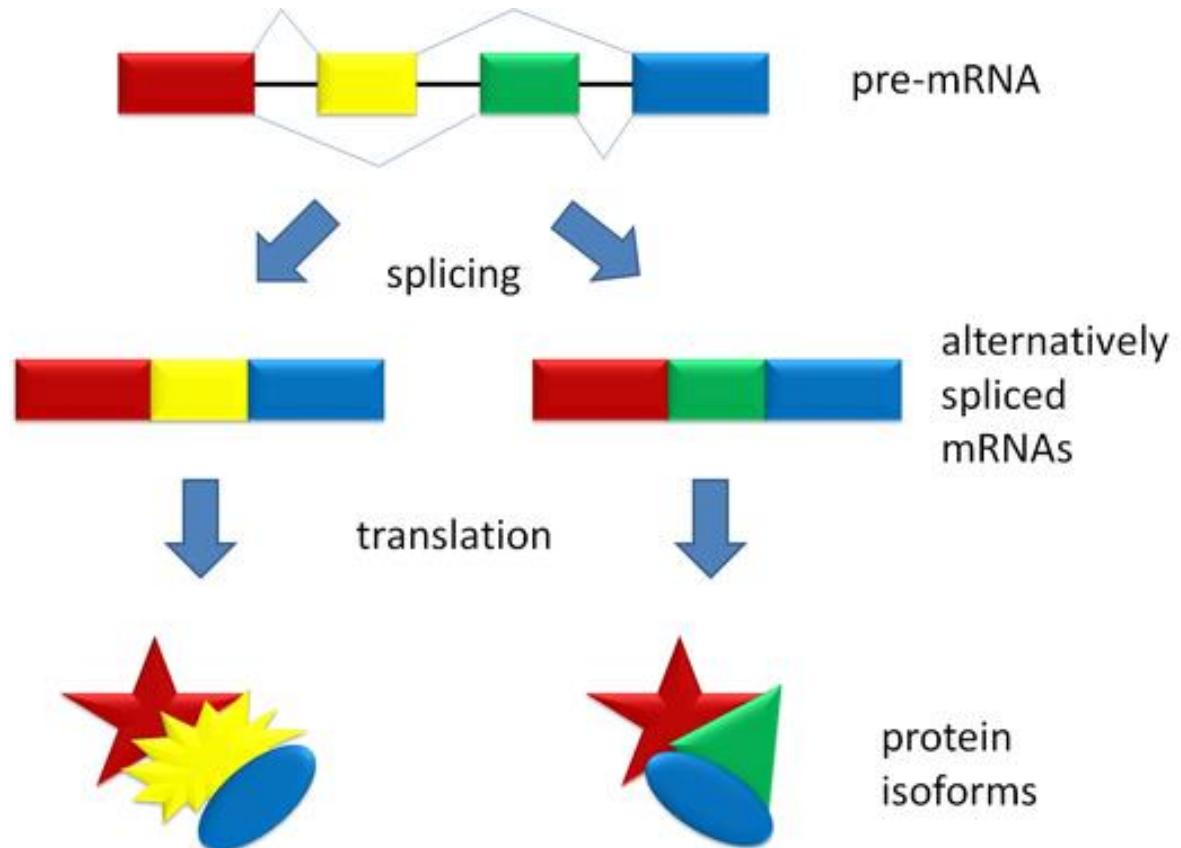


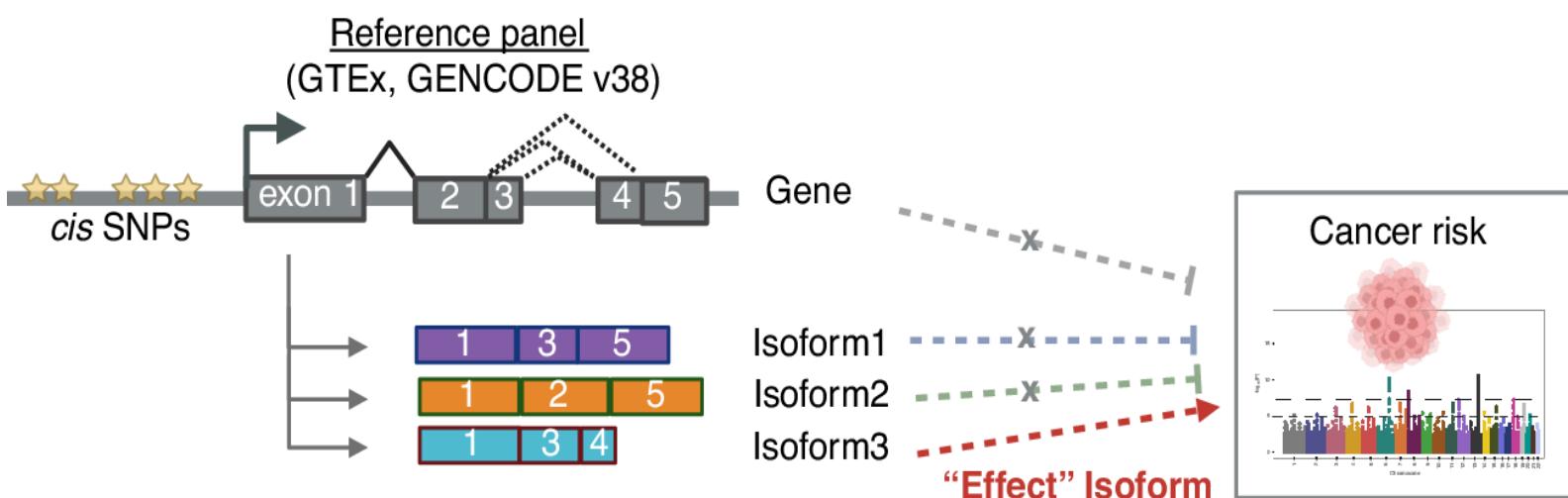
