joanne@msl.ubc.ca

Bioinformatics

Common Tools & Tricks of the Trade

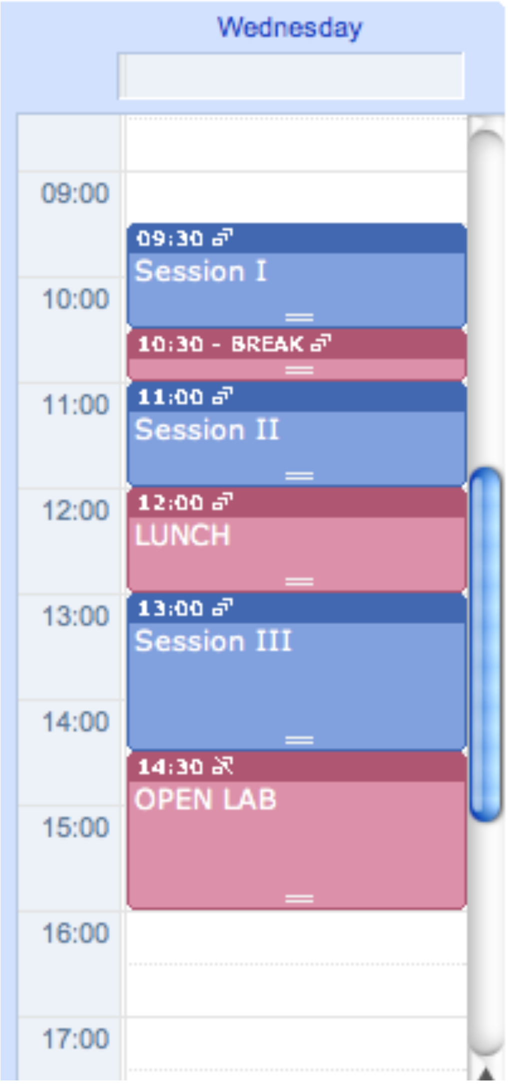
Welcome to Day 3
bioteach.ubc.ca/bioinfo2009

Workshop Schedule

- Laptops, available here for your use 9am - 4:30pm
- wireless login

  mslguest

  4myguest




Thank You
Thank You
Enjoy

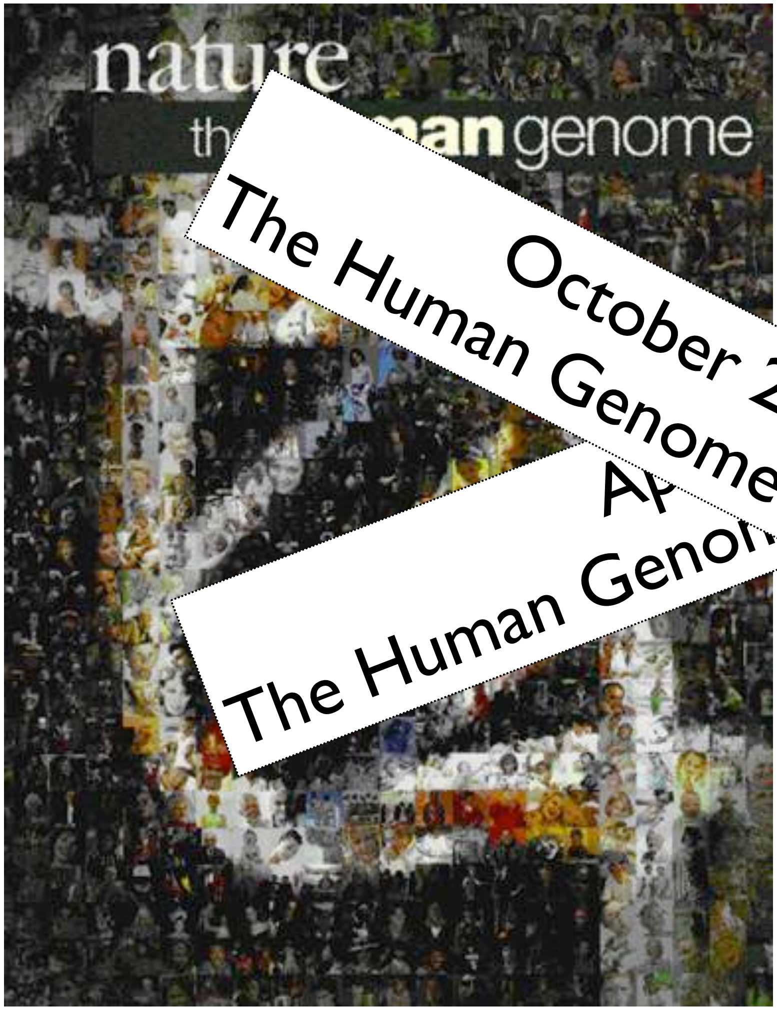
Today's Topics

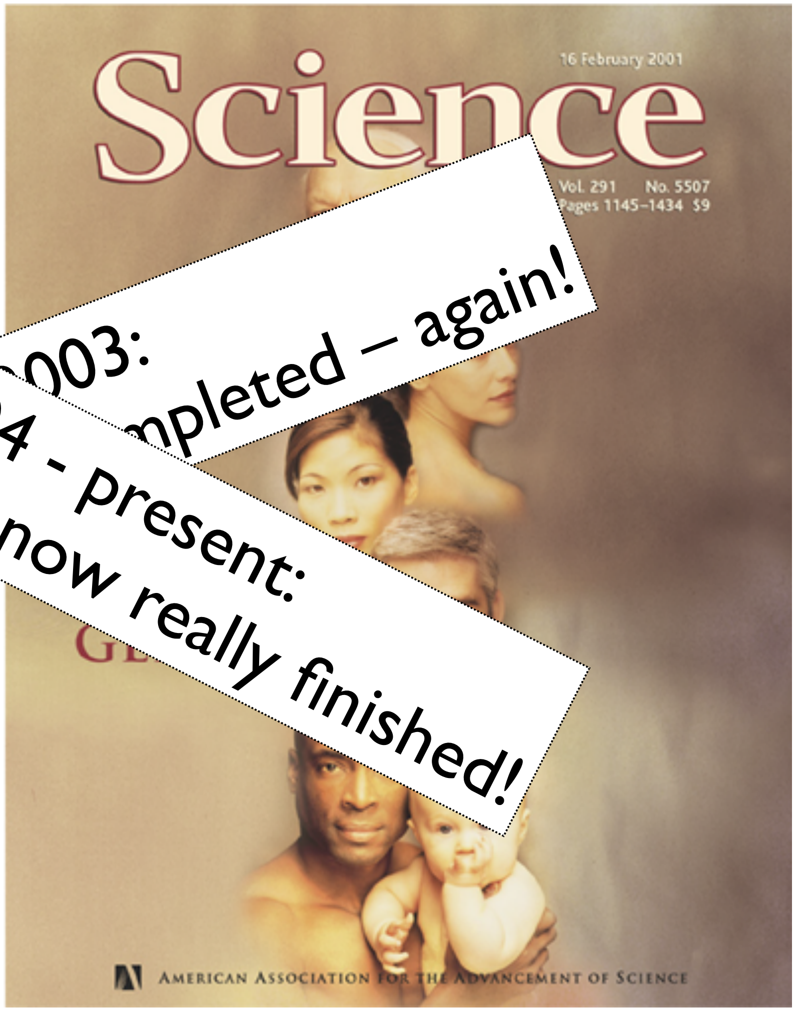
- **Genome Browsers**, Accessing Genome Annotations.
- **PRACTICAL EXERCISES**, three different views of the BRCA1 gene
- **Discovering GEO**, the Gene Expression Omnibus.
- **Pathway Resources** for Systems Biology
- **Bioinformatics Links Directory**, Conducting Research on the Web

Genome Browsers

Accessing Genome Annotations & PRACTICAL EXERCISE: Three Different Views of the BRCA1 Gene

The Human Genome Project





Ap 003:
The Human Genon mpleted – again!

October 2004 - present:
The Human Genome is now really finished!

Public HGP

Celera Genomics

February 2001: Completion of the Draft Human Genome

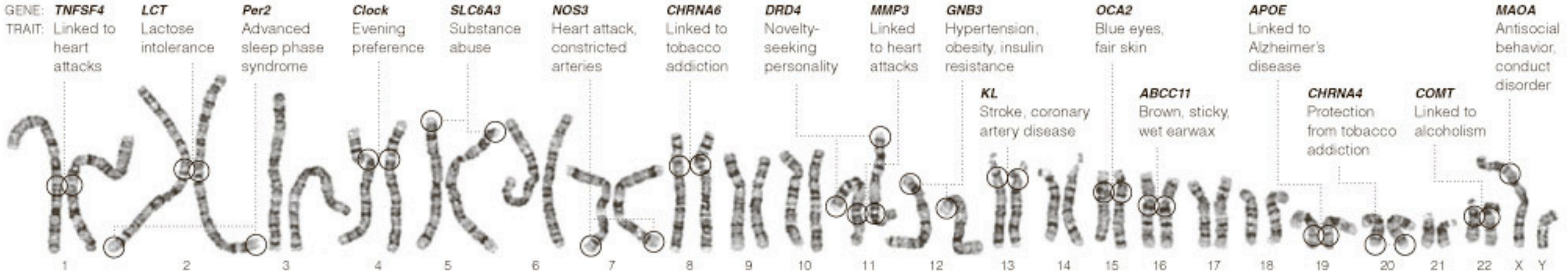
In the Genome Race, the Sequel Is Personal

Thor Swift for The New York Times

A team led by J. Craig Venter, above, has finished the first mapping of a full, or diploid, genome, made up of DNA inherited from both parents. The genome is Dr. Venter's own.

The New York Times September 3, 2007

DECODING HIMSELF A team led by J. Craig Venter, above, has finished the first mapping of a full, or diploid, genome, made up of DNA inherited from both parents. The genome is Dr. Venter's own.



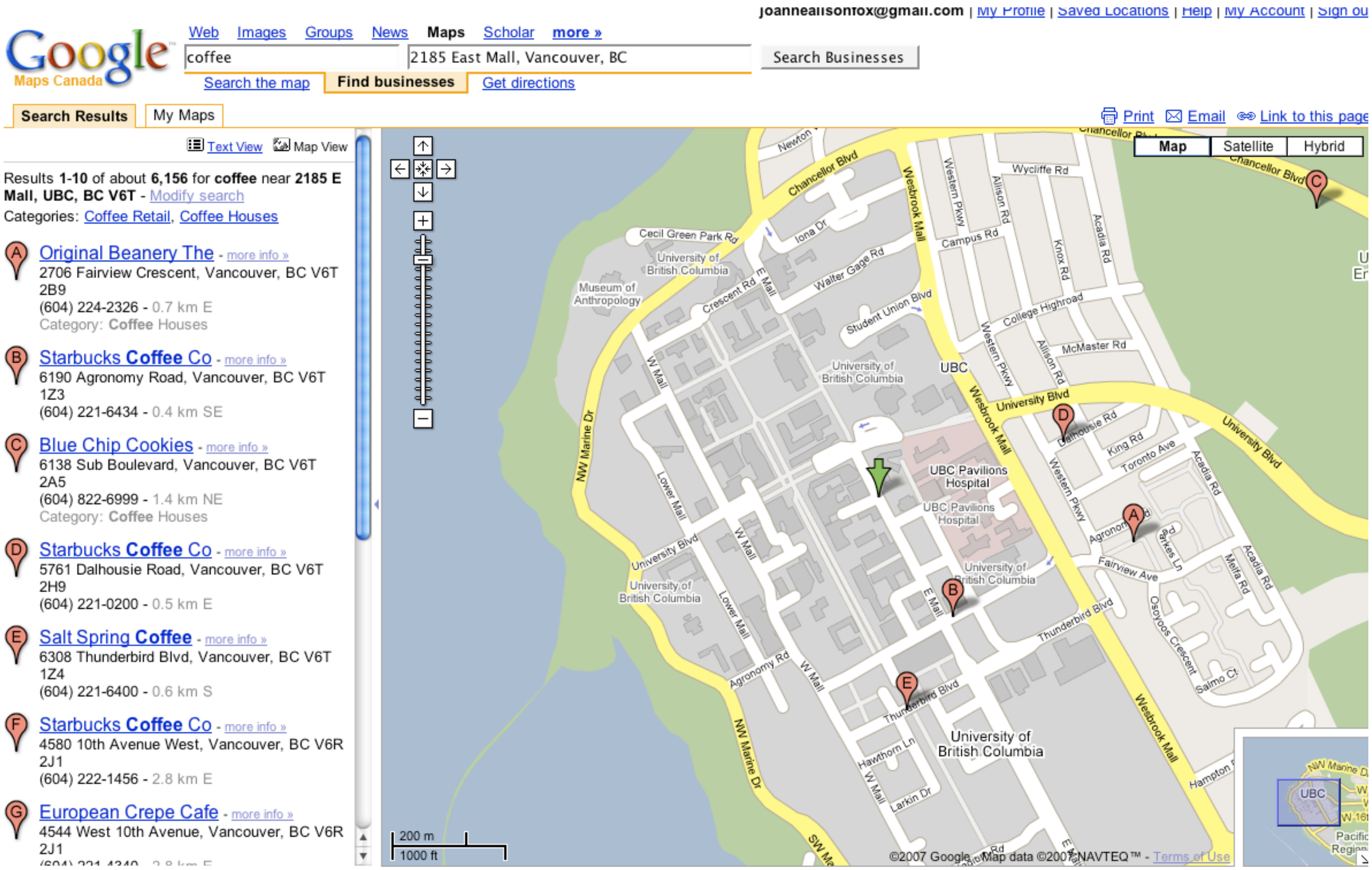
Sources: PLoS Biology; GeneCards

Genes highlighted above are linked to specific traits but do not mean the conditions are inevitable.

THE NEW YORK TIMES

Stanley Park
Stanley Park Shuttle
Stanley Park Shuttle (seasonal)
STANLEY PARK LOOP
19
Lost Lagoon
Deadmans Island
Coal Harbour
0
500
METRES
N
To North Vancouver and Lonsdale Quay
SEABUS
SEABUS TERMINAL
WATERFRONT
BURRARD
GRANVILLE
STADIUM - CHINATOWN
MAIN STREET - SCIENCE WORLD
West End
English Bay
Gastown
Chinatown
Yaletown
False Creek
Fairview
Mount Pleasant
Devonian Harbour Park
Coal Harbour Park
Harbour Green Park
Vancouver Convention & Exhibition Centre
CN IMAX Theatre
Vancouver Tourist InfoCentre
Sinclair Centre
Harbour Centre/SFU
Portside Park
West End Community Centre
Robson Market
Barclay Heritage Square
Royal Centre
Vancouver Art Gallery
Robson Square
UBC Downtown
Pacific Centre
Nelson Park
St. Paul's Hospital
Orpheum Theatre
Canada Post
Vancouver Library
Queen Elizabeth Theatre
VCC
Dr Sun Yat Sen Chinese Gardens
Andy Livingstone Park
G.M. Place
B.C. Place Stadium
Plaza of Nations
Creekside Park
Science World
Pacific Central Station (Bus & Railway)
Thornton Park
Aquabus
False Creek Ferry
Morton Park
Alexandra Park
English Bay Beach Park
Sunset Beach Park
Vanier Park
H.R. MacMillan Space Centre & Vancouver Museum
Gordon Southam Observatory
Vancouver Academy of Music
Granville Island Market
Arts Club Theatre
George Wainborn Park
David Lam Park
Coopers' Park
Emily Carr Institute of Art + Design
Waterfront Theatre
Sutcliffe Park
Charleson Park
LAGOON DR
ALBERNI
CHILCO
GILFORD
PENDRELL
DENMAN
BIDWELL
BEACH AVE
CARDERO
HARO
BARCLAY
BROUGHTON
COMOX
NICOLA
BURNABY
JERVIS
BUTE
THURLOW
DAVIE
HARWOOD
BEACH
PACIFIC
HORNBY
HOWE
PENDER
HASTINGS
GEORGIA
ROBSON
BURRARD
DUNSMUIR
GRANVILLE
SEYMOUR
RICHARDS
HOMER
HAMILTON
SMITHE
CAMBIE
NELSON
DRAKE
EXPO BLVD
PACIFIC BLVD
BEATTY
WATER
CORDOVA
CARRALL
ABBOTT
KEEFER
COLUMBIA
POWELL
MAIN
QUEBEC
GORE
PRIOR ST
NATIONAL
TERMINAL
BURRARD BRIDGE
GRANVILLE BRIDGE
CAMBIE BRIDGE
CHESTNUT
1 AVE
2 AVE
3 AVE
4 AVE
5 AVE
6 AVE
7 AVE
8 AVE
FIR
PINE
HEMLOCK
BIRCH
ALDER
SPRUCE
OAK
LAUREL
WILLOW
HEATHER
ASH
YUKON
ALBERTA
MANITOBA
ONTARIO
SCOTIA
BRUNSWICK
GT NORTHERN
LAMEY'S MILL
BROADWAY
240.241.242.246.247.250.251
252.253.254.257.258.N24
5 N6
6.C23
C21
2.22.32.44
2.C21
C21.N22
98
19.22.44
N6.N8.N9 N10.N15 N17.N19 N20.N35
10.44
5.6.8.20
4.7.10.16.17
4.6.7.10.16.17
N9.N10.N17
C21.C23
C23
4.7.8
8.10.16.20.N20
4.7.N8
19.N19
3.8.19.22.N8.N19
15.N15
15.N19
2.22.32.44.N22
4.7.10.16.17.50.98
N9.N10.N17
4.7.84
9.16.17.99.N17
9.17.99.N9
9.50.99.N9
9.99.N9
50.84
84
50
17
R 09.07

maps.google.ca



coffee

2185 East Mall, Vancouver, BC

Results **1-10** of about **6,156** for **coffee** near **2185 E Mall, UBC, BC V6T** - Modify search

Categories: Coffee Retail, Coffee Houses

(A) Original Beanery The - more info »
2706 Fairview Crescent, Vancouver, BC V6T 2B9
(604) 224-2326 - 0.7 km E
Category: **Coffee** Houses

(B) Starbucks **Coffee** Co - more info »
6190 Agronomy Road, Vancouver, BC V6T 1Z3
(604) 221-6434 - 0.4 km SE

(C) Blue Chip Cookies - more info »
6138 Sub Boulevard, Vancouver, BC V6T 2A5
(604) 822-6999 - 1.4 km NE
Category: **Coffee** Houses

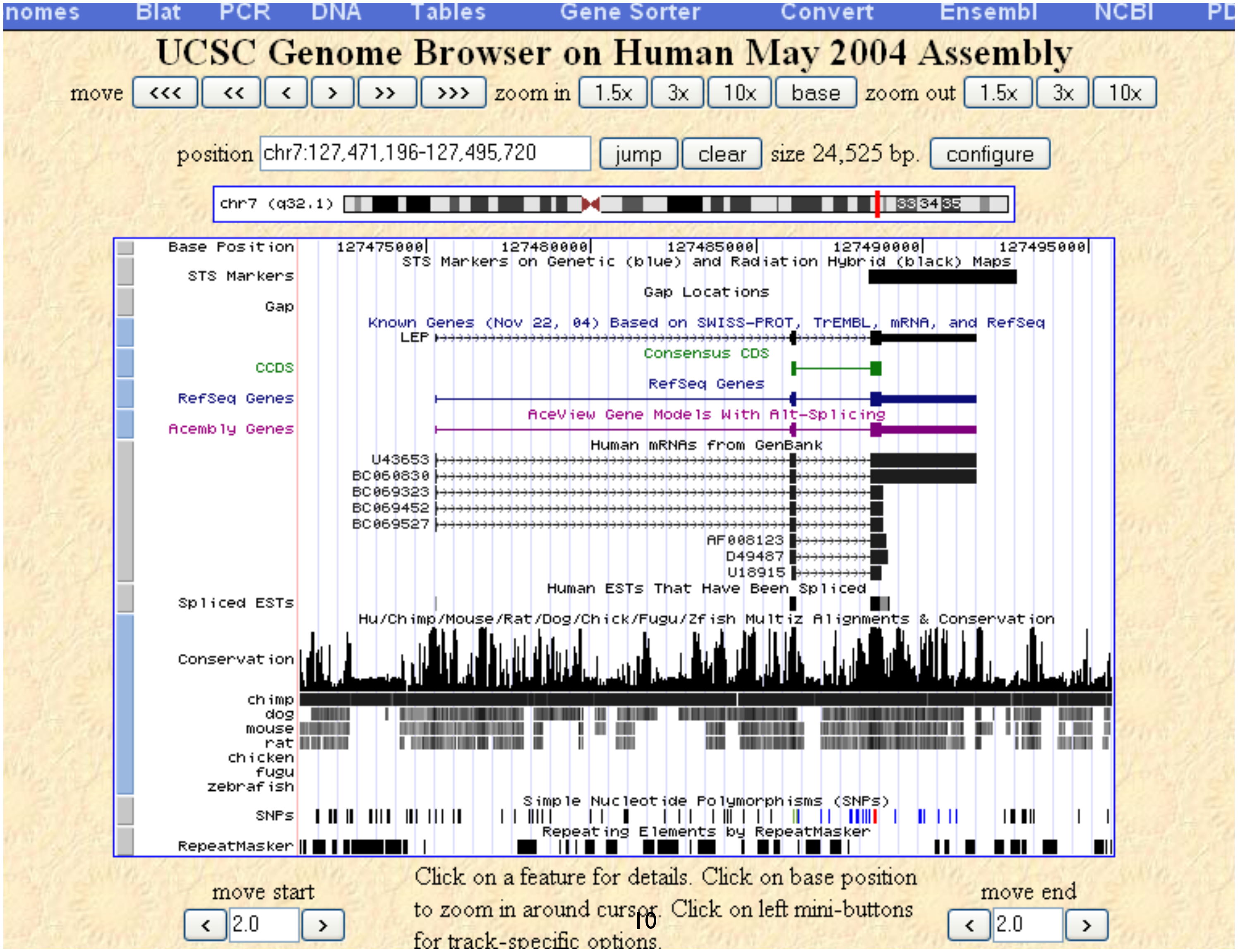
(D) Starbucks **Coffee** Co - more info »
5761 Dalhousie Road, Vancouver, BC V6T 2H9
(604) 221-0200 - 0.5 km E

(E) Salt Spring **Coffee** - more info »
6308 Thunderbird Blvd, Vancouver, BC V6T 1Z4
(604) 221-6400 - 0.6 km S

(F) Starbucks **Coffee** Co - more info »
4580 10th Avenue West, Vancouver, BC V6R 2J1
(604) 222-1456 - 2.8 km E

(G) European Crepe Cafe - more info »
4544 West 10th Avenue, Vancouver, BC V6R 2J1
(604) 221-4340 - 2.8 km E

Let’s Look at the Human Genome…



UCSC Genome Browser on Human May 2004 Assembly

position chr7:127,471,196-127,495,720

size 24,525 bp.

Click on a feature for details. Click on base position to zoom in around cursor. Click on left mini-buttons for track-specific options.

Objectives

- By the end of this module:

✓ You will be able to describe the following concepts: genome annotation, genome builds, and genome browsers.

✓ You will view the genomic location that contains the BRCA1 gene in the human genome using three different genome browsers.

✓ You will be able to compare and contrast the UCSC, Ensembl and MapViewer systems for visualizing genome information.

Genome Browsers

- What is a Genome Browser?
  - System for displaying, viewing, and accessing genome annotation data
- Genome annotations = knowledge attached to raw genome sequence.
  - Annotation information comes from many different sources
    - ✓ Computational pipelines
    - ✓ Research groups
    - ✓ Databases

The "Neopolitan Ice Cream" World of Genome Browsing:

- UCSC Genome Browser

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

- Ensembl

  http://www.ensembl.org/

- NCBI Map Viewer

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

The underlying data is common for all three “flavors” of Genome Browsers.

- NCBI, UCSC and Ensembl use the same human genome assembly that is generated by NCBI
  - release timing is different between sites.
- Note the version of genome assembly to which you are referring
  - available precomputed info and locations of features will be different between different assemblies.

Let's compare the view of the BRCA1 gene in all three genome browsers.

Viewing the genomic region containing BRCA1

- Common features:

✓ Coordinate system is based on the build

✓ Zoom in and out

✓ Annotations displayed – ie. Gene features

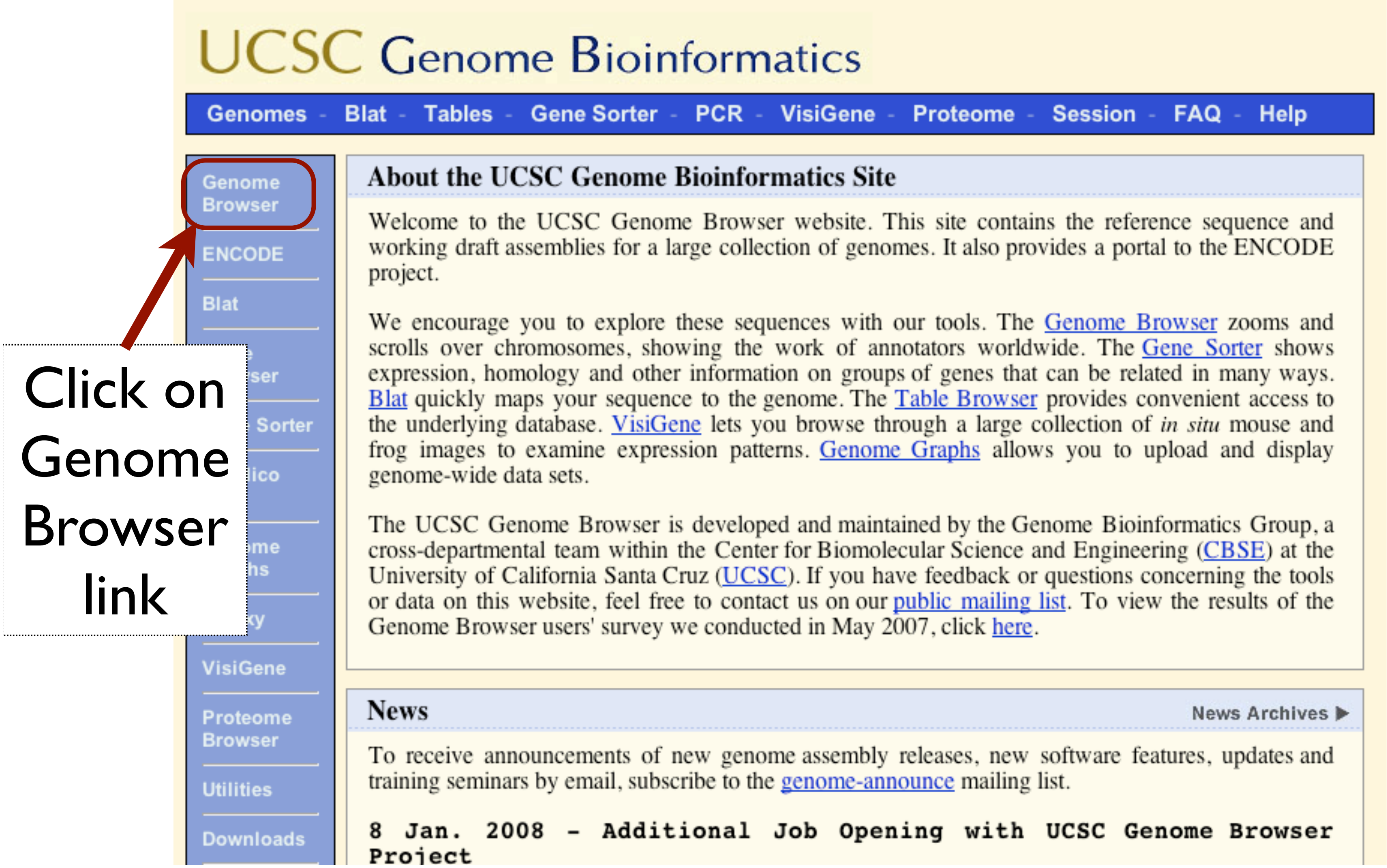
- Major Differences:

✓ Each Browser has a very different look and feel

✓ Annotation information displayed differently

✓ Different ways to navigate through the information

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



Click on Genome Browser link

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.

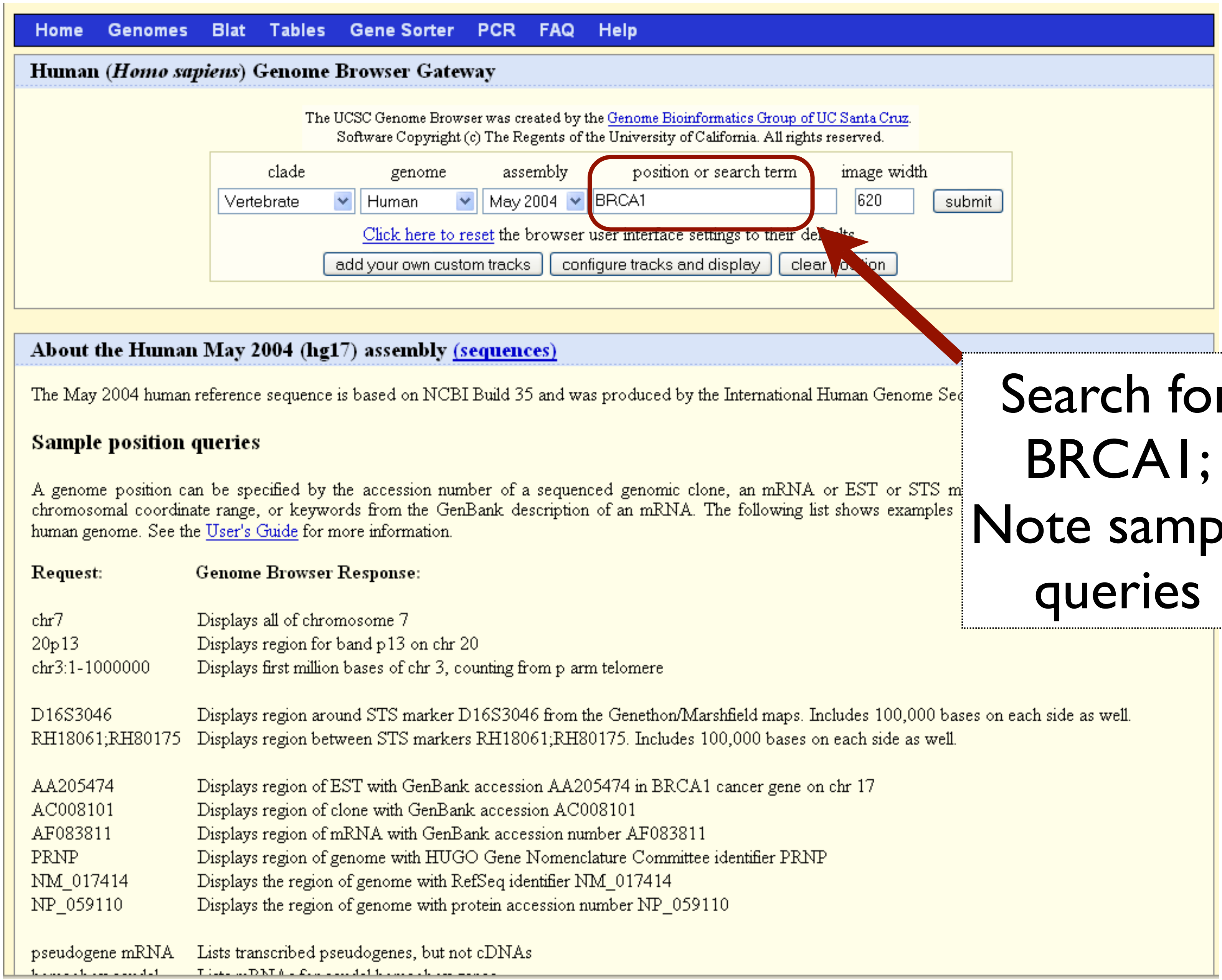
We encourage you to explore these sequences with our tools. The Genome Browser zooms and scrolls over chromosomes, showing the work of annotators worldwide. The Gene Sorter shows expression, homology and other information on groups of genes that can be related in many ways. Blat quickly maps your sequence 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



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)

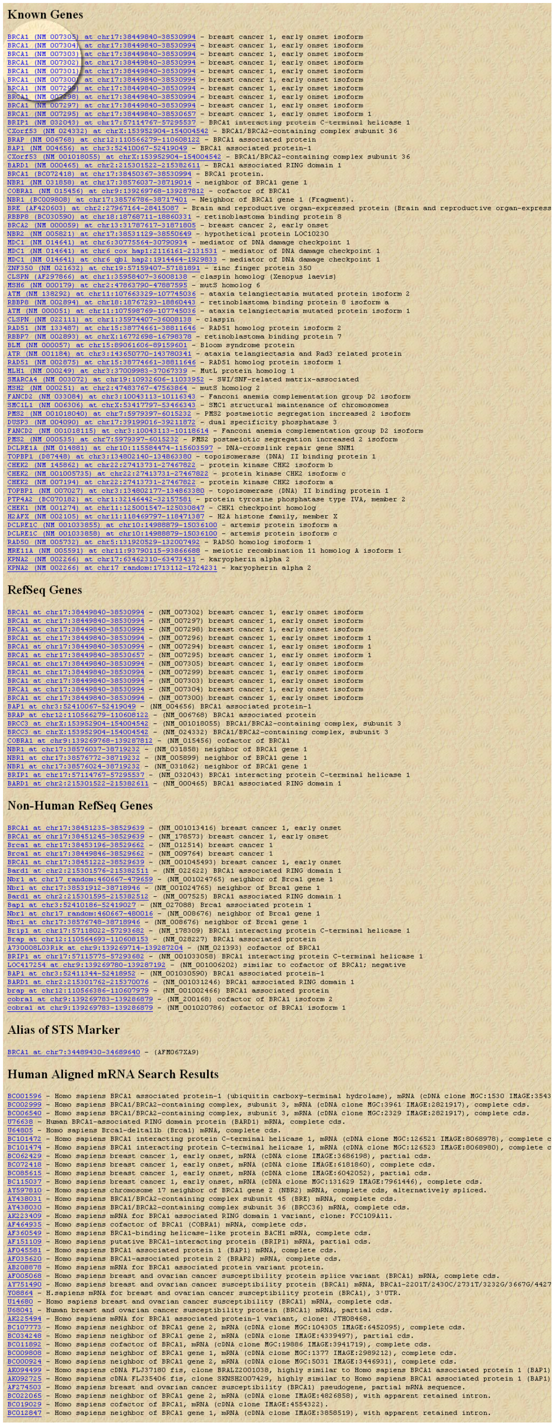
The May 2004 human reference sequence is based on NCBI Build 35 and was produced by the International Human Genome Se

Sample position queries

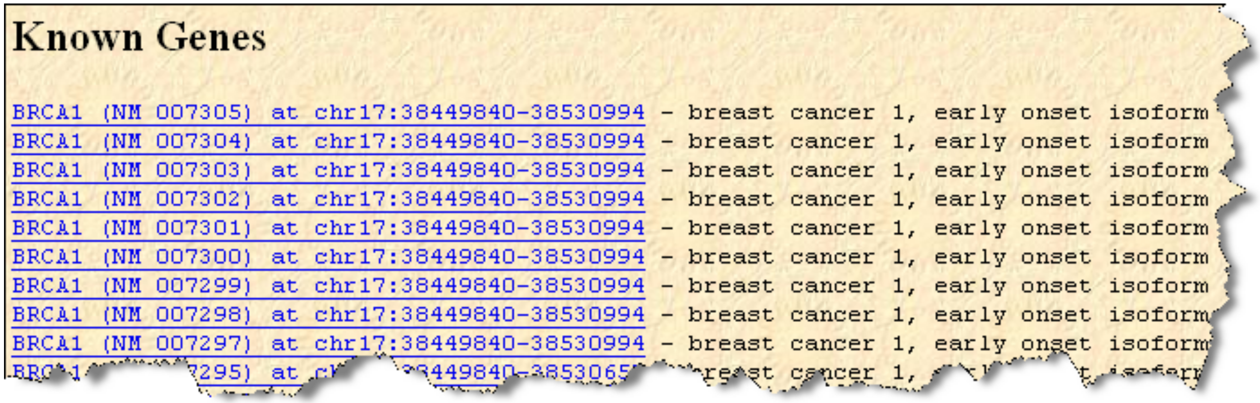
A genome position can be specified by the accession number of a sequenced genomic clone, an mRNA or EST or STS m chromosomal coordinate range, or keywords from the GenBank description of an mRNA. The following list shows examples human genome. See the User's Guide for more information.

