

Osservando la Genoma con Ensembl



Ensembl release 42 - Dec 2006 HOME · BLAST · BIOMART · SITEMAP **HELP**

Your Ensembl

- Show account · Log out
- Save bookmark

Help & Documentation

- Table of Contents
- Helpdesk
- What's New
- About Ensembl
- Downloading data
- Displaying your own data
- Ensembl software

Select a species

- Mammals
- Other chordates
- Other eukaryotes

Ensembl Archive

- View previous release of page in Archive!
- Stable Archive! link for this page

Ensembl tools

- Start a sequence search** →
Search Ensembl for nucleotide and peptide sequences with BLAST and SSAHA.
- Mine Ensembl with BioMart** →
Cross-reference Ensembl datasets with BioMart, a powerful data-mining tool.
- Customise Your Ensembl** →
Register with Ensembl to bookmark your favourite pages, customise your home page and much more!
- Fetch data with the Ensembl API** →
Learn how to extract data from the public Ensembl database with this tutorial.

You are logged in as **Giulietta Spudich**: [Account home page](#) · [Log out](#)

Ensembl headlines: Release 42 (December 2006)

- New - User accounts** (all species)
- New species - Duck-billed Platypus** (*Ornithorhynchus anatinus*)

Search Ensembl

Search: for

e.g. mouse chromosome 2 or X:10000..20000 or human gene BRCA2

Ensembl 42 Pre! species

Popular genomes · [Reorder](#)

- Homo sapiens**
NCBI 36 | Vega
- Mus musculus**
NCBI m36 | Vega
- Danio rerio**
Zv6 | Vega

More genomes · [Reorder](#)

- ▶ **Aedes aegypti** AaegL1
- ▶ **Anopheles gambiae** Ag amP3
- ▶ **Bos taurus** Btau 2.0
- ▶ **Caenorhabditis elegans** WS160
- ▶ **Canis familiaris** CanFam 2.0 **UPDATED!**
- ▶ **Ciona intestinalis** JGI 2
- ▶ **Ciona savignyi** CSAV 2.0
- ▶ **Dasylops novemcinctus** ARMA
- ▶ **Drosophila melanogaster** BDGP 4.3
- ▶ **Echinops telfairi** TENRF 0

Dr. Giulietta Spudich
Ensembl User Support

Ordine del Giorno

Polaris, Febbraio 2007

- **Introduzione**
- **Un ‘Worked Example’**
- **Parte Pratica**

- **Lettura del Programma ‘BioMart’**
- **Esercizi ed Applicazioni (Parte Pratica)**

- **Presentazione (a Scelta): ‘Variazioni’**

- **Parte Pratica e Domande**

Esplorando le Genome

Con i Browser di genome possiamo:

- **Sfogliare i geni nel contesto genomico**
- **Mostrare I 'features' vicino un gene**
- **Esplorare una regione genoma**
- **Trovare informazione di un gene oppure tanti geni**
- **Osservare l'ordine della genoma**
- **Fare un paragone tra le genome**

