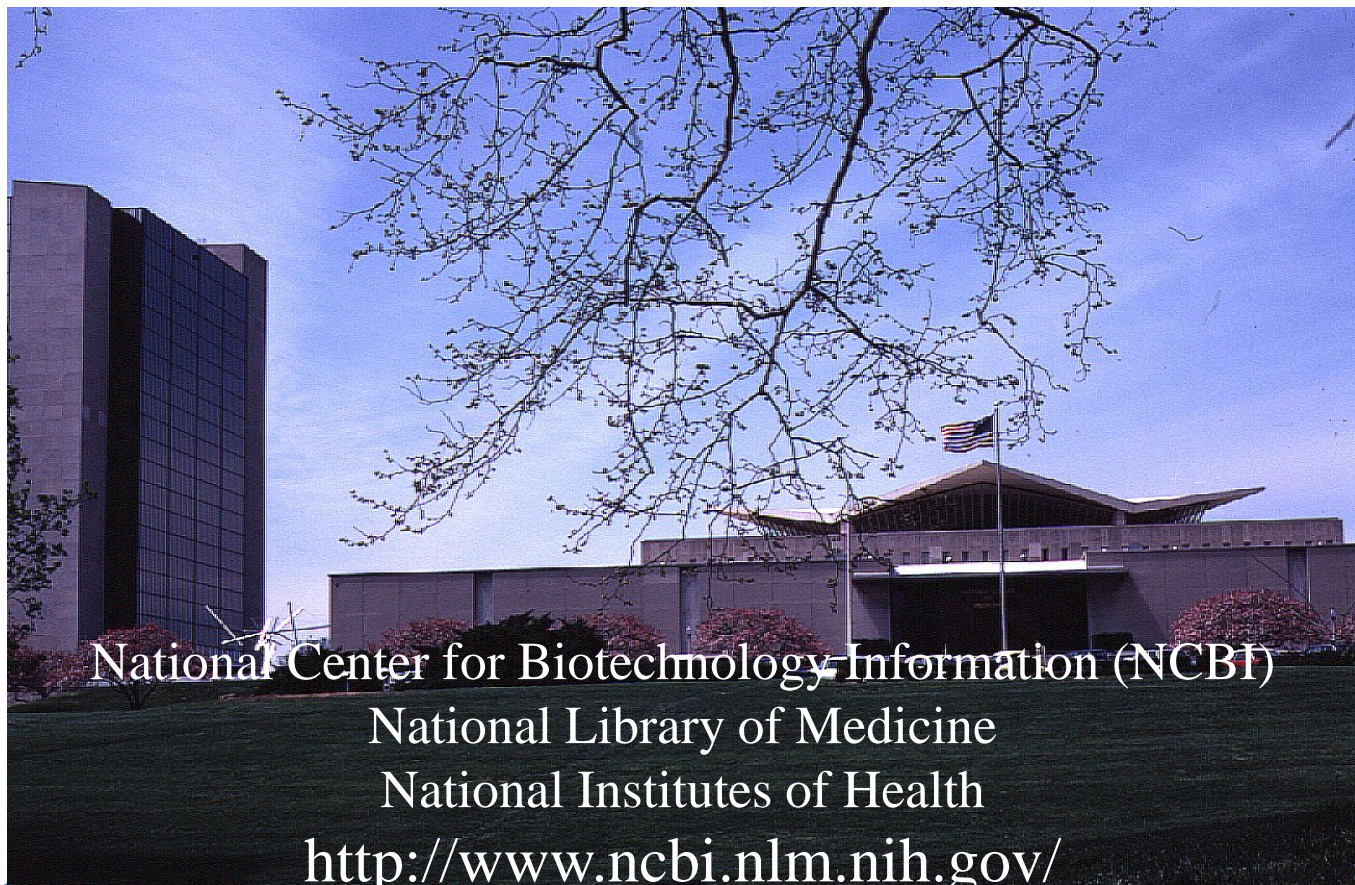


# Using NCBI Resources for Gene Discovery

Kim D. Pruitt  
Transcriptome 2002

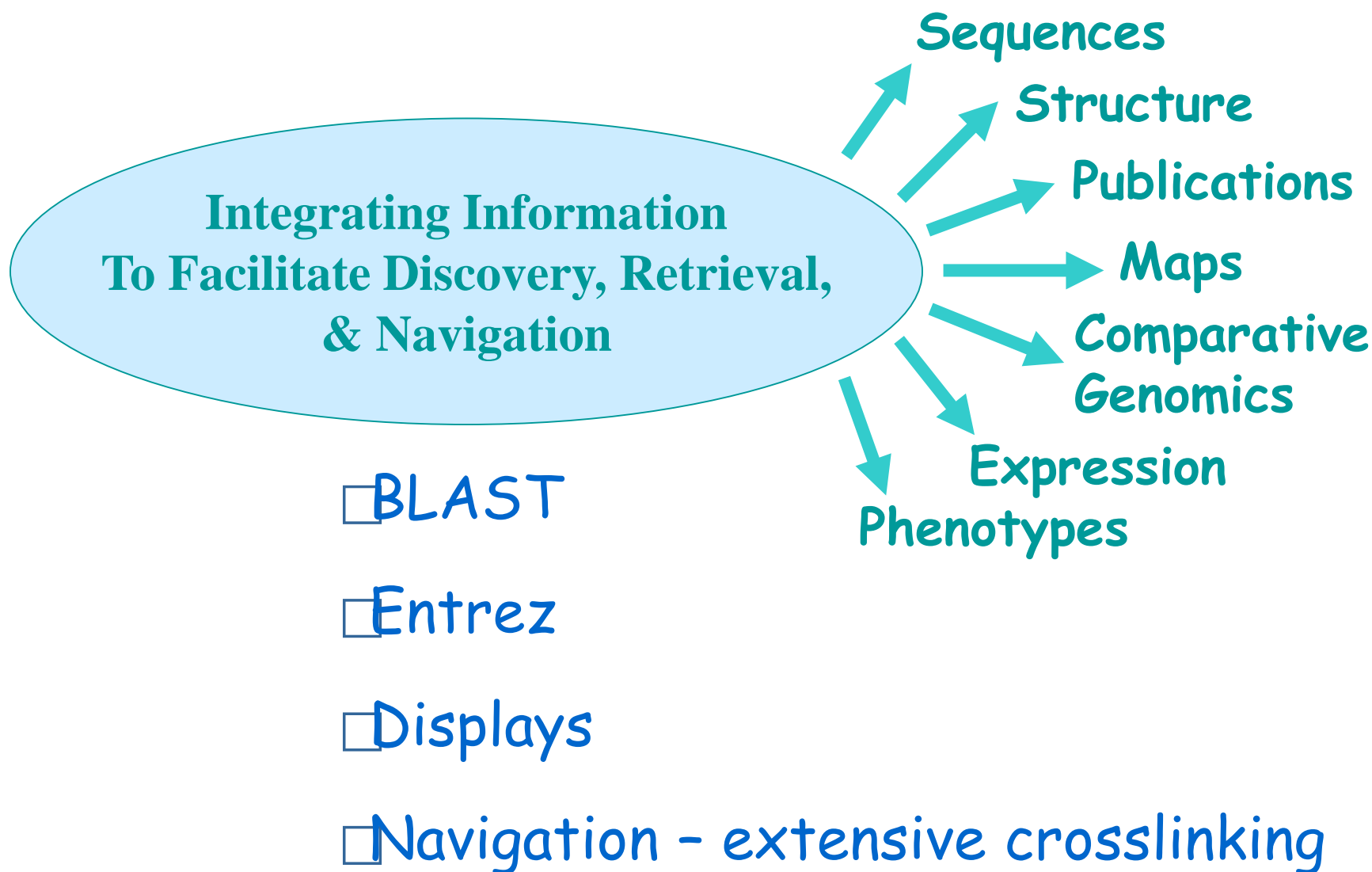


## Primary Databases - GenBank, dbEST, dbSTS, PubMed

- Archival - original data submissions
- Database staff organize, but don't add additional information

## Derivative Databases - RefSeq, LocusLink, UniGene, Map Viewer

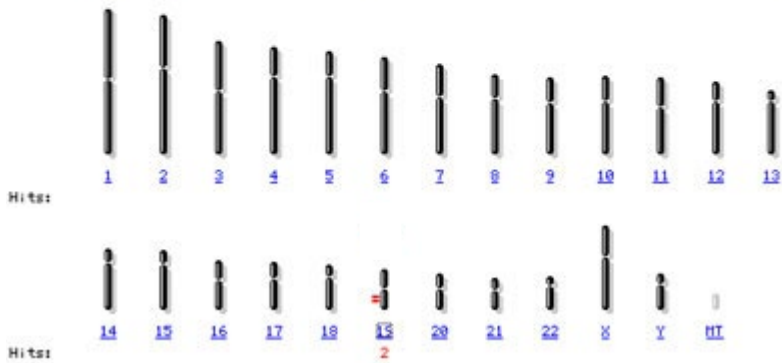
- Curated/expert review
  - compilation and correction of data
- Computationally Derived
- Combinations



