Python and Omics Data

Practice solving problems in Python, and an introduction to Rosalind

Class purpose

- Practise developing solutions to data processing problems using omics data as an example data type
- Figure out how to find answers to questions you might have about Python or generated error messages

Class format

1. Review data processing examples in Python
2. Learn about Rosalind and how to complete and submit answers to this platform

What is data processing?

- The collection and manipulation of items of data to produce meaningful information
- Other examples of data processing
  - General: recategorization or grouping of data
  - Text analysis: grouping words per article, per book, or per metadata element
  - Microbiology: changing DNA and RNA strings

Task: counting DNA nucleotides

Return four integers counting the respective number of times the symbols 'A', 'C', 'G', and 'T' occur in the sample DNA string

Sample data ('string'):

AGCTTTTCATTCTGACTGCAACGGGCAATATGTCTCTGTGTGGATTAAAAAAAGAGTGTCTGATAGCAGC

Sample output:

20 12 17 21

Task: transcribing DNA into RNA

An RNA string is a string formed from the alphabet containing 'A', 'C', 'G', and 'U'.

Replace all occurrences of 'T' in the sample dataset with 'U'.

Sample dataset:

`GATGGAACTTGACTACGTAAATT`

Sample output:

`GAUGGAACUUGACUACGUAAAUU`

Task: reverse complementing a strand of DNA

In DNA strings, symbols 'A' and 'T' are complements of each other, as are 'C' and 'G'. The reverse complement of string is formed by reversing the symbols of the strand, then taking the complement of each symbol.

Sample dataset:

`AAAACCCGGT`

Sample output:

`ACCGGGTTTT`

After this class

- Schedule a consultation meeting with me
- Continue to visit rosalind.info/problems/locations to find more bioinformatics problems/prompts

Feedback and Questions

Contact me at sawyer.newman@yale.edu

Use the camera on your phone to access
The survey behind this QR code