

Supervised detection of the same peaks jointly across several ChIP-seq samples

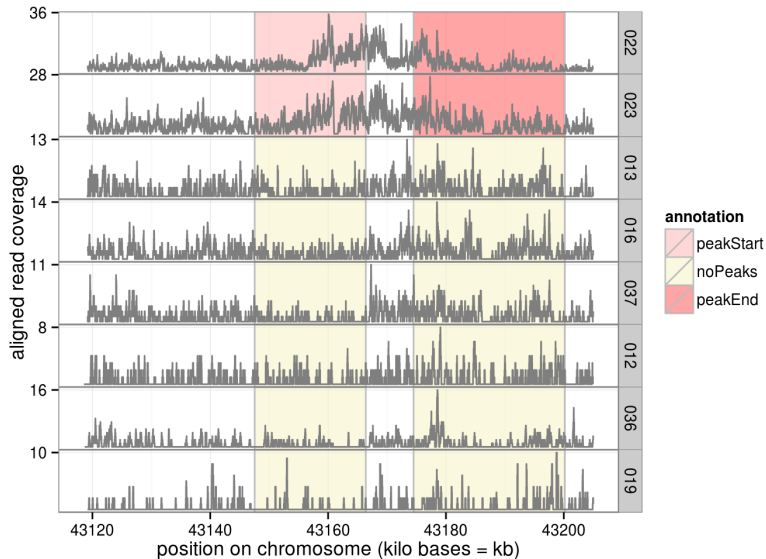
Toby Dylan Hocking

toby.hocking@mail.mcgill.ca

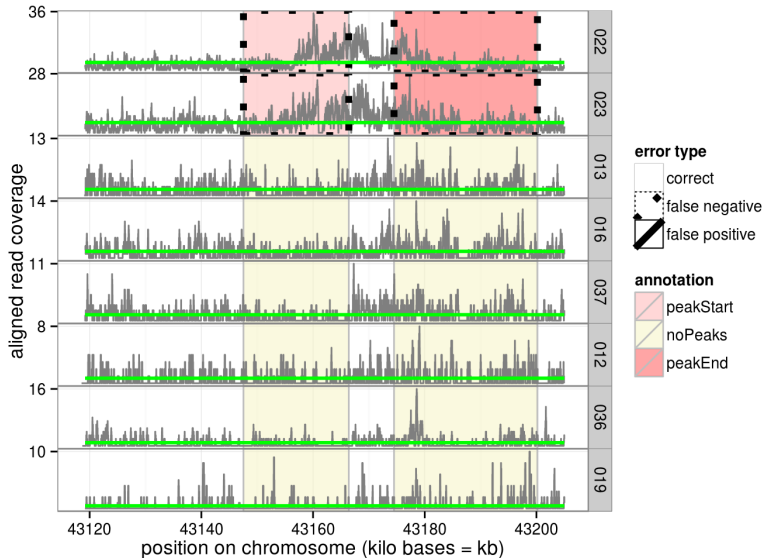
joint work with Guillem Rigau and Guillaume Bourque

