CATA: a comprehensive chromatin accessibility database for cancer

Jianyuan Zhou^, Xuecang Li^, Jianjun Xie*, Chunquan Li*

Corresponding author: Chunquan Li, School of Medical Informatics, Daqing Campus, Harbin Medical University, Daqing 163319, China, Tel.: +86-15004591078; E-mail: lcqbio@163.com; Jianjun Xie, Department of Biochemistry and Molecular Biology, Medical College of Shantou University, Shantou 515041, China Tel.: +86-754-86500847; E-mail: g_jjxie@stu.edu.cn.

*To whom correspondence should be addressed.

^These authors contribute equally to this work.

Abstract

Chromatin accessibility is a crucial epigenetic concept that plays a biological role in oncology. As humans become more involved in cancer research, a comprehensive database is required to identify and annotate tumor chromatin accessible regions (CARs). Here, CATA was developed to provide cancer related CAR annotation. Currently, CATA possesses 2,991,163 CARs, relevant clinical data and transcription factor binding predictions for cancer CARs from 410 tumor samples of 24 cancer types. Furthermore, CARs were annotated by SNPs, risk SNPs, eQTLs, linkage disequilibrium SNPs, transcription factors, CNV, SNV, enhancer and 450K methylation sites in our database. By combining all these resources, we believe that CATA will provide better service for researchers on oncology. Our database is accessible at http://bio.licpathway.net/cata/

Keywords: Cancer; chromatin accessibility; enhancer; transcription factors;

Introduction

Chromatin accessibility refers to the extent to which nuclear macromolecules can physically contact chromatinized DNA, as determined by the occupancy and topology of nucleosomes and other chromatin binding factors that impede DNA contact [1] and chromatin accessible regions (CARs) signal is like a bridge to link regulator elements with transcription factors. It can help us discover more key information about epigenetics and cancer mechanism. Researchers always used DNase-seq, to measure chromatin accessibility until that Wliliam J developed the Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq) technique in 2013[2]. Compared to DNase-seq, ATAC-seq has 4 major strengthens. First of all, the accuracy of ATAC-seq was consistent with that of DNase-seq but ATAC-seq experiment is easier to perform. Secondly, ATAC-seq kept simultaneous disclosure of accessible genomic location, DNA binding protein, transcriptional binding (TF) site interaction. Thirdly, ATAC-seq required fewer cell numbers. Fourthly, ATAC-seq showed great repeatability ($R = 0.98$) and also had good consistence with the DHS sequencing ($R>0.79$) [2].

ATAC-seq made it possible to assess the gene regulatory landscape in primary human cancers. Interestingly, chromatin accessibility changes are usually early cytological events in the context of various strain responses, stress responses, or developmental transitions [3]. In the early diagnosis and treatment of cancer, chromatin structure studies can provide valuable information [4]. However, the combinatorial effects of these phenomena in a specific biological function are poorly understood. Moreover, TFs are major players in the regulation of gene expression, and transcriptional regulation involves complex and detailed patterns of activity that bind to TFs. Numerous studies had indicated that TFs regulated genes by binding regulatory elements in the accessible genome [5]. There are a large number of regulatory elements, such as enhancers and promoters, located in ATAC-peak. Chromatin accessibility is diverse in different cancers, so there is a lot of transcriptional regulation information contained in CARs [6]. Identification of chromatin accessibility of cancer would provide ample insight into cancer mechanism.

The TCGA uses a large number of human cancer samples for ATAC-seq [1], including 410 primary samples of 23 cancer types which provided abundant information of cancer chromatin accessibility. However, it is difficult for researchers to obtain regulating information of chromatin accessibility from those ATAC-seq data

because annotation is lacked. Importantly, the cancer CAR usually contains a lot of regulatory elements [4,7] and could be bound by a lot of transcription factors, which might co-regulate target genes with the regulatory elements [8,9]. In this study, We integrated data from 12 databases including TCGA, FUNTOM [10], 1000GENOMES [11], Jaspar [12] and Xena to annotated enhancers and TFs as well as mutation and methylation sites in CAR and developed CATA database. CATA contains 2,991,163 CAR and corresponding annotations for 410 tumor samples of 24 cancer types and supplies clinical data and survival analysis. CATA is also a comprehensive cancer ATAC-seq database that provides multiple functions, including storage, browsing, annotation, and analysis. It could be a powerful work platform for mining potential functions of CAR and explore relevant regular patterns about cancer.

Materials and methods

The collection of chromatin accessibility: We download chromatin accessible region data (bed file) from TCGA. First, The ATAC-seq data processing and alignment was performed using the PEPATAC pipeline (http://code.databio.org/PEPATAC/). The hg38 genome build used for alignment was obtained using Refgenie (https://github.com/databio/refgenie). Precisely, Bowtie2 was used to align to the hg38 human reference genome using "--very-sensitive - X 2000 --rg-id" options. Picard (http://broadinstitute.github.io/picard/) was then used to remove duplicates. Peak calling for 796 ATAC-seq profiles and 23 cancer types was performed to ensure high quality fixed-width peaks with 501bp. For each sample, peak calling was performed on the Tn5-corrected single-base insertions using the MACS2 callpeak command with parameters "--shift -75 --extsize 150 --nomodel --call-summits --nolambda --keep-dup all -p 0.01" .The peak summits were then extended by 250 bp on either side to a final width of 501 bp, filtered by the ENCODE hg38 blacklist (https://www.encodeproject.org/annotations/ENCSR636HFF/), and filtered to remove peaks that extend beyond the ends of chromosomes. For overlapping peaks within a single sample, the most significant peak is kept and any peak that directly overlaps with that significant peak is removed. Then there is a set of fixed-width peaks for each sample. For each cancer, TCGA compiled a "cancer type-specific peak set" containing all of the reproducible peaks observed in an individual cancer type. For the overlapping peaks from different samples, TCGA kept the most significant peak. At last, using the same method, the 'pan-cancer peak set' was obtained from all cancer types that could then be used for cross-cancer comparisons.

