Mining Genomic Sequence Data

Tyra G. Wolfsberg, Ph.D.
NHGRI

Current Topics in Genome Analysis
September 16, 2003

Accessing the public genome sequence data

UCSC's Genome Browser ("Golden Path")
http://genome.ucsc.edu

NCBI's Map Viewer
http://www.ncbi.nlm.nih.gov/mapview/

Ensembl
http://www.ensembl.org

Types of data integrated in genome browsers

- Genomic sequence
- RefSeq mRNAs (non-redundant)
- GenBank mRNAs (redundant)
- ESTs
- Gene predictions
- SNPs
- Homologous sequences from other organisms
- STSs

NCBI Reference Sequences

Accession Format

RefSeq accession numbers can be distinguished from GenBank accessions by their distinct format of [2 characters|underbar|6 digits]. For example, a RefSeq protein accession is NP_015325.

Accession	Molecule	Method @	Note
NC_123456	Genomic	Curation	Complete genomic molecules including genomes, chromosomes, organelles, plasmids.
NG_123456	Genomic	Curation	Incomplete genomic region; primarily supplied for *Homo sapiens* and *Mus musculus* to support the NCBI Genome Annotation pipeline.
NM_123456	mRNA	Curation	
NR_123456	RNA	Curation	Non-coding transcripts including structural RNAs, transcribed pseudogenes, and others
NP_123456	Protein	Curation	
NT_123456	Genomic	Automated	Intermediate genomic assemblies of BAC sequence data
NW_123456	Genomic	Automated	Intermediate genomic assemblies of Whole Genome Shotgun sequence data
XM_123456	mRNA	Automated	*Homo sapiens* model mRNA provided by the Genome Annotation process; sequence corresponds to the genomic contig.
XR_123456	RNA	Automated	*Homo sapiens* model non-coding transcripts provided by the Genome Annotation process; sequence corresponds to the genomic contig.
XP_123456	Protein	Automated	*Homo sapiens* model proteins provided by the Genome Annotation process; sequence corresponds to the genomic contig.
NZ_ABCD12345678	Genomic	Automated	An ordered collection of whole genome shotgun sequence data, for incomplete bacterial genomes.Accessions are not tracked between releases.The first four characters following the underscore (e.g. 'ABCD') identifies a genome project.
ZP_12345678	Genomic	Automated	Proteins, annotated on NZ_ accessions.

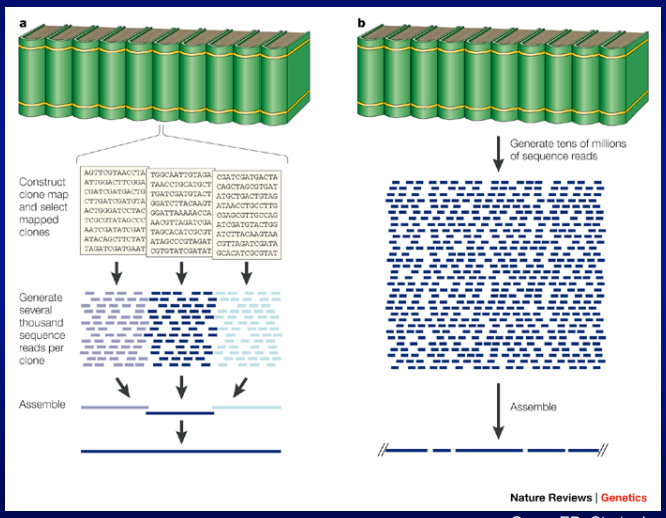
@ Method:
Curated: indicates the process flow includes expert review for some of the records; analysis may be provided either by NCBI staff or collaborators.
Automated: indicates records that are not individually reviewed; updates are released in bulk for a genome.

http://www.ncbi.nlm.nih.gov/RefSeq/key.html

Overview of sequencing strategies

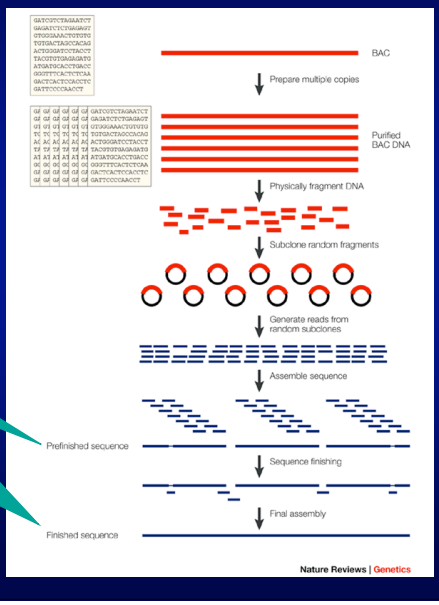
Clone-by-clone shotgun sequencing

Whole-genome shotgun sequencing



Green ED. Strategies for the systematic sequencing of complex genomes. Nat Rev Genet. 2001. 2:573-83.

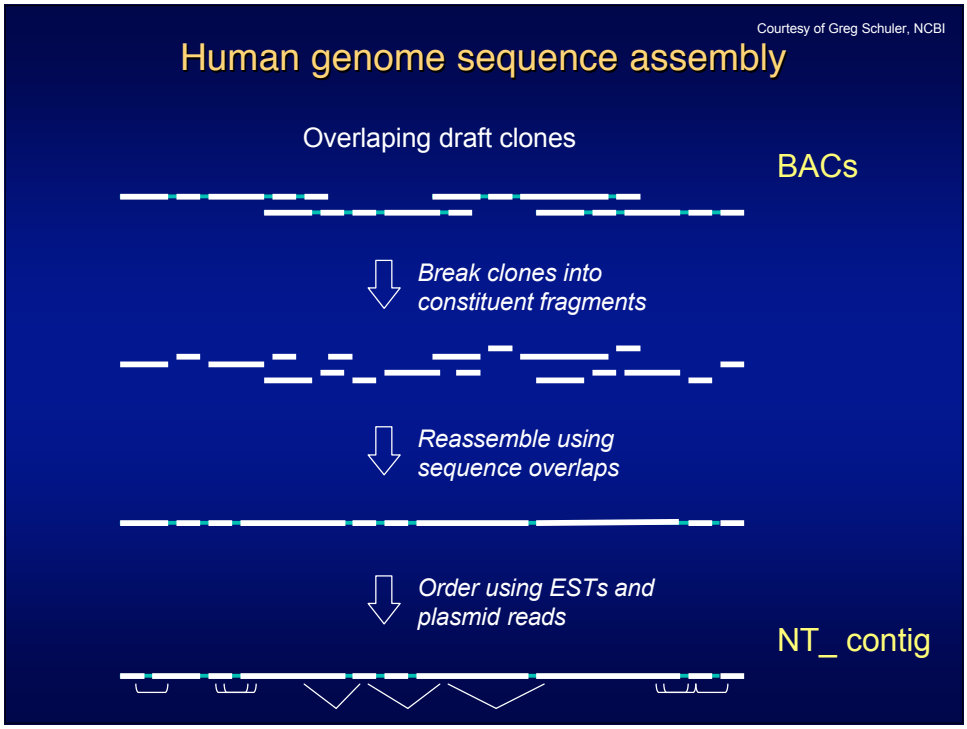
Clone-by-clone shotgun sequencing



"working draft" Phase 0,1, or 2 BAC

Finished Phase 3 BAC

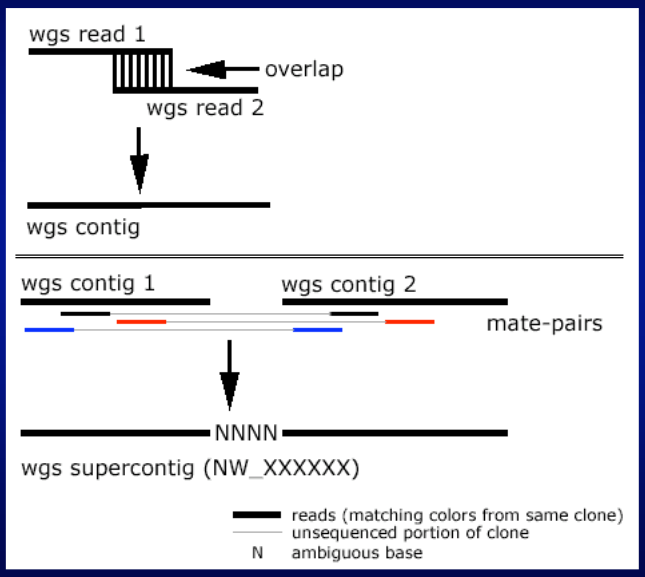
Green ED. Strategies for the systematic sequencing of complex genomes. Nat Rev Genet. 2001. 2:573-83.



Status of the human genome sequence

- All chromosomes are now considered finished
- Build 33; April 2003
  - <400 gaps, averaging <100 Kb, representing DNA regions with unusual structures that can't be reliably sequenced
    - 138 unplaced contigs each with sequence from a single clone
    - Assembly will be updated as gaps are closed
- Build 34; July 2003
  - 11 Mb (~0.4%) more finished nucleotides than build 33
  - Covers ~99% of gene-containing regions in the genome
- NCBI and Ensembl currently display build 33; UCSC features a partially annotated build 34, as well as older assemblies
- UCSC is usually the first to display new assemblies, followed by NCBI and then Ensembl.

Mouse and rat whole genome shotgun (WGS) assemblies



http://www.ncbi.nlm.nih.gov/genome/seq/NCBIContigInfo.html

Mouse genome sequencing

- Whole genome shotgun sequence (WGS) is now completed (7x coverage)
- "MGSC Version 3" is the current assembly of the WGS
- Sequence will be finished by sequencing individual BACs and incorporating WGS
- NCBI, UCSC, and Ensembl provide browsers based on an assembly that combines MGSCv3 with finished BAC sequence (called build 30 at NCBI and Ensembl, Feb 2003 at UCSC)

Rat genome sequencing

- Draft genome assembly produced by the Rat Genome Sequencing Consortium
- Hybrid approach combined clone by clone and whole genome shotgun methods
- Assembly covers more than 90% of the genome
- UCSC displays v. 3.1 (June 2003); not clear what assembly is shown by NCBI, or whether Ensembl shows v. 2.0 or 2.1

Where's the genome sequence at NCBI?

Status	BLAST database	Definition
Phase 0 HTG	htgs	single-few pass reads of a single clone (not contigs)
Phase 1 HTG	htgs	Unfinished, may be unordered, unoriented contigs, with gaps
Phase 2 HTG	htgs	Unfinished, ordered, oriented contigs, with or without gaps
Phase 3 HTG	nr	Finished, no gaps (with or without annotations)
NT_ contigs	human genome BLAST -->genome	assembled BACs from HTG phase 1-3
WGS	trace archive; mouse or rat genome BLAST -->WGS Traces	whole genome shotgun sequences
NW_contigs	rat genome BLAST -->genome	assembled WGS sequence (also called supercontigs or scaffolds)
WGS/HTG assembly	mouse genome BLAST -->genome	assembled WGS sequence, along with finished HTGs

Accessing the public genome sequence data

UCSC's Genome Browser ("Golden Path")
http://genome.ucsc.edu

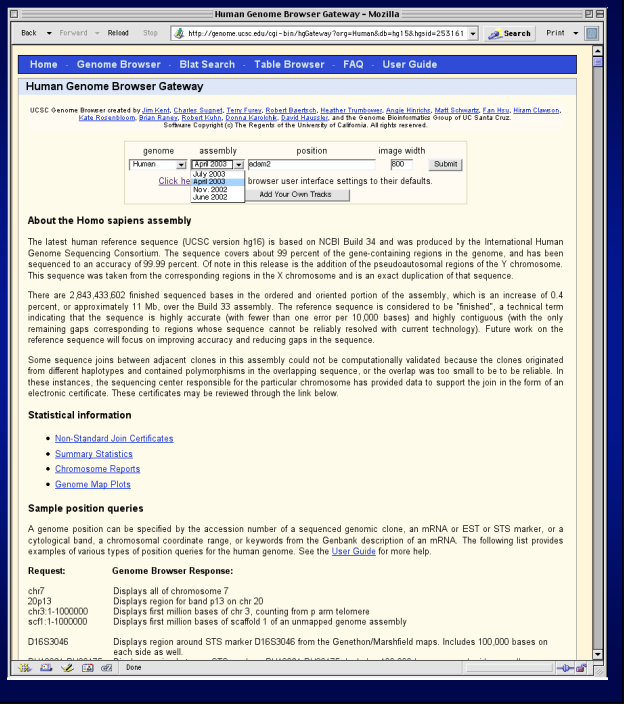
NCBI's Map Viewer
http://www.ncbi.nlm.nih.gov/mapview/

Ensembl
http://www.ensembl.org

UCSC

Genomes
- Human
- Mouse
- Rat
- C. elegans
- C. brigssae
- SARS
- Zoo

Human Assemblies
- Jun 2002: NCBI build 30
- Nov 2002: NCBI build 31
- Apr 2003: NCBI build 33
- July 3002: NCBI build 34



UCSC: Results for ADAM2 query

Known Genes

ADAM20 at chr14:68979118-68981543 - (AF029899) a disintegrin and metalloproteinase domain 20
ADAM21 at chr14:68914800-68916644 - (AF029900) a disintegrin and metalloproteinase domain 21
ADAM22 at chr7:87161961-87424669 - (AF073291) a disintegrin and metalloproteinase domain 22
ADAM29 at chr4:176436583-176483706 - (AF134708) a disintegrin and metalloproteinase domain 29
ADAM23 at chr2:207272561-207446902 - (AJ005580) a disintegrin and metalloproteinase domain 23
ADAM27 at chr8:39183088-39328434 - (AJ133004) a disintegrin and metalloproteinase domain 18
ADAM28 at chr8:23972184-24033174 - (AJ242015) a disintegrin and metalloproteinase domain 28
ADAM2 at chr8:39342200-39436675 - (BC034957) a disintegrin and metalloproteinase domain 2 (fertilin beta)

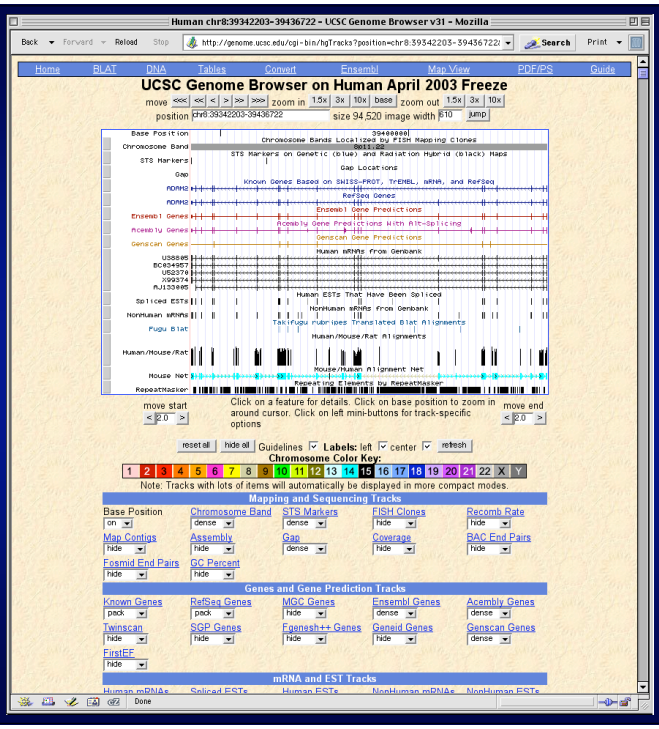
RefSeq Genes

ADAM29 at chr4:176481126-176706678 - (NM_021780) a disintegrin and metalloproteinase domain 29
ADAM29 at chr4:176481126-176483589 - (NM_021779) a disintegrin and metalloproteinase domain 29
ADAM28 at chr8:23972192-24032878 - (NM_021778) a disintegrin and metalloproteinase domain 28
ADAM28 at chr8:23972192-24014179 - (NM_021777) a disintegrin and metalloproteinase domain 28
ADAM22 at chr7:87161928-87424674 - (NM_021723) a disintegrin and metalloproteinase domain 22
ADAM22 at chr7:87161928-87424674 - (NM_021722) a disintegrin and metalloproteinase domain 22
ADAM22 at chr7:87161928-87409655 - (NM_021721) a disintegrin and metalloproteinase domain 22
ADAM20 at chr14:68979115-68981693 - (NM_003814) a disintegrin and metalloproteinase domain 20
ADAM2 at chr8:39342203-39436722 - (NM_001464) a disintegrin and metalloproteinase domain 2
ADAM29 at chr4:176436583-176483706 - (NM_014269) a disintegrin and metalloproteinase domain 29
ADAM22 at chr7:87161928-87424674 - (NM_016351) a disintegrin and metalloproteinase domain 22
ADAM28 at chr8:23972184-24033171 - (NM_014265) a disintegrin and metalloproteinase domain 28
ADAM22 at chr7:87161928-87409655 - (NM_004194) a disintegrin and metalloproteinase domain 22
ADAM21 at chr14:68914257-68916663 - (NM_003813) a disintegrin and metalloproteinase domain 21
ADAM23 at chr2:207272390-207446700 - (NM_003812) a disintegrin and metalloproteinase domain 23

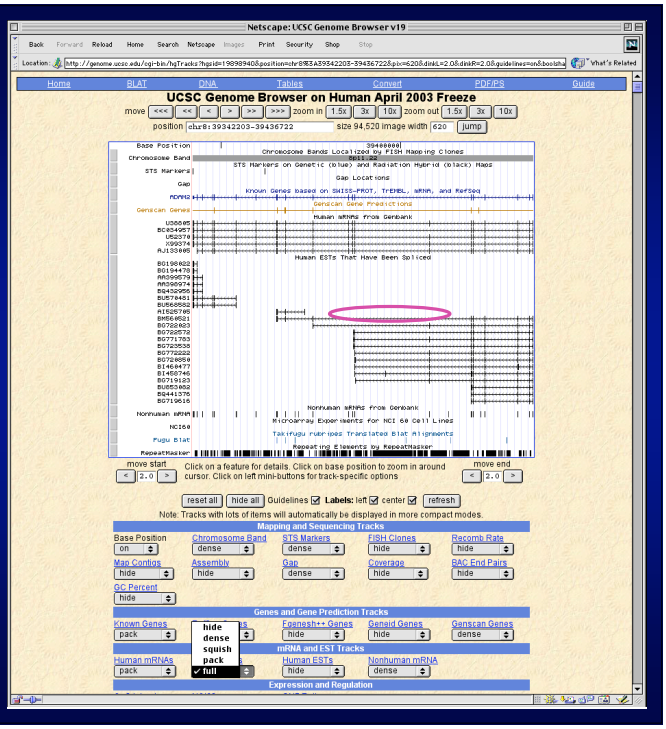
Human Aligned mRNA Search Results

AJ005580 - Homo sapiens mRNA for adam23 protein.
AF158637 - Homo sapiens metalloproteinase-disintegrin ADAM22-3 (ADAM22) mRNA, alternatively spliced, parti
AF134708 - Homo sapiens disintegrin and metalloproteinase domain 29 (ADAM29) mRNA, complete cds.
AF171929 - Homo sapiens metallaproteinase-disintegrin (ADAM29) mRNA, complete cds.
AF171930 - Homo sapiens metallaproteinase-disintegrin beta (ADAM29) mRNA, alternatively spliced, complete
AF171931 - Homo sapiens metallaproteinase-disintegrin gama (ADAM29) mRNA, alternatively spliced, complete
AF155381 - Homo sapiens metalloprotease-like, disintegrin-like, cysteine-rich protein 2 delta (ADAM22) mRN
AF155382 - Homo sapiens metalloprotease-like, disintegrin-like, cysteine-rich protein 2 epsilon (ADAM22) m

UCSC: Default view of ADAM2



UCSC:
Is ADAM2
alternatively
spliced?



