

# Welcome!



## *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)

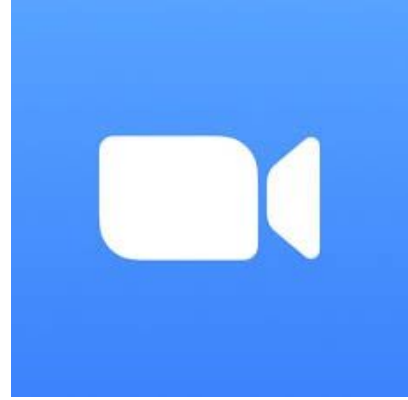
[Kim-Anh Lê Cao](#) (University of Melbourne)

***Workshop talks will be recorded & posted online***

***Live Stream*** <http://www.birs.ca/live>



2021?



# Emerging challenges in integrative sc'omics COVID edition

Bring in your best PJ, tracksuit pants and zoom background!

<http://www.birs.ca/live>



#BIRSBioIntegration

# Multi-omics Nature Methods Method of the Year

nature methods

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

Editorial | Published: 06 January 2020

## Method of the Year 2019: Single-cell multimodal omics

*Nature Methods* 17, 1(2020) | [Cite this article](#)

22k Accesses | 4 Citations | 120 Altmetric | [Metrics](#)

**Multimodal omics measurement offers opportunities for gaining holistic views of cells one by one.**

# What is this workshop?

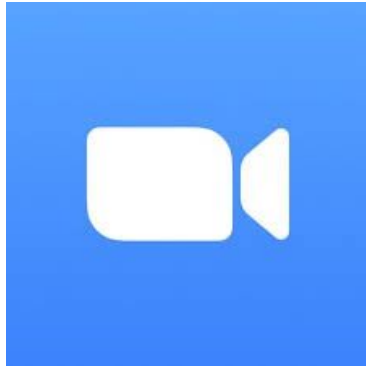
- **New technologies** yield new biological understanding at the RNA, DNA, cellular, and spatial resolution.
- **Integrating** diverse data from new technologies may provide a comprehensive multi-layer view of a biological system that cannot be obtained from single datasets.
- Multi-omics integration is an active area of research, but **lacks established performance benchmarks** and assessment standards.
- Bringing together **interdisciplinary computational scientists** will promote the development of cutting edge techniques for multi-omics.
- The workshop will deliver an **open source set of resources** for multi-platform analysis with common datasets to benchmark method performance.

# What should you expect at the workshop?

- Exciting set of **keynote speakers** discussing state of the art technologies, datasets, and analysis frameworks for multi-omics.
- Selected speakers featuring analysis on **curated** set of multi-omics data for standardized methods comparison.
- Opportunities to interact with experts in the field in small breakout brainstorm sessions.
- Brainstorm sessions on '**set of guidelines**' to develop and validate computational tools for cutting-edge biological data.
- **Publication** opportunity for white paper framing the computational challenges and future directions of multi-omics formulated at the workshop.

# Resources for virtual communication

## Participation



Zoom

## Communication



Slack

## Datasets, code, paper



GitHub



(details in BIRS page)

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

BIRSBioIntegration

<https://github.com/BIRSBioIntegration>

# Information about zoom sessions

- All talks will be **recorded** and made available through the BIRS website
- All talks will be **live-streamed** on <http://www.birs.ca/live>
- Questions are encouraged through use of the zoom chat and asked by the session chair
- Chairs will ring a **bell** for a 5 minute and 2 minute warning to help speakers track time
- Brainstorming sessions will be held in zoom breakout rooms and discussion will be encouraged through slack
- Everyone is encouraged to live Tweet the meeting

# Themes of the week

	<b>Monday</b>	<b>Tuesday</b>	<b>Wednesday</b>	<b>Thursday</b>	<b>Friday</b>
<b>Theme</b>	seq-FISH	sc Targeted Proteomics	scNMT-seq	Computational Challenges	Software Infrastructure   Future Directions
<b>Chair</b>	Elana Fertig	Aedín Culhane	Kim-Anh Lê Cao	Stephanie Hicks	Michael Love
<b>Theme leaders</b> (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



