Introduction to Variant Analysis

https://tinyurl.com/Intro-to-variant-analysis

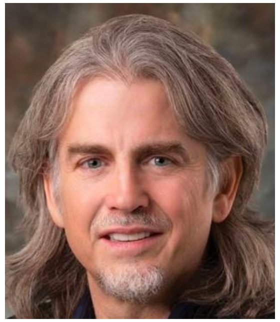
Harvard Chan Bioinformatics Core

Introductions!

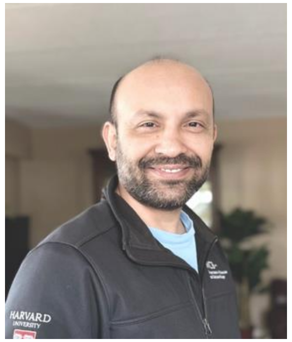
Shannan Ho Sui
Director

Meeta Mistry
Associate Director

Lorena Pantano
Director of Bioinformatics Platform



John Quackenbush
Faculty Advisor



Upen Bhattarai

Heather Wick

Will Gammerdinger

Noor Sohail

Alex Bartlett

Elizabeth
Partan

Emma Berdan

James Billingsley

Zhu Zhuo

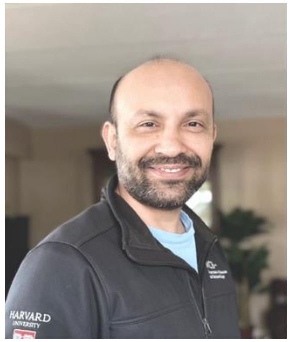
Maria Simoneau

Shannan Ho Sui
Director

Meeta Mistry
Associate Director

Lorena Pantano
Director of Bioinformatics Platform

John Quackenbush
Faculty Advisor



Upen Bhattarai

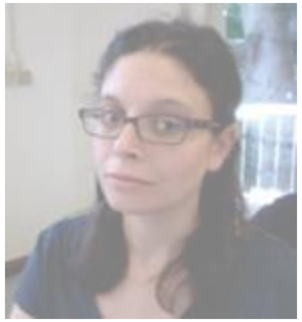
Heather Wick

Will Gammerdinger

Noor Sohail

Alex Bartlett

Elizabeth Partan



Emma Berdan

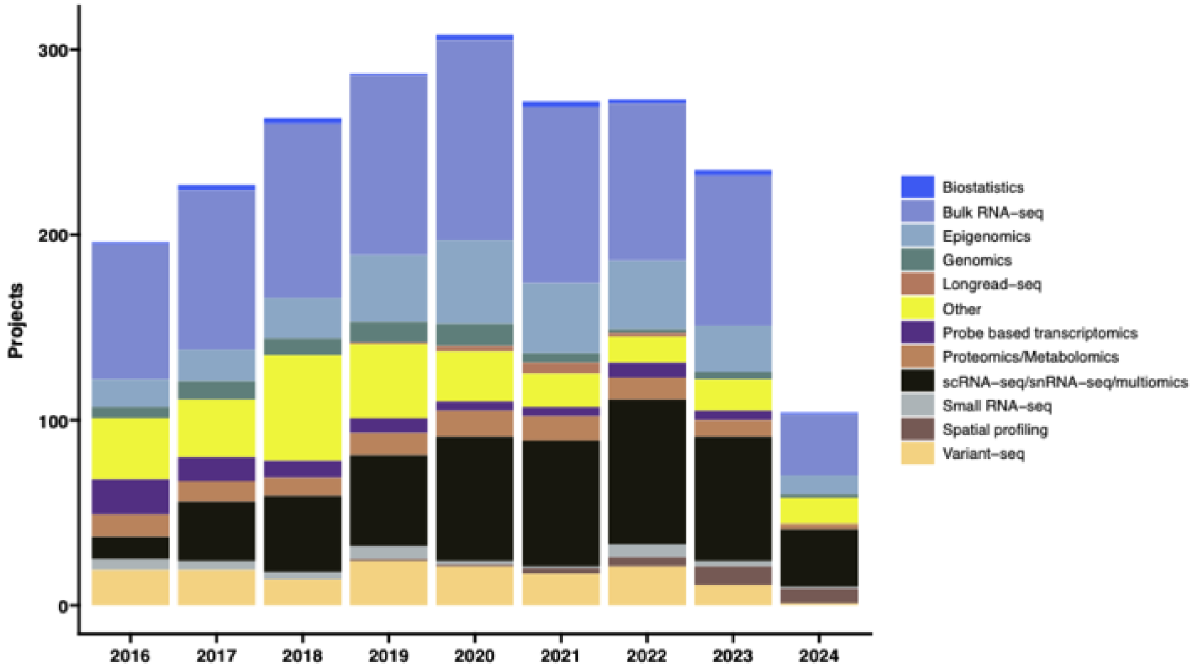
James Billingsley

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



