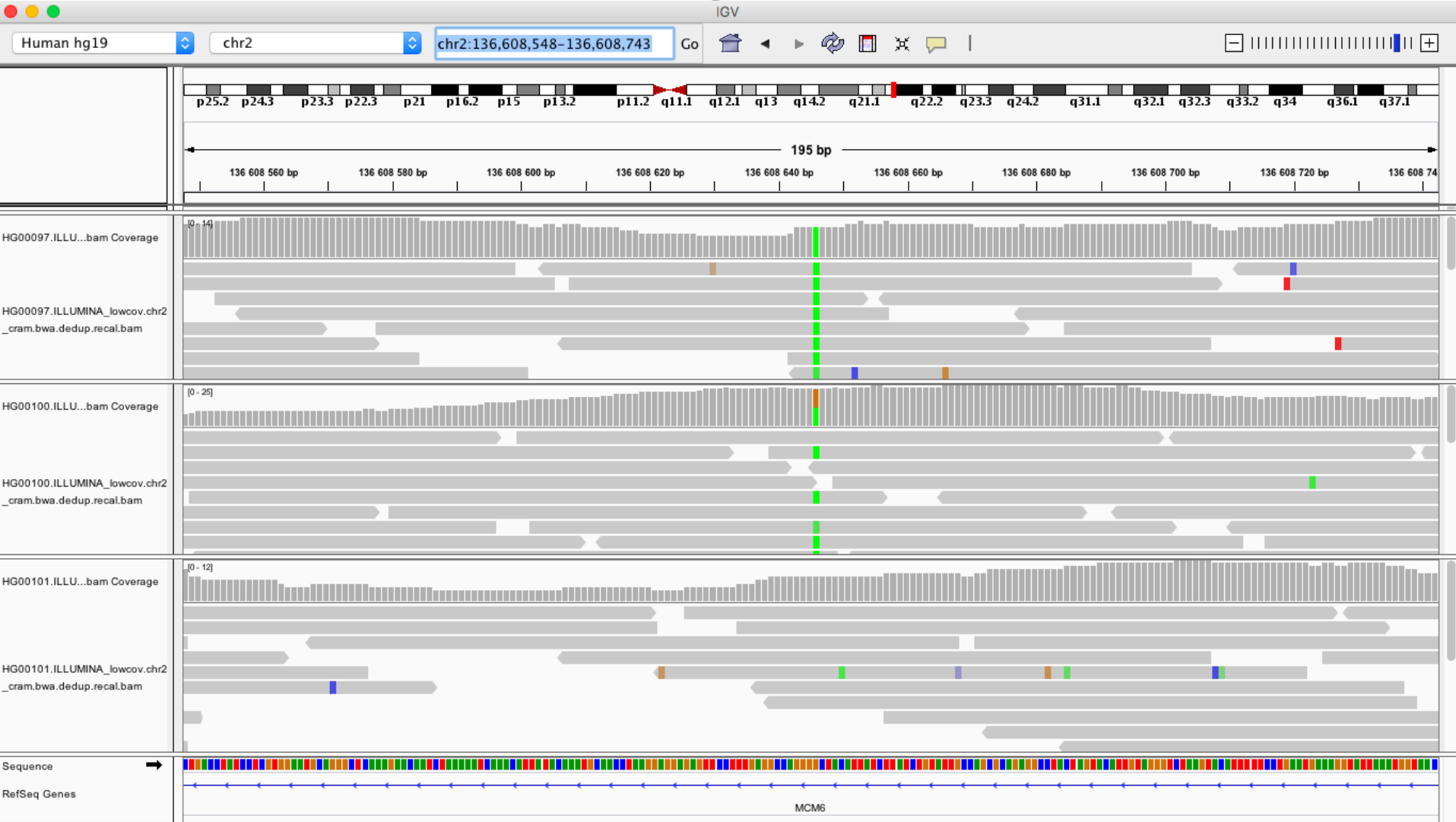