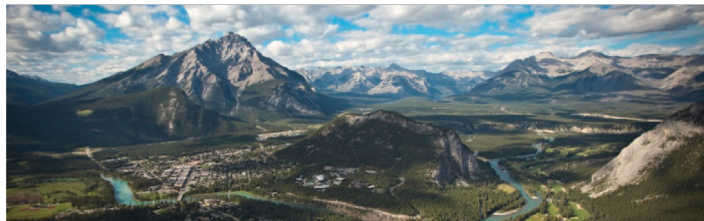


# *White paper discussions*



## *Mathematical Frameworks for Integrative Analysis of Emerging Biological Data Types*

June 15 - 19, 2020

Zoom from Banff International Research Station, Canada

[Aedin Culhane](#) (Dana-Farber Cancer Institute, Harvard TH Chan School of Public Health)

[Elana Fertig](#) (John Hopkins University)

[Kim-Anh Lê Cao](#) (University of Melbourne)

# Goals of this workshop white paper

- **Publication** opportunity to frame computational challenges and future directions of multi-omics.
- Overview of cutting-edge techniques for **integrating** diverse data from new technologies to provide a comprehensive multi-layer view of a biological systems.
- **Establish performance benchmarks** and assessment standards for the multi-omics integration research community.
- Deliver **open source set of resources** for multi-platform analysis with common datasets to benchmark method performance.
- Formulate **goals and future directions** to enable multi-omics research.

# Three hackathon multi-modal studies

- 3 data hackers
- 16 contributed talks focusing on analysis
- Data and GitHub code shared
- 5 keynotes setting up the scene
- Active community on Slack

This is **not** a regular virtual meeting  
How do we capitalize on these efforts?



# Format: Hybrid opinion piece, workshop summary, and hackathon analyses



HOME | ABOUT | ARCHIVE | SUBMIT | SUBSCRIBE | ADVERTISE | AUTHOR INFO | CON

## Hackathons as a means of accelerating scientific discoveries and knowledge transfer

Amel Ghouila<sup>1,9,10</sup>, Geoffrey Henry Siwo<sup>2,3,9</sup>,

Jean-Baka Domelevo Entfellner<sup>4,5,9,10</sup>, Sumir Panji<sup>6,9,10</sup>,

Katrina A. Button-Simons<sup>7</sup>, Sage Zenon Davis<sup>7</sup>, Faisal M. Fadlelmola<sup>8,10</sup>, The

DREAM of Malaria Hackathon Participants<sup>11</sup>, Michael T. Ferdig<sup>7</sup> and

Nicola Mulder<sup>6,10</sup>

[+](#) Author Affiliations

A screenshot of a BMC Genome Biology article page. The header includes the BMC logo and 'Part of Springer Nature'. The article title is 'Eleven grand challenges in single-cell data science' by David Lähnemann, Johannes Köster, and Alexander Schönhuth. It is published in Genome Biology, volume 21, article number 31 (2020). The page shows 22k accesses, 2 citations, and 276 Altmetric mentions. The article is published on 07 February 2020 and is available as Open Access. A navigation bar at the top includes links for Home, About, Articles, and Submission Guidelines. A blue banner at the bottom of the header area says 'We'd like to understand how you use our websites in order to improve our services for you.'

**BMC** Part of Springer Nature

## Genome Biology

Home About [Articles](#) Submission Guidelines

We'd like to understand how you use our websites in order to improve our services for you.

Review | [Open Access](#) | Published: 07 February 2020

### Eleven grand challenges in single-cell data science

[David Lähnemann](#), [Johannes Köster](#), [...] [Alexander Schönhuth](#)

*Genome Biology* **21**, Article number: 31 (2020) | [Cite this article](#)

**22k** Accesses | **2** Citations | **276** Altmetric | [Metrics](#)

External oversight; structured collaborations across institutions; Inspiring further challenges to tackle

# Resources for collaborative writing and contribution

## Participation



Zoom

## Communication



Slack

## Datasets, code, paper



Github



(details in BIRS page)

<http://www.birs.ca/events/2020/5-day-workshops/20w5197/schedule>

BIRSBioIntegration

Join the [#\\_themes](#)

<https://github.com/BIRSBioIntegration>

Demo by Casey Greene

 [#BIRSBioIntegration](#)

# What makes this white paper unique?

- Access to emerging data types
- Analyses and developments motivated by hackathons
- Complementary expertise across the group, highly multidisciplinary
- Spark new ideas for us and others
- Establish new vision for the 5 years to come
- Create new collaborations / opportunities to access new data sets
- Expand your international network without the jetlag (or, almost ...)

# Themes of the week orient the white paper

	Monday	Tuesday	Wednesday	Thursday	Friday
Theme	<b>seq-FISH</b>	<b>sc Targeted Proteomics</b>	<b>scNMT-seq</b>	<b>Computational Challenges: Interpretation, Common methods &amp; challenges, Benchmarking</b>	<b>Software Infrastructure</b>  <b>Future Directions</b>
Theme leaders (breakout room)	GC Yuan & Ruben Dries	Aedin Culhane & Olga Vitek	Oliver Stegle & Ricard Argelaguet	Susan Holmes  Casey Green & Kim-Anh Lê Cao  Michael Love & Matthew Ritchie	Vincent Carey  Elana Fertig

**Transparency**

**Collaboration**

**Open science**

**Fairness**

**Inclusion**

A unique opportunity for our trainees to learn the 'ropes' despite the publish or perish system.