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

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

CFAR
HARVARD UNIVERSITY
CENTER FOR AIDS RESEARCH

HARVARD
T.H. CHAN
SCHOOL OF PUBLIC HEALTH

NIEHS

HARVARD
CATALYST
THE HARVARD CLINICAL AND TRANSLATIONAL SCIENCE CENTER

HARVARD
MEDICAL SCHOOL

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
    - Bulk RNA-seq
    - Differential Gene Expression
    - 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
    - Bulk RNA-seq
    - Differential Gene Expression
    - scRNA-seq
    - Variant Calling
  - Current Topics in Bioinformatics

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

CFAR
HARVARD UNIVERSITY
CENTER FOR AIDS RESEARCH

HARVARD
T.H. CHAN
SCHOOL OF PUBLIC HEALTH

DF/HCC
DANA-FARBER / HARVARD CANCER CENTER

HARVARD
CATALYST
THE HARVARD CLINICAL AND TRANSLATIONAL SCIENCE CENTER

HARVARD
MEDICAL SCHOOL

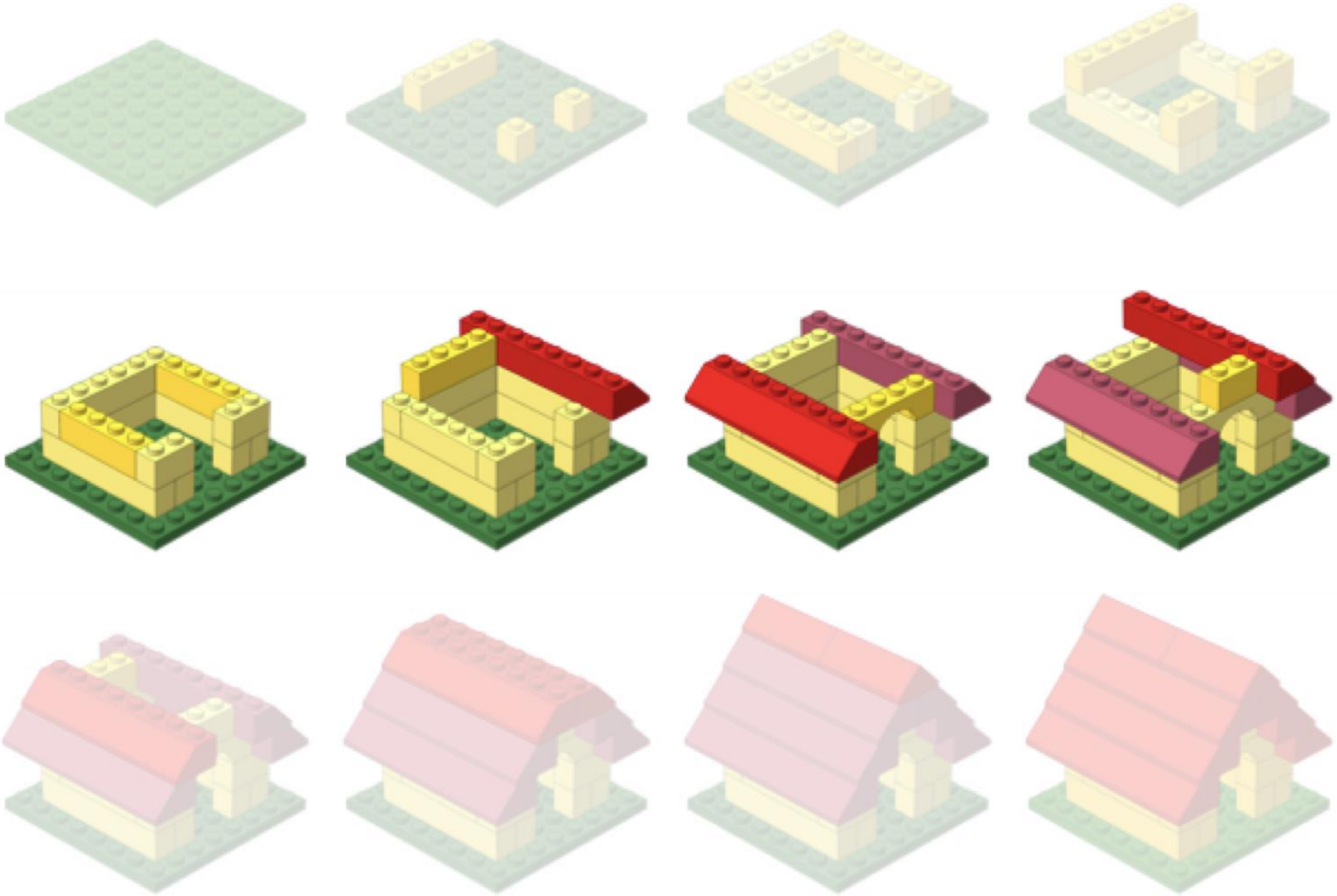
Join us for HBC Community Breakfast!