Annotation of chromatin accessibility regions: CARs were annotated both genetically and epigenetically through BEDTools [13], including common SNPs, risk SNPs, copy number variation, somatic mutation, motif changes, eQTLs, Transcription factors binding regions, methylation, enhancers and CRISPR/Cas9 target sites, which prompted discovering potential functions of chromatin accessibility. In addition, interactive tables were utilized to present further detail.

Enhancer: Totally 65,423 enhancers were collected from FANTOM5[10] and then converted to hg38 genome via LiftOver for the annotation.

Transcription factor related data: Transcription factors binding regions were gained from FIMO [14] prediction. Besides, 5,797,266 TFBS [15] was downloaded from the UCSC[15] and convert to hg38 genome through LiftOver.

CRISPR/Cas9 target sites: CRISPR/Cas9 target site [16], refers to RNA sequence within 200 bp of the genome region or DNA sequence in the transcription region, induction of which causes a precise cleavage of endogenous genomic loci. UCSC was employed to download CRISPR/Cas9 target sites, which were converted via LiftOver. Moreover, CRISPOR tool was used in designation, evaluation and clone for guidance sequence of CRISPR/Cas9 system.

Gene annotation: Three strategies were adopted to locate CAR associated genes. ROSE2 [17] gene-mapper method was applied in prediction of associated genes including overlap, proximal and closest.

Common SNPs/Linkage disequilibrium SNPs/Risk SNPs: Total 38,063,729 common SNPs were downloaded from dbSNP [18] version 1.50. Gene linkage disequilibrium (LD) 1000 genome [11] project was calculated by 3 phased genotype information. Besides, minimum filtered SNP allele frequency of less than 0.05 (MAF) was accepted through VCFTools (v0.1.13). Moreover, plink (v1.9) was utilized to calculate SNPs with MAF > 0.05 in LD (r2 = 0.8) for 5 supergroups (Africa, ad mix America, East Asia, Europe and South Asia). Risk SNP, genome-wide association studies (GWAS) resulting from GWAS catalog [19] and GWASdb [20] v2.0 set of planning were integrated into the table of human diseases/traits of SNP and insertion/deletion variant, which allowed functional annotation.

Motif changes: Annotation of motif mutations was based on TRANSFAC [21] weight JASPAR [12] collection position weight matrix. The binding affinity of motif mutations, CAR located on the MAF> 1000 Genome Project 0.05 SNP stage, and 30-bp regions upstream and downstream of the SNP were calculated through R-package

at-SNP. Finally, 254,545,586 motif changes were collected.

TCGA series data: TCGA-related data including methylation data, RNA expression profile data, and Somatic-mutation-variation, Copy-number-variation, Clinical Information, ATAC-seq raw counts numbers, CAR about 24 cancer types was obtained from UCSC XENA (http://xena.ucsc.edu/). Additionally, similar data from ATAC data involved in methylation and RNA expression profile was and other ATAC-raw count numbers were averaged based on cancer species and samples, respectively.

EQTL: Human eQTL was downloaded and merged from GTEx [22] v5.0, HaploReg [23] for comparision in various tissues and PancanQTL [24] with TCGA eQTL gene relationship (https://tcga-data.nci.nih.gov/tcga). SNP eQTL was plotted and CAR were annotated which indicated a SNP gene regulated by the target gene as a potential associated CAR.

Results

A search interface for retrieving CAR

CATA offers 4 search methods to help users find CAR of cancer of interest. 1. Search for cancer CAR based on tumor samples (enter the cancer of interest) 2. Search for cancer CAR based on gene name (i. Enter the gene name of interest, gene symbol format, ii. Select the cancer type of interest). 3. Search for cancer CAR by transcription factors (i. Enter the transcription factor of interest. ii. Select the cancer type of interest.), 4. Search for relevant tumor CARs based on genomic coordinates (i. Enter the genomic coordinate information of interest, the CATA database supports the latest hg38 reference genome. ii. Select the cancer species). In the search results, CATA provides summative information about the CARs, including SNP, SNV, CNV, methylation, Motif and so on. Click interest PEAK-ID can entering detail page. The Detail page consists of seven parts, including the CAR overview and annotation, RNA-expression, associative clinical data, survival analysis, methylation visualization, Upstream Transcription factor enrichment. 1. Accessible overview: CATA provides a preview of CAR, including raw counts number, presented by bar plots and some summary information about the region. 2. Accessible region annotation: Provides chromatin in tabular form Information on the CAR, including SNP, motif, CNV, SNV, TFBS, etc., users can click the relevant knob to browse the corresponding information. 3. RNA-expression: We visualized the gene expression of the CAR by barplot. The user can select the gene annotated by ROSE2 to be divided into close, proximal, and overlap. 4.

associative clinical data: We provide clinical information about patients with ATAC-seq, which is convenient for researchers to find potential cure targets. 5.survival analysis: CATA provides survival analysis of cancer CAR related genes to help researchers find genes with clinical value. 6.methation visualization: CATA uses barplot to visualize 450K methylation chip data, providing methylation in CAR, presented in 24 cancer forms. 7. Upstream Transcription factor enrichment: We used FIMO software to predict the binding factors of CAR and found some transcription factors and transcriptional cofactors that interact with transcription factors to form a visual network. Figure (contains: closet, proximal, overlap gene, combined motif, interact TF, and peak ID are marked by different colors).

