doi: 10.1093/nar/gkz302

SEanalysis: a web tool for super-enhancer associated regulatory analysis

Feng-Cui Qian[1,†], Xue-Cang Li[1,†], Jin-Cheng Guo[2,†], Jian-Mei Zhao[1], Yan-Yu Li[1], Zhi-Dong Tang[1], Li-Wei Zhou[1], Jian Zhang[1], Xue-Feng Bai[1], Yong Jiang[1], Qi Pan[1], Qiu-Yu Wang[2,1], En-Min Li[1], Chun-Quan Li[1,*], Li-Yan Xu[2,*] and De-Chen Lin[3,*]

[1]School of Medical Informatics, Daqing Campus, Harbin Medical University, Daqing 163319, China, [2]Institute of Oncologic Pathology, Medical College of Shantou University, Shantou 515041, China and [3]Guangdong Province Key Laboratory of Malignant Tumor Epigenetics and Gene Regulation, Sun Yat-Sen Memorial Hospital, Sun Yat-Sen University, Guangzhou 510120, China

Received January 25, 2019; Revised April 04, 2019; Editorial Decision April 13, 2019; Accepted April 18, 2019

ABSTRACT

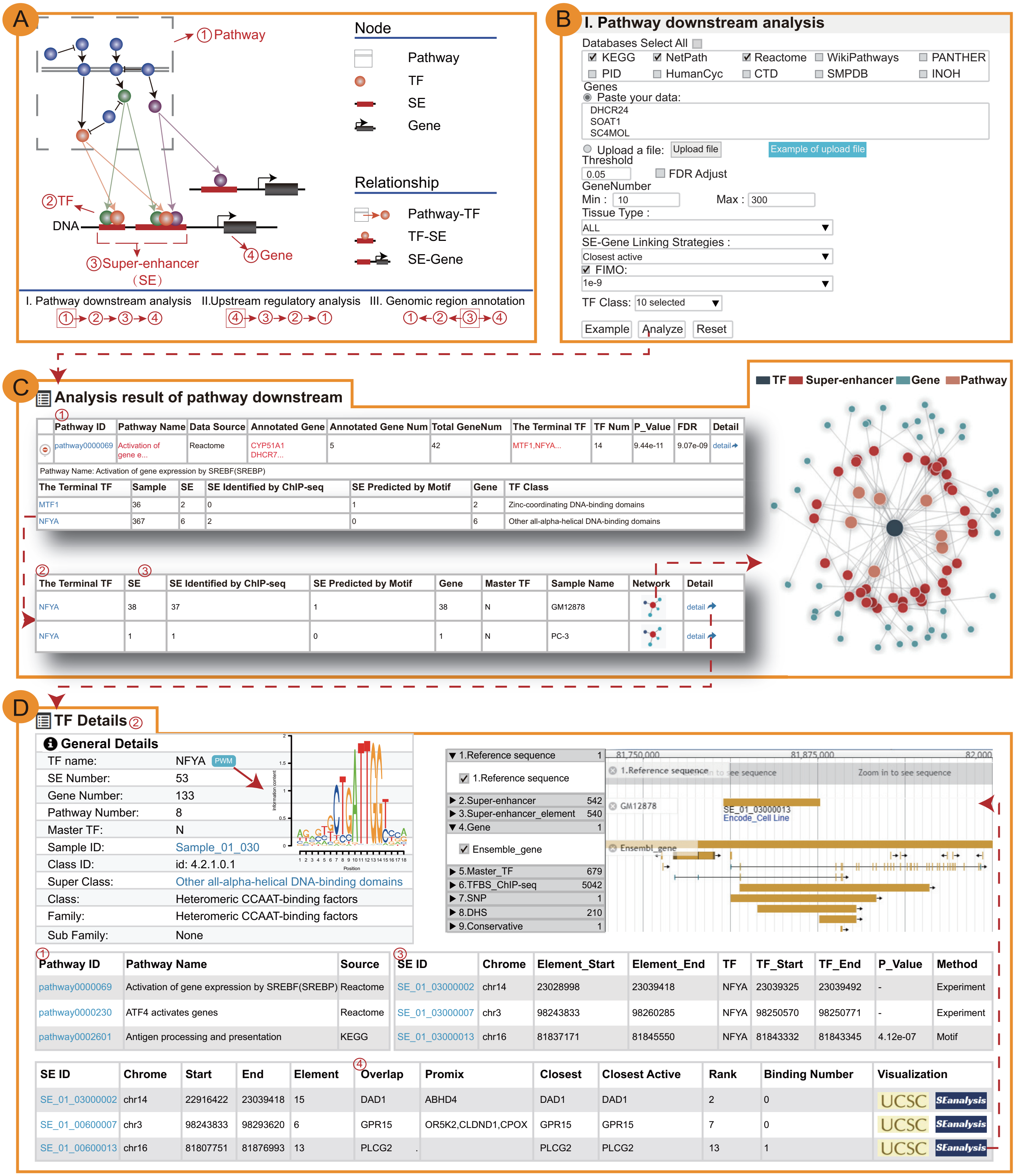
Super-enhancers (SEs) have prominent roles in biological and pathological processes through their unique transcriptional regulatory capability. To date, several SE databases have been developed by us and others. However, these existing databases do not provide downstream or upstream regulatory analyses of SEs. Pathways, transcription factors (TFs), SEs, and SE-associated genes form complex regulatory networks. Therefore, we designed a novel web server, SEanalysis, which provides comprehensive SE-associated regulatory network analyses. SEanalysis characterizes SE-associated genes, TFs binding to target SEs, and their upstream pathways. The current version of SEanalysis contains more than 330 000 SEs from more than 540 types of cells/tissues, 5042 TF ChIP-seq data generated from these cells/tissues, DNA-binding sequence motifs for ∼700 human TFs and 2880 pathways from 10 databases. SEanalysis supports searching by either SEs, samples, TFs, pathways or genes. The complex regulatory networks formed by these factors can be interactively visualized. In addition, we developed a customizable genome browser containing >6000 customizable tracks for visualization. The server is freely available at http://licpathway.net/SEanalysis.

