BIOBASE
BIOLOGICAL DATABASES

"Hot Topics"
BIOBASE: Proteome and HGMD

Fran Lewitter

2/25/2009

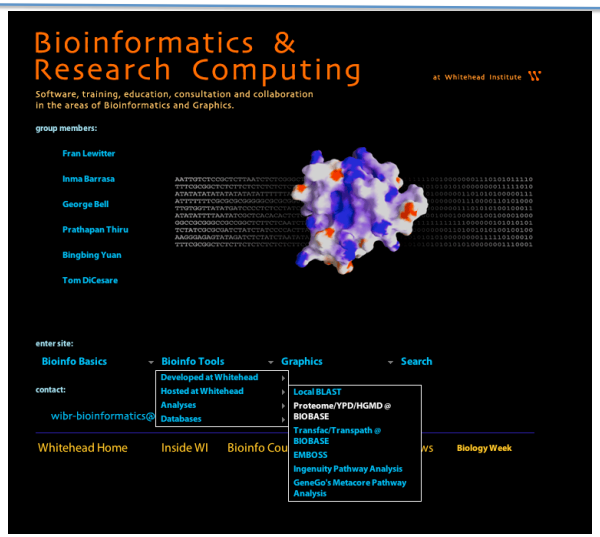
Our License

- **BIOBASE Knowledge Library (BKL)**
  - HumanPSD (Human Proteome Survey Database; human, rat, mouse)
  - GPCR-PD (G protein-coupled receptors; human, rat, mouse)
  - YPD suite and WormPD (PD stands for Proteome Database)
  - TRANSFAC (transcription factors)
  - TRANSPATH (signal transduction pathways)
  - ExPLAIN (Gene Expression Analysis System)
- HGMD
  - Human Gene Mutation Database

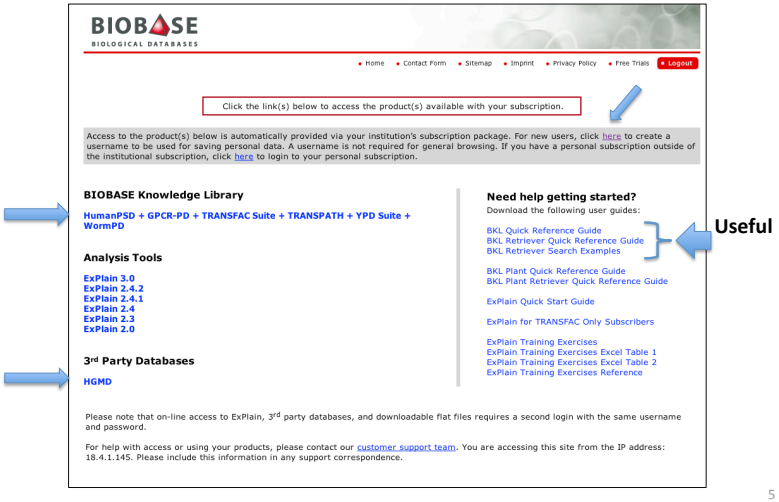
ExPlain™ Analysis System

- Facilitates analysis and biological interpretation of large scale data sets generated by DNA microarrays as well as proteomics, ChIP-on-chip, Chip-Seq and other high-throughput experiments.
- Flexible workflows systematically guide the user through the process of identifying the transcription factors and key signal transduction molecules that are unique for a data set

