

TRGN 527: Applied Data Science and Bioinformatics

UNIT I. Introduction and Basic Data Science

Week 4 – Lecture 4 – Case Study Part 4

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*

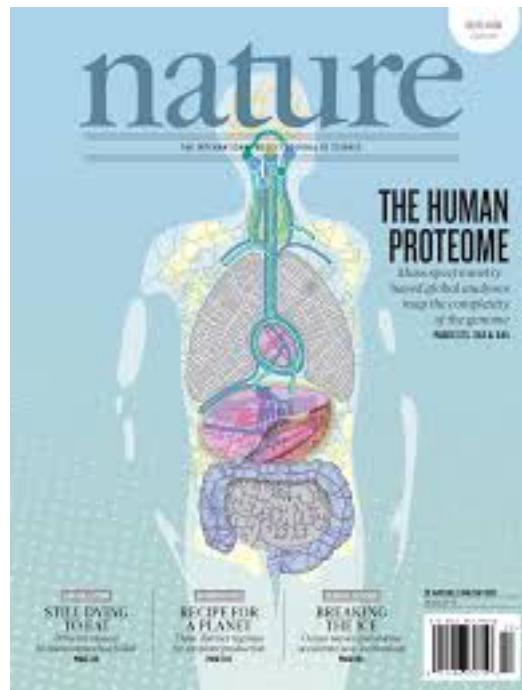
David W. Craig, Ph.D. | Professor and Vice Chair

Dept. of Translational Genomics

USC | Keck School of Medicine | Norris Comprehensive Cancer Center

Co-Director, Institute of Translational Genomics

Case Study – Proteomics Analysis



Case Study – Proteomics Analysis

- Getting protein-specific annotation

```
315  
316 # Getting protein-specific annotation  
317 ``{r}  
318  
319 # Getting protein-specific annotation  
320  
321 # Installing hpar  
322 #if (!requireNamespace("BiocManager", quietly = TRUE))  
323 #   install.packages("BiocManager")  
324 #BiocManager::install("hpar", version = "3.8")  
325  
326  
327 library(hpar)  
328 id <- "ENSG00000158813"  
329 eda.tissue <- getHpa(id, "hpaNormalTissue")  
330 eda.RnaGT <- getHpa(id, "rnaGeneTissue")  
331 eda.RnaGC <- getHpa(id, "rnaGeneCellLine")  
332 eda.locSB <- getHpa(id, "hpaSubcellularLoc")  
333 eda.locSBL <- getHpa(id, "hpaSubcellularLoc14")  
334 eda.locCA <- getHpa(id, "hpaCancer")  
335 ````  
336 ````
```

Case Study – Proteomics Analysis

- Getting protein-specific annotation

```
336  
337 # Visualizing first five rows of "rnaGeneTissue"  
338 ``{r}  
339  
340 head(eda.RnaGT)  
341  
342 ````
```

		Gene	Gene.name		Sample	Value	Unit
374478	ENSG00000158813		EDA	adipose tissue	3.0	TPM	
374479	ENSG00000158813		EDA	adrenal gland	8.9	TPM	
374480	ENSG00000158813		EDA	appendix	0.9	TPM	
374481	ENSG00000158813		EDA	bone marrow	0.1	TPM	
374482	ENSG00000158813		EDA	breast	2.4	TPM	
374483	ENSG00000158813		EDA	cerebral cortex	1.8	TPM	

Case Study – Proteomics Analysis

- Getting protein-specific annotation

```
343  
344 # Visualizing first five rows of "rnaGeneCellLine"  
345 ````{r}  
346  
347 head(eda.RnaGC)  
348  
349 ````
```

		Gene	Gene.name	Sample	Value	Unit
647745	ENSG00000158813		EDA	A-431	0.0	TPM
647746	ENSG00000158813		EDA	A549	0.5	TPM
647747	ENSG00000158813		EDA	AF22	0.3	TPM
647748	ENSG00000158813		EDA	AN3-CA	2.6	TPM
647749	ENSG00000158813		EDA	ASC diff	0.3	TPM
647750	ENSG00000158813		EDA	ASC TERT1	0.0	TPM

Case Study – Proteomics Analysis

- Getting protein-specific annotation

```
350
351 # Visualizing first five rows of "hpaSubcellularLoc"
352 ````{r}
353
354 head(eda.locSB)
355
356 ````
```

	Gene	Gene.name	Reliability	Enhanced	Supported	Approved	Uncertain
7074	ENSG00000158813	EDA	Supported		Vesicles	Lipid droplets	
		Single.cell.variation.intensity	Single.cell.variation.spatial		Cell.cycle.dependency		
7074				G0.id			
7074	Lipid droplets	(G0:0005811);Vesicles	(G0:0043231)				

Case Study – Proteomics Analysis

- Getting protein-specific annotation

```
357  
358 # Visualizing "hpaSubcellularLoc14"  
359 ````{r}  
360  
361 head(eda.locSBL)  
362  
363 ````
```

	Gene	Main.location	Other.location	Expression.type	Reliability
5588	ENSG00000158813		Vesicles		APE Supportive

Case Study – Proteomics Analysis

- Getting protein-specific annotation

```
364  
365 # Visualizing first five rows of "hpaCancer"  
366 ``{r}  
367  
368 head(eda.locCA)  
369  
370 ````
```

	Gene	Gene.name	Cancer	High	Medium	Low	Not.detected	prognostic.favourable	
202421	ENSG00000158813	EDA	breast cancer	NA	NA	NA	NA	NA	NA
202422	ENSG00000158813	EDA	carcinoid	NA	NA	NA	NA	NA	NA
202423	ENSG00000158813	EDA	cervical cancer	NA	NA	NA	NA	NA	NA
202424	ENSG00000158813	EDA	colorectal cancer	NA	NA	NA	NA	NA	NA
202425	ENSG00000158813	EDA	endometrial cancer	NA	NA	NA	NA	NA	NA
202426	ENSG00000158813	EDA	glioma	NA	NA	NA	NA	NA	NA
			unprognostic.favourable	prognostic.unfavourable	unprognostic.unfavourable				
202421		NA		NA			0.037600		
202422		NA		NA			NA		
202423		NA		NA			0.082860		
202424		0.1591		NA			NA		
202425		NA		NA			0.003589		
202426		NA		NA			0.102700		