Clinical Bioinformatics as a Service

Date and time: September 4, 2016, 09:00 - 17:00 (CEST)
Location: Room Everest, World Forum in The Hague, Netherlands

Agenda

Time	Description	Lead
0830-0900	Arrival & Registration	ECCB organizers
0900-0910	Welcome address	Daniel Stekhoven
0910-1000	**Drug response modelling for individualized systems medicine and precision oncology**	**Tero Aittokallio**
1000-1020	DrugTargetInspector: An assistance tool for patient treatment stratification	Lara Schneider
1020-1100	*Coffee break*	
1100-1120	NEXUS: Comprehensive Molecular Diagnostics in Swiss Clinics	Nora C. Toussaint
1120-1140	Integrating data, tools and infrastructure to enable efficient collaboration and management for personalized therapy of hepatocellular carcinoma	Stefan Czemmel
1140-1230	**Statistical analysis of high-throughput data: interdisciplinarity is the new black**	**Camille Stephan-Otto Attolini**
1230-1330	*Lunch break*	
1330-1420	**Challenges and concepts for bioinformatics services in translational research**	**Dieter Beule**
1420-1440	An infrastructure for sustainable, automated and unified benchmarking	Piotr Wojciech Dabrowski
1440-1455	Creation of a collaborative variant interpretation service for personalised medicine and translational research (*Sponsor talk*)	Christophe Roos BC Platforms
1455-1510	*Coffee break*	
1510-1600	**Implementing bioinformatics for infectious diseases into clinical service**	**Lydia Drumright**
1600-1730	Networking Apéro	Sponsor

Invited Speakers

Dieter Beule
Head of Bioinformatics Core Unit
Berlin Institute of Health

Challenges and Concepts for Bioinformatics Services in Translational Research

Bioinformatics is becoming indispensable for many areas of translational biomedical research. NGS and other omics technologies can be used to address a wide variety of questions with different assays and thus a multitude of different data analysis needs has emerged in clinical contexts. Bioinformatics service and core units are one way to cope with this demand, but the units operate in an area of conflicting requirements: Because both omics technologies and bioinformatics algorithms are quickly evolving the data analysis processes need constant evolution to stay up to date. Clinical applications on the other hand require proper process validation and benchmarking because reproducibility and reliability are key issues. We discuss the concepts used at the Berlin Institute of Health to cope with these challenges and provide insight into example projects from cancer genomics and rare genetic diseases.

Lydia Drumright
Clinical Informatics
University of Cambridge

Implementing bioinformatics for infectious diseases into clinical service

Whole genome sequencing (WGS) for microorganisms has become relatively inexpensive and holds extensive potential for improving diagnosis, treatment, and prevention with respect to infection, especially among hospitalised patients. However, integrating WGS into a clinical service comes with challenges, preventing adoption in most disease areas. Specifically, WGS of microorganisms exists at the intersection between clinical care, public health, research and antimicrobial and vaccination development, which involves many different stakeholders and brings into question responsibility for the work. Additionally, utility of WGS for microorganisms is different between chronic and acute infections, and standardisation of pipelines is complicated by the extensive genetic variation within and between pathogens. These issues will be explored with examples from our work at the Cambridge University Hospitals NHS Foundation Trust (CUH) in the UK.

Camille Stephan-otto Attolini
Bioinformatics-Biostatistics Unit
Institute for Research in Biomedicine (IRB) Barcelona

Statistical analysis of high-throughput data: interdisciplinarity is the new black

The correct analysis of the enormous amount of data generated by new technologies for biological research is a keystone for future scientific advances. Although more and more research is dealing with the processing and analysis of these data, there is still a long way to go regarding experimental design, statistical inference and interpretation of results. As new technologies are introduced to measure and describe biological phenomena it is necessary to understand the biases intrinsic to each of them. To this end statisticians, technicians and biologists must work hand to hand in order to distinguish between biological and technical effects. The same goes for proper interpretation of results, where the aim is to avoid missing biological information while avoiding over-interpretation of statistical conclusions.

