DEMULTIPLEXING:

Total reads:	564,660
Sample1	7,799
Sample2	4,957
Sample3	7,315
Sample4	4,258
Sample5	6,762
Sample6	7,914
Sample7	5,120
Sample8	7,666
Sample9	7,587
Sample10	7,368
Sample11	6,185
Sample12	6,380
Sample13	5,025
Sample14	8,302
Sample15	5,132
Sample16	4,262
Sample17	3,668
Sample18	3,862
Sample19	4,876
Sample20	4,982
Sample21	7,095
Sample22	3,755
Sample23	4,934
Sample24	8,094
Sample25	3,186
Sample26	4,825
Sample27	9,032
Sample28	7,536
Sample29	5,208
Sample30	6,693

Total reads after demultiplexing:
179,778

Unfortunately, the index sequences are simply not present in most of the reads

But there are still 5-9K reads

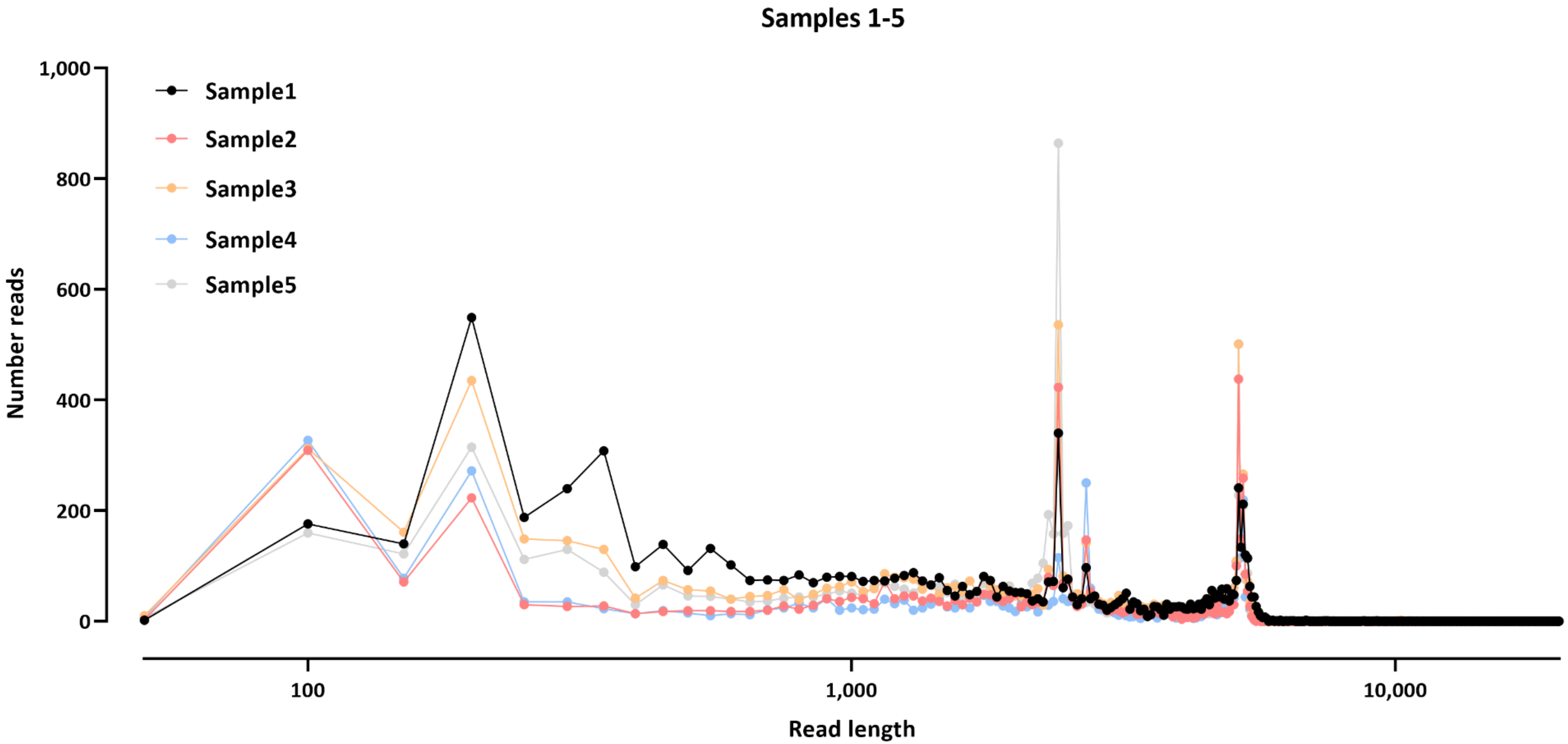
ANNOTATIONS

The next question is what annotation to use, as the different available annotations for the human genome have a different number of isoforms

