MICB 405 Bioinformatics
Mini-Lab #3 – Genome Browsers

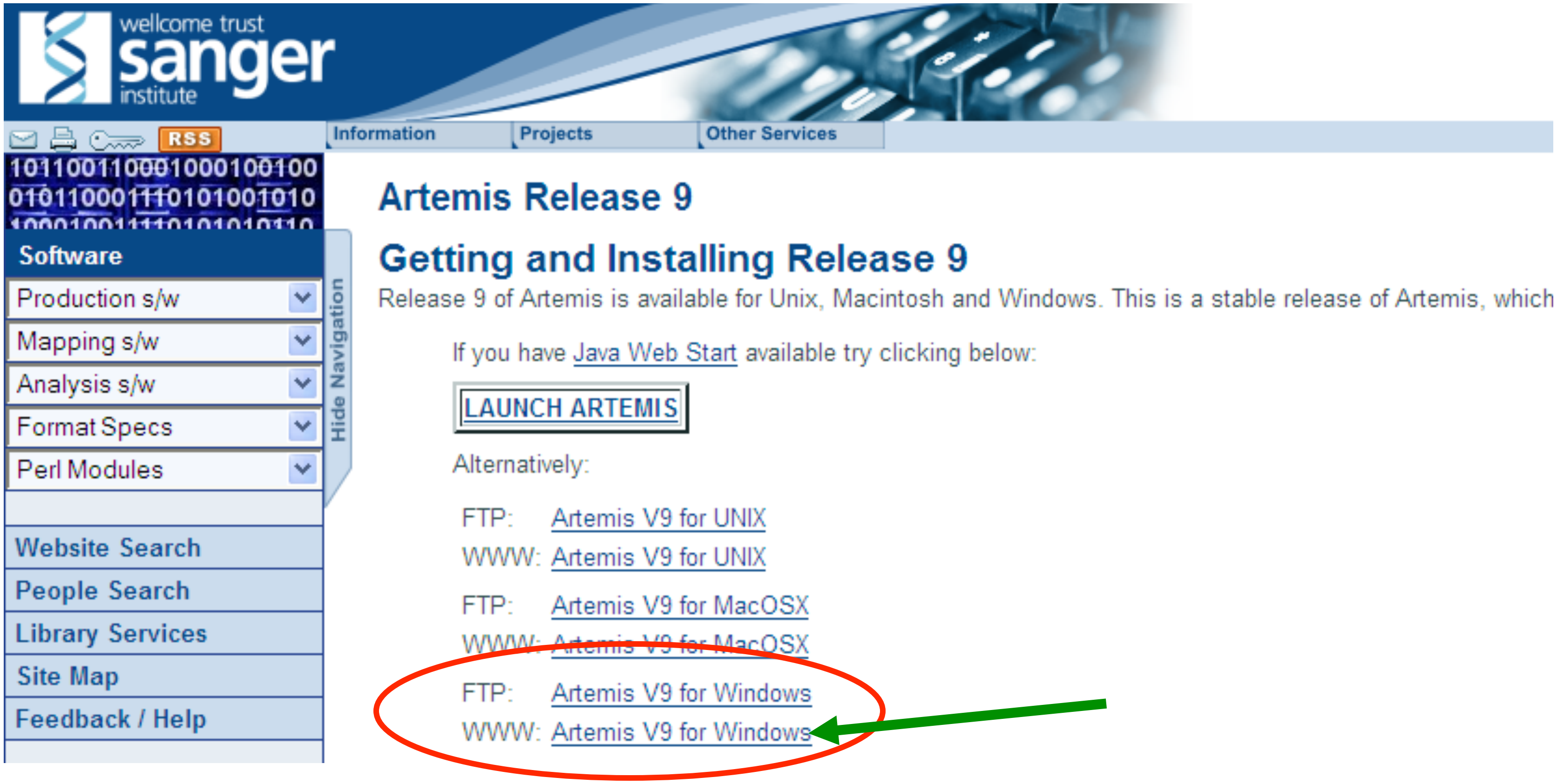
Dr. Joanne Fox
joanne@msl.ubc.ca

We gratefully acknowledge the funding for the development of these teaching materials provided by UBC students through the Teaching and Learning Enhancement Fund (TLEF) from 2005-2007.

Download Artemis to Desktop

http://www.sanger.ac.uk/Software/Artemis/v9/

Google: “artemis v9”



Artemis Release 9

Getting and Installing Release 9

Release 9 of Artemis is available for Unix, Macintosh and Windows. This is a stable release of Artemis, which

If you have Java Web Start available try clicking below:

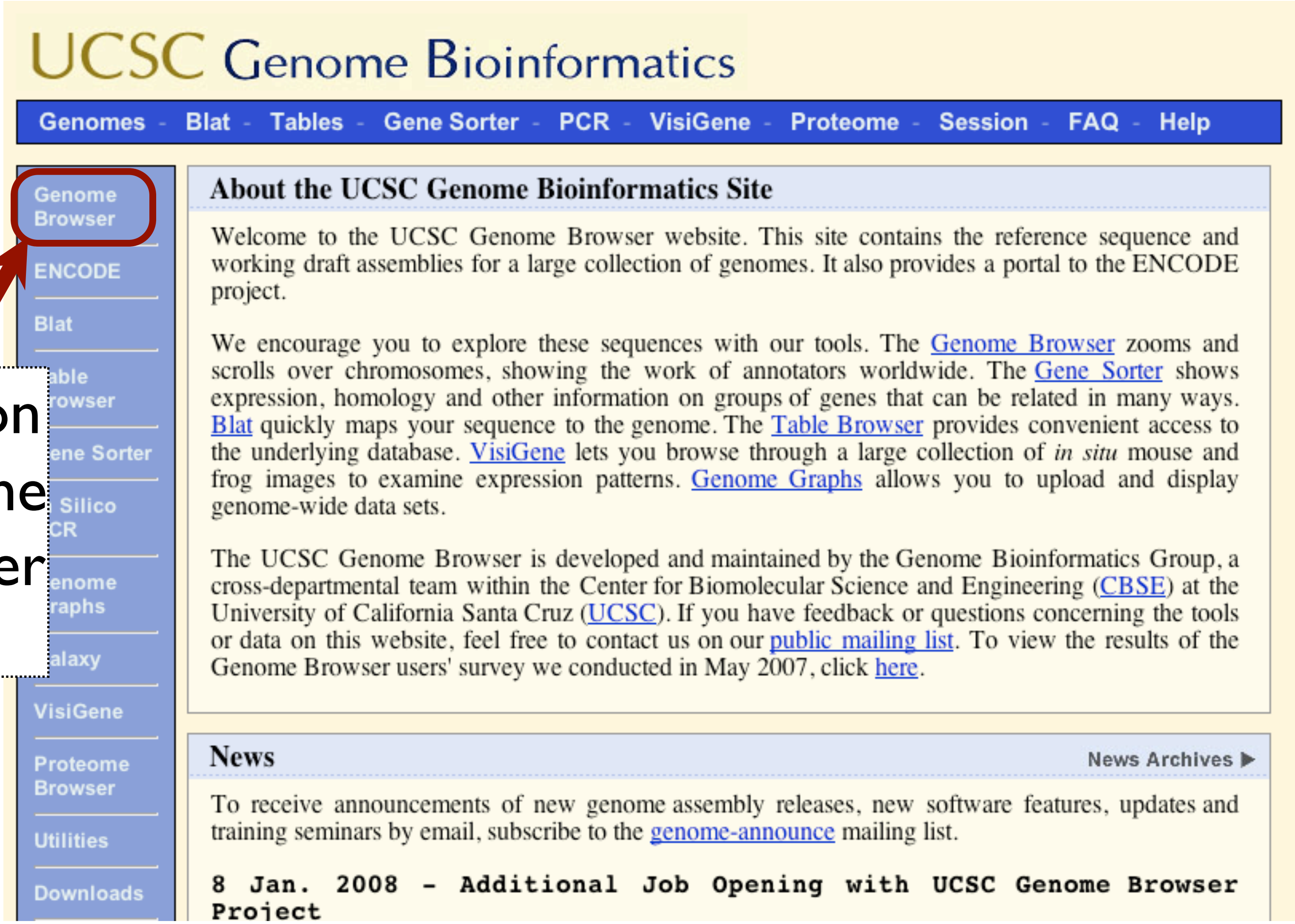
Alternatively:

FTP: Artemis V9 for UNIX
WWW: Artemis V9 for UNIX
FTP: Artemis V9 for MacOSX
WWW: Artemis V9 for MacOSX
FTP: Artemis V9 for Windows
WWW: Artemis V9 for Windows

Objectives

- By the end of today's lecture:
  - ✓You will practice using four different genome browsers.
  - ✓You will navigate the genomic region for BRCA1 using three different genome browsers.
  - ✓You will download genomic data for BRCA1
  - ✓You will learn about the photosynthetic gene cluster of *Rhodobacter sphaeroides* using Artemis

http://genome.cse.ucsc.edu/



About the UCSC Genome Bioinformatics Site

Welcome to the UCSC Genome Browser website. This site contains the reference sequence and working draft assemblies for a large collection of genomes. It also provides a portal to the ENCODE project.

We encourage you to explore these sequences with our tools. The Genome Browser zooms and scrolls over chromosomes, showing the work of annotators worldwide. The Gene Sorter shows expression, homology and other information on groups of genes that can be related in many ways. Blat quickly maps your sequence to the genome. The Table Browser provides convenient access to the underlying database. VisiGene lets you browse through a large collection of *in situ* mouse and frog images to examine expression patterns. Genome Graphs allows you to upload and display genome-wide data sets.

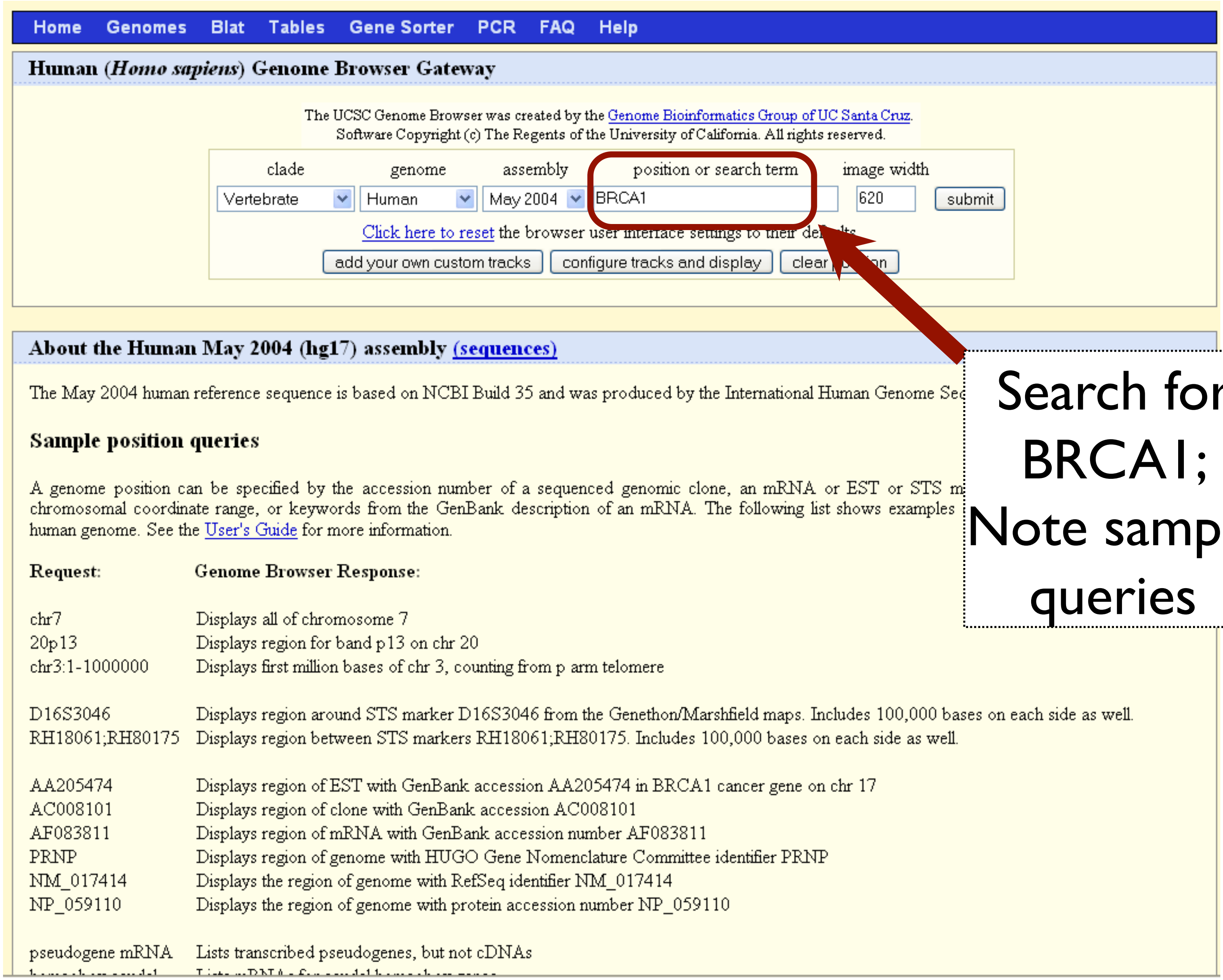
The UCSC Genome Browser is developed and maintained by the Genome Bioinformatics Group, a cross-departmental team within the Center for Biomolecular Science and Engineering (CBSE) at the University of California Santa Cruz (UCSC). If you have feedback or questions concerning the tools or data on this website, feel free to contact us on our public mailing list. To view the results of the Genome Browser users' survey we conducted in May 2007, click here.

News

To receive announcements of new genome assembly releases, new software features, updates and training seminars by email, subscribe to the genome-announce mailing list.

8 Jan. 2008 - Additional Job Opening with UCSC Genome Browser Project

Click on Genome Browser link



Human (*Homo sapiens*) Genome Browser Gateway

The UCSC Genome Browser was created by the Genome Bioinformatics Group of UC Santa Cruz.
Software Copyright (c) The Regents of the University of California. All rights reserved.

clade	genome	assembly	position or search term	image width
Vertebrate	Human	May 2004	BRCA1	620

Click here to reset the browser user interface settings to their defaults.

About the Human May 2004 (hg17) assembly (sequences)

The May 2004 human reference sequence is based on NCBI Build 35 and was produced by the International Human Genome Seq

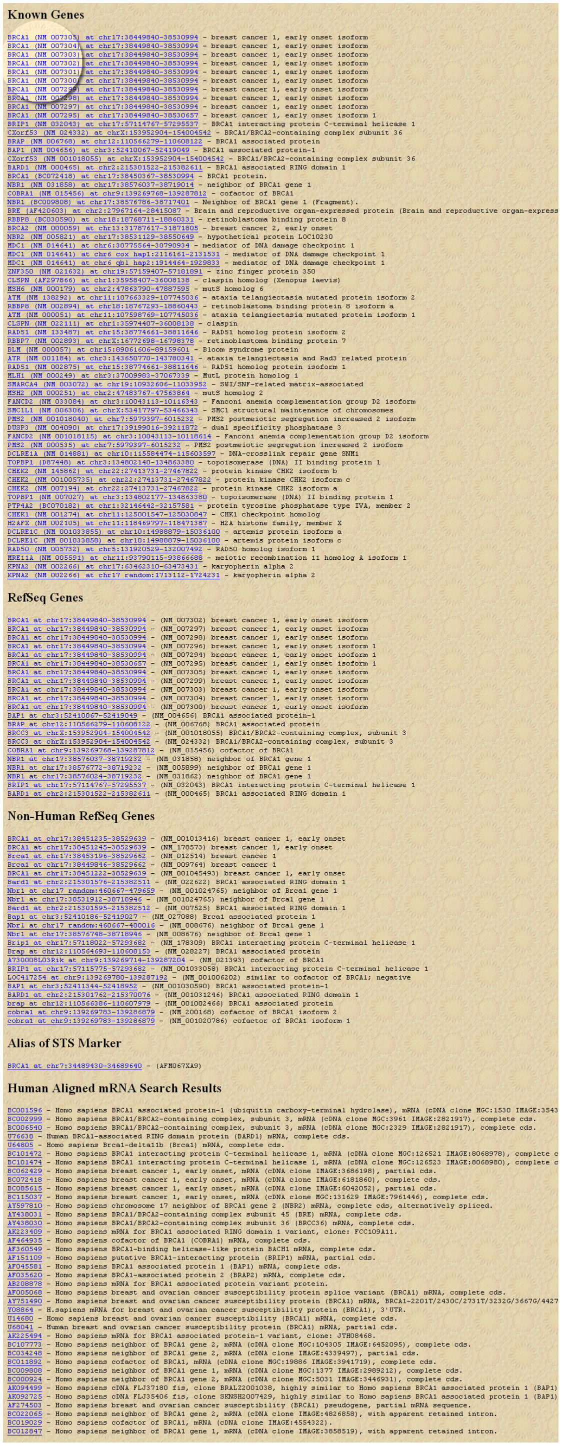
Sample position queries

A genome position can be specified by the accession number of a sequenced genomic clone, an mRNA or EST or STS m chromosomal coordinate range, or keywords from the GenBank description of an mRNA. The following list shows examples human genome. See the User's Guide for more information.

Request:	Genome Browser Response:
chr7	Displays all of chromosome 7
20p13	Displays region for band p13 on chr 20
chr3:1-1000000	Displays first million bases of chr 3, counting from p arm telomere
D16S3046	Displays region around STS marker D16S3046 from the Genethon/Marshfield maps. Includes 100,000 bases on each side as well.
RH18061;RH80175	Displays region between STS markers RH18061;RH80175. Includes 100,000 bases on each side as well.
AA205474	Displays region of EST with GenBank accession AA205474 in BRCA1 cancer gene on chr 17
AC008101	Displays region of clone with GenBank accession AC008101
AF083811	Displays region of mRNA with GenBank accession number AF083811
PRNP	Displays region of genome with HUGO Gene Nomenclature Committee identifier PRNP
NM_017414	Displays the region of genome with RefSeq identifier NM_017414
NP_059110	Displays the region of genome with protein accession number NP_059110
pseudogene mRNA	Lists transcribed pseudogenes, but not cDNAs

