



*Universidad Nacional
de Colombia*

UniGene

***Diana Marcela Castillo
Yamile Adriana Celis***

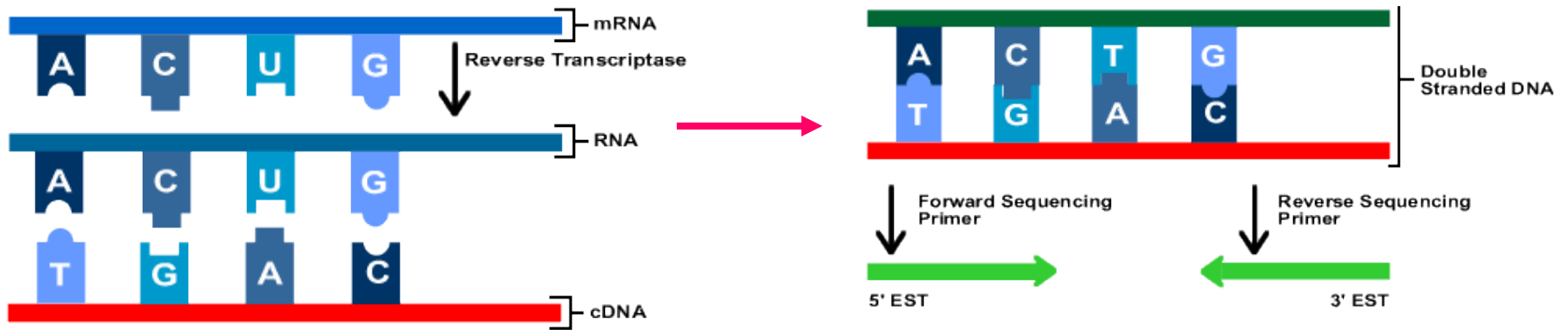
Bioinformática

Que es??

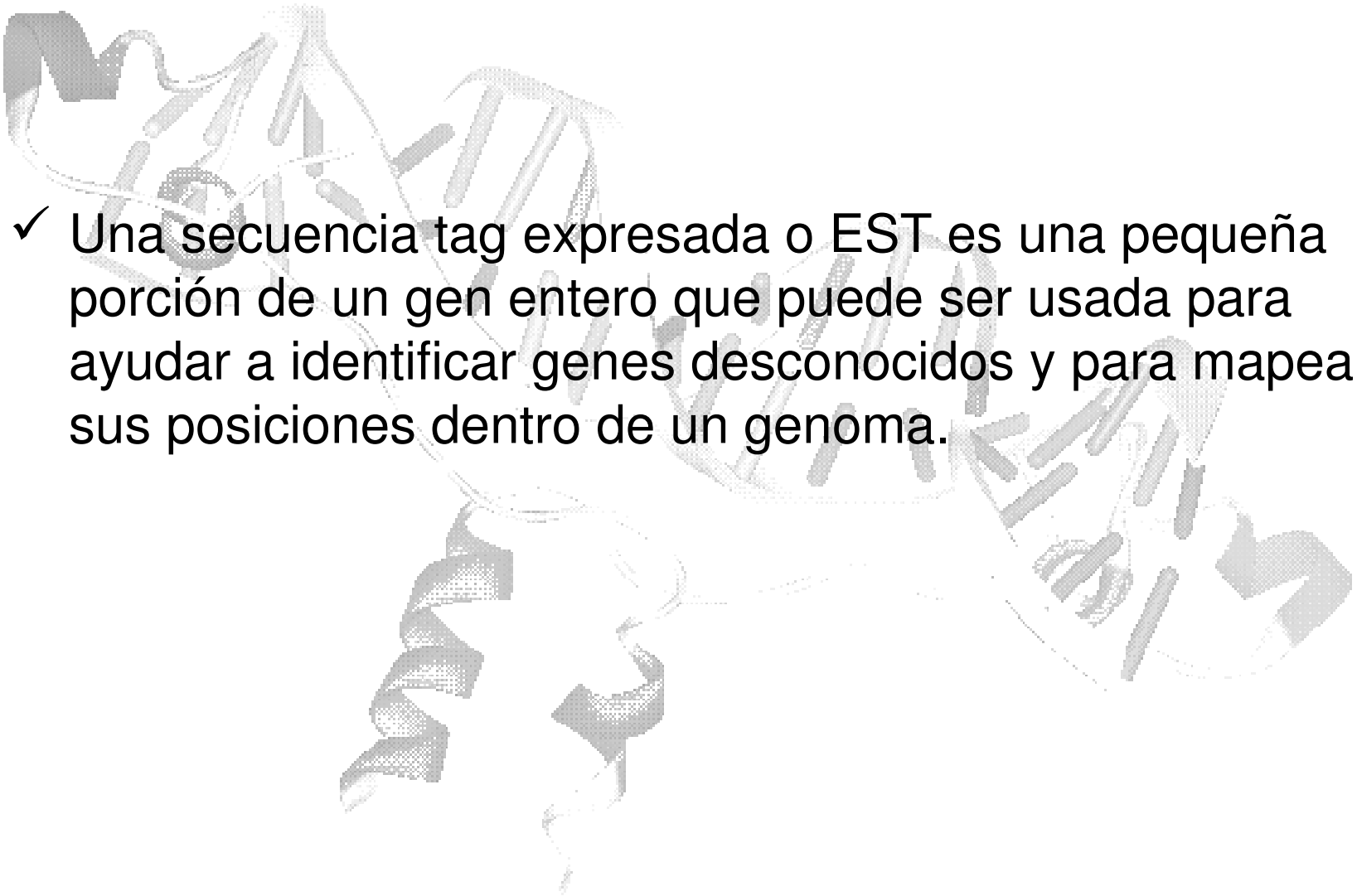
- Una manera organizada de ver transcriptomas.



