

HOWR-1

Summary

General

fastp version:	0.20.0 (https://github.com/OpenGene/fastp)
sequencing:	paired end (150 cycles + 150 cycles)
mean length before filtering:	150bp, 150bp
mean length after filtering:	139bp, 139bp
duplication rate:	12.360865%
Insert size peak:	150

Before filtering

total reads:	700.631888 M
total bases:	105.094783 G
Q20 bases:	102.521695 G (97.551650%)
Q30 bases:	98.830196 G (94.039107%)
GC content:	53.617147%

After filtering

total reads:	680.948918 M
total bases:	95.200473 G
Q20 bases:	93.657590 G (98.379333%)
Q30 bases:	90.771133 G (95.347355%)
GC content:	53.904791%

Filtering result

reads passed filters:	680.948918 M (97.190683%)
reads with low quality:	15.303720 M (2.184274%)
reads with too many N:	47.110000 K (0.006724%)
reads too short:	3.058662 M (0.436558%)
reads with low complexity:	1.273478 M (0.181761%)

Adapters

Adapter or bad ligation of read1

Sequence	Occurrences
A	2608629
AG	2562751
AGA	2514611
AGAT	2454053
AGATC	2539273

AGATCG	2620379
AGATCGG	2658024
AGATCGGA	2647892
AGATCGGAA	2571383
AGATCGGAAG	2576256
AGATCGGAAGA	2589648
AGATCGGAAGAG	2552533
AGATCGGAAGAGC	2603750
AGATCGGAAGAGCA	2563254
AGATCGGAAGAGCAC	2587161
AGATCGGAAGAGCAC	2508373
AGATCGGAAGAGCACAC	2556080
AGATCGGAAGAGCACACG	2517331
AGATCGGAAGAGCACACGT	2542562
AGATCGGAAGAGCACACGTC	2490694
AGATCGGAAGAGCACACGTCT	2481185
AGATCGGAAGAGCACACGTCTG	2406962
AGATCGGAAGAGCACACGTCTGA	2412035
AGATCGGAAGAGCACACGTCTGAA	2398910
AGATCGGAAGAGCACACGTCTGAAC	2355010
AGATCGGAAGAGCACACGTCTGAAC	2308832
AGATCGGAAGAGCACACGTCTGAAC	2250614
AGATCGGAAGAGCACACGTCTGAAC	2156605
AGATCGGAAGAGCACACGTCTGAAC	2162723
AGATCGGAAGAGCACACGTCTGAAC	2078170
AGATCGGAAGAGCACACGTCTGAAC	2101136
AGATCGGAAGAGCACACGTCTGAAC	2053762
AGATCGGAAGAGCACACGTCTGAAC	2016481
AGATCGGAAGAGCACACGTCTGAAC	1936410
AGATCGGAAGAGCACACGTCTGAAC	1903163
AGATCGGAAGAGCACACGTCTGAAC	1828288
AGATCGGAAGAGCACACGTCTGAAC	1771230
AGATCGGAAGAGCACACGTCTGAAC	1813572
AGATCGGAAGAGCACACGTCTGAAC	1726712
AGATCGGAAGAGCACACGTCTGAAC	1695351
AGATCGGAAGAGCACACGTCTGAAC	1657962
AGATCGGAAGAGCACACGTCTGAAC	1551756
AGATCGGAAGAGCACACGTCTGAAC	1512011
AGATCGGAAGAGCACACGTCTGAAC	1421265
AGATCGGAAGAGCACACGTCTGAAC	1357258
other adapter sequences	32837864

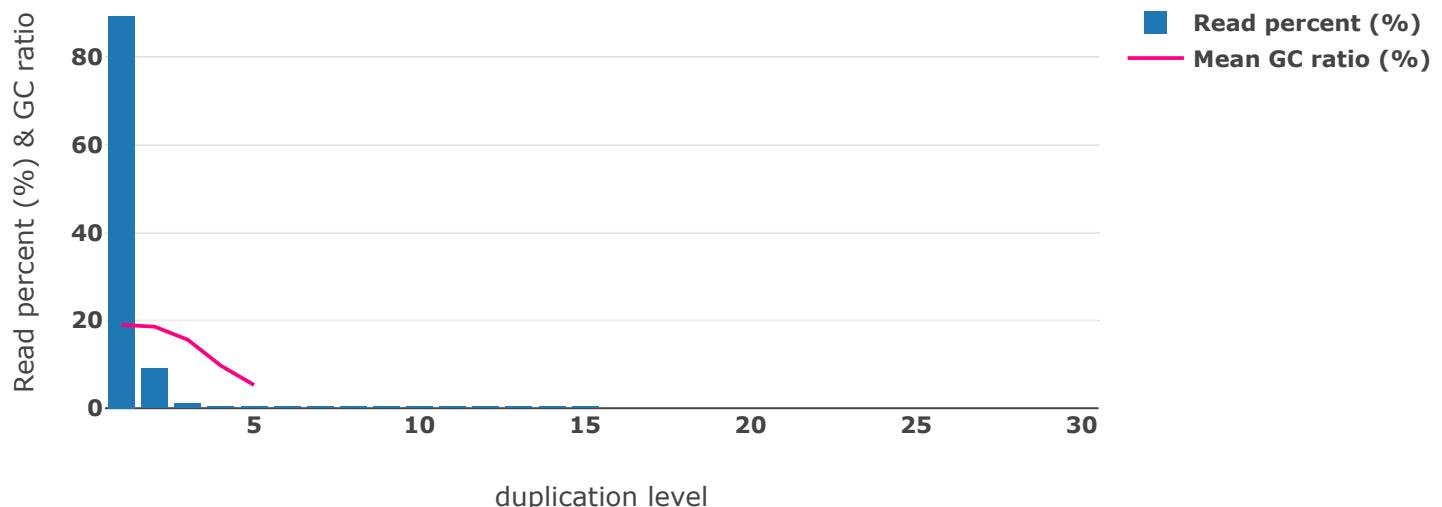
Adapter or bad ligation of read2

Sequence	Occurrences
A	2628992
AG	2583098
AGA	2544990
AGAT	2485040
AGATC	2578691
AGATCG	2603541
AGATCGG	2652394
AGATCGGA	2640287
AGATCGGAA	2555635
AGATCGGAAG	2558984
AGATCGGAAGA	2588428
AGATCGGAAGAG	2553728
AGATCGGAAGAGC	2634257
AGATCGGAAGAGCG	2605604
AGATCGGAAGAGCGT	2556577
AGATCGGAAGAGCGTC	2489142
AGATCGGAAGAGCGTCG	2504938
AGATCGGAAGAGCGTCGT	2424821
AGATCGGAAGAGCGTCGTG	2462370
AGATCGGAAGAGCGTCGTGT	2417558
AGATCGGAAGAGCGTCGTGT	2399977
AGATCGGAAGAGCGTCGTGTAG	2370289
AGATCGGAAGAGCGTCGTGTAGG	2411720
AGATCGGAAGAGCGTCGTGTAGGG	2392761
AGATCGGAAGAGCGTCGTGTAGGGAA	2319263
AGATCGGAAGAGCGTCGTGTAGGGAAAG	2278574
AGATCGGAAGAGCGTCGTGTAGGGAAA	2221458
AGATCGGAAGAGCGTCGTGTAGGGAAAG	2114771
AGATCGGAAGAGCGTCGTGTAGGGAAAGAA	2127562
AGATCGGAAGAGCGTCGTGTAGGGAAAGAG	2051628
AGATCGGAAGAGCGTCGTGTAGGGAAAGAGT	2059130
AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTG	2027982
AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTG	1995252
AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTA	1923806
AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGAG	1890419
AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGAGA	1815330
AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGAGAT	1752768
AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGAGATC	1792731
AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGAGATCT	1710737

AGATCGGAAGAGCGCTGTTAGGGAAAGAGTGTAGATCTC	1681294
AGATCGGAAGAGCGCTGTTAGGGAAAGAGTGTAGATCTCG	1657781
AGATCGGAAGAGCGCTGTTAGGGAAAGAGTGTAGATCTCGG	1524121
AGATCGGAAGAGCGCTGTTAGGGAAAGAGTGTAGATCTCGGT	1484987
AGATCGGAAGAGCGCTGTTAGGGAAAGAGTGTAGATCTCGTG	1426460
AGATCGGAAGAGCGCTGTTAGGGAAAGAGTGTAGATCTCGGTG	1350504
AGATCGGAAGAGCGCTGTTAGGGAAAGAGTGTAGATCTCGGTGATCATTAAAAA	1770778
AGATCGGAAGAGCGCTGTTAGGGAAAGAGTGTAGATCTCGGTGTCGCCGTATCATTAAAAA	1559186
other adapter sequences	29060295

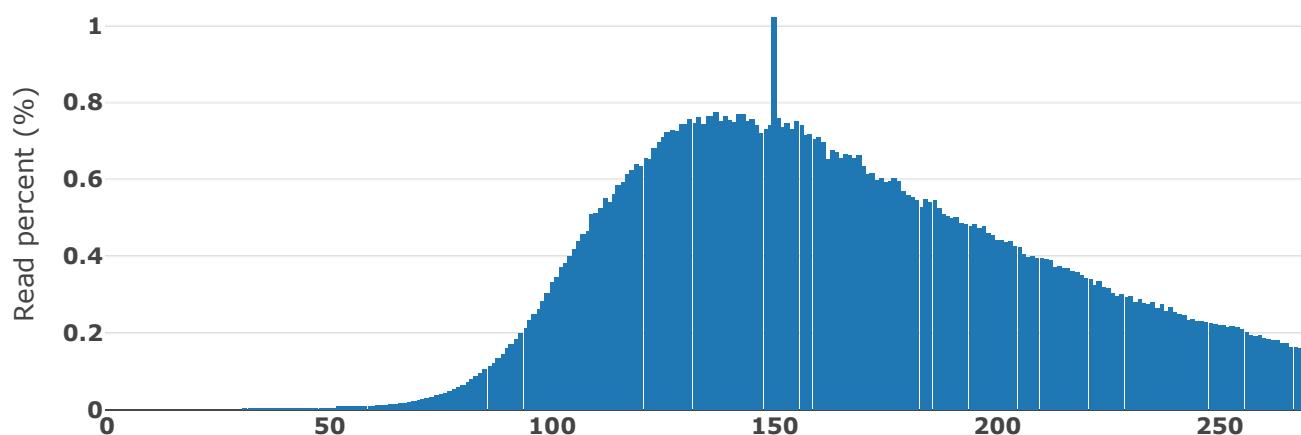
Duplication

duplication rate (12.360865%)



Insert size estimation

Insert size distribution (13.754308% reads are with unknown length)



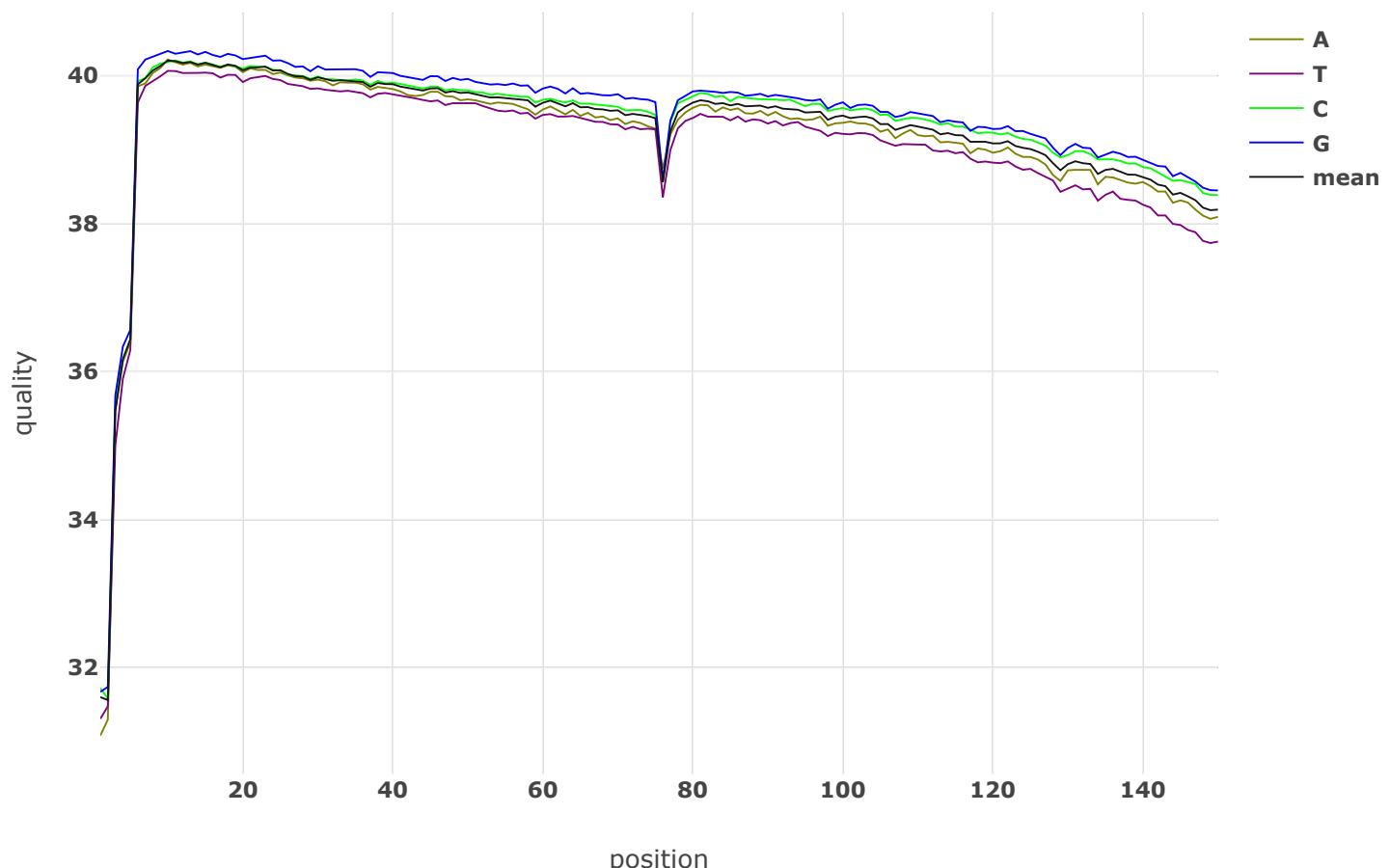
Insert size

This estimation is based on paired-end overlap analysis, and there are 13.754308% reads found not overlapped.
The nonoverlapped read pairs may have insert size <30 or >270, or contain too much sequencing errors to be detected as overlapped.

Before filtering

Before filtering: read1: quality

Value of each position will be shown on mouse over.



Before filtering: read1: base contents

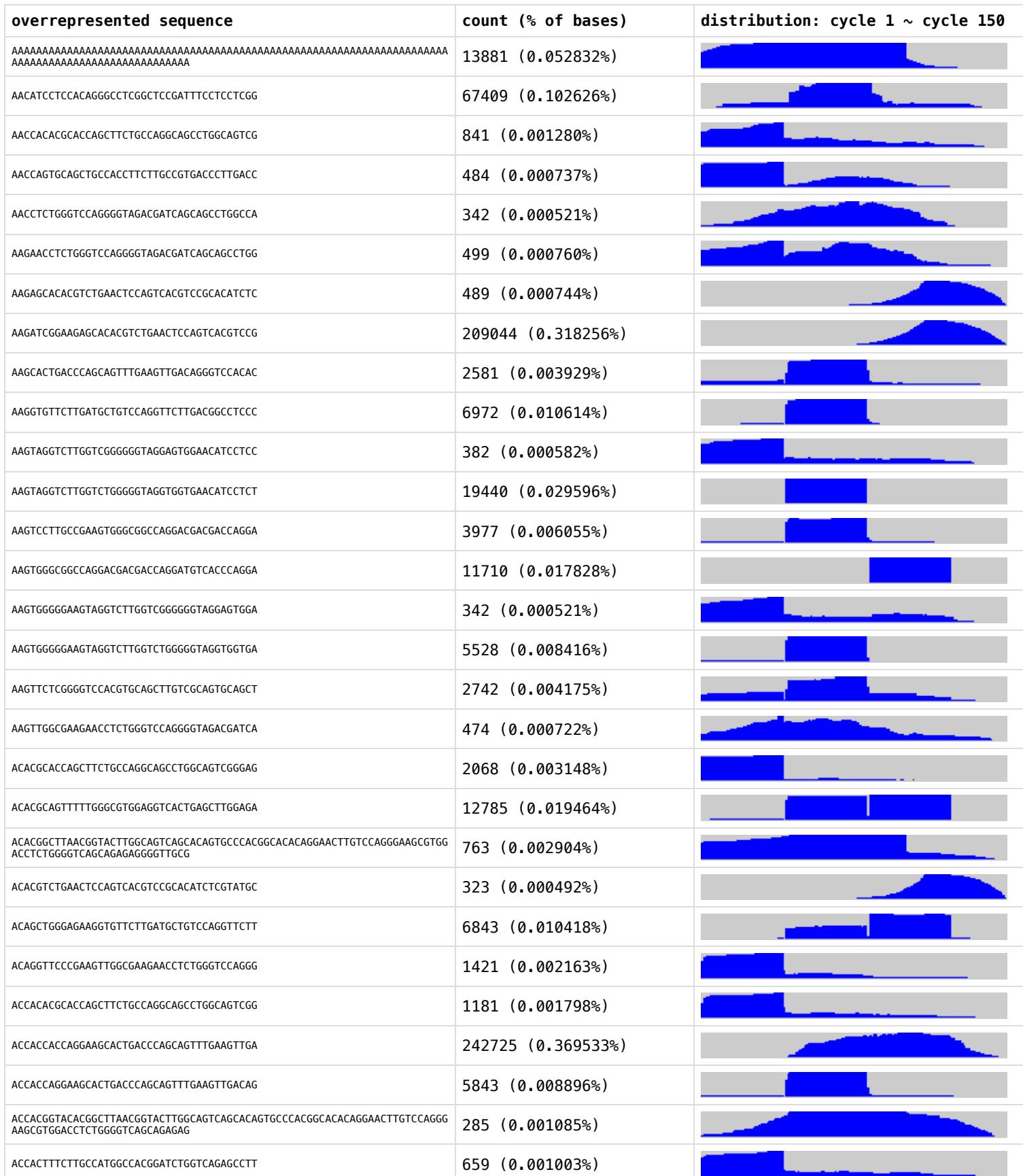
Value of each position will be shown on mouse over.

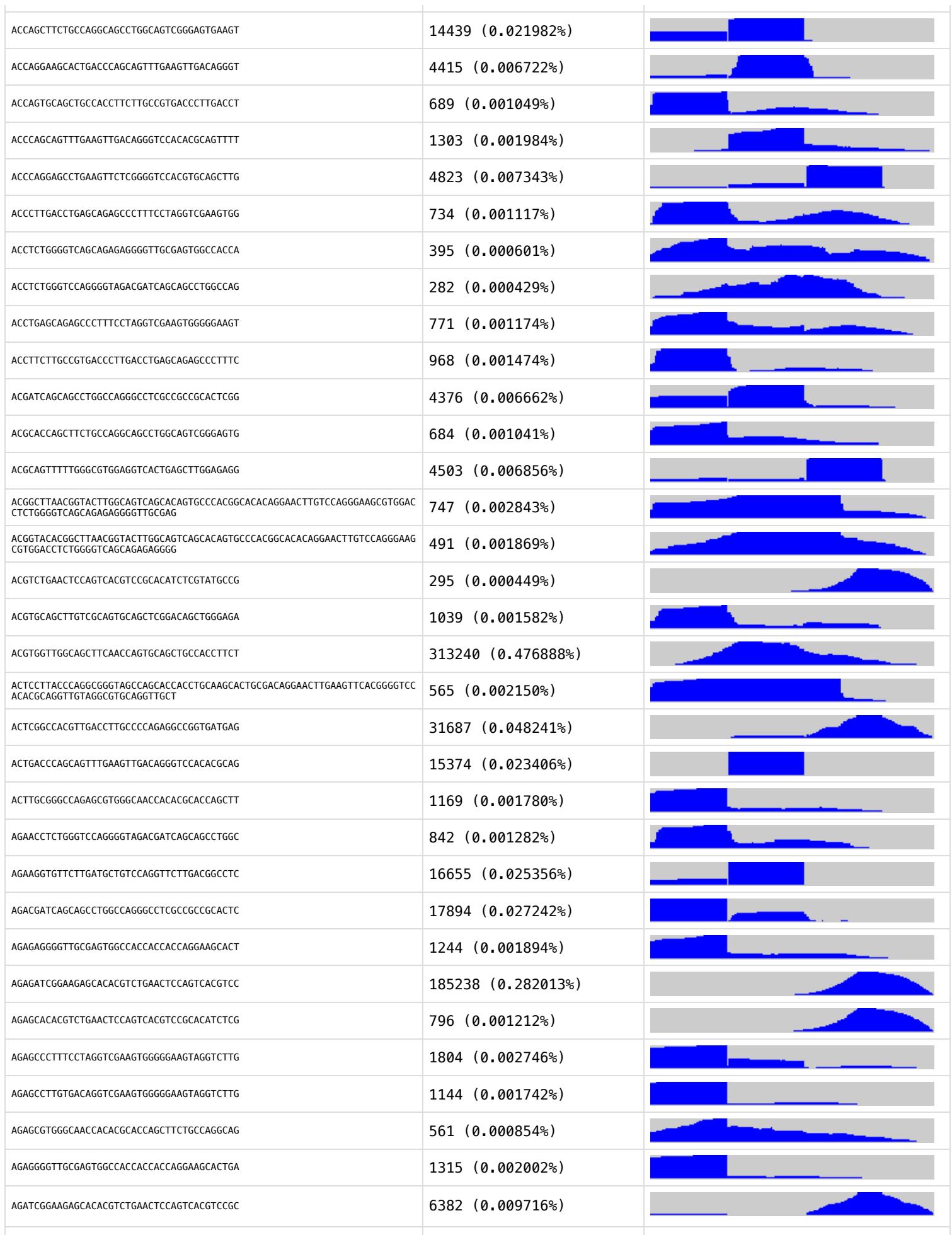


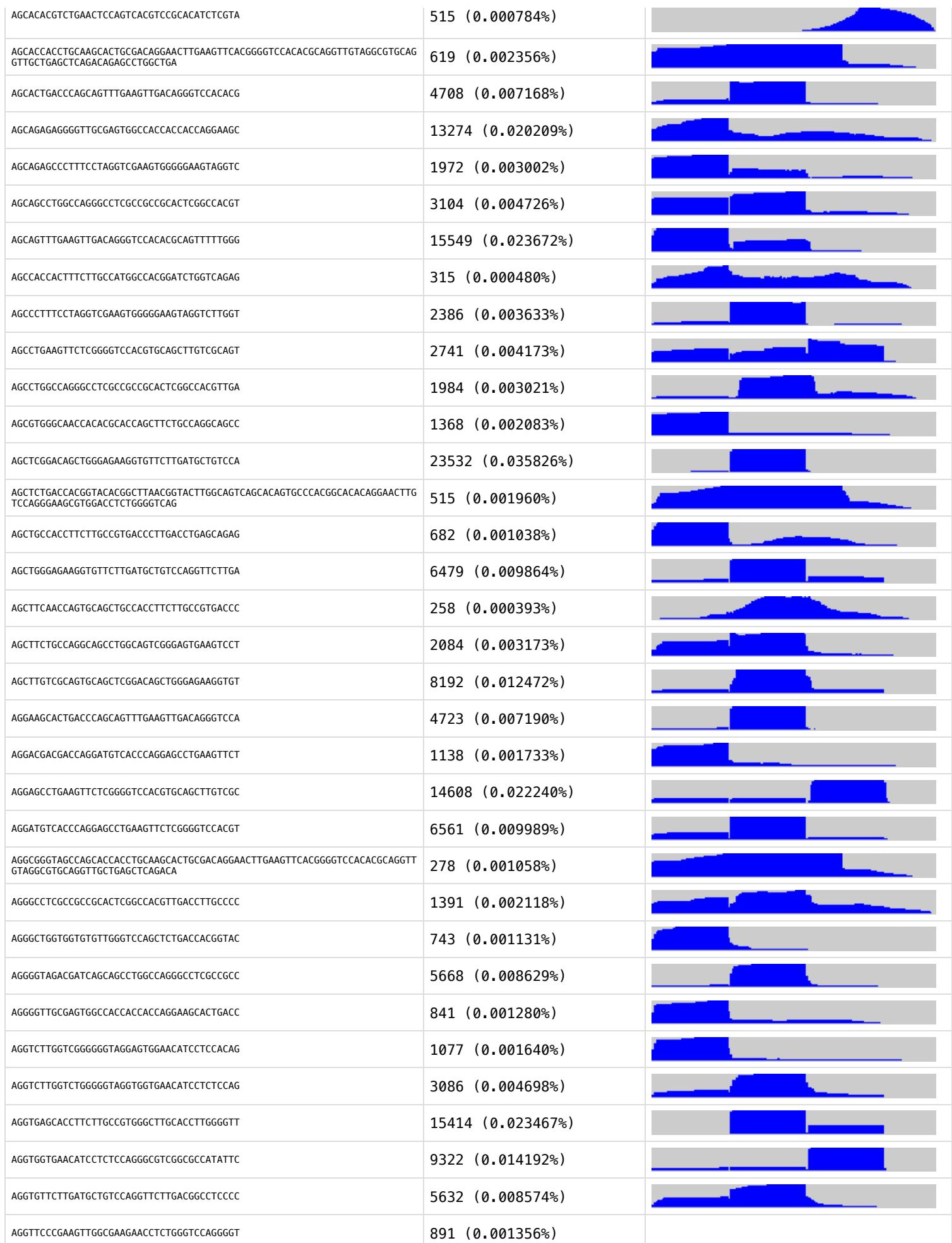
GCT	GCTAA	GCTAT	GCTAC	GCTAG	GCTTA	GCTTT	GCTTC	GCTTG	GCTCA	GCTCT	GCTCC	GCTCG	GCTGA	GCTGT	GCTGC	GCTGG
GCC	GCCAA	GCCAT	GCCAC	GCCAG	GCTTA	GCCTT	GCTTC	GCTTG	GCCC	GCCCC	GCCCG	GCCGA	GCCGT	GCCGC	GCCGG	
GCG	GCGAA	GCGAT	GCGAC	GCGAG	GCGTA	GCGTT	GCGTC	GCGTG	GCGCA	GCGCT	GCGCC	GCGCG	GCGGA	GCGGT	GCGGC	GCGGG
GGA	GGAAG	GGAACT	GGAAC	GGAGA	GGATA	GGATT	GGATC	GGATG	GGACA	GGACT	GGACC	GGACG	GGAGA	GGAGT	GGAGC	GGAGG
GGT	GGTAA	GGTAT	GGTAC	GGTAG	GGTTA	GGTTT	GGTTC	GGTTC	GGTCA	GGTCT	GGTCC	GGTCA	GGTGA	GGTGT	GGTGC	GGTGG
GGC	GGCAA	GGCAT	GGCAC	GGCAG	GGCTA	GGCTT	GGCTC	GGCTG	GGCCA	GGCCT	GGCCC	GGCCG	GGCGA	GGCGT	GGCGC	GGCGG
GGG	GGGAA	GGGAT	GGGAC	GGGAG	GGGTA	GGGTT	GGGTC	GGGTC	GGGCA	GGGCT	GGGCC	GGGCG	GGGGA	GGGGT	GGGGC	GGGGG

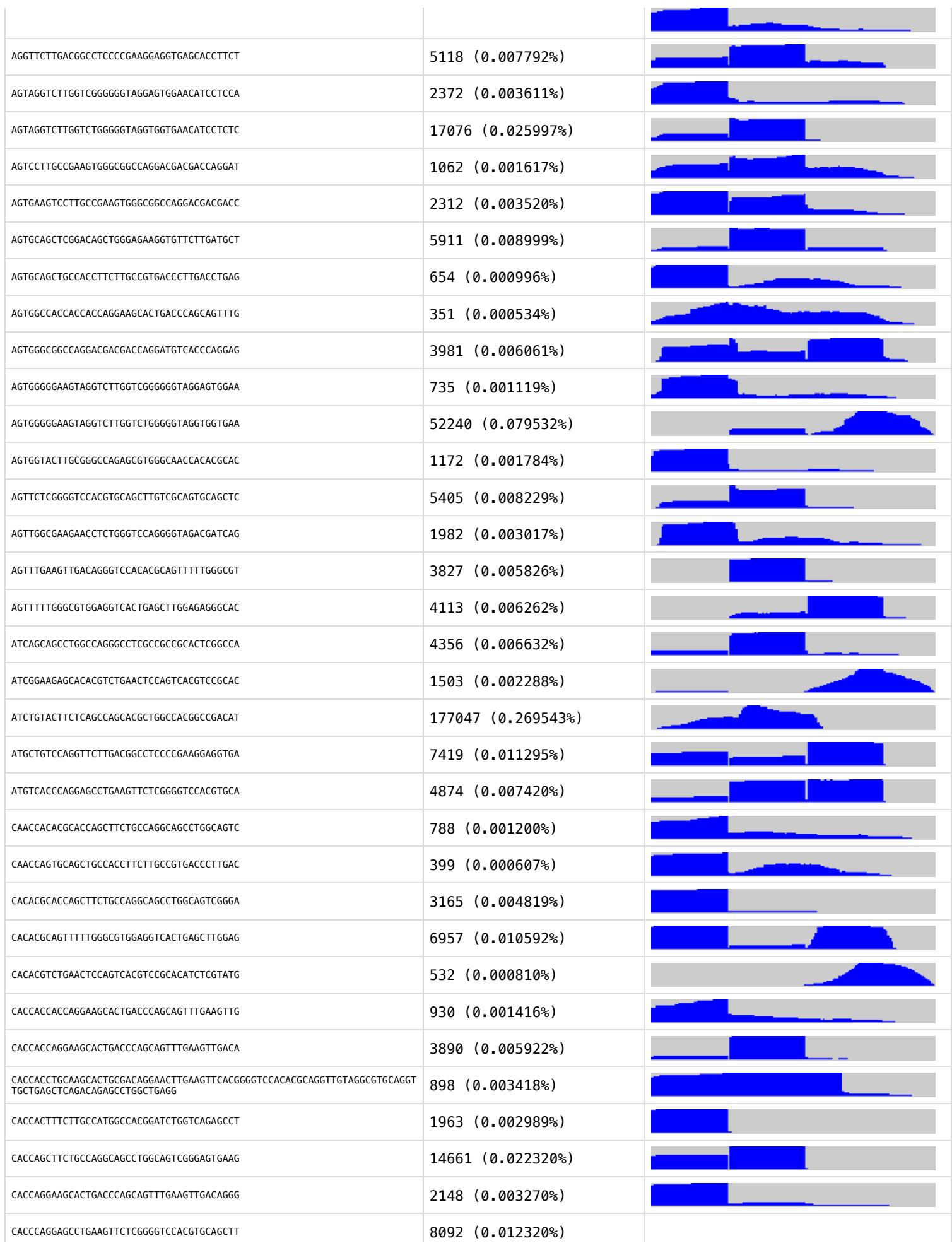
Before filtering: read1: overrepresented sequences

