Bioconductor
OPEN SOURCE SOFTWARE FOR BIOINFORMATICS

Introduction to Differential Gene Expression Analysis

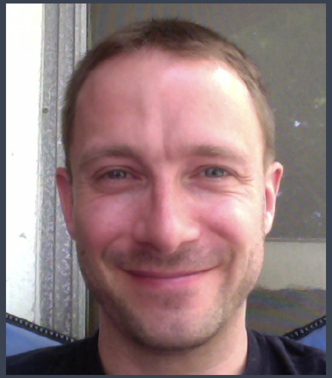
Harvard Chan Bioinformatics Core

http://tinyurl.com/hbc-intro-to-dge

Shannan Ho Sui

John Hutchinson

Brad Chapman


Rory Kirchner

Meeta Mistry

Radhika Khetani

Mary Piper

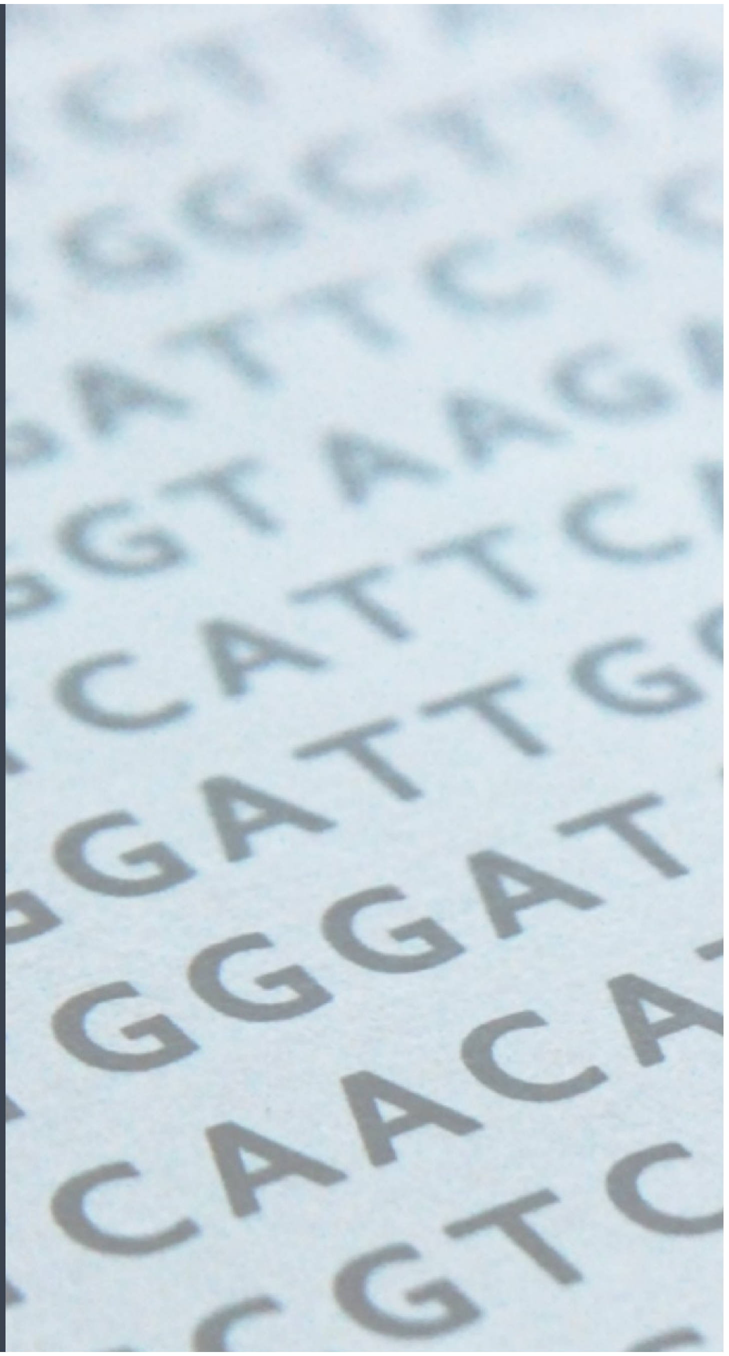
Victor Barrera

Lorena Pantano

Peter Kraft

Consulting

- RNA-seq, small RNA-seq and ChIP-seq analysis
- Genome-wide methylation
- WGS, resequencing, exome-seq and CNV studies
- Quality assurance and analysis of gene expression arrays
- Functional enrichment analysis
- Grant support



