



MartView

new
START
FILTER
OUTPUT
export

← back
next →

REGION:

Limit to (uncheck for entire genome):

Chromosome name:

From

To

Limit to ENCODE region

Type:

Region:

Entries in an ENCODE region

Only

Excluded

GENE:

Only

Excluded

Limit to genes with these IDs:
(Paste ID list, or upload file)

(NB) AFFY probe IDs should no longer contain the chip name prefix

Transcripts per gene:

Single

Multiple

Entries with a 5' UTR

Only

Excluded

Entries with a 3' UTR

Only

Excluded

GENE ONTOLOGY:

Evidence code for mapping:

Molecular Function
E.g. GO:0008083 or growth factor activity

Biological Process
E.g. GO:0008219 or cell death

Cellular Component
E.g. GO:0005623 or cell

EXPRESSION:

... eGenetics/SANBI data

bio::mart
count help

Summary

▶ **start**

- Focus:
- Ensembl
- Genes
- Species: Homo sapiens

① 21787 Genes Total

▶ **filter**

- Disease
- Genes Only
- Single
- transcript(s) per gene
- Has 5' UTR: Only
- Has 3' UTR: Only

① 537 Genes pass Filters

▶ **output**

① Not yet initialised

MartShell



```
C:\WINNT\System32\cmd.exe
MartShell: An Interactive User Interface to Mart based on Mart Query Language (MQL)
type 'help' for a list of available commands, or type 'help command' to get help for a particular command.

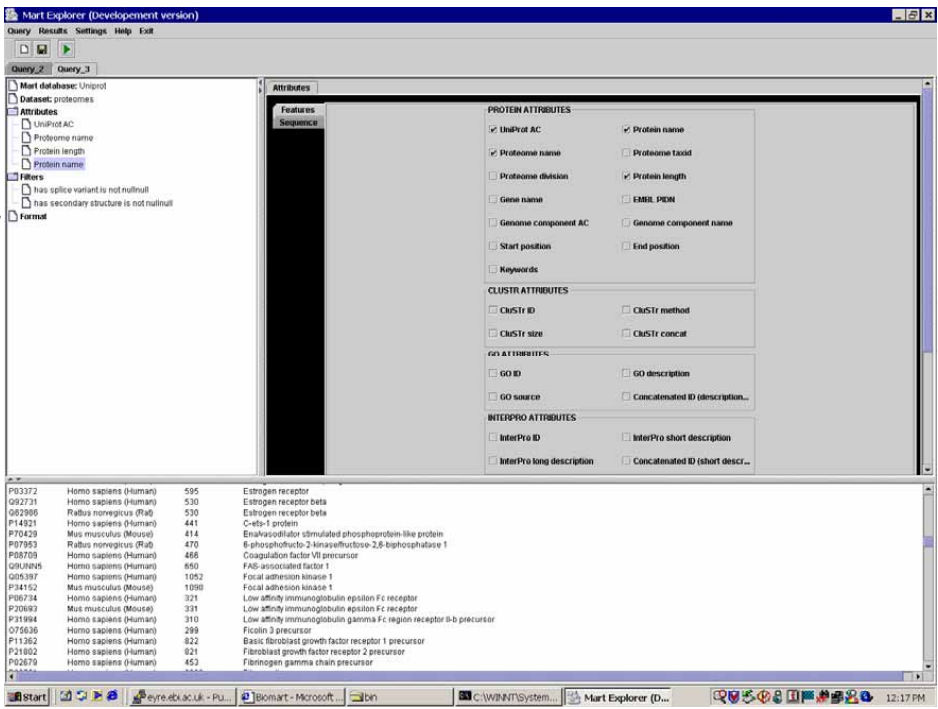
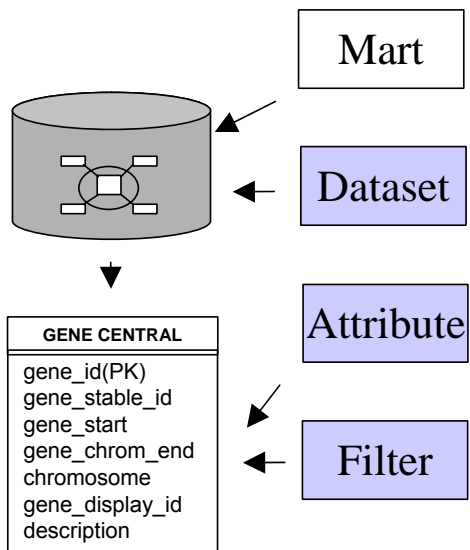
MartShell> list datasets all;

Ensembl.ensembl_genes_homo_sapiens
Ensembl.ensembl_genes_mus_musculus
Ensembl.snps_homo_sapiens
Ensembl.vega_genes_homo_sapiens
MSD_structures
Uniprot.proteomes

MartShell> using Ensembl.ensembl_genes_homo_sapiens get xref_swissprot_acc where disease_genes only and nonsynonymous_snps only
% and transmembrane_domains only and chromosome_name = 2 as ensembl_set;
MartShell>
MartShell> using Uniprot.proteomes get prot_seq where has_splice_variant only and sprot_list in ensembl_set
