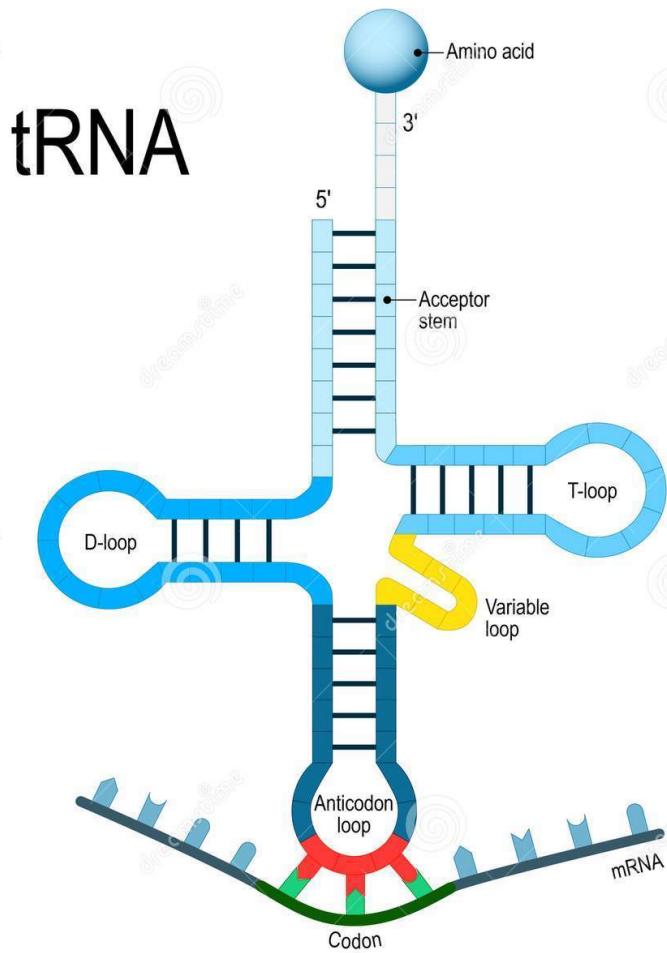


Investigating the Role of Relative Usage of Synonymous Codons in Breast Cancer Metastasis: tRNA and miRNA perspective



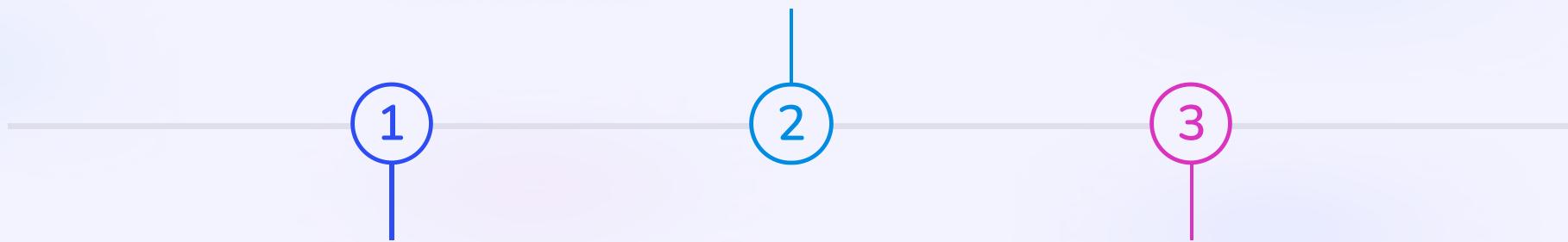
by Nischal Bhandari

Understanding Codon Synonymy

What is Codon Synonymy?

Codon synonymy refers to the **existence of multiple codons** that specify **the same amino acid**.

This redundancy impacts how genes are expressed.

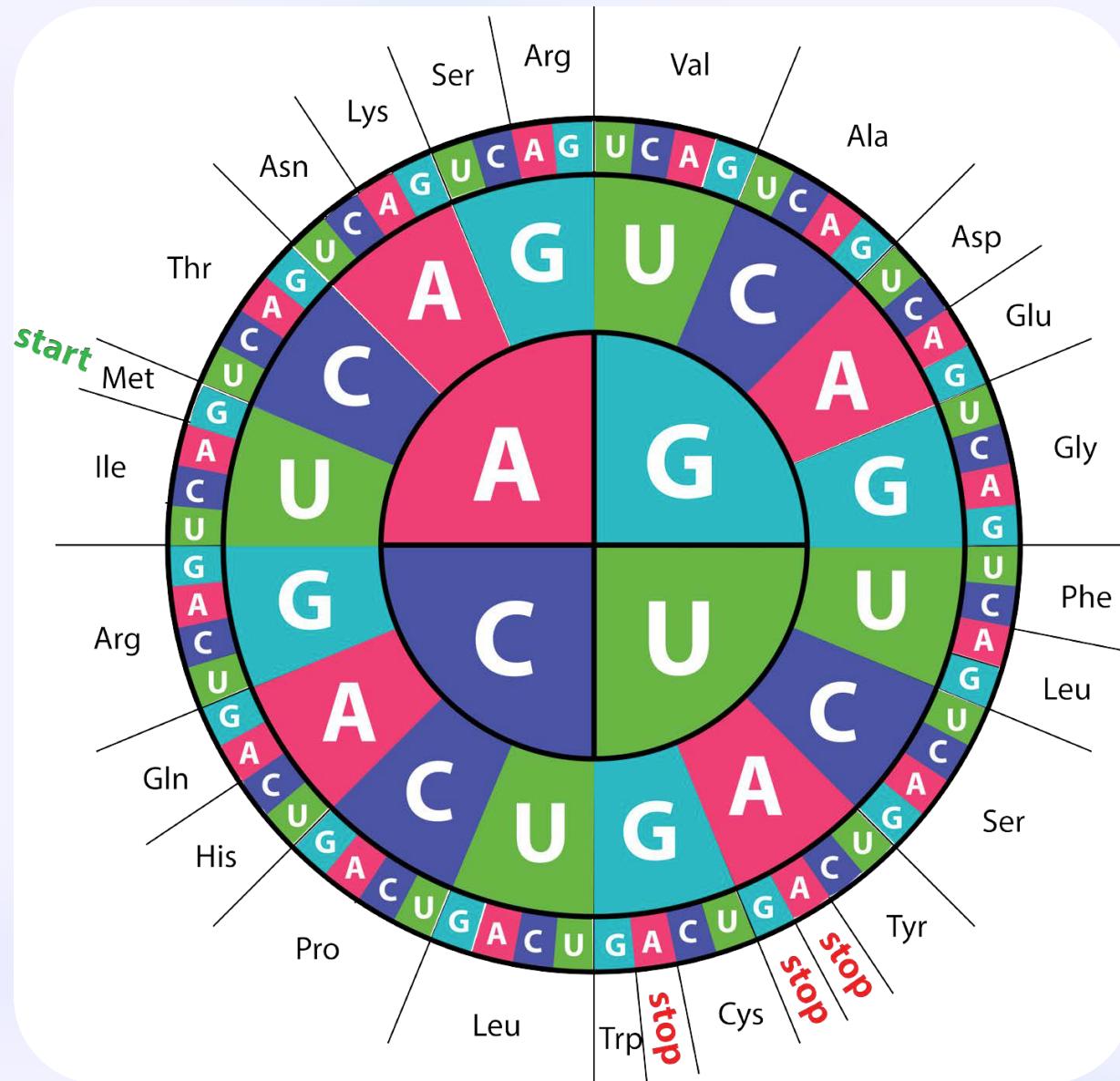


What are Codons?

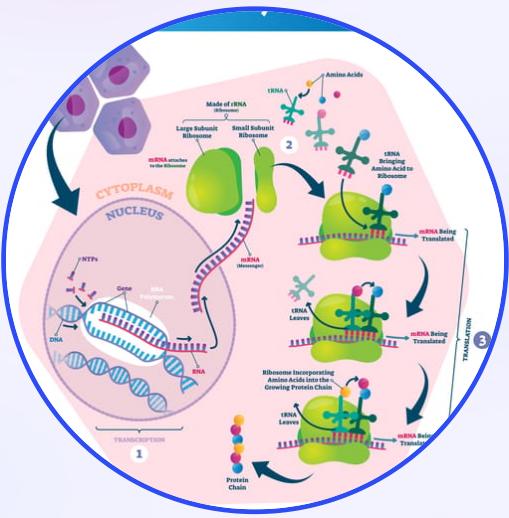
three nucleotides in mRNA that encode for a specific amino acid during protein synthesis.

Codon Usage Bias

Codon bias is a **phenomenon where certain codons are preferred over others in coding for the same amino acid in cells**. This preference can vary between species and can impact gene expression.

