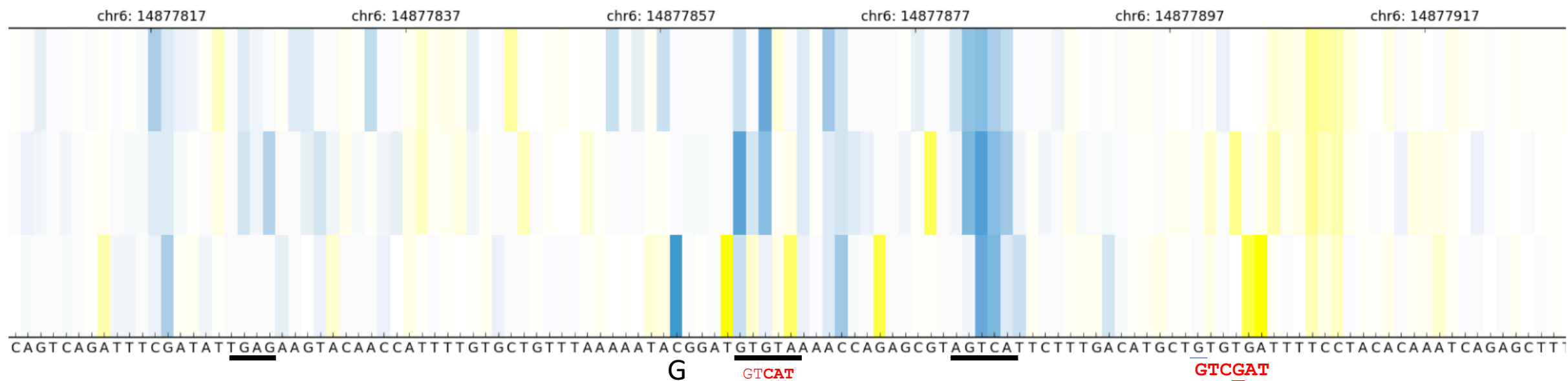


# ATF2

Each of these is a different genomic location used as input for DeepSEA. Each subsequent slide for this TF is of the specified loci, in order (i.e. chr6 is on the next slide, chr1 is on the following slide)

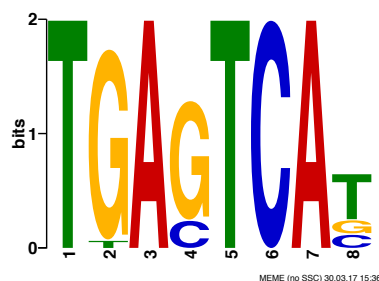
chr6:14877358-14877458	+	9	1.68e-05	GGATACGG TGAGTCAT
TTCAACTTCC				
chr1:38493616-38493716	+	20	1.68e-05	TGAATCAGCC TGAGTCAT
TTACAGCTTT				
chr18:49265616-49265716	+	41	1.68e-05	CAGGTGCGAA TGAGTCAT
GCACAGATTA				
chr2:145131768-145131868	+	62	1.68e-05	G TTCCTTATC TGAGTCAT
TGCTTAAATA				
chr11:18637681-18637781	+	35	1.68e-05	AAGTATCAAT TGAGTCAT
CTTTACATAA				

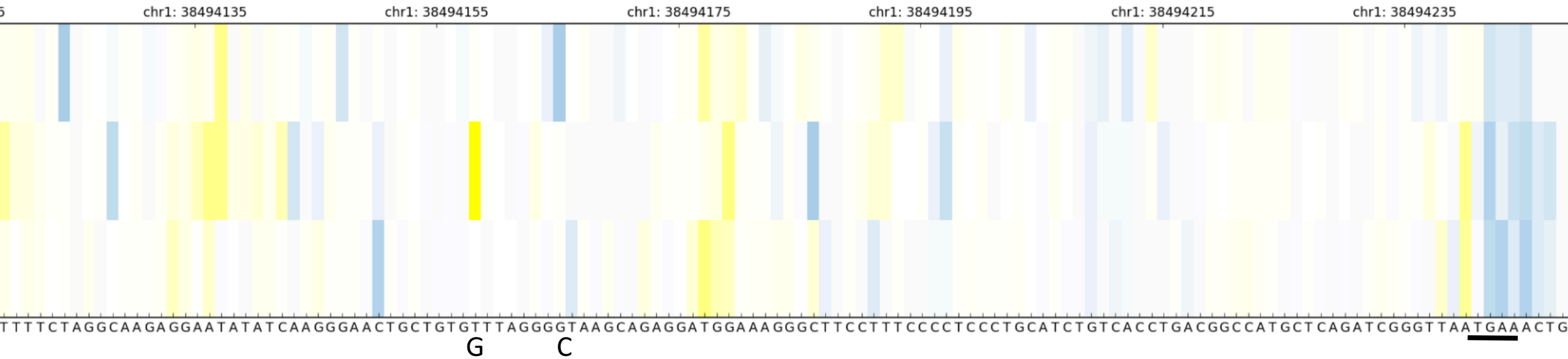


TGAGTCAT  
A>G>C>T>A

overrepresented  
mutations result in  
regained binding site

Potential obstruction of  
binding due to this G,  
otherwise binding site  
regained

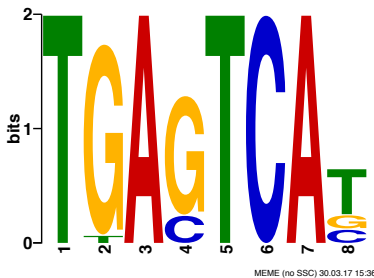


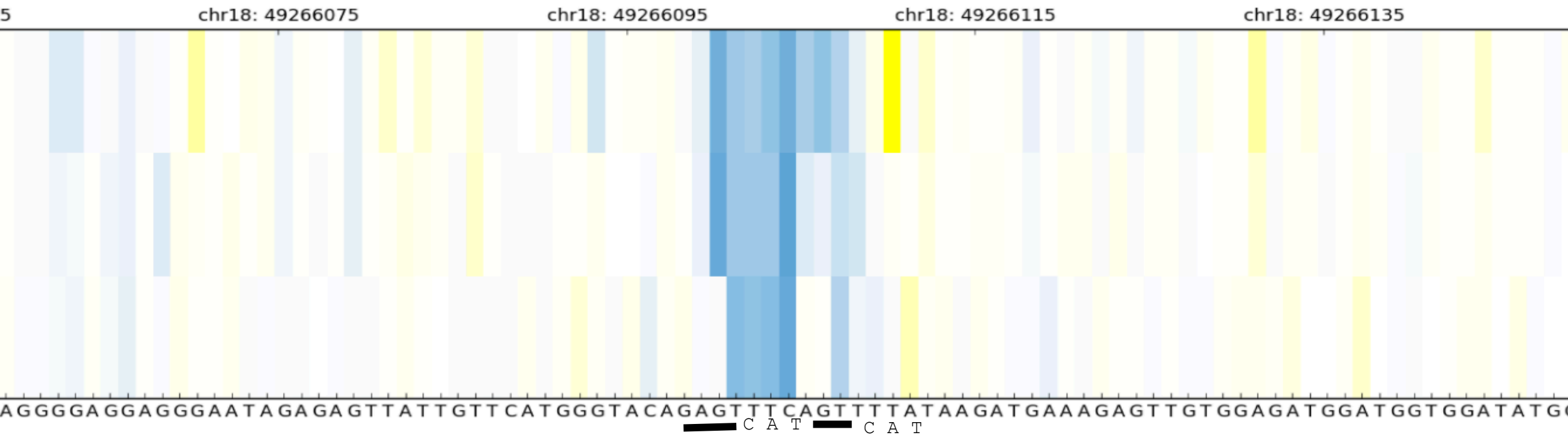


GM12878|DNase|None

TGAGTCAT

A>G>C>T>A



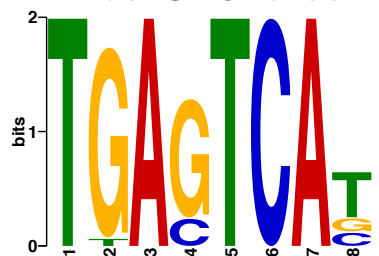


GM12878|DNase|None

Second GTTCAT  
Motif may be regained or  
potential obstruction of  
binding by extra T