HARVARD
T.H. CHAN
SCHOOL OF PUBLIC HEALTH

HSCI
HARVARD STEM CELL
INSTITUTE

HARVARD
CATALYST
THE HARVARD CLINICAL
AND TRANSLATIONAL
SCIENCE CENTER

HARVARD
MEDICAL SCHOOL

NIEHS / CFAR Bioinformatics Core

Center for Stem Cell Bioinformatics

Harvard Catalyst Bioinformatics Consulting

HMS Tools & Technology

HARVARD
T.H. CHAN
SCHOOL OF PUBLIC HEALTH

HSCI
HARVARD STEM CELL
INSTITUTE

HARVARD
CATALYST
THE HARVARD CLINICAL
AND TRANSLATIONAL
SCIENCE CENTER

HARVARD
MEDICAL SCHOOL

NIEHS / CFAR Bioinformatics Core

Center for Stem Cell Bioinformatics

Harvard Catalyst Bioinformatics Consulting

HMS Tools & Technology

Training

We have divided our short workshops into 2 categories:

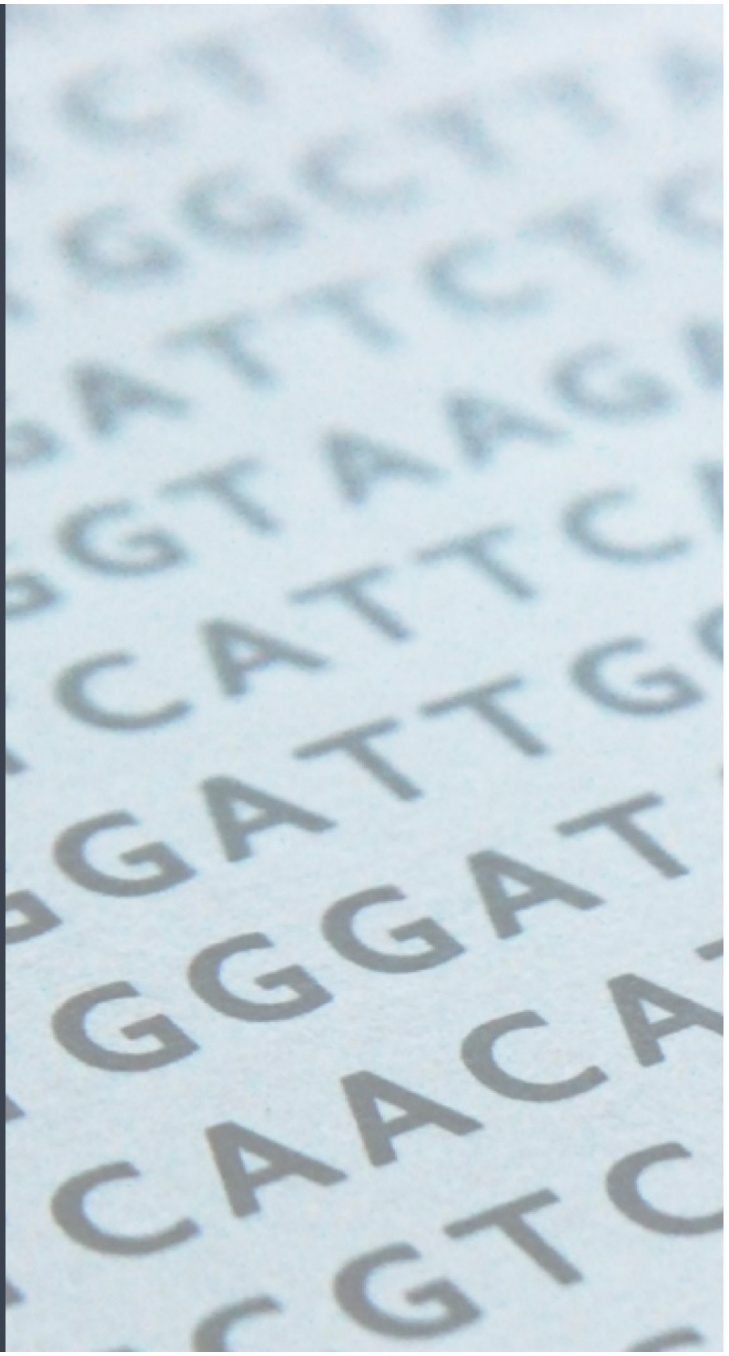
1. Basic Data Skills - No prior programming knowledge needed (no prerequisites)
2. Advanced Topics: Analysis of high-throughput sequencing (NGS) data - Certain “Basic” workshops required as prerequisites.