User's friendly Explorations

CATA provides a user-friendly interface to help users to navigate quickly and easily. On the left side of the page, user can distinguish between samples, via four options (Tissue type, Cancer type, Annotation, Chromosome), which can be clicked. The results are filtered and the user can click on the 'Peak ID' to jump to the details page of the CAR for more information.

Personalized genome browser and data visualization

CATA provides the latest genome browsing area GIVE[25] to help visualize the open chromatin region. We provide 24 types of cancer and a total of 796 bigwig visualization files. Users can jump on the detail page, and view detailed regional information, or enter a region in the navigation bar and load the corresponding track for visualization. We grouped 23 cancers in detail and named the samples according to the TCGA-patient ID.

Online analysis tools

CATA provides three analytical tools, including that: (1). Cell Pathway Analysis, in which the researchers analyzed pathways for binding transcription factors on the chromatin region and found upstream regulatory pathways. Users only need to input the PEAK-ID of interest to enrich. We provide ten choices about pathway databases (KEGG, Reactome, NetPath, WikiPathways, PANTHER, PID, HumanCyc, CTD, SMPDB and INO). According to the formula:

.

$$P = 1 - \sum_{i=0}^{a-1} \frac{\binom{n}{i}\binom{t-n}{z-i}}{\binom{t}{z}}$$

Where t is the number of genes of the entire genome, and z is number of genes of interest, of which a gene involved in the pathway containing n gene. The calculated P value is provided in the results page, along with the relevant genes, as well as the pathway ID (the pathway ID can be clicked into the view channel details); (2). For the relevant region analysis, the user can input one or more genomic coordinates, while CATA it is also supported to enter a standard BED file. In the results page, we will provide the associated chromatin accessible area information. The user can jump to the detail page by clicking PEKA-ID; (3). Survival analysis, CATA allows users to select relevant genes for analysis in the detail page, which was implemented by built-in gepia2[26] (python3 package).

Data download

Data from 24 types of cancer (ACC, BLCA, BRCA, CESC, CHOL, COAD, ESCA, GBM, HNSC, KIRC, KIRP, LGG, LIHC, LUAD, LUSC, MESO, PCPG, PRAD, SKCM, STAD, TGCT, THCA, UCEC, PAN) was available to download, including transcription factor information in CAR, and gene annotation information in CARs, users can click the icon to download interest data.

System design and Implementation

CATA was developed using MySQL 5.7.16 (http://www.mysql.com) and runs on a Linux-based Tomcat Web server (http://tomcat.apache.org/). We used JAVA 1.8 (https://www.java.com) for server-side scripting. Front end was designed and builted using Bootstrap v4.1.1 (https://v4.bootcss.com). The network visualization was accomplished through Echarts[27] (https://www.echartsjs.com/) Chromatin interaction visualization was supplied by Genomic Interaction Visualization Engine[25] (GIVE) (https://zhong-lab-ucsd.github.io/GIVE_homepage/). We recommend using a modern web browser that supports the HTML5 standard such as Firefox, Google Chrome for the best display.

Discussion

CATA is a tumor chromatin accessible regions database for cancer, storing 29,366,632 accessible chromatin regions from 24 types of cancers, including pan-cancer. Meanwhile, the CATA database annotates 2,936,663 CARs and stores the binding site of 1,678 transcription factors, which could be a good predictor for cancer-

specific transcription factors. CATA also integrates data of methylation, SNP, SNV from TCGA and CATA has a favorable interactive interface for users. Four kinds of searching methods are provided by the database, which are based on cancer type, transcription factor, gene symbol, and advance search through genomic location, respectively. CATA also provides survival analysis of some CAR genes via the built-in gepia2 (python package). In addition, CATA also supports pathway analysis of transcription factors binding to CAR. All of that were aim to help cancer researcher easy to mine potential information for cancer mechanism. However, there are still some deficiencies in CATA. Since the existing chip-seq data cannot perfect diverse for cancer species, It is believed that an explosion of these data will occur, or perhaps relevant chip-seq data and single-cell cancer ATAC-seq data will be added in the second edition to embellish the database, which enables better tumorigenesis mechanism and cancer markers mining.

Acknowledgements

We thank TCGA for sharing their cancer chromatin accessibility data.

We thank Richard A. Young and his colleagues for sharing ROSE program to this work.

We thank Xiaoyi Cao help us to deploy Give genome-browsers.

We thank Zemin Zhang and his colleagues for sharing GIPIA2(python package) to this work.

Key points

- CATA is the first comprehensive resource for chromatin accessibility for cancer.
- CATA provides TF-binding sites, methylation levels, enhancers, CNV, SNV and SNP for CAR.
- CATA documents 2,991,163 CARs from 410 cancer samples.
- CATA genome-browser provides visualization of chromatin accessibility for 23 types of cancer.

Funding:

This work was supported by grants from the National Natural Science Foundation of China (81871921); the Natural Science Foundation of Guangdong Province-Outstanding Youth Project (2019B151502059) and the Basic & Applied Basic Research Programs of Guangdong province (2018KZDXM033).