Search for BRCA1; Note sample queries

The Search Results



Known Genes

```
BRCA1 (NM 007305) at chr17:38449840-38530994 - breast cancer 1, early onset isoform
BRCA1 (NM 007304) at chr17:38449840-38530994 - breast cancer 1, early onset isoform
BRCA1 (NM 007303) at chr17:38449840-38530994 - breast cancer 1, early onset isoform
BRCA1 (NM 007302) at chr17:38449840-38530994 - breast cancer 1, early onset isoform
BRCA1 (NM 007301) at chr17:38449840-38530994 - breast cancer 1, early onset isoform
BRCA1 (NM 007300) at chr17:38449840-38530994 - breast cancer 1, early onset isoform
BRCA1 (NM 007299) at chr17:38449840-38530994 - breast cancer 1, early onset isoform
BRCA1 (NM 007298) at chr17:38449840-38530994 - breast cancer 1, early onset isoform
BRCA1 (NM 007297) at chr17:38449840-38530994 - breast cancer 1, early onset isoform
BRCA1 (NM 007295) at chr17:38449840-38530657 - breast cancer 1, early onset isoform 1
BRIP1 (NM 032043) at chr17:57114767-57295537 - BRCA1 interacting protein C-terminal helicase 1
CXorf53 (NM 024332) at chrX:153952904-154004542 - BRCA1/BRCA2-containing complex subunit 36
BRAP (NM 006768) at chr12:110566279-110608122 - BRCA1 associated protein
BAP1 (NM 004656) at chr3:52410067-52419049 - BRCA1 associated protein-1
CXorf53 (NM 001018055) at chrX:153952904-154004542 - BRCA1/BRCA2-containing complex subunit 36
BARD1 (NM 000465) at chr2:215301522-215382611 - BRCA1 associated RING domain 1
BRCA1 (BC072418) at chr17:38450367-38530994 - BRCA1 protein.
NBR1 (NM 031858) at chr17:38576037-38719014 - neighbor of BRCA1 gene 1
COBRA1 (NM 015456) at chr9:139269768-139287812 - cofactor of BRCA1
NBR1 (BC009808) at chr17:38576786-38717401 - Neighbor of BRCA1 gene 1 (Fragment).
BRE (AF420603) at chr2:27967164-28415087 - Brain and reproductive organ-expressed protein (Brain and reproductive organ-express
RBBP8 (BC030590) at chr18:18768711-18860331 - retinoblastoma binding protein 8
BRCA2 (NM 000059) at chr13:31787617-31871805 - breast cancer 2, early onset
NBR2 (NM 005821) at chr17:38531129-38550649 - hypothetical protein LOC10230
MDC1 (NM 014641) at chr6:30775564-30790934 - mediator of DNA damage checkpoint 1
MDC1 (NM 014641) at chr6 cox hap1:2116161-2131531 - mediator of DNA damage checkpoint 1
MDC1 (NM 014641) at chr6 qbl hap2:1914464-1929833 - mediator of DNA damage checkpoint 1
ZNF350 (NM 021632) at chr19:57159407-57181891 - zinc finger protein 350
CLSPN (AF297866) at chr1:35958407-36008138 - claspin homolog (Xenopus laevis)
MSH6 (NM 000179) at chr2:47863790-47887595 - mutS homolog 6
ATM (NM 138292) at chr11:107663329-107745036 - ataxia telangiectasia mutated protein isoform 2
RBBP8 (NM 002894) at chr18:18767293-18860443 - retinoblastoma binding protein 8 isoform a
ATM (NM 000051) at chr11:107598769-107745036 - ataxia telangiectasia mutated protein isoform 1
CLSPN (NM 022111) at chr1:35974407-36008138 - claspin
RAD51 (NM 133487) at chr15:38774661-38811646 - RAD51 homolog protein isoform 2
RBBP7 (NM 002893) at chrX:16772698-16798378 - retinoblastoma binding protein 7
BLM (NM 000057) at chr15:89061606-89159601 - Bloom syndrome protein
ATR (NM 001184) at chr3:143650770-143780341 - ataxia telangiectasia and Rad3 related protein
RAD51 (NM 002875) at chr15:38774661-38811646 - RAD51 homolog protein isoform 1
MLH1 (NM 000249) at chr3:37009983-37067339 - MutL protein homolog 1
SMARCA4 (NM 003072) at chr19:10932606-11033952 - SWI/SNF-related matrix-associated
MSH2 (NM 000251) at chr2:47483767-47563864 - mutS homolog 2
FANCD2 (NM 033084) at chr3:10043113-10116343 - Fanconi anemia complementation group D2 isoform
SMC1L1 (NM 006306) at chrX:53417797-53466343 - SMC1 structural maintenance of chromosomes
PMS2 (NM 001018040) at chr7:5979397-6015232 - PMS2 postmeiotic segregation increased 2 isoform
DUSP3 (NM 004090) at chr17:39199016-39211872 - dual specificity phosphatase 3
FANCD2 (NM 001018115) at chr3:10043113-10118614 - Fanconi anemia complementation group D2 isoform
PMS2 (NM 000535) at chr7:5979397-6015232 - PMS2 postmeiotic segregation increased 2 isoform
DCLRE1A (NM 014881) at chr10:115584474-115603597 - DNA-crosslink repair gene SNM1
TOPBP1 (D87448) at chr3:134802140-134863380 - topoisomerase (DNA) II binding protein 1
CHEK2 (NM 145862) at chr22:27413731-27467822 - protein kinase CHK2 isoform b
CHEK2 (NM 001005735) at chr22:27413731-27467822 - protein kinase CHK2 isoform c
CHEK2 (NM 007194) at chr22:27413731-27467822 - protein kinase CHK2 isoform a
TOPBP1 (NM 007027) at chr3:134802177-134863380 - topoisomerase (DNA) II binding protein 1
PTP4A2 (BC070182) at chr1:32146442-32157581 - protein tyrosine phosphatase type IVA, member 2
CHEK1 (NM 001274) at chr11:125001547-125030847 - CHK1 checkpoint homolog
H2AFX (NM 002105) at chr11:118469797-118471387 - H2A histone family, member X
DCLRE1C (NM 001033855) at chr10:14988879-15036100 - artemis protein isoform a
DCLRE1C (NM 001033858) at chr10:14988879-15036100 - artemis protein isoform c
RAD50 (NM 005732) at chr5:131920529-132007492 - RAD50 homolog isoform 1
MRE11A (NM 005591) at chr11:93790115-93866688 - meiotic recombination 11 homolog A isoform 1
KPNA2 (NM 002266) at chr17:63462310-63473431 - karyopherin alpha 2
KPNA2 (NM 002266) at chr17 random:1713112-1724231 - karyopherin alpha 2
```

RefSeq Genes

```
BRCA1 at chr17:38449840-38530994 - (NM_007302) breast cancer 1, early onset isoform
BRCA1 at chr17:38449840-38530994 - (NM_007297) breast cancer 1, early onset isoform
BRCA1 at chr17:38449840-38530994 - (NM_007298) breast cancer 1, early onset isoform
BRCA1 at chr17:38449840-38530994 - (NM_007296) breast cancer 1, early onset isoform 1
BRCA1 at chr17:38449840-38530994 - (NM_007294) breast cancer 1, early onset isoform 1
BRCA1 at chr17:38449840-38530657 - (NM_007295) breast cancer 1, early onset isoform 1
BRCA1 at chr17:38449840-38530994 - (NM_007305) breast cancer 1, early onset isoform
BRCA1 at chr17:38449840-38530994 - (NM_007299) breast cancer 1, early onset isoform
BRCA1 at chr17:38449840-38530994 - (NM_007303) breast cancer 1, early onset isoform
BRCA1 at chr17:38449840-38530994 - (NM_007304) breast cancer 1, early onset isoform
BRCA1 at chr17:38449840-38530994 - (NM_007300) breast cancer 1, early onset isoform
BAP1 at chr3:52410067-52419049 - (NM_004656) BRCA1 associated protein-1
BRAP at chr12:110566279-110608122 - (NM_006768) BRCA1 associated protein
BRCC3 at chrX:153952904-154004542 - (NM_001018055) BRCA1/BRCA2-containing complex, subunit 3
BRCC3 at chrX:153952904-154004542 - (NM_024332) BRCA1/BRCA2-containing complex, subunit 3
COBRA1 at chr9:139269768-139287812 - (NM_015456) cofactor of BRCA1
NBR1 at chr17:38576037-38719232 - (NM_031858) neighbor of BRCA1 gene 1
NBR1 at chr17:38576772-38719232 - (NM_005899) neighbor of BRCA1 gene 1
NBR1 at chr17:38576024-38719232 - (NM_031862) neighbor of BRCA1 gene 1
BRIP1 at chr17:57114767-57295537 - (NM_032043) BRCA1 interacting protein C-terminal helicase 1
BARD1 at chr2:215301522-215382611 - (NM_000465) BRCA1 associated RING domain 1
```

Non-Human RefSeq Genes

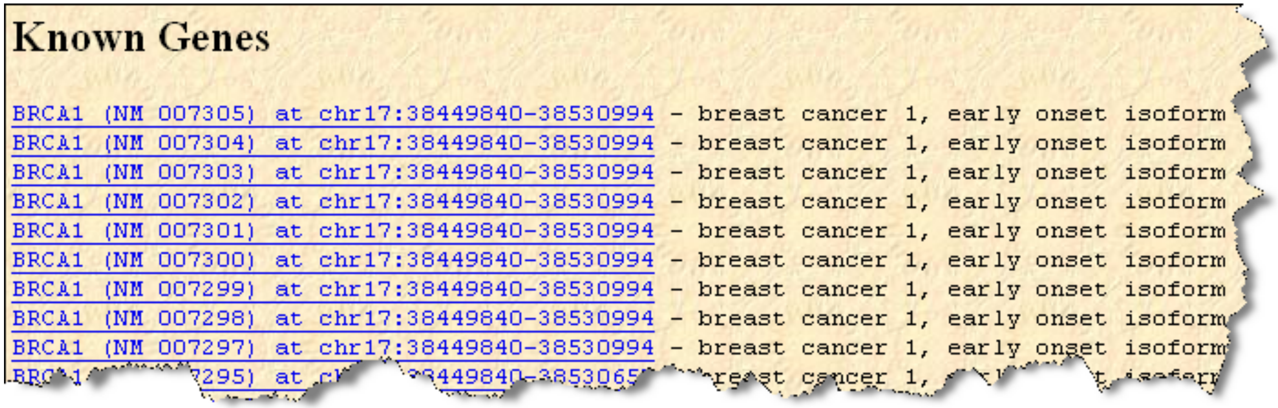
```
BRCA1 at chr17:38451235-38529639 - (NM_001013416) breast cancer 1, early onset
BRCA1 at chr17:38451245-38529639 - (NM_178573) breast cancer 1, early onset
Brca1 at chr17:38453196-38529662 - (NM_012514) breast cancer 1
Brca1 at chr17:38449846-38529662 - (NM_009764) breast cancer 1
BRCA1 at chr17:38451222-38529639 - (NM_001045493) breast cancer 1, early onset
Bard1 at chr2:215301576-215382511 - (NM_022622) BRCA1 associated RING domain 1
Nbr1 at chr17 random:460667-479659 - (NM_001024765) neighbor of Brca1 gene 1
Nbr1 at chr17:38531912-38718946 - (NM_001024765) neighbor of Brca1 gene 1
Bard1 at chr2:215301595-215382512 - (NM_007525) BRCA1 associated RING domain 1
Bap1 at chr3:52410186-52419027 - (NM_027088) Brca1 associated protein 1
Nbr1 at chr17 random:460667-480016 - (NM_008676) neighbor of Brca1 gene 1
Nbr1 at chr17:38576748-38718946 - (NM_008676) neighbor of Brca1 gene 1
Brip1 at chr17:57118022-57293682 - (NM_178309) BRCA1 interacting protein C-terminal helicase 1
Brap at chr12:110564693-110608153 - (NM_028227) BRCA1 associated protein
A730008L03Rik at chr9:139269714-139287204 - (NM_021393) cofactor of BRCA1
BRIP1 at chr17:57115775-57293682 - (NM_001033058) BRCA1 interacting protein C-terminal helicase 1
LOC417254 at chr9:139269780-139287192 - (NM_001006202) similar to cofactor of BRCA1; negative
BAP1 at chr3:52411344-52418952 - (NM_001030590) BRCA1 associated protein-1
BARD1 at chr2:215301762-215370076 - (NM_001031246) BRCA1 associated RING domain 1
brap at chr12:110566386-110607979 - (NM_001002466) BRCA1 associated protein
cobra1 at chr9:139269783-139286879 - (NM_200168) cofactor of BRCA1 isoform 2
cobra1 at chr9:139269783-139286879 - (NM_001020786) cofactor of BRCA1 isoform 1
```

Alias of STS Marker

```
BRCA1 at chr7:34489430-34689640 - (AFM067XA9)
```

Human Aligned mRNA Search Results

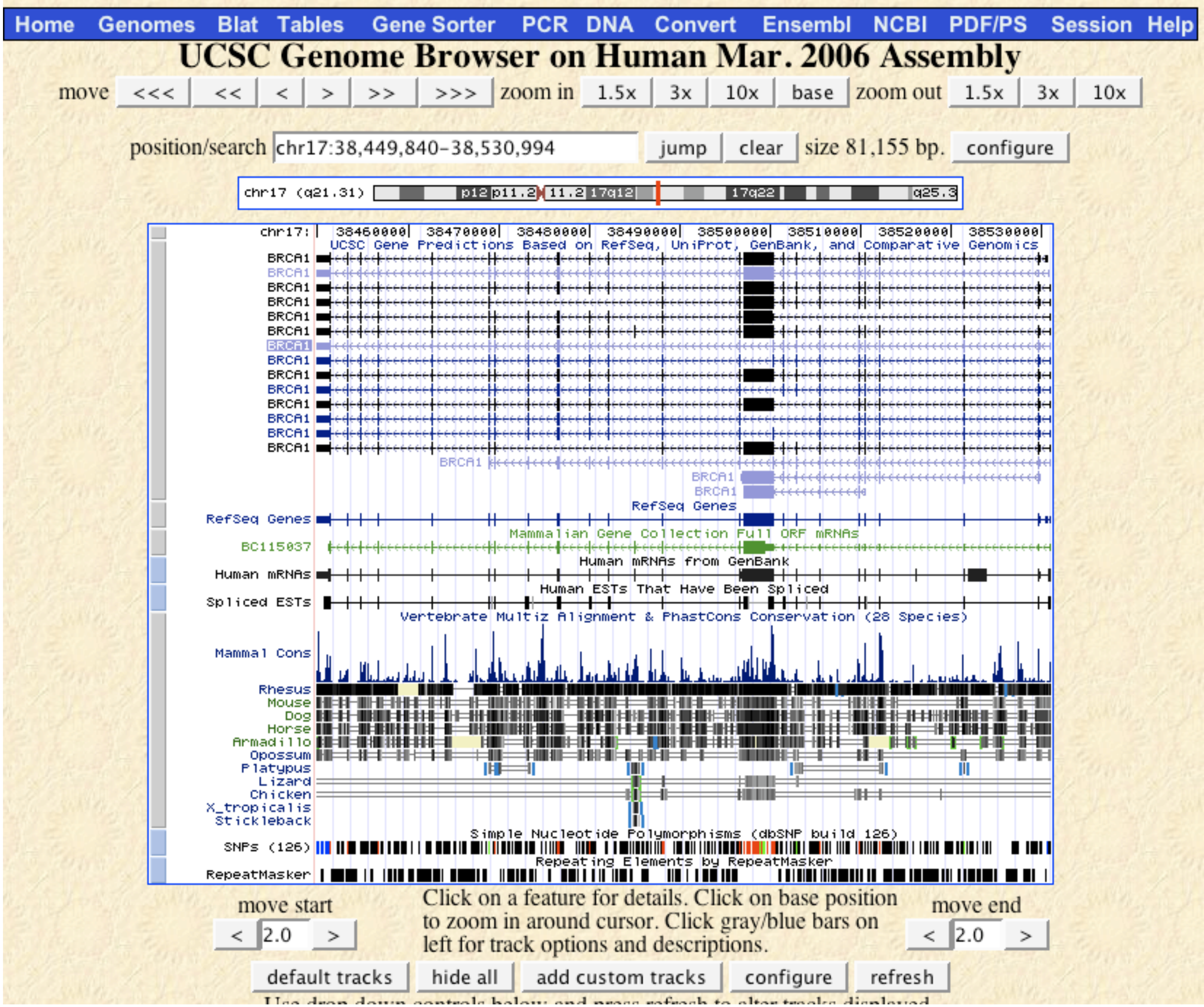
```
BC001596 - Homo sapiens BRCA1 associated protein-1 (ubiquitin carboxy-terminal hydrolase), mRNA (cDNA clone MGC:1530 IMAGE:3543
BC002999 - Homo sapiens BRCA1/BRCA2-containing complex, subunit 3, mRNA (cDNA clone MGC:3961 IMAGE:2821917), complete cds.
BC006540 - Homo sapiens BRCA1/BRCA2-containing complex, subunit 3, mRNA (cDNA clone MGC:2329 IMAGE:2821917), complete cds.
U76638 - Human BRCA1-associated RING domain protein (BARD1) mRNA, complete cds.
U64805 - Homo sapiens Brca1-delta11b (Brca1) mRNA, complete cds.
BC101472 - Homo sapiens BRCA1 interacting protein C-terminal helicase 1, mRNA (cDNA clone MGC:126521 IMAGE:8068978), complete c
BC101474 - Homo sapiens BRCA1 interacting protein C-terminal helicase 1, mRNA (cDNA clone MGC:126523 IMAGE:8068980), complete c
BC062429 - Homo sapiens breast cancer 1, early onset, mRNA (cDNA clone IMAGE:3686198), partial cds.
BC072418 - Homo sapiens breast cancer 1, early onset, mRNA (cDNA clone IMAGE:6181860), complete cds.
BC085615 - Homo sapiens breast cancer 1, early onset, mRNA (cDNA clone IMAGE:6042052), partial cds.
BC115037 - Homo sapiens breast cancer 1, early onset, mRNA (cDNA clone MGC:131629 IMAGE:7961446), complete cds.
AY597810 - Homo sapiens chromosome 17 neighbor of BRCA1 gene 2 (NBR2) mRNA, complete cds, alternatively spliced.
AY438031 - Homo sapiens BRCA1/BRCA2-containing complex subunit 45 (BRE) mRNA, complete cds.
AY438030 - Homo sapiens BRCA1/BRCA2-containing complex subunit 36 (BRCC36) mRNA, complete cds.
AK223409 - Homo sapiens mRNA for BRCA1 associated RING domain 1 variant, clone: FCC109A11.
AF464935 - Homo sapiens cofactor of BRCA1 (COBRA1) mRNA, complete cds.
AF360549 - Homo sapiens BRCA1-binding helicase-like protein BACH1 mRNA, complete cds.
AF151109 - Homo sapiens putative BRCA1-interacting protein (BRIP1) mRNA, partial cds.
AF045581 - Homo sapiens BRCA1 associated protein 1 (BAP1) mRNA, complete cds.
AF035620 - Homo sapiens BRCA1-associated protein 2 (BRAP2) mRNA, complete cds.
AB208878 - Homo sapiens mRNA for BRCA1 associated protein variant protein.
AF005068 - Homo sapiens breast and ovarian cancer susceptibility protein splice variant (BRCA1) mRNA, complete cds.
AY751490 - Homo sapiens breast and ovarian cancer susceptibility protein (BRCA1) mRNA, BRCA1-2201T/2430C/2731T/3232G/3667G/4427
Y08864 - H.sapiens mRNA for breast and ovarian cancer susceptibility protein (BRCA1), 3'UTR.
U14680 - Homo sapiens breast and ovarian cancer susceptibility (BRCA1) mRNA, complete cds.
U68041 - Human breast and ovarian cancer susceptibility protein (BRCA1) mRNA, partial cds.
AK225494 - Homo sapiens mRNA for BRCA1 associated protein-1 variant, clone: JTH08468.
BC107773 - Homo sapiens neighbor of BRCA1 gene 2, mRNA (cDNA clone MGC:104305 IMAGE:6452095), complete cds.
BC034248 - Homo sapiens neighbor of BRCA1 gene 2, mRNA (cDNA clone IMAGE:4339497), partial cds.
BC011892 - Homo sapiens cofactor of BRCA1, mRNA (cDNA clone MGC:19886 IMAGE:3941719), complete cds.
BC009808 - Homo sapiens neighbor of BRCA1 gene 1, mRNA (cDNA clone MGC:1377 IMAGE:2989212), complete cds.
BC000924 - Homo sapiens neighbor of BRCA1 gene 2, mRNA (cDNA clone MGC:5031 IMAGE:3446931), complete cds.
AK094499 - Homo sapiens cDNA FLJ37180 fis, clone BRALZ2001038, highly similar to Homo sapiens BRCA1 associated protein 1 (BAP1)
AK092725 - Homo sapiens cDNA FLJ35406 fis, clone SKNSH2007429, highly similar to Homo sapiens BRCA1 associated protein 1 (BAP1)
AF274503 - Homo sapiens breast and ovarian cancer susceptibility (BRCA1) pseudogene, partial mRNA sequence.
BC022065 - Homo sapiens neighbor of BRCA1 gene 2, mRNA (cDNA clone IMAGE:4826858), with apparent retained intron.
BC019029 - Homo sapiens cofactor of BRCA1, mRNA (cDNA clone IMAGE:4554322).
BC012847 - Homo sapiens neighbor of BRCA1 gene 1, mRNA (cDNA clone IMAGE:3858519), with apparent retained intron.
```



Known Genes

```
BRCA1 (NM 007305) at chr17:38449840-38530994 - breast cancer 1, early onset isoform
BRCA1 (NM 007304) at chr17:38449840-38530994 - breast cancer 1, early onset isoform
BRCA1 (NM 007303) at chr17:38449840-38530994 - breast cancer 1, early onset isoform
BRCA1 (NM 007302) at chr17:38449840-38530994 - breast cancer 1, early onset isoform
BRCA1 (NM 007301) at chr17:38449840-38530994 - breast cancer 1, early onset isoform
BRCA1 (NM 007300) at chr17:38449840-38530994 - breast cancer 1, early onset isoform
BRCA1 (NM 007299) at chr17:38449840-38530994 - breast cancer 1, early onset isoform
BRCA1 (NM 007298) at chr17:38449840-38530994 - breast cancer 1, early onset isoform
BRCA1 (NM 007297) at chr17:38449840-38530994 - breast cancer 1, early onset isoform
BRC 1      7295) at c    3449840-3853065   reast cancer 1,
```

- Many BRCA1 isoforms
  - ✓ All located on chr 17
  - ✓ same chr coordinates
  - ✓ different gene structures



UCSC Genome Browser on Human Mar. 2006 Assembly

position/search chr17:38,449,840-38,530,994

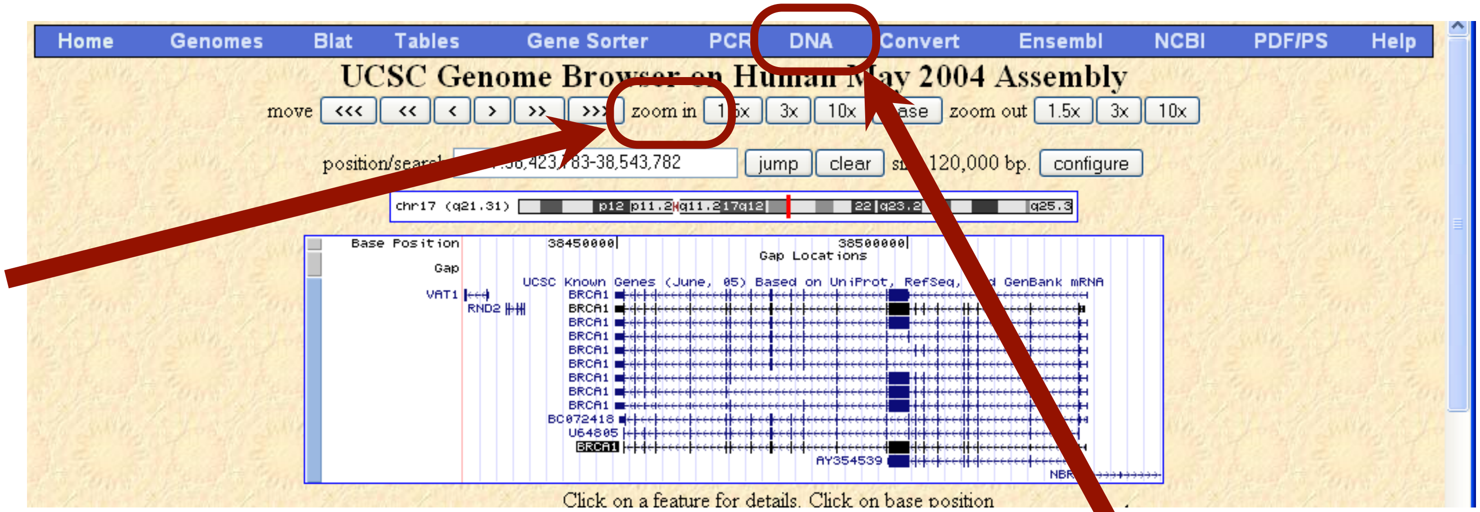
size 81,155 bp.

