Practical Introduction

Introduction to Linux for Sequencing Workshop

December 8, 2014

Mary Kate Wing
University of Michigan
Center for Statistical Genetics

Goals of This Session

- No food/drink in this room

- Learn how to logon to the workshop Linux machine
- Become familiar with basic Linux commands that will be used during the workshop

- Ask questions at any point

Logon to Windows

- In front of you are Windows machines
  - Username/password on right-hand monitor
- Will connect to the workshop Linux machine from Windows
  - Tools to connect already installed

Everyone logged in?

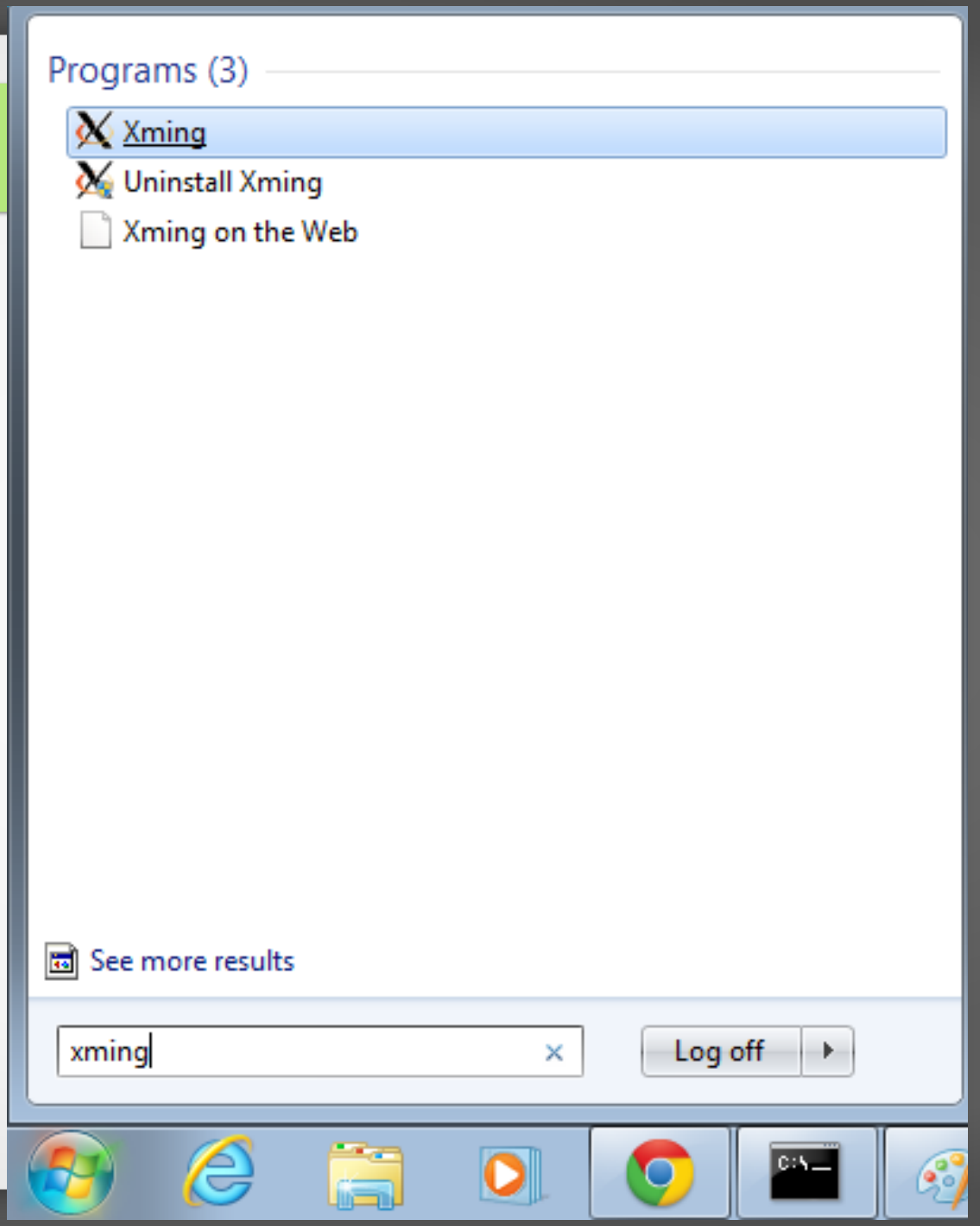
Connecting to Workshop Machine

- PuTTY - connect to Linux machine
  - terminal emulator
  - run commands
- xming - X11 display server
  - Allows you to open pdfs, etc from putty
- Instructions at start of each practical
  - So don’t worry about memorizing it

