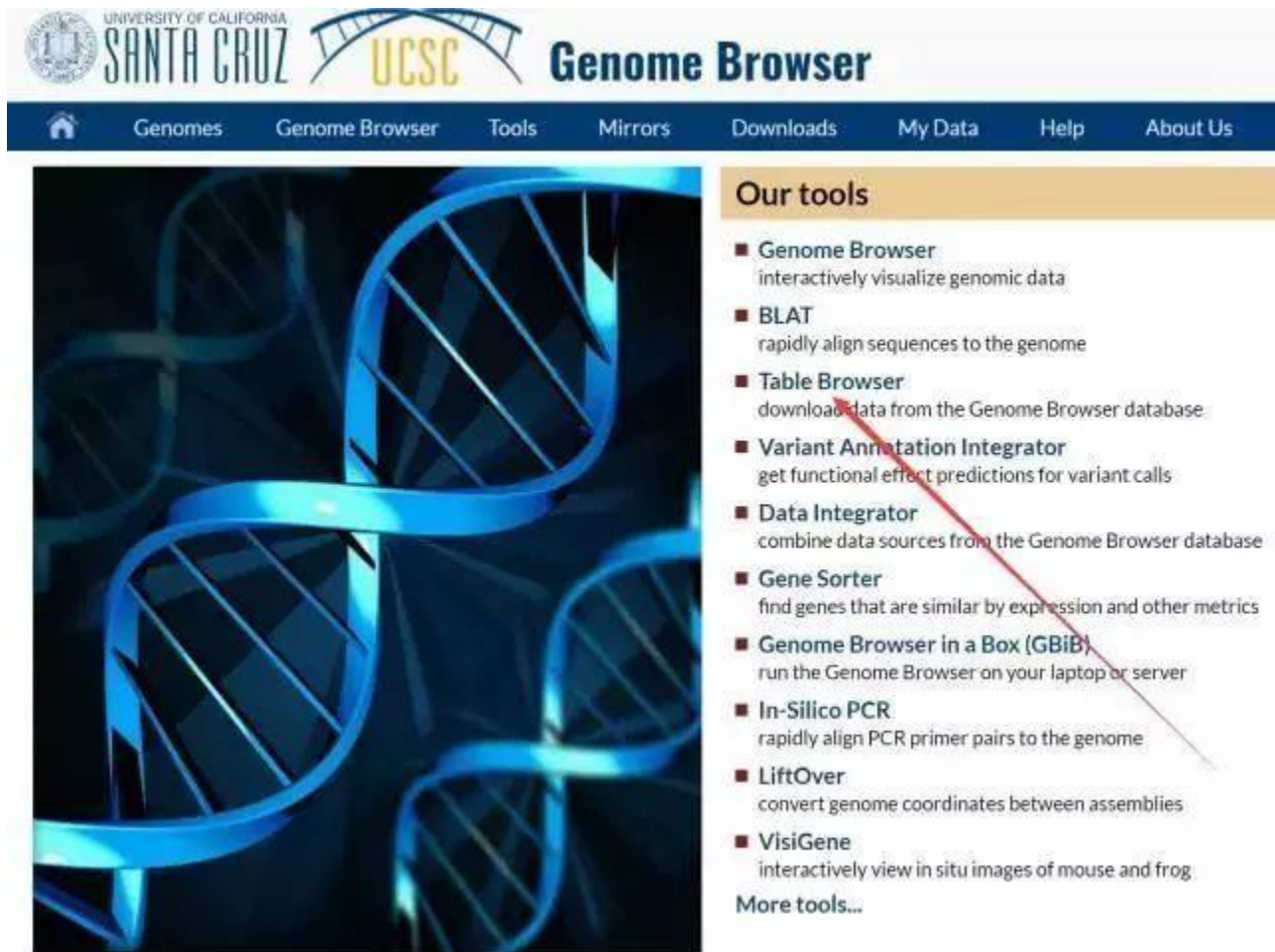


## Locate promoter sequence for a specific gene

1. Go to website UCSC: <http://www.genome.ucsc.edu/>, choose “Table Browser”:



UNIVERSITY OF CALIFORNIA  
SANTA CRUZ UCSC Genome Browser

Genomes Genome Browser Tools Mirrors Downloads My Data Help About Us

### Our tools

- **Genome Browser**  
interactively visualize genomic data
- **BLAT**  
rapidly align sequences to the genome
- **Table Browser**  
download data from the Genome Browser database
- **Variant Annotation Integrator**  
get functional effect predictions for variant calls
- **Data Integrator**  
combine data sources from the Genome Browser database
- **Gene Sorter**  
find genes that are similar by expression and other metrics
- **Genome Browser in a Box (GBiB)**  
run the Genome Browser on your laptop or server
- **In-Silico PCR**  
rapidly align PCR primer pairs to the genome
- **LiftOver**  
convert genome coordinates between assemblies
- **VisiGene**  
interactively view in situ images of mouse and frog

More tools...

2. Type into the gene information, click “get output”:

Genomes Genome Browser Tools Mirrors Downloads My Data Help About

### Table Browser

Use this program to retrieve the data associated with a track in text format, to calculate intersections between tracks, or to calculate the intersection of two tracks. For more information on this application see [Using the Table Browser](#) for a description of the controls in this form, the [User's Guide](#) for a tutorial for a narrated presentation of the software features and usage. For more complex queries, you may want to use the [function of your set](#) through annotation enrichments, send the data to [GREAT](#). Send data to [GenomeSpace](#) for a list of contributors and usage restrictions associated with these data. All tables can be downloaded in their entirety.