- ✓ Cada entrada es un grupo o una colección de secuencias transcriptas que parecen venir del mismo locus de transcripción (genes o pseudogenes expresados).
- ✓ Todos con información de similares proteínas, genes de expresión, cDNA y localización genómica.



EST o Expressed Sequence Tags, son secuencias simples de pequeños trozos de cDNA, que son generados a partir de la secuenciación de uno o de los dos extremos del gen expresado. Su tamaño suele ser corto, de entre 200 y 500 nucleótidos y pueden ser secuencias codificantes o no codificantes.

- 
- ✓ Una secuencia tag expresada o EST es una pequeña porción de un gen entero que puede ser usada para ayudar a identificar genes desconocidos y para mapear sus posiciones dentro de un genoma.

Características:



- ✓ Además de secuencias de genes bien caracterizadas cientos de miles de secuencias tag (EST) expresadas han sido incluidas.
- ✓ La colección puede ser usada por la comunidad como fuente para descubrir genes, mapeo de genes y análisis de expresión a gran escala.

- ✓ No esta hecha para producir secuencias consenso.
- ✓ Por otra parte los grupos contienen EST que aunque tiene secuencias 5' - 3' de un mismo cDNA la secuencia no siempre se sobrelapa.

- ✓ Actualmente las secuencias de humanos, ratas, ratones, vacas, cebras, anfibios, mosca de la fruta y mosquitos han sido procesadas.
- ✓ Plantas como trigo, arroz, cebada y maíz, estas especies tiene una gran cantidad de datos disponibles de EST y representan una variedad de especies.
- ✓ En el futuro otros organismos pueden se adicionados en el futuro.



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

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

Y! Buscar Guardar en Mi Web

NCBI HomePage + Agregar pestaña

NCBI National Center for Biotechnology Information
National Library of Medicine National Institutes of Health

PubMed All Databases BLAST OMIM Books TaxBrowser Structure

Search **UniGene** for

SITE MAP
Alphabetical List
Resource Guide

About NCBI
An introduction to NCBI

GenBank
Sequence submission support and software

Literature databases
PubMed, OMIM, Books, and PubMed Central

Molecular databases
Sequences, structures, and taxonomy

What does NCBI do?

Established in 1988 as a national resource for molecular biology information, NCBI creates public databases, conducts research in computational biology, develops software tools for analyzing genome data, and disseminates biomedical information - all for the better understanding of molecular processes affecting human health and disease. [More...](#)

Hot Spots

- ▶ Assembly Archive
- ▶ Clusters of orthologous groups
- ▶ Coffee Break, Genes & Disease, NCBI Handbook
- ▶ Electronic PCR
- ▶ Entrez Home
- ▶ Entrez Tools
- ▶ Gene expression omnibus (GEO)
- ▶ Human genome resources
- ▶ Influenza Virus

New dbGaP
NCBI's dbGaP Genome Wide Association Database

NCBI's **dbGaP** (database of Genotype and Phenotype) provides data from Genome Wide Association (GWA) studies. The resource is intended to help elucidate the link between genes and disease. For each study, users have access to detailed information about the phenotypic variables measured and pre-computed associations between subjects' phenotypes and genotypes. [Click here to read the press release.](#) To read more about GWA projects, see NCBI's [GWA resource page](#)



UniGene

ORGANIZED VIEW OF THE TRANSCRIPTOME

My NCBI

All Databases PubMed Nucleotide Protein Genome Structure PMC Taxonomy Books

Search UniGene for

UniGene: An Organized View of the Transcriptome.