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



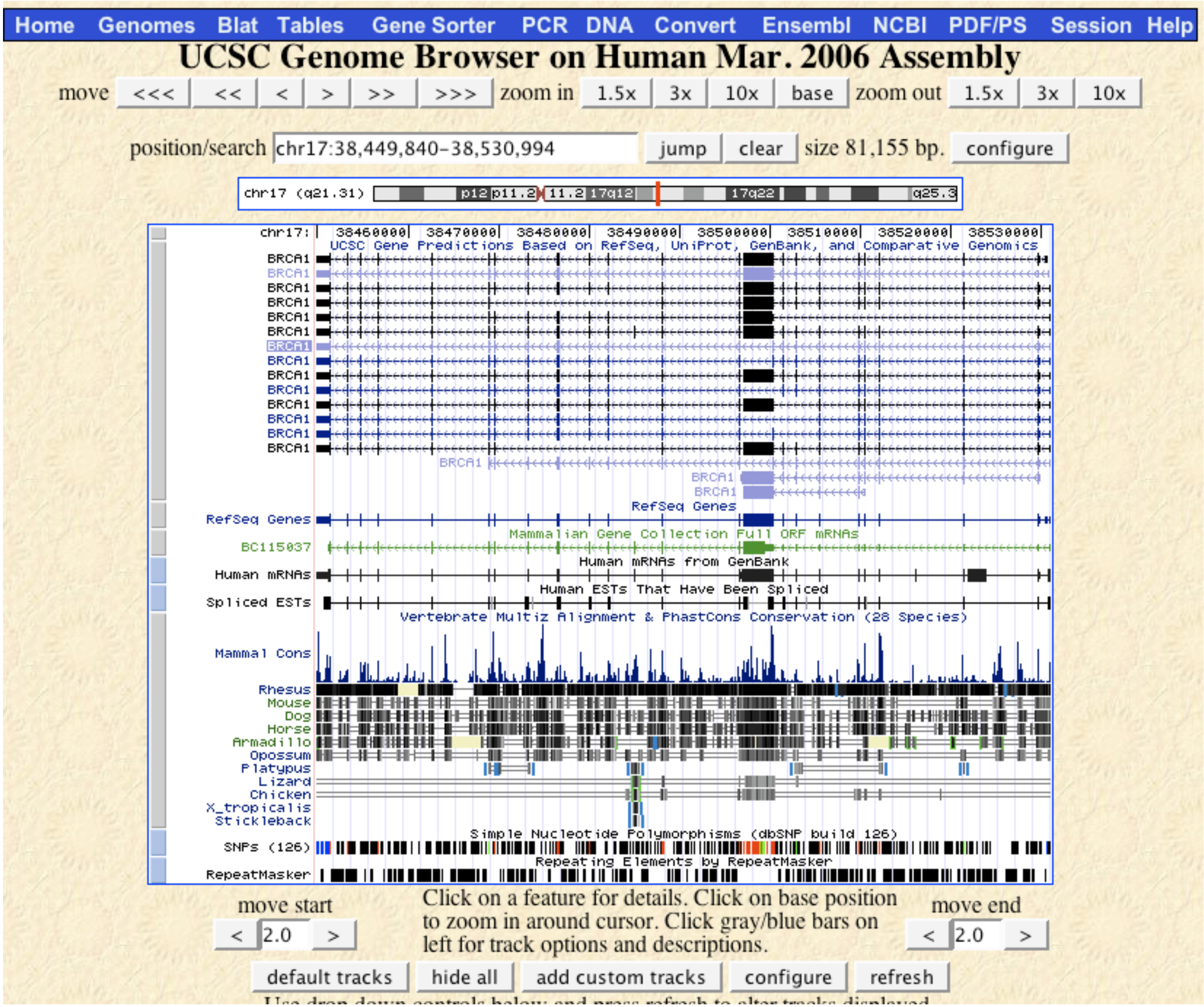
The Search Results



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
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.

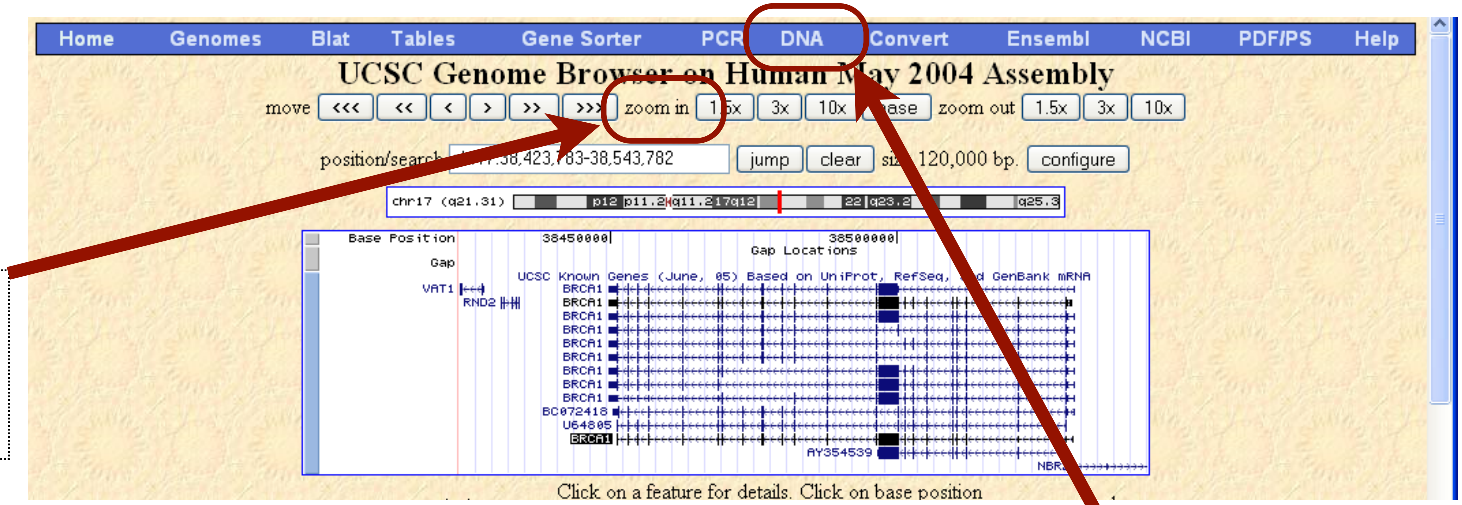
Click on a feature for details. Click on base position to zoom in around cursor. Click gray/blue bars on left for track options and descriptions.

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



Click on a feature for details. Click on base position

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')

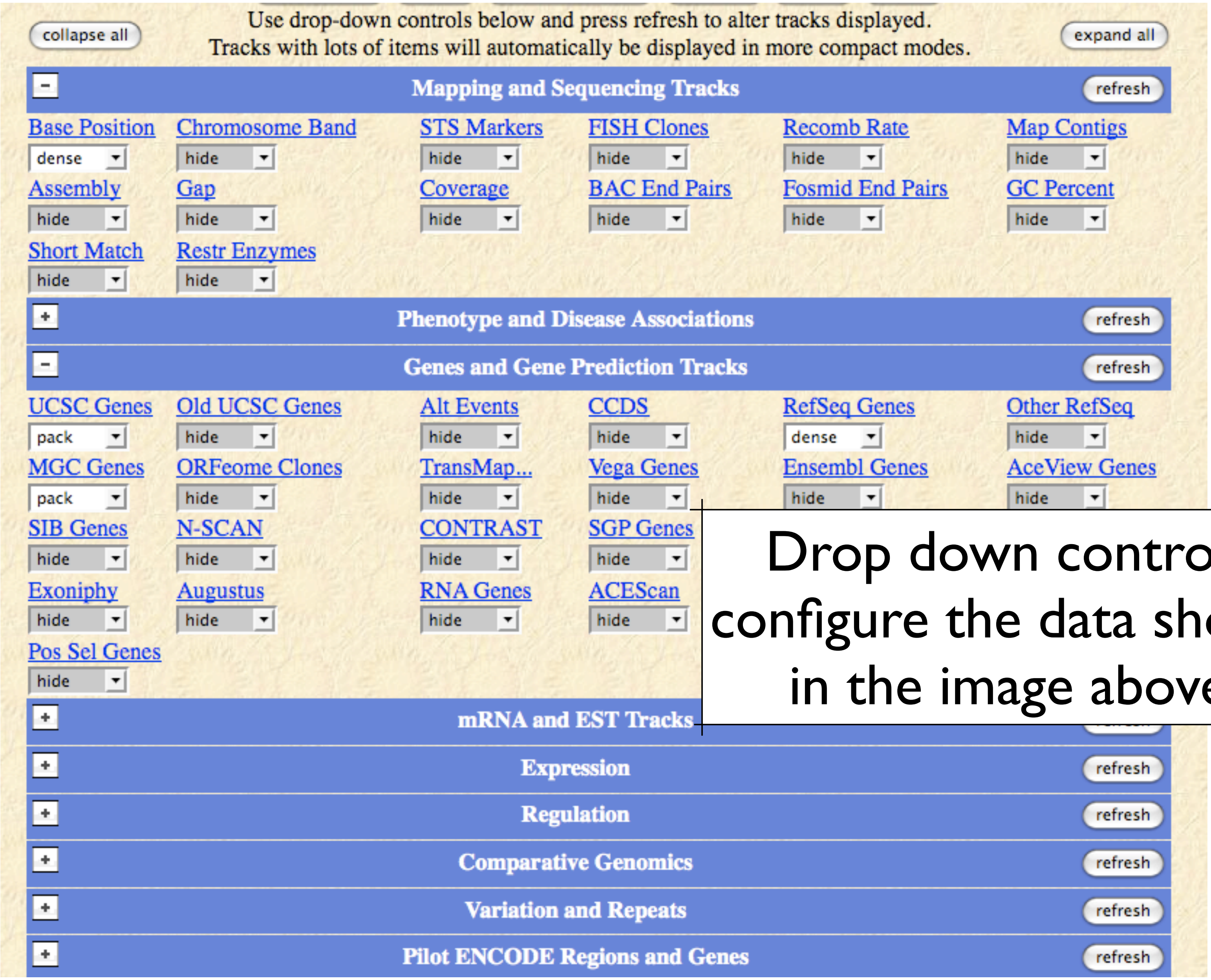
Note: if a feature is close to the beginning or end of a chromosome and upstream/downstream bases are added, they may be truncated in order to avoid extending past the edge of the chromosome.

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".

DNA link
Download
Sequence



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

Phenotype and Disease Associations

Genes and Gene Prediction Tracks

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

mRNA and EST Tracks

Expression

Regulation

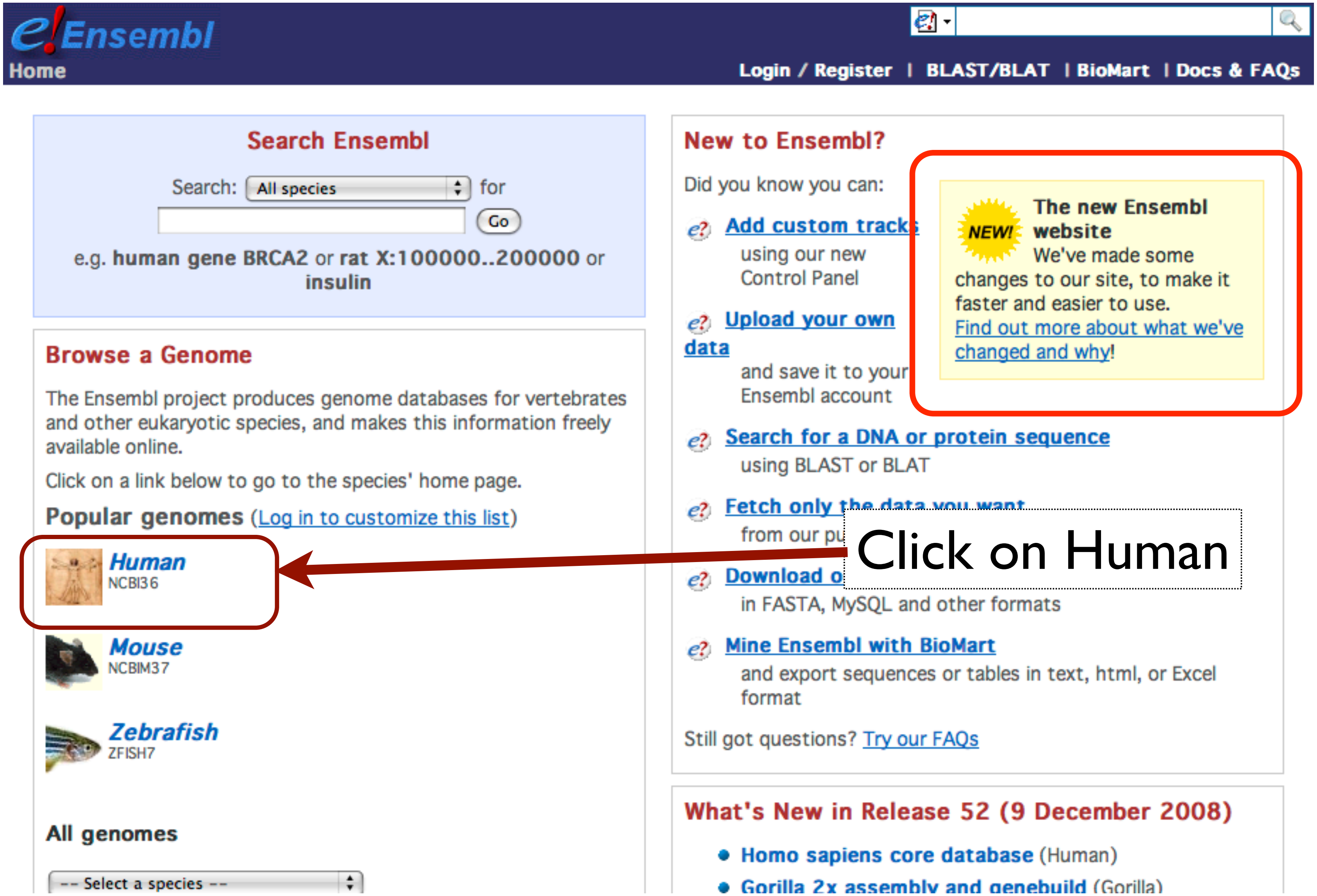
Comparative Genomics

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

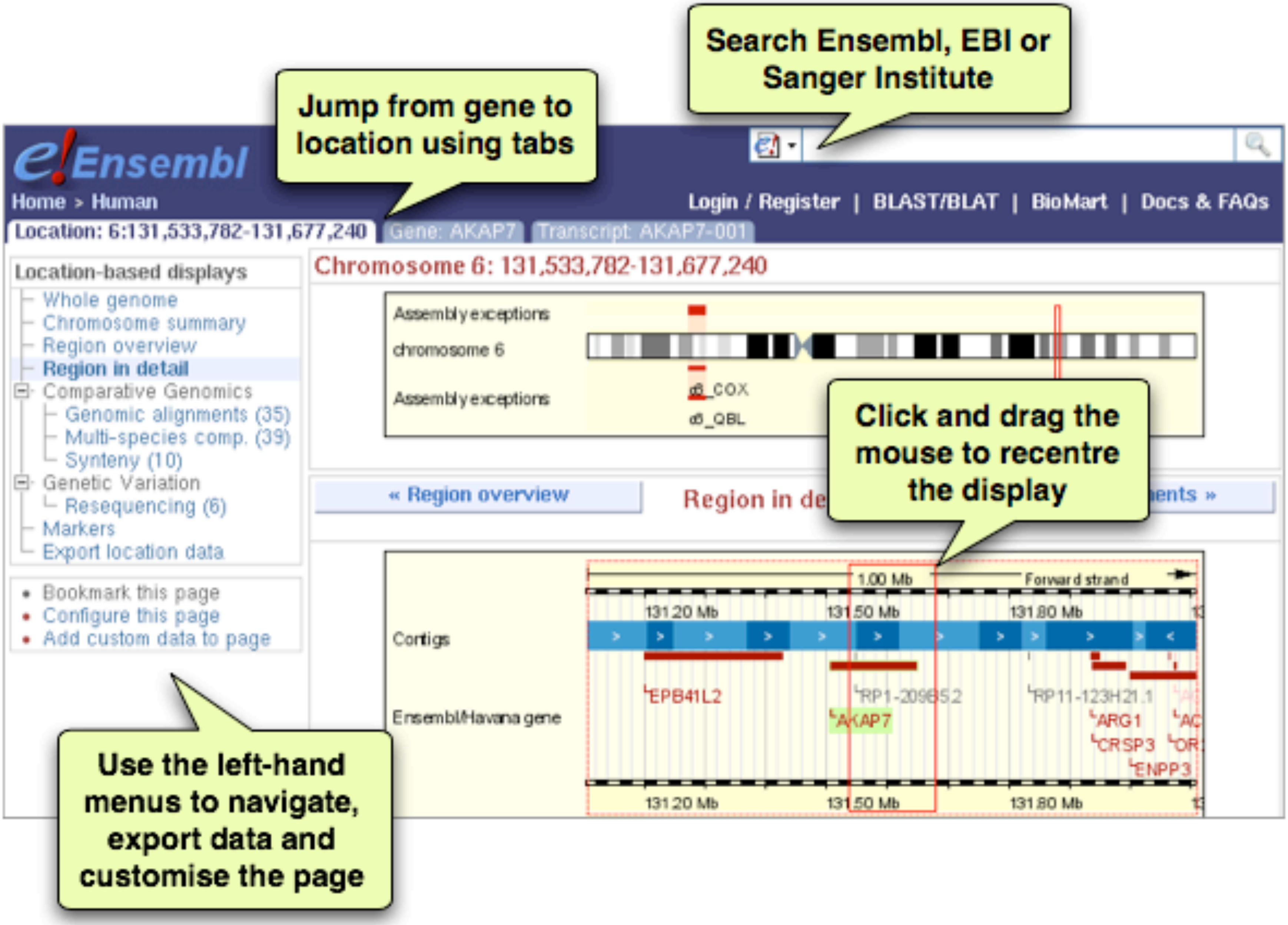
Still got questions? Try our FAQs

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

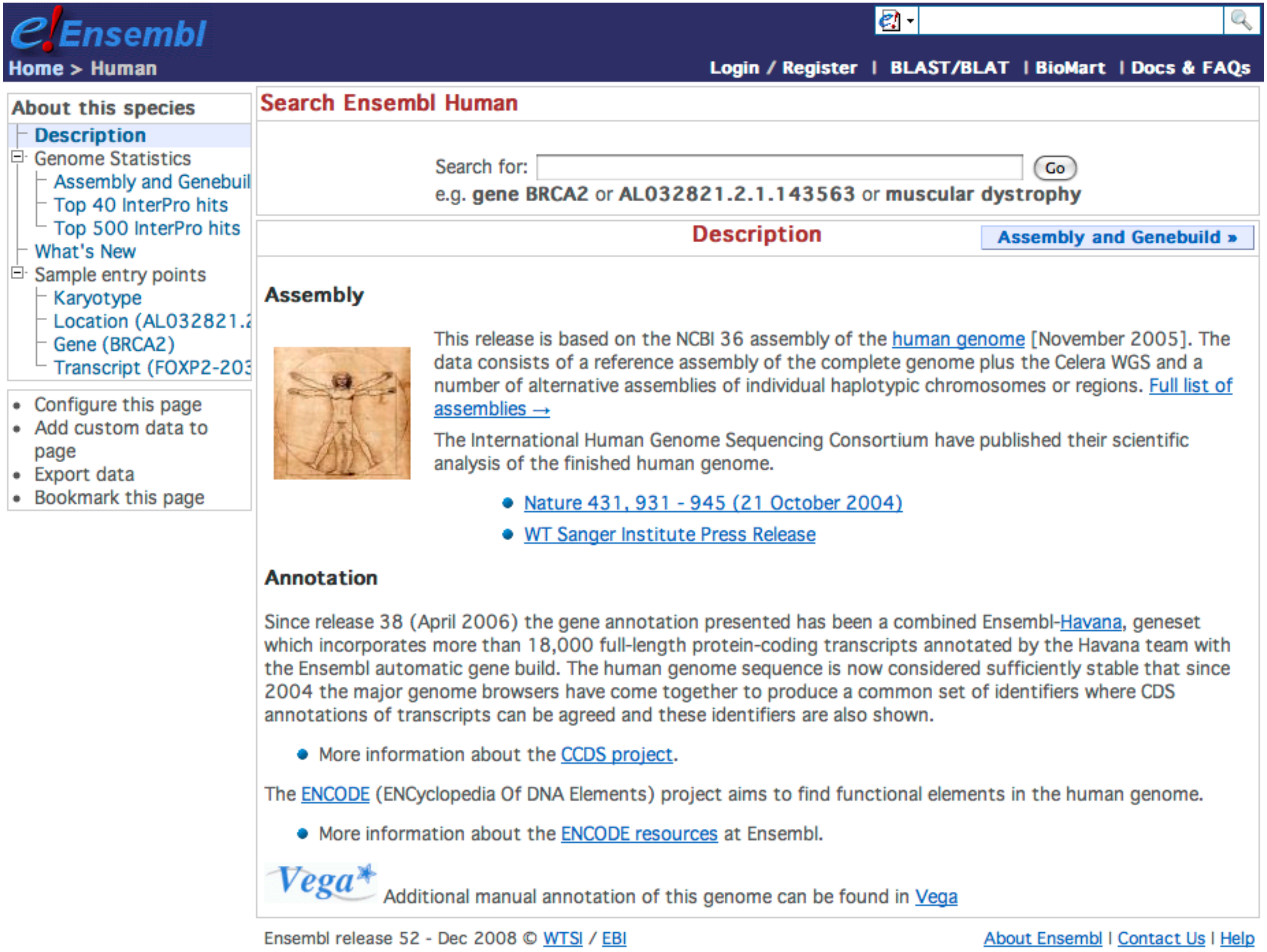


Search Ensembl, EBI or Sanger Institute

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.

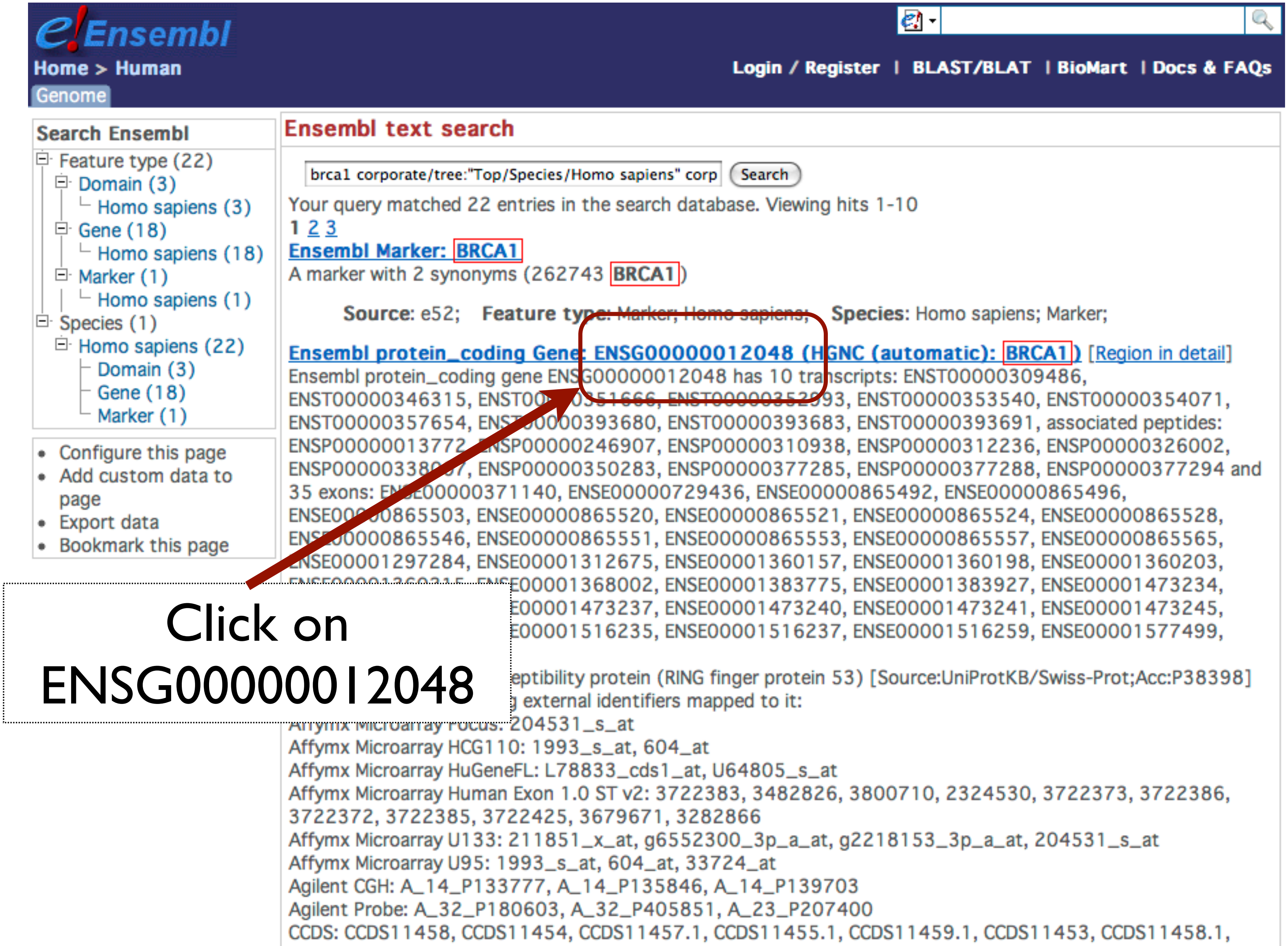
The ENCODE (ENCyclopedia Of DNA Elements) project aims to find functional elements in the human genome.

- 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)

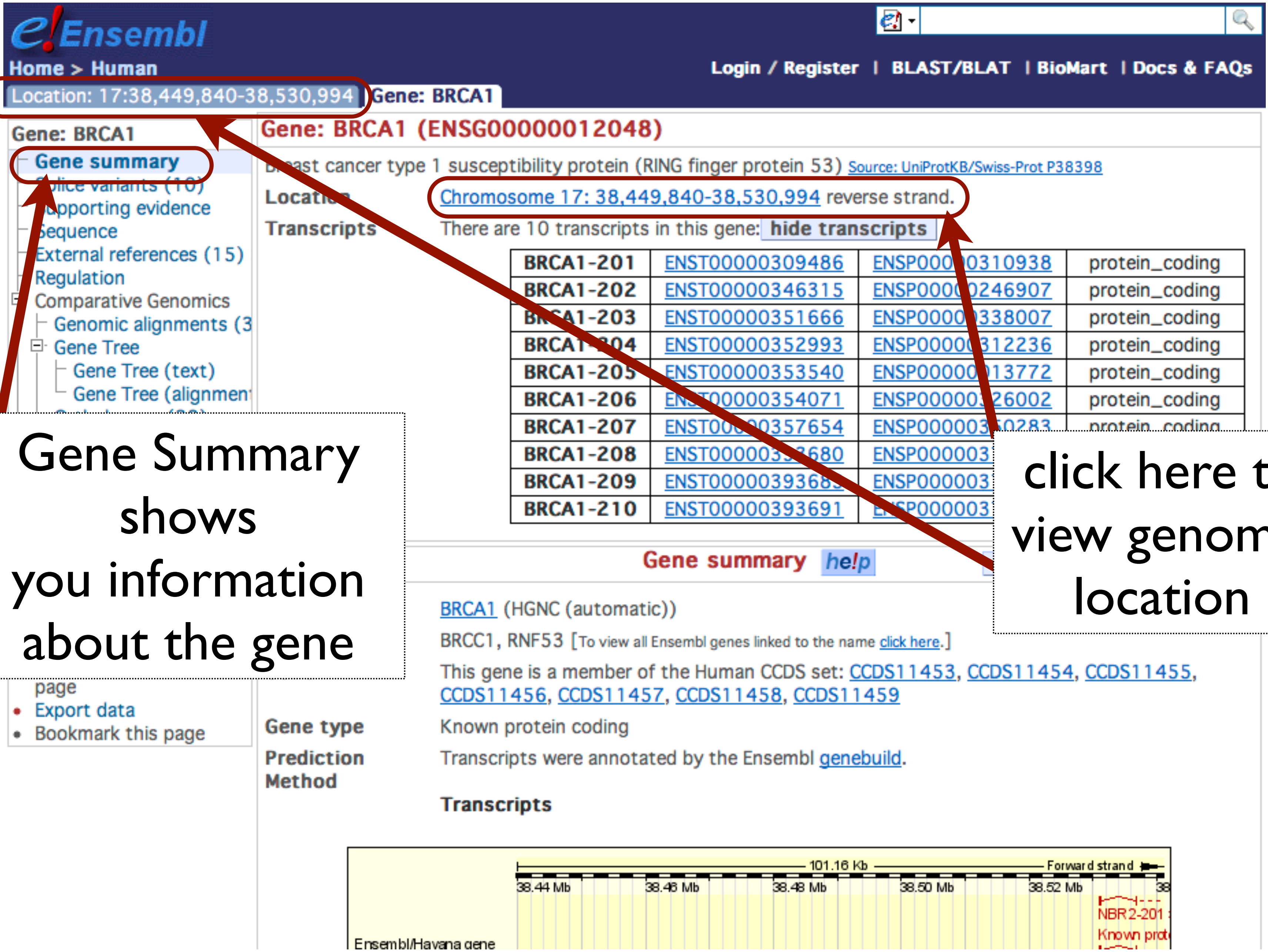
Source: e52; **Feature type:** Marker; Homo sapiens; **Species:** Homo sapiens; Marker;

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

Ensembl protein_coding gene ENSG00000012048 has 10 transcripts: ENST00000309486, ENST00000346315, ENST00000351666, ENST00000352993, ENST00000353540, ENST00000354071, ENST00000357654, ENST00000393680, ENST00000393683, ENST00000393691, associated peptides: ENSP00000013772, ENSP00000246907, ENSP00000310938, ENSP00000312236, ENSP00000326002, ENSP00000338007, ENSP00000350283, ENSP00000377285, ENSP00000377288, ENSP00000377294 and 35 exons: ENSE00000371140, ENSE00000729436, ENSE00000865492, ENSE00000865496, ENSE00000865503, ENSE00000865520, ENSE00000865521, ENSE00000865524, ENSE00000865528, ENSE00000865546, ENSE00000865551, ENSE00000865553, ENSE00000865557, ENSE00000865565, ENSE00001297284, ENSE00001312675, ENSE00001360157, ENSE00001360198, ENSE00001360203, ENSE00001368002, 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
- Splice variants (10)
- Supporting evidence
- Sequence
- 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)

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: **hide transcripts**

BRCA1-201	ENST00000309486	ENSP000[illegible]0310938	protein_coding
BRCA1-202	ENST00000346315	ENSP0000[illegible]0246907	protein_coding
BRCA1-203	ENST00000351666	ENSP0000[illegible]338007	protein_coding
BRCA1-204	ENST00000352993	ENSP0000[illegible]312236	protein_coding
BRCA1-205	ENST00000353540	ENSP00000[illegible]13772	protein_coding
BRCA1-206	ENST00000354071	ENSP00000[illegible]26002	protein_coding
BRCA1-207	ENST00[illegible]00357654	ENSP000003[illegible]0283	protein_coding
BRCA1-208	ENST000003[illegible]680	ENSP000003	
BRCA1-209	ENST0000039368[illegible]	ENSP000003	
BRCA1-210	ENST00000393691	ENSP000003	

Gene summary help

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.

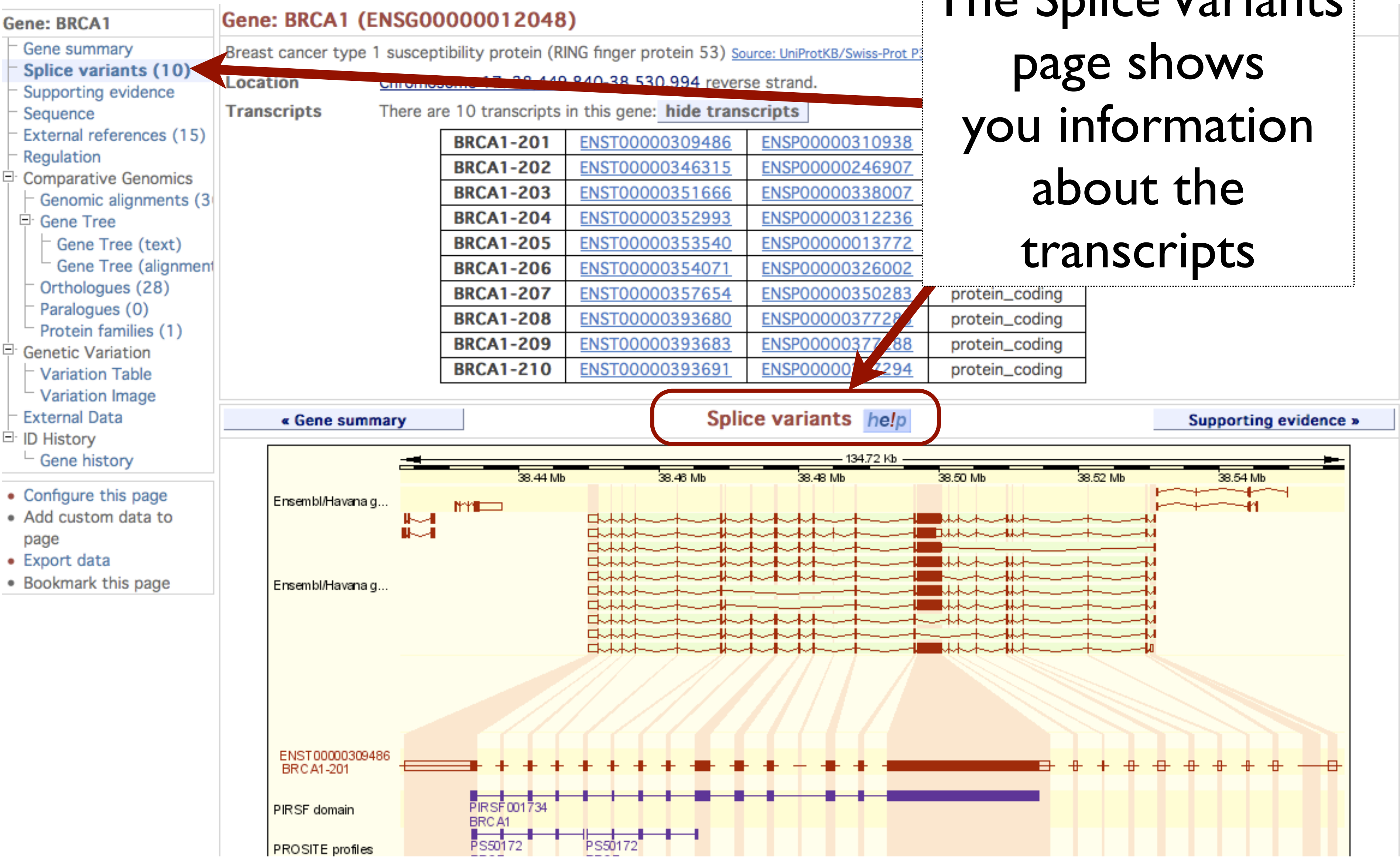
Transcripts

Gene Summary shows you information about 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)

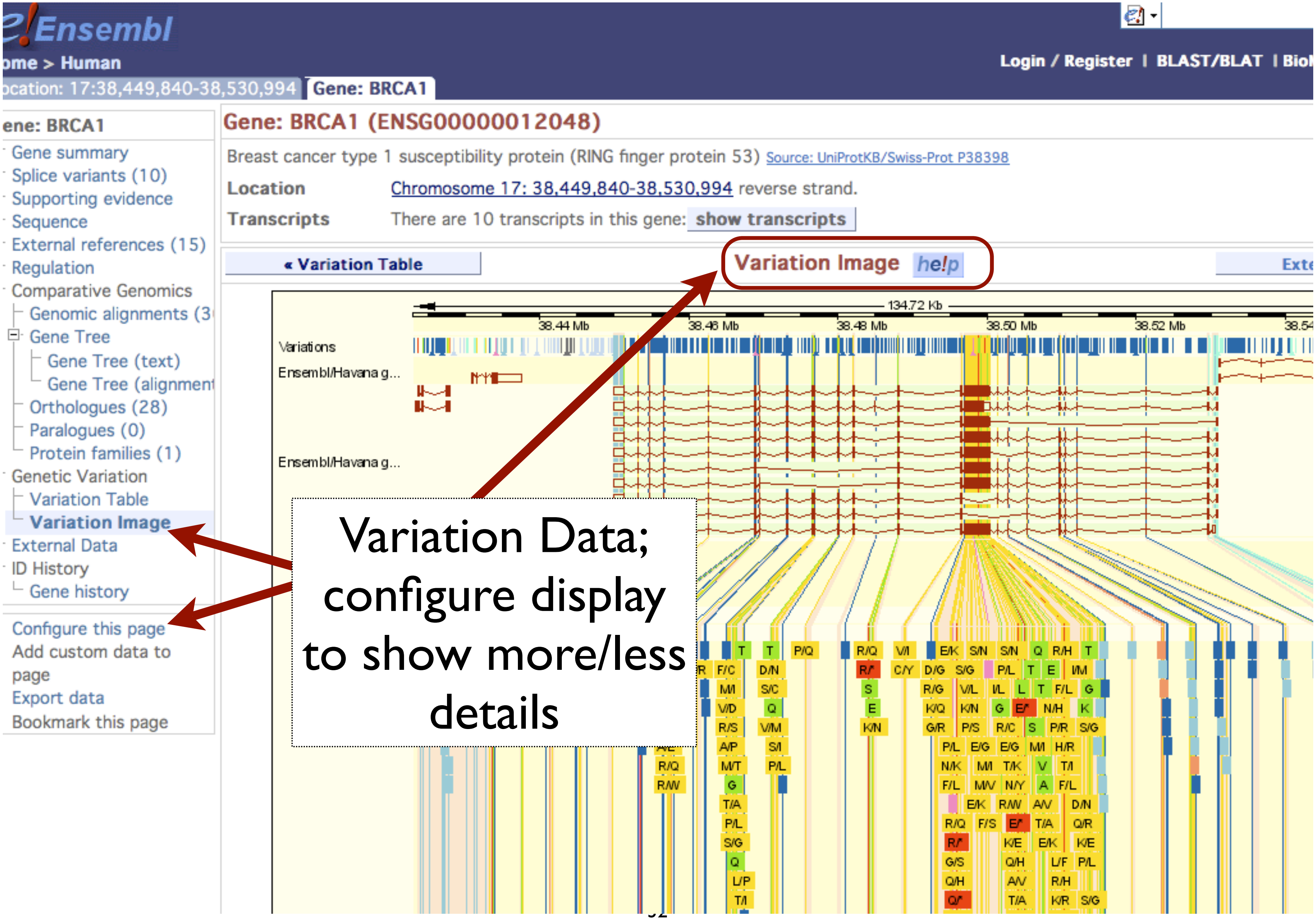
Breast cancer type 1 susceptibility protein (RING finger protein 53) Source: UniProtKB/Swiss-Prot P3

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	ENSP0000037728[illegible]	protein_coding
BRCA1-209	ENST00000393683	ENSP00000377[illegible]88	protein_coding
BRCA1-210	ENST00000393691	ENSP00000[illegible]7294	protein_coding

The Splice Variants page shows you information about the transcripts



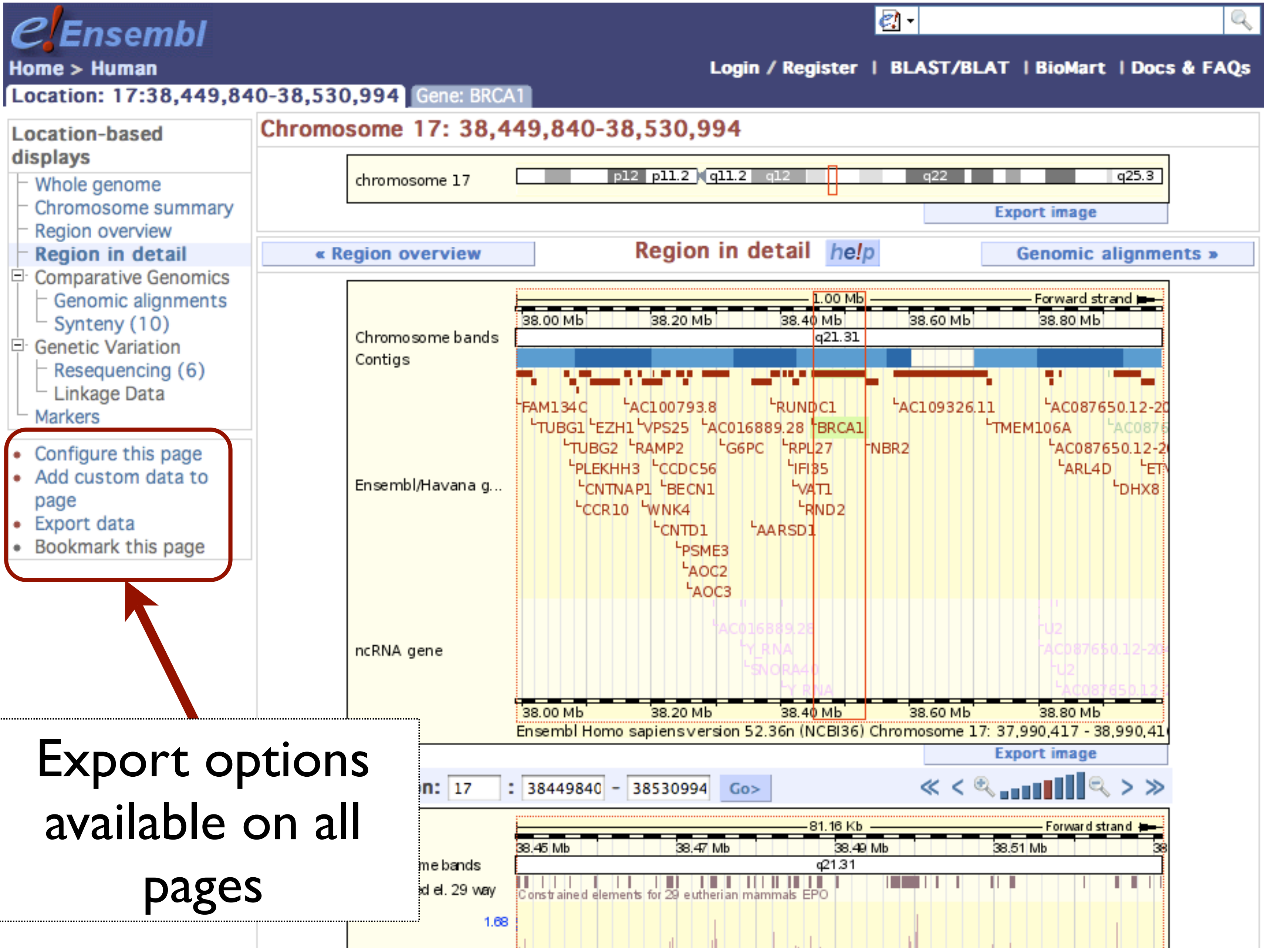
Gene: BRCA1 (ENSG00000012048)

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.

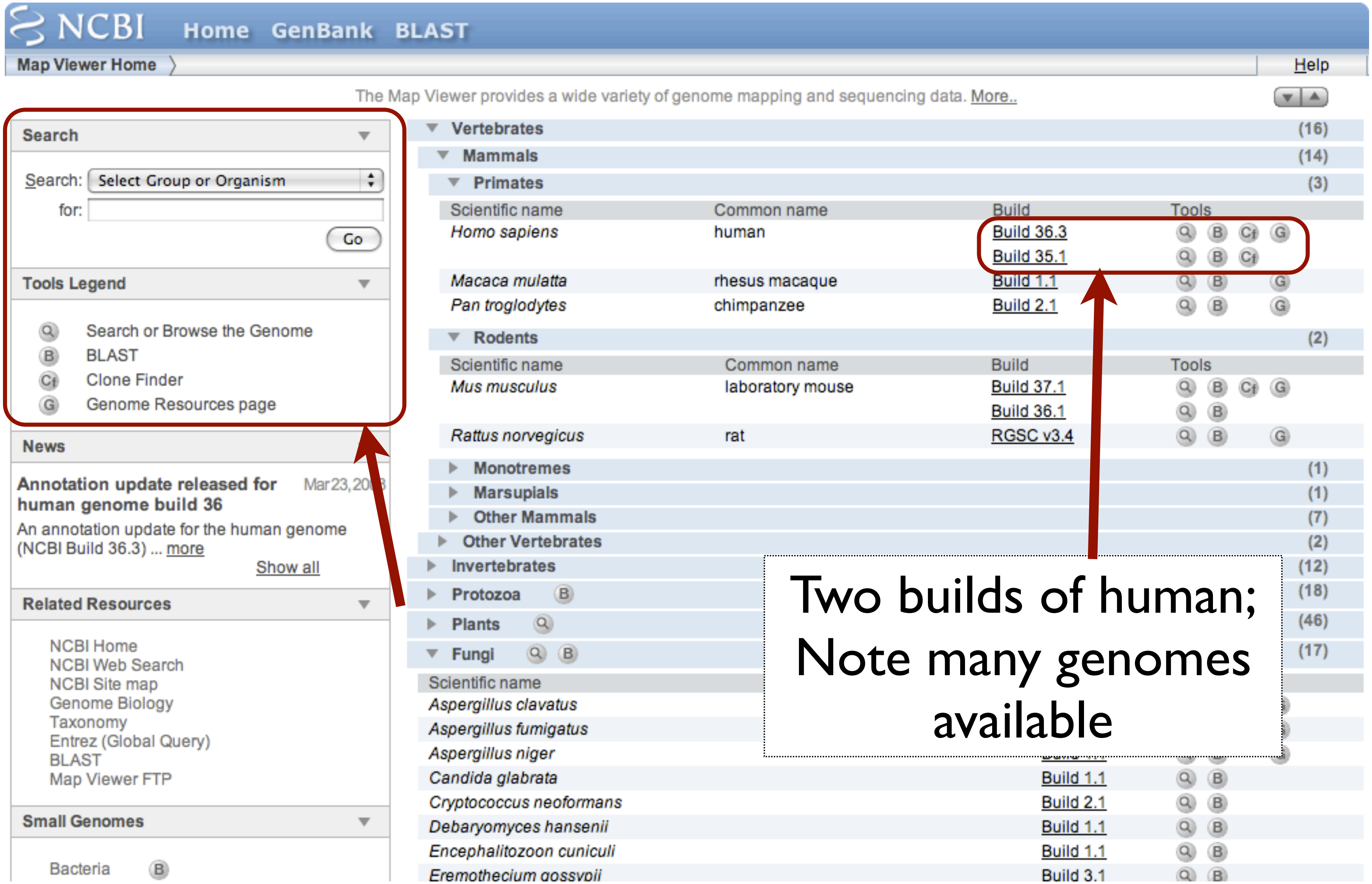
Transcripts There are 10 transcripts in this gene: **show transcripts**

Variation Data; configure display to show more/less details



Export options available on all pages

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, 2008

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)

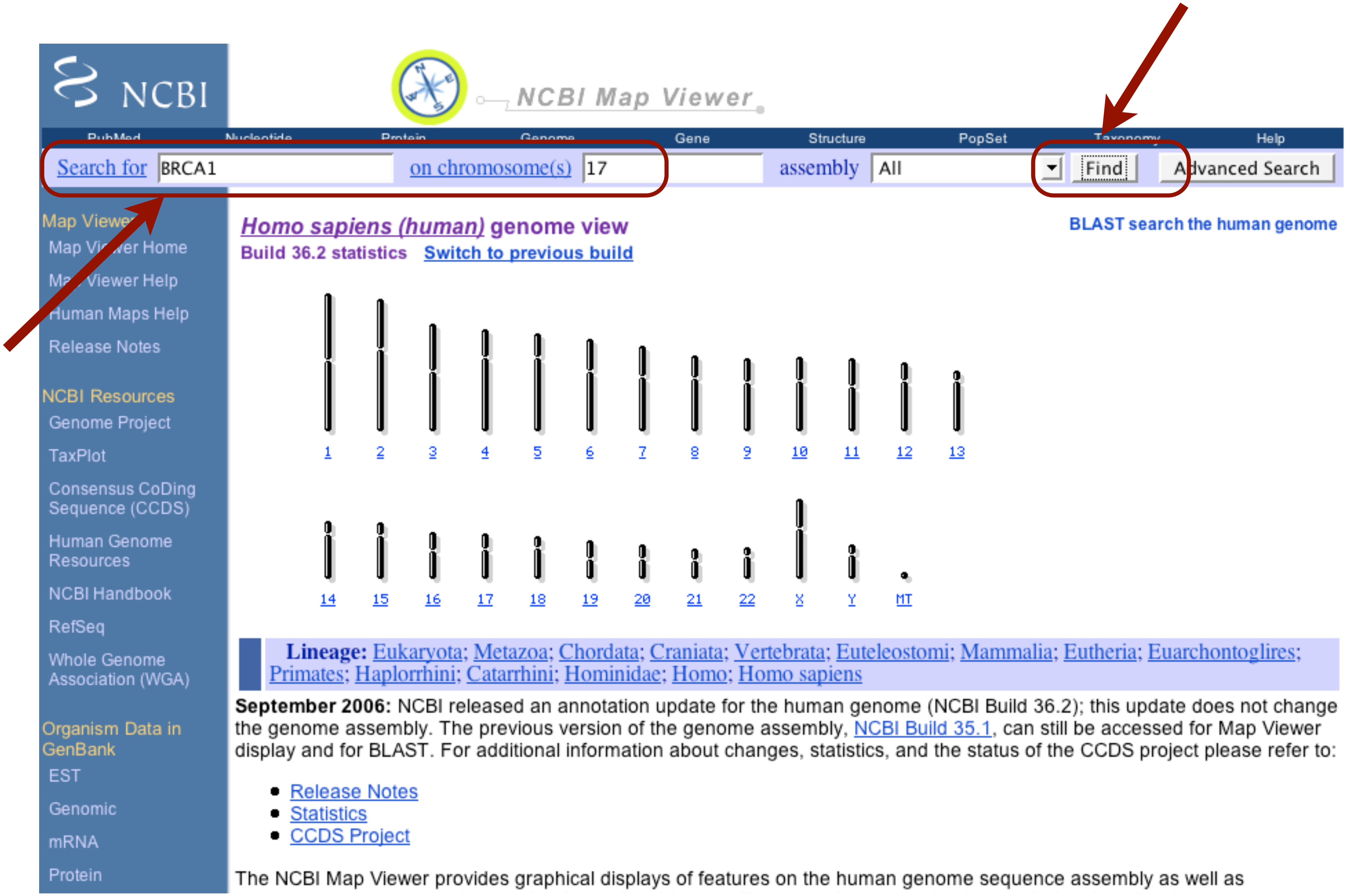
Protozoa B (18)

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



Homo sapiens (human) genome view

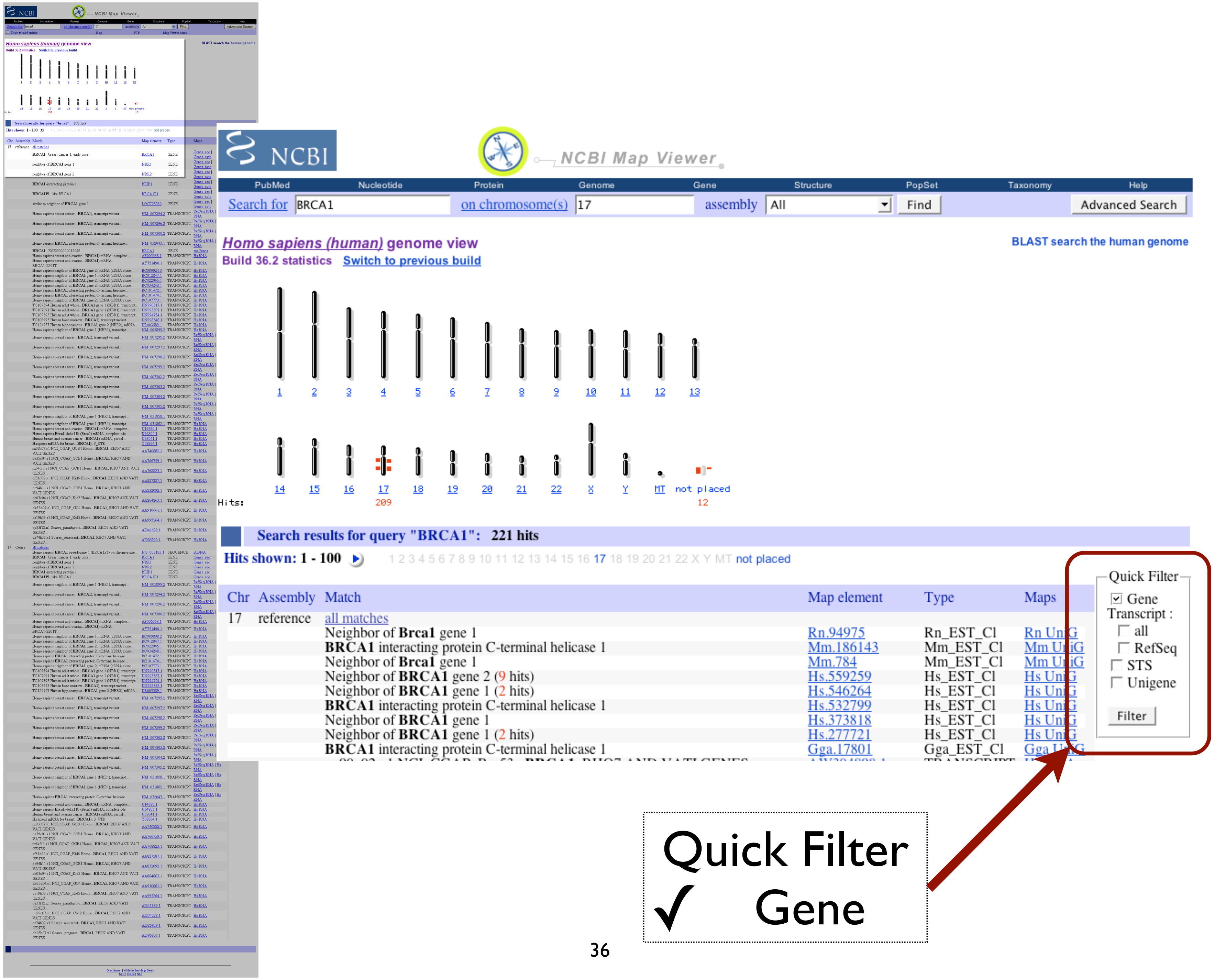
Build 36.2 statistics Switch to previous build

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



Search for: BRCA1

on chromosome(s): 17

assembly: All

Homo sapiens (human) genome view

Build 36.2 statistics

Search results for query "BRCA1": 221 hits

Hits shown: 1 - 100

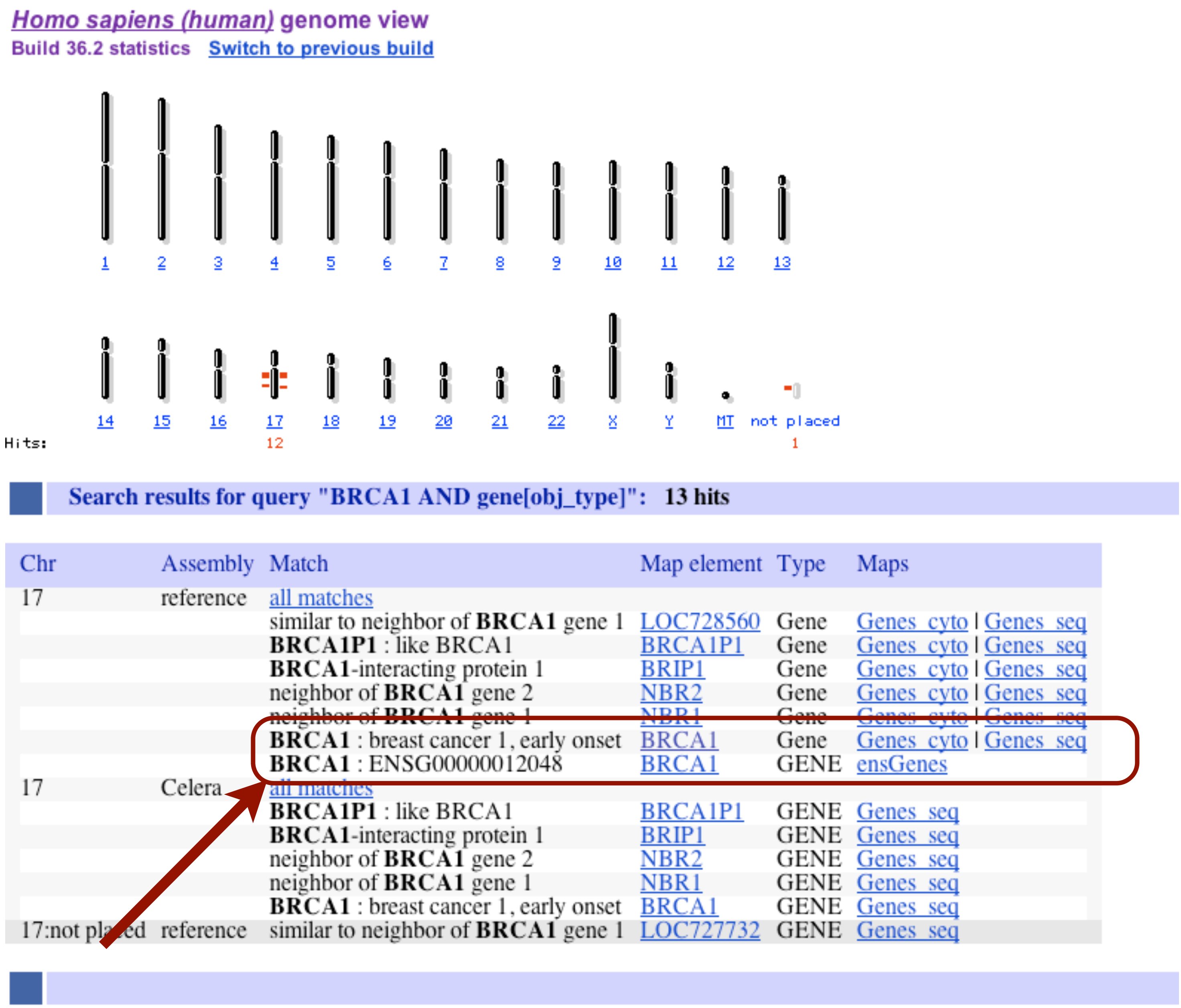
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X Y MT not placed

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

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

Quick Filter

✓ Gene

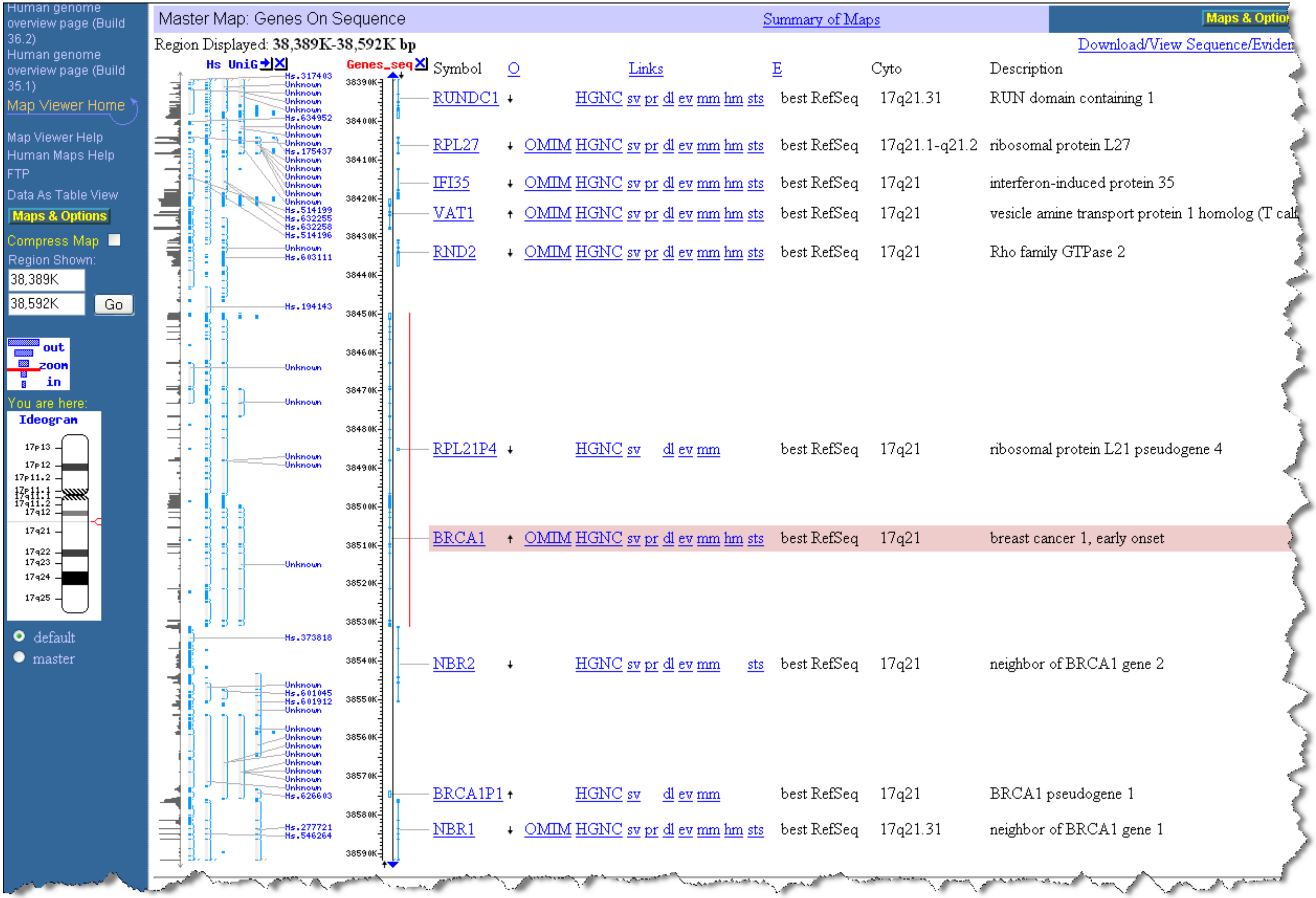


Homo sapiens (human) genome view

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



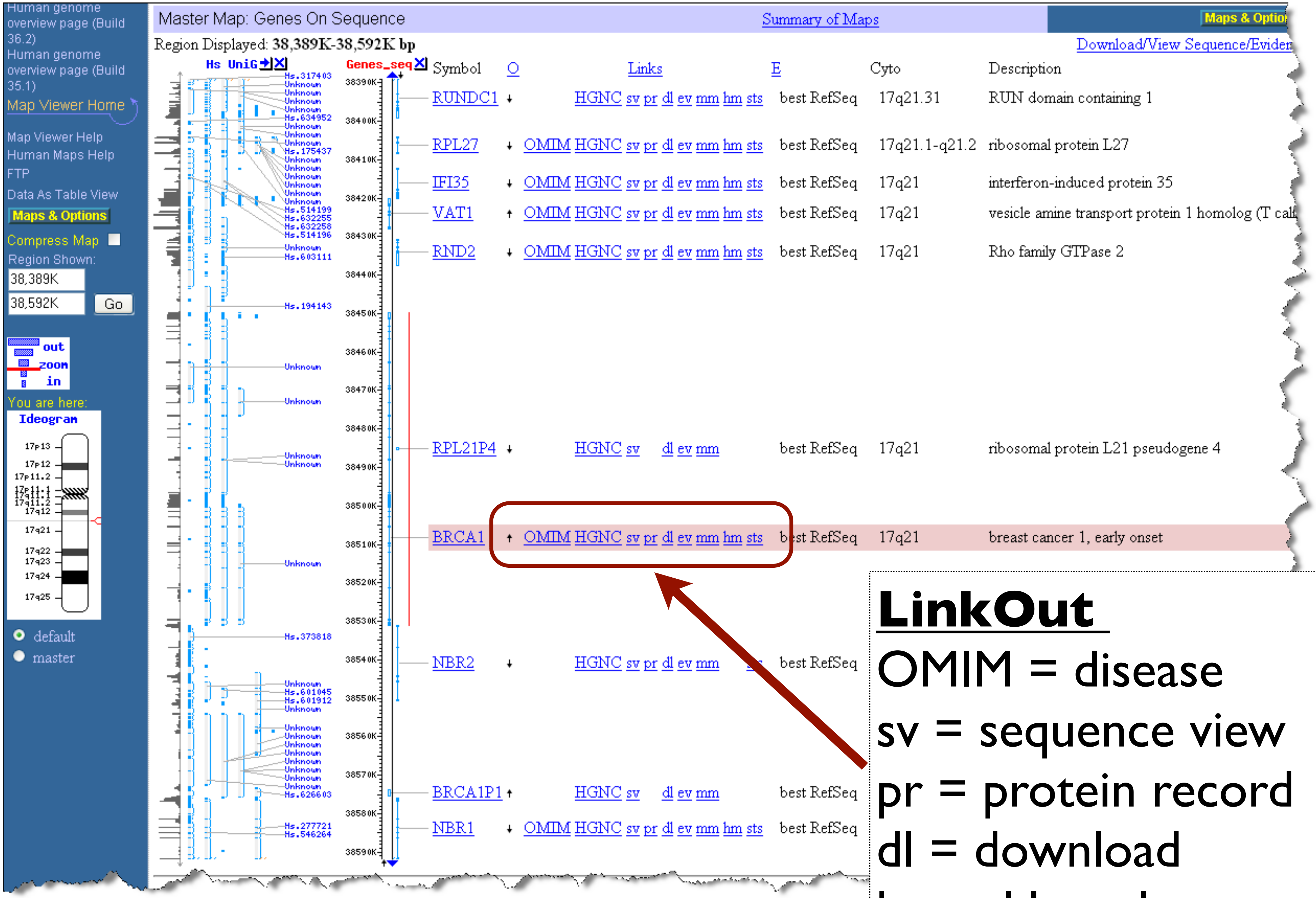
Human genome overview page (Build 36.2)
Human genome overview page (Build 35.1)
Map Viewer Home
Map Viewer Help
Human Maps Help
FTP
Data As Table View
Maps & Options
Compress Map ☐
Region Shown:
38,389K
38,592K
Go
You are here:
○ default
○ master

Master Map: Genes On Sequence
Summary of Maps
Maps & Optio
Region Displayed: 38,389K-38,592K bp
Download/View Sequence/Eviden

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 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	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

Bioinformatics

Session 3.1 - Discovering GEO, the Gene Expression Omnibus.

