Assessing the quality of peaks identified using mock IP

Jinrui Xu

modERN call

02-16-2017

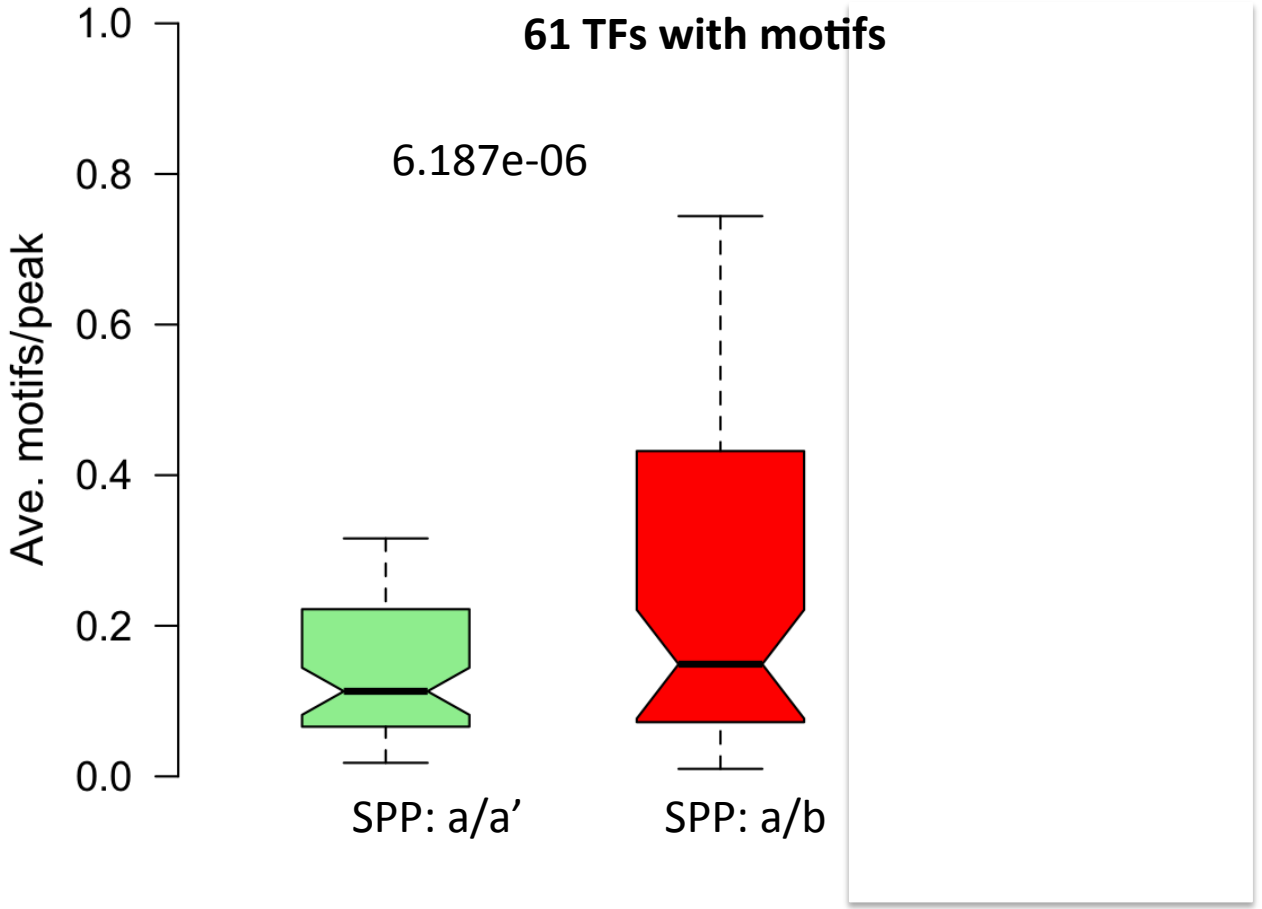
- Spurious peaks in modERN ChIP-Seq
  - Higher peak overlaps among modERN datasets than modENCODE
  - High peak overlap btw IP and mockIP

- Potential factors causing the spurious peaks
  - GFP-Ab can enrich some genomic regions, more than the Ab used in modEncode
  - In the TF-GFP, the GFP can also interact with DNA

- To control GFP-Ab
  - Mock IP with GFP-Ab (b), DNA input for mock IP (b’)
  - Treatment IP (a), DNA input (a’)

- To test if mock IP removes spurious peaks
  - Motif enrichment as a proxy of peak qualities
  - 61 TFs with motifs identified by Bacteria one hybrid

- With mock IP, SPP identify top 500* peaks (a/b) enriched more motifs than those with DNA input (a/a')



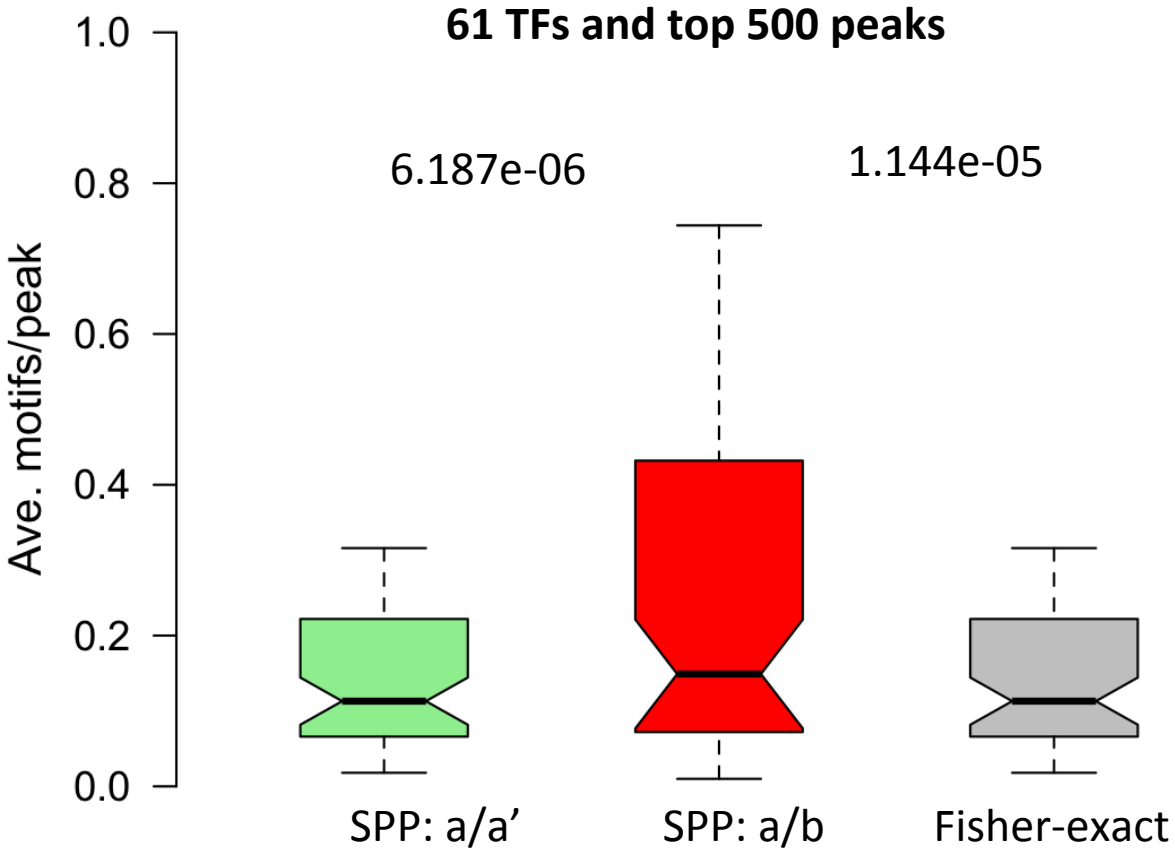
* Other cutoffs and IDR cutoff lead to same conclusion

- No. of sig. (IDR=0.01) peaks is much smaller than using DNA input
  - Median: 184 (a/b) vs 3550 (a/a’)

- Statistical models in my peak calling methods
  - Poisson
  - 2 modified Poisson models weighted by a'
  - Fisher-exact test (a/a' vs b/b')
  - Poisson-Gamma

- These methods all get more sig. peaks than SPP a/b, but lower quality

- The Fisher-exact test (a/a’ vs b/b’) provides good motif enrichment but still much less than SPP



- In mock IP, many genomic regions has no or only a couple of reads
  - Make my statistical models unreliable
  - Non-model based methods, such as SPP, perform better

- Solutions to this problem
  - Use more available mock IP controls
  - Use mock IP from strains expressing GFP

    Pros:

    It control both GFP and GFP-Ab

    If GFP contributes peaks, the mock IP will imunioprecipate more DNA, alleviating the sample size issue

    Cons:

    But need to consider differential expression of GFP and TF
  - Develop or modify non-model based methods

- Solutions to this problem
  - Tune IDR threshold in SPP to find a point that balances motif enrichment and number of sig. peaks

  "With shallow sequencing depths, you can use an IDR threshold as relaxed as 0.1 if you start with < 100K pre-IDR peaks"

  https://sites.google.com/site/anshulkundaje/projects/idr