

PB051 Výpočetní metody v bioinformatice a systémové biologii

Týden 1

Katedra informačních technologií
Masarykova Univerzita Brno

Jaro 2011

Tento projekt je spolufinancován Evropským sociálním fondem a státním rozpočtem České republiky.



Informace o kurzu

Genomové data prohlížeče

Informace o kurzu

Genomové data prohlížeče

- ▶ Dr. Matej Lexa, C506 (lexa@fi.muni.cz)
- ▶ Kurz: Út 10:00-11:50 (B116)
- ▶ Konzultace: Čt 13:00-15:00 (C506)
- ▶ <http://www.fi.muni.cz/~lexa/teaching.html>

- ▶ Hodnotí se
 - ▶ Úkoly 4 x 5 bodů
 - ▶ Cvičení 30 bodů
 - ▶ Zkouška 50 bodů
- ▶ Klasifikační stupnice
 - ▶ A 91 - 100
 - ▶ B 81 - 90
 - ▶ C 71 - 80
 - ▶ D 61 - 70
 - ▶ E 51 - 60
 - ▶ F méně než 51

- ▶ Analýza dat v genomovém kontextu
- ▶ Aplikace Markovovských modelů v bioinformatice
- ▶ Statická analýza sítí
- ▶ Dynamická analýza sítí

- ▶ Prohlížeče genomů (UCSC, Ensembl, Argo, GBrowse)
- ▶ Základní funkce UCSC Genome Browser
- ▶ Pokročilé funkce UCSC Genome Browser
- ▶ Programovatelný přístup ke genomu přes rozhraní Biomart

Aplikace Markovovských modelů v bioinformatice

- ▶ Markovovy řetězce
- ▶ Markovovy řetězce proměnného řádu
- ▶ Skryté Markovovské modely
- ▶ HMM profily
- ▶ HMM pro identifikaci genů

- ▶ metody a nástroje statické analýzy
 - ▶ integrace dat
 - ▶ rekonstrukce sítě genových interakcí z experimentálních dat
 - ▶ analýza interakční sítě jako obecného grafu
- ▶ metody a nástroje dynamické analýzy
 - ▶ deterministické simulační techniky
 - ▶ stochastické simulační techniky
 - ▶ metody odhadování parametrů

- ▶ integrace dat
 - ▶ nástroje: Genome-Matrix, Gaggle
- ▶ rekonstrukce sítí
 - ▶ nástroje: Genomica, GeneNetworks
- ▶ analýza sítí
 - ▶ nástroje: Cytoscape, VisAnt

- ▶ deterministické simulace
 - ▶ nástroje: Dizzy, COPASI
- ▶ stochastické simulace
 - ▶ nástroje: Dizzy, STOCKS
- ▶ odhadování parametrů
 - ▶ nástroje: COPASI, BioCHAM

