Report from Dagstuhl Seminar 13491

Computational Mass Spectrometry

Edited by

Ruedi Aebersold[1], Oliver Kohlbacher[2], and Olga Vitek[3]

1 **ETH Zürich / University of Zürich, CH,** aebersold@imsb.biol.ethz.ch
2 **Universität Tübingen, DE,** oliver.kohlbacher@uni-tuebingen.de
3 **Purdue University, US,** ovitek@stat.purdue.edu

Abstract

The last decade has brought tremendous technological advances in mass spectrometry, which in turn have enabled new applications of mass spectrometry in the life sciences. Proteomics, metabolomics, lipidomics, glycomics and related fields have gotten a massive boost, which also resulted in vastly increased amount of data produced and increased complexity of these data sets. An efficient and accurate analysis of these data sets has become the key bottleneck in the field. The seminar 'Computational Mass Spectrometry' brought together scientist from mass spetrometry and bioinformatics, from industry and academia to discuss the state of the art in computational mass spectrometry. The participants discussed a number of current topics, for example new and upcoming technologies, the challenges posed by new types of experiments, the challenges of the growing data volume ('big data'), or challenges for education in several working groups.

The seminar reviewed the state of the art in computational mass spectrometry and summarized the upcoming challenges. The seminar also led to the creation of structures to support the computational mass spectrometry community (the formation of an ISCB Community of Interest and a HUPO subgroup on computational mass spectrometry). This community will also carry on with some of the efforts initiated during the seminar, in particular with the establishment of a computational mass spectrometry curriculum that was drafted in Dagstuhl.

Seminar 1.–6. December, 2013 – www.dagstuhl.de/13491
1998 ACM Subject Classification J.3 Life and medical science
Keywords and phrases computational mass spectrometry, proteomics, metabolomics, bioinformatics
Digital Object Identifier 10.4230/DagRep.3.12.1

1 Executive Summary

All participants of Dagstuhl Seminar 13491

Motivation

Mass Spectrometry (MS) is an analytical technique of immense versatility. Detection of explosives at airports, urine tests for doping in sports, tests for cancer biomarkers in a clinic – all these rely on mass spectrometry as the key analytical technique. During the last decade, technological advances have resulted in a flood of mass spectrometric data (high-resolution mass spectrometry, mass spectrometry coupled to high-performance liquid chromatography – HPLC-MS). The publication of the first human genome in 2001 was a key even that enabled the explosive development of proteomics, which led to the conception of the Human

Proteome Project in 2010. Today, mass spectrometric techniques are an indispensible tool in the life sciences. Their development, however, is more and more hampered by the lack of computational tools for the analysis of the data. Modern instrumentation can easily produce data sets of hundreds of gigabytes from an individual sample. Most experimental groups are no longer able to deal with both the amount and the inherent complexity of these data. Computer science has the necessary tools to address these problems. It is thus necessary to intensify collaboration between the three key communities involved: life scientists applying MS; analytical chemists and engineers developing the instruments; computer scientists, bioinformaticians and statisticians developing algorithms and software for data analysis.

Goals

The seminar 'Computational Mass Spectrometry' is a follow-up seminar to the successful Dagstuhl seminars on 'Computational Proteomics (05471 and 08101). The different title was chosen to reflect the growing scope of computational mass spectrometry: from proteomics to metabolomic, lipidomics, and glycomics.

The goal of the seminar was thus to assess the state of the art for the field of computational mass spectrometry as a whole and to identify the challenges the field will be facing for the years to come. To this end we put together a list of participants covering both computational and experimental aspects of mass spectrometry from industry and academia from around the world. The result of these discussion should then be summarized in a joint status paper.

Results

The seminar was very productive and led to a number of tangible outcomes summarized below.