INTRODUCTION

Super-enhancers (SEs), composed of clusters of enhancers, regulate cell-type-specific expression programs through a unique transcriptional activity to drive expression of genes that define cell identity (1–3). Because of their prominent functions in transcriptional regulation, SEs have been annotated in numerous cell/tissue types. As a hallmark of cancer, the alterations of signaling pathways converge on regulating terminal DNA-bound transcription factors (TFs) (4,5). Importantly, SEs are more frequently occupied by terminal TFs of pathways than typical enhancers. Concordantly, SE-associated genes are also more responsive to signalling cues than typical enhancers (5). Pathways, TFs, SEs and SE-associated genes form complex regulatory networks (5–7). These regulatory networks allow SEs to act as a crucial platform for pathways to regulate gene expression programs with much higher potency than typical enhancers. Notably, the functional interplay between oncogenic pathways and SEs is particularly prominent in regulating cancer biology, which have been highlighted by numerous reports (5,8–10).

Several SE databases have been developed, including dbSUPER (11), SEA (12) and SEdb (13). These databases summarize and catalog SE regions for various tissue and cell types using an H3K27ac signal-based ranking method (ROSE) (14). However, none of the databases provide downstream or upstream regulatory analysis involving SEs. To address this need, we developed the SEanalysis web server to provide SE-associated regulatory analyses. Users can perform several SE-associated analyses in our web server. I. Pathway downstream analysis: with the input of a set of genes of interest, SEanalysis will identify pathways that they are significantly enriched in, the terminal TFs that are downstream of the identified pathways, the SEs and SE-associated genes occupied by the terminal TFs (①→②→③→④ in Figure 1A). II. Upstream regulatory analysis: with the input of gene(s) of interest, SEanalysis will identify associated SEs and determine which TFs occupy these SE regions and the upstream pathways of

[*]To whom correspondence should be addressed. Tel: +86 459 8153035; Fax: +86 459 8153035; Email: lcqbio@163.com
Correspondence may also be addressed to De-Chen Lin. Tel: +86 20 89205573; Fax: +86 20 81332601; Email: lindch5@mail.sysu.edu.cn
Correspondence may also be addressed to Li-Yan Xu. Tel: +86 754 88900460; Fax: +86 754 88900847; Email: lyxu@stu.edu.cn
[†]The authors wish it to be known that, in their opinion, the first three authors should be regarded as Joint First Authors.


A
①Pathway
②TF
DNA
③Super-enhancer (SE)
④Gene
Node
Pathway
TF
SE
Gene
Relationship
Pathway-TF
TF-SE
SE-Gene
I. Pathway downstream analysis
①→②→③→④
II.Upstream regulatory analysis
④→③→②→①
III. Genomic region annotation
①←②←③→④
B
C
TF Super-enhancer Gene Pathway
D
①
②
③
④

I. Pathway downstream analysis

Databases Select All □

☑ KEGG ☑ NetPath ☑ Reactome □ WikiPathways □ PANTHER
□ PID □ HumanCyc □ CTD □ SMPDB □ INOH

Genes
● Paste your data:
DHCR24
SOAT1
SC4MOL
○ Upload a file: Upload file Example of upload file
Threshold
0.05 □ FDR Adjust
GeneNumber
Min : 10 Max : 300
Tissue Type :
ALL ▼
SE-Gene Linking Strategies :
Closest active ▼
☑ FIMO:
1e-9 ▼
TF Class: 10 selected ▼

Example Analyze Reset

Analysis result of pathway downstream

	Pathway ID	Pathway Name	Data Source	Annotated Gene	Annotated Gene Num	Total GeneNum	The Terminal TF	TF Num	P_Value	FDR	Detail
⊖	pathway0000069	Activation of gene e...	Reactome	CYP51A1 DHCR7...	5	42	MTF1,NFYA...	14	9.44e-11	9.07e-09	detail

Pathway Name: Activation of gene expression by SREBF(SREBP)

The Terminal TF	Sample	SE	SE Identified by ChIP-seq	SE Predicted by Motif	Gene	TF Class
MTF1	36	2	0	1	2	Zinc-coordinating DNA-binding domains
NFYA	367	6	2	0	6	Other all-alpha-helical DNA-binding domains

The Terminal TF	SE	SE Identified by ChIP-seq	SE Predicted by Motif	Gene	Master TF	Sample Name	Network	Detail
NFYA	38	37	1	38	N	GM12878		detail
NFYA	1	1	0	1	N	PC-3		detail

TF Details

General Details

TF name:	NFYA PWM
SE Number:	53
Gene Number:	133
Pathway Number:	8
Master TF:	N
Sample ID:	Sample_01_030
Class ID:	id: 4.2.1.0.1
Super Class:	Other all-alpha-helical DNA-binding domains
Class:	Heteromeric CCAAT-binding factors
Family:	Heteromeric CCAAT-binding factors
Sub Family:	None

▼ 1.Reference sequence	1
☑ 1.Reference sequence	
▶ 2.Super-enhancer	542
▶ 3.Super-enhancer_element	540
▼ 4.Gene	1
☑ Ensemble_gene	
▶ 5.Master_TF	679
▶ 6.TFBS_ChIP-seq	5042
▶ 7.SNP	1
▶ 8.DHS	210
▶ 9.Conservative	1

Pathway ID	Pathway Name	Source
pathway0000069	Activation of gene expression by SREBF(SREBP)	Reactome
pathway0000230	ATF4 activates genes	Reactome
pathway0002601	Antigen processing and presentation	KEGG

SE ID	Chrome	Element_Start	Element_End	TF	TF_Start	TF_End	P_Value	Method
SE_01_03000002	chr14	23028998	23039418	NFYA	23039325	23039492	-	Experiment
SE_01_03000007	chr3	98243833	98260285	NFYA	98250570	98250771	-	Experiment
SE_01_03000013	chr16	81837171	81845550	NFYA	81843332	81843345	4.12e-07	Motif

SE ID	Chrome	Start	End	Element	Overlap	Promix	Closest	Closest Active	Rank	Binding Number	Visualization
SE_01_03000002	chr14	22916422	23039418	15	DAD1	ABHD4	DAD1	DAD1	2	0	UCSC SEanalysis
SE_01_00600007	chr3	98243833	98293620	6	GPR15	OR5K2,CLDND1,CPOX	GPR15	GPR15	7	0	UCSC SEanalysis
SE_01_00600013	chr16	81807751	81876993	13	PLCG2 .		PLCG2	PLCG2	13	1	UCSC SEanalysis

Figure 1. Main functions of SEanalysis. (**A**) Schematic diagram of SEanalysis core functions. (**B**) Input and parameter page of 'Pathway downstream analysis'. (**C**) Results page of 'Pathway downstream analysis'. (**D**) Detailed interactive table of results of 'Pathway downstream analysis'.

the identified TFs (④→③→②→① in Figure 1A). III. Genomic region annotation: users can input genomic region(s) of interest in bed format to discover SEs overlapping the region(s), SE-associated genes and TFs occupying the SE regions and the upstream pathways of identified TFs (①←②←③→④ in Figure 1A).