Click on a feature for details. Click on base position to zoom in around cursor. Click gray/blue bars on left for track options and descriptions.

Use drop down controls below and press refresh to alter tracks displayed.

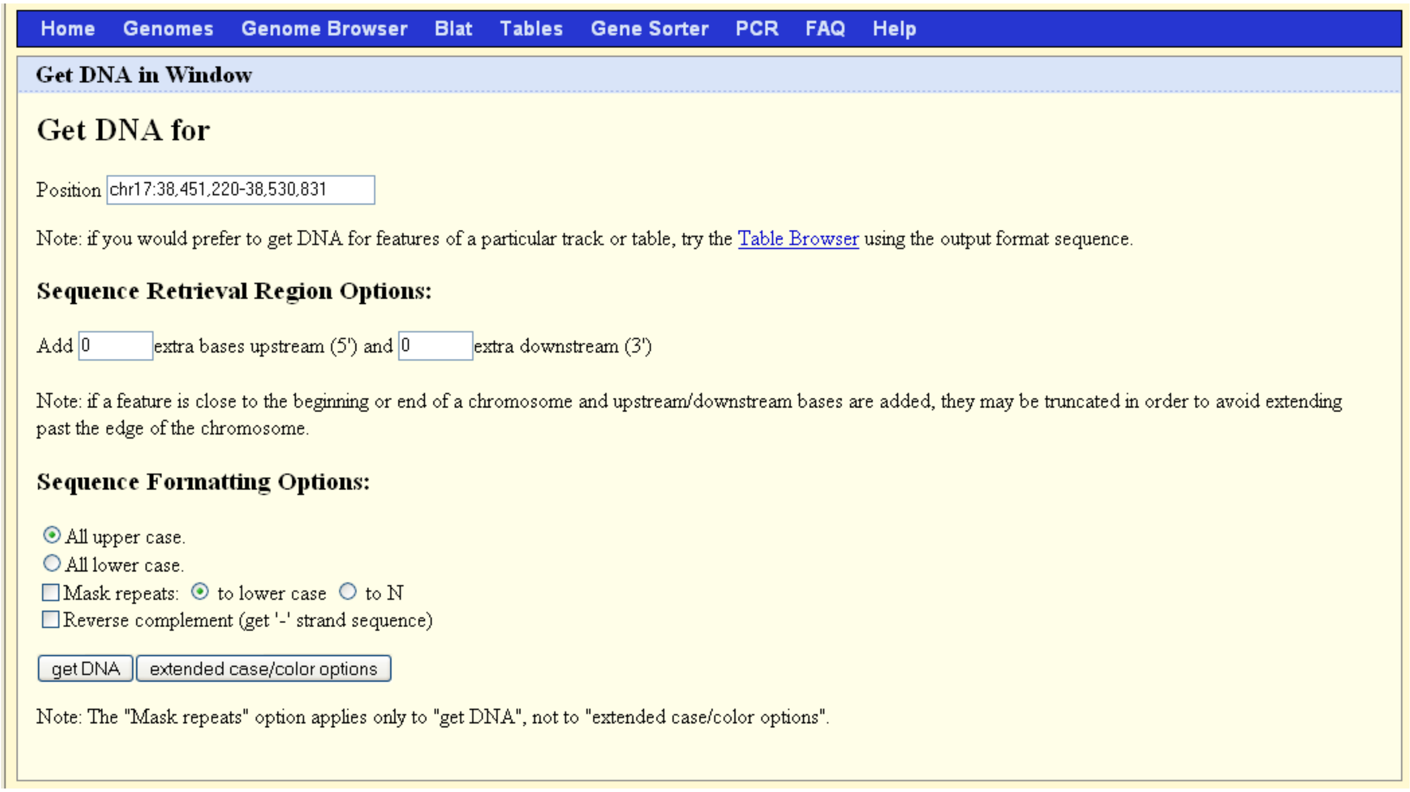
Tasks

- What genes are on either side of BRCA1 on chr 17?
- Can you figure out how to download the genomic sequence for the BRCA1 region?
- Can you figure the display to add/remove tracks that are (or are not) of interest to you?



Zoom in
Zoom out

DNA link
Download
Sequence



Get DNA in Window

Get DNA for

Position chr17:38,451,220-38,530,831

Note: if you would prefer to get DNA for features of a particular track or table, try the Table Browser using the output format sequence.

Sequence Retrieval Region Options:

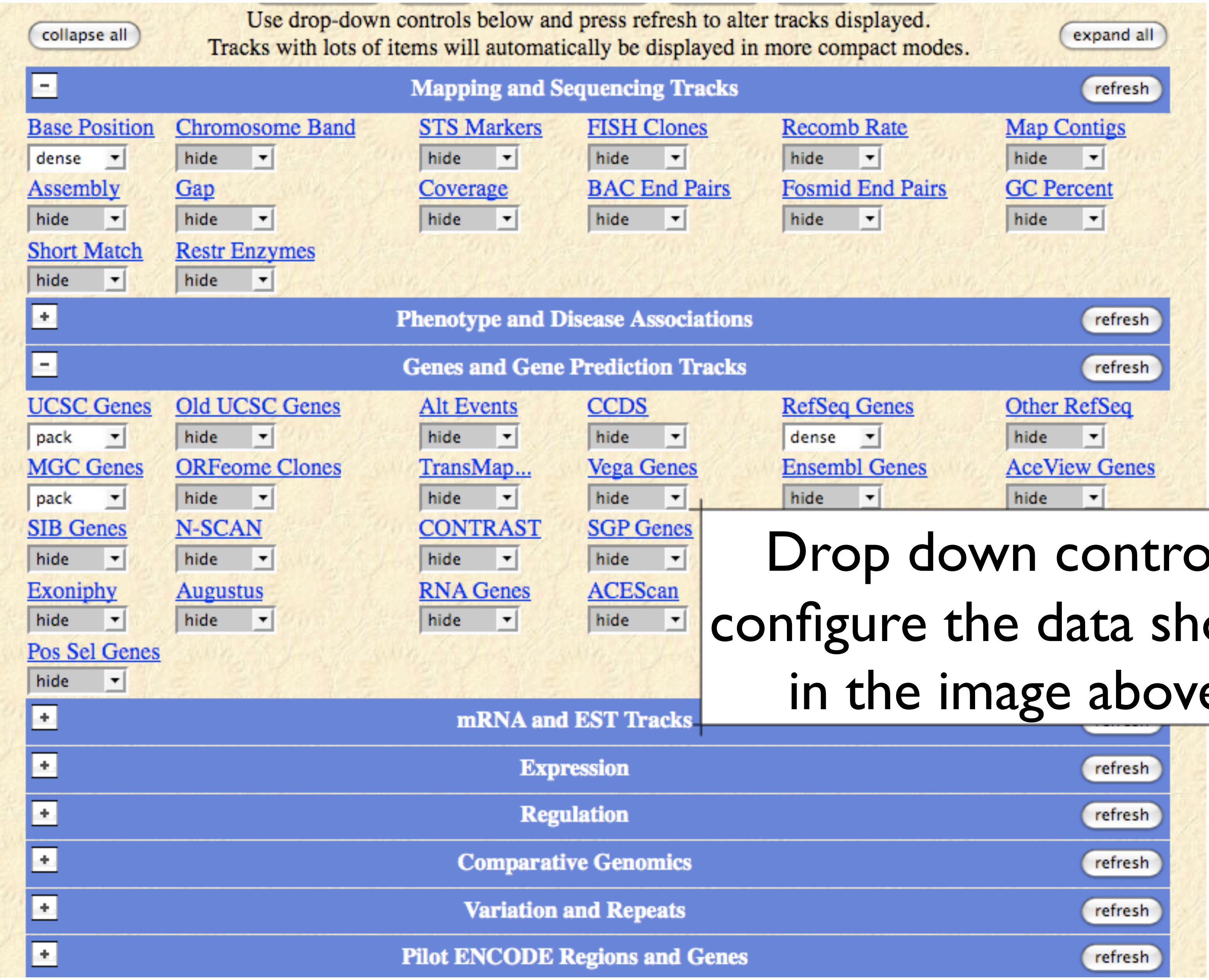
Add 0 extra bases upstream (5') and 0 extra downstream (3')

Note: if a feature is close to the beginning or end of a chromosome and upstream/downstream bases are added, they may be truncated in order to avoid extending past the edge of the chromosome.

Sequence Formatting Options:

● All upper case.
○ All lower case.
☐ Mask repeats: ● to lower case ○ to N
☐ Reverse complement (get '-' strand sequence)

Note: The "Mask repeats" option applies only to "get DNA", not to "extended case/color options".



Use drop-down controls below and press refresh to alter tracks displayed.
Tracks with lots of items will automatically be displayed in more compact modes.

Mapping and Sequencing Tracks

Base Position: dense
Chromosome Band: hide
STS Markers: hide
FISH Clones: hide
Recomb Rate: hide
Map Contigs: hide
Assembly: hide
Gap: hide
Coverage: hide
BAC End Pairs: hide
Fosmid End Pairs: hide
GC Percent: hide
Short Match: hide
Restr Enzymes: hide

Phenotype and Disease Associations

Genes and Gene Prediction Tracks

UCSC Genes: pack
Old UCSC Genes: hide
Alt Events: hide
CCDS: hide
RefSeq Genes: dense
Other RefSeq: hide
MGC Genes: pack
ORFeome Clones: hide
TransMap...: hide
Vega Genes: hide
Ensembl Genes: hide
AceView Genes: hide
SIB Genes: hide
N-SCAN: hide
CONTRAST: hide
SGP Genes: hide
Exoniphy: hide
Augustus: hide
RNA Genes: hide
ACEScan: hide
Pos Sel Genes: hide

mRNA and EST Tracks

Expression

Regulation

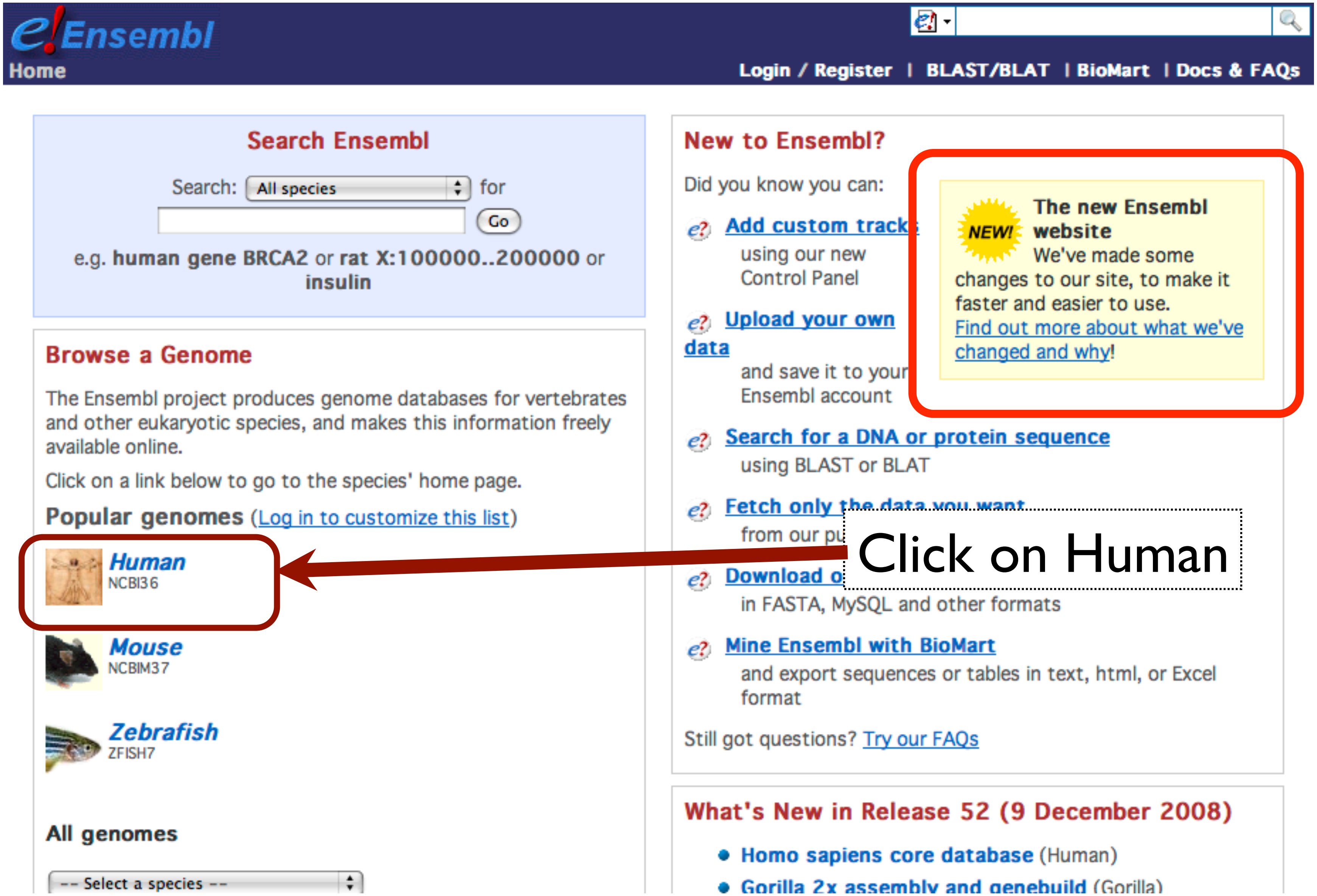
Comparative Genomics

Variation and Repeats

Pilot ENCODE Regions and Genes

Drop down controls configure the data shown in the image above

http://www.ensembl.org/



Search Ensembl

Search: All species for

Go

e.g. **human gene BRCA2** or **rat X:100000..200000** or **insulin**

Browse a Genome

The Ensembl project produces genome databases for vertebrates and other eukaryotic species, and makes this information freely available online.

Click on a link below to go to the species' home page.

Popular genomes (Log in to customize this list)

Human NCBI36

Mouse NCBIM37

Zebrafish ZFISH7

All genomes

-- Select a species --

New to Ensembl?

Did you know you can:

- **Add custom tracks** using our new Control Panel
- **Upload your own data** and save it to your Ensembl account
- **Search for a DNA or protein sequence** using BLAST or BLAT
- **Fetch only the data you want** from our pu
- **Download o** in FASTA, MySQL and other formats
- **Mine Ensembl with BioMart** and export sequences or tables in text, html, or Excel format

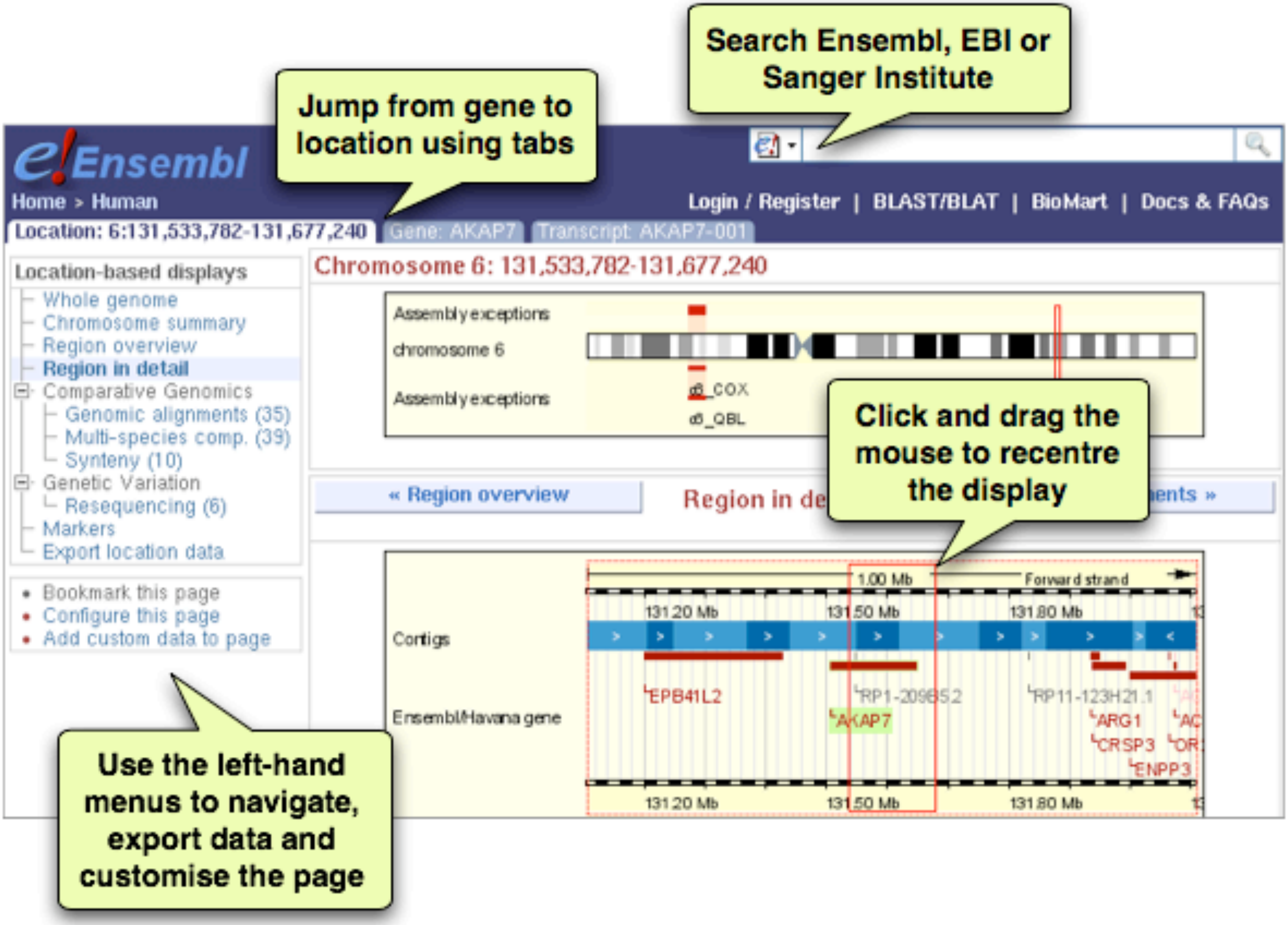
Still got questions? Try our FAQs

NEW! **The new Ensembl website** We've made some changes to our site, to make it faster and easier to use. Find out more about what we've changed and why!

