in life sciences by providing data analysis for large-scale and complex-design genomics experiments, by developing original software and methods, and by providing state-of-the art bioinformatics applications and genomic data through web portals. The facility currently focuses on complex high-density DNA arrays and high-throughput sequencing.
See also p. 75 and http://bbcf.epfl.ch.

Services
- Data analysis and statistical consulting
- Software and database development
- Design and implementation of data management solution
- Bioinformatics training for life science researchers

BUGFri

Bioinformatics Unravelling Group
Laurent Falquet's group, University of Fribourg
BUGFri supports life science researchers by providing expertise in data analysis of next-generation sequencing (NGS) experiments, or any large-scale biological experiment requiring bioinformatics resources.
See also p. 43 and www.unifr.ch/bugfri.

Services
- Project planning and grant writing
- Software testing and development
- Data management and analysis
- Training and teaching

IBU

Interfaculty Bioinformatics Unit
Rémy Bruggmann's group, University of Bern
IBU provides services and expertise to assist researchers in data analysis and project planning in the context of NGS.
See also p. 41 and www.bioinformatics.unibe.ch.

Services
- Discussions with researchers on their projects prior to starting with IBU's bioinformatics experts and – if required – experts from the wet lab (e.g. NGS lab)
- Provision of data analyses including short read mapping, *de novo* assembly, mutant screening, transcriptomics, epigenomics, whole exome and genome analysis
- Collaboration in large and complex projects

LEADING AND COORDINATING
BIOINFORMATICS IN SWITZERLAND

SIB is developing at rapid pace and its role on the bioinformatics scene is expanding. Efficient coordination of activities across SIB groups and with external stakeholders is more than ever crucial.

SIB has created four groups dedicated to the following activities:

- **Clinical bioinformatics group**
  which aims at developing bioinformatics for personalized health, to facilitate the analysis and the use of "big data" by clinicians, in order to support diagnosis as well as preventive and therapeutic approaches

- **Legal and technology transfer office**
  with the goal of enabling the scientific community to benefit from SIB's many innovations in the field of bioinformatics

- **Technology group**
  to optimize the coordination of all technology-related activities within SIB and across SIB groups

- **Training and outreach group**
  to respond to the ever-growing need for qualified bioinformaticians, as well as to provide professional training in bioinformatics to the life science community

CLINICAL BIOINFORMATICS: PAVING THE WAY TO A FINER KIND OF HEALTH

The medical world is about to undergo a revolution called personalized health. It all started with the development of technologies that produce gigantic quantities of data on almost all parts of a human being and tools to process them so as to sketch a very personal and specific portrait of an individual's state of health. These developments are expected to greatly assist clinicians in the tailoring of preventive measures and treatments for their patients. Bringing the fruits of biotechnology into medical practice is one of the roles of clinical bioinformatics.

Personalized health: a definition
The rapid development of technologies in the past decades and the huge volumes of data they produce on an individual level will give rise to a very different type of medical practice, also usually referred to as precision medicine, personalized medicine or personalized health.

Medical practitioners can already have access to data, which are unique to their patients – such as their genome sequence for instance – not to mention data supplied by various smartphone applications brought along by the patients themselves. Coupled with the individual's lifestyle and eating habits, the information made available to a clinician on a patient's state of health will be not only more and more personal but also unique.

As a result, in the not so distant future, such knowledge will assist clinicians in a patient's diagnosis and in deciding on tailored treatment, besides tracking down the propensity for a disease before its actual onset, and delaying, if not preventing, diseases.

Several countries, such as the UK, Korea, China and the USA, have already launched extensive programs to speed up the process.

Bioinformatics and the medical field – the future of medicine
Bioinformatics is, or will be, at the very heart of this medical revolution. For several years now, SIB has been involved in research whose applications are directly linked to the medical field.

SIB's high-performance computing centre Vital-IT, for example, has developed the algorithm for a non-invasive prenatal test which can detect the most frequent trisomies and chromosomal rearrangements from a pregnant woman's blood sample, and it has been commercialized. Another SIB group developed a model which predicts the evolution of an aneurysm, and is intended to support clinicians in deciding upon the best treatment to offer a patient. Yet another group offered their support during last year's outbreak of the Ebola virus in West Africa by estimating the infection's dynamics – information of crucial importance for controlling an epidemic.

These are just three of the many projects developed by SIB in the field of human health – not to mention those that are involved in the diagnosis and treatments of different kinds of cancer.

Jacques Beckmann

SIB has the know-how to develop standardized tools for diagnostic and clinical applications

Clinical bioinformatics, SIB's choice
As personalized health is gaining ground, a very particular field of bioinformatics, known as clinical bioinformatics, is emerging – it is bioinformatics **dedicated to the storage, organization and analysis of data pertaining to an individual's state of health, which can be used and understood by clinicians.**

SIB saw early on the role it could play within the realm of personalized health, and the subsequent support it could offer to clinicians. Consequently, back in 2012 the Institute created a group dedicated to clinical bioinformatics, under the leadership of Jacques Beckmann. The initial aim was to identify areas where the analyses of "big data" are most likely to enter the clinical arena, and then prioritize these domains and motivate stakeholders to contribute to them.

SIB has the know-how to develop standardized tools for diagnostic and clinical applications. There is already, for instance, a growing demand by clinicians for the interpretation of a patient's genome in the diagnosis of inherited disorders or the sequencing of tumour cells so as to offer customized treatment. It is also paramount to adopt common, cost-effective and inter-operable procedures that are both flexible and compatible with procedures used throughout the various Swiss diagnostics and medical centres.

The rigour and quality of SIB's services, software and databases offer major advantages since they can be grafted onto existing procedures and provide enhanced power of interpretation. What is more, the Institute ensures the continuous maintenance of state-of-the-art performances and options for new developments.

Such a revolution also entails ethical, security and educational issues. Clinicians will have to learn how to deal with databases, how to interpret the data and where the boundaries of privacy lie.

Clinical bioinformatics, crossing boundaries
Hospitals, diagnostic centres and medical practitioners need to cooperate – both on the regional and national scale – to build a pool of data that can be shared, so that comparisons of all sorts can be made across the population.

At the regional level, a section of SIB moved into Geneva Campus Biotech – a new centre of excellence in life science research – which brings together other groups from academic institutions involved in personalized health, such as the Universities of Geneva, Lausanne and the EPFL. **On the national level, SIB took an active part in elaborating a proposal for a federal personalized health programme.**

On the international front, in 2014 SIB became one of the members of the Global Alliance for Genomics and Health, itself founded in 2013 by 50 scientists from eight different countries. Today, the Global Alliance is represented by over 220 leading institutions involved in healthcare, research, disease advocacy, life sciences and information technology with a single aim: to create a global framework of approaches that will enable the responsible, voluntary and secure sharing of genomic and clinical data.

LEGAL AND TECHNOLOGY TRANSFER OFFICE: 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 showcase, manage and transfer SIB knowledge and know-how so that the scientific community can benefit from SIB's many innovations.

Marc Filliettaz

The LTTO, headed by Marc Filliettaz, strives to enhance the scientific and industrial visibility of SIB innovation by assisting SIB members in their contacts with external partners. The LTTO works closely with SIB's Group Leaders in order to offer innovative products and services.

Partnerships with industry
Companies involved in medicine and life sciences can collaborate with SIB to complement their internal capacity. Outsourcing certain activities and relying on a non-industrial, world-renowned network of scientists and bioinformaticians provides companies with additional flexibility and enhances their research.

SIB collaborates with industrial partners in the following domains:

- *Scientific support and data analysis:* SIB has developed in-house computational tools and acquired in-depth expertise in a broad range of fields and techniques, including cancer subtype discovery, biomarker selection, class discrimination, cross-platform analysis, and meta-analysis of publicly available clinical and genomics multiple datasets.

- *Education and practical training:* SIB provides scientists with educational support and practical training in the use of software and analysis methods. This includes the organization of seminars, workshops and training courses.

- *Text mining/web monitoring:* SIB develops text mining tools, e.g. enabling the creation of a patient cohort based on patient records in text format or the monitoring of social media platforms in the context of drug safety surveillance.

SIB TECHNOLOGY: OPTIMIZING TECHNOLOGY-RELATED ACTIVITIES

The SIB Technology group, headed by the Chief Technology Officer (CTO) Heinz Stockinger, is in charge of optimizing the coordination of all SIB technology-related activities. The group works in close cooperation with technology and infrastructure providers and competence centres such as Vital-IT/Swiss-Prot and sciCORE to combine forces wherever necessary and practical.

Heinz Stockinger

The SIB Technology group's core competencies are:

- Design, development, testing and operation of scientific, technical and administrative software in bioinformatics, in cooperation with various SIB groups, with a strong focus on web and internet technologies – mainly with the SIB Web Team.

- Technical coordination of topics that require an SIB-wide approach: web application deployment, security and related guidelines, code repositories, etc.

- Support and operation of SIB-wide services developed and/or deployed by the group, such as:
  - ExPASy SIB Bioinformatics Resource Portal and some scientific resources available on the portal
  - Request tracking operations for user support
  - Web applications for SIB course registration and administration
  - Web applications for SIB personnel management/administration

- Coordination of SIB technical activities within the ELIXIR project and operation (more information about ELIXIR on p. 30).

TRAINING AND OUTREACH: TRAINING THE NEXT GENERATION OF BIOINFORMATICIANS

It is essential to train the next generation of bioinformaticians and to ensure that life scientists make the best of bioinformatics and SIB resources. Under the leadership of Patricia Palagi, the Training and outreach team is in charge of promoting bioinformatics to the layman, as well as coordinating education and training in the field, both in Switzerland and internationally.

Patricia Palagi

Training in 2014

SIB's training strategy in 2014 followed the same trend as in past years, focusing on professional training, the PhD training network and international collaborations.

Professional training

The SIB professional training portfolio, which is continuously evolving to meet the scientific community's needs, proposes courses on several topics, such as bioinformatics techniques, computational biology methods, statistics and next-generation sequencing analysis.
To respond to an increasing demand for basic training, a new "First Steps with" course series was started in 2014, with a one-day course on the UNIX computer operating system. Similar courses on R, the statistical package, and Python are in development.
You can find more information on SIB courses on its training portal: www.isb-sib.ch/training.

PhD training network

The PhD training network specifically targets Swiss bioinformatics and computational biology PhD students. It fosters interactions among students and trains them on selected hot topics in bioinformatics. Special events of 2014 included the annual retreat and the summer school on "Systems Medicine and its applications" jointly organized with the Swiss Initiative in Systems Biology (SystemsX).
For more information on the PhD training network, visit:
www.isb-sib.ch/training/sib-phd-training-network.html.

International collaborations

SIB organized the Workshop in Education for Bioinformatics during the Intelligent Systems for Molecular Biology (ISMB) 2014 conference in Boston, in collaboration with the Global Organization for Bioinformatics Learning, Education & Training (GOBLET), of which SIB is a member. SIB has developed strong connections with the European bioinformatics training community as well, especially within ELIXIR, and various projects are underway.

SIB Fellowship programme

Through its Fellowship programme launched in 2012, SIB aims to create a pool of excellent young bioinformaticians. Thanks to the generous support of committed partners, the best students chosen worldwide have the opportunity to carry out their PhD research in one of the Institute's groups.
In the fall of 2014, a new call was launched. The selection committee, made up of internationally renowned professors and SIB Group Leaders, will choose the laureates in the first quarter of 2015.

SIB professional training in 2014 in figures:

23 short workshops organized

11 long block-courses coordinated

More than **650** participants trained

Involved experts and teachers

More than **50** SIB members

More than **25** invited speakers

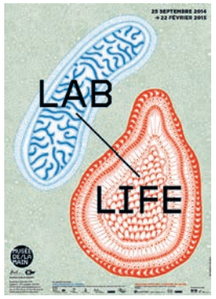
Outreach in 2014

Another SIB mission is to bring bioinformatics to the layman, contributing to a better understanding of this relatively new science.

At the national level:

SIB took part in various community events, such as:

- **Popular science fairs, including** "La Nuit de la Science" in Geneva, "Les Mystères de l'UNIL" in Lausanne, and the first "Swiss Health Fair" at EPFL, to promote public awareness and understanding of bioinformatics;
- **Workshops for high-school students** in the French-speaking part of Switzerland, some within the framework of a special collaboration with the Bioscope and the "Ramène Ta Science" initiative (both of which are University of Geneva projects), and during the TechDays in Lausanne, Köniz and Locarno;
- **Public scientific exhibitions**, such as the LAB/LIFE exhibition of the Musée de la Main (Museum of the Hand) in Lausanne;
- **"Music for a Gene"**: In a project directed by Lydie Lane (co-director of the SIB group CALIPHO), the French composer Olivier Calmel was asked to write a string quartet that could transpose the complexity of the human genome into musical emotions. The concert Opus 23 – Music for a Gene was performed by the Ramses Quartet in Geneva and attracted significant interest from the Swiss media.



At the international level:

- **"Chromosome Walk" exhibition**: In 2014, this exhibition went beyond the Swiss borders. It was shown in the "Village des Sciences de GENOPOLYS" in Montpellier, France.
  The virtual version is available on www.chromosomewalk.ch.
- **Workshops for high-school teachers**: SIB developed two bioinformatics workshops for high-school teachers organized by GOBLET, one in Toronto and the other in Boston.

SIB – THE SWISS NODE OF ELIXIR

The demand for bioinformatics support is exploding, and the need to coordinate all efforts and improve technology development in this field is, more than ever, necessary. Europe's awareness of the need for an international bioinformatics infrastructure is embodied by ELIXIR, the European Life Science Infrastructure for Biological information that was officially launched in Brussels in December 2013.

Vision
The amount of data produced by life science experiments is rapidly expanding. Some of these datasets are highly specialized and would previously only have been available to researchers within the country in which they were generated. Open access to the biological datasets will facilitate discoveries and support life science research and its translation to medicine, agriculture, bioindustry and society.

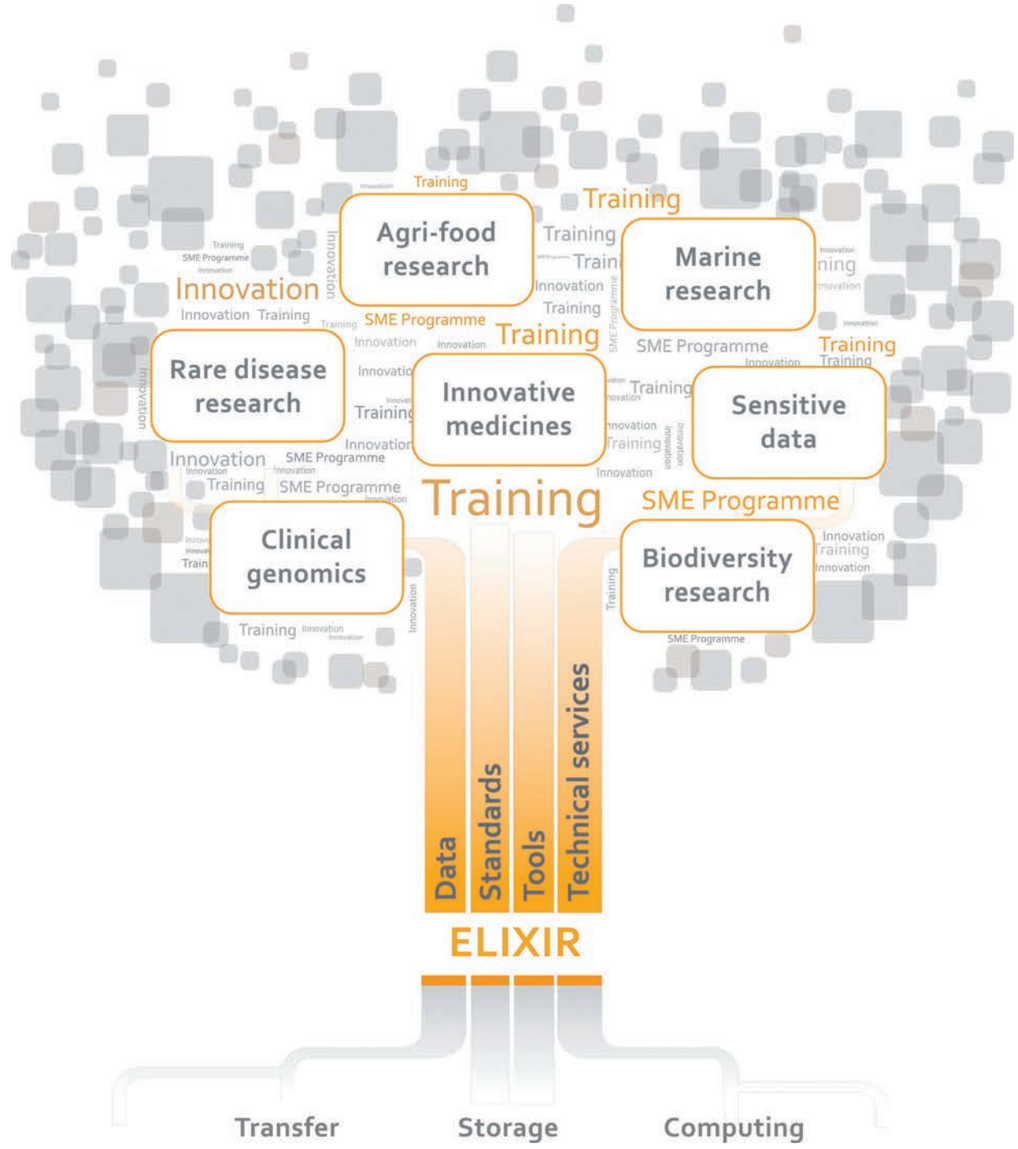
Mission
The goal of ELIXIR is to build a responsive and sustainable European infrastructure to orchestrate the collection, quality control and archiving of large biological datasets. ELIXIR integrates research data from different parts of Europe and ensures a service provision that is easily accessible to all.