Any participants wanting to take an advanced workshop will have to have taken the appropriate basic workshop(s) within the past 6 months.

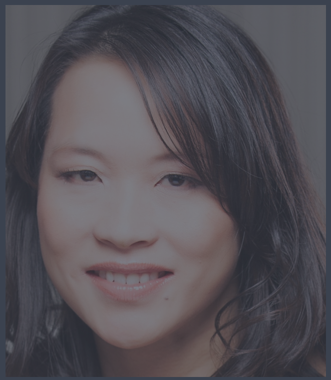
https://hbctraining.github.io/main/training_spring2019.html

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

https://hbctraining.github.io/main/

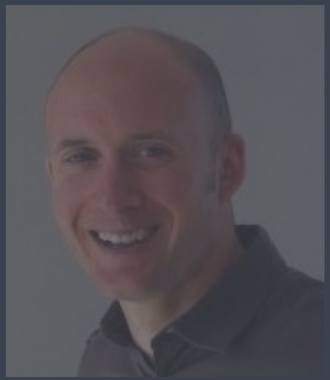


Introductions!

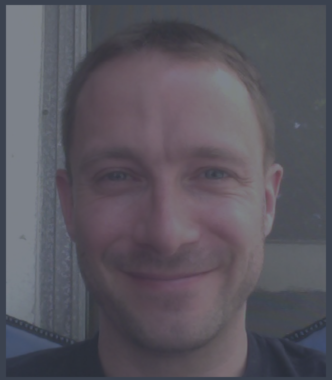


Shannan Ho Sui

John Hutchinson



Brad Chapman



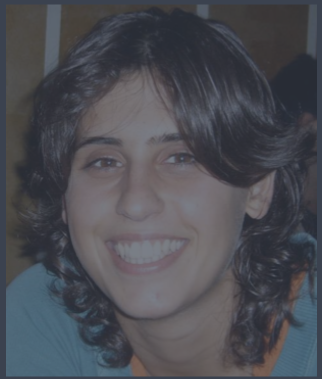
Rory Kirchner

Meeta Mistry

Radhika Khetani

Mary Piper

Victor Barrera

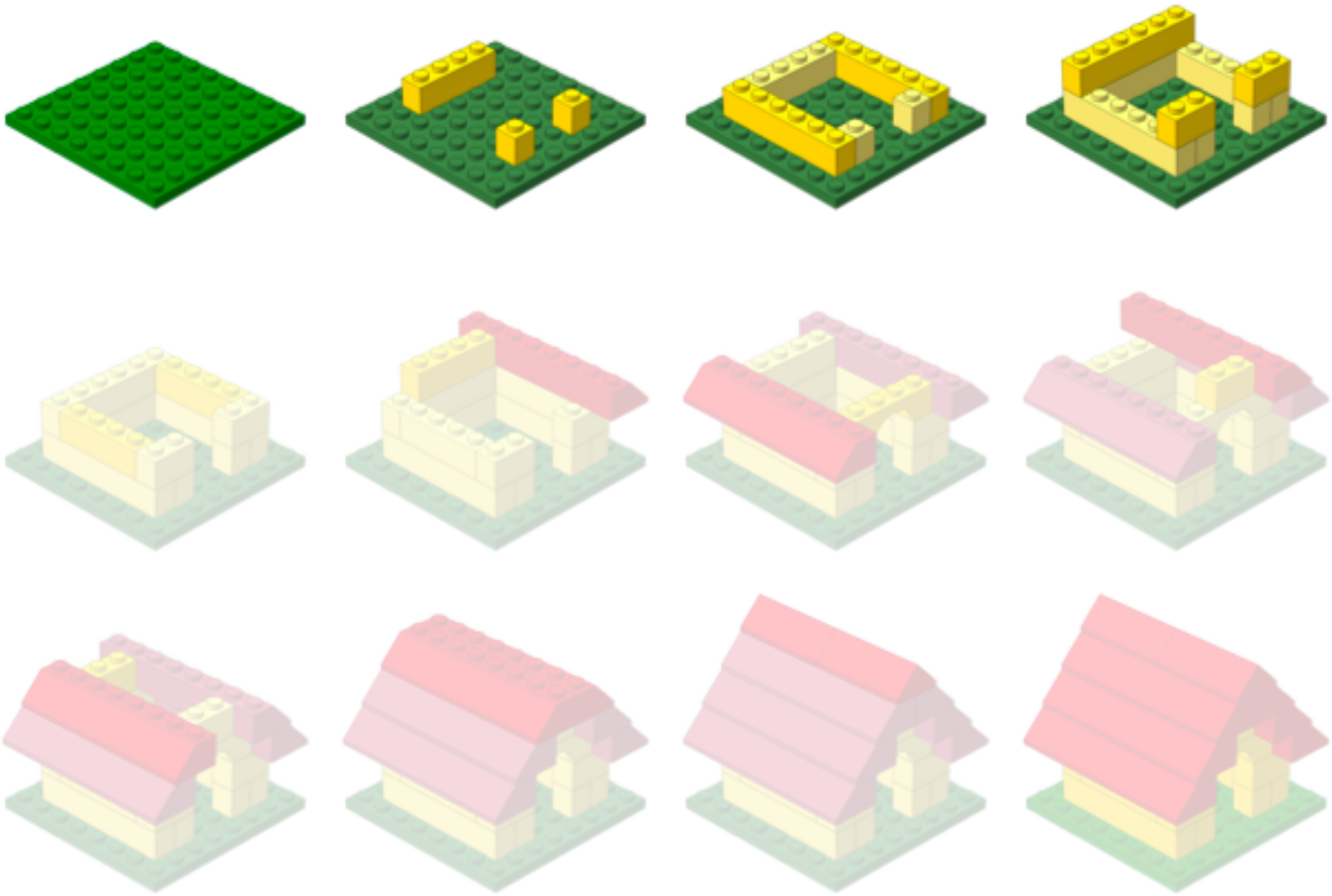


Lorena Pantano

Peter Kraft

Class Introductions!

