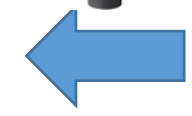
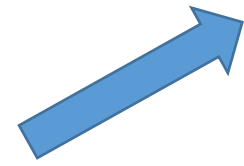
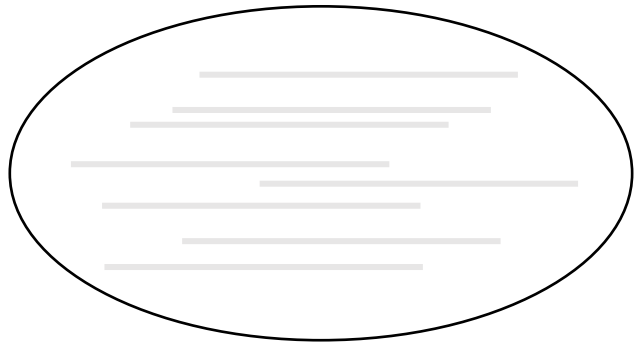


SNP filtering



hard filter



multiple samples

	Sample 1	Sample 2
Pos 1	AA	AT
Pos 3	AT	TT

	Sample 1	Sample 2
Pos 1	AA	AT
Pos 2	TT	AA
Pos 3	AT	TT

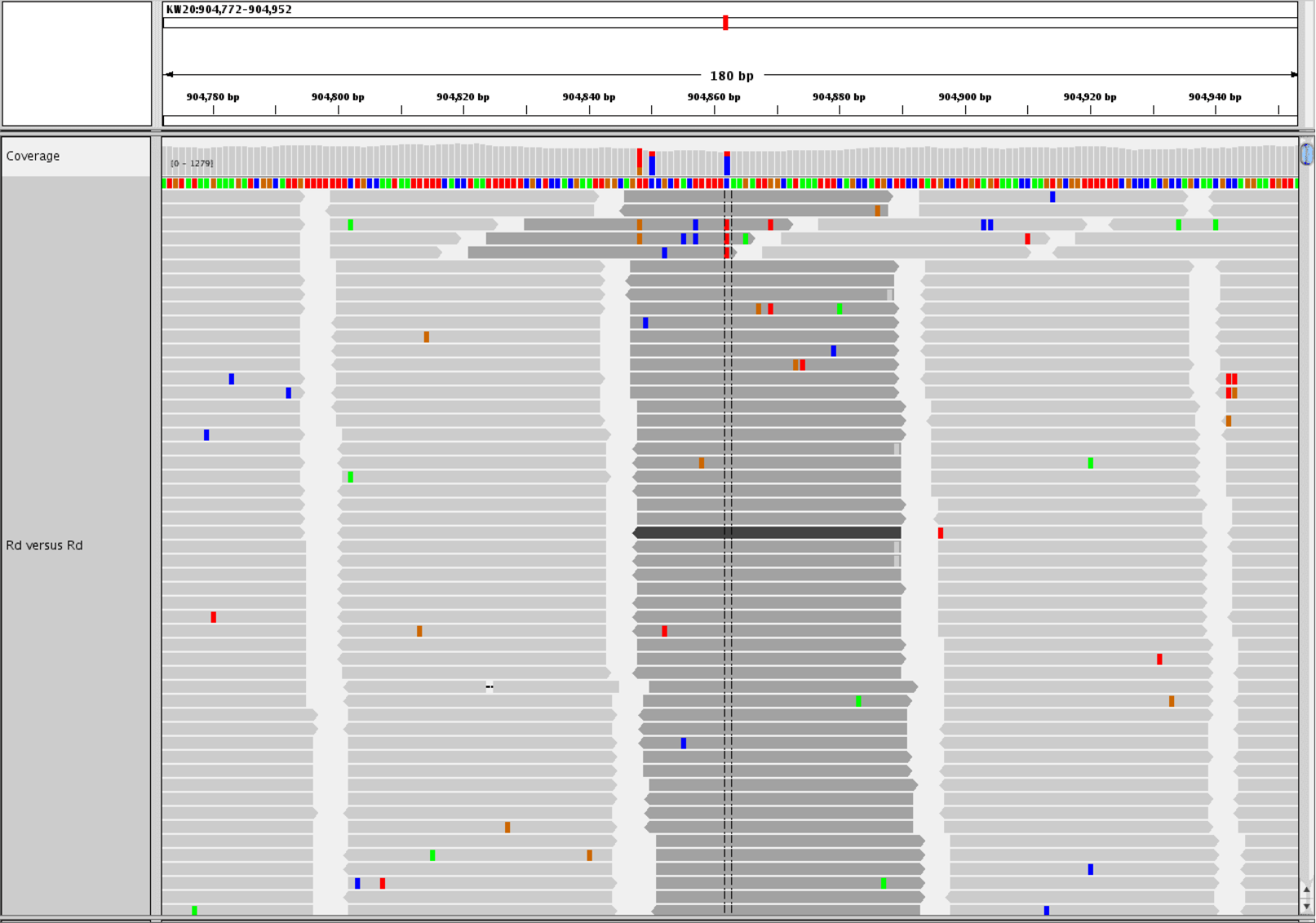


Tools for filtering

- Vcftools
- Vcflib
- Bcftools
- Rvcf

