File Types to research and present:

* Fasta/fastq (sanger and illumina)/fai
* GFF/GTF
* Sam/Bam/Bai
* Vcf/bcf/Pileup/interval/ROD
* Bed/bigBed
* Wig/bigWig

Questions you could answer in your presentation (5 – 7 min)

* What does each line in the file represent?\*
* What does each column represent?\*
* What character separates the columns? Can you write this character?\*
* Is there a file header?\*
* What online repositories are available to download this data and how can I download it?\*
* What type of user/database/online tools uses this file format/who came up with it?
* Show an example of the file.
* What is the difference between the files you’ve been given?
* Can you convert between one file format and another?
* Any other information interesting information?

\* Questions you should ask about any file format you have been given.

Unsure about jargon? Ask for help!

Useful links:

VCF

<http://www.1000genomes.org/wiki/analysis/variant-call-format/vcf-variant-call-format-version-42>

GFF

<http://www.ensembl.org/info/website/upload/gff.html>

<http://www.sanger.ac.uk/resources/software/gff/spec.html>

BED/BIGBED/WIG/BIGWIG

<http://genome.ucsc.edu/FAQ/FAQformat.html#format1>

<http://www.ensembl.org/info/website/upload/bed.html>

<http://www.ensembl.org/info/website/upload/wig.html>

SAM/BAM/BAI

<http://genome.sph.umich.edu/wiki/SAM>

<http://samtools.sourceforge.net/SAMv1.pdf>

<http://blog.nextgenetics.net/?e=18> = bitwise flags

FASTA/FASTQ

http://en.wikipedia.org/wiki/FASTQ\_format#Illumina\_sequence\_identifiers

<http://bioinf.comav.upv.es/courses/sequence_analysis/sequence_file_formats.html>

<http://blog.nextgenetics.net/?e=33> = phred quality

All files!

<http://www.broadinstitute.org/igv/FileFormats>

http://genome.ucsc.edu/FAQ/FAQformat

NEXT STEPS –

Look at the databases mentioned in the presentations and go through them in a live demo. Typical will be Ensembl, UCSC, NCBI, GATK, 1000 genomes

SAM very important – re-enforce QUAL score, CIGAR string, FLAG

Quality scores – make reference to illumina vs sanger (Sanger pretty standard now)