- An opportunity to get to know others in the community
- Free food and beverages
- Great conversations

More Info:

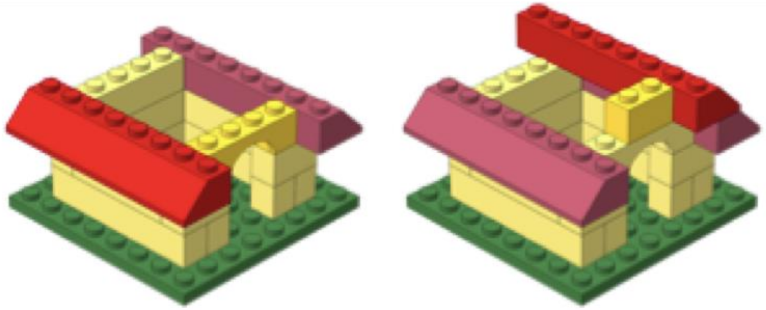
http://bioinformatics.sph.harvard.edu/breakfast/

Workshop scope



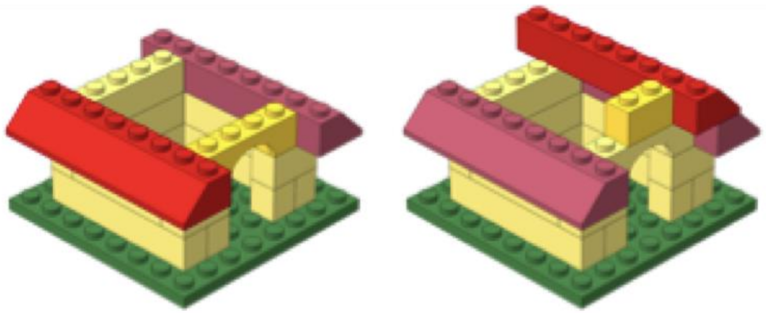
Bioinformatic Data Analysis

Variant Analysis

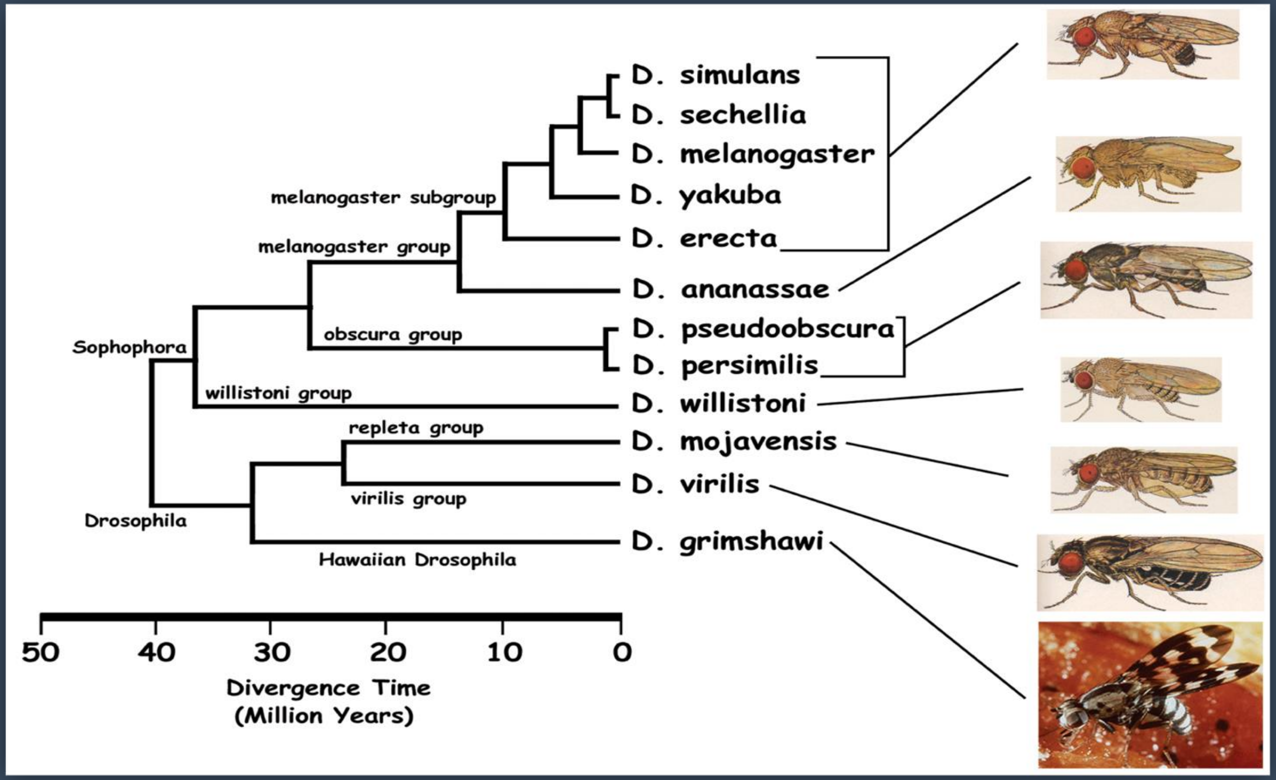


- *Variant Calling* identifies loci in the genome/exome where allelic variation exist
- *Variant Annotation* integrates outside databases and information to provide alleles with context on potential functional impacts
- *Variant Prioritization* allows a framework through which a researcher can filter through their annotated variants to find ones of potential interest