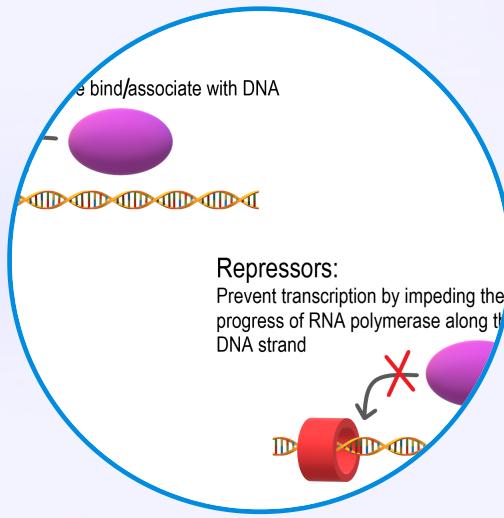


The Role of Codon Synonymy



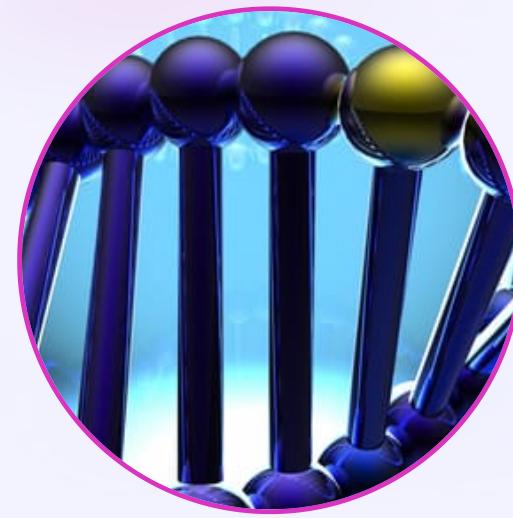
Translation

The use of synonymous codons **has been shown to affect the efficiency and accuracy of protein translation by the ribosome.**



Gene Expression

Codon usage can also influence gene expression levels, by affecting **how regulatory proteins interact with mRNA transcripts during transcription and translation.**



Molecular Biology

Studying codon usage patterns can provide insights into molecular mechanisms that underlie the regulation of gene expression and protein synthesis.

Impact on Translation Speed

Factors Affecting Translation Speed

The speed of translation can be influenced by **the codon usage in the mRNA transcript**, as well as **the availability of charged tRNAs**, **the ribosome binding site**, and **the energy status of the cell**.

The Codon Adaptation Index

The codon adaptation index (CAI) is a measure of how well a gene's codon usage matches the codon usage of the host organism. This can be used to predict the speed of translation and the level of gene expression for a specific gene.

Optimizing Codon Usage

Manipulating the codon usage in a gene

- whole genome codon optimization vs
- first 8 codon window

Regulation of Gene Expression

1

Post-Transcriptional Regulation

Codon usage can influence the binding of mRNA binding proteins, microRNAs, and other regulatory factors, affecting the stability of the mRNA and the efficiency of translation.

2

Translational Pausing

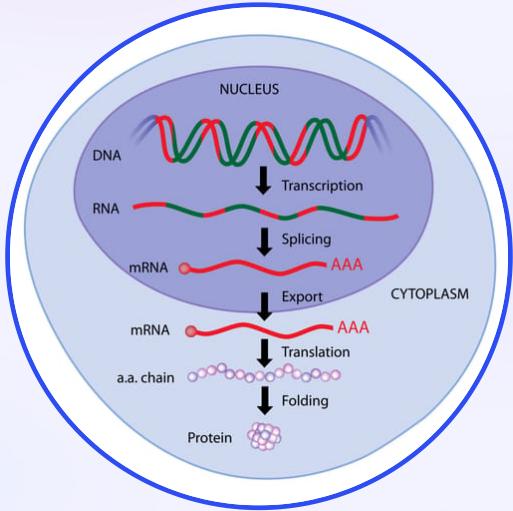
The use of certain **rare codons** **can induce the ribosome to pause during translation,** leading to the creation of truncated proteins and reduced protein expression levels.

3

Structural Features of mRNA

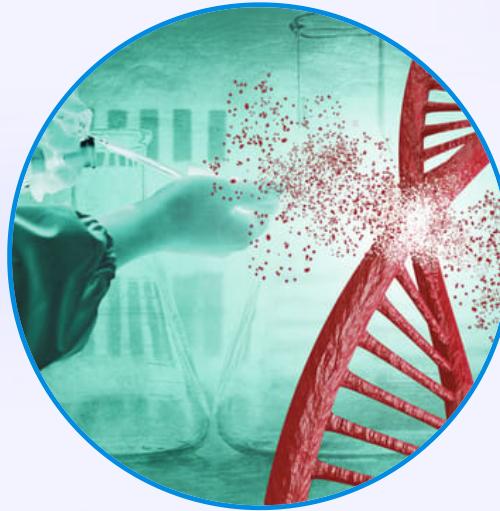
The **secondary structure** of mRNA can affect codon accessibility and ribosome movement, influencing the speed and efficiency of translation.

Applications of Codon Synonymy



Protein Production

Optimizing the codon usage in recombinant genes can improve protein production in bacterial and mammalian expression systems.



Gene Therapy

The manipulation of codon usage can improve the expression of therapeutic proteins in gene therapy applications, particularly for genetic disorders.

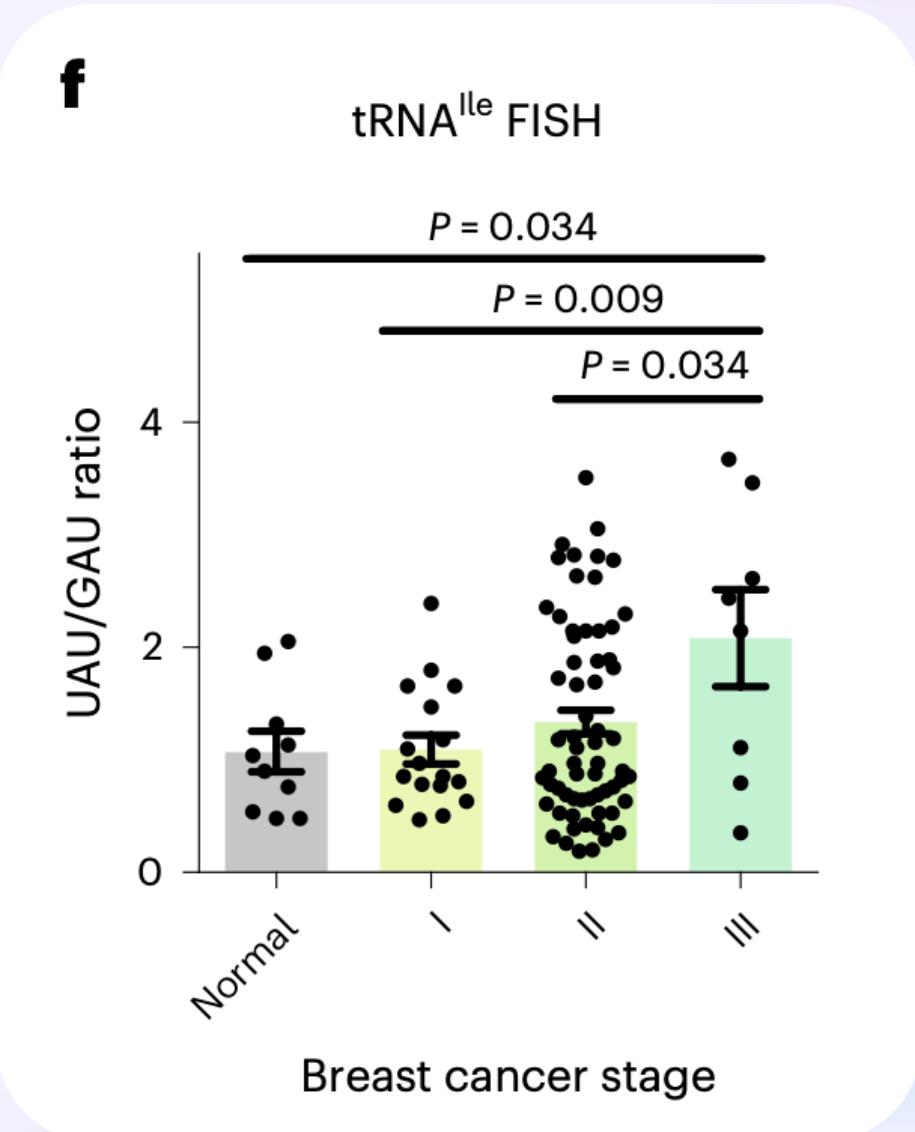


Synthetic Biology

Codon optimization can be used to design and engineer synthetic genes and pathways for novel biological functions and biotechnological applications.