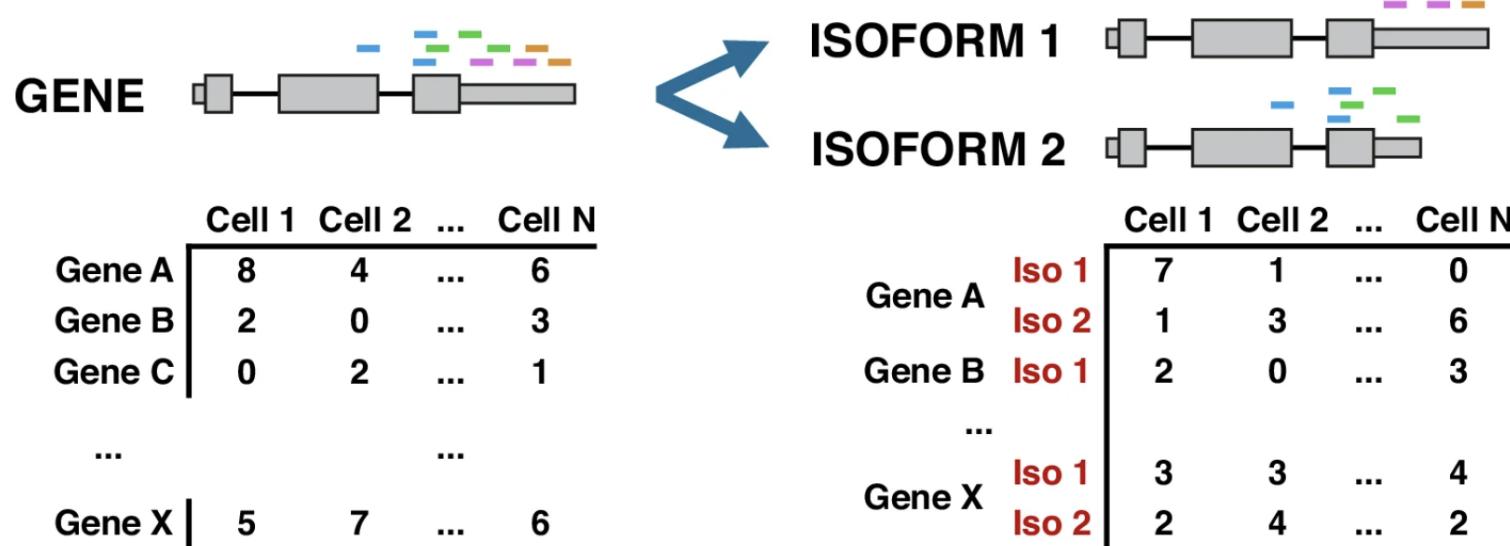
What is an isoform?



