Introduction to Peak Analysis

https://tinyurl.com/Peak-analysis

Harvard Chan Bioinformatics Core

Introductions!

Shannan Ho Sui
Director

Meeta Mistry
Associate Director

Lorena Pantano
Director of Bioinformatics Platform

John Quackenbush
Faculty Advisor

Upen Bhattarai

Heather Wick

Will Gammerdinger

Noor Sohail

Elizabeth Partan

Alex Bartlett

Emma Berdan

James Billingsley

Zhu Zhuo

Maria Simoneau

Shannan Ho Sui
Director

Meeta Mistry
Associate Director

Lorena Pantano
Director of Bioinformatics Platform

John Quackenbush
Faculty Advisor

Upen Bhattarai

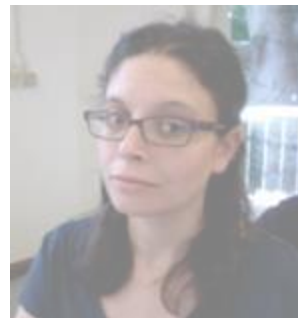Heather Wick

Will Gammerdinger

Noor Sohail

Elizabeth Partan
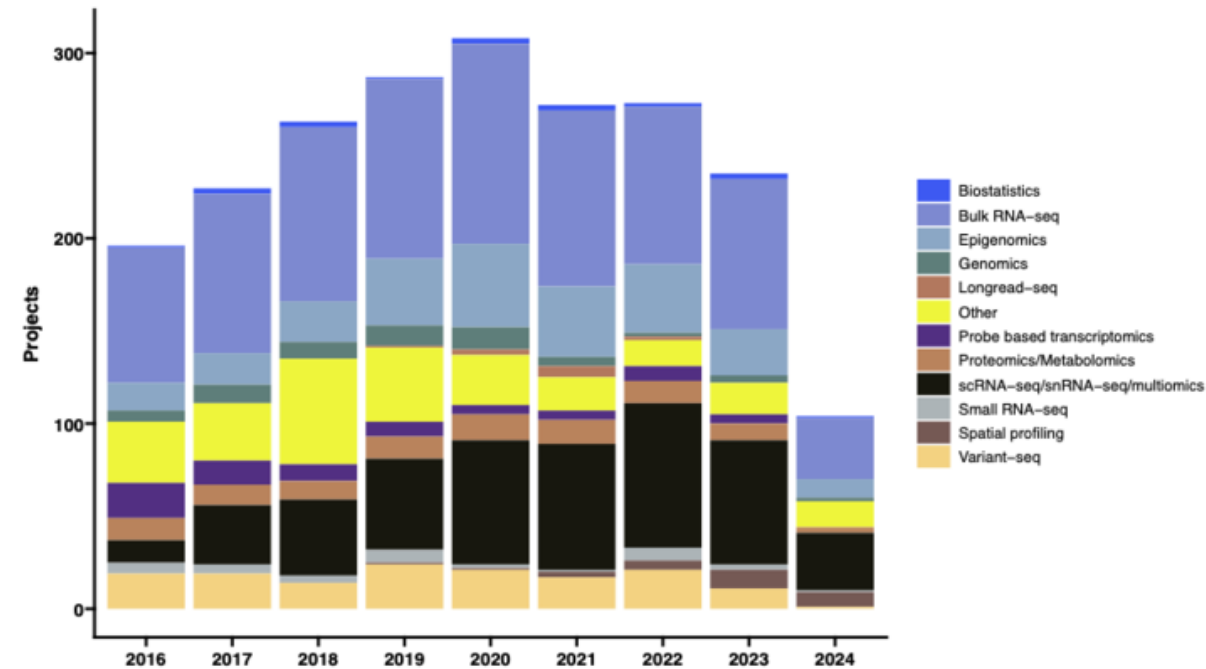
Alex Bartlett

Emma Berdan

James Billingsley

Zhu Zhuo

Maria Simoneau

Consulting

❖ Transcriptomics: Bulk, single cell, small RNA

❖ Epigenomics: ChIP-seq, CUT&RUN, ATAC-seq, DNA methylation

❖ Variant discovery: WGS, resequencing, exome-seq and CNV

❖ Multiomics integration

❖ Spatial biology

❖ Experimental design and grant support

Consulting

❖ Transcriptomics: Bulk, single cell, small RNA

❖ Epigenomics: ChIP-seq, CUT&RUN, ATAC-seq, DNA
methylation

❖ Variant discovery: WGS, resequencing, exome-seq and CNV

❖ Multiomics integration

❖ Spatial biology

❖ Experimental design and grant support

NIEHS

Training

❖ Hands-on workshops design to reflect best practices, reproducibility and an emphasis on experimental design

