# SIB Profile

## 2018

Empowering advances in life sciences and health for 20 years

20 YEARS
SIB
Swiss Institute of Bioinformatics

# Foreword

**Felix Gutzwiller**
President of the
Foundation Council

**Manuel Peitsch**
Chairman of the
Board of Directors

**Ron Appel**
Executive Director

The SIB Swiss Institute of Bioinformatics was founded 20 years ago. Today, it represents the largest national bioinformatics network in Europe. In this special edition of our SIB Profile, we are looking back over 20 years of services to the life sciences and health community.

Since 1998, SIB's activities have developed along two main axes: infrastructure for life scientists and the bioinformatics community. As we write, the Institute brings together 800 members, 67 research and service groups, 12 core facilities and over 150 resources, covering fields as diverse as proteomics, phylogeny and machine learning. Alongside these two main axes, SIB also has at heart a number of strategic activities: the sustainability of core data resources and personalized health, to name but two.

Defending the long-term sustainability of data resources is rooted in our Institute's DNA. SIB is part of the Global Life Sciences Data Resources working group, whose mission is to develop a global infrastructure to support and ensure the sustainability of core data resources, which are of crucial importance for the life sciences. An SIB-led study to identify a candidate funding model for knowledgebases has also been completed and will contribute to informing the working group.

On the personalized health front, the Institute initiated the implementation of a national secure and interoperable infrastructure, as part of the Swiss Personalized Health Network (SPHN). By the close of the year, the network had approved 15 projects – in fields as varied as oncology and inflammatory disease immunotherapy – in which SIB Group Leaders are well represented.

> **"Defending the long-term sustainability of data resources is rooted in our Institute's DNA"**

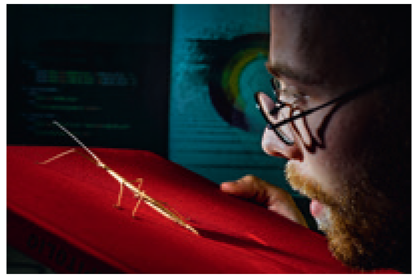
SIB is also involved in developing personalized health capacities in Swiss hospitals, and bringing bioinformaticians and clinical actors together to harmonize routine next-generation sequencing practices on a national scale. And along with the University Hospitals of Basel and Lausanne, SIB has just launched a Certificate of Advanced Studies in personalized molecular oncology for professionals who are faced with the new challenges posed by complex genomic data, for the benefit of patient care.

Over these years, SIB has come a very long way, with the critical support of many bodies: we would like to thank the State Secretariat for Education, Research and Innovation (SERI), the Federal Assembly, the Swiss National Science Foundation and all those in funding roles, as well as our partner institutions.

While there are still exciting challenges in front of us, we would like to express our heartfelt gratitude to all SIB members, without whom none of this could have happened, and whose expertise and dedication has brought Swiss bioinformatics to where it is today.

Two decades of Swiss bioinformatics are something to celebrate, and we hope you will join us for the festivities and actions planned throughout the year.

# Table of contents



**Providing life scientists with a complete bioinformatics infrastructure**

**Creating a strong bioinformatics community**

# SIB in brief

## OUR MISSION AND OBJECTIVES

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

**67**
research and service groups

Some
**800**
members, including
**200**
employees

**20**
institutional partners across Switzerland

**12**
core facilities and competence centres

Over
**150**
databases and software tools accessible via the ExPASy web portal

Over
**2,200**
peer-reviewed articles published since 1998

### Infrastructure

SIB provides the national and international life science community with a state-of-the-art bioinformatics infrastructure, including resources, expertise and services.

**DATABASES AND SOFTWARE TOOLS**
SIB creates, maintains and disseminates a large portfolio of databases and software tools worldwide.

**CORE FACILITIES AND COMPETENCE CENTRES**
It offers easily accessible, in-depth expertise and support in bioinformatics through core facilities and competence centres.

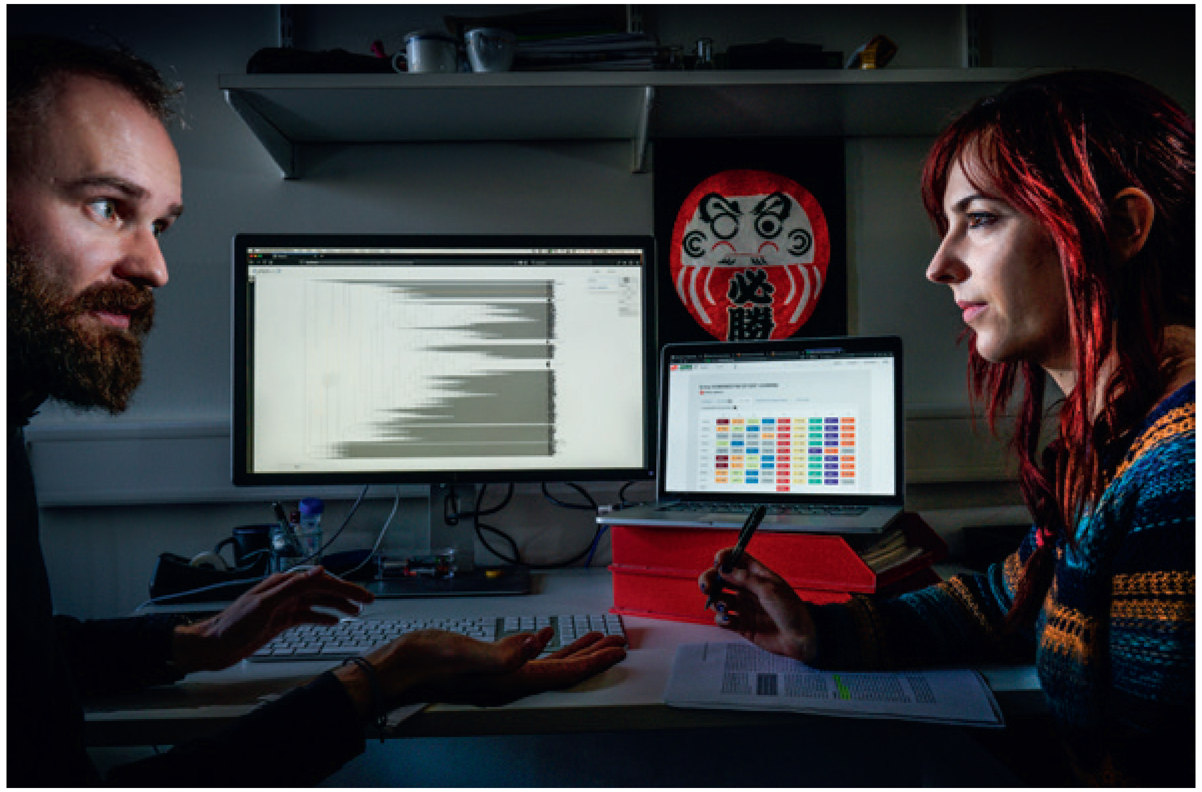
**PERSONALIZED HEALTH**
It supports researchers and clinicians with know-how, resources and a secure data infrastructure dedicated to personalized health.

### Community

SIB brings together world-class researchers based in Switzerland and delivers training in bioinformatics.

**SCIENTIFIC COLLABORATION**
SIB represents and promotes Swiss bioinformatics nationally and internationally, by bringing together bioinformatics research and service groups from Swiss universities and research institutes. SIB fosters scientific collaboration and innovation at the highest level of excellence.

**TRAINING IN BIOINFORMATICS**
It provides life scientists and clinicians with a large portfolio of courses and workshops. It fosters exchanges among bioinformatics and computational biology PhD students, and trains them using the most up-to-date methods necessary for their research.



**Top**
Life scientists rely on various types of infrastructure to perform their research: knowledge-based resources, computing power, software…

**Bottom**
Collaboration is at the heart of science. And it all starts with a strong community.

CHAPTER 1

# Why bioinformatics?

Answering biological questions using computer- and data science

## Bioinformatics: a definition

### FIGHTING DISEASES. DISCOVERING SPECIES. UNDERSTANDING LIFE.

Life scientists and clinicians have always tried to assemble data and evidence to find the right answers to fundamental questions. In 2018, data are no longer lacking. But a different kind of problem has emerged. Nowadays, new technologies are producing data at an unprecedented rate. Indeed, so much data - and of so many kinds (see right-hand examples) - that they can no longer be interpreted by the human mind alone.

**Enter bioinformatics.**

Bioinformatics is the application of computer technology to the understanding and effective use of biological and clinical data.

It is the discipline that stores, analyses and interprets the 'big data' generated by life science experiments, or clinical data, using computer science. And dedicated data experts. This multidisciplinary field brings together biologists, computer scientists and mathematicians, as well as statisticians and physicists.

Computer-based approaches have become indispensable to science, by allowing researchers to advance their understanding of complex systems.

Bioinformatics encompasses:

**DATA- AND KNOWLEDGEBASES** for storing, retrieving and organizing biological information, including biocuration, to maximize the value of biological data;

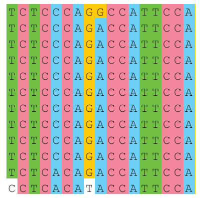
**SOFTWARE TOOLS** for modelling, visualizing, analysing, interpreting and comparing biological data;

**COMPUTING AND STORAGE INFRASTRUCTURE** to process large amounts of data;

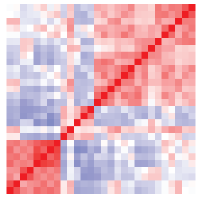
**ANALYSIS** of complex biological datasets or systems in the context of particular research projects;

**RESEARCH** in a wide variety of biological fields using computer- and data science and leading to applications in diverse areas, from agriculture to precision medicine. (SEE P.50-51)

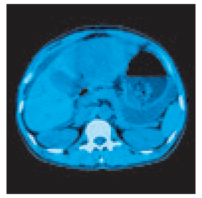
**What sort of data are we talking about?**
Bioinformatics is dealing with a broad spectrum of complex data types. Often all at once.



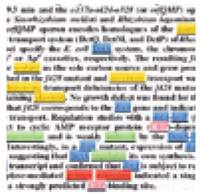
Sequence data



Expression data



Imagery data



Text data
And more...

# Converting biological questions …

Massive amount of data and data types: genetics, text, biochemical, imagery…

**SIB**
**Swiss Institute of Bioinformatics**

Creating databases and software tools

Analysing complex datasets and conducting research

Providing computing and storage infrastructure

Hospitals and clinics

Research institutes

Private sector

International institutions

Life sciences and health actors

## … into answers

with various applications

Basic research

Medicine

Ecology

Agriculture

Tailoring treatment to cancer patients

Understanding how chameleons change colour

Predicting the spread of the bird flu

# Some of the things bioinformatics has made possible since 1998

Though the beginnings of bioinformatics can be tracked back to the early 1950s, key biological and medical questions have been solved thanks to technological breakthroughs and bioinformatics advances in the past two decades.

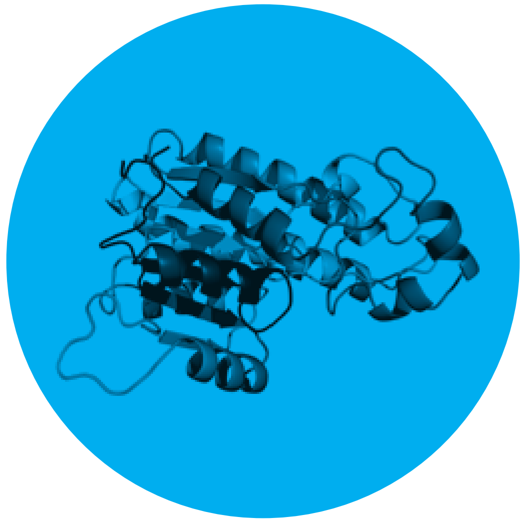
### Detecting susceptibility to a disease

Spotting our genetic Achilles' heel

Individuals share genetic peculiarities, or variations, in their DNA sequence, which can predispose them to conditions such as diabetes.

Today, it is possible to selectively sequence all the protein-coding regions in a genome in one go, thanks to a technique known as **WHOLE EXOME SEQUENCING (WES).**

- Using this technique, SIB Group Leader Jacques Fellay and his team identified a mutation which increases the severity of common colds among children.
  SEE P. 54



### Optimizing a protein's function

Tell me your shape and I'll tell you what you do

The structure of a protein determines how it functions and interacts with other proteins and molecules in order to perform its vital tasks. Twenty years ago, predicting a protein's function was tricky, as it was difficult to change a protein's sequence in order to find out how this would affect its structure and function.

Today, our knowledge on **3D PROTEIN STRUCTURES** has grown dramatically to reach 90% coverage in some model organisms, such as the bacterium *Escherichia coli*.

- SIB's core resource SWISS-MODEL, a platform for generating 3D protein models, can be used by scientists to engineer proteins in order to improve vaccines.
  SEE P. 64

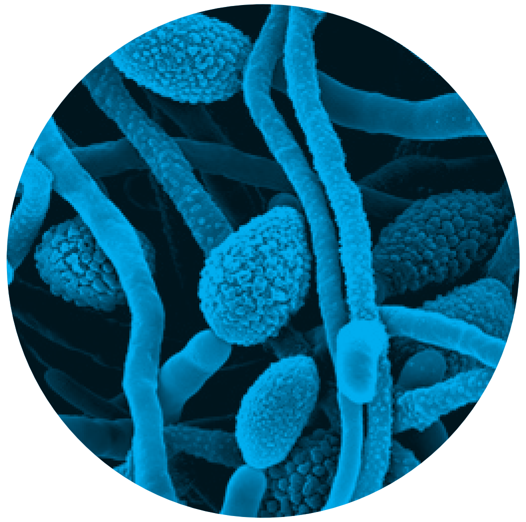
### Tracing our origins

Making do with less to see more

**ANCIENT DNA** has a lot to say about how organisms evolved and populated the world. Over time, though, DNA undergoes damage, and becomes a story in which paragraphs are missing. This has long impeded researchers' ability to read into our past.

Today, bioinformatics techniques account and to some extent, compensate for the less than ideal quality of the data. More knowledge can thus be extracted from less material.

- SIB Group Leader Laurent Excoffier and his team took part in a study that shed light on the social and reproductive behaviour of our Paleolithic ancestors, over 30,000 years ago.
  SEE P. 60



### Revealing the multitude

Many birds and one stone

Imagine our guts, a handful of soil, or a drop of blood. Everything is teeming with microscopic entities, including bacteria, viruses, fungi and more. How can you identify each one of them?

Today, the overall DNA present in a sample – taken for example from a patient infected by an unknown pathogen – can be retrieved thanks to **METAGENOMICS**, whether it represents a single or indeed several thousand species.

- SIB Group Leader Jérôme Goudet and his team used metagenomics to document the fungal biodiversity of alpine soils.
  SEE P. 61

### Unfolding the tree of life

Splitting paths

We have been building the tree of life for centuries, ever since scientists took an interest in how species – or populations – were related. For years, however, relationships were merely based on what could be seen with the naked eye or through the microscope.

Today, **PHYLOGENETIC TECHNIQUES** are able to delve into the invisible, and compare genotypes and phenotypes in far greater detail, thus refining the tree of life, but also predicting the evolution of a disease.

- SIB Group Leader Tanja Stadler developed a phylogenetic tool to model the dynamics of the Ebola virus during the 2014 outbreak in West Africa.
  SEE P. 62

> **“Bioinformatics is everywhere but no one can see it.”**

> **“Bioinformatics helps you find a needle in a haystack.”**

> **“Bioinformatics is another way of creating better drugs.”**

Bioinformatics defined
in visitors’ own words
at SIB’s *Planète Santé 2016*
booth

# Activities for the layperson

Bioinformatics is at the heart of life science research. As it is involved in issues pertaining to health, ethics and society, explaining its role to the public is crucial. SIB has been involved in communicating bioinformatics for 20 years, through a range of fun activities and at various events.

### In 2017, the SIB team met you at:

**A STAGE PLAY TO UNDERSTAND OUR ORIGINS**
Inspired by a research carried out by SIB Groups on the peopling of Australia, *Genome Odyssey* was produced at the Museum of Ethnography in Geneva.

**SIB HACKATHONS**
Children, developers and structural biologists used programming or augmented reality apps to tackle biological problems, as part of the 2017 Open Geneva Hackathons.

**Follow us on Facebook**
Our dedicated outreach channel in French and English.

### In 2018, meet the SIB team at:

**IMAGES DE SCIENCE**
*Geneva, Museum of Science History, until 26 August*
Science image exhibition in the grounds of the Museum, with two panels provided by SIB.

**MYSTÈRES DE L’UNIL**
*University of Lausanne, 2-3 June*
Discover the field of metagenomics with a workshop entitled *Bacteria, viruses, fungi and me…*

**NUIT DE LA SCIENCE**
*Geneva, Museum of Science History & Campus Biotech, 7-8 July*
What is the role of bioinformatics in drug design, personalized health and metagenomics?

**PLANÈTE SANTÉ LIVE**
*Geneva, Palexpo, 4-7 October*
Latest advances in clinical bioinformatics in the field of personalized health.

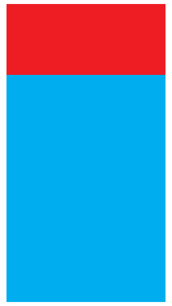
## BIOINFORMATICS IN THE CLASSROOM

SIB has designed several websites and workshops to introduce bioinformatics to schoolchildren, teachers and students:

**CHROMOSOMEWALK**
An award-winning website to explore human chromosomes and the world of DNA, genes and proteins.
www.chromosomewalk.ch

**DRUG DESIGN**
A web-based workshop to acquire a simple yet realistic picture of how drugs are designed with the help of computers.
www.drug-design-workshop.ch

**THE METAGENOMICS PIZZA**
A workshop to grasp the concept of metagenomics, and identify all the species present in a pizza, based on their DNA.



Each year
**2,100**
participants
including 1,600 students
(12-18 years old)

CHAPTER 2

# 20 years of commitment

In 2018, SIB celebrates its twentieth anniversary.

In the following pages, discover some of the Institute's highlights over the past two decades; learn about its role in federating a growing Swiss bioinformatics community; and find out more about its engagement to make open data sustainable.

# Federating Swiss bioinformatics

A few questions to Mauro Dell’Ambrogio, State Secretary for Education, Research and Innovation.

Switzerland is a small country benefiting from an excellent international position on the scientific, innovation and education fronts. Small, but highly decentralized, it is also one of the few countries in the world with a coordinated bioinformatics infrastructure, provided by SIB to life scientists and clinicians with the support of the Swiss government for 20 years.

**When did you first hear about bioinformatics?**

Mauro Dell’Ambrogio I was appointed State Secretary for Education and Research in 2008. Bioinformatics was a growing scientific field then and I quickly discovered SIB and its ambitions.

In particular, it was during my first official trips abroad to visit life science research institutions, that I noticed the ‘Swiss-Prot’ database being mentioned by different scientists, and in different parts of the world. For me, this was the best sign that this was an important initiative in Switzerland. The topic of bioinformatics has been with me for almost 10 years now.

**Every four years, the Swiss government allocates significant funding to SIB. Can you tell us a bit more about the underlying process, and the key elements that have contributed to making the Institute eligible for funding over the past 20 years?**

MDA First of all, the federal funding is primarily allocated to SIB for its function as a supporting institution for research in life science. When we prepare the fund allocation every four years, the applicants have to undergo a scientific evaluation. If the result is that something is indispensable for the development of a specific domain, it is of national importance and demonstrates a high level of quality, we can provide subsidiary funding – in addition to support from the universities and cantons.

Another key element that contributes to SIB’s eligibility is its governance model: it has to be a non-commercial research institution, which is independent from the universities. By including both the providers and users of bioinformatics services among its researchers, the Institute ensures it can lead and coordinate the field. So SIB is a good example of how we can reconcile competition and cooperation in research.

**How is Switzerland positioned internationally in terms of its bioinformatics infrastructure?**

MDA Thanks to SIB, Switzerland has been a first-mover in the bioinformatics field internationally. The Swiss government has been financing bioinformatics infrastructure since 1998. In 2013, the European Research Infrastructure ELIXIR was founded, with reference to the SIB model. This and other developments in Europe have since strengthened our leading position in bioinformatics, not only as pioneers in the field, but also in research as a whole.

**How do you expect the role played by bioinformatics to evolve over the next 20 years?**

MDA Personalized health and big data will be among the main challenges for science in the next years. I think the experience and possibilities SIB offers will be of utmost importance, in particular in the field of personalized health, where we gave a central role to the Institute in the framework of the national “Swiss Personalized Health Network” initiative.

> “SIB is a good example of how we can reconcile competition and cooperation in research”

**As you know, this year marks SIB’s 20th birthday. Would you like to express your wishes to the Institute, its employees and its members?**

MDA Switzerland is among the top countries for research and innovation. However, it is a continuous challenge to remain at the top, and institutions like SIB play a key role in supporting our country’s leading position.

I would therefore like to thank all those at SIB who have helped to strengthen Switzerland’s standing as a location for top research for the past 20 years. Their work benefits society and the economy – today and tomorrow.

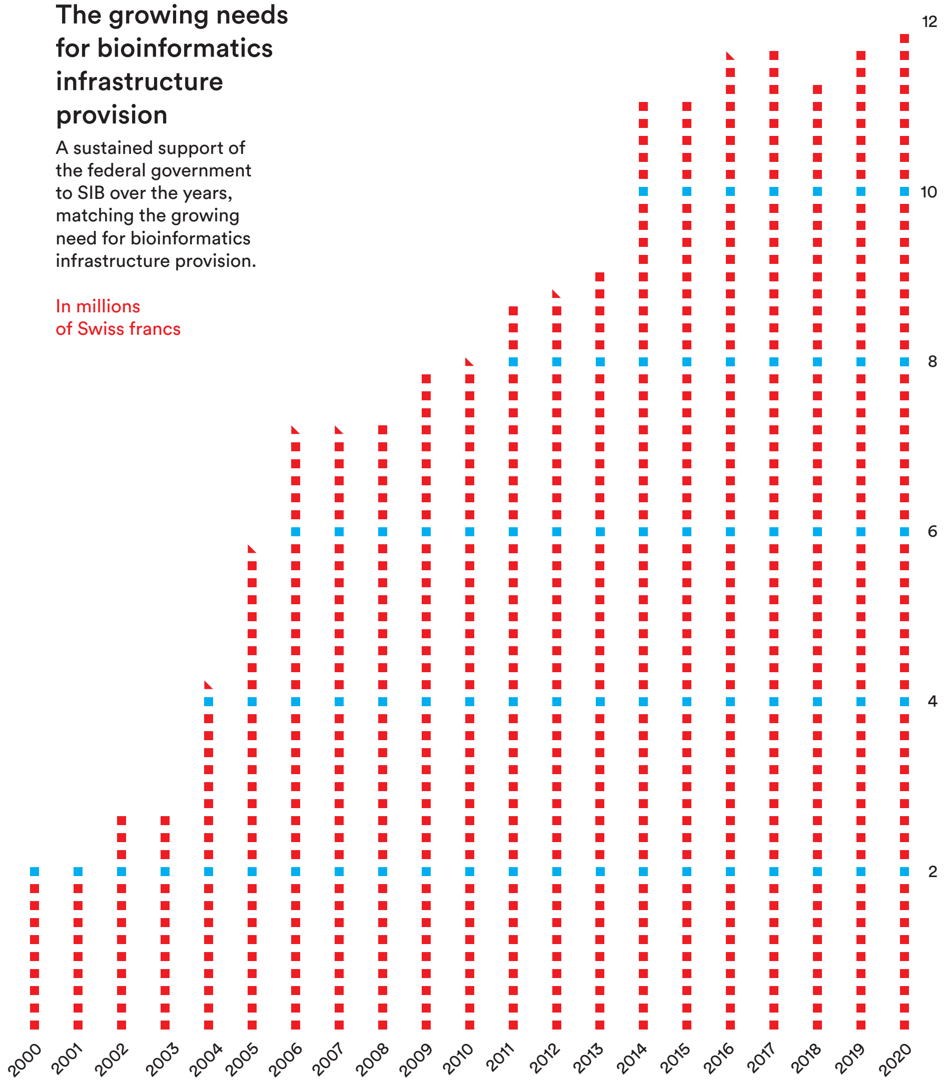
With this I wish you all every success for the future! ●

Watch this interview online on our Youtube channel: goo.gl/uR6WvB

### The growing needs for bioinformatics infrastructure provision

A sustained support of the federal government to SIB over the years, matching the growing need for bioinformatics infrastructure provision.

In millions of Swiss francs

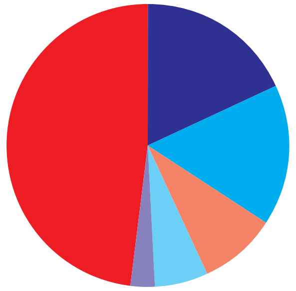


# Representing bioinformatics

## IN SWITZERLAND...

### All roads lead to bioinformatics

SIB Members come from a wide range of educational backgrounds



Biology
■ 48%

Other (engineering, bioinformatics, administrative...)
■ 18%

Informatics
■ 16%

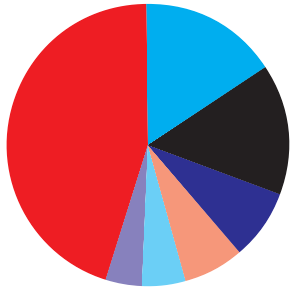
Physics
■ 9%

Chemistry
■ 6%

Mathematics
■ 3%

### Bioinformatics as a career

SIB Members' current positions are diverse too



Bioinformatician, Computational Biologist
■ 45%

Others (Biochemist, Structural Biologist, Mathematician...)
■ 16%

N/A
■ 15%

Biocurator
■ 8%

Software Developer
■ 7%

Administrative Staff
■ 5%

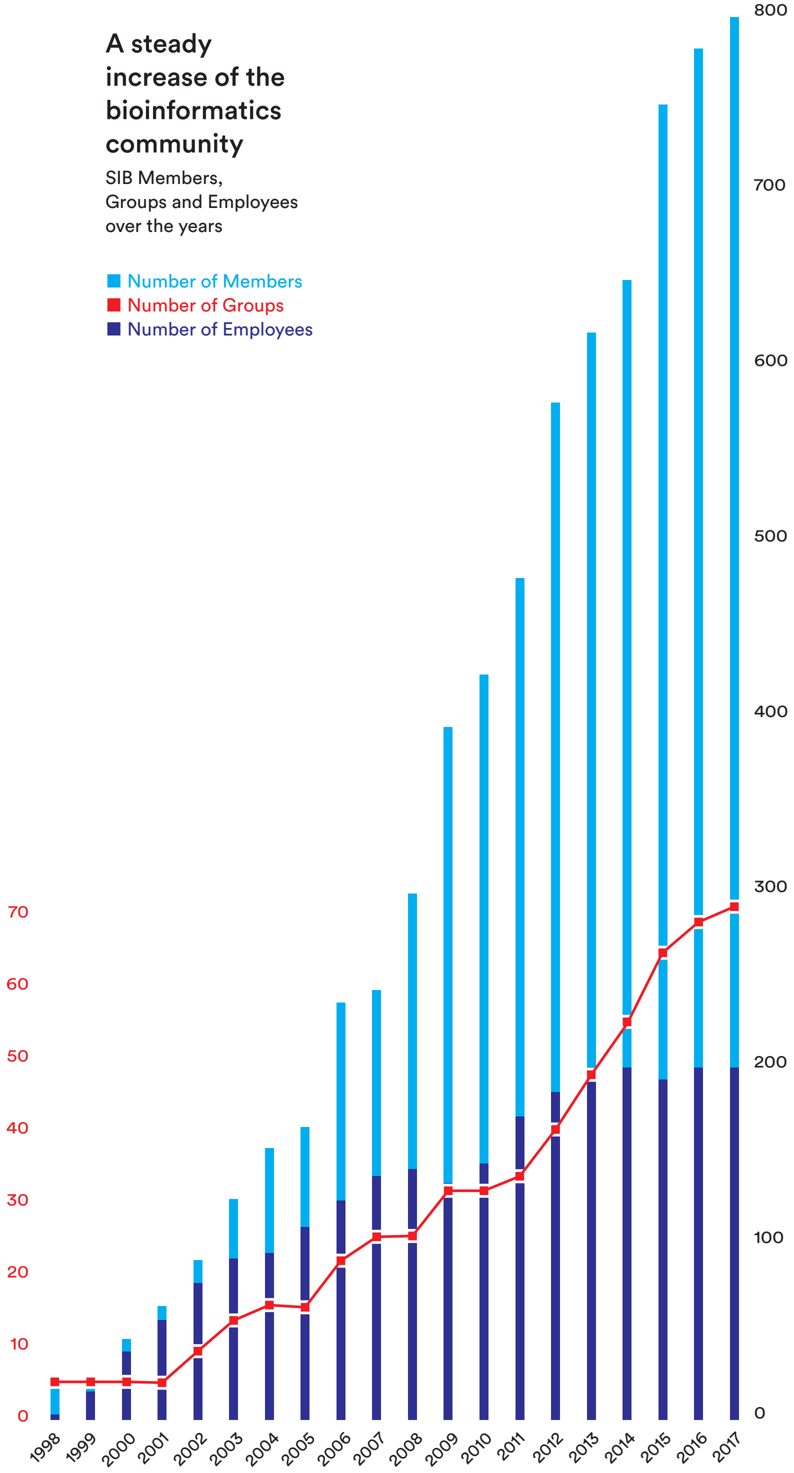
Statistician
■ 4%

* Data collected from 110 SIB Members as part of an internal survey (2017)

### A steady increase of the bioinformatics community

SIB Members, Groups and Employees over the years

■ Number of Members
■ Number of Groups
■ Number of Employees



## ...AND BEYOND

In addition to its history of collaboration with other national bioinformatics institutes, SIB also acts as the Swiss Node of ELIXIR, the organization for a sustainable infrastructure to support life science research across Europe.