# Our keynote speakers



**Prof. GC Yuan**  
Dana-Farber Cancer Institute,  
Harvard TH Chan School of  
Public Health

**Mon 7.30am EST**



**Prof. Bernd  
Bodenmiller**  
University of Zurich

**Tues 7.30am EST**



**Prof. Oliver Stegle**  
German Cancer Research  
Center & EMBL

**Wed 7.30am EST**



**Prof. Susan  
Holmes**  
Stanford University

**Thurs 8am EST**



**Prof. Vincent Carey**  
Harvard Medical School,  
Brigham & Women's Hospital

**Fri 8am EST**

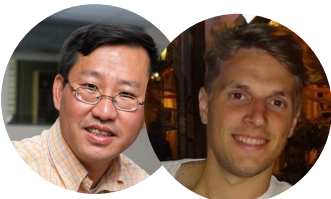
Full Schedule: <http://www.birs.ca/events/2020/5-day-workshops/20w5197>

 [#BIRSBioIntegration](https://twitter.com/BIRSBioIntegration)



# Brainstorming theme sessions: all welcome

More information on the [slack channels](#) 



seqfish\_theme

**Guo-Cheng Yuan &  
Ruben Dries**

Dana-Farber Cancer Institute,  
Harvard TH Chan School of Public  
Health & Boston University

**Mon 11am EST**



sc\_targ\_proteomics\_theme

**Aedin Culhane &  
Olga Vitek**

Dana-Farber Cancer Institute,  
Harvard TH Chan School of Public  
Health & Northeastern University

**Tues 10.30am EST**



scNMT-seq\_theme

**Ricard Arguelaget &  
Oliver Stegle**

German Cancer Research Center &  
EMBL

**Wed 10.30am EST**

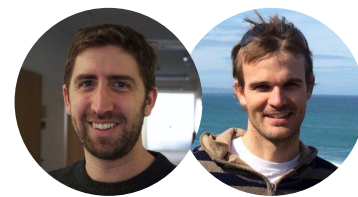


summary\_analyses\_theme

**Kim-Anh Lê Cao &  
Casey Green**

University of Melbourne & Uni  
Pennsylvania

**Wed 7pm EST**



benchmark\_theme

**Mike Love &  
Matt Ritchie**

University of North Carolina-Chapel Hill  
& Walter and Eliza Hall Institute

**Thurs 7pm EST**



**Susan Holmes**  
Stanford University

**Thurs 12pm EST**

interpretation\_theme



**Vincent Carey**  
Harvard Medical School and  
Brigham & Women's Hospital

**Fri 9am EST**

software\_theme



**Elana Fertig**  
Johns Hopkins University

**Fri 9am EST**

future\_theme

# White paper

- Hybrid opinion piece, workshop summary, and hackathon analyses.
- Manuscript aimed for Cell Systems, Genome Biology, or PLoS Computational Biology
- Manuscript sections mirrors brainstorming sessions and led by theme leaders.
- Manuscript drafted collaboratively with manubot.
- Analysis scripts must be in GitHub to be included in the manuscript.
- Workshop participants who actively contribute to analyses or themes will be added as co-authors.
- Criterion and commitment for authorship must be documented by theme leaders and recorded.

[github.com/BIRSBioIntegration/whitePaper](https://github.com/BIRSBioIntegration/whitePaper)

BIRSBioIntegration / whitePaper

Unwatch 5 Star 2 Fork 4

Code Issues 0 Pull requests 0 Actions Projects 0 Wiki Security 0

Insights

Manubot for white paper

339 commits 4 branches 0 packages 0 releases 1 environment 20 contributors View license

Branch: master - New pull request Create new file Upload files Clone or download - Find file

BIRSBioIntegration Merge pull request #4 from eifertig/patch-2 Latest commit a90bb72 14 days ago

