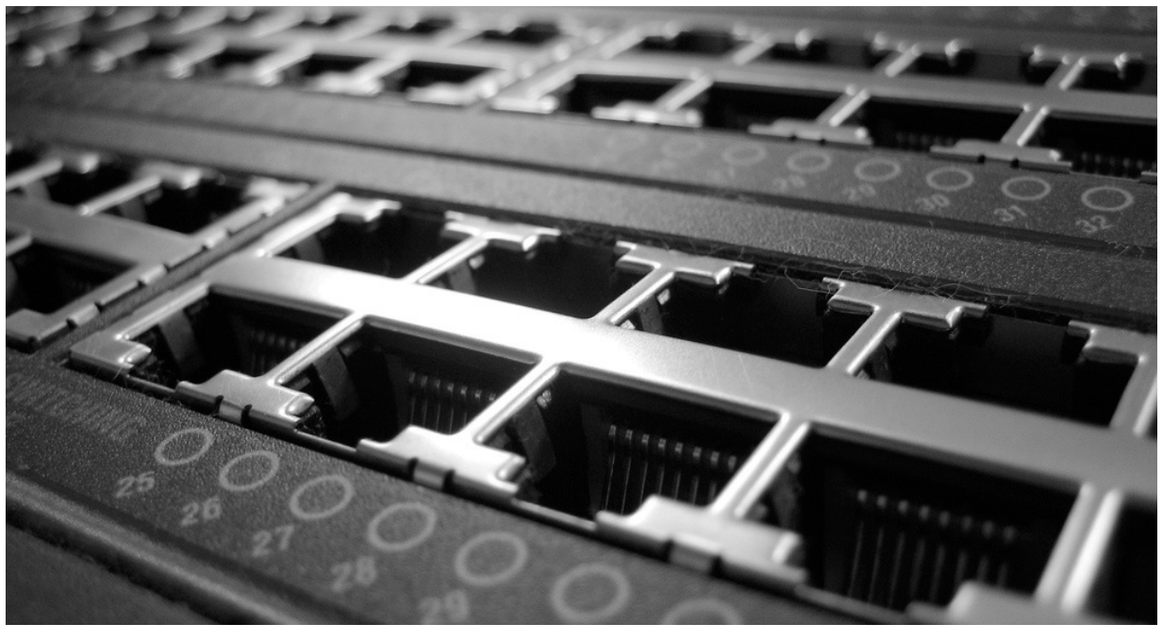


Introduction to the command-line interface

(shell)