Thanks to its pioneering role, SIB's expertise is renowned outside Switzerland and the Institute is regularly invited to share its experience with and present its activities to international institutes and organizations. Recent visits from abroad have included delegates from Bahrain, China, Greece, Iran and South Africa.

## Faces of Swiss bioinformatics

**Cristina Casals Casas, PhD**

UniProtKB/Swiss-Prot biocurator
Swiss-Prot Group

“I am a biocurator in SIB's Swiss-Prot Group, specializing in the manual annotation of proteins of human origin, as well as those from other vertebrates. My job is to ensure that knowledge is extracted from the best-quality biological data generated from the literature, and that such knowledge becomes accessible to the community.

I believe biocuration is the key for a better understanding of human genomic variation, which is fundamental for the future of precision medicine. To me, SIB's growth over the past 20 years shows the essential role played by bioinformatics in this new era of big data.”

**Yann Christinat, PhD**

Clinical bioinformatician
Clinical Pathology Division, HUG

“I am an SIB embedded bioinformatician in the Clinical Pathology Division at the Geneva University Hospitals. My background is in computer science with a solid grounding in biology, and thus allows me to understand and address biological questions by developing novel computing approaches, such as clinical diagnostics tools and pipelines. Bioinformatics has the power to bring the best of computer science to medicine and biology. For instance, many advances in cancer biology were only made possible through the development of bioinformatics tools and algorithms.

To me, SIB's growth over the past 20 years shows that bioinformatics has become a driving force in biological research. Furthermore, the recent addition of a “clinical bioinformatics” unit at SIB highlights a change of scope: bioinformatics is no longer limited to research but is now being used to treat patients on a regular basis.”

**Gerardo Tauriello, PhD**

Development team Leader
Computational Structural Biology Group

“I am an SIB software developer in the Computational Structural Biology Group of Torsten Schwede. I apply my programming skills to the development and optimization of bioinformatics resources such as SWISS-MODEL – an SIB core resource – for the life science community.

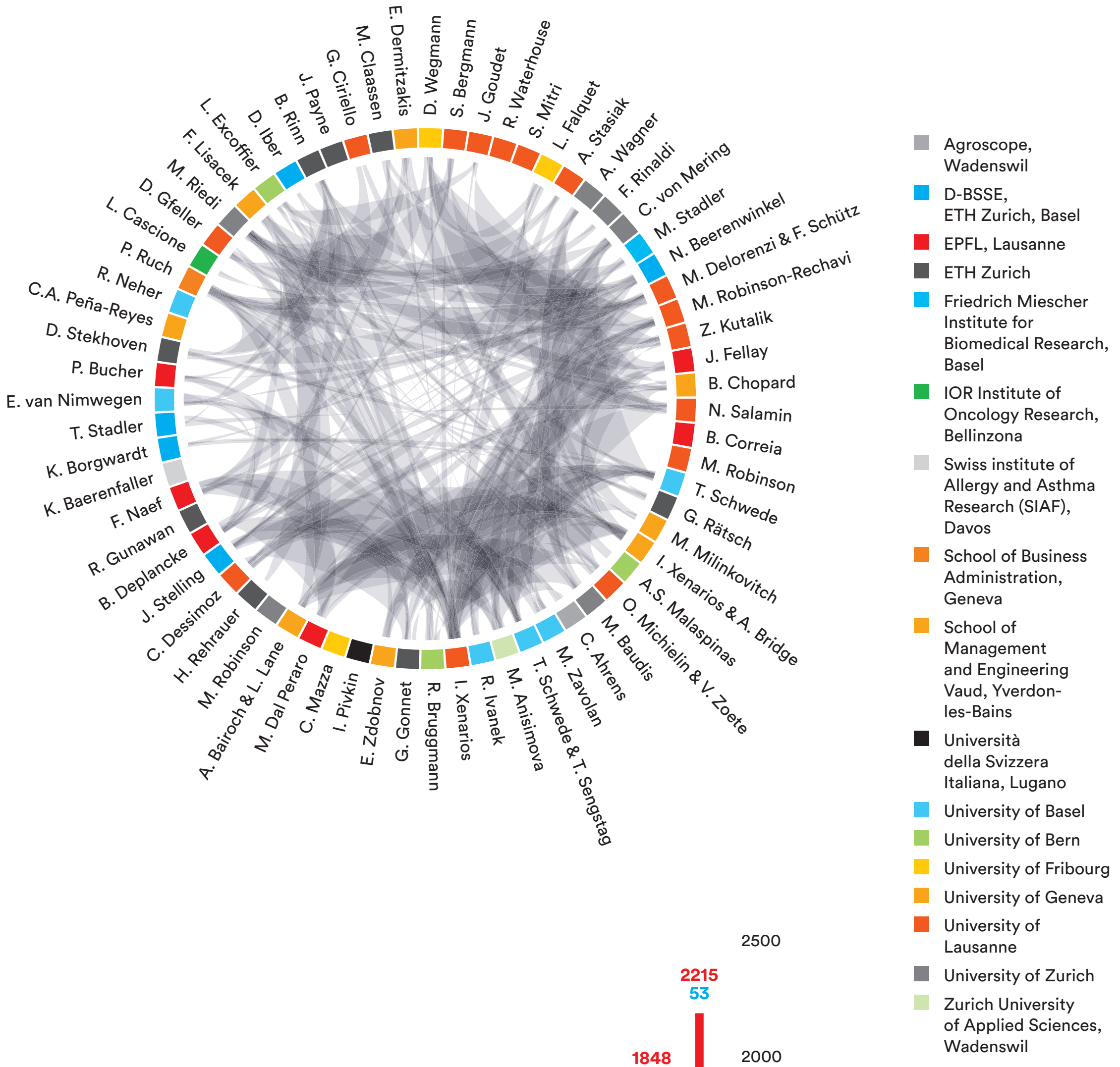
I believe bioinformatics has the power to provide the necessary tools and methods to improve our understanding of the complex machinery of life and enhance health. In that context, I am always excited to see how our users employ SWISS-MODEL-generated 3D models of proteins to help explain their function in pathogens, viruses, cancers and many more.

To me, SIB's growth over the past 20 years shows the importance of maintaining and improving reliable and top-quality services and resources to support life scientists in their research.”

# Fostering scientific collaboration…

## The SIB collaboration network

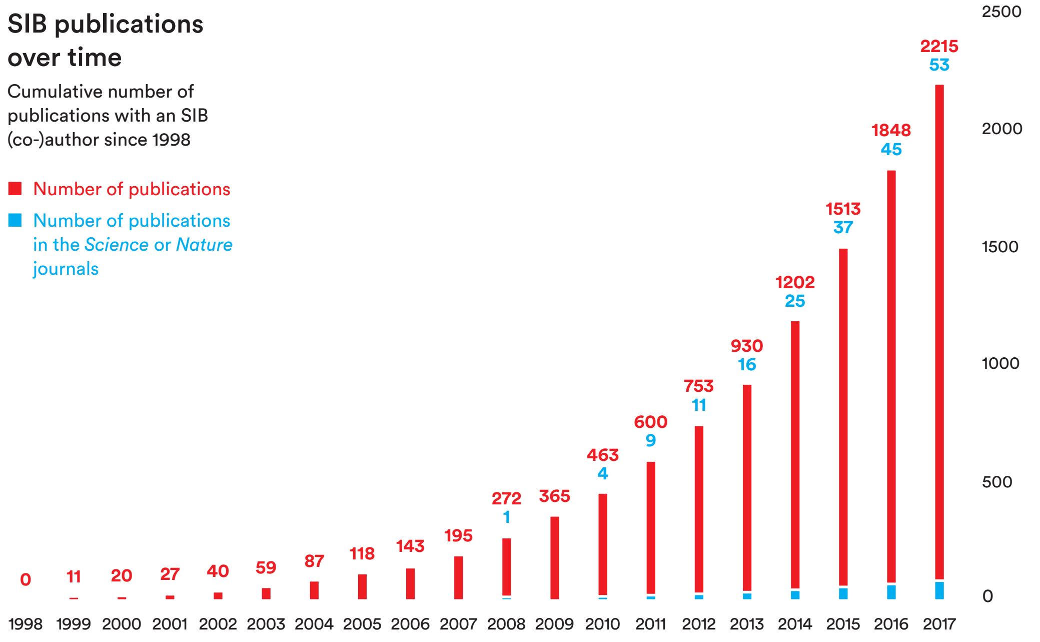
Past or ongoing collaborations between SIB Groups: the more collaborations, the broader the connection.



THIS COLLABORATION NETWORK WAS GENERATED WITH A PROGRAMME DEVELOPED BY SIB GROUP LEADER MICHAEL BAUDIS: PROGENETIX.ORG/COLLAPLOTS

## SIB publications over time

Cumulative number of publications with an SIB (co-)author since 1998



## Some examples of collaboration

### Nationwide working groups

Under the lead of its Clinical Bioinformatics Group, SIB is fostering collaborative working in the health sector and in particular, it has set up two working groups on “somatic mutation calling” and “microbe typing and characterization”, which bring together bioinformaticians and clinical actors from all major **SWISS HOSPITALS**, with the aim of harmonizing routine next-generation sequencing (NGS) practices at the national level.

---

### Innovation@hospitals

The **GENEVA UNIVERSITY HOSPITALS (HUG)** and SIB’s Clinical Bioinformatics and Vital-IT Groups are jointly developing and upgrading OncoBench™, a sample-to-report platform used by the HUG Clinical Pathology Division for the management and analysis of NGS patient data in routine cancer diagnosis. The **INSTITUTE OF MEDICAL GENETICS OF ZURICH** has recently joined the partnership, to further collaborate on the development of a tool serving the specific needs of medical genetics.

---

### Award-winning platform to make data discoverable

The SourceData platform, developed by EMBO in collaboration with SIB’s Vital-IT Group, provides an intuitive interface for researchers and publishers alike, enabling them to share figures and their underlying data in a machine-readable, searchable format. In addition, SourceData provides a public interface where scientists can efficiently find and re-use published results.

http://sourcedata.embo.org/

**Converting ideas into applications**
The SIB Legal and Technology Transfer Office ensures seamless collaboration between SIB Groups, academic institutions and the private sector, with over 100 contracts signed each year.

## Some papers with high impact

The following SIB (co-)authored papers have been cited over 1,000 times since their publication date*

**STRING v10: Protein-protein interaction networks, integrated over the tree of life.**
By Szklarczyk D *et al.*
NUCLEIC ACIDS RES. 2015

---

**STRING v9.1: Protein-protein interaction networks, with increased coverage and integration.**
By Franceschini A *et al.*
NUCLEIC ACIDS RES. 2013

---

**The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored.**
By Szklarczyk D *et al.*
NUCLEIC ACIDS RES. 2011

---

**Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows.**
By Excoffier L and Lischer H
MOL ECOL RESOUR. 2010

---

**Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index.**
By Speliotes E *et al.*
NAT GENET. 2010

---

**Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP.**
By Hafner M *et al.*
CELL 2010

---

**STRING 8 – a global view on proteins and their functional interactions in 630 organisms.**
By Jensen L *et al.*
NUCLEIC ACIDS RES. 2009

---

**A mammalian microRNA expression atlas based on small RNA library sequencing.**
By Landgraf P *et al.*
CELL 2007

---

**SWISS-MODEL: An automated protein homology-modeling server.**
By Schwede T *et al.*
NUCLEIC ACIDS RES. 2003

---

**The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003.**
By Boeckmann B *et al.*
NUCLEIC ACIDS RES. 2003

---

**ExPASy: The proteomics server for in-depth protein knowledge and analysis.**
By Gasteiger E *et al.*
NUCLEIC ACIDS RES. 2003

* SOURCE: EUROPEPMC EXTRACTION FROM 30.03.1998 TO 17.01.2018. CITATIONS ARE BASED ON OPEN CITATIONS DATA.

# … and bioinformatics skills development

One of SIB's missions is to train the next generation of bioinformaticians and to ensure that life and health scientists make the best use of bioinformatics resources, many of which are developed by SIB Groups. The SIB Training team promotes and coordinates training in bioinformatics, both in Switzerland and internationally.

### PHD TRAINING NETWORK

The PhD Training Network was created as a community support for doctoral students carrying out their research in bioinformatics or computational biology in Switzerland. The Network aims to foster interactions and exchanges among SIB PhD students and train them in the most up-to-date methods necessary for their doctoral research. At the time of its creation in 2007, the Network had 19 PhD students in Lausanne, Basel and Geneva. Ten years later, it has seen close to 170 alumni defending their PhD thesis. Today, it has 240 members spread across eight Swiss universities.

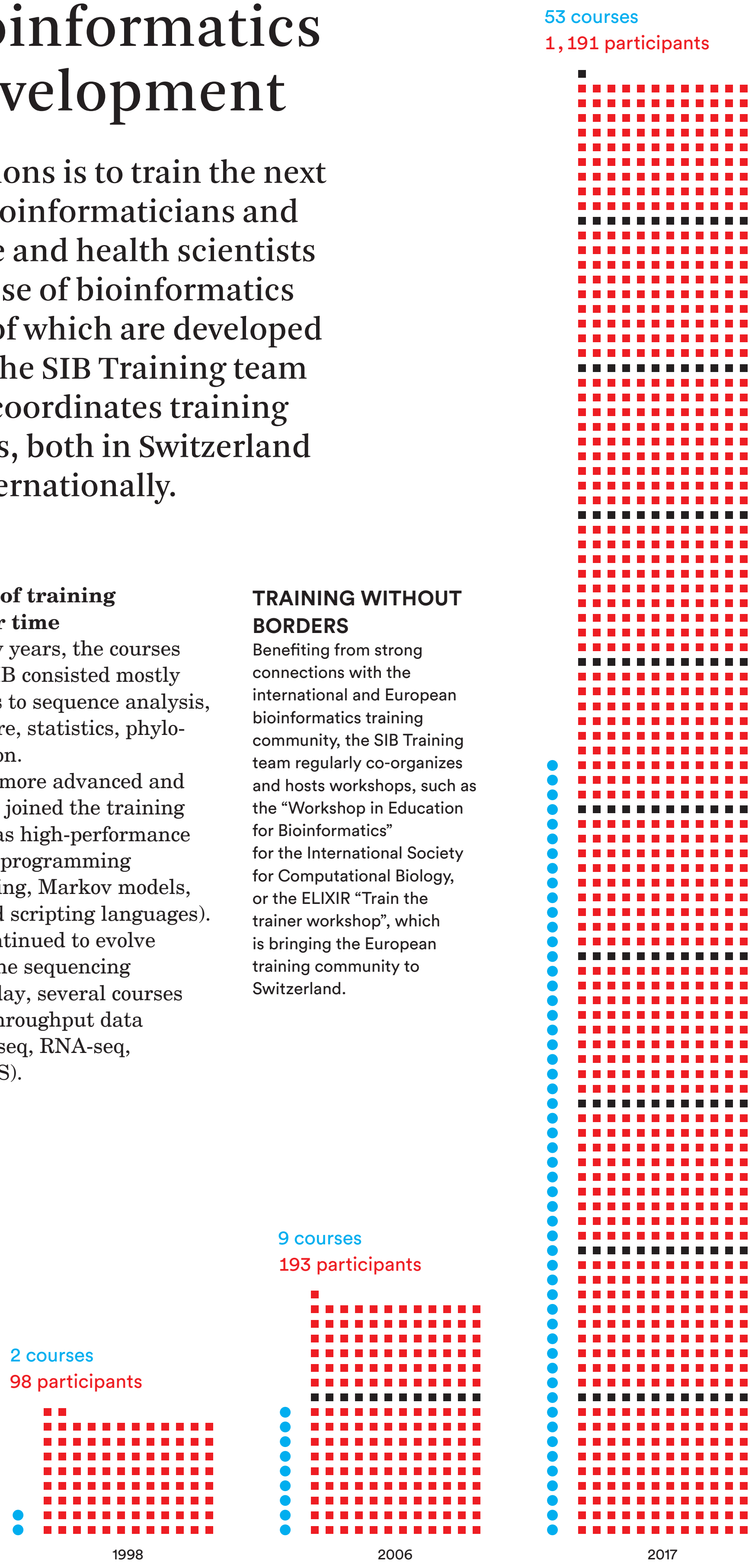
### Development of training activities over time

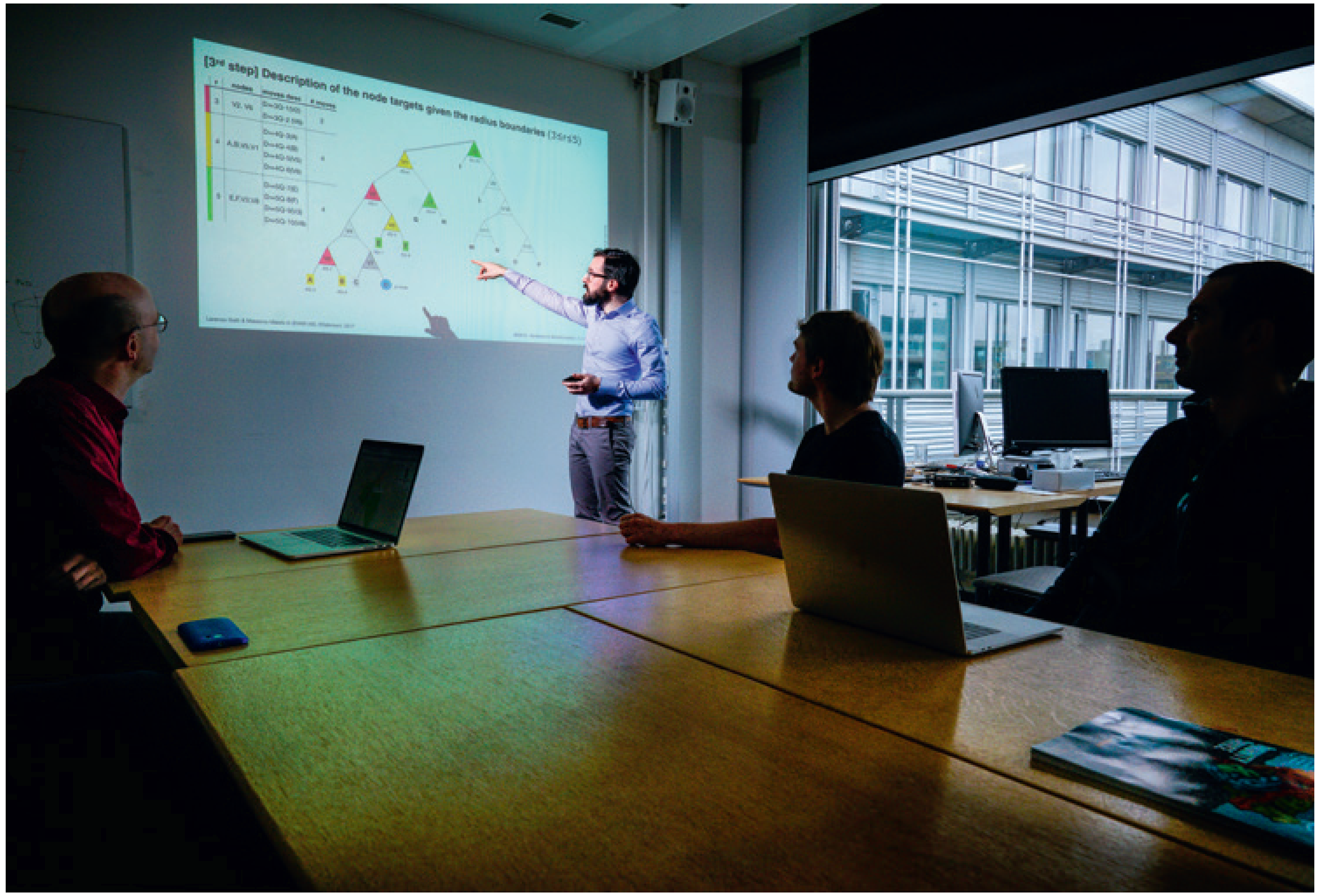
In the early years, the courses organized by SIB consisted mostly of introductions to sequence analysis, protein structure, statistics, phylogeny or evolution.

Over time, more advanced and technical topics joined the training portfolio, such as high-performance computing and programming (machine learning, Markov models, advanced R and scripting languages). The courses continued to evolve together with the sequencing techniques. Today, several courses focus on high-throughput data analysis (Chip-seq, RNA-seq, GWAS and NGS).

### TRAINING WITHOUT BORDERS

Benefiting from strong connections with the international and European bioinformatics training community, the SIB Training team regularly co-organizes and hosts workshops, such as the "Workshop in Education for Bioinformatics" for the International Society for Computational Biology, or the ELIXIR "Train the trainer workshop", which is bringing the European training community to Switzerland.





**What do the participants say?**

**96%**
consider that SIB courses are excellent or very good

**94%**
consider that SIB courses have the right balance between theoretical and practical content

**93%**
consider that SIB courses are relevant to their work

FIGURES FROM 2017

Some of the most taught skills

Best coding practices HPC Machine learning
ChIP-seq R software Metabolomics Protein databases
Statistics Long read sequencing
Experimental design Variant annotation Regular expressions
Next-generation sequencing RNA-Seq
UNIX Python Proteogenomics and immunopeptidomics
Computational biology
Computer-aided drug design Snakemake
Sensitive data management Perl 6

# Making open data sustainable

Open access has been a buzzword in the scientific community for some years now. SIB is committed to making this concept a reality for the information and data contained in its resources, thereby enabling science to make the most of bioinformatics discoveries. But the Institute is also concerned about the sustainability of open data in the long run.

**Christine Durinx, PhD**
SIB Associate Director

**Chiara Gabella, PhD**
SIB Project Manager

Open access does not mean free of cost, especially for databases that need significant human resources. With the current rate of data production, the question is a pressing one, and SIB has conducted a study to identify the most suitable funding model to ensure that open access to essential biological information can be maintained.

### What are core data resources for the life sciences?

**Christine Durinx** Imagine a centralized encyclopaedia of the function of hundreds of thousands of proteins, made accessible to life scientists throughout the world. It allows you to discover, for example, within a few minutes, a new form of an enzyme for the industrial production of biofuel. With 900,000 user requests per month, the UniProtKB/Swiss-Prot knowledgebase is the most widely used protein information database in the world. The database is manually curated by a team of SIB experts who facilitate data reuse and maximize the value of biological data. This is an example of what is referred to as a core data resource on the international scene. Such resources are of fundamental importance to the broad life science community and the long-term preservation of biological data. They provide complete collections of generic value to life science, and show high levels of usage, scientific quality and service.

### What is the threat they are facing?

**CD** Today's life science research cannot be conducted without data resources. The issue is that most open-access databases are currently financed by research grants and institutional funds. If a database applies for a research grant, it will compete against research projects. Moreover, such grants are cyclical, so they are not stable over time.

> **"The situation is actually quite dramatic: a recent survey revealed that two-thirds of databases have one year or less of guaranteed funding."**

The situation is actually quite dramatic: a recent survey revealed that two-thirds of databases have one year or less of guaranteed funding.

We should bear in mind that scientists have entrusted their research results to these databases. So the fact that their existence is so unstable could result in the loss of an immense wealth of biological information, and the associated investments.

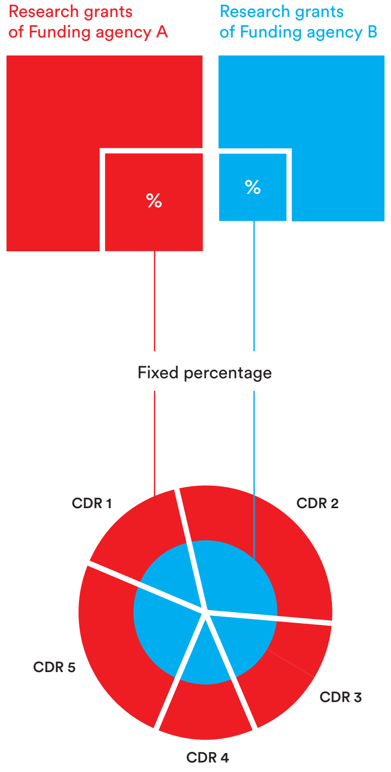
### What is the study about and what is the outcome?

**Chiara Gabella** In this study, we analysed 12 funding models for databases, and applied them to the UniProt knowledgebase.

The objective was to find the ideal funding model, by comparing the pros and cons of each model, for example, in terms of open access, equity between users, the potential to generate sufficient income, or the stability of income over time.

We have selected what we have called the "Infrastructure model" as a sustainable funding scheme that can be applied to other core data resources in the life sciences.

## The Infrastructure model - a sustainable funding scheme for core data resources (CDR)



Core data resources' funding allocation

### Can you give us some more details about this model?

**CG** Under the Infrastructure model, funding agencies would set aside a fixed percentage of their research grants. The amount would subsequently be re-distributed to core data resources according to well-defined selection criteria (SEE ABOVE). Based on existing data, we have been able to estimate that less than 1% of the total amount dedicated to research grants in the life sciences would be sufficient to cover the costs of the core data resources worldwide.

Of course, this model requires major changes in how funds are now distributed and necessitates collaboration at an international level.

### Why is SIB preoccupied with this issue?

**CD** Caring for sustainability really is in our Institute's DNA and raison d'être. SIB was founded following a funding crisis, to ensure the long-term sustainability of the Swiss-Prot database, which is now part of the UniProt consortium. Twenty years after SIB's creation, the topic of the long-term sustainability of its knowledgebases is still a hot topic.

SIB is working closely with ELIXIR, an intergovernmental organization that is building a sustainable infrastructure for biological information across Europe. This study was done in the context of the ELIXIR Data Platform and benefited from ELIXIR's support. SIB also plays an active role in an international coalition that is working on a mechanism to provide global support for core data resources. Joining forces globally is the key to the long-term sustainability of knowledge preservation. ●

**A blog to go further**
Read the discussion between C. Gabella and reviewer Helen Berman, Rutgers, The State University of New Jersey
https://blog.f1000.com/2018/02/07/how-best-to-fund-knowledgebases/

**Reference**
Funding knowledgebases: Towards a sustainable funding model for the UniProt use case
By Gabella C *et al.*
*F1000Research* 2018, 6(ELIXIR):2051
[version 2; referees: 3 approved]

Watch this interview online on our Youtube channel:
goo.gl/GWmZSV

**Expert curation is sustainable: the example of UniProtKB/Swiss-Prot**
While the importance of knowledgebases is increasingly recognized, the question of their sustainability is frequently raised. How can a tight number of curators keep up with the rapid growth in biomedical literature, i.e. currently over one million papers published a year? A study led by SIB's Swiss-Prot Group reveals that in fact, no more than 2-3% of all publications indexed in PubMed each year appear to be relevant for UniProt curation. The sheer volume of biomedical papers published each year is therefore a poor way to evaluate biocuration effort.
These results demonstrate, for one thing, that expert curation in UniProt can keep up with the increasing number of biomedical publications. They also underline the fact that, now more than ever, curators play a crucial role in shedding light on biological knowledge, to support researchers in identifying the most relevant and substantiated evidence.

