modERN Call

White Lab

04-28-2016

modERN fly ChIP-seq datasets

24 new since last report, 72/80 for FY.

218 lines complete data sets

- ab, **Abd-B**, achi, acj6[!], ADD1-mimic, Antp-mimic, Atf-2, Atf3, Bab2, br, brk, Bteb2, btn, **cad**, Camta-mimic, cato, CG10462, CG10565, CG10631, CG10654, CG11398, CG11762, CG11902, CG12155, CG12236, CG12744, CG12942, CG13123, CG13624, CG14965, CG15073, CG1602, CG1620, CG1647, CG16863, CG1792, CG18476, CG18764, CG2120, CG3065, CG31388, CG32264, CG31627, CG3163, CG33213, CG3838, CG3995, CG4282, CG4402(aka CG34406), CG4424, CG4617, CG4820, CG4854, CG5204, CG5245, CG6792(aka Plzf), CG7368, CG7786, CG8089, CG8319, CG8944, CG9139, CG9305, CG9609, CG9876, CHES-1-like, chif*, chn, Chrac-16, **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, Elba2, emc, ems, en, ERR, esn, E(spl)m3, E(spl)mγ-HLH, Ets21C, Ets65A(ets3), Ets97D, E(var)3-9, **eve**, **exd**, ey, eyg, E(z), FoxP, foxo(&MiMIC), 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, ind, insv, jim, **jing**, Jra*, jumu-mimic, kay, 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, six4, slou, slp2, Smox, Sox102F, Sox14, Sox15, Sry-delta, **Stat92E**, Su(H), Su(var)3-7, sv, svp, tai, tin, **tio**, **tll**, toe, topi, trh, trl-mimic, tup, tx, **usp**, vfl, vri, Vsx2, woc, Zfh2, ZnT49B.

- **14 from modENCODE**
- 204 from modERN
- * multiple isoforms run
- ! Multiple time-points collected
- XX: verifying data with rerun.

Current ChIP-seq progress

MiMIC being expanded for ChIP-seq (CG16779, CG9727, snoo, dsx[†])

91 tagged lines being expanded for ChIP-seq
Target Stages:

Embryo: ac, ash1, Az2, bcd, Beaf-32, CG10274, CG10543[‖], CG11723, CG12104, CG13775[‖], CG15011, CG16815, CG15479(Mabi)[‖], CG15601, CG16629, CG17181, CG17568, CG18011, CG1832, CG18619(REPTOR), CG2116, CG30431, CG30403, CG3281, CG6654, CG6683, CG6765, CG7556, CG7928, CG7987, CG8359, CG8388, CG9727, CG9817, CG9948, CrebB, CTCF, dalao, E(bx), E(spl)m-beta, fd96Cb, Fer1, fkh, Hey, HmgZ, Kr[‖], l(3)neo38, meics, mor, MTF-1, nau, oc, odd, org-1, pita, run, sc , Sp1, su(Hw), su(var)2-10-RH, TFAM, unpg, Usf, Vsx1, Xbp1, YL-1[‖].

W3L: CG7839, dmrt93B, HLH106[§], rgr.

WPP: bsh, caup[‖], CG32006, CG33017, CG4318, CG4374, CG6276, CG6808, CG6854, CG9139(Rabex-5), net[†], NfI, salr.

AM: CG11617, CG33017, CG7045, Clk.

AF: CG30403, CG8159, pad, tj.

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

*: Probably need to recollect at better time point
[†]: Collect at different time-point if fail
[‖]: failed once

34: expanding
16: Chromatin extracted
05: IPed
26: Libraried
 21: @ HGAC
 00: Processing
 05: awaiting reps or
  recollecting

Control ChIP-seq

- nls-GFP Controls 5629/30:
  - E0-24: X X X X Need: 0 more
  - W3L: X X   Need: 1 more
  - WPP: X X   Need: 1 more
  - AF: X X X   Need: 0.5 more Ad
- W1118 goatV:
  - E0-24: X   Need: 12 more for
  - W3L: X X   Need: 1 more
  - WPP: X   Need: 2 more
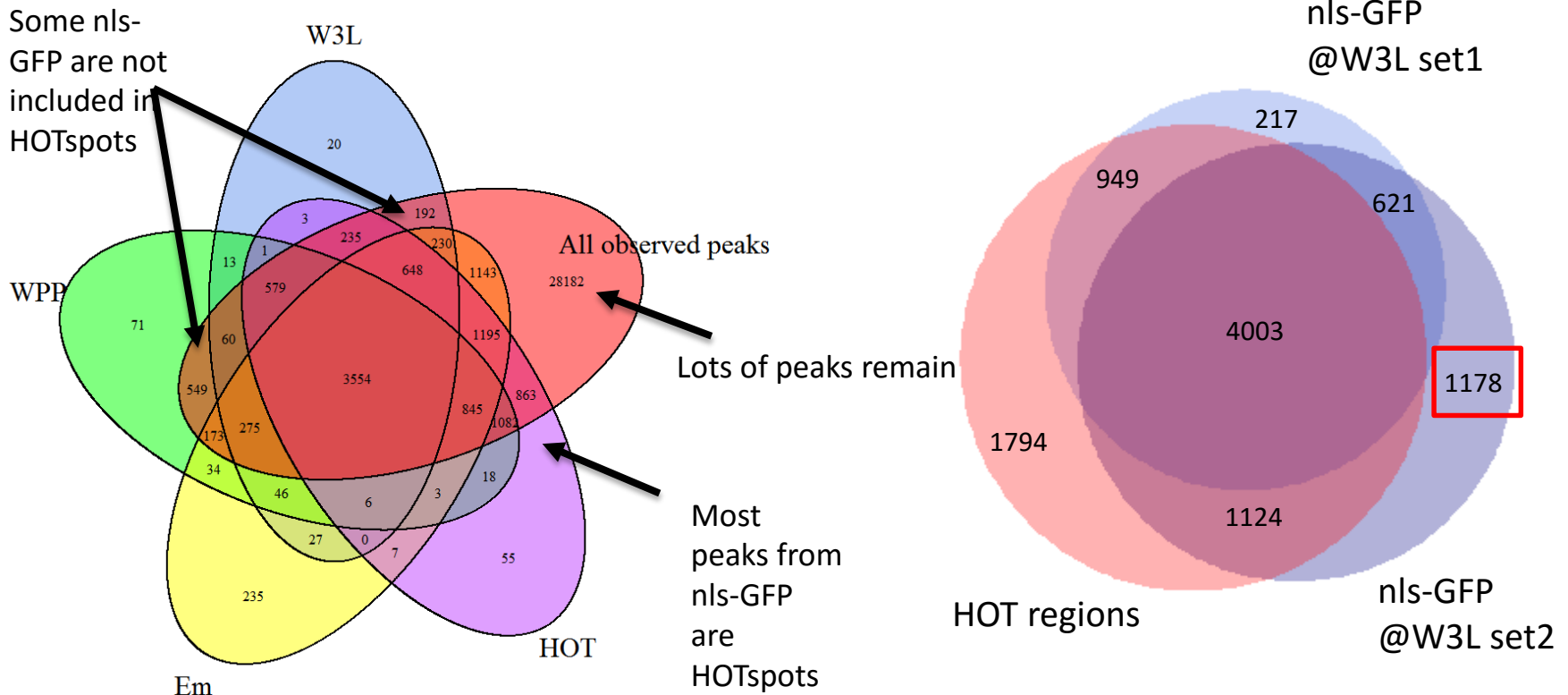 - Ad: Need: 3 more

- Potential W1118 TF Abs:
  - cnc, disco, eve, h, Hr39, Jra, Jumu, kn, Med, mod(mdg4), p53, slp2, Trl, psq, CG1832, kr

XX: Chromatin extracted
XX: IPed
XX: Libraried

Reproducibility of controls

- Good overlap between nls-GFP, but additional peaks!

