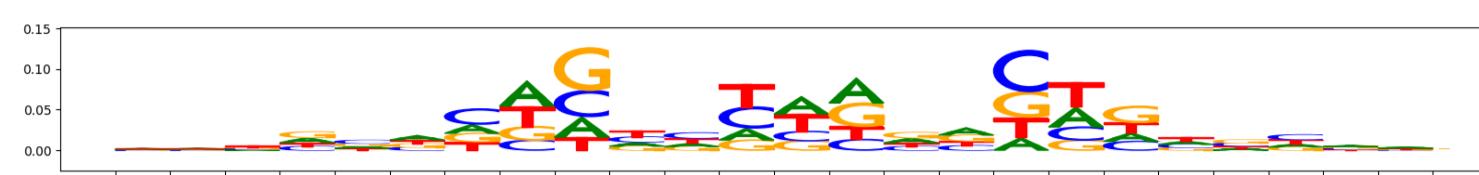


**Bias factorized ChromBPNet training and quality check report**

## 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.

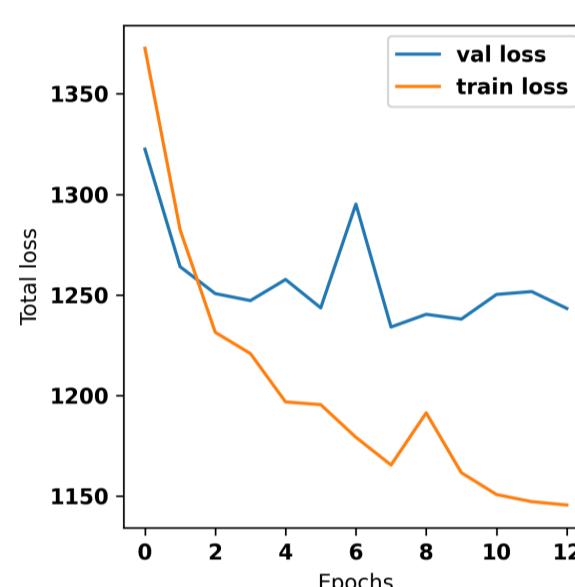
**Profile Metrics:** Median JSD (Jensen Shannon Divergence between observed and predicted) lower the better. Median norm JSD is median of the min-max normalized JSD where min JSD is the worst case JSD i.e JSD of observed with uniform profile and max JSD is the best case JSD i.e 0. Median norm JSD is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

**What to do if your pearsonr in peaks is less than -0.3?** In the range of -0.3 to -0.5 please be wary of your chrombpnet\_wo\_bias.h5 TFModisco results showing lots of GC rich motifs (> 3 in the top-10). If this is not the case you can continue using the chrombpnet\_wo\_bias.h5. If you end up seeing a lot of GC rich motifs it is likely that bias model has learnt a different GC distribution than your GC-content in peaks. If you are transferring a bias model from a different sample you can consider using a different bias model or [training a bias model](#) for this sample. If you have trained a bias model for this sample and encounter this you might have to increase the bias\_threshold\_factor argument input to the *chrombpnet bias pipeline* or *chrombpnet bias train* command used in training the bias model and retrain a new bias model. For more intuition about this argument refer to the [FAQ](#) section in wiki. If the value is less than -0.5 the pipeline will automatically throw an error.

	<b>peaks.pearsonr</b>	<b>peaks.mse</b>
<b>counts_metrics</b>	-0.158301	8.777323
	<b>peaks.median_jsd</b>	<b>peaks.median_norm_jsd</b>
<b>profile metrics</b>	0.480217	0.258033

## **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.

	<b>peaks.pearsonr</b>	<b>peaks.mse</b>
<b>counts_metrics</b>	0.641842	0.527226
	<b>peaks.median_jsd</b>	<b>peaks.median_norm_jsd</b>
<b>profile_metrics</b>	0.438246	0.314343

