

# Marginal Footprinting (2wk vs 8wk)

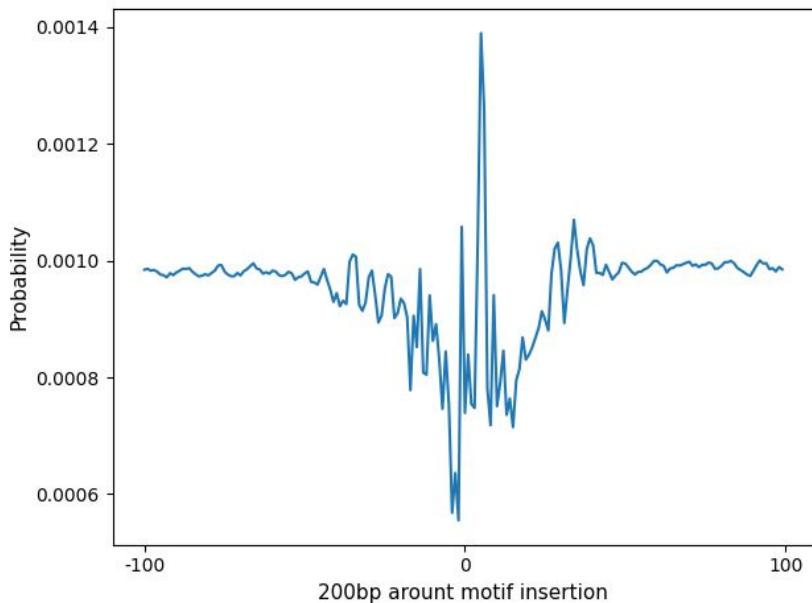
# TFs we examined:

From motif\_to\_pwm.TF tsv file below:

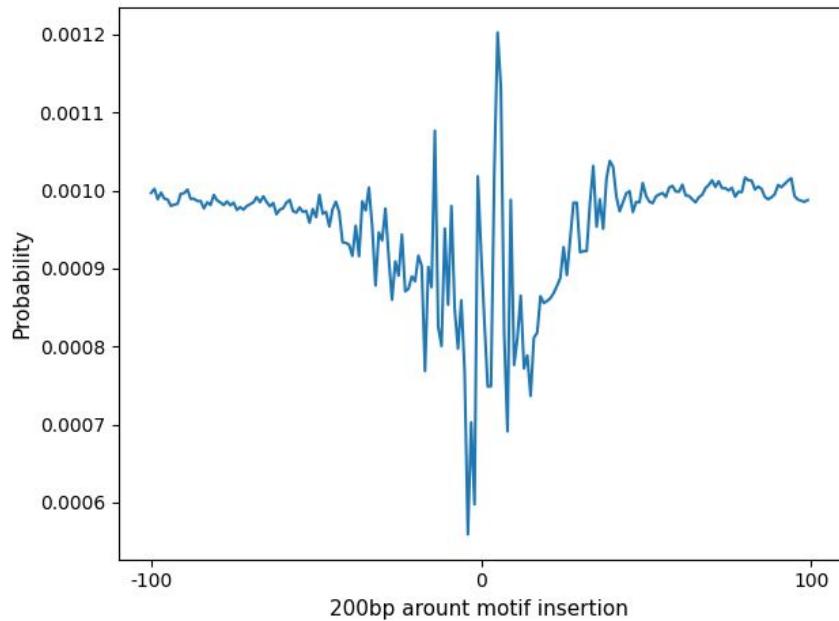
NRF1	GCGCATGCGC
AP1	CGATATGACTCATCCC
CTCF	TTGGCCACTAGGGGGCGCTAT
ETS	CCGAAAGCGGAAGTGAGAC
SP1	AAGGGGGCGGGGCCTAA
RUNX	CCCTAACACAGCCC
NFKB	GCAAGGGAAATTCCCCAGG
GATA+TAL	GGCTGGGGGGGGCAGATAAGGCC
TAL	GGCTGGG
NFYB	CCAGCCAATCAGAGC
GABPA	GAAACCGGAAGTGGCC
BACH1+MAFK	AACTGCTGAGTCATCCCG
NRF1	CCCCGCGCATGCGCAGTGC
HNF4G	CCGTTGGACTTGGACCCTG

