

Contents

User Manual for CENTDIST	1
Login	2
Input	3
Custom PWM.....	5
Job Management	7
Output page.....	8

User Manual for CENTDIST

The web application CENTDIST is available at

<http://compbio.ddns.comp.nus.edu.sg/~chipseq/centdist/>.

It is simple. The user is just required to input a set of ChIP-seq peak locations and select a set of TRANSFAC motifs, grouped by our family classification (see Figure 2). Advanced users can create their own motifs and family classification using the “Use custom motif database” option.

After pressing the submit button, the user will be directed to the viewjobs page, which will show the status of submitted jobs (Pending, Running, Complete).

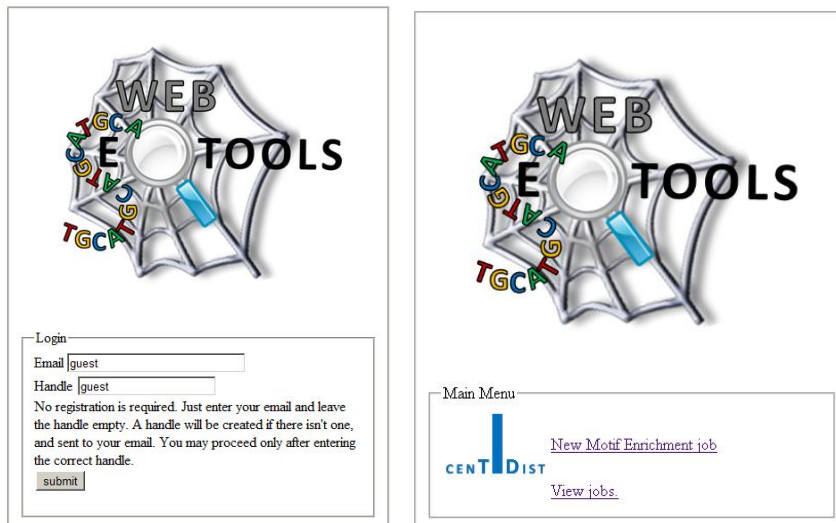
Once the status turn to be “Complete”, the user can click the “View Result” link, and then will be directed to the result page where he/she can obtain the list of candidate co-TFs (see Figure 5).

User also can refine the display and select motifs for further analysis.

Login

CENTDIST is the first born tool in our “Webseqtools” web server, and there will be more developed tools here. Before reaching CENTDIST, you will face the login page of “Webseqtools”. There is no registration requirement, just use guest account to login. After login, you can see the CENTDIST logo and the link ”New Motif Enrichment job” which leads you to CENTDIST input page. Also, there is a link called “View jobs”, which leads you to job management page.

Although no registration requirement, we encourage you to login using your email. If you login using your email address, then you can manage your jobs in any place and enable the email notification of job completion.



The figure displays two side-by-side screenshots of the WebSeqTools web interface. Both screenshots feature a large, stylized logo at the top that reads "WEB SEQ TOOLS" with a circular motif containing DNA base pairs (A, T, C, G) and a blue rectangular element. Below the logo, the left screenshot shows a "Login" section with two input fields labeled "Email" and "Handle", both containing the text "guest". Below these fields is a "submit" button and a paragraph of text: "No registration is required. Just enter your email and leave the handle empty. A handle will be created if there isn't one, and sent to your email. You may proceed only after entering the correct handle." The right screenshot shows a "Main Menu" section with the CENTDIST logo (a blue vertical bar next to the text "CENTDIST") and two links: "New Motif Enrichment job" and "View jobs".

Figure 1 Login and main menu of WebSeqTools

Input

For first time users, you can try CENTDIST using the sample dataset by clicking the single button **”Load sample ChIP-seq peaks”** which fill in all the necessary input area.

For general usage, users can follow the following 4 steps (see Figure 2):

- (1) User can paste their ChIP-seq peak coordinates as “chromosome-position pairs” or [bed format](#) genome regions into the input text area. CENTDIST is designed for high-through ChIP-seq experiment, and it can support more than 1 million ChIP-seq peaks.
- (2) User need to select the corresponding reference genome of your peak coordinates.
- (3) User can select the 849 matrices of TransFac database (version 6.0) or subgroups of them as enriched PWM candidates for further scanning and enrichment analysis.