Reference：

1. Corces MR, Granja JM, Shams S, et al. The chromatin accessibility landscape ofprimary human cancers. Science (80-.). 2018; 362:

2. Buenrostro JD, Wu B, Chang HY, et al. ATAC-seq: A method for assaying chromatin accessibility genome-wide. Curr. Protoc. Mol. Biol. 2015; 2015:21.29.1-21.29.9

3. John S, Sabo PJ, Thurman RE, et al. Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. Nat. Genet. 2011; 43:264–268

4. Klemm SL, Shipony Z, Greenleaf WJ. Chromatin accessibility and the regulatory epigenome. Nat. Rev. Genet. 2019; 29–35

5. Tsompana M, Buck MJ. Chromatin accessibility: A window into the genome. Epigenetics and Chromatin 2014; 7:1–16

6. Lambert SA, Jolma A, Campitelli LF, et al. The Human Transcription Factors. Cell 2018; 172:650–665

7. Mao S. Cancer chromatin accessibility landscape. Science (80-.). 2018; 362:417B

8. Yao L, Shen H, Laird PW, et al. Inferring regulatory element landscapes and transcription factor networks from cancer methylomes. Genome Biol. 2015; 16:105

9. Xie J-J, Jiang Y-Y, Jiang Y, et al. Increased Expression of the Long Non-coding RNA LINC01503, Regulated by TP63, in Squamous Cell Carcinoma and Effects on Oncogenic Activities of Cancer Cell Lines. Gastroenterology 2018;

10. Andersson R, Gebhard C, Miguel-Escalada I, et al. An atlas of active enhancers across human cell types and tissues. Nature 2014; 507:455–461

11. Altshuler DM, Durbin RM, Abecasis GR, et al. An integrated map of genetic variation from 1,092 human genomes. Nature 2012; 491:56–65

12. Khan A, Fornes O, Stigliani A, et al. Erratum: JASPAR 2018: Update of the open-

access database of transcription factor binding profiles and its web framework (Nucleic Acids Research (2017) DOI: 10.1093/nar/gkx1126). Nucleic Acids Res. 2018; 46:D1284
13. Quinlan AR, Hall IM. BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics 2010; 26:841–842
14. Grant CE, Bailey TL, Noble WS. FIMO: Scanning for occurrences of a given motif. Bioinformatics 2011; 27:1017–1018
15. Karolchik D, Baertsch R, Diekhans M, et al. The UCSC Genome Browser Database. Nucleic Acids Res. 2003; 31:51–54
16. Bae S, Kweon J, Kim HS, et al. Microhomology-based choice of Cas9 nuclease target sites. Nat. Methods 2014; 11:705–706
17. Lovén J, Hoke HA, Lin CY, et al. Selective inhibition of tumor oncogenes by disruption of super-enhancers. Cell 2013; 153:320–334
18. Sherry ST. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. 2001; 29:308–311
19. Welter D, MacArthur J, Morales J, et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res. 2014; 42:1001–1006
20. Li MJ, Liu Z, Wang P, et al. GWASdb v2: An update database for human genetic variants identified by genome-wide association studies. Nucleic Acids Res. 2016; 44:D869–D876
21. Matys V. TRANSFAC(R) and its module TRANSCompel(R): transcriptional gene regulation in eukaryotes. Nucleic Acids Res. 2006; 34:D108–D110
22. Carithers LJ, Moore HM. The Genotype-Tissue Expression (GTEx) Project. Biopreserv. Biobank. 2015; 13:307–308
23. Ward LD, Kellis M. HaploReg: A resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res. 2012; 40:930–934
24. Gong J, Mei S, Liu C, et al. PancanQTL: Systematic identification of cis -eQTLs and trans -eQTLs in 33 cancer types. Nucleic Acids Res. 2018; 46:D971–D976
25. Cao X, Yan Z, Wu Q, et al. GIVE: Portable genome browsers for personal websites. Genome Biol. 2018; 19:1–8
26. Tang Z, Li C, Kang B, et al. GEPIA: A web server for cancer and normal gene expression profiling and interactive analyses. Nucleic Acids Res. 2017; 45:W98–W102

