# Australasian Computational Proteomics Community Meeting

### **Public document**

### Zoom link

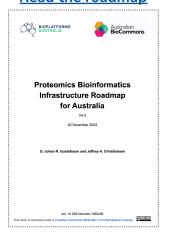
<u>Join the Google Group</u> to receive regular updates and automatic community meeting invitations.

### Why does this community meet?

This group meets so that members of the community can:

- Hear from other researchers about how they are approaching and leveraging computational proteomics
- Learn more about recent developments in bioinformatics that impact proteomics
- Share knowledge, make connections, and collaborate!
- Contribute to discussions that will inform how the BioCommons addresses bioinformatics challenges facing proteomics researchers
- Ask BioCommons anything!

### Read the roadmap



# **Tuesday 30th of September 2025**

11:00 am (AWST), 12:30 pm (ACST), 13:00 pm (AEST)

Chair: Johan Gustafsson (Australian BioCommons)

Minutes: Collectively recorded

### Zoom link

### In attendance

Please add your name and affiliation here		
Johan Gustafsson, BioCommons	Julia Broadbent, UoM	
Will Klare, APAF	Michela Mitchell, UoA and WCHN	
Ignatius Pang, APAF	Yuheng Zhu, USYD	
Narjis Fatima, Kolling	Ben Crossett, USyd	
Angela Juhasz, ECU		
Yaan Kit Ng, UoM		
Dr Shivalaxmi Manchukonda		

ltem	Item / Description
Welcome to meeting & intros Johan Gustafsson	An introduction to the <u>Australian BioCommons</u> The <u>Proteomics bioinformatics community</u>
Presentation	"User friendly proteogenomics with GenomeProt"  A/Prof Benjamin L Parker (University of Melbourne)  Goal: to fully understand the regulatory processes converting genotypes to phenotypes Proteomics used to improve annotation of protein coding genes (Jaffe et al. 2004) Building patient specific databases of expressed ORFs You could use standard or extended databases (e.g. OpenProt), but these may contain isoforms that aren't actually expressed Several proteogenomics tools exist Missing GUI Lack intuitive annotations required for downstream proteomics No built in visualisation GenomeProt workflow Database generation -> FASTA -> proteomics analysis -> integration Steps Generate custom proteogenomic database based on RNAseq Download FASTA and metadata files GenomeProt will map all ORFs to UniProt, then OpenProt, and create detailed FASTA headers along with detailed metadata Examples of potential ORFs detected with GenomeProt Protein coding IncRNA Perform proteomics / database search and upload results to the viewer

- Testing
  - o Mouse osteoblast cell line
  - Long read RNA-seq (nanopore)
  - GenomeProt database gen
  - Whole cell lysate -> multi-dimensional LC-MS/MS
  - Question: Impact of number of ORFs in the database?
    - #peptides and #proteins look similar, for all databases
    - As database gets larger, greater score cut off needed to achieve 1% FDR
    - Small database: more "sensitive", more proteins at lower scores but misses ORFs
    - Large database: less "sensitive", require higher scores, but gain in ORFs

#### Questions

 If I have a whole genome sequencing data, can I create a library with only sequences (between two tryptic digestion points) that have been mutated? I would like to use this mutated library to detect genetic mutations using proteomic.

•

#### **Community resources**

- Join the <u>Slack group</u>
- GitHub organisation <a href="https://github.com/Australian-Computational-Proteomics">https://github.com/Australian-Computational-Proteomics</a>
- Proteomics Learning Library Collection
  - Example <a href="https://australianbiocommons.github.io/Community-Library/structural\_biology">https://australianbiocommons.github.io/Community-Library/structural\_biology</a>
  - You can add resources to <u>this spreadsheet</u>
- Inviting international experts in applications of AI to mass spectrometry
  - ► Al in the life sciences Dr Tine Claeys and Dr Veit Schwämmle
  - DONE, revisit for next year as well?

### **Updates**

- Other BioCommons events, including webinars, workshops and community meetings
- ACF: <a href="https://australasiancorefacilities.org/">https://australasiancorefacilities.org/</a>
- WEBINAR SERIES: Al in the life sciences: Exploring possibilities, inspiring change <a href="https://www.biocommons.org.au/events/ai-series-2025">https://www.biocommons.org.au/events/ai-series-2025</a>
  - ► Al in the life sciences Dr Tine Claeys and Dr Veit Schwämmle
- WEBINAR SERIES: Leveraging deep learning to design custom protein-binding proteins <a href="https://www.biocommons.org.au/events/protein-binder-series">https://www.biocommons.org.au/events/protein-binder-series</a>
- APAF is currently hiring for a six-months fixed-term full-time bioinformatics role starting in January 2026.
   Please find the hyperlink to apply here:

Closing date of Wednesday 22<sup>nd</sup> of October at 11:59pm (AEST).

Macquarie Careers (internal)	<u>View Job Posting Details - Workday</u>
Macquarie Careers (external)	<u>Careers at Macquarie</u>
Seek	Analytical Technician Job in North Ryde, Sydney NSW - SEEK

LinkedIn	(24) Analytical Technician - Bioinformatics   Macquarie University   LinkedIn
Add your item(s) for discussion here Anyone can add an item here!	<ul> <li>Reminder: Bioinformatics Forum after this meeting</li> <li>Immediately after the meeting (2 pm AEST), the community will be able to virtually join a Bioinformatics Forum where A/Prof Matt Padula will talk about the "challenges of proteome and proteoform analysis, discuss the short-comings of current methodologies, and attempt to propose potential solutions". Zoom link</li> </ul>
The next meeting is in 2026!	

# Tuesday 22nd of July 2025 - cancelled

11:00 am (AWST), 12:30 pm (ACST), 13:00 pm (AEST)

Chair: Johan Gustafsson (Australian BioCommons)

Minutes: Collectively recorded

### In attendance

Please add your name and affiliation here		re
Johan Gustafsson, BioCommons		

Item	Item / Description
Welcome to meeting & intros Johan Gustafsson	An introduction to the <u>Australian BioCommons</u> The <u>Proteomics bioinformatics community</u> Community leadership and invitation to join the effort   Contact Johan if you would like to be involved in these discussions johan@biocommons.org.au
AusOMICS debrief	"AusOMICS debrief" Group discussion and round table, facilitated by Johan

- Key takeaways
- New developments?
- Challenges raised?

