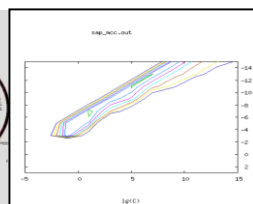
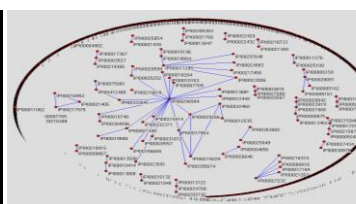
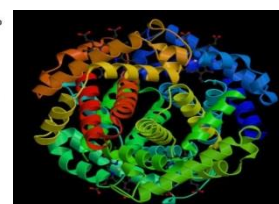
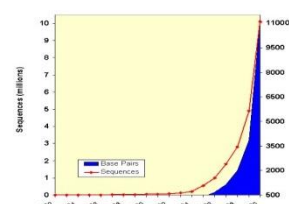


北京大学生物信息学中心 魏丽萍

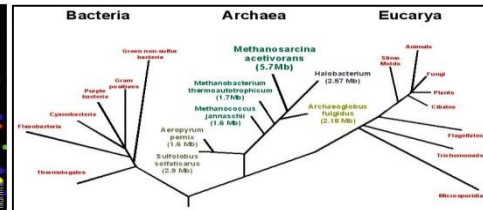
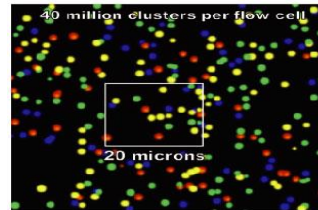
Liping Wei, Ph.D.

Center for Bioinformatics, Peking University





TAACCCTAACCCTAACCCTAACCCTAACCCTA
 CCTAACCCCTAACCCCTAACCCCTAACCCCTAACCC
 CCCTAACCCCTAACCCCTAACCCCTAACCCCTAAC
 AACCCCTAACCCCTAACCCCTAACCCCTAACCCCTA
 ACCCTAACCCCAACCCCAACCCCAACCCCAAC
 CTACCCTAACCCCTAACCCCTAACCCCTAACCCCTA
 ACCCTAACCCCTAACCCCTAACCCCTAACCCCTAA

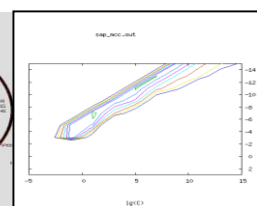
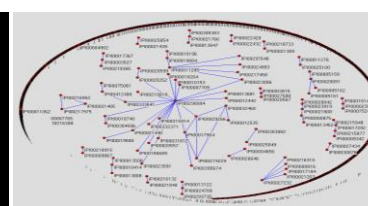
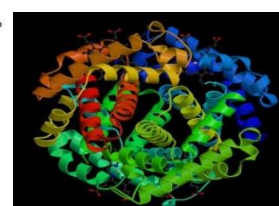
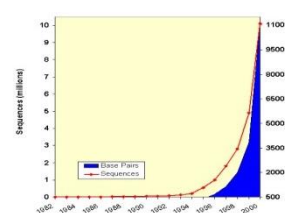


Unit 4: UCSC Genome Bioinformatics

北京大学生物信息学中心 魏丽萍

Liping Wei, Ph.D.

Center for Bioinformatics, Peking University



UCSC Genome Bioinformatics (<http://genome.ucsc.edu/>)

UCSC Genome Bioinformatics

[Genomes](#) - [Blat](#) - [Tables](#) - [Gene Sorter](#) - [PCR](#) - [VisiGene](#) - [Session](#) - [FAQ](#) - [Help](#)

[Genome Browser](#)

[ENCODE](#)

[Neandertal](#)

[Blat](#)

[Table Browser](#)

[Gene Sorter](#)

[In Silico PCR](#)

[Genome Graphs](#)

[Galaxy](#)

[VisiGene](#)

[Utilities](#)

[Downloads](#)

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.

