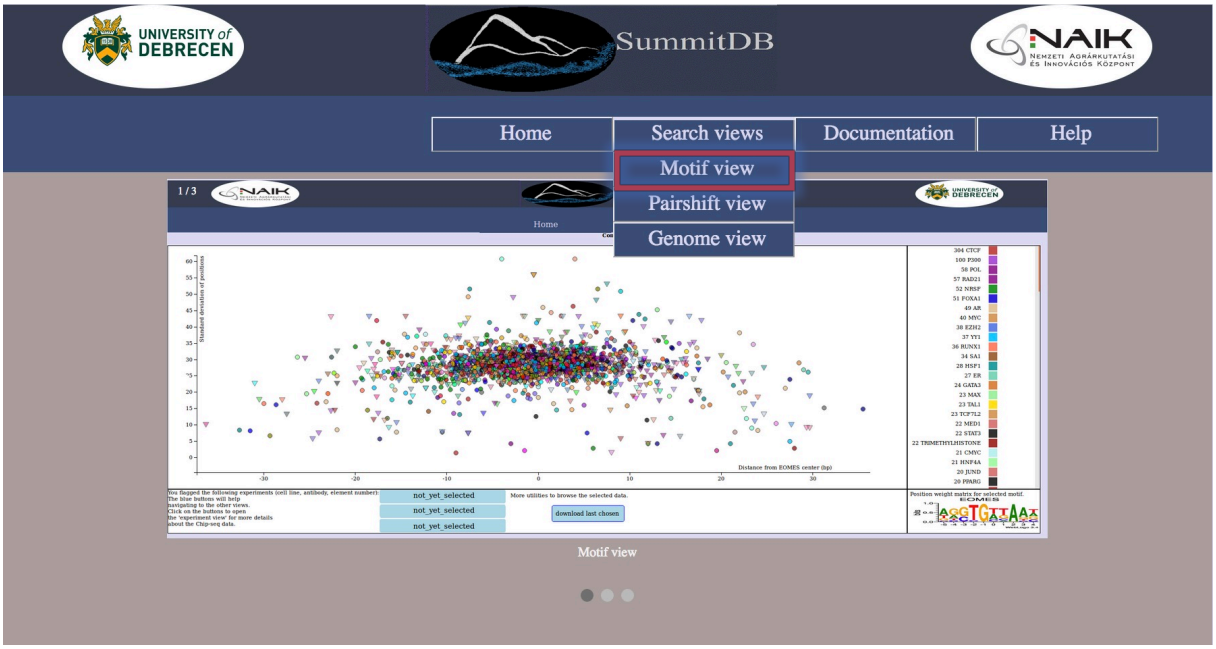
Tutorial:

Motif view

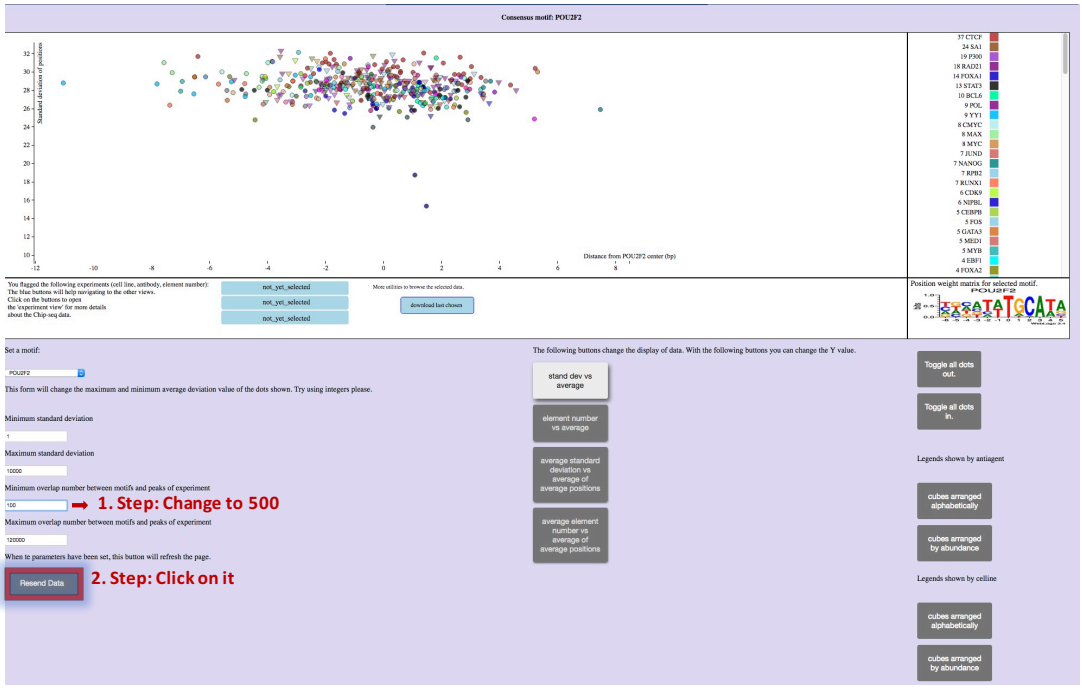
1. Go to the home page (http://summit.med.unideb.hu/summitdb/main.html). Click on the “Search views” menu and then click the “Motif view” button.



