

# Introduction to R

Harvard Chan Bioinformatics Core

<https://tinyurl.com/intro-to-r-1-5>

Sponsored by HMS (TnT) and HSCI



Shannan Ho Sui  
*Director*



John Hutchinson  
*Associate Director*



Victor Barrera



Rory Kirchner



Zhu Zhuo



Preetida Bhetariya



Meeta Mistry



Mary Piper



Jihe Liu



Radhika Khetani  
*Training Director*



Ilya Sytchev



James Billingsley



Sergey Naumenko



Joon Yoon

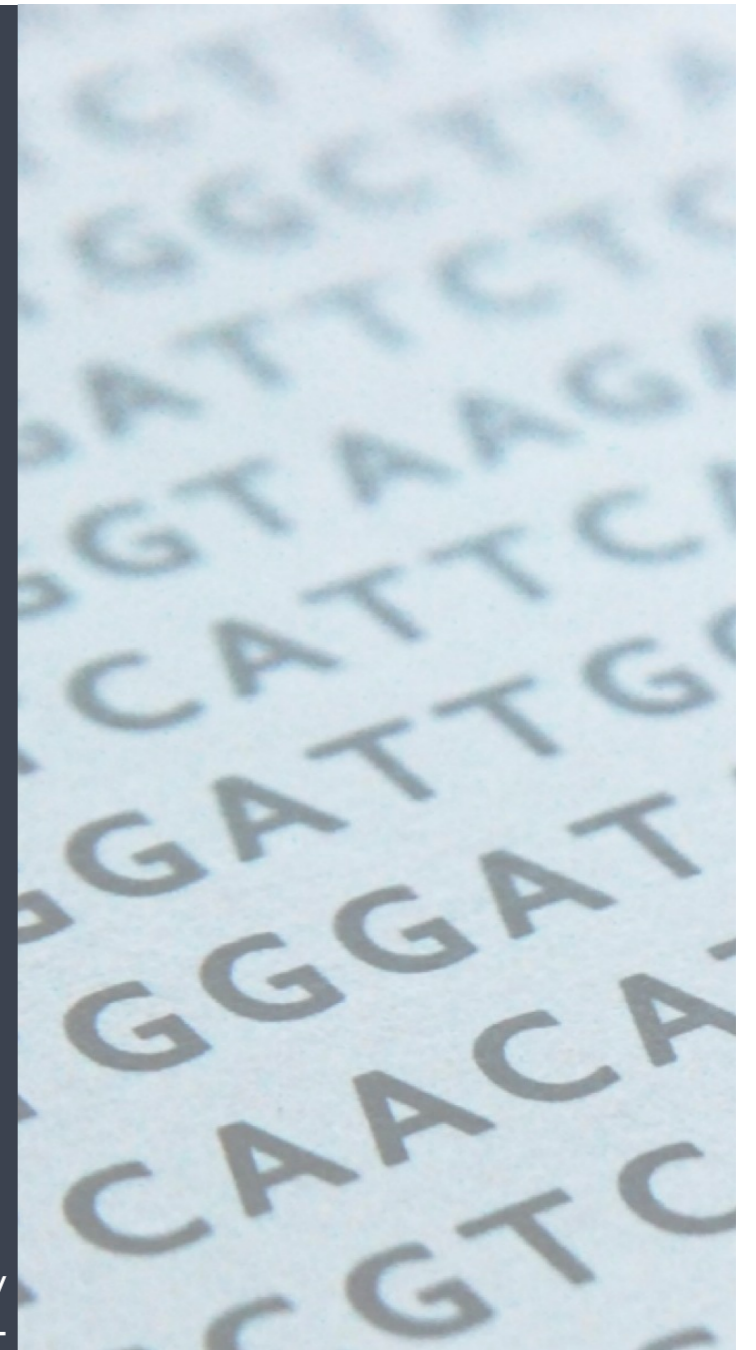


Peter Kraft  
*Faculty Advisor*

# Consulting

- RNA-seq analysis: bulk, single cell, small RNA
- ChIP-seq and ATAC-seq analysis
- Genome-wide methylation
- WGS, resequencing, exome-seq and CNV studies
- QC & analysis of gene expression arrays
- Functional enrichment analysis
- Grant support

