hmChIP Manual

Analysis Procedure

1. Query available experiments and samples

1.1 Choose "species", "platform type", "GSE or Lab" and "ChIP type".

1.2 Provide protein names (e.g., Sox17) and cell or tissue types (e.g., stem cell). If no key words are provided, all experiments in the database will be returned.

1.3 (Optional) In the box titled "Overlap With Regions Below", provide a list of query genomic regions as a bait. The regions should be stored in a COD file (e.g., Sox17_peak_2rep.cod) or a BED file (e.g., UTA_cMyc_Combined.bed). Use the "Browse" button to choose the file. If a list of query genomic regions is provided, the returned experiments will be rank ordered based on the degree of overlap with the query genomic regions.

1.4 Click the "Query" button. Experiments matching the search criteria will be returned.

1.5 The query results will contain the following information:

(i) A download link to the peak list of each experiment (a COD file);

(ii) A download link to the log2 fold change profile of each experiment (a BAR file);

(iii) Download links to sample binding intensity profiles (BAR files).

Note: BAR files can be visualized by CisGenome Browser, and can be converted to TXT or WIG files by CisGenome. COD files can be used as input for a number of CisGenome analysis functions.

(iv) Quality measures for each sample. For ChIP-chip, the measures contain two numbers: percentage of probes covered by the peaks, and log2 (average probe intensity in peak regions) – log2 (average probe intensity in non-peak regions). For ChIP-seq, the measures are three numbers: total read count of the sample, percentage of genome covered by the peaks, log2 (average binding intensity in peak regions) – log2 (average binding intensity in non-peak regions).

(v) Links to the original GEO/SRA/ENCODE data set.

(vi) If a list of query genomic regions is used to search the database, the returned experiments are rank ordered based on the degree of overlap between the peak lists and the uploaded binding regions. In this scenario, there will be an additional column titled "Ranking Statistics" in the query results. This column shows the enrichment ratios

computed by comparing the observed overlap and the expected overlap, a p-value for testing if the observed overlap is significantly different from the expected overlap, and a multiple testing adjusted p-value (FDR, adjusted using Benjamini-Hochberg procedure).

2. Retrieve data

2.1 From the query results, check all samples you want to study.

2.2 Go to the section titled "Retrieve Data from Selected Samples". Provide a list of genomic regions for which you want to retrieve data. The regions should be stored in a COD file (e.g., Sox17_peak_2rep.cod) or a BED file (e.g., UTA_cMyc_Combined.bed). Choose this file using the "Browse" button next to the box titled "Upload Region".

2.3 Provide an email address. If no email address is provided, results will be returned on a webpage.

2.4 Click "Run". Data for input genomic regions will be retrieved from the selected samples.

2.5 Five files will be returned.

(i) A text file "result*.txt" that contains the raw binding intensities for user-provided genomic regions. "-1000" stands for missing value, i.e. no probes are found in the provided region.

(ii) A text file "resultnorm*.txt" that contains the normalized binding intensities.

(iii) A text file "fcbar*.txt" that contains the log2 IP/control fold changes

(iv) A text file "codIndex*.txt" that contains the binary binding indicators. For a genomic region and a peak list, 1 means the region is bound by protein in that peak list, and 0 means unbound.

(v) If your input region file contains less than 3000 lines and the number of selected samples is less than 60, a PDF file named "cluster*.pdf" will be returned. The file contains a heat map showing the hierarchical clustering of the samples and genomic regions based on the normalized intensities. The clustering uses average linkage, and 1-Pearson's correlation as the distance measure.

Note: If the number of your input regions is more than 3000 lines or the number of selected samples is more than 60, we will not return the PDF file with the heat map, since

the clustering may take significant amount of computational resource and may affect other users/jobs. We recommend you to do your own clustering using dChip.

EXAMPLES

Example S1. A step-by-step illustration of the analysis of mouse Sox17, Oct4, Sox2 and Nanog data. The Sox17 binding regions used in this analysis are obtained from Niakan et al. (2010) and can be downloaded from the hmChIP website. They are stored in a COD file. A COD file is a tab-delimited text file with at least five columns. The first five columns are region identifier, chromosome, region start, region end, and region strand information (the strand is not used in hmChIP analysis). Below is a sample line in the file:

geneA chr1 2000 2300 +

One can perform the hmChIP analysis as follows.

(a) Query mouse ChIP-chip samples. Choose species = mouse_mm8, platform type = ChIPchip_AffyAll, GSE or lab = All, ChIP Type = TF or DNA-binding proteins, protein name = Sox17 | Oct4 | Sox2 | Nanog, cell or tissue type = stem cell, then click the "Query" button. Note that multiple proteins and cell types can be queried together by using '|' (i.e. OR) to separate them.

(b) In the query results, all mouse ChIP-chip experiments and samples matching the key words are listed. Links to peak lists, log2 fold change profiles of experiments, and binding intensity profiles of samples are provided to download these data. In the quality measure column, two numbers are provided. They are percentage of genome covered by peaks, and log2 signal-to-noise ratio.

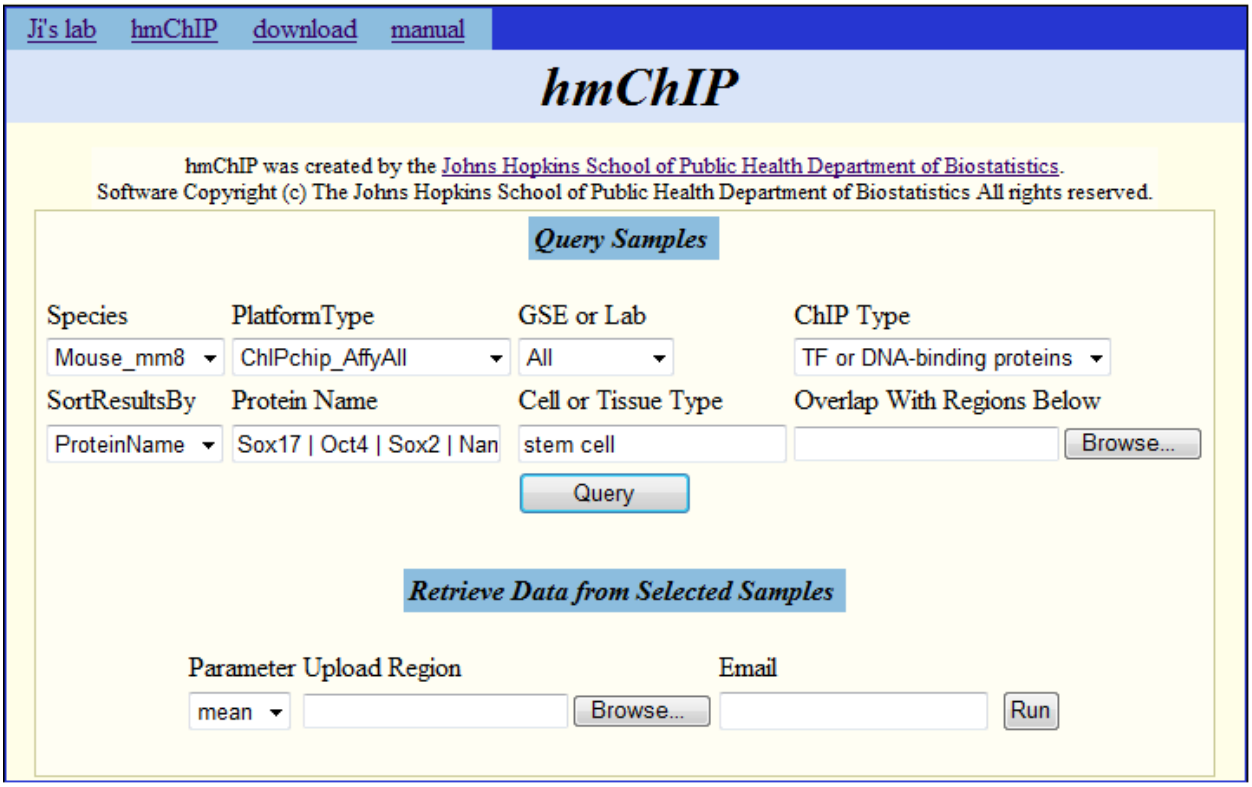
Select samples of interest from the query results. These include Sox17 IP and control samples in extraembryonic endoderm (XEN) stem cells (i.e., samples labeled with X-S and X-I), all Oct4, Sox2 and Nanog IP samples, and BirA and Input controls. Note that control samples labeled by "BirA" are shared by several experiments. Only one copy needs to be selected. The Sox17 samples labeled with M-F, M-S and M-I are not used in this analysis, since they are from mESC in which Sox17-DNA binding is weak.