Applications

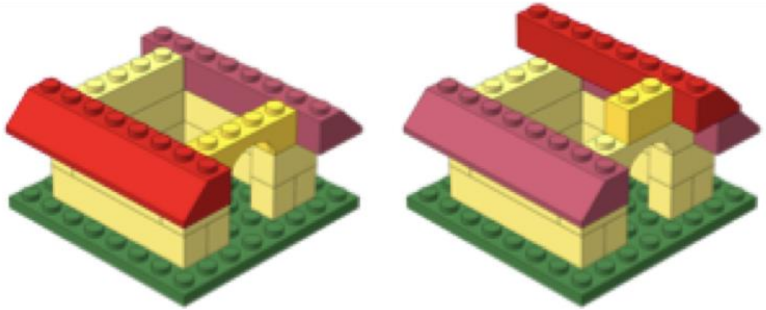


- Evolutionary Biology

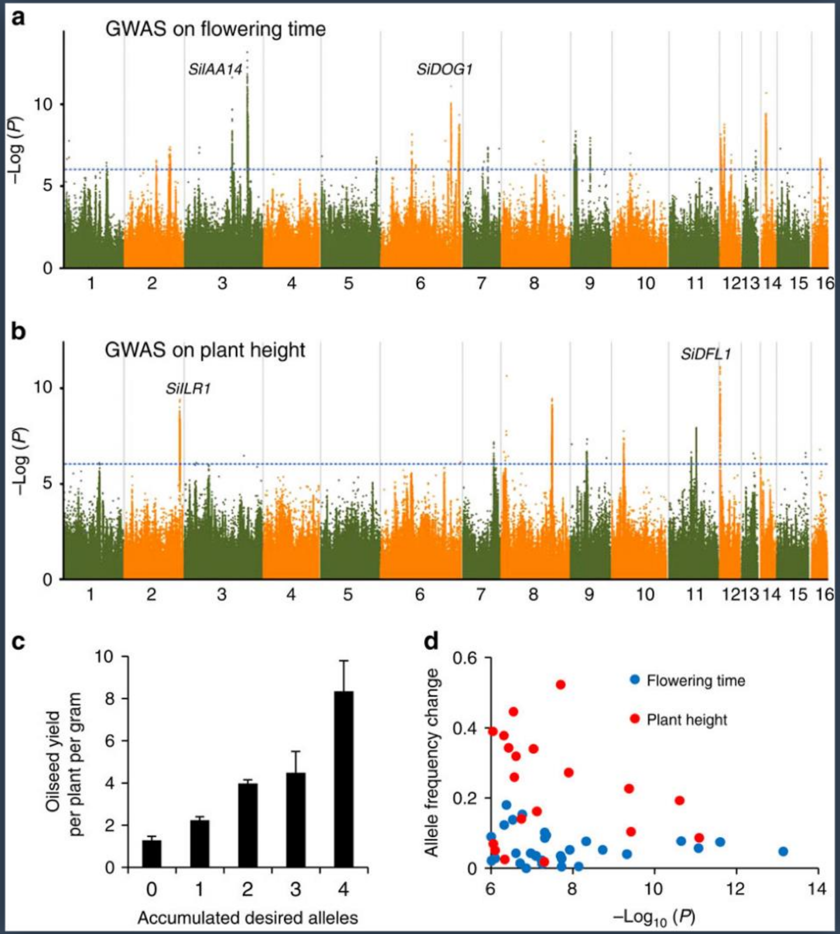


http://insects.eugenes.org/DroSpeGe/

Applications

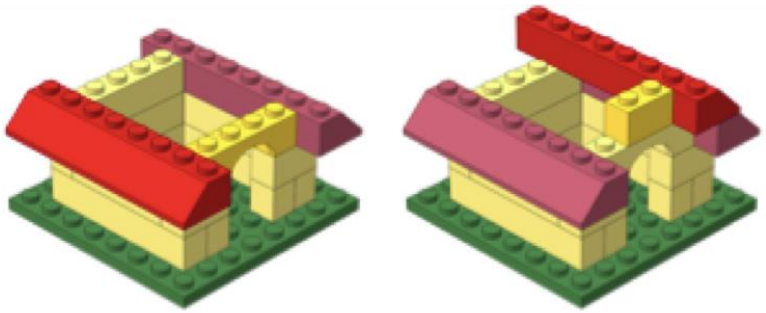


- Evolutionary Biology
- Agriculture

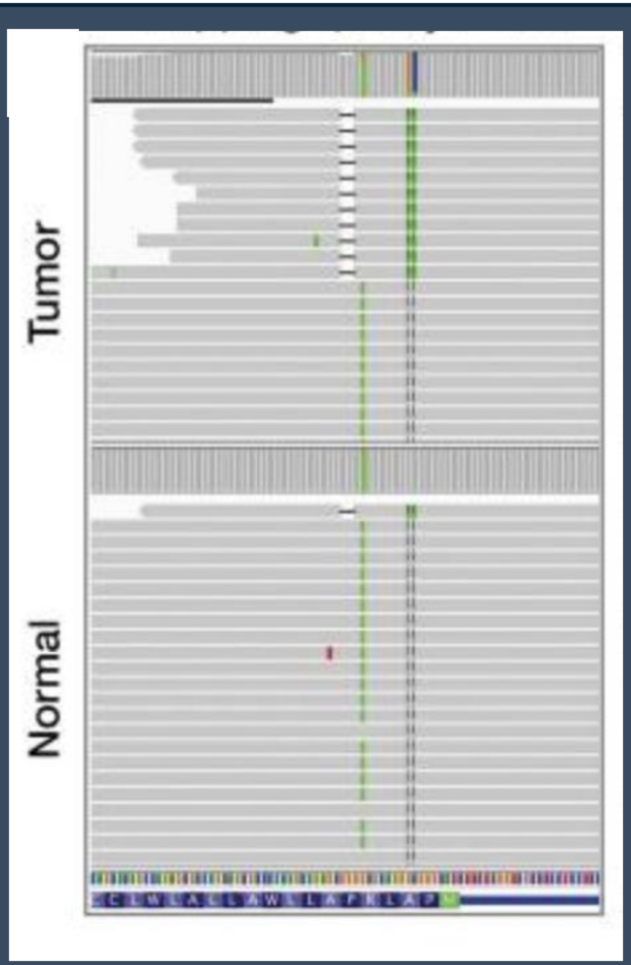


http://www.nature.com/ncomms/2015/151019/ncomms9609/fig_tab/ncomms9609_F6.html

Applications

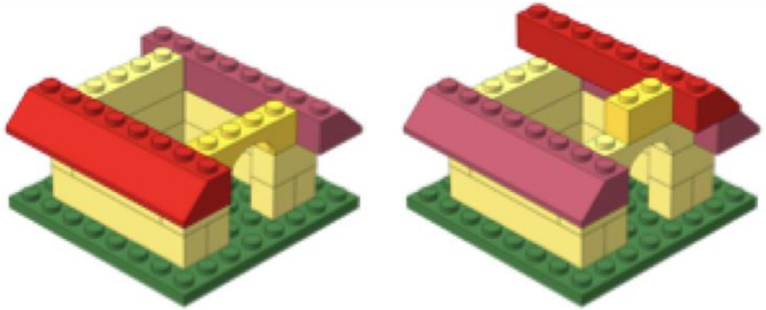


- Evolutionary Biology
- Agriculture
- Clinical Applications
  - Tracking infectious disease
  - Identifying alleles responsible for heredity diseases
  - Interrogating potential drivers of cancer from paired tumor-normal samples