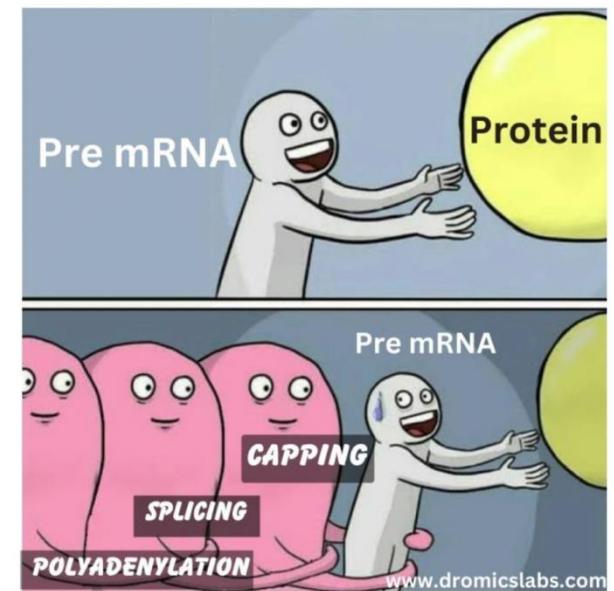
Each person has about **20,000** coding genes

X5

Proteins are essential for life, and the human body has over **100,000** different proteins performing many different functions.



Are isoforms good for us?



ALGORITHM ;)

assign random values to all $f(i)$

while not converged **do**

E-step:

initialize all $n(j)$ to 0

for each read r **do**

sum = $\sum_{j:w_{r,j}>0} w_{r,j} f(j)$

for each isoform j with $w_{r,j} > 0$ **do**

$n(j) += w_{r,j} f(j)/\text{sum}$

end for

end for

M-step:

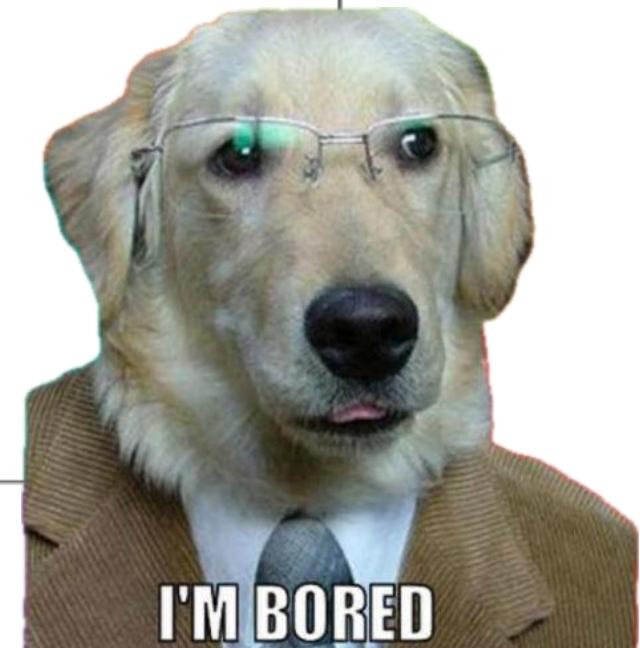
$s = \sum_j n(j)/(l(j) - \mu + 1)$

for each isoform j **do**

$f(j) = \frac{n(j)/(l(j)-\mu+1)}{s}$

end for

end while



I'M BORED

DPI calculation (Delta Percent Isoform)

$$DPI = \frac{\sum |\Delta PI_i|}{2}$$

2 DPI= all of the isoforms pi
 Delts PI = absolute value of pi's

pos → acceptor (3' splice site)

neg → donor (5' splice site)

2 most found isoforms are shown, where are the others??