(c) Go to the section titled "Retrieve Data from Selected Samples". Provide genomic coordinates of Sox17 binding sites in a COD file, and provide an email address. Click the "Run" button.

(d) The raw and normalized binding intensities (for each region and sample), binary binding indicators (one number for each region and peak list), and log2 IP/control fold changes (one for each region and peak list) will be returned by hmChIP through email. The data can be opened using Excel. Note that for some of the Sox17 binding regions, the binary indicators are zero even for the Sox17 peak list. This is because the Sox17 binding regions used in this analysis were obtained from Niakan et al. (2010), and they are slightly different from the Sox17 peak list in hmChIP which was generated using a different analysis procedure.

(e) The clustering heat map returned by hmChIP. In the heat map, a subset of Sox17 binding sites is co-bound by Sox2 and Nanog, whereas another subset is bound by Sox17 alone.

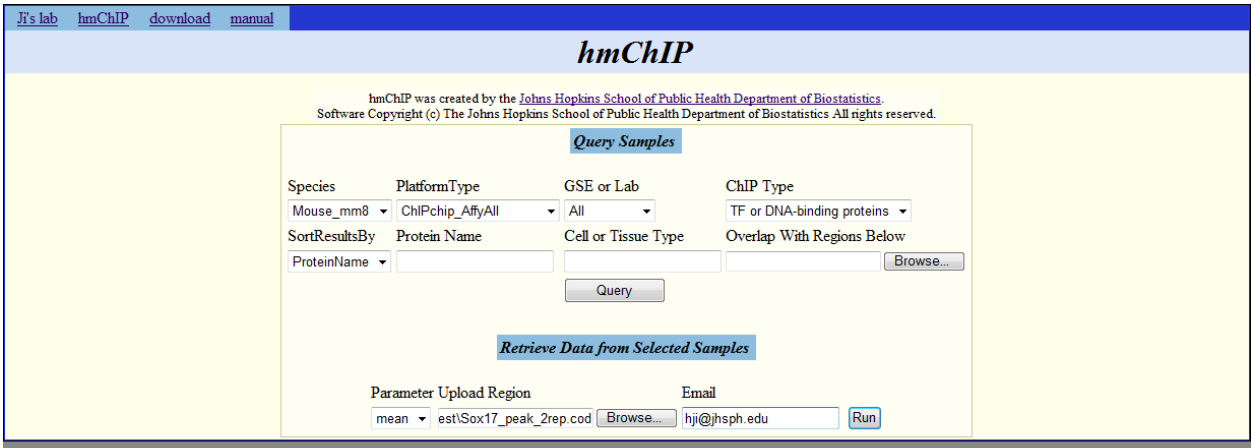
(a)



(b)