Local Access info

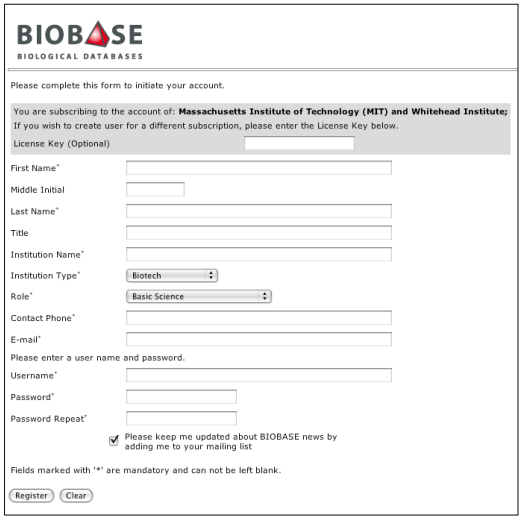


Local Access info



Useful

Register for Login



PROTEOME

- **Complete Mammalian coverage**
  - Protein characteristics on the full proteomes of human, mouse and rat, including:
  - Biological processes, molecular function, localization, expression
  - Protein modifications and interactions
  - Gene expression regulation
- **Yeast**
  - Quickly access experimental details from the literature on *S. cerevisiae*, *S. pombe*, *C. albicans* and 17 other human fungal pathogens
  - Mutant phenotypes and genetic interactions
  - Protein interaction networks
  - Environmental, chemical, and molecular gene regulators
- **Worm**
  - Integrates gene and protein interaction data, mutant phenotype, expression, function and protein family data
- **Disease**
  - Biomarker associations
  - Drug-target interactions
  - Mouse disease models

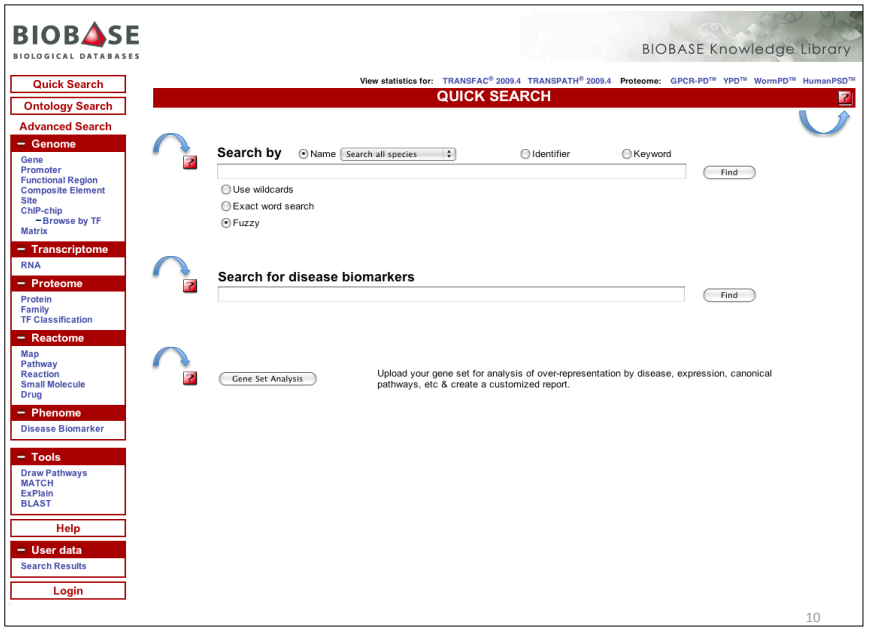
Statistics for the Winter 2009 release

HUMAN PSD	Human	Mouse	Rat
Proteins	18,985	21,125	11,410
Diseases/Disease Models	3,018	1,299	-
Disease-Gene Assignments	36.671	5,769	-
GO Assignments	181,359	183,359	99,773
Expression Assignments	149,251	153,560	93,738
Phenotype Assignments	-	66,032	-
References	151,858	77,669	48,086

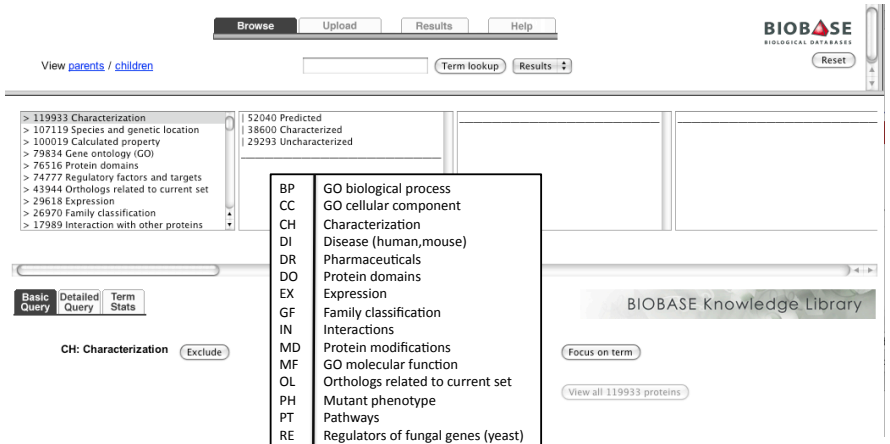
Other Organisms	YPD	PombePD	MycoPathPD	WormPD
Proteins	6,654	5,001	19,932	20,749
References	30,333	4,237	4,427	5,325
GO Assignments	65,771	25,178	93,290	76,035
Phenotype Assignments	38,632	5,866	2,722	18,845
Annotations	624,654	84,619	70,721	260,568

