

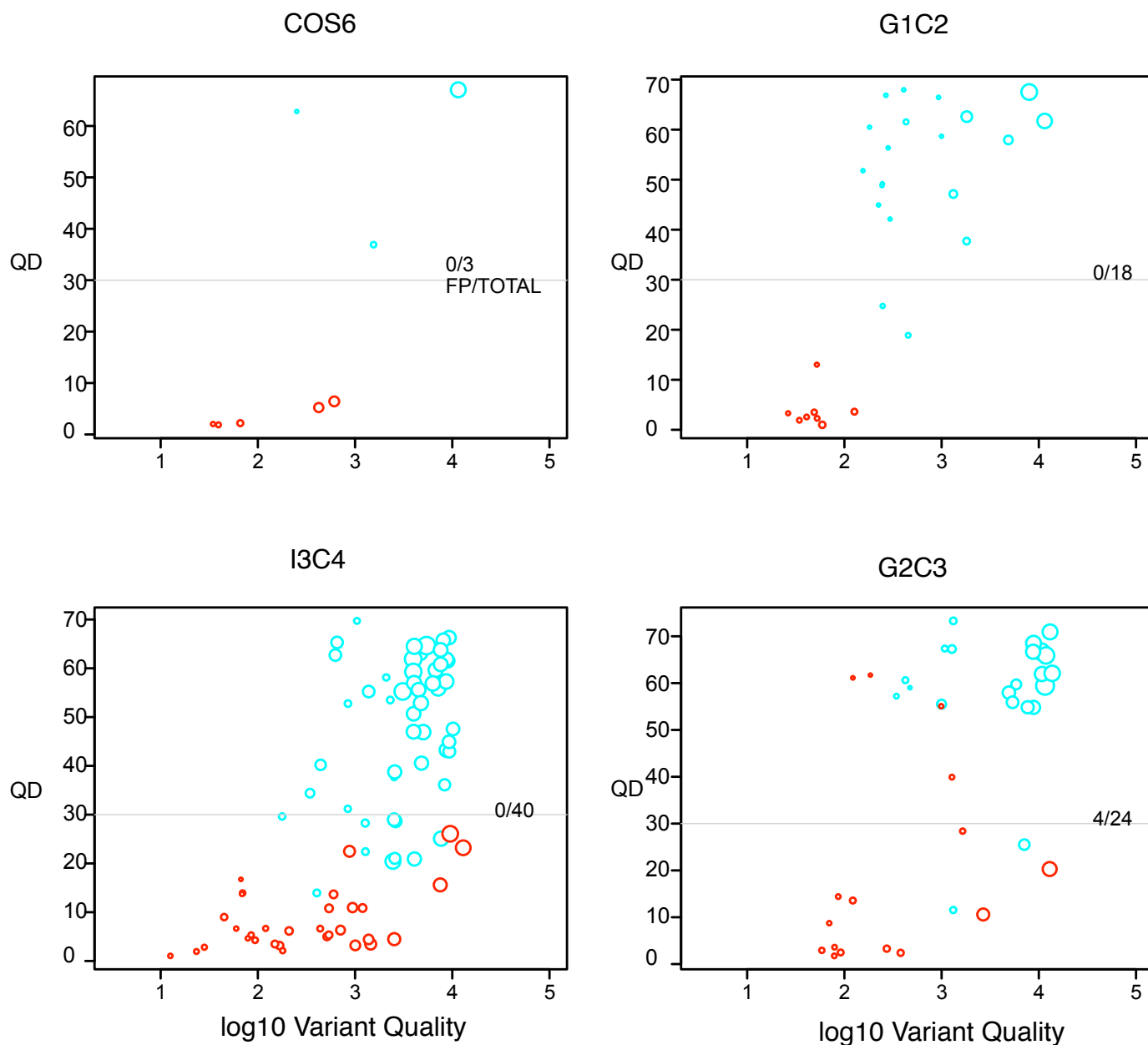


Figure S8. Sanger sequence data used to set empirical thresholds for SNP-calling in Illumina data. Four loci were sequenced in all forty TN lines in both directions (see Methods). **RED**: Not a Sanger SNP (False Positive). **CYAN**: A Sanger SNP (True Positive). Size of point: allele frequency in 40 TN lines. QD: GATK quality scaled by depth, line at QD=30 is the selected threshold for SNP calling and the number of false positives/true positives is given for each of the four loci.