gid	p_val	dpi	pos_iso_1	pos_iso_2	pos_iso_1_dpi	pos_iso_2_dpi	neg_iso_1	neg_is_0_2	neg_iso_1_dpi	neg_iso_2_dpi	adj_p_val	gname
ENSG0000079150	0.0014534 99946493 2115	28.125	ENST000004191 ENST0004709 84.1	ENST000004247 7011108 45.1	14.5432 7949066 3984	6.91105 162 162	ENST000004247 00004247 85.6	ENST000004247 00004 64248. 1	-23.4375 -4.6875	-4.6875	0.0116 43364 38036 7663	FKBP7

$$\begin{aligned}\sum |\Delta PI| &= 14.54327 + 6.91106 + 23.4375 + 4.6875 \\ &= 49.57933\end{aligned}$$

Now divide by 2:

$$DPI = \frac{49.57933}{2} = 24.79$$

$$\sum |\Delta PI_{all}| = 2 \times DPI = 2 \times 28.125 = 56.25$$

$$\sum |\Delta PI| = 49.57933 + 6.67067 = 56.25$$

$$DPI = \frac{56.25}{2} = 28.125$$

Motivation

Standard DPI / isoform analyses:

- Typically report only the top 2 positive and top 2 negative isoforms

What if I am interested in other isoforms of genes??

- Remaining isoforms are often treated as noise or biologically negligible

This study introduces **Dpi₃**, a quantitative metric defined as:

$$\text{Dpi}_3 = (2 \times \text{dpi}) - \sum |\text{dominant isoform DPI values}|$$

<u>Gen</u>	<u>Dpi_3</u>	<u>Significance</u>
CFLAR	~30.9	Very powerful alternative splicing/apoptotic switch
CALD1	~16.5	Cytoskeletal reorganization
SEC31A	~16.0	ER–Golgi transport regulation
TPM1	~10.2	Actin filament dynamics
RTKN	~11.2	Rho-GTPase signal

Output file called “third_iso.csv”

gid	p_val	Dpi_3	gname
ENSG00000079 150	0.001453499946493211 5	21.35 ≥ 10	FKBP7



- Write this gene to the visualization code



corb
@awshuqs

is ChatGPT down for anyone else? i'm a
neurosurgeon in the middle of brain
surgery 

mRNA and Protein(s)

1. [NM_001135212.2](#) → [NP_001128684.1](#) peptidyl-prolyl cis-trans isomerase FKB P7 isoform b precursor

[See identical proteins and their annotated locations for NP_001128684.1](#)

Status: REVIEWED

Description	Transcript Variant: This variant (2) uses an alternate in-frame splice site in the coding region, compared to variant 1. The resulting isoform (b) is 1 amino acid shorter than isoform 1.
Source sequence(s)	AA447628 , AC009948 , AY353086 , BC009711 , DA020387
Consensus CDS	CCDS46462.1
UniProtKB/Swiss-Prot	Q9Y680
Related	ENSP00000415486.2 , ENST00000434643.6

Conserved Domains (3) [summary](#)

	pfam00254 Location: 46 → 140	FKBP_C; FKB-type peptidyl-prolyl cis-trans isomerase
	pfam13499 Location: 151 → 213	EF-hand_7; EF-hand domain pair
	c108302 Location: 151 → 212	EFh; EF-hand, calcium binding motif; A diverse superfamily of calcium sensors and calcium signal modulators; most examples in this alignment model have 2 active canonical EF hands. Ca ²⁺ binding induces a conformational change in the EF-hand motif, leading to ...

2. [NM_001410972.1](#) → [NP_001397901.1](#) peptidyl-prolyl cis-trans isomerase FKB P7 isoform c precursor

Status: REVIEWED

Source sequence(s)	AC009948 , AK299215
Consensus CDS	CCDS92911.1
UniProtKB/TrEMBL	B4DRE2
Related	ENSP00000509396.1 , ENST00000464248.1

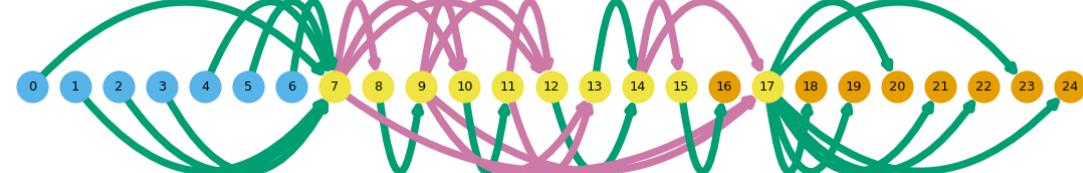
3. [NM_181342.3](#) → [NP_851939.1](#) peptidyl-prolyl cis-trans isomerase FKB P7 isoform a precursor

