



# Course organization

- **Course introduction ( Week 1)**
  - Code editor: Emacs (Week 2)
- **Part I: Introduction to C programming language (Week 3 - 12)**
  - Chapter 1: Overall Introduction (Week 3-4)
  - Chapter 2: Types, operators and expressions (Week 5)
  - Chapter 3: Control flow (Week 6)
  - Chapter 4: Functions and program structure (Week 7)
  - Chapter 5: Pointers and arrays (Week 8)
  - Chapter 6: Structures (Week 10)
  - Chapter 7: Input and Output (Week 11)
- **Part II: Skills others than programming languages (Week 12)**
  - Debugging tools (Week 12)
- **Part III: Reports from the battle field (student forum) (Week 13– 16)**
  - Presentation (week 13-14)
  - Demo (week 15-16)



上海交通大学  
SHANGHAI JIAO TONG UNIVERSITY



# The final project

Chaochun Wei

Shanghai Jiao Tong University

Spring 2019



# Contents



## Final project

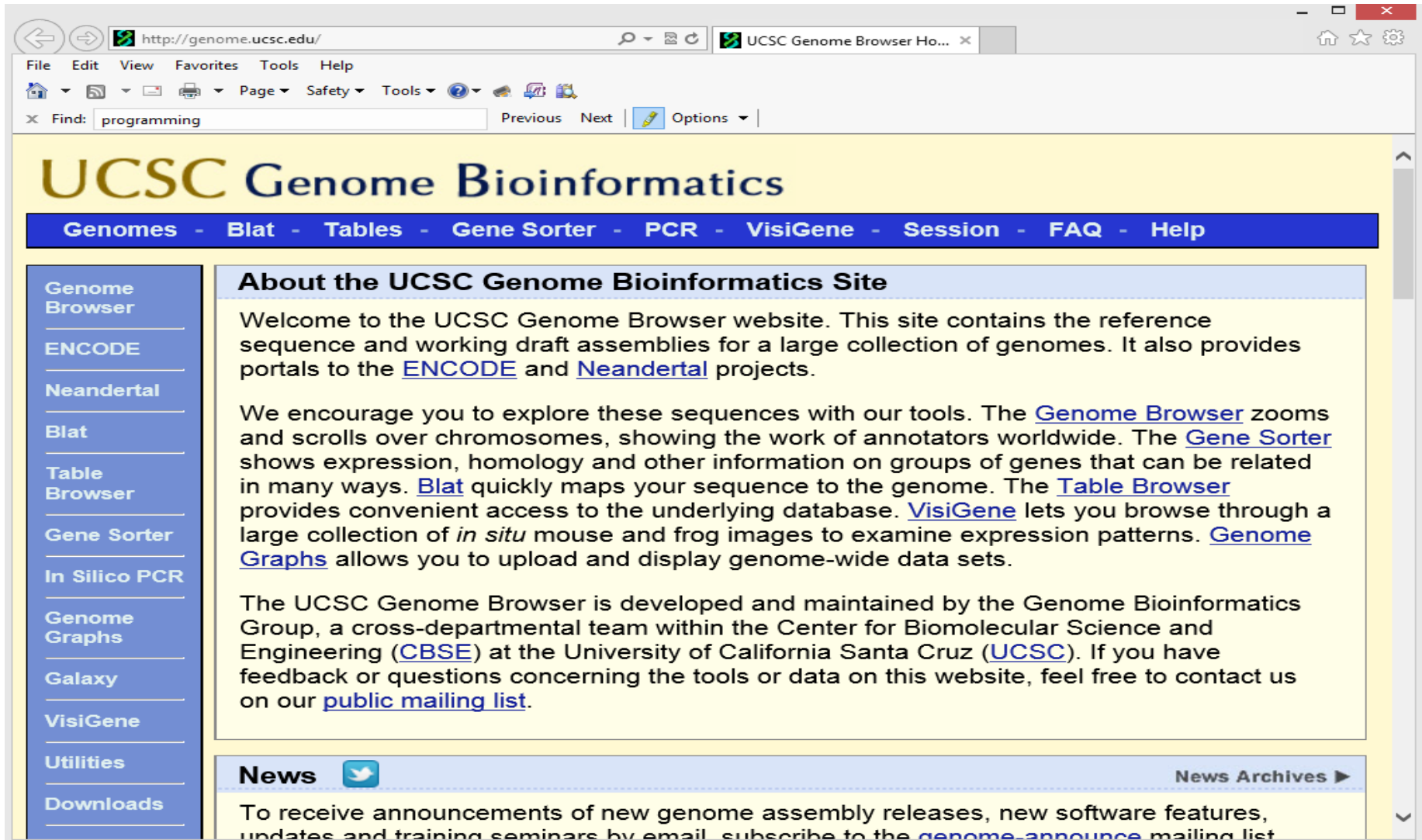
- Background
- Presentation content
- Presentation arrangement