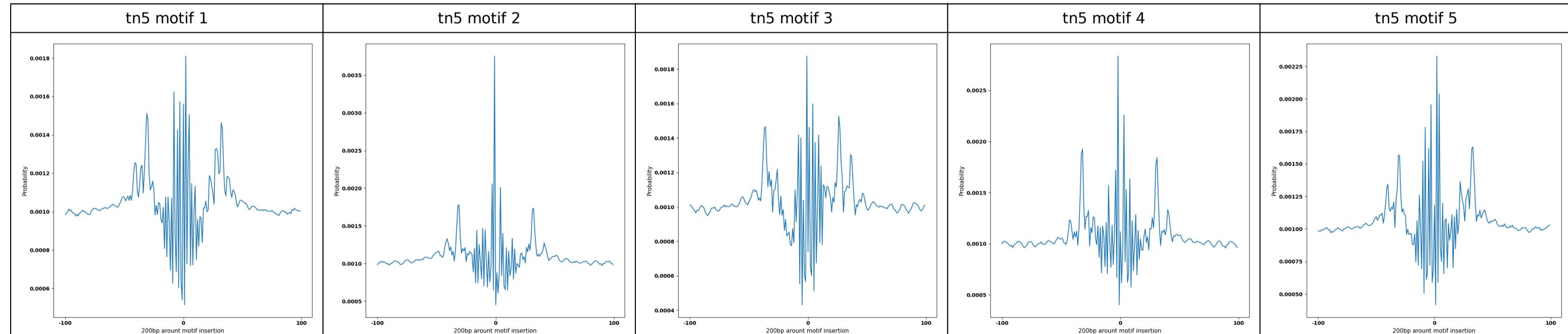
## ChromBPNet marginal footprints on tn5 motifs

The marginal footprints are the response of the ChromBPNet no bias model to the heterogeneous 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: 0.003 And the model according to this is **uncorrected**