On expert curation and scalability: UniProtKB/Swiss-Prot as a case study
By Poux S *et al.*
Bioinformatics, 2017

# Two decades of SIB

## 1986

### BIRTH OF SWISS-PROT

Amos Bairoch, then a PhD student at the University of Geneva, creates what has since become the world reference resource for protein knowledge, UniProtKB/Swiss-Prot. At that time, the various versions are distributed on magnetic tapes!

## 1993

### CREATION OF THE FIRST LIFE SCIENCE WEB PORTAL

To exchange information within the community, Ron Appel launches the first life science server, which is one of the 150 first websites in the world. It is accessed 7,295 times even during its first month of existence, and over 2 million times per month a few years later.

## 1996

### A FUNDING CRISIS LEADING TO SIB'S CREATION IN 1998

The Swiss-Prot knowledgebase, which by then has around 100,000 users, is threatened with extinction due to a lack of financial support. Thousands of support messages from the international scientific community flow in. As a result, the Swiss authorities recommend that a stable long-term funding mechanism be sought to ensure the survival of Swiss-Prot and bioinformatics, in the form of a publicly funded foundation.

## 1998

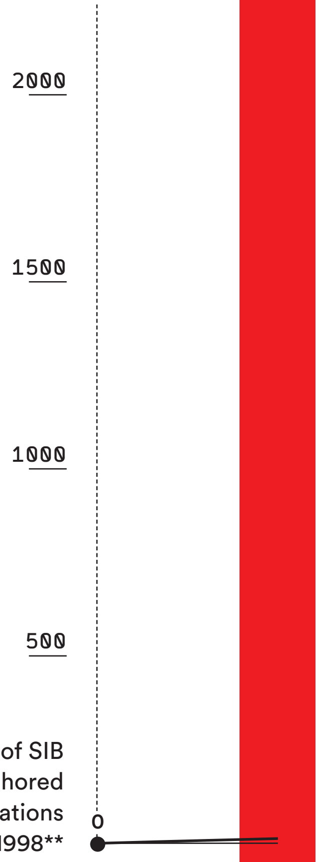
### CREATION OF THE SIB SWISS INSTITUTE OF BIOINFORMATICS

A team of visionary Swiss scientists (Ron Appel, Amos Bairoch, Philipp Bucher, Victor Jongeneel and Manuel Peitsch) unite to create one sole bioinformatics organization. It includes five founding groups from the Universities of Geneva and Lausanne, the Geneva University Hospitals (HUG), the Ludwig Institute for Cancer Research, the Institut Suisse de Recherche Expérimentale sur le Cancer (ISREC) and Glaxo Wellcome Experimental Research.*

**5**
groups

**20**
members

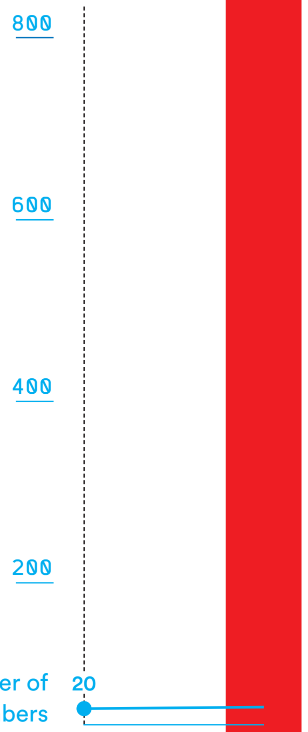
**6**
institutional partners



Number of SIB (co-)authored publications since 1998**

### CREATION OF GENEBIO

Created to facilitate fund matching between industries and the Swiss government, GeneBio, SIB's commercial entity, was first focused on protein identification and characterization as well as proteome imaging.



Number of SIB Members

---

* The HUG remained an institutional partner of SIB until 2006, before joining again in 2016; ISREC remained until 2008 when it was integrated into EPFL, and Glaxo Wellcome Experimental Research until 2002.

** SOURCE: EUROPEPMC EXTRACTION FROM 30.03.1998 TO 17.01.2018

## 2018

### 20 YEARS OF SIB!

A year of festivities and projects to celebrate 20 years of commitment to life sciences and health.

### CERTIFICATE OF ADVANCED STUDY IN PERSONALIZED MOLECULAR ONCOLOGY

First of its kind in Switzerland, the CAS is launched jointly by the University Hospitals of Basel and Lausanne and SIB.

**New Partner**

### SWISS INSTITUTE OF ALLERGY AND ASTHMA RESEARCH (SIAF), DAVOS

**67**
groups

some
**800**
members

**20**
institutional partners

### A WEB SERIES, A MOBILE GAME AND A BOOK...

To mark our 20th anniversary in a surprising way, different projects were launched to promote bioinformatics within and outside the research community during the year:

Share the everyday life and challenges of a bioinformatician in a 7-episode humorous web series;

With the mobile game "Genome Jumper", explore our genetic variability and why we have curly or straight hair, or lack the capability to digest milk;

Discover the colourful and surprising side of SIB's Group Leaders, captured by Swiss photographer Nicolas Righetti, while learning about the history of bioinformatics in a science and art book.

Stay tuned for the release of these projects on **www.sib.swiss/20years**

**2,238th**
publication with an SIB (co-)author, including 54 in *Nature* or *Science*

## 1999

### FIRST POST-GRADUATE DIPLOMA IN BIOINFORMATICS

Created by SIB together with the Universities of Geneva and Lausanne

## 2000

20th publication with an SIB (co-)author

### LAUNCH OF *PROTEIN SPOTLIGHT*

The online popular science magazine, launched by SIB's Swiss-Prot Group and written by Vivienne Baillie Gerritsen, science writer, tells the tale of one protein each month.

## 2001

New Partner

### UNIVERSITY OF BASEL

## 2002

New Partner

### EPFL, LAUSANNE

### CREATION OF THE UNIPROT CONSORTIUM

SIB, the EMBL-European Bioinformatics Institute (EMBL-EBI) in the UK and the Protein Information Resource (PIR) in the US unite to jointly maintain and develop the most widely used protein information resource in the world, which includes the Swiss-Prot database.

### LAUNCH OF THE FIRST SWISS PHD PROGRAMME IN BIOINFORMATICS

Created by SIB together with the University of Geneva

## 2003

### FIRST [BC]$^2$ CONFERENCE "LIFE SCIENCES MEET IT"

SIB, the Biozentrum of the University of Basel and the Friedrich Miescher Institute launch the first Basel Computational Biology conference, to foster collaborations between academic, biomedical and pharmaceutical research institutes and to share knowledge, resources and services in the field of computational biology and bioinformatics.

### CREATION OF THE FIRST SWISS BIOINFORMATICS CORE FACILITY

Launch of the high-performance computing centre Vital-IT by SIB, the Universities of Geneva and Lausanne as well as EPFL, in cooperation with Hewlett-Packard, Intel Corporation and Oracle.

**Key articles***

"SWISS-MODEL: An automated protein homology-modeling server" by Schwede T *et al.* in Nucleic Acids Res.

"The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003" by Boeckmann B *et al.* in Nucleic Acids Res.

"ExPASy: The proteomics server for in-depth protein knowledge and analysis" by Gasteiger E *et al.* in Nucleic Acids Res.

## 2004

### FIRST SIB DAYS IN LEYSIN

The first edition of SIB's internal scientific conference at the foot of the snowy slopes of Leysin

## 2005

New Partner

### ETH ZURICH

## 2006

## 2007

New Partners

### UNIVERSITY OF ZURICH

### FRIEDRICH MIESCHER INSTITUTE FOR BIOMEDICAL RESEARCH, BASEL

**Key article***

"A mammalian microRNA expression atlas based on small RNA library sequencing" by Landgraf P *et al.* in Cell

### CREATION OF THE SIB PHD TRAINING NETWORK

The network includes 19 PhD students in Lausanne, Basel and Geneva.

## 2008

New Partner

### UNIVERSITY OF BERN

### SIB'S 10TH ANNIVERSARY

To mark the occasion, SIB launches the SIB Young Bioinformatician Award to recognize talented early career scientists.

272th publication with an SIB (co-)author

**24** groups

**300** members

**9** institutional partners

### FULLY CURATED HUMAN PROTEOME

UniProtKB/Swiss-Prot releases the first version of the fully curated complete human proteome, i.e approximately 20,000 putative human protein-coding genes.

## 2009

**Key article***

"STRING 8-a global view on proteins and their functional interactions in 630 organisms" by Jensen L *et al.* in Nucleic Acids Res.

## 2010

**Key articles***

"Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows" by Excoffier L & Lischer H in Mol Ecol Resour.

"Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index" by Speliotes E *et al.* in Nat Genet.

"Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP" by Hafner M *et al.* in Cell.

## 2011

New Partner

### UNIVERSITY OF FRIBOURG

**Key article***

"The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored" by Szklarczyk D *et al.* in Nucleic Acids Res.

## 2012

New Partner

### UNIVERSITÀ DELLA SVIZZERA ITALIANA, LUGANO

### CREATION OF THE CLINICAL BIOINFORMATICS GROUP

The group aims to provide health professionals with expertise and support for the organization, analysis and interpretation of patient molecular data, so that they may be used for diagnostic purpose.

### A NON-INVASIVE PRENATAL DIAGNOSTIC TEST

Vital-IT develops the algorithms behind a test that allows the detection of chromosomal abnormalities in the fœtus from the mother's blood - the test is the first of its kind in Europe.

## 2013

New Partner

### HEG-GE, GENEVA

### CREATION OF ELIXIR

Switzerland signs the Memorandum of Understanding to catalyse the creation of ELIXIR, the European infrastructure organization whose aim is to guarantee the future of biological data.

**Key article***

"STRING v9.1: protein-protein interaction networks, with increased coverage and integration" by Franceschini A *et al.* in Nucleic Acids Res.

### CHROMOSOMEWALK, THE WEBSITE

A virtual, award-winning online exhibition introducing the human genome and bioinformatics to the public.

## 2014

New Partners

### SWISS TROPICAL AND PUBLIC HEALTH INSTITUTE, BASEL

### AGROSCOPE, WÄDENSWILL

### CREATION OF THE SIB TRAINING GROUP

The group integrates and harmonizes existing initiatives, and provides a more comprehensive and complete training portfolio. From this moment on, over 50 courses are organized per year.

## 2015

New Partners

### INSTITUTE OF ONCOLOGY RESEARCH, BELLINZONA

### ZURICH UNIVERSITY OF APPLIED SCIENCE (ZHAW)

### SIB BECOMES FAO REFERENCE CENTRE FOR BIOINFORMATICS

The Food and Agriculture Organization of the United Nations (FAO) is thus recognizing SIB's expertise and state-of-the-art services, for example through open-access databases such as Viralzone, maintained by the SIB Swiss-Prot Group.

**Key article***

"STRING v10: protein-protein interaction networks, integrated over the tree of life" by Szklarczyk D *et al.* in Nucleic Acids Res.

### THE DRUG DESIGN WORKSHOP

SIB launches a web-based platform to introduce non-specialists to computer assisted drug design and bioinformatics.

### FIRST CLINICAL BIOINFORMATICS WORKING GROUP

Bringing together clinical and data science experts nationwide, the *Somatic mutation calling* working group aims to benchmark next-generation sequencing practices in Swiss hospitals and to define common guidelines for the analysis and interpretation of genetic variants in cancer diagnosis.

## 2016

New Partners

### HEIG-VD, YVERDON-LES-BAINS

### UNIVERSITY HOSPITALS OF GENEVA (HUG)

### ESPERARE, GENEVA

### LAUNCH OF A CANCER DIAGNOSTIC PLATFORM

The Geneva University Hospitals (HUG) and SIB develop OncoBench™, a sample-to-report tool used by HUG for faster routine cancer diagnosis.

### SIB TAKES PART IN A GLOBAL COALITION TO SUSTAIN CORE DATA RESOURCES

The coalition, which gathers representatives from funding organizations, key core data resources, professional societies and business leaders, is setting out to ensure the sustainability of data resources of crucial importance for the life sciences.

### 30TH BIRTHDAY FOR TWO SIB CORE RESOURCES

UniProtKB/Swiss-Prot, the protein knowledgebase and EPD, the eukaryotic promoter database, celebrate their anniversary with dedicated events and symposiums.

## 2017

### SIB IS RANKED 14TH/132 MOST INNOVATIVE INSTITUTIONS IN SWITZERLAND

The Nature Index Innovation reflects the use of references to academic literature in patents to show concrete links between discovery and its economic potential.

### FIRST SIB BIOINFORMATICS HACKATHONS

Organized as part of the Open Geneva Hackathons, children, developers and structural biologists tackle biological problems using programming and augmented reality apps.

### LAUNCH OF THE SWISS PERSONALIZED HEALTH NETWORK (SPHN) INITIATIVE

SIB plays a pivotal role as Data Coordination Centre (DCC) and as the provider of a secure infrastructure for biomedical research in Switzerland (BioMedIT).

### TWO SIB RESOURCES ESSENTIAL FOR THE LIFE SCIENCES

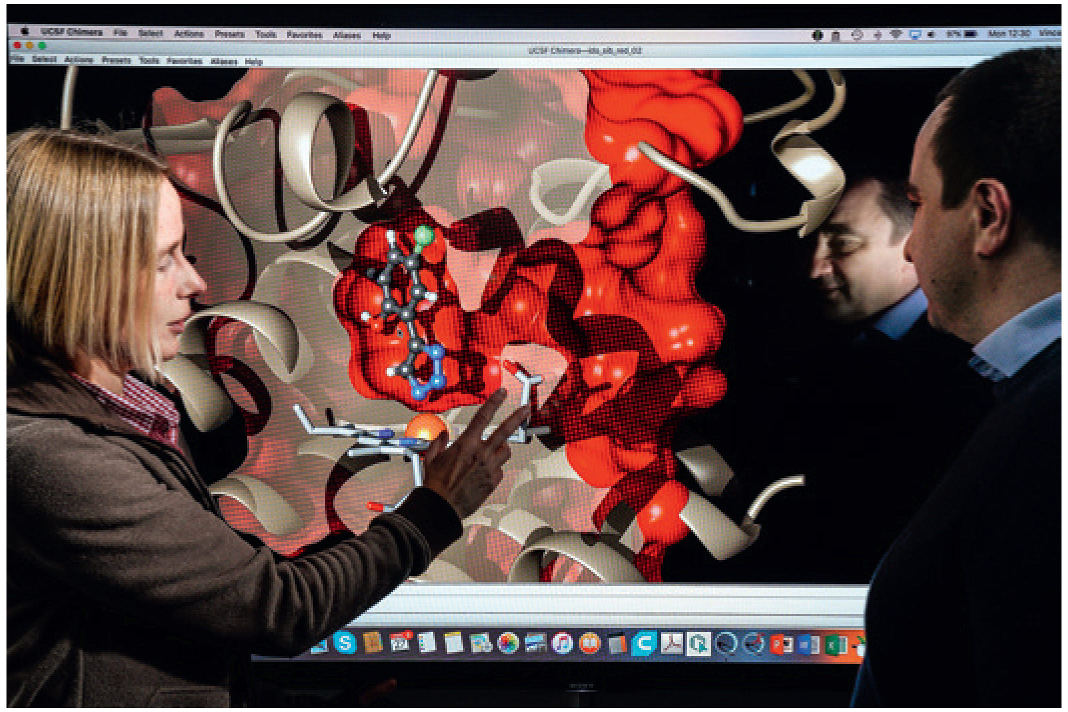
UniProt and STRING, two of SIB's core data resources, are singled out by ELIXIR as of crucial importance for the life sciences.

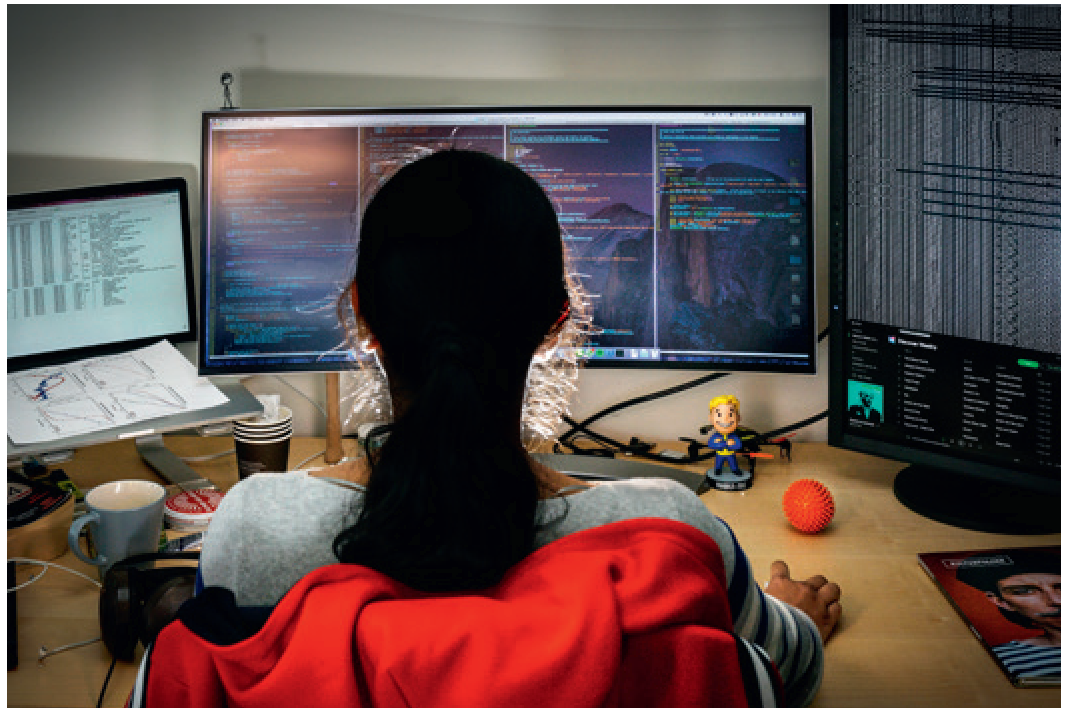
---

* SIB (CO-)AUTHORED PAPER CITED OVER 1000 TIMES SINCE ITS PUBLICATION, SOURCE: EUROPEPMC EXTRACTION FROM 30.03.1998 TO 17.01.2018. CITATIONS ARE BASED ON OPEN CITATIONS DATA.

CHAPTER 3

# Activities 2017





SIB's activities revolve
around five pillars:

**Infrastructure**

I.
DATABASES
AND SOFTWARE TOOLS

II.
CORE FACILITIES
AND COMPETENCE CENTRES

III.
PERSONALIZED HEALTH

**Community**

IV.
TRAINING IN
BIOINFORMATICS

V.
SCIENTIFIC
COLLABORATION

Discover our latest developments and achievements on each of these fronts in the following pages.

# Providing life scientists with a complete bioinformatics infrastructure

## I.
## DATABASES AND SOFTWARE TOOLS

Most of the resources developed by SIB Groups are available in open access via the SIB bioinformatics resource portal ExPASy (www.expasy.org) and cover a broad range of domains. Here are a few examples:

| DOMAIN | SUB-DOMAIN | DATABASE | SOFTWARE TOOL |
|---|---|---|---|
| **Genes and genomes** | Sequence alignment | | LALIGN, Newick Utilities, T-Coffee |
| | Similarity search | | BLAST, BUSCO, LALIGN, Phylogibbs |
| | Characterization/ annotation | EPD, miROrtho, OMA, OpenFlu, OrthoDB, SwissRegulon | ChIP-Seq, ISA, PACMAN, QuasR, V-pipe |
| | Transcriptomics | Bgee, CleanEx, SwissRegulon | ISMARA, PPA, SNP2TFBS, TopAnat |
| **Proteins and proteomes** | Protein sequences and identification | neXtProt, UniProtKB/ Swiss-Prot, ViralZone | HAMAP, PeptideMass, Terminus, Translate |
| | Mass spectrometry and 2-DE data | SWISS-2D PAGE, WORLD-2D PAGE Repository | FindPept, GlycoMod, MSight |
| | Protein characterization and function | ENZYME, neXtProt, iPtgxDBs, Rhea, Swiss-Lipids, UniProtKB/Swiss-Prot | AACompSim, Biochemical Pathways, ProtScale |
| | Families, patterns and profiles | MyHits, PROSITE | MyDomains, pftools, PRATT |
| | Post-translational modification | SugarBind, UniCarb-DB, UniProtKB/Swiss-Prot | FindMod, GlycanMass, ISMARA |
| | Protein-protein interaction | STRING, UniProtKB/Swiss-Prot | |
| | Similarity search/ alignment | UniProtKB/Swiss-Prot | BLAST, ClustalW, MyHits |
| | Imaging | | ImageMaster / Melanie, MSIght |
| **Evolution and phylogeny** | | Bgee, ImmunoDB, miROrtho, Swiss Orthology (OMA, OrthoDB), Selectome | |
| **Structural biology** | | SWISS-MODEL Repository, SwissSideChain, SwissBioIsostere | CAMEO, Click2Drug, SWISS-MODEL Workspace, Swiss-PdbViewer, SwissADME, SwissDock, SwissParam, SwissSimilarity, SwissTargetPrediction |
| **Systems biology** | | Cellosaurus, Progenetix, SwissRegulon | arrayMap, CRUNCH, MetaNetX, The Systems Biology Research Tools |

## SIB core resources

### A stringent process to ensure the highest level of excellence

SIB's External Scientific Advisory Board (SAB) is in charge of providing recommendations on the selection of SIB's core resources every four years. Assessing competitive criteria, including the scientific impact and return-on-investment, along with SIB's global portfolio management and strategic orientation, SAB members select best-in-class resources and provide guidelines for their continuous development. Once validated by the Board of Directors, the SIB management team then assists resource managers in implementing their action plan and exchanging best practices.

**In 2017, SIB core resources were used by about six million users worldwide**

Source: Google Analytics

## Proteins & proteomes

### UniProtKB/Swiss-Prot

**Protein knowledgebase**

TYPE Knowledgebase with manual expert curation
DESCRIPTION Hundreds of thousands of protein descriptions, including function, domain structure, subcellular location, post-translational modifications and functionally characterized variants.
HIGHLIGHT UniProtKB/Swiss-Prot is the expert curated part of UniProt, the most widely used protein information resource in the world, with over six million page views per month. Selected as an ELIXIR core resource.
SEE P. 59

### neXtProt

**Human protein knowledgebase**

TYPE Knowledgebase with manual expert curation
DESCRIPTION Information on human protein biology such as function, involvement in diseases, mRNA/protein expression, protein/protein interactions, post-translational modifications, protein variations and their phenotypic effects.
HIGHLIGHT Advanced search options combining data at genomic, transcriptomic, proteomic and phenotypic levels.
SEE P. 58

### Glycomics@ExPASy

**Glyco(proteo)mics resource portal**

TYPE Web tools and knowledgebases with manual expert curation
DESCRIPTION Extensive set of resources on sugar structures (UniCarb-DB), their relationships with the proteins that carry them (GlyConnect) and the proteins that bind them, with a special focus on interactions between pathogen proteins and mammalian glycans (SugarBind).
HIGHLIGHT Unique portal for glyco(proteo)mics data analysis and interpretation.
SEE P. 59

### STRING

**Protein-protein interactions**

TYPE Knowledgebase
DESCRIPTION Resource for known and predicted protein-protein interactions, including direct (physical) and indirect (functional) associations derived from various sources, such as genomic context, high-throughput experiments, (conserved) co-expression and the literature.
HIGHLIGHT Currently includes 9,643,763 proteins from 2,031 organisms and is the most used resource of its kind. Selected as an ELIXIR core resource.
SEE P. 59

### SWISS-MODEL

**Protein structure homology-modelling**

TYPE Software tool and repository
DESCRIPTION Automated protein structure homology-modelling platform for generating 3D models of a protein using a comparative approach, and database of annotated models for key reference proteomes based on UniProtKB.
HIGHLIGHT Easy-to-use web-based platform processing over one million model requests per year and also providing model information for non-specialists. Celebrates its 25th birthday in 2018!
SEE P. 64

## Structural biology

### SwissDrugDesign

**Drug design**

**TYPE** Software tools
**DESCRIPTION** Web-based computer-aided drug design tools, from molecular docking (SwissDock) to pharmacokinetics and druglikeness (SwissADME), through virtual screening (SwissSimilarity) and target prediction of small molecules (SwissTargetPrediction).
**HIGHLIGHT** Comprehensive and integrated web-based drug design environment.
SEE P. 63

## Lipids

### SwissLipids NEW

**A knowledge resource for lipids**

**TYPE** Knowledgebase
**DESCRIPTION** Information about known lipids, including knowledge of lipid structures, metabolism, interactions, and subcellular and tissular localization, providing a framework for the integration of lipid and lipidomic data with biological knowledge and models.
**HIGHLIGHT** Contains information on over 400,000 lipid structures from over 180 major lipid classes.
SEE P. 59

## Genes and genomes

### Bgee

**Gene expression evolution**

**TYPE** Knowledgebase with manual expert curation
**DESCRIPTION** Information on gene expression evolution (incl. all types of transcriptomes), allowing retrieval and comparison of expression patterns between animals including human, model organisms and diverse species of evolutionary or agronomical relevance.
**HIGHLIGHT** Only resource to provide homologous gene expression between species.
SEE P. 61

### SwissRegulon Portal

**Tools and data for regulatory genomics**

**TYPE** Software tools and knowledgebases
**DESCRIPTION** Web portal for regulatory genomics, including genome-wide annotations of regulatory sites and motifs, the webserver ISMARA for automated inference of regulatory networks and CRUNCH for automated analysis of ChIP-seq data.
**HIGHLIGHT** User can upload raw micro-array, RNA-seq or ChIP-seq data to automatically infer the core regulatory networks acting in the system of interest.
SEE P. 57

Swiss Orthology

### OrthoDB

**The hierarchical catalogue of orthologs**

**TYPE** Automated phylogenomic database and software tool
**DESCRIPTION** Comprehensive online catalogue of orthologs, including functional and evolutionary gene annotations and enabling the inference of putative gene functions.
**HIGHLIGHT** Largest orthology resource, enabling users to generate publication-quality comparative genomics charts, and to upload, analyse and interactively explore their own pre-publication data.
SEE P. 58

### V-Pipe NEW

**Viral genomics pipeline**

**TYPE** Software tool
**DESCRIPTION** Pipeline integrating existing and in-house pieces of software, designed for the clinical application of next-generation sequencing to viral pathogens.
**HIGHLIGHT** Enabling reliable and comparable viral genomics and epidemiological studies and facilitating clinical diagnostics of viruses.
SEE P. 52

