

B4M36SAN

Final projects

General data-science task

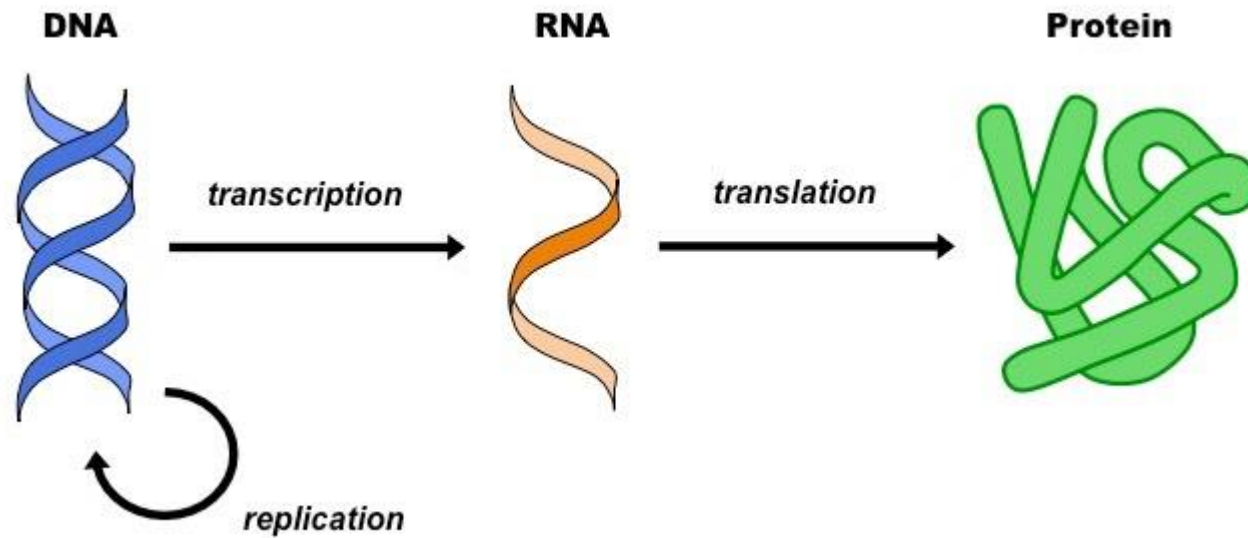
Bioinformatics task

Biotask: Intron detection

- To understand introns, we need some background first:
 - How DNA encodes the functioning of organisms?
 - What are the *intron sequences*?
 - what role they play in this process?
- Task
 - Distinguish introns from coding sequences of DNA

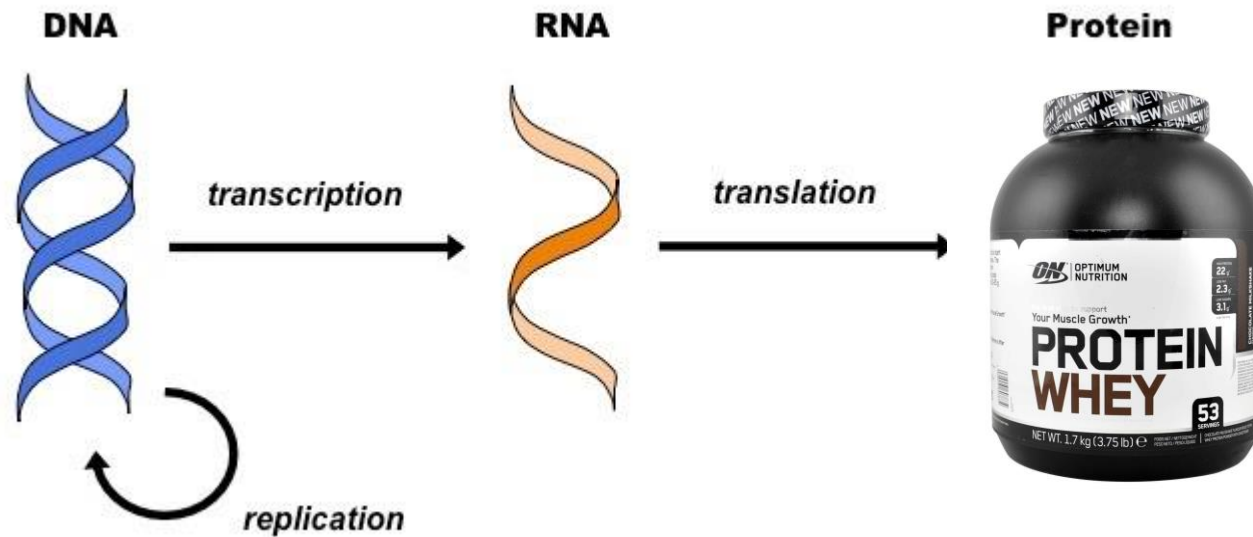
Biotask: Intron detection

- How DNA encodes functioning of organisms?