%
% ;
MEPWPLLLLSLCSAGLULGSEHETRLUAKLFKDYSSUURPUEDHRQQUEUTUGLQLIQLINUDEUNQIUTTNURLKQGDMDLPRPSCUTLGLUPLFSLHNEQWUDYNLKWNPDDYGGUKKIHIPSEKIWRPDLUL
VNNADGDFAIUKFTKULLQYTGHTITWPPAIFKSYCEIIVTHFPFDEQNCMSKLGWTYDGSUVAINPESDQDPLSNFMESGEWUIKESRGWKHSUTYSCCPDTPYLDITYHFUMQRLPLVFIUNUIIPCLLFSFLT
GLUFVLPDTSGEKMTLSISULLSLTUFLLUVELIPSTSSAULIGKVMFLTHUFUIASIIITUIUINTHHRSPSTHUMPNURKUFIDITIPNIMFFSTMKRPSREKQDKKIFTEIDISDISGKPPPPHGFHSPL
IKHPEUKSAIEGKIVIAETMKSQDSNNAAREMKYUAMUMDHILLGUFMLUCIIGTLAUFAGRLIELNQQG
MALLUSSLAFLLSLGSCCHRRICHSNRUFLCOESKUTEIPSDLPANAIELRFLTKLRUIQKGAFSGFQDLEKTEISONDULEIEADUFSLPKLHEIRIEKANLNLVTNPEAFQNLNLOVLLISNTGKIKHLPD
VHKIHSLOKQLLDIQDNIHHTIERNFSUGLSPFESUILLWLNKNGIQEIHNCANFGTQDELNLSDNNLEELPNDFHAGSAPVILDISRTRIHSLPSVYGLNKKLARSTVNLKLPLEKLUALMEASLTYPSPH
CCAFANURRQISELHPICNKSILRQEUVMYTRGRQSSLAEDNESSYRQFDMTYTEFYDLCNEUDDUTCSPKDPAFNPCEDIMGVNLRLIWFISILAITGNIULUILLTTSQYKLTUPRFLMCNLAFAADLCI
GIYLLLIASUDIHTKSQYHNYAIDWQTGAGCDAAGFFTFUFASEL SUYTLTAITLERWHTITHAMQLDCKUQLRHAASUMUMGWIFAFAAALFP IFGISSYMKUSICLPHMIDSPLSQLYUMSLLULNULAFUUICGC
VIHIYLVTURNPNIUSSSDTRIARKMAMLIFTDFLCMAPIISFFAISASLKUPLITUSKAKILLLULFHPINSCANPFLYAIFTKNFRDRFFLLSKCGCVENQAQIVRTETSSTUHNTHPRNGHCSSAPRNTUGSTYI
LUPLSHLAGN
MAAAGQLCLLYL SAGLLSRLGAANLDTREDNUIRKYVDPGSLFGFSLAMHWOLQPEDKRLLLUGAPRGEALPLORANRTGGLYSCDITARGPCTRIEFDNDADPTSESKEDAMMUTUOSQGGKUUOCAHRYEK
RQHUNTKQESRDIFGRCVYLSQNLRIEDDMGGDWSFCGRLRGHEKFGSCQQGUAATFTKDFHYIUFAGPPTYNWKGIURVEQKNNTFFDMNIFEDGPPYEUGETEHEDESUPANSYVLLGLFLTUSUYTDPDQF
UYKTRPPREQDTPFDUMMNSYLGFLSDSGKGIUSKDEITFUSGAPRANHSGAUULLKRDMSAHLLPEHIFDGEGLASSFGYDUUVDLNDKGDQDIUIGAPQYFDRDGEUGGAUYUYMNOQGRWNNKUIRNLGT
KDSMFGIAUKNIGDINODGVPDIAGUAPYDDLKUFIVHGSANGINTKPTQULGKISPYFVGSYIAGNMDLRNYSVPDUAUGSLSDSUTIFRSPRPUINIQKTIUTPNRIDLRQKTACGAPSGICLQKUSCFEVTANP
AGYNPSISIUIGTLEAEKERRKSGLSRROUFRNQGSEPKYTOELTLKROKQKUCMEETLWLDNIRDKLRPIPTASUEIQEPSRRRNSLPEULPILNSDEPKTAHIDUHFLEKCGGDDNVCNSNLKLEYKFTCE