# AP1 footprint - similar.

2 week

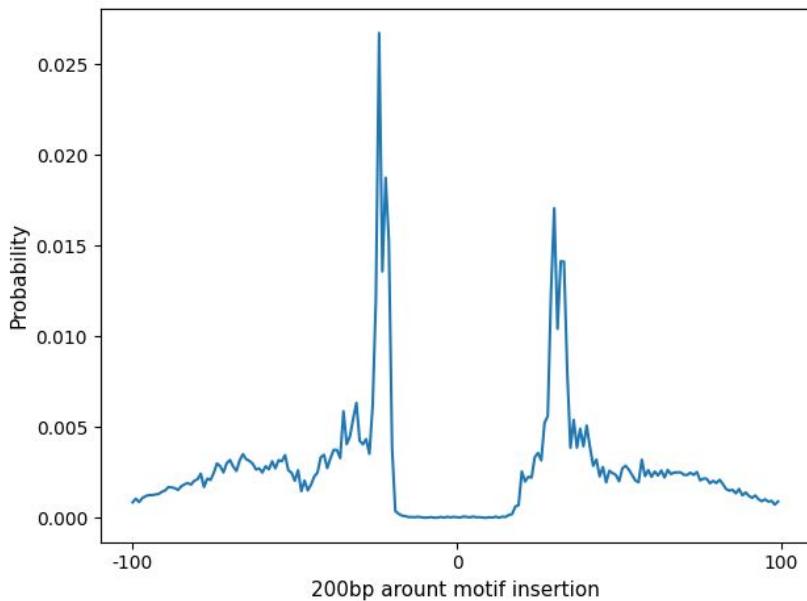


8 week

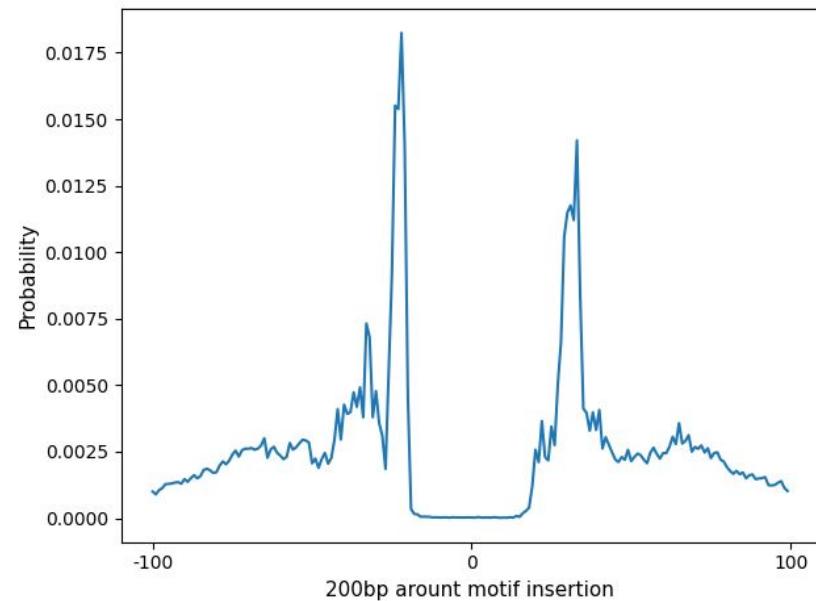


# CTCF footprint - “largest” footprints in terms of probability

a) 2 Week

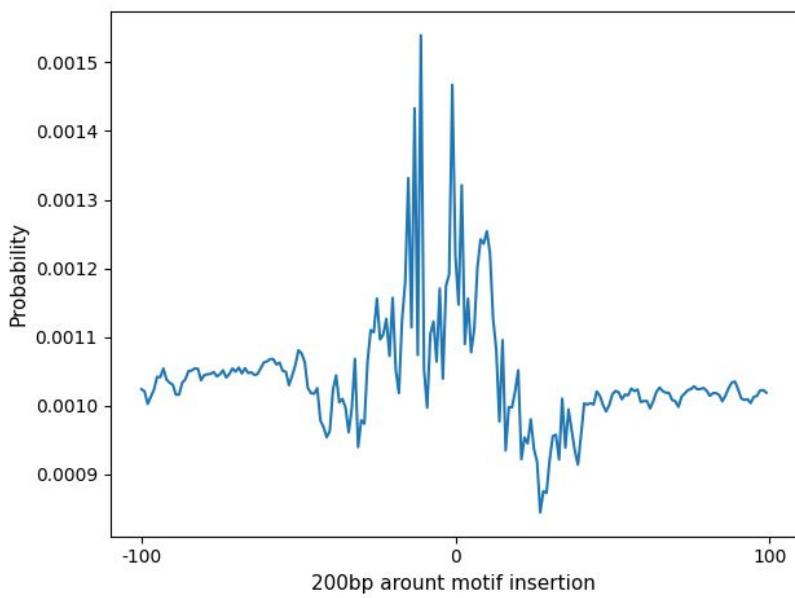


b) 8 Week

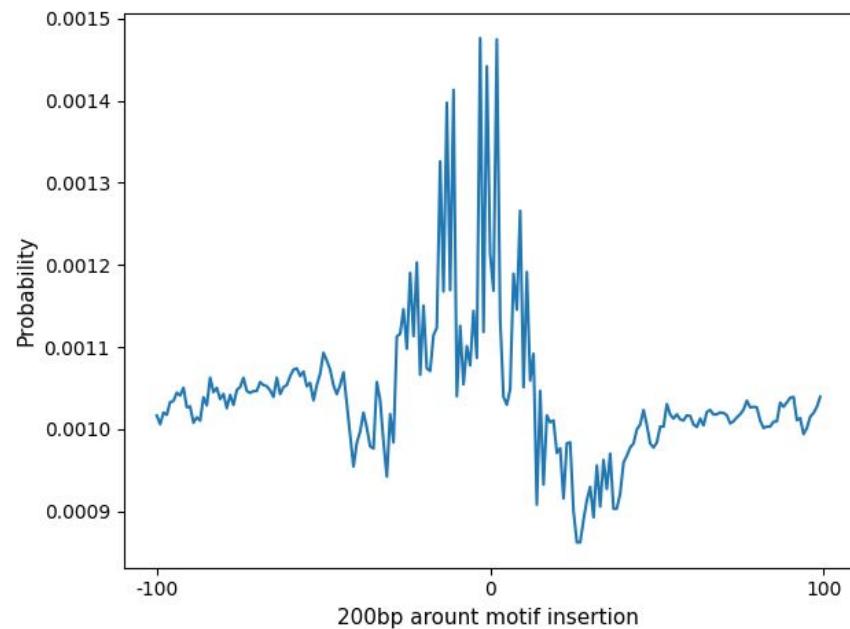


# ETS footprint - highly similar.

2 week

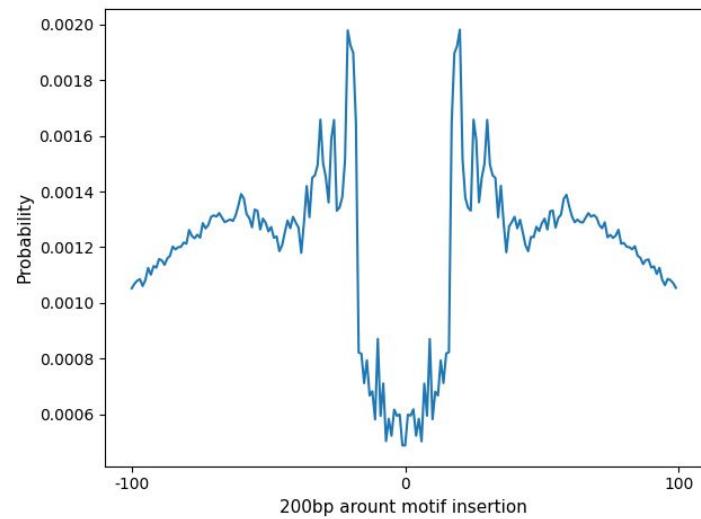


8 week

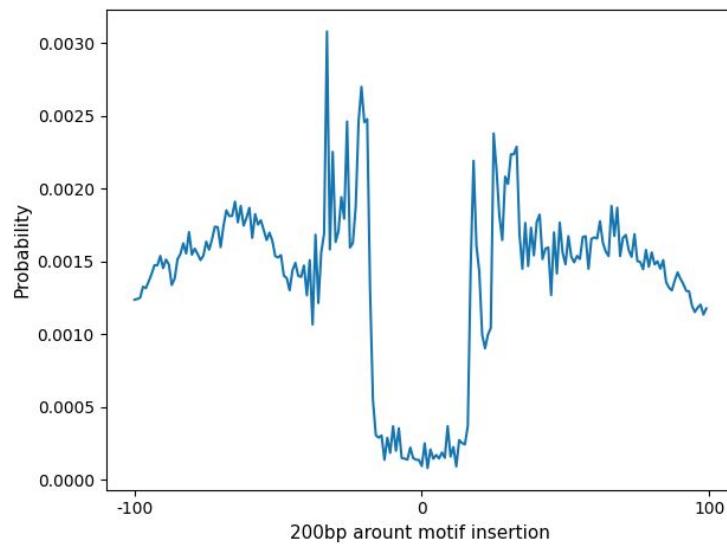


# NRF1 footprint - 0.001 difference here.

2 week

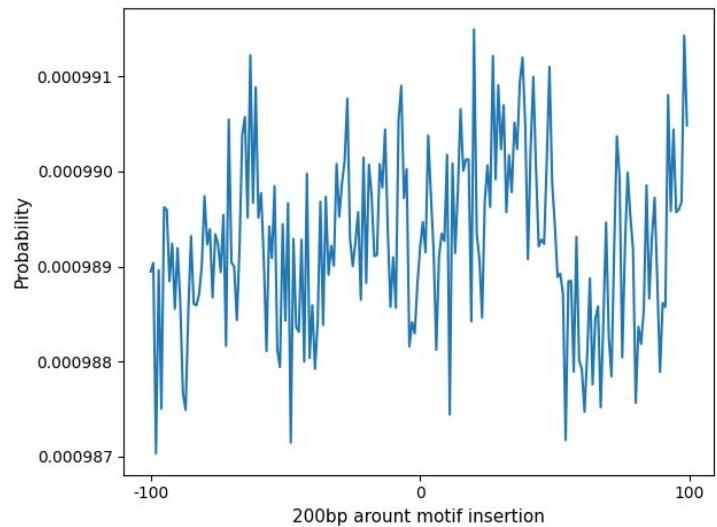


8 week

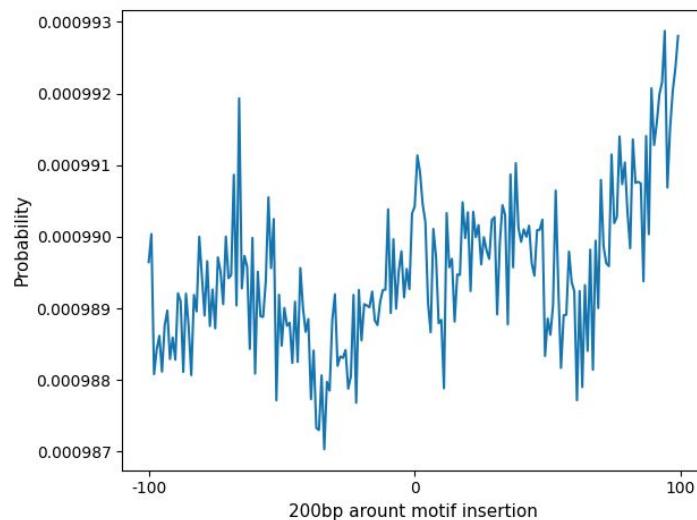


# Control footprint

2 week

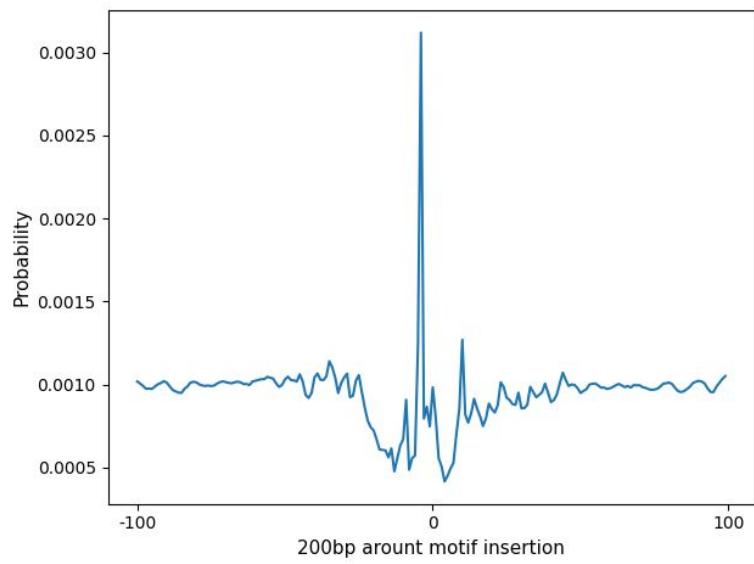


8 week

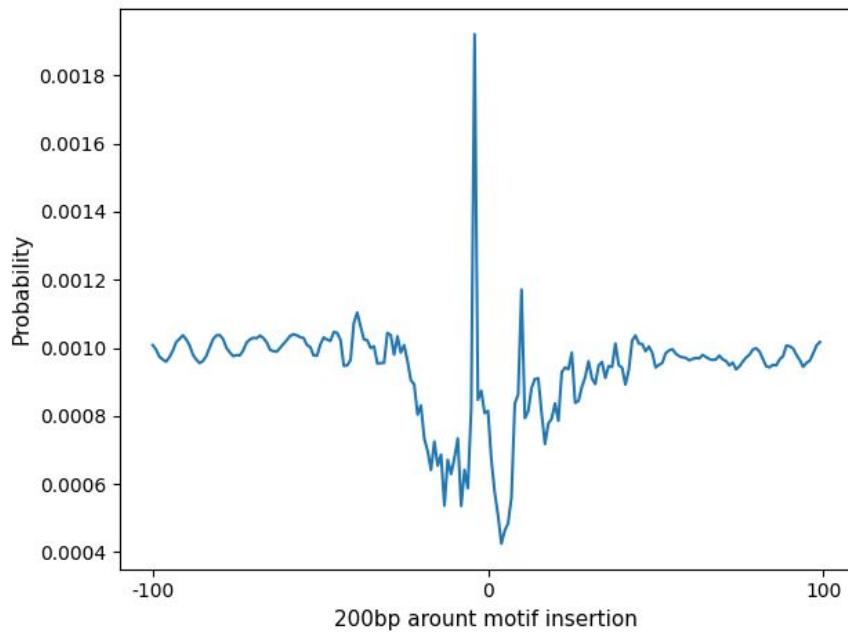


# GATA+TAL - 0.001 difference here.

2 Week

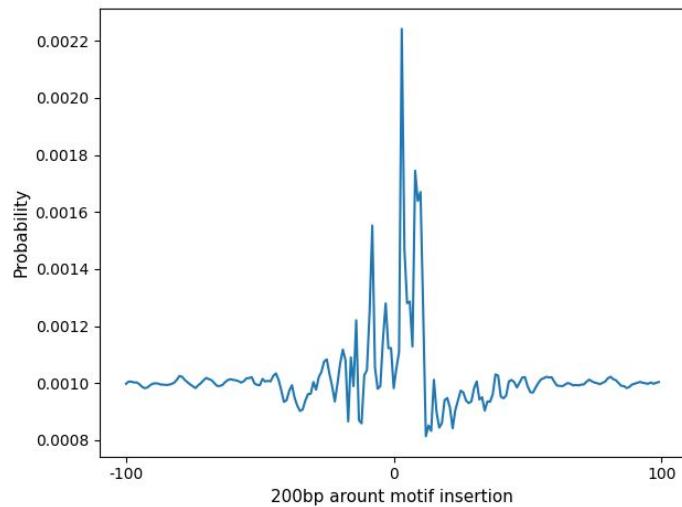


8 week

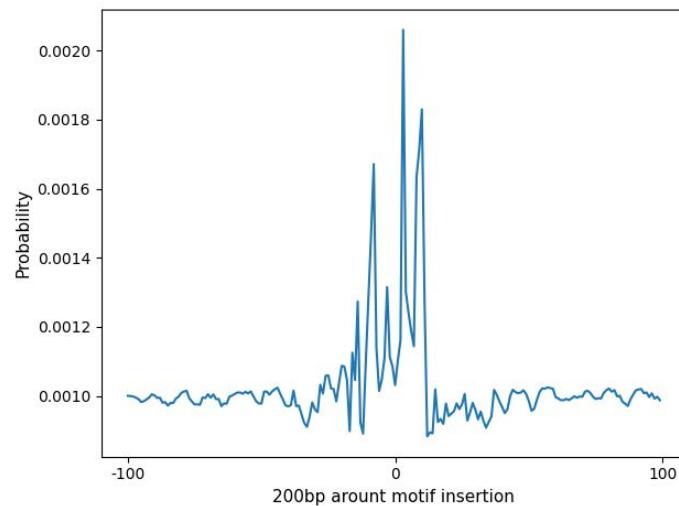


# NFKB - similar.

Week 2

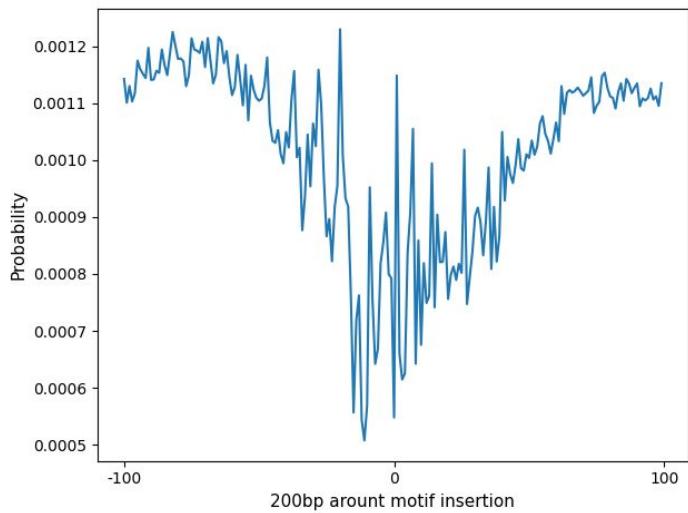


Week 8

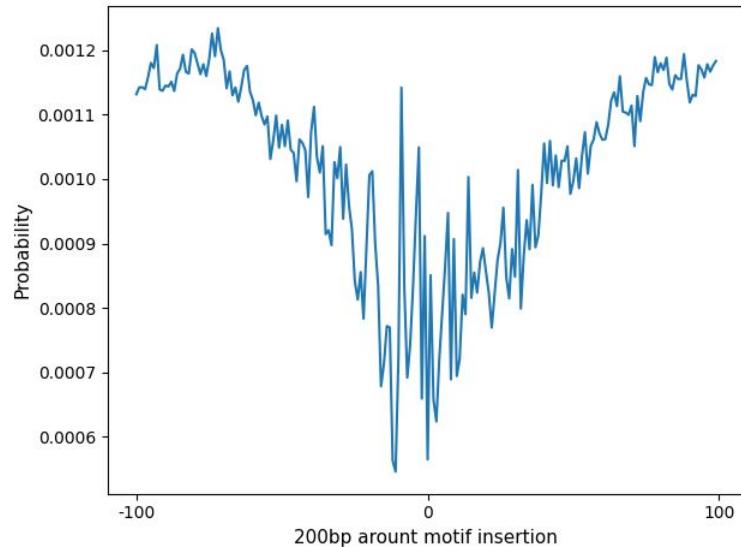


# NFYB - highly similar.

Week 2

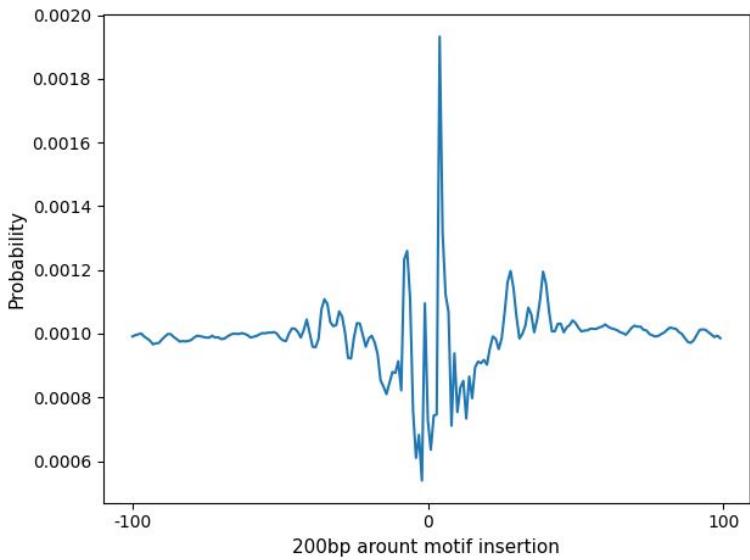


Week 8

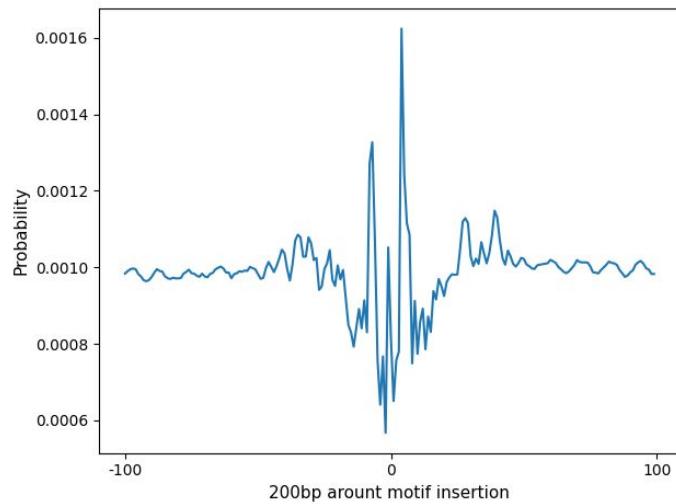


# RUNX - similar.

Week 2

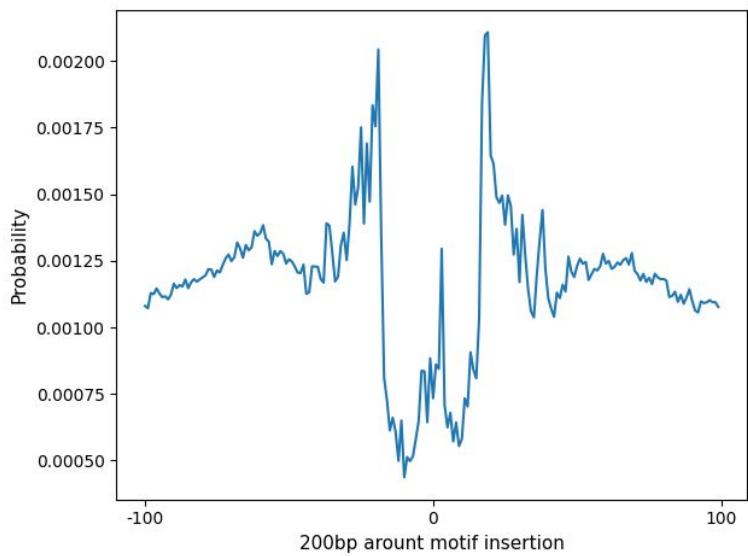


Week 8

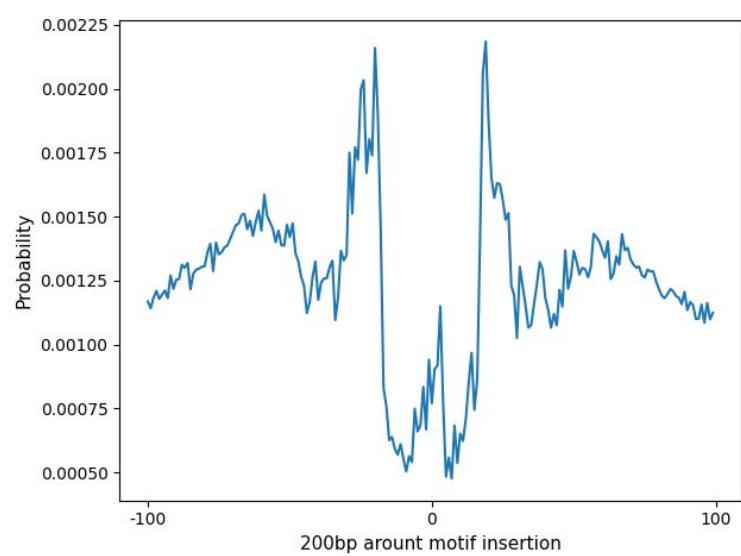


# SP1 - similar.

Week 2

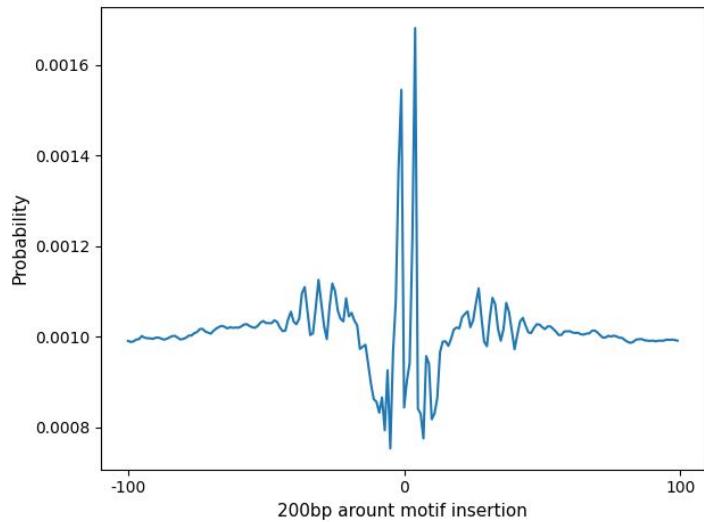


Week 8

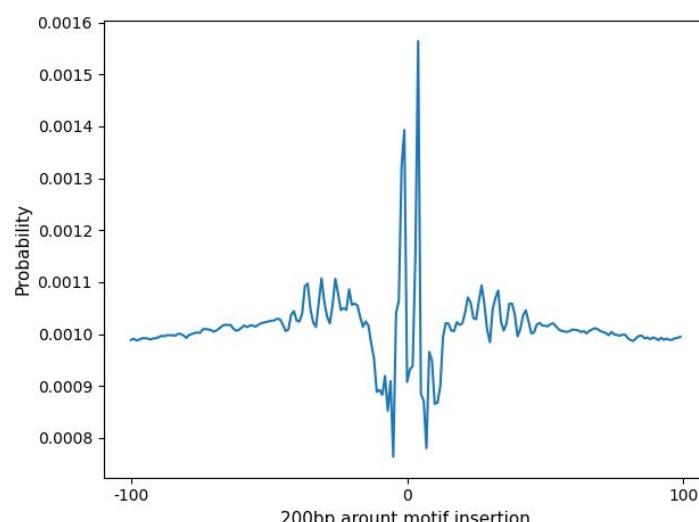


# TAL - highly similar.

2 Week



8 Week



# Predict effect of variant on 2wk, 8wk profiles

- Perform variant scoring to look for significant absolute LFC p-values
  - Calculate SHAP scores
  - Create logo plots
- Compare predicted effect of normal allele vs variant allele

SNPs we used in variant\_scoring and got shap scores for:

chr5	37859972	A	C	8*10^-6
chr7	94075721	C	G	3*10^-6
chr7	94075721	C	A	3*10^-6
chr15	22937754	A	G	9*10^-6
chr15	22937754	A	C	9*10^-6
chr15	57786063	T	A	2*10^-9
chr3	155998632	A	G	9*10^-9
chr17	66282726	C	T	4*10^-8
chr14	81144598	C	T	9*10^-8
chr1	60952057	G	T	2*10^-6
chr16	7575790	A	C	1*10^-5
chr16	7575790	A	G	1*10^-5
chr16	7575790	A	T	1*10^-5
chr15	22927754	A	G	9*10^-6
chr15	22927754	A	C	9*10^-6
chr2	3170550	T	C	2*10^-6

# Steps for variant scoring and motifs

1. Score variants using `variant_scoring.py` here: [https://github.com/kundajelab/variant-scorer/blob/main/src/variant\\_scoring.py](https://github.com/kundajelab/variant-scorer/blob/main/src/variant_scoring.py)
2. Get shap scores using `variant_shap.py` here: [https://github.com/kundajelab/variant-scorer/blob/main/src/variant\\_shap.py](https://github.com/kundajelab/variant-scorer/blob/main/src/variant_shap.py)
3. Visualize shap scores for SNPs with `abs_logfc.pval` or `abs_logfc_x_max_percentile.pval` score < 0.01 using the library here:  
[https://github.com/jbkinney/logomaker/blob/master/logomaker/tutorials/4\\_saliency\\_logos.ipynb](https://github.com/jbkinney/logomaker/blob/master/logomaker/tutorials/4_saliency_logos.ipynb)
4. For those SNPs which overlap what looks like a high scoring region (TF motif), type the high scoring sequence into this tool to find what known TF motif it matches to: <https://meme-suite.org/meme/tools/tomtom>