### EPD

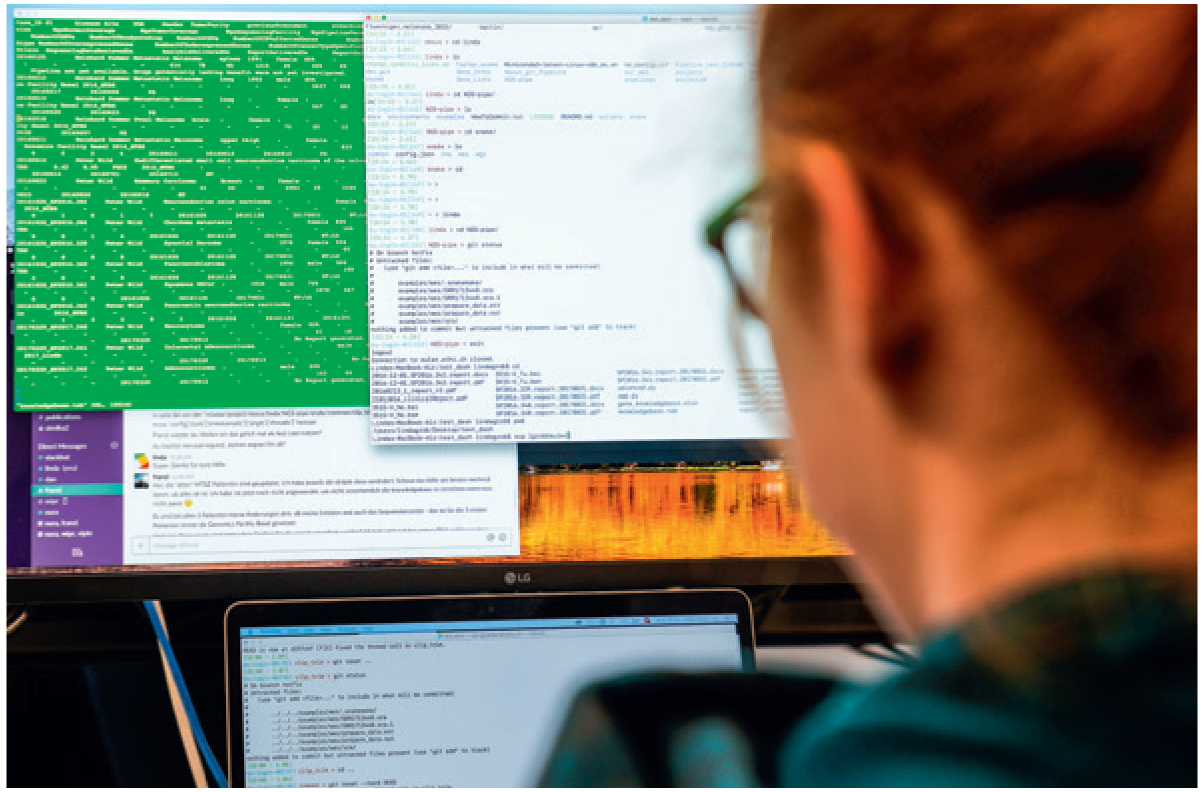
**Eukaryotic Promoter Database**

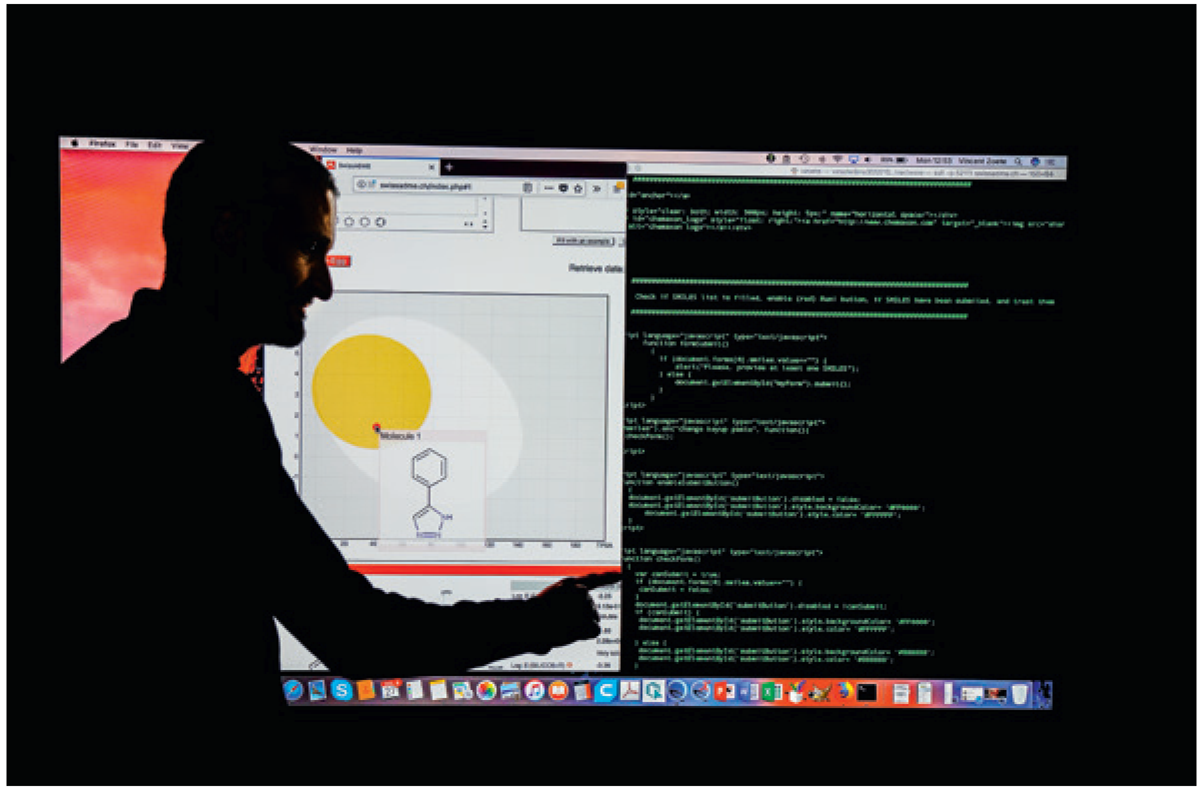
**TYPE** Knowledgebase with manual expert curation and software tools
**DESCRIPTION** Quality-controlled information on experimentally defined promoters of higher organisms, as well as web-based tools for promoter analysis.
SEE P. 53

### OMA

**Orthology MAtrix browser**

**TYPE** Automated phylogenomic database and software tool
**DESCRIPTION** High-quality orthology predictions among complete genomes.
**HIGHLIGHT** Broad scope and size, feature-rich web interface, availability in a wide range of formats and interfaces, frequent update schedule.
SEE P. 60





**Top**
Programming is at the heart of many of our bioinformaticians' activity. It enables them to analyse data more swiftly or to develop softwares...

**Bottom**
...such as computer-aided drug design tools, which in turn enable life scientists to create the drugs of tomorrow.

## II.
## CORE FACILITIES AND COMPETENCE CENTRES

Swiss life scientists with computational needs can count on a tight mesh of expert hubs in Switzerland whether they need access to high-performance computing resources, software development services, cutting-edge multidisciplinary bioinformatics expertise, specialized training – or all these at once.

### Competence centres for life scientists

Vital-IT, an SIB-staffed group, started off as a small-scale high-performance computing centre in 2007. Now covering Western Switzerland and comprising a multidisciplinary team of scientists and technical staff, the group is involved in a wide diversity of projects on a national and international scale.

**IOANNIS XENARIOS**
SIB's Swiss-Prot and Vital-IT Group Leader
Professor at the Centre for Integrative Genomics, University of Lausanne and at the University of Geneva

**What is the added value of competence centres such as Vital-IT for life scientists?**

**Ioannis Xenarios** Scientists have several options for keeping up with the increase in the amount and diversity of data generated today.

Some groups – originally 'experimentalists at heart' – have, for example, been able to successfully expand their competencies by hiring computational biologists, bioinformaticians or IT staff internally.

Those who do not have the 'luxury' of having a bioinformatician in their own premises can turn to competence centres, which can address a diverse range of needs. In addition to taking the computational workload off the shoulders of the research groups, competence centres such as Vital-IT also provide software, storage and long-term data maintenance.

If these tasks were to be performed by the individual groups themselves, it would multiply the costs for their respective institutions. Competence centres thus offer an economy of scale and a coordinated environment, where expertise is shared amongst their members.

**What sort of activities do you typically conduct?**

**IX** Depending on the needs of the groups, we provide three types of activities. If the need is purely related to computing power or storage, we provide the necessary IT infrastructure. For groups that do not have the means to actually hire a bioinformatician, we also provide up to a dozen different types of expertise, according to the project. Interdisciplinarity is the key word here.

> **"Competence centres thus offer an economy of scale and a coordinated environment, where expertise is shared amongst their members."**

Finally, for scientists and groups that are not quite there yet in terms of generating their own data, but know that there are data out there and want to 'probe' the field, we help them to identify whether the investment and move towards a computational approach is worth it. This is therefore more a consulting and strategic counselling type of activity.

**Since the beginning of 2017, the Swiss National Science Foundation (SNSF) allows researchers to include the costs of infrastructure usage for their projects in their grant application. What does this move represent, in your view?**

**IX** Overall, the decision of the SNSF – and nearly all European funding agencies – is emphasizing the fact that, without provision for access to computational infrastructure and bioinformatics expertise, life scientists would not be able to function in the 21st century. However, the road to making all generated data compliant with Findable, Accessible, Interoperable and Re-usable ('FAIR') principles has a cost that is often underestimated. When writing this part of their grant application, life scientists should, for example, include not only the data generation aspects, but also the data and resource maintenance aspects beyond the end of the project, as these are likely to exceed the duration of their project grant by a long way. This is even more important in the light of the reproducible science arena, where this cost is rarely taken into account.

While the SNSF's decision is a first step in the right direction, competence centres will play a crucial role in sensitizing and training scientists with whom they collaborate on the importance of data management – a role Vital-IT has been playing for the last 15 years. ●

**Top**
From high-performance computing resources…

**Bottom**
…to multidisciplinary expertise, gathered under a single roof: Swiss bioinformatics core facilities and competence centres offer a match to every need from life scientists.

**How does it work at… the DBM Bioinformatics Core Facility?**

"We provide a special type of scientific services, which focus on the complete analysis of next-generation sequencing data – or indeed any other type of high-throughput data, such as flow or mass cytometry. Each of these projects starts with a discussion with the users to define the appropriate experimental design (number of replicates, selection of technology or protocol). Once the data are generated, we perform an initial analysis and meet the users again to discuss the results. Most often there are new questions or additional analyses. Such projects typically last for months – sometimes even years."

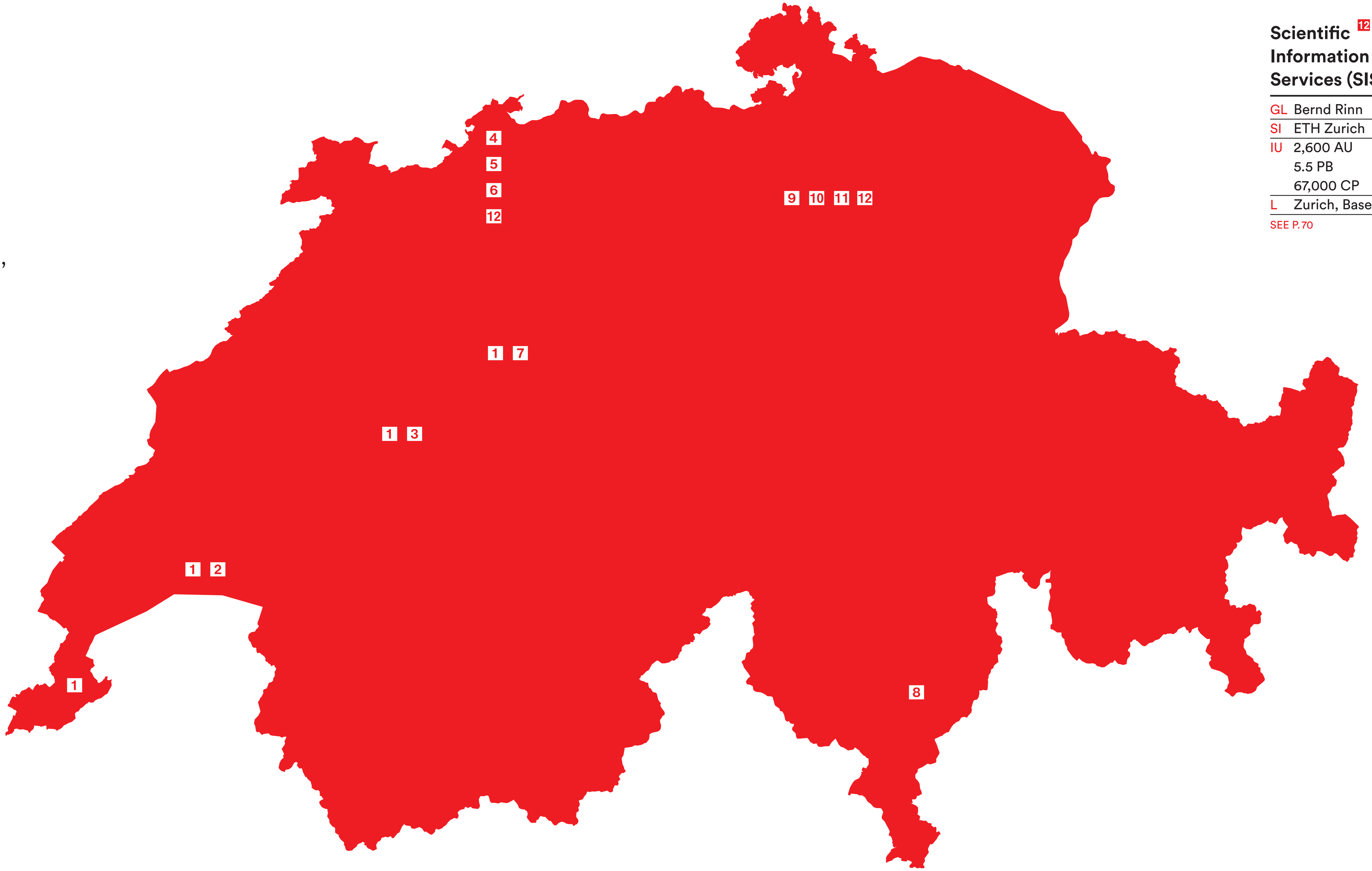
**ROBERT IVANEK**
SIB Group Leader,
DBM Bioinformatics Core Facility

# A national network to support the computational needs of life scientists

## A 2017 snapshot

Twelve core facilities and competence centres are distributed in all the major Swiss academic institutions, and coordinated by SIB.

GL = Group leader(s)
SI = Supporting institution(s)
IU = Infrastructure usage, when applicable
  *AU* Active users
  *PB* Petabytes of storage capacity
  *CP* Core processing units
L = Location(s)



### Vital-IT 1

GL Ioannis Xenarios
SI SIB, EPFL,
University of Lausanne,
University of Geneva,
University of Bern,
University of Fribourg
IU 1,200 AU
8 PB
9,900 CP
L Lausanne, Geneva, Bern,
Fribourg

SEE P. 71

### Bioinformatics Core Facility (BCF) 2

GL Mauro Delorenzi
and Frédéric Schütz
SI SIB, University
of Lausanne
L Lausanne

SEE P. 53

### Bioinformatics Unravelling Group (BUGFri) 3

GL Laurent Falquet
SI University
of Fribourg
L Fribourg

SEE P. 69

### FMI Computational Biology Group (FMICBG) 4

GL Michael Stadler
SI Friedrich Miescher
Institute
IU 170 AU
1 PB
144 CP
L Basel

SEE P. 56

### DBM Bioinformatics Core Facility 5

GL Robert Ivanek
SI University of Basel,
University
Hospital Basel
L Basel

SEE P. 69

### Scientific Information Services (SIS) 12

GL Bernd Rinn
SI ETH Zurich
IU 2,600 AU
5.5 PB
67,000 CP
L Zurich, Basel

SEE P. 70

### Service and Support for Science IT (S3IT) facility 11

GL Marcel Riedi
SI University of Zurich
IU 1,000 AU
7 PB
7,000 CP
L Zurich

SEE P. 70

### NEXUS Personalized Health Technologies 10

GL Daniel Stekhoven
SI ETH Zurich
IU 50 AU
0.2 PB
192 CP
L Zurich

SEE P. 71

### FGCZ Genome Informatics (FGCZ-GI) 9

GL Hubert Rehrauer
SI ETH Zurich,
University of Zurich
IU 1,081 AU
0.3 PB
480 CP
L Zurich

SEE P. 69

### Centre for Scientific Computing (sciCORE) 6

GL Torsten Schwede
and Thierry Sengstag
SI SIB, University of Basel
IU 750 AU
5 PB
6,500 CP
L Basel

SEE P. 70

### Interfaculty Bioinformatics Unit (IBU) 7

GL Rémy Bruggmann
SI University of Bern
IU 100 AU
2 PB
600 CP
L Bern

SEE P. 68

### Bioinformatics Core Unit (BCU) 8

GL Luciano Cascione
SI Institute of Oncology
Research
IU 20-40 AU
0.05 PB
80 CP
L Bellinzona

SEE P. 68

# III.
# PERSONALIZED HEALTH

SIB is playing a leading role in the Swiss personalized health landscape, by sharing its expertise with hospitals and by building the nationwide infrastructure that will enable patients' data to be used for research.

Infrastructure, expertise and collaboration between bioinformatics and health experts, for the patients' benefit.

## Clinical bioinformatics

NAME **SIB CLINICAL BIOINFORMATICS GROUP**

TAGS personalized health; hospital collaborations; pipeline optimization; clinical training; practice harmonization; interoperability; genomics; oncology; infectious diseases; medical genetics.

The group provides expertise and support for the organization, analysis and interpretation of patient-related data (e.g. omics data), converting them into clinically useful information for health professionals.

It notably contributes to:

Establishing consensus and good practices for diagnostic omics analyses across Swiss hospitals;

Building partnerships with Swiss hospitals to develop, implement and sustain state-of-the-art diagnostic tools;

Providing clinical bioinformatics training across Switzerland;

Facilitating interactions between SIB's research groups and the medical realm for clinical research projects.

### 2017 highlights

WORKING GROUPS

**THE SOMATIC MUTATION CALLING** working group has started to benchmark Swiss hospitals next-generation sequencing (NGS) practices and define common guidelines for the analysis and interpretation of genetic variants in cancer diagnosis.

The group is also a partner of SIB in the Swiss Variant Interpretation Platform, part of the BioMedIT project (SPHN), which aims to provide a central resource for clinicians and researchers to annotate and interpret clinically verified variants in cancer (SEE NEXT PAGE).

**THE MICROBE TYPING AND CHARACTERIZATION** working group has also started benchmarking NGS practices in the diagnostic of infectious diseases, in both virology and bacteriology.

PARTNER HOSPITALS

At the Geneva University Hospitals, OncoBench™ is now running in version 2, with new features and major improvements.

The Institute of Medical Genetics of Zurich has also entered the partnership to jointly develop a tool for the management and analysis of medical genetics data.

CLINICAL RESEARCH COLLABORATIONS

The group has been involved in several clinical projects, focusing on research and/or the development of resources of clinical utility. In particular, it is a partner in a multi-site NRP72 project led by the University Hospital of Basel, on the development of a Swiss surveillance database for molecular epidemiology of multi-drug resistant pathogens.

**Launch of a Certificate of Advanced Studies**

SIB has partnered with the University Hospitals of Basel and Lausanne to launch a unique, multi-disciplinary and multi-site Certificate of Advanced Studies (CAS) on personalized molecular oncology. The first of its kind in Switzerland, this programme aims to train the next generation of professionals from various backgrounds in this rapidly evolving field. www.pmo.unibas.ch

**In 2017, SIB Clinical Bioinformatics group was also involved in several outreach-related events, to foster discussions and exchange ideas on personalized health with non-specialists: Expand Your Horizons, Café Scientifique and Leenaards SantéPerso.**

## SIB and the Swiss Personalized Health Network initiative (SPHN)

NAME **SIB PERSONALIZED HEALTH INFORMATICS GROUP (PHI)**

TAGS interoperability; personalized health; data management; standardization; precision therapy; information security

The group contributes to the SPHN initiative in a number of ways. As the group in charge of the Data Coordination Centre (DCC), it establishes standards for the interoperability of health-related data. These standards will allow for efficient nationwide collaborative research across hospitals and academic research institutions. The group is also coordinating a network of bioinformatics core facilities at Swiss universities, to provide secure IT infrastructure and support personalized health research projects (BioMedIT project directly managed by SIB, SEE BELOW).

### 2017 highlights

**FIRST INFRASTRUCTURE STEP FOR BIOMEDIT**

In November 2017, the green light was given to launch the first step of the infrastructure implementation under the leadership of SIB Clinical Bioinformatics, together with the Vital-IT and SIS (ETH Zurich) Groups. Swiss hospitals and SIB have decided to join forces to develop a common infrastructure allowing hospitals to share and unify the clinical interpretation of the genetic variants identified in their patients all over Switzerland.

**WORKSHOP ON INFORMATION SECURITY AND DATA PROTECTION**

The 2-days workshop took place in December and gathered about 40 SPHN stakeholders from Swiss university hospitals and research institutions, including specialists in IT security, data protection, legal and governance aspects.

**In addition to the involvement of the PHI Group, several other SIB Groups are playing an active role in the SPHN initiative via project calls and the implementation of BioMedIT.**

The results of SPHN's first project call were issued in December 2017: of the 15 projects selected, five have SIB Group Leaders as co-applicants or co-Principal Investigators and three are led by SIB Group Leaders. These SIB-led projects include:

**An "infrastructure development" project on cancer variants annotation**

SIB PIs DANIEL STEKHOVEN AND PATRICK RUCH

The Swiss Variant Interpretation Platform for Oncology (SVIP-O) aims to provide a central resource for clinicians and researchers to annotate and interpret clinically verified variants in cancer. The database and the governance mechanisms have been launched in the context of BioMedIT (see 'First infrastructure step for BioMedIT').

---

**A "driver" project for oncology treatment and research**

SIB PI OLIVIER MICHIELIN (IN ASSOCIATION WITH SIB'S JACQUES FELLAY, PATRICH RUCH AND SYLVAIN PRADERVAND)

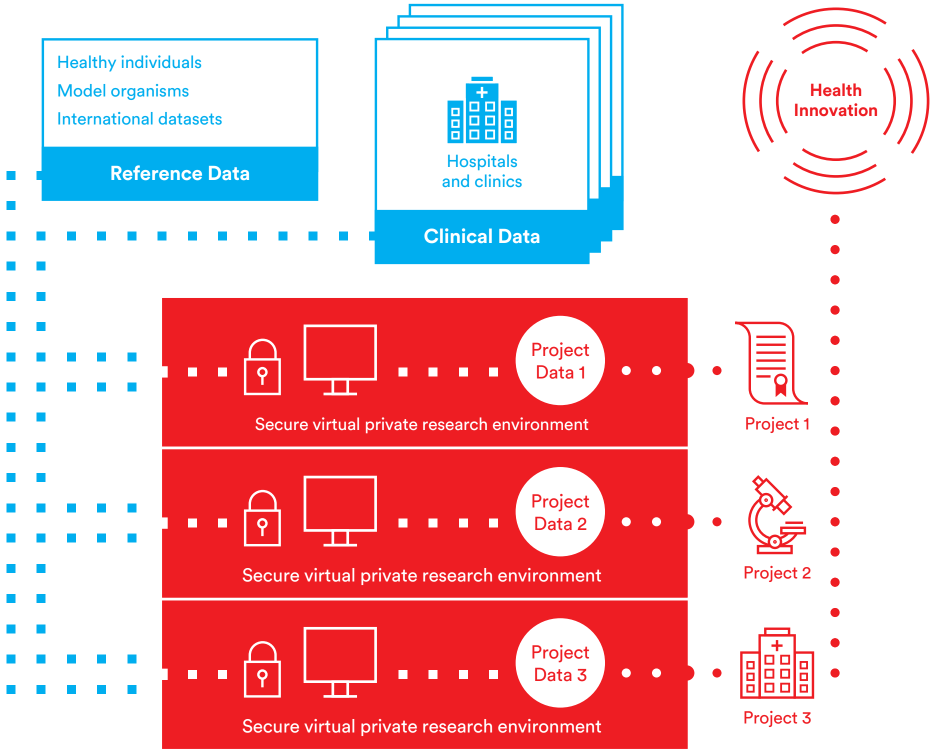
This "driver" project aims to adapt the clinical information flow in order to capture essential data, which can then be used for treatment and research. It will be collaborating with the "Multidisciplinary Multicentre Molecular and Cellular Cancer Consortium" project.

---

**A "driver" project for inflammatory disease immunotherapy**

SIB PI MANFRED CLAASSEN

This Swiss-wide SPHN/PHRT* driver project aims to identify cell population biomarkers and therapeutic targets in inflammatory disease immunotherapy.



**BioMedIT**
**Network of IT competence centres led by SIB**

---

* PHRT: Personalized Health and Related Technology www.sfa-phrt.ch

# Creating a strong bioinformatics community

## IV.
## TRAINING IN BIOINFORMATICS

**23**
SIB groups engaged in teaching activities

**1,191**
trainees

**81**
experts and trainers

**53**
courses and workshops

**94**
training days

### Stronger governance for training

The governance of the group was strengthened in 2017 to determine its strategy in terms of course content and audience reach.

Two SIB Group Leaders, Mark Robinson and Marc Robinson-Rechavi, have been elected by their peer Group Leaders to represent them on the SIB Training Committee. They have joined the four Group Leaders whose groups have long been involved in training activities: Rémy Bruggmann, Mauro Delorenzi, Torsten Schwede and Ioannis Xenarios.

Moreover, three external scientists have joined the Committee to represent the academic and industry-based life scientists: Pierre Farmer (Novartis), Florian Steiner (UNIGE) and Christoph Handschin (Biozentrum).

Marc Robinson-Rechavi has agreed to serve as chair of the Committee and works closely with Patricia Palagi, Team Leader of the Training Group.

### SIB PhD Training Network

In 2017, the network celebrated its 10th anniversary with a special conference for its 240 members in Biel/Bienne. Other events in 2017 included the autumn school in "Machine Learning Applied to Systems Biology", jointly organized with SystemsX.ch, and the course on "Best practices in programming", which were both very well attended.

### Reaching out to new training audiences

SIB hosted a half-day workshop to present its computational biology resources to the Biology17 conference's participants and is repeating this positive experience in 2018 (SEE PAGE 47).

Scientists from the industry are also in need of quality training in bioinformatics. Reaching out to them is part of the new training strategy and action plan.

**Who took part in SIB courses in 2017?**

- 🟥 43% PhD students
- 🟦 25% Postdocs
- ■ 2.5% Principal Investigators
- 🟧 4% Laboratory technicians
- ▨ 11% Other scientists
- ☐ 14.5% Other

The SIB Training portfolio is constantly evolving to meet the scientific community's needs: find the full list of courses at **www.sib.swiss/training**

# V.
# SCIENTIFIC COLLABORATION

SIB is a unique ecosystem, in which Swiss bioinformatics service providers rub shoulders with users – researchers working at the interface between experimental and computational biology – on a daily basis. It thus ensures that the two types of scientists understand each other and collaborate optimally. This is an essential ingredient in SIB's core mission, which is to provide cutting-edge bioinformatics infrastructure in line with the fast-paced evolution of life science.

## Key scientific events and research outcomes in 2017

At SIB, fostering scientific collaboration means creating opportunities for our scientists to exchange know-how, as well as promoting their work. Some of the actions implemented in 2017 are listed below, together with a digest of some of the research results released during the year.

**14.12.2017**

2,000th follower on Twitter

### [BC]$^2$ - Basel Computational Biology Conference

Co-organized by SIB and the Biozentrum University of Basel.

Key computational biology event of the year in Switzerland.

This 13th conference attracted over 500 participants from all over the globe. Among them were world-leading researchers alongside promising junior scientists, who returned home with prestigious awards.

### LS$^2$ Annual meeting, Zurich

International scientists from all backgrounds gather to explore the diverse range of life science topics.

A symposium on personalized medicine in the context of the Swiss Personalized Health Network was chaired by SIB Group Leader Torsten Schwede.

### Biology17, Bern

The Swiss conference on organismic biology and joint congress of the Swiss Zoological Society, the Swiss Botanical Society, and the Swiss Systematics Society.

First incursion of computational biology at the conference, with an SIB workshop during which Patricia Palagi, Joshua Payne, Alexandra Gavryushkina, and Christophe Dessimoz respectively presented the institution and the following resources: Genonets, BEAST2 and OMA.

### Congratulations to the winners of the SIB Awards and [BC]$^2$ Best Posters Awards 2017

**SIB International Young Bioinformatician Award – CHF 10,000**

**SEBASTIAN WASZAK,** EMBL (Germany), for his outstanding work in computational and statistical biology.

Following a PhD at EPFL in Bart Deplancke's Lab, he is now a postdoc at EMBL in the group led by Jan Korbel.

**SIB Bioinformatics Resource Innovation Award – CHF 10,000**

The drug discovery resource **REPURPOSEDB - REFERENCE DATABASE OF DRUG REPOSITIONING INVESTIGATIONS,** represented by Khader Shameer.

This resource includes computational drug discovery tools such as drug similarity searches, and was developed at the Dudley Lab, led by Joel Dudley, at the Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai (USA).

**SIB Best Swiss Bioinformatics Graduate Paper Award – CHF 5,000**

**JOSÉ AGUILAR-RODRIGUEZ,** University of Zurich, for his paper "A thousand empirical adaptive landscapes and their navigability", published in January 2017 in *Nature Ecology & Evolution*.

This paper, which explores the topography of adaptive landscapes in transcription regulation using high-throughput empirical data, was published as a part of his PhD thesis in Andreas Wagner's group.

**[BC]$^2$ Best Poster Award, 1st prize**

**ANNA MARCIONETTI,** SIB & University of Lausanne, with "Investigating the genetic basis of clownfish adaptive radiation using comparative genomics"

**[BC]$^2$ Best 'Shared Poster' Award, 1st prize**

**LUKAS FOLKMAN,** SIB & ETH Zurich, with "Kernelized Rank Learning for Personalized Drug Recommendation"

# Research at SIB in 2017

### A toolbox to improve genome annotation

With thousands of prokaryotic genomes newly sequenced each year, solutions for accurately annotating these genomes and thereby providing the basis for uncovering the function(s) of genes and their encoded proteins are urgently needed. SIB scientists have developed an open proteogenomics toolbox to allow researchers to obtain more accurate and complete genome annotations.

Published in *Genome Research*

SIB Groups involved

| | |
|---|---|
| Bioinformatics and Proteogenomics Group | P. 58 |
| Statistical Bioinformatics Group | P. 56 |

### A gene increases the severity of common colds

Colds that are not linked to influenza are generally benign. Nonetheless, 2% of each generation of children have to go to hospital following a virulent infection. An international research collaboration led by scientists from SIB and EPFL has discovered the reason for some of these infections: they are caused by mutations of a gene that plays a part in recognizing certain cold-inducing viruses.

Published in *Proceedings of the National Academy of Sciences*

SIB Group involved

| | |
|---|---|
| Host-Pathogen Genomics Group | P. 54 |

### Harmful mutations and range expansion: computers got it right

Organisms that are expanding their spatial range suffer from a drastic decrease of fitness over time due to the accumulation of harmful mutations. A study led by a team of scientists from SIB and the University of Bern now provides the first experimental evidence of this theory.

Published in *Genetics*

SIB Groups involved

| | |
|---|---|
| Computational and Molecular Population Genetics Group | P. 60 |
| Interfaculty Bioinformatics Unit (IBU) | P. 68 |

### Supercoiling pushes molecular handcuffs along chromatin fibres

As it squeezes down the chromatin fibre, the cohesin protein complex extrudes a growing loop of DNA – a bit like the quick-lacing system of trail-running shoes. But what is powering the movement of the protein? A team of SIB scientists has found that the driving force could be the supercoiling of upstream DNA. Their research is thereby adding a key piece to the puzzle of gene expression regulation.

Published in *Nucleic Acids Research*

SIB Group involved

| | |
|---|---|
| DNA and Chromosome Modelling Group | P. 57 |

### A thorough ‘catalogue’ of an aggressive type of childhood cancer

Improving cancer precision therapy requires a greater ability to identify and describe groups of patients who share the same molecular and clinical particularities. An SIB group recently played a pivotal role in building the most thorough ‘catalogue’ of an aggressive type of childhood cancer –thus providing a basis for novel therapeutic treatments.

Published in *Cancer Cell*

SIB Group involved

| | |
|---|---|
| Computational Oncogenomics Group | P. 52 |

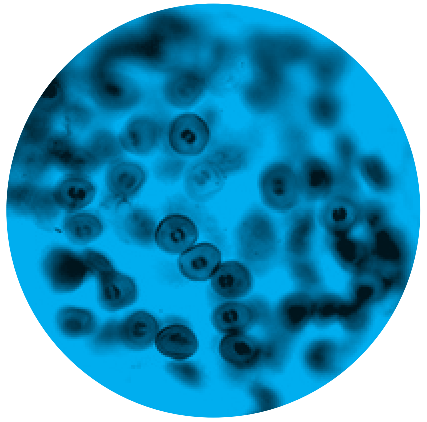
### From cancer evolution to personalized therapies

A study led by researchers at SIB, the University of Lausanne and EPFL provides a promising framework to anticipate drug resistance in cancer, by predicting the co-occurrence of tumour alterations and their response to common cancer drugs. The research showed that while some co-occurrences between genomic alterations confer a resistance on tumour cells to particular drugs, they also make them sensitive to others.

Published in *Cancer Cell*

SIB Group involved

| | |
|---|---|
| Computational Systems Oncology | P. 64 |



### The genome of a 234-year-old oak tree, young as ever

A study carried out by researchers at the University of Lausanne and SIB has revealed that the genome of an emblematic tree of the University campus, the Napoleon oak, has changed little during its first 234 years of existence. This surprising result suggests that the tree is capable of protecting itself from an accumulation of harmful genetic mutations.

Published in *Nature Plants*

SIB Groups involved

| | |
|---|---|
| Vital-IT | P. 71 |
| Evolutionary Bioinformatics Group | P. 61 |
| Bioinformatics Core Facility (BCF) | P. 53 |

### The quick-change artist strategy of a fungus that causes pneumonia

A study led by a team of researchers from the Institute of Microbiology of CHUV-UNIL and SIB reveals the mechanisms of antigenic variation used by the pathogen *Pneumocystis jirovecii* to escape the human immune system.

Published in *mBio*

SIB Group involved

| | |
|---|---|
| Vital-IT | P. 71 |

### The 16 genetic markers that can cut a life story short

The answer to how long each of us will live is partly encoded in our genome. Researchers have identified the largest set of markers of lifespan uncovered to date. About 10% of the population carries some configurations of these markers that shorten their life by over a year compared with the population average. Spearheaded by scientists from SIB, the Lausanne University Hospital (CHUV), the University of Lausanne and EPFL, the study provides a powerful computational framework to uncover the genetics of our time of death, and ultimately of any disease.

Published in *Nature Communications*

SIB Groups involved

| | |
|---|---|
| Statistical Genetics Group | P. 55 |
| Laboratory of Systems Biology and Genetics | P. 53 |
| Evolutionary Bioinformatics Group | P. 61 |

Full references of the papers mentioned are available at www.sib.swiss/about-us/media-homepage/news-2017

# Fields of activity of SIB Groups

Bioinformatics is an interdisciplinary field, where the encounter of genetics, physiology, chemistry or physics leads to many fields of activities and applications.

## Genes and genomes

Life's instruction manual

A genome is the sum of genetic material of an organism, including all of its genes. It is composed of DNA and contains all the information needed to create and maintain an organism, as well as the instructions on how this information should be expressed.

Bioinformatics develops tools able to read genomes, and store, analyse and interpret the resulting data.

**SEE P. 52**

## Proteins and proteomes

More than meets the eye

A proteome is the sum 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 carried out within an organism – from carrying oxygen to fighting off pathogens.

Bioinformatics develops tools to understand the role of proteins.

**SEE P. 58**

## Evolution and phylogeny

Splitting ends

Changes that occur in genomes tell life scientists how an organism has evolved over time. Comparisons made between genomes from different species or populations tell them how species are related to one another – this is the field of phylogenetics.

Bioinformatics develops tools able to compare the genomes of organisms, as well as computing methods to reconstruct their past and build their 'family' trees.

**SEE P. 60**

## Structural biology

The third dimension

Macromolecules such as DNA and proteins have specific 3D structures that are dictated by their sequence. A protein's function is defined by its 3D structure, or architecture, which in turn defines the way it reacts with other molecules.

Bioinformatics develops software to create 3D models of proteins to study their interactions with other molecules, such as drugs.

**SEE P. 63**

## Systems biology

Never alone

Life occurs and is sustained by a mesh of interactions within and between cells, tissues, organisms, and their environment. Understanding how these complex systems function allows scientists to predict what happens if one of the components changes or the conditions are altered.

Bioinformatics develops models to delineate metabolic pathways.

**SEE P. 64**

## Core facilities and competence centres

The means to an end

The quantity of data generated by the life sciences has grown exponentially over the years, and needs to be stored and processed. Researchers also need help in interpreting their data. Core facilities centralize research resources, and provide tools, technologies, services and expert consultation to this end.

SIB coordinates 12 core facilities and competence centres located in the major Swiss academic institutions.

**SEE P. 68**

## Text mining and machine learning NEW

Rise of the machines

Text mining extracts high-quality information from texts. Algorithms are designed to recognize patterns and trends within text so that computers can read and extract the information required. Computers can also be given the ability to learn without being explicitly programmed: this is the promising field of machine learning.

Bioinformatics develops text-mining tools that can be of immense value to complement expert biocuration. It can also rely on machine learning techniques to predict the outcome of complex interactions such as in antibiotic resistance.

**SEE P. 71**

## Many areas of applications, including…

**…AGRICULTURE**
from predicting the spread of bird flu outbreaks and understanding the lifecycle of agricultural pests, to improving crop productivity.

**…BASIC RESEARCH**
from unravelling the evolutionary processes that have shaped today's biodiversity, to solving the equation behind a lizard's scale colour pattern.

**…MEDICINE AND HEALTH**
from designing optimized proteins in cancer immunotherapy, to creating biomedical decision-support tools using text-mining.

**…ECOLOGY**
from understanding how insects adapt to specific ecological niches, to documenting the biodiversity of fungi in the soil.

## Michael BAUDIS

GROUP NAME **COMPUTATIONAL ONCOGENOMICS GROUP**
INSTITUTION UNIVERSITY OF ZURICH
CITY ZURICH

The Computational Oncogenomics Group works on the computational analysis of structural genome variations in cancer. Our work centres around our arrayMap and Progenetix resources of curated molecular cytogenetic and sequencing data. Specific projects explore computational methods, genomics of selected tumour entities as well as genomic variant patterns across malignancies. As a member of the Global Alliance for Genomics and Health, the group is developing standards in biocuration and data sharing for genomic variants and phenotypic data. Other projects are related to genome data epistemology, e.g. geographic and diagnostic sampling biases.

MAIN DOMAIN OF ACTIVITY
Genes and genomes

SECONDARY DOMAINS OF ACTIVITY
Biocuration; Biomarkers; Data mining; Knowledgebase; Microarrays; Next-generation sequencing; Oncology; Precision therapy; Software development; Text mining; APIs

DOMAINS OF APPLICATION

RESOURCES
Progenetix
arrayMap
Beacon+

KEY PAPER IN 2017

Mackay *et al.* Integrated molecular meta-analysis of 1,000 pediatric high-grade and diffuse intrinsic pontine glioma.
Cancer Cell

## Niko BEERENWINKEL

GROUP NAME **COMPUTATIONAL BIOLOGY GROUP**
INSTITUTION ETH ZURICH, D-BSSE
CITY ZURICH & BASEL

The Computational Biology Group is involved in research and teaching in the areas of computational biology, biostatistics, and systems biology. Our activities include the development of mathematical and statistical models, their implementation in computer programs, and their application to biomedical problems. We are conducting active research projects on HIV drug resistance, the somatic evolution of cancer, haplotype inference from ultra-deep sequencing data, reconstruction of signalling pathways from perturbation screens, and computational analysis of single-cell sequencing data.

MAIN DOMAIN OF ACTIVITY
Genes and genomes

SECONDARY DOMAINS OF ACTIVITY
Biostatistics; Drug resistance; Evolution; Infectious diseases; Next-generation sequencing; Oncology; Phylogeny; Precision therapy; Single-cell biology; Software development

DOMAIN OF APPLICATION

RESOURCES
V-pipe
SCITE
NGS-pipe

KEY PAPERS IN 2017

Mohammadi P *et al.* Automated design of synthetic cell classifier circuits using a two-step optimization strategy.
Cell Systems

Kuipers J *et al.* Single-cell sequencing data reveals widespread recurrence and loss of mutational hits in the life histories of tumors.
Genome Res.

Crona K *et al.* Inferring genetic interactions from comparative fitness data.
eLife

## Sven BERGMANN

GROUP NAME **COMPUTATIONAL BIOLOGY GROUP**
INSTITUTION UNIVERSITY OF LAUSANNE
CITY LAUSANNE

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

MAIN DOMAIN OF ACTIVITY
Genes and genomes

SECONDARY DOMAINS OF ACTIVITY
Biostatistics; Epigenetics; Genetic networks; GWAS; Metabolomics; Microarrays; Protein networks; Software development; Systems biology; Transcriptomics

DOMAINS OF APPLICATION

RESOURCES
Pascal
Metabomatching
Iterative signature algorithm

KEY PAPERS IN 2017

Lamparter D *et al.* Genome-wide association between transcription factor expression and chromatin accessibility reveals regulators (...).
PLoS Comput Biol.

Yihong Tan J *et al.* cis-acting complex-trait-associated lincRNA expression correlates with modulation of chromosomal architecture.
Cell Rep.

Rueedi R *et al.* Metabomatching: using genetic association to identify metabolites in proton NMR spectroscopy.
PLoS Comput Biol.

## Philipp BUCHER

GROUP NAME **COMPUTATIONAL CANCER GENOMICS GROUP**
INSTITUTION EPFL
CITY LAUSANNE

In the Computational Cancer Genomics Group, we are interested in gene regulation in both healthy and diseased cells. Breakthroughs in genomics technologies have led to the production of large volumes of data that could potentially tell us something about how gene regulatory instructions are encoded in our DNA. Our group develops new algorithms, computer programs, web services and databases that will help us and others to extract knowledge and understanding from such data. We also collaborate with experimental biologists on interdisciplinary projects and organize postgraduate courses in genomic data analysis.

MAIN DOMAIN OF ACTIVITY
Genes and genomes

SECONDARY DOMAINS OF ACTIVITY
Comparative genomics; Epigenetics; Genome structure; Machine learning; Software development

DOMAINS OF APPLICATION

RESOURCES
Eukaryotic Promoter Database EPD
ChIP-Seq tools and web server
PWMTools

KEY PAPERS IN 2017

Isakova A *et al.* SMiLE-seq identifies binding motifs of single and dimeric transcription factors.
Nat Methods

Dreos R *et al.* The eukaryotic promoter database in its 30th year: focus on non-vertebrate organisms.
Nucleic Acids Res.

Kumar S *et al.* SNP2TFBS - a database of regulatory SNPs affecting predicted transcription factor binding site affinity.
Nucleic Acids Res.

## Mauro DELORENZI
## Frédéric SCHÜTZ

GROUP NAME **BIOINFORMATICS CORE FACILITY (BCF)**
INSTITUTION UNIVERSITY OF LAUSANNE
CITY LAUSANNE

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

MAIN DOMAINS OF ACTIVITY
Genes and genomes
Core facilities and competence centres

SECONDARY DOMAINS OF ACTIVITY
Biomarkers; Biostatistics; Data mining; Immunology; Machine learning; Metabolomics; Next-generation sequencing; Oncology; Single-cell biology; Transcriptomics

DOMAINS OF APPLICATION

KEY PAPERS IN 2017

Barras D *et al.* BRAF V600E mutant colorectal cancer subtypes based on gene expression.
Clin Cancer Res.

Dobay MP *et al.* Integrative clinicopathological and molecular analyses of angioimmunoblastic T-cell lymphoma and other nodal lymphomas of follicular helper T-cell origin.
Haematologica

Frismantas V *et al.* Ex vivo drug response profiling detects recurrent sensitivity patterns in drug-resistant acute lymphoblastic leukemia.
Blood

## Bart DEPLANCKE

GROUP NAME **LABORATORY OF SYSTEMS BIOLOGY AND GENETICS**
INSTITUTION EPFL
CITY LAUSANNE

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

MAIN DOMAIN OF ACTIVITY
Genes and genomes

SECONDARY DOMAINS OF ACTIVITY
Genome structure; Functional genomics; Microfluidics; Proteomics; Single-cell biology; Systems biology

DOMAIN OF APPLICATION

RESOURCES
ASAP
GETPrime

KEY PAPERS IN 2017

Schwalie PC *et al.* Cross-tissue identification of somatic stem and progenitor cells using a single-cell RNA-seq derived gene signature.
Stem Cells

Gardeux V *et al.* ASAP: a web-based platform for the analysis and interactive visualization of single-cell RNA-seq data.
Bioinformatics

Isakova A *et al.* SMiLE-seq identifies binding motifs of single and dimeric transcription factors.
Nat methods

## Emmanouil DERMITZAKIS

GROUP NAME **GENOMICS OF COMPLEX TRAITS GROUP**
INSTITUTION UNIVERSITY OF GENEVA
CITY GENEVA

Our group has a strong interest in population genomics and genetics of complex traits. We use various methodologies to understand the role of genetic variation in phenotypic variation and what fraction of genetic variation is harboured within functional elements of the human genome. Our main focus is on genome-wide analysis of gene expression and cellular phenotypes and their association with nucleotide variation. We attempt to detect functional genetic variation in regulatory elements and then use functional variation and accurately measured gene expression variation to bridge the genotype with disease phenotypes in association studies.

MAIN DOMAIN OF ACTIVITY
Genes and genomes

SECONDARY DOMAINS OF ACTIVITY
Comparative genomics; Population genetics; Epigenetics; Functional genomics; Genome structure; GWAS; Next-generation sequencing; Oncology; Systems biology; Transcriptomics

DOMAIN OF APPLICATION

RESOURCE
QTLtools

KEY PAPERS IN 2017

Ongen H *et al.* Estimating the causal tissues for complex traits and diseases. Nat Genet.

Brown AA *et al.* Predicting causal variants affecting expression by using whole-genome sequencing and RNA-seq from multiple human tissues. Nat Genet.

Delaneau O *et al.* A complete tool set for molecular QTL discovery and analysis. Nat Commun.

## Jacques FELLAY

GROUP NAME **HOST-PATHOGEN GENOMICS GROUP**
INSTITUTION EPFL
CITY LAUSANNE

In the Host-Pathogen Genomics Group, we explore the genetic roots of inter-individual differences in response to infections, with a particular focus on the genomic interactions between pathogens and their human hosts. At the crossroads between basic science and the clinical world, we are committed to translational genomic research, aimed at identifying, validating and bringing to clinical use genetic markers of susceptibility to infectious diseases – e.g. host genomics of HIV infection, joint analyses of interactions between human and viral genomes, and exome sequencing on patients with extreme infectious disease phenotypes.

MAIN DOMAIN OF ACTIVITY
Genes and genomes

SECONDARY DOMAINS OF ACTIVITY
GWAS; Immunology; Infectious diseases; Next-generation sequencing; Precision therapy

DOMAIN OF APPLICATION

KEY PAPERS IN 2017

Asgari S *et al.* Severe viral respiratory infections in children with IFIH1 loss-of-function mutations. Proc Natl Acad Sci USA

Bartha I *et al.* Estimating the respective contributions of human and viral genetic variation to HIV control. PLoS Comput Biol.

McLaren PJ *et al.* Evaluating the impact of functional genetic variation on HIV-1 control. J Infect Dis.

## David GFELLER

GROUP NAME **COMPUTATIONAL CANCER BIOLOGY GROUP**
INSTITUTION UNIVERSITY OF LAUSANNE
CITY LAUSANNE

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

MAIN DOMAIN OF ACTIVITY
Genes and genomes

SECONDARY DOMAINS OF ACTIVITY
Immunology; Machine learning; Peptidomics; Single-cell biology; Transcriptomics

DOMAINS OF APPLICATION

RESOURCES
EPIC
MixMHCp
TILAtlas

KEY PAPERS IN 2017

Racle *et al.* Simultaneous enumeration of cancer and immune cell types from bulk tumor gene expression data. elife

Bassani-Sternberg *et al.* Deciphering HLA-I motifs across HLA peptidomes improves neo-antigen predictions (...). PLoS Comp Bio.

Carmona *et al.* Single-cell transcriptome analysis of fish immune cells provides insight into the evolution of vertebrate immune cell types. Gen Research

## Zoltán KUTALIK

GROUP NAME **STATISTICAL GENETICS GROUP**
INSTITUTION CHUV / UNIVERSITY OF LAUSANNE
CITY LAUSANNE

In the Statistical Genetics Group, we are interested in the development of statistical methodologies in order to decipher the genetic architecture of complex human traits related to obesity. To do this, we efficiently combine large-scale genome-wide association studies (GWAS) with various -omics data. Our methods improve genetic fine-mapping, reveal gene-environment interactions, dissect genetic subtypes of obesity, enhance causal effect estimation and detect various statistical artefacts. Furthermore, we are involved in large consortia researching the genetic basis of anthropometric traits (GIANT) and longevity (LifeGen).

MAIN DOMAIN OF ACTIVITY
Genes and genomes

SECONDARY DOMAIN OF ACTIVITY
Biomarkers; Biostatistics; Population genetics; Epigenetics; Genetic networks; GWAS; Metabolomics; Next-generation sequencing; Transcriptomics

DOMAIN OF APPLICATION

RESOURCES
Quickest
SS-imp

KEY PAPERS IN 2017

Marouli E *et al.* Rare and low-frequency coding variants alter human adult height. Nature

McDaid AF *et al.* Bayesian association scan reveals loci associated with human lifespan and linked biomarkers. Nat Commun.

Mace A *et al.* CNV-association meta-analysis in 191,161 European adults reveals new loci associated with anthropometric traits. Nat Commun.

## Anna-Sapfo MALASPINAS

GROUP NAME **EVOLUTIONARY GENOMICS**
INSTITUTION UNIVERSITY OF LAUSANNE
CITY LAUSANNE

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

MAIN DOMAIN OF ACTIVITY
Genes and genomes

SECONDARY DOMAIN OF ACTIVITY
Population genetics; Paleogenomics; Evolution

DOMAIN OF APPLICATION

KEY PAPERS IN 2017

Ziegler D *et al.* Genome Odyssey, Chapter 1: Australia. Musée d'ethnographie Genève

Dortch J and Malaspinas AS. Madjedbebe and genomic histories of Aboriginal Australia. Australian Archaeology

## Richard NEHER

GROUP NAME **MICROBIAL EVOLUTION**
INSTITUTION UNIVERSITY OF BASEL
CITY BASEL

Our goal is to understand how evolution works at the molecular level and how organisms adapt to changing conditions through random mutations and recombination. Pathogens provide excellent model systems for studying these processes. We apply modern sequencing techniques to decipher the genome of thousands of HIV particles and develop new algorithms to elucidate the interactions between HIV and the immune system. We have developed a method that can predict the evolution of influenza viruses based on their phylogenetic tree (nextflu.org). Such predictions can help to ensure that the seasonal influenza vaccine matches the viruses in circulation.

MAIN DOMAIN OF ACTIVITY
Genes and genomes

SECONDARY DOMAIN OF ACTIVITY
Drug resistance; Evolution; Infectious diseases; Next-generation sequencing; Phylogeny; Software development

DOMAIN OF APPLICATION

RESOURCES
TreeTime
nextstrain.org
pangenome.de

KEY PAPERS IN 2017

Puller V *et al.* Estimating time of HIV-1 infection from next-generation sequence diversity. PLOS Comput Biol.

Zanini F *et al.* In vivo mutation rates and the landscape of fitness costs of HIV-1. Virus Evol.

Ding W *et al.* panX: pan-genome analysis and exploration. Nucleic Acids Res.

## Mark ROBINSON

GROUP NAME **STATISTICAL BIOINFORMATICS GROUP**
INSTITUTION UNIVERSITY OF ZURICH
CITY ZURICH

In the Statistical Bioinformatics Group, we develop robust data analysis solutions, including new or improved methods, for the analysis of genome-scale data. We develop statistical methods for interpreting data from high-throughput sequencing and other technologies in the context of genome sequencing, gene expression and regulation and analysis of epigenomes. We are largely data- and problem-driven, and ultimately the methods we develop are geared to the characteristics of the technology platform generating the data. We develop publicly available open-source software tools, generally through the Bioconductor project.

MAIN DOMAIN OF ACTIVITY
Genes and genomes

SECONDARY DOMAINS OF ACTIVITY
Biostatistics; Epigenetics; Experimental biology; Immunology; Microarrays; Next-generation sequencing; Oncology; Single-cell biology; Software development; Transcriptomics

DOMAINS OF APPLICATION

RESOURCES
Cytometry dATa anALYSis Tools (CATALYST)
CrispRVariants
conquer

KEY PAPERS IN 2017

Nowicka *et al.*
CyTOF workflow: differential discovery in high-throughput high-dimensional cytometry datasets.
F1000 Research

Soneson C and Robinson MD
Towards unified quality verification of synthetic count data with countsimQC.
Bioinformatics

Sharma *et al.*
Male sex in houseflies is determined by Mdmd, a paralog of the generic splice factor gene CWC22.
Science

## Matthew ROBINSON

GROUP NAME **COMPLEX TRAIT GENETICS GROUP**
INSTITUTION UNIVERSITY OF LAUSANNE
CITY LAUSANNE

We develop and apply statistical methodologies to large human phenotype-genotype datasets, to address long-standing questions in population and quantitative genetics. These have improved the prediction accuracy of disease risk in personalized medicine, quantified the contribution of genotype-environment interaction effects to obesity risk, and provided frameworks to examine the genetic basis of phenotypic differences among people. Our current work focuses on improved testing for sex-, age-, or environment-specific genetic effects and investigating the role of interactions between microbial and host genotypes in shaping human phenotypes.

MAIN DOMAIN OF ACTIVITY
Genes and genomes

SECONDARY DOMAINS OF ACTIVITY
Data mining; Population genetics; GWAS; Machine learning; Selection; Software development

DOMAIN OF APPLICATION

RESOURCES
smtpred
Bayesian random regression models

KEY PAPERS IN 2017

Robinson MR *et al.*
Genotype–covariate interaction effects and the heritability of adult body mass index.
Nat Genet.

Robinson MR *et al.*
Genetic evidence of assortative mating in humans.
Nat Hum Behav.

Tropf C *et al.*
Hidden heritability due to heterogeneity across seven populations.
Nat Hum Behav.

## Michael STADLER

GROUP NAME **FMI COMPUTATIONAL BIOLOGY GROUP**
INSTITUTION FRIEDRICH MIESCHER INSTITUTE
CITY BASEL

We study gene regulation through the analysis and modelling of genome-wide datasets. We collaborate closely with experimental researchers on various biological topics, including cancer progression, chromatin biology and cellular differentiation, using data from various aspects of gene expression, such as DNA methylation and RNA transcription in single cells, measured by high-throughput sequencing. Our aim is to better understand how the different layers of epigenetic, transcriptional and post-transcriptional regulation interact and contribute to the control of gene expression.

MAIN DOMAINS OF ACTIVITY
Genes and genomes
Core facilities and competence centres

SECONDARY DOMAINS OF ACTIVITY
Epigenetics; Functional genomics; Genetic networks; Microarrays; Next-generation sequencing; Oncology; Single-cell biology; Software development; Systems biology; Transcriptomics

DOMAINS OF APPLICATION

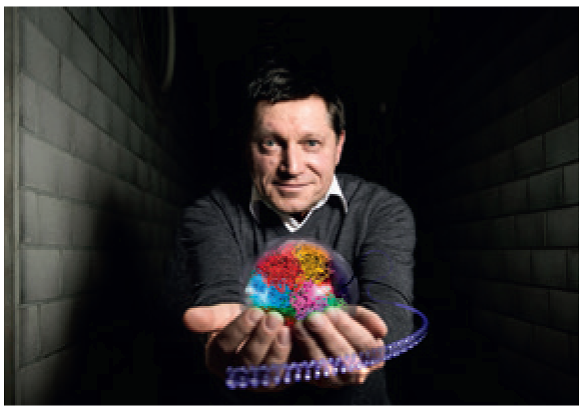
RESOURCES
QuasR
MethylSeekR
griph

KEY PAPERS IN 2017

Minoux M *et al.*
Gene bivalency at Polycomb domains regulates cranial neural crest positional identity.
Science

Horvathova I *et al.*
The dynamics of mRNA turnover revealed by single-molecule imaging in single cells.
Mol Cell

Krebs AR *et al.*
Genome-wide single-molecule footprinting reveals high RNA polymerase II turnover at paused promoters.
Mol Cell



## Andrzej STASIAK

GROUP NAME **DNA AND CHROMOSOME MODELLING GROUP**
INSTITUTION UNIVERSITY OF LAUSANNE
CITY LAUSANNE

Using the mathematical concept of knotoids, we were able to characterize the topology of protein structures without the necessity of imposing a closure of analysed protein chains. New simulation approaches permitted us to emulate the effect of torsional tension introduced during transcription into chromatin fibres, as well as the effects of torsional stress relaxation by DNA topoisomerases. Our modelling indicated that TAD-like self-interacting chromatin domains in chromosomes of *S. pombe* can simply arise as a result of torsional stress accumulated during diverging transcription.

MAIN DOMAIN OF ACTIVITY
Genes and genomes

SECONDARY DOMAINS OF ACTIVITY
Genome structure; Epigenetics; Genome reconstruction; Protein structure and interaction

DOMAIN OF APPLICATION

RESOURCE
KnotProt

KEY PAPERS IN 2017

Goundaroulis D *et al.*
Studies of global and local entanglements of individual protein chains using the concept of knotoids.
Sci Rep.

