modERN analysis call

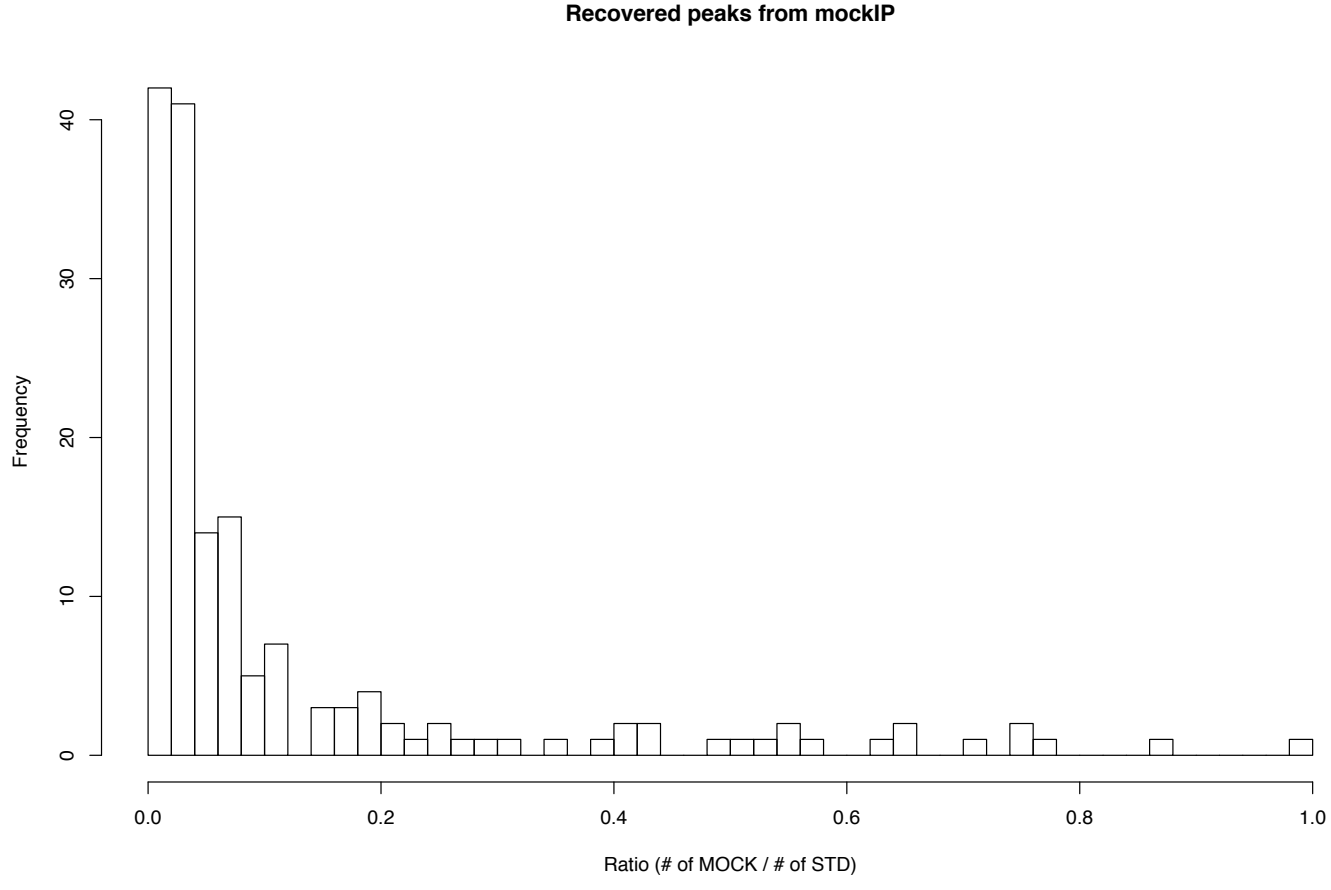
Lijia Ma

03/30/2017

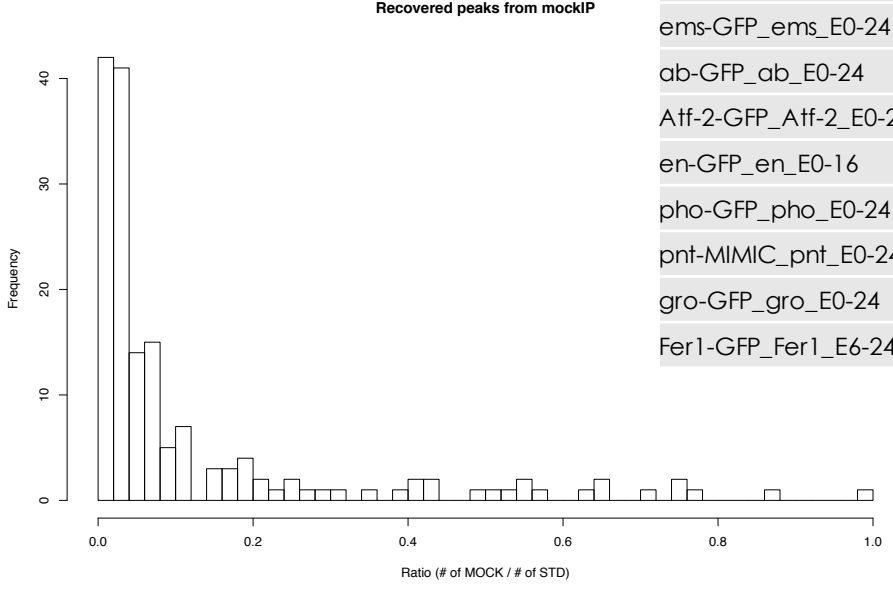
Re-cap

- In fly, most dataset can recover less than 10% peaks of the standard pipeline *(STD)* when using mockIP as input *(MOCK)*
- The possible explanation could be a mixture of:
  - **open chromatin regions were enriched in the mock-IP experiment?** For each particular factor, some of these regions might have real enrichment while some are not.
  - **The Goat-V antibody has significant non-specificity issue?** WB/IF shows good results. Should also see how the worm *MOCK* peaks look like.
- We want to have a *revised-STD* pipeline
  - Still follow the *STD* pipeline, but adjust the cutoff for calling a peak in order to maximize the number of peaks that overlaps with the *MOCK* pipeline.
  - We hope the cutoff could be uniform across datasets so that the tagging strategy and the *STD* pipeline can be widely applied.

Most of modERN data recovered less than 10% of *STD* peaks from the *MOCK* pipeline

