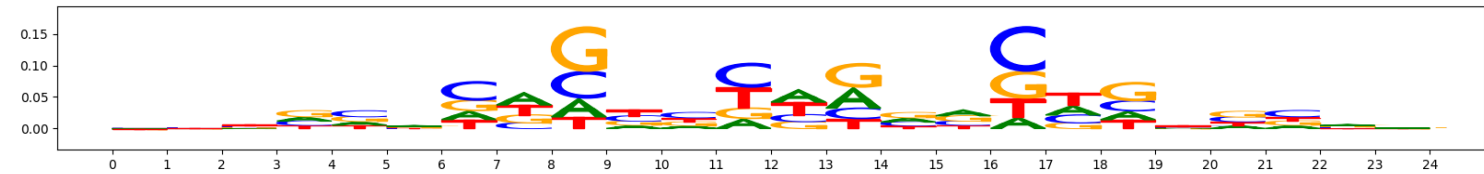


Bias factorized ChromBPNet training and quality check report

Preprocessing report

The image below should look closely like a Tn5 or DNase bias enzyme motif.



Bias model performance in peaks

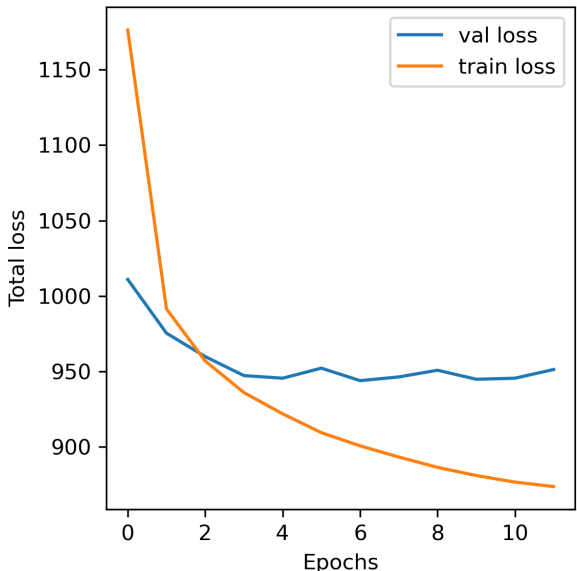
Counts Metrics: The pearsonr in peaks should be greater than -0.3 (otherwise the bias model could potentially be capturing AT bias). MSE (Mean Squared Error) will be high in peaks.

Profile Metrics: Median JSD (Jensen Shannon Divergence between observed and predicted) lower the better. Median norm JSD is median of the min-max normalized JSD where min JSD is the worst case JSD i.e JSD of observed with uniform profile and max JSD is the best case JSD i.e 0. Median norm JSD is higher the better. Both JSD and median norm JSD are sensitive to read-depth. Higher read-depth results in better metrics.

		peaks.pearsonr	peaks.mse
counts_metrics		-0.26979	10.435034
	peaks.median_jsd	peaks.median_norm_jsd	
profile_metrics	0.393967	0.366759	

Training report

The val loss (validation loss) will decrease and saturate after a few epochs.



ChromBPNet model performance in peaks

Counts Metrics: The pearsonr in peaks should be greater than 0.5 (higher the better). MSE (Mean Squared Error) will be low in peaks.

Profile Metrics: Median JSD (Jensen Shannon Divergence between observed and predicted) lower the better. Median norm JSD is median of the min-max normalized JSD where min JSD is the worst case JSD i.e JSD of observed with uniform profile and max JSD is the best case JSD i.e 0. Median norm JSD is higher the better. Both JSD and median norm JSD are sensitive to read-depth. Higher read-depth results in better metrics.

		peaks.pearsonr	peaks.mse
counts_metrics		0.697164	0.625264
	peaks.median_jsd	peaks.median_norm_jsd	
profile_metrics	0.34265	0.448128	

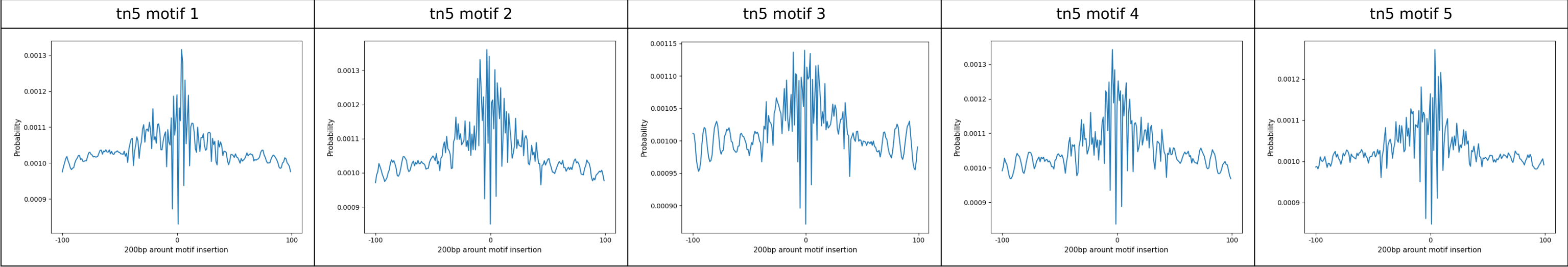
ChromBPNet marginal footprints on tn5 motifs