What's New in Release 52 (9 December 2008)

- **Homo sapiens core database** (Human)
- **Gorilla 2x assembly and genebuild** (Gorilla)

Click on Human

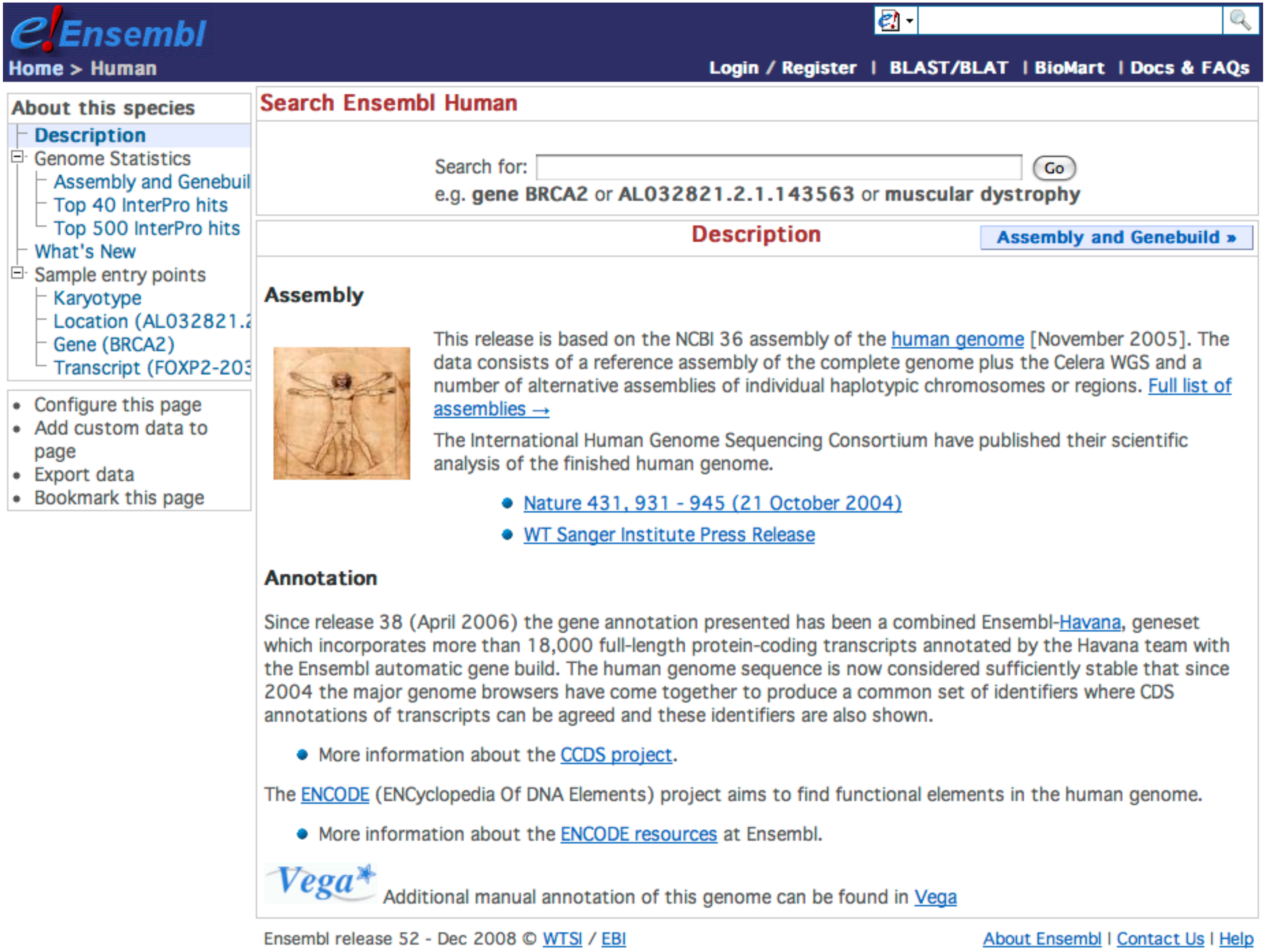


Search Ensembl, EBI or Sanger Institute

Jump from gene to location using tabs

Click and drag the mouse to recentre the display

Use the left-hand menus to navigate, export data and customise the page



Search Ensembl Human

Search for:

e.g. **gene BRCA2** or **AL032821.2.1.143563** or **muscular dystrophy**

Description

Assembly

This release is based on the NCBI 36 assembly of the human genome [November 2005]. The data consists of a reference assembly of the complete genome plus the Celera WGS and a number of alternative assemblies of individual haplotypic chromosomes or regions. Full list of assemblies →

The International Human Genome Sequencing Consortium have published their scientific analysis of the finished human genome.

- Nature 431, 931 - 945 (21 October 2004)
- WT Sanger Institute Press Release

Annotation

Since release 38 (April 2006) the gene annotation presented has been a combined Ensembl-Havana, geneset which incorporates more than 18,000 full-length protein-coding transcripts annotated by the Havana team with the Ensembl automatic gene build. The human genome sequence is now considered sufficiently stable that since 2004 the major genome browsers have come together to produce a common set of identifiers where CDS annotations of transcripts can be agreed and these identifiers are also shown.

- More information about the CCDS project.

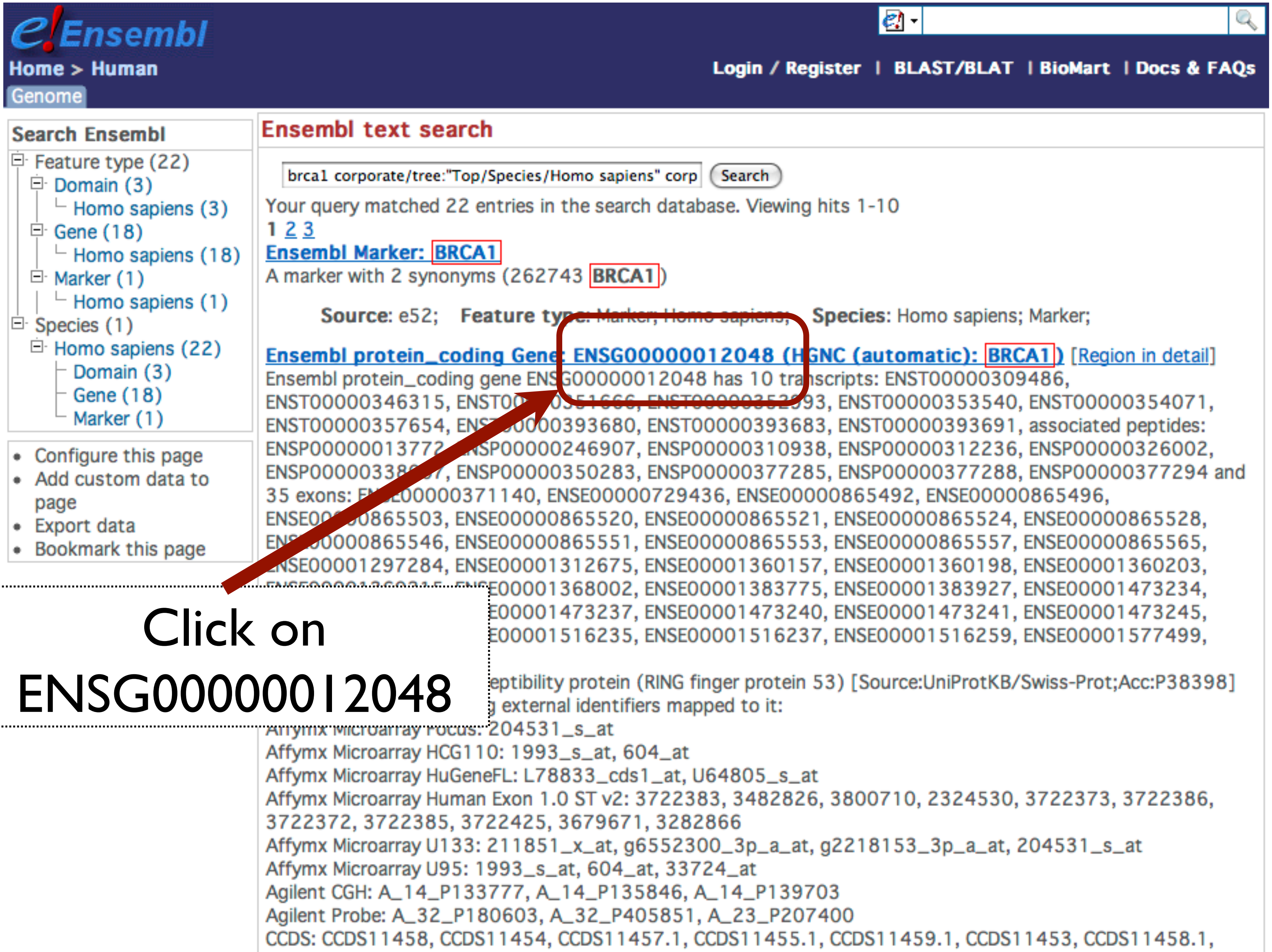
The ENCODE (ENCyclopedia Of DNA Elements) project aims to find functional elements in the human genome.

- More information about the ENCODE resources at Ensembl.

Additional manual annotation of this genome can be found in Vega

Ensembl release 52 - Dec 2008 © WTSI / EBI

About Ensembl | Contact Us | Help



Ensembl text search

brca1 corporate/tree:"Top/Species/Homo sapiens" corp

Your query matched 22 entries in the search database. Viewing hits 1-10

1 2 3

Ensembl Marker: BRCA1

A marker with 2 synonyms (262743 **BRCA1**)

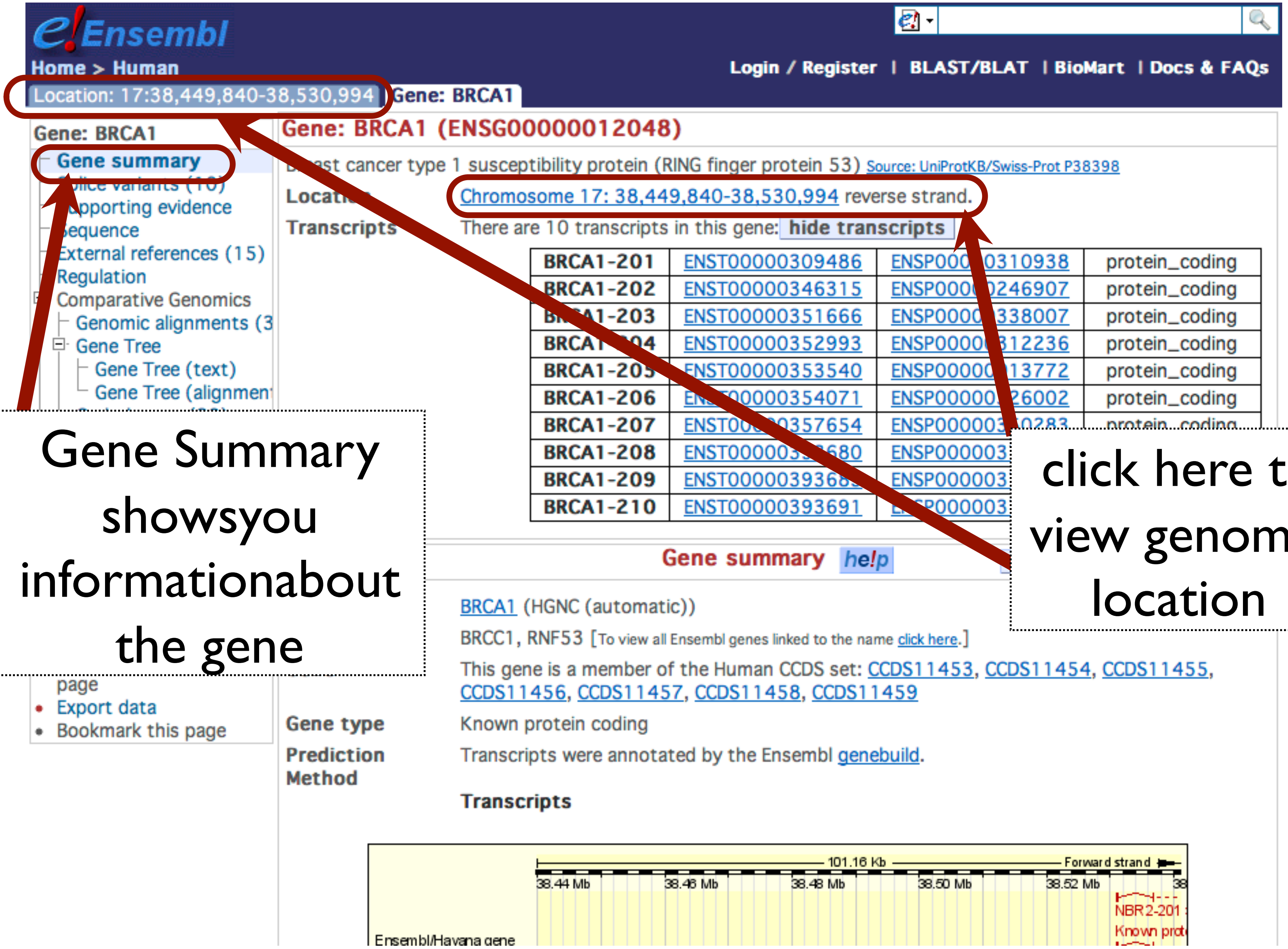
Source: e52; **Feature type:** Marker; Homo sapiens; **Species:** Homo sapiens; Marker;

Ensembl protein_coding Gene: ENSG00000012048 (HGNC (automatic): BRCA1) [Region in detail]

Ensembl protein_coding gene ENSG00000012048 has 10 transcripts: ENST00000309486, ENST00000346315, ENST00 0351666, ENST00000352 93, ENST00000353540, ENST00000354071, ENST00000357654, EN 0000393680, ENST00000393683, ENST00000393691, associated peptides: ENSP0000013772 NSP00000246907, ENSP00000310938, ENSP00000312236, ENSP00000326002, ENSP0000033 7, ENSP00000350283, ENSP00000377285, ENSP00000377288, ENSP00000377294 and 35 exons: E 00000371140, ENSE00000729436, ENSE00000865492, ENSE00000865496, ENSE00 0865503, ENSE00000865520, ENSE00000865521, ENSE00000865524, ENSE00000865528, EN 0000865546, ENSE00000865551, ENSE00000865553, ENSE00000865557, ENSE00000865565, NSE00001297284, ENSE00001312675, ENSE00001360157, ENSE00001360198, ENSE00001360203, E00001368002, ENSE00001383775, ENSE00001383927, ENSE00001473234, E00001473237, ENSE00001473240, ENSE00001473241, ENSE00001473245, E00001516235, ENSE00001516237, ENSE00001516259, ENSE00001577499,

eptibility protein (RING finger protein 53) [Source:UniProtKB/Swiss-Prot;Acc:P38398]
g external identifiers mapped to it:
Affymx Microarray Focus: 204531_s_at
Affymx Microarray HCG110: 1993_s_at, 604_at
Affymx Microarray HuGeneFL: L78833_cds1_at, U64805_s_at
Affymx Microarray Human Exon 1.0 ST v2: 3722383, 3482826, 3800710, 2324530, 3722373, 3722386, 3722372, 3722385, 3722425, 3679671, 3282866
Affymx Microarray U133: 211851_x_at, g6552300_3p_a_at, g2218153_3p_a_at, 204531_s_at
Affymx Microarray U95: 1993_s_at, 604_at, 33724_at
Agilent CGH: A_14_P133777, A_14_P135846, A_14_P139703
Agilent Probe: A_32_P180603, A_32_P405851, A_23_P207400
CCDS: CCDS11458, CCDS11454, CCDS11457.1, CCDS11455.1, CCDS11459.1, CCDS11453, CCDS11458.1,

Click on
ENSG00000012048



Gene: BRCA1

- Gene summary
- plice variants (10)
- pporting evidence
- equence
- External references (15)
- Regulation
- Comparative Genomics
  - Genomic alignments (3
  - Gene Tree
    - Gene Tree (text)
    - Gene Tree (alignmen

page

- Export data
- Bookmark this page

Gene: BRCA1 (ENSG00000012048)

ast cancer type 1 susceptibility protein (RING finger protein 53) Source: UniProtKB/Swiss-Prot P38398

Locati Chromosome 17: 38,449,840-38,530,994 reverse strand.

Transcripts There are 10 transcripts in this gene: **hide transcripts**

BRCA1-201	ENST00000309486	ENSP000 0310938	protein_coding
BRCA1-202	ENST00000346315	ENSP000 0246907	protein_coding
BR A1-203	ENST00000351666	ENSP0000 338007	protein_coding
BRCA1 04	ENST00000352993	ENSP0000 312236	protein_coding
BRCA1-20	NST00000353540	ENSP0000 13772	protein_coding
BRCA1-206	EN 00000354071	ENSP00000 26002	protein_coding
BRCA1-207	ENST00 357654	ENSP00000 0283	protein_coding
BRCA1-208	ENST00000 680	ENSP000003	
BRCA1-209	ENST00000 9368	ENSP000003	
BRCA1-210	ENST00000393691	P000003	

Gene summary he!p

BRCA1 (HGNC (automatic))

BRCC1, RNF53 [To view all Ensembl genes linked to the name click here.]

This gene is a member of the Human CCDS set: CCDS11453, CCDS11454, CCDS11455, CCDS11456, CCDS11457, CCDS11458, CCDS11459

Gene type Known protein coding

Prediction Method Transcripts were annotated by the Ensembl genebuild.

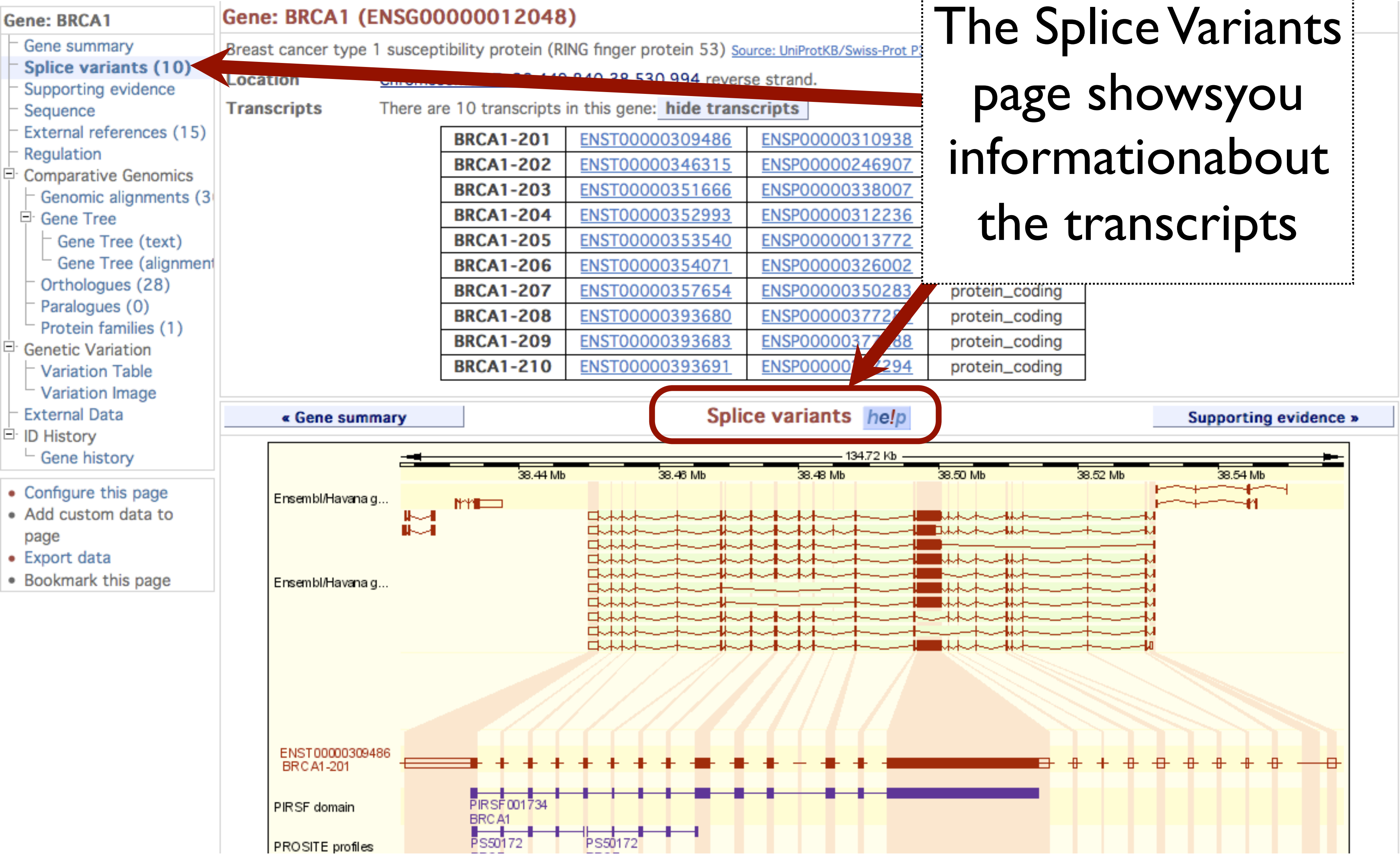
Transcripts

Gene Summary showsyou informationabout the gene

click here to view genomic location

Tasks

- Explore the information presented in the Gene Summary views.
  - Can you figure out how to visualize the alternatively spliced isoforms for BRCA1?
  - What can you find out about known variations in this gene?
- Using the Location Based Displays, can you figure out how to download the genomic sequence for the BRCA1 region?



