3rd Annual Workshop
Practical Genomics: From Biology to Biostatistics

August 26-28, 2013
Room 202 Kennedy Krieger Institute
801 N. Broadway, Baltimore MD 21205

Coordinator: Lauren Ciotti (lschenk2@jhmi.edu; 443-287-7341)
Parking: Ashland Street Garage (enter on Ashland off of North Broadway) (coupons available)

Before the workshop starts		
We will use software called VirtualBox to create a Linux environment, complete with software programs and data that we'll use during the workshop. We are emailing all participants instructions on how to install VirtualBox on your PC or Mac laptop. Here's a checklist of things "to do" before the workshop.		
	1. Install VirtualBox and Ubuntu	
	2. Connect to your virtual Ubuntu server using SSH	
	3. Install IGV (registration required)	
	4. Install R and RStudio	
	5. Install MEGA (registration required)	
All these are free and generally easy to install. All are very useful. **Allow at least ***TWO HOURS*** to download and complete the full installation. Please arrive with everything in working order.** If you have any trouble, or if you need any help, we've set aside a little bit of time during the workshop to help you finish your installations. If you have any questions before the workshop, email Frederick Tan (frederick.j.tan@gmail.com) and the workshop team will try to help.		

Monday, August 26th, 2013	
8:30-9:00	Breakfast (provided); set up VirtualBox and connect via SSH
9:00-10:00	Introduction to high throughput biology
10:00-10:30	Linux: Shell prompt, directory structure, command line interfaces, finding help, navigation, organizing files
10:30-10:45	Break
10:45-11:55	Quality control for sequencing data
12-12:45	Lunch (provided). Go to room 510 of the 707 Building. Directions***
12:45-2:00	Experimental design considerations: control samples in genetic analyses
2:00-2:30	Break
2:30-3:00	Linux: Viewing file contents, grep, pipes, paths, nano
3:00-3:30	Data analysis: FastQ format, quality control with FastQC
3:30-5:30	Data analysis: Using Python to parse alignment output

5:30-6:30	Scientific networking and informal meetings with Hopkins core directors	
6:30-9	Dinner: on campus at the Daily Grind. Directions: Enter the Miller Research Building (on N. Broadway, across the street (south) of the Kennedy Krieger 801 Building. Walk past the guard station. Go straight to the first set of elevators. Take them to 3. Proceed left/straight about 50 paces.	

*** Exit the 801 Building past the guard station. Cross Madison and go to the main Kennedy Krieger building at 707 N. Broadway (at the intersection of Monument). Take the main bank of elevators to 5; room 510 is immediately to the right. We will use this room ONLY from noon-2 on Monday afternoon.

Tuesday, August 27th, 2013	
8:30-9:00	Breakfast (provided); set up R and RStudio
9:00-10:00	Variation in high dimensional sequence datasets: biological considerations. Kasper's paper.
10:00-10:30	Break
10:30-12:00	Gene sets
12:00-12:30	Lunch (provided)
12:30-1:30	Reproducible research in high throughput biology
1:30-2:20	Hands-on programming: alignment, samtools view/sort/index, idxstats, depth, mpileup
2:20-3:30	Hands-on programming: variant calling with samtools/vcftools, and comparison with IGV
3:30-4:00	Break
4:00-5:30	Using R for gene set analysis

Wednesday, August 28th, 2013	
8:30-9:00	Breakfast (provided); set up MEGA
9:00-9:30	Overview of amplicon sequencing for metagenomics using rDNA (16S)
9:30-10:45	Multiple sequence alignment and phylogeny
10:45-11:00	Break
11:00-12:15	Phylogeny and clustering methods: statistics
12:15-12:45	Lunch (provided)
12:45-1:45	Hands-on programming: align amplicon data, mafft
1:45-3:30	Hands-on programming: phylogeny, MEGA
3:30	Workshop ends