Sampling rate: 1 / 20

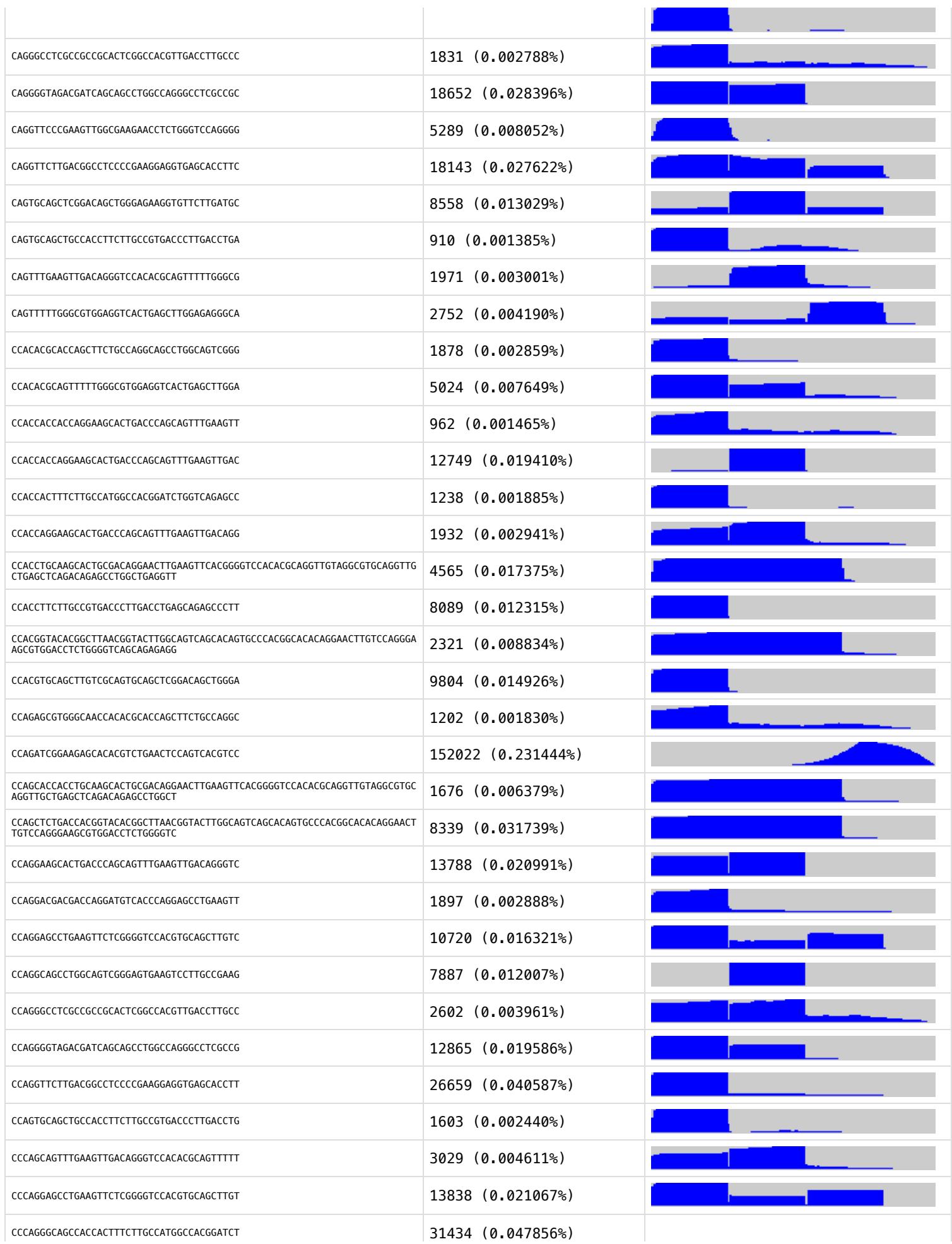


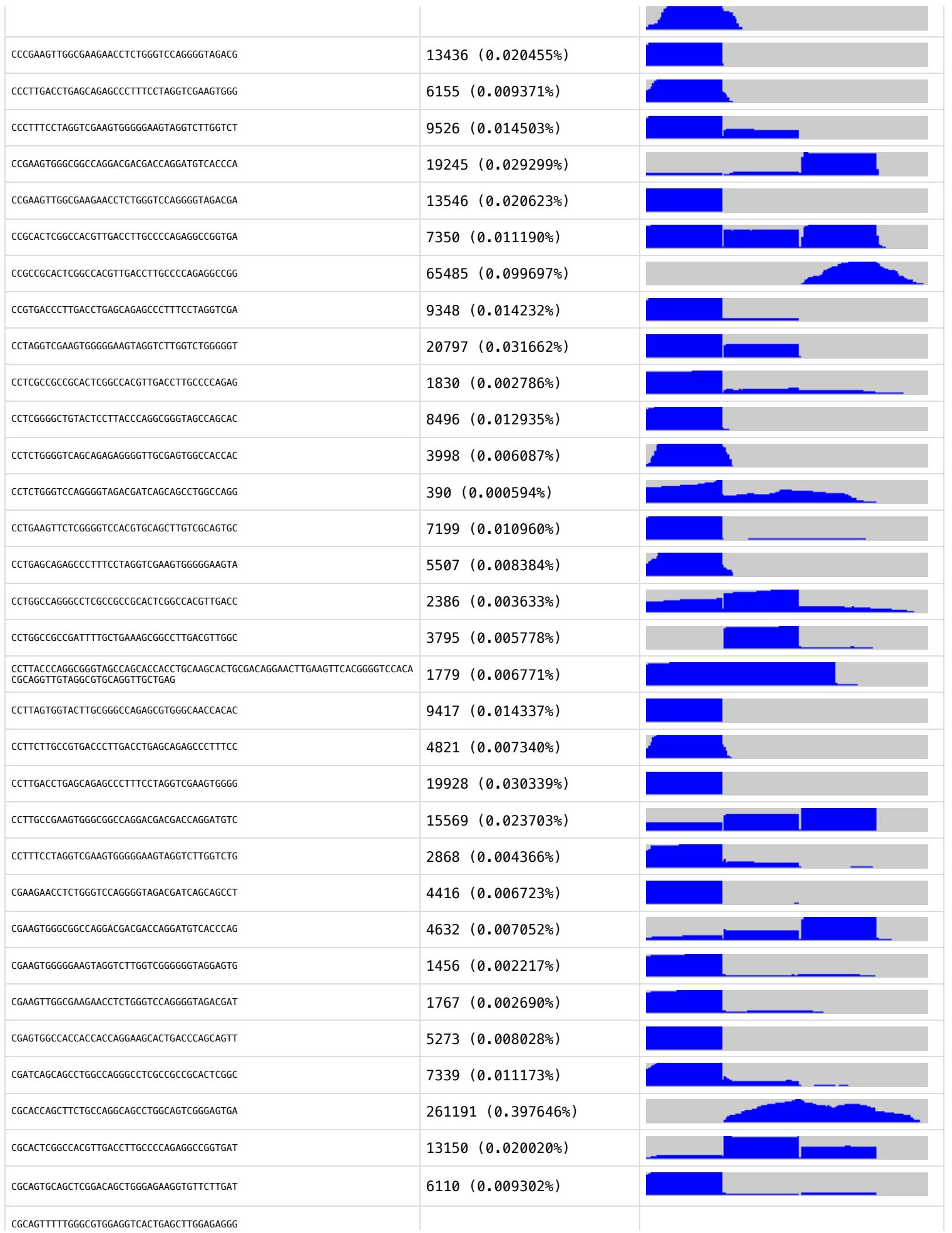


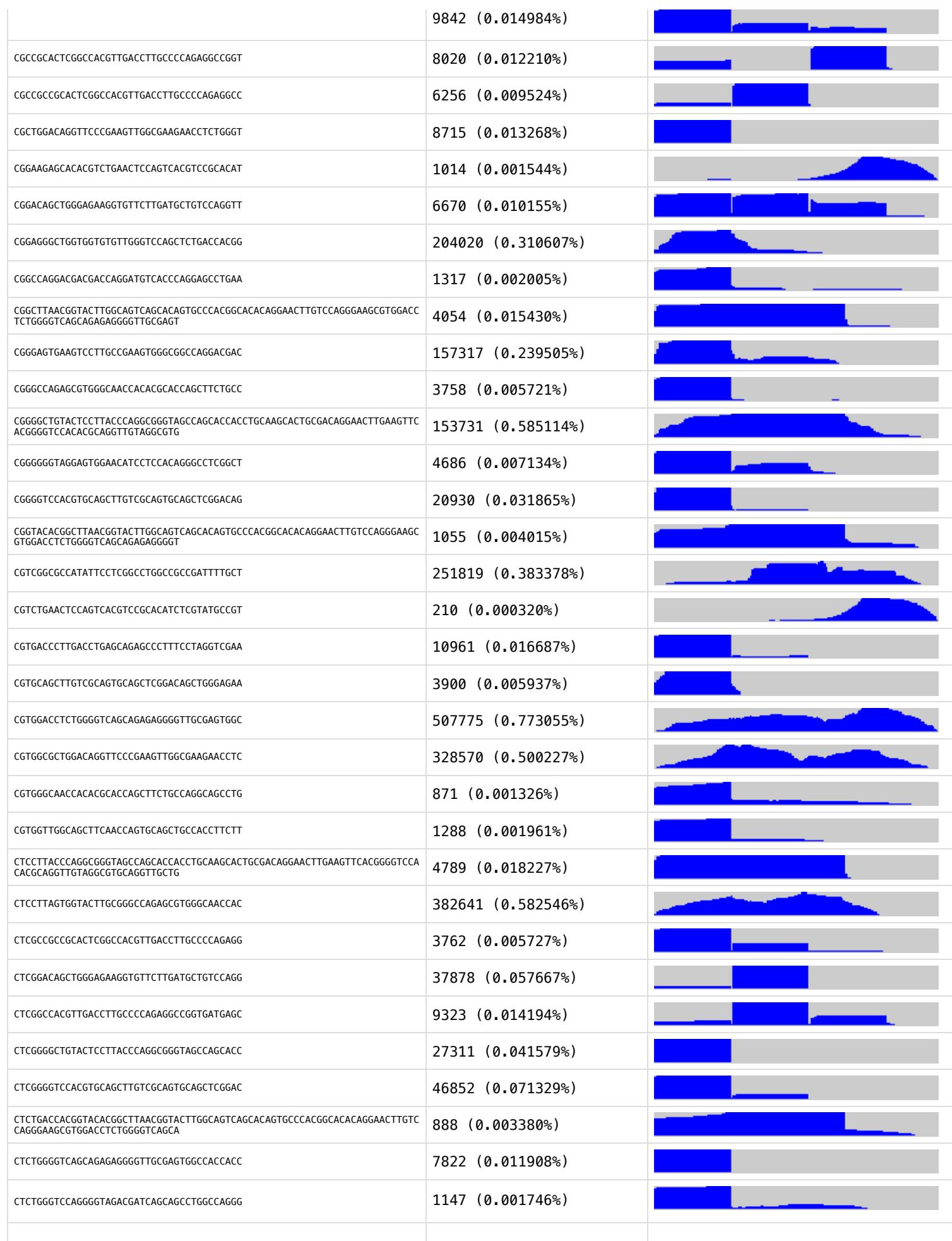


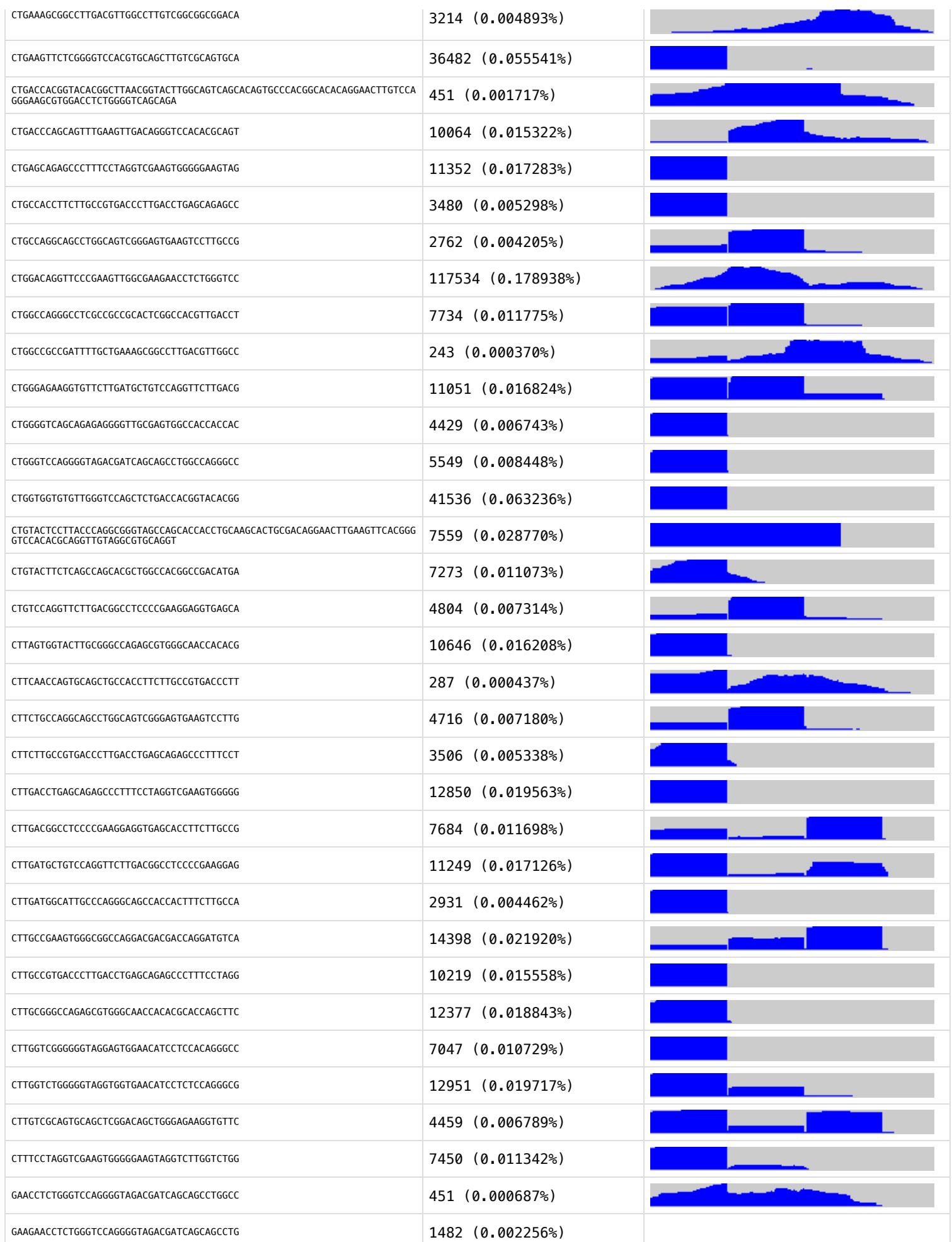






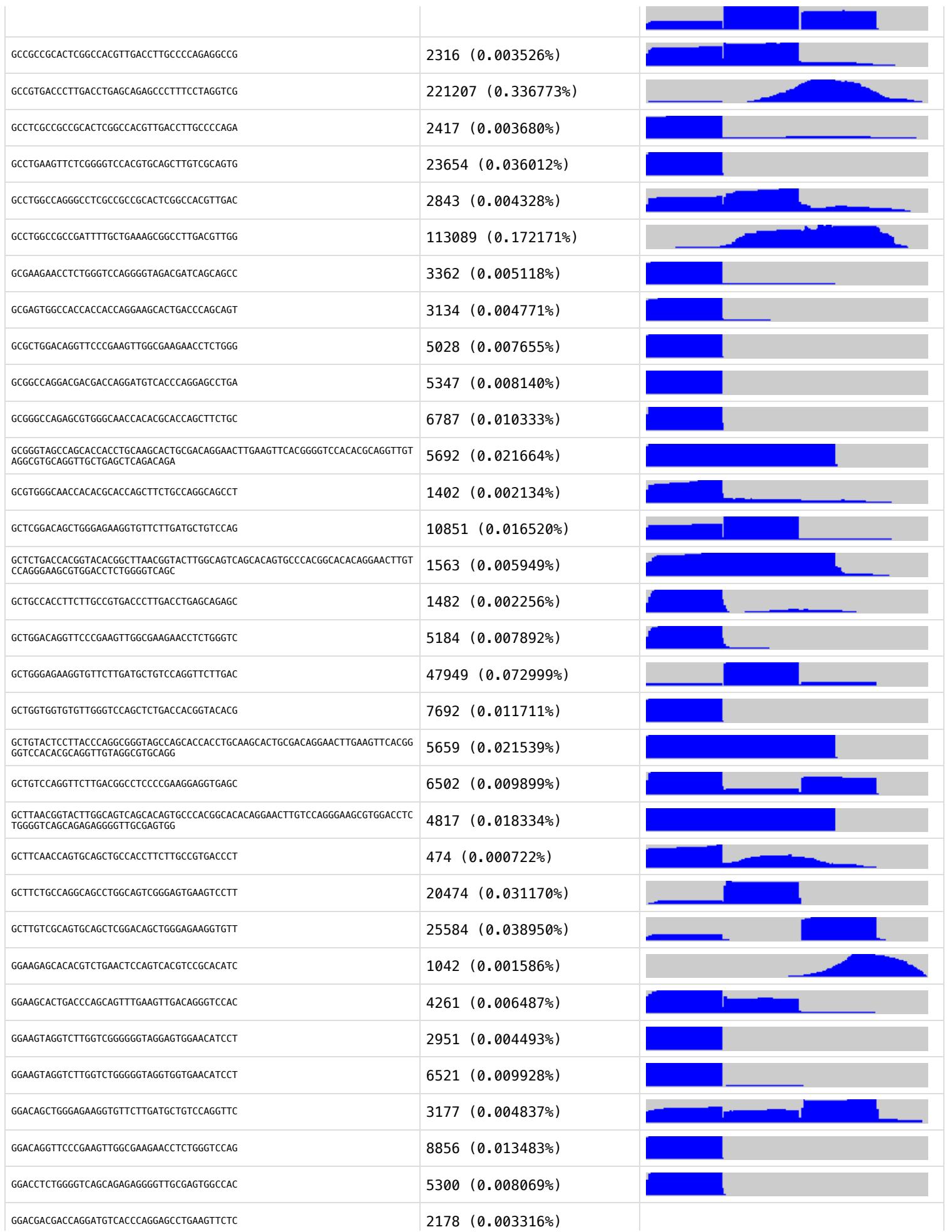




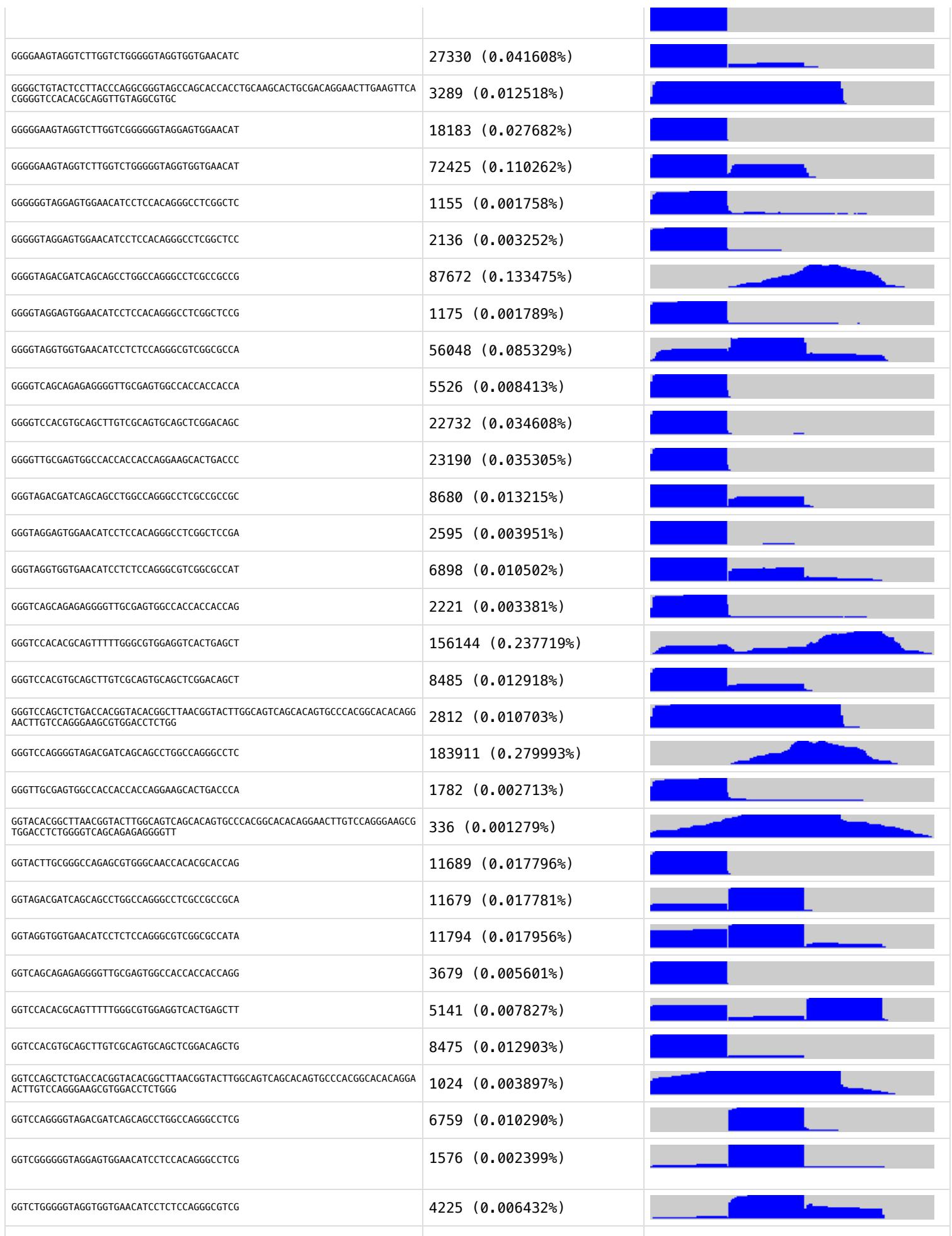


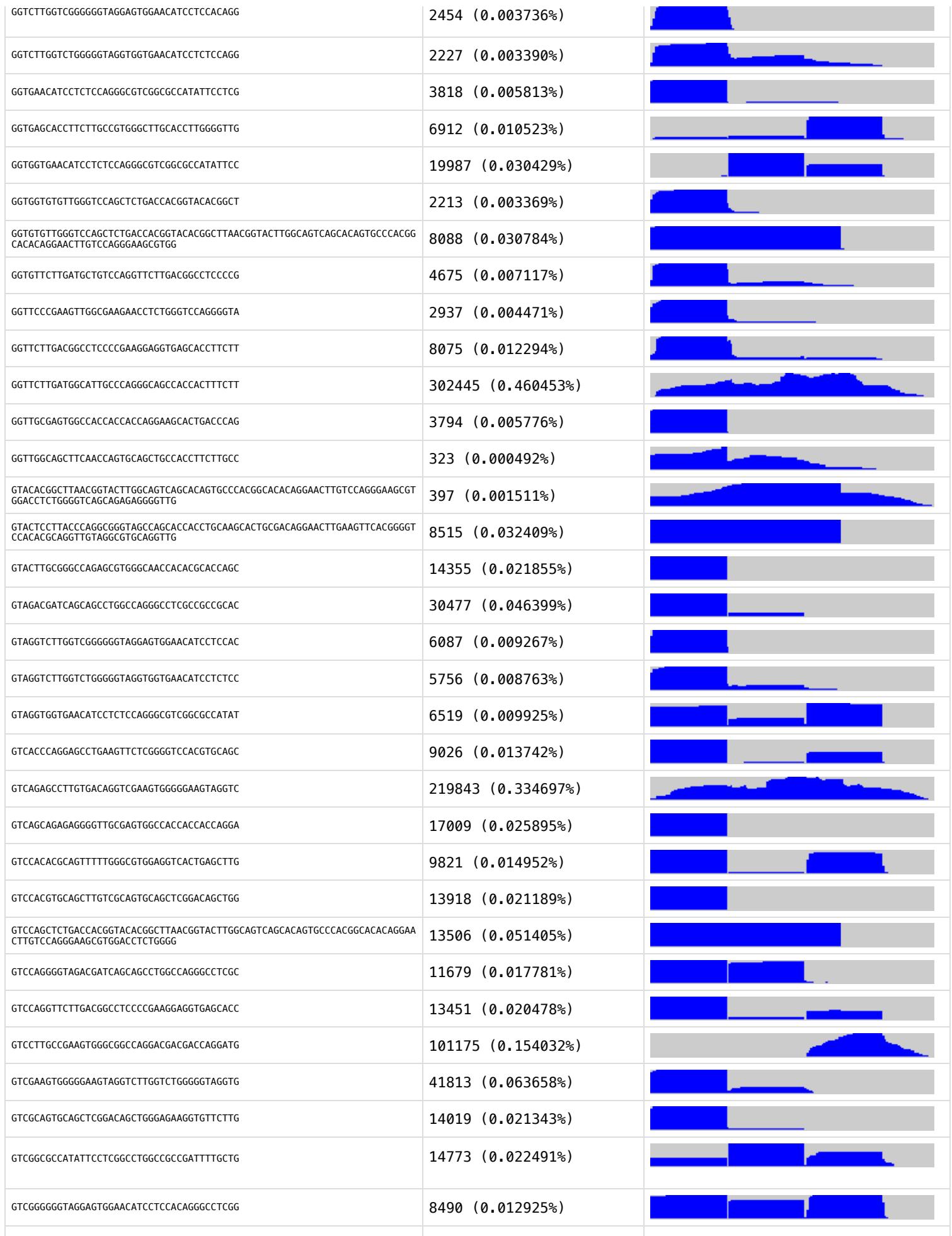


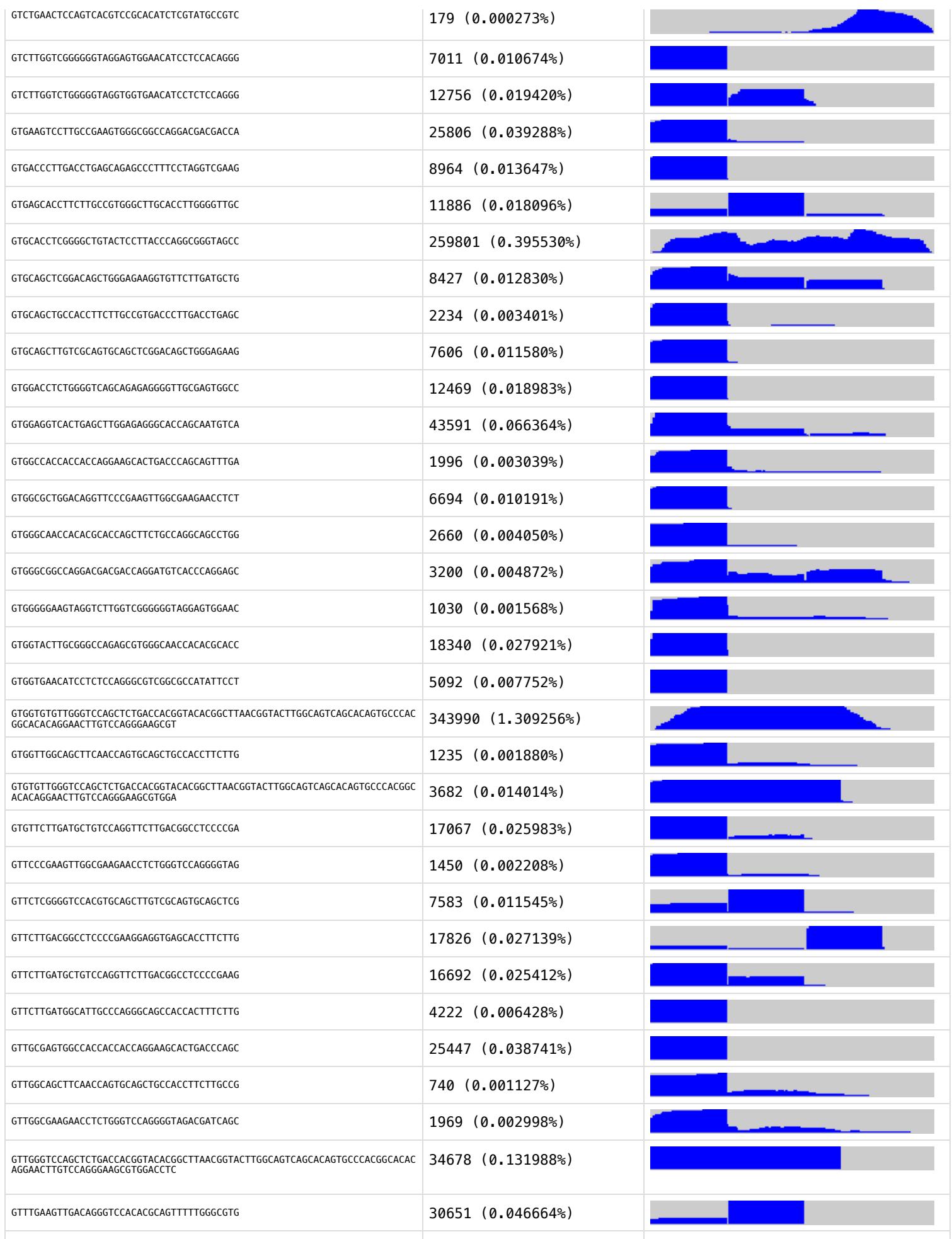


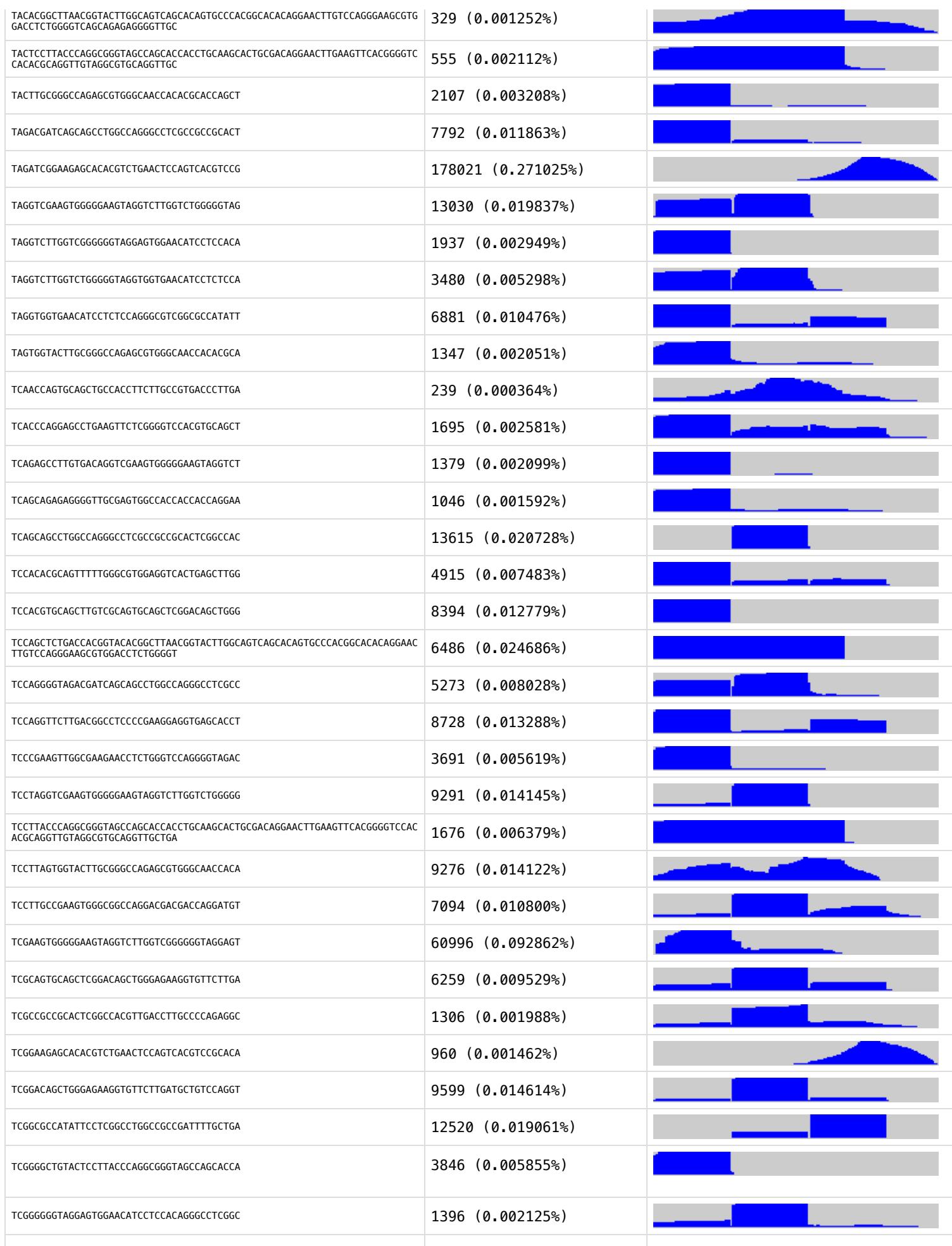


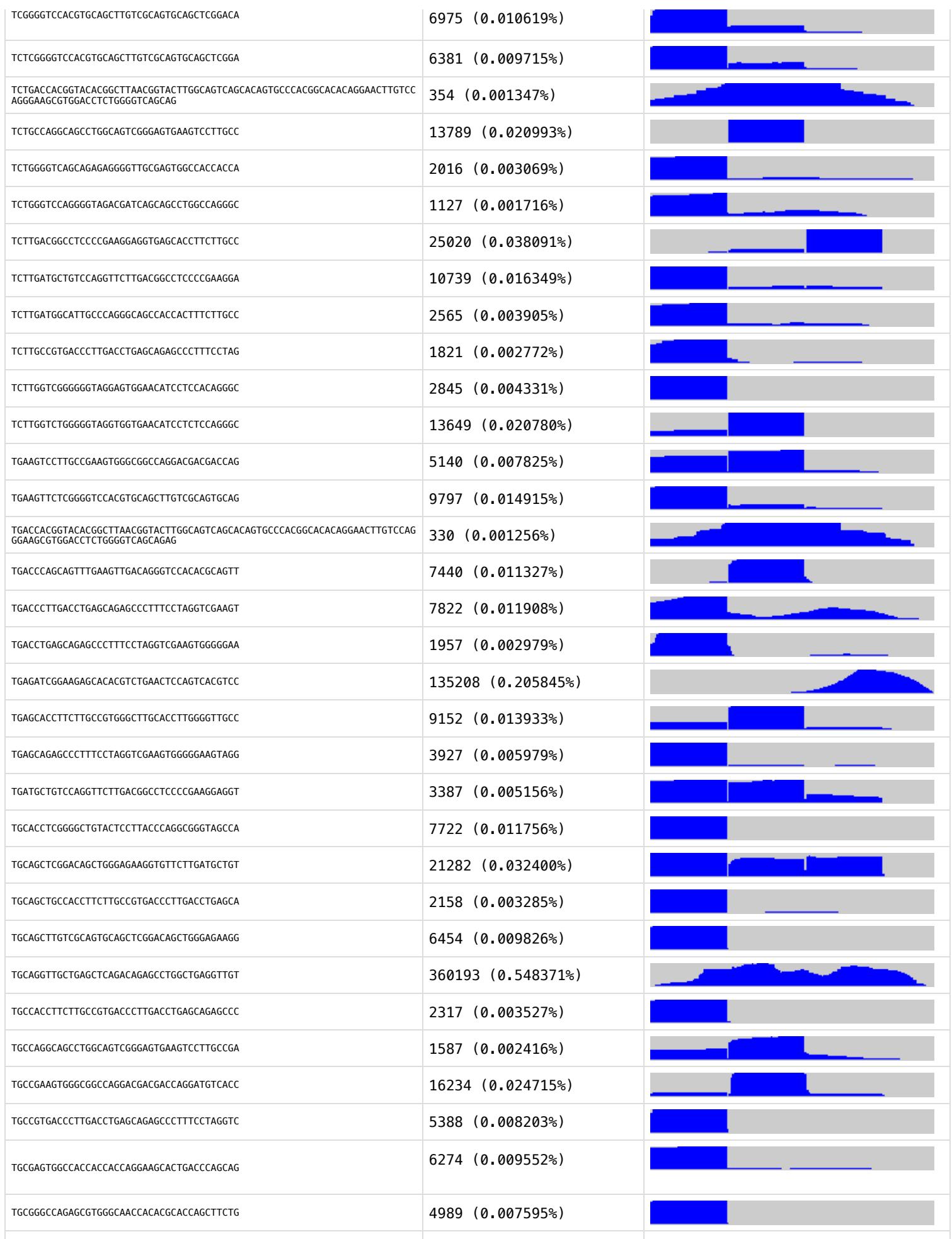




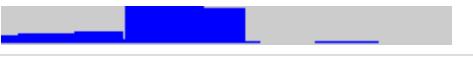
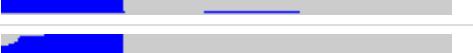
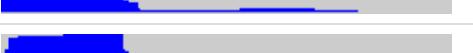
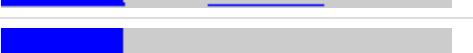
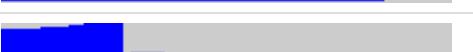






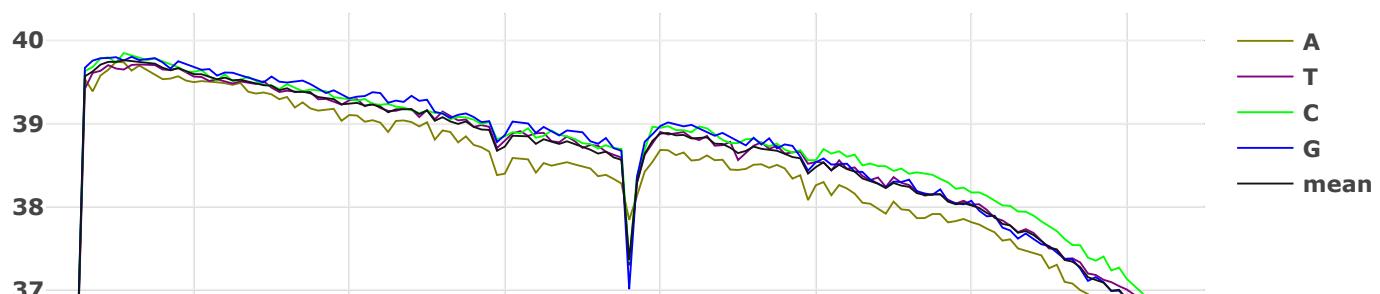


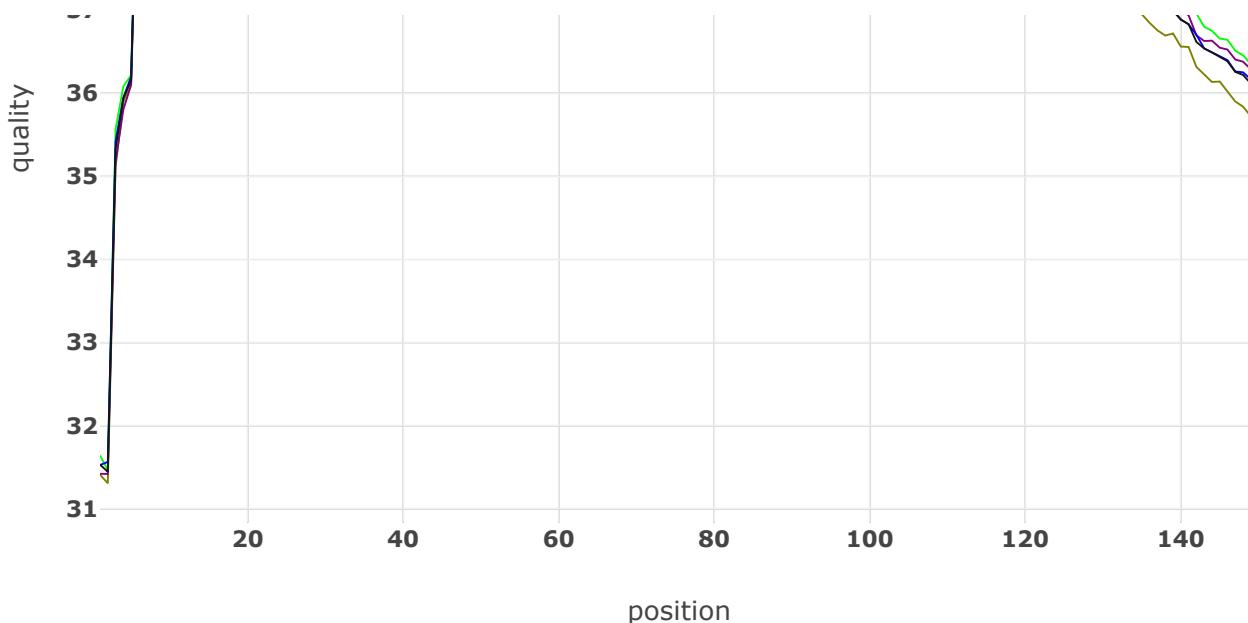


TTCTTAGTCGAAGTGGGGAAGTAGGTCTTGGTCTGGG	3007 (0.004578%)	
TTCTCGGGTCCACGTGCAGCTTGTGCACTGCAGCTCGG	2117 (0.003223%)	
TTCTGCCAGGCAGCCTGGCAGTCGGAGTGAAAGTCTTC	11426 (0.017395%)	
TTCTTGACGCCCTCCCAGAAGGAGGTGAGCACCTCTTC	10185 (0.015506%)	
TTCTTGATGCTGTCCAGGTTCTTGACGCCCTCCCGAAGG	40932 (0.062316%)	
TTCTTGATGGCATTGCCAGGGCAGCCACACTTCTTC	1555 (0.002367%)	
TTCTTGCGGTGACCCCTGACCTGAGCAGGCCCTTCTTA	1539 (0.002343%)	
TTGACCTTGAGCAGAGCCTTCTAGGTGAAAGTGGGGG	1012 (0.001541%)	
TTGACGCCCTCCCCAGAAGGAGGTGAGCACCTCTTGCGG	4268 (0.006498%)	
TTGATGCTGTCCAGGTTCTTGACGCCCTCCCGAAGGAGG	5277 (0.008034%)	
TTGATGGCATTGCCAGGGCAGCCACACTTCTTGCCAT	475 (0.000723%)	
TTGCCGAAGTGGCCGCCAGGACGACGACCAGGATGTAC	903 (0.001375%)	
TTGCCGTGACCTTGACCTGAGCAGAGGCCCTTCTTAGGT	2188 (0.003331%)	
TTGCAGTGCCACCAACCAGGAAGCACTGACCCAGCA	4611 (0.007020%)	
TTGGGCCAGAGCGTGGCAACCACACGCACCAGCTCT	7810 (0.011890%)	
TTGGCAGCTTCAACCAAGTGCACTGCCACCTTCTTGCGG	268 (0.000408%)	
TTGGCGAACCTCTGGGTCAAGGGTAGACGATCAGCA	1315 (0.002002%)	
TTGGTCCAGCTCTGACACGGTACACGGCTAACGGTACTTGGCAGTCAGCACAGTGCCACGGCACACA GGAACTTGTCCAGGGAGCGTGGACCTCT	749 (0.002851%)	
TTGGTCTGGGGTAGGGTGTGAACATCCTCTCCAGGGCGT	625 (0.000952%)	
TTGTCGAGCTGCAGCTCGGACAGCTGGAGAAGGTGTTCT	3939 (0.005997%)	
TTTCTAGTCGAAGTGGGGAGTAGGTCTTGGCTGG	5711 (0.008695%)	
TTTGGCGGTGGAGGTCACTGAGCTGGAGAGGGACCAGC	4352 (0.006626%)	
TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	39907 (0.060756%)	
	91485 (0.139280%)	

Before filtering: read2: quality

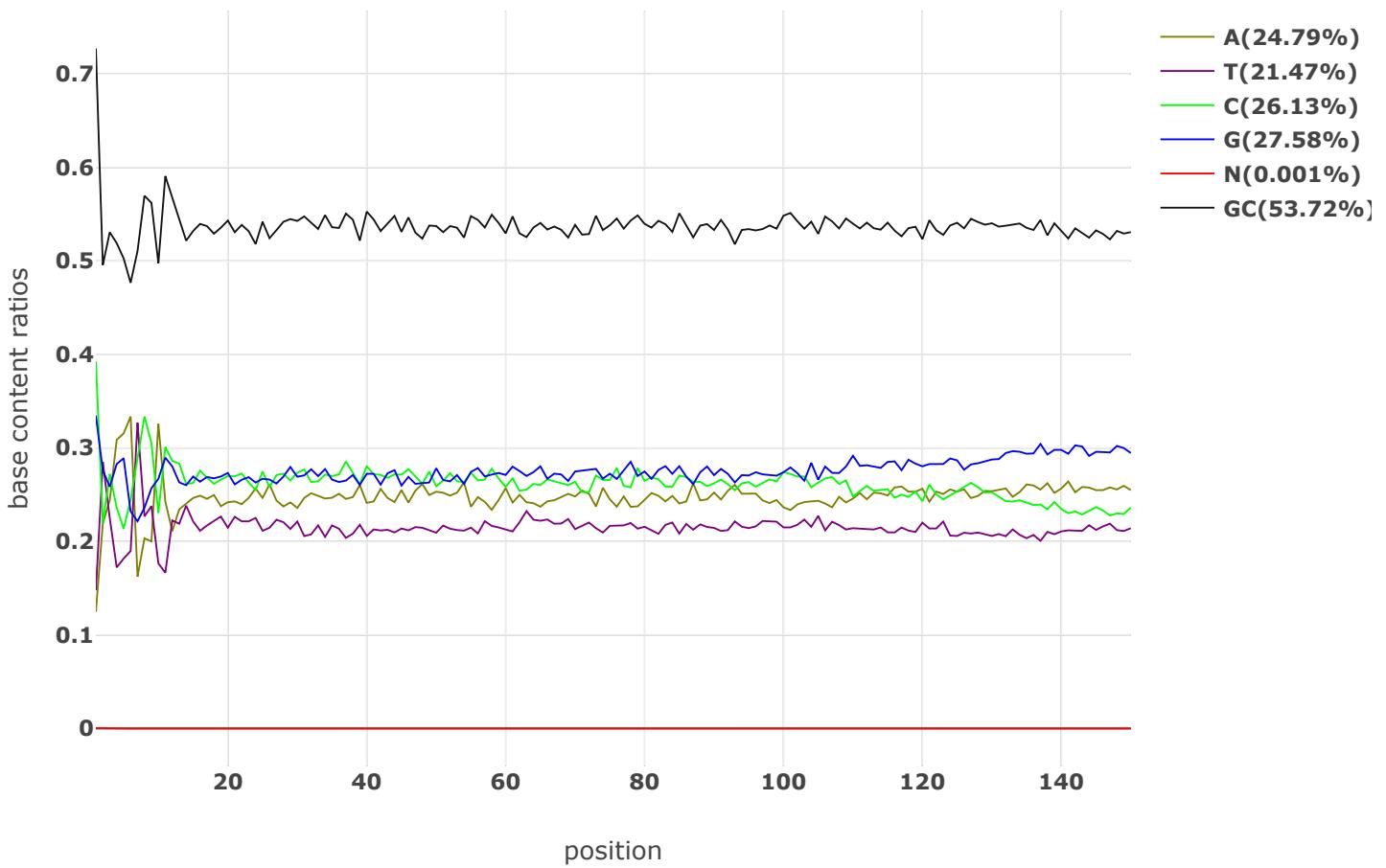
Value of each position will be shown on mouse over.



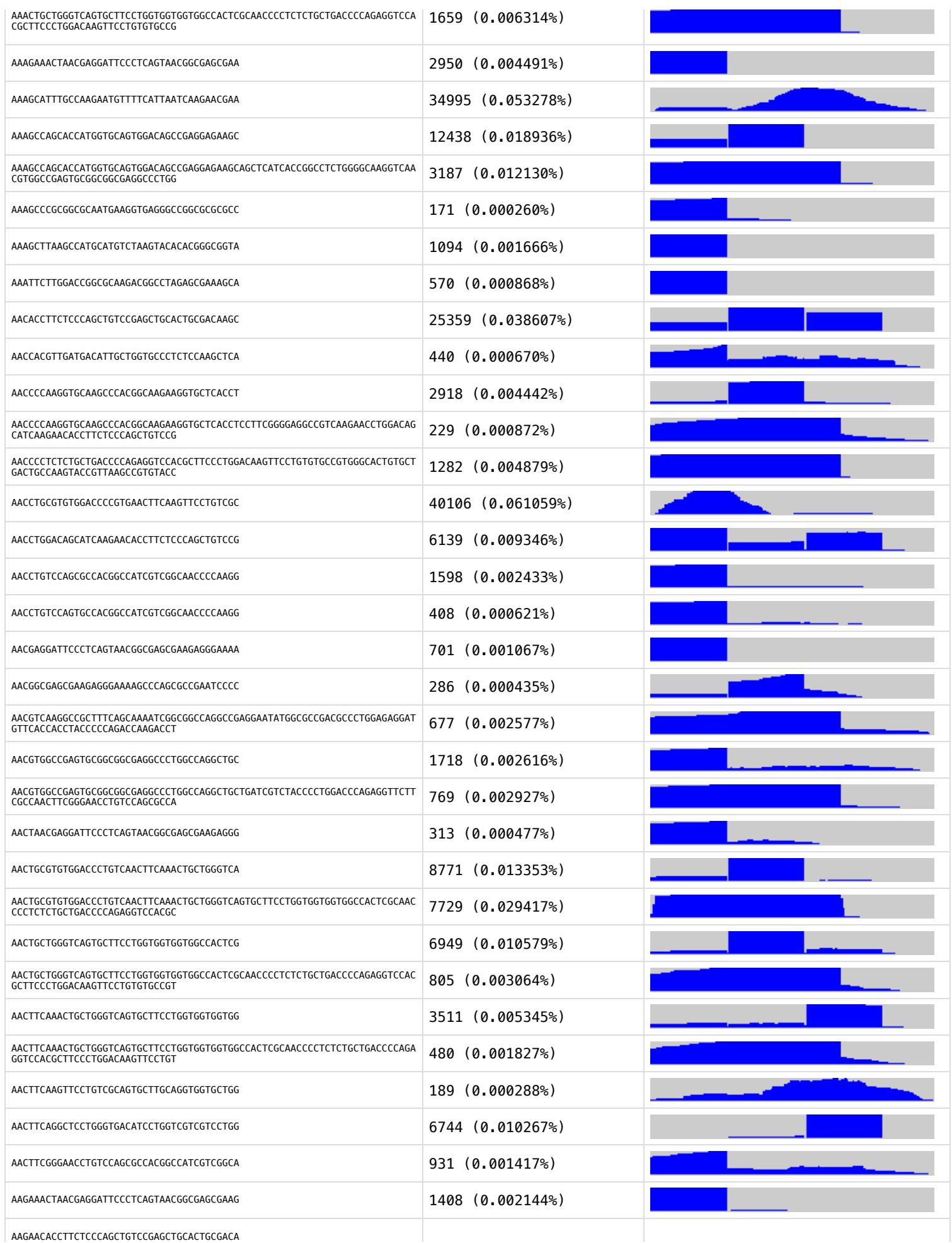


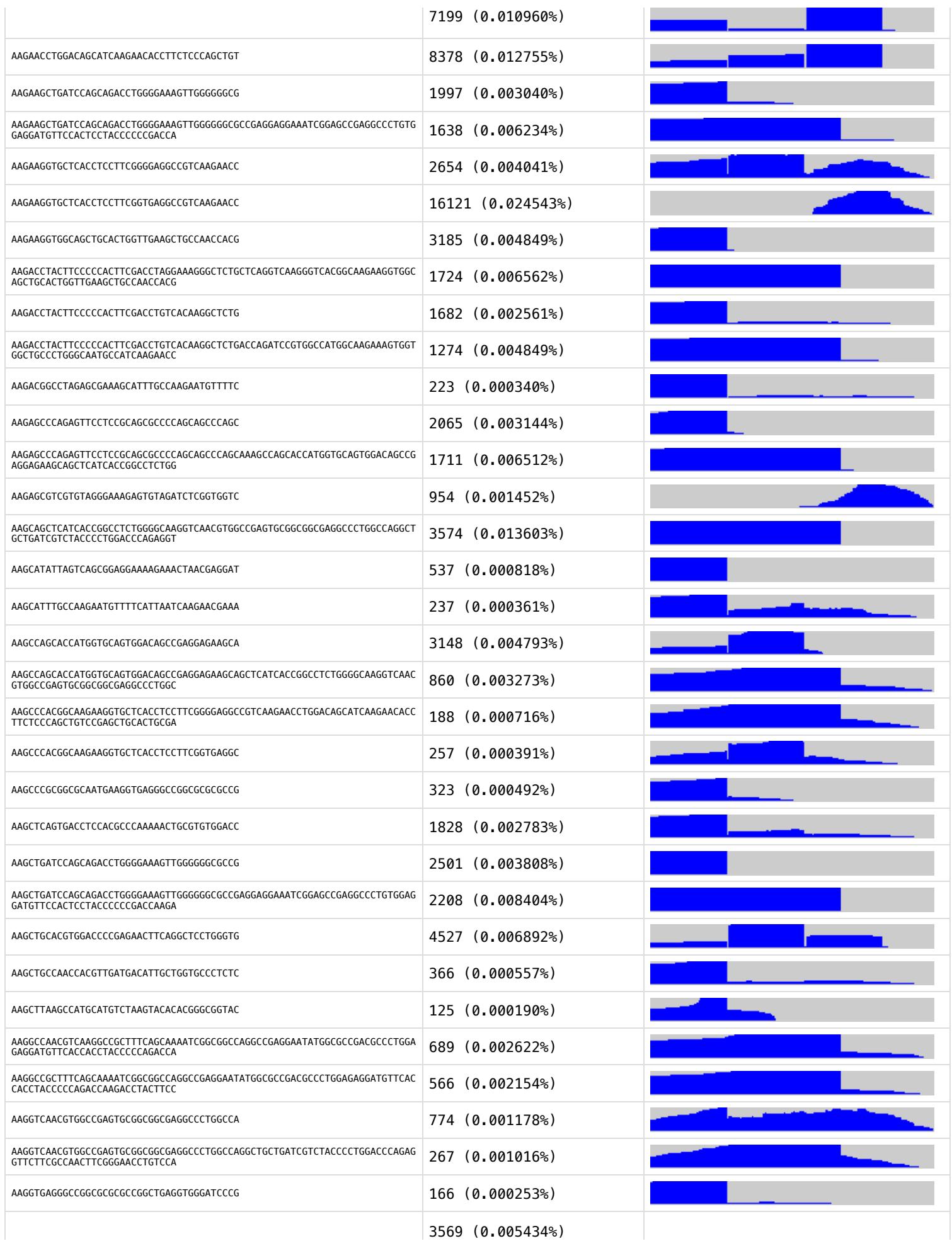
Before filtering: read2: base contents

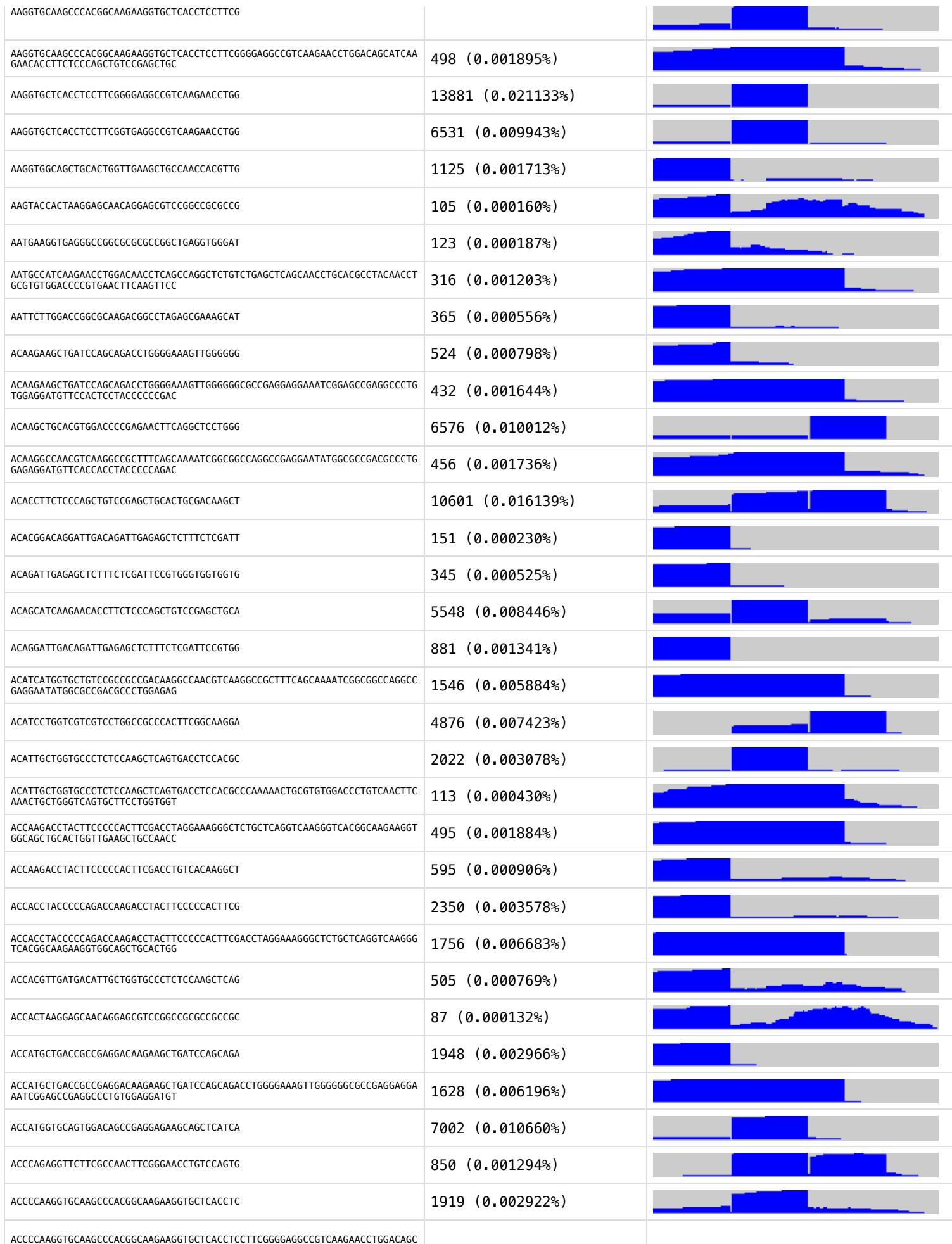
Value of each position will be shown on mouse over.