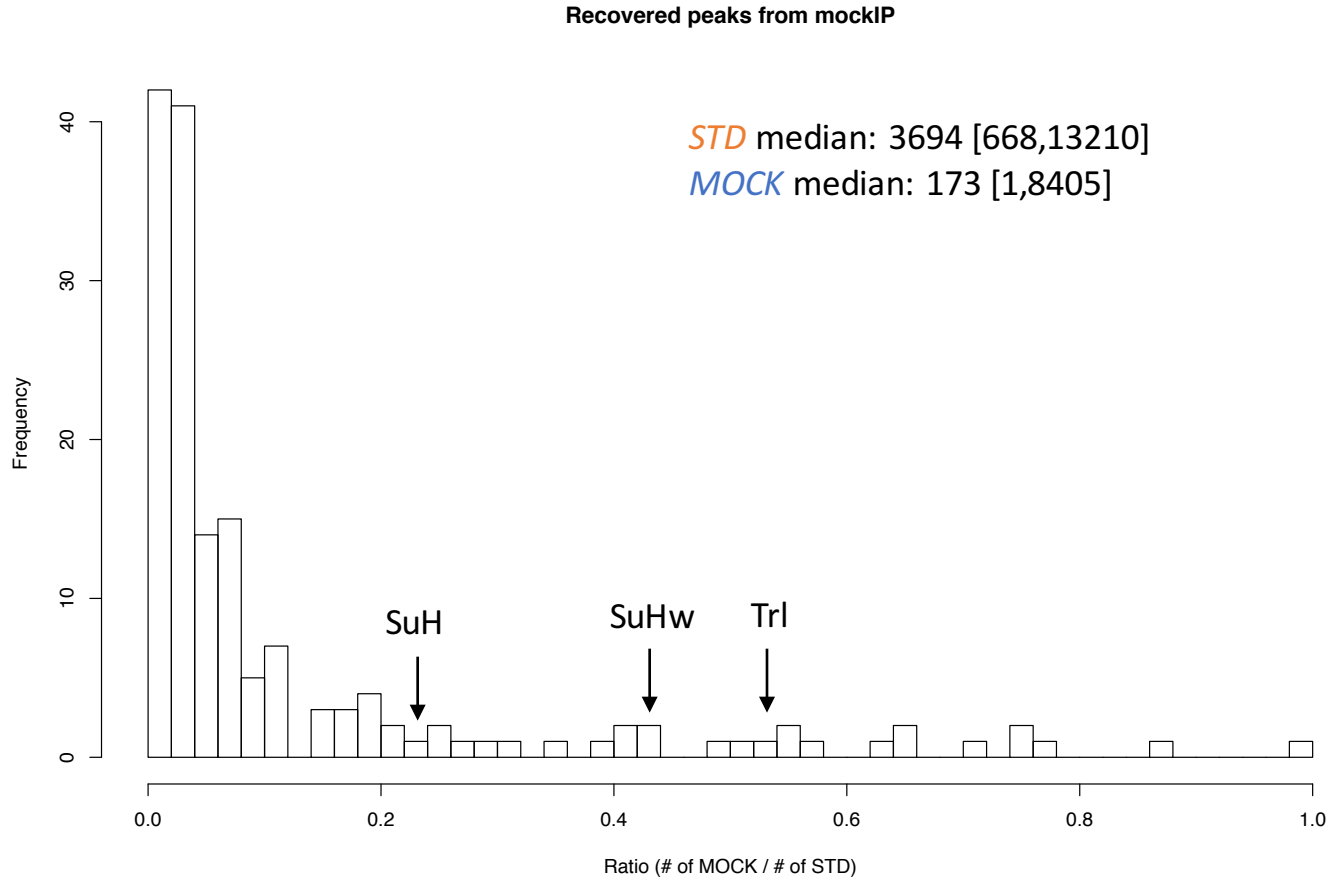


Picked up ten datasets representing different peak recovery rate to test



Dataset	# STD	# MOCK	#(MOCK_in_STD)	Ratio(MOCK/STD)	Ratio(MOCK_in_STD/MOCK)
tio-GFP_tio_E4-24	3733	21	7	0.006	0.333
disco-GFP_disco_E0-24	2250	14	7	0.006	0.500
ems-GFP_ems_E0-24	3963	33	6	0.008	0.182
ab-GFP_ab_E0-24	1546	70	25	0.045	0.357
Atf-2-GFP_Atf-2_E0-24	3067	34	11	0.011	0.324
en-GFP_en_E0-16	2259	32	10	0.014	0.313
pho-GFP_pho_E0-24	4204	78	38	0.019	0.487
pnt-MIMIC_pnt_E0-24	3543	95	25	0.027	0.263
gro-GFP_gro_E0-24	4533	484	244	0.107	0.504
Fer1-GFP_Fer1_E6-24	7477	2179	1864	0.291	0.855

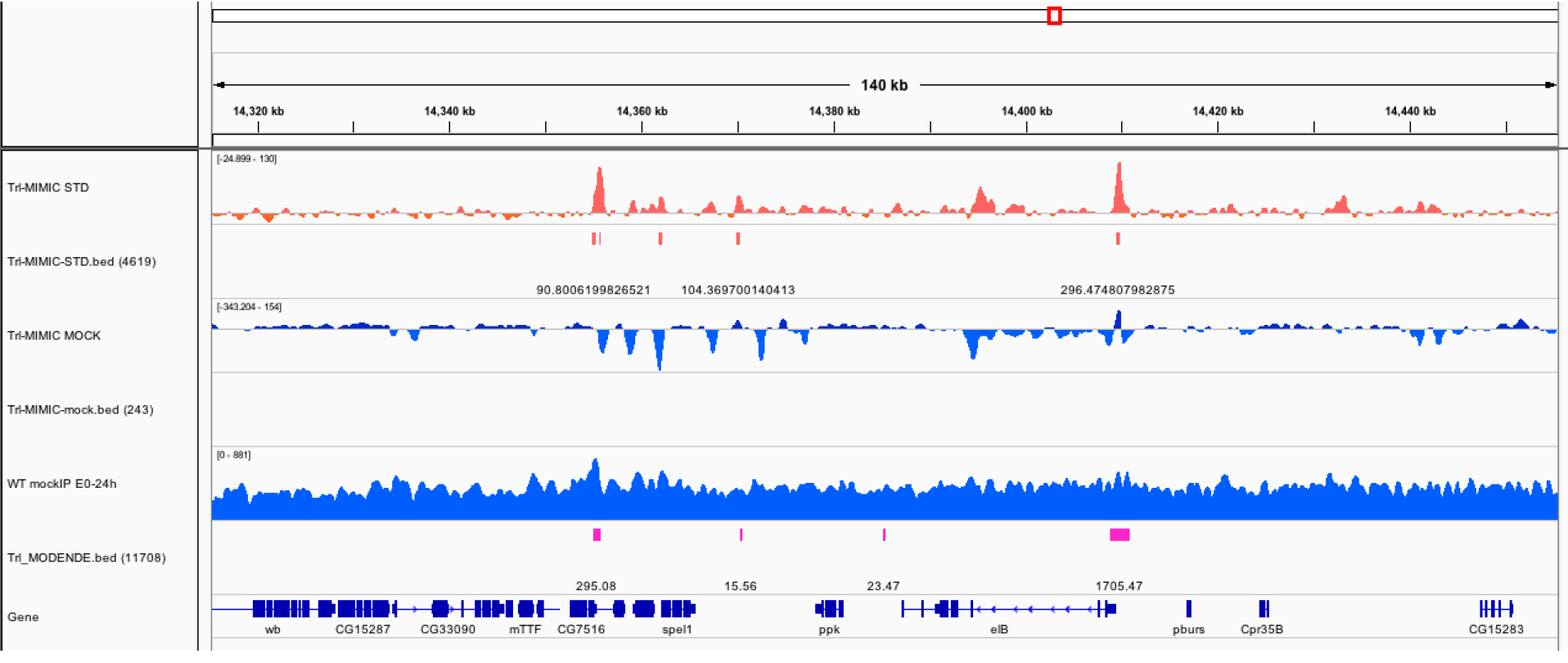
Picked up three candidates with modENCODE matched data to double check the *MOCK* pipeline