Annotazione in Ensembl: di Base

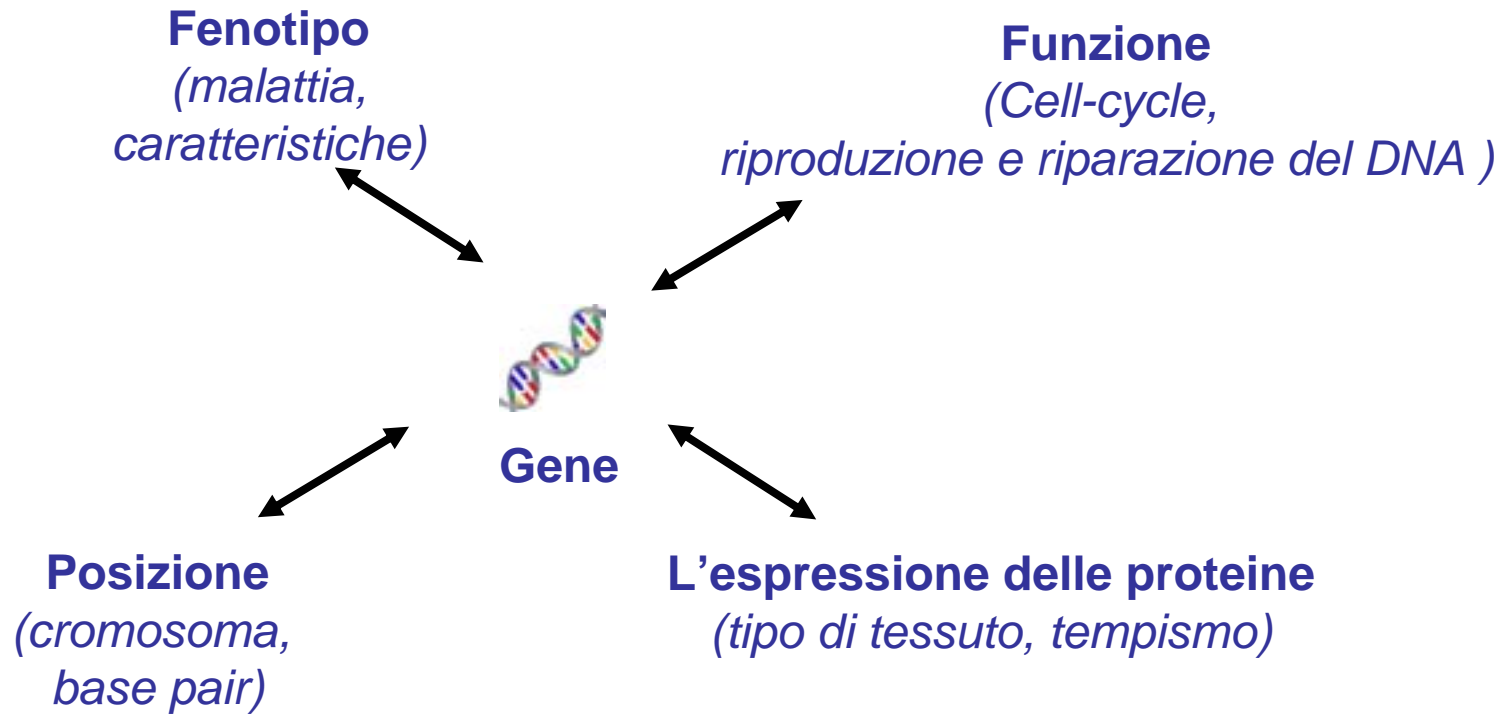
Geni

- **Posizione nelle genome**
- **Strutture**
 - **Exons**
 - **Introns**
 - **UTRs (Untranslated Regions)**
- **Transcript(s)**
 - **Pseudogenes**
 - **Non-coding RNA**
- **Proteine**

Annotazione in Ensembl: di Piu

- **Le bande citogenetiche**
- **Polymorphic markers**
- **Variazioni genetiche**
- **Sequenze che si ripetano (repeats)**
- **Expressed Sequence Tags (ESTs)**
- **cDNAs o mRNAs dei speci vicini**
- **Regioni simili tra genome**