2 April 2015

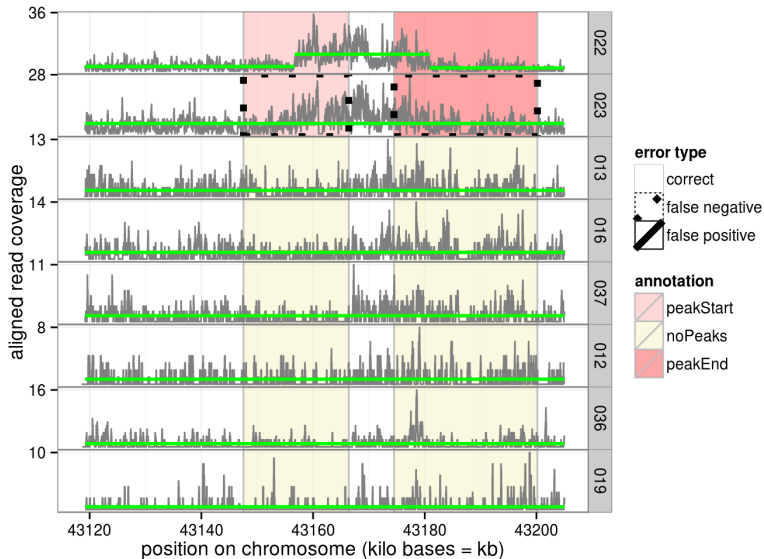
H3K36me3 data and labels



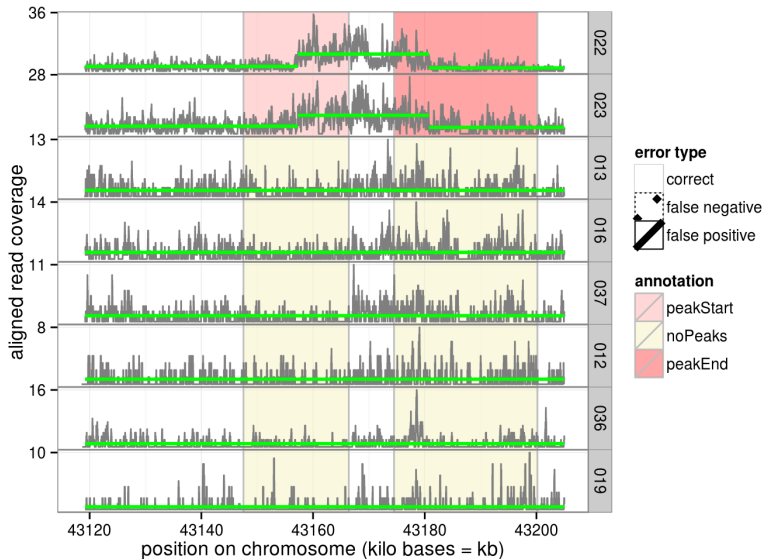
PeakSegJoint model with 0 peaks



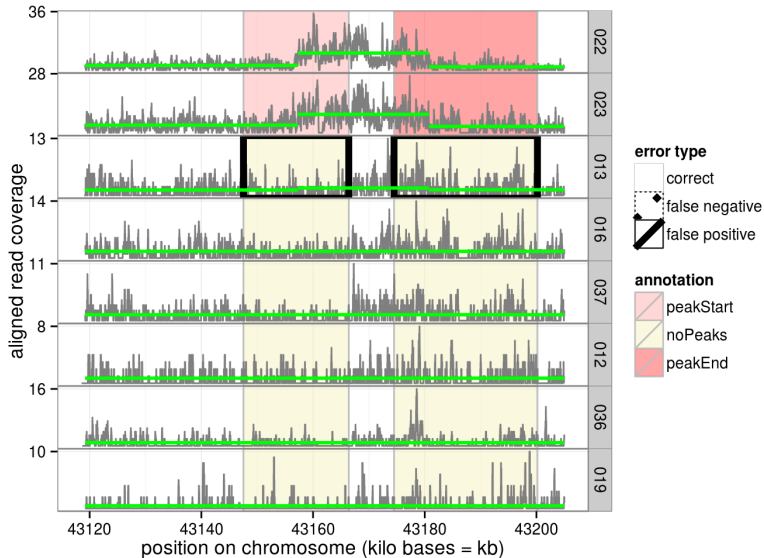
PeakSegJoint model with 1 peak



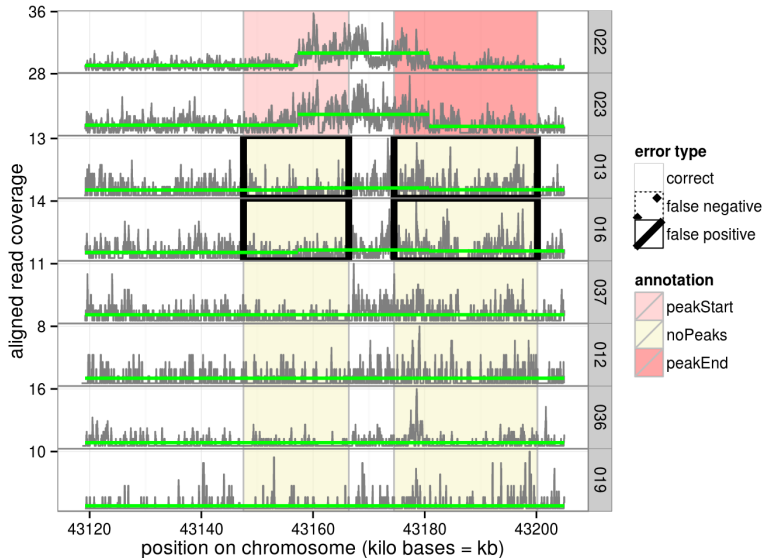
PeakSegJoint model with 2 peaks



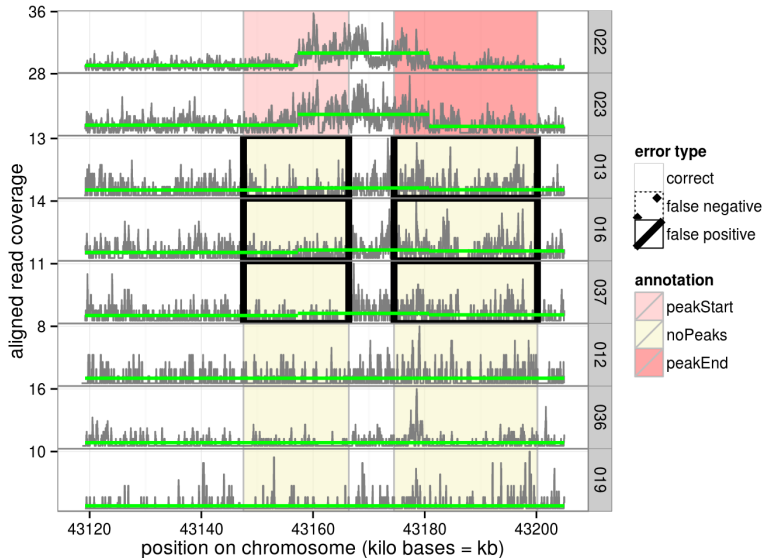