Functional Genomics

- What kinds of questions can you ask with microarray data?

✓ basic research

✓ drug target discovery

✓ biomarker discovery

✓ pharmacology & toxicogenomics

✓ clinical diagnosis - prognosis, diagnosis, & disease classification

✓ gene regulatory networks

✓ protein-DNA binding

✓ + more

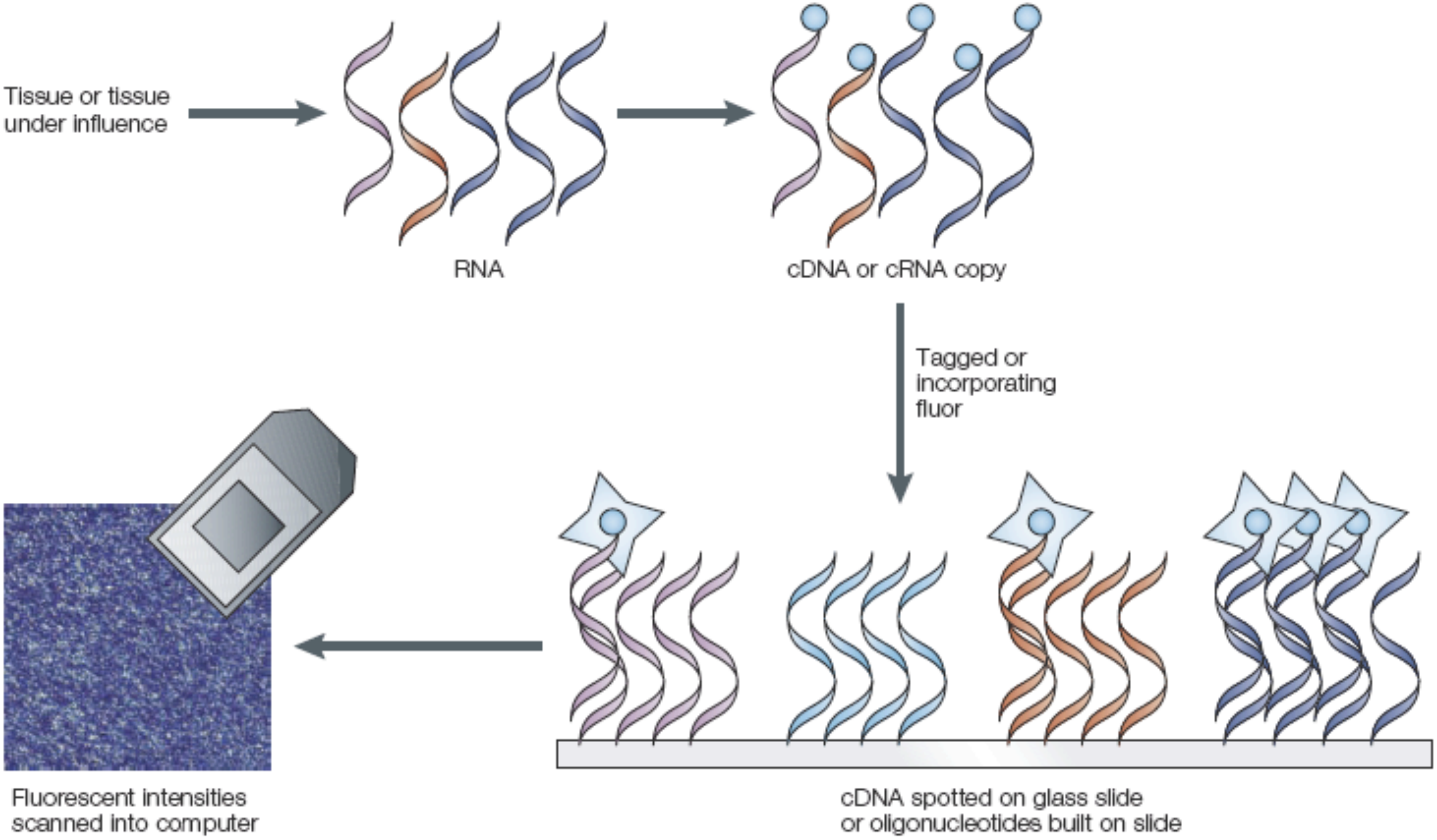


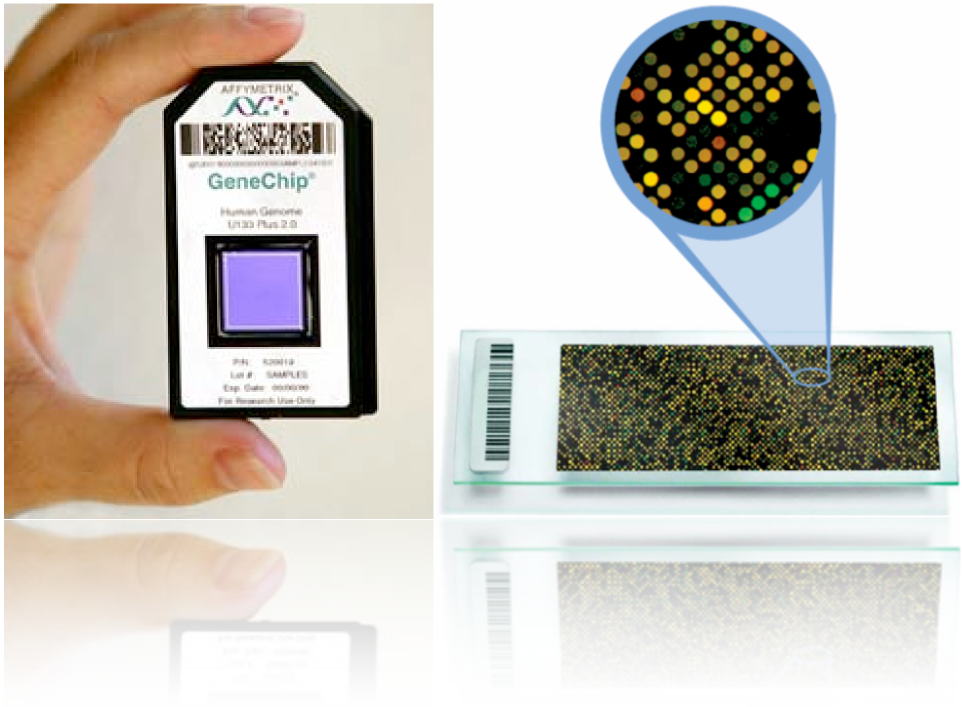
Figure 1 | **Schematized experimental process using a microarray.** Although the specific protocols differ, the microarray approach first involves isolating RNA or messenger RNA from appropriate biological samples, making the RNA (or a copy of it) fluorescent, hybridizing it to the microarray, washing off the excess and scanning the microarray under laser light.

source: Butte, A Nature Reviews 2002

Different Platforms

in situ oligonucleotide
single sample, absolute levels

spotted DNA/cDNA
two samples, relative levels



Microarray Experiment

- Design
- Collect
- Pre-Process
- Analyze
- Interpret
- Submit
- Publish

Example = Normalization

Examples = Distance measures, data classification, clustering, + more

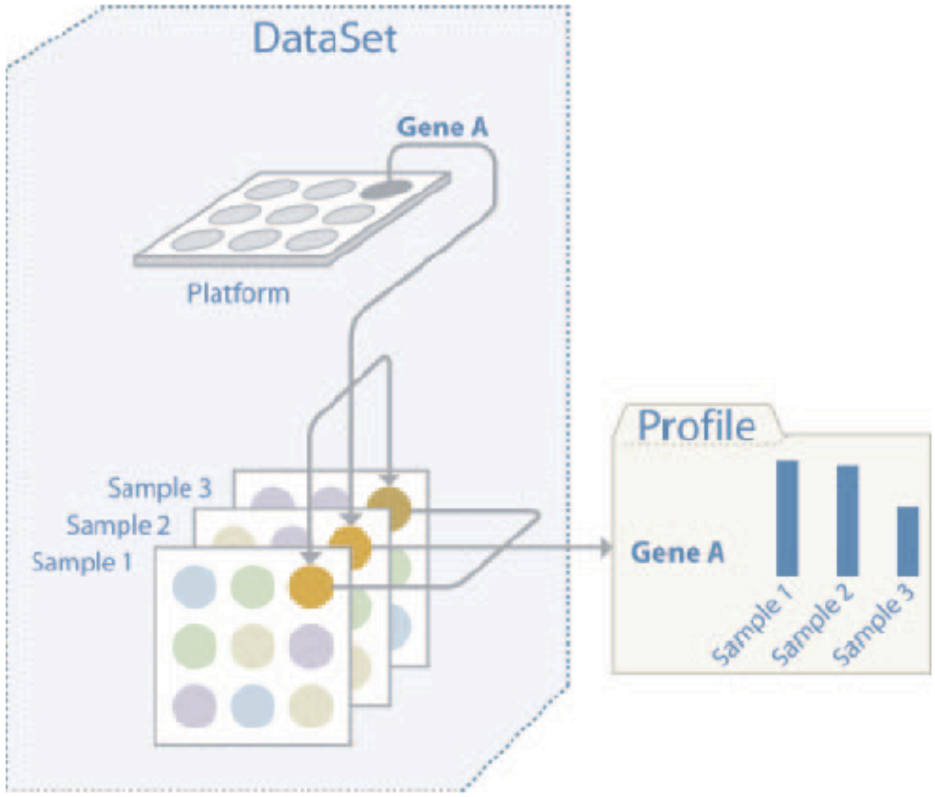
Rate Limiting Step = What do these results actually mean?

Public Microarray Data

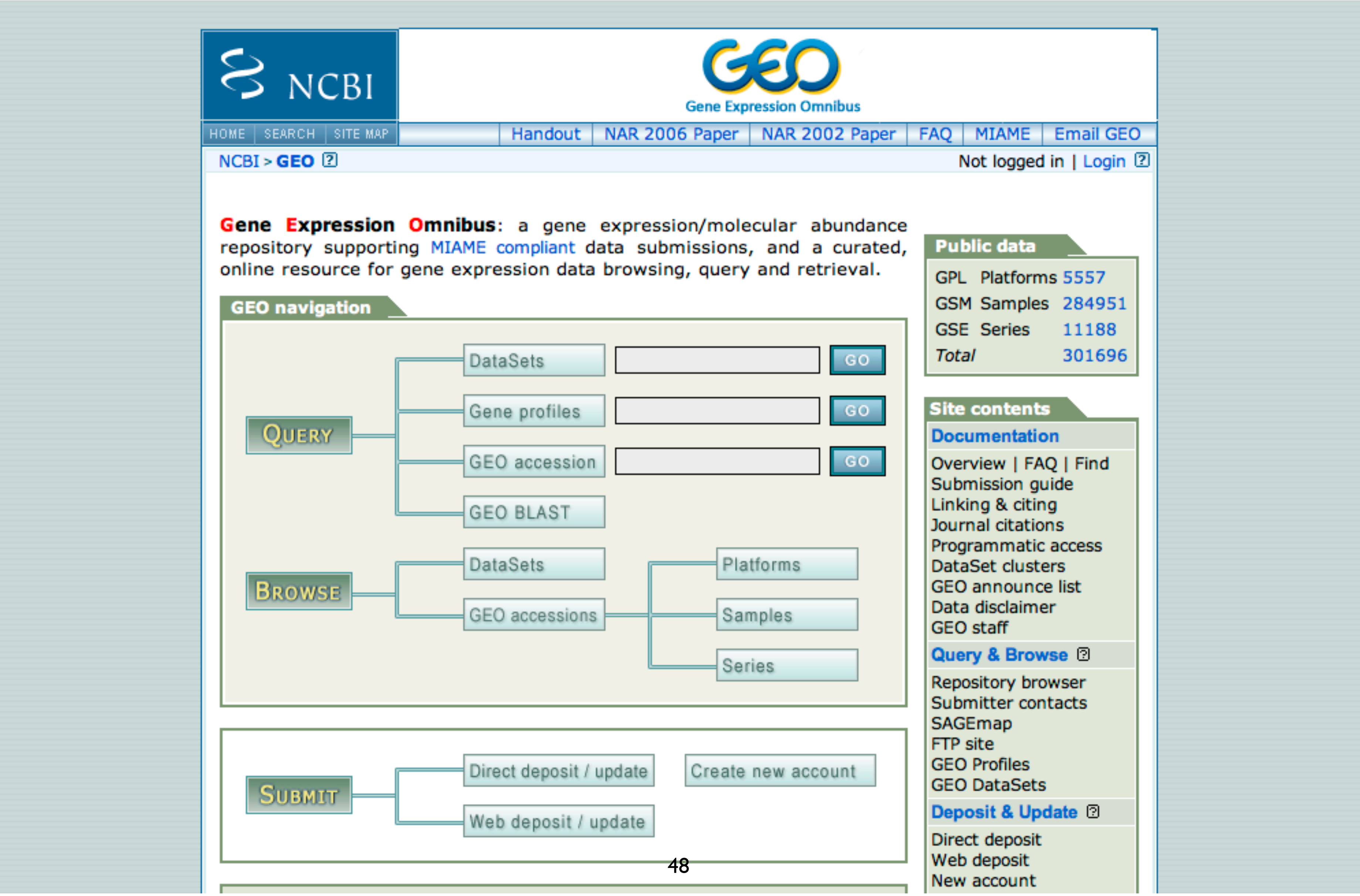
- The Gene Expression Omnibus (GEO)
    - ◉ repository/archive gene expression data
- data submitted by the research community in fulfillment of journal requirements
- this public data represents an untapped resource; potential discovery from existing data sets is at your fingertips

GEO Database

Organized by:
Platform
Sample
Series/DataSet
Profile



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



Gene Expression Omnibus: a gene expression/molecular abundance repository supporting MIAME compliant data submissions, and a curated, online resource for gene expression data browsing, query and retrieval.

Public data

GPL	Platforms	5557
GSM	Samples	284951
GSE	Series	11188
Total		301696

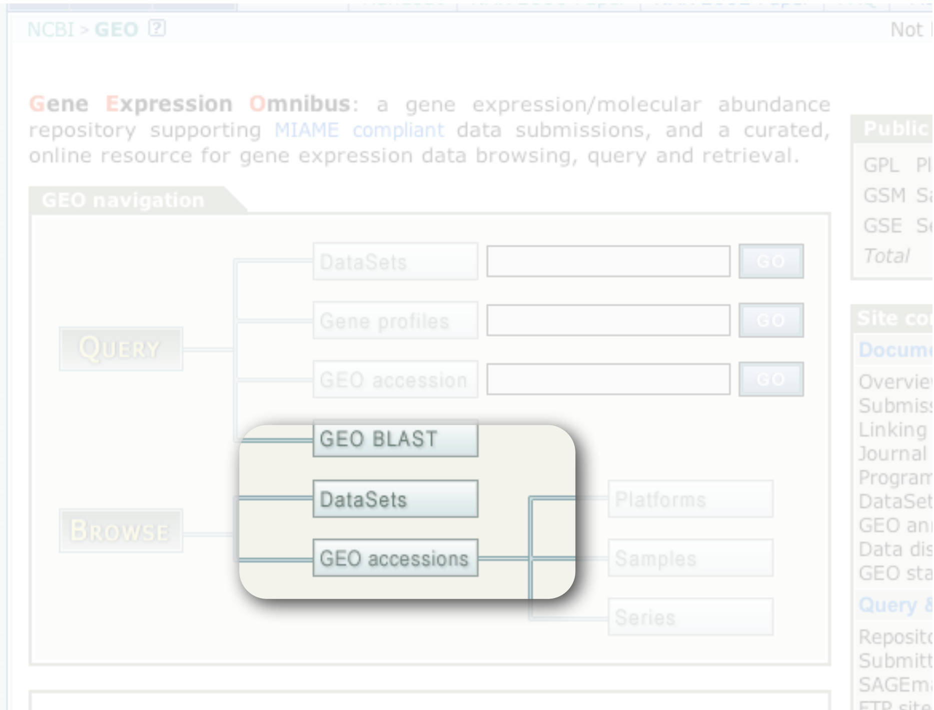
Searching GEO

- Are you interested in a particular type of expt?

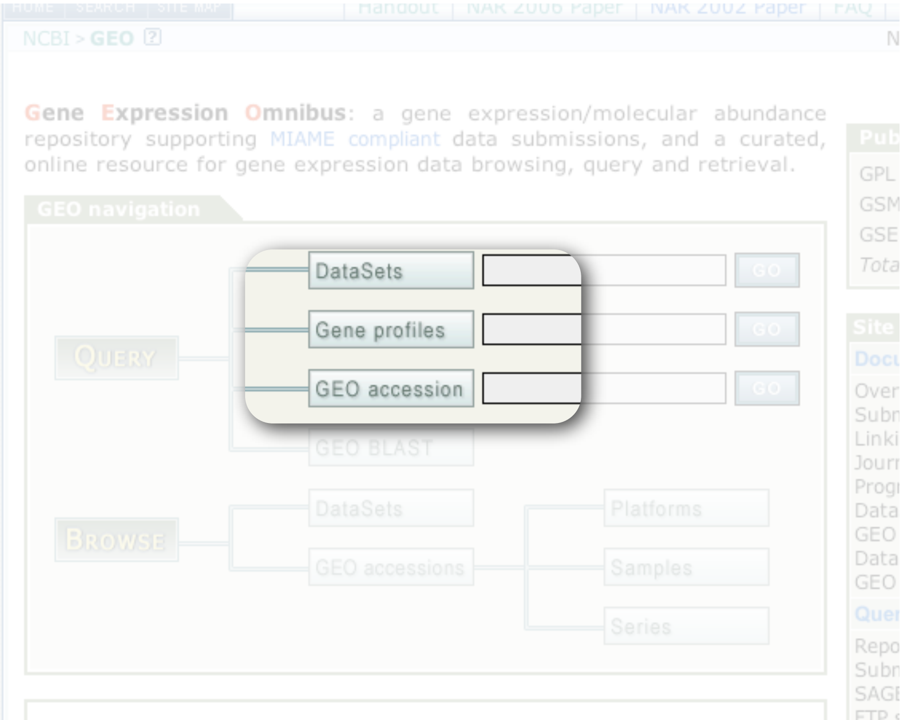
✓ GEO DataSets

- Are you looking for your favorite gene?

✓ GEO Profiles



Gene Expression Omnibus: a gene expression/molecular abundance repository supporting MIAME compliant data submissions, and a curated, online resource for gene expression data browsing, query and retrieval.

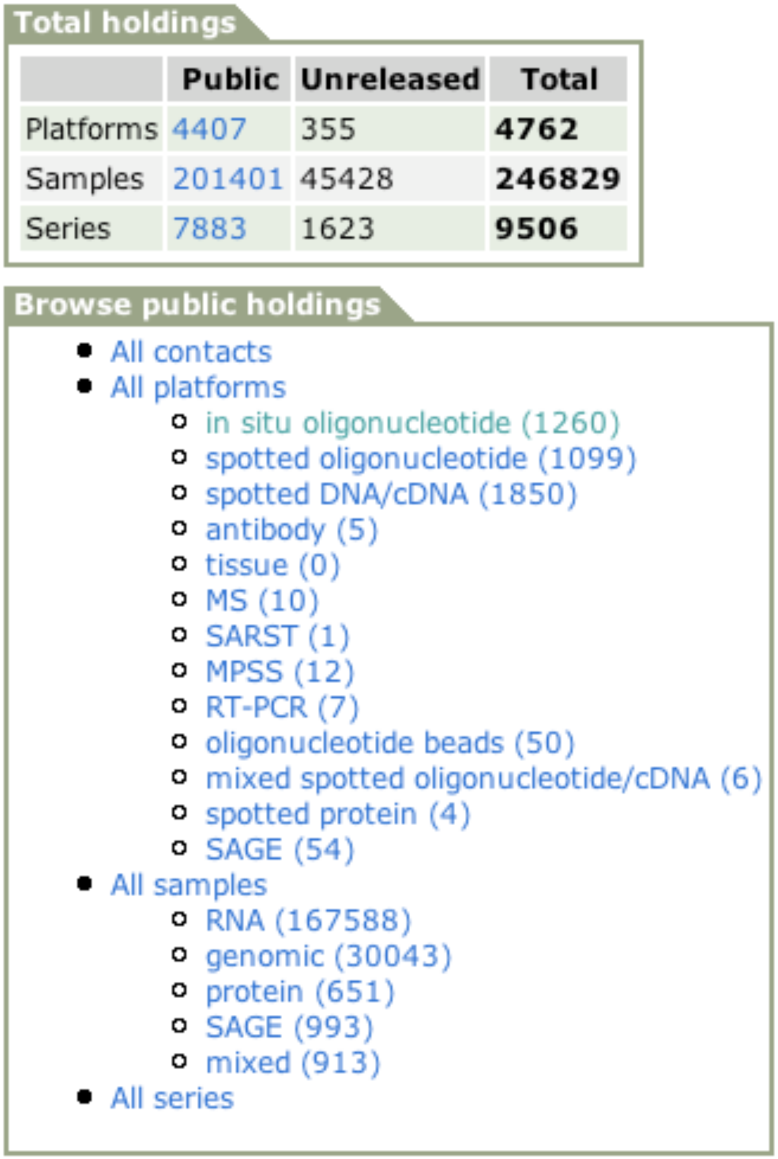


Gene Expression Omnibus: a gene expression/molecular abundance repository supporting MIAME compliant data submissions, and a curated, online resource for gene expression data browsing, query and retrieval.

Data in GEO

>120,000 samples
>3.2 billion measurement
200+ organisms
from >2000 labs

freely available online
ftp downloads

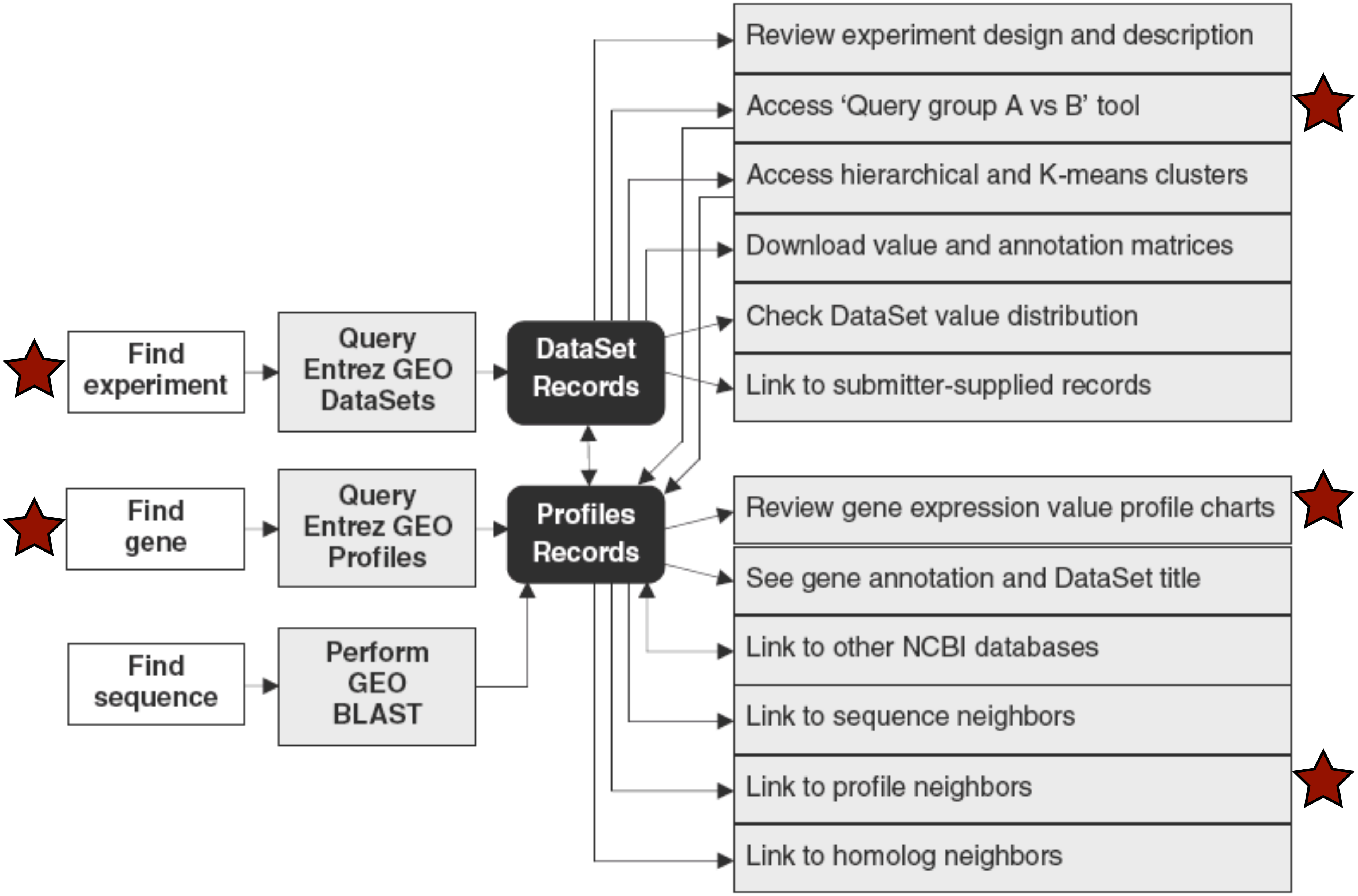


Total holdings

	Public	Unreleased	Total
Platforms	4407	355	**4762**
Samples	201401	45428	**246829**
Series	7883	1623	**9506**

Browse public holdings

- All contacts
- All platforms
    - in situ oligonucleotide (1260)
    - spotted oligonucleotide (1099)
    - spotted DNA/cDNA (1850)
    - antibody (5)
    - tissue (0)
    - MS (10)
    - SARST (1)
    - MPSS (12)
    - RT-PCR (7)
    - oligonucleotide beads (50)
    - mixed spotted oligonucleotide/cDNA (6)
    - spotted protein (4)
    - SAGE (54)
- All samples
    - RNA (167588)
    - genomic (30043)
    - protein (651)
    - SAGE (993)
    - mixed (913)
- All series



Find experiment

Query Entrez GEO DataSets

DataSet Records

Find gene

Query Entrez GEO Profiles

Profiles Records

Find sequence

Perform GEO BLAST

Review experiment design and description

Access ‘Query group A vs B’ tool

Access hierarchical and K-means clusters

Download value and annotation matrices

Check DataSet value distribution

Link to submitter-supplied records

Review gene expression value profile charts

See gene annotation and DataSet title

Link to other NCBI databases

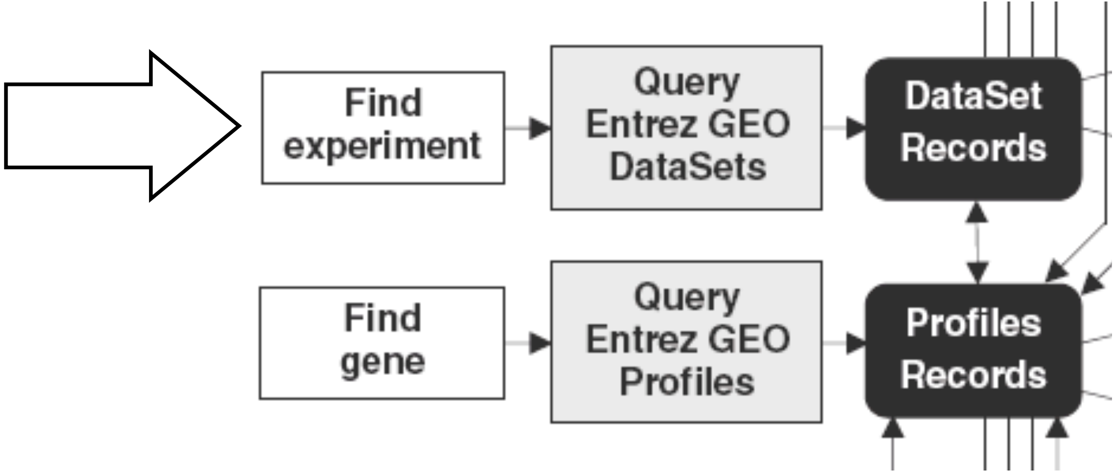
Link to sequence neighbors

Link to profile neighbors

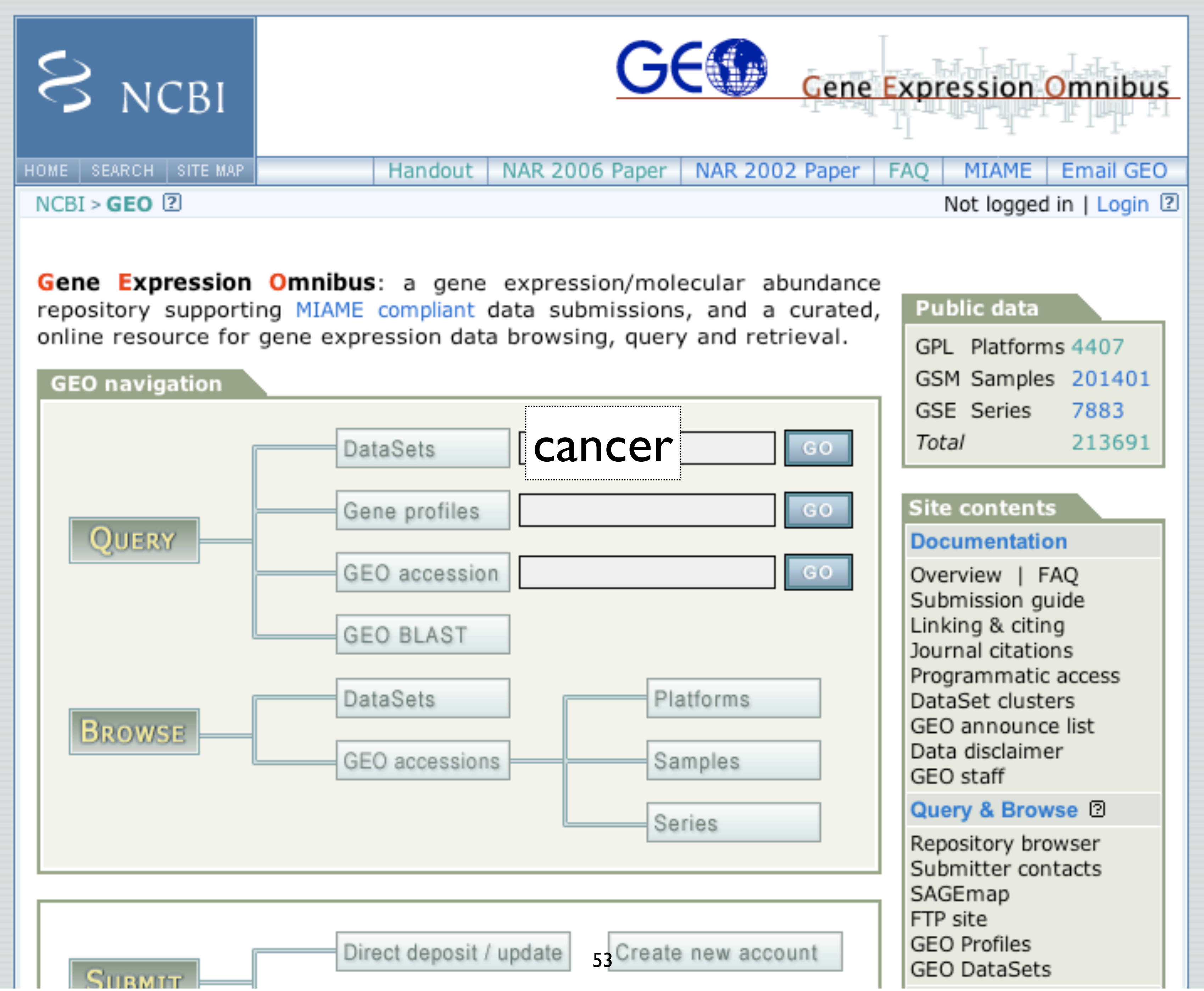
Link to homolog neighbors

An Example

- Find microarray experiments that look at the expression of genes in cancer



You can use these GEO data mining tools for quick and easy identification of relevant & noteworthy data sets. For serious analyses, you should download the data and use a microarray data analysis software suite.