Benedetti F *et al.*
Transcription-induced supercoiling explains formation of self-interacting chromatin domains in *S. pombe.*
Nucleic Acids Res.

Goundaroulis D *et al.*
Topological models of open-knotted protein chains using the concepts of knotoids and bonded knotoids.
Polymers

## Erik VAN NIMWEGEN

GROUP NAME **GENOME SYSTEMS BIOLOGY GROUP**
INSTITUTION UNIVERSITY OF BASEL
CITY BASEL

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

MAIN DOMAIN OF ACTIVITY
Genes and genomes

SECONDARY DOMAINS OF ACTIVITY
Comparative genomics; Population genetics; Epigenetics; Evolution; Genetic networks; Next-generation sequencing; Phylogeny; Systems biology; Systems modelling; Transcriptomics
Gene regulation; Single-cell biology

DOMAINS OF APPLICATION

RESOURCES
SwissRegulon
ISMARA
CRUNCH

KEY PAPER IN 2017

Omidi S *et al.*
Automated incorporation of pairwise dependency in transcription factor binding site prediction using dinucleotide weight tensors.
PLoS Comput biol.

## Andreas WAGNER

GROUP NAME **EVOLUTIONARY SYSTEMS BIOLOGY GROUP**
INSTITUTION UNIVERSITY OF ZURICH
CITY ZURICH

In the Evolutionary Systems Biology Group, we study the evolution and evolvability of biological systems at all levels of biological organization, from genes and genomes to biological networks and whole organisms. We develop bioinformatics tools to integrate data from a variety of sources, including comparative whole-genome sequence data, microarray expression data, and high-throughput protein interaction data. Our work uses comparative analysis of genomic data, laboratory evolution experiments and mathematical modelling.

MAIN DOMAIN OF ACTIVITY
Genes and genomes

SECONDARY DOMAINS OF ACTIVITY
Evolution; Experimental biology; Genetic networks; Molecular evolution

DOMAIN OF APPLICATION

KEY PAPERS IN 2017

Aguilar Rodriguez J *et al.* 1000 empirical adaptive landscapes and their navigability.
Nat Ecol Evol.

Bratulic S *et al.*
Mistranslation benefits adaptive evolution through purging of deleterious mutations.
Nat Commun.

Fortuna MA *et al.*
The genotype-phenotype map of an evolving digital organism.
PLoS Comput Biol.

## Evgeny ZDOBNOV

GROUP NAME **COMPUTATIONAL EVOLUTIONARY GENOMICS GROUP**
INSTITUTION UNIVERSITY OF GENEVA
CITY GENEVA

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

MAIN DOMAIN OF ACTIVITY
Genes and genomes

SECONDARY DOMAINS OF ACTIVITY
Genome structure; Comparative genomics; Evolution; Functional genomics; Gene ontology; Metagenomics; Next-generation sequencing; Phylogeny

DOMAINS OF APPLICATION

RESOURCES
OrthoDB
BUSCO
miRmap

KEY PAPERS IN 2017

Waterhouse RM *et al.* BUSCO applications from quality assessments to gene prediction and phylogenomics. Mol Biol Evol.

Djordjevic M *et al.* Scoring targets of transcription in bacteria rather than focusing on individual binding sites. Front Microbiol.

Ioannidis P *et al.* Genomic features of the damselfly *calopteryx* splendens representing a sister clade to most insect orders. Genome Biol Evol.

## Christian AHRENS

GROUP NAME **BIOINFORMATICS AND PROTEOGENOMICS GROUP**
INSTITUTION AGROSCOPE
CITY WÄDENSWIL

Our research revolves around the bioinformatic integration and analysis of datasets from state-of-the-art omics technologies, which we obtain through close collaboration with experimental biologists. These datasets include genome sequences, gene and protein expression, as well as metabolomics data. One particular focus is to exploit the unique advantages of proteomics data, including strategies to identify all proteins encoded in a genome (proteogenomics). Recently, we started to study the role of microbiomes – e.g. for plant protection – by applying metagenomics, genomics and transcriptomics approaches.

MAIN DOMAIN OF ACTIVITY
Proteins and proteomes

SECONDARY DOMAINS OF ACTIVITY
Comparative genomics; Data mining; Drug resistance; Functional genomics; Genome structure; Infectious diseases; Metagenomics; Next-generation sequencing; Proteomics; Transcriptomics

DOMAINS OF APPLICATION

RESOURCES
iPtgxDB
PeptideRank
Protter

KEY PAPERS IN 2017

Omasits U *et al.* An integrative strategy to identify the entire protein coding potential of prokaryotic genomes by proteogenomics. Genome Res.

Montes Vidal D *et al.* Long-chain alkyl cyanides: unprecedented volatile compounds released by *pseudomonas* (...). Angew Chem Int Ed Engl.

Hilber-Bodmer M *et al.* Competition assays and physiological experiments of soil and phyllosphere yeasts identify Candida subhashii (...). BMC Microbiol.

## Amos BAIROCH
## Lydie LANE

GROUP NAME **COMPUTER AND LABORATORY INVESTIGATION OF PROTEINS OF HUMAN ORIGIN (CALIPHO)**
INSTITUTION UNIVERSITY OF GENEVA
CITY GENEVA

We aim to use a combination of bioinformatics and experimental methodologies to increase knowledge about the function of the 20,000 protein-coding genes in the human genome. Our main mission is the development of neXtProt, a human protein knowledge resource. We also annotate the effects of human protein variations in the context of cancers and genetic diseases. We are part of the HUPO Human Protein Project, which aims to validate the existence of all predicted human proteins in biological samples by mass spectrometry. We are active in the development of ontologies/standardization resources such as the Cellosaurus and ICEPO.

MAIN DOMAIN OF ACTIVITY
Proteins and proteomes

SECONDARY DOMAINS OF ACTIVITY
Biochemistry; Biocuration; Biomarkers; Data mining; Experimental biology; Knowledgebase; Precision therapy; Proteomics; Systems biology; Text mining

DOMAINS OF APPLICATION

RESOURCES
neXtProt
Cellosaurus

KEY PAPERS IN 2017

Hinard V *et al.* Annotation of functional impact of voltage-gated (...). Hum Mutat.

Ommen GS *et al.* Progress on the HUPO draft human proteome: 2017 Metrics (...). J Proteome Res.

Schaeffer M *et al.* The neXtProt peptide uniqueness checker: a tool for the (...). Bioinformatics

## Frédérique LISACEK

GROUP NAME **PROTEOME INFORMATICS GROUP**
INSTITUTION UNIVERSITY OF GENEVA
CITY GENEVA

In the Proteome Informatics Group (PIG), we are involved in software and database development for the benefit of the proteomics and glycomics communities. These resources are made available through the ExPASy server. Software tools support experimental mass spectrometry data analysis, mainly for the detection of post-translational modifications, with a strong focus on glycosylation. We also develop integrated tools and databases to query knowledge of carbohydrates attached to proteins, carbohydrate structures and protein-carbohydrate interactions.

MAIN DOMAIN OF ACTIVITY
Proteins and proteomes

SECONDARY DOMAINS OF ACTIVITY
Biocuration; Glycomics; Infectious diseases; Mass spectroscopy; Proteomics

DOMAIN OF APPLICATION

RESOURCES
Glyconnect
Glynsight
UniCarb-DB

KEY PAPERS IN 2017

Horlacher O *et al.* Glycoforest 1.0 Anal Chem.

Domagalski *et al.* PepSweetener: a web-based tool to support manual annotation (...). Proteomics Clinic appl.

Campbell *et al.* Chapter 52: Glycoinformatics. Essentials of Glycobiology, 3rd Edition

## Christian VON MERING

GROUP NAME **BIOINFORMATICS / SYSTEMS BIOLOGY**
INSTITUTION UNIVERSITY OF ZURICH
CITY ZURICH

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

MAIN DOMAIN OF ACTIVITY
Proteins and proteomes

SECONDARY DOMAINS OF ACTIVITY
Comparative genomics; Metagenomics; Phylogeny; Protein networks; Proteomics; Software development

DOMAIN(S) OF APPLICATION

RESOURCES
STRING
PAX-db
MAPseq

KEY PAPERS IN 2017

Matias Rodrigues JF *et al.* MAPseq: highly efficient k-mer search with confidence estimates, for rRNA (...). Bioinformatics

Feigelman R *et al.* Sputum DNA sequencing in cystic fibrosis: non-invasive access to the lung microbiome (...). Microbiome

Szklarczyk D *et al.* The STRING database in 2017: quality-controlled protein-protein association networks (...). Nucleic Acids Res.

## Ioannis XENARIOS
## Alan BRIDGE

GROUP NAME **SWISS-PROT**
INSTITUTION SIB GENEVA
CITY GENEVA

In the Swiss-Prot Group, we develop, annotate and maintain the UniProtKB/Swiss-Prot protein sequence database, the most widely used protein information resource in the world (and an ELIXIR Core Data Resource), as well as a number of other internationally renowned expert-curated knowledge resources. These include the HAMAP and PROSITE databases of protein families and domains, the ENZYME database of enzyme nomenclature, the Rhea database of biochemical reactions, the SwissLipids database of lipid structures and biological knowledge, and the ViralZone portal. The group also participates in the development and maintenance of the ExPASy website.

MAIN DOMAIN OF ACTIVITY
Proteins and proteomes

SECONDARY DOMAINS OF ACTIVITY
Biocuration; Gene ontology; Knowledgebase; Lipids; Metabolomics; Proteomics; Systems biology

DOMAIN(S) OF APPLICATION

RESOURCES
UniProtKB/Swiss-Prot
Rhea
ViralZone

KEY PAPERS IN 2017

Hulo C *et al.* Bacterial virus ontology; coordinating across databases. Viruses

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

Poux S *et al.* On expert curation and scalability: UniProtKB/Swiss-Prot as a case study. Bioinformatics

## Maria ANISIMOVA

GROUP NAME **APPLIED COMPUTATIONAL GENOMICS TEAM**
INSTITUTION ZHAW
CITY WÄDENSWIL

The Applied Computational Genomics Team focuses on theoretical and computational aspects of modelling the process of genome evolution and adaptive change. With the growing size and complexity of molecular data, we strive to keep pace by providing accurate, scalable and practical computational solutions that enable a wide range of scientists to analyse patterns of evolution and natural selection in large genomic and omics data. We develop bioinformatic methods for real applications, ranging from biotechnology to biomedical research, ecology and agriculture.

MAIN DOMAIN OF ACTIVITY
Evolution and phylogeny

SECONDARY DOMAINS OF ACTIVITY
Comparative genomics; Evolution; Phylogeny; Selection; Software development; Text mining; Semantic web

DOMAINS OF APPLICATION

RESOURCES
CodonPhyML
TRAL
ProGraphMSA
ProGraphMSA+TR

KEY PAPERS IN 2017

Nater *et al.* Morphometric, behavioral and genomic evidence for a new orangutan species. Current Biology.

Maiolo *et al.* Progressive multiple sequence alignment with the Poisson indel process. BioRxiv

Gatti *et al.* Cross-reactive immunity drives global oscillation and opposed alternation patterns of seasonal influenza A viruses. BioRxiv.

## Christophe DESSIMOZ

GROUP NAME **COMPUTATIONAL EVOLUTIONARY BIOLOGY AND GENOMICS GROUP**
INSTITUTION UNIVERSITY OF LAUSANNE
CITY LAUSANNE

At the interface of biology and computer science, we seek to better understand evolutionary and functional relationships between genes, genomes and species. Key underlying questions are: 1) How can we extrapolate current biological knowledge, concentrated in a few model organisms, to the rest of life? 2) Conversely, how can we exploit the wealth and diversity of life to better grasp an organism of interest? 3) Can we summarize the evolutionary history of species as a sparse mixture of tree topologies? Our activities are divided between bioinformatics methods and resource development, and their application – typically with experimentalists.

MAIN DOMAIN OF ACTIVITY
Evolution and phylogeny

SECONDARY DOMAINS OF ACTIVITY
Comparative genomics; Evolution; Gene ontology; Machine learning; Next-generation sequencing; Phylogeny; Software development

DOMAINS OF APPLICATION

RESOURCES
OMA Browser
Phylo.io
ALF

KEY PAPERS IN 2017

Train *et al.* Orthologous matrix (OMA) algorithm 2.0: more robust to asymmetric evolutionary rates (...). Bioinformatics

Jeffares *et al.* Transient structural variations have strong effects on quantitative traits and reproductive isolation in fission yeast. Nat Commun.

Dessimoz C and Škunca N. The Gene Ontology Handbook. Methods in Molecular Biology

## Laurent EXCOFFIER

GROUP NAME **COMPUTATIONAL AND MOLECULAR POPULATION GENETICS GROUP**
INSTITUTION UNIVERSITY OF BERN
CITY BERN

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

MAIN DOMAIN OF ACTIVITY
Evolution and phylogeny

SECONDARY DOMAINS OF ACTIVITY
Population genetics; Evolution; Selection; Software development

DOMAINS OF APPLICATION

RESOURCES
Arlequin
fastsimcoal2
splatche

KEY PAPERS IN 2017

Sikora M *et al.* Ancient genomes show social and reproductive behavior of early Upper Paleolithic foragers. Science

Bosshard L *et al.* Accumulation of deleterious mutations during bacterial range expansions. Genetics

Gouy A *et al.* Detecting gene subnetworks under selection in biological pathways. Nucleic acids Res.

## Gaston GONNET

GROUP NAME **COMPUTATIONAL BIOCHEMISTRY RESEARCH GROUP**
INSTITUTION ETH ZURICH
CITY ZURICH

In the Computational Biochemistry Research Group, we are interested in the modelling and analysis of biological problems at the molecular level. In particular, our expertise lies in searching algorithms, optimizing algorithms, mathematical modelling, and computational systems. Most of our research efforts are concentrated on the Orthologous MAtrix (OMA) project. This particular project aims to produce, automatically, reliable orthologous groups of proteins that are derived from entire genomes.

MAIN DOMAIN OF ACTIVITY
Evolution and phylogeny

SECONDARY DOMAINS OF ACTIVITY
Comparative genomics; Evolution; Metagenomics; Phylogeny; Software development

DOMAINS OF APPLICATION

RESOURCES
OMA Browser

KEY PAPERS IN 2017

Train CM *et al.* Orthologous Matrix (OMA) algorithm 2.0: more robust to asymmetric evolutionary rates and more (...). Bioinformatics

Altenhoff AM *et al.* The OMA orthology database in 2018: retrieving evolutionary relationships among all domains of life (...). Nucleic acids res.

Forslund K *et al.* Gearing up to handle the mosaic nature of life in the quest for orthologs. Bioinformatics

## Jérôme GOUDET

GROUP NAME **POPULATION GENETICS AND GENOMICS GROUP**
INSTITUTION UNIVERSITY OF LAUSANNE
CITY LAUSANNE

Our group's interest is focused on understanding how the interplay of population structure, trait architecture and selection can be disentangled. To this end, we use different approaches, from theory and the development of statistical tools to field observations. The main biological models currently used are humans and the barn owl. On the theoretical side, we investigate the dynamics of multilocus genetic systems under the influence of selection, migration and drift, and develop comprehensive individual-based models as well as statistical methods to infer selection, mating systems and population structure.

MAIN DOMAIN OF ACTIVITY
Evolution and phylogeny

SECONDARY DOMAINS OF ACTIVITY
Biostatistics; Population genetics; Evolution; GWAS; Next-generation sequencing; Selection; Software development

DOMAINS OF APPLICATION

RESOURCES
quantiNemo
Hierfstat

KEY PAPERS IN 2017

Weir BS and Goudet J. A unified characterization of population structure and relatedness. Genetics

Paradis E *et al.* Linking genomics and population genetics with R. Mol Ecol Resour.

Jombart T *et al.* Lapp H. apex: phylogenetics with multiple genes. Mol Ecol Resour.

## Marc ROBINSON-RECHAVI

GROUP NAME **EVOLUTIONARY BIOINFORMATICS GROUP**
INSTITUTION UNIVERSITY OF LAUSANNE
CITY LAUSANNE

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

MAIN DOMAIN OF ACTIVITY
Evolution and phylogeny

SECONDARY DOMAINS OF ACTIVITY
Biocuration; Comparative genomics; Evo-Devo; Evolution; Functional genomics; Knowledgebase; Selection; Software development; Systems biology; Transcriptomics

DOMAINS OF APPLICATION

RESOURCES
Bgee
Selectome

KEY PAPERS IN 2017

Roux J *et al.* Selective constraints on coding sequences of nervous system genes are a major determinant of duplicate gene retention in vertebrates. Mol Biol Evol.

Daub JT *et al.* Detection of pathways affected by positive selection in primate lineages ancestral to humans. Mol Biol Evol.

Maitre D *et al.* Sex differentiation in grayling *(Salmonidae)* goes through an all-male stage and is delayed in genetic males who instead grow faster. Scientific Reports

## Nicolas SALAMIN

GROUP NAME **COMPUTATIONAL PHYLOGENETICS GROUP**
INSTITUTION UNIVERSITY OF LAUSANNE
CITY LAUSANNE

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

MAIN DOMAIN OF ACTIVITY
Evolution and phylogeny

SECONDARY DOMAINS OF ACTIVITY
Comparative genomics; Evolution; Functional genomics; Machine learning; Next-generation sequencing; Phylogeny; Selection; Software development; Transcriptomics

DOMAINS OF APPLICATION

RESOURCE
coevweb

KEY PAPERS IN 2017

Meyer X *et al.*
Accelerating Bayesian inference for evolutionary biology models.
Bioinformatics

Davydov I *et al.*
State aggregation for fast likelihood computations in molecular evolution.
Bioinformatics

Meyer X *et al.*
Scheduling finite difference approximations for DAG-modeled (...). Proceedings of the Platform for Advanced Scientific Computing Conference

## Tanja STADLER

GROUP NAME **COMPUTATIONAL EVOLUTION GROUP**
INSTITUTION ETH ZURICH, D-BSSE
CITY ZURICH, BASEL

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

MAIN DOMAIN OF ACTIVITY
Evolution and phylogeny

SECONDARY DOMAINS OF ACTIVITY
Biostatistics; Evolution; Infectious diseases; Next-generation sequencing; Phylogeny; Single-cell biology; Software development

DOMAINS OF APPLICATION

RESOURCES
BEAST add-ons

KEY PAPERS IN 2017

Barido-Sottani J *et al.*
BEAST—A community teaching material resource for BEAST 2,
Systematic Biology

Müller NF *et al.*
The structured coalescent and its approximations.
Mol Biol Evol.

Rasmussen DA *et al.*
Phylodynamics on local sexual contact networks.
PLoS Comput Biol.

## Daniel WEGMANN

GROUP NAME **STATISTICAL AND COMPUTATIONAL EVOLUTIONARY BIOLOGY GROUP**
INSTITUTION UNIVERSITY OF FRIBOURG
CITY FRIBOURG

When observing nature, it is easy to be impressed by the huge diversity seen on any biological scale. Our primary aim is to better understand the underlying evolutionary and ecological processes that have been shaping this diversity over the course of evolution on our planet. To achieve this, we design and evaluate new statistical and computational approaches to infer complex evolutionary histories. For this, we develop and apply machine learning algorithms, with a particular focus on likelihood-free methods. We then apply these approaches to the wealth of data currently being generated.

MAIN DOMAIN OF ACTIVITY
Evolution and phylogeny

SECONDARY DOMAINS OF ACTIVITY
Biostatistics; Population genetics; Evolution; GWAS; Machine learning; Next-generation sequencing; Paleogenomics; Selection; Software development; Human genetics

DOMAINS OF APPLICATION

RESOURCES
ABCtoolbox
ATLAS
RASPberry

KEY PAPERS IN 2017

Gros-Balthazard M *et al.* The discovery of wild date palms in Oman reveals a complex domestication history involving centers in the Middle East and Africa.
Current Biology

Kousathanas A *et al.*
Inferring heterozygosity from ancient and low coverage genomes.
Genetics

Duchen P *et al.*
Inference of evolutionary jumps in large phylogenies using Lévy processes.
Systematic Biology

## Bruno CORREIA

GROUP NAME **LABORATORY OF PROTEIN DESIGN AND IMMUNOENGINEERING**
INSTITUTION EPFL
CITY LAUSANNE

The Laboratory of Protein Design & Immunoengineering started in March 2015. We have established both the computational and experimental branches of our research.We currently perform our computational calculations at the High Performance Computing facility at the EPFL. The experimental arm of the laboratory is prepared to perform molecular biology, protein expression and purification, and protein biophysical characterization.

MAIN DOMAIN OF ACTIVITY
Structural biology

SECONDARY DOMAINS OF ACTIVITY
Biochemistry; Immunology; Infectious diseases; Protein structure and interaction; Vaccines

DOMAINS OF APPLICATION

RESOURCES
Rosetta Fold From Loops
TopoBuilder
SEEDER

KEY PAPERS IN 2017

Matthews ML *et al.*
Chemoproteomic profiling and discovery of protein electrophiles in human cells.
Nat Chem.

Parker CG *et al.*
Ligand and target discovery by fragment-based screening in human cells.
Cell

## Matteo DAL PERARO

GROUP NAME **LABORATORY FOR BIOMOLECULAR MODELING**
INSTITUTION EPFL
CITY LAUSANNE

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

MAIN DOMAIN OF ACTIVITY
Structural biology

SECONDARY DOMAINS OF ACTIVITY
Biochemistry; Drug resistance; Lipids; Molecular dynamics; Protein structure and interaction; Software development

DOMAINS OF APPLICATION

RESOURCES
pow$^{er}$
LipidBuilder

KEY PAPERS IN 2017

Tamò G *et al.*
Disentangling constraints using viability evolution principles in integrative modeling of macromolecular assemblies.
Sci Rep.

Abriata LA *et al.*
Signal sensing and transduction by histidine kinases as unveiled through studies on a temperature sensor.
Acc Chem Res.

Reidenbach AG *et al.*
Conserved cipid and small-molecule modulation of COQ8 reveals regulation of the ancient kinase-like UbiB family.
Cell Chem Biol.

## Olivier MICHIELIN
## Vincent ZOETE

GROUP NAME **MOLECULAR MODELLING GROUP**
INSTITUTION UNIVERSITY OF LAUSANNE
CITY LAUSANNE

The Molecular Modelling Group studies molecular interactions using modelling techniques such as homology modelling, molecular dynamics, docking and free energy simulations. Our main activity consists in developing and applying methods for computer-aided protein engineering and drug design. Notably, we develop novel inhibitors of important targets for cancer therapy and design optimized proteins, such as T-cell receptors, for cancer immunotherapy. We develop and maintain web tools for drug design, such as SwissDock, SwissTargetPrediction and SwissADME. We also run the Protein Modelling Facility of the University of Lausanne.

MAIN DOMAIN OF ACTIVITY
Structural biology

SECONDARY DOMAINS OF ACTIVITY
Drug design; Immunology; Molecular dynamics; Oncology; Precision therapy; Protein structure and interaction; Software development; Systems biology

DOMAIN OF APPLICATION

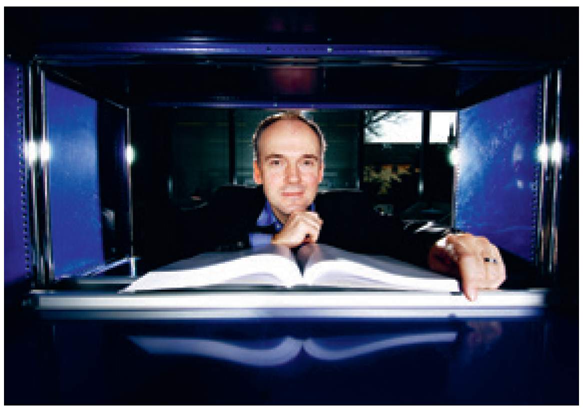
RESOURCES
SwissDock
SwissTargetPrediction
SwissADME

KEY PAPERS IN 2017

Daina A *et al.*
SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small (...).
Sci Rep.

Chaskar P *et al.*
On-the-Fly QM/MM docking with attracting cavities.
J Chem Inf Model.

Röhrig UF *et al.*
The binding mode of N-hydroxyamidines to indoleamine 2,3-dioxygenase 1 (IDO1).
Biochemistry



## Torsten SCHWEDE

GROUP NAME **COMPUTATIONAL STRUCTURAL BIOLOGY**
INSTITUTION UNIVERSITY OF BASEL
CITY BASEL

In the Computational Structural Biology (CSB) Group, we focus on the development of methods and algorithms to model, simulate and analyse three-dimensional protein structures and their molecular properties in order to apply these techniques to the understanding of biological processes at a molecular level. Our main emphasis is on homology modelling – using evolutionary information to model protein tertiary and quaternary structures. Applications in biomedical research include the study of protein-ligand interactions, drug discovery, structure-guided protein engineering, and the interpretation of disease-causing mutations.

MAIN DOMAIN OF ACTIVITY
Structural biology

SECONDARY DOMAINS OF ACTIVITY
Data mining; Drug design; Knowledgebase; Machine learning; Protein structure and interaction; Software development; Homology modelling; Benchmarking; Model quality; Protein engineering; Biotechnology

DOMAINS OF APPLICATION

RESOURCES
SWISS-MODEL
CAMEO
Model Archive

KEY PAPERS IN 2017

Bertoni M *et al.*
Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology.
Sci Rep.

Bienert S *et al.*
The SWISS-MODEL Repository - new features and functionality.
Nucleic Acids Res.

Haas J *et al.*
Continuous automated model evaluation (CAMEO) complementing the critical assessment of structure prediction in CASP12.
Proteins

## Bastien CHOPARD

GROUP NAME **SCIENTIFIC AND PARALLEL COMPUTING GROUP**
INSTITUTION UNIVERSITY OF GENEVA
CITY GENEVA

In the Scientific and Parallel Computing (SPC) Group, we develop new algorithms and methods to better understand and/or predict various phenomena in biology. We focus on multiscale modelling and computing, high-performance computing, cellular automata, lattice Botzmann methods, multi-agent systems, and optimizing techniques and machine learning. A core activity of our group is the modelling and simulation of complex systems. In bioinformatics, the objective is to develop advanced numerical methodologies to model biological processes.

MAIN DOMAIN OF ACTIVITY
Systems biology

SECONDARY DOMAINS OF ACTIVITY
Data mining; Evo-Devo; Proteomics; Software development; Systems biology

DOMAINS OF APPLICATION

RESOURCE
Terminus

KEY PAPERS IN 2017

Merzouki A *et al.*
Influence of cell mechanics and proliferation on the buckiling of simulated tissues using a vertex model.
NACO

Li S *et al.*
Model for pressure drop and flow deflection in the numerical simulation of stents in aneurysm.
Int J Numer Method Biomed Eng.

Dutta R *et al.*
Parameter estimation of platelets deposition: Approximate Bayesian computation with high performance computing.
arXiv

## Giovanni CIRIELLO

GROUP NAME **COMPUTATIONAL SYSTEMS ONCOLOGY**
INSTITUTION UNIVERSITY OF LAUSANNE
CITY LAUSANNE

The Computational Systems Oncology lab integrates algorithmic design, numerical modelling, and molecular biology approaches to address relevant questions in cancer biology and therapeutics. We explore single and combinations of genetic and epigenetic alterations that emerge and are selected during cancer evolution, to understand cancer driving mechanisms and inform precision medicine approaches.

MAIN DOMAIN OF ACTIVITY
Systems biology

SECONDARY DOMAINS OF ACTIVITY
Genome structure; Epigenetics; Evolution; Oncology; Precision therapy; Signaling pathways

DOMAIN OF APPLICATION

RESOURCE
SELECT

KEY PAPERS IN 2017

Mina M *et al.*
Conditional Selection of Genomic Alterations Dictates Cancer Evolution and Oncogenic Dependencies.
Cancer Cell

Oricchio E *et al.*
Genetic and epigenetic inactivation of SESTRIN1 controls mTORC1 and response to EZH2 inhibition in follicular lymphoma.
Science Trans Med.