PeakSegJoint model with 3 peaks



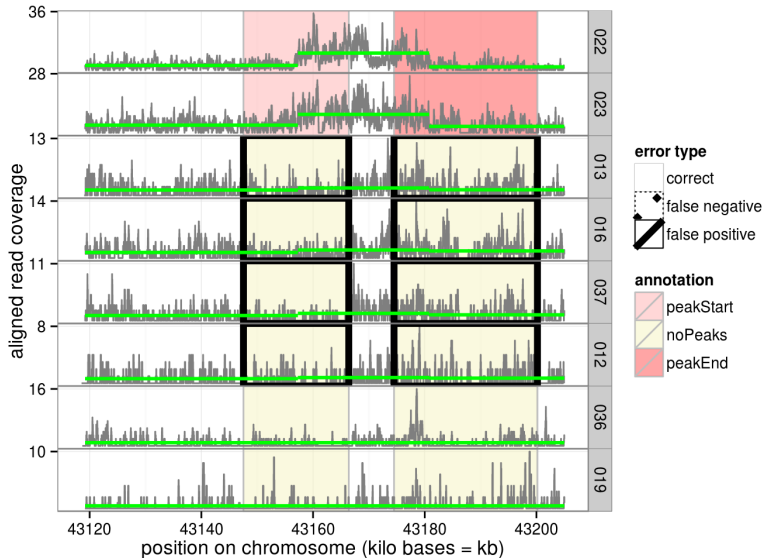
PeakSegJoint model with 4 peaks



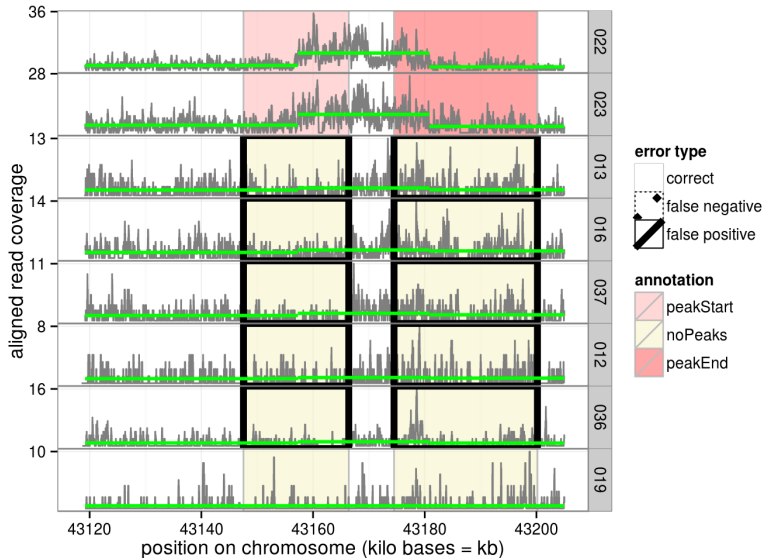
PeakSegJoint model with 5 peaks



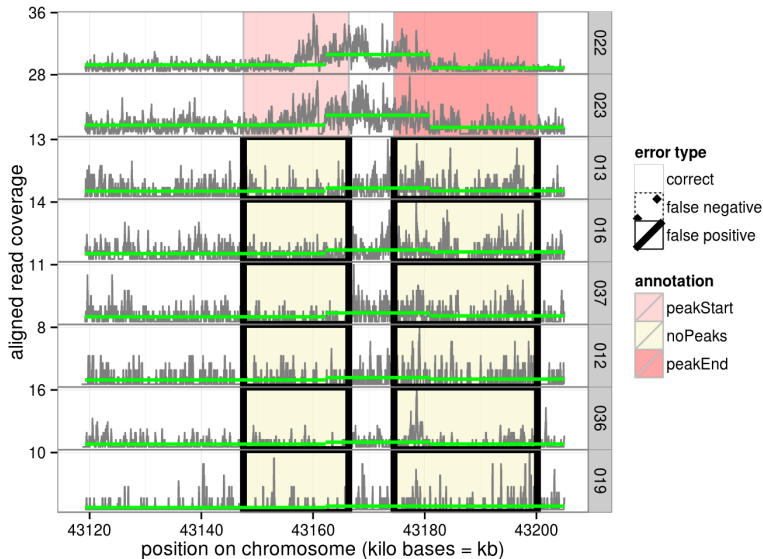
PeakSegJoint model with 6 peaks



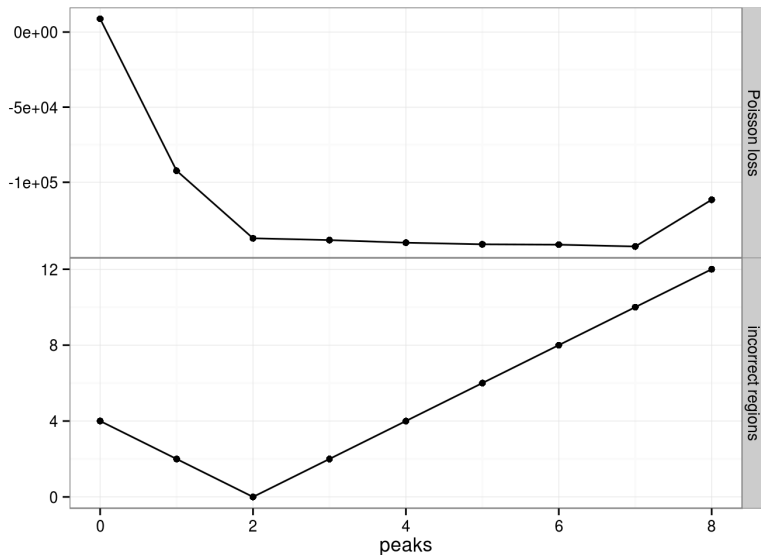
PeakSegJoint model with 7 peaks



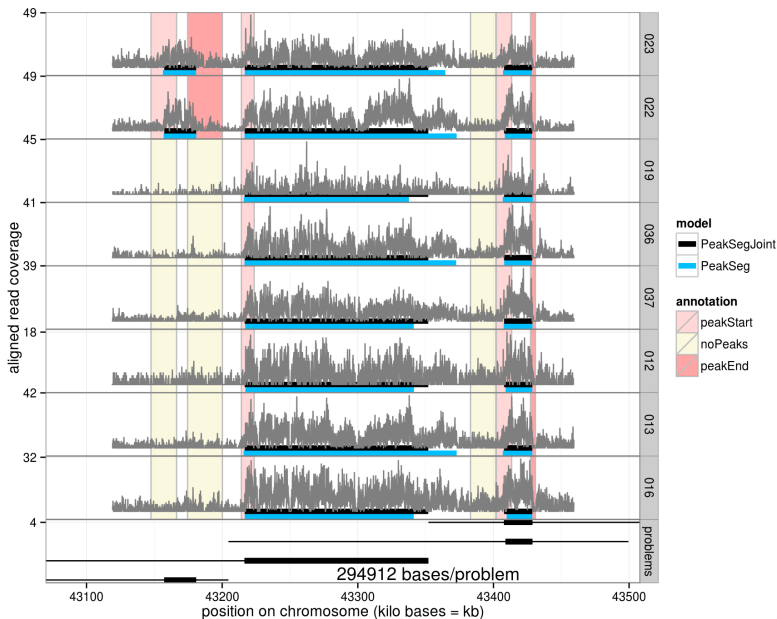