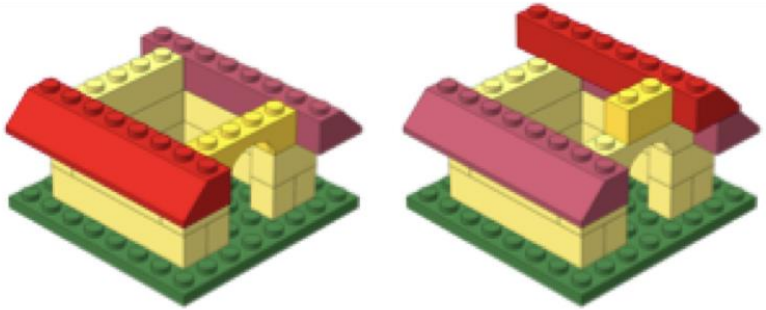
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5678989/

Learning Objectives



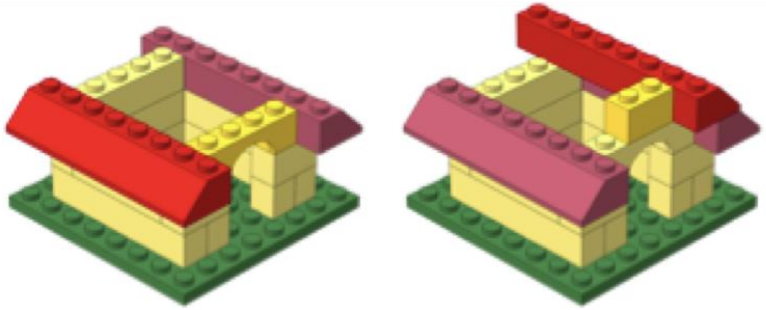
- Evaluate QC metrics for variant calling
- Call variants using GATK
- Filter variants to retain only high-quality variant calls
- Annotate variants using SnpEff and dbSNP
- Prioritize variants by their impact
- Visualize variants in IGV
- Implement cBioPortal to analyze variants

Tools for Variant Analysis

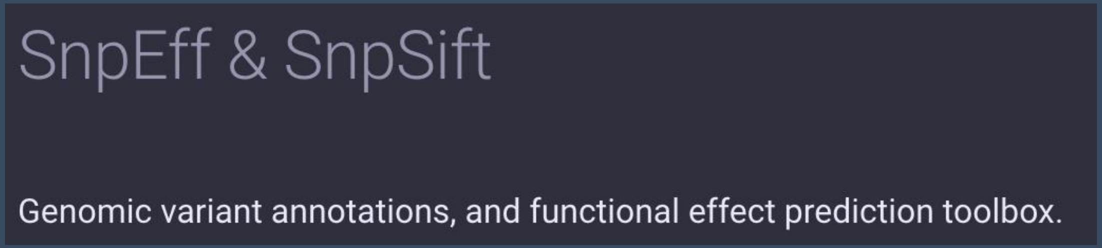


- Variant Calling
  - Genomic Analysis Toolkit (GATK)

Tools for Variant Analysis



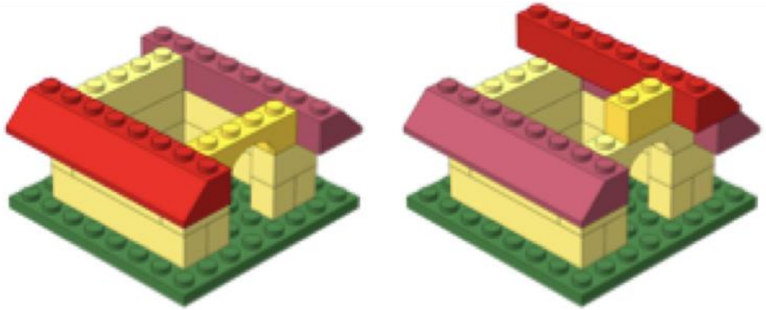
- Variant Calling
  - Genomic Analysis Toolkit (GATK)
- Variant Annotation and Prioritization
  - SnpEff and SnpSift



SnpEff & SnpSift

Genomic variant annotations, and functional effect prediction toolbox.