The Big Challenges

Not unexpectedly, it turned out to be difficult to identify the *big challenges* of the coming years and views on this differed quite a bit. After lengthy discussions, we were able to categorize the challenges. We are currently in the process of finalizing the draft of a paper on these challenges for computational mass spectrometry, which is supposed to be submitted by end of March 2014. The paper is a joint work of all the participants and will document the current state of the field. The challenges identified were the following:

Challenges of computational and statistical interpretation of mass spectra

- *Identification*
  Identification of analytes is still a challenge. In proteomics, the identification of post-translational modifications and of different proteoforms pose problems. Also the identification of non-tryptic peptides (peptidomics, MHC ligands) are interesting problems. Estimation of false-discovery rates based on target-decoy approaches has been criticized, but there is still a distinct lack of established alternatives. With the increasing interest in small-molecule mass spectrometry, the identification of metabolites, glycans, and lipids is increasingly becoming an issue and the algorithmic support for this is currently still lacking.
- *Quantification*
  Quantification faces challenges due to the – still-growing – diversity of experimental methods for analyte quantifications that necessitate a permanent development of new

computational approaches. There are also more fundamental, statistical problems, for example, inferring the absence of an analyte based on the absence of a signal. Quantification is also expected to contribute to the understanding of protein complexes and their stoichiometry.

Challenges arising from new experimental frontiers

- *Data-independent acquisition*
  The recent developments of data-independent acquisition techniques resulted in a set of entirely new computational challenges due to the different structure of the underlying data.
- *Imaging*
  Imaging mass spectrometry has become mature on the experimental side. The analysis of spatially resolved MS data, however, poses entirely new problems for computational mass spectromtry with increased complexity and data volume.
- *Single-cell mass spectrometry*
  Multi-parameter single cell mass spectrometry enables the characterization of rare and heterogeneous cell populations and prevents the typical averaging across a whole tissue/cell population. The key challenge will be the development of new computational tools able to define biologically meaningful cell types and then model the dynamic behaviour of the biological processes.
- *Top-down proteomics*
  Despite its obvious advantages of top-down approaches for functional proteomics, isoform identification and related topics, the approach suffers from unmet challenges on the computational side. Methods for mass spectrum deconvolution need to be improved and algorithms for the identification of multiple PTM sites are required.

Challenges of extracting maximal information from datasets

- *Democratization of data*
  Public availability of large datasets enables novel types of studies in computational mass spectrometry (data mining). The standardized deposition is and reliable repositories handling this data is still a major problem that needs to be addressed.
- *Integration of MS data with different technologies*
  Increasingly, computational biologists face data from multiple omics technologies. Integrating data from computational mass spectrometry across omics levels (genomics with transcriptomics, transcriptomics with proteomics, proteomics with metabolomics) poses a difficult data integration challenge, but will be essential for a more comprehensive view of the biological systems under study.
- *Visualization of heterogeneous data sets*
  The amount, structure and complexity of large-scale mass spectometric data turns out to be a challenging issue. While some end-users of theses methods tend to be interested in a final, aggregated result of a complex data analysis pipeline, it is often essential to analyze the data conveniently down to the raw spectra. Tools navigating these data sets on all levels are currently not yet available.

Community Building

It was felt among participants that computational mass spectrometry is lacking a structured community. Researchers in computational mass spectrometry come from diverse backgrounds: statistics, computer science, analytical sciences, biology, or medicine. Traditionally they are thus organized in different scientific societies, for example the International Society

on Computational Biology (ISCB), the American Society of Mass Spectrometry (ASMS), the Human Proteome Organization (HUPO), the Metabolomics Society, and of course various national societies. Many participants attend both computational and experimental conferences in the area of mass spectrometry organized by these different organizations. Participants suggested to form subgroups for computational mass spectrometry in different societies. At the same time, in order to avoid duplication of structures and efforts, it was planned to share these subgroups across the different societies and establish joint chairs of these groups, organize joint workshops, and coordinate educational activities.

After the Dagstuhl seminar we contacted ISCB and HUPO to discuss the formation of these subgroups. After intensive discussion with the societies, HUPO and ISCB both agreed to this plan. A HUPO subgroup CompMS on computational mass spectrometry was formed. In parallel, ISCB agreed to form a *Community of Special Interest* (CoSI) CompMS. Both subgroups share a joint structure. A joint steering committee (Steering Committee (Oliver Kohlbacher, Olga Vitek, Shoba Ranganathan, Henning Hermjakob, and Ruedi Aebersold) has been established to guide both groups through their formation period. The groups have set up a joint mailing list, a website, and are currently planning initial kick-off meetings as satellite workshops to ISMB 2014 (Boston) and HUPO 2014 (Madrid).