### Questions

•

### Community resources & Open discussion

- Join the <u>new Slack group</u>
- GitHub organisation <a href="https://github.com/Australian-Computational-Proteomics">https://github.com/Australian-Computational-Proteomics</a>
- Proteomics Learning Library Collection
  - $\circ \quad \textbf{Example} \ \underline{\textbf{https://australianbiocommons.github.io/Learning-Library/structural\_biology}}$
  - You can add resources to <u>this spreadsheet</u>
- Inviting international experts in applications of AI to mass spectrometry in progress

### Community feedback

- Resources required, and other challenges?
- •

Updates	<ul> <li>Other BioCommons events, including webinars, workshops and community meetings</li> <li>ACF: <a href="https://australasiancorefacilities.org/">https://australasiancorefacilities.org/</a></li> <li>WEBINAR SERIES: AI in the life sciences: Exploring possibilities, inspiring change <a href="https://www.biocommons.org.au/events/ai-series-2025">https://www.biocommons.org.au/events/ai-series-2025</a></li> <li>WEBINAR SERIES: Leveraging deep learning to design custom protein-binding proteins <a href="https://www.biocommons.org.au/events/protein-binder-series">https://www.biocommons.org.au/events/protein-binder-series</a></li> <li>New roadmap         <ul> <li>Michie, K. A., Litfin, T., Beecroft, S. J., Collins, B., Czabotar, P., Downton, M., Doyle, M. T., Ghosal, D., Grinter, R., Knott, G. J., Samaha, G., Christiansen, J. H., &amp; Gustafsson, O. J. R. (2025). Australian Structural Biology Deep-Learning Infrastructure Roadmap. Zenodo. <a href="https://doi.org/10.5281/zenodo.15786982">https://doi.org/10.5281/zenodo.15786982</a></li> </ul> </li> </ul>
Add your item(s) for discussion here Anyone can add an item here!	<ul> <li>Sanders, J., Wen, B., Rudnick, P.A. et al. A transformer model for de novo sequencing of data-independent acquisition mass spectrometry data. Nat Methods (2025). https://doi.org/10.1038/s41592-025-02718-y</li> </ul>

Next meeting is on 30th September 2025

# Tuesday 13th of May 2025

11:00 am (AWST), 12:30 pm (ACST), 13:00 pm (AEST)

### Zoom link

Chair: Johan Gustafsson (Australian BioCommons)

**Minutes:** Collectively recorded

### In attendance

Please add your name and affiliation here		
Johan Gustafsson, BioCommons	Desmond Li, SydneyMS, USYD	
	Keiran Rowell, UNSW	
Nikeisha Caruana, Unimelb	Julia Broadbent, Unimelb	
Rob Goode, CSIRO	Joel Steele, MPMP - Monash	
Caitlin Finney, Westmead Institute	Ignatius Pang, APAF	
Artur Shvetcov, Westmead Institute	Will Klare, APAF	
Kristina Gagalova, Curtin University		
Mailie Gall, Standardbiotools		

ltem	Item / Description
Welcome to meeting & intros Johan Gustafsson	An introduction to the <u>Australian BioCommons</u> The <u>Proteomics bioinformatics community</u> Community leadership and invitation to join the effort   Contact Johan if you would like to be involved in these discussions johan@biocommons.org.au
Presentation	"Proteomics with explainable machine learning: a pathway to experimental and clinical

#### translation"

Dr. Artur Shvetcov (Translational Dementia Research Group, Westmead Institute for Medical Research)

- CSF proteome of APOE4 carriers with and without impairment
- Alzheimer"s Disease Neuroimaging Initiative (ADNI) dataset
- https://adni.loni.usc.edu/
- See publication: Proteome profiling of cerebrospinal fluid using machine learning shows a unique protein signature associated with APOE4 genotype <a href="https://doi.org/10.1111/acel.14439">https://doi.org/10.1111/acel.14439</a>
- GNPC <a href="https://www.neuroproteome.org/">https://www.neuroproteome.org/</a>

#### Questions

- Does the plasma proteome give an indication of what effects or pathways are upregulated in the brain?
  - o Just showing main points from the study
  - Didn't show enrichments, which was significant work
  - Did find connections / similarities between plasma and CSF
- Advice for others that would like to apply ML methods?
  - Keep it simple
  - Focus on the question, then pick the method
  - Also focus on the end user? Is it biomarker development? Is it for clinical diagnostics? Something else?

# "Proteomics and Machine Learning-Based Approach to Decipher Subcellular Proteome of Mouse Heart"

Dr. David Greening (Baker Heart and Diabetes Institute)

- Publication: <a href="https://www.mcponline.org/article/S1535-9476(25)00050-7/fulltext">https://www.mcponline.org/article/S1535-9476(25)00050-7/fulltext</a>
  - o DOI https://doi.org/10.1016/j.mcpro.2025.100952
- Multiomics to illuminate EC vesicle biology
- Look at abundance and ultrasmall amounts
- Apply ML and other algorithms to understand complex biology in more detail
- Heart
  - Regional -> Cellular -> Subcellular
  - Covering limited profiling through to comprehensive mapping
  - Molecular spatial composition of heart is missing
- 450 markers associated with 16 subcellular niches
- How can this map be improved?
- Evaluated support vector machine (SVM), random forest (RF), extreme gradient boosting (XGB)
  - SVM: high dimensional space, memory efficient
  - RF: high accuracy, less overfitting, handles missing data, effective with a wide range of features
  - o XGB: high performance, flexible, handling missing values, limits overfitting
- Choice of method depends on data and task
- Model evaluation metrics -> accuracy, precision, recall, F1-score

#### Presentation

- Always ask: why are you using a model? What is it you are actually trying to do?
- Selected SVM and RF for the study

### Questions

- Features? Proteins and their localisation
- Number of n's?
  - o 6 hearts for dev of fractionation
  - Replicates across 2 broad groups (6 and 6)
  - o Validation in independent hearts using targeted experimental approaches
  - o Total is 24
  - o Age matched male mice
  - No injury, no pertubation
  - o Applying to multiple other groups now

