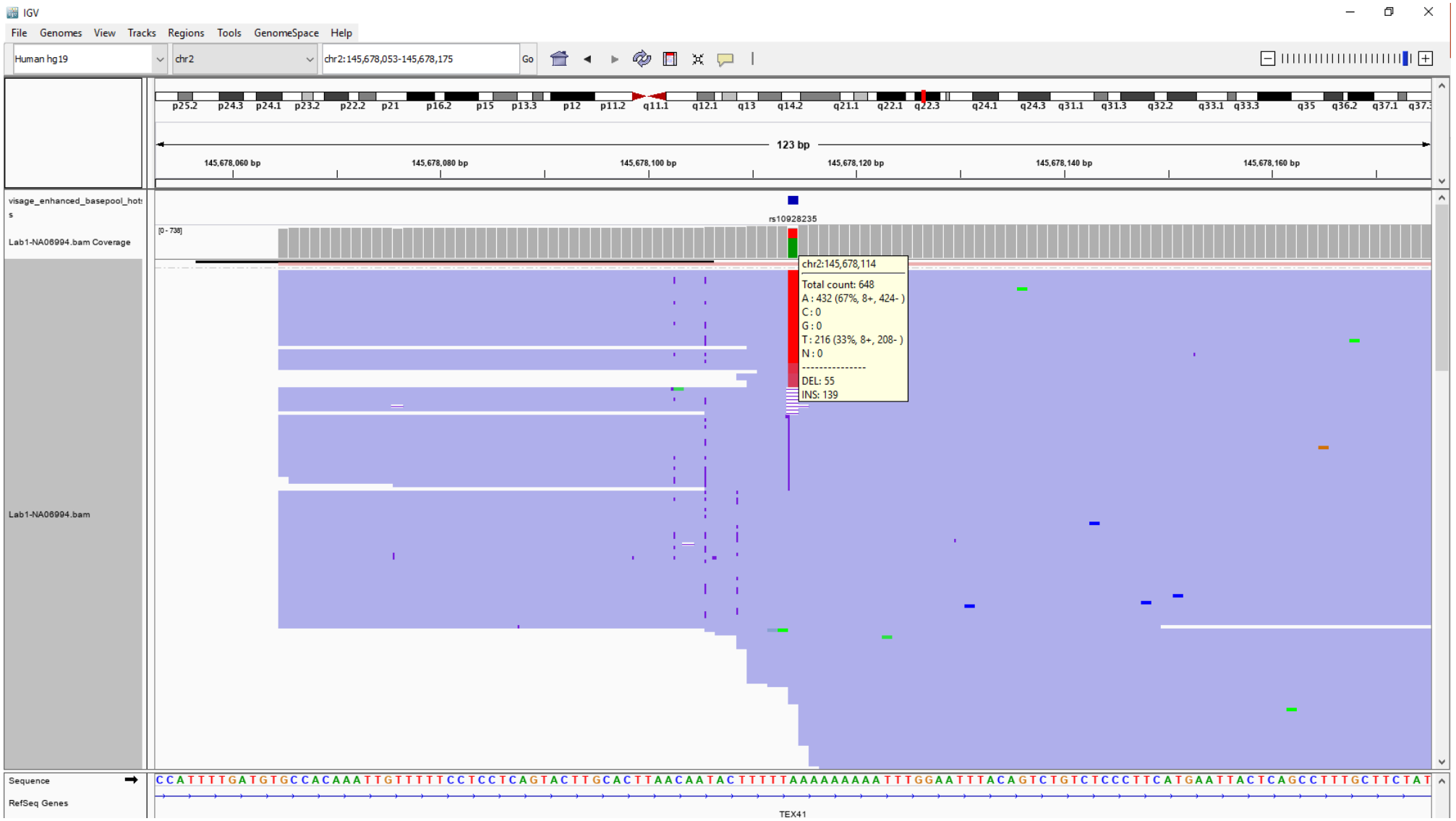
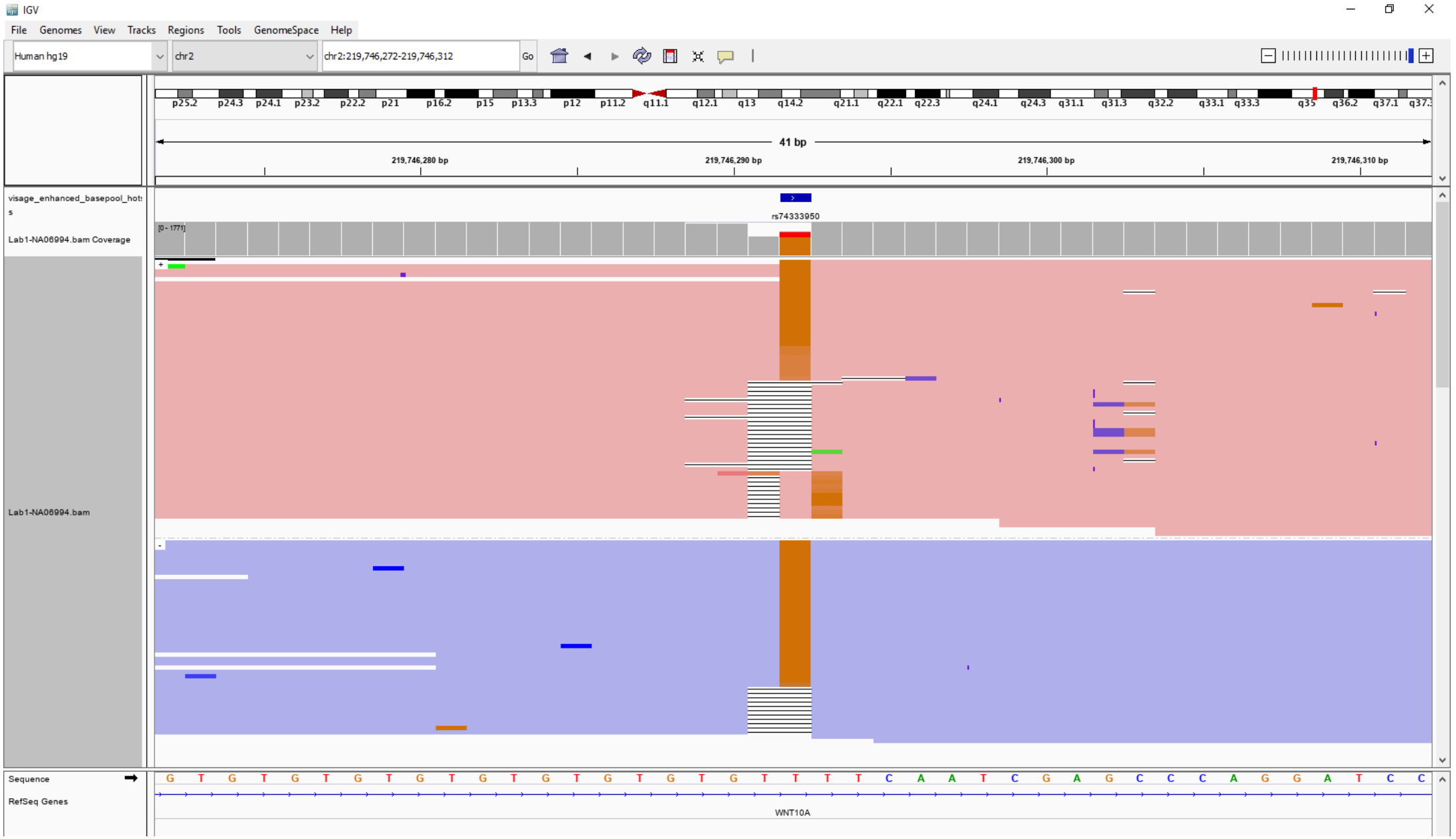
Development and inter-laboratory evaluation of the VISAGE Enhanced Tool for appearance and ancestry inference

Supplementary File S2: IGV screenshots of incoherencies between Coriell replicates of different laboratories.

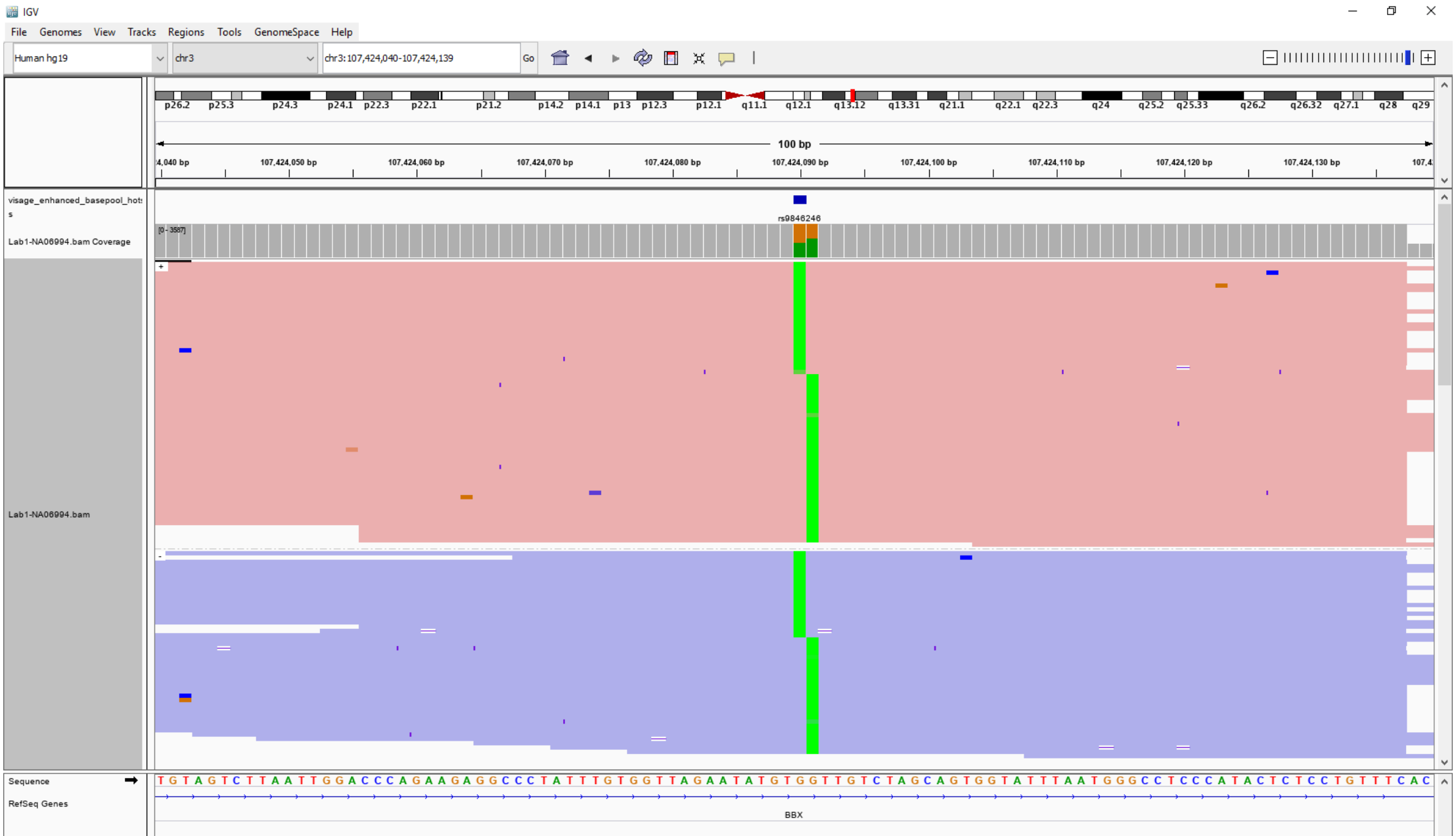
Lab1-NA06994 rs10928235 – No call (NN) by the HID-SNP-Genotyper led to a discordance between compared genotypes of different laboratories. Manual inspection with IGV showed the accurate genotype (AT) could be called, even if imbalanced.



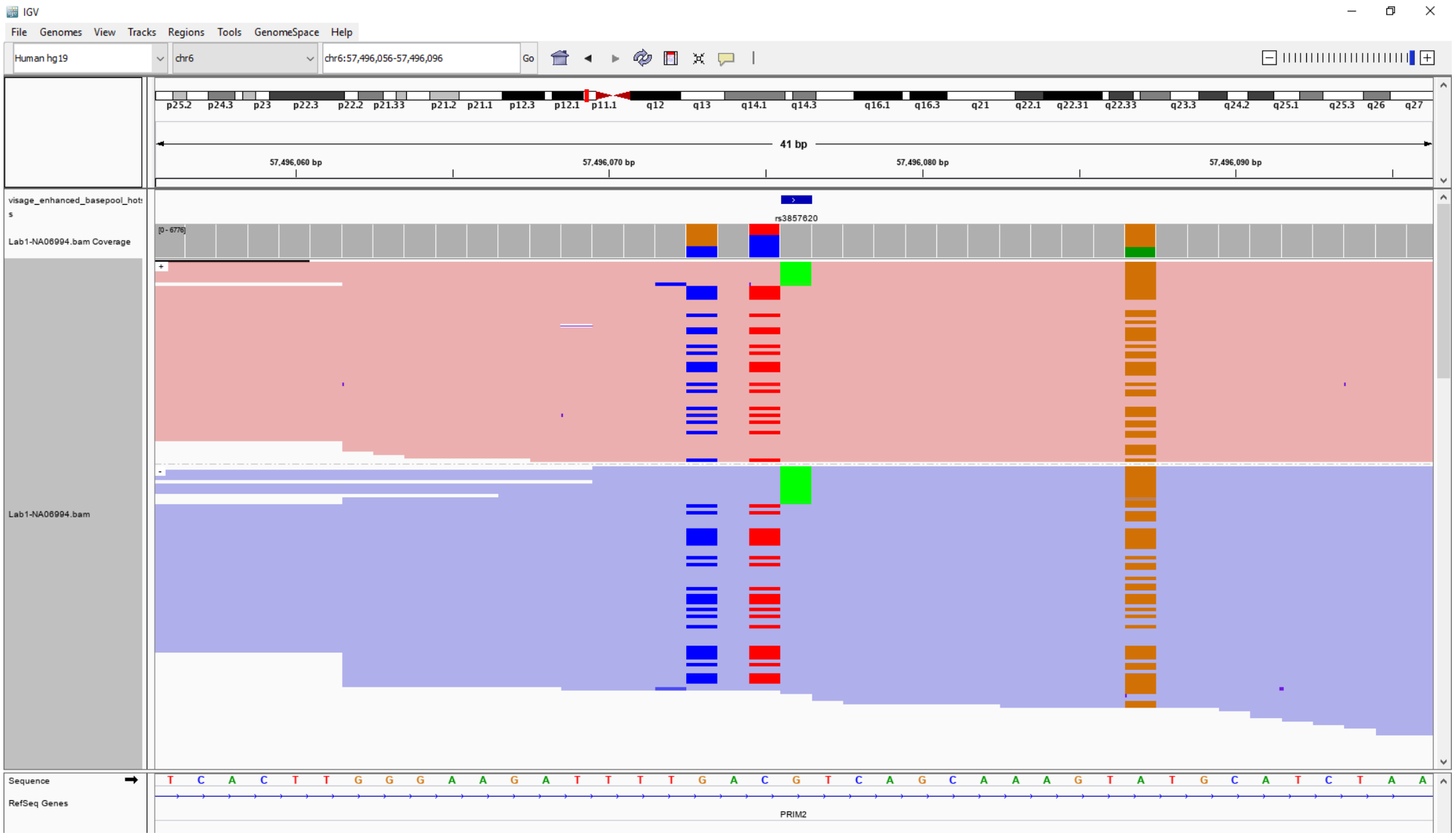
Lab1-NA06994 rs74333950 – Misaligned reads due to highly repetitive region upstream of the SNP lead to the erroneous incorporation of T reads. Here an example from Lab 1, but the same was observed for all laboratories.



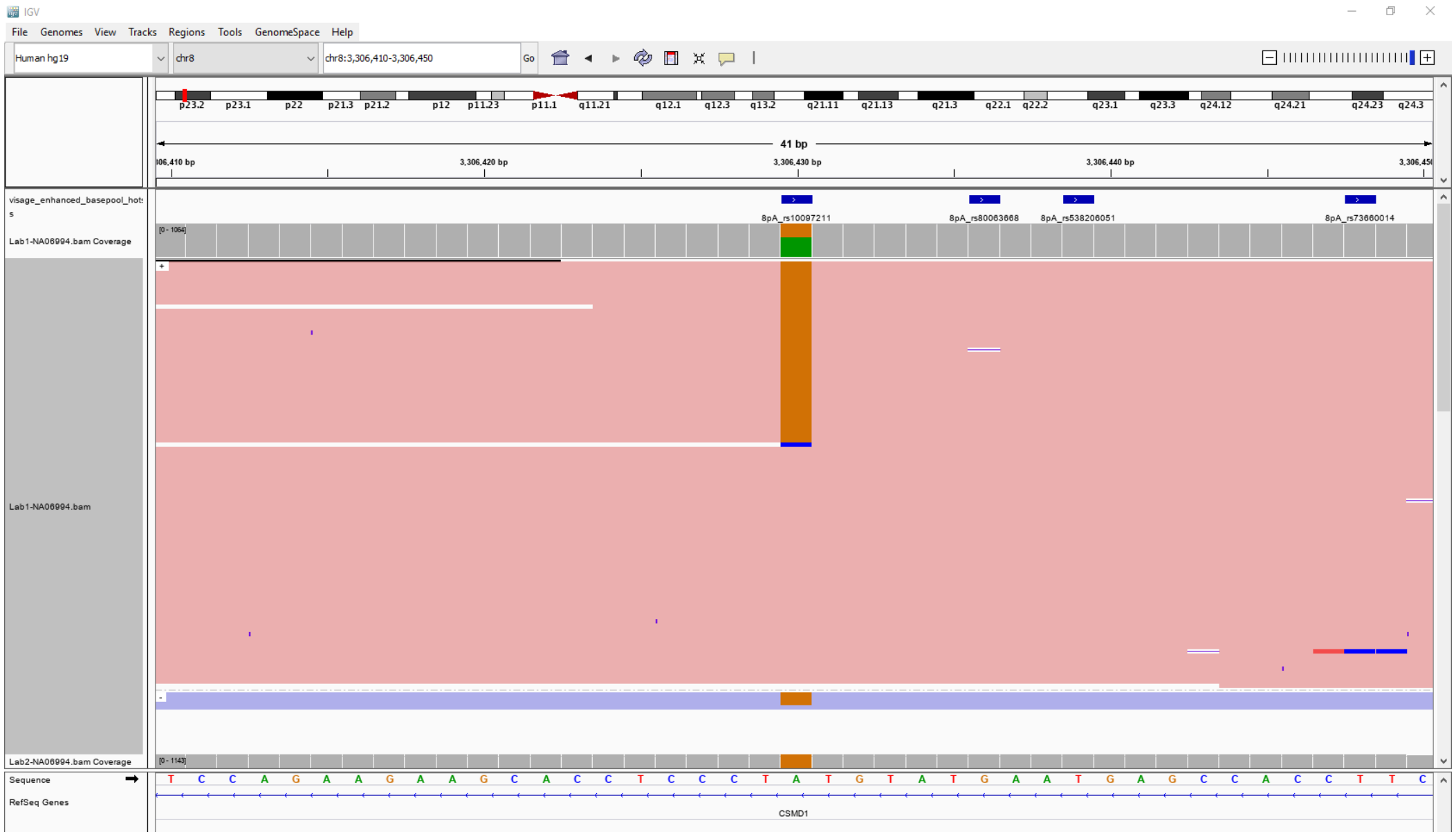
Lab1-NA06994 rs9846246 – example of AG genotype observed in Lab1 replicate. The same was observed for Labs 2 and 3.



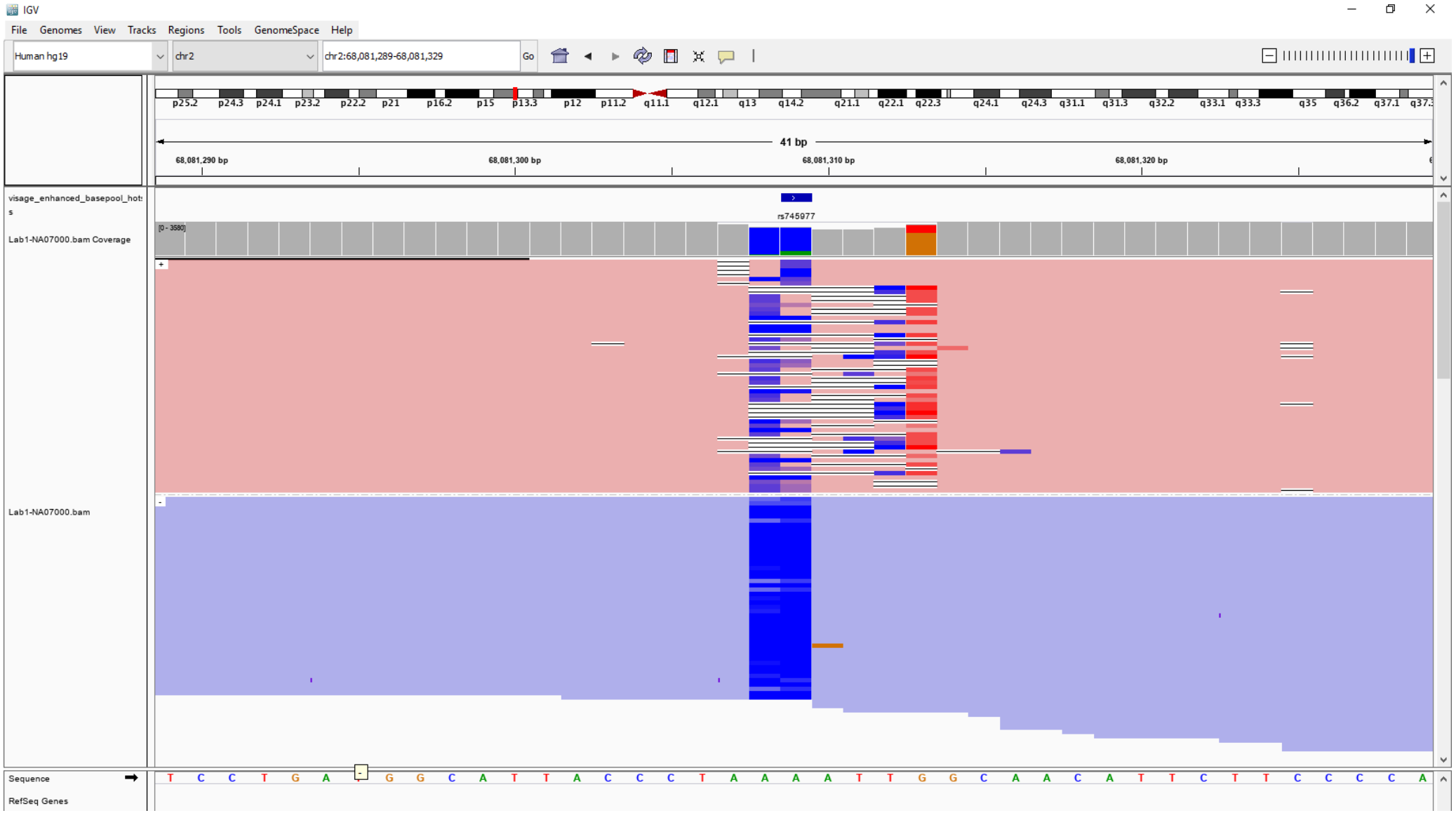
Lab1-NA06994 rs3857620 – example of the imbalanced AG genotype observed in Lab1 replicate. The same was observed for Labs 2 and 3.



Lab1-NA06994 8pA_rs10097211 – example of the AG genotype observed in Lab1 replicate. The same was observed for Labs 2 and 3.

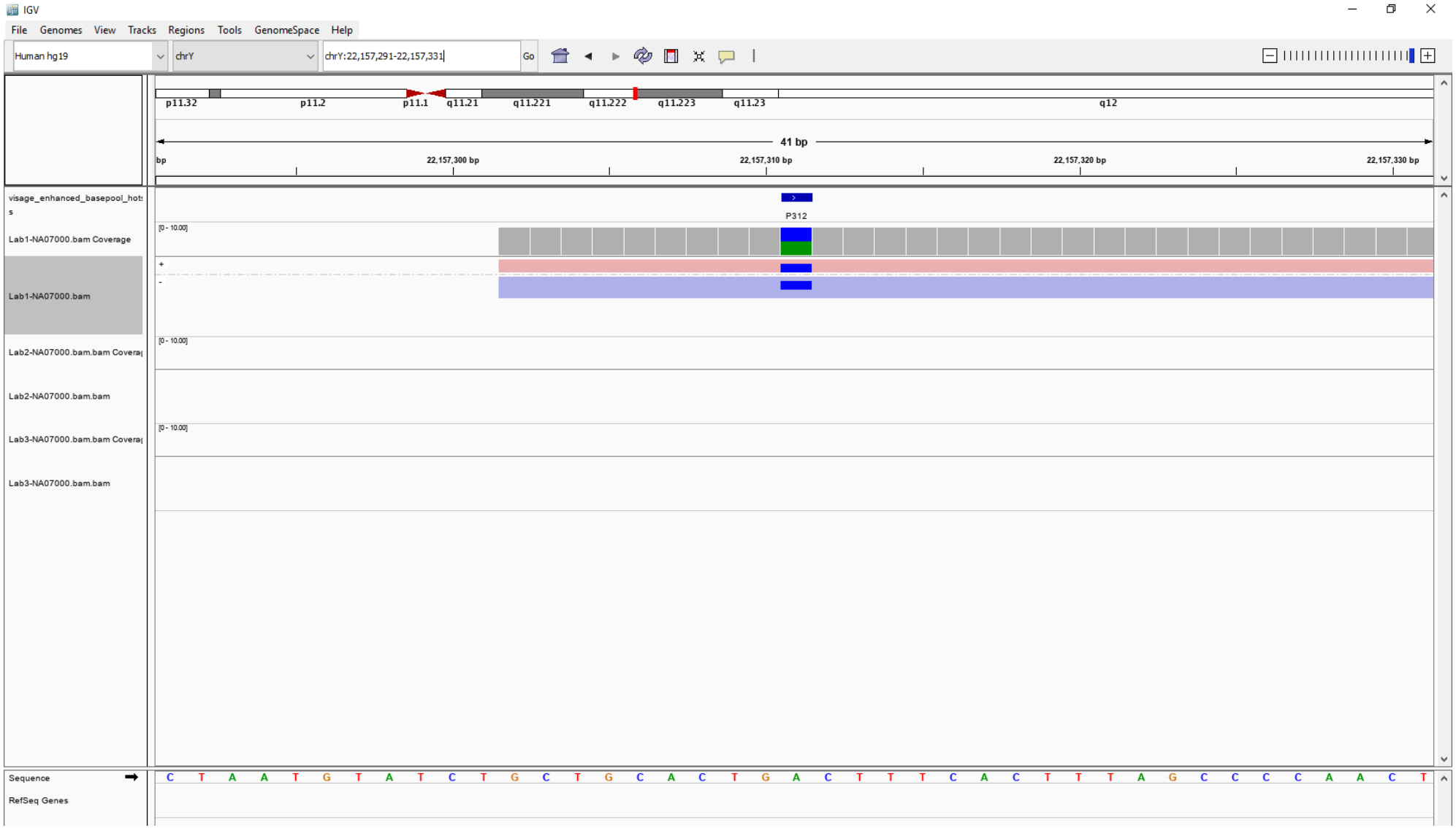


Lab1-NA0700 rs745977 – Misaligned reads due to poly-A tract at the SNP site lead to the erroneous incorporation of A reads (see reverse strand reads). The other laboratories produced an NN result which would result in a failed prediction. In this particular case, as genotype calling is difficult, the obtained NN genotype would prevent false results.

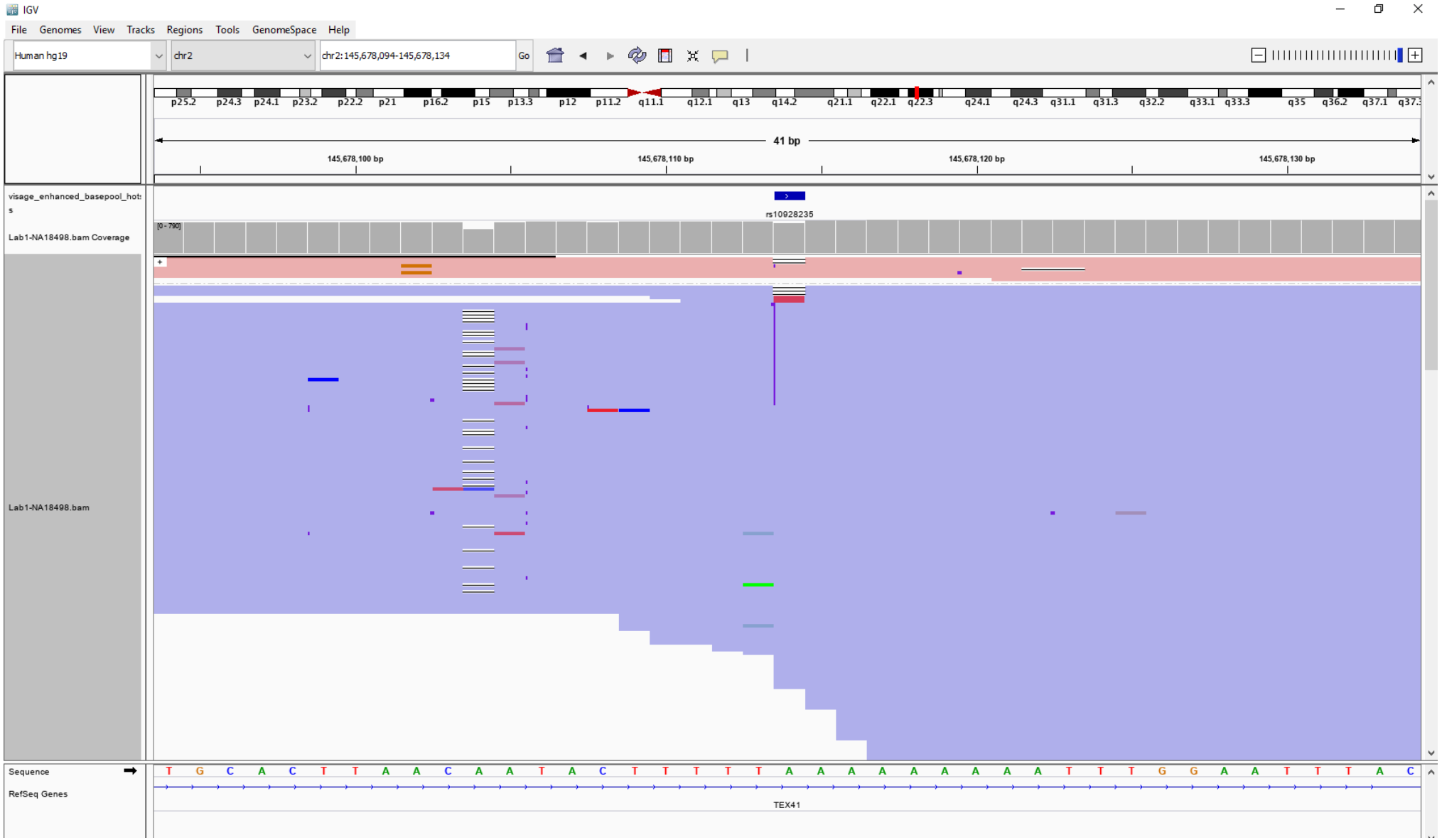


Lab1-NA07000 rs3857620 – example of the imbalanced AG genotype observed in Lab1 replicate. The same was observed for Labs 2 and 3.

Lab1-NA07000 P312 – dropin at Y chromosome marker in female sample.

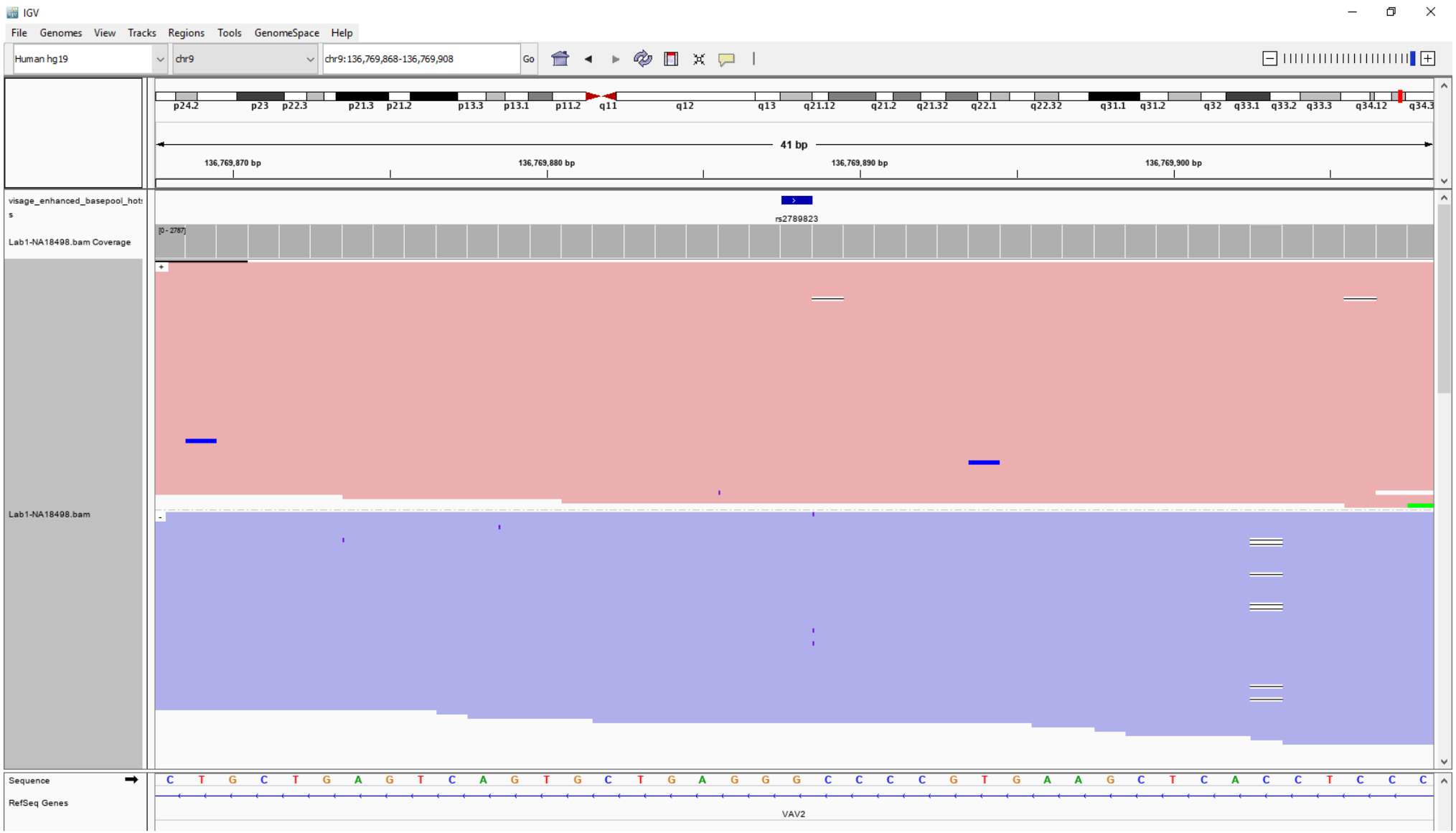


Lab1-NA18498 rs10928235 – A no call (NN) by the HID-SNP-Genotype was obtained for all replicates of the inter-laboratory tests. A manual inspection with IGV showed the presence of A reads at the SNP position. The no call might have resulted from the presence of a poly-A and a poly-T tract at the SNP site that prevent the sequencing of the forward reads.