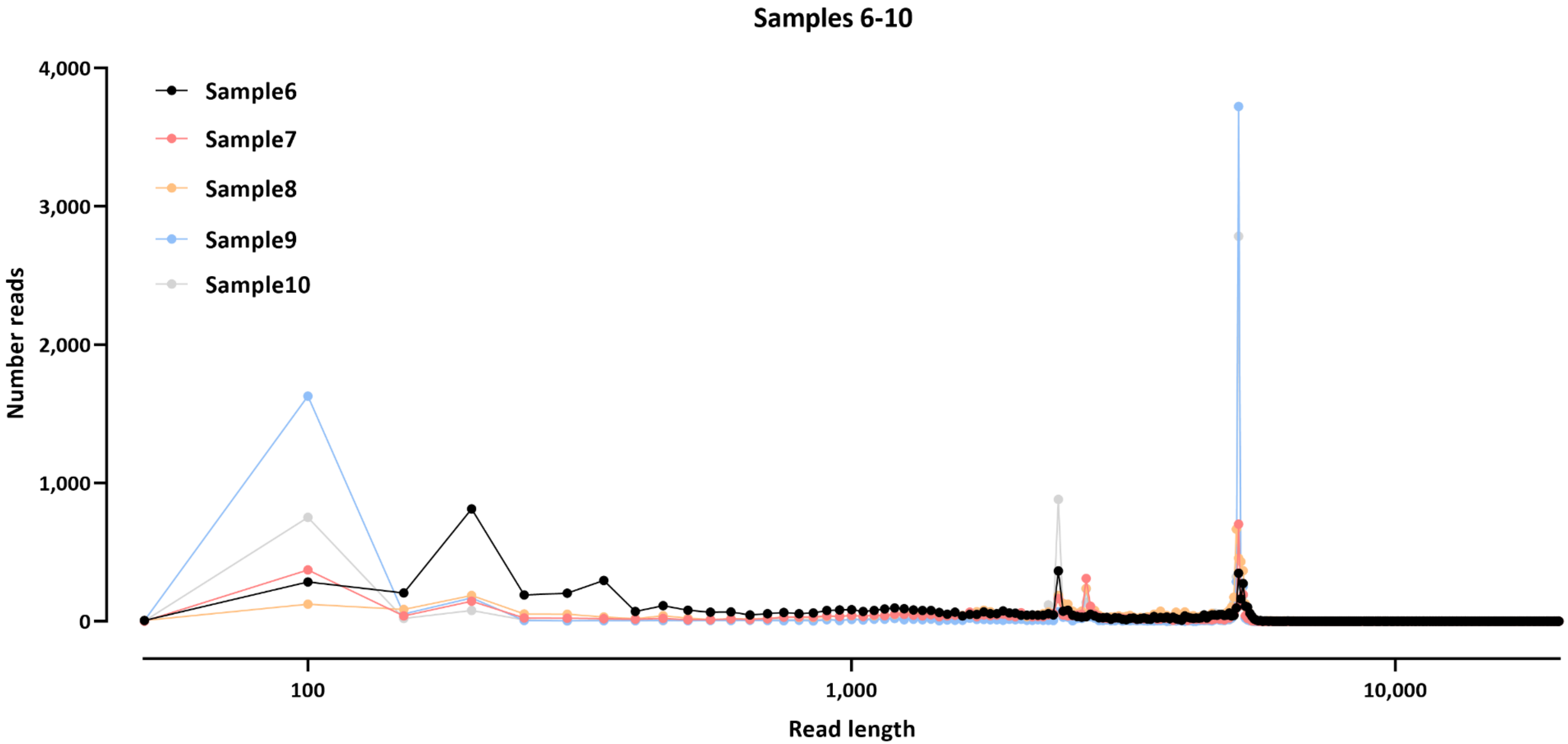
See further below for more details

I ran it on the three annotations that I have, but we can quickly rerun on anything else