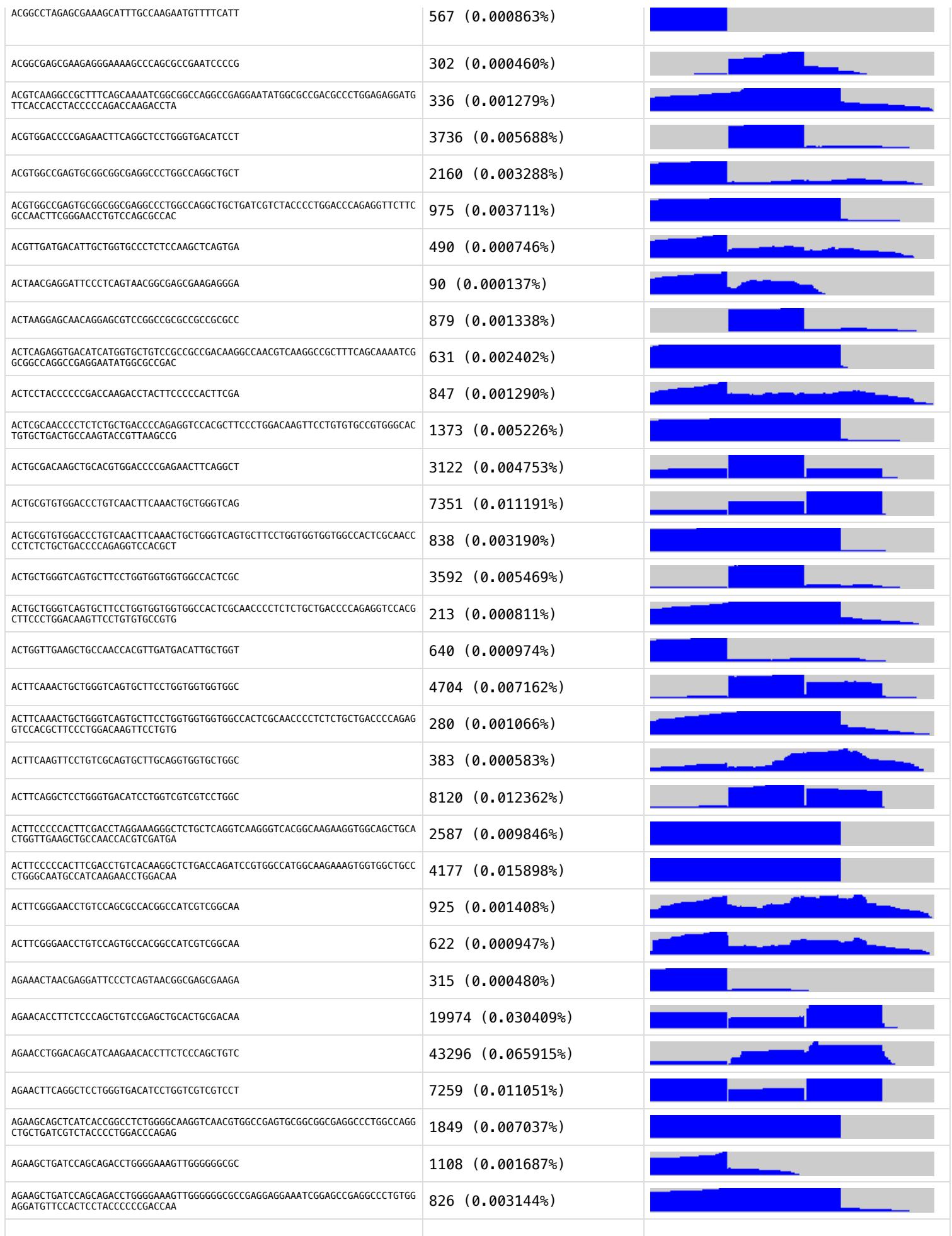
Before filtering: read2: KMER counting

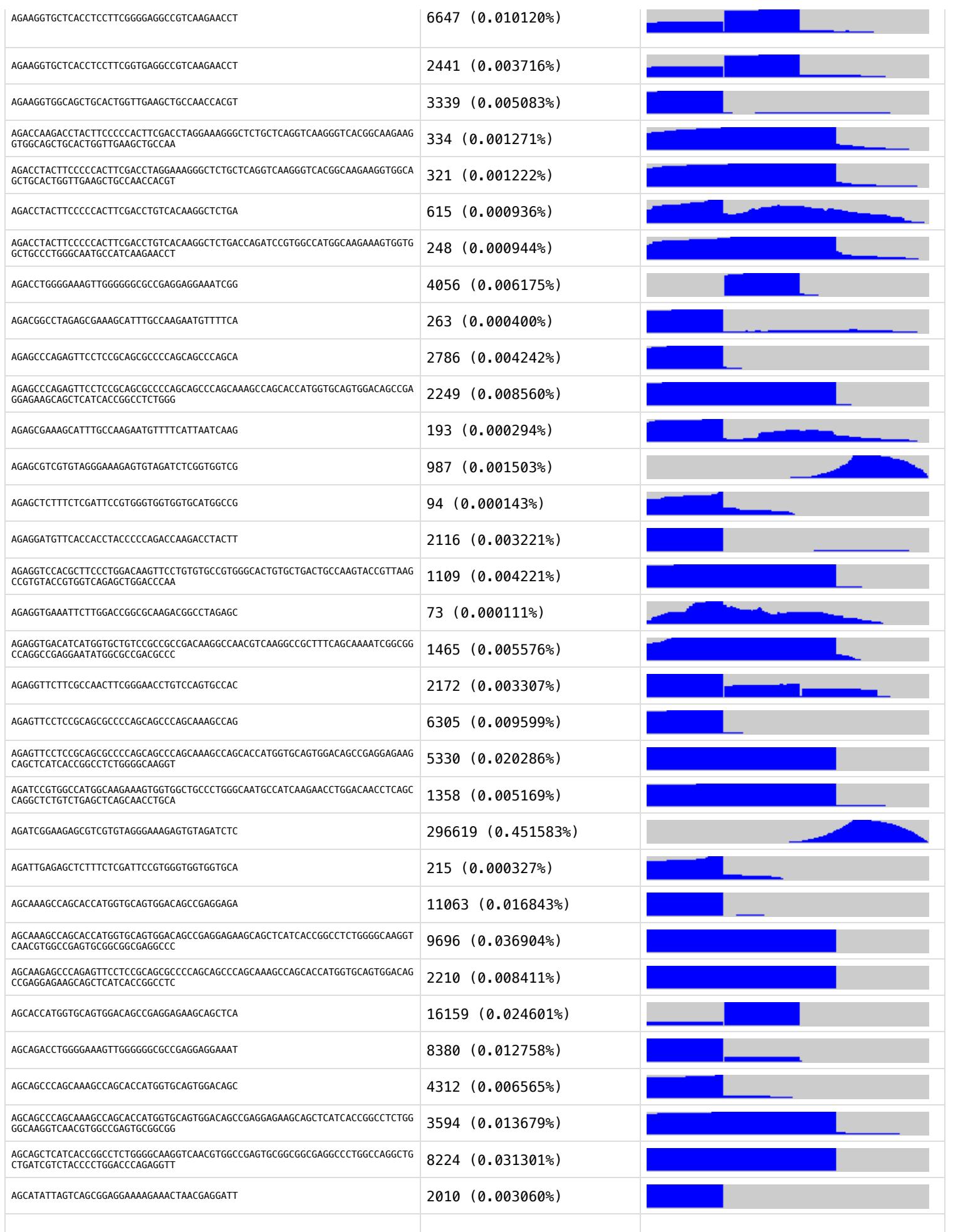


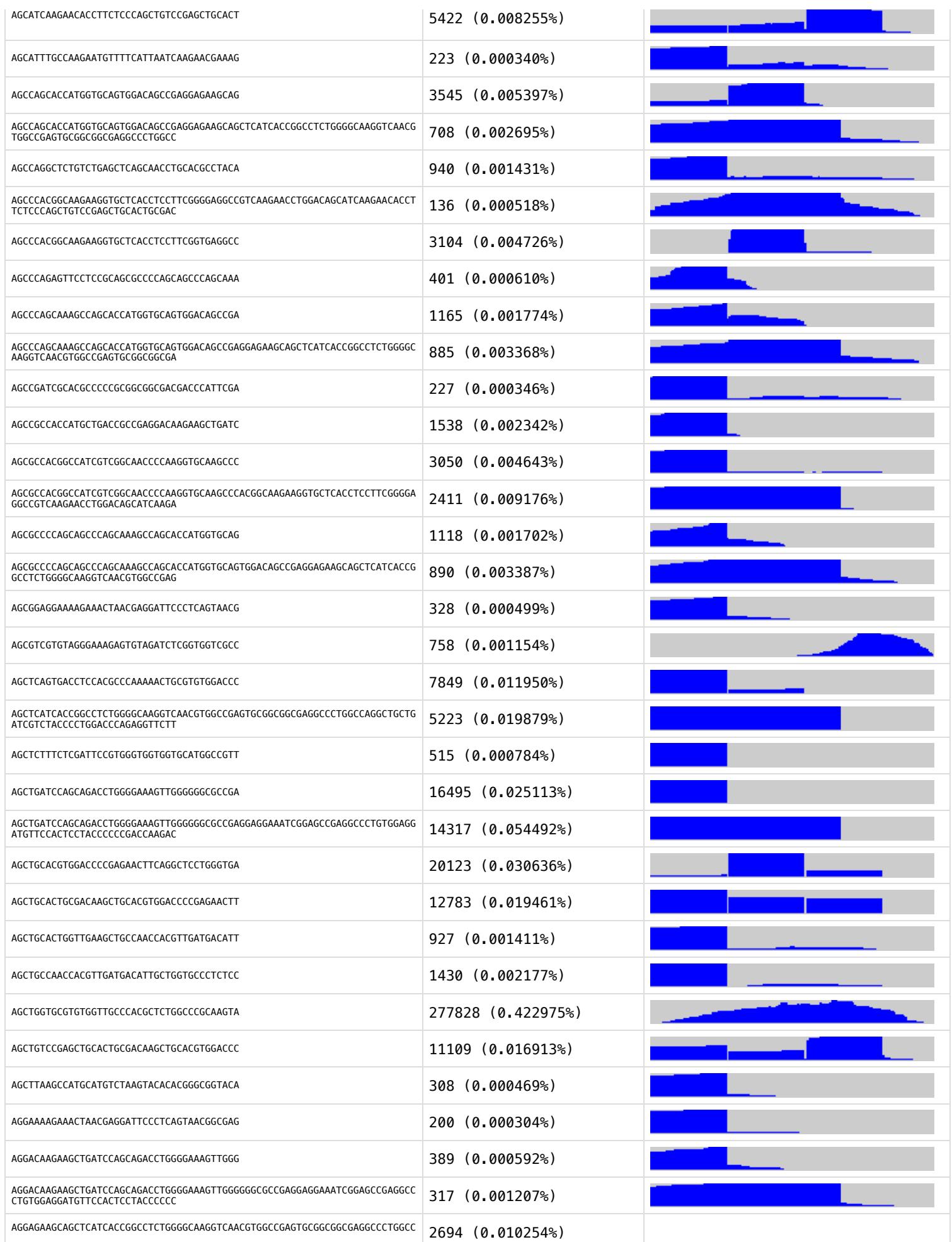


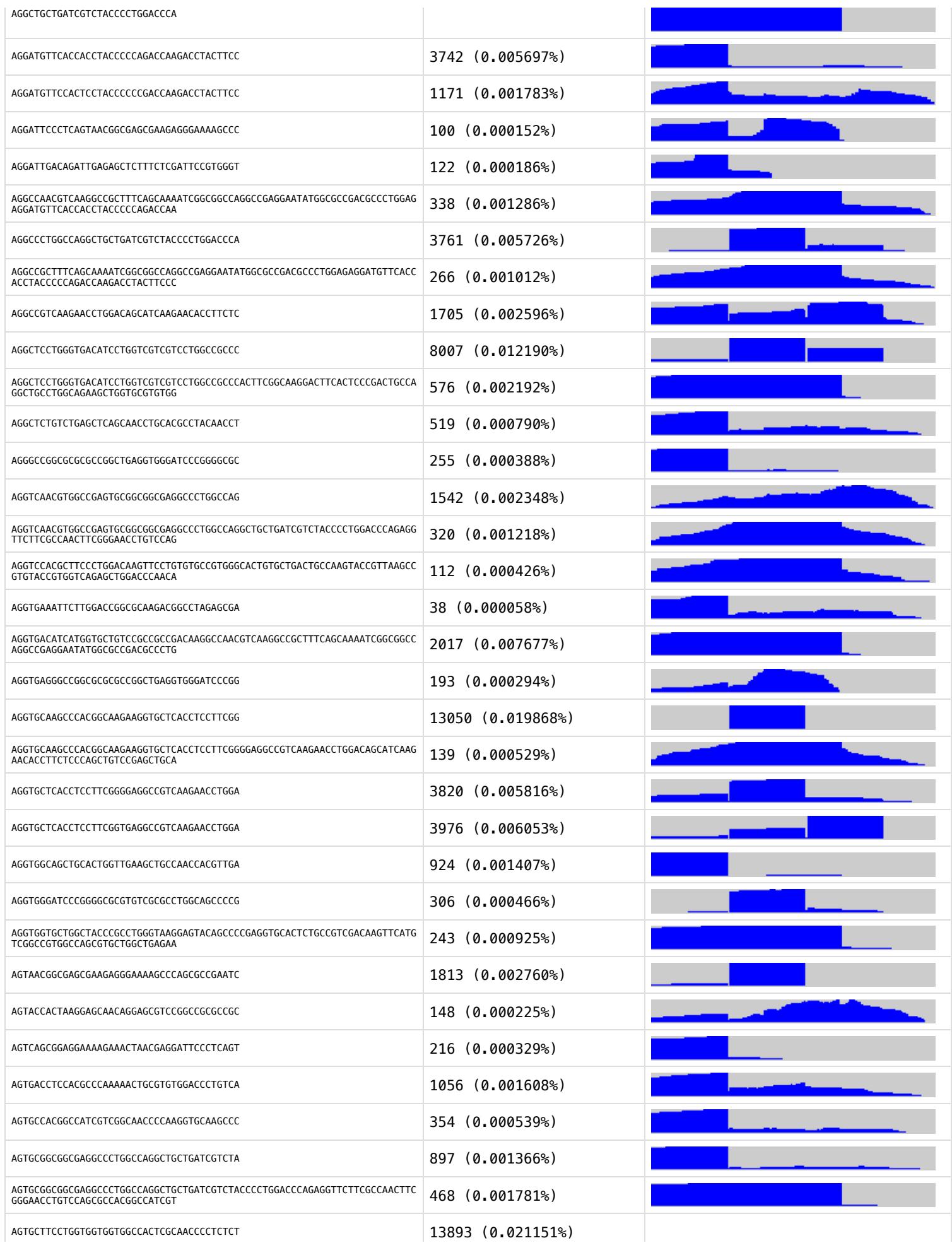


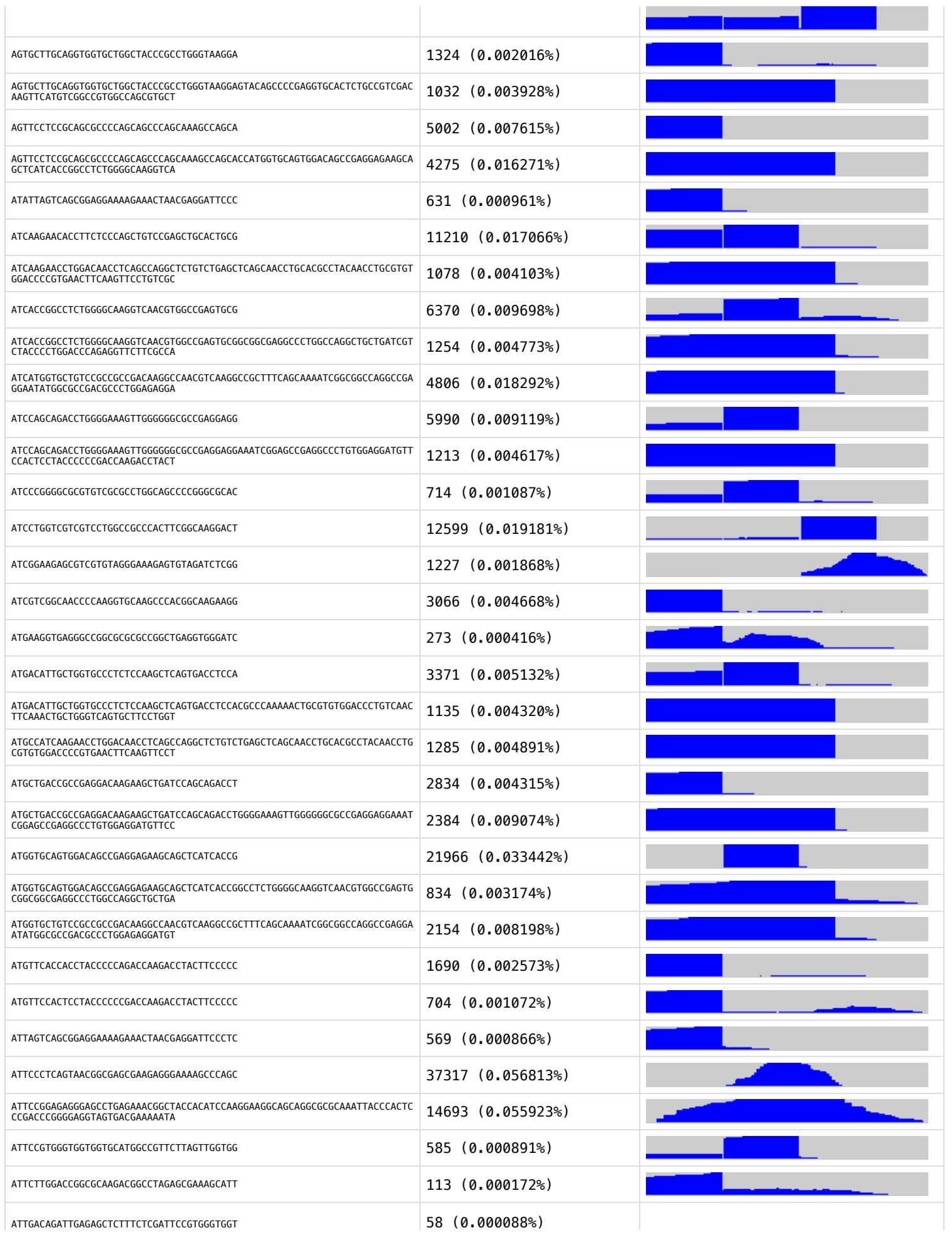


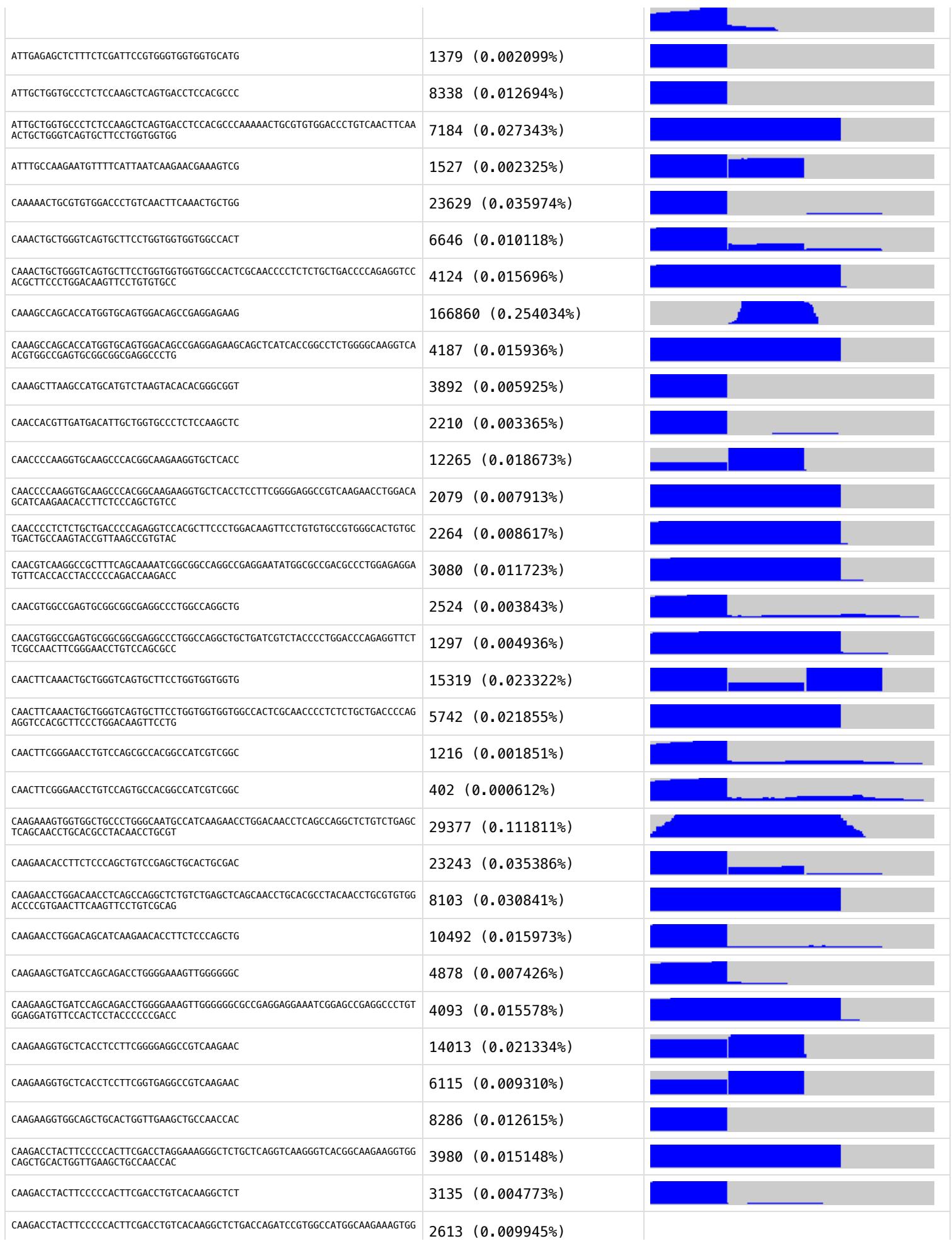


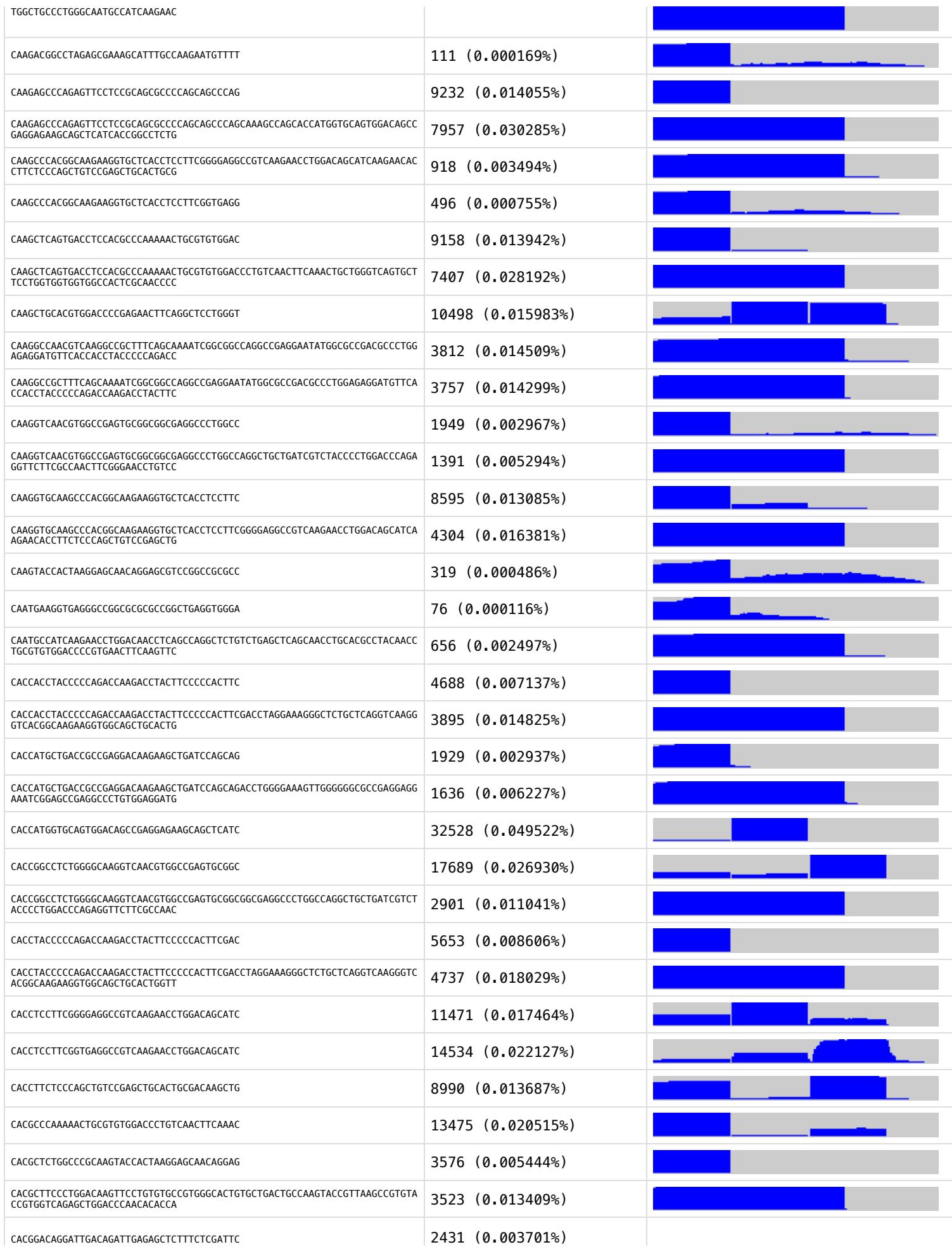


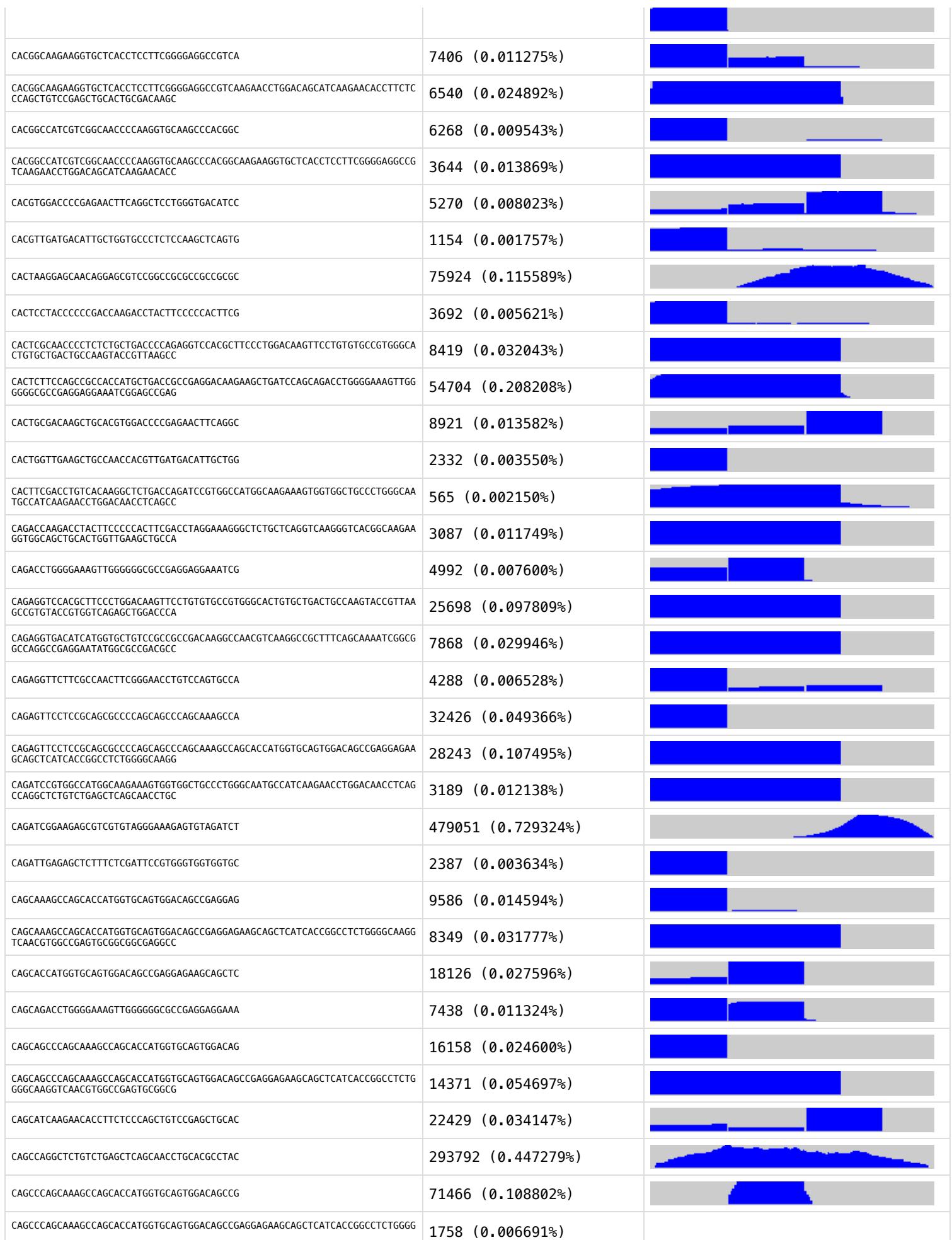


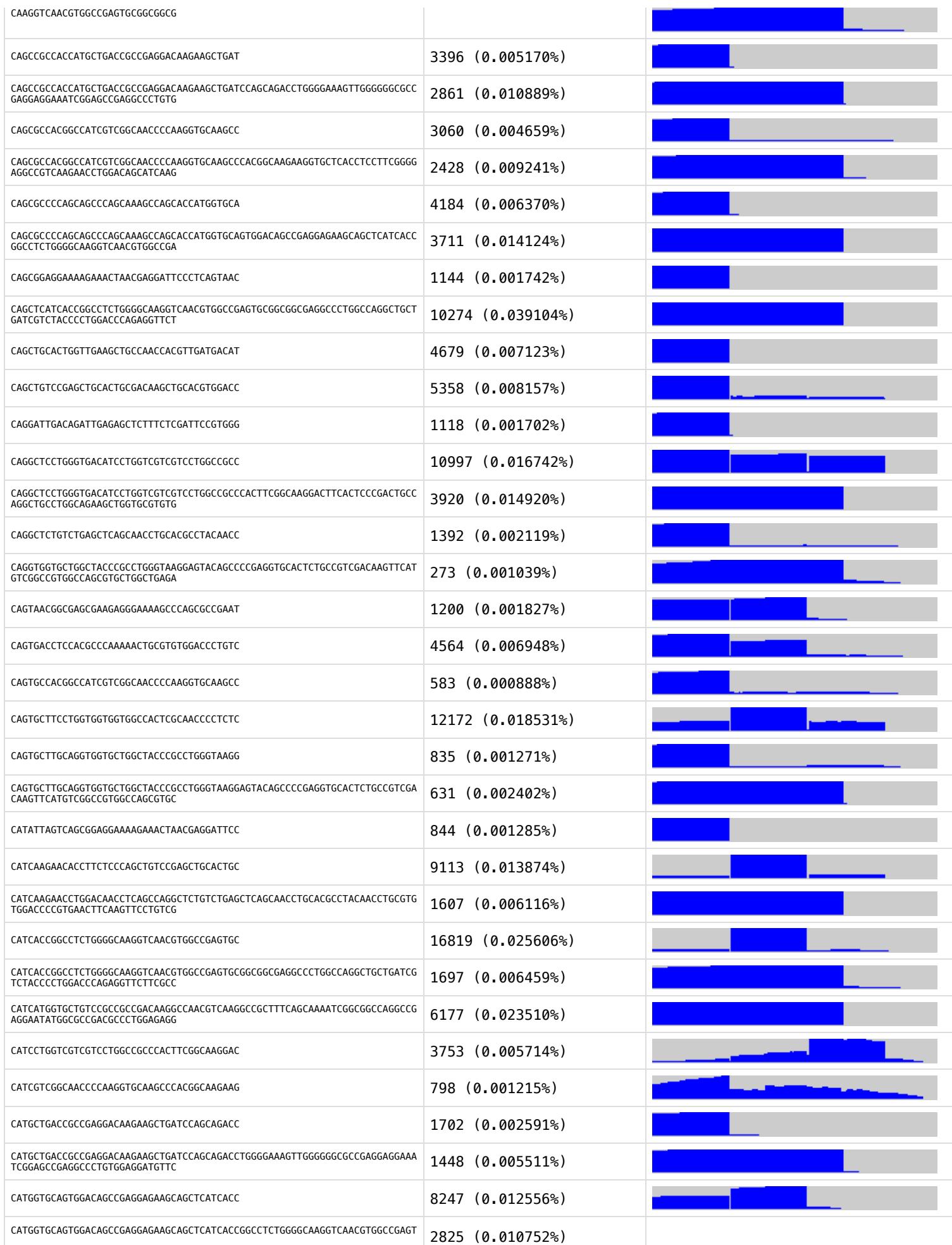


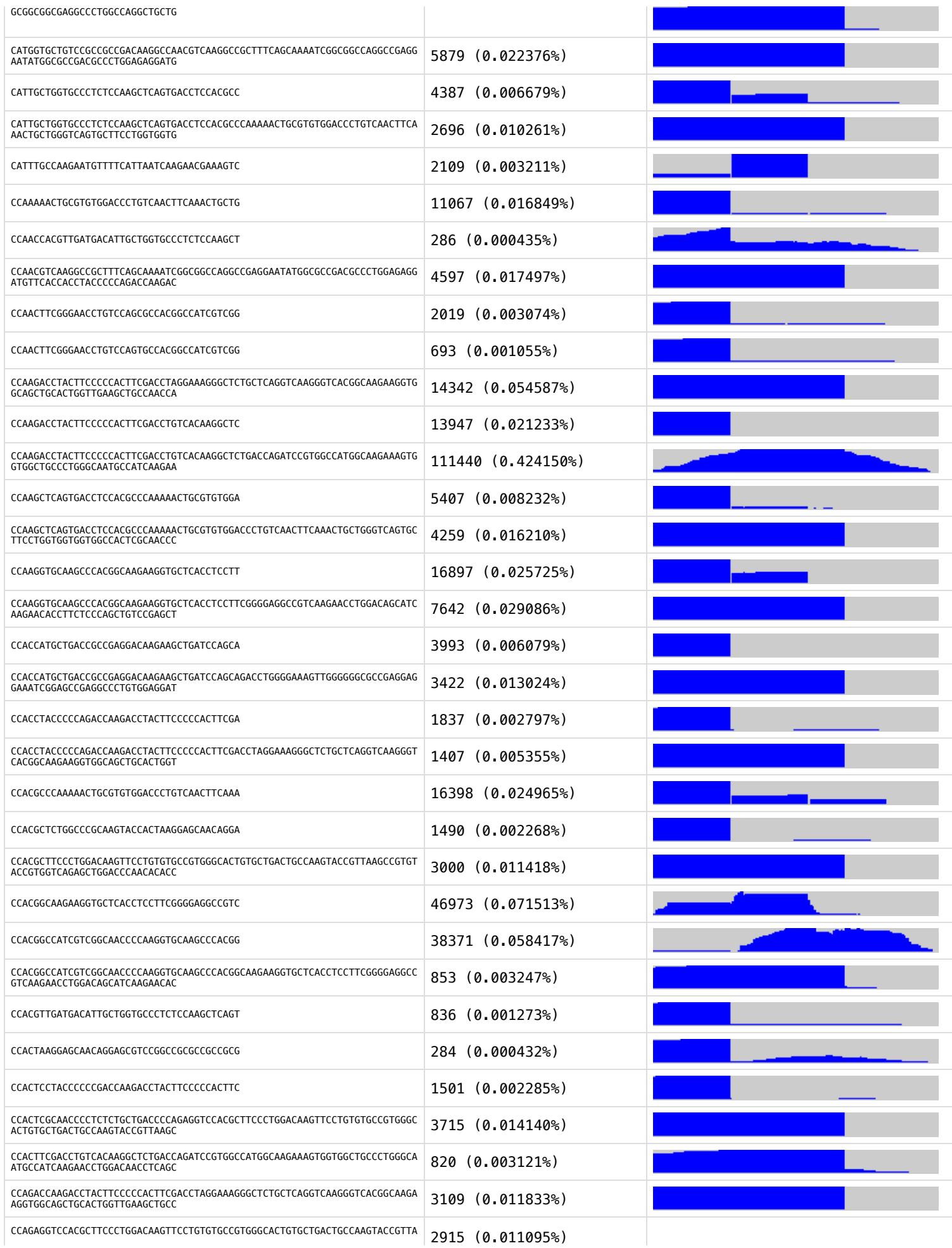


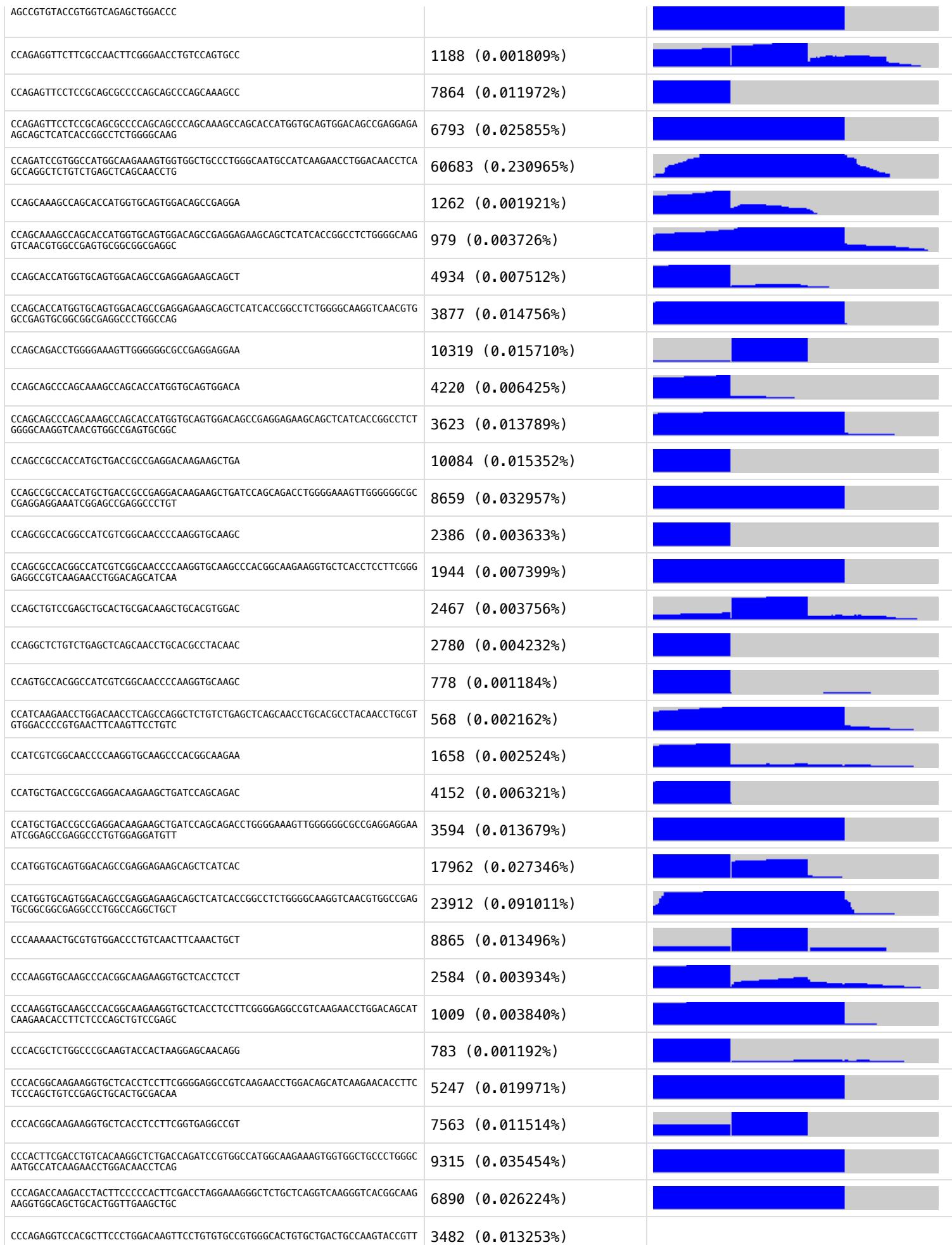




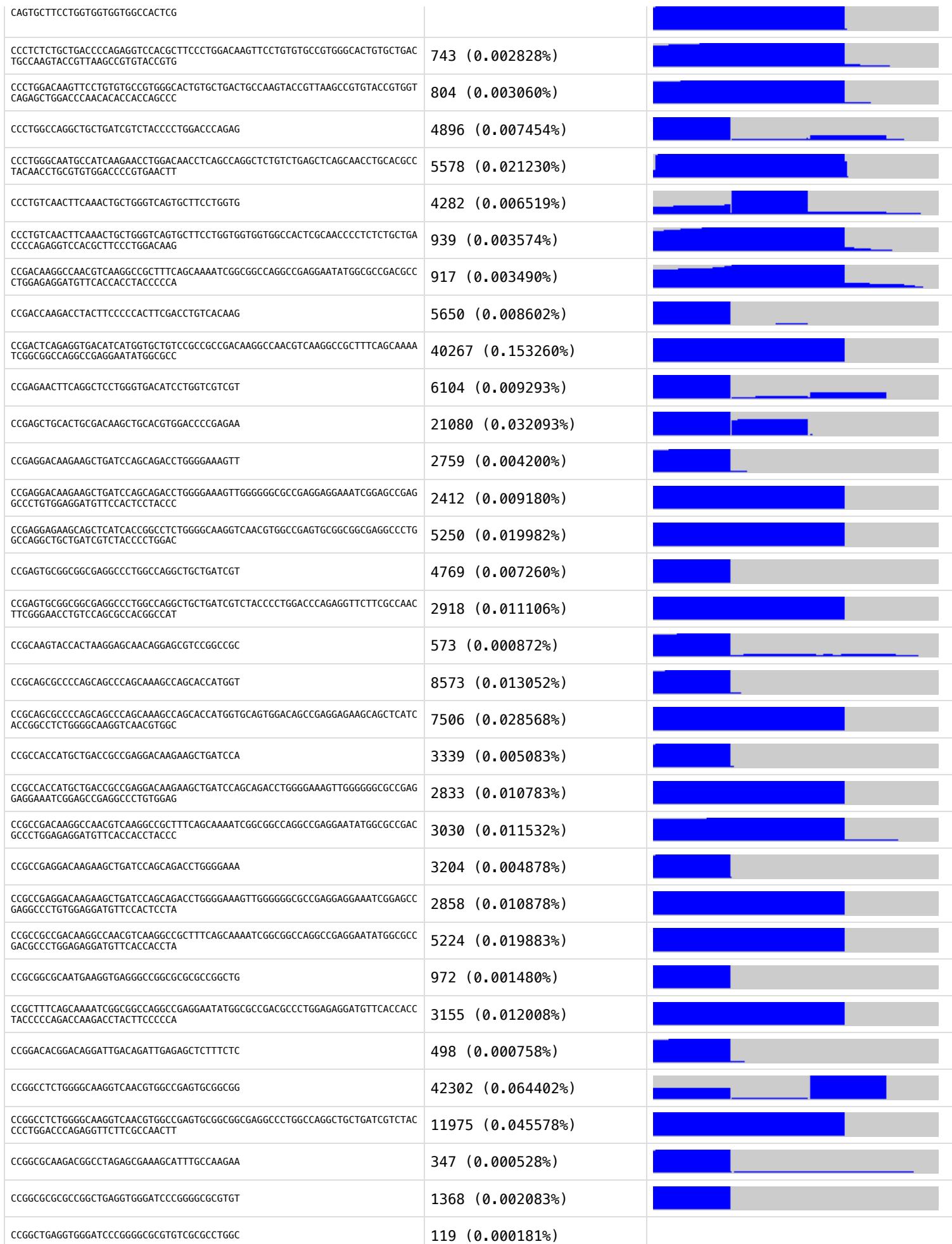


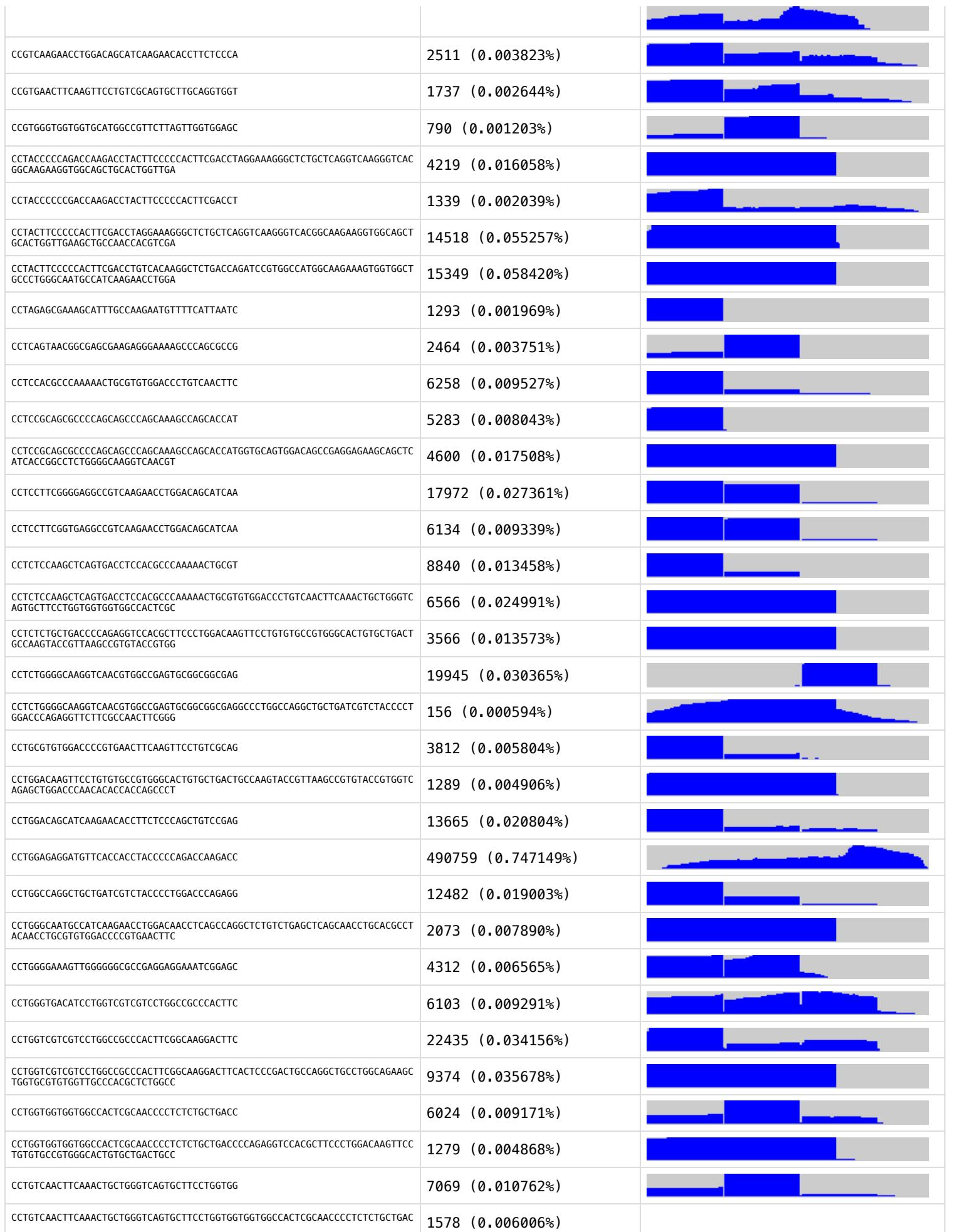


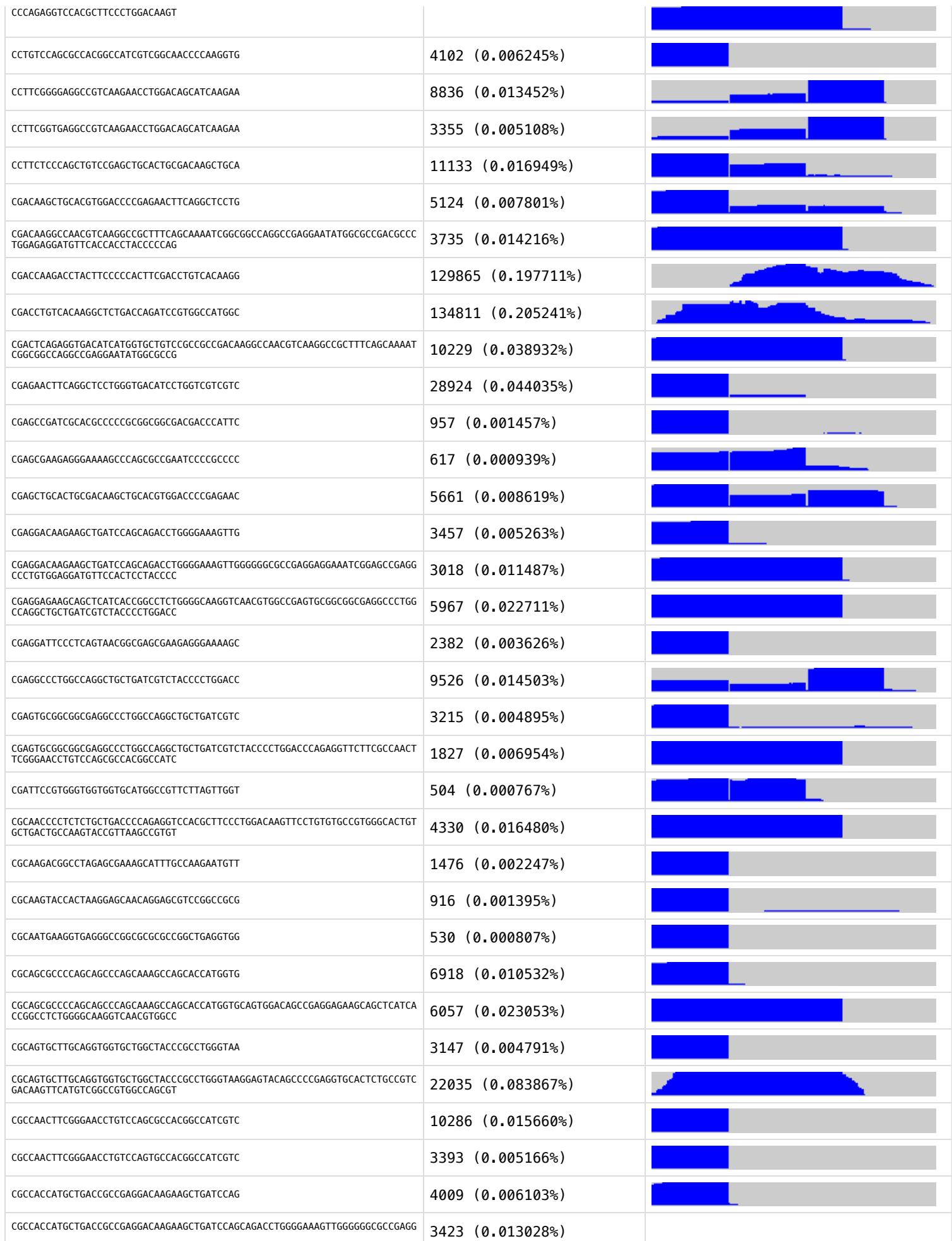




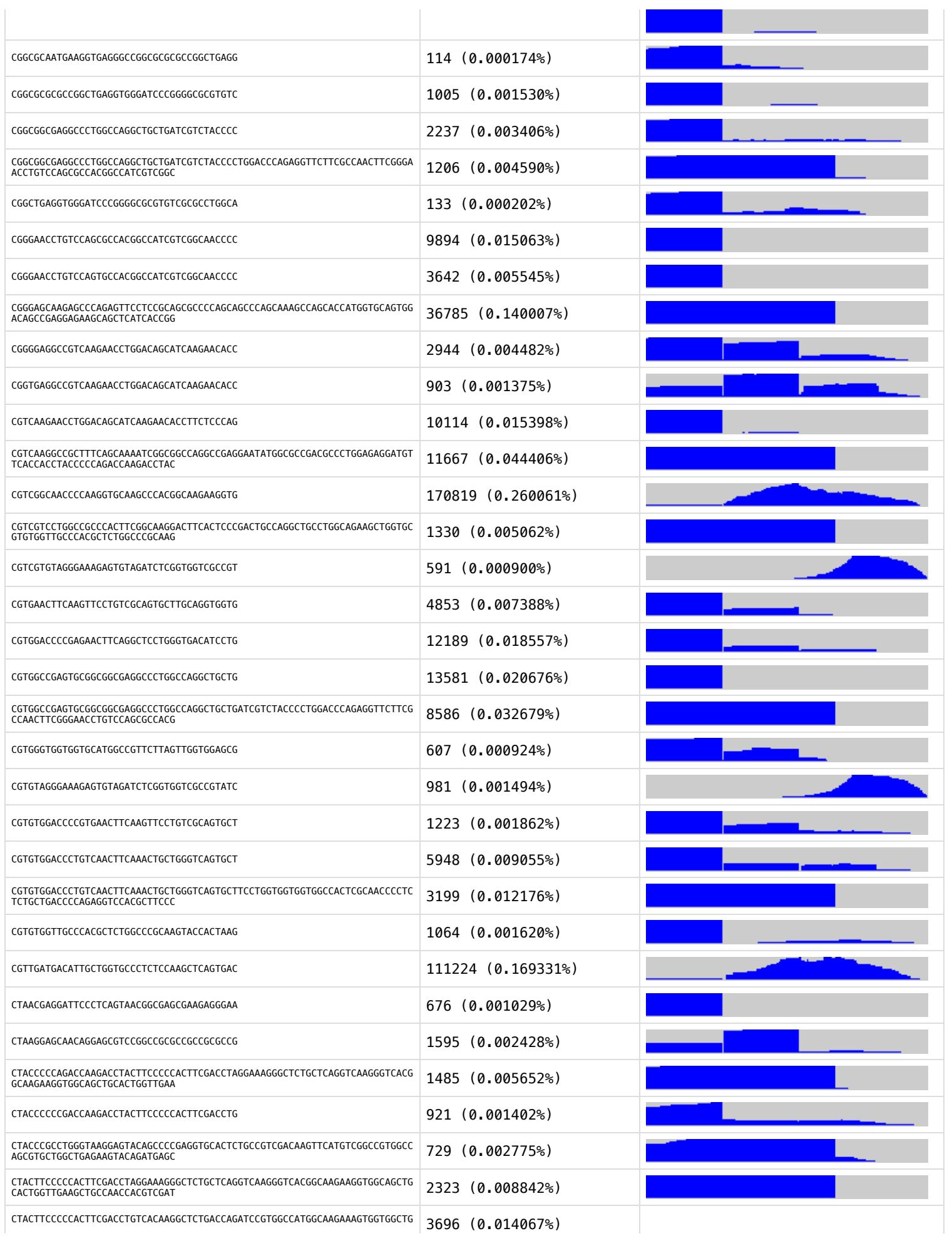


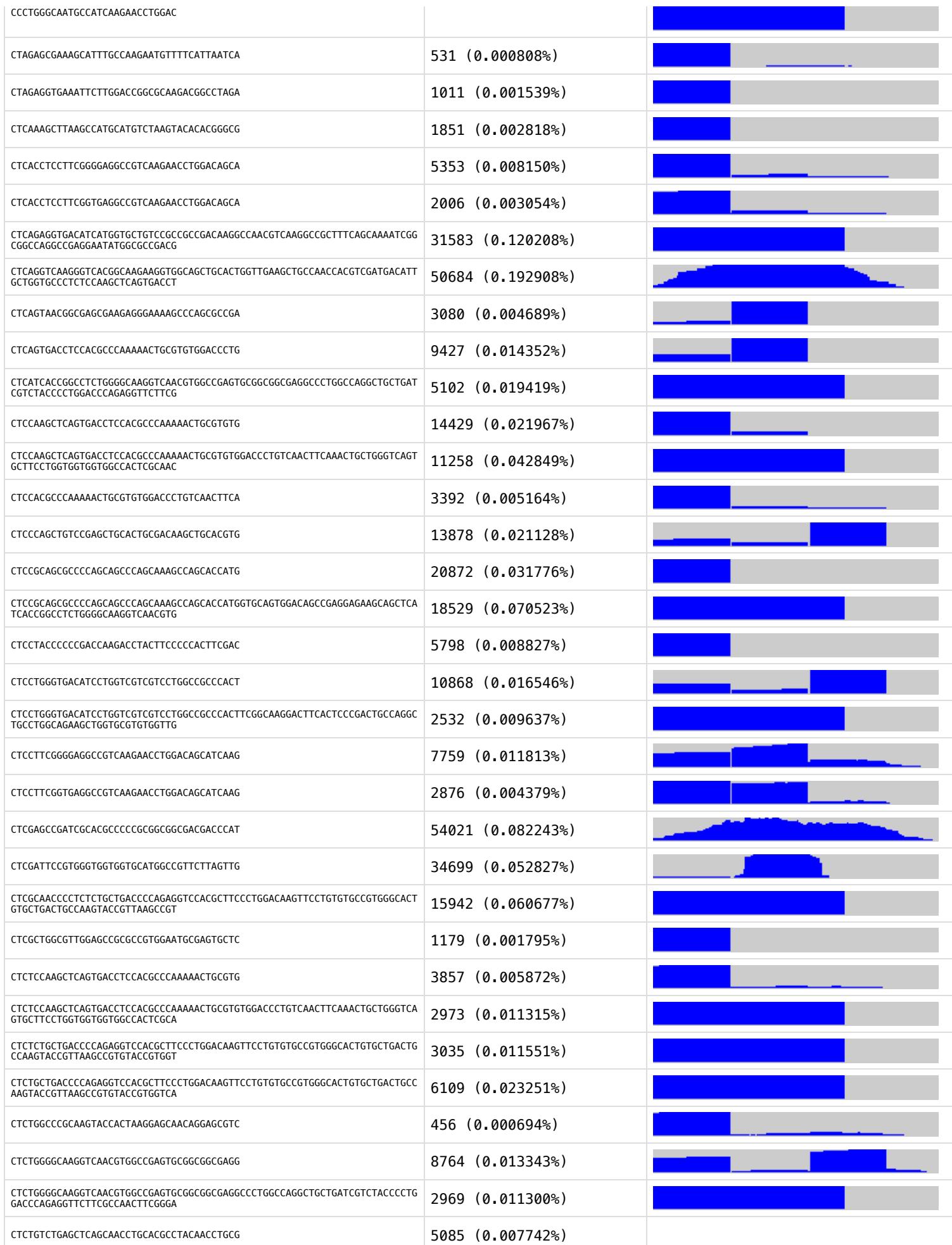