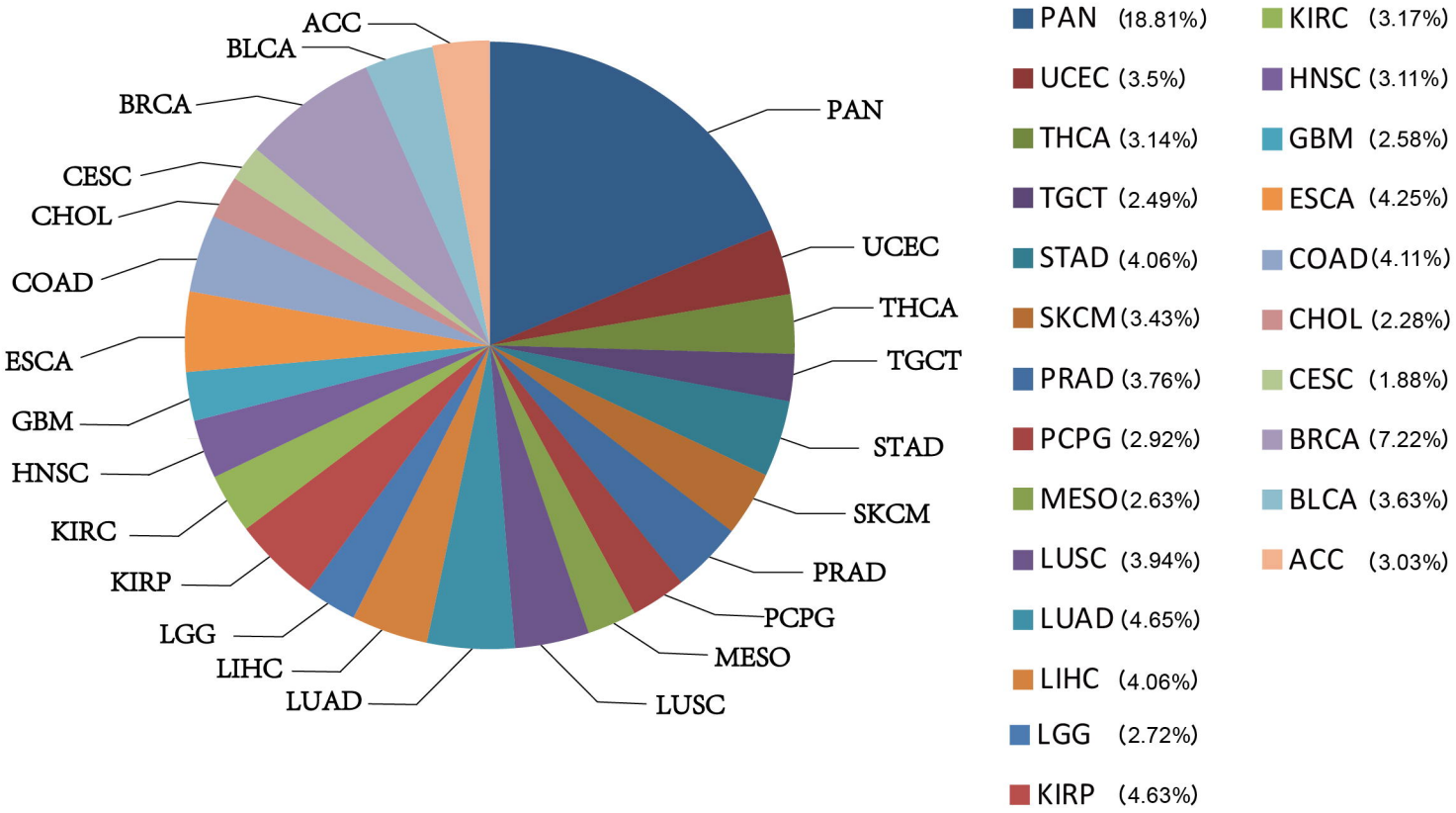
27. Li D, Mei H, Shen Y, et al. ECharts: A declarative framework for rapid construction of web-based visualization. Vis. Informatics 2018; 2:136–146

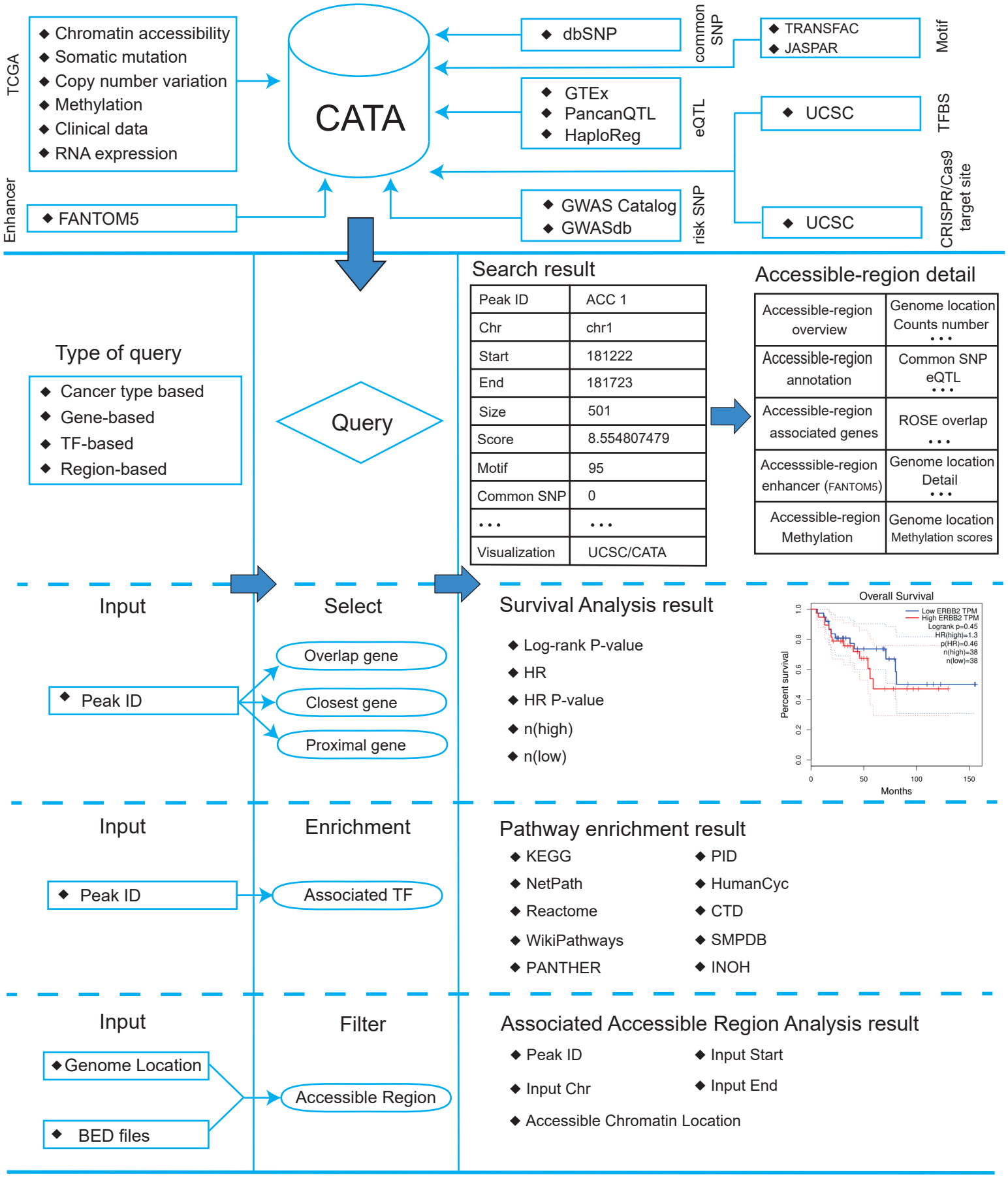
FIGURE LEGENDS

Figure 1. The percentage of chromatin accessible region for per cancer type.

