BioMart Central Portal

Database, Vol. 2011, Article ID bar041, doi:10.1093/database/bar041

Original article

BioMart Central Portal: an open database network for the biological community

Jonathan M. Guberman[1], J. Ai[2], O. Arnaiz[3], Joachim Baran[1], Andrew Blake[4], Richard Baldock[5], Claude Chelala[6], David Croft[7], Anthony Cros[1], Rosalind J. Cutts[6], A. Di Génova[8], Simon Forbes[9], T. Fujisawa[10], E. Gadaleta[6], D. M. Goodstein[11,12], Gunes Gundem[13], Bernard Haggarty[5], Syed Haider[14], Matthew Hall[15], Todd Harris[16], Robin Haw[1], S. Hu[2], Simon Hubbard[17], Jack Hsu[1], Vivek Iyer[18], Philip Jones[7], Toshiaki Katayama[19], R. Kinsella[7], Lei Kong[20], Daniel Lawson[21], Yong Liang[1], Nuria Lopez-Bigas[13], J. Luo[7], Michael Lush[22], Jeremy Mason[15], Francois Moreews[23], Nelson Ndegwa[7], Darren Oakley[18], Christian Perez-Llamas[13], Michael Primig[24], Elena Rivkin[1], S. Rosanoff[7], Rebecca Shepherd[9], Reinhard Simon[25], B. Skarnes[18], Damian Smedley[7], Linda Sperling[3], William Spooner[16,26], Peter Stevenson[5], Kevin Stone[15], J. Teague[9], Jun Wang[20], Jianxin Wang[1], Brett Whitty[1], D. T. Wong[2], Marie Wong-Erasmus[1], L. Yao[1], Ken Youens-Clark[16], Christina Yung[1], Junjun Zhang[1] and Arek Kasprzyk[1,*]

[1]Ontario Institute for Cancer Research, Toronto, M5G 0A3, Canada, [2]School of Dentistry and Dental Research Institute, University of California Los Angeles (UCLA), Los Angeles, 90095-1668, USA, [3]Centre de Génétique Moléculaire UPR3404, Centre National de la Recherche Scientifique (CNRS), F-75794 Paris cedex 16 Paris, France, [4]MRC Harwell, Harwell Science and Innovation Campus, Oxfordshire, OX11 0RD, UK, [5]MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, Western General Hospital, Edinburgh, EH4 2XU, UK, [6]Centre for Molecular Oncology and Imaging, Barts Cancer Institute, Queen Mary University of London, E1 2AD, UK, [7]EMBL-European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SD, UK, [8]Center for Mathematical Modeling and Center for Genome Regulation, University of Chile, Blanco Enclada, Chile, [9]Cancer Genome Project, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Cambridge, CB10 1SA, UK, [10]Kasuza DNA Research Institute, Chiba, 292–0818, Japan, [11]Department of Energy, Joint Genome Institute, Walnut Creek, USA, [12]Genomics Division, Lawrence Berkeley National Laboratory, Berkeley, 94720, USA, [13]Research Unit on Biomedical Informatics, Department of Experimental and Health Sciences, Universitat Pompeu Fabra, Barcelona, 08003, Spain, [14]Computer Laboratory, University of Cambridge, Cambridge, CB3 0FD, UK, [15]Mouse Genomic Informatics Group, The Jackson Laboratory, Bar Harbor, 04609, USA, [16]Cold Spring Harbor Laboratory, Cold Spring Harbor, 11724, USA, [17]Faculty of Life Sciences, The University of Manchester, Manchester, M13 9PL, UK, [18]Wellcome Trust Sanger Institute, Welcome Trust Genome Campus, Hinxton, CB10 1SD, UK, [19]University of Tokyo, Tokyo, Japan, 14 University of Tokyo 7-3-1, Hongo, Bunkyo-ku, Tokyo, Japan 113-0033, [20]Center for Bioinformatics, State Key Laboratory of Protein and Plant Gene Research, College of Life Sciences, Peking University, Beijing, 100871, P.R. China, [21]VectorBase, European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, CB10 1SD, UK, [22]HUGO Gene Nomenclature Committee (HGNC), European Bioinformatics Institute (EMBL-EBI) Wellcome Trust Genome Campus, Hinxton, CB10 1SD, UK, [23]Sigenae, Institute National de la Recherche Agronomique (INRA), St-Gilles, 75338, France, [24]Inserm U625, University of Rennes, Rennes, 35043, France, [25]International Potato Center (CIP), Lima, 1558, Peru and [26]Eagle Genomics Ltd., Babraham Research Campus, Cambridge, CB22 3AT, UK

***Corresponding author:** Tel: +647 258 4321; Fax: 647 258 4321; Email: arek.kasprzyk@gmail.com

Submitted 16 May 2011; Revised 8 August 2011; Accepted 10 August 2011

BioMart Central Portal is a first of its kind, community-driven effort to provide unified access to dozens of biological databases spanning genomics, proteomics, model organisms, cancer data, ontology information and more. Anybody can contribute an independently maintained resource to the Central Portal, allowing it to be exposed to and shared with the research community, and linking it with the other resources in the portal. Users can take advantage of the common interface to quickly utilize different sources without learning a new system for each. The system also simplifies cross-database searches that might otherwise require several complicated steps. Several integrated tools streamline common tasks, such as converting between ID formats and retrieving sequences. The combination of a wide variety of databases, an easy-to-use interface, robust programmatic access and the array of tools make Central Portal a one-stop shop for biological data querying. Here, we describe the structure of Central Portal and show example queries to demonstrate its capabilities.
Database URL: http://central.biomart.org.