# Background of the final project



## UCSC Genome Browser:



The screenshot shows a web browser window displaying the UCSC Genome Browser homepage. The browser's address bar shows the URL <http://genome.ucsc.edu/>. The page features a navigation menu with links to Genomes, Blat, Tables, Gene Sorter, PCR, VisiGene, Session, FAQ, and Help. A sidebar on the left lists various tools and resources, including Genome Browser, ENCODE, Neandertal, Blat, Table Browser, Gene Sorter, In Silico PCR, Genome Graphs, Galaxy, VisiGene, Utilities, and Downloads. The main content area is titled "About the UCSC Genome Bioinformatics Site" and contains a welcome message and a description of the site's resources. At the bottom, there is a "News" section with a Twitter icon and a "News Archives" link.

**UCSC Genome Bioinformatics**


Genomes - Blat - Tables - Gene Sorter - PCR - VisiGene - Session - FAQ - Help

**About the UCSC Genome Bioinformatics Site**

Welcome to the UCSC Genome Browser website. This site contains the reference sequence and working draft assemblies for a large collection of genomes. It also provides portals to the [ENCODE](#) and [Neandertal](#) projects.

We encourage you to explore these sequences with our tools. The [Genome Browser](#) zooms and scrolls over chromosomes, showing the work of annotators worldwide. The [Gene Sorter](#) shows expression, homology and other information on groups of genes that can be related in many ways. [Blat](#) quickly maps your sequence to the genome. The [Table Browser](#) provides convenient access to the underlying database. [VisiGene](#) lets you browse through a large collection of *in situ* mouse and frog images to examine expression patterns. [Genome Graphs](#) allows you to upload and display genome-wide data sets.

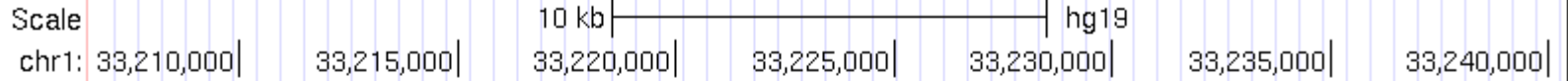
The UCSC Genome Browser is developed and maintained by the Genome Bioinformatics Group, a cross-departmental team within the Center for Biomolecular Science and Engineering ([CBSE](#)) at the University of California Santa Cruz ([UCSC](#)). If you have feedback or questions concerning the tools or data on this website, feel free to contact us on our [public mailing list](#).

**News**  [News Archives](#) ▶

To receive announcements of new genome assembly releases, new software features, updates and training seminars by email, subscribe to the [genome-announce](#) mailing list



# mRNA-seq mapping (alignments)



UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)