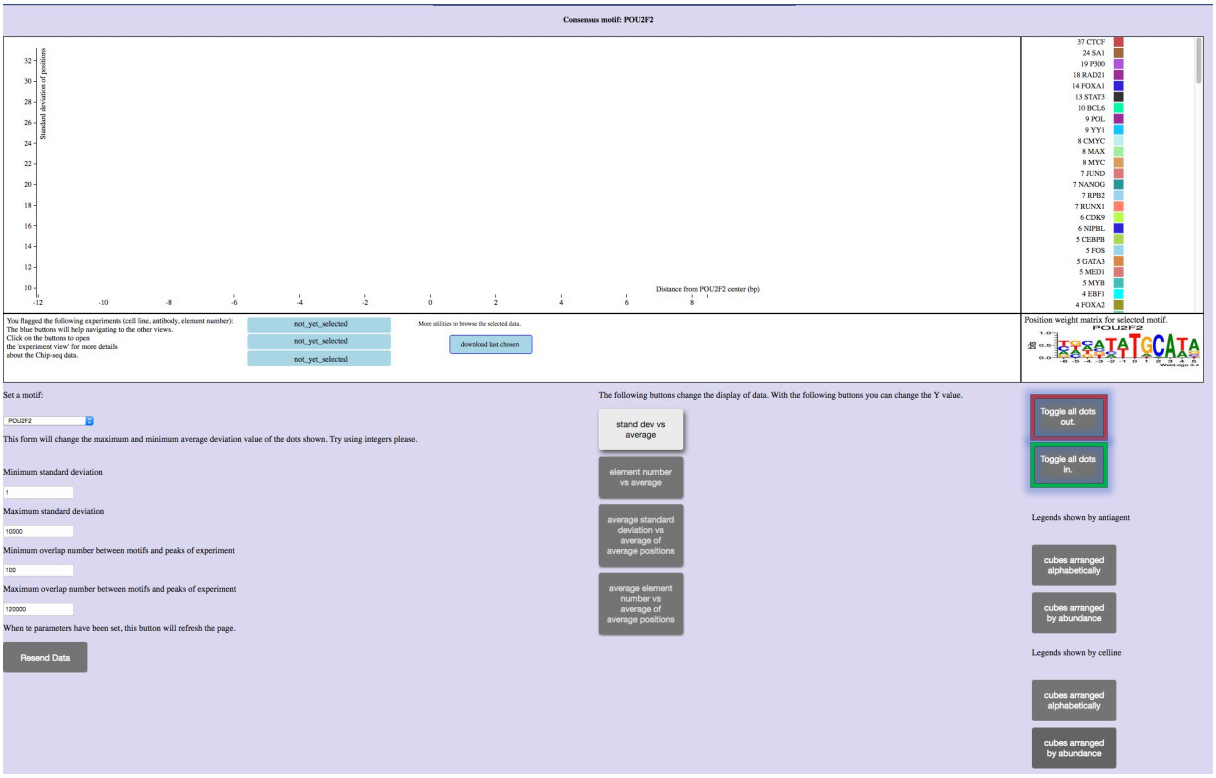
2. To guide through the different steps of using SummitDB the POU2F2 motif was used as enxample because it has a relatively small set of overlapping factors. To display the results of POU2F2 click on the dropdown box under the “Set a motif” subtitle.