select	peakID	ProteinName	GSEorLab	peakfile	fcbar	Sample	QualityMeasure	Description
☐	1-IP	Sox17	GSE19026	peak	log2FC	GSM470844_M-S.bar.tpnorm.bar	0.003350/1.793589	Sox17 antibody ChIP in Sox17-induced mouse embryonic stem cell
☐	1-CT					GSM470844_M-I.bar.tpnorm.bar	0.003350/-0.282756	Sox17 antibody ChIP in Sox17-induced mouse embryonic stem cell
☐	2-IP	Sox17	GSE19026	peak	log2FC	GSM470845_M-F.bar.tpnorm.bar	0.003019/1.689871	Sox17 FLAG ChIP in Sox17-induced mouse embryonic stem cell
☐	2-CT					GSM470845_M-I_2.bar.tpnorm.bar	0.003019/-0.290433	Sox17 FLAG ChIP in Sox17-induced mouse embryonic stem cell
☑	3-IP	Sox17	GSE19026	peak	log2FC	GSM470843_X-S_1.bar.tpnorm.bar	0.005013/1.749960	Sox17 antibody ChIP in mouse extraembryonic endoderm stem cells
☑	3-IP					GSM470843_X-S.bar.tpnorm.bar	0.005013/1.835563	Sox17 antibody ChIP in mouse extraembryonic endoderm stem cells
☑	3-CT					GSM470843_X-I.bar.tpnorm.bar	0.005013/0.119038	Sox17 antibody ChIP in mouse extraembryonic endoderm stem cells
☑	3-CT					GSM470843_X-I_1.bar.tpnorm.bar	0.005013/0.169406	Sox17 antibody ChIP in mouse extraembryonic endoderm stem cells
☑	4-IP	Oct4	GSE11329	peak	log2FC	bOct4_1.bar.tpnorm.bar	0.001950/2.290329	Biotin-mediated Oct4 ChIP in mouse J1 embryonic stem cells
☑	4-IP					bOct4_2.bar.tpnorm.bar	0.001950/2.394215	Biotin-mediated Oct4 ChIP in mouse J1 embryonic stem cells
☑	4-IP					bOct4_3.bar.tpnorm.bar	0.001950/2.348637	Biotin-mediated Oct4 ChIP in mouse J1 embryonic stem cells
☑	4-CT					BirA_1.bar.tpnorm.bar	0.001950/0.125043	Biotin-mediated Oct4 ChIP in mouse J1 embryonic stem cells
☑	4-CT					BirA_2.bar.tpnorm.bar	0.001950/0.073325	Biotin-mediated Oct4 ChIP in mouse J1 embryonic stem cells
☑	4-CT					BirA_3.bar.tpnorm.bar	0.001950/0.002012	Biotin-mediated Oct4 ChIP in mouse J1 embryonic stem cells
☑	4-CT					BirA_4.bar.tpnorm.bar	0.001950/0.064496	Biotin-mediated Oct4 ChIP in mouse J1 embryonic stem cells
☑	5-IP	Sox2	GSE11329	peak	log2FC	bSox2_1.bar.tpnorm.bar	0.005551/1.693830	Biotin-mediated Sox2 ChIP in mouse J1 embryonic stem cells
☑	5-IP					bSox2_2.bar.tpnorm.bar	0.005551/0.985870	Biotin-mediated Sox2 ChIP in mouse J1 embryonic stem cells
☑	5-IP					bSox2_3.bar.tpnorm.bar	0.005551/2.533835	Biotin-mediated Sox2 ChIP in mouse J1 embryonic stem cells
☑	5-IP					bSox2_4.bar.tpnorm.bar	0.005551/1.771573	Biotin-mediated Sox2 ChIP in mouse J1 embryonic stem cells
☑	5-IP					bSox2_5.bar.tpnorm.bar	0.005551/2.158800	Biotin-mediated Sox2 ChIP in mouse J1 embryonic stem cells
☐	5-CT					BirA_1.bar.tpnorm.bar	0.005551/-0.012312	Biotin-mediated Sox2 ChIP in mouse J1 embryonic stem cells
☐	5-CT					BirA_2.bar.tpnorm.bar	0.005551/-0.029146	Biotin-mediated Sox2 ChIP in mouse J1 embryonic stem cells
☐	5-CT					BirA_3.bar.tpnorm.bar	0.005551/-0.149720	Biotin-mediated Sox2 ChIP in mouse J1 embryonic stem cells
☐	5-CT					BirA_4.bar.tpnorm.bar	0.005551/-0.023620	Biotin-mediated Sox2 ChIP in mouse J1 embryonic stem cells
☑	6-IP	Nanog	GSE11329	peak	log2FC	abNanog_1.bar.tpnorm.bar	0.010302/1.974034	Antibody Nanog ChIP in mouse J1 embryonic stem cells
☑	6-IP					abNanog_2.bar.tpnorm.bar	0.010302/2.383844	Antibody Nanog ChIP in mouse J1 embryonic stem cells
☑	6-IP					abNanog_3.bar.tpnorm.bar	0.010302/2.234422	Antibody Nanog ChIP in mouse J1 embryonic stem cells
☑	6-CT					Input_J1_1.bar.tpnorm.bar	0.010302/-0.136272	Antibody Nanog ChIP in mouse J1 embryonic stem cells
☑	6-CT					Input_J1_2.bar.tpnorm.bar	0.010302/-0.037026	Antibody Nanog ChIP in mouse J1 embryonic stem cells
☑	6-CT					Input_J1_3.bar.tpnorm.bar	0.010302/0.006191	Antibody Nanog ChIP in mouse J1 embryonic stem cells
☑	7-IP	Nanog	GSE11329	peak	log2FC	bioNanog_1.bar.tpnorm.bar	0.004940/2.587122	Biotin-mediated Nanog ChIP in mouse J1 embryonic stem cells
☑	7-IP					bioNanog_2.bar.tpnorm.bar	0.004940/3.057984	Biotin-mediated Nanog ChIP in mouse J1 embryonic stem cells
☑	7-IP					bioNanog_3.bar.tpnorm.bar	0.004940/2.813412	Biotin-mediated Nanog ChIP in mouse J1 embryonic stem cells
☐	7-CT					BirA_1.bar.tpnorm.bar	0.004940/0.101987	Biotin-mediated Nanog ChIP in mouse J1 embryonic stem cells
☐	7-CT					BirA_2.bar.tpnorm.bar	0.004940/0.066249	Biotin-mediated Nanog ChIP in mouse J1 embryonic stem cells
☐	7-CT					BirA_3.bar.tpnorm.bar	0.004940/-0.038397	Biotin-mediated Nanog ChIP in mouse J1 embryonic stem cells
☐	7-CT					BirA_4.bar.tpnorm.bar	0.004940/0.032157	Biotin-mediated Nanog ChIP in mouse J1 embryonic stem cells

(c)



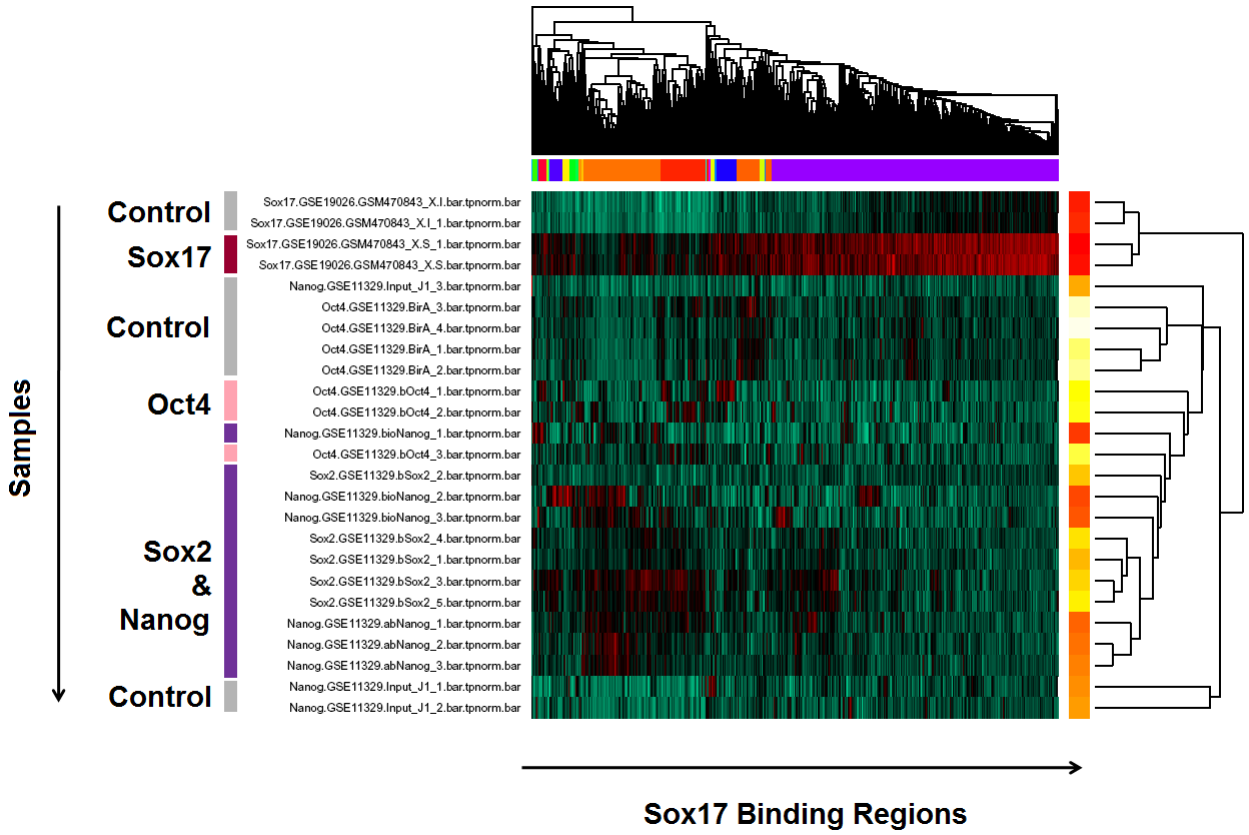
Search Result

select	peakID	ProteinName	GSEorLab	peakfile	fcbar	Sample	QualityMeasure	Description
☐	1-IP	Sox17	GSE19026	peak	log2FC	GSM470844_M-S.bar.tpnorm.bar	0.003350/1.793589	Sox17 antibody ChIP in Sox17-induced mouse embryonic stem cell
☐	1-CT					GSM470844_M-I.bar.tpnorm.bar	0.003350/-0.282756	Sox17 antibody ChIP in Sox17-induced mouse embryonic stem cell
☐	2-IP	Sox17	GSE19026	peak	log2FC	GSM470845_M-F.bar.tpnorm.bar	0.003019/1.689871	Sox17 FLAG ChIP in Sox17-induced mouse embryonic stem cell
☐	2-CT					GSM470845_M-I_2.bar.tpnorm.bar	0.003019/-0.290433	Sox17 FLAG ChIP in Sox17-induced mouse embryonic stem cell
☑	3-IP	Sox17	GSE19026	peak	log2FC	GSM470843_X-S_1.bar.tpnorm.bar	0.005013/1.749960	Sox17 antibody ChIP in mouse extraembryonic endoderm stem cells

