

Database

Open Access

## Pancreatic Expression database: a generic model for the organization, integration and mining of complex cancer datasets

Claude Chelala\*<sup>1</sup>, Stephan A Hahn<sup>2</sup>, Hannah J Whiteman<sup>1</sup>, Sayka Barry<sup>1</sup>, Deepak Hariharan<sup>1</sup>, Tomasz P Radon<sup>1</sup>, Nicholas R Lemoine<sup>†1</sup> and Tatjana Crnogorac-Jurcevic<sup>†1</sup>

Address: <sup>1</sup>Centre for Molecular Oncology, Institute of Cancer & CR-UK Clinical Centre, Barts & The London School of Medicine (QMUL), Charterhouse Square London EC1M 6BQ, UK and <sup>2</sup>Molecular GI-Onkologie (MGO), University of Bochum, Germany

Email: Claude Chelala\* - [Claude.Chelala@cancer.org.uk](mailto:Claude.Chelala@cancer.org.uk); Stephan A Hahn - [Stephan.Hahn@rub.de](mailto:Stephan.Hahn@rub.de); Hannah J Whiteman - [Hannah.Whiteman@cancer.org.uk](mailto:Hannah.Whiteman@cancer.org.uk); Sayka Barry - [Sayka.Barry@cancer.org.uk](mailto:Sayka.Barry@cancer.org.uk); Deepak Hariharan - [deepak.hariharan@cancer.org.uk](mailto:deepak.hariharan@cancer.org.uk); Tomasz P Radon - [t.p.radon@qmul.ac.uk](mailto:t.p.radon@qmul.ac.uk); Nicholas R Lemoine - [Nick.Lemoine@cancer.org.uk](mailto:Nick.Lemoine@cancer.org.uk); Tatjana Crnogorac-Jurcevic - [T.C.Jurcevic@qmul.ac.uk](mailto:T.C.Jurcevic@qmul.ac.uk)

\* Corresponding author †Equal contributors

Published: 28 November 2007

Received: 8 August 2007

*BMC Genomics* 2007, **8**:439 doi:10.1186/1471-2164-8-439

Accepted: 28 November 2007

This article is available from: <http://www.biomedcentral.com/1471-2164/8/439>

© 2007 Chelala et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

**Background:** Pancreatic cancer is the 5th leading cause of cancer death in both males and females. In recent years, a wealth of gene and protein expression studies have been published broadening our understanding of pancreatic cancer biology. Due to the explosive growth in publicly available data from multiple different sources it is becoming increasingly difficult for individual researchers to integrate these into their current research programmes. The Pancreatic Expression database, a generic web-based system, is aiming to close this gap by providing the research community with an open access tool, not only to mine currently available pancreatic cancer data sets but also to include their own data in the database.

**Description:** Currently, the database holds 32 datasets comprising 7636 gene expression measurements extracted from 20 different published gene or protein expression studies from various pancreatic cancer types, pancreatic precursor lesions (PanINs) and chronic pancreatitis. The pancreatic data are stored in a data management system based on the BioMart technology alongside the human genome gene and protein annotations, sequence, homologue, SNP and antibody data. Interrogation of the database can be achieved through both a web-based query interface and through web services using combined criteria from pancreatic (disease stages, regulation, differential expression, expression, platform technology, publication) and/or public data (antibodies, genomic region, gene-related accessions, ontology, expression patterns, multi-species comparisons, protein data, SNPs). Thus, our database enables connections between otherwise disparate data sources and allows relatively simple navigation between all data types and annotations.

**Conclusion:** The database structure and content provides a powerful and high-speed data-mining tool for cancer research. It can be used for target discovery i.e. of biomarkers from body fluids, identification and analysis of genes associated with the progression of cancer, cross-platform meta-analysis, SNP selection for pancreatic cancer association studies, cancer gene promoter analysis as well as mining cancer ontology information. The data model is generic and can be easily extended and applied to other types of cancer. The database is available online with no restrictions for the scientific community at <http://www.pancreasexpression.org/>.

## Background

Pancreatic ductal adenocarcinoma (PDAC) usually presents at an advanced stage so that surgical cure is rarely achieved and conventional chemotherapy and radiotherapy have little impact, resulting in a very low 5-year survival rate (0.5%–5%) [1]. Thus a number of laboratories have focused on studying the evolution of pancreatic cancer from its earliest stages (pancreatic intraepithelial neoplasias or PanINs), putting pancreatic cancer among the best studied tumour tissue types at the molecular level. Thus a wealth of information regarding mutated and aberrantly expressed genes, miRNAs and proteins is now available, not only significantly boosting our biological understanding of the disease but also helping to identify new (early) diagnostic and therapeutic targets. Unfortunately, the huge and still rising volume and diversity of public pancreatic datasets makes it increasingly difficult for researchers to integrate this information into their current research efforts. In this report, we describe a dedicated Pancreatic Expression database [2] aiming to overcome this restriction, and furthermore propose it as a generic model for the organization, integration and presentation of complex cancer research data. The model is designed to address various research problems, ranging from the specimen origin and type, through cancer development stages to expression patterns. By bringing complex profiling data together, the Pancreatic Expression database should enable scientists worldwide to perform a whole range of user-friendly queries, from deciphering the biological mechanisms underlying pancreatic disease to target discovery.

## Construction and Content

### Construction

The aim of the Pancreatic Expression database is to provide a comprehensive mining tool for large-scale genomic, transcriptomic and proteomic data sets. In order to achieve this, we designed a robust internal structure encompassing specific pre-defined modules (which can be found under the "Filters" section in the database) including "*pancreatic specimen/cell type*", "*pancreatic differential expression information*", "*genes differentially expressed in*" and "*genes expressed in*" modules. Our design enables uploading of any available (pancreatic) datasets that comply with the structure of the pre-defined modules. Each module contains a number of subcategories related to the module name, which are fundamental to store and retrieve user-defined sub-datasets from the database by setting filters to the specific subcategories within each module. The "*pancreatic specimen/cell type*" module covers categories such as normal (microdissected ductal cells (ND) or bulk normal pancreas (NP), acinar cells, islet cells, stromal cells and pancreatic stellate cells), and disease specimens from both exocrine (pancreatic intraepithelial neoplasias (PanIN-1A, PanIN-1B, PanIN-2, PanIN-