3. Click on the “Go to motif view” button to update the page.

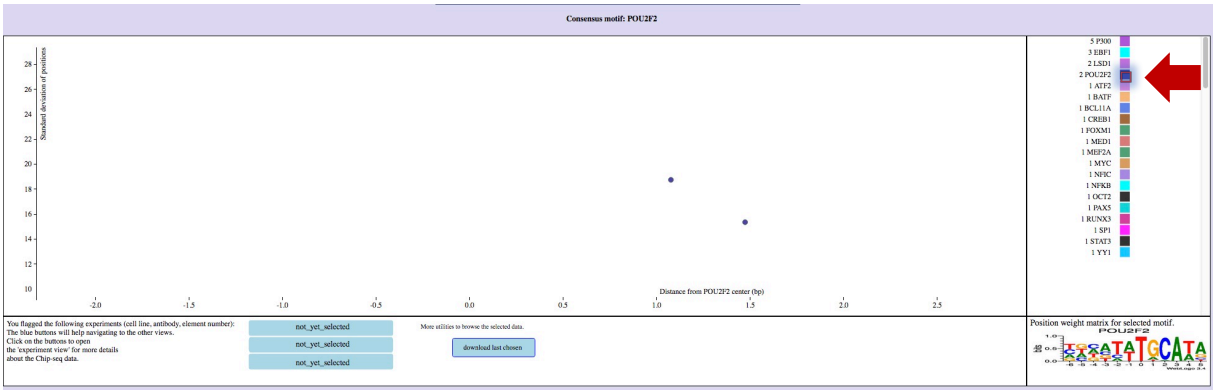
4. After updating the page a scatterplot for the POU2F2 motif can be seen. Filter out the experiments, which overlaps with POU2F2 less than 500 times. Change the default value, which is 100,) to 500 by clicking on the “Minimum overlap number between motifs and peaks of experiment” box. Click on the „Resend data” button, which will decrease the number of scatters to 28. This means that the identified POU2F2 binding sites overlap with peaks from 28 ChIP-seq experiments. Twenty colors are represented on the scatterplot, which is visible in the legend of the right-hand side of the page. Each scatter represent a target protein in the ChIp-seq experiment thus in this case 20 different Tfs from the 28 experiments can be distinguished.



5. For simplicity just investigate the TFs individually.. To hide all scatters from the plot, click on the “Clear all scatters out” button, which results in an empty graph. To undo the hiding click on the “Bring all scatters back” button. The user also or we canchoose individual data points to display by selecting the TF’s name.

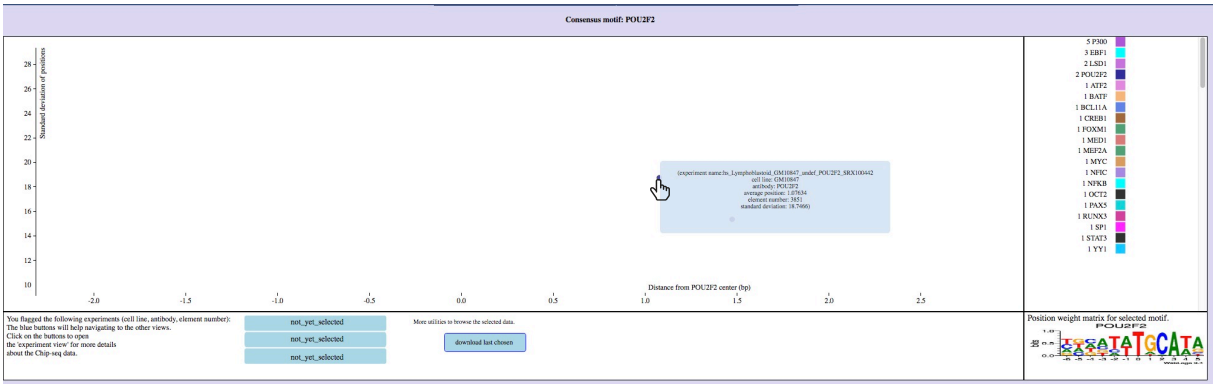


6. The legend in the right windows shows, that data from two POU2F2 ChIP-seq experiements are available. By clicking on the purple square next to“POU2F2”two scatters will appear on the graphbetween the values 0 and -1 on the x axis.



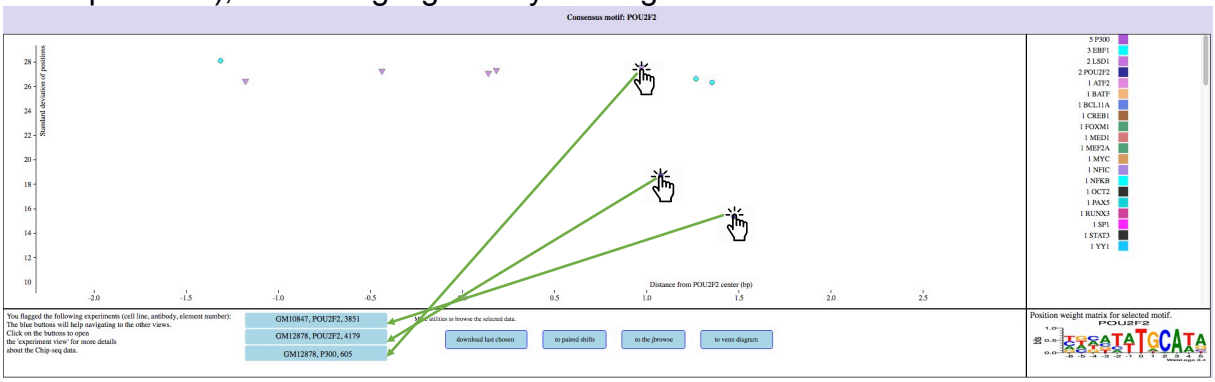
7. By moving the cursor over any of the two scatters,detailed information about the experiment will be seen in the appearing tool tip. The description indicates, that the data is from a ChIP-seq experiments in which lymphoblastoid cells,GM10847 and GM12878 cell lines, were used. The description highlights that the POU2F2 peaks from cell line GM12878 show an occupancy with 4,179 elements on the POU2F2 motif. It means that 4,179 POU2F2-binding motives can be found in the genome, on which sites the ChIP-seq experiment from GM12878 cells shows peaks. The tooltip shows both the punctual average position values and the standard deviation values. The average position is the average distance between the highest ChIP-seq read coverages, i.e. peak summits, and the motif centers. The motif centers are a fix