KIAA1522  
KIAA1522  
KIAA1522  
KIAA1522

H1-hESC 200 bp paired read RNA-seq Alignments Rep 1 from ENCODE/Caltech

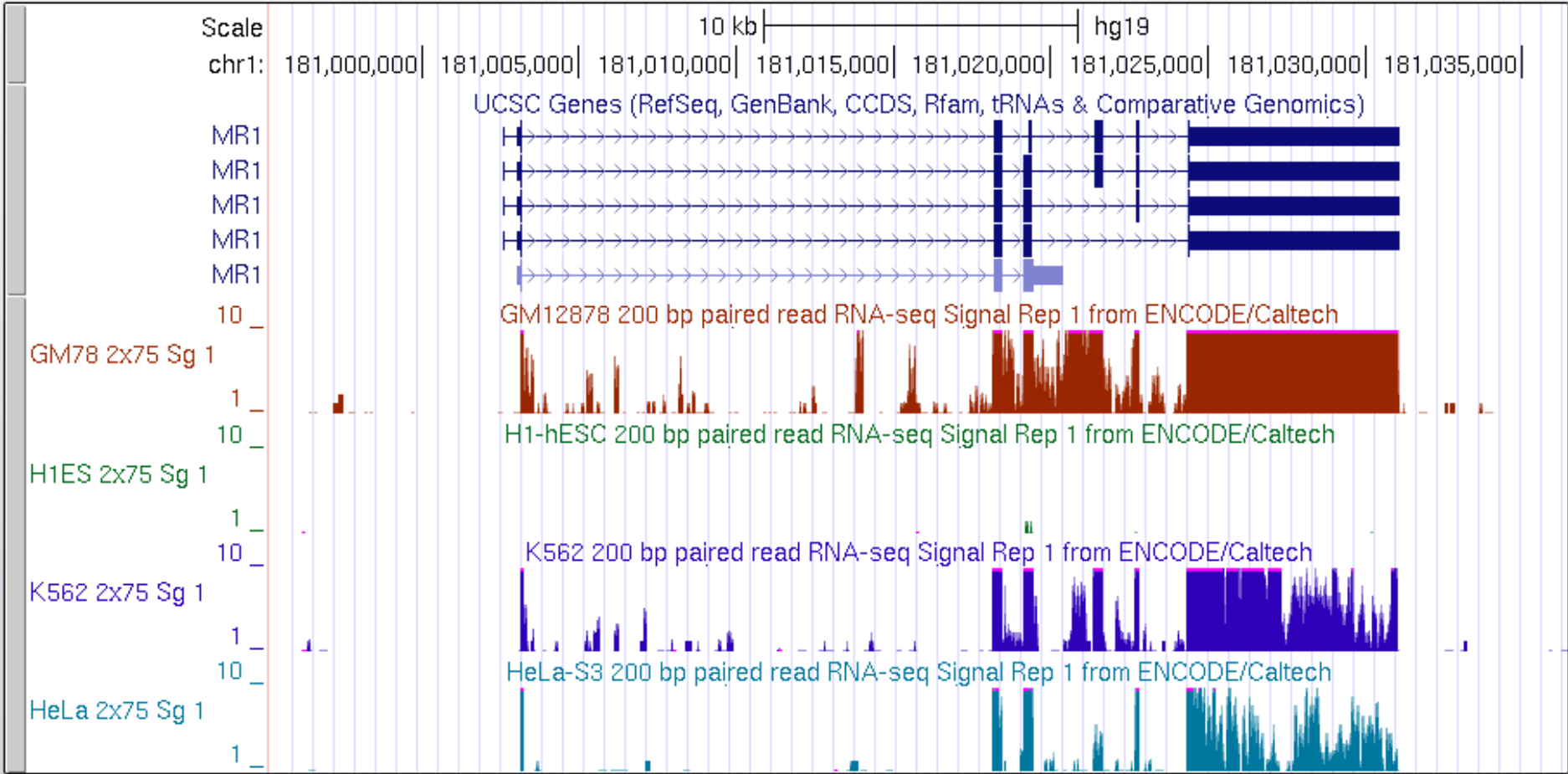