❖Basic Data Skills

❖ Shell

❖ R

❖Advanced Topics: Analysis of high-throughput sequencing data

❖ Chromatin Biology & Peak Analysis

❖ Bulk RNA-seq

❖ Differential Gene Expression (Bulk & scRNA)

❖ scRNA-seq

❖ Variant Calling

❖Current Topics in Bioinformatics

https://bioinformatics.sph.harvard.edu/training

Training

❖ Hands-on workshops design to reflect best practices, reproducibility and an emphasis on experimental design

  ❖ Basic Data Skills

    ❖ Shell

    ❖ R

  ❖ Advanced Topics: Analysis of high-throughput sequencing data

    ❖ Chromatin Biology & Peak Analysis

    ❖ Bulk RNA-seq

    ❖ Differential Gene Expression (Bulk & scRNA)

    ❖ scRNA-seq

    ❖ Variant Calling

  ❖ Current Topics in Bioinformatics

https://bioinformatics.sph.harvard.edu/training

Workshop scope

Bioinformatic Data Analysis
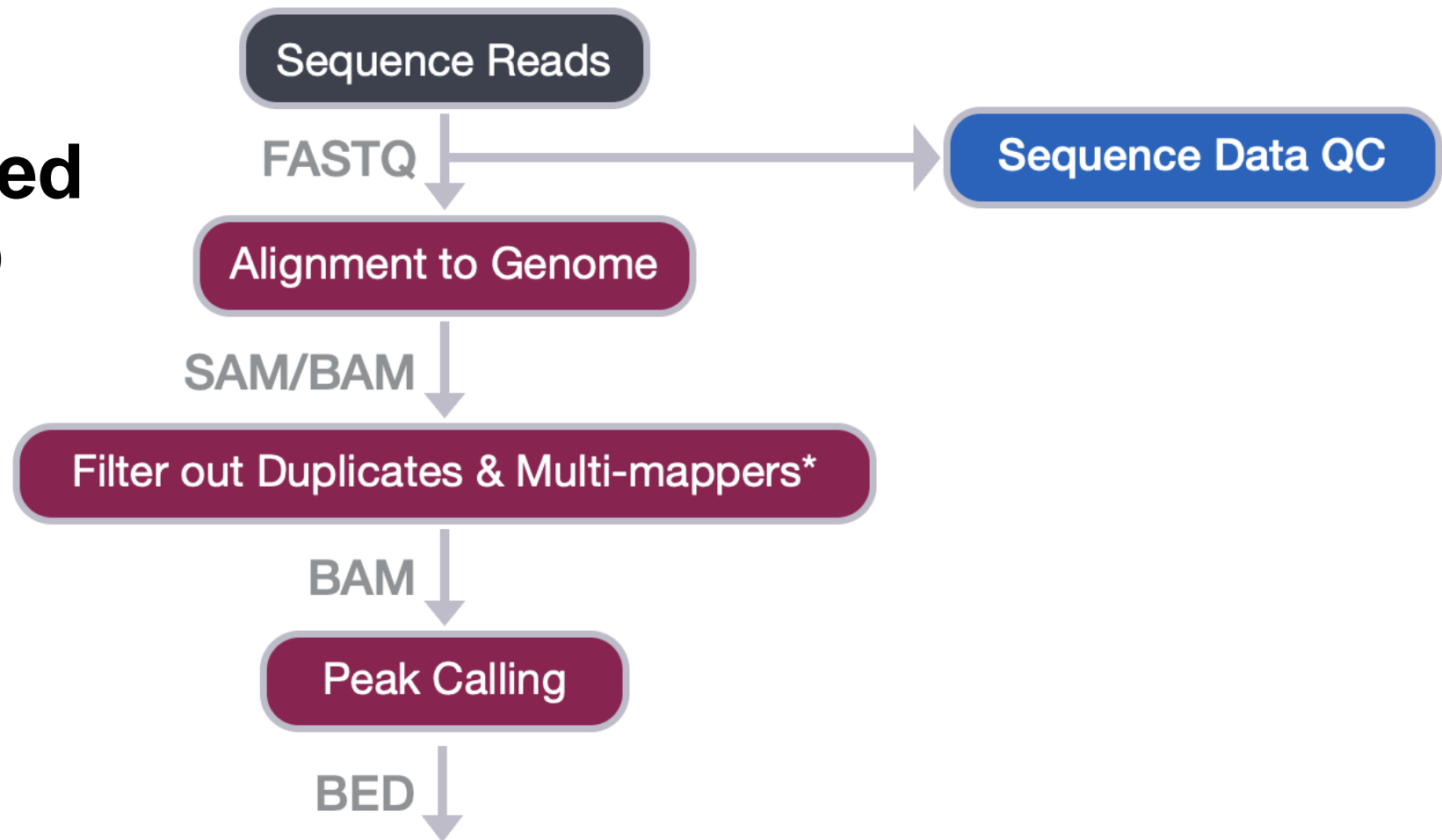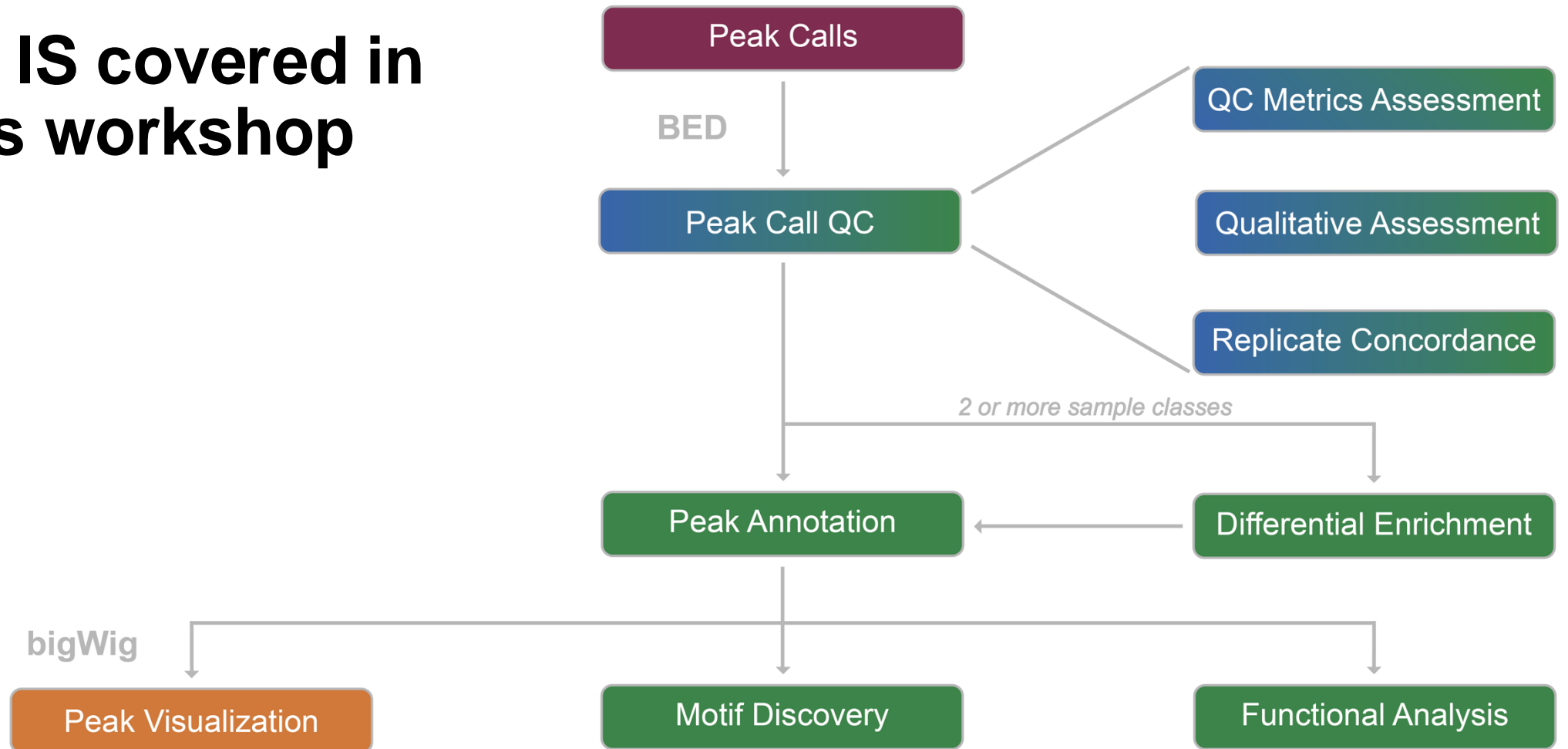
Workshop Scope
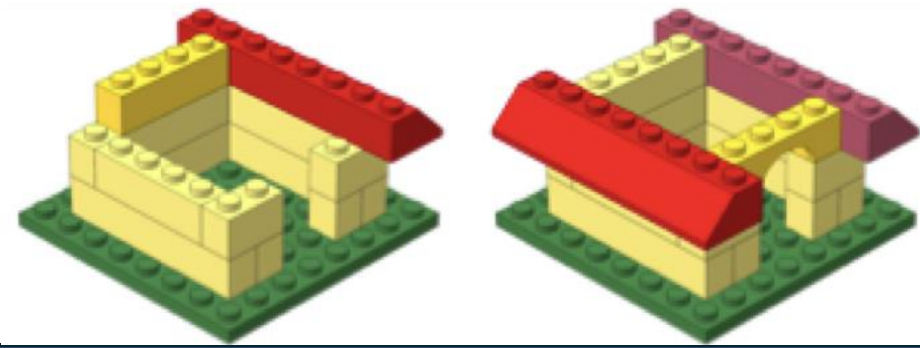
❖ Describe peak data and different file formats generated from peak calling algorithms