reference points on the genome, and the summits are the centers of regions covered by proteins. . By comparing these two positions, the position preferences of the protein can be elucidated. Regarding to the position values,1.07 and 1.47, in the example it can be concluded that POU2F2 covers the DNA downstream of the motif to which it binds, and also binds to its own regulatory region. The summit positions have a relatively low standard deviation (< 20). This indicates a close (tight?) DNA element–protein connection. It is supposed that a direct protein-DNA binding is more stabilized less mobile than an indirect binding between a co-factor and DNA. This postulate may explain the small variability between protein positions and is demonstrated by another example. If data from five experiments involving the P300 co-factor were displayed (which consists of 5 experiments) it could be seen that its binding positions have no side-specific preference, and the standard deviation is significantly higher. It is because P300 indirect binding to the motif is mediated by POU2F2. Another example for this phenomenon is protein EBF1. It has a different binding motif asPOU2F2, so itcan be hypothesized that P300 and EBF1bind to POU2F2and then thes protein complex is able to bind directly to the DNA binding site of POU2F2.



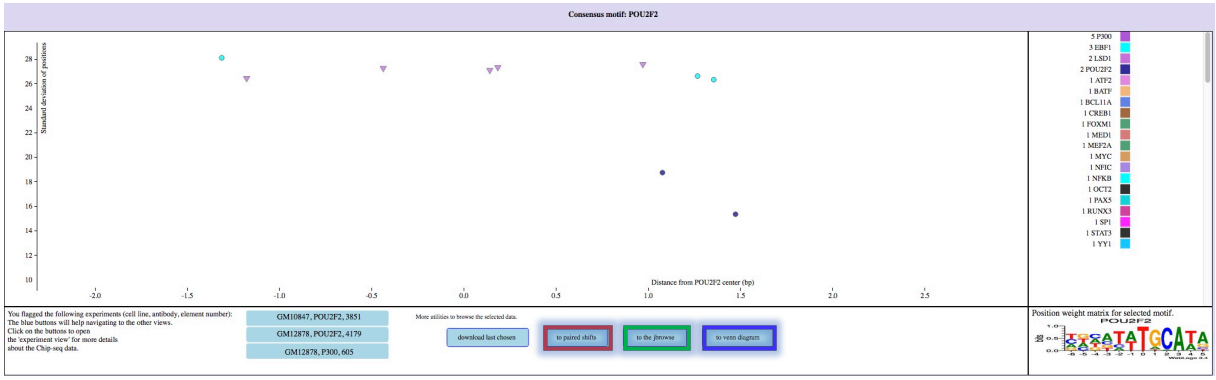
Interoperability between the different views

The motif view provides a global picture about the transcription factor binding sites and their occupying ChIP-seq signals on an interactive graph.
The scatters on the diagram, each representing one particular ChIP-seq experiment), can be highlighted by clicking on them.



The attributes, such as the cell line, antibody, SRA ID, and element number of maximum three selected experiments appear below the chart. To examine the the highlighted data, the following options can be used.

