

**CMPSC 300**  
**Bioinformatics**  
**Fall 2019**

**Lab 4:**  
**Genomic Regions Associated with Parkinson's Disease**

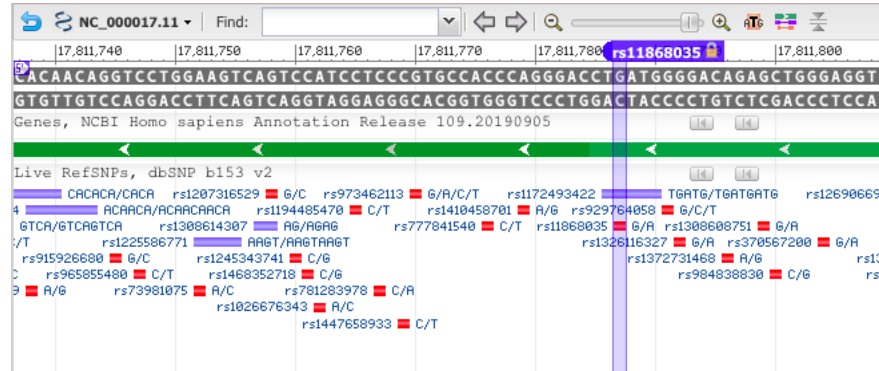


Figure 1: NCBI's Genomic regions, transcripts, and products.

**GitHub starter link**

<https://classroom.github.com/a/k4uhy6MT>

To use this link, please follow the steps below.

- Click on the link and accept the assignment.
- Once the importing task has completed, click on the created assignment link which will take you to your newly created GitHub repository for this lab.
- Clone this repository (bearing your name) and work on the practical locally.
- As you are working on your practical, you are to commit and push regularly. You can use the following commands to add a single file, you must be in the directory where the file is located (or add the path to the file in the command):

```
- git add -A
- git commit -m 'Your notes about commit here'
- git push
```

Alternatively, you can use the following commands to add multiple files from your repository:

```
- git commit <nameOfFile> -m 'Your notes about commit here'
```

– git push

Be sure to read the README.md file in the GitHub Classroom repository for instructions on how to complete your first assignment.

## Objectives

- To learn how to use a Web-based genomic databases and tools.
- To understand some of the types of information that are stored in PubMed databases.
- To learn how to use different interfaces to find and retrieve genomic information.

## Locate your data on PubMed

**PubMed** (available at: <https://www.ncbi.nlm.nih.gov/pubmed/>), is a large, online, database resource containing information regarding DNA, genes, RNA, protein products, literature and graphical resources such as, *Genomic regions, transcripts, and products*, as shown in Figure 1. These resources are curated for bioinformatics and biological science meaning that the information comes from an educated and factual basis. As written on their website, PubMed comprises more than 30 million citations for biomedical literature from MEDLINE, life science journals, and online books. Citations may include links to full-text content from PubMed Central and publisher web sites. This great resource provides much of the data that bioinformaticians may use during the course of their work in research.

**Single-nucleotide polymorphisms** (SNPs, pronounced *snips*) are DNA sequence variations occurring when a single nucleotide (i.e., adenine (A), thymine (T), cytosine (C), or guanine (G) in the genome, or other shared sequence) differs between members of a species or paired chromosomes in an individual. As mentioned in class, the appearance of these SNPs may be used to separate DNA samples from each other based on the presence or absence of a particular type of nucleotide.

For instance, the research group, Do *et al.* [1] genotyped 3,426 PD patients and 29,624 healthy control individuals for 522,782 known simple nucleotide polymorphisms (SNPs). They identified 11 SNP sites where one allele was correlated with PD with a statistically significant frequency. Known (SNPs) in the human genome are recorded in the primary genomic database dbSNP, where each SNP has a unique accession number that identifies it. In this laboratory assignment you will investigate the genomic neighborhood of one of these SNPs with the accession number *rs11868035*, which was identified by Do *et al.*

## Part 1: PubMed's Online Resource

Please locate the markdown file, `writing/stepsAndQuestions.md` and follow the instructions to complete the thirteen (13) questions found in this file.

## Part 2: The Ethics of Online Information

PubMed is a resource from which biological and health-sciences data may be located. A majority of the data on this site has been *curated* by researchers, meaning that there is some form of credibility behind each piece of the site's information. For instance, the literature that PubMed offers has

all been peer-reviewed to imply that a particular article's foundations, methods and conclusions were created using an educated, informed and logical conduct. Researchers who use the PubMed resource are therefore able to assume that the facts that they gain from the PubMed resource are real and worthy of their considerations for the determination of other types of truth from research.

There are sites online who freely offer media which has not been curated for faults and flaw. For this information there is no approval of any kind that the information may be used for any reason. Sadly, many consumers of this information apply this unconfirmed knowledge to their daily lives.

Please read the Washington Post article entitled, *They turn to Facebook and YouTube to find a cure for cancer and get sucked into a world of bogus medicine*, ([article link](#), [2]), and complete the questions found in the file `ethics/reflections.md`.

### Required Deliverables

- Your report from your adventures with NCBI's online resource: `writing/stepsAndQuestions.md`
- Your reflection document to respond to the questions in the article: `ethics/reflections.md`

Please see the instructor if you have questions about assignment or its submission.

### References

- [1] Chuong B Do, Joyce Y Tung, Elizabeth Dorfman, Amy K Kiefer, Emily M Drabant, Uta Francke, Joanna L Mountain, Samuel M Goldman, Caroline M Tanner, J William Langston, et al. Web-based genome-wide association study identifies two novel loci and a substantial genetic component for parkinson's disease. *PLoS genetics*, 7(6):e1002141, 2011.
- [2] Abby Ohlheiser. They turn to facebook and youtube to find a cure for cancer - and get sucked into a world of bogus medicine. *The Washington Post*, June 2019.