❖ Assess various metrics used to assess the quality of peak calls

❖ Compare peak calls across samples within a dataset

❖ Create visualizations to evaluate peak annotations

❖ Evaluate differentially enriched regions between two sample groups

Logistics

Course schedule

Day 1

Time	Topic	Instructor
09:30 - 09:45	Workshop Introduction	Meeta
09:45 - 10:15	Pre-reading discussion	Meeta
10:15 - 11:00	Understanding peaks and peak file formats	Meeta
11:00- 11:05	Break	
11:05 - 12:00	Assessing peak quality metrics	Will

Before the next class:

I. Please **study the contents** and **work through all the code** within the following lessons:

1. Assessing sample similarity and identifying potential outliers
   Click here for a preview of this lesson
2. Concordance across replicates using peak overlaps
   Click here for a preview of this lesson
3. **Complete the exercises**:
   ○ Each lesson above contains exercises; please go through each of them.
   ○ Copy over your solutions into the Google Forms the **day before the next class**.

https://tinyurl.com/Peak-analysis

Course materials

❖ We continuously update our materials to reflect changes in the field/software

Intro to dataset and peak call file formats

An introduction to various methods/approaches for the analysis of peaks generated from ChIP-seq / CUT&RUN / ATAC-seq

View on GitHub

Contributors: Heather Wick, Upendra Bhattarai, Meeta Mistry, Will Gammerdinger

Approximate time: 40 minutes

Learning Objectives

In this lesson, we will:

* Explain the dataset and the biological context
* Define peaks as genomic coordinate data
* Describe file formats for peak data

Introduction to the dataset

For this workshop we will be working with ChIP-seq data from a publication in Neuron by *Baizabal et al., 2018* 1.

Course participation

❖ Mandatory review of self-learning lessons and assignments

❖ Attendance required for all classes

❖ Your questions and active participation drive learning

❖ **We look forward to all of your questions!**

Course participation

❖ At-home lessons and exercises after each session

❖ Cover material not previously discussed

❖ Provides us feedback to help pace the course appropriately

❖ 3-5 hours to complete

Using AI for Assignments

❖ Do

    ❖ Try to resolve error messages with it

    ❖ Test code written by AI on a dataset where you have expected results

    ❖ Take the time to review the generated code line-by-line

❖ Don't

    ❖ Implement it in replacement to learning

    ❖ Write code that you don't understand

    ❖ Assume the output from an AI process is correct

Odds & Ends

❖ Quit/minimize all applications that are not required for class

❖ Name tags

❖ Post-its

  ❖ 🟩 green - I am all set

  ❖ 🟥 red - I need time/help

❖ Phones on vibrate/silent

❖ Bathrooms

Contact Us

- ❖ *HBC training team:* hbctraining@hsph.harvard.edu

- ❖ *HBC consulting:* bioinformatics@hsph.harvard.edu