# Variant scoring - 8 week

8wkvaroutput.variant\_scores

chr	pos	allele1	allele2	rsid	allele1_pred_counts	allele2_pred_counts	logfc	abs_logfc	jsd	original_jsd	abs_logfc_x_jsd	abs_logfc_pval	jsd_pval	abs_logfc_x_jsd_pval
chr5	37859972	A	C	8*10^-6	404.78613	396.46832	-0.029954247	0.029954247	0.0101217	0.0101217	0.000303188	0.470588235	0.588235294	0.588235294
chr7	94075721	C	G	3*10^-6	138.89102	138.59021	-0.003127969	0.003127969	0.006948687	0.006948687	2.17E-05	0.941176471	0.882352941	0.941176471
chr7	94075721	C	A	3*10^-6	138.89102	139.41235	0.00540508	0.00540508	0.007427283	0.007427283	4.01E-05	0.941176471	0.882352941	0.941176471
chr15	22937754	A	G	9*10^-6	806.12537	794.64014	-0.020702597	0.020702597	0.011196986	0.011196986	0.000231807	0.705882353	0.529411765	0.705882353
chr15	22937754	A	C	9*10^-6	806.12537	793.8063	-0.022217253	0.022217253	0.007814653	0.007814653	0.00017362	0.705882353	0.882352941	0.705882353
chr15	57786063	T	A	2*10^-9	121.61077	123.30304	0.019937309	0.019937309	0.008273219	0.008273219	0.000164946	0.705882353	0.705882353	0.705882353
chr3	155998632	A	G	9*10^-9	393.64886	361.6845	-0.12217753	0.12217753	0.01266249	0.01266249	0.001547072	<b>0.058823529</b>	0.352941176	0.058823529
chr17	66282726	C	T	4*10^-8	339.12805	327.20016	-0.051656656	0.051656656	0.012515042	0.012515042	0.000646485	0.294117647	0.352941176	0.294117647
chr14	81144598	C	T	9*10^-8	342.9475	322.4348	-0.08898021	0.08898021	0.01087231	0.01087231	0.00096742	<b>0.058823529</b>	0.529411765	0.176470588
chr1	60952057	G	T	2*10^-6	837.3866	799.312	-0.06713503	0.06713503	0.010582278	0.010582278	0.000710442	<b>0.117647059</b>	0.529411765	0.294117647
chr16	7575790	A	C	1*10^-5	153.49127	153.45074	-0.000380992	0.000380992	0.014290824	0.014290824	5.44E-06	0.941176471	0.235294118	0.941176471
chr16	7575790	A	G	1*10^-5	153.49127	153.80984	0.002991287	0.002991287	0.013361039	0.013361039	4E-05	0.941176471	0.294117647	0.941176471
chr16	7575790	A	T	1*10^-5	153.49127	153.55571	0.000605596	0.000605596	0.013198125	0.013198125	7.99E-06	0.941176471	0.294117647	0.941176471
chr15	22927754	A	G	9*10^-6	185.98972	184.50473	-0.011565027	0.011565027	0.006726093	0.006726093	7.78E-05	0.764705882	0.882352941	0.764705882
chr15	22927754	A	C	9*10^-6	185.98972	184.76782	-0.00950932	0.00950932	0.007122014	0.007122014	6.77E-05	0.764705882	0.882352941	0.823529412
chr2	3170550	T	C	2*10^-6	449.52136	450.308	0.002522499	0.002522499	0.008258193	0.008258193	2.08E-05	0.941176471	0.764705882	0.941176471

# Variant scoring - 2 week

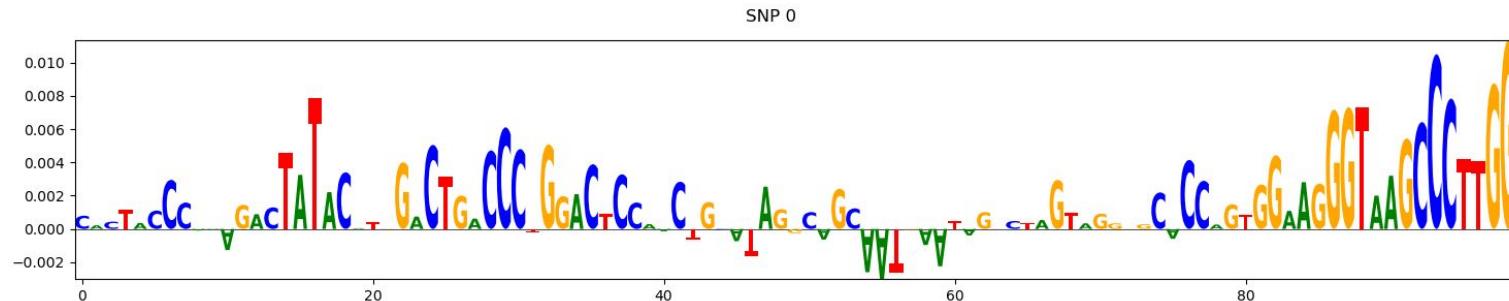
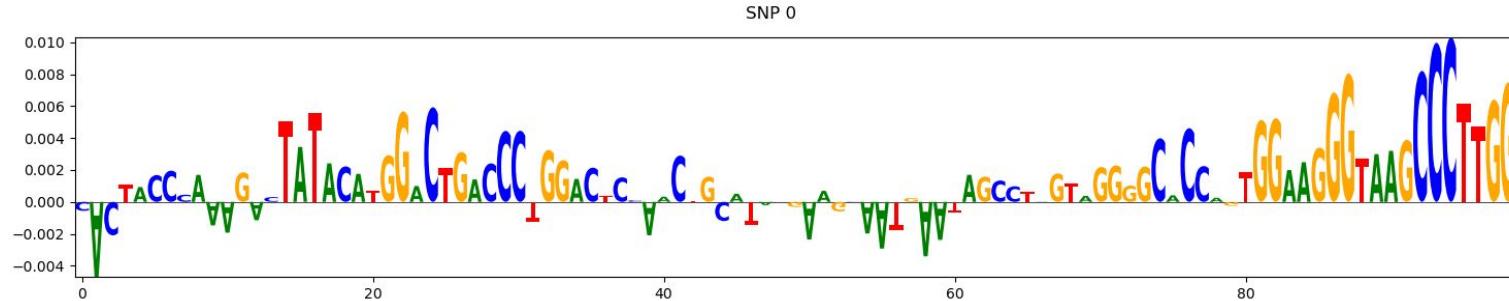
chr	pos	allele1	allele2	rsid	allele1_pred_counts	allele2_pred_counts	logfc	abs_logfc	jsd	original_jsd	abs_logfc_x_jsd	abs_logfc_pval	jsd.pval	abs_logfc_x_jsd.pval
chr5	37859972	A	C	8*10^-6	314.88074	319.07062	0.019070277	0.019070277	0.010169313241326400	0.010169313241326400	0.00019393162031558800	0.7058823529411770	0.5294117647058820	0.5294117647058820
chr7	94075721	C	G	3*10^-6	121.17729	119.89203	-0.015383626	0.015383626	0.007577417705651670	0.007577417705651670	0.00011656816256317800	0.7647058823529410	0.8235294117647060	0.8235294117647060
chr7	94075721	C	A	3*10^-6	121.17729	119.26181	-0.022987235	0.022987235	0.010830422014154700	0.010830422014154700	0.0002489614596437700	0.5882352941176470	0.5294117647058820	0.47058823529411800
chr15	22937754	A	G	9*10^-6	584.13306	588.7528	0.011364963	0.011364963	0.01301203470181480	0.01301203470181480	0.0001478812917130950	0.8235294117647060	0.35294117647058800	0.7647058823529410
chr15	22937754	A	C	9*10^-6	584.13306	589.6132	0.013471768	0.013471768	0.011860704194924500	0.011860704194924500	0.00015978465804935	0.8235294117647060	0.4117647058823530	0.7058823529411770
chr15	57786063	T	A	2*10^-9	102.85942	103.72259	0.012056203	0.012056203	0.007712856584559810	0.007712856584559810	9.29877618219855E-05	0.8235294117647060	0.8235294117647060	0.8235294117647060
chr3	155998632	A	G	9*10^-9	221.29166	217.61926	-0.024142763	0.024142763	0.00494890925240193	0.00494890925240193	0.00011948034143772400	0.5294117647058820	0.9411764705882350	0.8235294117647060
chr17	66282726	C	T	4*10^-8	287.71063	282.7006	-0.025343673	0.025343673	0.008828053479411310	0.008828053479411310	0.00022373530328999800	0.5294117647058820	0.7058823529411770	0.5294117647058820
chr14	81144598	C	T	9*10^-8	154.81073	150.60007	-0.039783075	0.039783075	0.008377144679491830	0.008377144679491830	0.0003332685788552950	0.4117647058823530	0.7647058823529410	0.4117647058823530
chr1	60952057	G	T	2*10^-6	511.49805	491.05804	-0.05883514	0.05883514	0.013013738944625000	0.013013738944625000	0.0007656651704390330	0.29411764705882400	0.35294117647058800	0.29411764705882400
chr16	7575790	A	C	1*10^-5	135.1637	136.41113	0.013253659	0.013253659	0.01670777568007060	0.01670777568007060	0.0002214391616781150	0.8235294117647060	0.17647058823529400	0.5294117647058820
chr16	7575790	A	G	1*10^-5	135.1637	135.27805	0.0012200448	0.0012200448	0.017023361695360700	0.017023361695360700	2.0769263852895E-05	0.9411764705882350	0.11764705882352900	0.9411764705882350
chr16	7575790	A	T	1*10^-5	135.1637	135.18523	0.00022975054	0.00022975054	0.016813303189575500	0.016813303189575500	3.86286551424118E-06	0.9411764705882350	0.17647058823529400	0.9411764705882350
chr15	22927754	A	G	9*10^-6	146.21786	146.87427	0.006462029	0.006462029	0.009019599801327990	0.009019599801327990	5.82849171209187E-05	0.9411764705882350	0.5882352941176470	0.9411764705882350
chr15	22927754	A	C	9*10^-6	146.21786	147.01263	0.007820486	0.007820486	0.006394143565760460	0.006394143565760460	5.00053108203813E-05	0.8823529411764710	0.8823529411764710	0.9411764705882350
chr2	3170550	T	C	2*10^-6	356.4243	368.6994	0.04884936	0.04884936	0.006868141814854450	0.006868141814854450	0.00033550432623924700	0.29411764705882400	0.8235294117647060	0.4117647058823530

8 Week SNPs (Allele 1 normal vs. Allele 2 variant).

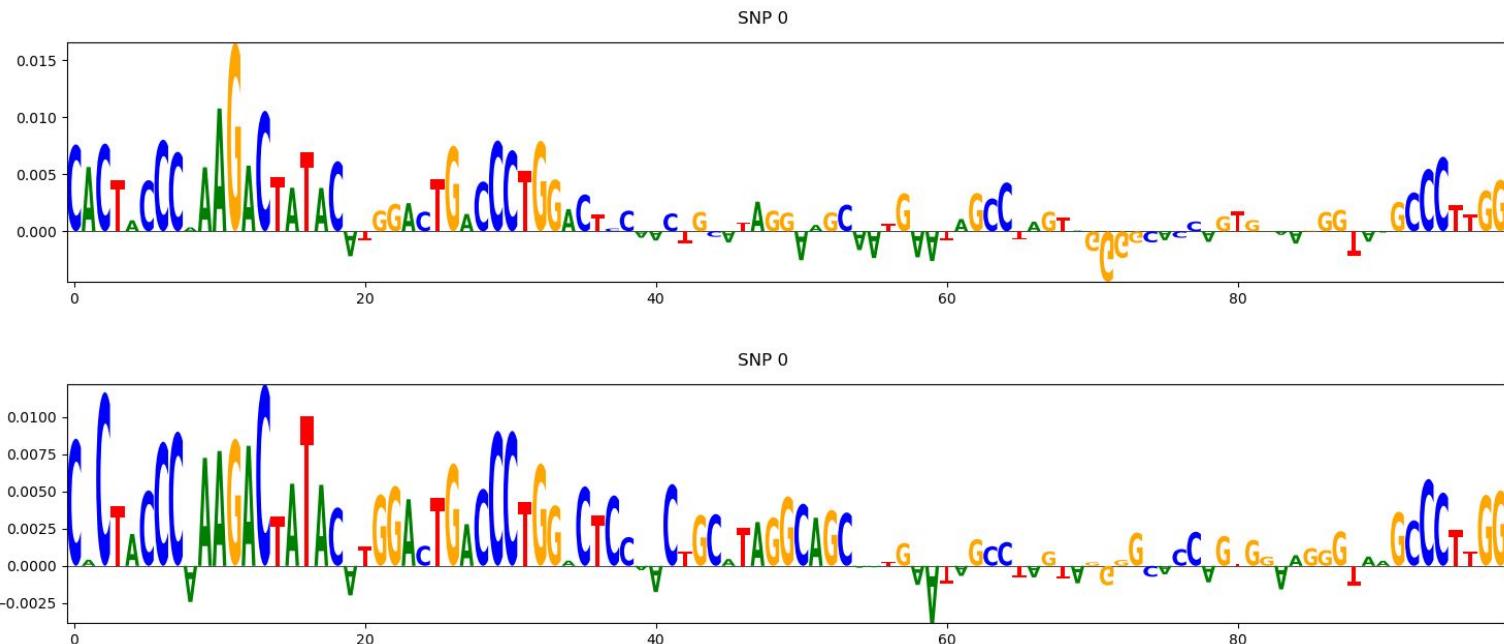
Q: What changes does variant have on motif?

# SNP 0 - 8 week

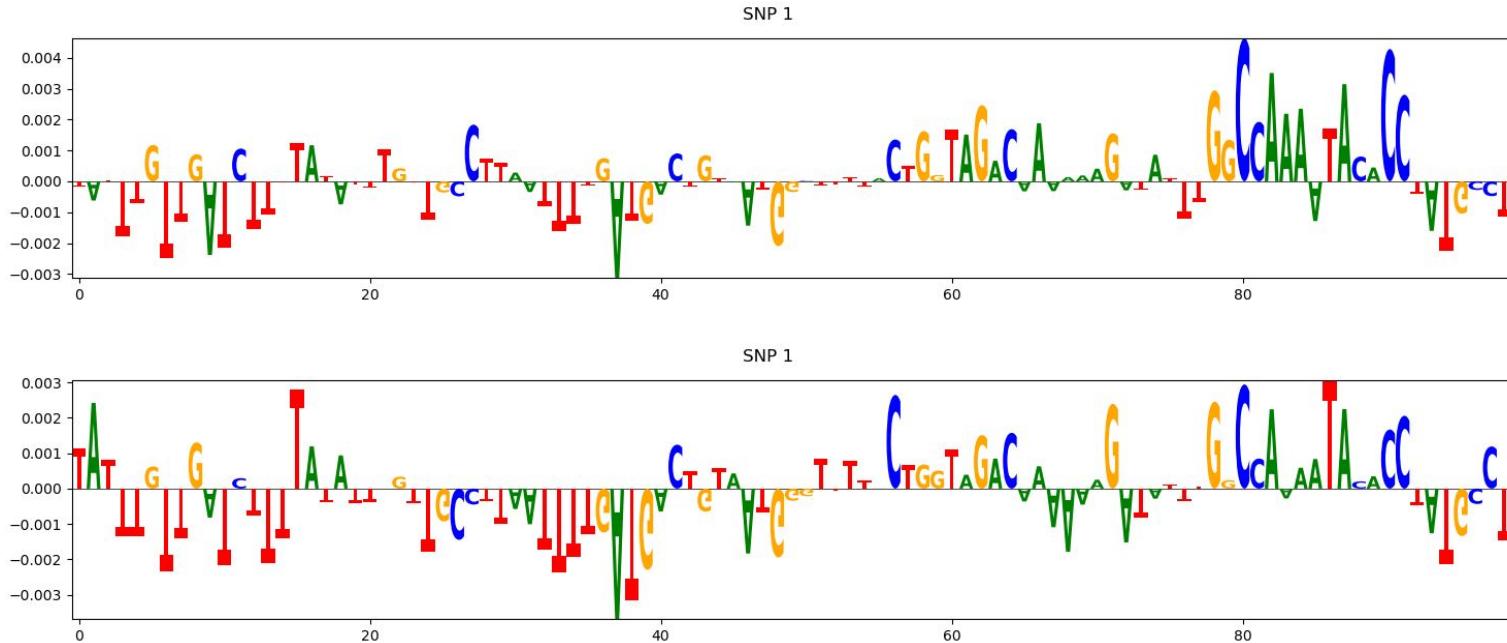
- i.e. Index 0, so this is chr5 SNP in the chart on slide #15.
- Note: Allele 1 (top) vs 2 (variant allele - bottom) for all slides.