Figure 2. Database content and construction. CATA provide chromatin accessible region of cancer based on TCGA ATAC-seq data, Genetic and epigenetic annotations of accessible regions were collected or calculated including common SNPs, eQTLs, risk SNPs, LD SNPs, TFs, CNV, SNV, methylation sites, Enhancer. CATA also provide ATAC-seq samples associated clinical data. CATA integrates multiple functions including storage, search, download, statistics, visualization, browse and analysis.

Figure 3 The main function and usage of CATA. (A) The navigation bar of CATA. (B) CATA provide user's friendly Explorations. (C) Users query using five methods: 'Search by Cancer type', 'Search by GENE', 'Search by TF', 'and 'Advanced search by genome location'. (D) The search results. (E) Overview of chromatin accessible region. (F) Interactive table of chromatin accessible region, related annotation information. (G) The visualization of RNA-expression.(H) The table of clinical data. (I) Associated gene survival analysis. (J) The visualization of Methylation level. (K) Upstream Transcription factors enrichment. (L) Data download page. (M) Personalized genome browser-GIVE


ACC
BLCA
BRCA
CESC
CHOL
COAD
ESCA
GBM
HNSC
KIRC
KIRP
LGG
LIHC
LUAD
LUSC
MESO
PCPG
PRAD
SKCM
STAD
TGCT
THCA
UCEC
PAN
PAN (18.81%)
UCEC (3.5%)
THCA (3.14%)
TGCT (2.49%)
STAD (4.06%)
SKCM (3.43%)
PRAD (3.76%)
PCPG (2.92%)
MESO (2.63%)
LUSC (3.94%)
LUAD (4.65%)
LIHC (4.06%)
LGG (2.72%)
KIRP (4.63%)
KIRC (3.17%)
HNSC (3.11%)
GBM (2.58%)
ESCA (4.25%)
COAD (4.11%)
CHOL (2.28%)
CESC (1.88%)
BRCA (7.22%)
BLCA (3.63%)
ACC (3.03%)


TCGA
Chromatin accessibility
Somatic mutation
Copy number variation
Methylation
Clinical data
RNA expression
Enhancer
FANTOM5
CATA
dbSNP
common SNP
TRANSFAC
JASPAR
Motif
GTEx
PancanQTL
HaploReg
eQTL
UCSC
TFBS
GWAS Catalog
GWASdb
risk SNP
UCSC
CRISPR/Cas9 target site
Type of query
Cancer type based
Gene-based
TF-based
Region-based
Query
Search result
Peak ID
ACC 1
Chr
chr1
Start
181222
End
181723
Size
501
Score
8.554807479
Motif
95
Common SNP
0
Visualization
UCSC/CATA
Accessible-region detail
Accessible-region overview
Genome location
Counts number
Accessible-region annotation
Common SNP
eQTL
Accessible-region associated genes
ROSE overlap
Accesssible-region enhancer (FANTOM5)
Genome location
Detail
Accessible-region Methylation
Genome location
Methylation scores
Input
Select
Peak ID
Overlap gene
Closest gene
Proximal gene
Survival Analysis result
Log-rank P-value
HR
HR P-value
n(high)
n(low)
Overall Survival
Low ERBB2 TPM
High ERBB2 TPM
Logrank p=0.45
HR(high)=1.3
p(HR)=0.46
n(high)=38
n(low)=38
Percent survival
Months
Input
Enrichment
Peak ID
Associated TF
Pathway enrichment result
KEGG
NetPath
Reactome
WikiPathways
PANTHER
PID
HumanCyc
CTD
SMPDB
INOH
Input
Filter
Genome Location
BED files
Accessible Region
Associated Accessible Region Analysis result
Peak ID
Input Chr
Input Start
Input End
Accessible Chromatin Location

A

Home | Exploration | Search | Analysis | Genome-Browser | Download | Holp

B

Tissue type

Kidney 411439

Bladder 108473

Cancer

ACC 90777

BLCA 108473

Peak ID	CHR Name	start	End	Score	Annotation	Tissue
ACC_1	chr1	181222	181723	8.5548	Distal	Kidney
ACC_10	chr1	906012	906513	7.1711	Intron	Kidney
ACC_100	chr1	1890895	1891396	17.5500	Promoter	Kidney
ACC_1000	chr1	21669756	21669756	1.1321	Promoter	Kidney
ACC_10000	chr2	112275250	112275251	63.5500	Promoter	Kidney

C

Search accessible regions by tumor category-bases

Search accessible regions by gene

Search accessible regions by TF

Advanced search by genome location

D

Peak ID	CHR	Start	End	Score	Annotation	Motif	TFBs	Enhancer	CNV	SNP	SNV	Tissue
ACC_39334	chr8	13133107	13133608	358.09	Promoter	124	3	0	90	21	0	Kidney
ACC_40348	chr8	37762313	37762814	231.63	Promoter	13	8	0	90	7	0	Kidney
ACC_22973	chr5	34929111	34929612	213.07	Promoter	32	5	0	90	3	0	Kidney
ACC_58625	chr12	71663734	71664235	208.62	Promoter	9	7	0	90	6	0	Kidney
ACC_22888	chr5	33440615	33441116	207.26	Promoter	17	2	0	90	7	0	Kidney

