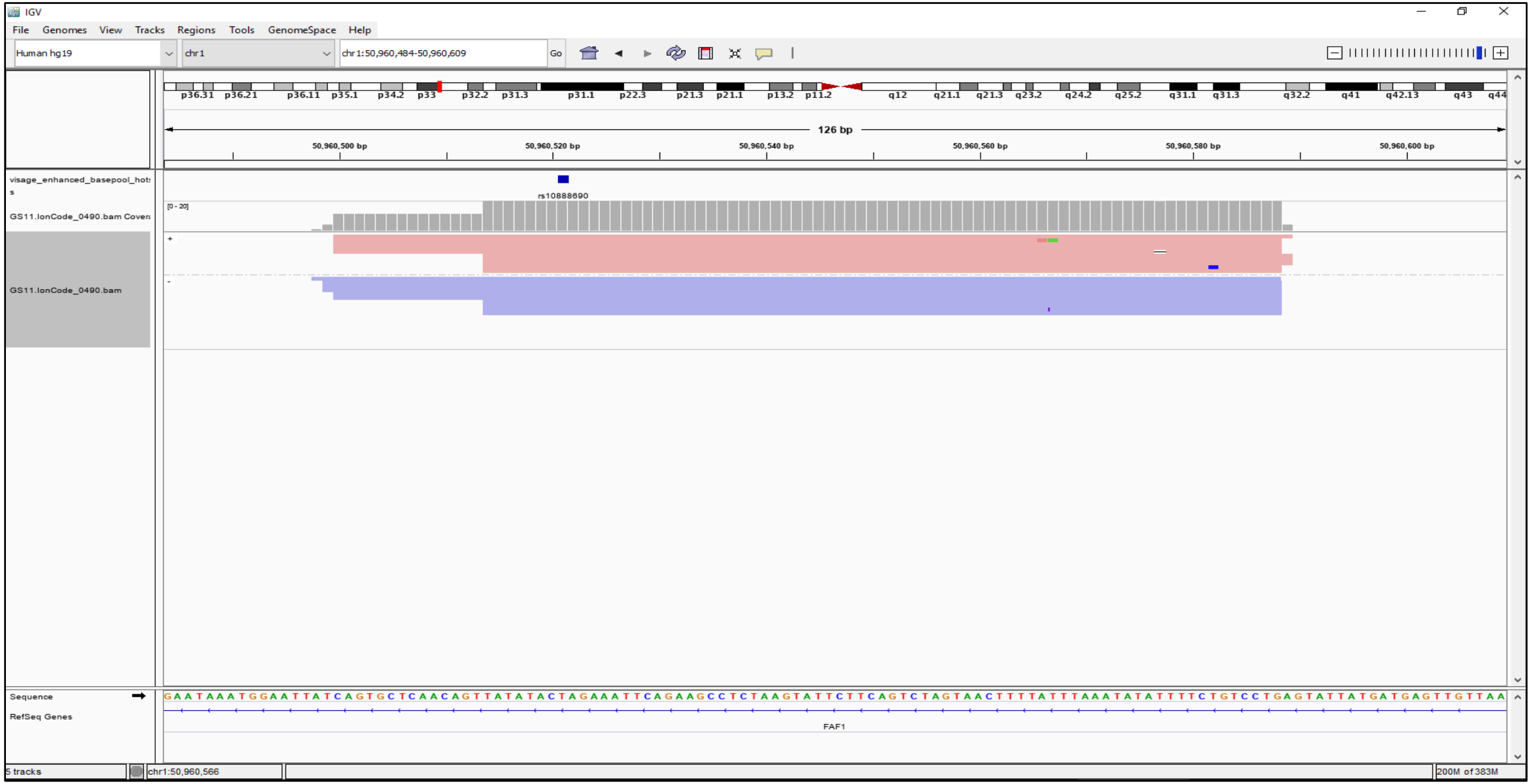
Development and inter-laboratory evaluation of the VISAGE Enhanced Tool for appearance and ancestry inference

Supplementary File S1: IGV screenshots of problematic SNPs and analysis recommendations.