Mapping of RNA-seq data of H1-hESC cell line in ENCODE Project



# mRNA-seq mapping (signals)

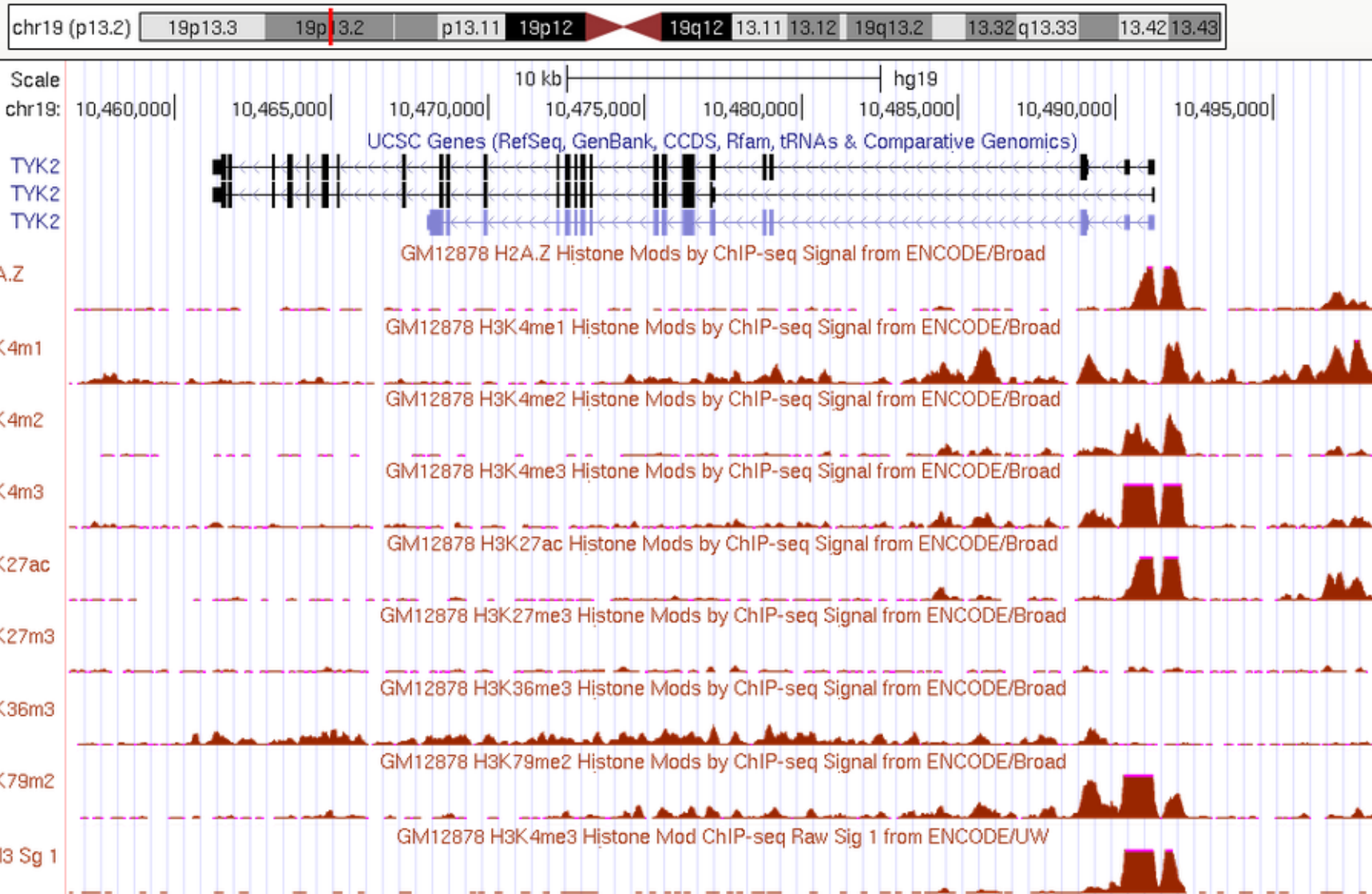
chr1 (q25.3) p31.1 1q12 q41 43