GNQDKFSYLPQKUPULKQDKDIALEITUTNSPNSRNPNTKDDGDAHEAKLIATFPDPTLTSAYRELRAFPEKQLSCUANONGSQADCELGPNPKRNSNUTFYULSTTEUTFDPTDLDINLLETTSNQDNLA
PITAKAKUUELELLSUSGUAKPSOUVFGGTUUGEQAMKSEDEUGSLIEVEFRUINL GKPLTNLGTATLNTIQPKESINGKMLLYLUKUESKLEKUTCEPAKEINSLNLTESHNSRKKREITEKOIDDNKFSLFAE
RKYQTLNCSUNUNUNIRCLRGLDSKASLILASRLWNSTFLEEYSKLNVLDILMRAFDIUTAARENIRLPNAGTQURUTUFPKSTUAQYSGUPWMIILUAILAGILMALLUFLUWKCGFFKRSRYDSDUPRYHAU
RIRKEEREIKDEKYIDNLEKKQWITKWNRESYS
MKQRFSAQLKLLKLLLLQPLPRALREALCPEPCNCPDGLRCPGPTAGLTRLSLAYLPUKUIPSQAFFRGLNEUIKIEISQIDSLERIEANAFDNLNLSEILIQNTKNLRYIEPGAFINLRLKYLKICNTGIRK
FPDUTKUFSSNSFILEICDNLHITIPGNAGMNNESUTLKYVNGFEEUQSHAFNGTTLTSLLEKENUHLEKMHNGAFRGATGPKTLDISSTKLQALPSYGLSEIQRILATSSYSYKLPRETFUNLLEATLT
VPSHCCAFRNLPTKEQNFSHSISENFSKOCESTURKUSNKTLYSSMLAESELSGWDVEYGFCLPKTPRCAPEPDAFNPCEDIMGVDFLRULIWLINILAIHGNMTULFULLTSRYKLTUPRFLMCNLSFADFCMGLY
LLL IASUDSOTKGOVYVNHAIWMTGSGCSTAGFFTFUFASEL SUYTLTUITLERWHTIYAIHLQKLLRHAAILIMLGGMLFSSLIAHMLPUGUSNYMKUSICFPMDUETTLTQVYILITILNLUVAFFIICACVYK
IYFAURNPELMAHNTKDIKAKMAILIFTDFTCMAPIISFFAISAAFKUPLITUTNSKULLULFYVINSCANPFLYAIFTKTFQRDFLLSKFGCCKRAELVRRKDFSAVTSNCKNGFTGSKNPSQSTLKLSTLHC
QGTALLDKTRYTEC
MALLIHLKTUSELRGGRDRIAKUTFRGQSFYSRULENCEDUADFDTEFRWPUASSIDRNEMLEIQUNFYKUSFNKLIIGTFRMULQKVUEESHUEUTDITLIDNNNAIKTSLCUEURYQATDGTUGSIDDGDFLGD
SLQEEKDSQETDGLLPGSRPSSRPPGKESFRRAGRSUFSAMKLGKNSHKEEPORPDEPAULEMEDLHLAIRLDGLDPPDSUSLASUTALTTNUSNKRKSPDIKMEPSAGRPMQYQUSITUIEARQLUGLNMOPU
UCUEUGDDKYSKSTNCPVYNEVUFDFHUSPDUHFKDIKISUIHKNLLRSGLTUGSFKMDUGTUYSQPEHOFHKKMAILSDPDDISSGLKGYKCDUAUUGKGDNIKTPHKANETDEDDIEGNLLLEPUG
PERQMARFYUKTYRAEGLPRMNTSLMANUKKAFIGENKDLUDPYUOUFFAGQKGTISUOKSSYPLWNEQUUFTDLPPLCKRNMUQIRDSKUNDUVAIGTHFDLARKISNDGKGFPLTGPAMUNNYGSTRNVTL
LDEHQDLNEGLGEGUSFRARLLLGLAUEIUDTNPELTSSTEUQEATPISESCAGKMEEFFLFGAFLEASIDRRNGDKPITFEUTIGNYNEUDGLSRPQRPRKPEGDEEVDLIGNASDDEAGDAGLASU
```