#### Community resources & Open discussion

- Join the <u>new Slack group</u>
- GitHub organisation <a href="https://github.com/Australian-Computational-Proteomics">https://github.com/Australian-Computational-Proteomics</a>
- Proteomics Learning Library Collection
  - Example <a href="https://australianbiocommons.github.io/Learning-Library/structural\_biology">https://australianbiocommons.github.io/Learning-Library/structural\_biology</a>
  - You can add resources to this spreadsheet
- Inviting international experts in applications of AI to mass spectrometry in progress

#### **Community feedback**

- Resources required, and other challenges?
- To be discussed in detail at AusOMICS as well

Updates	<ul> <li>Galaxy Training Academy 2025 - happening this week!</li> <li>Other BioCommons events, including webinars, workshops and community meetings         <ul> <li>CO-WORKING: Sharing your bioinformatics workflows to make them findable and citable</li> </ul> </li> <li>AusOMICS 2025: <a href="https://www.ausomics.com/">https://www.ausomics.com/</a></li> <li>ACF: <a href="https://australasiancorefacilities.org/">https://australasiancorefacilities.org/</a></li> </ul>
Add your item(s) for discussion here Anyone can add an item here!	•

Next meeting is on 22nd July 2025

# Tuesday 4th of March 2025

10:00 am (AWST), 12:00 pm (AEST), 12:30 pm (ACST), 13:00 pm (AEDT)

### Zoom link

Chair: Johan Gustafsson (Australian BioCommons)

Minutes: Collectively recorded

### In attendance

Please add your name and affiliation here		
Johan Gustafsson, BioCommons	Clifford Young, UniSA	Abarna Murugan, APAF
William Klare, APAF	Muhammad (Maz) ALI, VCCRI, UNSW	Desmond Li, SydneyMS, USYD
Mahya Bahmani, CSIRO	Tiff Nelson, Australian BioCommons	Laura Dagley, WEHI
Ignatius Pang, APAF	Muhammad Zenaidee, APAF	Jumana Yousef, WEHI
Yunqi Wu, APAF	Jeff CHristiansen, BioCommons	
Rob Goode, CSIRO	Ruth Wang, APAF	
Prashanth Gouda, UWA(Student)	Michela Mitchell, WCHN	
Kristina Gagalova, Curtin University	Hossein V Kahrood, Monash Proteomics	
Nikeisha Caruana, Unimelb	Ben Goudey, BioCommons	
Nicholas DeBono (Institute for Biomedicine and Glycomics, Griffith University)	Joel Steele, Monash Proteomics and Metabolomics Platform	

ltem	Item / Description
Welcome to meeting & intros Johan Gustafsson	An introduction to the <u>Australian BioCommons</u> The <u>Proteomics bioinformatics community</u> Community leadership and invitation to join the effort

#### "From APAF to CSIRO, life as a proteomics enjoyer in an epigenetics group"

Dr Dana Pascovici, CSIRO, Macquarie University

- Group: Cell and Molecular Health
- Proteomics @ CSIRO
  - o MS capability in proteomics / metabolomics (Canberra, Brisbane, Melbourne)
  - Leverage internal expertise + connections to external Astral, Olink, Somascan etc.
- Multiomics why?
  - Samples can be precious
  - De-risk by having alternatives
  - Proteomics looks good / moves quickly 'golden age of plasma proteomics'
  - Boosting sensitivity of methylation assays with proteins?
- Case studies presented
  - Oesophageal cancer multiomics tissue and plasma cross-facility work (Monash, ProCan, AGRF)
  - Detection of disease in plasma
  - Saliva Alzheimer"s diagnostics https://doi.org/10.1002/alz.14420
- Multiomics how?
  - Still working on that one
  - Analysis side
    - Combining (i.e. mixOmics)
    - Side-by-side (e.g. WGCNA)
    - Data reduction and ML
    - 2 omics is more than twice as difficult as 1
  - Translation side
    - Validation / assay dev is expensive
    - Multiple techs in clinic is expensive
    - Makes sense to combine tech (e.g. proteomics, metabolomics)
- ML
- SWATH opened door to ML for low level analysis (library gen, protein ID)
- Diagnostics panels
  - Traditional ML (LASSO, elastic net, boosted trees, SVM) in variable selection / biomarker ranking
- Multiple cohorts, filtering / normalisation, variable selection, small panels + model dev
- BioCommons have a multi-omics community group, you can join this group here: https://groups.google.com/a/biocommons.org.au/g/multi-omics
  - We meet quarterly with our next meeting scheduled for Wednesday, 19th March,
     13:00 14:30 AEDT, see in your timezone, add to google calendar, add to calendar via .ics file. Join the google group to be invited.
  - Our recent meeting saw a presentation from Dr Ignatius Pang (also here at this meeting!) covering <u>Integrative multi-omics analysis for exploring rare disease</u> <u>mechanisms and biomarker discovery: a case study with Rett Syndrome</u>. You can <u>watch the recording here</u>.

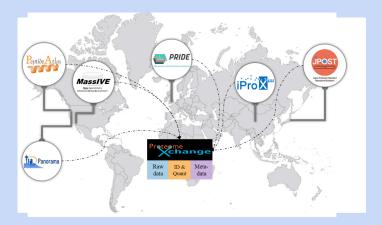
#### Questions

- Are we able to work as a group to define the term Multi-Omics? May be able to publish an opinion/definition article on this.
  - More than one omics?
  - Less about the omics, and more about the question? What are you hoping to get from a multiomics approach?
- Step change for proteomics as a result of ML / DL?
  - $\circ \quad \text{ Might still be coming, not here yet} \\$
  - o Can invite others working in this space as well to get their perspective
  - Maybe someone from ProCan?