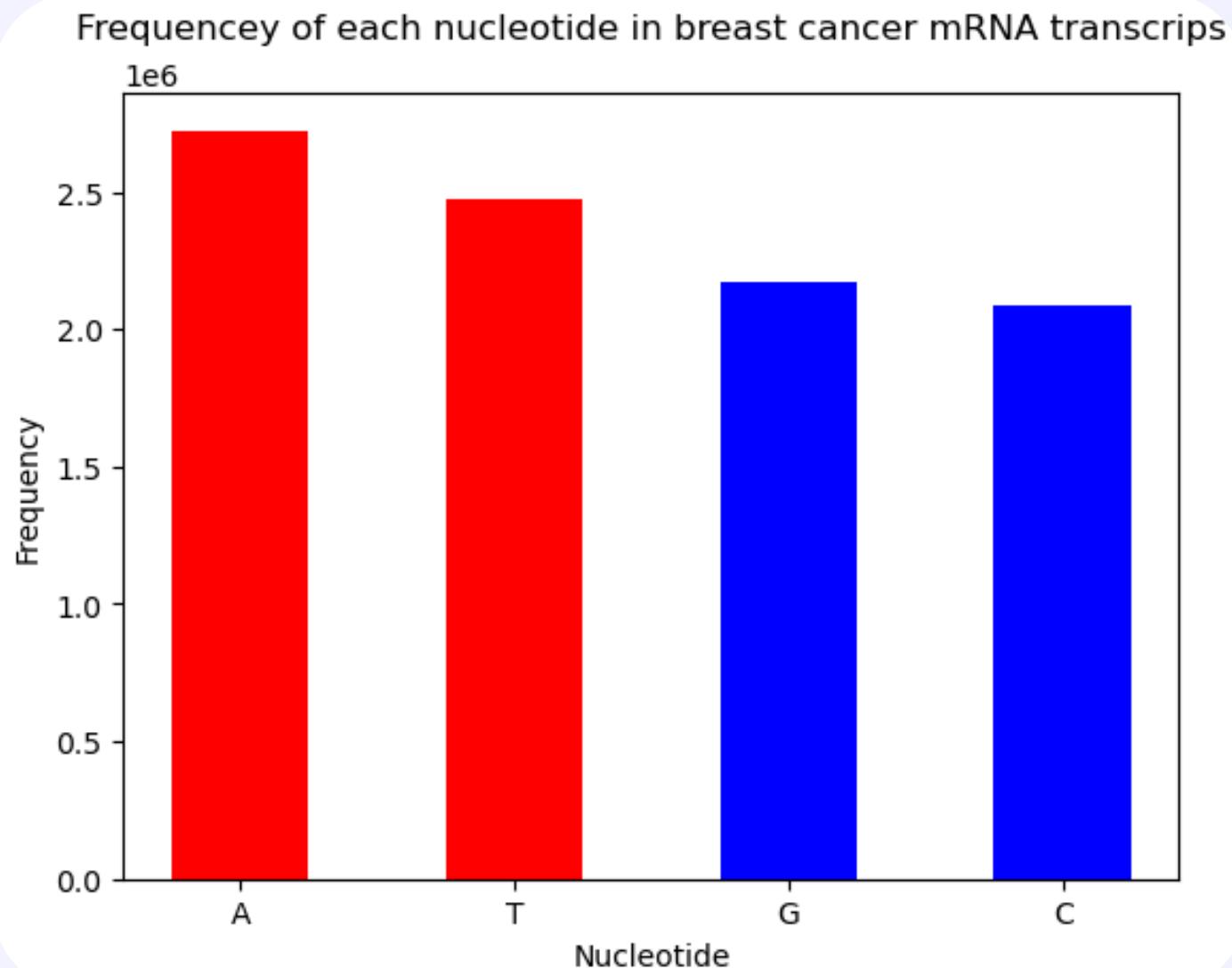
Codon Bias in Cancer Transcript

Abundance of tRNA for 5'AUA3' compared to 5'AUC3' isoleucine codons.

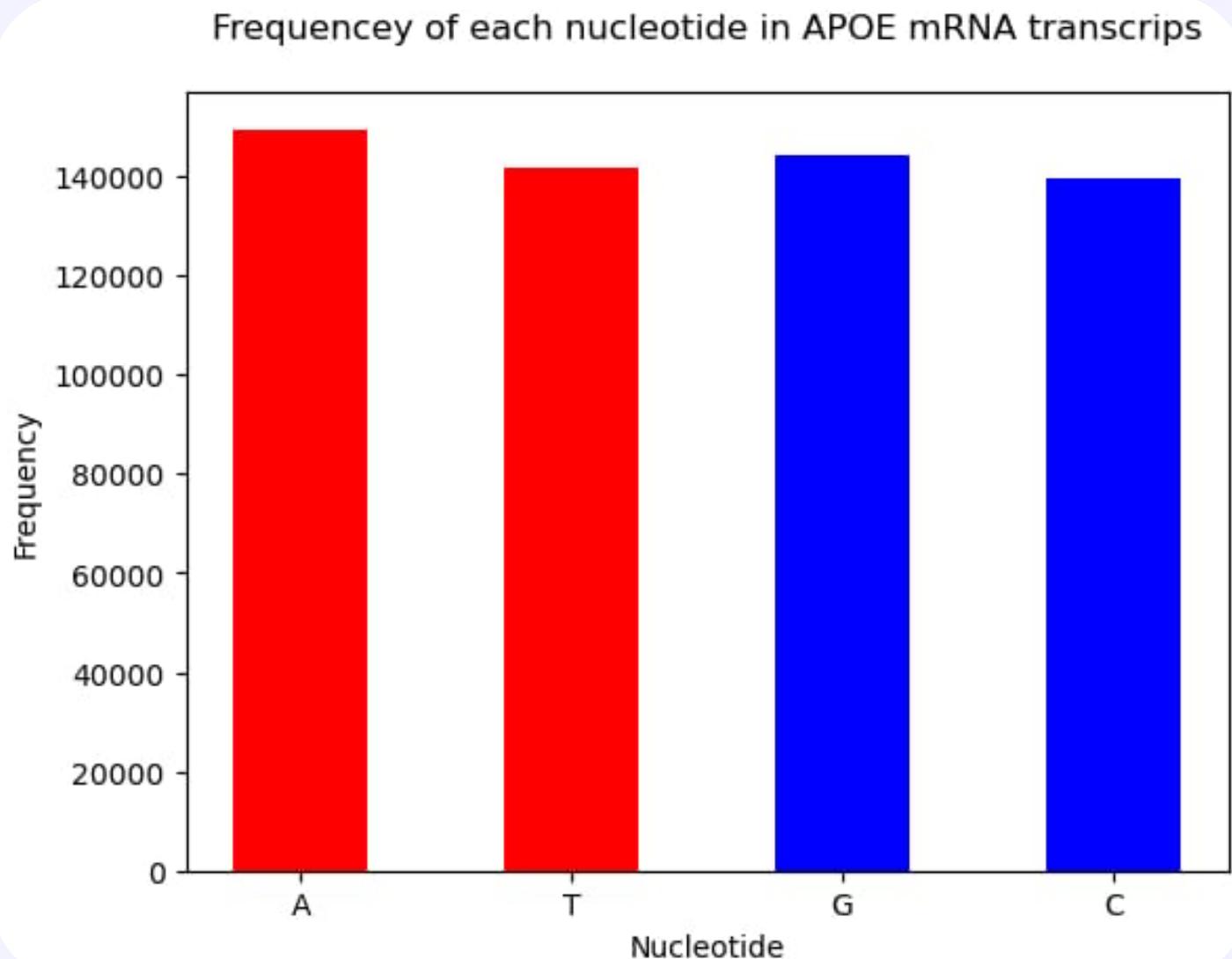


How are nucleotide sequences being used in Breast cancer transcripts?

- the presence of G or C bases tends to be negatively correlated with protein production levels *in this region*
- downloaded from NCBI using e-utils
- filtered for "NM" only
- total 1918 mRNA RefSeq sequences