CENTDIST allows user to filter the PWM name by their first characters, e.g. in the Figure, the PWMs in left box are filtered by first character equals to “V”, which mean vertebrate PWMs. User can enter the their filter criteria in the right textbox.

Moreover, CENTDIST allows user to choose the PWMs in one specified family. In the “Filter Group” dropdown, menu, you can select the TF family you are interested in.

After PWMs selection, you need to click “Add selected”, and make sure the PWMs you want appear in the right box. Also, you simply click “Add all” to add all PWMs in left box to the right box.

If users want to add their own PWMs, this can be done by selecting the checkbox ”Use custom motif database”. More detail can be found in the section “custom PWM”.

- (4) click “Submit Job” button.



CENTDIST

Title: Run Incap_AR sample

Upload peak coordinates

Load sample ChIP-seq peaks: Incap_AR

chr11 113555375
chr8 39808935
chr1 8241935
chr1 94982285
chr7 103613225
chr1 68783115
chr12 71734125
chr4 78704515
chr11 130791055
chr5 122761285
chr12 92732095
chr8 124487115
chr2 189146535
chr6 11058455
chr15 77857965
chr5 16600755
chr10 123870705
chr2 168843475
chr2 119812015
chr17 58157075

Genome Build hg18

Motif

Select

Filter group (ALL)

☒ Show only motif whose first characters are V

V\$ACAAT_B
V\$AFP1_Q6
V\$AHRARNT_01
V\$AHRARNT_02
V\$AHRHIF_Q6
V\$AHR_01
V\$AHR_Q5
V\$AIRE_01
V\$AIRE_02
V\$ALPHACP1_01

Add All

Add selected

Clear

V\$ACAAT_B
V\$AFP1_Q6
V\$AHRARNT_01
V\$AHRARNT_02
V\$AHRHIF_Q6
V\$AHR_01
V\$AHR_Q5
V\$AIRE_01
V\$AIRE_02
V\$ALPHACP1_01

☐ Use custom motif database

Submit Job

Figure 2 Input interface of CENTDIST

Custom PWM

After you select the checkbox "Use custom motif database", two more textbox will show up. The left box is the text content of the current motif database, and the right box is the text content of the current mapping between PWM and TF family.

☒ Use custom motif database
[Generate motif from matrix](#)

Motif Database

DE	V\$MYOD_01			
PO	A	C	G	T
01	1	2	2	0
02	2	1	2	0
03	3	0	1	1
04	0	5	0	0
05	5	0	0	0
06	0	0	4	1
07	0	1	4	0
08	0	0	0	5
09	0	0	5	0
10	0	1	2	2
11	0	2	0	3
12	1	0	3	1
XX				
DE	V\$E47_01			
PO	A	C	G	T
01	4	4	3	0
02	2	5	4	0
03	3	2	4	2

Motif Group File

V\$CI2_01	AAAAA
F\$ABF_C ABF	
F\$ABF1_01	ABF
F\$TAF_Q6	ABF
V\$ACAAAT_B	ACAAAT
I\$ADF1_Q6	ADF1
I\$ADF1_Q6_01	ADF1
P\$ABI4_01	ADF1
P\$ERF2_01	ADF1
V\$AFP1_Q6	AFP1
F\$MCM1_01	AGL
F\$MCM1_02	AGL
V\$AHR_01	AHR
V\$AHR_Q5	AHR
V\$AHRARNT_01	AHR
V\$AHRARNT_02	AHR
V\$AHRHIF_Q6	AHR
V\$AIRE_01	AIRE
V\$AIRE_02	AIRE
V\$ALX4_01	ALX4

Restore Default Motif Database

Restore Default Motif Group

Update Motif Select List

Figure 3 Add custom PWM library to CENTDIST

If you want to add your own PWMs, you need follow 4 steps (Figure 4):

