Proteome Informatics Research Group (iPRG)

Educating the scientific community on best application and practice of bioinformatics toward accurate and comprehensive analysis of proteomics data.

CURRENT MEMBERSHIP

Viktoria Dorfer (Co-Chair) – *University of Applied Sciences Upper Austria*
Melanie Föll - *University Medical Center Freiburg, Germany*
Michael Hoopmann (Co-Chair) - *Institute for Systems Biology, Seattle, WA*
Pratik Jagtap - *University of Minnesota, Minneapolis, MN*
Magnus Palmblad - *Leiden University Medical Center, The Netherlands*
Yasset Perez-Riverol - *European Bioinformatics Institute, Hinxton, UK*
Susan T. Weintraub (EB Liaison) - *University of Texas Health Science Center at San Antonio*

VISIT:
https://abrf.org/research-groups/proteomics-metabolomics-mass-spectrometry/proteomics/

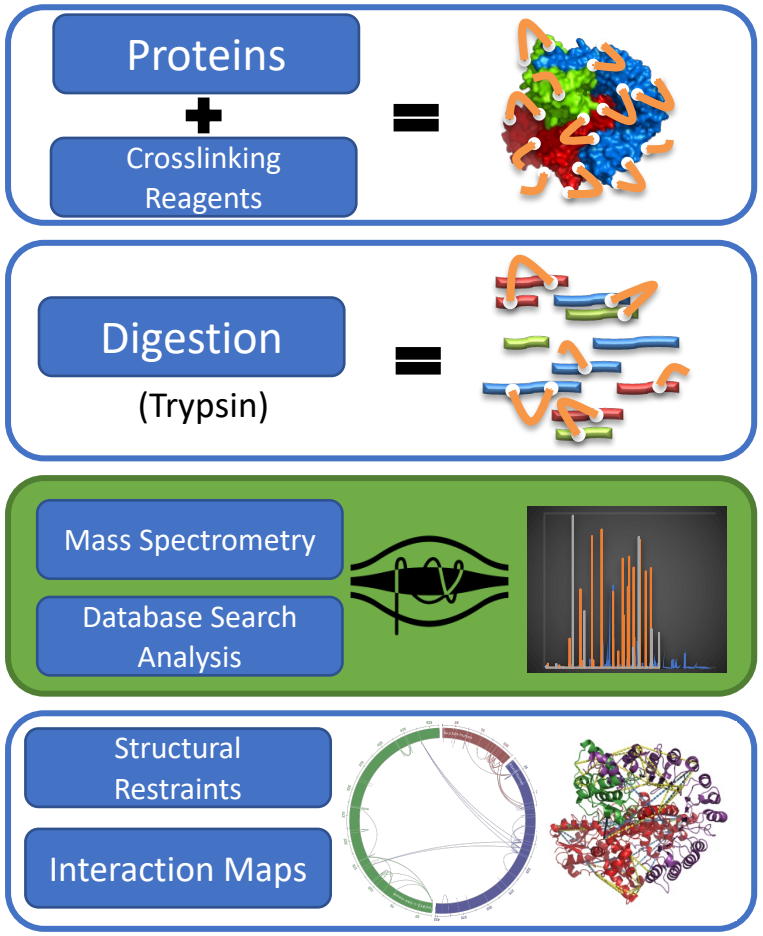
The iPRG

The mission of the ABRF iPRG is to educate ABRF members and the scientific community on the best application and practice of bioinformatics toward accurate and comprehensive analysis of proteomics data. The iPRG actively supports and participates in the development and advancement of new algorithms, software tools and strategies for proteome informatics with the goal of both educating and introducing these technologies to the membership.