SIB represents Swiss bioinformatics at ELIXIR
To date, 11 countries, including Switzerland, and the European Molecular Biology Laboratory-European Bioinformatics Institute (EMBL-EBI) are members of the ELIXIR consortium. Switzerland was one of the first countries to sign the ELIXIR Consortium Agreement and has been on board since the very first proposals concerning this impressive undertaking. SIB acts as the Swiss node of ELIXIR and is the biggest national node within the organization. As a provider of various renowned resources to the international life science community, SIB plays an important role in the ELIXIR project.

For more information about ELIXIR, please visit www.elixir-europe.org.

Highlights ELIXIR 2014

- ELIXIR was identified by the European Council as one of Europe's three priority research infrastructures that are pushing the boundaries of scientific excellence. As a result, ELIXIR was invited to respond to the Horizon 2020 call INFRADEV-3, which will fund development of new world-class research infrastructures. The project that has been submitted is called ELIXIR-EXCELERATE and SIB is closely involved.
- The scientific programme for 2014-2018, prepared in collaboration with more than 100 scientists from all of ELIXIR's national nodes, was published.
- Torsten Schwede, leader of the SIB Computational Structural Biology Group and member of the SIB Board of Directors, was appointed Chair of the ELIXIR Board, starting from 1 January 2015.



ELIXIR provides the facilities necessary for life science researchers to make the most of the rapidly growing store of information about living systems, which is the foundation on which the understanding of life is built.

IN THE SPOTLIGHT IN 2014

SIB and its director honoured with the BioAlps Award 2014

Ron Appel received the **BioAlps Award 2014**, recognizing his major contribution to the creation and successful development of SIB. The purpose of the BioAlps Award is to honour a person without whom western Switzerland would not enjoy its extraordinary international reputation in the field of life sciences.

Benoît Dubuis, President of BioAlps, presented the award to Ron Appel and SIB at the traditional BioAlps Networking Day, stating that "as co-founder and Executive Director of SIB since 2007, Ron Appel has been a key figure in enabling Swiss bioinformatics to rank among the world leaders in this discipline. He is leading one of the rare life science organizations, which is able to bring together so many entities with so much elegance".

"I am proud to receive this award which honours SIB's expertise and the contribution of Swiss bioinformaticians to the progress of life sciences."

Ron Appel, winner of BioAlps Award 2014

Cloëtta Prize 2014 awarded to Henrik Kaessmann

The Professor Max Cloëtta Foundation, based in Zurich, promotes medical research and related disciplines in the natural sciences in Switzerland. In 2014, two researchers were awarded the annual Cloëtta Prize, one of whom was Henrik Kaessmann, leader of the SIB group "Functional Evolutionary Genomics". Henrik Kaessmann was recognized in the fundamental research area for his discoveries in the field of molecular genetics, in particular the nature of diversification in higher mammals.

Three SIB Group Leaders elected members of EMBO

The world-renowned European Molecular Biology Organization (EMBO) announced that 106 outstanding researchers in life sciences, **including three SIB Group Leaders** – Emmanouil Dermitzakis, Henrik Kaessmann and Andreas Wagner – were newly elected to its Membership. EMBO Members make invaluable contributions to the organization by providing suggestions and feedback on the activities of EMBO. They serve on selection committees for EMBO programmes and mentor young scientists. The EMBO Membership currently comprises more than 1600 life scientists.

Winners of SIB Awards 2014

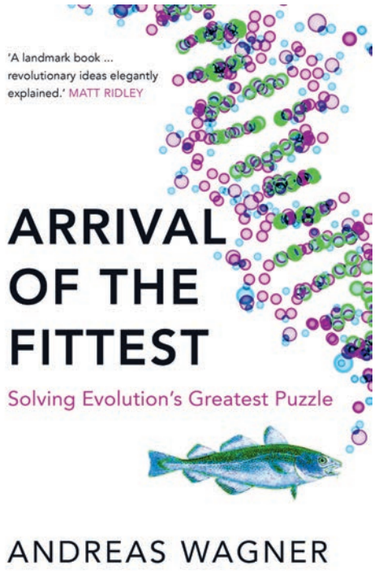
Joshua Payne received the **SIB Young Bioinformatician Award** for his promising work on "The robustness and evolvability of transcription factor binding sites", which he conducted at the University of Zurich, in the SIB group Evolutionary Systems Biology led by Andreas Wagner.

The **SIB Best Graduate Paper Award** went to Josephine Daub. She was honoured for her work on "Evidence for polygenic adaptation to pathogens in the human genome", carried out at the University of Bern, in the SIB group Computational Population Genetics led by Laurent Excoffier.

Andreas Wagner's book selected among the best popular science books of 2014 by The Guardian

Darwin's theory of natural selection explains how useful adaptations are preserved over time. But the biggest mystery about evolution eluded him. In his book, SIB Group Leader Andreas Wagner draws on over 15 years of research to present the missing piece in Darwin's theory. Using experimental and computational technologies that were heretofore unimagined, he found that adaptations are not just driven by chance, but by a set of laws that allow nature to discover new molecules and mechanisms in a fraction of the time that random variation would take.

The book was selected among the best popular science books of 2014 by The Guardian.
More information on http://arrival-of-the-fittest.com



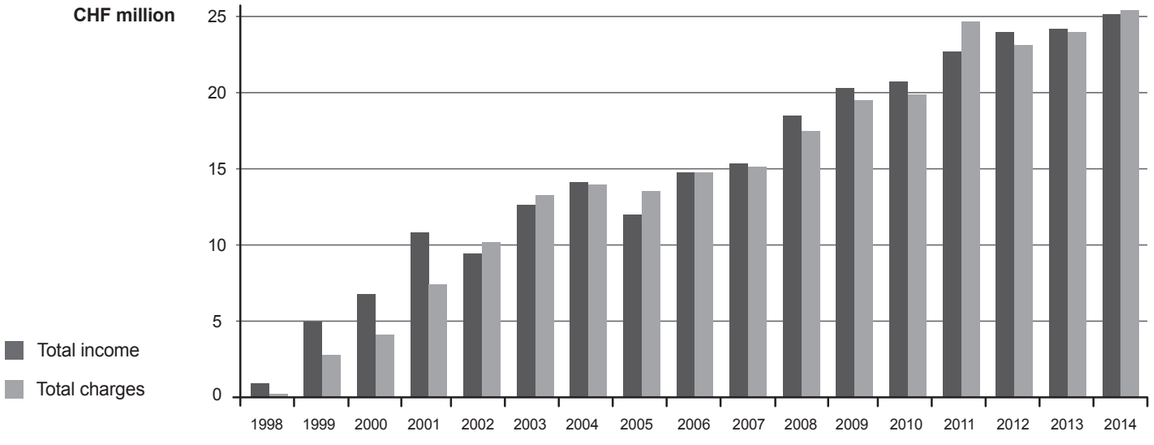
Christine Durinx joins SIB as Associate Director

To support Ron Appel in his role as Executive Director, SIB reinforced its leadership by appointing Christine Durinx as Associate Director. Christine Durinx holds a PhD in pharmacy and has ten years' experience in the pharmaceutical industry. Before joining SIB, she was Director of International Medical Communication for a major firm.

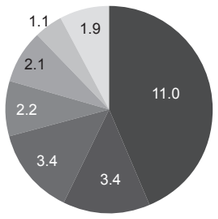
FINANCE

In a context of challenging times for research funding worldwide, SIB funds remained stable in 2014, thanks to the continued support of its funders.

In 2014, the total income of SIB groups reached CHF 67.9 million, of which **25.1 were managed by SIB**.



The first source of SIB funds is the Swiss government (11.0 CHF million, 43.8%), followed by the Swiss National Science Foundation/European funds and the industry (CHF 3.4 million, 13.5% each).



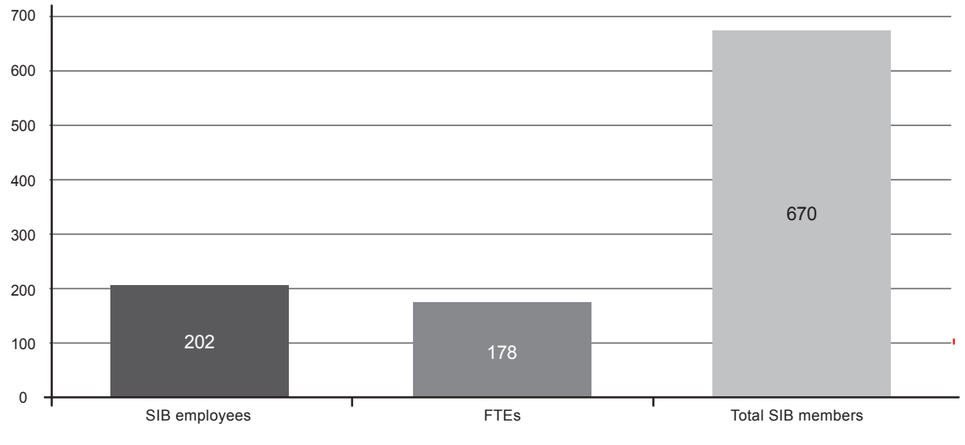
Funds managed by SIB (CHF million)

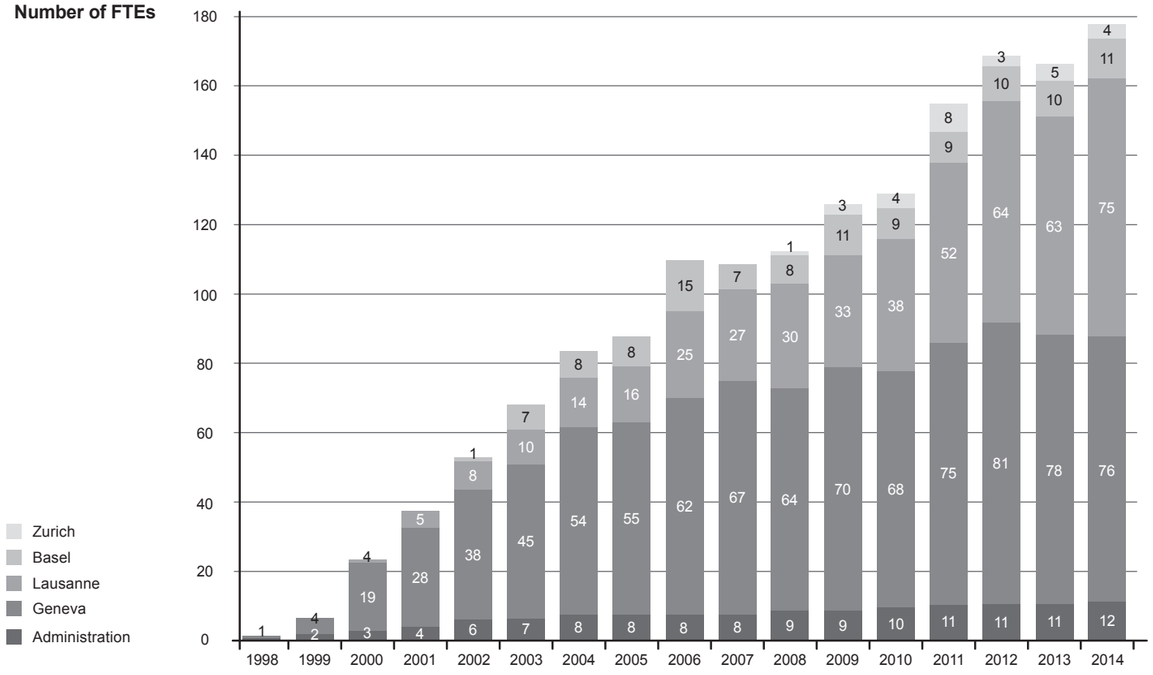
	Funds managed by SIB		Funds managed by the partner institutions		Total number of grants/ contracts
	in CHFmillions	%	in CHFmillions	%	
Swiss government (article 16)	11.0	43.8%			
SNSF and European funds	3.4	13.5%	9.7	22.7%	117
Industry	3.4	13.5%	1.1	2.6%	23
NIH	2.2	8.8%	0.1	0.2%	3
Universities and hospitals	2.1	8.4%	26.2	61.2%	59
SystemsX.ch	1.1	4.4%	4.8	11.2%	53
Other	1.9	7.6%	0.9	2.1%	17
Expenditures					
Wages and expenses		86.1%			
Facilities and functioning		13.9%			
	25.1		42.8		

STAFF

SIB continues to grow in terms of the number of staff members.

As of 31 December 2014, SIB staff was composed of **670 SIB members**, of whom 202 had a contract with SIB, which represented 178 full-time equivalents (FTEs).





RESEARCH AND SERVICE ACTIVITIES

SIB COLLABORATIVE NETWORK

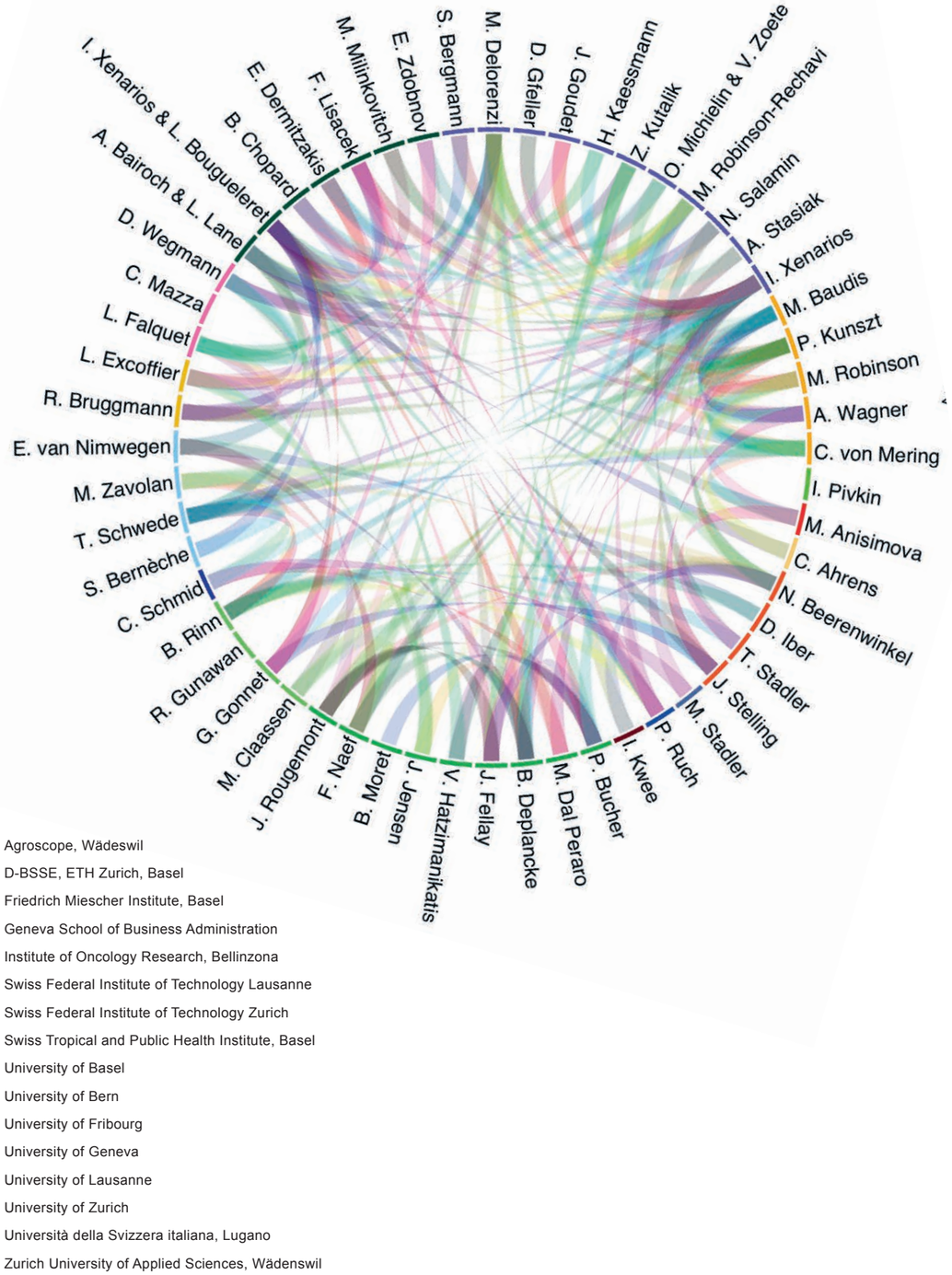
As of today, SIB counts 56 research and service groups and over 650 scientists from Swiss universities and research institutes located in the cantons of Basel, Bern, Geneva, Fribourg, Ticino, Vaud and Zurich.

New SIB groups
From five groups at its inception, SIB has been growing exponentially ever since. Since 1 January 2014, 11 new groups joined the Institute:

- Maria Anisimova, Zurich University of Applied Sciences, Wädenswil
- Christian Ahrens, Agroscope, Wädenswil
- Manfred Claassen, ETH Zurich
- David Gfeller, University of Lausanne
- Rudiyanto Gunawan, ETH Zurich
- Peter Kunszt, University of Zurich
- Ivo Kwee, Institute of Oncology Research, Bellinzona
- Michel Milinkovitch, University of Geneva
- Christoph Schmid, Swiss Tropical and Public Health Institute, Basel
- Tanja Stadler, D-BSSE, ETH Zurich, Basel
- Andrzej Stasiak, University of Lausanne

One of SIB's missions is to lead and coordinate the field of bioinformatics in Switzerland by fostering collaborations. Over time, an intense network of collaborative links has been established between the 56 SIB groups (**see collaboration network on the facing page**).