- (1) Add you own PWMs follows the format below into the left box "Motif Database"

TRANSFAC-like	<p>This format is a simplification of the TRANSFAC format. The format of the matrix remains the same, but each matrix is directly preceded by a line beginning with a "DE" tag and followed by the motif name</p>	<pre>DE Mync PO A C G T 01 0 31 0 0 C 02 29 0 0 2 A 03 0 30 0 1 C 04 2 1 28 0 G 05 0 3 0 28 T 06 0 0 31 0 G XX</pre>
----------------------	---	--

- (2) [optional] define the mapping for your added PWMs to the TF families.
- (3) Click the “**Update Motif Select List**” button, you will see the refresh of the upper left box, which originally shows your filtered PWMs.
- (4) As usual, you need to add the PWMs you want to the upper right box by clicking “Add All” or “Add selected”.

Motif
Select

Filter group (ALL) ☐ Show only motif whose first characters are V

V\$ZIC3_01
V\$ZID_01
V\$ZNF219_01
V\$ZTA_Q2
custom_01
custom_02
custom_03
custom_04
custom_05
custom_06

Add All
Add selected
Clear

4

☒ Use custom motif database
[Generate motif from matrix](#)

Motif Database

05	1	2	0	2
06	0	5	0	0
07	5	0	0	0
08	0	0	5	0
09	0	5	0	0
10	0	0	1	4
11	0	1		0
12	0	2		2
13	1	0		1
14	0	0		0
15	1	1		2
16	1	4		0
17	2	1	1	1
18	0	0	3	2
XX				
DE	custom_06			
PO	A	C	G	T
01	0	3	1	1
02	0	0	1	4
03	0	4	0	1

Motif Group File

custom_01 AR
custom_02 AR
custom_03 ER
custom_04 ER
custom_05 FOX
custom_06 GATA

Restore Default Motif Database

Restore Default Motif Group

Update Motif Select List

3

Figure 4 Steps for using your custom PWM library in CENTDIST

Job Management

After submitting a job, you will be redirected to the job management page, which is the same page by click “view job” link after the login.

In this page, you will see a table showing the jobs that you have submitted.

Table 1 Definitions of each column in the view job page

Title	The title given to the job. You may change it by clicking the change job title link.
Type	Shows the type of job submitted. Currently, CENTDIST which has job type Motif_Enrichment is the only job type publicly available.
Submit Time	The time of submission. The page is sorted by submission time
Status	There are 4 statuses for a job: (1)PENDING: your job has already been assigned to the job queue, but not yet executed. (2)RUNNING: your job is running now. (3)COMPLETED: congratulation, your job is finished. Please click “ View Result ”. (4)ERROR: sorry, something wrong with your data or CENTDIST, please click “View log” to see the problem or send the log text to administrator to get help.
Actions	Here, you may View Log: Shows the execution log to get an idea of the progress of execution or the details of an error. View Result: Displays the results of the job Share Result: Share the result with others, not worrying about others getting access to your handle. Edit Job: Proceed to the submit page with the same set of parameters Delete Job: Delete unwanted jobs

At the top, you have the option to turn on auto-refresh which refreshes the page every 5 seconds, by clicking the button. Turn it off by clicking it again.

Below it shows the current load of the server.