Database, Vol. 2011, Article ID bar041, doi:10.1093/database/bar041

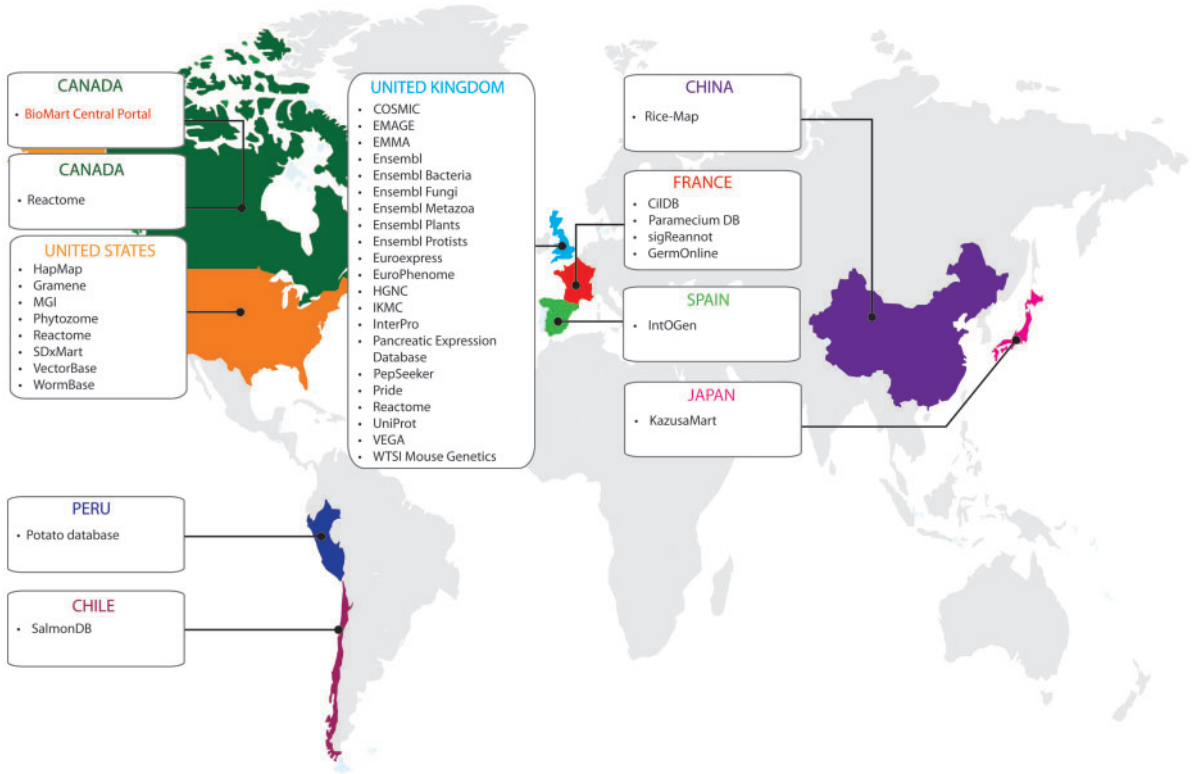

Figure 1. Databases available on the BioMart Central Portal and their host countries (April 2011).

Project description

Introduction

BioMart is a free, open-source, federated database system (1–3). It is cross-platform and supports many popular relational database managements systems, including MySQL, Oracle, PostgreSQL, SQL Server and DB2. The software is data-agnostic, and can therefore be easily adapted to existing data sets. It is expandable and customizable through a plug-in system, and is open-source so the community can participate in deeper development. Furthermore, BioMart can seamlessly connect geographically disparate databases, facilitating collaboration between different groups. These features have catalyzed the creation of BioMart Central Portal, a first of its kind community-supported effort to create a single access point integrating many different, independently administered biological databases (Figure 1).

For administrators, participation in Central Portal offers several benefits. Central Portal can provide an instantly available and automatically updated source of annotations for other projects, as is done in the International Cancer Genome Consortium Data Portal (4). Being part of the community can also expose a database to a wide user base. Furthermore, because the BioMart software allows administrators to easily create their own plug-ins, joining the community allows administrators to take advantage of the tools that others have created, thereby enhancing their own databases. Central Portal passes queries directly to the individual member servers, so administrators retain full control of their databases and their data (Figure 2).

