```python
In [1]:  import numpy as np
         from CalcMRate2 import CalcMRate

         c = CalcMRate("./bed_data/illumina_hg38_exome.bed", "./fasta_data/hg38.fa", "./extractio
```

```python
In [2]:  trinlist, exome_spec32vec = c.get_spec32()
```

```
GENERATION OF 32-vec DESCRIPTOR | EXTRACTED SEQUENCES
100%|████████████████████████████████████████████████████████████████
██████| 287879/287879 [00:39<00:00, 7360.37it/s]
```

```python
In [4]:  import matplotlib.pyplot as plt

         f, ax = plt.subplots(figsize=(32,5)) # set the size that you'd like (width, height)
         plt.bar(trinlist[:16], exome_spec32vec[:16])
         plt.bar(trinlist[16:], exome_spec32vec[16:])
         plt.title('EXOME TRINUCLEOTIDE FREQUENCIES')
         plt.xlabel('Trinucleotides')
         plt.ylabel('Frequency')
         ax.legend(["C Trinucleotide Frequencies", "T Trinucleotide Frequencies"], fontsize = 14)
```

Out[4]:   <matplotlib.legend.Legend at 0x10923b700>
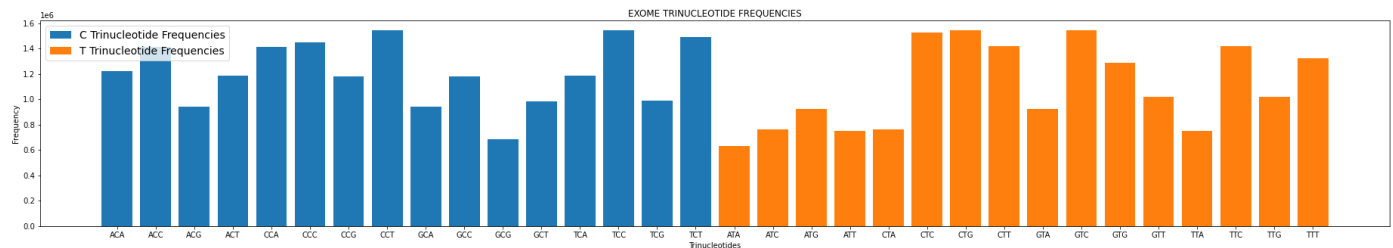

```python
In [5]:  hg38_spec32vec = np.load('hg38-32vec.npy')

         f, ax = plt.subplots(figsize=(32,5)) # set the size that you'd like (width, height)
         plt.bar(trinlist[:16], hg38_spec32vec[:16])
         plt.bar(trinlist[16:], hg38_spec32vec[16:])
         plt.title('HUMAN GENOME (HG38) TRINUCLEOTIDE FREQUENCIES')
         plt.xlabel('Trinucleotides')
         plt.ylabel('Frequency')
         ax.legend(["C Trinucleotide Frequencies", "T Trinucleotide Frequencies"], fontsize = 14)
```