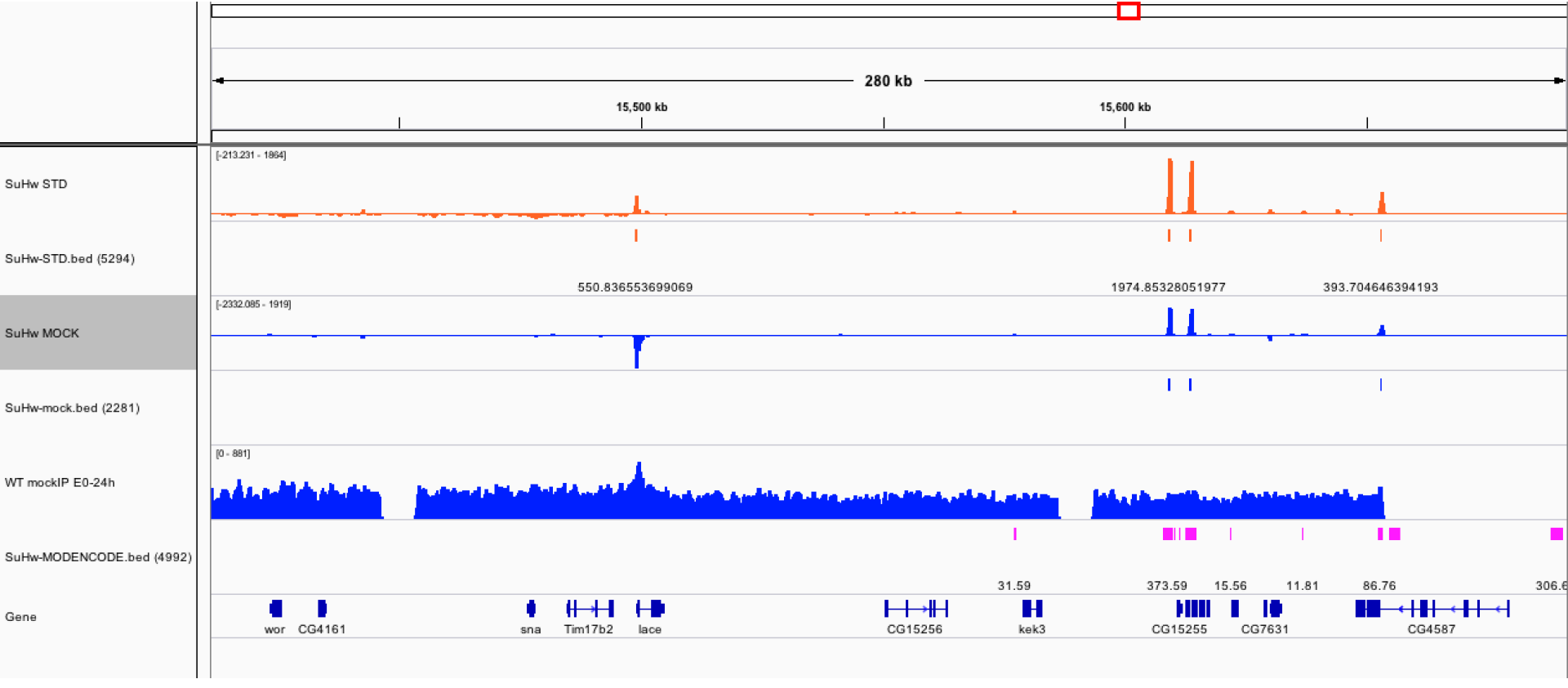
Overlap between modENCODE, *STD* and *MOCK* peaks varied.

modENCODE	modERN	modENCODE uniq	modENCODE common	modERN uniq	modERN common	# modENCODE	% modENCODE common	# modERN	% B modERN
SuH	SuH-GFP-MOCK	9529	18	62	22	9547	0.2%	84	21.4%
SuH	SuH-GFP-STD	8172	1375	2080	1486	9547	16.8%	3566	38.6%
SuHw	SuHw-GFP-MOCK	3330	1662	235	2046	4992	49.9%	2281	72.9%
SuHw	SuHw-GFP-STD	3202	1790	3115	2179	4992	55.9%	5294	33.8%
Trl	Trl-MIMIC-MOCK	11627	81	159	84	11708	0.7%	243	33.3%
Trl	Trl-MIMIC-STD	10325	1383	2919	1700	11708	13.4%	4619	29.9%