For users, Central Portal offers a central repository for a vast array of biological data. BioMart can interoperate with other web sites, because results can be configured to link to outside resources; examples in Central Portal include KEGG pathway information (5–7) and Pancreatic Expression Database entries (8). The intuitive interface is consistent across all databases, so users familiar with one source can immediately transfer their skills to another data source. Since Central Portal is constantly updated, users are immediately exposed to new resources as they become available. In addition to the web-based interface, Central Portal also offers a wide variety of other access methods for more advanced querying, including application programming interfaces (APIs) for Java, SPARQL, REST and SOAP.

Moreover, both users and administrators benefit from the value gained by having individual databases connected in a central access point. By allowing data sets to be linked together, resources can be combined in novel ways,

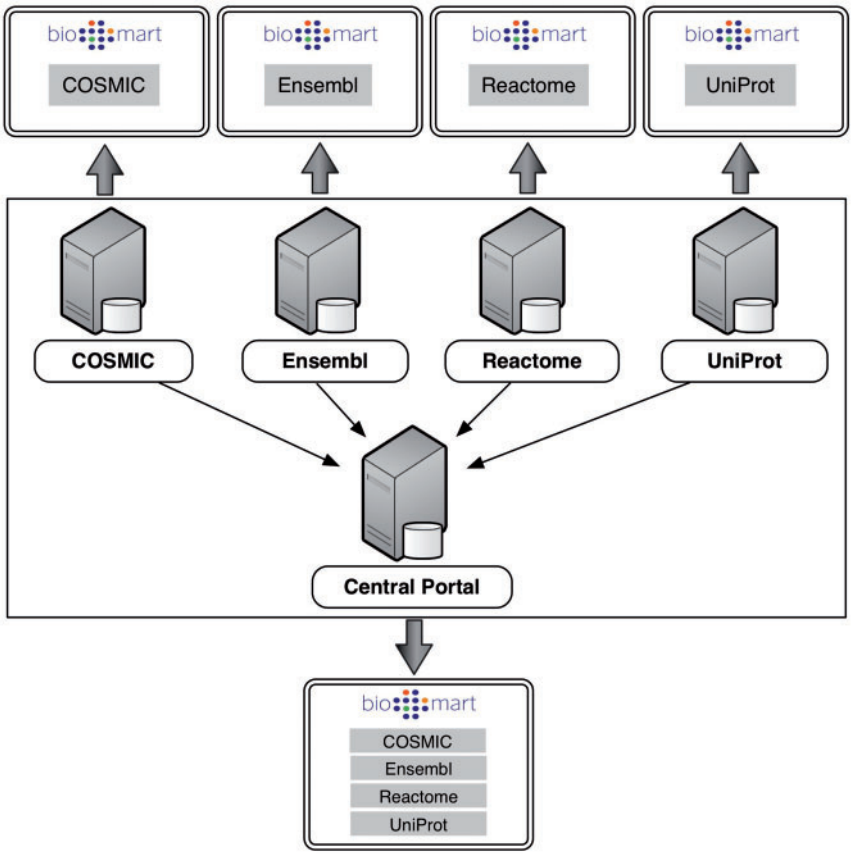

Figure 2. Each individual server hosts its own instance of BioMart retrieving data from its own local database backend. Central Portal offers a unified access point to all of these databases, distributing queries to the appropriate servers.

potentially revealing unexpected connections or suggesting new avenues of inquiry. The strength of the Central Portal comes from the fact that it is created and supported by a large community, and, as a whole, it is greater than the sum of its parts.

Interface

When viewing the Central Portal home page, users are presented with the main querying section, which is divided into three subsections: Identifier Search, Tools and Database Search (Figure 3).

The Identifier Search (Figure 3A) allows users to input gene identifiers in a number of formats (e.g. Gene name, Ensembl IDs, RefSeq IDs, etc.) and search for it across all of the member databases in the Portal. The result of the search links to a report page for the identifier, which summarizes key information about the search term taken from several sources (Figure 4). With this function users can quickly find information about a single identifier, and perhaps even locate resources that they did not realize were applicable to the target of their query.

The Tools section (Figure 3B) contains links to various data analysis tools in four categories: Gene retrieval, Variant retrieval, Sequence retrieval and ID Converter. The first two sections allow quick access to some of the largest and most popular databases contained in Central Portal. The third section, Sequence retrieval, allows easy querying of genomic and protein sequences in any of several formats (Figure 5). The fourth section, the ID Converter tool, allows users to enter or upload a list of identifiers in any format supported by a BioMart database, and retrieve the same list converted to any other supported format.

In the Database Search section (Figure 3C), users can access the individual member databases for querying through the BioMart interface. To make finding the relevant database easier, users can choose to browse databases by the type of information contained therein (Search by type) or by the organism with which the database is concerned (Search by organism). Browse by type is further subdivided into several categories such as Genome [e.g. Ensembl databases (9)], Gene annotation [e.g. HGNC (10)], Protein sequence and structure [e.g. InterPro (11)], Interactions and pathways [e.g. Reactome (12)], Gene

Database, Vol. 2011, Article ID bar041, doi:10.1093/database/bar041

expression [e.g. EMAGE (13)], Cancer [e.g. COSMIC (14)] and Model organism databases [e.g. Gramene (15)], Search by organism is subdivided into categories for bacteria, plants, protists, invertebrates and vertebrates. After choosing a data set, users can construct queries using the basic BioMart concepts of attributes, which indicate what information should be returned, and filters, which restrict the database entries that are retrieved.