mRNA-seq signals for 4 cell lines in ENCODE Project



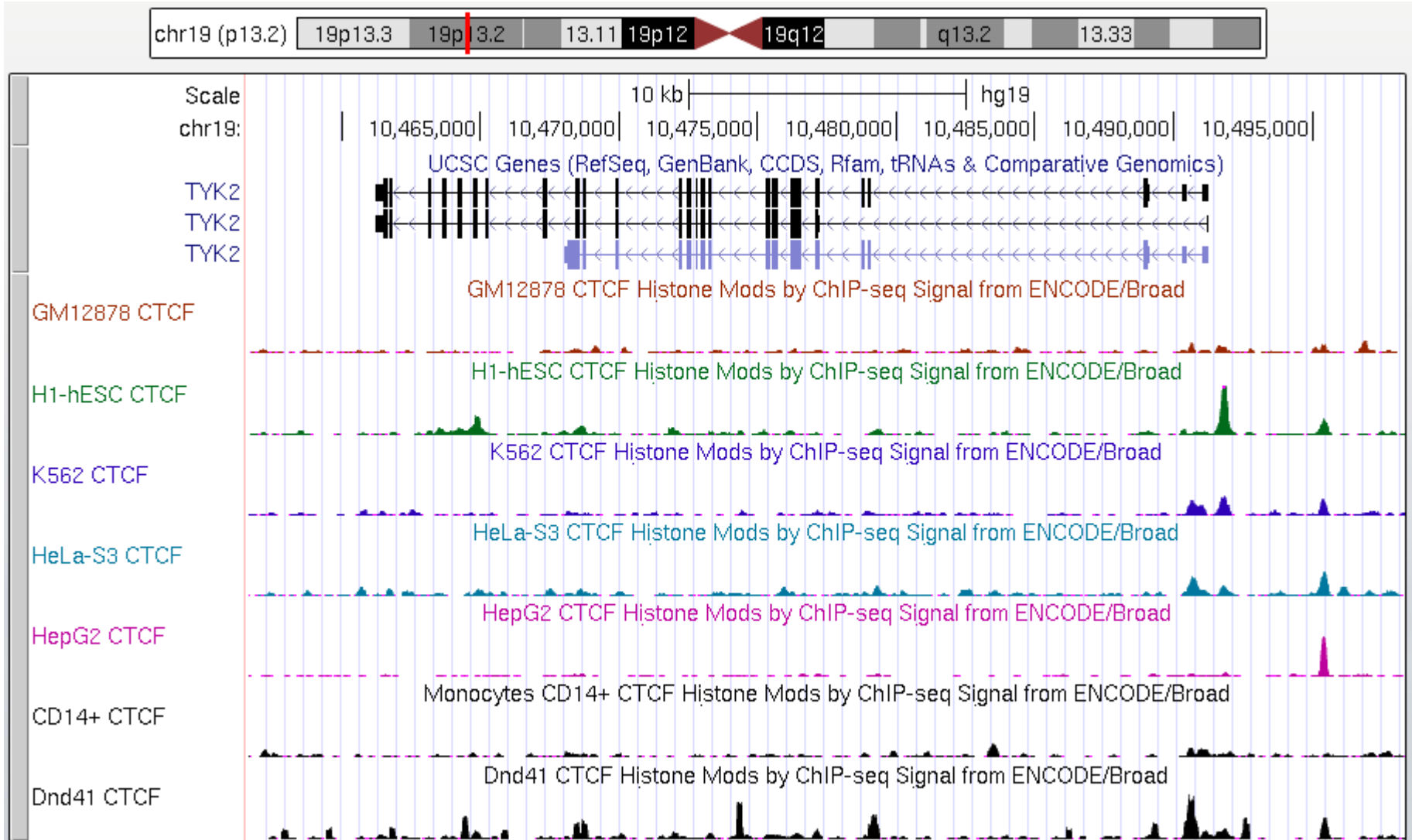
# Histone Modification (Chip-seq signals)



