Supervised detection of the same peaks jointly across several ChIP-seq samples
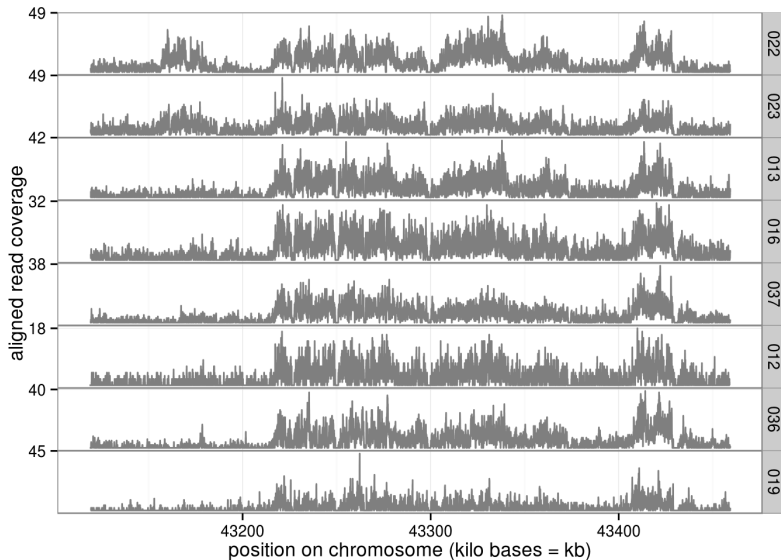
Toby Dylan Hocking
toby.hocking@mail.mcgill.ca
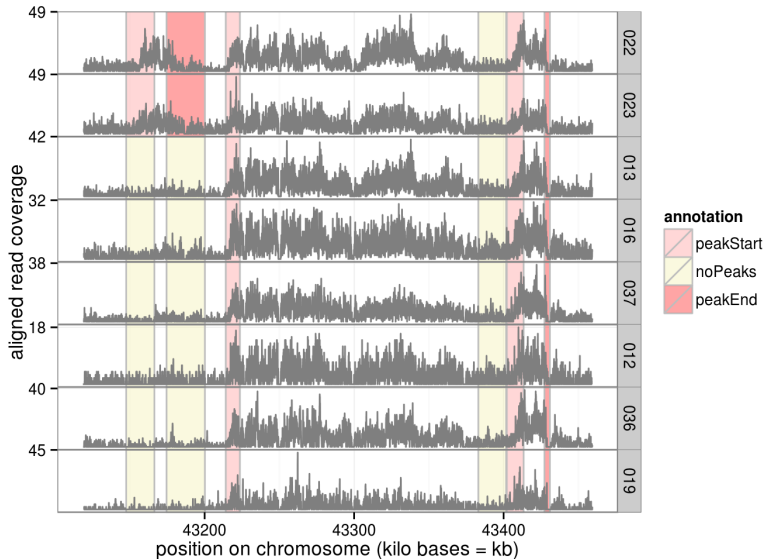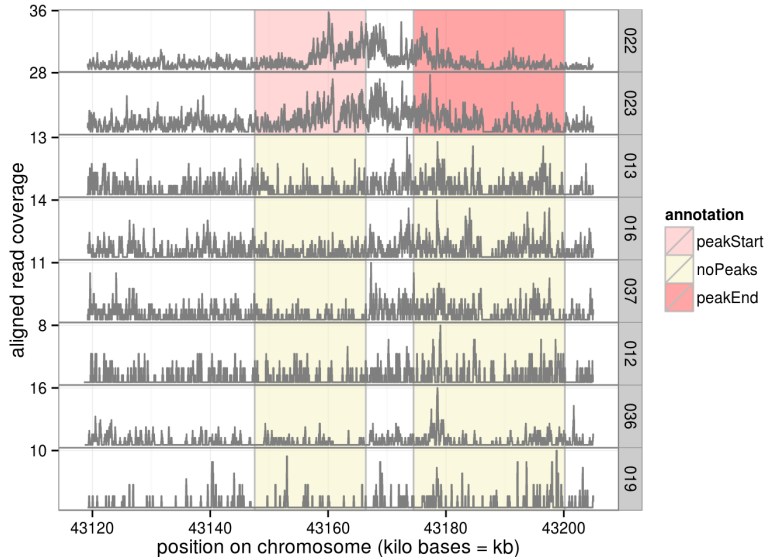joint work with Guillem Rigaill and Guillaume Bourque

2 April 2015
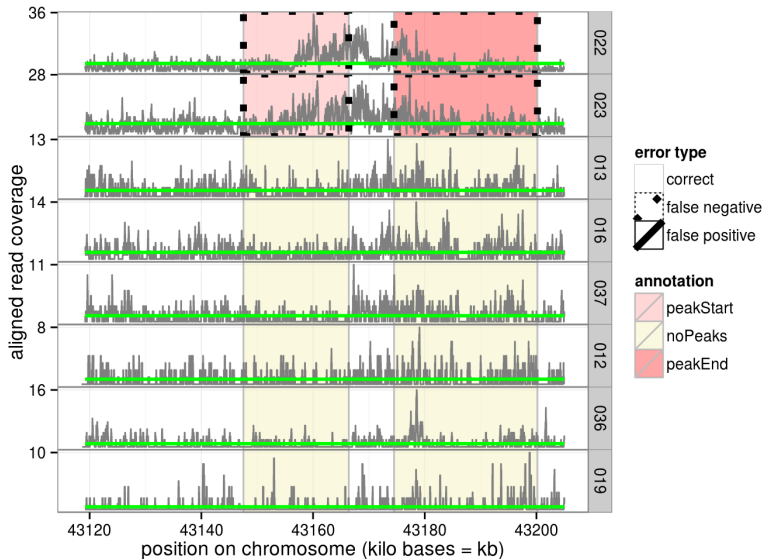
Peaks visually obvious in H3K36me3 data

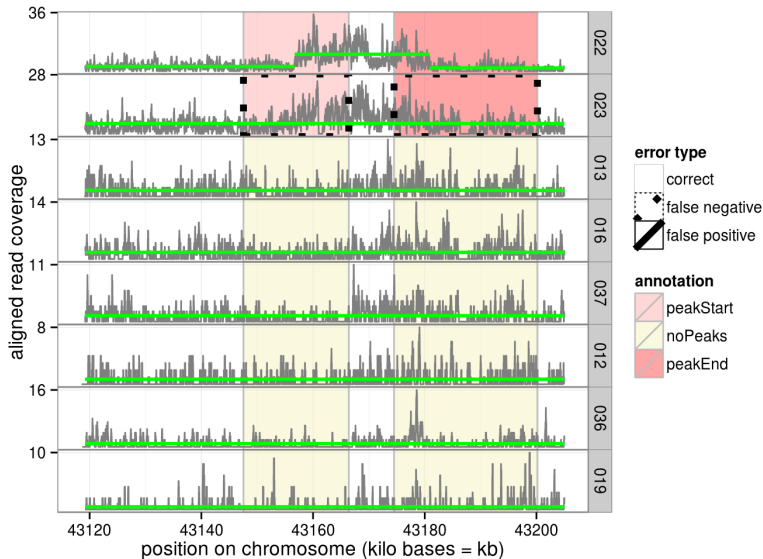# H3K36me3 data and visually determined labels

H3K36me3 data and labels (zoom to one peak)

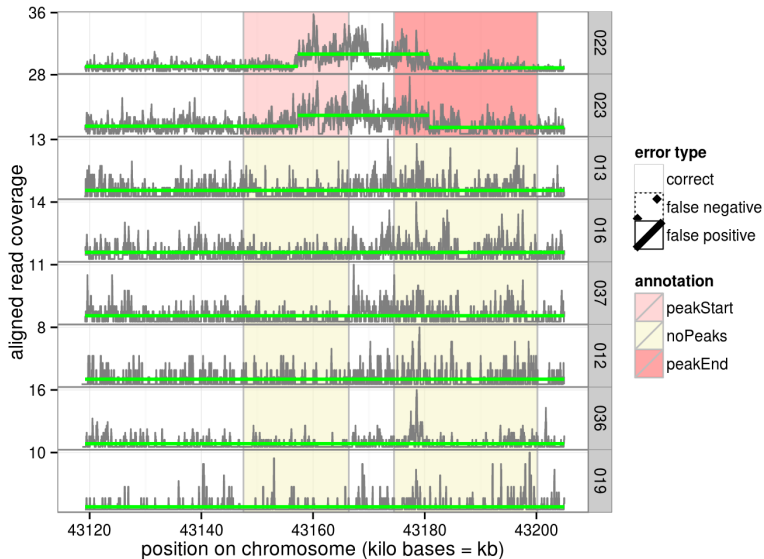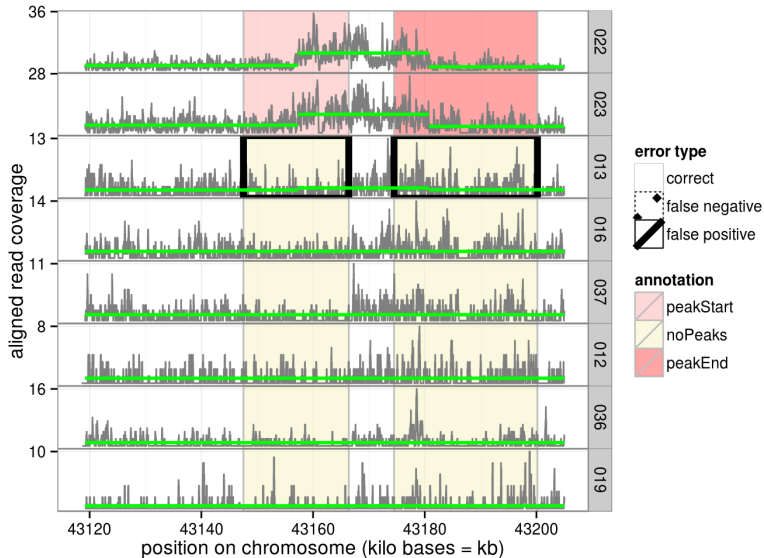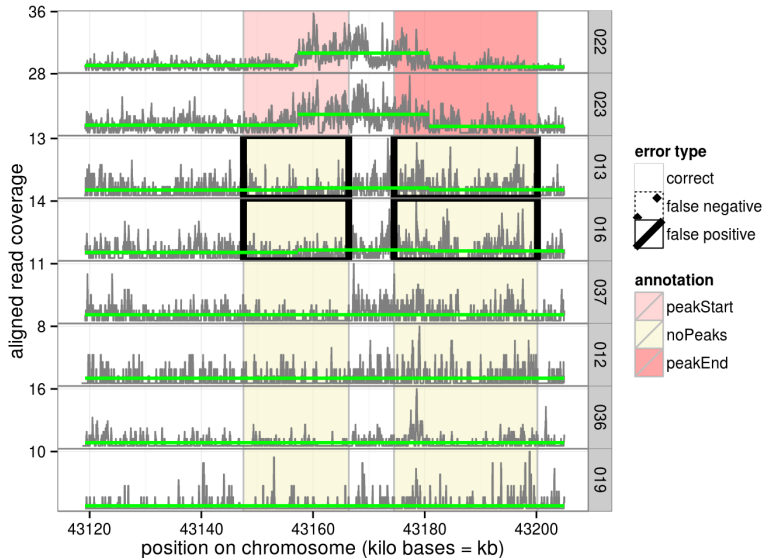# PeakSegJoint model with 0 peaks

PeakSegJoint model with 1 peak

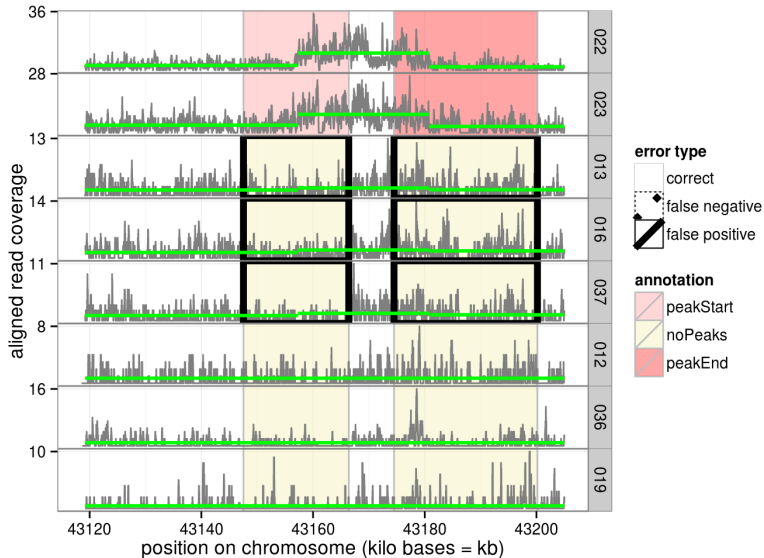# PeakSegJoint model with 2 peaks

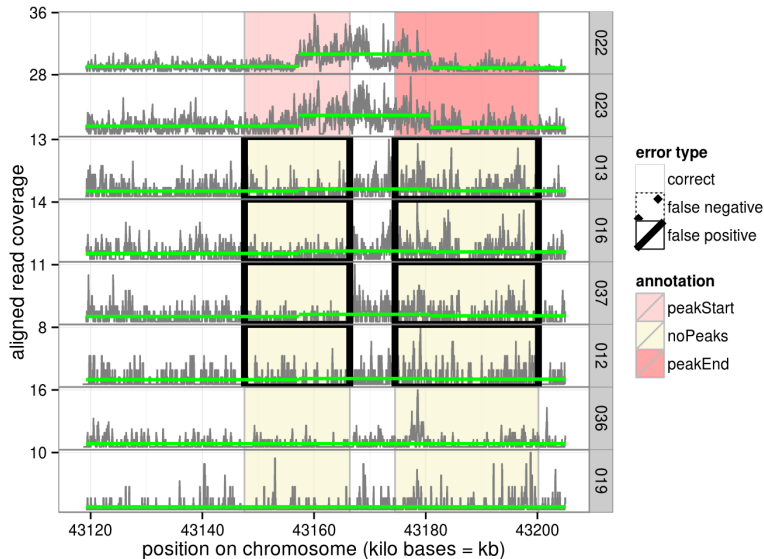# PeakSegJoint model with 3 peaks

PeakSegJoint model with 4 peaks

PeakSegJoint model with 5 peaks

PeakSegJoint model with 6 peaks

PeakSegJoint model with 7 peaks
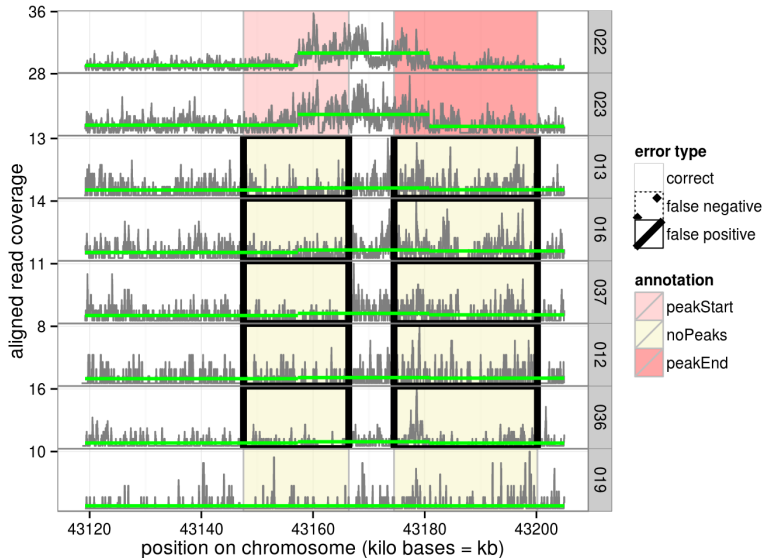
PeakSegJoint model with 8 peaks
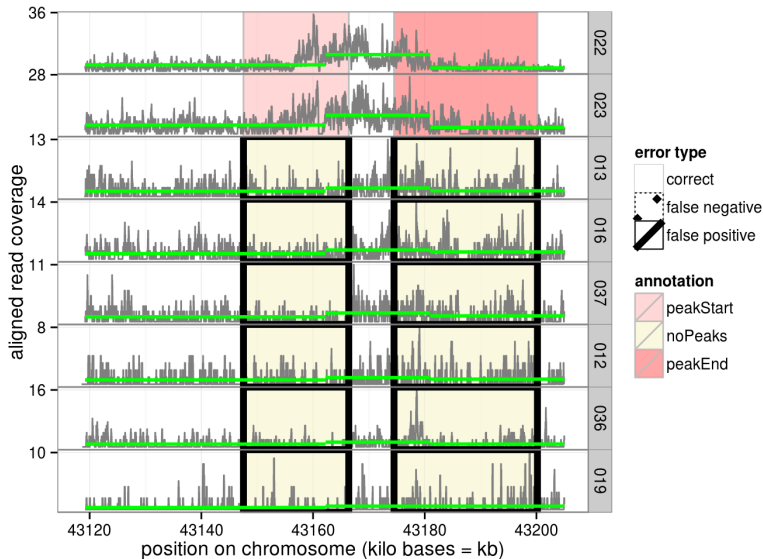
Select model with minimal number of incorrect regions