Tools for Variant Analysis



- Variant Calling
  - Genomic Analysis Toolkit (GATK)
- Variant Annotation and Prioritization
  - SnpEff and SnpSift
- Variant Visualization
  - Integrative Genomics Viewer
  - cBioPortal

gatk

SnpEff & SnpSift

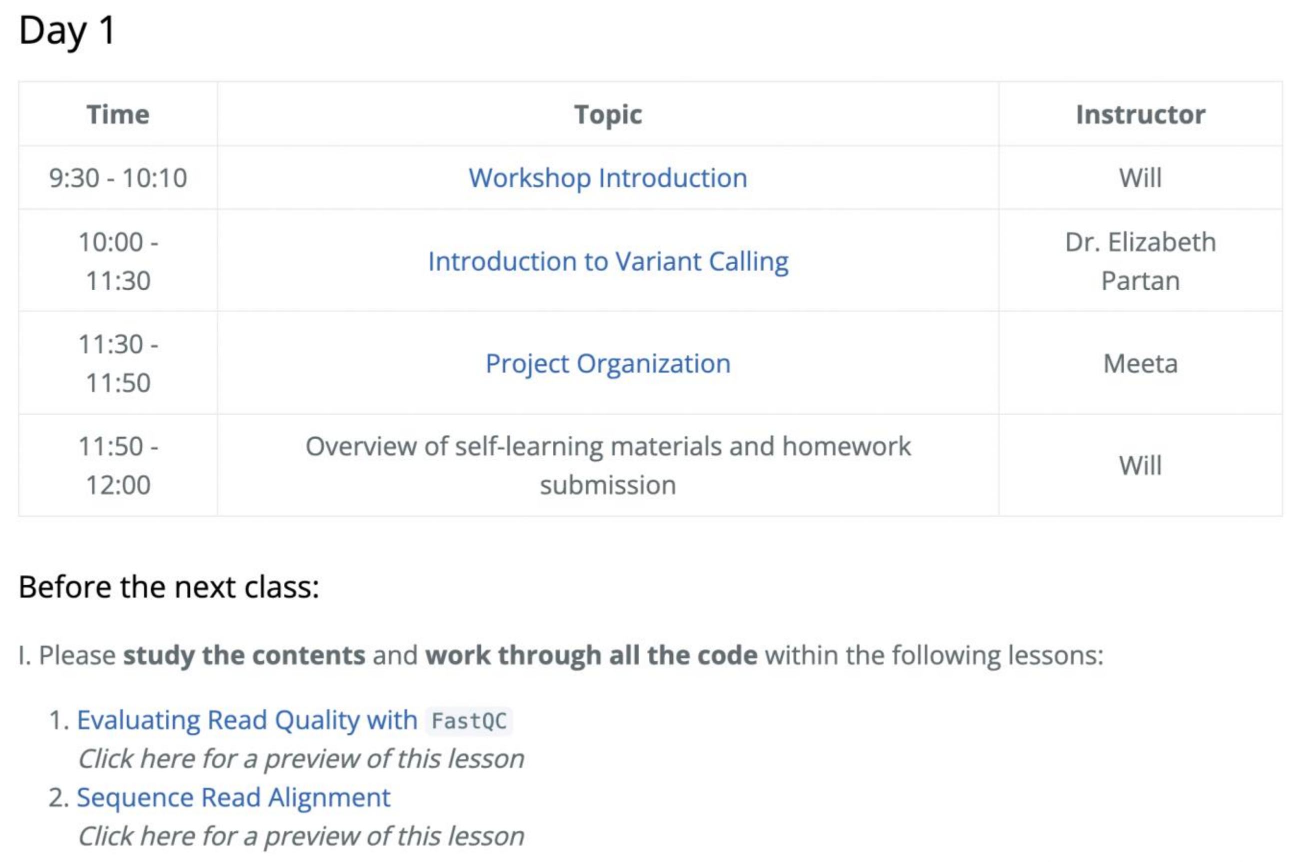
Genomic variant annotations, and functional effect prediction toolbox.

igv

cBioPortal
FOR CANCER GENOMICS

Logistics

Course schedule



Day 1

Time	Topic	Instructor
9:30 - 10:10	Workshop Introduction	Will
10:00 - 11:30	Introduction to Variant Calling	Dr. Elizabeth Partan
11:30 - 11:50	Project Organization	Meeta
11:50 - 12:00	Overview of self-learning materials and homework submission	Will

Before the next class:

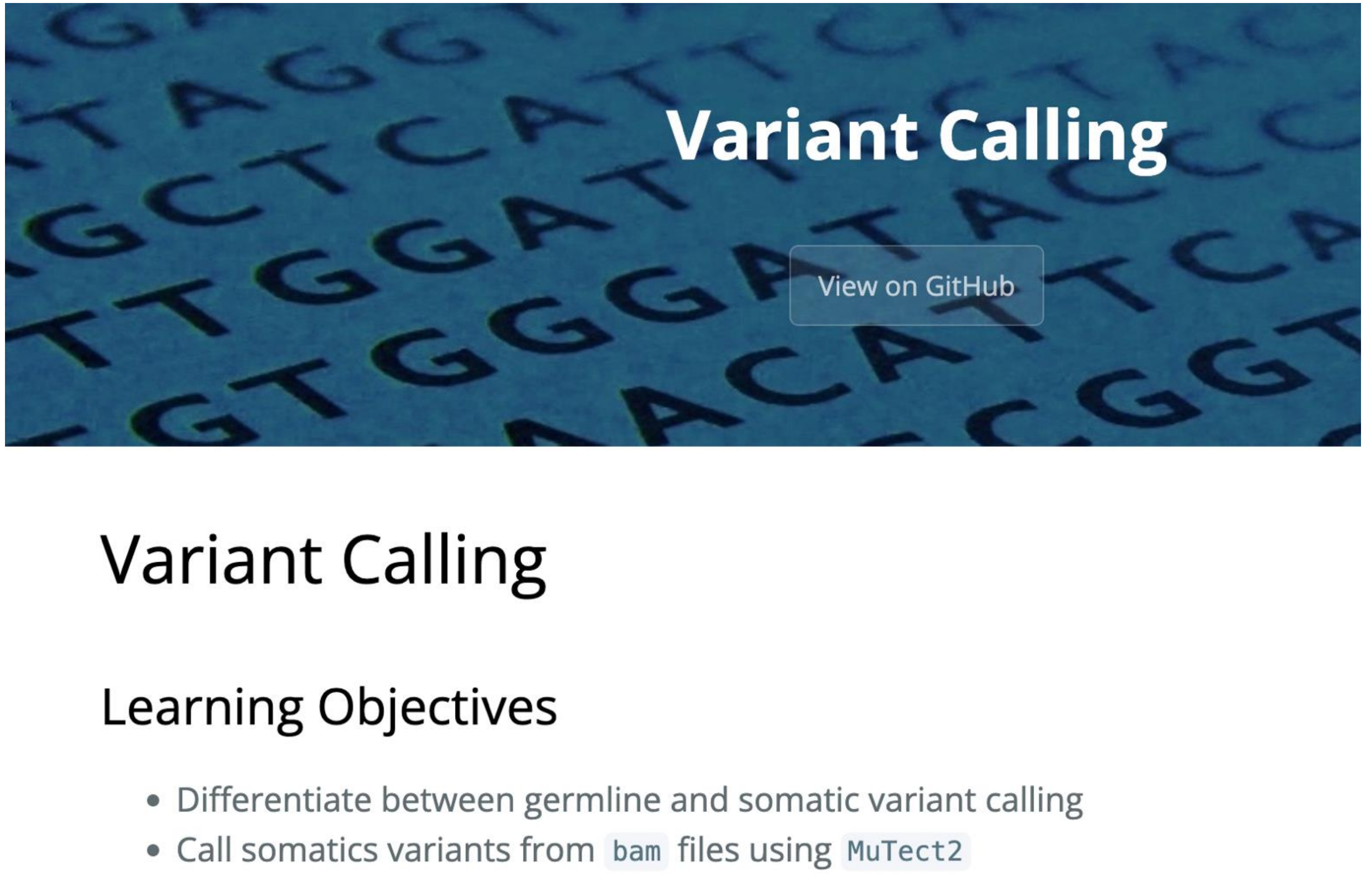
I. Please **study the contents** and **work through all the code** within the following lessons:

1. Evaluating Read Quality with `FastQC`
   Click here for a preview of this lesson
2. Sequence Read Alignment
   Click here for a preview of this lesson

https://tinyurl.com/Intro-to-variant-analysis

Course materials

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



Variant Calling

Learning Objectives

- Differentiate between germline and somatic variant calling
- Call somatics variants from `bam` files using `MuTect2`

https://tinyurl.com/Intro-to-variant-analysis

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
- Homework load is heavier in the beginning of this workshop series and tapers off

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

Thanks!

- Kathleen Chappell and Andy Bergman from HMS-RC
- Dr. Tali Mazor from DFCI
- Data Carpentry

Contact Us

- *HBC training team:* hbctraining@hsph.harvard.edu
- *HBC consulting:* bioinformatics@hsph.harvard.edu
- *O2 (HMS-RC):* rchelp@hms.harvard.edu