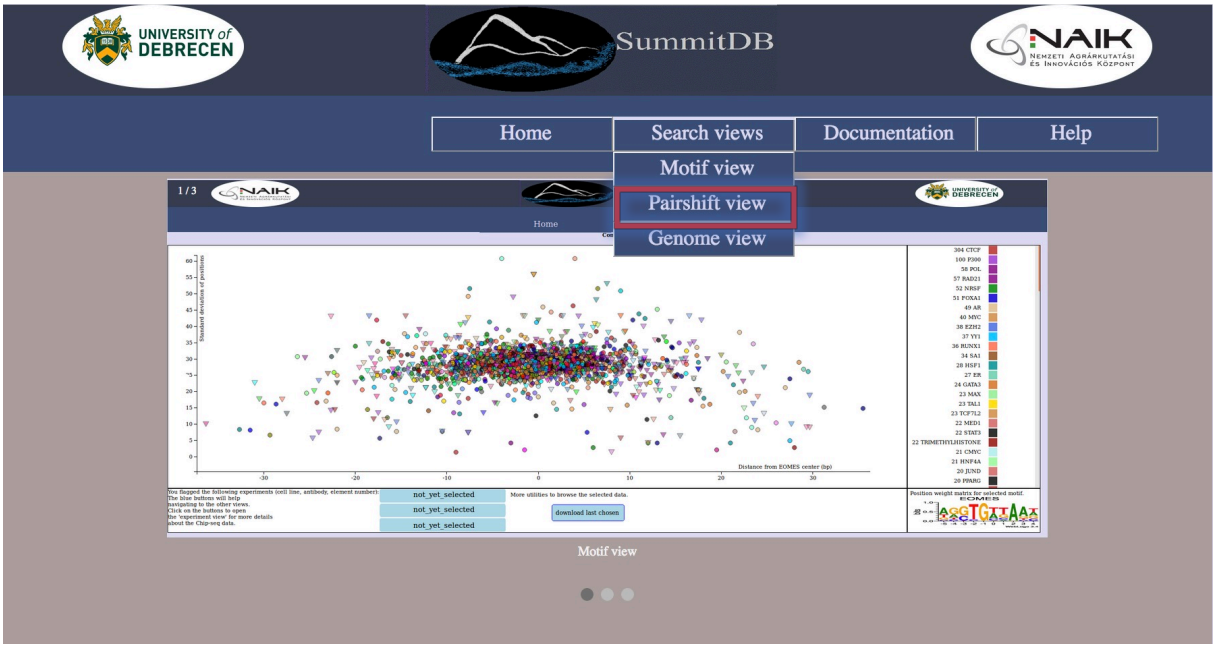
- A click on the “to pair shift view” button, will navigate to the distance distribution chart of the summit positions of the selected experiments compared to the adjusted motif, which isthe same motif that was displayed in the motif view.
- To browse the genomic locations of peak-motif co-occurrences in the genome browser, click on the “to the jbrowse” button.
- To see the frequency of co-occupancy between the selected factors, click on the “to Venn diagram” button. A Venn diagram will appear, in which the motif related co-appearance of summits can be seen.



Pair Shift view

To demonstrate this tool, the relation between the CTCF motif and the summits of two experiments from the MCF7 cell line will be sued as an example.

1. First, take the cursor over the “Search views” menu and click on the “Pair shift view“ button.

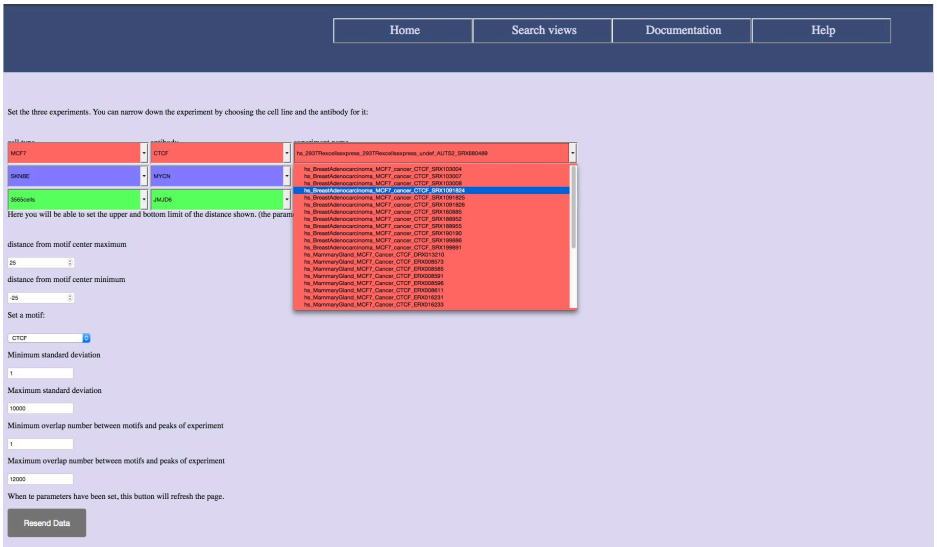


2. Select the CTCF motif under “Set a motif”.

3. Then click on the first upper left (red) dropdown box and select the “MCF7” cell type from the list. Click on the second box from the left in the row of red boxes, and select CTCF from the list. In the most rightward red box select the name of the experiment to “hs_BreastAdenocarcinoma_MCF7_cancer_CTCF_SRX1091824”.