2023: Crosslinking

2022: Cloud Computing Workshop (Online Now!)
2020: Metaproteomics
2016: Inferring Proteoforms from Bottom-up Proteomics Data
2015: Label-Free Quantitative Proteomics
2013: Using RNA-Seq Data to Refine Proteomic Data Analysis
2012: Detecting Modified Peptides in a Complex Mixture
2011: Identification of ETD Mass Spectra.
2010: Phosphopeptide Identification
2009: Differentially expressed proteins

Crosslinking Study



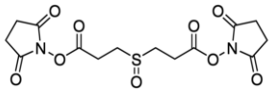
Specific Aims:

1. **Provide participants with crosslinking mass spectrometry datasets (with static or cleavable crosslinkers) along with example workflows and tutorials.**
2. **Collect analyzed datasets and highlight the differences and challenges unique to static or cleavable crosslinkers when used to study the same proteins.**
3. **Educate community on common practices for computational analysis of cross-linking mass spectrometry data.**

Crosslinking Study Details

The Study Design

Two protein mixture or
Three protein mixture



DSSO (cleavable)

MS2 method

Participants Receive

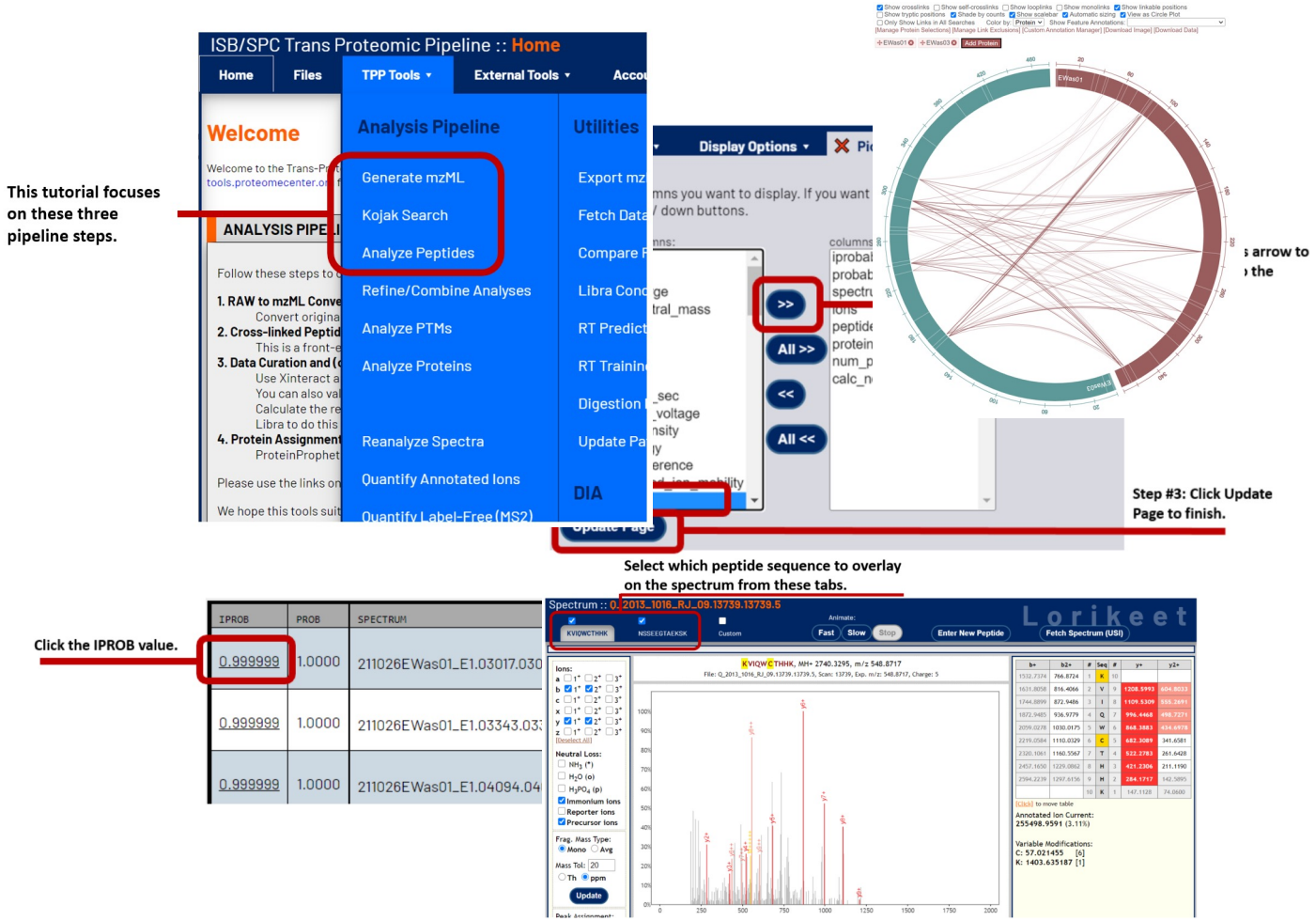
1. **Mass spectra in RAW and mzML format.**
2. **FASTA sequence file.**
3. **Crosslinker mass and structure information.**
4. **Generalized tutorial for the MS2-based analysis method.**
5. **Template form for results.**

