# HANDBOOK
## OF THE GERMAN NETWORK FOR BIOINFORMATICS INFRASTRUCTURE

**BIG DATA EXPLOITATION IN LIFE SCIENCES**

www.denbi.de

# TABLE OF CONTENTS

# PREFACE

The **'German Network for Bioinformatics Infrastructure' (de.NBI)** has been initiated by the Federal Ministry of Education and Research (BMBF) to meet the bioinformatic challenges in modern life sciences due to the rapid technological progress in analytical areas such as sequencing, 'omics' and imaging techniques. These technologies generate huge amounts of data (big data), and thus require the access to well-maintained databases, bio-informatics tools, workflows and computing capacities.
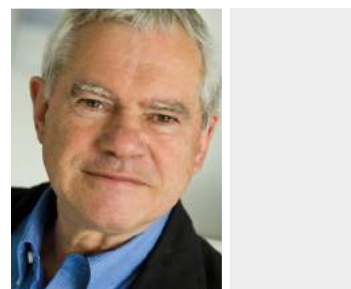
**Alfred Pühler**
*de.NBI Coordinator*
puehler@cebitec.uni-bielefeld.de

The 'German Network for Bioinformatics Infrastructure', consists of eight Service Centers located at German universities and research institutions as well as an Administration Office. The broad spectrum of bioinformatics services ranges from human, microbial and plant research to software libraries, data management provision and cloud computing.

As coordinator of the 'German Network for Bioinformatics Infrastructure', it is my pleasure to present the de.NBI handbook providing basic information on the network, its Service Centers and individual projects. The goal is to facilitate networking within the life science community, to pool existing bioinformatics expertise and resources and to support sustainable provision of infrastructure. In particular, by providing service and training activities, the network aims to support research groups working experimentally in analyzing big data sets produced in life science experiments.

The introduction of the handbook describes the mission and structure of the network. It informs about the Service Centers forming the backbone of the network and also on the decision-making Central Coordination Unit which is supported by Special Interest Groups. In addition, the Administration Office and the international Scientific Advisory Board are introduced. The main part of the handbook presents basic information on all Service Centers and their individual projects. The Service Centers specify their scientific topics for bioinformatics services and user training, the individual projects focus on services offered to de.NBI users.

With the de.NBI handbook the network hopes to increase its publicity and to provide helpful information for users interested in services, tools and training courses. For specific questions e-mail contacts are included. Another way of getting in touch with the network is the de.NBI web page (**http://www.denbi.de**) and its help desk. Last but not least, I would like to see the de.NBI handbook widely used by German and European scientists interested in starting a cooperation with members or institutions of the network.

**Alfred Pühler**
de.NBI Coordinator

# INTRODUCTION
## TO THE GERMAN NETWORK FOR BIOINFORMATICS INFRASTRUCTURE

# ESTABLISHMENT AND MISSION STATEMENT

## Forming a National Bioinformatics Infrastructure

The de.NBI is a distributed bioinformatics service and training network funded by the BMBF. It consists of 30 project partners with 39 individual projects organized in 8 thematically focused Service Centers distributed all over Germany. The project partners provide bioinformatics services and organize training events to assist researchers in life sciences in the exploitation of their experimental data.

## Establishment of de.NBI

- In **May 2013** the German Federal Ministry of Education and Research (BMBF) published a call for a network for bioinformatics infrastructure.
- Until **July 2014** the concept of the network constituted of eight Service Centers was elaborated.
- In **March 2015** the de.NBI initiative was officially launched by the BMBF.
- In **November 2016** the de.NBI initiative was extended by partner projects.

## Mission statement of de.NBI

**The de.NBI program provides**
- Comprehensive first-class bioinformatics **services** to users in basic and applied life sciences research from academia, industry and medicine
- Bioinformatics **training** to users in Germany and Europe through a wide range of workshops and courses
- **Cooperation** of the German bioinformatics community with international network structures

**SERVICE**

**TRAINING**

**INTERNATIONAL COOPERATION**
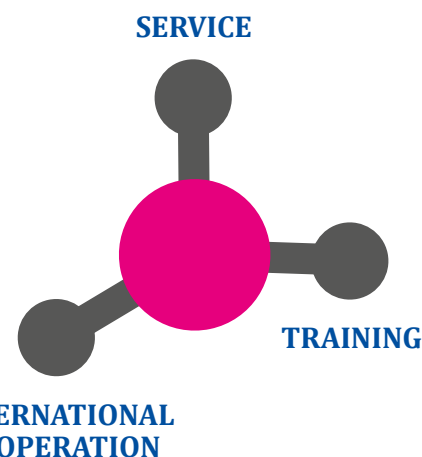
▼ *de.NBI members* *participating in the 2nd Plenary Meeting in Berlin in November 2016.*

# COORDINATION AND ORGANIZATION

## The Acting Units of de.NBI

**Right from the beginning de.NBI has been designed as a distributed and coordinated infrastructure.**

The de.NBI Coordinator Alfred Pühler, Center for Biotechnology at Bielefeld University, has been appointed by the BMBF and is responsible for the overall organization and the management of the network.

**Alfred Pühler**
*de.NBI Coordinator*
puehler@cebitec.uni-bielefeld.de

**The network is composed of the following units:**
- The eight **Service Centers (SCs)** are responsible for service and training.
- The **Central Coordination Unit (CCU)** is the decision-making body of de.NBI.
- The **Special Interest Groups (SIGs)** prepare decisions to be taken by the CCU.
- The **Administration Office (AO)** supports all organizational issues of de.NBI.
- The **Scientific Advisory Board (SAB)** advises the CCU and the Service Centers in developing the network.

## de.NBI organizational chart
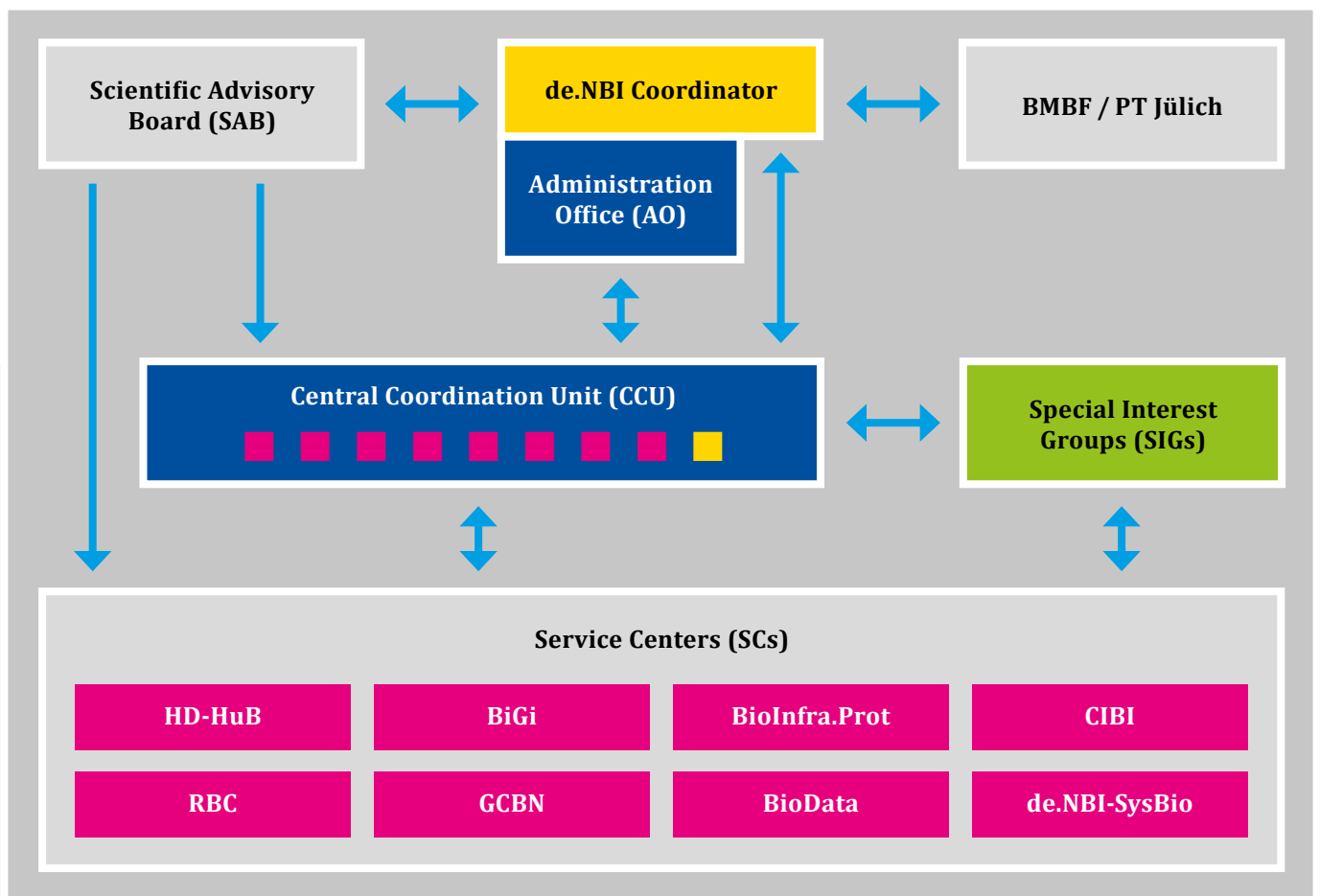
# THE EIGHT SERVICE CENTERS (SCs)

## The Backbone of de.NBI

A main goal of de.NBI is to provide excellent bioinformatics services and training for life sciences research in academia and industry. de.NBI maintains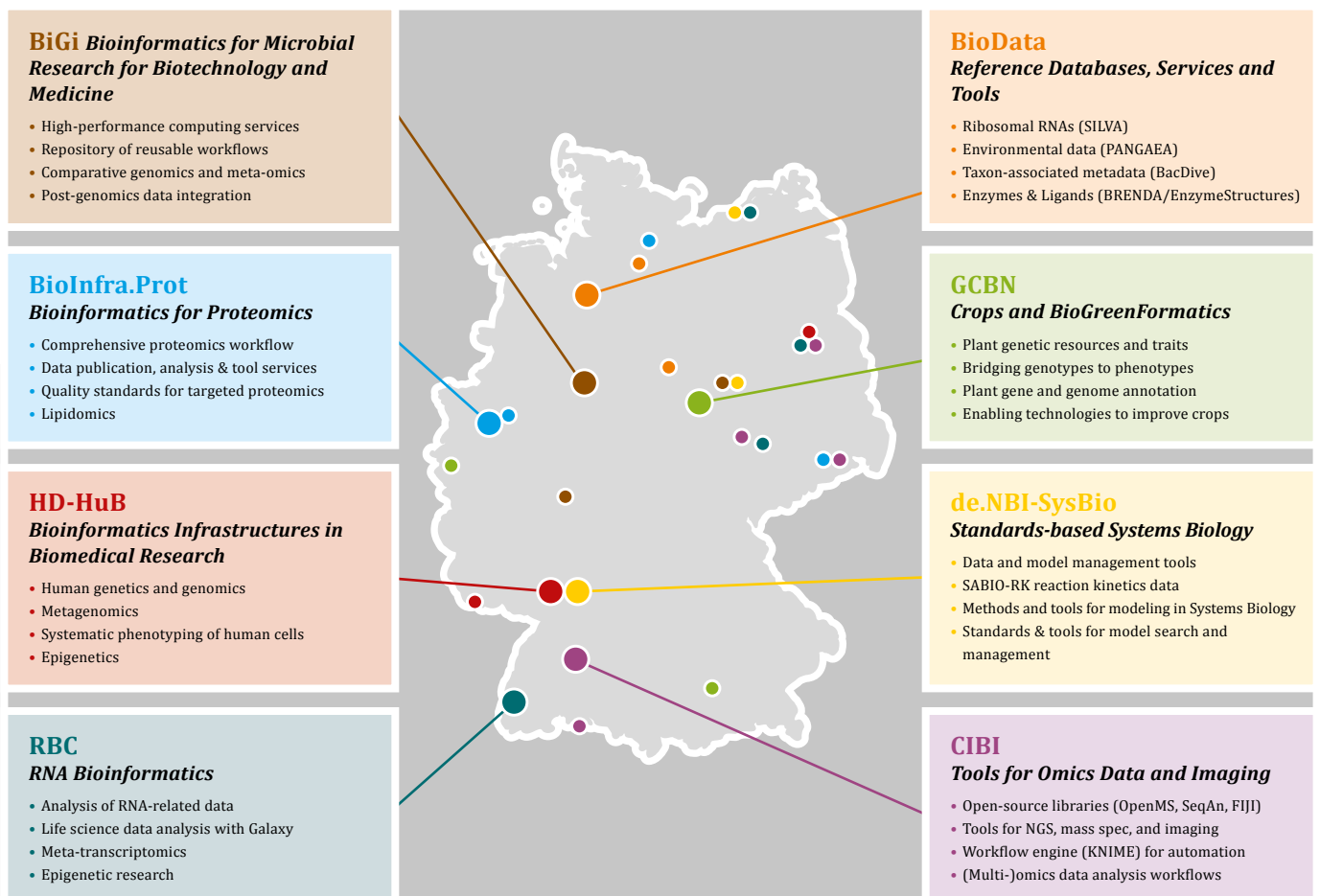 and develops almost 100 bioinformatics services for the human, plant and microbial research fields. de.NBI services are registered and searchable in the ELIXIR Tools and Data Services Registry. Furthermore, the de.NBI Cloud offers compute resources to the research community. de.NBI training activities include one to two-day training courses, webinars, online training, and themed one-week summer schools.

**The eight de.NBI Service Centers and their coordinators are:**

- **Heidelberg Center for Human Bioinformatics** (**HD-HuB**) — **Peer Bork**, EMBL, Heidelberg
- **Bielefeld-Gießen Resource Center for Microbial Bioinformatics** (**BiGi**) — **Jens Stoye**, Bielefeld University
- **Bioinformatics for Proteomics** (**BioInfra.Prot**) — **Martin Eisenacher**, Ruhr-Universität Bochum
- **Center for Integrative Bioinformatics** (**CIBI**) — **Oliver Kohlbacher**, University of Tübingen
- **RNA Bioinformatics Center** (**RBC**) — **Rolf Backofen**, University of Freiburg
- **German Crop BioGreenformatics Network** (**GCBN**) — **Uwe Scholz**, IPK Gatersleben
- **Center for Biological Data** (**BioData**) — **Frank Oliver Glöckner**, Jacobs University Bremen
- **de.NBI Systems Biology Service Center** (**de.NBI-SysBio**) — **Wolfgang Müller**, HITS Heidelberg

## de.NBI service topics at a glance



**BiGi** *Bioinformatics for Microbial Research for Biotechnology and Medicine*

- High-performance computing services
- Repository of reusable workflows
- Comparative genomics and meta-omics
- Post-genomics data integration

**BioInfra.Prot**
*Bioinformatics for Proteomics*

- Comprehensive proteomics workflow
- Data publication, analysis & tool services
- Quality standards for targeted proteomics
- Lipidomics

**HD-HuB**
*Bioinformatics Infrastructures in Biomedical Research*

- Human genetics and genomics
- Metagenomics
- Systematic phenotyping of human cells
- Epigenetics

**RBC**
*RNA Bioinformatics*

- Analysis of RNA-related data
- Life science data analysis with Galaxy
- Meta-transcriptomics
- Epigenetic research

**BioData**
*Reference Databases, Services and Tools*

- Ribosomal RNAs (SILVA)
- Environmental data (PANGAEA)
- Taxon-associated metadata (BacDive)
- Enzymes & Ligands (BRENDA/EnzymeStructures)

**GCBN**
*Crops and BioGreenFormatics*

- Plant genetic resources and traits
- Bridging genotypes to phenotypes
- Plant gene and genome annotation
- Enabling technologies to improve crops

**de.NBI-SysBio**
*Standards-based Systems Biology*

- Data and model management tools
- SABIO-RK reaction kinetics data
- Methods and tools for modeling in Systems Biology
- Standards & tools for model search and management

**CIBI**
*Tools for Omics Data and Imaging*

- Open-source libraries (OpenMS, SeqAn, FIJI)
- Tools for NGS, mass spec, and imaging
- Workflow engine (KNIME) for automation
- (Multi-)omics data analysis workflows

# CENTRAL COORDINATION UNIT (CCU)

## The Decision-Making Body of de.NBI

The **Central Coordination Unit** (CCU) is the decision-making body of de.NBI and responsible for the effective operation of the scientific, technical and administrative structures. The board of the CCU consists of nine de.NBI members, i.e. the coordinators of the eight Service Centers and the de.NBI coordinator. The CCU is on duty to control that the consortium's activities comply with the mission statement to provide state-of-the-art bioinformatics services and training to users in basic and applied life sciences research.

## Members of the CCU are:

**Rolf Backofen**
*RBC*
Freiburg

**Peer Bork**
*HD-HuB*
Heidelberg

**Martin Eisenacher**
*BioInfra.Prot*
Bochum

**Frank Oliver Glöckner**
*BioData*
Bremen

**Jens Stoye**
*BiGi*
Bielefeld

**Alfred Pühler**
*Coordinator*
Bielefeld

**Oliver Kohlbacher**
*CIBI*
Tübingen

**Wolfgang Müller**
*de.NBI-SysBio*
Heidelberg
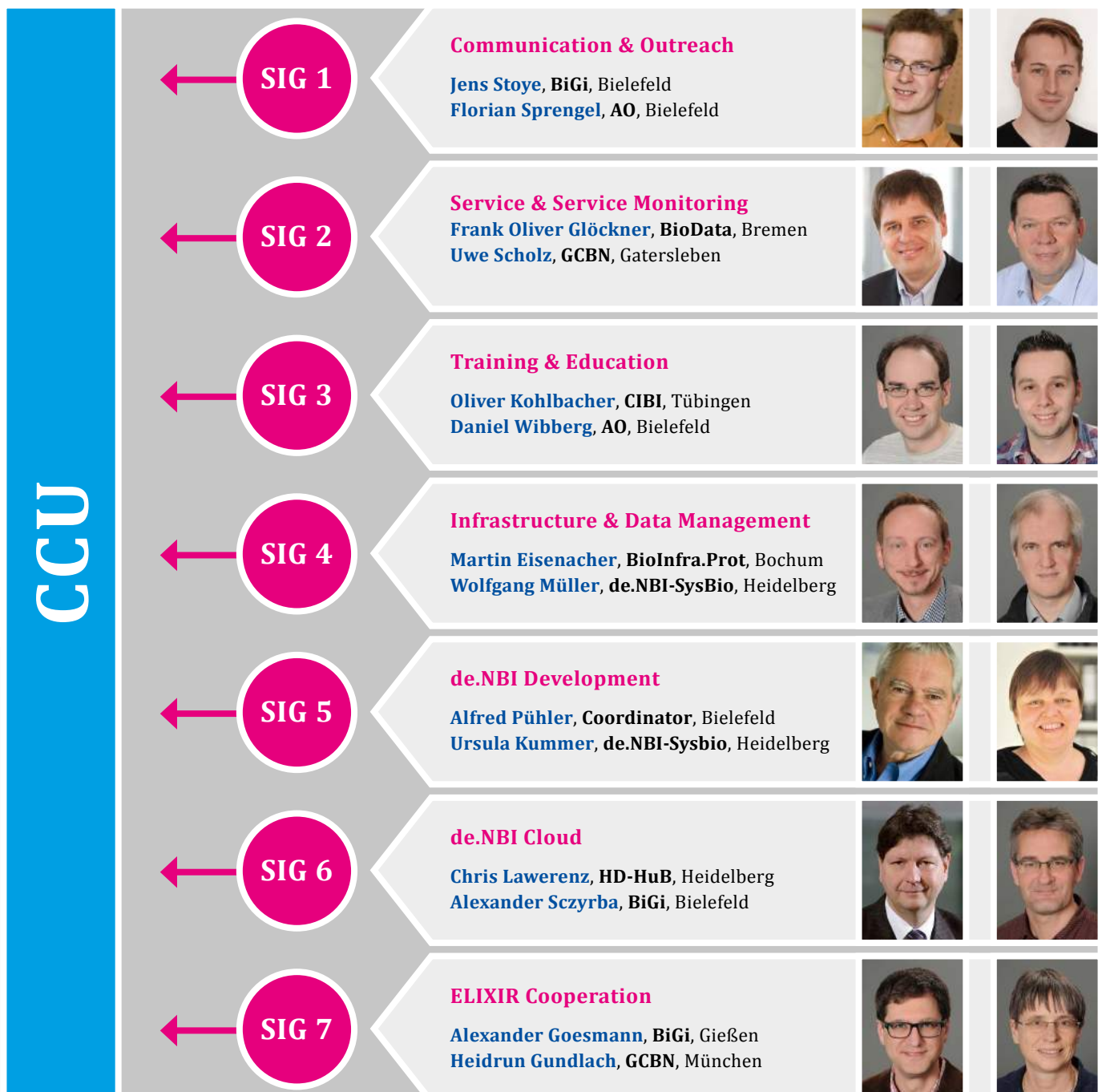
**Uwe Scholz**
*GCBN*
Gatersleben

# SPECIAL INTEREST GROUPS (SIGs)

## Preparing Operational and Strategic Issues for the CCU

**Special Interest Groups** (SIGs) are small discussion groups of de.NBI experts from all Service Centers. These subcommittees are established by the Central Coordination Unit (CCU) and focus on topics relevant for operational and strategic decisions by the de.NBI management.

## The current SIGs and their chairpersons:

**CCU**

**SIG 1**
### Communication & Outreach
Jens Stoye, **BiGi**, Bielefeld
Florian Sprengel, **AO**, Bielefeld

**SIG 2**
### Service & Service Monitoring
Frank Oliver Glöckner, **BioData**, Bremen
Uwe Scholz, **GCBN**, Gatersleben

**SIG 3**
### Training & Education
Oliver Kohlbacher, **CIBI**, Tübingen
Daniel Wibberg, **AO**, Bielefeld

**SIG 4**
### Infrastructure & Data Management
Martin Eisenacher, **BioInfra.Prot**, Bochum
Wolfgang Müller, **de.NBI-SysBio**, Heidelberg

**SIG 5**
### de.NBI Development
Alfred Pühler, **Coordinator**, Bielefeld
Ursula Kummer, **de.NBI-Sysbio**, Heidelberg

**SIG 6**
### de.NBI Cloud
Chris Lawerenz, **HD-HuB**, Heidelberg
Alexander Sczyrba, **BiGi**, Bielefeld

**SIG 7**
### ELIXIR Cooperation
Alexander Goesmann, **BiGi**, Gießen
Heidrun Gundlach, **GCBN**, München

# de.NBI ADMINISTRATION OFFICE (AO)

## Taking Care of the Administration of de.NBI

The **Administration Office** (AO) is the central support entity for the de.NBI coordinator and the Central Coordination Unit (CCU). It provides a range of administrative and management services to the consortium and serves as a connecting element between the coordinator and the CCU. The AO facilitates the communication within the consortium, with users of de.NBI services or training activities and the public.