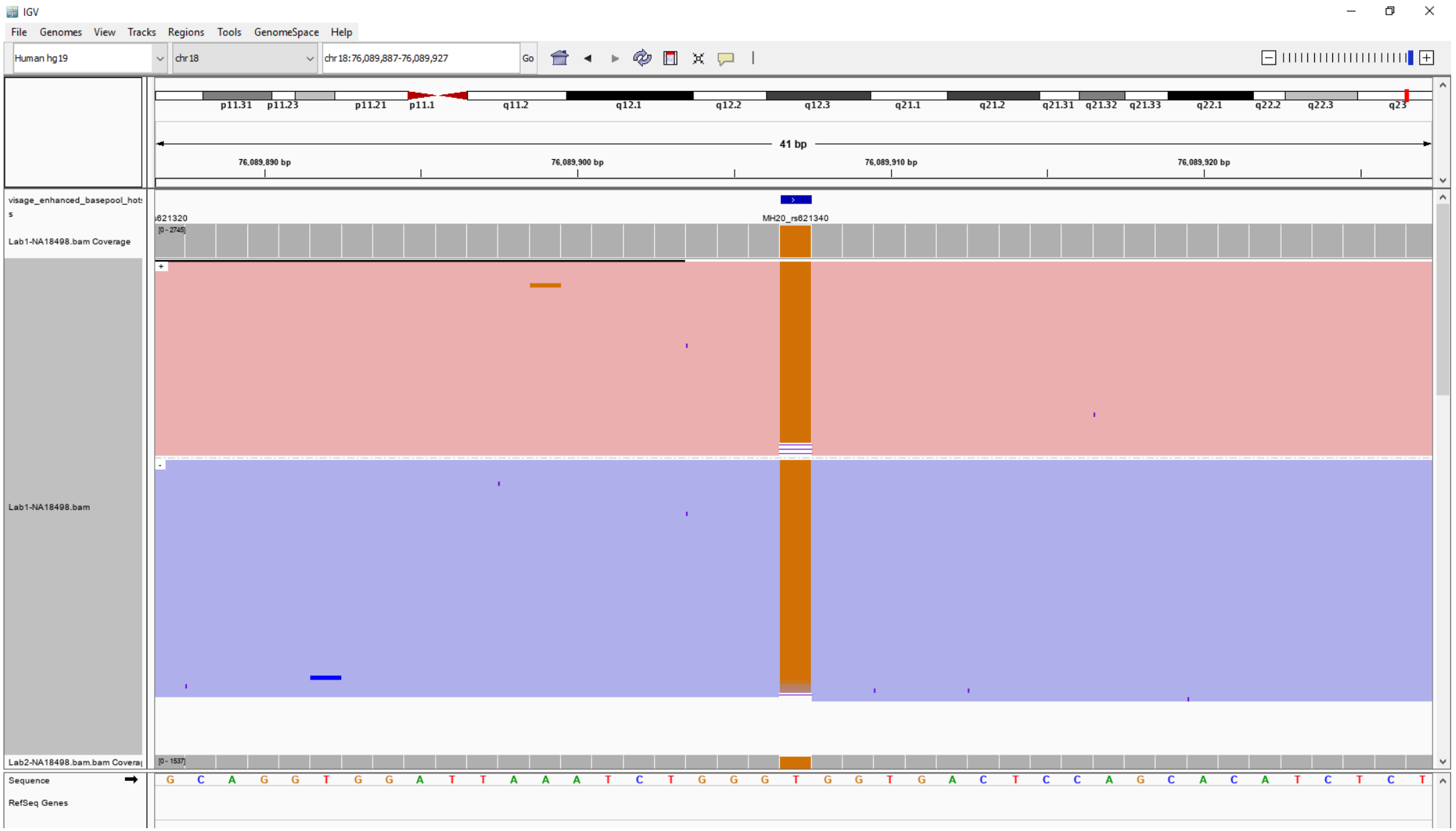


Lab1-NA18498 rs71530654 – Imbalanced AG genotype for Lab 1 and GG genotype for Labs 2-3. Difficulties in genotype calling probably due to highly repetitive region upstream of the SNP. Due to alignment issues this SNP was removed from the panel.

Lab1-NA18498 rs2789823 – Clear GG genotype for Labs 1-3. Here the example of Lab 1, but the same was observed for the other replicates.