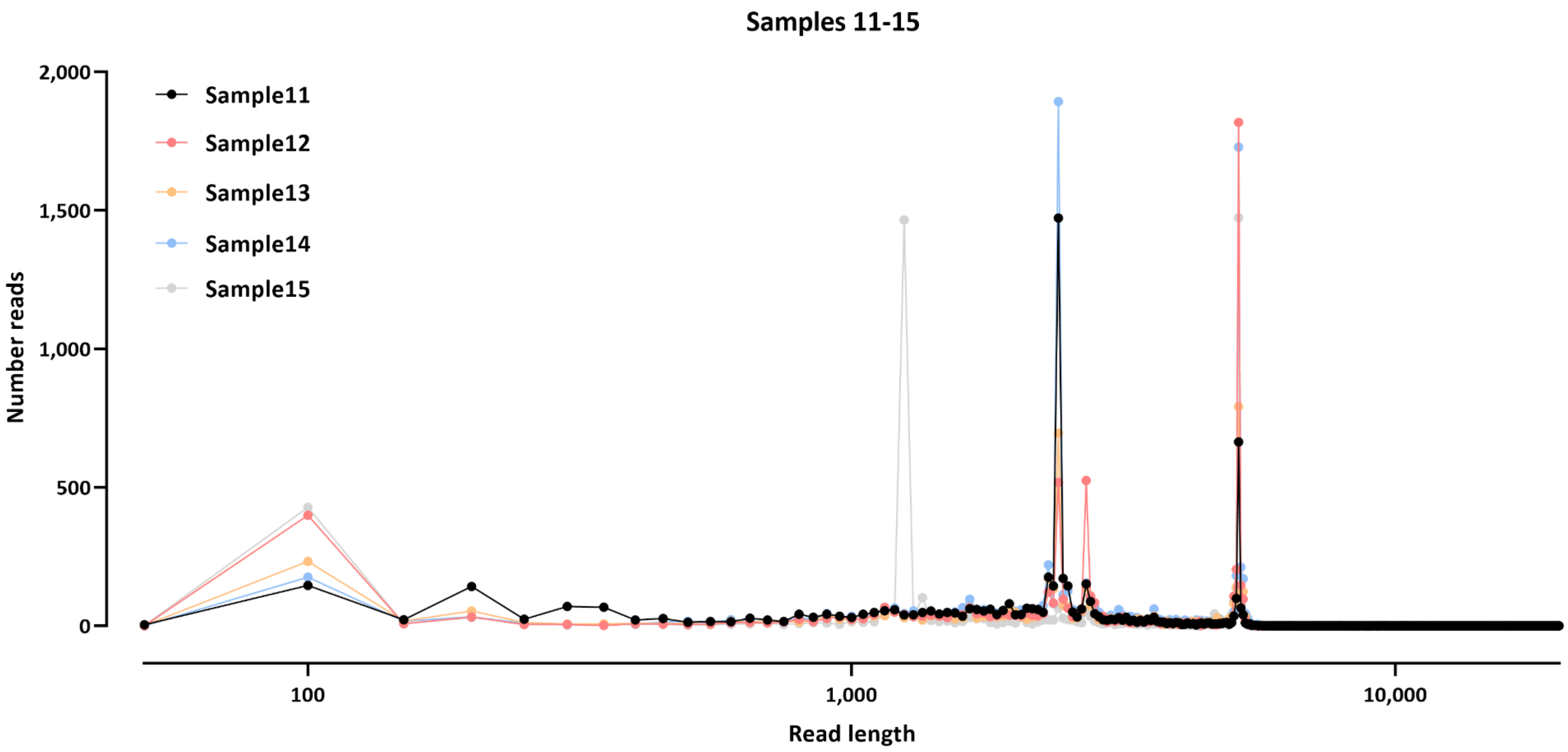
READ LENGTH DISTRIBUTION



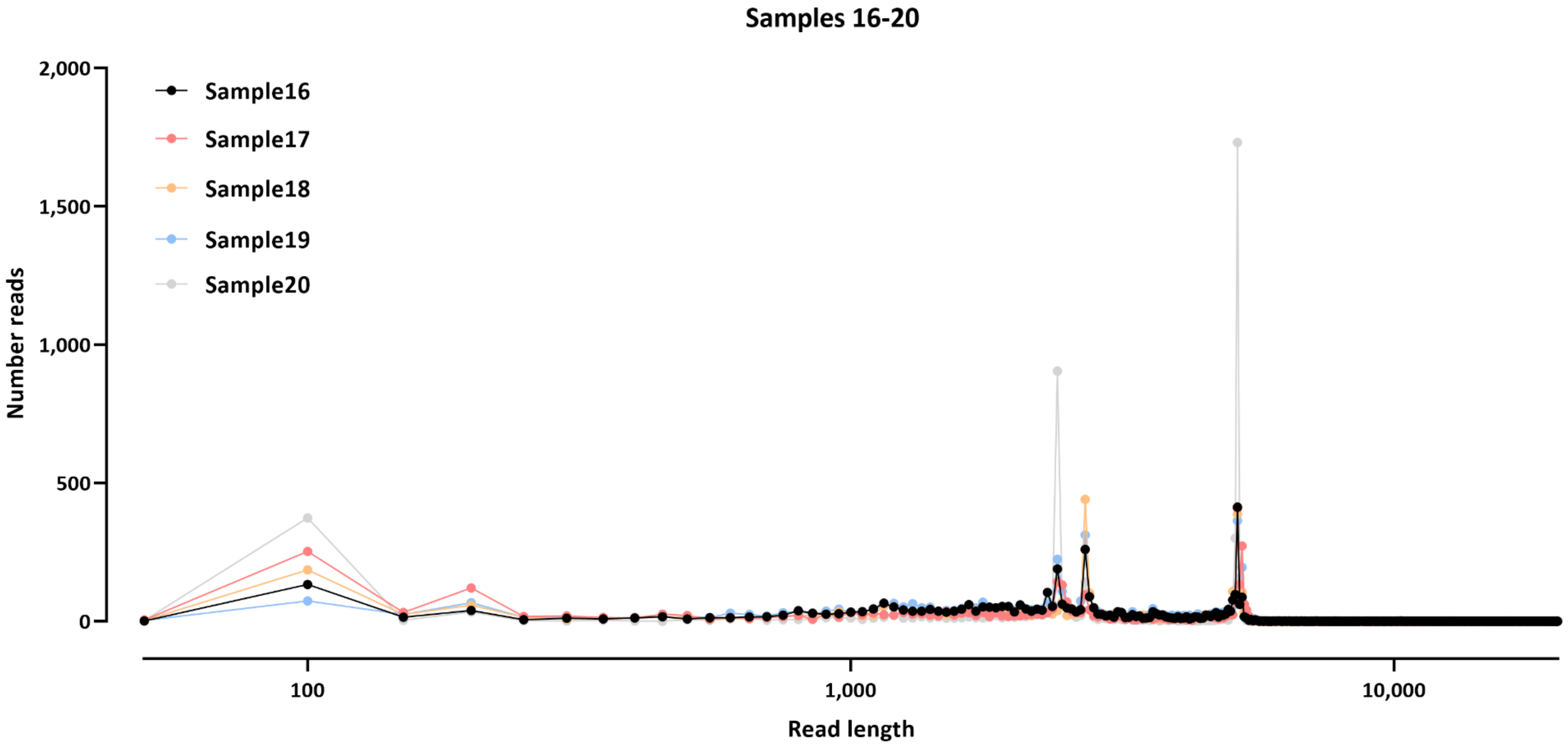
READ LENGTH DISTRIBUTION



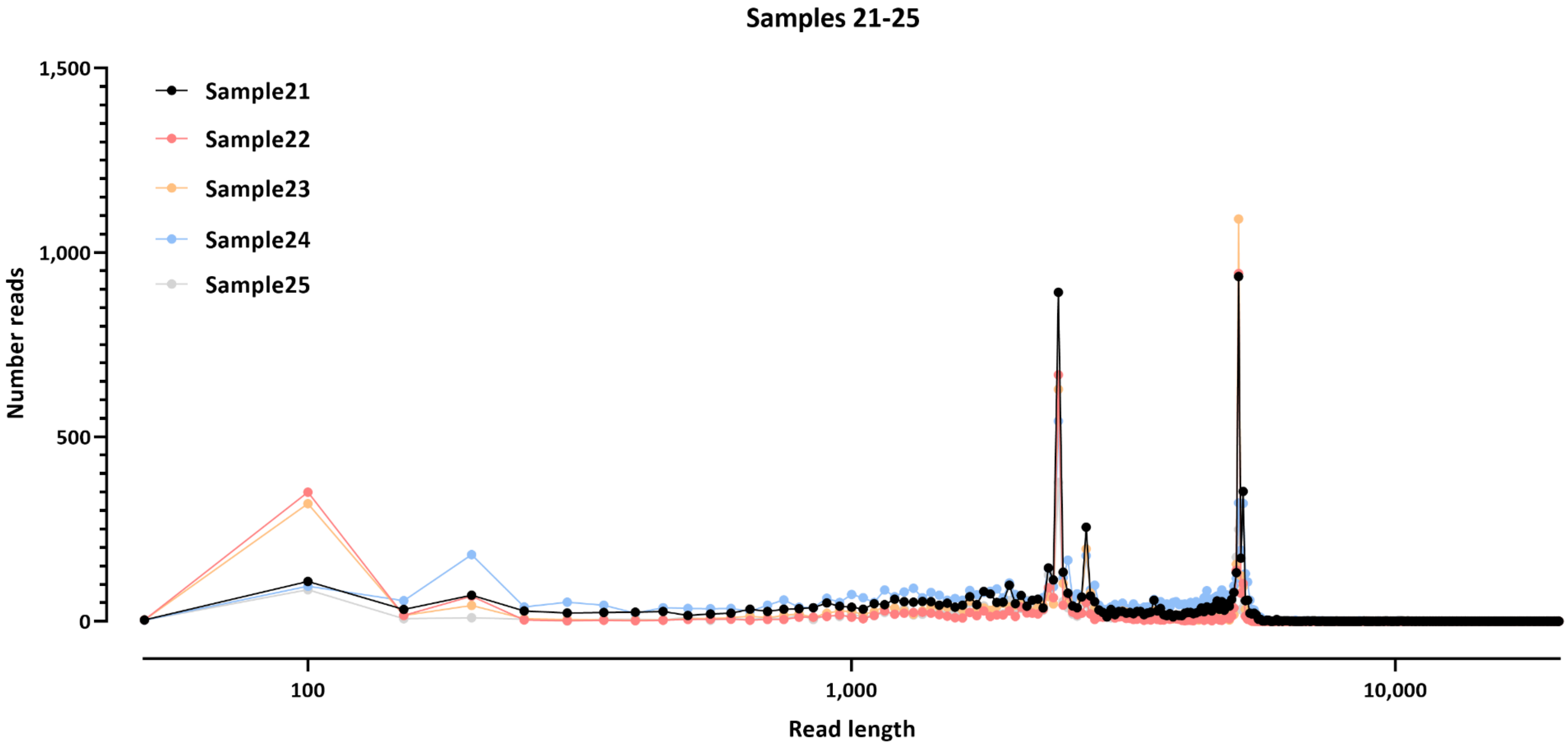
READ LENGTH DISTRIBUTION



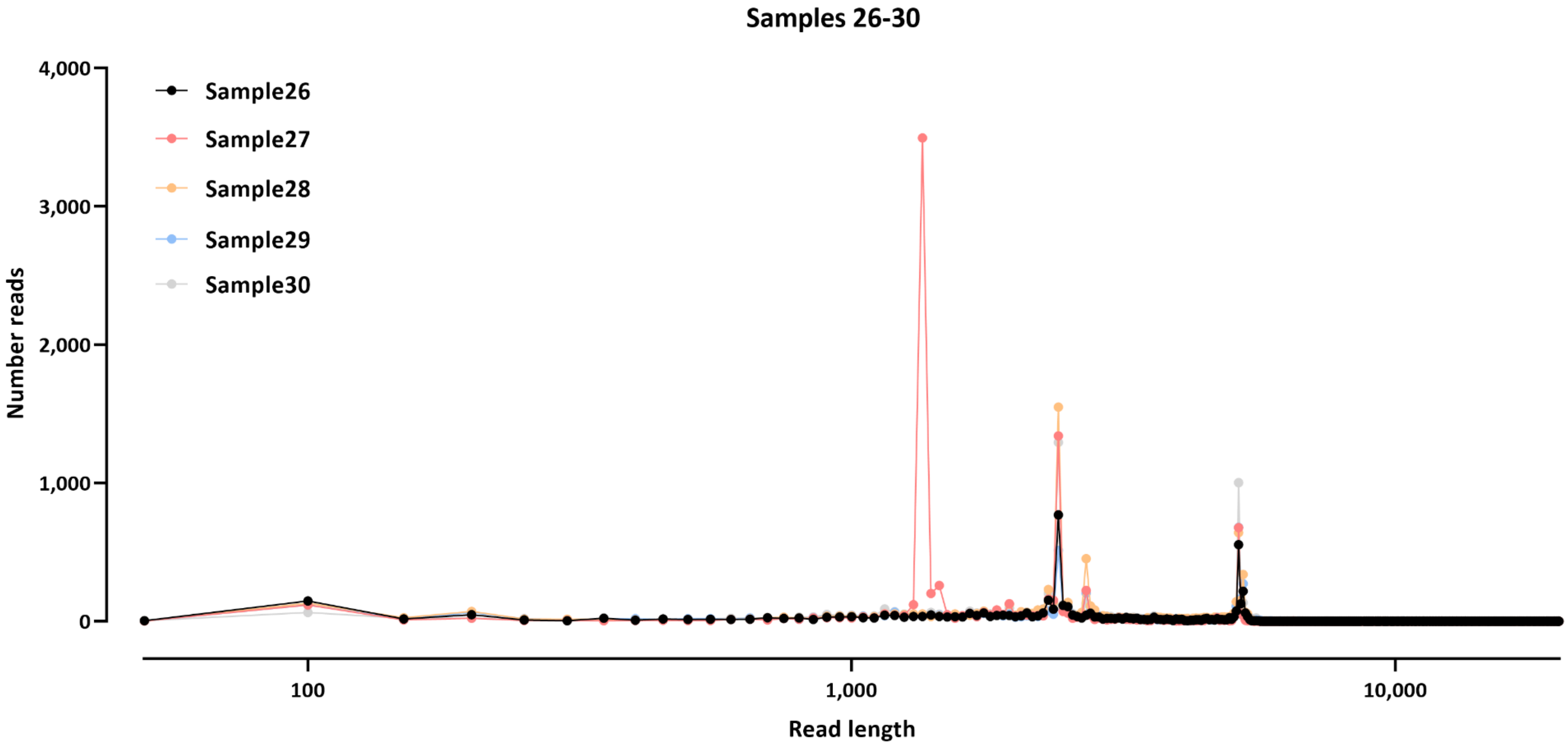
READ LENGTH DISTRIBUTION



READ LENGTH DISTRIBUTION



READ LENGTH DISTRIBUTION



CACNA1C ISOFORMS IN GENCODE

Isofrom	transcript length
CACNA1C-201	6,616
CACNA1C-202	6,968
CACNA1C-203	6,634
CACNA1C-204	6,655
CACNA1C-205	6,535
CACNA1C-206	6,535
CACNA1C-207	6,511
CACNA1C-208	6,511
CACNA1C-209	7,377
CACNA1C-210	6,571
CACNA1C-211	7,482
CACNA1C-212	6,568
CACNA1C-213	6,562
CACNA1C-214	7,590
CACNA1C-215	6,568
CACNA1C-216	6,595
CACNA1C-217	6,511
CACNA1C-218	6,511
CACNA1C-219	6,529
CACNA1C-220	13,433
CACNA1C-221	6,568
CACNA1C-222	7,590
CACNA1C-223	831
CACNA1C-224	581
CACNA1C-225	6,811
CACNA1C-226	592
CACNA1C-227	669
CACNA1C-228	819
CACNA1C-229	545
CACNA1C-230	577
CACNA1C-231	2,833

CACNA1C ISOFORMS IN REFSEQ 2012

Isofrom	transcript length
CACNA1C_dup9	13,531
CACNA1C_dup10	13,480
CACNA1C_dup12	13,480
CACNA1C_dup17	13,504
CACNA1C_dup13	13,537
CACNA1C_dup7	13,624