Participants Return

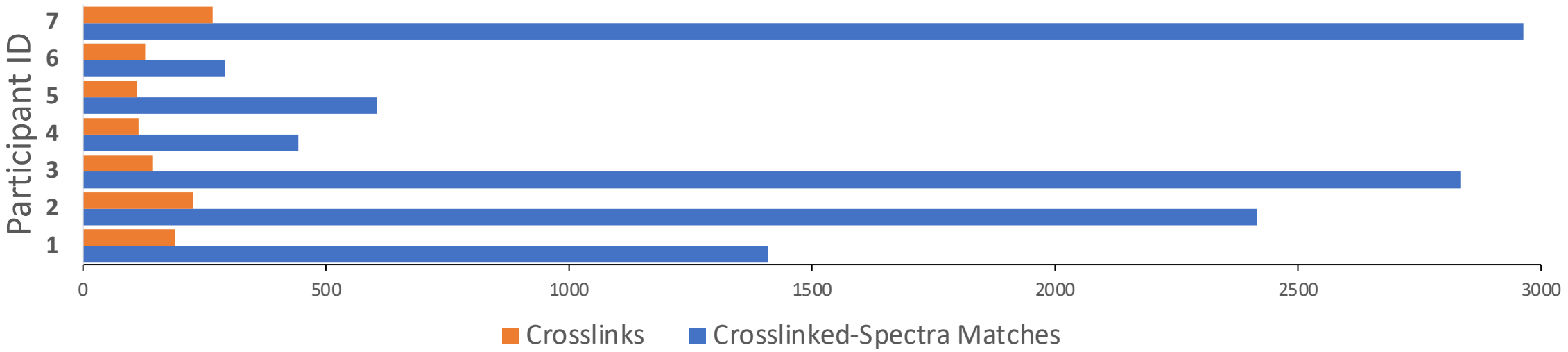
1. **List of proteins and the linked positions between them.**
2. **Statistical probability or error rate associated with each link interaction.**
3. **Computational software/pipeline and parameters used for static and cleavable crosslinker analyses.**

A Crosslinking Study For Everyone

- Study divided into two parts.
- Part 1: Participants given self-guided, graphical tutorials (Mascot, Kojak, or MS Annika) that instruct them on how to complete data analysis on half the study data.
- Part 2: Participants apply what they learned in Part 1 to the remaining data. They can use the same analysis pipelines, or explore new tools.
- No prior knowledge of crosslinking analysis is required to participate.

