Optional: viewing mapped reads

1) Bam files are binary and cannot be read with less, head, etc. Use samtools view to investigate a bam file:
   a. What are options?   samtools view --help
   b. Header only   samtools view -H <sample.bam>
   c. First few rows   samtools view < sample.bam> | head   #note this does not show the header!
   d. Scroll, incl header   samtools view -h < sample.bam> | less

2) Bam files are very large, so we will subset the bam file to a part of one chromosome for visualization with IGV and/or as a custom track on the UCSC genome browser.

3) First, a bam file needs to be indexed – run this program in the same folder that contains the bam file of interest:
        samtools index <sample.bam>
   When it completes, there should be a new file called sample.bam.bai.

4) You may want to create a new directory for the for the gene- or region-specific bam files, such as
        mkdir /workdir/netid/bam_GOI
   and cd to this directory.

5) Subset bam files (for ease of handling in the workshop). To compare multiple samples, run samtools view for each sample of interest. To include any sample in the workshop, use bam files in /workdir/shared_data/STAR.
   a. Option 1: HRAS   samtools view -b </path-to/sample.bam> 11:530000-537000 > sample_HRAS.bam &
   b. Option 2: ACTB   samtools view -b </path-to/sample.bam> 7:5526000-5532000 > sample_ACTB.bam &
   c. Option 3: MFG   find the genomic interval for MyFavoriteGene, modify examples above
   You can count the number of lines in the file (number of mapped reads – note may not be unique read count!) directly with samtools:
        samtools view -c <sample_GOI.bam>
   or with wc -l:
        samtools view <sample_GOI.bam> | wc -l.
   You can also generate a .sam file without '-b', which does not require samtools view to read.

6) Index each GOI bam file, as in step 3.

7) To view in the UCSC genome browser, first create a bedgraph file using bedtools genomecov:
        bedtools genomecov -bg -trackopts 'name="<sample>"' -split -ibam <sample_GOI.bam> > sample_GOI.bg &
   You will need to edit the chromosome name for the UCSC browser to recognize the coordinates, and it can help to compress the file with gzip:
        awk '{if(NR>1) print "chr"$0; else print $0}' <sample_GOI.bg> | gzip > sample_GOI_chrName.bg.gz

8) To view reads in IGV, you will need to transfer the sample_GOI.bam and sample_GOI.bam.bai (index) files to your laptop with Filezilla.
   For the workshop, we will use the IGV web app, but you can also download/install IGV to your computer from https://software.broadinstitute.org/software/igv/download.
   View in IGV (web app):
         a. Open the app in a browser on your laptop: https://igv.org/app/
         b. In the top drop-down menu, change Genome to Human (GRCh38/hg38)
         c. In the top drop-down menu, select Tracks: Local File and point to your sample_GOI.bam and sample_GOI.bam.bai (index) files.
         d. Jump to the gene of interest using the gene symbol or genomic coordinates.
         e. Right-click for lots of options! Change visualization: color, sort, etc

9) To view coverage in the UCSC genome browser, you will need to transfer the sample_GOI_chrName.bg.gz (gzipped begraph) files to your laptop with Filezilla.
   View in UCSC genome browser:
   a. Open the genome browser web site: https://genome.ucsc.edu/cgi-bin/hgGateway
   b. Select the Human genome (wg38) and click GO
   c. Below the browser view, click the middle button 'Add custom tracks'
   d. Above the top entry window, click Browse: and find the bedgraph file sample_GOI_chrName.bg.gz, and Upload
   e. Add more tracks (if desired) – must upload one bedgraph file at a time.
   f. When uploaded, click Go (after Add to genome browser…GO)
   g. Search for gene symbol or enter genome coordinates for GOI.
   h. Modify tracks, visualization with right-click menu to the left of each track or in the table below.