(d)

	A	B	C	D	E	F	G	H	I	J
1	#seq_id	chromoso	start	end	strand	Sox17=GSE19026	Sox17=GSE1902	Sox17=GSE19026	Sox17=GSE19026	Nanog=GSE11329
2	1	chr8	11308223	11309596	+	2.430354	2.650095	0.582786	0.658697	0.459217
3	2	chr11	59424378	59424890	+	4.010035	4.206688	0.681039	0.656236	-0.355861
4	3	chr14	53572000	53572847	+	1.863982	2.292804	0.214124	0.080405	0.134098
5	4	chr7	44716885	44718184	+	1.215891	1.255014	-0.171202	-0.223582	0.6831
6	5	chr10	94753278	94754335	+	1.449498	1.609936	-0.116781	-0.129828	0.274486
7	6	chr16	92381143	92381630	+	2.480817	2.393408	0.543634	0.555498	-0.146107
8	7	chr4	35334997	35335849	+	2.479593	2.769188	0.381452	0.607598	-0.260825
9	8	chr15	96749917	96750821	+	3.010269	3.158312	0.82295	0.876143	0.703882
10	9	chr2	136583070	136584071	+	1.809312	1.990057	-0.149859	0.084732	0.449342
11	10	chr17	78413572	78414771	+	1.873767	1.946013	0.375127	0.318645	0.482946
12	11	chr14	92598244	92598984	+	1.993325	2.109167	0.273227	0.377136	-0.366827
13	12	chr11	95126067	95127111	+	1.507463	1.636327	0.031536	0.03849	1.462014

(e)



EXAMPLE S2. A step-by-step illustration of using Sox17 binding regions as input to search the database. The Sox17 binding regions used in this analysis are the same as previous.

(a) Query all mouse transcription factor data. Here we choose species = mouse_mm8, platform type = ChIPchip_or_ChIPseq, GSE or lab = All, ChIP Type = TF or DNA-binding proteins. We then choose a file in the box titled "Overlap With Regions Below" using the "Browse" button. For this example, we choose the file "Sox17_peak_2rep.cod" which contains Sox17 binding regions in XEN cells. Next, click the "Query" button. Now the Sox17 binding regions will be used as a bait to search the database.

(b) In the query results, all mouse TF ChIP-chip and ChIP-seq experiments are listed. These experiments are rank ordered based on the degree of overlap between their peak lists and the Sox17 binding regions. The "Ranking Statistics" column shows the enrichment ratios when comparing the observed overlap to the expected overlap. A p-value is computed for each experiment to test if the observed overlap is significantly different from the expected overlap. P-values are adjusted for multiple testing using the Benjamini-Hochberg procedure. The p-values and adjusted p-values (i.e. FDR) are shown as well.

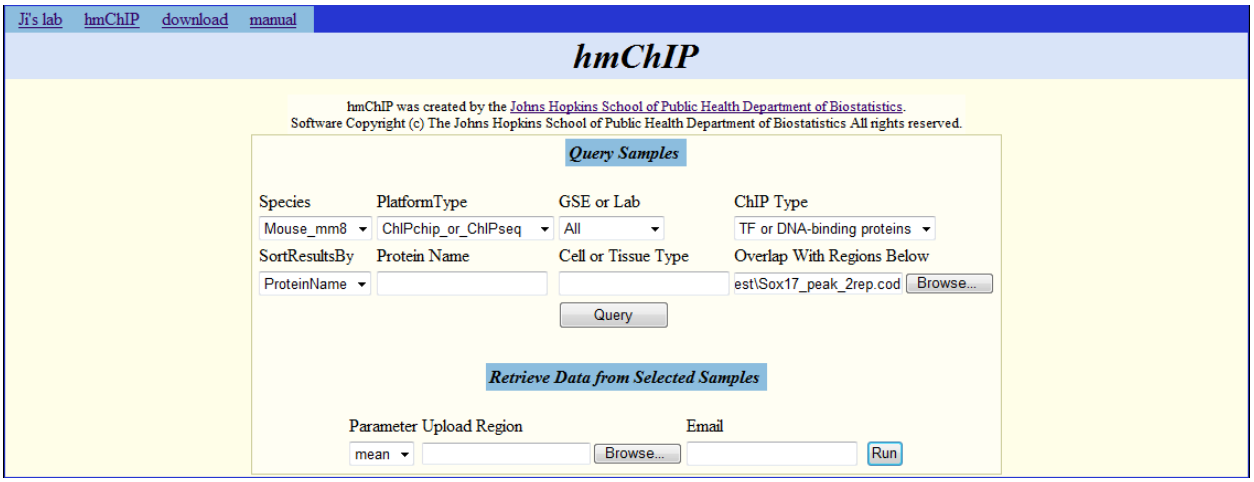
The results show that, other than Sox17 itself, the two TFs that overlap most with Sox17 binding sites are Sox2 and Nanog. From the query results, select samples from the experiments ranked at top 1 (Sox17 in XEN cells), top 4 (Sox2 in mESC), and top 5 (Nanog in mESC). These samples are all generated using the same platform (Affymetrix mouse promoter arrays) and can be meaningfully compared.

(c) Go to the "Retrieve Data from Selected Samples" section. Provide genomic coordinates of Sox17 binding regions, but do not provide an email address. Click "Run". Note: in general, the bait regions provided for database query in step (a) and the genomic regions provided for data retrieval in step (c) do not necessary need to be the same. Users can use one set of regions as bait to query the database, and retrieve data for another set of regions.

(d) Since no email address is provided, the data will be returned by hmChIP through a web page.

(e) The returned heat map again shows that a significant fraction of Sox17 binding sites in XEN cells are bound by Sox2 and Nanog in mESC.