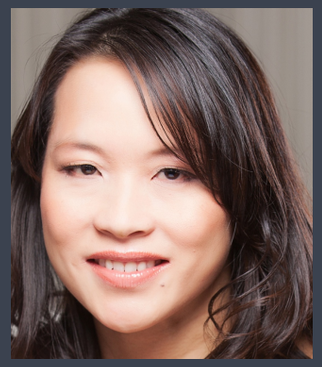
Harvard Chan Bioinformatics Core

in collaboration with

HMS Research Computing

Jan 17, 2019

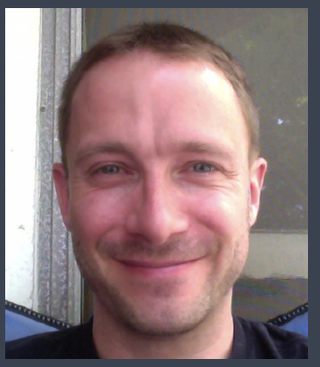
https://tinyurl.com/hbc-shell



Shannan Ho Sui

John Hutchinson

Brad Chapman



Rory Kirchner

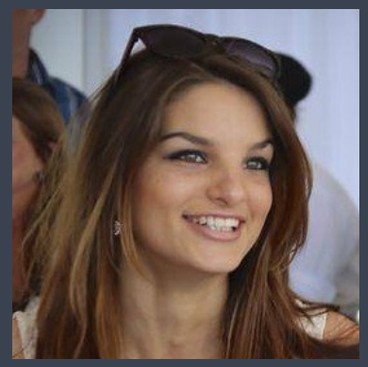
Meeta Mistry

Radhika Khetani