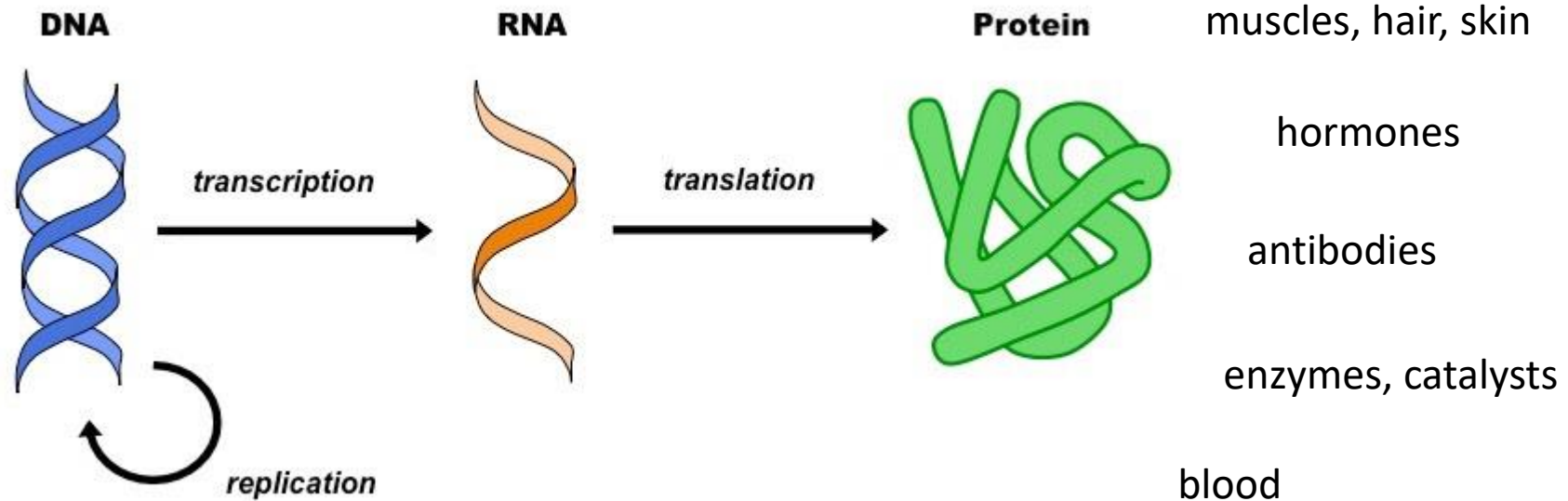
Biotask: Intron detection

- How DNA encodes functioning of organisms?