UCSC: Skipped exons deleted smoothly

Alt. spliced EST (BM560521)

```
00000389 gggcgtactacagtttgaaaatgttagttatggaatagaacccctggagt 00000438
>>>>>>>> |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<<
39419636 gggcgtactacagtttgaaaatgttagttatggaatagaacccctggagt 39419587

00000439 cttcagttggctttgaacatgtaatttaccaagtaaaacataagaaagca 00000488
>>>>>>>> |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<<
39419586 cttcagttggctttgaacatgtaatttaccaagtaaaacataagaaagca 39419537

00000489 gatgtttccttatataatgagaaggatattgaatcaagagatctgtcctt 00000538
>>>>>>>> |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<<
39419536 gatgtttccttatataatgagaaggatattgaatcaagagatctgtcctt 39419487

00000539 taaattacaaagcgtagag 00000557
>>>>>>>> |||||||||||||||||| <<<<<<<<
39419486 taaattacaaagcgtagag 39419468
```

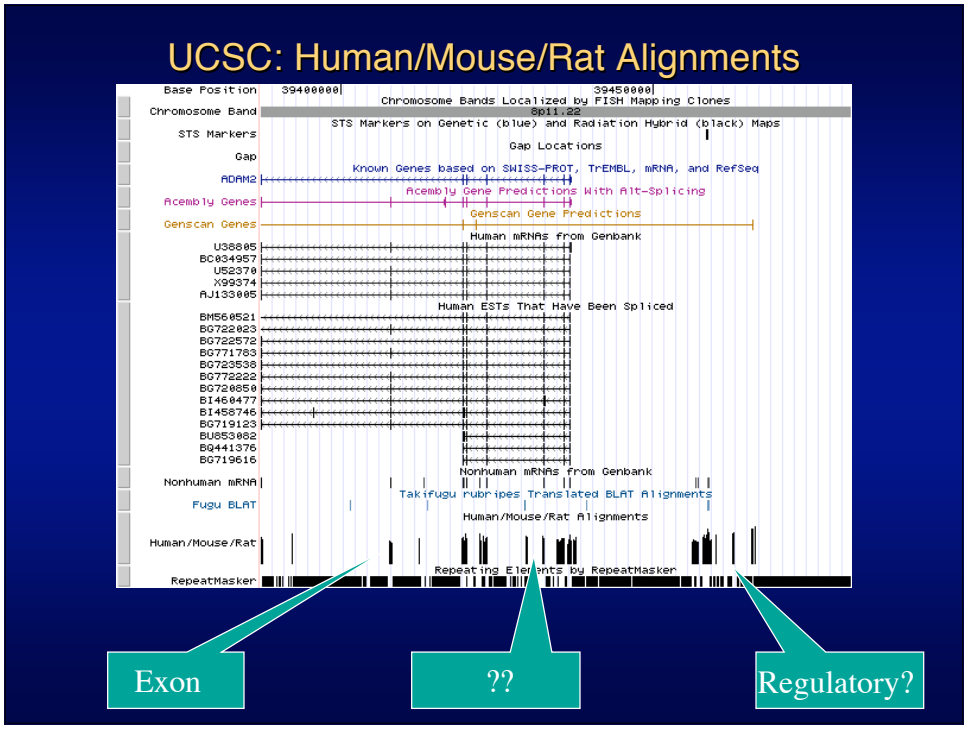
```
00000558 caccccagaaccataagtctggaatcacttgcagttattttagctcaatt 00000607
>>>>>>>> |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<<
39375627 caccccagaaccataagtctggaatcacttgcagttattttagctcaatt 39375578

00000608 attgagccttagtatggggatcacttatgatgacattaacaaatgccagt 00000657
>>>>>>>> |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<<
39375577 attgagccttagtatggggatcacttatgatgacattaacaaatgccagt 39375528

00000658 gctcaggagctgtctgcattatgaatccagaagcaat 00000694
>>>>>>>> |||||||||||||||||||||||||||||||||||| <<<<<<<<
39375527 gctcaggagctgtctgcattatgaatccagaagcaat 39375491
```

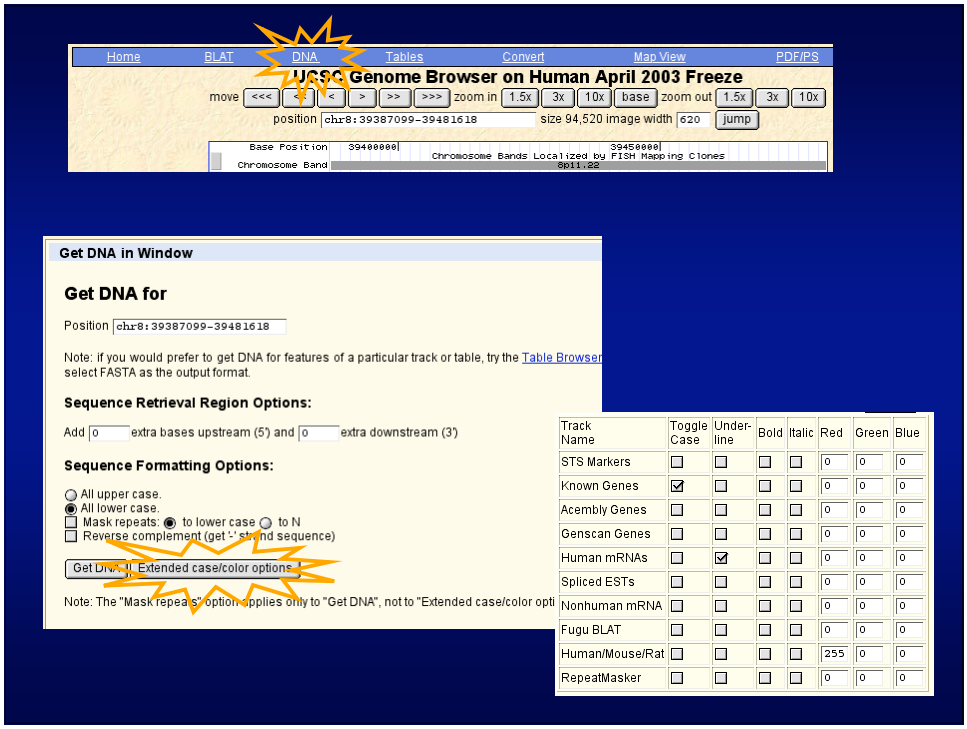
cDNA (U38805)

```
00000413 gggcgtactacagtttgaaaatgttagttatggaatagaacccctggagt 00000462
>>>>>>>> |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<<
39419636 gggcgtactacagtttgaaaatgttagttatggaatagaacccctggagt 39419587

00000463 cttcagttggctttgaacatgtaatttaccaagtaaaacataagaaagca 00000512
>>>>>>>> |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<<
39419586 cttcagttggctttgaacatgtaatttaccaagtaaaacataagaaagca 39419537

00000513 gatgtttccttatataatgagaaggatattgaatcaagagatctgtcctt 00000562
>>>>>>>> |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<<
39419536 gatgtttccttatataatgagaaggatattgaatcaagagatctgtcctt 39419487

00000563 taaattacaaagcgcagag 00000581
>>>>>>>> |||||||||||||||||| <<<<<<<<
39419486 taaattacaaaggctagag 39419468
```

4 exons

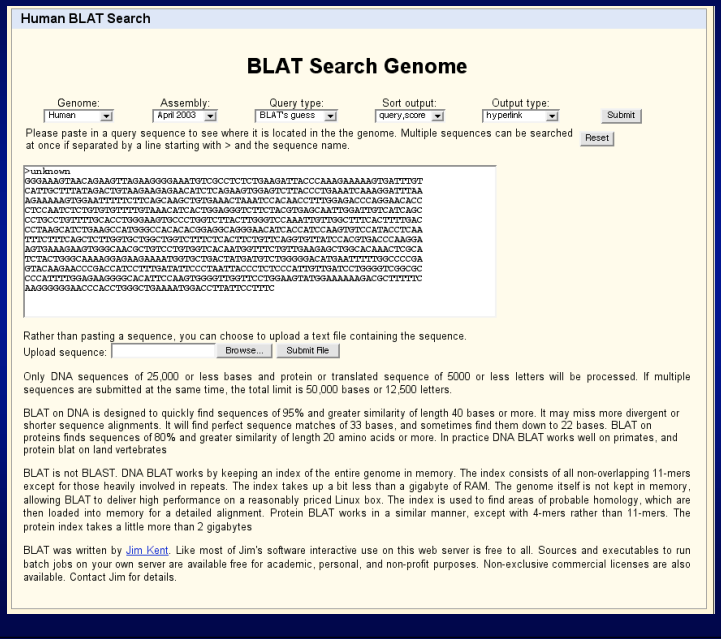
```
00000960 caccccagaaccataagtctggaatcacttgcagttattttagctcaatt 00001009
>>>>>>>> |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<<
39375627 caccccagaaccataagtctggaatcacttgcagttattttagctcaatt 39375578

00001010 attgagccttagtatggggatcacttatgatgacattaacaaatgccagt 00001059
>>>>>>>> |||||||||||||||||||||||||||||||||||||||||||||||||| <<<<<<<<
39375577 attgagccttagtatggggatcacttatgatgacattaacaaatgccagt 39375528

00001060 gctcaggagctgtctgcattatgaatccagaagcaat 00001096
>>>>>>>> |||||||||||||||||||||||||||||||||||| <<<<<<<<
39375527 gctcaggagctgtctgcattatgaatccagaagcaat 39375491
```


UCSC: Human/Mouse/Rat Alignments
Base Position
Chromosome Band
STS Markers
Gap
ADAM2
Acembly Genes
Genscan Genes
U38805
BC034957
U52370
X99374
AJ133005
BM560521
BG722023
BG722572
BG771763
BG723538
BG772222
BG720850
BI460477
BI458746
BG719123
BU853062
BQ441376
BG719616
Nonhuman mRNA
Fugu BLAT
Human/Mouse/Rat
RepeatMasker
Chromosome Bands Localized by FISH Mapping Clones
STS Markers on Genetic (blue) and Radiation Hybrid (black) Maps
Gap Locations
Known Genes based on SWISS-PROT, TrEMBL, mRNA, and RefSeq
Acembly Gene Predictions With Alt-Splicing
Genscan Gene Predictions
Human mRNAs from Genbank
Human ESTs That Have Been Spliced
Nonhuman mRNAs from Genbank
Takifugu rubripes Translated BLAT Alignments
Human/Mouse/Rat Alignments
Repeating Elements by RepeatMasker
Exon
??
Regulatory?


Home
BLAT
DNA
Tables
Convert
Map View
PDF/PS
UCSC Genome Browser on Human April 2003 Freeze
move
zoom in
zoom out
position chr8:39387099-39481618
size 94,520 image width 620
jump
Get DNA in Window
Get DNA for
Position chr8:39387099-39481618
Sequence Retrieval Region Options:
Add 0 extra bases upstream (5') and 0 extra downstream (3')
Sequence Formatting Options:
All upper case.
All lower case.
Mask repeats: to lower case to N
Reverse complement (get '-' strand sequence)
Extended case/color options
Track Name
Toggle Case
Under-line
Bold
Italic
Red
Green
Blue
STS Markers
Known Genes
Acembly Genes
Genscan Genes
Human mRNAs
Spliced ESTs
Nonhuman mRNA
Fugu BLAT
Human/Mouse/Rat
RepeatMasker

Known genes in upper case; Human mRNAs underlined; Human/Mouse/Rat alignments in red

UCSC BLAT search



UCSC BLAT results

Human BLAT Results

BLAT Search Results

ACTIONS	QUERY	SCORE	START	END	QSIZE	IDENTITY	CHRO	STRAND	START	END
browser details	unknown	664	4	767	814	98.1%	3	-	120544667	120559679

Alignment of unknown and chr3:120544667-120559679

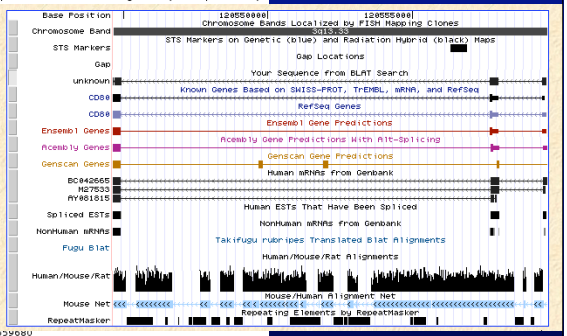
Click on links in the frame to left to navigate through alignment. Matching bases in cDNA and genomic sequences are colored blue and capitalized. Light blue bases mark the boundaries of gaps in either side of the alignment (often splice sites).

cDNA unknown

```
gggAAAGTAA CAGAAGTTAG AAGGGGAAAT GTCGCCTCTC TGAAGATTAC   50
CCAAAGAAAA AGTGATTTGT CATTGCTTTA TAGACTGTAA GAAGAGAACA  100
TCTCAGAAGT GGAGTCTTAC CCTGAAATCA AAGGATTTAA AGAAAAAGTG  150
GAATTTTTCT TCAGCAAGCT GTGAAACTAA ATCCACAACC TTTGGAGACC  200
CAGGAACACC CTCCAATCTC TGTGTGTTTT GTAAACATCA CTGGAGGGTC  250
TTCTACGTGA GCAATTGGAT TGTCATCAGC CCTGCCTGTT TTGCACCTGG  300
GAAGTGCCCT GGTCTTACTT GGGTCCAAAT TGTTGGCTTT CACTTTTGAC  350
CCTAAGCATC TGAAGCCATG GGCCACACAC GGAGGCAGGG AACATCACCA  400
TCCAAGTGTC CATACCTCAA TTTCTTTCAG CTCTTGGTGC TGGCTGGTCT  450
TTCTCACTTC TGTTCAGGTG TTATCCACGT GACCCAAGGA AGTGAAAGAA  500
GTgGGCAACG CTGTCCTGTG GTCACAATgG TTTCTGTTGA AGAGCTGGCA  550
CAAACTCGCA TCTACTgGGC AAAAGGAGAA GAAAATGTTG CTGACTATGA  600
TGTCTGGGGg ACATGAATtt ttggcCCGA GTACAAGAAC CeGACCATeC  650
TTTGATATtc cctaattacC CTCTCCeATT GtTGATCCTG Gggtcgggcgc  700
cccattttgg agaaggggca cattccaagt ggggtggtt cctgGAAGTA  750
TGaAAAAAG ACGCTTTttc aaggggggaa cccacctggg ctgaaaatgg  800
accttattcc tttc
```

Genomic chr3 (reverse strand):

```
acaaaagcaa ctagaagaag acggcagcag agggcgcttc ttcaaacacc  120
ctgtccaact cctggcactg acaagtactg agtgaactca aaccctctgt  120559680
AAGTAACAG AAGTTAGAAG GCGAAATGTC GCCTCTTCTGA AGATTACCCA  120559630
AAGAAAAAGT GATTTGTCAT TGCTTTATAG ACTGTAAGAA GAGAACATCT  120559580
CAGAAGTGGA GTCTTACCCT GAAATCAAAG GATTTAAAGA AAAAGTGGAA  120559530
TTTTTCTTCA GCAAgtaagt acatttacta ttatcttact atgggatagc  120559480
ttggtagaac tagttaccag caacctcact gtttttaaa caaacaatac  120559430
ctctgctacc acctcctgga tccaggcttg ccctcaaagg aacccttagg  120559380
```

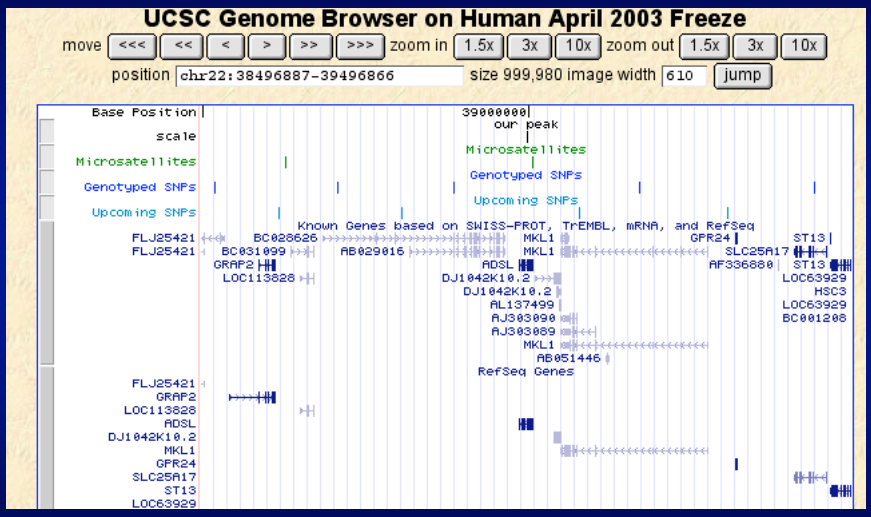


"Add your own tracks" to UCSC

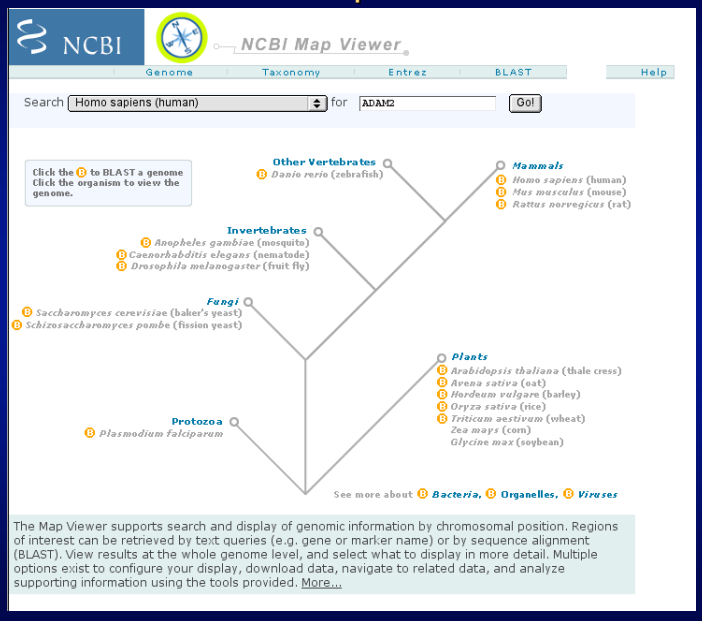
```
browser position chr22:38496887-39496866
browser hide cytoBand
browser hide stsMap
browser hide gap
browser hide clonePos
browser full refGene
browser dense mrna
track name="scale" description="our peak"
chr22 38996887 38996888 peak
track name="Microsatellites" description="Microsatellites" color=0,128,0
chr22 38627059 38627060 D22S276
chr22 39005417 39005418 D22S307
track name="Genotyped SNPs" description="Genotyped SNPs" color=0,0,255
chr22 38518342 38518343 ss146131
chr22 38705963 38705964 ss2941443
chr22 38884157 38884158 ss141110
chr22 39171390 39171391 ss22916
chr22 39438769 39438770 ss1479794
track name="Upcoming SNPs" description="Upcoming SNPs" color=0,128,192
chr22 38615712 38615713 ss86855
chr22 38804838 38804839 ss85533
chr22 39077895 39077896 ss141190
chr22 39305065 39305066 ss137027
```

Nature Genetics User's Guide, Question 7

"Add your own tracks" to UCSC

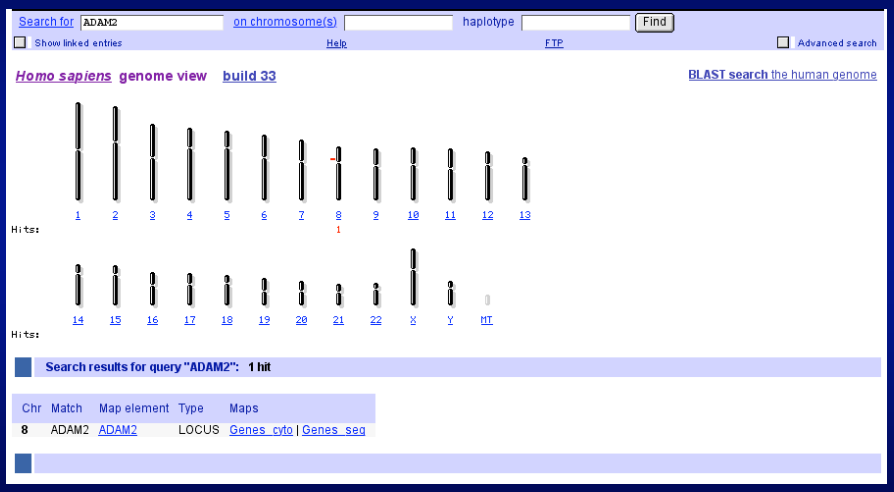


NCBI Map Viewer

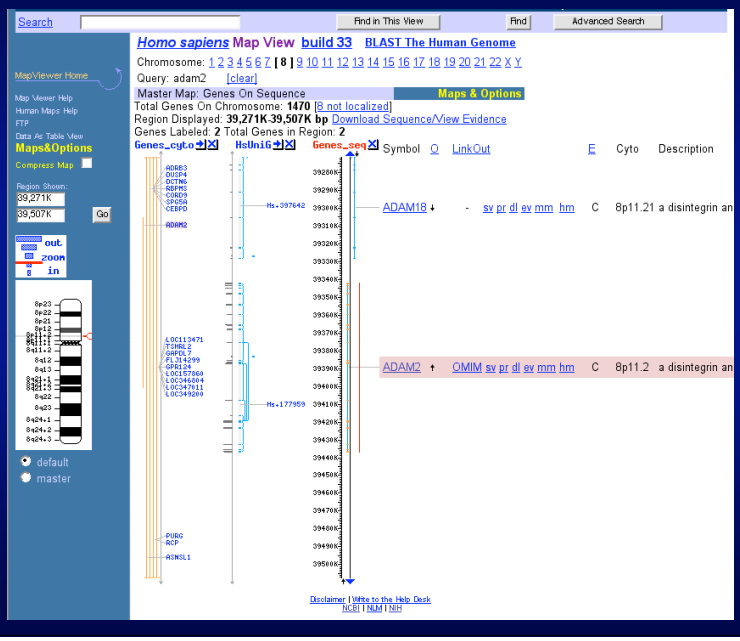


http://www.ncbi.nlm.nih.gov/mapview/

NCBI Map Viewer search results



NCBI default view of ADAM2



NCBI download (dl)