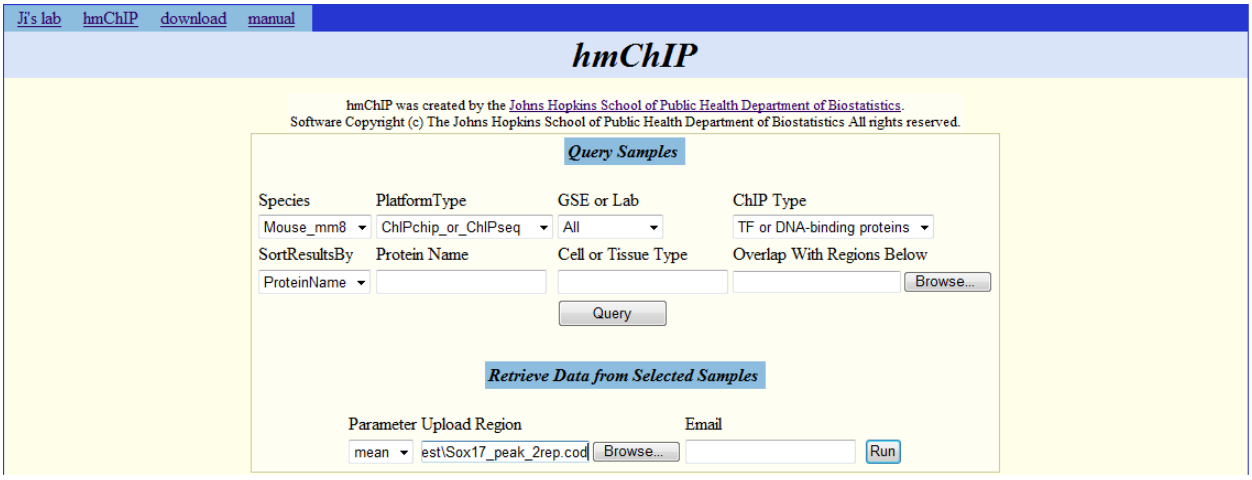
(a)



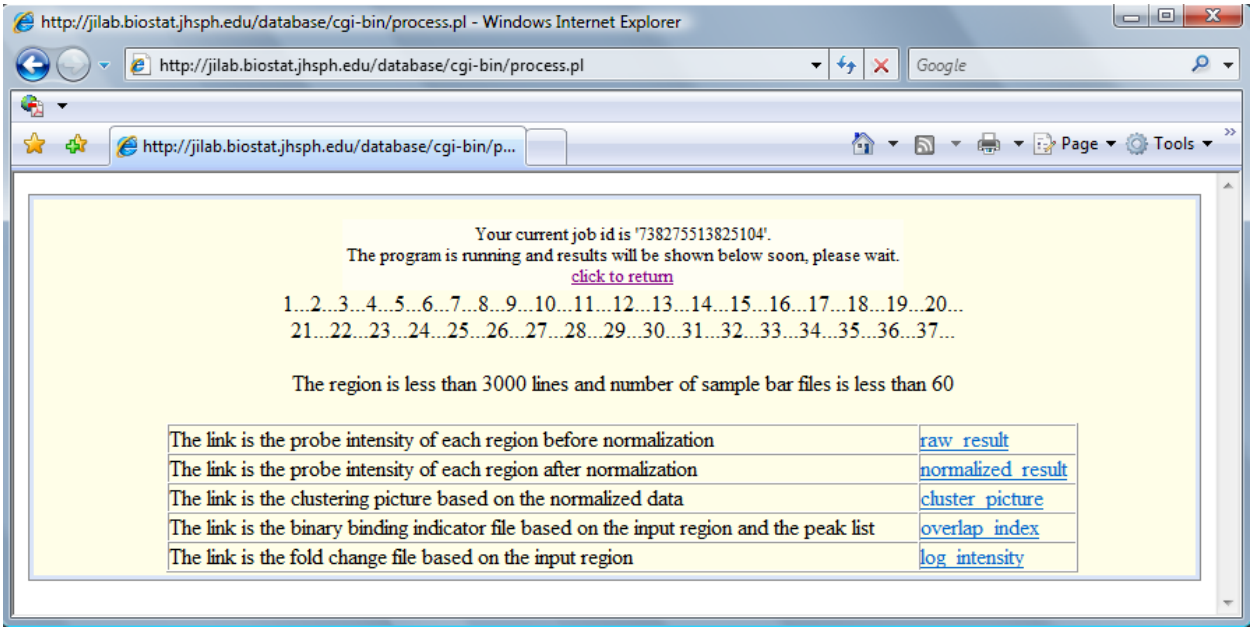
(b)

select	peakID	ProteinName	GSEorLab	peakfile	RankingStatistics	fcbar	Sample	Quality Measure	Description
☑	1-IP	Sox17	GSE19026	peak	r=73.690421 pvalue=0.000e+00 FDR=0.000e+00	log2FC	GSM470843_X-S_1.bar.tpnorm.bar	0.005013/1.749960	Sox17 antibody ChIP in mouse extraembryonic endoderm stem cells
☑	1-IP						GSM470843_X-S.bar.tpnorm.bar	0.005013/1.835563	Sox17 antibody ChIP in mouse extraembryonic endoderm stem cells
☑	1-CT						GSM470843_X-I.bar.tpnorm.bar	0.005013/0.119038	Sox17 antibody ChIP in mouse extraembryonic endoderm stem cells
☑	1-CT						GSM470843_X-I_1.bar.tpnorm.bar	0.005013/0.169406	Sox17 antibody ChIP in mouse extraembryonic endoderm stem cells
☐	2-IP	Sox17	GSE19026	peak	r=12.885457 pvalue=2.107e-28 FDR=2.239e-28	log2FC	GSM470844_M-S.bar.tpnorm.bar	0.003350/1.793589	Sox17 antibody ChIP in Sox17-induced mouse embryonic stem cell
☐	2-CT						GSM470844_M-I.bar.tpnorm.bar	0.003350/-0.282756	Sox17 antibody ChIP in Sox17-induced mouse embryonic stem cell
☐	3-IP	Sox17	GSE19026	peak	r=11.167574 pvalue=5.222e-26 FDR=5.765e-26	log2FC	GSM470845_M-F.bar.tpnorm.bar	0.003019/1.689871	Sox17 FLAG ChIP in Sox17-induced mouse embryonic stem cell
☐	3-CT						GSM470845_M-I_2.bar.tpnorm.bar	0.003019/-0.290433	Sox17 FLAG ChIP in Sox17-induced mouse embryonic stem cell
☑	4-IP	Sox2	GSE11329	peak	r=9.872396 pvalue=1.040e-33 FDR=1.091e-33	log2FC	bSox2_1.bar.tpnorm.bar	0.005551/1.693830	Biotin-mediated Sox2 ChIP in mouse J1 embryonic stem cells
☑	4-IP						bSox2_2.bar.tpnorm.bar	0.005551/0.985870	Biotin-mediated Sox2 ChIP in mouse J1 embryonic stem cells
☑	4-IP						bSox2_3.bar.tpnorm.bar	0.005551/2.533835	Biotin-mediated Sox2 ChIP in mouse J1 embryonic stem cells
☑	4-IP						bSox2_4.bar.tpnorm.bar	0.005551/1.771573	Biotin-mediated Sox2 ChIP in mouse J1 embryonic stem cells
☑	4-IP						bSox2_5.bar.tpnorm.bar	0.005551/2.158800	Biotin-mediated Sox2 ChIP in mouse J1 embryonic stem cells
☑	4-CT						BirA_1.bar.tpnorm.bar	0.005551/-0.012312	Biotin-mediated Sox2 ChIP in mouse J1 embryonic stem cells
☑	4-CT						BirA_2.bar.tpnorm.bar	0.005551/-0.029146	Biotin-mediated Sox2 ChIP in mouse J1 embryonic stem cells
☑	4-CT						BirA_3.bar.tpnorm.bar	0.005551/-0.149720	Biotin-mediated Sox2 ChIP in mouse J1 embryonic stem cells
☑	4-CT						BirA_4.bar.tpnorm.bar	0.005551/-0.023620	Biotin-mediated Sox2 ChIP in mouse J1 embryonic stem cells
☑	5-IP	Nanog	GSE11329	peak	r=7.378352 pvalue=1.729e-43 FDR=1.771e-43	log2FC	abNanog_1.bar.tpnorm.bar	0.010302/1.974034	Antibody Nanog ChIP in mouse J1 embryonic stem cells
☑	5-IP						abNanog_2.bar.tpnorm.bar	0.010302/2.383844	Antibody Nanog ChIP in mouse J1 embryonic stem cells
☑	5-IP						abNanog_3.bar.tpnorm.bar	0.010302/2.234422	Antibody Nanog ChIP in mouse J1 embryonic stem cells
☑	5-CT						Input_J1_1.bar.tpnorm.bar	0.010302/-0.136272	Antibody Nanog ChIP in mouse J1 embryonic stem cells
☑	5-CT						Input_J1_2.bar.tpnorm.bar	0.010302/-0.037026	Antibody Nanog ChIP in mouse J1 embryonic stem cells
☑	5-CT						Input_J1_3.bar.tpnorm.bar	0.010302/0.006191	Antibody Nanog ChIP in mouse J1 embryonic stem cells
☐	6-IP	Sox2	SRP000217	peak	r=6.634372 pvalue=6.714e-24 FDR=7.609e-24	log2FC	SRR002023_mm8_b35e150.bar	2222134/0.009825/1.223348	Sox2 binding in mouse embryonic stem cells (mESC