## Manfred CLAASSEN

GROUP NAME **COMPUTATIONAL SINGLE CELL BIOLOGY GROUP**
INSTITUTION ETH ZURICH
CITY ZURICH

Our research in the Computational Single Cell Biology Group aims to elucidate the composition of heterogeneous cell populations and how these implement function in the context of cancer and immune biology. To accomplish this task, we build on concepts from statistics, machine learning and mathematical optimization to develop probabilistic approaches to describe biological systems, learn these descriptions from data, and design experiments to validate hypotheses following computational analyses. Our research can be used to pinpoint therapeutic targets with a view to designing drugs.

MAIN DOMAIN OF ACTIVITY
Systems biology

SECONDARY DOMAINS OF ACTIVITY
Experimental biology; Immunology; Machine learning; Oncology; Precision therapy; Signaling pathways; Single-cell biology; Systems biology; Systems modelling; Transcriptomics

DOMAINS OF APPLICATION

RESOURCES
CellCnn
Reactionet lasso
TreeTop

KEY PAPERS IN 2017

Arvaniti E and Claassen M.
Sensitive detection of rare disease-associated cell subsets via representation learning.
Nat Commun.

## Rudiyanto GUNAWAN

GROUP NAME **CHEMICAL AND BIOLOGICAL SYSTEMS ENGINEERING LABORATORY**
INSTITUTION ETH ZURICH
CITY ZURICH

In the Chemical and Biological Systems Engineering Laboratory, we develop tools for systems modelling and the analysis of chemical and biological networks. Our mission is to create enabling theories and computational methods for the generation of systems insights, as well as for understanding and acquiring knowledge in chemical, biological and medical applications. Our research spans multiple length and time-scales of cell biology, from gene/ signalling/metabolic networks in single cells to the aging process in human and cell culture bioreactors in the pharmaceutical industry.

MAIN DOMAIN OF ACTIVITY
Systems biology

SECONDARY DOMAINS OF ACTIVITY
Genetic networks; Glycomics; Machine learning; Metabolomics; Microarrays; Protein networks; Signaling pathways; Single-cell biology; Systems biology; Systems modelling

DOMAINS OF APPLICATION

RESOURCES
SINCERITIES
Glycosylation flux analysis

KEY PAPERS IN 2017

Hutter S *et al.* Glycosylation flux analysis reveals dynamic changes of intracellular glycosylation flux distribution in Chinese Hamster Ovary fed-batch cultures.
Metab Eng.

Gunawan R and Hutter S. Assessing and resolving model misspecifications in metabolic flux analysis.
Bioengineering

Liu Y and Gunawan R. Bioprocess optimization under uncertainty using ensemble modeling.
J Biotech.

## Dagmar IBER

GROUP NAME **COMPUTATIONAL BIOLOGY GROUP**
INSTITUTION ETH ZURICH, D-BSSE
CITY ZURICH & BASEL

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

MAIN DOMAIN OF ACTIVITY
Systems biology

SECONDARY DOMAINS OF ACTIVITY
Experimental biology; Genetic networks; Next-generation sequencing; Precision therapy; Signaling pathways; Single-cell biology; Software development; Systems biology; Systems modelling In silico organogenesis; Mathematical modelling

DOMAINS OF APPLICATION

RESOURCES
LBIBCell
MOCCASIN

KEY PAPERS IN 2017

Boareto M *et al.*
Differential interactions between Notch and ID factors control neurogenesis by modulating Hes factor autoregulation.
Development

Vollmer J *et al.*
Growth control in the Drosophila eye disc by the cytokine unpaired.
Development

Vollmer J *et al.*
Growth and size control during development.
Open Biol.

## Christian MAZZA

GROUP NAME **BIOMATHEMATICS AND COMPUTATIONAL BIOLOGY GROUP**
INSTITUTION UNIVERSITY OF FRIBOURG
CITY FRIBOURG

Mathematical modelling is becoming more and more instrumental in life sciences; the data complexity and the high number of interacting components, from molecules to animals, render intuitive reasoning very difficult. The idea consists in formulating mathematically how certain biological units affect each other and how these interactions affect the whole system. Such quantitative approaches produce mathematical models, which are then implemented using computer software. We study phyllotaxis by trying, for example, to explain the formation of geometrically regular patterns in plants, such as spirals in sunflowers.

MAIN DOMAIN OF ACTIVITY
Systems biology

SECONDARY DOMAINS OF ACTIVITY
Biostatistics; Evolution; Genetic networks; Signaling pathways; Synthetic biology; Systems biology

DOMAINS OF APPLICATION

KEY PAPERS IN 2017

Dougoud M *et al.* Ultrasensitivity and sharp threshold theorems for multisite systems. Journal of Physics A: Mathematical and Theoretical

Roggo C. *et al.* Quantitative chemical biosensing by bacterial chemotaxis in microfluidic chips. Environ Microbiol.

## Michel MILINKOVITCH

GROUP NAME **ARTIFICIAL & NATURAL EVOLUTIONARY DEVELOPMENT OF COMPLEXITY GROUP**
INSTITUTION UNIVERSITY OF GENEVA
CITY GENEVA

We combine Evolutionary and Developmental Biology as well as genomic approaches with the study of physical processes to understand the mechanisms generating life's complexity and diversity. The self-organizational capabilities of tissues are pertinent to EvoDevo at multiple spatial scales. We aim to understand the interactions between physical (e.g., mechanics, reaction diffusion) and biological (e.g., cell signalling, proliferation, migration) parameters, which generate patterns and shapes during development. Our multidisciplinary team includes Evolutionary and Developmental Biologists, Computer Scientists, Engineers, and Physicists.

MAIN DOMAIN OF ACTIVITY
Systems biology

SECONDARY DOMAINS OF ACTIVITY
Comparative genomics; Evo-Devo; Evolution; Experimental biology; Machine learning; Next-generation sequencing; Phylogeny; Software development; Systems modelling; Transcriptomics; Biophysics

DOMAIN OF APPLICATION

RESOURCE
MetaPIGA
Reptilomics

KEY PAPERS IN 2017

Manukyan L *et al.* A living mesoscopic cellular automaton made of skin scales. Nature

Daljit Singh J *et al.* Bifurcation analysis of reaction diffusion systems on arbitrary surfaces. Bull Math Biol.

## Sara MITRI

GROUP NAME **EVOLUTIONARY MICROBIOLOGY GROUP**
INSTITUTION UNIVERSITY OF LAUSANNE
CITY LAUSANNE

Our group focuses on studying social interactions in microbial communities and how they evolve. We follow an interdisciplinary approach, where we quantify and characterize interactions within small bacterial communities in the lab, and then use computational and mathematical models to explain and predict how these interactions will change over ecological and evolutionary time-scales.

MAIN DOMAIN OF ACTIVITY
Systems biology

SECONDARY DOMAINS OF ACTIVITY
Comparative genomics; Evolution; Experimental biology; Metabolomics; Selection; Synthetic biology; Systems modelling

DOMAINS OF APPLICATION

KEY PAPERS IN 2017

Niehus R *et al.* The evolution of siderophore production as a competitive trait. Evolution

## Félix NAEF

GROUP NAME **COMPUTATIONAL SYSTEMS BIOLOGY GROUP**
INSTITUTION EPFL
CITY LAUSANNE

Systems biology is aimed at a quantitative and dynamic understanding of cellular networks by combining experimental data with theoretical and computational methodologies. Our interest lies in the regulatory and cellular networks involved in oncogenic signalling, cell-cycle regulation, and molecular oscillators. Data obtained from technologies such as microarrays, chromatin-immunoprecipitation (ChIP) and genome sequencing are brought together to discover regulatory dependencies between genes and regulatory proteins involved in cell proliferation. One thematic focus is the study of biomolecular oscillators, in particular the circadian clock.

MAIN DOMAIN OF ACTIVITY
Systems biology

SECONDARY DOMAINS OF ACTIVITY
Functional genomics; Single-cell biology; Systems biology; Transcriptomics

DOMAIN OF APPLICATION

KEY PAPERS IN 2017

Yeung J *et al.* Transcription factor activity rhythms and tissue-specific chromatin interactions explain circadian gene (...). Genome Res.

Mauvoisin D *et al.* Circadian and feeding rhythms orchestrate the diurnal liver acetylome. Cell Rep.

Sobel JA *et al.* Transcriptional regulatory logic of the diurnal cycle in the mouse liver. PLoS Biol.

## Igor PIVKIN

GROUP NAME **SCIENTIFIC COMPUTING GROUP**
INSTITUTION UNIVERSITÀ DELLA SVIZZERA
CITY LUGANO

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

MAIN DOMAIN OF ACTIVITY
Systems biology

SECONDARY DOMAINS OF ACTIVITY
Cell mechanobiology; Computational modelling; Particle-based methods

DOMAIN OF APPLICATION

KEY PAPERS IN 2017

Lykov K *et al.* Probing eukaryotic cell mechanics via mesoscopic simulations. PLOS Comp Biol.

Christel S *et al.* Multi-omics reveal the lifestyle of the acidophilic (...). Applied Environ Microbiol.

## Jörg STELLING

GROUP NAME **COMPUTATIONAL SYSTEMS BIOLOGY GROUP**
INSTITUTION ETH ZURICH, D-BSSE
CITY ZURICH, BASEL

We develop and apply computational and – most recently – experimental methods to analyse and design complex cellular networks, with a focus on large-scale mechanistic approaches. The group comprises biologists, computer scientists, engineers, and mathematicians, who perform interdisciplinary research in systems and synthetic biology. We focus on developing and applying computational methods and mechanistic mathematical models to study complex cellular networks and their operating principles, and to enable their rational re-design. Applications rely on our experimental lab, which uses budding yeast as a model organism, and on collaboration.

MAIN DOMAIN OF ACTIVITY
Systems biology

SECONDARY DOMAINS OF ACTIVITY
Synthetic biology; Systems modelling

DOMAIN(S) OF APPLICATION

RESOURCES
metanetx.org (collaboration with Vital-IT)

KEY PAPERS IN 2017

Otero-Muras I. Chemical reaction network theory elucidates sources of multistability in interferon signaling. PLOS.

Lormeau C *et al.* Multi-objective design of synthetic biological circuits. IFAC

Hass H *et al.* Predicting ligand-dependent tumors from multi-dimensional signaling features. NPJ Syst Biol Appl.

## Mihaela ZAVOLAN

GROUP NAME **RNA REGULATORY NETWORKS GROUP**
INSTITUTION UNIVERSITY OF BASEL
CITY BASEL

The individual cells of a body exhibit a stunning diversity of phenotypes, despite carrying a largely identical genetic makeup. This is due to the distinct ways in which the same genetic information can be read, interpreted and translated into function. At the RNA Regulatory Networks (RRN) Group at the Biozentrum in Basel, we combine computational modelling with big data and experimental analysis to discover and understand the regulatory networks governing the interpretation of genetic information at the level of tissues and single cells.

MAIN DOMAIN OF ACTIVITY
Systems biology

SECONDARY DOMAINS OF ACTIVITY
Data mining; Experimental biology; Genetic networks; Machine learning; Next-generation sequencing; Single-cell biology; Software development; Systems biology; Systems modelling; Transcriptomics

DOMAIN(S) OF APPLICATION

RESOURCES
MIRZA
polyAsite

KEY PAPERS IN 2017

Gruber AJ *et al.* Discovery of physiological and cancer-related regulators of 3' UTR processing with KAPAC. bioRxiv

Riba A *et al.* Explicit Modeling of siRNA-Dependent On-and Off-Target Repression Improves the Interpretation (...). Cell Systems.

Gumienny R *et al.* High-throughput identification of C/D box snoRNA targets with CLIP and RiboMeth-seq. Nucleic Acids Res.

## Rémy BRUGGMANN

GROUP NAME **INTERFACULTY BIOINFORMATICS UNIT (IBU)**
INSTITUTION UNIVERSITY OF BERN
CITY BERN

In the Interfaculty Bioinformatics Unit of the University of Bern (IBU), we provide services and expertise to assist researchers in the three “Life-Science” Faculties (i.e. Sciences, Medicine, and VetSuisse) in data analysis and project planning for large-scale experiments (e.g. next-generation sequencing, genome assembly). Furthermore, we have our own research programme and collaborate on large and complex projects. We develop methods to analyse high-throughput data. We have a high-performance computing cluster and a data storage system that we use for our own research, collaborations and service projects.

MAIN DOMAIN OF ACTIVITY
Core facilities and competence centres

SECONDARY DOMAIN OF ACTIVITY
Biostatistics; Comparative genomics; Metagenomics; Next-generation sequencing; Single-cell biology; Systems biology; Transcriptomics; Precision medicine

DOMAIN(S) OF APPLICATION

RESOURCES
SetRank

KEY PAPERS IN 2017

Wüthrich D *et al.* The histidine decarboxylase gene cluster of *Lactobacillus parabuchneri* was gained by horizontal gene transfer and is mobile within the species. Front Microbiol.

Simillion C *et al.* Avoiding the pitfalls of gene set enrichment analysis with SetRank. BMC Bioinformatics

Boujon CL *et al.* Indication of cross-species transmission of astrovirus associated with encephalitis in sheep and cattle. Emerging Infectious Diseases

## Luciano CASCIONE

GROUP NAME **BIOINFORMATICS CORE UNIT**
INSTITUTION INSTITUTE OF ONCOLOGY RESEARCH
CITY BELLINZONA

The Bioinformatics Core Unit supports the research groups with computational and statistical services. Importantly, more than just a supporting role, we proactively identify and develop novel bioinformatics projects that can complement and in many cases drive our biologic research. We develop innovative data analysis tools, visualization software and database resources for genomics research in collaboration with the Dalle Molle Institute of Artificial Intelligence (IDSIA) and the SIB.

MAIN DOMAIN OF ACTIVITY
Core facilities and competence centres

SECONDARY DOMAIN OF ACTIVITY
Biomarkers; Biostatistics; Drug resistance; Epigenetics; Machine learning; Microarrays; Next-generation sequencing; Oncology; Software development; Transcriptomics

DOMAIN(S) OF APPLICATION

RESOURCES
PCSF
mBPCR

KEY PAPERS IN 2017

Murodzhon A *et al.* PCSF: An R-package for network-based interpretation of high-throughput data. PLoS Comput Biol.

Vázquez R *et al.* The bromodomain inhibitor OTX015 (MK-8628) exerts anti-tumor activity in triple-negative breast cancer models as single agent (...). Oncotarget

Akhmedov A *et al.* OmicsNet: integration of multi-omics data using path analysis in multilayer networks. BioRxiv

## Laurent FALQUET

GROUP NAME **BIOINFORMATICS UNRAVELLING GROUP (BUGFRI)**
INSTITUTION UNIVERSITY OF FRIBOURG
CITY FRIBOURG

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

MAIN DOMAIN OF ACTIVITY
Core facilities and competence centres

SECONDARY DOMAIN OF ACTIVITY
Comparative genomics; Epigenetics; Functional genomics; Genome structure; Genome reconstruction; Metagenomics

DOMAIN(S) OF APPLICATION

RESOURCES
PACMAN

KEY PAPERS IN 2017

Rychener L *et al.* *Clostridium chauvoei*, an Evolutionary Dead-End Pathogen. Front Microbiol.

Gros-Balthazard *et al.* The discovery of wild date palms in oman reveals a complex domestication history involving centers in the middle east and africa. Curr Biol.

Saavedra *et al.* Whole-genome sequence of a colombian *acinetobacter baumannii* strain, a coproducer of OXA-72 and OXA-255-like carbapenemases. Genome Announc.

## Robert IVANEK

GROUP NAME **DBM BIOINFORMATICS CORE FACILITY**
INSTITUTION UNIVERSITY OF BASEL AND UNIVERSITY HOSPITAL BASEL
CITY BASEL

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

MAIN DOMAIN OF ACTIVITY
Core facilities and competence centres

SECONDARY DOMAINS OF ACTIVITY
Biostatistics; Epigenetics; Functional genomics; Single-cell biology; Software development; Systems biology; Transcriptomics

DOMAIN OF APPLICATION

RESOURCE
Gviz

KEY PAPERS IN 2017

Alberti-Servera L *et al.* Single-cell RNA sequencing reveals developmental heterogeneity among early lymphoid progenitors. EMBO J.

Diepenbruck M *et al.* miR-1199-5p and Zeb1 function in a double-negative feedback loop potentially coordinating EMT and tumour metastasis. Nat Commun.

## Hubert REHRAUER

GROUP NAME **FUNCTIONAL GENOMICS CENTER ZURICH**
INSTITUTION ETH ZURICH, UNIVERSITY OF ZURICH
CITY ZURICH

We are dedicated to the processing, analysis and interpretation of next-generation sequencing data. We interact closely with research groups, and provide tailored comprehensive bioinformatics solutions. Additionally, we provide standard analysis pipelines for the most frequent research questions. We train researchers and bioinformaticians on various aspects of data analysis, and provide access to our computing infrastructure for running analyses.

MAIN DOMAIN OF ACTIVITY
Core facilities and competence centres

SECONDARY DOMAINS OF ACTIVITY
Biostatistics; Data management plan; Data mining; Functional genomics; GWAS; Machine learning; Next-generation sequencing; Single-cell biology; Software development; Transcriptomics

DOMAINS OF APPLICATION

RESOURCE
sushi

KEY PAPERS IN 2017

Hatakeyama M *et al.* Multiple hybrid *de novo* genome assembly of finger millet, an orphan allotetraploid crop. DNA Res.

Qi W *et al.* RNA-Seq data analysis: from raw data quality control to differential expression analysis. Methods Mol Biol.

Frank-Bertoncelj M *et al.* Epigenetically-driven anatomical diversity of synovial fibroblasts guides joint-specific fibroblast functions. Nat Commun.

## Marcel RIEDI

GROUP NAME **SERVICE AND SUPPORT FOR SCIENCE IT (S3IT)**
INSTITUTION UNIVERSITY OF ZURICH
CITY ZURICH

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

MAIN DOMAIN OF ACTIVITY
Core facilities and competence centres

SECONDARY DOMAINS OF ACTIVITY
Data management plan; Data mining; Infectious diseases; Machine learning; Next-generation sequencing; Protein structure and interaction; Proteomics; Software development; Systems biology

DOMAIN OF APPLICATION

RESOURCES
S3IT compute and storage infrastructure
ElastiCluster/GC3Pie
S3IT data manager

KEY PAPERS IN 2017

Graf U *et al.* Pramel7 mediates ground-state pluripotency through proteasomal-epigenetic combined pathways. Nat Cell Biol.

Ulman V *et al.* An objective comparison of cell-tracking algorithms. Nat methods

Schmid M.W. RNA-Seq data analysis protocol: combining in-house (...). In: Schmidt A. (eds) Plant Germline Development. Methods in Molecular Biology

## Bernd RINN

GROUP NAME **SCIENTIFIC IT SERVICES (SIS)**
INSTITUTION ETH ZURICH, D-BSSE
CITY ZURICH & BASEL

Our group in Scientific IT Services (SIS) is an interdisciplinary bioinformatics and scientific IT group, which develops computational tools and services. We provide high performance computing infrastructure and services, improve and port scientific software, develop data management solutions and provide associated services. We also integrate and operate data analysis pipelines, and provide training and consulting in databases, scientific software development, high-performance and cloud computing. With our services, we support life science and biomedical research, as well as other basic and applied research.

MAIN DOMAIN OF ACTIVITY
Core facilities and competence centres

SECONDARY DOMAINS OF ACTIVITY
Data management plan; Machine learning; Software development; Biochemistry; Data Pipelines; Data Visualization; Scientific Programming Support

DOMAINS OF APPLICATION

RESOURCES
openBIS
emzed
YouScope

KEY PAPERS IN 2017

Rosenberger G *et al.* Statistical control of peptide and protein error rates in large-scale targeted data-independent acquisition analyses. Nat methods

Singer J *et al.* NGS-pipe: a flexible, easily extendable, and highly configurable framework for NGS analysis. Bioinformatics

Graf U *et al.* Pramel7 mediates ground-state pluripotency through proteasomal-epigenetic combined pathways. Nat Cell Biol.

## Torsten SCHWEDE
## Thierry SENGSTAG

GROUP NAME **SCICORE**
INSTITUTION UNIVERSITY OF BASEL
CITY BASEL

sciCORE is a centre of competence in scientific computing located at the University of Basel. Over 170 research groups in the fields of bioinformatics, computational chemistry, physics, biology, medicine, and economics use our services. Our expertise ranges from high-performance computing infrastructure, large-scale storage resources, scientific software and databases to scientific consulting and education. We help scientists with computational needs achieve their research goals. We operate the IT infrastructure behind several key SIB bioinformatics resources and, as a BioMedIT node, we actively support Swiss biomedical research.

MAIN DOMAIN OF ACTIVITY
Core facilities and competence centres

SECONDARY DOMAINS OF ACTIVITY
Biostatistics; Data management plan; Data mining; Machine learning; Next-generation sequencing; Software development

DOMAINS OF APPLICATION

KEY PAPERS IN 2017

Heck A *et al.* Exome sequencing of healthy phenotypic extremes links TROVE2 to emotional memory and PTSD. Nat Human Behav.

Egli T *et al.* Exhaustive search for epistatic effects on the human methylome. Sci Rep.

Cabezón R *et al.* SPHYNX: an accurate density-based SPH method for astrophysical applications. Astronomy & Astrophysics

## Daniel STEKHOVEN

GROUP NAME **CLINICAL BIOINFORMATICS UNIT**
INSTITUTION ETH ZURICH
CITY ZURICH

The Clinical Bioinformatics Unit is the dry lab of the ETH Zurich core facility NEXUS Personalized Health Technologies. CBU offers customized bioinformatics and biostatistics services for projects in the field of biomedical research. We maintain close collaborations with hospitals in Switzerland, enabling and accelerating the use of state-of-the-art multi-omics data analysis. Based on a fee-for-service model, our aim is to meet individual project goals and establish close interaction with our customers. Our strongest asset is the experience in reporting comprehensive bioinformatics results to clinicians, especially in oncology.

MAIN DOMAIN OF ACTIVITY
Core facilities and competence centres

SECONDARY DOMAINS OF ACTIVITY
Biostatistics; Drug resistance; Oncology; Precision therapy; Proteomics; Signaling pathways; Software development; Transcriptomics; Molecular diagnostics

DOMAINS OF APPLICATION

RESOURCES
richnet
MTB

KEY PAPERS IN 2017

Singer J et al. NGS-pipe: a flexible, easily extendable and highly configurable framework for NGS analysis. Bioinformatics

Harlander S et al. Combined mutation in Vhl, Trp53 and Rb1 causes clear cell renal cell carcinoma in mice. Nature Medicine

Singer J *et al.* Bioinformatics for precision oncology. Brief Bioinform.

## Ioannis XENARIOS

GROUP NAME **VITAL-IT**
INSTITUTION SIB LAUSANNE
CITY LAUSANNE

With a multidisciplinary team of scientists and technical staff, at Vital-IT we maintain a competence centre in bioinformatics and computational biology. Our infrastructure currently spreads across six institutions that maintain bio-technological platforms: SIB, the Universities of Geneva, Lausanne, Fribourg and Bern, and EPFL. At Vital-IT we enable scientists to access state-of-the-art computational infrastructure (processing, storage and archiving) as well as expertise in data analysis and algorithmic development.

MAIN DOMAIN OF ACTIVITY
Core facilities and competence centres

SECONDARY DOMAINS OF ACTIVITY
Genome structure; Data management plan; Data mining; Genome reconstruction; Machine learning; Mass spectroscopy; Next-generation sequencing; Software development; Systems biology; Systems modelling: Advanced software optimization

DOMAINS OF APPLICATION

RESOURCES
MetaNetX
OpenFlu EMPRES-i
SourceData platform

KEY PAPERS IN 2017

Schmid-Siegert E *et al.* Low number of fixed somatic mutations in a long-lived oak tree. Nat plants

Solimena M *et al.* Systems biology of the IMIDIA biobank from organ donors and pancreatectomised patients defines a novel transcriptomic signature of islets (...). Diabetologia

Liechti R *et al.* SourceData: a semantic platform for curating and searching figures. Nat methods

## Karsten BORGWARDT

GROUP NAME **MACHINE LEARNING AND COMPUTATIONAL BIOLOGY LAB**
INSTITUTION ETH ZURICH
CITY ZURICH

Our lab acts as the bridge between big data analysis and biomedical research. We develop novel data mining algorithms to detect patterns and statistical dependencies in large datasets from the fields of biology and medicine. Our major goals are twofold: 1) to enable the automatic generation of new knowledge from big data through machine learning, and 2) to gain an understanding of the relationship between biological systems and their molecular properties. Such an understanding is of fundamental importance for personalized medicine, which tailors medical treatment to the molecular properties of a person.

MAIN DOMAIN OF ACTIVITY
Text mining and machine learning

SECONDARY DOMAINS OF ACTIVITY
Biomarkers; Biostatistics; Data mining; GWAS; Machine learning; Software development; Systems biology

DOMAIN OF APPLICATION

RESOURCES
easyGWAS
R and Python packages for graph comparison
AraPheno

KEY PAPERS IN 2017

Grimm D *et al.* easyGWAS: a cloud-based platform for comparing the results of genome-wide association studies. Plant Cell

Llinares-Lopez F *et al.* Genome-wide genetic heterogeneity discovery with categorical covariates. Bioinformatics

Sugiyama M *et al.* graphkernels: R and Python packages for graph comparison. Bioinformatics

## Carlos PEÑA

GROUP NAME **COMPUTATIONAL INTELLIGENCE FOR COMPUTATIONAL BIOLOGY (CI4CB)**
INSTITUTION HEIG-VD
CITY YVERDON-LES-BAINS

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

MAIN DOMAIN OF ACTIVITY
Text mining and machine learning

SECONDARY DOMAINS OF ACTIVITY
Biomarkers; Data mining; Machine learning; Metagenomics; Software development; Systems modelling

DOMAINS OF APPLICATION

RESOURCE
FUGE

KEY PAPERS IN 2017

Leite D *et al.* Computational prediction of host-pathogen interactions through omics data analysis (...), IWBBIO 2017; Bioinformatics and Biomedical Engineering

Mungloo Z *et al* A meta-review of feature selection techniques in the context of microarray data, IWBBIO 2017; Bioinformatics and Biomedical Engineering

Despraz J *et al.* Towards a better understanding of deep neural networks representations using deep generative networks.

## Gunnar RÄTSCH

GROUP NAME **BIOMEDICAL INFORMATICS**
INSTITUTION ETH ZURICH
CITY ZURICH

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

MAIN DOMAIN OF ACTIVITY
Text mining and machine learning

SECONDARY DOMAINS OF ACTIVITY
Comparative genomics; Genome structure; Machine learning; Metagenomics; Next-generation sequencing; Oncology; Software development; Text mining; Transcriptomics; eHealth Records; Deep sequencing data

DOMAIN OF APPLICATION

RESOURCES
RiboDiff
SplAdder
BRCA Exchange

KEY PAPERS IN 2017

Locatello F. *et al.* Greedy algorithms for cone constrained optimization with convergence guarantees. NIPS

Hyland S L *et al.* Real-valued (Medical) Time Series Generation with Recurrent conditional GANs. arXiv

Zhong Y. *et al.* RiboDiff: detecting changes of mRNA translation efficiency from ribosome footprints. Bioinformatics

## Fabio RINALDI

GROUP NAME **BIOMEXT: BIOMEDICAL INFORMATION EXTRACTION**
INSTITUTION UNIVERSITY OF ZURICH
CITY ZURICH

The BioMeXT group specializes in Information Extraction from the biomedical literature and other textual sources. Information extraction consists in automatically extracting structured information from textual documents. We specialize in the extraction of domain-specific entities (such as genes, drugs, diseases), and their semantic relationships (e.g. gene-disease associations). Our tools are often evaluated through participation in community-run evaluation challenges (e.g. BioCreAtIvE). We provide an environment for Assisted Curation (ODIN), which is used in the curation pipeline of the RegulonDB database, in an NIH-funded project.