SIB also collaborates at the international level with renowned institutions:
- in Europe: e.g. the European Molecular Biology Laboratory-European Bioinformatics Institute (EMBL-EBI, UK), the Bioinformatics Services to Swedish Life Science (BILS), the Spanish National Bioinformatics Institute (INB), the Netherlands Bioinformatics Centre (NBIC)
- in the US: e.g. the National Institutes of Health (NIH), the National Center for Biotechnology Information (NCBI), the Protein Information Resource (PIR)
- elsewhere: e.g. SOKA University (Japan), Macquarie University (Australia), the University of Cape Town (South Africa), the Weizmann Institute of Science (Israel)



The SIB collaboration network was generated from a programme developed by Michael Baudis, SIB Group Leader: http://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 SIB groups are therefore involved in numerous domains.

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

	Genes and genomes pp. 40-46	Proteins and proteomes pp. 47-50	Medicine and health pp. 51-55	Evolution and phylogeny pp. 56-62	Structural biology pp. 63-65	Systems biology pp. 66-72	Bioinformatics infrastructure pp. 73-75
Sven Bergmann	■		□				□
Rémy Bruggmann	■	□	□			□	□
Philipp Bucher	■		□				□
Bart Deplancke	■		□	□		□	□
Emmanouil Dermitzakis	■		□			□	
Laurent Falquet	■	□	□				□
Zoltán Kutalik	■		□	□			□
Erik van Nimwegen	■			□		□	□
Mark D. Robinson	■		□				
Michael Stadler	■		□			□	□
Andrzej Stasiak	■	□	□		□		
Evgeny Zdobnov	■	□	□	□			□
Christian Ahrens	□	■				□	□
Amos Bairoch & Lydie Lane		■	□				□
Frédérique Lisacek		■	□			□	□
Christian von Mering	□	■		□		□	□
Ioannis Xenarios & Lydie Bougueleret	□	■	□	□	□	□	□
Michael Baudis	□		■			□	□
Niko Beerenwinkel	□		■	□		□	□
Mauro Delorenzi	□		■			□	□
Jacques Fellay	□		■			□	□
David Gfeller	□	□	■		□		
Ivo Kwee	□		■			□	
Patrick Ruch		□	■				□
Christoph Schmid	□		■				
Maria Anisimova	□	□		■			□
Laurent Excoffier	□			■			□
Gaston Gonnet	□	□		■			□
Jérôme Goudet	□			■			□
Jeffrey D. Jensen	□			■			□
Henrik Kaessmann	□			■			
Bernard Moret	□			■			
Marc Robinson-Rechavi	□			■			□
Nicolas Salamin	□			■			□
Tanja Stadler	□		□	■			□
Andreas Wagner	□			■			
Daniel Wegmann	□		□	■			□
Simon Bernèche		□			■		
Matteo Dal Peraro		□	□		■		□
Olivier Michielin & Vincent Zoete		□	□		■	□	□
Torsten Schwede		□			■		□
Bastien Chopard			□			■	□
Manfred Claassen			□			■	
Rudiyanto Gunawan	□		□			■	□
Vassily Hatzimanikatis	□	□	□			■	□
Dagmar Iber			□			■	□
Christian Mazza	□					■	
Michel Milinkovitch	□			□		■	□
Félix Naef	□	□	□			■	
Igor V. Pivkin			□			■	
Jörg Stelling						■	□
Mihaela Zavolan	□	□	□			■	
Peter Kunszt	□	□	□	□	□	□	■
Bernd Rinn							■
Jacques Rougemont	□			□		□	■
Ioannis Xenarios	□	□	□	□	□	□	■

GENES AND GENOMES

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

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

Sven Bergmann
Computational Biology Group
University of Lausanne

What do they do?
Sven Bergmann and his group develop and apply algorithms for the analysis of large sets of biomedical data. The focus is on the integration of information on DNA variability ("genotypes") and on molecular and organismal traits ("phenotypes").

2014 highlights
During 2014, the group made very good progress in developing LAS-VEGAS, a powerful tool for computing gene and pathway scores from single nucleotide polymorphisms (SNP)-phenotype association summary statistics. For gene score calculation, they use analytical and numerical solutions to derive various aggregate test-statistics such as the sum and the maximum of chi-square statistics at high speed and precision. For pathway scoring, their method uses a modified Fisher method, which offers not only significant power improvement over more traditional enrichment strategies but also eliminates the problem of arbitrary threshold selection inherent in any binary membership-based pathway enrichment approach. The group demonstrates the increase in power by analysing summary statistics from a large number of different meta-analyses.

Sven Bergmann and his team also developed MetaboMatching, which is a simple method that uses nuclear magnetic resonance data from a cohort to link genetic variants to chemical compounds – if their concentration has a genetic component.

Selected publications:
Rueedi R, *et al.* Genome-wide association study of metabolic traits reveals novel gene-metabolite-disease links. PLoS Genet 2014;10:e1004132.
Hohm T, *et al.* Plasma membrane H+-ATPase regulation is required for auxin gradient formation preceding phototropic growth. Mol Syst Biol 2014;10:751.
Hersch M, *et al.* Light intensity modulates the regulatory network of the shade avoidance response in Arabidopsis. Proc Natl Acad Sci USA 2014; 111:6515-20.

Rémy Bruggmann
Interfaculty Bioinformatics Unit – IBU
University of Bern

What do they do?
IBU provides bioinformatics services, expertise and computational resources to assist researchers in "big data" analyses and project planning for large-scale experiments (see also p. 22). Rémy Bruggmann's group also has its own research programme and develops methods to analyse high-throughput data.

2014 highlights
Resource: SetRank
Vital-IT and Rémy Bruggmann's team implemented a web interface to provide access to the gene set enrichment analysis tool SetRank. The SetRank algorithm was developed at IBU and is useful for prioritizing the results of expression studies. Its key principle is to discard gene sets that were initially considered significant, where their significance was solely based on parts of genes also involved in other processes.

The web implementation allows the visualization of pathways in a network, as well as browsing and further prioritizing the results. SetRank is in principle not limited to model organisms but can be applied to any organism for which pathway annotation data exists.

Research: novel astrovirus
One of the research highlights in 2014 was the discovery of a novel virus probably associated with encephalitis in cattle. Viral encephalitis is diagnosed in approximately 10-15% of cattle in Europe with neurological clinical signs. A high proportion of viral encephalitis in cattle is of unknown etiology. Attempts to culture and to identify the infectious agents associated with these conditions have yielded largely inconclusive results. To overcome these limitations, Rémy Bruggmann's group together with Torsten Seuberlich's team successfully performed a viral metagenome study using next-generation sequencing and identified a novel astrovirus [BoAstV]-CH13 in the brain of a cow with encephalitis. In a retrospective study, Seuberlich's group found this virus in five out of 22 animals with encephalitis of unknown etiology.

Selected publications
Bouzalas IG, *et al.* Neurotropic astrovirus in European cattle with non-suppurative encephalitis. J Clin Microbiol 2014;52:3318-24.
Hilty M, *et al.* Global phylogenomic analysis of nonencapsulated Streptococcus pneumoniae reveals a deep-branching classic lineage that is distinct from multiple sporadic lineages. Genome Biol Evol 2014;6:3281-94.
Greminger MP, *et al.* Generation of SNP datasets for orangutan population genomics using improved reduced-representation sequencing and direct comparisons of SNP calling algorithms. BMC Genomics 2014;15:16.

Philipp Bucher
Computational Cancer Genomics Group
EPFL, Lausanne

What do they do?
Philipp Bucher and his team are set on cracking the regulatory code of the human genome by analysing DNA sequences and experimental data which reveal the function of these sequences. The current focus is on developing methods for exploiting the so-called next-generation sequencing (NGS) data to this end.

2014 highlights
The group develops EPD, an annotated database of experimentally defined eukaryotic promoters (see also p. 17). Thanks to large amounts of newly available transcript mapping data, the team was able to more than double the promoter coverage for mouse – reaching 21,239 promoters from a previous 9,773. With this new increase, EPD has reached over 90% gene coverage for the two most important organisms: human and mouse. The team also introduced a comprehensive promoter collection for a new model organism, the worm *C. elegans*.

EPD is generated from data stored in Mass Genome Annotation (MGA), a publicly accessible directory that contains various kinds of NGS datasets. Reflecting the rapid accumulation of NGS in public repositories, the contents of the MGA repository almost doubled during 2014 to exceed 8,000 samples. Among the newly added samples are ChIP-seq data from the Roadmap Epigenomics Project.

The web servers for programmatic access to eukaryotic promoters have been greatly improved and made more user-friendly. Relevant programs from the ChIP-Seq analysis and Signal Search Analysis servers can now be accessed via direct links from the EPD homepage. A new tool named PWMscan has been added to the EPD promoter analysis suite, allowing for rapid whole genome scans for promoter elements defined by a position weight matrix or consensus sequence.

Philipp Bucher and his group also develop methods for the analysis of ChIP-Seq and related data, as well as methods for automatic classification and diagnosis of molecularly characterized clinical samples.

Selected publications

Nair NU, *et al.* Study of cell differentiation by phylogenetic analysis using histone modification data. BMC Bioinformatics 2014;15:269.
Polychronopoulos D, *et al.* Classification of selectively constrained DNA elements using feature vectors and rule-based classifiers. Genomics 2014;104:79-86.
Nair NU, *et al.* Probabilistic partitioning methods to find significant patterns in ChIP-Seq data. Bioinformatics 2014;30:2406-13.

Bart Deplancke
Laboratory of Systems Biology and Genetics
EPFL, Lausanne

What do they do?
Bart Deplancke's laboratory uses high-throughput sequencing, microfluidics, large-scale yeast screens and computational approaches to characterize the regulatory code in *Drosophila melanogaster* and mammals, as well as to examine how variations in this code affect molecular and organismal diversity.

2014 highlights
Bart Deplancke and his team participated in a consortium-driven effort to comprehensively characterize naturally occurring genetic variation in *D. melanogaster* Genetic Reference Panel (DGRP), consisting of 205 sequenced inbred lines. This resource will be of great value for analysing the genetic and molecular basis of quantitative traits with relevance to human biology. In this consortium, the group employed an integrated genotyping strategy to identify almost 5 million single nucleotide polymorphisms (SNPs) and 1.3 million non-SNPs at high resolution. This level of naturally occurring genetic variation is about 10-fold larger than that found in humans, thus constituting a powerful resource to molecularly dissect quantitative trait loci down to the nucleotide level.

Interestingly, the group found that their own local assembly-based variant detection tool PrInSeS performed particularly well when compared to other variant mapping approaches – such as GATK and the ensemble of tools used for the 1000 (human) genomes project – in that it contributed both the largest number of uniquely called variants as well as validated variants. This probably reflects the fact that PrInSeS was specifically developed for the analysis of *Drosophila* genomes which, as indicated, are highly polymorphic. This in turn may lower the efficacy of tools that were developed for human genome analyses. The group is now testing the efficacy of PrInSeS in detecting variants in both human exome and whole genome data.

Selected publications

Gubelmann C, *et al.* Identification of ZEB1 as a central component of the adipogenic gene regulatory network. eLife 2014;3:e03346.
Huang W, *et al.* Natural variation in genome architecture among 205 Drosophila melanogaster genetic reference panel lines. Genome Res 2014;24:1193-208.
Waszak SM, *et al.* Identification and removal of low-complexity sites in allele-specific analysis of ChIP-seq data. Bioinformatics 2014;30:165-71.

Emmanouil Dermitzakis
Genomics of Complex Traits Group
University of Geneva

What do they do?
The team led by Emmanouil Dermitzakis is interested in the genomics of complex traits, and is using various methodologies to understand the role of genetic variation in phenotypic variation. The group's focus is on the genetics and genomics of cellular phenotypes, genomic medicine and the development of methods to analyse the genetics of molecular phenotypes.

2014 highlights
In 2014, the group developed a number of algorithms to improve the discovery of regulatory variants in the human genome and link those involved in disease risk.

In addition, methodologies and pipelines developed in previous years were used to address, among other issues, two fundamental problems:

1) The contribution of non-coding regulatory variants in cancer progression. The team combined information from the genome and gene expression levels in tumours of colorectal cancer to implicate changes in gene regulation of 71 genes in cancer progression, and is extending this analysis to additional samples and cancer types.

2) The group quantified the contribution of interaction among genetic variants (GxG), and between genetic variants and the environment (GxE) in the variability of gene expression. This was a stepping stone to understand the contribution of GxG and GxE in disease.

Dermitzakis' group participated in a number of other projects, on themes such as the analysis of gene expression quantitative trait loci QTLs (eQTLs) in multiple tissues, the genetic and epigenetic contribution to variability in gene regulation, the contribution of regulatory variation to colorectal cancer risk, and the analysis of the pancreatic beta cell transcriptome.

Selected publications

Buil A, *et al.* Gene-gene and gene-environment interactions detected by transcriptome sequence analysis in twins. Nat Genet 2015;47:88-91.
Ongen H, *et al.* Putative cis-regulatory drivers in colorectal cancer. Nature 2014;512:87-90.
Bryois J, *et al.* Cis and trans effects of human variants on gene expression. PLoS Genet 2014;e1004461.

Laurent Falquet
Bioinformatics Unravelling Group – BUGFri
University of Fribourg

What do they do?
BUGFri supports life science researchers by providing expertise in data analysis of next-generation sequencing experiments, or any large-scale biological experiment that requires bioinformatics resources (see also p. 22).

The group focuses on genome assembly, annotation and comparison as well as on mutant and structure variant identification by resequencing. It also concentrates on RNA sequencing data analysis, proteome clustering and ortholog/paralog classification, and gene set enrichment analysis.

2014 highlights
During the course of 2014, BUGFri expanded its activities and collaborations with lab researchers of the University. A first article was published reporting the discovery of genes involved in the mycorrhization process – a symbiotic collaboration between plants and fungi. This work was carried out completely by *in silico* mining of existing databases, and applying comparative proteomics among dozens of plant species.

The first edition of the BEFRI Genomics Day was organized in June – an event the group plans to organize every other year, in collaboration with the bioinformatics core facility of Bern.

Selected publications

Favre P, *et al.* A novel bioinformatics pipeline for gene discovery based on conservation of the protein coding sequence. BMC Plant Biol 2014;14:333.
Cannarozzi G, *et al.* Genome and transcriptome sequencing identifies breeding targets in the orphan crop tef (Eragrostis tef). BMC Genomics 2014;15:581.
Miyazaki R, *et al.* Comparative genome analysis of Pseudomonas knackmussii B13, the first bacterium known to degrade chloroaromatic compounds. Environ Microbiol 2014;17:91-104.

Zoltán Kutalik
Statistical Genetics Group
University of Lausanne

What do they do?
Zoltán Kutalik and his team develop statistical methods to gain a greater understanding in the genetic mechanisms which influence complex human traits, such as obesity, or which trigger diseases, such as infectious or autoimmune diseases. The group is particularly interested to learn how the human environment modifies the impact of these genetic factors.

2014 highlights
Genome-wide association studies (GWAS) measure the correlation between genetic and phenotypic variation in large groups of individuals. Almost all previous studies assumed that the effect of all genetic variants is the same, regardless of whether they are inherited from the mother or the father. Kutalik's team discovered that it is possible to detect genetic effects that are different according to their parental origin, even in samples of unrelated individuals where the parental origin is unknown. This is because, at the genetic markers producing these effects, an increase in the variability of body mass index (BMI) is observed among people who have inherited a different genetic code from their mother and father compared to those who have inherited the same genetic code from both parents.

The group applied this method to discover genetic markers with parent-of-origin effects (POE) on overweight. This resulted in six candidate markers showing strong POE association. In family-based studies (where the parental origin of the variants can be inferred), two of the candidates showed a significant association. Surprisingly, they found that the same genetic code at these markers may increase BMI when inherited from one parent, but decrease when inherited from the other.

Kutalik's group extended the QuickTest software developed for GWA scans to identify loci with POE effect based on the newly developed method. Thanks to this software, many collaborators applied it to their data and the combined results yielded a major discovery.

Selected publications

Rueedi R, *et al.* Genome-wide association study of metabolic traits reveals novel gene-metabolite-disease links. PLoS Genet 2014;10:e1004132.
Hoggart CJ, *et al.* Novel approach identifies SNPs in SLC2A10 and KCNK9 with evidence for parent-of-origin effect on body mass index. PLoS Genet 2014;10:e1004508.
Rüeger S, *et al.* Impact of common risk factors of fibrosis progression in chronic hepatitis C. Gut; Epub ahead of print.

Erik van Nimwegen
Genome Systems Biology Group
University of Basel

What do they do?
Erik van Nimwegen's group uses both theoretical and experimental approaches to study the structure, function and evolution of the regulatory networks that cells use to control the expression of their genes. Another key interest of the group is the identification of general quantitative laws in genome evolution.

2014 highlights
SwissRegulon is a database of genome-wide annotations of regulatory sites (see also p. 16). Besides genome-wide annotations of regulatory sites, the group's ISMARA web server allows any researcher to reconstruct genome-wide regulatory networks from their own microarray, RNA-sequencing, or ChIP-seq data in a completely automated fashion.

In 2014, there were three key developments and highlights related to SwissRegulon:

1) The most important current resource within SwissRegulon is the ISMARA web server, and the main article describing the methodologies underlying this resource was published in Genome Research.

2) With the continuing dramatic drop in costs of next-generation sequencing, many groups are now routinely sequencing large numbers of genomes of related microbial strains, but no easily useable methods for reconstructing phylogenies from such data are currently available. Erik van Nimwegen's group recently developed a completely automated procedure called REALPHY for reconstructing phylogenies directly from raw sequencing data, and has made it available as a web server within SwissRegulon.