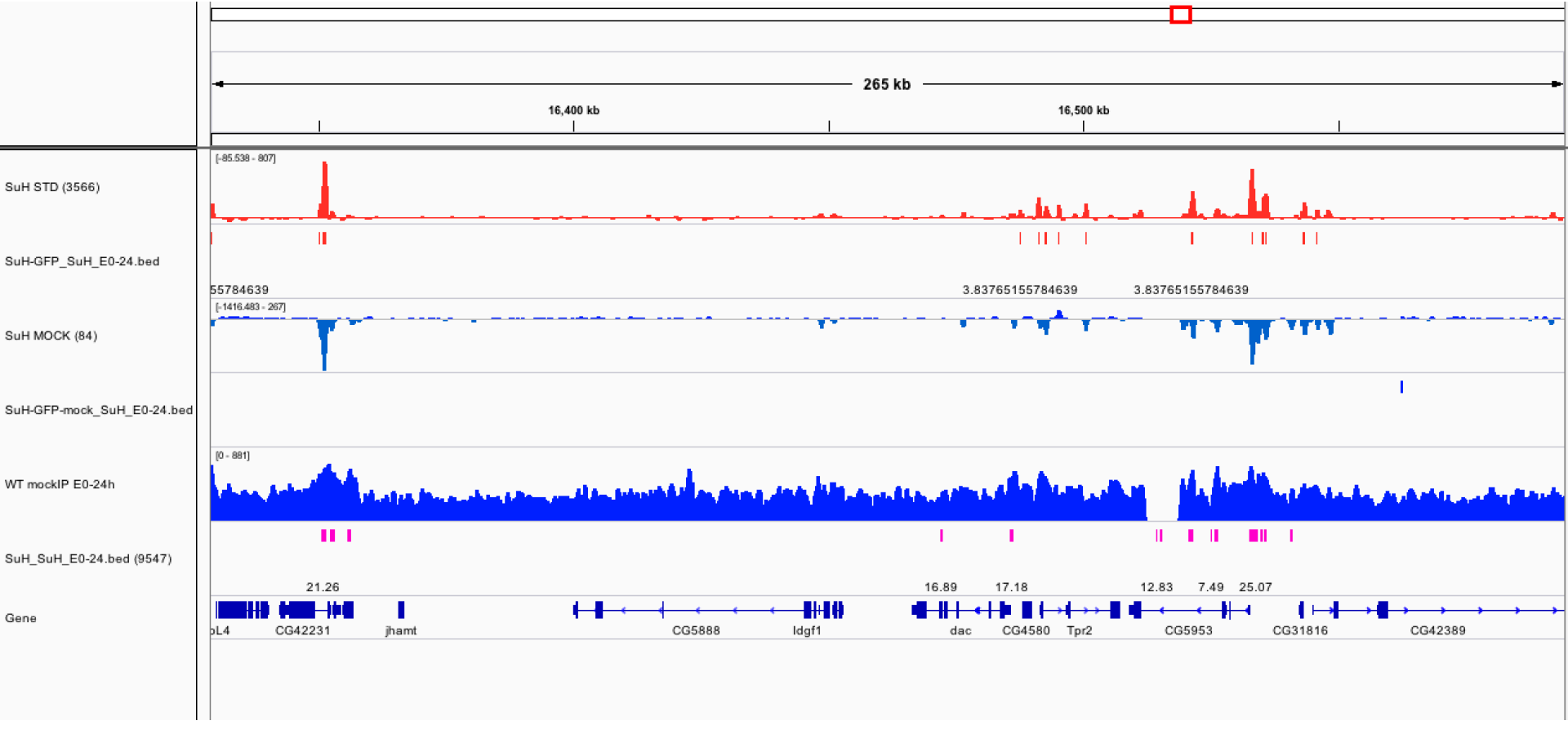
Trl: *MOCK* removed most of modENCODE and *STD* peaks



Su(Hw): Good consistency



Su(H): *MOCK* removed most of modENCODE and *STD* peaks



Open question

- Should we just go ahead with the *STD* pipeline?