

BioMart

Datamining Databases

Jane Loveland
WTSI

What is BioMart?

- Data retrieval tool
- Originally developed for Ensembl (EnsMart)
- Now used by many large data resources
- Integrated with several widely used software packages
- Joint project between the European Bioinformatics Institute (EBI) and the Ontario Institute for Cancer Research (OICR)
- Central portal: <http://www.biomart.org>

Bio Portal : 41 databases in 4 continents and growing



UK - 21
USA - 8
France - 4
Canada - 1
Chile - 1
Peru - 1
Spain - 1
Italy - 1
China - 1
South Korea - 1
Japan - 1

DOWNLOAD

Version 0.8



Proceed to **Bio Portal** ▶
Proceed to **Cancer Portal** ▶
Proceed to **Mouse Portal** ▶

General or Specific Data-Tables

- All the genes for one species
- Or... only genes on one specific region of a chromosome
- Or... make BioMart select genes
(I.e. all transcripts that match a microarray probe set, GO term, or InterPro domain)

The first step: Choose the dataset

The second step: Filters

Third step: Attributes

Finally: Results

Other Export Options

- Sequences: UTRs, flanking sequences, cDNA and peptides, etc
- Gene IDs from Ensembl and external sources (MGI, Entrez, etc)
- Microarray data
- Protein Functions/descriptions (Interpro, GO)
- Orthologous gene sets
- SNP/ Variation Data

BioMart - Summary

- BioMart is a highly flexible tool for data retrieval
- Queries are defined in just 4 steps: 'Dataset', 'Filters', 'Attributes' and 'Results'
- Genomic regions, Gene identifiers, Gene
- Ontology terms and many other sources of information can serve as filters
- BioMart heavily links to data sets within Ensembl and provides links to external resources