Mary Piper

Lorena Pantano

Victor Barrera

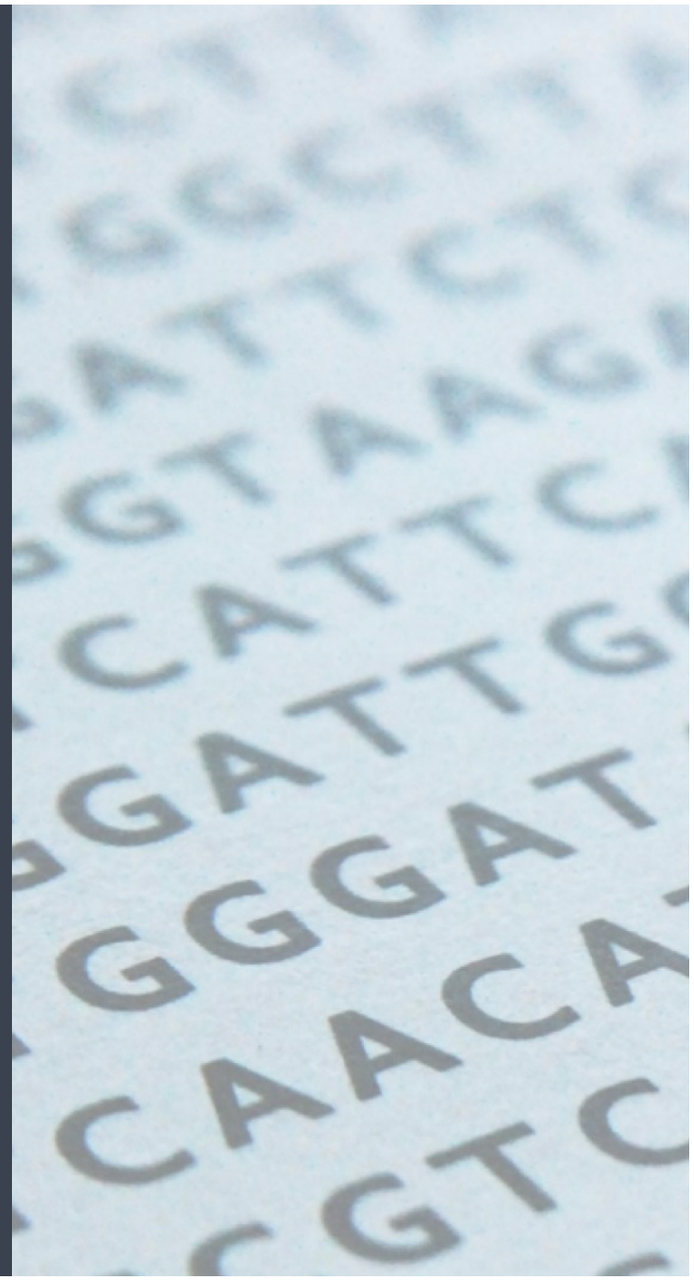


Kayleigh Rutherford

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

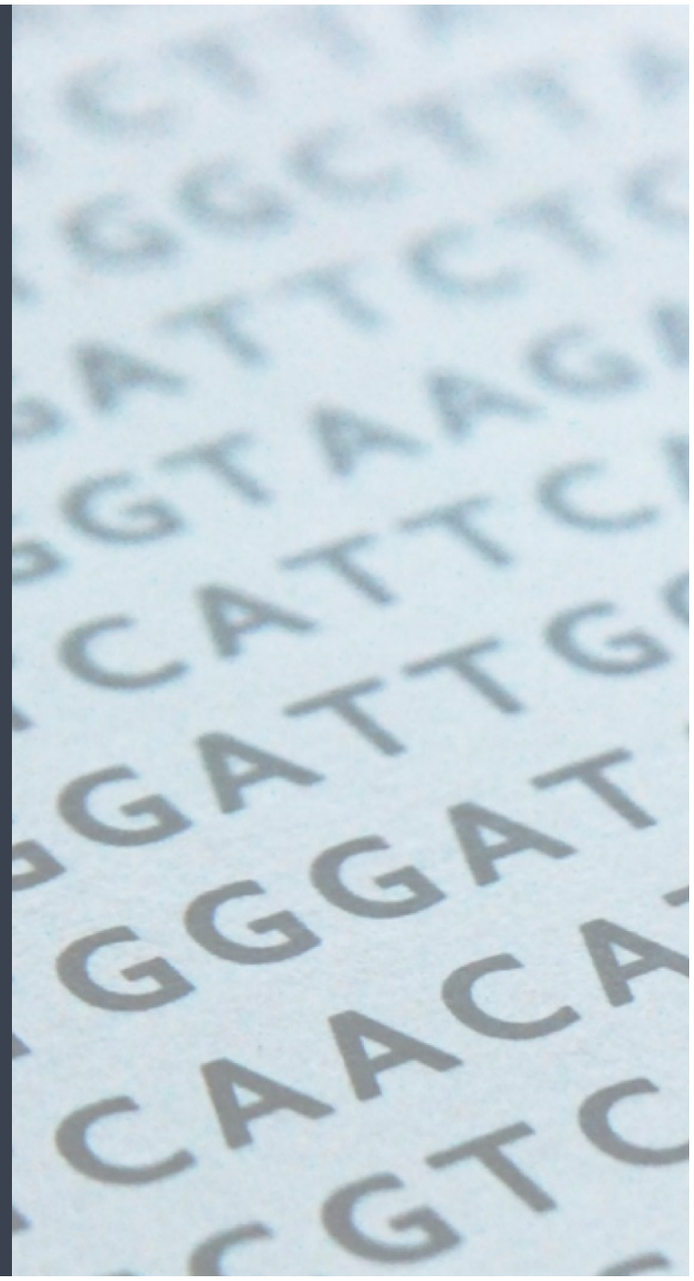


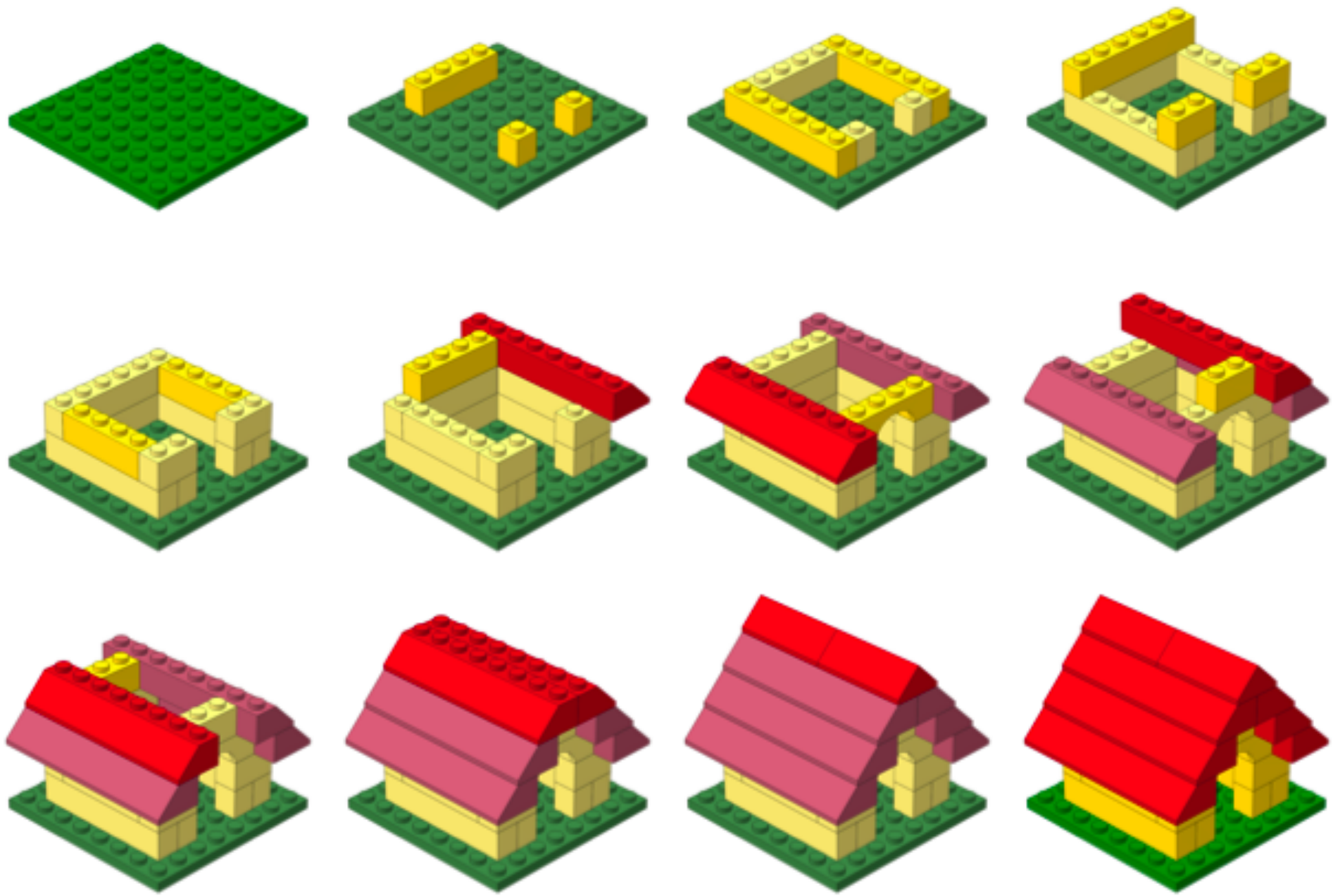
Training

- Short workshops on introductory, intermediate and advanced topics related to NGS data analysis