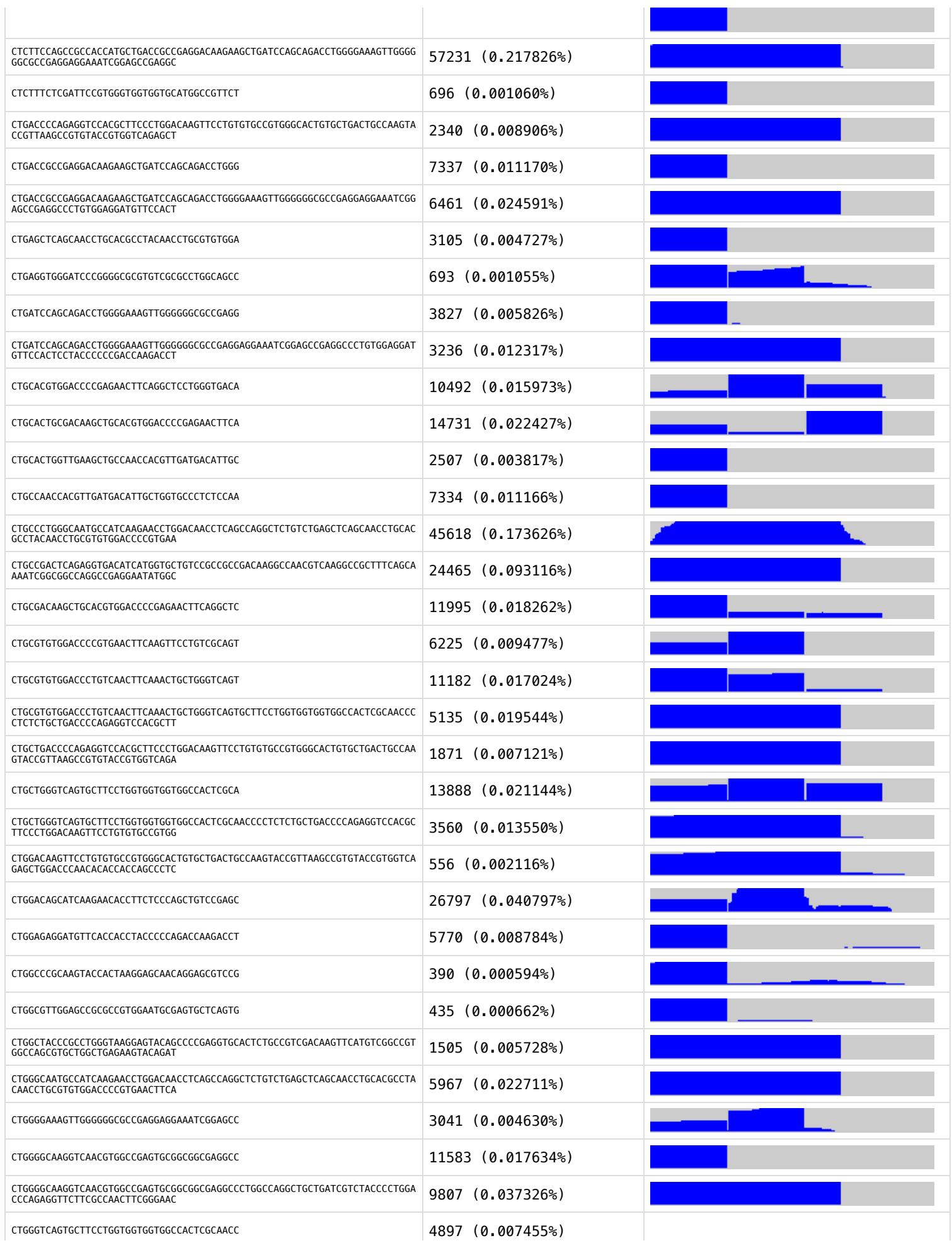


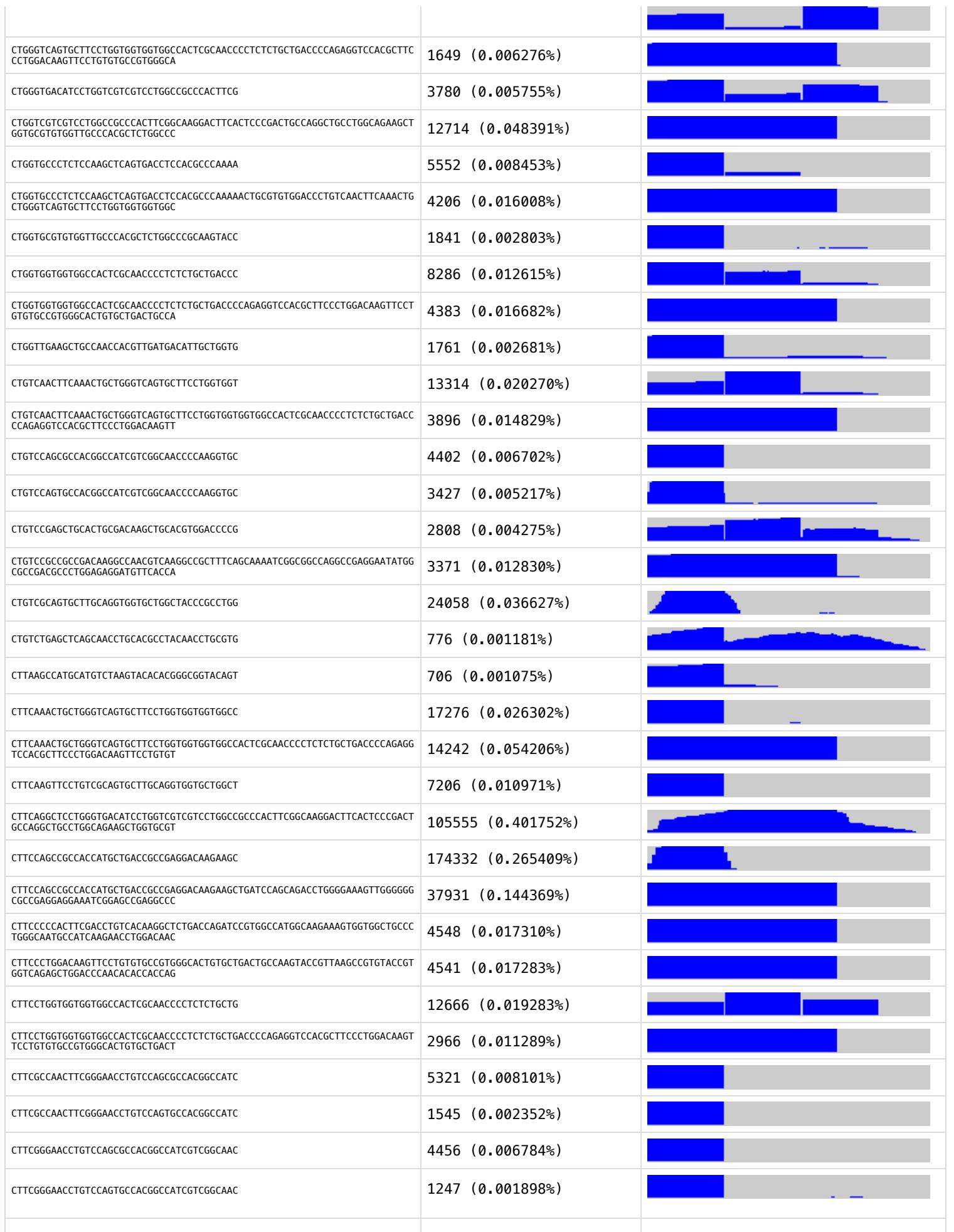


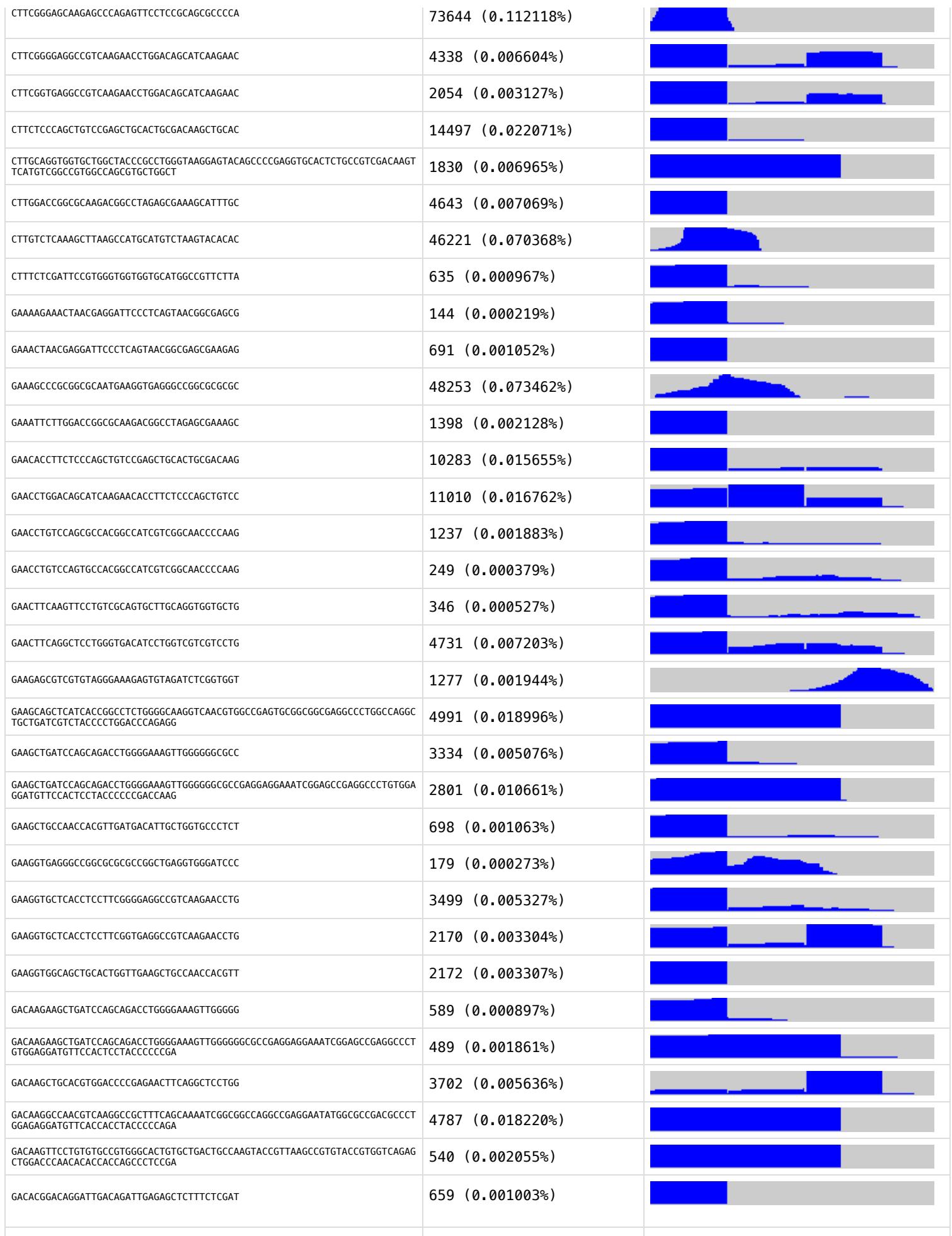


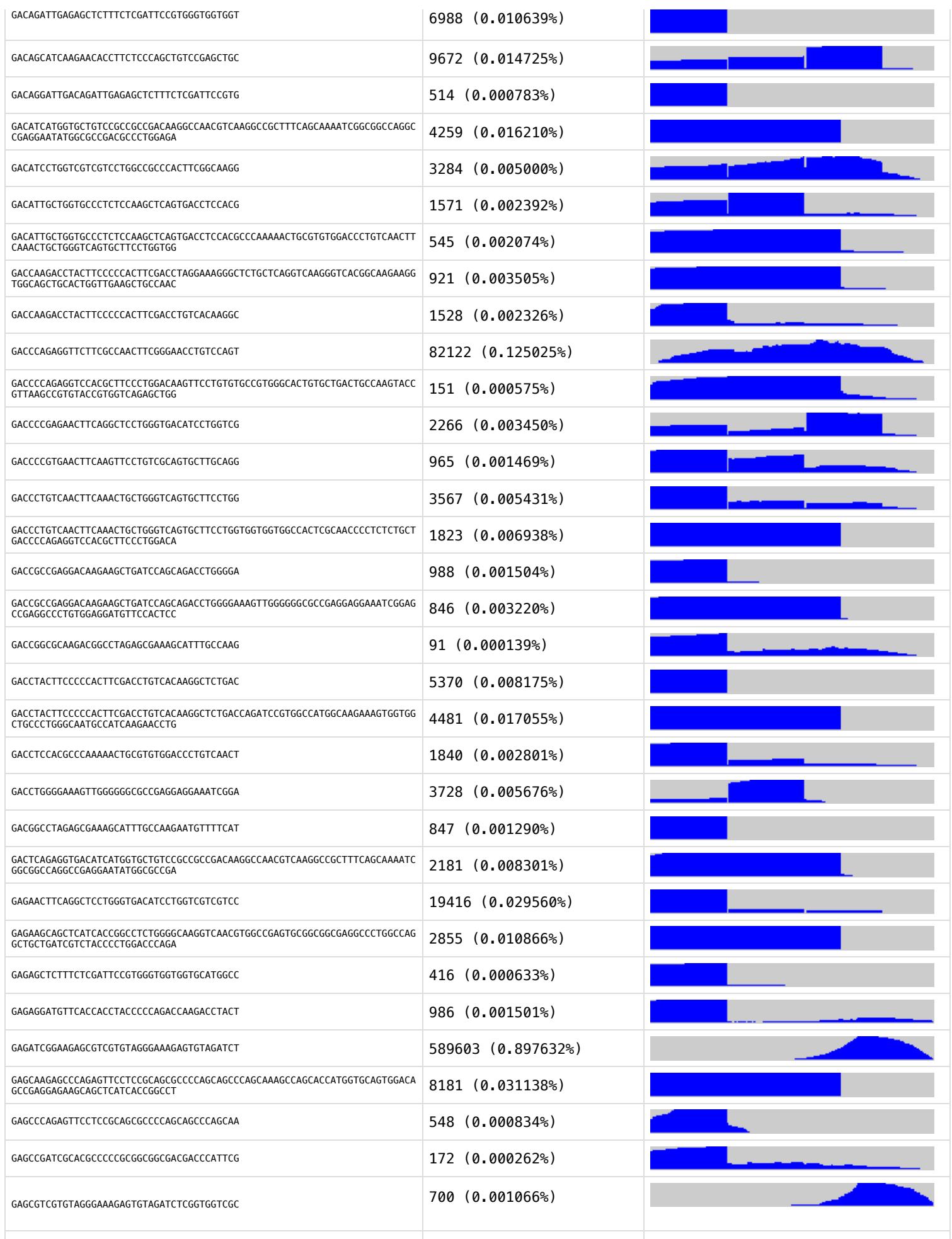


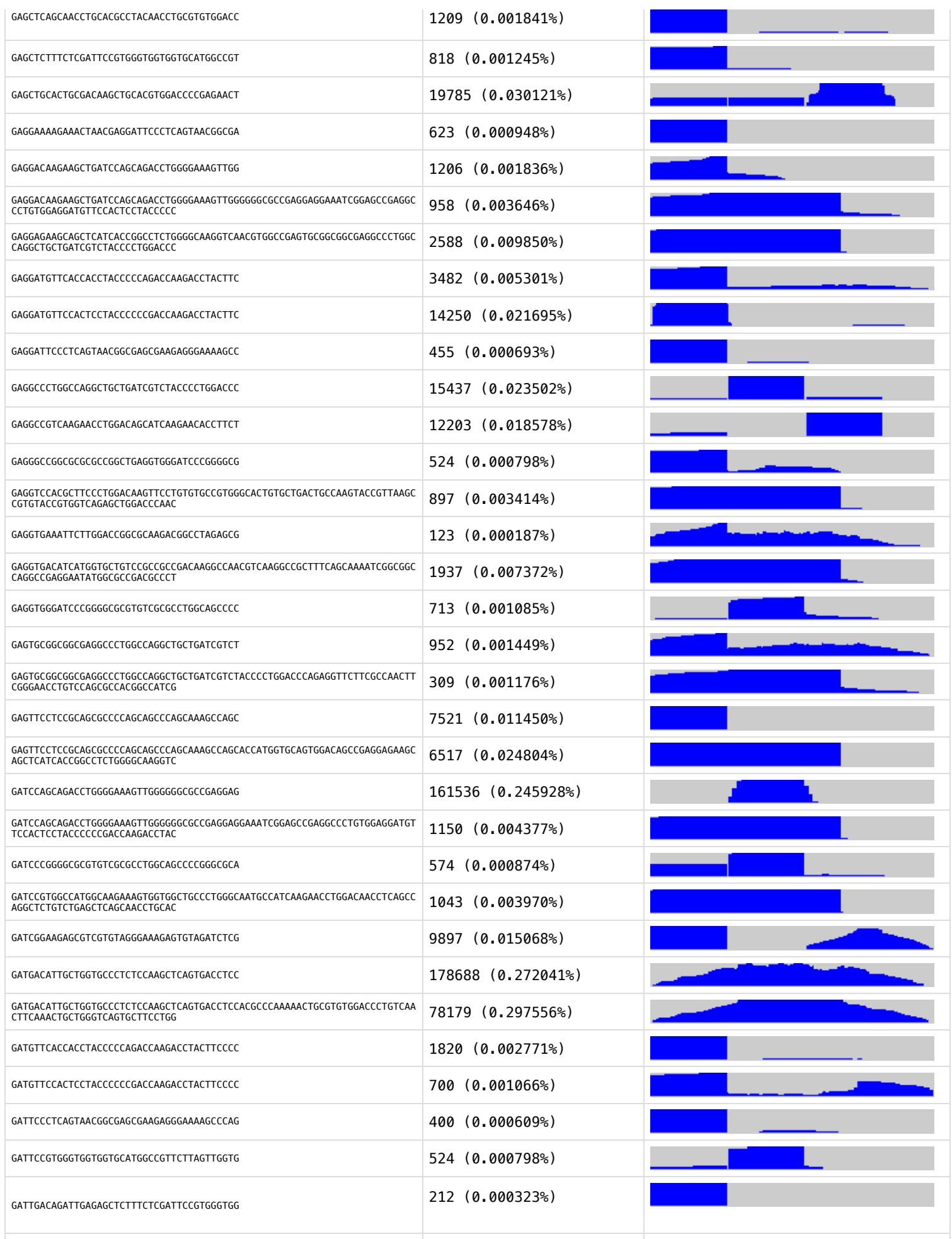


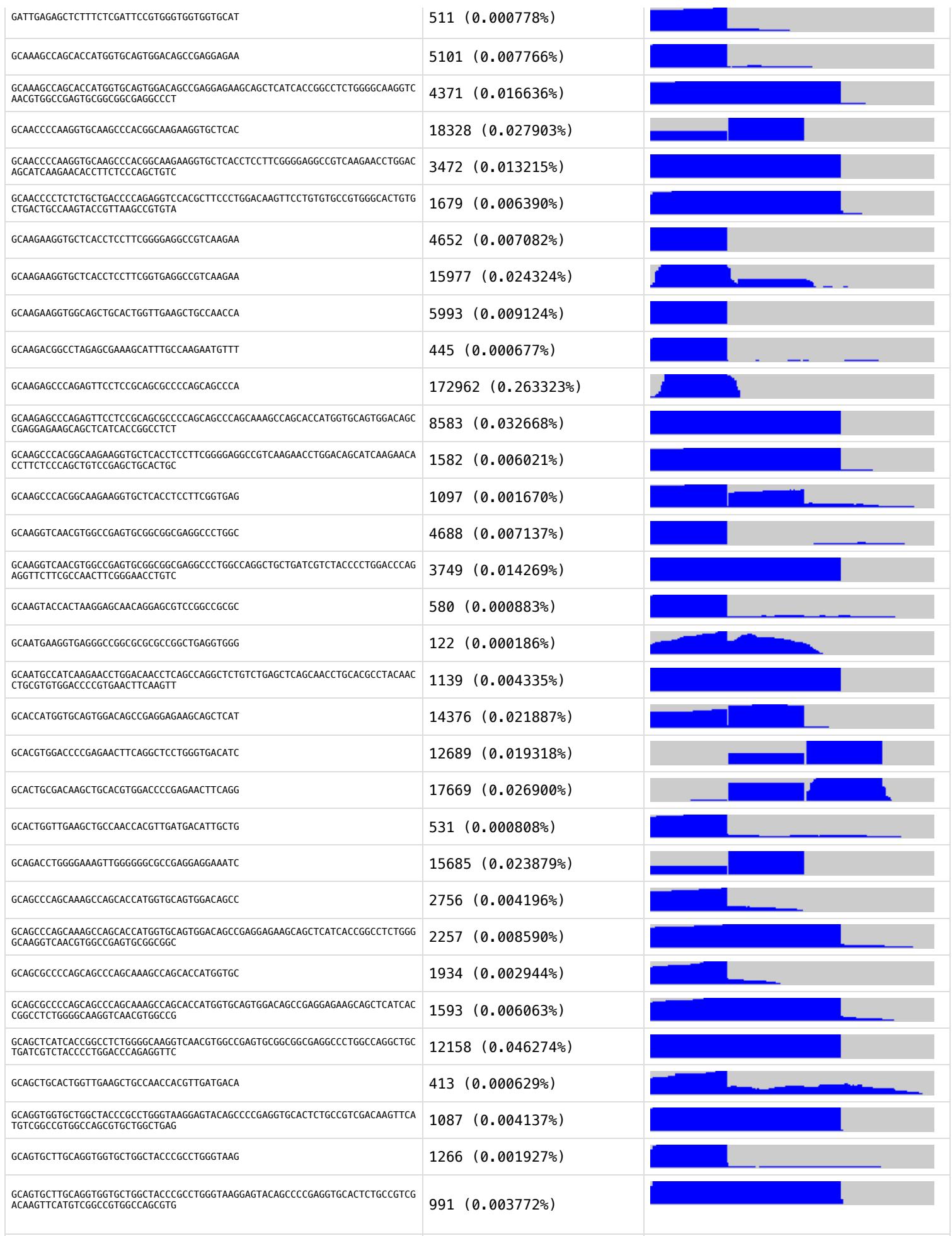


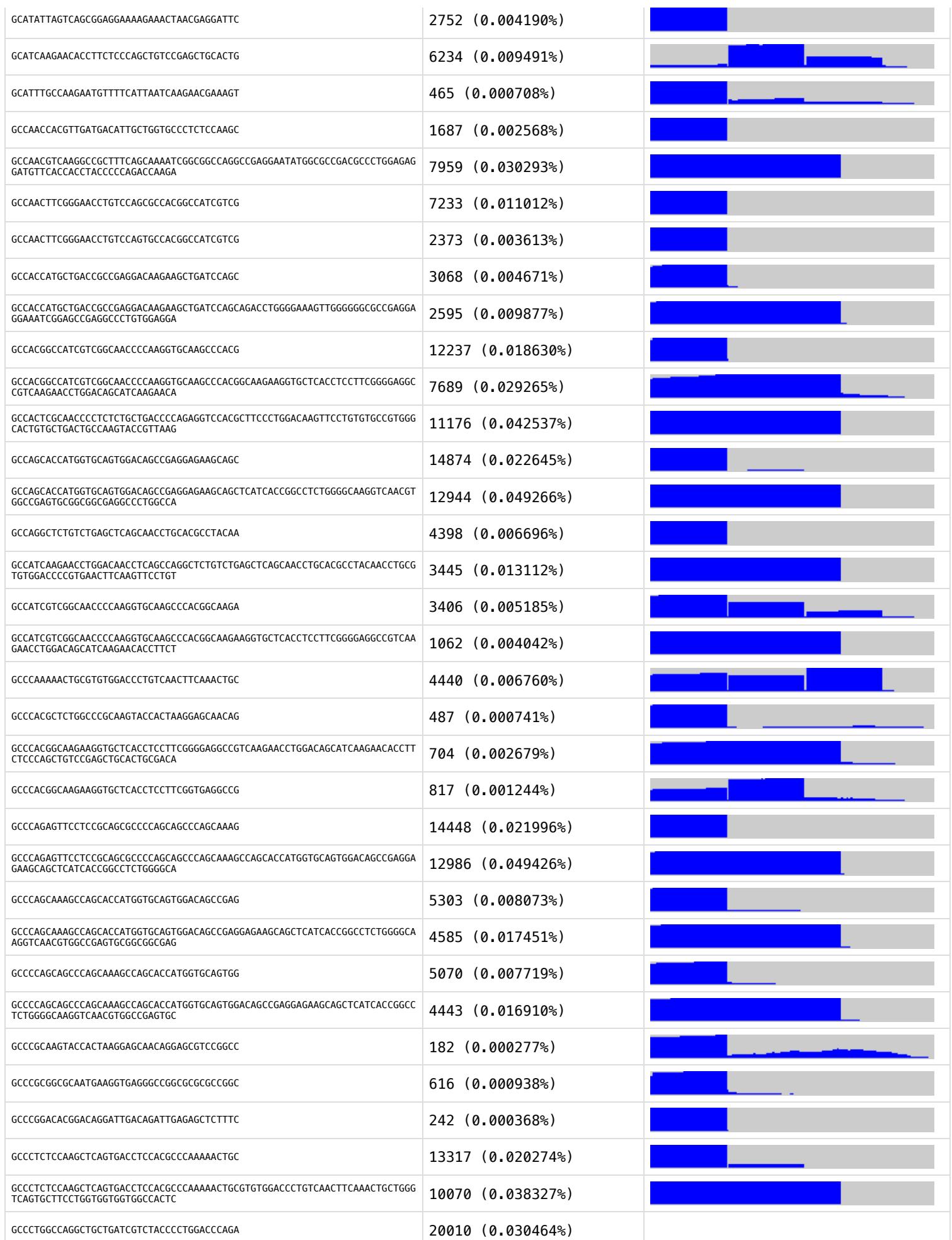


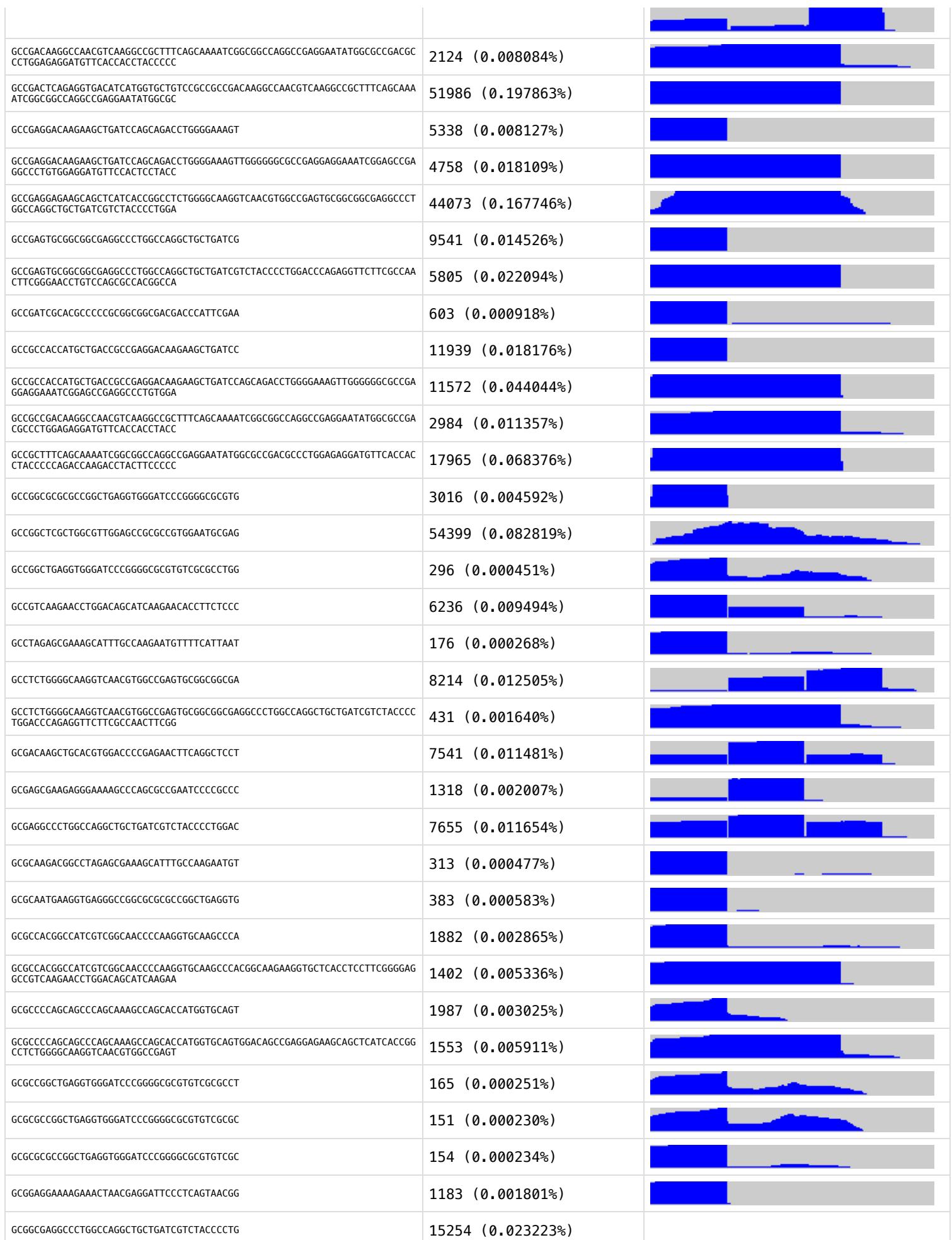


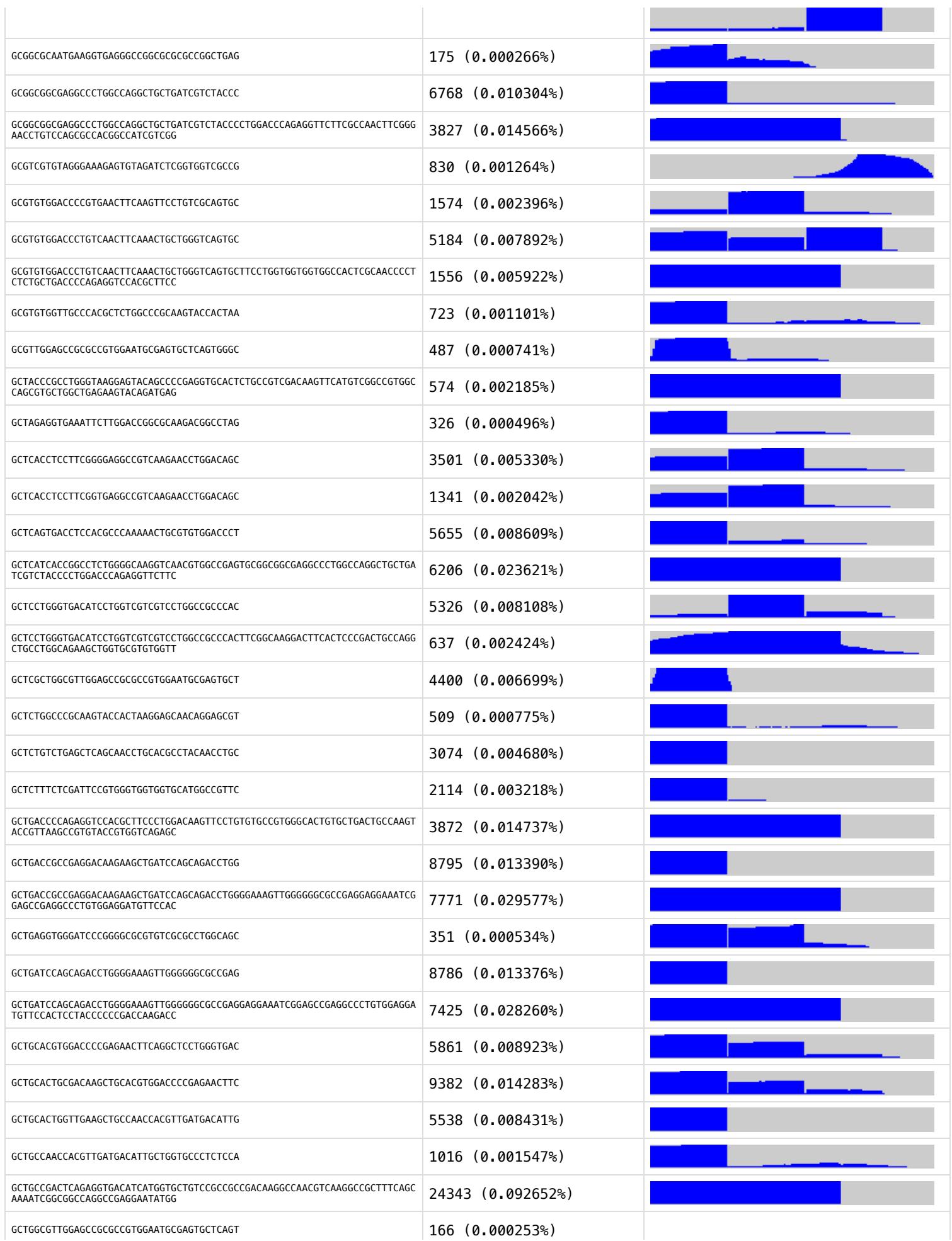


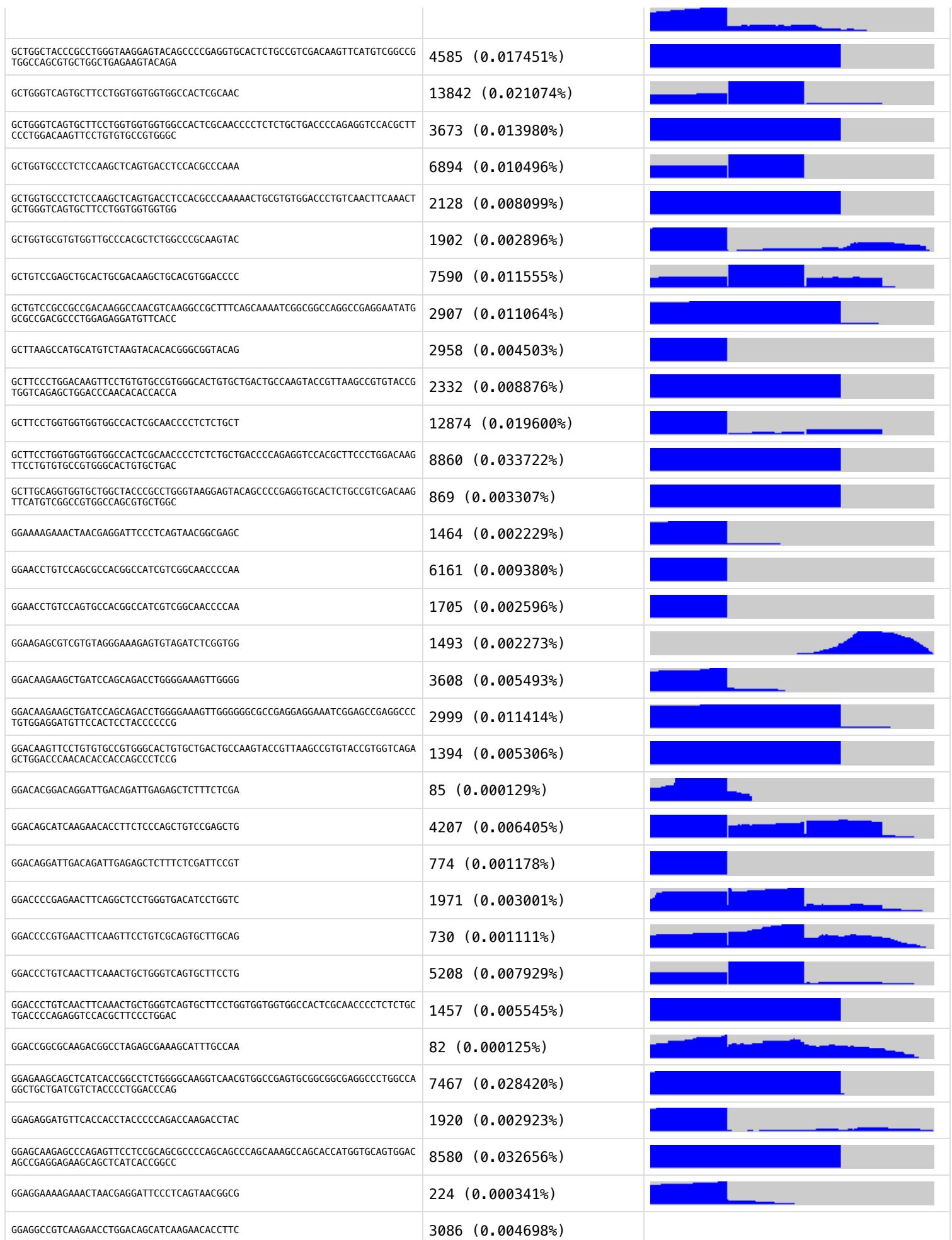


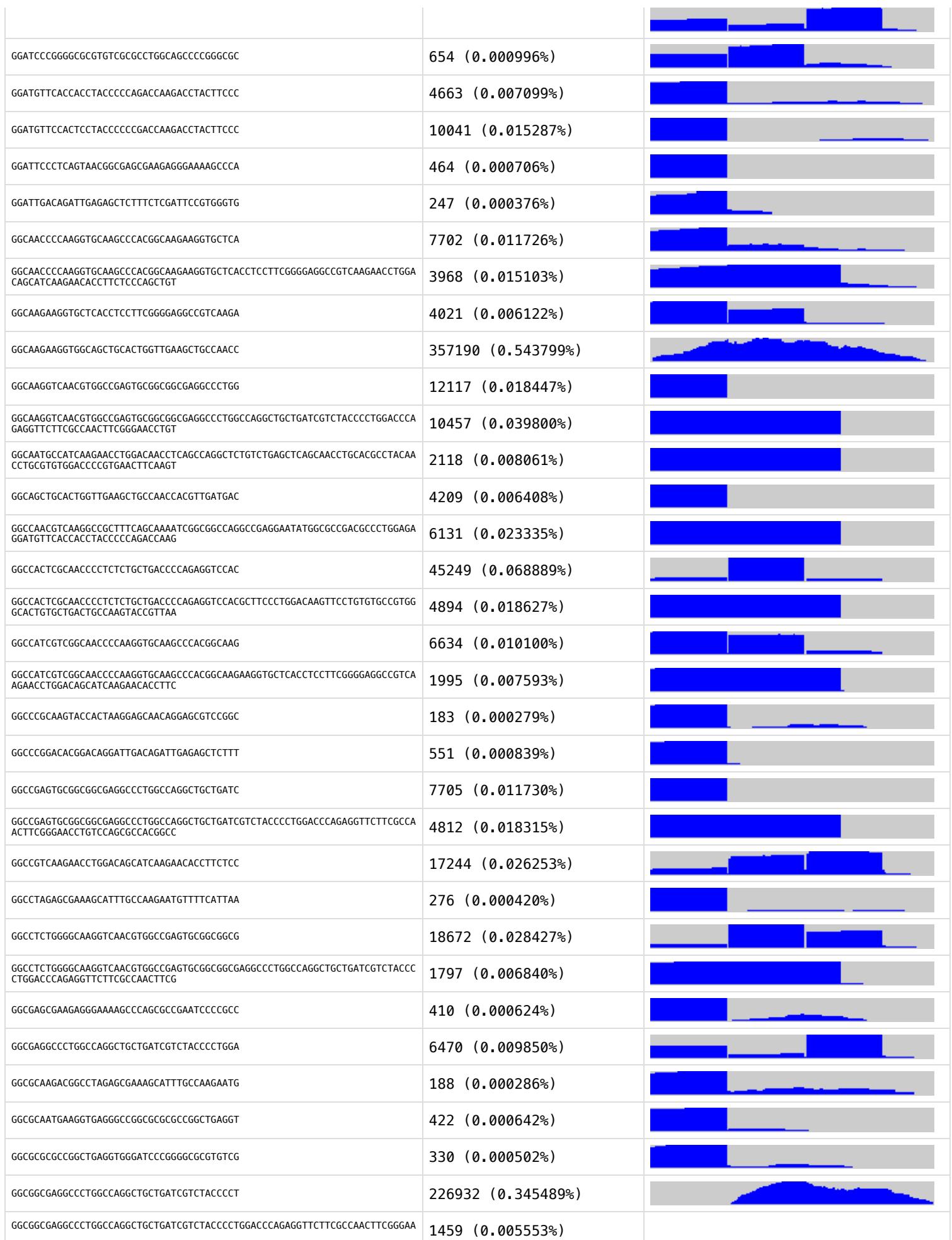


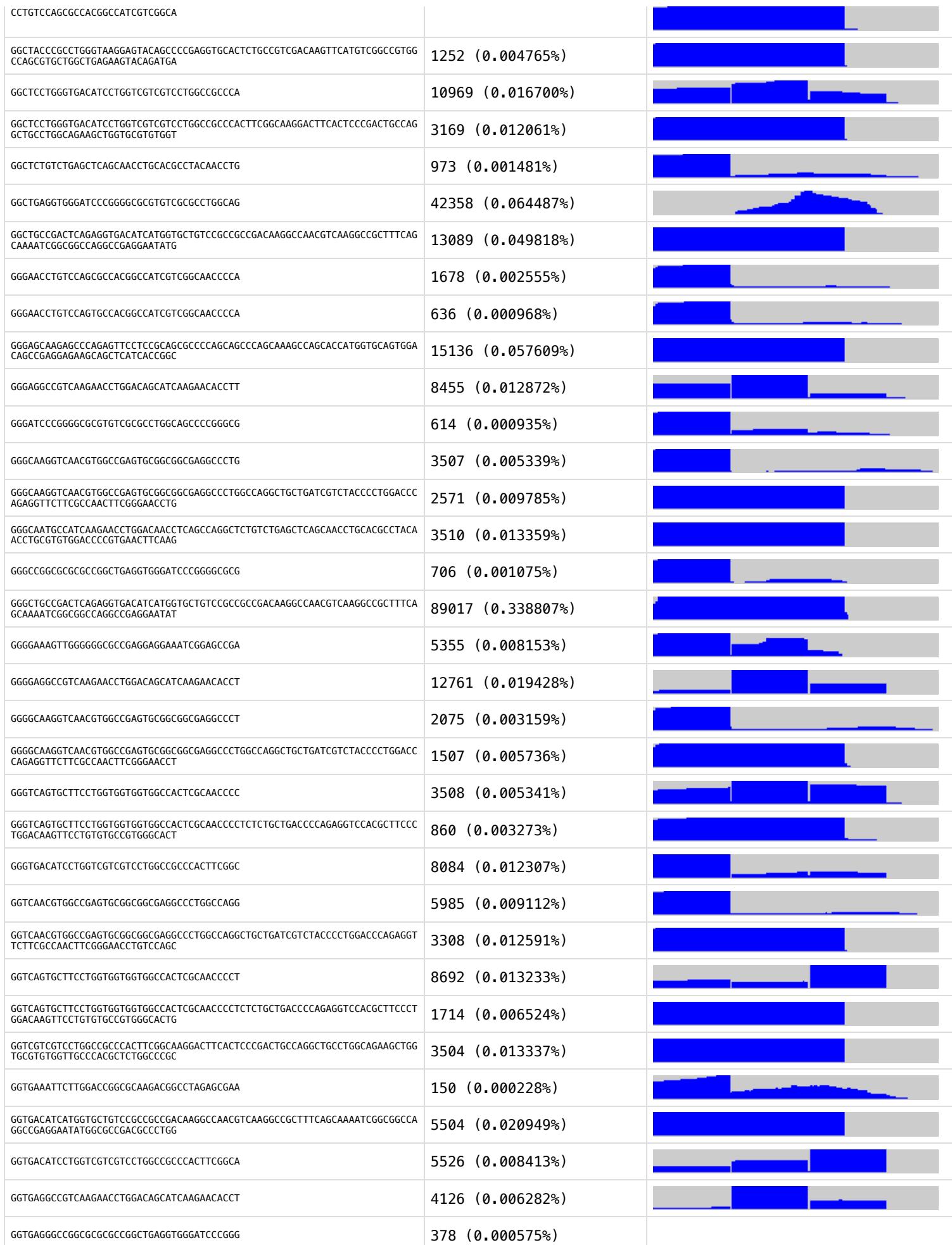


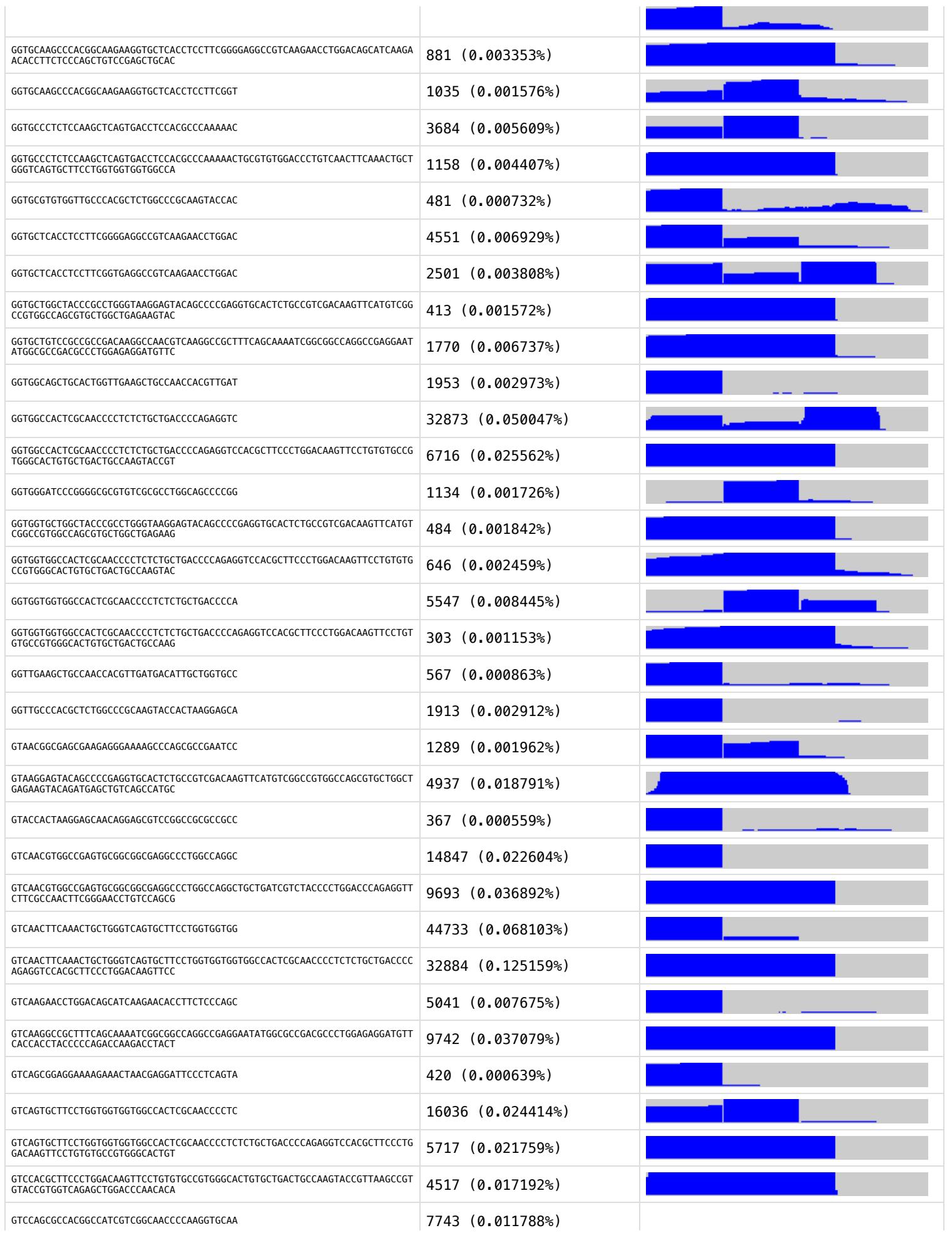


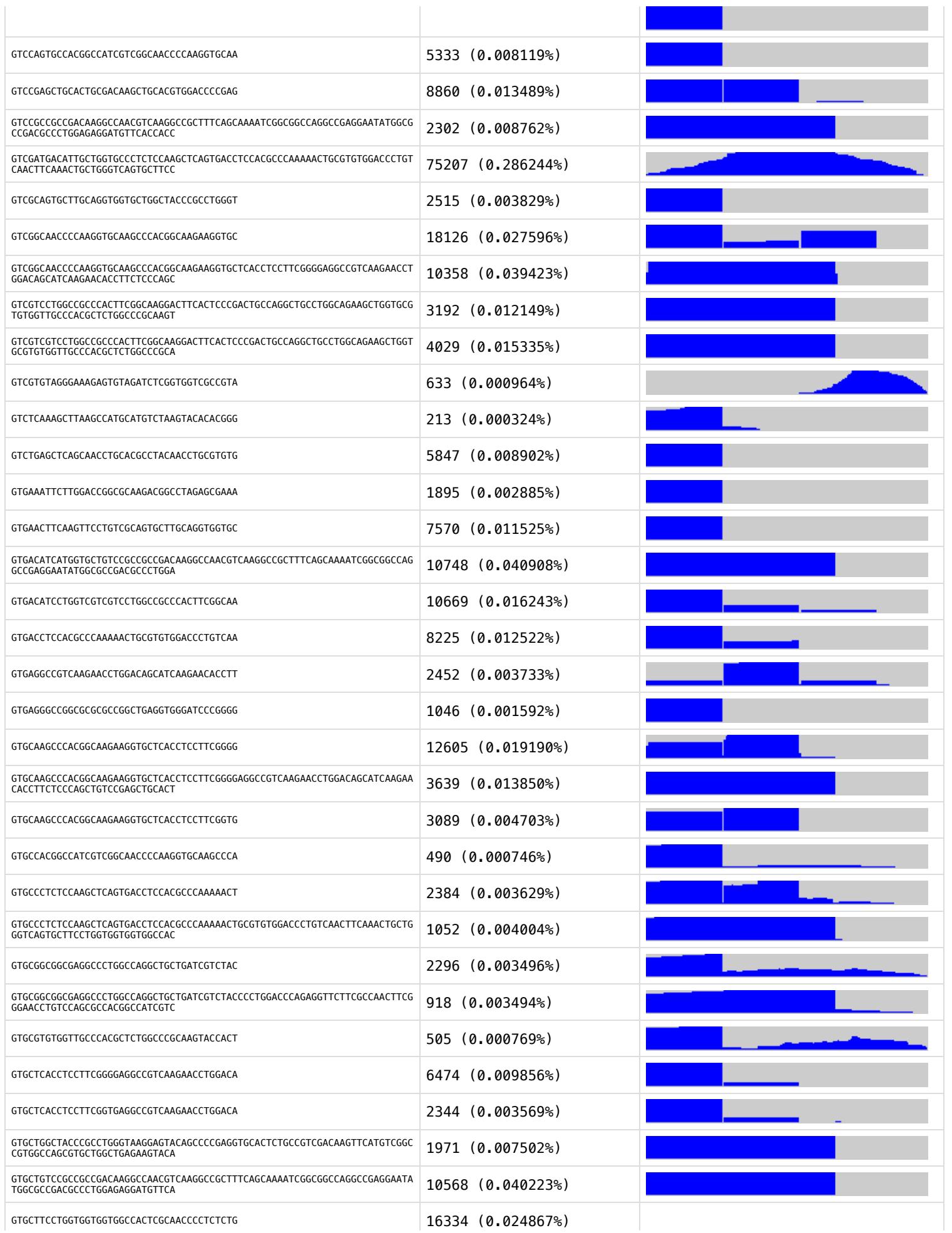


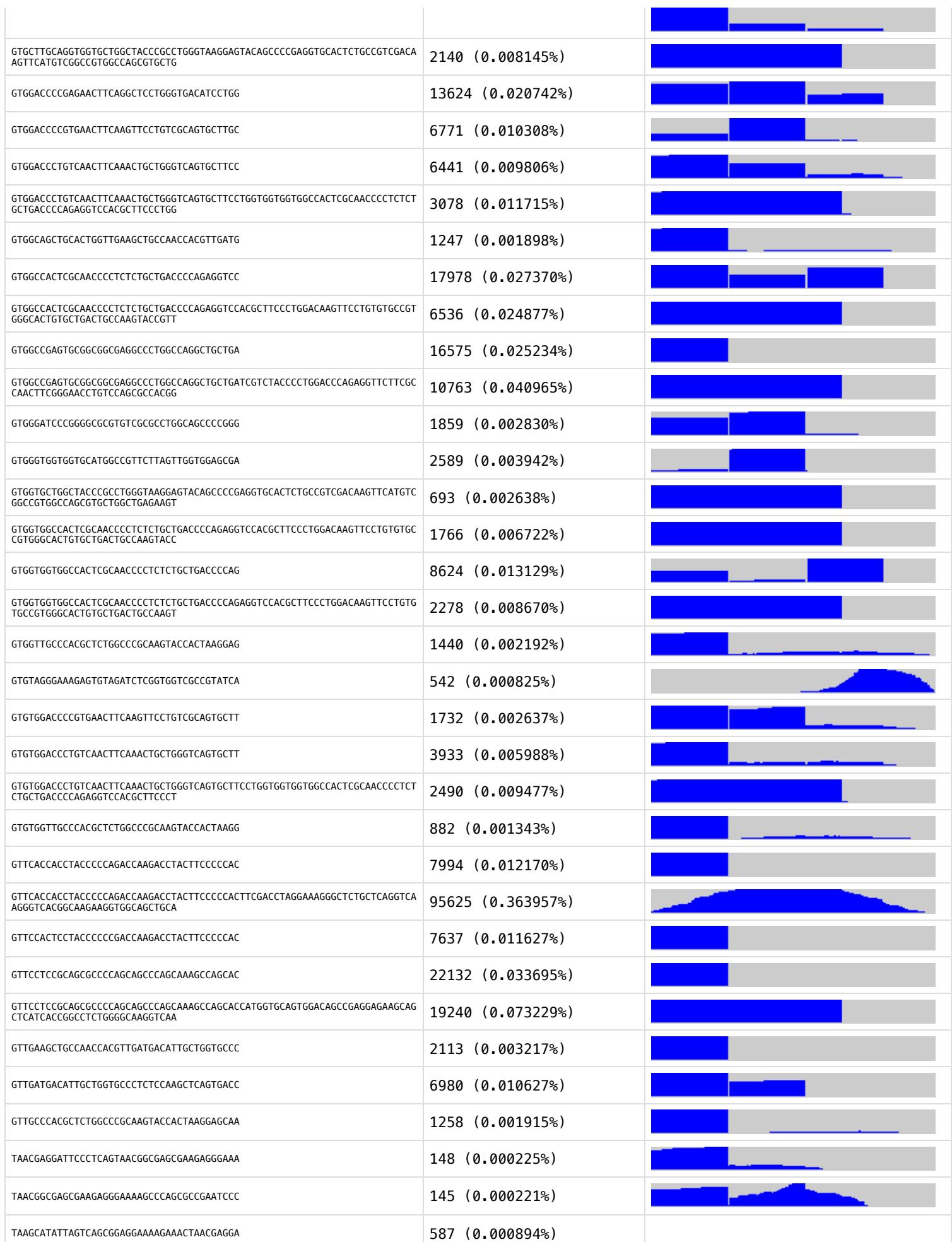


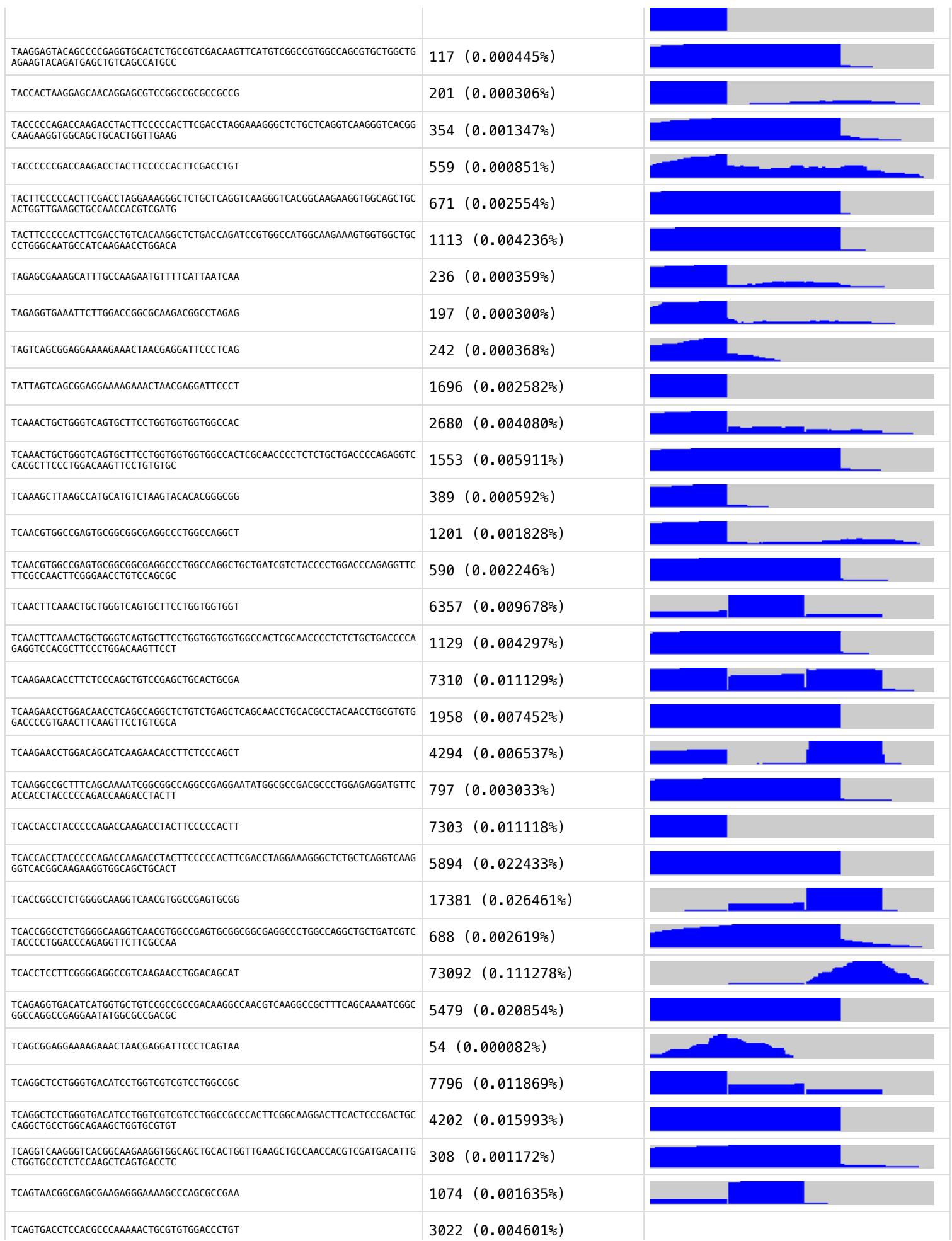