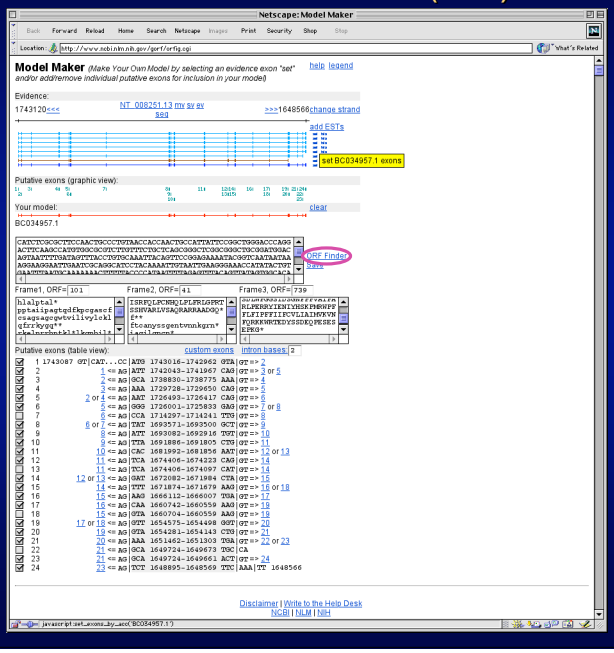
Homo sapiens Genome (build 33)
Region to retrieve (in chromosome coordinates):
Chromosome: 8 Strand: plus
from: 39342201 adjust by: -0K
to: 39436755 adjust by: +0K Change Region/Strand

Sequence Format: FASTA

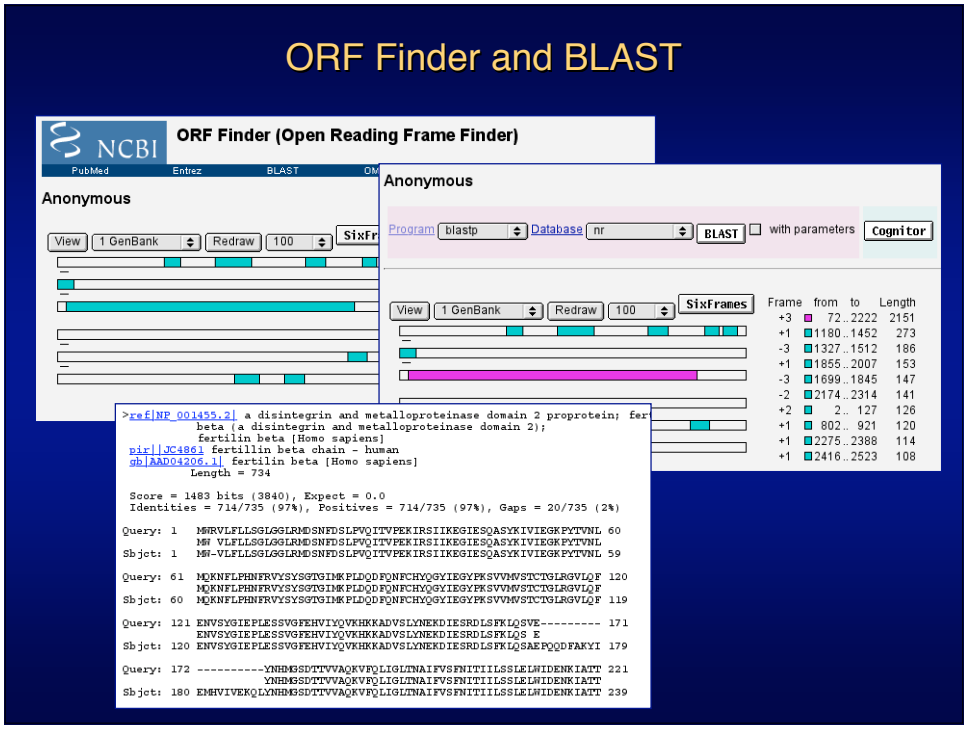
This chromosome region corresponds to the contig region(s):

Contig	start	stop	strand	
NT_008251.13	1648566	1743120	+	Display Save to Disk View Evidence ModelMaker

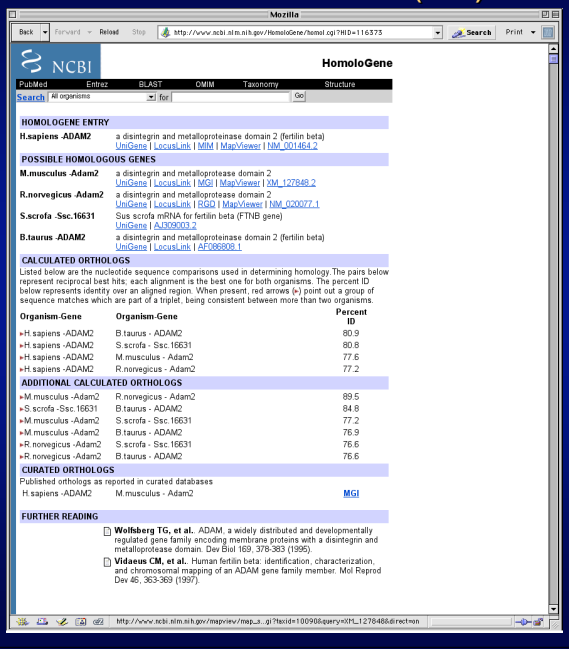
NCBI model maker (mm)



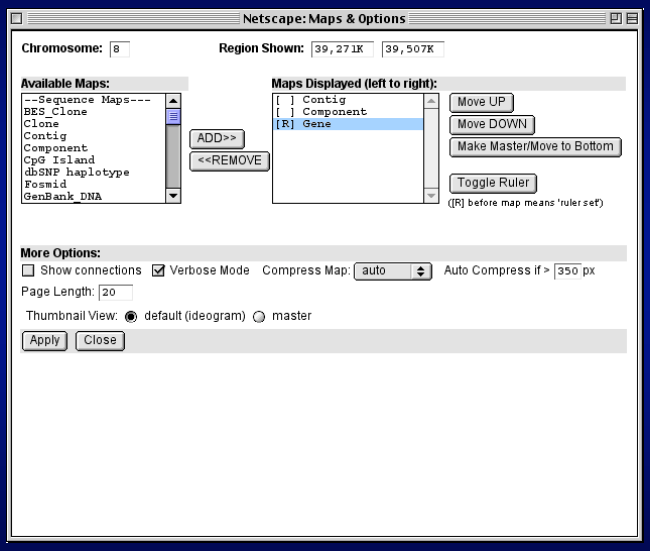
ORF Finder and BLAST



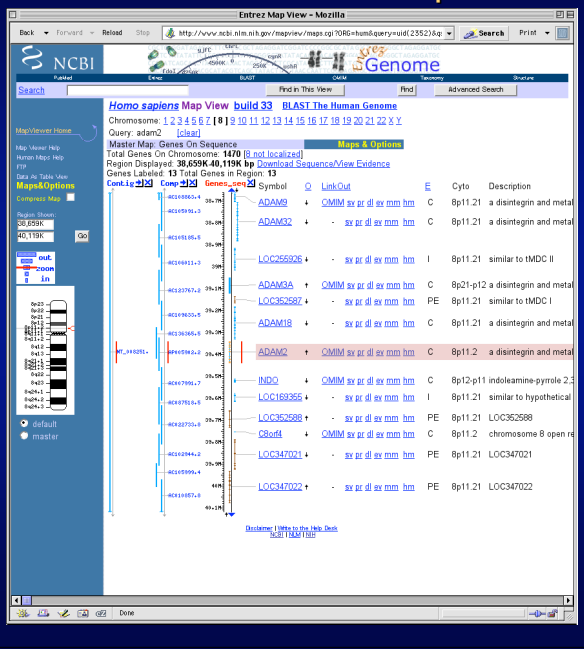
NCBI HomoloGene (hm)