E

Accessinle Region Overview

Peak ID : ACC_40348

Cancer type : ACC

Tissue : Kidney

Average counts : 1208.7778

Gene symbol : OVERLAP_GENES : HSPA9

Gene symbol : PROXIMAL_GENES : ETF1 ,SNORD63

Gene symbol : CLOSERST_GENES : SNORD63

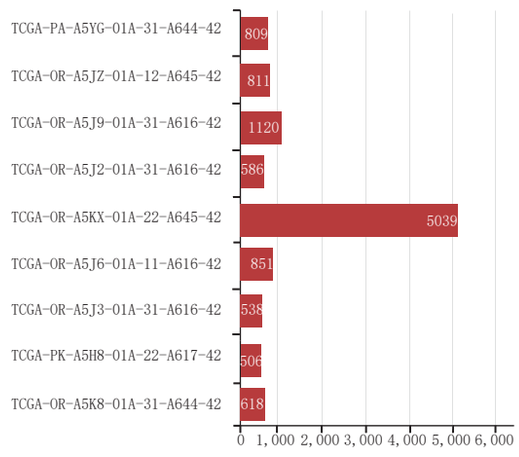
Genome location : chr8: 37762313-37762814

SCORE : 231.6377127

RNA average FPKM : 0.000000

Genome browser : UCSC CATA

Raw counts numbers



F

Accessinle Region Overview

SNP | CNV | TFBS conserved | Motif | SNV | Same sample SNV | Same sample SNV | CRISPR/Cas9 target site | Enhancer

Copy | Excel | PDF | CSV | Print

SNP ID	chr	position	Ref	Alt	AFR	AMR	EAS	EUR	SAS	eQTL	Risk SNP	Motif change
rs116231061	chr8	37762538	T	C	0	0	0	0	0	0	0	0
rs142026785	chr8	37762548	C	T	0	0	0	0	0	0	0	0
rs182986556	chr8	3776265	C	A	0	0	0	0	0	0	0	0
rs35423325	chr8	37762729	G	A	0	0	0	0	0	0	0	0
rs367563760	chr8	37762636	G	T	0	0	0	0	0	0	0	0
rs371407450	chr8	37762651	C	A, T	0	0	0	0	0	0	0	0
rs528390640	chr8	37762502	G	A	0	0	0	0	0	0	0	0
rs6468438	chr8	37762558	A	C, G	0	0	0	0	0	0	0	0

showing 1 to 8 of 8 entries

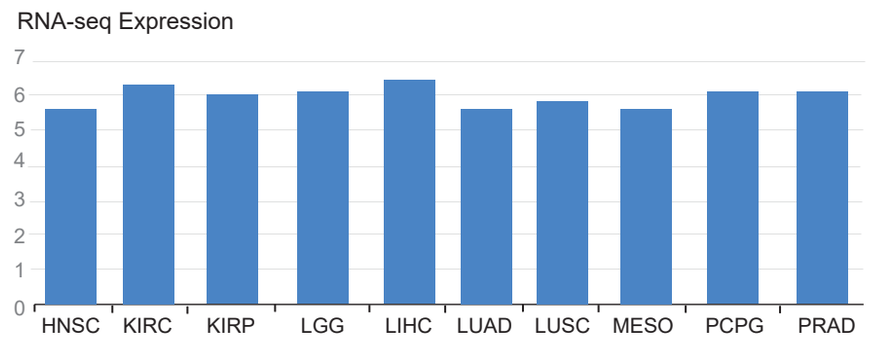
First | Previous | 1 | Next | Last

G

RNA-seq Expression

Overlap gene | HSPA9

RNA-seq Expression



H

Associative clinical data

Copy | Excel | PDF | CSV | Print

samples id	age	anatomic treatment site	batch number	brc	bcr drug barcode
TCGA-L7-V6AZ-01A	62	Primary Tumor Field	339.55.0	Nationwide Children's Hospital	TCGA-L7-A6VZ-D52184
TCGA-IG-A4QS-01A	71		272.64.0	Nationwide Children's Hospital	
TCGA-IG-A3YA-01A	53		254.64.0	Nationwide Children's Hospital	
TCGA-IG-A7DP-01A	50		339.55.0	Nationwide Children's Hospital	TCGA-IG-A7DP-D62774
TCGA-LN-A8HZ-01A	56	Primary Tumor Field	374.55.0	Nationwide Children's Hospital	TCGA-LN-A8HZ-D69839
TCGA-LN-A4A2-01A	57	Primary Tumor Field	272.64.0	Nationwide Children's Hospital	TCGA-LN-A4A2-D69650
TCGA-LN-A49W-01A	73	Primary Tumor Field	272.64.0	Nationwide Children's Hospital	TCGA-LN-A49W-D69630
TCGA-L5-A4OE-01A	81		272.64.0	Nationwide Children's Hospital	

I

Survival analysis

Proximal gene | ETF1

Methods: Overall Survival | Disease Free Survival (RFS)