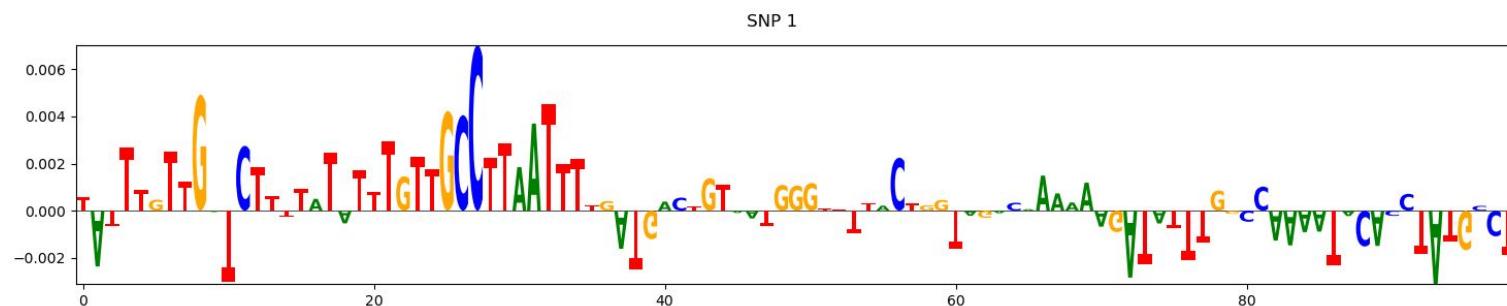
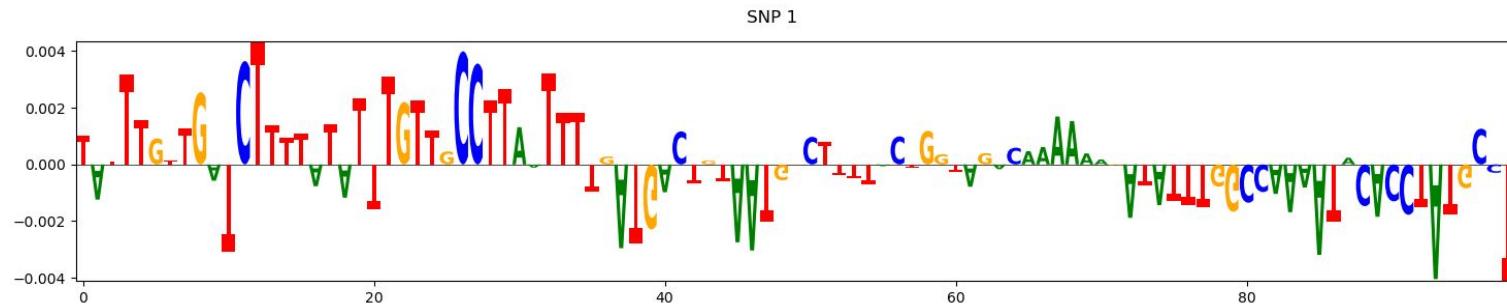
# SNP 0 - 2 week



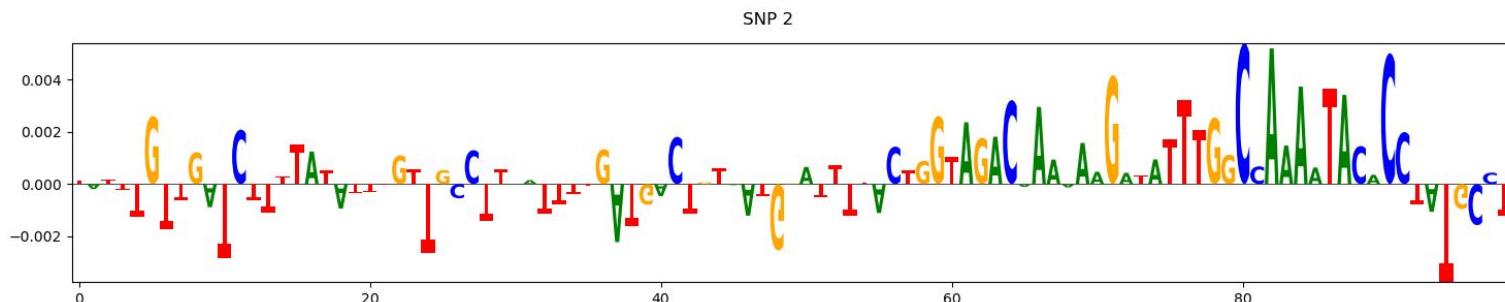
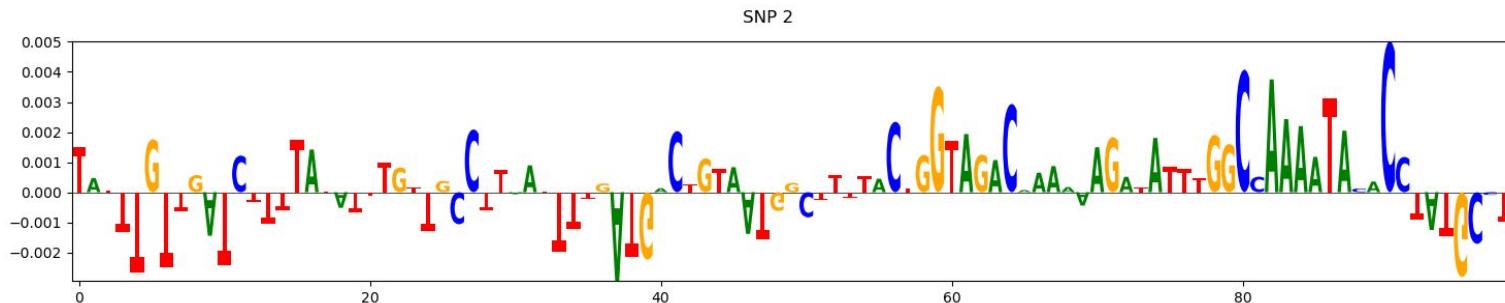
# SNP 1 - 8wk allele 1 (top) vs 2 (variant - bottom).



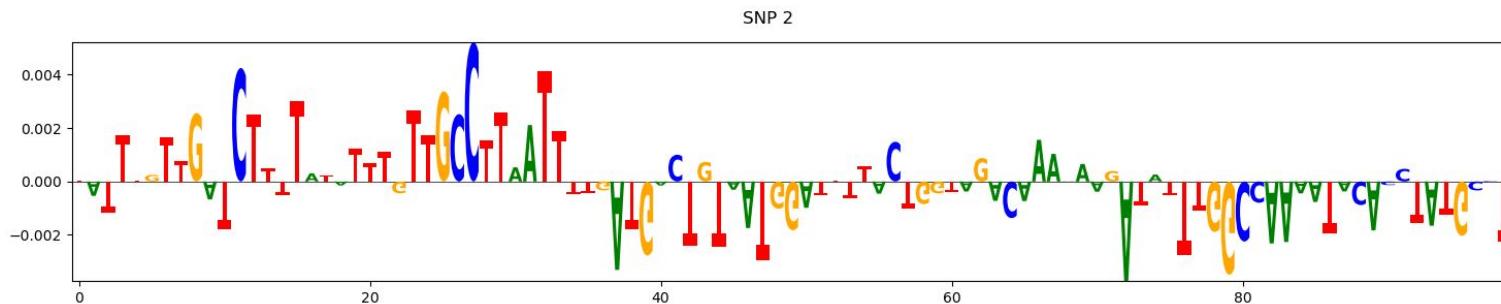
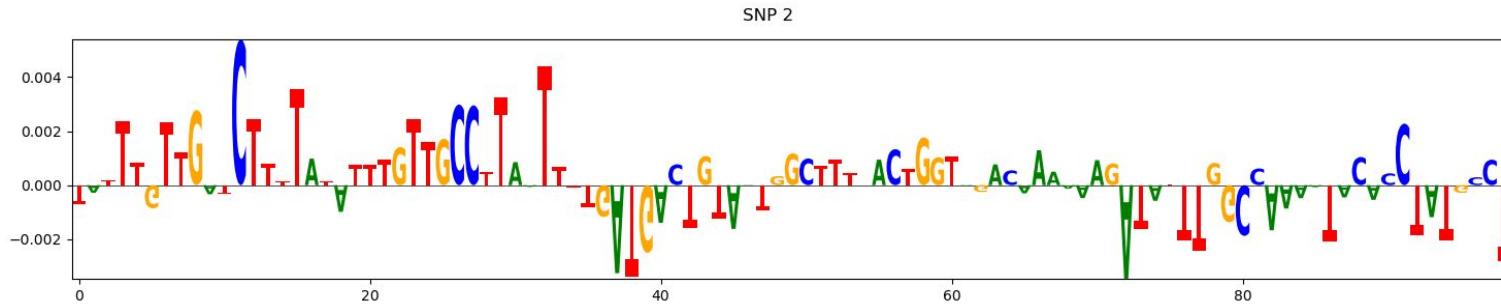
## SNP 1 - 2 week



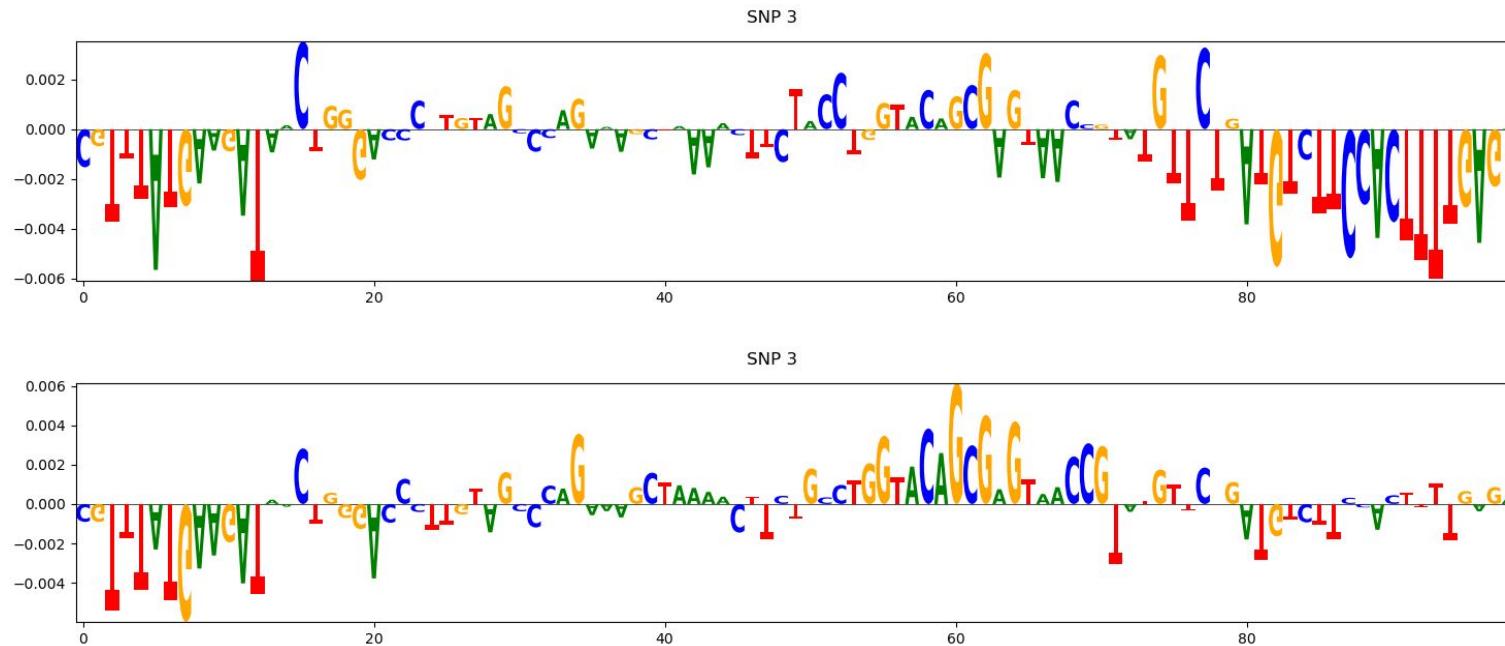
# SNP 2 - 8 week



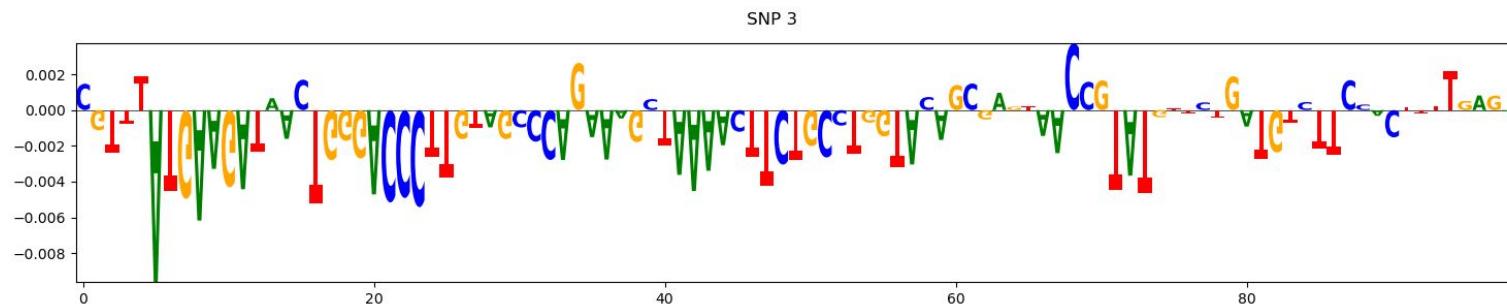
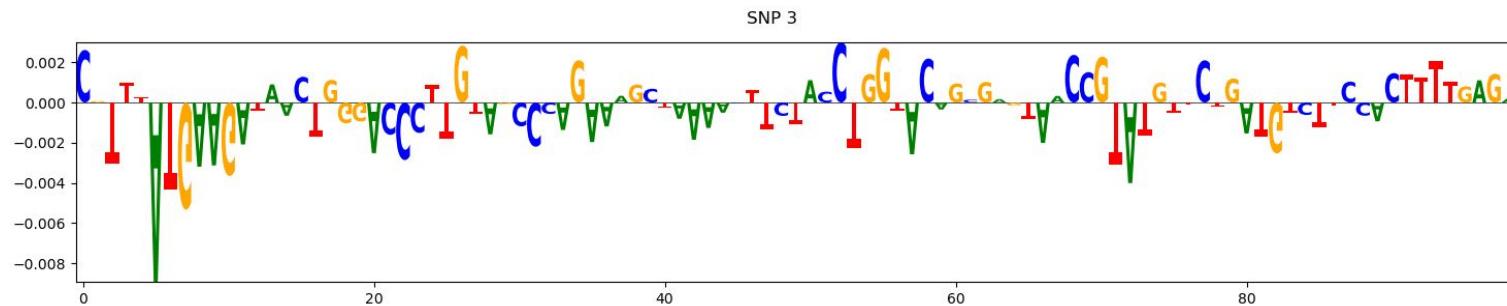
## SNP 2 - 2 week



