HELENA
Helmholtz Graduate School Environmental Health

GSISH TUM Graduate School of Information Science in Health

Invitation to participate in the 14th Bioinformatics Spring School

Bertinoro, Italy

22.03. - 26.03. 2014

HELENA
Helmholtz Graduate School Environmental Health

GSISH TUM Graduate School of Information Science in Health

Outline of the Course

The 14th edition of the Spring School of Bioinformatics for Molecular Biologists will be held from 22nd through 26th of March 2014. It is organized by the Institute of Bioinformatics and Systems Biology of the Helmholtz Centre München (IBIS). The course is part of the HMGU Ph.D. programme (Graduate School) but also open to other Ph.D. students and post-graduates.

Experiments generate data and from data information is extracted. The experimental design today is –omics driven, generates massive amounts of data and faces a serious challenge of a systemic, rationale, and reproducible approach to analyse and translate raw data into biological information.

Even a simple biological system such as a prokaryotic cell consists of thousands of variables describing the cellular information, their values are conditionally dependent from genetics, regulatory processes and the environment. To answer any question about function and malfunction of biological systems requires not only the observation of data but also their interpretation in the context of any related biological knowledge.

Bioinformatics provides the tools to manage, analyse and evaluate biological data. Systems Biology compiles, parameterizes and quantifies abstract models mostly represented as network graphs. Directed edges represent causal dependencies, but causality linking correlating states is very difficult to catch. As said by Dennis Noble "there is no single level of causality". We face an unexpected complexity; we do not find simple answers to our questions. The clue, why we suffer from diseases and how to cure them is hard to find, and we are often biased by established but misleading assumptions. And sometimes we do not even know if there is any such simple clue. Knowing and understanding the methods, ideas and tools to analyse data for your work will not only help to stay efficient and competitive using bioinformatics tools but also elucidate their power and limitations.

The course will give a fresh look on the data and what can be done in terms of "omics", sequences, systems biology and even the basic foundations of the way we generate, analyse and interpret data. What is the role of algorithms, models, and the biological knowledge? Why do we need network analysis and what are the main tools that we need for generating our next publication?

These days, nobody can follow his own field of interest by reading the publications. We all are aware that life science is now data driven, we face a tsunami of highly complex information from different omics-technologies as well as medical and clinical information. Not only management of the data and running sequence similarity searches are black boxes to many biologists, but the fundamental question how to transform data into knowledge must become part of the strategy of almost any Ph.D. project.

Having an intuitive look at the data cannot perform the exploration of the experiment while ignoring the existing knowledge. Often, "the more the better" in data collection comes first and the nightmare of data analysis follows. The course will try to demystify "computational biology", it will introduce basic principles and advances of bioinformatics and systems biology and, at the same time, foster interdisciplinary ideas and stimulate discussion between experts and students. An open discussion on the world of data we live in and how we will perform our work in the future in distributed teams will be part of the course.

Expert faculty members will introduce the basics in tutorials but also discuss new developments in the interdisciplinary fields of Bioinformatics and Systems Biology. The concept of the course is tailored for Molecular Biologists and aims to support the important dialogue between the experimentalist interested in a specific biological problem and the theoretician eager to solve problems in a generic, reusable way.

The course aims to provide a head start in this important interdisciplinary field by presenting basic concepts as well as recent applications. The intended audience are Ph.D. students and junior scientists. Its main subjects will be the integration of information, biological networks, and concepts in Systems Biology. The course is supported by the Helmholtz Graduate School (HELENA).

Please register as soon as possible since the number of available places is limited (30)!

Please distribute this invitation to all interested Ph.D. students and young scientists. For reservations please contact the IBIS Office (ibis.office@helmholtz-muenchen.de).

H. Werner Mewes

HMGU, Inst. f. Bioinformatics and Systems Biology;
TUM Chair f. Bioinformatics, Center for Life and Food Science, Weihenstephan

Prerequisites for participation:
Participants should have some very basic knowledge in bioinformatics, basic knowledge of genome structure, and be familiar with the very basics of molecular databases. There are no formal requirements.

Fee and registration:
Ca. € 650 (including tuition, course material, room and board for the whole duration of the event, bus transfer from Munich/Neuherberg to Bertinoro). The course fee for HELENA Graduate School students will be covered by the respective Graduate Schools.