DESCRIPTION OF WEB SERVER

Annotation of SEs and SE-associated genes

We obtained more than 330 000 SE regions involving 542 cells/tissues from the SEdb database (13) that was developed by our group using H3K27ac ChIP-seq data from NCBI GEO/SRA (15), ENCODE (16), Roadmap (16,17) and GGR (Genomics of Gene Regulation Project) (16). Raw sequencing reads were aligned to hg19 reference genomes with Bowtie (v0.12.9) (1,18), peaks were called using MACS14 (v1.4.2) (19), and SE regions were annotated using ROSE (14) software. Four different strategies were used to annotate SE-associated genes: closest active genes (20), overlapping genes, proximal genes and the closest genes (14).

Identification of TF occupancy in SE regions

To identify TFs binding to SEs, we collected a total of 5042 TF ChIP-seq datasets from ENCODE (16), Remap (21), Cistrome (22), ChIP-Atlas (http://chip-atlas.org) and GTRD (23) (Figure 2, top panel). For the uniformity of format and version, these peak datasets were converted to the hg19 genome using liftOver (http://genome.ucsc.edu/cgi-bin/hgLiftOver) tool of UCSC (24), and peaks that were failed to be converted were discarded. We used the 'cat' shell command to merge files of different samples for the same TF from the same tissue to generate union sets of peaks. TF binding peaks overlapping with constituent enhancers of SEs in matched cell/tissue types were identified using BEDTools (v2.25.0) (25). Motif occurrences in constituent enhancers of SEs for ∼700 TFs were identified using FIMO (Find Individual Motif Occurrences) (26) from the MEME (Multiple Em for Motif Elicitation) suite (27). More than 3000 DNA binding motifs for ∼700 TFs were compiled from the TRANSFAC (28) and MEME suite (20,27), based on the following collections: JASPAR CORE 2014 vertebrates (29), Jolma2013 (30), Homeodomains (31), UniPROBE (32), Wei2010 (33). Finally, TF motif occurrence within SE constituents was identified with a *P*-value threshold of 1e–5.

Identification of master TFs and classification of TFs

Saint-André *et al.* developed CRC_Mapper program to efficiently reconstruct cell-type-specific core regulatory circuitry (CRC) models based on the identification of SE-associated master TFs in a number of cell types (20). In this program, master TFs are defined as auto-regulated TFs encoded by SE-associated genes (1,2,20) that bind to at least three DNA sequence motifs at SEs associated with their own gene, and form fully interconnected auto-regulatory loops with other auto-regulated TFs by binding to SEs associated with other TFs within the loop (34–37). We identified master TFs for each cell/tissue using this program and provided interactive visualization of the CRC model. In addition, we manually assigned four generic level classifications (superclass, class, family and subfamily) of TFs according to TFClass database (38), based on their DNA-binding domains.

Construction of SE-associated regulatory networks

The data above were combined to construct an SE-associated regulatory network (Figure 2, top panel). Nodes of this network were composed of pathways, TFs, SEs and SE-associated genes. First, we established relationships between SEs and occupying TFs by either direct evidence generated from TF ChIP-seq data or by prediction based on motif analysis. Next, we obtained 2880 pathways with their pathway components from 10 pathway databases: KEGG, Reactome, NetPath, WikiPathways, PANTHER, PID, HumanCyc, CTD, SMPDB and INOH (39,40). We built relationships between a TF and a pathway if the TF was a component of the pathway. Finally, we constructed SE-associated regulatory networks by merging all relationships between all nodes, including (i) SEs-TFs, (ii) pathways-TFs and (iii) SEs-genes.

SEanalysis core functions

We designed three types of analyses to determine SE-associated regulatory networks (Figure 2, middle panel):

I. Pathway downstream analysis (①→②→③→④ in Figure 1A). With the input of a set of genes of interest and the selection of at least one pathway database (e.g. KEGG), SEanalysis will identify significantly enriched pathways, downstream TFs, SEs occupied by TFs and SE-associated genes (Figure 1B–D). SEanalysis will begin with the identification of the pathways in which these genes are significantly enriched using hypergeometric test (41). For each pathway assuming the entire genome has a total of *n* genes, of which *k* are components of the pathway under investigation, and the set of genes of interest has a total of *s* genes, of which *i* are involved in the same pathway, the enrichment significance *P*-value for that pathway is calculated as:

$$p = 1 - \sum_{x=0}^{i-1} \frac{\binom{k}{x}\binom{n-k}{s-x}}{\binom{n}{s}}$$

The false discovery rate (FDR) method is used to correct for multiple testing. Users can adjust the number of genes required to be enriched and set thresholds of *P*-values or FDRs to control the stringency of analysis. SEanalysis offers a 'FIMO' option to allow users to set different statistical thresholds to control for false positivity. The 'SE-Gene Linking Strategies' option allows users to select different annotation strategies to link SEs with target genes. In addition, 'Tissue Type' option allows user to perform targeted analysis in tissues of interest.

The output table contains basic information of identified pathways (Pathway ID, Pathway name, Pathway source,