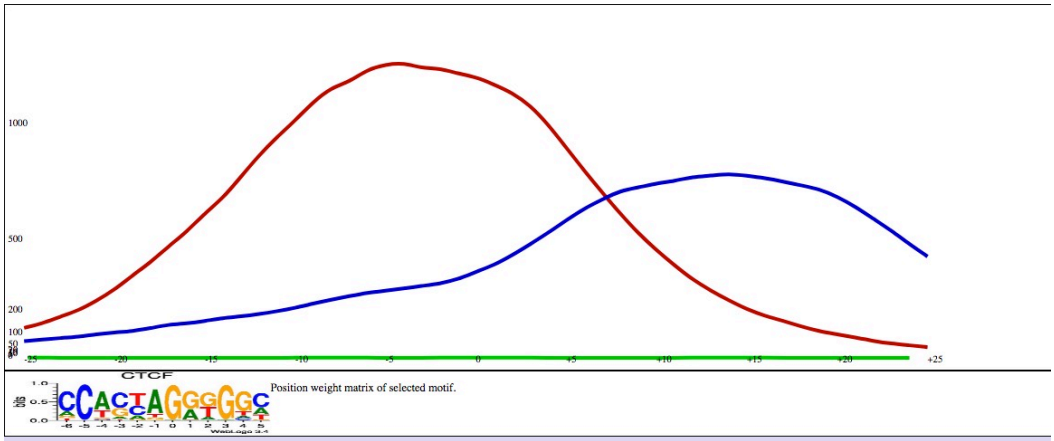

Home
Search views
Documentation
Help
LNCAP
LNCaPpbl
Luvcy
LoVo
LOVIMVI
LS174T
LS180
lung
lungcancercellline
LX2
LY1
LY2
lymphocyteCD20
MDTeC
MGH
macrophage
MCF
MCF10A
MCF10Acell
MCF7
AUTS2
hs_293TRescellexpress_293TRescellexpress_undef_AUTS2_SRX660469
SKNBE
MYCN
hs_2_SKNBE_undef_MYCN_SRX1178181
3565cells
JMJD6
hs_3565cells_3565cells_undef_JMJD6_SRX2270409
Here you will be able to set the upper and bottom limit of the distance shown. (the parameter set here is the max x and the negative of the minimum x value)
distance from motif center maximum
25
distance from motif center minimum
-25
Set a motif:
CTCF
Minimum standard deviation
1
Maximum standard deviation
10000
Minimum overlap number between motifs and peaks of experiment
1
Maximum overlap number between motifs and peaks of experiment
12000
When te parameters have been set, this button will refresh the page.
Resend Data


Home
Search views
Documentation
Help
AGO1
AHR
AHRR
AP2G
APOBEC3B
ARNT
BAF155
BORIS
CBP
CEBPB
CJUN
CMYC
CTCF
E2F1
ELF1
ER
ERA
ERARA
ESTROGEN
JMJD6
Set the three experiments. You can narrow d
and the antibody for it:
MCF7
SKNBE
3565cells
hs_293TRescellexpress_293TRescellexpress_undef_AUTS2_SRX660469
hs_2_SKNBE_undef_MYCN_SRX1178181
hs_3565cells_3565cells_undef_JMJD6_SRX2270409
Here you will be able to set the upper and bottom limit of the distance shown. (the parameter set here is the max x and the negative of the minimum x value)
distance from motif center maximum
25
distance from motif center minimum
-25
Set a motif:
CTCF
Minimum standard deviation
1
Maximum standard deviation
10000
Minimum overlap number between motifs and peaks of experiment
1
Maximum overlap number between motifs and peaks of experiment
12000
When te parameters have been set, this button will refresh the page.
Resend Data


Home
Search views
Documentation
Help
Set the three experiments. You can narrow down the experiment by choosing the cell line and the antibody for it:
MCF7
CTCF
hs_293TRescellexpress_293TRescellexpress_undef_AUTS2_SRX660469
SKNBE
MYCN
3565cells
JMJD6
distance from motif center maximum
25
distance from motif center minimum
-25
Set a motif:
CTCF
Minimum standard deviation
1
Maximum standard deviation
10000
Minimum overlap number between motifs and peaks of experiment
1
Maximum overlap number between motifs and peaks of experiment
12000
When te parameters have been set, this button will refresh the page.
Resend Data

4. In the next (blue) row of boxes repeat the previous steps, but select RRAD21 as the antibody, and “hs_BreastAdenocarcinoma_MCF7_cancer_RAD21_ERX004452” as the experiment name.
5. When you have finished setting the parameters, click on the “Resend Data” button.