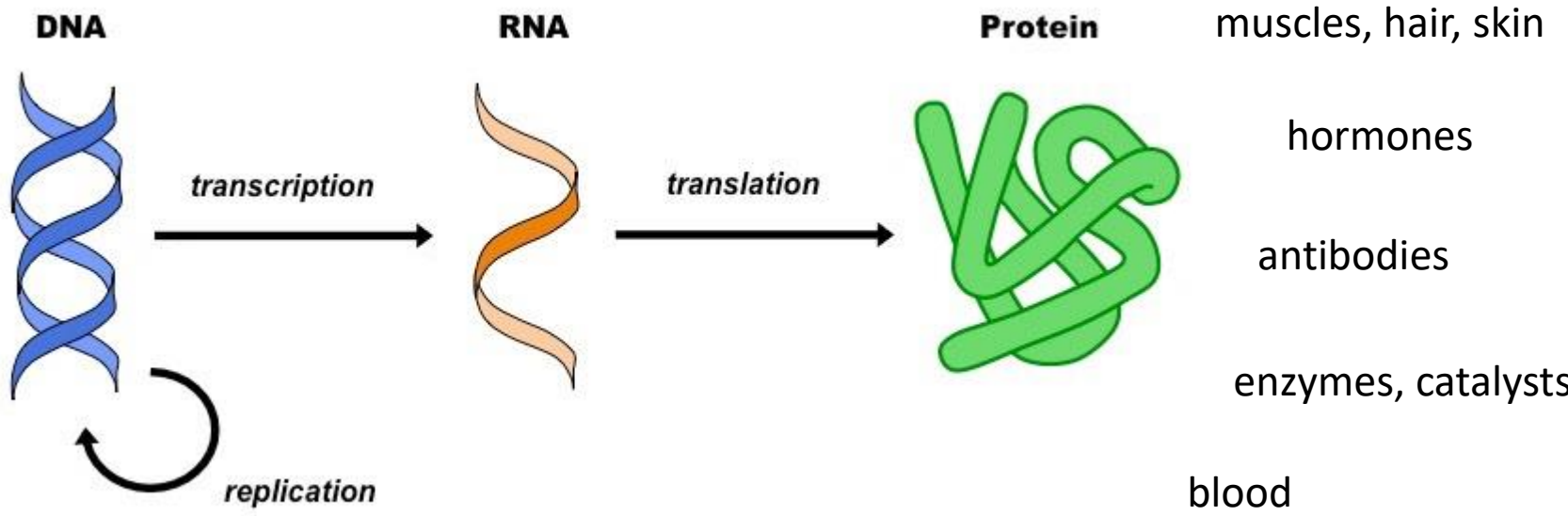
Biotask: Intron detection

- How DNA encodes functioning of organisms?

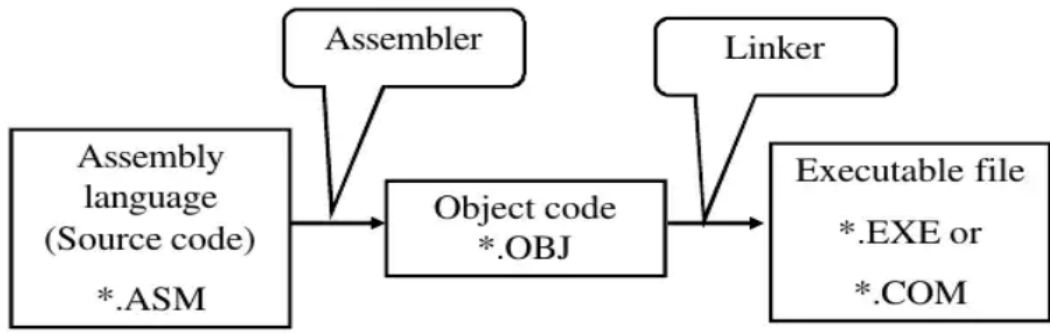


Biotask: Intron detection

- How DNA encodes functioning of organisms?



CTGGGAGTCCAA
TTCACGTACTT...



LHEMQGILEED
NRDCAQHLKKI...

Biotask: Intron detection

- DNA as source code

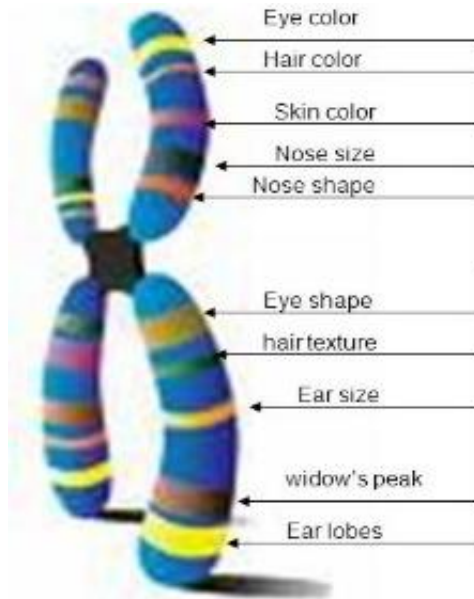
```
ATGAACTCGGAATCAGTACGCATTTATCTCGTTGCTGCGATGGGAGCCAATCGGGTTATTGGCAATGGTC
CTAATATCCCCTGGAAAATTCCGGGTGAGCAGAAGATTTTTTCGCAGACTCACTGAGGGAAAAGTCGTTGT
CATGGGGCGAAAGACCTTTGAGTCTATCGGCAAGCCTCTACCGAACCGTCACACATTGGTAATCTCACGC
CAAGCTAACTACCGCGCCACTGGCTGCGTAGTTGTTTCAACGCTGTCGCACGCTATCGCTTTGGCATCCG
AACTCGGCAATGAACTCTACGTCGCGGGCGGAGCTGAGATATACACTCTGGCACTACCTCACGCCACGG
CGTGTTCCTATCTGAGGTACATCAAACCTTCGAGGGTGACGCCTTCTTCCCAATGCTCAACGAAACAGAA
TTCGAGCTTGTCTCAACCGAAACCATTCAAGCTGTAATTCCGTACACCCACTCCGTTTATGCGCGTCGAA
```