The advent of high throughput sequencing allowed for very detailed and precise gene expression measurements, at the cost of increasing the amount of data in several orders of magnitude. We have developed a statistical model and associated software to extract confident isoform expression estimates from RNA-seq experiments. Moreover, when dealing with RNA-seq several experimental parameters must be decided (e.g. read length, coverage, fragment size) which must be contrasted with the number of biological replicates in order to maintain low costs but meaningful answers. The right balance between precision of measurement and estimation of biological and technical variability is a problem that up to our knowledge has never been addressed in RNA-seq setups. We have extended our previous tool to perform sample size calculations and personalized experimental design.

In recent years it has become possible to measure gene expression at the cell level. Single-Cell RNA-seq is a very promising field with very complex and demanding analytical problems. We have dealt with some of these allowing for informative biological conclusions.

Once seen as a luxury or rarity, interdisciplinary collaborations are nowadays the only option to perform conclusive research in an era of big data and ever more complex biological knowledge.

Tero Aittokallio
Computational Systems Medicine
Institute for Molecular Medicine Finland (FIMM)

Drug response modelling for individualized systems medicine and precision oncology

Comprehensive testing of the response of patient-derived samples to a wide collection of chemical compounds is increasingly being used to tailor drug treatments for individual cancer patients. However, high-throughput drug testing experiments result in high-dimensional datasets, with inherent measurement noise and technical variability, which hinders many downstream analyses, such as detection of differential drug sensitivities or stratification of patients based on their response patterns.

This talk describes the systems medicine modeling framework developed and used at FIMM, which enables reliable quantification of differential drug sensitivities, mapping of target addiction networks behind the individual response profiles, prediction of targeted drug combinations for relapsed patients, as well as identification of biomarkers that are predictive of drug sensitivities. The pipeline is illustrated in several case studies involving both hematological cancers as well as solid tumors.

Contributed Talks

Lara Schneider
Center for Bioinformatics, Saarland University

DrugTargetInspector: An assistance tool for patient treatment stratification

One of the Hallmarks of Cancer is the acquisition of genome instability and mutations. In combination with high proliferation rates and failure of repair mechanisms, this leads to clonal evolution within a tumor, and hence to a high genotypic and phenotypic diversity. As a consequence, successful treatment of malignant tumors is still a grand challenge. Moreover, under selective pressure, e.g. caused by chemotherapy, resistant subpopulations may emerge that in turn can cause relapse. In order to minimize the risk of developing multi-drug resistant tumor cell populations, optimal (combination) therapies have to be determined on the basis of an in-depth characterization of the tumor's genetic and phenotypic makeup, a process that is an important aspect of stratified medicine and precision medicine. To this end, we present DrugTargetInspector (DTI), an interactive assistance tool for treatment stratification. DTI analyzes genomic, transcriptomic and proteomic datasets and provides information on deregulated drug targets, enriched biological pathways and deregulated subnetworks, as well as mutations and their potential effects on drugs, drugs targets, and genes of interest. Using DTI's powerful web-based toolsuite allows users to characterize the tumor under investigation based on patient-specific -omics datasets and to elucidate putative treatment options based on clinical decision guidelines, but also proposing additional points of intervention that might be neglected otherwise. DTI can be freely accessed at https://dti.bioinf.uni-sb.de.

Nora C. Toussaint
NEXUS Personalized Health Technologies, ETH Zurich

NEXUS: Comprehensive Molecular Diagnostics in Swiss Clinics

High-throughput genomics and screening technologies have changed the way biomedical research is performed. The transition from directed testing of a few specific targets, selected based on prior knowledge, to analyzing comprehensive high-throughput data promises tremendous possibilities but also introduces new challenges to the clinical world. Despite the great potential, particularly for the treatment of patients with rare diseases, with tumors lacking known targetable mutations, and of those considered end-of-treatment line, the use of high-throughput techniques to go beyond standard diagnostics is not fully established in the clinics yet. Establishing high-throughput molecular diagnostics for clinical use requires specific protocols accounting for stringent quality control, privacy issues, and thorough process documentation. To this end, NEXUS, a core facility at ETH Zurich, provides state-of-the-art bioinformatics, statistical analyses, and screening of FDA-approved drugs combined with high standards for quality control, data privacy, and reproducibility. We present a workflow for the molecular profiling of matched tumor and normal samples from sequencing to clinical decision support. In addition to the identification of somatic variants, our workflow links alterations to possible treatment options, both cancer type-specific and off-label. The analysis results are summarized in a concise and clearly structured report designed to form the basis for discussions in a clinical molecular tumor board.