3) Over the last few years, the group has been developing a large collection of methodologies for the analysis of genome-wide binding data (ChIP-seq), including various methods for inferring regulatory motifs and predicting regulatory sites. They have now integrated all these procedures into a comprehensive pipeline called CRUNCH, also available as a webserver, which takes raw ChIP-seq data as input, and performs all analysis steps from quality control, read mapping, peak calling, up to comprehensive regulatory motif analysis in an automated way.

Selected publications

Balwierz PJ, *et al.* ISMARA: automated modeling of genomic signals as a democracy of regulatory motifs. Genome Res 2014;24:869-84.
Bertels F, *et al.* Automated reconstruction of whole-genome phylogenies from short-sequence reads. Mol Biol Evol 2014;31:1077-88.
FANTOM5 Consortium, *et al.* A promoter-level mammalian expression atlas. Nature 2014;507:462-70.

Mark D. Robinson
Statistical Bioinformatics Group
University of Zurich

What do they do?
Mark Robinson and his team develop robust data analysis solutions for the analysis of genome-scale data. They develop statistical methods for interpreting data from high-throughput sequencing and other technologies, primarily in the domain of gene expression and epigenetic regulation.

The group is largely data- and problem-driven, and the methods they develop are catered to the characteristics of the technology platform that is generating the data. They develop open-source software tools via the Bioconductor project.

Where needed, the team also designs experiments and collects data to compare the performance of competing methods and platforms.

Selected publications

Zhou X, *et al.* Robustly detecting differential expression in RNA sequencing data using observation weights. Nucleic Acids Res 2014;42:e91.
Riebler A, *et al.* BayMeth: improved DNA methylation quantification for affinity capture sequencing data using a flexible Bayesian approach. Genome Biol 2014;15:R35.
Robinson MD, *et al.* Statistical methods for detecting differentially methylated loci and regions. Front Genet 2014;5:324.

Michael Stadler
FMI Computational Biology Group
Friedrich Miescher Institute, Basel

What do they do?
Michael Stadler and his team study gene regulation through the analysis and modelling of genome-wide datasets. Most analysed data are generated by high-throughput sequencing. The group collaborates closely with experimental researchers on experimental models including cancer progression and cellular differentiation. Most of the projects measure various aspects of gene expression, including DNA methylation, single cell transcription, protein-binding to DNA, and translation.

2014 highlights
The group has been developing QuasR, an R/Bioconductor package with a collection of high-throughput sequencing analysis tools. It is fully documented and very easy to install, although it contains powerful algorithms and is designed to support both beginners and experts. www.bioconductor.org

The QuasR package was published in the Bioinformatics journal, and has been widely used to analyse in-house and external data in over a dozen projects. Since its publication, the package has been downloaded over 4,000 times from the main Bioconductor server. QuasR supports various types of high-throughput sequencing experiments including RNA sequencing (spliced alignment and de novo splice site detection), bisulphite sequencing to measure DNA methylation and allele-specific analysis. With QuasR and R/Bioconductor, it is possible to perform a complex analysis in the form of a single R script, thus simplifying exchange between collaborators and improving reproducibility, even on different computer platforms (Windows, OS X and Unix/Linux).

Selected publications

Gaidatzis D, *et al.* QuasR: quantification and annotation of short reads in R. Bioinformatics; Epub ahead of print.
Tocchini C, *et al.* The TRIM-NHL protein LIN-41 controls the onset of developmental plasticity in Caenorhabditis elegans. PLoS Genet 2014;10:e1004533.
Gaidatzis D, *et al.* DNA sequence explains seemingly disordered methylation levels in partially methylated domains of Mammalian genomes. PLoS Genet 2014;10:e1004143.

Andrzej Stasiak
DNA and Chromosome Modelling Group
University of Lausanne

What do they do?
Andrzej Stasiak's team uses numerical simulations to shed light on the organization of chromosomes in yeasts and higher eukaryotes such as humans. The group is especially interested in understanding chromosome structure and organization during interphase. Interphase is the longest stage of the cell cycle, and the stage during which genes are expressed.

2014 highlights
Thanks to the progress of high-resolution 3C (Chromosome Conformation Capture) methods, in 2012 scientists discovered that interphase chromosomes of higher eukaryotes are composed of sequential blocks with a high frequency of internal contacts. The average size of these blocks – also known as topological domains – are about 1 Mb. It is not yet known how chromatin fibres are arranged in these blocks nor is the mechanism responsible for their formation known. Inspired by the fact that bacterial chromosomes are composed of supercoiled topological domains, Stasiak and his team performed Brownian dynamics simulations of supercoiled chromatin fibres that are found between borders of individual topological domains. Their simulations revealed that transcription-induced supercoiling explains the formation and all known properties of topological domains.

Another important development consisted in modifying the chromatin fibre model the group uses, by introducing a torsional resistance into it. Thanks to this modification, Stasiak's group was able to simulate the effect of supercoiling on chromosome structure. Supercoiling arises during transcription and has important consequences on overall chromosome organization.

In a collaborative project involving researchers in Poland and the USA, the group worked on the database of proteins with polypeptide chains that form knots and slipknots.

Selected publications:

Benedetti F, *et al.* Models that include supercoiling of topological domains reproduce several known features of interphase chromosomes. Nucleic Acids Res 2014;42:2848-55.

Benedetti F, *et al.* Effects of supercoiling on enhancer-promoter contacts. Nucleic Acids Res 2014;42:10425-32.

Jamroz M, *et al.* KnotProt: a database of proteins with knots and slipknots. Nucleic Acids Res 2015;43:D306-14.

Evgeny Zdobnov
Computational Evolutionary Genomics Group
University of Geneva

What do they do?
Evgeny Zdobnov and his group use comparative genomics techniques to interpret genome sequencing data and learn about how genes and genomes evolve, as well as to investigate the possible functions of protein-coding genes, microRNA genes and conserved non-coding sequences.

2014 highlights
The group continues to develop OrthoDB resource – a catalogue of "equivalent" genes among species, called orthologs (see also p. 17). Each updated release of OrthoDB includes more species with available sequenced genomes. The current release holds 61 vertebrates, 87 arthropods, 25 basal metazoans, 227 fungi and 2,627 bacteria. The website has been recently redesigned to display the information more clearly to users. Moreover, the underlying ortholog clustering software is now freely available to researchers who need to find orthologs in custom data sets. By identifying equivalent genes from many different organisms, OrthoDB helps biologists to find information on the evolutionary history and possible functions of a specific gene.

The team also collaborates actively with clinical colleagues to bring the sequencing revolution to medical practice, in particular in the field of viral metagenomics. The group is also contributing to the i5K initiative to analyse the genomes of 5,000 insects. This will provide an exceptional opportunity to explore the evolutionary processes that act on genes and genomes, since insects are the most diverse and successful terrestrial animals, having by far the largest number of species.

Selected publications

Kriventseva EV, *et al.* OrthoDB v8: update of the hierarchical catalog of orthologs and the underlying free software. Nucleic Acids Res 2015;43:D250-6.

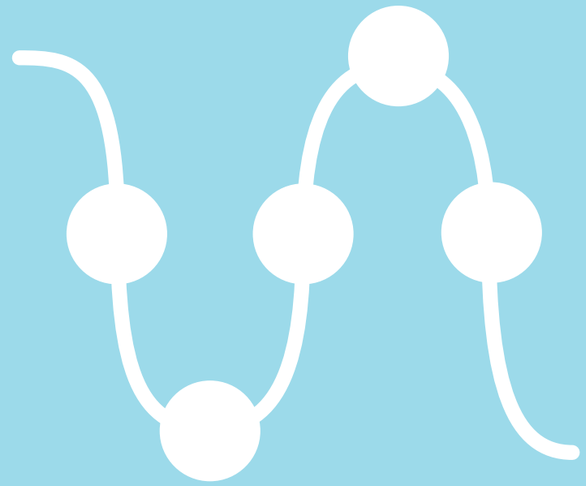
Neafsey DE, *et al.* Highly evolvable malaria vectors: the genomes of 16 Anopheles mosquitoes. Science 2015;347:43.

Petty TJ, *et al.* Comprehensive human virus screening using high-throughput sequencing with a user-friendly representation of bioinformatics analysis: a pilot study. J Clin Microbiol 2014;52:3351-61.

PROTEINS AND PROTEOMES

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

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



Christian Ahrens

Bioinformatics and Proteogenomics Group
Agroscope, Wädenswil

What do they do?
Christian Ahrens's group uses and develops bioinformatics tools for the integration and analysis of datasets provided by experimental biologists (genome sequences, gene and protein expression data, metabolomics data). One focus is to exploit the unique advantages of protein expression data; another is on proteogenomics, i.e. the identification of all proteins encoded in a genome.

The group contributed to the development of Protter, a software tool that allows users to visualize the topology of membrane proteins and to integrate annotations and experimental evidence in the form of publication-ready plots.

Protter addresses an important and previously unmet need of the research community, namely to visualize the predicted topology of membrane proteins in the context of additional information on a protein's function, while concomitantly highlighting parts of the protein that have been identified in proteomics experiments. http://wlab.ethz.ch/protter/start/

2014 highlights
Ulrich Omasits, a PhD student co-supervised by Christian Ahrens and Prof. Bernd Wollscheid (ETH Zurich), finalized the development of the resource PeptideRank, which allows users to select the best-suited peptides to quantitatively measure protein amounts in diverse organisms. This particular resource performs better than other currently available software solutions. http://wlab.ethz.ch/peptiderank/

Selected publications

Qeli E, *et al.* Improved prediction of peptide detectability for targeted proteomics using a rank-based algorithm and organism-specific data. J Proteomics 2014;108:269-83.
Stekhoven DJ, *et al.* Proteome-wide identification of predominant subcellular protein localizations in a bacterial model organism. J Proteomics 2014;99:123-37.
Omasits U, *et al.* Protter: interactive protein feature visualization and integration with experimental proteomic data. Bioinformatics 2014;30:884-6.

Amos Bairoch & Lydie Lane

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

What do they do?
Amos Bairoch, Lydie Lane and their team are set on broadening our knowledge and understanding of the function of the 20,000 or so protein-coding genes that exist in the human genome. Their main focus is on building neXtProt, a knowledge resource on human proteins, annotating the effects of human protein variations in the context of cancers and genetic diseases, and analysing results of high-throughput experiments to shed light on aspects of protein function.

2014 highlights
neXtProt includes curated information on various aspects of human protein biology such as function, mRNA/protein expression, protein/protein interactions, post-translational modifications and protein variations (see also p. 16). neXtProt is the reference resource for the international HUPO c-HPP consortium, whose aim is to experimentally validate all the proteins predicted from the analysis of the human genome. CALIPHO is involved in defining the guidelines to curate and integrate all the data generated within the framework of these projects. www.nextprot.org

In 2014, the CALIPHO team focused their efforts on the development of a new powerful search functionality for neXtProt, based on a software technology called SPARQL. SPARQL allows users to make very complex queries that use the wealth of data encapsulated in neXtProt. The software also allows users to seamlessly merge information from neXtProt with data from other resources that also use SPARQL.

The team also pursued their laboratory activities targeted towards the functional characterization of human proteins assisted by bioinformatics data-mining activities. In 2014, they published two papers, one describing a protein involved in cilia formation, and another on stress granule formation.

Selected publications

Lane L, *et al.* Metrics for the human proteome project 2013-2014 and strategies for finding missing proteins. J Proteome Res 2014;13:15-20.
Bontems F, *et al.* C2orf62 and TTC17 are involved in actin organization and ciliogenesis in zebrafish and human. PLoS One 2014;9:e86476.
Salleron L, *et al.* DERA is the human deoxyribose phosphate aldolase and is involved in stress response. Biochim Biophys Acta 2014;1843:2913-25.

Frédérique Lisacek

Proteome Informatics Group – PIG
University of Geneva

What do they do?
The group led by Frédérique Lisacek is involved in software and database development for the proteomics and glycomics communities. Their focus is on mass spectrometry data analysis, the discovery of posttranslational modifications, and the functional study of glycoconjugates and carbohydrates.

2014 highlights
SugarBind database covers knowledge of interactions between pathogens and carbohydrate ligands of mammalian hosts (see also p. 17 and http://sugarbind.expasy.org). The new redesigned version of SugarBindDB was released in 2014.

The framework was chosen to match that of UniCarbKB, the universal glycan structure and glycoprotein database co-developed by an international consortium of which PIG is a major player. http://unicarbkb.org

Information in SugarBindDB was extended to cover properties of pathogens including the diseases they cause, as well as to describe in detail the tissue source and the constitutive parts of the glyconjugate carrying the ligand. Naming conventions of all entities involved were adopted. Standard representations of the glycan molecules were introduced. Cross-links were put in place to expand exploration of the functional role of cell-surface carbohydrates. Internal links support exploration of the database content. External links with bioinformatics resources provide background information on sugar-binding proteins. Graphic tools were also added to track the specificity of a particular ligand across a variety of strains.

Furthermore, PIG develops software for mass spectrometry analysis, which is structured and stored in a library renamed MzJava. http://mzjava.expasy.org

The group also develops software specifically for analysing glycan structural data and, in 2014, released GlycoDigest. http://glycodigest.org

Selected publications

Mariethoz J, *et al.* SugarBindDB, a resource of pathogen lectin-glycan interactions. In Glycoscience: Biology and Medicine 2014. Springer, Japan.
Gotz L, *et al.* GlycoDigest: a tool for the targeted use of exoglycosidase digestions in glycan structure determination. Bioinformatics 2014;30:3131-3.
Bilbao A, *et al.* Processing strategies and software solutions for proteomics studies using mass spectrometry data-independent acquisition. Proteomics; Epub ahead of print.

Christian von Mering

Bioinformatics / Systems Biology Group
University of Zurich

What do they do?
Christian von Mering and his group are interested in biological networks and the quantitative, high-throughput data that define them. As networks evolve, their impact on the molecular players within the cell is studied at the levels of protein expression, sequence evolution and genome architecture.

2014 highlights
The group maintains STRING, a popular database of protein-protein interactions (see also p. 16). In 2014, STRING was updated to version 10, a major new release that almost doubled the number of organisms. STRING now covers 2,031 organisms from all three domains of life, including many newly sequenced genomes.

This new version includes an important change with respect to the "transfer" of interaction knowledge from one organism to another. Interaction transfer is necessary because many molecular studies are conducted in model organisms, which can be applied to human biology. This transfer can be done by applying interactions from model organisms to humans and vice versa, assuming that the proteins in question serve the same overall functions. This assumption holds best for so-called "orthologs", i.e. pairs of proteins that have followed a similar evolutionary trajectory as they split from the last common ancestor of both organisms. STRING now executes orthology-informed interaction transfer systematically, in an all-against-all fashion.

Other changes include improvements to the user interface. For example, users can upload their own interaction and protein data and view them privately in the context of all integrated protein network data in STRING.

The group continues the development of its PaxDB database dedicated to systematic protein abundance measurements in model organisms. It also initiated a number of collaborations focusing on microbial communities relevant in diseases, such as cystic fibrosis, salmonella infection and inflammatory bowel disease.

Selected publications

Szklarczyk D, *et al.* STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res 2014;43:D447-52.
Franceschini A, *et al.* Specific inhibition of diverse pathogens in human cells by synthetic microRNA-like oligonucleotides inferred from RNAi screens. Proc Natl Acad Sci USA 2014;111:4548-53.
Schmidt TS, *et al.* Ecological consistency of SSU rRNA-based operational taxonomic units at a global scale. PLoS Comput Biol 2014;10:e1003594.

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

What do they do?
The Swiss-Prot group produces and maintains a number of internationally renowned databases for the life science community, namely UniProtKB/Swiss-Prot, HAMAP, PROSITE, Rhea, UniPathway and ENZYME. All these databases are manually curated and continuously updated.

2014 highlights
One of the group's main activities is the production and development of the UniProtKB/Swiss-Prot knowledgebase. UniProtKB/Swiss-Prot provides comprehensive information on protein sequences and their function sourced from the scientific literature by expert curators. It is an internationally renowned resource with over 900,000 requests per month from all over the world (see also p. 16).

In 2014, the group released a completely redesigned version of the UniProt website – the culmination of an extensive user-centred design process carried out over two years. The new UniProt website features improved search functionalities and more intuitive navigation, with individual protein pages restructured to make links between related knowledge and data more obvious. The provenance of information is rendered transparent through improved "evidence tags" that link each annotation to its source – such as a scientific publication or homologous protein.

The first public version of the SwissLipids knowledgebase for lipid chemistry and biology was released. SwissLipids features a library of over 240,000 possible structures for common lipid classes with extensive curated links to proteins (UniProtKB) and metabolism (Rhea).

Selected publications
Morgat A, *et al.* Updates in Rhea-a manually curated resource of biochemical reactions. Nucleic Acids Res 2015;43:D459-64.
Pedruzzi I, *et al.* HAMAP in 2015: updates to the protein family classification and annotation system. Nucleic Acids Res 2015;43:D1064-70.
Poux S, *et al.* Expert curation in UniProtKB: a case study on dealing with conflicting and erroneous data. Database (Oxford) 2014:bau016.

MEDICINE AND HEALTH

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

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

Michael Baudis
Computational Oncogenomics Group
University of Zurich

What do they do?
The group focuses on the exploration of changes in the genomes of cancer cells. They use their continuously expanding data collections to detect both specific mutation events that affect individual tumours, as well as complex genomic aberration patterns and their relationships to cancer entities. Collaborative projects – e.g. in association with specialists in the fields of childhood brain tumours or skin diseases – are aimed at the biological characterization of these diseases, to facilitate targeted therapeutic approaches.