Each UniGene entry is a set of transcript sequences that appear to come from the same transcription locus (gene or expressed pseudogene), together with information on protein similarities, gene expression, cDNA clone reagents, and genomic location.

Species	UniGene Entries
Chordata	
Mammalia	
Bos taurus (cattle)	45,417
Canis familiaris (dog)	22,349
Equus caballus (horse)	5,499
Homo sapiens (human)	84,738
Macaca fascicularis (crab-eating macaque)	12,402
Macaca mulatta (rhesus monkey)	10,533
Mus musculus (mouse)	64,756
Oryctolagus cuniculus (rabbit)	5,915
Ovis aries (sheep)	10,944
Rattus norvegicus (Norway rat)	52,702
Sus scrofa (pig)	39,941

- UniGene
- Homepage
- News
- FAQs
- Query Tips
- Library Browser
- DDD
- Download UniGene
- Related Databases
- Gene
- HomoloGene
- dbEST
- Trace Archive
- NIH cDNA Projects
- MGC | ZGC | XGC
- Finding cDNAs

Dirección <http://www.ncbi.nlm.nih.gov/UniGene/JGOrg.cgi?TAXID=9615>

Canis familiaris-UniGene Build #17

NCBI UniGene Canis familiaris


PubMed Protein Genome Structure PopSet Taxonomy OMM

Search UniGene Canis familiaris[organism] Go Clear

Limits Index History ClipBoard Details

Canis familiaris: UniGene Build #17

Lineage: cellular organisms; Eukaryota; Fungi/Metazoa group; Metazoa; Eumetazoa; Bilateria; Coelomata; Deuterostomia; Chordata; Craniata; Vertebrata; Gnathostomata; Teleostomi; Euteleostomi; Sarcopterygii; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Laurasiatheria; Carnivora; Caniformia; Canidae; Canis; Canis familiaris



Sequences Included in UniGene

Known genes are from GenBank 30 Oct 2006
ESTs are from dbEST through 30 Oct 2006

2,331	mRNAs
0	Models
0	HTC
125,578	EST, 3'reads
22,469	EST, 5'reads
139,759	EST, other/unknown
290,137	total sequences in clusters

Build Method: Transcript Based

Alignments between all transcript sequences are used to generate clusters of sequences originating from the same gene.
[More...](#)

Final Number of Clusters (sets)

22,349	sets total
1,236	sets contain at least one mRNA
0	sets contain at least one HTC sequence
748	sets contain both mRNAs and ESTs

Histogram of cluster sizes for UniGene Cfa build 17

4097-8192	4
2049-4096	10
1025-2048	13
513-1024	18
257-512	41
129-256	122
65-128	305
33-64	752
17-32	1469
9-16	2522
5-8	3939
3-4	5453
2	2881
1	4820

UniGene Links

- [Library Browser](#)
- [DDD](#)
- [Query Tips](#)
- [FAQ](#)
- [Finding cDNAs](#)

Related Databases

- [Gene](#)
- [HomoloGene](#)
- [dbEST](#)
- [Trace Archive](#)

NIH cDNA Projects

- [CGAP](#)
- [MGC](#)
- [ZGC](#)
- [XGC](#)



Información de UniGen



Otras herramientas



Search UniGene for dog [Save Search](#)

Display Summary Show 20 Sort by Send to

All: 22357 Fungi: 0 Insects: 0 Mammals: 22353 Plants: 0

Items 1 - 20 of 22357

Page 1 of 1118 [Next](#)

- 1:** [Hs.646478](#) [Links](#)
LOC392900: Similar to ribosome binding protein 1; Ribosome binding protein-1; ribosome binding protein 1 (dog 180kD homolog); ribosome binding protein 1 homolog 180kD (dog); hES
Homo sapiens, 3 sequence(s)
- 2:** [Cfa.15828](#) [Links](#)
RRBP1: Ribosome binding protein 1 homolog 180kDa (dog)
Canis familiaris, 28 sequence(s)
- 3:** [Hs.472213](#) [Order cDNA clone, Links](#)
RRBP1: Ribosome binding protein 1 homolog 180kDa (dog)
Homo sapiens, 864 sequence(s)
- 4:** [Str.1240](#) [Order cDNA clone, Links](#)



[PubMed](#)
[Nucleotide](#)
[Protein](#)
[Genome](#)
[Structure](#)
[Popset](#)
[Taxonomy](#)

Search

[Limits](#)
[Preview/Index](#)
[History](#)
[Clipboard](#)
[Details](#)

- NCBI
- UniGene
 - Homepage
 - News
 - FAQs
 - Query Tips
 - Library Browser
 - DDD
 - Download UniGene
- Related Databases
 - Gene
 - HomoloGene
 - dbEST
 - Trace Archive
- NIH cDNA Projects
 - MGC | ZGC | XGC
 - Finding cDNAs

UniGene Hs.646478 *Homo sapiens* LOC392900

[Links](#)

Similar to ribosome binding protein 1; Ribosome binding protein-1; ribosome binding protein 1 (dog 180kD homolog); ribosome binding protein 1 homolog 180kD (dog); hES (LOC392900)

GENE EXPRESSION

Tissues and development stages from this gene's sequences survey gene expression. Links to other NCBI expression resources.

no information :

SEQUENCES

Sequences representing this gene; mRNAs, ESTs, and gene predictions supported by transcribed sequences.

Model sequences (3)

XR_018062.1	PREDICTED: Homo sapiens similar to Ribosome-binding protein 1 (Ribosome receptor protein) (180 kDa ribosome receptor homolog) (ES/130-related protein) (LOC392900), mRNA
XR_018385.1	PREDICTED: Homo sapiens similar to Ribosome-binding protein 1 (Ribosome receptor protein) (180 kDa ribosome receptor homol



UniGene

ORGANIZED VIEW OF THE TRANSCRIPTOME

PubMed Nucleotide Protein Genome Structure Popset Taxonomy

Search UniGene

[Limits](#) [Preview/Index](#) [History](#) [Clipboard](#) [Details](#)

NCBI

UniGene

[Homepage](#)

[News](#)

[FAQs](#)

[Query Tips](#)

[Library Browser](#)

[DDD](#)

[Download UniGene](#)

Related Databases

[Gene](#)

[HomoloGene](#)

[dbEST](#)

[Trace Archive](#)

Hs#S34546543 1118 bases *Homo sapiens* UniGene [Hs.646478](#)

PREDICTED: Homo sapiens similar to Ribosome-binding protein 1 (Ribosome receptor protein) (180 kDa ribosome receptor homolog) (ES/130-related protein) (LOC392900), mRNA

SEQUENCE INFORMATION

GenBank entry: [XR_018062.1](#)

Sequence length: 1118 bases

FEATURES



[NLN|NIH|UniGene|](#)

[Privacy Statement](#)

[Disclaimer](#)

[NCBI Help](#)

PubMed Nucleotide Protein Genome Structure PMC Taxonomy OMIM Books

Search for

Limits Preview/Index History Clipboard Details

Display Show Send to Hide: sequence all but gene, CDS and mRNA features

Range: from to Reverse complemented strand Features:

1: [XR_018062](#). Reports PREDICTED: Homo s...[gi:113419526] [Links](#)

[Comment](#) [Features](#) [Sequence](#)

LOCUS XR_018062 1118 bp mRNA linear PRI 29-AUG-2006
DEFINITION PREDICTED: Homo sapiens similar to Ribosome-binding protein 1 (Ribosome receptor protein) (180 kDa ribosome receptor homolog) (ES/130-related protein) (LOC392900), mRNA.
ACCESSION XR_018062
VERSION XR_018062.1 GI:113419526
KEYWORDS .
SOURCE Homo sapiens (human)
ORGANISM [Homo sapiens](#)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo.
COMMENT MODEL [REFSEQ](#): This record is predicted by automated computational analysis. This record is derived from an annotated genomic sequence ([NT_079592](#)) using gene prediction method: GNOMON.
 Also see:
[Documentation](#) of NCBI's Annotation Process

- ✓ Actualización: entre una semana y menos de un mes.
- ✓ Las secuencias pueden ser retiradas:
 - Por que se han encontrado contaminantes
 - Por que dos grupos se relacionan y se pierde la identidad de uno de los dos grupos
 - Un grupo puede ser dividido en diferentes grupos por no encontrar relación.