#### Presentation

#### Community resources & Open discussion

- Join the <u>new Slack group</u>
- GitHub organisation <a href="https://github.com/Australian-Computational-Proteomics">https://github.com/Australian-Computational-Proteomics</a>
- Proteomics Learning Library Collection
  - Example <a href="https://australianbiocommons.github.io/Learning-Library/structural\_biology">https://australianbiocommons.github.io/Learning-Library/structural\_biology</a>
  - You can add resources to this spreadsheet
- Databases
  - Standardisation?
  - o ProteomicsDB (<a href="https://www.proteomicsdb.org/about">https://www.proteomicsdb.org/about</a>)
  - o PRIDE (<a href="https://www.ebi.ac.uk/pride/">https://www.ebi.ac.uk/pride/</a>)
  - o ProteomeXchange (<a href="https://proteomecentral.proteomexchange.org/">https://proteomecentral.proteomexchange.org/</a>)
  - MassIVE (https://massive.ucsd.edu/ProteoSAFe/static/massive.jsp)
  - o PanoramaWeb ( <a href="https://panoramaweb.org/home/project-begin.view">https://panoramaweb.org/home/project-begin.view</a>?)
  - Chorus ( <a href="https://chorusproject.org/pages/index.html">https://chorusproject.org/pages/index.html</a>)
  - o JPostDb (https://repository.jpostdb.org/entry/JPST002257)
  - o <a href="https://gen3.org/">https://gen3.org/</a>
  - MediaFlux



#### Community feedback

- 2022 Roadmap and 2024 forum report
- 2025-03-04 Revisiting proteomics roadmaps
- Resources required, and other challenges?

#### Updates

- Other BioCommons events, including webinars, workshops and community meetings
- AusOMICS 2025: <a href="https://www.ausomics.com/">https://www.ausomics.com/</a>
- ACF: <a href="https://australasiancorefacilities.org/">https://australasiancorefacilities.org/</a>
- AusOmics Workshops (including multi-omics R workshop), in-person booking full, online places still available: <a href="https://www.ausomics.com/pre-conference-workshops">https://www.ausomics.com/pre-conference-workshops</a>

	Next meeting 13th May
Add your item(s) for discussion here Anyone can add an item here!	•
Next meeting is on 13th May 2025	

# Monday 2nd of December 2024

10:30 am (AWST), 11:30 pm (AEST), 12:00 pm (ACST), 12:30 pm (AEDT)

# **Zoom link**

Chair: Johan Gustafsson (Australian BioCommons)

Minutes: Collectively recorded

### In attendance

Please add your name and affiliation here		
Johan Gustafsson, BioCommons	Tiff Nelson, Australian BioCommons	
Ziad Al Bkhetan, BioCommons	Giorgia Mori, Australian BioCommons	
Matt Padula, UTS.	Keiran Rowell, Structural Biology Facility, UNSW	
Will Klare, APAF, MQ	Georgie Samaha, SIH, USyd	
Nikeisha Caruana, Unimelb		
Michela Mitchell, Womens and Childrens Hospital		
Muhammad Zenaidee, APAF, MQ		
Joel Steele, Monash, MPMP	Monash Proteomics and Metabolomics Platform	
Stephanie Town, UTS		

ltem	Item / Description
------	--------------------

Welcome to meeting & intros Johan Gustafsson	An introduction to the <u>Australian BioCommons</u> <u>Proteomics bioinformatics community</u>
Presentation	"Folding entire proteomes - get a molecular inventory of your organism"  Dr. Keiran Rowell, Structural Biology Facility UNSW  Alphafold2  Using the tools, not necessarily how they work Novel organism -> sequencing -> folding / prediction Models in chemical biology - function follows chemical form Calculated structures are increasing and are now predominant in databases (i.e. PDB) The "XFold" programs (not just AlphaFold) Does not include physics Co-evolutionary correlations (i.e. residues that mutate together are likely closely associated structurally) Evoformers: AlphaFold, OpenFold, RoseTTAfold, AbFold, and others! Protein-LMs: ESMFold, ProtTrans Structural alignment: FoldSeek, Foldmason Need GPUs to run this software / code NVIDIA and AMD Folding proteomes N = 613 proteins -> 8 hours using ESMFold on GPU (22 days using AlphaFold) N = 4622 proteins -> 10 days using ESMFold on GPU (setimate of 2 years using AlphaFold) N = 4622 proteins -> 10 days using ESMFold on GPU (estimate of 2 years using AlphaFold) N = 4622 proteins -> 10 days using ESMFold on GPU (setimate of 2 years using AlphaFold) Australian AlphaFold service <a href="https://www.biocommons.org.au/alphafold">https://www.biocommons.org.au/alphafold</a> At UNSW: dev / using nf-core proteinfold workflow: <a href="https://inf-core/proteinfold/10.0/">https://inf-core/proteinfold/10.0/</a> Open On Demand (OOD) for proteinfold workflow: <a href="https://inf-core/proteinfold/10.0/">https://inf-core/proteinfold/10.0/</a> Open On Demand (OOD) for proteinfold workflow: <a href="https://inf-core/proteinfold/10.0/">https://inf-core/proteinfold/10.0/</a> Open On Demand (OOD) for proteinfold workflow: <a href="https://inf-core/proteinfold/10.0/">https://inf-core/proteinfold/10.0/</a> Open On Demand (OOD) for proteinfold workflow: <a href="https://inf-core/proteinfold/10.0/">https://inf-core/proteinfold/10.0/</a> Open On Demand (OOD) for proteinfold workflow: <a href="https://inf-core/proteinfold/10.0/">https://inf-core/proteinfold/10.0/</a> Open On Demand (OOD) for

	<ul> <li>Questions</li> <li>How do you see this scaling in future? Given areas like proteo-genomics and the thousands of organisms being sequenced</li> <li>How do you determine if the structures are bioactive? Possible?         <ul> <li>Depends on the structure</li> <li>Annotations tell you what it's related to</li> <li>Will often be a variant of an existing protein</li> </ul> </li> <li>Full tertiary structures with AlphaFold? Yes!         <ul> <li>AlphaFold doesn't work well with point mutations</li> </ul> </li> <li>DBs for non-model organisms and structures - to begin working out structure-function relationships?         <ul> <li>Single point mutations example - AlphaMissense for Human only (training on highly curated human and primate data)</li> <li>Not for non-model organisms</li> </ul> </li> </ul>
Presentation	<ul> <li>"ABLeS and proteomics"</li> <li>Dr. Johan Gustafsson, Dr. Ziad Al Bkhetan, Australian BioCommons</li> <li>https://www.biocommons.org.au/ables</li> <li>Main motivation: difficulty of accessing high end compute for bioinformatics</li> <li>3 aspects: <ul> <li>Reference data generation,</li> <li>Production bioinformatics,</li> <li>Software accelerator</li> </ul> </li> <li>Questions <ul> <li>Questions</li> </ul> </li> </ul>
Proteomics Learning Library Collection	<ul> <li>Example <a href="https://australianbiocommons.github.io/Learning-Library/data_management">https://australianbiocommons.github.io/Learning-Library/data_management</a></li> <li>You can add resources to <a href="mailto:this.spreadsheet">this.spreadsheet</a></li> </ul>
Updates	<ul> <li>Join the Slack group</li> <li>Other events, including webinars, workshops and community meetings</li> <li>AUS-oMicS 2025 abstract submission open <a href="https://www.ausomics.com/abstracts">https://www.ausomics.com/abstracts</a></li> </ul>
Add your item(s) for discussion here	AusOMICS event? Discuss offline
	Novt mooting is in 2025

Next meeting is in 2025 Happy Holidays!