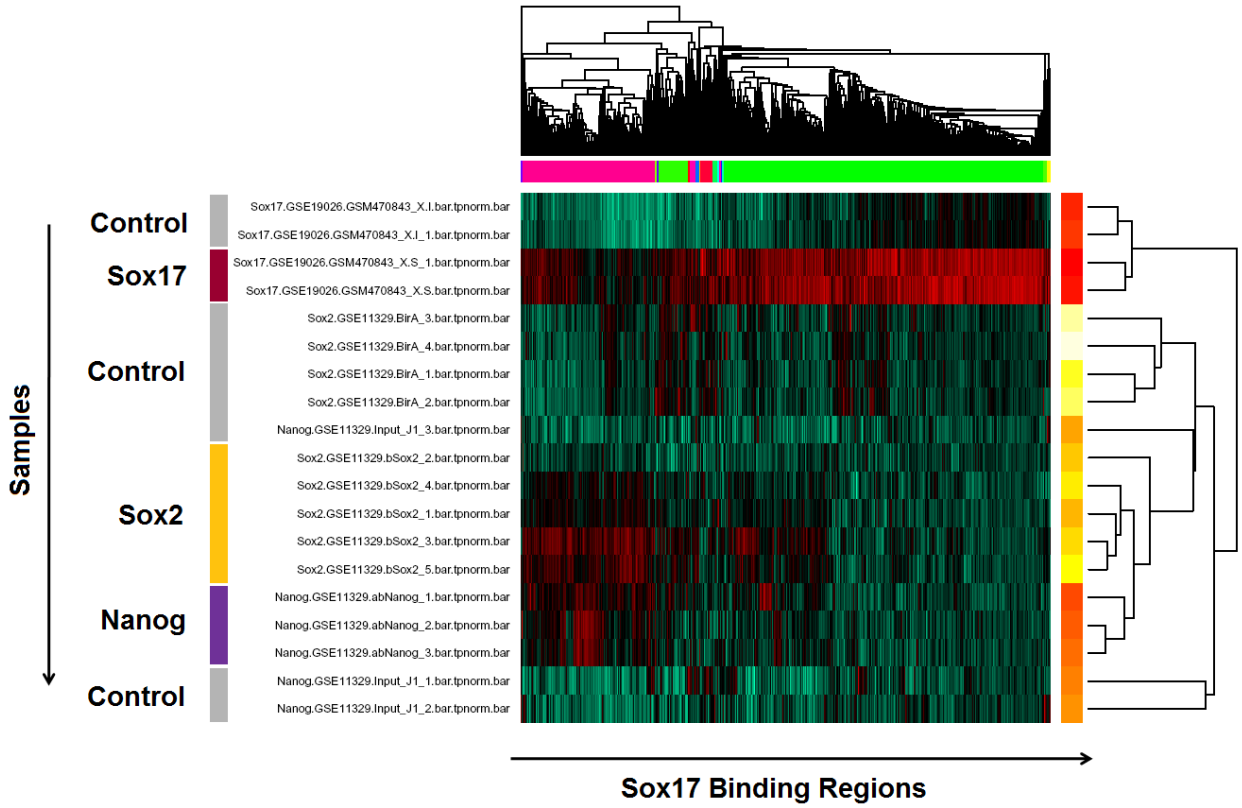
(c)



(d)


http://jilab.biostat.jhsph.edu/database/cgi-bin/process.pl - Windows Internet Explorer
http://jilab.biostat.jhsph.edu/database/cgi-bin/process.pl
Google
http://jilab.biostat.jhsph.edu/database/cgi-bin/p...
Page
Tools
Your current job id is '738275513825104'.
The program is running and results will be shown below soon, please wait.
click to return
1...2...3...4...5...6...7...8...9...10...11...12...13...14...15...16...17...18...19...20...
21...22...23...24...25...26...27...28...29...30...31...32...33...34...35...36...37...
The region is less than 3000 lines and number of sample bar files is less than 60
The link is the probe intensity of each region before normalization
raw_result
The link is the probe intensity of each region after normalization
normalized_result
The link is the clustering picture based on the normalized data
cluster_picture
The link is the binary binding indicator file based on the input region and the peak list
overlap_index
The link is the fold change file based on the input region
log_intensity

(e)


Samples
Control
Sox17
Control
Sox2
Nanog
Control
Sox17.GSE19026.GSM470843_X.I.bar.tpnorm.bar
Sox17.GSE19026.GSM470843_X.I_1.bar.tpnorm.bar
Sox17.GSE19026.GSM470843_X.S_1.bar.tpnorm.bar
Sox17.GSE19026.GSM470843_X.S.bar.tpnorm.bar
Sox2.GSE11329.BirA_3.bar.tpnorm.bar
Sox2.GSE11329.BirA_4.bar.tpnorm.bar
Sox2.GSE11329.BirA_1.bar.tpnorm.bar
Sox2.GSE11329.BirA_2.bar.tpnorm.bar
Nanog.GSE11329.Input_J1_3.bar.tpnorm.bar
Sox2.GSE11329.bSox2_2.bar.tpnorm.bar
Sox2.GSE11329.bSox2_4.bar.tpnorm.bar
Sox2.GSE11329.bSox2_1.bar.tpnorm.bar
Sox2.GSE11329.bSox2_3.bar.tpnorm.bar
Sox2.GSE11329.bSox2_5.bar.tpnorm.bar
Nanog.GSE11329.abNanog_1.bar.tpnorm.bar
Nanog.GSE11329.abNanog_2.bar.tpnorm.bar
Nanog.GSE11329.abNanog_3.bar.tpnorm.bar
Nanog.GSE11329.Input_J1_1.bar.tpnorm.bar
Nanog.GSE11329.Input_J1_2.bar.tpnorm.bar
Sox17 Binding Regions

EXAMPLE S3. A step-by-step illustration of the analysis of human c-Myc ChIP-seq data. The c-Myc binding regions used in this analysis can be downloaded from the hmChIP website. They were obtained by taking the union of the top 500 c-Myc binding regions reported by the ENCODE Iyer/UT-Austin lab in Helas3, Hepg2 and K562 cancer cell lines. The regions are stored in a BED file, which contains three columns separated by tabs. The columns are chromosome, region start, and region end.