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

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

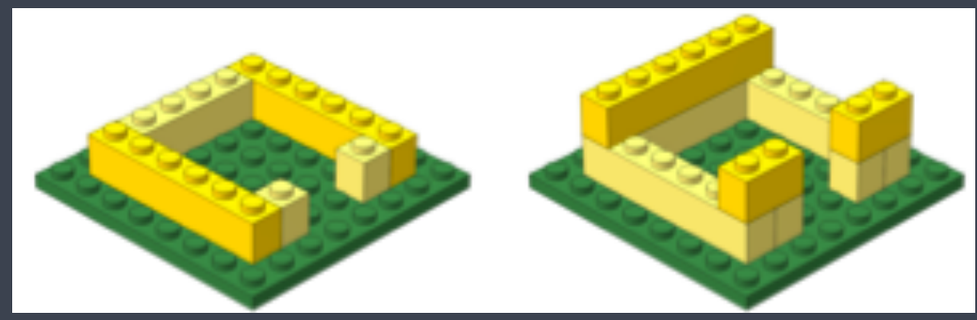




http://anoved.net/tag/lego/page/3/

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

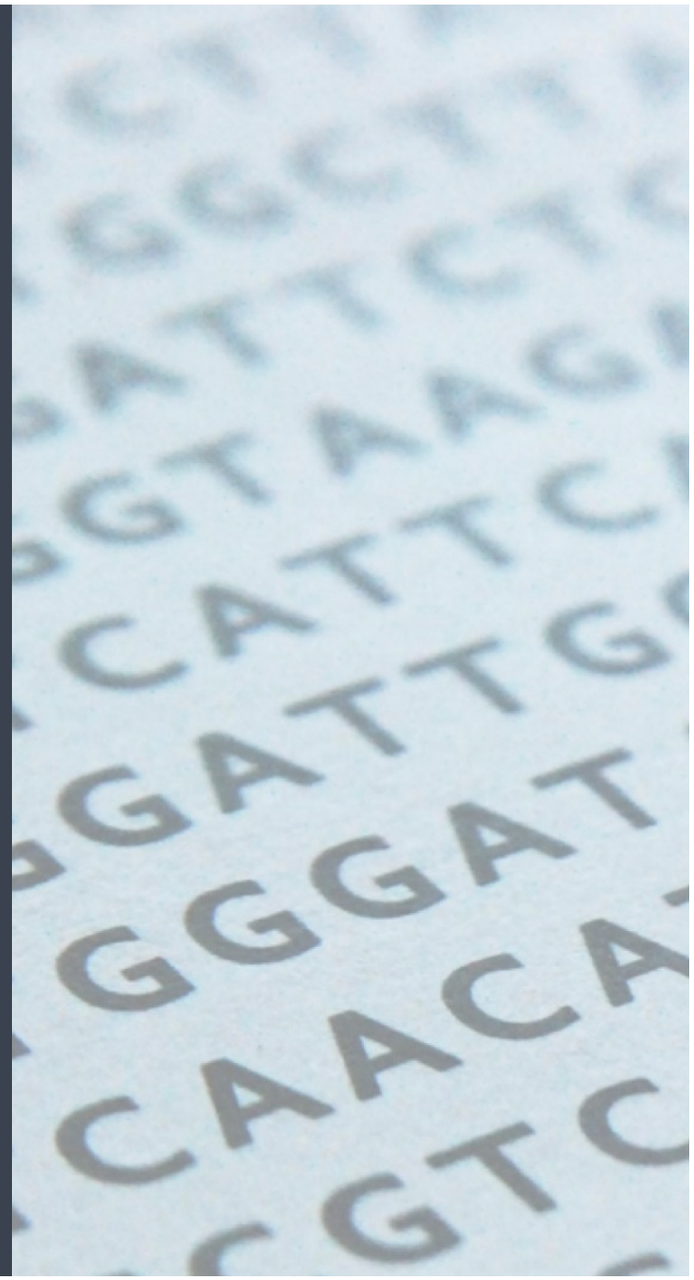
Training

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

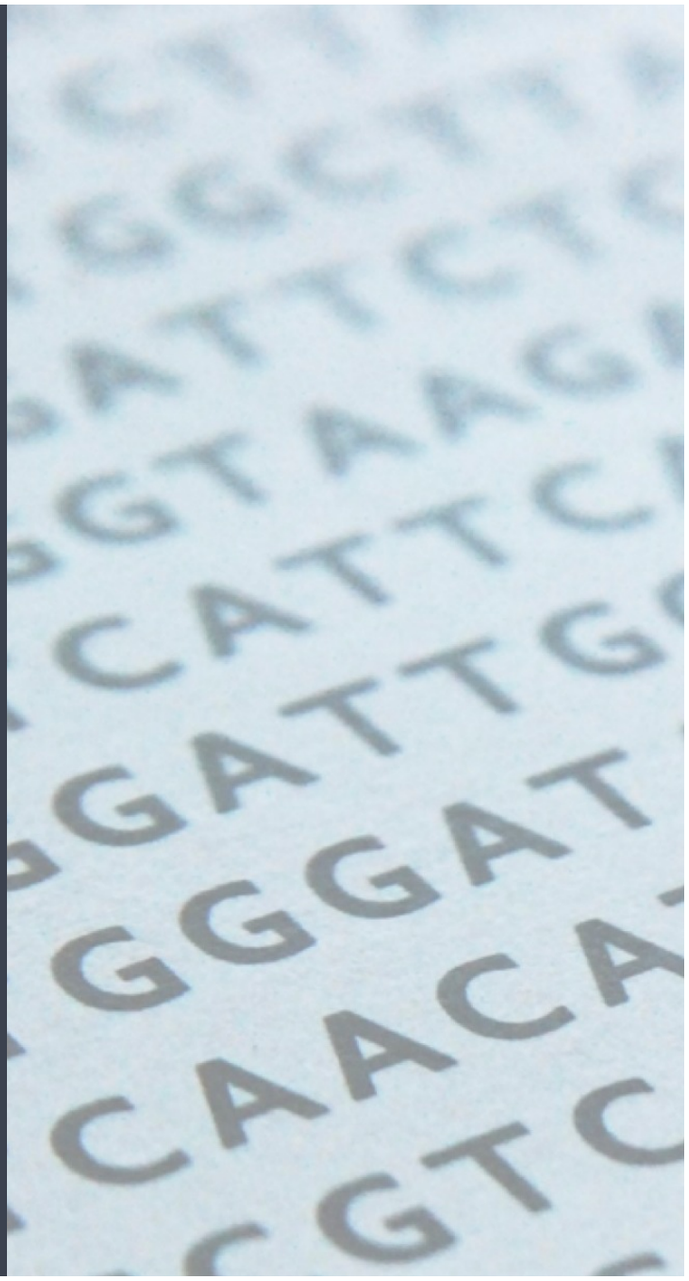


Training

Spring 2019 Schedule

We are structuring the training schedule such that it gives interested researchers several opportunities to take the basic workshops.

Topic	Category	Date	Duration	Prerequisites
Introduction to the command-line interface (shell)	Basic	January 17th	1 day	None
Introduction to R	Basic	January 28th & 29th	1.5 days	None
Introduction to the command-line interface (shell)	Basic	February 13th	1 day	None
Introduction to (bulk) RNA-seq	Advanced	Late February	2 days	Intro to shell
Introduction to R	Basic	Early March	1.5 days	None
Introduction to differential gene expression analysis (bulk RNA-seq)	Advanced	Late March	2 days	Intro to R
Introduction to the command-line interface (shell)	Basic	Early April	1 day	None
Introduction to ChIP-seq analysis	Advanced	Late April	2 days	Intro to R + shell

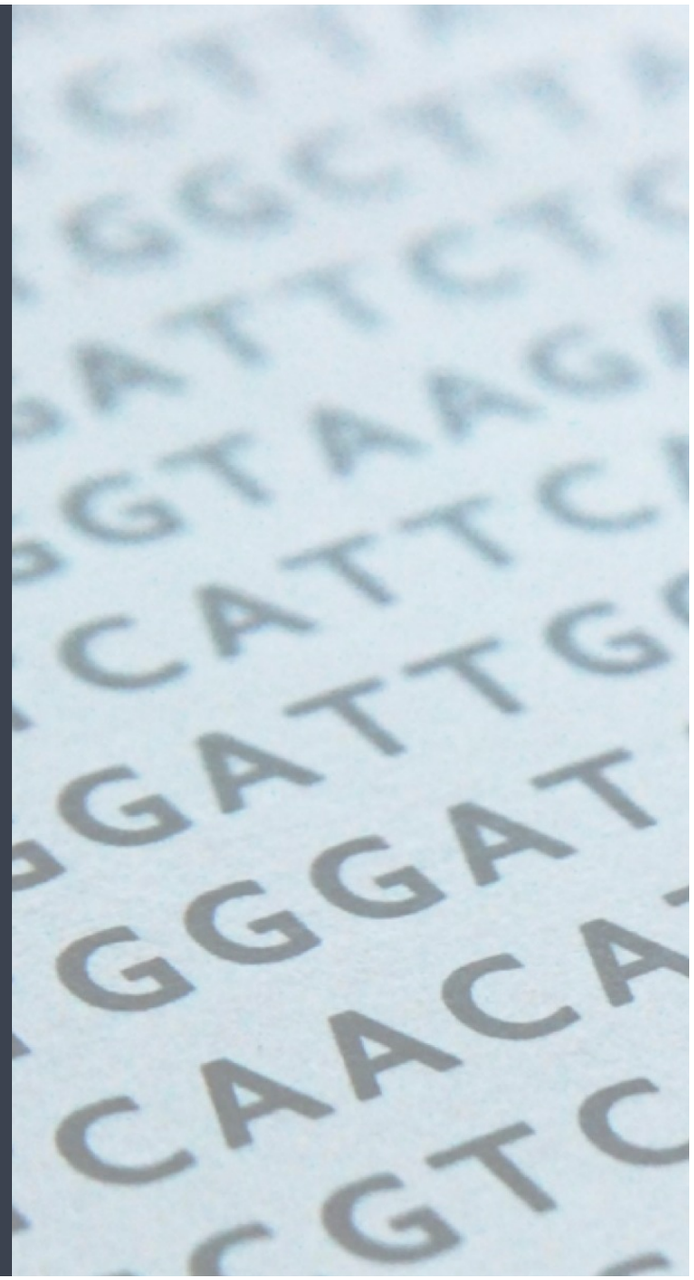


Training

- Short workshops on introductory, intermediate and advanced topics related to NGS data analysis
- Monthly, 2-3 hour, hands-on and free workshops on “Current Topics in Bioinformatics”