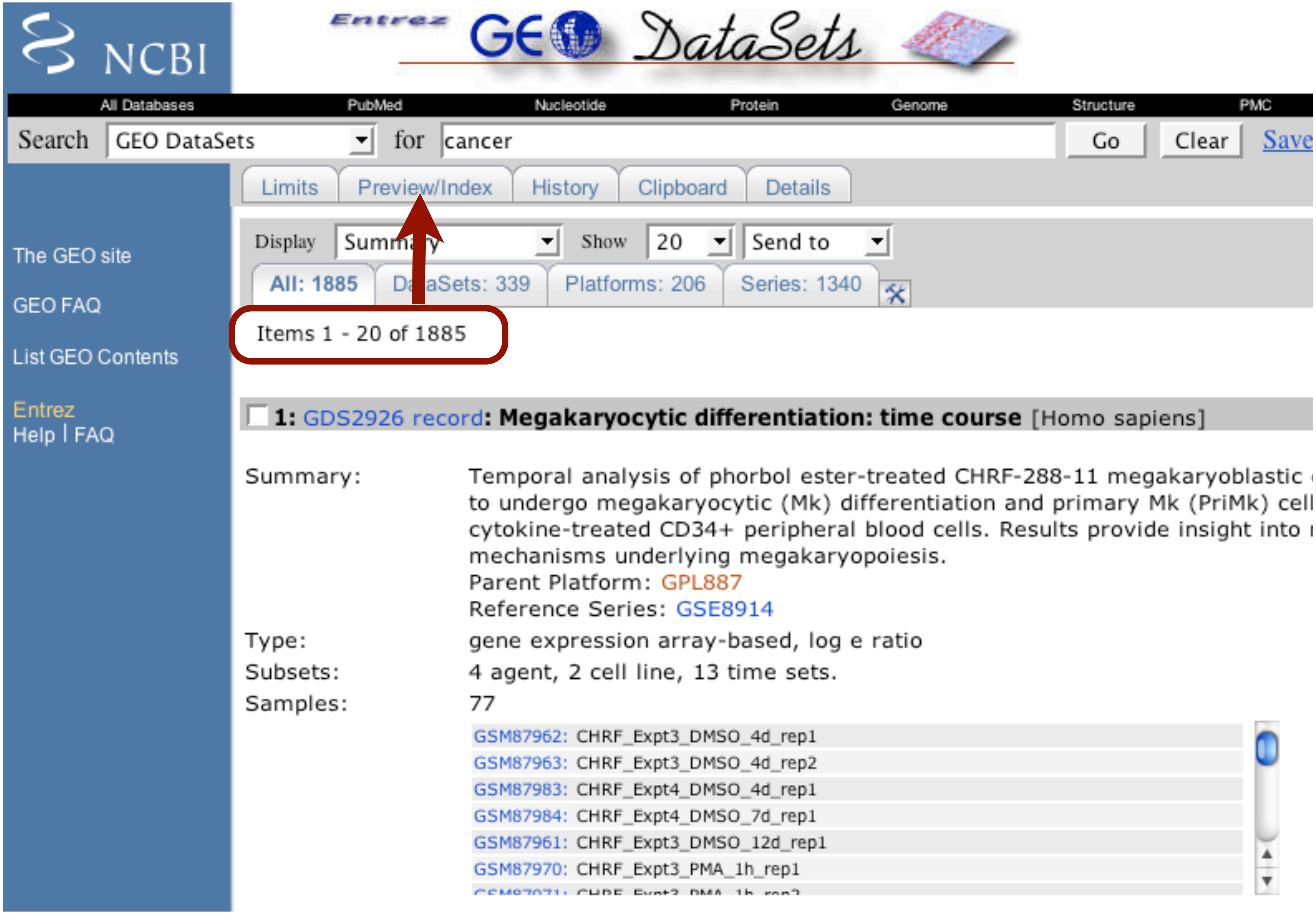


Gene Expression Omnibus: a gene expression/molecular abundance repository supporting MIAME compliant data submissions, and a curated, online resource for gene expression data browsing, query and retrieval.

DataSets: cancer

Public data

GPL	Platforms	4407
GSM	Samples	201401
GSE	Series	7883
Total		213691



Search GEO DataSets for cancer

Display Summary Show 20 Send to

All: 1885 DataSets: 339 Platforms: 206 Series: 1340

Items 1 - 20 of 1885

☐ **1:** GDS2926 record**: Megakaryocytic differentiation: time course** [Homo sapiens]

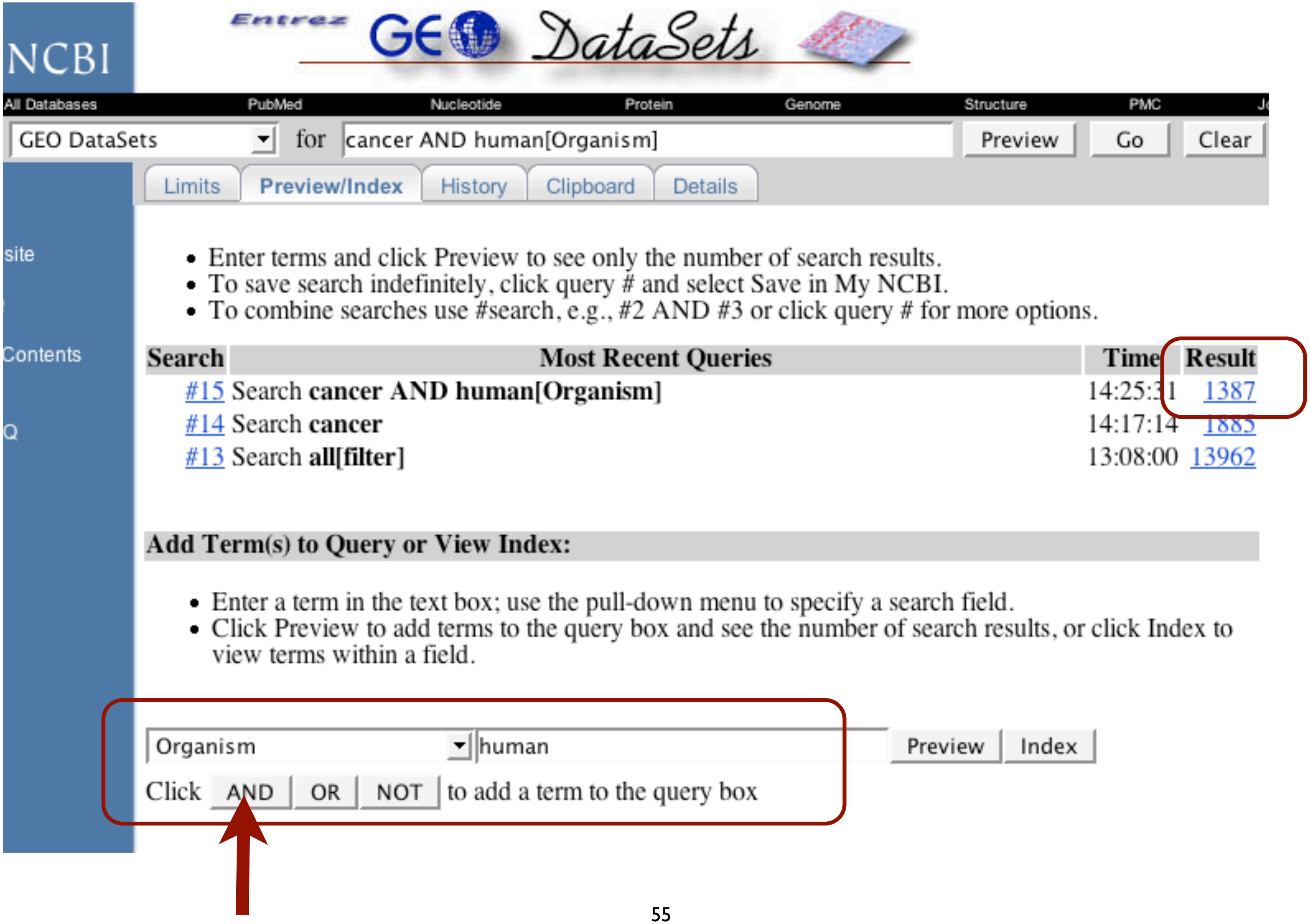
Summary: Temporal analysis of phorbol ester-treated CHRF-288-11 megakaryoblastic to undergo megakaryocytic (Mk) differentiation and primary Mk (PriMk) cell cytokine-treated CD34+ peripheral blood cells. Results provide insight into mechanisms underlying megakaryopoiesis.
Parent Platform: GPL887
Reference Series: GSE8914

Type: gene expression array-based, log e ratio

Subsets: 4 agent, 2 cell line, 13 time sets.

Samples: 77

GSM87962: CHRF_Expt3_DMSO_4d_rep1
GSM87963: CHRF_Expt3_DMSO_4d_rep2
GSM87983: CHRF_Expt4_DMSO_4d_rep1
GSM87984: CHRF_Expt4_DMSO_7d_rep1
GSM87961: CHRF_Expt3_DMSO_12d_rep1
GSM87970: CHRF_Expt3_PMA_1h_rep1



GEO DataSets for cancer AND human[Organism]

- Enter terms and click Preview to see only the number of search results.
- To save search indefinitely, click query # and select Save in My NCBI.
- To combine searches use #search, e.g., #2 AND #3 or click query # for more options.

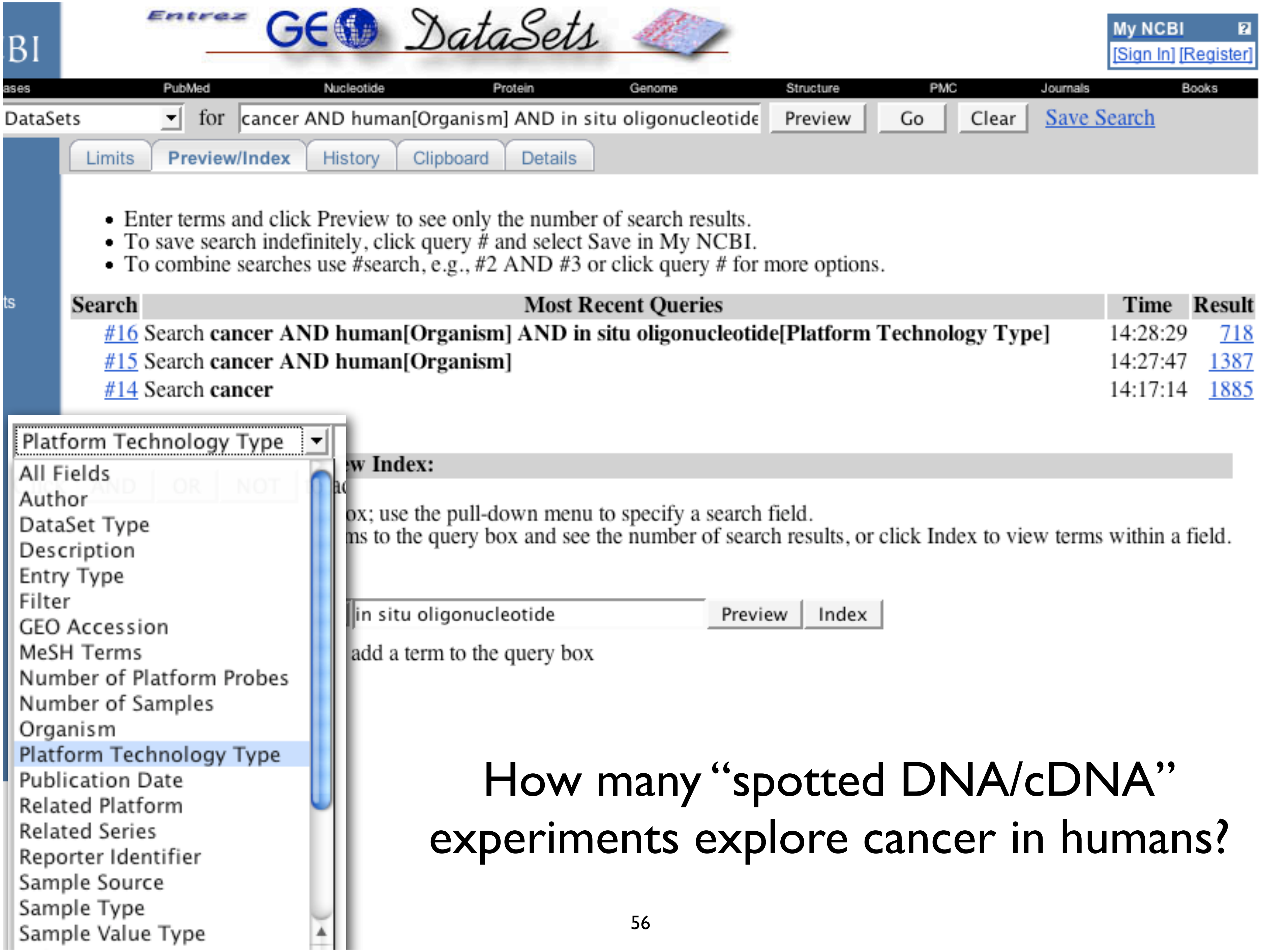
Search	Most Recent Queries	Time	Result
#15	Search **cancer AND human[Organism]**	14:25:31	1387
#14	Search **cancer**	14:17:14	1885
#13	Search **all[filter]**	13:08:00	13962

Add Term(s) to Query or View Index:

- Enter a term in the text box; use the pull-down menu to specify a search field.
- Click Preview to add terms to the query box and see the number of search results, or click Index to view terms within a field.

Organism human

Click AND OR NOT to add a term to the query box



- Enter terms and click Preview to see only the number of search results.
- To save search indefinitely, click query # and select Save in My NCBI.
- To combine searches use #search, e.g., #2 AND #3 or click query # for more options.

Search	Most Recent Queries	Time	Result
#16	Search **cancer AND human[Organism] AND in situ oligonucleotide[Platform Technology Type]**	14:28:29	718
#15	Search **cancer AND human[Organism]**	14:27:47	1387
#14	Search **cancer**	14:17:14	1885

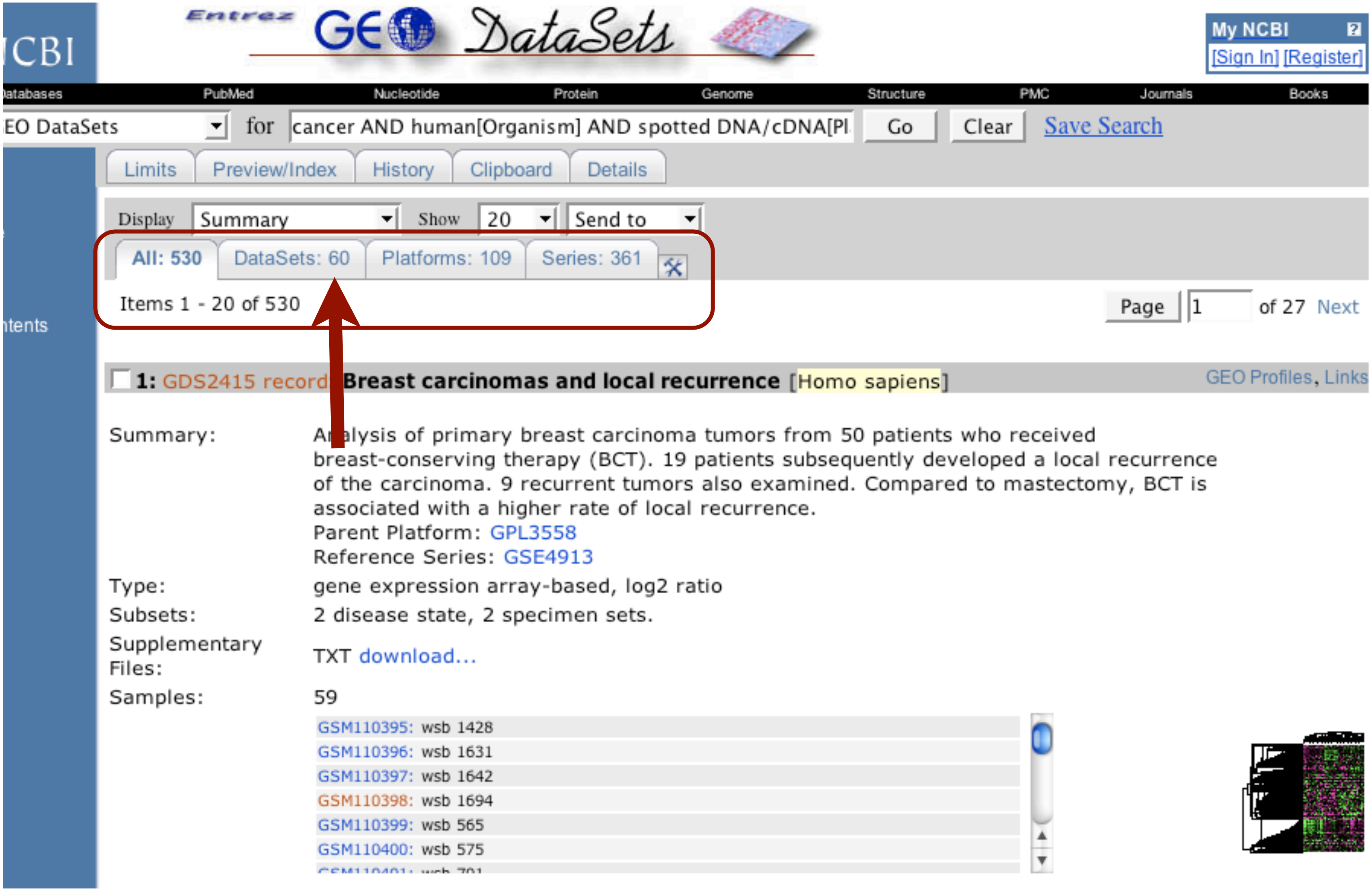
w Index:

ox; use the pull-down menu to specify a search field.

ms to the query box and see the number of search results, or click Index to view terms within a field.

add a term to the query box

How many “spotted DNA/cDNA” experiments explore cancer in humans?



EO DataSets for cancer AND human[Organism] AND spotted DNA/cDNA[Pl

Display Summary Show 20 Send to

All: 530 DataSets: 60 Platforms: 109 Series: 361

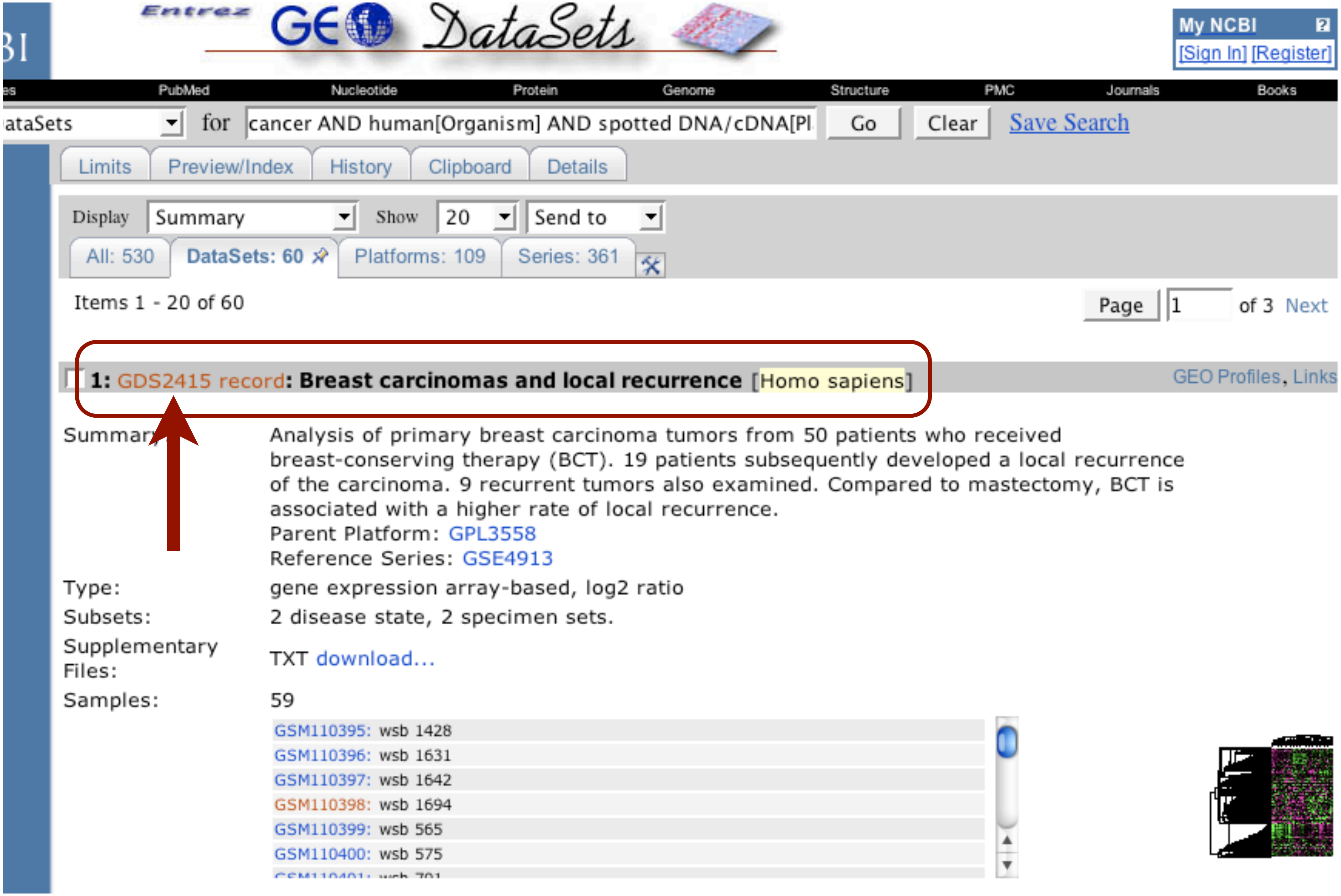
Items 1 - 20 of 530

Page 1 of 27 Next

☐ **1:** GDS2415 record: **Breast carcinomas and local recurrence** [Homo sapiens]

Summary:	Analysis of primary breast carcinoma tumors from 50 patients who received breast-conserving therapy (BCT). 19 patients subsequently developed a local recurrence of the carcinoma. 9 recurrent tumors also examined. Compared to mastectomy, BCT is associated with a higher rate of local recurrence. Parent Platform: GPL3558 Reference Series: GSE4913
Type:	gene expression array-based, log2 ratio
Subsets:	2 disease state, 2 specimen sets.
Supplementary Files:	TXT download...
Samples:	59

GSM110395: wsb 1428
GSM110396: wsb 1631
GSM110397: wsb 1642
GSM110398: wsb 1694
GSM110399: wsb 565
GSM110400: wsb 575



ataSets for cancer AND human[Organism] AND spotted DNA/cDNA[Pl

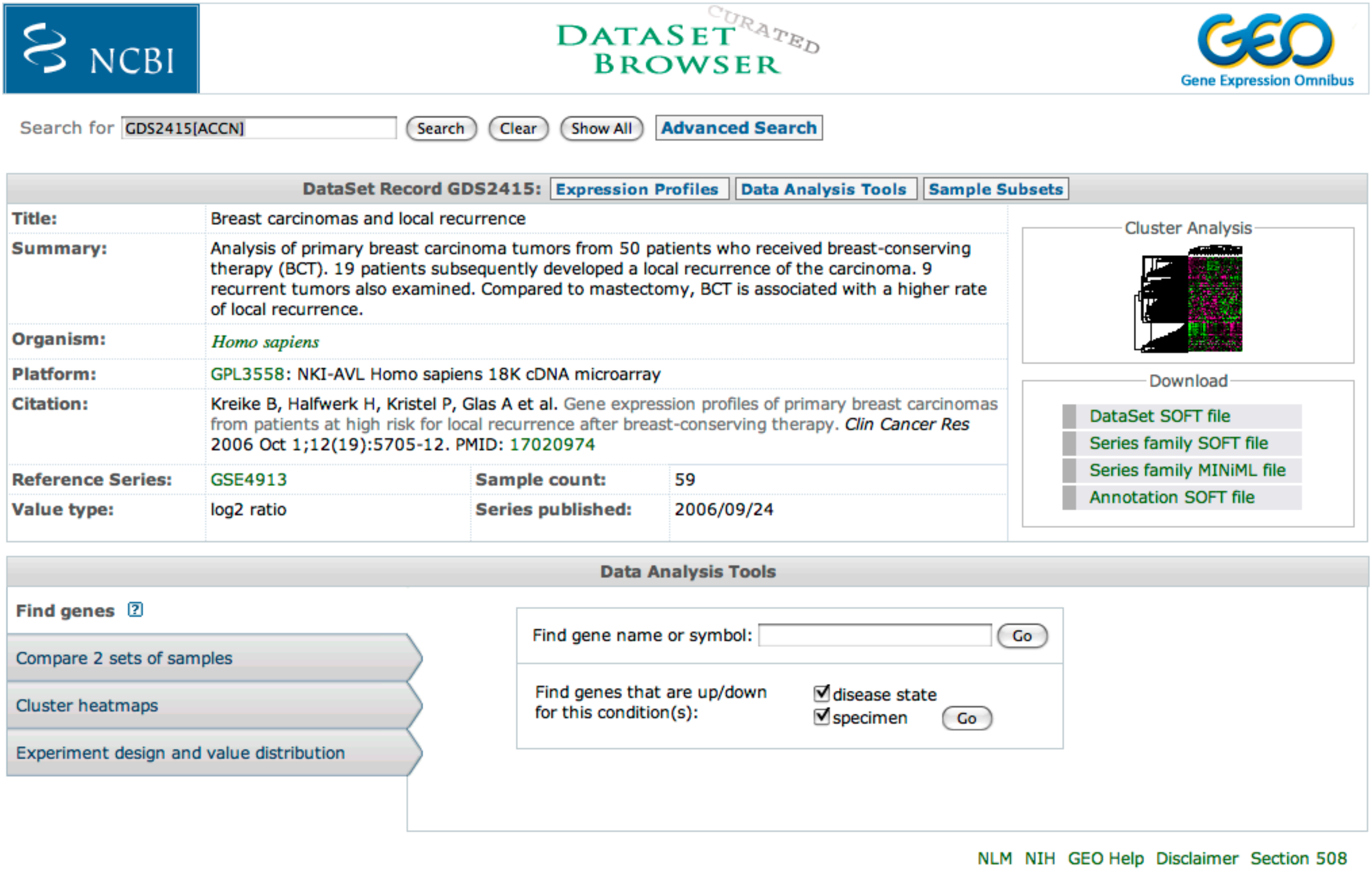
Display Summary Show 20

Items 1 - 20 of 60 Page 1 of 3

☐ **1: GDS2415 record: Breast carcinomas and local recurrence** [Homo sapiens]

Summary:	Analysis of primary breast carcinoma tumors from 50 patients who received breast-conserving therapy (BCT). 19 patients subsequently developed a local recurrence of the carcinoma. 9 recurrent tumors also examined. Compared to mastectomy, BCT is associated with a higher rate of local recurrence. Parent Platform: GPL3558 Reference Series: GSE4913
Type:	gene expression array-based, log2 ratio
Subsets:	2 disease state, 2 specimen sets.
Supplementary Files:	TXT download...
Samples:	59

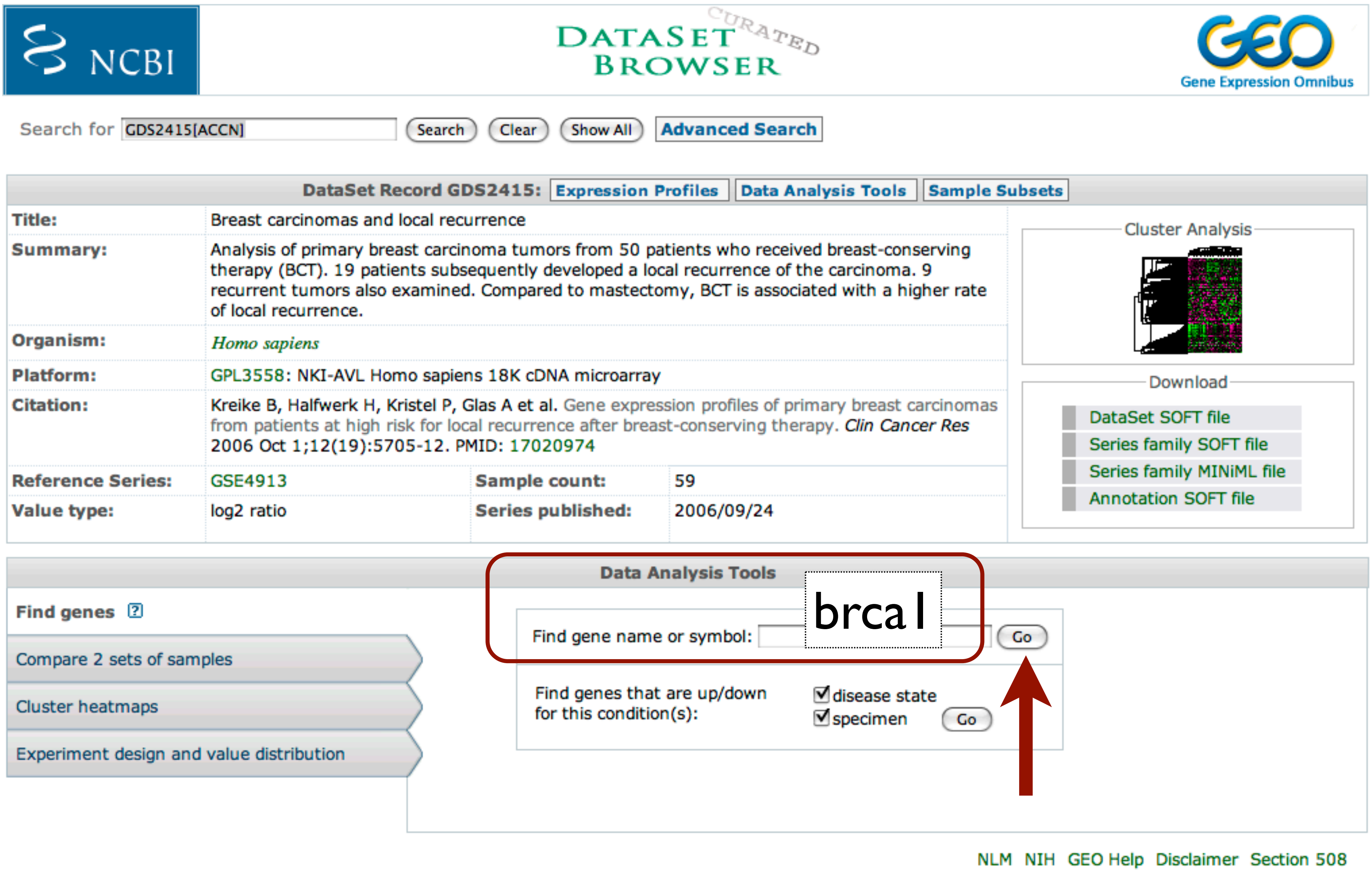
GSM110395: wsb 1428
GSM110396: wsb 1631
GSM110397: wsb 1642
GSM110398: wsb 1694
GSM110399: wsb 565
GSM110400: wsb 575



DataSet Record GDS2415: Expression Profiles | Data Analysis Tools | Sample Subsets

Title:	Breast carcinomas and local recurrence		
Summary:	Analysis of primary breast carcinoma tumors from 50 patients who received breast-conserving therapy (BCT). 19 patients subsequently developed a local recurrence of the carcinoma. 9 recurrent tumors also examined. Compared to mastectomy, BCT is associated with a higher rate of local recurrence.		
Organism:	*Homo sapiens*		
Platform:	GPL3558: NKI-AVL Homo sapiens 18K cDNA microarray		
Citation:	Kreike B, Halfwerk H, Kristel P, Glas A et al. Gene expression profiles of primary breast carcinomas from patients at high risk for local recurrence after breast-conserving therapy. *Clin Cancer Res* 2006 Oct 1;12(19):5705-12. PMID: 17020974		
Reference Series:	GSE4913	**Sample count:**	59
Value type:	log2 ratio	**Series published:**	2006/09/24

Data Analysis Tools



Search for GDS2415[ACCN]

DataSet Record GDS2415: Expression Profiles | Data Analysis Tools | Sample Subsets

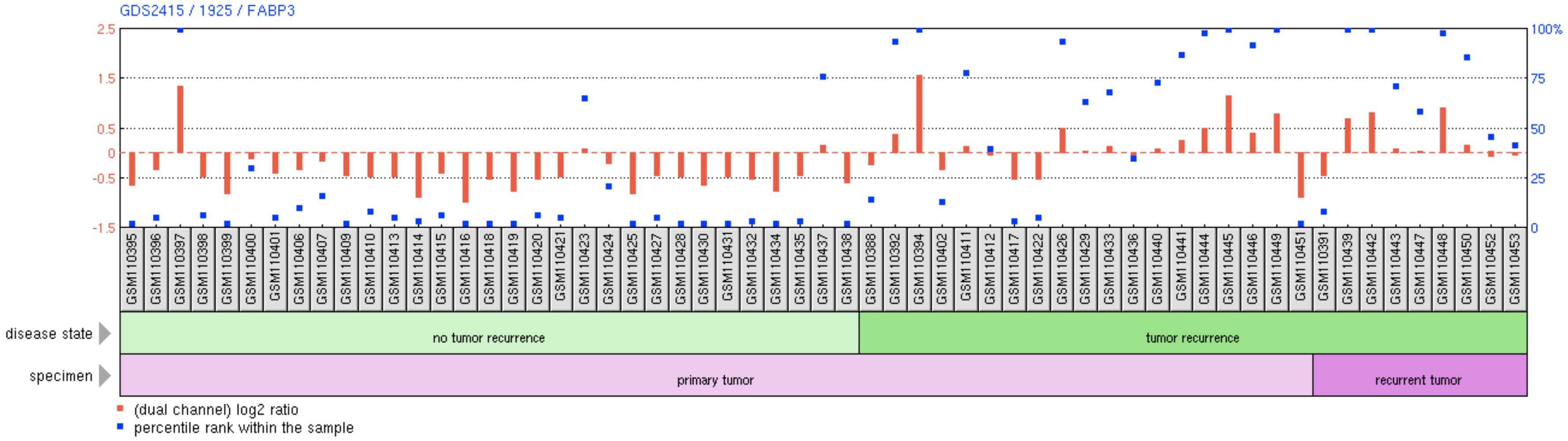
Title:	Breast carcinomas and local recurrence
Summary:	Analysis of primary breast carcinoma tumors from 50 patients who received breast-conserving therapy (BCT). 19 patients subsequently developed a local recurrence of the carcinoma. 9 recurrent tumors also examined. Compared to mastectomy, BCT is associated with a higher rate of local recurrence.
Organism:	*Homo sapiens*
Platform:	GPL3558: NKI-AVL Homo sapiens 18K cDNA microarray
Citation:	Kreike B, Halfwerk H, Kristel P, Glas A et al. Gene expression profiles of primary breast carcinomas from patients at high risk for local recurrence after breast-conserving therapy. *Clin Cancer Res* 2006 Oct 1;12(19):5705-12. PMID: 17020974

Reference Series:	GSE4913	**Sample count:**	59
Value type:	log2 ratio	**Series published:**	2006/09/24

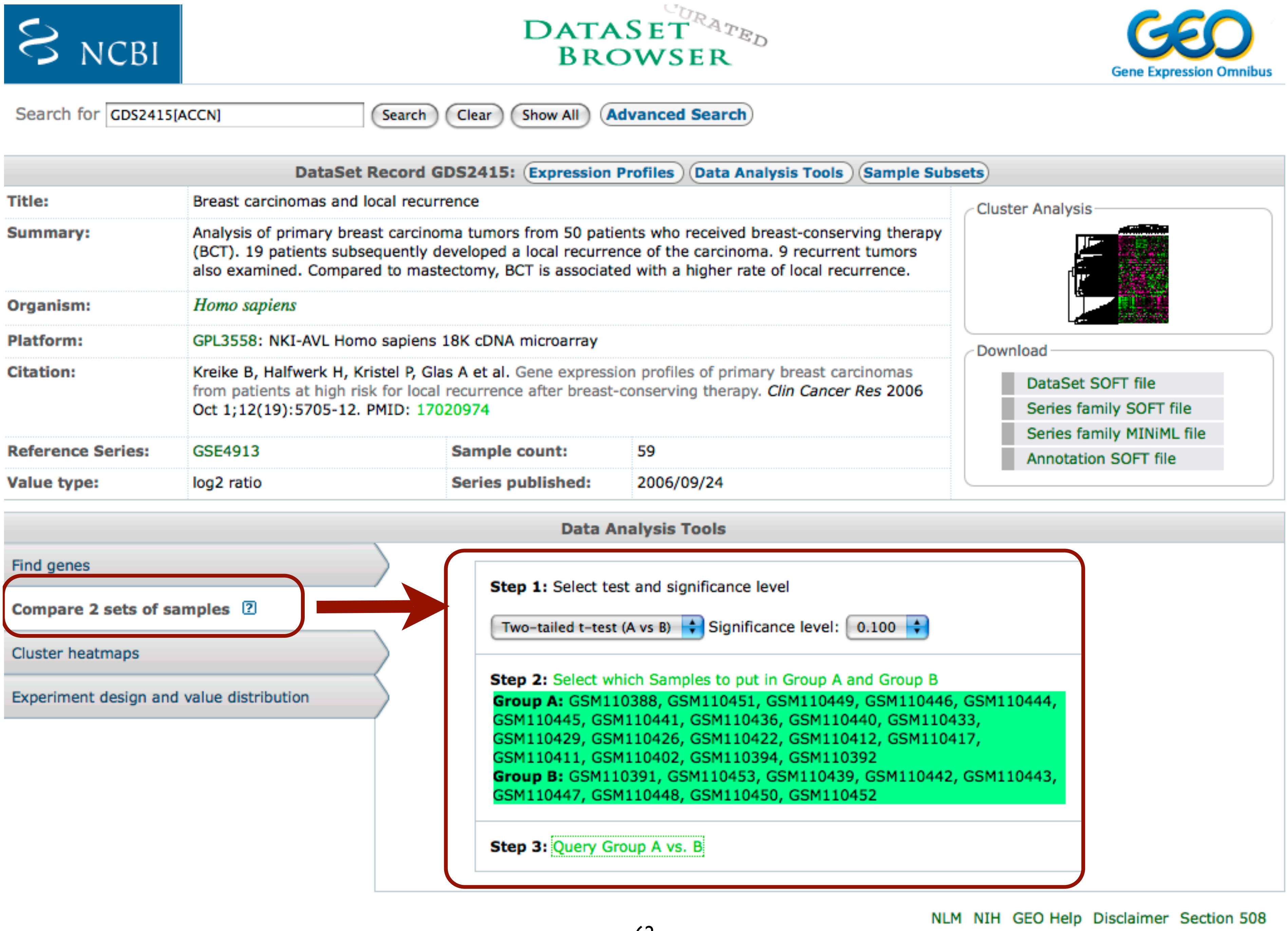
Data Analysis Tools

Find gene name or symbol: brca1 Go

Find genes that are up/down for this condition(s): ☑ disease state ☑ specimen Go



- thumbnail image represents the abundance profile for an individual gene across each Sample in a DataSet
- bars at the bottom of the chart represent experimental subsets within the DataSet.
- Red bar: measured level of abundance
- Blue square: indication of where the expression of that gene falls with respect to all other genes on that array



Search for GDS2415[ACCN]

DataSet Record GDS2415:

Title:	Breast carcinomas and local recurrence		
Summary:	Analysis of primary breast carcinoma tumors from 50 patients who received breast-conserving therapy (BCT). 19 patients subsequently developed a local recurrence of the carcinoma. 9 recurrent tumors also examined. Compared to mastectomy, BCT is associated with a higher rate of local recurrence.		
Organism:	*Homo sapiens*		
Platform:	GPL3558: NKI-AVL Homo sapiens 18K cDNA microarray		
Citation:	Kreike B, Halfwerk H, Kristel P, Glas A et al. Gene expression profiles of primary breast carcinomas from patients at high risk for local recurrence after breast-conserving therapy. *Clin Cancer Res* 2006 Oct 1;12(19):5705-12. PMID: 17020974		
Reference Series:	GSE4913	**Sample count:**	59
Value type:	log2 ratio	**Series published:**	2006/09/24

Data Analysis Tools

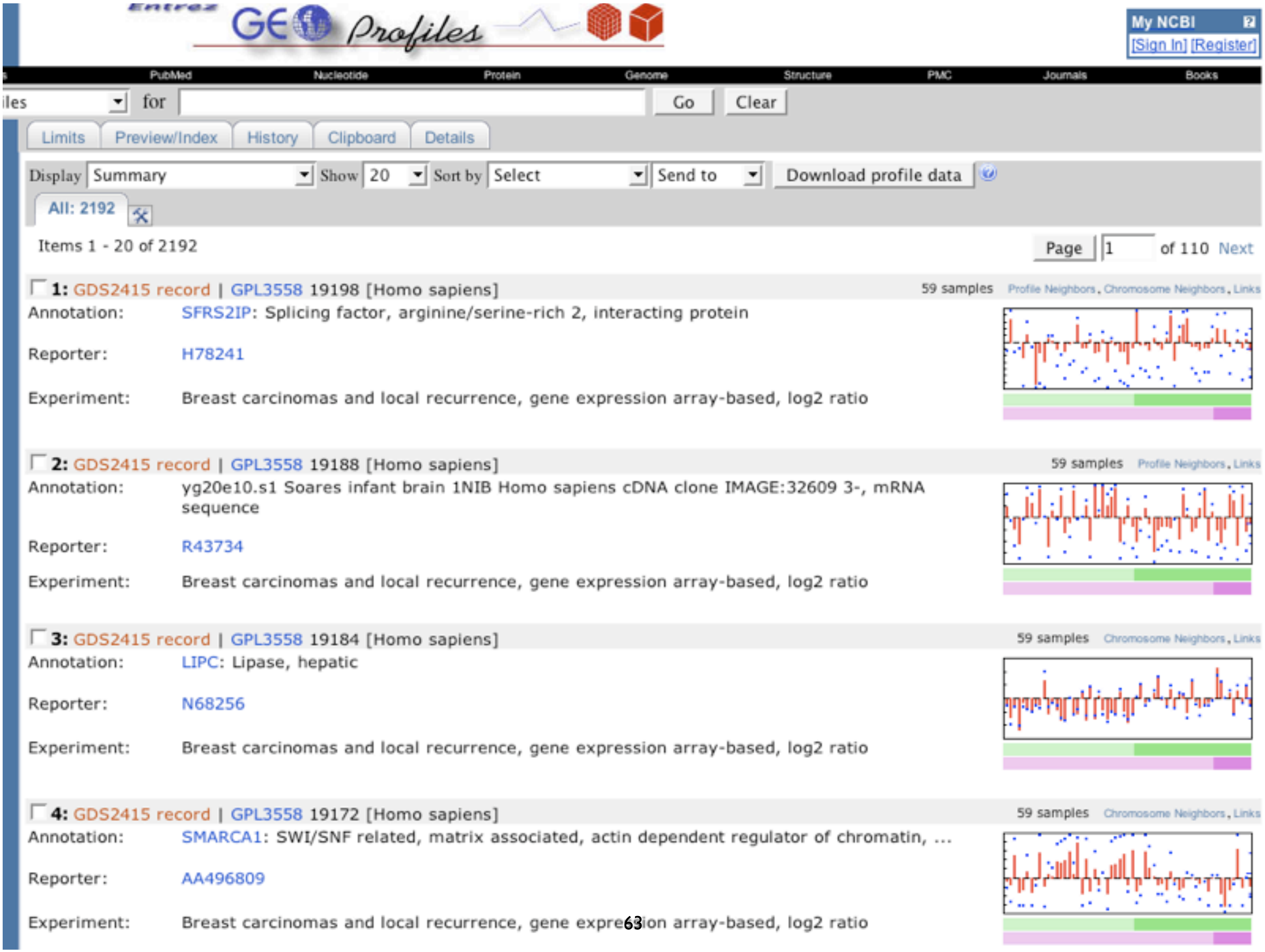
Step 1: Select test and significance level

Two-tailed t-test (A vs B) Significance level: 0.100

Step 2: Select which Samples to put in Group A and Group B

Group A: GSM110388, GSM110451, GSM110449, GSM110446, GSM110444, GSM110445, GSM110441, GSM110436, GSM110440, GSM110433, GSM110429, GSM110426, GSM110422, GSM110412, GSM110417, GSM110411, GSM110402, GSM110394, GSM110392
Group B: GSM110391, GSM110453, GSM110439, GSM110442, GSM110443, GSM110447, GSM110448, GSM110450, GSM110452

Step 3: Query Group A vs. B



Items 1 - 20 of 2192

☐ **1:** GDS2415 record | GPL3558 19198 [Homo sapiens] 59 samples Profile Neighbors, Chromosome Neighbors, Links

Annotation: SFRS2IP: Splicing factor, arginine/serine-rich 2, interacting protein

Reporter: H78241

Experiment: Breast carcinomas and local recurrence, gene expression array-based, log2 ratio

☐ **2:** GDS2415 record | GPL3558 19188 [Homo sapiens] 59 samples Profile Neighbors, Links

Annotation: yg20e10.s1 Soares infant brain 1NIB Homo sapiens cDNA clone IMAGE:32609 3-, mRNA sequence

Reporter: R43734

Experiment: Breast carcinomas and local recurrence, gene expression array-based, log2 ratio

☐ **3:** GDS2415 record | GPL3558 19184 [Homo sapiens] 59 samples Chromosome Neighbors, Links

Annotation: LIPC: Lipase, hepatic

Reporter: N68256

Experiment: Breast carcinomas and local recurrence, gene expression array-based, log2 ratio

☐ **4:** GDS2415 record | GPL3558 19172 [Homo sapiens] 59 samples Chromosome Neighbors, Links

Annotation: SMARCA1: SWI/SNF related, matrix associated, actin dependent regulator of chromatin, ...

Reporter: AA496809

Experiment: Breast carcinomas and local recurrence, gene expression array-based, log2 ratio

A vs B Query Tool

Take home message: GEO data analysis tools are great for quick identification of interesting leads; you download the data to carry out more robust statistical analyses

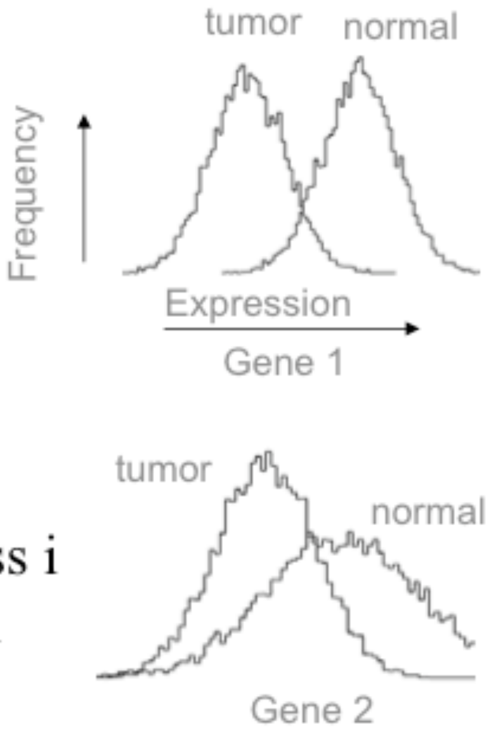
- **Purpose:** To help identify gene profiles that display marked differences in expression level between two subsets of experimental factors (e.g. tissue, strain, time, dose, etc).
- **Caveats:** The "mean group A vs B" is perhaps the most rudimentary means of filtering data; t-test is well established but comes with a set of basic assumptions.

A Simple Test

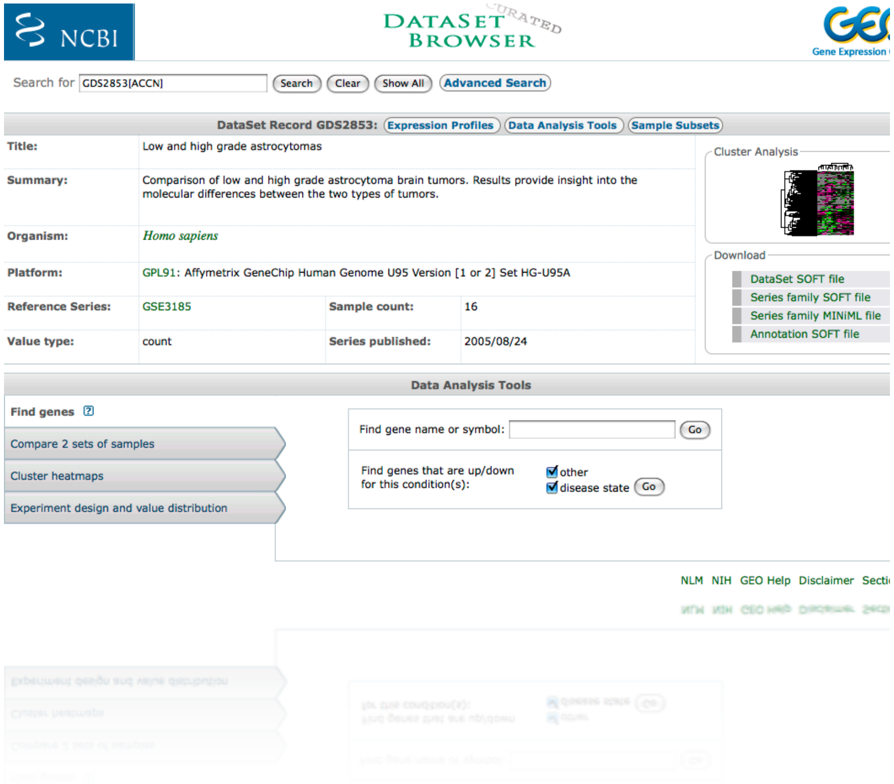
- Student’s t-test
  - Assumptions: Normality, equal variance

$$t = \frac{m_{tumor} - m_{normal}}{\sqrt{\frac{s_{tumor}^2}{N_{tumor}} + \frac{s_{normal}^2}{N_{normal}}}}$$

m_i = mean expression value in class i
N_i = number of examples in class i
s_i^2 = variance



Using GEO for differential expression



DataSet Record GDS2853:

Title: Low and high grade astrocytomas

Summary: Comparison of low and high grade astrocytoma brain tumors. Results provide insight into the molecular differences between the two types of tumors.

Organism: *Homo sapiens*

Platform: GPL91: Affymetrix GeneChip Human Genome U95 Version [1 or 2] Set HG-U95A

Reference Series: GSE3185 **Sample count:** 16

Value type: count **Series published:** 2005/08/24

Data Analysis Tools

Find gene name or symbol:

Find genes that are up/down for this condition(s): ☑ other ☑ disease state

http://www.ncbi.nlm.nih.gov/sites/GDSbrowser?acc=GDS2853

GEO limitations

- Differential expression can only be done for “Datasets”(GDS****)
- T-tests only
- Very little control over parameters
- Output is not that easy to use

Be careful with $p < 0.05$

- In GDS2853 example, 2912 genes met $p<0.05$
- This is 11% of the genes on the array
- Expect 5% by chance (this is what $p < 0.05$ means)
- Probably ~1/2 of the selected 2912 are false positives
- This is the “multiple testing” problem

Download Data Types

SOFT - text based
MiNiML* - xml based

tumors. Results provide insight into the
rs.

rsion [1 or 2] Set HG-U95A

16

2005/08/24

Contains DataSet information, experiment variable subsets and expression value measurements (plain text, tab-delimited format).

*MIAME Notation in Markup Language

all GEO data are available for bulk download:
ftp://ftp.ncbi.nih.gov/pub/geo/DATA

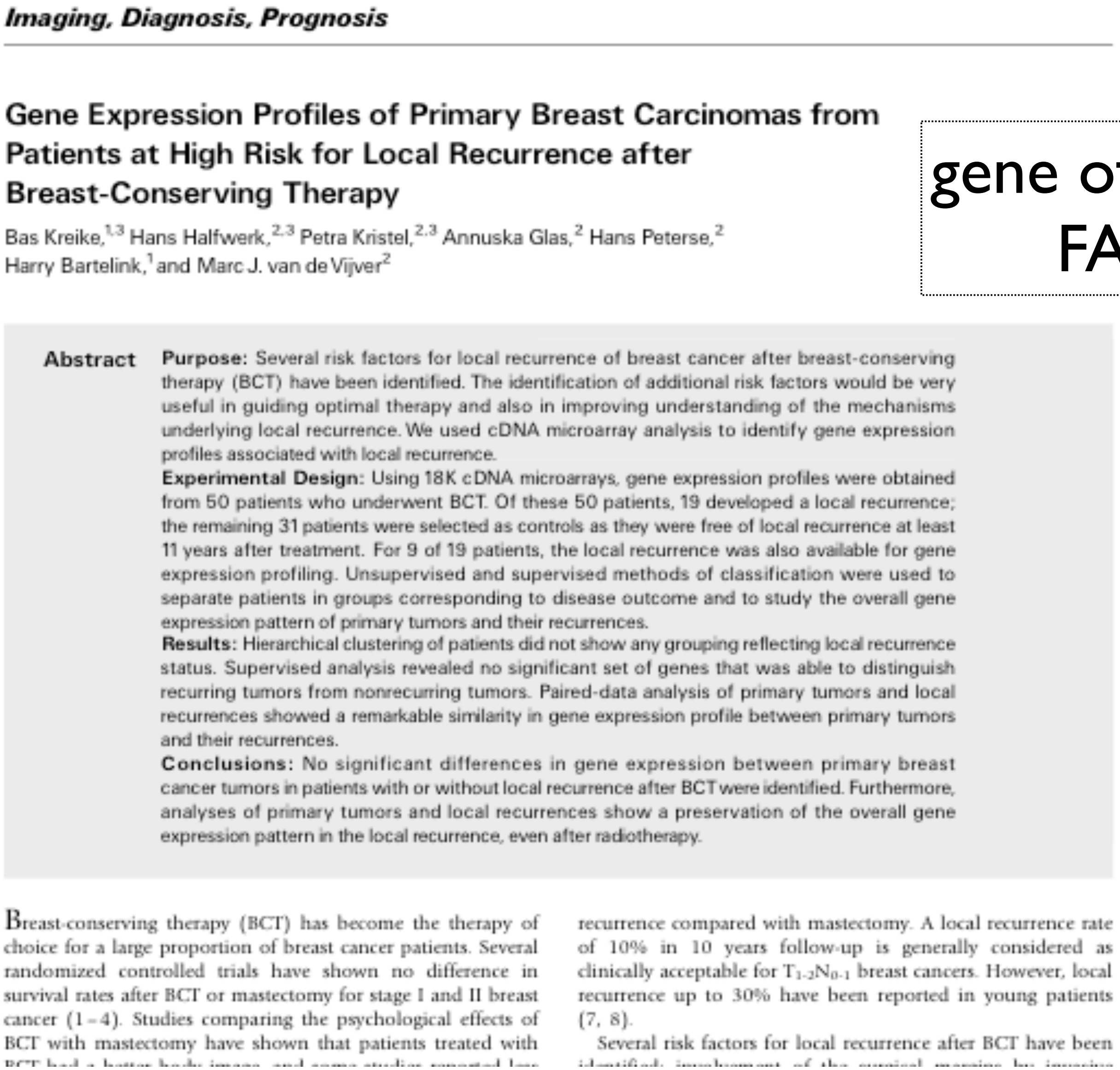
More Serious Tools

- Free
  - R + Bioconductor
  - TIGR MultiExperimentViewer (MeV)

  ...

- Commerical
  - Genespring–ArrayAssist
  - Rosetta Resolver

  ...



Imaging, Diagnosis, Prognosis

Gene Expression Profiles of Primary Breast Carcinomas from Patients at High Risk for Local Recurrence after Breast-Conserving Therapy

Bas Kreike,[1,3] Hans Halfwerk,[2,3] Petra Kristel,[2,3] Annuska Glas,[2] Hans Peterse,[2]
Harry Bartelink,[1] and Marc J. van de Vijver[2]

Abstract **Purpose:** Several risk factors for local recurrence of breast cancer after breast-conserving therapy (BCT) have been identified. The identification of additional risk factors would be very useful in guiding optimal therapy and also in improving understanding of the mechanisms underlying local recurrence. We used cDNA microarray analysis to identify gene expression profiles associated with local recurrence.

Experimental Design: Using 18K cDNA microarrays, gene expression profiles were obtained from 50 patients who underwent BCT. Of these 50 patients, 19 developed a local recurrence; the remaining 31 patients were selected as controls as they were free of local recurrence at least 11 years after treatment. For 9 of 19 patients, the local recurrence was also available for gene expression profiling. Unsupervised and supervised methods of classification were used to separate patients in groups corresponding to disease outcome and to study the overall gene expression pattern of primary tumors and their recurrences.

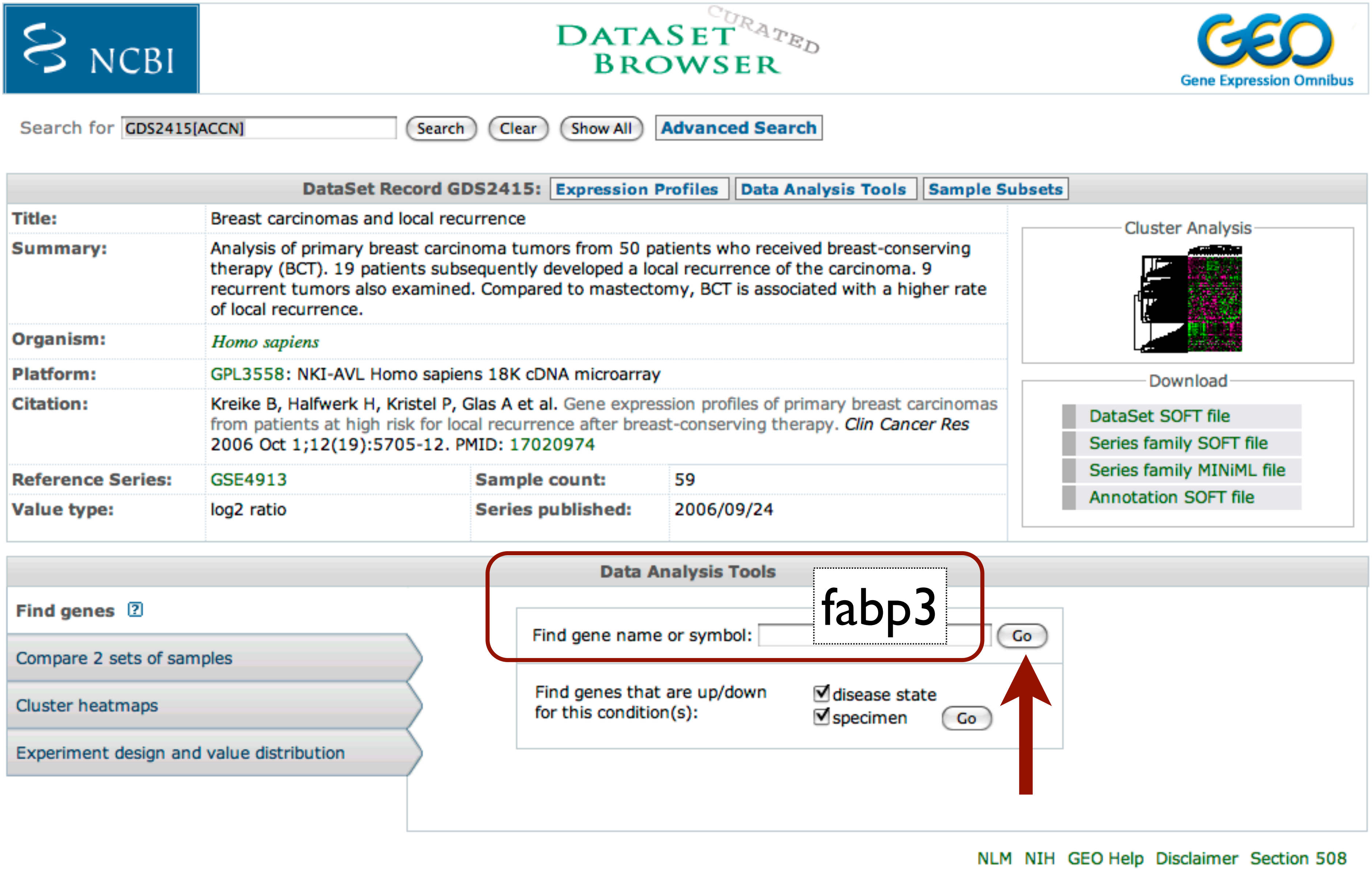
Results: Hierarchical clustering of patients did not show any grouping reflecting local recurrence status. Supervised analysis revealed no significant set of genes that was able to distinguish recurring tumors from nonrecurring tumors. Paired-data analysis of primary tumors and local recurrences showed a remarkable similarity in gene expression profile between primary tumors and their recurrences.

Conclusions: No significant differences in gene expression between primary breast cancer tumors in patients with or without local recurrence after BCT were identified. Furthermore, analyses of primary tumors and local recurrences show a preservation of the overall gene expression pattern in the local recurrence, even after radiotherapy.

Breast-conserving therapy (BCT) has become the therapy of choice for a large proportion of breast cancer patients. Several randomized controlled trials have shown no difference in survival rates after BCT or mastectomy for stage I and II breast cancer (1–4). Studies comparing the psychological effects of BCT with mastectomy have shown that patients treated with recurrence compared with mastectomy. A local recurrence rate of 10% in 10 years follow-up is generally considered as clinically acceptable for $T_{1\text{-}2}N_{0\text{-}1}$ breast cancers. However, local recurrence up to 30% have been reported in young patients (7, 8).

Several risk factors for local recurrence after BCT have been

gene of interest
FABP3



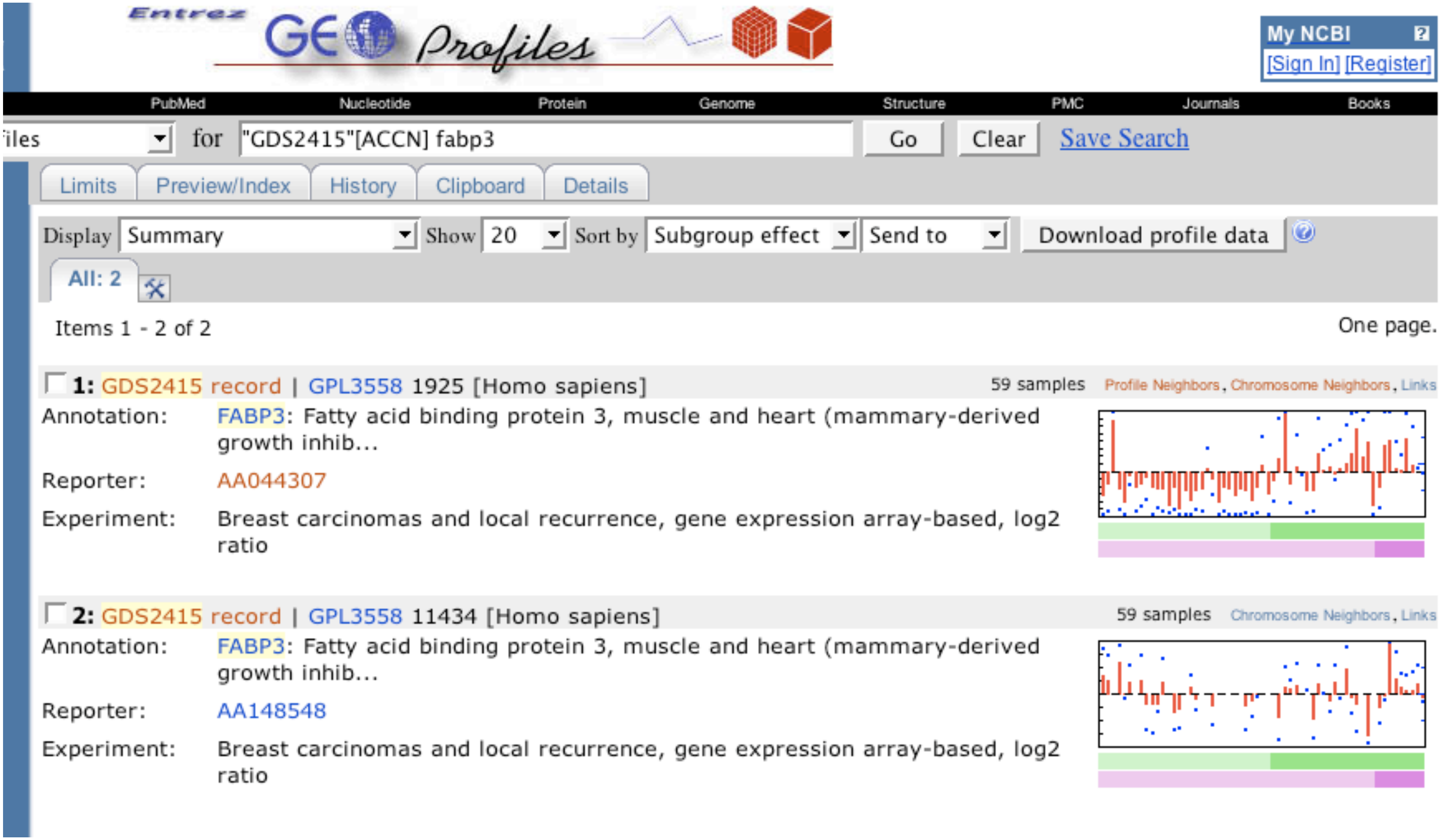
Search for: GDS2415[ACCN]

DataSet Record GDS2415:

Title:	Breast carcinomas and local recurrence		
Summary:	Analysis of primary breast carcinoma tumors from 50 patients who received breast-conserving therapy (BCT). 19 patients subsequently developed a local recurrence of the carcinoma. 9 recurrent tumors also examined. Compared to mastectomy, BCT is associated with a higher rate of local recurrence.		
Organism:	*Homo sapiens*		
Platform:	GPL3558: NKI-AVL Homo sapiens 18K cDNA microarray		
Citation:	Kreike B, Halfwerk H, Kristel P, Glas A et al. Gene expression profiles of primary breast carcinomas from patients at high risk for local recurrence after breast-conserving therapy. *Clin Cancer Res* 2006 Oct 1;12(19):5705-12. PMID: 17020974		
Reference Series:	GSE4913	**Sample count:**	59
Value type:	log2 ratio	**Series published:**	2006/09/24

Find gene name or symbol: fabp3

Find genes that are up/down for this condition(s): ☑ disease state ☑ specimen



iles for "GDS2415"[ACCN] fabp3

Display Summary Show 20 Sort by Subgroup effect Send to

All: 2

Items 1 - 2 of 2

One page.

☐ **1:** GDS2415 record | GPL3558 1925 [Homo sapiens] 59 samples Profile Neighbors, Chromosome Neighbors, Links

Annotation: FABP3: Fatty acid binding protein 3, muscle and heart (mammary-derived growth inhib...