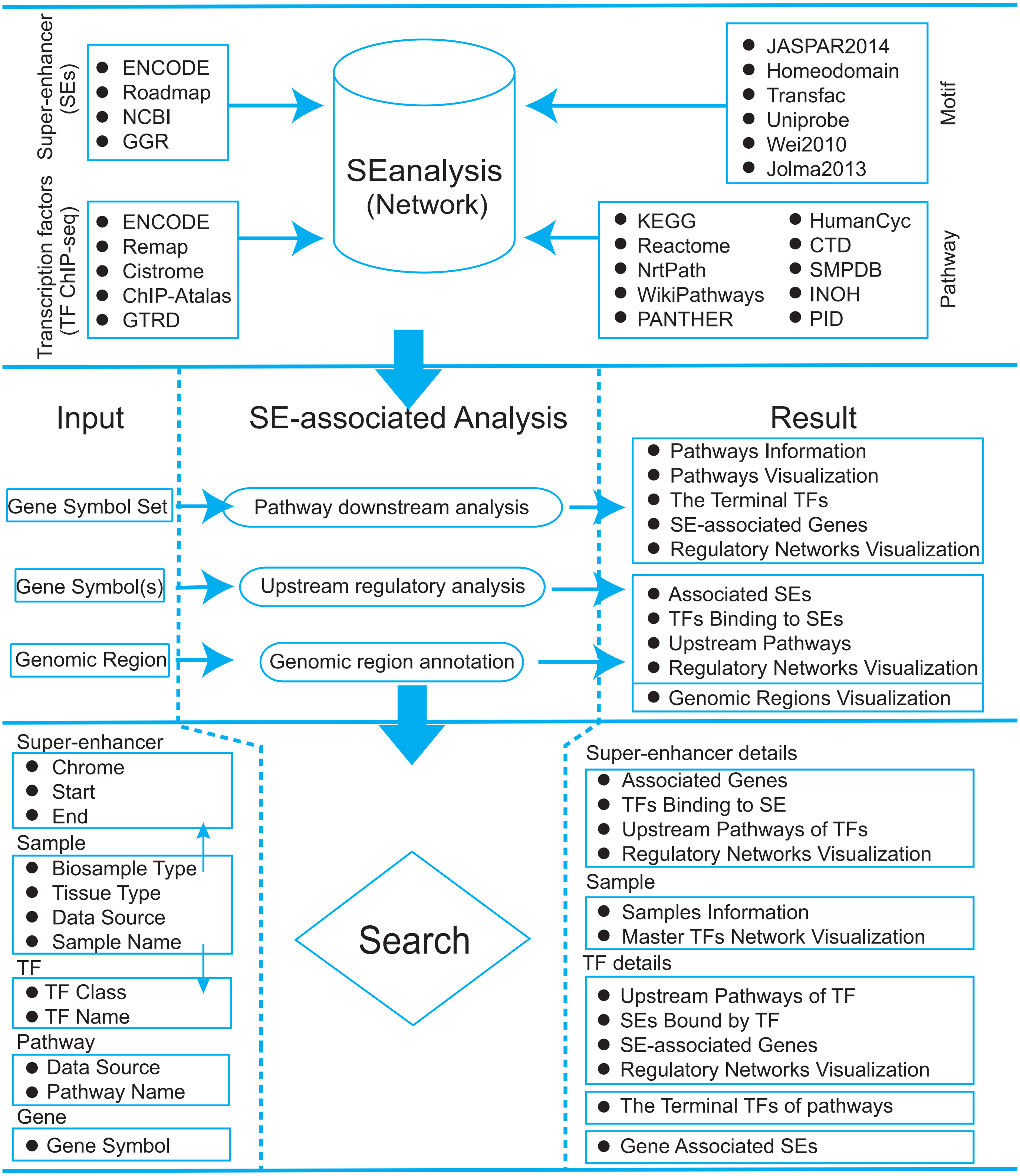


Figure 2. SEanalysis content and construction. SEanalysis contains a large number of SEs, TF ChIP-seq data, and DNA-binding sequence motifs as well as pathway information. Users can perform the following SE-associated analyses in our web server: I. Pathway downstream analysis, II. Upstream regulatory analysis, and III. Genomic region annotation. SEanalysis supports five searching modes, including ‘Searching by SE’, ‘Searching by Sample’, ‘Searching by TF’, ‘Searching by Pathway’ and ‘Searching by Gene’.

Annotated gene, Annotated gene number, Total gene number, The terminal TF and TF number), P-value and FDR of the enrichment score (Figure 1C). Pathways can be further displayed by clicking the 'Pathway ID' button. The TF related statistics will be further viewed by clicking the '+' button, including the number of SEs bound by TF (based on either ChIP-seq data or predicted by motif analysis), the number of genes associated with these SEs and visualization of regulatory networks based on the TF (Figure 1C and D). Furthermore, the 'detail' page provides the detailed description of the regulatory relationship between TFs involved in the current pathway, SEs bound by these TFs, and genes associated with SEs (Figure 3).

II. Upstream regulatory analysis (④→③→②→① in Figure 1A). With the input of gene(s) of interest and the setting of 'Tissue Type', 'SE-Gene Linking Strategies', 'FIMO' and 'Pathway Enrichment Threshold' options, SEanalysis will first identify the associated SEs, then determine the TFs occupying the SE regions and the enriched upstream pathways of the TFs. The output table will show: (i) the relationships between input genes and identified SEs, (ii) the number and names of TFs binding to the SEs (based on either ChIP-seq data or predicted by motif analysis), (iii) master TFs binding to these SEs (predicted by CRC_Mapper) (20) and (iv) upstream pathways and the sample information. The regulatory network base on SEs can be interactively visualized. The 'detail' page provides the full description of the regulatory relationship.