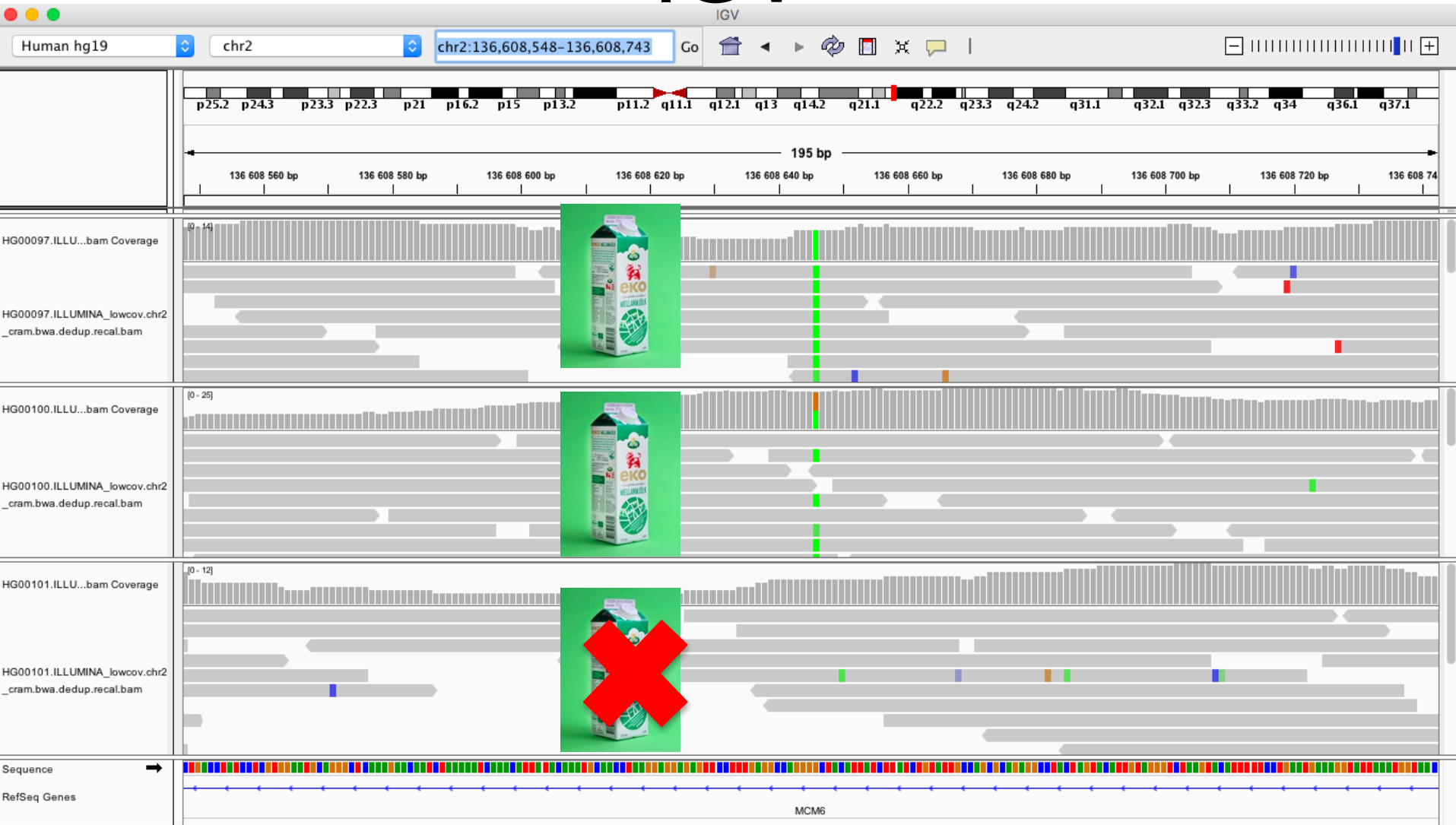# NGS workflow