# Increasing Discovery Space



## NCBI Map Viewer



## Genome Oriented Resource

- A sequence for each macromolecule of Central Dogma
- Linked on a residue by residue basis
- Objectively non-redundant and comprehensive

## Curated Resource

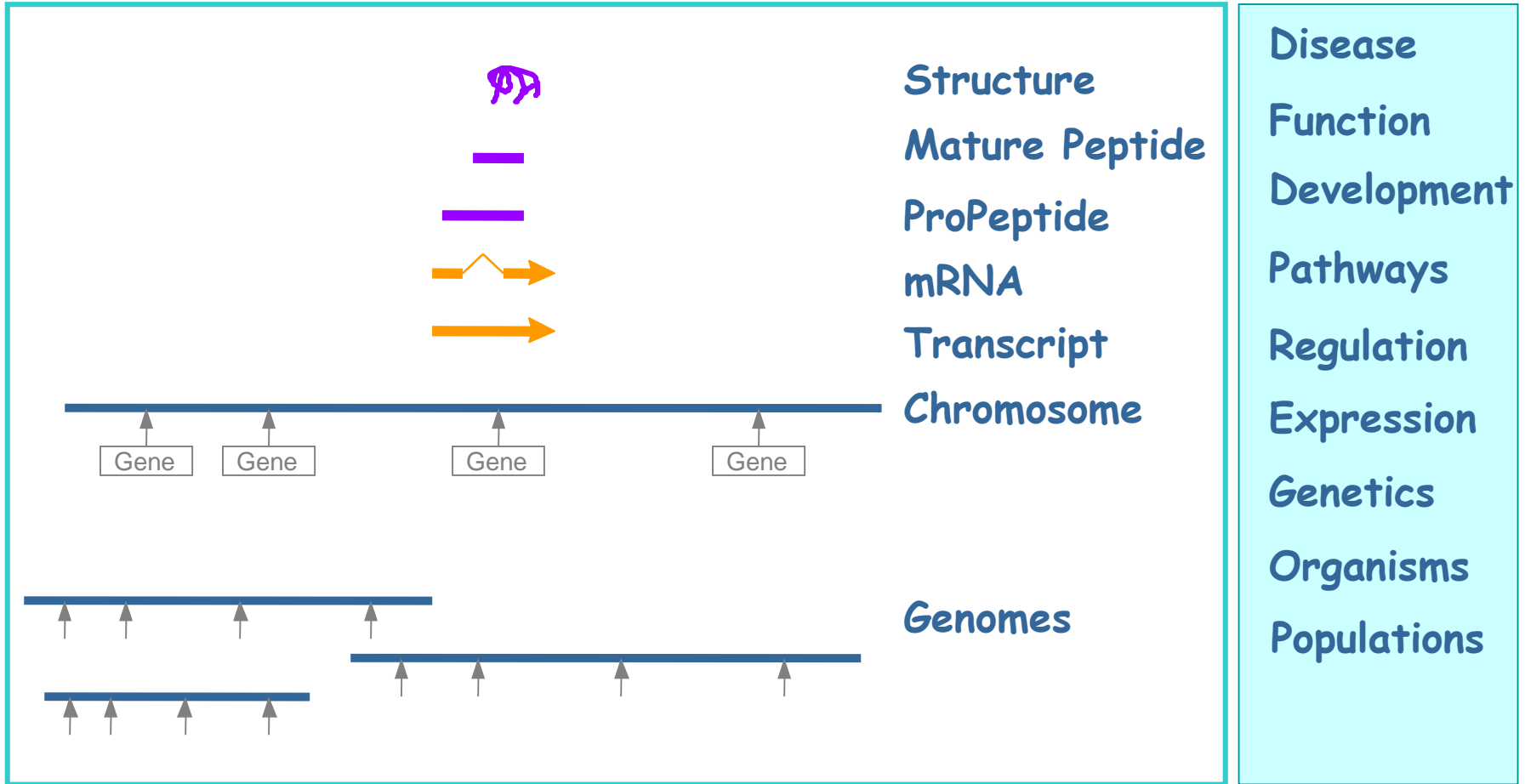
- Authoritative source by genome
- Derivative of GenBank but corrected, merged, extended
- Publicly distributed

## Reagents for Genome Annotation and Analysis

## Substrate for Functional Genomics

# The Basic Model

A framework to anchor other information...



# RefSeq: Scope

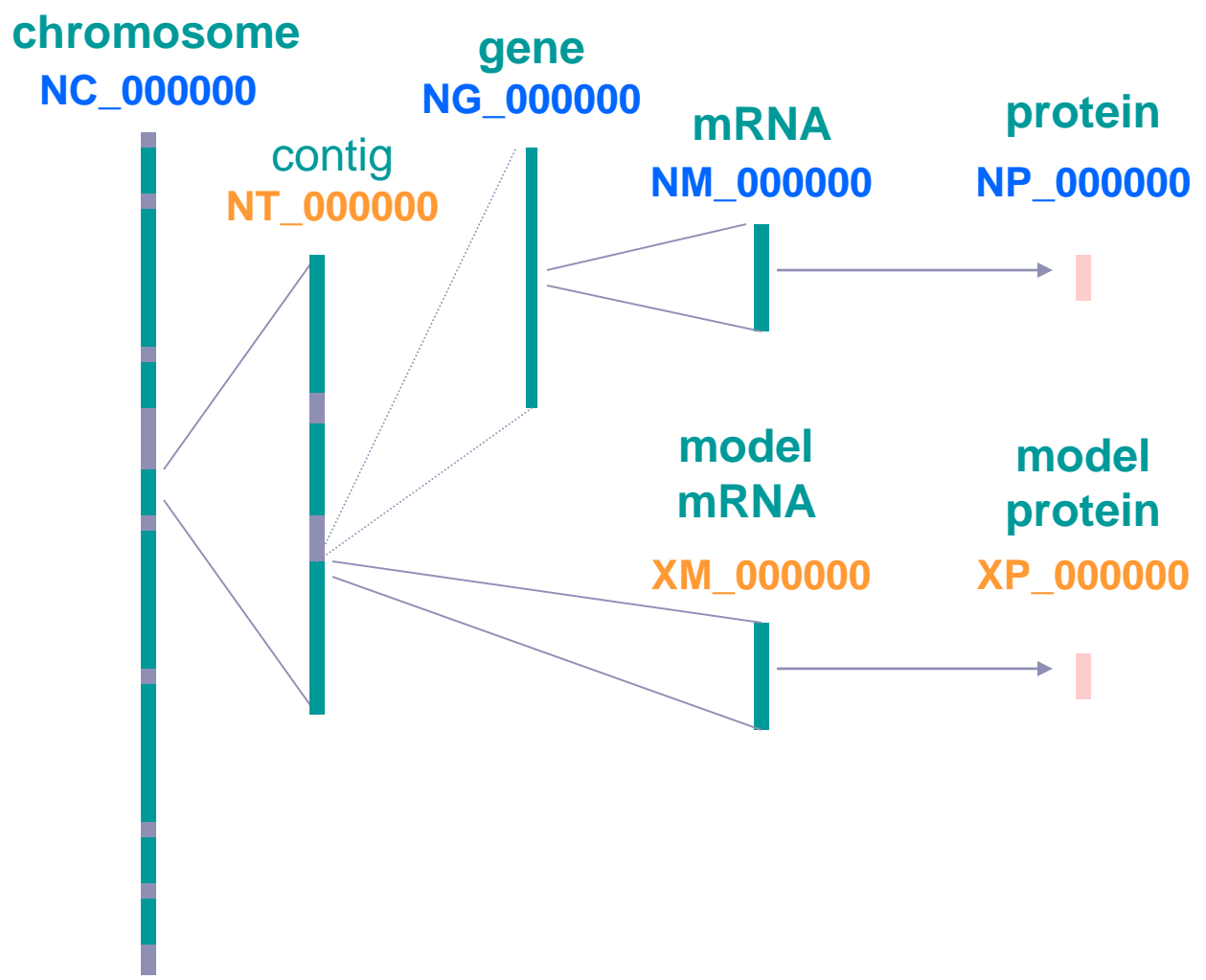
The NCBI Reference Sequence (RefSeq) project provides non-redundant sequence data including bacterial and viral genomes, mitochondrion, chromosomes, constructed genomic contigs, transcripts, and proteins.