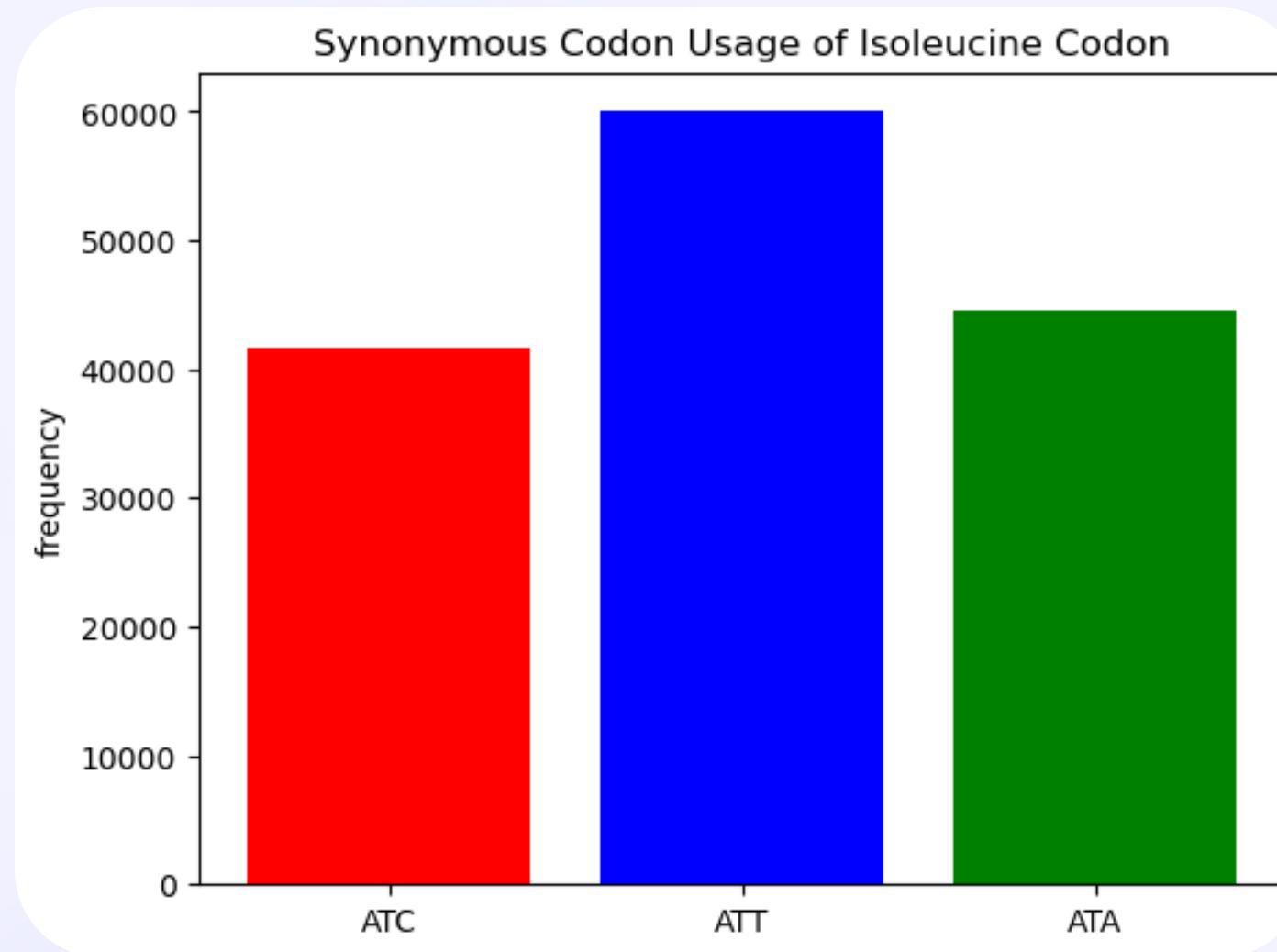


What about in another transcripts with relatively low mutations? #1.2



Do we have any indiscriminate usage of synonymous isoleucine codons in Breast cancer transcripts? #2

Isoleucine : ATT, ATA, ATC



expectation was: ATA » ATC.

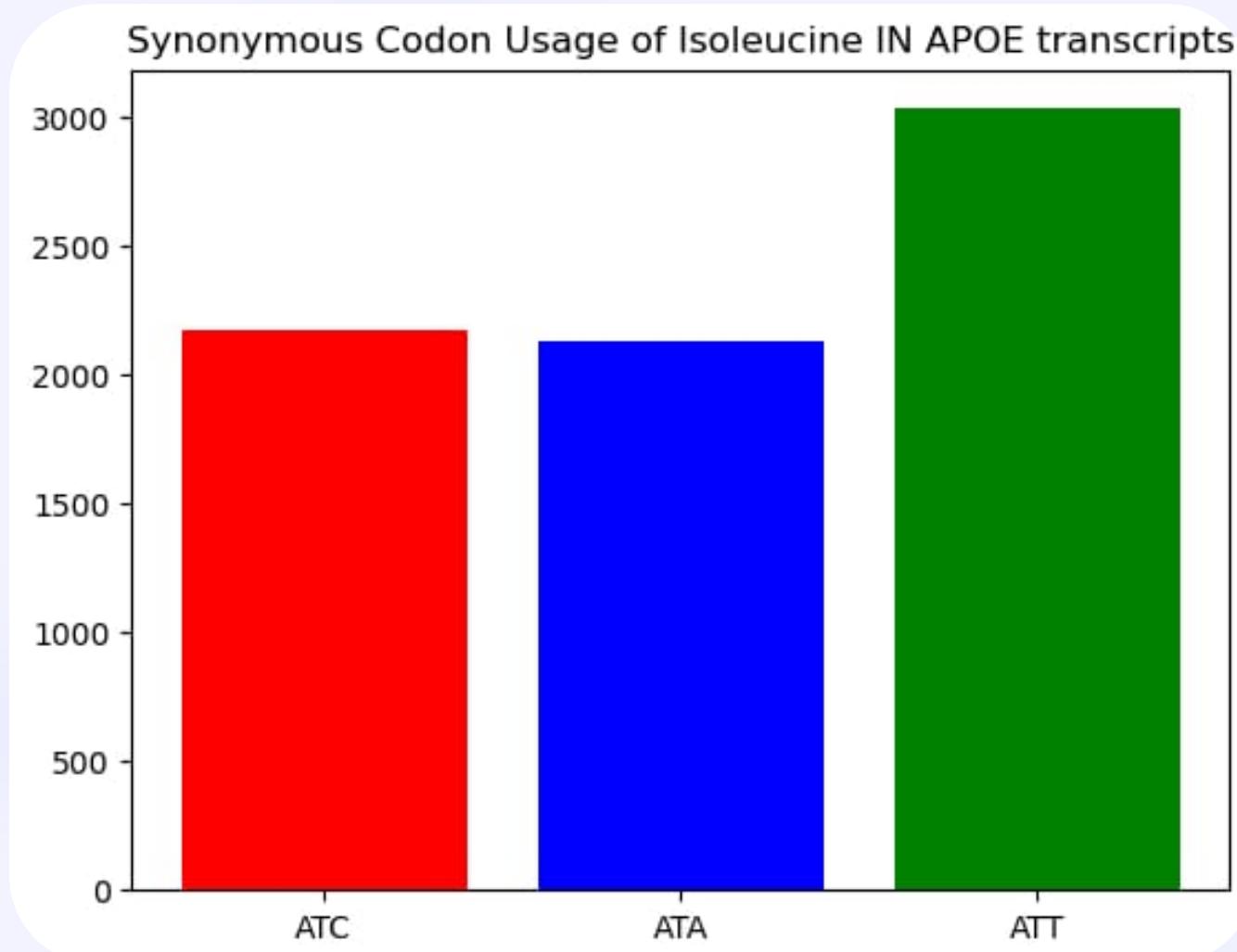
Isoleucine codon frequency:

ATA: 44477

ATC: 41576

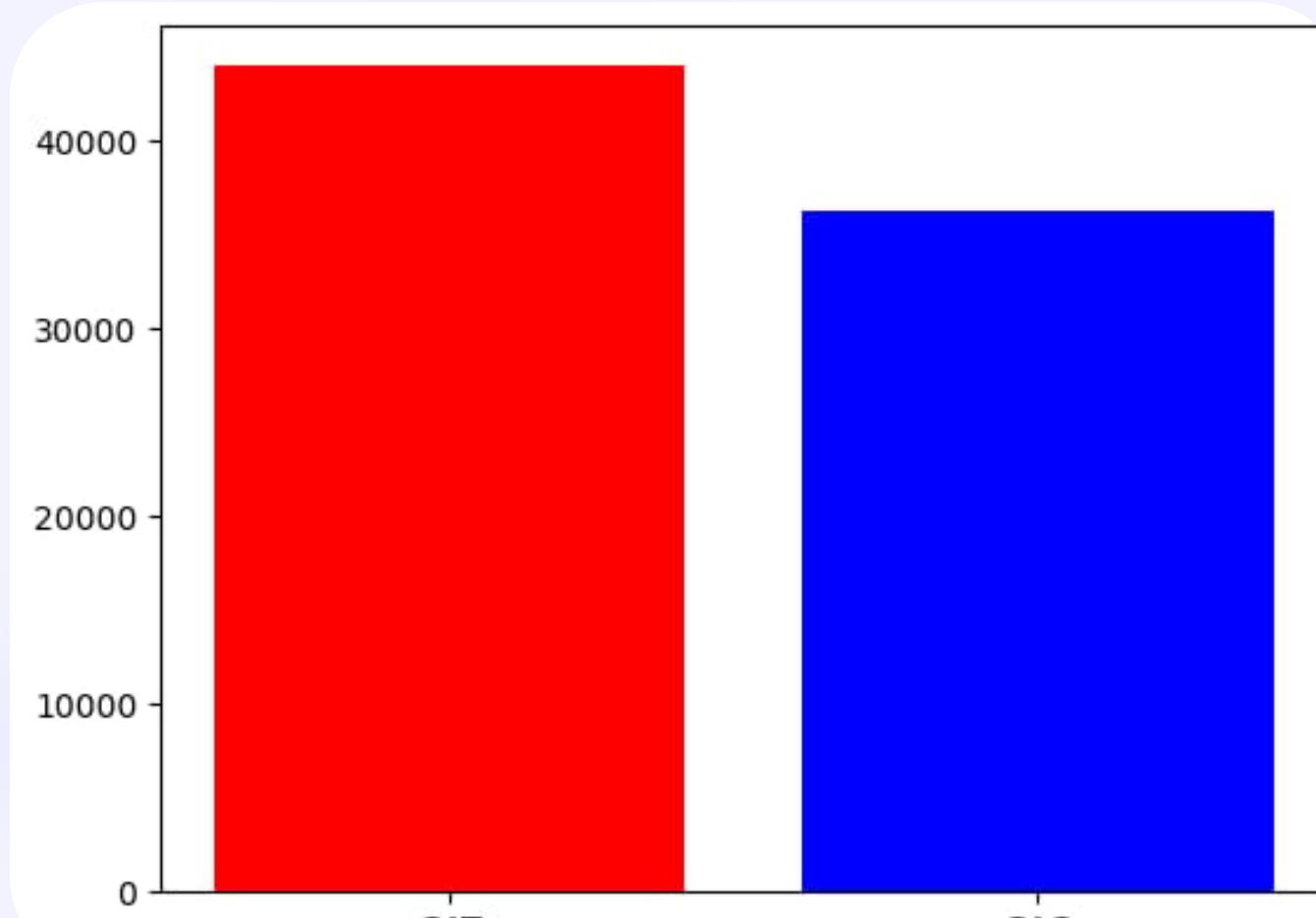
ATT: 60022

#2.1 what about synonymous codon usages among amino acids of APOE gene?