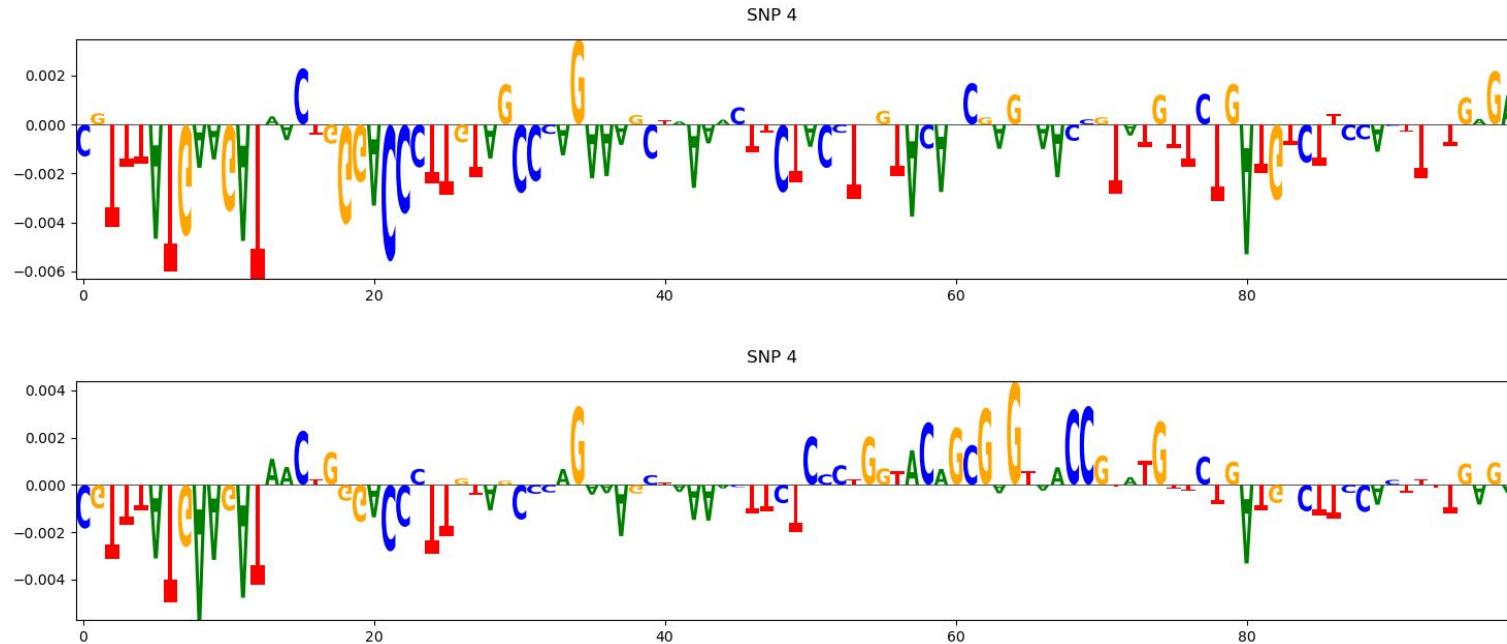
# SNP 3 - 8 week



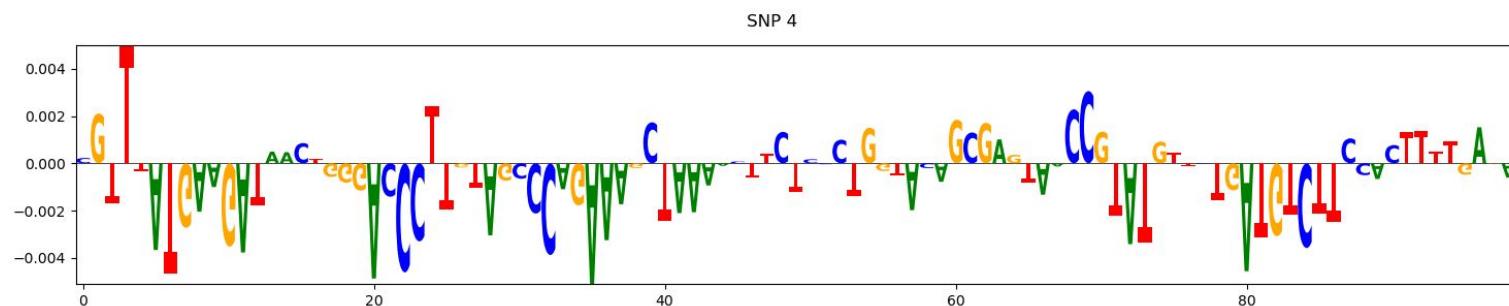
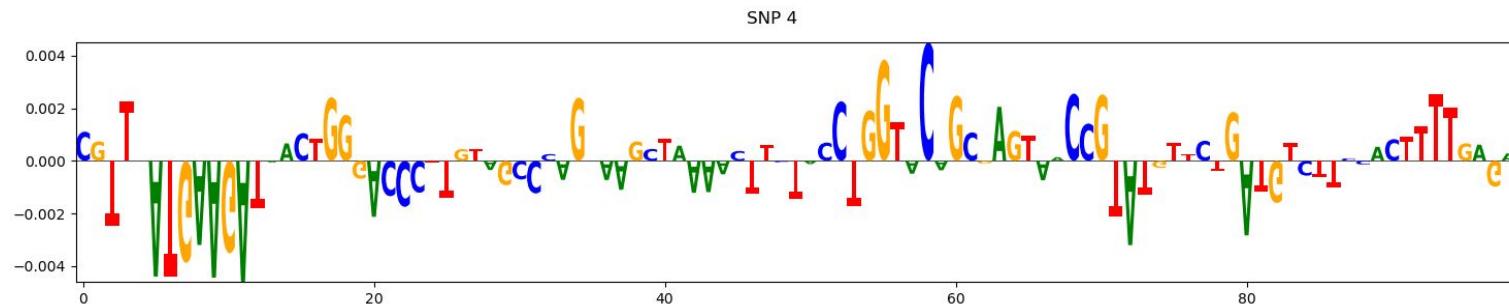
## SNP 3 - 2 week