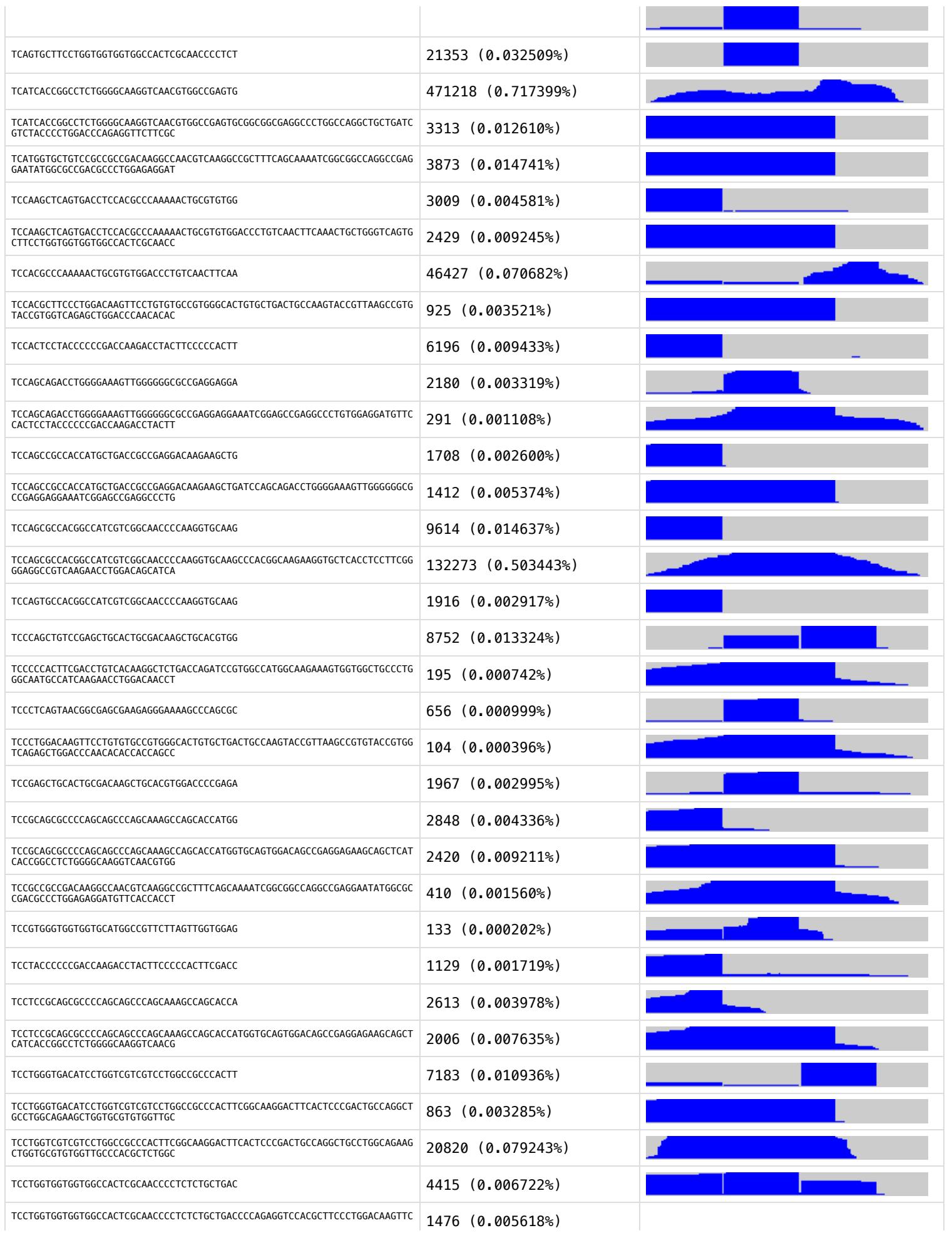


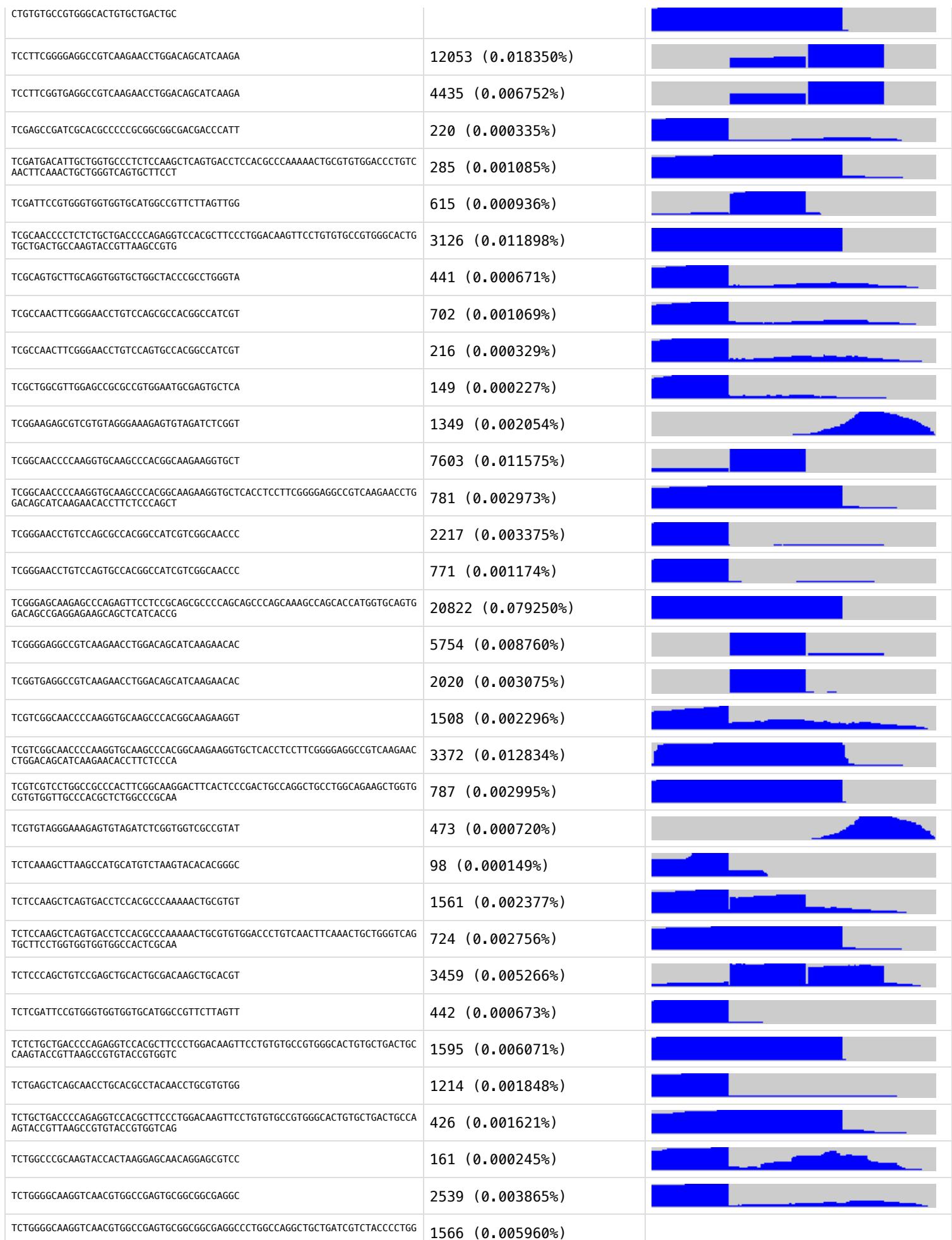


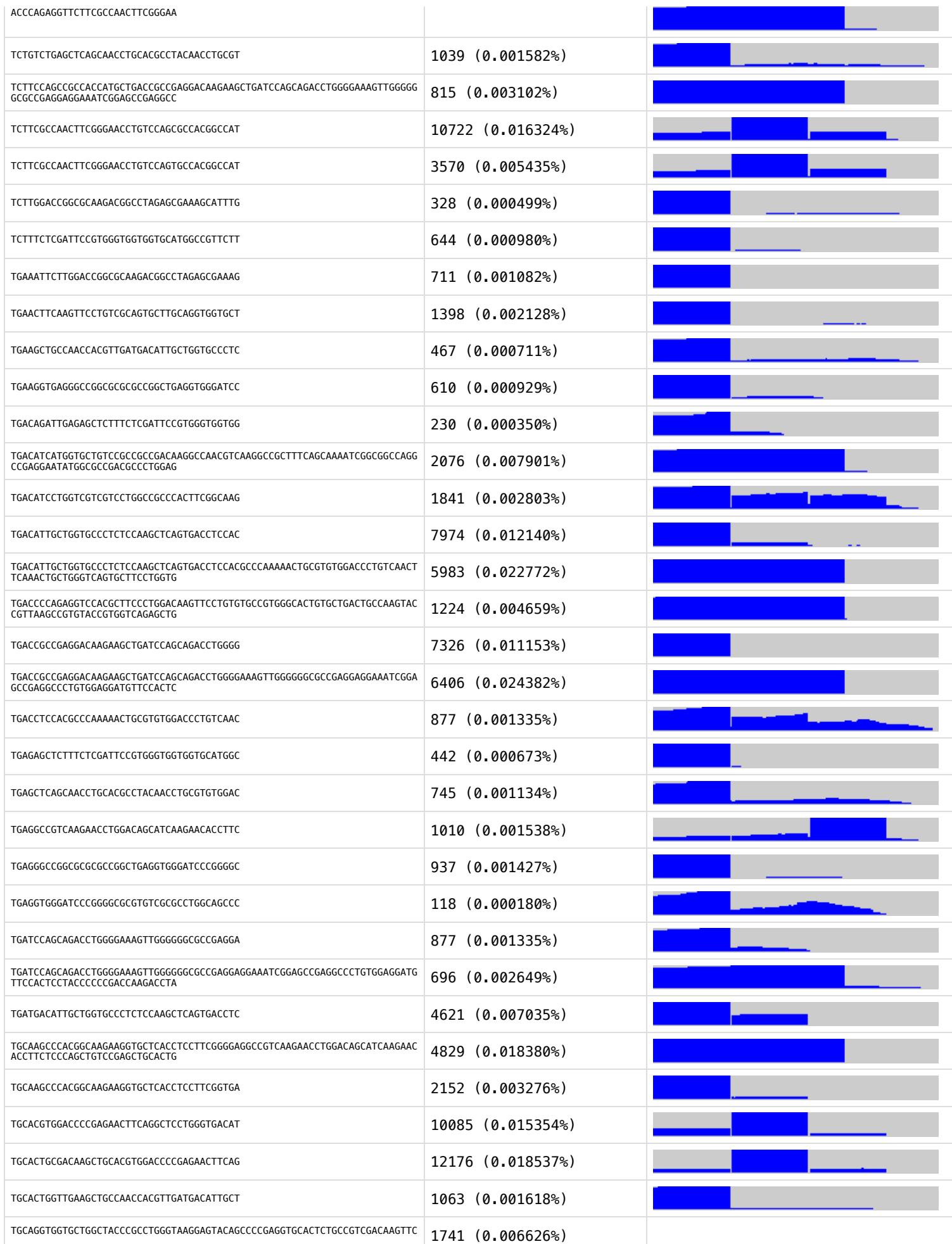


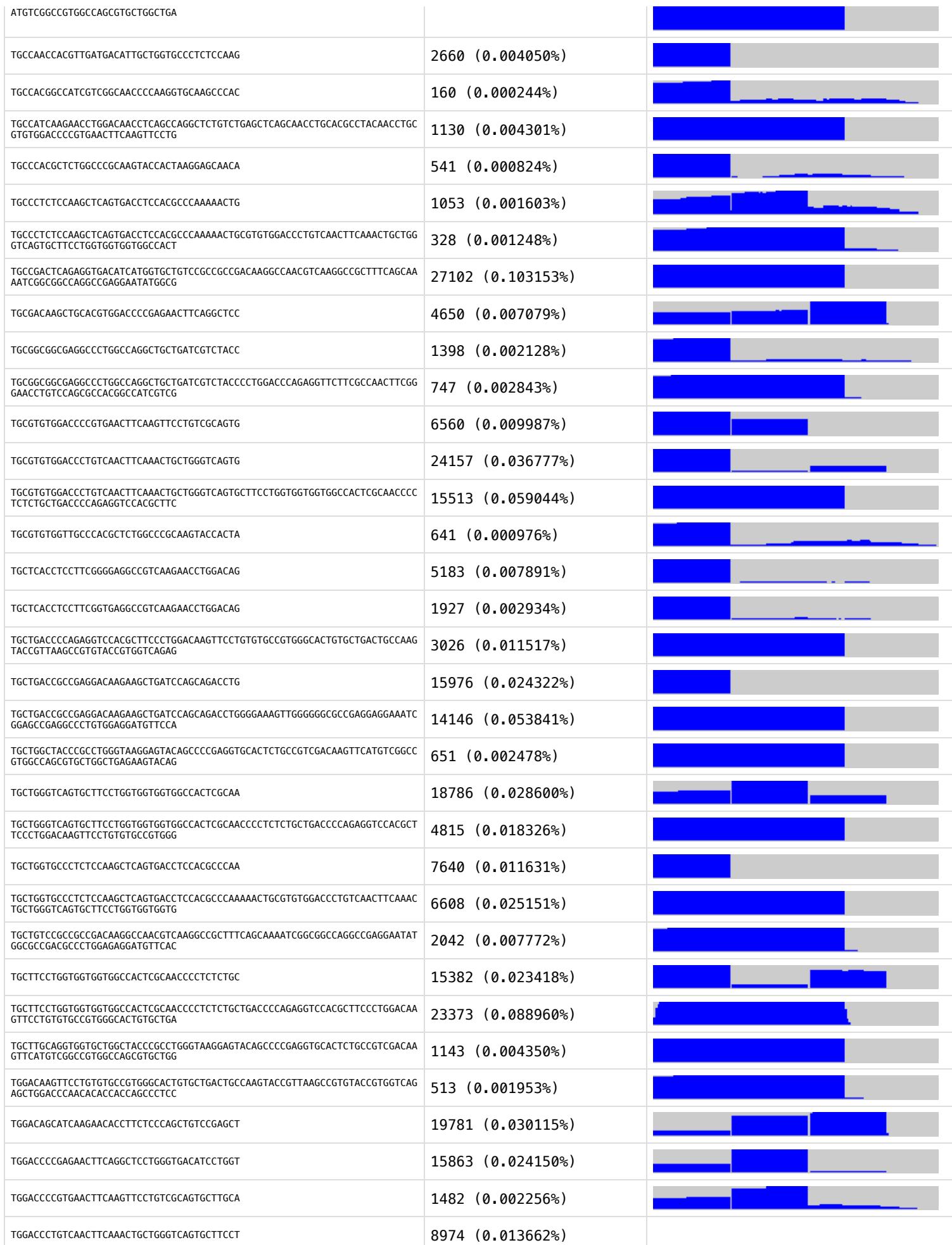


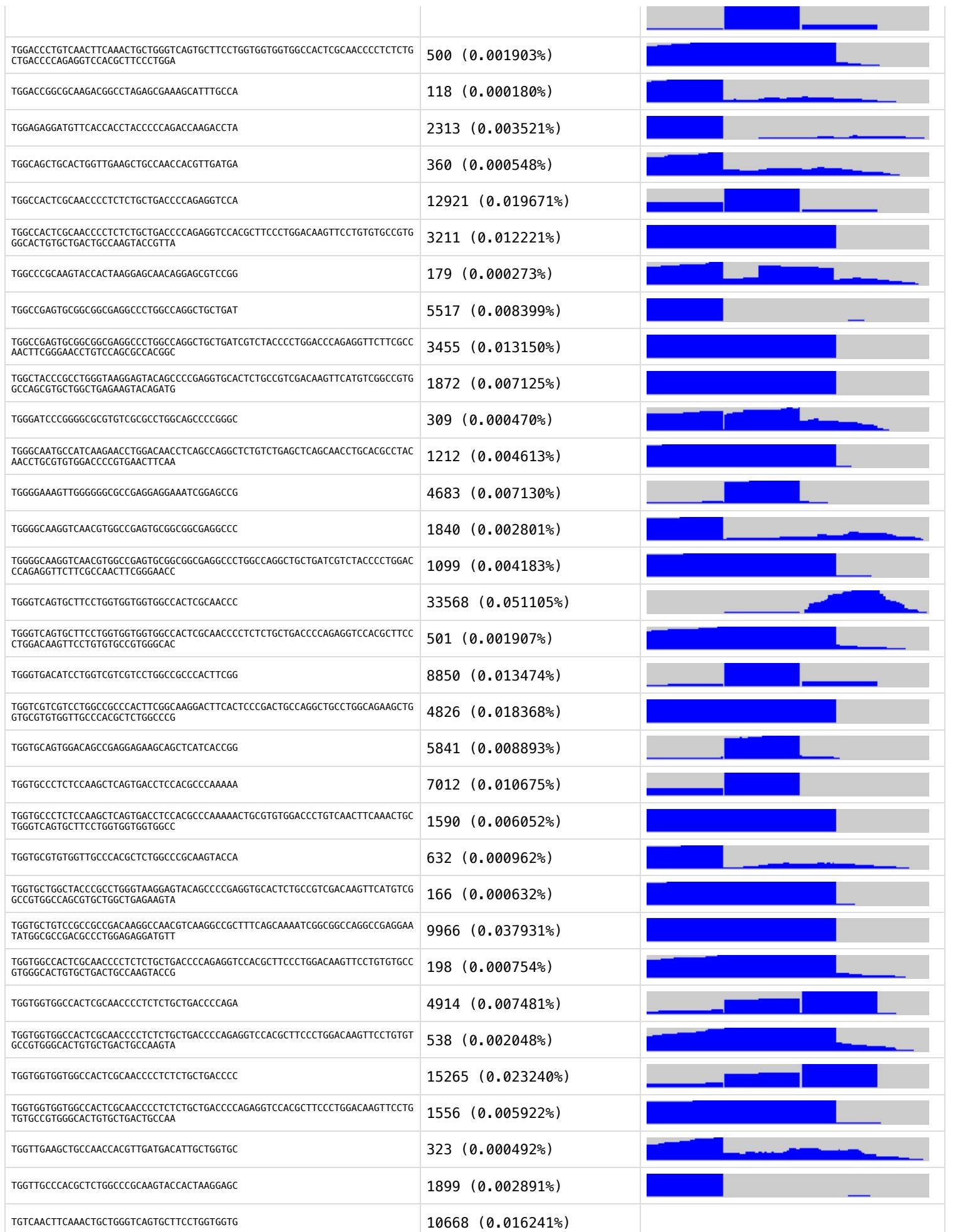


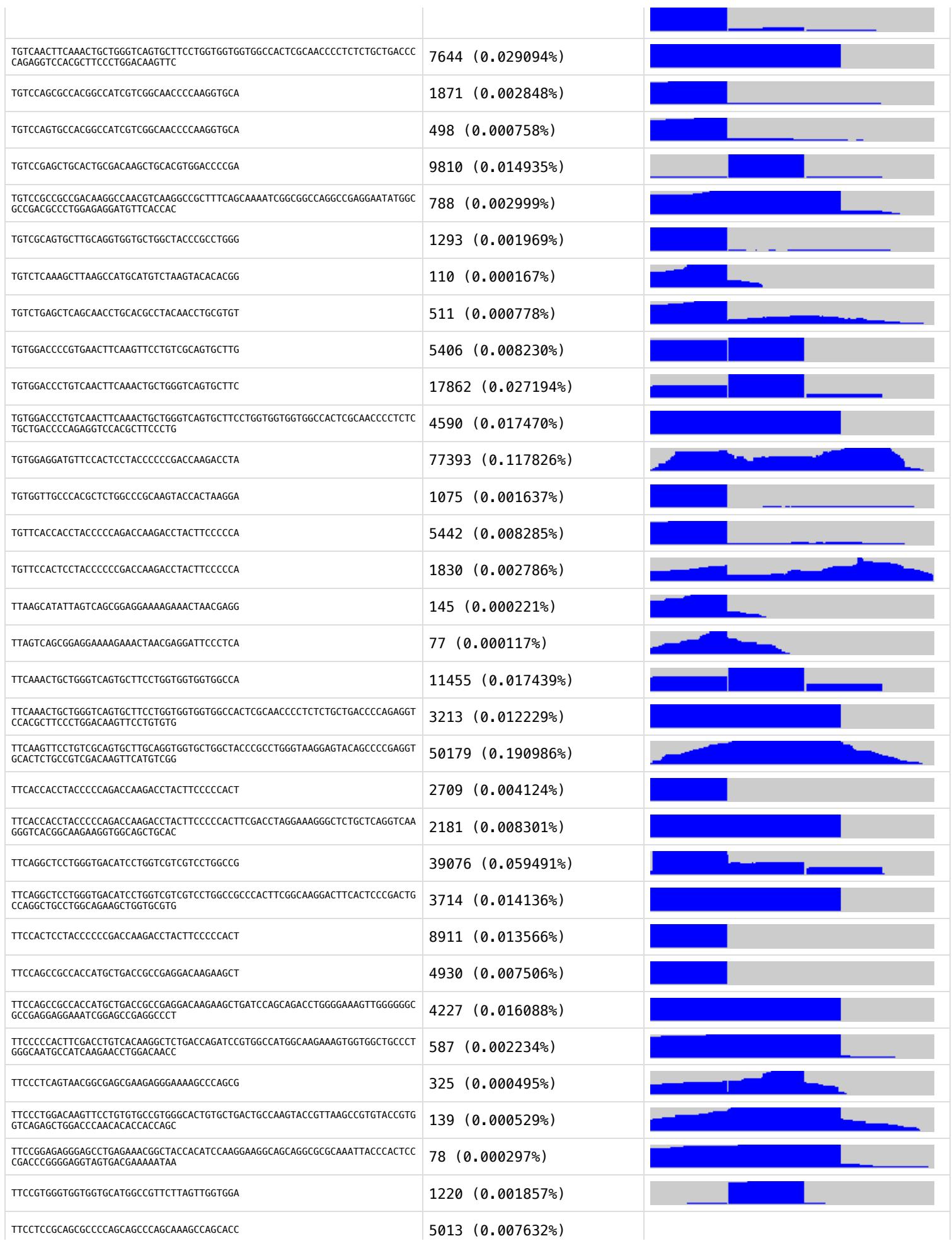


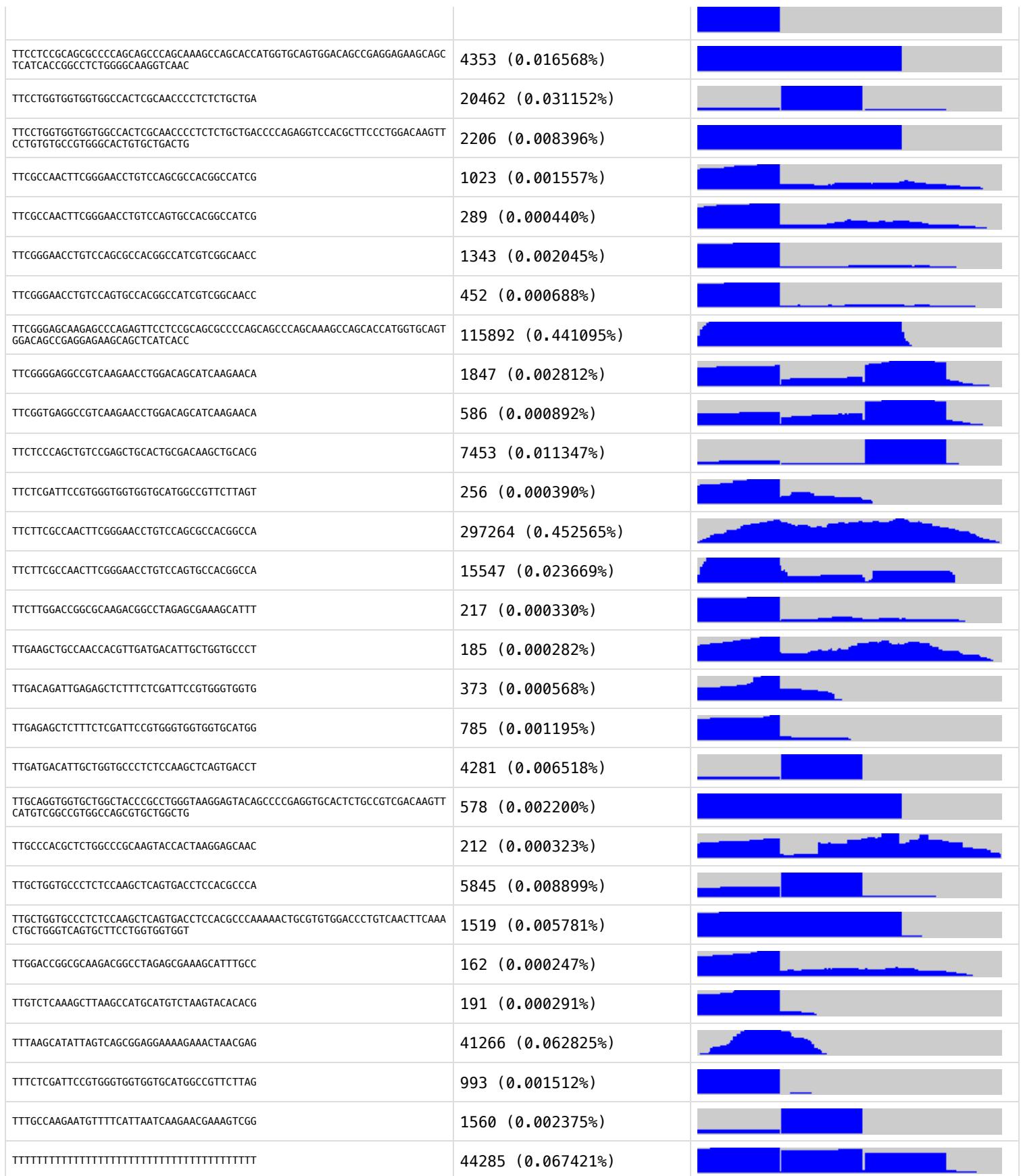








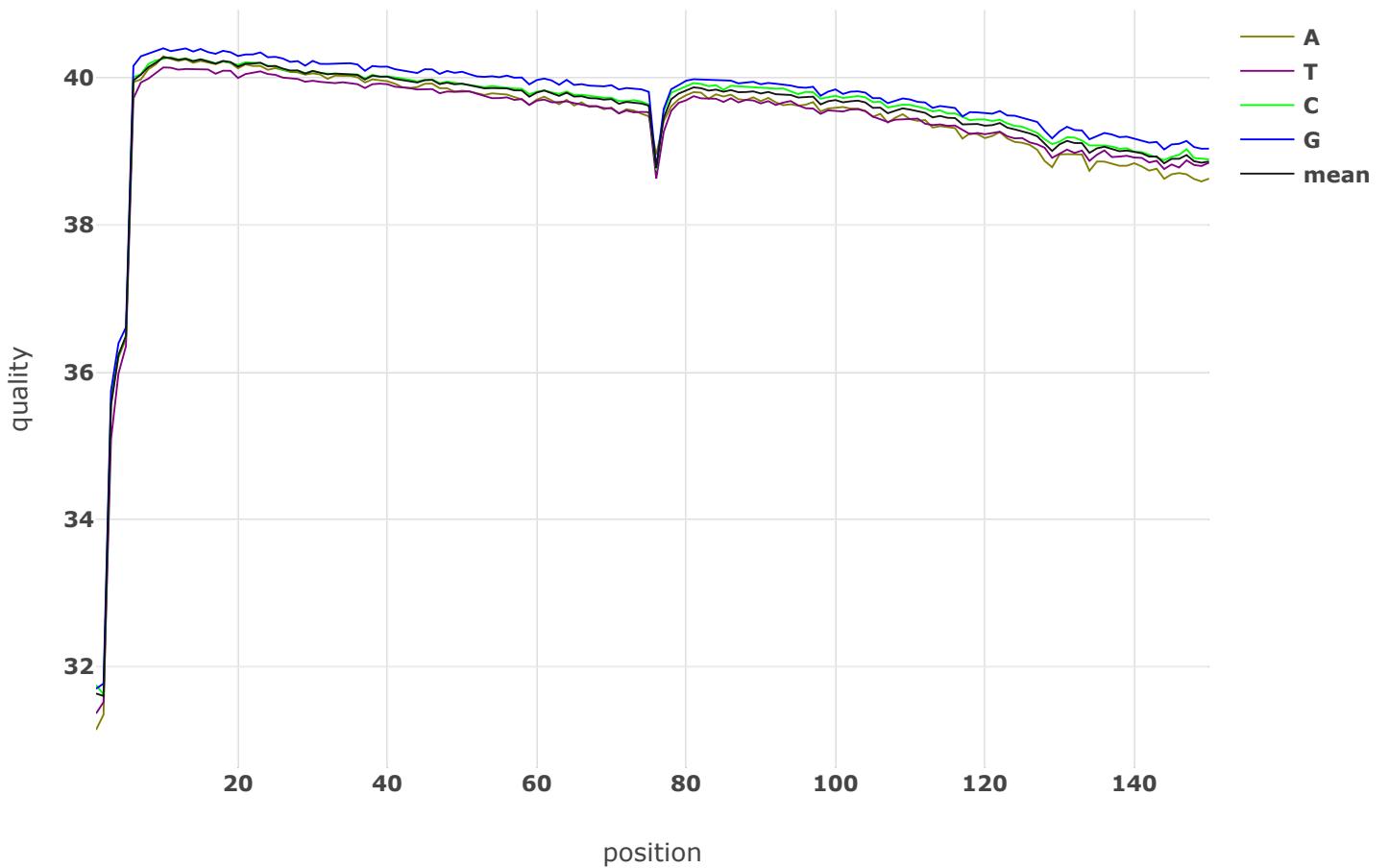




After filtering

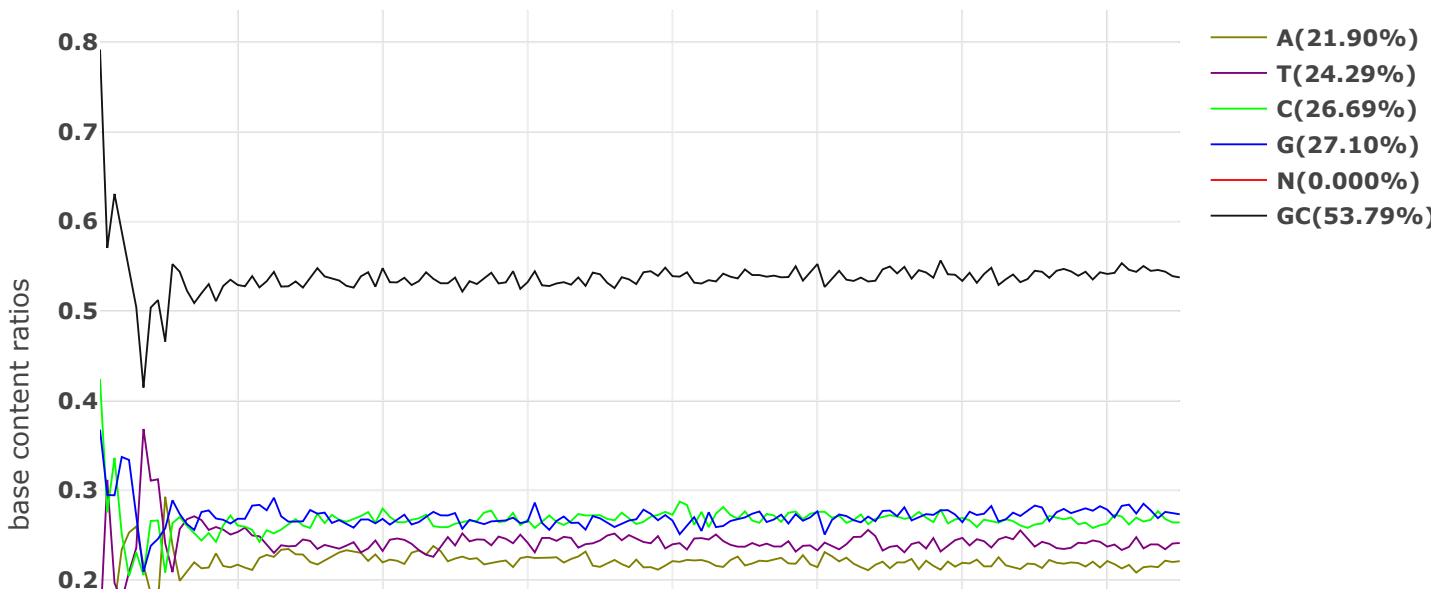
After filtering: read1: quality

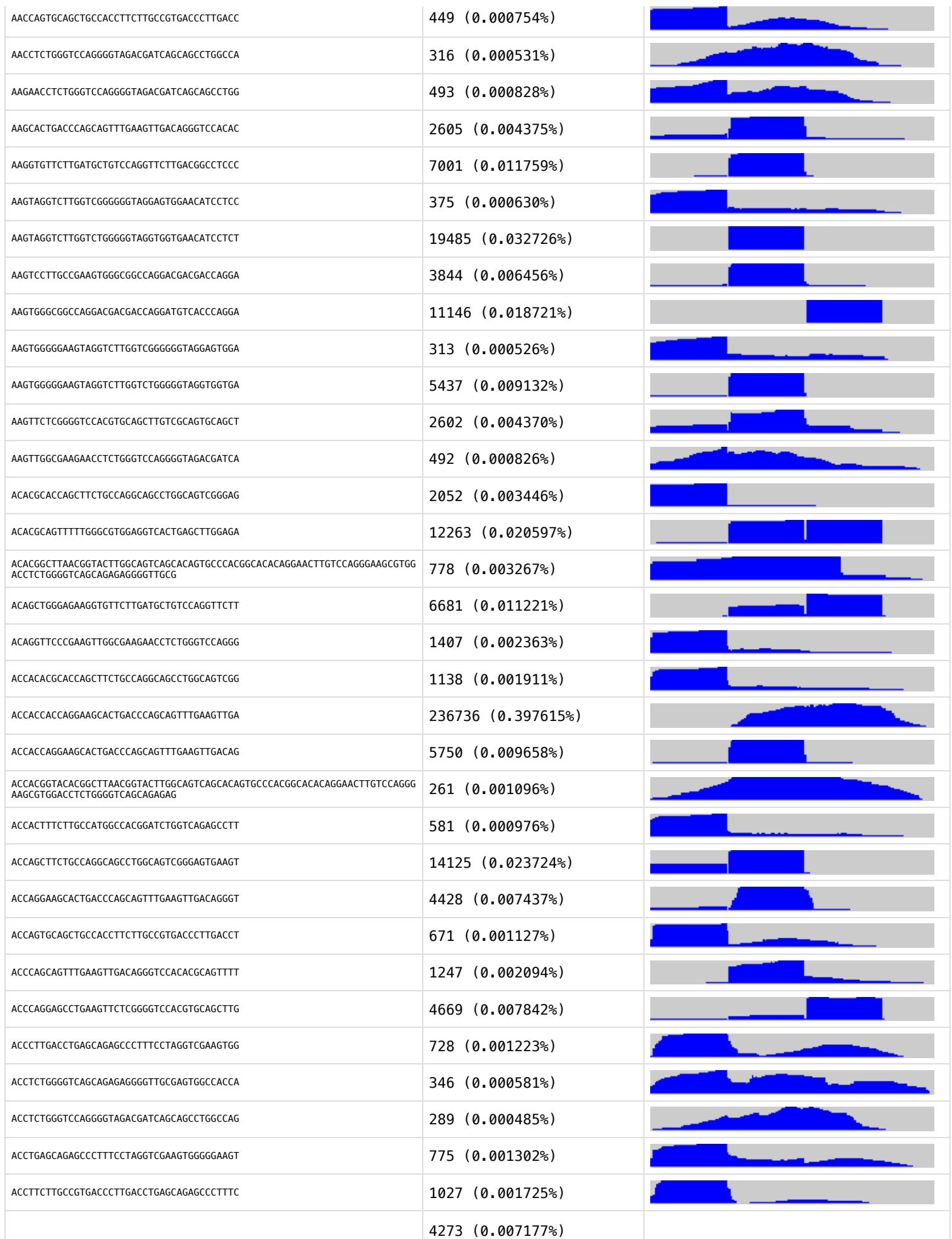
Value of each position will be shown on mouse over.

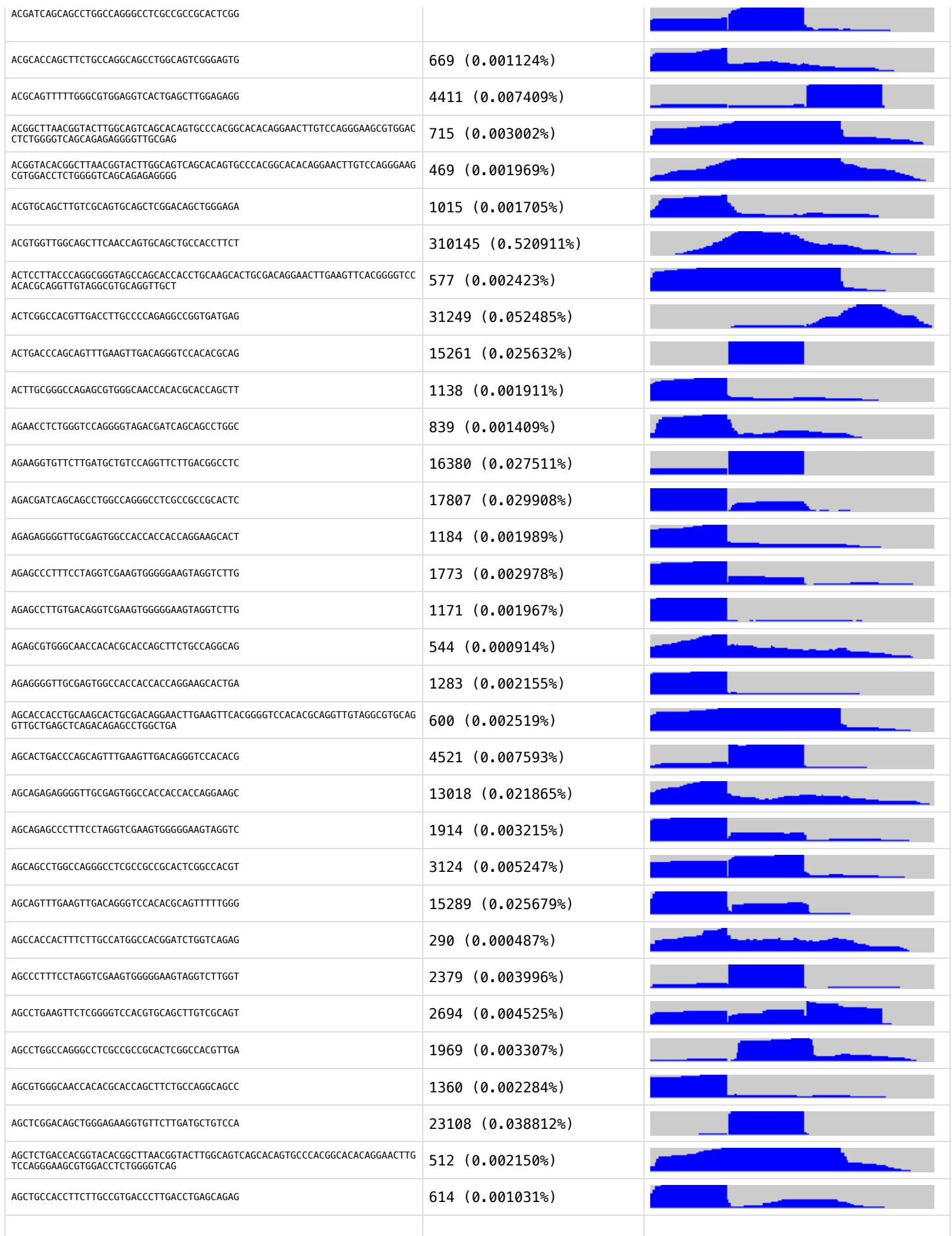


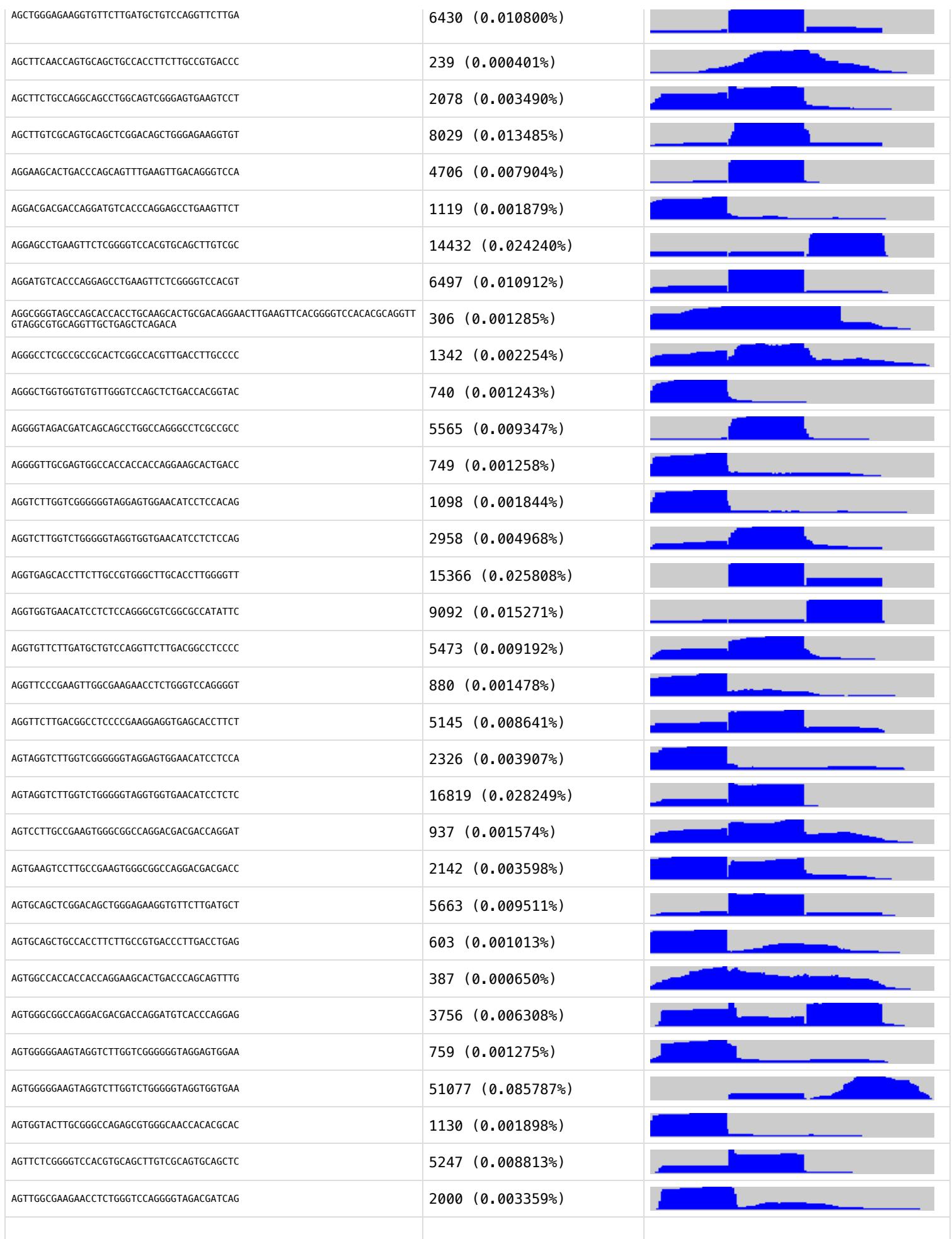
After filtering: read1: base contents

Value of each position will be shown on mouse over.