Start xming

1. Start->search
2. Type: xmingSelect : Xming

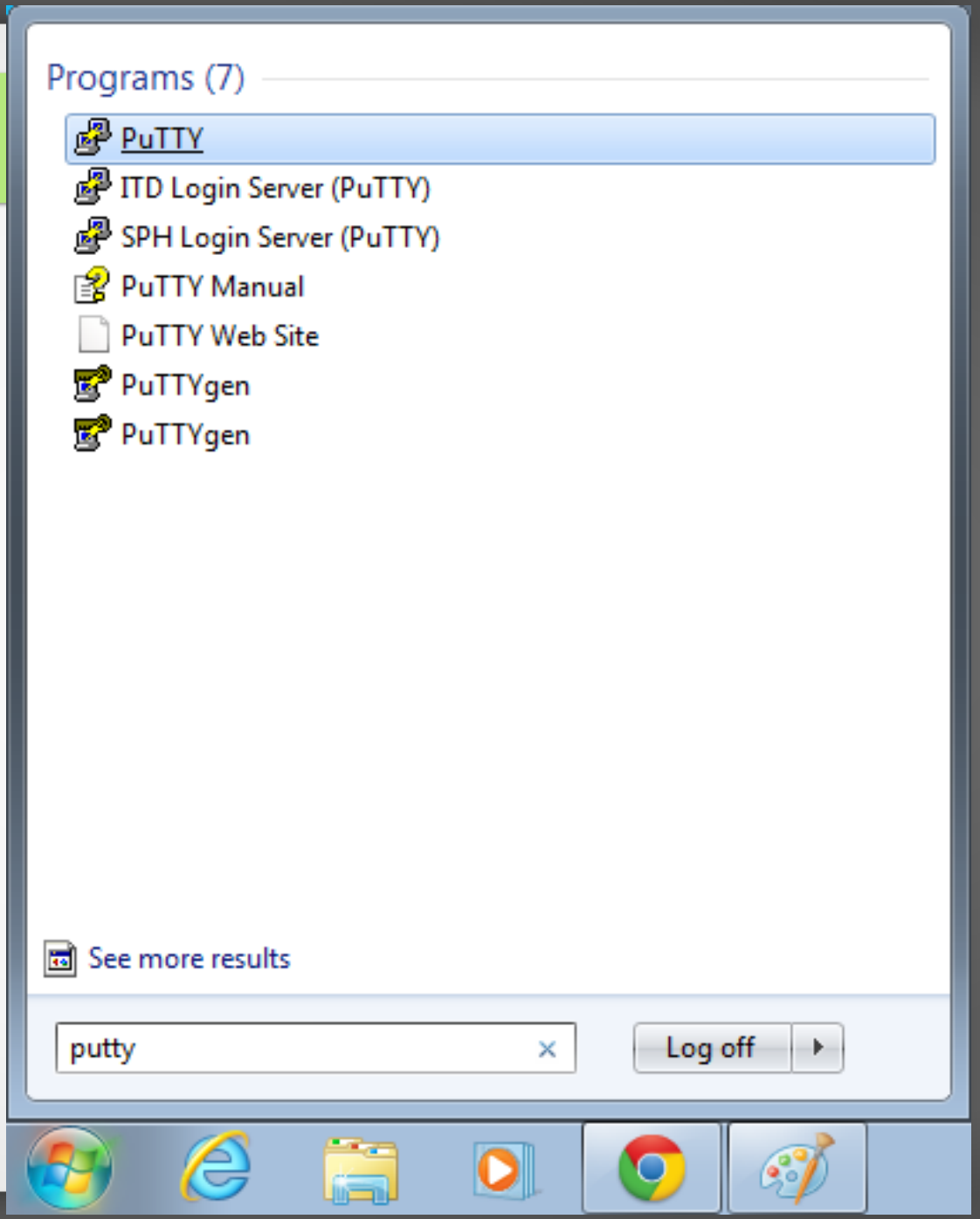
Nothing will happen, but it is started



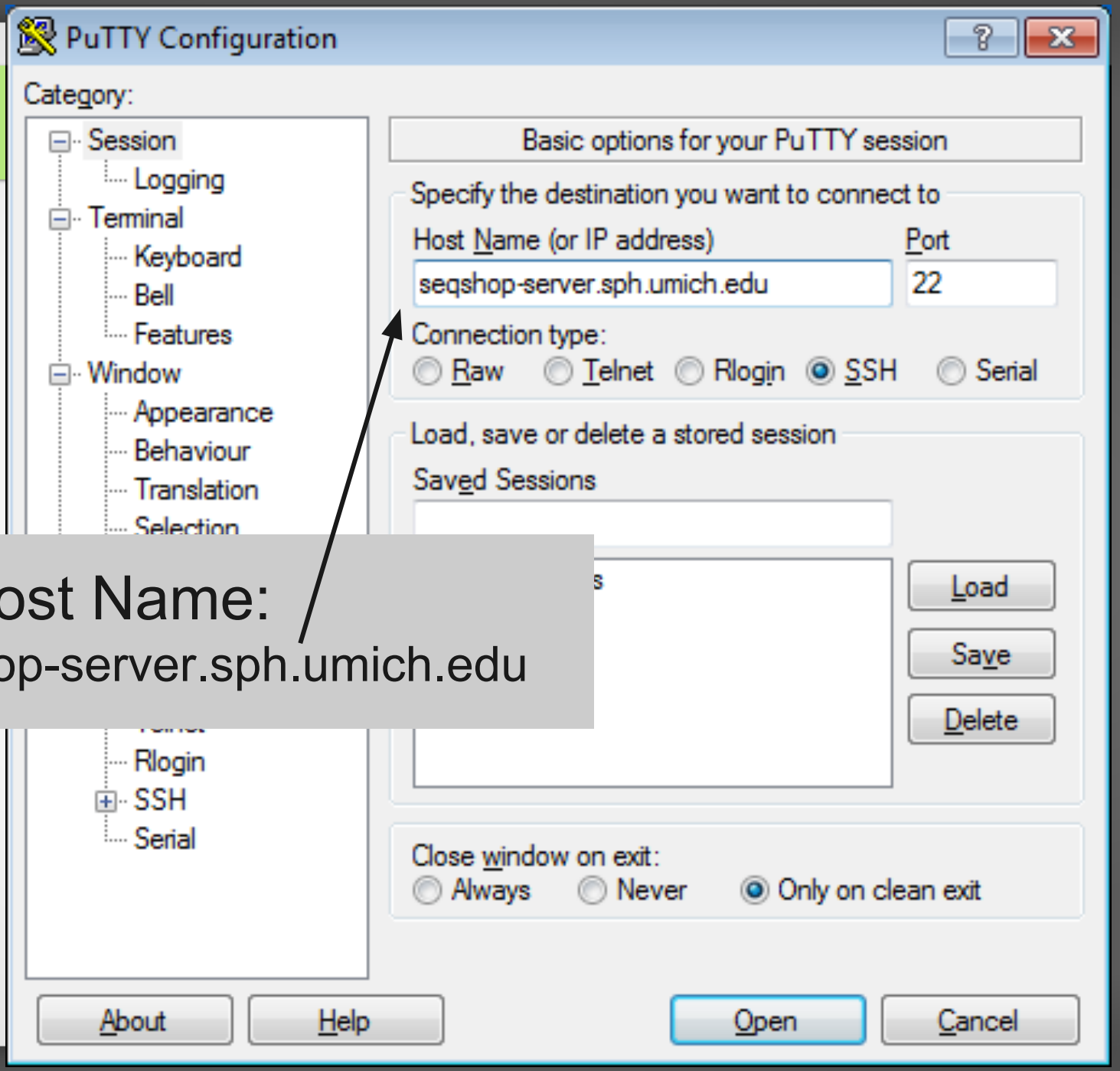
Open PuTTY

1. Start->search
2. Type: putty
3. Select : PuTTY