TGAGTCAT

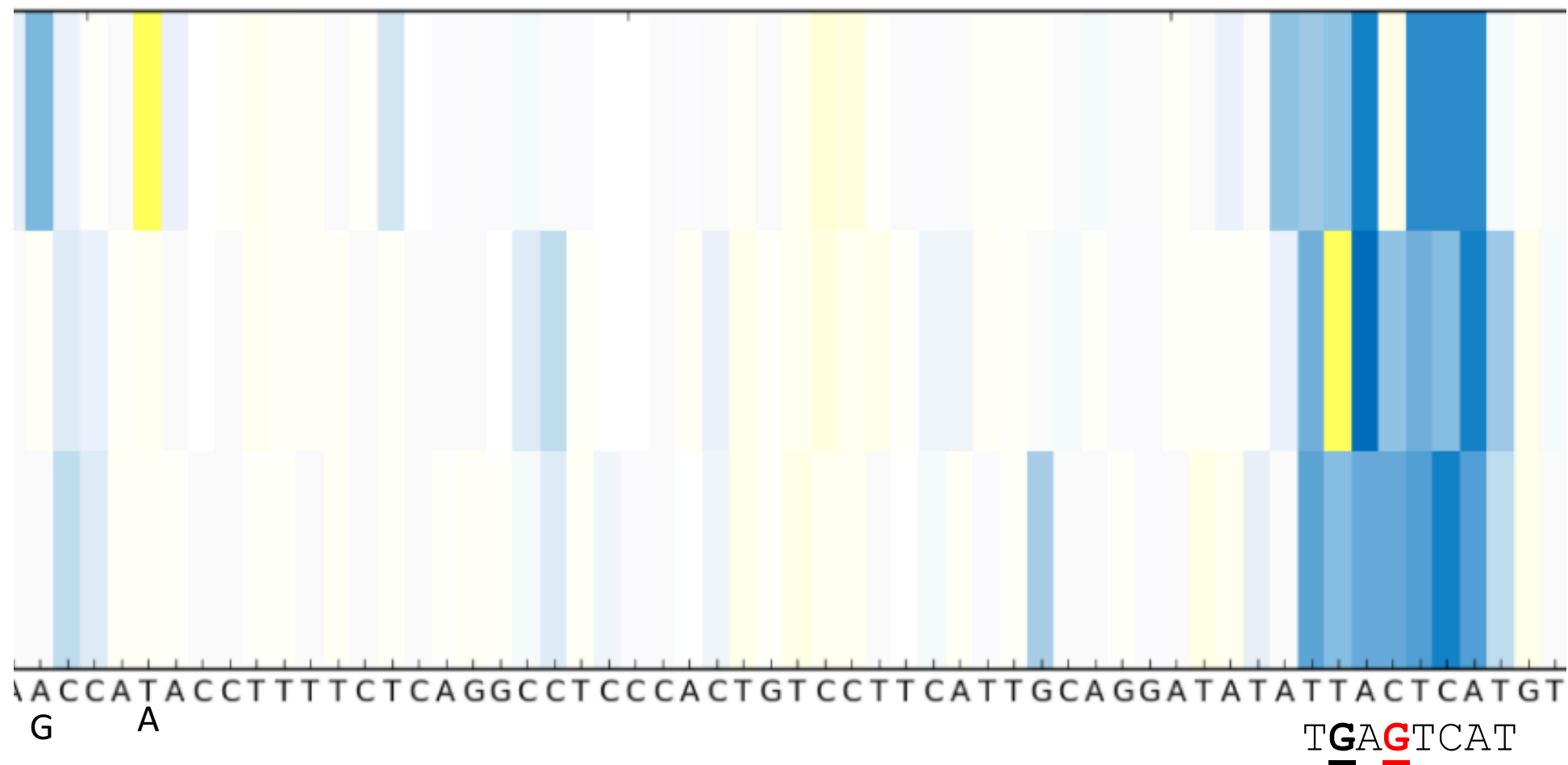
A>G>C>T>A



2: 145132247

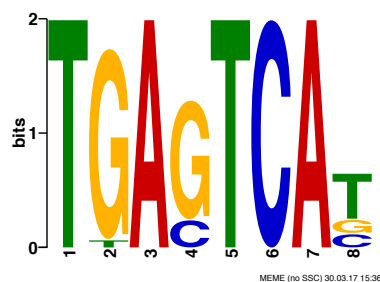
chr2: 145132267

chr2: 145132287

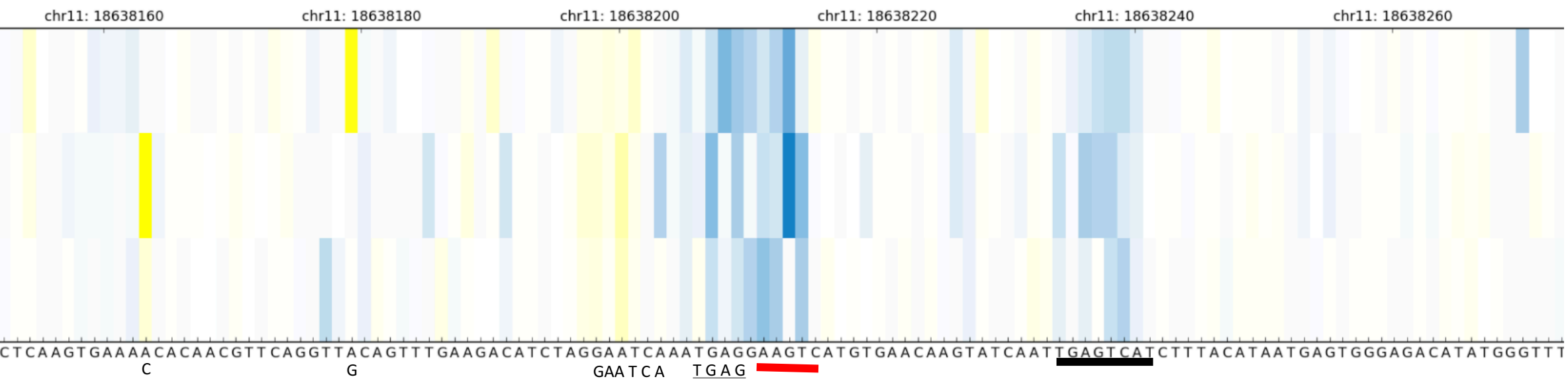


A&gt;G&gt;C&gt;T&gt;A

TGAGTCAT



Regain of motif here, except for potential  
obstruction by second guanine



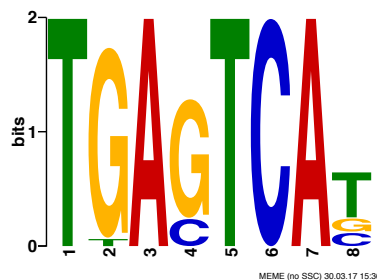
overrepresented  
mutations result in  
regained binding site

GM12878|DNase|None

Unclear what the conservation is here. May be  
obstruction of the end of the binding motif?

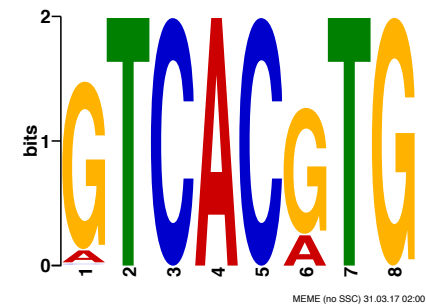
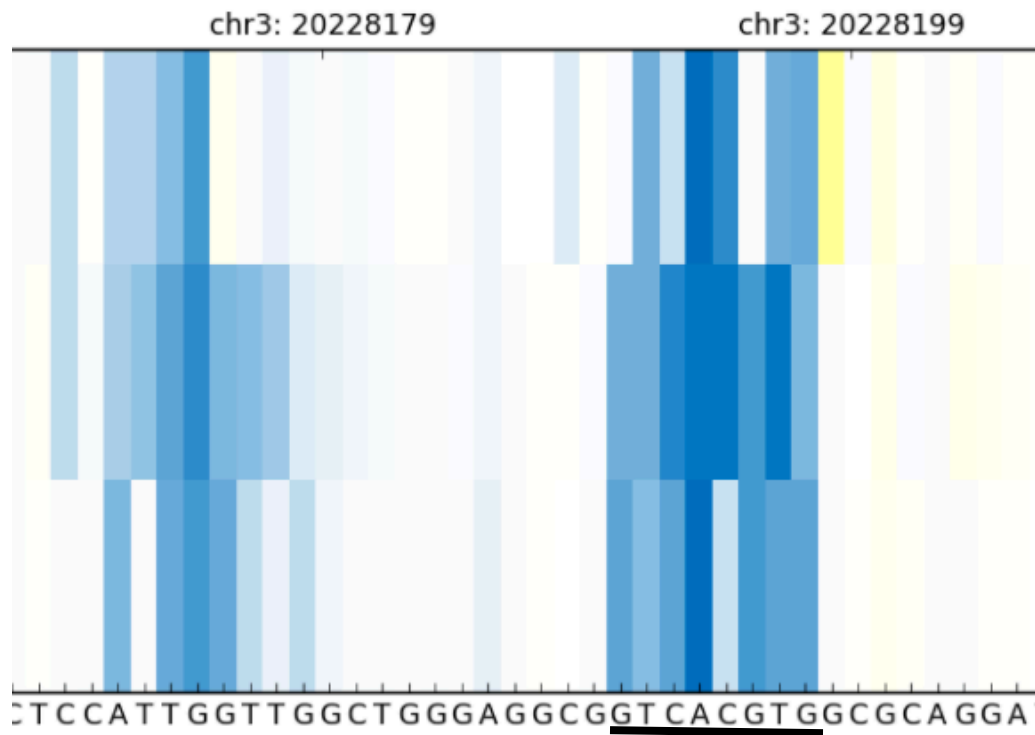
TGAGTCAT

A>G>C>T>A

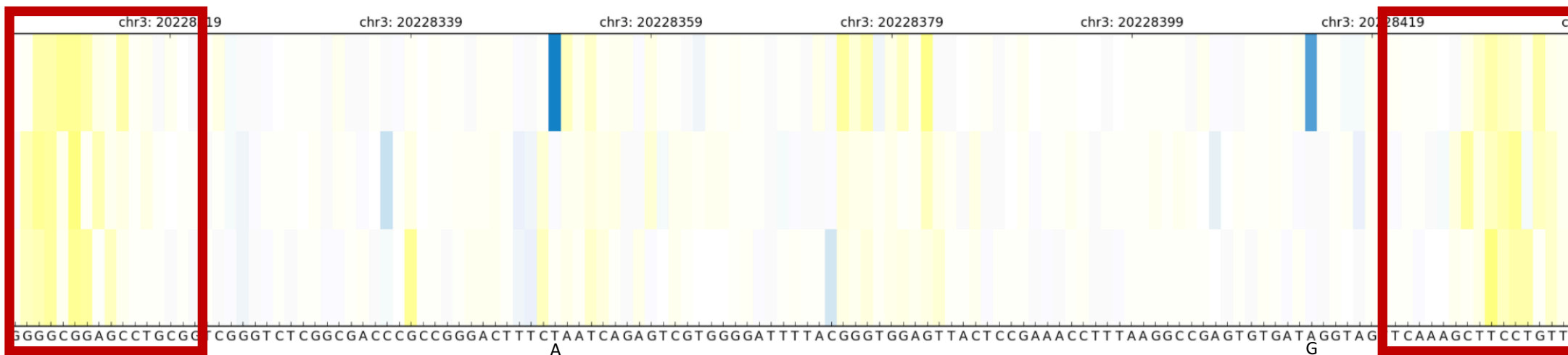


# ATF3

chr3:20227700-20227800	+	6	1.95e-05	AGGCG	GTCACGTG
GCGCAGGATG					
chr4:47465648-47465748	+	9	1.95e-05	GGGCTTGG	
GTCACGTG TCAAGGCTGG					
chr1:202896372-202896472	+	55	1.95e-05	CAGAGTCTGC	
GTCACGTG AGGAGGTGGT					
chr10:46090300-46090400	+	61	1.95e-05	TGACGCGGGC	
GTCACGTG TCCGGCCGAG					
chr1:78148570-78148670	+	28	1.95e-05	CAGCCTCCCT	
GTCACGTG GAAACATTTT					