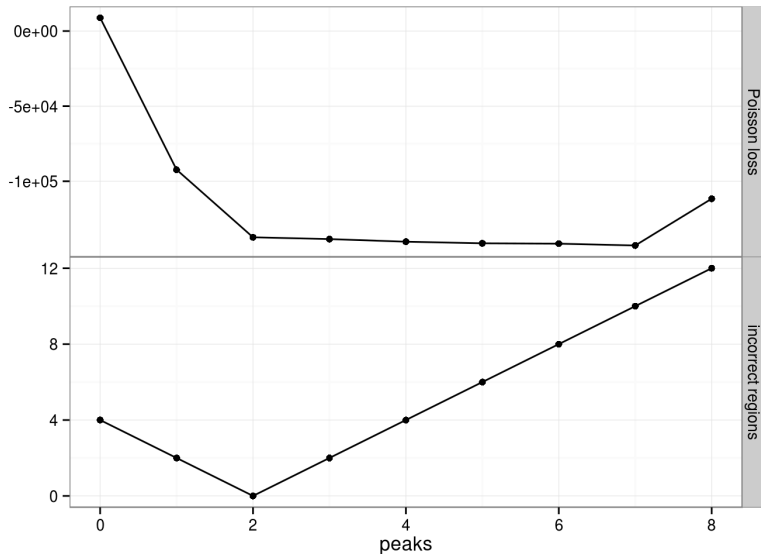
H3K36me3 data, PeakSeg and Joint model
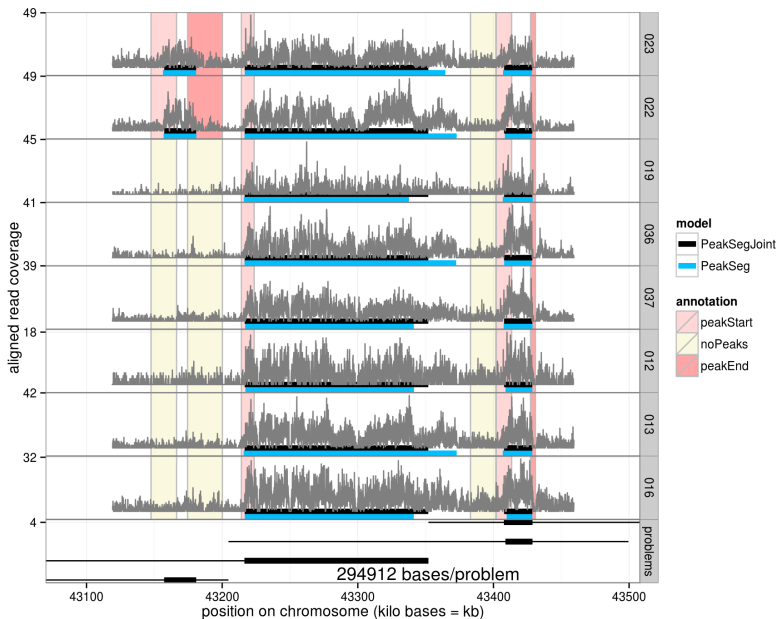
Timings on example H3K36me3 data

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

seconds	sample.id
0.77	McGill0023
0.77	McGill0022
0.76	McGill0019
0.77	McGill0036
0.77	McGill0037
0.77	McGill0012
0.77	McGill0013
0.77	McGill0016
6.14	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	0.11
chr21:43204608-43499520	206437	0.11
chr21:43352064-43646976	67903	0.06
total		0.31