.github/workflows	GitHub Actions: cache manubot files in ci/cache	3 months ago
build	upgrade manubot to fix webpage subprocess handling	2 months ago
ci	Export environment variables needed for gh-pages readme	3 months ago
content	Update metadata.yaml	14 days ago
output	GitHub Actions workflow for building and deployment	5 months ago
webpage	GitHub Actions workflow for building and deployment	5 months ago
.appveyor.yml	.appveyor.yml: note about skipping branches with PR	4 months ago
.gitignore	Dependency upgrade on 2019-06-03 with multiple ref file s...	12 months ago
LICENSE-CC0.md	Dual license code and data under CC0	3 years ago
LICENSE.md	Switch CC BY license to markdown	3 years ago
README.md	slight re-work of the readme	28 days ago
SETUP.md	Simplify setup by creating branches later	4 months ago
USAGE.md	metadata: use list for author.funders	2 months ago
screenshot_pull_reque...	Add files via upload	2 months ago

README.md

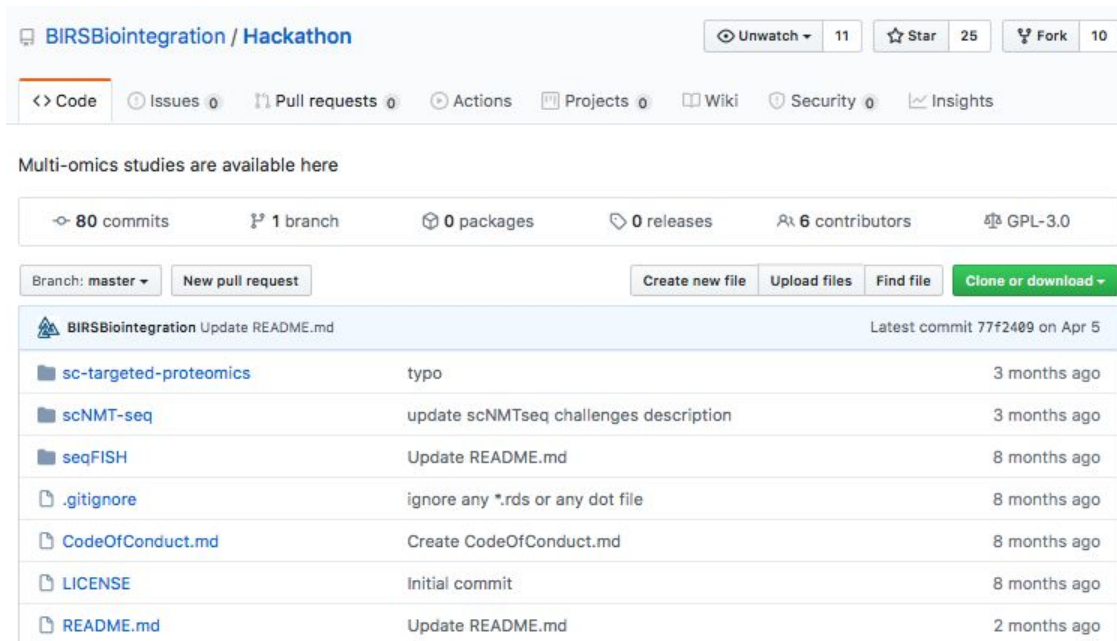
## White Paper for the Mathematical Frameworks for Integrative Analysis of Emerging Biological Data Types Workshop

Click on the following badges to access the mauscript:

About this Manuscript

# Datasets and analyses

- Easy access to 3 multi-omics studies and challenges
  - Cortex seq-FISH
  - Single cell proteomics
  - Gastrulation scNMT-seq
- Abstract speakers selected based on proposed analysis methods.
- Add your analysis code to the GitHub throughout the workshop <https://github.com/BIRSBioIntegration/Hackathon>
- Analyses will be highlighted in the white paper!