Teaching Initiative

Recognizing the great need for educational materials for various audiences (bioinformaticians, biologists, computer scientists) some participants initiated an initiative to put these materials together as online courses. Discussions of this initiative have come quite far. It is currently planned to come up with a core curriculum for mass spectrometry. This core curriculum will be open for discussion within the computational mass spectrometry community. After the contents of the core curriculum has been established, tutorial papers will be solicited for the various modules of the curriculum. These papers will refer to each other, will use a coherent vocabulary and notation and will appear as a paper collection online in PLoS Computational Biology (edited by Theodore Alexandrov). Additional materials will be included, for example, online courses and lecture videos. An initial tutorial workshop is currently in planning to kickstart the further development of the curriculum.

Reviewing Guidelines

A working group (see Section 4.5) discussed the problems that computational papers face in the reviewing process. The main driver for this discussion was expediting the review process, specifically in terms of reducing the number of review cycles. It is worth noting that the Journal of Proteome Research (JPR), published by the American Chemical Society (ACS), presents a special case since this journal is the only one in the field that does not have a regular mechanism for the reviewers to see the comments of the other reviewers and the corresponding responses of the authors after each round of review. The proposal initially on the table was to share all reviews among reviewers and invite comments and changes before the first editorial decision is made for the first round of review. This system is, for instance, already in place via EasyChair (software used for the RECOMB meetings, but not for proteomics journals). After discussion, it was decided that it would clearly be beneficial if the JPR distributed all reviews among reviewers after each stage of revision. But it was felt that it would only be necessary to collect comments and feedback from the reviewers (based on sending them all reviews) before the editor reached an initial decision in cases where there was substantive disagreement among reviewers on critical points. These ad

hoc communications can be handled in a semi-manual way within the existing manuscript management systems used by the proteomics journals, with the added benefit of maintaining an audit trail for the process. The reviewing guidelines developed by the participants in Dagstuhl are currently being discussed by with the editorial boards of different journals (currently J. Proteome Res and Mol. Cell. Prot.).

2 Table of Contents

3 Structure of the Seminar

The seminar was structured in an unusual way in so far as we omitted introductory talks by participants on purpose. While the organizers had prepared some material on the most obvious topics, we started the seminar in a very open manner and solicited proposals for break-out groups. The participants then voted on these topics.

3.1 Planned Topics

The topics initially planned to be covered were:

- **Identification from MS/MS spectra.** Although these topics received a lot of attention from the computational community in the past, many aspects of spectral identification are still in need of computational solutions. Open problems include the identification of high-quality spectra that are not identified by conventional methods, identification of metabolites, identification of post-translational modifications, MS/MS spectra from a mixture of sources, proteogenomics and metagenomics, metaproteomics, cross-linking data. These problems require advanced algorithmic solutions, in particular efficient search algorithms to deal with the combinatorial nature of the search space that needs to be explored during database searches.
- **Quantification.** There is currently a great diversity of sample preparation workflows and of mass spectrometry-based quantitative techniques, and these produce quantitative measurements with diverse properties and structure. Although statistical models and algorithms for analysis increasingly appear, there is currently no consensus on the appropriate analysis of these data. Open issues include the choice of normalization, handling of missing data, features of low quality and features with uncertain identity, and treatments of interferences (e.g., due to post-translational modifications).
- **Downstream interpretation of the identified and quantified proteins.** Many biological investigations require interpretations of mass spectra beyond the lists of identified and quantified proteins. Relatively little work has been done on the downstream interpretation of proteomic experiments so far. Open problems in biological applications include integration of data from heterogeneous sources (e.g., from transcriptomic, metabolomic and interactomic experiments) and with functional annotations, and dealing with novel data structures such as spectral images. Open problems in biomedical applications include the development of diagnostic and prognostic rules from the quantitative protein profiles, integration of the protein predictors with the clinical records, and assessment of the predictive ability. Solutions involve statistical inference and machine learning techniques.
- **Experimental design.** The choice of experimental design is critical for the success of all proteomic experiments, however many practitioners do not approach this choice from the computational and statistical point of view. The issues in need of computational development are the choice of type and number of the biological samples, and allocation of the experimental resources in space and time, guiding the choice of labeling strategies in label-based experiments, and the choice of target proteins, peptides and transitions in targeted experiments. Solutions to these problems involve approaches from statistical modeling and stochastic optimization.
- **Data management.** A major bottleneck in the analysis of proteomic experiments is the volume and complexity of the generated data. Although a lot of progress has been recently achieved in developing proteome-centric formats, there is an urgent need for the infrastructure for integrating measurements across experiments and data types. Most

