

# TRGN 599: Applied Data Science and Bioinformatics

## Publish Rmarkdowns in ITG

### Week 14 - Lecture 2

**Enrique I. Velazquez Villarreal, M.D., Ph.D., M.P.H., M.S. | Assistant Professor**

Dept. of Translational Genomics

USC | Keck School of Medicine | Norris Comprehensive Cancer Center

Leader of the USC Bioinformatics Core – *USC CaRE2 Health Equity Center*

**David W. Craig, Ph.D. | Professor and Vice Chair**

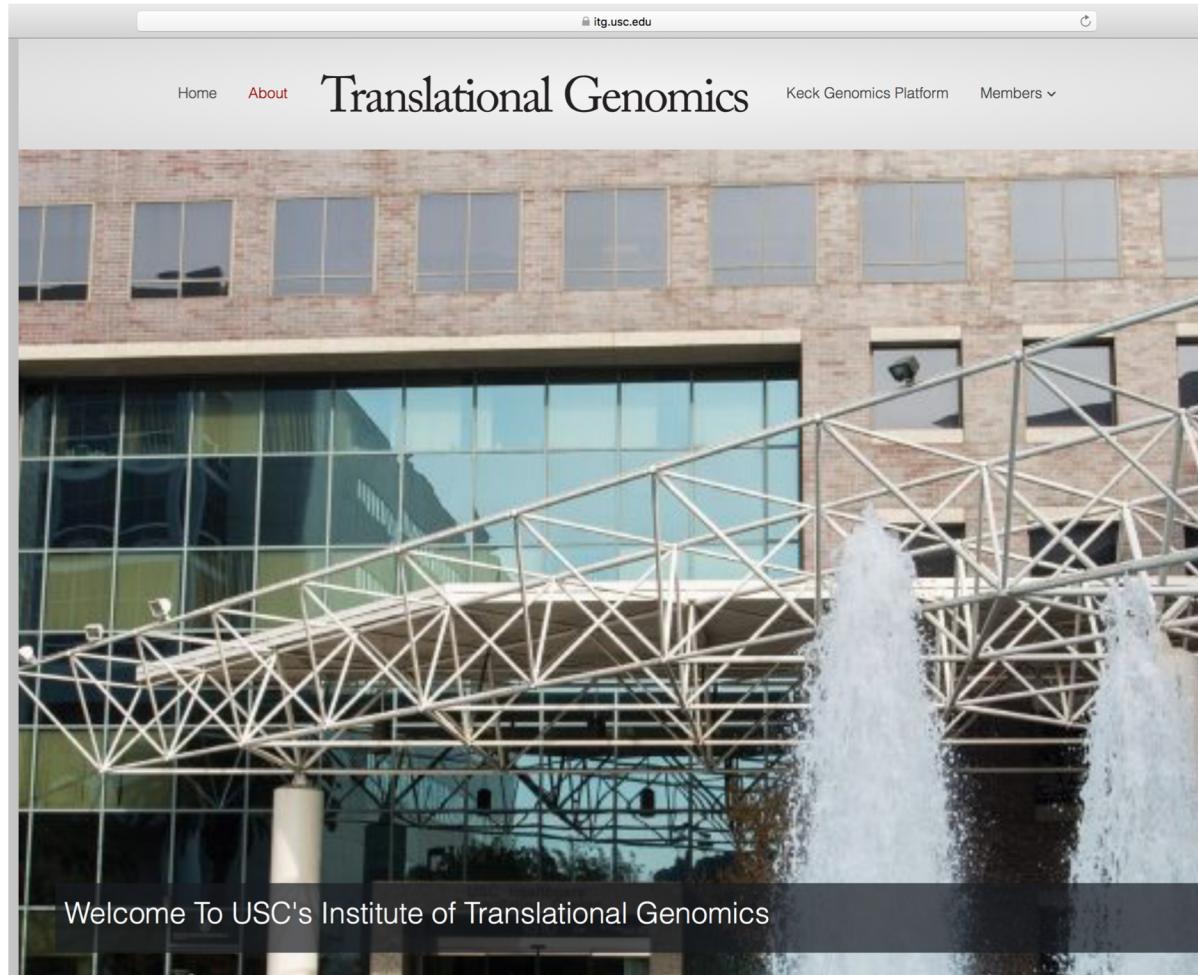
Dept. of Translational Genomics

USC | Keck School of Medicine | Norris Comprehensive Cancer Center

Co-Director, Institute of Translational Genomics

# Topics

- Publishing your Rmarkdown in our WEBSITE itg.usc.edu



# Steps to publish your HTML Rmarkdowns in our website

1. You would like to frame your R markdown before knit your code, probably as follow:

---

```
title: "RNAseq_Analysis_YOUR_NAME"
```

```
author: "Student: *YOUR NAME*, Professors: Dr. Enrique I. Velazquez Villarreal & Dr. David W. Craig "
```

```
date: "4/15/2019"
```

```
output: html_document
```

---

2. Include the **name of the analysis** and **your name** in your already generated **.html** Rmarkdowns

Example: Rmarkdown\_TRGN599\_RNAseq\_Analysis\_Enrique\_Velazquez\_Villarreal.HTML

3. Please send me ([eivelazq@usc.edu](mailto:eivelazq@usc.edu)) all your **.html** Rmarkdowns that want to be published in our WEBSITE [itg.usc.edu](http://itg.usc.edu).

# Steps to publish your HTML Rmarkdowns in our website

```
1 ---  
2 title: "Rmarkdown_Linear_Regression_Analysis_*YOUR NAME*"  
3 author: "Student: *YOUR NAME*, Professors: Dr. Enrique I. Velazquez Villarreal & Dr. David W. Craig "  
4 date: "4/15/2019"  
5 output: html_document  
6 ---  
7  
8 ## R Markdown  
9 ## Linear regression analysis  
10  
11 ## In this example we will model the relationship between percent body fat, siri, and height, using a simple linear regression model.  
12  
13 # Installing required packages  
14 # Making the data bodyfat available for analysis  
15 ````{R}  
16 #install.packages("mfp", dependencies = TRUE)  
17 #install.packages("survival")  
18 library(mfp)  
19 data(bodyfat)  
20 ````  
21  
22 # Fitting the least-squares regression model  
23 ````{R}  
24 fit <- lm(siri ~ height, data = bodyfat)  
25 ````  
26  
27 # fit object now stores all the output from the linear regression model  
28 # Get the estimates of the estimates of Alpha and Beta  
29 ````{R}  
30 fit  
31 ````  
32  
33 # Summarizing data  
34 ````{R}  
35 summary(fit)  
36 ````  
37  
38 # Checking all the information information contained in the fit object:  
39 ````{R}  
40 names(fit)  
41 ````  
42  
43 # Accessing coefficients  
44 ````{R}  
45 fit$coefficients  
46 ````  
47  
48 # Accesing the estimates for the first five people  
49 ````{R}  
50 fit$fitted.values[1:5]  
51 ````  
52  
53 # Accesing the residuals of the first five people  
54 ````{R}  
55 fit$residuals[1:5]  
56 ````  
57  
58 # Adding the least-squares line to the scatterplot is easy with the abline() function:  
59 ````{R}  
60 plot(bodyfat$height, bodyfat$siri, main = "Scatterplot for Percent Body Fat by Height", xlab = "Height", ylab = "Percent Body Fat")  
61 abline(fit)  
62 ````  
63  
64 # Multiple linear regression model  
65 # We can also fit a multiple linear regression model to the bodyfat data using abdomen circumference and height as explanatory variables.  
66
```

# Steps to publish your HTML Rmarkdowns in our website

A screenshot of a web browser window. The address bar shows the URL [itg.usc.edu](http://itg.usc.edu). The main content area displays the title "Index of /trgn599/EnriqueVelazquezVillarreal/" in large, bold, black font. Below the title, there is a horizontal line. Underneath the line, there are three entries: a blue link "[..](#)/Rmarkdown\_TRGN599\_Week\_11\_Lecture\_1.html", the date and time "18-Apr-2019 01:33", and the file ID "2766166".

---

[..](#)/  
[Rmarkdown\\_TRGN599\\_Week\\_11\\_Lecture\\_1.html](#)      18-Apr-2019 01:33      2766166

# Rmarkdown\_TRGN599\_Week\_11\_Lecture\_1

Enrique I. Velazquez Villarreal, MD, PhD, MPH, MS

3/25/2019

## Next-Generation Sequencing in Transcriptomics

### RNAseq

#### Preparing count tables

```
# install package: http://bioconductor.org/packages/release/data/experiment/html/RNAseqData.HNRRNPC.bam.chr14.html
# if (!requireNamespace("BiocManager", quietly = TRUE))
#   install.packages("BiocManager")
BiocManager::install("RNAseqData.HNRRNPC.bam.chr14", version = "3.8")

library(RNAseqData.HNRRNPC.bam.chr14)

alnFiles <- RNAseqData.HNRRNPC.bam.chr14_BAMFILES

exp.des <- factor(rep(c('Control','HNRRNPC knockdown'),each=4))

names(exp.des) <- RNAseqData.HNRRNPC.bam.chr14_RUNNAMES
```

#### Alignment files to read counts

```
# install package: http://bioconductor.org/packages/release/bioc/html/GenomicFeatures.html
# if (!requireNamespace("BiocManager", quietly = TRUE))
#   install.packages("BiocManager")
BiocManager::install("GenomicFeatures", version = "3.8")

# install package: https://bioconductor.org/packages/release/data/annotation/html/TxDb.Hsapiens.UCSC.hg19.knownGene.html
# if (!requireNamespace("BiocManager", quietly = TRUE))
#   install.packages("BiocManager")
BiocManager::install("TxDb.Hsapiens.UCSC.hg19.knownGene", version = "3.8")

library(TxDb.Hsapiens.UCSC.hg19.knownGene)

## Loading required package: GenomicFeatures

## Loading required package: BiocGenerics

## Loading required package: parallel

##
## Attaching package: 'BiocGenerics'

## The following objects are masked from 'package:parallel':
## 
##   clusterApply, clusterApplyLB, clusterCall, clusterEvalQ,
##   clusterExport, clusterMap, parApply, parCapply, parLapply,
##   parLapplyLB, parRapply, parSapply, parSapplyLB

## The following objects are masked from 'package:stats':
## 
##   IQR, mad, sd, var, xtabs

## The following objects are masked from 'package:base':
## 
##   anyDuplicated, append, as.data.frame, basename, cbind,
##   colMeans, colnames, colSums, dirname, do.call, duplicated,
```