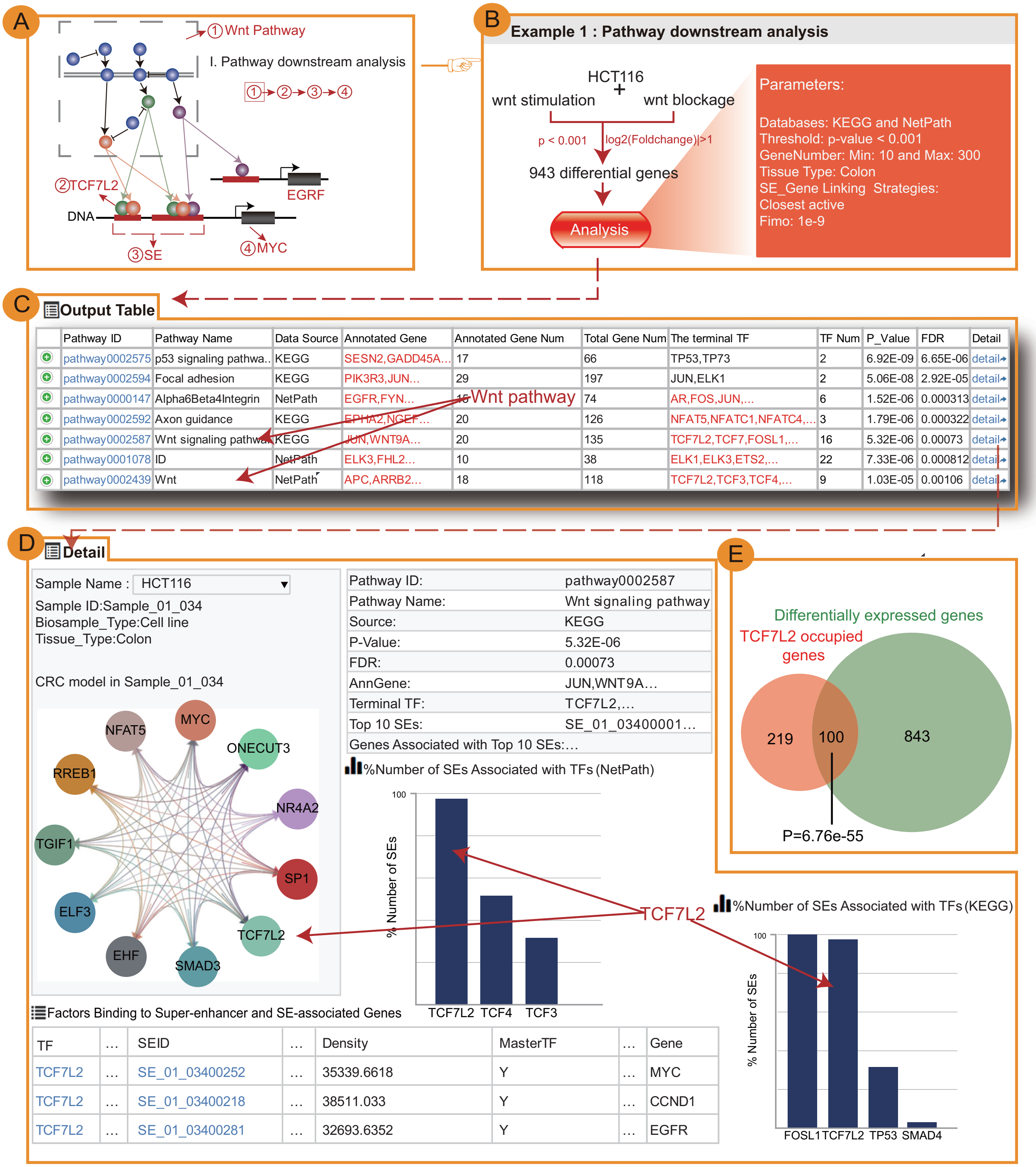
III. Genomic region annotation (①←②←③→④ in Figure 1A). Users can upload either a 'bed' format file or a list of genomic regions to identify SEs overlapping with the queried regions using Bedtools (25). Furthermore, users can set multiple options, including 'Tissue Type', 'SE-Gene Linking Strategies', 'FIMO' and 'Pathway Enrichment Threshold'. The output table includes: (i) the identified SEs overlapping with the queried regions and SE-associated genes, (ii) the number and names of TFs binding to the identified SEs (based on either ChIP-seq data or predicted by motif analysis), (iii) the number and names of master TFs binding to the identified SEs and (iv) the number and names of upstream pathways and sample information. The detailed description of the regulatory relationship is provided in the 'detail' page.

Case studies

We used the experimental data from two different studies to validate the key predictions of SEanalysis. For 'Pathway downstream analysis' (Figure 3A), we re-analyzed the work wherein a colon cancer cell line (HCT116, known to be dependent on Wnt activation for proliferation) was treated with Wnt inhibitor or stimulator followed by RNA-seq (5). We first obtained 943 differentially expressed genes upon treatment with Wnt modulators from Array Express experiment E-MTAB-651 (P-value < 0.001, $|\log_2(\text{Foldchange})| > 1$, Figure 3B) (5,42). These genes were used as input for our webserver for 'Pathway downstream analysis' (parameters: Databases: KEGG and NetPath, Threshold: P-value < 0.001, GeneNumber: Min: 10 and Max: 300, Tissue Type: Colon, SE-Gene Linking Strategies: Closest active and FIMO: 1e–9). The output table showed that Wnt pathway was not only significantly enriched (Hypergeometric test; P-value = 5.32e–06) but it was also the sole pathway identified by both pathway sources (KEGG and NetPath), and furthermore, it ranked highly as fifth and seventh among all pathways identified (Figure 3C). The webserver next identified a number of terminal TFs downstream of Wnt signaling pathway, including TCF7L2, TCF3, TCF4 and FOSL1. Importantly, using TCF7L2 ChIP-seq generated in HCT116 cells, our analysis showed that TCF7L2 occupied the vast majority of HCT116 SEs (98% of total SEs) (histogram in Figure 3D), which is consistent with the result of Hnisz *et al.* (5) and validated the prediction of webserver. Compared to other terminal TFs of Wnt pathway, TCF7L2 occupied a greater percentage of SEs in both KEGG and NetPath (histogram in Figure 3D). Lastly, we tested whether these SE-associated genes occupied by TCF7L2 were responsive to the manipulation of the Wnt pathway. Notably, these SE-associated genes occupied by TCF7L2 were significantly enriched in those exhibiting expression changes after disruption of Wnt pathway (Hypergeometric test; P-value = 6.76e–55) (Figure 3E), again confirming the previous report (5). Some of these TCF7L2-occupied, SE-associated genes included well-established Wnt targets, such as MYC, CCND1 and EGFR. Considering the well-established role of TCF7L2 in mediating Wnt signaling pathway through occupying super-enhancers, these results suggest the value and usefulness of our webserver in linking pathways, terminal TFs and super-enhancer activity.

To validate the prediction of 'Upstream regulatory analysis' (Supplementary Figure S1A), we studied luminal breast cancer which is known to be highly and uniquely dependent on estrogen signaling (5). Specifically, we used an ER-positive cell line, MCF-7, wherein a super-enhancer of ESR1 gene has been shown to be occupied by the TF estrogen receptor alpha (ERα). With input of the ESR1 gene in the 'Upstream regulatory analysis' (parameters: Tissue Type: Mammary Gland, SE-Gene Linking Strategies: Closest active, FIMO: 1e–9 and Pathway Enrichment Threshold: FDR corrected P-value < 0.001) (Supplementary Figure S1B), the output table predicted that the SEs associated with ESR1 gene were indeed occupied by estrogen receptor ERα in almost all ER-positive breast cancer cell lines (Supplementary Figure S1C). Moreover, ERα was further identified as a master TF in multiple ER-positive breast cancer cell lines, along with other well-established ERα interacting TFs, such as XBP1, FOXA1 and GATA3 (Supplementary Figure S1C and D). In the next step of prediction of enriched pathways, the webserver identified that the TFs associated with this ESR1 SE were significantly enriched in pathways including 'Nuclear receptor transcription pathway (ranked second of all pathways, Hypergeometric test; FDR corrected P-value = 5.7e–11)' and 'Validated nuclear estrogen receptor alpha network (ranked sixth of all pathways, Hypergeometric test; FDR corrected P-value = 5.62e–07)' (Supplementary Figure S1D, bottom panel). These predictions are congruent with the key role of ERα in mediating nuclear estrogen receptor signaling to the regulation of super-enhancer activity in luminal breast cancer.

