CS 660 Lab 1
Professor Johnson

Group Members:
Saffat Hasan
Lakshmi Maguluri
Zexi Yu
Manoj Venkatachalaiah

The Size of the Human Genome

The **human genome** is the complete set of nucleic acid sequences for **humans**, encoded as DNA within the 23 chromosome pairs (about 3 billion base pairs long and contains around 30,000 genes) in cell nuclei and in a small DNA molecule found within individual mitochondria. Four distinct base pairs represented by the characters A (Adenine), T (Thymine), G (Guanine), C (Cytosine)

Each letter can be represented by 2 bits (00, 01, 10, 11), so the 3 billion base pairs of the haploid human genome correspond to a maximum of about 725 megabytes of data. Since individual genomes vary by less than 1% from each other, they can be losslessly compressed to roughly 4 megabytes.[1]

Assuming there are approximately 8 billion people on earth, we would need the following amount of storage space to record each individual's genome:

4 megabytes (compressed) * 8 billion people = 12 terabytes of data (compressed)

725 megabytes (uncompressed) * 8 billion people = 5.8 exabytes of data (uncompressed)

Sources Cited

1. https://en.wikipedia.org/wiki/Human_genome