Ways to search BKL

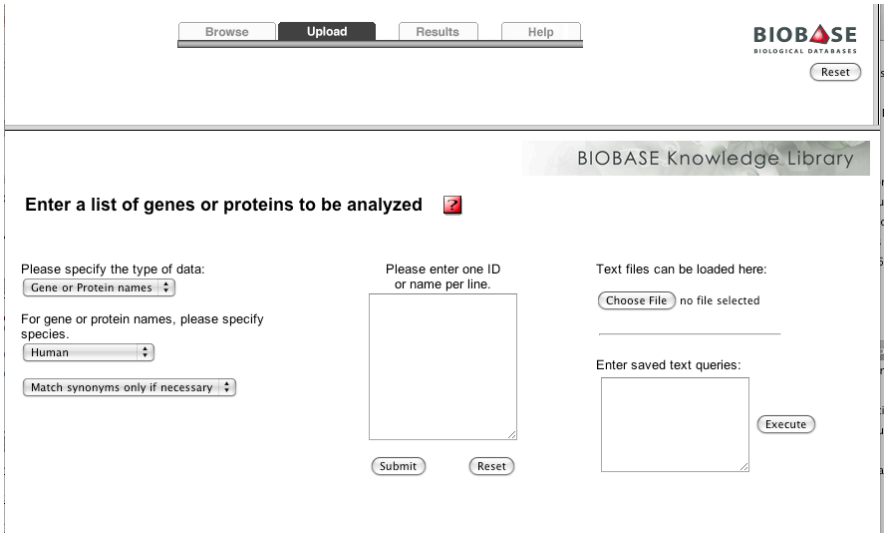
- Quick Search
  - –Locus report
- Ontology Search (Retriever)
  - –Upload data
- Advanced Search