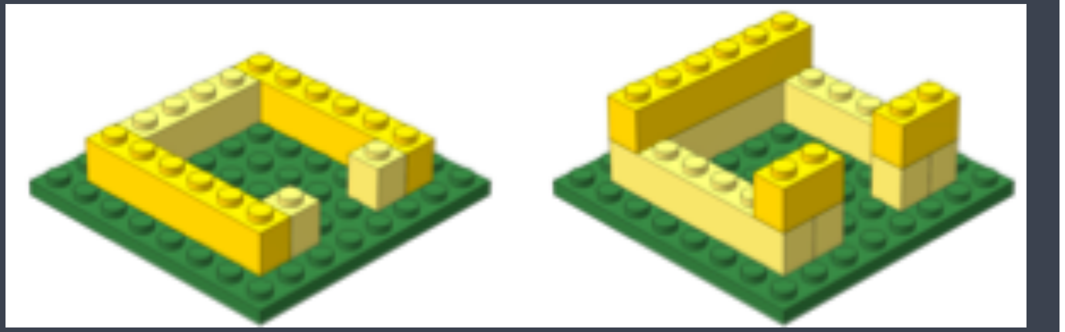
Workshop Scope…



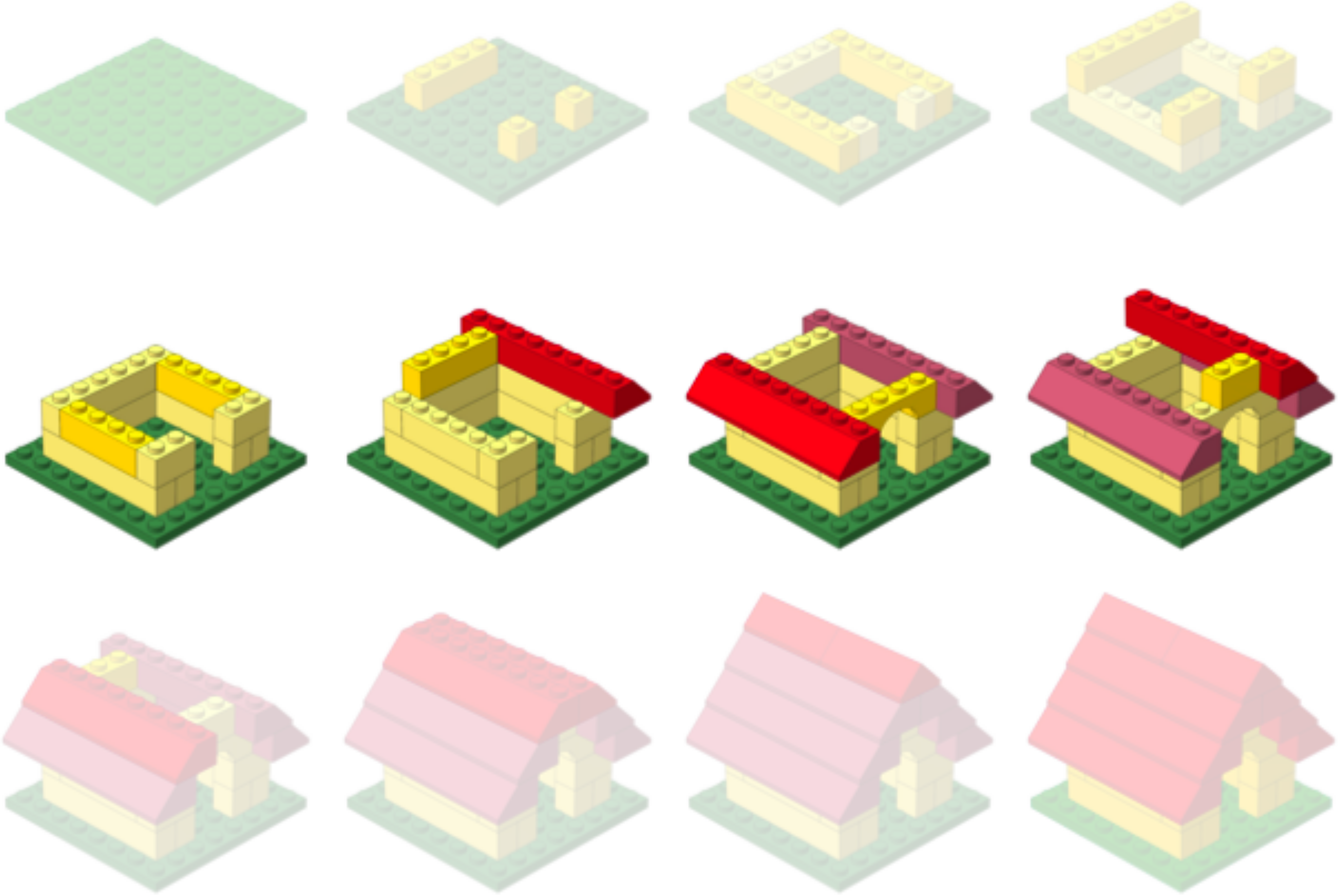
http://anoved.net/tag/lego/page/3/

Setting up to perform Bioinformatics analysis

Setting up…



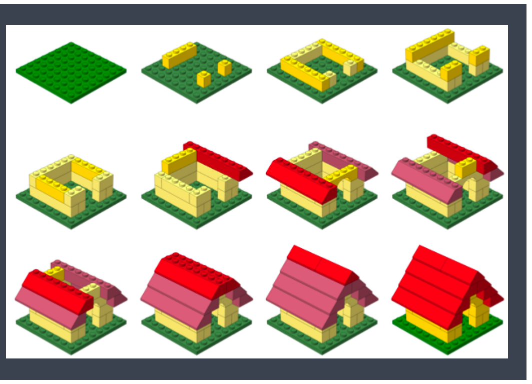
- ✓ Introduction to the command-line interface (shell, Unix, Linux)
    - Dealing with large data files
    - Performing bioinformatics analysis
        - Using tools
        - Accessing and using compute clusters
- ✓ R
    - Parsing and working with smaller results text files
    - Statistical analysis, e.g. differential expression analysis
    - Generating figures from complex data