MAIN DOMAIN OF ACTIVITY
Text mining and machine learning

SECONDARY DOMAINS OF ACTIVITY
Biocuration; Knowledgebase; Machine learning; Text mining

DOMAINS OF APPLICATION

RESOURCE
Bio Term Hub

KEY PAPERS IN 2017

Basaldella M *et al.* Entity recognition in the biomedical domain using a hybrid approach. J Biomed Semantics

Balderas-Martínez YI *et al.* Improving biocuration of microRNAs in diseases: a case study in idiopathic pulmonary fibrosis. Database

Rinaldi F *et al.* Strategies towards digital and semi-automated curation in RegulonDB. Database

## Patrick RUCH

GROUP NAME **TEXT MINING**
INSTITUTION HES-SO - GENEVA SCHOOL OF BUSINESS ADMINISTRATION (HEG)
CITY GENEVA

In the Text Mining Group, we carry out activities in semantic interoperability and text analytics applied to the health and life sciences. Previously hosted by the Radiology and Medical Informatics Department of the University Hospitals of Geneva, our group moved to the University of Applied Sciences Geneva (HES-SO – HEG Geneva) in 2008. We develop text mining solutions to support both the annotation of curated databases and the work of a wide range of biomedical professionals from drug designers to clinicians. We thus develop specific biomedical decision-support systems (hypothesis generators, search engines, classifiers, etc.).

MAIN DOMAIN OF ACTIVITY
Text mining and machine learning

SECONDARY DOMAINS OF ACTIVITY
Data management plan; Data mining; Drug metabolism; Functional genomics; Gene ontology; Knowledgebase; Machine learning; Oncology; Precision therapy; Software development

DOMAIN OF APPLICATION

RESOURCES
GOCat
neXtA5
EAGLi

KEY PAPERS IN 2017

Teodoro D *et al.* Improving average ranking precision in user searches for biomedical research datasets. Database

Mottin L *et al.* Triage by ranking to support the curation of protein interactions. Database

Venkatesan A *et al.* SciLite: a platform for displaying text-mined annotations as a means to link research articles with biological data. Wellcome Open Res.

## NEW GROUPS AND GROUP LEADERS

## Katja BAERENFALLER

GROUP NAME **MOLECULAR ALLERGOLOGY**
INSTITUTION SIAF
CITY DAVOS

In the Molecular Allergology group we are aiming to unravel the molecular basis of allergic diseases, antigen presentation, immunotherapies, immunotolerance and tolerance breakage with a combination of functional genomics techniques, including proteomics, peptidomics and immunopeptidomics. The acquired large-scale molecular data are analysed with the aim to identify molecular markers with which risk factors, different disease phenotypes, their underlying pathogenic mechanisms, and different patient response groups can be distinguished.

MAIN DOMAIN OF ACTIVITY
Proteins and proteomes

SECONDARY DOMAIN OF ACTIVITY
Biomarkers; Biostatistics; Data mining; Functional genomics; Immunology; Peptidomics; Proteomics; Systems biology; Transcriptomics; Mass spectrometry

DOMAIN OF APPLICATION

## Joshua PAYNE

GROUP NAME **COMPUTATIONAL BIOLOGY**
INSTITUTION ETH ZURICH
CITY ZURICH

We are interested in understanding the design constraints, robustness, and evolution of gene regulatory systems, particularly at the level of transcription, using both modeling and data-driven approaches.

MAIN DOMAIN OF ACTIVITY
Evolution and phylogeny

SECONDARY DOMAINS OF ACTIVITY
Evolution; Functional genomics; Genetic networks; Microarrays; Next-generation sequencing; Systems biology; Systems modelling

DOMAINS OF APPLICATION

## Robert WATERHOUSE

GROUP NAME **EVOLUTIONARY-FUNCTIONAL GENOMICS**
INSTITUTION UNIVERSITY OF LAUSANNE
CITY LAUSANNE

The group's research is focused on elucidating interactions between gene evolution and gene function through developing computational approaches to interrogate evolutionary and functional genomics data. We are developing quantifications of gene evolutionary dynamics and functional properties to build models that link the underlying genetics to observable biological features. Our research projects focus on insects and other arthropods as their countless adaptations to exploit ecological niches mean that they are ideal for investigating how conservation or divergence of functional genomic elements give rise to the splendour of animal biology.

MAIN DOMAIN OF ACTIVITY
Genes and genomes

SECONDARY DOMAINS OF ACTIVITY
Biocuration; Comparative genomics; Evolution; Functional genomics; Genome structure; Phylogeny; Selection; Software development; Systems biology; Transcriptomics

DOMAIN OF APPLICATION

## Luciano CASCIONE

GROUP NAME **BIOINFORMATICS CORE UNIT**
INSTITUTION INSTITUTE OF ONCOLOGY RESEARCH
CITY BELLINZONA

Luciano is taking over the BCU group from Ivo Kwee.
SEE P. 70

## Frédéric SCHÜTZ

GROUP NAME **BIOINFORMATICS CORE FACILITY (BCF)**
INSTITUTION UNIVERSITY OF LAUSANNE
CITY LAUSANNE

Frédéric is joining Mauro Delorenzi as co-group leader.
SEE P. 53

CHAPTER 4

# Focus on the Institution

Take a backstage tour of the SIB Swiss Institute of Bioinformatics : governance, partner institutions, finances, members and staff.

# Organization

## A ROBUST GOVERNANCE

As a non-profit foundation unifying bioinformatics across Switzerland and with 20 partner institutions, SIB has a robust governance mechanism ensuring both its scientific independence and optimal internal functioning.

A Foundation Council, a Board of Directors, an external Scientific Advisory Board and a Council of Group Leaders complement the management and support teams of the Institute.

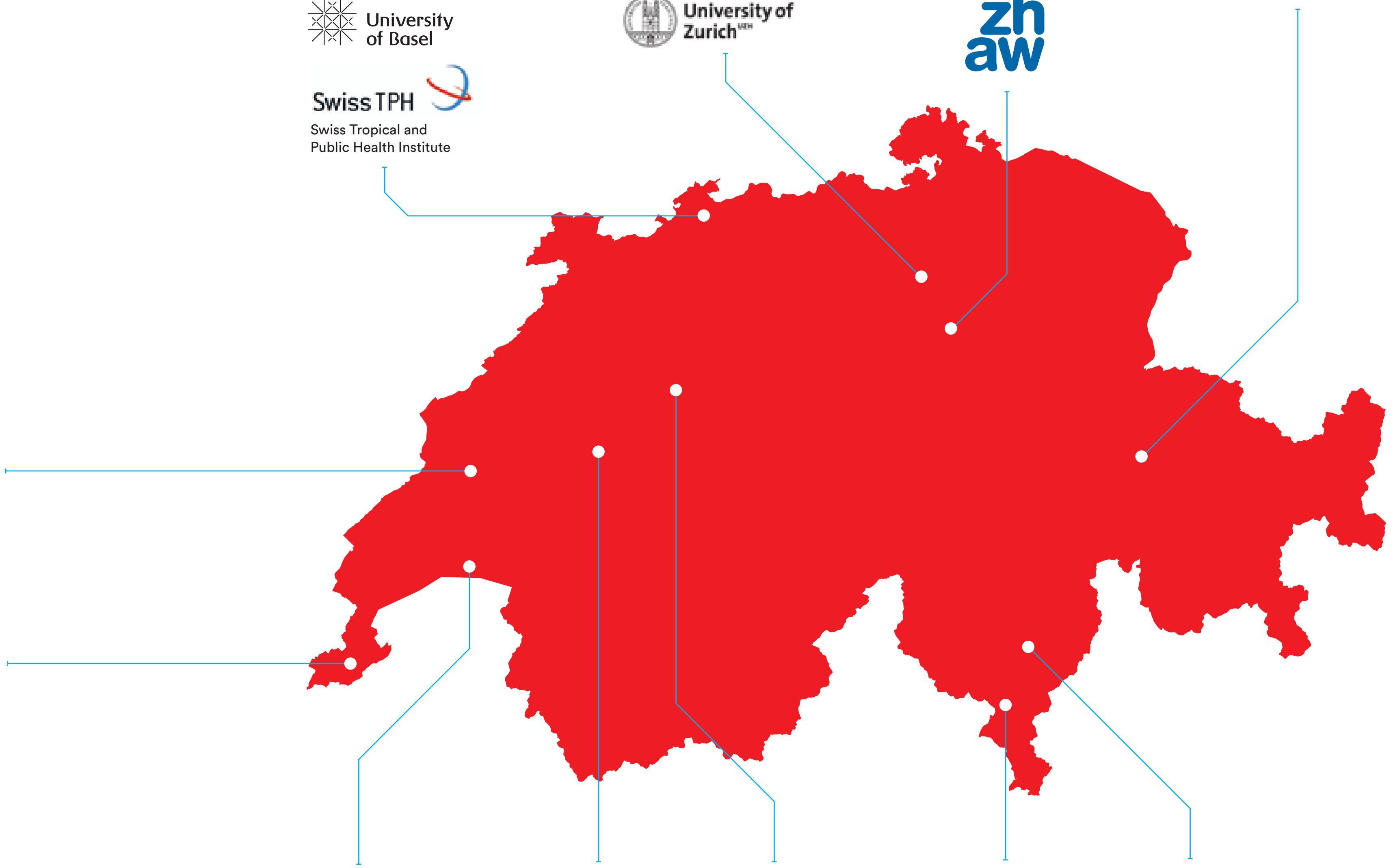
SIB Internal Groups, which are fully staffed by SIB, include two of its research and service groups, Swiss-Prot and Vital-IT, as well as the Clinical Bioinformatics, Personalized Health Informatics, Training and Technology Groups. The latter is in charge of coordinating and optimizing technical activities across SIB.

### SIB, a national bioinformatics organization model

SIB has become a reference throughout Europe with regard to its structure and organization. The ingredients to 'make an SIB' were the focus of a 2017 publication by Vivienne Baillie Gerritsen (science writer at SIB), Patricia Palagi (team leader of SIB's Training Group) and Christine Durinx (SIB's associate director), in the journal *Briefings in Bioinformatics*.

**Becoming an SIB Group Leader: some criteria**
Aspiring Group Leaders in the field of bioinformatics must have a professional appointment in one of SIB's 20 partner institutions, or be a senior scientist leading her/his own group. Since SIB does not fund Group Leaders, they also must have a sufficient level of financial independence, such as that provided by research grants.

## SIB'S PARTNER INSTITUTIONS



FMI
Friedrich Miescher Institute for Biomedical Research

University of Basel

Swiss TPH
Swiss Tropical and Public Health Institute

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

University of Zurich

Agroscope

Zurich University of Applied Sciences
zh aw

SIAF
Swiss Institute of Allergy and Asthma Research

heig-vd
HAUTE ÉCOLE D'INGÉNIERIE ET DE GESTION DU CANTON DE VAUD
www.heig-vd.ch
School of Business and Engineering Vaud

UNIVERSITÉ DE GENÈVE
University of Geneva

HUG
Hôpitaux Universitaires Genève

espeRare

h e g
Haute école de gestion de Genève
Geneva School of Business Administration

EPFL
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Unil
UNIL | Université de Lausanne

LUDWIG CANCER RESEARCH

UNI FR
UNIVERSITÉ DE FRIBOURG
UNIVERSITÄT FREIBURG

u b
UNIVERSITÄT BERN

USI
Università della Svizzera italiana

IOR
Institute of Oncology Research

# LIST OF GOVERNING BODIES

## The Foundation Council

Each of SIB's partner institutions is represented in this Council.

**President**
Prof. Felix Gutzwiller
Former Senator

**Founding Members**
Prof. Ron Appel
SIB Executive Director, University of Geneva

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

Dr Philipp Bucher
Group Leader, SIB and EPFL

Prof. Denis Hochstrasser
Vice-Rector, University of Geneva

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

*Ex officio Members*
Mr Thomas Baenninger
Chief Financial Officer
Ludwig Institute for Cancer Research

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

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

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

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

Prof. Daniel Candinas
Vice-Rector, University of Bern

Prof. Carlo Catapano
Director, IOR Institute of Oncology Research

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

Prof. Boas Erez
Rector, Università della Svizzera Italiana

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

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

Prof. Antoine Geissbühler
Head of eHealth and Telemedicine Division, Geneva University Hospitals

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

Prof. Denis Hochstrasser
Vice-Rector, University of Geneva

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

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

Dr Caroline Kant
Executive Director, EspeRare Foundation Switzerland

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

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

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

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

Prof. Michael Schaepman
Vice-President for Veterinary Medicine and Natural Sciences, University of Zurich

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

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

## The Board of Directors (BoD)

The BoD consists of two Group Leaders elected jointly by the Council of Group Leaders and the BoD, two external members elected by the Foundation Council on the recommendation of the BoD, and the Executive Director. Members of the BoD are appointed for a renewable five-year period.

Prof. Ron Appel
SIB Executive Director, University of Geneva

Ms Martine Brunschwig Graf
Former National Councillor

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

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

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

## The Scientific Advisory Board (SAB)

The SAB is made up of at least five members, who must be internationally renowned scientists from the Institute's fields of activity.

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

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

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

Prof. Melissa Haendel
Director of the Ontology Development Group, Oregon Health & Science University, Portland, USA

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

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

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

## Council of Group Leaders

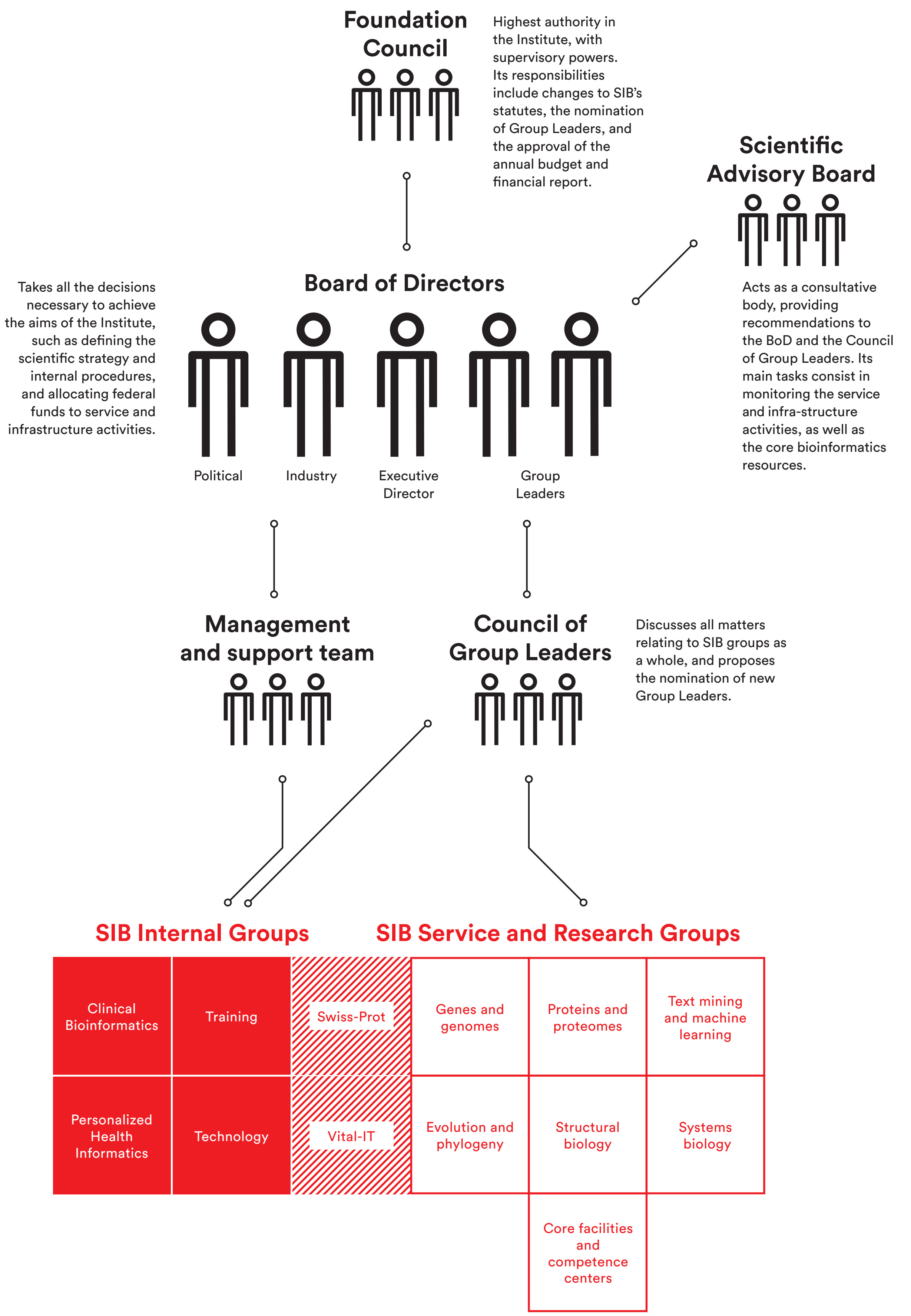
The Council consists of the Group Leaders, the Affiliate Group Leaders and the Executive Director.

OsiriX, led by Prof. Osman Ratib, is an Affiliate Group.

### Honorary Members

Prof. Ernest Feytmans
Honorary Director

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



Highest authority in the Institute, with supervisory powers. Its responsibilities include changes to SIB's statutes, the nomination of Group Leaders, and the approval of the annual budget and financial report.

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 infra-structure activities, as well as the core bioinformatics resources.

Takes all the decisions necessary to achieve the aims of the Institute, such as defining the scientific strategy and internal procedures, and allocating federal funds to service and infrastructure activities.

Discusses all matters relating to SIB groups as a whole, and proposes the nomination of new Group Leaders.

# Key figures

## FINANCES

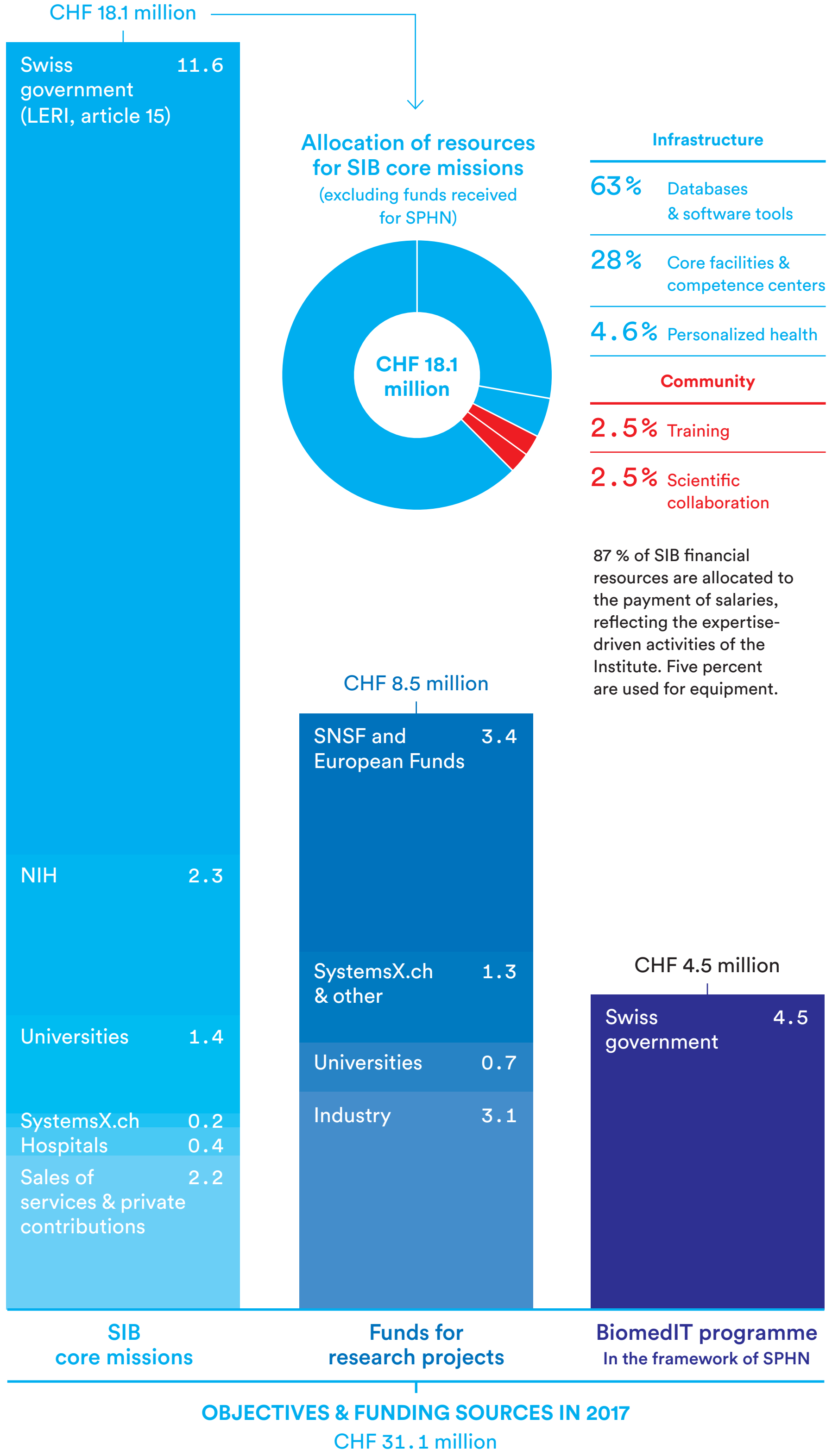
**SIB funds remained stable in 2017, thanks to the continued support of its funders.**

In 2017, the total amount of funds managed by SIB has reached CHF 26.6 million. This amount remained stable compared to last year.

---

SIB has also received funds from the Swiss Government in the framework of the Swiss Personalized Health Network (SPHN): Being in charge of the BioMedIT project, CHF 18M have been allocated to SIB for the period 2017-2020. This amount goes to the core facilities of SIB and the Universities, for the work performed as data nodes in the project over these 4 years.

The Government has granted SIB with CHF 20M for the SPHN programme over the same period. This amount is however directly transferred to the Swiss Academy of Medical Sciences to be allocated to the projects as per the decisions of the National Steering Board. It therefore does not appear in SIB's accounts.

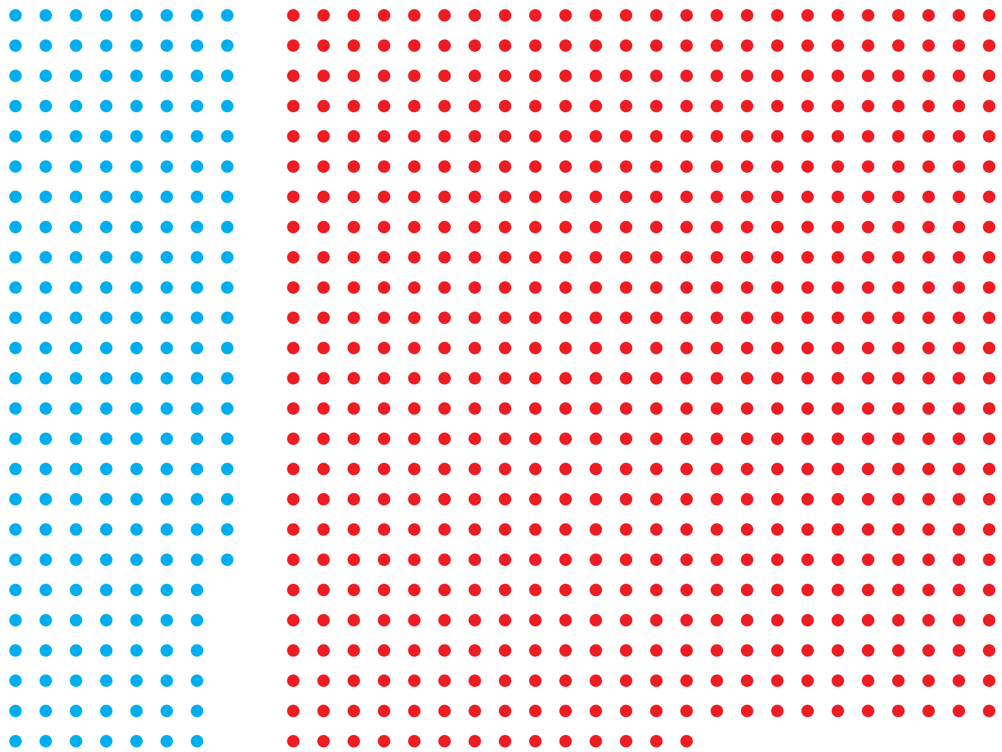


87 % of SIB financial resources are allocated to the payment of salaries, reflecting the expertise-driven activities of the Institute. Five percent are used for equipment.

## SIB MEMBERS

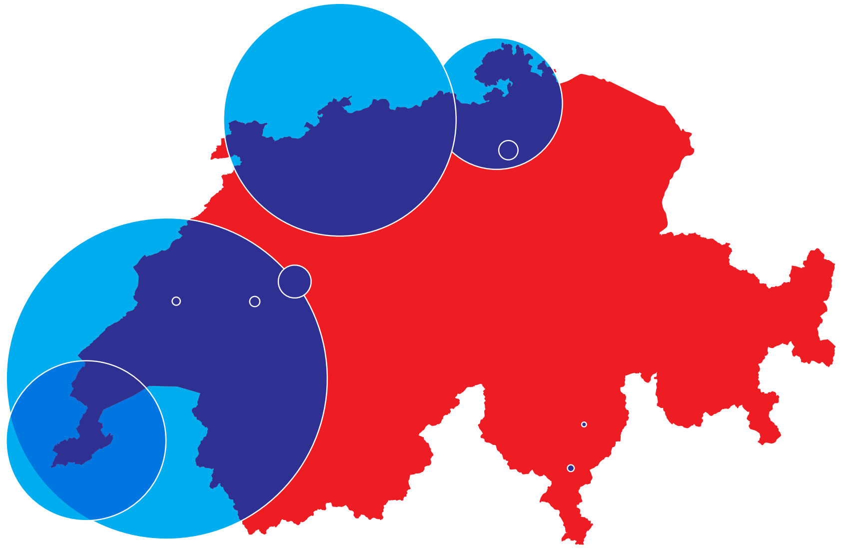
### Members and Employees*

| | |
|---|---|
| Employees | 194 |
| Members with another type of work contract | 590 |
| *Nationalities* | *25* |



### Geographical distribution of SIB Members

| | |
|---|---|
| Basel | 171 |
| Bellinzona | 6 |
| Berne | 29 |
| Lausanne | 283 |
| Fribourg | 9 |
| Geneva | 141 |
| Lugano | 4 |
| Wädenswil | 17 |
| Yverdon | 8 |
| Zurich | 116 |



---

* As of 13 February 2018

# INDEX OF SIB GROUP LEADERS

# ACKNOWLEDGEMENTS

**We gratefully acknowledge the following funders, sponsors and partners for their financial support and encouragement in helping us fulfil our mission in 2017.**

- The Swiss government and in particular:
  - The State Secretariat for Education, Research and Innovation (SERI)
  - The Swiss National Science Foundation (SNSF)
  - Innosuisse (CTI)
- Alfred and Annemarie von Sick grant
- Our institutional partners
- The European Commission
- The Leenaards Foundation
- The Medic Foundation
- The National Institutes of Health (NIH)
- The Research for Life Foundation
- SystemsX.ch

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

- The R. Geigy Foundation
- The University of Geneva
- The University of Lausanne

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

## IMPRESSUM

SIB GROUP LEADERS AND ACTIVITIES PICTURES BY
Nicolas Righetti, www.lundi13.ch

DESIGN AND LAYOUT BY
Bogsch & Bacco, www.bogsch-bacco.ch

COVER IMAGE

**Fireworks… and some birthday cake to celebrate SIB's 20th anniversary!**

While most people would gladly enjoy a slice – or two – of birthday cake, some others have to restrain themselves. Why is it that an extra dose of sugar leads to diabetes in some people? And once there, how does the disease evolve over time?

At SIB, data scientists are working together with experimental biologists to better understand the progression of diabetes over time, using advanced computational methods and visualization tools. This image represents a network of relationships (arcs) between genes, cellular pathways and traits associated with diabetes.

SOURCE: MARK IBBERSON, SENIOR SCIENTIST AT SIB'S VITAL-IT COMPETENCE CENTRE (CC BY-NC 4.0)

IMAGES CREDITS
(from top to bottom and from left to right)

P. 7 Franziska Gruhl - SIB. All rights reserved
Franziska Gruhl - SIB. All rights reserved
Sutthaburawonk / iStock
Fabio Rinaldi - SIB. All rights reserved
P. 10 Estrada Anton / Shutterstock
Creative commons
P. 11 Natural History Museum, London / SPL
Dennis Kunkel Microscopy / SPL
Creative commons
P. 49 DR
Creative commons
Strannik_fox / Shutterstock

SIB Swiss Institute of Bioinformatics
Quartier Sorge
Bâtiment Génopode
CH – 1015 Lausanne
T. +41 21 692 40 50

www.sib.swiss