The marginal footprints are the response of the ChromBPNet no bias model to the heterogenous bias motifs. If the bias correction is complete the max of the profiles below should be below 0.003 on all the bias motifs.

For your convenience we calculate here the average of the max of the profiles: 0.001 And the model according to this is **corrected**

What to do if your model looks uncorrected (i.e max of profiles is greater than 0.003)?

Look at the motifs below captured by TFModisco and you should be able to see motifs that closely look like the bias motifs showing incomplete bias correction. This indicates that your bias model was not completely capturing the response of the bias. We recommend that you use a different pre-trained bias model. For more intuition on choosing the correct pre-trained model or retraining your bias model refer to [FAQ](#) section in wiki.



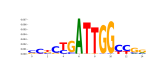
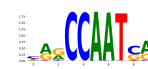
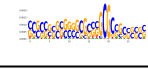
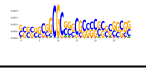
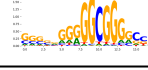
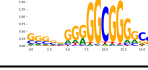
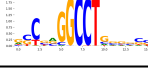
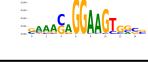
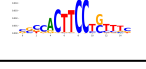
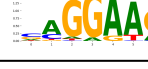
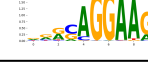
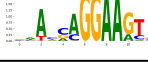
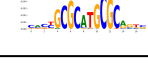
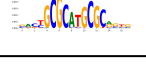
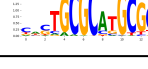
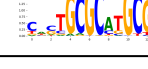
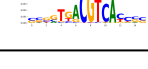
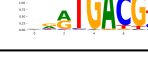
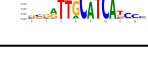
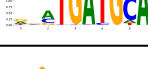
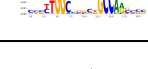
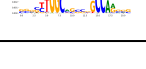
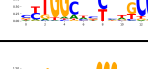
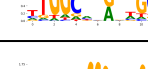
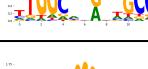
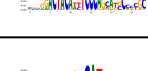
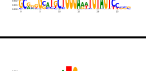
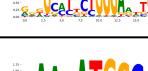
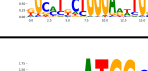
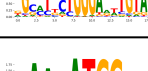
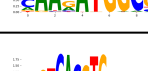
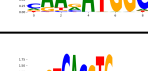
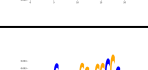
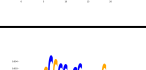
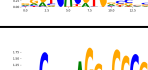
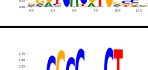
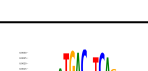
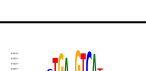
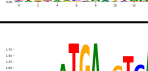
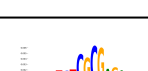
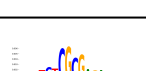
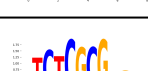
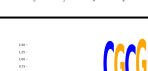
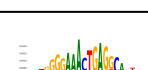
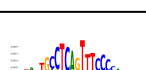
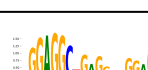
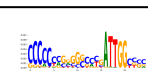
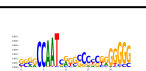
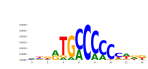
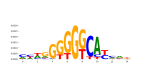
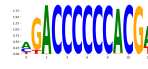
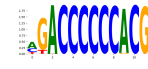
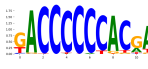
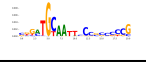
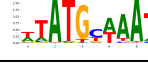
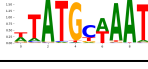
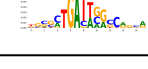
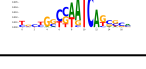
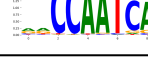
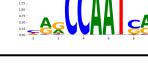
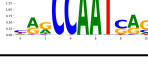
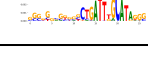
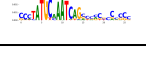
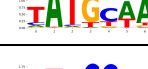
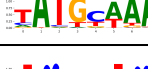
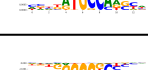
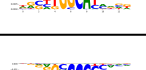
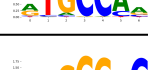
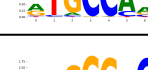
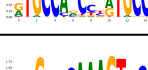
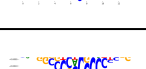
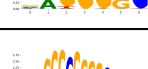
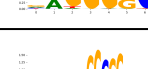
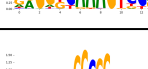
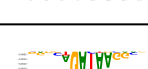
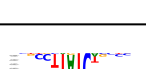
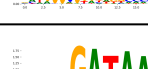
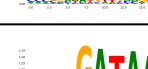
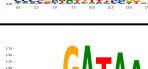
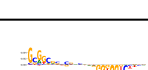
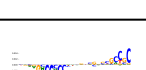
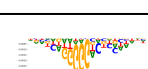
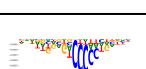
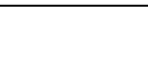
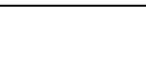
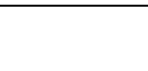
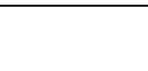
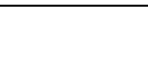
TFModisco motifs learnt from ChromBPNet after bias correction (chrombpnet_nobias.h5) model