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

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

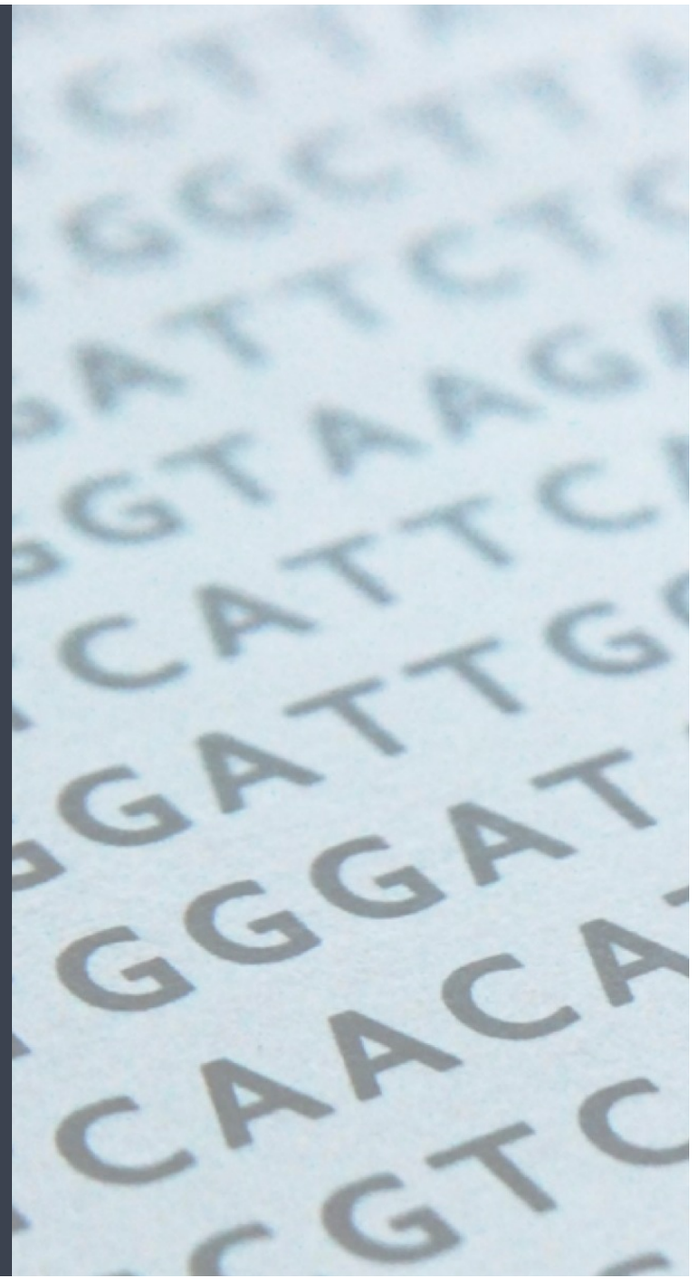


Training

https://tinyurl.com/spring2019-modules

Spring 2019 Schedule:

Topic	Date	Time	Location	Pre requisites
Introduction to R	January 10th	1 PM	Kresge (HSPH), room 502	None
Introduction to tidyverse and visualizations with ggplot2	February 6th	1 PM	FXB (HSPH), room G12	Beginner R
Gene annotations and functional analysis of gene lists	March 6th	1 PM	FXB (HSPH), room G12	Beginner R
Generating research analysis reports with RMarkdown	April 3rd	1 PM	FXB (HSPH), room G12	Beginner R

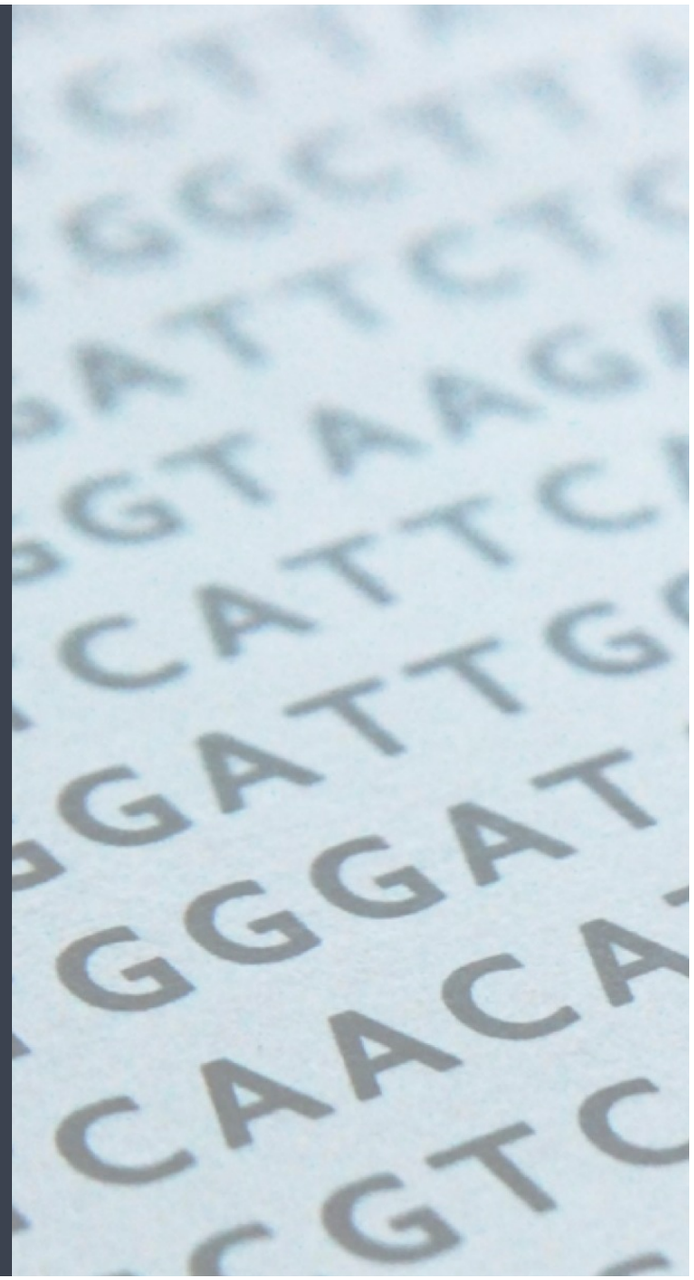


Training

- Short workshops on introductory, intermediate and advanced topics related to NGS data analysis
- Monthly, 2-3 hour, hands-on and free workshops on “Current Topics in Bioinformatics”
- In-depth courses (8- or 12-day formats)

