Intro to R

Nick Mancuso
Workshop logistics

Target: beginners
• most of you indicated zero/little programming experience
• a few have a stronger background – please be patient

Questions
• ask anytime
• don't be shy, if you miss some steps you soon won't be able to follow anything else

Time
• 2-3 hrs/day; 1 break if we go past 2 hrs
• 6-9 hrs overall: reasonable goal is to get you started, overcoming fear of programming
Workshop approach

- Zillions of resources/websites/slides in Internet on R
  - my suggestion: Quick-R [www.statmethods.net](http://www.statmethods.net)
    - comes with 38% discount on useful book
    - but website is sufficient
  - yet another one?
- Practice vs Slides
  - you can learn philosophy by reading, but to learn math you need to do your homework
  - programming is like math
  - we know you are a busy student/postdoc with no time for homework
  - so let's do the homework HERE!
Workshop materials: R file

- commands and explanation next to each other
- can be executed directly
- keep it as your reference for the future, and extend it when you learn new functions
- build your own reference during class
- to program, you need to **write** code
  - please please please type, do not copy/paste
- sometimes I will introduce errors purposely
  - I need you to spot the error (= debugging)
Why R for data mining?

• R language per se is not really elegant (personal opinion, but shared by many)
• Interactive
• Open source
• Thousands of ready-to-go functions
  • especially statistical packages
• Active community
Intro to R – Day 1

- Simple math
- Variables
- Vectors
- Indexing
- Matrices
- Types
- Boolean operations
- NAs
- Data frames
- Loading and saving data
Intro to R – Day 2

- Getting help
- Package installation
- Factors
- Lists
- Conditionals
- Loops
- Apply functions
- Custom functions
- Correlation
- t-tests
- Multiple comparison correction
• Bioconductor's edgeR for differential gene expression in RNAseq
• Bioconductor annotations to link gene names to gene IDs
• Plotting results as Volcano plots
• Hierarchical clustering and heat maps
Intro to R – Day 1

- Simple math
- Variables
- Vectors
- Subsetting
- Matrices
- Types
- Boolean operations
- NAs
- Data frames
- Loading and saving data