[See identical proteins and their annotated locations for NP_851939.1](#)

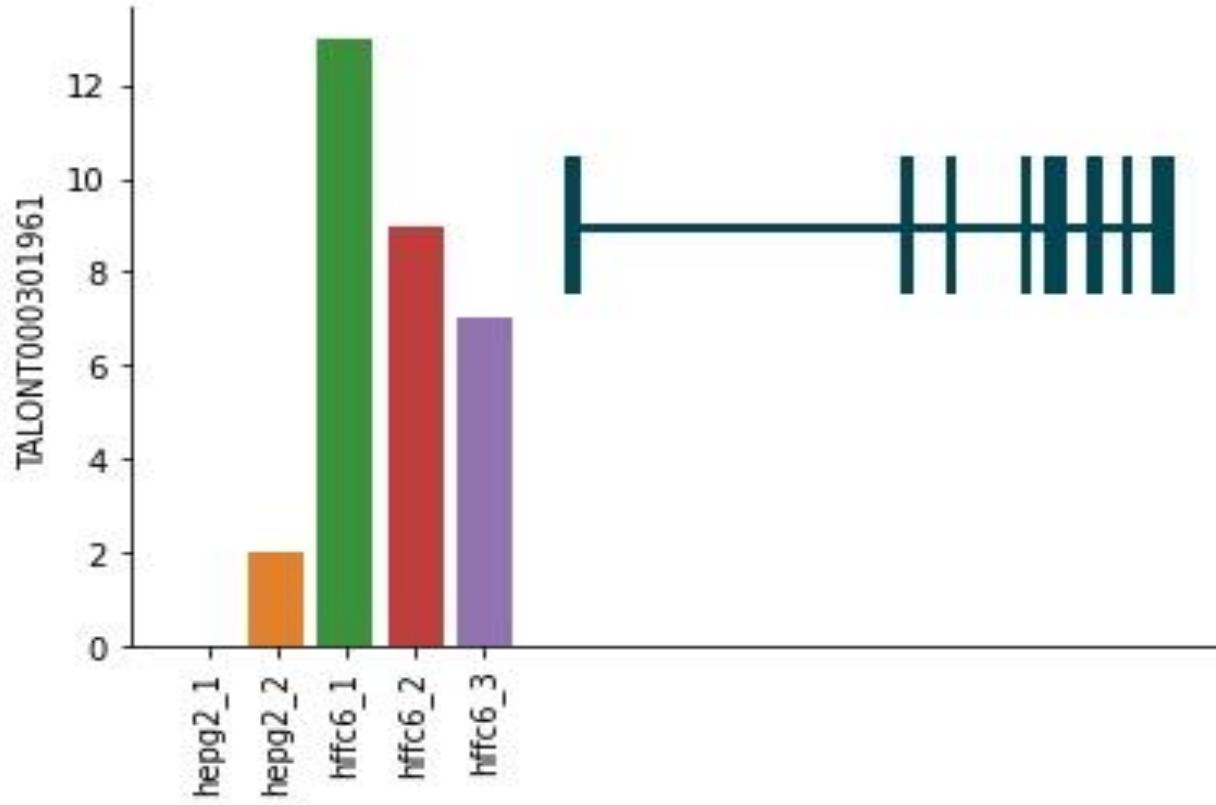
Status: REVIEWED

Description	Transcript Variant: This variant (1) represents the longer transcript and encodes the longer isoform (a).
Source sequence(s)	AA447628 , AC009948 , BC009711 , DA020387
Consensus CDS	CCDS2280.1
UniProtKB/Swiss-Prot	Q4ZG70 , Q6V3B2 , Q86U65 , Q96DA4 , Q9Y680 , Q9Y6B0
Related	ENSP00000413152.2 , ENST00000424785.7

Conserved Domains (2) [summary](#)



Swan assigns a node to each unique splice site.