PeakSegJoint model with 8 peaks



Select model with minimal number of incorrect regions



H3K36me3 data, PeakSeg and Joint model



Timings on example H3K36me3 data

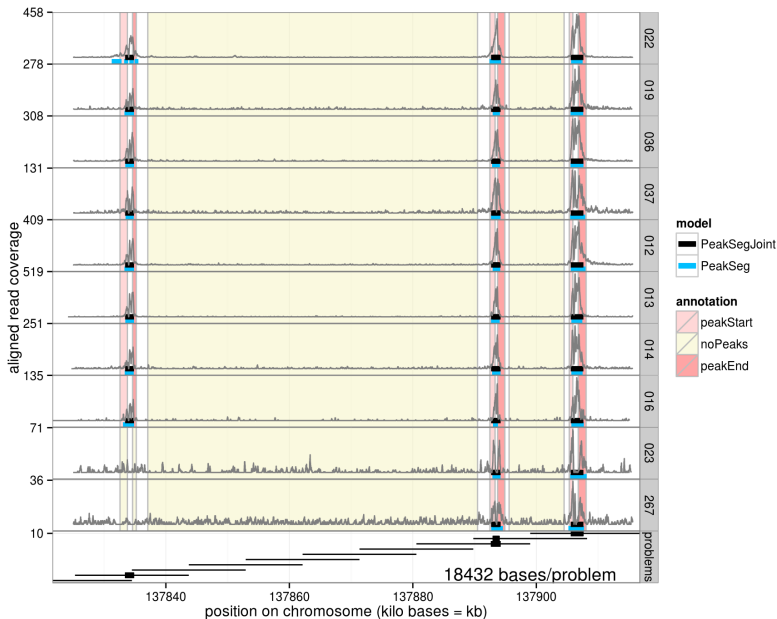
Find best 0,...,9 peaks in each of 8 samples (80 PeakSeg models):

seconds	sample.id
1.07	McGill0023
1.06	McGill0022
1.05	McGill0019
1.07	McGill0036
1.09	McGill0037
1.06	McGill0012
1.08	McGill0013
1.06	McGill0016
8.54	total

Find best common peak in 0,...,8 samples in each of 4 genomic regions (36 PeakSegJoint models):

	data	seconds
chr21:42909696-43204608	23595	0.04
chr21:43057152-43352064	162129	0.09
chr21:43204608-43499520	206437	0.11
chr21:43352064-43646976	67903	0.06
total		0.30