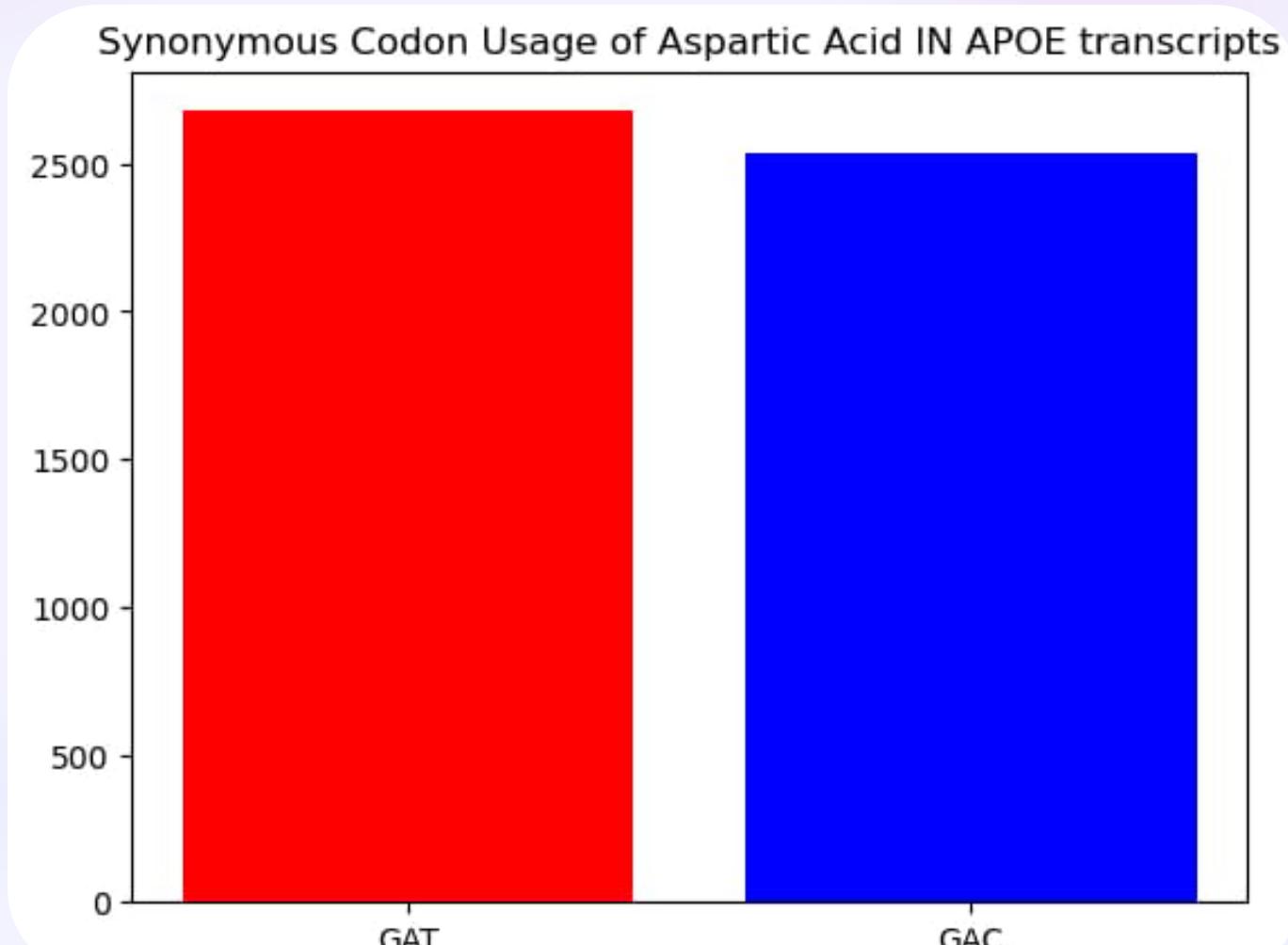


#2.2 what about in another amino acid in Breast Cancer transcripts?

Aspartic Acid Codon Usage: GAT: 44008 GAC: 36321



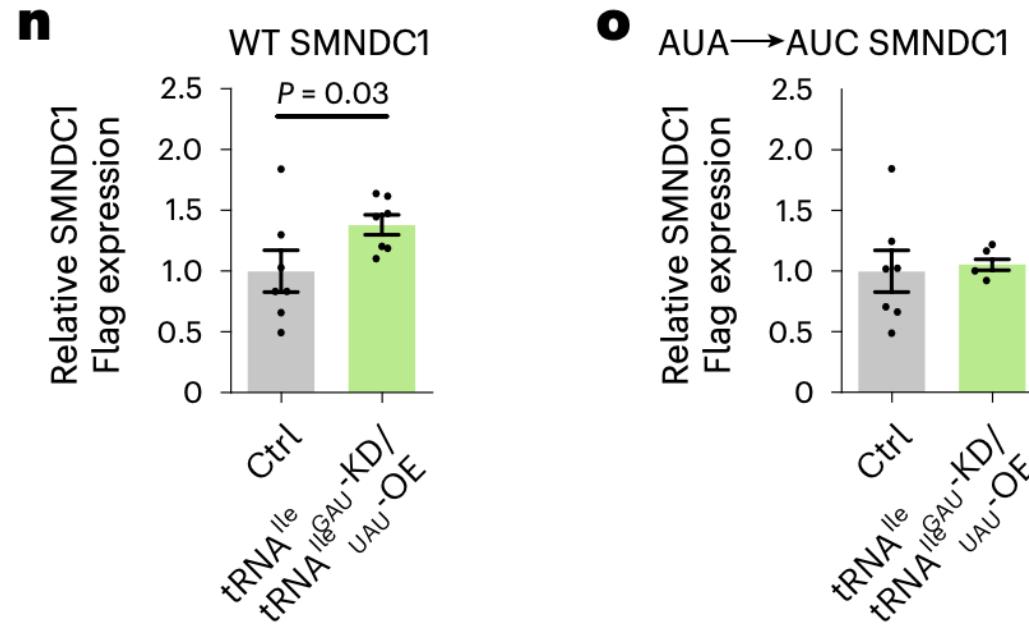
#2.3 what about in APOE?



What about codon usage among growth promoting genes? # 3

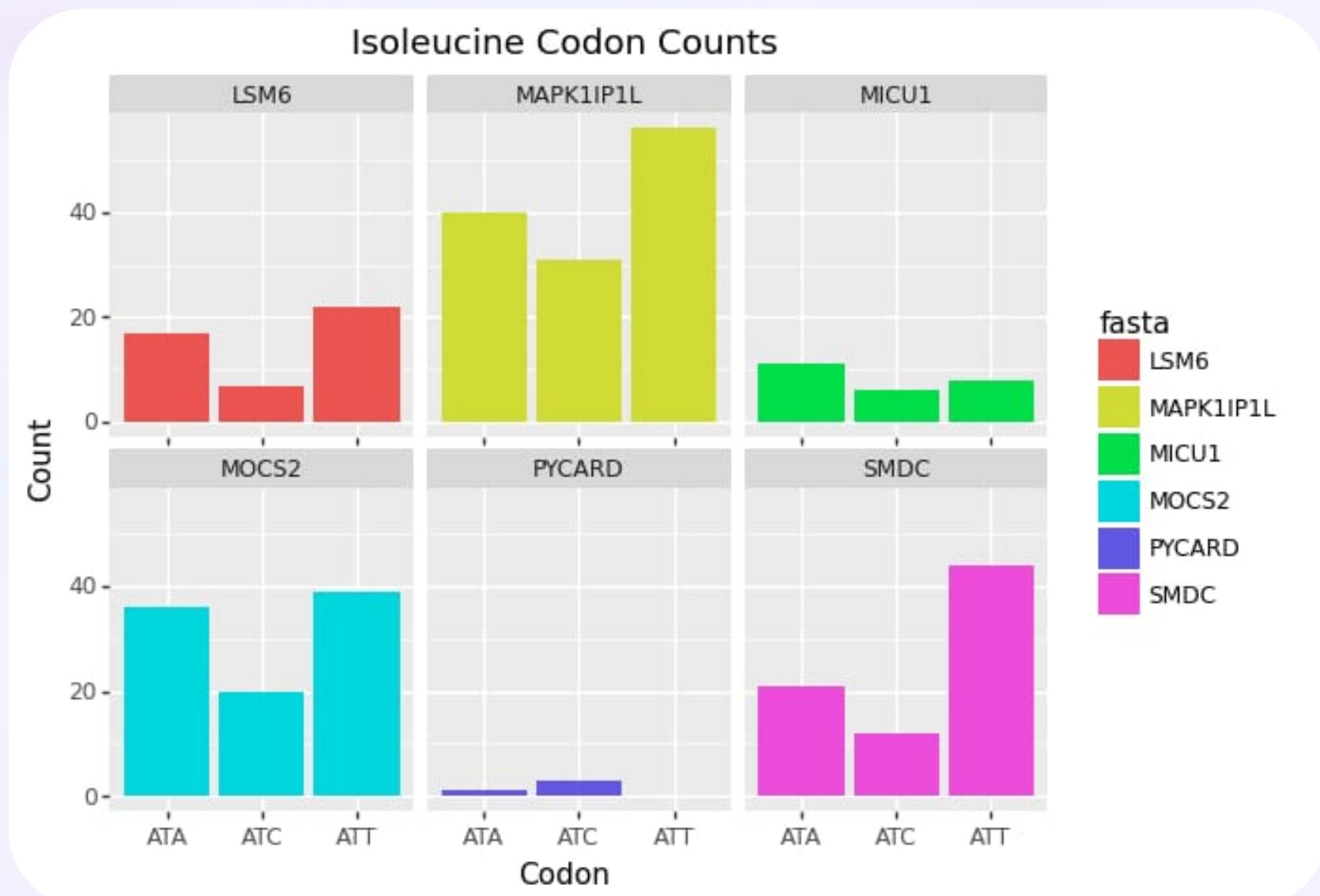
Genes:

“**SMNDC1** (Survival Motor Neuron Domain Containing 1), **LSM6** (Like-Sm Protein 6), **MOCs2** (Molybdenum Cofactor Synthesis 2), **MAPK1Ip1L** (Mitogen-Activated Protein Kinase 1 Interacting Protein 1-Like), **MICU1** (Mitochondrial Calcium Uptake 1), and **PYCARD** (PYD and CARD domain-containing protein)



for codons: ATA/ATC higher

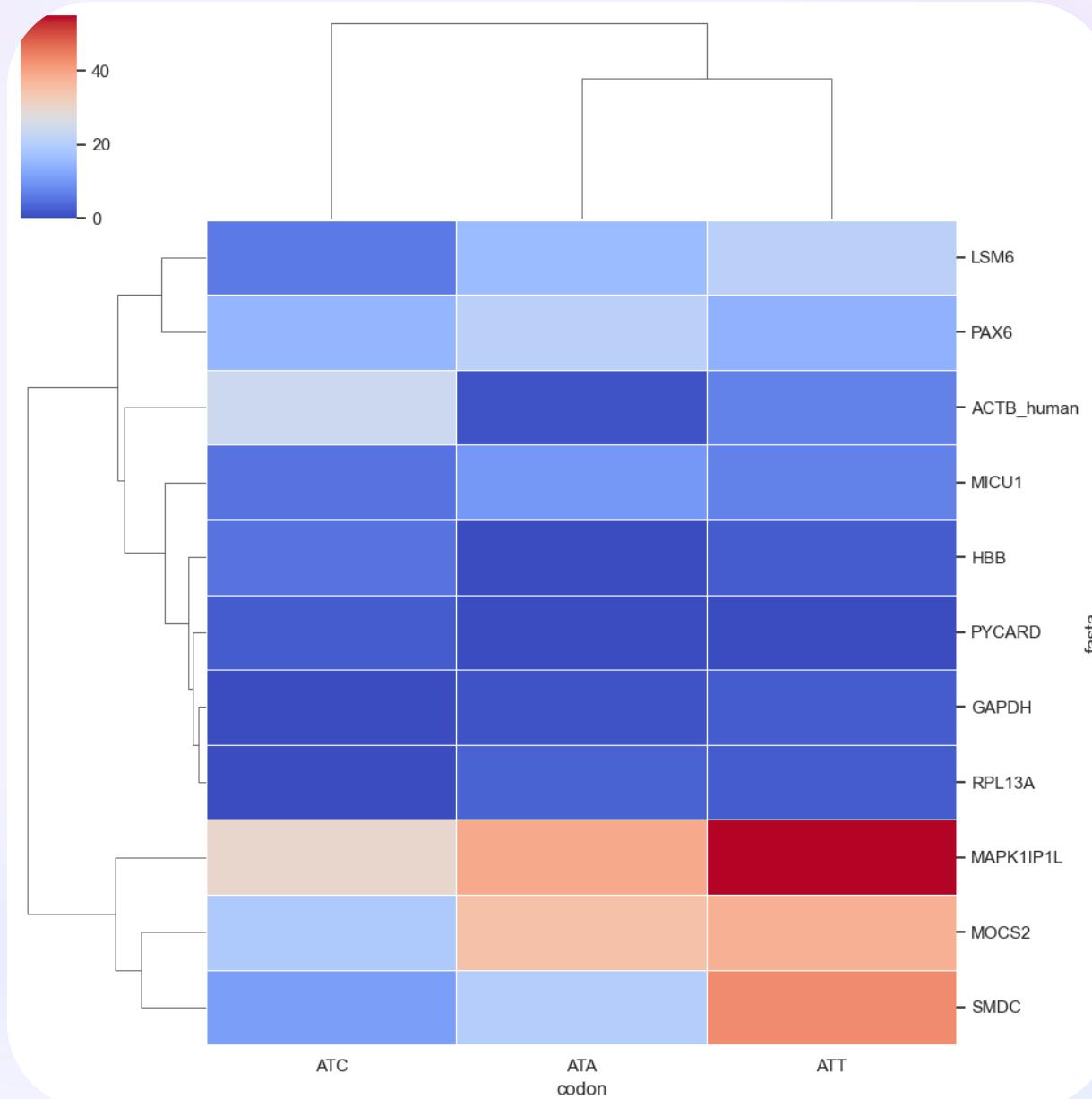
#3



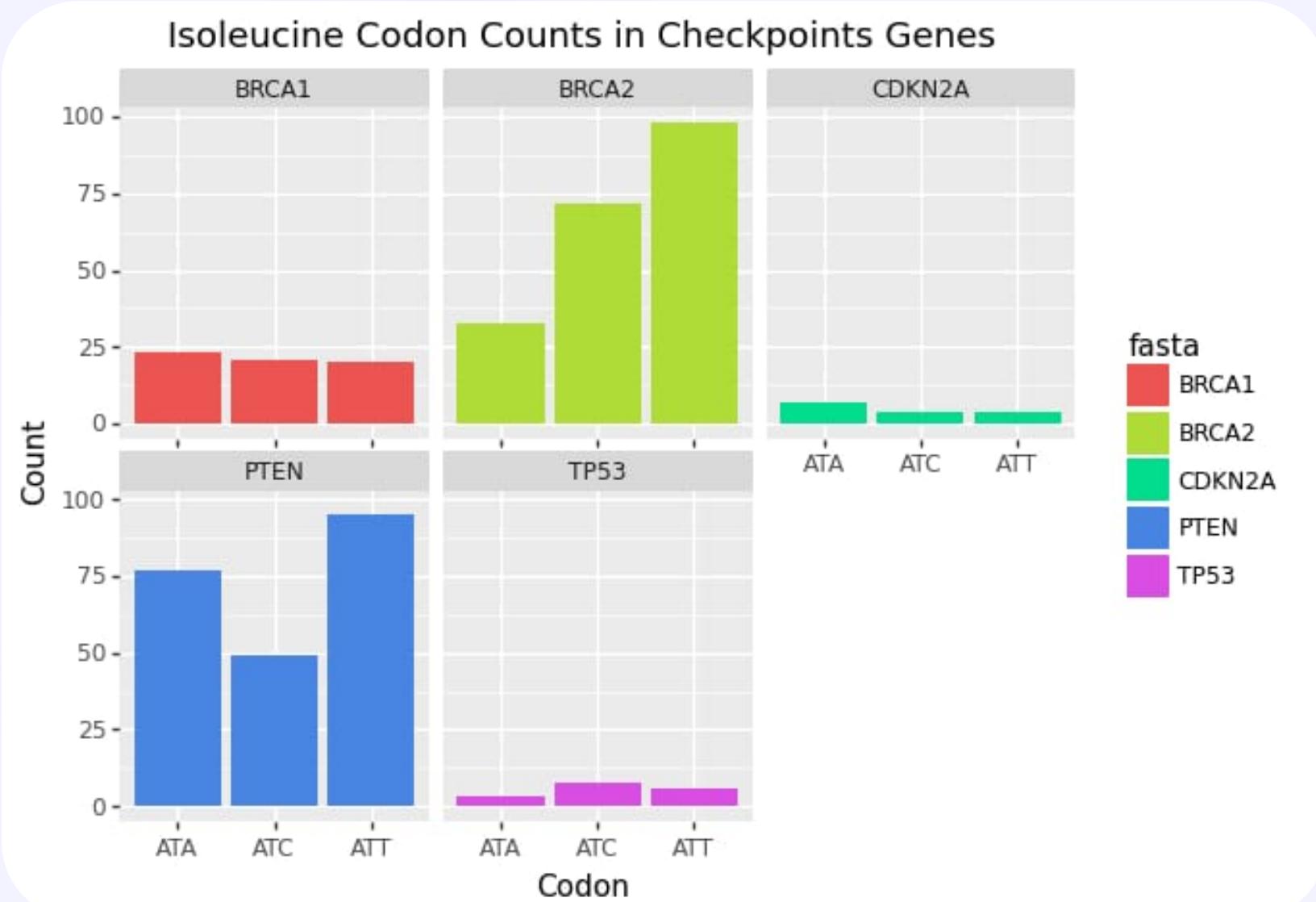
#3.1 what about isoleucine codon usage in cancer promoter genes compared to control genes?