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

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



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

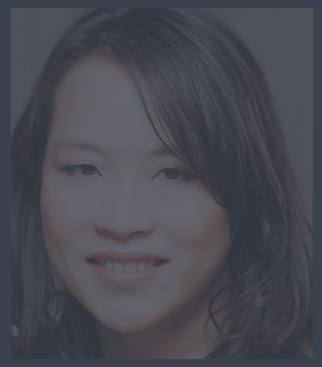
NIEHS / CFAR Bioinformatics Core

Center for Stem Cell Bioinformatics

Harvard Catalyst Bioinformatics Consulting

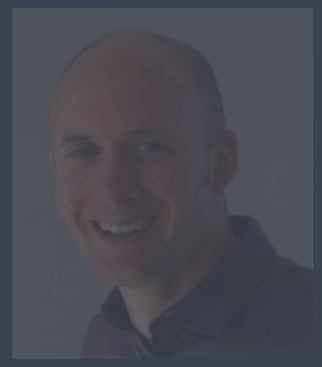
HMS Tools & Technology

Introductions!

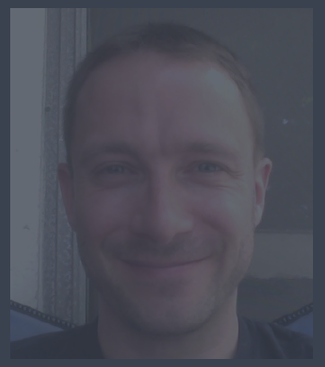


Shannan Ho Sui

John Hutchinson



Brad Chapman

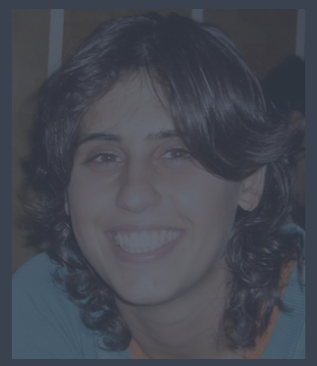


Rory Kirchner

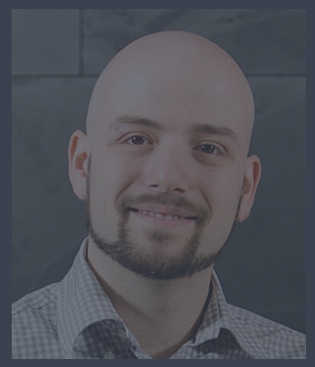
Meeta Mistry

Radhika Khetani

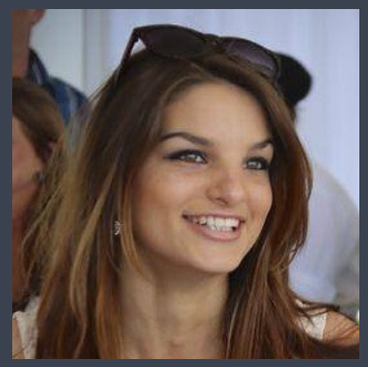
Mary Piper



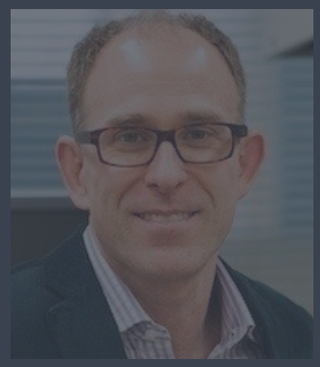
Lorena Pantano



Victor Barrera

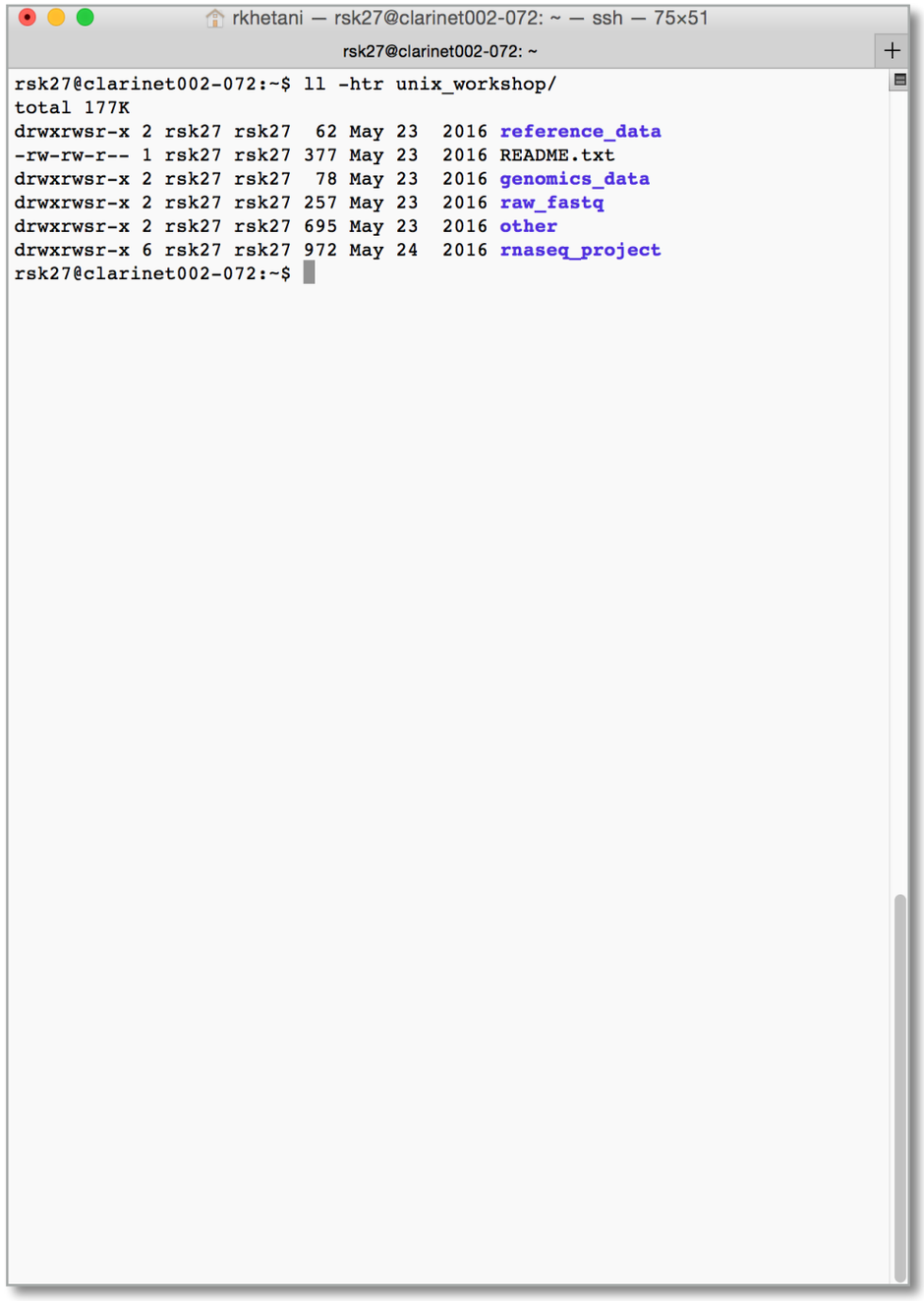


Kayleigh Rutherford



Peter Kraft

Workshop scope



“Unix is user-friendly.

It's just very selective about who its friends are.”

http://en.wikipedia.org/wiki/Unix, http://en.wikipedia.org/wiki/Tux

The Unix command-line interface

- Unix is a stable, efficient and powerful operating system
- It can easily coordinate the use and sharing of a computer's (or a system's) resources, i.e. built to allow multi-user functionality