Reporter: AA044307

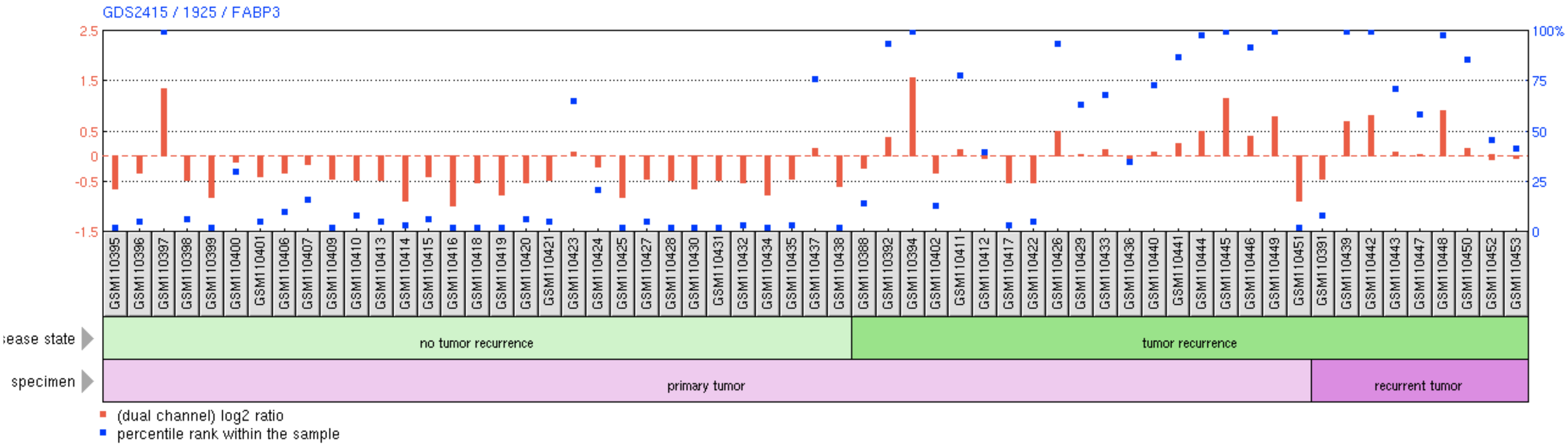
Experiment: Breast carcinomas and local recurrence, gene expression array-based, log2 ratio

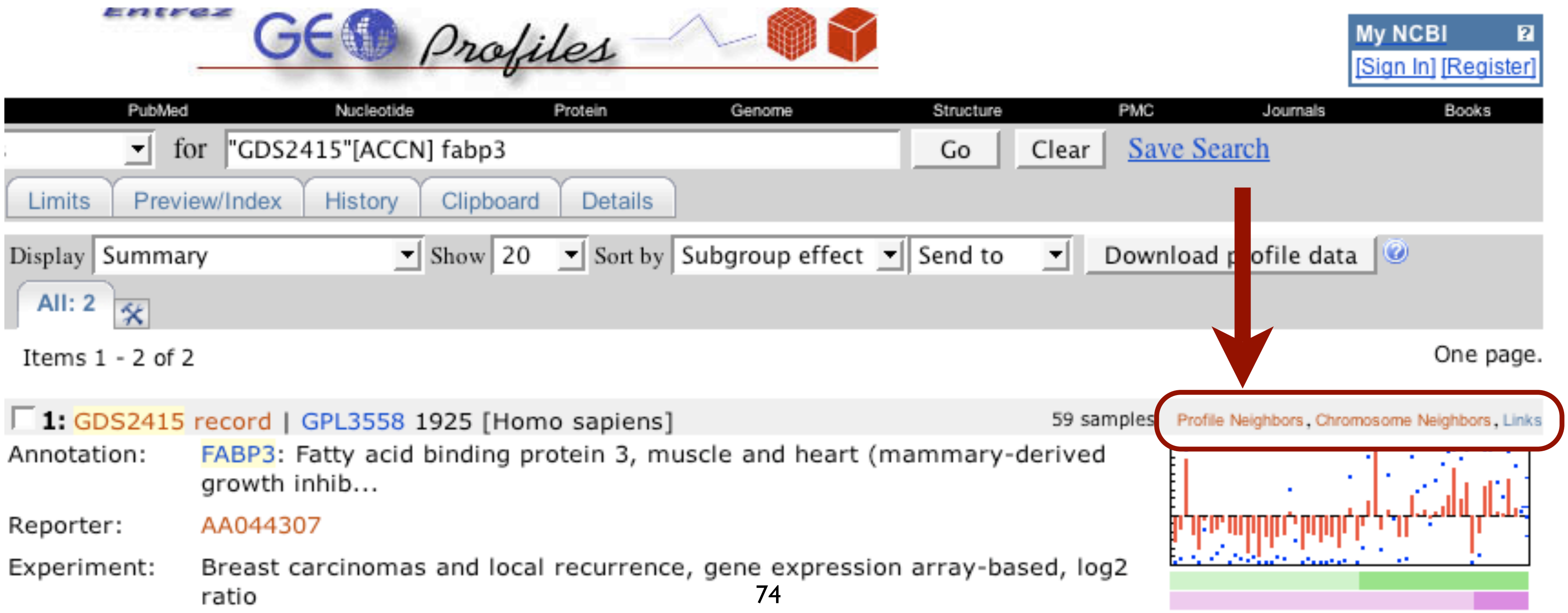
☐ **2:** GDS2415 record | GPL3558 11434 [Homo sapiens] 59 samples Chromosome Neighbors, Links

Annotation: FABP3: Fatty acid binding protein 3, muscle and heart (mammary-derived growth inhib...

Reporter: AA148548

Experiment: Breast carcinomas and local recurrence, gene expression array-based, log2 ratio





for "GDS2415"[ACCN] fabp3

Display Summary Show 20 Sort by Subgroup effect

Items 1 - 2 of 2

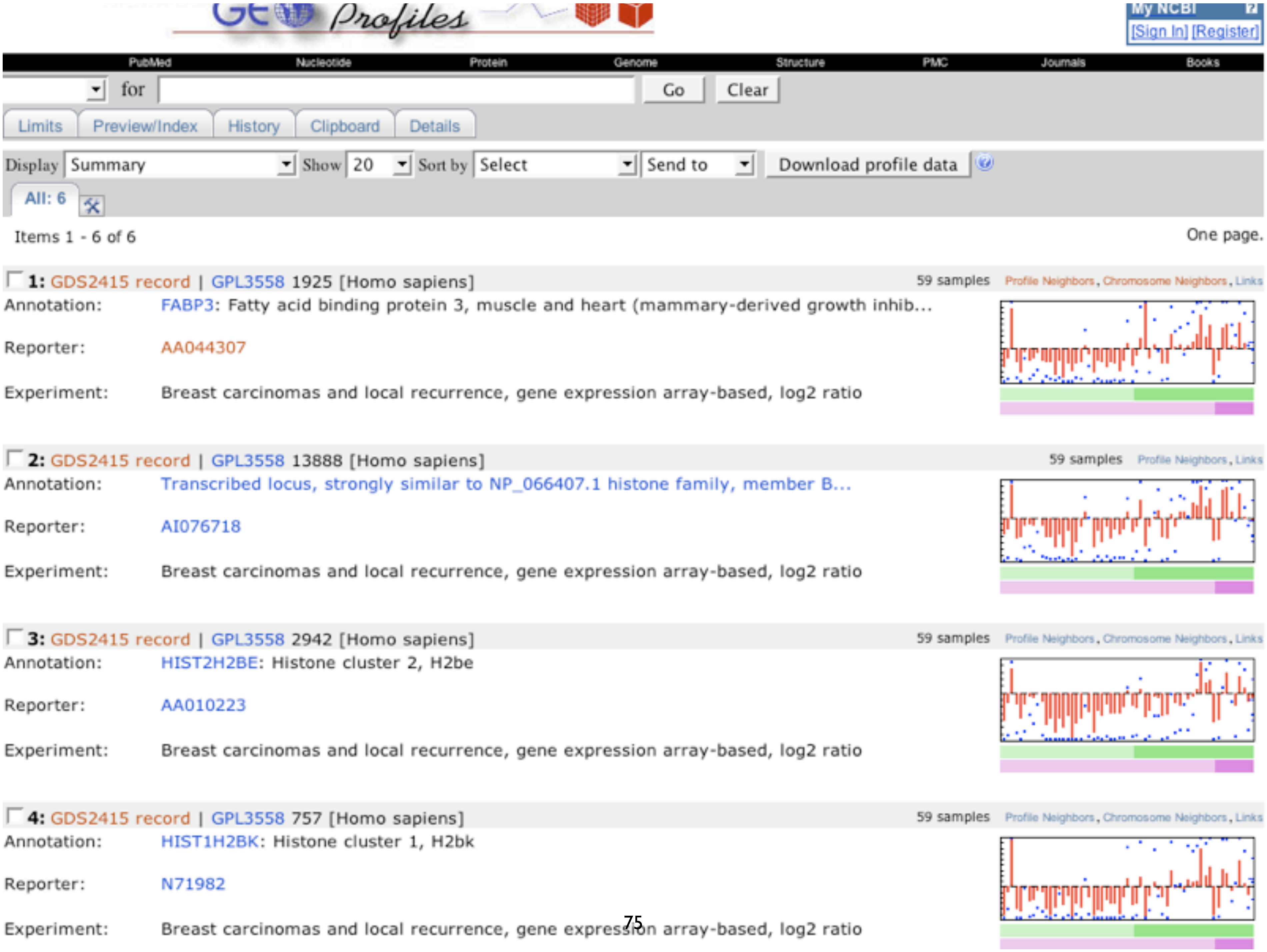
One page.

☐ **1:** GDS2415 record | GPL3558 1925 [Homo sapiens] 59 samples Profile Neighbors, Chromosome Neighbors, Links

Annotation: FABP3: Fatty acid binding protein 3, muscle and heart (mammary-derived growth inhib...

Reporter: AA044307

Experiment: Breast carcinomas and local recurrence, gene expression array-based, log2 ratio



Items 1 - 6 of 6

One page.

☐ **1:** GDS2415 record | GPL3558 1925 [Homo sapiens] 59 samples Profile Neighbors, Chromosome Neighbors, Links

Annotation: FABP3: Fatty acid binding protein 3, muscle and heart (mammary-derived growth inhib...

Reporter: AA044307

Experiment: Breast carcinomas and local recurrence, gene expression array-based, log2 ratio

☐ **2:** GDS2415 record | GPL3558 13888 [Homo sapiens] 59 samples Profile Neighbors, Links

Annotation: Transcribed locus, strongly similar to NP_066407.1 histone family, member B...

Reporter: AI076718

Experiment: Breast carcinomas and local recurrence, gene expression array-based, log2 ratio

☐ **3:** GDS2415 record | GPL3558 2942 [Homo sapiens] 59 samples Profile Neighbors, Chromosome Neighbors, Links

Annotation: HIST2H2BE: Histone cluster 2, H2be

Reporter: AA010223

Experiment: Breast carcinomas and local recurrence, gene expression array-based, log2 ratio

☐ **4:** GDS2415 record | GPL3558 757 [Homo sapiens] 59 samples Profile Neighbors, Chromosome Neighbors, Links

Annotation: HIST1H2BK: Histone cluster 1, H2bk

Reporter: N71982

Experiment: Breast carcinomas and local recurrence, gene expression array-based, log2 ratio

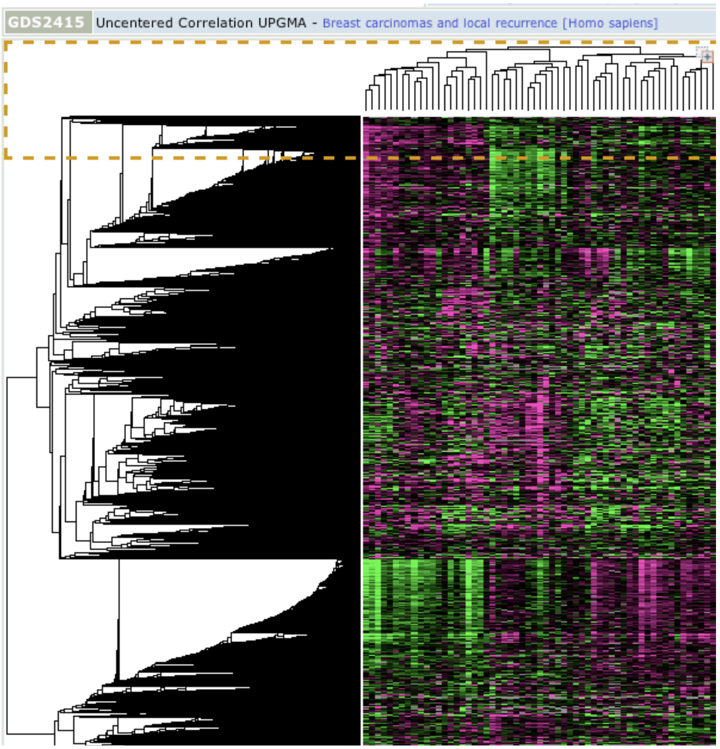
Profile Neighbors

Take home message: GEO data analysis tools are great for quick identification of interesting leads; you download the data to carry out more robust statistical analyses

- Connects groups of genes that have similar expression profiles within a DataSet
- pre-computed
- calculated by Pearson correlation coefficients

Other Features

- Cluster Heat Maps
  - precomputed sample and gene hierarchical cluster heat maps provided
  - different methods available; can select, expand, download



- GEO BLAST
  - retrieve gene expression profiles by sequence similarity

GEO, the gene expression omnibus

- public repository of expression data from many different experimental platforms
- Main uses
  - ✓ search for experiments of interest
  - ✓ search for expression information about gene of interest
- submit, search, analyses tools available
- data standards required MIAME, MiNiML

Credits & References

- NCBI GEO: mining tens of millions of expression profiles–database and tools update. Barrett T, et al. Nucleic Acids Res. 35 (2007) D760-5. [PMID: 17099226]
- GEO: the Gene Expression Omnibus

  http://www.ncbi.nlm.nih.gov/projects/geo/info/GEOHandoutFinal.pdf
- Dr. Paul Pavlidis, UBC Bioinformatics Centre

Bioinformatics

Session 3.2 - Pathway Resources for Systems Biology

Proteomics

- How large is the human proteome, anyway?

Class	Size	Description
Non Redundant Proteins	20,000-25,000	representative protein from every gene locus
Variants	50,000-500,000	different proteins obtained by splicing or proteolysis
Combinatorial Variants	>10,000,000	different proteins generated by somatic DNA rearrangements
Protein Species	>100,000	proteins that differ in chemical composition due to PTM
Protein Alleles	75,000-150,000	proteins that differ by genetic variation (coding SNPs)

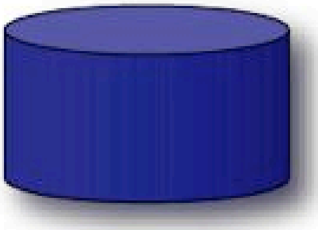
source: Uhlen M et al, Molecular & Cellular Proteomics (2005)

Cellular Pathways

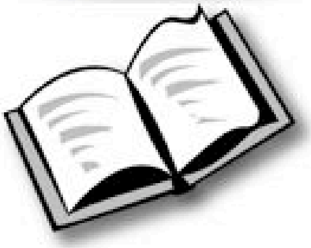
- A striking similarity between intracellular signaling pathways and the Tokyo subway system

Pathway Information

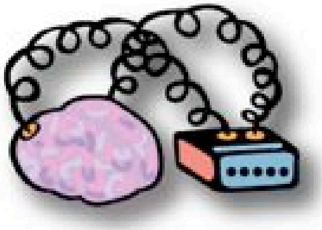
- Databases
  - Fully electronic
  - Easily computer readable



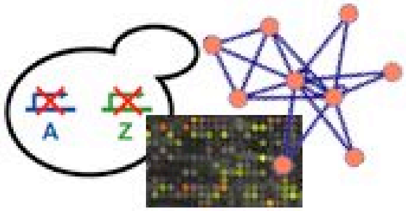
- Literature
  - Increasingly electronic
  - Human readable



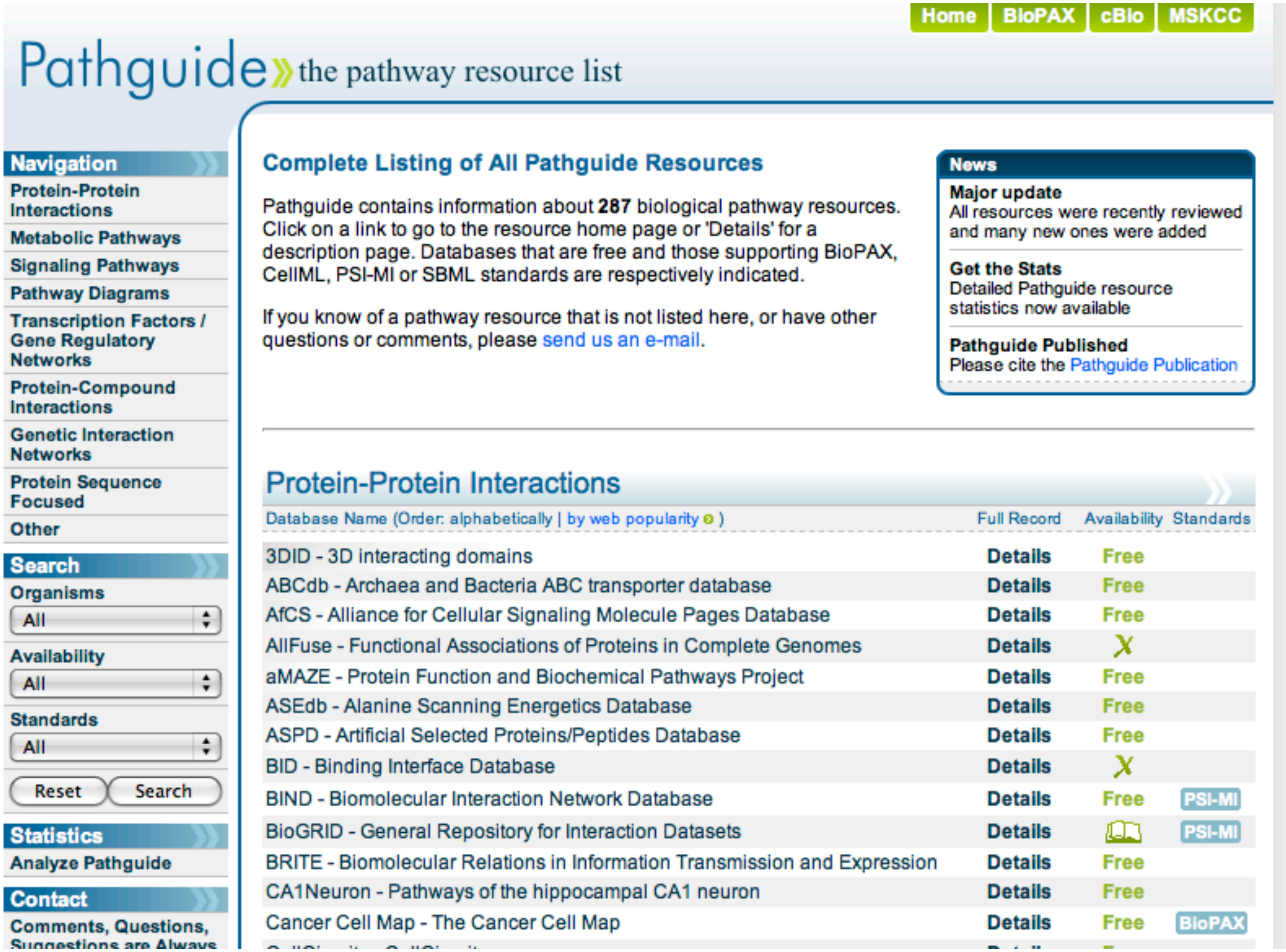
- Biologist's brains
  - Richest data source
  - Limited bandwidth access



- Experiments
  - Basis for models



http://www.pathguide.org/



Complete Listing of All Pathguide Resources

Pathguide contains information about **287** biological pathway resources. Click on a link to go to the resource home page or 'Details' for a description page. Databases that are free and those supporting BioPAX, CellML, PSI-MI or SBML standards are respectively indicated.

If you know of a pathway resource that is not listed here, or have other questions or comments, please send us an e-mail.

News

Major update
All resources were recently reviewed and many new ones were added

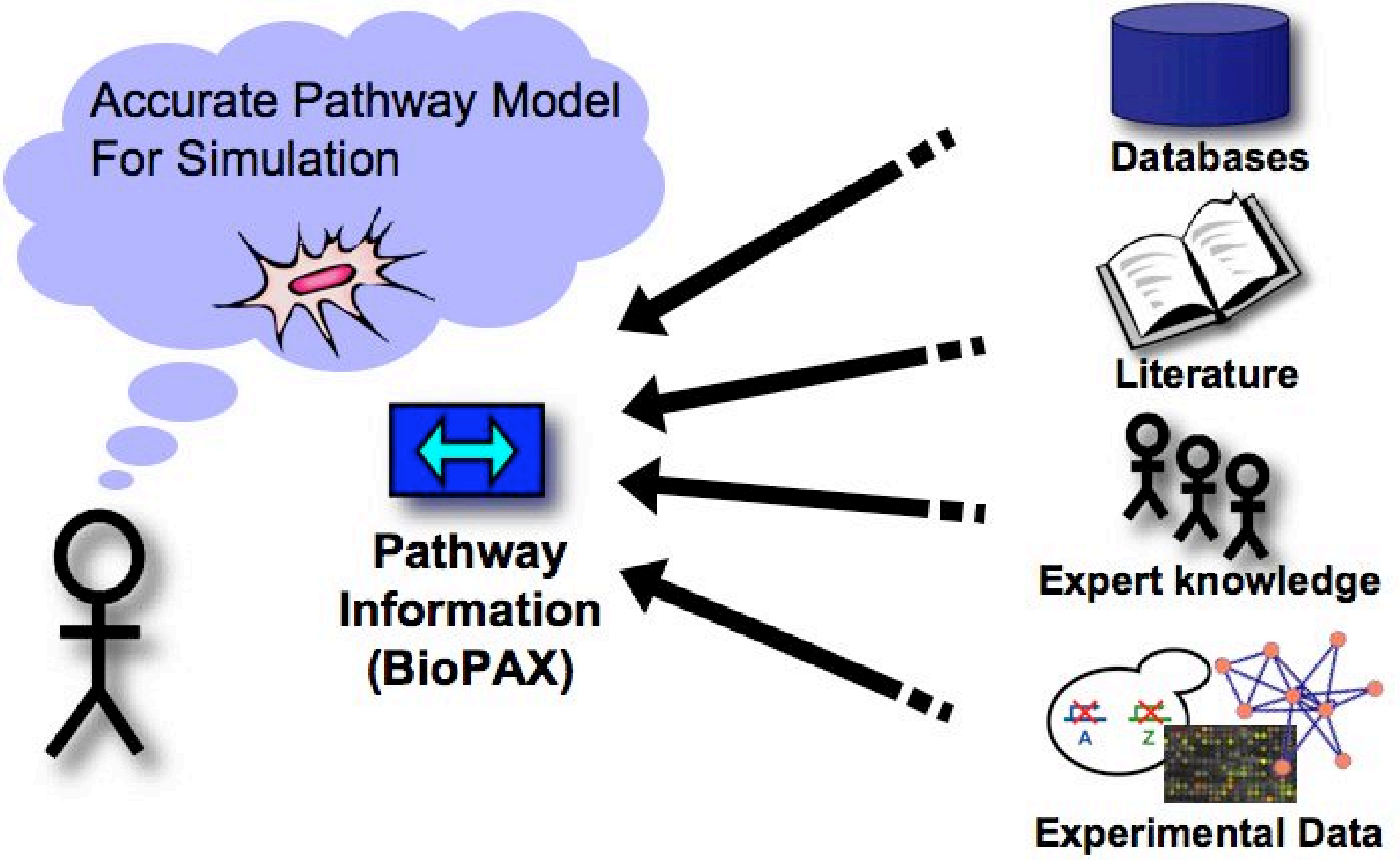
Get the Stats
Detailed Pathguide resource statistics now available

Pathguide Published
Please cite the Pathguide Publication

Protein-Protein Interactions

Database Name (Order: alphabetically | by web popularity)	Full Record	Availability	Standards
3DID - 3D interacting domains	Details	Free	
ABCdb - Archaea and Bacteria ABC transporter database	Details	Free	
AfCS - Alliance for Cellular Signaling Molecule Pages Database	Details	Free	
AllFuse - Functional Associations of Proteins in Complete Genomes	Details	X	
aMAZE - Protein Function and Biochemical Pathways Project	Details	Free	
ASEdb - Alanine Scanning Energetics Database	Details	Free	
ASPD - Artificial Selected Proteins/Peptides Database	Details	Free	
BID - Binding Interface Database	Details	X	
BIND - Biomolecular Interaction Network Database	Details	Free	PSI-MI
BioGRID - General Repository for Interaction Datasets	Details	📖	PSI-MI
BRITE - Biomolecular Relations in Information Transmission and Expression	Details	Free	
CA1Neuron - Pathways of the hippocampal CA1 neuron	Details	Free	
Cancer Cell Map - The Cancer Cell Map	Details	Free	BioPAX

Using Pathway Information



Accurate Pathway Model
For Simulation

Pathway Information (BioPAX)

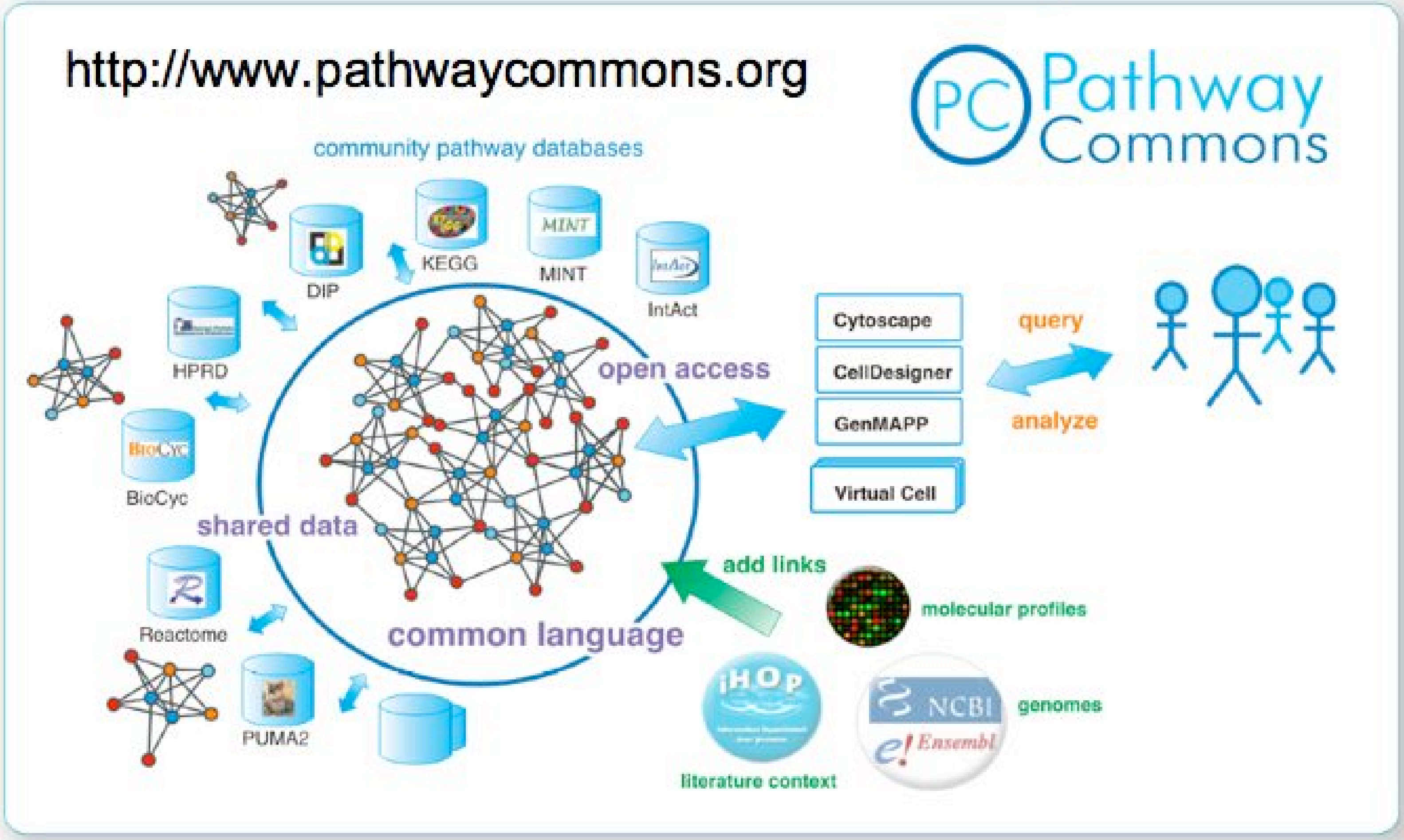
Databases

Literature

Expert knowledge

Experimental Data

Aim: Convenient Access to Pathway Information

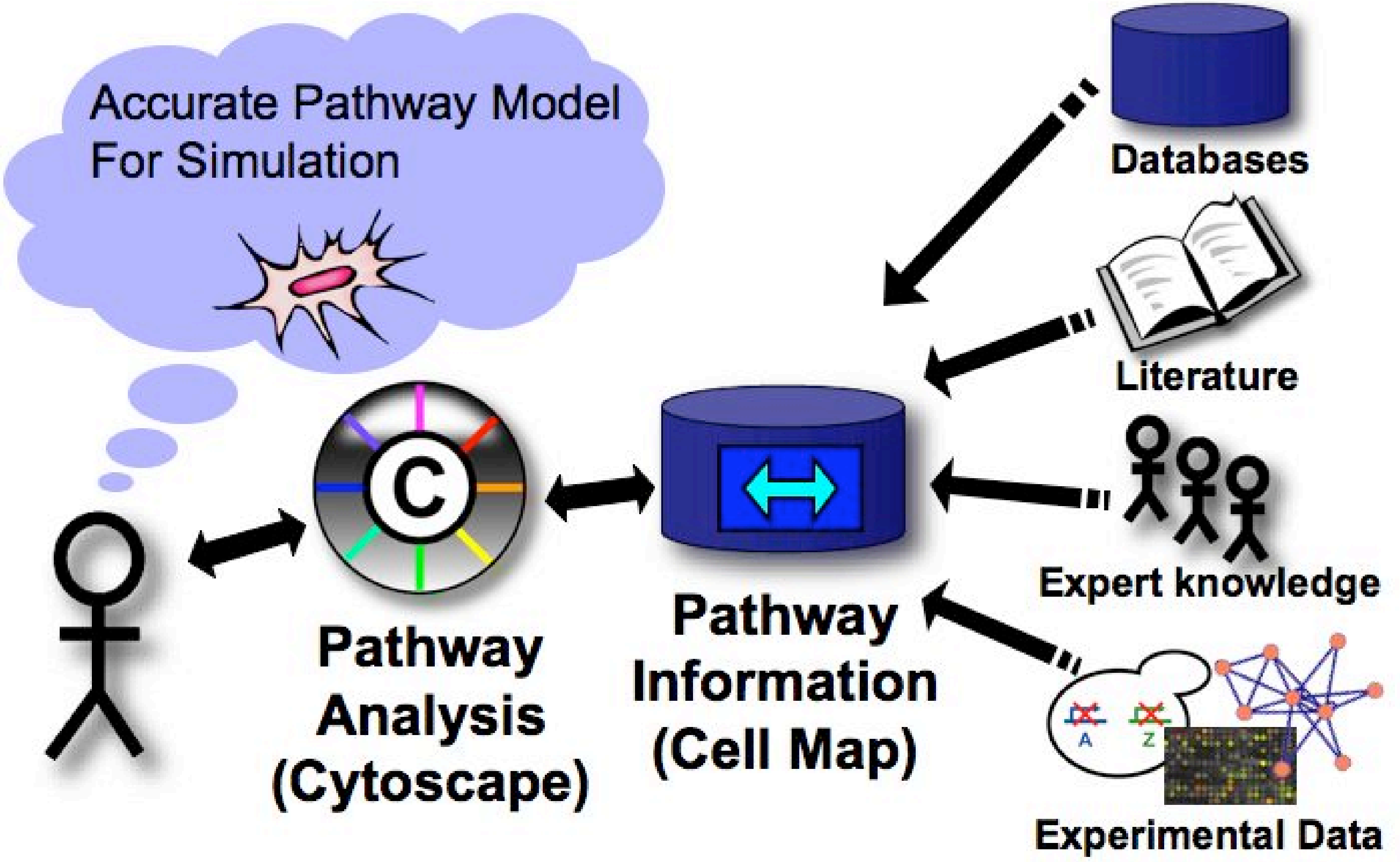


http://www.pathwaycommons.org

Facilitate creation and communication of pathway data
Aggregate pathway data in the public domain
Provide easy access for pathway analysis

Long term: Converge to integrated cell map

Using Pathway Information



Accurate Pathway Model
For Simulation

Pathway Analysis (Cytoscape)

Pathway Information (Cell Map)