Taken together, these data validated all of the key webserver predictions including: (i) pathway enrichment; (ii)



Parameters:

Databases: KEGG and NetPath
Threshold: p-value < 0.001
GeneNumber: Min: 10 and Max: 300
Tissue Type: Colon
SE_Gene Linking Strategies:
Closest active
Fimo: 1e-9

Output Table

	Pathway ID	Pathway Name	Data Source	Annotated Gene	Annotated Gene Num	Total Gene Num	The terminal TF	TF Num	P_Value	FDR	Detail
⊕	pathway0002575	p53 signaling pathwa..	KEGG	SESN2,GADD45A...	17	66	TP53,TP73	2	6.92E-09	6.65E-06	detail→
⊕	pathway0002594	Focal adhesion	KEGG	PIK3R3,JUN...	29	197	JUN,ELK1	2	5.06E-08	2.92E-05	detail→
⊕	pathway0000147	Alpha6Beta4Integrin	NetPath	EGFR,FYN...	15	74	AR,FOS,JUN,...	6	1.52E-06	0.000313	detail→
⊕	pathway0002592	Axon guidance	KEGG	EPHA2,NGEF...	20	126	NFAT5,NFATC1,NFATC4,...	3	1.79E-06	0.000322	detail→
⊕	pathway0002587	Wnt signaling pathwa	KEGG	JUN,WNT9A...	20	135	TCF7L2,TCF7,FOSL1,...	16	5.32E-06	0.00073	detail→
⊕	pathway0001078	ID	NetPath	ELK3,FHL2...	10	38	ELK1,ELK3,ETS2,...	22	7.33E-06	0.000812	detail→
⊕	pathway0002439	Wnt	NetPath	APC,ARRB2...	18	118	TCF7L2,TCF3,TCF4,...	9	1.03E-05	0.00106	detail→

Detail

Sample Name : HCT116 ▾
Sample ID:Sample_01_034
Biosample_Type:Cell line
Tissue_Type:Colon

Pathway ID:	pathway0002587
Pathway Name:	Wnt signaling pathway
Source:	KEGG
P-Value:	5.32E-06
FDR:	0.00073
AnnGene:	JUN,WNT9A...
Terminal TF:	TCF7L2,...
Top 10 SEs:	SE_01_03400001...
Genes Associated with Top 10 SEs:...	

Factors Binding to Super-enhancer and SE-associated Genes

TF	...	SEID	...	Density	MasterTF	...	Gene
TCF7L2	...	SE_01_03400252	...	35339.6618	Y	...	MYC
TCF7L2	...	SE_01_03400218	...	38511.033	Y	...	CCND1
TCF7L2	...	SE_01_03400281	...	32693.6352	Y	...	EGFR

Figure 3. Validation results for 'Pathway downstream analysis'. (**A**) Schematic diagram of 'Pathway downstream analysis'. (**B**) Input exemplary data and parameters of 'Pathway downstream analysis' using HCT116/Wnt pathway example. (**C**) The output table of 'Pathway downstream analysis' generated by our webserver. (**D**) The detailed page of the output table. The page provided the detailed description of the regulatory relationship between TFs involved in the predicted pathway, SEs bound by these TFs, genes associated with SEs, as well as the CRC model. (**E**) SE-associated genes occupied by TCF7L2 were significantly enriched in those exhibiting expression changes upon disruption of Wnt pathway.

terminal TF ranking; (iii) identification of downstream SEs and (iv) annotation of SE-associated genes.

User-friendly searching and browsing functions

SEanalysis supports five different searching modes: 'Searching by SE', 'Searching by Sample', 'Searching by TF', 'Searching by Pathway' and 'Searching by Gene'. SEanalysis provides data browsing, which is an interactive table with a sorting function that allows users to quickly search for samples and customize filters including 'Data Sources', 'Biosample Type', 'Tissue Type' and 'Biosample Name'. To further view the SE of a given sample, users can click 'Sample ID' button.

Visualization of regulatory network and customizable genome browser

As mentioned above, SEs, SE-associated genes, TFs binding to SEs and upstream pathways of TFs form complex networks. To facilitate the understanding of the network, SEanalysis supports interactive visualization of networks using the visualization plugin Echarts (http://echarts.baidu.com).

To view SEs along the genome, we developed a customizable genome browser using JBrowse (http://jbrowse.org) (43) containing more than 6,000 tracks. This browser allows viewing the genomic coordinates of SEs, TF binding sites (TFBS) identified by ChIP-seq, SNPs, DHSs and conservation score. SEanalysis can also link the data to the UCSC genome browser (24).

Implementation

SEanalysis is freely available to the research community at http://www.licpathway.net/SEanalysis and requires no registration or login. The main framework of SEanalysis was developed based on Java 1.8.0 (https://www.oracle.com/technetwork/java/) and MySQL 5.7.16 (https://www.mysql.com/). JQuery 3.3.1 (http://jquery.com) and Bootstrap 3.3.7 (https://getbootstrap.com/) (an open source front-end framework) were used to design the front-end web interface. Google Chrome, Mozilla Firefox, Opera and Safari are the preferred browsers for display.

SUMMARY

To provide comprehensive analysis of SE-associated regulatory networks, we designed and developed a novel web server, SEanalysis, with the following functions: (i) Pathway downstream analysis, (ii) Upstream regulatory analysis, (iii) Genomic region annotation. Compared with other SE databases, this webserver focuses on constructing and analyzing the networks between pathways, TFs, SEs, and SE-associated genes. SEanalysis also allows users to readily download SEs for different cells/tissues, in both bed and csv format. The output results of analyses can also be downloaded. In addition, SEanalysis supports external analytical tools of genomic regions such as GREAT (44) and UCSC (24). SEanalysis also links to additional external resources including NCBI Gene (45), GeneCards (46) and UniProt (47).

The rapid development of high-throughput sequencing technology leads to the accelerated accumulation of a large number of epigenomic datasets. SEanalysis will be updated and maintained accordingly. Our effort to establish this web server was prompted by the great need of researchers to understand the biology of epigenomic network regulation. These researchers include cell and molecular biologists, geneticists and data scientists. Moreover, the field of epigenomics is rapidly progressing, and the integrative analysis of epigenomic regulatory networks is one of the most investigated areas. Therefore, SEanalysis will be a valuable resource for experimental and computational biologists in the field of epigenomics.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

National Natural Science Foundation of China [81572341, 81872306, 61601150, 81772532] (in part); Natural Science Foundation of Heilongjiang Province [JJ2016ZR1232/F2016031]; YuWeihan Outstanding Youth Training Fund of Harbin Medical University. Funding for open access charge: National Natural Science Foundation of China [81572341, 81872306, 61601150, 81772532].