Access methods

In addition to the graphical user interfaces, Central Portal also offers programmatic access to allow for automated querying. Several programming interfaces are available: an XML querying method that can be accessed via REST or SOAP requests, a full Java API and RDF querying via SPARQL. The syntax of any of the APIs is easy to use for programmers familiar with the basic BioMart concepts of attributes, filters and data sets. For example, to retrieve a list of filters for a given data set, a client could use the REST API and access the URL /martservice/filters?datasets=*datasetname*. Alternatively and equivalently, the client could use the Java API using the method getFilters (*datasetname*) to accomplish the same result. Because, there are a variety of APIs available, developers can choose the access method that makes the most sense for their specific applications and use cases.

To further ease the adoption of the APIs, the equivalent code of any query constructed in the web GUI can be retrieved in any of the API formats by clicking on the appropriate button on the query page; in this way, queries can be saved, modified and easily transferred from one format to another. It also provides a readily available graphical method of constructing complex API calls, which could be of use in certain tools or scripts.

Data content

BioMart Central Portal contains a constantly growing list of data sources accessible by a wide variety of methods and tools. The following table reflects the contents of the portal as of May 2011:

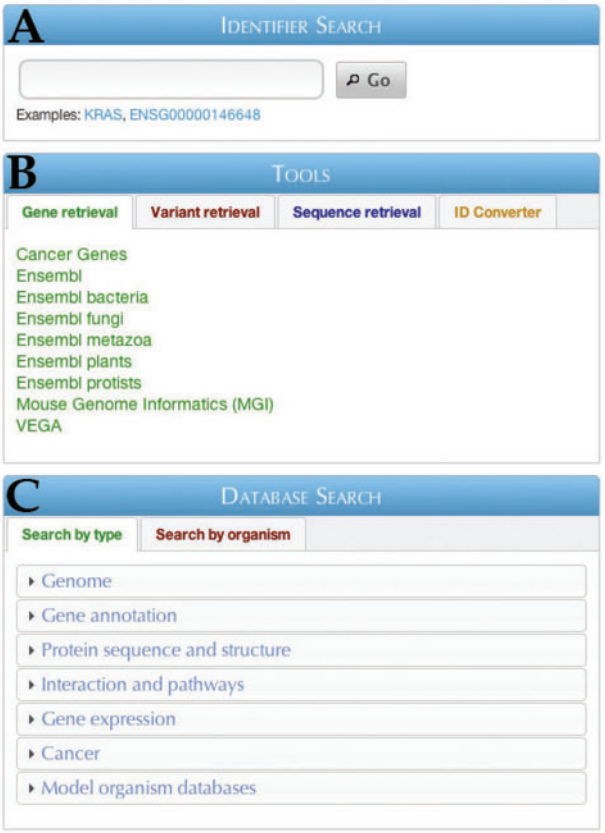

Figure 3. The BioMart Central Portal home page. Three main entry points are available: (**A**) Identifier search, (**B**) Tools and (**C**) Database search.

Database	Location	Description	References
Cildb	CNRS, France	Database for eukaryotic cilia and centriolar structures, integrating orthology relationships for 33 species with high-throughput studies and OMIM	(16)
COSMIC	WTSI, UK	Somatic mutation information relating to human cancers	(14)
EMAGE	MRC HGU, UK	*In situ* gene expression data in the mouse embryo	(13)
EMMA	EBI, UK	Mouse mutant strain information	(17)
Ensembl	WTSI/EBI, UK	Genome databases for vertebrates and other eukaryotic species	(9)
Ensembl Bacteria	EBI, UK	Genome databases for bacteria	(9)
Ensembl Fungi	EBI, UK	Genome databases for fungi	(9)
Ensembl Metazoa	EBI, UK	Genome databases for metazoa	(9)

(continued)