2014 highlights
The arrayMap database provides over 60,000 pre-processed copy number profiles from cancer genome studies to biomedical researchers. This resource, which is the largest of its kind, facilitates the identification of cancer type specific mutation patterns, and allows for a simplified association of potential target genes with affected tumour entities.

In 2014, Michael Baudis' team launched a new edition of the arrayMap resource, which almost doubled its data content and introduced a range of new features, including an HTTP-based data API. This allows users to access directly the group's pre-processed cancer genome data and to use them in their own downstream applications.

In 2014 the group also published an extensive study analysing "chromothripsis"-like genome patterns in cancer, based on datasets collected in arraymap.org. As part of the results, the study showed that the genomic aberrations, commonly referred to as "chromothripsis", probably constitute a group of phenomena with variable pathophysiologies and not all refer to the core definition of synchronous genome shattering events.

The group continues to extend the Progenetix database (progenetix.org) and the DIPG repository for childhood glioma data (dipg.progenetix.org), and participates in a collaborative project to delineate inflammatory dermatologic diseases from cutaneous lymphomas.

Selected publications

Cai H, *et al.* Chromothripsis-like patterns are recurring but heterogeneously distributed features in a survey of 22,347 cancer genome screens. BMC Genomics 2014;15:82.
Cai H, *et al.* Progenetix: 12 years of oncogenomic data curation. Nucleic Acids Res 2014;42:D1055-62.
Baderca F, *et al.* Biopsying parapsoriasis: quo vadis? Are morphological stains enough or are ancillary tests needed? Rom J Morphol Embryol 2014;55:1085-92.

Niko Beerenwinkel
Computational Biology Group
D-BSSE, ETH Zurich, Basel

What do they do?
Niko Beerenwinkel and his team develop and apply statistical methods and algorithms to analyse high-throughput molecular data. Their goal is to predict the effect of genetic alterations and to support the diagnosis and treatment of diseases.

The group has four main focuses:
1) computational biology and bioinformatics;
2) medical systems biology and computational medicine;
3) statistical methods for high-throughput molecular profiling data;
4) evolutionary dynamics of cancer and pathogen populations.

2014 highlights
HaploClique is an efficient haplotype multi-assembly software for the analysis of within-host heterogeneous virus populations. It is essential for a reliable detection of all viral variants present within individual patients, and hence for improved diagnosis and prediction of response to antiviral therapy.

HaploClique has been optimized for both short-read and long-read sequencing technologies. It can handle paired-end data and is particularly useful for detecting long-range deletions in viral haplotype sequences.

Selected publications

Giallonardo FD, *et al.* Full-length haplotype reconstruction to infer the structure of heterogeneous virus populations. Nucleic Acids Res 2014;42:e115.
Töpfer A, *et al.* Viral quasispecies assembly via maximal clique enumeration. PLoS Comput Biol 2014;10:e1003515.
Szczurek E, Beerenwinkel N. Modeling mutual exclusivity of cancer mutations. PLoS Comput Biol 2014;10:e1003503.

Mauro Delorenzi
Bioinformatics Core Facility – BCF
University of Lausanne

What do they do?
Mauro Delorenzi and his group provide bioinformatics-statistical expertise to the life science community and promote collaborations (see also p. 21). They perform the analysis of biomedical-genomics data with a focus on biomarker studies in cancer research and statistical methods for genomics. Recently, the group concentrated on molecular heterogeneity and pathway activation patterns in cancer subtypes.

2014 highlights
The importance of competences in data analysis continues to increase. Many researchers obtain results they consider as "big data", which prove to be very challenging to analyse. Statisticians have been working with such data for a long time and Delorenzi and his team are well positioned to tackle the challenges.

In order to meet these challenges, the group developed their offer in two directions in 2014:

1) The group gave courses on state-of-the-art methods for statistical analysis of biological data. In particular, they organized several advanced courses on topics for which training has rarely been offered to the Swiss life science community.

2) The group continues to provide consulting services to the life science community, enabling researchers to gain access to cutting-edge statistical analysis methods in their work. Today, many research groups, both in the academic world and in industry, have projects that require statistical help. However, they are often unable to hire dedicated data analysts with the right competences, and rely on the Delorenzi group's services to analyse data.

The Biostatistics Service of BCF provides consulting services, focusing on the analysis of data produced by modern high-throughput biological methods. The expertise they provided led to the publication of several scientific articles over the course of the year, in particular in the field of cancer biology.

Selected publications

Di Narzo AF, *et al.* Test of four colon cancer risk-scores in FFPE microarray gene expression data. J Natl Cancer Inst 2014;106(10).
Ragusa S, *et al.* PROX1 promotes metabolic adaptation and fuels outgrowth of Wnt(high) metastatic colon cancer cells. Cell Rep 2014;8:1957-73.
Soneson C, *et al.* Batch effect confounding leads to strong bias in performance estimates obtained by cross-validation. PLoS One 2014;9:e100335.

Jacques Fellay
Host-Pathogen Genomics Group
EPFL, Lausanne

What do they do?
Research in the Fellay lab focuses on human genomics of infection. Using genome-wide association analysis, exome sequencing and transcriptomics, the group explores the genetic roots of inter-individual differences in response to pathogens. Host genomics of HIV infection, joint analyses of interactions between human and viral genomes, and exome sequencing in patients with extreme infectious disease phenotypes are some of the important research directions.

2014 highlights
The group is leading large international projects to understand how human genetic variation impacts HIV disease progression. In collaboration with over 20 cohorts or centres studying HIV progression, the group has collected genome-wide genotyping data on about 11,000 HIV-infected individuals with clinical follow-up.
This project (the International Collaboration for the Genomics of HIV – ICGH) has three main goals:
1) the identification of common genetic markers (>1% frequency) associated with HIV disease progression;
2) the identification of likely causal variants underlying associated regions through fine-mapping;
3) the quantification of the proportion of variation in HIV disease progression that is heritable.

The group also investigates the influence of rare gene polymorphisms. This project uses exome sequencing to identify mutations that alter protein sequences and test them for an impact on HIV viral load individually, in combination within a gene and across relevant gene sets. So far the group has obtained exome sequence data on about 1,000 HIV-infected patients.

The main goal of both projects is to improve the understanding of human-HIV interactions at the genomic level, which may assist in the development of new therapeutics and/or vaccines.

In addition, the team developed an online compendium of host genomic data in HIV biology and disease. This intuitive web interface allows queries, and supports validation of the rapidly growing body of host genomic information pertinent to HIV research. www.guavah.org

Selected publications

Regoes RR, *et al.* Disentangling human tolerance and resistance against HIV. PLoS Biol 2014;12:e1001951.
Rausell A, *et al.* Analysis of stop-gain and frameshift variants in human innate immunity genes. PLoS Comput Biol 2014;10:e1003757.
Bartha I, *et al.* GuavaH: A compendium of host genomic data in HIV biology and disease. Retrovirology 2014;11:6.

David Gfeller

Computational Cancer Biology
University of Lausanne

What do they do?
Recent developments in immunotherapy are revolutionizing cancer treatments. Using large genomics data, David Gfeller and his team hope to gain a better understanding of fundamental properties of tumours, such as how tumours are recognized by the immune system and how this recognition can be exploited towards clinical benefits. To reach this goal, the team develops computational tools from statistics, machine learning and modelling.

2014 highlights
Tumours are characterized by changes at the genetic and epigenetic level, which are often detrimental to patients. But these changes also provide a way of distinguishing cancer cells from normal cells. Among the proteins that are specifically expressed or mutated in cancer cells, only a given fraction is detected by the immune system. This is because certain proteins need to be displayed on the cell's membrane so as to become visible to the immune system – a process that is highly regulated in cells.

David Gfeller's group is interested in computational tools to predict which antigens have the greatest potential to elicit an immune reaction that – together with immunotherapy drugs – could help to redirect the immune system against cancer cells. The group is also working on optimizing tools based on genomics data to predict and characterize the presence of immune cells in tumour samples.

Proteins perform their function by interacting with each other. Experimentally mapping all possible interactions is challenging, and computational approaches are promising to narrow the list of interesting candidates. Therefore a second interest of the group is to develop protein-protein interaction prediction tools by combining statistical approaches with structural modelling. They are also interested in understanding better how cancer somatic mutations affect protein interactions.

Selected publications

Gfeller D, *et al.* SwissTargetPrediction: a web server for target prediction of bioactive small molecules. Nucleic Acids Res 2014;42:W32.
Gfeller D, *et al.* Prediction and experimental characterization of nsSNPs altering human PDZ-binding motifs. PLoS One 2014;9:e94507.

Ivo Kwee

Bioinformatics Core Unit
Institute of Oncology Research, Bellinzona

What do they do?
Ivo Kwee and his team are involved in cancer bioinformatics. Their unit supports the research groups at the Institute of Oncology Research with computational and statistical services for genomic profiling, sequencing analysis, functional genomics, pharmacogenomics, and clinical study support. They develop new computational methods and tools to complement and drive cancer research.

2014 highlights
Cancer is not caused by a single gene but through complex interactions of multiple genes. The team's efforts are to go beyond single gene statistics and pursue functional analysis on gene sets and interaction networks. The group is using optimization on Prize Collecting Steiner Trees (PCST) to find functional correlated subnetworks that may point to driving oncogenetic pathways. In a large interaction network, PCST attempts to find a neighbourhood – or subnetwork – where genetic aberrations are most concentrated. For example, a connected neighbourhood where many genes are differentially expressed can be found; or, combined with survival data, a subnetwork that locates interacting genes that are differentially expressed and correlated with patient survival can be found. The optimal solution is NP-hard. The group has developed a fast heuristic for PCST that can handle tens of thousands of nodes with hundreds of thousands of edges. Furthermore, the group aims to integrate different datatypes such as copy number, methylation and expression in a single network.

The group received a number of grants for bioinformatics projects in collaboration with the Artificial Intelligence institute IDSIA. As an example, they developed a Bayesian algorithm for copy number segmentation, and in a running project they are developing an optimization method for functional analysis of genetic networks using PCST.

Selected publications

Ahkmedov M, *et al.* A fast heuristic for the prize-collecting Steiner tree problem. In Lecture Notes in Management Science 2014;6:207-16.
Mian M, *et al.* Genome-wide DNA profiling identifies clonal heterogeneity in marginal zone lymphomas. Br J Haematol 2014:164:896-9.
Ronchetti D, *et al.* Distinct patterns of global promoter methylation in early stage chronic lymphocytic leukemia. Genes Chromosomes Cancer 2014;53:264-73.

Patrick Ruch

Text Mining Group
Geneva School of Business Administration (HEG)

What do they do?
Patrick Ruch and his team transform the wealth of implicit and hidden information buried in the biomedical literature, into human-actionable and machine-readable knowledge items (e.g. answers to questions, functional annotation).

The group maintains an updated data infrastructure to store, access and analyse the bibliome, i.e. biomedical literature contents such as MEDLINE, biomedical patent libraries and other web contents, including clinical practice guidelines and patient communities. The bibliome is then normalized and enriched with unambiguous semantic descriptors, i.e. database accession numbers and onto-terminological descriptors.

2014 highlights
The Gene Ontology Categorizer (GOCat), a tool used to assign gene ontology functional descriptors to any input text (abstract, full-text article…) received about 1 million visits in 2014! http://eagl.unige.ch/GOCat/

GOCat was further developed for Full-Texts (GOCat4FT).
http://eagl.unige.ch/GOCat4FT/

GOCat was customized to participate in the BioCreative competition, which it won. The system is now able to accept full-text papers. From this basic input, it generates a curation-driven summary, which is ultimately sent to GOCat to generate a ranked list of functional descriptors to support the annotator. The service is available for demonstration and is currently being integrated into the curation workflow of neXtProt within the context of the neXtPresso project.

For Switzerland, the group maintains the biomedical terminological resources of the EXPAND project. In particular, they support the cross-border exchange of electronic health records between Switzerland and European countries. http://www.expandproject.eu/

In collaboration with Dr Rodolphe Meyer, from the Geneva University Hospital (HUG), Ruch and his team began the TransCoD (Translational Coding Database) project, whose aim is to build a holistic clinical knowledge repository to support clinical research.

Selected publications

Gobeill J, *et al.* Closing the loop: from paper to protein annotation using supervised gene ontology classification. Database (Oxford) 2014;2014:bau088.
Vishnyakova D, *et al.* Electronic processing of informed consents in a global pharmaceutical company environment. Stud Health Technol Inform 2014;205:995-9.
Gobeill J, *et al.* Instance-based learning for tweet categorization in CLEF REPLAB 2014. Proc CLEF 2014;1491-9.

Christoph Schmid

Computational PathoGenOmics Group
Swiss TPH, Basel

What do they do?
The Computational PathoGenOmics group focuses on data analysis. Methods used in biomedical research increasingly include high-throughput assays to conduct millions of chemical, genetic and pharmacological tests on biological samples. The group leads corresponding projects and also supports collaborating research groups in the development and application of computational methods to address research questions in the fields of infection biology and public health.

2014 highlights
Chemical modifications in DNA have emerged as an additional superposed layer of information defining some of the properties of a living cell. The roles of such epigenetic modifications, including DNA methylation, are incompletely characterized. By applying a recently developed genome-wide sequencing assay, the group assessed the modifications in the genomes of *Neisseria meningitidis* from African epidemics. The applied PacBio sequencing method readily detected unexpected diversity in epigenetic modifications, and confirms the potential of epigenetic changes to contribute to rapidly adapting properties of bacterial cells.

The group is involved in several other projects to assess genome sequences of a variety of pathogens. Pathogenic organisms such as *Plasmodium*, the causative agent of malaria, have evolved specific mechanisms to evade the human defence systems. In collaboration with groups at the Swiss TPH, Schmid and his team have identified a specific portion in the DNA sequence that controls the production of proteins.

The group is also contributing to a European project assessing the potential effects of electromagnetic fields. Human cells were exposed in culture to electromagnetic fields and epigenetic modifications were assessed using high-throughput sequencing methods. Ongoing analysis approaches are comparing epigenetic profiles of exposed to unexposed control cells.

Selected publications

Brancucci NM, *et al.* A var gene upstream element controls protein synthesis at the level of translation initiation in Plasmodium falciparum. PLoS One 2014;9:e100183.
Begitt A, *et al.* STAT1-cooperative DNA binding distinguishes type 1 from type 2 interferon signaling. Nat Immunol 2014;15:168-76.
Seguín-Estévez Q, *et al.* Extensive remodeling of DC function by rapid maturation-induced transcriptional silencing. Nucleic Acids Res 2014;42:9641-55.

EVOLUTION AND PHYLOGENY

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

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

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

What do they do?
Maria Anisimova and her team develop computational methods to study genome evolution and adaptation. Keeping pace with growing data size and complexity, they provide efficient solutions for the analysis of phylogenetic patterns and selection in genomic sequences. Their goal is to develop bioinformatics applications for real problems in biotechnology, medicine, ecology and agriculture.

2014 highlights
Molecular phylogenies provide a test-base for biological hypotheses or support downstream bioinformatics analyses. Selection acting on protein-coding DNA affects phylogenetic shapes and should be included during phylogeny inference. CodonPhyML allows for fast maximum likelihood phylogeny inference from protein-coding genes under codon substitution models.

The latest version of CodonPhyML includes multiple partition models that can be used for phylogeny inference from multiple loci or multiple domain proteins. Different models can share parameters, providing there is a convenient framework for hypotheses testing. Furthermore, based on fast approximation algorithms, the group aims to include the alignment uncertainty during phylogeny estimation. This will allow for more accurate phylogenetic inferences from vast high-throughput data.
http://sourceforge.net/projects/codonphyml/

Their recent methods for analysing genomic tandem repeats are implemented in the Tandem Repeat Annotation Library (TRAL).
https://github.com/elkeschaper/TandemRepeats/tree/master

ProGraphMSA implements fast probabilistic graph-based phylogeny-aware alignment with tandem repeats. http://sourceforge.net/projects/prographmsa

Selected publications
Schaper E, *et al.* The majority of human protein tandem repeats are highly conserved within the Eukaryotes. Mol Biol Evol 2014;31:1132-48.
Schaper E, Anisimova M. The evolution and roles of protein tandem repeats in plants. New Phytol 2015;206:397-410.
Dimitrieva S, Anisimova M. Unraveling patterns of site-to-site synonymous rates variation and associated gene properties in protein domains and families. PLoS One 2014;9:e102721.

Laurent Excoffier
Computational Population Genetics Group
University of Bern

What do they do?
Laurent Excoffier and his team use genomic data to understand how populations evolve under the joint effects of demography and selection. They also develop 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.

2014 highlights
The group carried out:
1) empirical studies of hybrid zones in small rodents;
2) the modelling of functional genomic evolution during range expansions;
3) the development of statistical methods to evidence signals of convergent adaptations at the genomic level;
4) experimental evolution with bacteria;
5) human genome scans.

In addition, the team develops Arlequin, a resource which has become very popular. Arlequin allows empirical population geneticists to extract information on genetic and demographic features from a collection of population samples, using a large set of population genetics methods and statistical tests. http://cmpg.unibe.ch/software/arlequin35/

Selected publications
Foll M, *et al.* Widespread signals of convergent adaptation to high altitude in Asia and America. Am J Hum Genet 2014;95:394-407.
Sousa V, *et al.* Impact of range expansions on current human genomic diversity. Curr Opin Genet Dev 2014;29:22-30.
Lischer HEL, *et al.* Ignoring heterozygous sites biases phylogenomic estimates of divergence times: implications for the evolutionary history of Microtus voles. Mol Biol Evol 2014;31:817-31.