BIRSBioIntegration / Hackathon

Unwatch 11 Star 25 Fork 10

<> Code Issues 0 Pull requests 0 Actions Projects 0 Wiki Security 0 Insights

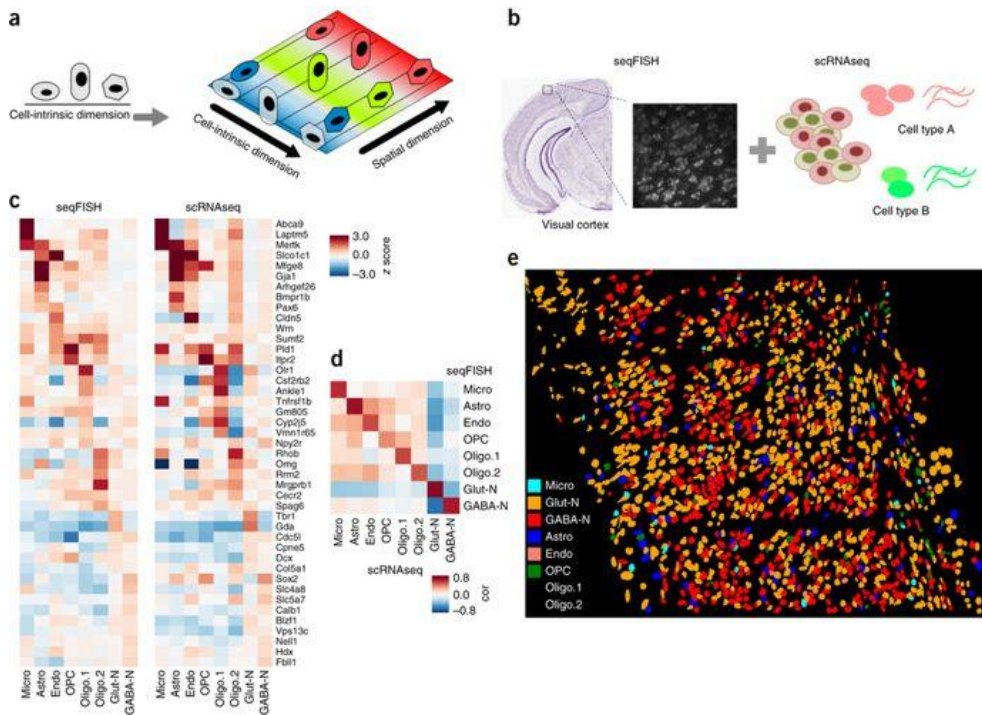
Multi-omics studies are available here

80 commits 1 branch 0 packages 0 releases 6 contributors GPL-3.0

Branch: master New pull request Create new file Upload files Find file Clone or download

BIRSBioIntegration Update README.md		Latest commit 77f2409 on Apr 5
sc-targeted-proteomics	typo	3 months ago
scNMT-seq	update scNMTseq challenges description	3 months ago
seqFISH	Update README.md	8 months ago
.gitignore	ignore any *.rds or any dot file	8 months ago
CodeOfConduct.md	Create CodeOfConduct.md	8 months ago
LICENSE	Initial commit	8 months ago
README.md	Update README.md	2 months ago

# Spatial transcriptomics: seqFISH + scRNA-seq



Spatial and non-spatial single cell transcriptional data for the adult mouse cortex

seqFISH data of 125 genes from Zhu et al 2018

scRNA-seq data for ~1,600 cells from Tasic et al.

Combined data enable cell type mapping with spatial information of non-spatial, high-resolution transcriptional data

# seqFISH: questions and challenges

- Identification of multi-omics signatures that characterise cell type, spatial region, or both.
- Can scRNA-seq data be overlaid onto seqFISH for resolution enhancement?
- Can higher resolution imaging-based transcriptional data be used to overcome missing data in scRNA-seq?
- What are the minimal number of genes needed for data integration?
- Are there signatures of cellular co-localization or spatial coordinates in non-spatial scRNA-seq data?