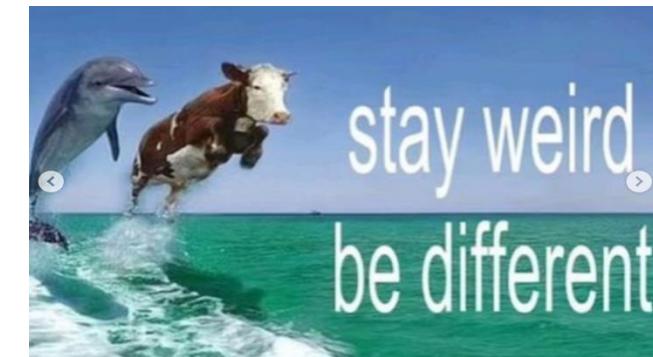
TALONT00311961L isoform shows:

Very low expression in hepg2_1 and hepg2_2 cell lines

High expression in hffc6_2 and hffc6_3 cell lines

This shows the **change in usage** across different cell lines → **isoform switch potential**

Gene is the same
Isoform is the same
Isoform function is different! Why?

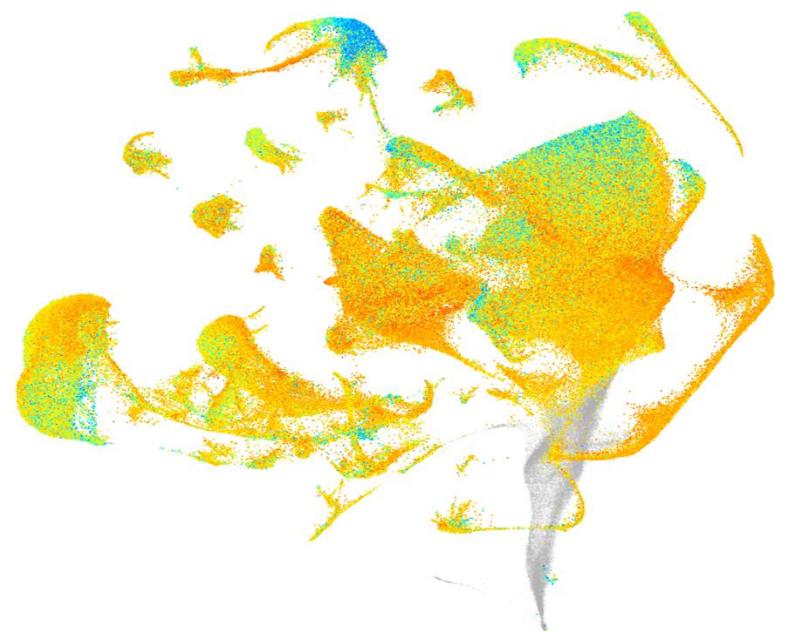
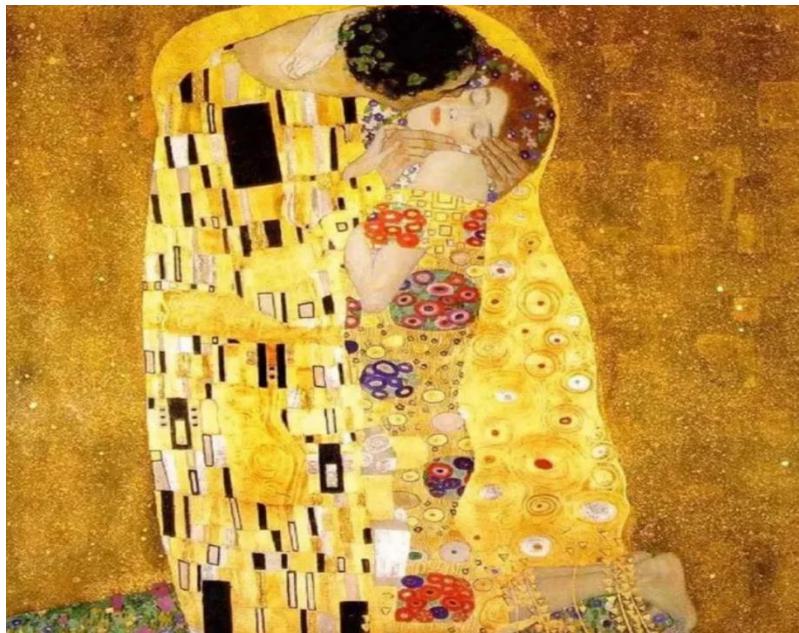