# Friday 16th of August 2024

10:30 am (AWST) 12:00 pm (ACST) 12:30 pm (AEST)

# Zoom link

Chair: Johan Gustafsson (Australian BioCommons)

Minutes: Collectively recorded

### In attendance

Please add your name and affiliation here		
Johan Gustafsson, BioCommons	Will Klare, APAF, Macquarie University	Ziad Al-Bkhetan, Australian BioCommons
Murtada D. Naser	University of Basra, Iraq	nasergriffith@gmail.com
Amaal Gh. Yasser	University of Basra, Iraq	a.ghaziyasser@gmail.com
ELA HAVARY	ISTANBUL BILGI UNIVERSITY	alaa.bilgi@gmail.com
James Chan	SA Health	james.chan@sa.gov.au
Dr Michela Mitchell, CF	Women's and Children's Health Network, SA	michela.mitchell@sa.gov.au
Magnus Palmblad	Leiden University Medical Center, The Netherlands	n.m.palmblad@lumc.nl
Jeff Christiansen	Australian BioCommons	
Ignatius Pang	APAF, Macq U	ignatius.pang@mq.edu.au
Andrew Isaac	Metabolomics Australia	aisaac@unimelb.edu.au
Harry Wang	Metabolomics Australia	harry.wang2@unimelb.edu.au
Nigel Ward	Australian BioCommons	
Alex Ip	AARNet	alex.ip@aarnet.edu.au
Melissa Burke	Australian BioCommons	melissa@biocommons.org.au
Matthew Downton	NCI	matthew.downton@anu.edu.au
Joel Steele	Monash Proteomics and Metabolomics Platform	joel.steele@monash.edu

Please add your name and affiliation here		
Johan Gustafsson, BioCommons	Will Klare, APAF, Macquarie University	Ziad Al-Bkhetan, Australian BioCommons
Murtada D. Naser	University of Basra, Iraq	nasergriffith@gmail.com
Amaal Gh. Yasser	University of Basra, Iraq	a.ghaziyasser@gmail.com
Georgie Samaha	SIH, University of Sydney	georgina.samaha@sydney.edu.au
Cali Willet	SIH, University of Sydney	cali.willet@sydney.edu.au
Samitha Amarapathy	Australian BioCommons	samitha@biocommons.org.au
Ben Crossett	SydneyMS, USyd	ben.crosssett@sydney.edu.au
Conrad Leonard	Australian BioCommons	
Jerry li		qiyerensheng@gmail.com
Riya Shrestha	UNSW Sydney	riya.shrestha@unsw.edu.au

ltem	Item / Description
Welcome to meeting & intros Johan Gustafsson	An introduction to the <u>Australian BioCommons</u> <u>Proteomics bioinformatics community</u> Welcome to the other communities joining today's meeting.
Presentation	"Workflomics" Magnus Palmblad, Leiden University Medical Center, Netherlands  Presenting Workflomics, "a platform used to help bioinformaticians to automatically generate and compare computational workflows." One of the key use cases for this platform is proteomics, and in this presentation you will have the opportunity to hear more about this platform and how it can be used to streamline workflow creation and benchmarking.
Questions	<ul> <li>What would be really great is to "close the loop" after benchmarking by storing the scores of the generated workflows so that next time someone wants to evaluate the parameter space they can see what worked well before         <ul> <li>Public benchmarks increasing efficiency - reducing random walk through parameter space</li> </ul> </li> <li>Working with the OpenEBench team, including literature mining etc.</li> <li>What about the individual parameter choices within each tool chosen for the workflows?         <ul> <li>Domain knowledge critical</li> <li>Reasonably good default inputs</li> </ul> </li> </ul>

Comparing default settings at the moment Click on the 2 "best looking" workflows and compare a few parameter changes Can the workflows be exported and run locally? • Yes, download Docker containers, in future also HPC compatible (maybe Singularity) Which data is being used? • User provides data for benchmarking, so this is tailored to your use case Some standard data available (i.e. ground truth) Also working on generating synthetic data for evaluating workflows What is the infrastructure behind the benchmarking? Eg a handful of thermos raw files can be a few GB, or in the genomics space, very large fastq pairs. Can run on national HPC in NL Benchmarking on big data - need bigger compute resources Should only benchmark on enough data Thinking about a link to Galaxy Workflomics available to use now? Yes! General Join the Slack group Join the community (for updates) Other events, including webinars, workshops and community meetings WORKSHOP: Online data analysis for biologists o Top Tips from Galaxy Australia WORKSHOP: Hello Nextflow! **Updates** Proteomics Lab update https://proteomics.usegalaxy.org.au/ ■ Proteomics Lab community working group • Proteomics Lab being presented at IMSC2024 next week! Tools and workflows Fragpipe on Galaxy • DIA-NN on Galaxy - if you have feedback, or something isn't working, please let us know https://site.usegalaxy.org.au/request **Proteomics training** We'll continue working together to collate a list of proteomics-themed learning resources that explain a concept or demonstrate how to perform a task. These should be resources that we already know about and would recommend to a friend or colleague. If you're extra keen or you can't join us you can add resources to this spreadsheet. Co-working This collection will ultimately be included in the Australian BioCommons' Learning Library. **Discussion notes** Example https://australianbiocommons.github.io/Learning-Library/data\_management

Action(s)	•
AOB?	Add your item(s) for discussion here
Next meeting is 18th November 2024	

# Monday 3rd of June 2024

10:30 am (AWST) 12:00 pm (ACST) 12:30 pm (AEST)