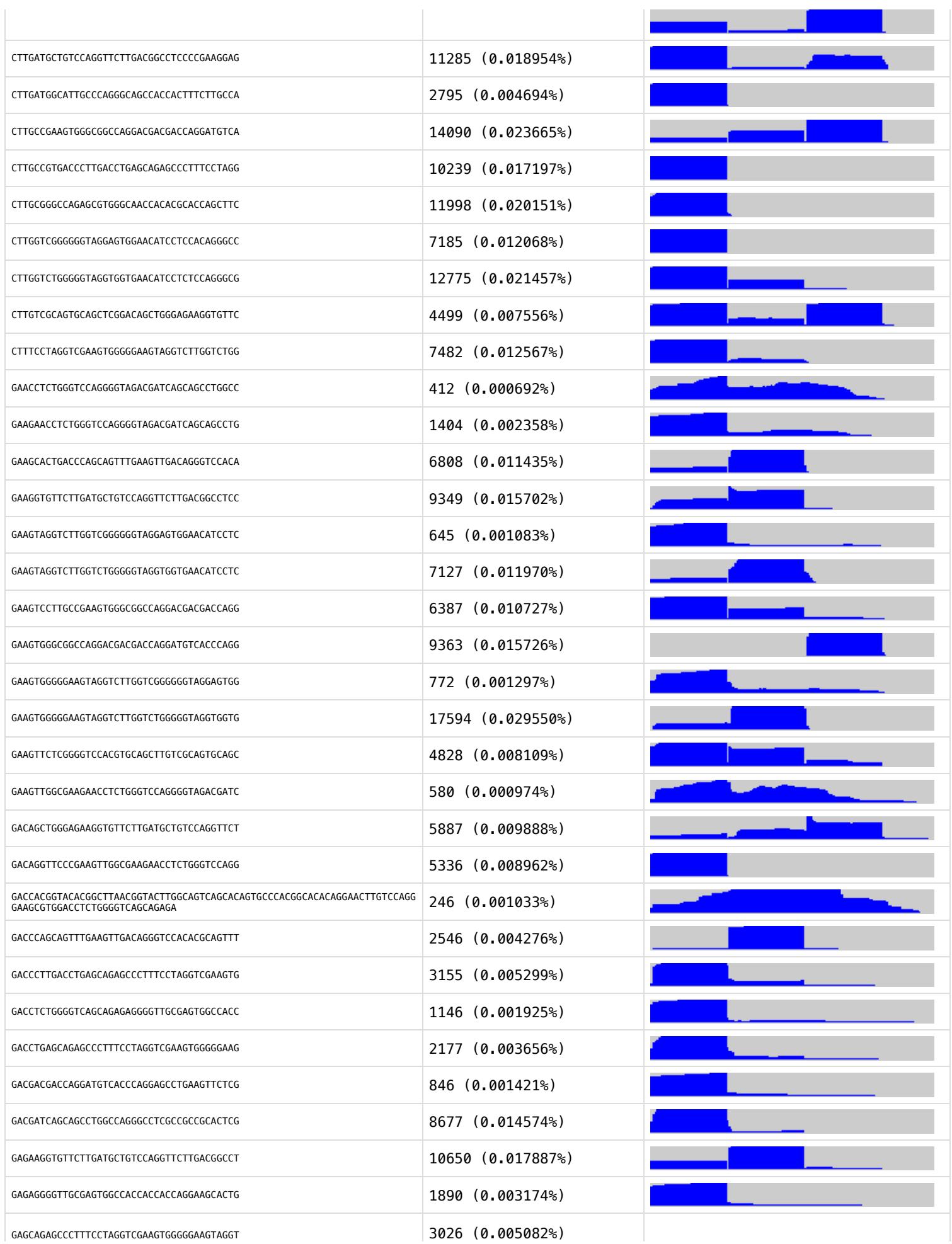


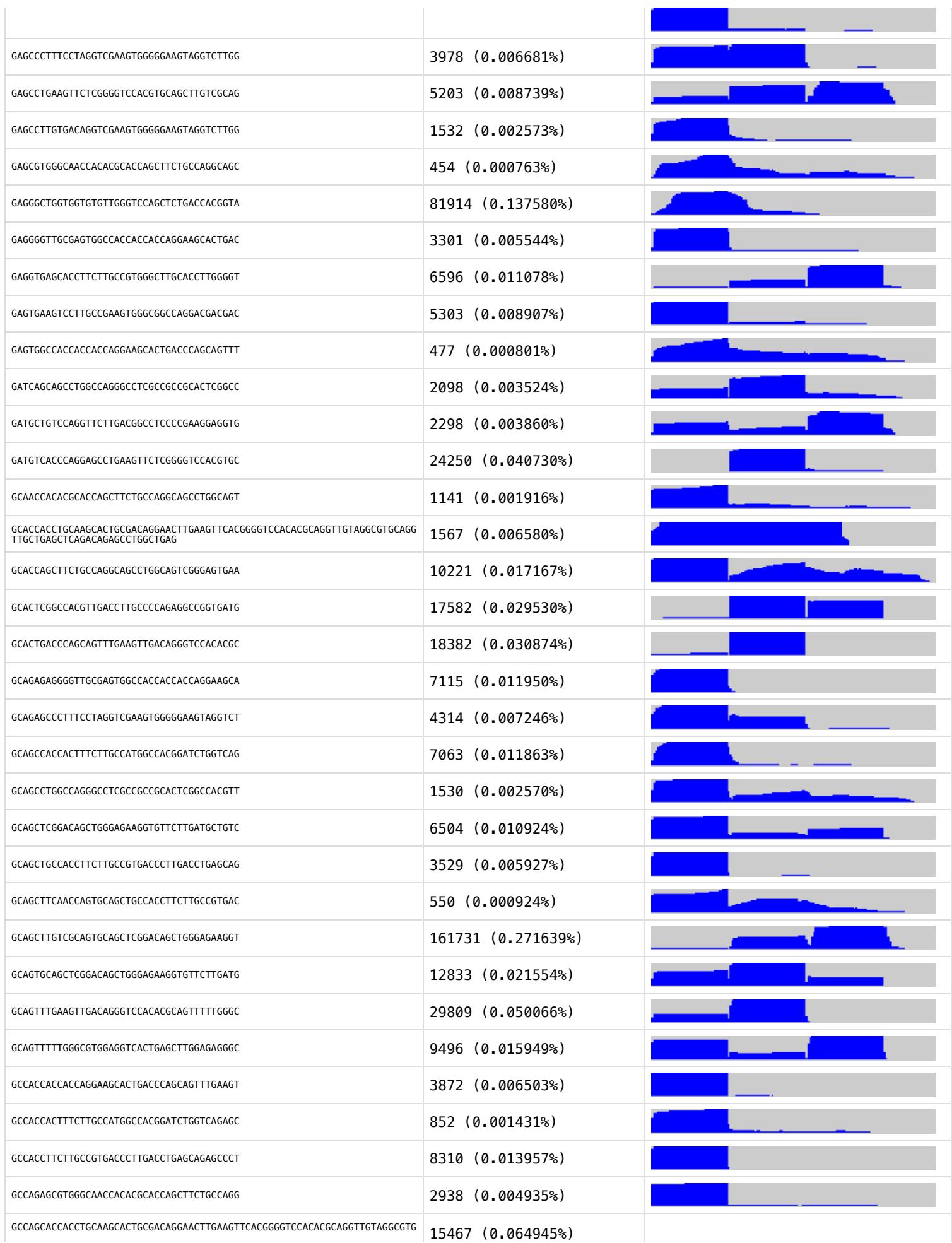






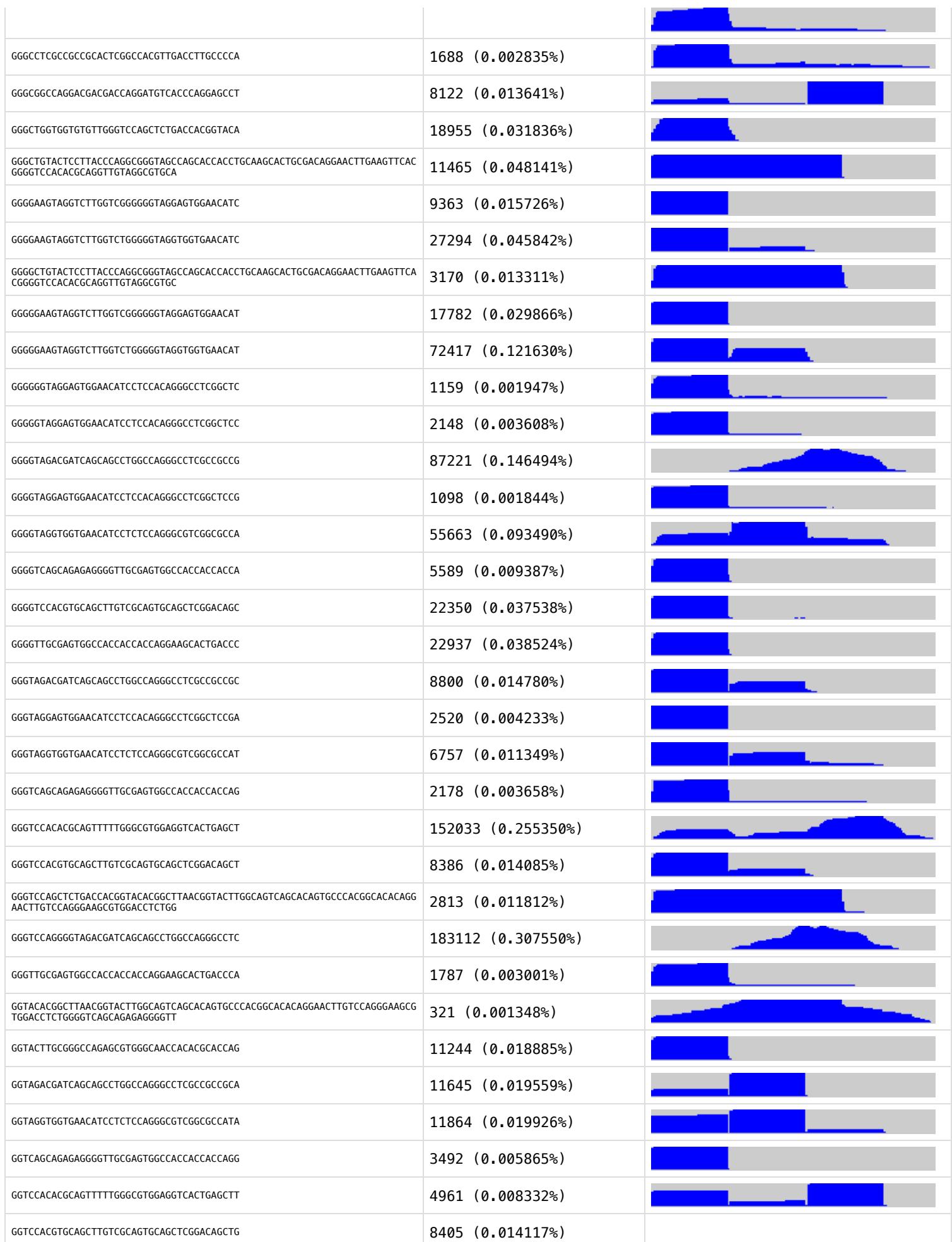


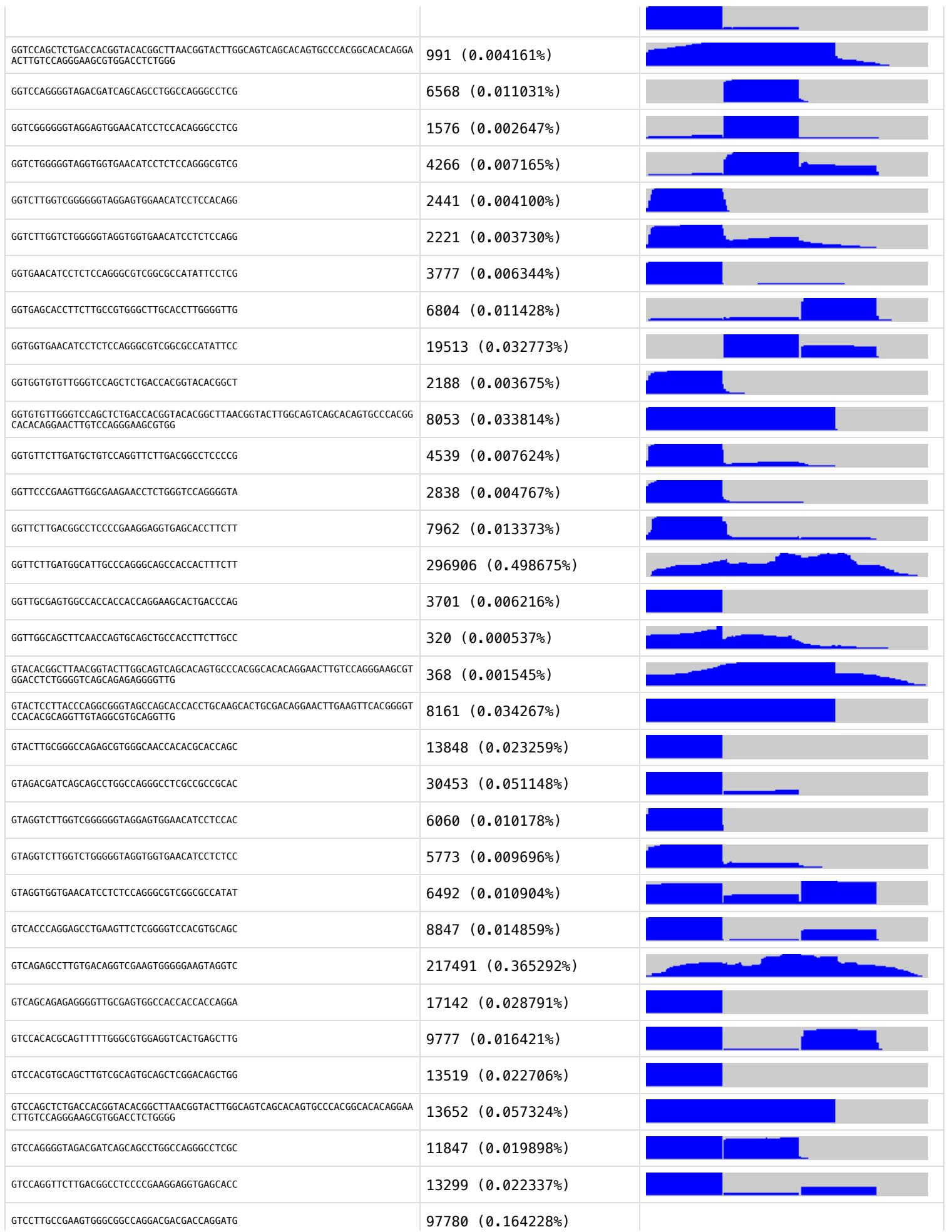




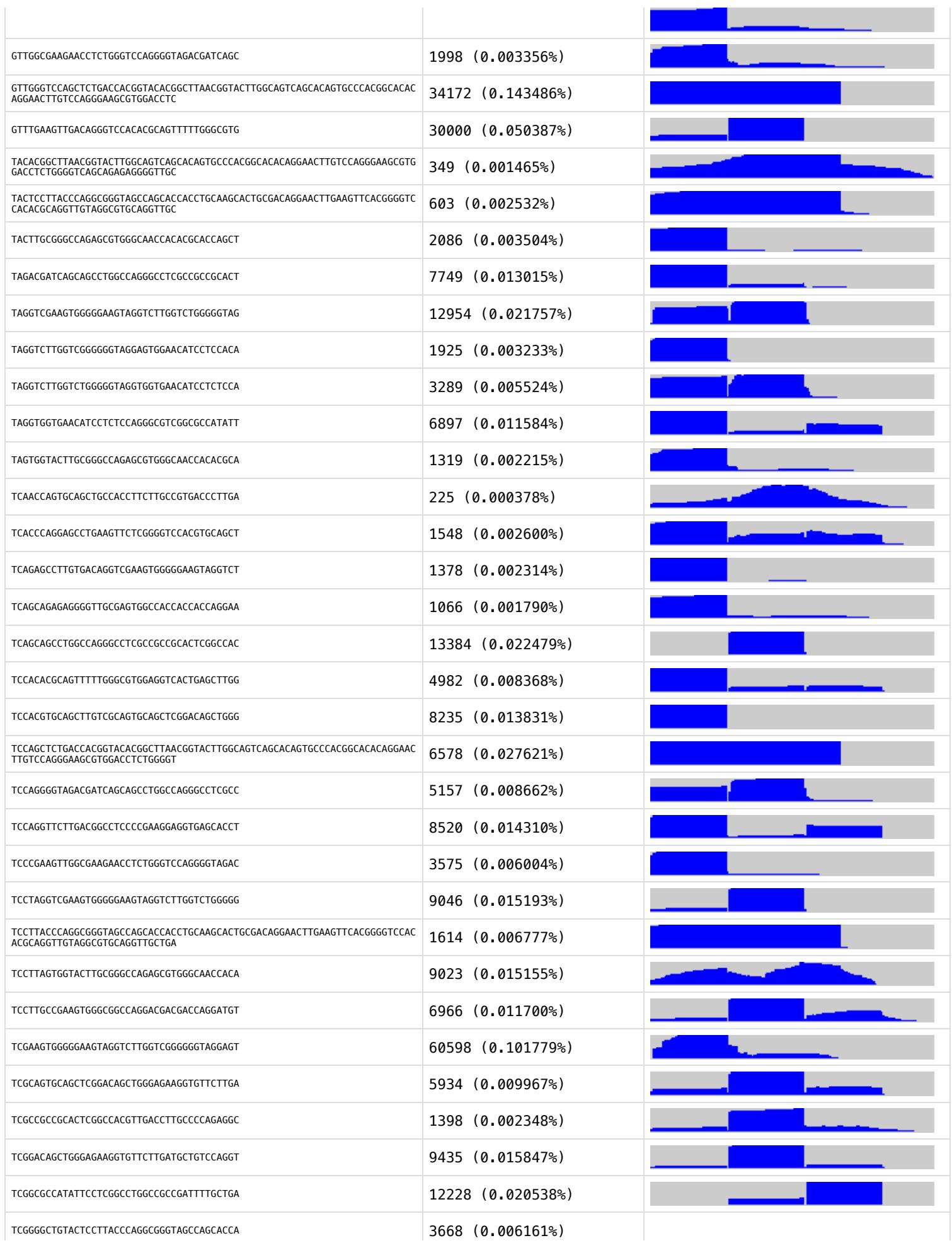




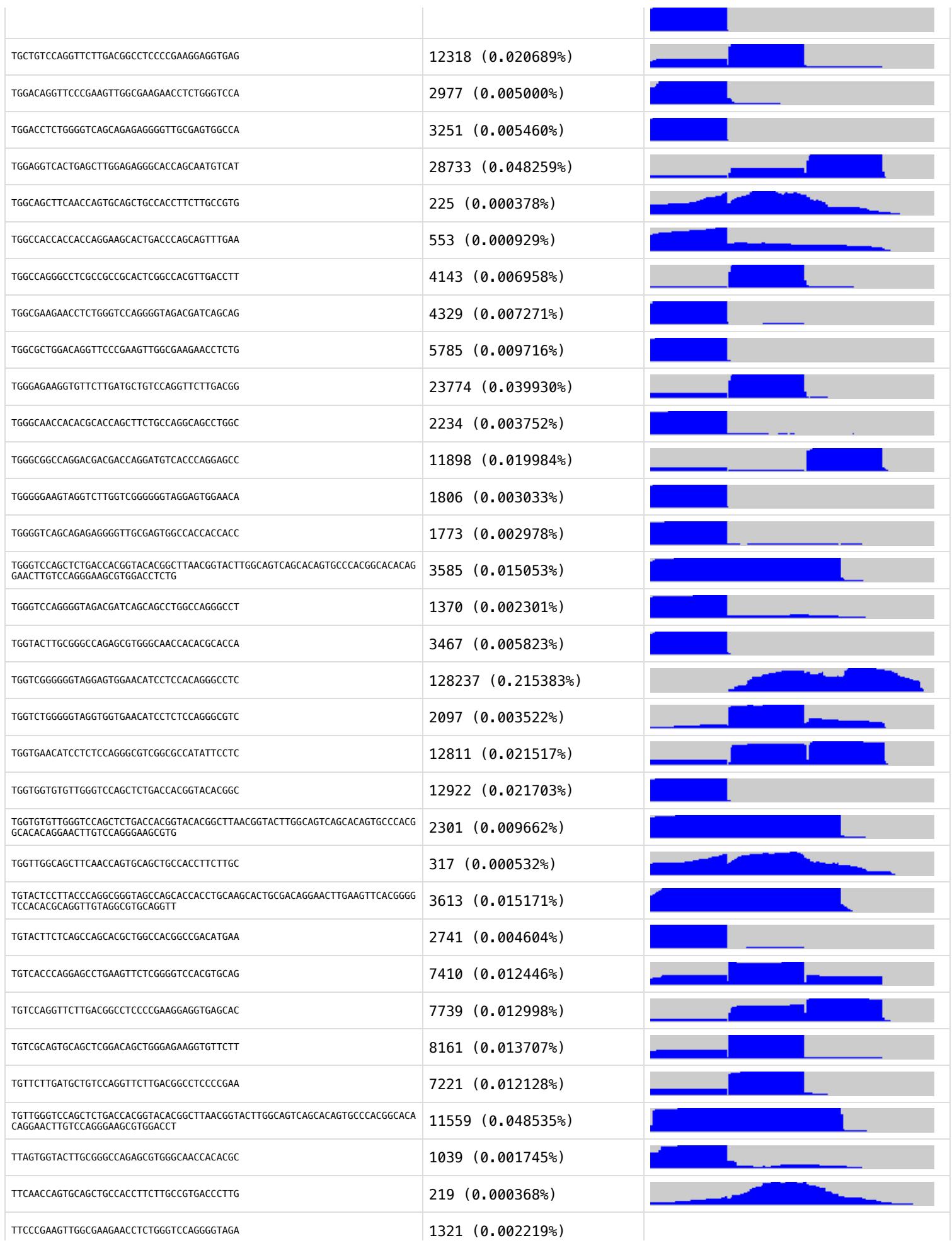


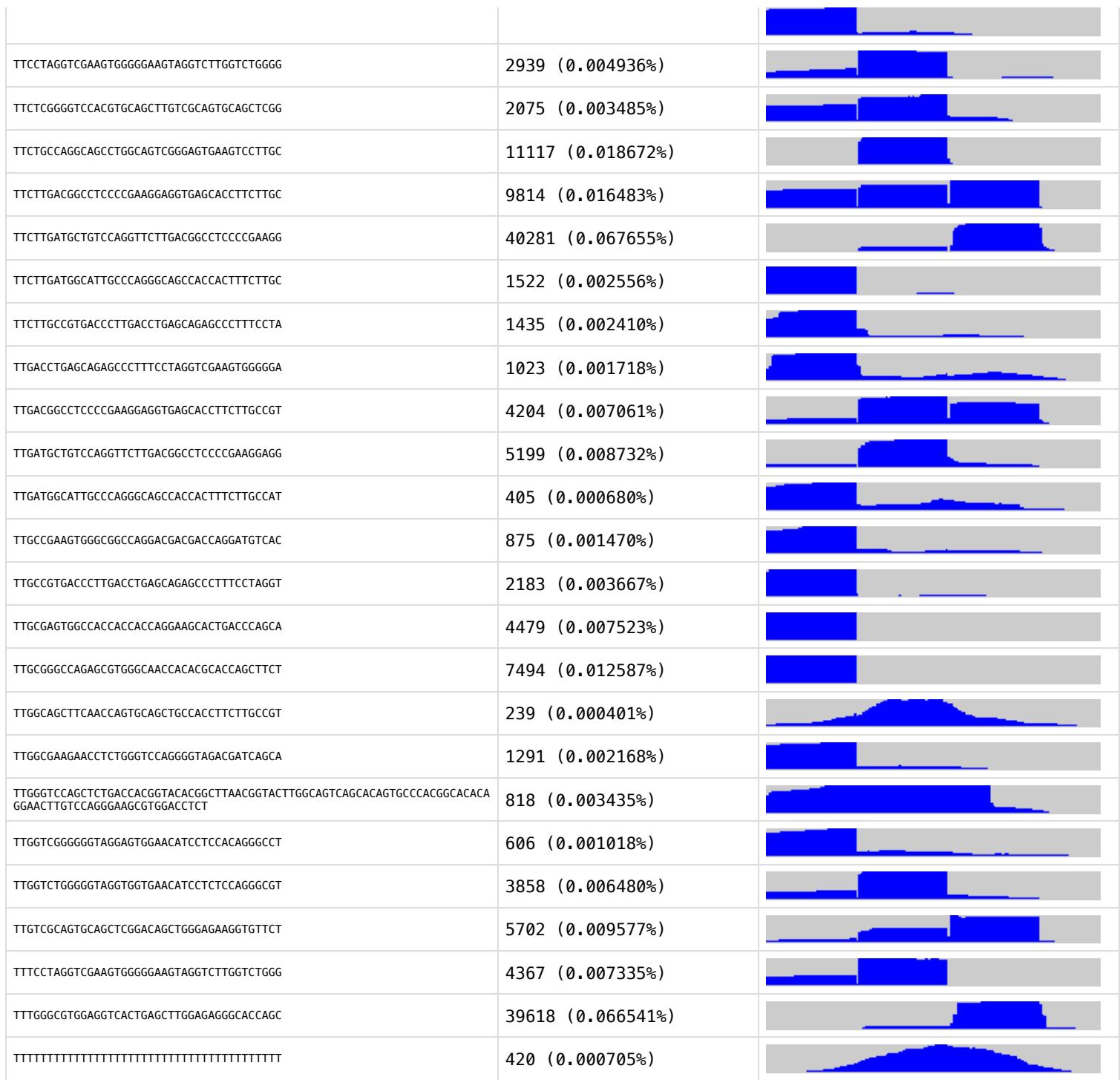






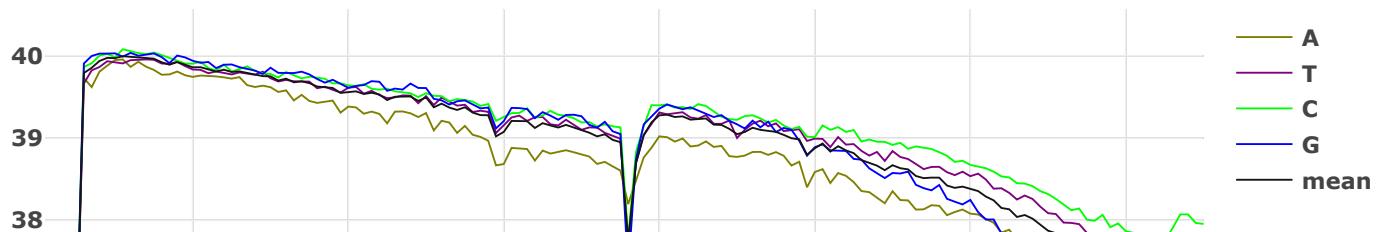


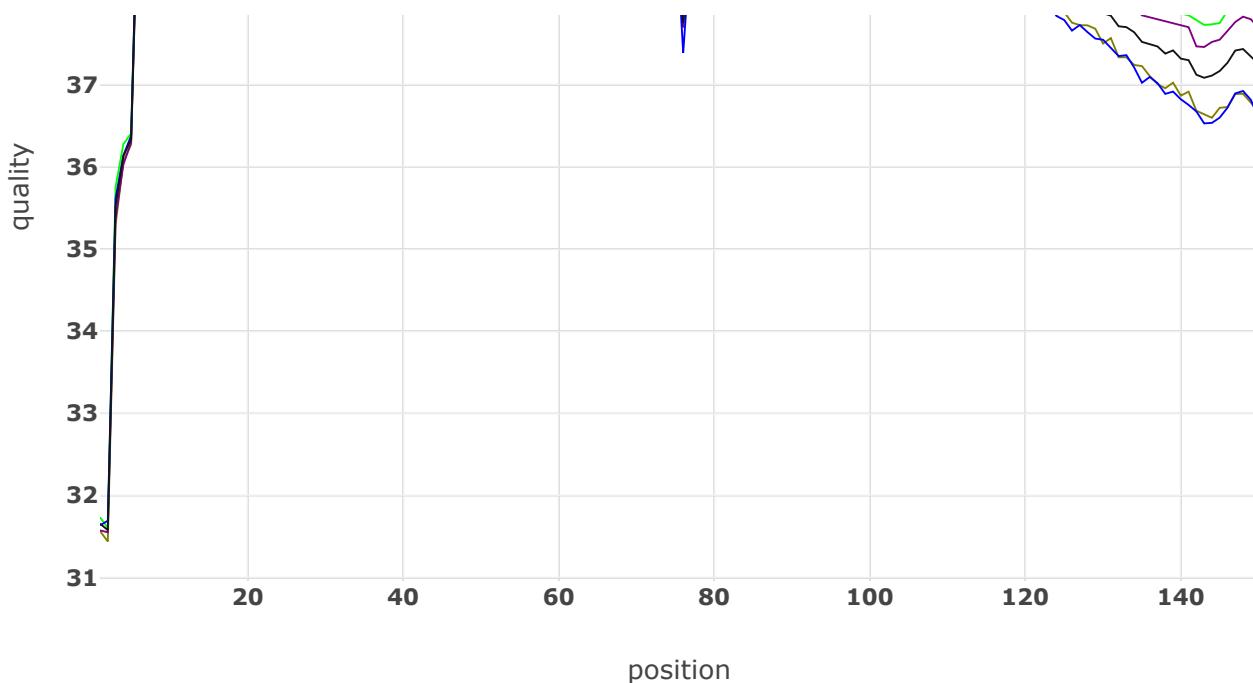




After filtering: read2: quality

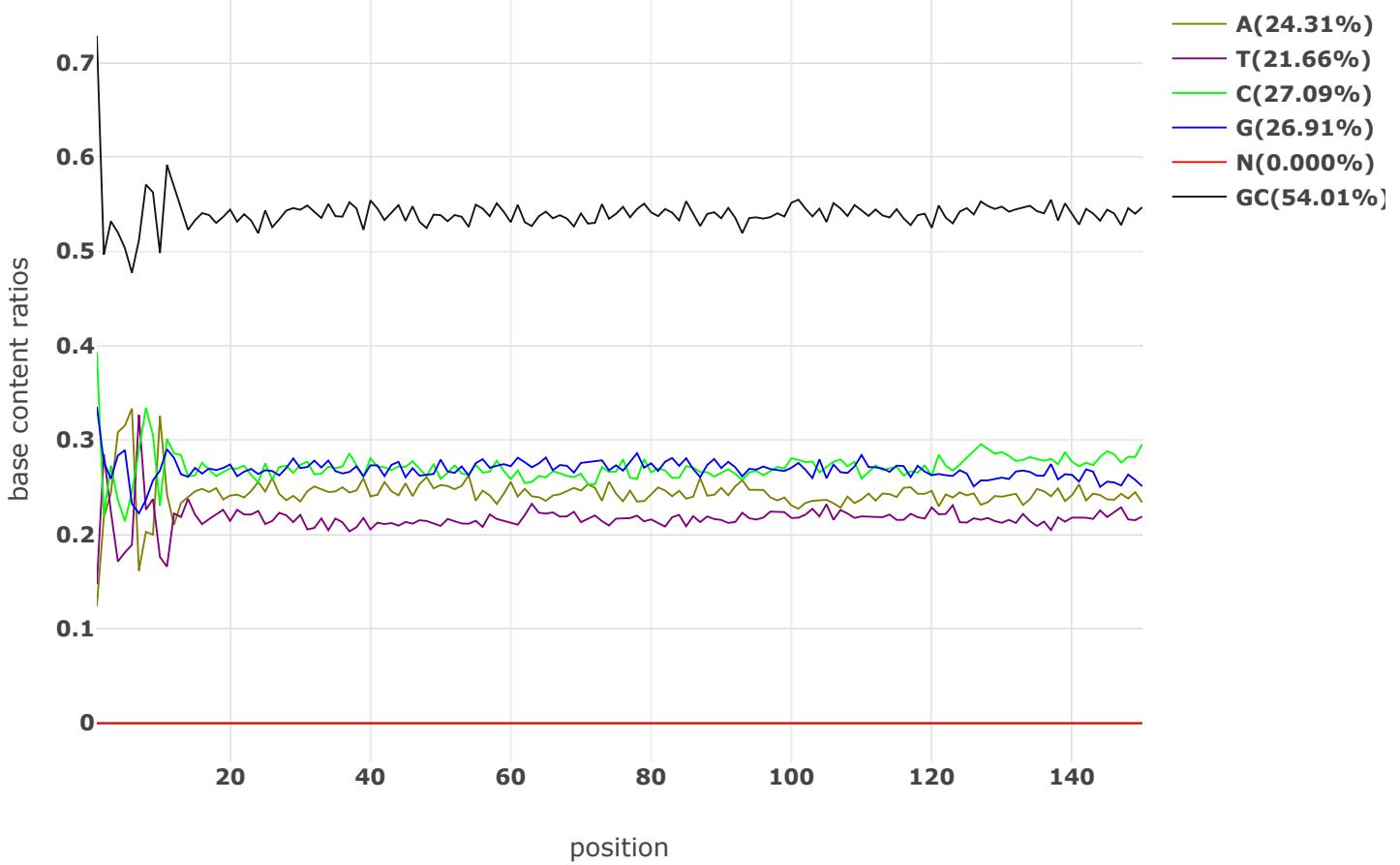
Value of each position will be shown on mouse over.

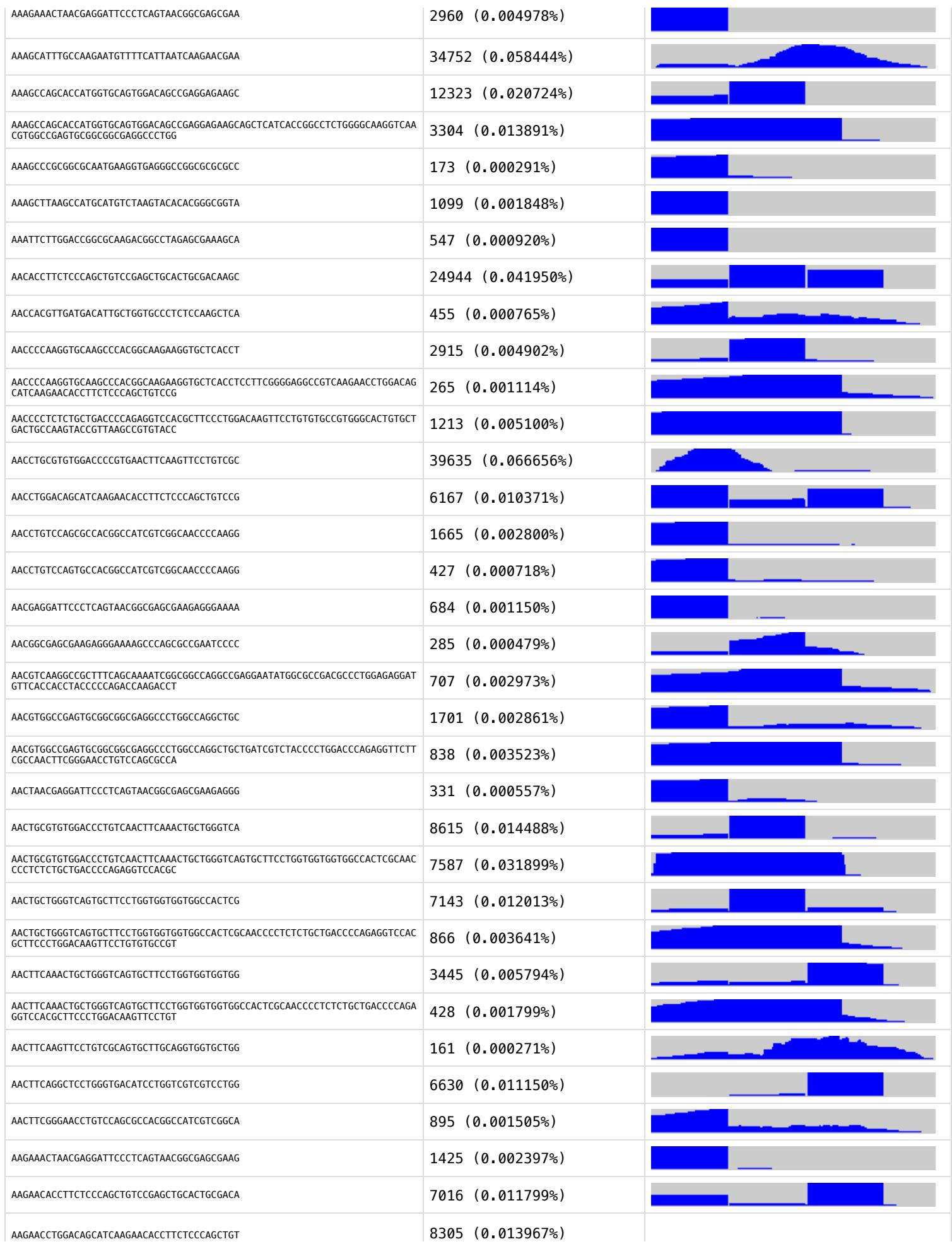


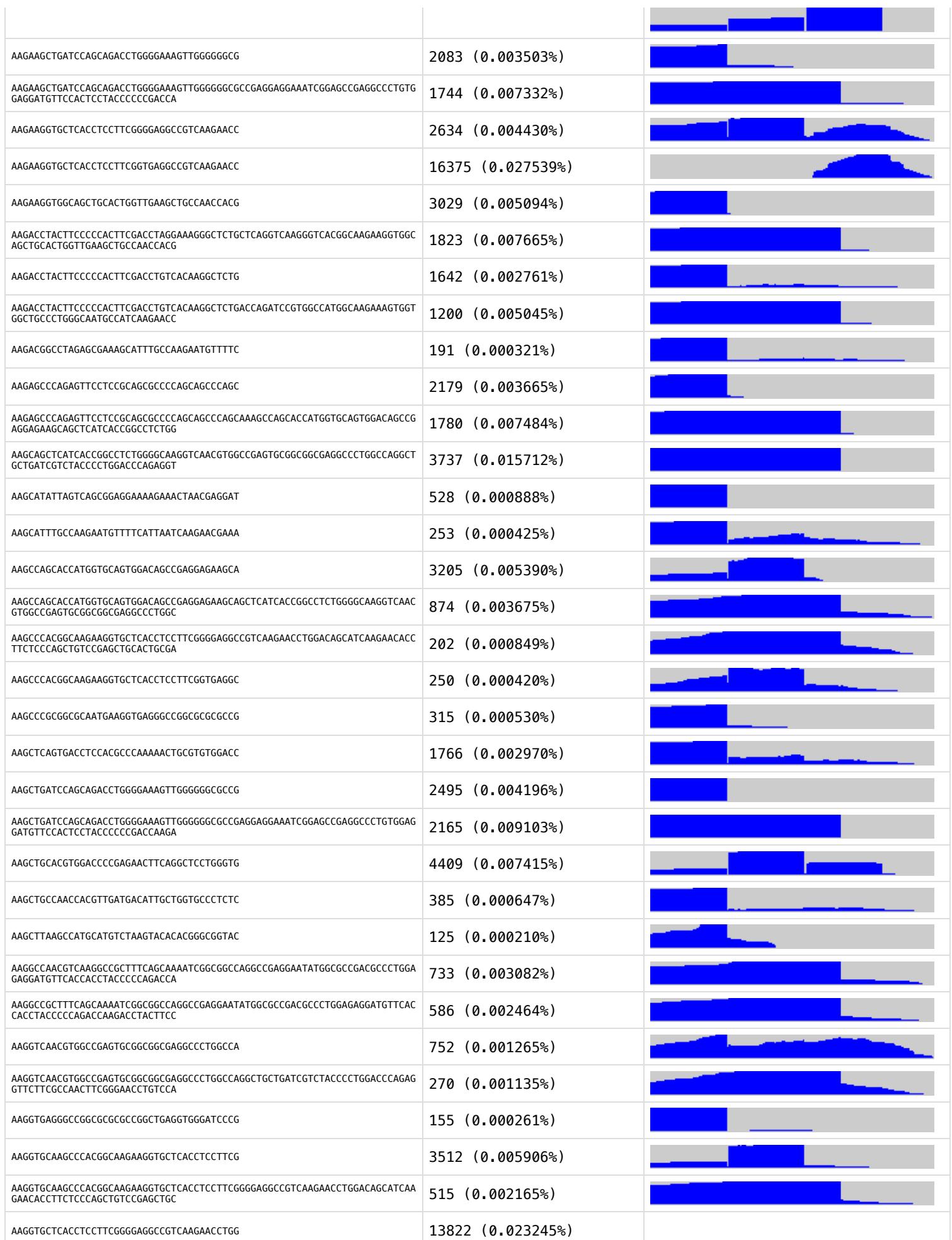


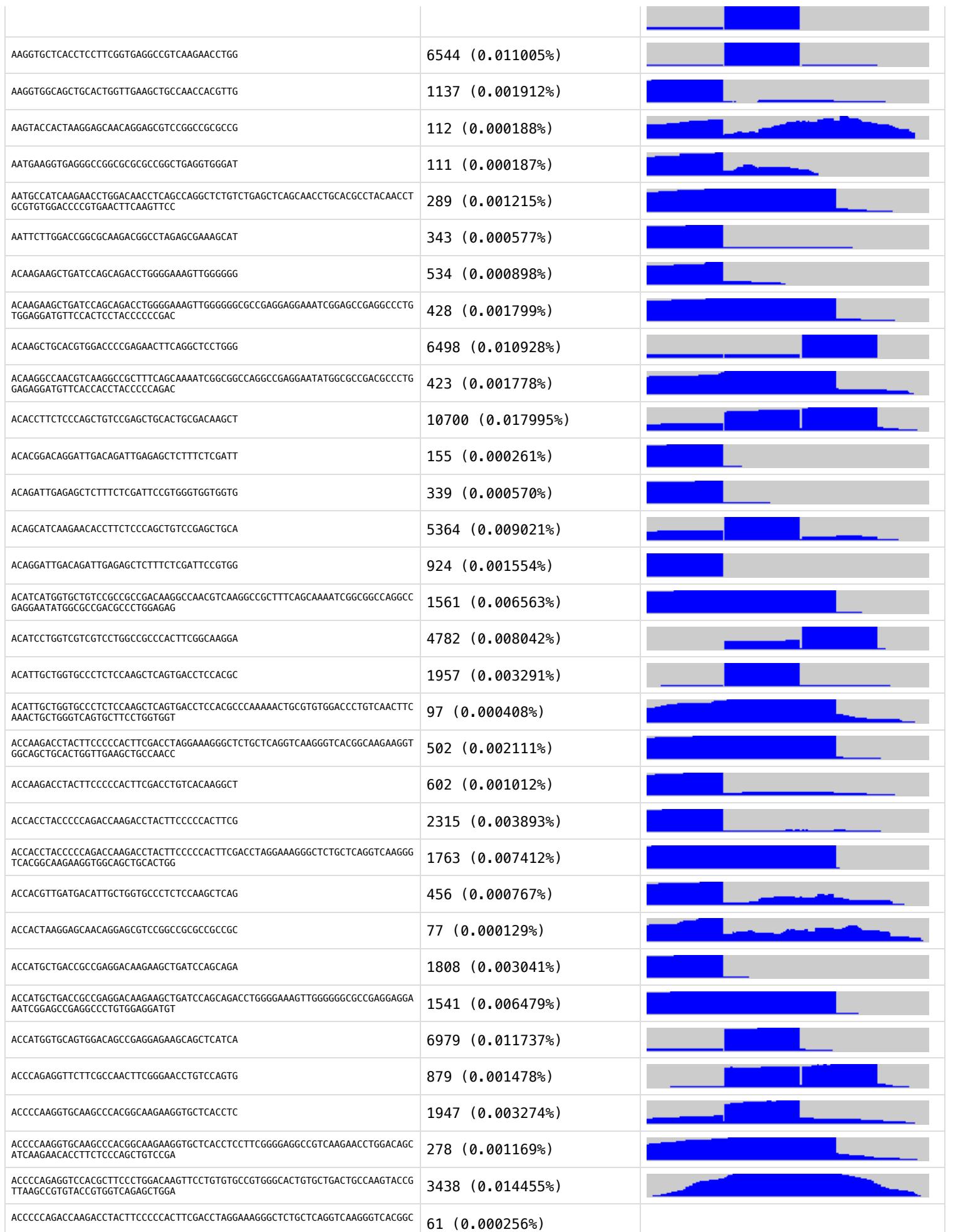
After filtering: read2: base contents

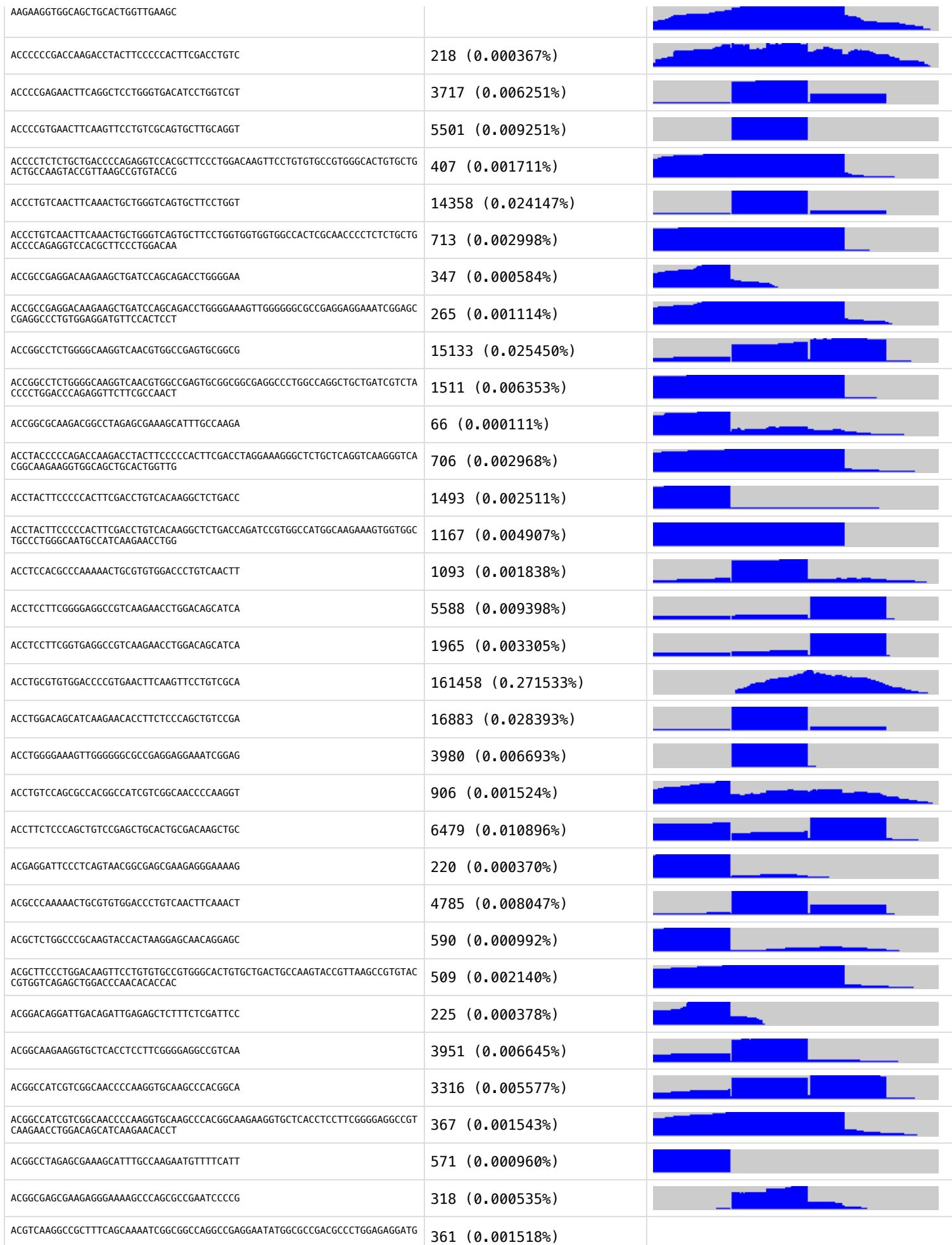
Value of each position will be shown on mouse over.

