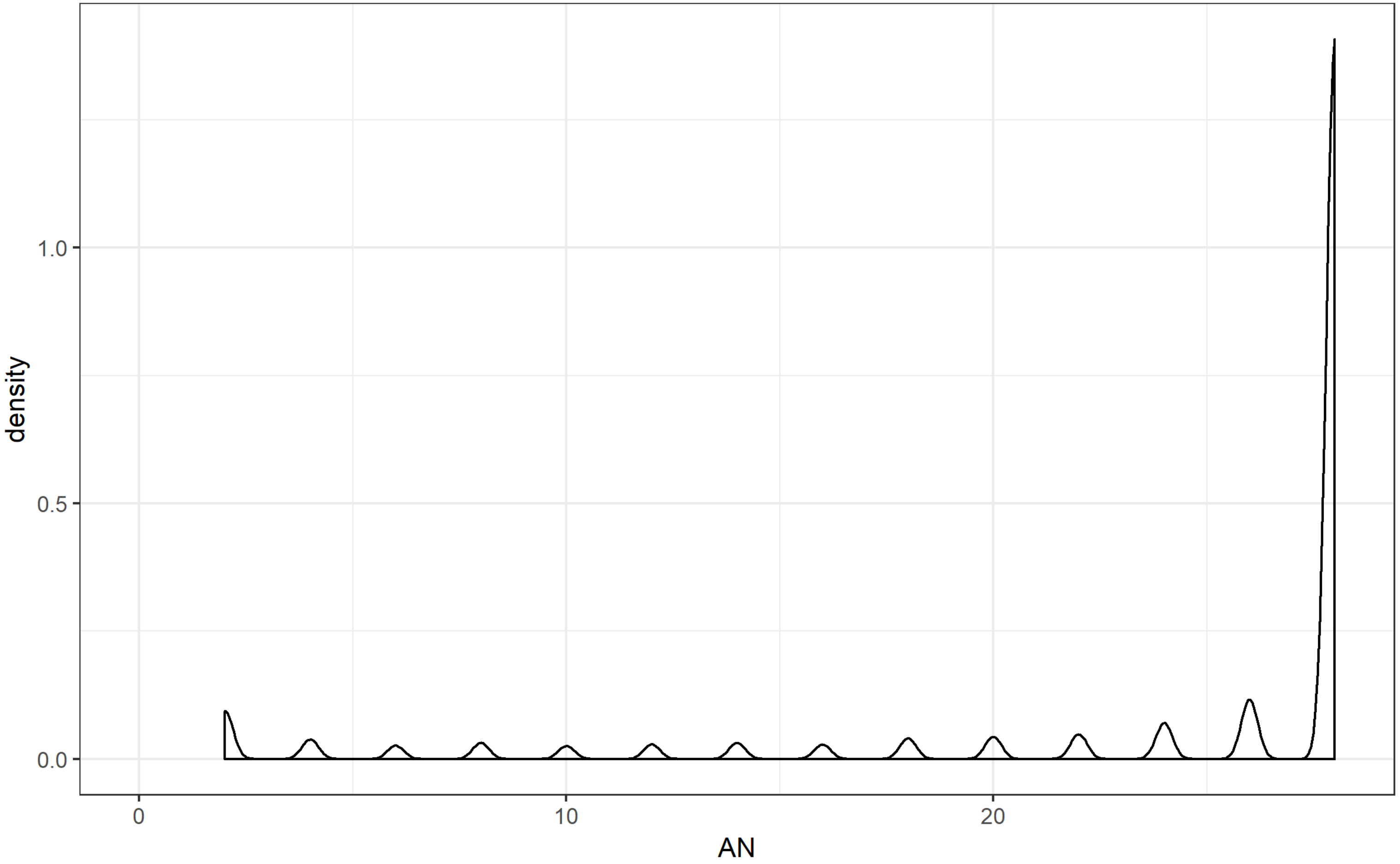
Hard-filter of Cayo exome data on chrX

Number of alleles that are genotyped AN

- There are 14 exomes → the maximum AN is 28, which means that a variant is genotyped across all 14 exomes