H3K4me3 data, PeakSeg and Joint model

18432 bases/problem
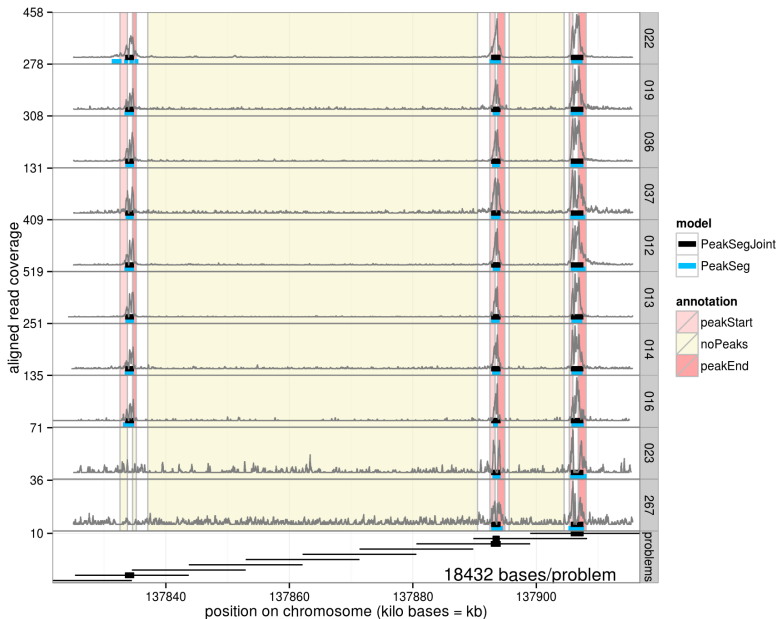
Timings on example H3K4me3 data

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

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

seconds	sample.id		data	seconds
0.75	McGill0022	chr21:137816064-137834496	7603	0.01
0.75	McGill0019	chr21:137825280-137843712	12420	0.05
0.75	McGill0036	chr21:137834496-137852928	7023	0.01
0.75	McGill0037	chr21:137843712-137862144	3915	0.04
0.75	McGill0012	chr21:137852928-137871360	3597	0.03
0.75	McGill0013	chr21:137862144-137880576	3588	0.03
0.75	McGill0014	chr21:137871360-137889792	4255	0.05
0.75	McGill0016	chr21:137880576-137899008	13317	0.05
0.75	McGill0023	chr21:137889792-137908224	26436	0.05
0.76	McGill0267	chr21:137899008-137917440	19644	0.05
7.52	total	total		0.38

NRSF transcription factor data, PeakSeg and Joint model
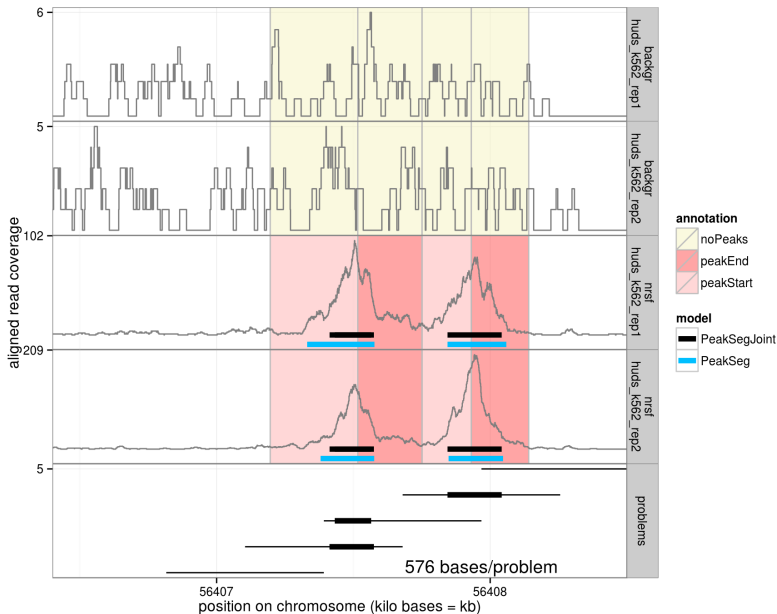
Timings on example transcription factor data