DNA

```
ComputewithinScatter <- function(data, n)
{
  # ADD YOUR CODE
}

ComputeBetweenScatter <- function(data, n, meanOverall)
{
  # The between-classes scatter matrix
  return(betweenMatrix)
}

SolveEigenProblem <- function(withinMatrix, betweenMatrix, prior)
{
  # solve the generalized eigenvalue problem to obtain the linear discriminants
  return(eivectors)
}
```



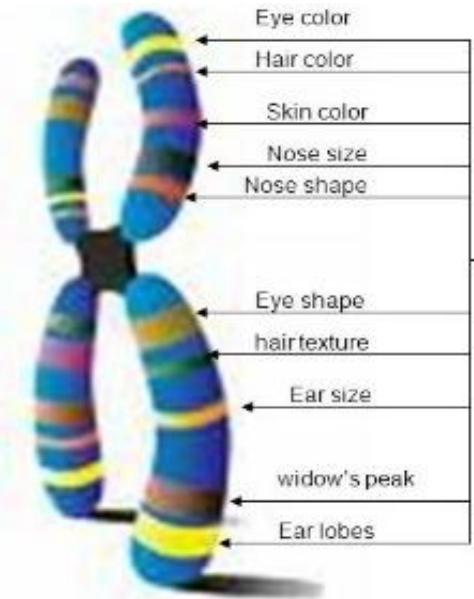
Chromosome

genes

Biotask: Intron detection

- DNA as source code

```
ATGAACTCGGAATCAGTACGCATTTATCTCGTTGCTGCGATGGGAGCCAATCGGGTTATTGGCAATGGTC
CTAATATCCCCTGGAAAATTCCGGGTGAGCAGAAGATTTTTTCGCAGACTCACTGAGGGAAAAGTCGTTGT
CATGGGGCGAAAGACCTTTGAGTCTATCGGCAAGCCTCTACCGAACCGTCACACATTGGTAATCTCACGC
CAAGCTAACTACCGCGCCACTGGCTGCGTAGTTGTTTCAACGCTGTCGCACGCTATCGCTTTGGCATCCG
AACTCGGCAATGAACTCTACGTCGCGGGCGGAGCTGAGATATACACTCTGGCACTACCTCACGCCACGG
CGTGTTCCTATCTGAGGTACATCAAACCTTCGAGGGTGACGCCTTCTTCCCAATGCTCAACGAAACAGAA
TTCGAGCTTGTCTCAACCGAAACCATTCAAGCTGTAATTCCGTACACCCACTCCGTTTATGCGCGTCGAA
```