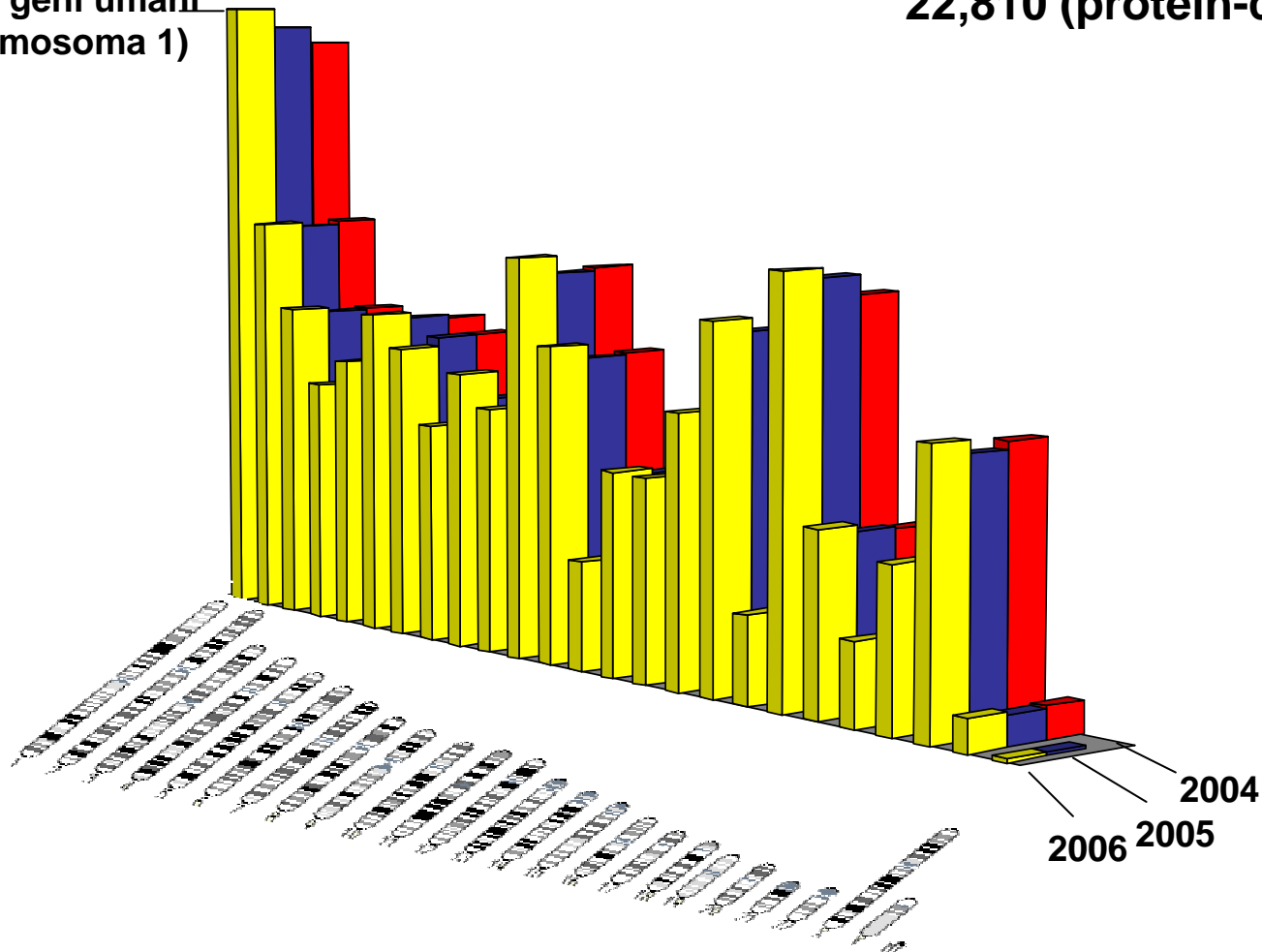
Diversi Modi di Vedere un Gene:



Una Specie Ha Tanti Geni

2,200 geni umani
(cromosoma 1)

Geni umani totale:
22,810 (protein-coding)



27 Genome con Annotazioni nel Ensembl



Anopheles gambiae
AgamP3



Gallus gallus
WASHUC2 **UPDATED!**



Macaca mulatta
MMUL 1.0



Aedes aegypti
AaegL1



Homo sapiens
NCBI 36 | Vega



Bos taurus
Btau 2.0 | **UPDATED!** *pre!*



Caenorhabditis elegans
WS160



Canis familiaris
CanFam 2.0 **UPDATED!** | Vega



Ciona intestinalis
JGI 2



Ciona savignyi
CSAV 2.0



Dasyus novemcinctus
ARMA



Drosophila melanogaster
BDGP 4.3



Echinops telfairi
TENREC



Loxodonta africana
BROAD E1



Gasterosteus aculeatus
BROAD S1



Ornithorhynchus anatinus
Dana-5.0 **NEW!**



Oryctolagus cuniculus
RABBIT



Monodelphis domestica
MonDom 4.0



Rattus norvegicus
RGSC 3.4



Saccharomyces cerevisiae
SGD1.01 **UPDATED!**



Takifugu rubripes
FUGU 4.0



Xenopus tropicalis
JGI 4.1



Tetraodon nigroviridis
TETRAODON 7



Pan troglodytes
PanTro 2.1 | **UPDATED!** *pre!*



Oryzias latipes
HdR



Mus musculus
NCBI m36 | Vega



Danio rerio
Zv6 | Vega

La maggioranza sono vertebrati come:



ci sono anche organismi interessanti per la ricerca, come:

Caenorhabditis elegans



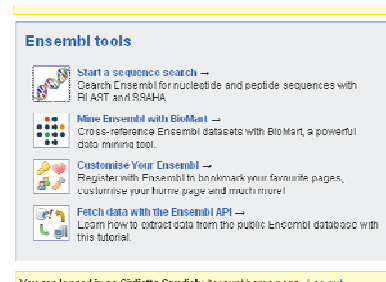
Saccharomyces cerevisiae



***Anopheles gambiae*
(malaria)**

Com'è Organizzata per Prendere l'Informazione?

- **Ensembl Views (Sito Web)**



- **Ensembl Database (gratuito)**
- **BioMart (Applicazione per predeare informazione del database facilmente)**

Ensembl

- **Fu fondata nel 1999**
- **Progetto tra EBI (Istituto Europeo d'Informatica) e il Centro di Sanger**
- **Finanziato principalmente dal Wellcome Trust, anche dal EMBL, NIH-NIAID and BBSRC**
- **Gruppo di circa 40 persone, i capi sono Ewan Birney (EBI) and Tim Hubbard (Sanger)**
- **Usa uno dei piu grandi sistemi di computer dedicato alla biologia in Europa!**

Geni nel Ensembl:

Tutti i geni (protein coding) hanno una base biologica.

- **UniProt/Swiss-Prot**

Un database 'manually curated'

- **NCBI RefSeq**

Un database 'manually curated' in parte

- **UniProt/TrEMBL**

'Translations' delle proteine da EMBL coding sequences (CDS)

Che Cos' Altro?

- **Pseudogeni (geni falsi)**
- **RNA genes (rRNA, tRNA, miRNA, etc)**
- **Informazioni degli altri 'database'**

Interpro



UniProt
Hinxton/Geneva/Washington



PRINTS
Manchester, UK



Gene3D
London, UK



PROSITE
SIB, Switzerland



SUPERFAMILY
Cambridge, UK



Pfam
Cambridge, UK



SCOP
Cambridge, UK



ProDom
Toulouse, France



CATH
London, UK



SMART
Oxford, UK



MSD
Hinxton, UK



TIGRFAMs
Rockville, USA



Swiss-Model
Basel, Switzerland



PANTHER
Foster City, USA



MODBASE
San Francisco, USA



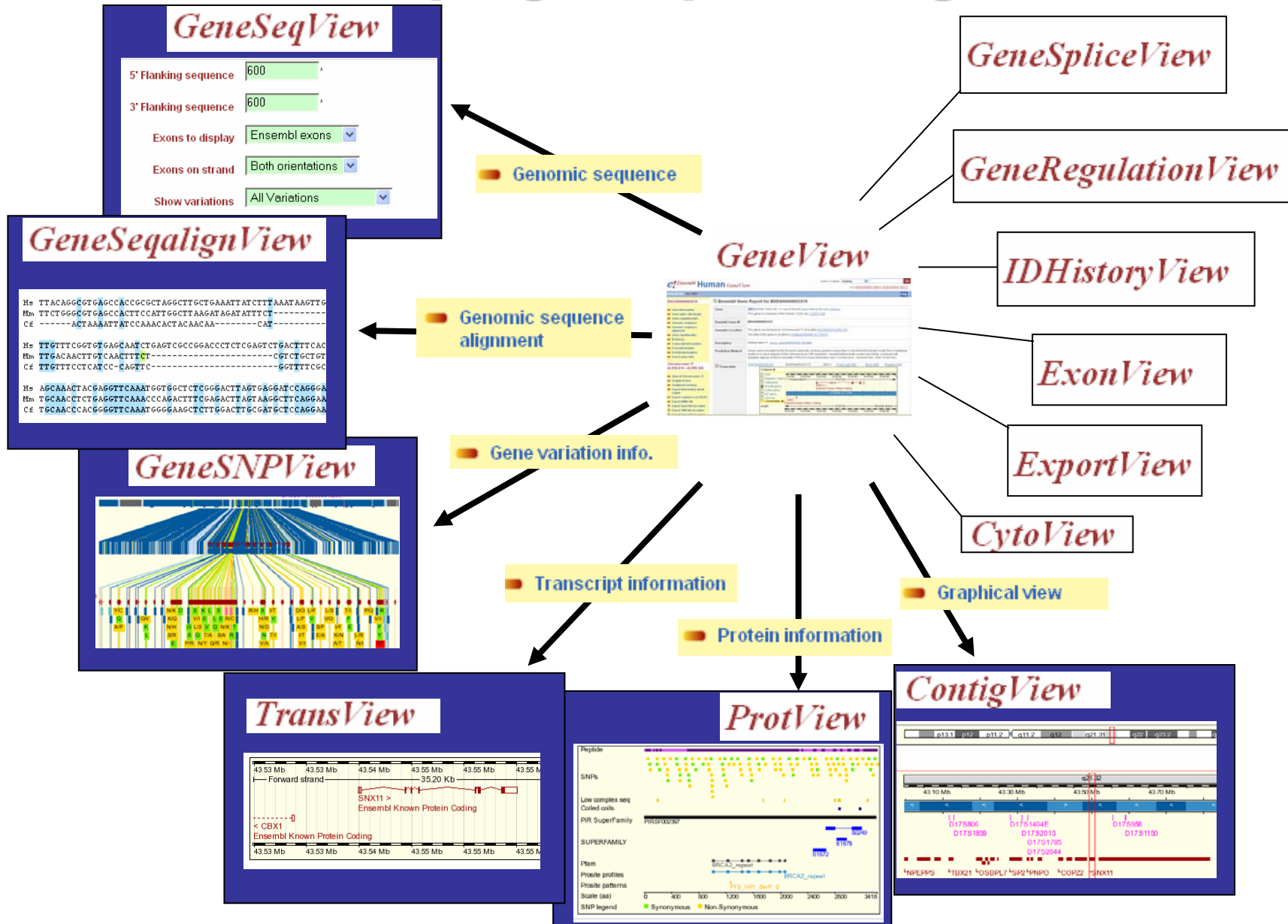
PIRSF
Washington, USA

‘InterPro is a database of protein families, domains and functional sites in which identifiable features found in known proteins can be applied to unknown protein sequences.’

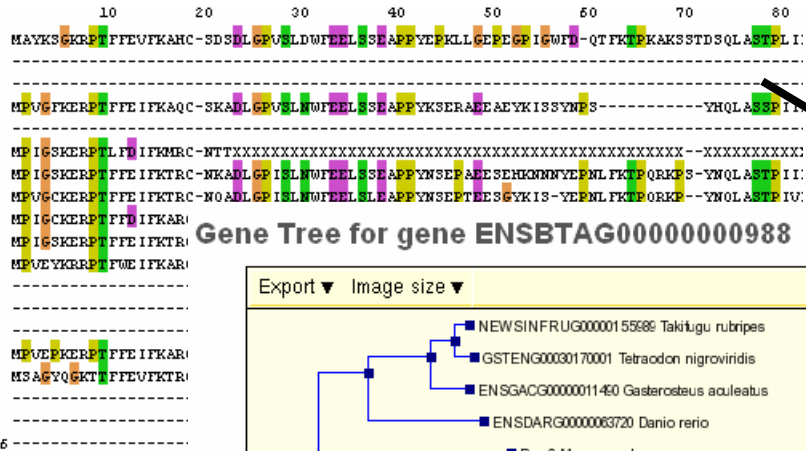
Altre Fonti di Informazione:

- **Chips (Affimetrix, Illumina, Agilent)**
- **GO (Gene Ontology)**
- **MIM (human diseases and phenotypes)**
- **Identifiers nel Entrez, UniProt, Refseq, etc**
- **PDB, MSD (structural databases)**

Nel Ensembl: pagine per un gene



Ensembl: Compara (un paragone tra le genome)



Sequence Alignment
(species to species)

Gene Tree for gene ENSBTAG00000000988



Gene Trees
(Evolutionary Information)

Ensembl: Variazioni

Choose SNP type

SNPs and variations in region of gene ENSG00000130427

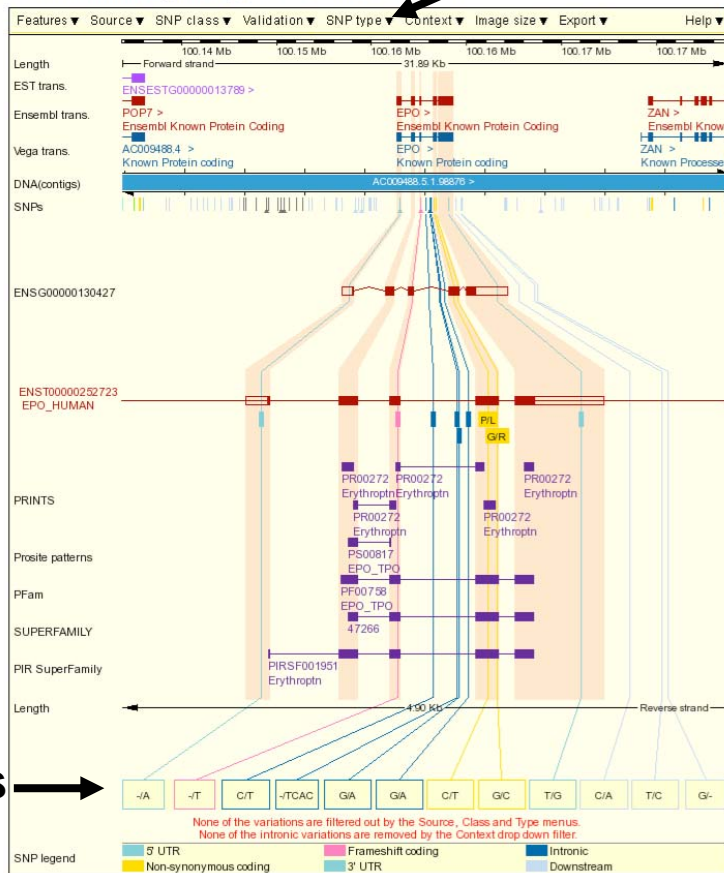


Table of Variations

Variations in ENST00000252723

ID	Type	Chr: bp	Alleles	Ambiguity	AA change	AA co-ordinate	Class	Source	Validation
rs34144627	5PRIME_UTR	7: 100156486-100156485	-/A	-	-	-	insertion	dbSNP	-
rs34937405	FRAMESHIFT_CODING	7: 100157590-100157589	-/T	-	-	77 (1)	insertion	dbSNP	-
rs507392	INTRONIC	7: 100157872	C/T	Y	-	-	snp	HGVbase, dbSNP	cluster, doublehit
rs33978705	INTRONIC	7: 100158068-100158067	-/TCAC	-	-	-	insertion	dbSNP	-
rs484199	INTRONIC	7: 100158087	G/A	R	-	-	snp	HGVbase, dbSNP	-
rs7789679	INTRONIC	7: 100158157	G/A	R	-	-	snp	dbSNP	-
rs11976235	NON_SYNONYMOUS_CODING	7: 100158317	C/T	Y	P/L	114 (2)	snp	dbSNP	cluster
rs1126887	NON_SYNONYMOUS_CODING	7: 100158394	G/C	S	G/R	140 (1)	snp	dbSNP	-
rs564449	3PRIME_UTR	7: 100159074	T/G	K	-	-	snp	HGVbase, dbSNP, TSC, Affy GeneChip 500K Mapping Array	-
rs551238	DOWNSTREAM	7: 100159464	C/A	M	-	-	snp	HGVbase, dbSNP, TSC	cluster, freq, doublehit
rs4729606	DOWNSTREAM	7: 100159726	T/C	Y	-	-	snp	HGVbase, dbSNP	cluster, doublehit
rs35970223	DOWNSTREAM	7: 100159780	G/-	-	-	-	deletion	dbSNP	-

© 2006 WTSI / EBI. Ensembl is available to [download for public use](#) - please see the [code licence](#) for details.

- Gene tree info
- Gene variation info.
- ID history

SNPs

Aiuto e Documenti al Sito Web

Aiuto...
Contatti
Glossario
Publicazioni
Workshops

Help & Documentation

- Table of Contents
- Helpdesk
- What's New
- About Ensembl
- Downloading data
- Displaying your own data
- Ensembl software

Select a species

- Mammals
- Other chordates
- Other eukaryotes

Ensembl Archive

e! View previous release of page in Archive!
e! Stable Archive! link for this page

BLAST contra una genoma

Ensembl tools

-  **Start a sequence search** →
Search Ensembl for nucleotide and peptide sequences with BLAST and SSAHA.
-  **Mine Ensembl with BioMart** →
Cross-reference Ensembl datasets with BioMart, a powerful data-mining tool.
-  **Customise Your Ensembl** →
Register with Ensembl to bookmark your favourite pages, customise your home page and much more!
-  **Fetch data with the Ensembl API** →
Learn how to extract data from the public Ensembl database with this tutorial.

You are logged in as **Giulietta Spudich**: [Account home page](#) · [Log out](#)

Ensembl headlines: Release 42 (December 2006)

-  **New - User accounts** (all species)
-  **New species - Duck-billed Platypus** (*Ornithorhynchus anatinus*)

Pagine Customabile
(User Logins)

Download
data

Notizie
(con ogni 'release')

I Geni e 'Transcripts' di Ensembl

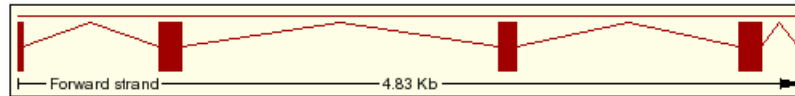
- Ensembl **known** genes or transcripts
Il base è evidenza della stessa specie
Ensembl **novel** genes or transcripts
Il base è evidenza della specie simile
- Ensembl **EST** genes or transcripts
Solamente usa l'evidenza EST
- **GENSCAN** transcripts
Senza evidenza: predizioni '*ab initio*'
- **VEGA/HAVANA** transcripts
Un' altro gruppo che fanno le predizioni manualmente
(manual curators)

Nomi di Ensembl

- **ENSG###** Ensembl Gene ID
 - **ENST###** Ensembl Transcript ID
 - **ENSP###** Ensembl Peptide ID
 - **ENSE###** Ensembl Exon ID
 - **ENSF###** Ensembl Family ID
-
- **Per le specie diverse da quella umana, ci sono tre lettere di piu:**
MUS (*Mus musculus*) per topolino: ENSMUSG###
DAR (*Danio rerio*) per zebrafish: ENSDARG###, etc.

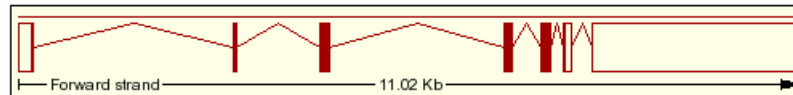
Strutture dei geni nel Ensembl

**Senza UTRs
(untranslated regions)**



Calmodulin- gallo

**Con
l'annotazione di
UTRs**



Calmodulin-umano

Termini di Ensembl

- I siti **Pre!** e **Archive**: *nuovi assemblaggi e siti ‘vecchi’ (ricordate: un nuovo ‘release’ ogni due mesi)*
- Geni **‘known e novel’**: *tutti hanno un base biologico*
- **‘Views’**: pagine specifiche nel sito web
- **‘BioMart’**: un’ applicazione per prendere il data di un gene o tanti geni/ regioni delle chromosome.

Gruppo Ensembl



Cambridge, UK

Leaders Ewan Birney (EBI), Tim Hubbard (Sanger Institute)

Database Schema and Core API Glenn Proctor, Andreas Kähäri, Ian Longden, Patrick Meidl

BioMart Arek Kasprzyk, Syed Haider, Richard Holland, Damian Smedley

Distributed Annotation System (DAS) Eugene Kulesha

Outreach & QC Xosé M Fernández, Bert Overduin, Michael Schuster, Giuletta Spudich

Web Team James Smith, Fiona Cunningham, Anne Parker, Stephen Rice, Steve Trevanion, Matt Wood

Comparative Genomics Abel Ureta-Vidal, Benoit Ballester, Kathryn Beal, Stephen Fitzgerald, Javier Herrero Sánchez, Albert Vilella

Analysis and Annotation Pipeline Val Curwen, Steve Searle, Bronwen Aken, Julio Banet, Laura Clarke, Sarah Dyer, Kevin Howe, Felix Kokocinski, Jan-Hinnerck Vogel, Simon White

Functional Genomics Paul Flicek, Yuan Chen, Stefan Gräf, Nathan Johnson, Daniel Rios

Zebrafish Annotation Kerstin Howe, Mario Caccamo, Ian Sealy

Vectorbase Annotation Martin Hammond, Dan Lawson, Karyn Megy

Systems & Support Guy Coates, Tim Cutts, Shelley Goddard

Research Damian Keefe, Ernesto Lowy, Guy Slater, Michael Hoffman, Alison Meynert, Benedict Paten, Dace Ruklisa, Daniel Zerbino

Adesso: il 'Worked Example'

- **Worked Example**
- **BioMart**
- **Le Variazioni**
- **Esercizi ed applicazioni dopo ogni lettura**

Tutti le mie presentazioni sono a questo sito:

http://www.ebi.ac.uk/~gspudich/workshop_presentations/pula