Stefan Czemmel
Quantitative Biology Center, University of Tübingen

Integrating data, tools and infrastructure to enable efficient collaboration and management for personalized therapy of hepatocellular carcinoma

The Multiscale HCC project is a multi-omics level project at the University of Tübingen that aims to combine imaging, clinical data and molecular profiling of tumor tissue in order to improve diagnosis and therapy of advanced hepatocellular carcinoma (HCCs).

The underlying experiments, ranging from tissue extraction from patient's biobank material and sample preparation to the final step, the data generation can be very complex and involves several steps that are essential to adequately analyze the data.

As part of the project we developed a data and project management infrastructure that facilitates the modeling of this complex experimental design, the interplay with different omics data acquisition facilities and the bioinformatics analysis.

In order to establish individualized therapy concepts, the generated multi-omic profiles (exome, transcriptome, proteome and metabolome) of patients with HCC will be stored together with metadata centrally at the Quantitative Biology Center (QBiC) and raw and results data are made accessible to all consortium/project members through a central web interface.

For the bioinformatics analysis, we established an IT infrastructure that allows for automated execution of data processing and data analysis pipelines. The pipeline functionality is integrated into a central data navigation and can be executed on a one-click-basis.

Piotr Wojciech Dabrowski
Robert Koch Institute

An infrastructure for sustainable, automated and unified benchmarking

Benchmarking of tools is an integral part of bioinformatics analyses on several levels. For users such as bioinformatics core facilities, benchmarks are necessary to make informed decisions when selecting tools for their specific research question and dataset. For developers, benchmarking allows the identification of gaps in existing solutions, the resulting definition of goals for new tools, and finally the verification whether new tools reach these goals. In contrast to this importance of benchmarking, there is no agreement on how to approach this task based on unified technical concepts. Benchmarks of small selections of tools are published on a regular basis. However, these mostly use different metrics and different datasets making a reliable comparison of different tools possible only when they appear in the same benchmark. We propose to remedy this by developing an infrastructure for automated, continuous benchmarking of bioinformatics tools through leveraging current developments in bioinformatics infrastructure. Tools can be automatically pulled from a central registry (bio.tools) and matched (EDAM annotations) to appropriate benchmarking datasets (GMI) and scripts for the calculation of appropriate metrics. Thanks to machine-readable interface descriptions (CWL) and portable containers (Docker/BioShaDock, Debian Med), these components can be automatically combined into complete, executable (CWLtools) workflows. The resulting metrics can be visualized for end users in a web interface (SeqWiki) that allows the comparison of tools' performance over a broad range of datasets, enabling users to select the tools most appropriate for their specific problems and giving developers access to an objective way of assessing their tool's performance.

Sponsor Talk

Christophe Roos
BC Platforms

Creation of a collaborative variant interpretation service for personalised

University hospitals, especially with national scale genomics or biobank initiatives need the full gamut of clinical bioinformatics services, spanning both research and clinical healthcare needs. However, although many companies exist that offer clinical sequence data pre-processing and interpretation, these services are only fitted to the US healthcare model, or to private clinical units.

When molecular genetics in hospitals is moving from targeted panels to whole-exome sequencing, and eventually to whole genome sequencing, there comes a need to build genome data banks, which can be utilised throughout every patient's whole life, which can accumulate curated data and which can serve diagnostics, preventive medicine and research. From an IT and data security perspective, genome data needs to be managed locally, and closely integrated to other clinical information.

In this model, it would still be possible to use 3rd party interpretation services if they can deliver information and reports in computer readable format.

Workshop Organizers

Niko Beerenwinkel
Department for Biosystems, Science and Engineering
ETH Zurich

SIB Swiss Institute of Bioinformatics

Wolfgang Huber
Genome Biology
European Molecular Biology Laboratory, Heidelberg

Simon Tavaré
Cancer Research UK
Cambridge Institute

Daniel Stekhoven
NEXUS Personalized Health Technologies
ETH Zurich

SIB Swiss Institute of Bioinformatics