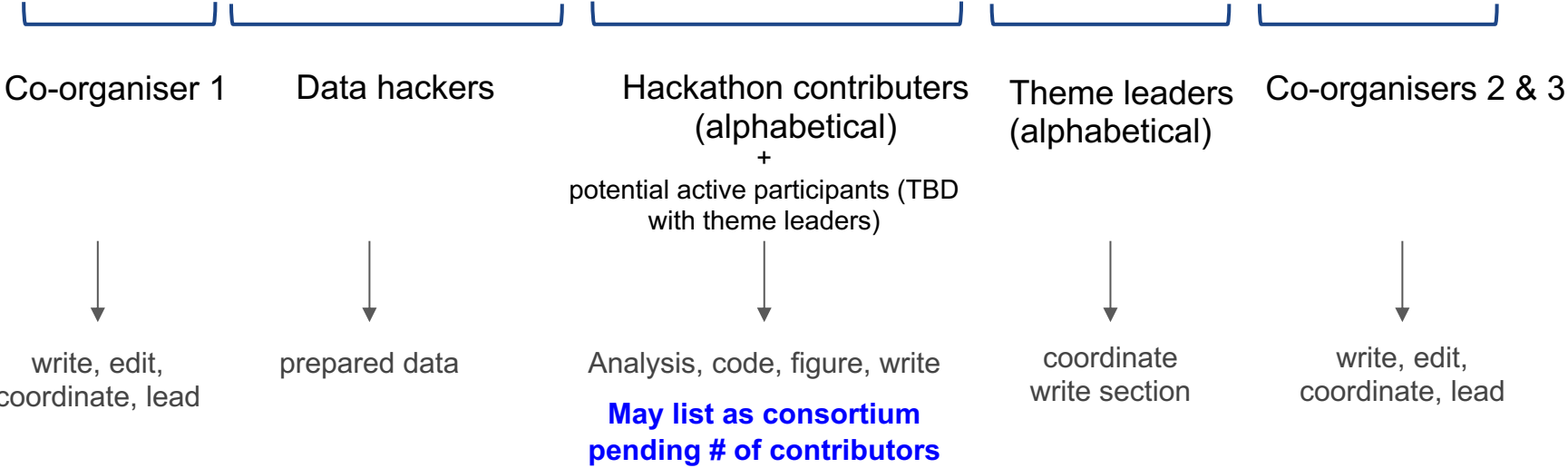


# Transform a brainstorming session into a digestible section

- Focus on three **high level** themes for your section
- One key figure (panel) or Table
- Discuss / coordinate with other theme leaders to avoid overlap (monitor Slack and shared doc). *Overshare* with others.
- Make sure everyone is heard (not necessarily the loudest!)
- Use gitHub for reproducible analyses + Manubot

# Co-authorship & roles

**Author A, Author B, ..... , Author Z**



A form will be sent **this Friday** to state your current contribution.  
The form will be sent a **second time** prior to submission to update your contribution

# Journals

We have approached:

1. **Genome Biology**: requires full submission
2. **PLoS Computational Biology**: requires full submission
3. **Cell Systems**: positive

Priority to:

- Open Access
- Featuring reproducible code

# Current state

## Four brainstorming sessions completed:

1. **Spatial**: #seqFish\_theme
2. **Targeted Proteomics**: #scTarg\_Proteomics\_theme
3. **RNA - DNA**: #scNMT-seq\_theme
4. **Summary of analytical methods**: #summary\_Analyses\_theme

Hackathons

## Four more to go:

1. **Interpretation challenges**: #interpretation\_theme
2. **Benchmarking**: #benchmark\_theme
3. **Software infrastructure**: software\_theme
4. **Future Directions**: #future\_theme

What's next  
(short term)

What's next  
(long term &  
vision)

# Current state: 4 brainstorming sessions completed

1. **Spatial**: Expt design, Platform Specific bias, Objective Assessment, Inclusion of spatial information
1. **Targeted Proteomics**: Normalisation, Partial feature overlap, Inherent spatial nature of biological data, from single cell to cell communities, atlases and maps for benchmarking
1. **RNA - DNA**: Binary data, DNA features summary, transfer learning or imputation using other atlases, Non-linear integration
1. **Summary of common challenges**: Non-overlapping features and/or cells, Incorporate prior knowledge, from data-driven towards mechanistic driven, Generic towards context specific methods

# Example for hackathon studies?

## **scRNA-seq + FISH as a case study for spatial transcriptomics**

### **Overview and biological question**

### **Computational challenges**

- Can scRNA-seq data be overlaid onto seqFISH for resolution enhancement
- What is the minimal number of genes needed for data integration?
- Are there signatures of cellular co-localization or spatial coordinates in the non-spatial scRNA-seq data?

### **Methods for stats/maths analyses and results summary**



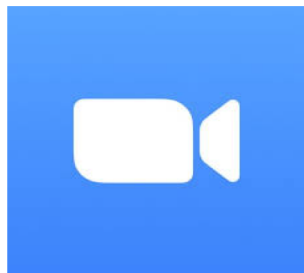
# Optimistic Timeline

- Week 1 (June 26): theme leaders push **outline** to Manubot
- Week 2 (July 3): **full section** written ( ~ 1 page + 1 Figure)
- Week 4 (July 17): **first draft** distributed to all for comments
- Week 6 (July 31): **comments back** from *all* co-authors
- Week 8 (August 14): finalise and submission

<https://birsbiointegration.github.io/whitePaper/>

# Communication will be key in the coming weeks!

Live



Zoom

Communication



Slack

Datasets, code, paper



Github



BIRSBioIntegration

<https://github.com/BIRSBioIntegration>

Monitor these tools and make good use of them!