3), chronic pancreatitis (CP), pancreatic adenocarcinoma (PDAC), intraductal papillary mucinous neoplasms (IPMN), mucinous cystic tumours and ampullary carcinoma) and endocrine (functioning and non-functioning tumours) origin. Moreover, pancreatic juice, plasma, urine, serum, and fine needle aspirates are included as additional options to further broaden future expansion of the database. The "*pancreatic differential expression information*" module provides information on direction of regulation (up- and down-regulation), fold-change, SAGE tag number and whether a gene or protein was found to be expressed only in pancreatic adenocarcinoma (PDAC) or in normal pancreas. The "*genes differentially expressed in*" module enables more defined selection of comparison methods such as pancreatic adenocarcinoma (PDAC) versus normal pancreas (bulk tissue or microdissected normal ductal cells), chronic pancreatitis (CP) versus normal pancreas (bulk tissue or microdissected normal ductal cells), chronic pancreatitis (CP) versus pancreatic adenocarcinoma (PDAC), pancreatic intraepithelial neoplasias (PanIN-1A, PanIN-1B, PanIN-2 or PanIN-3) versus normal pancreas (ND) or microdissected normal ductal cells, etc. The "*genes expressed in*" module lists the genes expressed in the tissue types defined in the pancreatic specimen/cell module, irrespective of their mode of regulation (whether they are differentially expressed or not). The "*platform technology*" module enables the selection of the technology used, such as Affymetrix arrays, cDNA arrays, Sanger human 10K cDNA arrays version 1.2.1, Sanger custom 5K1 cDNA arrays, Clontech Atlas Human Cancer cDNA Expression Array, SAGE, Agilent Human Genome CGH array, 2D PAGE, SELDI, etc. The data is stored in a data management system created using MySQL [3] and based on the open-source BioMart technology [4], a simple, federated query system designed specifically for use with large datasets. We imported the available Ensembl [5] human genome annotations (Ensembl release 41) for genes and proteins, SNP information, sequences, gene structure and multi-species data enabling the integration and annotation of heterogeneous pancreatic cancer data. In order to avoid integration and annotations errors, we used the pre-established Ensembl annotations and microarray probe set mapping. Ensembl links to UniProt/Swiss-Prot, RefSeq and UniProt/TrEMBL databases are made on the basis of sequence similarity. All other subsequent links are inferred from these mappings. Ensembl also establishes mappings to microarray probe set identifiers by matching probe set sequences to Ensembl transcripts [6]. We also integrated the antibody data from the Human Protein Atlas [7] based on Ensembl gene ID.

### Content

The Pancreatic Expression database currently contains 32 datasets from 20 different published sources, from 14

international laboratories encompassing 22 different platforms (Affymetrix GeneChip Human Full Length Array HuGeneFL, Affymetrix GeneChip Human Genome U95 Set (HG-U95A, HG-U95B, HG-U95C, HG-U95D, HG-U95E), Affymetrix GeneChip Human Genome U133 Array Set (HG-U133A, HG-U133B), 2D PAGE, cDNA arrays, SAGE, Operon oligo array version 2.0, Clontech Atlas Human Cancer cDNA Expression Array, immunohistochemistry, in situ hybridisation, Oligo array, MALDI, mass spectrometry, Sanger human 10K cDNA arrays version 1.2.1, Sanger custom 5K1 cDNA arrays, United Gene Technique Ltd, BD PowerBlot Western array and qRT-PCR) [8-27]. These initial datasets provide valuable information about 7636 gene expression measurements from a first-pass selection of relevant papers in the field of pancreatic research; however, the inclusion of additional relevant datasets will be a continuous and ongoing process. All the datasets were manually processed, checked for accuracy and consistency and loaded into our relational database alongside annotations from several public resources such as Ensembl, GO, dbSNP, UniProt and the Human protein atlas. Currently, several modules are present for which data are either not yet incorporated or not available (ICAT, iTRAQ), but these will be populated as it is our intention to continuously extend the current data content and cover all the existing modules as and when the data becomes available.

## Utility and Discussion

### Data access

The Pancreatic Expression database provides access not only to bioinformatic and biostatistic experts but also to bench researchers with a limited knowledge of bioinformatics. The database allows multiple levels of access. Firstly, access to the data is provided through a customized version of MartView, a BioMart web-based query interface based on Perl API [28]. The interface is navigated using the left panel with user selections taking place in the right panel. A summary of user choices is also displayed in the left panel. A simple query involves choosing attributes (or using the default ones) and optionally filters if one wants to restrict the query (Figure 1). Secondly, the Pancreatic Expression database is available from the BioMart central server (Figure 2A) where it is exposed to third party software such as the Bioconductor [29] package biomaRt [30] therefore allowing its easy interrogation within the open source R statistical environment [31] and its integration into any expression profiling experiment (Figure 2B). In addition, the database is exposed to the Taverna workflow system [32] (Figure 2C) and to the Galaxy framework (Figure 2D) [33]. The data can be also accessed programmatically through web services (Figure 3) [34]. A query constructed in the web-based query interface can be easily converted into an xml or perl template for future bioinformatics expansion and use. Finally, the Pancreatic Expression

database is a DAS server providing a Pancreatic Expression DAS annotation available at the Ensembl GeneView (Figure 4) [5].

### Examples of use

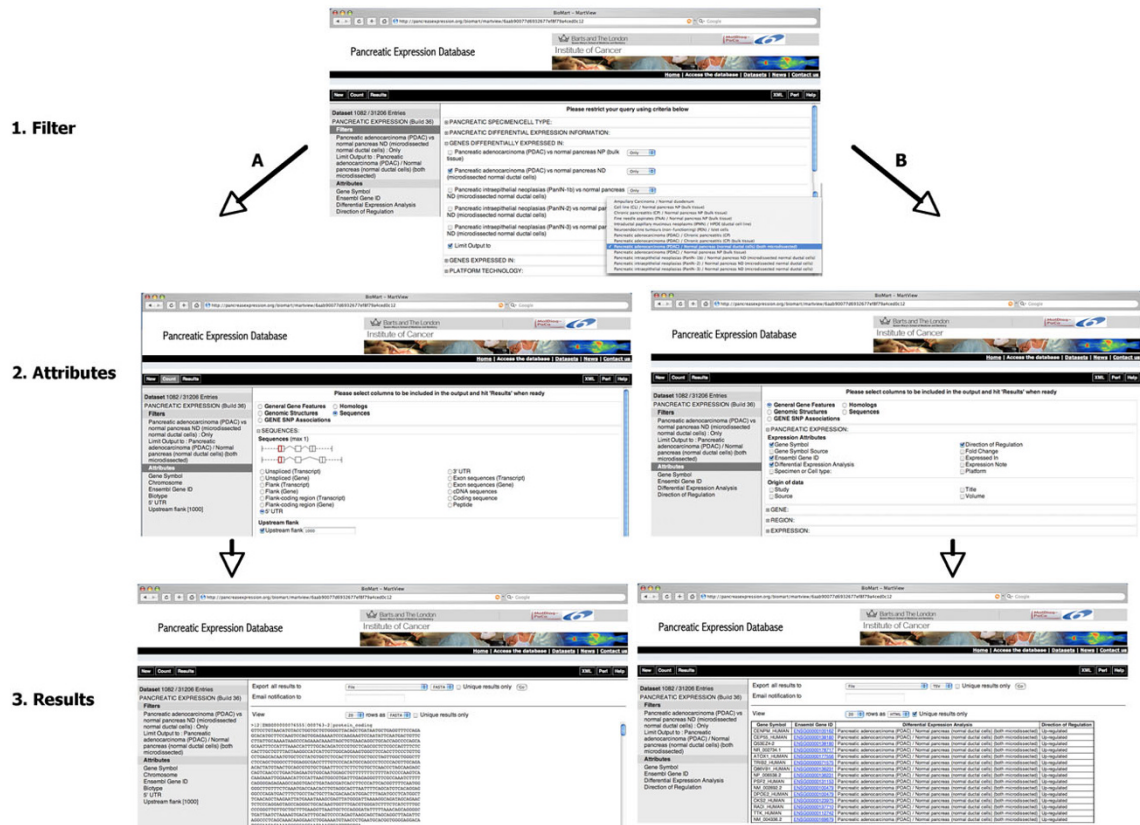
Navigation between all data types is simple and user-friendly; a variety of possible query combinations allow researchers quickly to determine the most de-regulated genes and proteins across all platforms.

Using the Pancreatic Expression database, it is possible to search and retrieve genes/proteins expressed only in pancreatic cancer and not in chronic pancreatitis and then ask which of these are present in urine and/or plasma. Such a query would be a first step for the discovery of non-invasive pancreatic cancer biomarkers from body fluids (Figure 5).

Researchers interested in the genes involved in the progression of pancreatic cancer can select the corresponding information among the differential expression datasets for the various tumour stages and retrieve the genes found to be de-regulated in the progression of pancreatic cancer (Figure 6). In the same way, one can search for genes specific to certain types of pancreatic cancer (Figure 1B).

Our database also allows cross-platform meta-analysis. Scientists can investigate pancreatic expression profiling performed across a wide range of different platforms (such as cDNA arrays or oligo arrays) to detect the most consistent sets of de-regulated genes (Figure 7). Importantly, scientists can also retrieve the sets of overlapping genes between their own results obtained by their particular platform (Proteomics, Affymetrix, Illumina etc.) and annotation method (UniProt, RefSeq, HGNC Hugo etc.) and those reported in the studies stored in the Pancreatic Expression database (Figure 8).

As they are available through the BiomaRt package [30], annotations can be added to any disease expression profiling experiment, which will allow detection of genes de-regulated in both pancreatic cancer and any other disease (Figure 2B). One can also obtain the gene ontology classification of the retrieved datasets or mine the genes of interest for a specific ontology term (Figure 9). Investigations using association studies can be designed using the Pancreatic Expression database by selecting a specific category of functional consequences (coding non-synonymous, 3' UTR, 5' UTR, splice site etc.) for SNPs associated with genes involved in pancreatic cancer (Figure 10). Expression data from a specific anatomical site can also be retrieved (Figure 11). Researchers can obtain immunohistochemistry data, where available, by selecting the antibodies filter (Figure 12). Scientists interested in promoter analysis can easily combine the gene search with the

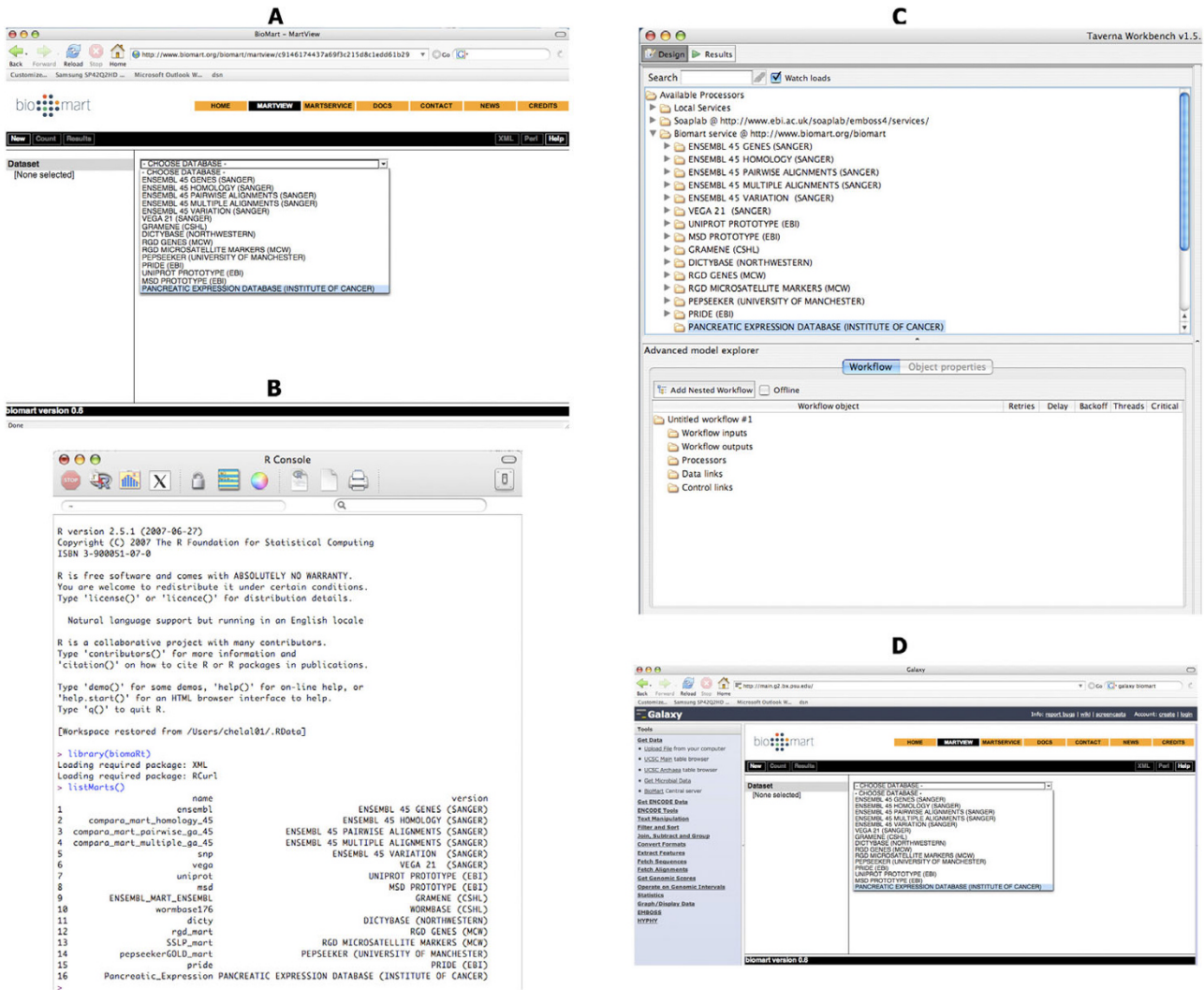


**Figure 1**  
**The Pancreatic Expression database web-based query interface.** In this first example our goal was to identify differentially regulated genes in pancreatic cancer found in microdissected samples only and/or to extract the upstream sequence data for these genes. This query starts by choosing the "Pancreatic Expression Database" from the "Choose Database" drop-down selection in the right panel. Users will be automatically directed to the "Pancreas Expression (Build 36)" dataset with the left panel displaying the "Filters" and "Attributes" nodes with their default settings. The next step involves choosing the appropriate attributes and filters to restrict the query. Clicking on the "Filters" or "Attributes" nodes on the left will display the related page where "Filters" or "Attributes" are arranged into sections, which can be expanded/collapsed using the "+/-" box. To choose an attribute or a filter, users can simply click on the checkbox next to its description. A summary of the selected filters and attributes is automatically displayed in the left panel. Clicking on the "Count" button in the tool bar at any time when constructing the query will return the number of genes satisfying the pre-selected criteria. One can select the option to download the results to a file at the top of the result page and export them using the "GO" button. Again, there are options to change the format ("HTML", "CSV" for comma separated values, "TSV" for tab separated values, "XLS" for Excel, "ADF" for array description format) and whether to make the results unique. One can select a compressed file output and the query will run in the background to be downloaded later. One needs to provide an email address to receive a URL in a notification email that allows the query results to be downloaded. 1. Filter: in this example, the query was restricted by selecting the "Filters" node on the left and choosing the filter options on the right. Next, the filter section "GENES DIFFERENTIALLY EXPRESSED IN:" was expanded in order to select the "Pancreatic adenocarcinoma (PDAC) vs normal pancreas ND (microdissected normal ductal cells)" and limited the output to the "Pancreatic adenocarcinoma (PDAC) vs normal pancreas ND (microdissected normal ductal cells)" from the "Limit Output to" drop down menu. This is to avoid retrieving all the information about the genes satisfying the search criteria. 2. Attributes: these are arranged into five pages for "General Gene Features", "Genomic Structures", "GENE SNP Associations", "Homologs" and "Sequences". In order to select the output content, the "Attributes" node on the left needs to be clicked on and the attribute page on the right needs to be chosen. A. from the "Sequences" attribute page, the section "SEQUENCES" was expanded and "5' UTR" selected by typing in "1000" basepair as an "Upstream flank". B. from the "General Gene Features" attribute page, one can expand the attribute section "PANCREATIC EXPRESSION" where the "Gene Symbol", "Ensembl Gene ID", "Differential Expression Analysis" and "Direction of Regulation" default options are displayed. 3. Results: clicking on the "Count" button displays 1082 genes satisfying the search criteria. Clicking on the "Results" button on the toolbar retrieves the final query results. By default, this shows the first 10 results in fasta format (A) and in hyper-linked HTML format (B) but the number and format can be altered using the drop-downs above this preview, as described above. Checking the "Unique results only" in (B) displays the unique rows only from the result table.

human genome upstream sequences and therefore collect the promoter sequences in a fast and simple way allowing further analysis of transcription factor-binding sites (Figure 1A).

**Conclusion**  
 Our integration model brings together relevant pancreatic cancer datasets and annotations from public sources and enables scientists to perform a wide variety of complex





**Figure 2**  
**Access through third party software.** The Pancreatic Expression database allows multiple levels of access. A – Access from the BioMart central server. B – Access within the open source R statistical environment through the biomaRt bioconductor package. C – Access from Taverna workflow system. D – Access with Galaxy.

queries on various types of data. The design of the database allows easy integration of additional modules and annotations from new public databases.

The Pancreatic Expression database constitutes a unique and valuable resource for the wider cancer research community, and is in rapid and constant development. We aim to continuously import new data sources and update the database on a regular basis, and invite scientists worldwide to deposit and share their data.

Although initially constructed using pancreatic cancer expression datasets, we have designed and implemented a

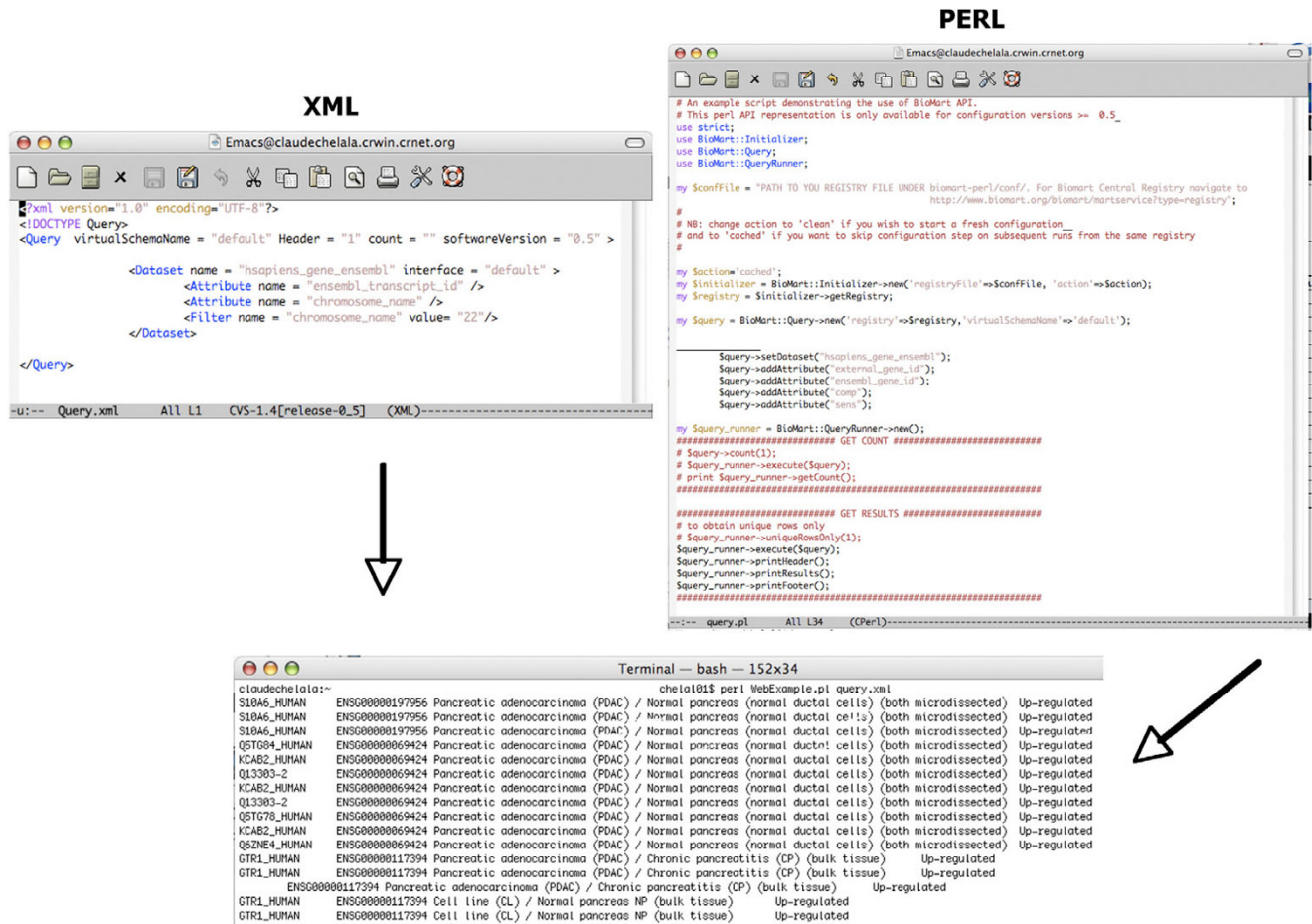
generic system that can be easily modified and applied to any other type of cancer. The system is available for collaboration with all interested research groups either by extending it to include other cancer data or by sharing our model should they want to adopt it for their data.

**Availability and requirements**

Project name: Pancreatic Expression database

Project home page: <http://www.pancreasexpression.org>

Operating system(s): Platform independent; Standard WWW browser (Safari, Firefox)



**Figure 3**  
**Access provided programmatically through web services.** In this example we show how to submit a query using web services. One can generate an XML or Perl script by building up a query using the web-based user interface (as described in Figure 1) and hitting the XML or Perl button on the right side of the toolbar. One can save the query as query.xml or query.pl and then use it to query the database.

Programming language: Perl, SQL, BioMart data management system

Licence: The database is freely available to academic and non-academic users. However, should you find the Pancreatic Expression database useful to your work, please cite this paper.

**Authors' contributions**

CC designed and implemented the web site and database, annotated and integrated the data, contributed to the data collection and wrote the manuscript. SAH defined the pancreatic cancer modules, collected pancreatic data, tested the database, contributed to the revision of the manuscript and continuous discussion. HJW, SB, DH and TPR were involved in the pancreatic data collection. TCJ

contributed to the definition of the pancreatic cancer modules and pancreatic data collection. TCJ and NRL provided valuable guidance and expertise on pancreatic cancer, contributed to the critical revision of the manuscript and continuous discussion. All authors read the final manuscript.

The screenshot displays the Ensembl GeneView interface for the gene ENSG0000038427 (VCAN). The main content area includes the following sections:

- Gene:** VCAN (HGNC Symbol). This gene is a member of the Human CCDS set: CCDS4060.
- Ensembl Gene ID:** ENSG0000038427
- Genomic Location:** This gene can be found on Chromosome 5 at location 82,803,339-82,912,737. The start of this gene is located in Contig AC016568.5.1.220932.
- Description:** Versican core protein precursor (Large fibroblast proteoglycan) (Chondroitin sulfate proteoglycan core protein 2) (PG-M) (Glial hyaluronate-binding protein) (GHAP). Source: Uniprot/SWISSPROT P13611.
- Prediction Method:** Genes were annotated by the Ensembl automatic analysis pipeline using either a GeneWise/Exonerate model from a database protein or a set of aligned cDNAs followed by an ORF prediction. GeneWise/Exonerate models are further combined with available aligned cDNAs to annotate UTRs (For more information see V.Curwen et al., Genome Res. 2004 14:342-50.)
- Transcripts:** A table listing transcripts with links to transcript, exon, and peptide information.
 

ENST00000265077	ENSP00000265077	CSPG2_HUMAN	[Transcript info]	[Exon info]	[Peptide info]
ENST00000342785	ENSP00000342785	P13611-3	[Transcript info]	[Exon info]	[Peptide info]
ENST00000343200	ENSP00000340062	P13611-2	[Transcript info]	[Exon info]	[Peptide info]
ENST00000348134	ENSP00000345946	P13611-4	[Transcript info]	[Exon info]	[Peptide info]
- Genomic Map:** A diagram showing the gene structure on Chromosome 5, with exons represented by red boxes and introns by lines with arrows. It includes a scale from 82,800 Mb to 82,900 Mb and a 120.40 kb scale bar.
- Alignments, Orthologue Prediction, Parologue Prediction:** Each section has a '+' icon to expand the information.
- Gene DAS Report:** A section for displaying data from various databases. The 'Pancreatic Expression' source is checked, and its results are shown:
 

PANCREATIC EXPRESSION (Build 36)	Up-regulated	Chronic pancreatitis (CP) / Normal pancreas NP (bulk tissue)
PANCREATIC EXPRESSION (Build 36)	Up-regulated	Pancreatic adenocarcinoma (PDAC) / Normal pancreas NP (bulk tissue)
PANCREATIC EXPRESSION (Build 36)	Up-regulated	Fine needle aspirates (FNA) / Normal pancreas NP (bulk tissue)
PANCREATIC EXPRESSION (Build 36)	Up-regulated	Pancreatic adenocarcinoma (PDAC) / Normal pancreas (normal ductal cells) (both microdissected)
PANCREATIC EXPRESSION (Build 36)	Up-regulated	Ampullary Carcinoma / Normal duodenum
- DAS Sources:** A list of databases with checkboxes. 'Pancreatic Expression' is checked. Other sources include AltSplice, AltTrans, ArrayExpress, GAD, HGNC, HUGO\_text, OMA, Phenotypes, Protonef, RZPD\_verif\_cDNA, RZPD\_esiRNA, RZPD\_Prot\_Exp, Reactome, UniProt, and aewTest. There are 'Update' and 'Manage Sources' buttons at the bottom.

**Figure 4**  
**The Pancreatic Expression DAS annotation available at the Ensembl GeneView.** The figure shows the Ensembl GeneView report for the ENSG0000038427 gene. The Pancreatic Expression database query results return Ensembl Gene ID by default with a hyper-link to the Ensembl GeneView website. In the Gene DAS Report section, which can be expanded/collapsed using the "+/-" box, a list of all the DAS sources including the Pancreatic Expression database is provided. To display the Pancreatic Expression database annotations, one can simply click on the checkbox next to the Pancreatic Expression and press the Update button at the end of the DAS sources.



### 1. Filter

Please restrict your query using criteria below

PANCREATIC SPECIMEN/CELL TYPE:

PANCREATIC DIFFERENTIAL EXPRESSION INFORMATION:

GENES DIFFERENTIALLY EXPRESSED IN:

Pancreatic adenocarcinoma (PDAC) vs normal pancreas NP (bulk tissue) Only

Pancreatic adenocarcinoma (PDAC) vs normal pancreas ND (microdissected normal ductal cells) Only

Chronic pancreatitis (CP) vs normal pancreas ND (microdissected normal ductal cells) Only

Chronic pancreatitis (CP) vs normal pancreas NP (bulk tissue) Only

Chronic pancreatitis (CP) vs pancreatic adenocarcinoma (PDAC) Only

Pancreatic adenocarcinoma (PDAC) vs chronic pancreatitis (CP) (bulk tissue) Only

GENES EXPRESSED IN:

Normal pancreas ND (microdissected normal ductal cells) Only

Plasma Only

Urine Only

### 2. Attributes

Please select columns to be included in the output and hit 'Results' when ready

General Gene Features  Homologs

Genomic Structures  Sequences

GENE SNP Associations

PANCREATIC EXPRESSION:

**Expression Attributes**

Gene Symbol  Direction of Regulation

Gene Symbol Source  Fold Change

Ensembl Gene ID  Expressed In

Differential Expression Analysis  Expression Note

Specimen or Cell type:  Platform

**Origin of data**

Study  Title

Source  Volume

GENE:

**Ensembl Attributes**

Transcript ID  % GC content

Peptide ID  Description

CDS length  Biotype

cDNA length  Source

Peptide length  Status

Transcript count

### 3. Results

**New** **Count** **Results** XML Perf Help

Dataset 37 / 31206 Entries  
PANCREATIC EXPRESSION (Build 36)

**Filters**

Pancreatic adenocarcinoma (PDAC) vs normal pancreas NP (bulk tissue) : Only  
Pancreatic adenocarcinoma (PDAC) vs chronic pancreatitis (CP) (bulk tissue) : Only  
Plasma : Only

**Attributes**

Ensembl Gene ID  
Description

Export all results to  (TSV)  Unique results only

Email notification to

View  rows as HTML  Unique results only

Ensembl Gene ID	Description
ENSG00000157901	Interferon-induced GTP-binding protein Mx1 (Interferon-regulated resistance GTP-binding protein MxA) (Interferon-induced protein p78) (IFI-78k). [Source:Uniprot/SWISSPROT/ Acc:P20591]
ENSG00000122611	Inhibin beta A chain precursor (Activin beta A chain) (Erythroid differentiation protein) (EDP). [Source:Uniprot/SWISSPROT/ Acc:P08476]
ENSG00000146674	Insulin-like growth factor-binding protein 3 precursor (IGFBP-3) (IGF-binding protein 3). [Source:Uniprot/SWISSPROT/ Acc:P17936]
ENSG00000106366	Plasminogen activator inhibitor 1 precursor (PAI-1) (Endothelial plasminogen activator inhibitor) (PAI). [Source:Uniprot/SWISSPROT/ Acc:P05121]
ENSG00000126396	Trypsin-1 precursor (EC 3.4.21.4) (Trypsin I) (Cationic trypsinogen). [Source:Uniprot/SWISSPROT/ Acc:P07477]
ENSG00000115414	Fibronectin precursor (FN) (Cold-insoluble globulin) (CIG). [Source:Uniprot/SWISSPROT/ Acc:P02751]
ENSG00000113461	Insulin-like growth factor-binding protein 5 precursor (IGFBP-5) (IGF-binding protein 5). [Source:Uniprot/SWISSPROT/ Acc:P24593]
ENSG00000109072	Vitronectin precursor (Serum-spreading factor) (S-protein) (V75) (Contains: Vitronectin V65 subunit; Vitronectin V10 subunit; Somatomedin B). [Source:Uniprot/SWISSPROT/ Acc:P04004]
ENSG00000185624	Protein disulfide-isomerase precursor (EC 5.3.4.1) (PDI) (Prolyl 4-hydroxylase subunit beta) (Cellular thyroid hormone-binding protein) (p55). [Source:Uniprot/SWISSPROT/ Acc:P07237]
ENSG00000117299	Endothelin-converting enzyme 1 (EC 3.4.24.71) (ECE-1). [Source:Uniprot/SWISSPROT/ Acc:P42892]

Figure 5

**Biomarker discovery.** The goal of this example was to retrieve genes expressed only in pancreatic cancer, not expressed in chronic pancreatitis and present in plasma. 1. Filter: the query is restricted by clicking on the "Filters" node on the left, expanding the filter section "GENES DIFFERENTIALLY EXPRESSED IN:" and selecting for genes differentially expressed in "Pancreatic adenocarcinoma (PDAC) vs normal pancreas NP (bulk tissue)" as well as "Pancreatic adenocarcinoma (PDAC) vs chronic pancreatitis (bulk tissue)" and expanding the filter section "GENES EXPRESSED IN:" and selecting for genes expressed in "Plasma". 2. Attributes: selected by clicking on the "Attributes" node on the left, choosing the "General Gene Features" attribute page, expanding the attribute section "PANCREATIC EXPRESSION" to remove "Gene Symbol", "Differential Expression Analysis" and "Direction of Regulation" default options, expanding the attribute section "GENE" and selecting "Description" under "Ensembl Attributes". 3. Results: Clicking the "Count" button in the tool bar returns 37 genes satisfying the filters criteria. Clicking the "Results" button on the toolbar will retrieve a result table. Checking the "Unique results only" will display only the unique rows of the result table.

### 1. Filter

Dataset 10 / 31206 Entries  
PANCREATIC EXPRESSION (Build 36)

Please restrict your query using criteria below

Filters

- Pancreatic intraepithelial neoplasias (PanIN-2) vs normal pancreas ND (microdissected normal ductal cells) : Only
- Pancreatic intraepithelial neoplasias (PanIN-3) vs normal pancreas ND (microdissected normal ductal cells) : Only
- Pancreatic adenocarcinoma (PDAC) vs normal pancreas NP (bulk tissue) : Only

Attributes

- Ensembl Gene ID
- Description

GENES DIFFERENTIALLY EXPRESSED IN:

- Pancreatic adenocarcinoma (PDAC) vs normal pancreas NP (bulk tissue) Only
- Pancreatic adenocarcinoma (PDAC) vs normal pancreas ND (microdissected normal ductal cells) Only
- Chronic pancreatitis (CP) vs normal pancreas ND (microdissected normal ductal cells) Only
- Chronic pancreatitis (CP) vs normal pancreas NP (bulk tissue) Only
- Chronic pancreatitis (CP) vs pancreatic adenocarcinoma (PDAC) Only
- Pancreatic intraepithelial neoplasias (PanIN-2) vs normal pancreas ND (microdissected normal ductal cells) Only
- Pancreatic intraepithelial neoplasias (PanIN-3) vs normal pancreas ND (microdissected normal ductal cells) Only

### 2. Attributes

Dataset 10 / 31206 Entries  
PANCREATIC EXPRESSION (Build 36)

Please select columns to be included in the output and hit 'Results' when ready

Filters

- Pancreatic intraepithelial neoplasias (PanIN-2) vs normal pancreas ND (microdissected normal ductal cells) : Only
- Pancreatic intraepithelial neoplasias (PanIN-3) vs normal pancreas ND (microdissected normal ductal cells) : Only
- Pancreatic adenocarcinoma (PDAC) vs normal pancreas NP (bulk tissue) : Only

Attributes

- Ensembl Gene ID
- Description

General Gene Features

- Genomic Structures
- GENE SNP Associations

PANCREATIC EXPRESSION:

Expression Attributes

- Gene Symbol
- Gene Symbol Source
- Ensembl Gene ID
- Differential Expression Analysis
- Specimen or Cell type:

Origin of data

- Study
- Source

GENE:

Ensembl Attributes

- Transcript ID
- Peptide ID
- CDS length
- rDNA length
- Peptide length
- Transcript count
- % GC content
- Description
- Biotype
- Source
- Status

### 3. Results

Dataset 10 / 31206 Entries  
PANCREATIC EXPRESSION (Build 36)

Export all results to  TSV  Unique results only

Email notification to

View  rows as   Unique results only

Ensembl Gene ID	Description
<a href="#">ENSG00000169910</a>	Argininosuccinate lyase (EC 4.3.2.1) (Argininosuccinase) (ASAL). [Source:UniProt/SWISSPROT;Acc:P04424]
<a href="#">ENSG00000142789</a>	Elastase-3B precursor (EC 3.4.21.70) (Elastase IIIB) (Protease E). [Source:UniProt/SWISSPROT;Acc:P08861]
<a href="#">ENSG00000184292</a>	Tumor-associated calcium signal transducer 2 precursor (Pancreatic carcinoma marker protein GA733-1) (Cell surface glycoprotein Trop-2). [Source:UniProt/SWISSPROT;Acc:P09758]
<a href="#">ENSG00000168925</a>	chymotrypsinogen B2 [Source:RefSeq;peptide;Acc:NP_001020371]
<a href="#">ENSG00000165949</a>	Interferon-alpha-induced 11.5 kDa protein (p27) (ISG12(a) protein). [Source:UniProt/SWISSPROT;Acc:P40305]
<a href="#">ENSG00000175535</a>	Pancreatic triacylglycerol lipase precursor (EC 3.1.1.3) (Pancreatic lipase) (PL). [Source:UniProt/SWISSPROT;Acc:P16233]
<a href="#">ENSG00000165862</a>	Pancreatic lipase-related protein 2 precursor (EC 3.1.1.3). [Source:UniProt/SWISSPROT;Acc:P54317]
<a href="#">ENSG00000196565</a>	Hemoglobin subunit epsilon (Hemoglobin epsilon chain) (Epsilon-globin). [Source:UniProt/SWISSPROT;Acc:P02100]
<a href="#">ENSG00000170835</a>	Bile salt-activated lipase precursor (EC 3.1.1.3) (EC 3.1.1.13) (BAL) (Bile salt-stimulated lipase) (BSSL) (Carboxyl ester lipase) (Sterol esterase) (Cholesterol esterase) (Pancreatic lysophospholipase). [Source:UniProt/SWISSPROT;Acc:P19835]
<a href="#">ENSG00000183993</a>	Protein S100-P (S100 calcium-binding protein P). [Source:UniProt/SWISSPROT;Acc:P25815]

Figure 6

**Genes involved in the progression of pancreatic cancer.** The goal of this example was to query the differentially expressed information for pancreatic tumour stages (pancreatic intraepithelial neoplasias and pancreatic adenocarcinoma) and to retrieve the genes found to be de-regulated in the progression of pancreatic cancer. 1. Filter: one can choose "Pancreatic intraepithelial neoplasias (PanIN-2) vs normal pancreas ND (microdissected normal ductal cells)", "Pancreatic intraepithelial neoplasias (PanIN-3) vs normal pancreas ND (microdissected normal ductal cells)" and "Pancreatic adenocarcinoma (PDAC) vs normal pancreas NP (bulk tissue)" from the filter section "GENES DIFFERENTIALLY EXPRESSED IN:". 2. Attributes: one can choose the "General Gene Features" attribute page, expand the attribute section "PANCREATIC EXPRESSION" to remove "Gene Symbol", "Differential Expression Analysis" and "Direction of Regulation" default options, expand the attribute section "GENE" and select "Description" under "Ensembl Attributes". 3. Results: Clicking the "Count" button in the tool bar returns 10 genes involved in the progression of pancreatic cancer.