C. elegans ChIP-seq

- Being Sequenced:
  - POP-1, MADF-2, CEH-9, XND-1, TBX-9, LIN-11, SOX-4 (LE and L3), T26A5.8, TBX-2, LIR-3.

- In the process of submitting:
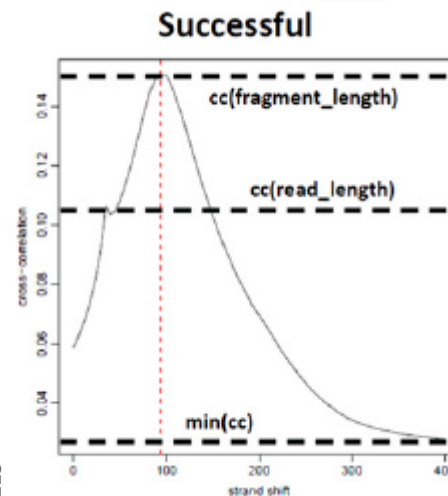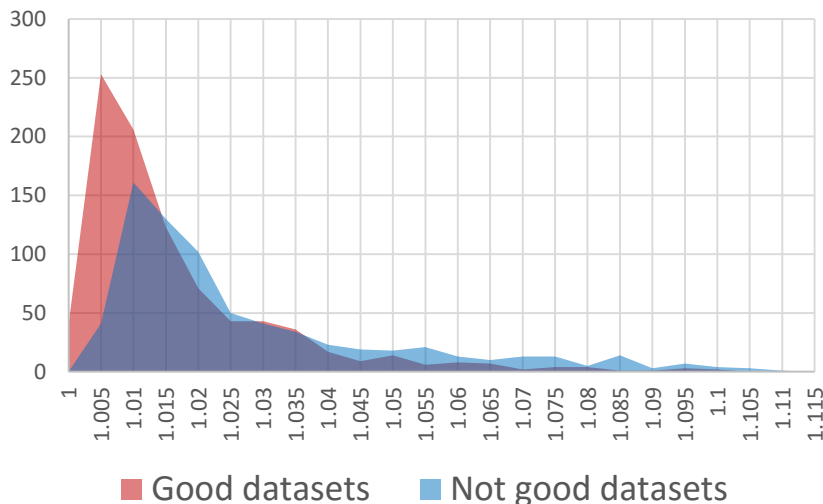 - F13C5.2, T22D7AL.16, UNC-120, HND-1

DCC

- New data, needs to be sent:
  - 9 Worm TFs
  - 11 Fly TFs to upload

- Finish uploading Ab comparisons for validation
  - 24 worm datasets pending

- Good for release:
  - 109 Worm experiments
  - 203 Fly experiments

- Stuck on dm6 uploads.
  - UCSC program giving error when creating bigWigs.

DCC ENCODE QC metrics

- Most metrics seem fine, eg # reads, #reads in peaks, #duplicate reads et ct.

- RSC and NSC problematic.
  - Plan was to provide those values but decide on threshold later.
  - Don't think we'll ever get a threshold based on quick check below.
    - Larger NSC and RSC better, but modERN good datasets have lower values!!!

Histogram of NSC values between "good" datasets and cursorly subtracted "not good" datasets

■ Good datasets   ■ Not good datasets

Successful

$$RSC = \frac{cc(fragment\ length) - min(cc)}{cc(read\ length) - min(cc)} \quad >1$$

$$NSC = \frac{cc(fragment\ length)}{min(cc)} \quad >1.05$$

Landt SG, Marinov GK, Kundaje A, Kheradpour P, Pauli F, Batzoglou S, et al. 2012. Genome Res.

Future work

- Upload QC metrics to DCC and continue with dm6 and new datasets

- Continue with control experiments and TFs