Histone modification of GM12878 cell line in ENCODE Project



# Transcript Factor Binding Sites (Chip-seq signals)

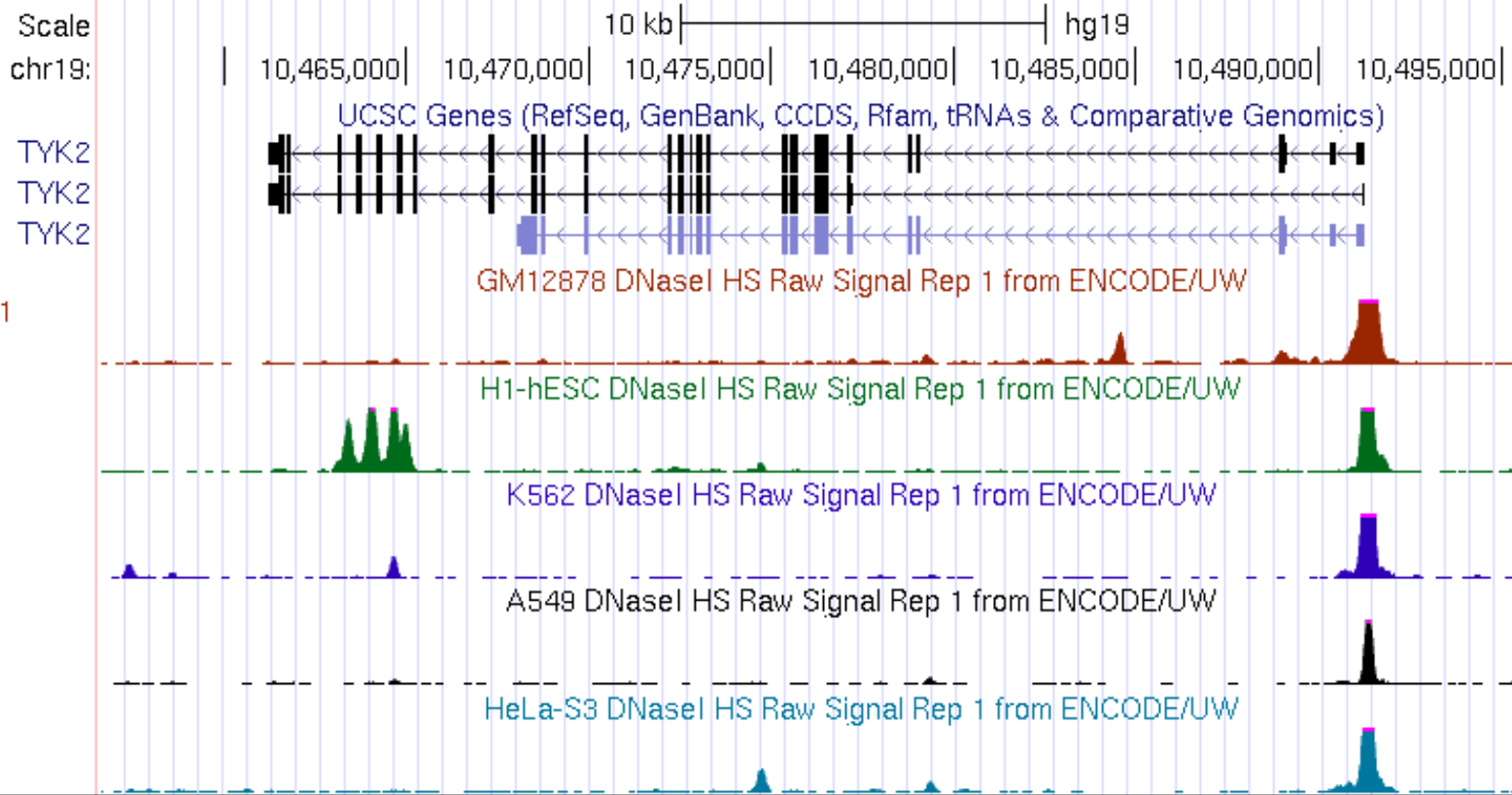


CTCF ChIP-seq signals for 7 cell lines in ENCODE Project





# DNase-seq Signals



DNase-seq signals for 5 cell lines in ENCODE Project



# GENCODE gene file format

```
#name chrom strand txStart txEnd cdsStart cdsEnd exonCount exonStarts exonEnds proteinID alignID
uc057cfx.1 chr1 - 11131593 11145025 11131593 11131593 5
11131593,11134350,11139303,11139532,11144647, 11133197,11134466,11139435,11139658,11145025,
ENST00000495435.1
uc057cfy.1 chr1 - 11138953 11139660 11138953 11138953 2 11138953,11139532, 11139435,11139660,
ENST00000476768.1
uc031plf.2 chr1 + 11143897 11149537 11143897 11143897 4 11143897,11144666,11145152,11149269,
11143987,11144886,11145279,11149537, ENST00000445982.5 uc057cfz.1 chr1 + 11144423 11149537 11144423
11144423 4 11144423,11144666,11145152,11149269, 11144481,11144886,11145279,11149537,
ENST00000420480.1
uc057cga.1 chr1 + 11152349 11152452 11152349 11152349 1 11152349, 11152452, ENST00000517277.1
uc001ase.5 chr1 + 11189340 11195981 11189579 11195023 5
11189340,11192269,11193579,11194460,11194853, 11189955,11192370,11193774,11194659,11195981, 043827
ENST00000376819.3
uc057cgb.1 chr1 + 11193626 11195178 11193626 11193626 3 11193626,11194525,11194853,
11193774,11194659,11195178, ENST00000476934.1
uc057cgc.1 chr1 + 11226253 11226360 11226253 11226253 1 11226253, 11226360, ENST00000384720.
```