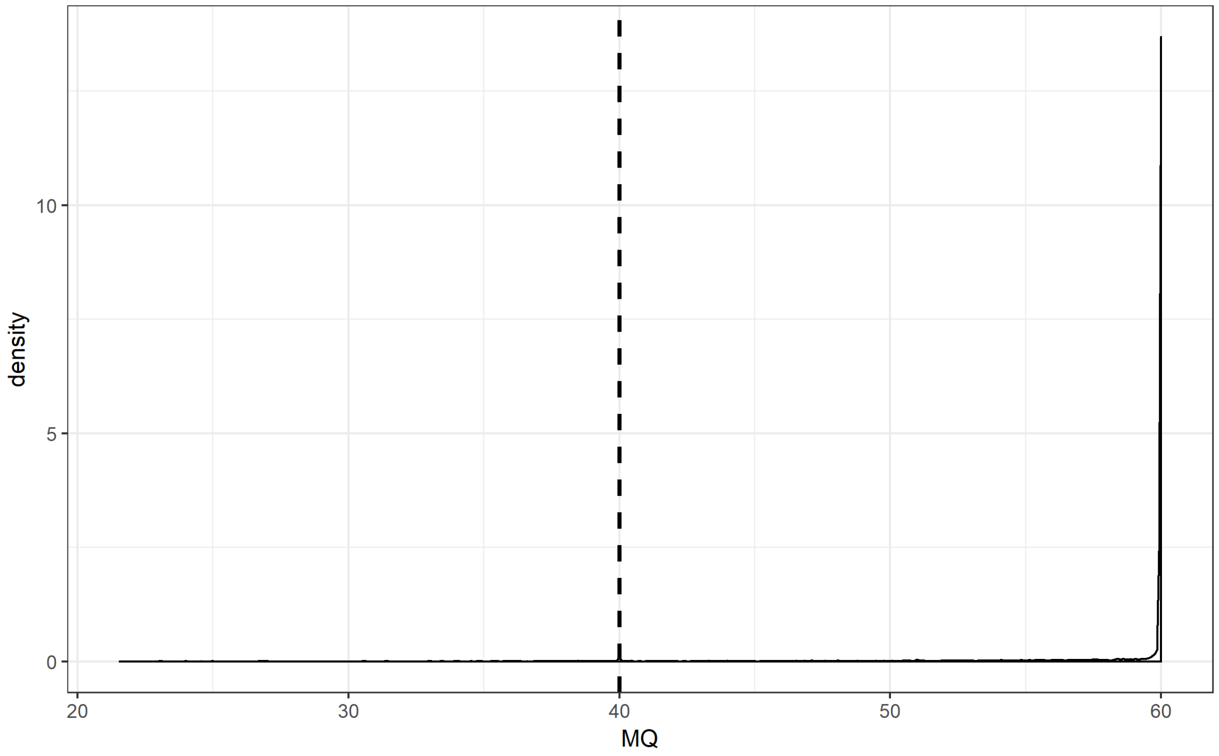


Investigating AN

		Number of variants (total = 16,299)
AN >= 2	Called in 1 (out of 14) exomes or more	16,299
AN >= 4	Called in 2 (out of 14) exomes or more	15,540
AN >= 6	Called in 3 (out of 14) exomes or more	15,232
AN >= 8	Called in 4 (out of 14) exomes or more	15,023
AN >= 10	Called in 5 (out of 14) exomes or more	14,769
AN >= 12	Called in 6 (out of 14) exomes or more	14,563
AN >= 14	Called in 7 (out of 14) exomes or more	14,336
AN >= 16	Called in 8 (out of 14) exomes or more	14,085
AN >= 18	Called in 9 (out of 14) exomes or more	13,859
AN >= 20	Called in 10 (out of 14) exomes or more	13,535
AN >= 22	Called in 11 (out of 14) exomes or more	13,184
AN >= 24	Called in 12 (out of 14) exomes or more	12,795
AN >= 26	Called in 13 (out of 14) exomes or more	12,232
AN = 28	Called in 14 (out of 14) exomes	11,298