**clade:** Mammal **genome:** Human **assembly:** Dec. 2013 (GRCh38/hg38)  
**group:** Genes and Gene Predictions **track:** GENCODE v24 [add custom tracks](#) [track hubs](#)  
**table:** knownGene [describe table schema](#)  
**region:**  genome  position chr6:52186387-52190638 [lookup](#) [define regions](#)  
**identifiers (names/accessions):** [paste list](#) [upload list](#)  
**filter:** [create](#)  
**intersection:** [create](#)  
**correlation:** [create](#)  
**output format:** sequence [Send output to](#)  Galaxy  GREAT  GenomeSpace  
**output file:**  (leave blank to keep output in browser)  
**file type returned:**  plain text  gzip compressed  
[get output](#) [summary/statistics](#)

To reset **all** user cart settings (including custom tracks), [click here](#).

2. Choose “genomic”:

Genomes Genome Browser Tools

### Select sequence type for GENCODE v24

genomic  
 protein  
 mRNA  
[submit](#) [cancel](#)

4. Choose “Promoter/Upstream by”, change to “2000 bases”, then click “get sequence”:



## GENCODE v24 Genomic Sequence

### Sequence Retrieval Region Options:

- Promoter/Upstream by  bases
  - 5' UTR Exons
  - CDS Exons
  - 3' UTR Exons
  - Introns
  - Downstream by  bases
  - One FASTA record per gene.
  - One FASTA record per region (exon, intron, etc.) with 
    - Split UTR and CDS parts of an exon into separate FA
- Note: if a feature is close to the beginning or end of a chromosome.

### Sequence Formatting Options:

- Exons in upper case, everything else in lower case.
- CDS in upper case, UTR in lower case.
- All upper case.
- All lower case.
- Mask repeats:  to lower case  to N

5. The sequence from beginning to the first capital letter is the promoter sequence.

## Predict transcription factor binding site on given promoter sequence

1. go to website jaspar: <http://jaspar.genereg.net/>;
2. Search your gene such as TBX3 (human), the possible binding site will be displayed ;
3. Click Scan, type the promoter sequence (FASTA type) into this region. For example PLD1 promoter sequence, Click “scan”;

The screenshot shows the JASPAR website interface. At the top, there is a search bar with "TBX3" entered and a "Search" button. Below the search bar, it says "Examples: SPI1, P17676, CHIP-seq, Homo sapiens". The main content area shows "1 profile(s) found". Below this, there is a table with columns: ID, Name, Species, Class, Family, and Logo. The table contains one row with ID "MA1566.1", Name "TBX3", Species "Homo sapiens", Class "T-Box factors", and Family "TBX2-related factors". A red box highlights the "Add to cart" button (a green checkmark) in the first column of the table. To the right of the table is a panel titled "Analyze selected profiles". It contains instructions: "Please select matrix profiles on the left side to add to your cart to perform the following analysis." Below this are buttons for "Add to cart" and "Scan". There is a text input field for "Enter FASTA sequence here: (955 nucleotides left)". The input field contains the following sequence: 

```
> PLD1 Promoter seq
aaaaatcacaaaattctctcaggaaaataaaataattacatacaaat
catactacttctaaattcattgtttggagttcc
ataagcttcagttctctctaaacaaacataggcattctctatttctct
ttgaaaggaggaagaattcagita
ttgaaaggaggaagaattcagita
```

 Below the input field is a "Relative profile score threshold" set to 80 and a "Scan" button. Two red arrows point to the "Scan" button in the right panel and the "Add to cart" button in the table.

ID	Name	Species	Class	Family	Logo
MA1566.1	TBX3	Homo sapiens	T-Box factors	TBX2-related factors	

4. Click “CSV” to download results, choose the possible binding sites according score.

Total 21 putative site(s) were predicted with relative profile score threshold 80%.

Show FASTA Sequence

Display 10 profiles

Filter:

Matrix ID	Name	Score	Relative score	Sequence ID	Start	End	Strand	Predicted sequence
MA1566.1	TBX3	10.831	0.960361764321	PLD1	470	479	+	CAGGTGTGAG
MA1566.1	TBX3	8.31373	0.897431301934	PLD1	1517	1526	-	GAAGTGTAT
MA1566.1	TBX3	7.74044	0.8830995629	PLD1	567	576	-	TAAGTGCAC
MA1566.1	TBX3	7.23403	0.870439837637	PLD1	661	670	+	TACGTGTGTG
MA1566.1	TBX3	7.01705	0.865015591515	PLD1	1875	1884	-	GGGGCGTCGC
MA1566.1	TBX3	6.99837	0.864548463788	PLD1	1890	1899	-	GCGGTGCCAG
MA1566.1	TBX3	6.9468	0.863259180769	PLD1	1584	1593	-	GAGGTGACAT
MA1566.1	TBX3	6.66929	0.856321868515	PLD1	887	896	+	GCTGTGTGAC
MA1566.1	TBX3	6.66929	0.856321868515	PLD1	1462	1471	+	GCTGTGTGAC
MA1566.1	TBX3	6.37353	0.848928001998	PLD1	455	464	+	AAAGTGTGG

Showing 1 to 10 of 21 entries

Previous 1 2 3 Next

Copy CSV