Database	Location	Description	References
Ensembl Plants	EBI, UK	Genome databases for plants	(9)
Ensembl Protists	EBI, UK	Genome databases for protists	(9)
Eurexpress	MRC HGU, UK	Transcriptome atlas database for mouse embryo	(18)
EuroPhenome	MRC Harwell, UK	Mouse phenotyping data	(19)
GermOnline	Inserm, France	Cross-species microarray expression database focusing on germline development, meiosis and gametogenesis as well as the mitotic cell cycle	(20)
Gramene	CSHL, USA	Agriculturally important grass genomes	(15)
HapMap	NCBI, USA	Multi-country effort to identify and catalog genetic similarities and differences in human beings	(21)
HGNC	EBI, UK	Repository of human gene nomenclature and associated resources	(10)
IKMC	WTSI, UK	Data on mutant products (mice, ES cells and vectors) generated and made available by members of the International Knockout Mouse Consotium	(22)
InterPro	EBI, UK	Integrated database of predictive protein 'signatures' used for the classification and automatic annotation of proteins and genomes	(11)
IntOGen	UPF, Spain	Integrated multi-dimensional data for the identification of genes and groups of genes involved in cancer development	(23)
KazusaMart	Kazusa, Japan	Cyanobase, rhizobia and plant genome databases	(24)
MGI	Jackson Laboratory, USA	Mouse genome features, locations, alleles and orthologues	(25)
Pancreatic Expression Database	Barts Cancer Institute, UK	Results from published pancreatic cancer papers	(8)
Paramecium DB	CNRS, France	Paramecium genome database	(26)
PepSeeker	University of Manchester, UK	Database of proteome peptide identifications for investigating fragmentation patterns	(27)
Phytozome	JGI/CIG, USA	Comparative genomics of green plants	(28)
Potato Database	CIP, Peru	Potato and sweet potato phenotypic and genomic information	(29)
PRIDE	EBI, UK	Repository for protein and peptide identifications	(30)
Reactome	OICR, Canada; EBI, UK; NYU Medical Center, USA	Curated pathway annotation database	(12)
Rice-Map	Peking University, China	Rice (*japonica* and *indica*) genome annotation database	(31)
SalmonDB	CMM, Chile	Genomic information for Atlantic salmon, rainbow trout and related species	(32)
SDxMart	UCLA, USA	Saliva diagnostics for high-impact human diseases	(33)
sigReannot	Rennes, France	Aquaculture and farm animal species EST contigs	(34)
UniProt	EBI, UK	Protein sequence and functional information	(35)
VectorBase	University of Notre Dame, USA	Genome information for invertebrate vectors of human pathogens	(36)
VEGA	WTSI, UK	Manual annotation of vertebrate genome sequences	(37)
WormBase	California Institute of Technology, USA; CSHL, USA; EBI, UK; Washington University, USA	*Caenorhabditis elegans* and related nematode genomic information	(38)
WTSI Mouse Genetics	WTSI, UK	Mouse phenotyping and expression data captured from mutant mouse lines	(39)

Database, Vol. 2011, Article ID bar041, doi:10.1093/database/bar041

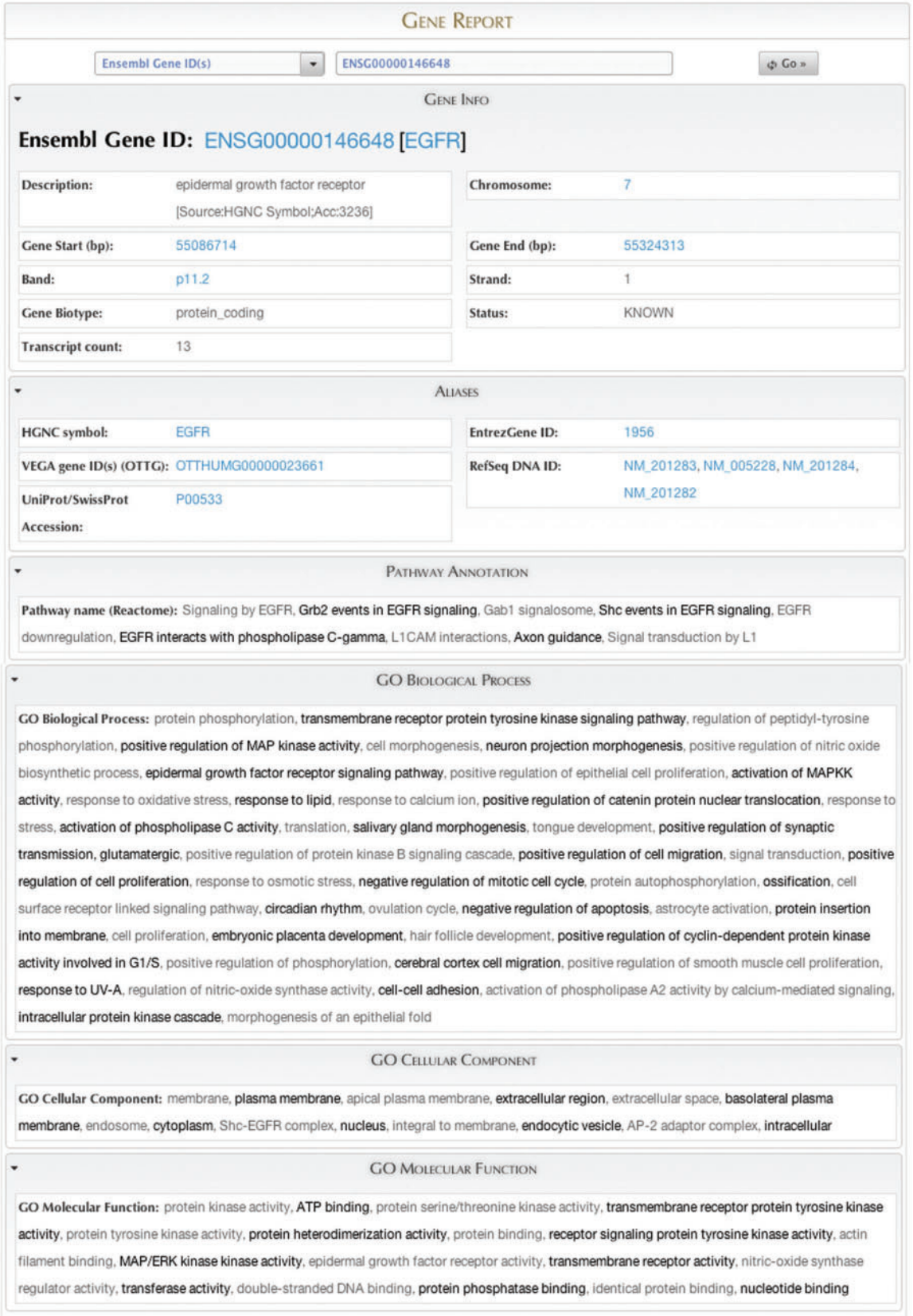

Figure 4. The Gene Report page for EGFR, displaying data federated from several sources.

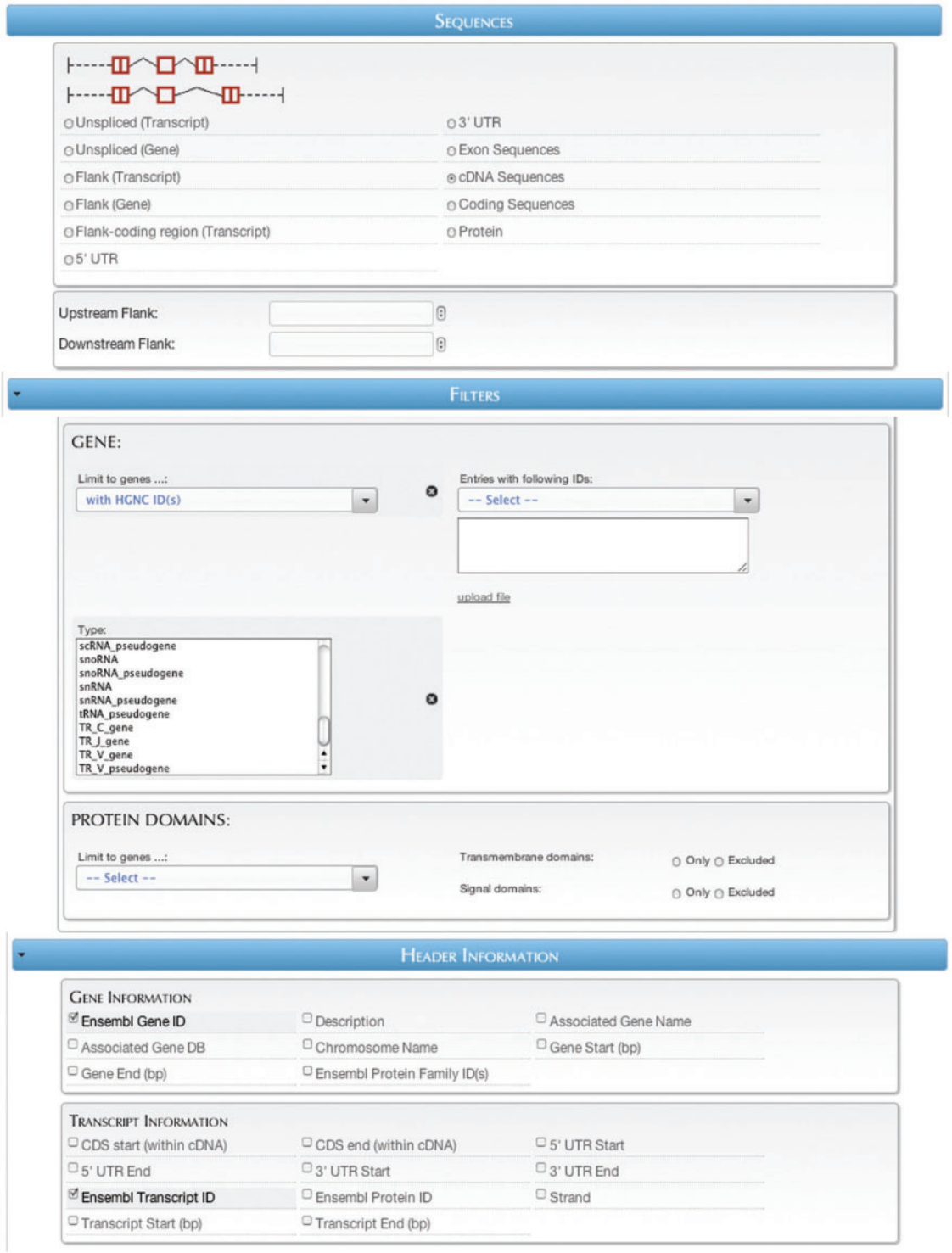

Figure 5. The sequence retrieval plug-in page.

