modERN Call

White Lab

02-11-2016

modERN fly ChIP-seq datasets

17 new since last report, 48/80 for FY.

194 lines complete data sets

- ab, Abd-B, achi, acj6[!], Antp-mimic, Atf-2, Atf3, Bab2, br, brk, Bteb2, btn, cad, Camta-mimic, cato, CG10462, CG10565, CG10631, CG11398, CG11762, CG11902, CG12236, CG12744, CG13624, CG14965, CG1602, CG1620, CG1647, CG16863, CG1792, CG18476, CG2120, CG32264, CG31627, CG3163, CG33213, CG3838, CG4402(aka CG34406), CG4282, CG4820, CG4854, CG5204, CG6792(aka Plzf), CG7786, CG8089, CG8319, CG8944, CG9305, CG9609, CG9876, CHES-1-like, chif*, chn, cic, cnc, corto, crebA-mimic, crc-mimic, crg-1, crp, cyc, da, dac, Dad, Dfd, Dif, Dip3, disco, dl*, dm, dpn-mimic, dsf, dsx, EcR, Eip75B-MiMIC, Eip78C, Eip93F, emc, ems, en, ERR, esn, E(spl)m3, Ets21C, Ets65A(ets3), Ets97D, E(var)3-9, eve, exd, ey, eyg, foxo, ftz-f1, fru-mimic, fu2, GATAd, gcm2, gfzf, grh, grn, gro, h, her, HLH54F, HLHm7, HmgD, Hmx-MiMIC, Hnf4, Hr38, Hr39, Hr4[!], Hr46, Hr51, Hr78, Hr83, Hr96, hsf, insv, jim, jing, Jra*, jumu-mimic, kn, lbe, lilli , lola*, Lpt, luna-mimic, lz, Mad, maf-s, mam-mimic, Max, med, Mes4, Met, Mio, Mnt, mod(mdg4), myb, N, NC2alpha, NC2beta, Neu2, NK7.1, nmo, OdsH[!], ovo, p53-mimic, pb, pdm3, pdp1(&mimic), pho, Pif1A, Pif1B, pnt-MiMIC, psq[!], pum-mimic, Rel, repo, sage, sens, shn, side-mimic, sima, slou, slp2, Smox, Sox102F, Sox14, Sox15, Sry-delta, Stat92E, Su(H), Su(var)3-7, sv, svp, tai, tio, tll, toe, topi, trh, trl-mimic, tup, tx, usp, vfl, vri, Vsx2, woc, Zfh2, ZnT49B.

- 14 from modENCODE
- 180 from modERN

* multiple isoforms run

! Multiple time-points collected

XX: verifying data with rerun.

Current fly ChIP-seq

MiMIC being expanded for ChIP-seq (ADD1, CG16779, CG9727, dsx†, foxo, Snoo)

87 tagged lines being expanded for ChIP-seq, 48 new lines from last report.

Target Stages:

Embryo: Az2, bcd, Beaf-32, CG10543, CG10654, CG11617, CG11723, CG12104, CG12155‼, CG12942, CG13123, CG16815, CG13775, CG15073, CG15479(aka Mabi), CG15601, CG16629, CG17568, CG1832, CG18011, CG18764, CG30403, CG3065, CG31388, CG4617, CG4424, CG5245, CG6683, CG7556, CG8388, CG9727, CG9817, CG9883‼, Chrac-16, Creb17-A, dmrt93B, dpn, E(bx), E(spl)m-beta, E(spl)m-gamma, fd96Cb, Fer1, fkh, HmgZ, ind, kay‼, Kr‼, meics, oc, odd, org-1, pita, Six4, Sp1, su(Hw), su(var)2-10-RH, TFAM, tin, unpg, Usf, Vsx1, Xbp1, YL-1.

W3L: CG7839, HLH106, net, rgr.

WPP: bsh, caup‼, CG18619, CG32006, CG30431, CG33017, CG4318, CG5245, CG7368, CG9139, E(z), fd85E, Nfl, salr.

AM: CG33017, CG3919, CG7045, Clk.

AF: CG30403, tj.

Repeat: Abd-B(for sue), cic, crp, dac, grn, insv, jing, nmo-small, tio

†: Collect at different time-point if fail

‼: failed once

07: Chromatin extracted
05: IPed
30: Librariied
- 17: @ HGAC
- 00: Processing
- *04*: awaiting reps or PCR

DCC

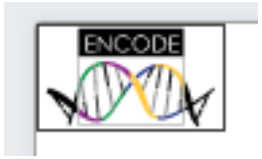
- Most current datasets for dm3, ce10 and ce11, have been uploaded
  - > 6k files.
- Currently reprocessing data with dm6
- Look into a threshold for RSC/NSC values
- Finish uploading Ab comparisons for validation

DCC

- Items from Last weeks call:
  - Icons added to pages.
    - Do we want one?

current icon

modENCODE



ENCODE

  - Anyone attending TAGC?
    - Esther willing to present poster describing available data

  - ENCODE epitope tagging standards:
    - DCC wants official answer if we'll comply with ENCODE line validation.
    - Would require IP/WB of each line.

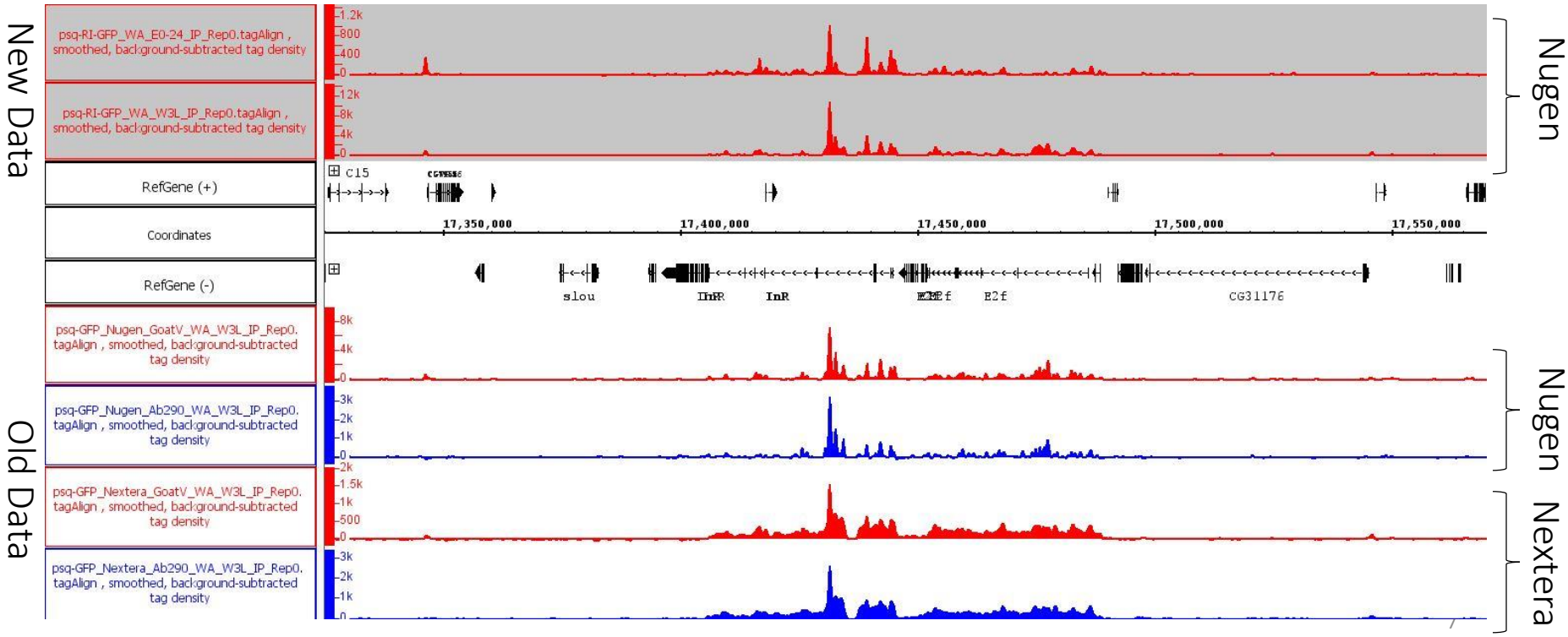
Control Experiments

- Precipitated by increase in overlapped peaks in new datasets

- Possible causes:
  - Library kit effect
  - GoatV effect
  - Comparison of TF antibody vs GFP antibody
  - Background peaks

Library kit and GoatV controls

- Not caused by Library or goatV
  - Ab290 (blue tracks) show similar traces to goatV (red tracks)



TF antibody vs GFP antibody

Difficult to tell with the small number of datasets with same factor and time point

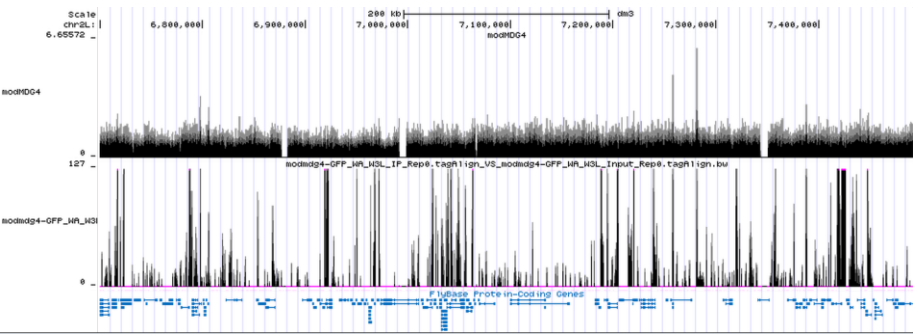
- mod(MDG4)
  - GFP: 5580
  - Iso1: 235
  - 207 in common.

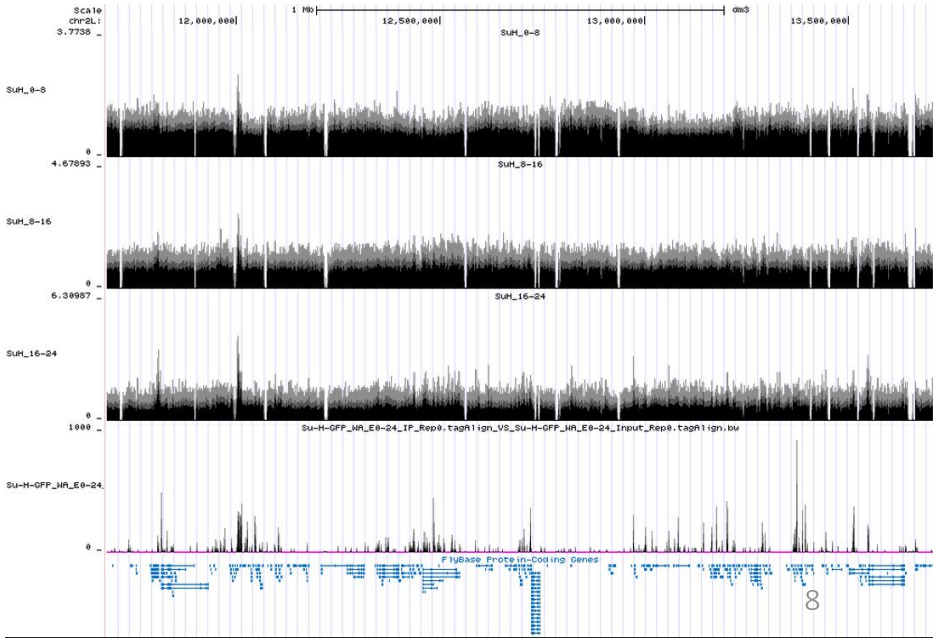
Su(H)

GFP: 3762

Iso1: 347

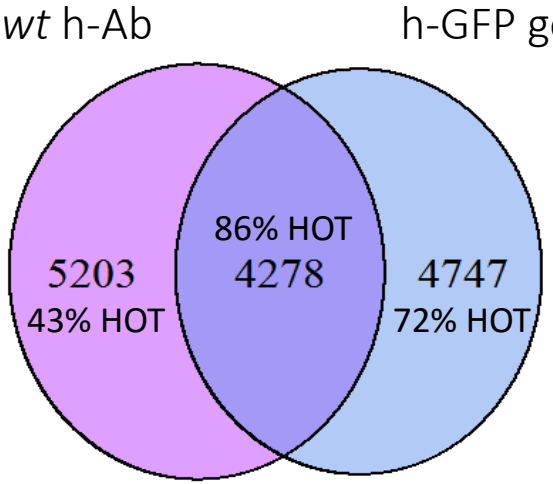
114 in common.

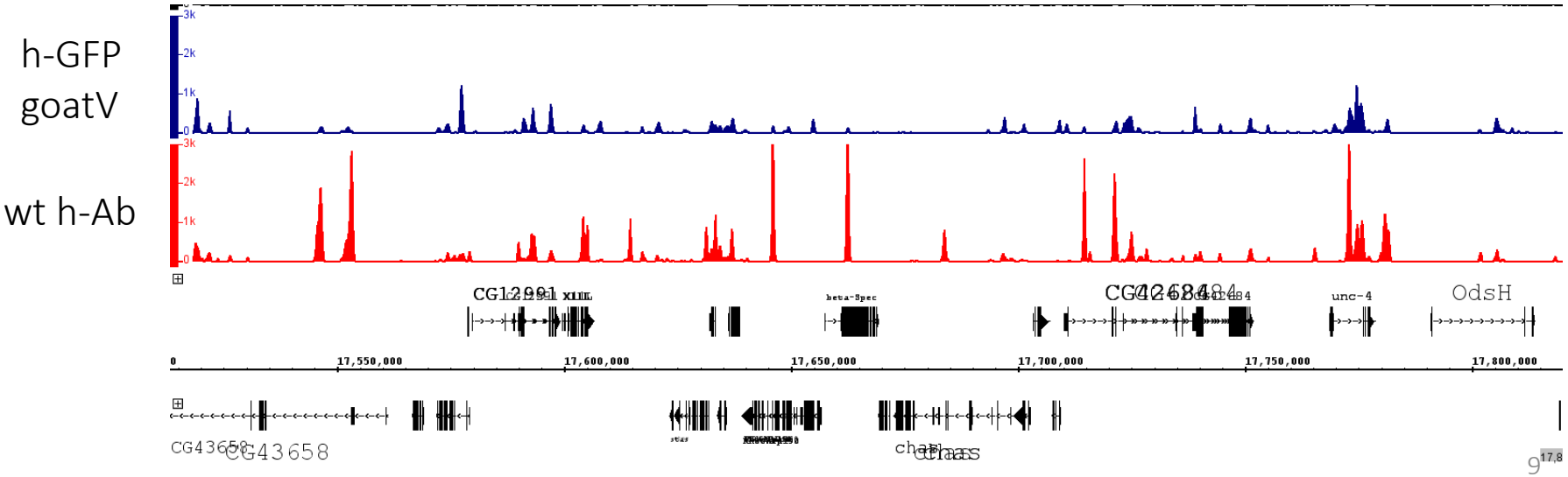




TF antibody vs GFP antibody

- Using current protocol, repeated experiment with TF Ab used for modENCODE
  - ~45% overlap.
  - Only 548 of overlapped peaks are not HOTregions, tend to be smaller peaks

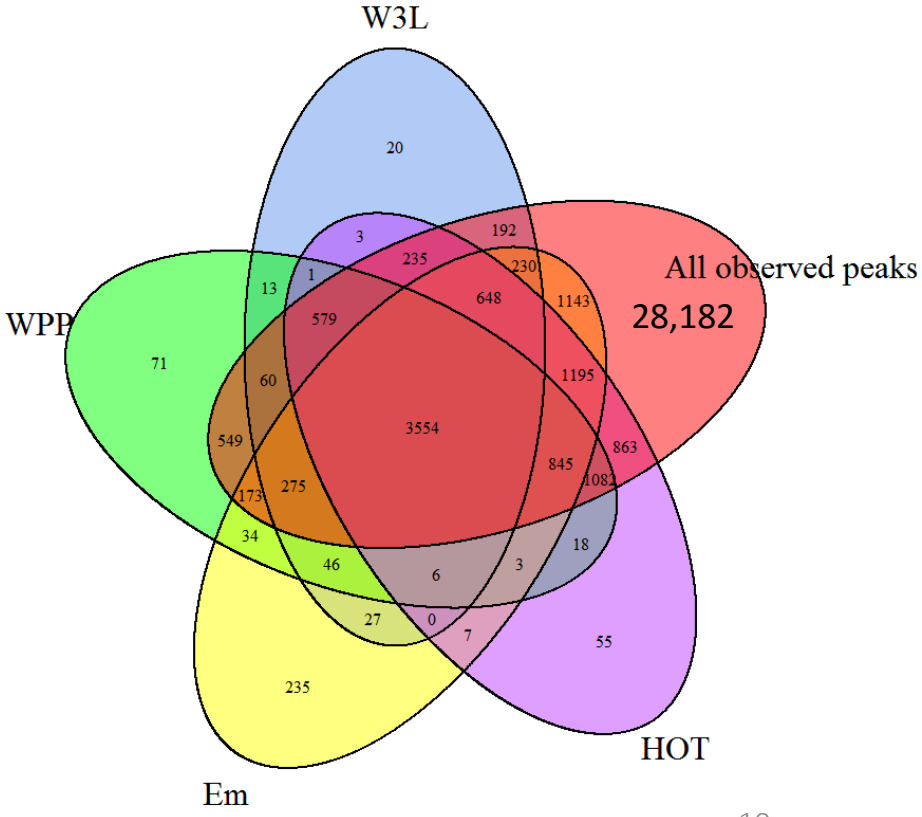




Background Peaks

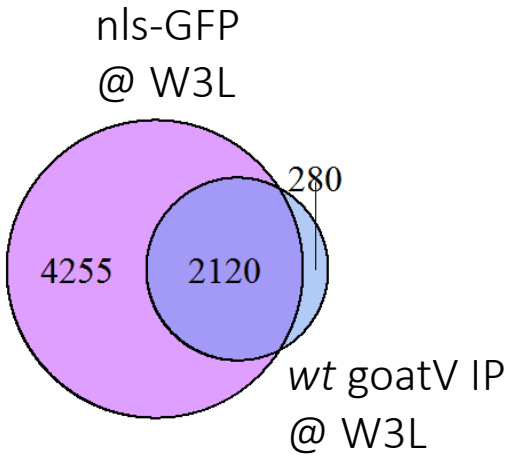
- Over-expression of nls-GFP yielded ChIP-seq peaks
  - Most were initially called as HOTspots