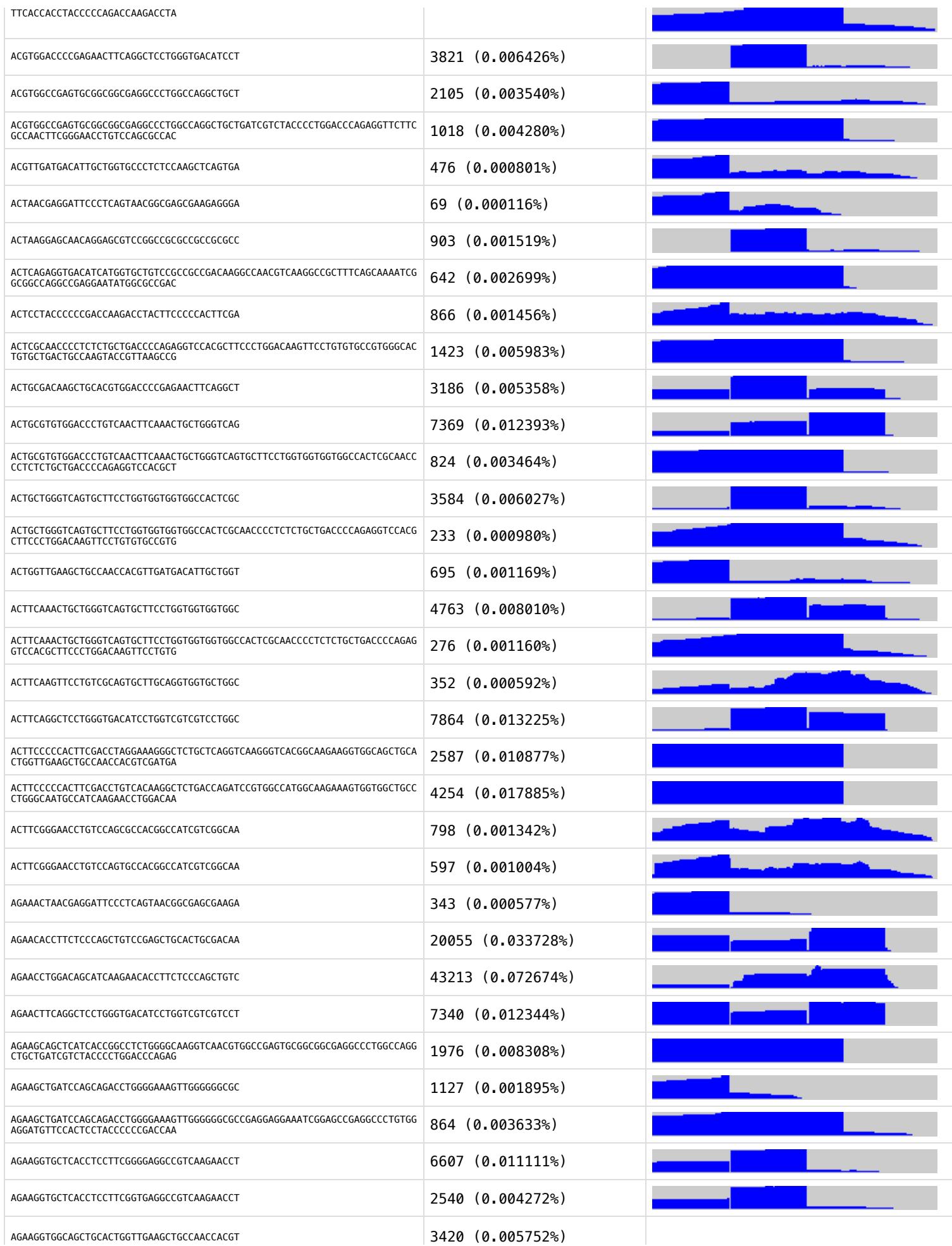


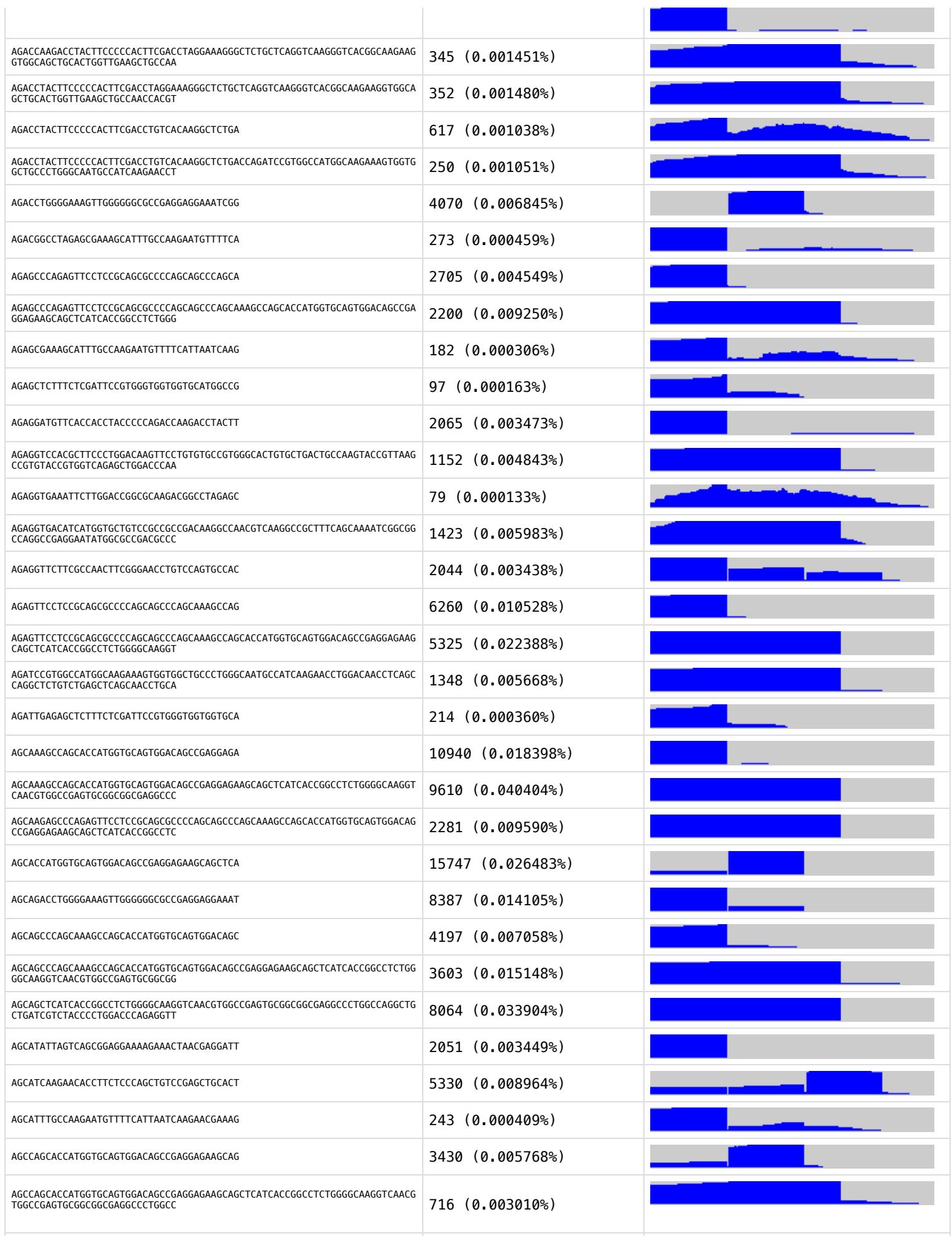


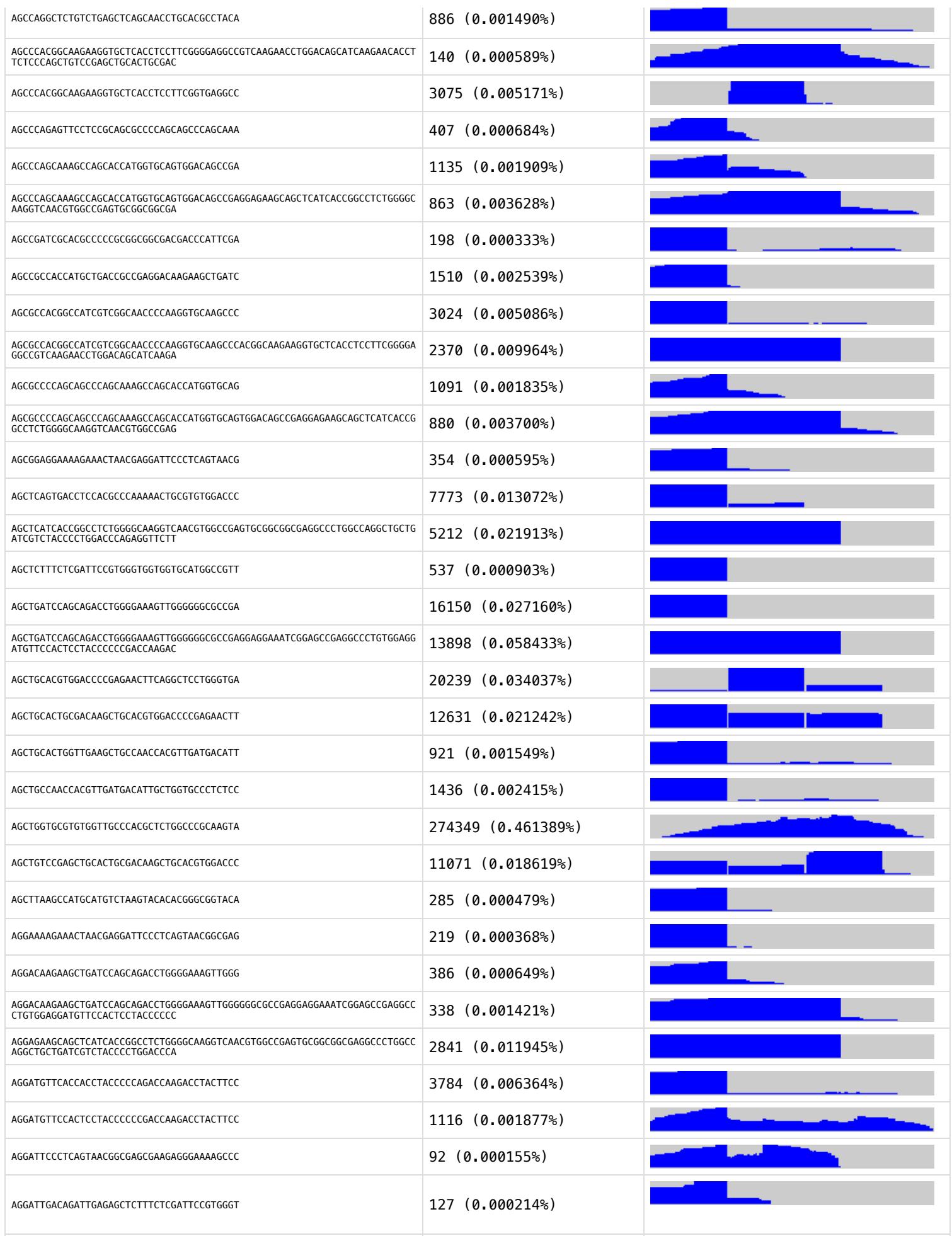


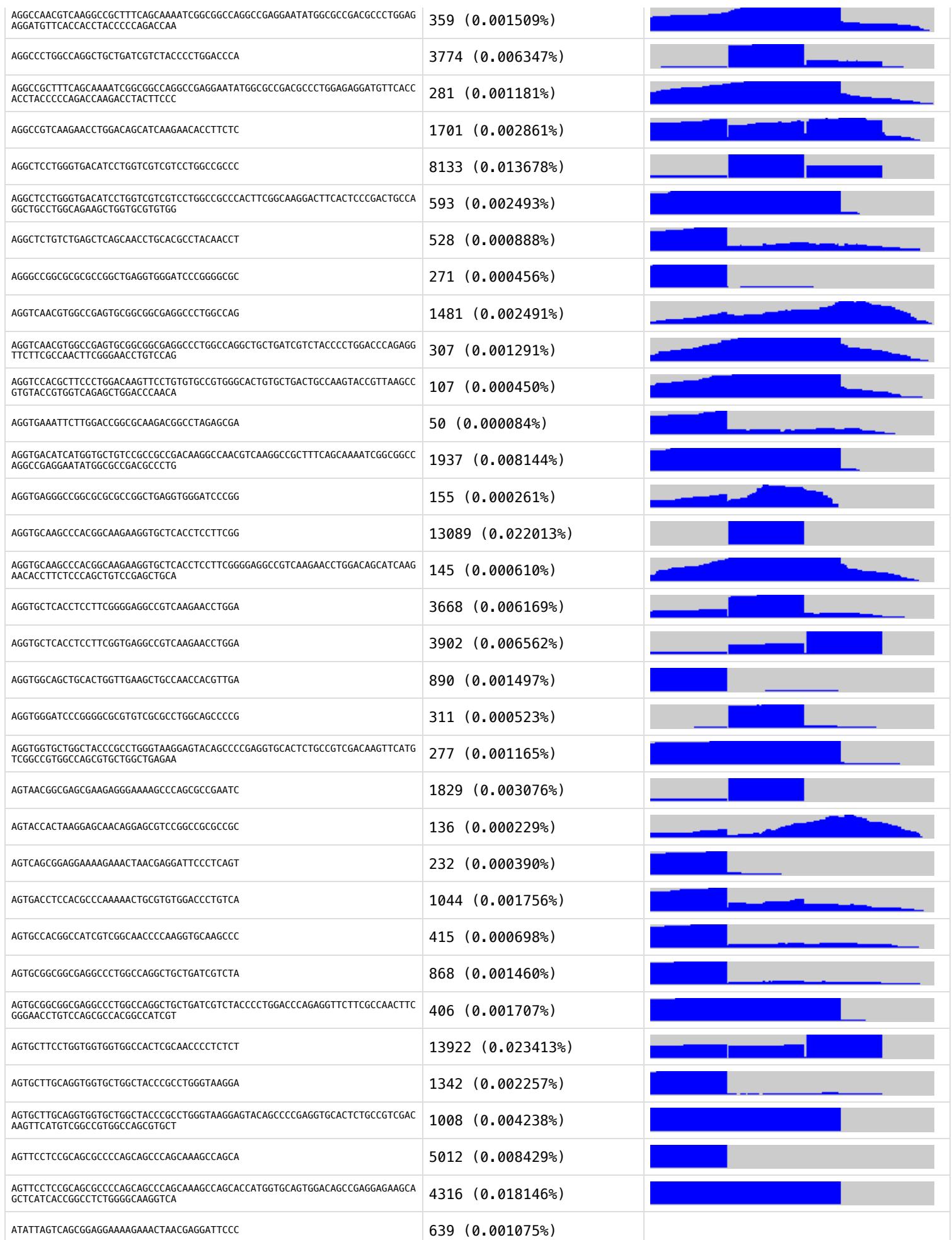


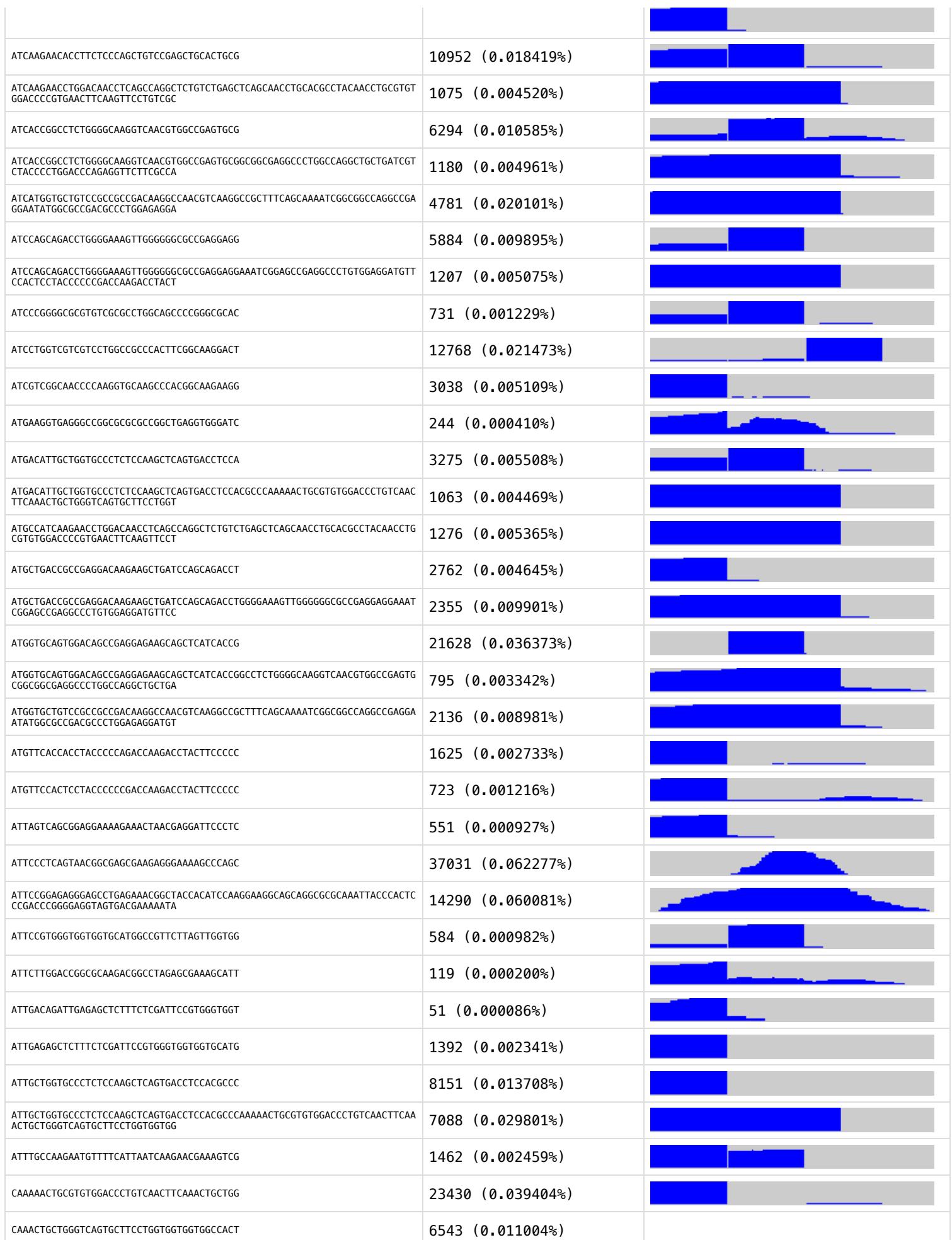


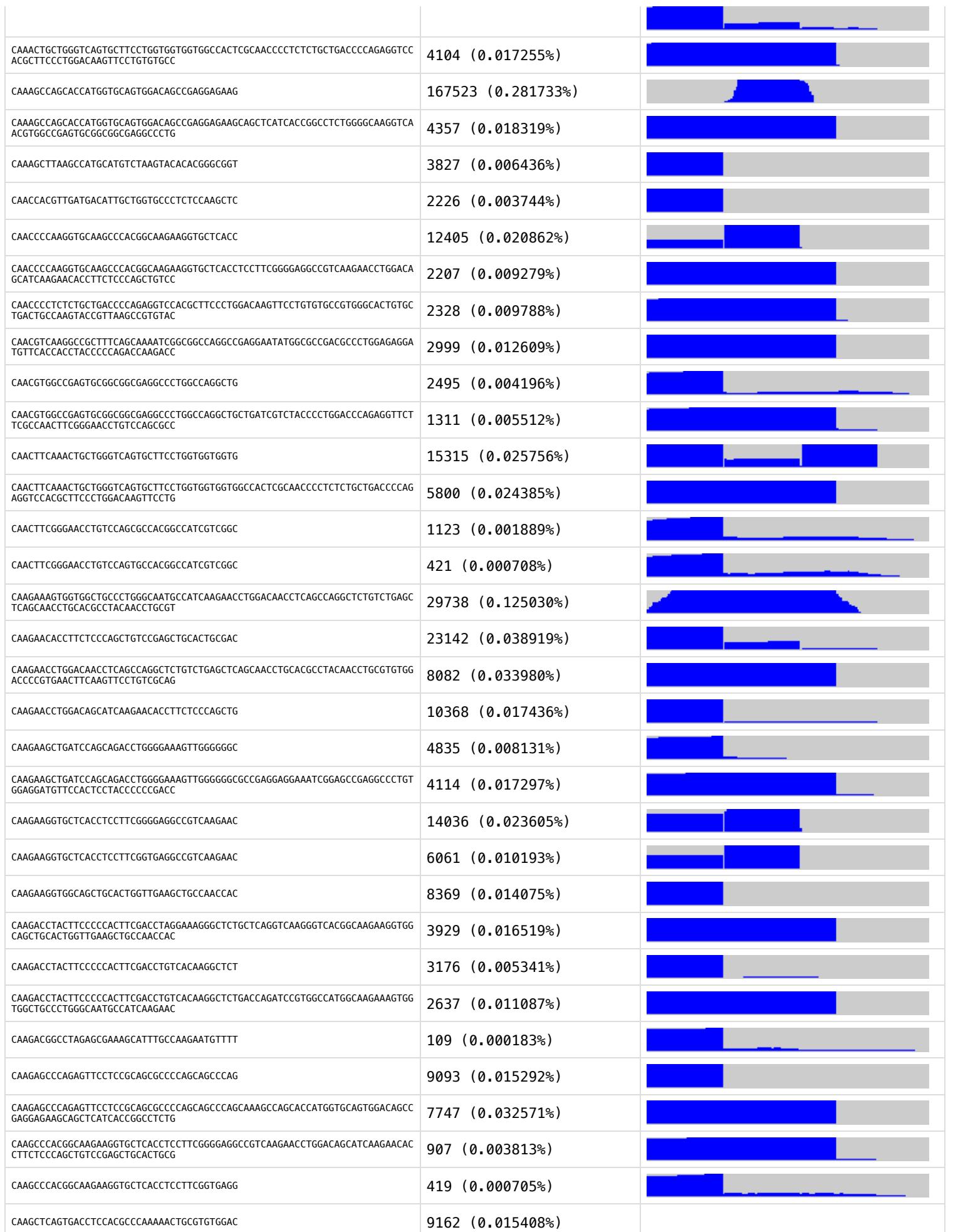


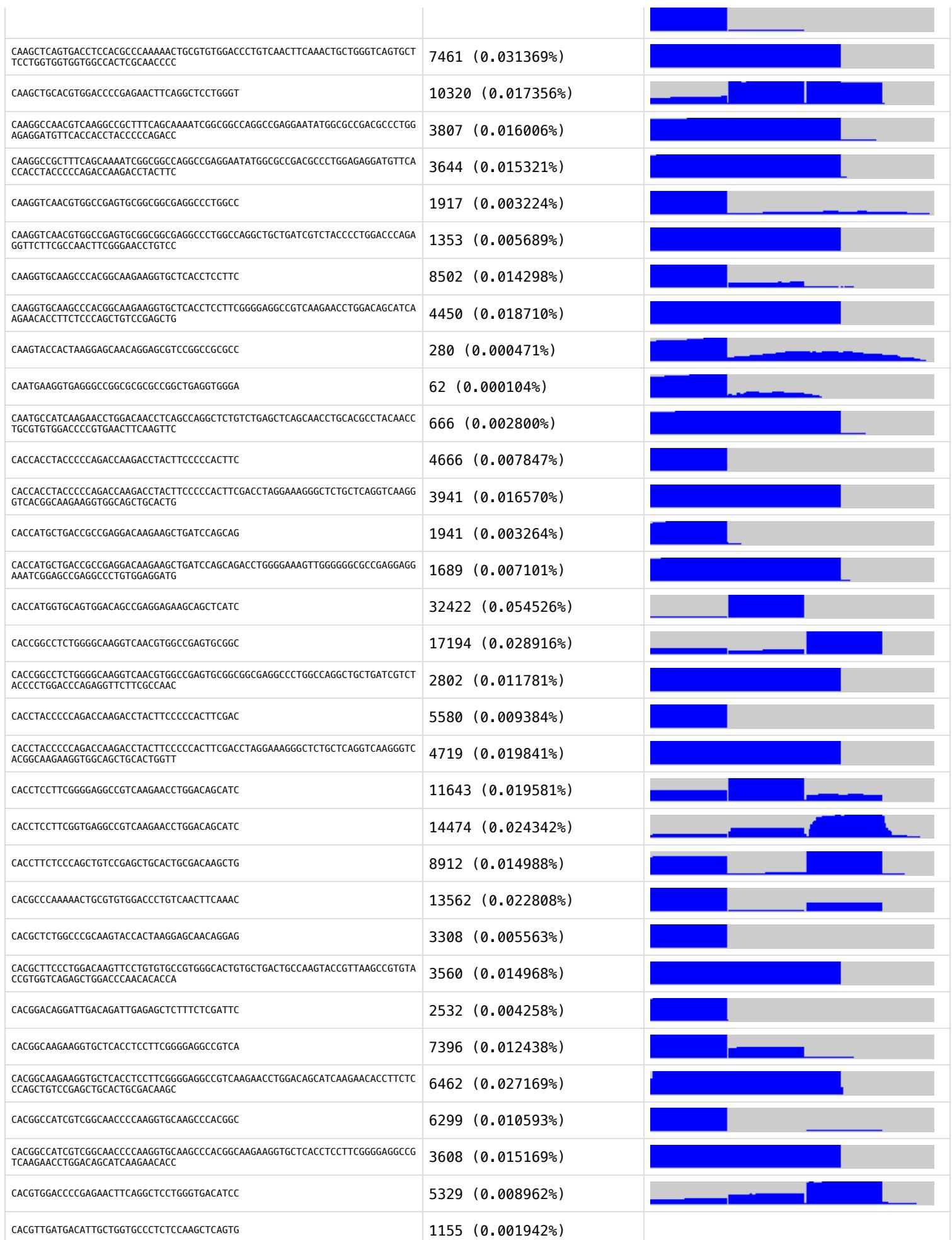


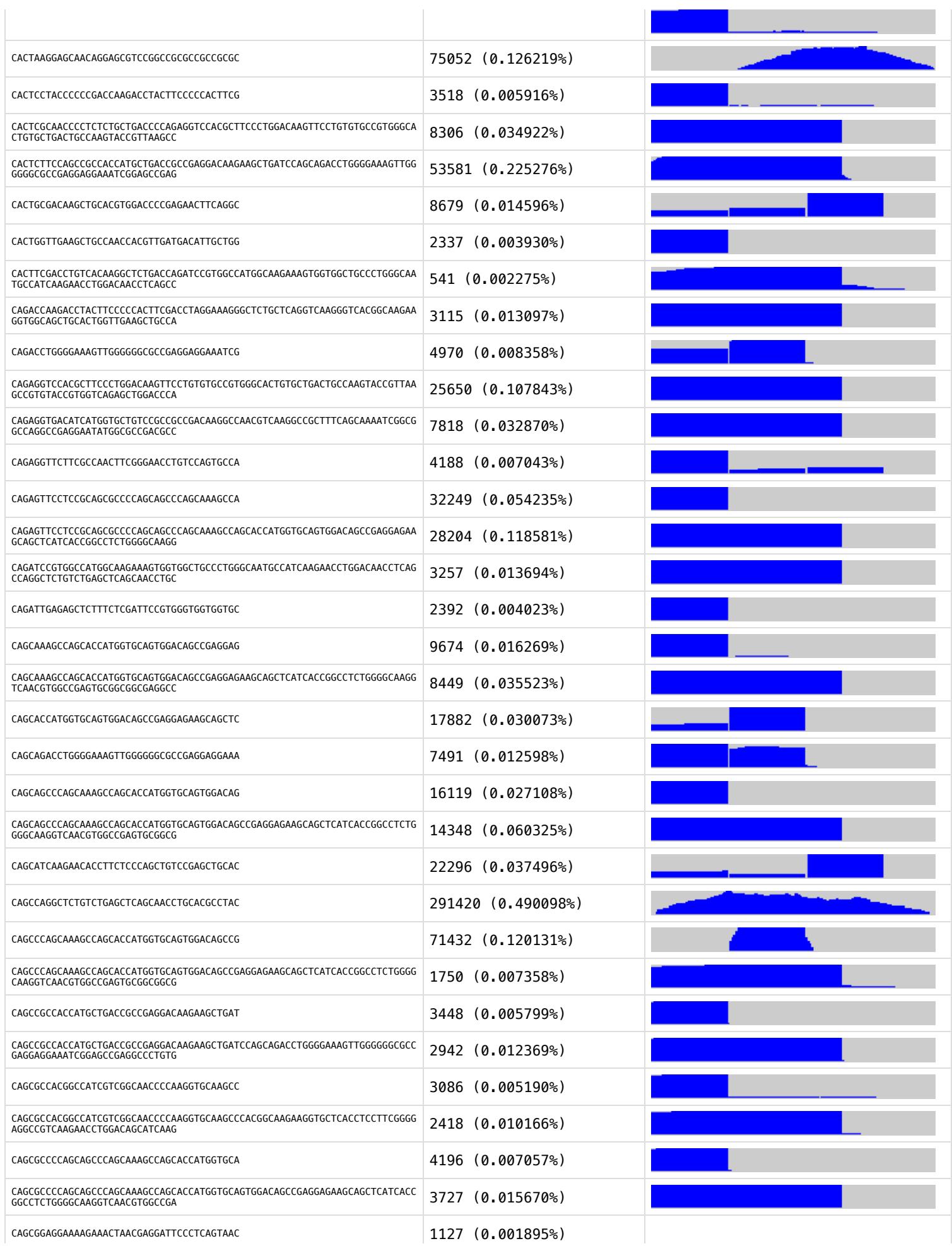


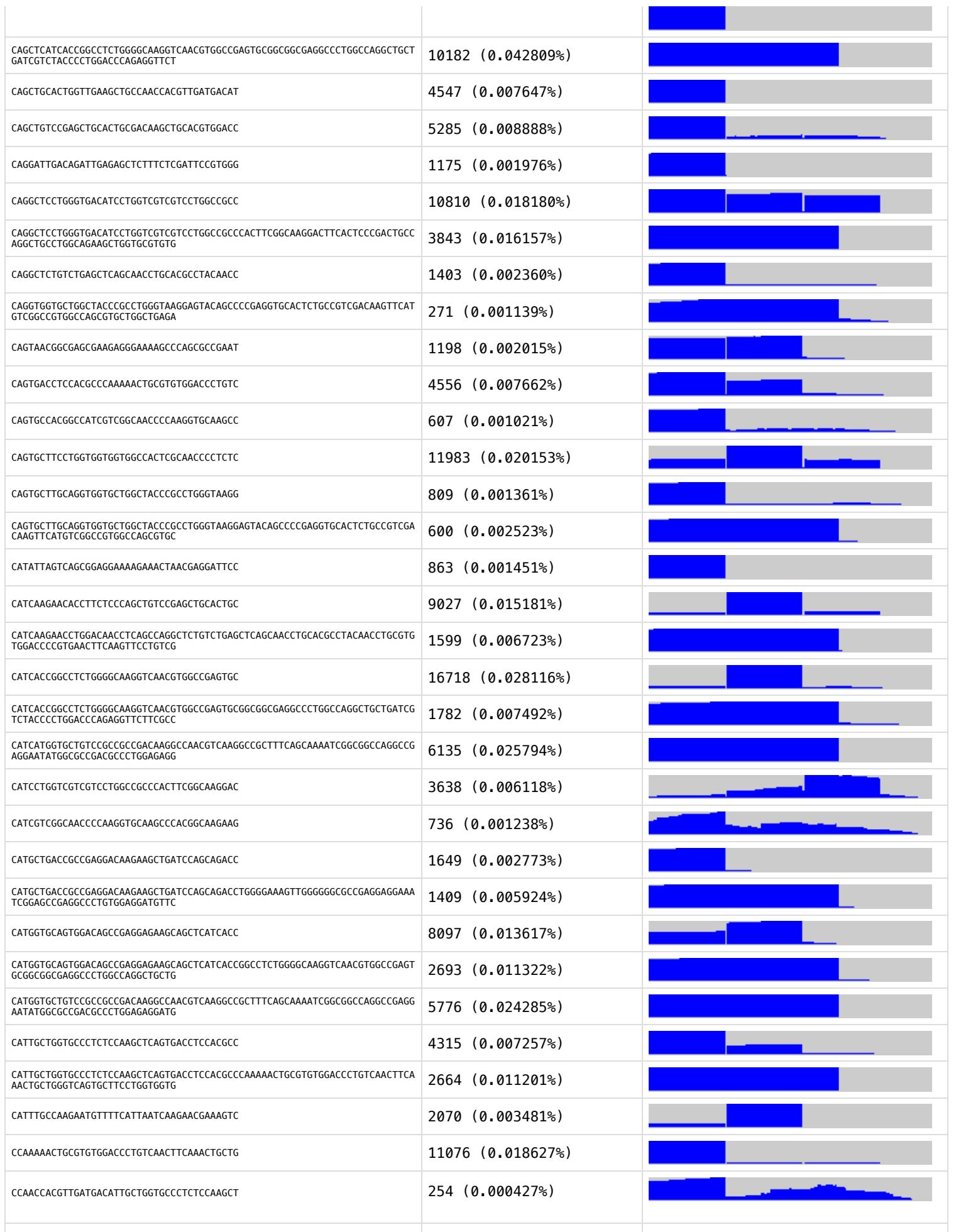


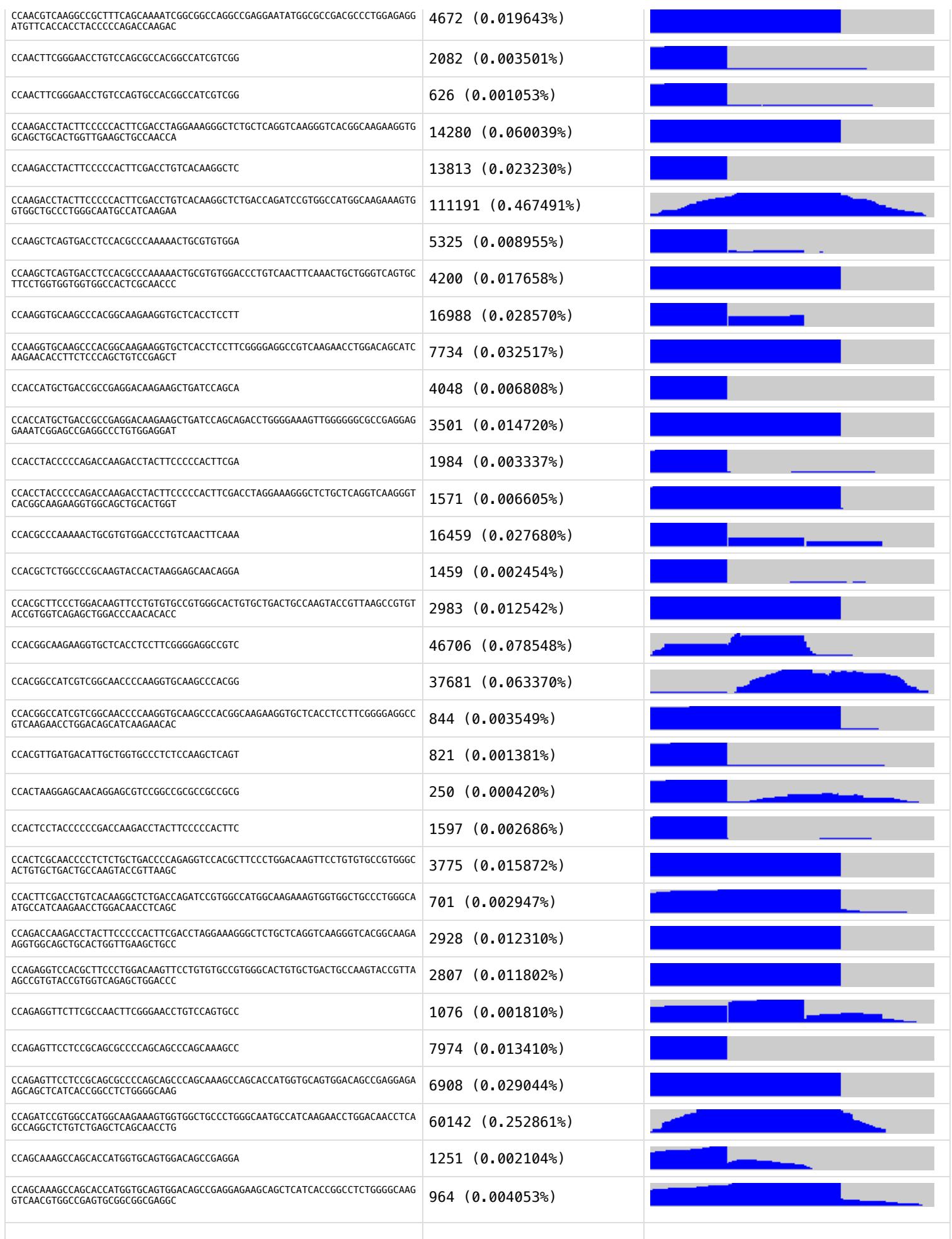


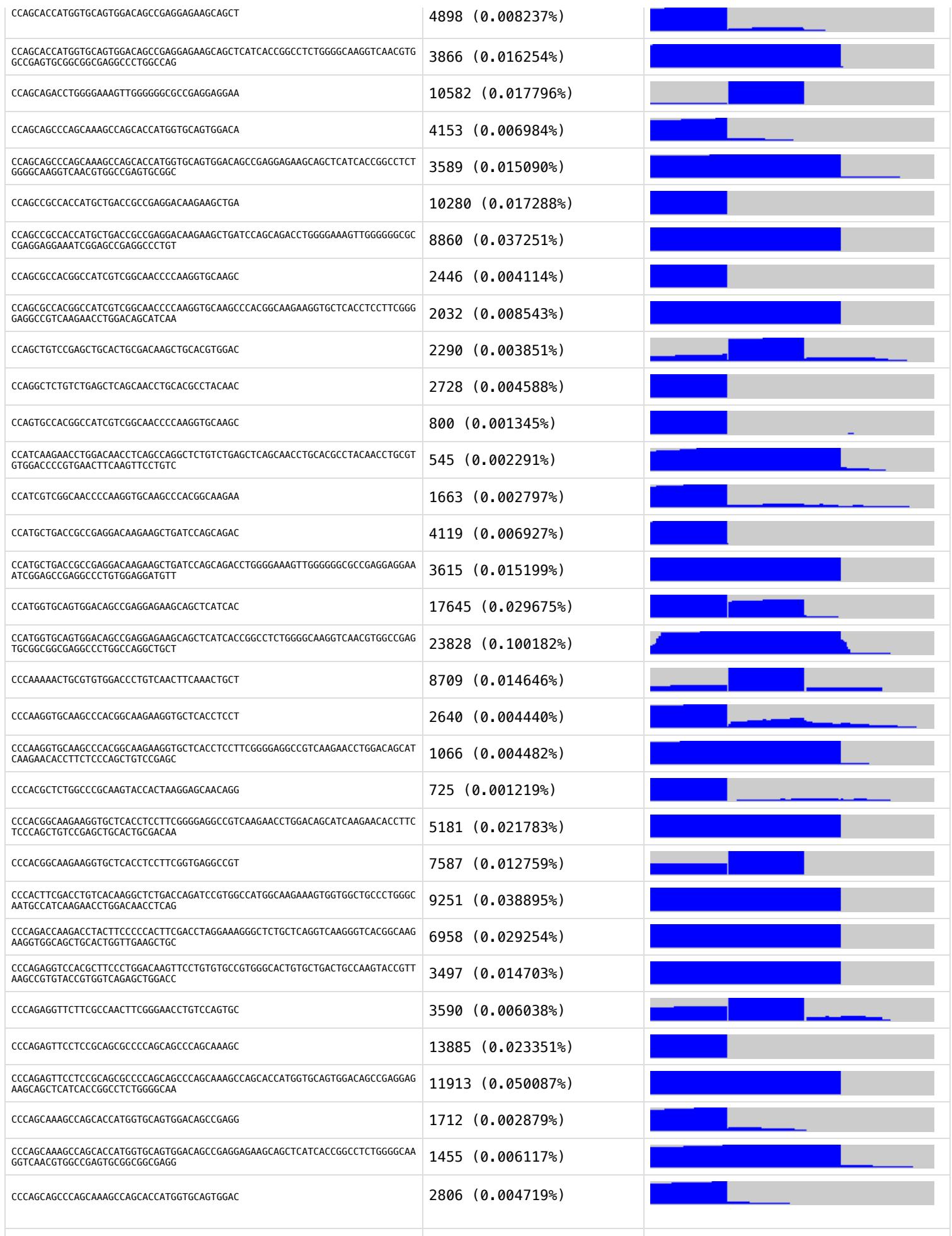




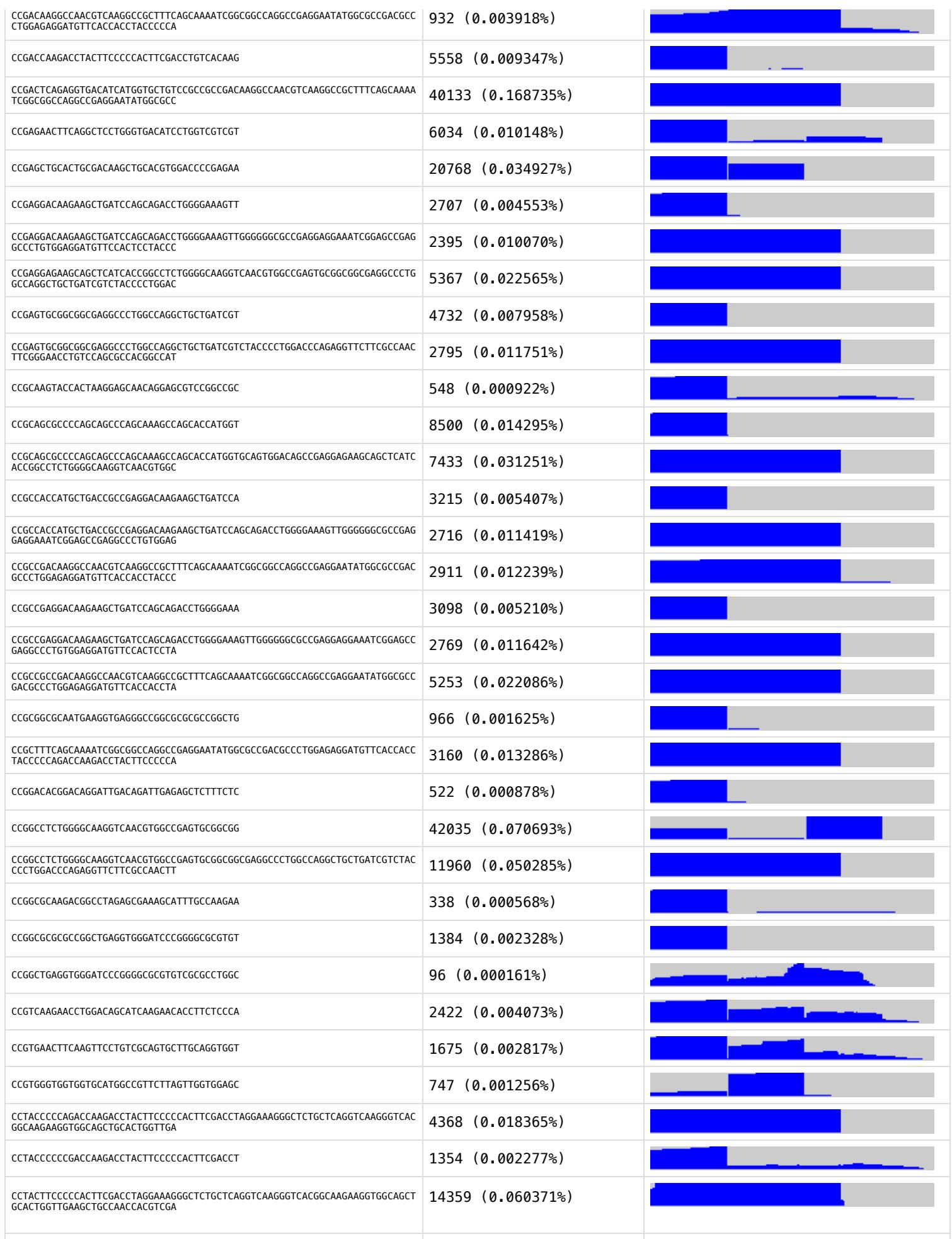


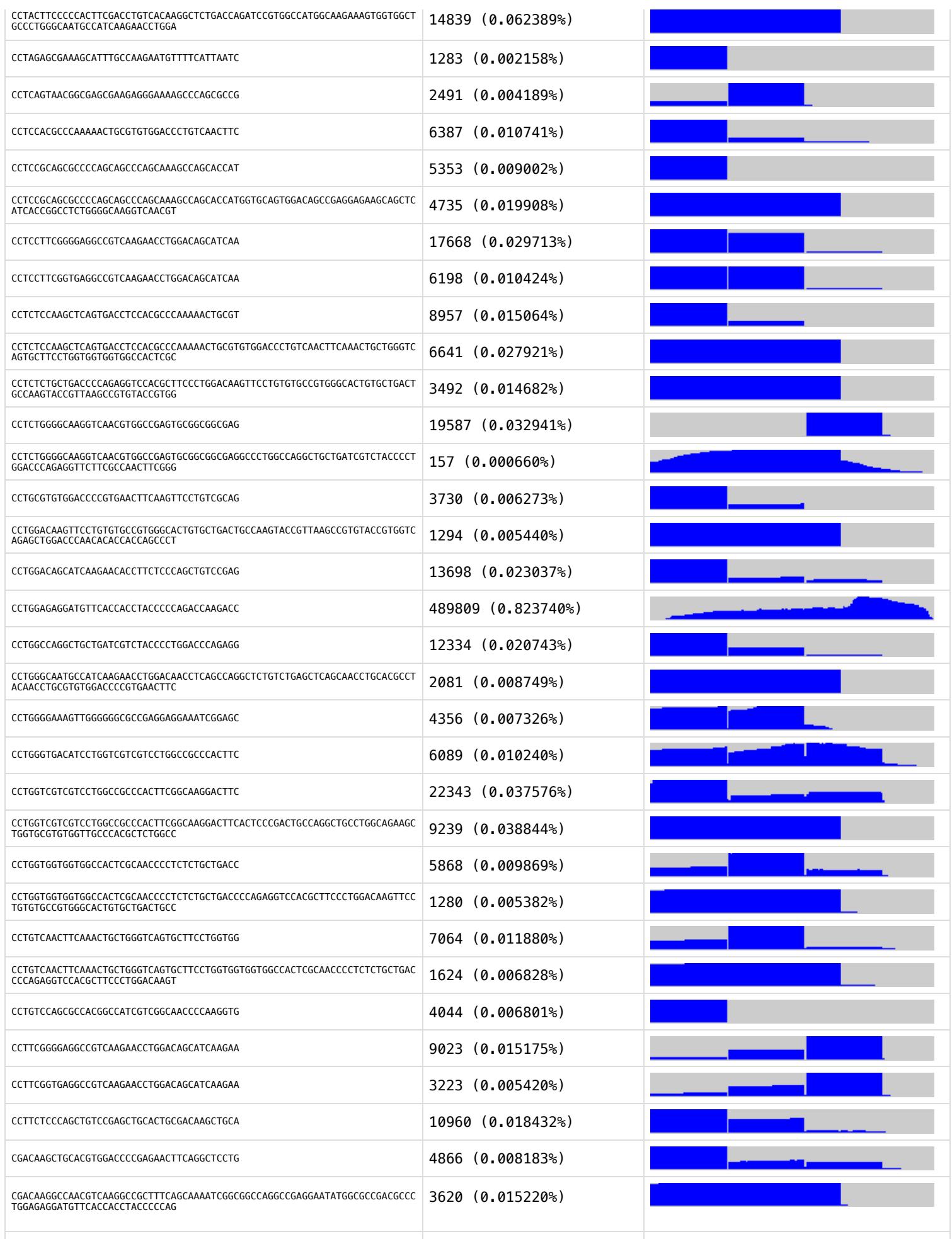




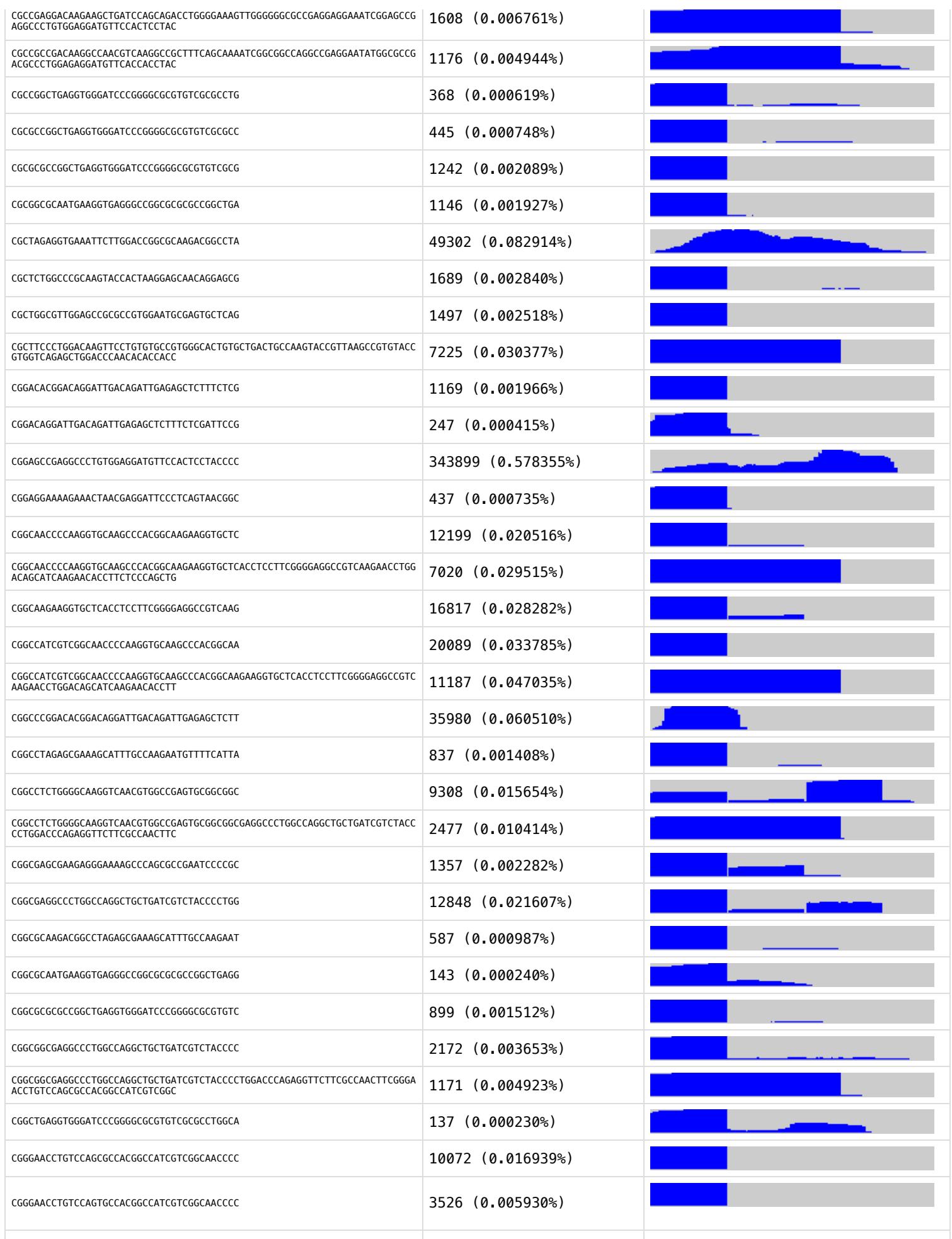


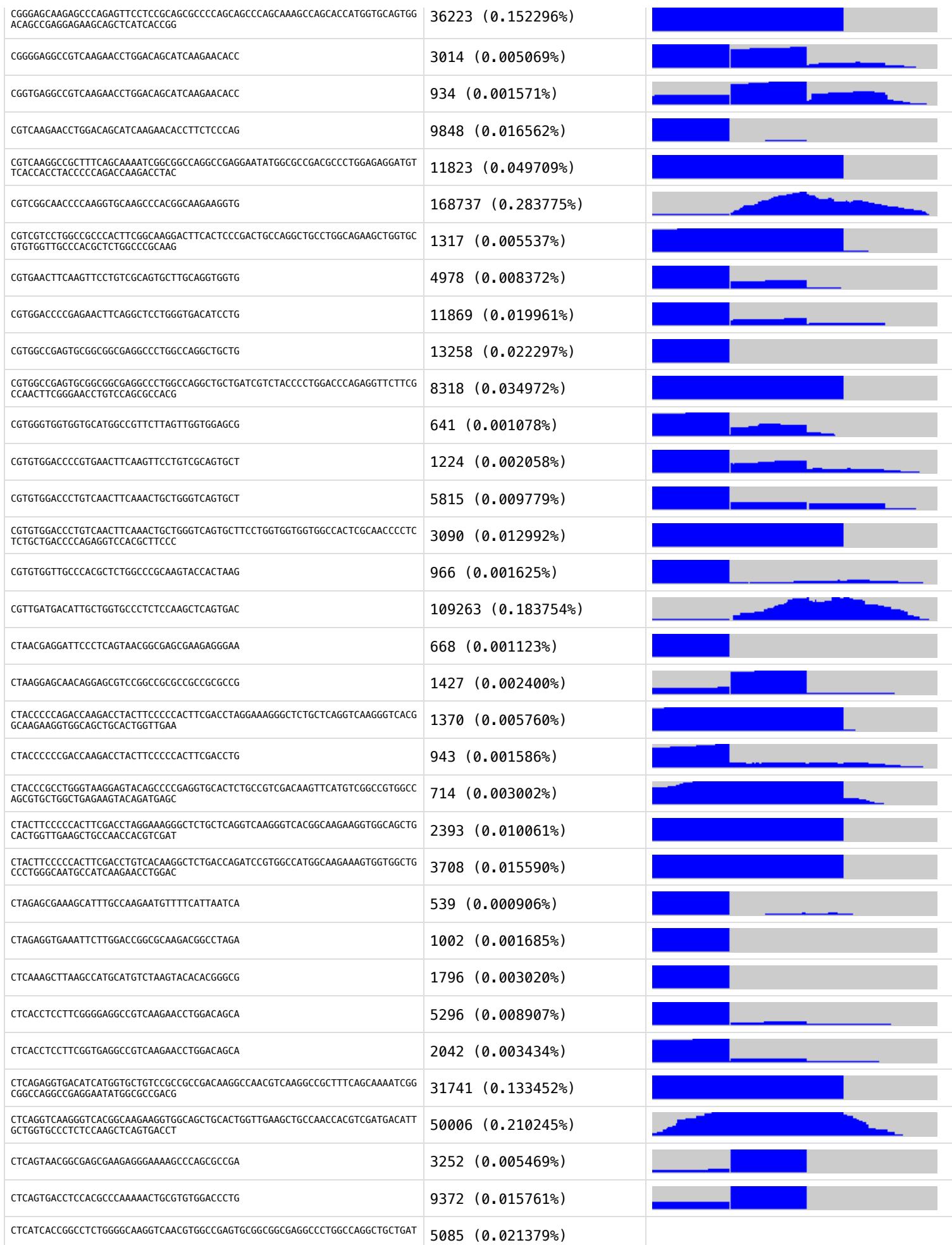


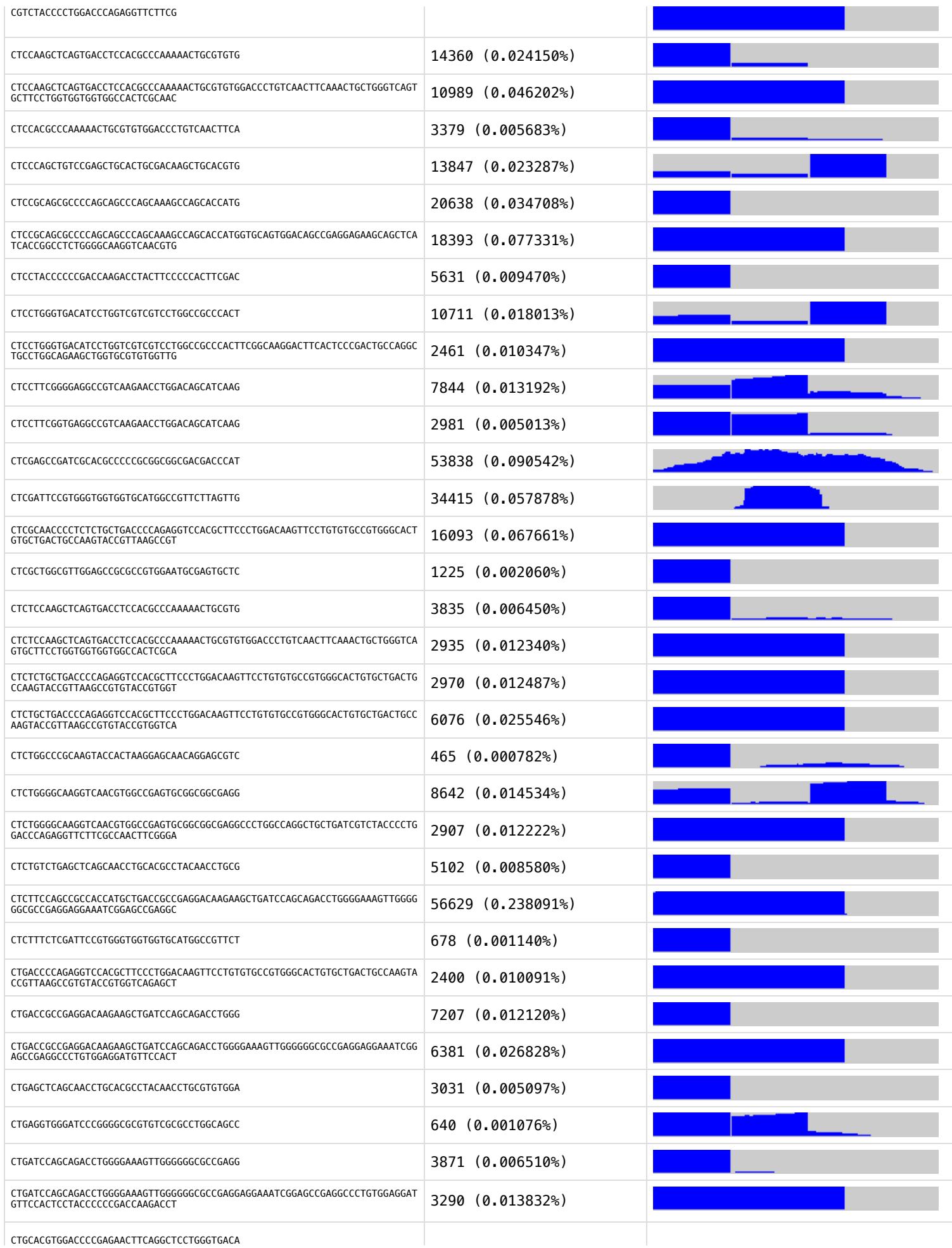


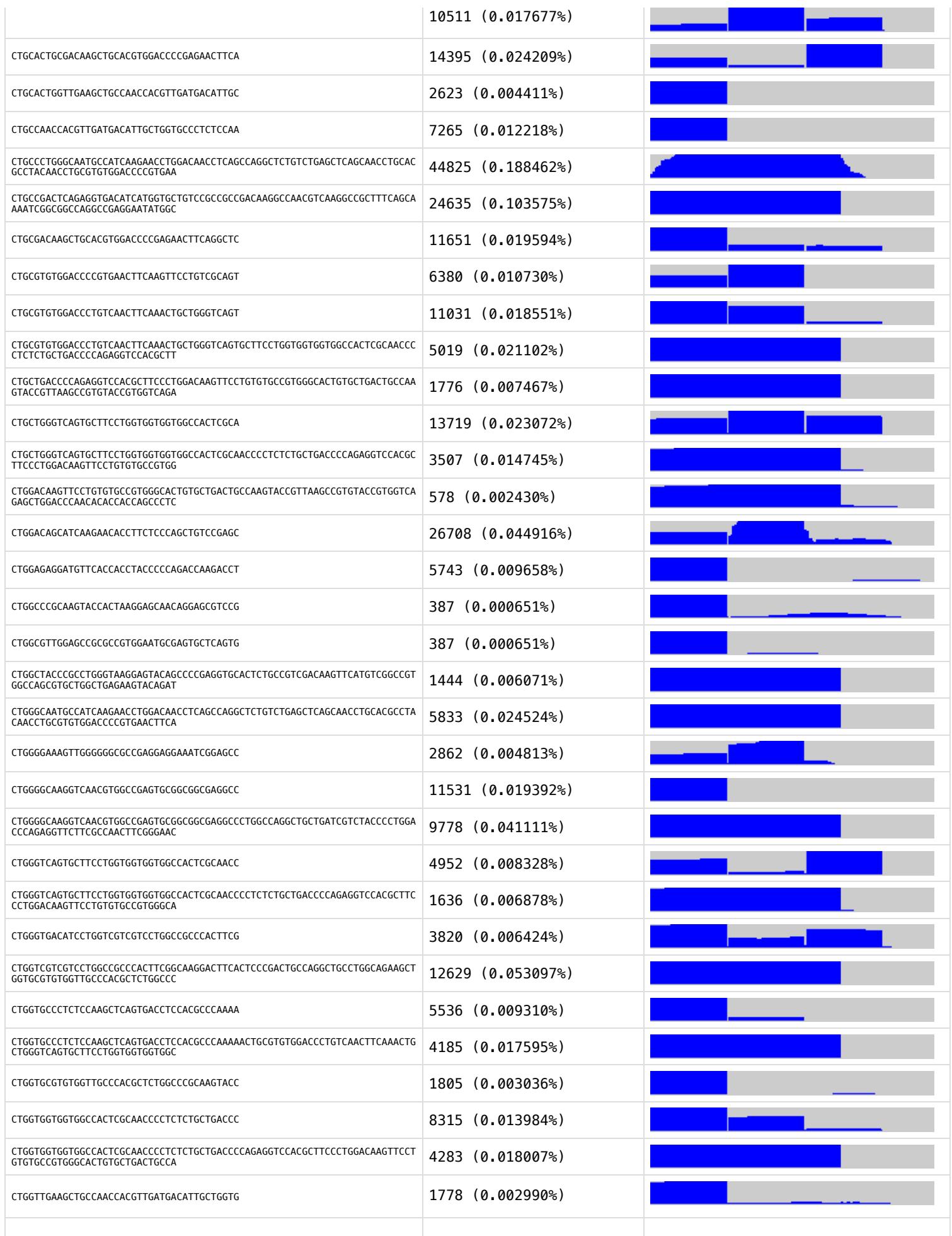




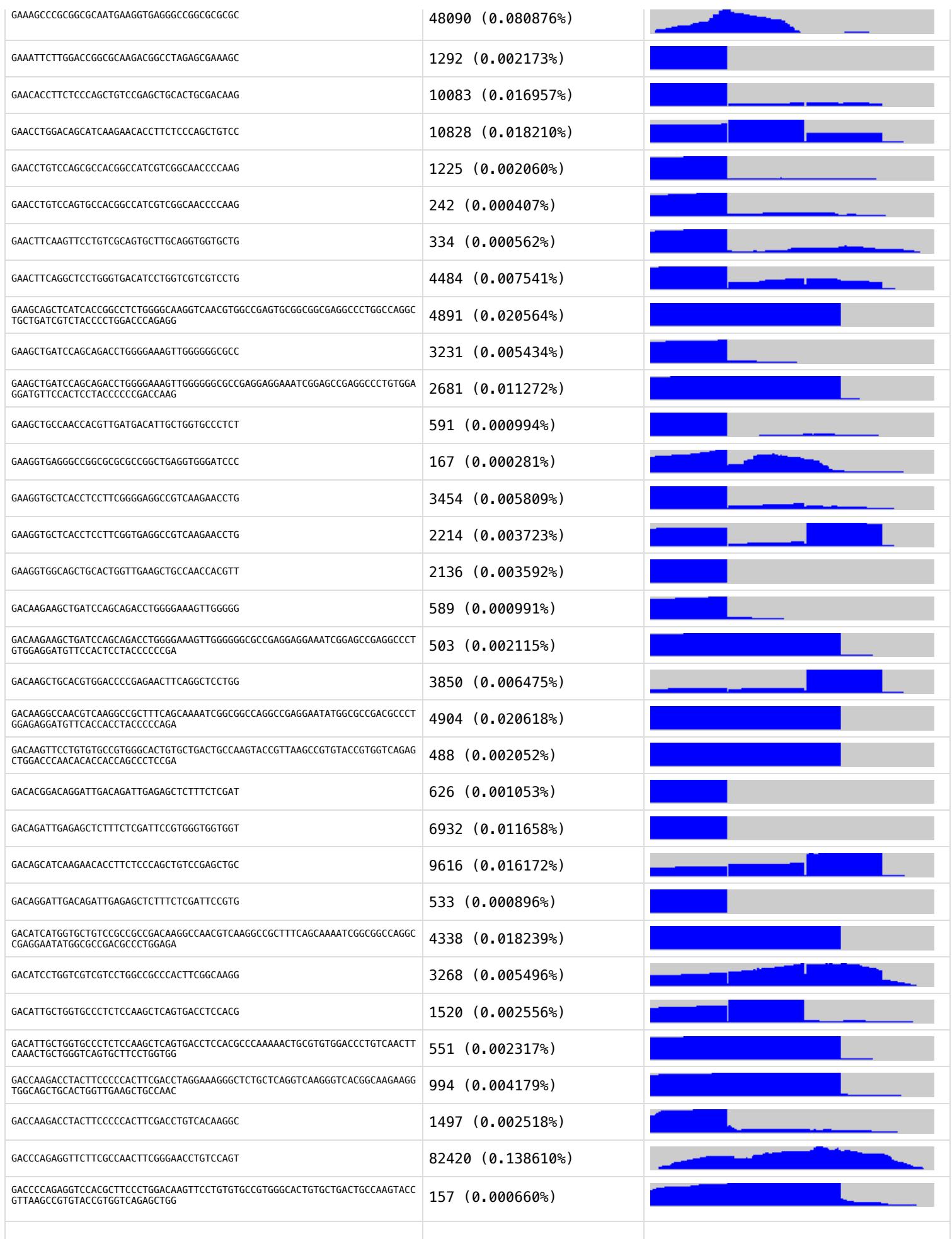


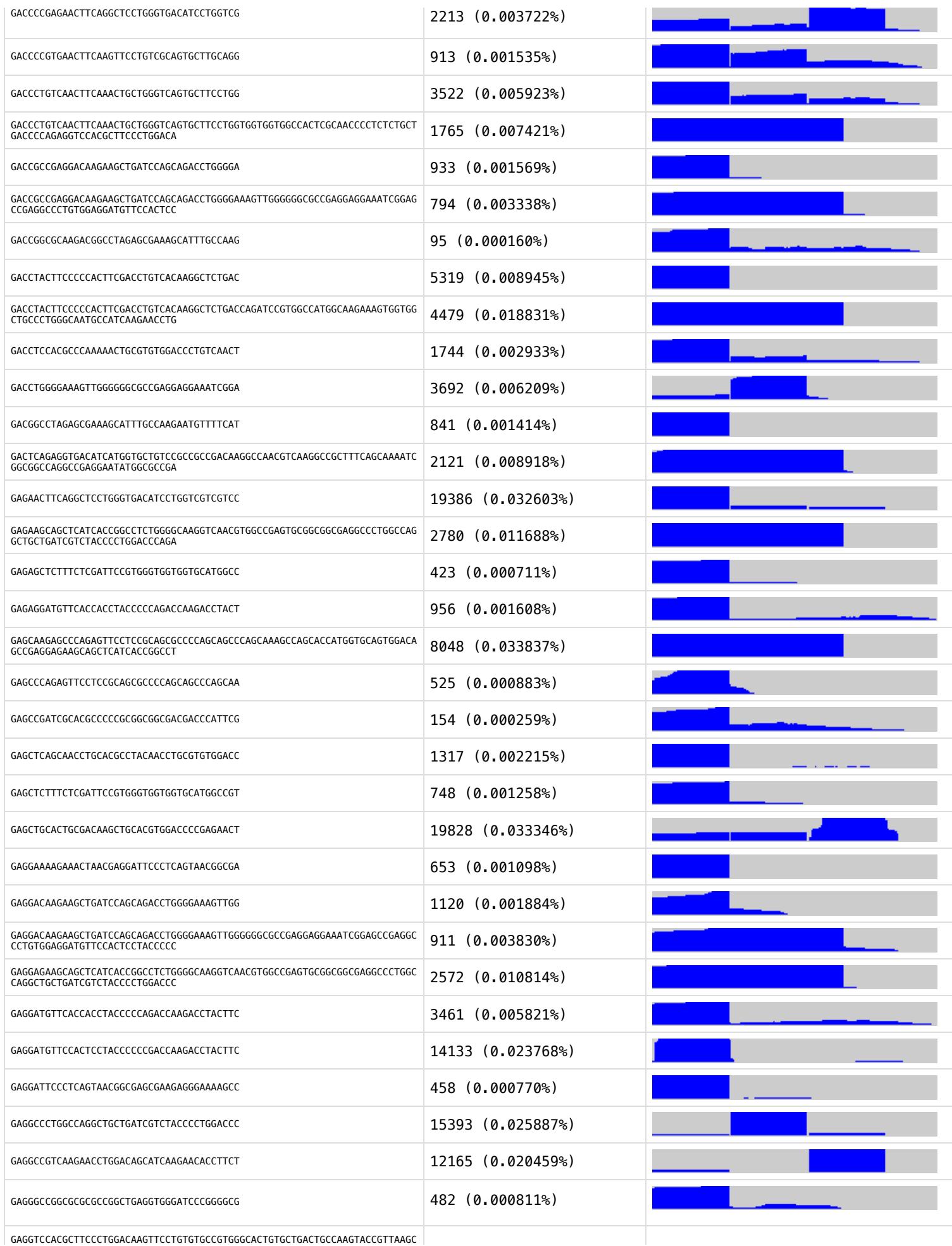




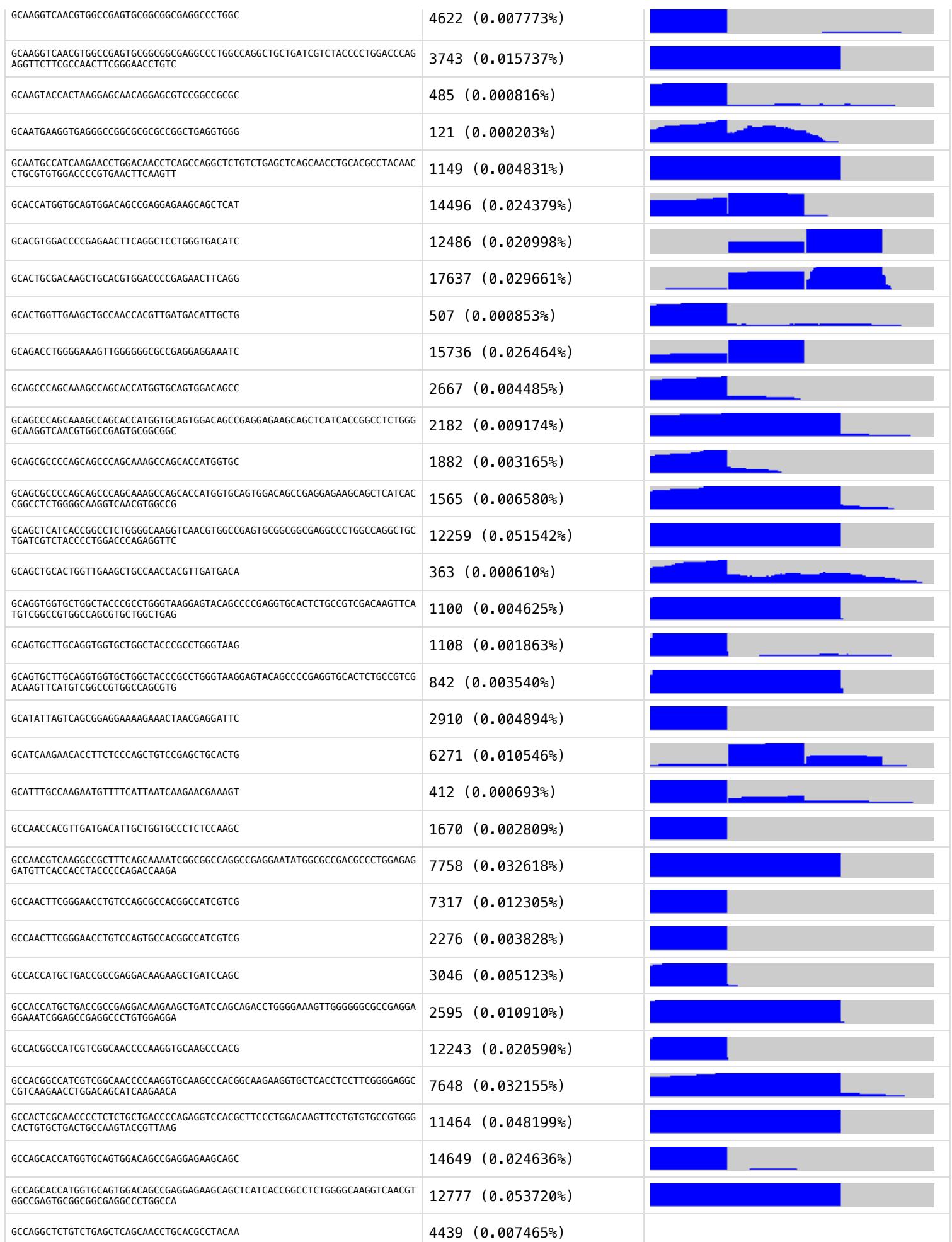


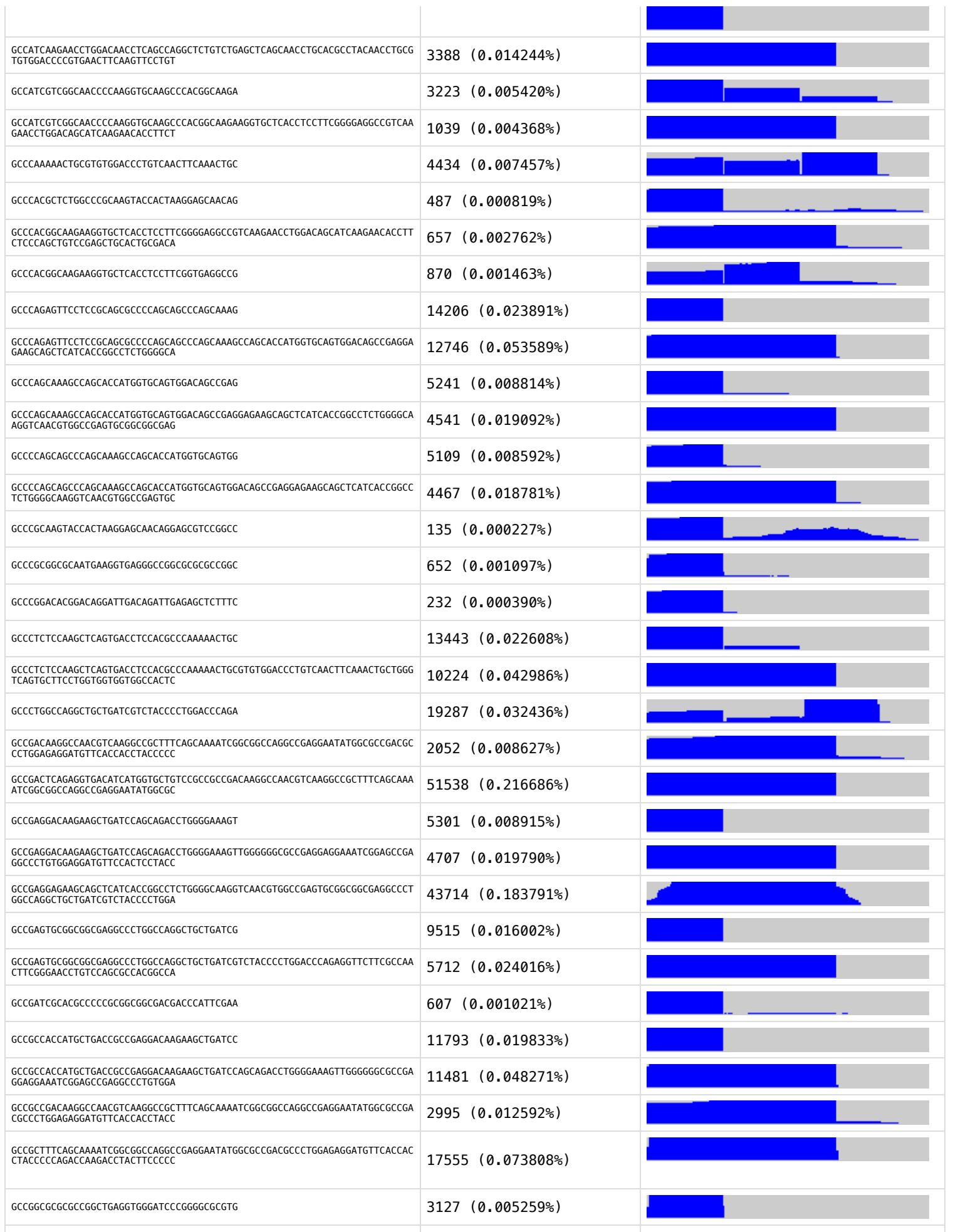


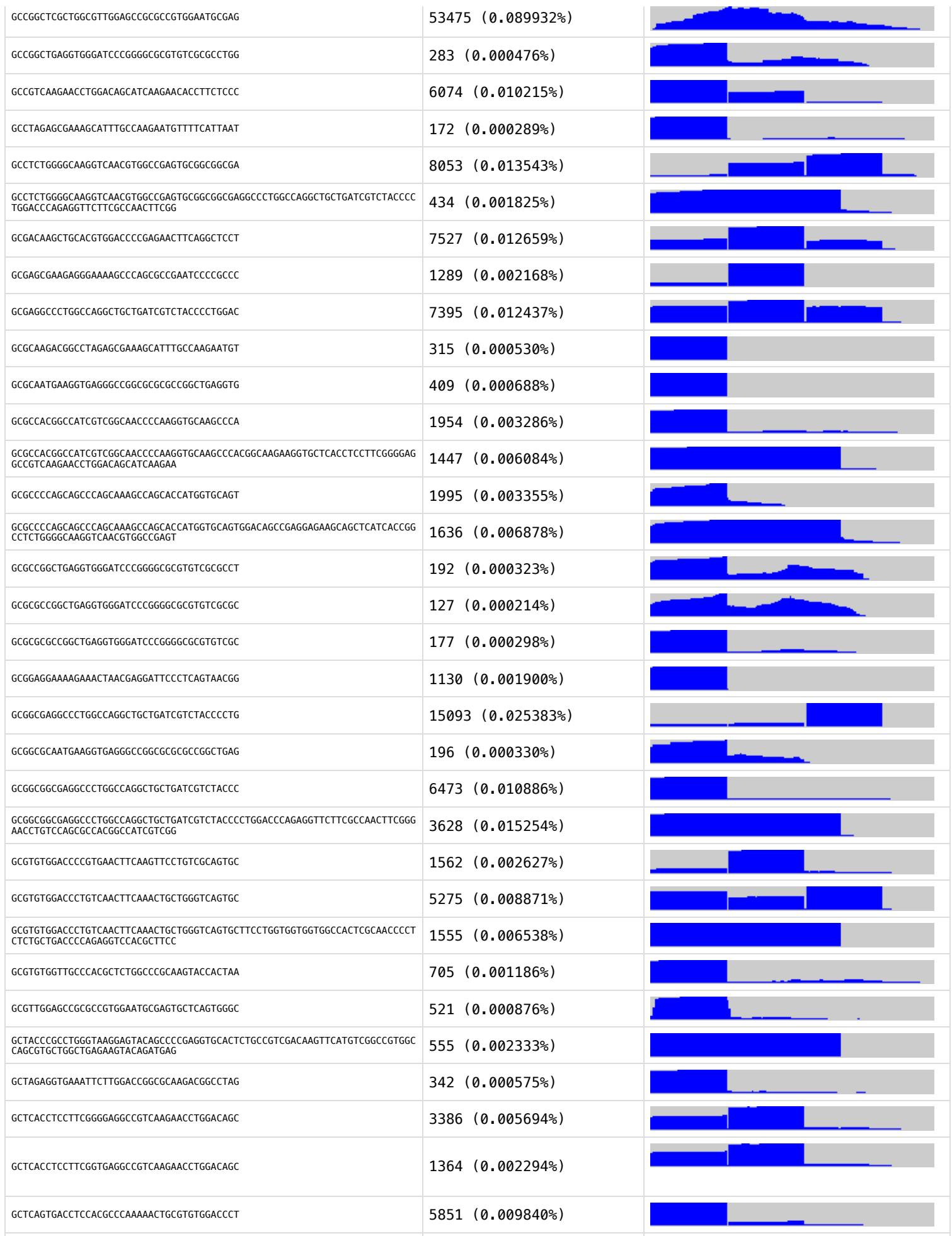


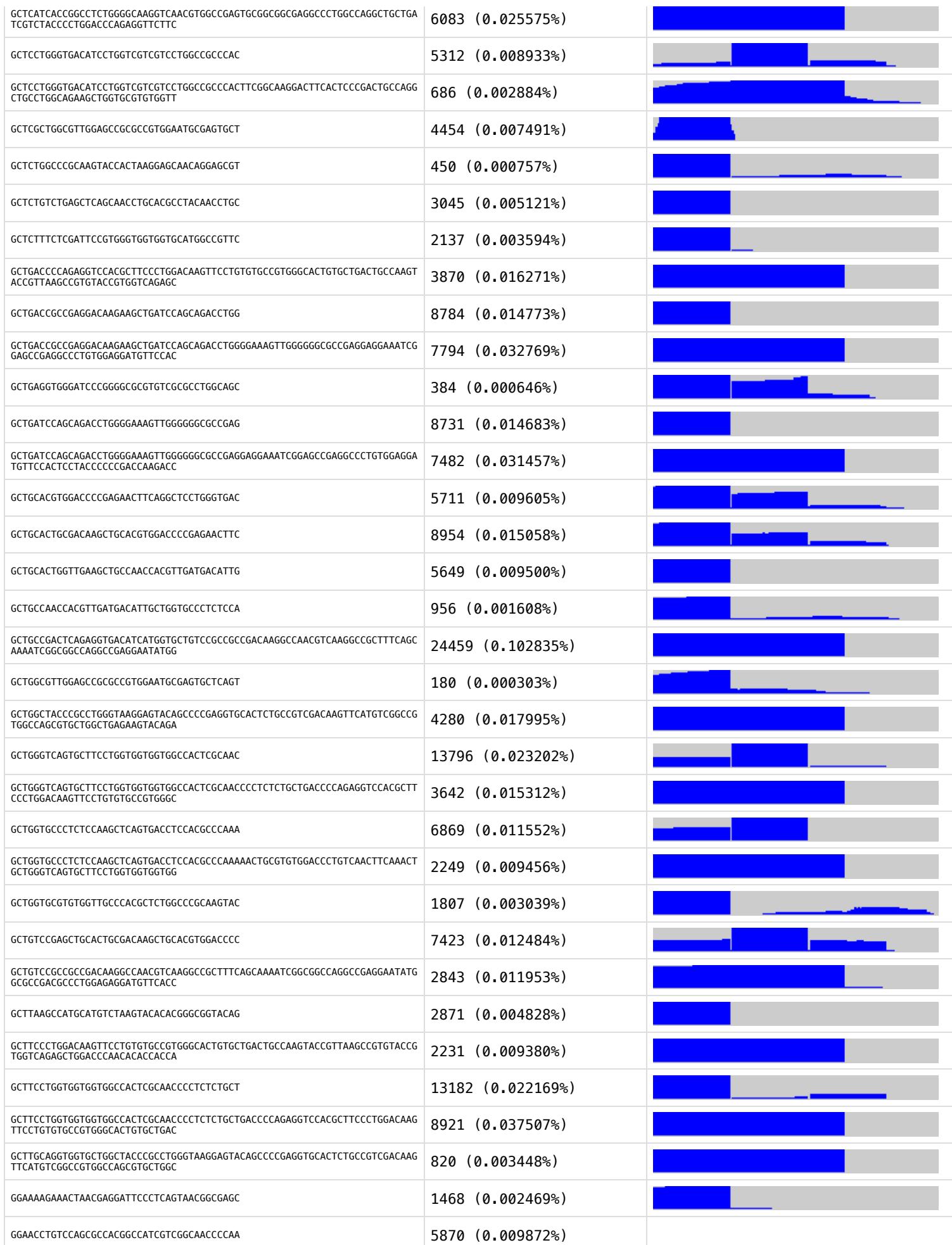


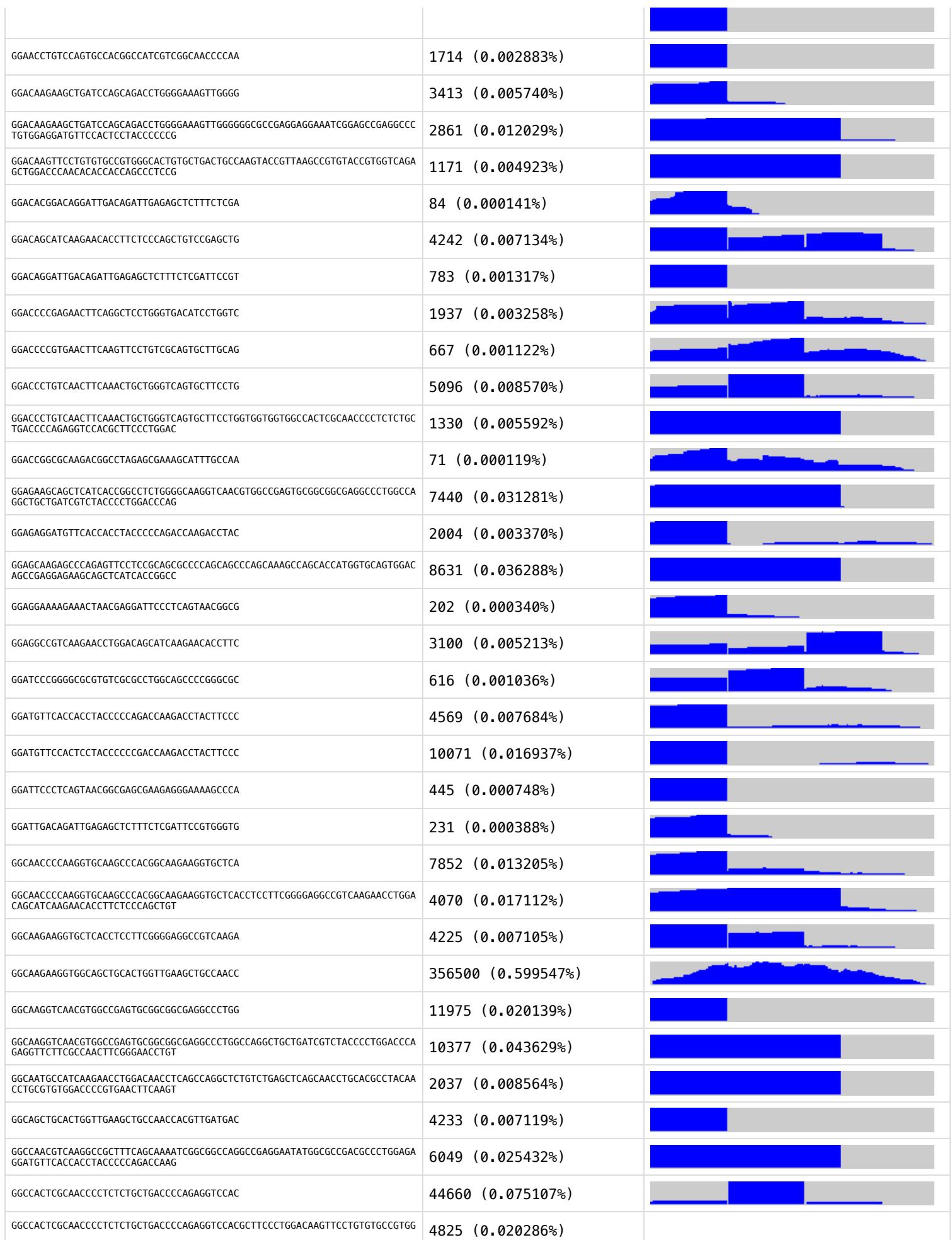