<http://bioinformatics.sph.harvard.edu/>



# Training

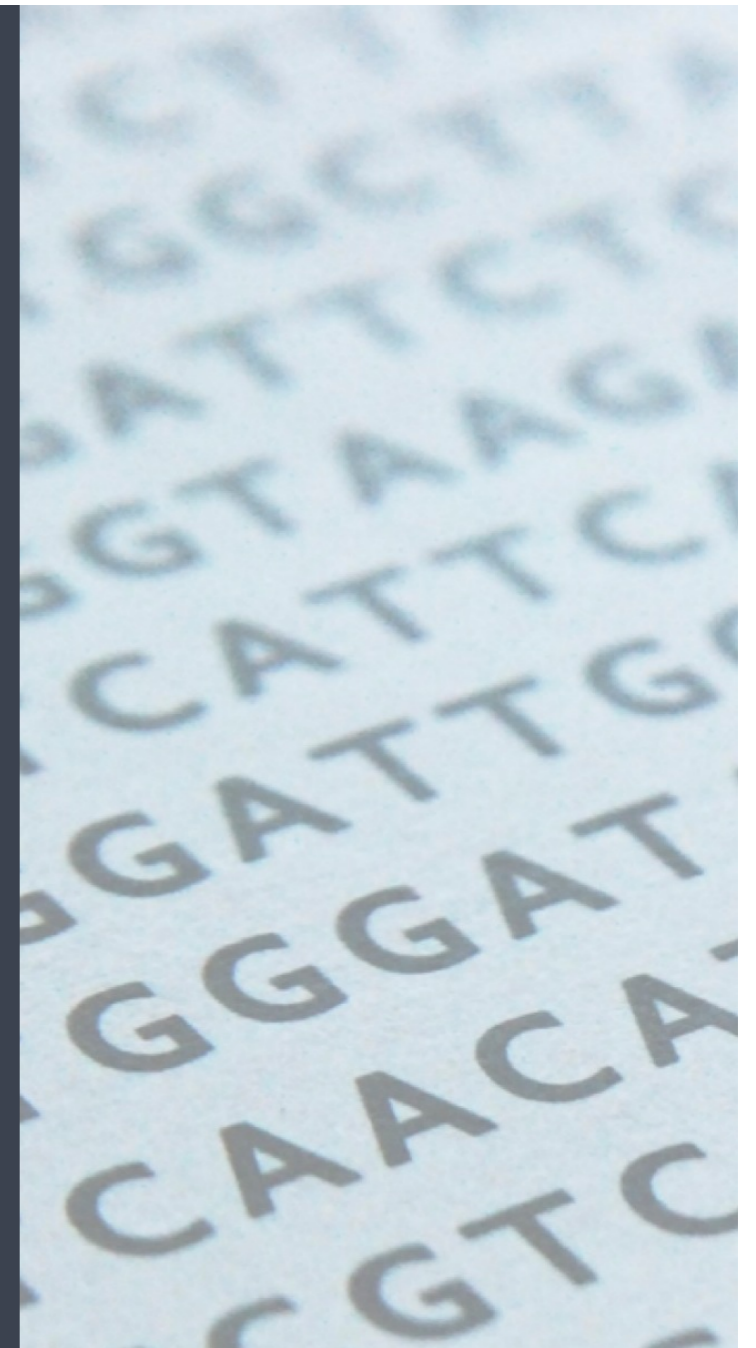
We have divided our short workshops into 2 categories:

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.*

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

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





**HARVARD**  
**T.H. CHAN**

**SCHOOL OF PUBLIC HEALTH**

**HSCI**  
HARVARD STEM CELL  
INSTITUTE



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  
Committee



**HARVARD**  
**T.H. CHAN**

SCHOOL OF PUBLIC HEALTH

**HSCI**  
HARVARD STEM CELL  
INSTITUTE



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  
Committee

# Class Introductions!



Shannan Ho Sui  
*Director*



John Hutchinson  
*Associate Director*



Victor Barrera



Rory Kirchner



Zhu Zhuo



Preetida Bhetariya



Meeta Mistry



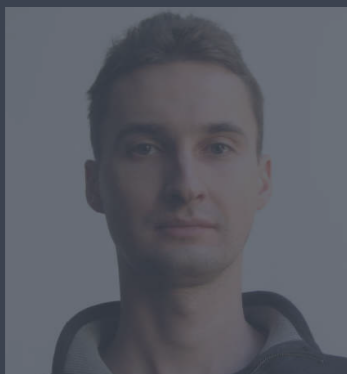
Mary Piper



Jihe Liu



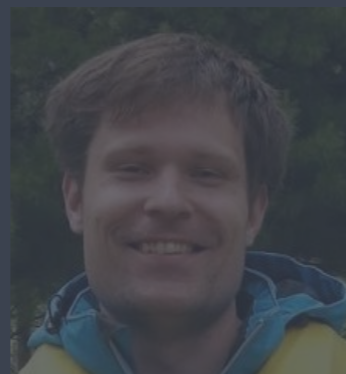
Radhika Khetani  
*Training Director*



Ilya Sytchev



James Billingsley



Sergey Naumenko



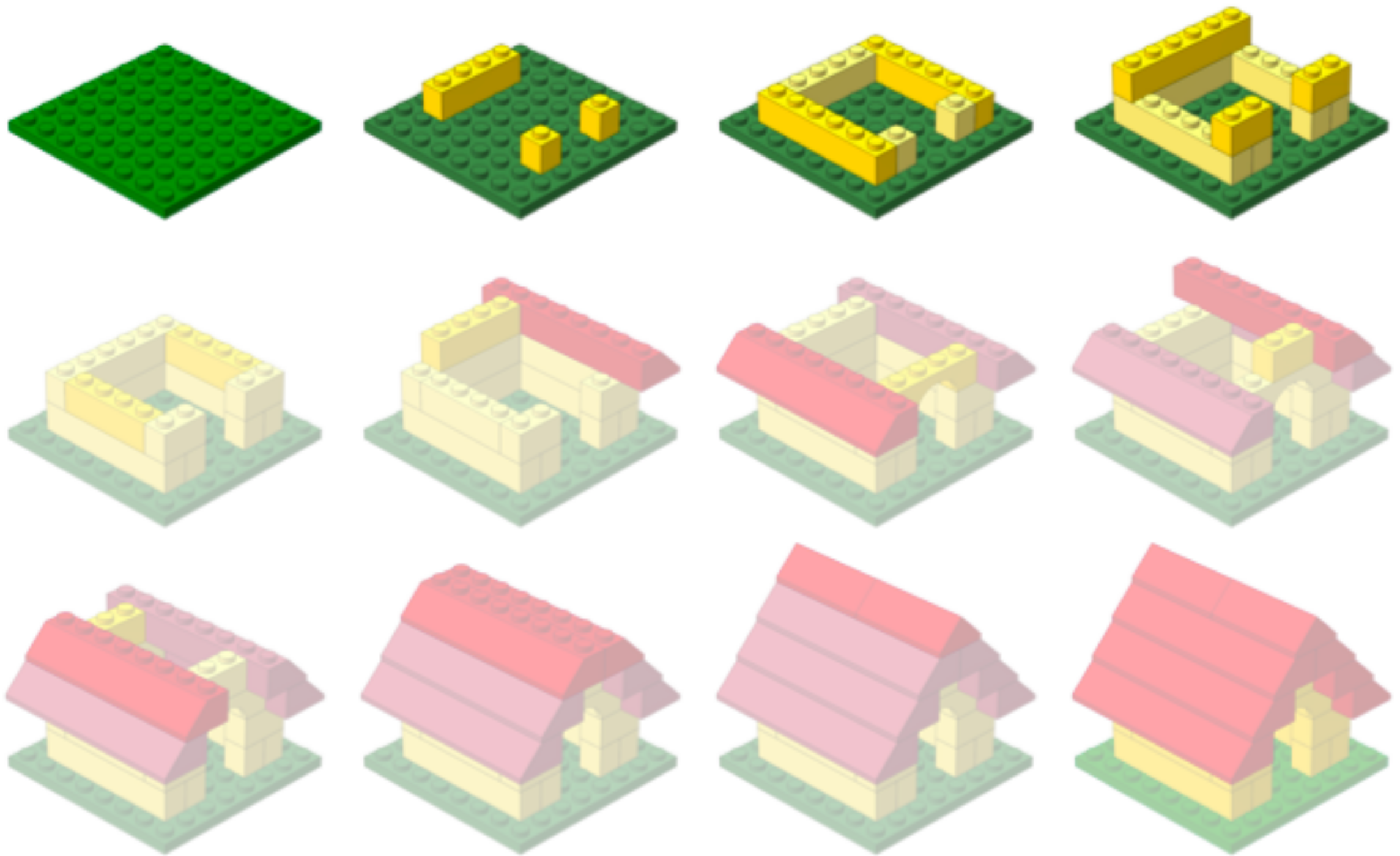
Joon Yoon



Peter Kraft  
*Faculty Advisor*

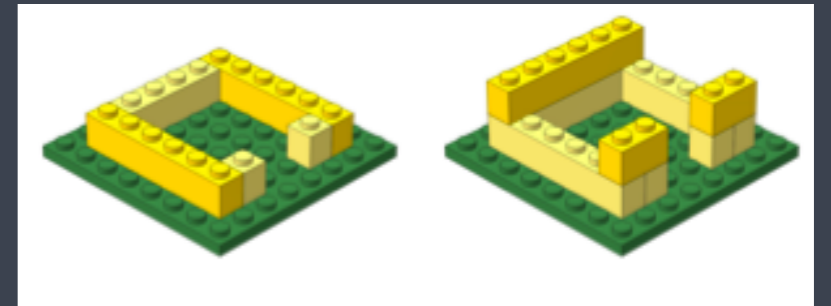


Workshop Scope...



# Learning R

# Workshop Scope



- ✓ Comfortably use RStudio (a graphical interface for R)
- ✓ Fluently interact with R using RStudio
- ✓ Become familiar with R syntax
- ✓ Understand data structures in R
- ✓ Inspect and manipulate data structures
- ✓ Install packages and use functions in R
- ✓ Visualize data using *ggplot2*

# Logistics

# Course webpage (wiki)

<https://tinyurl.com/intro-to-r-1-5>

# Course materials online

## Introduction to R

[View on GitHub](#)

Approximate time: 70 min

### Learning Objectives

- Employ variables in R.
- Describe the various data types used in R.
- Construct data structures to store data.

### The R syntax

Now that we know how to talk with R via the script editor or the console, we want to use R for something more than adding numbers. To do this, we need to know more about the R syntax.

Below is an example script highlighting the many different “parts of speech” for R (syntax):