Group Cutoff: Median | Quartile | Custom

Cutoff-High(%) 50 | Cutoff-High(%) 50

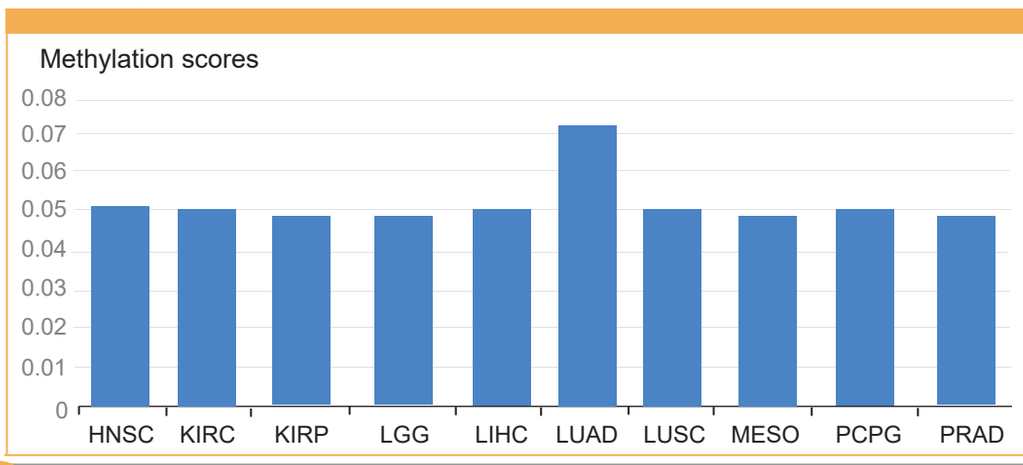
Hazards Ratio (HR): Yes | 95% Confidence Interval: Yes | Axis Units: Months | plot

Calculate the hazards ratio based on Cox PH Model. Add the 95% CI as dotted line.

J

Methylation visualization

Methylation scores



K

Upstream Transcription Factor Enrichment

Peak ID : ACC_40348

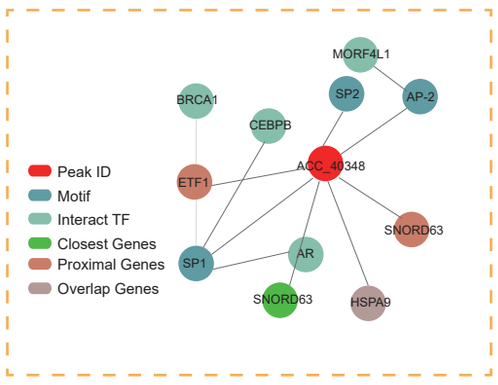
Rose Overlap: HSPA9

Rose Proximal: ETF1,SNORD63

Rose closest: SNORD63

Motif: EBF1, EBF1_full, Sp1, SP1, YY1, AP-2, CTCF, SP2, HINFP

Interact Transcription Factor: BCL6, CREBBP, HOMEZ, ZNF423, AHR, AR, ARNT, ASH2L, BCL11B, BRCA1, CEBPB, CRX, CTCFL, DLX4, DNMT1, E2F1, E2F2, E2F3, EGR1, ELF1, EPAS1, ESR1, ESR2,.....



L

Chromatin accessible region data download

Months

10 entries

Search:

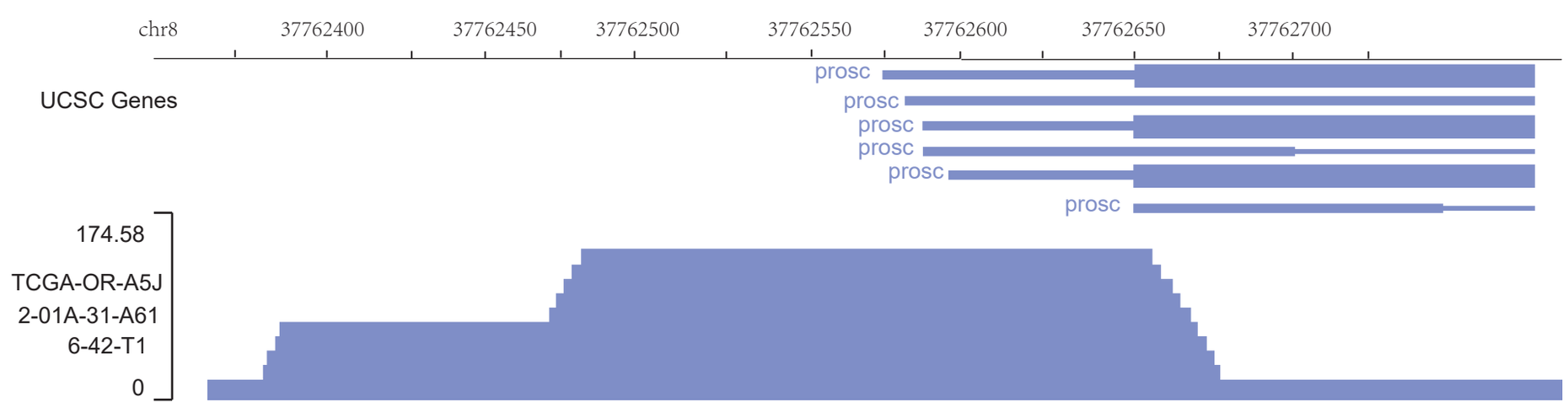
Cancer Type	Tissue Type	Class	
ACC	Kidney	Gene annotation	Associated TF
BLCA	Bladder	Gene annotation	Associated TF
BRCA	Breast	Gene annotation	Associated TF
CESC	Uterus	Gene annotation	Associated TF
CHOL	Liver	Gene annotation	Associated TF
COAD	Colon	Gene annotation	Associated TF
ESCA	Esophagus	Gene annotation	Associated TF
GBM	Brain	Gene annotation	Associated TF
HNSC	Head	Gene annotation	Associated TF
KIRC	Kidney	Gene annotation	Associated TF

M

CANCEL | UPDATE COORDINATE

Selected tracks (hg38)

UCSC Genes

TCGA-OR-A5J2-01A-31-A...-T1

DEFAULT TRTACKS | + ADD TRACKS