Registration:

Secretariat of the Helmholtz-Institute for Bioinformatics and Systems Biology

(ibis.office@helmholtz-muenchen.de).

Location: Bertinoro, Italy (close to Forli, about 40 km from Bologna)

Transport: A bus will be organized from Munich to Bertinoro leaving in the morning of the 22nd of March.

Course Schedule and additional information:

See http://www.helmholtz-muenchen.de/en/mips/events/

DIRECTOR:

Prof. Dr. H.W. Mewes (Helmholtz Zentrum München and TU München)

HELENA
Helmholtz Graduate School Environmental Health

GSISH TUM Graduate School of Information Science in Health

Lecturers and Topics

Lecturer	Affiliation	Topic
M. Campillos	Helmholtz Zentrum München (HMGU)	Systems Biology of Small Molecules
D. Frishman	TUM	Genome Evolution
J. Hoser	Helmholtz Zentrum München	Next Generation Sequencing
L.J. Jensen	Univ. Kopenhagen, Panum Inst.	Large-scale integration of data and text
G. Kastenmüller	HMGU	Metabolomics
J. Krumsiek	HMGU, CMB	Systems Biology and Models
R. Küffner	HMGU	Transcriptional Biological Networks
H.W. Mewes	TUM & HMGU	Systems Biology of Diseases
A. Pfeufer	HMGU	Human Genetics
T. Rattei	Univ. Vienna	Resources in Bioinformatics Environmental Genomes
T. Werner	Consultant & U Mich, Ann Arbor	Translational Transcriptomics

HELENA
Helmholtz Graduate School Environmental Health

GSISH TUM Graduate School of Information Science in Health

Bioinformatics Spring School
22.03 – 26.03. Preliminary Schedule

Saturday, 22. 03.
Arrival Bertinoro

Sunday, 23. 03.

9:00 – 10:00	H.W. Mewes	Introduction to the Course Presentation of Staff Members Presentation of Participants
10:30 – 12:30	T. Rattei	Bioinformatics Resources
13:30 – 15:00	T. Rattei	Microbial Genomes
15:30 – 17:00	L.J. Jensen	Large-scale integration of data and text

Monday, 24. 03.

9:00 – 10:30	T. Werner	Translational Transcriptomics
11:00 – 12:30	G. Kastenmüller	Metabolomics
14:00 – 15:30	J. Hoser	Next Generation Sequencing
15:30 – 17:00	M. Campillos	Systems Biology of Small Molecules
17:00 – 18:30	A. Pfeufer	Genetics and Genomics

Tuesday, 25. 03.

9:00 – 10:30	H.W. Mewes	Epistemology and Personalized Medicine
11:00 – 12:30	R. Küffner	Regulatory Networks
13:30 – 15:00	H.W. Mewes	Systems Biology of Diseases
15:30 – 17:00	D. Frishman	Genome evolution

Wednesday, 26. 03.

09:00 - 11:00	J. Krumsiek	Quantitative Models and Simulation for Systems Biology
11:15 - 12:00	H.W. Mewes	Final Roundtable
13:00	Departure	

HELENA
Helmholtz Graduate School Environmental Health

GSISH TUM Graduate School of Information Science in Health

Faculty members of the 14th Course on Bioinformatics and Systems Biology

HELENA
Helmholtz Graduate School Environmental Health

GSISH TUM Graduate School of Information Science in Health

Prof. Dr. H.W. Mewes,
TU München & Helmholtz Zentrum München

Professor for Bioinformatics at the TU München, Weihenstephan and Head of the Inst. f. Bioinformatics and Systems Biology at the Helmholtz-Zentrum München. Educated as a chemist (Univ. Marburg) he worked in bioenergetics at the Univ. of Heidelberg, in protein chemistry at the EMBL and joined the Max-Planck-Inst. f. Biochemistry in 1985. From 1988 he led the MIPS group active in data collection for protein sequences and genome analysis (yeast, A. thaliana). He has initiated curriculum in bioinformatics started in 2001 as a joint activity of LMU and TUM. He is Co-Founder of Biomax Informatics (1997) and Clueda (2012).

Main interests: genome analysis, human genotype/, disease biology, qualitative models and knowledge representation, philosophy of (life)-science.

Epistemology for Life Sciences:
Science today is a practical, success-oriented industry generating information. The scientific discovery leading to a publication is based on scientific principles one is often not aware and not always the established foundations of sciences are well understood. The recent development of high-dimensional data generation based on omics-technologies marks a transition from individual, hypothesis driven discovery to systemic investigation without fixing any hypothesis at the beginning of the observation. Epistemology is needed for the critical evaluation of our methods applied. The gene/environment concept involves thousands of variables generated by genetic variance, ageing, and everything that we can subsume as “environment”, namely nutrition, life style, drugs, exercise, stress and all other physical, chemical, and mental factors we interact with. How we deal with complex systems comprising trillions of cells that can have more states than there are molecules in the universe, how we can achieve prediction and gain control to cure diseases is very much a methodological challenge. Over the enthusiasm that we can generate large data sets we often forget about the logic of our experiments and the way we meet the challenges rationally and not by trial and error.

Systems Biology of Diseases: Diseases represent a fundamental concept of human health care. A huge classification scheme documented in sophisticated collections such as IDC10 and DSM-V are based on phenotypic evidence but not on biological causes. Only recently rationale insight into disease aitioloy based on genetic and functional molecular information became feasible. This development has substantially changed research concepts, publications and the perspectives for individual treatment in the future. I will discuss how Systems Biology and its applications might help to translate knowledge into health care.

HELENA
Helmholtz Graduate School Environmental Health

GSISH TUM Graduate School of Information Science in Health

Prof. Dmitrij Frishman, TU München

Dmitrij Frishman received a M.S. in Biomedical Electronics from the Saint Petersburg Electrotechnical University in 1984 and a Ph.D. in Biochemistry from the Russian Academy of Sciences in 1991. An Alexander von Humboldt Research Fellowship he received at the end of 1991 allowed him to join the Pat Argos group at the Biocomputing Department of EMBL in Heidelberg, where he pursued postdoctoral training in structural bioinformatics until 1996. He subsequently joined the Munich Information Center for Protein Sequences as a senior scientist and later became Deputy Director of the Institute for Bioinformatics at the German Research Center for Health and Environment. Since 2003, Dmitrij Frishman has been Professor for Bioinformatics at the Technical University of Munich. His current research interests focus on genome annotation, prediction and analysis of protein interactions, and structural genomics

Genome development

Genes generally evolve by duplication and subsequent divergence of sequence. Additional mechanisms for increasing the functional spectrum of proteins include internal duplications, oligomerisation, domain shuffling and swapping, and circular permutations, to name just a few. In my talk I will review the currently known variety of protein folds and the dominant theories of their emergence. I will put these insights into genomic and evolutionary perspective and review the factors that determine the evolutionary status of proteins in the cell and their functional plasticity.

Prof. Lars Juhl Jensen, Panum Inst. Copenhagen

Lars Juhl Jensen started his research career in Søren Brunak's group at the Technical University of Denmark (DTU), from which he in 2002 received the Ph.D. degree in bioinformatics for his work on non-homology based protein function prediction. During this time, he also developed methods for visualization of microbial genomes, pattern recognition in promoter regions, and microarray analysis. From 2003 to 2008, he was at the European Molecular Biology Laboratory (EMBL) where he worked on literature mining, integration of large-scale experimental datasets, and analysis of biological interaction networks. Since the beginning of 2009, he has continued this line of research as a professor at the Novo Nordisk Foundation Center for Protein Research at the Panum Institute in Copenhagen and as a co-founder and scientific advisor of Intomics A/S. He is a co-author of more than 100 scientific publications that have in total received more than 8000 citations.

Large-scale integration of data and text

High-throughput technologies such as DNA sequencing, microarrays, mass spectrometry, and protein interaction assays have resulted in an ever-increasing flood of data. To make the most of such data sets, it is important to integrate them both with each other and with the existing scientific literature. In my presentation, I will explain how heterogeneous data types are integrated in STRING to construct a network of functional associations. I will also introduce STITCH, which extends STRING with protein-chemical interactions as well as a suite of three new web-based resources that use similar techniques, including text mining, to associate the proteins in the STRING network with cellular compartments, tissues, and diseases. Finally, I will talk about how we currently apply similar techniques to mine data and text from electronic health records and registries.

Jonathan Hoser,
Ph.D. Student, Helmholtz Zentrum München

Jonathan Hoser has graduated from the Bioinformatics Curriculum at the Munich Universities. During his Masterthesis, he had 'first contact' with Next Generation Sequencing, the then first generation of technologies. During his time as a PhD-Student, technologies evolved, and with them the projects changed. Always at (or close to) the edge of technology, with new datatypes, new datasamples, new applications and of course new questions. Jonathan managed to obtain deep insights into the technologies, their advantages, uniquenesses and disadvantages as well as into many applications and answers to questions answerable by NGS.

Next-Generation-Sequencing: From Methods, Applications and Data Analysis

Being the cause for the recent years' explosive growth of biological data, Next-Generation-Sequencing (or NGS) has seen technologies come and go, with a few constant key players. Each of the technologies has different advantages and disadvantages and as a result, different optimal applications.

But technology and applications are only the tip of the iceberg, and are well understood; the analysis of the resulting data however, is where current discoveries are made. If the analysis fits the application, b) the technology, and c) the discovery to be made.

While many pairings of a) and b) are readily available, c) is were most of the current work has to go;

This lecture will try to provide a good overview of the technologies and optimal applications, that aims to provide a sound base layer to make educated choices of a)'s and b)'s, and will allow glances into the broad field of c)'s in dependence on a) .

Dr. Robert Küffner, Helmholtz Zentrum München

Robert Küffner is a group leader and lecturer currently at the Helmholtz Zentrum München. He habilitated in informatics in 2010 and has been, since 2003, orking in the computer science and bioinformatics department at the Ludwig-Maximilians Universität München, Germany. Between the years 2000 and 2003, he was head of software development at the National Center for Genome Resources (NCGR) in New Mexico, USA. He received his PhD in molecular biology in 1998 at the Heinrich-Heine Universität Düsseldorf, Germany. Küffner's main interests include the investigation and reconstruction of biological networks via Petri Nets as well as research in the areas of text mining, expression analysis, gene regulation, and systems biology. Approaches and tools resulting from this research have been applied in many projects to provide systematic bioinformatics support; e.g. for the pre-processing, analysis and integration of large-scale transcriptomics and proteomics datasets. Recently, his team was recognized as best performer in two international community-wide challenges where comprehensive blinded assessments of network inference approaches have been conducted. He is also challenge organizer, for instance aiming at the prediction of disease progression of Amyotrophic lateral sclerosis (ALS) patients.

Regulatory Networks: The reconstruction of gene regulatory networks from mRNA expression data is an important but difficult problem. Available approaches are characterized by their specific performances, data requirements, and inherent biases. The so far largest assessments of inference approaches has been conducted via recent community-wide challenges, the Dialogue on Reverse Engineering Assessment and Methods (DREAM). Predictions are evaluated against experimentally supported interactions in procaryote and eukaryote model organisms. Many independently contributed methods were analyzed including well-known and novel approaches. Further analyses revealed limitations of the current methods, for instance that gene expression data alone is insufficient to understand the regulatory circuitries. Required additional experiments can for instance be obtained from the ENCODE project that aimed to identify all regions of transcription, transcription factor association, chromatin structure and histone modification in the human genome sequence.

Jan Krumsiek,
Helmholtz Zentrum München

Jan Krumsiek studied Bioinformatics at the TU Munich and LMU Munich, receiving his diploma in 2009. He obtained his PhD in Bioinformatics from the TU Munich in 2012, working on network methods for metabolomics data at the Institute of Bioinformatics and Systems Biology (IBIS), Helmholtz Zentrum München. From 2012 to 2013, he worked as a Postdoctoral Fellow at the IBIS. Since 2013, he leads the 'Systems Metabolomics' research team at the Institute of Computational Biology, Helmholtz Zentrum München. Moreover, he is a visiting fellow at the Weill Cornell Medical College, New York City, USA.

Quantitative models and simulation for Systems Biology

In Systems Biology, experimentalists and theoreticians make a concerted effort to unravel the functionality of complex biological systems: Experimental knowledge allows the creation and refinement of mathematical models, which, in turn, help to design and optimize further experiments. In this talk we will go through basic principles of mathematical modeling, covering differential equations, mass-action kinetics, enzyme-catalyzed reactions, stability analysis, discrete modeling approaches, network biology and biostatistics. We will discuss all methods in the context of real biological examples like, for example, stem cell differentiation mechanisms and metabolomics measurements.