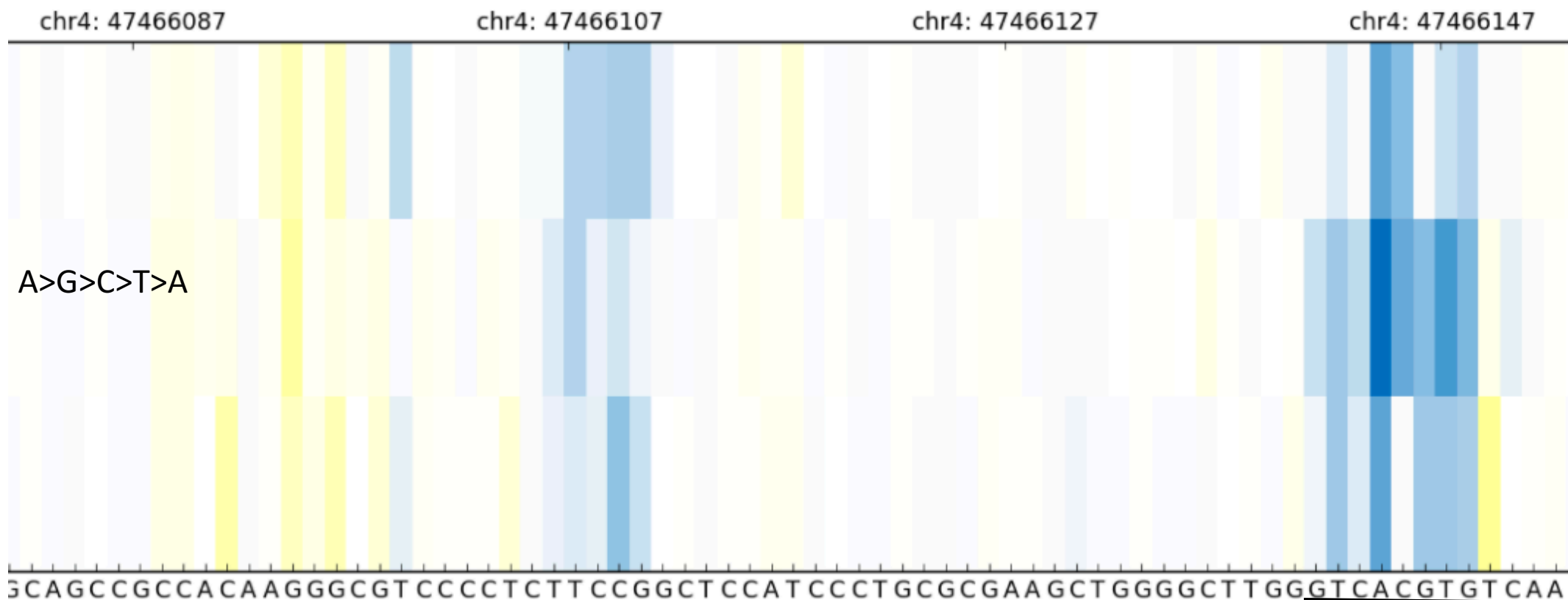


A>G>C>T>A  
 GTCACGTG  
 GTCACATG



Area of overrepresented mutations





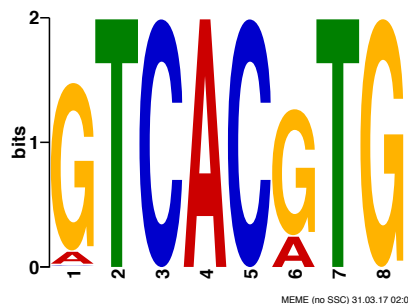
**CACAT**

overrepresented  
mutations

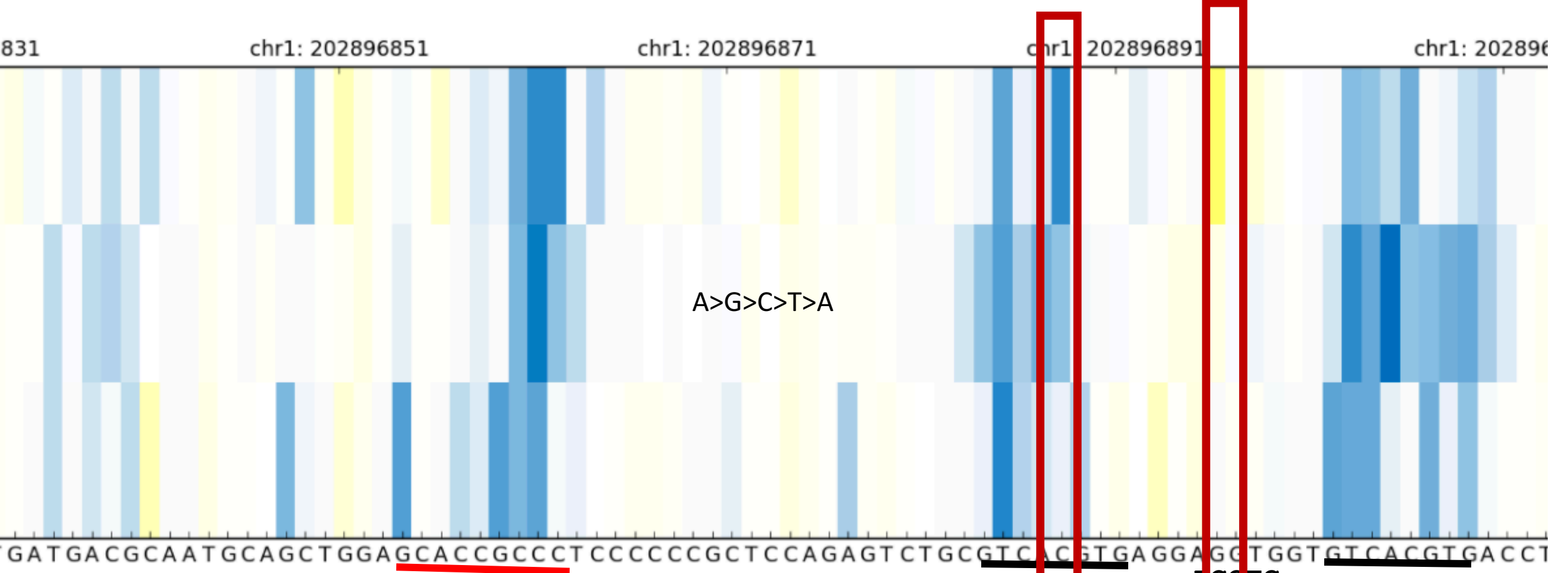
**TCCG**

Missing A4  
A conserved site forbidding binding?

underrepresented  
mutations

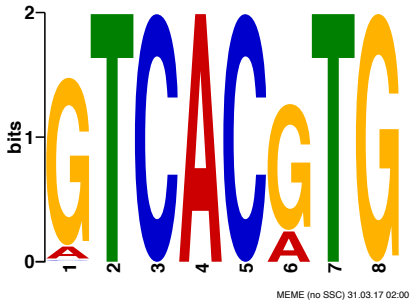


GTCACGTG  
GTCACATG



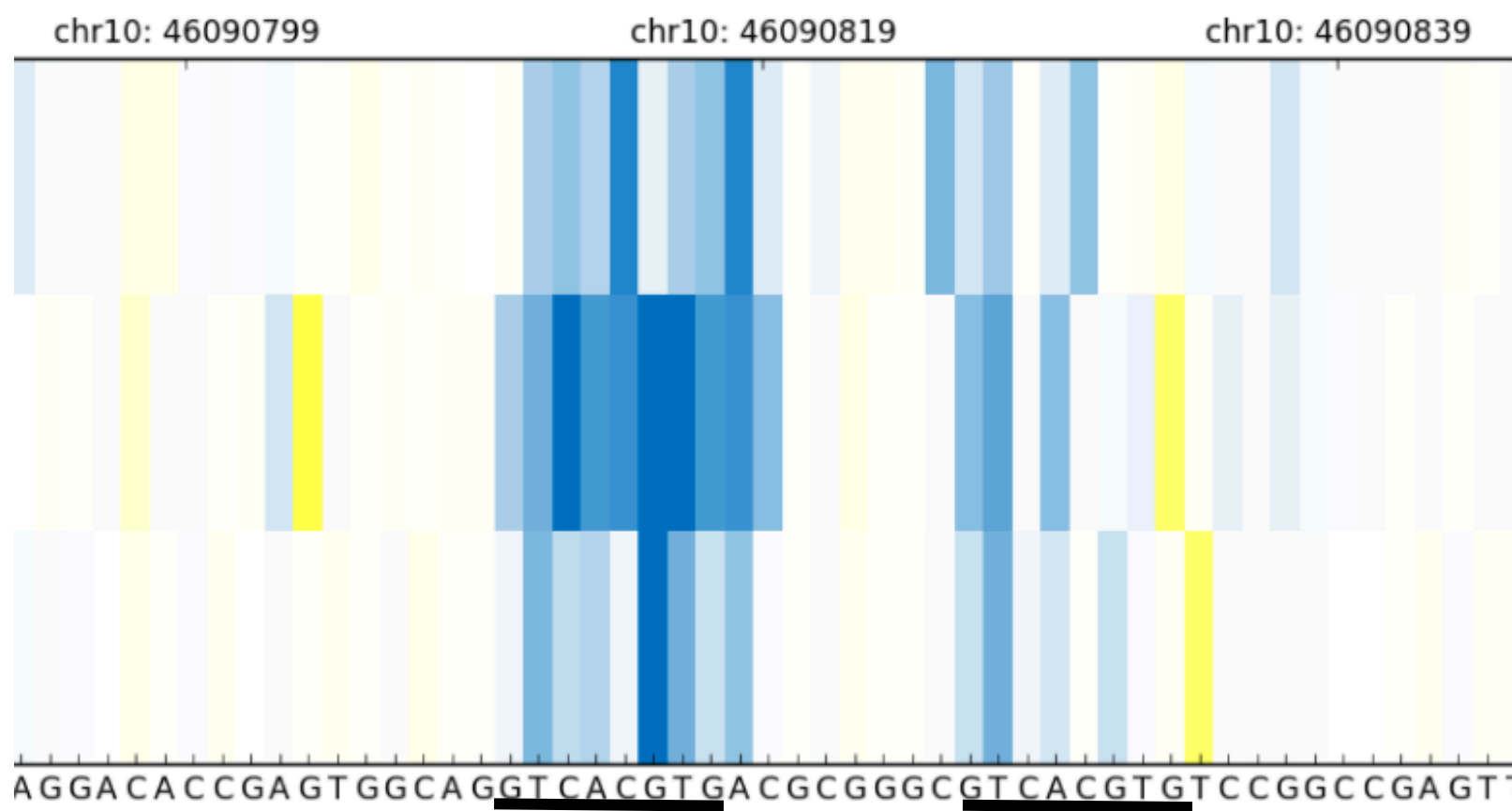
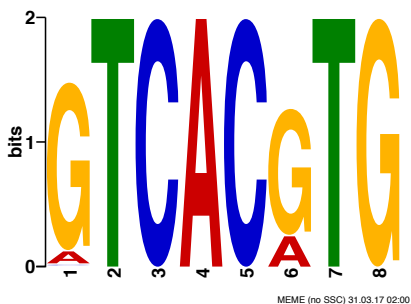
? → conserved site that is not a binding site

overrepresented mutation



May be that C5 is critical site

GTCACGTG  
GTCACATG



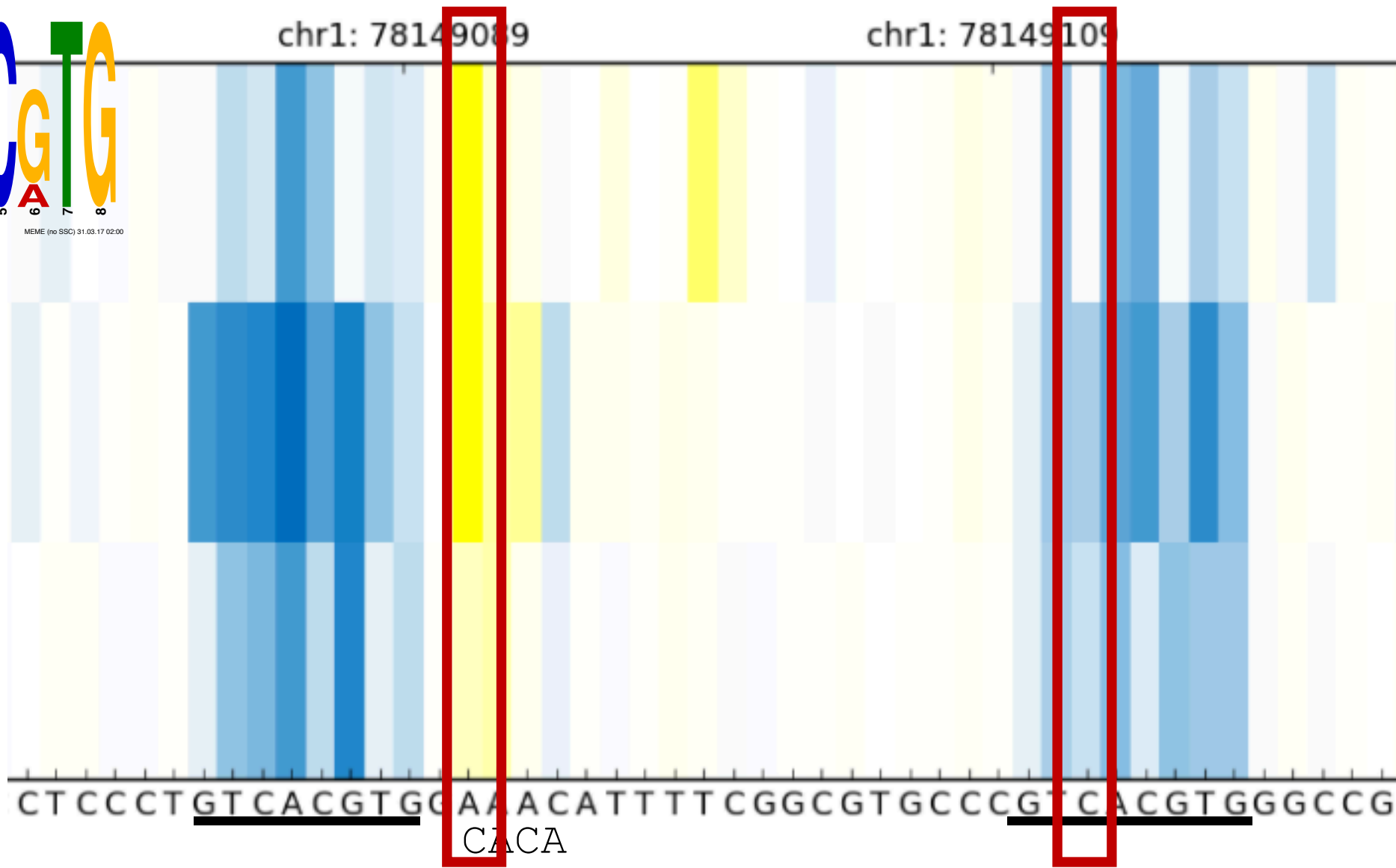
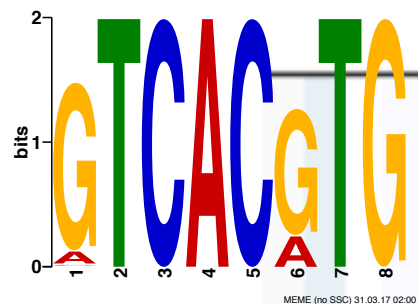
T

TCACG

overrepresented  
mutation

GTCACGTG  
GTC**CAC**ATG

A>G>C>T>A



CACA

overrepresented (partial)