Out[5]:   <matplotlib.legend.Legend at 0x117cc0190>
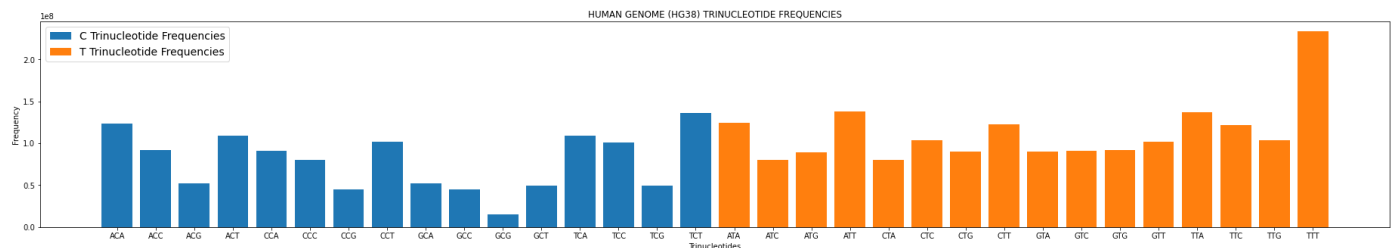

```python
In [6]:  np.save('illumina_hg38_exome-32vec.npy', exome_spec32vec)
```

```python
In [7]:  normalized_exome_spec32vec = exome_spec32vec/np.sum(exome_spec32vec)
         normalized_hg38_spec32vec = hg38_spec32vec/np.sum(hg38_spec32vec)
```
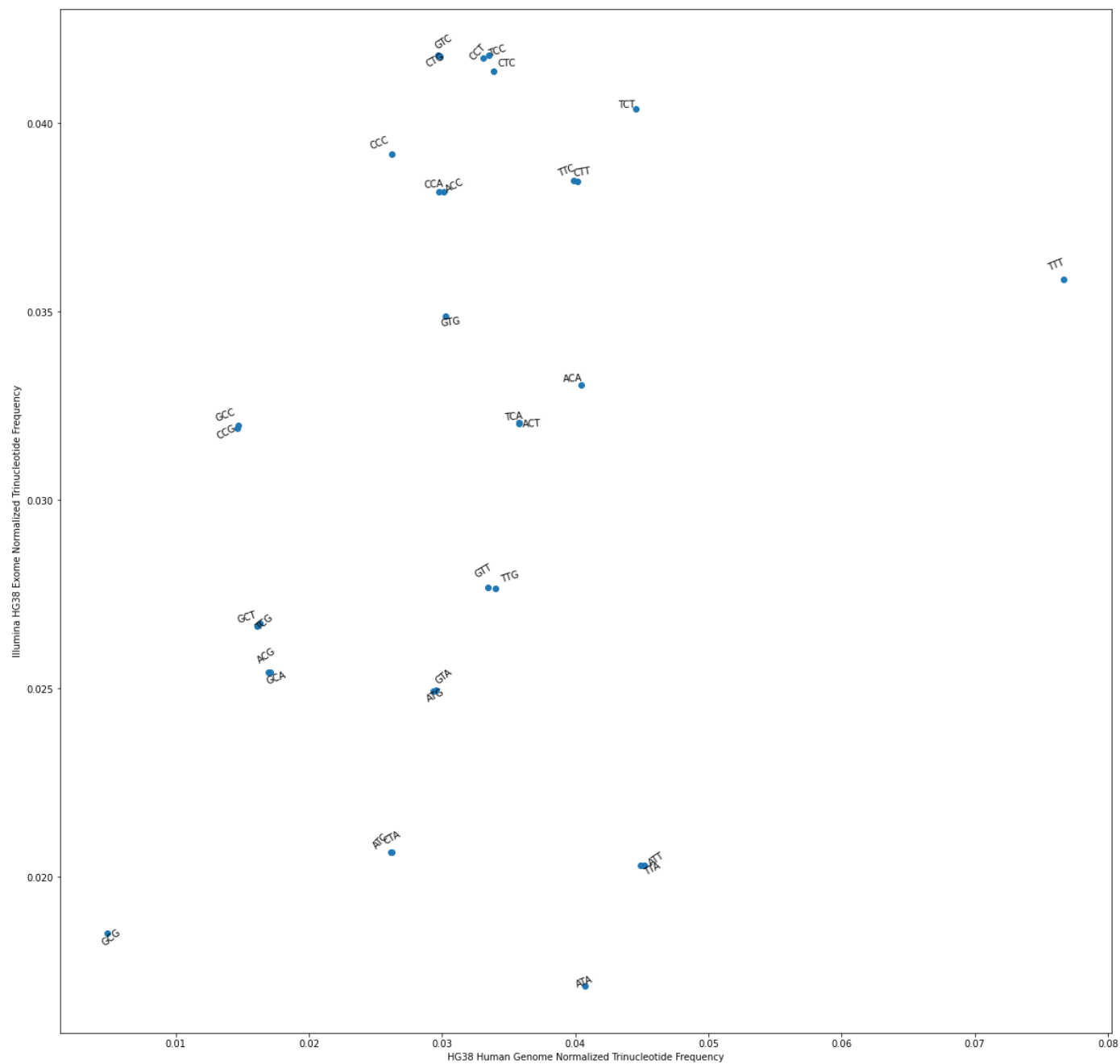
```python
In [49]: import random

         plt.figure(figsize=(20, 20))
         plt.scatter(normalized_hg38_spec32vec, normalized_exome_spec32vec, marker='o')
         plt.xlabel('HG38 Human Genome Normalized Trinucleotide Frequency')
```

```
plt.ylabel('Illumina HG38 Exome Normalized Trinucleotide Frequency')

for i, (x_val, y_val) in enumerate(zip(normalized_hg38_spec32vec.tolist(), normalized_ex
        plt.annotate(trinlist[i], (x_val, y_val), textcoords="offset points", xytext=(ra
#       plt.text(x_val, y_val, trinlist[i], ha='center', va='bottom')
```


In [ ]: