

Schedule

Tuesday, September 5th

- Connecting to servers
- Genomics datasets

Thursday, September 7th

- Substitutions and regular expressions
- Introduction to shell scripting

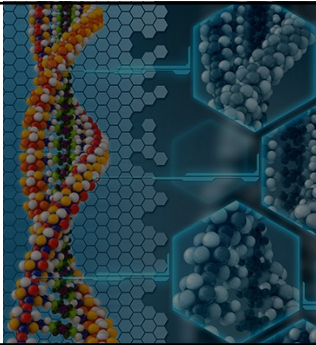
Tuesday, September 12th

- Processing data with awk and sed

Thursday, September 14th

- Conclusions and future directions
- Comprehensive final exam

CSU NSG 58044



Searching and manipulating genomic datasets

Common Linux tools for text searching and processing datasets:

grep...search for patterns within a file and return lines containing the pattern.
usage... \$ grep "pattern" file

tr...deletes or replaces one set of characters in a file with another set of characters.
usage... \$ tr 'ACTG' 'TGAC' file

sed...for more complex substitutions than tr.
usage... \$ sed 's/"old_pattern"/"new_pattern"/g' file

awk...allows for more complex text editing and processing.
usage... \$ awk 'pattern { action }' file