May be that C3 is also a critical site (in addition to C5)

GTCACGTG

GTCACATG

A>G>C>T>A

# BATF

chr3:10262400-10262500	+	46	1.12e-06	AACCTACCCC	ATGACTCATC	AGCTGGCCTC
chrX:74287866-74287966	+	10	1.12e-06	TTTCTCTAA	ATGACTCATC	CTGACTAACA
chr9:37994680-37994780	+	39	2.02e-06	TCTGGTATAA	ATGACTCACC	AGTCAATCGG
chr1:12116570-12116670	+	57	2.02e-06	GAGAGTCTAA	ATGACTCACC	TCCAATCAC
chr17:53522513-53522576	+	47	2.02e-06	GGCCTGAGGG	ATGACTCACC	CTTTAAA

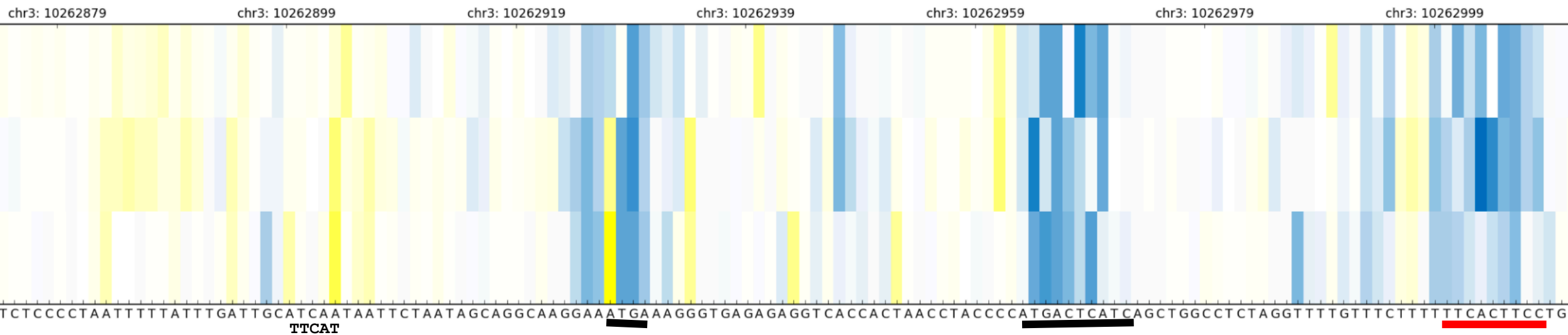
Multilevel  
consensus  
sequence

ATGACTCATC  
C CA  
GT

Multilevel  
consensus  
sequence

TTTCACTTCCTCTTT  
C T CTC  
GA

A>G>C>T>A



partial overrepresentation

A->C in the yellow region still  
means it is a binding motif

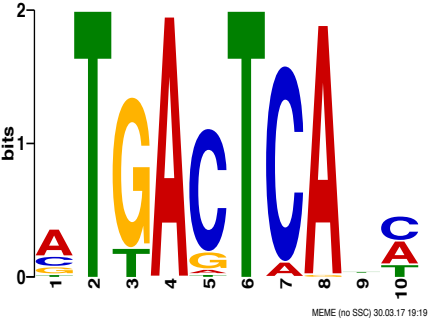
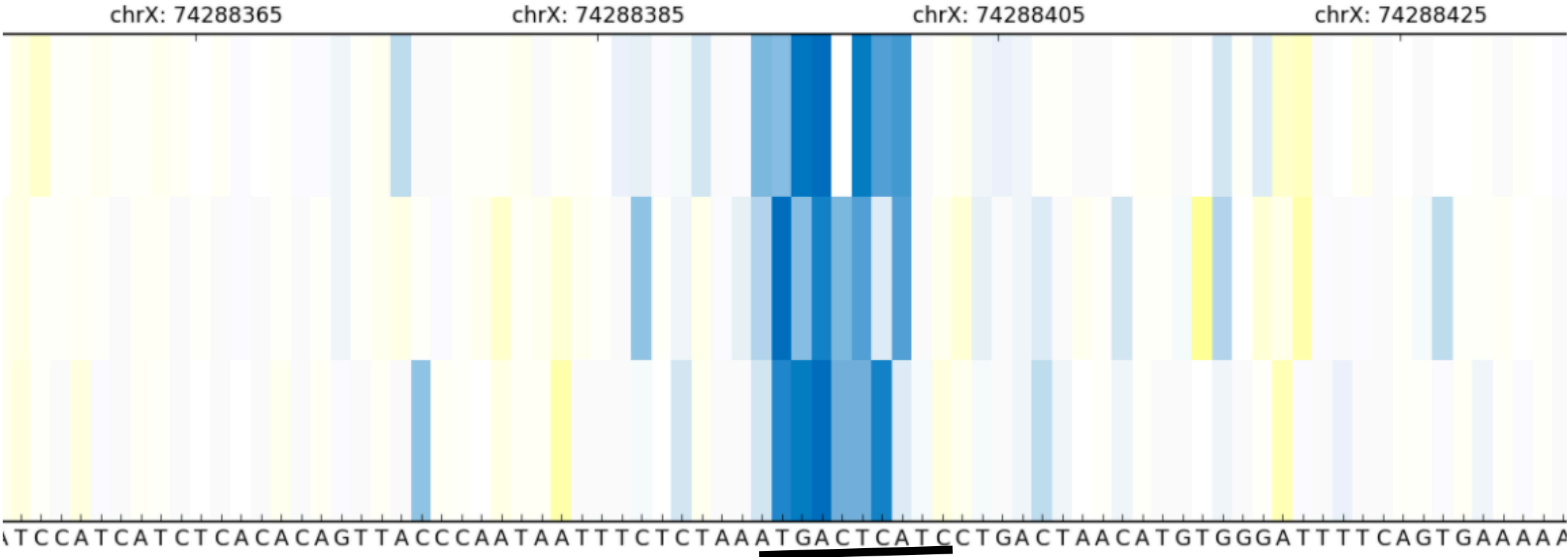
Secondary motif  
identified by  
MEME



Multilevel  
consensus  
sequence

ATGACTCATC  
C            CA  
             GT|

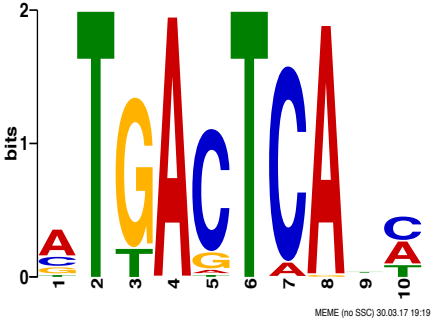
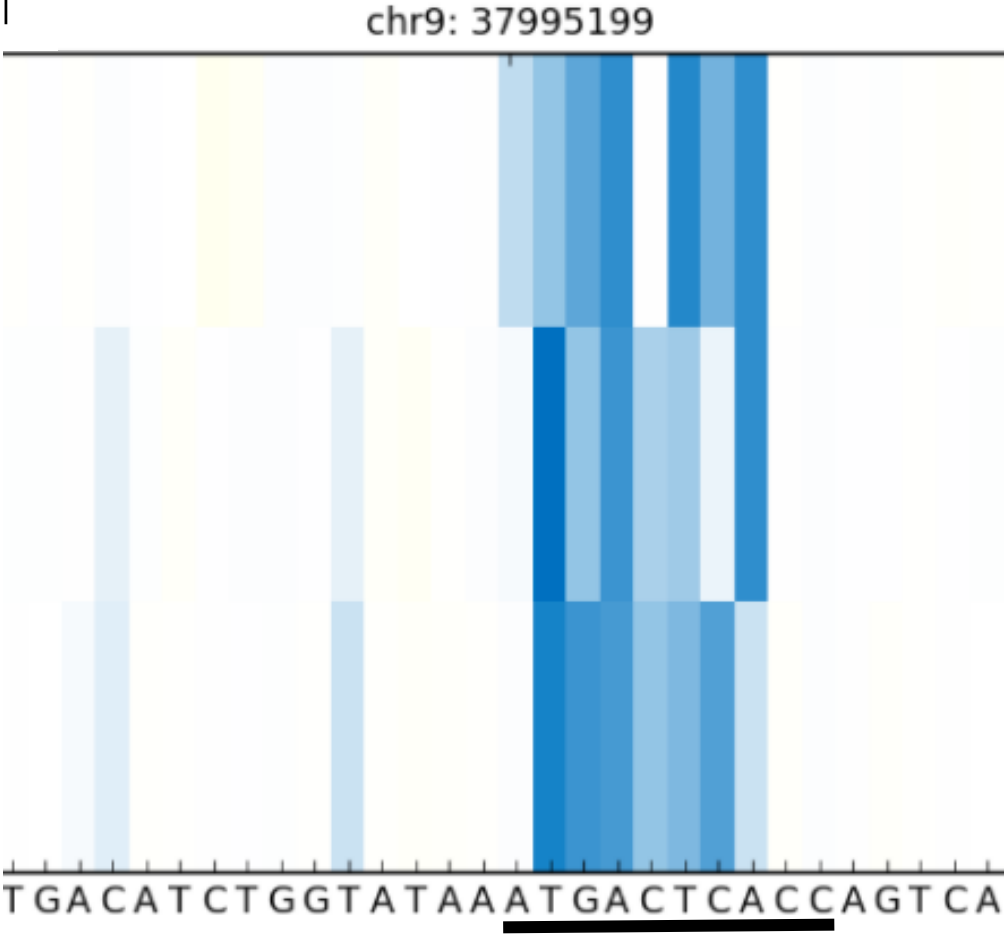
A>G>C>T>A



Multilevel  
consensus  
sequence

ATGACTCATC  
C            CA  
             GT|

A>G>C>T>A

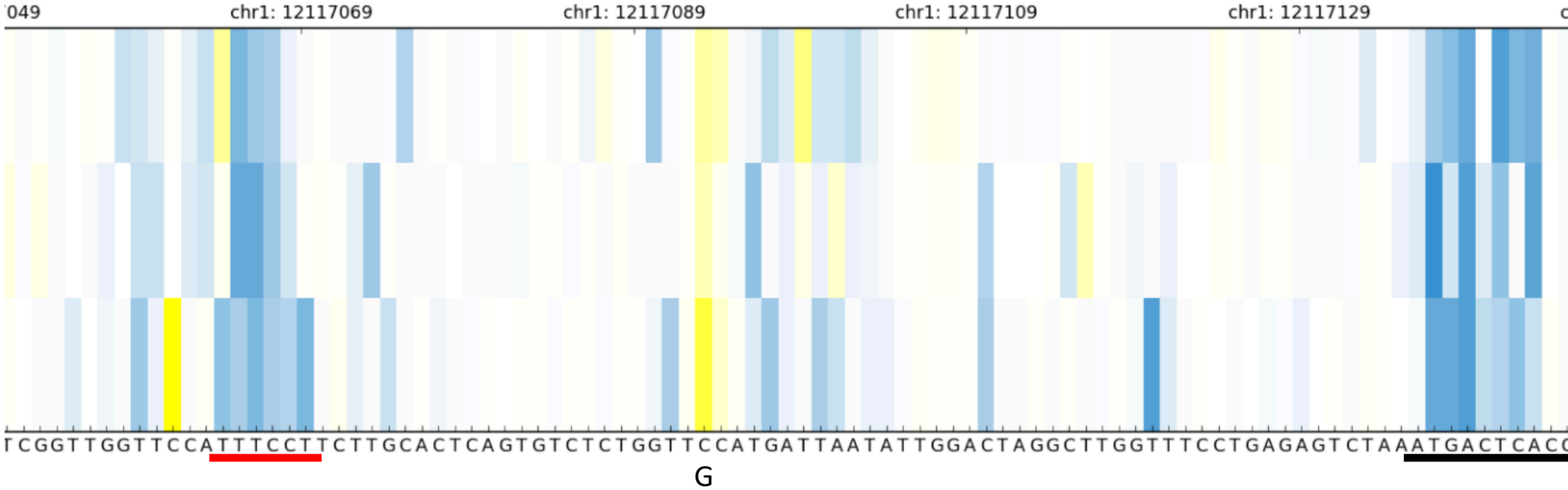




Multilevel  
consensus  
sequence

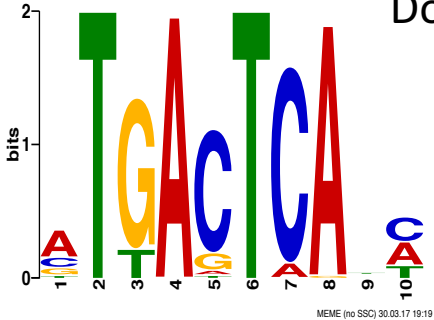
ATGACTCATC  
C            CA  
             GT|

A>G>C>T>A



?