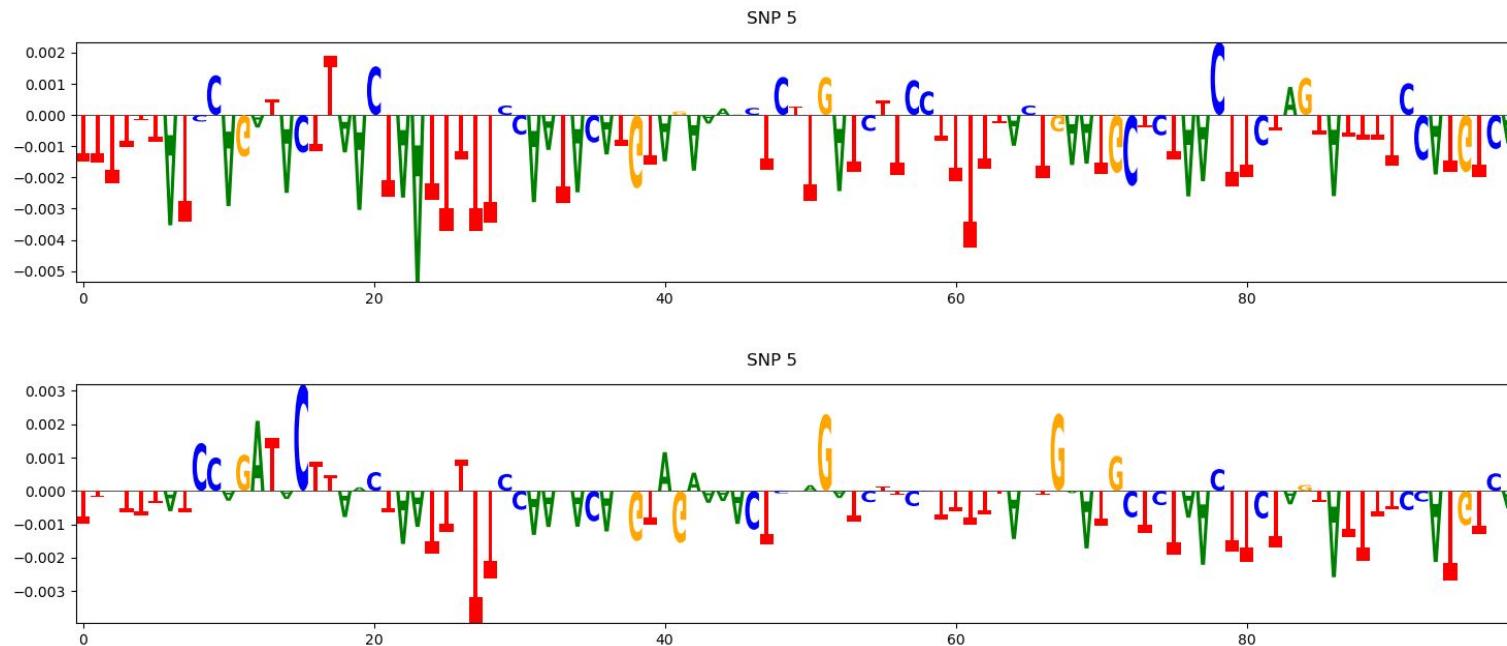
# SNP 4 - 8 week



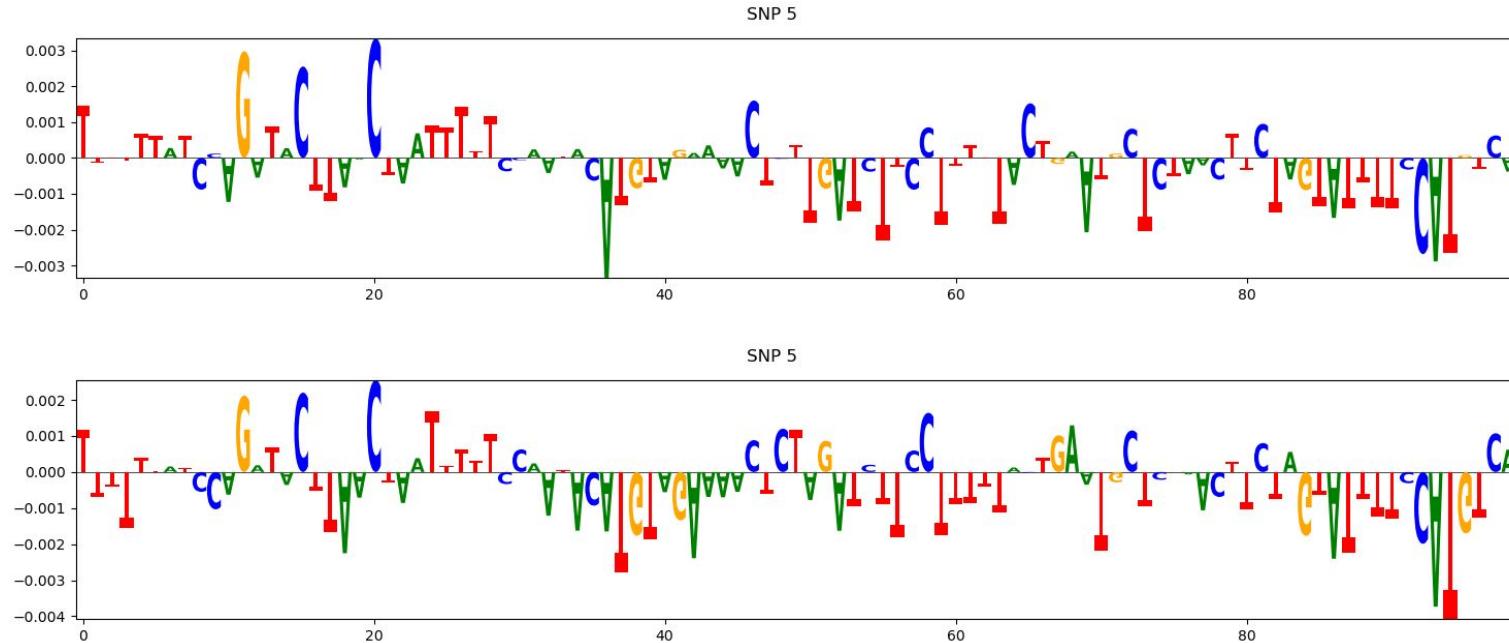
## SNP 4 - 2 week



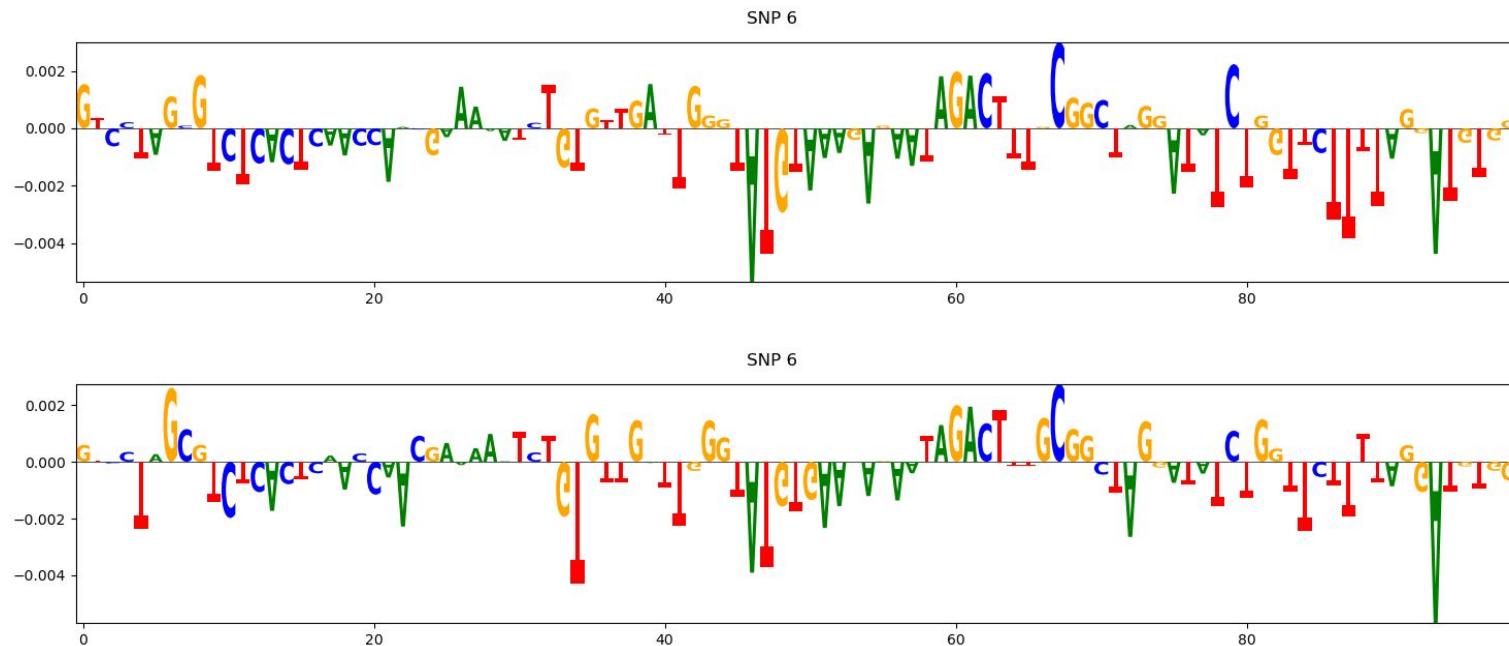
# SNP 5 - 8 week



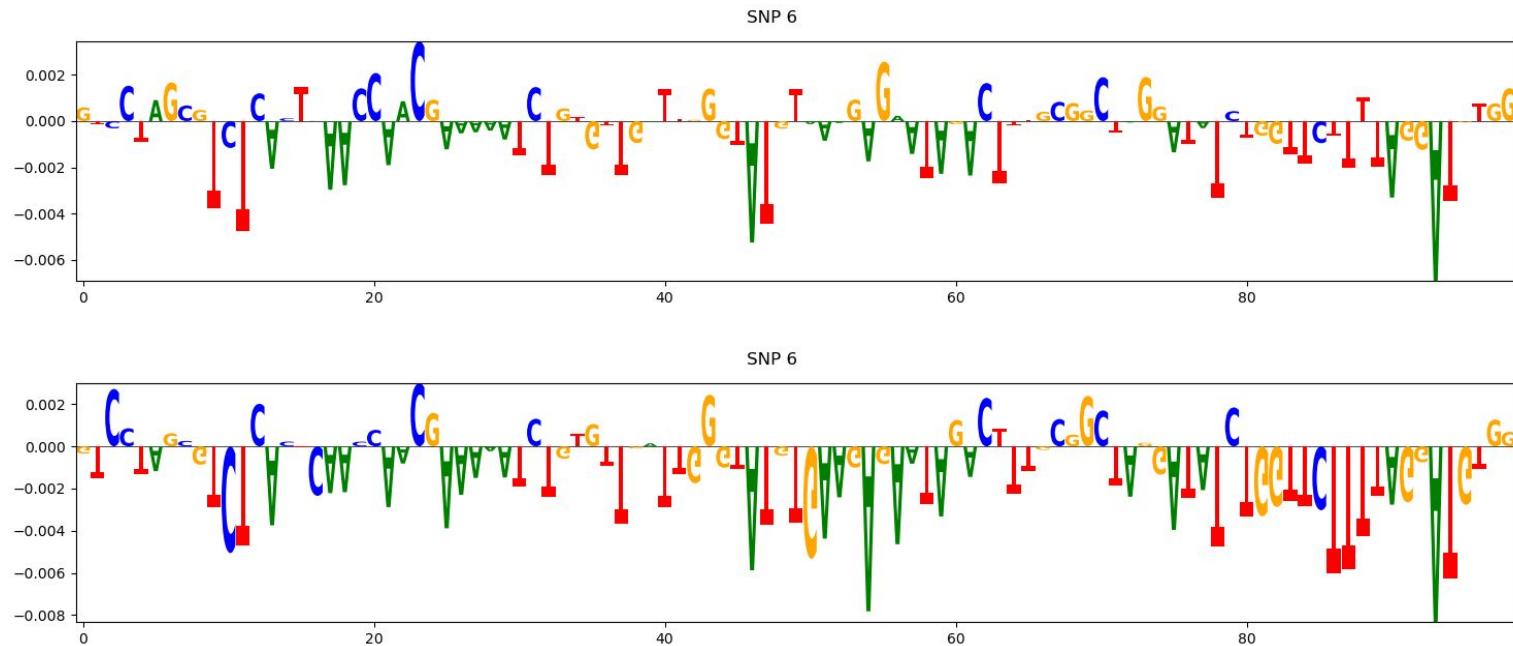
## SNP 5 - 2 week



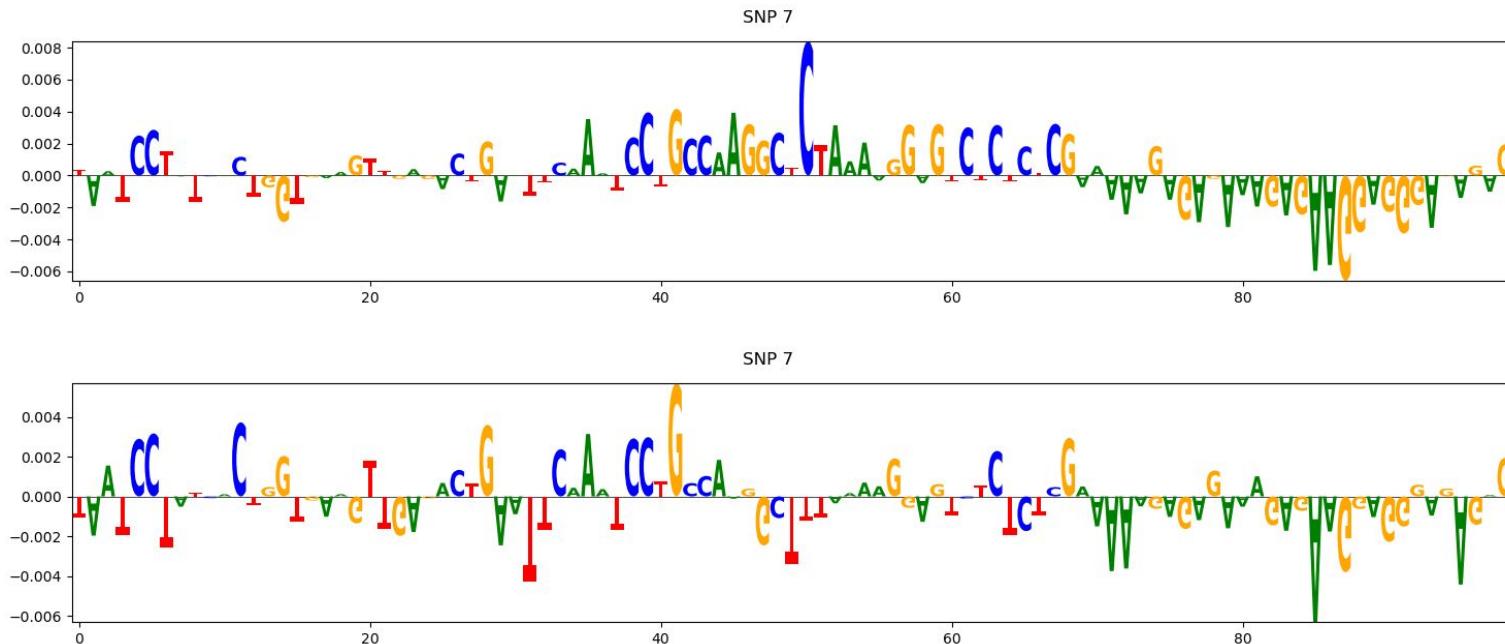
# SNP 6 - 8 week



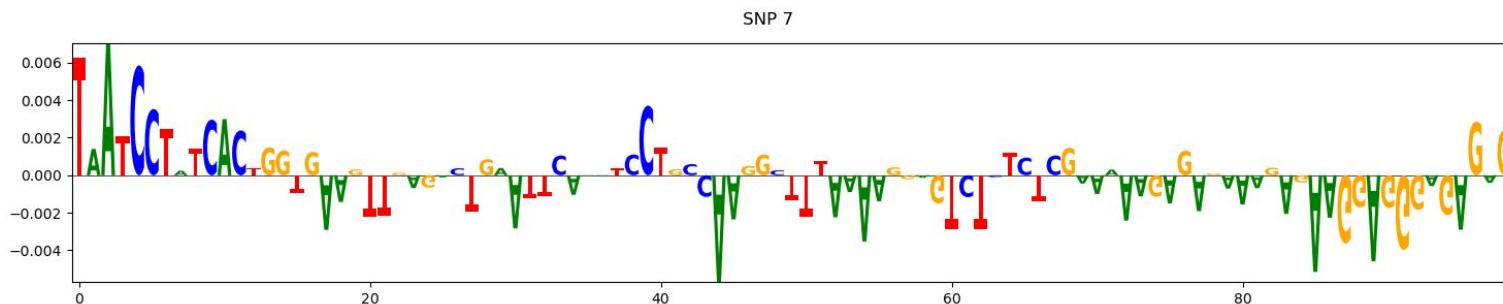
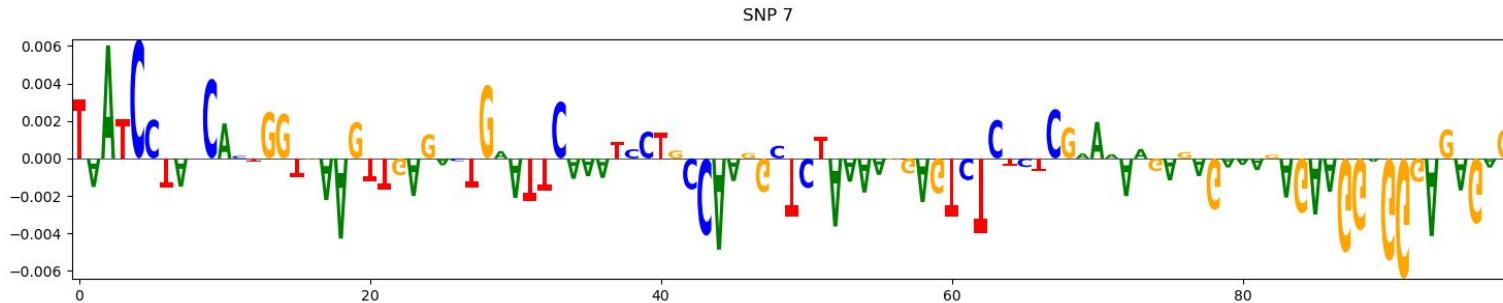
## SNP 6 - 2 week



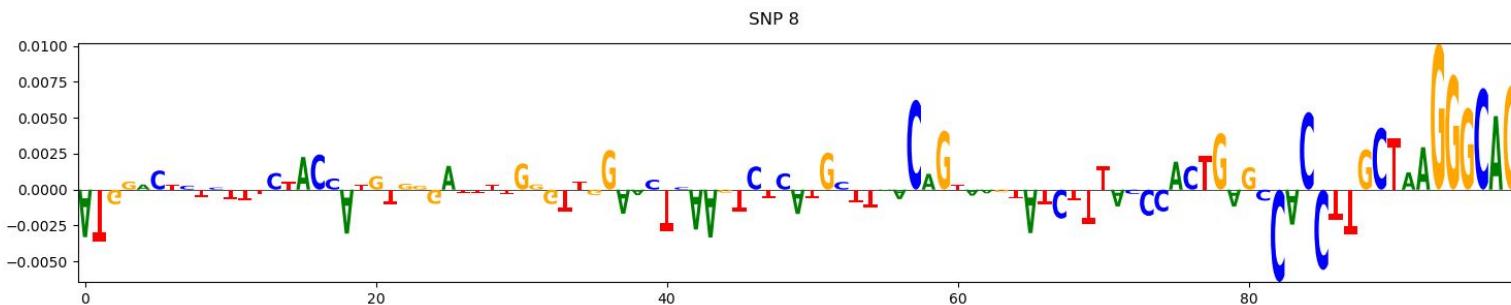
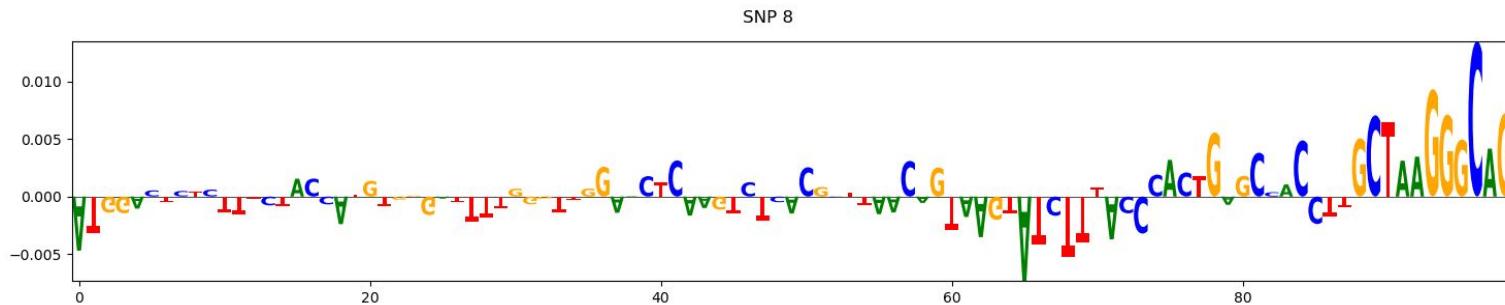
# SNP 7 - 8 week



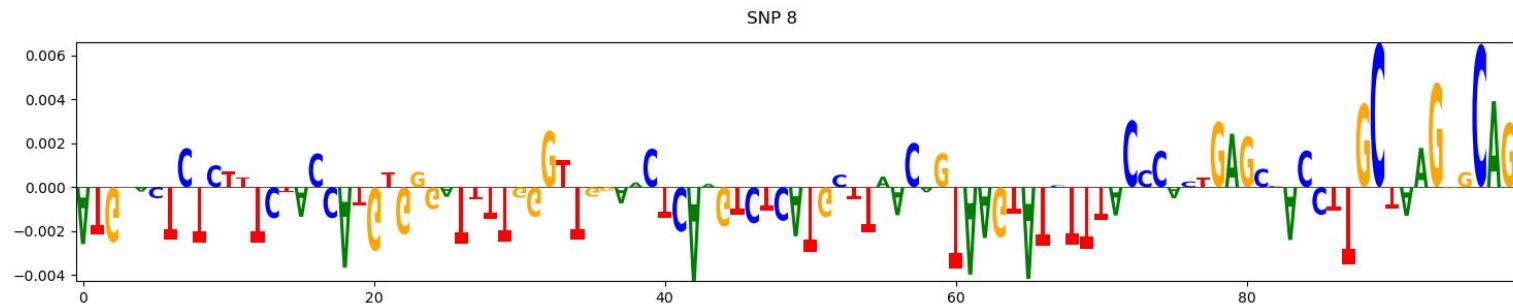
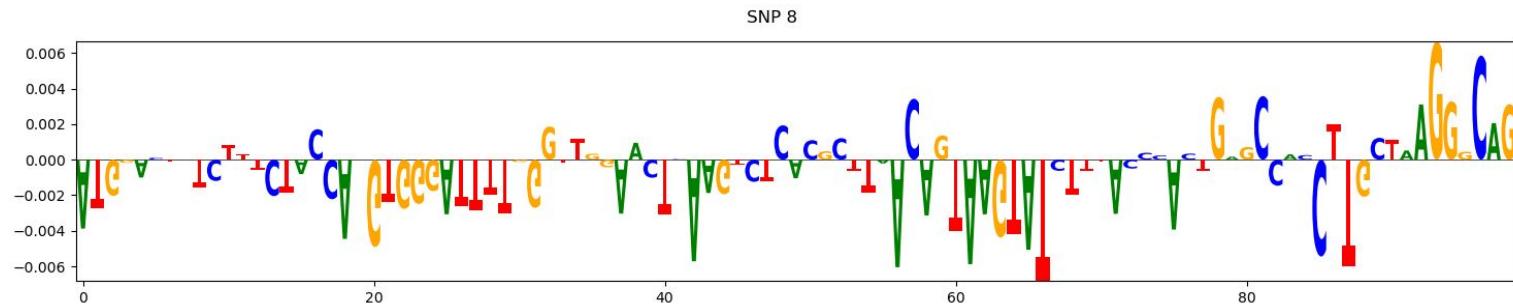
## SNP 7 - 2 week



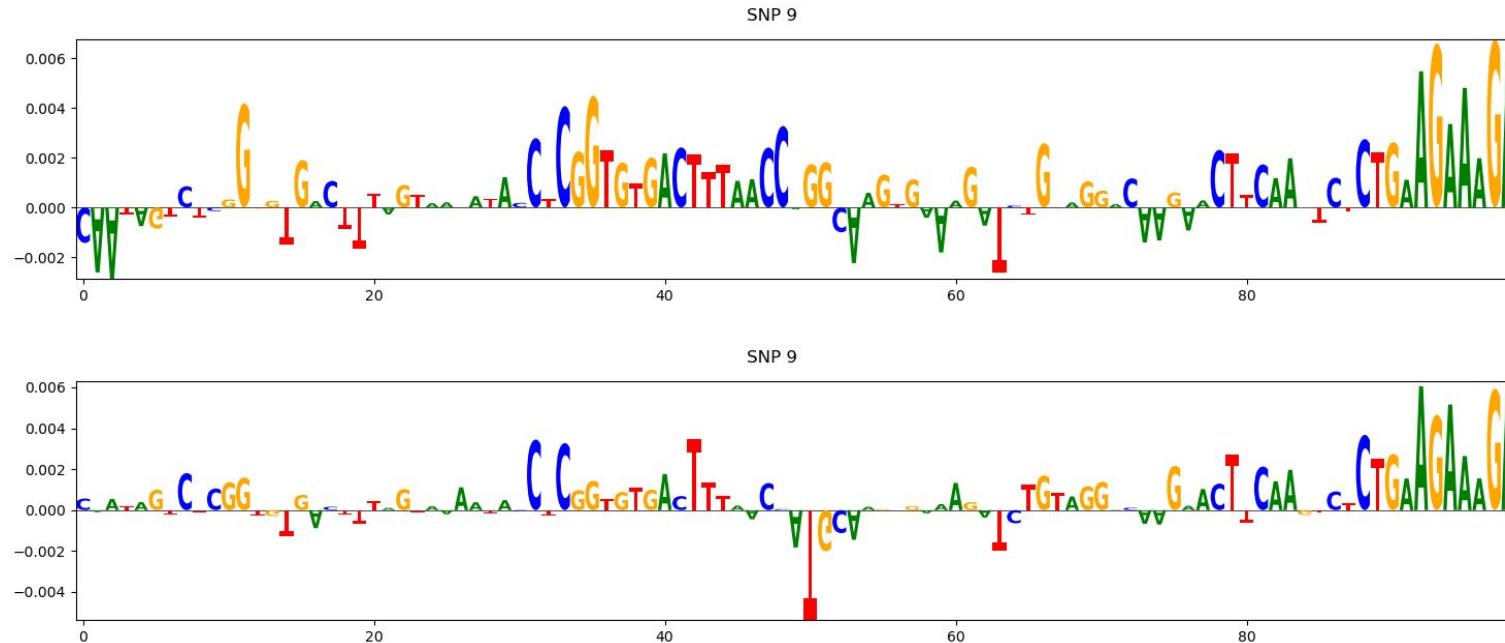
# SNP 8 - 8 week



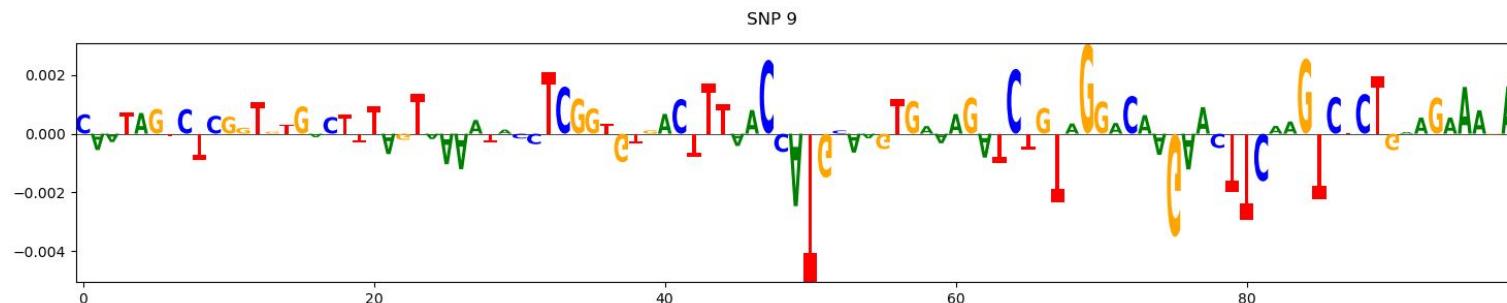
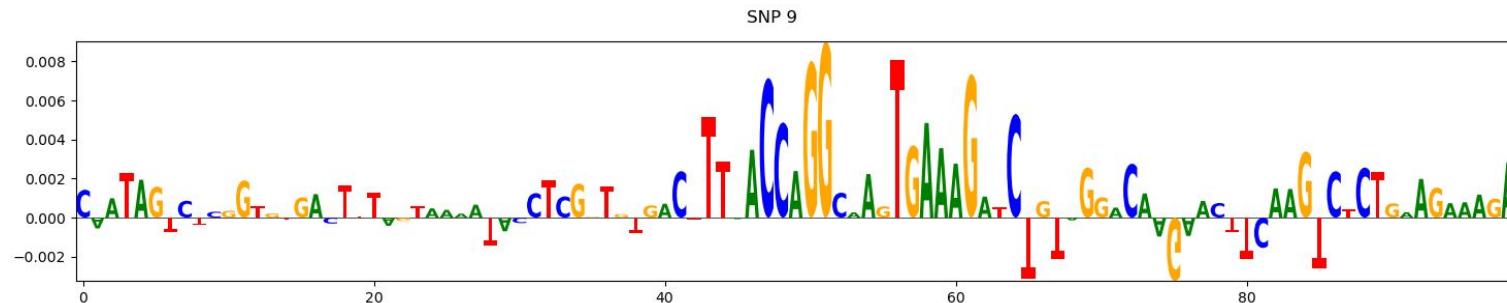
## SNP 8 - 2 week



