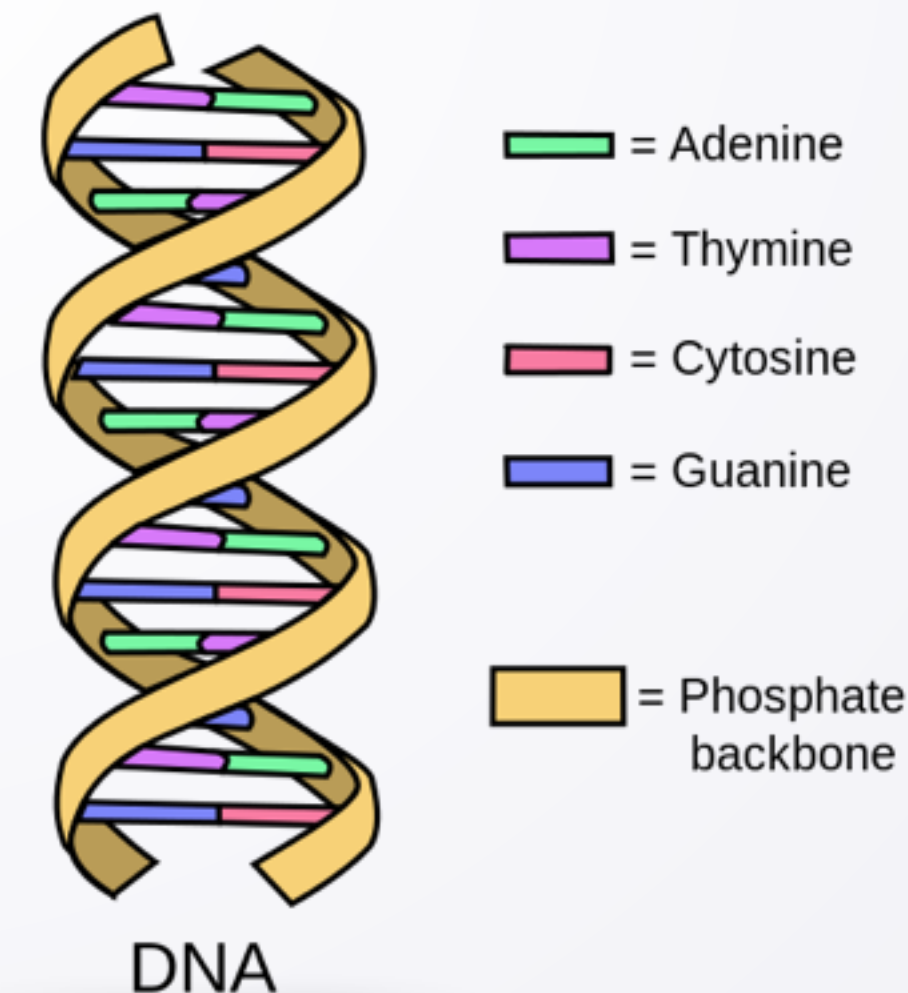# StorageResource Class
Methods for Finding and Storing

Finding and Processing Data

- You've seen code for finding start and stop codons: simple extension to finding genes

  - What if genes must have 50+ nucleotides?

  - What if genes must have CG content > 60%?

  - How do we process all the genes

= Adenine

= Thymine

= Cytosine

= Guanine

= Phosphate backbone

DNA

Finding and Processing Data

- You've seen code for finding start and stop codons: simple extension to finding genes
    - What if genes must have 50+ nucleotides?
    - What if genes must have CG content > 60%?
    - How do we process all the genes
- Simple extension to web crawling
    - How do we find links in a web page
    - Search for href=".." instead of ATG..TAA

Finding All Genes

- Printing all genes found in DNA
  - **findStopIndex()** discussed in earlier lesson

```java
public void printAll(String dna){
    int start = 0;
    while (true){
        int tag = dna.indexOf("atg",start);
        if (tag == -1) {
            break;
        }
        int end = findStopIndex(dna,tag+3);

        if (end != dna.length()) {
            System.out.println(dna.substring(tag,end+3));
            start = end+3;
        }
        else {
            start = start + 3;
        }
    }
}
```

Duke
UNIVERSITY

Finding All Genes

- Printing all genes found in DNA

  - **findStopIndex()** discussed in earlier lesson

```java
public void printAll(String dna){

    int tag = dna.indexOf("atg",start);
    if (tag == -1) {
        break;
    }
}

        if (end != dna.length()) {
            System.out.println(dna.substring(tag,end+3));
            start = end+3;
        }
        else {
            start = start + 3;
        }
    }
}
```

Finding All Genes

- Printing all genes found in DNA

  - **findStopIndex()** discussed in earlier lesson

```java
public void printAll(String dna){
    int start = 0;
    while (true){
        int tag = dna.indexOf("atg",start);
        if (tag == -1) {
            int end = findStopIndex(dna,tag+3);

            if (end != dna.length()) {
                System.out.println(dna.substring(tag,end+3));
                start = end+3;
            }
            else {
                start = start + 3;
            }
        }
    }
}
```

Storing Strings: Post-Processing

- Modifying the gene finding/printing code

    - To find/print genes that are long

    - To find/print genes that are good

    - Combines finding and printing, separating hard

- StorageResource class helps here

    - Can add strings with `.add(str)`

    - Can ask for count of added with `.size()`

    - Can use as iterable with `.data()`

Duke
UNIVERSITY

Post-Processing Example

- If we had stored all genes in a StorageResource object we could …

  - Print those long enough

```
public void printGenes(StorageResource sr){
    for(String gene : sr.data()){
        if (gene.length() > 60) {
            System.out.println(gene.length()+"\t"+gene);
        }
    }
}
```

Post-Processing Example

- If we had stored all genes in a StorageResource object we could …

  - Print those long enough

  - Print those rich enough in CG content

```java
public void printGenes(StorageResource sr){
    for(String gene : sr.data()){
        if (cgRatio(gene) > 0.35) {
            System.out.println(gene.length()+"\t"+gene);
        }
    }
}
```

Adding to a StorageResource

- Adding words to StorageResource object
  - Simply call `.add()` to add String to collection
  - Suppose tracking only words not seen before?

```java
public StorageResource getWords(){
    FileResource fr = new FileResource("/data/confucius.txt");
    StorageResource store = new StorageResource();
    for(String w : fr.words()){
        if (! store.contains(w)){
            store.add(w);
        }
    }
    return store;
}
```