Does not seem to match any of MEME motifs for BATF



Multilevel  
consensus  
sequence

ATGACTCATC  
C CA  
GT|

A>G>C>T>A

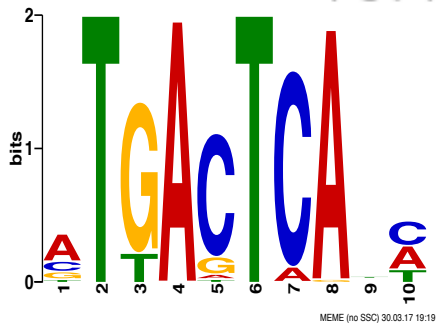
chr17: 53522952

chr17: 53522972

chr17: 53522992

c

T C T T G G T G G A A A T G A A A T T C A G G A A G T G A A G G T T G A G A C T T G A A C T T A G C A T G C C A A T C C T



?

Does not seem to match any of MEME motifs for BATF

# BCL11A

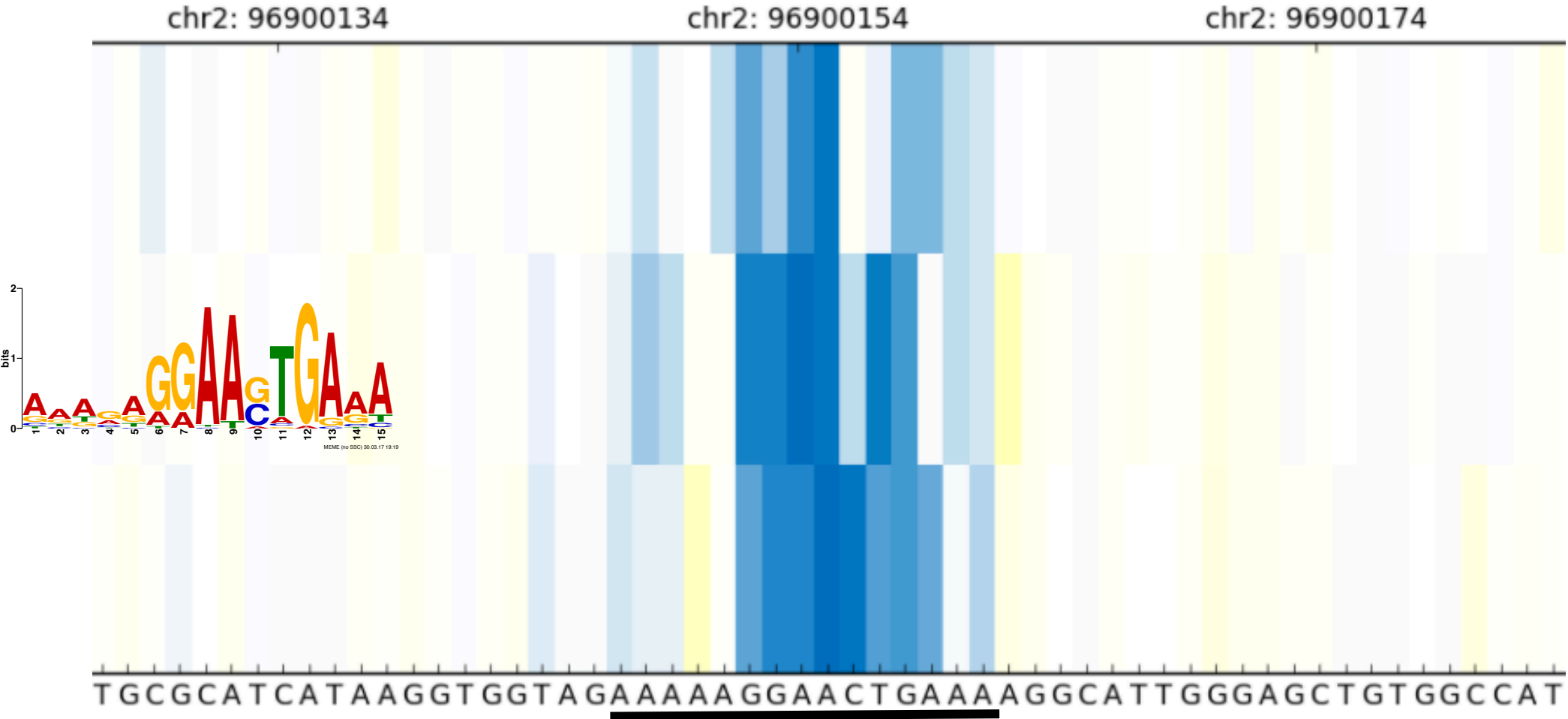
chr2:96899595-96899695	+	13	8.60e-09	AAGGTGGTAG	AAAAAGGAAGTGAAA
AGGCATTGGG					
chr4:111790937-111791037	+	67	3.78e-08	GTAGAGTCAC	AGAAAGGAAGTGAAA
CTGAAATCAA					
chr5:180232173-180232273	+	63	5.86e-08	CTGAAAAGAT	GAAAAGGAAGTGAAA
ACTGTTCCAG					
chr3:46580646-46580746	+	42	8.22e-08	ATTATTTTA	AAACAGGAAGTGAGA
AAGCAGACAC					
chr2:242314039-242314139	+	16	1.15e-07	GTGCTTATCT	GAGGAGGAAGTGAAA
GATAATTTC					