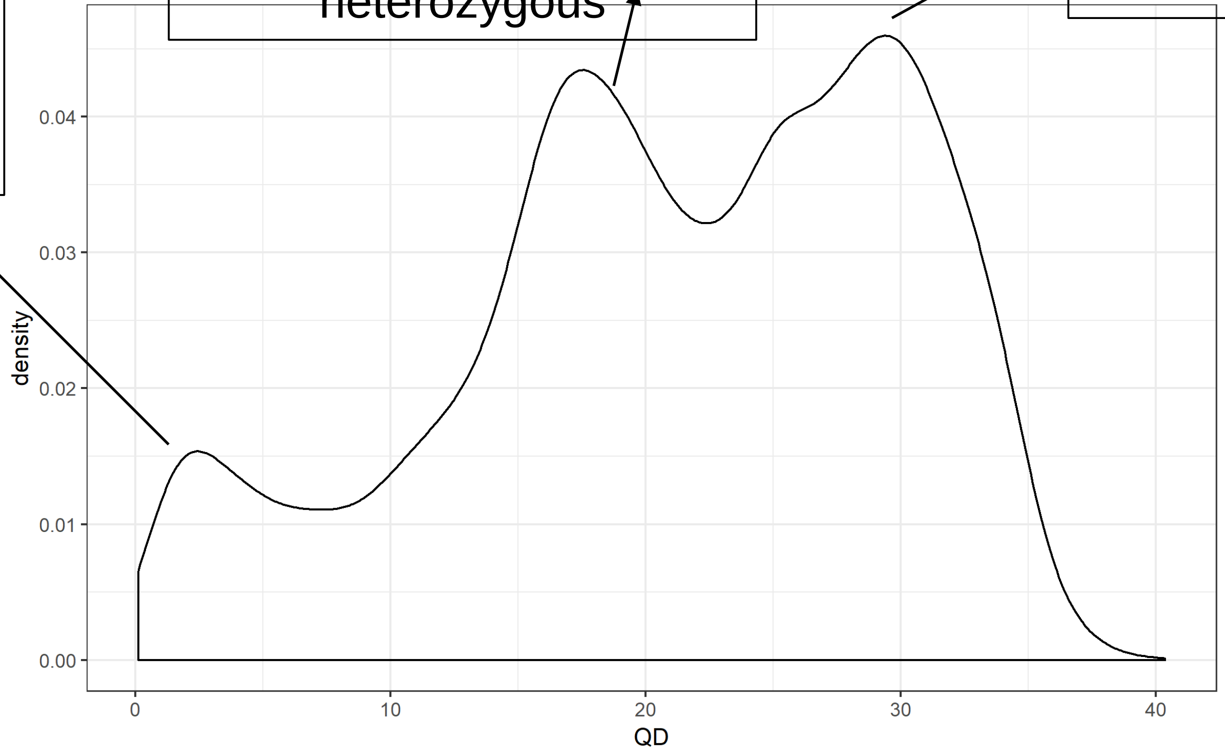
RMSMapping quality (MQ)

- From GATK: “This is the root mean square mapping quality over all the reads at the site. Instead of the average mapping quality of the site, this annotation gives the square root of the average of the squares of the mapping qualities at the site.” and “When the mapping qualities are good at a site, the MQ will be around 60.”