## Contacts

Emily Davis-Marcisak (Fertig lab): prepared the data

Guocheng Yuan provided the data

Elana Fertig

**Slack: #seqfish\_theme**

Easy access, guidelines:

<https://github.com/BIRSBioIntegration/Hackathon>

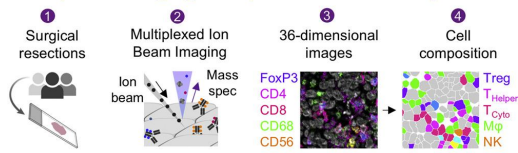
# Single-cell targeted proteomics across technologies

Data from 2 studies characterizing the **breast cancer tumor immune microenvironment**, using:

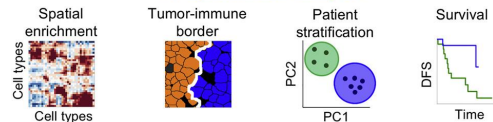
## Multiplexed ion beam imaging by time-of-flight (MIBI-TOF)

- 36 proteins
- 41 TN breast cancer patients

### Multiplexed imaging of 36 proteins in 41 TNBC patients

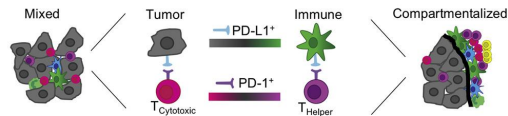


### Computational analysis



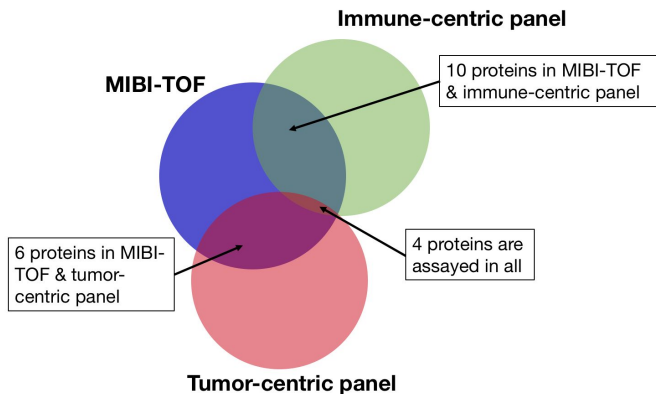
### Structured tumor-immune microenvironment in TNBC

Immune organization ↔ Histology ↔ Checkpoint expression



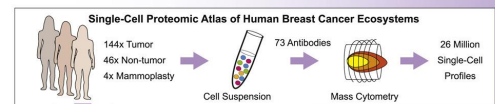
Keren, et al. (2018)

... with a total of 20 overlapping proteins



## Mass cytometry (aka mass tag, CyTOF)

- 73 proteins across 2 panels: tumor and immune
- 140 breast cancer patients, 6 TN



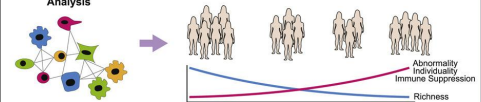
### Tumor Heterogeneity



### Immunoenvironment



### Ecosystem Relationship Analysis



Wagner, et al. (2019)

# Single-cell targeted proteomics across technologies

Some ideas on where to get started...

## Integrating data :::

- How should we approach integrating **partially-overlapping proteomic data** collected on different patients with similar phenotypes?
- Can we integrate **other 'omics datasets** (e.g., scRNA-seq) to support the results of these proteomic analyses?
- What additional information can we learn about the different **macrophage and immune populations** in breast cancer by conducting integrated analyses of these datasets?

## Spatial analysis :::

- Without including the spatial x-y coordinate data, how well can we predict **cell co-location**?
- Can we predict the **spatial expression patterns** of proteins measured on mass-tag but not measured in the MIBI-TOF data? There is dataset to be released in 2020, which may allow scoring of these predictions.

**Slack: #sc\_targ\_proteomics\_theme**

**Contacts:** Aedín Culhane and Lauren Hsu prepared these datasets  
*Data from Mendeley (Wagner 2019); Ionpath & Angelo Lab Websites (Keren, 2018)*