## Members of the AO:

| | |
|---|---|
| **Andreas Tauch** | *Head of the Administration Office* |
| **Tanja Dammann-Kalinowski** | *Project Management* |
| **Daniel Wibberg** | *Training Coordination & Outreach* |
| **Florian Sprengel** | *Technical Project Management & Media* |
| **Irena Maus** | *Service Coordination* |
| **Doris Jording** | *ELIXIR Cooperation & Partner Projects* |
| **Peter Belmann** | *de.NBI Cloud Governance* |

**Contact:**

**de.NBI Administration Office**
*Center for Biotechnology (CeBiTec)*
*Bielefeld University*
*Universitätsstraße 27*
*33615 Bielefeld*
*contact@denbi.de*



◄ *Members of the AO. From left to right: Alfred Pühler, Peter Belmann, Florian Sprengel, Tanja Dammann-Kalinowski, Irena Maus, Doris Jording, Andreas Tauch, Daniel Wibberg (Juli 2017).*

# SCIENTIFIC ADVISORY BOARD (SAB)

## Giving Advice for the Future Development of de.NBI

The **Scientific Advisory Board** (SAB) is appointed by the BMBF and provides recommendations to the de.NBI partners and the Central Coordination Unit on technical, organizational and strategic issues related to the de.NBI mission. Evaluation of the de.NBI consortium by the SAB is scheduled as an annual process that is associated with de.NBI workshops. The SAB is composed of six members who are internationally renowned scientists in bioinformatics.

## Six European experts have been appointed by the BMBF as SAB members:

**Janet Thornton**
*SAB Chairperson*
*Director Emeritus of EMBL-EBI and Senior Scientist*
Hinxton, United Kingdom

**Reinhard Schneider**
*Head of Bioinformatics Core Facility University of Luxembourg*
Belvaux, Luxembourg

**Ron D. Appel**
*Executive Director of Swiss Institute of Bioinformatics*
Lausanne, Switzerland

**Alice C. McHardy**
*Head of Department Computational Biology of Infection Research Helmholz Centre for Infection Research*
Braunschweig, Germany

**Ivo Hofacker**
*Head of Research Group Bioinformatics and Computational Biology University of Vienna*
Vienna, Austria

**Ivo G. Gut**
*Director of Centro Nacional de Análisis Genómico*
Barcelona, Spain

# THE de.NBI SERVICE CENTERS

### HD-HuB                                    13 ff.
***Heidelberg Center for Human Bioinformatics***

- DKFZ Heidelberg
- EMBL Heidelberg
- Heidelberg University
- Saarland University
- Charité Berlin/BIH

### BiGi                                      19 ff.
***Bielefeld-Gießen Resource Center for Microbial Bioinformatics***

- Bielefeld University
- Justus Liebig University Giessen
- University of Magdeburg

### BioInfra.Prot                             24 ff.
***Bioinformatics for Proteomics***

- Ruhr-Universität Bochum
- ISAS Dortmund
- Forschungszentrum Borstel
- MPI of Molecular Cell Biology and Genetics, Dresden

### CIBI                                      30 ff.
***Center for Integrative Bioinformatics***

- University of Tübingen
- Freie Universität Berlin
- University of Konstanz
- Center for Systems Biology, Dresden
- Leibniz Institute of Plant Biochemistry, Halle

### RBC                                       36 ff.
***RNA Bioinformatics Center***

- University of Freiburg
- Leipzig University
- MDC for Molecular Medicine, Berlin
- University of Rostock

### GCBN                                      45 ff.
***German Crop BioGreenformatics Network***

- IPK Gatersleben
- Helmholtz Zentrum München
- Forschungszentrum Jülich

### BioData                                   49 ff.
***Center for Biological Data***

- Jacobs University Bremen
- University Bremen
- DSMZ, Braunschweig
- Technische Universität Braunschweig
- Universität Hamburg

### de.NBI-SysBio                             55 ff.
***de.NBI Systems Biology Service Center***

- Heidelberg Institute for Theoretical Studies (HITS)
- University of Rostock
- Heidelberg University
- MPI for Dynamics of Complex Technical Systems, Magdeburg

# THE de.NBI SERVICE CENTER HD-HuB

## Heidelberg Center for Human Bioinformatics

**The HD-HuB Service Center integrates the expertise of four major institutions: The German Cancer Research Center (DKFZ), European Molecular Biology Laboratory (EMBL) and Heidelberg University, and Charité/Berlin Institute of Health (BIH).**

In addition, the de.NBI-epi partner project consisting of partners from DKFZ, Saarland University, and MPI for Informatics has been associated with HD-HuB since November 2016. The goal of the HD-HuB Service Center is to provide access to bioinformatics services, compute resources and training for the manage-ment, analysis and interpretation of sequencing and image-based data to facilitate data-driven discoveries in key fields of human biology ranging from basic to clinical research. HD-HuB is a node of the de.NBI cloud within the de.NBI network.

### TOPICS OF THE SERVICE CENTER HD-HuB:

- Human genomics and epigenomics
- Human microbiomics
- Systematic phenotyping of human cells

### SELECTION OF TRAINING COURSES CARRIED OUT BY HD-HuB:

- Microscopy Image Analysis
- Data Interpretation of RNASeq and Bisulfite Sequencing Data in Cancer Research
- Data Interpretation of Whole-Genome and Exome Sequencing Data in Cancer Research
- Software Carpentry

### CURRENT PROJECTS:

- **HD-HuB 1** at **DKFZ Heidelberg and EMBL Heidelberg**: Human genetics and genomics
- **HD-HuB 2** at **EMBL Heidelberg**: Human microbiomics
- **HD-HuB 3** at **Heidelberg University and DKFZ Heidelberg**: Systematic phenotyping of human cells
- **de.NBI-epi 1** at **Heidelberg University** and **de.NBI-epi 2** at **Saarland University**: Analysis tools for epigenomic data



**Peer Bork**
*Coordinator*
*EMBL Heidelberg*
bork@embl.de



**Roland Eils**
*Deputy coordinator*
*Charité/BIH*
roland.eils@charite.dedkfz.de

▼ *Members of the Service Center HD-HuB. From left to right: Y. Assenov, N. Habermann, S. Yakhnin, U. Trojahn, G. Zeller, M. Smith, R. Russel, P. Bork, C. Lawerenz, I. Buchhalter, B. Rauscher, K. Rohr, P. Kensche, J. Puyol, M. Sharan, N. Ishaque, J. Korbel, T. Wollmann, M. Lang, P. Lustik, K. Nordstroem, M. Gunkel (October 2018)*

# THE de.NBI PROJECT HD-HuB 1

## Human Genetics and Genomics

**The HD-HuB 1 project at the German Cancer Research Center (DKFZ) and European Molecular Biology Laboratory (EMBL) and Charité/Berlin Institute of Health (BIH) offers tools, services and courses for the application area of human genetics and genomics.**

▼ *The HD-HuB 1 team at Heidelberg University, DKFZ, EMBL and BIH.*

**Rob Russell**
*Project leader*
robert.russell@bioquant.
uni-heidelberg.de

**Ivo Buchhalter**
i.buchhalter@dkfz.de

**Wolfgang Huber**
whuber@embl.de

**Jan Korbel**
korbel@embl.de

**Roland Eils**
roland.eils@charite.de

**Naveed Ishaque**
naveed.ishaque@charite.de

### COLLECTION OF TOOLS AND SERVICES OFFERED BY HD-HuB 1:

- **Butler** – A computational framework for orchestrating genomics analysis workflows in the cloud.
- **rhdf5/Rhdf5lib** – The package provides an interface between HDF5 and R. In particular HDF5 is being used to address the rapid growth of single-cell datasets.
- **biomaRt** – Provides an R interface to BioMart services.
- **DELLY** – Workflow for the discovery of germline and somatic structural variants.
- **DESeq2 & DEXSeq** – Suite of R/Bioconductor packages for differential expression analysis of high-throughput sequencing assays.
- **PanCancer alignment workflow** – Based in the Pan-Cancer project for the alignment of NGS data to reference genomes .
- **ACESeq** – Detection of somatic copy number variations from matching tumor/control whole genome sequencing data pairs.
- **SNV calling pipeline** – Calling of high confidence somatic SNVs based on set from tumor and control samples.
- **Indel calling pipeline** – Detection of high confidence indels (1-20 bp) from tumor and control samples.
- **Roddy** – Framework for large scale NGS processing pipelines.
- **OTP** – Automation platform for management and processing of NGS data.
- **Mechismo** – Rapid predictions of how genetic variants impact on biomolecular function employing pre-computed collections of protein-protein, protein-DNA/RNA, protein-chemical interactions.

**A**

**B**



A) Overview of data processing in OTP. For details please refer to J Biotechnol. 2017 Nov 10; 261:53-62. doi: 10.1016/ j.jbiotec.2017.08.006.
B) Visualisation of differential exon usage in the Apoe gene as quantified by the DEXSeq method (y-axis of the top panel). For details see Nat Methods. 2015 Feb;12(2):115-21. doi: 10.1038/nmeth.3252.
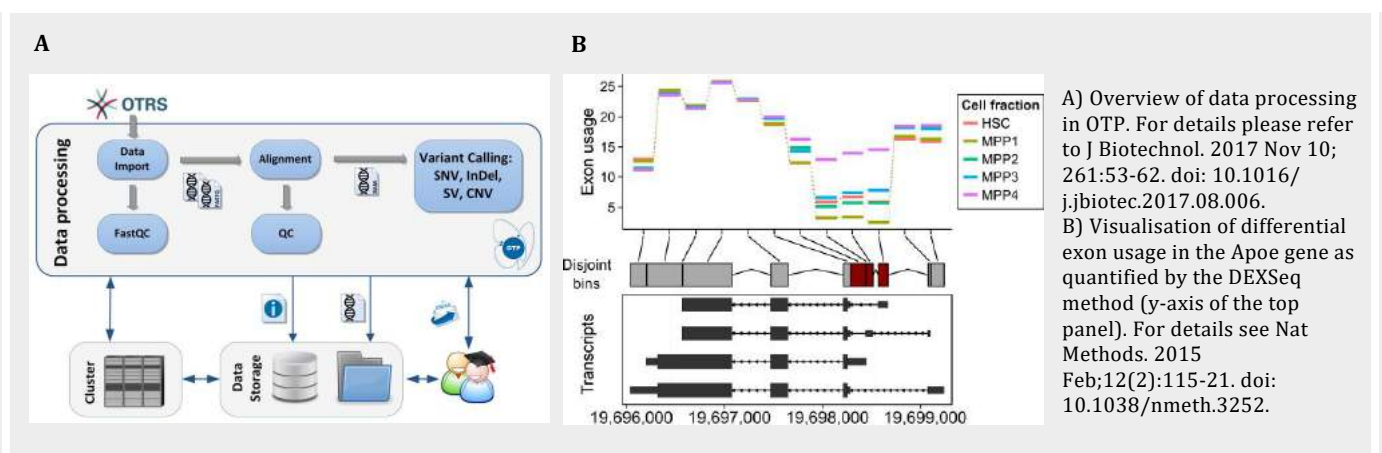
# THE de.NBI PROJECT HD-HuB 2

## Human Microbiomics

The **HD-HuB 2** project at the European Molecular Biology Laboratory (EMBL) covers the whole range of analysis tasks typically applied in the field of micro-biomics.
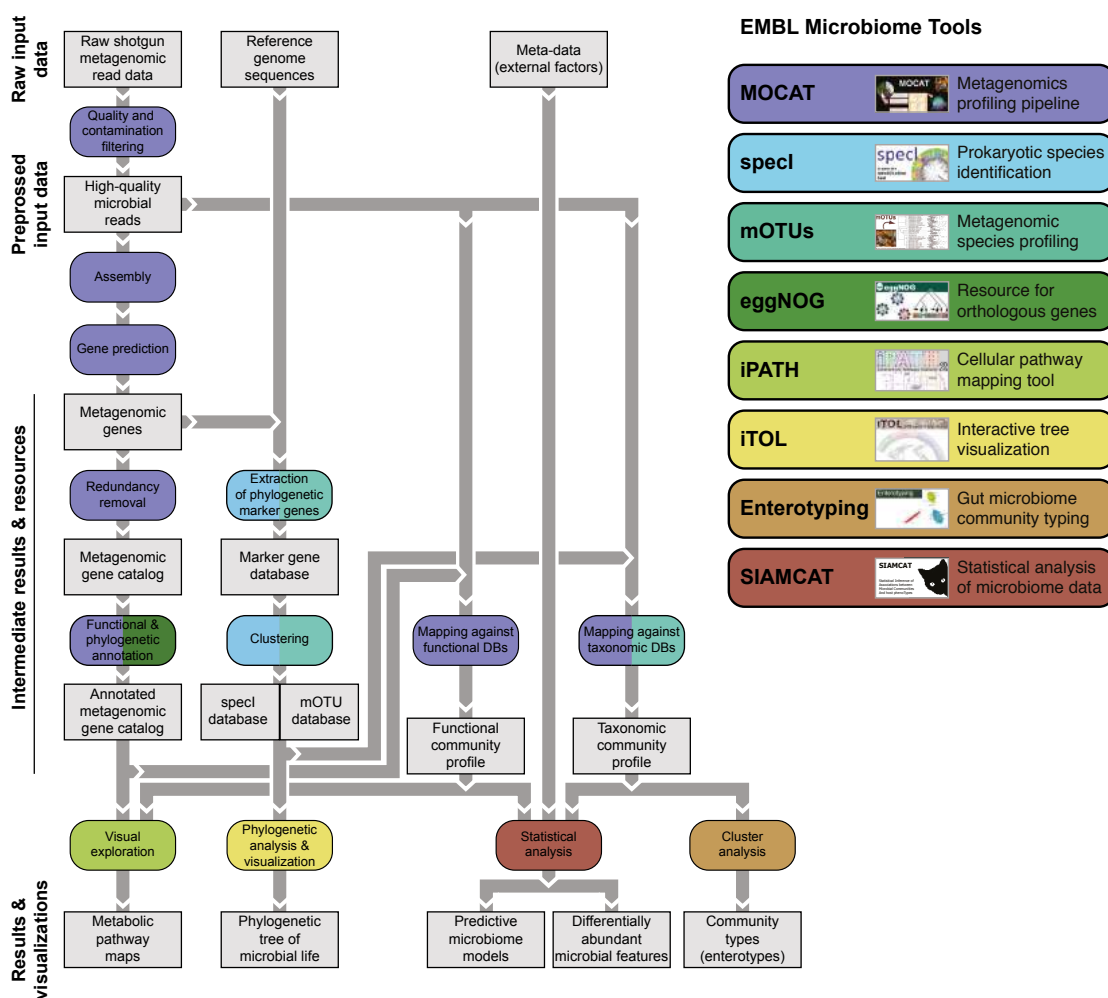
▼ *The HD-HuB 2 team at EMBL.*

**Peer Bork**
*Project leader*
bork@embl.de

**Georg Zeller**
zeller@embl.de

**COLLECTION OF TOOLS AND SERVICES OFFERED BY HD-HuB 2:**



**EMBL Microbiome Tools**

| Tool | Description |
|------|-------------|
| MOCAT | Metagenomics profiling pipeline |
| specI | Prokaryotic species identification |
| mOTUs | Metagenomic species profiling |
| eggNOG | Resource for orthologous genes |
| iPATH | Cellular pathway mapping tool |
| iTOL | Interactive tree visualization |
| Enterotyping | Gut microbiome community typing |
| SIAMCAT | Statistical analysis of microbiome data |

# THE de.NBI PROJECT HD-HuB 3

## Systematic Phenotyping of Human Cells

**The HD-HuB 3 project at DKFZ and Heidelberg University offers tools, services and training for the analysis and interpretation of microscopy image data for automated phenotyping of human cells.**

▼ *The HD-HuB 3 team at DKFZ and Heidelberg University.*

**Michael Boutros**
*Project leader*
m.boutros@dkfz.de

**Oliver Stegle**
o.stegle@dkfz.de

**Holger Erfle**
holger.erfle@bioquant.
uni-heidelberg.de

**Karl Rohr**
karl.rohr@bioquant.
uni-heidelberg.de

### COLLECTION OF TOOLS AND SERVICES OFFERED BY HD-HuB 3:

- **KNIME-based cellular phenotyping of microscopy image data** – a data mining platform including manifold libraries for image processing (based on Fiji/ImageJ) and data exploration. The platform provides access to workflows and pipelines for (large-scale) automated phenotype analysis of microscopy image data.
- **Galaxy Image analysis** for systematic phenotyping of human cells using microscopy image data and the Galaxy platform.
- **GenomeRNAi** – Provides RNA interference (RNAi) phenotype data extracted by manual curation from the literature, or submitted by data producers directly. In addition to the RNAi screening data of human and Drosophila cells, it contains RNAi reagent information, along with an assessment as to their efficiency and specificity.
- **GenomeCRISPR** – Database for high-throughput screening experiments performed by using the CRISPR/Cas9 system. A dynamic web interface guides users through the process of finding information about published CRISPR screens. The database contains detailed data about observed hits and phenotypes. Moreover, it provides knowledge about performance of individual single guide RNAs (sgRNAs) used under various experimental conditions.
- **E-CRISP** – E-CRISP is a computational tool to design and evaluate guide RNAs for use with CRISPR/Cas9. The design process incorporates different parameters of how CRISPR constructs can be used in experimental applications, such as knock out and tagging experiments.



▼ *Processing workflows of human imaging data.*
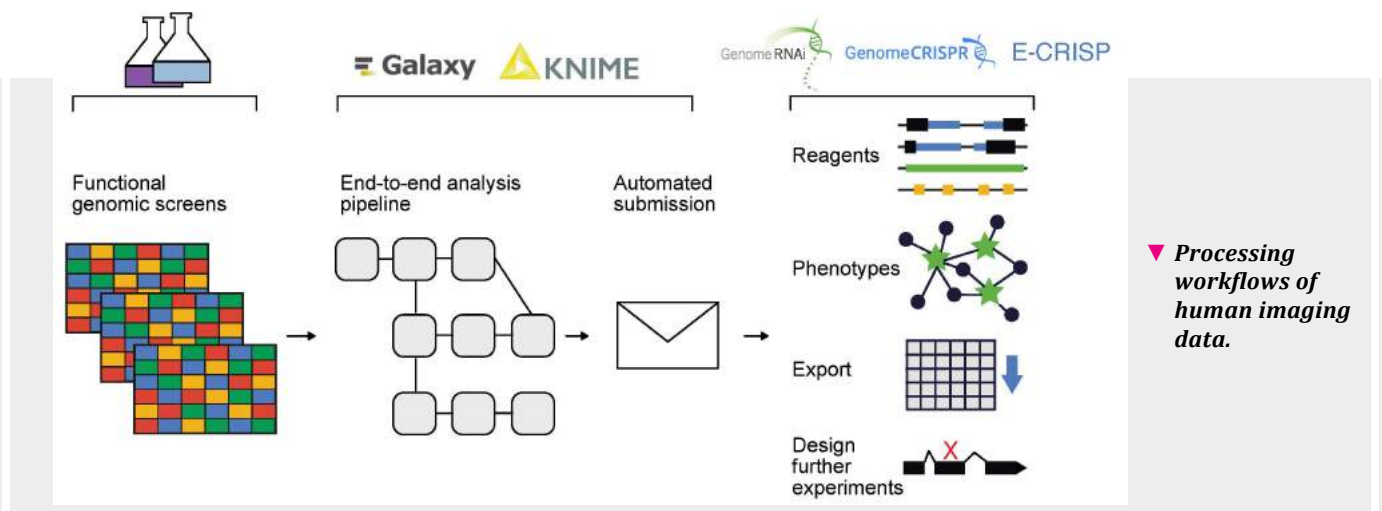
# THE de.NBI PROJECT de.NBI-epi 1

## Analysis of DNA Methylation

**The de.NBI-epi 1 project at the German Cancer Research Center (DKFZ) focuses on methods to analyze data from whole-genome bisulfite sequencing (WGBS) and array-based hybridization to Illumina Infinium methylation arrays.**

▼ *The de.NBI-epi 1 team at DKFZ.*

**Benedikt Brors**
*Project leader*
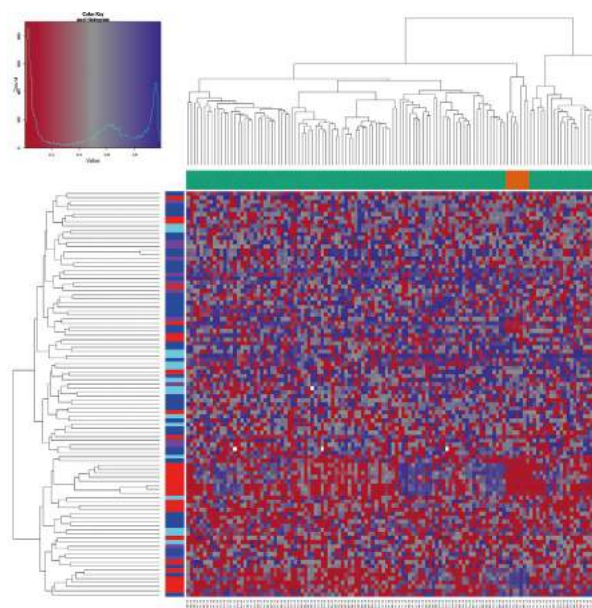b.brors@dkfz.de

**Yassen Assenov**
y.assenov@dkfz.de

### COLLECTION OF TOOLS AND SERVICES OFFERED BY de.NBI-epi 1:

- **RnBeads** – A R/Bioconductor library for analysing data from Illumina Infinium methylation arrays (Infinium 450k and EPIC arrays). It includes methods for data import, different normalization methods, quality assessment, exploratory analysis and calling of differentially methylated nucleotides.

▼ *RnBeads analysis workflow (Assenov et al., 2014).*



**Data Import**
This report describes the loading of the data into RnBeads.

**Quality Control**
This report performs assay quality validation.

**Preprocessing**
This report presents the filtering and normalization steps applied to the dataset.

**Tracks and Tables**
This report provides information on exported data, generated genome browser tracks and sample summary tables.

**Exploratory Analysis**
This reports describes sample subgroups, methylation profiles and associations with sample annotations.