Gene: BRCA1

- Gene summary
- **Splice variants (10)**
- Supporting evidence
- Sequence
- External references (15)
- Regulation
- Comparative Genomics
  - Genomic alignments (3
  - Gene Tree
    - Gene Tree (text)
    - Gene Tree (alignment
  - Orthologues (28)
  - Paralogues (0)
  - Protein families (1)
- Genetic Variation
  - Variation Table
  - Variation Image
- External Data
- ID History
  - Gene history

- Configure this page
- Add custom data to page
- Export data
- Bookmark this page

Gene: BRCA1 (ENSG00000012048)

Breast cancer type 1 susceptibility protein (RING finger protein 53) Source: UniProtKB/Swiss-Prot P

Location Chromosome 17: 38,449,840-38,530,994 reverse strand.

Transcripts There are 10 transcripts in this gene: **hide transcripts**

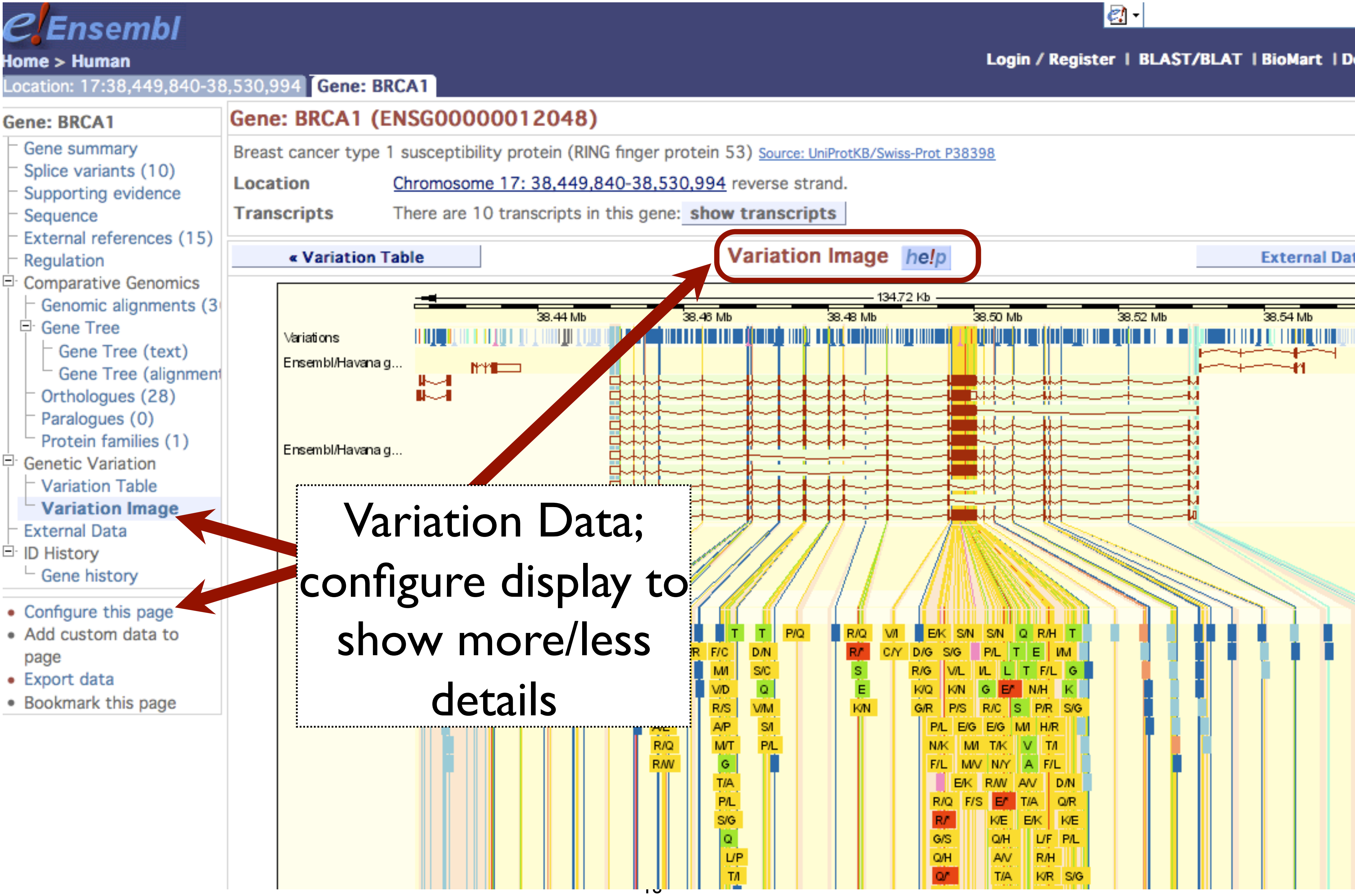
BRCA1-201	ENST00000309486	ENSP00000310938	
BRCA1-202	ENST00000346315	ENSP00000246907	
BRCA1-203	ENST00000351666	ENSP00000338007	
BRCA1-204	ENST00000352993	ENSP00000312236	
BRCA1-205	ENST00000353540	ENSP00000013772	
BRCA1-206	ENST00000354071	ENSP00000326002	
BRCA1-207	ENST00000357654	ENSP00000350283	protein_coding
BRCA1-208	ENST00000393680	ENSP0000037772[illegible]	protein_coding
BRCA1-209	ENST00000393683	ENSP0000037[illegible]88	protein_coding
BRCA1-210	ENST00000393691	ENSP00000[illegible]294	protein_coding

« Gene summary

Splice variants he!p

Supporting evidence »

The Splice Variants page showsyou informationabout the transcripts



Gene: BRCA1

- Gene summary
- Splice variants (10)
- Supporting evidence
- Sequence
- External references (15)
- Regulation
- Comparative Genomics
  - Genomic alignments (3
  - Gene Tree
    - Gene Tree (text)
    - Gene Tree (alignment
  - Orthologues (28)
  - Paralogues (0)
  - Protein families (1)
- Genetic Variation
  - Variation Table
  - **Variation Image**
- External Data
- ID History
  - Gene history

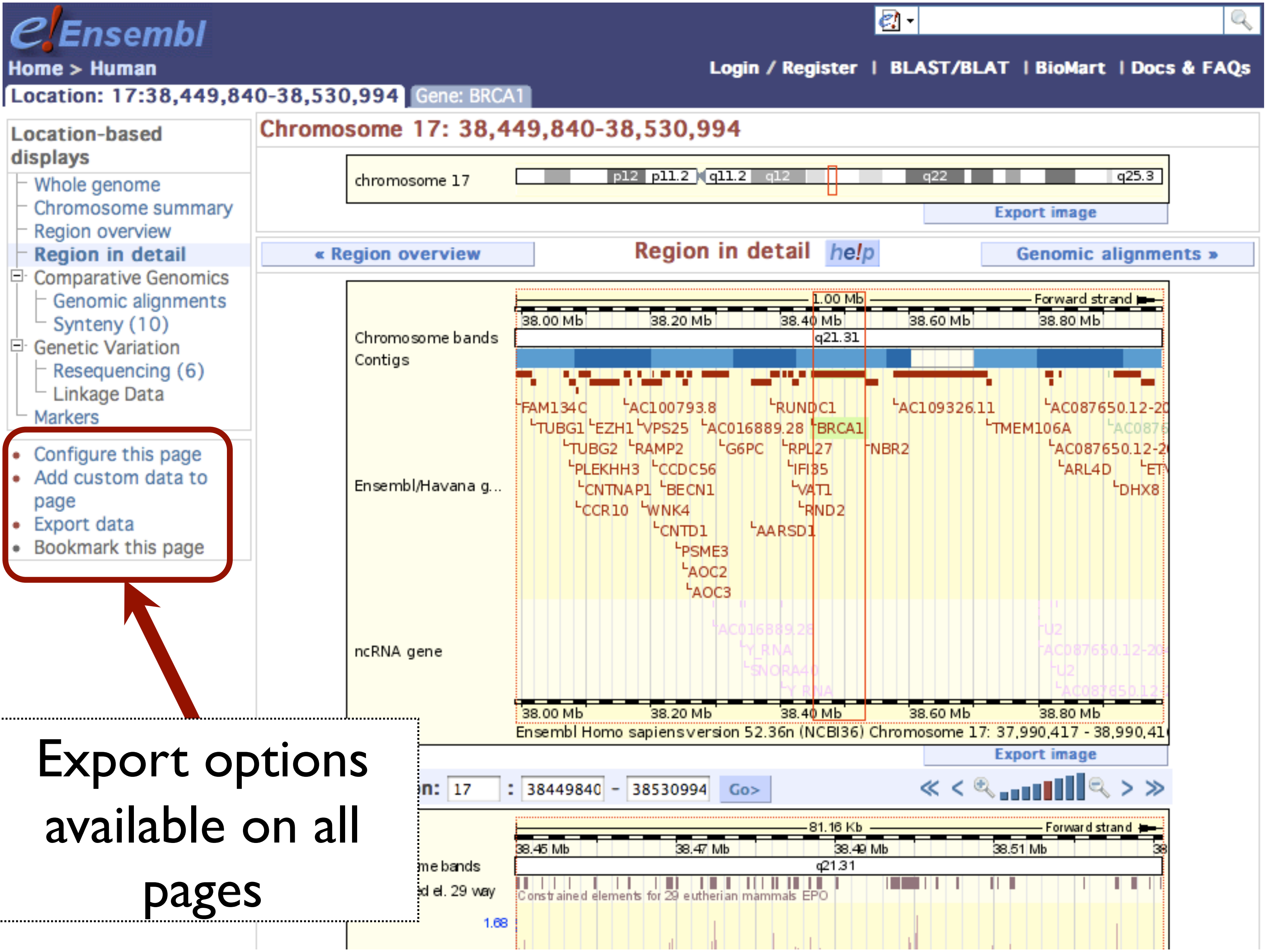
- Configure this page
- Add custom data to page
- Export data
- Bookmark this page

Gene: BRCA1 (ENSG00000012048)

Breast cancer type 1 susceptibility protein (RING finger protein 53) Source: UniProtKB/Swiss-Prot P38398

Location	Chromosome 17: 38,449,840-38,530,994 reverse strand.
Transcripts	There are 10 transcripts in this gene: show transcripts

Variation Data; configure display to show more/less details



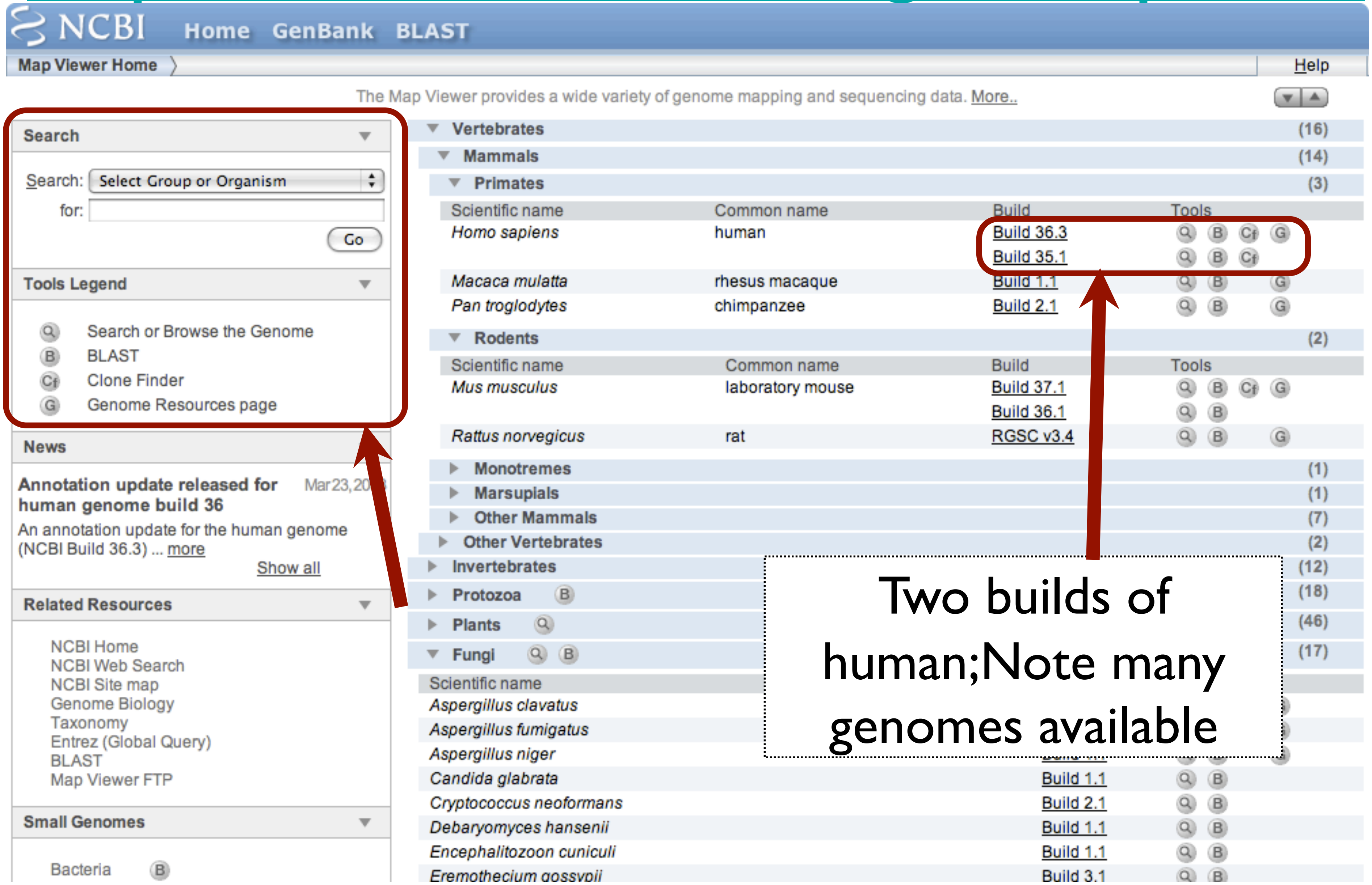
Export options available on all pages

On your own….

Follow the next few slides to explore the genomic region for BRCA1 using the

NCBI MapViewer system.

http://www.ncbi.nlm.nih.gov/mapview/



The Map Viewer provides a wide variety of genome mapping and sequencing data. More..

Search

Search: Select Group or Organism

for:

Go

Tools Legend

🔍	Search or Browse the Genome
B	BLAST
Cf	Clone Finder
G	Genome Resources page

News

Annotation update released for human genome build 36 Mar 23, 20

An annotation update for the human genome (NCBI Build 36.3) ... more

Show all

Related Resources

- NCBI Home
- NCBI Web Search
- NCBI Site map
- Genome Biology
- Taxonomy
- Entrez (Global Query)
- BLAST
- Map Viewer FTP

Small Genomes

Bacteria B

Vertebrates (16)

Mammals (14)

Primates (3)

Scientific name	Common name	Build	Tools
Homo sapiens	human	Build 36.3	🔍 B Cf G
		Build 35.1	🔍 B Cf
Macaca mulatta	rhesus macaque	Build 1.1	🔍 B G
Pan troglodytes	chimpanzee	Build 2.1	🔍 B G

Rodents (2)

Scientific name	Common name	Build	Tools
Mus musculus	laboratory mouse	Build 37.1	🔍 B Cf G
		Build 36.1	🔍 B
Rattus norvegicus	rat	RGSC v3.4	🔍 B G

Monotremes (1)

Marsupials (1)

Other Mammals (7)

Other Vertebrates (2)

Invertebrates (12)

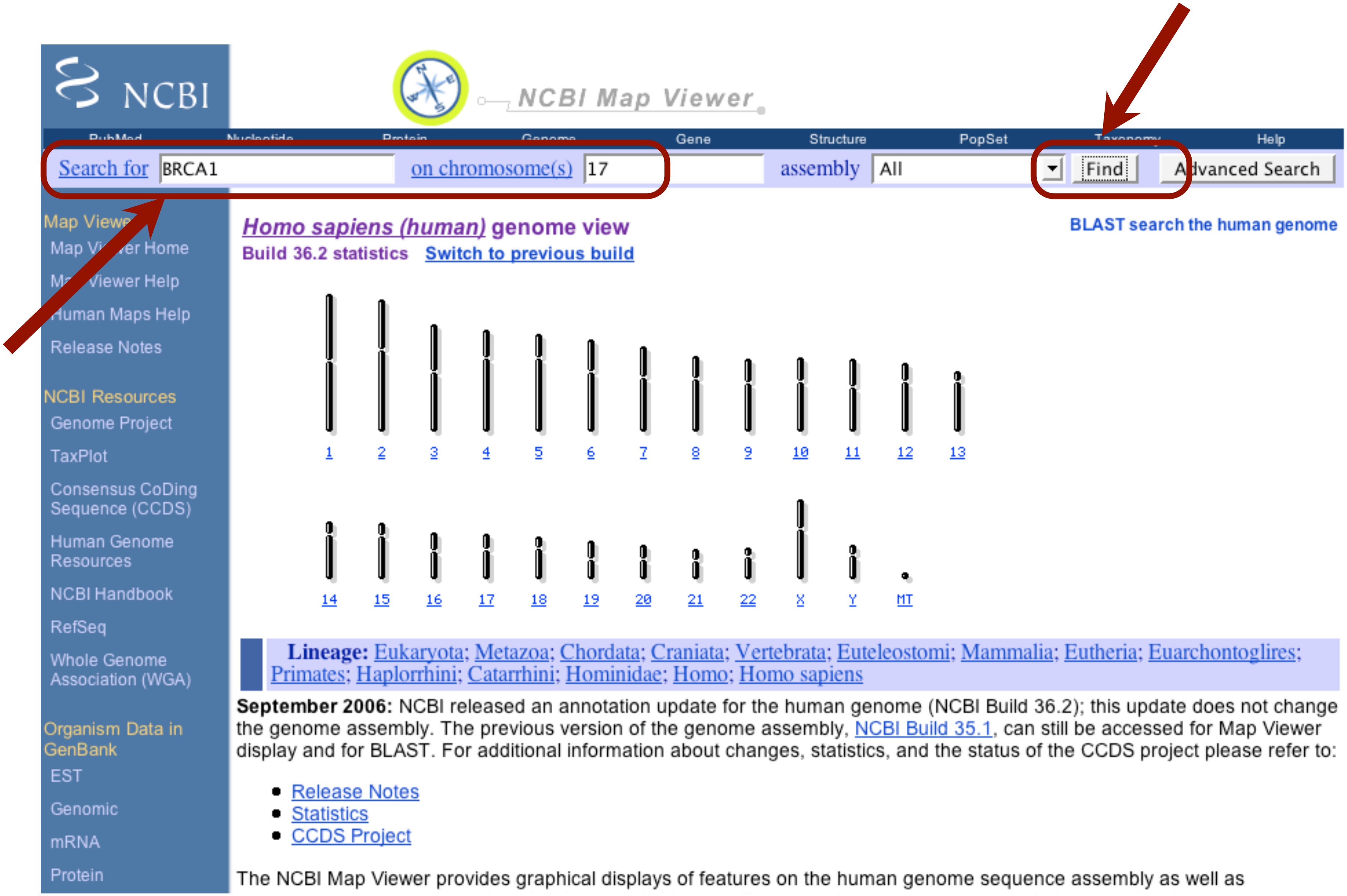
Protozoa B (18)

Plants 🔍 (46)

Fungi 🔍 B (17)

Scientific name	Build	Tools
Aspergillus clavatus		
Aspergillus fumigatus		
Aspergillus niger		
Candida glabrata	Build 1.1	🔍 B
Cryptococcus neoformans	Build 2.1	🔍 B
Debaryomyces hansenii	Build 1.1	🔍 B
Encephalitozoon cuniculi	Build 1.1	🔍 B
Eremothecium gossypii	Build 3.1	🔍 B

Two builds of human;Note many genomes available



Search for BRCA1 on chromosome(s) 17 assembly All Find

Homo sapiens (human) genome view

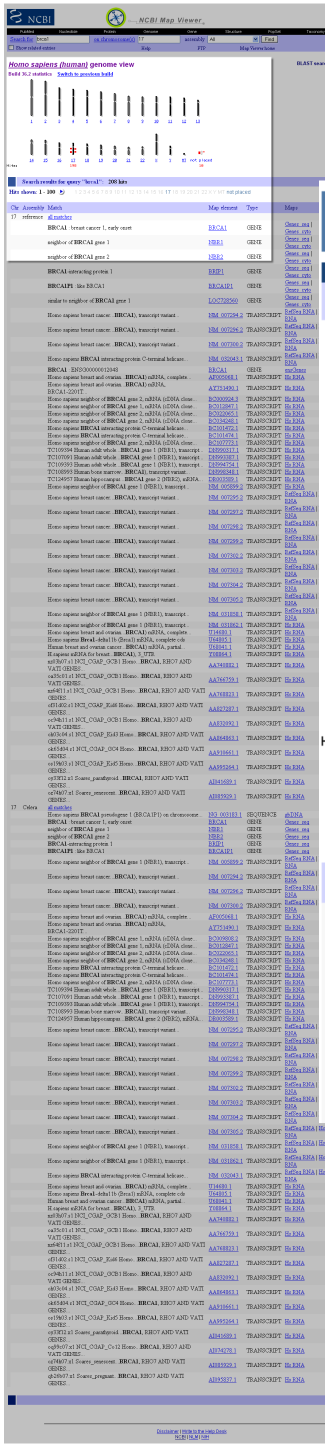
Build 36.2 statistics Switch to previous build