currently available data repositories are genome-centric, and development of proteome-centric repositories is a complex but important task. Open issues in this field are related to databases and data mining.
- **Software engineering.** Even the best algorithms are of a limited practical use if they are not accompanied with the biologist-friendly software. The complexity of the analytical tasks creates new opportunities for development of comprehensive but modular pipelines, and standardization of analytical workflows.
- **Training of interdisciplinary experts.** The current and future computational approaches can rarely be used blindly, and should often be adapted to the experiment at hand. There is a strong need for training of interdisciplinary experts, and a Dagstuhl seminar can help identify the opportunity for such training.

3.2 Seminar Schedule

After a brief introduction the participants decided to form working groups to discuss and assess the state of the art as well as upcoming challenges for various topics. These working groups formed partially *ad hoc* and partially based on suggestion by previous working groups. The schedule of the seminar is shown below. Brief abstracts describing the conclusions of the individual working groups are reproduced in Section 6.

- **Monday**
  - Welcome and introductions
  - Formation of working groups (WGs)
  - WG 'New instrumentation challenges, quantitative proteomics, and data-independent acquisition'
  - WG 'Multiple Omics, Integration, and Metabolomics'
  - WG 'Democratization of proteomics data and big data'
- **Tuesday**
  - WG 'Teaching and Outreach'
  - WG 'Data-independent acquisition'
  - WG 'Guidelines for sustainable software and reviewing of computational MS' papers
  - WG 'Single-cell technologies and imaging'
  - WG 'Biological applications'
- **Wednesday**
  - Discussion of a first version of the challenges paper
  - Continuing discussions in the various working groups
  - Outing The usual Dagstuhl outing went to Trier, where the participants had a guided tour of the city, visited the Weihnachtsmarkt, and finally a wine tasting at one of the local wineries.
- **Thursday**
  - WG 'Non-DIA Quantification'
  - WG 'Algorithmic issues and big data in mass spectrometry'
  - WG 'Statistical methods and experimental design'
  - WG 'Categorization of challenges'
- **Friday**
  - Final discussion of the status paper, distribution of work
  - Discussions on the formation of a computational mass spectrometry community

Figure 1 Some impressions from the seminar and the outing in Trier (photos: Oliver Kohlbacher).

4 Working Groups

The various working groups formed and re-formed throughout the whole seminar. Each group reported on its results at the beginning of each day and several working groups came up with proposals for the formation of other groups. While some groups (e.g., data-independent acquisition) were active throughout the seminar, others were active for shorter periods of time only.

4.1 New Technologies and Data-independent Acquisition

Much of the discussion centered on whether data-independent acquisition (DIA) could replace data-driven acquisition (DDA) for identification purposes, or whether it was best suited as a targeted proteomics approach. It was felt by the group that at the moment, DIA is already a great alternative to SRM approaches, and indeed much of the published work is using DIA in this context. This does not mean that all the issues in using DIA as an alternative to SRM have been resolved, and signal processing is still difficult. A better framework to integrate data from different experiments (technical, biological replicates, etc.) is still needed.

Sample complexity and the type of experiments performed may drive the selection of using DIA for identification, SRM-like quantification, or both. The outcome of these discussions is that we should be thinking about different types of workflows as separate, and try to work out examples of the kind of computational challenges that are posed by the different types of experimental designs. Examples of such experimental samples to be worked out are clinical samples and other samples of high complexity, affinity purified samples and samples enriched for PTMs for which peptide level information is needed. Dataset size (meaning whether you are only analyzing a few samples, or hundreds of them) may also bring us to consider different computational approaches. This should help framing the next small group discussions.

We also briefly discussed how to discover, score and quantify PTMs from DIA data, but this is certainly a point that should be further discussed. An ongoing discussion is how starting from a peptide and looking for its presence (or evidence of absence) in a sample was different (conceptually and in terms of FDR calculation) from starting from the spectra (which could be virtual spectra) and performing a database search.

Identified issues are:

- signal processing issues (dynamic range estimates)?
- effects of complexity on identification and quantification?
- how to handle datasets of different sizes?
- how to score PTMs?
- what are the limitations due to current instrumentation and which of them will be overcome in the next generation(s) of instrumentation?

4.2 Multiple Omics, Integration, and Metabolomics

For a full understanding of biological complexity and functionality, system-wide identifications and quantification of all relevant players in a cell would be highly desirable. However, the various entities (genes, transcripts, proteins, metabolites, etc ...) all require different technological approaches for their assessment and quantification. Consequently, different research communities and scientific approaches have developed around each of these. “Multi-omics” aims to bridge and re-connect these different worlds, in order to take advantage

of synergies and complementaries. For example, concomitant/parallel measurements of transcripts and proteins might help in removing false-positive hits, and can reveal which transcriptional events/isoforms gave rise to mature proteins and are hence perhaps more relevant functionally. Likewise, environmental proteomics experiments often require DNA sequencing to be conducted in parallel from the same sample, in order to construct matched ORF databases for peptide identifications. Most importantly, mechanistic and quantitative modeling of functional processes inside cells often requires precise quantifications of players of various types. What are currently the biggest obstacles for multi-omics integration? Mainly, common ground needs to be established with respect to nomenclature, statistics and reference datasets. Which representation can encompass all players? Which entities will be the "central" ones, where the others are attached? Rather than "linear" genome browsers, networks are likely to be an apropriate representation. In addition, novel software tools need to be developed for visualization purposes and dissemination, and algorithms developed for scoring datasets jointly, not separately. Most importantly, however, since few scientists are experts on multiple dimensions of -omics, renewed efforts need to go into providing meta-sites resource directories on available software, pipelines, tools and procedures.

4.3 Democratization of Proteomics Data and Big Data

The session on the democratization of proteomics data and big data discussed several topics related to overarching issues around computational proteomics. The focus was on teaching and outreach materials, logistics, quality control, peer reviewing and ethics and privacy issues. Two additional working groups were proposed as a result of this session, one on drafting a curriculum for teaching computational mass spectrometry and the other on drafting guidelines and manuscript types for computational contributions in proteomics journals.

4.4 Teaching and Outreach

The participants of the seminar felt that at the moment there is a lack of teaching material to bring interested students and experimentalists quickly into the field of computational proteomics. In addition it was felt that this is also detrimental for the external perception of the field. The group thought it would be beneficial to come up with a curriculum for an online course which:

- allows people from various backgrounds to enter the field
- explains the main concepts as opposed to operating certain programs
- is set up as a backbone of main lectures complemented by some prerequisite courses to homogenize assumed knowledge in biochemistry or mathematics and by sidetrack courses that go in-depth (e.g., details of HPLC or how to solve a specific algorithmic problem efficiently)

In the discussion it was decided to discuss the content of such a curriculum (draft is in the seminar wiki). In order to extend this, the plan was to identify potential funding souces for an experienced postdoc as an editor of this course (homogenize nomenclature, provide templates, technically implement it, approach contributors and reviewers, etc.). This effort should be coordinated within the scientific societies, for example with existing efforts at HUPO (or elsewhere if they are presently not known). The group produced a first draft of the core topics of such a curriculum (see Appendix A).

4.5 Guidelines for Sustainable Software and for Reviewing of Computational MS papers

The session on guidelines for sustainable software and reviewing of computational mass spectrometry papers focused on the specific topic of improving the guidelines (for authors as well as for reviewers) that apply to computational contributions to proteomics journals. An important aspect of the discussion concerned the ability to provide a place for all possible contributions, including highly innovative but immature approaches, highly mature but incrementally innovative tools, and both highly innovative as well as mature contributions. This resulted in three manuscript types, matched to already existing manuscript categories for logistic convenience at the journals, and clarity resulting from the already defined scope: Brief Communication, Application/Technical Note, and Research Article. Furthermore, a request was made to introduce a new subtype for two of these types, that would bear the 'stamp of approval' resulting from passing the checklist. The ability of journals to accommodate this request is to be investigated, and the specific title they would carry is to be decided.

4.6 Single-cell Technologies and Imaging

Understanding molecular characteristics in dynamic and heterogeneous cell populations on the levels of single-cell and cell populations is an essential challenge of biology. Mass cytometry is an emerging technique for single-cell analysis allowing analysis of up to tens of molecules simultaneously either in suspension or directly on tissue. Imaging MS is a maturing technique for spatially-resolved mass spectrometry analysis directly on tissue. Both techniques provide tools for molecular imaging and allow for unprecedented insight into biological organization of biological processes on the levels of individual cells and cell populations and paving the way to spatial pathway analysis. We outlined the challenges which are faced by these young and promising techniques.

4.7 Biological Applications

After discussing some definitions (who are our 'customers' – example biologist name 'Steve', who are the 'computational mass spectrometrists'), current challenges on the biological side were identified:

1. Communication about the biological question
   This is a challenge for us and Steve. He has to take time to educate us, we'll end up very experienced biologists. This is as important for the data collectors as the data processors. Challenges for bioinformatician: communication with Steve, establishing end point, educating future Steves.
2. Triaging
   Will it ever work? Is there a better method? Therefore we have to have a good idea of alternative technologies.
3. Experimental design
   How can you design the critical experiment to give a clear answer? We have to manage Steve's expectation – this is a challenge in itself. Define bioinformatics pipeline and endpoint to determine whether the design is sufficient and the experiment is feasible.
4. Knowing and selling our unique selling points
   - The read out of translational control and protein degradation: (a) is a protein translated, if so, how much of it and (b) is a protein isoform translated

- Subcellular/organelle (complementary methods would be: immunocytochemistry, confocal microscopy)
- post translational modifications/truncations
- protein interactions

These biological applications result in several grand challenges for computational MS:

- System wide probing of fluxes
- Dynamics of everything (multiple layers)
- Coverage, comprehensive measurement of the proteome and isoformome
- Combinatorial nature of PTMs – going beyond the single PTM
- Epitope mapping – MHC ligands demand de novo identification

4.8 Non DIA-quantification

Quantitation is fundamental to characterizing biological material with mass spectrometry. Despite major advances made over the past 15 years, particularly with respect to proteomics, significant obstacles remain that inhibit our ability to comprehensively and unambiguously evaluate our data, and share them with others. We describe here five critical computational challenges that underly all quantitative techniques. Solving them will promote wider dissemination of high-confidence results, and deeper biological understanding.

4.9 Algorithmic Issues and Big Data in MS

One key issue arises from big data in computational mass spectrometry: finding structure in the big data sets. In proteomics, a key statistical challenge is still the determination of FDRs beyond the target-decoy approach. Also the inclusion of structural data as part of protein inference is a big challenge. Identification of proteins can be improved by inclusion of genomic variants, algorithms to deal with PTMs, and annotation problems. New algorithms are also required to analyze non-shotgun data and top-down data. Finally, the visualization of all levels of the data poses interesting problems.

4.10 Statistical Issues and Experimental Design

Statistics is key for reproducible research. It is central to experimental design, data processing and downstream analysis. In some applications, there are available statistical workflows that can still be improved, but are not perceived as a grand statistical challenges, for example, confidence in peptide identification in a DDA-type workflow. There are other areas where even basic tools do not exist, and the need for these is more pressing (e.g., new types of experimental workflows are in need of solutions). Furthermore, perhaps one of the most important challenges is the communication and training of good statistical principles.

A A draft curriculum for computational mass spectrometry

- Intro, Goals, of (Computational) Proteomics and Metabolomics
  - What is the proteome, metabolome, transcriptome
  - What is it good for? (Protein ID, Quantitation for pathway analysis, Biomarker analysis, maybe concrete examples, e.g. protein interaction)
  - What skills do you need?
- Basic Biochemistry (PR)
  - DNA, Proteins, PTMs
  - Lipids, Metabolites
  - Degradation
- Analytical Chemistry
  - Precision, accuracy
  - Dynamic range
  - LOD/LOQ (noise)
  - Profiling vs. Targeted
- Statistics
  - Distributions
  - p-value
- Needed Computational concepts
  - Algorithms, run time, space requirements
  - Machine learning
- Mass spectrometry
  - LC/GC/CE
  - MS (precurscor selection)
  - MS/MS (ion types, ion ladders
  - Resolution, accuracy
  - Isotopes (de-isotoping, deconvolution)
  - Fragmentation patterns (CID, ECD, peptides vs. lipids, glycans, etc..)
  - MS of biomolecules
- Peptide ID methods
  - DB Search
    - perfect spectra, ion ladders
    - what can wrong: noise, missing ions, PTMs
  - DeNovo (DeNovo vs DB search)
  - Spectral libraries
  - Common mistakes
- Protein inference
  - Coverage, shared peptides, one-hit wonders
  - standard inference approaches
- Quantification
  - different strategies (labeled, unlabeled (in-vivo, in-vitro))
  - absolute vs relative
  - main algorithmic concepts (alignment, model fitting, peak integration)
- Validation/Error control
  - FDR, decoy DB
  - Quality control
- Interpretation of results

- Pathways enrichment
- How to extract biology of the data

- Data management/DBs
  - Repositories
  - Standard formats (mzML, etc.)
  - Metadata
- Strategies
  - Targeted vs non targeted
  - Top-Down, Bottom-up
- Experimental design (ST)
  - technical, biological replicate
  - variability, number of samples
  - multiple testing, pooling

Participants

- Rudolf Aebersold
ETH Zürich, CH
- Theodore Alexandrov
Universität Bremen, DE
- Dario Amodei
Stanford University, US
- Bernd Bodenmiller
Universität Zürich, CH
- Sebastian Böcker
Universität Jena, DE
- Karsten Boldt
Univ.-Klinikum Tübingen, DE
- Daniel R. Boutz
University of Texas – Austin, US
- Julia Burkhart
ISAS – Dortmund, DE
- Manfred Claassen
ETH Zürich, CH
- John Cottrell
Matrix Science Ltd. – London, GB
- Eric Deutsch
Institute for Systems Biology – Seattle, US
- Joshua Elias
Stanford University, US
- David Fenyö
New York University, US
- Anne-Claude Gingras
University of Toronto, CA
- Henning Hermjakob
EBI – Cambridge, GB
- Lukas Käll
KTH Royal Institute of Technology, SE
- Sangtae Kim
Pacific Northwest Nat'l Lab., US
- Oliver Kohlbacher
Universität Tübingen, DE
- Theresa Kristl
Universität Salzburg, AT
- Bernhard Küster
TU München, DE
- Henry Lam
The Hong Kong Univ. of Science & Technology, HK
- Wolf D. Lehmann
DKFZ – Heidelberg, DE
- Kathryn Lilley
University of Cambridge, GB
- Michal Linial
The Hebrew University of Jerusalem, IL
- Mike MacCoss
University of Washington – Seattle, US
- Brendan MacLean
University of Washington – Seattle, US
- Alexander Makarov
Thermo Fisher GmbH – Bremen, DE
- Lennart Martens
Ghent University, BE
- Sara Nasso
ETH Zürich, CH
- Alexey Nesvizhskii
University of Michigan – Ann Arbor, US
- Steffen Neumann
IPB – Halle, DE
- William Stafford Noble
University of Washington – Seattle, US
- Paola Picotti
ETH Zürich, CH
- Knut Reinert
FU Berlin, DE
- Bernhard Renard
RKI – Berlin, DE
- Hannes Röst
ETH Zürich, CH
- Stephen Tate
AB SCIEX – Concord, CA
- Andreas Tholey
Universität Kiel, DE
- Henning Urlaub
MPI für Biophys. Chemie – Göttingen, DE
- Olga Vitek
Purdue University, US
- Christian von Mering
Universität Zürich, CH
- Susan T. Weintraub
The University of Texas Health Science Center, US
- Witold E. Wolski
ETH Zürich, CH
- René Zahedi
ISAS – Dortmund, DE