6. The page will update and then the diagram will be displayed. The distribution curve for CTCF will be represented by red color. The curve for the second row will be in blue. If the curves do not fit properly in the diagram, then adjust the minimum and maximum values of the axes below the graph.



Set the three experiments. You can narrow down the experiment by choosing the cell line and the antibody for it:

cell type	antibody	experiment name
MCF7	CTCF	hs_BreastAdenocarcinoma_MCF7_cancer_CTCF_SRX1091824
MCF7	RAD21	hs_BreastAdenocarcinoma_MCF7_cancer_RAD21_ERX004452
3565cells	JMJD6	hs_3565cells_3565cells_undef_JMJD6_SRX2270409

Here you will be able to set the upper and bottom limit of the distance shown. (the parameter set here is the max x and the negative of the minimum x value)

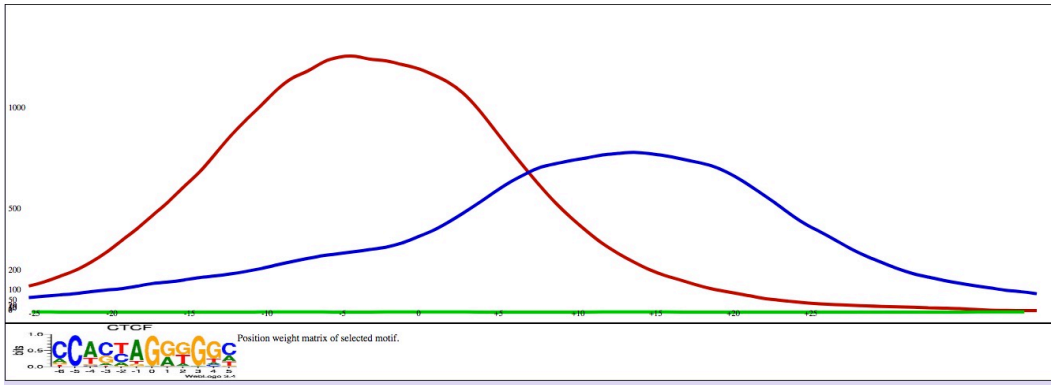
distance from motif center maximum

25 ➡ **Change to 40**

distance from motif center minimum

-25

Set a motif:



Set the three experiments. You can narrow down the experiment by choosing the cell line and the antibody for it:

cell type	antibody	experiment name
MCF7	CTCF	hs_BreastAdenocarcinoma_MCF7_cancer_CTCF_SRX1091824
MCF7	RAD21	hs_BreastAdenocarcinoma_MCF7_cancer_RAD21_ERX004452
3565cells	JMJD6	hs_3565cells_3565cells_undef_JMJD6_SRX2270409

Here you will be able to set the upper and bottom limit of the distance shown. (the parameter set here is the max x and the negative of the minimum x value)

distance from motif center maximum

40

distance from motif center minimum

-25

Set a motif:

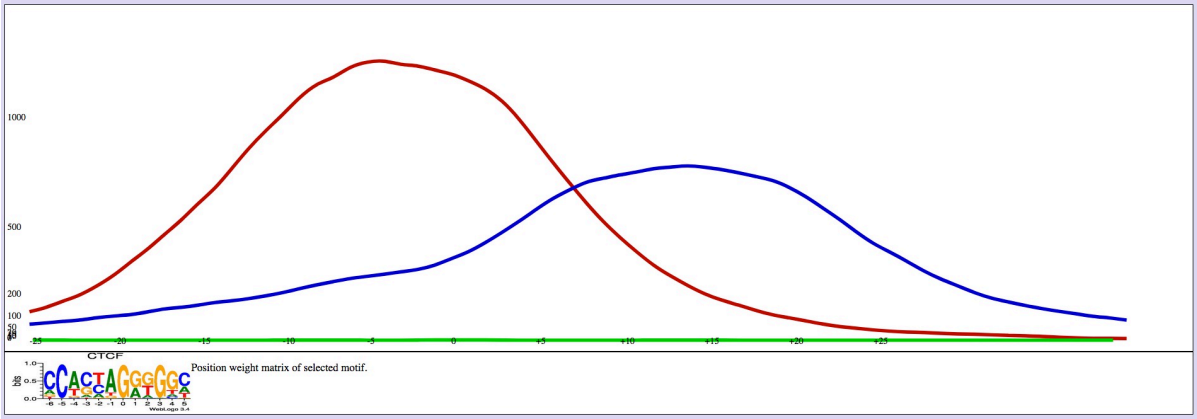
CTCF

The page will update and then the diagram will be displayed. The distribution curve for CTCF will be represented by red color. The curve for the second row will be in blue. If the curves do not fit properly in the

diagram, then adjust the minimum and maximum values of the axes below the graph.

