

PB051 Výpočetní metody v bioinformatice a systémové biologii

Týden 1

Katedra informačních technologií
Masarykova Univerzita Brno

Podzim 2010

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 8:00-9:50 (B130)
- ▶ Konzultace: Čt 13:00-15:00 (C506)
- ▶ <http://www.fi.muni.cz/~lexa/teaching.html>

- ▶ Hodnotí se
 - ▶ Cvičení 2 x 25 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

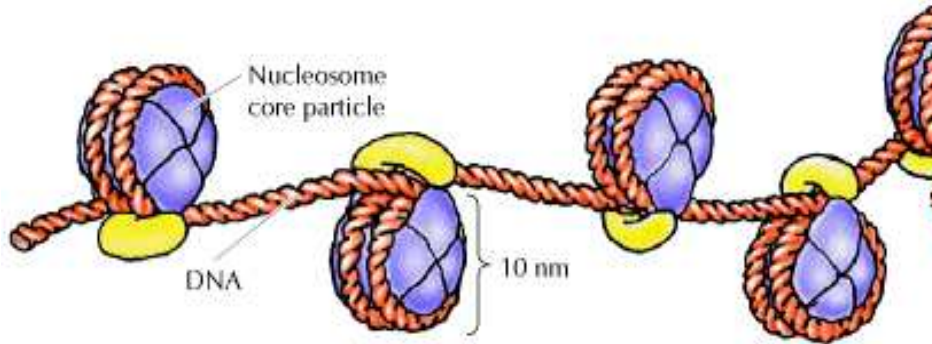
Informace o kurzu

Genomové data prohlížeče



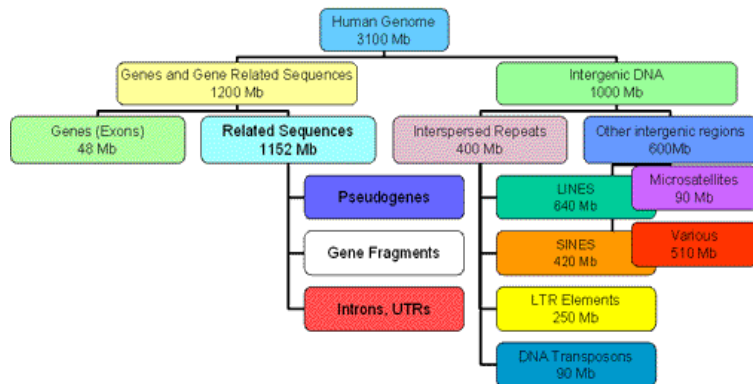
Informace o kurzu

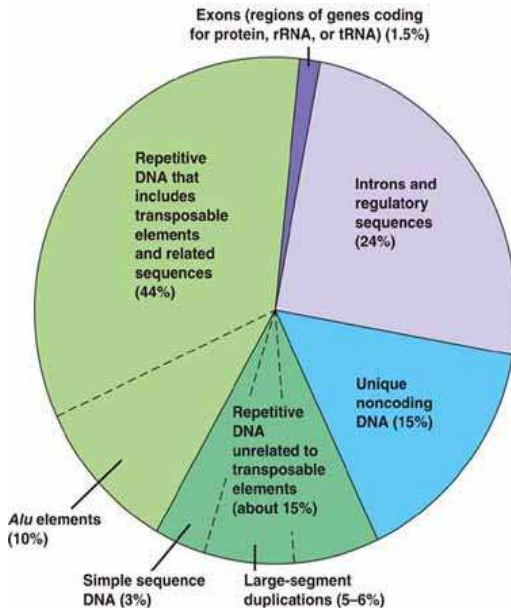
Genomové data prohlížeče



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

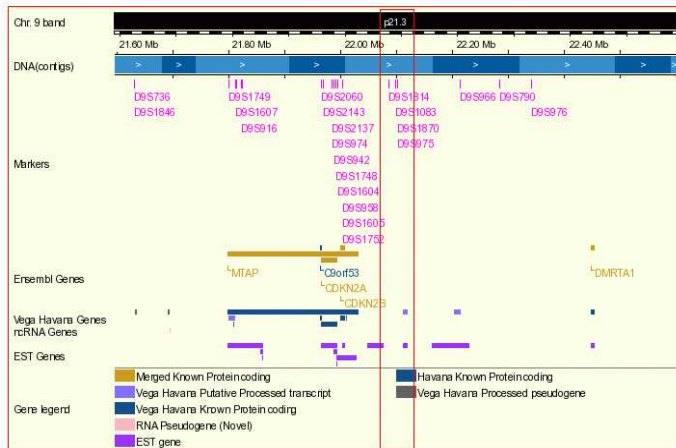
mouse
rat
rabbit
dog
armadillo
elephant
opossum
chicken

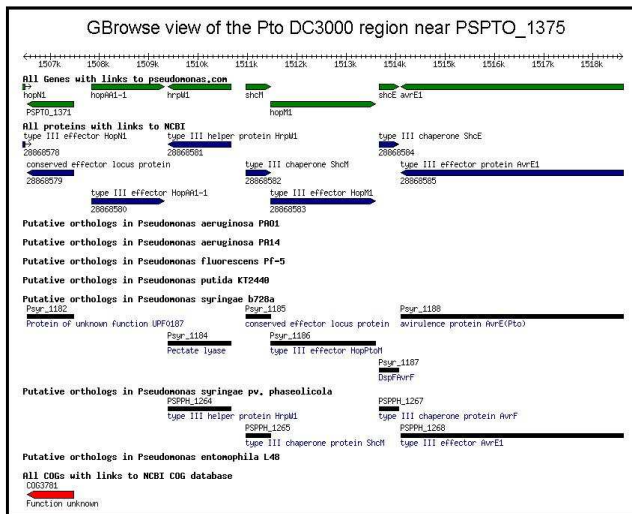
<http://genome.ucsc.edu/cgi-bin/hgc?hgid=733...523&r=70284592&db=hg18&pix=620>

Ensembl Genome Browser

Informace o kurzu

Genomové data prohlížeče





Informace o kurzu

Genomové data prohlížeče

The screenshot displays the Argo genome browser interface. The main window shows a genomic track for Human Chr15, contig 1.1 (1,399,746:1-100,000). 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 with their coordinates and scores:

Signal	Score	Start	End
AATACA	(1.28)	415-420	82808-82813
AATACA	(1.28)	470-475	86534-86539
AATAAA	(14.91)	804-809	89548-89553

The 'Inspector' panel at the bottom left shows the 'Properties' tab with the following sequence:

```

AGTAATTATAGA...TGGACAGATGAA...AAAATAGCAATGTCATAAGGACGTCCTACACCTGGGCTGGG
CAATGGCCCTGACGAGAGTGTATACCTCTCTGGTATGATGAGAGTGTGACGCTGTACATCTTGAAGTGTGAAGAG
CAACCTCTG...AAGCGTGGCAATGCCAGAGGGAGGTTGTGCAATATTCCTCATGATCTTGCTGTATCCAAAT
ATTGTAGATGTATGACCAACACAGCTGTCCATTATGCTCTTTATGGT...Nucleotide C 127/80162...AATGCT
GTCTTGTGGTCAACATCGAAGTGTGAAGAACAGGCTGGCCACACACCC...GAGAGTGTGAGAGGTTGAGAGT
MCKAAATGTGGAAATTTTACAGAGAAATACAAATGCAATACAGCTAGTGTATATCTTAAATCATTTCATCAACAA
CTTTTGG...AATACA...AACAAAGATATCTAAAAATCTCAAAATAGCAATCCACAGAGGAGATCTGAGGAGACCTG
CGAGCTGCACCTTGGTGGAAAGACACTGAC...GCTGAAGCTTGGTGGAAAGACACTGACGA...GATA
CAGTGAATCTCTTCAAGAAATTTAGCCTGTAAAAAATCTTTAAAAATTCAGAGAGGGGTTTGAATACAGTGAATCT

```

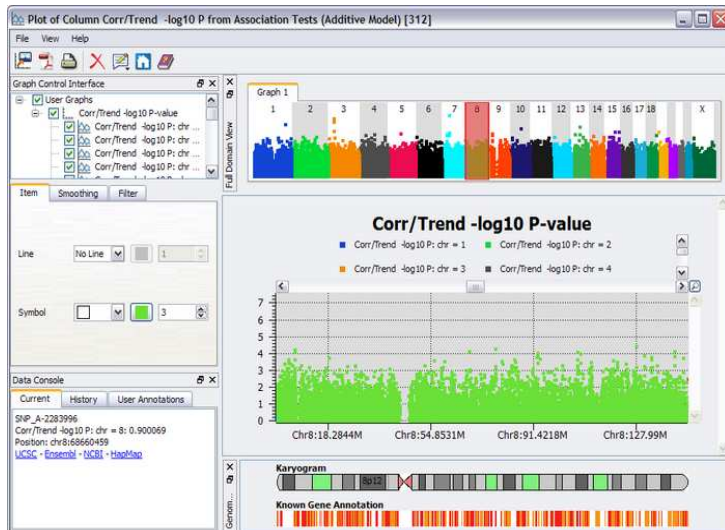
The 'Finder' panel at the bottom right allows for searching features based on labels, protein length, and mRNA sequence. The search criteria are:

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

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

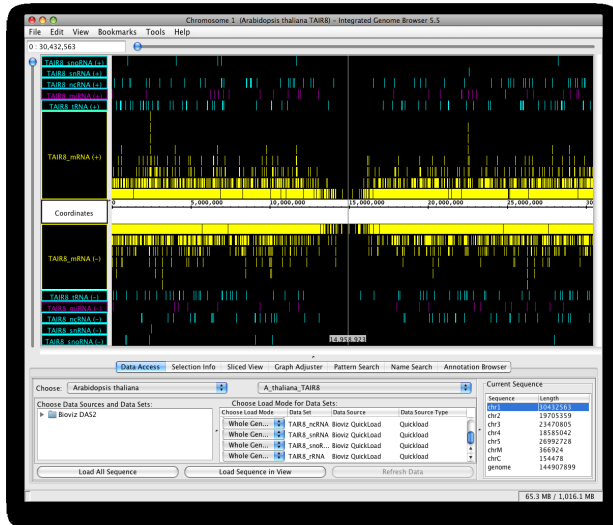


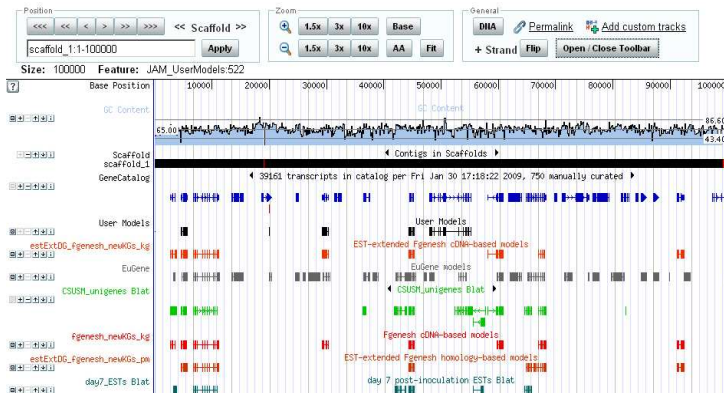
Golden Helix Genome Browser



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 annotated with red circles 1 through 6:

- 1**: Database list on the left side, including Mouse miRBase version 5, Mouse FANTOM3, Mouse Fantom3 Mega Ge, Mouse CAGE, Mouse Mutant Resources, Mouse Ensembl Transp, Mouse Ensembl VegaGen, Mouse GTOP, Mouse RefSeq Peptide, Mouse RefSeq DNA, Mouse UniProtKB SwissP, Mouse UniProtKB TrEMBL, Mouse Riken Transcription, Mouse dbSNP NCBI, Mouse Ensembl Gene 40, Mouse Ensembl VegaGene, Mouse Ensembl Gene NCBI, Mouse MGI Gene, and Mouse RefSeq.
- 2**: Expert's set on the left side, including Mouse, Human Homology, Catagans Homology, Other Homology, All, dbSNP, Medline, PostMed (Positional Medline), Transcriptome, FANTOM, and CAGE.
- 3**: Search and interval registration area, including "Go to Search page", "Register current interval", and a filter by keyword.
- 4**: Gene track for *Mus musculus*: *Irf1*, showing the gene structure with exons and introns, and coordinates (82,100,466 bp to 82,182,103 bp).
- 5**: List of tracks including Mouse FANTOM, Mouse CAGE (all tissues) Expression (TAG=249, TPM=34.62), Mouse Ensembl Transcript 45.36a, Mouse Ensembl VegaGene Transcript 45.36a, Mouse GTOP, Mouse RefSeq Peptide, Mouse RefSeq DNA, and Mouse UniProtKB SwissProt.
- 6**: Detailed view of the *Irf1* gene structure and associated data, including the gene symbol, marker symbols, and coordinates.

Informace o kurzu

Genomové data prohlížeče

