

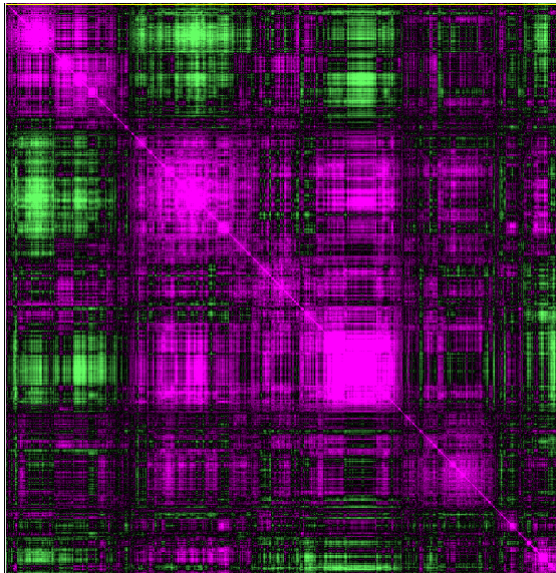
Practical Bioinformatics

Mark Voorhies

5/1/2017

```
dictionary = {"A": "T", "T": "A", "G": "C", "C": "G"}  
dictionary["G"]  
dictionary["N"] = "N"  
dictionary.has_key("C")
```

Clustering exercises – Visualizing the distance matrix



Homework

- 1 Download the mouse transcriptome GTF file linked on the website. This is a gzipped, tab-delimited text file.
- 2 Filter the file for lines with “transcript” in the third column.
- 3 The first column gives the chromosome. transcript_id in the ninth column gives the transcript_id, as used in the count matrices that we’ve been working with. Create a dictionary mapping transcript_ids to chromosomes.
- 4 Filter est_counts.cdt from the course website to remove transcripts on the X and Y chromosomes.
- 5 Bonus: columns 4, 5, and 7 of the GTF give 1-based start, stop, and strand coordinates for the transcripts. Try writing a simple matplotlib-based genome browser for the mouse genome. Try including exon rows to plot detailed gene structures.