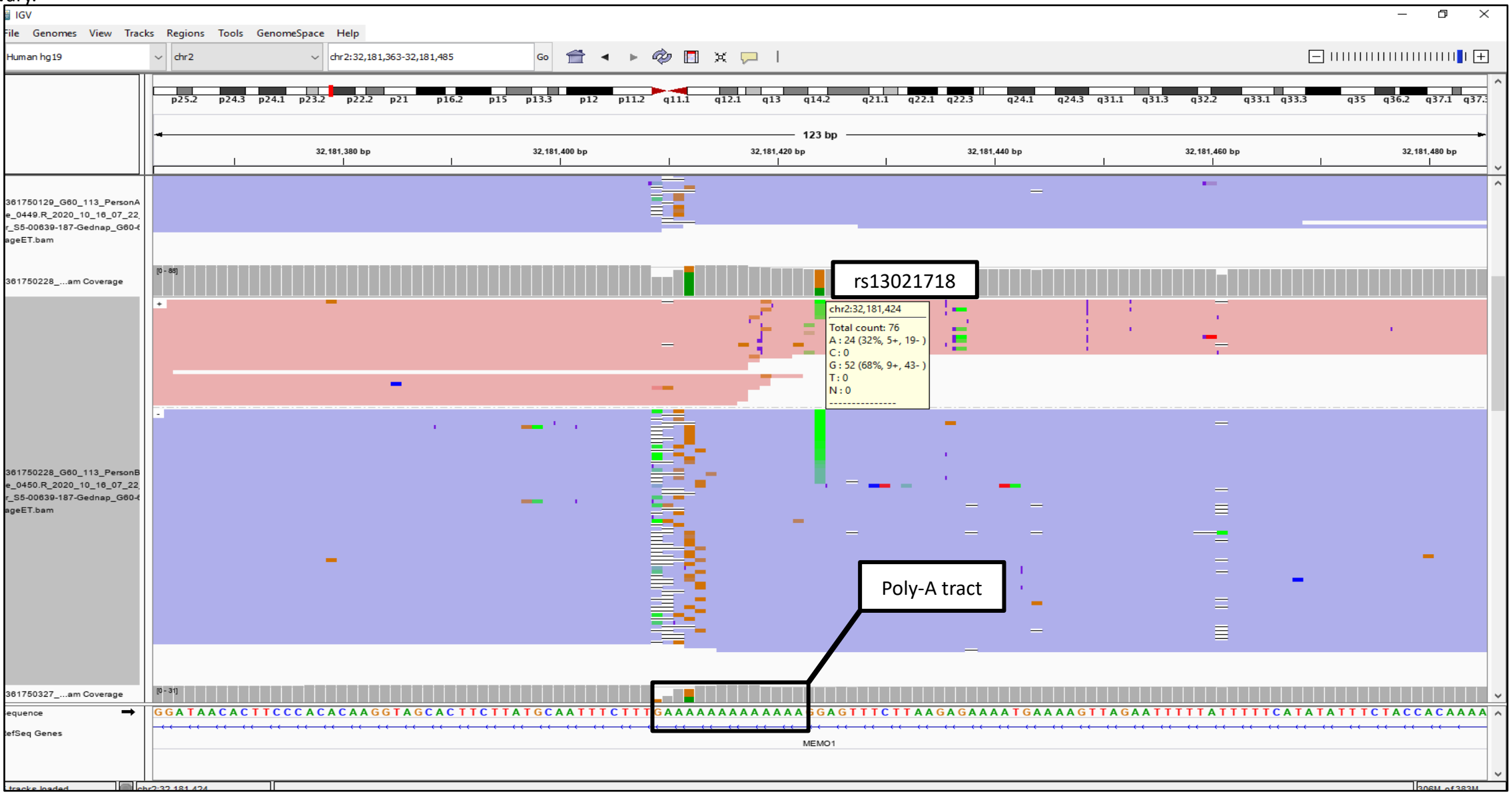
rs10888690 – No alignment problems are reported. The user should verify read quality and authenticity (forward and reverse) before manually calling the genotype. The user should input all possible genotypes (CC, CT and TT) and observe how the prediction probabilities vary.