**Differential Methylation**
This report identifies differentially methylated sites and regions between sample groups.

▼ *RnBeads DNA methylation heatmap.*

# THE de.NBI PROJECT de.NBI-epi 2

## Analysis of Epigenomic Data

**The de.NBI-epi 2 project at Max Planck Institute for Informatics and Saarland University provides a series of high-end tools for analysis of epigenomic data with focus on DNA methylation data.**

▼ *The de.NBI-epi 2 team at Max Planck Institute for Informatics and Saarland University.*

**Jörn Walter**
*Project leader*
j.walter@mx.uni-saarland.de

**Thomas Lengauer**
lengauer@mpi-inf.mpg.de

**Michael Scherer**
mscherer@mpi-inf.mpg.de

**Karl Nordström**
karl.nordstroem@uni-saarland.de

### COLLECTION OF TOOLS AND SERVICES OFFERED BY de.NBI-epi 2:

- **BioQ Analyzer HiMod** – Processing and primary analysis of data obtained in standard targeted bisulfite sequencing experiments, including Tet-assisted (TAB-Seq) and chemical modification-based bisulfite sequencing (Ox-BS) approaches.

- **MeDeCom –** A reference free method for decomposition of multi-cellular, heterogeneous DNA methylomes into quantified latent components approximating distinct cell population.

- **EpiExplorer** – Integration of multiple epigenetic and genetic annotations made explorable via an interactive interface.

- **RnBeads** – R package for comprehensive analysis of DNA methylation data obtained by Infinium microarrays and bisulfite sequencing protocols. RnBeads implements a series of QC steps and presents a number of analyses levels as annotated, easy accessible and readable hypertext report. RnBeads is the standard preparation tool for DNA-methylation data obtained by the DEEP (German Epigenome Programme).

- **The DeepBlue Epigenomic Data Server** provides a data access hub for large collections of epigenomic data. It organizes the data using controlled vocabularies and ontologies. The data is stored on our server where users can access the data programmatically or through our web interface.
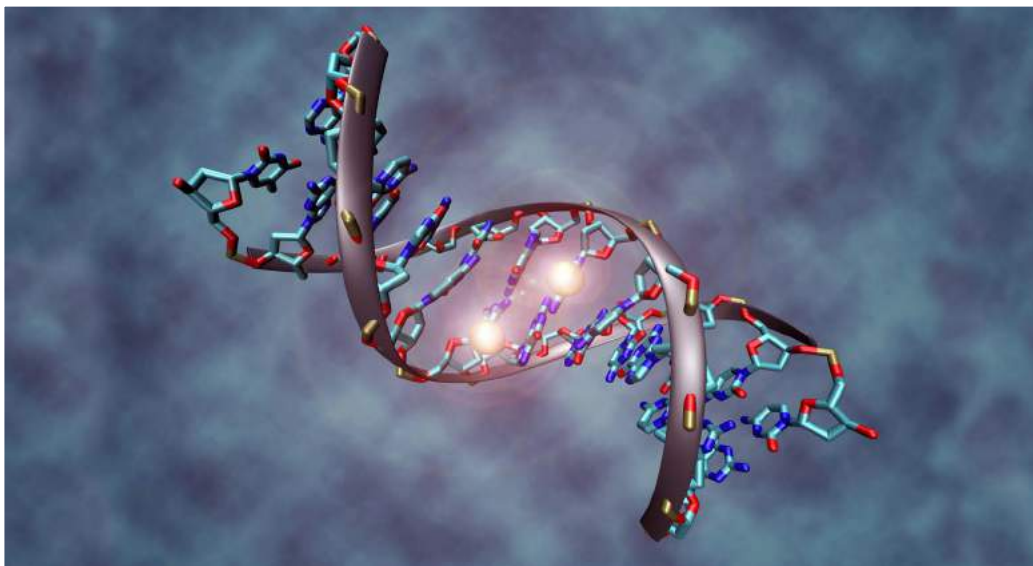
◄ *DNA segment with highlighted methylation sites.*

# THE de.NBI SERVICE CENTER BiGi

## Bielefeld-Gießen Resource Center for Microbial Bioinformatics

**The BiGi Service Center offers tools, services and training for microbial genome, metagenome, and post-genome research that is complemented by a large-scale hardware infrastructure.**

The BiGi Service Center concentrates on the bioinformatics analysis of genomics and postgenomics data of isolated microbes as well as of microbial communities. The BiGi Service Center is equipped with large-scale computing and storage resources and both, Bielefeld and Giessen, are involved in establishing a dedicated de.NBI cloud.

### TOPICS OF THE SERVICE CENTER BiGi:

- Genomics and metagenomics: Assembly, annotation and comparative analysis of microbial genome and metagenome sequences
- Postgenomics and post-metagenomics: Analysis of transcriptome, proteome and metabolome data of microbes and microbial communities

### SELECTION OF TRAINING COURSES CARRIED OUT BY BiGi:

- High-throughput genome analysis and comparative genomics
- Introduction into targeted and untargeted metagenome analysis
- Introduction to cloud computing in the de.NBI cloud

### CURRENT PROJECTS:

- **BiGi 1** at **Bielefeld University**: Postgenomics of isolated microbes and meta-genomics of microbial communities
- **BiGi 2** at **Giessen University**: Assembly, annotation and comparative analysis of microbial genome sequences; short-read mapping data analysis; read-based metagenomics
- **MetaProtServ 1** at **University of Magdeburg**: Protein identification, functional and taxonomic investigation in microbial communities
- **MetaProtServ 2** at **University of Magdeburg**: Web-based meta-proteomics as a highly-scalable service



**Jens Stoye**
*Coordinator*
*Bielefeld University*
jens.stoye@uni-bielefeld.de



**Alexander Goesmann**
*Deputy coordinator*
*Gießen University*
alexander.goesmann@computational.
bio.uni-giessen.de

▼ *Members of the Service Center BiGi. From left to right: Oliver Rupp, Sebastian Jünemann, Jochen Blom, Sebastian Jaenicke, Alexander Goesmann, Jens Stoye, Alexander Sczyrba (November 2015).*

# THE de.NBI PROJECT BiGi 1
## Microbial Postgenome and Metagenome Research

**The BiGi 1 project at Bielefeld University offers tools, services and training for postgenome and metagenome research.**

▼ *The BiGi 1 team at Bielefeld University and CeBiTec.*

**Jens Stoye**
*Project leader*
jens.stoye@uni-bielefeld.de

**Stefan Albaum**
alu@cebitec.uni-bielefeld.de

**Alexander Sczyrba**
asczyrba@cebitec.uni-bielefeld.de

**Sebastian Jünemann**
jueneman@cebitec.uni-bielefeld.de

**Nils Kleinbölting**
nkleinbo@cebitec.uni-bielefeld.de

**COLLECTION OF TOOLS AND SERVICES OFFERED BY BiGi 1:**

- **ProMeTra/Fusion** – Integrative analysis and visualization of multi-omics data
- **MeltDB** – Individual and combinatorial analysis of raw GC- and LC-MS metabolome datasets
- **QuPE** – A web application for metabolic labeling-based quantitative proteomics
- **EMMA2** – Evaluation of microarray experiment datasets
- **CARMEN** – Functional and comparative analysis of metabolic networks
- **BiBiServ Tools** – Includes more than 50 software tools and various educational media

▼ *Reconstruction of microbial diversity* on sugar beet surface using Krona-Plot.
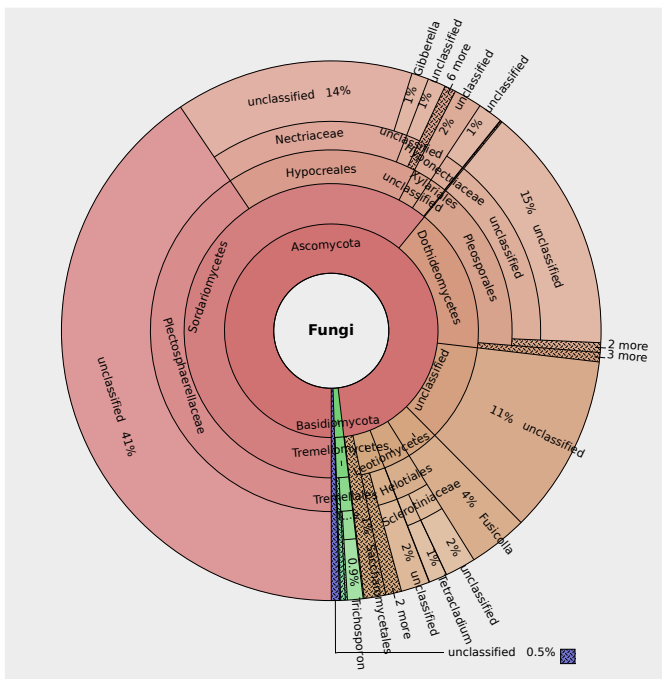


▼ *Omics Fusion:* Mapping of transcriptome and metabolome data on TCA cycle (Brink et al. 2016).
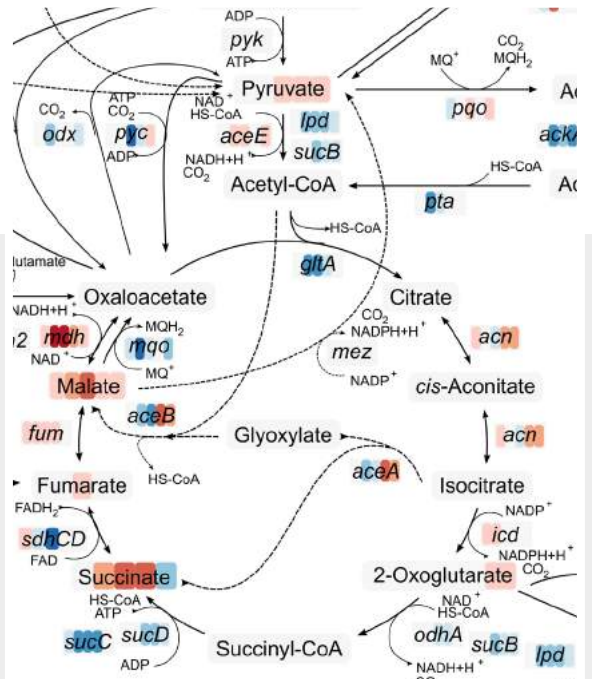
# THE de.NBI PROJECT BiGi 2

## Microbial Genome and Metagenome Research

**The BiGi 2 project at Justus Liebig University in Giessen offers tools, services and training for microbial genome and metagenome research. Services include microbial genome annotation, comparative analyses, short-read-mapping data evaluation and read-based metagenome analysis.**

▼ *The BiGi 2 team at Justus Liebig University Giessen. Contact: bigi@computational.bio*

**Alexander Goesmann**
*Project leader*
alexander.goesmann@computational.
bio.uni-giessen.de

**Oliver Rupp**
oliver.rupp@computational.bio.
uni-giessen.de

**Sebastian Jaenicke**
sebastian.jaenicke@computational.
bio.uni-giessen.de

### COLLECTION OF TOOLS AND SERVICES OFFERED BY BiGi 2:

- **GenDB** – Automatic and manual annotation of microbial genomes
- **GenDBE** – Interactive genome browser for eukaryotic genomes
- **EDGAR** – Enhanced software platform for comparative gene content analyses
- **ReadXplorer** – Visualization and analysis of short-read mapping data
- **MGX** – Taxonomic and functional characterization of metagenome sequencing reads
- **Conveyor** – Versatile workflow engine for automated bioinformatics analyses
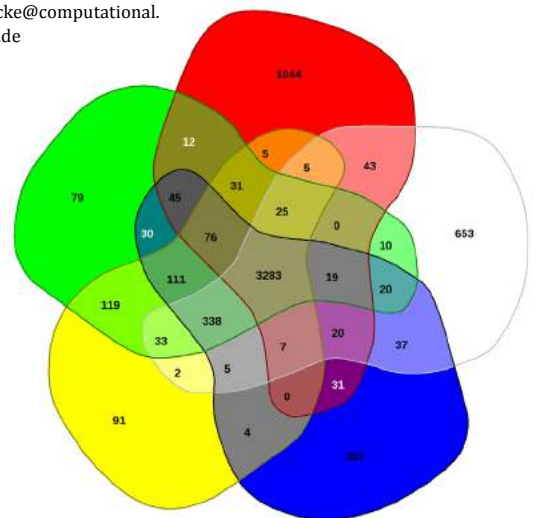


▲ *Venn diagram* generated with EDGAR to illustrate shared and unique gene sets of five Xanthomonas campestris genomes (Blom et al., 2016).
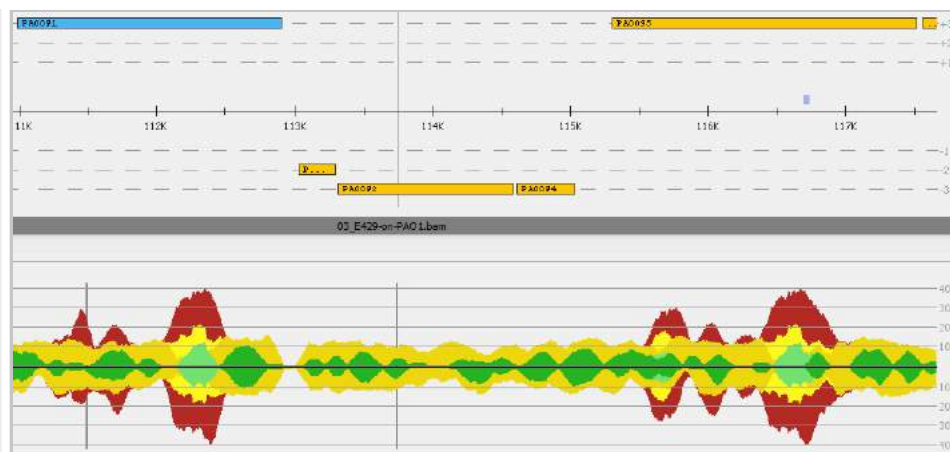


◄ *Screenshot of a short read mapping data set* visualized with the ReadXplorer software (Hilker et al., 2016). The different colors in the lower part display read mappings of varying quality according to the ReadXplorer classification.

# THE de.NBI PROJECT MetaProtServ 1

## Protein Identification, Functional and Taxonomic Investigation of Microbial Communities

**The MetaProtServ 1 and 2 projects at Otto von Guericke University Magdeburg offer the MetaProteomeAnalyzer software as well as services and training for the analysis of microbial communities with metaproteomics.**

▼ *The MetaProtServ 1 team at Otto von Guericke University Magdeburg.*

**Dirk Benndorf**
*Project leader*
benndorf@mpi-magdeburg.mpg.de

**Kay Schallert**
*Application developer*
kay.schallert@ovgu.de

### COLLECTION OF TOOLS AND SERVICES OFFERED BY MetaProtServ 1:

- Graphical user interface for easy processing of data by wet-lab scientists
- Database searches using multiple search engines (X!Tandem, OMMSA, Crux, InsPect, MASCOT)
- BLAST search of protein hits from non-annotated metagenomes against Uni ProtKB/Swiss-Prot
- Grouping of redundant protein hits to metaproteins (shared peptide, peptide sets, UniProt Reference Clusters)
- Functional and taxonomical (lowest common ancestor) assignment of meta-proteins based on UniProtKB/Swiss-Prot metadata
- Export and visualization of data (KEGG maps, KronaPlots)
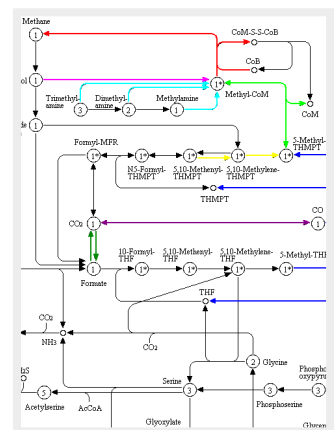


▲ *Plotting of identified proteins into KEGG maps.*
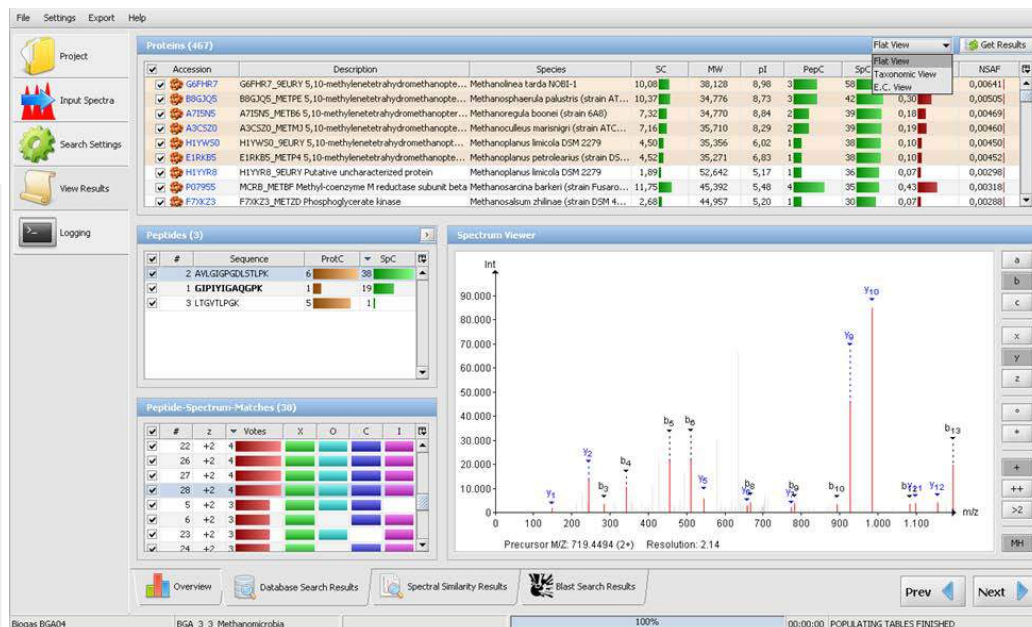


◄ *Graphical user interface of MetaProteomAnalyzer Software (Muth et al., 2015).*

# THE de.NBI PROJECT MetaProtServ 2

## Web-Based Metaproteomics as Highly-Scalable Service

**The MetaProtServ 1 and 2 projects at Otto von Guericke University Magdeburg offer the MetaProteomeAnalyzer software as well as services and training for the analysis of microbial communities with metaproteomics.**

▼ *The MetaProtServ 2 team at Otto von Guericke University Magdeburg.*

**Gunter Saake**
*Project leader*
saake@iti.cs.uni-magdeburg.de

**Roman Zoun**
*Application developer*
Roman.zoun@ovgu.de

### COLLECTION OF TOOLS AND SERVICES OFFERED BY MetaProtServ 2:

- Graphical user interface for easy processing of data by wet-lab scientists
- Storage of raw data and search results in a SQL database at server
- Result processing and export at client
- SQL database is highly scalable regarding the amount and open for incorporation of new types of data
- Easy reprocessing of search results based on data and previous results stored in the database
- Provision of MetaProteomAnalyzer as web service is intended

▼ *Scheme of MetaProteomAnalyzer Software* (Muth et al., 2015).
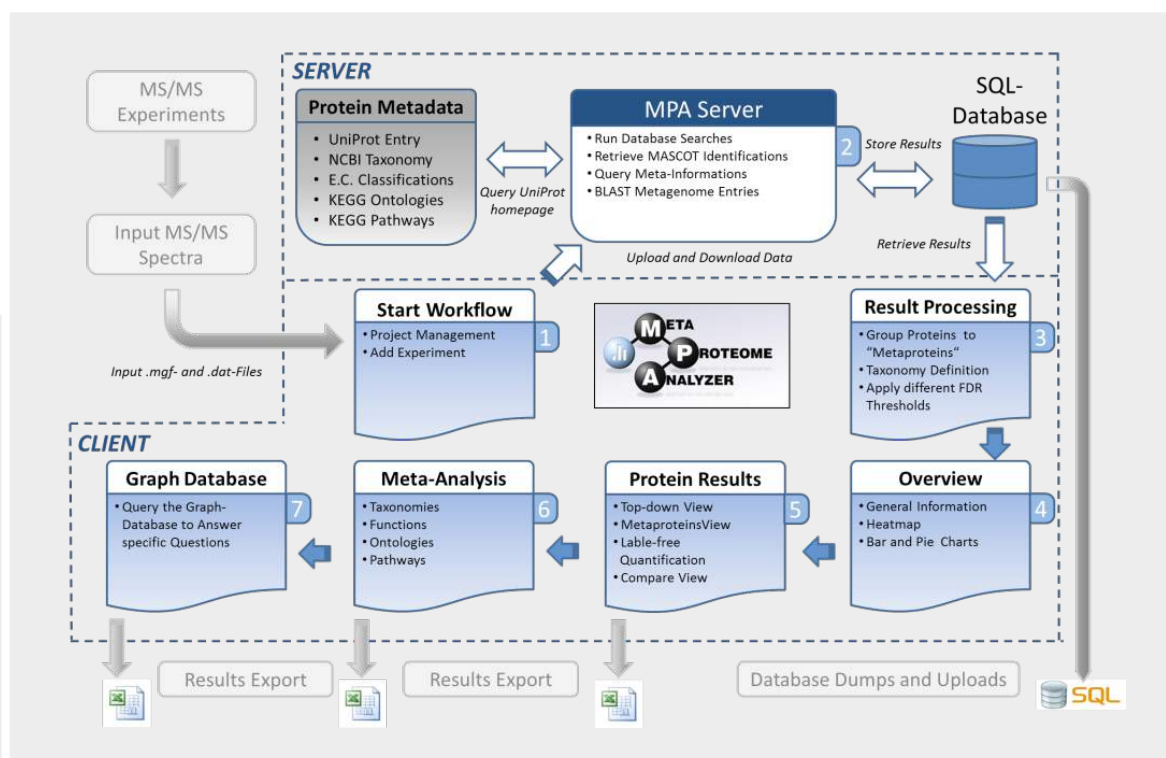
# THE de.NBI SERVICE CENTER BioInfra.Prot

## Bioinformatics for Proteomics

**The Service Center BioInfra.Prot provides standardization, consultancy and analysis services as well as tools, training and quality standards for proteomics and lipidomics.**

The Service Center BioInfra.Prot offers a comprehensive portfolio of services for mass spectrometry-based proteomics and lipidomics. This includes data standardization and conversion services, bioinformatical and statistical consulting and data analysis, software tools, quality standards, hardware sharing as well as training courses.