Beta-actin (ACTB) gene, Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) gene , Hemoglobin subunit beta (HBB), Paired box 6 (PAX6) gene ,Ribosomal protein L13a (RPL13A)





#3.2 what about synonymous codon usage in growth suppressors/ cellular checkpoints genes?



4 What are some of the mutations patterns in TCGA database?

- for example, is ATC codon changed into ATA from C → A nucleotide conversion.

```
sub_count['DNA Change'].value_counts()  
| ✓ 0.0s  
G>A    27  
C>T    15  
T>C     8  
A>G     4  
C>A     4  
A>C     3  
G>C     3  
T>A     2  
G>T     2  
A>T     1  
C>G     1  
Name: DNA Change, dtype: int64
```

for isoleucine codons,

ATA (C → A, T → A)
ATC (A → C, T → C)
ATT (A → T, C → T)

```
# substitutions in CDKN2A gene
cdk_count['DNA Change'].value_counts()

✓ 0.0s

C>T      18
G>A      17
C>A      15
G>T       5
T>C       4
G>C       3
C>G       3
T>G       3
T>A       1
A>C       1
A>T       1

Name: DNA Change, dtype: int64
```

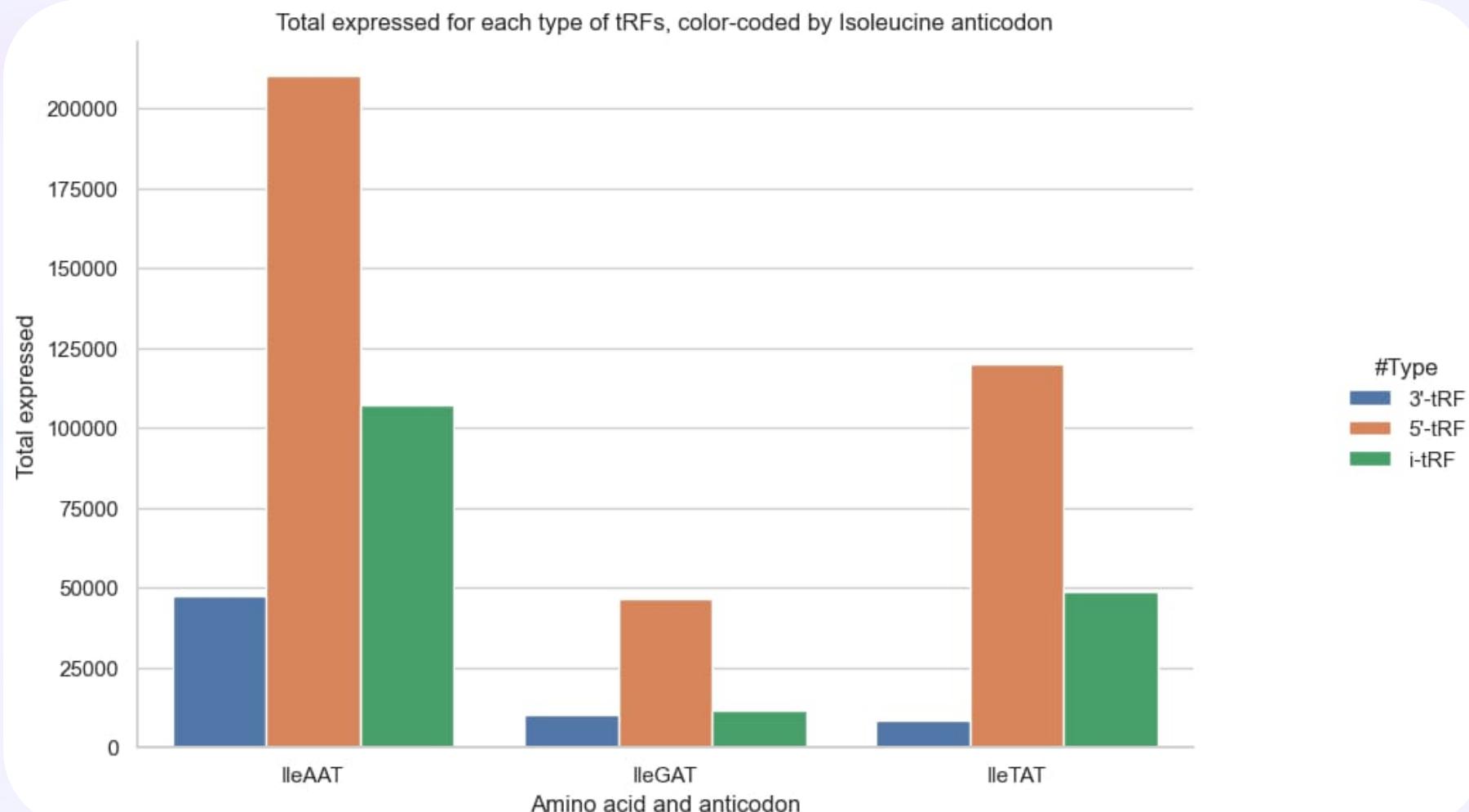
#5 tRNA: what kind of tRNA do we have in cancer cells?

- tRFs (tRNA-derived RNA fragments) are small non-coding RNA molecules derived from the processing of mature or precursor tRNAs. (MINTbase)
- found to play roles in various cellular processes, including gene regulation, stress response, and RNA silencing.
- tRFs can be classified into several types based on their origin and biogenesis, such as tRF-5, tRF-3, and tRF-1 series.

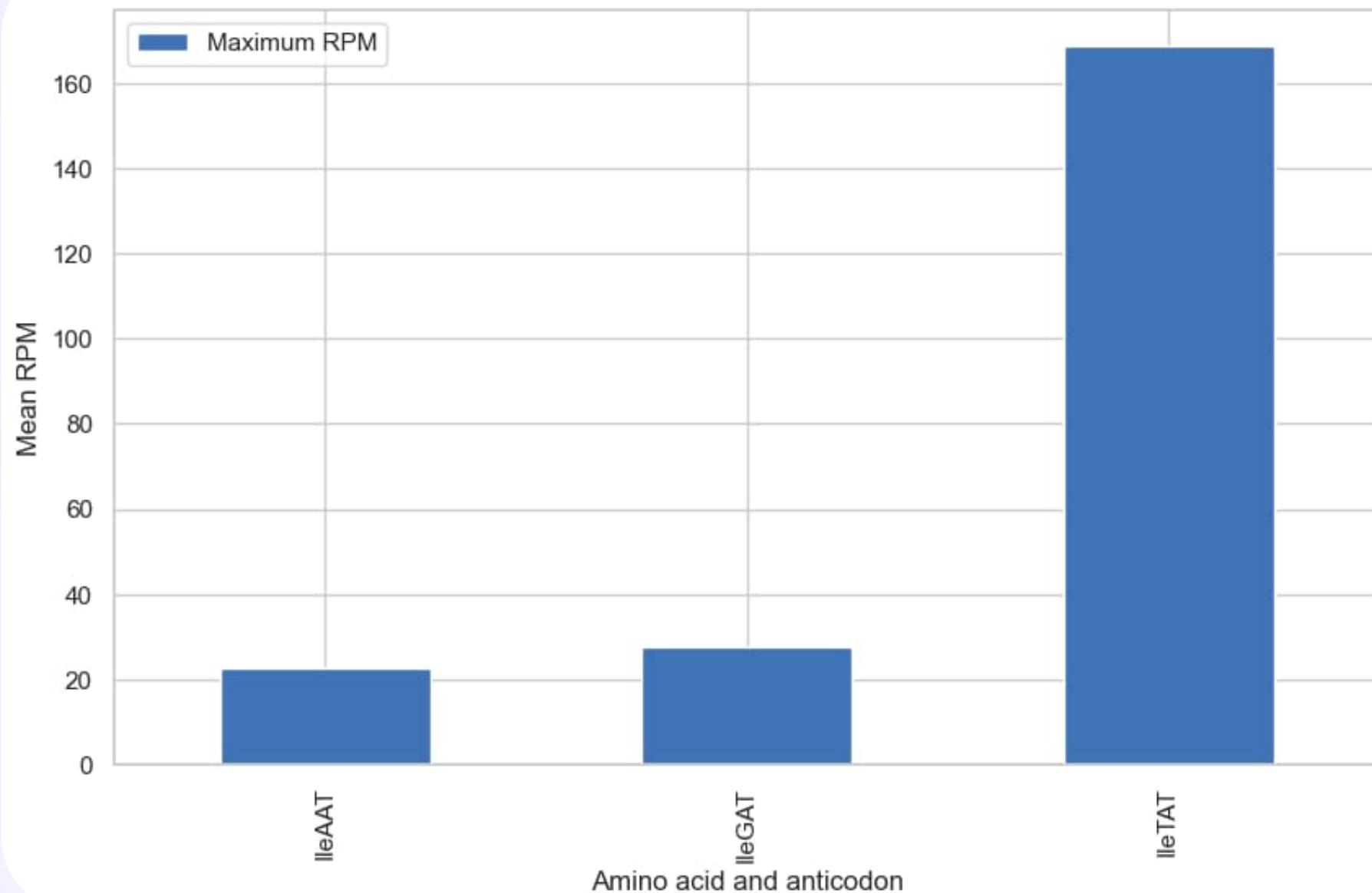
```
IleAAT      2088  
IleTAT      579  
IleGAT      402
```

```
Name: Amino acid and anticodon, dtype: int64
```

so, we're expecting 5'ATA3' isoleucine codon to be dominant. In that case, anticodon tRNA cognate, 5'TAT3' should have been found more, but **that is not the case.**