CACNA1C_dup18	13,564
CACNA1C_dup11	13,447
CACNA1C-AS4	737
CACNA1C_dup22	13,550
CACNA1C_dup2	13,480
CACNA1C_dup8	13,504
CACNA1C-AS2	609
CACNA1C_dup5	13,537
CACNA1C_dup14	13,660
CACNA1C_dup6	13,471
CACNA1C-IT2	588
CACNA1C_dup20	13,694
CACNA1C	13,480
CACNA1C_dup19	13,480
CACNA1C_dup4	13,540
CACNA1C-AS1	2,230
CACNA1C_dup1	13,498
CACNA1C_dup15	13,480
CACNA1C_dup3	13,537
CACNA1C-IT3	430
CACNA1C_dup21	13,550
CACNA1C_dup16	13,603

CACNA1C ISOFORMS IN REFSEQ 2022

Isofrom	transcript length
NM_000719.7_2	13,744
NM_001129827.2_2	13,888
NM_001129829.2_2	13,867
NM_001129830.3_2	13,849
NM_001129831.2_2	13,828
NM_001129832.2_2	13,804
NM_001129833.2_2	13,801
NM_001129834.2_2	13,801
NM_001129835.2_2	13,801
NM_001129836.2_2	13,795
NM_001129837.2_2	13,768
NM_001129838.2_2	13,768
NM_001129839.2_2	13,762
NM_001129840.2_2	13,744
NM_001129841.2_2	13,744
NM_001129842.2_2	13,744
NM_001129843.2_2	13,744
NM_001129844.2_2	13,735
NM_001129846.2_2	13,711
NM_001167623.2_2	13,744
NM_001167624.3_2	13,849
NM_001167625.2_2	13,924
NM_199460.4_2	13,993
NR_045725.1_2	2,230
NR_046578.1_2	737
NR_046579.1_2	609
NR_046768.1_2	589
NR_046769.1_2	430