# MGC gene file format

```
#bin matches misMatches repMatches nCount qNumInsert qBaseInsert tNumInsert  
tBaseInsert strand qName qSize qStart qEnd tName tSize tStart tEnd blockCount  
blockSizes qStarts tStarts  
83 8146 6 0 0 0 0 56 144275 - BC117166 8155 3 8155 chr1 248956422 11106997 11259424 57  
503, 106, 81, 81, 66, 136, 75, 73, 83, 123, 148, 136, 175, 135, 183, 123, 99, 97, 101, 249, 118, 116, 132, 12  
6, 108, 78, 116, 101, 140, 76, 146, 163, 143, 147, 93, 163, 113, 168, 87, 251, 130, 135, 93, 90, 123, 206, 21  
6, 245, 129, 187, 109, 276, 135, 201, 233, 109, 177,  
0, 503, 609, 690, 771, 837, 973, 1048, 1121, 1204, 1327, 1475, 1611, 1786, 1921, 2104, 2227, 2326, 2423,  
2524, 2773, 2891, 3007, 3139, 3265, 3373, 3451, 3567, 3668, 3808, 3884, 4030, 4193, 4336, 4483, 4576, 4  
739, 4852, 5020, 5107, 5358, 5488, 5623, 5716, 5806, 5929, 6135, 6351, 6596, 6725, 6912, 7021, 7297, 74  
32, 7633, 7866, 7975,  
11106997, 11108180, 11109289, 11109648, 11112851, 11114317, 11114812, 11115395, 11117003, 11121  
245, 11121978, 11124497, 11126621, 11127009, 11127623, 11128003, 11128453, 11128854, 11129737, 1  
1130528, 11133079, 11134350, 11139303, 11139532, 11144647, 11144967, 11146675, 11150125, 111571  
51, 11167441, 11199257, 11199540, 11204560, 11209311, 11210813, 11212311, 11212795, 11213398, 11  
216147, 11228667, 11230924, 11231299, 11232435, 11233397, 11234142, 11237842, 11238401, 1124030  
2, 11241552, 11243113, 11247624, 11247818, 11253838, 11255991, 11256932, 11258484, 11259247,  
  
670 2238 0 0 0 0 0 4 4403 + BC001881 2255 0 2238 chr1 248956422 11189340 11195981 5  
615, 101, 195, 199, 1128, 0, 615, 716, 911, 1110,  
11189340, 11192269, 11193579, 11194460, 11194853,
```



# UCSC genome browser: table browser

Table Browser

Use this program to retrieve the data associated with a track in text format, to calculate intersections between tracks, and to retrieve DNA sequence covered by a track. For help in using this application see [Using the Table Browser](#) for a description of the controls in this form, the [User's Guide](#) for general information and sample queries, and the OpenHelix Table Browser [tutorial](#) for a narrated presentation of the software features and usage. For more complex queries, you may want to use [Galaxy](#) or our [public MySQL server](#). To examine the biological function of your set through annotation enrichments, send the data to [GREAT](#). Refer to the [Credits](#) page for the list of contributors and usage restrictions associated with these data. All tables can be downloaded in their entirety from the [Sequence and Annotation Downloads](#) page.

**clade:** Mammal **genome:** Human **assembly:** Feb. 2009 (GRCh37/hg19)

**group:** Genes and Gene Predictions **track:** UCSC Genes

**table:** knownGene

**region:**  genome  ENCODE Pilot regions  position chr16:28468329-28474878

**identifiers (names/accessions):**

**filter:**

**intersection:**

**correlation:**

**output format:** BED - browser extensible data  Send output to  [Galaxy](#)  [GREAT](#)

**output file:** (leave blank to keep output in browser)

**file type returned:**  plain text  gzip compressed



- **Comparison and evaluation of these different data sets**
  - **Use UCSC Table browser, intersection function**
  - **Write your own program**



# The Final Project

- **Final project (50)**
  - Report (30)
  - **Demo**
  - Presentation (20)
    - content
- **Presentation and demo arrangement**



# Presentation and demo



## Presentation

- **45% about your project ( project design and/or implementation)**
- **45% about the C program language**
  - A chapter assigned to everyone
  - Your memorable C programming experience
- **Time: 6 minutes (4 + 2) (10% Timing)**



## Demo

- **show and tell your project**
- **Test your program with different input files**
- **Time: 4 minutes**



# Questions?