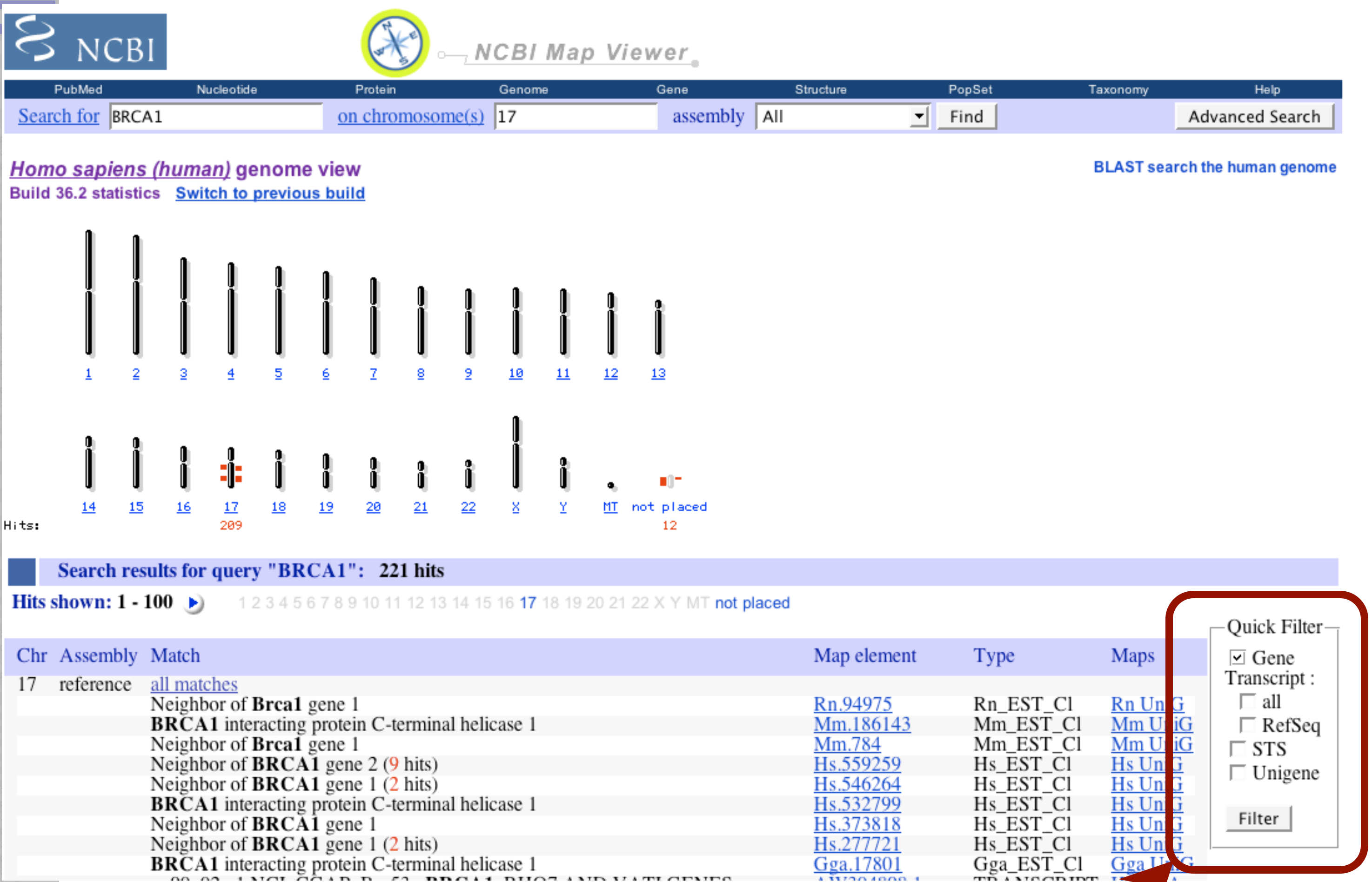
Lineage: Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo; Homo sapiens

September 2006: NCBI released an annotation update for the human genome (NCBI Build 36.2); this update does not change the genome assembly. The previous version of the genome assembly, NCBI Build 35.1, can still be accessed for Map Viewer display and for BLAST. For additional information about changes, statistics, and the status of the CCDS project please refer to:

- Release Notes
- Statistics
- CCDS Project

The NCBI Map Viewer provides graphical displays of features on the human genome sequence assembly as well as





Homo sapiens (human) genome view

Build 36.2 statistics Switch to previous build

BLAST search the human genome

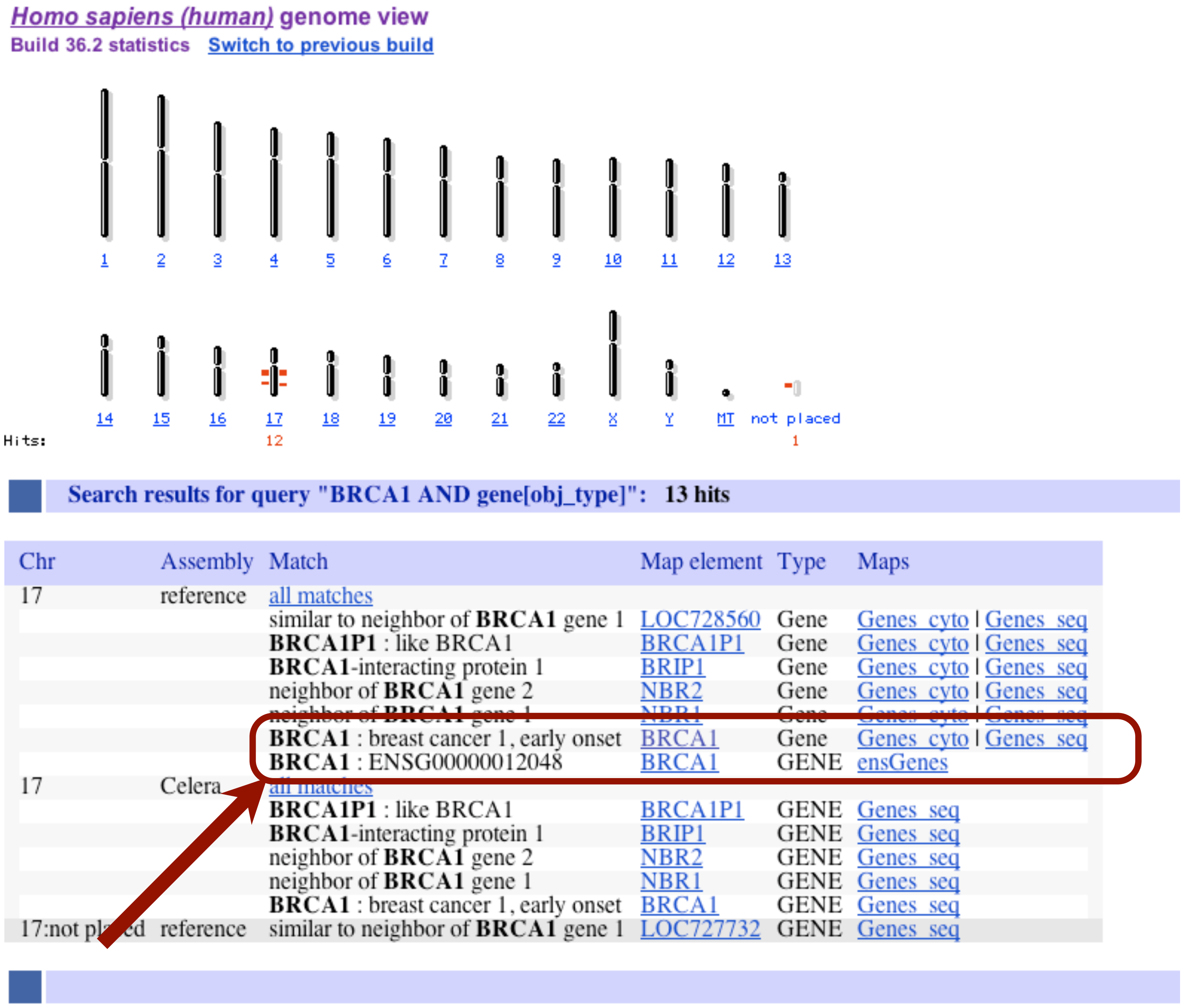
Search results for query "BRCA1": 221 hits

Hits shown: 1 - 100 ▶ 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X Y MT not placed

Chr	Assembly	Match	Map element	Type	Maps
17	reference	all matches			
		Neighbor of **Brca1** gene 1	Rn.94975	Rn_EST_Cl	Rn Un G
		BRCA1 interacting protein C-terminal helicase 1	Mm.186143	Mm_EST_Cl	Mm U iG
		Neighbor of **Brca1** gene 1	Mm.784	Mm_EST_Cl	Mm U iG
		Neighbor of **BRCA1** gene 2 (9 hits)	Hs.559259	Hs_EST_Cl	Hs Un G
		Neighbor of **BRCA1** gene 1 (2 hits)	Hs.546264	Hs_EST_Cl	Hs Un G
		BRCA1 interacting protein C-terminal helicase 1	Hs.532799	Hs_EST_Cl	Hs Un G
		Neighbor of **BRCA1** gene 1	Hs.373818	Hs_EST_Cl	Hs Un G
		Neighbor of **BRCA1** gene 1 (2 hits)	Hs.277721	Hs_EST_Cl	Hs Un G
		BRCA1 interacting protein C-terminal helicase 1	Gga.17801	Gga_EST_Cl	Gga U G

Quick Filter
☑ Gene
Transcript :
☐ all
☐ RefSeq
☐ STS
☐ Unigene
Filter

Quick Filter
✓ Gene

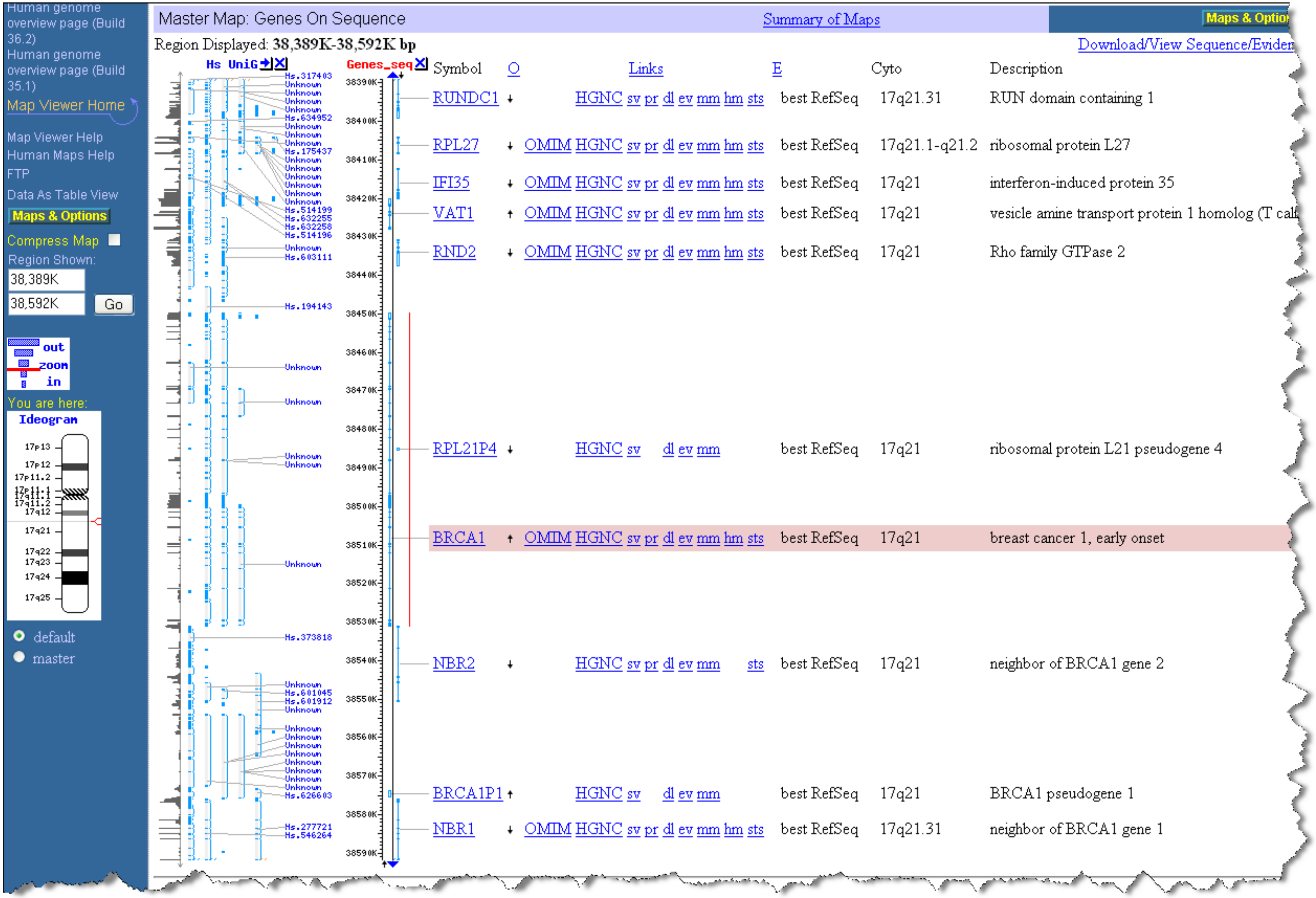


Homo sapiens (human) genome view

Build 36.2 statistics   Switch to previous build

Search results for query "BRCA1 AND gene[obj_type]": 13 hits

Chr	Assembly	Match	Map element	Type	Maps
17	reference	all matches			
		similar to neighbor of **BRCA1** gene 1	LOC728560	Gene	Genes_cyto | Genes_seq
		BRCA1P1 : like BRCA1	BRCA1P1	Gene	Genes_cyto | Genes_seq
		BRCA1-interacting protein 1	BRIP1	Gene	Genes_cyto | Genes_seq
		neighbor of **BRCA1** gene 2	NBR2	Gene	Genes_cyto | Genes_seq
		neighbor of **BRCA1** gene 1	NBR1	Gene	Genes_cyto | Genes_seq
		BRCA1 : breast cancer 1, early onset	BRCA1	Gene	Genes_cyto | Genes_seq
		BRCA1 : ENSG00000012048	BRCA1	GENE	ensGenes
17	Celera	all matches			
		BRCA1P1 : like BRCA1	BRCA1P1	GENE	Genes_seq
		BRCA1-interacting protein 1	BRIP1	GENE	Genes_seq
		neighbor of **BRCA1** gene 2	NBR2	GENE	Genes_seq
		neighbor of **BRCA1** gene 1	NBR1	GENE	Genes_seq
		BRCA1 : breast cancer 1, early onset	BRCA1	GENE	Genes_seq
17:not placed	reference	similar to neighbor of **BRCA1** gene 1	LOC727732	GENE	Genes_seq



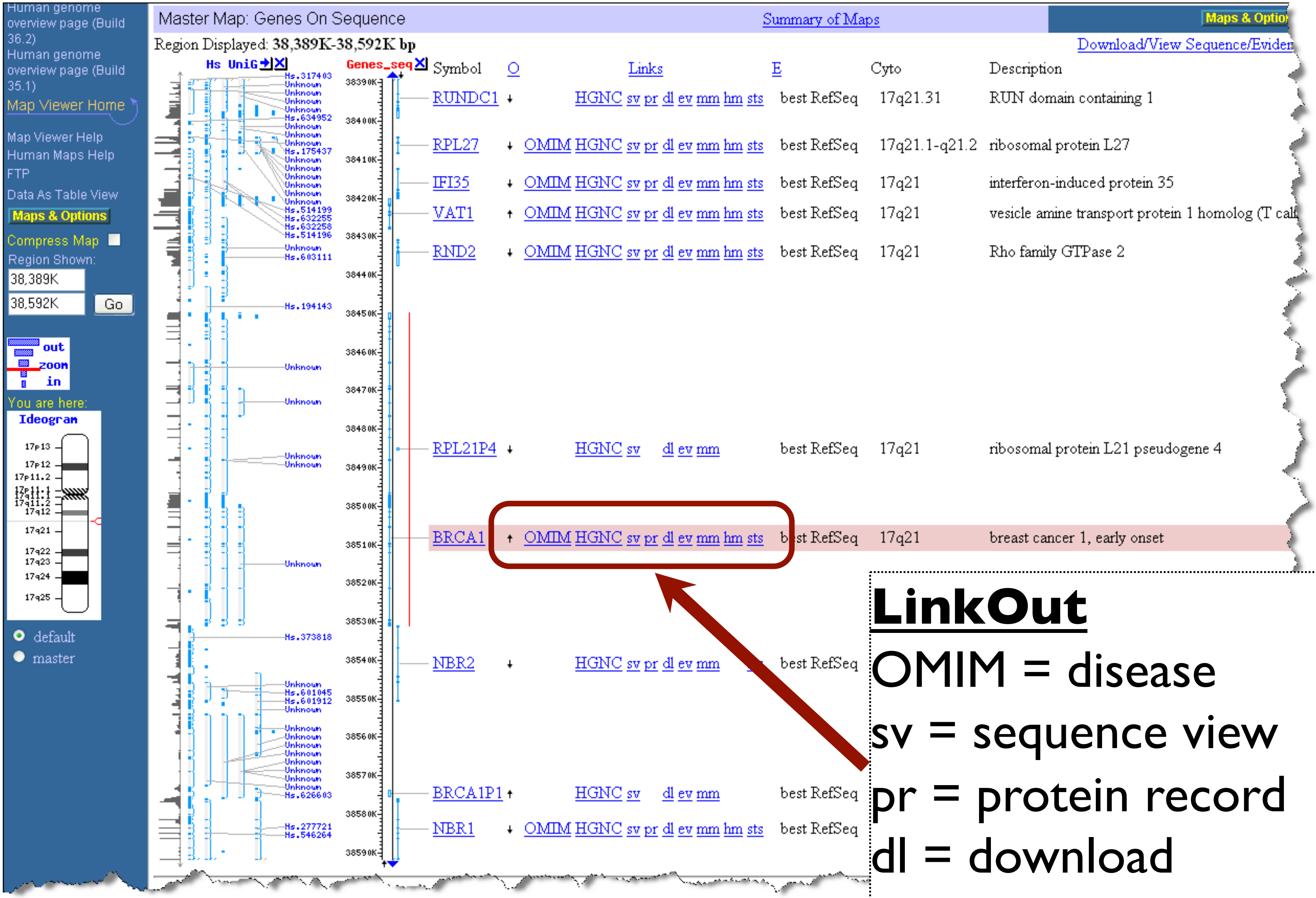
Master Map: Genes On Sequence

Region Displayed: **38,389K-38,592K bp**

Symbol	O	Links	E	Cyto	Description
RUNDC1	↓	HGNC sv pr dl ev mm hm sts	best RefSeq	17q21.31	RUN domain containing 1
RPL27	↓	OMIM HGNC sv pr dl ev mm hm sts	best RefSeq	17q21.1-q21.2	ribosomal protein L27
IFI35	↓	OMIM HGNC sv pr dl ev mm hm sts	best RefSeq	17q21	interferon-induced protein 35
VAT1	↑	OMIM HGNC sv pr dl ev mm hm sts	best RefSeq	17q21	vesicle amine transport protein 1 homolog (T cali
RND2	↓	OMIM HGNC sv pr dl ev mm hm sts	best RefSeq	17q21	Rho family GTPase 2
RPL21P4	↓	HGNC sv dl ev mm	best RefSeq	17q21	ribosomal protein L21 pseudogene 4
BRCA1	↑	OMIM HGNC sv pr dl ev mm hm sts	best RefSeq	17q21	breast cancer 1, early onset
NBR2	↓	HGNC sv pr dl ev mm sts	best RefSeq	17q21	neighbor of BRCA1 gene 2
BRCA1P1	↑	HGNC sv dl ev mm	best RefSeq	17q21	BRCA1 pseudogene 1
NBR1	↓	OMIM HGNC sv pr dl ev mm hm sts	best RefSeq	17q21.31	neighbor of BRCA1 gene 1

Two tasks

- Can you figure out how to LinkOut to the OMIM and/or Homologene entries for BRCA1?
- Can you figure out how to download the genomic sequence for the BRCA1 region?