(a) Query human ChIP-seq samples. Here we choose species = human_hg18, platform type = ChIPseq, GSE or lab = All, ChIP Type = TF or DNA-binding proteins, Protein Name = MYC. Click "Query".

(b) In the query results, all human MYC ChIP-seq experiments and samples are listed. Download links are provided for peak lists, log2 fold change profile of each experiment, and binding intensity profile of each sample. In the quality measure column, three numbers are provided. They are total read count of each sample, percentage of genome covered by peaks, and log2 signal-to-noise ratio respectively.

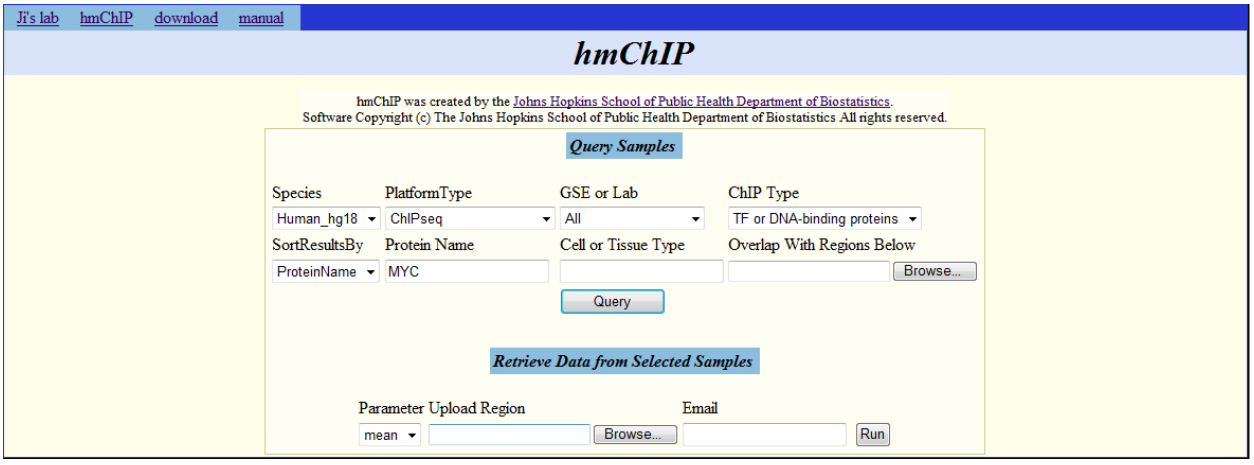
Now from the query results, select all MYC samples and their corresponding control samples from the two labs below: ENCODE_UTA and ENCODE_Yale.

(c) Go to the section titled "Retrieve Data from Selected Samples". Provide genomic coordinates of c-Myc binding regions, and provide an email address. Click "Run".

(d) The raw and normalized binding intensities, binary binding indicators, and log2 IP/control fold changes are returned through email. The data can be opened using Excel.

(e) The clustering heat map returned by hmChIP. The analysis automatically clusters ChIP-seq samples based on their cell types rather than lab origins (e.g. Helas3 samples produced by Yale and UTA are clustered together). The analysis also reveals different classes of Myc binding sites based on their cell type dependencies. A few classes are highlighted as examples.

(a)



(b)