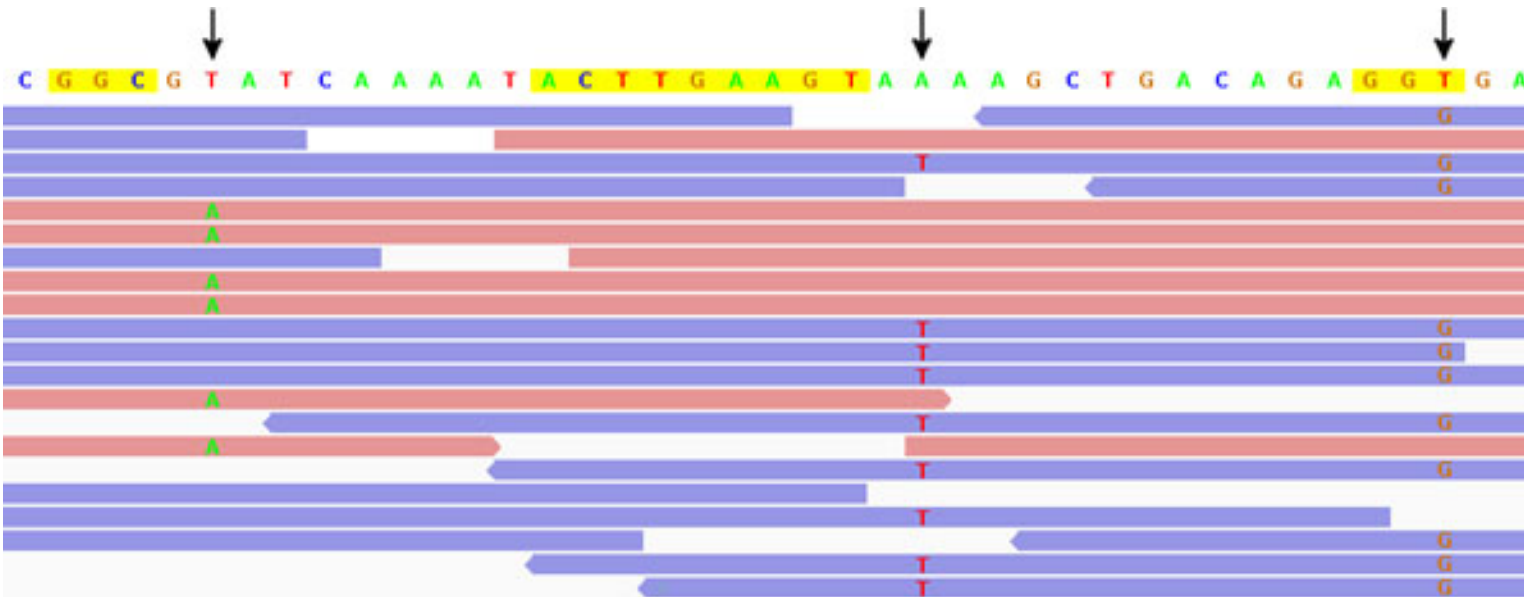


Sequencing errors





Context dependant errors



Rank	Context	FER [%]	RER [%]	ERD [%]
1	ACGGCGGT	26.1	0.5	25.6
2	GTGGCGGT	25.1	0.7	24.4
3	GCGGCGGT	22.9	0.7	22.2
4	GTGGCTGT	22.4	0.6	21.8
5	ATGGCGGT	21.2	1.0	20.3
6	NCGGCGGT	20.0	0.7	19.3
7	GTGGCTTG	20.2	1.2	19.0
8	GNGGCGGT	19.2	0.7	18.5
9	GCGGCTGT	18.8	0.7	18.1
10	ACGGCTGT	18.6	0.8	17.7