XM_006719017.2	13,531
XM_011521020.2	13,606
XM_011521023.3	13,517
XM_017019926.2	15,497
XM_017019927.2	15,392
XM_017019928.2	15,179
XM_017019929.2	15,122
XM_017019930.2	8,375
XM_017019931.2	8,291
XM_017019932.2	9,236
XM_017019933.2	8,180
XM_017019934.2	8,171
XM_017019935.2	14,918
XM_017019936.2	8,153
XM_017019937.2	8,153
XM_017019938.2	8,153
XM_017019939.2	8,147
XM_017019940.2	8,120
XM_017019941.2	8,114
XM_017019942.2	8,096
XM_017019943.2	8,096
XM_017019944.2	8,096
XM_017019945.2	8,096
XM_017019946.2	9,056
XM_017019947.2	8,096
XM_017019948.2	8,096
XM_017019949.2	8,096
XM_017019950.2	8,090
XM_017019951.2	8,087
XM_017019952.2	8,063
XM_017019953.1	13,531
XM_017019954.1	13,531
XM_017019955.2	14,918
XR_001749433.1	8,374

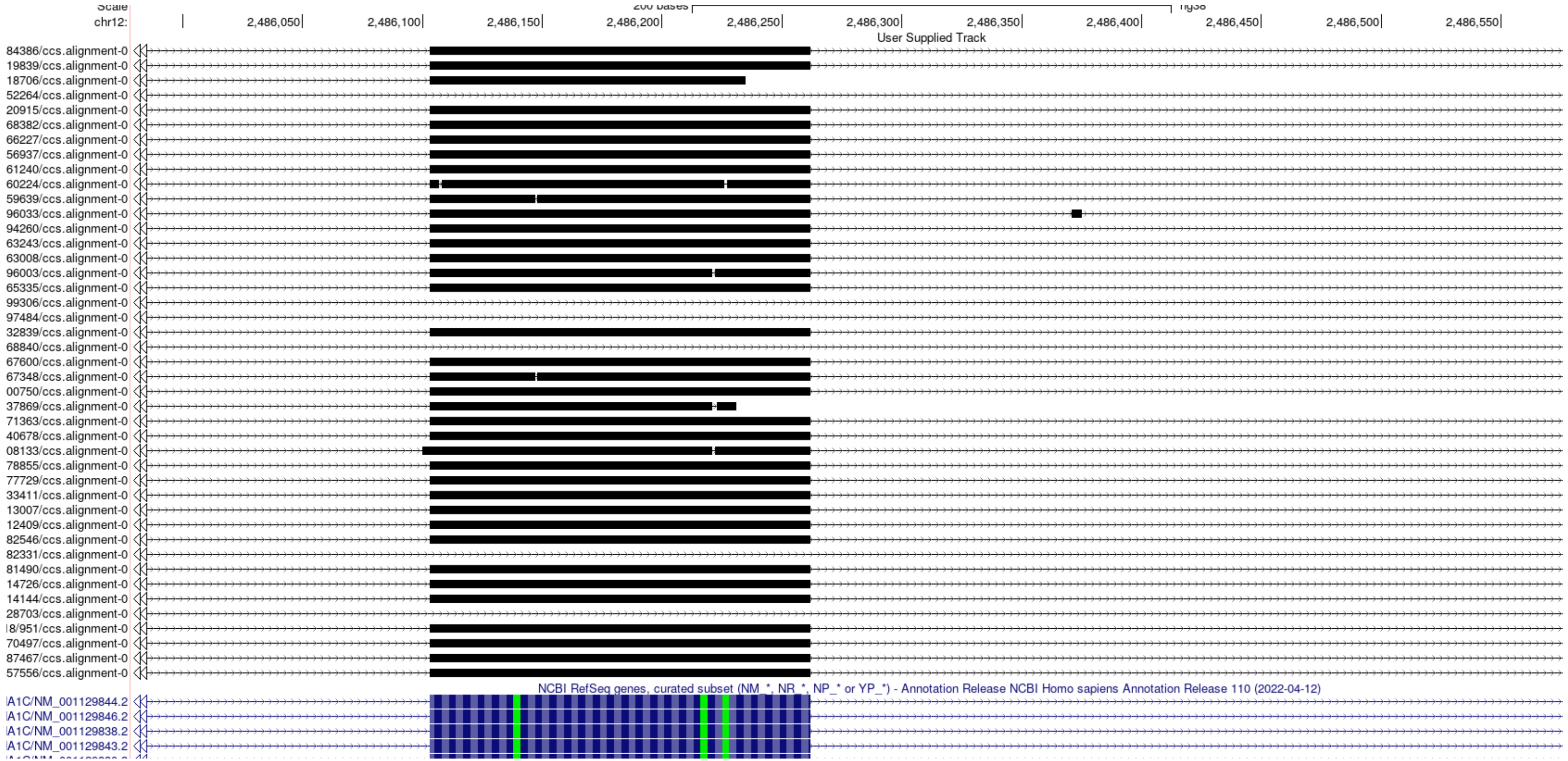
ALIGNMENT

Did three different alignments:

1) BLAT
2) Minimaps2
3) GMAP

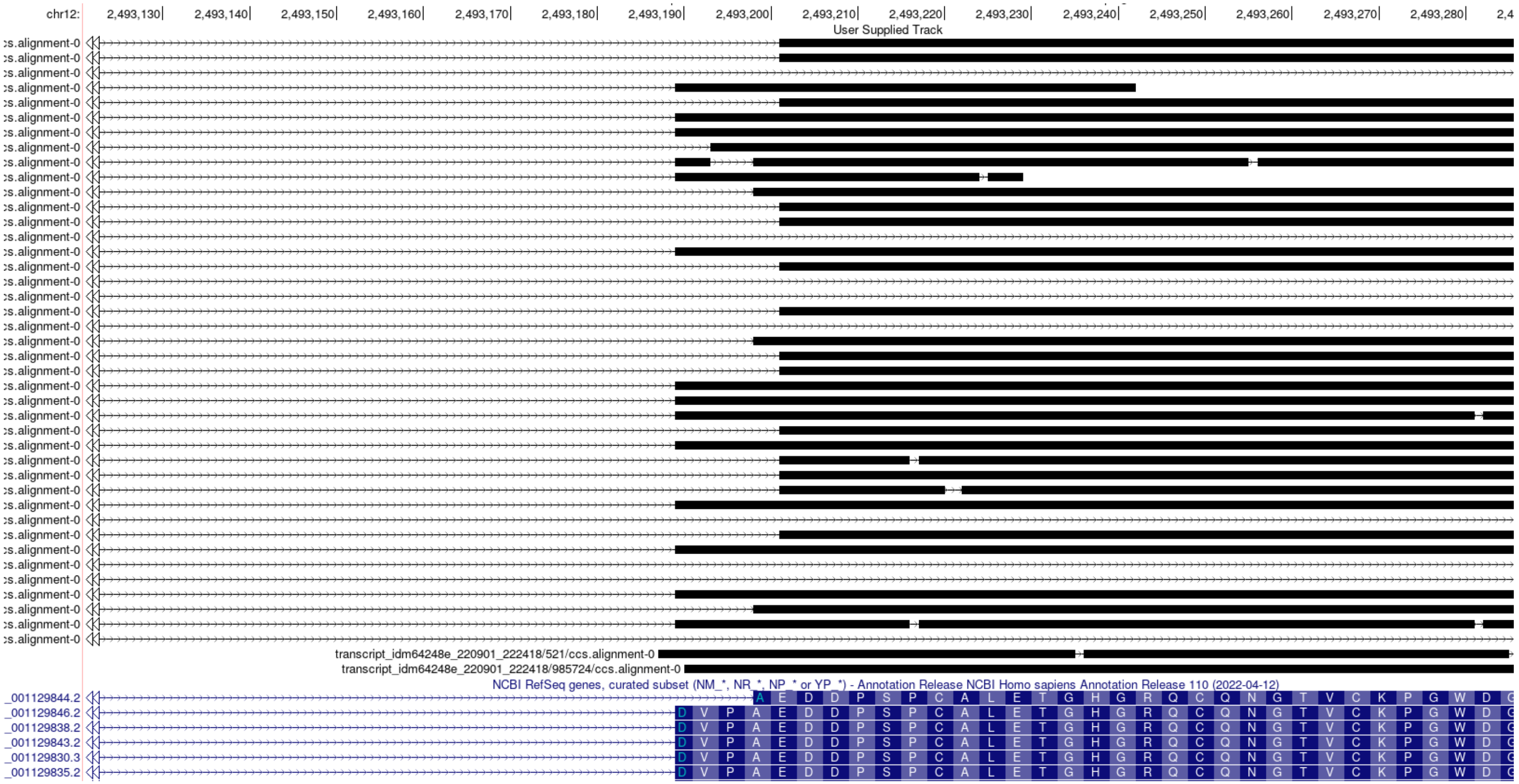
BLAT ALIGNMENT (FIRST 100 READS FROM SAMPLE1)

ALIGNMENT ISSUE #1 (BLAT)



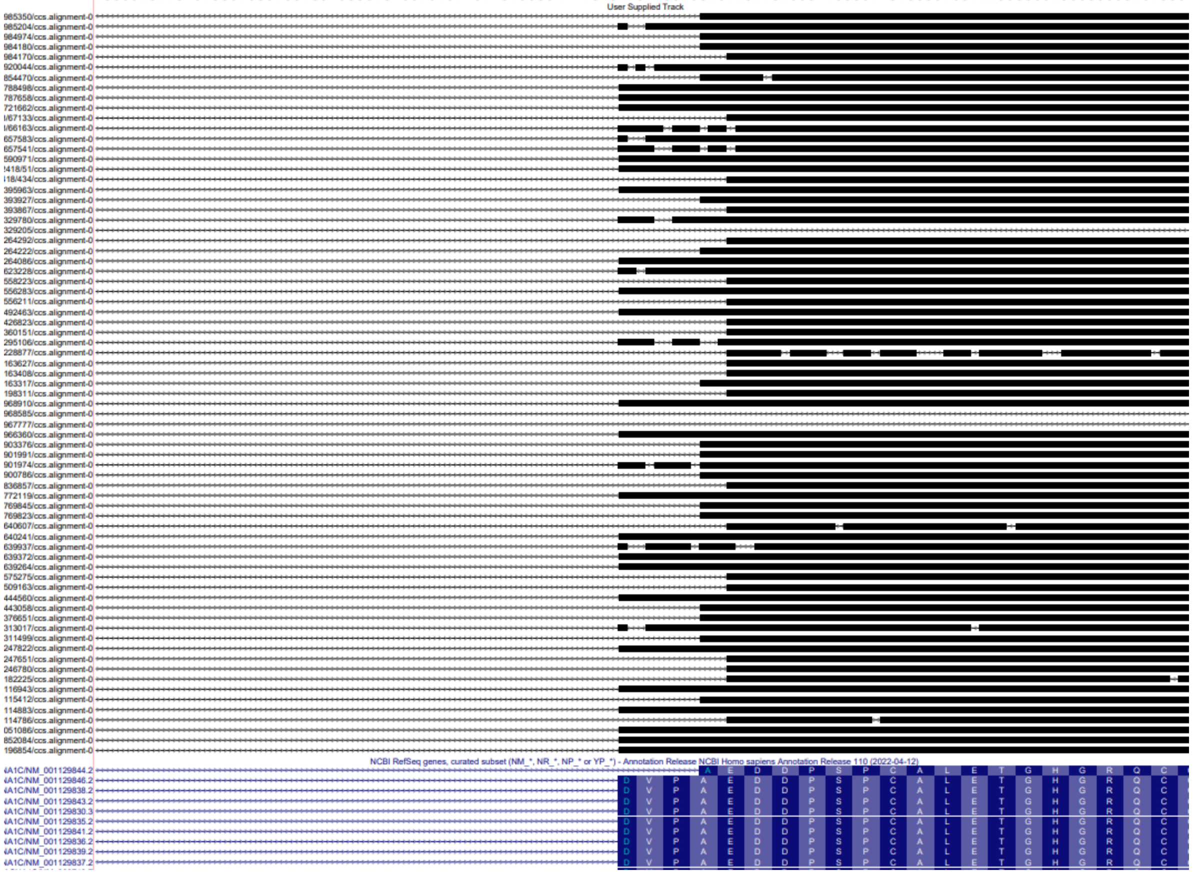
BLAT exhibits a tendency to create microexons for this gene

ALIGNMENT ISSUE #2 (BLAT)