### In attendance

Please add your name and affiliation here		
Johan Gustafsson, BioCommons	Lisa Phippard, BioCommons	Melissa Burke, BioCommons
Asif Ahmed, CSIRO ACDP	Will Klare, APAF, Macquarie Uni	Uwe Winter, BioCommons
Maz Ali, VCCRI, UNSW	Cali Willet, USyd	Ilgnatius Pang, APAF Macquarie U
Joel Steele, Monash, MPMP	Georgie Samaha, USyd	Muhammad Zenaidee, APAF, Macquarie Uni
Denise Tran, USyd	Chingyu Lu, USyd	Adam Carroll, JMSF, ANU
Matthew Downton, NCI	Ziad Al-Bkhetan, BioCommons	Ben Crossett, SydneyMS, USyd
Hossein Valipour, Monash, MPMP	Fozia Masood, Macquarie University	

ltem	Item / Description
Welcome to meeting & intros Johan Gustafsson	An introduction to the Australian BioCommons The context for these meetings  • Australian BioCommons  • Proteomics bioinformatics community page  • Proteomics Analysis Infrastructure Roadmap for Australia
Presentation	"Scaling DIA-NN to HPC for accelerated large-scale proteomics" Cali Willet, Sydney Informatics Hub

	Dr. Cali Willet (Sydney Informatics Hub) will present her work to speed up DIA-NN analyses using high performance computing (HPC) at the National Computational Infrastructure (NCI), reducing processing time for large proteomics experiments from days / weeks to just hours.
Questions	<ul> <li>What were the hardware specifications i.e. Core count, Ram, Storage space and speed?         <ul> <li>Gadi normal nodes ~ 150,000 cores</li> <li>Could run with fewer cores as well</li> </ul> </li> <li>mzML effect - closer look at spectra? Do they look different after conversion? Explanation for the difference in results?         <ul> <li>Dev for ProteoWizard - issue</li> <li>Recommended software to check converted mzML files: SeeMS</li> <li>Converted mzML had same values as wiff EXCEPT for Number of data points: this was much lower for mzML than wiff</li> <li>Suggested that DIA-NN might be doing peak picking differently to ProteoWizard (issue available on GitHub repository)</li></ul></li></ul>
Updates	General  Join the Slack group  Join the community (for updates)  Other events, including webinars, workshops and community meetings  Gustafsson, O. J. R. (2024). FORUM Report: Connections in Computational Proteomics. Zenodo. https://doi.org/10.5281/zenodo.10929402  Proteomics Lab update https://proteomics.usegalaxy.org.au/  Proteomics Lab community working group  Tools and workflows  Fragpipe on Galaxy - major elements are being merged at the moment - JG will keep the group up-to-date  MSFragger part of this work: https://github.com/bioconda/bioconda-recipes/pull/45127  DIA-NN on Galaxy - if you have feedback, or something isn't working, please let us know https://site.usegalaxy.org.au/request
Workshop	Proteomics training Melissa Burke, Australian BioCommons  The proteomics bioinformatics community recognises the need to train more life scientists in proteomics. Collating and sharing collections of recommended training resources is one way that we might address this as a community. Melissa Burke, Australian BioCommons Training

	and Communications Officer will join us to facilitate a discussion on how we might create a collection of recommended proteomics training resources and share them with Australian life scientists. Everyone is welcome to join the discussion. No training experience is needed.  • <u>Discussion notes</u>	
Action(s)	•	
AOB?	Add your item(s) for discussion here	
Next meeting is 16th August 2024		

# Monday 11th of March 2024

10:30 am (AWST)

12:30 pm (AEST)

13:00 pm (ACDT)

13:30 pm (AEDT)

### In attendance

Please add your name and affiliation here			
Johan Gustafsson Australian BioCommons			
Winnie Mok Australian Biocommons			
Joel Steele Monash (MPMP)			

ltem	Item / Description	Action(s)
Welcome to meeting & intros Johan Gustafsson	An introduction to the Australian BioCommons The context for these meetings	
Presentation	"A summary of the FORUM: Connections in computational proteomic	s"

Johan Gustafsson	
Questions	•
General updates	Join the Slack group Join the community (for updates)  Other events, including webinars, workshops and community meetings  https://zenodo.org/communities/australianbiocommons-training/records?q=&l=list&p=1&s=10&sort=newes  https://www.youtube.com/c/AustralianBioCommonsChannel  UniProt survey  Are you using Uniprot in your work?  If yes, we would love to hear your thoughts as part of this survey designed by economists to assess its value:     https://ec.europa.eu/eusurvey/runner/use_value_asses_sment_of_UniProt  If you have industry contacts who use Uniprot, we would be delighted if you could forward the opportunity to them!  Add your item for discussion here
Services	ABLeS: <a href="https://www.biocommons.org.au/ables">https://www.biocommons.org.au/ables</a> Alphafold2: <a href="https://www.biocommons.org.au/alphafold">https://www.biocommons.org.au/alphafold</a> Training infrastructure: <a href="https://www.biocommons.org.au/training-infrastructure">https://www.biocommons.org.au/training-infrastructure</a> Galaxy Australia: <a href="https://www.biocommons.org.au/galaxy-australia">https://www.biocommons.org.au/galaxy-australia</a>
Community activities	Proteomics Lab update <a href="https://proteomics.usegalaxy.org.au/">https://proteomics.usegalaxy.org.au/</a> Proteomics Lab community working group  Tools and workflows  DIA-NN on NCI  Fragpipe on Galaxy - major elements are being merged at the moment - JG will keep the group up-to-date  MSFragger part of this work: <a href="https://github.com/bioconda/bioconda-recipes/pull/45">https://github.com/bioconda/bioconda-recipes/pull/45</a> 127  DIA-NN on Galaxy - if you have feedback, or something isn't working, please let us know

	<ul> <li>https://site.usegalaxy.org.au/request</li> <li>Pathway mapping - this work will continue early next year https://australianbiocommons.github.io/pathway mapping to ols/</li> </ul>		
Next meeting topic - sharing training resources	The proteomics bioinformatics community recognises the need to train more life scientists in proteomics. Collating and sharing recommended training resources is one way that we might address this as a community.  Our next community meeting will focus on how we might collectively address this challenge, and help make high quality proteomics training resources more visible to the Australian life scientists. Melissa Burke, Australian BioCommons Training and Communications Officer will join us to facilitate the discussion.  Everyone is welcome to join the discussion. No training experience is needed.		
Add your item here!			
AOB?	•		
Next meeting is 3rd June 2024			