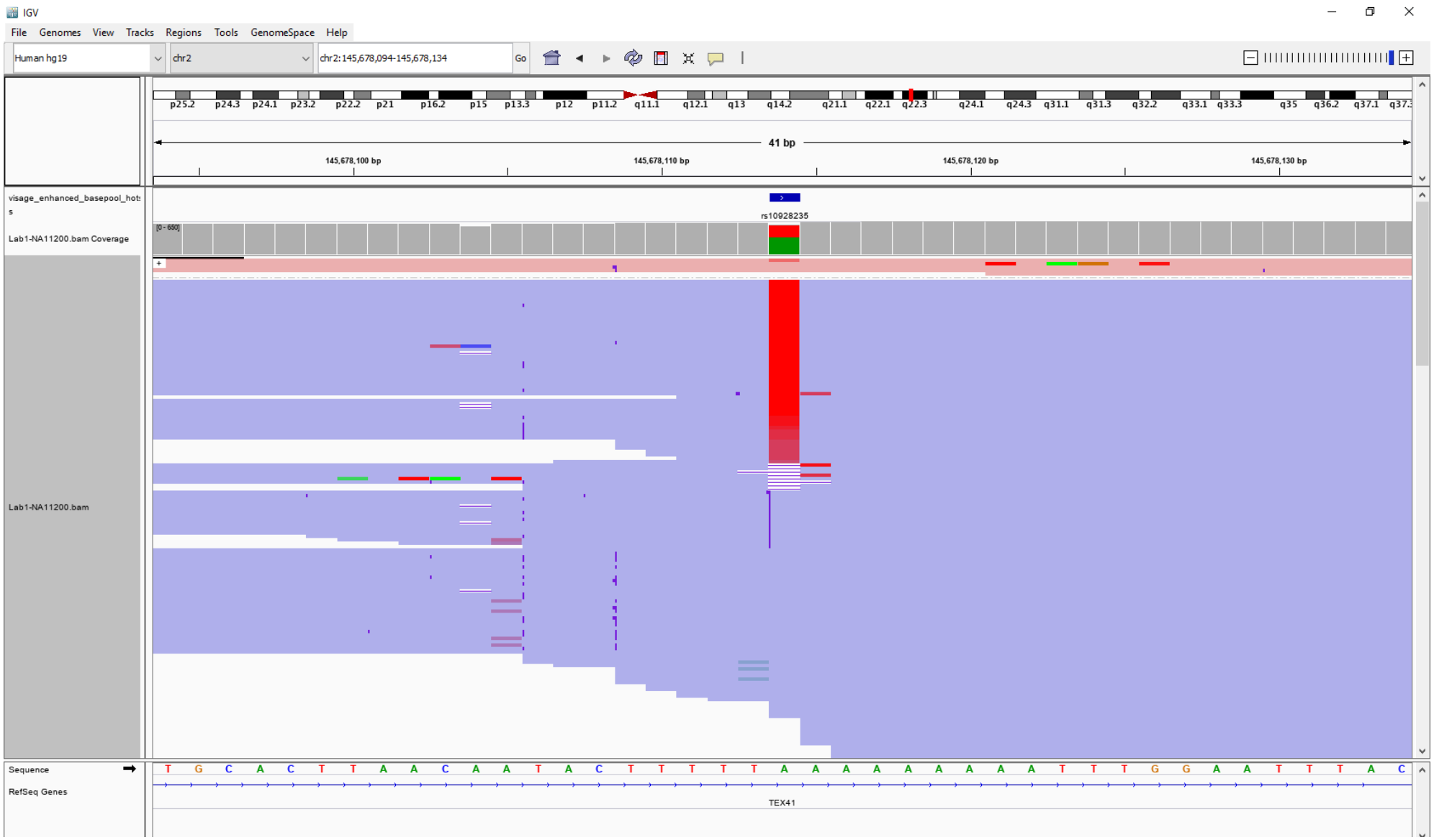


Lab1-NA18498 MH20_rs621340 – Clear GG genotype for Labs 1-3. Here the example of Lab 1, but the same was observed for the other replicates.

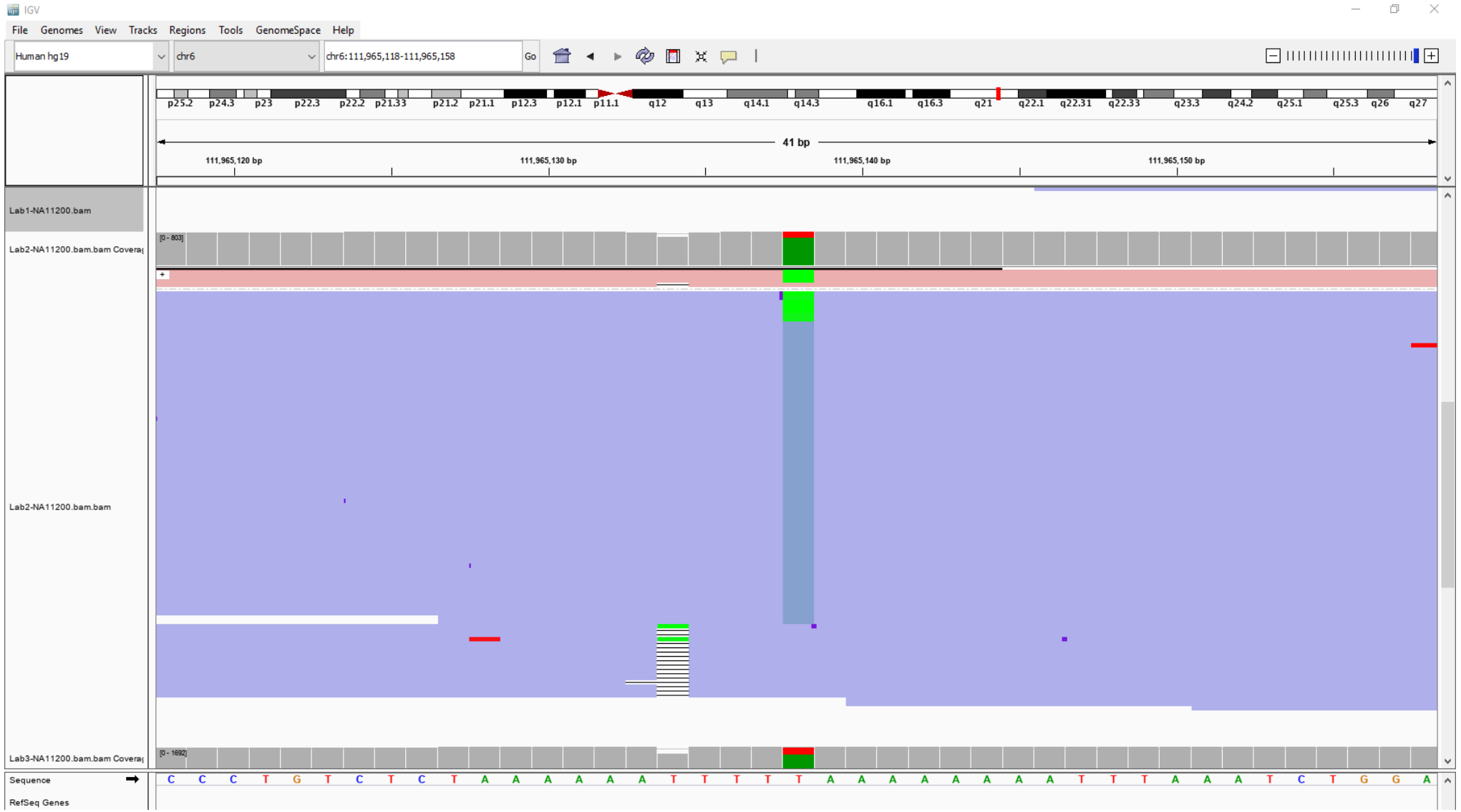


Lab1-NA18498 rs112348497 – This is an InDel variant, which causes problems for genotype calling when the insertion variant -CCC- is present (reference is homozygote for CC). InDel variants should be called with the Variant Caller Plugin.