Turn on autorefresh

Server status:
0 jobs running
0 jobs pending

Title	Type	Submit Time	Status	Actions
Run loci_p300 sample (Change title)	Motif_Enrichment	30/11/10 : 11:31:02	COMPLETED	View Log View Result Share Result Edit Job Delete Job
Run Incap_AR sample (Change title)	Motif_Enrichment	29/11/10 : 11:42:55	COMPLETED	View Log View Result Share Result Edit Job Delete Job
Run Incap_AR sample (Change title)	Motif_Enrichment	29/11/10 : 11:17:28	COMPLETED	View Log View Result Share Result Edit Job Delete Job
ctcf test (Change title)	Motif_Scan	26/11/10 : 13:26:54	COMPLETED	View Log View Result Share Result Edit Job Delete Job
Run Incap_AR sample (Change title)	Motif_Enrichment	24/11/10 : 18:02:05	COMPLETED	View Log View Result Share Result Edit Job Delete Job
p53 control combine (Change title)	Motif_Enrichment	24/11/10 : 13:02:42	COMPLETED	View Log View Result Share Result Edit Job Delete Job
Run loci_ctcf sample (Change title)	Motif_Enrichment	24/11/10 : 11:49:30	COMPLETED	View Log View Result Share Result Edit Job Delete Job
p53common peaks (Change title)	De_Novo	05/11/10 : 17:34:24	COMPLETED	View Log View Result Share Result Edit Job Delete Job
p53 control combine (Change title)	Motif_Enrichment	05/11/10 : 16:45:57	COMPLETED	View Log View Result Share Result Edit Job Delete Job
p53_control combine (Change title)	Motif_Scan	05/11/10 : 16:42:04	COMPLETED	View Log View Result Share Result Edit Job Delete Job
p53center with hbx (Change title)	Motif_Enrichment	04/11/10 : 16:51:01	COMPLETED	View Log View Result Share Result Edit Job Delete Job
				View Log

Figure 5 ViewJobs page of CENTDIST

Output page

For output, it is a table of PWM motifs ranked according to the center enrichment score. Each row in the output page presents the enriched TF family, and user can click the link of each TF family to browse the result of each member inside the family. The exact meanings of columns are shown below:

Table 2 Definitions of each column in the output page

1 st column	The rank of the given tf family
Name	The name of the best motif within the given tf family
Family	The name of the tf family
Logo	The seqlogo of the best motif within the given tf family
LogoR	The reverse complement of seqlogo of the best motif within the given tf family
Score	The center distribution score, which is used to rank by default, also is the sum of Z0 score and Z1 score.
Distribution	The motif distribution of the best motif (in the given tf family) around the peak center, the left side is center view, and the right side is folding view.
Range	The enriched window size (around the peak center) of the best motif (in the given tf family) detected by CENDIST.
Threshold	The best PWM threshold maximizing the center distribution score detected by CENDIST
%Motif above optimal cutoff within optimal binding range	When input peaks are sorted by quality(e.g, ChIP intensity), graph shows the moving average of the fraction of peaks containing a motif with score above the best PWM threshold within Range(in the Range column) from the center, as the quality of peak varies.
%Motif above 1e-4 cutoff within 200bp	When input peaks are sorted by quality(e.g, ChIP intensity), graph shows the moving average of the fraction of peaks containing a motif with score above the 1e-4 cutoff within 200bp from the center, as the quality of peak varies.
Z0Score	The z-score based on frequency graph of the motif distribution (the best motif of given tf family).
Z1Score	The z-score based on velocity graph of the motif distribution (the best motif of given tf family).
P-value	The empirical p-value of the given center distribution score.

Results for Run Incap_AR sample

606 TFs

Show top 50 Families Go Download As Text



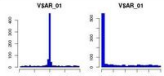




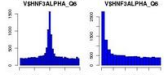




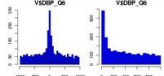




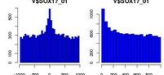




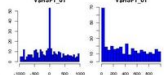




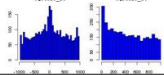


RANK	Name	Family	Logo	LogoR	Score	Distribution	%Motif above best cutoff within best binding range	%Motif above 1e-4 cutoff within 200bp	Binding Range	Threshold	ZScore	ZIScore
1	VSAR_01	AR			92.6652				40	3.0919	91.6751	0.990115
2	VSHNF3ALPHA_Q6	FOX			91.579				120	3.2831	85.5479	6.03114
3	V'SDBP_Q6	DBP			35.0132				80	2.754	35.3498	-0.336667
4	V'SOX17_01	SOX			23.1131				200	3.1747	20.0296	3.08344
5	V'SHSF1_01	HSE			17.111				40	3.4013	16.959	0.151926
6	V'SFAC1_01	EAC1			15.7941				160	2.881	14.125	1.66914

Figure 6 Output page of CENTDIST