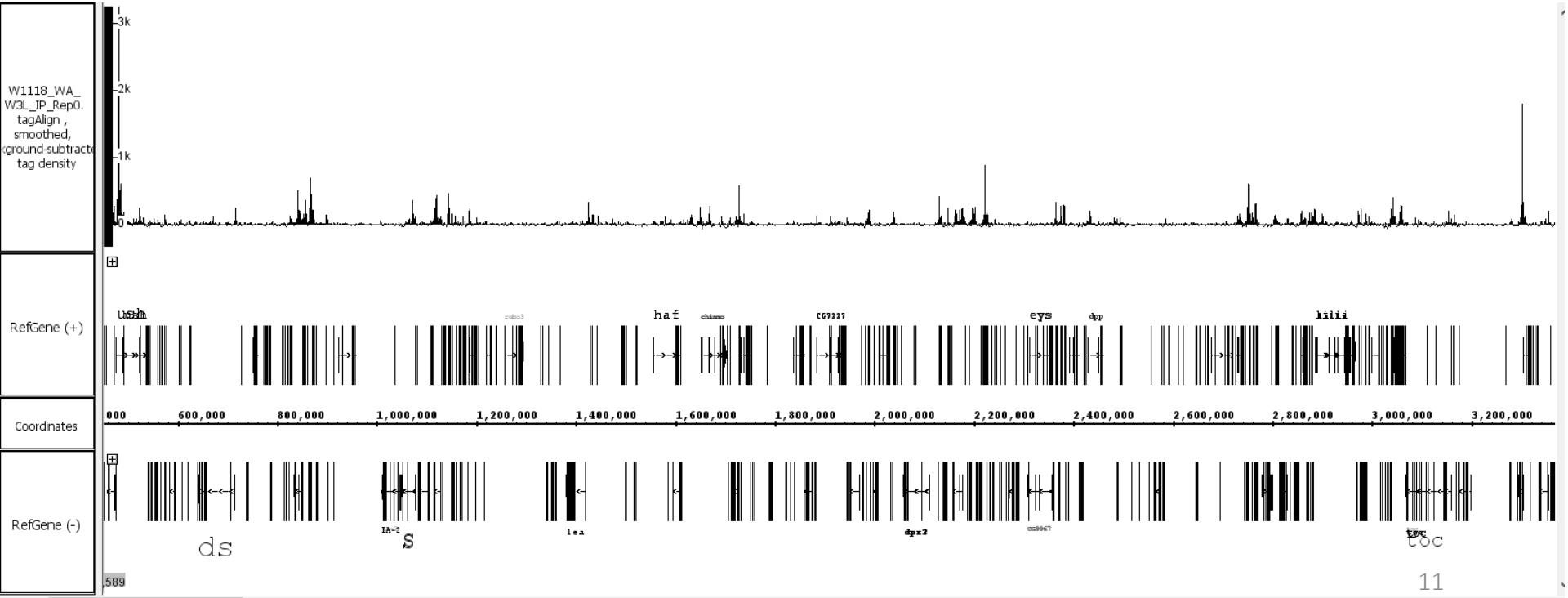
- Unclear if this was due to overexpression of GFP



Background Peaks

- IP using goatV on *wt Drosophila*
  - 2400 peaks called.
  - Fewer than nls-GFP
  - 97% peaks are HOTspots

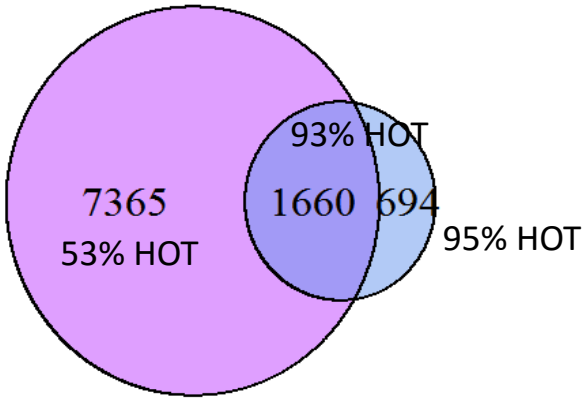




Background Peaks

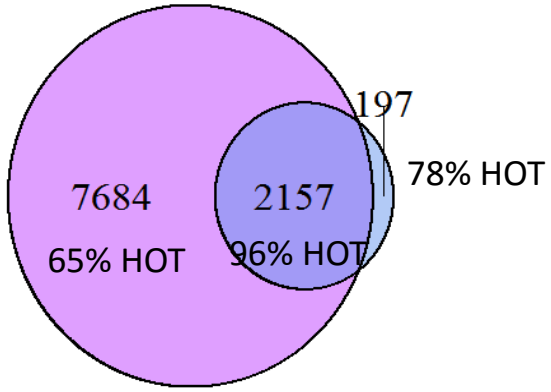
- GFP background peaks seen in non-GFP datasets
  - not really GFP background?

wt h-Ab vs *wt* goatV



wt h-Ab 60% HOT

h-GFP goatV vs *wt* goatV



h-GFP goatV 74% HOT

Future work

- Repeat nls-GFP and *wt*-goatV tests again
  - Will the same peaks be observed?
- Are the *wt*-goatV peaks a subset of the nls-GFP peaks?
- Are these regions the most intense when see in TF datasets?
- Can “hyper-ChIP-able” regions be subtracted from data prior to peak calling?
- Could check background with another Ab, eg aFLAG
- Test a few more TF-Ab vs GoatV.
  - Both comparing to old data and repeating with TF Ab.