# Monday 13th of November 2023

10:30 am (AWST) 13:00 pm (ACDT) 13:30 pm (AEDT)

### In attendance

Please add your name and affiliation here			
Johan Gustafsson Australian BioCommons	Luke Carroll APAF	Pawel Sadowski QUT-CARF	
Jairo A. Pinzon-Cortes Monash University	Matthew Downton NCI		
Michael Dausmann Children's Medical	Cali Willet Sydney Informatics Hub		

Research Institute	University of Sydney	
Joel Steele (MPMP) Monash University	Ben Crossett SydneyMS, USyd	
Ignatius Pang Children's Medical Research Institute	Matt Padula UTS	
Dorothy Loo Oey Translational Research Institute, Qld	Clifford Young University of South Australia	

ltem	Item / Description	Action(s)
Welcome to meeting & intros Johan Gustafsson	An introduction to the Australian BioCommons The context for these meetings	
Presentation Ori Livson ProCan / CMRI	Scaling DIA-NN and Cross Cohort Data Exploration	
Questions	<ul> <li>SWON vs. Nextflow?         <ul> <li>Could Nextflow achieve the same thing as SWON?</li> <li>Running of workflows - yes</li> <li>Producing same DB - no</li> <li>If you have graph like requirements for exploring data, you won't get this with nextflow</li> <li>OL: It turns out Nextflow has SQL plugins now for managing workflow metadata but that didn't exist at the time ProCan developed swon.</li> <li>For QC, we hash files</li> <li>Bespoke system at the moment</li> </ul> </li> <li>POCKET aspect run in house - clinical data</li> <li>Possible to replicate the workflow approach?         <ul> <li>Not straightforward</li> <li>Not easy to set up, but technically feasible</li> </ul> </li> </ul>	

	<ul> <li>VIsualisation tools - compute heavy? What are they doing?</li> <li>Visualising data</li> <li>Example, high quality read access - large csv or MS data</li> <li>POCKET query engine - 100s of GB or RAM</li> <li>Lightweight, but high mem available</li> <li>Visualisation platforms lightweight in principle, with post-processing completed before hand if possible</li> <li>K8 with unrestricted resources are helpful for heavier jobs</li> </ul>
General updates	<ul> <li>Join the Slack group</li> <li>Join the community (for updates)</li> <li>Other events, including webinars, workshops and community meetings</li> <li>Add your item for discussion here</li> </ul>
Community activities	Proteomics Lab update <a href="https://proteomics.usegalaxy.org.au/">https://proteomics.usegalaxy.org.au/</a> <ul> <li>Tools list update (in progress)</li> <li>Centre page update (in progress)</li> <li>Proteomics Lab community working group</li> </ul> Tools and workflows <ul> <li>DIA-NN on Galaxy - if you have feedback, or something isn't working, please let us know <a href="https://site.usegalaxy.org.au/request">https://site.usegalaxy.org.au/request</a></li> </ul> Pathway mapping - this work will continue early next year <a href="https://australianbiocommons.github.io/pathway_mapping_to_ols/">https://australianbiocommons.github.io/pathway_mapping_to_ols/</a>
FORUM: Connections in computational proteomics 30-31 Jan 2024	Computational proteomics forum in Melbourne on 30th and 31st of January - right before Lorne Proteomics 2024  If you would like to attend, please register
Plan for 2024	<ul> <li>Quarterly meetings will continue</li> <li>Will continue working on Proteomics Lab - if you would like to join the working group, let Johan know</li> <li>Proteomics Lab community working group</li> </ul>

Add your item here!			
AOB?	•		
Next meeting is in 2024 Happy Holidays!!!			

# Monday 21st of August 2023

11:30 am (AWST) 13:00 pm (ACDT) 13:30 pm (AEST)

### In attendance

Please add your name and affiliation here			
Johan Gustafsson Australian BioCommons	Matt Padula UTS		
Matthew Downton NCI	Nikeisha Caruana, Unimelb		
Luke Carrol APAF			
Rohan Lowe, La Trobe U			
Clifford Young, UniSA			

ltem	Item / Description	Action(s)
Welcome to meeting & intros Johan Gustafsson	An introduction to the Australian BioCommons The context for these meetings	
<b>Presentation #1</b> Nikeisha Caruana	"RDMassSpec Explorer: An online platform for the interrogation of rare disease proteomics data."	
Questions	•	

<b>Proteomics Lab</b> Johan Gustafsson	"Proteomics Lab: <a href="https://proteomics.usegalaxy.org.au/">https://proteomics.usegalaxy.org.au/</a> " Also inviting community members to join a working group that will direct and co-create content for Proteomics Lab.  •	
General updates	<ul> <li>Proteomics BoF outcome Galaxy Community Conference 2023 (Brisbane)         <ul> <li>GCC 2023 Global Galaxy MS planning</li> </ul> </li> <li>Survey on MS imaging standards being conducted by https://www.imagingmssociety.org/</li> <li>Other events, including webinars, workshops and community meetings</li> <li>PMV conference: 25th October 2023         <ul> <li>Together with ECR conference</li> <li>Free event</li> <li>Registration open until close to conference</li> </ul> </li> </ul>	
Community activities	Training  WEBINAR: Getting started with proteomics  Event possibilities before Lorne 2024  First community F2F meeting in Melbourne?  Open discussion about this idea  Tools and workflows  Pathway mapping https://australianbiocommons.github.io/pathway mapping to ols/  © 2023 Evaluation of pathway mapping software  Fragpipe workflows  DIA-NN installed on Galaxy Australia Has anyone used the following tools?  Prolfqua https://doi.org/10.1021/acs.jproteome.2c00441  Einprot https://doi.org/10.1101/2023.07.27.550821	Action Johan: summarise plan for Lorne and send to group for feedback
Add your item here!		
AOB?	•	
Next meeting is 13th November 2023		

# Monday 29th of May 2023

10:30 am (AWST) 12:30 pm (AEST) 13:00 pm (ACDT) 13:30 pm (AEDT)

### In attendance

Please add your name and affiliation here		
Johan Gustafsson Australian BioCommons		
Joel Steele, Monash (MPMP)		
Matt Padula - UTS.		
Keiran Rowell - MQ		
Clifford Young (UniSA)		
Dorothy Loo-Oey (TRI)		