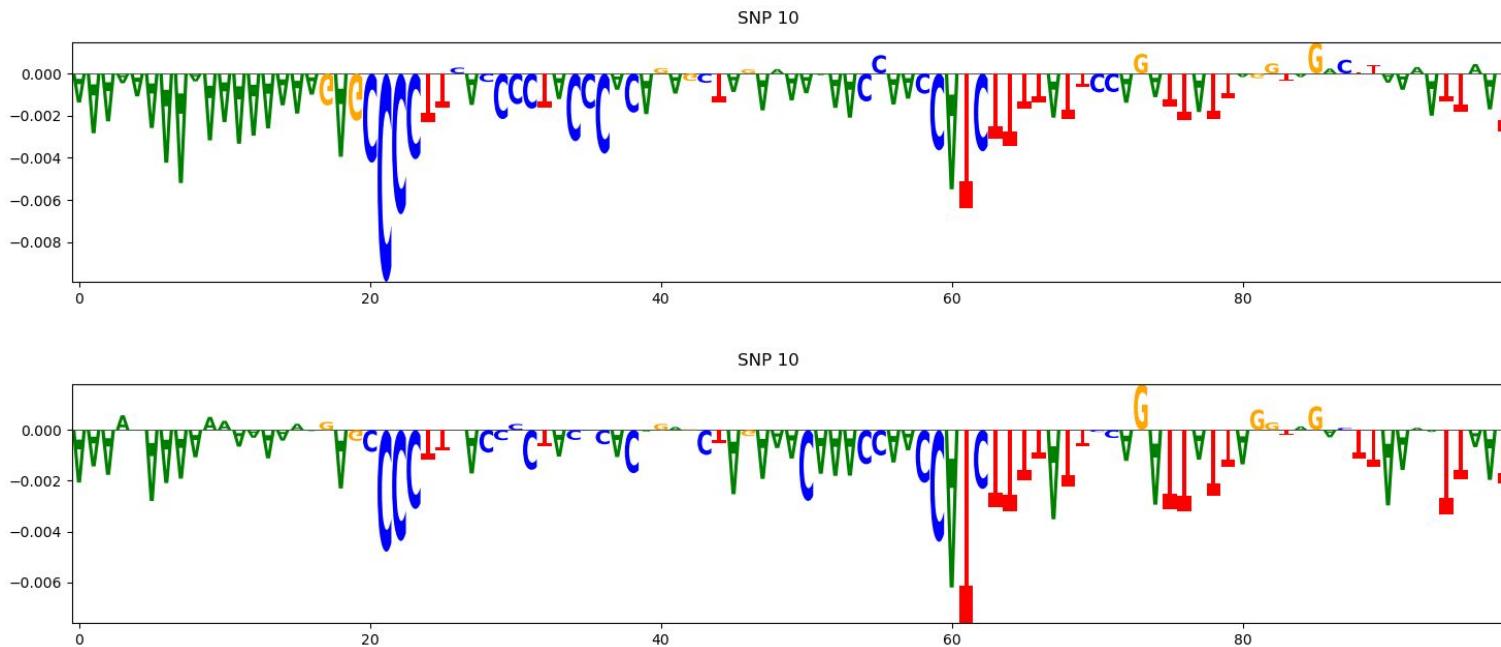
# SNP 9 - 8 week



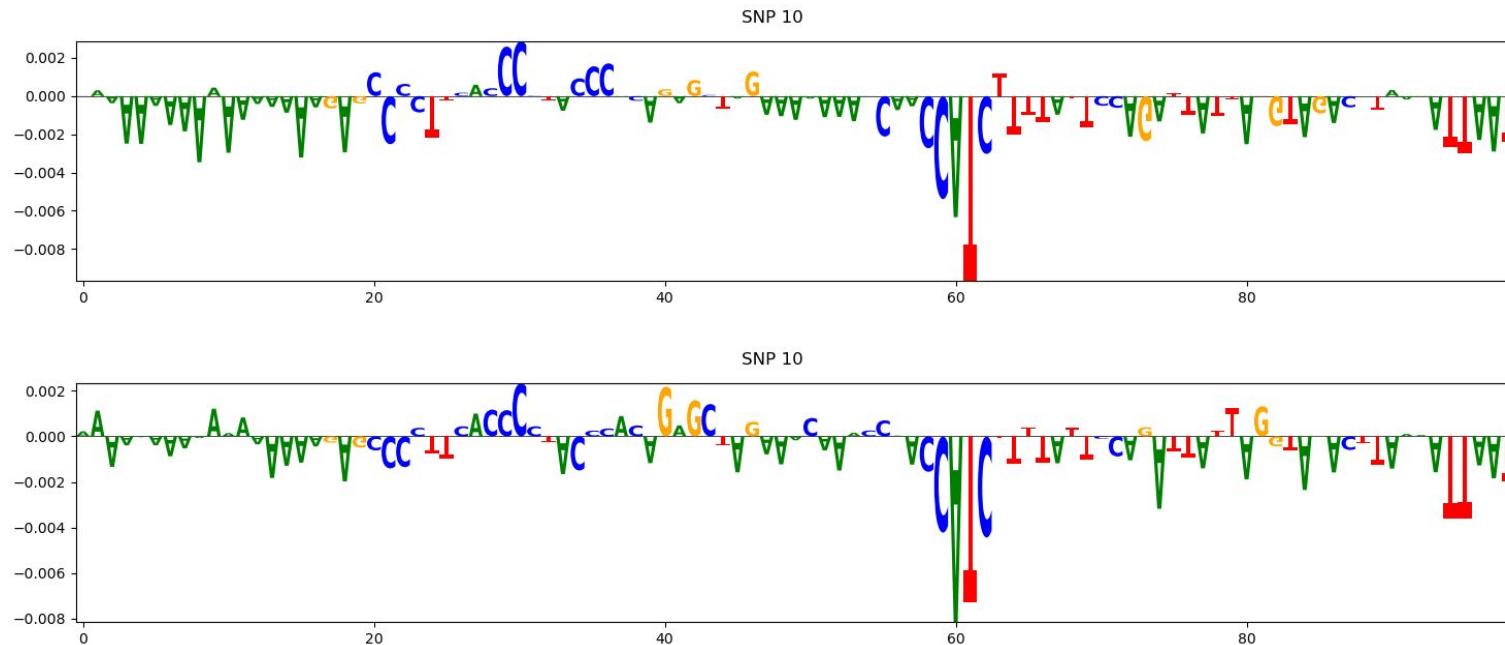
# SNP 9 - 2 week



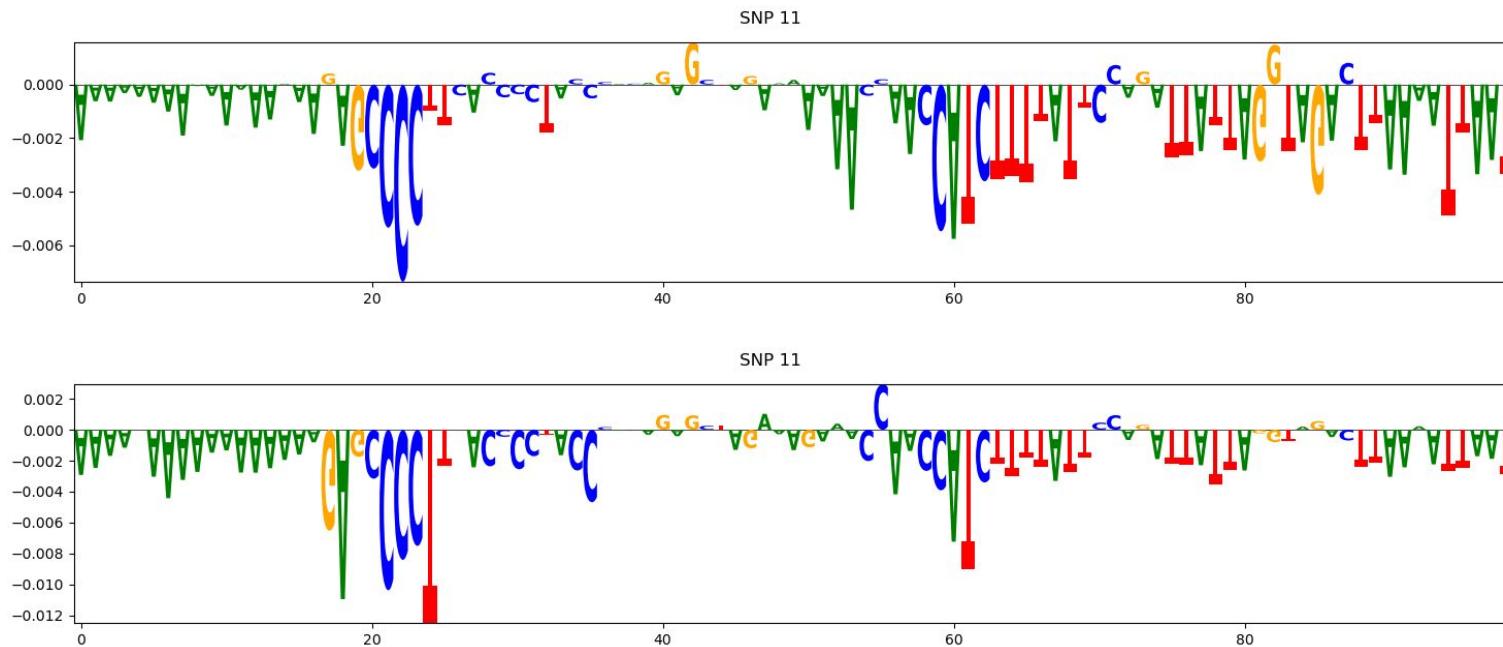
# SNP 10 - 8 week



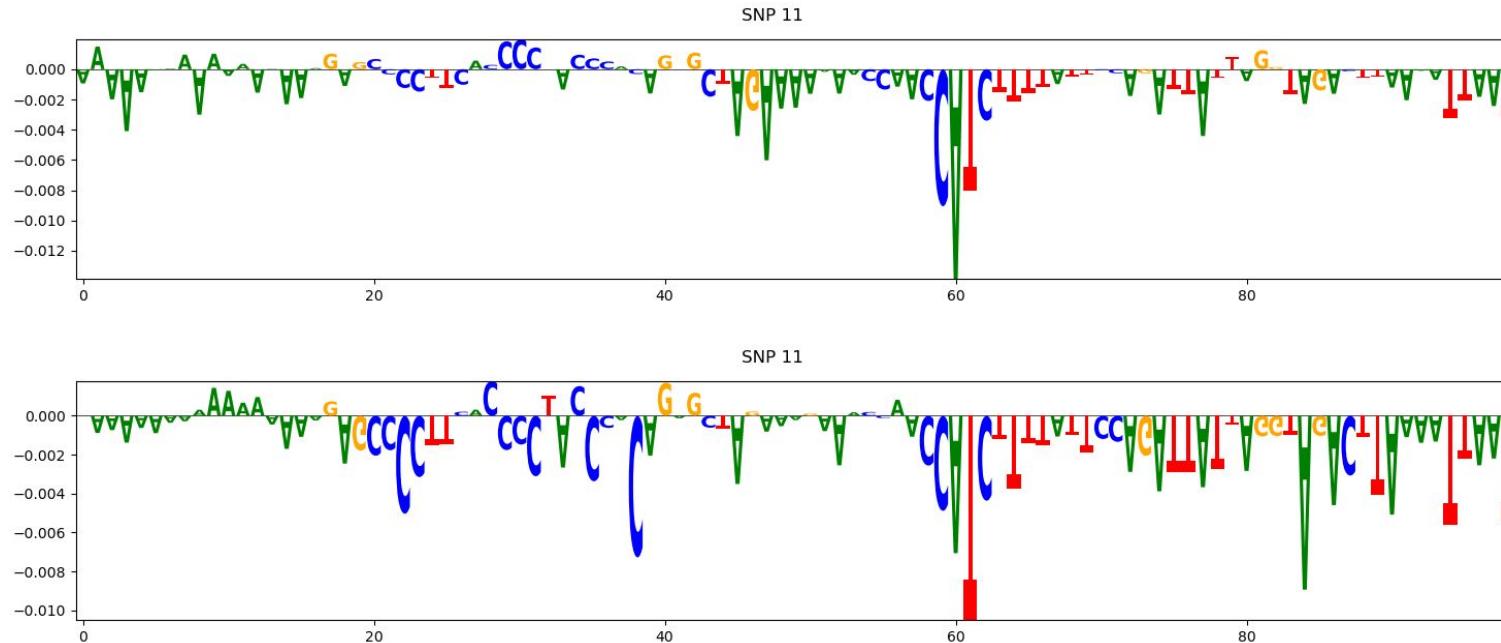
# SNP 10 - 2 week



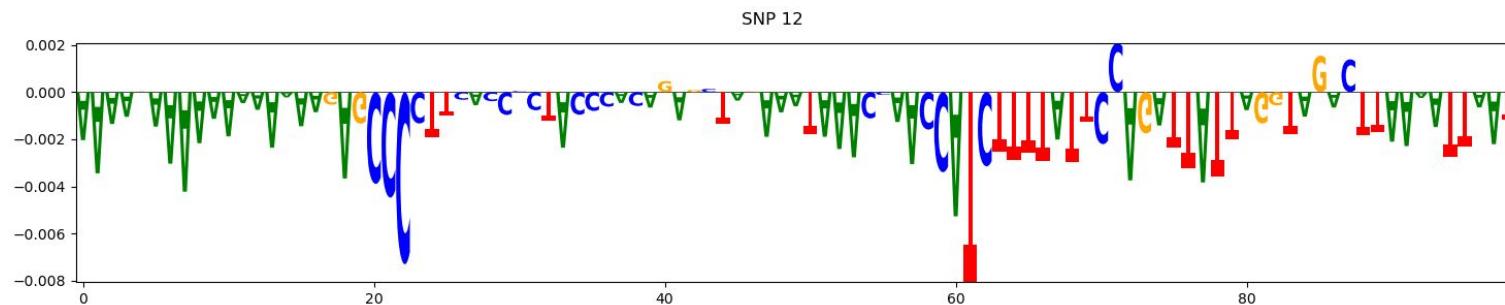
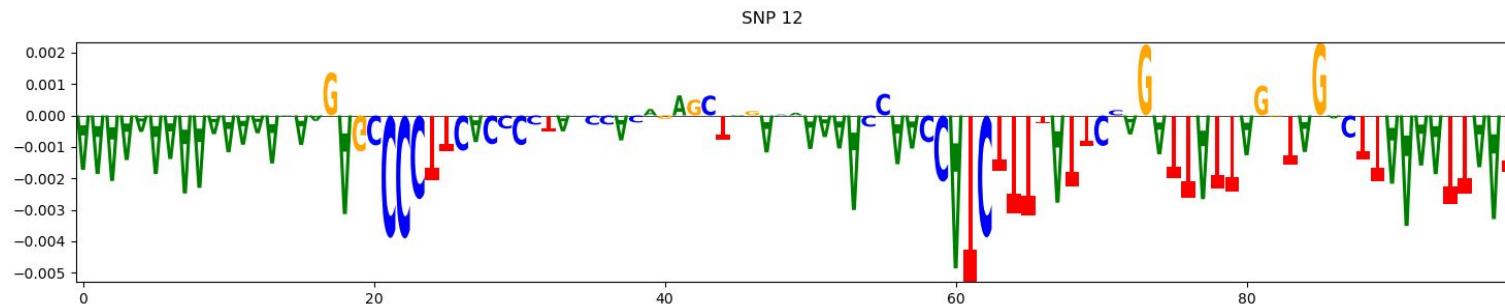
# SNP 11 - 8 week