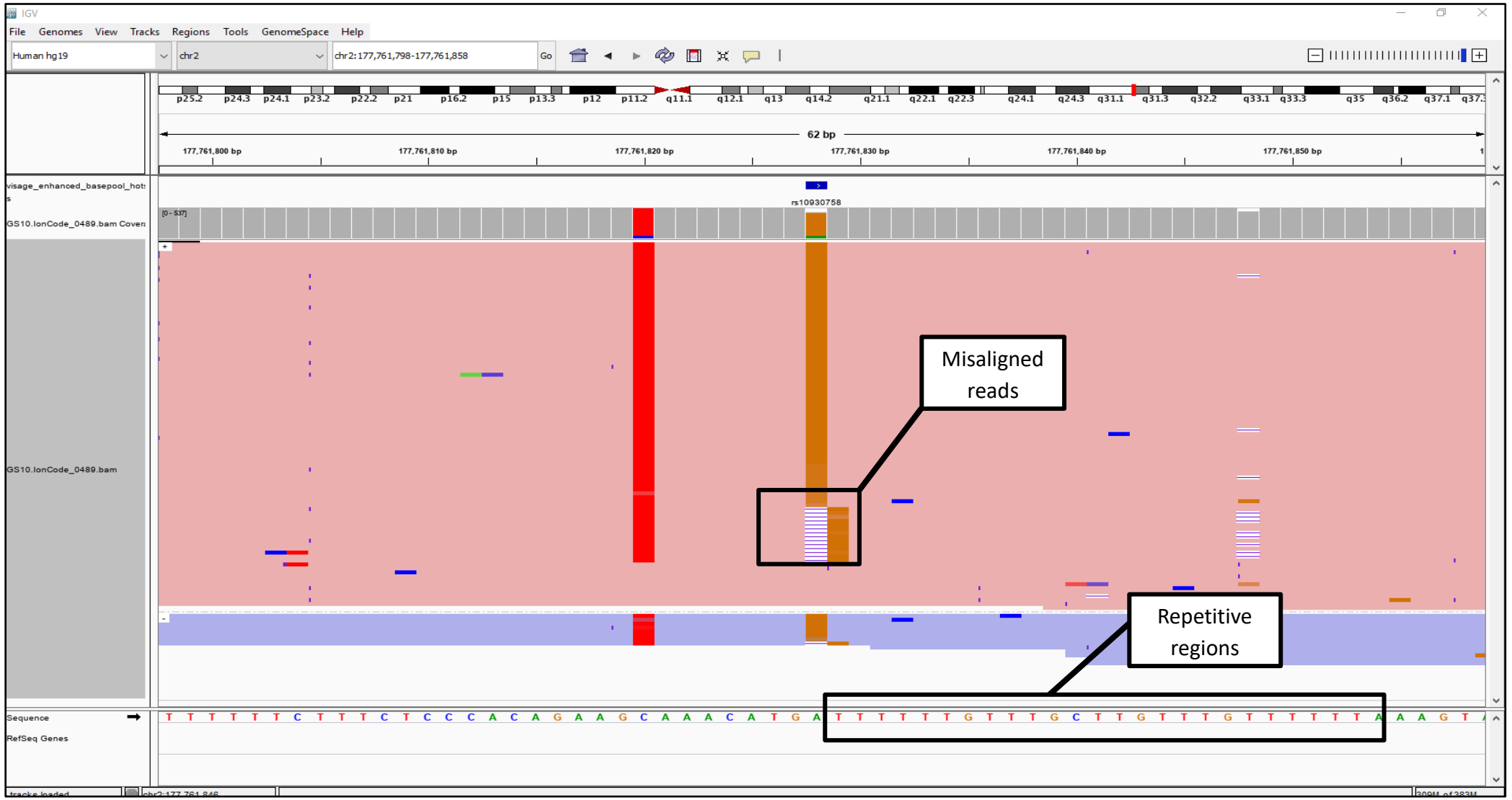
rs3862700 – The low efficiency of this amplicon probably originates in the presence of long poly A and poly-G stretches downstream. The user should input all possible genotypes (GG, GT and TT) and observe how the prediction probabilities vary.

rs9388490 – The low efficiency of this amplicon probably originates in the presence of long poly A stretch upstream. The user should verify read quality and authenticity (forward and reverse) before manually calling the genotype. The user should input all possible genotypes (CC, CT and TT) and observe how the prediction probabilities vary.