**TOPICS OF THE SERVICE CENTER BioInfra.Prot:**

- Proteomics: Data standardization and conversion, protein inference, quality standards, expression analysis and bioinformatical and statistical consulting
- Lipidomics: Bioinformatics for targeted and comparative lipidomics including lipid identification and discovery

**SELECTION OF TRAINING COURSES CARRIED OUT BY BioInfra.Prot:**

- ´8th OpenMS User Meeting´ at MPC (Bochum)
- Fundamentals of Proteome Bioinformatics Revisited
- Spezielle Aspekte von Bioinformatik-Methoden in der MS-basierten Proteomik
- de.NBI-Summer School 2016, R course and EUBIC Winter School 2017

**CURRENT PROJECTS:**

- **BioInfra.Prot 1** at **Ruhr-Universität Bochum**: Comprehensive proteomics workflow including data standardization, protein inference, expression analysis and data publication
- **BioInfra.Prot 2** at **ISAS Dortmund**: Proteomics quality standards
- **LIFS 1** at **ISAS Dortmund**: Targeted lipidomics
- **LIFS 2** at **Research Center Borstel**: Informatics for comparative lipidomics
- **LIFS 3** at **MPI Dresden**: Lipid identification and discovery

▼ *Members of the Service Center BioInfra.Prot. From left to right: Robert Ahrends, Julian Uszkoreit, Maike Ahrens, Dominik Kopczynski, Michael Turewicz, Martin Eisenacher, Michael Kohl and Gerhard Mayer (November 2016).*



**Martin Eisenacher**
*Coordinator*
*Ruhr-Universität Bochum*
martin.eisenacher@rub.de

**Albert Sickmann**
*Deputy coordinator*
*ISAS Dortmund*
sickmann@isas.de

**Robert Ahrends**
*LIFS coordinator*
*ISAS Dortmund*
robert.ahrends@isas.de

# THE de.NBI PROJECT BioInfra.Prot 1

## Bioinformatics for Proteomics Services

**The BioInfra.Prot 1 project at Ruhr-Universität Bochum provides a comprehensive proteomics workflow including data standardization, protein inference, expression analysis and data publication as well as bioinformatical and statistical consulting.**

▼ *The BioInfra.Prot 1 team at Medizinisches Proteom-Center at Ruhr-Universität Bochum.*



**Martin Eisenacher**
*Project leader*
martin.eisenacher@rub.de

▲ *From left to right: Michael Krafzik, Michael Turewicz, Julian Uszkoreit, Michael Kohl, Martin Eisenacher, Gerhard Mayer and Maike Ahrens*

### COLLECTION OF TOOLS AND SERVICES OFFERED BY BioInfra.Prot 1:

- **ProCon** – Tool for the conversion of proteomics data formats into standard data formats
- **Data standardization and conversion service** – Consultancy and service regarding the conversion and repository upload of proteomics data
- **Bioinformatical and statistical consulting and analysis of proteomics data**
- **PIA** – Tool for flexible protein inference and identification analysis using different algorithms
- **PAA** – R/Bioconductor tool for protein microarray data analysis for biomarker discovery
- **CrossPlatformCommander** – Multi omics data analysis workflow within KNIME
- **ProLiC** – Different algorithms for the comparison of protein identification result lists

▼ *The ProCon conversion workflow (Mayer et al., 2015).*
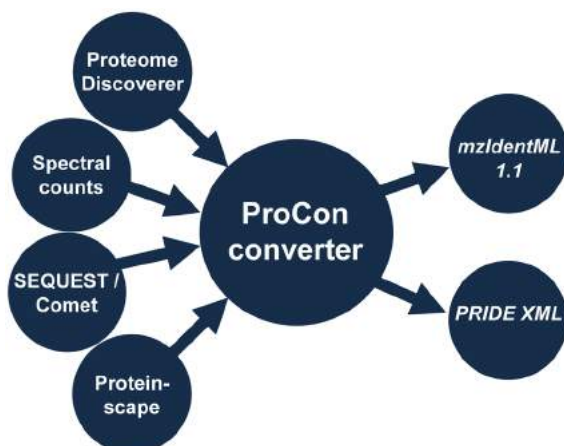
▼ *The PAA workflow for biomarker discovery with protein microarray data (Turewicz et al., 2016).*
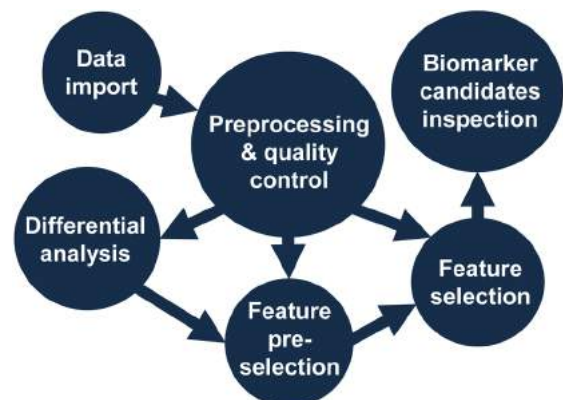
# THE de.NBI PROJECT BioInfra.Prot 2
## Bioinformatics for Proteomics Services

**The BioInfra.Prot 2 project at ISAS Dortmund provides a variety of bioinformatics services and tools for computational proteomics.**

▼ *The BioInfra.Prot 2 team at ISAS.*

**Albert Sickmann**
*Project leader*
sickmann@isas.de

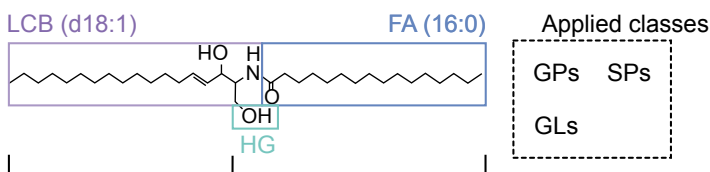▲ *From left to right: Jens Hinrichs, Jannis Schulte, Robert Ahrends, Dominik Kopczynski*

### COLLECTION OF TOOLS AND SERVICES OFFERED BY BioInfra.Prot 2:

- **Quality Standard Database (QSDB)** – Protein database for creating spectral libraries
- **SearchGui** – Graphical user interface for proteomics identification search engines
- **PeptideShaker** – Interpretation of proteomics identification results
- **IceLogo** – Next-generation visualization of protein consensus sequences
- **Web-based PeptideShaker** – Interpretation of proteomics results as a web service
- **Easy QC** – Toolbox for measuring the quality of different MS measurements from one experiment

▼ *QSDB is a database containing high-resolution reference spectra of peptides for targeted proteomics.*
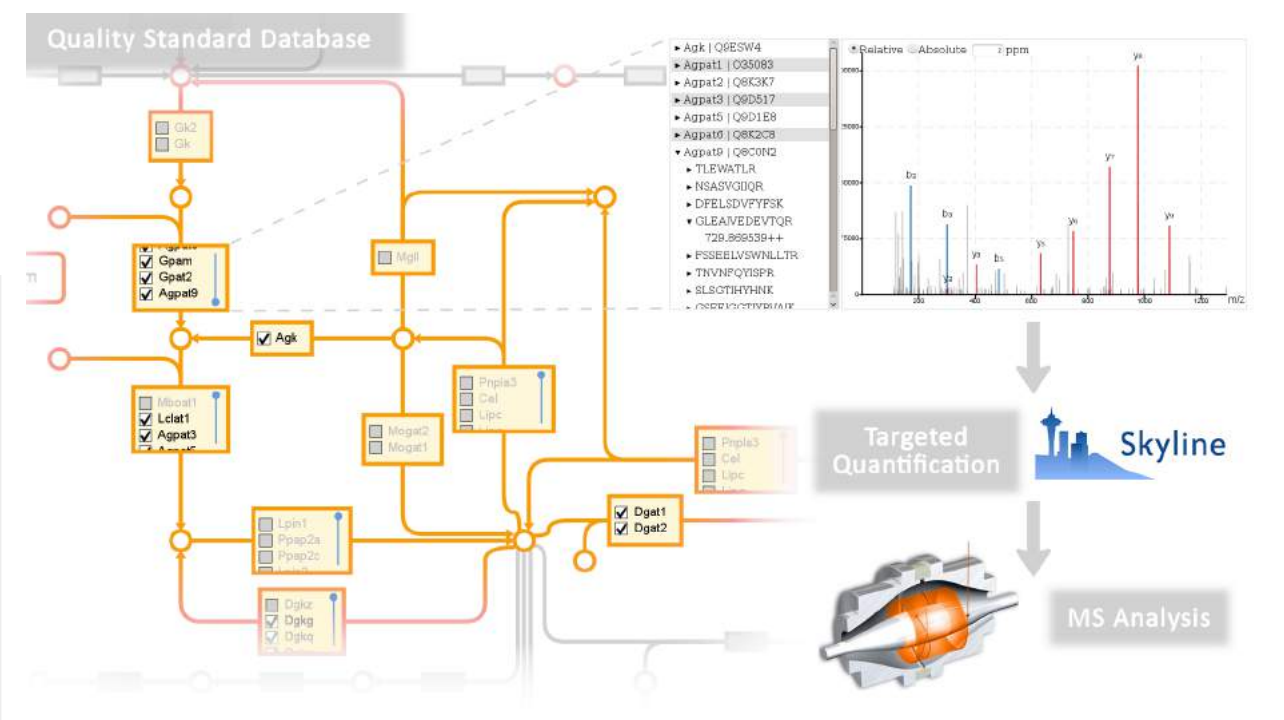
# THE de.NBI PROJECT LIFS 1

## Bioinformatics for Lipidomics Services

**The LIFS 1 project at Leibniz-Institut für Analytische Wissenschaften - ISAS e.V. in Dortmund provides a variety of bioinformatics services and tools for computational lipidomics.**

▼ *The LIFS 1 team at ISAS Dortmund.*

**Robert Ahrends**
*Project leader*
robert.ahrends@isas.de

▲ *From left to right: Jens Hinrichs, Jannis Schulte, Robert Ahrends, Dominik Kopczynski*

**COLLECTION OF TOOLS AND SERVICES OFFERED BY LIFS 1:**

- **LipidHome** – Structural database for registration of identified lipids
- **Skyline for Lipidomics** – Method development for targeted lipidomics
- **LIFS web portal** – Service platform for tools and service requests

▼ *Modification principle for Skyline: Lipid building blocks will be utilized to create transitions for targeted lipidomics.*

LCB or FA — HG — FA

- Positive mode
- Negative mode
- Positive or negative mode

Example: Cer d18:1/16:0

LCB (d18:1)     FA (16:0)

HO  H
N
O
OH
HG

Applied classes

GPs     SPs

GLs

Sequence: Q (Glutamine) C (Cysteine) G (Glycine)

Calculation of structural modification by modifying amino acids masses (in pos/neg mode):

LCB part : Modified mass [Q] = Mass [LCB] - Mass [Q] + FA Constant 1 (or FA)
HG part  : Modified mass [C] = HG Constant 1
FA part   : Modified mass [G] = Mass [FA] - Mass [G] + FA Constant 2

Category
Glycerophospholipids

Main Class
Glycerophospholine

Sub Class
Diacylglycerophosphocholine

Species
PC 36:2

Fatty Acid Chain Species
PC 18:0_18:2

Sub Species
PC 18:0/18:2

Isomer
PC 18:0/18:2[3,6]
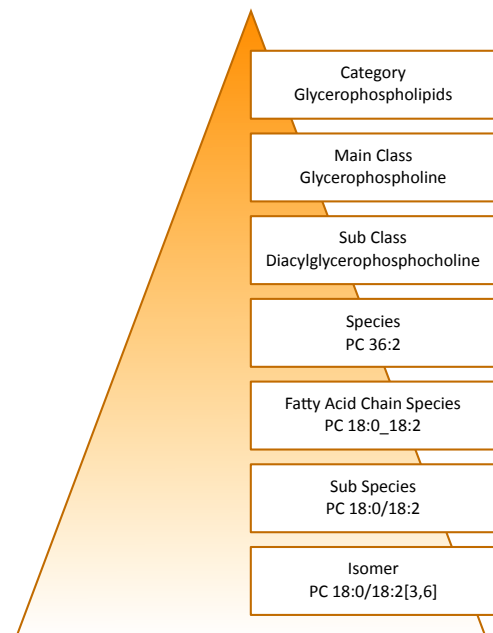
▲ *Supposed nomenclature for the LipidHome database.*

# THE de.NBI PROJECT LIFS 2

## Bioinformatics for Lipidomics Services

**The LIFS 2 project at Research Center Borstel provides bioinformatics services and tools for systematic lipidome comparisons and shotgun lipidomics.**

▼ *The LIFS 2 team at Research Center Borstel.*

**Dominik Schwudke**
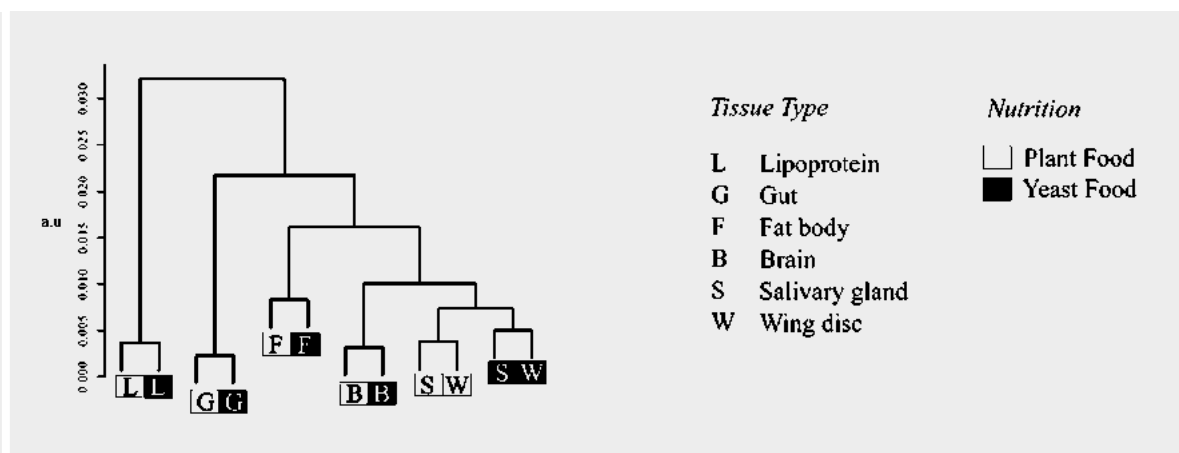*Project leader*
dschwudke@fz-borstel.de

### COLLECTION OF TOOLS AND SERVICES OFFERED BY LIFS 2:

- **Lipidhome** – Resource on lipids, identification, ontology and structures
- **LUX Score** – Systematic lipidome comparison
- **Scorelip** – A general approach to verify lipid IDs
- **LIFS web portal** – Service platform for tools and service requests
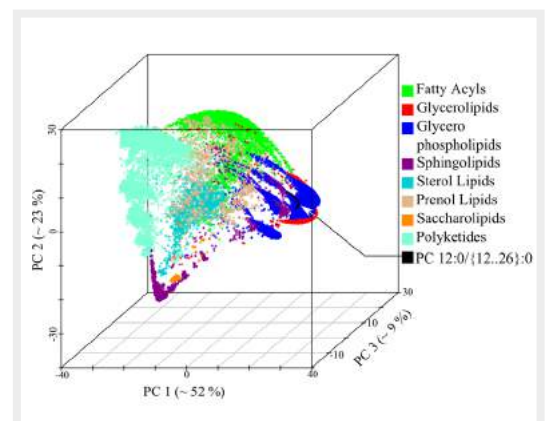


▲ *Representation of structural relation of 30.150 lipids in a chemical space model obtained from the lipid maps structure database (LMSD). Structural related phosphatidylcholines are highlighted.*

▼ *Clustering of Drosophila tissue lipidomes using the* **LUX score** *as metric (Marella et al., 2015).*



Tissue Type
L   Lipoprotein
G   Gut
F   Fat body
B   Brain
S   Salivary gland
W   Wing disc

Nutrition
☐ Plant Food
■ Yeast Food

# THE de.NBI PROJECT LIFS 3

## Bioinformatics for Lipidomics Services

**The LIFS 3 project at Max Planck Institute of Molecular Cell Biology and Genetics in Dresden is a partner of the LIFS service unit, which provides software and support for lipidomics analysis by high-resolution shotgun mass spectrometry.**

▼ *The LIFS 3 team at MPI Dresden.*

**Andrej Shevchenko**
*Project leader*
shevchenko@mpi-cbg.de

### COLLECTION OF TOOLS AND SERVICES OFFERED BY LIFS 3:

● **LipidXplorer** – Universal software for shotgun lipidomics on any mass spectrometry platform
● Software for pre-processing shotgun mass spectra and removal chemical noise
● Software for quantification of polyunsaturated lipids
● **LIFS web portal** – Service platform for tools and service requests

▼ **LipidXplorer software** *identifies lipids by user-defined queries encompassing essential features of their mass spectra (Herzog et al., 2011).*
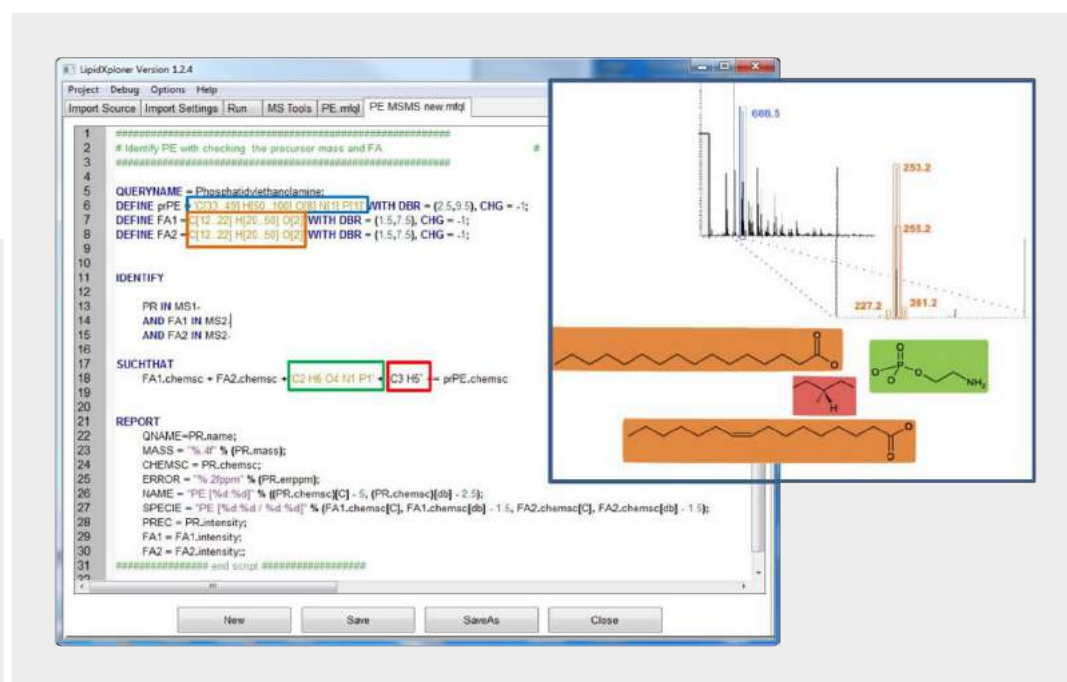
# THE de.NBI SERVICE CENTER CIBI

## Center for Integrative Bioinformatics

**The CIBI Service Center offers computational tools, workflows and training for integrated omics and image data analysis.**

The BioInfra.Prot Center provides users cutting-edge and integrative tools for proteomics, metabolomics, NGS and image data analysis as well as workflow engine to integrate tools into coherent solutions for reproducible analysis of large-scale biological data.

### TOPICS OF THE SERVICE CENTER CIBI:

- Implementation of reliable, robust, open, and interoperable workflow tools permitting a complete modeling of data analysis
- Providing collection of standards for reproducible analysis of mass spectrometric, NGS and metabolomics data

### SELECTION OF TRAINING COURSES CARRIED OUT BY CIBI:

- User and developer meetings for all tools supported by CIBI
- KNIME Summit Workshops
- Courses on omics and image data analysis associated with larger national and international life science conferences

### CURRENT PROJECTS:

- **CIBI 1** at **University of Tübingen**: Software tools and libraries for the analysis of mass spectrometric data (OpenMS) and node of the de.NBI cloud
- **CIBI 2** at **FU Berlin**: Software tools and libraries for NGS analysis (SeqAn)
- **CIBI 3** at **University of Konstanz**: Software for workflow management and large-scale data analysis (KNIME)
- **DAIS** at **MPI Dresden**: Software and libraries for image analysis
- **MASH** at **Leibniz Institute of Plant Biochemistry, Halle (Saale)**: Analysis of metabolomics data

▼ *Members of the Service Center CIBI. From left to right: Eugene Myers, Tobias Pietsch, Michael Berthold, Florian Jug, René Meier, Oliver Kohlbacher, Pavel Tomancak, Knut Reinert, Svenja Mehringer, René Rahn, Julianus Pfeuffer, Alexander Fillbrunn, Steffen Neumann, Timo Sachsenberg (November 2016).*



**Oliver Kohlbacher**
*Coordinator*
*University of Tübingen*
oliver.kohlbacher@uni-tuebingen.de



**Knut Reinert**
*Deputy coordinator*
*FU Berlin*
knut.reinert@fu-berlin.de

# THE de.NBI PROJECT CIBI 1

## OpenMS – An Open-Source Framework for Mass Spectrometry Data Analysis

**The CIBI 1 project at University of Tübingen offers computational tools, workflows and training for the analysis of mass spectrometry-based proteomics and metabolomics data.**

▼ *The CIBI 1 team at University of Tübingen.*



**Oliver Kohlbacher**
*Project leader*
oliver.kohlbacher@uni-tuebingen.de

**Julianus Pfeuffer**
j.pfeuffer@fu-berlin.de

**Timo Sachsenberg**
sachsenb@informatik.uni-tuebingen.de

### CIBI 1 OFFERS THE OpenMS SOFTWARE, WHICH PROVIDES USERS WITH:

- Open-source mass spectrometry data analysis framework
- Large library of proven algorithms for rapid software development
- Builds on open, standardized data formats
- Designed for high-throughput data processing and analysis
- More than 180 tools for proteomic, metabolomics, and lipidomics analysis tasks
- Available for Microsoft Windows, Mac OS and Linux
- Chaining of tools allows building powerful workflows
- Integrated into various workflow systems and portals
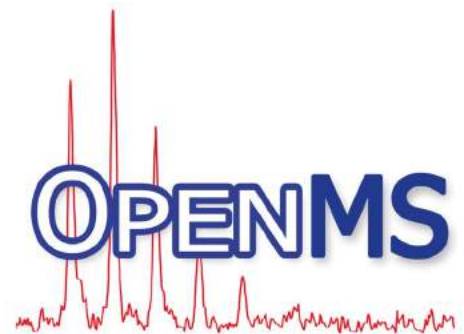


