Intro to R and Rstudio

UT Spring 2016 Biocomputing
Sean Leonard
3/30/16

What is  ?

- A statistical programming language based on “S” from Bell labs

- First appeared in 1993, now a GNU project and completely open source

- Managed by CRAN (Comprehensive R Archive Network) - https://cran.r-project.org

Why not  ?

- Syntax less immediately intuitive than Python

- Some decisions made to support statistics do not make broad sense for data analysis (stringsAsFactors….)

- A little quirky- some things that rub experienced programmers the wrong way: ‘<-‘ for assignment

Why  ?

- Extremely powerful and easy to use for data analysis and beautiful visualization

- Huge community of scientists and academics that develop packages (probably within your niche field)

- Good knowledge base here at UT

- “Bioconductor” is a separate repository from CRAN, specializing in packages for biological data analysis (sequencing and non-sequencing)

- Lots of other “data scientists” use R right now, so there are a lot of introductory courses and learning resources out there

- Powerful IDE (Integrated development environment) for R

- Supports code completion, integrated help and plotting

- **easy to script and save your workflow**

- (also integrates with git and Rmarkdown, if you like that kind of thing)

RStudio

Go to file/function

Project: (None)

ay_play_in_R.R | fig4_pathway_heatmaps.R | 2016_03_22_reporting_tools.R* | Untitled25*

Source on Save | Run | Source

```
16  full_data <- read_tsv("results_combined.tsv")
17  #keep only variables we care about
18  full_data <- full_data %>% select(old_locus_tag, Fitness.log2FoldChange, RNAseq.l2fc) %
    >% mutate(Fitness.scaled = as.numeric(scale(Fitness.log2FoldChange)), RNAseq.scaled = as
    .numeric(scale(RNAseq.l2fc)), id = old_locus_tag) %>% select(id, Fitness.scaled, RNAseq
    .scaled)
19  #make wide data long for better plotting
20  melted <- full_data %>% gather("measurement", "result", 2:3)
21  ##use KEGGREST to pull gene lists for modules/pathways
22
23  #supply pathway number for ones you care about
24  pathway <- "salv00290"
25  query <- keggGet(pathway)
26  #get genes
27  pathway_genes <- query[[1]]$GENE
28  #use stringr to match SALWKB2 to return "old locus tags"
29  grep <- "SALWKB2_...."
30  gene_list <- str_subset(pathway_genes, grep)
31  reduced_data <- melted %>% filter(id %in% gene_list)
32  heatmap <- ggplot(reduced_data, aes(x = measurement, y = id))
33  heatmap + geom_tile(aes(fill = result)) + scale_fill_distiller(type = "div", palette = 3
    , limits = c(-4,4)) + ylab("Locus Tag") + xlab("")
34
35  ##method to plot pathway
```

35:1 (Top Level) R Script

Console ~/Dropbox/articles/2015_tnseq_snod/analysis_rerun/

```
> query <- keggGet(pathway)
> #get genes
> pathway_genes <- query[[1]]$GENE
> #use stringr to match SALWKB2 to return "old locus tags"
> grep <- "SALWKB2_...."
> gene_list <- str_subset(pathway_genes, grep)
> reduced_data <- melted %>% filter(id %in% gene_list)
> heatmap <- ggplot(reduced_data, aes(x = measurement, y = id))
> heatmap + geom_tile(aes(fill = result)) + scale_fill_distiller(type = "div", palette = 3, lim
its = c(-4,4)) + ylab("Locus Tag") + xlab("")
>
```

Environment | History

Import Dataset | List

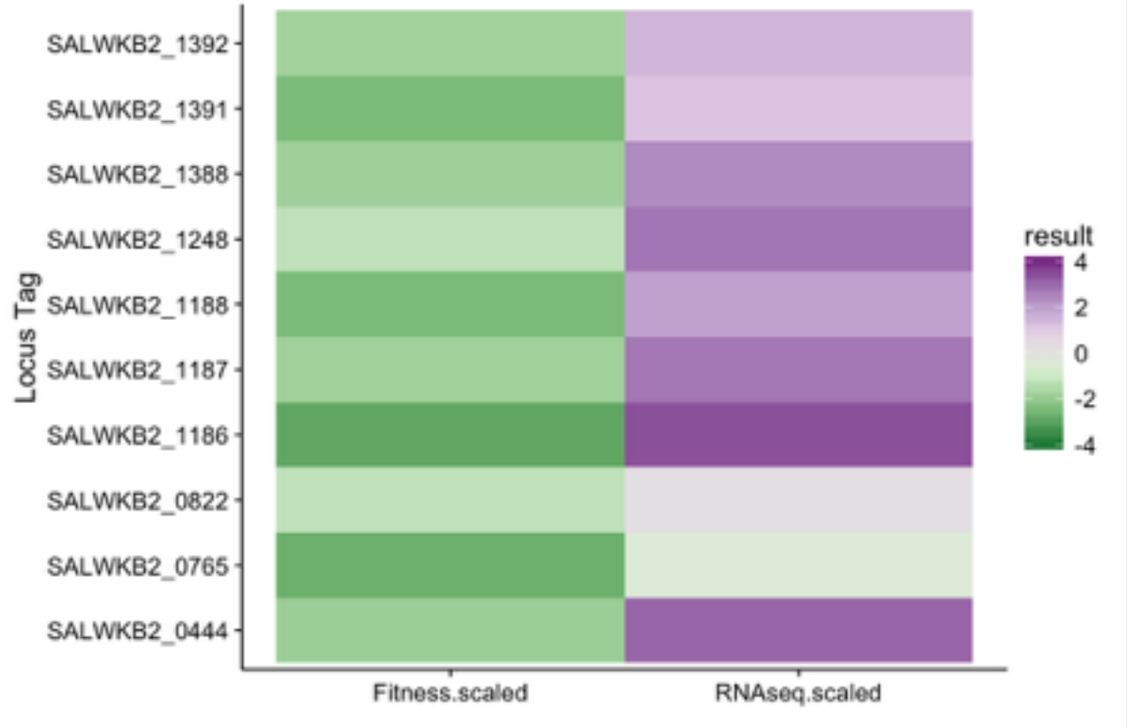
Global Environment

```
    normal.2 : Factor w/ 1566 levels "0","1","1.1.1.102",..: 263 1 1488 1 …
    gingivitis.1: Factor w/ 1503 levels "0","1","1.1.1.102",..: 916 1 1 1 …
    gingivitis.2: Factor w/ 1029 levels "0","1","1.1.1.102",..: 698 1 2 1 …
```

full_data	2298 obs. of 3 variables
melted	4596 obs. of 3 variables
reduced_data	20 obs. of 3 variables
Values	
gene_list	chr [1:10] "SALWKB2_0822" "SALWKB2_1388" "SALWKB2_1…
grep	"SALWKB2_...."
heatmap	List of 9

Files | Plots | Packages | Help | Viewer

Zoom | Export | Publish



Learning R: Resources

- "Swirl" - Learn R, in R. http://swirlstats.com
- Code School. https://www.codeschool.com/courses/try-r
- Coursera- Specialization in Data Science: https://www.coursera.org/specializations/jhu-data-science
- Software Carpentry Lessons - http://software-carpentry.org/lessons/