**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.

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

pattern	NumSeqs	cwm_fwd	cwm_rev	match0	qval0	match0_logo	match1	qval1	match1_logo	match2	qval2	match3
pos_0	2809			HNF4G_MA0484.1	9.077310e-06		NR1H3_MOUSE.H11MO.0.A	3.592490e-05		Hnf4a.mouse_nuclearreceptor_1	3.592490e-05	
pos_1	2111			CTCF_MA0139.1	2.616790e-11		CTCF_HUMAN.H11MO.0.A	1.248500e-09		CTCF_MOUSE.H11MO.0.A	1.279240e-08	
pos_2	1600			KLF12_HUMAN.H11MO.0.C	1.363870e-04		SP3_HUMAN.H11MO.0.B	1.363870e-04		SP3_MOUSE.H11MO.0.B	1.363870e-04	
pos_3	1519			CEBPA_MA0102.3	2.872760e-09		CEBPA_MOUSE.H11MO.0.A	7.953850e-08		CEBPB_MOUSE.H11MO.0.A	7.953850e-08	
pos_4	1243			RREB1_MA0073.1	1.000000e+00		BRAC_MOUSE.H11MO.0.B	1.000000e+00		None	NaN	
pos_5	1038			NFIA_HUMAN.H11MO.0.C	2.653820e-04		NFIA_MOUSE.H11MO.0.C	2.653820e-04		NFIC_HUMAN.H11MO.0.A	2.653820e-04	
pos_6	1023			SP2_HUMAN.H11MO.0.A	1.965730e-06		SP2_MOUSE.H11MO.0.B	1.965730e-06		SP1_HUMAN.H11MO.0.A	1.522470e-05	
pos_7	1021			NFIX_MA0671.1	3.434730e-02		NFIX_NFI_2	3.434730e-02		NFIA_MA0670.1	1.893120e-01	
pos_8	999			FOXM1_HUMAN.H11MO.0.A	3.754490e-06		FOXM1_MOUSE.H11MO.0.B	3.754490e-06		FOXA1_HUMAN.H11MO.0.A	6.466380e-03	
pos_9	882			SP3_HUMAN.H11MO.0.B	2.861780e-04		SP3_MOUSE.H11MO.0.B	2.861780e-04		SP4_HUMAN.H11MO.0.A	2.861780e-04	
pos_10	807			NR2F1_MA0017.2	1.997360e-04		NR2F1_nuclearreceptor_4	1.997360e-04		RARG_nuclearreceptor_3	1.997360e-04	
pos_11	710			HNF1B_MA0153.2	3.550220e-07		HNF1B_homeodomain_1	3.550220e-07		HNF1B_homeodomain_2	7.100450e-07	
pos_12	618			ZN331_HUMAN.H11MO.0.C	7.425500e-04		TN5_8	4.222260e-03		TN5_7	1.636020e-02	
pos_13	520			NFIC_HUMAN.H11MO.0.A	1.472160e-01		EWSR1-FLI1_MA0149.1	1.472160e-01		PATZ1_HUMAN.H11MO.0.C	2.059260e-01	
pos_14	515			IRF3_HUMAN.H11MO.0.B	7.732750e-02		IRF3_MOUSE.H11MO.0.A	7.732750e-02		MAZ_HUMAN.H11MO.0.A	1.557910e-01	
pos_15	495			MAZ_HUMAN.H11MO.0.A	1.334550e-01		MAZ_MOUSE.H11MO.0.A	1.334550e-01		VEZF1_HUMAN.H11MO.0.C	1.334550e-01	
pos_16	447			PRDM6_HUMAN.H11MO.0.C	8.599690e-02		ZNF384_MA1125.1	8.599690e-02		STAT1_MOUSE.H11MO.0.A	2.152680e-01	
pos_17	426			CUX2_MOUSE.H11MO.0.C	1.357960e-02		HNF6_HUMAN.H11MO.0.B	1.357960e-02		HNF6_MOUSE.H11MO.0.A	1.357960e-02	
pos_18	381			SP5_MOUSE.H11MO.0.C	3.127320e-01		MAZ_HUMAN.H11MO.0.A	3.127320e-01		MAZ_MOUSE.H11MO.0.A	3.127320e-01	
pos_19	336			DLX1_homeodomain_1	1.000000e+00		Dlx1_MA0879.1	1.000000e+00		None	NaN	
pos_20	323			ZN331_HUMAN.H11MO.0.C	4.319510e-01		TN5_1	4.319510e-01		NR2C1_HUMAN.H11MO.0.C	4.319510e-01	
pos_21	290			NFAC4_HUMAN.H11MO.0.C	4.404050e-01		NFAC4_MOUSE.H11MO.0.C	4.404050e-01		NFAC2_HUMAN.H11MO.0.B	1.000000e+00	
pos_22	289			CTCF_HUMAN.H11MO.0.A	3.458410e-09		CTCF_MA0139.1	3.723560e-09		CTCF_MOUSE.H11MO.0.A	5.109270e-09	
pos_23	286			RREB1_MA0073.1	1.000000e+00		KLF9_MA1107.1	1.000000e+00		BRAC_MOUSE.H11MO.0.B	1.000000e+00	
pos_24	262			RREB1_MA0073.1	5.414900e-03		RUNX2_HUMAN.H11MO.0.A	7.359730e-01		KLF9_MA1107.1	7.359730e-01	
pos_25	212			ZN467_HUMAN.H11MO.0.C	1.236360e-07		VEZF1_HUMAN.H11MO.0.C	1.236360e-07		ZNF263_MA0528.1	1.236360e-07	
pos_26	209			NRF1_MOUSE.H11MO.0.A	5.730160e-06		NRF1_HUMAN.H11MO.0.A	6.558620e-06		NRF1_NRF_1	1.469990e-05	
pos_27	186			SIX2_HUMAN.H11MO.0.A	1.000000e+00		None	NaN		None	NaN	
pos_28	167			ZN502_HUMAN.H11MO.0.C	8.306670e-01		NFAC4_HUMAN.H11MO.0.C	1.000000e+00		NFAC4_MOUSE.H11MO.0.C	1.000000e+00	
pos_29	159			FOXJ2_HUMAN.H11MO.0.C	4.192720e-02		FOXJ2_MOUSE.H11MO.0.C	4.192720e-02		FOXM1_HUMAN.H11MO.0.A	4.192720e-02	
pos_30	157			FOXC1_forkhead_2	7.639470e-04		FOXD2_forkhead_1	7.639470e-04		FOXD3_forkhead_1	7.639470e-04	
pos_31	156			RUNX2_MOUSE.H11MO.0.A	4.368480e-01		RUNX1_MA0002.2	4.368480e-01		RUNX1_MOUSE.H11MO.0.A	4.368480e-01	
pos_32	153			SP1_HUMAN.H11MO.0.A	1.295390e-01		KLF3_HUMAN.H11MO.0.B	1.295390e-01		KLF3_MOUSE.H11MO.0.A	1.295390e-01	
pos_33	152			MAZ_HUMAN.H11MO.0.A	1.642010e-04		MAZ_MOUSE.H11MO.0.A	1.642010e-04		VEZF1_HUMAN.H11MO.0.C	2.105850e-04	
pos_34	130			RARA_HUMAN.H11MO.0.A	4.786180e-01		RXRA_MOUSE.H11MO.0.A	7.994380e-01		RARA		