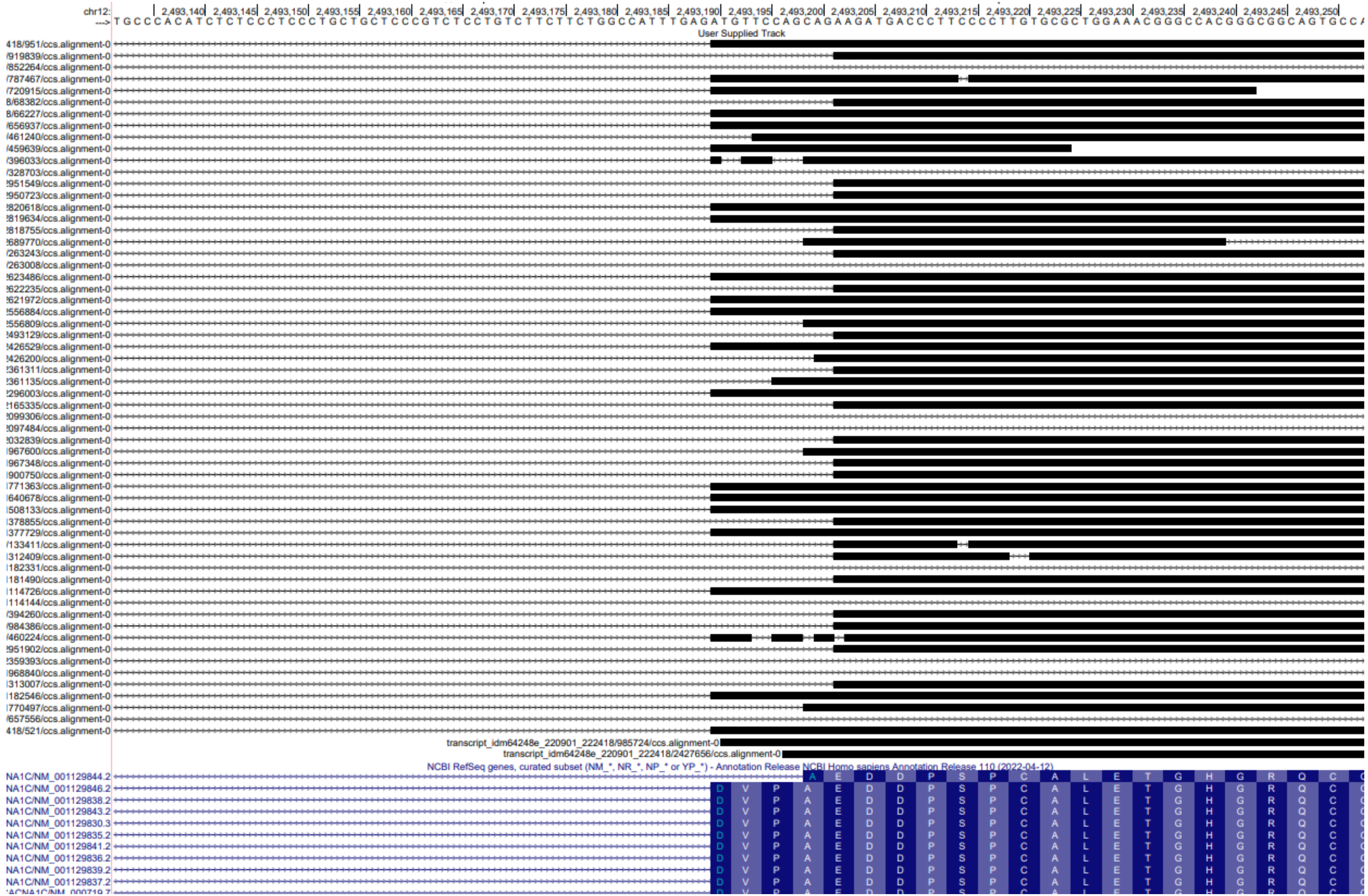
Second, there is an alternative 3’ site on this exon, and the alignments don’t get it always right

ALIGNMENT ISSUE #2 (MINIMAP2)



Same issue with Minimap (though Minimap does not create spurious microexons, which is a plus)

ALIGNMENT ISSUE #2 (GMAP)



QUANTIFICATION SUMMARY:

Total reads: 564,660

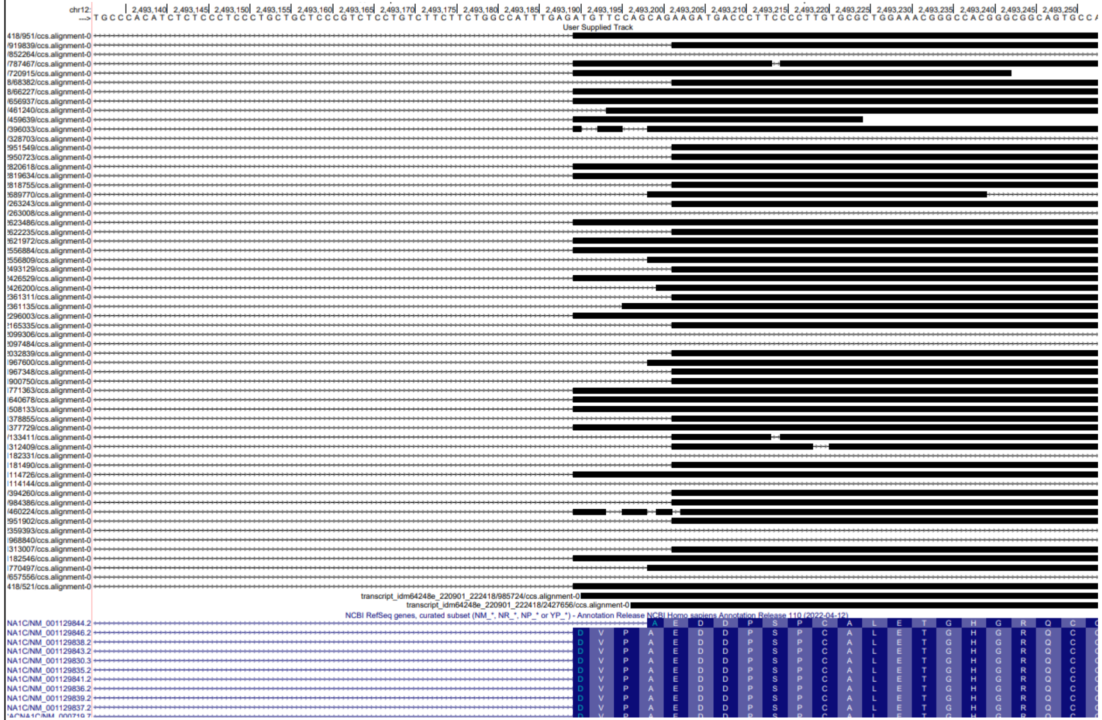
Total reads after demultiplexing:
179,778