Informace o kurzu

Genomové data prohlížeče

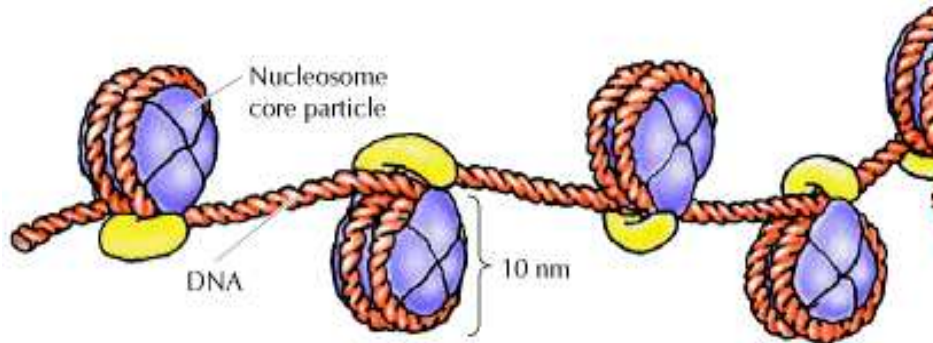
Informace o kurzu

Genomové data prohlížeče



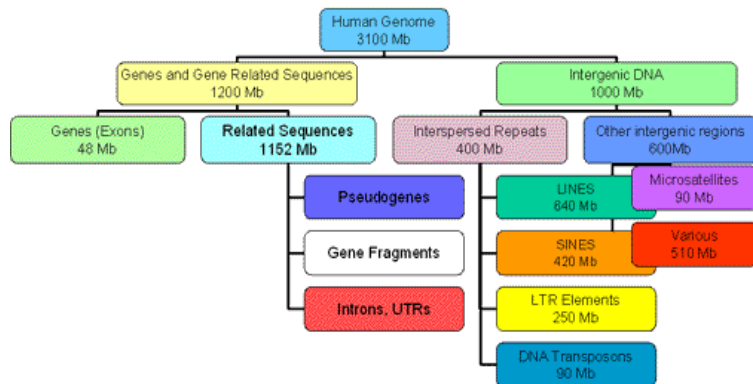
Informace o kurzu

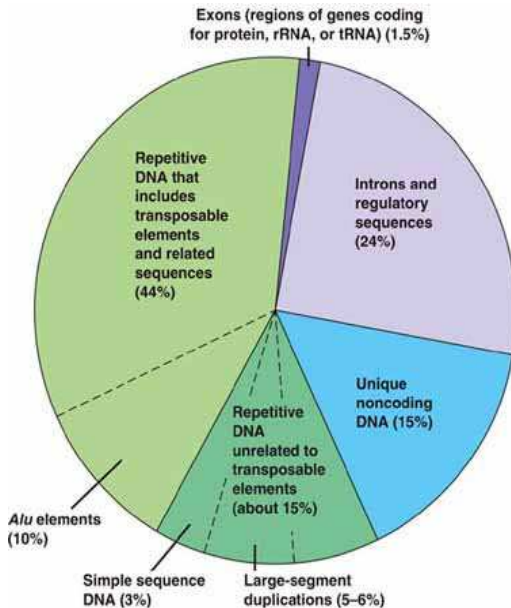
Genomové data prohlížeče



10-nm fiber

30-nm





- ▶ organizace (konfokální a elektronová mikroskopie)
- ▶ sekvenace
- ▶ mapování metylace, nukleozomů
- ▶ měření transkripce (RNA-Seq, DNA čipy)
- ▶ identifikace regulačních sekvencí (Chip-Seq)
- ▶ funkce genů - podrobný výzkum

- ▶ predikce genů
- ▶ homologie (zjišťování podobnosti sekvencí)

- ▶ Geny kódující
 - ▶ proteiny (exon, intron)
 - ▶ RNA
- ▶ Regulační sekvence
 - ▶ promotory
 - ▶ enhancery
 - ▶ jiné
- ▶ Repetitivní sekvence
- ▶ Cizí sekvence
 - ▶ viry
 - ▶ transpozony
- ▶ Oblasti podobnosti (homology)
 - ▶ SNP
 - ▶ delší strukturní variace

UCSC Genome Browser

Informace o kurzu

Genomové data prohlížeče

Human chr5: 70,256,524-70,284,592 - UCSC Genome Browser v134 - Konqueror

Location Edit View Bookmarks Tools Settings Help

id=73350821&knownGene=full

Human chr5:70,256,524-70,28...

Home Genomes Blat Tables Gene Sorter PCR DNA Convert PDF/PS Help

UCSC Genome Browser on Human Mar. 2006 Assembly

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

10x

position/search chr5:70,256,524-70,284,592 jump clear size 28,069 bp. configure

chr5 (q13.2)

chr5: 70260000 70265000 70270000 70275000 70280000

STS Markers STS Markers on Genetic (blue) and Radiation Hybrid (black) Maps

UCSC Known Genes Based on UniProt, RefSeq, and GenBank mRNA

RefSeq Genes

Human mRNAs From GenBank

Human ESTs That Have Been Spliced

Vertebrate Multiz Alignment & Conservation (17 Species)

Conservation

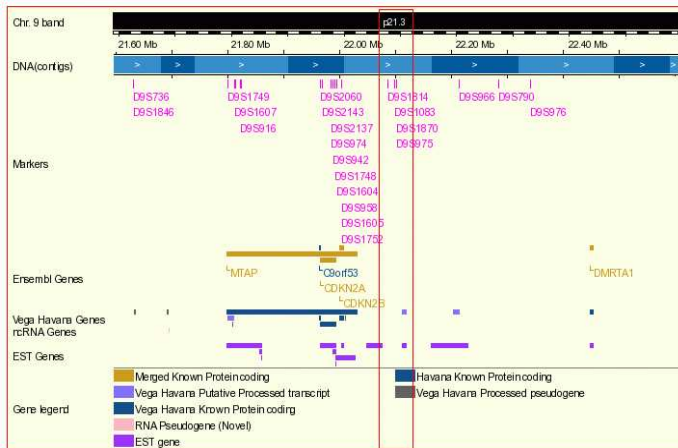
mouse
rat
rabbit
dog
armadillo
elephant
opossum
chicken

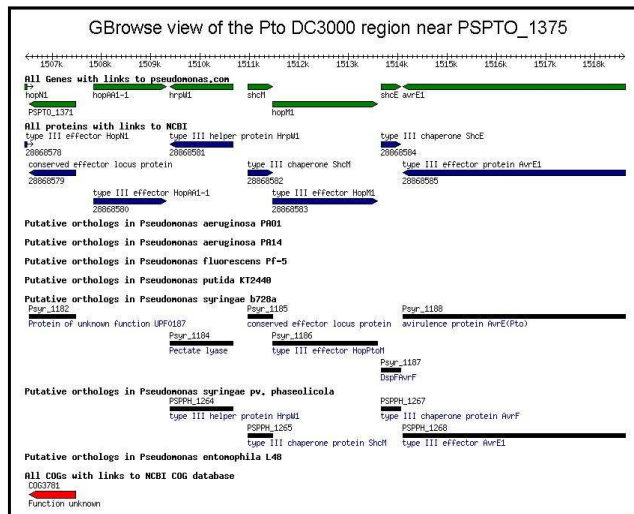
<http://genome.ucsc.edu/cgi-bin/hgc?hgsid=733...523&r=70284592&db=hg18&pix=620>