- Bioinformatics Tutorials -> <https://ngs101.com/tutorials/>

For advanced isoform analysis ->

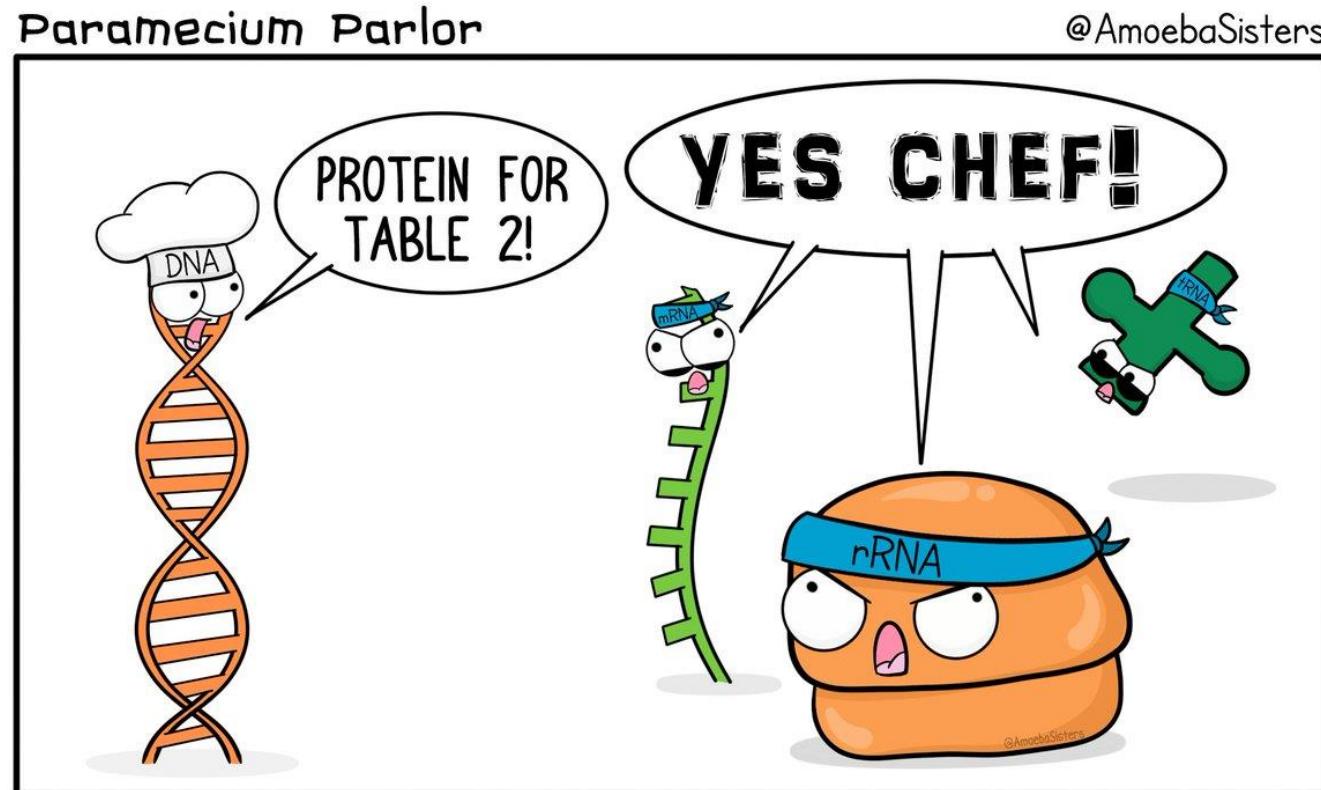
<https://bioconductor.org/packages/release/bioc/vignettes/IsoformSwitchAnalyzeR/inst/doc/IsoformSwitchAnalyzeR.html#importing-data-from-salmon-via-tximeta>

Github101 -> <https://github.com/nuryagli/git-github-101-networking>



Thanks for attending!!

Any questions???



RNA: If you can't help make proteins,
get out of the kitchen.