Database, Vol. 2011, Article ID bar041, doi:10.1093/database/bar041

Query examples

One of the great strengths of Central Portal is that it allows cross-database searches that any individual resource would not. Here are some examples of the possibilities afforded by this feature.

Query #1: 'Find insertion-frameshift mutations in the COSMIC database that affect genes involved in Apoptosis'.

Entry point	Filters
Gene retrieval > cancer genes	COSMIC: Mutation type-AA: Insertion-frameshift KEGG: KEGG Pathway: apoptosis

By integrating data from the COSMIC and KEGG databases, Central Portal allows users to identify COSMIC mutations specific to their pathways of interest. The Pathway title links back to the KEGG web site and mutation ID links back to the COSMIC web site, providing the ability to obtain more detailed information on the pathway or on the mutation, respectively.

Query #2: 'Retrieve the cDNA sequences of protein-coding human genes that have HGNC IDs' (Figure 5).

Entry point	Data sets	Filters/attributes
Sequence retrieval > Ensembl	*Homo sapiens* gene (GRCh37.p2)	Sequences: cDNA sequences
		Filters:
		Limit to genes: with HGNC ID(s)
		Type: protein_coding
		Header information:
		Ensembl Gene ID
		Ensembl Transcript ID

By combining the sequence retrieval tool with search capabilities, BioMart reduces what is often a two-step process—retrieving a list of genes, and then retrieving the sequences of those genes—into a single query.

Future directions

BioMart Central Portal is constantly evolving thanks to the efforts of the community that supports it and contributes data. To make joining Central Portal easier, we are creating BioMart Central Registry. With this resource, database administrators will be able to create an account, add their data sources and suggest categorization for them. Once registered, participants will also be able to make changes to their databases and notify Central Portal of updates.

In addition to including new data sets, Central Portal will evolve, as new tools are developed and added. Such tools will perform deeper analysis, such as detecting enrichment of certain properties (e.g. GO terms) within a given set of genes or calculating consequences given a list of SNP terms. BioMart plug-ins developed by other community members may also be incorporated, further strengthening the project as a whole.

Acknowledgements

BioMart Central Portal is a collaborative, community effort and as such it is the product of the efforts of dozens, if not hundreds, of people. Creating a biological database is a multi-step process: experimenters must collect the data, database managers must create data models and administer databases and bioinformaticians must create methods for analysing the data. Additionally, over the years many programmers have contributed to the BioMart project codebase. We would like to acknowledge all the hard work of the many contributors to the projects that BioMart comprises.

Funding

The development of the BioMart software and the creation and hosting of BioMart Central Portal was supported by the Ontario Institute for Cancer Research and the Ontario Ministry for Research and Innovation. The individual data sources that Central Portal comprises are funded separately and independently.

Conflict of interest. None declared.

References

1. Haider,S., Ballester,B., Smedley,D. *et al*. (2009) BioMart Central Portal–unified access to biological data. *Nucleic Acids Res.*, **37**(Web Server issue), W23–W27.
2. Smedley,D., Haider,S., Ballester,B. *et al*. (2009) BioMart–biological queries made easy. *BMC Genom.*, **10**, 22.
3. Zhang,J., Haider,S., Guberman,J.M. *et al*. (2011) BioMart: a data federation framework for large collaborative projects. *Database*, (This issue).
4. Zhang,J., Baran,J., Cros,A. *et al*. (2011) International Cancer Genome Consortium Data Portal: a One Stop-Shop for Cancer Genomics Data. *Database*, (This issue).
5. Kanehisa,M. and Goto,S. (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.*, **28**, 27–30.
6. Kanehisa,M., Goto,S., Furumichi,M. *et al*. (2010) KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res.*, **38**(Database issue), D355–D360.
7. Kanehisa,M., Goto,S., Hattori,M. *et al*. (2006) From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res.*, **34**(Database issue), D354–D357.