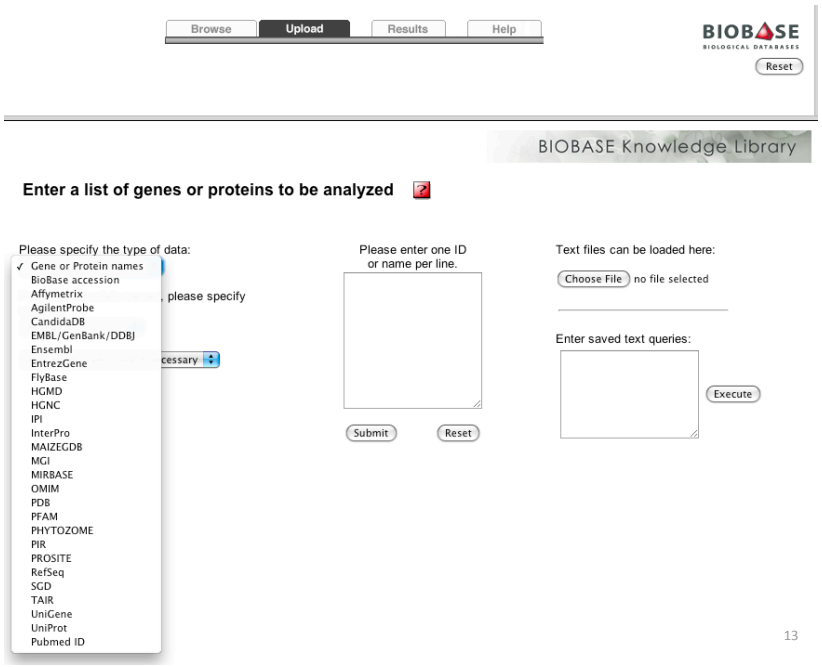


Ontology/Retriever Search

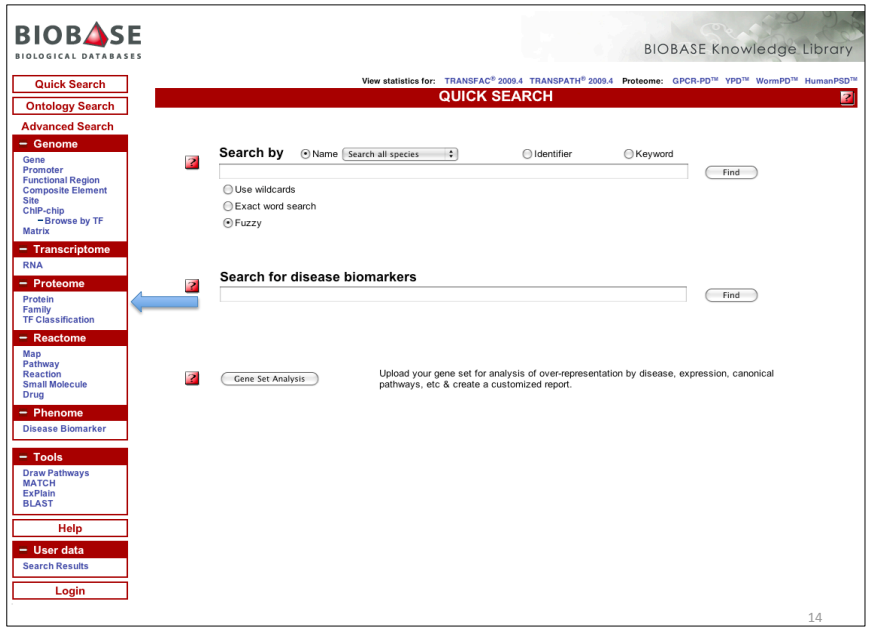


BP	GO biological process
CC	GO cellular component
CH	Characterization
DI	Disease (human,mouse)
DR	Pharmaceuticals
DO	Protein domains
EX	Expression
GF	Family classification
IN	Interactions
MD	Protein modifications
MF	GO molecular function
OL	Orthologs related to current set
PH	Mutant phenotype
PT	Pathways
RE	Regulators of fungal genes (yeast)
RG	Calculated properties
SP	Species and genetic location
TF	Regulatory factors and targets

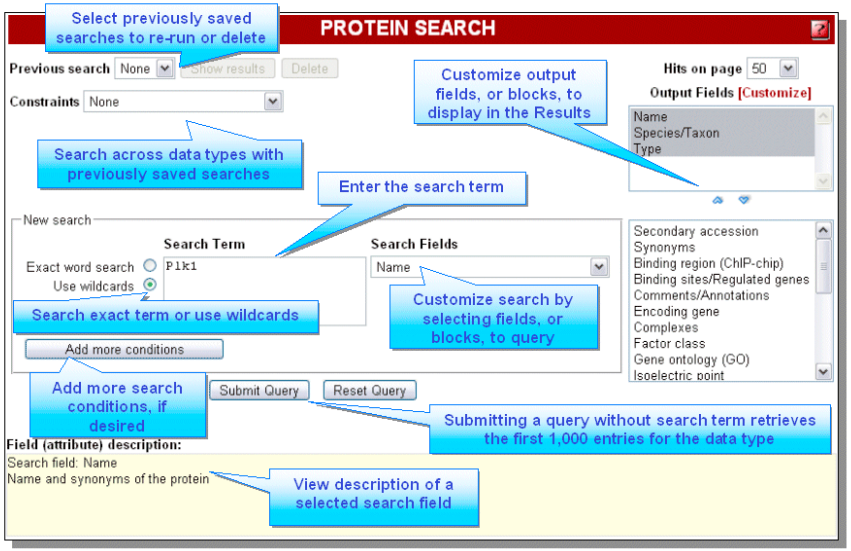



Browse
Upload
Results
Help
BIOBASE
BIOLOGICAL DATABASES
Reset
BIOBASE Knowledge Library
Enter a list of genes or proteins to be analyzed
Please specify the type of data:
Gene or Protein names
BioBase accession
Affymetrix
AgilentProbe
CandidaDB
EMBL/GenBank/DDBJ
Ensembl
EntrezGene
FlyBase
HGMD
HGNC
IPI
InterPro
MAIZEGDB
MGI
MIRBASE
OMIM
PDB
PFAM
PHYTOZOME
PIR
PROSITE
RefSeq
SGD
TAIR
UniGene
UniProt
Pubmed ID
, please specify
cessary
Please enter one ID
or name per line.
Submit
Reset
Text files can be loaded here:
Choose File
no file selected
Enter saved text queries:
Execute