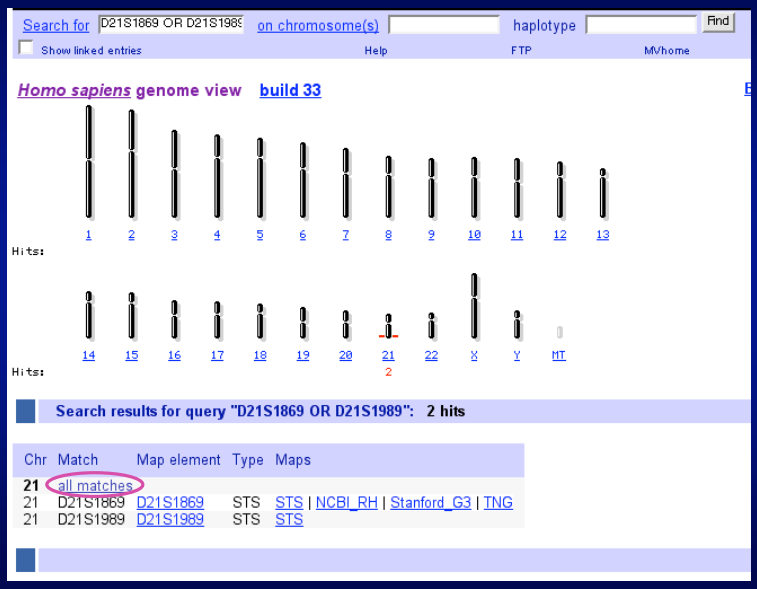
NCBI *Maps & Options*



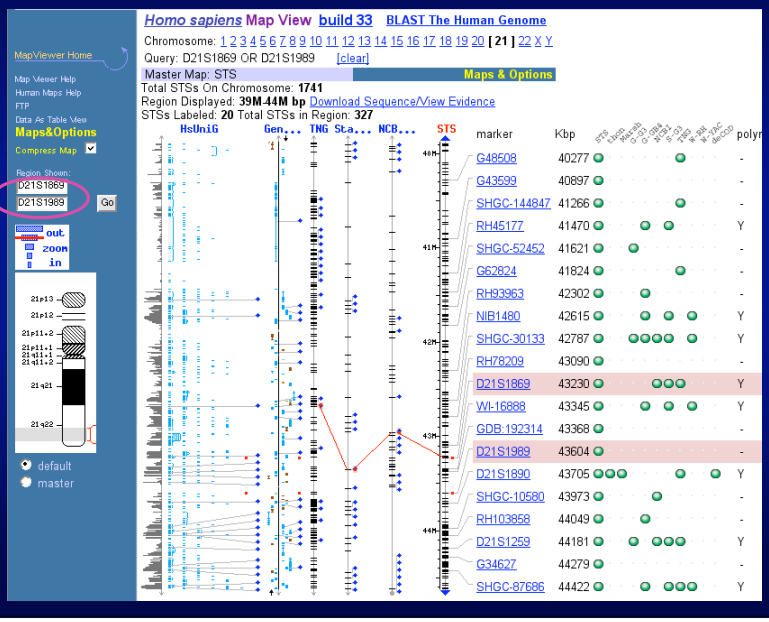
Additional NCBI maps



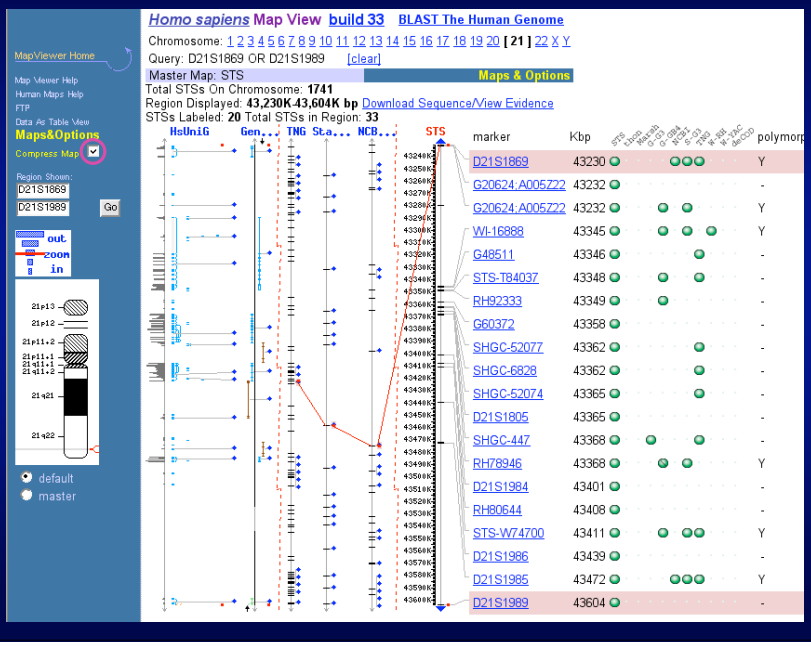
NCBI: Search for region between 2 STS markers



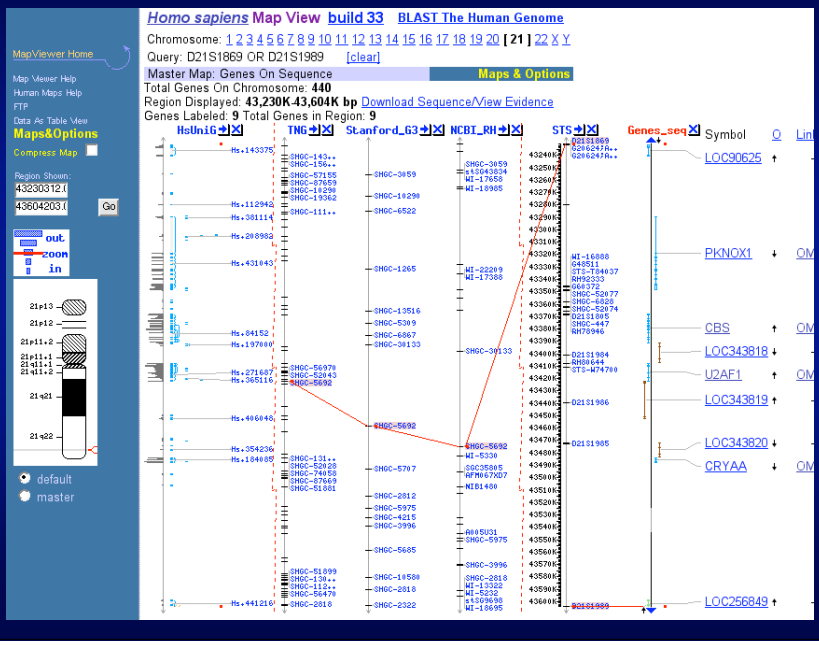
NCBI: Default view of region



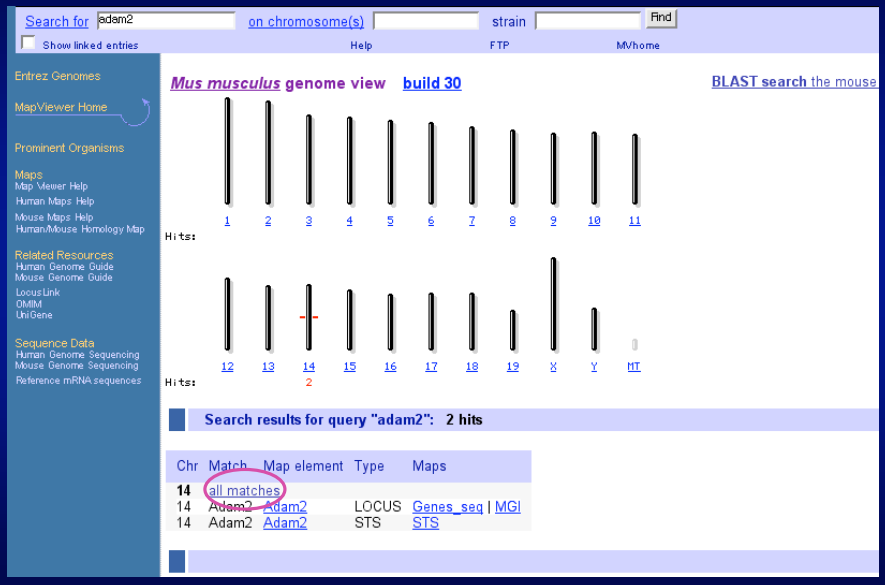
NCBI: Zoomed-in view of region



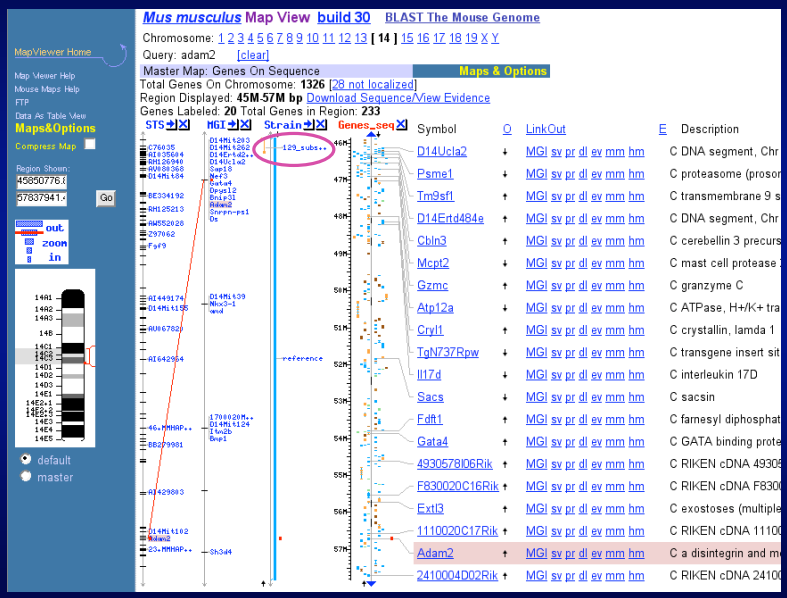
NCBI: Genes between the 2 STS markers



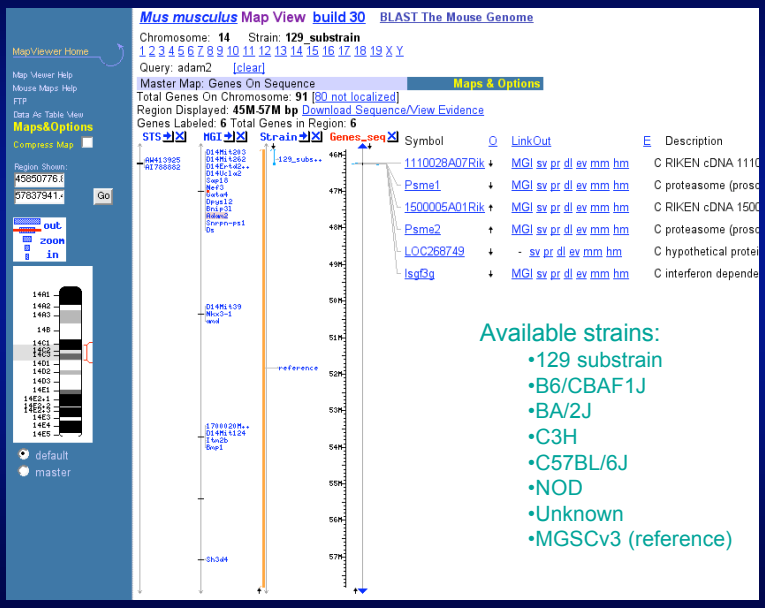
NCBI: Search for ADAM2 in the mouse genome