Does not seem to be much overrepresentation in any of these loci

A>G>C>T>A

Multilevel  
consensus  
sequence

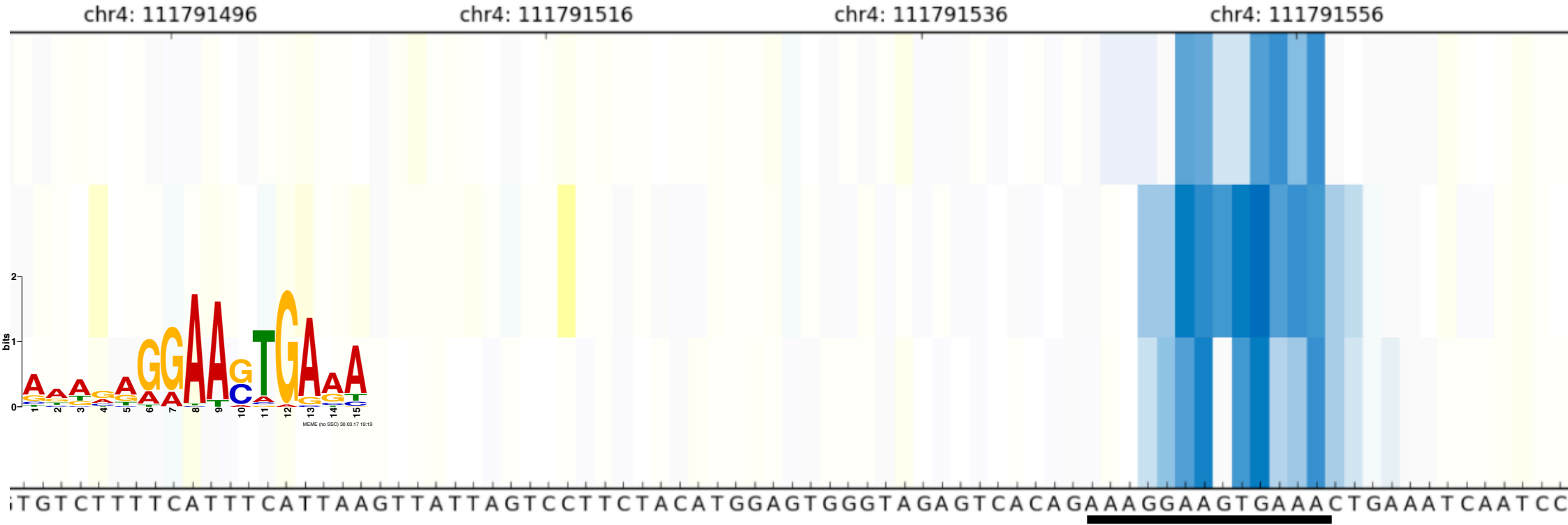
AAAGAGGAAGTGAAA  
AG C G



A>G>C>T>A

Multilevel  
consensus  
sequence

AAAGAGGAAGTGAAA  
AG C G

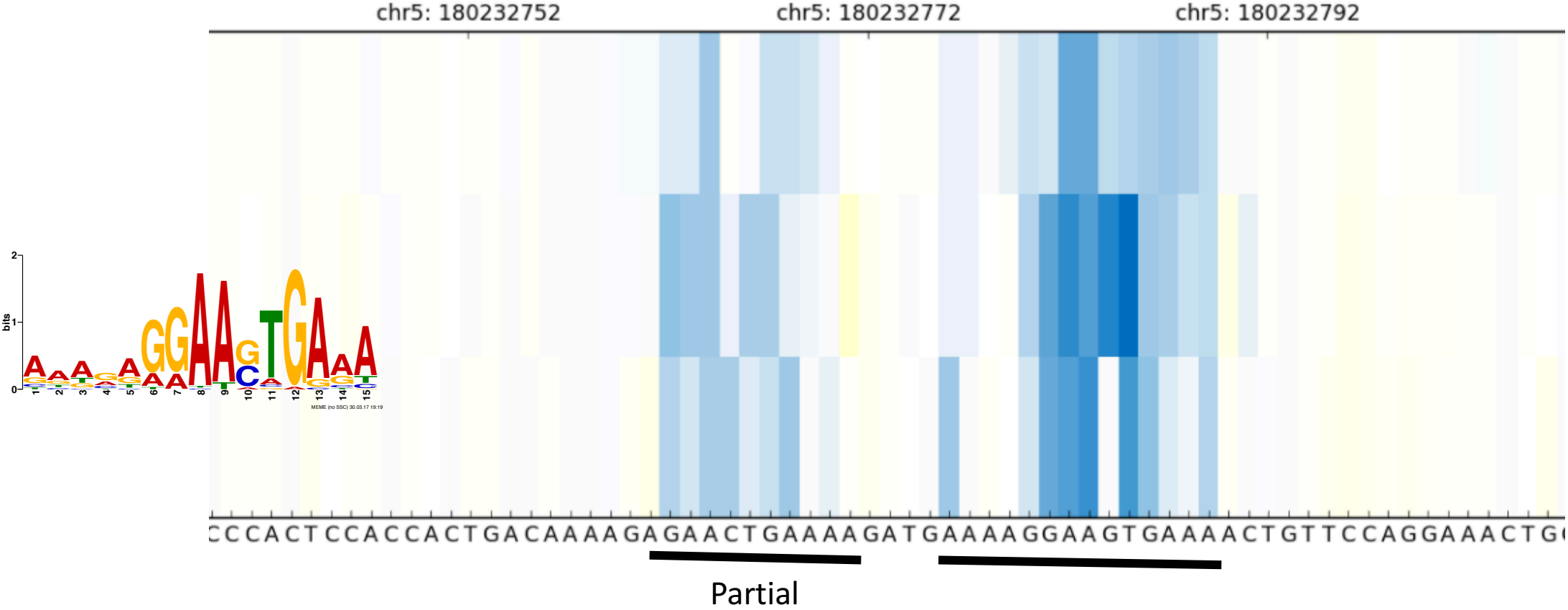


Partial – missing first 2 nucleotides

A>G>C>T>A

Multilevel  
consensus  
sequence

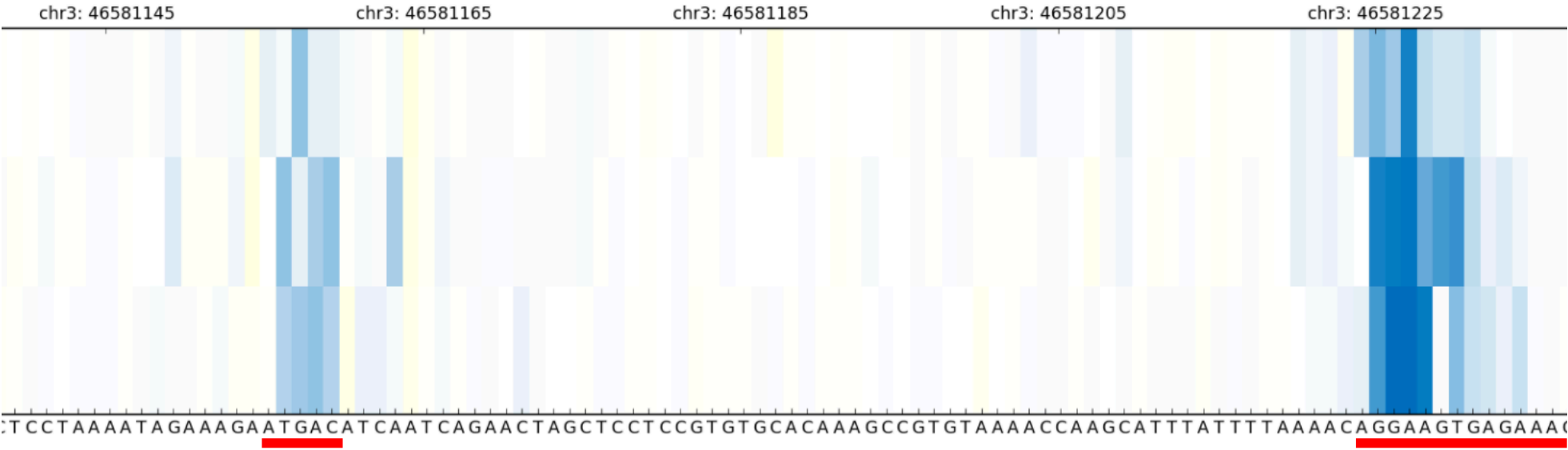
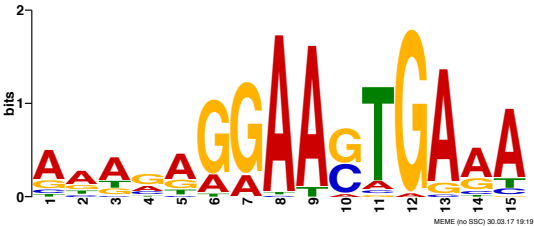
AAAGAGGAAGTGAAA  
AG C G



A>G>C>T>A

Multilevel  
consensus  
sequence

AAAGAGGAAGTGAAA  
AG C G



Just a secondary binding motif found by MEME

Multilevel  
consensus  
sequence

ATGACTCA  
G A

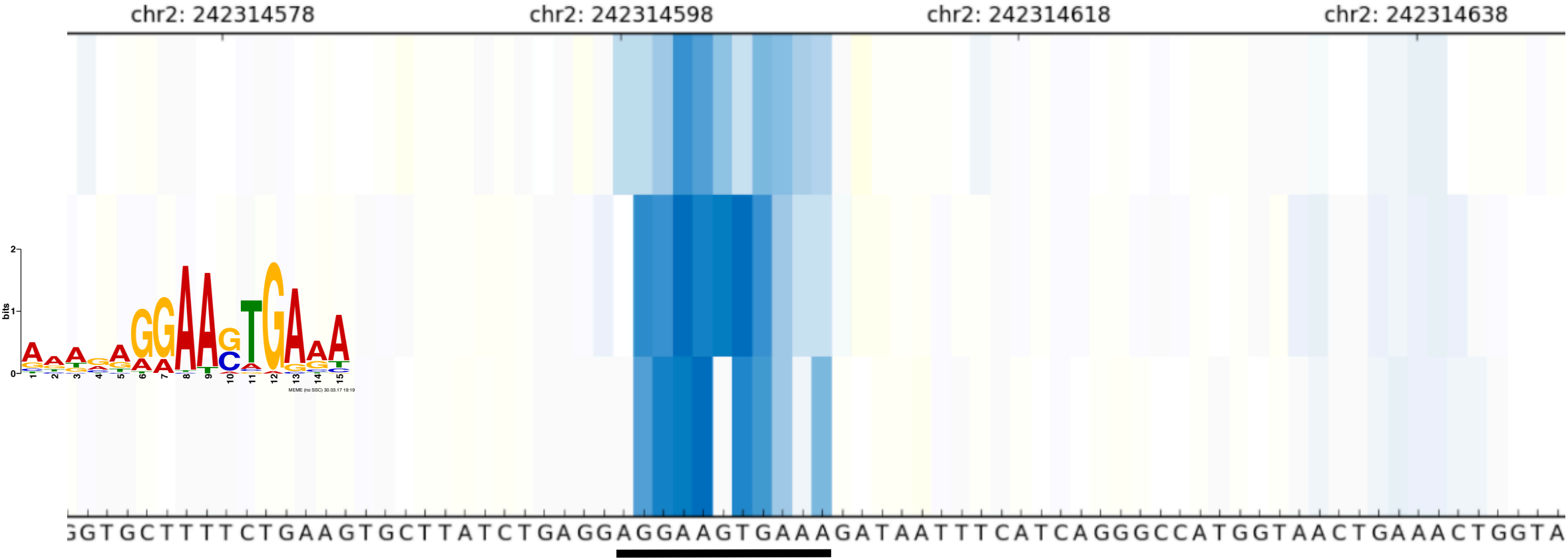
Obstruction by one G  
May be indicative of loss of binding (but conserved  
to potentially decrease gene expression nearby)

AGGAAGTGAGAAA

A>G>C>T>A

Multilevel  
consensus  
sequence

AAAGAGGAAGTGAAA  
AG C G





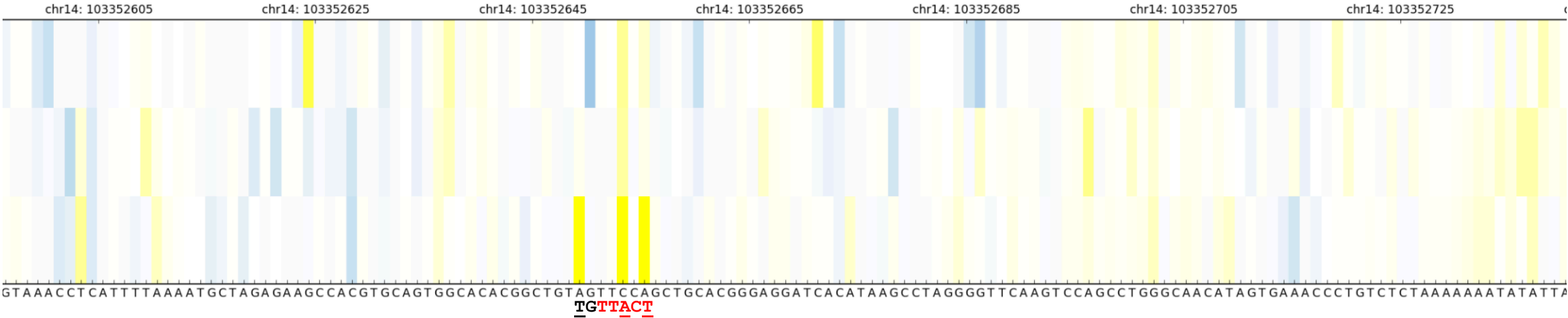
# BCL3

chr14:103352206-10335230	+	43	1.44e-05	CGTGACAATA ATGACTCA GCCATACGTA
chr1:28471014-28471114	+	52	1.44e-05	TCTTGGGTCT ATGACTCA TCTCAACGAG
chr7:130523512-130523612	+	78	1.44e-05	TTTCTCAGAA ATGACTCA AGGCACGATC
chr17:65418034-65418134	+	33	1.44e-05	TGCAGGAAGT ATGACTCA ATTGTGACTG
chr19:14696120-14696220	+	23	1.44e-05	GTCTACATAT ATGACTCA TATGACTCTC