Master Map: Genes On Sequence

Region Displayed: 38,389K-38,592K bp

Symbol	O	Links	E	Cyto	Description
RUNDC1 ↓		HGNC sv pr dl ev mm hm sts	best RefSeq	17q21.31	RUN domain containing 1
RPL27 ↓	OMIM	HGNC sv pr dl ev mm hm sts	best RefSeq	17q21.1-q21.2	ribosomal protein L27
IFI35 ↓	OMIM	HGNC sv pr dl ev mm hm sts	best RefSeq	17q21	interferon-induced protein 35
VAT1 ↑	OMIM	HGNC sv pr dl ev mm hm sts	best RefSeq	17q21	vesicle amine transport protein 1 homolog (T cal
RND2 ↓	OMIM	HGNC sv pr dl ev mm hm sts	best RefSeq	17q21	Rho family GTPase 2
RPL21P4 ↓		HGNC sv dl ev mm	best RefSeq	17q21	ribosomal protein L21 pseudogene 4
BRCA1 ↑	OMIM	HGNC sv pr dl ev mm hm sts	best RefSeq	17q21	breast cancer 1, early onset
NBR2 ↓		HGNC sv pr dl ev mm	best RefSeq		
BRCA1P1 ↑		HGNC sv dl ev mm	best RefSeq		
NBR1 ↓	OMIM	HGNC sv pr dl ev mm hm sts	best RefSeq		

LinkOut

OMIM = disease

sv = sequence view

pr = protein record

dl = download

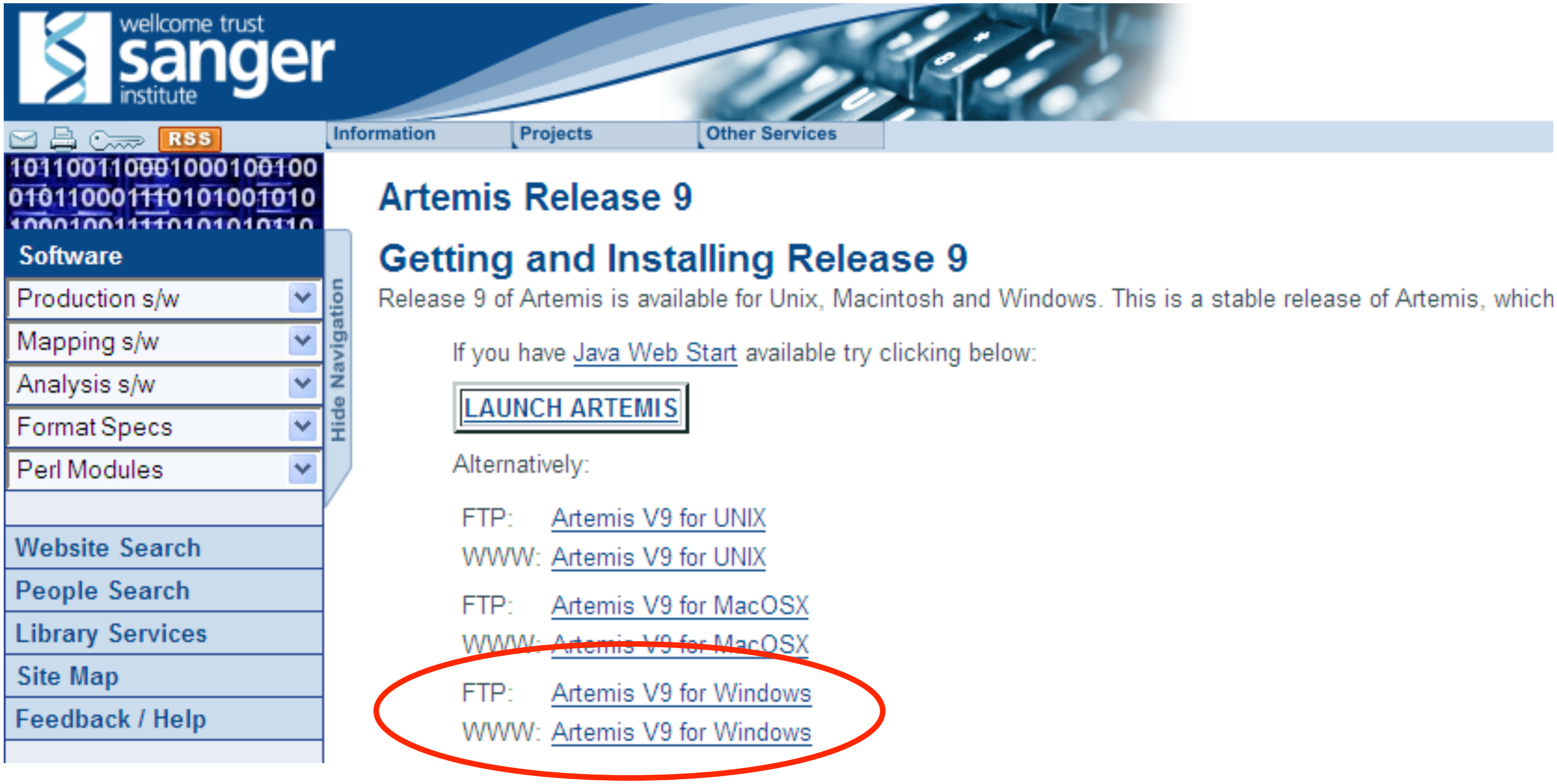
hm = Homologene

Artemis genome browser for prokaryotic genomes

Download Artemis to Desktop

http://www.sanger.ac.uk/Software/Artemis/v9/

Google: “artemis v9”



Artemis Release 9

Getting and Installing Release 9

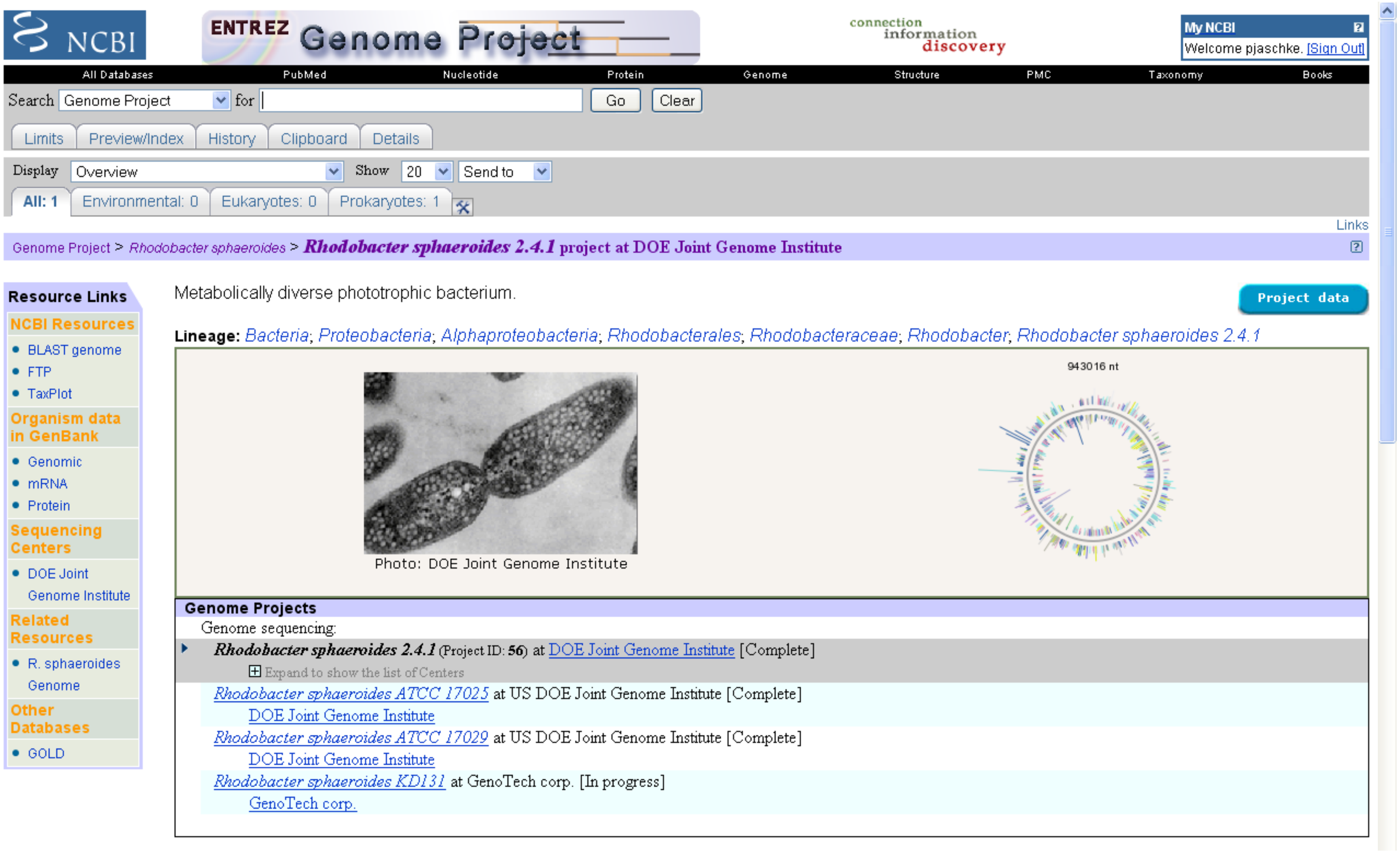
Release 9 of Artemis is available for Unix, Macintosh and Windows. This is a stable release of Artemis, which

If you have Java Web Start available try clicking below:

Alternatively:

FTP: Artemis V9 for UNIX
WWW: Artemis V9 for UNIX
FTP: Artemis V9 for MacOSX
WWW: Artemis V9 for MacOSX
FTP: Artemis V9 for Windows
WWW: Artemis V9 for Windows

Rhodobacter sphaeroides



Genome Project > *Rhodobacter sphaeroides* > ***Rhodobacter sphaeroides 2.4.1*** **project at DOE Joint Genome Institute**

Metabolically diverse phototrophic bacterium.

Lineage: *Bacteria*; *Proteobacteria*; *Alphaproteobacteria*; *Rhodobacterales*; *Rhodobacteraceae*; *Rhodobacter*; *Rhodobacter sphaeroides 2.4.1*

Photo: DOE Joint Genome Institute

Genome Projects

Genome sequencing:

Rhodobacter sphaeroides 2.4.1 (Project ID: **56**) at DOE Joint Genome Institute [Complete]

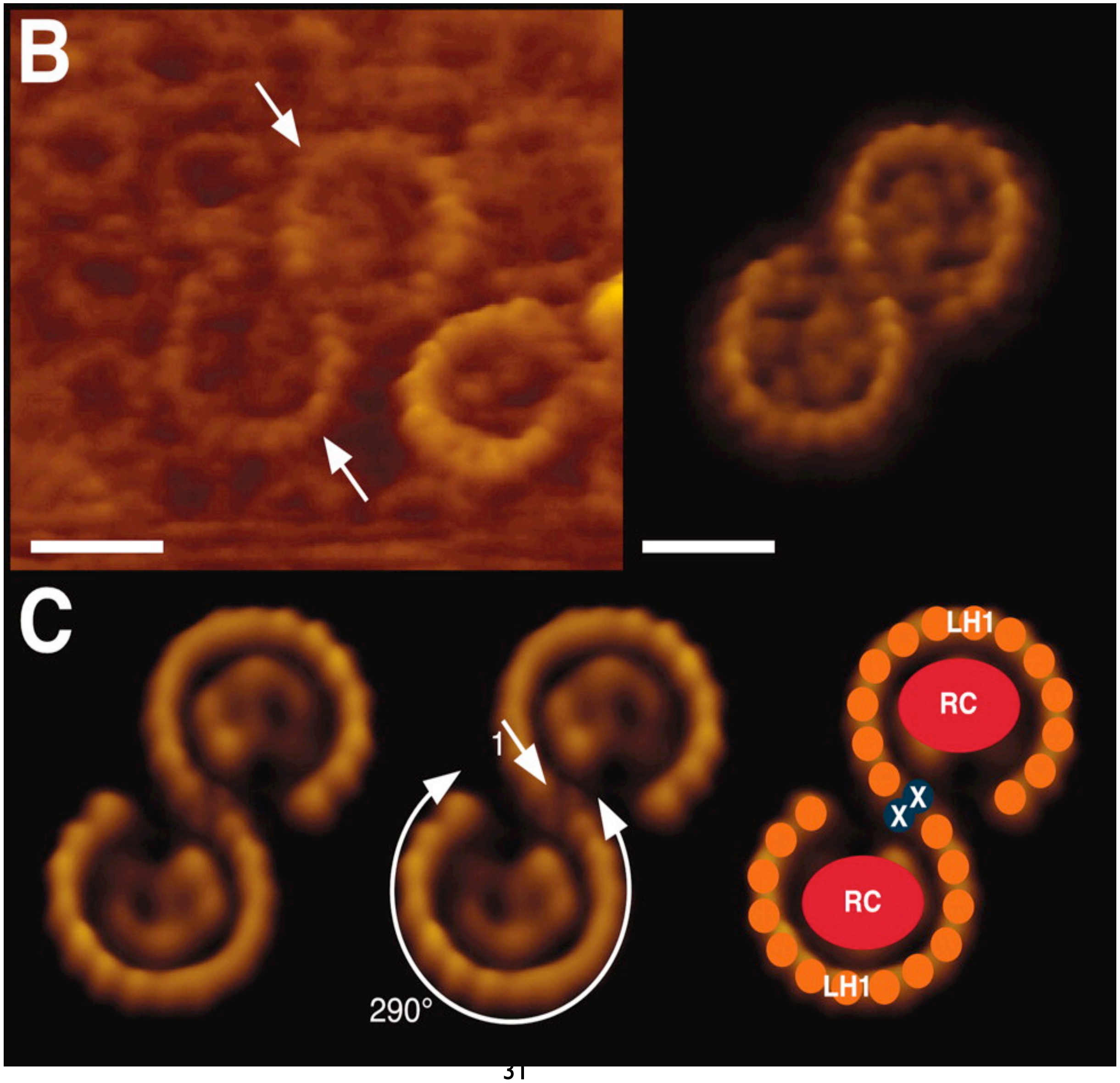
⊞ Expand to show the list of Centers

Rhodobacter sphaeroides ATCC 17025 at US DOE Joint Genome Institute [Complete]
DOE Joint Genome Institute

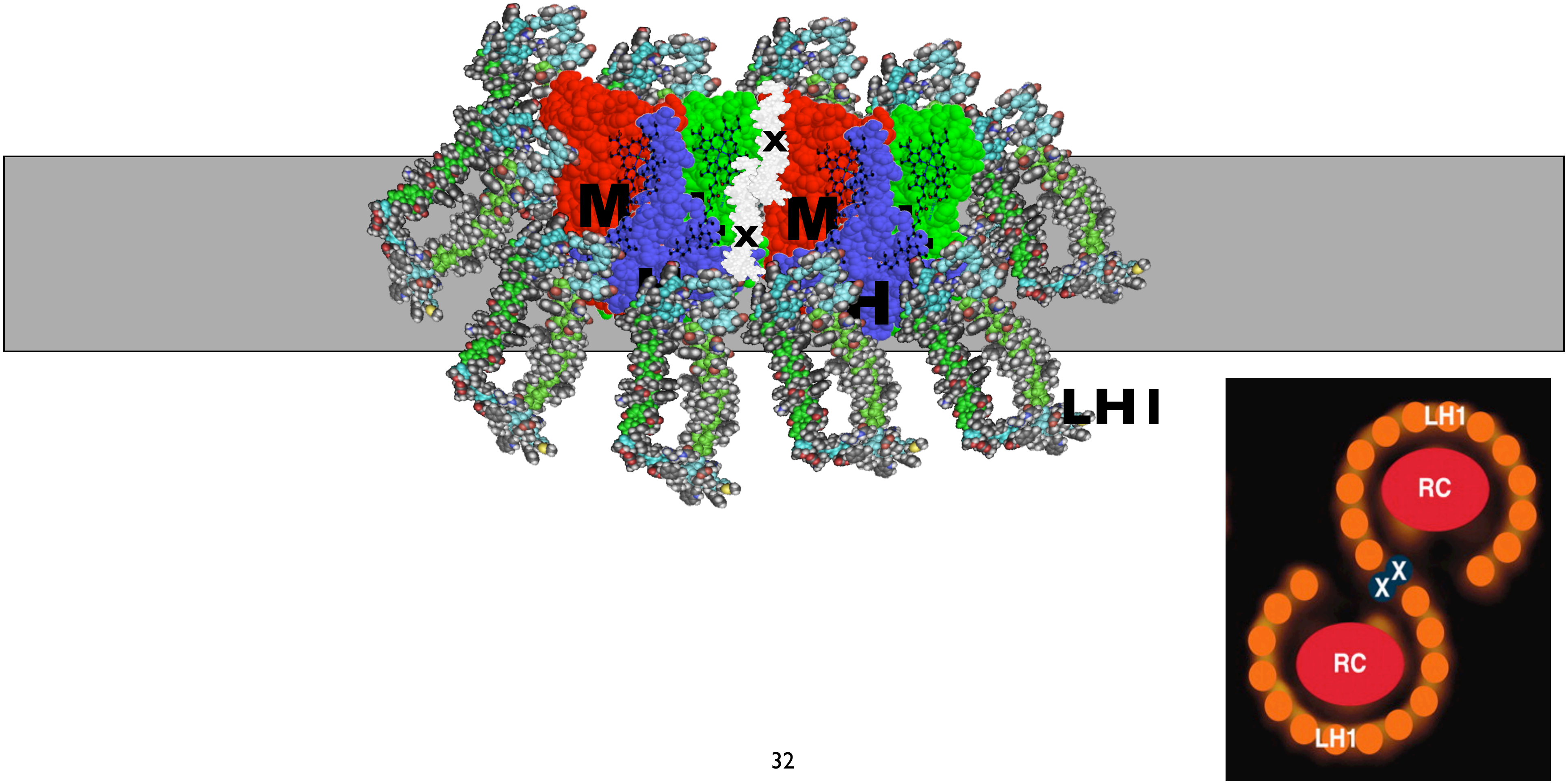
Rhodobacter sphaeroides ATCC 17029 at US DOE Joint Genome Institute [Complete]
DOE Joint Genome Institute

Rhodobacter sphaeroides KD131 at GenoTech corp. [In progress]
GenoTech corp.

