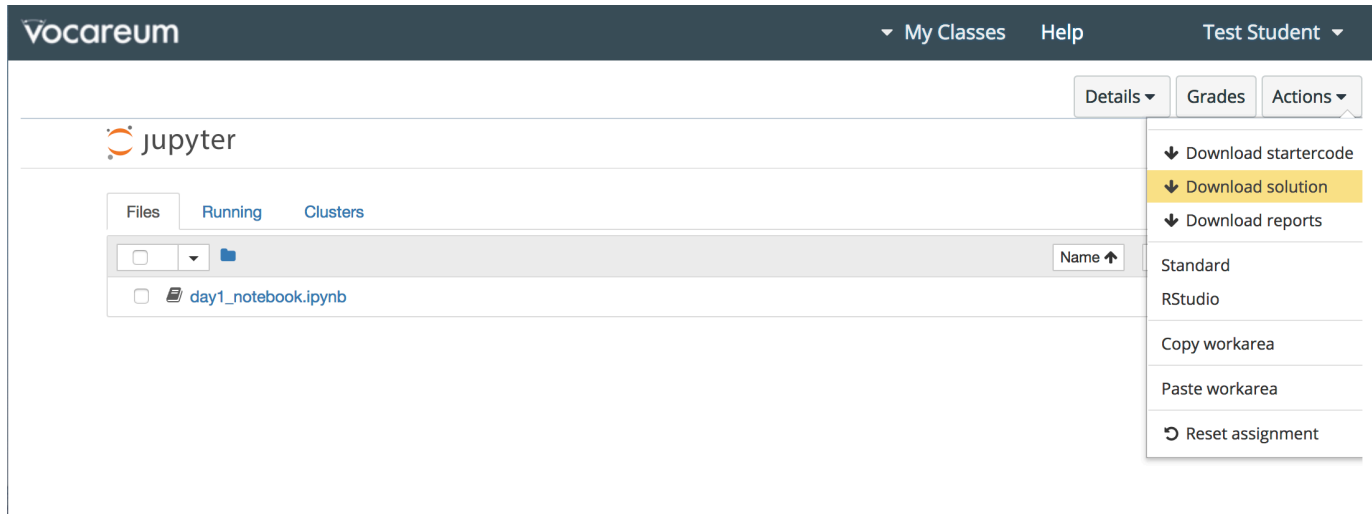


BMI/CS 576 – Day 3

- Today
 - Genome sequencing technology
- Thursday
 - Fragment assembly
 - Graphs
- Day 1 & 2 notebook solutions available via Vocareum

Accessing notebook solutions



The screenshot shows the Vocareum web interface. At the top is a dark navigation bar with the Vocareum logo, a 'My Classes' dropdown, a 'Help' link, and a 'Test Student' dropdown. Below this is a light gray header with 'jupyter' and a 'Details' dropdown. The main content area has tabs for 'Files', 'Running', and 'Clusters'. Under 'Files', there is a file named 'day1_notebook.ipynb'. To the right of the file list is an 'Actions' dropdown menu. The menu is open, showing options: 'Download startercode', 'Download solution' (highlighted in yellow), 'Download reports', 'Standard', 'RStudio', 'Copy workarea', 'Paste workarea', and 'Reset assignment'.

Options for viewing notebook solutions:

1. Upload into your Vocareum sandbox space
2. Install Jupyter on your personal computer (I recommend the [Anaconda distribution](#))

Quiz

- Reading whole genomes is hard because chromosomes are very long!

Muddiest points

- Again, great questions (and answers)
- Thursday's muddiest point will be last required
- If your question was not answered
 - Feel free to pull an instructor aside during the class
 - Ask your neighbors!

Python Exceptions

- Coding style:
 - “Easier to ask for forgiveness than permission” (EAFP)
 - Write code assuming best case (all assumptions met)
 - Deal with errors if they come up
- Alternative to “Look before you leap” (LBYL)
 - Test all assumptions up front

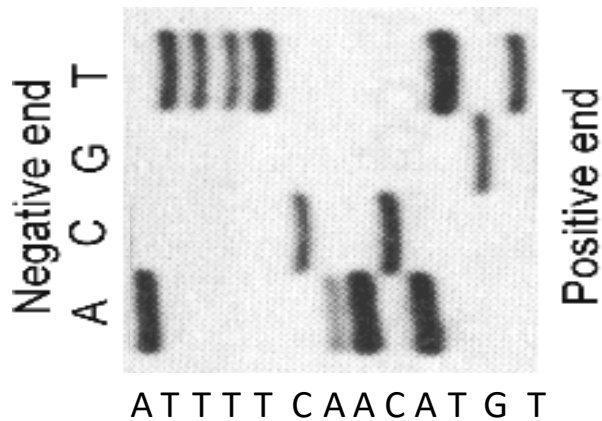
Exception Example

```
def sum_of_numbers_in_file(filename):  
    try:  
        f = open(filename)  
        numbers = [int(line) for line in f]  
    except ValueError:  
        return None  
    else:  
        return sum(numbers)  
    finally:  
        f.close()
```

Comparison of technologies

Technology	Read length	Throughput (# reads/run)	Cost (\$/base)
Illumina	Very short	Enormous	Very cheap
Pacific Biosciences	long	low	expensive
Oxford Nanopore	Very long	low	expensive

Reading a Sanger gel



Sequencing products:

T

TG

TGTA

TGTAC

TGTACA

TGTACAA

TGTACAAC

TGTACAAC T

TGTACAAC TT

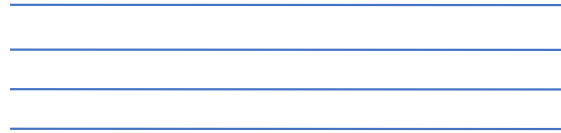
TGTACAAC TTT

TGTACAAC TTTT

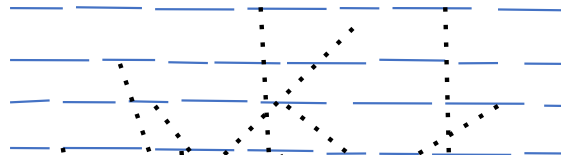
TGTACAAC TTTTA

Shotgun Sequencing Fragment Assembly

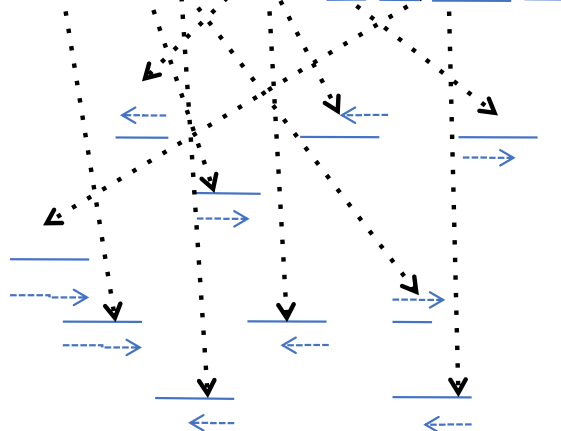
Multiple copies of sample DNA



Randomly fragment DNA



Sequence sample of fragments



Assemble reads