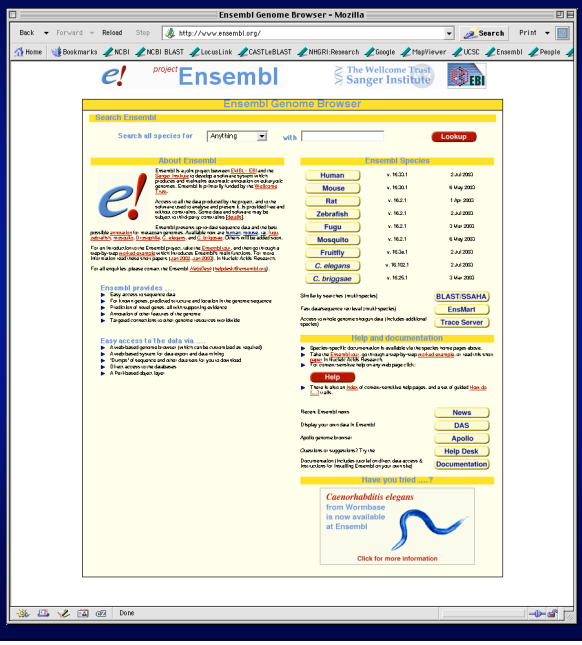
NCBI: Default view of ADAM2 in mouse



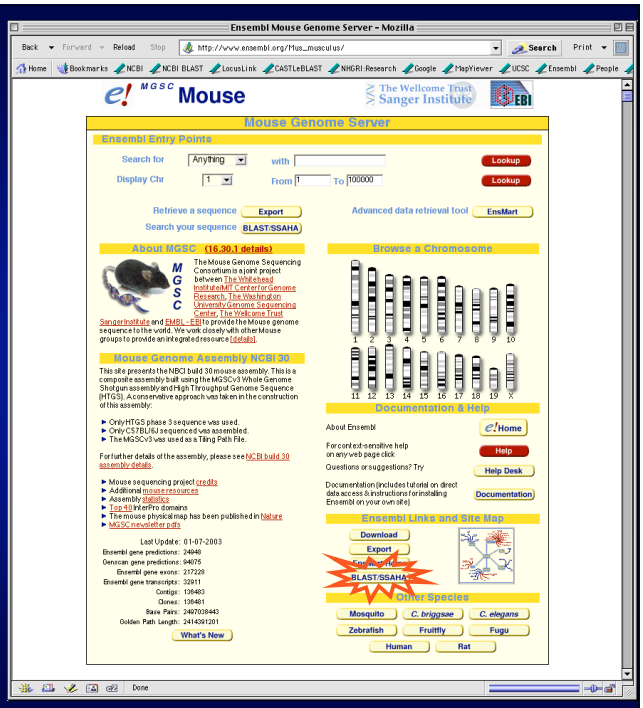
NCBI: View features annotated on sequence from other mouse strains



Ensembl Home

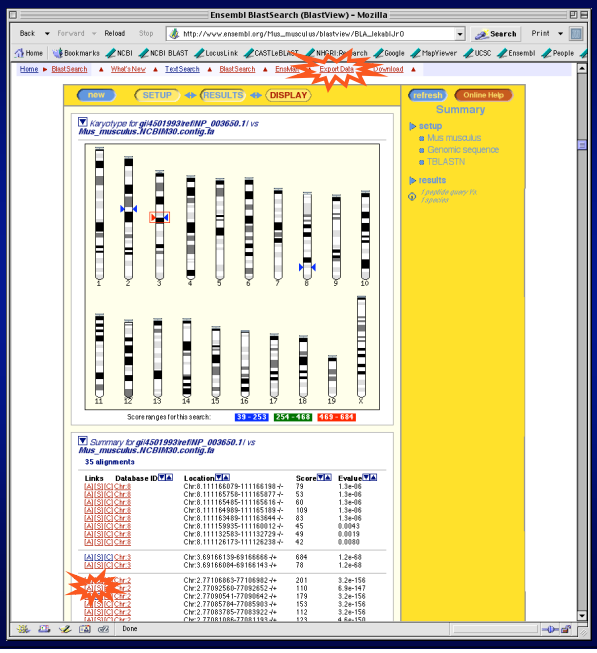


Ensembl Mouse Home



Ensembl: BLAST human protein against mouse genome

Ensembl BLAST results



Ensembl BLAST details

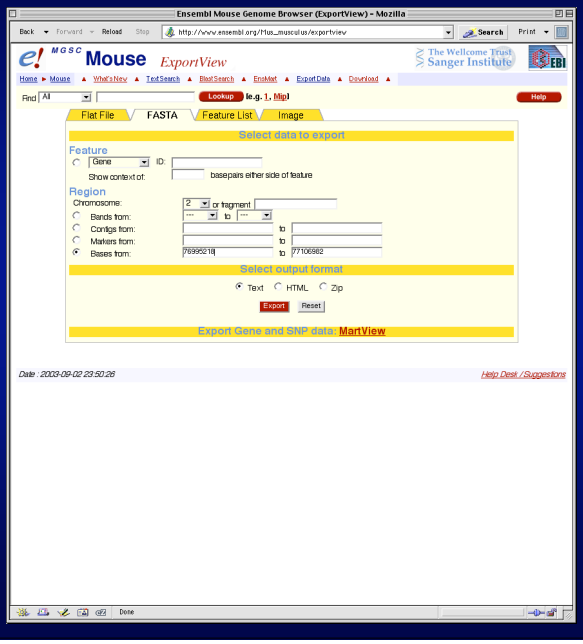
A (alignment) link

```
Query localtion    : gi|4501993|ref|NP_003650.1|      566 to 599
Database location  : 2                          77090541 to 77090642
Genomic location   : 2                          77090541 to 77090642

Alignment score    : 179
E-value            : 3.2e-156
Alignment length   : 34
Percentage identity: 97.06

Query:      566 RVTQTYDAGACIYFYFAFNYRGISDPLTVFEQTE 599
                RVTQTYDAGACIYFYFAFNYRGISDPLTVFE TE
Sbjct: 77090541 RVTQTYDAGACIYFYFAFNYRGISDPLTVFEHTE 77090642
```

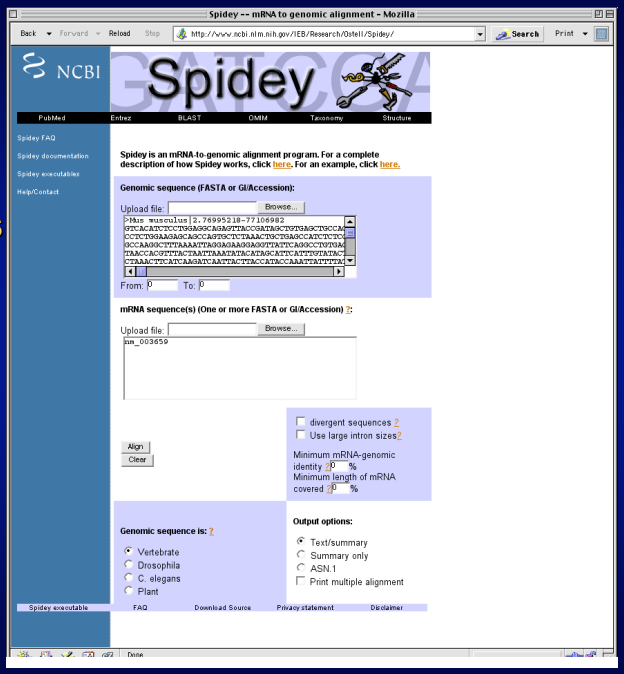
S (query sequence) link

```
Query localtion    : gi|4501993|ref|NP_003650.1|      566 to 599
Database location  : 2                          77090541 to 77090642
Genomic location   : 2                          77090541 to 77090642

Alignment score    : 179
E-value            : 3.2e-156
Alignment length   : 34
Percentage identity: 97.06

THIS COLOUR: Matching bases for selected HSP
THIS COLOUR: Matching bases for other HSPs in selected hit

>gi|4501993|ref|NP_003650.1|
MAEAAAAAGGTGLGAGASYGSAADRDRDPDPDRAGRRLRYLSGHLLGRPREALSTNECKA
RFAASAATAAPTATPAAQESGTIPKKRQEVMKWNGWGYNDSKFIFNKKGQIELTGKRYPL
SGMGLPTFKEWIQNTLGVNVEHKTTSKASLNFSDTPPSVVNEDFLHDLKETNISYSQEAD
DRVFRANGHCLHEIFLLREGMFERIPDIVLWPTCHDDVVKIVNLACKYNLCIIPIGGGTS
VSYGLMCPADETRTIISLDTSQMNRILWVDENNLTAHYEAGITGQELERQLKESGYCTGH
EPDSLEFSTVGGWVSTRASGMKKNIYGNIEDLVVHIKMVTPRGIIEKSCQGPRMSTGPDI
HHFIMGSEGTLGVITEATIKIRPVPEYQKYGSVAFPNFEQGVACLREIAKQRCAPASIRL
MDNKQFQFGHALKPQVSSIFTSFLDGLKKFYITKFKGFDPNQLSVATLLFEGDREKTLQH
EKQVYDIAAKFGGLAAGEDNGQRGYLLTVVIAYIRDLALEYYVLGESFETSAPWDRVVDL
CRNVKEKITRECKEKGVQFAPFSTCRVTQTYDAGACIYFYFAFNYRGISDPLTVFEQTEA
AAREEILANGGSLSHHHGVGKLRKQWLKESISDVGFGMLKSVKEYVDPNNIFGNRNLL
```

Ensembl: Export genomic sequence of hit



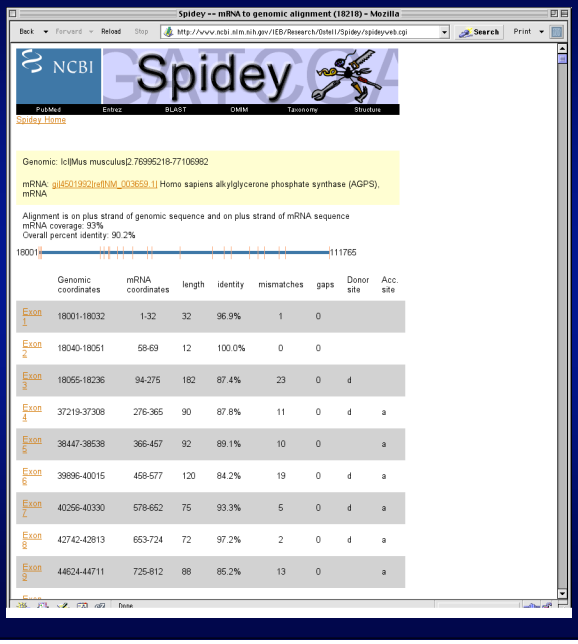
NCBI's Spidey: mRNA/Genomic alignments to determine positions of exons



Wheelan SJ, Church DM, Ostell JM.
Spidey: a tool for mRNA-to-genomic alignments.
Genome Res. 2001 Nov;11(11):1952-7.