Lab discussion

IGV

Read length?
Do we have decent coverage across the variant position?
Homozygotes/heterozygotes?
Who can drink milk?

IGV

Vcf files

Descriptions of each field can be found in the header section (lines starting with '#')

CHROM POS   ID   REF   ALT   QUAL   FILTER INFO   FORMAT   HG00097 HG00100 HG00101

2   136608646   .   G   A   628.02 PASS AC=3;AF=0.500;AN=6;BaseQRankSum=-3.410e-01;ClippingRankSum=0.00;DP=38;ExcessHet=1.5490;FS=3.211;MLEAC=3;MLEAF=0.500;MQ=60.00;MQRankSum=0.00;QD=20.26;ReadPosRankSum=-3.970e-01;SOR=1.037;set=snp

GT:AD:DP:GQ:PL 1/1:0,9:9:27:316,27,0   0/1:11,11:22:99:348,0,382 0/0:7,0:7:21:0,21,223

UCSC

- Most species have the G allele, however squirrels and naked mole rat apparently have the A allele
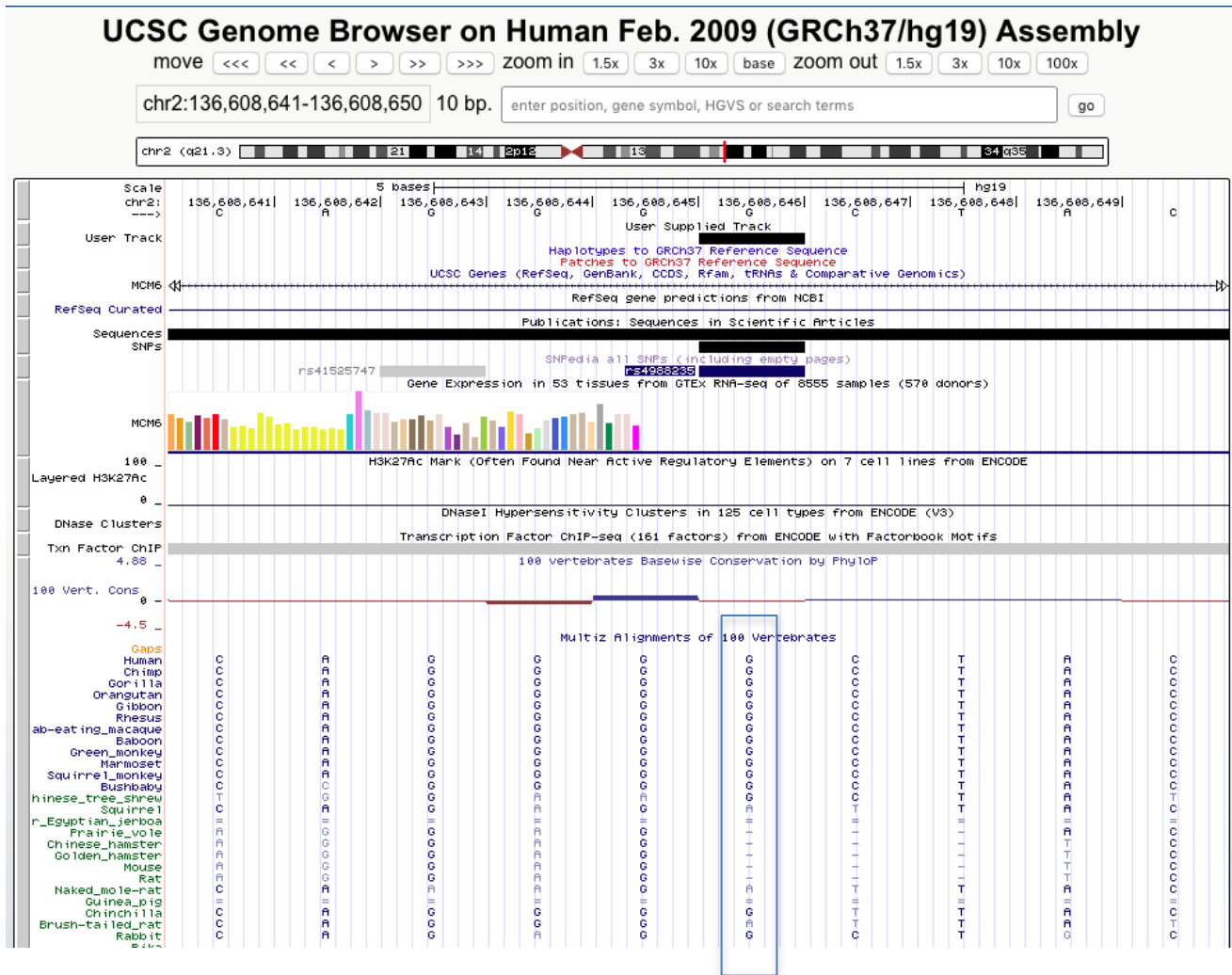
Questions

- MarkDuplicates: 588 duplicate reads
- 697 of the 711 variants passed the filters
- 108 variants are indels (use: --keep-only-indels)

Warnings

- Warnings are common, google them if you are not sure what they mean
  - JointGenotyping: Some filters require heterozygote variants
  - You probably got a warning that optical duplicates were not discovered. This is because the read names are not properly formatted to be able to do this
  - Annovar: "1 invalid alternative alleles found in input file" (it is a '*', which is a symbol used by JointGenotyping when some samples have a SNP and others have an indel covering that site.