8. Cutts,R.J., Gadaleta,E., Lemoine,N.R. *et al*. (2011) Using BioMart as a framework to create a cancer-specific database. *Database*, Published online 11 June 2011, doi:10.1093/database/bar024.
9. Kinsella,R., Kahari,A., Haider,S. *et al*. (2011) Ensembl BioMarts: a hub for data retrieval across the taxonomic space. *Database*, Published online 23 July 2011, doi:10.1093/database/bar030.
10. Povey,S., Lovering,R., Bruford,E. *et al*. (2001) The HUGO Gene Nomenclature Committee (HGNC). *Hum. Genet.*, **109**, 678–680.
11. Jones,P., Binns,D., McMenamin,C. *et al*. (2011) The InterPro BioMart: Powerful, federated query and web-service access to the InterPro resource. *Database*, Published online 23 July 2011, doi:10.1093/database/bar033.
12. Haw,R., Croft,D., Yung,C.K. *et al*. (2011) The Reactome BioMart. *Database*, (This issue).
13. Stevenson,P., Richardson,L., Venkataraman,S. *et al*. (2011) The BioMart interface to the eMouseAtlas gene expression EMAGE. *Database*, (This issue).
14. Shepherd,R., Forbes,S.A., Beare,D. *et al*. (2011) Data mining using the Catalogue of Somatic Mutations in Cancer BioMart (COSMICMart). *Database*, Published online 23 May 2011, doi:10.1093/database/bar018.
15. Spooner,W., Youens-Clark,K. and Ware,D. (2011) GrameneMart: The BioMart Data Portal for the Gramene Project. *Database*, This issue.
16. Arnaiz,O., Malinowska,A., Klotz,C. *et al*. (2009) Cildb: a knowledge-base for centrosomes and cilia. *Database*, **2009**, bap022 Published online 7 December 2009, doi: 10.1093/database/bap022.
17. Wilkinson,P., Sengerova,J., Matteoni,R. *et al*. (2010) EMMA–mouse mutant resources for the international scientific community. *Nucleic Acids Res.*, **38**(Database issue), D570–D576.
18. Diez-Roux,G., Banfi,S., Sultan,M. *et al*. (2011) A high-resolution anatomical atlas of the transcriptome in the mouse embryo. *PLoS Biol.*, **9**, e1000582.
19. Blake,A. (2011) The EuroPhenome BioMart: A mouse phenotyping resource. *Database*, (This issue).
20. Lardenois,A., Gattiker,A., Collin,O. *et al*. (2010) GermOnline 4.0 is a genomics gateway for germline development, meiosis and the mitotic cell cycle. *Database*, **2010**, baq030 Published online 10 December 2010, doi:10.1093/database/baq030.
21. International HapMap Consortium. (2003) The International HapMap Project. *Nature*, **426**, 789–796.
22. Oakley,J., Iyer,V., Skarnes,W.C. *et al*. (2011) BioMart as an integration solution for the International Knockout Mouse Consorium. *Database*, (This issue).
23. Perez-Llamas,C., Gundem,G. and Lopez-Bigas,N. (2011) Integrative Cancer Genomics (IntOGen) in BioMart. *Database*, (This issue).
24. KazusaMart. http://mart.kazusa.or.jp/biomart/martview/ (5 August 2011, date last accessed).
25. Shaw,D.R. (2009) Searching the Mouse Genome Informatics (MGI) resources for information on mouse biology from genotype to phenotype. *Curr. Protoc. Bioinformatics*, **Chapter** 1, Unit 1 7.
26. Arnaiz,O. and Sperling,L. (2011) ParameciumDB in 2011: new tools and new data for functional and comparative genomics of the model ciliate Paramecium tetraurelia. *Nucleic Acids Res.*, **39**(Database issue), D632–D636.
27. McLaughlin,T., Siepen,J.A., Selley,J. *et al*. (2006) PepSeeker: a database of proteome peptide identifications for investigating fragmentation patterns. *Nucleic Acids Res.*, **34**(Database issue), D649–D654.
28. Phytozome. http://www.phytozome.net/biomart/martview (5 August 2011, date last accessed).
29. International Potato Center. http://germplasmdb.cip.cgiar.org/biomart/martview (5 August 2011, date last accessed).
30. Vizcaino,J.A., Cote,R., Reisinger,F. *et al*. (2010) The proteomics identifications database: 2010 update. *Nucleic Acids Res.*, **38**(Database issue), D736–D742.
31. Rice Map. http://ricemart.cbi.edu.cn/biomart/martview (5 August 2011, date last accessed).
32. SalmonDB. http://genomicasalmones.dim.uchile.cl:9002/biomart/martview/ (5 August 2011, date last accessed).
33. Ai,J., Hu,S., Kasprzyk,A. *et al*. (2011) SDxMart: The BioMart Data Portal for Saliva Diagnostics. *Database*, This Issue.
34. Moreews,F., Klopp,C., Rauffet,G. *et al*. (2011) SigReannot-mart: A query environment for expression microarray probe re-annotations. *Database*, (This issue).
35. UniProt Consortium. (2010) The Universal Protein Resource (UniProt) in 2010. *Nucleic Acids Res.*, 38 (Database issue), D142–D148.
36. Lawson,D., Arensburger,P., Atkinson,P. *et al*. (2007) VectorBase: a home for invertebrate vectors of human pathogens. *Nucleic Acids Res.*, **35**(Database issue), D503–D505.
37. Wilming,L.G., Gilbert,J.G., Howe,K. *et al*. (2008) The vertebrate genome annotation (Vega) database. *Nucleic Acids Res.*, **36**(Database issue), D753–D760.
38. Harris,T.W., Antoshechkin,I., Bieri,T. *et al*. (2010) WormBase: a comprehensive resource for nematode research. *Nucleic Acids Res.*, **38**(Database issue), D463–D467.
39. Mouse Resources Portal. http://www.sanger.ac.uk/htgt/biomart/martview (5 August 2011, date last accessed).