

Group 6:

Tiffany Cook

Drashti Mehta

Joshua Mikombo

### **BINF 6201 Lab Report 1**

- **Part 1: Find Genes in Candidate Regions**
  - 1A. What is the total number of protein-coding genes that fall within a candidate region?
    - 17 protein-coding genes in the candidate regions
  - 1B. How did you find these genes?
    - (NCBI database) ("Calypste anna"[Organism] OR ("Calypste anna"[Organism] OR ("Calypste anna"[Organism] OR Calypste anna[All Fields]))) AND ("genotype protein coding"[Properties] AND (NC\_044245[nucl\_accn] AND 000056181726[CHRPOS] : 000056876451[CHRPOS]) OR (NC\_044274[nucl\_accn] AND 000043020578[CHRPOS] : 000043645624[CHRPOS])) OR (NC\_044248[nucl\_accn] AND 000032923716[CHRPOS] : 000033354545[CHRPOS]))
- **Part 2: Find Known Genes Related to Pigmentation/Melanin**
  - 2A. What is the total number of protein-coding genes from humans and mice with known phenotypes related to melanin and/or pigmentation?
    - Ans. 35 protein coding genes in humans that have known phenotypes and/or diseases related to pigmentation and melanin
  - 2B. How did you find these genes?
    - (NCBI database) Performed filter with “disease/phenotype” for all the words which were related to what we were searching for.
    - Specified organism as “Homo sapiens”
    - Query: (pigment[Disease/Phenotype] OR pigmentation[Disease/Phenotype] OR pigmented[Disease/Phenotype] OR hyperpigmented[Disease/Phenotype] OR hyperpigmentation[Disease/Phenotype] OR hypopigmented[Disease/Phenotype] OR hypopigmentation[Disease/Phenotype] OR melanocyte[Disease/Phenotype] OR melanosome[Disease/Phenotype] OR melanin[Disease/Phenotype]) AND (homo sapiens[Organism])
    - Specified for protein-coding genes only

- Part 3: Find the Overlap Between Gene Sets
  - 3A. Which gene(s) from the candidate gene regions overlap with the known pigmentation genes?
    - **TYRP1**
  - 3B. How did you find the overlap?
    - **Using the following python script**

```
import pandas as pd

pigment = pd.read_csv("pigment.csv")
hummingbird = pd.read_csv("hummingbird.csv")

pigment1 = pigment['Symbol']
hummingbird1 = hummingbird['Symbol']

for i in pigment1:
    for j in hummingbird1:
        if i == j:
            print(i)
```

- Part 4: Find More Details on the Top Candidate Gene
  - 4A. Where (i.e., in which tissue) is this gene most highly expressed in humans?
    - **The skin tissue (NCBI database)**
  - 4B. What biological processes is this gene involved in?
    - **Melanin biosynthetic and acetoacetic acid metabolic processes, along with melanocyte differentiation and melanosome organization (NCBI database)**
  - 4C. What family of proteins does the product of this gene belong to, and how many other members of this family are found in humans?
    - **Tyrosinase family**
    - **15 members**
    - **(UniProt database)**
  - 4D. What (if any) human diseases are associated with mutations in this gene/protein?
    - **Rufous oculocutaneous albinism and oculocutaneous albinism type III (NCBI database)**