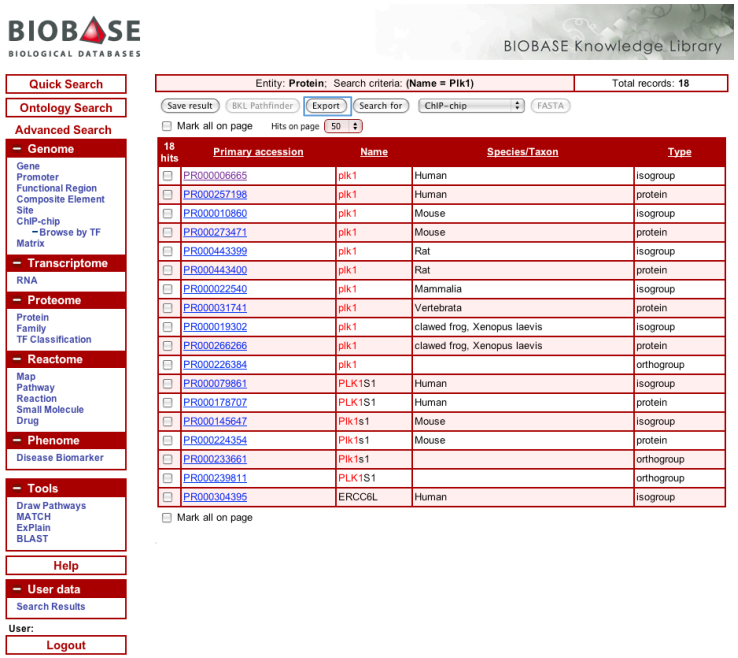
13


BIOBASE
BIOLOGICAL DATABASES
BIOBASE Knowledge Library
Quick Search
Ontology Search
Advanced Search
Genome
Gene
Promoter
Functional Region
Composite Element
Site
ChIP-chip
Browse by TF
Matrix
Transcriptome
RNA
Proteome
Protein
Family
TF Classification
Reactome
Map
Pathway
Reaction
Small Molecule
Drug
Phenome
Disease Biomarker
Tools
Draw Pathways
MATCH
ExPlain
BLAST
Help
User data
Search Results
Login
View statistics for: TRANSFAC® 2009.4 TRANSPATH® 2009.4 Proteome: GPCR-PD™ YPD™ WormPD™ HumanPSD™
QUICK SEARCH
Search by
Name
Search all species
Identifier
Keyword
Find
Use wildcards
Exact word search
Fuzzy
Search for disease biomarkers
Find
Gene Set Analysis
Upload your gene set for analysis of over-representation by disease, expression, canonical pathways, etc & create a customized report.

14


Select previously saved searches to re-run or delete
PROTEIN SEARCH
Previous search
None
Show results
Delete
Constraints
None
Customize output fields, or blocks, to display in the Results
Hits on page
50
Output Fields [Customize]
Name
Species/Taxon
Type
Search across data types with previously saved searches
Enter the search term
New search
Search Term
Search Fields
Exact word search
Use wildcards
Plk1
Name
Secondary accession
Synonyms
Binding region (ChIP-chip)
Binding sites/Regulated genes
Comments/Annotations
Encoding gene
Complexes
Factor class
Gene ontology (GO)
Isoelectric point
Search exact term or use wildcards
Customize search by selecting fields, or blocks, to query
Add more conditions
Add more search conditions, if desired
Submit Query
Reset Query
Submitting a query without search term retrieves the first 1,000 entries for the data type
Field (attribute) description:
Search field: Name
Name and synonyms of the protein
View description of a selected search field

15


BIOBASE
BIOLOGICAL DATABASES
BIOBASE Knowledge Library
Quick Search
Ontology Search
Advanced Search
Genome
Gene
Promoter
Functional Region
Composite Element
Site
ChIP-chip
Browse by TF
Matrix
Transcriptome
RNA
Proteome
Protein
Family
TF Classification
Reactome
Map
Pathway
Reaction
Small Molecule
Drug
Phenome
Disease Biomarker
Tools
Draw Pathways
MATCH
ExPlain
BLAST
Help
User data
Search Results
User:
Logout
Entity: Protein; Search criteria: (Name = Plk1)
Total records: 18
Save result
BKL Pathfinder
Export
Search for
ChIP-chip
FASTA
Mark all on page
Hits on page
50
18 hits
Primary accession
Name
Species/Taxon
Type
PR000006665 plk1 Human isogroup
PR000257198 plk1 Human protein
PR000010860 plk1 Mouse isogroup
PR000273471 plk1 Mouse protein
PR000443399 plk1 Rat isogroup
PR000443400 plk1 Rat protein
PR000022540 plk1 Mammalia isogroup
PR000031741 plk1 Vertebrata protein
PR000019302 plk1 clawed frog, Xenopus laevis isogroup
PR000266266 plk1 clawed frog, Xenopus laevis protein
PR000226384 plk1 orthogroup
PR000079861 PLK1S1 Human isogroup
PR000178707 PLK1S1 Human protein
PR000145647 Plk1s1 Mouse isogroup
PR000224354 Plk1s1 Mouse protein
PR000233661 Plk1s1 orthogroup
PR000239811 PLK1S1 orthogroup
PR000304395 ERCC6L Human isogroup
Mark all on page

16

HGMD Professional Winter 2009

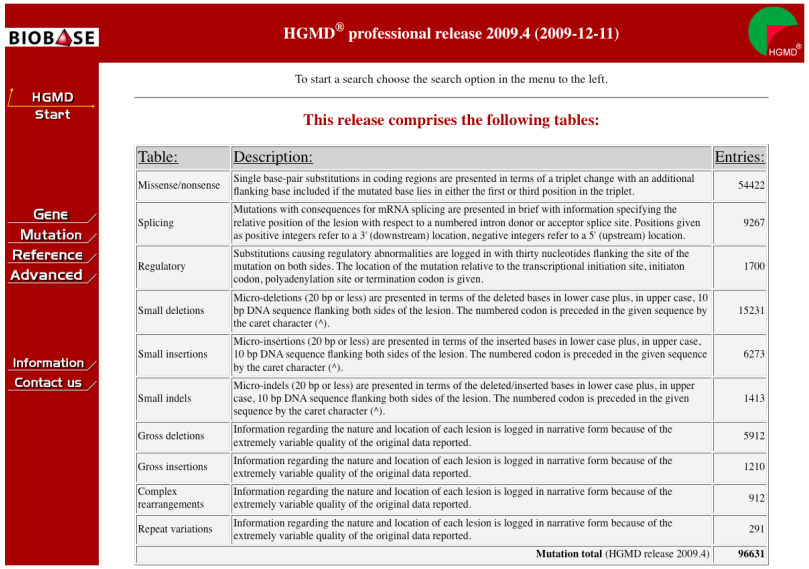
- Contains **96,631** mutations and disease associated/functional polymorphisms in **3,526** genes taken from **28,466** journal articles and provides **3,456** reference cDNA sequences. This represents a growth of **2,214** mutations over the previous release.
- Micro-lesions – missense, splicing, etc.
- Gross lesions – repeats, indels, rearrangements