		Assigned reads								
		GMAP			BLAT			Minimap2		
		2016-11-16-refSeq	2021-12-08-RefSeq	gencode.v30	2016-11-16-refSeq	2021-12-08-RefSeq	gencode.v30	2016-11-16-refSeq	2021-12-08-RefSeq	gencode.v30
Sample1	7,799	0.44	0.45	0.44	0.46	0.47	0.46	0.58	0.60	0.57
Sample2	4,957	0.35	0.36	0.35	0.34	0.35	0.34	0.47	0.50	0.47
Sample3	7,315	0.39	0.40	0.39	0.40	0.41	0.40	0.51	0.53	0.50
Sample4	4,258	0.39	0.40	0.38	0.40	0.42	0.40	0.47	0.51	0.46
Sample5	6,762	0.29	0.30	0.29	0.36	0.37	0.35	0.40	0.43	0.39
Sample6	7,914	0.46	0.47	0.46	0.51	0.52	0.50	0.55	0.58	0.55
Sample7	5,120	0.36	0.37	0.35	0.32	0.32	0.31	0.46	0.47	0.44
Sample8	7,666	0.25	0.26	0.25	0.23	0.24	0.23	0.36	0.37	0.35
Sample9	7,587	0.33	0.33	0.33	0.32	0.32	0.32	0.38	0.39	0.38
Sample10	7,368	0.21	0.21	0.21	0.26	0.27	0.25	0.32	0.35	0.31
Sample11	6,185	0.25	0.26	0.25	0.31	0.33	0.31	0.39	0.41	0.38
Sample12	6,380	0.27	0.28	0.27	0.27	0.29	0.27	0.34	0.38	0.34
Sample13	5,025	0.24	0.24	0.24	0.22	0.22	0.22	0.37	0.39	0.36
Sample14	8,302	0.18	0.19	0.18	0.19	0.19	0.19	0.32	0.33	0.29
Sample15	5,132	0.18	0.18	0.18	0.17	0.17	0.17	0.21	0.22	0.21
Sample16	4,262	0.33	0.35	0.33	0.31	0.32	0.31	0.40	0.43	0.40
Sample17	3,668	0.28	0.29	0.28	0.29	0.31	0.29	0.36	0.39	0.35
Sample18	3,862	0.40	0.42	0.39	0.35	0.37	0.35	0.45	0.48	0.45
Sample19	4,876	0.32	0.33	0.32	0.29	0.30	0.29	0.41	0.43	0.40
Sample20	4,982	0.15	0.16	0.15	0.15	0.15	0.15	0.31	0.31	0.29
Sample21	7,095	0.23	0.24	0.23	0.23	0.23	0.22	0.38	0.40	0.36
Sample22	3,755	0.21	0.22	0.21	0.22	0.23	0.22	0.35	0.36	0.34
Sample23	4,934	0.25	0.25	0.25	0.23	0.24	0.23	0.39	0.40	0.36
Sample24	8,094	0.28	0.29	0.27	0.27	0.29	0.27	0.38	0.40	0.37
Sample25	3,186	0.19	0.20	0.19	0.18	0.19	0.18	0.26	0.28	0.26
Sample26	4,825	0.22	0.23	0.22	0.25	0.26	0.25	0.35	0.36	0.34
Sample27	9,032	0.11	0.11	0.11	0.12	0.13	0.12	0.19	0.20	0.19
Sample28	7,536	0.24	0.25	0.23	0.23	0.24	0.23	0.38	0.40	0.36
Sample29	5,208	0.26	0.27	0.26	0.26	0.27	0.25	0.39	0.42	0.38
Sample30	6,693	0.19	0.20	0.19	0.19	0.20	0.19	0.34	0.36	0.32

CAVEAT:

Most reads are in fact not uniquely assigned because the gene is very long and the reads do not fully cover it, so if the read does not cover some of the exons that make an isoform unique, then it can only be assigned provisionally to two isoforms.

BACK TO THE PROBLEMATIC EXON:



Most reads are actually lost because of the mess here.

Most transcripts have this exon starting at position 2,493,190, but there are reads mapping to 2,493,196, 2,493,199, and 2,493,202

GENCODE does not have this at all

I made new transcripts with these coordinates and requantified

QUANTIFICATION SUMMARY, NEW ANNOTATIONS

Total reads: 564,660

Total reads after demultiplexing: 179,778

		Assigned reads					
		GMAP		BLAT		Minimap2	
		2016-11-16-refSeq	gencode.v30	2016-11-16-refSeq	gencode.v30	2016-11-16-refSeq	gencode.v30
Sample1	7,799	0.45	0.45	0.49	0.49	0.68	0.68
Sample2	4,957	0.35	0.35	0.37	0.37	0.59	0.60
Sample3	7,315	0.40	0.40	0.43	0.43	0.63	0.63
Sample4	4,258	0.39	0.39	0.45	0.45	0.56	0.56
Sample5	6,762	0.30	0.30	0.44	0.45	0.53	0.54
Sample6	7,914	0.47	0.47	0.57	0.57	0.65	0.65
Sample7	5,120	0.36	0.36	0.33	0.33	0.56	0.56
Sample8	7,666	0.27	0.26	0.25	0.25	0.52	0.51
Sample9	7,587	0.33	0.33	0.34	0.34	0.53	0.53
Sample10	7,368	0.22	0.22	0.33	0.33	0.46	0.46
Sample11	6,185	0.26	0.26	0.43	0.44	0.57	0.57
Sample12	6,380	0.27	0.27	0.33	0.34	0.45	0.46
Sample13	5,025	0.25	0.24	0.24	0.24	0.49	0.50
Sample14	8,302	0.19	0.19	0.23	0.22	0.54	0.54
Sample15	5,132	0.18	0.18	0.17	0.17	0.29	0.29
Sample16	4,262	0.33	0.33	0.33	0.33	0.46	0.47
Sample17	3,668	0.29	0.29	0.33	0.33	0.47	0.47
Sample18	3,862	0.40	0.40	0.38	0.38	0.52	0.52
Sample19	4,876	0.33	0.33	0.31	0.31	0.53	0.53
Sample20	4,982	0.16	0.16	0.18	0.18	0.47	0.47
Sample21	7,095	0.24	0.24	0.25	0.25	0.52	0.52
Sample22	3,755	0.22	0.22	0.25	0.25	0.50	0.51
Sample23	4,934	0.25	0.25	0.26	0.26	0.52	0.53
Sample24	8,094	0.29	0.29	0.30	0.30	0.50	0.51
Sample25	3,186	0.20	0.20	0.23	0.23	0.40	0.41
Sample26	4,825	0.23	0.23	0.34	0.34	0.53	0.54
Sample27	9,032	0.11	0.11	0.16	0.16	0.28	0.29
Sample28	7,536	0.24	0.24	0.27	0.27	0.54	0.54
Sample29	5,208	0.26	0.26	0.27	0.27	0.52	0.53
Sample30	6,693	0.20	0.20	0.24	0.24	0.53	0.53