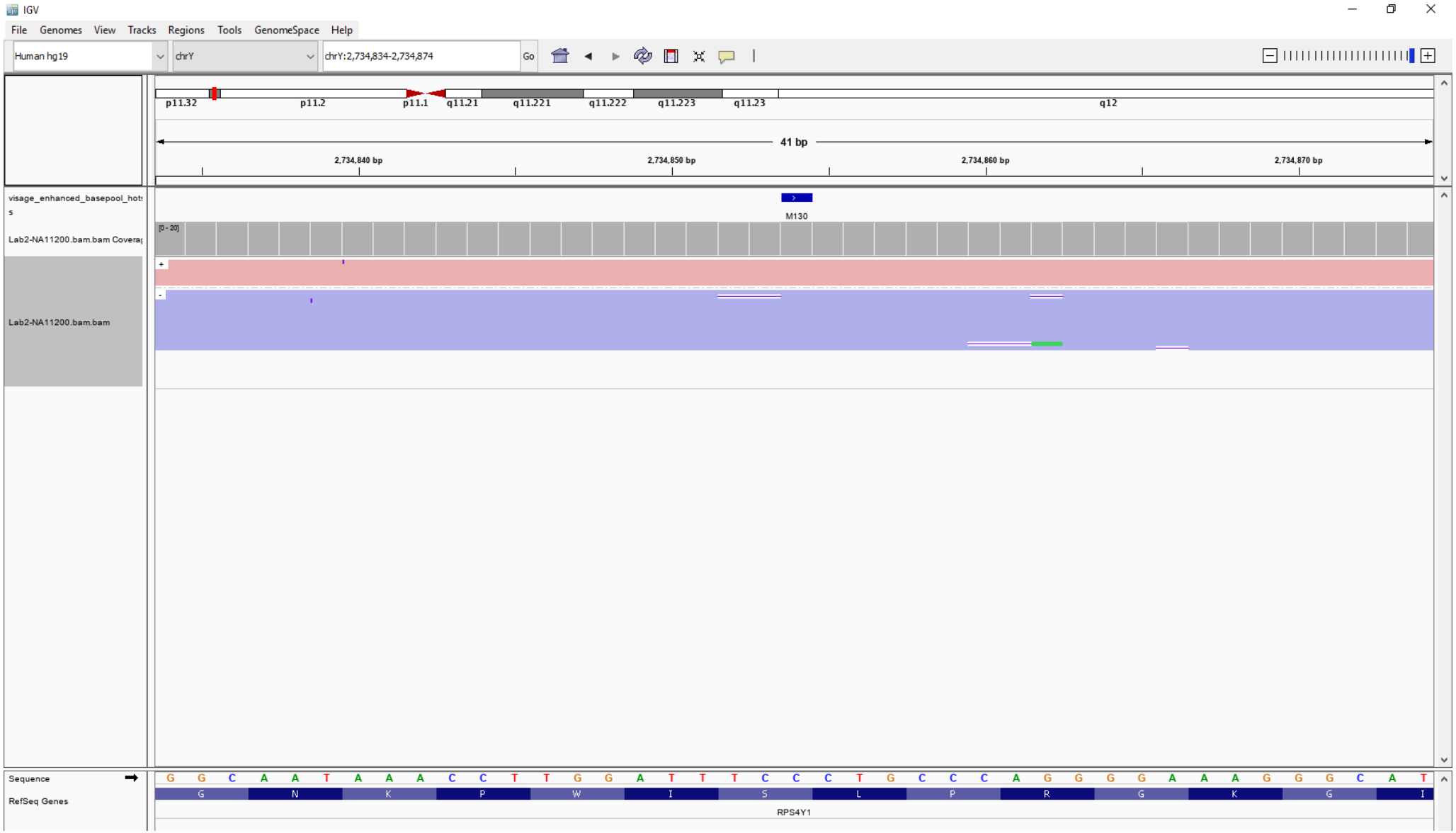
Lab1-NA11200 rs10928235 – A no call (NN) was observed for Lab1, probably due to the poly A and T tracts at the SNP site. Manual inspection with IGV showed the accurate genotype (AT) could be called, even if imbalanced.



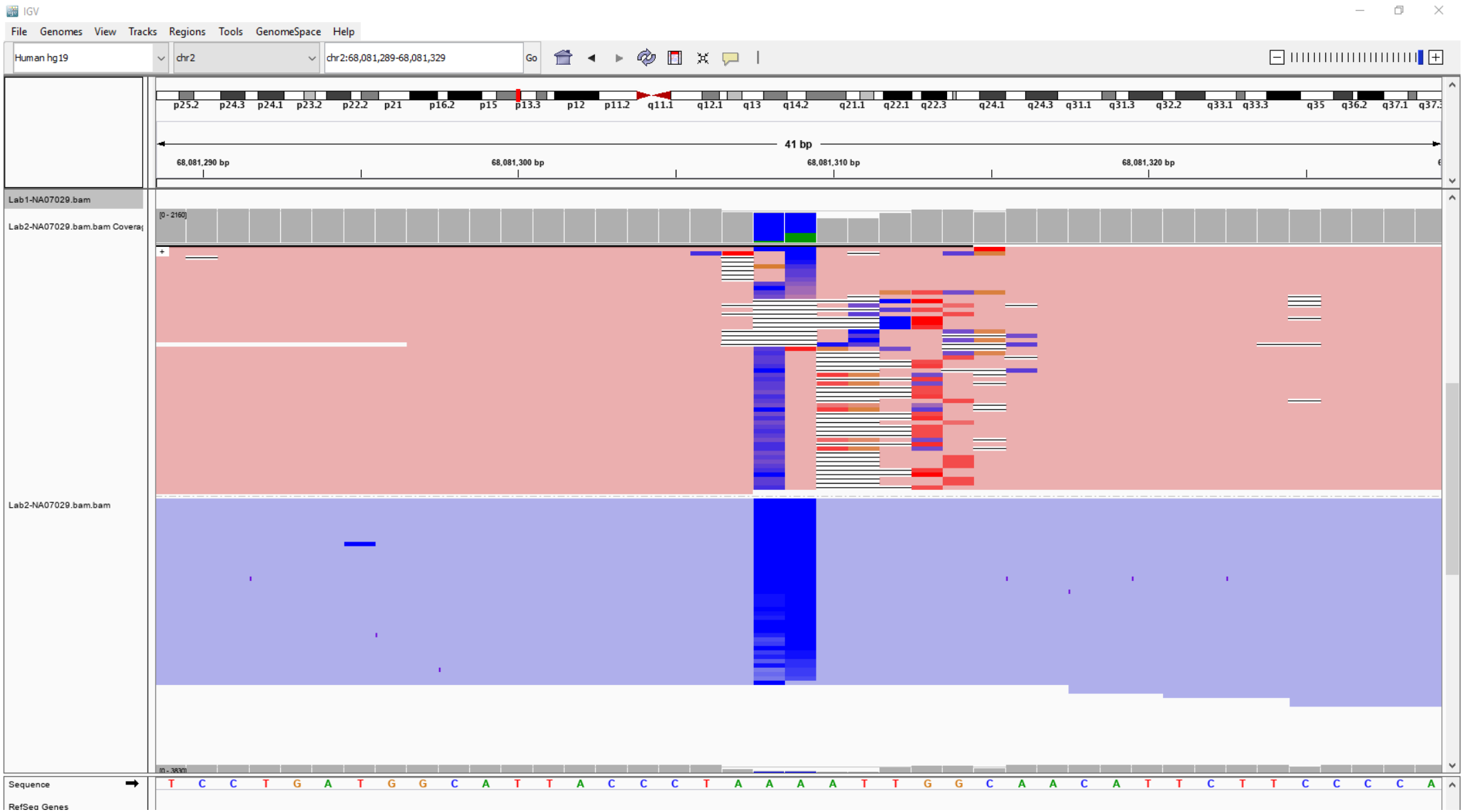
Lab2-NA11200 rs2344704 – AT genotypes were called for all replicates (Labs 1-3). The higher percentage of A reads on the reverse strand (low quality reads, lighter shade) could result due to the poly-A tract downstream of the SNP. Due to the presence of three poly tracts near the SNP and alignment difficulties, this SNP was removed from the final panel.



Lab2-NA11200 M130 – A no call was observed for this position. The manual inspection using IGV shown 20 C reads at this SNP site.



Lab2-NA07029 rs745977 – Misaligned reads due to poly-A tract at the SNP site lead to the erroneous incorporation of A reads (see reverse strand reads).



Lab2-NA07029 rs10928235 – A no call (NN) by the HID-SNP-Genotype was obtained for two replicates of the Labs 2-3. A manual inspection with IGV showed the presence of A reads at the SNP position. The no call might have resulted from the presence of a poly-A and a poly-T tract at the SNP site that prevent the sequencing of the forward reads.