Chromosome

genes

DNA

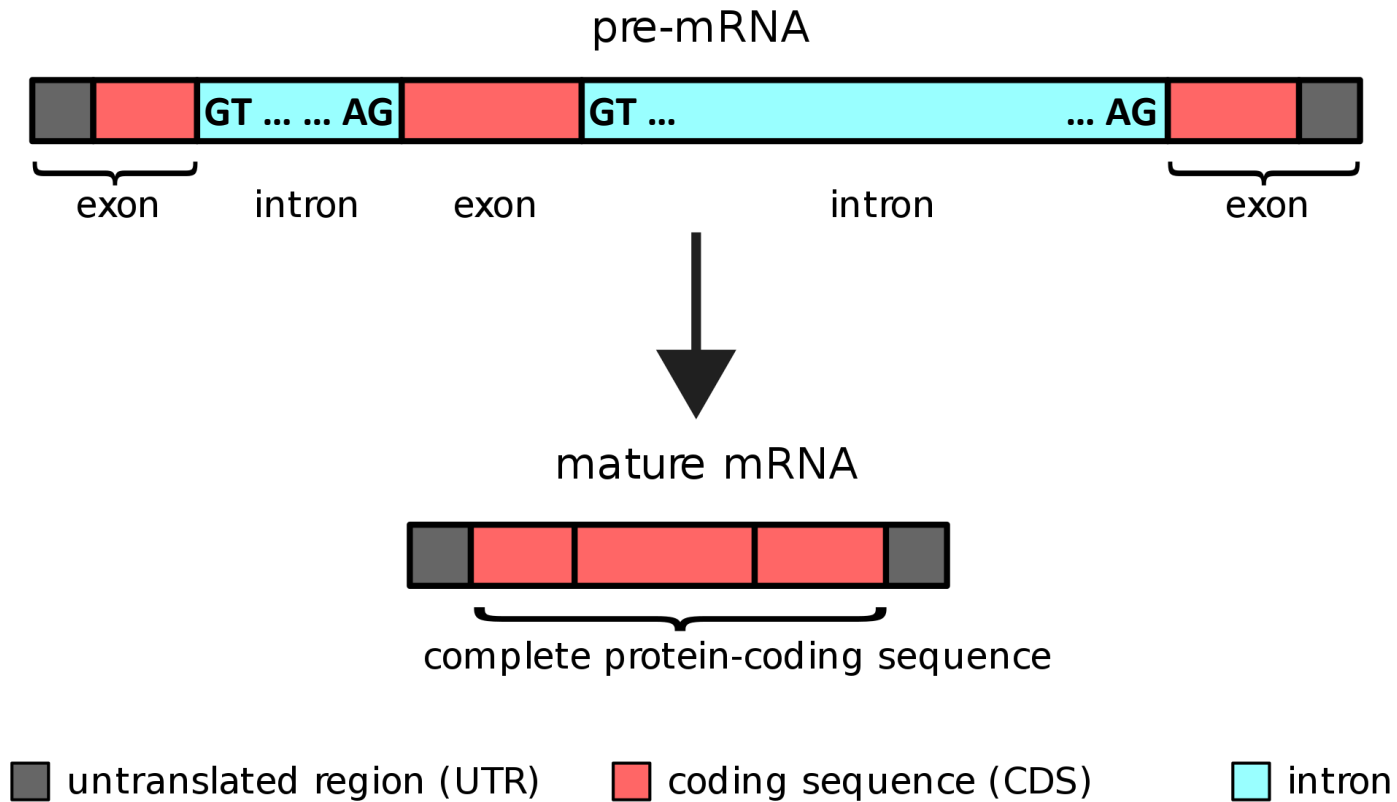
introns

```
ComputewithinScatter <- function(data, n)
{
  # ADD YOUR CODE
}

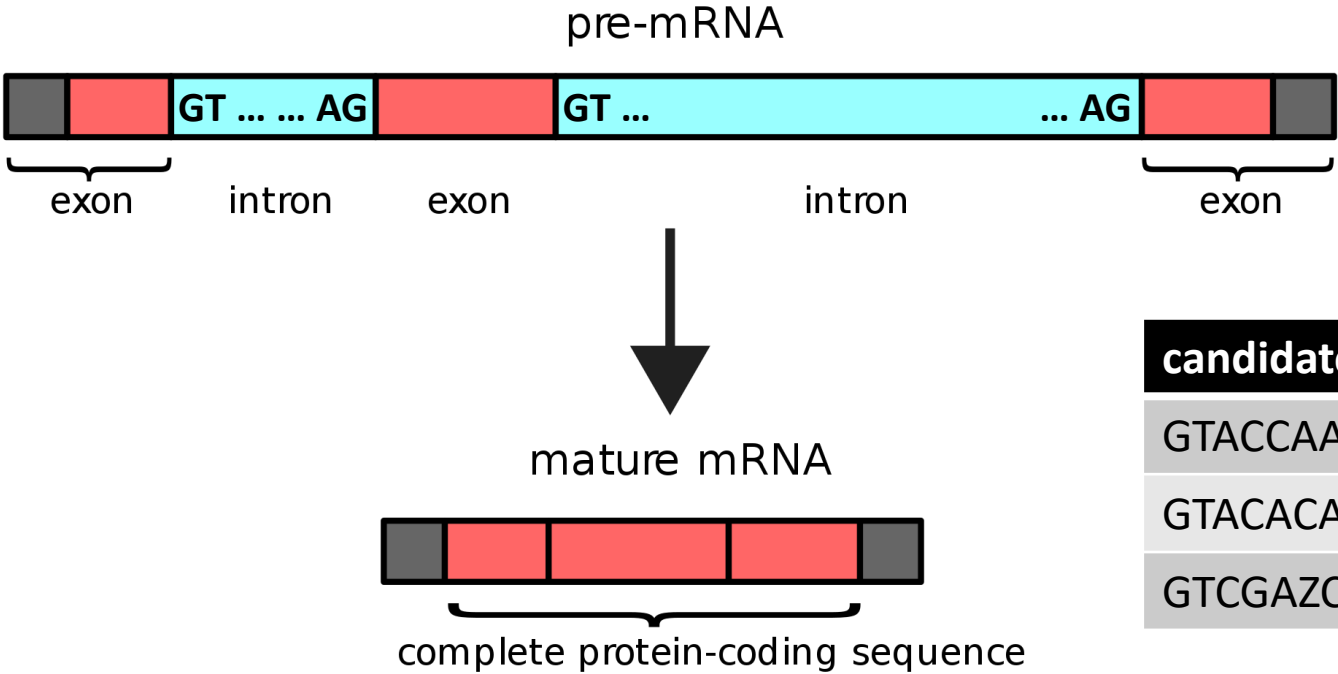
ComputeBetweenScatter <- function(data, n, meanOverall)
{
  # The between-classes scatter matrix
  return(betweenMatrix)
}

SolveEigenProblem <- function(withinMatrix, betweenMatrix, prior)
{
  # solve the generalized eigenvalue problem to obtain the linear discriminants
  return(eivectors)
}
```