### Chair

Johan Gustafsson (Australian BioCommons)

### Minutes

Collectively recorded

ltem	Item / Description	Action(s)
Welcome to meeting & intros Johan Gustafsson	An introduction to the Australian BioCommons  The context for these meetings  BioCommons: <a href="https://www.biocommons.org.au/">https://www.biocommons.org.au/</a> Proteomics community: <a href="https://www.biocommons.org.au/proteomics">https://www.biocommons.org.au/proteomics</a> Proteomics Analysis Infrastructure Roadmap for Australia:	

	https://doi.org/10.5281/zenodo.7460248
Questions	Please add questions here ::
Presentation Gareth Price Rohan Lowe	Introducing TlaaS – Training Infrastructure as a Service - for proteomics
Questions / open discussion	<ul> <li>Pre-computed results for environmental computing</li> <li>Reducing the unnecessary expenditure of compute cycles</li> <li>How to teach based on mistakes, when pre-computed results are available?</li> </ul>
Updates	Community activities  WEBINAR: Getting started with proteomics  Proteomics.usegalaxy.au Fragpipe workflows  CONFERENCE: Galaxy Community Conference 2023 (Brisbane) Proteomics BoF together with Galaxy-Proteomics from US  TRAINING: Galaxy Training Network Smorgasbord 2023 (Online)  MEETING: Australian metabolomics community  Other events, including webinars, workshops and community meetings
Add your item here!	
AOB?	Pre Conference workshop possibilities? For Lorne 2024?
Next meeting is 29th May 2023	

# Monday 6th of March 2023

10:30 am (AWST) 12:30 pm (AEST) 13:00 pm (ACDT)

13:30 pm (AEDT)

### In attendance

Please add your name and affiliation here		
Johan Gustafsson Australian BioCommons	Ali Raza	Ralf Schittenhelm
Jeff Christiansen	Luke Carroll	Dorothy Loo-Oey
Harry Wang	Igy Pang	Rob Goode
Hailey Zhang	Ben Crossett	Amar Inder
Clifford Young	Pawel Sadowski	Matt Padula
Shanice	Tara Pukala	

# Chair

Johan Gustafsson (Australian BioCommons)

### Minutes

Collectively recorded

ltem	Item / Description	Action(s)
Welcome to meeting & intros Johan Gustafsson	An introduction to the Australian BioCommons The context for these meetings  BioCommons: <a href="https://www.biocommons.org.au/">https://www.biocommons.org.au/</a> Proteomics community: <a href="https://www.biocommons.org.au/proteomics">https://www.biocommons.org.au/proteomics</a> Proteomics Analysis Infrastructure Roadmap for Australia: <a href="https://doi.org/10.5281/zenodo.7460248">https://doi.org/10.5281/zenodo.7460248</a>	
Questions	Please add questions here	
<b>Presentation</b> Hailey Zhang	Phospho-Analyst: visualising and analysing phosphoproteomics data with "one-click"	

	https://analyst-suites.org/apps/phospho-analyst/
Questions / open discussion	<ul> <li>When will PathwayAnalyst be available?         <ul> <li>Being integrated with Rshiny now</li> </ul> </li> <li>Is it available on Galaxy?         <ul> <li>Not yet, hopefully it will be</li> </ul> </li> </ul>
Updates	<ul> <li>Looking for speakers for the next community meeting (29th May 2023)!</li> <li>We need your help! Help us improve the ToolFinder service by chatting with us about your experience: http://tinyurl.com/finding-tools</li> <li>Do you struggle to obtain data from international repositories? Let us know what infrastructure would make your life easier</li> <li>WORKSHOP: Make your bioinformatics workflows findable and citable</li> <li>CONFERENCE: Galaxy Community Conference 2023 (Brisbane)</li> <li>TRAINING: Galaxy Training Network Smorgasbord 2023 (Online)</li> <li>Other events, including webinars, workshops and community meetings         <ul> <li>MaxQuant training</li> </ul> </li> </ul>
Tool availability	<ul> <li>MSfragger on Galaxy</li> <li>Meeting with Fragpipe Galaxy dev team soon</li> </ul>
Add your item here!	
AOB?	
Next meeting is 29th May 2023	

# **Other Relevant Information**

### More information

- Visible Research Software Interest Group
  - o <a href="https://sites.google.com/ardc.edu.au/visible-research-software">https://sites.google.com/ardc.edu.au/visible-research-software</a>
  - o <a href="https://github.com/au-research/Visible-Research-Software/discussions">https://github.com/au-research/Visible-Research-Software/discussions</a>
- International Workflows Community Initiative: <a href="https://workflows.community/">https://workflows.community/</a>

- Research Software Engineer (RSE) Association of Australia and New Zealand
  - o https://rse-aunz.github.io/

### Australian BioCommons Community Engagement

- This community is one of several research community groups within the Australian
  BioCommons that we're engaging to understand the community-scale bioinformatics
  infrastructure challenges. It's our mission over the next few years to deliver a more coherent
  national infrastructure for Bioinformatics that will allow researchers in these communities to
  spend more time doing their research and less time on that bioinformatics infrastructure set
  up and management.
- The Australian BioCommons Project.
- The Australian BioCommons Code of Conduct
- Other **Communities** in the Australian BioCommons.
- BioCommons strategy documents:
  - Nelson, T. et al, March 2021, <u>The Australian BioCommons Community Engagement</u>
     <u>Strategy: Understanding community-scale challenges to inform solution delivery</u>, conference presentation, eResearch New Zealand.
  - Lonie, A. and Francis, R. 2020, <u>Australian BioCommons Strategic Plan 2021-2023</u>
- Example Infrastructure Roadmaps from other communities:
  - Nelson, T., Griffin, P. and Christiansen, J. 2020, <u>Genome Annotation Infrastructure</u> <u>Roadmap for Australia</u>.
  - Nelson, T. and Christiansen, J. 2020, <u>Genome Assembly Infrastructure Roadmap for Australia</u>.
- Hall, C. et al, 2021, <u>Application of a bioinformatics training delivery method for reaching dispersed and distant trainees</u>, PLoS Computational Biology, https://doi.org/10.1371/journal.pcbi.1008715
- Have something to say but didn't get the chance? Please contact us at communities@biocommons.org.au.