▼ **OpenMS** *provides a wide range of distinct tools that can be combined into complex data analysis workflows.*
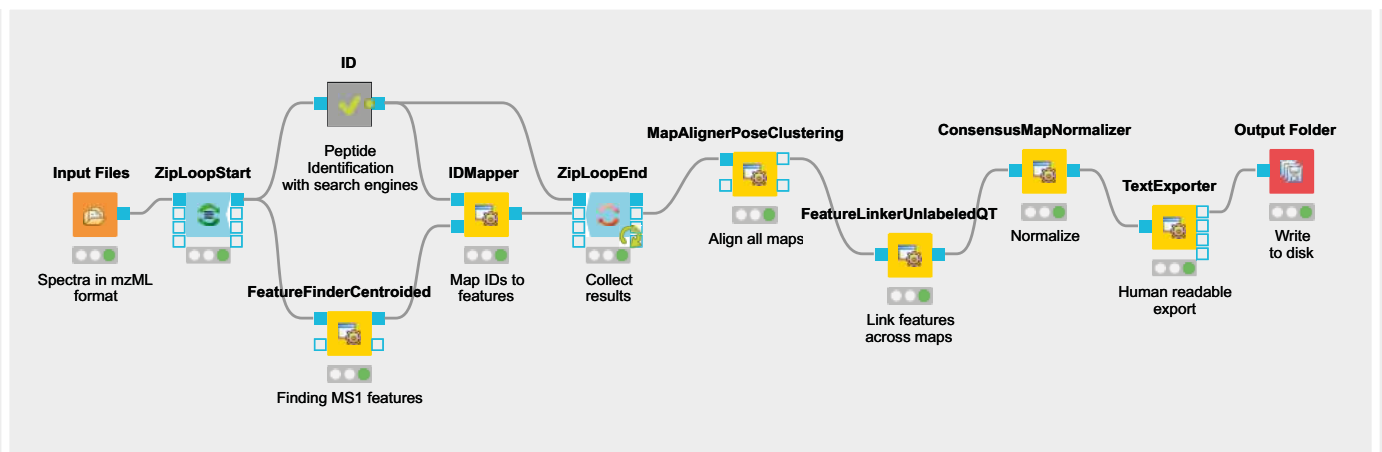
# THE de.NBI PROJECT CIBI 2

## SeqAn: A C++ Template Library for Biological Sequence Analysis

**The CIBI 2 project at Freie Universität Berlin develops a generic C++ library of efficient algorithms and data structures for the analysis of sequences with the focus on biological data.**

▼ *The CIBI 2 team at Freie Universität Berlin.*

**Knut Reinert**
*Project leader*
knut.reinert@fu-berlin.de

**René Rahn**
rene.rahn@fu-berlin.de

**Svenja Mehringer**
svenja.mehringer@fu-berlin.de

### CIBI 2 OFFERS THE SeqAn SOFTWARE, WHICH PROVIDES USERS WITH:

- Open source C++ library
- Support for C++11/14 and different platforms and compilers: gcc, clang, VisualStudio, icc
- Unique generic design that guarantees high performance, generality, extensibility, and integration with other libraries
- Simplifies development of new software tools
- Generic parallelization of core algorithms
- Rang of state-of-the-art tools for NGS and data analysis of biological data
- Exhaustive support for biological sequence, alignment and annotation formats, including fastq, sam, bam, gtf, blast, …

▼ *Journaled String Tree – A scalable data structure for analyzing thousands of similar genomes on your laptop.*
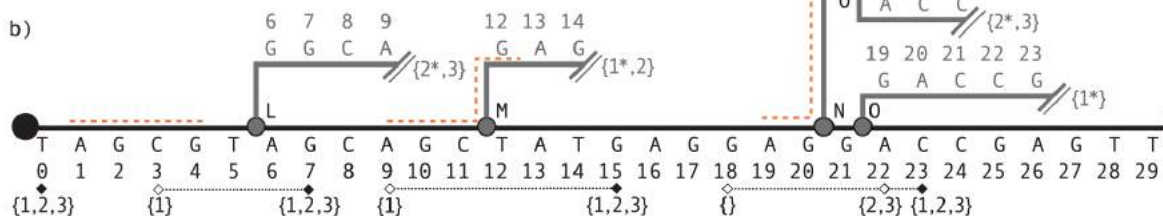
# THE de.NBI PROJECT CIBI 3

## For Data Processing and Mining (KNIME)

**The CIBI 3 project at University of Konstanz develops and maintains the open source analytics platform KNIME, which provides an extensible toolbox for data processing, data mining and bioinformatics.**

▼ *The CIBI 3 team at University of Konstanz.*

**Michael Berthold**
*Project leader*
michael.berthold@uni-konstanz.de

**Alexander Fillbrunn**
alexander.fillbrunn@uni-konstanz.de

**CIBI 3 OFFERS THE KNIME SOFTWARE, WHICH PROVIDES USERS WITH:**

- Open-source data analysis platform
- More than 1,000 data processing nodes
- Easily extensible through plugin system
- Connectors for all major file formats and databases
- Tool blending for SQL, Java, Weka, and many more
- Extensions, among others, for mass spectrometry (OpenMS), sequence analysis (SeqAn, pf2), image processing (ImageJ, …), cheminformatics (RDKit, CDK), and Integrations of R, Python, JavaScript, and more

Open for Innovation ®
**KNIME**

▼ *The KNIME Analytics Platform – Thousands of data processing nodes for Advanced Analytics at your fingertips.*
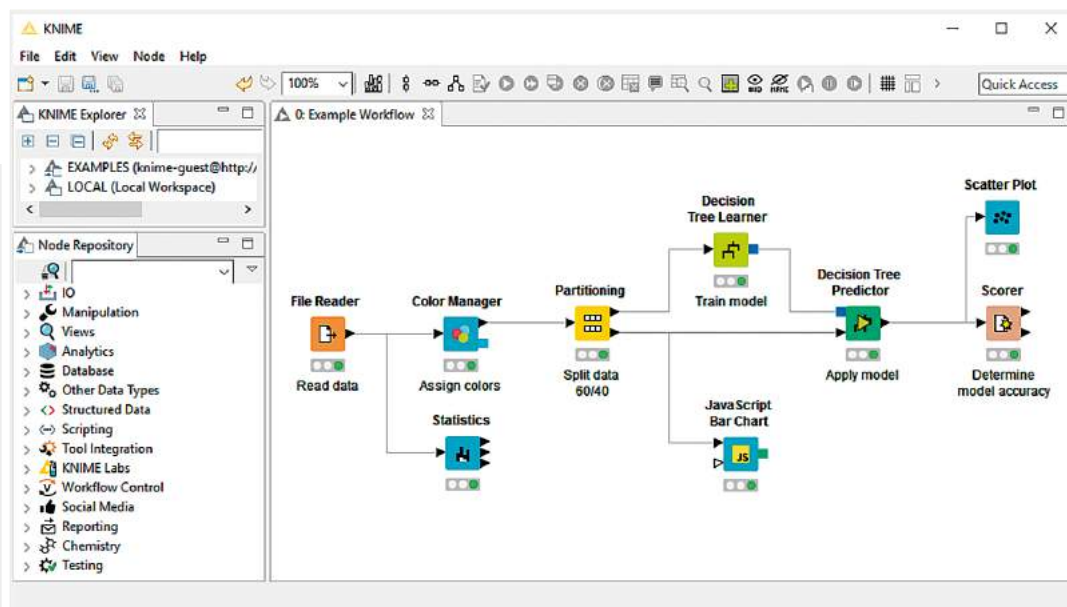
# THE de.NBI PROJECT DAIS

## The Dresden Analysis-of-Images Suite

**The DAIS project at the Center for Systems Biology Dresden offers services and training for the analysis of and quantification of biological image data.**

▼ *The DAIS team at the Center for Systems Biology Dresden.*

**Eugene Myers**
*Project leader*
myers@mpi-cbg.de

**Pavel Tomancak**
tomancak@mpi-cbg.de

**Florian Jug**
jug@mpi-cbg.de

**Tobias Pietzsch**
pietzsch@mpi-cbg.de

**DAIS OFFERS A SOFTWARE STACK FOR BIO-IMAGE ANALYSIS BASED ON THE Fiji AND KNIME PLATFORMS THAT PROVIDES THE USER WITH:**

- A large library of basic and advanced image analysis tools that is maintained, improved, and expanded upon as the field advances.
- User-friendly versions of cutting-edge image processing methods developed within our research labs in Dresden.
- For developers, we provide, maintain, and extend essential and powerful software libraries such as ImgLib2 – the modern image container library used in Fiji and in KNIME Image Processing.

Crucially, the Fiji community is actively working towards tight integration with the KNIME workflow engine, which is maintained and developed by the de.NBI project partner **CIBI 3**.



| Nr. of citations (Schindelin et al. 2012) | Nr. of citations (8 Tomancak lab plugin papers) | Nr. of downloads (2015) | Nr. of users (estimate) | Nr. of unique wiki visits per day | Nr. of wiki visits per month |
|---|---|---|---|---|---|
| **2,395** | 719 | 589,840 | 50,000 | >13,000 | >330,000 |

◀ *Fiji usage statistics: Geo-locations of unique IP addresses that updated Fiji (A) world-wide within one calendar day and (B) in Germany within one representative week in 2015. The numbers of citations are based on Google Scholar.*

# THE de.NBI PROJECT MASH
## Metabolite Annotation and Sharing – Halle

**The MASH project at the Leibniz Institute of Plant Biochemistry in Halle provides training, maintenance and workflow integration of metabolite annotation software, and support for standards-compliant metabolomics and data sharing.**

▼ *The MASH team at the Leibniz Institute of Plant Biochemistry Halle.*

**Steffen Neumann**
*Project leader*
sneumann@ipb-halle.de

**René Meier**
rmeier@ipb-halle.de

**Hendrik Treutler**
htreutle@ipb-halle.de

### COLLECTION OF TOOLS AND SERVICES OFFERED BY MASH:

- **MassBank** – Software platform for Metabolomics spectral reference libraries
- **MetFrag** – In silico identification of Metabolites from mass spectrometry data
- **MetFusion** – Integrated identification of Metabolites using MassBank and MetFrag
- **MetaboLights** – Submission to metabolomics repository @ EMBL-EBI
- **ISA-Tab** – Dataset description format used in MetaboLights

▼ *Services offered by MASH: Metabolite annotation services (blue), Training on metabolomics annotation, data standards and sharing (green) and related de.NBI services (yellow).*
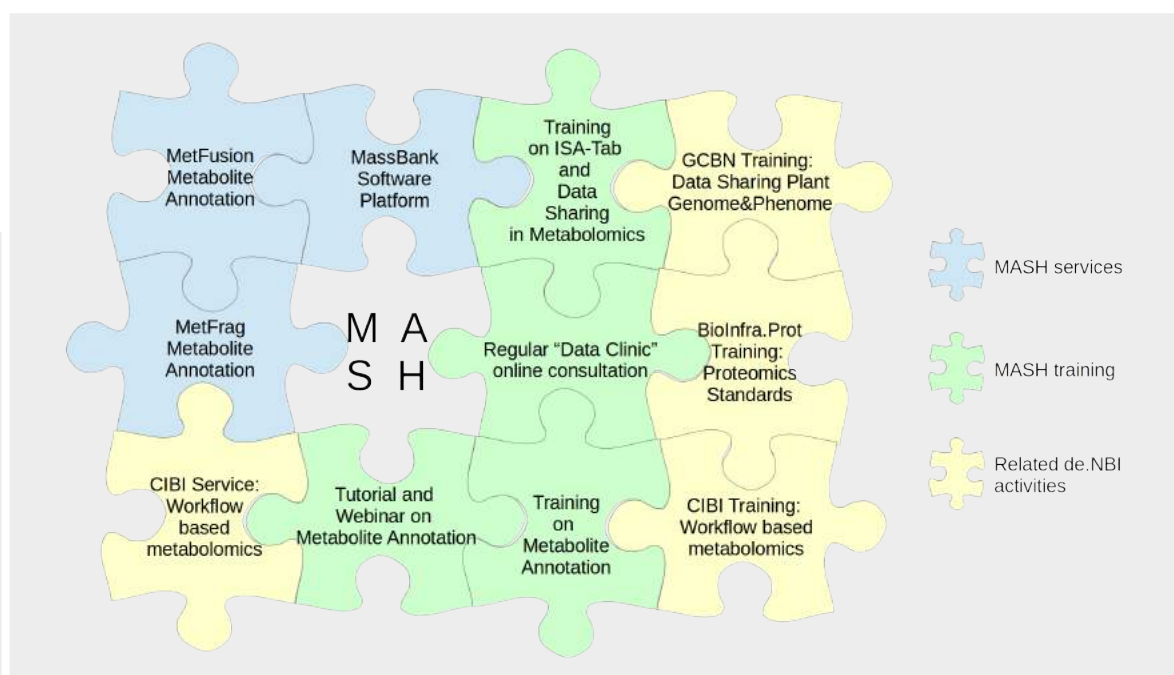
# THE de.NBI SERVICE CENTER RBC
## RNA Bioinformatics Center

**The RBC Service Center offers tools, services and training covering all aspects of RNA bioinformatics, meta-transcriptomics and epigenetic research.**

The RBC Service Center establishes an integrated, easily accessible scientific workbench based on GALAXY covering a wide range of RNA-related and epigenetic analysis tasks. The workbench can be used on a local cluster or deployed on any HPC-like environment. The RBC Service Center is equipped with extended compute clusters and participates in the de.NBI Cloud.

### TOPICS OF THE SERVICE CENTER RBC:

- Ready-to-use pipelines for the analysis of RNA-related high-throughput sequencing data
- RNA-protein interactions: advanced approaches for the analysis of genome-wide protein-RNA interaction experiments
- Analysis of high-throughput data related to epigenetic modifications

### SELECTION OF TRAINING COURSES CARRIED OUT BY RBC:

- Galaxy Workshop on high-throughput sequencing data analysis
- Computational Genomics hands-on workshop
- ELIXIR/GOBLET/GTN Hackathon on Galaxy training material re-use

### CURRENT PROJECTS:

- **RBC 1** at **University of Freiburg**: Analysis of RNA interactions
- **RBC 2** at **Leipzig University**: Analysis of non-coding RNAs
- **RBC 3 & 4** at **MDC Berlin**: RNA-binding proteins and post-transcriptional regulation
- **de.STAIR 1** at **Leipzig University**: Alternative splicing and differential expression
- **de.STAIR 2** at **University of Freiburg**: Regulatory RNA and data
- **de.STAIR 3** at **University of Rostock**: Customized workflows
- **de.NBI-epi 3** at **MDC Berlin**: Epigenomics data integration and analysis
- **de.NBI-epi 4** at **University of Freiburg**: The epigenetic workbench

▼ ***Members of the Service Center RBC.*** *From left to right: Dilmurat Yusuf, Jörg Fallmann, Björn Grüning, Sebastian Will, Joachim Wolff, Olaf Wolkenhauer, Andrea Bagnacani, Torsten Houwaart, Markus Wolfien, Altuna Akalin, Rolf Backofen, Vedran Franke, Anika Erxleben, Bérénice Batut, Wolfgang R. Hess, Uwe Ohler, Bora Uyar, Steve Hoffmann (November 2016).*

**Rolf Backofen**
*Coordinator*
*University of Freiburg*
backofen@informatik.uni-freiburg.de

**Uwe Ohler**
*Deputy coordinator*
*MDC Berlin*
Uwe.Ohler@mdc-berlin.de

# THE de.NBI PROJECT RBC 1

## Analysis of RNA Interactions

**The RBC 1 project at the University of Freiburg offers tools, services and training for the analysis of RNA-RNA and RNA-protein interactions.**

▼ *The RBC 1 team at University of Freiburg.*

**Rolf Backofen**
*Project leader*
backofen@informatik.uni-freiburg.de

**Björn Grüning**
gruening@informatik.uni-freiburg.de

### COLLECTION OF TOOLS AND SERVICES OFFERED BY RBC 1:

- **Freiburg Galaxy Server** – The largest Galaxy instance in Europe with currently 620 users and more than 800 tools; 500 tools out of 3000 world-wide have been integrated by the RBC.
- **IntaRNA** – Fast and accurate prediction of RNA-RNA interactions
- **CopraRNA** – Prediction of small RNA target networks and interaction domains
- **GraphProt** – Modeling binding preferences of RNA-binding proteins
- **GraphClust** – Fast alignment-free structural clustering of local RNA secondary structures
- **BlockClust** – Efficient clustering and classification of ncRNAs from RNA-seq profiles
- **LocaRNA** – Global and local alignment of RNAs
- **ExpaRNA-P** – Simultaneous exact pattern matching and folding of RNAs
- **Training courses** on HTS data analysis, genome annotation and Galaxy

▼ *CopraRNA: Interacting regions plot for the sRNA GcvB showing accessible regions R1 and R2 in 8 homologs (Wright et al., 2013).*
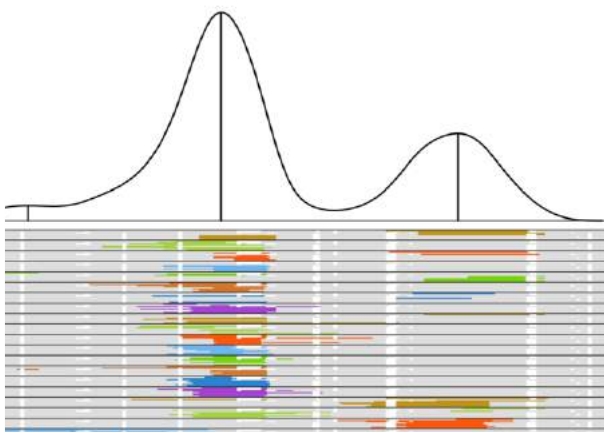
▼ *GraphProt: Encoding of a region identified by CLIP. The secondary structure is represented as a graph and encoded by graph-kernel features (Maticzka et al., 2014).*
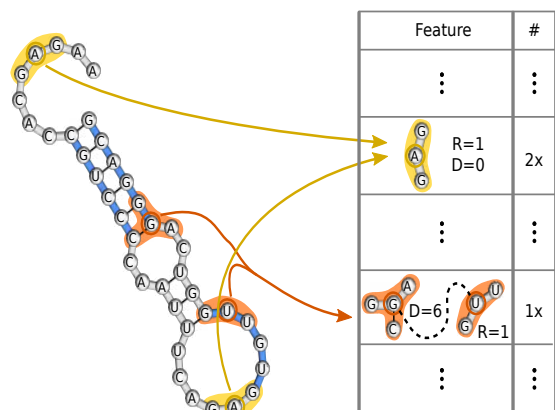
# THE de.NBI PROJECT RBC 2

## Analysis of non-coding RNAs

The **RBC 2** project at the Leipzig University offers tools, services and training for the analysis of non-coding RNAs specializing on advanced secondary structure analysis.

▼ *The RBC 2 team at Leipzig University.*

**Peter Stadler**
*Project leader*
stadler@bioinf.uni-leipzig.de

**Jörg Fallmann**
fall@bioinf.uni-leizig.de

**Sebastian Will**
will@bioinf.uni-leizig.de

### COLLECTION OF TOOLS AND SERVICES OFFERED BY RBC 2:

- **Vienna RNA package** – RNA secondary structure prediction and analysis
- **LocARNA package** – Multiple comparative analysis of RNAs
- **RNAz** – Genome-wide *de novo* prediction of non-coding RNAs
- **CARNA** – Alignment and folding of RNA ensembles and pseudoknotted RNAs
- **RNAplex** – Fast screening for RNA-RNA interactions
- **Kinwalker** – Prediction of the co-transcriptional folding kinetics of large RNAs
- **snoSTRIP** – Analysis of small nucleolar RNAs (snoRNAs) in fungi

- **PLEXY** – Predicting interactions of C/D-box snoRNAs with target RNAs
- **MITOS** – Annotation of metazoan mitochondrial genomes
- **tRNAdb** – Comprehensive database of tRNAs
- **DARIO** – Analysis of short RNAs from high-throughput sequencing data
- **Sierra Platinum** – Peak calling for NGS data with advanced quality control
- **Blockbuster** – Detection of read blocks in NGS mapping data
- **ViennaNGS** – Tools for next generation sequencing analysis

▼ *Vienna RNA and LocARNA package: Example results from selected tools.* **A)** *Multiple structure-based alignment of five tRNAs by LocARNA.* **B)** *Sequence and structure reliability plot of the alignment (STAR plot) of Subfigure A by LocARNA-P.* **C)** *Optimal structure of the alignment by RNAalifold and* **D)** *probability dot plot of the alignment by partition function folding of RNAalifold.*
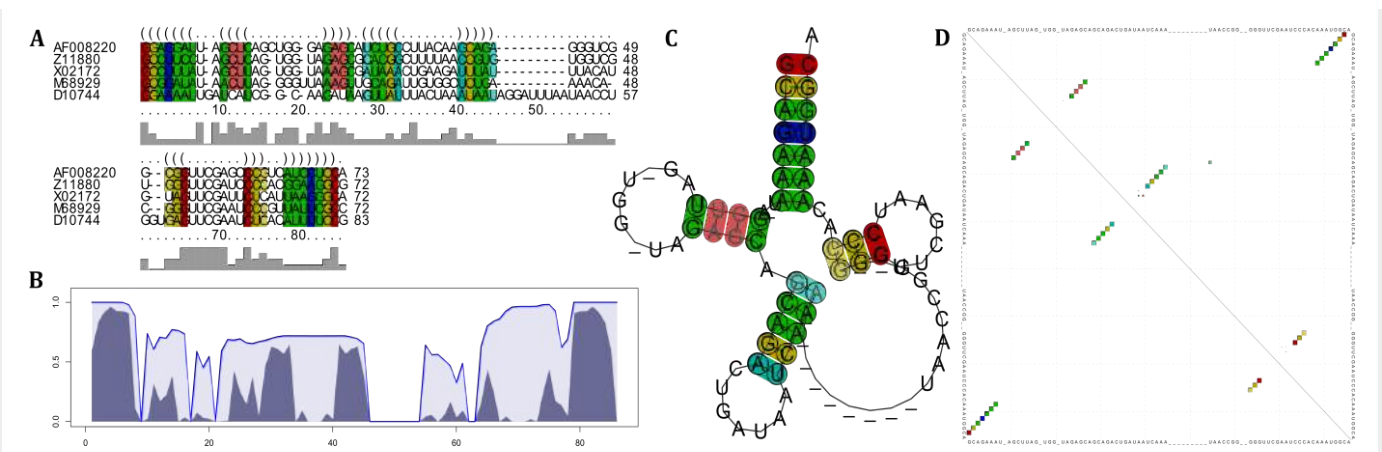
# THE de.NBI PROJECT RBC 3

## RNA-Binding Proteins and Post-Transcriptional Regulation

**The RBC 3 project at Max Delbrück Center for Molecular Medicine offers tools, services and training for the analysis of RNA-binding proteins and post-transcriptional regulation.**

▼ *The RBC 3 team at MDC Berlin.*

**Uwe Ohler**
*Project leader*
uwe.ohler@mdc-berlin.de

**Nikolaus Rajewsky**
*Project leader*
rajewsky@mdc-berlin.de

### COLLECTION OF TOOLS AND SERVICES OFFERED BY RBC 3:

- **DoRiNA database** – Database for RNA-binding protein binding sites and miRNA binding sites from various resources including our in-house tool PicTar2
- **RiboTaper** – Toolbox for analysis of ribosome profiling data
- **RCAS** – Pipeline for annotation of RNA-specific features such as RNA-modification sites and RBP binding sites
- **microMummie** – MicroRNA target site identification by integrating sequence and binding information
- **PARalyzer** – Definition of RNA binding sites from PAR-CLIP short-read sequence data
- **Local Galaxy server** – Aimed for wetlab researchers being able to run bioinformatics tools and analysis pipelines

▼ *DoRiNA: integrates experimental and computational datasets for RBP and miRNA binding sites (Blin et al., 2015).*
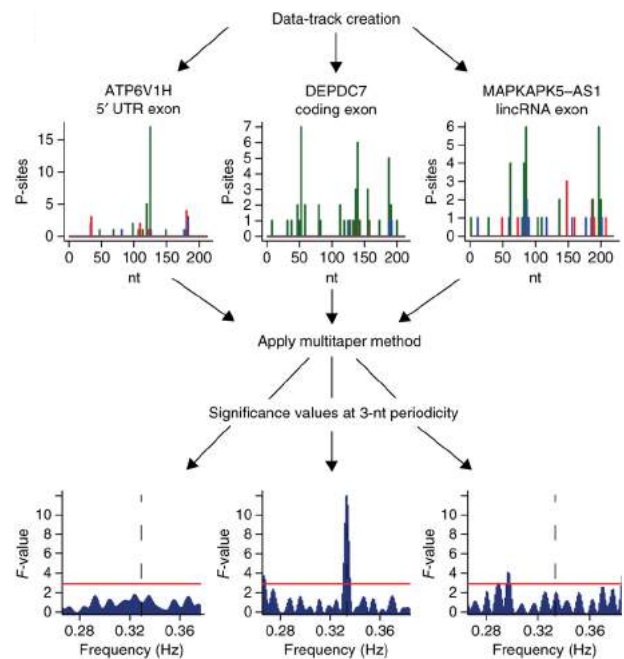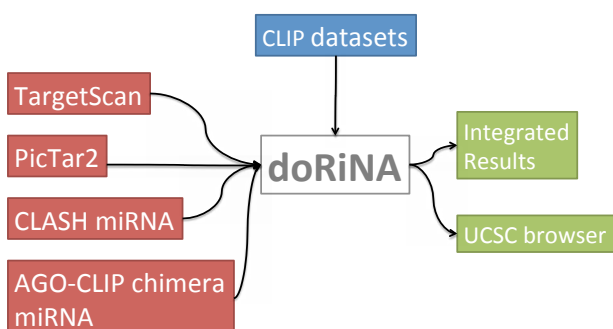




▲ *RiboTaper: P-site positions are color-coded by the frame; the method estimates the significance of each frequency component. Exons from different transcript classes are shown with their periodicity significance (Calviello et al., 2016).*

# THE de.NBI PROJECT de.STAIR 1

## Alternative Splicing and Differential Expression

**The de.STAIR 1** project at Leipzig University implements workflows for the (split)alignment of RNA-Seq reads, detection of transcription changes and the integration of transcriptomic data with epigenetic marks.