- GATK’s hard-filter suggests to remove variants with MQ < 40
- 15,740 out of 16,299 variants with MQ > 40

Quality by depth (QD)



GATK recommends to remove this peak

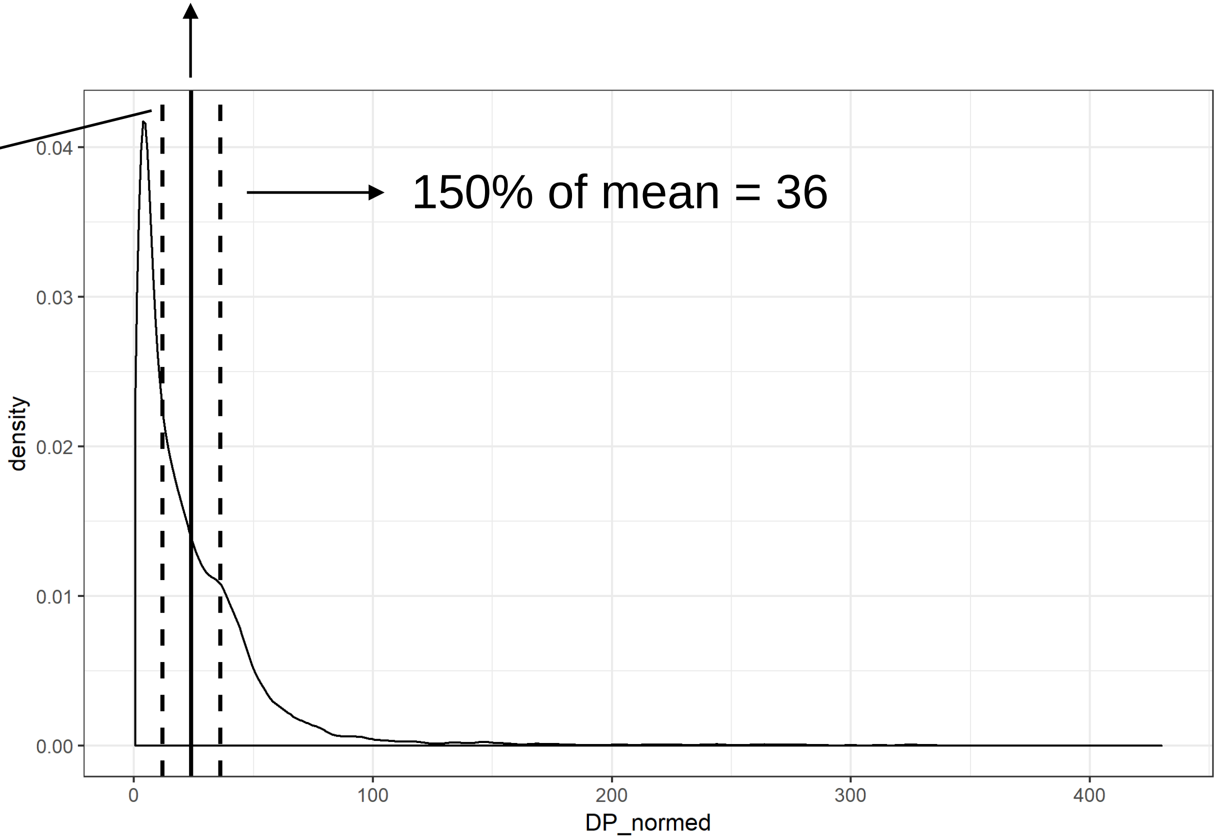
This peak represents variants that are mostly heterozygous

This peak represents variants that are mostly homozygous

- Use QD threshold of 5
- There are 15,152 out of 16,299 variants with QD > 5

Total depth of coverage over all sample (DP)

- DP from the INFO field from VCF file is summed across all samples. To get the mean, I divided DP by (AN/2) for that variant



mean = 24

150% of mean = 36

50% of mean = 12

Investigating DP

	Number of variants
0-25X	10,613
25-50X	3,979
50-100X	1,366
100-150X	185
150-200X	63
200-250X	43
250-300X	32
300-350X	15
350-400X	2
>400X	1