TFModisco motifs generated from profile contribution scores of the ChromBPNet after bias correction model. cwm_fwd, cwm_rev are the forward and reverse complemented consolidated motifs from contribution scores in subset of random peaks. These CWM motifs should be free from any bias motifs and should contain only Transcription Factor (TF) motifs. For each of these motifs, we use TOMTOM to find the top-3 closest matches (match_0, match_1, match_2) from a database consisting of both MEME TF motifs and heterogenous enzyme bias motifs that we have repeatedly seen in our datasets. The qvals (qval0,qval1,qval2) should be low (< 0.0001) for most of the closest TF motif hits (i.e indicating that the closest match is the correct match) - this is also generally verifiable by eye as the closest match will look closely like the CWMs (atleast part of it in case of heterodimers). All the motifs in the list should look nothing like the enzyme bias motif.

What to do if you find an obvious bias motif in the list?

This indicates that your bias model was not completely capturing the response of the bias. We recommend that you use a different pre-trained bias model. For more intuition on choosing the correct pre-trained model or retraining your bias model refer to [FAQ](#) section in wiki.

pattern	NumSeqs	cwm_fwd	cwm_rev	match0	qval0	match0_logo	match1	qval1	match1_logo	match2	qval2	match2_logo
pos_0	5727			CTCF_MA0139.1	5.134900e-15		CTCF_HUMAN.H11MO.0.A	1.122370e-11		CTCF_MOUSE.H11MO.0.A	2.250130e-11	
pos_1	4601			SP1_MA0079.3	1.030370e-05		KLF3_HUMAN.H11MO.0.B	1.030370e-05		KLF3_MOUSE.H11MO.0.A	1.030370e-05	
pos_2	3736			GATA2_HUMAN.H11MO.0.A	8.844520e-03		GATA4_HUMAN.H11MO.0.A	1.436420e-02		GATA4_MOUSE.H11MO.0.A	1.436420e-02	
pos_3	2479			BACH2_MOUSE.H11MO.0.A	4.232450e-06		BACH2_HUMAN.H11MO.0.A	5.237360e-06		BACH2_MA1101.1	6.208500e-06	