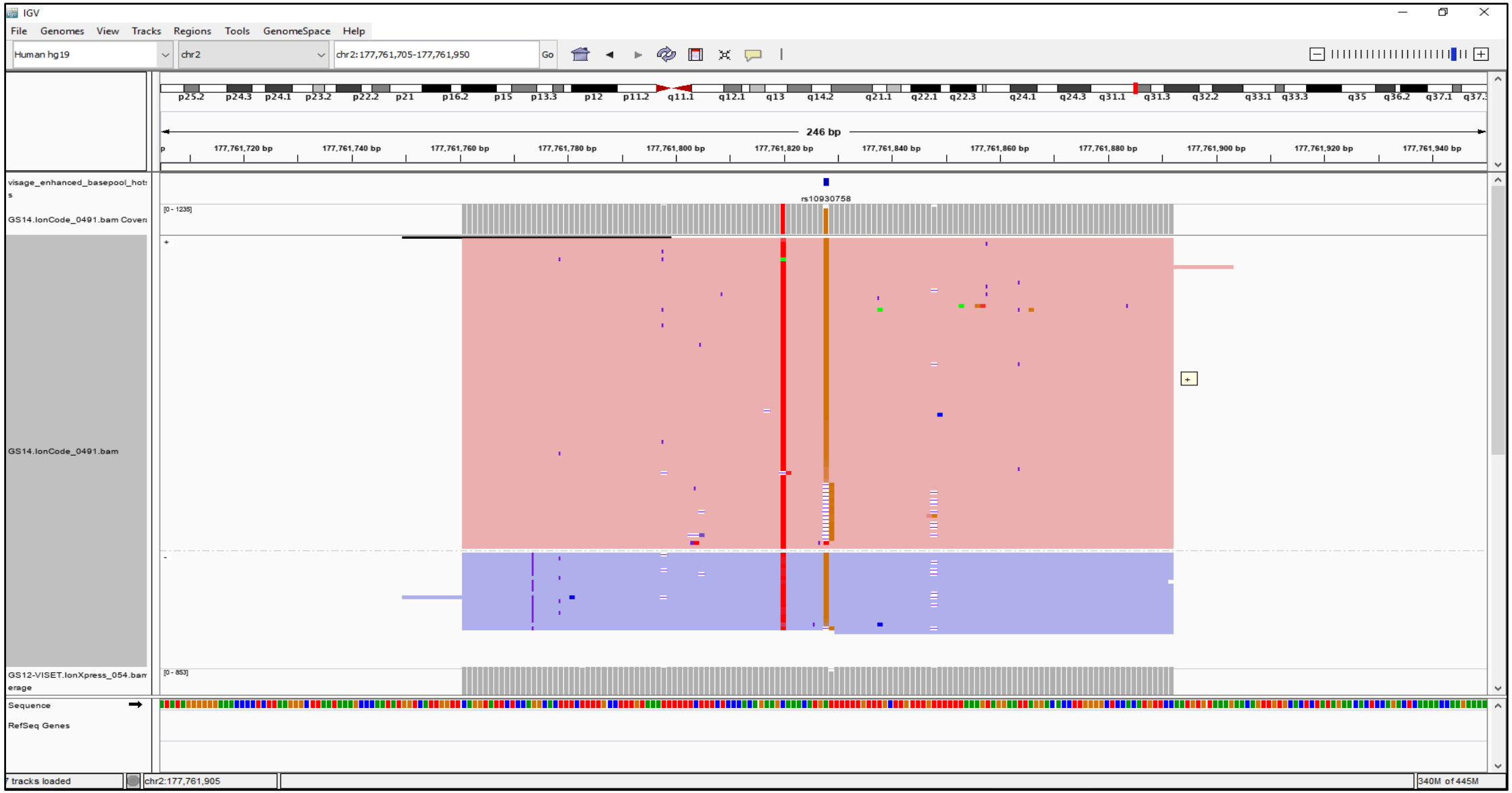
rs13021718 – The low efficiency of this amplicon probably originates in the presence of long poly A stretch upstream. The user should verify read quality and authenticity (forward and reverse) before manually calling the genotype. Here, the user should input all possible genotypes (AA, AG and GG) and observe how the prediction probabilities vary.