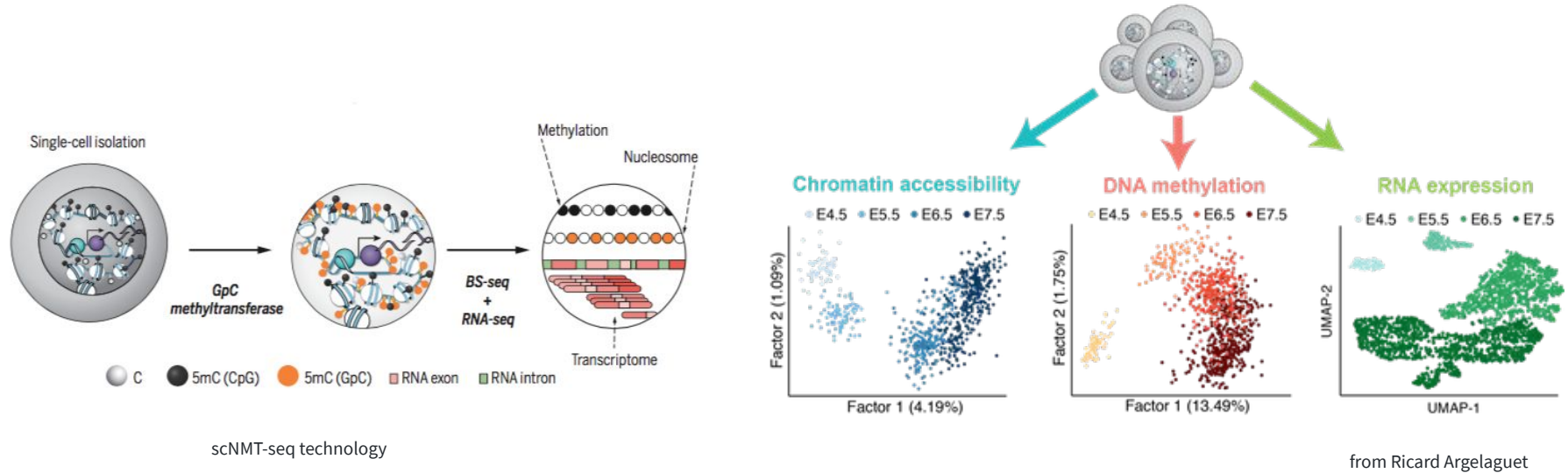
Easy access data and guidelines:  
<https://github.com/BIRSBioIntegration/Hackathon>



# scNMT-seq: transcriptome, DNA accessibility & DNA methylation

Gastrulation in mouse from [Argelaguet et al. 2019](#) (accepted for publication).

Total of 826 cells *matching* across all data sets



Various subsets of DNA methylation (5 data sets) and DNA chromatin accessibility (5 data sets) can be investigated as they have been summarised across specific genic regions (P300, CTCF and DHS)

# scNMT-seq: questions and challenges

## Types of analysis

- Identification of multi-omics signatures that characterise lineage, stage or both.
- Handling missing values
- Do epigenetic changes in some genomic contexts affect cell fate decision more than others and how?

## Challenges with epigenome data

- All epigenome data include a **large amount** of missing values (a proportion of 35-40%!).
- Methylome and chromatin accessibility data are generally far less predictive of cellular behaviour than transcriptome but are complementary to it. These data may not be able to highlight phenotypic heterogeneity, or identify outliers.

### Contacts

Al Abadi (Lê Cao lab): prepared the data

Ricard Argelaguet: provided the data

Kim-Anh Lê Cao

**Slack: #scnmt-seq\_theme**

Easy access, guidelines:

<https://github.com/BIRSBioIntegration/Hackathon>

On behalf of the fearless organizers, have a wonderful time!



**Aedín Culhane**  
Senior Research Scientist  
Data Sciences, Dana-Farber  
Cancer Institute, and  
Biostatistics, Harvard TH Chan  
School of Public Health  
aedin@ds.dfci.harvard.edu



**Elana Fertig**  
Associate Professor of  
Oncology, Biomedical  
Engineering, and Applied  
Mathematics and Statistics  
Johns Hopkins University  
ejfertig@jhmi.edu



**Kim-Anh Lê Cao**  
Associate Professor  
Melbourne Integrative  
Genomics, School of  
Mathematics and Statistics  
The University of Melbourne  
kimanh.lecao@unimelb.edu.au