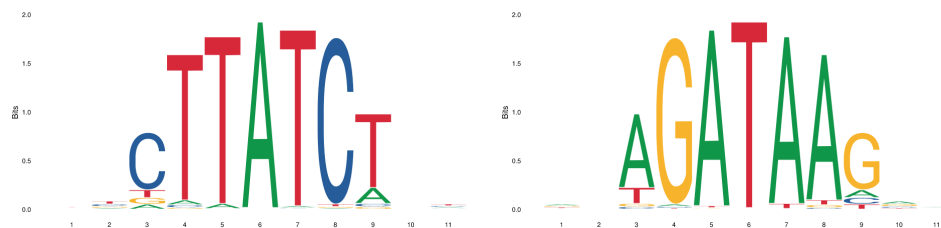# Motif discovery results

Input data

- GATA2
- GATA4
- NR3C1 time-series data:
    - 0h
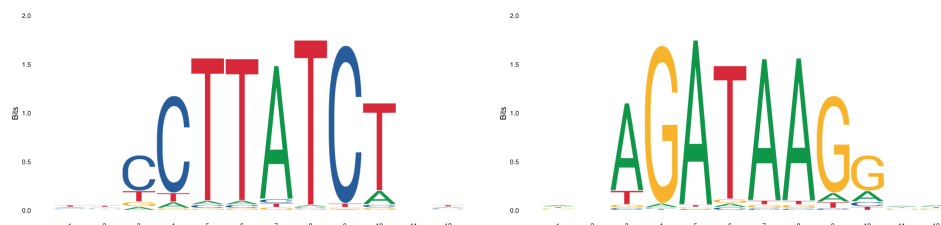   - 10min
    - 8h

GATA transcription factors

The expected motifs can be found in jaspar.genereg.net.

**GATA2** (MA0036.3)

GATA4 (MA0482.2)


NR3C1 transcription factor

Glucocorticoid receptor (GR), bound by cortisol and other glucocorticoids. Transactivation - upon binding of glucocorticoid, the receptors form homodimers, moves from cytosol to nucleus and binds DNA to activate transcription.

Dexamethasone is an agonist that binds to NR3C1 to activate its binding to the DNA.

The given data is a time-series, where the cells are observed 0h, 10min and 8h after treatment with dexamethasone.
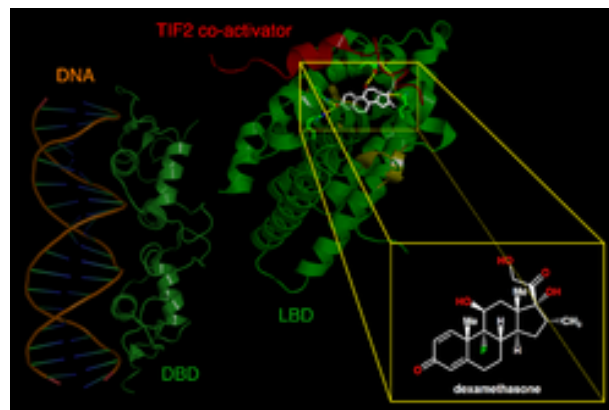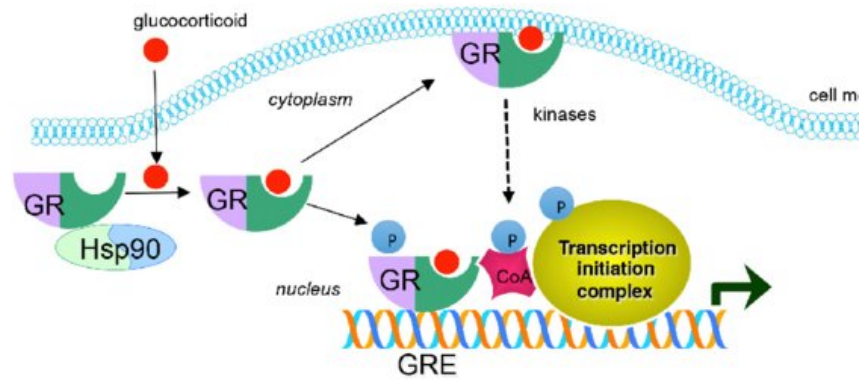
The expected motif in JASPAR under the matrix ID MA00113.3 (palindromic motif).
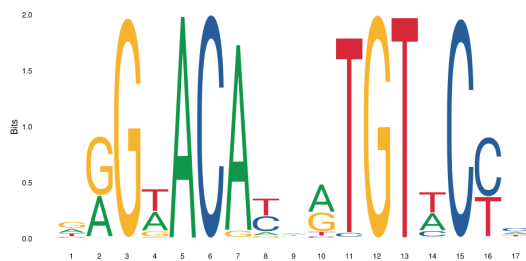

This canonical binding can be a combination of GR signals (capturing monomer or dimer) when looking into ChIP-exo data.
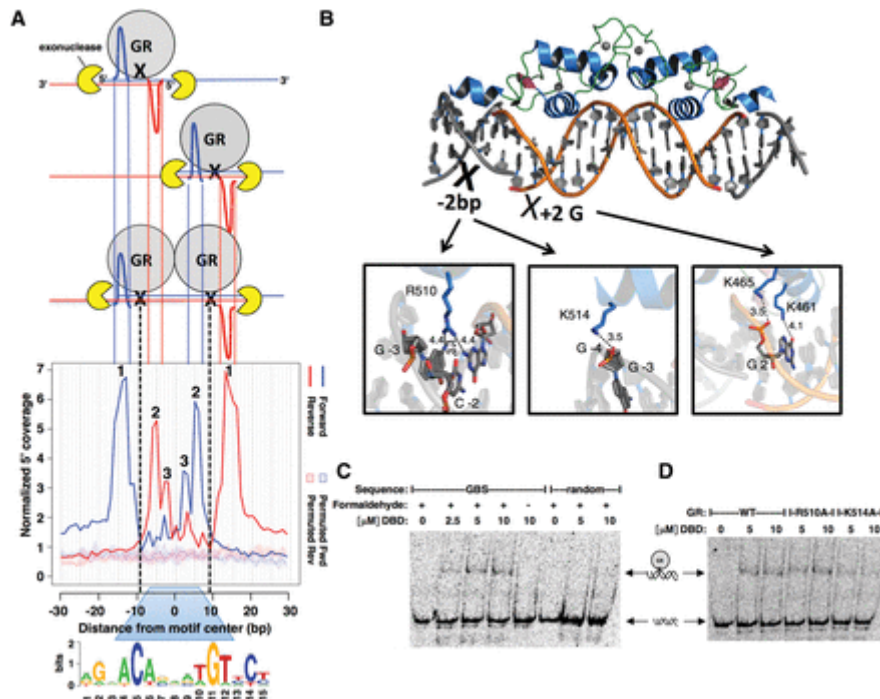
Adapted from Starick SR et al. 2015. ChIP-exo signal associated with DNA-binding motifs provides insight into the genomic binding of the glucocorticoid receptor and cooperating transcription factors

However, GR can produce non-GBS footprint profiles, which then can give a different DNA binding motif. This depends on what GR is partnering with.
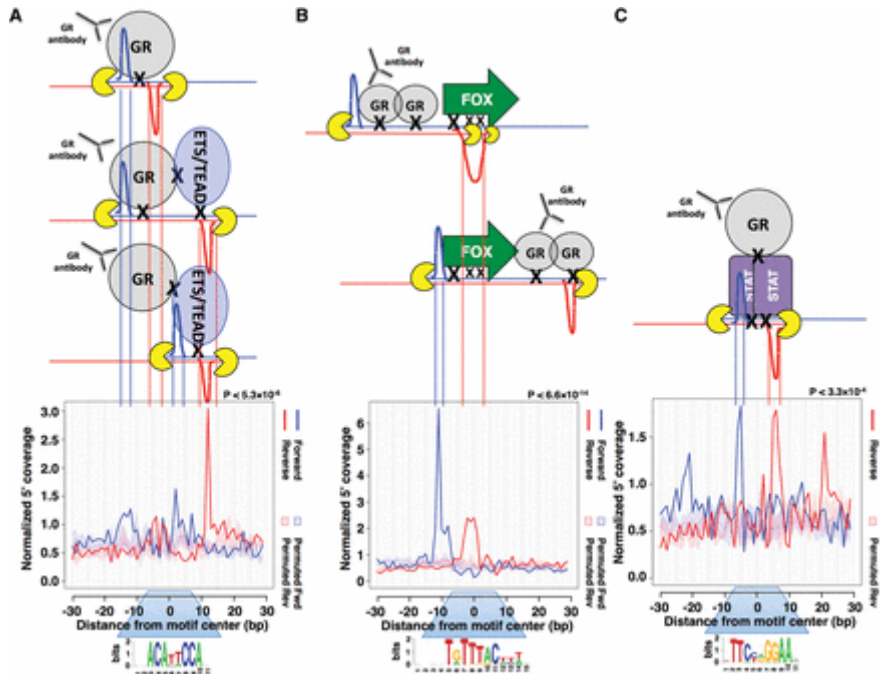
Adapted from Starick SR et al. 2015. ChIP-exo signal associated with DNA-binding motifs provides insight into the genomic binding of the glucocorticoid receptor and cooperating transcription factors

GATA2

Results

[Download all results (peak-motifs_archive.zip)] [Download all matrices (transfac format)]

Sequence composition & statistics
Discovered motifs (by algorithm) - 10
Discovered motifs (with motif comparison) - 10

Motif 1: [matrix] cayACACAyACACAc
Motif 2: [matrix] gasAGACAGAsAs
Motif 3: [matrix] acACATGCACryac
Motif 4: [matrix] mmycmwyCATCCATCaCCAyCwyy
Motif 5: [matrix] yaCACTCACay

Motif 6: [matrix] AyCATCACCATCAyc
Motif 7: [matrix] CACACATrCACAcA
Motif 8: [matrix] yyATCATCAyc
Motif 9: [matrix] ayATWCACAya
Motif 10: [matrix] rtGGGTGGAkr

RSAT peak-motif: no expected GATA2 motif was found.

STREME: no expected GATA2 motif found.

GATA4

Results

[Download all results (peak-motifs_archive.zip)] [Download all matrices (transfac format)]

Sequence composition & statistics
Discovered motifs (by algorithm) - 10
Discovered motifs (with motif comparison) - 10

Motif 1: [matrix] ccaCcCCACCCcs
Motif 2: [matrix] ssCCCGCCCCGCss
Motif 3: [matrix] gsACGCACGcv
Motif 4: [matrix] tyCTTATCTTtw
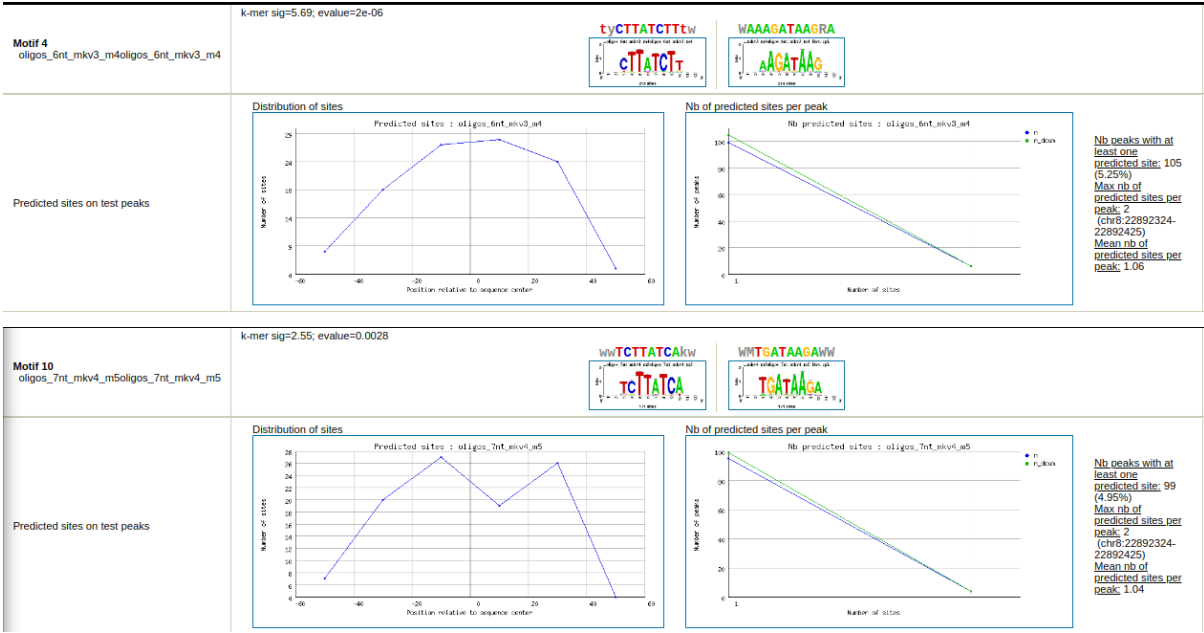Motif 5: [matrix] sgCGGAGCGss

Motif 6: [matrix] csGCCCCGCCCCcbs
Motif 7: [matrix] ssCGAGGCGsg
Motif 8: [matrix] sgCCCCACCCys
Motif 9: [matrix] sgGCGGAGCGsg
Motif 10: [matrix] wwTCTTATCAkw

Motif 4 oligos_6nt_mkv3_m4oligos_6nt_mkv3_m4	k-mer sig=5.69; evalue=2e-06	tyCTTATCTTtw WAAAGATAAGRA

Predicted sites on test peaks

Distribution of sites   Nb of predicted sites per peak

Nb peaks with at least one predicted site: 105 (5.25%)
Max nb of predicted sites per peak: 2 (chr8:22892324-22892425)
Mean nb of predicted sites per peak: 1.06

Motif 10 oligos_7nt_mkv4_m5oligos_7nt_mkv4_m5	k-mer sig=2.55; evalue=0.0028	wwTCTTATCAkw WMTGATAAGAWW

Predicted sites on test peaks

Distribution of sites   Nb of predicted sites per peak

Nb peaks with at least one predicted site: 99 (4.95%)
Max nb of predicted sites per peak: 2 (chr8:22892324-22892425)
Mean nb of predicted sites per peak: 1.04

RSAT peak-motif: expected motif found (Motif 4 and Motif 10).

DESCRIPTION

GATA4

DISCOVERED MOTIFS

Motif	Logo	RC Logo	P-value	E-value	Sites	More	Submit/Download	Positional Distribution	Matches per Sequence
1 - SGCCCCGCCYCC			9.4e-003	1.0e-001	147 (7.3%)				
2 - GCGCCTGCGCA			1.5e-002	1.7e-001	65 (3.3%)				
3 - GCMATAAAAWGYTW			1.5e-002	1.7e-001	57 (2.9%)				
4 - HGCTGGGAGCTGG			1.8e-002	2.0e-001	68 (3.4%)				
5 - AAAGCAAACA			2.2e-002	2.4e-001	175 (8.8%)				
6 - ATGTTTAC			3.0e-002	3.4e-001	76 (3.8%)				
7 - CAGCCCCACCCC			3.4e-002	3.7e-001	73 (3.6%)				
8 - AGATAAKR			3.5e-002	3.9e-001	109 (5.5%)				
9 - AGGAAGCA			3.8e-001	4.1e+000	255 (12.8%)				

STREME: expected motif found (Motif 8).

NR3C1

0h

Results

[Download all results (peak-motifs_archive.zip)] [Download all matrices (transfac format)]

Sequence composition & statistics
Discovered motifs (by algorithm) - 10
Discovered motifs (with motif comparison) - 10

Motif 1: [matrix] ATYrAATGGAATCrwmd   Motif 6: [matrix] aATyGAATGGAATCatCrdm
Motif 2: [matrix] kwCTGTACtkr   Motif 7: [matrix] TCCATTCGAtkm
Motif 3: [matrix] avaGTACACam   Motif 8: [matrix] rmATCAAATgg
Motif 4: [matrix] ssCCTCCCvb   Motif 9: [matrix] kwyGATTCCATTCgATk
Motif 5: [matrix] yCATTCGAkk   Motif 10: [matrix] rvTGACGCAsss

DESCRIPTION

NR3C1-0h

DISCOVERED MOTIFS

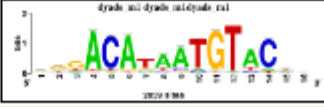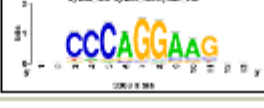| Motif	Logo	RC Logo	P-value	E-value	Sites	More	Submit/Download	Positional Distribution	Matches per Sequence
1-CTTTGTGWTGTKTGY			5.0e-001	1.5e+000	15 (19.5%)				
2-ATMAAATR			1.0e+000	3.0e+000	33 (42.9%)				
3-AATCCCAGCACTTT			1.0e+000	3.0e+000	14 (18.2%)				

Expected motif not found.
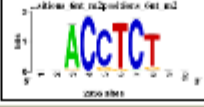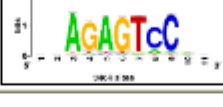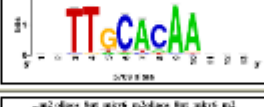
Immediately after treatment the NR3C1 is not bound to the DNA. That can be indicated by the absence of the motif and a very small number of peaks.

10min

DISCOVERED MOTIFS

Motif	Logo	RC Logo	P-value	E-value	Sites	More	Submit/Download	Positional Distribution	Matches per Sequence
1 - NRTGASTCAYN			3.8e-044	5.7e-043	2570 (25.7%)		→		
2 - RGRACABNVTGTYCY			4.7e-043	7.1e-042	2716 (27.2%)		→		
3 - WRARYAAAYAWWN			6.9e-019	1.0e-017	1832 (18.3%)		→		
4 - RTTGCAHAAY			1.3e-013	1.9e-012	1690 (16.9%)		→		
5 - CCCCRCCC			2.5e-009	3.8e-008	1021 (10.2%)		→		

Expected motif is found (STREME Motif 2).

Motif	Logo	3 Top hits in databases
oligos_6nt_mkv4_m1oligos_6nt_mkv4_m1		
oligos_6nt_mkv4_m2oligos_6nt_mkv4_m2		
oligos_6nt_mkv4_m3oligos_6nt_mkv4_m3		
local_words_6nt_m1local_words_6nt_m1		
local_words_6nt_m2local_words_6nt_m2		
local_words_6nt_m3local_words_6nt_m3		
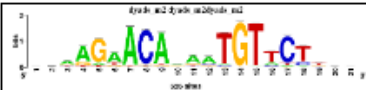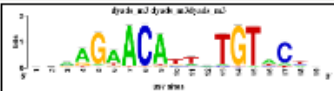
positions_6nt_m1positions_6nt_m1		
positions_6nt_m2positions_6nt_m2		
oligos_7nt_mkv5_m1oligos_7nt_mkv5_m1		
oligos_7nt_mkv5_m2oligos_7nt_mkv5_m2		
oligos_7nt_mkv5_m3oligos_7nt_mkv5_m3		
local_words_7ntlocal_words_7nt_m1		
positions_7nt_m1positions_7nt_m1		
positions_7nt_m2positions_7nt_m2		
positions_7nt_m3positions_7nt_m3		
oligos_8nt_mkv6_m1oligos_8nt_mkv6_m1		
oligos_8nt_mkv6_m2oligos_8nt_mkv6_m2		
oligos_8nt_mkv6_m3oligos_8nt_mkv6_m3		
local_words_8ntlocal_words_8nt_m1		
positions_8ntpositions_8nt_m1		
dyads_m1dyads_m1		
dyads_m2dyads_m2		
dyads_m3dyads_m3		

With RSAT peak-motif the expected motif is found with dyad algorithm.

8h

DISCOVERED MOTIFS

Motif	Logo	RC Logo	P-value	E-value	Sites	More	Submit/Download	Positional Distribution	Matches per Sequence
1 - RTTDWGMAATH			2.6e-033	3.7e-032	1715 (33.2%)				
2 - RGWACANNNTGTWCY			3.0e-030	4.2e-029	1689 (32.7%)				
3 - NVTGASTCABN			2.1e-017	3.0e-016	1140 (22.0%)				
4 - MAAARYAAACA			3.8e-008	5.4e-007	699 (13.5%)				
5 - AGTAAATAT			3.0e-006	4.2e-005	458 (8.9%)				

Expected motif found (STREME Motif 2), but pay attention to the number of sites! Less sites in the 8h timepoint.

Motif	Logo
oligos_6nt_mkv3_m1oligos_6nt_mkv3_m1	
oligos_6nt_mkv3_m2oligos_6nt_mkv3_m2	
oligos_6nt_mkv3_m3oligos_6nt_mkv3_m3	
positions_6ntpositions_6nt_m1	
oligos_7nt_mkv4_m1oligos_7nt_mkv4_m1	
oligos_7nt_mkv4_m2oligos_7nt_mkv4_m2	
oligos_7nt_mkv4_m3oligos_7nt_mkv4_m3	
oligos_8nt_mkv5_m1oligos_8nt_mkv5_m1	
oligos_8nt_mkv5_m2oligos_8nt_mkv5_m2	
oligos_8nt_mkv5_m3oligos_8nt_mkv5_m3	
dyads_m1dyads_m1	
dyads_m2dyads_m2	
dyads_m3dyads_m3	

Expected motif found with RSAT peak-motif as well, but full expected motif only with dyads algorithm.

Take home message

Always run your analysis at least two times, either using two different algorithms or changing your parameters, it is better to be more comprehensive and then reduce the redundant results that limit your analysis to a single run.