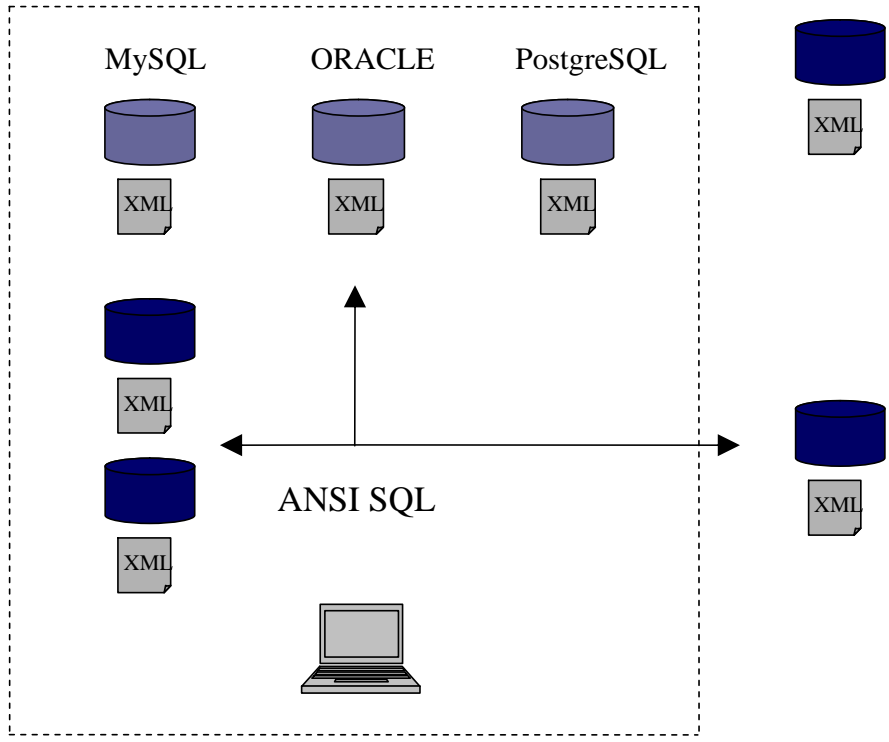


Key abstractions of generic system





BioMart - a distributed architecture





MartShell examples

MartShell> using MSD.msd get pdb_id where resolution_less < 1.5 and has_ec_info only;

193l
194l
1arb ...

MartShell> using MSD.msd get pdb_id where resolution_less < 1.5 and has_ec_info only as q;

MartShell> using

Ensembl.hsapiens_gene_ensembl get sequence transcript_flanks+1000 where pdb in q;

```
ENST00000270142.2    ENSG00000142168.2
    strand=forward   chr=21 assembly=NCBI34
    downstream flanking sequence of transcript
only
    AA ACTAAATTAGCTCTGATACTTATTTATATAAACAGCTTCAGTGGAA
    ....
```



MartShell examples (cont)

```
MartShell> using
Ensembl.hsapiens_gene_ensembl
get gene_stable_id, hugo, go_description
where chr_name = 3 and
      3.band_start = q22.1 and
      3.band_end = q22.3
and est.anatomical_site = retina;
```

```
ENSG00000051382  PIK3CB phosphoinositide 3-kinase
complex
```

```
ENSG00000163914  RHO G-protein coupled photoreceptor
activity ...
```