▼ *The de.STAIR 1 team at University of Leipzig.*

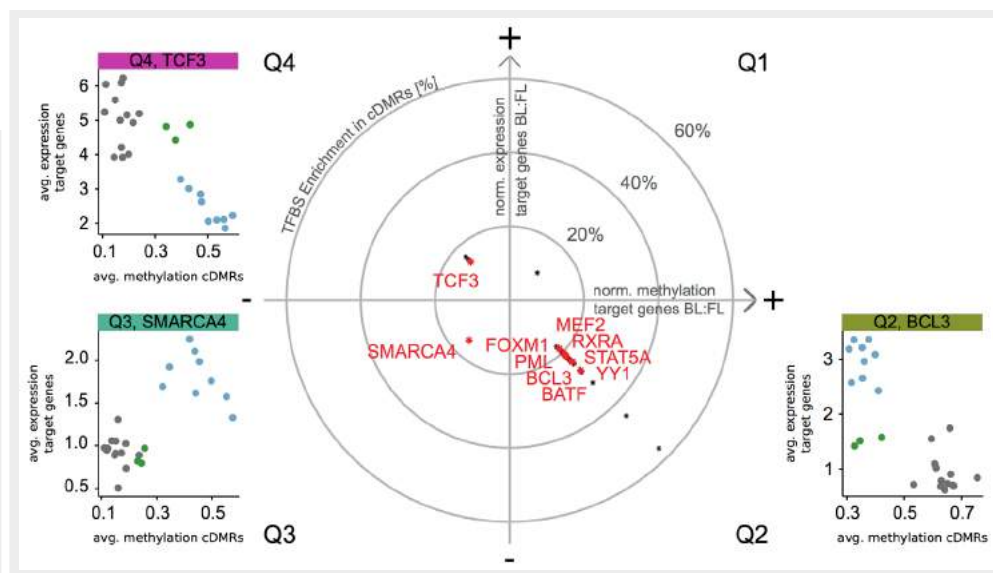**Steve Hoffmann**
*Project leader*
steve@bioinf.uni-leipzig.de

**Gero Doose**
gero@bioinf.uni-leipzig.de

### COLLECTION OF TOOLS AND SERVICES OFFERED BY de.STAIR 1:

- **DoDAS** – A tool for the fast detection of alternative splicing in large RNA-Seq data sets
- **segemehl** – Versatile NGS aligner that supports various protocols such as mRNA-Seq (split-read mapping) or bisulfite conversions
- **Metilene** and **BAT** – Ultra-fast detection of differentially methylated regions and the joint analysis of transcriptomic and epigenomic data

▼ *Integration of RNA-Seq data with Whole Genome Bisulfite sequencing data. The radar plot shows the enrichment of transcription factor binding sites in differentially methylated regions that are correlated with the expression (cDMRs) of target genes. Circles indicate levels of transcription factor biding site (TFBS) enrichment measured as the percentage of binding sites of a particular transcription factor found in cDMRs relative to all binding sites of this transcription factor found in DMRs. The unexpected positve correlation of enriched TFBS in cDMRs (quadrants 1 and 3) may to the discovery of molecular mechanisms for cancer development (Kretzmer et al., Nature Genetics 2015).*

# THE de.NBI PROJECT de.STAIR 2

## Regulatory RNA and Data

**The de.STAIR 2 project at University of Freiburg deals with specific aspects of the workflow design, its technical integration and pre-processing. de.STAIR 2 offers services for the analysis of complex RNA-Seq datasets, the development, benchmarking and accommodation of novel analysis methods with the Galaxy and RBC workbench, and the integration of results into the regulatory context.**

▼ *The de.STAIR 2 team at Freiburg University.*

**Wolfgang R. Hess**
*Project leader*
wolfgang.hess@biologie.uni-freiburg.de

**Steffen Lott**
steffen.lott@biologie.uni-freiburg.de

### COLLECTION OF TOOLS AND SERVICES OFFERED BY de.STAIR 2:

- **GLASSgo** – GLASSgo stands for Global Automatic SRNA Search go. It takes only one sRNA into account and returns possible homologous sequences.
- **CopraRNA** – By using a set of sRNA homologs (coming, e.g., from GLASSgo), CopraRNA tries to predict the corresponding targets within a given genome.
- **CoVennTree** – With high dimensional datasets coming from metatranscriptomic experiments, traditional visualization tools cannot handle these in a proper way. CoVennTree fills this gap and allows a comparing up to three data sets at the same time.
- **Galaxy workflow** – Preprocessing of metatranscriptomic data
- **Galaxy workflow** – Advanced analysis and visualization of metatranscriptomic data
- **Galaxy workflow** – sRNA analysis

▼ *With **de.STAIR 2** specific expertise in the work with regulatory RNA is offered. We combine wet-lab results with theoretical approaches and vice versa. This combination allows far-reaching assumptions as well as specific outcomes. The project provides experience with customized tool development and the visualization of complex datasets (e.g., GLASSgo and CoVennTree).*

**REGULATORY RNA - METATRANSCRIPTOMICS - METAGENOMICS**



de.STAIR 2
- biological expertise (RNA)
- wet and dry lab equipment
- experimental bioinformatics

# THE de.NBI PROJECT de.STAIR 3

## Customized Workflow Development and Data Integration Concepts for RNA-Sequencing

**The de.STAIR 3 project at University of Rostock offers services for the analysis of RNA sequencing data and is specialized in tailor-made workflows and training for their use.**

▼ **The de.STAIR 3 team at Rostock University.** *Contact: www.sbi.uni-rostock.de*

**Olaf Wolkenhauer**
*Project leader*
olaf.wolkenhauer@uni-rostock.de

**Andrea Bagnacani**
andrea.bagnacani@uni-rostock.de

**Markus Wolfien**
markus.wolfien@uni-rostock.de

### COLLECTION OF TOOLS AND SERVICES OFFERED BY de.STAIR 3:

- Guidance for the design of RNA-sequencing experiments
- Workflow design and technical integration into the RBC workbench
- Providing alternative workflow solutions, like Docker containers
- Evaluation of preprocessing modules for sequencing data analysis
- Facilitating the annotation and characterization of transcripts
- Data retrieval interfaces for knowledge-base integration in Galaxy
- Training courses for structural data integration and workflow development
- Deployment of interactive Galaxy tours for our developed workflows
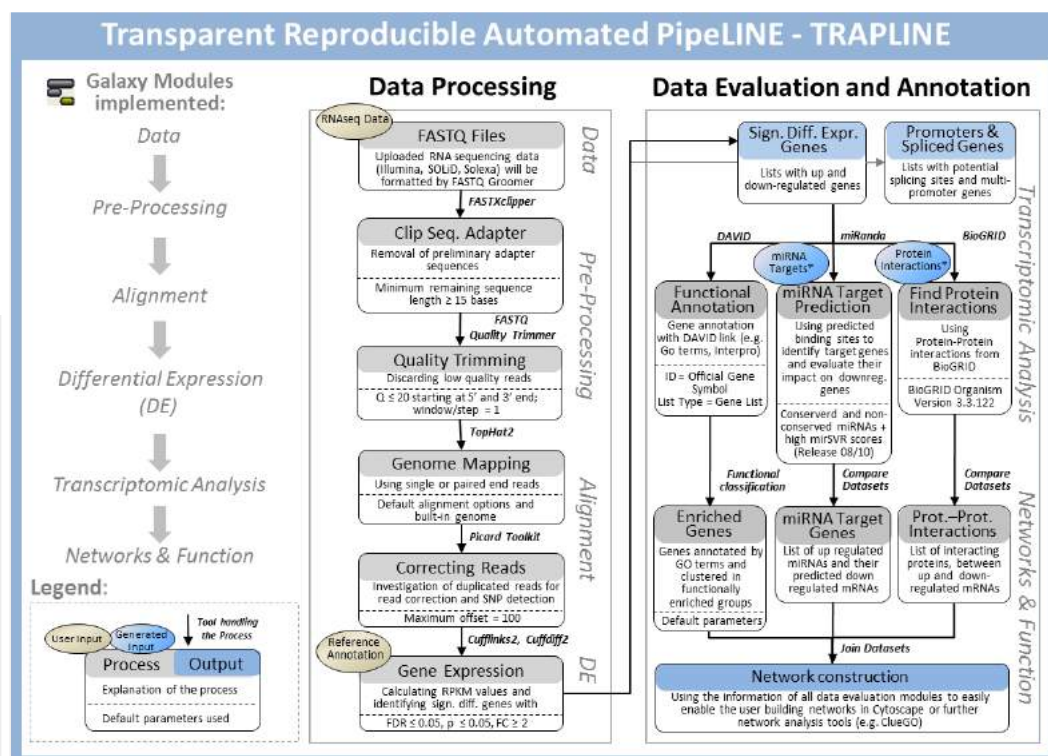


◄ *The TRAPLINE workflow (Wolfien et al.,2016).*

# THE de.NBI PROJECT de.NBI-epi 3

## Epigenomics Data Integration and Analysis

**The de.NBI-epi 3 project at Max Delbrück Center for Molecular Medicine offers tools, services and training for the epigenomics data integration and analysis.**

▼ *The de.NBI-epi 3 team at MDC Berlin.*

**Altuna Akalin**
*Project leader*
altuna.akalin@mdc-berlin.de

**Vedran Franke**
vedran.franke@mdc-berlin. de

**COLLECTION OF TOOLS AND SERVICES OFFERED BY de.NBI-epi 3:**

- **methylKit** – Toolkit for analysis of DNA modifications from bisulfite-sequencing data
- **Chromosome segmentation algorithms** – Application of various algorithms for segmentation of epigenomic profiles
- **Genomation** - Epigenomic data integration and visual summaries
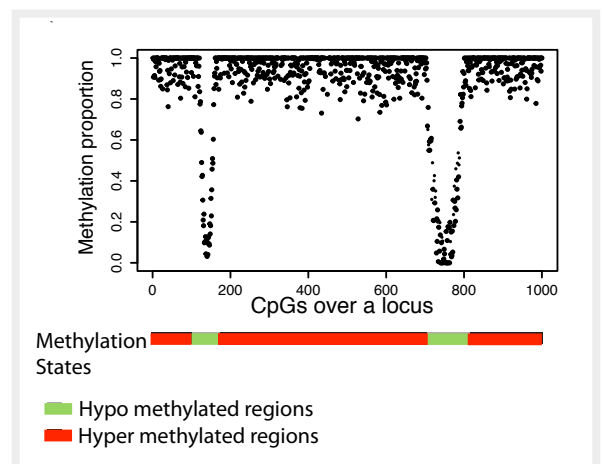


▲ *methylKit segmentation algorithm for methylation.*

▼ *Genomation: Integration of different ChIP-seq datasets using R (Akalin et al., 2015).*
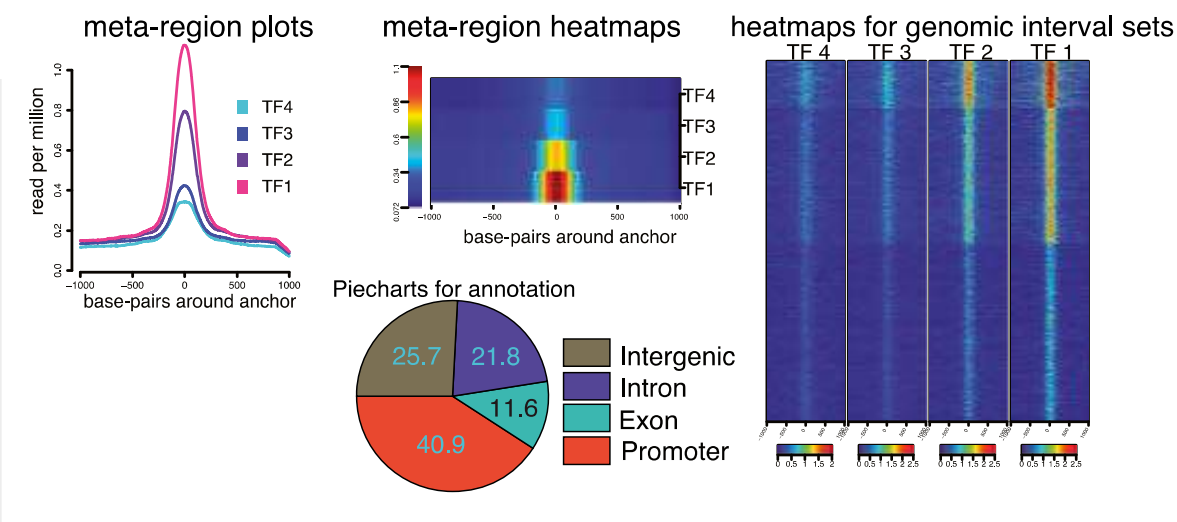
# THE de.NBI PROJECT de.NBI-epi 4

## The Epigenetic Workbench

**The de.NBI-epi 4 project at University of Freiburg offers tools, services and training for the analysis of epigenetic modifications.**

▼ *The de.NBI-epi 4 team at University of Freiburg.*

**Björn Grüning**
*Project leader*
gruening@informatik.uni-freiburg.de

**Joachim Wolff**
wolffj@informatik.uni-freiburg.de

### COLLECTION OF TOOLS AND SERVICES OFFERED BY de.NBI-epi 4:

- Integration, development and maintenance of **deepTools** in collaboration with the MPI Freiburg
- Development, maintenance and services around **Galaxy**
- Development and services around the **bioconda** build system
- Containerization (Docker, rkt) with layer-donning and **BioDocker**
- Maintenance and support for the Galaxy Docker Project and Galaxy flavors
- A tool library for the analysis of bisulfite sequencing data and ChIP-seq
- **Galaxy Workflows** and ready-to-use pipelines including QC, normalization and visualization
- **Galaxy tours** as new technique for scalable and interactive training
- **Training courses** and training material on data analysis of epigenetic modifications in close cooperation with GTN and GOBLET

▼ *Customized, scalable, production-ready workbenches powered by **Galaxy** flavors.*
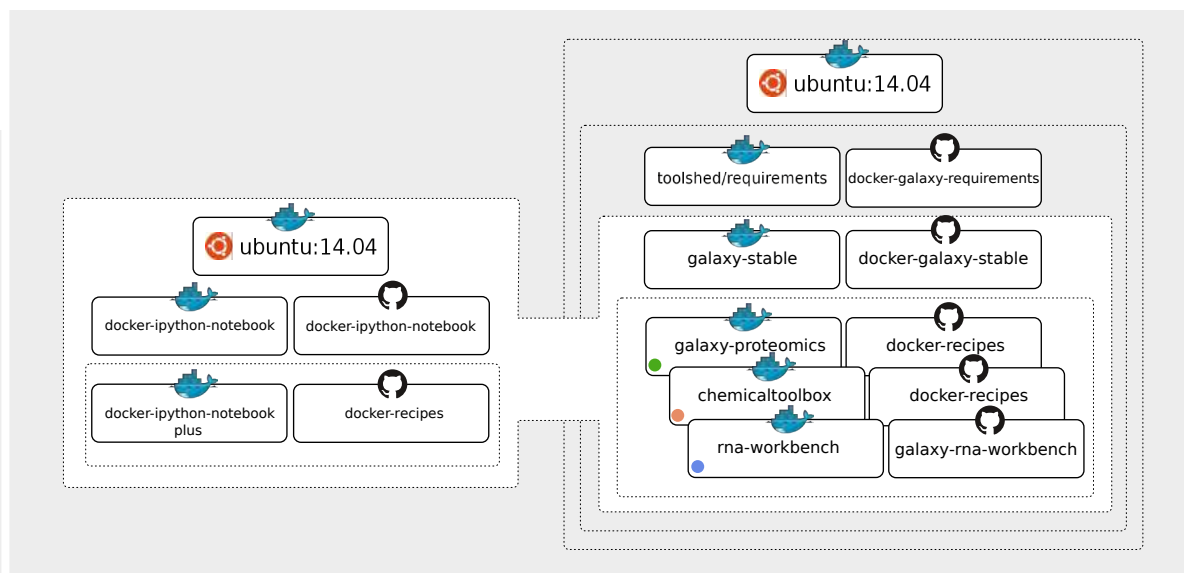
# THE de.NBI SERVICE CENTER GCBN

## German Crop BioGreenformatics Network

**The GCBN Service Center focuses on crops and provides both tailored plant-specific data and infrastructure to the crop plant research community.**

The GCBN Service Center consists of three partners with vast experience in different fields of green bioinformatics. As services to the plant research community, we provide transparent access to germplasm data, improve gene and genome annotation and bridge genotypes with phenotypes. GCBN offers access to various crop and plant specific information systems and tools.

**TOPICS OF THE SERVICE CENTER GCBN:**
- Provision of transparent access to germplasm data
- Improvement of gene and genome annotation for complex crop genomes
- Bridging genotypes with phenotypes

**SELECTION OF TRAINING COURSES CARRIED OUT BY GCBN:**
- Data Management for Plant Genomics & Phenomics: Standards, Tools and Infrastructures
- Big Data Training Course in Plant Genomics

**CURRENT PROJECTS:**
- **GCBN 1** at **IPK Gatersleben**: Evolving versatile levels to access germplasm data using database systems (GBIS/I and EURISCO); development and maintenance of scientific data management system (e!DAL)
- **GCBN 2** at **Helmholtz Zentrum München**: Exploration of complex Triticeae genomes and support with genomic tools and database systems (PlantsDB and PlantsDB Tools); Establishment and maintenance of a structured plant repeat collection (REdat and REcat) as transposon annotation repository
- **GCBN 3** at **Forschungszentrum Jülich**: Integration and access to genomic data in plant database resources (PlaBiPD); visualization concepts aiming for a thorough and well-structured presentation of plant omics data using ontology categories (MapMan)

**Uwe Scholz**
*Coordinator*
*IPK Gatersleben*
scholz@ipk.gatersleben.de

**Klaus Mayer**
*Deputy coordinator*
*Helmholtz Zentrum München*
k.mayer@helmholtz-muenchen.de

▼ *Members of the Service Center GCBN. From left to right: Matthias Lange, Marie Bolger, Thomas Schmutzer, Heidrun Gundlach, Uwe Scholz, Jinbo Chen (November 2016).*

# THE de.NBI PROJECT GCBN 1

## Access Germplasm Data and Phenotypes

**The GCBN 1 project at the Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) Gatersleben offers tools, services and training for accessing plant genetic resources data and to connect genotypes and phenotypes.**

▼ *The GCBN 1 team at the IPK Gatersleben.*

**Uwe Scholz**
*Project leader*
scholz@ipk-gatersleben.de

**Matthias Lange**
lange@ipk-gatersleben.de

### COLLECTION OF TOOLS AND SERVICES OFFERED BY GCBN 1:

- **GBIS** – The information system of the German Genebank. This ex-situ collection is a safeguard of plant biodiversity and in this infrastructure GBIS is an important source of information for plant researchers as well as crop breeders.
- **EURISCO** – The search catalogue providing information about ex-situ plant collections maintained in Europe and hosted at the IPK.
- **LAILAPS** – The life science search engine, which provides an exploration of plant genomic data for forward genetic research. It comprises millions of quarterly updated and indexed information from major life science databases as well as millions of evidence ranked gene annotations to associate these features with plant genome loci.
- **e!DAL** – The lightweight open source software framework for publishing and sharing research data. Using the e!DAL technology, a Plant Genomics and Phenomics Research Data Repository is currently provided at the IPK Gatersleben.
- **IPK BLAST Server** – Enables a homology search against the current barley genome sequences resources as well as related data sets.
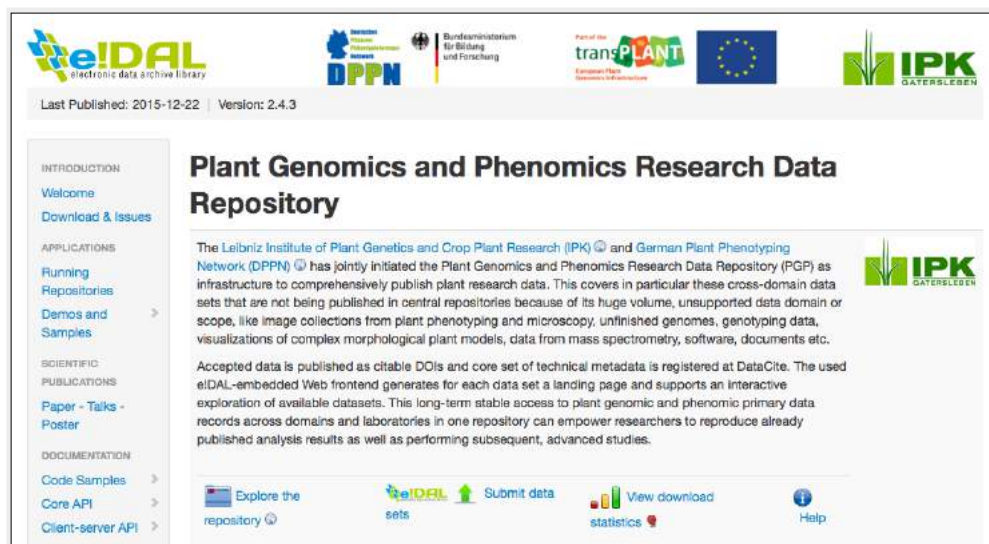


◄ *Screenshot of the Plant Genomics and Phenomics Research Data Repository – PGP, an infrastructure to publish research data (Arend et al., 2016).*

# THE de.NBI PROJECT GCBN 2

## Plant Gene and Genome Annotation

**The GCBN 2 project at the Helmholtz Zentrum München provides infrastructure, services and training for analyzing (complex) plant genomes and accessing genomic data.**

▼ *The GCBN 2 team at the Helmholtz Zentrum München.*

**Klaus Mayer**
*Project leader*
k.mayer@helmholtz-muenchen.de

**Heidrun Gundlach**
h.gundlach@helmholtz-muenchen.de

### COLLECTION OF TOOLS AND SERVICES OFFERED BY GCBN 2:

- **PlantsDB** and **PlantsDB Tools** – Plant genome Information resource structuring and and communicating complex datasets, with a focus on large Triticeae genomes like barley, wheat and rye; assists comparative analyses of both model and crop plants.
- **REdat** and **REcat** – A comprehensive and well characterized plant transposon collection structured by taxonomy and a detailed repeat classification catalog (PGSB-REcat).
- Plant genome annotation and analysis infrastructure, including BLAST server, synteny viewer (CrowsNest) or expression browser.
- Plant genome resources registry with more than 300 different databases/entities included by now.
- Extensive user training material including video tutorials and hands-on training.