Biotask: Intron detection



Biotask: Intron detection



candidate sequence	is_intron
GTACCAAAGTCTACCATGACAAG	0
GTACACACCTGTATAGCACGACCCTTACAG	1
GTCGAZCTTCTATCGCTAGCTACGACATCGCAG	0

untranslated region (UTR)
 coding sequence (CDS)
 intron

Data Science task: Employee attrition

- IBM artificial dataset accessible from Kaggle ([IBM HR Analytics Employee Attrition & Performance | Kaggle](#))
- Major subtasks
 - EDA (Exploratory Data Analysis)
 - Dimensionality reduction
 - Prediction models
 - Verify assumptions (if any)
 - Proper training, parameter tuning, testing
 - Interpretation, feature transformations (e.g. standardization)



Scientists Have Discovered Hidden Annotations In DNA Code Left By God And It Will Terrify You

@welcometomymemepage
@wtmmp

```
human_7605232288.exe
File Edit Help
//made he mouth way too big lmao
//im leavin it in LOL
    CGTAGCTAGTCATGCAGATGGTATGGTCCTTA
    GAAATCGCCGTTGGGCAGAAATTTATCCGGAG
//this guy is gonna be so DUMB
    CCGTTACGCTTGATGATATGGTGGAGTGCCGT
    GAATCGCCGAAGTAGCGTTGGGTCGGATCGAT
    AAGTCCTAGTGCGTATAACAGTTTTCCCGCGTA
//im just put some baboon code
in here n see what happens
    CGAGTGCTGATCGTGGGTAATTGGTCCGGTCAT
//spent 7 hours writing this part
just to fuck up his toes
    CCGTTTAGCGGTGAGCGAGTTGCGATTCGTGA
    GGTGGAGTCGAGTGGGTGAGCGTAGCCGTAGC
```

- [image.png \(698×315\) \(steemitimages.com\)](#)
- [on-protein-whey.jpg \(800×800\) \(predatornutrition.com\)](#)
- [8086-architecture-43-728.jpg \(728×563\) \(slidesharecdn.com\)](#)
- [Pre-mRNA to mRNA MH - Intron – Wikipedia](#)
- [Chromosomes and Genes - W3spoint](#)
- [\(3\) Scientists have discovered hidden annotations in DNA code left by God and it will terrify you : ProgrammerHumor \(reddit.com\)](#)