Gaston Gonnet

Computational Biochemistry Research Group
ETH Zurich

What do they do?
Gaston Gonnet and his team are interested in bioinformatics problems, in particular the modelling and simulation of molecular sequence data. The group is pursuing large-scale computational problems, such as the Orthologous MAtrix Project (OMA). This particular project aims to produce, automatically, reliable orthologous groups of proteins that are derived from entire genomes.

2014 highlights
The OMA Browser provides orthology predictions among publicly available genomes (see also p. 17). The resource includes a web interface ("OMA Browser"), DAS and SOAP programmatic interfaces, and downloadable data and meta-data in various standard formats. Recently, OMA also provided an efficient stand-alone version that makes it easy to combine custom user data with pre-existing reference genomes.

In 2014, two new releases of the OMA database were made public, in which the number of covered genomes has steadily increased. Over the course of the year, the resource covered an additional 162 genomes, including 18 new plant genomes. A complete redesign of the website was also implemented, and several new functionalities were added to the service. Of note are the gene function predictions based on OMA orthologs and the homeolog relations for polyploid genomes. http://omabrowser.org

As part of the ongoing SIB infrastructure grant for OMA and with the OrthoDB group, Gonnet's group develops the SIBLINGs database. This new service will allow computationally inferred homologs among complete genomes to be available to the public. This particular step is the most time consuming part of the OMA and OrthoDB pipeline. The development of this web service was again a focus over the year and a beta version of the service will be available in early 2015.

Selected publications
Altenhoff AM, *et al.* The OMA orthology database in 2015: function predictions, better plant support, synteny view and other improvements. Nucleic Acids Res 2014;43:D240-9.
Wittwer LD, *et al.* Speeding up all-against-all protein comparisons while maintaining sensitivity by considering subsequence-level homology. Peer J 2014;2:e607.
Coman D, *et al.* Distinct evolutionary strategies in the GGPPS family from plants. Front Plant Sci 2014;5:230.

Jérôme Goudet

Population Genetics and Genomics Group
University of Lausanne

What do they do?
The group led by Jérôme Goudet strives to understand how the interplay between population structure, trait architecture and selection can be disentangled. It focuses on the simulation of individuals and their genomes through time in a realistic landscape, inferring historical events and selection from genomic and phenotypic data, and developing statistical tools for population genomics.

2014 highlights
The focus in 2014 was along the following lines:
1) Vignettes were written to illustrate how to simulate metapopulations, neutral genetic traits, quantitative traits as well as how to estimate population parameters using Approximate Bayesian Computation.
2) Sex can now be determined using a genetic basis rather than being allocated at random.
3) The size of the genetic map can evolve – chromosomes can either grow or shrink.
4) Implementation of a two-stage dispersal, with a gametic dispersal phase and a zygotic dispersal phase. This is an important feature of plant mating systems where not only seeds but also pollen disperse.

Resources
How diversity among genomes has been shaped is a huge challenge for researchers, and will impact fields from evolutionary biology to medical genetics. quantiNEMO has been designed to tackle such questions. It is an individual-based, genetically explicit stochastic simulation program which allows the investigation of the effects of selection, mutation, recombination and drift on quantitative traits.

Besides quantiNEMO, the team is developing HierFstat, a collection of statistical functions for the analysis of population genomic data.

Selected publications
Antoniazza S, *et al.* Natural selection in a postglacial range expansion: the case of the colour cline in the European barn owl. Mol Ecol 2014;23:5508-23.
Pellissier L, *et al.* Soil fungal communities of grasslands are environmentally structured at a regional scale in the Alps. Mol Ecol 2104;23:4274-90.
Alcala N, *et al.* On the transition of genetic differentiation from isolation to panmixia: what we can learn from GST and D. Theor Popul Biol 2014;93:75-84.

Jeffrey D. Jensen

Population Genetics Group
EPFL, Lausanne

What do they do?
The primary research theme of Jensen's group is to draw statistical inference from DNA polymorphism data, and more specifically, to describe the processes which determine the amount and distribution of genetic variation within and between populations. Lab members work on both applied and theoretical problems, in fields ranging from population genomics to medical genetics.

The group focuses 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.

2014 highlights
In 2014, Jensen's team had a particular focus on estimating selection parameters from time-sampled population data, be it via ancient DNA or clinically/experimentally sampled populations. This resulted in the program WFABC (hosted on their lab website: http://jensenlab.epfl.ch), which has been demonstrated to outperform other existing time-sampled based approaches. This method was further utilized to characterize the evolution of drug-resistance evolution (to the antiviral drug oseltamivir) in time-sampled data from influenza virus.

Selected publications
Foll M, *et al.* Influenza virus drug resistance: a time-sampled population genetics perspective. PLoS Genet 2014;10:e1004185.
Bank C, *et al.* A Bayesian MCMC approach to assess the complete distribution of fitness effects of new mutations: uncovering the potential for adaptive walks in challenging environments. Genetics 2014;196:841-52.
Jensen JD. On the unfounded enthusiasm for soft selective sweeps. Nat Commun 2014;5:5281.

Henrik Kaessmann

Functional Evolutionary Genomics Group
University of Lausanne

What do they do?
Henrik Kaessmann and his group pursue integrated bioinformatics projects regarding the functional evolution of mammalian genomes based on publicly available genomic data as well as various experimental data sets generated by the wet lab unit of the group.

The main focus is on the evolution of gene expression levels, the origin and evolution of long non-coding RNA genes, the birth and functional evolution of microRNA genes, the evolution of sex chromosomes – as, for example, the evolution of X dosage compensation and of the Y chromosome – the evolution of germ cell transcriptomes, the functional evolution of new protein-coding genes, and the evolution of alternative splicing.

2014 highlights
Changes in gene expression are thought to underlie many of the phenotypic differences between species, but large-scale analyses of gene expression were, until recently, prevented by technological limitations. As sequels to their major initial study published in 2011, Kaessmann's team assessed the functional evolution of long non-coding RNAs, microRNA editing and Y chromosomes.

Other projects pursued in 2014 – based on unique transcriptomes generated by their group – include the evolution of newly emerged coding genes, untranslated regions (UTRs) and lizard sex chromosomes.

Selected publications
Cortez D, *et al.* Origins and functional evolution of Y chromosomes across mammals. Nature 2014;508:488-93.
Necsulea A, *et al.* The evolution of lncRNA repertoires and expression patterns in tetrapods. Nature 2014;505:635-40.
Necsulea A, Kaessmann H. Evolutionary dynamics of coding and noncoding transcriptomes. Nat Rev Genet 2014;15:734-48.

Bernard Moret

Laboratory for Computational Biology and Bioinformatics
EPFL, Lausanne

What do they do?
Bernard Moret and his group develop, implement, and assess models and algorithms for genome and network evolution. They conduct foundational research in optimization as well as in stochastic models and associated algorithms; software produced in the course of the research is available to the community in source form on its web site.

2014 highlights
In 2014, Moret's team made substantial progress on the very difficult problem of computing optimal solutions for the evolutionary history of groups of large genomes under various models of rearrangements and duplication of genes. Their approaches now allow them to compute maximally parsimonious solutions for positional orthology assignments in groups of 100 or more genomes, with the number of genes ranging from 2,000 to 30,000 (to appear in RECOMB'15, the 19th Annual International Conference on Research in Computational Molecular Biology).

Selected publications
Shao M, *et al.* An exact algorithm to compute the DCJ distance for genomes with duplicate genes. J Comput Biol; Epub ahead of print.
Ghiurcuta CG, Moret BM. Evaluating synteny for improved comparative studies. Bioinformatics 2014;30:i9-18.
Nair NU, *et al.* Study of cell differentiation by phylogenetic analysis using histone modification data. BMC Bioinformatics 2014;15:269.

Marc Robinson-Rechavi

Evolutionary Bioinformatics Group
University of Lausanne

What do they do?
Research in Marc Robinson-Rechavi's group is mainly focused on linking the evolution of animal development to genome evolution. The group develops databases for evolutionary biology, and studies genome evolution in vertebrates.

2014 highlights
The team develops Bgee, a database of gene expression evolution (see also p. 17 and http://bgee.org/). Bgee has grown from 5 to 17 species, with a wealth of RNA sequencing data that is quality controlled, mapped and precomputed for expression calls. The increase in species numbers was a result of the group's collaboration with the Gene Ontology and other projects on the anatomical Uberon ontology, as well as the group's new methods for the rapid development of new developmental and anatomical annotations. All resources and developments linked to Bgee are being made publicly available on GitHub. https://github.com/BgeeDB

With other biocuration groups, Robinson-Rechavi's team produced an ontology of confidence information, for the better use of annotated data from all databases (CIO, available on Bgee GitHub).

The group also continues to develop high-performance computing tools for the detection of natural selection in protein coding genes, and uses them to enrich their database of positive selection.

Selectome: http://selectome.unil.ch/

Selected publications
Haendel MA, *et al.* Unification of multi-species vertebrate anatomy ontologies for comparative biology in Uberon. J Biomed Semantics 2014;5:21.
Rosikiewicz M, *et al.* IQRray, a new method for Affymetrix microarray quality control, and the homologous organ conservation score, a new benchmark method for quality control metrics. Bioinformatics 2014;30:1392-9.
Roux J, *et al.* Patterns of positive selection in seven ant genomes. Mol Biol Evol 2014;31:1661-85.

Nicolas Salamin

Computational Phylogenetics Group
University of Lausanne

What do they do?
Nicolas Salamin and his group estimate and use the evolutionary relationships between species to investigate the processes affecting species evolution. In particular, they are looking at the ecological, genomic and morphological factors that limit and constrain speciation and adaptation.

The group focuses 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.

2014 highlights
The group is developing new approaches to estimate the rate of diversification of species. They use complex Bayesian approaches to incorporate fossil information and link them with phylogenetic methods to better estimate the tempo of species evolution through time. These developments are important to understand the factors that are influencing the appearance and extinction of species over time. The method was implemented into a Python software called pyRate.

In 2014, the models implemented in pyRate were extended by developing a new Bayesian approach that can estimate standard heterogeneous birth-death process on fossil data. The advantage of such an approach is that the birth-death process is directly comparable with estimates done on phylogenetic trees, and it could provide additional information to infer the divergence times of the nodes in a phylogenetic tree estimated by molecular data.

The Bayesian approach is very flexible and Salamin's team incorporated several models to fully account for the heterogeneity in the tempo of species evolution. This includes allowing for shifts of speciation and extinction rates during species evolution, but also the association with external factors such as climate variation or competition from other clades.

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

Selected publications
Dib L, *et al.* Evolutionary footprint of coevolving positions in genes. Bioinformatics 2014;30:1241-9.
Silvestro D, *et al.* Bayesian estimation of speciation and extinction from incomplete fossil occurrence data. Syst Biol 2014;63:349-67.
Zaheri M, *et al.* A generalized mechanistic codon model. Mol Biol Evol 2014;31:2528-41.

Tanja Stadler

Computational Evolution Group
D-BSSE, ETH Zurich, Basel

What do they do?
The team led by Tanja Stadler develops phylogenetic tools to understand evolutionary, epidemiological and ecological processes from genetic sequencing data. They focus on defining and analysing stochastic models, implementing computational methods, analysing empirical data and discussing new insights with clinicians, public health policy makers, ecologists and palaeontologists.

The group addresses questions ranging from epidemiology, public health and medicine, to ecology, evolution and language evolution.

2014 highlights
Tanja Stadler's group developed the add-on "phylodynamics" for the phylogenetic software platform BEAST. The main purpose of this add-on is to quantify epidemiological and evolutionary dynamics using genetic sequencing data from pathogens. For example, the group used it to quantify the early spread of Ebola in Sierra Leone. The results are published in PLoS Currents: Outbreaks (see Selected publications), and were largely covered by the media.

For more news:
www.bsse.ethz.ch/cevo/research/west-africa-ebov-epidemic.html.

Selected publications
Heath T, *et al.* The fossilized birth-death process for coherent calibration of divergence-time estimates. Proc Natl Acad Sci USA 2014;111:E2957-66.
Stadler T, *et al.* Insights into the early epidemic spread of Ebola in Sierra Leone provided by viral sequence data. PLoS Curr 2014;6.
Bošková V, *et al.* Inference of epidemiological dynamics based on simulated phylogenies using birth-death and coalescent models. PLoS Comput Biol 2014;10:e1003913.

Andreas Wagner
Evolutionary Systems Biology Group
University of Zurich

What do they do?
Andreas Wagner and his team study the evolution and evolvability of biological systems at all levels of biological organization – from genes and genomes to biological networks and whole organisms. The group develops bioinformatics tools for the integration of data from a variety of sources, such as comparative whole-genome sequence data, microarray expression data and high-throughput protein interaction data.

2014 highlights
Gene regulation is central to the development of all organisms, and dysregulation of genes is behind many genetic diseases. The group studies the principles by which proteins called transcription factors (TF) bind DNA and regulate genes, in order to understand how changes in transcription factor binding sites create novel patterns of gene regulation.

The variation in genetic sequences which determine the binding of TF is an important facet of evolution. However, the degree to which a genome is robust or in other words able to withstand changes and how robustness affects evolution is unclear.

In a paper published in *Science*, the team investigated the empirical support for mutational robustness by examining TF binding in the mouse and yeast genomes. A network analysis of the degree of variation revealed that the sites with the highest affinity for TF binding exhibit the greatest tolerance for mutations, whereas low-affinity sites exhibit greater sensitivity to mutation. Thus, while mutational robustness and evolvability are antagonistic at the genotypic level, they are synergistic at the phenotypic level.

Selected publications

Payne JL, Wagner A.The robustness and evolvability of transcription factor binding sites. Science 2014;343:875-7.
Wagner A. A genotype network reveals homoplastic cycles of convergent evolution in influenza A (H3N2) evolution. Proc Biol Sci 2014;281:20132763.
Szövényi P, *et al.* Efficient purging of deleterious mutations in plants with haploid selfing. Genome Biol Evol 2014;6:1238-52.

Daniel Wegmann
Statistical and Computational Evolutionary Biology Group
University of Fribourg

What do they do?
Daniel Wegmann and his group characterize the evolutionary and ecological processes which shape the realm of biological diversity observed today. To achieve this, they design and evaluate new statistical and computational approaches to infer complex evolutionary histories, and apply them to the wealth of data that is currently generated.

They focus on four themes of research: numerical algorithms for inference problems, inferring evolutionary histories, the evolution of morphological characters and human genetics and disease.

2014 highlights
The group develops and maintains ABCtoolbox, a versatile and user-friendly software package for Approximate Bayesian Computation (ABC). ABC is a numerical inference method that bypasses analytical evaluations of likelihood functions, and is hence particularly suited in situations in which the likelihood function is either unknown or cannot be solved numerically.

The group has just finished developing an ABC algorithm suitable for inference problems of very high dimensionality, and which is available through ABCtoolbox. This algorithm will be particularly useful in the field of genomics, where many parameters are inferred for each locus.

As an example, the group is currently interested in understanding the strength of selection that acts upon each individual position in the genome of viruses when exposed to an antiviral drug. Since the demographic history of the virus population is a confounding factor affecting all positions, Wegmann and his team would like to learn about that history while learning about the strength of selection at each individual locus. Thanks to the group's new algorithm, such a joint inference is now possible and computationally efficient.

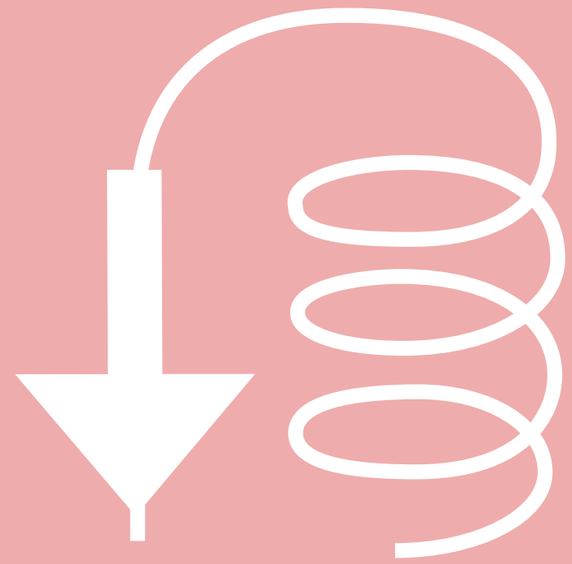
Selected publications

Adrion JR, *et al.* Drosophila suzukii: the genetic footprint of a recent, world-wide invasion. Mol Biol Evol 2014;31:3148-63.
Zawistowski M, *et al.* Analysis of rare variant population structure in Europeans explains differential stratification of gene-based tests. Eur J Hum Genet 2014;22:1137-44.
Dussex N, *et al.* Postglacial expansion and not human influence best explains the population structure in the endangered kea (Nestor notabilis). Mol Ecol 2014;23:2193-209.

STRUCTURAL BIOLOGY

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

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



Simon Bernèche

Computational Biophysics Group
University of Basel

What do they do?
Simon Bernèche's group uses molecular mechanics and statistical physics approaches to understand how the functions of proteins emerge from their 3D structure. They are especially interested in the mechanisms that regulate ion permeation in potassium channels and the resulting impact on neuron signalling.

The group focuses on elementary mechanisms of protein folding, permeation and gating mechanisms in ion channels, transport mechanisms in passive and active transporters, and stochastic simulation of membrane activity.

2014 highlights
The relation between the amino acid (aa) sequence of a protein and its 3D structure remains largely unknown. A lasting dream is to elucidate the aa-dependent driving forces that govern the protein folding process, and understand the structural properties of intrinsically disordered proteins, which play fundamental functional roles in human physiology.

By comparing molecular dynamics (MD) simulations and NMR residual dipolar coupling (RDC) data, the group studied the structure and dynamics of 9-residue long peptides in which only the central residue was changed.

The originality of their approach was to use the RDC data to validate the all-atom MD simulations by clustering the sampled conformations according to their matching with the RDC data. This allowed them to extract from the simulations the relevant microscopic information that provides a detailed view of the folding mechanism.

At the atomic scale, the group showed that a single aa determines the local hydration level of the backbone. Bulkier aa prevent water molecules from interacting properly with the backbone polar groups, favouring the formation of hydrogen bonds between these polar groups and thus the folding of the peptide.

These findings are consistent with the general understanding that protein folding evolves from states in which the backbone polar groups are shielded from the solvent and thus preferably interact together. For the first time, such a mechanism was shown to take place at the level of a single aa. The calculations revealed how a short stretch of aa can act as a nucleus point from which folding is initiated. This local mechanism effectively reduces the conformational search space and explains how protein folding can be so efficient.

Selected publications
Thomson AS, *et al.* Initial steps of inactivation at the K+ channel selectivity filter. Proc Natl Acad Sci USA 2014;111:E1713-22.

Matteo Dal Peraro

Laboratory for Biomolecular Modelling
EPFL, Lausanne

What do they do?
The group led by Matteo Dal Peraro strives to understand the physical and chemical properties of complex biological systems, in particular their function with regard to structure. To this end, they use and develop multiscale molecular simulation methods and dynamic integrative modelling to investigate large molecular assemblies, mimicking conditions found in the cellular environment.

2014 highlights
In order to study the biophysics of biological membranes using models that closely mimic their real composition, the group has developed LipidBuilder, a web-based tool that allows the design of lipids of different nature and their assembly into models of cellular membrane portions. http://lipidbuilder.epfl.ch

LipidBuilder capabilities grew in 2014 with the possibility to:
1) design and store a large variety of lipid species;
2) construct asymmetric membrane bilayers of different composition, size and shape;
3) use several molecular force fields for molecular simulations of biological systems.

This framework enabled the lab to build more realistic models of membrane compartments, like those of the mitochondrion, the endoplasmic reticulum, or the bacterial membrane.

Another emerging line of research in the lab is the investigation of crowding effects on biomolecules. The cell is a crowded place, whose cytoplasm contains myriads of different molecules. Such a crowded environment affects molecular diffusion, and hence the molecular properties and events that depend on it, such as collision frequencies and subcellular localization. By using molecular simulations and NMR techniques, the group is investigating the impact of crowding agents of different nature on the structure, dynamics and function of proteins.

Selected publications
Spiga E, *et al.* Dissecting the effects of concentrated carbohydrate solutions on protein diffusion, hydration and internal dynamics. J Phys Chem B 2014;118:5310-21.
Lemmin T, *et al.* Perturbations of the straight transmembrane α-helical structure of the amyloid precursor protein affect its processing by γ-secretase. J Biol Chem 2014;289:6763-74.
Seitz P, *et al.* ComEA is essential for the transfer of external DNA into the periplasm in naturally transformable Vibrio cholerae cells. PLoS Genet 2014;10:e1004066.

Olivier Michielin & Vincent Zoete

Molecular Modelling Group
University of Lausanne

What do they do?
The Molecular Modelling group develops and employs techniques for the computer-aided rational design of proteins or small molecules, for research and treatment of human diseases, mostly in the field of oncology.

2014 highlights
SwissDrugDesign is a large collection of tools covering all aspects of computer-aided drug design (see also p. 16). Among them:
1) SwissDock predicts the molecular interactions that may occur between a target protein and a small molecule. swissdock.ch
2) SwissParam provides topology and parameters for the molecular modelling of drug-like molecules. swissparam.ch
3) SwissSidechain gathers information on hundreds of commercially available non-natural sidechains for *in silico* peptide design. swisssidechain.ch
4) SwissBioisostere collects several million molecular substructural replacements extracted from the literature. swissbioisostere.ch

In 2014, the group released a new web service, SwissTargetPrediction, to predict the targets of bioactive small molecules. This is useful to understand the molecular mechanisms which underlie a given bioactivity, to rationalize possible side effects and to predict off-targets. swisstargetprediction.ch

The development of a new tool, SwissADME, was initiated to predict the physicochemical, PK and PD properties and drug likeness of real or virtual organic compounds.

The group is also active in the field of protein engineering and engineered several T-cell receptors with enhanced binding properties for two melanoma-related antigens. T-cells expressing this new receptor exhibit an increased ability to kill cancer cells. These T-cell receptors are now being tested in a mouse model.

The team is also developing inhibitors of IDO, a cancer target. They obtained several compounds showing a significant activity in a cancer mouse model. The group recently discovered new scaffolds of potential interest for the design of IDO inhibitors.

Selected publications
Gfeller D, et *al.* SwissTargetPrediction: a web server for target prediction of bioactive small molecules. Nucleic Acids Res 2014;42:W32-8.
Daina A, et *al.* iLOGP: a simple, robust, and efficient description of n-octanol/water partition coefficient for drug design using the GB/SA approach. J Chem Inf Model 2014;54:3284-301.
Röhrig UF, et *al.* Detailed analysis and follow-up studies of a high-throughput screening for indoleamine 2,3-dioxygenase 1 (IDO1) inhibitors. Eur J Med Chem 2014;84:284-301.

Torsten Schwede

Computational Structural Biology Group
University of Basel

What do they do?
Torsten Schwede and his team focus on the development of methods for modelling 3D protein structures and their applications in biomedical research. The main emphasis is on homology modelling approaches – using evolutionary information to model protein tertiary and quaternary structures, as well as protein-ligand interactions.

2014 highlights
SWISS-MODEL is a widely used server for generating 3D models of a protein (see also p. 16 and swissmodel.expasy.org). It uses information from homologous protein structures as templates to build models for target protein sequences. Over the course of evolution, the quaternary structure of proteins is less conserved than their tertiary structure. Therefore, even in the same protein family, a range of different oligomeric assemblies can frequently be observed, which poses a challenge for the modelling and prediction of protein structures. The team has developed an improved approach to predict the oligomeric structure of target proteins.

The accuracy of this new approach is being evaluated by CAMEO, a project developed by the group to benchmark the accuracy of protein structure prediction servers. cameo3d.org

Homology models find widespread applications in biomedical research where experimental structures are not available. The suitability of a model for a specific application depends on its accuracy, and model quality estimation is an essential step in structure prediction. To this end, the group is developing QMEANBrane as a statistical potential of mean force. Most current approaches for estimating model quality are limited to soluble proteins. However, as membrane proteins play crucial roles in many biological processes and are important drug targets, the group has extended their QMEAN approach to membrane protein structures.

Torsten Schwede's team is developing the Protein Model Portal as part of the Structural Biology Knowledgebase. proteinmodelportal.org

Torsten Schwede is also responsible for the sciCORE center (see p. 20), and for organizing the SIB's annual [BC]² Basel Computational Biology Conference. bc2.ch

Selected publications
Biasini M, *et al.* SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res 2014;42:W252-8.
Studer G, *et al.* Assessing the local structural quality of transmembrane protein models using statistical potentials (QMEANBrane). Bioinformatics 2014;30:i505-11.
Schmidt T, *et al.* Modelling three-dimensional protein structures for applications in drug design. Drug Discov Today 2014;19:890-7.

SYSTEMS BIOLOGY

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

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

Bastien Chopard
Scientific and Parallel Computing Group
University of Geneva

What do they do?
Bastien Chopard and his team model and simulate complex systems and natural phenomena. They focus on multiscale modelling and computing, high performance computing, cellular automata, lattice Botzmann methods, multi-agent systems, and optimizing techniques and machine learning.

The group works on massively parallel computers, like the Scylla and Baobab clusters at the University of Geneva, or the CADMOS supercomputer. They also develop resources needed to perform simulations of large natural systems.

Finally, Bastien Chopard's group develops the Lattice Boltzmann solver on massively parallel machines, as well as a multiscale modelling and computing infrastructure (MAPPER FP7 project).

2014 highlights
During the course of 2014, the group developed a numerical model of epithelium to study its elastic properties.

They also provided a physical description of the adhesion and aggregation of platelets in the presence of shear induced diffusion, and determined the adhesion and aggregation rate of platelets by comparing experimental and numerical simulations.

Finally, they determined the threshold of wall shear stress in aneurysms to determine whether cerebral aneurysms in given patients will thrombose or not.

Selected publications
Montandon S, *et al.* Two waves of anisotropic growth generate enlarged follicles in the spiny mouse. EvoDevo 2014;5:33.
Florez-Valencia L, *et al.* A new model for the construction of virtual fully resolved flow-diverters and their effect in blood flow simulations for studying intracranial aneurysms. Submitted.
Anzai H, *et al.* Combinational optimization of strut placement for intracranial stent using a realistic aneurysm. J Flow Control 2014;2:66-76.

Manfred Claassen
Computational Single Cell Biology Group
ETH Zurich

What do they do?
The group led by Manfred Claassen carries out research to understand the composition of heterogeneous cell populations and their function in the context of cancer and immune biology. Their research can then be used to pinpoint therapeutic targets with a perspective to designing drugs.

2014 highlights
Manfred Claassen's lab develops statistics, machine learning and dynamic systems methods to describe the cellular dynamics and functionally relevant cell types in cancer and immune biology. These approaches are implemented as program prototypes and ideally further refined to widely applicable software.

In 2014, the group has developed an approach to describe heterogeneous cell populations such as tissues by means of cell-transcending networks that make explicit the differentiation-induced relationships among related cell types. This approach, for instance, facilitates the detection of rare and yet disease relevant cell types, such as stem cells.

Selected publications
Arvaniti E, Claassen M. Markov network structure learning via ensemble-of-forests models. Uncertain Artif Intell 2014;2014:42-51.
De Vargas Roditi L, Claassen M. Computational and experimental single cell biology techniques for the definition of cell type heterogeneity, interplay and intracellular dynamics. Curr Opin Biotechnol 2014;34C:9-15.

Rudiyanto Gunawan

Chemical and Biological Systems Engineering
Group / ETH Zurich

What do they do?
The group led by Rudi Gunawan develops mathematical tools for systems modelling and analysis of chemical and biological networks. Their research spans many aspects of biology, from gene and metabolic networks in single cells to the ageing process in humans and animals and to bioreactors of cell culture fermentation in the pharmaceutical industry.

2014 highlights
Gunawan and his team developed TRaCE (Transitive Reduction and Closure Estimation) as a novel inference method for gene regulatory networks (GRNs) from expression data. They designed TRaCE to specifically address undetermined GRN inference by employing an ensemble inference strategy. TRaCE produces two networks – the most complicated (largest) and the least complicated (smallest) – among a sum of networks that agree with expression data from gene perturbation experiments. The ensemble of consistent networks can be derived from these bounds, and the difference between the largest and the smallest networks describes the uncertainty in the network inference. TRaCE outperformed top algorithms in DREAM4 and DREAM5 network inference challenges in inferring GRN from multi-gene knock-out data.

The group also developed a new assessment strategy for comparing network inference performance based on an ensemble inference framework, by addressing the question: how to compare inference methods when the inference problem is underdetermined and, thus, does not have a (unique) solution? In the new assessment, methods are not penalized for errors (false positive/false negative) which are associated with gene regulatory interactions that are not inferable from gene perturbation data. The inferability of regulatory interactions from a given perturbation experiment was determined using TRaCE.

The group also developed a new tool, REDEMPTION, for the creation of kinetic models of biological networks.

Selected publications

Siegenthaler C, Gunawan R. Assessment of network inference methods: how to cope with an underdetermined problem. PLoS One 2014;9:e90481.
Ud-Dean SM, Gunawan R. Ensemble inference and inferability of gene regulatory networks. PLoS One 2014;9:e103812.
Liu Y, Gunawan R. Parameter estimation of dynamic biological network models using integrated fluxes. BMC Syst Biol 2014;8:127.

Vassily Hatzimanikatis

Laboratory of Computational Systems Biotechnology
EPFL, Lausanne

What do they do?
The team concentrates on the mathematical modelling of complex cellular processes and develops computational methods for the integration of experimental information from multiple levels. The aim is to provide experimentally testable hypotheses and targets for purposeful redesign and manipulation of these processes.

2014 highlights
BNICE is a computational framework for the identification of novel metabolic reactions and pathways. It uses a database of over 580 generalized enzyme rules, and employs a pathway generation algorithm for the generation of all possible routes from known metabolic intermediates to target natural and man-made chemicals. These routes are composed of generalized enzyme reactions. Hence, any of the novel reactions in these routes can be associated with a class of known enzymes, which catalyse similar reactions. BNICE enables researchers to explore the potential of nature's chemical toolbox, as well as discover and exploit its diversity.

In 2014, using the principles from BNICE, the group introduced a computational framework for the atom-level reconstruction of metabolic networks from *in silico* labelled substrates, which allows the tracking of the atoms' fate through the reconstructed metabolic network. The method was applied for the reconstruction of an atom-level representation of core metabolic networks of *E. coli*.

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

The group also develops ORACLE, a framework for the development of large-scale and towards-genome-scale kinetic models that account for uncertainty in available experimental data.

Selected publications

Hadadi N, *et al.* A computational framework for integration of lipidomics data into metabolic pathways. Metab Eng 2014;23:1-8.
Almquist J, *et al.* Kinetic models in industrial biotechnology – improving cell performance. Metab Eng 2014;24:38-60.
Soh KC, Hatzimanikatis V. Constraining the flux space using thermodynamics and integration of metabolomics data. Methods Mol Biol 2014;1191:49-63.

Dagmar Iber

Computational Biology Group
D-BSSE, ETH Zurich, Basel

What do they do?
Dagmar Iber and her team focus on the development of data-based models of biological signalling networks to gain a greater understanding of the dynamics and evolution of cellular signalling. The experimentally validated models are used to investigate the biological system, as well as to address more general questions regarding the evolution and the design of cellular signalling networks. The group is particularly interested in the mechanisms that enable the spatiotemporal control of developmental processes such as organ development in the growing embryo.

2014 highlights
The Iber group focuses on key processes in developmental biology. Highlights from the last year include the delineation of a number of developmental mechanisms. These include:
1) a mechanism by which developmental patterns can scale according to the size of a (growing) embryonic domain;
2) the mechanism by which digit asymmetry is lost in bovine limbs, but not in murine or human limbs;
3) a receptor-ligand based mechanism that allows Turing patterns to emerge for virtually any physiological parameter value.

In addition, Dagmar Iber and her team developed a computational framework to simulate models of organogenesis on growing embryonic domains. To enable the coupled simulation of signalling and tissue dynamics, the group also developed LBIBCell, a tightly coupled cell-based mechano-regulatory simulation tool. A cellular resolution of the tissue domain is important to describe adequately the impact of cell-based events, such as cell division, cell-cell interactions and spatially restricted signalling events.

Selected publications

Menshykau D, *et al.* An interplay of geometry and signaling enables robust lung branching morphogenesis. Development 2014;141:4526-36.
Fried P, Iber D. Dynamic scaling of morphogen gradients on growing domains. Nat Commun 2014;5:5077.
Lopez-Rios J, *et al.* Attenuated sensing of SHH by Ptch1 underlies evolution of bovine limbs. Nature 2014;511:46-51.

Christian Mazza

Biomathematics and Computational Biology Group
University of Fribourg

What do they do?
Christian Mazza and his team study biological networks by focusing both on their geometrical structure (graphs, patterns) and on their underlying dynamics (deterministic and stochastic). Typical examples are Lotka-Volterra dynamics on complex ecological networks and formation of the plant vascular system.

The group also provides services in statistics for the life science community in Fribourg and is working on several projects related to chemical reaction networks.

2014 highlights
Christian Mazza published a book with a fellow colleague, Michel Benaïm, Professor of mathematics at the University of Neuchâtel: *Stochastic Dynamics for Systems Biology* (CRC Press). The book describes a systematic study of the many varied stochastic models used in systems biology today, and shows how mathematical models can lead to conceptual insights on cellular processes, such as signalling and metabolic pathways, phosphorylation processes, genetic switches and gene transcription.

Selected publications

Feller C, *et al.* Pattern formation in auxin flux. J Math Biol 2014;68:879-909.
Mazza C, Banaïm M. Stochastic dynamics for systems biology. CRC Press, Taylor & Francis Group, 2014.

Michel Milinkovitch

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

What do they do?

The laboratory led by Michel Milinkovitch combines evolutionary development biology (EvoDevo) and the study of physical processes to understand the mechanisms which generate complexity and diversity in the living world.

The group specializes in non-classical model species in reptiles and mammals and integrates data and analyses from comparative genomics, molecular developmental genetics, physical experiments, as well as from computer modelling and numerical simulations.

Most of the group's research requires intensive high-performance computing such as the comparative analysis of genome sequences among species, the reconstruction of super-high resolution coloured 3D geometry of biological objects, and the simulation of the interaction between skin cells and photons for the generation of structural colours.

2014 highlights

In 2014, Milinkovitch's team finalized the development of R2OBBIE-3D, an integrated system that combines a robotic arm, a high-resolution digital colour camera, an illumination basket of high-intensity light-emitting diodes and state-of-the-art 3D reconstruction approaches.

They demonstrated that R2OBBIE generates accurate 3D models of biological objects between 1 and 100 cm, with colour-texture and geometric resolutions greater than 15 µm without the use of magnifying lenses. R2OBBIE has the potential to greatly improve quantitative analyses of pattern formation in living organisms in addition to providing multiple new applications in biomedical and forensic sciences. The results are to be published this year.

Selected publications

Montandon SA, *et al.* Two waves of anisotropic growth generate enlarged follicles in the spiny mouse. EvoDevo 2014;5:33.

Martins AF, *et al.* R2OBBIE-3D, a fast robotic high-resolution system for quantitative phenotyping of surface geometry and colour-texture. In press.

Teyssier J, *et al.* Photonic crystals cause active colour change in chameleons. Nat Commun 2015;6:6368.

Félix Naef

Computational Systems Biology EPFL, Lausanne

What do they do?

Félix Naef's lab works in the field of computational and systems biology, in the areas of gene regulation, stochastic transcription and circadian rhythms. While their research is largely computational, they started an independent wet lab working on circadian gene expression and stochastic transcription using imaging in single cells.

2014 highlights

Circadian and cell cycles are two periodic, noisy, cell-autonomous processes with a period of about one day. Consequently, when these cycles run in parallel in the same cell, their coupling may lead to resonances or even synchronization. Observations of circadian variations in mitotic indices in mammalian cells and on the daytime-dependence of cell division in mouse liver have led to the hypothesis that the circadian cycle might gate cell-cycle progression. A better understanding of how the two systems mutually interact is currently of great interest, notably with regard to the role of circadian clocks in proliferating tissues, such as the epidermis, immune or stem cells.

As part of the group's ongoing analyses of circadian cycles and transcriptional bursting in single mammalian cells, Félix Naef's team has become very interested in the interactions of circadian and cell cycle oscillators. In particular, they hope to gain a better understanding of the dynamic consequences of possible mutual coupling between the two oscillators, which may lead to resonance or synchronization phenomena. The group developed a combined experimental and modelling approach to investigate the problem.

The group's recent quantitative time-lapse imaging study of circadian cycles in dividing mammalian NIH3T3 cells clearly indicated that both oscillators tick in a tightly synchronized state. Contrary to their expectations, they showed that, in NIH3T3 cells, the cell cycle progression exerts a unilateral influence on the circadian clock, and not the opposite.

Selected publications

Bieler J, *et al.* Robust synchronization of coupled circadian and cell cycle oscillators in single mammalian cells. Mol Syst Biol 2014;10:739.

Mauvoisin D, *et al.* Circadian clock-dependent and -independent rhythmic proteomes implement distinct diurnal functions in mouse liver. Proc Natl Acad Sci USA 2014;111:167-72.

Quinodoz M, *et al.* Characteristic bimodal profiles of RNA polymerase II at thousands of active mammalian promoters. Genome Biol 2014;15:R85.

Igor V. Pivkin

Scientific Computing Group Università della Svizzera italiana, Lugano

What do they do?

Igor Pivkin and his team's interests lie in the area of multiscale/multiphysics modelling, and parallel large-scale simulations of biological and physical systems. Their focus is on the development of new computational models and corresponding numerical methods suitable for the next generation of supercomputers.

The group's current projects include stochastic multiscale modelling of motion, interaction, deformation and aggregation of cells under physiological flow conditions, biofilm growth, coarse grained molecular dynamics simulations, as well as modelling of transport processes in healthy and tumour-induced microcirculation.

2014 highlights

Electrostatic and polarization effects play an important role in many biological and physical systems, including membranes, vesicles and cells. The modelling of such effects on large length and time scales is of great interest to many scientific communities – from biophysics to bioengineering and applied mathematics to name but three.

Dissipative Particle Dynamics (DPD) is an efficient method for modelling mesoscale behaviour of systems at timescales larger than are currently accessible by atomistic methods. Polarizability has been introduced into some coarse-grained molecular dynamics simulation methods. However, it has not been done with DPD before.

The group developed a new polarizable coarse-grained water model for DPD, which employs long-range electrostatics and Drude oscillators. They expect their model to be used in a wide range of studies of physical and biological systems, where polarization effects are important and not negligible.

Selected publications

Peter EK, Pivkin IV. A polarizable coarse-grained water model for dissipative particle dynamics. J Chem Phys 2014;141:164506.

Peter EK, et *al.* How water layers on graphene affect folding and adsorption of TrpZip2. J Chem Phys 2014;141:22D511.

Cruz-Chu E, et *al.* Structure and response to flow of the glycocalyx layer. Biophys J 2014;106:232-43.

Jörg Stelling

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

What do they do?

Jörg Stelling's group studies complex cellular networks to elucidate their design and operational principles. This involves the development of concepts and tools for network inference, system modelling and analysis, and the design of biological circuits with new functionalities. Bioinformatics methods are used for the large-scale analysis of cellular networks.

2014 highlights

With Vital-IT, the group developed the web resource MetaNetX.org for accessing, analysing and manipulating genome-scale metabolic networks (GSMs) and biochemical pathways. This resource integrates data from various public resources in a standardized format. The methods enable researchers to construct and analyse GSMs efficiently.

With regard to metabolic network construction and analysis methods, progress in 2014 included the development of a novel pattern-based approach for inferring features of metabolic reactions.

Identifying suitable patterns in complex biological interaction networks helps in understanding network functions and allows for predictions at the pattern level. By recognizing a known pattern, one can assign its previously established function. However, former approaches failed for previously unseen patterns, when patterns overlap, or when they are embedded into a new network context.

The team showed how to conceptually extend pattern-based approaches. A probabilistic framework decodes the implicit information in the networks' metabolite patterns to predict metabolic functions. They demonstrated the predictive power by identifying "indicator patterns"; e.g. for enzyme classification, by predicting directions of novel reactions and known reactions in new network contexts, and by ranking candidate network extensions for filling gaps in the network.

Such methods can be used to improve genome annotations as well as metabolic network models. Other theoretical approaches for the modularization of biological networks, in view of experimental design for topology inference for example, will help to exploit structural network properties beyond metabolism.

Selected publications

Ausländer D, *et al.* A synthetic multifunctional mammalian pH sensor and CO2 transgene-control device. Mol Cell 2014;55:397-408.

Lang M, *et al.* Cutting the wires: modularization of cellular networks for experimental design. Biophys J 2014;106:321.

Ganter M, *et al.* Predicting network functions with nested patterns. Nat Commun 2014;5:3006.

Mihaela Zavolan
RNA Regulatory Networks Group
University of Basel

What do they do?
Mihaela Zavolan and her team combine experimental and computational approaches to study the regulation of gene expression in development, ageing and disease.

They develop:
1) experimental methods to isolate and quantify various types of RNAs;
2) computational methods to analyse data and make the results accessible to experimental biologists through web servers;
3) computational methods to predict the regulatory elements in genomes;
4) computational models to understand the dynamics of gene regulation in various conditions.

2014 highlights
Mihaela Zavolan's group is one of the first to develop combined, experimental and computational approaches to capture regulatory interactions within cells, in particular between proteins and RNAs.

In this regard, they made substantial progress in characterizing the process of the 3' end formation of mRNAs. Almost all cellular mRNAs rely on a specific protein complex for their termination. Most human genes can terminate transcription at multiple places, a process called alternative termination. In 2014, the group described the sequence-specific component of the 3' end processing complex, which had remained elusive for over 20 years. By carrying out high-throughput measurements of transcript and protein expression levels, the group further discovered that alternative polyadenylation does not serve to systematically increase or decrease gene expression, but rather that the effects are gene-specific. Finally, the group built a comprehensive catalogue of 3' end processing sites in the human genome.

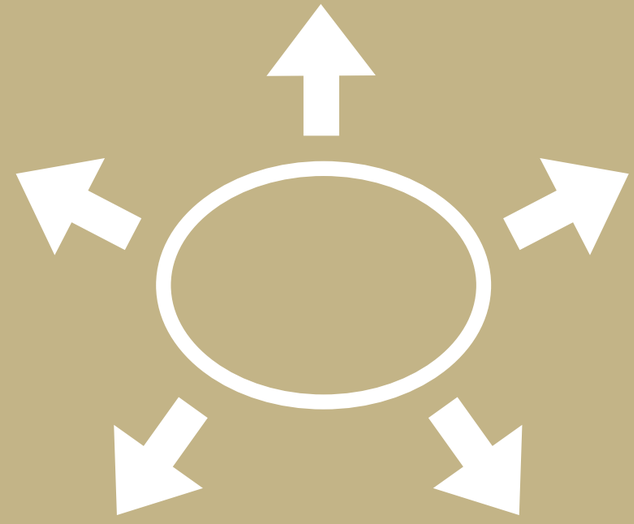
Another of the group's main projects concerns gene regulation by the small regulatory RNAs called miRNAs. In 2014, the group developed a new method to predict miRNA target sites genome-wide and is currently extending the method to other small RNAs. They also contributed to the development of a novel experimental method for isolating miRNA binding sites *in vivo*.

Selected publications
Imig J, *et al.* miR-CLIP capture of a miRNA targetome uncovers a lincRNA H19-miR-106a interaction. Nat Chem Biol 2015;11:107-14.
Gruber AR, *et al.* Global 3' UTR shortening has a limited effect on protein abundance in proliferating T cells. Nat Commun 2014;5:5465.
Hausser J, Zavolan M. Identification and consequences of miRNA-target interactions--beyond repression of gene expression. Nat Rev Genet 2014;15:599-612.

BIOINFORMATICS INFRASTRUCTURE

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

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



Peter Kunszt

Service and Support for Science IT – S3IT
University of Zurich

What do they do?
The S3IT unit led by Peter Kunszt provides IT infrastructure, software, tools and services to the research groups of the University of Zurich, including life sciences and medicine (see also p. 21). S3IT also takes part in national projects and cooperates with similar technology-oriented groups to ensure that its expertise is always up-to-date.

2014 highlights
The group is an infrastructure and service provider group. Their “resource” is the local cluster and storage infrastructure, including scientists with bioinformatics expertise who they provide to research groups as “embedded bioinformaticians”.

The resource is a specialized application and user support; a collaborative model, where IT experts from S3IT work closely with end users to address their computational and storage needs. To this end, S3IT provides access to a large-scale private cloud infrastructure for storing and processing research data.

In only one year, the team has been 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 took part in two projects that are highly specialized in 3D imaging, making use of new Light Sheet Microscopy technology, and developed 3D segmentation and genealogy tracing algorithms in these projects.

With regard to infrastructure, Peter Kunszt and his team finished the public procurement process for their local ScienceCloud infrastructure. The new infrastructure – to be installed during the course of 2015 – will have over 3,000 CPU cores and 1.5 PB of usable storage.

Selected publications

Wiewiórka MS, *et al.* SparkSeq: fast, scalable, cloud-ready tool for the interactive genomic data analysis with nucleotide precision. Bioinformatics 2014;30:2652-3.
Quandt A, *et al.* Using synthetic peptides to benchmark peptide identification software and search parameters for MS/MS data analysis. EuPA Open Proteom 2014;5:21-31.
Rost H, *et al.* pyOpenMS: A Python-based interface to the OpenMS mass-spectrometry algorithm library. Proteomics 2014;14:74-7.

Bernd Rinn

Scientific IT Services – SIS
ETH Zurich

What do they do?
The SIS group provides services and resources for scientific computing in computing- and data-intensive fields of research, with a strong focus on providing services for life science research (see also p. 21).

SIS operates high-performance computing (HPC) clusters, provides solutions for research data management and develops data analysis workflows. The group also offers platforms for sharing information, and provides training and documentation for selected topics of scientific programming and software packages.

2014 highlights
SIS provides important HPC resources developed with a focus on flexibility and usability. Users have access to a large portfolio of compilers, libraries, open-source and commercial analysis applications.

In 2014, phase 1 of the new general purpose HPC cluster Euler with over 10,000 CPU cores, over 400 TB of fast, scalable disk space, and a peak performance of about 250 TFlops was successfully established and brought into production. It features a classical user interface for batch queuing, special support for very long-running jobs, and support for “computing clouds”. The cloud interface brings user flexibility to a new level as it allows running any platform and operating system users might need. The need for computing power continues to increase strongly, which is why SIS immediately started with the planning, design and procurement of phase 2.

The scientific workflow manager “BeeWM”, developed together with the University of Basel and the TargetInfectX project, reached a usable state in 2014. This system is “data-driven” by design and has advanced features to support selective re-computations of parts of the workflow when workflow nodes are changing, or when new data become available. It uses openBIS as a data management backbone and will be shipped with the next release version of openBIS. BeeWM is already being used in several projects which require support for heavy computational workflows in imaging and genomics, enabling both a high level of computational efficiency and reproducibility of results.

Selected publications

Rämö P, *et al.* Simultaneous analysis of large-scale RNAi screens for pathogen entry. BMC Genomics 2014;15:1162.
Osterwalder M, *et al.* HAND2 targets define a network of transcriptional regulators that compartmentalize the early limb bud mesenchyme. Dev Cell 2014;31:345-57.

Jacques Rougemont

Bioinformatics and Biostatistics Core Facility – BBCF
EPFL, Lausanne

What do they do?
The BBCF offers support for the analysis of high-throughput genomic data for life science researchers in the Geneva-Lausanne area (see also p. 22). The group develops tools and methods to facilitate, automate and extend the analyses. In particular, it focuses on all aspects of gene regulation, functional genomics and the analysis of DNA sequencing data.

2014 highlights
The expansion of high-throughput sequencing technologies forces biologists and bioinformaticians to find efficient solutions for managing and storing their data. BioRepo (Biological data Repository) addresses those needs by allowing the storage, management and sharing of genomic data among collaborators, but also by facilitating visualization in publicly available genome browsers.

BioRepo is a web-based application developed in Python, SQL and Javascript, and runs on the Vital-IT infrastructure. It manages the archiving and retrieval of large files from the Vital-IT hierarchical file system, and uses Shibboleth as a Swiss-wide authentication system to identify users and their host institution. Although there is a single back-end system, the user interface is personalized for every lab via simple configuration files that can be edited at any time.

In 2014, major usability improvements were made with the implementation of an Application Programming Interface, of a Dropbox-like loading system and bridging to the UCSC genome browser by interactively building track hubs. The system currently holds over 5 Tb of data from 6 different labs at EPFL and University of Geneva. http://biorepo.epfl.ch/

In 2014, Rougemont’s team also upgraded and extended their high-throughput sequencing data analysis platform. http://htsstation.epfl.ch

Selected publications

Woltering JM, *et al.* Conservation and divergence of regulatory strategies at Hox loci and the origin of tetrapod digits. PLoS Biol 2014;12:e1001773.
Dos Santos AX, *et al.* Systematic lipidomic analysis of yeast protein kinase and phosphatase mutants reveals novel insights into regulation of lipid homeostasis. Mol Biol Cell 2014;25:3234-46.
Knight B, *et al.* Two distinct promoter architectures centered on dynamic nucleosomes control ribosomal protein gene transcription. Genes Dev 2014;28:1695-709.

Ioannis Xenarios

Vital-IT Group

What do they do?
Vital-IT maintains a competence centre in bioinformatics and computational biology (see also p. 19). The group enables and supports life and medical science research in multiple domains, such as behaviour, ecology, genetics, genomics, metagenomics, pharmacodynamics, phylogeny, population genetics, oncology, proteomics, structural biology, systems biology.

2014 highlights
Scientific support and collaborations
The services and support were provided to various collaborators in Swiss universities and international research groups through joint research projects, such as:
1) the development of the OpenFlu database which, in collaboration with the Food and Agriculture Organization (FAO), enables the monitoring of the flu pandemic with genomic information. In 2015, SIB was designated as the bioinformatics reference centre for FAO.
2) the optimization of software enabling heavy computational calculations on the high-performance computer to search for new genes related to the symptoms of Alzheimer’s disease, in the context of the European project AgedBrainSysBio.

www.vital-it.ch/projects/project_list.php

Computational infrastructure
To meet the users’ needs, and remain up-to-date with the new generation of computers, Vital-IT experts maintain and develop new computer algorithms. For example, Vital-IT optimized the parallel computing of the Fastepistasis software that is used to search for biomarkers in Alzheimer’s disease.

Education
Given the growing need to train highly qualified personnel and to develop, coordinate and maintain high-performance level computational resources enabling (big) large-scale data analysis and management, Vital-IT’s pool of experts is deeply involved in the continued education of researchers, in particular via SIB’s training programme.

ELIXIR
Through its active participation in the European programme ELIXIR, Vital-IT along with all the SIB groups is contributing to the development of a sustainable support for life and medical science research in Europe.

Selected publications

Genolet R, *et al.* Duality of the murine CD8 compartment. Proc Natl Acad Sci USA 2014;111:E1007-15.
Hauser PM, *et al.* Microbiota present in cystic fibrosis lungs as revealed by whole genome sequencing. PLoS One 2014;9:e90934.
Seguín-Estévez Q, *et al.* Extensive remodeling of DC function by rapid maturation-induced transcriptional silencing. Nucleic Acids Res 2014;1:9641-55.

ACKNOWLEDGEMENTS

We gratefully acknowledge the following funders, sponsors and partners for their financial support and encouragement in helping us fulfil our mission in 2014:

- The Swiss government and in particular:
  - The State Secretariat for Education, Research and Innovation (SERI)
  - The Swiss National Science Foundation (SNSF)
  - The Commission for Technology and Innovation (CTI)
- Our institutional members (see list pp. 10-11)
- The National Institutes of Health (NIH)
- SystemsX.ch
- Oncosuisse
- The Cogito Foundation
- The Force Foundation
- The Solidar Immun Foundation
- Sophia Genetics SA
- GeneBio SA

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

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

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

SIB profile 2015 written by the SIB Communications Department, with contributions by SIB Members.

SIB Group Leaders' pictures: Nicolas Righetti, rezo.ch
Cover image: Artist's view of how DNA supercoiling stimulates enhancer-promoter interaction, by Julien Dorier, Vital-IT, SIB.

Design and layout: Atelier Valenthier, valenthier.ch

SWISS WORLD-CLASS BIOINFORMATICS FOR MEDICINE AND LIFE SCIENCES

Swiss Institute of Bioinformatics

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

www.isb-sib.ch

elixir SWITZERLAND