<i>Source</i>	<i>Molecule</i>	<i>Organism</i>
<i>Microbial Genomes</i>		
<i>Complete Genomes</i>		
<i>Genome Build Pipeline</i>	Genomic Transcript Protein	Human Mouse
<i>Collaboration</i>	Genomic Transcript Protein	Yeast Drosophila Arabidopsis Miscellaneous other
<i>LocusLink-based</i>	Genomic Transcript Protein	Human Mouse Rat Drosophila Zebrafish

RefSeq as a protein database  
over 280,000 proteins

# RefSeq: Products

Goal: One sequence entry for each naturally occurring molecule



Multiple products for one gene are instantiated as separate RefSeqs with the same LocusID.

**Key:**  
 curated  
 calculated

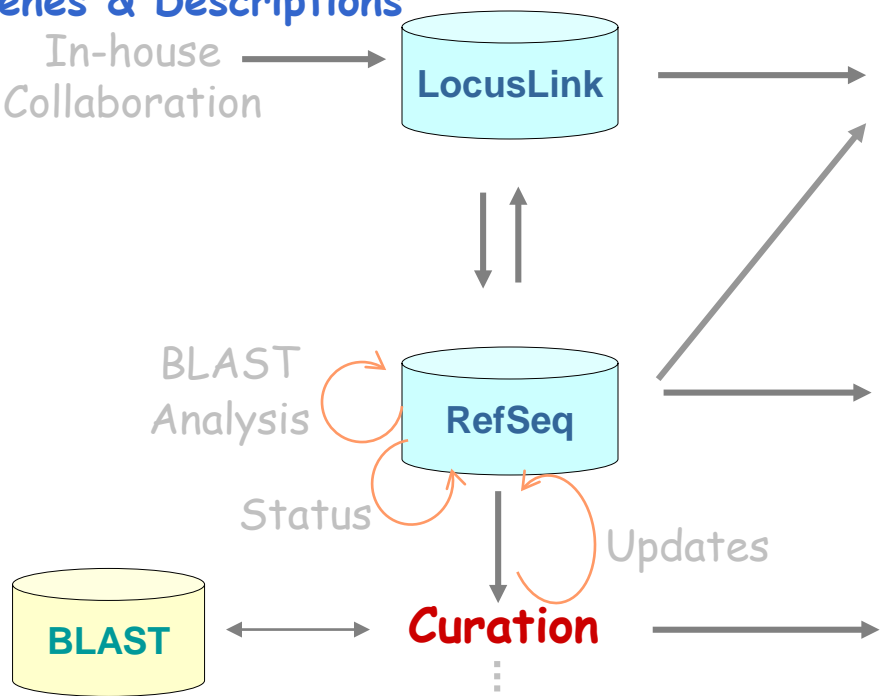


# Process Flow

Database support, automated steps, manual curation

## New Genes & Descriptions

In-house  
Collaboration



### Public Release:

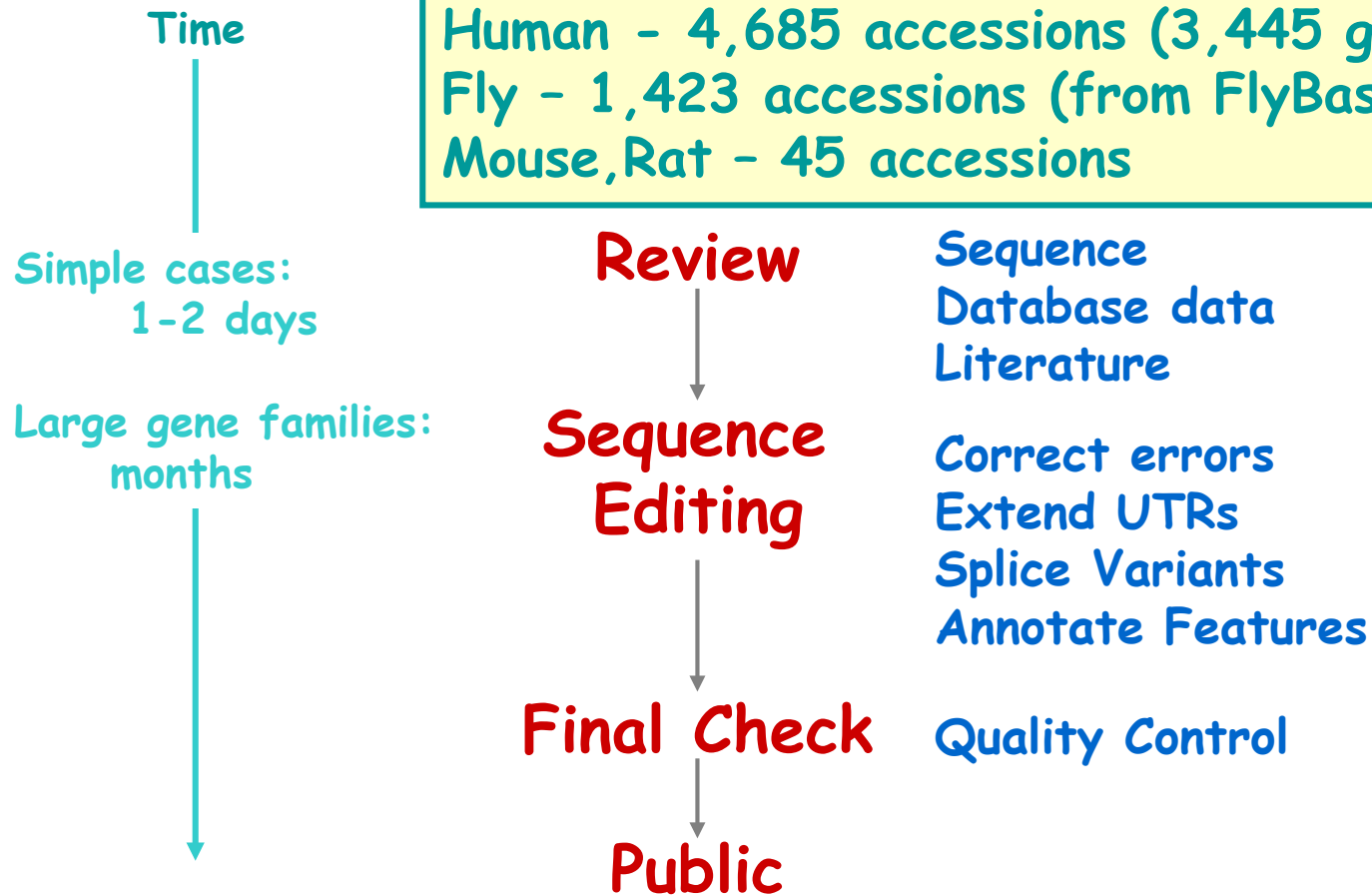
- LocusLink Web Site
- Provisional & Predicted Records (Transcripts & Proteins)
- Reviewed Records (Genomic Regions, Transcripts, Proteins)

### Accessible in:

- BLAST
- BLink
- Entrez
- FTP
- LocusLink

# RefSeq: Curation

**Reviewed Records:**  
 Human - 4,685 accessions (3,445 genes)  
 Fly - 1,423 accessions (from FlyBase)  
 Mouse,Rat - 45 accessions



## Why?

Correct Assembly through Duplications, Paralogous Gene Clusters



Optimize Annotation in Gene Clusters



Used in Genome Annotation Process

## New Genes:

GenBank  
UniGene  
Genome Annotation  
Collaboration  
e-mail

## Updates:

GenBank updates  
Collaboration  
Ongoing curation  
Genome Annotation  
e-mail

➡ We welcome feedback, suggestions, collaborations ←

## Why Look For RefSeqs?

### Enhanced Discovery Space:

What do we already know?

Predicted RefSeqs - where do we need to know more?

Genome Annotation Products (Model RefSeqs)

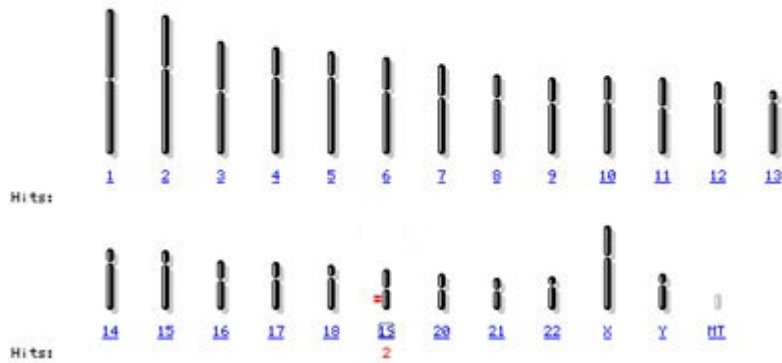
### Analysis:

transcript, protein, annotation, gene index

# Increasing Discovery Space



## NCBI Map Viewer



Gene centered  
Integrated data  
RefSeq <-> LocusLink  
Navigation

CGCTCAGGAT ACGCTCAGGAT ACGCTCAGGAT ACGCTCAGGAT ACGCTCAGGAT ACGCTCAGGAT ACGCTCAGGAT ACGCTCAGGAT ACGCTCAGGAT ACGCTCAGGAT  
TTCTCTATATATTCTC  
CCCCATCAGCACCCCA  
CACAGACTGCACAG

PubMed Entrez BLAST OMIM Taxonomy

Link Display: Brief Organism: Human Go Clear

LocusLink Home

A B C D E F G H I J K L M N O P Q R S T U V

17 loci found

LocusID	Org	Symbol	Description	Position	Links
6899	Hs	TBX1	T-box 1	22q11.21	PORGHUV
6910	Hs	TBX10	T-box 10	11q13	PORGHUV
6913	Hs	TBX15	T-box 15	1p13	PORGHUV
9096	Hs	TBX2	T-box 2	17q23	PORGHUV
9095	Hs	TBX21	T-box 21	17	PORGHUV

PubMed

OMIM

RefSeq

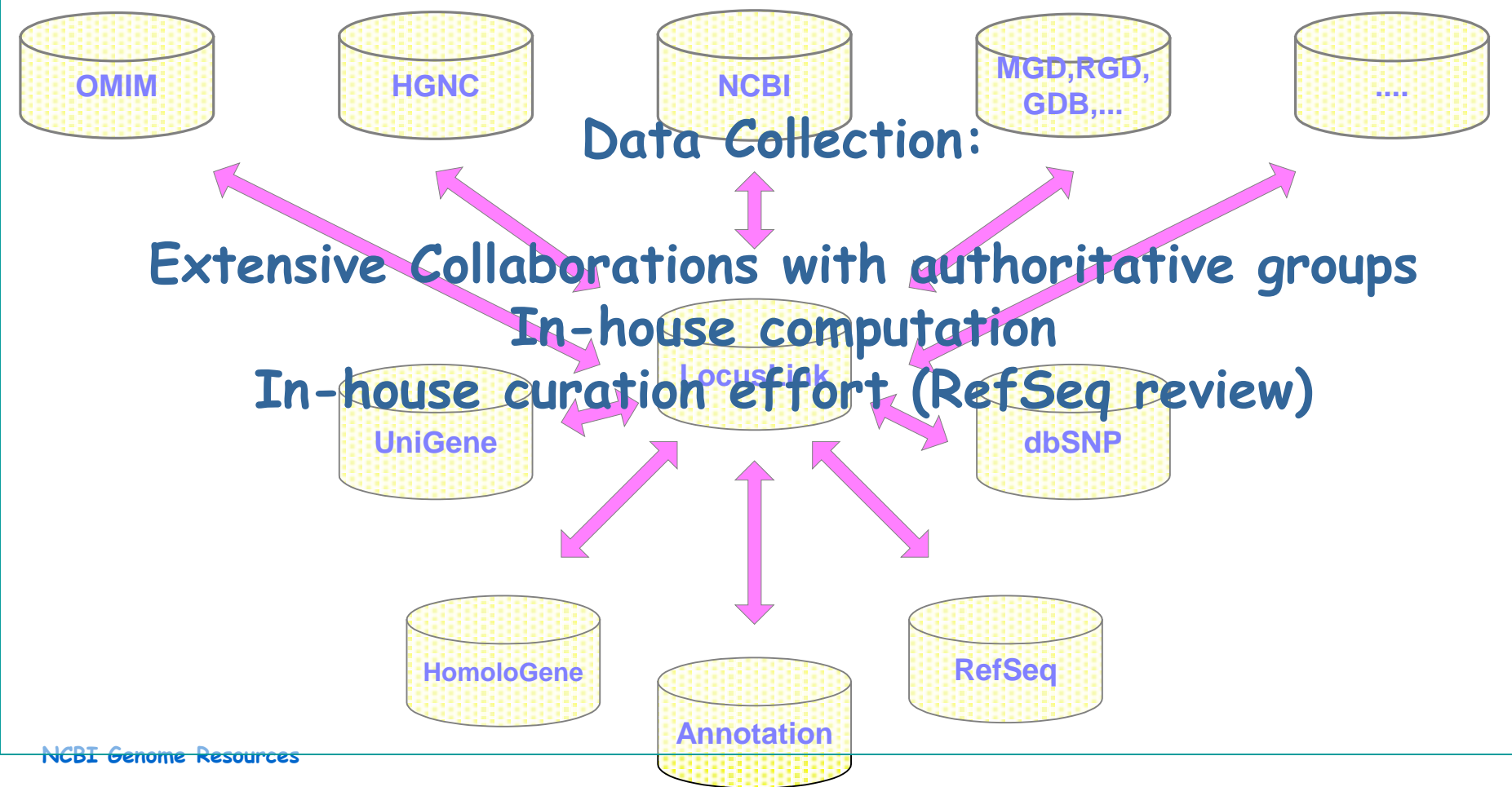
GenBank

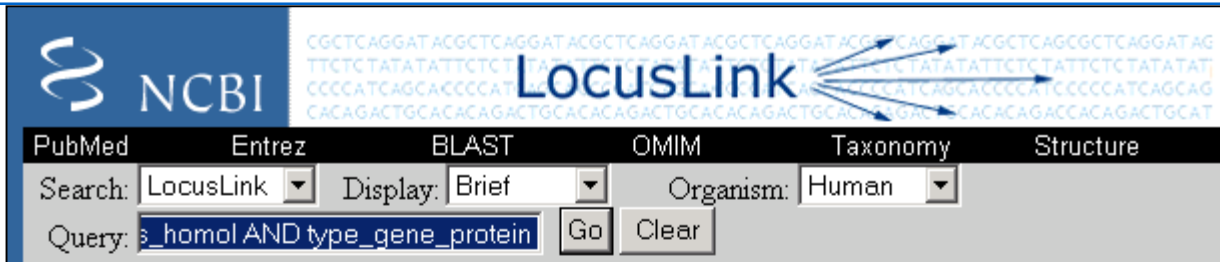
HomoloGene

Scope:  
Human  
Mouse  
Rat  
Fly  
Zebrafish  
HIV-1

# LocusLink: Maintenance

MIM numbers phenotypes Sequence links New genes	Symbols Descriptions Sequence links New Genes	Protein names Database links Sequence links New Genes	Symbols Descriptions Sequence links New Genes	Collaborators (Sequence links) More genomes
--	--	--	--	---





Find novel uncharacterized genes on a finished chromosome

QUERY= 21[chr] NOT has\_omim AND has\_homol AND type\_gene\_protein

AND predicted

AND model

AND provisional

AND C21orf\* OR MGC\*

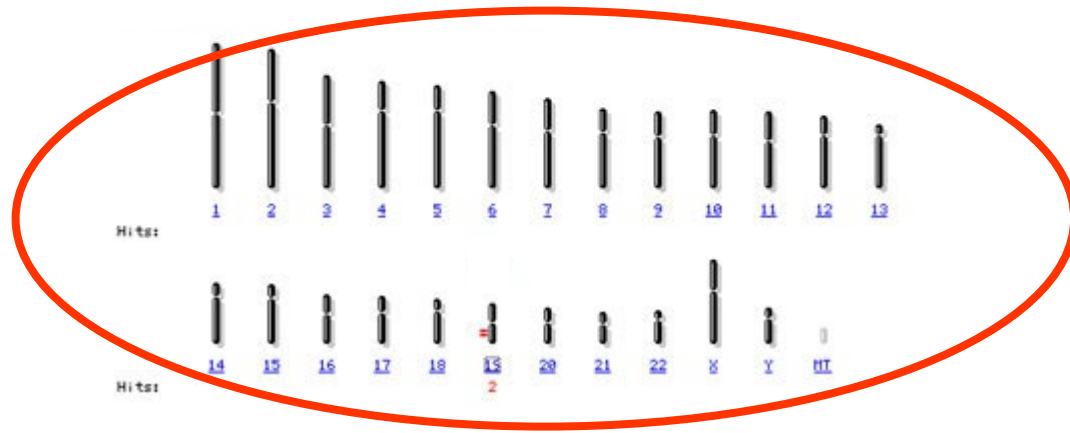
<input type="checkbox"/>	54094	<i>Hs</i>	C21orf15	reading frame 11 chromosome 21 open	21q11	P	GPHU
<input type="checkbox"/>	54093	<i>Hs</i>	C21orf18	reading frame 15 chromosome 21 open	21q22.13	P	RGPHUV
<input type="checkbox"/>	54090	<i>Hs</i>	C21orf21	reading frame 18 chromosome 21 open	21q22.3	P	G HU
<input type="checkbox"/>	54089	<i>Hs</i>	C21orf22	reading frame 21 chromosome 21 open	21q22.3	P	G HU
<input type="checkbox"/>	54087	<i>Hs</i>	C21orf24	reading frame 22 chromosome 21 open	21q22.2	P	G HU
				reading frame 24			



# Increasing Discovery Space



## NCBI Map Viewer



## What?

Genome Assembly

Genome Annotation

Integrated map data (genetic, cytogenetic, RH)

## Scope?

Human

Drosophila

Mouse

(model genomes)

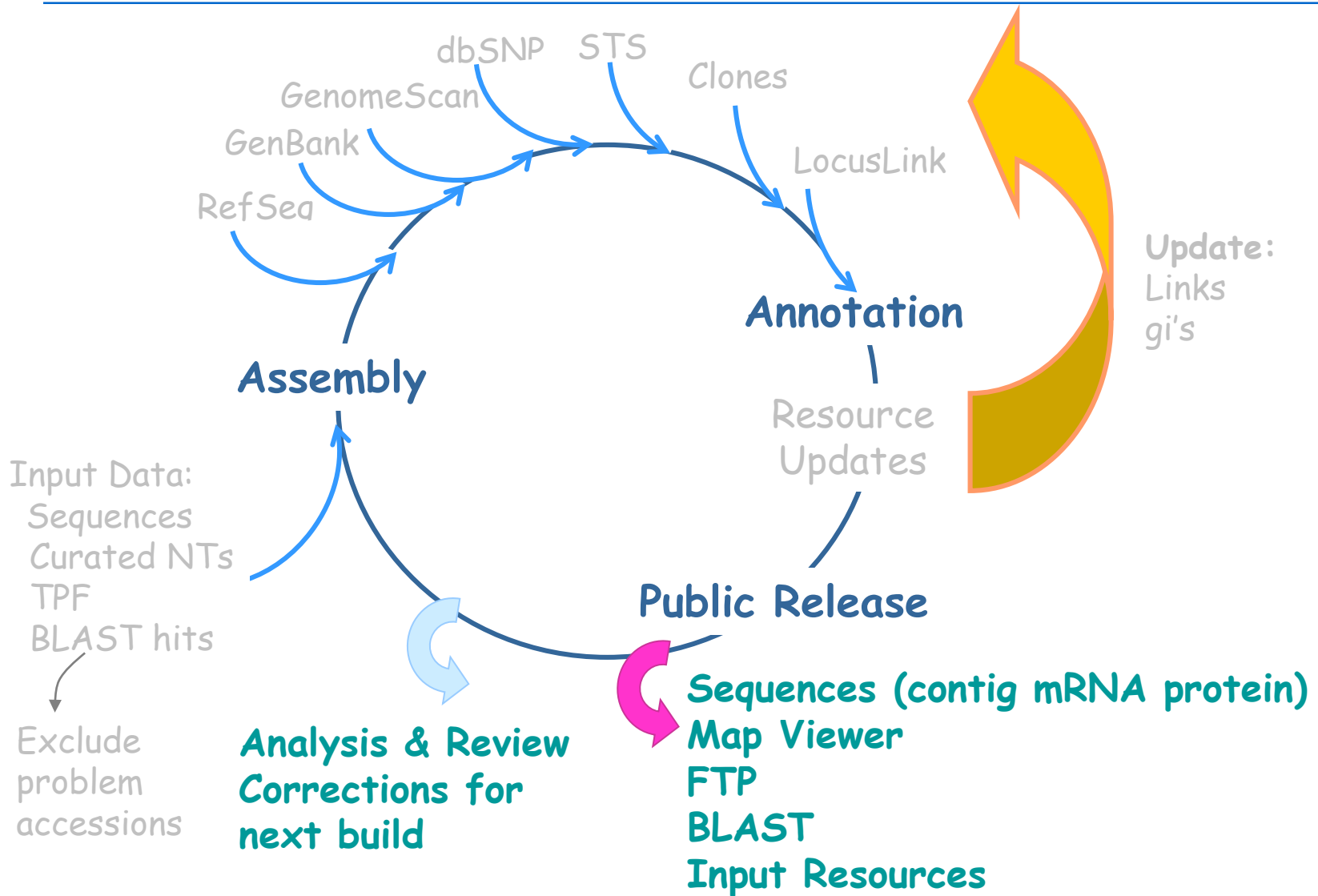
## Why?

Facilitate discovery (genes, variation ...)

Facilitate navigation

Facilitate use of genomic sequence information

# Genome Build Process



# RefSeq: a reagent for Annotation



RefSeq mRNAs



GenBank mRNAs



ESTs



TBLASTN



RPSBLAST



GenomeScan



## Potential Problems:

- Gene Families
- Partial sequences
- Chimeric
- Intron read-through
- Linker
- Vector
- Wrong organism

## RefSeq Advantages:

- Separate Gene Families
- Not Partial
- Means to correct problem sequences

## Quality Control:

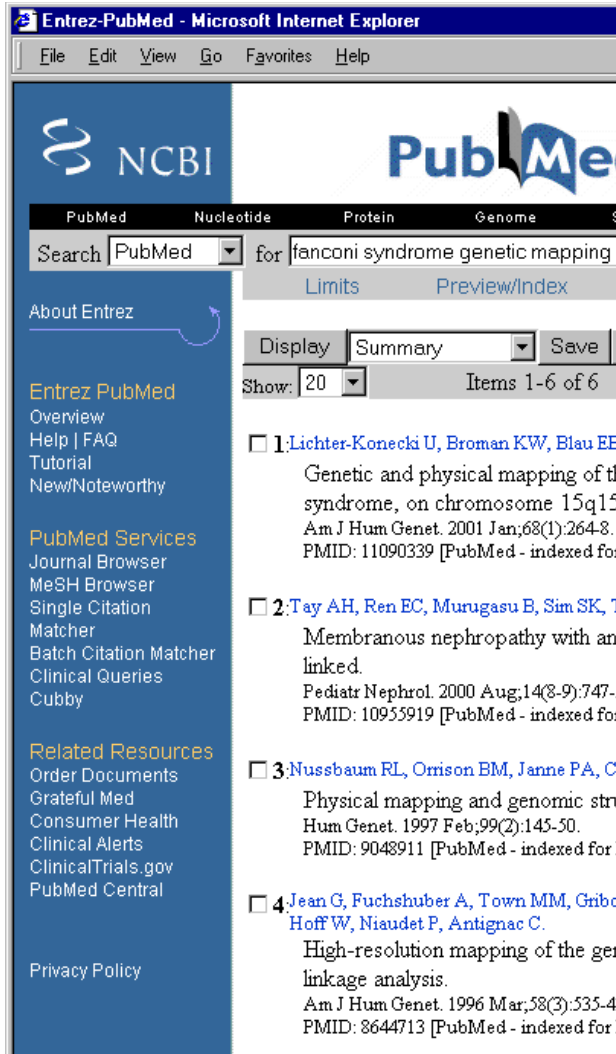
RefSeq review results in excluding problem GenBank sequences from annotation pipeline

# What questions can be asked?

---

- What genes (markers, SNPs) are between 2 markers?
- What BAC clones are available on Xq28?
- Where are there serine kinases?
- I've cloned gene xyz in my favorite organism. What is related in human?
- What is the evidence that there is a gene at position n?
- I have found a phenotype of interest between markers x and y; what is known about this region?

## Pathology in proximal renal tubular transport



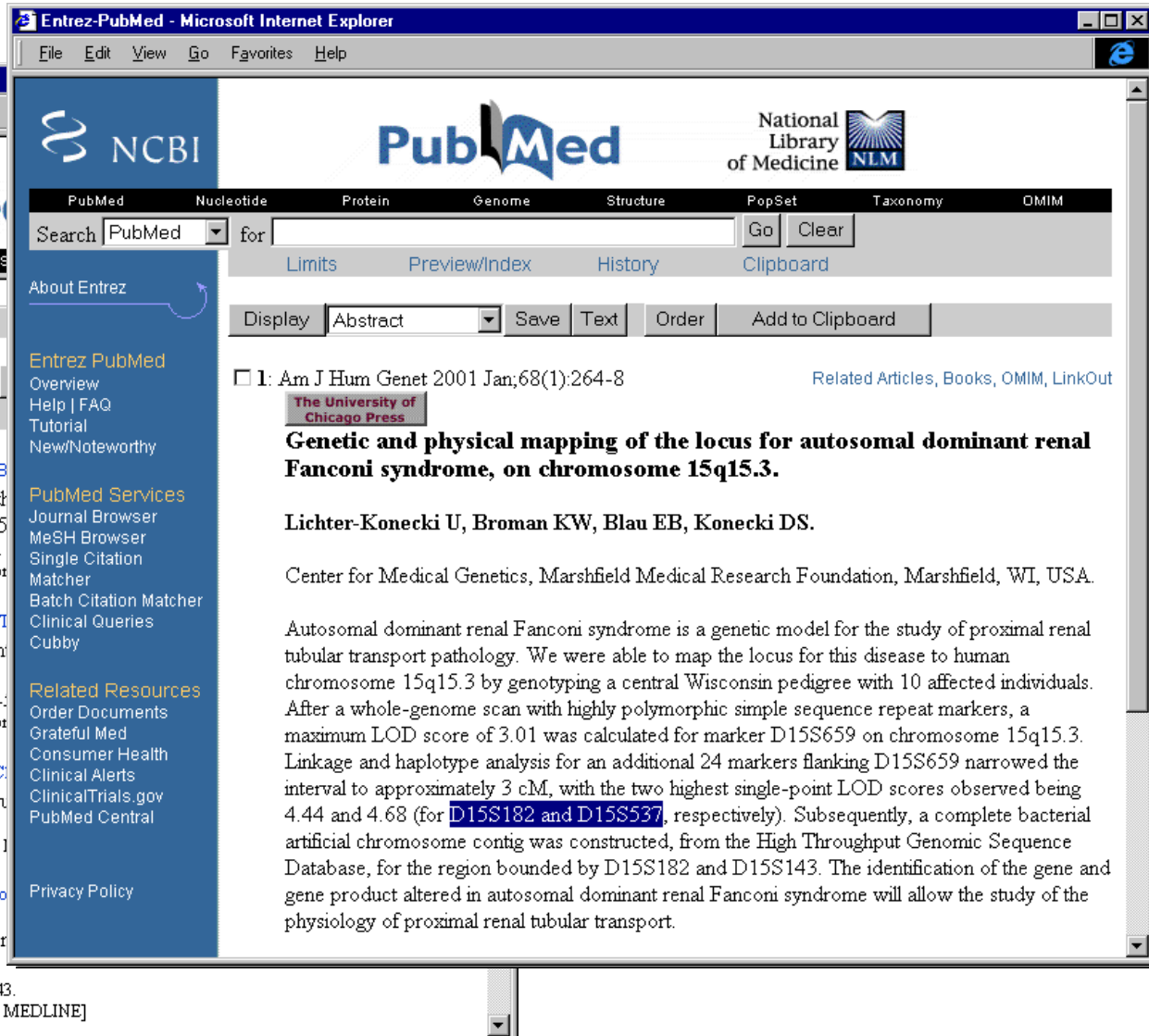
Entrez-PubMed - Microsoft Internet Explorer

Search PubMed for fanconi syndrome genetic mapping

Display Summary Save

Show: 20 Items 1-6 of 6

- 1: **Lichter-Konecki U, Broman KW, Blau EB**  
Genetic and physical mapping of the syndrome, on chromosome 15q15.3. *Am J Hum Genet.* 2001 Jan;68(1):264-8. PMID: 11090339 [PubMed - indexed for MEDLINE]
- 2: **Tay AH, Ren EC, Murugasu B, Sim SK, Iyer R**  
Membranous nephropathy with an unusual immunofluorescence pattern. *Pediatr Nephrol.* 2000 Aug;14(8-9):747-50. PMID: 10935919 [PubMed - indexed for MEDLINE]
- 3: **Nussbaum RL, Orrison BM, Janne PA, Cawthon JM**  
Physical mapping and genomic structure of the Fanconi anemia gene. *Hum Genet.* 1997 Feb;99(2):145-50. PMID: 9048911 [PubMed - indexed for MEDLINE]
- 4: **Jean G, Fuchshuber A, Town MM, Gribouval L, Hoff W, Niaudet P, Antignac C**  
High-resolution mapping of the gene for Fanconi anemia. *Am J Hum Genet.* 1996 Mar;58(3):535-43. PMID: 8644713 [PubMed - indexed for MEDLINE]



Entrez-PubMed - Microsoft Internet Explorer

Search PubMed for fanconi syndrome genetic mapping

Display Abstract Save Text Order Add to Clipboard

1: **Am J Hum Genet** 2001 Jan;68(1):264-8  
The University of Chicago Press  
**Genetic and physical mapping of the locus for autosomal dominant renal Fanconi syndrome, on chromosome 15q15.3.**  
**Lichter-Konecki U, Broman KW, Blau EB, Konecki DS.**  
Center for Medical Genetics, Marshfield Medical Research Foundation, Marshfield, WI, USA.  
Autosomal dominant renal Fanconi syndrome is a genetic model for the study of proximal renal tubular transport pathology. We were able to map the locus for this disease to human chromosome 15q15.3 by genotyping a central Wisconsin pedigree with 10 affected individuals. After a whole-genome scan with highly polymorphic simple sequence repeat markers, a maximum LOD score of 3.01 was calculated for marker D15S659 on chromosome 15q15.3. Linkage and haplotype analysis for an additional 24 markers flanking D15S659 narrowed the interval to approximately 3 cM, with the two highest single-point LOD scores observed being 4.44 and 4.68 (for **D15S182** and **D15S537**, respectively). Subsequently, a complete bacterial artificial chromosome contig was constructed, from the High Throughput Genomic Sequence Database, for the region bounded by D15S182 and D15S143. The identification of the gene and gene product altered in autosomal dominant renal Fanconi syndrome will allow the study of the physiology of proximal renal tubular transport.



# Query Map

## Look for genes in the region

Entrez Genome view - Microsoft Internet Explorer

File Edit View Go Favorites Help

NCBI

gJfC TNPL cgnR ushR

4500K 250K

PubMed Nucleotide Protein Genome Structure PopSet Taxonomy

Search for  on chromosome(s)  Find

Show linked entries Help FTP

### Homo sapiens genome view build 24

Hits:

1 2 3 4 5 6 7 8 9 10 11 12 13

14 15 16 17 18 19 20 21 22 X Y MT

Search results for query "D15S182 OR D15S537": 2 hits

Chr	Match	Map element	Type	Maps
15	D15S182	UT477	sts	Marshfield   STS
15	UT7098	UT7098	sts	Marshfield   STS

Entrez Map View - Microsoft Internet Explorer

File Edit View Go Favorites Help

NCBI

gJfC TNPL cgnR ushR

4500K 250K

PubMed Entrez BLAST OMIM Taxonomy Structure

Search  Find in This View Find

Advanced Search

Map Viewer Help  
Human Maps Help  
FTP  
Chr. 15 Resource

### Homo sapiens Map View build 24

Chromosome: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 [15] 16 17 18 19 20 21 22 X Y

Query: D15S182 OR D15S537 [clear]

Master: STS Map [Display settings](#)

Total STSs On Chromosome: 3089 [23 not localized]  
Region Displayed: 33,325K-34,326K bp [Download/View Sequence](#)  
STSs Labeled: 20 Total STSs in Region: 31

Genes\_seq  Marshfield  STS  marker Kbp

SORD	33502
D15S516	33502
stSG51836	33558
D15S799	33572
mp0554	33572
sts-S82300	33575
GDB:454836	33575
WI-11756	33623
SHGC-11064	33668
SGC33828	33716
stSG4054	33726
SHGC-13414	33727
SHGC-30973	33755
D15S537	33766
sts-AA018839	33797
D15S182	33845
RH27235	33874
SHGC-6160	33993
SHGC-79674	34083
SHGC-5901	34228

symb: HSPC129  
 orient.: -  
 links: sv  
 : ev  
 cyto.: 15q11.2  
 full name: hypothetical protein

RY43E11  
 AFM070x47  
 AFM0519c1  
 AFM002x49  
 AFM0153uF1  
 UT7098  
 AFM0123xx5  
 UT477



**Blast Result - Microsoft Internet Explorer**

File Edit View Go Favorites Help

---

**NCBI Blast 2 Sequences results**

PubMed Entrez **BLAST** OMLM Taxonomy Structure

**BLAST 2 SEQUENCES RESULTS VERSION BLASTP 2.1.2 [Oct-19-2000]**

Matrix:  gap open:  gap extension:

x\_dropoff:  expect:  wordsize:   Filter

---

**Sequence 1** gi:7705461 Length 466 (1..466)

**Sequence 2** gi:6323047 Length 397 (1..397)

**NOTE:** The statistics (bitscore and expect value) is calculated based on the size of nr database

Score = 159 bits (399), Expect = 5e-38  
 Identities = 79/153 (51%), Positives = 104/153 (67%), Gaps = 1/153 (0%)

```

Query: 290 LVLDDLDETLVHCSLNELEDAALTFPVLFQDVVIYQVYVRLRPFREFLERMSQMYEILFT 349
      L+LDLDETLVH S + A PV D ++ VYV RP EFL R+SQ+YE+++FT
Sbjct: 230 LILDDLDETLVHSSFKYMHSADFVLPVEIDDQVHNVYVIKRPGVDFLNRVSQLYEVVFT 289

Query: 350 ASKKVYADKLLNILDPKKQLVRHRLFREHCVCVQGNYYIKDLNILGRDLSKTIIDNSPQA 409
      AS YA+ LL+ LDP + HRLFRE C +GNYIK+L+ +GR LS+TII+DNSP +
Sbjct: 290 ASVSRYANPLLDTLDPNGT-IHHRLFREACYNYYEGNYIKNLSQIGRPLSETIILDNSPAS 348

Query: 410 FAYQLSNGIPIESWFMKNDNELLKLIPIFLEKL 442
      + + + +PI SWF D +DNELL +IP LE L
Sbjct: 349 YIFHPQHAVPISSWFS DTHDNELLDIIPLEDL 381
  
```

NCBI Sequence Viewer - Microsoft Internet Explorer

gil7705461 view - Microsoft Internet Explorer

File Edit View Go Favorites Help

**BLAST**

Query: gi:7705461 hypothetical protein  
 Matching gi: 6841480  
 Lineage: Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Primates; Hominidae; Homo

Best hits Common hits

116 BLAST hits to 14 unique databases

0 Archaea 0 Bacteria

Keep only

466 aa

Microsoft Internet Explorer

ony Help

etical proteins [Caenorhabditis elegans]

Scnaromyces cerevisiae]

rotein YLR019w - y6us]

myces cerevisiae] us]

factor isoform T

human

erevisiae)

racting factor [Gallus gallus]

erevisiae)

ns]

nal





PubMed

PubMed Nucleotide Protein Gene

Search PubMed for Psr2p

Limits Preview/

Abc

## Psr2p - sodium stre

### Entrez PubMed

Overview  
Help | FAQ  
Tutorial  
New/Noteworthy

### PubMed Services

Journal Browser  
MeSH Browser  
Single Citation  
Matcher  
Batch Citation Matcher  
Clinical Queries  
Cubby

### Related Resources

Order Documents  
Grateful Med  
Consumer Health  
Clinical Alerts  
ClinicalTrials.gov  
PubMed Central

Privacy Policy

1. Siniossoglou S, Hurt EC, Pelham HR.

Psr1p/Psr2p, two plasma membrane phosphatases with an essential DXDX(T/V) motif required for sodium stress response in yeast. *J Biol Chem.* 2000 Jun 23;275(25):19352-60. PMID: 10777497 [PubMed]



PubMed



PubMed Nucleotide Protein Genome Structure PopSet Taxonomy OMIM

Search PubMed for

Limits Preview/Index History Clipboard

Display Abstract Save Text Order Add to Clipboard

1. *J Biol Chem* 2000 Jun 23;275(25):19352-60

[Related Articles, Books, Links](#)

FREE full text article at  
[www.jbc.org](http://www.jbc.org)

### Psr1p/Psr2p, two plasma membrane phosphatases with an essential DXDX(T/V) motif required for sodium stress response in yeast.

Siniossoglou S, Hurt EC, Pelham HR.

Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, United Kingdom.

Regulation of intracellular ion concentration is an essential function of all cells. In this study, we report the identification of two previously uncharacterized genes, PSR1 and PSR2, that perform an essential function under conditions of sodium ion stress in the yeast *Saccharomyces cerevisiae*. Psr1p and Psr2p are highly homologous and were identified through their homology with the endoplasmic reticulum membrane protein Nem1p. Localization and biochemical fractionation studies show that Psr1p is associated with the plasma membrane via a short amino-terminal sequence also present in Psr2p. Growth of the psr1psr2 mutant is severely inhibited under conditions of sodium but not potassium ion or sorbitol stress. This growth defect is due to the inability of the psr1psr2 mutant to properly induce transcription of ENA1/PMR2, the major sodium extrusion pump of yeast cells. We provide genetic evidence that this regulation is independent of the phosphatase calcineurin, previously implicated in the sodium stress response in yeast. We show that Psr1p contains a DXDX(T/V) phosphatase motif essential for its function in vivo and that a Psr1p-PtA fusion purified from yeast extract exhibits phosphatase activity. Based on these data, we suggest that Psr1p/Psr2p, members of an emerging class of eukaryotic phosphatases, are novel regulators of salt stress response in yeast.

### About Entrez

### Entrez PubMed

Overview  
Help | FAQ  
Tutorial  
New/Noteworthy

### PubMed Services

Journal Browser  
MeSH Browser  
Single Citation  
Matcher  
Batch Citation Matcher  
Clinical Queries  
Cubby

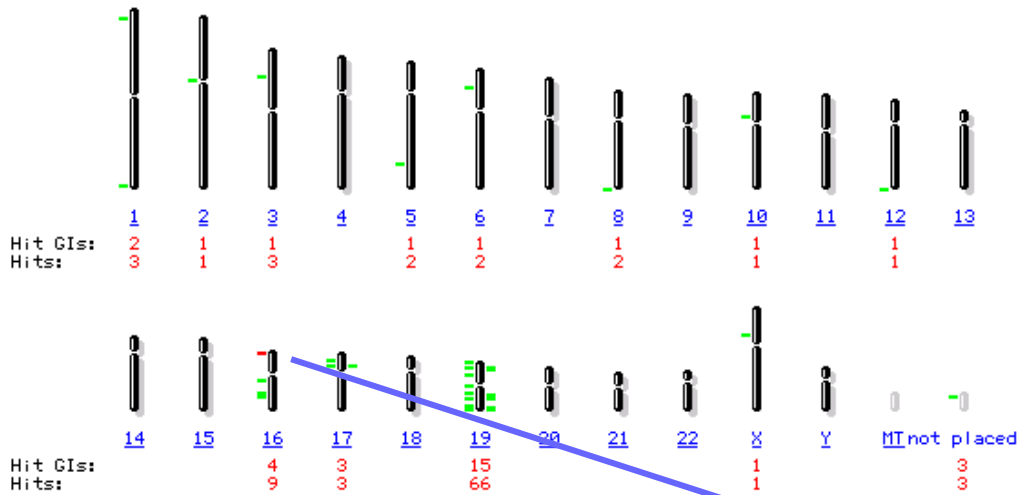
### Related Resources

Order Documents  
Grateful Med  
Consumer Health  
Clinical Alerts  
ClinicalTrials.gov  
PubMed Central

Privacy Policy

[Homo sapiens genome view build 25](#)

[BLAST search the human genome](#)



Result from a BLAST query of a zinc finger protein

Color key for scores: < 40 40-50 50-80 80-200 >= 200 [Back to BLAST alignments page](#)

**BLAST search results: 35 BLAST hits found**  
 g|4508012|ref|NM\_003454.1| Homo sapiens zinc finger protein 200 (ZNF200), mRNA

[Sort results by score](#)

Chr	Hit GI	Hits	Score	E value	Map element
1	<a href="#">15297148</a>	2	<a href="#">54.0</a>	3e-04	<a href="#">NT_004536</a> Homo sapiens chromosome 1 working draft sequence segment
1	<a href="#">15296234</a>	1	<a href="#">50.1</a>	0.005	<a href="#">NT_004359</a> Homo sapiens chromosome 1 working draft sequence segment
2	<a href="#">15294983</a>	1	<a href="#">60.0</a>	5e-06	<a href="#">NT_022300</a> Homo sapiens chromosome 2 working draft sequence segment
3	<a href="#">15296047</a>	3	<a href="#">54.0</a>	3e-04	<a href="#">NT_005563</a> Homo sapiens chromosome 3 working draft sequence segment
5	<a href="#">15297054</a>	2	<a href="#">56.0</a>	8e-05	<a href="#">NT_006951</a> Homo sapiens chromosome 5 working draft sequence segment
6	<a href="#">15305200</a>	2	<a href="#">50.1</a>	0.005	<a href="#">NT_007592</a> Homo sapiens chromosome 6 working draft sequence segment
8	<a href="#">15299853</a>	2	<a href="#">52.0</a>	0.001	<a href="#">NT_007997</a> Homo sapiens chromosome 8 working draft sequence segment
10	<a href="#">15298811</a>	1	<a href="#">58.0</a>	2e-05	<a href="#">NT_008583</a> Homo sapiens chromosome 10 working draft sequence segment
12	<a href="#">15302779</a>	1	<a href="#">54.0</a>	3e-04	<a href="#">NT_009455</a> Homo sapiens chromosome 12 working draft sequence segment
16	<a href="#">15318559</a>	5	<a href="#">2680</a>	0.0	<a href="#">NT_015360</a> Homo sapiens chromosome 16 working draft sequence segment

Best match

# Review the alignment

Map Viewer Help  
 Human Maps Help  
 FTP  
 Chr. 16 Resource

**Data As Table View** UPDATED

Region Shown:

out  
 zoom  
 in

Contig

Color Key for Alignment Scores: <40 40-50 50-80 80-200 >=200

**Master: Contig Map** [Display settings](#)

Total Contigs On Chromosome: 110 [12 not localized]  
 Region Displayed: 3,829K-3,844K bp [Download/View Sequence/Evidence](#) UPDATED  
 Contigs Labeled: 6 Total Contigs in Region: 6

[Genes\\_seq](#)  [Contig](#)  accession orient.

3830K  
 3831K **Blast hit** Identity=100% query: 2053..702  
 3832K  
 3833K  
 3834K  
 3835K  
 3836K **ZNF200** [NT\\_015360.5](#)  
 3837K  
 3838K  
 3839K  
 3840K **Blast hit** Identity=100% query: 702..574  
**Blast hit** Identity=100% query: 576..485  
 3841K **Blast hit** Identity=100% query: 486..154  
 3842K **Blast hit** Identity=100% query: 155..1  
 3843K

[Disclaimer](#) | [Write to the Help Desk](#)

## A click away:

- Alignments (BLAST hit)
- Gene Description
- Report of all features in the region
- Sequence in the region
- other mRNAs aligning in the region
- Homology Maps
- Model Maker - Define your own gene model based on alignments in the region

## Genes in regions of conserved synteny

●	4p12	TXK	5	●	Txk	40	●
●	4p12	TEC	5	●	Tec	41	
●	4	KIAA1458	5	●	5033405K12Rik		
●	4	KIAA0826	5	●	2310004H21Rik		
●	4q12	SGCB	5	●	Sgcb		●
●	4	DKFZP586K0717	5	●	1300019H17Rik		
●	4q11	CHIC2	5	●	Chic2		
●	4q11-q13	PDGFRA	5	●	Pdgfra	42	●
●	4q11-q12	KIT	5	●	Kit	42	●
●	4q11-q12	KDR	5	●	Kdr	42	●
●	4	LOC55858	5	●	Tpar1	42	●
●	4q12	CLOCK	5	●	Clock	43	●
●	4q12-q13.3	SEC3	5	●	2810407P21Rik		
●	4q12	IGFBP7	5	●	Igfbp7		●
●	4q12-q13.3	REST	5	●	Rest		●
●	4q11-q12	SMAP31 *	5	●	1110018K11Rik		●
●	4	LEC3	5	●	5430402I23Rik		
●	4	EPHA5	5	●	Epha5		●
●	4q13.3	BRDG1	5	●	Brdg1-pending		

Anchored by  
human gene  
order

●	40	Txk	4	●	TXK	4p12	●
●	41	Cnecg	4	●	CNGA1	4p12-cen	●
●	41	Tec	4	●	TEC	4p12	●
●	42	Pdgfra	4	●	PDGFRA	4q11-q13	●
●	42	Kit	4	●	KIT	4q11-q12	●
●	42	Kdr	4	●	KDR	4q11-q12	●
●	42	Tpar1	4	●	LOC55858	4	●
●	42	Ubc2n-ps1	12	●	UBE2N	12	●
●	43	Clock	4	●	CLOCK	4q12	●
●	44	Gnrhr	4	●	GNRHR	4q21.2	●
●	44	Ste	4	●	STE	4q13.1	●
●	44	Gc	4	●	GC	4q12-q13	●
●	44.9	Csna	4	●	CSN1	4q21.1	●
●	45	Csnb	4	●	CSN2	4q21.1	
●	45	Csnk	4	●	CSN10	4q21.1	

Anchored by  
mouse gene  
order

# Map Viewer: Model Maker

Mode  
 exon "se  
 in your m

Evidence  
 207152

Putative e  
 1

Your mo

Putative e  
 1 20  
 2  
 3  
 4

Putative exons (graphic view):

Your model:

clear  
 BG714075.1  
 AA203194.1  
 XM\_086586

AGGCGCAGGCGCAGGCGAGGGGCTGGGTGGCGGTTGAGACAGCGGGCGGTACTGGGAGGGC  
 TAGGTGAGGGTGCAGAGGCTGCCCGAGCTTCTGAGCGAGCGCGGTGCTTTTGGGAACGG  
 GGACGGGCGATCTGCGGCGCCAGGAGCTGGGCCGAGGCGCGGGCGGGCTGCCGGCTG  
 CCCTGTGAATGGGAAGTTACGCGAAGTCCACCCAGCGTTTCTGAGGCAATCTGAAGGCAA

ORF Finder  
 Save

Frame1, ORF=56      Frame2, ORF=59      Frame3, ORF=112

Frame1,  
 rrrrrrgagwrlrqrrywea  
 \*VRVARLPELLSERGAFGNA  
 GRAICGARSWAEARRRGRL  
 PCEWEVTRSPPSVSEAI\*rq

gagageglggg\*dsggtgrr  
 r\*gsrgcpsf\*ASAVLLGTR  
 DGRSAAPGAGPRRGAAAGC  
 PVNGKLREVHPAFLRQSEGK

aqaqargwvavetaavlggv  
 gegreaaraserarcfwerg  
 tgdllrrqelgrgaaarlpaa  
 l\*mgisyakstqrf\*gnlkan

Putative exons (table view):

Exon	Start	End	Start Codon	Stop Codon	ORF
<input checked="" type="checkbox"/> 1	207152	GC AGG...GC GAG	207056-206928	GAG GT =>	3 or 4
<input type="checkbox"/> 2		CA GGA	207010-206984	GCG GC =>	4
<input type="checkbox"/> 3		1 <=AG GCA	202369-202114	GAG CT	
<input checked="" type="checkbox"/> 4		1 <=AG GCA	202369-202098	CAG GT =>	6 or 10 or 18
<input type="checkbox"/> 5		GC ACT	202259-202098	CAG GT =>	18

Make your own gene model

# Increasing Discovery Space

Entrez

GenBank

PubMed



Expression (GEO)



UniGene

Clones (Clone Registry)

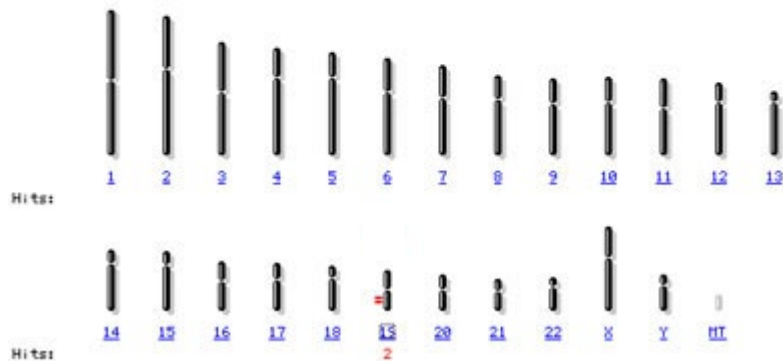
HomoloGene

NCBI Map Viewer

Markers (UniSTS)

OMIM

Variation (dbSNP)



Homology Maps



# Acknowledgments

RefSeq Curator Staff  
BLAST Team  
Entrez Team  
NCBI Service Desk Staff

## **Collaborators:**

Human Gene Nomenclature Committee  
OMIM Staff  
The Jackson Laboratory  
Rat Genome Database

<http://www.ncbi.nlm.nih.gov/>

## **Genome Build Team:**

Richa Agarwala  
Hsiu-Chuan Chen  
Slava Chetvernin  
Deanna Church  
Olga Ermolaeva  
Renata Geer  
Wratko Hlavina  
Wonhee Jang  
Jonathan Kans  
Ken Katz  
Paul Kitts  
David Lipman  
Adam Lowe  
Donna Maglott  
Jim Ostell  
Kim Pruitt  
Sergey Resenchuk  
Victor Sapojnikov  
Greg Schuler  
Steve Sherry  
Andrei Shkeda  
Tatiana Tatusova  
Lukas Wagner  
Sarah Wheelan