HELENA
Helmholtz Graduate School Environmental Health


GSISH TUM
Graduate School of
Information Science
in Health

HELENA
Helmholtz Graduate School Environmental Health

GSISH TUM Graduate School of Information Science in Health

Dr. Gabi Kastenmüller, Helmholtz Zentrum München

Dr. Kastenmüller holds a master's degree in chemistry and a master's degree in computer science from the Ludwig-Maximilians Universität München. In 2009, she obtained her PhD in Bioinformatics from the Technische Universität München for her work on *in silico* prediction and comparison of metabolic capabilities from sequenced genomes. In the same year, she joined Karsten Suhre's group focusing on metabolomics at the Institute of Bioinformatics and Systems Biology (IBIS), Helmholtz Zentrum München, where she was involved in the analysis of various large-scale metabolomics experiments. In 2010, she spent four months at the metabolomics company Metabolon, Inc, USA as a visiting scientist, delving into metabolite identification and the interpretation of spectral data. Since 2011, she has headed the Metabolomics Group at IBIS. Her group mainly focuses on the analysis and interpretation of high-throughput metabolomics data sets by the means of bioinformatics and systems biology. In particular, they are interested in the inborn metabolic individuality in human populations and how this individuality affects predisposition to diseases and response to treatments.

Metabolomics
is the systematic study of ideally all small molecules (metabolites) in a biological system such as a cell, tissue, or a complete organism. In contrast to genomics, transcriptomics, and proteomics, metabolomics describes the end point of all regulatory and enzymatic processes in a biological system, which is influenced by genetic and environmental factors. The metabolite profile of a system thus represents a snapshot of its physiological state.

In my talk, I will first introduce the different approaches that are currently used for quantifying small organic molecules in a high-throughput metabolomics setting. Subsequently, I will give an overview of all steps in the analysis of data produced by these approaches. Finally, we will go through selected examples of current metabolomics projects allowing the participants to find out how metabolomics may be useful for their own research.

HELENA
Helmholtz Graduate School Environmental Health

GSISH TUM Graduate School of Information Science in Health

PD Dr. Arne Pfeufer, HMGU and TU München

Arne Pfeufer, MD, MSc obtained his M.S. in biochemistry in 1992 at Hannover University and his medical degree in 1997 at Humboldt University Berlin. In his medical thesis he characterized causal and modifier genes in hypertrophic cardiomyopathy, a monogenic heart disease. Between 1998 and 2000 he did residency training in internal medicine at the Charité in Berlin and the Kerckhoff-Klinik in Bad Nauheim. In 2001 he started his specialty training in human genetics at the Department of Human Genetics at the Klinikum Rechts der Isar of TU München and at the Helmholtz Zentrum München ((HMGU). There he shifted his scientific interest more towards common genetic determinants of quantitative cardiac traits and common heart diseases. In 2008 he was board certified in human genetics and in 2011 appointed lecturer in human genetics at the TU München. In 2012 he became group leader for predictive and preventive medicine at the Institute for Bioinformatics (IBIS) of the Helmholtz Zentrum München. He continues to work on the same topics including patient and population based studies of cardiovascular disease risk with a special focus on translation of research findings into prediction and prevention.

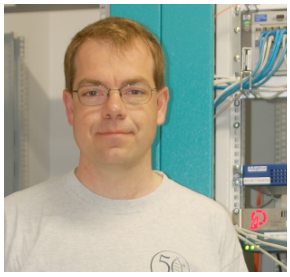
Computational Genetics, Genomics and Genetic Epidemiology

With recent technological advances (Human Genome Project, HapMap,"1000 Dollar Genome") genetics and genomics have become central disciplines in molecular biomedicine and life sciences. They are key to disease gene identification through linkage and association ("gene mapping") and play important roles in their causal validation ("functional genomics"). In my presentation I will provide a framework of unifying genetic and genomic knowledge integrating concepts from human genetics, population genetics, genetic epidemiology and genomics. This framework comprise the essentials scientists working in genomic applications of bioinformatics need to be familiar with whether they work with Next Generation sequencing (NGS), rare variants, genome-wide associations studies (GWAS), common genome variants, tumor genome variants of structural variants of the genome.