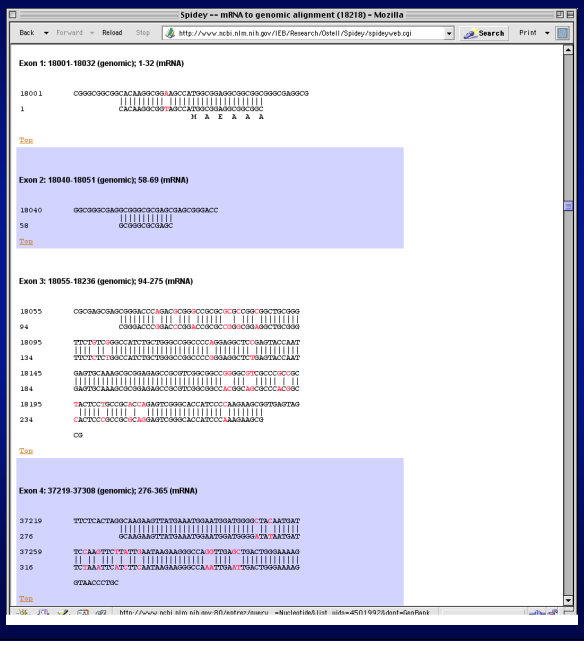
http://www.ncbi.nlm.nih.gov/spidey

NCBI: Spidey results overview

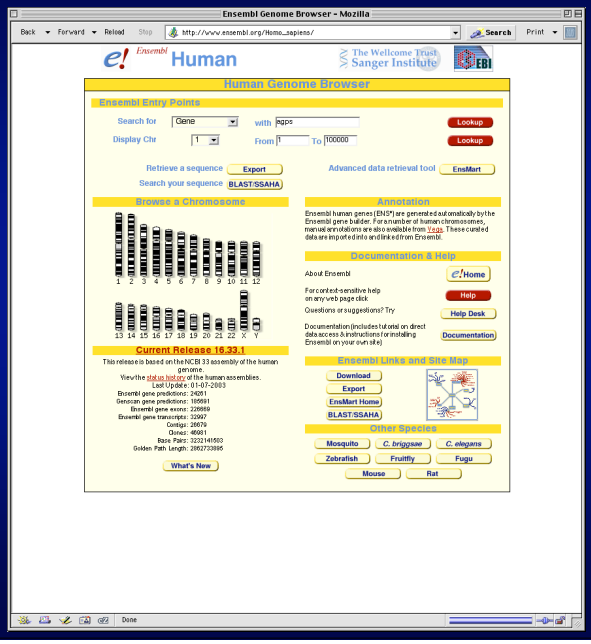


	Genomic coordinates	mRNA coordinates	length	identity	mismatches	gaps	Donor site	Acc. site
Exon 1	18001-18032	1-32	32	96.9%	1	0		
Exon 2	18040-18051	58-69	12	100.0%	0	0		
Exon 3	18055-18236	94-275	182	87.4%	23	0	d	
Exon 4	37219-37308	276-365	90	87.8%	11	0	d	a
Exon 5	38447-38538	366-457	92	89.1%	10	0		a
Exon 6	39896-40015	458-577	120	84.2%	19	0	d	a
Exon 7	40256-40330	578-652	75	93.3%	5	0	d	a
Exon 8	42742-42813	653-724	72	97.2%	2	0	d	a
Exon 9	44624-44711	725-812	88	85.2%	13	0		a

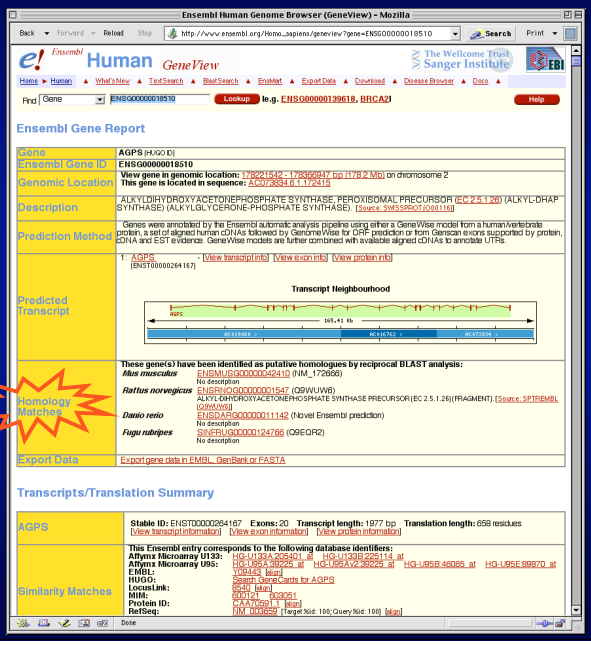
NCBI: Spidey results detail



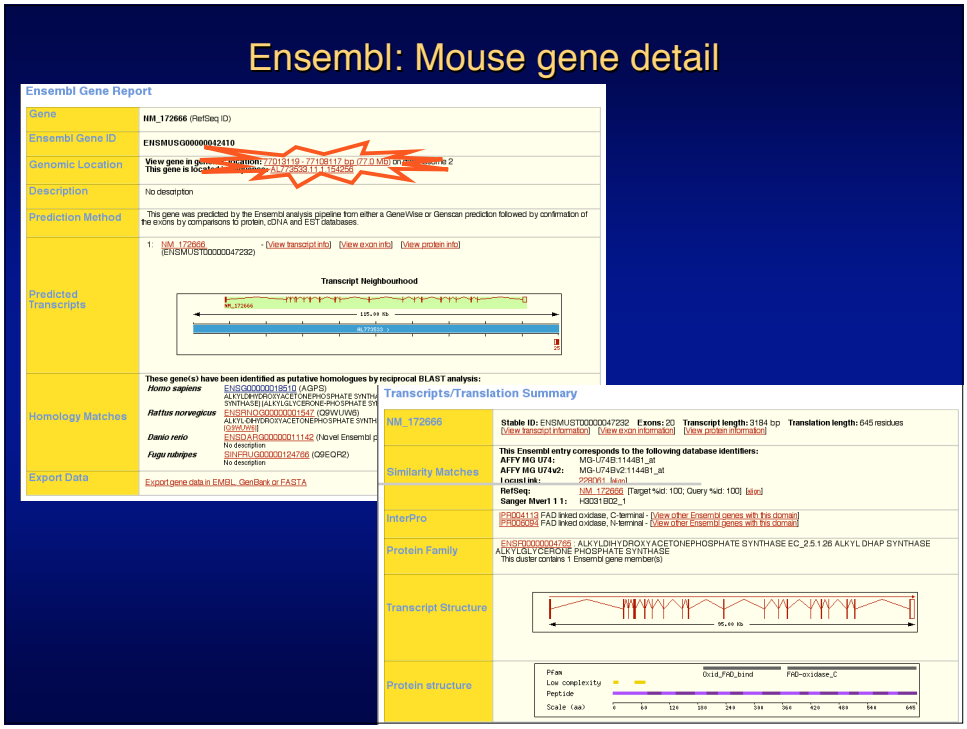
Ensembl: Search for human gene name

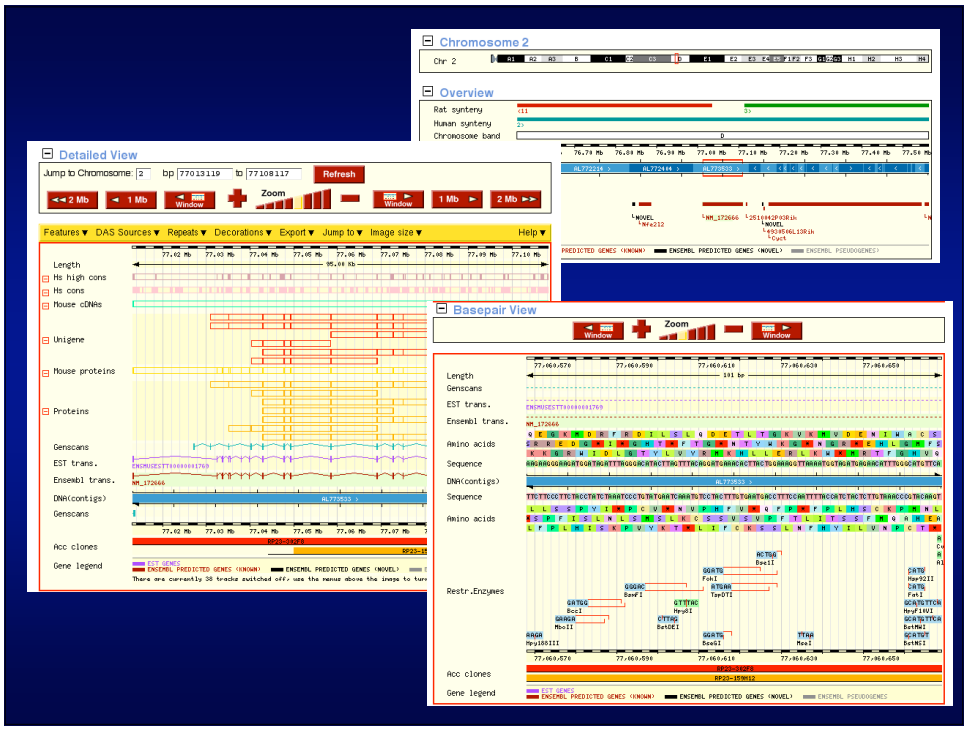


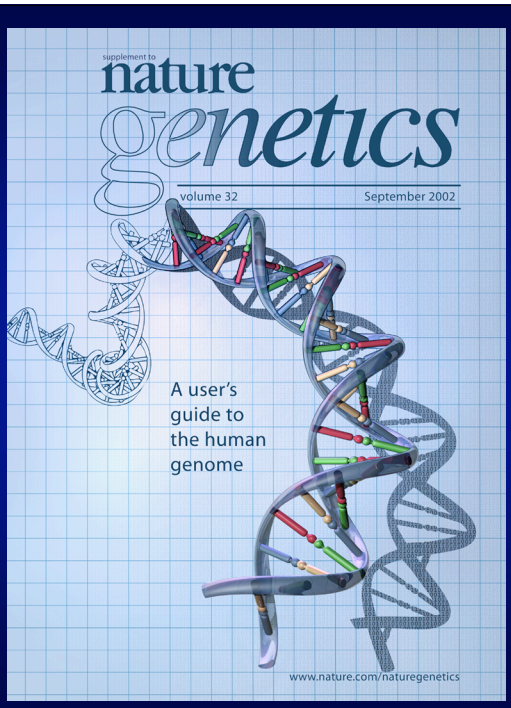
Ensembl: Human gene detail



Ensembl: Mouse gene detail







http://www.nature.com/genomics

Additional resources

- UCSC Human Genome Browser User Guide
  http://genome.ucsc.edu/goldenPath/help/hgTracksHelp.html
- NCBI Genomic Biology
  http://www.ncbi.nih.gov/Genomes/
- NCBI MapViewer Help
  http://www.ncbi.nlm.nih.gov/mapview/static/MapViewerHelp.html
- Ensembl Tour
  http://www.ensembl.org/Docs/enstour/