Conflict of interest statement. None declared.

REFERENCES

1. Hnisz,D., Abraham,B.J., Lee,T.I., Lau,A., Saint-Andre,V., Sigova,A.A., Hoke,H.A. and Young,R.A. (2013) Super-enhancers in the control of cell identity and disease. *Cell*, **155**, 934–947.
2. Whyte,W.A., Orlando,D.A., Hnisz,D., Abraham,B.J., Lin,C.Y., Kagey,M.H., Rahl,P.B., Lee,T.I. and Young,R.A. (2013) Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell*, **153**, 307–319.
3. Parker,S.C., Stitzel,M.L., Taylor,D.L., Orozco,J.M., Erdos,M.R., Akiyama,J.A., van Bueren,K.L., Chines,P.S., Narisu,N., Black,B.L. *et al.* (2013) Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 17921–17926.
4. Hanahan,D. and Weinberg,R.A. (2011) Hallmarks of cancer: the next generation. *Cell*, **144**, 646–674.
5. Hnisz,D., Schuijers,J., Lin,C.Y., Weintraub,A.S., Abraham,B.J., Lee,T.I., Bradner,J.E. and Young,R.A. (2015) Convergence of developmental and oncogenic signaling pathways at transcriptional super-enhancers. *Mol. Cell*, **58**, 362–370.
6. Betancur,P.A., Abraham,B.J., Yiu,Y.Y., Willingham,S.B., Khameneh,F., Zarnegar,M., Kuo,A.H., McKenna,K., Kojima,Y., Leeper,N.J. *et al.* (2017) A CD47-associated super-enhancer links pro-inflammatory signalling to CD47 upregulation in breast cancer. *Nat. Commun.*, **8**, 14802.
7. Gunnell,A., Webb,H.M., Wood,C.D., McClellan,M.J., Wichaidit,B., Kempkes,B., Jenner,R.G., Osborne,C., Farrell,P.J. and West,M.J. (2016) RUNX super-enhancer control through the notch pathway by Epstein-Barr virus transcription factors regulates B cell growth. *Nucleic Acids Res.*, **44**, 4636–4650.
8. Katerndahl,C.D.S., Heltemes-Harris,L.M., Willette,M.J.L., Henzler,C.M., Frietze,S., Yang,R., Schjerven,H., Silverstein,K.A.T., Ramsey,L.B., Hubbard,G. *et al.* (2017) Antagonism of B cell enhancer networks by STAT5 drives leukemia and poor patient survival. *Nat. Immunol.*, **18**, 694–704.
9. Kandaswamy,R., Sava,G.P., Speedy,H.E., Bea,S., Martin-Subero,J.I., Studd,J.B., Migliorini,G., Law,P.J., Puente,X.S., Martin-Garcia,D. *et al.* (2016) Genetic predisposition to chronic lymphocytic leukemia

Is mediated by a BMF super-enhancer polymorphism. *Cell Rep.*, **16**, 2061–2067.

