

How-to: Functional annotation

a super-quick guide to use InterPro Scan

What do we need?

- We have a list of genes with its theoretical translations.
- We have no idea what those genes are...

Annotate against a well known database



Search SUPERFAMILY

SUPERFAMILY  +44 Recommend this on Google

 Follow @SUPERFAMILY

SUPERFAMILY is a database of structural and functional annotation for all proteins and genomes.

The SUPERFAMILY annotation is based on a collection of **hidden Markov models**, which represent structural protein domains at the [SCOP](#) superfamily level. A superfamily groups together domains which have an evolutionary relationship. The annotation is produced by scanning protein sequences from over [2,478 completely sequenced genomes](#) against the hidden Markov models.

[HOME](#) | [SEARCH](#) | [BROWSE](#) | [FTP](#) | [HELP](#) | [ABOUT](#)

Pfam 27.0 (March 2013, 14831 families)

The Pfam database is a large collection of protein families, each represented by **multiple sequence alignments** and **hidden Markov models (HMMs)**. [Less...](#)

Proteins are generally composed of one or more functional regions, commonly termed **domains**. Different combinations of domains give rise to the diverse range of proteins found in nature. The identification of domains that occur within proteins can therefore provide insights into their function.

Welcome to Gene3D v12.0

Gene3D is primarily a database of CATH v4.0 protein domain assignments for ENSEMBL and UniProt sequences

[Learn more »](#)



Annotate against a **bunch** of well known databases

InterProScan 5

[Input form](#) | [Web services](#) | [Help & Documentation](#)

[Tools](#) > [Protein Functional Analysis](#) > InterProScan 5

InterProScan 5 Sequence Search

This form allows you to scan your sequence for matches against the InterPro collection of protein signature databases.

STEP 2 - Select the applications to run

Select All Clear All

- | | | | | |
|---|---|--|---|---|
| <input checked="" type="checkbox"/> BlastProDom | <input checked="" type="checkbox"/> FPrintScan | <input checked="" type="checkbox"/> HMMPiR | <input checked="" type="checkbox"/> HMMPfam | <input checked="" type="checkbox"/> HMMSmart |
| <input checked="" type="checkbox"/> HMMTigr | <input checked="" type="checkbox"/> ProfileScan | <input checked="" type="checkbox"/> HAMAP | <input checked="" type="checkbox"/> PatternScan | <input checked="" type="checkbox"/> SuperFamily |
| <input checked="" type="checkbox"/> SignalPHMM | <input checked="" type="checkbox"/> TMHMM | <input checked="" type="checkbox"/> HMMPanther | <input checked="" type="checkbox"/> Gene3D | <input checked="" type="checkbox"/> Phobius |
| <input checked="" type="checkbox"/> Coils | | | | |

Example: Gene model 1 of *P. unknowinensis*

>Gene_model_1_aa

MIQIQTKVKVNDNSGIKIGQCICKIYKKKKVVGKIGDTI
LISAKKLRLNQKKKIKIVKGD LFKALIIHTTYQKQS
TIGNMVKFDKNCIILNNQNKPLGTRIFGPITSEFR
KQKNFKILSLASNIL

Example: Gene model 1 of *P. unknowinensis* - Ribosomal protein!

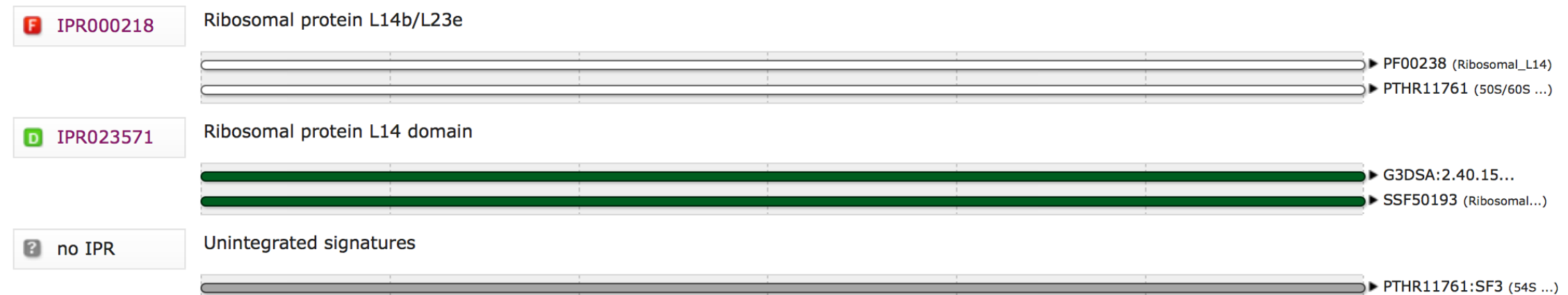
Protein family membership

 Ribosomal protein L14b/L23e (IPR000218)

Domains and repeats



Detailed signature matches



GO term prediction

Biological Process

[GO:0006412](#) translation

Molecular Function

[GO:0003735](#) structural constituent of ribosome

Cellular Component

[GO:0005840](#) ribosome