Search Result

select	peakID	ProteinName	GSEorLab	peakfile	fcbar	Sample	QualityMeasure	Description
☑	1-IP	MYC	ENCODE_UTA	peak	log2FC	wgEncodeUtaChIPseqAlignmentsRep1Gm12878Cmyc_hg18.bar	10836823/0.010727/1.817533	c-Myc binding in CEPH sample GM12878 lymphoblastoid
☑	1-IP					wgEncodeUtaChIPseqAlignmentsRep2Gm12878Cmyc_hg18.bar	12032534/0.013654/1.594445	c-Myc binding in CEPH sample GM12878 lymphoblastoid
☑	1-CT					wgEncodeUtaustinChIPseqAlignmentsGm12878Input_hg18.bar	16161546/0.002004/-0.227702	c-Myc binding in CEPH sample GM12878 lymphoblastoid
☑	2-IP	MYC	ENCODE_UTA	peak	log2FC	wgEncodeUtaChIPseqAlignmentsHuvecCmyc_hg18.bar	12930115/0.010799/3.092577	c-Myc binding in HUVEC umbilical vein endothelial cells
☑	2-CT					wgEncodeUtaustinChIPseqAlignmentsHuvecInput_hg18.bar	13022336/0.002882/-0.376737	c-Myc binding in HUVEC umbilical vein endothelial cells
☑	3-IP	MYC	ENCODE_UTA	peak	log2FC	wgEncodeUtaChIPseqAlignmentsRep1Helas3CmycV2_hg18.bar	5736024/0.008548/1.934243	c-Myc binding in HeLa-S3 cervical carcinoma cells
☑	3-IP					wgEncodeUtaChIPseqAlignmentsRep2Helas3CmycV2_hg18.bar	6082670/0.008605/2.223525	c-Myc binding in HeLa-S3 cervical carcinoma cells
☑	3-CT					wgEncodeUtaustinChIPseqAlignmentsHelas3Input_hg18.bar	11361912/0.002936/-0.077083	c-Myc binding in HeLa-S3 cervical carcinoma cells
☑	4-IP	MYC	ENCODE_UTA	peak	log2FC	wgEncodeUtaChIPseqAlignmentsRep1Hepg2CmycV2_hg18.bar	8922221/0.014825/2.411414	c-Myc binding in HepG2 liver carcinoma cells
☑	4-IP					wgEncodeUtaChIPseqAlignmentsRep2Hepg2CmycV2_hg18.bar	5940327/0.018172/1.831657	c-Myc binding in HepG2 liver carcinoma cells
☑	4-IP					wgEncodeUtaChIPseqAlignmentsRep3Hepg2CmycV2_hg18.bar	5682849/0.019756/2.268820	c-Myc binding in HepG2 liver carcinoma cells
☑	4-CT					wgEncodeUtaustinChIPseqAlignmentsHepg2Input_hg18.bar	11452102/0.004621/-0.274919	c-Myc binding in HepG2 liver carcinoma cells
☑	5-IP	MYC	ENCODE_UTA	peak	log2FC	wgEncodeUtaChIPseqAlignmentsRep1K562CmycV2_hg18.bar	10603177/0.018626/1.977667	c-Myc binding in K562 leukemia cells
☑	5-IP					wgEncodeUtaChIPseqAlignmentsRep2K562CmycV2_hg18.bar	8688233/0.025489/2.104446	c-Myc binding in K562 leukemia cells
☑	5-IP					wgEncodeUtaChIPseqAlignmentsRep3K562CmycV2_hg18.bar	9522101/0.019449/2.205117	c-Myc binding in K562 leukemia cells
☑	5-CT					wgEncodeUtaustinChIPseqAlignmentsK562Input_hg18.bar	15960569/0.005732/0.204944	c-Myc binding in K562 leukemia cells
☑	6-IP	MYC	ENCODE_UTA	peak	log2FC	wgEncodeUtaChIPseqAlignmentsRep1Mcf7Cmyc_hg18.bar	16744594/0.006249/3.860773	c-Myc binding in MCF-7 adenocarcinoma cells
☑	6-IP					wgEncodeUtaChIPseqAlignmentsRep2Mcf7Cmyc_hg18.bar	10742069/0.007899/3.305770	c-Myc binding in MCF-7 adenocarcinoma cells
☑	6-CT					wgEncodeUtaustinChIPseqAlignmentsMcf7_hg18.bar	24538043/0.003726/1.108719	c-Myc binding in MCF-7 adenocarcinoma cells
☑	7-IP	MYC	ENCODE_Yale	peak	log2FC	wgEncodeYaleChIPseqAlignmentsRep1Helas3CmycV2_hg18_b35e150.bar	5575852/0.030790/2.428627	c-Myc binding in HeLa-S3 cervical carcinoma cells
☑	7-IP					wgEncodeYaleChIPseqAlignmentsRep2Helas3CmycV2_hg18_b35e150.bar	8911883/0.022673/3.280352	c-Myc binding in HeLa-S3 cervical carcinoma cells
☑	7-CT					wgEncodeYaleChIPseqAlignmentsHelas3Input_hg18_b35e150.bar	29840987/0.009625/1.525979	c-Myc binding in HeLa-S3 cervical carcinoma cells
☑	8-IP	MYC	ENCODE_Yale	peak	log2FC	wgEncodeYaleChIPseqAlignmentsRep1K562CmycIfna30V2_hg18_b35e150.bar	11069143/0.003480/3.640877	c-Myc binding in K562 leukemia cells (IFNa30)
☑	8-IP					wgEncodeYaleChIPseqAlignmentsRep2K562CmycIfna30V2_hg18_b35e150.bar	8944143/0.004036/2.803604	c-Myc binding in K562 leukemia cells (IFNa30)
☑	8-CT					wgEncodeYaleChIPseqAlignmentsRep1K562InputIfna30V2_hg18_b35e150.bar	16413656/0.002345/1.384908	c-Myc binding in K562 leukemia cells (IFNa30)
☑	9-IP	MYC	ENCODE_Yale	peak	log2FC	wgEncodeYaleChIPseqAlignmentsRep1K562CmycIfna6hV2_hg18_b35e150.bar	10818470/0.005770/3.600277	c-Myc binding in K562 leukemia cells (IFNa6h)
☑	9-IP					wgEncodeYaleChIPseqAlignmentsRep2K562CmycIfna6hV2_hg18_b35e150.bar	9970223/0.005838/2.860015	c-Myc binding in K562 leukemia cells (IFNa6h)
☑	9-CT					wgEncodeYaleChIPseqAlignmentsRep1K562InputIfna6hV2_hg18_b35e150.bar	17185067/0.003539/1.198873	c-Myc binding in K562 leukemia cells (IFNa6h)
☑	10-IP	MYC	ENCODE_Yale	peak	log2FC	wgEncodeYaleChIPseqAlignmentsRep1K562CmycIfng6hV2_hg18_b35e150.bar	7569091/0.013101/2.584348	c-Myc binding in K562 leukemia cells (IFNg6h)
☑	10-IP					wgEncodeYaleChIPseqAlignmentsRep2K562CmycIfng6hV2_hg18_b35e150.bar	10955943/0.012048/3.842483	c-Myc binding in K562 leukemia cells (IFNg6h)
☑	10-CT					wgEncodeYaleChIPseqAlignmentsRep1K562InputIfng6hV2_hg18_b35e150.bar	14897110/0.007093/1.211787	c-Myc binding in K562 leukemia cells (IFNg6h)
☑	11-IP	MYC	ENCODE_Yale	peak	log2FC	wgEncodeYaleChIPseqAlignmentsRep1K562Cmyc_hg18_b35e150.bar	5917263/0.018392/2.683290	c-Myc binding in K562 leukemia cells
☑	11-IP					wgEncodeYaleChIPseqAlignmentsRep2K562Cmyc_hg18_b35e150.bar	6175053/0.020628/3.323572	c-Myc binding in K562 leukemia cells
☑	11-CT					wgEncodeYaleChIPseqAlignmentsRep1K562InputV3_hg18_b35e150.bar	17519475/0.007960/1.498407	c-Myc binding in K562 leukemia cells
☑	11-CT					wgEncodeYaleChIPseqAlignmentsRep2K562InputV3_hg18_b35e150.bar	16267199/0.008563/1.506334	c-Myc binding in K562 leukemia cells

Select All | Clear All

(c)

Ji's lab | hmChIP | download | manual

hmChIP

hmChIP was created by the Johns Hopkins School of Public Health Department of Biostatistics.
Software Copyright (c) The Johns Hopkins School of Public Health Department of Biostatistics All rights reserved.

Query Samples

Species	PlatformType	GSE or Lab	ChIP Type
Human_hg18	ChIPseq	All	TF or DNA-binding proteins

SortResultsBy	Protein Name	Cell or Tissue Type	Overlap With Regions Below
ProteinName	MYC		Browse...

Query

Retrieve Data from Selected Samples

Parameter Upload Region: mean | JTA_cMyc_Combined.bed | Browse...

Email: hji@jhsph.edu | Run

Search Result

select	peakID	ProteinName	GSEorLab	peakfile	fcbar	Sample	QualityMeasure	Description
☑	1-IP	MYC	ENCODE_UTA	peak	log2FC	wgEncodeUtaChIPseqAlignmentsRep1Gm12878Cmyc_hg18.bar	10836823/0.010727/1.817533	c-Myc binding in CEPH sample GM12878 lymphoblastoid

(d)

	A	B	C	D	E	F	G	H	I
1	#seq_id	chromoso	start	end	strand	MYC=ENCODE_Yale=	MYC=ENCODE_Yale	MYC=ENCODE_Ya	MYC=ENCODE_Yal
2	1	chr7	4647775	4648617	+	2.859512	3.170291	0.497713	0.712729
3	2	chr17	39503373	39503917	+	2.849976	3.209385	0.42497	0.776677
4	3	chrX	9391009	9391496	+	3.037474	3.371766	0.179526	0.494252
5	4	chr5	154114320	154114991	+	2.777633	2.94468	0.456201	0.805428
6	5	chr17	46585666	46586365	+	3.030274	3.301313	0.933392	0.750249
7	6	chr1	157251790	157252243	+	1.408602	2.100786	-0.648345	0.207311
8	7	chr17	74369645	74370312	+	2.245247	2.806379	0.498017	0.362151
9	8	chr7	154589347	154589807	+	0.949942	1.220329	0.004728	-0.171605
10	9	chr6	45521547	45521990	+	2.657684	2.983398	0.819505	0.595195
11	10	chr3	185449822	185450358	+	2.801525	2.935857	0.535085	0.833683
12	11	chr11	9737695	9738181	+	1.300227	2.017859	-0.105243	0.612093
13	12	chr9	92999283	92999743	+	1.907902	2.095313	-0.027936	-0.291591

(e)