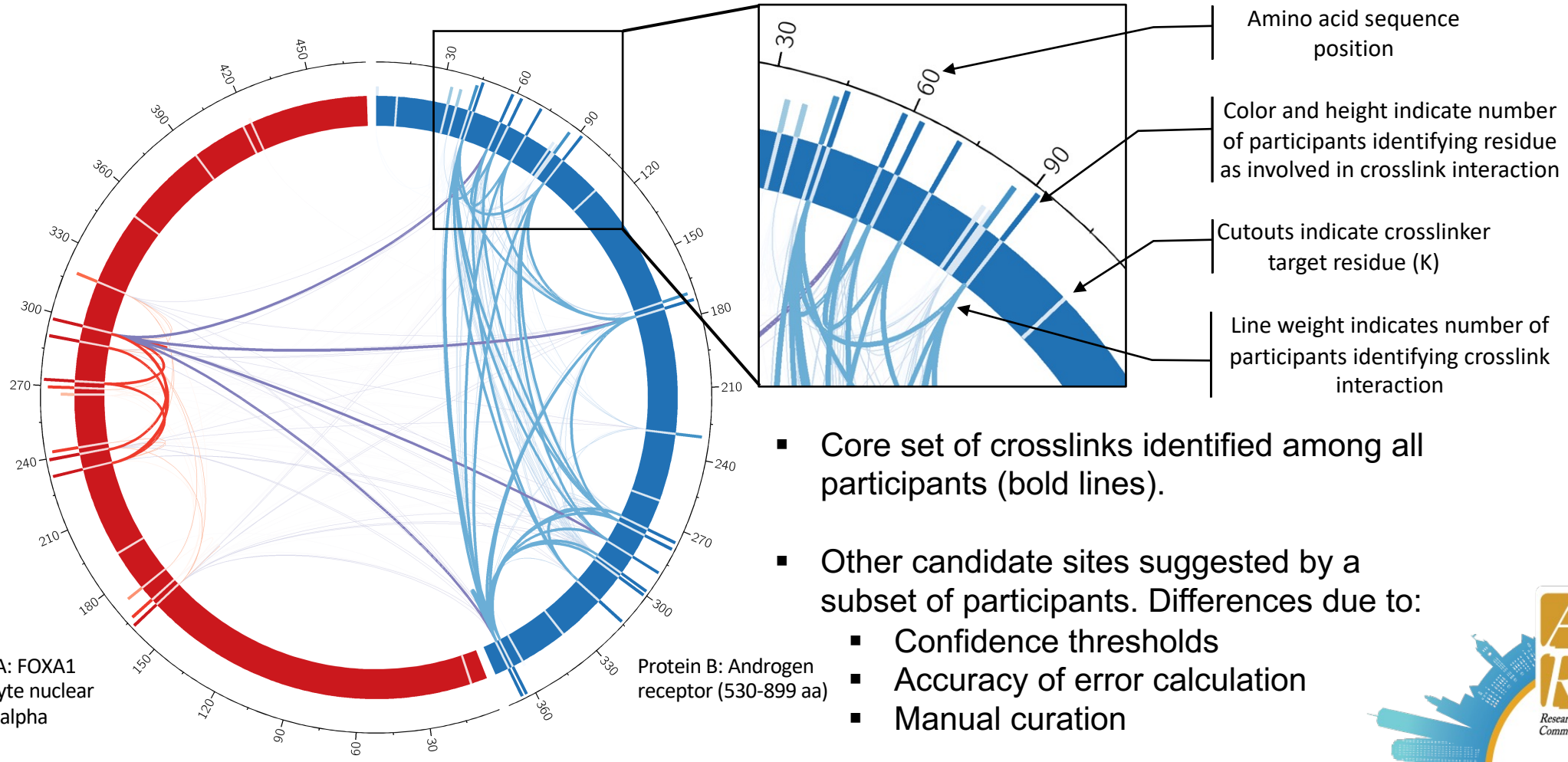


Summary of Study Participants



- Seven participants
- Seven different analytical pipelines
  - Different search algorithms + multiple search algorithms
  - Different validation algorithms
  - Comparative results at different error thresholds
  - Noticeable diversity in quantity of results

Preliminary Results: Phase 1



- Core set of crosslinks identified among all participants (bold lines).
- Other candidate sites suggested by a subset of participants. Differences due to:
  - Confidence thresholds
  - Accuracy of error calculation
  - Manual curation

Conclusions

- Participants were invited to take part in a two-part study on protein cross-linking.
- For Part 1, seven participants sent us seven different pipelines, which led to a diverse set of results. Despite the diversity, there was a core set of crosslinks everyone identified.
- Preliminary analysis Part 1: Only a single site of protein A interacts with protein B among the core set of identified crosslinks. We are investigating on why this core set stands out: most abundant? most solvent accessible regions? most representative alignment of proteins? etc.
- We plan to proceed on Phase 2 of the study, wherein participants apply what they learned in Part 1 to the remaining data.