pattern	NumSeqs	cwm_fwd	cwm_rev	match0	qval0	match0_logo	match1	qval1	match1_logo	match2	qval2	match2_logo
pos__4	1591			NFYB_HUMAN.H11MO.0.A	5.590780e-05		NFYB_MOUSE.H11MO.0.A	5.590780e-05		NFYA_HUMAN.H11MO.0.A	5.590780e-05	
pos__5	782			EHF_ETS_1	9.407300e-05		EHF_MA0598.2	9.407300e-05		ELF1_ETS_1	9.407300e-05	
pos__6	684			SP2_HUMAN.H11MO.0.A	3.893670e-04		SP2_MOUSE.H11MO.0.B	3.893670e-04		ZFX_MOUSE.H11MO.0.B	5.119710e-04	
pos__7	586			ETV4_MOUSE.H11MO.0.B	1.554980e-03		ERG_HUMAN.H11MO.0.A	2.713150e-03		EHF_HUMAN.H11MO.0.B	2.713150e-03	
pos__8	519			NRF1_HUMAN.H11MO.0.A	3.855420e-07		NRF1_MOUSE.H11MO.0.A	3.855420e-07		NRF1_NRF_1	6.437460e-06	
pos__9	483			ATF1_HUMAN.H11MO.0.B	3.962340e-05		CREB1_HUMAN.H11MO.0.A	3.962340e-05		CREB1_MOUSE.H11MO.0.A	3.962340e-05	
pos__10	461			CEBPG_HUMAN.H11MO.0.B	7.713130e-05		ATF4_HUMAN.H11MO.0.A	1.661730e-04		ATF4_MOUSE.H11MO.0.A	1.661730e-04	
pos__11	413			NFIC_HUMAN.H11MO.0.A	3.747750e-06		NFIA_HUMAN.H11MO.0.C	6.508160e-05		NFIA_MOUSE.H11MO.0.C	6.508160e-05	
pos__12	342			ZNF76_HUMAN.H11MO.0.C	9.867250e-14		ZN143_MOUSE.H11MO.0.A	1.890470e-12		ZN143_HUMAN.H11MO.0.A	4.907230e-12	
pos__13	322			TYY1_HUMAN.H11MO.0.A	1.597450e-07		TYY1_MOUSE.H11MO.0.A	3.222440e-06		YY1_MA0095.2	2.038790e-05	
pos__14	302			USF2_HUMAN.H11MO.0.A	4.977890e-06		USF2_MOUSE.H11MO.0.A	4.977890e-06		MITF_HUMAN.H11MO.0.A	2.394940e-05	
pos__15	138			CTCF_MOUSE.H11MO.0.A	1.164030e-01		CTCF_C2H2_1	1.164030e-01		CTCF_HUMAN.H11MO.0.A	1.164030e-01	
pos__16	121			ATF2_HUMAN.H11MO.0.B	2.006450e-04		ATF2_MOUSE.H11MO.0.A	2.006450e-04		JUND_MA0492.1	2.006450e-04	
pos__17	84			ZBTB33_MA0527.1	2.426620e-03		KAISO_HUMAN.H11MO.0.A	2.426620e-03		KAISO_MOUSE.H11MO.0.B	2.426620e-03	
pos__18	72			ZN770_HUMAN.H11MO.0.C	2.454330e-01		RORG_HUMAN.H11MO.0.C	2.454330e-01		RORG_MOUSE.H11MO.0.B	2.454330e-01	
pos__19	62			ZNF740_C2H2_1	1.655770e-02		ZNF740_C2H2_2	1.655770e-02		ZNF740_MA0753.1	1.655770e-02	
pos__20	51			INSM1_HUMAN.H11MO.0.C	2.363180e-01		INSM1_MOUSE.H11MO.0.C	2.363180e-01		HIC2_C2H2_1	2.363180e-01	
pos__21	50			Gabpa_MA0062.2	2.722880e-03		ELK1_HUMAN.H11MO.0.B	1.080280e-02		ERG_ETS_3	1.080280e-02	
pos__22	45			GLIS1_C2H2_1	1.305860e-01		GLIS1_MA0735.1	1.305860e-01		GLIS3_C2H2_1	1.305860e-01	
pos__23	45			POU3F1_MA0786.1	8.075680e-02		POU3F1_POU_1	8.075680e-02		POU3F2_MA0787.1	8.075680e-02	
pos__24	31			NFYA_MA0060.3	4.131910e-01		NFYB_HUMAN.H11MO.0.A	4.131910e-01		NFYB_MOUSE.H11MO.0.A	4.131910e-01	
pos__25	23			Pou2f2.mouse_POU_2	1.023160e-01		POU2F3_POU_1	1.023160e-01		POU5F1B_MA0792.1	1.023160e-01	
pos__26	23			Hic1.mouse_C2H2_2	1.495200e-04		Hic1_MA0739.1	1.495200e-04		Hic1.mouse_C2H2_1	3.239590e-04	
neg_0	199			ZBT7A_HUMAN.H11MO.0.A	7.966680e-03		ZBT7A_MOUSE.H11MO.0.B	7.966680e-03		HNF4A_nuclearreceptor_2	9.953220e-02	
neg_1	50			SP1_HUMAN.H11MO.0.A	4.635800e-03		SP2_HUMAN.H11MO.0.A	4.635800e-03		SP2_MOUSE.H11MO.0.B	4.635800e-03	
neg_2	45			GATA4_HUMAN.H11MO.0.A	1.033760e-02		GATA4_MOUSE.H11MO.0.A	1.033760e-02		GATA6_MOUSE.H11MO.0.A	1.033760e-02	
neg_3	34			GLI1_MOUSE.H11MO.0.C	2.284750e-02		GLI2_C2H2_1	3.056210e-02		GLI3_HUMAN.H11MO.0.B	6.660000e-02	
neg_4	30			SP1_MA0079.3	7.732270e-02		SP3_C2H2_1	7.732270e-02		SP3_MA0746.1	7.732270e-02	