▼ *Screenshots of the **CrowsNest tool** (part of PGSB PlantsDB) to visualize syntenic relationships between plant genomes. The syntenic regions between the organisms (Brachypodium and Sorghum in this example) can be browsed in a hierarchical way from macro-synteny **(A)** over chromosomes **(B)** down to micro-synteny views **(C)**.*
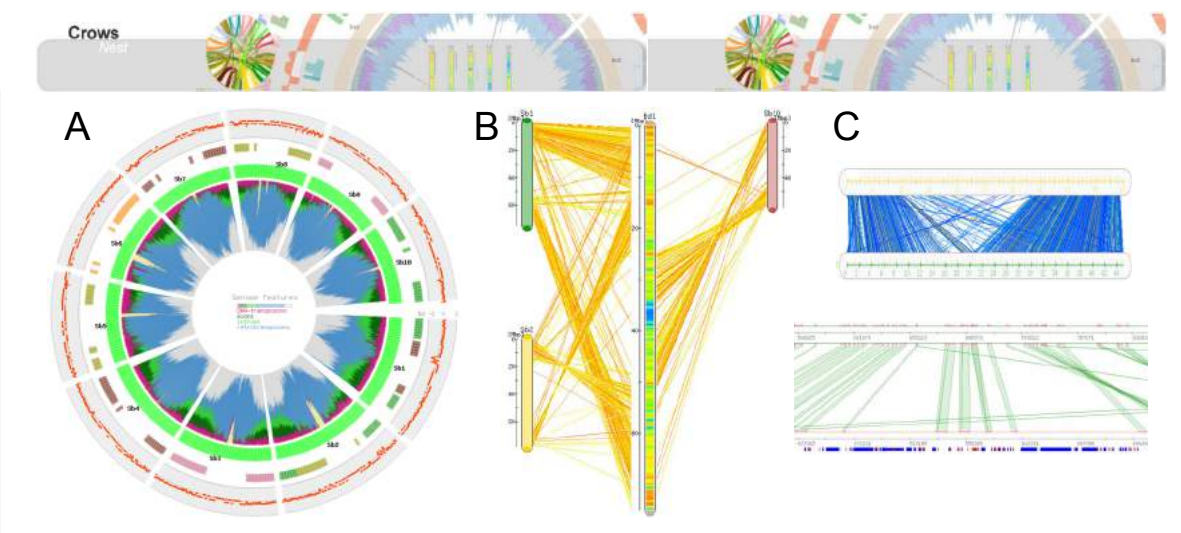
# THE de.NBI PROJECT GCBN 3

## Bridging Genotypes with Phenotypes

**The GCBN 3 project at the Forschungszentrum Jülich provides infrastructure, services and training for functional plant genome annotation and phenotypic integration.**

▼ *The GCBN 3 team at the Forschungszentrum Jülich.*

**Björn Usadel**
*Project leader*
b.usadel@fz-juelich.de

**Marie Bolger**
m.bolger@fz-juelich.de

### COLLECTION OF TOOLS AND SERVICES OFFERED BY GCBN 3:

- **PlabiPD** – Analysis, visualization and integration of plant genomes, omics data and whole plant phenotypes. PlabiPD functions as portal for tools and services.
- **Mercator** – Automatic gene functional annotation using a dedicated high performance computer cluster.
- Access to hand curated reference plant functional annotations.
- A selection of downloadable software tools and suites for NGS data preprocessing (**Trimmomatic**), analysis (**RobiNA**) and visualization (**MapMan**).
- Extensive user hands-on training including downloadable virtual machines with all course material.
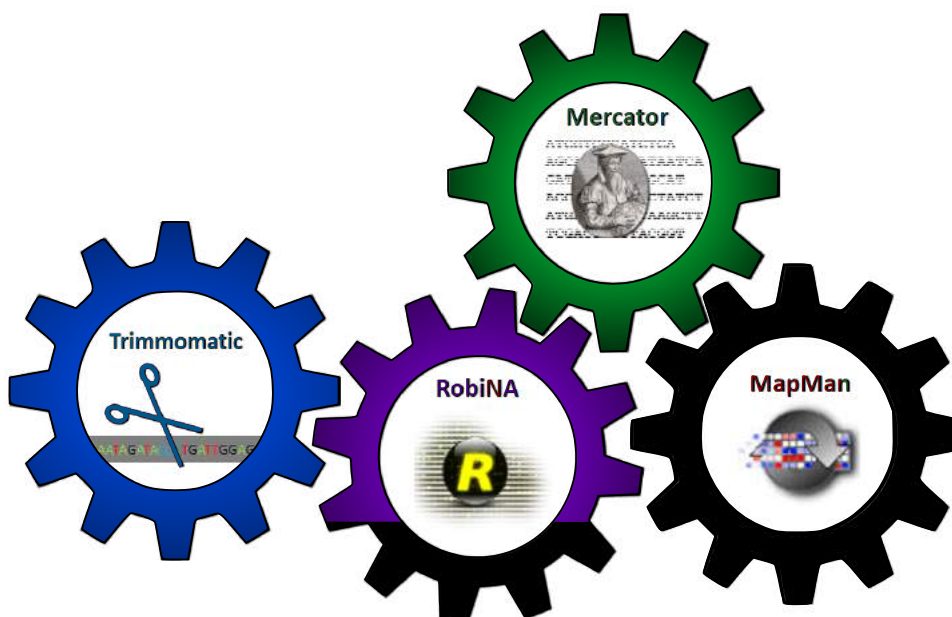


◄ *Trimmomatic workflow overview* of tools provided by GCBN 3 for NGS data from raw fastq files processing using to data visualization using MapMan.

# THE de.NBI SERVICE CENTER BioData

## Center for Biological Data

**The BioData Service Center, consisting of the data services SILVA, PANGAEA, Bac*Dive*, BRENDA and EnzymeStructures, provides highly curated reference datasets for academia and industry.**

The BioData Service Center facilitates access to and services for ribosomal RNA genes (SILVA), georeferenced data from earth system research (PANGAEA), detailed strain-linked information on the different aspects of bacterial and archaeal biodiversity (Bac*Dive*), comprehensive information on all aspects of enzyme functions (BRENDA), as well as easily accessible protein structure data (EnzymeStructures).

**TOPICS OF THE SERVICE CENTER BioData:**
- Data products: Reference databases for taxonomy, phylogeny, biotechnology, biochemistry, pharmacy, medicine, quality control, diagnostics and environmental research
- Research and service: Enzyme structure and function, taxonomy of cultivated/uncultivated strains, data mobilization and presentation

**SELECTION OF TRAINING COURSES CARRIED OUT BY BioData:**
- Training in BRENDA/Bac*Dive* modules
- ARB/SILVA and Bac*Dive*: Training on microbial phylogeny and diversity analysis

**CURRENT PROJECTS:**
- **BioData 1** at **Jacobs University Bremen**: Quality-controlled databases of aligned ribosomal RNA gene sequences (SILVA)
- **BioData 2** at **University Bremen**: A globally leading information system archive and data publisher for geoscientific, biological and environmental data (PANGAEA)
- **BioData 3** at **Leibniz Institute DSMZ Braunschweig**: Strain-linked information about bacterial and archaeal biodiversity (Bac*Dive*)
- **BioData 4** at **TU Braunschweig**: Comprehensive information on all aspects of enzyme function and property data (BRENDA)
- **BioData 5** at **Universität Hamburg**: Easy accessible and usable protein structures for life science research and applications (EnzymeStructures)

▼ *Members of the Service Center BioData. From left to right: Stefan Bietz, Matthias Rarey, Dietmar Schomburg, Alan Beccati, Anna Vetcininova, Lisa Jeske, Frank Oliver Glöckner, Sandra Placzek, Jan Gerken, Janine Felden, Christian Quast, Lorenz Reimer (November 2016).*

**Frank Oliver Glöckner**
*Coordinator*
*Jacobs University Bremen*
f.gloeckner@jacobs-university.de

**Michael Diepenbroek**
*Deputy coordinator*
*University of Bremen*
mdiepenbroek@pangaea.de

# THE de.NBI PROJECT BioData 1

## SILVA – High Quality Ribosomal RNA gene Databases, Tools and Services

**The BioData 1** project at Jacobs University Bremen and Max Planck Institute for Marine Microbiology offers with **SILVA** a comprehensive web resource for up to date, quality-controlled databases of aligned ribosomal RNA (rRNA) gene sequences from the *Bacteria, Archaea* and *Eukaryota* domains and supplementary online services.

▼ *The BioData 1 team at Jacobs University Bremen.*

**Frank Oliver Glöckner**
*Project leader*
f.gloeckner@jacobs-university.de

▲ *From left to right: Gerrit Bruns, Andreea Ciuprina, Jan Gerken, Pelin Yilmaz, Christian Quast, Jörg Peplies, Frank Oliver Glöckner.*

### COLLECTION OF TOOLS AND SERVICES OFFERED BY BioData 1/SILVA:

- **Taxonomic Browser** – Explore the SILVA content by taxonomy.
- **Advanced Search** – Find the sequences you need for your research or application.
- **Sequence Search** – Extract related sequences from the SILVA databases.
- **Aligner** – Align and classify your sequences against the SILVA references.
- **TestProbe** – Evaluate the quality of your probes.
- **TestPrime** – Evaluate the quality and coverage of your primer pairs.
- **The All-Species Living Tree** – Get the best type strains for your taxonomy.
- **SILVAngs** – The user-friendly, web-based data analysis service for rDNA amplicon reads from Next-Generation Sequencing (NGS). It classifies your rDNA reads and provides a wealth of results for download and further downstream processing.

▼ *SILVA´s data products and applications.*
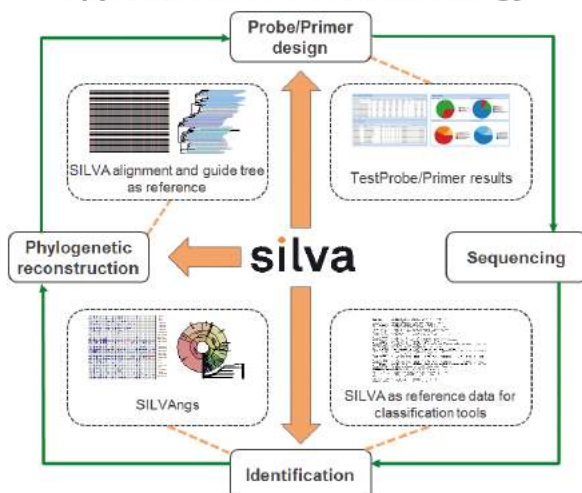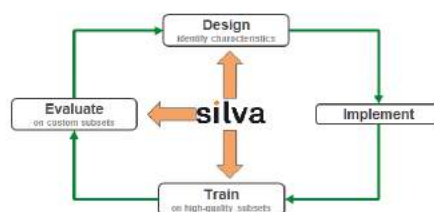


**SILVA Release 128 - Statistics**

| | SSU | | LSU | |
|---|---|---|---|---|
| | Parc | Ref NR 99 | Parc | Ref |
| Archaea | 200,557 | 24,315 | 1528 | 1271 |
| Bacteria | 4,660,106 | 552,377 | 176,194 | 130,965 |
| Eukaryota | 460,766 | 68,996 | 557,769 | 22,105 |
| Total | 5,616,941 | 645,151 | 735,238 | 154,297 |

# THE de.NBI PROJECT BioData 2

## PANGAEA – The Data Publisher for Earth and Environmental Science

**The BioData 2** project offers with PANGAEA a globally leading information system, long term archive, and data publisher for geoscientific, biological, and environmental data. PANGAEA is hosted by the MARUM-Center for Marine Environmental Sciences (University Bremen) and the Alfred Wegener Institute – Helmholtz Centre for Polar and Marine Research.

▼ *The BioData 2 team at MARUM, University Bremen.*

**Michael Diepenbroek**
*Project leader*
mdiepenbroek@pangaea.de

**Janine Felden**
jfelden@marum.de

**Melanie Buß**
mbuss@marum.de

### COLLECTION OF TOOLS AND SERVICES OFFERED BY BioData 2 / PANGAEA:

- **Project data management** – Let PANGAEA manage your project data. Experienced data curators supported more than 200 national and international multidisciplinary projects.
- **Long term archive** – PANGAEA accepts data from individual scientists, institutes, and science projects. Most data are in open access and enable the scientific community to share and reuse them.
- **Data publication** – Get your data published and citable (including assignment of DOIs) in open access in order to increase the impact of your work.
- **Data infrastructures** – Share your standard compliant data and metadata with global data infrastructures such as Global Earth Observation System of Systems (GEOSS) and Global Biodiversity Information Facility (GBIF).
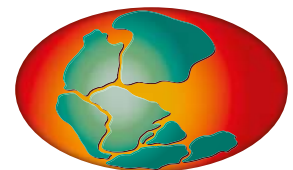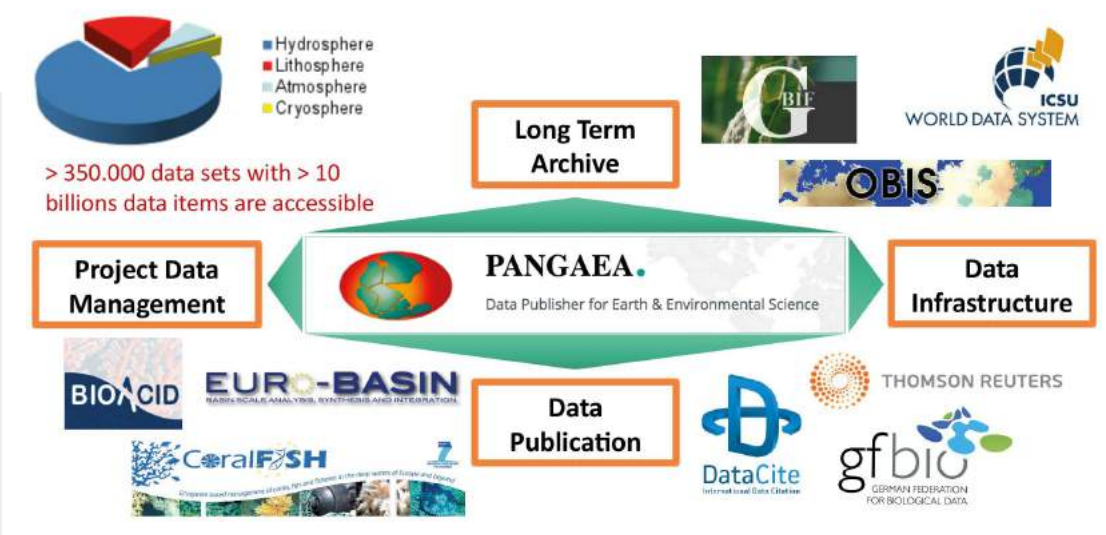


◄ *Services offered by PANGAEA.*

# THE de.NBI PROJECT BioData 3

## Bac*Dive* – The Bacterial Diversity Meta-Database

**The BioData 3 project at the Leibniz Institute DSMZ in Braunschweig offers with Bac*Dive* a database providing strain-linked information about bacterial and archaeal biodiversity for 54,610 strains.**

▼ *The BioData 3 team at Leibniz Institute DSMZ.*

**Jörg Overmann**
*Project leader*
joerg.overmann@dsmz.de

**Anna Vetcininova**
anna.vetcininova@dsmz.de

**Adam Podstawka**
adam.podstawka@dsmz.de

**Lorenz Reimer**
lorenz.reimer@dsmz.de

**Carola Söhngen**
carola.soehngen@dsmz.de

### COLLECTION OF TOOLS AND SERVICES OFFERED BY BioData 3/Bac*Dive*:

- Data encompass the fields of taxonomy, morphology, physiology, sampling and environmental conditions as well as molecular biology.
- Majority of the data are manually annotated and curated.
- Data sources are primary literature, internal descriptions of culture collections and expert compiled compendia.
- Continuously growing data content (overall increase in 2016: 23%).
- Implementiation of new data fields and features (e.g. the taxonomy browser TAXplorer).
- Data accession via GUI (bacdive.dsmz.de) or via the RESTful web service (bacdive.dsmz.de/api/).

| BacDive Section | Release 09/2016 | Increase 2015-2016 |
|---|---|---|
| Name and taxonomic classification | 108,226 | 4% |
| Morphology and physiology | 54,129 | 28% |
| Culture and growth conditions | 83,496 | 71% |
| Isolation, sampling and environmental information | 23,077 | 3% |
| Application and interaction | 15,873 | 1% |
| Molecular biology | 20,029 | 13% |
| Strain availability | 68,798 | 3% |
| Overall entries | 373,628 | 23% |

◀ **Bac*Dive*** *release 2016 – Statistics.*

# THE de.NBI PROJECT BioData 4

## BRENDA – World´s largest Enzyme Information Database

**BRENDA is the world's largest and most widely used enzyme information system, providing comprehensive information on all aspects of enzyme function and properties. Used by up to 100,000 users per month from biotechnology, biochemistry, pharmacy, medicine, and systems biology it has evolved into an indispensable member of the most important websites in the life sciences.**

▼ *The BioData 4 team at Technische Universität Braunschweig.*



**Dietmar Schomburg**
*Project leader*
d.schomburg@tu-bs.de
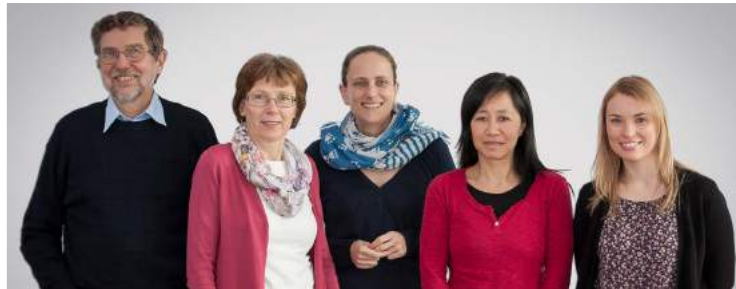
▲ *From left to right: Dietmar Schomburg, Ida Schomburg, Sandra Placzek, Antje Chang, Lisa Jeske.*

### COLLECTION OF TOOLS AND SERVICES OFFERED BY BioData 4/BRENDA:

- As the enzyme data are complex, including text, numeric, graphical and structural data, a complex query tool engine has been developed to allow the users an **efficient access to the data**: This includes a „Quick Search", text-based, full text, and combinatorial queries, structure-based queries e.g. for chemical substructures and metabolic pathways, and explorers for Enzyme Classification, Taxonomy, Protein Folding and various ontologies.
- **Visualization tools** have been developed including the visualization of Word Maps, the Genome Explorer, Functional Parameter statistics, Metabolic Pathways, distribution of sequence patterns, and membrane helices within the enzyme classes.
- The **website** also includes supporting tools like a complex enzyme function prediction tool for genome sequences, as well as the BRENDA Tissue Ontology and a non-redundant biochemical reactions database (BKM).
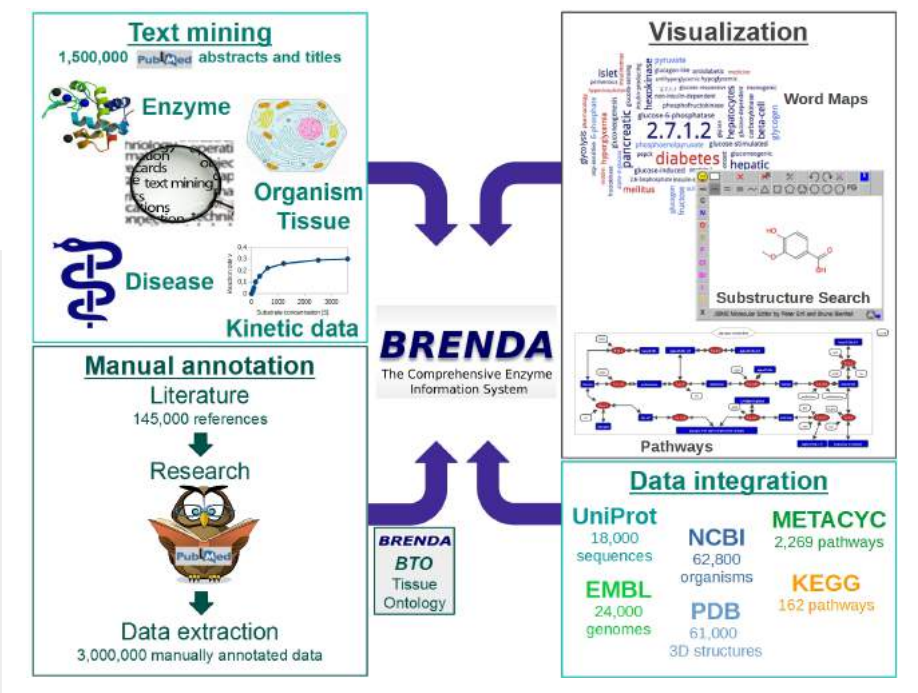