HELENA
Helmholtz Graduate School Environmental Health

GSISH TUM Graduate School of Information Science in Health

Prof. Dr. Thomas Rattei, University of Vienna



Professor for Computational Biology and head of the Department of Computational Systems Biology at the University of Vienna. Main interest: microbial genomics and systems biology, genomics of pathogens, molecular inter-species interactions, metagenomics, large scale sequence analysis.

Bioinformatics Databases and Resources

Recent years have seen an explosive growth in biological data, which is usually no longer published in a conventional sense, but deposited in a database and assigned a unique identifying number for quotation in publications. Sequence data from mega-sequencing projects may not even be linked to a conventional publication. This trend and the need for computational management of the data have made databases essential tools for biological research.

This lecture will survey the most important primary databases, which deposit data provided directly by experiments, as nucleotide and protein sequences or 3D structures of macromolecules. It will be discussed which type of information is contained and which sources of errors are most relevant.

Many secondary databases have been implemented during the last decade to interpret primary data and provide predictions of biologically relevant features in un-annotated sequences. There are so many specialized databases, that it is reasonable to discuss the different types and retrieval strategies of those databases and resources. Additionally selected examples, like InterPro and KEGG, are presented in detail.

With the breakthrough of next generation DNA sequencing methods and their application to environmental samples, metagenomic and metatranscriptomic sequences have indelibly expanded the known DNA and protein sequence space to non-culturable microorganisms and environmental communities. The last part of the lecture will introduce the major computational methods necessary to process, analyze and interpret these sequences.

Struggling with the jigsaw puzzle: Bioinformatics for environmental genomics

Microbial life is populating many different habitats (including ourselves – the human body) on earth. Many of these habitats exhibit an unseen diversity of microorganisms, of which most cannot be grown in the lab using cultivation techniques. Applying high-throughput sequencing of DNA and RNA to environmental samples, metagenomic and metatranscriptomic sequences have revolutionized

microbiology and microbial ecology during the last years and allow for an understanding of environmental communities on the molecular level.

This talk introduces the major computational methods necessary to process, analyse and interpret these sequences. The talk will discuss how to pre-process environmental sequences (i.e. estimation and correction of sequencing errors), how to estimate taxonomic diversity in metagenomic and metatranscriptomic samples and how to predict the function and taxonomy of environmental sequences. Finally, the perspectives of upcoming new sequencing and single-cell techniques on environmental genomics will be discussed.

Prof. Dr. Thomas Werner

Adjunct Professor for Internal Medicine-Nephrology Division at the University of Michigan, Ann Arbor, USA. Educated as a chemist at the LMU Munich he completed his PhD in Biochemistry at the same University (1986) and "habilitated" at the TU Munich in the fields of Genetics (2003). He founded a bioinformatics group at the Helmholtz-Zentrum (formerly known as GSF) in 1988 focusing of gene regulation, especially promoter analysis, and founded Genomatix Software GmbH in 1997, where he served as CEO & CSO until 2009. He is now a member of the Board of Genomatix and works as an independent Scientific & Business consultant.

Main interest: gene regulation, comparative genomics, pattern recognition, next generation sequencing, translational omics-based medicine.

Topic: Translational Transcriptomics

We have learned from the ENCODE project and a large body of recent literature that disease-relevant gene regulation is tightly integrated including genomic mutations, epigenetic remodeling, transcription, post-transcriptional regulation and translation and post-translational control. Not all aspects are easily assessed in routine analysis but the availability of microarrays and increasingly also NGS-methods such as DNA-genotyping and RNA-seq anaylses have opened new opportunities to apply omics-based methods in order to gain medically relevant insights. However, there is still a large gap between even the most sophisticated statistical methods, their results, and the biomedical functional information clinicians will require to finally be able to base their diagnosis and therapeutic decisons on such omics-data analyses results.

I will focus on an overview of how a high-through data analysis pipeline for transcriptomics data from the viewpoint of application in translational medicine should look like, where we (at U Mich and at Genomatix) stand today and how to tackle the meta-analysis probpems of integrating other omics data into the pipleine (genotyping, epigenetic data, proteomics data).

For more information follow the pages on
http://www.helmholtz-muenchen.de/en/ibis