H3K4me3 data, PeakSeg and Joint model



Timings on example H3K4me3 data

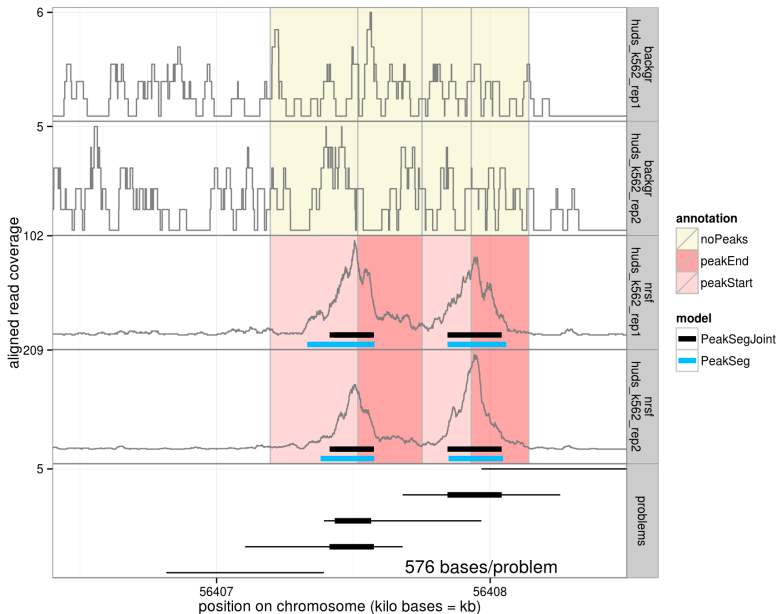
Find best
0,...,9 peaks
in each of 10 samples
(100 PeakSeg models)

seconds	sample.id
0.76	McGill0022
0.74	McGill0019
0.74	McGill0036
0.75	McGill0037
0.75	McGill0012
0.73	McGill0013
0.74	McGill0014
0.74	McGill0016
0.74	McGill0023
0.75	McGill0267
7.44	total

Find best common peak
in 0,...,10 samples
in each of 10 genomic regions
(110 PeakSegJoint models)

	data	seconds
chr21:137816064-137834496	7603	0.01
chr21:137825280-137843712	12420	0.05
chr21:137834496-137852928	7023	0.01
chr21:137843712-137862144	3915	0.04
chr21:137852928-137871360	3597	0.03
chr21:137862144-137880576	3588	0.03
chr21:137871360-137889792	4255	0.05
chr21:137880576-137899008	13317	0.05
chr21:137889792-137908224	26436	0.06
chr21:137899008-137917440	19644	0.06
total		0.39

NRSF transcription factor data, PeakSeg and Joint model



Timings on example transcription factor data

Find best

0,...,9 peaks

in each of 4 samples

(40 PeakSeg models)

seconds	sample.id
0.23	backgr huds_k562_rep1
0.25	backgr huds_k562_rep2
0.30	nrsf huds_k562_rep1
0.31	nrsf huds_k562_rep2
1.09	total

Find best common peak

in 0,...,4 samples

in each of 5 genomic regions

(25 PeakSegJoint models)

	data	seconds
chr21:56406816-56407392	345	0.01
chr21:56407104-56407680	761	0.02
chr21:56407392-56407968	975	0.01
chr21:56407680-56408256	709	0.02
chr21:56407968-56408544	298	0.01
total		0.07

Conclusions

Thanks for your attention!

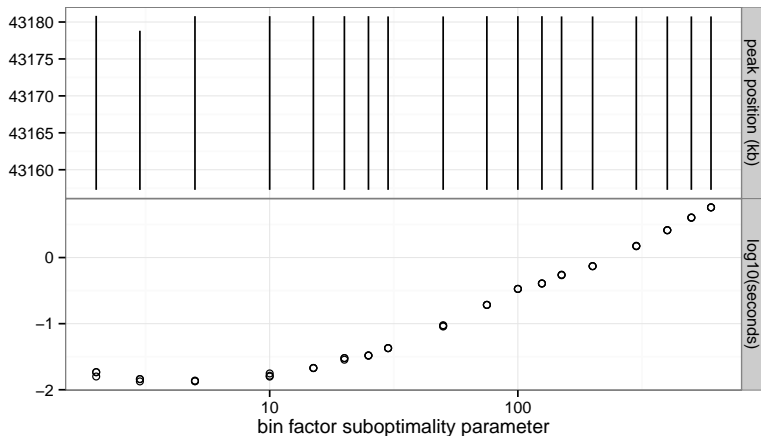
Write me at toby.hocking@mail.mcgill.ca to collaborate!

Source code for slides, figures, paper online!

<https://github.com/tdhock/PeakSegJoint-paper>

Supplementary slides appear after this one.

Bin factor parameter controls optimality and speed



H3K36me3 example data set, PeakSegJoint model with 2 peaks.