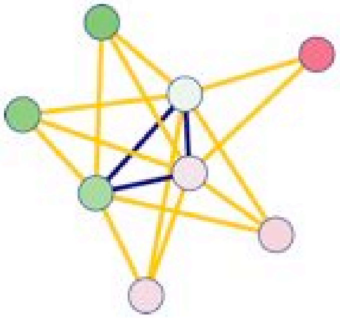
Databases

Literature

Expert knowledge

Experimental Data

Cytoscape - Network Visualization and Analysis

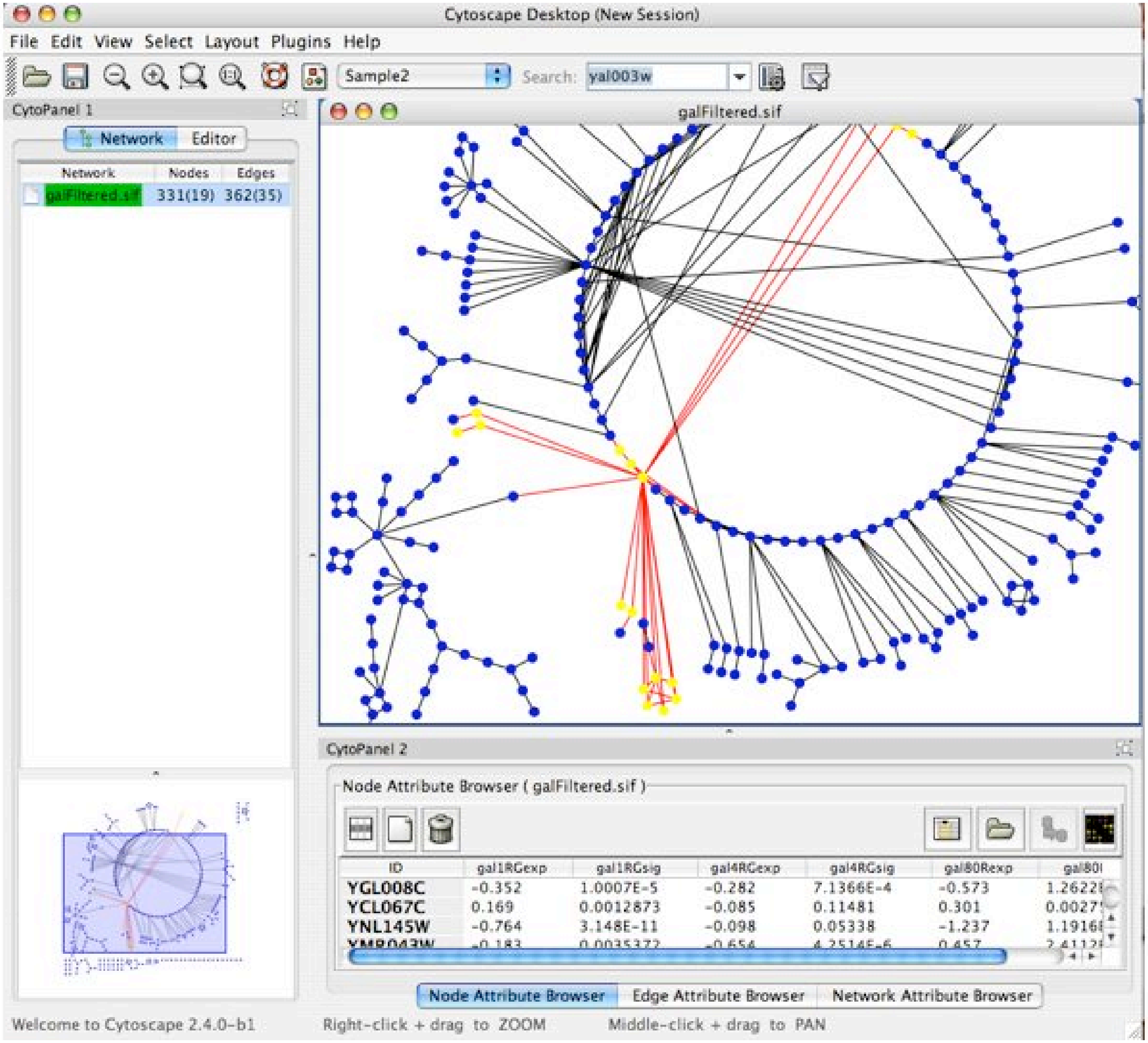


http://cytoscape.org

- Freely-available (open-source, java) software
- Visualizing biological networks (e.g. molecular interaction networks)
- Analyzing networks with gene expression profiles and other cell state data

UCSD, ISB, Agilent, MSKCC, Pasteur, UCSF, UToronto

Other software: Osprey, BioLayout, VisANT, Navigator, PIMWalker, ProViz



Network	Nodes	Edges
galFiltered.sif	331(19)	362(35)

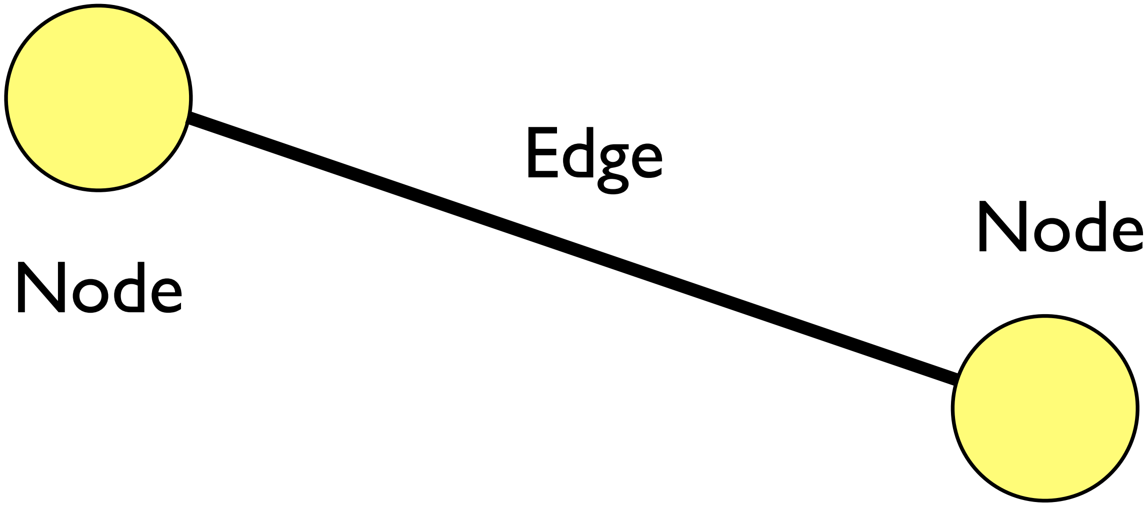
ID	gal1RGexp	gal1RGsig	gal4RGexp	gal4RGsig	gal80Rexp	gal80I
YGL008C	-0.352	1.0007E-5	-0.282	7.1366E-4	-0.573	1.2622
YCL067C	0.169	0.0012873	-0.085	0.11481	0.301	0.0027
YNL145W	-0.764	3.148E-11	-0.098	0.05338	-1.237	1.1916
YMR043W	-0.183	0.0035372	-0.654	4.2514E-6	0.457	2.4112

Welcome to Cytoscape 2.4.0-b1

Right-click + drag to ZOOM

Middle-click + drag to PAN

Pathway Graphs

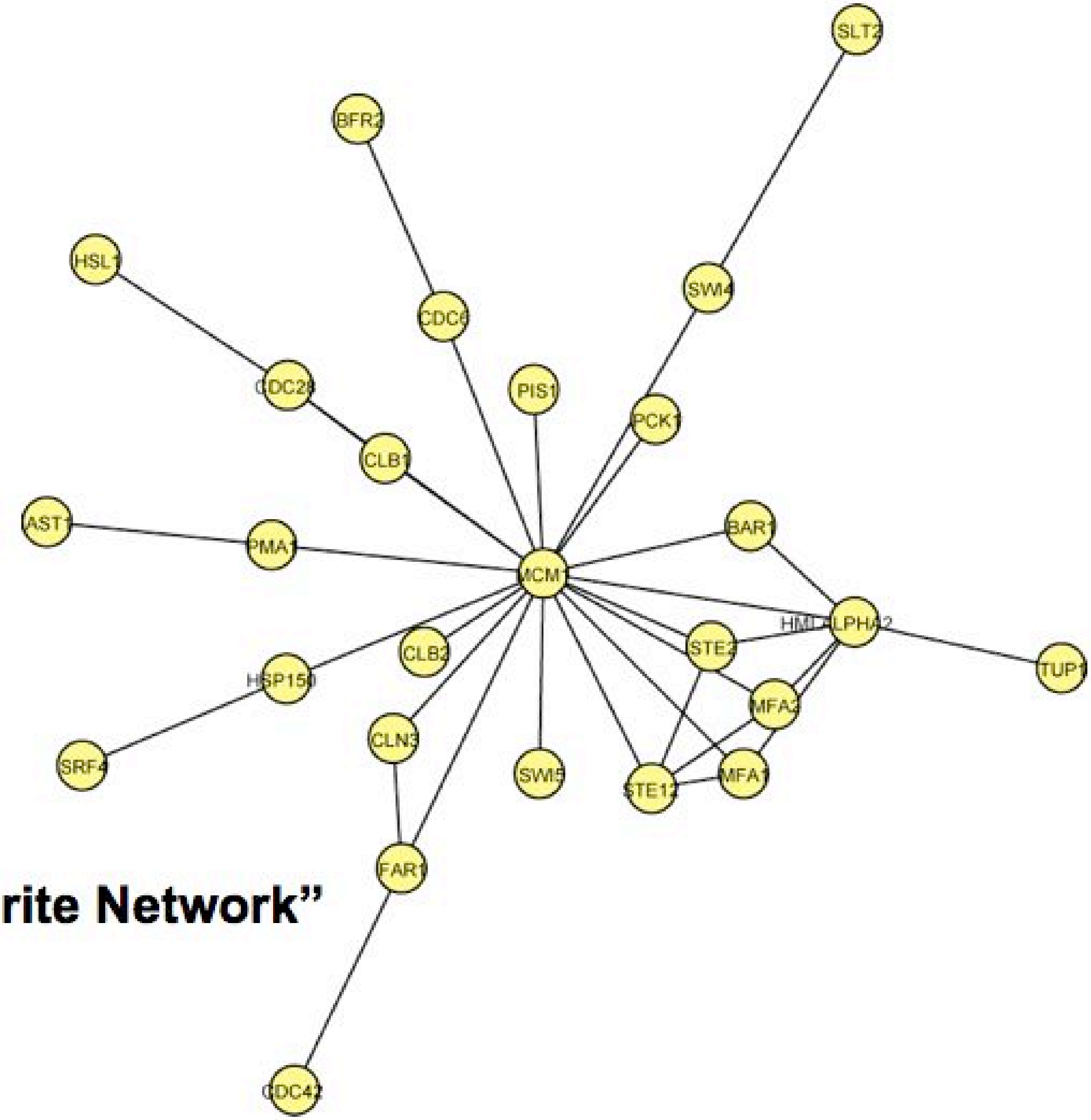


- In addition to describing the network topology, nodes and edges can each have their own attributes

Visual Style

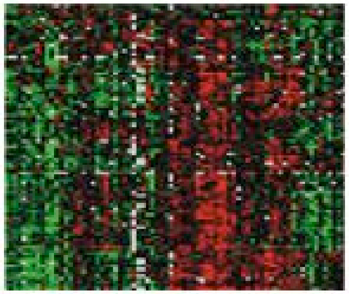
- Customized views of experimental data in a network context
- Network has node and edge attributes
    - E.g. expression data, interaction type, GO function
- Mapped to visual attributes
    - E.g. node/edge size, shape, colour…
- E.g. Visualize gene expression data as node colour gradient on the network

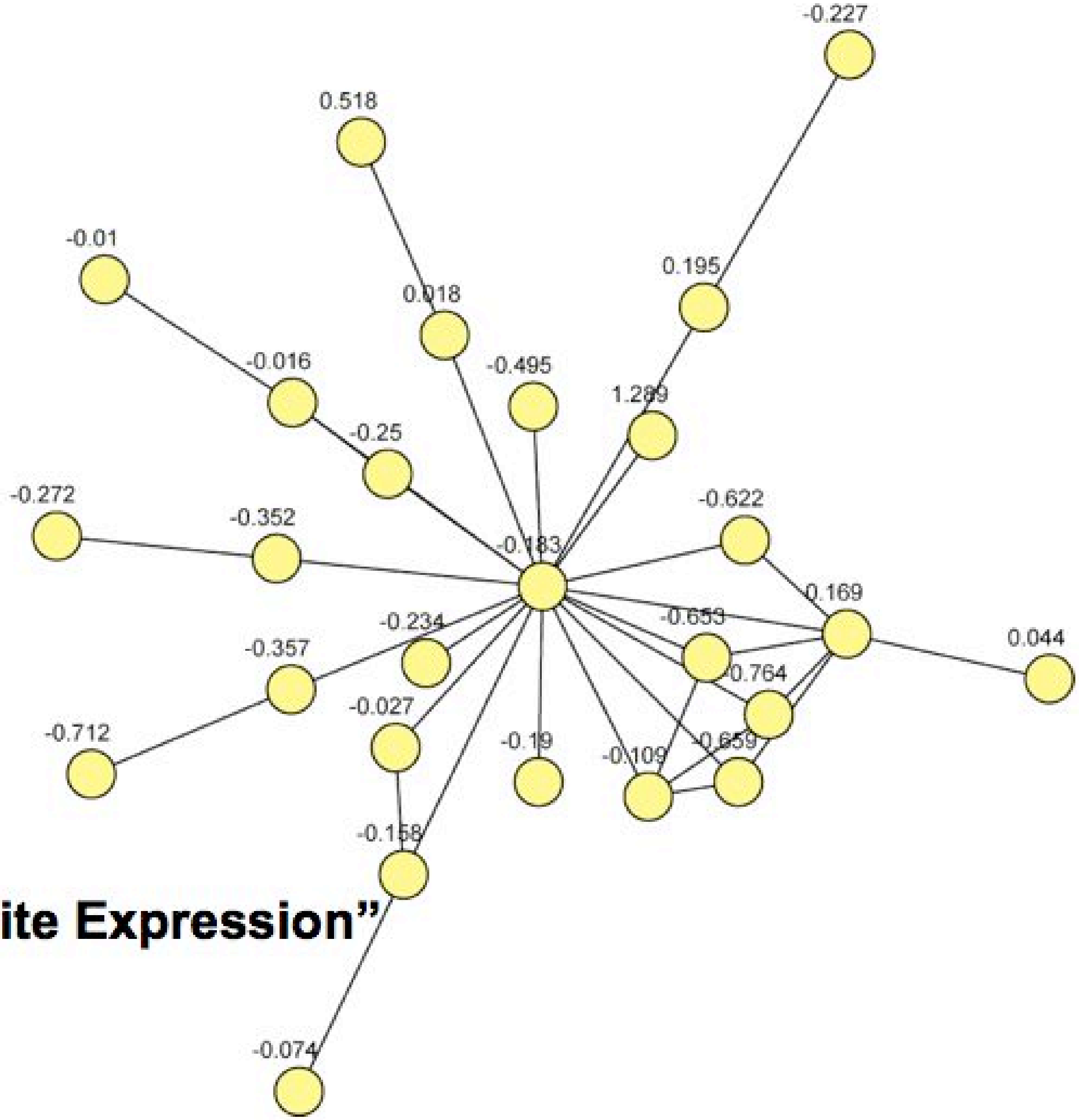
Visual Style



Load “Your Favorite Network”

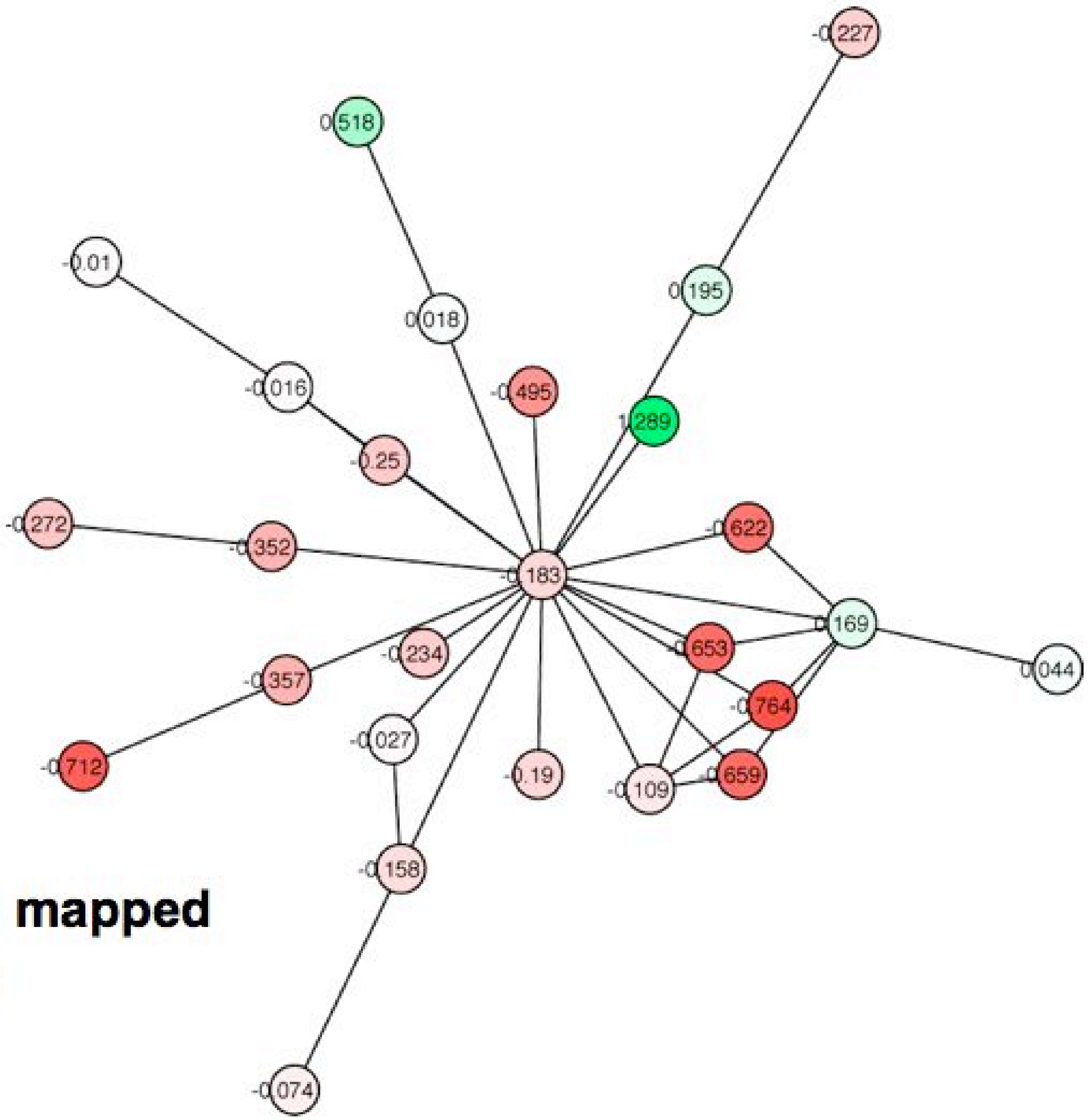
Visual Style





Load “Your Favorite Expression” Dataset

Visual Style



Expression data mapped to node colours

Systems Biology

- Goals:
  - ✓ integrating diverse data types, pathways
  - ✓ cellular simulations
- Community approaches:
  - ✓ pathguide, pathway commons, cytoscape
- Open data exchange key to success

Credits & References

- Dr. Gary Bader, DCCBR, UofT

  slides/images used with permission

- Cary MP, Bader GD, Sander C "Pathway Information for Systems Biology", FEBS Letters (2005)

Bioinformatics Links Directory

Finding online tools & resources for Life Sciences research

Conducting Research on the Web: 2007 Update for the Bioinformatics Links Directory

Joanne A. Fox, Scott McMillan and B. F. Francis Ouellette*

UBC Bioinformatics Centre (http://bioinformatics.ubc.ca/),
Vancouver, British Columbia, Canada, V6T 1Z4

Received May 18, 2007; Accepted May 22, 2007

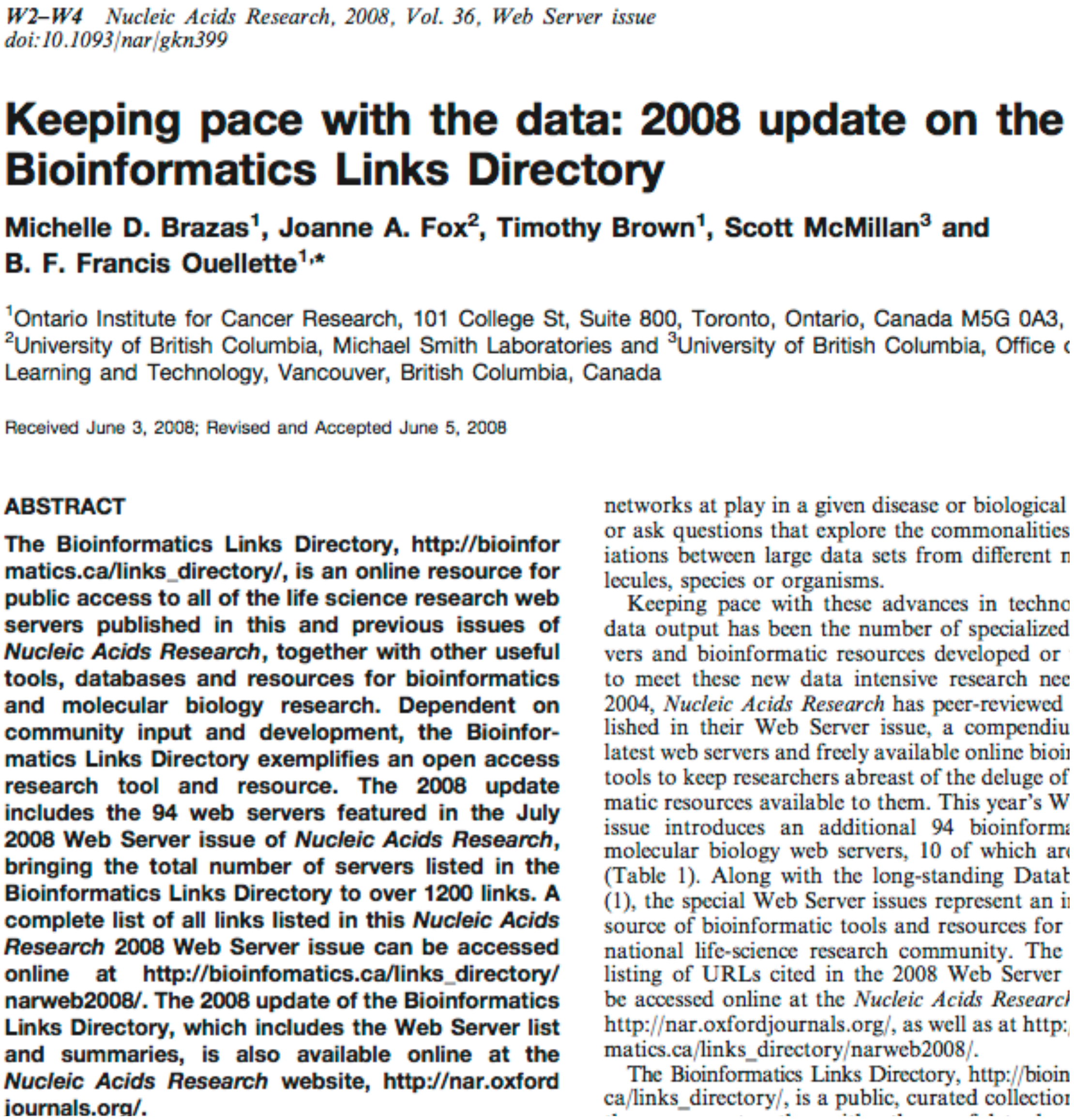
ABSTRACT

The Bioinformatics Links Directory, http://bioinformatics.ca/links_directory, is an actively maintained compilation of servers published in this and previous issues of *Nucleic Acids Research* issues together with many other useful tools, databases and resources for life sciences research. The 2007 update includes the 130 websites highlighted in the July 2007 Web Server issue of *Nucleic Acids Research* and brings the total number of servers listed in the Bioinformatics Links Directory to just under 1200 links. In addition to the updated content, the 2007 update of the Bioinformatics Links Directory includes new features for improved navigation, accessibility and open data exchange. A complete listing of all links listed in this *Nucleic Acids Research* 2007 Web Server issue can be accessed online at, http://bioinformatics.ca/links_directory/narweb2007. The 2007 update of the Bioinformatics Links Directory, which includes the Web Server list and summaries is also available online, at the *Nucleic Acids Research* web site, http://nar.oupjournals.org.

COMMENTARY

With the publication of the 2007 *Nucleic Acids Research* Web Server issue, we have a chance to reflect on how the web has transformed the way we conduct scientific

---



W2–W4 Nucleic Acids Research, 2008, Vol. 36, Web Server issue
doi:10.1093/nar/gkn399

Keeping pace with the data: 2008 update on the Bioinformatics Links Directory

Michelle D. Brazas[1], Joanne A. Fox[2], Timothy Brown[1], Scott McMillan[3] and B. F. Francis Ouellette[1,*]

[1]Ontario Institute for Cancer Research, 101 College St, Suite 800, Toronto, Ontario, Canada M5G 0A3,
[2]University of British Columbia, Michael Smith Laboratories and [3]University of British Columbia, Office
Learning and Technology, Vancouver, British Columbia, Canada

Received June 3, 2008; Revised and Accepted June 5, 2008

ABSTRACT

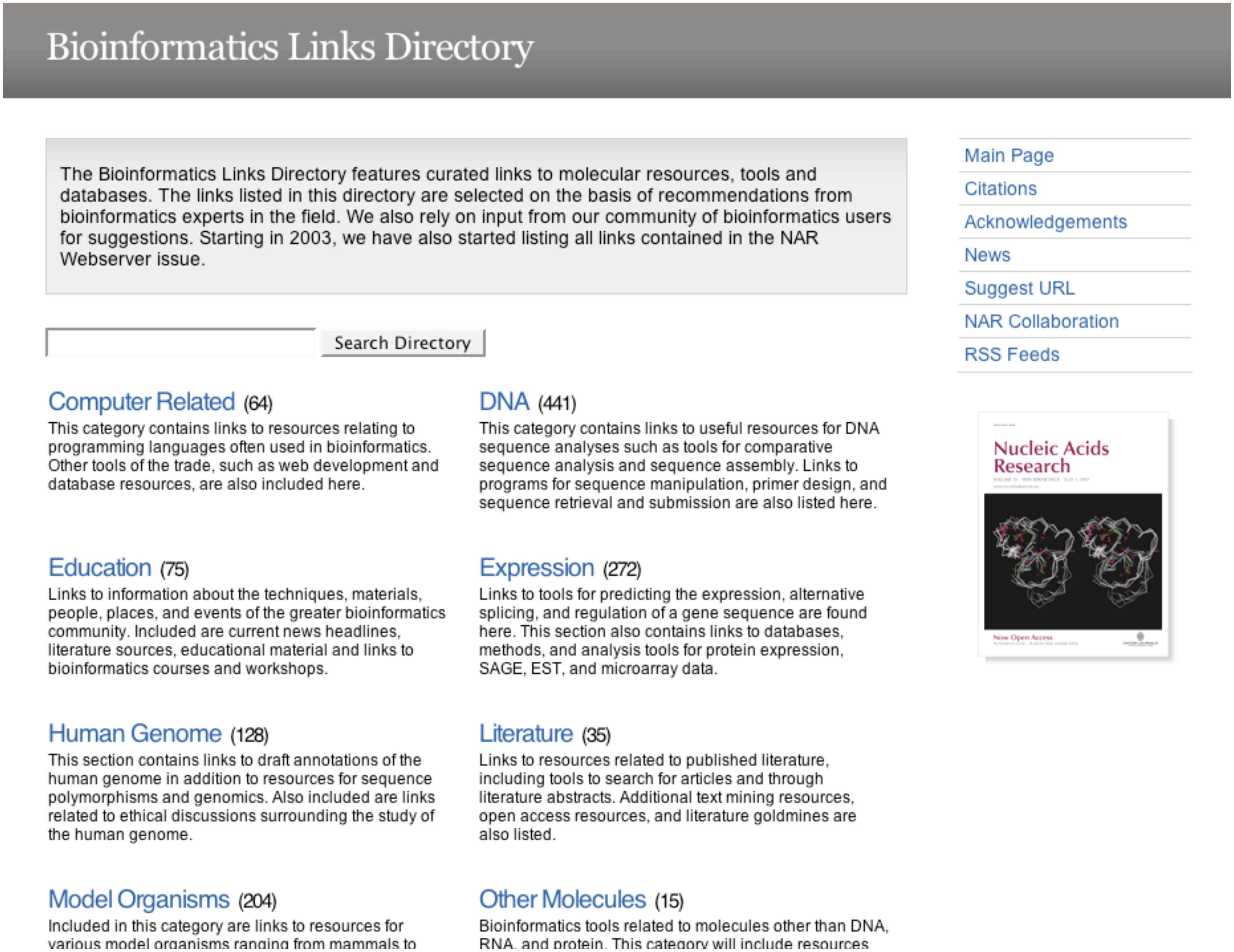
The Bioinformatics Links Directory, http://bioinformatics.ca/links_directory/, is an online resource for public access to all of the life science research web servers published in this and previous issues of *Nucleic Acids Research*, together with other useful tools, databases and resources for bioinformatics and molecular biology research. Dependent on community input and development, the Bioinformatics Links Directory exemplifies an open access research tool and resource. The 2008 update includes the 94 web servers featured in the July 2008 Web Server issue of *Nucleic Acids Research*, bringing the total number of servers listed in the Bioinformatics Links Directory to over 1200 links. A complete list of all links listed in this *Nucleic Acids Research* 2008 Web Server issue can be accessed online at http://bioinfomatics.ca/links_directory/narweb2008/. The 2008 update of the Bioinformatics Links Directory, which includes the Web Server list and summaries, is also available online at the *Nucleic Acids Research* website, http://nar.oxfordjournals.org/.

networks at play in a given disease or biological
or ask questions that explore the commonalities
iations between large data sets from different
lecules, species or organisms.

Keeping pace with these advances in techno
data output has been the number of specialized
vers and bioinformatic resources developed or
to meet these new data intensive research nee
2004, *Nucleic Acids Research* has peer-reviewed
lished in their Web Server issue, a compendiu
latest web servers and freely available online bioi
tools to keep researchers abreast of the deluge of
matic resources available to them. This year's W
issue introduces an additional 94 bioinform
molecular biology web servers, 10 of which ar
(Table 1). Along with the long-standing Data
(1), the special Web Server issues represent an
source of bioinformatic tools and resources for
national life-science research community. The
listing of URLs cited in the 2008 Web Server
be accessed online at the *Nucleic Acids Researc*
http://nar.oxfordjournals.org/, as well as at http:
matics.ca/links_directory/narweb2008/.

The Bioinformatics Links Directory, http://bioin
ca/links_directory/, is a public, curated collectio

http://bioinformatics.ca/links_directory/



Bioinformatics Links Directory

The Bioinformatics Links Directory features curated links to molecular resources, tools and databases. The links listed in this directory are selected on the basis of recommendations from bioinformatics experts in the field. We also rely on input from our community of bioinformatics users for suggestions. Starting in 2003, we have also started listing all links contained in the NAR Webserver issue.

Computer Related (64)

This category contains links to resources relating to programming languages often used in bioinformatics. Other tools of the trade, such as web development and database resources, are also included here.

Education (75)

Links to information about the techniques, materials, people, places, and events of the greater bioinformatics community. Included are current news headlines, literature sources, educational material and links to bioinformatics courses and workshops.

Human Genome (128)

This section contains links to draft annotations of the human genome in addition to resources for sequence polymorphisms and genomics. Also included are links related to ethical discussions surrounding the study of the human genome.

Model Organisms (204)

Included in this category are links to resources for various model organisms ranging from mammals to

DNA (441)

This category contains links to useful resources for DNA sequence analyses such as tools for comparative sequence analysis and sequence assembly. Links to programs for sequence manipulation, primer design, and sequence retrieval and submission are also listed here.

Expression (272)

Links to tools for predicting the expression, alternative splicing, and regulation of a gene sequence are found here. This section also contains links to databases, methods, and analysis tools for protein expression, SAGE, EST, and microarray data.

Literature (35)

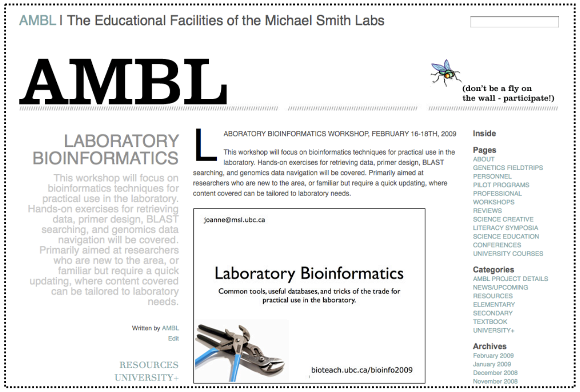
Links to resources related to published literature, including tools to search for articles and through literature abstracts. Additional text mining resources, open access resources, and literature goldmines are also listed.

Other Molecules (15)

Bioinformatics tools related to molecules other than DNA, RNA, and protein. This category will include resources

Your Feedback is Important!

navigate to:
bioteach.ubc.ca/bioinfo2009



AMBL | The Educational Facilities of the Michael Smith Labs

AMBL

(don't be a fly on the wall - participate!)

LABORATORY BIOINFORMATICS

This workshop will focus on bioinformatics techniques for practical use in the laboratory. Hands-on exercises for retrieving data, primer design, BLAST searching, and genomics data navigation will be covered. Primarily aimed at researchers who are new to the area, or familiar but require a quick updating, where content covered can be tailored to laboratory needs.

Written by AMBL
Edit

LABORATORY BIOINFORMATICS WORKSHOP, FEBRUARY 16-18TH, 2009

This workshop will focus on bioinformatics techniques for practical use in the laboratory. Hands-on exercises for retrieving data, primer design, BLAST searching, and genomics data navigation will be covered. Primarily aimed at researchers who are new to the area, or familiar but require a quick updating, where content covered can be tailored to laboratory needs.

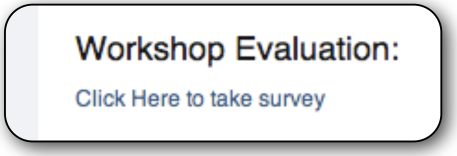
joanne@msl.ubc.ca

Laboratory Bioinformatics

Common tools, useful databases, and tricks of the trade for practical use in the laboratory.

bioteach.ubc.ca/bioinfo2009

➡



Thanks for attending the AMBL Laboratory Bioinformatics Workshop

- Questions? Please contact:

Dr. Joanne Fox
Michael Smith Laboratories
joanne@msl.ubc.ca