This will open a PuTTY Configuration window

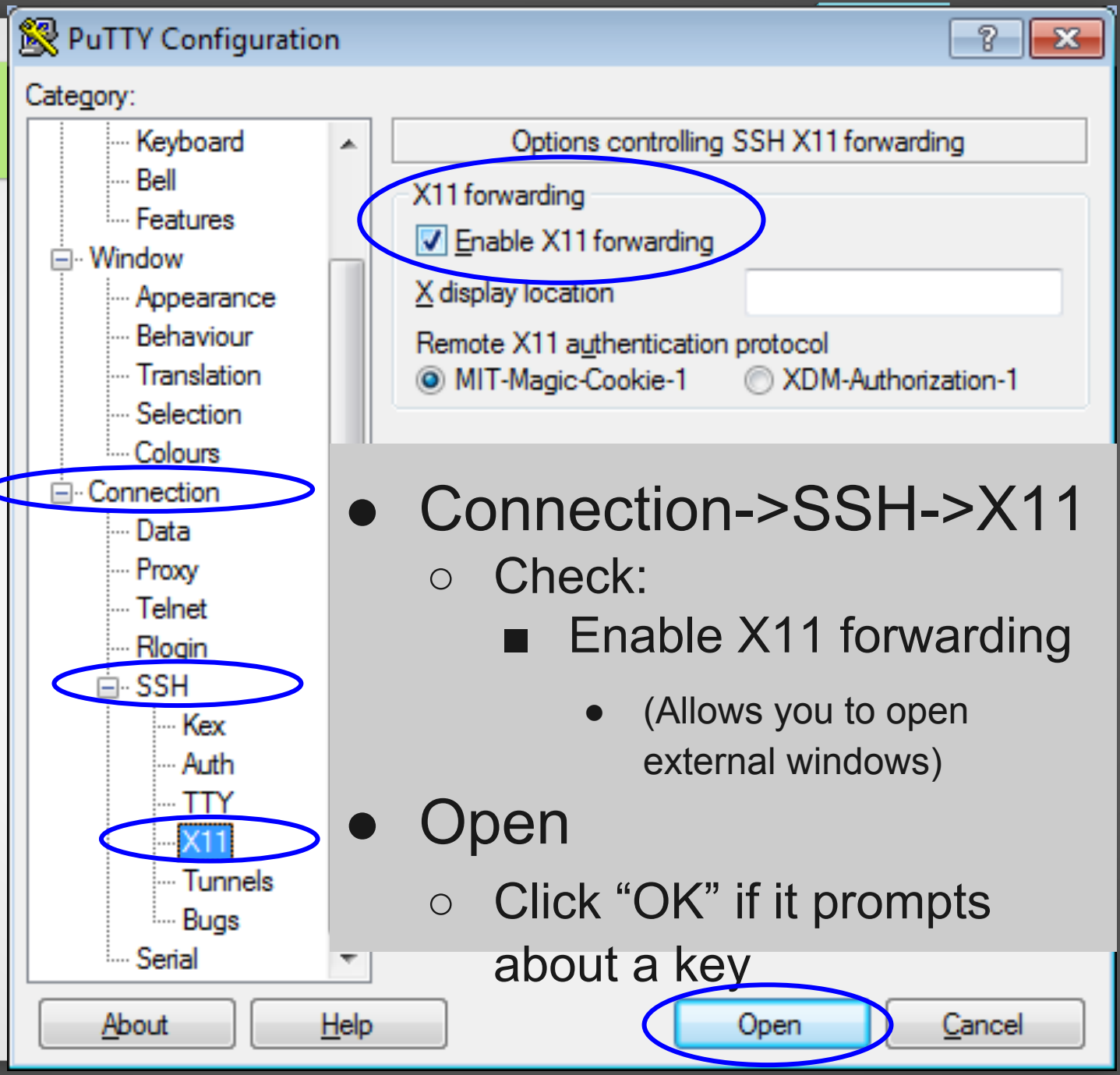


PuTTY



- Enter Host Name:
  - seqshop-server.sph.umich.edu

PuTTY



- Connection->SSH->X11
  - Check:
    - Enable X11 forwarding
      - (Allows you to open external windows)
- Open
  - Click “OK” if it prompts about a key

Logon to seqshop-server

- Enter your provided Linux username and password
- You should now have a terminal on the seqshop-server Linux machine
  - If you ever want/need a 2nd terminal, redo the PuTTY steps
- Does everyone have a terminal?