- the **comments** `#` and how they are used to document function and its content
- **variables** and **functions**
- the **assignment operator** `<-`

# The 2 Window problem...

The screenshot shows the RStudio interface with the following components:

- Script Editor:** Contains R code for creating a dataframe and a vector.
- Console:** Shows the R prompt and workspace loading message.
- Environment Pane:** Lists loaded objects: 'animals' (6 obs. of 4 variables) and 'behavior' (6 obs. of 3 variables).
- Files Pane:** Shows a directory listing of files in the current project.

```
351
352 animals[4, 2] <- "Gray"
353
354 animals$color <- factor(animals$color)
355 animals$new2 <- c(1,2,3)
356
357 vector1 <- c(6:11)
358 data.frame(animals[, 1:2], vector1, animals[, 3:4])
359
```

Console output:

```
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[Workspace loaded from ~/Dropbox (HBC)/HBC Team Folder (1)/Teaching/Intro-to-R/.RData]
>
```

Name	Size	Modified
..		
.RData	5.8 MB	May 3, 2018, 1:41
.Rhistory	17.4 KB	Nov 15, 2018, 1:21
data		
de_sleuth.R	2.6 KB	Oct 10, 2018, 10:11
figures		
Intro-to-R.R	11.9 KB	May 1, 2018, 3:31

```
rownames(metadata)
metadata[c("sample10", "sample12"),]
```

Selecting using indices with logical operators

With dataframes, similar to vectors, we can use logical vectors for specific columns in the dataframe to select only the rows in a dataframe with TRUE values at the same position or index as in the logical vector. We can then use the logical vector to return all of the rows in a dataframe where those values are TRUE.

# Odds and Ends

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



# Contact us!

*HBC training team:* [hbctraining@hsph.harvard.edu](mailto:hbctraining@hsph.harvard.edu)

*HBC consulting:* [bioinformatics@hsph.harvard.edu](mailto:bioinformatics@hsph.harvard.edu)

Twitter

[@bioinfocore](https://twitter.com/bioinfocore)