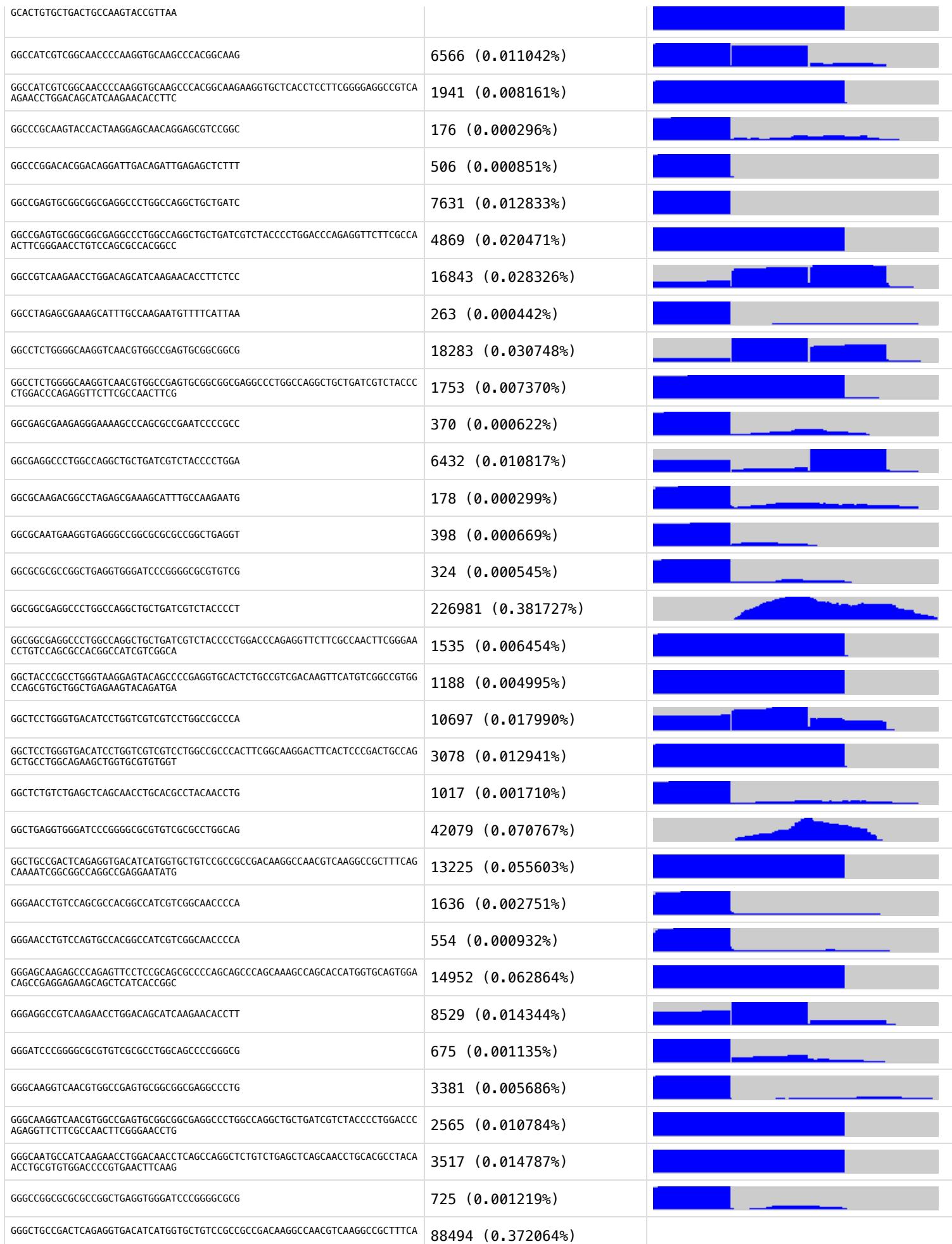


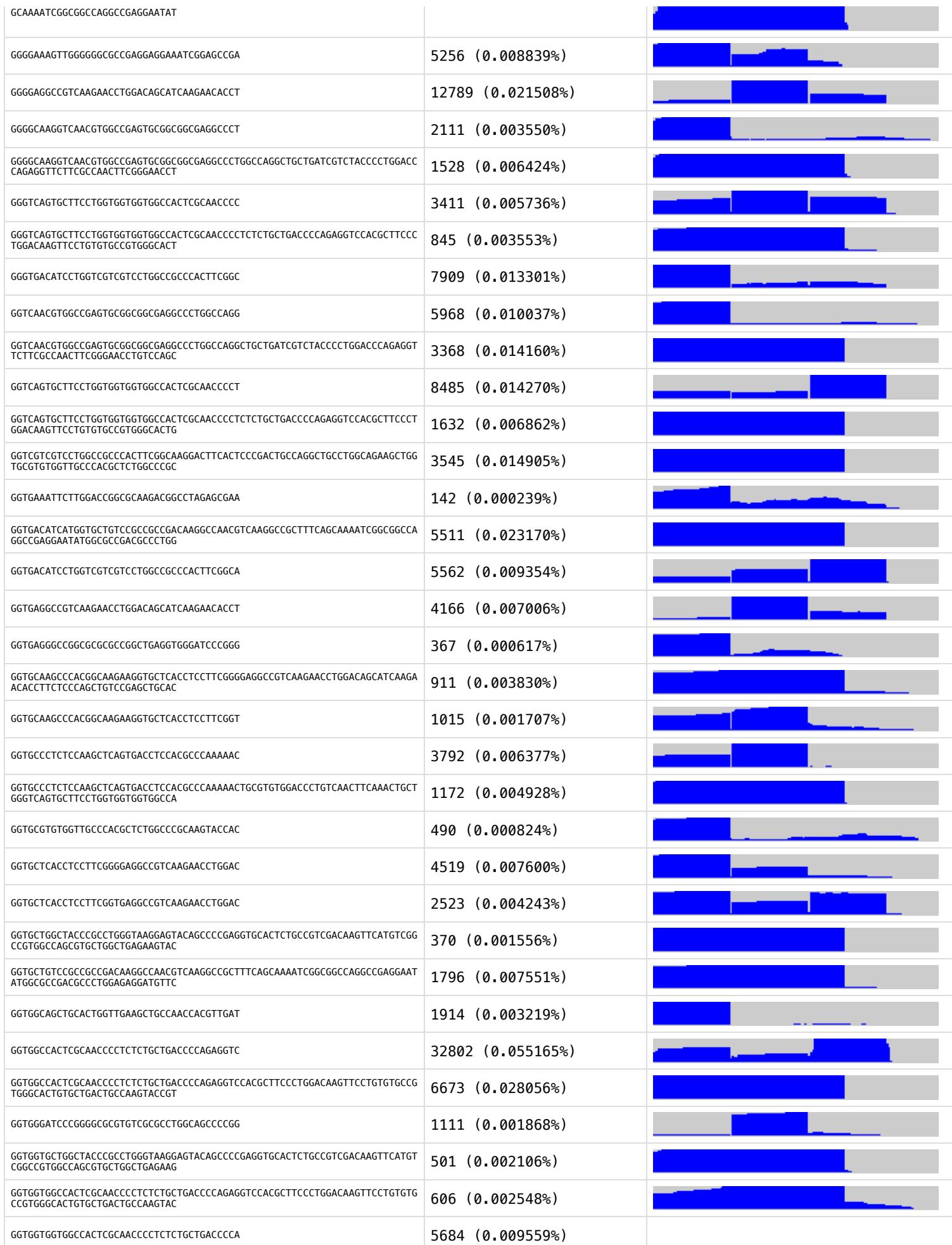


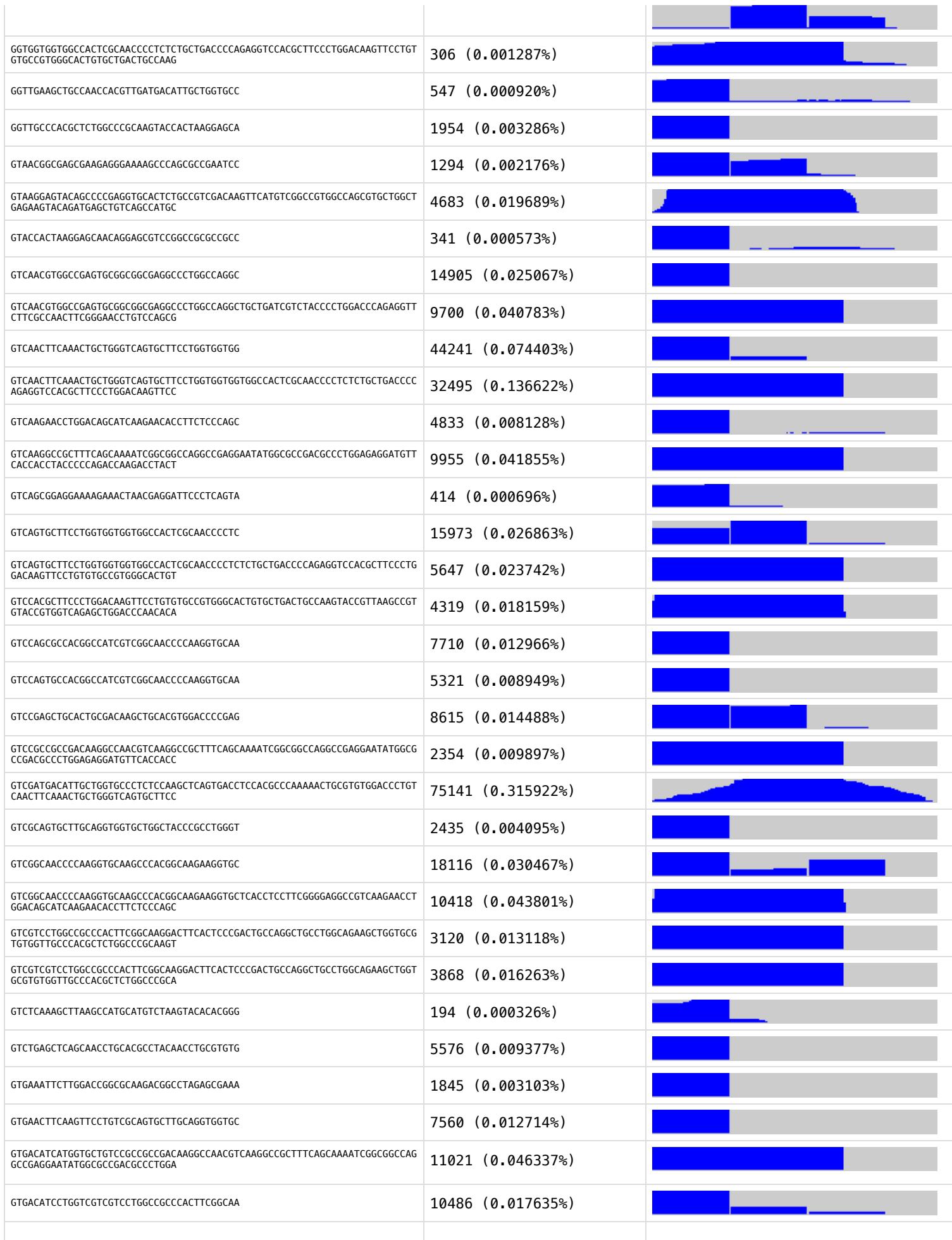


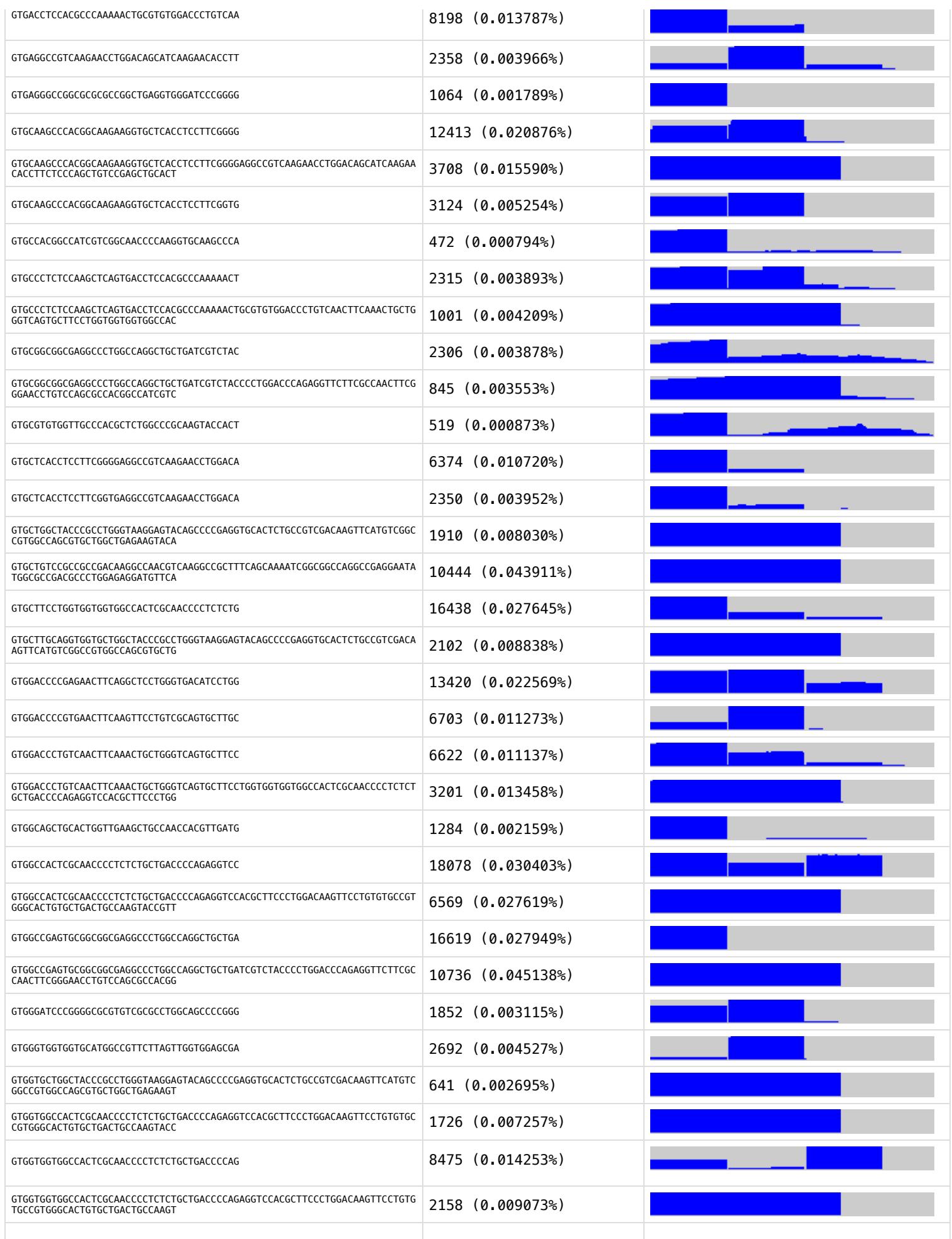


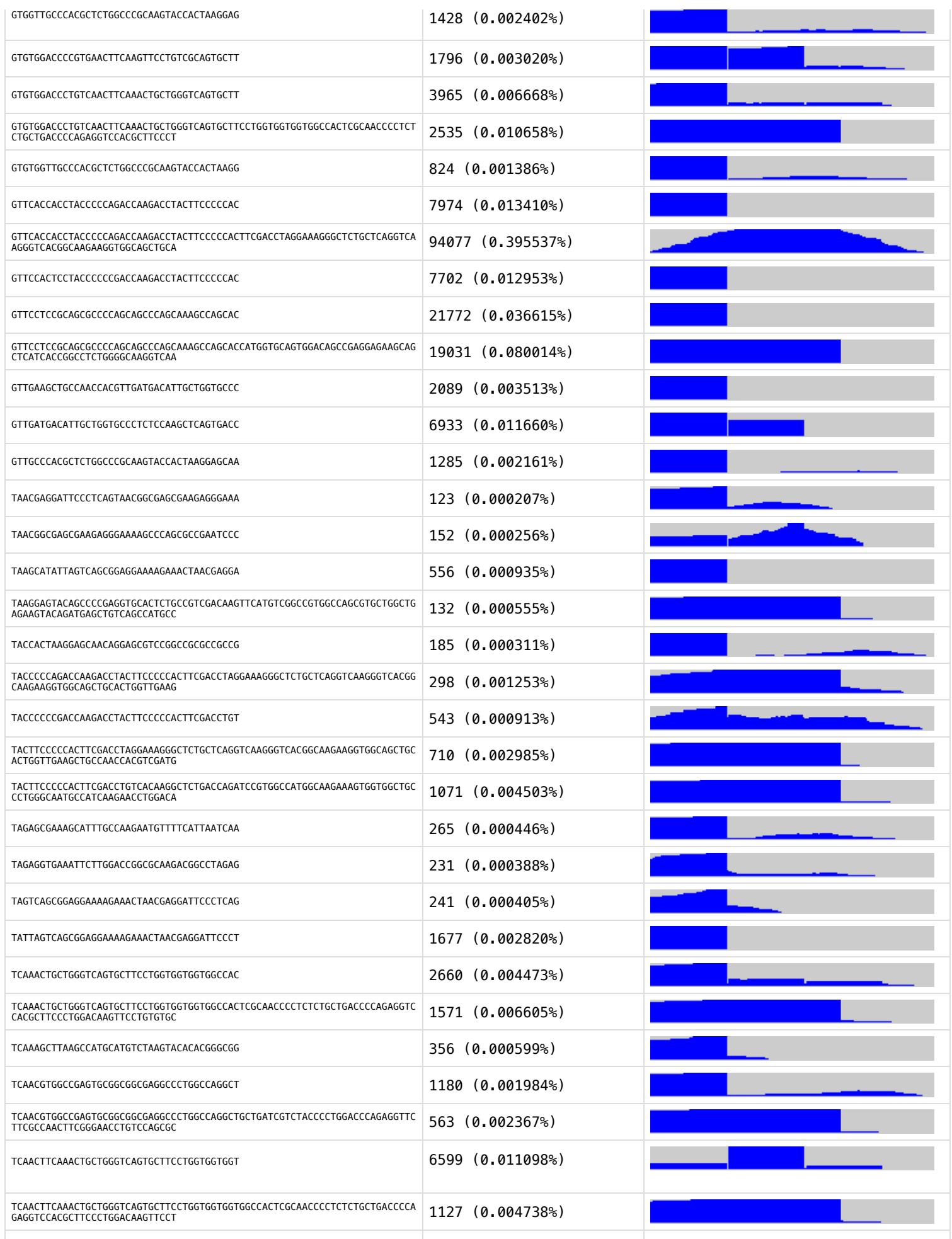


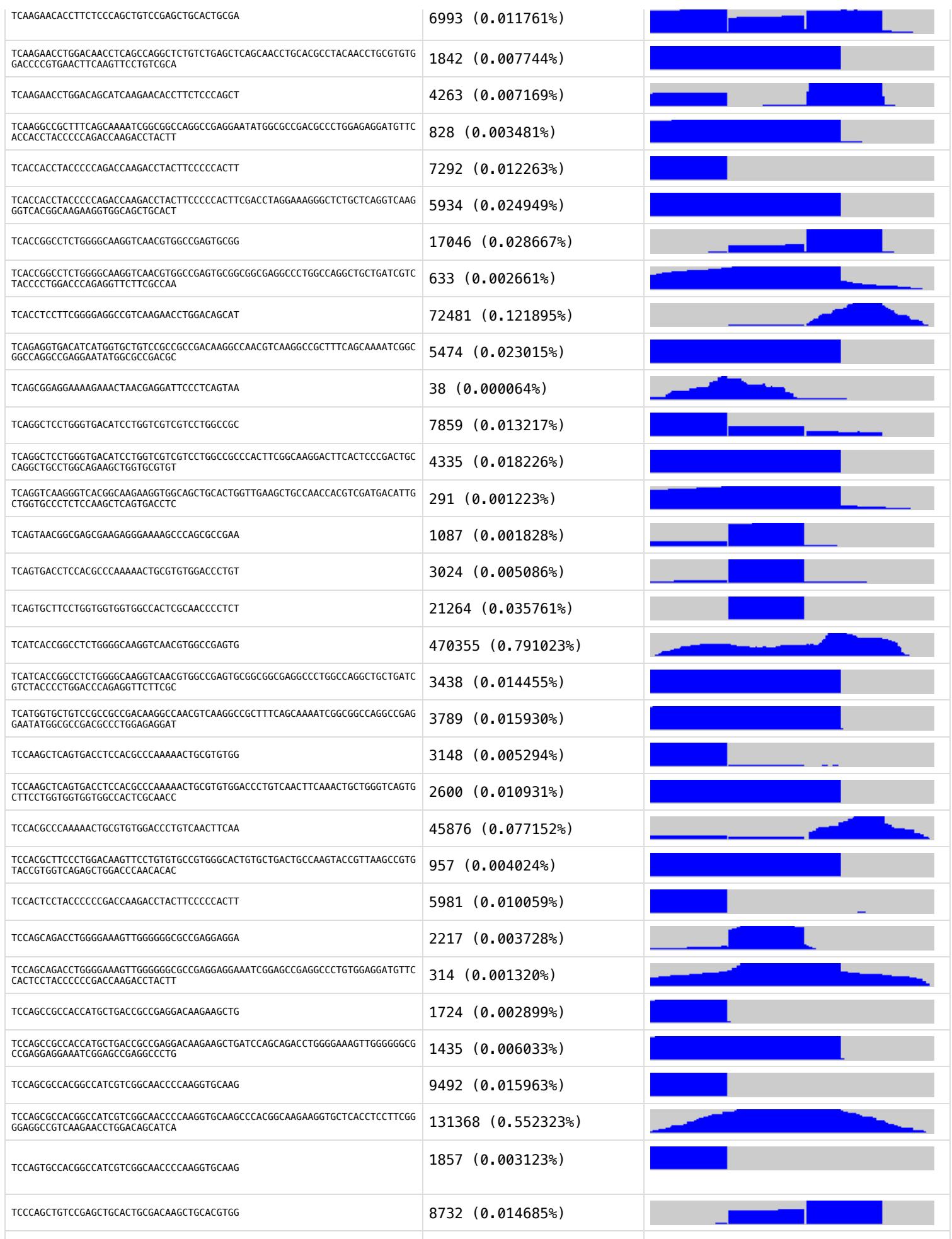




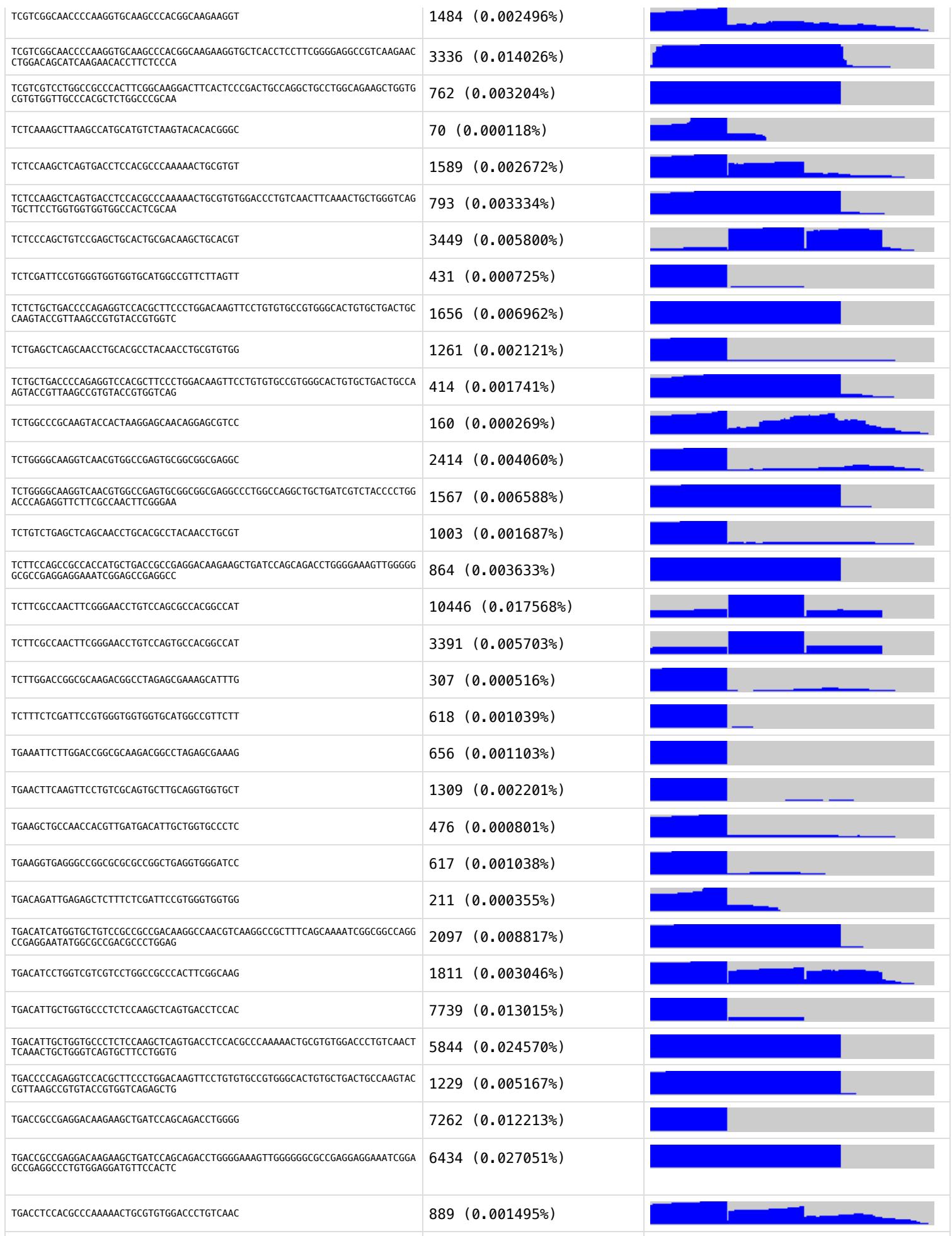


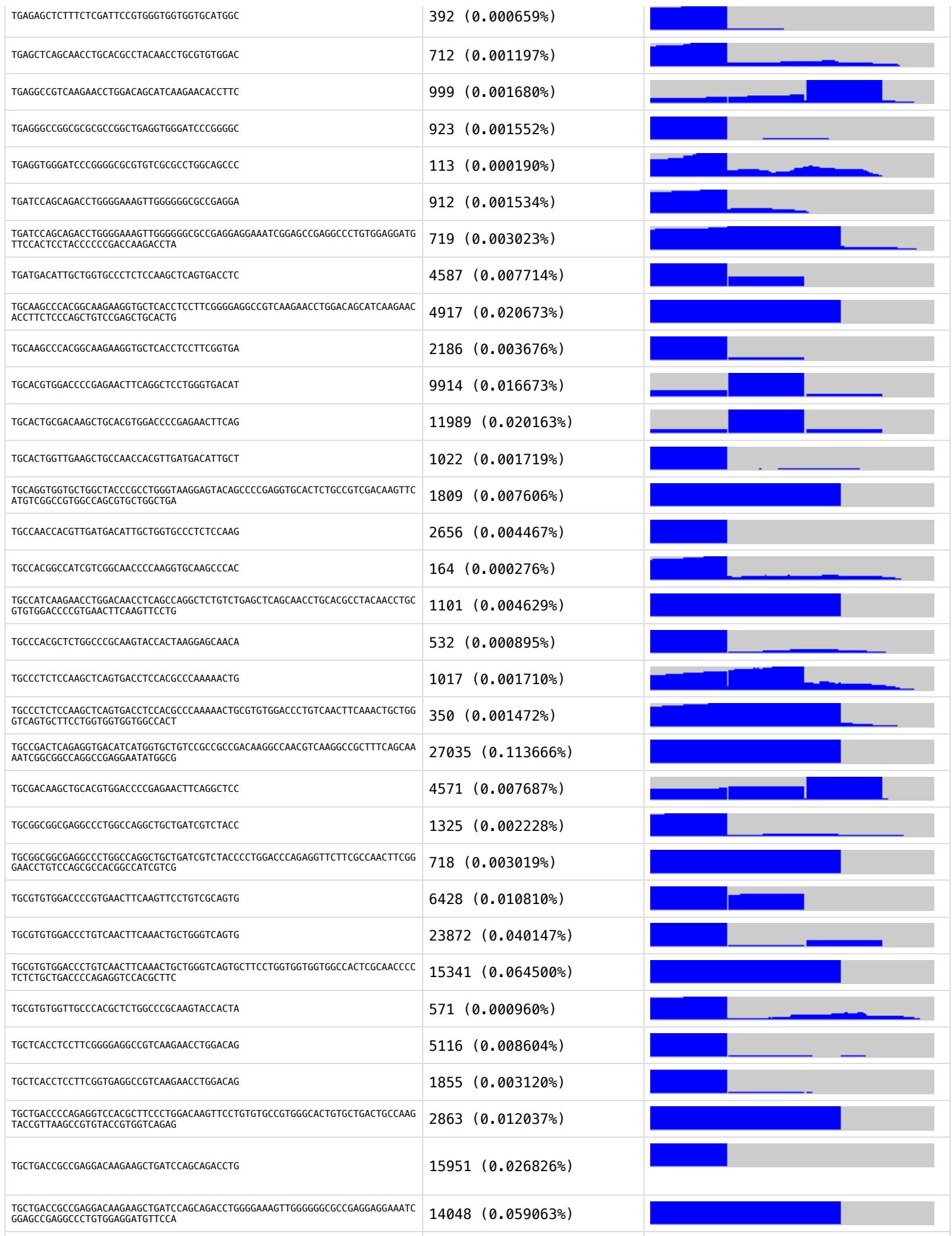




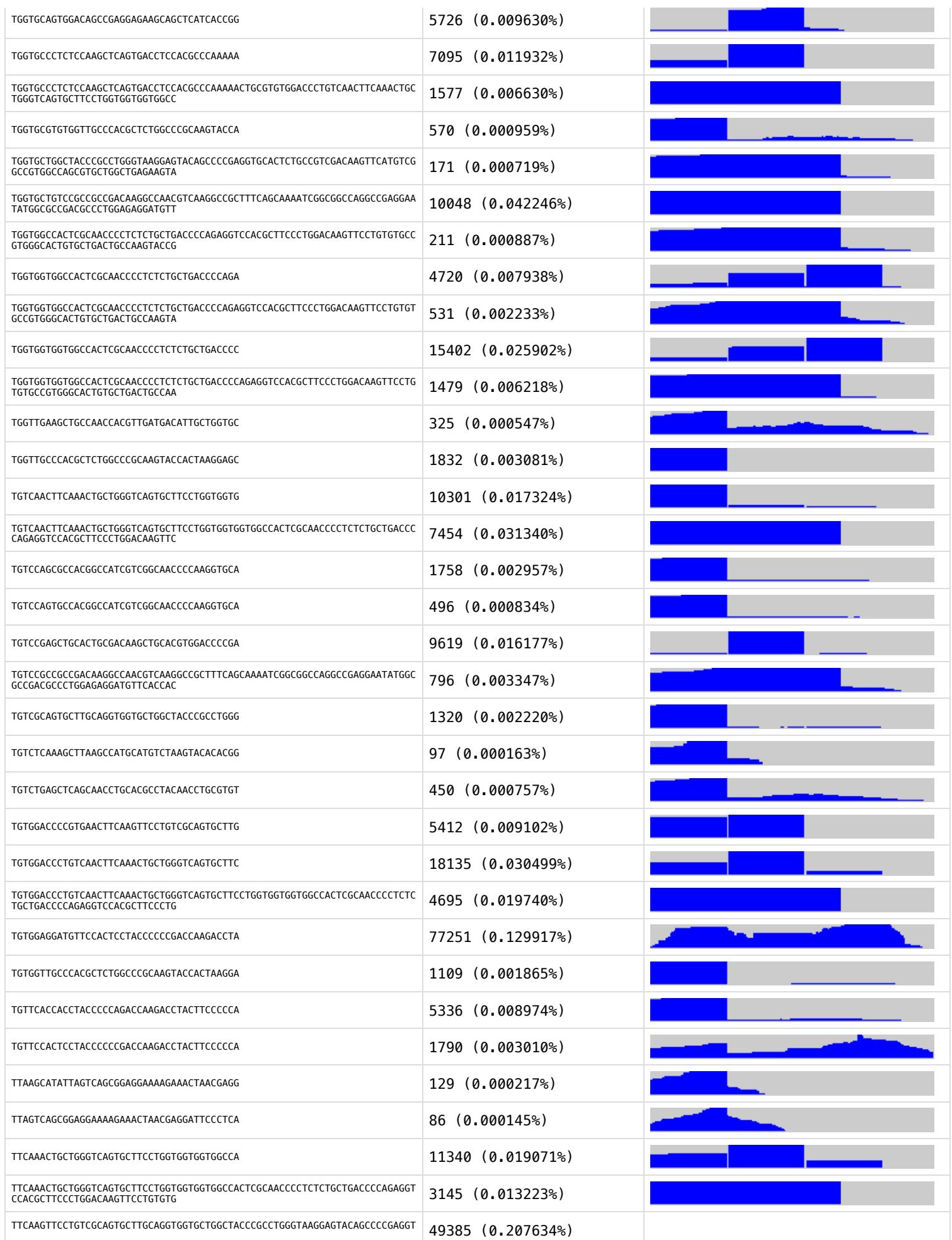


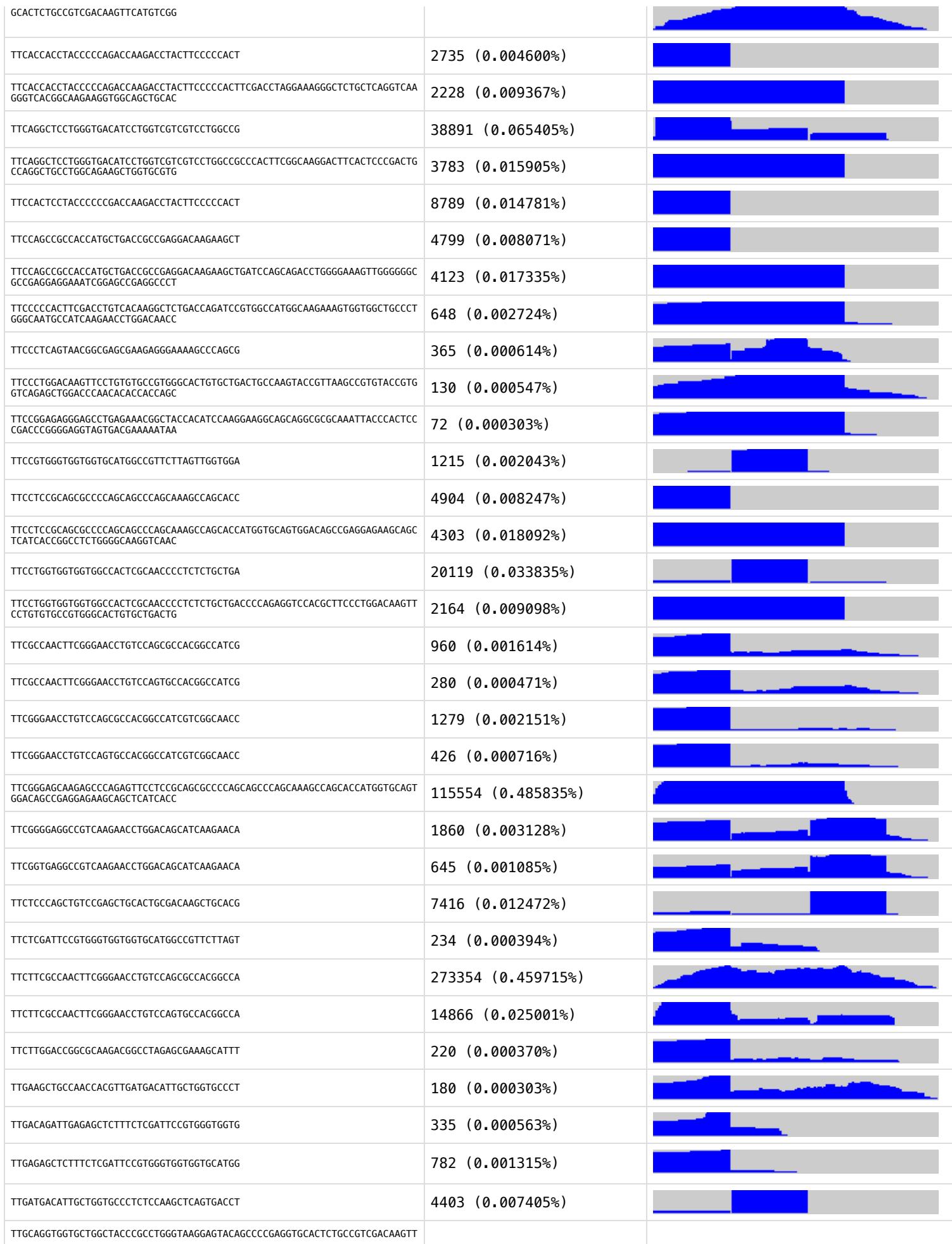


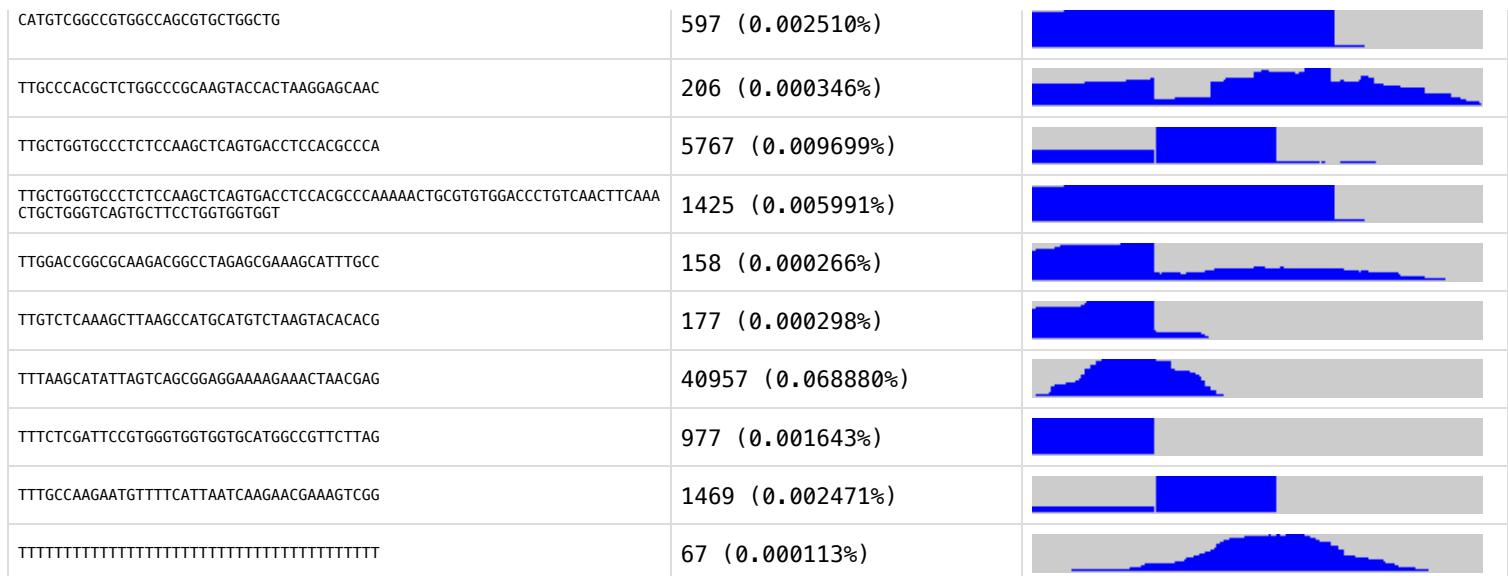












```
fastp --in1=H0WR-1_R1_001.fastq.gz --in2=H0WR-1_R2_001.fastq.gz --out1=H0WR-1_cleaned.R1.fastq.gz --out2=H0WR-  
1_cleaned.R2.fastq.gz --adapter.fasta=/home/rfitak/.bin/adapters.fa --cut_front --cut_tail --cut_window_size=4 --  
cut_mean_quality=20 --qualified_quality_phred=20 --average_qual=20 --unqualified_percent_limit=30 --n_base_limit=5 --  
length_required=50 --low_complexity_filter --complexity_threshold=30 --overrepresentation_analysis --trim_poly_x --  
poly_x_min_len=10 --html=H0WR-1.html --json=.json --report_title=H0WR-1 --thread=16
```

fastp 0.20.0, at 2020-04-09 10:45:01