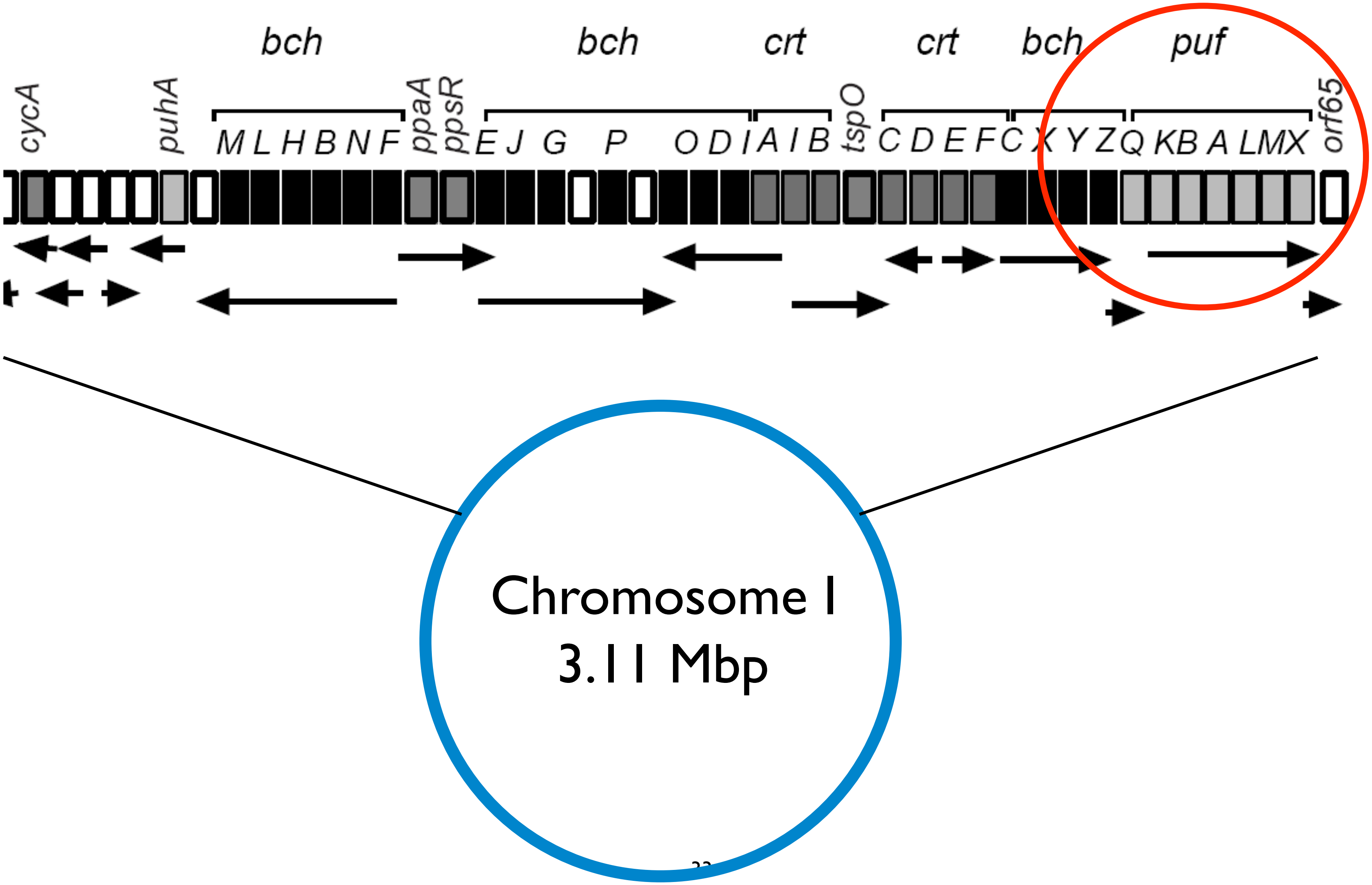
The photosynthetic co-complex



Co-complex Structure



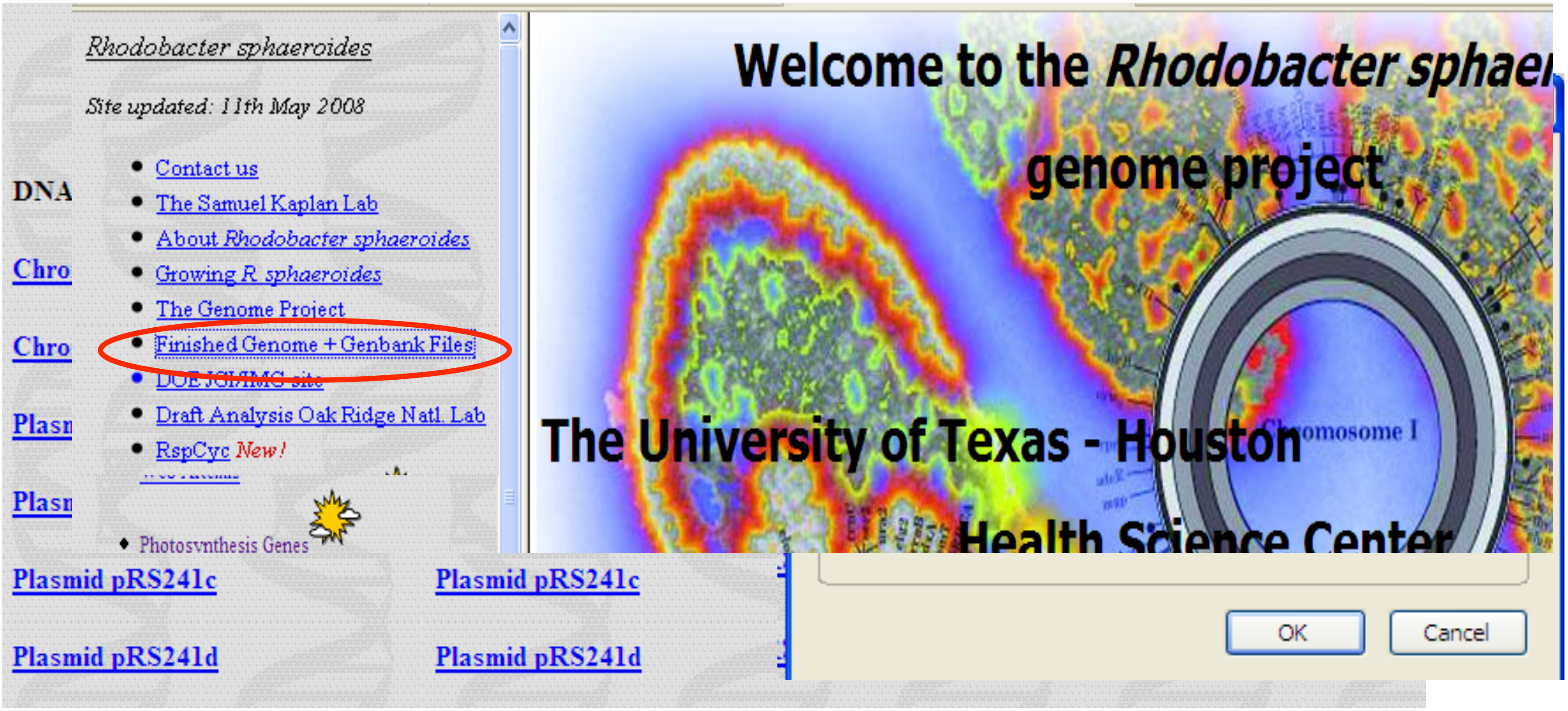
The photosynthetic gene cluster



Chromosome I
3.11 Mbp

Download and Extract Sequence File for Chromosome I

1. Go to www.rhodobacter.org
2. Click 'Finished Genome + GenBank files'



Rhodobacter sphaeroides

Site updated: 11th May 2008

Welcome to the *Rhodobacter sphae*

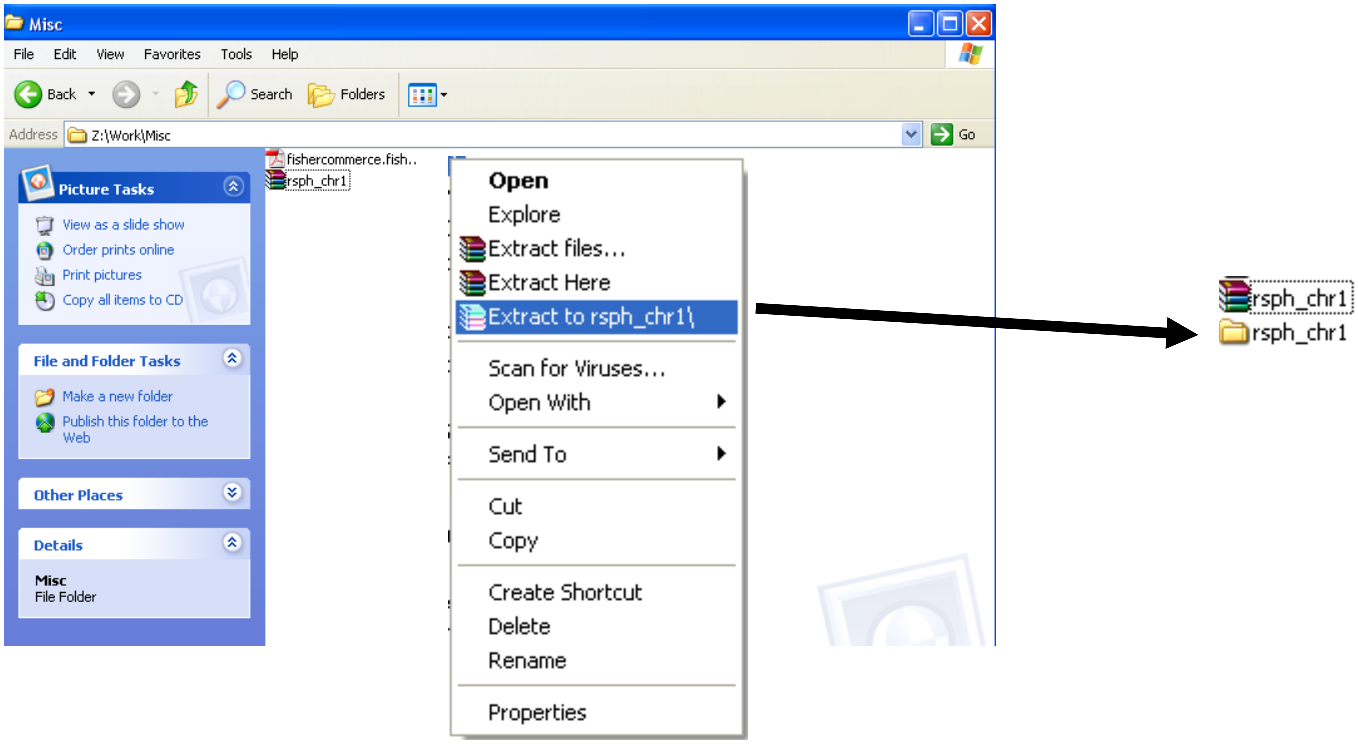
genome project

The University of Texas - Houston

Health Science Center

Download and Extract Sequence File for Chromosome I

3. Find ‘rsph_chr1.zip’ file.
4. Right click on file and then choose **‘Extract to rsph_chr1\’**



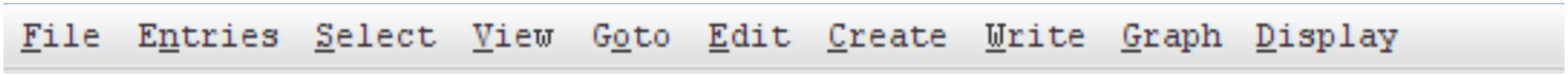
Start Artemis and Load Chromosome I File

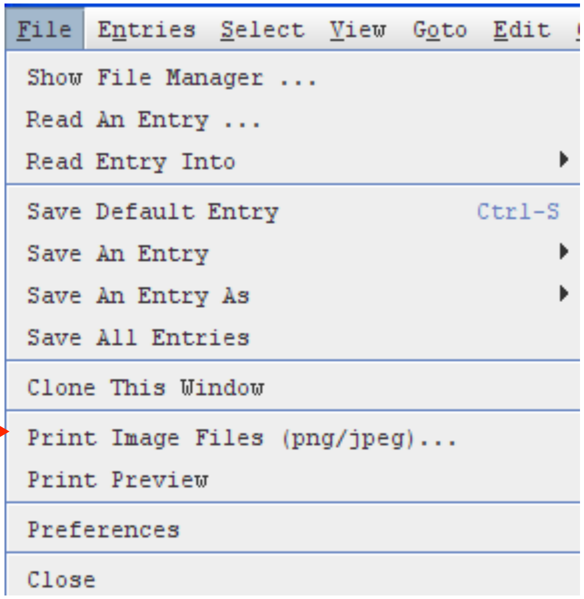
Artemis
Release 9
1. Standard

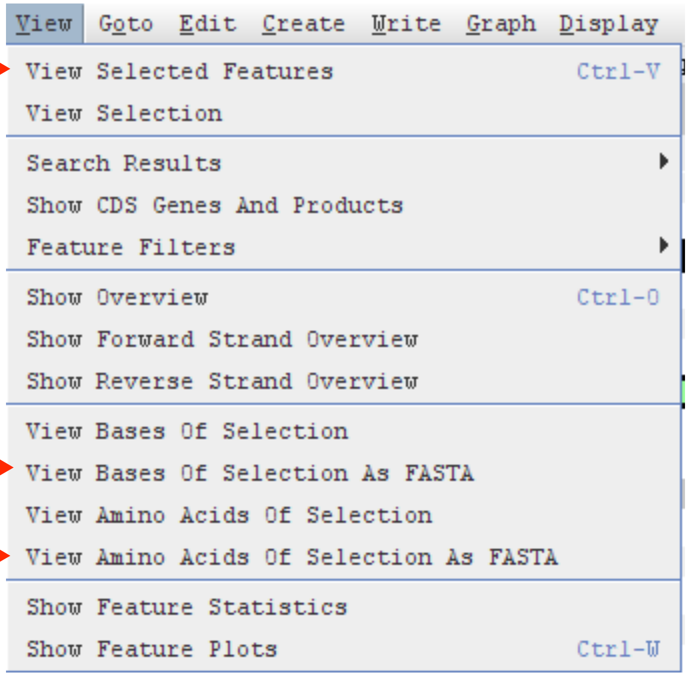
Copyright 1998 - 2007
Genome Research Limited

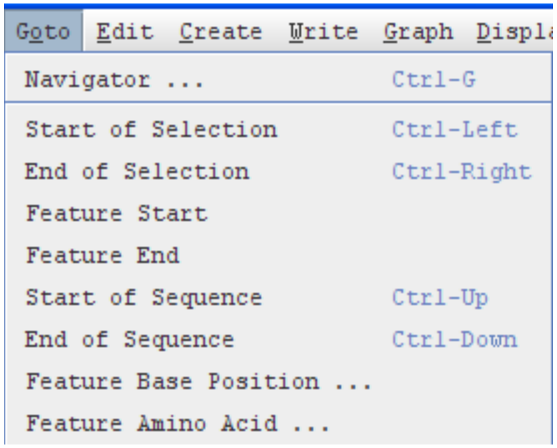
The Sanger Institute

Menu Bar

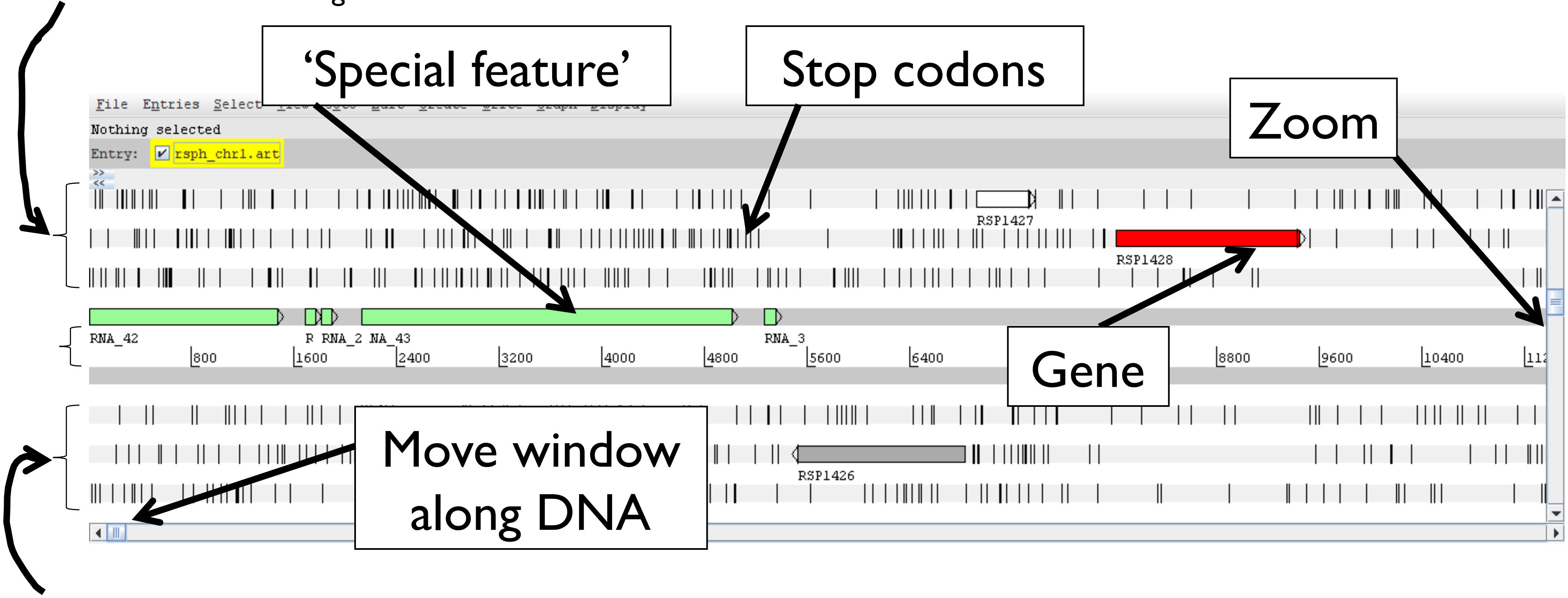

File
Entries
Select
View
Goto
Edit
Create
Write
Graph
Display


File
Entries
Select
View
Goto
Edit
Show File Manager ...
Read An Entry ...
Read Entry Into
Save Default Entry Ctrl-S
Save An Entry
Save An Entry As
Save All Entries
Clone This Window
Print Image Files (png/jpeg)...
Print Preview
Preferences
Close


View
Goto
Edit
Create
Write
Graph
Display
View Selected Features Ctrl-V
View Selection
Search Results
Show CDS Genes And Products
Feature Filters
Show Overview Ctrl-O
Show Forward Strand Overview
Show Reverse Strand Overview
View Bases Of Selection
View Bases Of Selection As FASTA
View Amino Acids Of Selection
View Amino Acids Of Selection As FASTA
Show Feature Statistics
Show Feature Plots Ctrl-W


Goto
Edit
Create
Write
Graph
Navigator ... Ctrl-G
Start of Selection Ctrl-Left
End of Selection Ctrl-Right
Feature Start
Feature End
Start of Sequence Ctrl-Up
End of Sequence Ctrl-Down
Feature Base Position ...
Feature Amino Acid ...

Top Pane: Overview



3 Forward strand reading frames

‘Special feature’

Stop codons

Zoom

Gene

Move window along DNA

3 Reverse strand reading frames

Middle Pane: Sequence Detail

Forward Strand
3 Frame Translation

```
<<
 P  S  G  F  R  G  G  A  P  A  N  G  S  A  W  V  G  A
  L  R  G  S  G  A  E  P  Q  R  T  A  A  R  G  +  G
A  F  G  V  Q  G  R  S  P  S  E  R  Q  R  V  G  R  G
```

Forward Strand (5’→3’)

```
CCTTCGGGGTTCAGGGGCGGAGCCCCAGCGAACGGCAGCGCGTGGGTAGGGGC
```

Genome Coordinates

```
      |7800                |7820                |7840
```

Reverse Strand (3’←5’)

```
GGAAGCCCCAAGTCCCCGCCTCGGGGTCGCTTGCCGTCGCGCACCCATCCCCG
```

Reverse Strand
3 Frame Translation

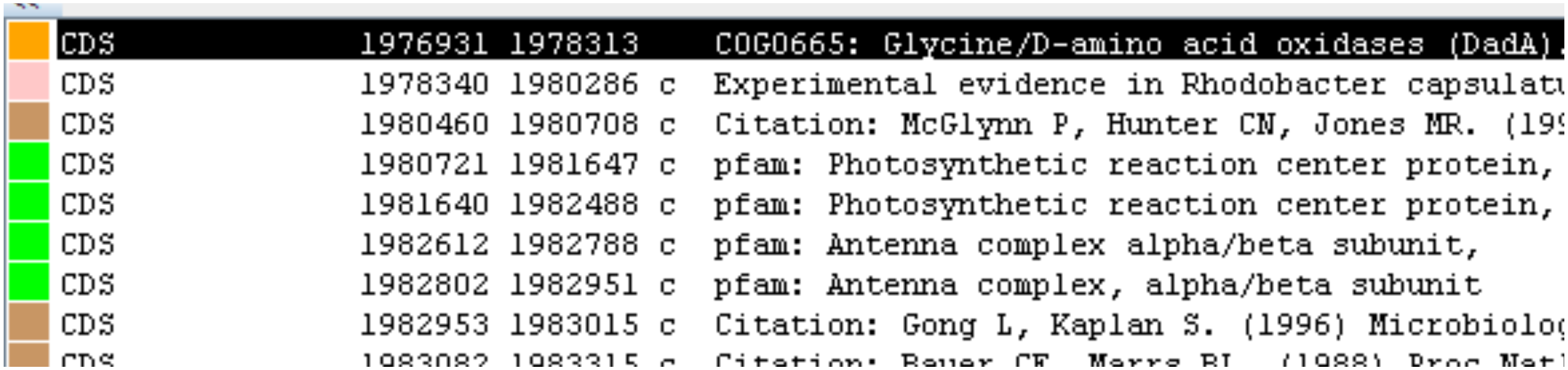
```
 R  R  P  E  P  A  S  G  W  R  V  A  A  R  P  Y  P  G
  K  P  T  *  P  R  L  G  L  S  R  C  R  T  P  L  F
G  E  P  N  L  P  P  A  G  A  F  P  L  A  H  T  P  A
```

Scroll & Zoom

Four Tasks

- Find the *pufM* and *pufL* genes of the reaction center (part of the *puf* operon)
- How many nucleotides overlap between the coding sequence of these genes?
- Are the genes in the same frame?
- Export a picture of the *puf* operon (*pufQBALMX*)

Bottom Pane: Feature List



```
CDS        1976931 1978313    COG0665: Glycine/D-amino acid oxidases (DadA)
CDS        1978340 1980286 c  Experimental evidence in Rhodobacter capsulatu
CDS        1980460 1980708 c  Citation: McGlynn P, Hunter CN, Jones MR. (19
CDS        1980721 1981647 c  pfam: Photosynthetic reaction center protein,
CDS        1981640 1982488 c  pfam: Photosynthetic reaction center protein,
CDS        1982612 1982788 c  pfam: Antenna complex alpha/beta subunit,
CDS        1982802 1982951 c  pfam: Antenna complex, alpha/beta subunit
CDS        1982953 1983015 c  Citation: Gong L, Kaplan S. (1996) Microbiolo
CDS        1983082 1983315 c  Citation: Bauer CE, Marrs BL. (1988) Proc Nat
```

Links

- UCSC Genome Browser

  http://genome.cse.ucsc.edu/

- Ensembl Genome Browser

  http://www.ensembl.org/index.html

- NCBI MapViewer

  http://www.ncbi.nlm.nih.gov/mapview/index.html

- Artemis

  http://www.sanger.ac.uk/Software/Artemis/