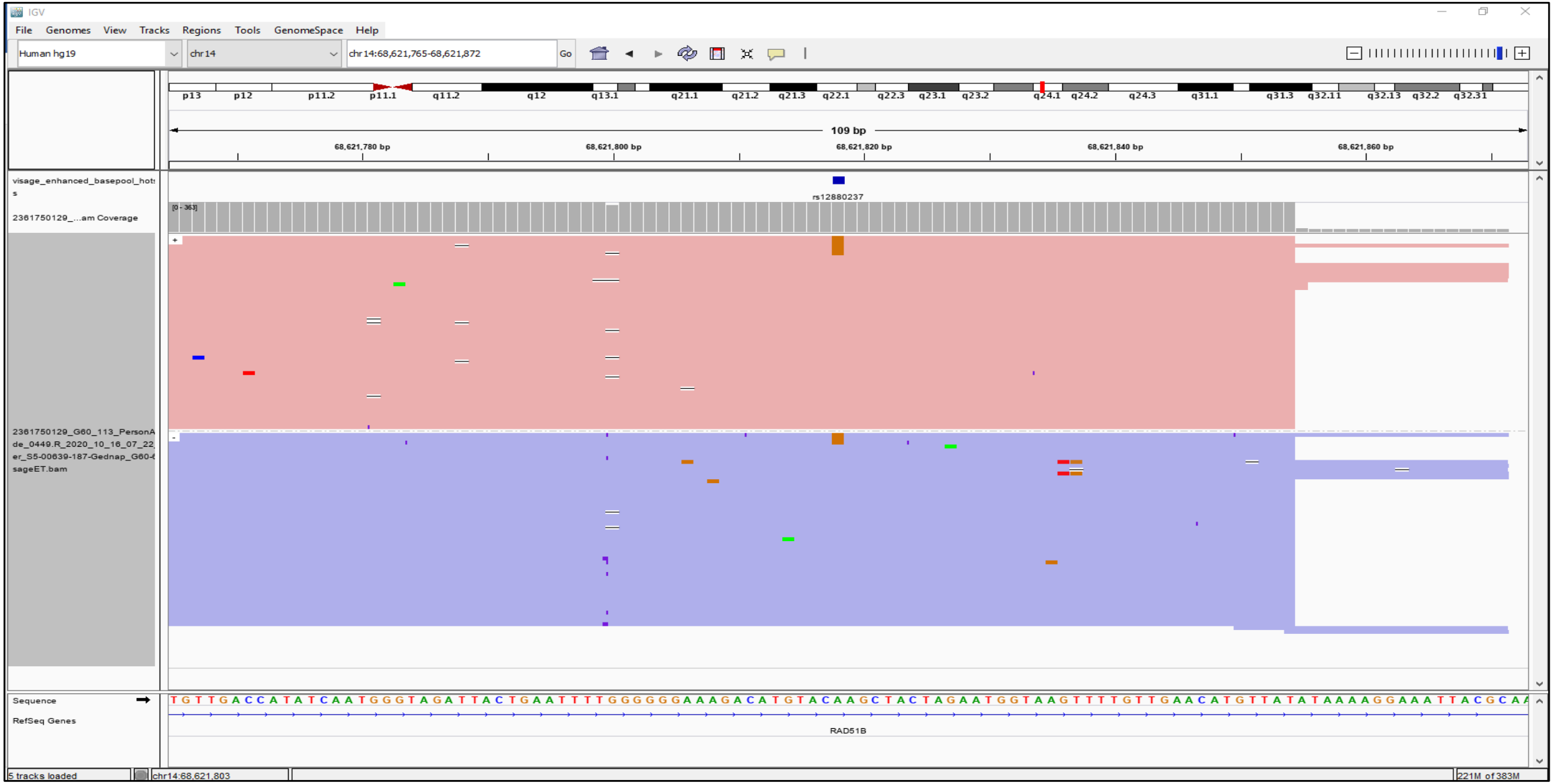
rs10930758 – A poly-T tract and repetitive regions downstream of the SNP leads to sequence incompleteness for the reverse reads. The SNP genotyper calls a very unbalanced AG genotype because the A reads are observed in both strands (forward and reverse). However, this highly unbalanced heterozygote does not fulfill the MAF threshold limits and thus, leads to a failed baldness prediction. This SNP presents a high misincorporation rate (1.5%), however we would expect the misincorporated reads to be Ts (due to the poly-T tract) and not As, thus the genotype AG is plausible, even if highly unbalanced. Here, the user should input all possible genotypes (AA, AG and GG) and observe how the prediction probabilities vary.



rs10930758 – Example of a true homozygote.



rs12880237 – The amplicon shows no alignment problem. Indeed, the presence of G reads looks reliable (see below), even if at a low frequency (~14%). Nevertheless, the SNP does not comply with the MAF intervals set in the software. Here, the user should input all possible genotypes (AA, AG and GG) and observe how the prediction probabilities vary.



rs11803731 – A poly-G tract downstream of the SNP location leads to misaligned reads (see below). Here, on the contrary to the rs10930758 example, the A reads are being misincorporated as an artifact from the poly-G. The most plausible genotype would be TT instead of the AT the genotyper has called. Again, manual observation of the data should be performed by the forensic analyst. The input all possible genotypes (AA, AG and GG) and verification of how the prediction probabilities vary is highly recommended.