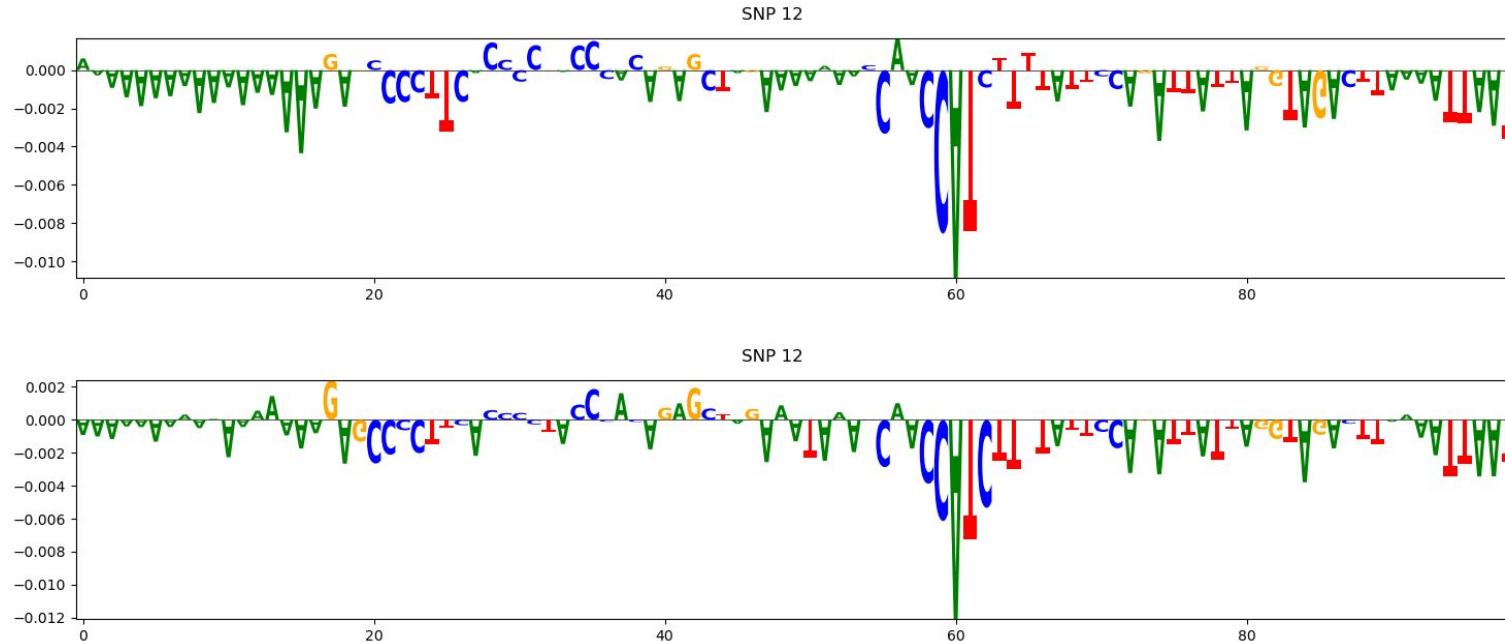
# SNP 11 - 2 week



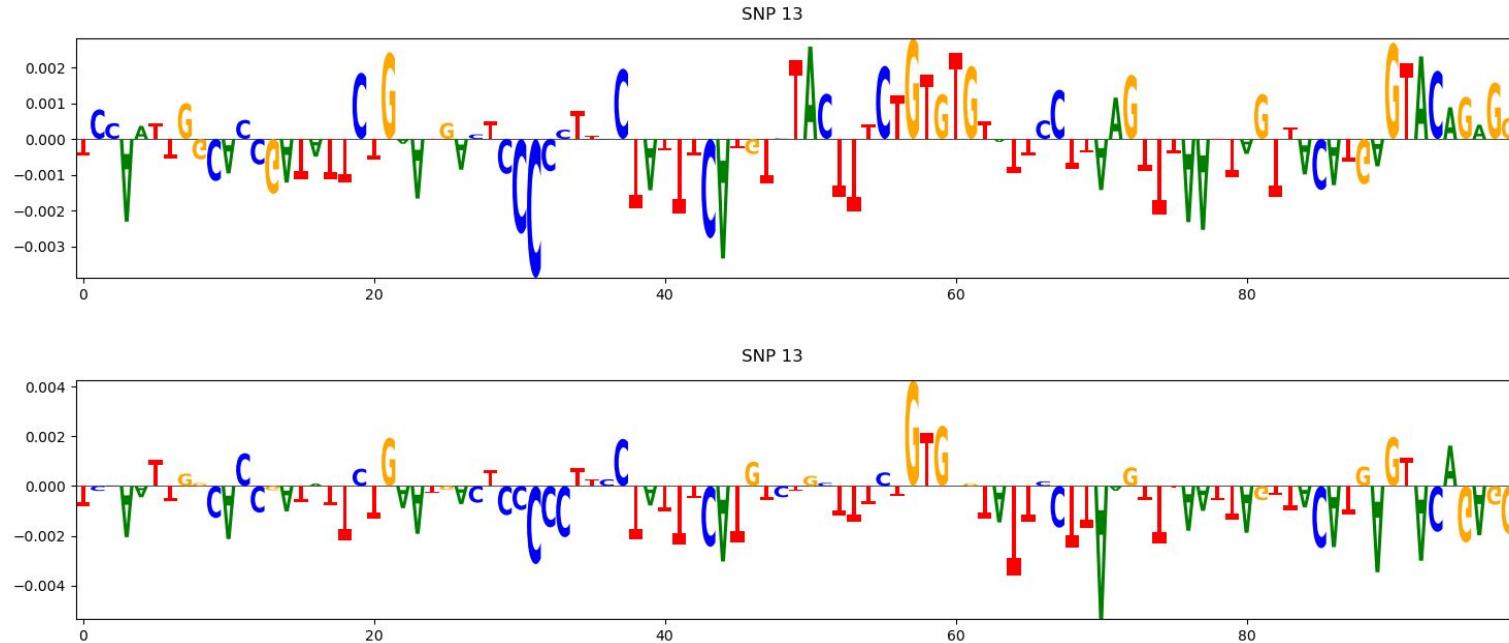
# SNP 12 - 8 week



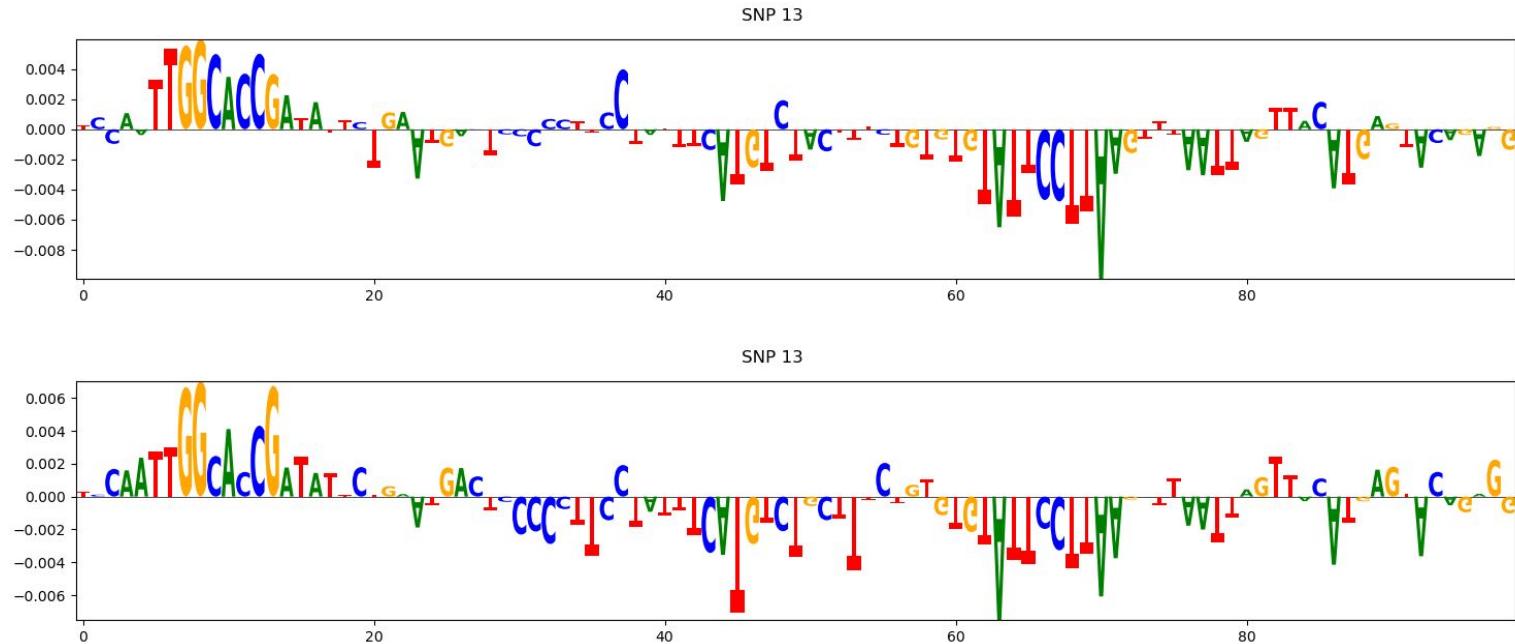
# SNP 12 - 2 week



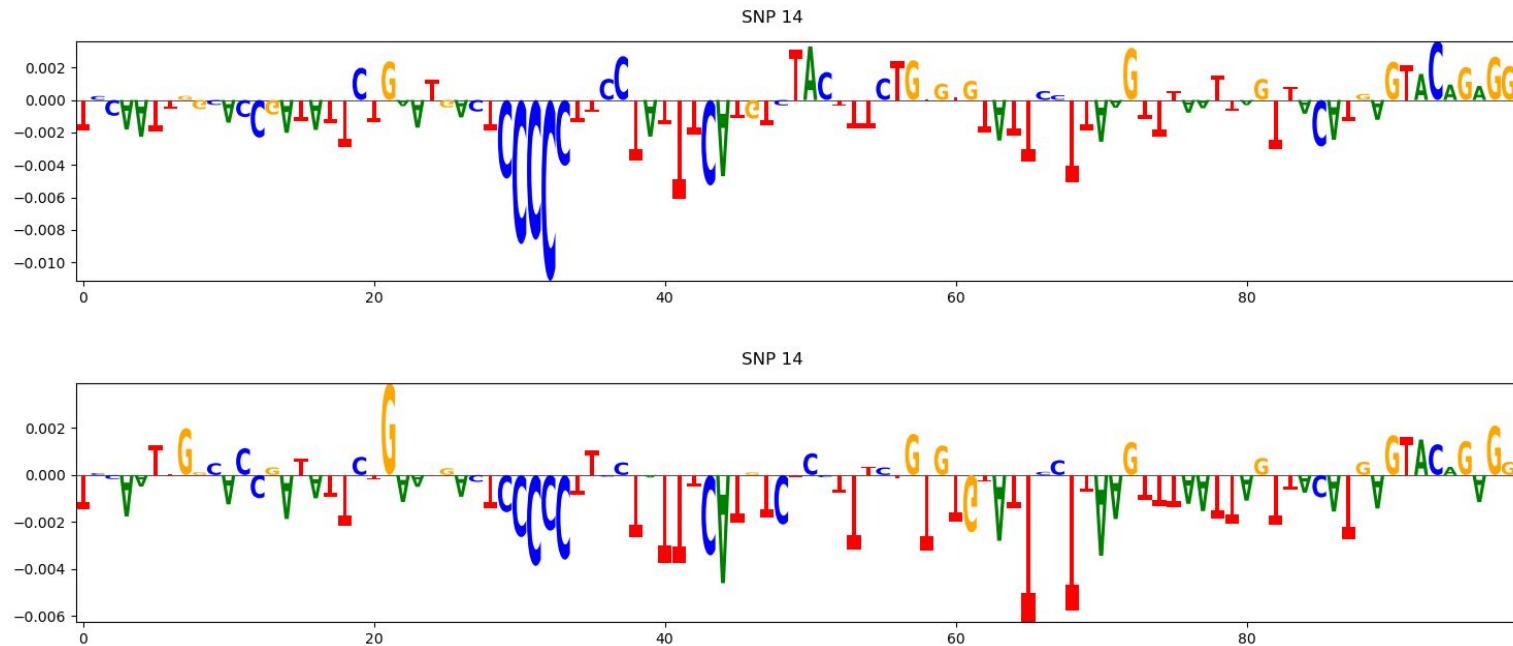
# SNP 13 - 8 week



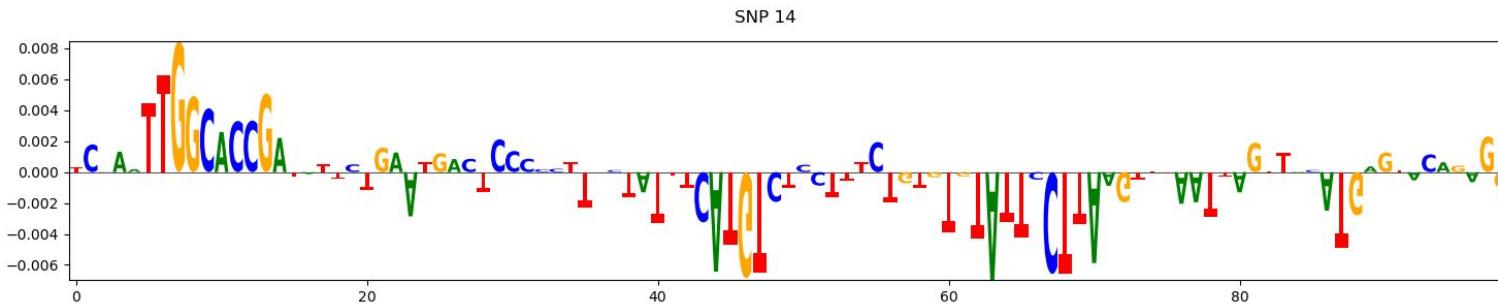
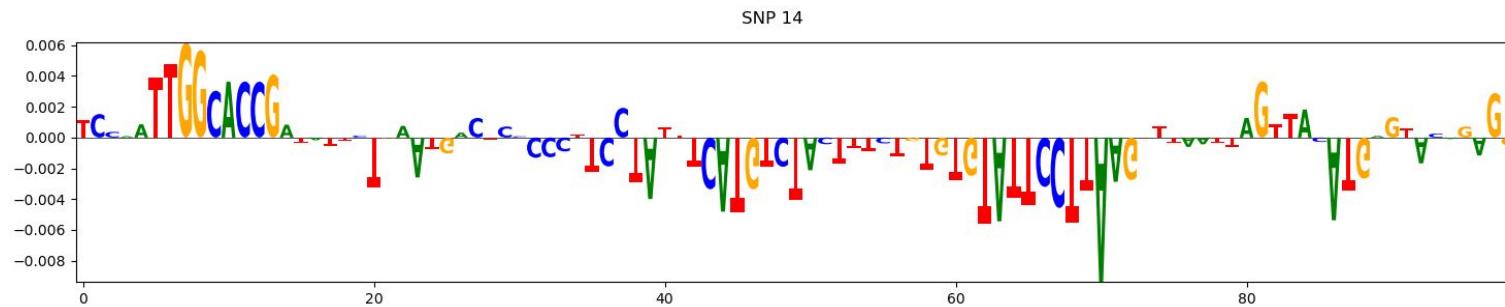
# SNP 13 - 2 week



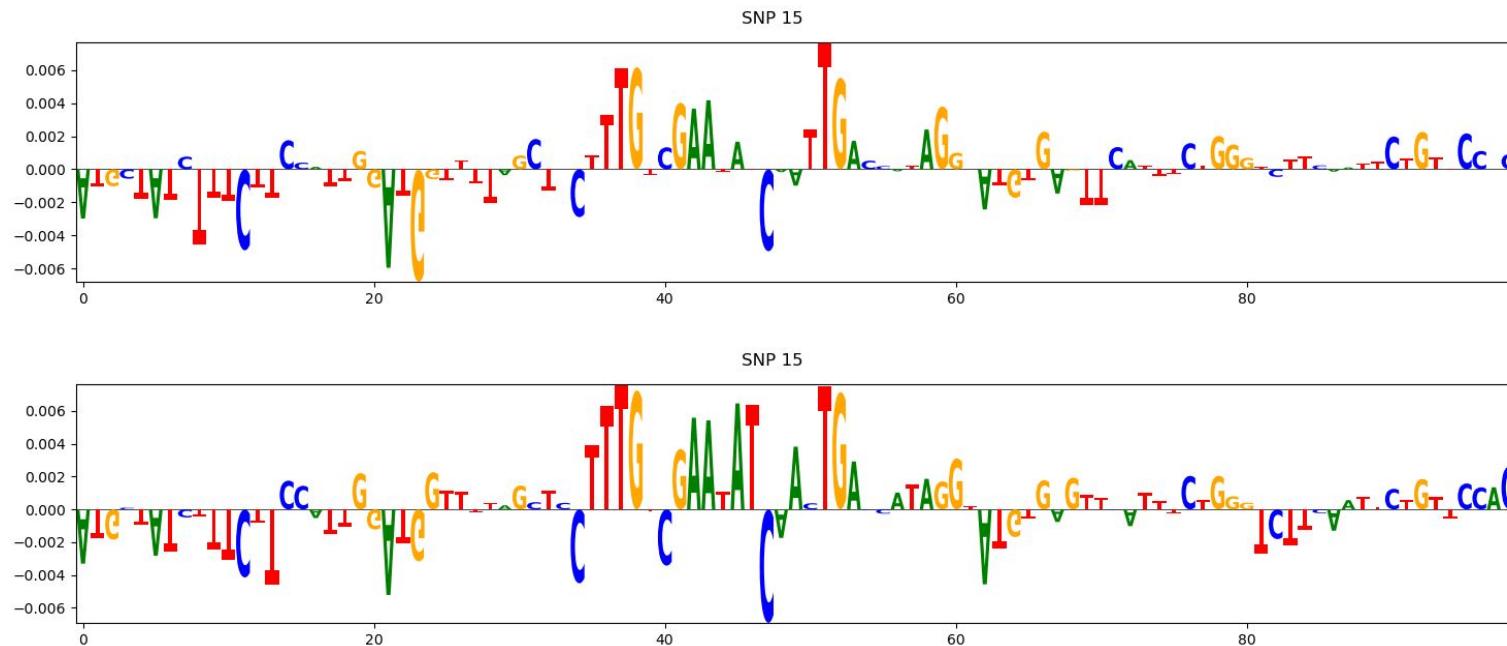
# SNP 14 - 8 week



# SNP 14 - 2 week



# SNP 15 - 8 week



# SNP 15 - 2 week

