

MOLECULAR METHODS AND BIOINFORMATICS

LM Evolutionary Biology, University of Padova
AA 2024/2025

Enrico Gaffo, Silvia Orsi,
Prof. Stefania Bortoluzzi

“Genome Browser”

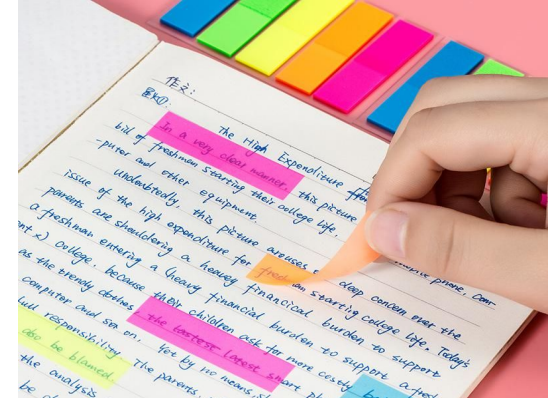
Padova, October 29, 2024

✉ Silvia Orsi: silvia.orsi@unipd.it

Key questions:

1. What does it mean *to annotate a genome*?
2. What is a Genome Browser?
3. How to use the UCSC Genome Browser

1. What does it mean to annotate a genome?



1. What does it mean to *annotate* a genome?



Genome Assembly



Genomic Features



Annotations

1. What does it mean to *annotate a genome*?

“Genome annotation is the process of finding and designating locations of individual genes and other features on raw DNA sequences, called **assemblies**. Annotation gives meaning to a given sequence and makes it much easier for researchers to view and analyze its contents. “

NIH, National Library of Medicine

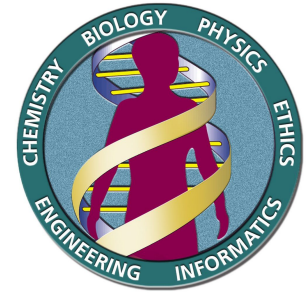
1. What does it mean to *annotate a genome*?

All annotations are based on a complex mix of gene expression data, computational predictions, and evidence from sequence similarity with other organisms (based on BLAST).

The annotation can be **structural** (assigning coordinates relative to a chromosome) or **functional** (what name should be given to a particular gene? What biological role does it have?)"

2. What is a Genome Browser?

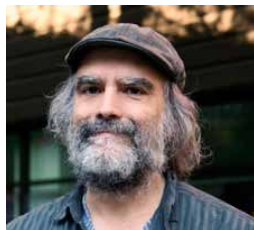
Let's start from the beginning!



1990–2003: The Human Genome Project

The Human Genome Project was a landmark global scientific effort whose signature goal was to generate **the first sequence of the human genome**. In 2003, the Human Genome Project produced a genome sequence that accounted for **over 90% of the human genome**. It was as close to complete as the technologies for sequencing DNA allowed at the time.

2. What is a Genome Browser?



*“Well, it has a lot of G, C, A and Ts”
Jim Kent*

Let's keep in mind that the assembled human genome is exactly 3,609,003,417 base pairs in size, corresponding to 20,418 protein-coding genes, 22,107 non-coding genes, and 15,195 pseudogenes for over 200,000 transcripts!

2. What is a Genome Browser?

Given the vast size of the genome, it becomes essential to create a comprehensive database that not only contains the genome sequence but also its detailed annotations, enabling to:

- explore chromosomal regions;
- explore regulatory regions flanking genes;
- perform genome-wide searches for elements;
- compare genome architecture across different organisms.

2. What is a Genome Browser?

Given the vast size of the genome, it becomes essential to create a comprehensive database that not only contains the genome sequence but also its detailed annotations, enabling to:

- explore chromosomal regions;
- explore regulatory regions flanking genes;
- perform genome-wide searches for elements;
- compare genome architecture across different organisms.



Let's create a Genome Browser!

2. What is a Genome Browser?

In bioinformatics, a genome browser is a graphical interface for displaying information from a biological database for genomic data.

Genome browsers enable users to visualize and browse entire genomes with annotated data. The software allows users to navigate the genome, view numerous features, analyze and investigate the relationships between various genomic elements.

- **BASIC** ANNOTATIONS: Genes, Transcripts, Pseudogenes, ncRNAs and so on...

2. What is a Genome Browser?

In bioinformatics, a genome browser is a graphical interface for displaying information from a biological database for genomic data.

Genome browsers enable users to visualize and browse entire genomes with annotated data. The software allows users to navigate the genome, view numerous features, analyze and investigate the relationships between various genomic elements.

- **BASIC** ANNOTATIONS: Genes, Transcripts, Pseudogenes, ncRNAs and so on...
- **ADVANCED** ANNOTATIONS:
 - Genetic variants (SNPs, indels...)
 - Repetitive sequences (STRs, LINE, SINE, DNA transposons)
 - Expression data (ESTs, RNA-seq...)
 - Alignments with homologous genomic regions in other species (tools for comparative genomics studies)"
 - and many others....

2. What is a Genome Browser?

GENOME BROWSER ACCESSIBILITY:

- ENSEMBL

<http://www.ensembl.org/index.html>

- UCSC genome Browser Gateway

<https://genome-euro.ucsc.edu/cgi-bin/hgGateway>

- NCBI Genome Data Viewer

<https://www.ncbi.nlm.nih.gov/genome/gdv/>

- Customized Genome browser

2. What is a Genome Browser?

UCSC Genome Browser

UNIVERSITY OF CALIFORNIA
SANTA CRUZ

Genomics
Institute

UCSC

Genome Browser

Genomes | Genome Browser | Tools | Mirrors | Downloads | My Data | Projects | Help | About Us

Try our new clinical tutorial!

Search genes, data, help docs and more...

Search

hg38

hg19

mm39

- **Genome Browser** - Interactively visualize genomic data
- **BLAT** - Rapidly align sequences to the genome
- **In-Silico PCR** - Rapidly align PCR primer pairs to the genome
- **Table Browser** - Download and filter data from the Genome Browser
- **LiftOver** - Convert genome coordinates between assemblies
- **REST API** - Returns data requested in JSON format
- **Variant Annotation Integrator** - Annotate genomic variants
- **More tools...**

News

Oct. 9, 2024 - **CADD v1.7** and **ClinGen CSPEC** for hg19 and hg38

Oct. 1, 2024 - **New clinical tutorial**

Sep. 30, 2024 - **New gnomAD v4.1 tracks** for hg38

Sep. 24, 2024 - **Nuclear mitochondrial DNA segments (NuMTs)** for hg38

Sep. 10, 2024 - **Two new public hubs: ENIGMA VCEP and BRCAExchange**

Aug. 29, 2024 - **Download track data in view**

More news...

Subscribe

Meetings and Workshops: Come see us in person!

- University of Pennsylvania School of Veterinary Science -- Philadelphia, PA. September 19, 2024
- **American Society of Human Genetics (ASHG)** -- Denver, CO. November 5, 2024

- **Variant Effect Prediction Training Course (VEPTC)** -- Palermo, Italy. October 14-16, 2024
- **Plant and Animal Genome Conference (PAG32)** -- San Diego, CA. January 10-15, 2025.

Feel free to **contact us** if you are interested in attending a workshop, or meeting someone from the team to collaborate, get help, or ask any questions at the meetings.

Sharing data

Learning

UCSC Geno...
Basics
Part 1: G...
Browser

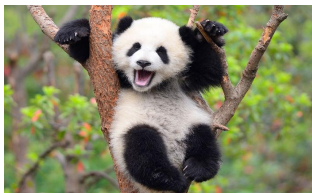
Introduction...
COVID...
latest i...
rom...

Saving and S...
Saving...
on the...
Sessions...
Browser

https://genome.ucsc.edu/cgi-bin/hgTracks?hg5_doOtherUser-submit&hg5_otherUser=mc.sierant&hg5_otherUserSessionName=RNU4-2_cons

2. What is a Genome Browser?

The UCSC Genome Browser was initially established to host the human genome, but it currently includes a collection of 46 genomes.



Evolutionary (primates, but not only), basic biology (Xenopus), economic (fish, chicken or sheep) or conservation (pandas) interest

2. What is a Genome Browser?



3. How to use the UCSC Genome Browser

UCSC Genome Browser on Human Dec. 2013 (GRCh38/hg38) Assembly

move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x 100x

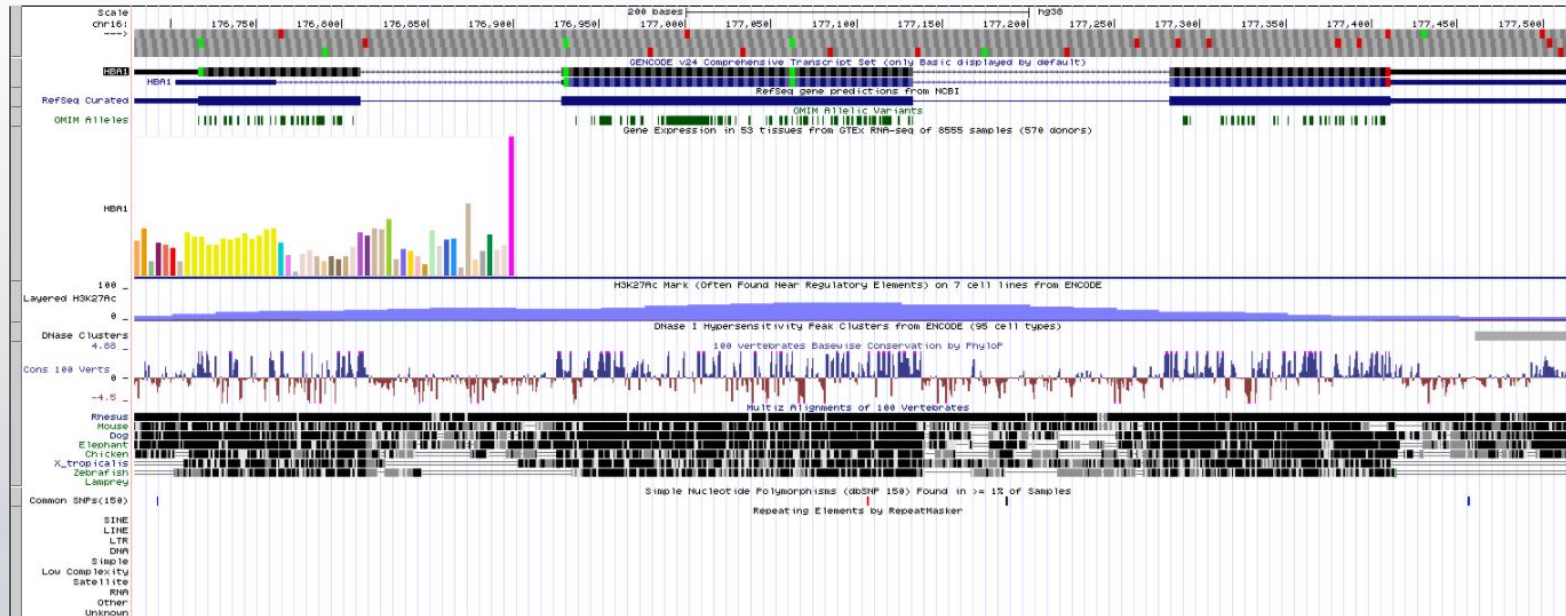
chr16:176,680-177,522 843 bp.

enter position, gene symbol, HGVS or search terms

go

ZOOM IN /OUT
BROWSER BAR

POSITION ON THE CHROMOSOME



mRNA (INTRONS/EXONS)

OMIM VARIANTS

GENE EXPRESSION TRACK

CONSERVATION TRACK

REPEATED ELEMENTS TRACK

Connect to:

http://compngen.bio.unipd.it/~stefania/Didattica/AA2024-2025/MMOL_BIOINFO_EB/MMOL_BIOINFO_EB.html

Program and materials of the course

PART OF BIOINFORMATICS

Teacher: Prof. STEFANIA BORTOLUZZI

Collaborators: Dr. Enrico Gaffo, Dr. Silvia Orsi

	LECTURES	MATERIALS
	Introduction to the course	Slides
	Database and data retrieval (biosequences, secondary databases, knowledgebases, molecular structures).	Slides
	Sequence alignment, scoring matrices, exact and heuristic methods. Similarity search and BLAST.	Slides Slides
	Multiple sequence alignment: from classic progressive methods to recent advances.	Slides
	Molecular structure prediction. Protein and RNA folding. Fold prediction approaches (ab initio, comparative modeling and threading). Neural networks.	Slides
	Genome sequencing and annotation. Genomic databases. Second and third generation methods for massive DNA sequencing (NGS). Genome sequencing (de novo) e resequencing. Introduction to NGS data analysis.	Slides
	RNA-seq: transcriptome sequencing and analysis in model and non-model species	Slides
	PRACTICAL SESSIONS	MATERIALS
	I Bioinformatics practical session	Guide
	<u>II Bioinformatics practical session</u>	Guide
	III Bioinformatics practical session	Guide
	IV Bioinformatics practical session	Guide
	V Bioinformatics practical session	Guide
	VI Bioinformatics practical session	Guide
	VII Bioinformatics practical session	Guide

Download
and unzip