http://anoved.net/tag/lego/page/3/

Bioinformatics data analysis

Workshop Scope



Differential Gene Expression analysis

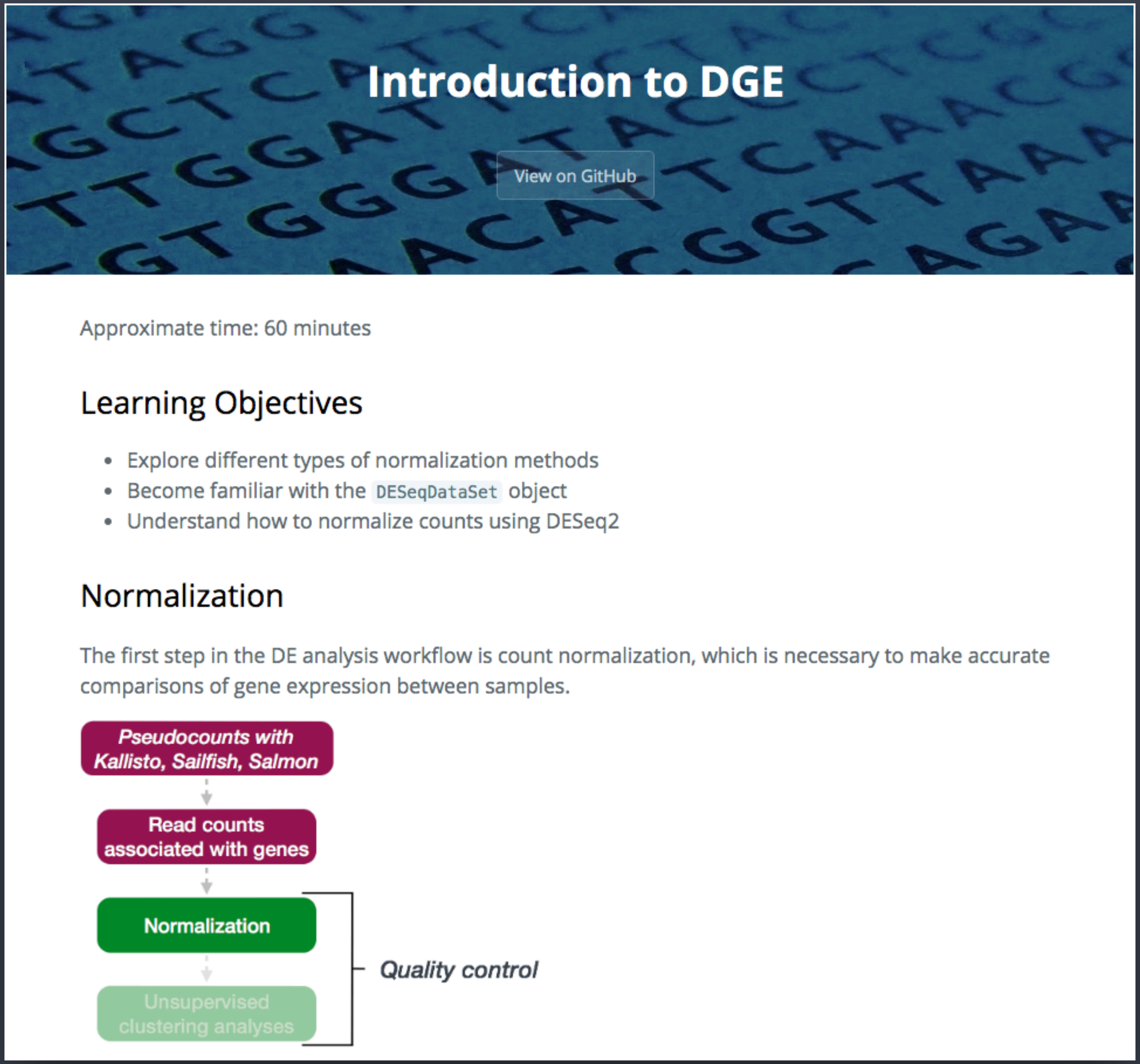
- ✓ Understand the considerations for performing statistical analysis on RNA-seq data
- ✓ Start with gene counts (after alignment and counting)
- ✓ Perform QC on count data
- ✓ Use DESeq2 to perform differential expression analysis on the count data and obtain a list of significantly different genes
- ✓ Visualize results of the analysis
- ✓ Perform functional analysis on the lists of differentially expressed genes

Logistics

Course webpage (wiki)

http://tinyurl.com/hbc-intro-to-dge

Course materials online



Introduction to DGE

View on GitHub

Approximate time: 60 minutes

Learning Objectives

- Explore different types of normalization methods
- Become familiar with the `DESeqDataSet` object
- Understand how to normalize counts using DESeq2

Normalization

The first step in the DE analysis workflow is count normalization, which is necessary to make accurate comparisons of gene expression between samples.

Pseudocounts with Kallisto, Sailfish, Salmon

Read counts associated with genes

Normalization

Unsupervised clustering analyses

Quality control

Odds and Ends

- Name tags: Tent Cards
- Post-its
- Wi-Fi: **HMS Public** or **HMS Secure**
- Lunch locations
- Bathrooms
- Water Fountain
- Phones on vibrate/silent!

Contact us!

HBC training team: hbctraining@hsph.harvard.edu

HBC consulting: bioinformatics@hsph.harvard.edu

Twitter

@bioinfocore