- Can easily handle complex and repetitive tasks easily on large and small datasets
- Usually, written commands are used to work with this OS, instead of the pointing and clicking used with operating systems like Windows and OSX

The Unix command-line interface

- Unix is a stable, efficient and powerful operating system
- It can easily coordinate the use and sharing of a computer's (or a system's) resources, i.e. built to allow multi-user functionality
- Can easily handle complex and repetitive tasks easily on large and small datasets
- Usually, written commands are used to work with this OS, instead of the pointing and clicking used with operating systems like Windows and OSX

Bioinformatics:

- A lot of NGS-analysis tools are created for the Unix OS
- High-performance compute clusters which are necessary to analyze large datasets require a working knowledge of Unix

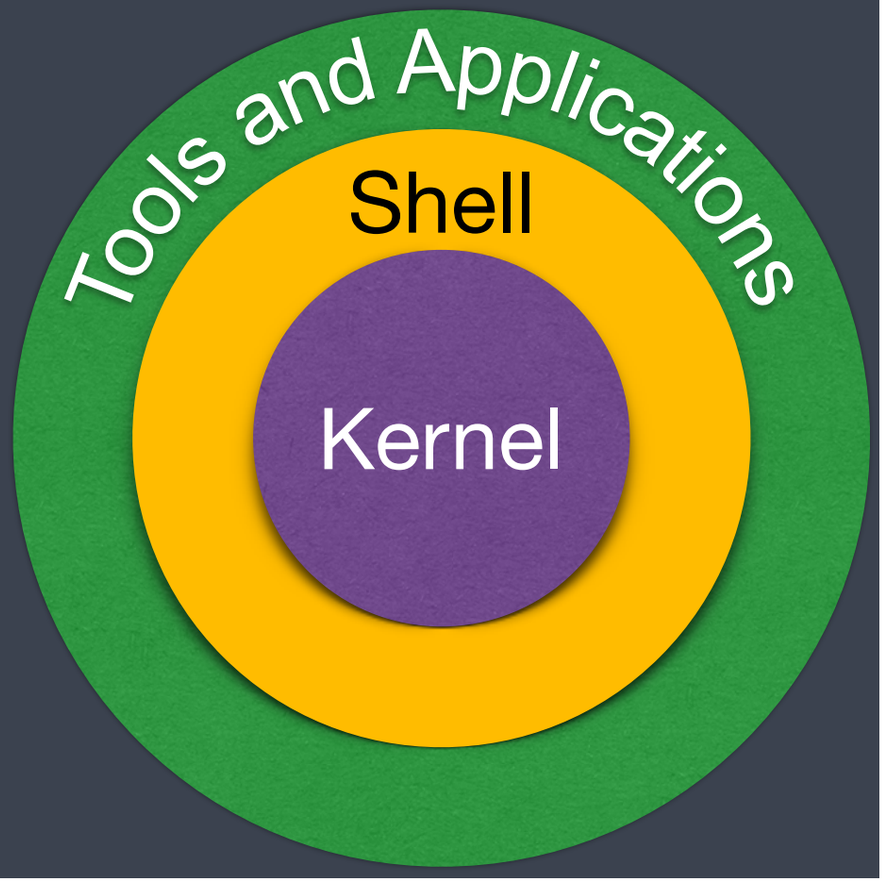
Linux

- Linux is a free, open-source operating system based on Unix
- It has the same components as the original, but the open source community is involved in active development of various distinct distributions of Linux

Components

The Unix/Linux system is functionally organized at 3 levels:

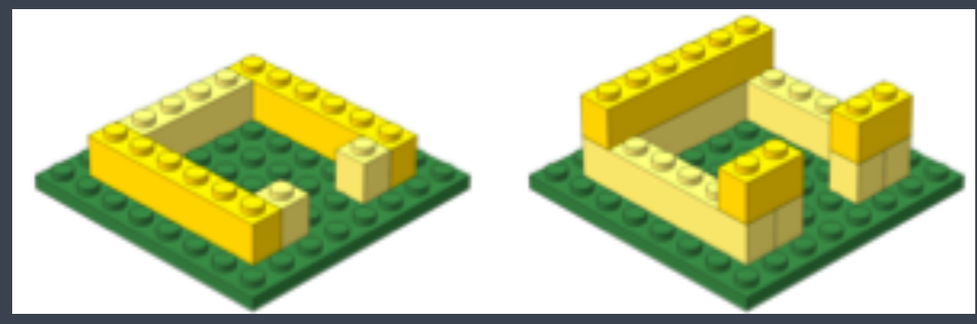
- The kernel, which schedules tasks and manages storage: ***the brain of the system***
- The shell, ***an interpreter*** that helps interprets our input for the kernel
- Utilities, tools and applications, which use the shell to communicate with the kernel



The “shell”

- The shell is **an interpreter**
- It is independent of the operating system
- Dozens of shells have been developed throughout UNIX history, and a lot of them are still in use
- The most commonly used shell is **bash**

Learning Objectives



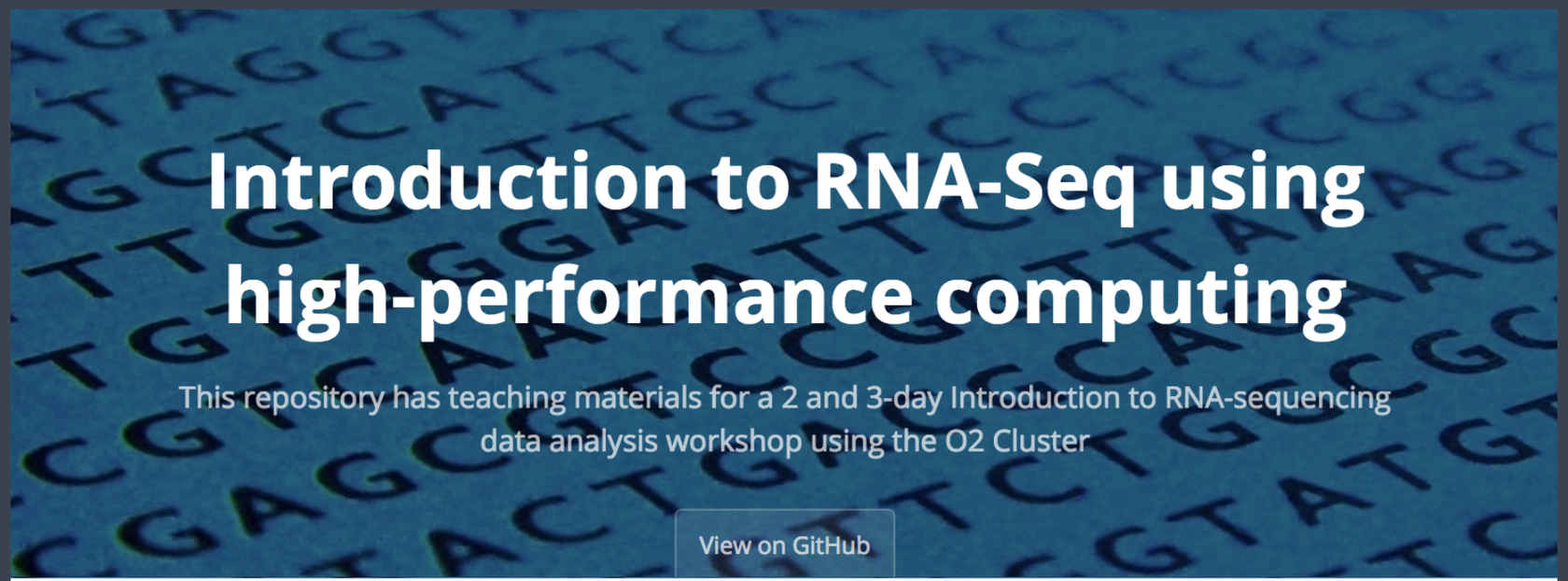
- ✓ Learn what a “shell” is and become comfortable with the command-line interface
  - Find your way around a filesystem using written commands
  - Work with small and large data files
  - Become more efficient when performing repetitive tasks
- ✓ Understand what a computational cluster is and why we need it

Logistics

Course webpage (wiki)

https://tinyurl.com/hbc-shell

Course materials online



Learning Objectives

- How do you access the shell?
- How do you use it?
  - Getting around the Unix file system
  - looking at files
  - manipulating files
  - automating tasks
- What is it good for?

Setting up

We will spend most of our time learning about the basics of the shell by exploring experimental data.

Since we are going to be working with this data on our remote server, **Orchestra 2 (O2)**, we first need to log onto the server. After we're logged on, we will each make our own copy of the example data folder.

The 2 Window problem…

rkhetani — rsk27@clarinet002-072: ~ — ssh — 141×21

rsk27@clarinet002-072: ~

```
rsk27@clarinet002-072:~$ ll -htr unix_workshop/
total 177K
drwxrwsr-x 2 rsk27 rsk27  62 May 23  2016 reference_data
-rw-rw-r-- 1 rsk27 rsk27 377 May 23  2016 README.txt
drwxrwsr-x 2 rsk27 rsk27  78 May 23  2016 genomics_data
drwxrwsr-x 2 rsk27 rsk27 257 May 23  2016 raw_fastq
drwxrwsr-x 2 rsk27 rsk27 695 May 23  2016 other
drwxrwsr-x 6 rsk27 rsk27 972 May 24  2016 rnaseq_project
rsk27@clarinet002-072:~$
```

Starting with the shell

We have each created our own copy of the example data folder into our home directory, **unix_workshop**. Let's go into the data folder and explore the data using the shell.

```
$ cd unix_workshop
```

> 'cd' stands for 'change directory'

Let's see what is in here. Type:

```
$ ls
```

Odds and Ends

- Name tags: Tent Cards
- Post-its
- Phones on vibrate/silent

Thanks!

- **Kristina Holton** and **Andy Bergman** from HMS-RC
- Data Carpentry

Contact us!

HBC training team: hbctraining@hsph.harvard.edu

O2 (HMS-RC): rchelp@hms.harvard.edu

HBC consulting: bioinformatics@hsph.harvard.edu

Twitter

HBC: @bioinfocore

HMS-RC: @hms_rc