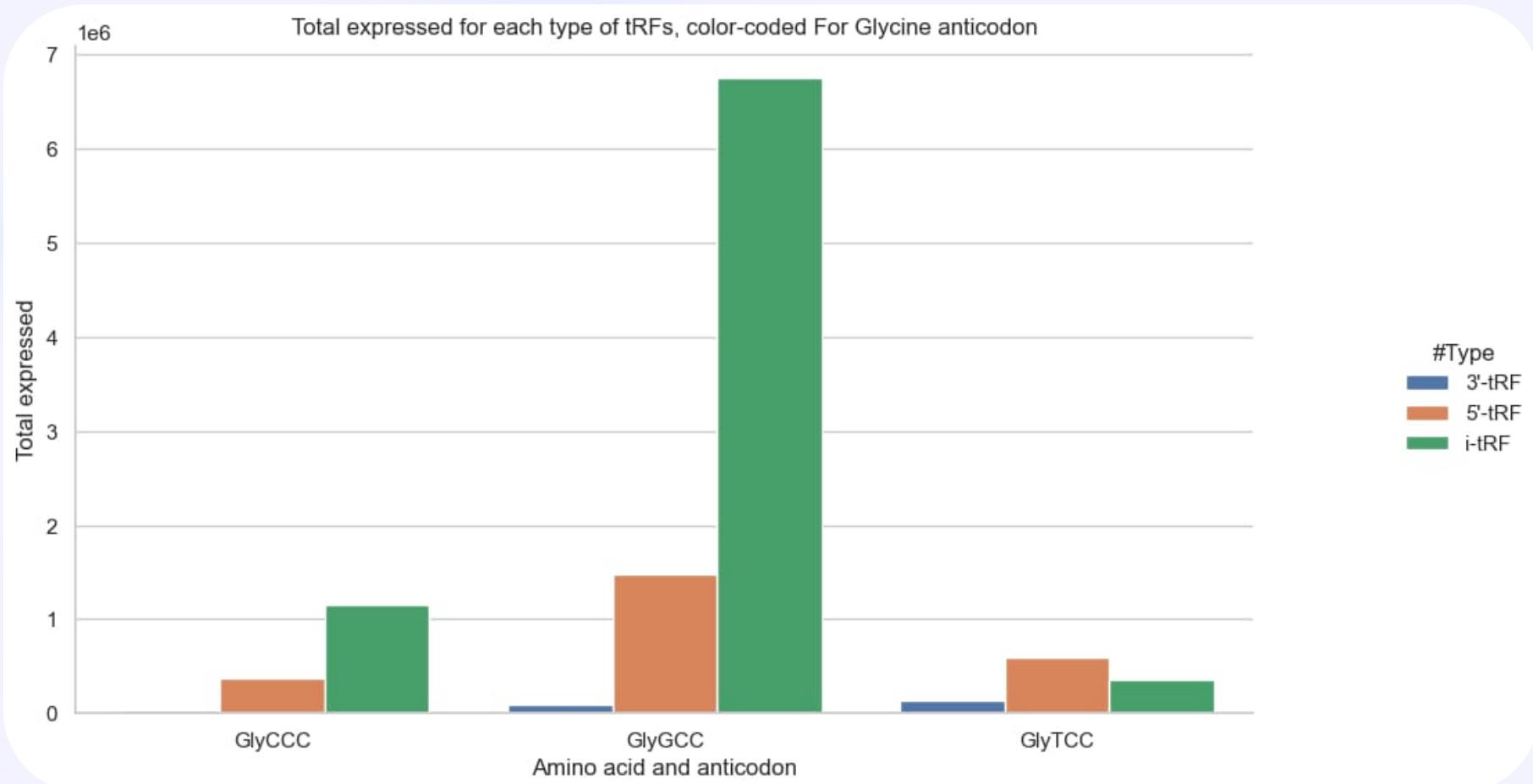
Mean RPM for each Amino acid and anticodon



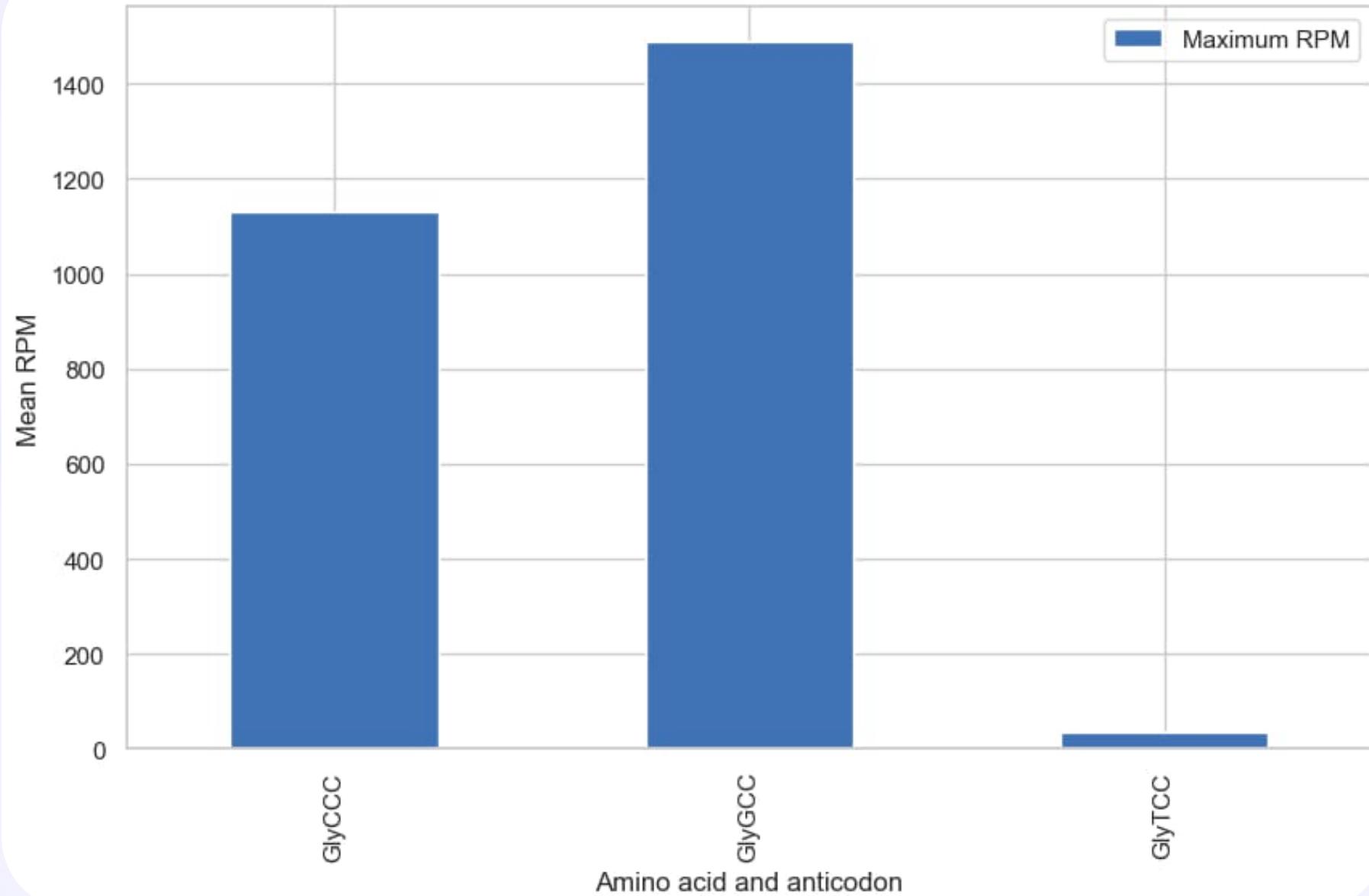
#5.1

glycine trfs

- codons all have relatively similar usage frequencies, and there is not a strong bias towards using one codon over the others. This makes glycine one of the amino acids with the most stable codon usage patterns.



Mean RPM for each Amino acid and anticodon



References

1. Thijs Nieuwkoop, Barbara R Terlouw, Katherine G Stevens, Richard A Scheltema, Dick de Ridder, John van der Oost, Nico J Claassens, Revealing determinants of translation efficiency via whole-gene codon randomization and machine learning, *Nucleic Acids Research*, Volume 51, Issue 5, 21 March 2023, Pages 2363–2376,
<https://doi.org/10.1093/nar/gkad035>
2. Earnest-Noble, L. B., Hsu, D., Chen, S., Asgharian, H., Nandan, M., Passarelli, M. C., Goodarzi, H., & Tavazoie, S. F. (2022). Two isoleucyl tRNAs that decode synonymous codons divergently regulate breast cancer metastatic growth by controlling translation of proliferation-regulating genes. *Nature cancer*, 3(12), 1484–1497.
<https://doi.org/10.1038/s43018-022-00469-9>