◀ *BRENDA content, tools and integrated databases.*

# THE de.NBI PROJECT BioData 5

## EnzymeStructures – Employing Protein Structures for Biochemical Research

**The BioData 5 project at Universität Hamburg makes protein structure data easily accessible and usable for life science research. The focus is on active sites and molecular interactions to small molecules like substrates, cofactors and inhibitors.**

▼ *The BioData 5 team at Universität Hamburg.*

**Matthias Rarey**
*Project leader*
rarey@zbh.uni-hamburg.de

**Stefan Bietz**
bietz@zbh.uni-hamburg.de

**COLLECTION OF TOOLS AND SERVICES OFFERED BY BioData 5/EnzymeStructures:**

- **Protoss** – Hydrogen assignment and placement, determination of ligand's tautomeric and protonation state
- **DoGSiteScorer** – Active site detection and druggability assessment
- **PoseView** – 2D schematic visualization of protein-ligand complexes and interactions
- **SIENA** – Searching and aligning active sites across the Protein Data Bank
- **PPI** – Distinguishing permanent from transient protein complexes
- **TrixP** – Binding site similarity search based on physico-chemical features, available in the near future
- **iRAISE** – Inverse screening for the detection of proteins binding a specific small molecule, available in the near future

▼ *Fully automatic binding site search and alignment created by SIENA (Protein: Aldose Reductase).*
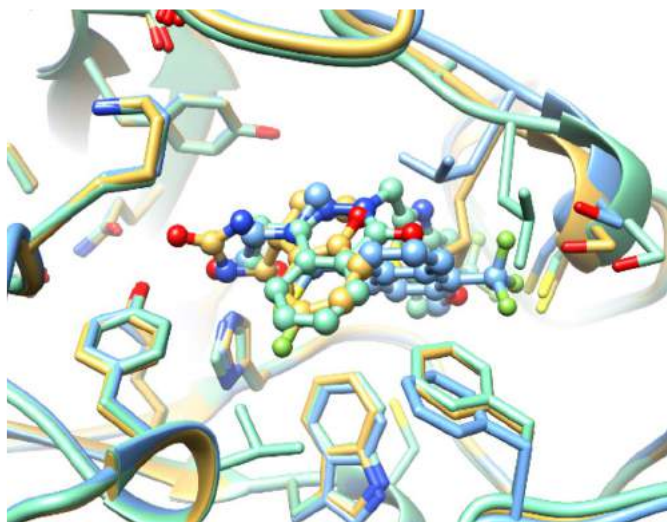
▼ *Binding site detection and characterization using DoGSite-Scorer (Protein: Urokinase-Type Plasminogen Activator).*
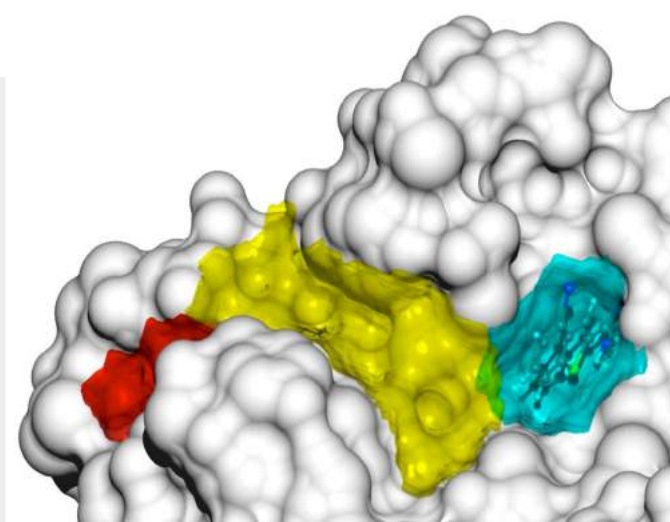
# THE de.NBI SERVICE CENTER de.NBI-SysBio

## de.NBI Systems Biology Service Center

**The de.NBI-SysBio Service Center offers support for data and model management, curated kinetic data, as well as two systems biology modelling tools. Its partners are strongly involved in related standardisation activities, as well as teaching and outreach activities.**

Via its FAIR (Findable Accessible Interoperable Reusable) data and model management tools integrated into SEEK and FAIRDOM Hub; with SABIO-RK, its data source for reaction kinetics data, and with the best-of-breed modelling tools COPASI and CellNetAnalyzer, the center provides tools, teaching and services that cover the whole systems biology cycle.

**TOPICS OF THE SERVICE CENTER de.NBI-SysBio:**

- Standardisation and reproducibility of simulation studies
- Methods and tools for modeling in systems biology
- Model management tools and strategies

**SELECTION OF TRAINING COURSES CARRIED OUT BY de.NBI-SysBio:**

- Workshop on Citable and Reproducible Data and Models
- de.NBI/FAIRDOM Workshop „Kinetics on the Move" about SEEK and SABIO-RK
- Participation in the annual standardisation meetings COMBINE and HARMONY
- Modeling workshops are in preparation

**CURRENT PROJECTS:**

- **de.NBI-SysBio 1** at **HITS Heidelberg**: Self-curated data and model management for systems biology projects; expert curated kinetic data via SABIO-RK
- **de.NBI-SysBio 2** at **University of Rostock**: Standards and tools for model management systems
- **NBI-ModSim 1** at **Heidelberg University** & **NBI-ModSim 2** at **MPI Magdeburg**: Mathematical modeling of biological networks via COPASI and CellNetAnalyzer

▼ *de.NBI SysBio tools and standards in the Systems Biology Cycle.*
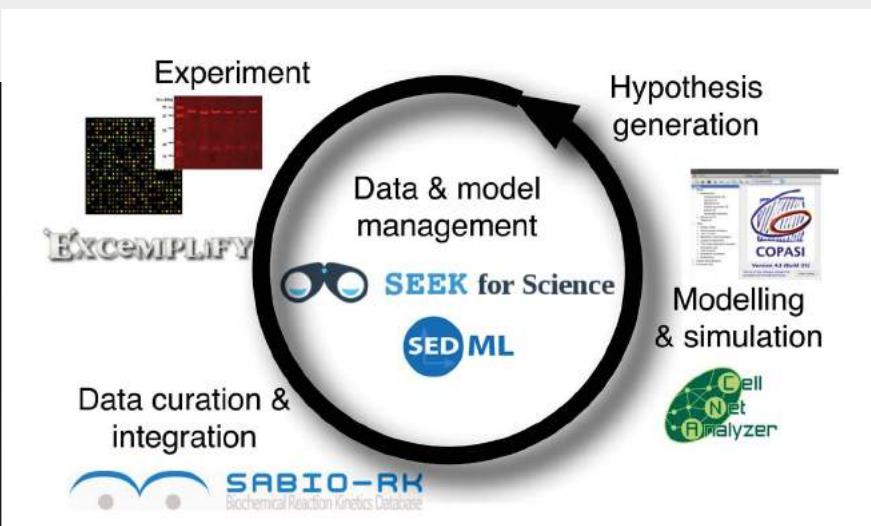
**Wolfgang Müller**
*Coordinator*
*HITS Heidelberg*
wolfgang.mueller@h-its.de

**Ursula Kummer**
*Deputy coordinator*
*Heidelberg University*
ursula.kummer@bioquant.uni-heidelberg.de

# THE de.NBI PROJECT de.NBI-SysBio 1
## FAIR Data and Models Management via (Self-)Curated Data

**The de.NBI-SysBio 1** project at the Heidelberg Institute for Theoretical Studies offers data and model management via self-curated data in the SEEK data management system, as well as the expert-curated high-quality data source for reaction kinetics data, SABIO-RK.

▼ *The de.NBI-SysBio 1 team at Heidelberg Institute for Theoretical Studies (HITS).*



**Wolfgang Müller**
*Project leader*
wolfgang.mueller@h-its.de

▲ *From left to right: Andreas Weidemann, Dagmar Waltemath, Maja Rey, Renate Kania, Wolfgang Müller, Ron Henkel.*

### COLLECTION OF TOOLS AND SERVICES OFFERED BY de.NBI-SysBio 1:

- **SEEK4Science, Rightfield, FAIRDOM Hub** – Codevelopers of SEEK data management system, host of FAIRDOM Hub, part of FAIRDOM
- **SABIO-RK** – High quality expert-curated reaction kinetics data for modelers and experimentalists; curation prioritized via user feedback
- **Excemplify** – Excel-based support of experimental workflows and metadata collection for immunoblot and QPCR in the wet lab
- **Excemplify Antibody Database** – Manage antibodies in your lab.
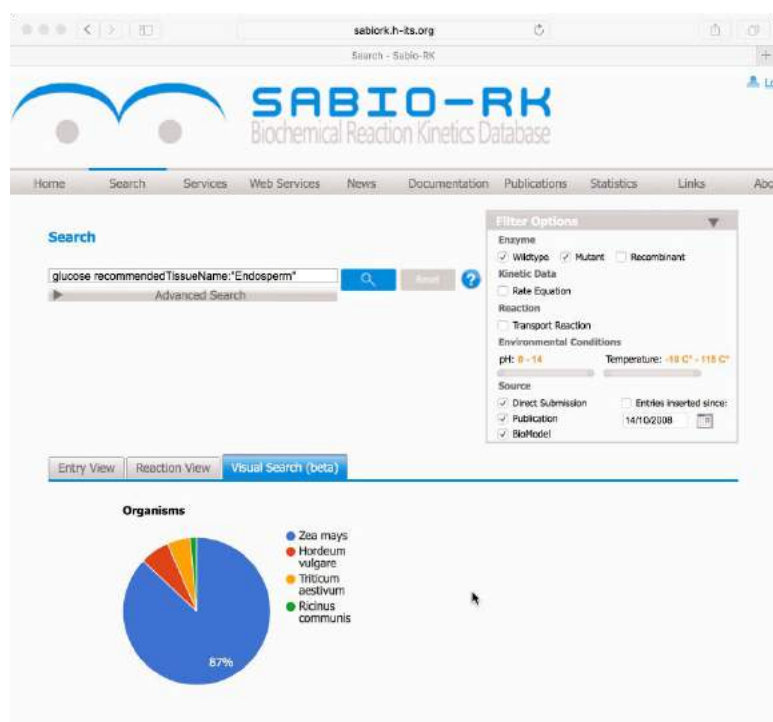


◄ *The SABIO-RK biochemical reaction kinetics database.*

# THE de.NBI PROJECT de.NBI-SysBio 2

## Standards and Tools for Model Management

**The de.NBI-SysBio 2** project at the University of Rostock develops standards, tools and algorithms for sophisticated model management. We furthermore offer training for COMBINE standards and tools that generate reproducible simulation studies.

▼ *The de.NBI-SysBio 2 team at the University of Rostock.*

**Dagmar Waltemath**
*Project leader*
dagmar.waltemath@uni-rostock.de

**Ron Henkel**
ron.henkel@uni-rostock.de

### COLLECTION OF TOOLS AND SERVICES OFFERED de.NBI-SysBio 2:

- **COMBINE** – Standards development for modeling in biology, including SBML
- **MASYMOS** – Graph database to integrate data related to simulation studies
- **MORRE** – Model retrieval engine to search for and compare models
- **BiVeS** – Tool for difference detection in model versions
- **COMBINE Archive Toolkit** – Suite of tools and libraries to generate, edit, and share COMBINE archives
- **SED-ML Database** – Repository of reproducible simulation studies

▼ *SED-ML Database: Launching reproducible simulation studies directly from the JWS Online Model Repository (Fig.: Martin Peters, SEMS).*
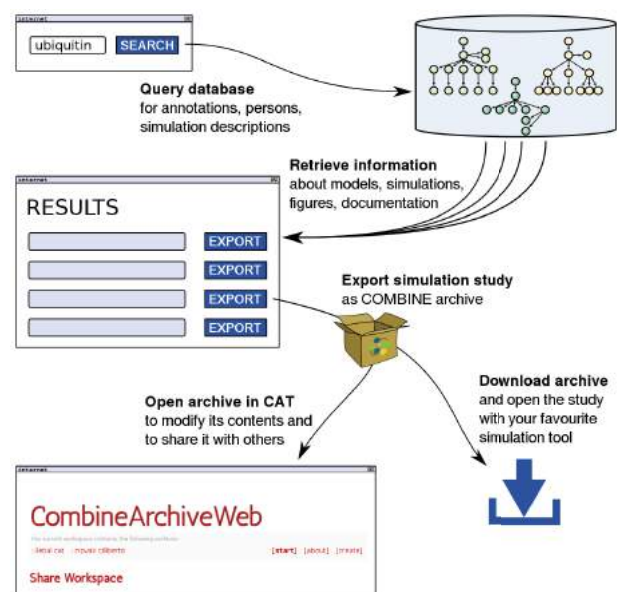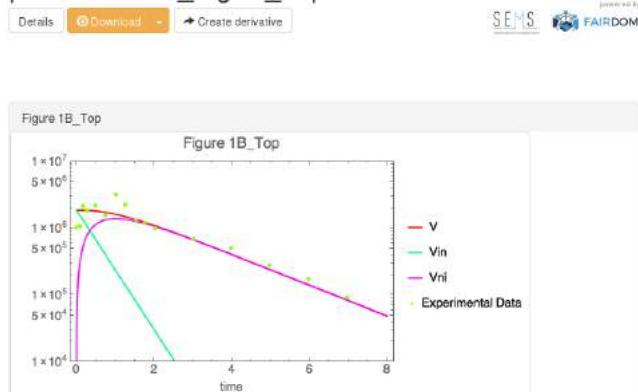




▲ *M2CAT: Retrieving reproducible simulation studies as COMBINE Archives (Fig.: Martin Scharm & Martin Peters, SEMS).*

# THE de.NBI PROJECT NBI-ModSim 1
## Kinetic, Dynamical Modeling and Analysis of Biological Processes

The **NBI-ModSim 1** project at Heidelberg University offers tools, services, and training for dynamical, quantitative modeling of biochemical networks, including e.g. modeling based on ODEs or stochastic formalism, steady and dynamical state analysis, parameter fitting, and sensitivity calculations. Extended user support and training courses will be offered for the COPASI software package.

▼ *The NBI-ModSim 1 team at Heidelberg University.*

**Ursula Kummer**
*Project leader*
ursula.kummer@bioquant.uni-heidelberg.de

**Frank Bergmann**
*Project leader*
frank.bergmann@bioquant.uni-heidelberg.de

### NBI-ModSim 1 OFFERS THE *COPASI* SOFTWARE, WHICH PROVIDES THE USERS WITH:

● Dynamical, kinetic modeling based on ODEs or stochastic, discrete-particle-based formalisms
● Steady state analysis, Metabolic Control Analysis, general sensitivity analysis, complexity reduction, parameter fitting, optimisation, computation of nonlinear dynamics features
● Additionally, the software is SBML compliant open source and offers APIs in different languages.

▼ *Model analysis with **COPASI.***
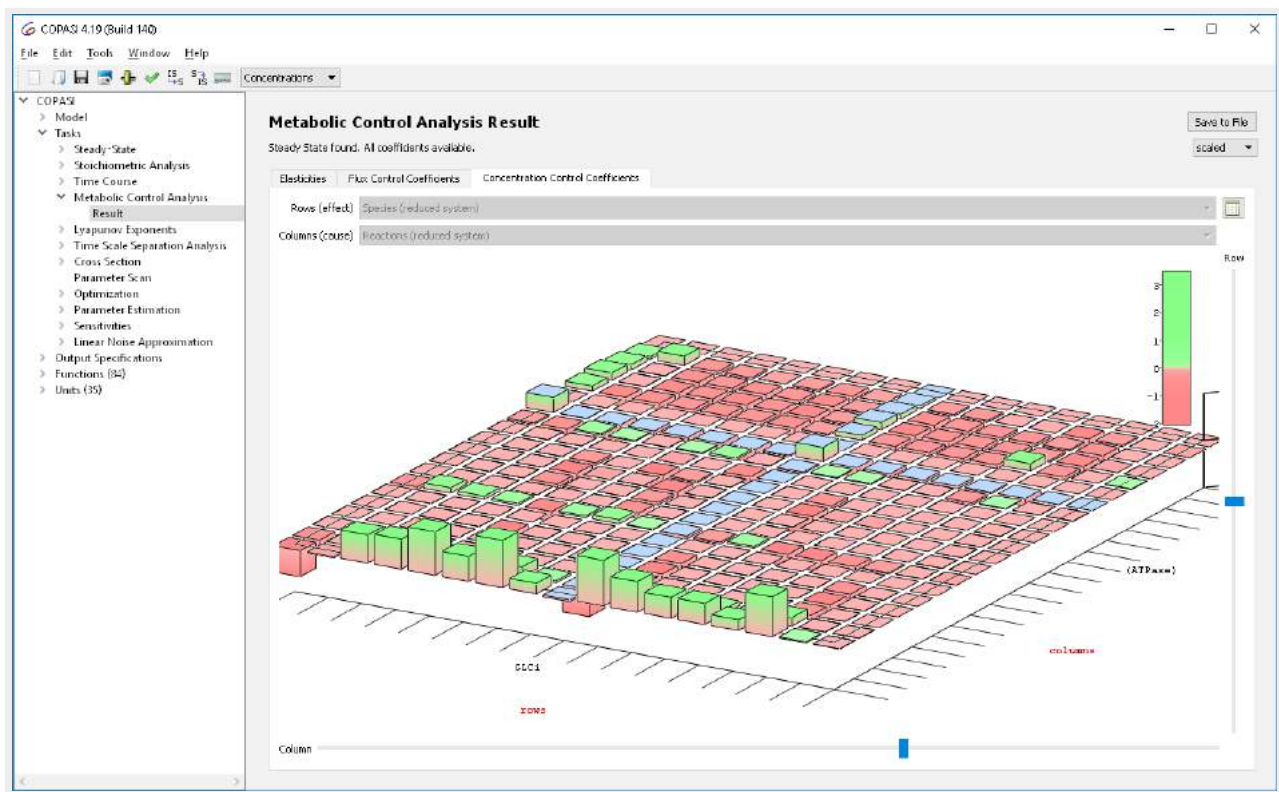
# THE de.NBI PROJECT NBI-ModSim 2
## Structural and Functional Analysis of Biological Networks

**The NBI-ModSim 2** project at the Max Planck Institute for Dynamics of Complex Technical Systems Magdeburg offers tools, services, and training for topological, stoichiometric, constraint-based, qualitative (logical) and semi-quantitative modeling of cellular networks. An extended user support and training courses for the *CellNetAnalyzer* software is provided.

▼ *The NBI-ModSim 2 team at MPI Magdeburg.*



**Steffen Klamt**
*Project leader*
klamt@mpi-magdeburg.mpg.de

**Sven Thiele**
thieles@mpi-magdeburg.mpg.de

### NBI-ModSim 2 OFFERS THE *CellNetAnalyzer* SOFTWARE, WHICH PROVIDES USERS WITH:

- Stoichiometric and constraint-based modeling of genome-scale metabolic networks (metabolic flux analysis, flux optimization and flux balance analysis, elementary modes, targeted modification and metabolic engineering (strain design), etc.
- Topological, logical (Boolean) and semi-quantitative modeling of signaling and regulatory networks (interaction graphs, feedback loops, Boolean dynamics, logical steady states and qualitative input/output behavior, intervention strategies, logic-based ODEs, data-driven inference, etc.)
- The software can be used via GUI or command line and supports the model standard SBML.
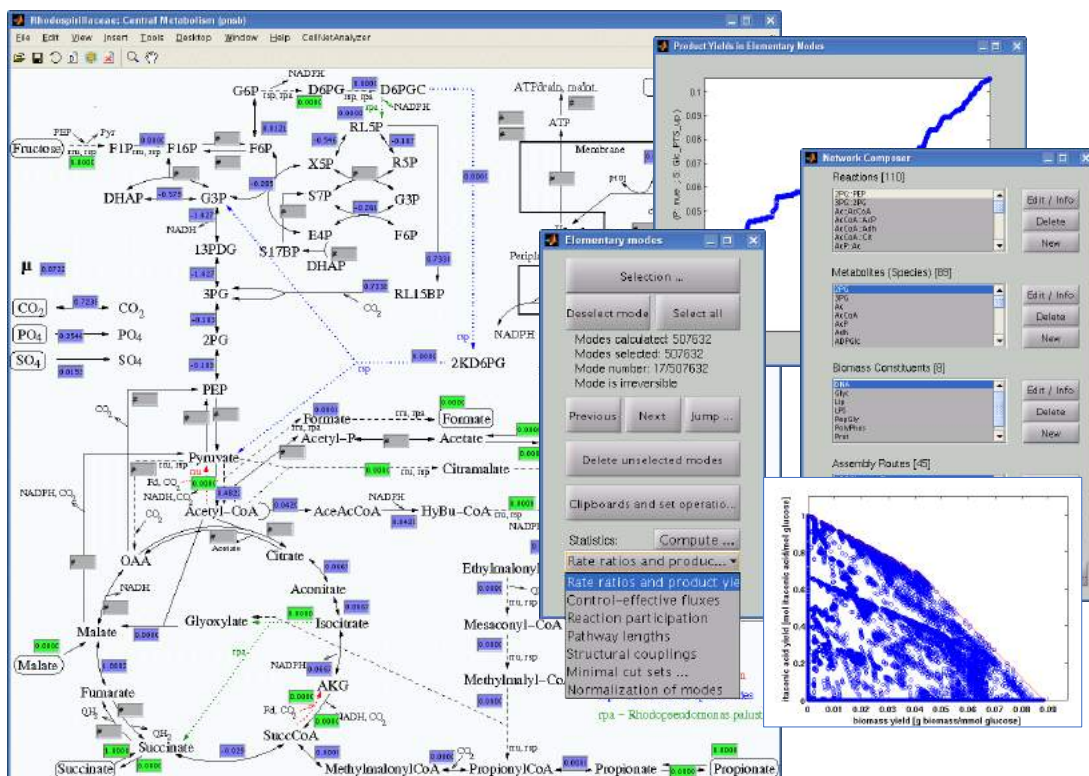


◄ *Metabolic network analysis in CellNetAnalyzer.*

# Contact Data and Imprint

**German Network for Bioinformatics Infrastructure (de.NBI)**
**de.NBI Coordinator and Administration Office**
c/o Center for Biotechnology
Universitätsstraße 27
33615 Bielefeld

**Tel:** +49 (0)521 106 8750
**Fax:** +49 (0)521 106 89046
**E-Mail:** contact@denbi.de
**Homepage:** www. denbi.de
🐦 #denbi@denbiOffice

**Design and Layout:**
LANGEundPFLANZ Werbeagentur GmbH, Speyer