It can be seen that the red curve (CTCF) is shifted towards the left-hand side of the CTCF motif, with a peak around position -5. In contrast, the RAD21 peak is shifted to the 3' direction of the motif, and has a local maxima around the position 15. Underlying these considerations was the assumption that the fine positional shifts that may exist between the contact points of cohesin proteins (CTCF, RAD21, SMC1/3 and STAG1/2) may reflect the 3D position of the components within the complex. Since CTCF is the only known specific DNA binder among the components of the CTCF/cohesin complex, we expected that the corresponding ChIP-seq peaks will point to the same position with respect to CTCF motif. In contrast, the fact that we can observe a positional shift suggests that RAD21 proteins occupy conserved – relatively fixed – positions that are close enough to the DNA so as to allow DNA-protein crosslinks to form during the ChIP-seq procedure.
The JARID1B is known as a histone demethylase enzyme. It is already described that a high fraction of JARID1B peaks overlap with CTCF binding sites in basal breast cancer cells (Yamamoto et al., 2014). The relationship between these two factors has also been investigated, as well as their relative effects on each other. According to the knock-down experiments, they found clear evidence for CTCF-JARID1B interactions, which suggests that the two proteins are present in the same complex.

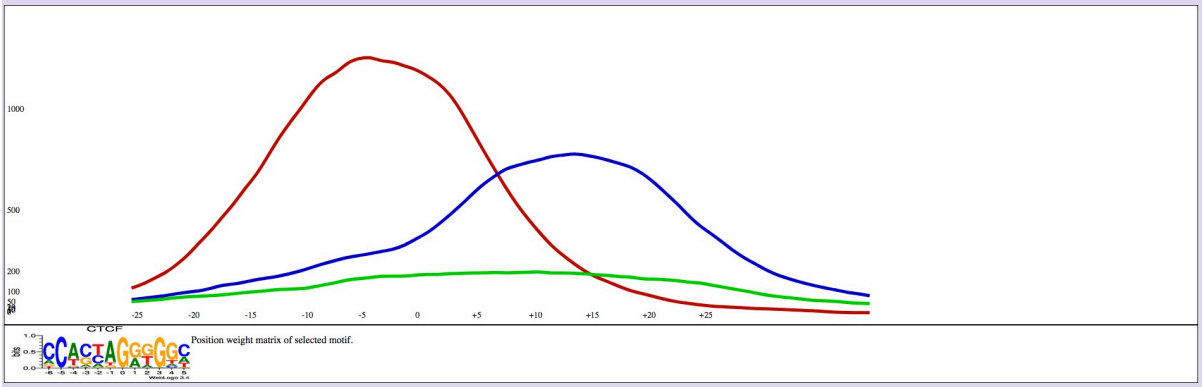
7. As another example, if "hs_BreastAdenocarcinoma_MCF7_cancer_JARID1B_SRX265412" was selected in the third row of green boxes as the third experiment in Pair shift view,a normal distribution curve in green can be seen, which indicates frequent CTCF-JARID1B co-appearance. Comparing this curve with the CTCF or RAD21 curves, it can be noticed that it is very flattened and broad. Consequently, the standard deviation of the positions for the JARID1B curve is higher (20.55), than that of the other two curves. This indicates that the interaction between the CTCF-binding motif and the JARID1B protein is very likely not direct, the enzyme connects to these binding sites through several other proteins.



Set the three experiments. You can narrow down the experiment by choosing the cell line and the antibody for it:

cell type	antibody	experiment name
MCF7	CTCF	hs_BreastAdenocarcinoma_MCF7_cancer_CTCF_SRX1091824
MCF7	RAD21	hs_BreastAdenocarcinoma_MCF7_cancer_RAD21_ERX004452
MCF7	JARID1B	hs_BreastAdenoCarcinoma_MCF7_cancer_JARID1B_SRX265412

Here you will be able to set the upper and bottom limit of the distance shown. (the parameter set here is the max x and the negative of the minimum x value)



Set the three experiments. You can narrow down the experiment by choosing the cell line and the antibody for it:

cell type	antibody	experiment name
MCF7	CTCF	hs_BreastAdenocarcinoma_MCF7_cancer_CTCF_SRX1091824
MCF7	RAD21	hs_BreastAdenocarcinoma_MCF7_cancer_RAD21_ERX004452
MCF7	JARID1B	hs_BreastAdenoCarcinoma_MCF7_cancer_JARID1B_SRX265412

Here you will be able to set the upper and bottom limit of the distance shown. (the parameter set here is the max x and the negative of the minimum x value)

distance from motif center maximum

40

distance from motif center minimum

-25

Set a motif: