NGS PRACTICALS

NEXT GENERATION SEQUENCE DATA ANALYSIS USING GALAXY

We are going to analyze NGS data, annotate and interpret SNPs from pig IGF1R gene.

Files available:

* Ind1.fq.gz: small example of reads from one lane of Illumina
* IGF1R.h.bam: bam file containing assembly of one Iberian pig in IGF1R region
* IGF1R.IB.vcf: vcf file containing SNPs of an Iberian pig in IGF1R region
* IGF1R.LW-WB.vcf: vcf file containing SNPs of Large White and wild boars in IGF1R region

The files are available at: Files

1. Get familiar with formats: fasta, fastq
2. Get familiar with galaxy (https://usegalaxy.org/): open an account, browse the wikis.
3. Upload files available
4. Check quality (fastqc option) of Ind1.fq.gz & IGF1R.h.bam files
5. Get familiar with genome browsers:

UCSC browser: <https://genome.ucsc.edu/cgi-bin/hgGateway>

IGV: <http://software.broadinstitute.org/software/igv/>

1. Visualize IGF1R.bam with IGV within galaxy.
2. Get familiar with vcf format, compare with plink format.
3. Identify SNPs from vcf into bam file with IGV. Check relevance of quality.
4. Annotate SNPs with variant effect predictor (use file IGF1R.LW-WB.vcf):

<http://www.ensembl.org/info/docs/tools/vep/index.html>

1. Get familiar with 1000 genome project:

<http://www.internationalgenome.org/>