pattern	NumSeqs	cwm_fwd	cwm_rev	match0	qval0	match0_logo	match1	qval1	match1_logo	match2	qval2	match2_logo
pos_55	28			HNF4G_MA0484.1	2.061890e-01		HNF4A_MOUSE.H11MO.0.A	2.061890e-01		NR1H2+RXRA_MA0115.1	2.061890e-01	
pos_56	28			CUX2_CUT_1	1.000000e+00		BRAC_MOUSE.H11MO.0.B	1.000000e+00		PBX2_HUMAN.H11MO.0.C	1.000000e+00	
pos_57	24			RARA_nuclearreceptor_4	1.000000e+00		NR2F1_nuclearreceptor_1	1.000000e+00		NR2C2_MA0504.1	1.000000e+00	
pos_58	22			RUNX2_MOUSE.H11MO.0.A	4.440880e-01		RUNX1_MA0002.2	4.440880e-01		RUNX1_MOUSE.H11MO.0.A	4.440880e-01	
pos_59	20			ZSC22_HUMAN.H11MO.0.C	1.000000e+00		ZN281_HUMAN.H11MO.0.A	1.000000e+00		NGN2_MOUSE.H11MO.0.C	1.000000e+00	
neg_0	495			None	NaN		None	NaN		None	NaN	
neg_1	278			SP1_HUMAN.H11MO.0.A	6.069050e-04		SP2_HUMAN.H11MO.0.A	6.069050e-04		SP2_MOUSE.H11MO.0.B	6.069050e-04	
neg_2	122			ZBT7A_HUMAN.H11MO.0.A	1.430650e-01		ZBT7A_MOUSE.H11MO.0.B	1.430650e-01		PAX5_MA0014.3	1.430650e-01	
neg_3	103			None	NaN		None	NaN		None	NaN	
neg_4	92			FOXO4_HUMAN.H11MO.0.C	1.677410e-01		FOXO4_MOUSE.H11MO.0.C	1.677410e-01		JUNB_HUMAN.H11MO.0.A	3.111080e-01	
neg_5	91			RELB_HUMAN.H11MO.0.C	7.903270e-02		RELB_MOUSE.H11MO.0.C	7.903270e-02		ZBT7A_HUMAN.H11MO.0.A	4.295090e-01	
neg_6	81			POU3F2_POU_1	1.000000e+00		CUX2_CUT_1	1.000000e+00		POU3F3_POU_2	1.000000e+00	
neg_7	77			AP2C_MOUSE.H11MO.0.A	2.882630e-02		TFAP2B_MA0811.1	2.882630e-02		TFAP2B_TFAP_1	2.882630e-02	
neg_8	71			None	NaN		None	NaN		None	NaN	
neg_9	70			IRF3_HUMAN.H11MO.0.B	1.227820e-01		IRF3_MOUSE.H11MO.0.A	1.227820e-01		ZFP82_HUMAN.H11MO.0.C	1.227820e-01	
neg_10	70			Foxg1.mouse_forkhead_2	1.000000e+00		Foxk1.mouse_forkhead_1	1.000000e+00		SPDEF_ETS_2	1.000000e+00	
neg_11	58			Z354A_HUMAN.H11MO.0.C	1.000000e+00		POU3F2_MA0787.1	1.000000e+00		POU3F2_POU_2	1.000000e+00	
neg_12	56			POU2F2_POU_2	2.780000e-01		POU2F1_POU_2	2.780000e-01		PO3F2_MOUSE.H11MO.0.A	2.780000e-01	
neg_13	54			FOXB1_forkhead_2	7.498770e-01		POU4F1_MA0790.1	7.498770e-01		POU4F1_POU_1	7.498770e-01	
neg_14	52			CTCF_MA0139.1	4.592450e-02		CTCF_HUMAN.H11MO.0.A	4.592450e-02		CTCF_MOUSE.H11MO.0.A	5.934680e-02	
neg_15	45			NR1D1_HUMAN.H11MO.0.B	3.954760e-03		NR1D1_MOUSE.H11MO.0.A	3.954760e-03		NR1D2_MOUSE.H11MO.0.A	4.325010e-03	
neg_16	42			ATF7_MA0834.1	1.000000e+00		ATF7_bZIP_1	1.000000e+00		Tp53.mouse_p53l_2	1.000000e+00	
neg_17	36			FOXA3_HUMAN.H11MO.0.B	2.050380e-01		FOXA3_MOUSE.H11MO.0.A	2.050380e-01		FOXA2_MOUSE.H11MO.0.A	3.995720e-01	
neg_18	35			GBX2_homeodomain_2	3.461910e-01		EN1_homeodomain_2	3.461910e-01		EN1_homeodomain_4	3.461910e-01	
neg_19	32			RREB1_MA0073.1	8.220130e-02		EGR2_HUMAN.H11MO.0.A	1.879240e-01		KLF5_HUMAN.H11MO.0.A	1.879240e-01	
neg_20	28			SP2_MA0516.1	1.936990e-03		SP1_MA0079.3	2.459960e-03		SP3_HUMAN.H11MO.0.B	2.910330e-03	
neg_21	28			None	NaN		None	NaN		None	NaN	