Ensembl Genome Browser

Informace o kurzu

Genomové data prohlížeče





Informace o kurzu

Genomové data prohlížeče

The screenshot displays the Argo genome browser interface. At the top, there is a menu bar with options: Argoc, File, Track, Edit, Select, View, Zoom, Rulers, Analyze, User, Bookmarks, Window, Help. The main window shows a genomic track for Human Chr15, coordinates 1399746:1-100000. The track includes various features such as genes, transcripts, and repeats. A 'PolyA Signal View: Novel Transcrip...' window is open, showing a list of polyA signals:

Signal	Score	Start	End	Start	End
AATACA	(1.28)	415-420	8208-82613		
AATACA	(1.28)	470-475	86534-86539		
A77AAA	(14.91)	804-809	89548-89553		

Below the PolyA signal view, there is an 'Inspector' panel with tabs for Properties, DNA, mRNA, and Protein. The DNA tab is selected, showing a sequence of nucleotides. A 'Nucleotide C 127/80162' is highlighted. To the right, there is a 'Finder' panel with search criteria:

Select Features whose:

- Label: contains repeats
- Protein length: > 50
- mRNA Sequence: contains gataca

A 'Search' button is located at the bottom of the Finder panel.

Informace o kurzu

Genomové data prohlížeče



Golden Helix Genome Browser

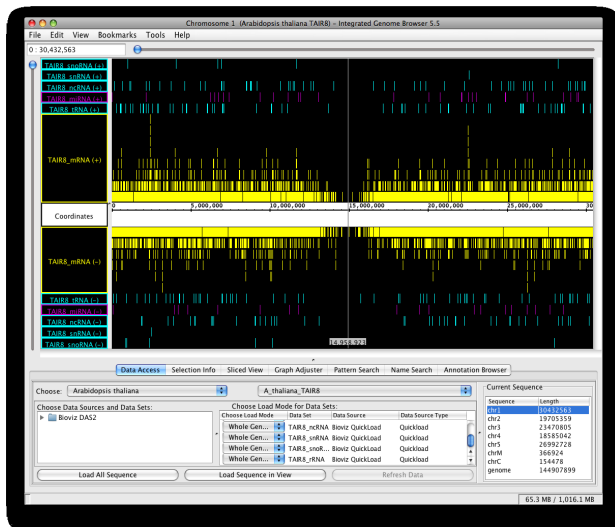
Informace o kurzu

Genomové data prohlížeče



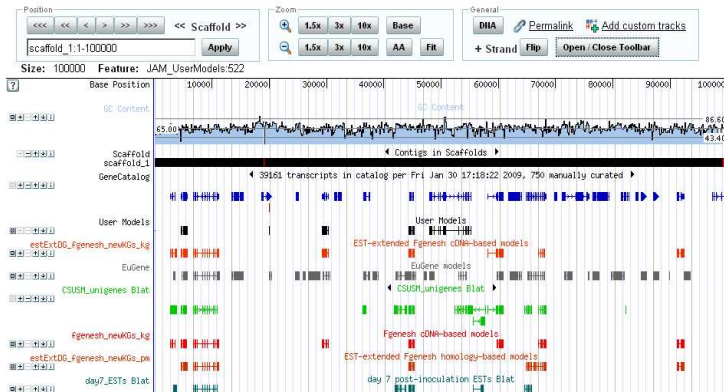
Informace o kurzu

Genomové data prohlížeče



Informace o kurzu

Genomové data prohlížeče



RIKEN Genome Browser

The screenshot displays the RIKEN Genome Browser interface for the *Mus musculus* genome. The interface is divided into several numbered sections:

- 1**: Navigation and search options, including a search bar and a "Register current interval" button.
- 2**: A sidebar menu for database selection, listing various databases such as Mouse miRBase, FANTOM3, CAGE, and RefSeq Peptide.
- 3**: A genomic map showing the current interval (82,100,466 bp to 82,182,103 bp) with a specific marker (81,637 bp) highlighted.
- 4**: A detailed view of the genomic region, showing the marker symbol (IRS1_MOUSE) and the RefSeq peptide (NP_034700.2) and RefSeq DNA (NM_010570.2) sequences.
- 5**: A series of tracks showing various genomic features, including FANTOM3 tracks (DNA, CAGE), Ensembl transcripts (43.36a), and RefSeq tracks (Peptide, DNA, UniProtKB).
- 6**: An "Expert's set" sidebar menu, listing various genomic features such as Mouse, Human Homology, C elegans Homology, and RefSeq Peptide.