27 November 2013 - 100 Species Conservation Track now available on hg19

After 15.4 years of CPU run time is 0.005504 individual label and 00 cluster time for last-minute alignment, we are excited to announce

UCSC Genome Browser (<http://genome.ucsc.edu/>)

[Genomes](#)[Genome Browser](#)[Tools](#)[Mirrors](#)[Downloads](#)[My Data](#)[About Us](#)[Help](#)

Human (*Homo sapiens*) Genome Browser Gateway

The UCSC Genome Browser was created by the [Genome Bioinformatics Group of UC Santa Cruz](#).
Software Copyright (c) The Regents of the University of California. All rights reserved.

group

- Mammal
- Vertebrate
- Deuterostome
- Insect
- Nematode
- Other

genome

Human

assembly

Feb. 2009 (GRCh37/hg19)

position

chr21:33,031,597-33,041,570

search term

enter position, gene symbol or search terms

submit

[Click here to reset](#) the browser user interface settings to their defaults.

[track search](#)[add custom tracks](#)[track hubs](#)[configure tracks and display](#)

Human Genome Browser – hg19 assembly ([sequences](#))

The February 2009 human reference sequence (GRCh37) was produced by the [Genome Reference Consortium](#). For more information about this assembly, see [GRCh37](#) in the NCBI Assembly database.

Sample position queries

A genome position can be specified by the accession number of a sequenced genomic clone, an mRNA or EST or STS marker, a chromosomal coordinate range, or keywords from the GenBank description of an mRNA. The following list shows examples of valid position queries for the human genome. See the [User's Guide](#) for more information.

Request:

Genome Browser Response:



UCSC Genome Browser (<http://genome.ucsc.edu/>)

[Home](#) [Genomes](#) [Genome Browser](#) [Tools](#) [Mirrors](#) [Downloads](#) [My Data](#) [About Us](#) [Help](#)

Human (*Homo sapiens*) Genome Browser Gateway

UCSC Genome Browser was created by the [Genome Bioinformatics Group of UC Santa Cruz](#).
Software Copyright (c) The Regents of the University of California. All rights reserved.

group

Mammal

Human

Chimp

Gorilla

Orangutan

Gibbon

Rhesus

Baboon

Squirrel monkey

Marmoset

Tarsier

Mouse lemur

Bushbaby

Tree shrew

Mouse

Rat

Kangaroo rat

Naked mole-rat

Guinea pig

Rabbit

Squirrel

Pika

Pig

Alpaca

Dolphin

Sheep

Cow

Horse

White rhinoceros

Cat

Ferret

assembly

Feb. 2009 (GRCh37/hg19)

position

chr21:33,031,597-33,041,570

search term

enter position, gene symbol or search terms

submit

[Click here to reset](#) the browser user interface settings to their defaults.

track search

add custom tracks

track hubs

configure tracks and display

Human Genome Browser

The February 2009 human genome assembly (GRCh37) was produced by the [Genome Reference Consortium](#). For more information about the GRCh37 in the NCBI Assembly database.

Sample position query

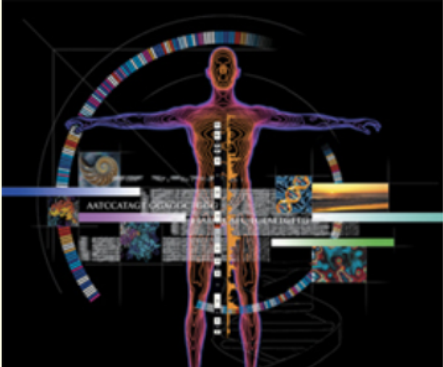
A genome position can be queried by accession number of a sequenced genomic clone, an mRNA or EST or STS or keywords from the GenBank description of an mRNA. The following list shows examples of valid queries for the human genome. See the [User's Guide](#) for more information.

Request:

chr7

Response:

Displays all of chromosome 7



UCSC Genome Browser

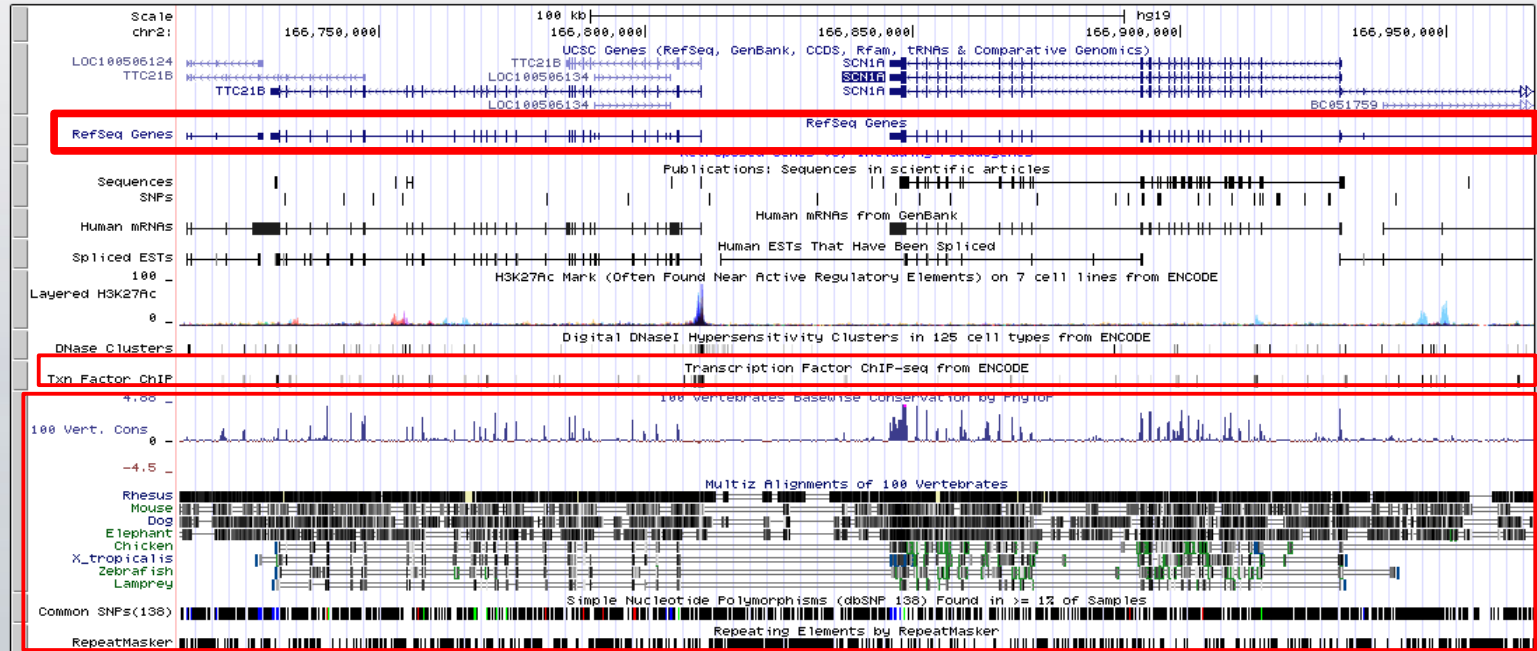
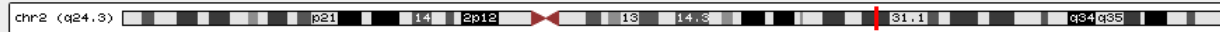
<http://genome.ucsc.edu/cgi-bin/hgTracks>

UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly

move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x

chr2:166,712,755-166,966,287 53,533 bp enter position, gene symbol or search terms

go



move start

Click on a feature for details. Click or drag in the base position track to zoom in. Click side bars for track

move end

ENCODE Project (<http://www.genome.gov/10005107>)



[Home](#) > [Research Funding](#) > [Research Funding Divisions](#) > [Division of Genome Sciences](#) > **ENCODE Project**

Division of Genome Sciences

- [Centers of Excellence in Genomic Science](#)
- [Division Staff](#)
- [ENCODE Project](#)**
- [Functional Analysis Program](#)
- [Genetic Variation Program](#)
- [Genome Informatics and Computational Biology Program](#)
- [Genome Technology Program](#)
- [NHGRI Genome Sequencing Program \(GSP\)](#)


The ENCODE Project: ENCyclopedia Of DNA Elements

[+ Share](#) [Print](#)



- [Overview](#)
- [Publications, Features and Press Releases](#)
- [Consortium Membership](#)
- [Data Release Policy](#)
- [Accessing ENCODE Data](#)
- [ENCODE Tutorials](#)
- [Common Cell Types](#)
- [Requests for Application \(RFAs\)](#)
- [Program Staff](#)

See Also:

[Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project](#) 
Nature, June 13, 2007

[Major Findings from The ENCODE Pilot Project](#)
June 2007

[The modENCODE Project](#)

[Grants Home](#)

On Other Sites:

[The ENCODE \(ENCyclopedia Of DNA Elements\) Project](#)

Follow the ENCODE Project on:

[Facebook](#)

[Twitter](#)

ENCODE data portal at UCSC

(<http://genome.ucsc.edu/encode/>)



Encyclopedia of DNA Elements

General

Resources & FAQ

Publications

Software Tools

Data Standards

Human

Downloads

Experiment Matrix

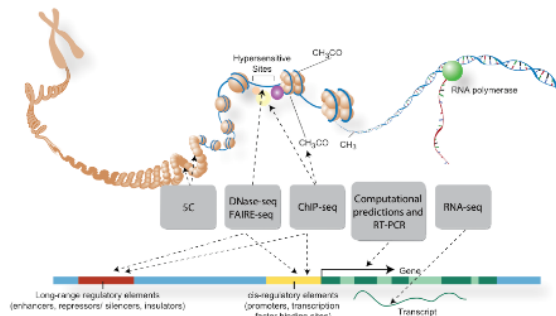
Search

Genome Browser (hg19)

Integrative Analysis


About ENCODE Data

The [Encyclopedia of DNA Elements](#) (ENCODE) Consortium is an international collaboration of research groups funded by the National Human Genome Research Institute ([NHGRI](#)). The goal of ENCODE is to build a comprehensive parts list of functional elements in the human genome, including elements that act at the protein and RNA levels, and regulatory elements that control cells and circumstances in which a gene is active.



[enlarge](#)

ENCODE data are now available for the entire human genome. **All ENCODE data are free and available for immediate use via :**

- [Search](#) for displayable tracks and downloadable files
- [Download](#) of data files
- [Visualization](#) in the UCSC Genome Browser (ENCODE data marked with the )
- [Data mining](#) with the UCSC Table Browser and other [UCSC Genome Bioinformatics tools](#)

[Click to](#)

To search for ENCODE data related to your area of interest and set up a browser view, use the UCSC [Experiment Matrix](#) or [Track Search tool](#) (Advanced features). The [Experiment List \(Human\)](#) and [Experiment List \(Mouse\)](#) links provide comprehensive listings of ENCODE data that is released or in preparation.

Neandertal Genome (<http://genome.ucsc.edu/Neandertal/>)

UCSC Genome Bioinformatics

[Home](#) - [Genomes](#) - [Blat](#) - [Tables](#) - [Gene Sorter](#) - [PCR](#) - [VisiGene](#) - [Session](#) - [FAQ](#) - [Help](#)

Genome
Browser

ENCODE

Blat

Table
Browser

Gene Sorter

In Silico
PCR

Genome
Graphs

Galaxy

VisiGene

Utilities

Downloads

Neandertal Genome Analysis Consortium Tracks at UCSC

Neandertals are the closest extinct relatives of humans. They lived from several hundred thousand years ago until their disappearance approximately 30,000 years ago. The Neandertal genome sequence (published by Green *et al.* in [Science](#) May 2010) consists of short sequence fragments, usually about 50 base pairs long, mapped to the human reference genome. The DNA was extracted largely from three Neandertal bones, each about 40,000 years old, from the Vindija Cave in Croatia: Vi33.16, Vi33.25, and Vi33.26. The bulk sequencing was carried out on the Illumina GAII platform. Neandertal DNA was identified from among the background of microbial sequences in the bone by similarity to the human or chimpanzee genomes.

This portal provides access to the sequence data and alignments to the reference human genome (NCBI Build 36/[hg18](#), GRCh37/[hg19](#)) as well as the reference chimpanzee genome (CGSC 2.1/[panTro2](#)) and several associated analyses (see [Downloads](#) and [References](#)).

The following annotations are available on the human hg18 and hg19 genome assemblies, except as noted:

- [Neandertal Alleles in Human/Chimp Coding Non-synonymous Differences in Human Lineage](#): Displays Neandertal alleles for human-chimp protein-coding differences on the human lineage using orangutan as the outgroup to determine which allele is more likely to be ancestral.
- [Selective Sweep Scan \(S\) on Neandertal vs. Human Polymorphisms](#): Shows the S score, an estimate



Artist's rendering of Neandertal man,
from
Neandertal museum in Mettmann,
Germany

Copyright: Johannes Krause, [Max Planck Institute
for Evolutionary Anthropology](#). All rights reserved.

<http://genome.ucsc.edu/cgi-bin/hgTracks>

Neandertal Assembly and Analysis

18 [H-C Coding Diffs](#)
hide ▼

18 [Sel Swp Scan \(S\)](#)
hide ▼

18 [5% Lowest S](#)
hide ▼

18 [S SNPs](#)
hide ▼

18 [Cand. Gene Flow](#)
hide ▼

18 [Neandertal Mito](#)
[No data-chr22]

Neandertal Seq
hide
dense
squish
pack
full

Denisova Assembly and Analysis

[Denisova Variants](#)
hide ▼

[Mod Hum Variants](#)
hide ▼

[Denisova Seq](#)
hide ▼

[Modern Derived](#)
hide ▼

Variation and Repeats

[Common SNPs\(138\)](#)
dense ▼

[Flagged SNPs\(138\)](#)
hide ▼

[Mult. SNPs\(138\)](#)
hide ▼

[All SNPs\(138\)](#)
hide ▼

[Common SNPs\(137\)](#)
hide ▼

[Flagged SNPs\(137\)](#)
hide ▼

[Mult. SNPs\(137\)](#)

[All SNPs\(137\)](#)

[Common SNPs\(135\)](#)

[Flagged SNPs\(135\)](#)

[Mult. SNPs\(135\)](#)

[All SNPs\(135\)](#)

UCSC BLAT (<http://genome.ucsc.edu/cgi-bin/hgBlat>)

Human BLAT Search

BLAT Search Genome

Genome:

Human

Assembly:

Feb. 2009 (GRCh37/hg19)

Query type:

BLAT's guess

Sort output:

query,score

Output type:

hyperlink

submit

I'm feeling lucky

clear

Paste in a query sequence to find its location in the the genome. Multiple sequences may be searched if separated by lines starting with '>' followed by the sequence name.


File Upload: Rather than pasting a sequence, you can choose to upload a text file containing the sequence.

Upload sequence: [选择文件](#) 未选择文件

[submit file](#)

Only DNA sequences of 25,000 or fewer bases and protein or translated sequence of 10000 or fewer letters will be processed. Up to 25 sequences can be submitted at the same time. The total limit for multiple sequence submissions is 50,000 bases or 25,000 letters.

UCSC *In-Silico* PCR (<http://genome.ucsc.edu/cgi-bin/hgPcr>)

 Genomes Genome Browser Tools Mirrors Downloads My Data About Us Help

UCSC In-Silico PCR

Genome:

Human

Assembly:

Feb. 2009 (GRCh37/hg19)

Target:

genome assembly

Forward Primer:

Reverse Primer:

submit

Max Product Size:

4000

Min Perfect Match:

15

Min Good Match:

15

Flip Reverse Primer: ☐

About In-Silico PCR

In-Silico PCR searches a sequence database with a pair of PCR primers, using an indexing strategy for fast performance.

Configuration Options

Genome and Assembly - The sequence database to search.

Target - If available, choose to query transcribed sequences.

Forward Primer - Must be at least 15 bases in length.

Reverse Primer - On the opposite strand from the forward primer. Minimum length of 15 bases.

Max Product Size - Maximum size of amplified region.

Min Perfect Match - Number of bases that match exactly on 3' end of primers. Minimum match size is 15.

Min Good Match - Number of bases on 3' end of primers where at least 2 out of 3 bases match.

Flip Reverse Primer - Invert the sequence order of the reverse primer and complement it.

Output

When successful, the search returns a sequence output file in fasta format containing all sequence in the database that lie between and include the primer pair. The fasta header describes the region in the database and the primers. The fasta body is capitalized in areas where the primer sequence matches the database sequence and in lower-case elsewhere. Here is an example from human:

```
>chr22:31000551+31001000 TAACAGATTGATGTCATGAAATGGG CCCATGAGTGGCTCCTAAAGCAAGCTGC  
TtACAGATTGATGTCATGAAATGGGzzzztggccagzzzztzzzztga
```

生物信息学：导论与方法

Bioinformatics: Introduction and Methods

Ge Gao 高歌 & Liping Wei 魏丽萍

Center for Bioinformatics, Peking University



<https://www.coursera.org/course/pkubioinfo>