10. Bojcsuk,D., Nagy,G. and Balint,B.L. (2017) Inducible super-enhancers are organized based on canonical signal-specific transcription factor binding elements. *Nucleic Acids Res.*, **45**, 3693–3706.
11. Khan,A. and Zhang,X. (2016) dbSUPER: a database of super-enhancers in mouse and human genome. *Nucleic Acids Res.*, **44**, D164–D171.
12. Wei,Y., Zhang,S., Shang,S., Zhang,B., Li,S., Wang,X., Wang,F., Su,J., Wu,Q., Liu,H. *et al.* (2016) SEA: a super-enhancer archive. *Nucleic Acids Res.*, **44**, D172–D179.
13. Jiang,Y., Qian,F., Bai,X., Liu,Y., Wang,Q., Ai,B., Han,X., Shi,S., Zhang,J., Li,X. *et al.* (2019) SEdb: a comprehensive human super-enhancer database. *Nucleic Acids Res.*, **47**, D235–D243.
14. Loven,J., Hoke,H.A., Lin,C.Y., Lau,A., Orlando,D.A., Vakoc,C.R., Bradner,J.E., Lee,T.I. and Young,R.A. (2013) Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell*, **153**, 320–334.
15. Barrett,T., Troup,D.B., Wilhite,S.E., Ledoux,P., Evangelista,C., Kim,I.F., Tomashevsky,M., Marshall,K.A., Phillippy,K.H., Sherman,P.M. *et al.* (2011) NCBI GEO: archive for functional genomics data sets–10 years on. *Nucleic Acids Res.*, **39**, D1005–D1010.
16. Consortium., E.P. (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature*, **489**, 57–74.
17. Bernstein,B.E., Stamatoyannopoulos,J.A., Costello,J.F., Ren,B., Milosavljevic,A., Meissner,A., Kellis,M., Marra,M.A., Beaudet,A.L., Ecker,J.R. *et al.* (2010) The NIH roadmap epigenomics mapping consortium. *Nat. Biotechnol.*, **28**, 1045–1048.
18. Langmead,B., Trapnell,C., Pop,M. and Salzberg,S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, **10**, R25.
19. Zhang,Y., Liu,T., Meyer,C.A., Eeckhoute,J., Johnson,D.S., Bernstein,B.E., Nusbaum,C., Myers,R.M., Brown,M., Li,W. *et al.* (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol.*, **9**, R137.
20. Saint-Andre,V., Federation,A.J., Lin,C.Y., Abraham,B.J., Reddy,J., Lee,T.I., Bradner,J.E. and Young,R.A. (2016) Models of human core transcriptional regulatory circuitries. *Genome Res.*, **26**, 385–396.
21. Cheneby,J., Gheorghe,M., Artufel,M., Mathelier,A. and Ballester,B. (2018) ReMap 2018: an updated atlas of regulatory regions from an integrative analysis of DNA-binding ChIP-seq experiments. *Nucleic Acids Res.*, **46**, D267–D275.
22. Mei,S., Qin,Q., Wu,Q., Sun,H., Zheng,R., Zang,C., Zhu,M., Wu,J., Shi,X., Taing,L. *et al.* (2017) Cistrome Data Browser: a data portal for ChIP-Seq and chromatin accessibility data in human and mouse. *Nucleic Acids Res.*, **45**, D658–D662.
23. Yevshin,I., Sharipov,R., Valeev,T., Kel,A. and Kolpakov,F. (2017) GTRD: a database of transcription factor binding sites identified by ChIP-seq experiments. *Nucleic Acids Res.*, **45**, D61–D67.
24. Karolchik,D., Barber,G.P., Casper,J., Clawson,H., Cline,M.S., Diekhans,M., Dreszer,T.R., Fujita,P.A., Guruvadoo,L., Haeussler,M. *et al.* (2014) The UCSC Genome Browser database: 2014 update. *Nucleic Acids Res.*, **42**, D764–D770.
25. Quinlan,A.R. and Hall,I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.
26. Grant,C.E., Bailey,T.L. and Noble,W.S. (2011) FIMO: scanning for occurrences of a given motif. *Bioinformatics*, **27**, 1017–1018.
27. Bailey,T.L., Boden,M., Buske,F.A., Frith,M., Grant,C.E., Clementi,L., Ren,J., Li,W.W. and Noble,W.S. (2009) MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.*, **37**, W202–W208.
28. Matys,V., Kel-Margoulis,O.V., Fricke,E., Liebich,I., Land,S., Barre-Dirrie,A., Reuter,I., Chekmenev,D., Krull,M., Hornischer,K. *et al.* (2006) TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res.*, **34**, D108–D110.
29. Mathelier,A., Zhao,X., Zhang,A.W., Parcy,F., Worsley-Hunt,R., Arenillas,D.J., Buchman,S., Chen,C.Y., Chou,A., Ienasescu,H. *et al.* (2014) JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. *Nucleic Acids Res.*, **42**, D142–D147.
30. Jolma,A., Yan,J., Whitington,T., Toivonen,J., Nitta,K.R., Rastas,P., Morgunova,E., Enge,M., Taipale,M., Wei,G. *et al.* (2013) DNA-binding specificities of human transcription factors. *Cell*, **152**, 327–339.
31. Berger,M.F., Badis,G., Gehrke,A.R., Talukder,S., Philippakis,A.A., Pena-Castillo,L., Alleyne,T.M., Mnaimneh,S., Botvinnik,O.B., Chan,E.T. *et al.* (2008) Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. *Cell*, **133**, 1266–1276.
32. Robasky,K. and Bulyk,M.L. (2011) UniPROBE, update 2011: expanded content and search tools in the online database of protein-binding microarray data on protein-DNA interactions. *Nucleic Acids Res.*, **39**, D124–D128.
33. Wei,G.H., Badis,G., Berger,M.F., Kivioja,T., Palin,K., Enge,M., Bonke,M., Jolma,A., Varjosalo,M., Gehrke,A.R. *et al.* (2010) Genome-wide analysis of ETS-family DNA-binding in vitro and in vivo. *EMBO J.*, **29**, 2147–2160.
34. Odom,D.T., Zizlsperger,N., Gordon,D.B., Bell,G.W., Rinaldi,N.J., Murray,H.L., Volkert,T.L., Schreiber,J., Rolfe,P.A., Gifford,D.K. *et al.* (2004) Control of pancreas and liver gene expression by HNF transcription factors. *Science*, **303**, 1378–1381.
35. Odom,D.T., Dowell,R.D., Jacobsen,E.S., Nekludova,L., Rolfe,P.A., Danford,T.W., Gifford,D.K., Fraenkel,E., Bell,G.I. and Young,R.A. (2006) Core transcriptional regulatory circuitry in human hepatocytes. *Mol. Syst. Biol.*, **2**, 2006 0017.
36. Boyer,L.A., Lee,T.I., Cole,M.F., Johnstone,S.E., Levine,S.S., Zucker,J.P., Guenther,M.G., Kumar,R.M., Murray,H.L., Jenner,R.G. *et al.* (2005) Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell*, **122**, 947–956.
37. Sanda,T., Lawton,L.N., Barrasa,M.I., Fan,Z.P., Kohlhammer,H., Gutierrez,A., Ma,W., Tatarek,J., Ahn,Y., Kelliher,M.A. *et al.* (2012) Core transcriptional regulatory circuit controlled by the TAL1 complex in human T cell acute lymphoblastic leukemia. *Cancer Cell*, **22**, 209–221.
38. Wingender,E., Schoeps,T., Haubrock,M. and Donitz,J. (2015) TFClass: a classification of human transcription factors and their rodent orthologs. *Nucleic Acids Res.*, **43**, D97–D102.
39. Kanehisa,M., Sato,Y., Kawashima,M., Furumichi,M. and Tanabe,M. (2016) KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.*, **44**, D457–D462.
40. Cerami,E.G., Gross,B.E., Demir,E., Rodchenkov,I., Babur,O., Anwar,N., Schultz,N., Bader,G.D. and Sander,C. (2011) Pathway commons, a web resource for biological pathway data. *Nucleic Acids Res.*, **39**, D685–D690.
41. Li,C., Li,X., Miao,Y., Wang,Q., Jiang,W., Xu,C., Li,J., Han,J., Zhang,F., Gong,B. *et al.* (2009) SubpathwayMiner: a software package for flexible identification of pathways. *Nucleic Acids Res.*, **37**, e131.
42. Moffa,G., Erdmann,G., Voloshanenko,O., Hundsrucker,C., Sadeh,M.J., Boutros,M. and Spang,R. (2016) Refining Pathways: A Model Comparison Approach. *PloS one*, **11**, e0155999.
43. Buels,R., Yao,E., Diesh,C.M., Hayes,R.D., Munoz-Torres,M., Helt,G., Goodstein,D.M., Elsik,C.G., Lewis,S.E., Stein,L. *et al.* (2016) JBrowse: a dynamic web platform for genome visualization and analysis. *Genome Biol.*, **17**, 66.
44. McLean,C.Y., Bristor,D., Hiller,M., Clarke,S.L., Schaar,B.T., Lowe,C.B., Wenger,A.M. and Bejerano,G. (2010) GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.*, **28**, 495–501.
45. Coordinators,N.R. (2016) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, **44**, D7–19.
46. Safran,M., Dalah,I., Alexander,J., Rosen,N., Iny Stein,T., Shmoish,M., Nativ,N., Bahir,I., Doniger,T., Krug,H. *et al.* (2010) GeneCards Version 3: the human gene integrator. *Database*, **2010**, baq020.
47. Consortium.,U. (2015) UniProt: a hub for protein information. *Nucleic Acids Res.*, **43**, D204–D212.