Cancer Gene Index End User Documentation

The information and links on this page are no longer being updated and are provided for reference purposes only.

Page Contents

- Overview of the Cancer Gene Index
- Means of Accessing Cancer Gene Index Data
- Selecting the Best Way for You to Access Cancer Gene Index Data
- Examples of How the Cancer Gene Index Facilitates Translational Research
  - Translational Medicine Research
  - Biomarker Discovery

Documentation Table of Contents

- Documentation Main Page
- Creation of the Cancer Gene Index
- Data, Metadata, and Annotations
- Cancer Gene Index Gene-Disease and Gene-Compound XML Documents
- caBIO APIs
- Cancer Gene Index Shared Parsed Data and Code
- caBIO Portlet Templated Searches
- caBIO Home Page
- caBIO iPhone Application
- caBIO Portlet Simple Searches
- Glossary
- Credits and Resources

To Print the Guide

We recommend you print one wiki page of the guide at a time. To do this, click the printer icon at the top right of the page; then from the browser File menu, choose Print. Printing multiple pages at one time is more complex. For instructions, refer to Exporting Multiple Pages to PDF.

Having Trouble Reading the Text?

Resizing the text for any web page is easy. For information on how to do this in your web browser, refer to this W3C tutorial

Overview of the Cancer Gene Index

There are nearly 2.5 million cancer-related publications in MEDLINE as of December 2009, and this number is rapidly increasing. Scientists cannot manually identify all known cancer genes, and it is even more difficult to uncover the relationships between these genes and various human cancers, for example. In theory, one could exhaustively search PubMed and compile a list of the genes related to a given disease or compound, but this would take many weeks, and it is highly likely that such a manual search would still miss some genes and relationships. The National Cancer Institute (NCI) recognized that a publicly-available resource that combined these gene-disease and gene-compound data with relevant annotations would greatly facilitate research, and as part of its caBIG® initiative, it created the Cancer Gene Index Project.

The goal of the Cancer Gene Index is to further translational cancer research by providing a high quality data resource consisting of genes that have been experimentally associated with human cancer diseases and/or pharmacological compounds, the evidence of these associations, and relevant annotations on the data. Thus, scientists can use the data resource to quickly discover and evaluate all of the genes associated with a disease, all of the genes associated with a compound, or all of the diseases and compounds associated with a gene. This extremely valuable resource was created through a unique process that coupled automated linguistic text analysis of millions of MEDLINE abstracts with manual validation and annotation of the extracted data by expert human curators. Details on this process are found in the section Creation of the Cancer Gene Index.

The Cancer Gene Index includes data on 6,955 unique human genes, nearly 12,000 NCI Thesaurus cancer disease terms, and 2,180 unique pharmacological compounds from the NCI Thesaurus. The gene-disease and gene-compound associations were extracted from over 92 million analyzed sentences of nearly 20 million abstracts. The resource was last updated in June, 2009.

Means of Accessing Cancer Gene Index Data

The Cancer Gene Index is available as computer-readable Gene-Disease and Gene-Compound data files. To effectively use these files, you must be a bioinformaticist or computer programmer-scientist, or you must collaborate with someone who has this expertise. Ideally, intuitive graphical user interfaces (GUIs) would allow all scientists to quickly and easily access these data and exploit the full power of the Cancer Gene Index. Already, geWorkbench and the Cancer Molecular Analysis Portal both allow end users to view some Cancer Gene Index data.

Several caBIO interfaces, on the other hand, expose all of the Cancer Gene Index data, and these can give you an appreciation for the full potential of the data resource. Many of these interfaces, however, require more experience with computer programming than the average bench scientist may have. Data within caBIO may be programmatically accessed through a variety of Application Programming Interfaces (APIs). The caBIO GUIs include the caBIO Portlet Templated Search, the caBIO Home Page, and Simple Search of the caBIO Portlet on the caGrid Portal.These caBIO GUIs are similar to PubMed in that queries will retrieve many results that you must sift through, examining each to determine whether or not it is useful. Unlike PubMed, caBIO is much more likely to return the information that you want.

Selecting the Best Way for You to Access Cancer Gene Index Data

The following section will help you select the best means to access Cancer Gene Index data based on your experience with bioinformatics and computer programming.

Bioinformaticists and Scientist Programmers

- **If you have limited knowledge of the caBIO object model and caBIG®**, you should use the Cancer Gene Index Gene-Disease and Gene-Compound XML documents with the Cancer Gene Index XML guide. The format of these documents is extremely simple, making them very easy with which to work. To download the XML documents, you must have a computer with at least 720 MB of free disk space, an internet connection, and a web browser; other system requirements depend upon the way in which you intend to use the data resource.
- **If you are familiar with the caBIO object model and caBIG®**, you may wish to refer to the documentation on the caBIO APIs. The caBIO APIs allow you to uncover associations within the Cancer Gene Index data set and to find additional information linking these data with associated pathways, protein annotations, clinical protocols, and other biomedical entities. For information on system requirements, please refer to the links for each API on the caBIO wiki page.

Scientists with No Programming or Bioinformatics Experience

You should use the step-by-step guide for the caBIO Portlet Templated Search tool. All that is required to access this web-based GUI is a computer with an internet connection and a web browser. Although it is easy to uncover gene-disease and gene-compound associations with this tool, it does not allow you to limit your search results and thus can return genes-disease and gene-compound associations that were not validated by human curators. Also, it does not necessarily return all of the data you would like. Because of these issues, you must use this tool in conjunction with the

- caBIO Object Graph Browser and, potentially, the
- NCI Thesaurus

Scientists Familiar with Classes, Objects, and Object Models

You can use the step-by-step guide to the caBIO Home Page, which has the Freestyle Lexical Mine and Search for Biological Entities tools. Although these interfaces expose the entirety of the Cancer Gene Index and the rest of caBIO, they require knowledge of the caBIO object model. All that is required to access these web-based caBIO search tools is a computer with an internet connection and a web browser.

caBIO Portlet Simple Search

The caBIO Portlet also has a Simple Search tool, which provides an overview of the data in caBIO in a way that does not require knowledge of the caBIO object model and can be useful for a quick look of the kinds of data within caBIO. Because the Simple Search tool does not allow you to differentiate Cancer Gene Index data from other data in caBIO, it is recommended that you instead use the XML, caBIO Portlet Templated Search, or even the caBIO Home Page. In the event that you would prefer to use the Simple Search a guide is provided.

Scientists On the Go

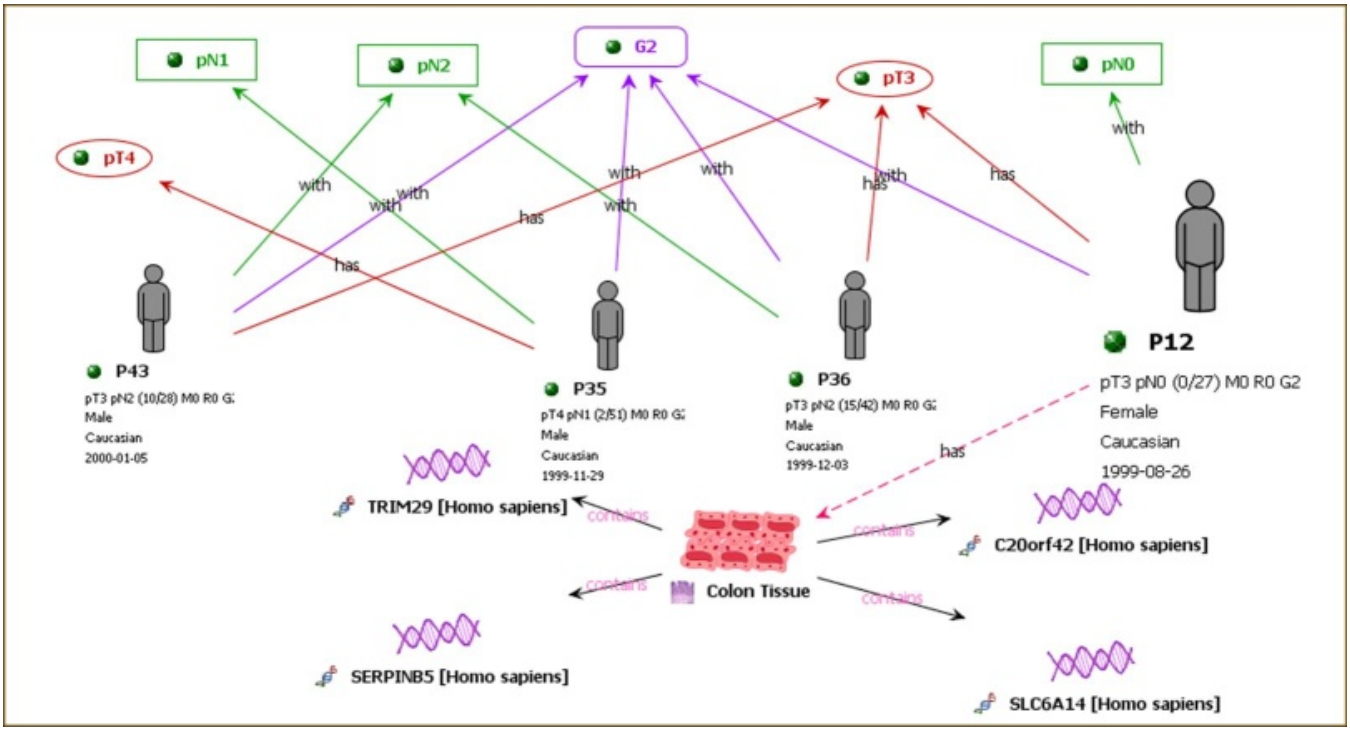
If you would like to view Cancer Gene Index data on the go, you can use the caBIO iPhone Application. A limited guide to accessing Cancer Gene Index data is provided here.

Examples of How the Cancer Gene Index Facilitates Translational Research

The Cancer Gene Index can facilitate many different types of cancer research. Two examples are given below.

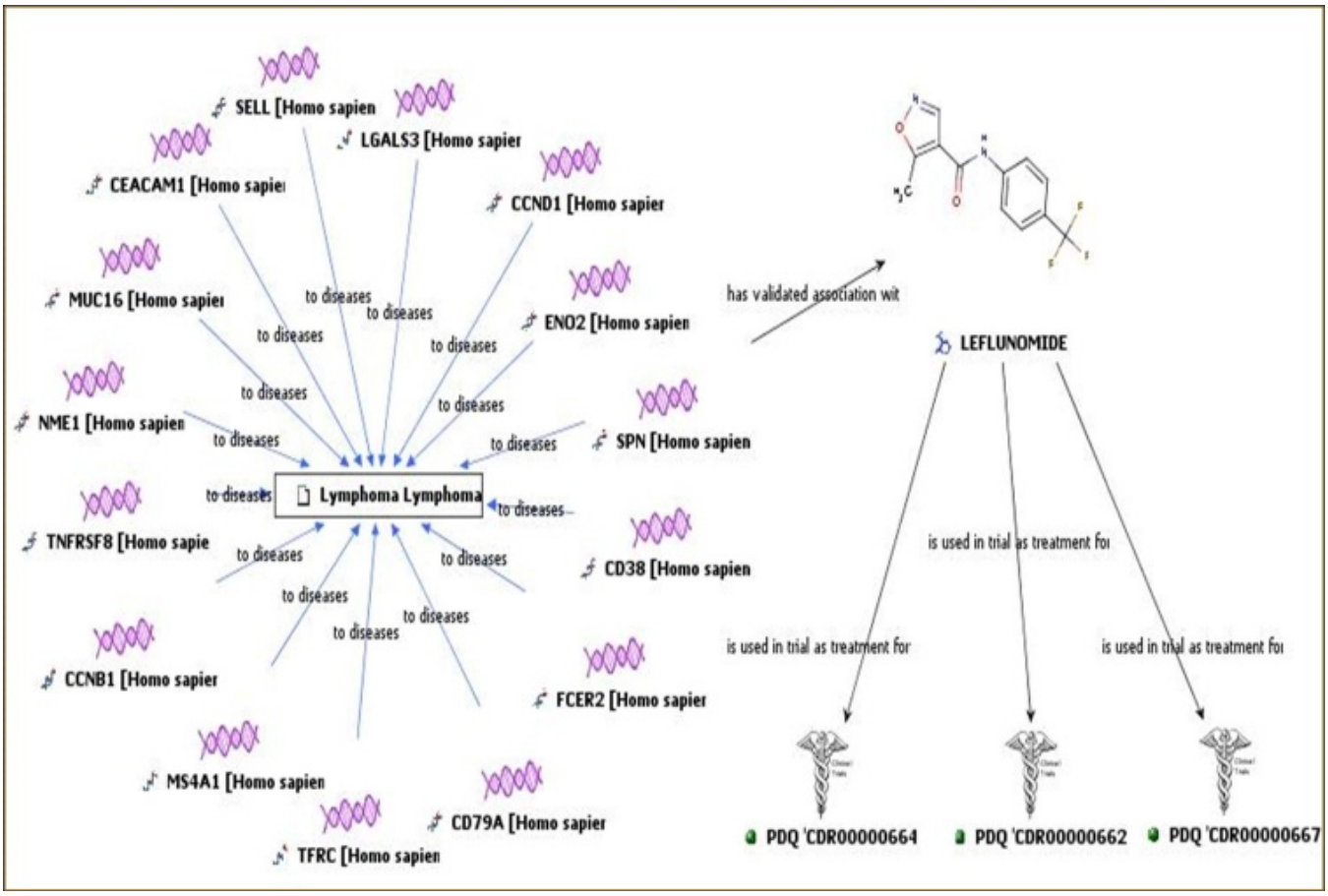
Translational Medicine Research

In this first example from the Cancer Gene Index Project poster, the data resource is used to validate colon cancer translational medicine research data. Here, scientists have obtained access to deidentified demographic data, histopathology data (lymph node [ICR:pN], tumor size [ICR:pT], and degree of metastasis [ICR:G]), and tumor tissue biospecimens from patients, which are represented by gray figures. The scientists perform gene expression microarrays on each colon cancer biospecimen (pink and red colon tissue cells). The genes (purple DNA fragments) with significantly altered expression are validated by cross-referencing the Cancer Gene Index.



Biomarker Discovery

The Cancer Gene Index also may be used for lymphoma biomarker discovery. This example from the Cancer Gene Index Project poster illustrates that researchers can use the data resource to quickly identify the genes (purple DNA fragments) that are associated with and may be biomarkers for Lymphoma. Here, gene-disease concept pair associations are shown as blue "to diseases" arrows. By searching the Cancer Gene Index for therapeutic compounds that are associated with these genes, scientists easily uncover which of these candidate disease biomarkers are also associated with lymphoma-related compounds. An association between the gene encoding SPN, also known as sialophorin or CD43, and the compound leflunomide is represented by a black "has validated association with" arrow. Cancer Gene Index data can be cross-referenced to other resources, such as the clinical trial protocol database Physician Data Query® (PDQ) to obtain information about trials that link these data.


SELL [Homo sapien
LGALS3 [Homo sapier
CEACAM1 [Homo sapie
CCND1 [Homo sapier
MUC16 [Homo sapie
ENO2 [Homo sapien
NME1 [Homo sapien
SPN [Homo sapien
TNFRSF8 [Homo sapie
CD38 [Homo sapien
CCNB1 [Homo sapier
FCER2 [Homo sapier
MS4A1 [Homo sapien
CD79A [Homo sapier
TFRC [Homo sapien
to diseases
Lymphoma Lymphoma
has validated association wit
LEFLUNOMIDE
is used in trial as treatment for
is used in trial as treatment for
is used in trial as treatment for
PDQ 'CDR00000664
PDQ 'CDR00000662
PDQ 'CDR00000667