Find best
0,...,9 peaks
in each of 4 samples
(40 PeakSeg models)

seconds	sample.id
0.26	backgr huds_k562_rep1
0.24	backgr huds_k562_rep2
0.30	nrsf huds_k562_rep1
0.31	nrsf huds_k562_rep2
1.10	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

H3K27ac and Input data, PeakSeg and Joint model
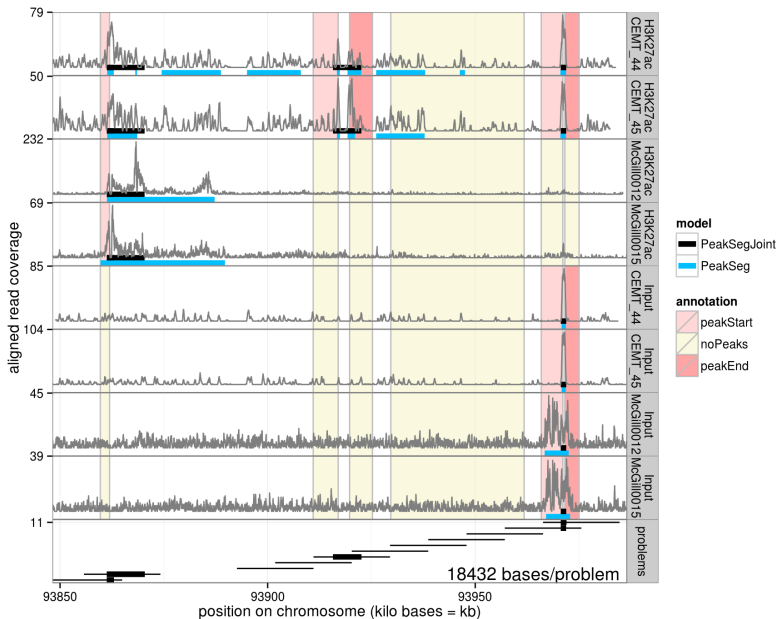
Timings on example H3K27ac data

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

seconds	sample.id
0.99	H3K27ac CEMT_44
0.98	H3K27ac CEMT_45
0.99	H3K27ac McGill0012
1.00	H3K27ac McGill0015
1.00	Input CEMT_44
1.01	Input CEMT_45
1.00	Input McGill0012
1.00	Input McGill0015
7.96	total

Find best common peak
in 0,...,8 samples
in each of 11 genomic regions
(99 PeakSegJoint models)

	data	seconds
chr11:93846528-93864960	7510	0.03
chr11:93855744-93874176	11675	0.04
chr11:93892608-93911040	5619	0.03
chr11:93901824-93920256	6236	0.03
chr11:93911040-93929472	5559	0.03
chr11:93920256-93938688	5149	0.04
chr11:93929472-93947904	4359	0.01
chr11:93938688-93957120	3071	0.03
chr11:93947904-93966336	3030	0.02
chr11:93957120-93975552	7184	0.04
chr11:93966336-93984768	7446	0.04
total		0.34

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
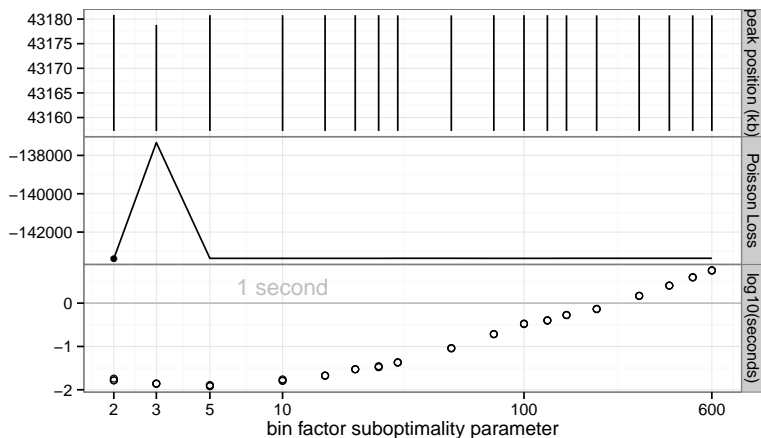
H3K36me3 example data set, PeakSegJoint model with 2 peaks.