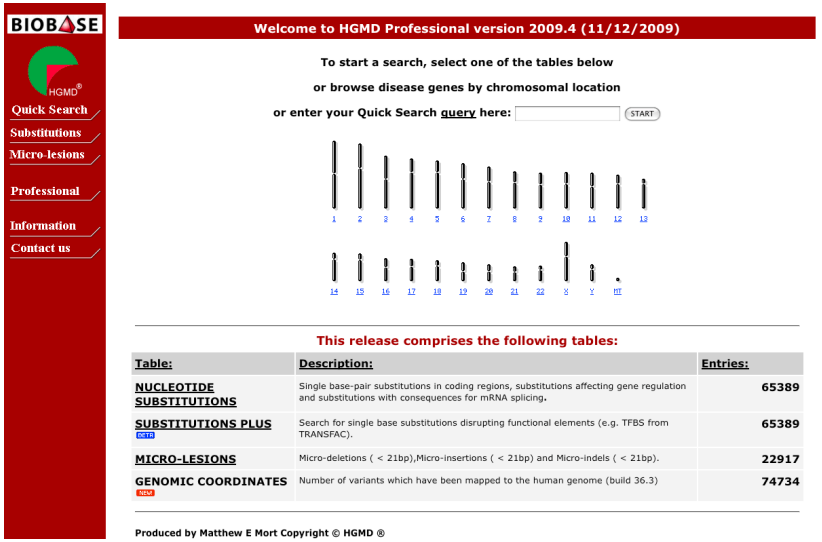
Access to HGMD data

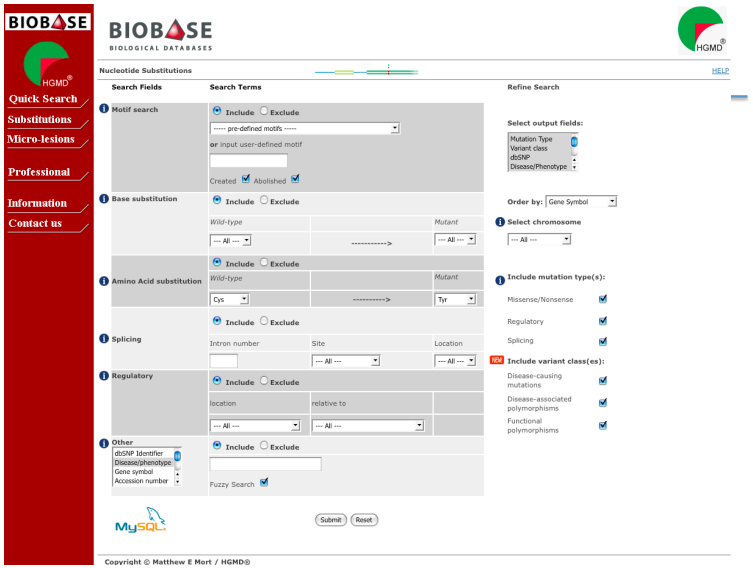
- Gene
  - Browse or Search
- Mutation
  - Search
- Reference
  - Search
- Advanced
  - Browse by chromosome or complex search

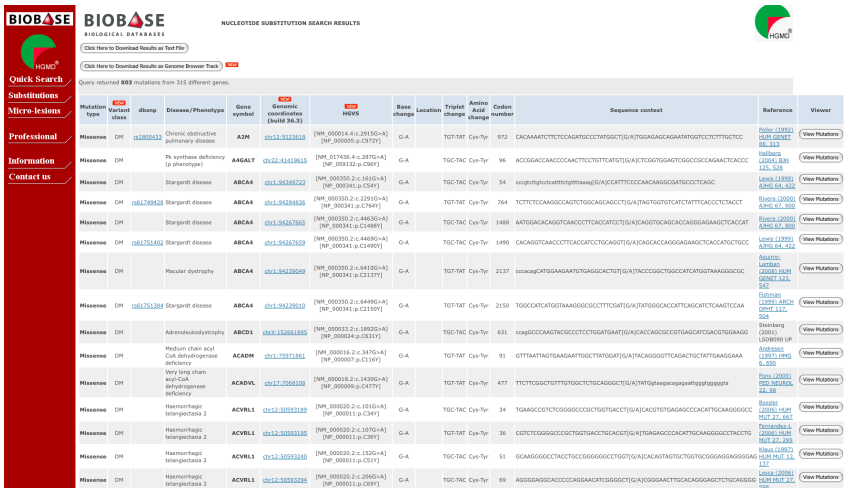
HGMD – Winter 2009

Type of Mutation	Number
Missense/nonsense	54,422
Splicing	9,267
Regulatory	1,700
Small deletions	15,231
Small insertions	6,273
Small indels	1,413
Gross deletions	5,912
Gross insertions	1,210
Complex rearrangements	912
Repeat variations	29
	96,631









Demo

- Simple Search and Locus Report
- Upload and export data
- Search in BKL
  - What yeast proteins act as structural components of the cytoskeleton but are not lethal when knocked out?
  - Which genes are associated with an alteration of the coding region in Alzheimer Disease?
  - Which mouse proteins are known to be associated with cardiovascular disease?
- Search in HGMD
  - What mutations are found more than 200 bp upstream of the initiation codon in humans?
  - What disease/phenotype is associated with a change of Cys->Tyr in humans?