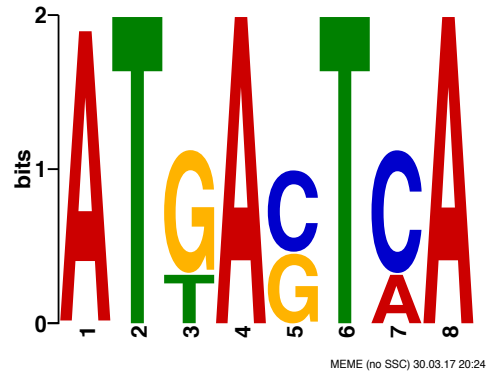
Multilevel  
consensus  
sequence|

ATGACTCA  
T G A

A>G>C>T>A



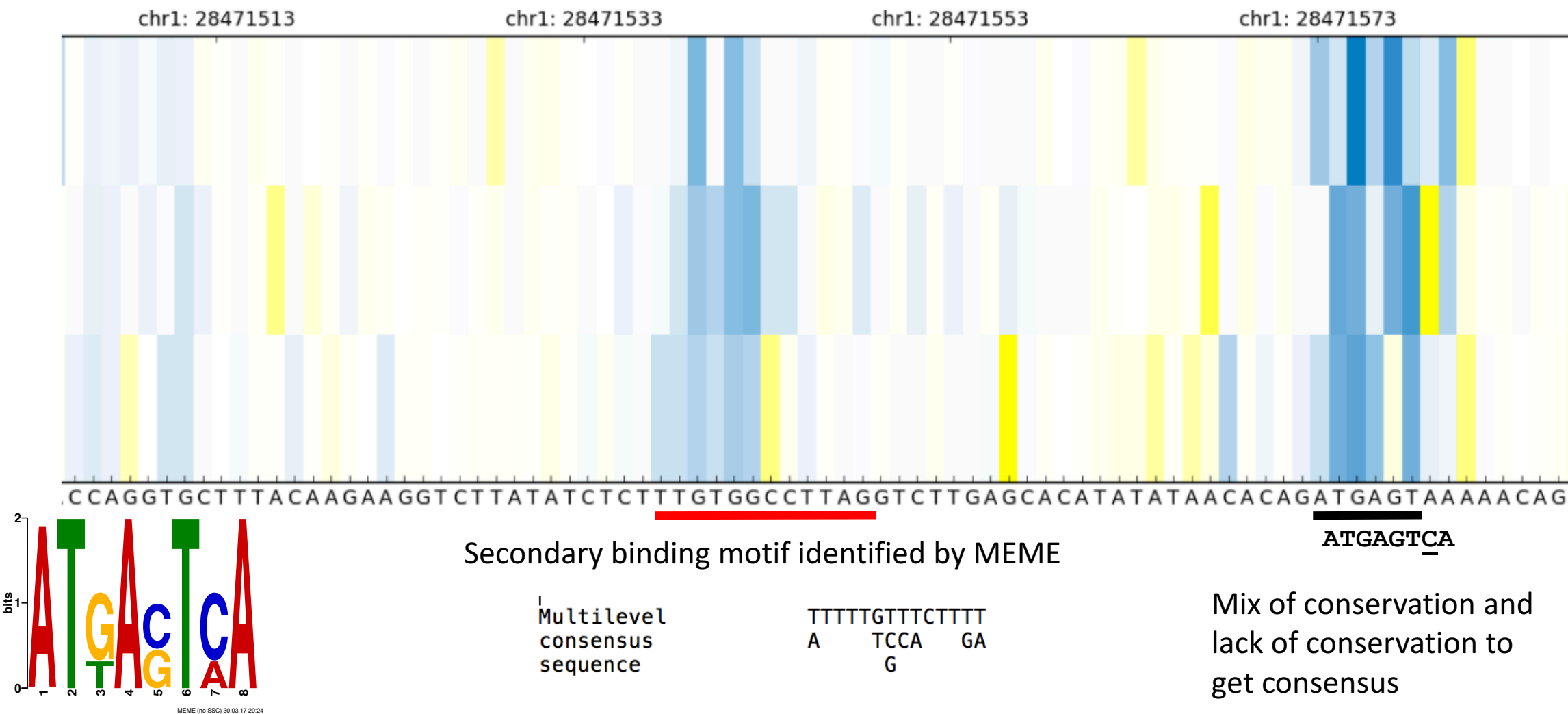
Partial overexpression



Multilevel  
consensus  
sequence|

ATGACTCA  
T G A

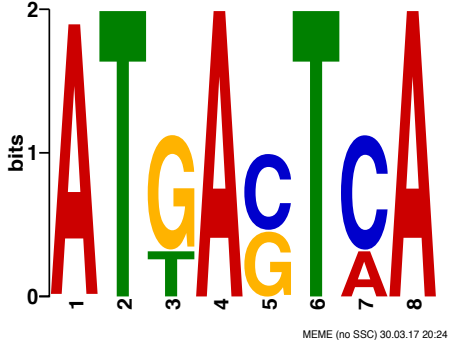
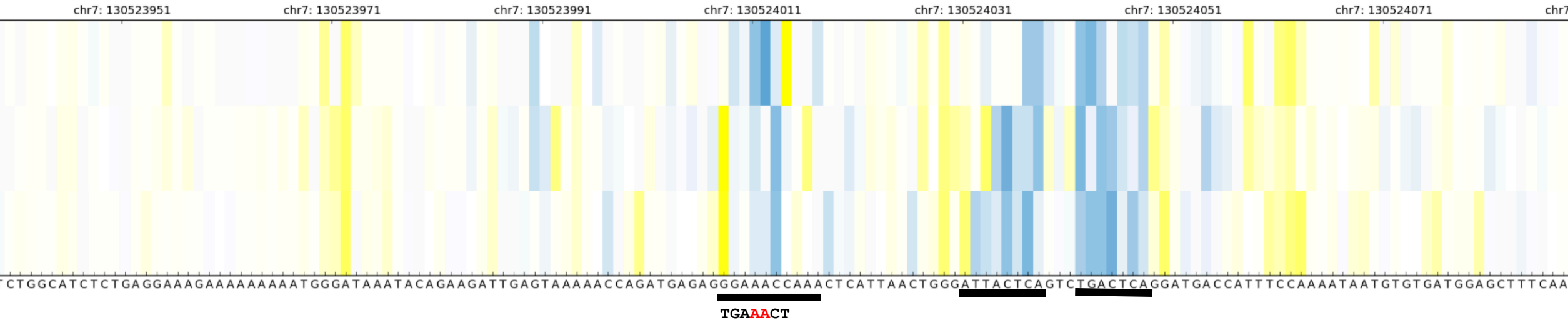
A>G>C>T>A



Multilevel  
consensus  
sequence|

ATGACTCA  
T G A

A>G>C>T>A

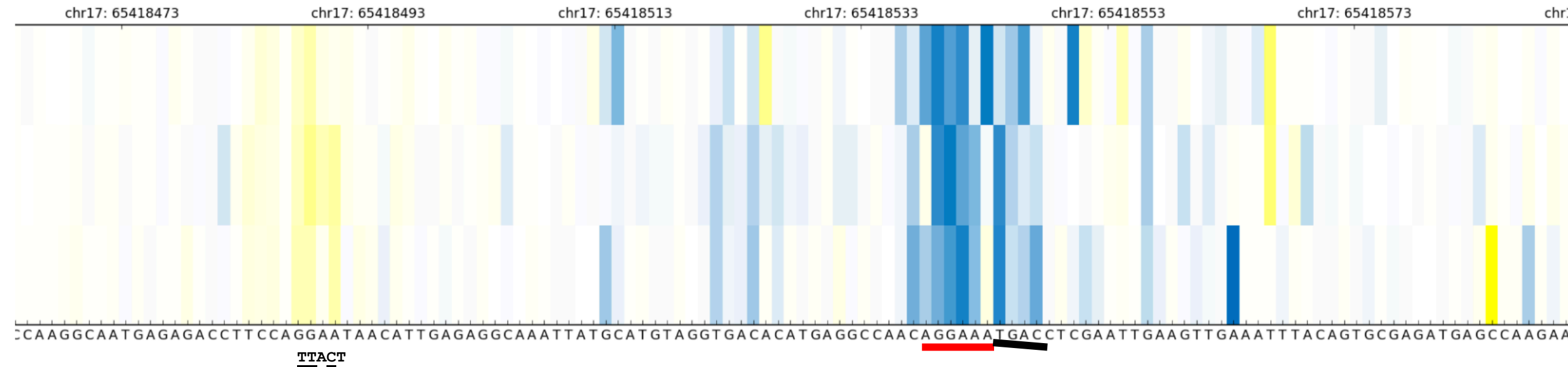
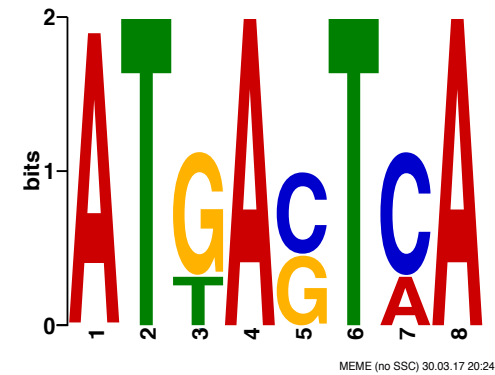


Two extra A  
A partially conserved/not  
conserved site forbidding  
binding?

Multilevel  
consensus  
sequence|

ATGACTCA  
T G A

A>G>C>T>A



Overexpression of mutation  
to get revival of binding  
motif

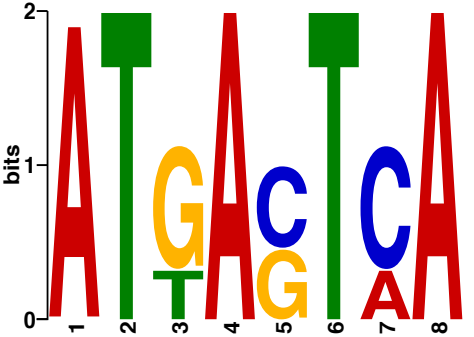
RED IS SAME SEQUENCE AS PREVIOUS SLIDE

A partially conserved/not conserved site  
forbidding binding?

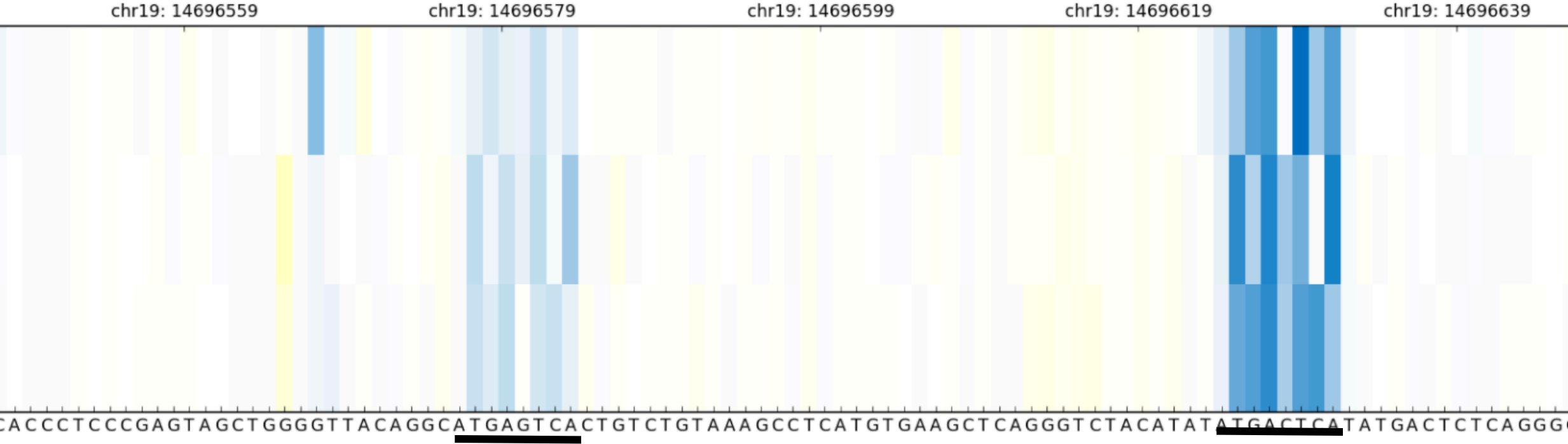
Multilevel  
consensus  
sequence|

ATGACTCA  
T G A

A>G>C>T>A



MEME (no\_SSC) 30.03.17 20:24

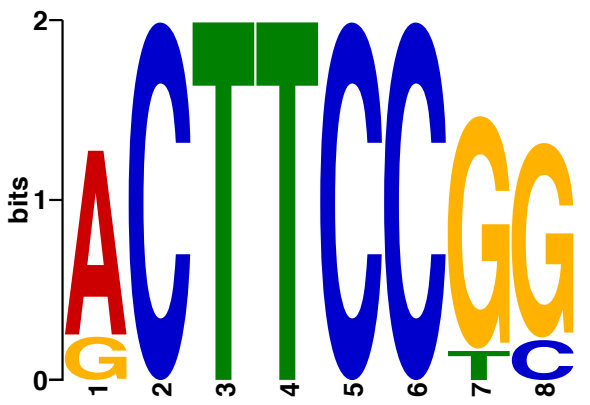
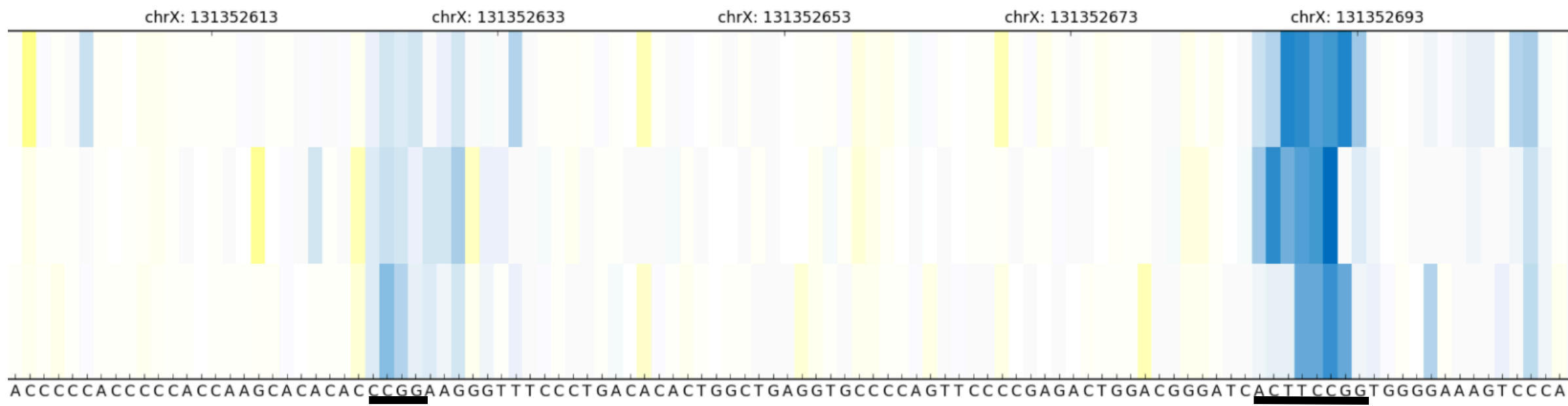


# BCLAF1

chrX:131352134-131352234	+	61	1.85e-05	GGACGGGATC	ACTTCCGG	TGGGGAAAGT
chr11:31391223-31391323	+	71	1.85e-05	CGGAACAGTG	ACTTCCGG	CGGAAGAAGA
chr11:60928700-60928800	+	58	1.85e-05	CCGCGCCCAC	ACTTCCGG	GGTCCGCCAG
chr7:43768917-43769017	+	60	1.85e-05	AGCAGTAGAG	ACTTCCGG	CGTCGGCTGC
chr18:33552485-33552585	+	87	1.85e-05	TGGGAAGCCG	ACTTCCGG	CGCCCT

