Using 1000 Genomes Tutorial Exercises 12th April 2012

Finding and Handling Data

1. How many Omni VCF files can you find on the ftp site (Omni is a high throughput genotyping platform from Illumina on which all 1000 genomes samples are being genotyped)

2. Find the most recent Omni VCF file on GRCh37 from the 31st January 2012

3. Use the Website search box found in the top right hand corner of all pages to find the FAQ question about getting subsections of VCF files.

4. Which exome sample from 20110521 has the highest percentage of targets covered at 20x or greater

5. Find the exome bam file for this sample

6. Get a slice of this exome bam file between 7:114173990-114175942 (exon of FOXP2)

Command Line Tools

7. Get the 7:114304000-114305000 (FoxP2 exon) section of the 20110521 release chr 7 genotypes file

8. Use vcftools vcf-stats to specify which SNP transition happens most in this section

9. Use this piece with tools, the variant effect predictor, the vcf pattern finder

10. Are there any snps with deleterious sift/polyphen consequences?

11. What is the most common pattern of variation in this region?

12. Use the vcf to ped script with 6:31830700-31840700

13. How many different haplotype blocks does the section contain?