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3622629/>



Filter Criteria

- low quality of mapping (and bases)
- read depth (coverage)
- missing sites
- remove samples with a lot of missign sites
- "quality / coverage"(higher coverage sites should have higher quality)
- keep SNPs only(= no indels)
- keep biallelic SNPs only(= no multiallelic SNPs)
- alleles that are found only on one strand (or reads)
- remove SNPs with large discrepancy between reference and alternative allele
- filter for minimum minor allele frequency (MAF), e.g. 5%
- filter for quality ratio of forward and reverse reads

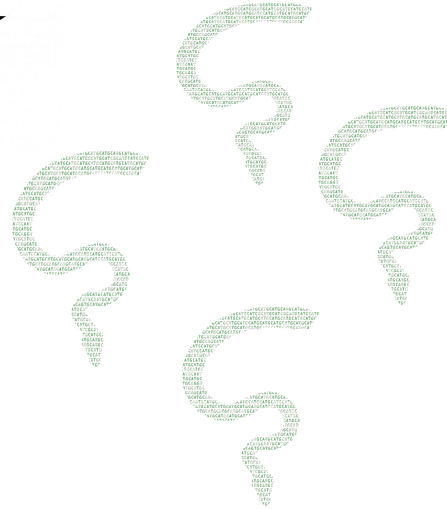
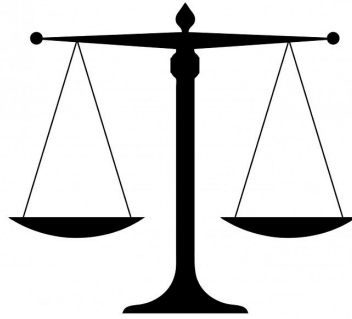
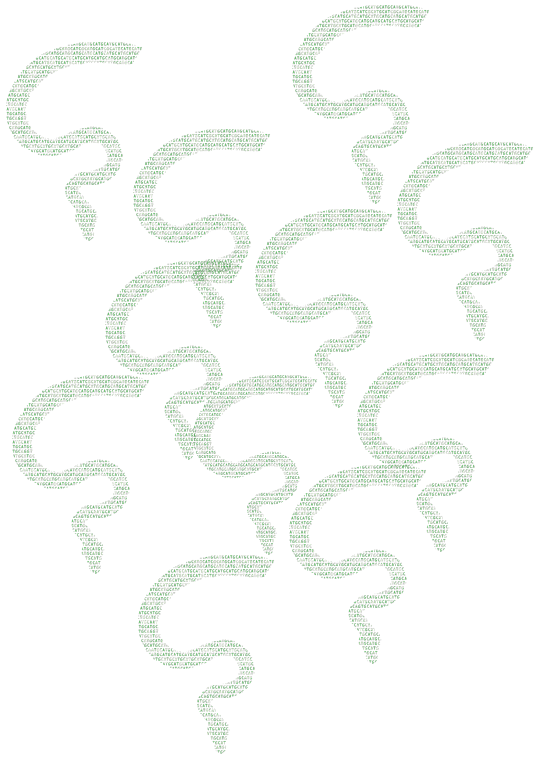


Genotype likelihoods

```
scaffold3/99|ref0000050|ref0000027 285 . A G 539.871 .
AB=0;ABP=0;AC=12;AF=0.146341;AN=82;AO=83;CIGAR=1X;DP=4295;DPB=42
95;DPRA=0.359959;EPP=28.1523;EPPR=23.8636;GTI=6;LEN=1;MEANALT=1.1
6667;MQM=30.4096;MQMR=59.131;NS=41;NUMALT=1;ODDS=2.47263;PAIR
ED=1;PAIREDR=0.941051;PAO=0;PQA=0;PQR=0;PRO=0;QA=3137;QR=150052
;RO=4207;RPL=2;RPP=166.289;RPPR=13.8625;RPR=81;RUN=1;SAF=59;SAP=3
5.0591;SAR=24;SRF=2074;SRP=4.80704;SRR=2133;TYPE=snp;technology.Illu
mina=1 GT:DP:DPR:RO:QR:AO:QA:GL 0/0:151:151,0:151:5286:0:0:0,-
45.4555,-473.625 0/0:43:43,1:42:1397:1:34:0,-10.3232,-123.369
0/0:53:53,1:52:1826:1:34:0,-12.5878,-160.796 1/1:20:20,20:0:0:20:807:-
55.4598,-6.0206,0 0/1:61:61,10:51:1695:10:354:-8.41372,0,-131.988
```



Ultra low coverage sequencing



Angsd (<http://www.popgen.dk/angsd/index.php/ANGSD>)