A>G>C>T>A

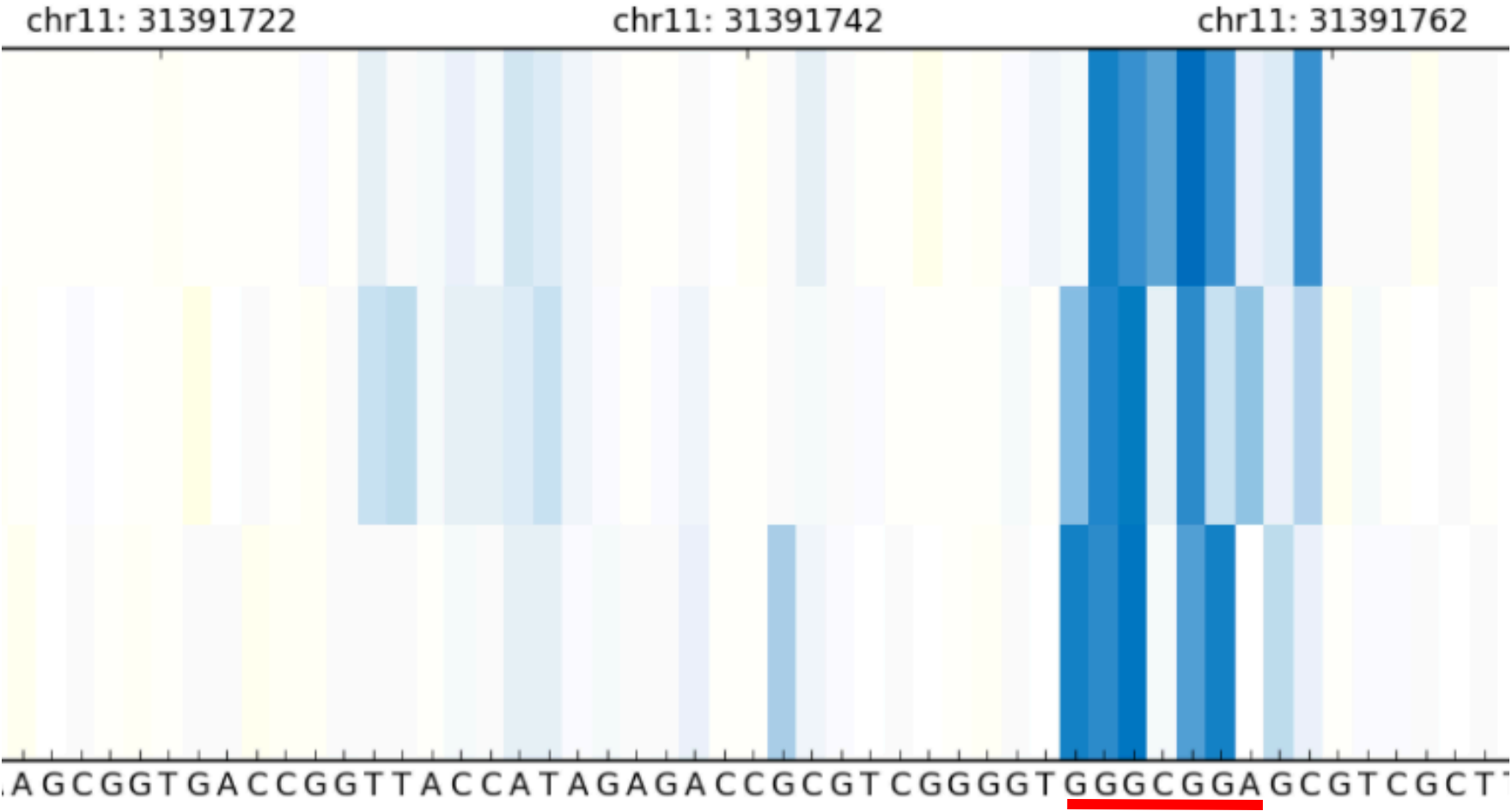
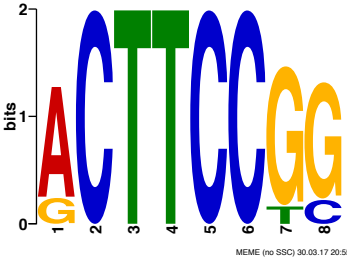
ACTTCCGG





A>G>C>T>A

ACTTCCGG



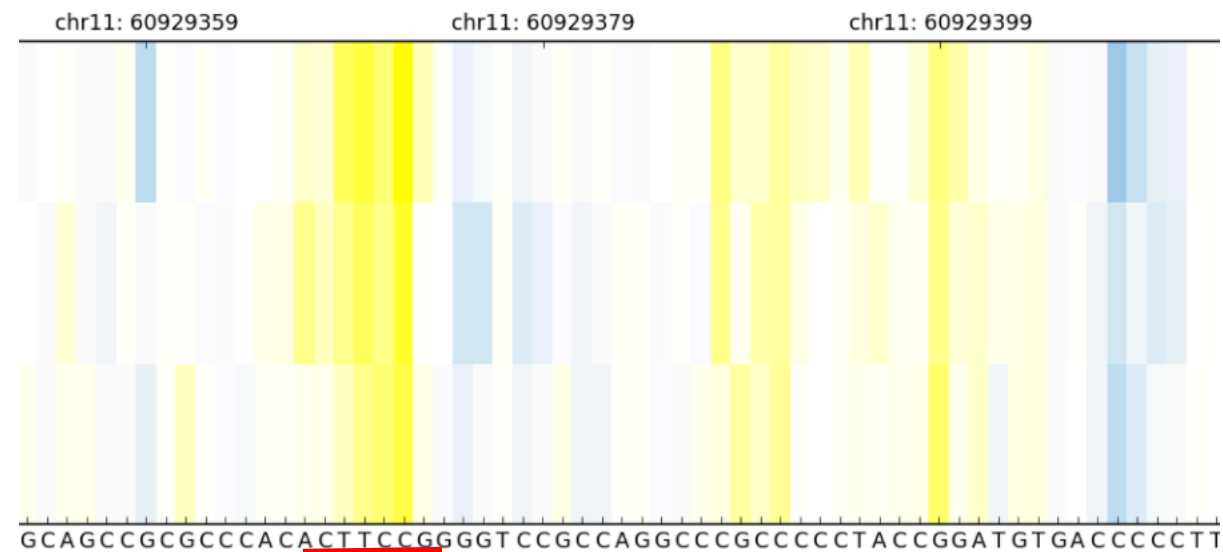
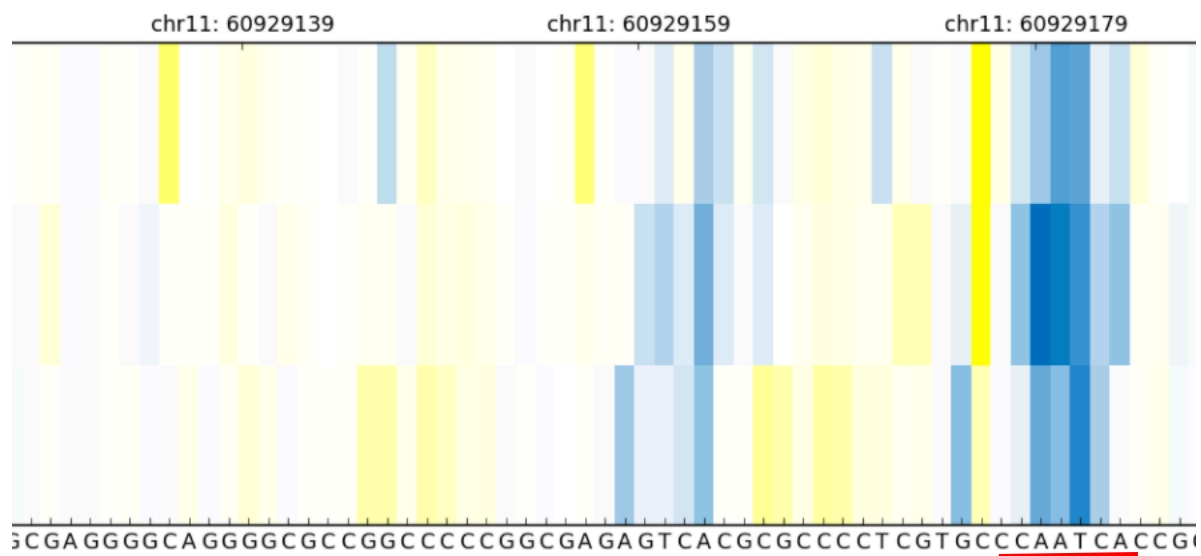
Multilevel  
consensus  
sequence

GGGCGGGGCGGAGCCGACGCGGGGGCCGG  
CCCACC CCG GCGGG GCCC GA  
A C AT G  
T

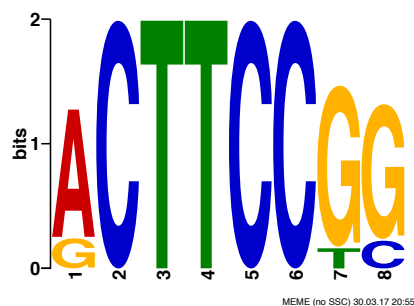
Likely this secondary motif found by MEME

A>G>C>T>A

ACTTCCGG

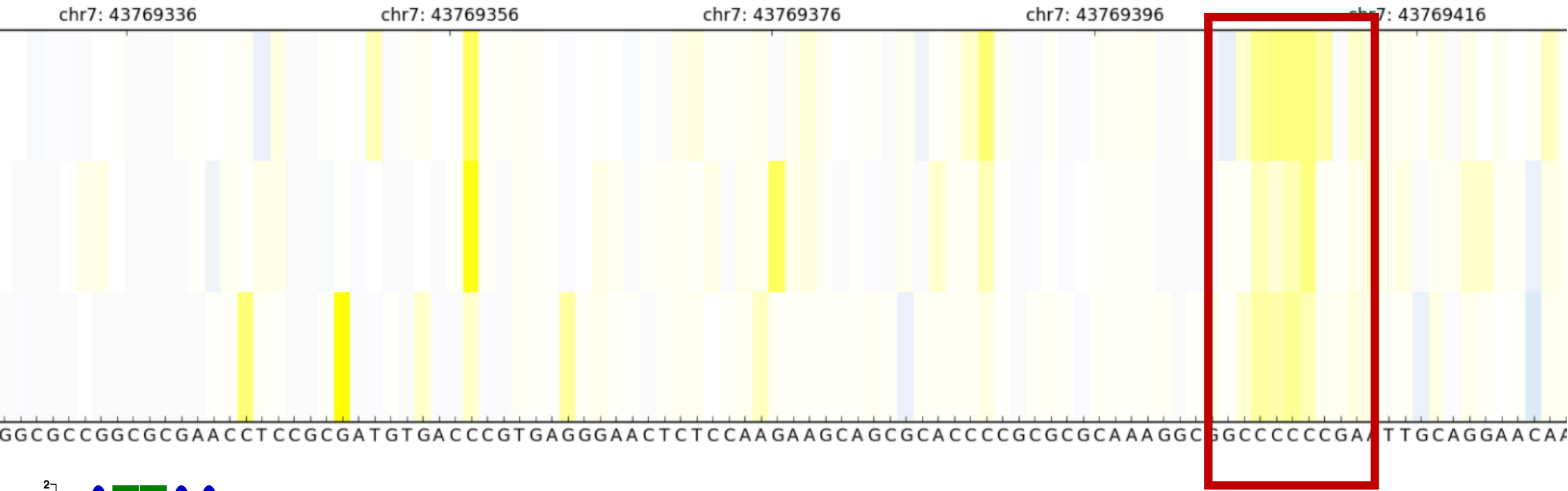


Found exact binding motif that has overexpressed mutations. Clear forbidding of binding!



A>G>C>T>A

ACTTCCGG



Area of overrepresented mutations