Terminal Basics

- Command-line prompt
  - You will start in your “home” directory
    - /net/seqshop-server/home/*mktrost*
      - (your username instead)
    - ~
  - Each have your own
    - You have permission to read/write
    - Your inputs & outputs go here

- How to figure out where you are:
  - `pwd`
    - /net/seqshop-server/home/*mktrost*

Directory Contents

- `ls`
  - Lists files/sub-directories
  - What do you have in your directory?
    - examples.desktop
    - Personal Genome directory: SampleXX/NA12878

Up-arrow will go to the previous command

Directory/File Permissions

- `ls -l`
  - ‘d’ - directory; ‘-’ - file
  - r/w/x: read, write, or execute
  - 3 sets of values
    - user/group/everyone
      - USER & GROUP are both set to you
  - Personal Data is read only by you

Looking at a Different Directory

- Enter the path of the directory you want to look at
  - `ls ~mktrost/seqshop/`
  - `ls Sample*`
    - * is wildcard
- Does not change you to a different directory
  - Still in same place you were before
  - Check with `pwd`

Change/Move to New Directory

- `cd Sample*`
  - (or `cd NA12878`)
  - `ls`
    - We will look at this more later

- `cd`
  - takes you back to your home directory

Create New Directory

- `mkdir testDir`
- `mkdir -p testDir/dir1/dir2`
  - Create parent directories if they don't exist
  - Do not generate error if directory exists
  - `ls testDir`
  - `ls testDir/dir1`

- Rename/Move
  - `mv testDir testDir1`
  - `ls`

Up a Directory

- . - current directory
- .. - up 1 directory
  - `ls ../..`
    - Lists contents of 2 directories up
  - `cd ..`
    - Move up 1 directory
  - `cd`
    - Move to home directory

Variables

- Save from typing long path/command
  - In BASH (the shell you are running):
    - export VARNAME=value
      - Allows you to use ${VARNAME} as shorthand
      - Can easily change the value without changing commands
        - Useful for scripts/seqshop copy/paste commands
  - Try it
    - `ls $TEST`
      - ${TEST} is not set, so is blank, so just lists current directory
    - `export TEST=~/testDir1`
    - `ls ${TEST}`
      - Should see contents of testDir1

Look at a file

- `less`
  - Use arrow (up/down/left/right) keys to scroll through file
  - Use *space bar* to jump down a page
  - -S - prevents line wrap
  - Use ‘q’ to exit
  - zless - read compressed
- `cat`
  - `zcat`
- `more`

Look at parts of a file

- `head` - look at start of the file
  - `-n K` - print first K lines (default 10)
  - `-n -K` - print all but the last K lines
- `tail` - look at end of the file
  - `-n K` - print last K lines (default 10)
  - `-n +K` - print starting with the Kth line

Extracting Specific Columns

- `cut` - remove sections from each line of files
  - `-d DELIM` - use DELIM as delimiter rather than TAB
  - `-f N1,N2` - print fields N1 & N2 (and lines with no delimiter)

Additional commands

- Search
  - `grep` - print lines matching a pattern
    - `-v` : lines without pattern
- Word Count
  - `wc`
    - `-l` : line count

String commands together

- `|` - pipe - use output of previous command as input to next command
  - `tail -n+300 file |head -n 5`
    - Output starting at line 300, but only read 5 lines

String commands together

- `zcat ${OUT}/split/chr22/chr22.filtered.PASS.vcf.gz |grep -v "^#"| cut -f 7| grep -v "PASS"`
  - `zcat` ...: uncompress the zipped VCF
  - '`|`' : takes the output of one command and sends it as input to the next
  - `grep -v "^#"` : exclude any lines that start with "#"
  - `cut -f 7` : extract the 7th column
  - `grep -v "PASS"` : exclude any rows that have a "PASS" in the 7th column

Screen

- Leave commands running after you logout
  - Start Screen: `screen`
  - Detach from screen session: `Ctrl-a d`
  - Logout as normal
- Resume session later
  - login
  - Resume: `screen -r`
- Exit
  - `exit`
- Scroll
  - Scroll: `Ctrl-a Esc`
  - Exit Scroll: `Esc`

Run Workshop Later

- Instructions/tutorial data will be on wiki by end of next week
  - `wget http://www.sph.umich.edu/csg/mktrost/seqshopExample.tar.gz`
  - `tar xvf seqshopExample.tar.gz`
- Build C++
  - `make`

Useful Links

- Helpful reference to many tools:
  - http://infoplatter.wordpress.com/2014/04/06/bioinformaticians-pocket-reference/
    - links to "cheat-sheets", including, Unix, screen, and vi
- Our wiki with some brief description of how to do some basic commands
  - http://genome.sph.umich.edu/wiki/Basic_Linux_Intro
- Commands to use screen
  - http://genome.sph.umich.edu/wiki/Screen_Commands