### 1. Filter

Please restrict your query using criteria below

PANCREATIC SPECIMEN/CELL TYPE:

PANCREATIC DIFFERENTIAL EXPRESSION INFORMATION:

GENES DIFFERENTIALLY EXPRESSED IN:

Pancreatic adenocarcinoma (PDAC) vs normal pancreas NP (bulk tissue) Only ▾

Pancreatic adenocarcinoma (PDAC) vs normal pancreas ND (microdissected normal ductal cells) Only ▾

PLATFORM TECHNOLOGY:

Affymetrix array Affymetrix GeneChip Human Full Length Array HuGeneFL ▾ Only

SAGE Only ▾

aRNA\_longSAGE Only ▾

2D PAGE Only ▾

cDNA array Only ▾

Oligo array Only ▾

### 2. Attributes

Please select columns to be included in the output and hit 'Results' when ready

General Gene Features  Homologs

Genomic Structures  Sequences

GENE SNP Associations

PANCREATIC EXPRESSION:

Expression Attributes

Gene Symbol  Direction of Regulation

Gene Symbol Source  Fold Change

Ensembl Gene ID  Expressed In

Differential Expression Analysis  Expression Note

Specimen or Cell type:  Platform

Origin of data

Study  Title

Source  Volume

GENE:

Ensembl Attributes

Transcript ID  % GC content

Peptide ID  Description

CDS length  Biotype

cDNA length  Source

Peptide length  Status

Transcript count

### 3. Results

New Count Results
XML Perf Help

Dataset 141 / 31206 Entries

PANCREATIC EXPRESSION (Build 36)

**Filters**

cDNA array: Only

Oligo array: Only

Pancreatic adenocarcinoma (PDAC) vs normal pancreas NP (bulk tissue) : Only

**Attributes**

Ensembl Gene ID

Description

Export all results to  TSV  Unique results only

Email notification to

View  rows as HTML  Unique results only

Ensembl Gene ID	Description
<a href="#">ENSG00000160180</a>	Trefoil factor 3 precursor (intestinal trefoil factor) (hP1.B). [Source:Uniprot/SWISSPROT;Acc:Q07654]
<a href="#">ENSG00000160181</a>	Trefoil factor 2 precursor (Spasmodytic polypeptide) (SP) (Spasmolysin). [Source:Uniprot/SWISSPROT;Acc:Q03403]
<a href="#">ENSG00000160182</a>	Trefoil factor 1 precursor (pS2 protein) (HP1.A) (Breast cancer estrogen-inducible protein) (PNR-2). [Source:Uniprot/SWISSPROT;Acc:P04155]
<a href="#">ENSG00000160255</a>	Integrin beta-2 precursor (Cell surface adhesion glycoproteins LFA- 1/CR3/p150,95 subunit beta) (Complement receptor C3 subunit beta) (CD18 antigen). [Source:Uniprot/SWISSPROT;Acc:P05107]

**Figure 7**

**Cross-platform analysis of pancreatic cancer data in the database.** The goal of this example was to investigate pancreatic expression profiling data obtained using cDNA and oligo arrays and to retrieve the common set of de-regulated genes. One can proceed as follows: 1. Filters: restrict the query by clicking on the "Filters" node on the left. Choose the cancer subtype, for example by expanding the filter section "GENES DIFFERENTIALLY EXPRESSED IN:" and selecting for genes differentially expressed in "Pancreatic adenocarcinoma (PDAC) vs normal pancreas NP (bulk tissue)", expanding the filter section "PLATFORM TECHNOLOGY:" and selecting both "Oligo array" and "cDNA array". 2. Attributes: select attributes by clicking on the "Attributes" node on the left, choosing the "General Gene Features" attribute, expanding the attribute section "PANCREATIC EXPRESSION" to remove "Gene Symbol", "Differential Expression Analysis" and "Direction of Regulation" default options, expanding the attribute section "GENE" and selecting "Description" under "Ensembl Attributes". 3. Results: hitting the "Count" button retrieves 141 de-regulated genes in cancer when using both oligo and cDNA platforms. One can proceed to saving the whole result set by choosing the appropriate action(s) at the top of the page, for example in an Excel format, and clicking the "Go" button to export the results.



### 1. Filter

Please restrict your query using criteria below

PANCREATIC SPECIMEN/CELL TYPE:

PANCREATIC DIFFERENTIAL EXPRESSION INFORMATION:

GENES DIFFERENTIALLY EXPRESSED IN:

GENES EXPRESSED IN:

PLATFORM TECHNOLOGY:

ANTIBODIES:

GENOMIC REGION:

GENE RELATED ACCESSIONS:

ID list filters with CCDS ID(s)  Only  Excluded

ID list limit IPI ID(s)  Only  Excluded

Test

- Unigene.txt
- expression.txt
- IPI.txt**
- genelid.txt
- affymetrix.133A.txt

IPI.txt

- IPI00456830
- IPI00456238
- IPI00456623
- IPI00456942
- IPI00457056
- IPI00457307
- IPI00464973

### 2. Attributes

Please select columns to be included in the output and hit 'Results' when ready

General Gene Features  Homologs

Genomic Structures  Sequences

GENE SNP Associations

**PANCREATIC EXPRESSION:**

**Expression Attributes**

Gene Symbol  Direction of Regulation

Gene Symbol Source  Fold Change

Ensembl Gene ID  Expressed In

Differential Expression Analysis  Expression Note

Specimen or Cell type:  Platform

**Origin of data**

Study  Title

Source  Volume

**GENE:**

**Ensembl Attributes**

Transcript ID  % GC content

Peptide ID  Description

CDS length  Biotype

cDNA length  Source

Peptide length  Status

Transcript count

**GO Attributes**

GO ID  GO evidence code

GO description

**External References (max 3) (max 3)**

CCDS ID  Protein ID

Codelink ID  RefSeq DNA ID

EMBL ID  RefSeq Predicted DNA ID

EntrezGene ID  RefSeq Peptide ID

Havana ID  Rfam ID

HGNC Symbol  Unigene ID

Illumina ID  UniProt/SPTREMBL ID

IPI ID  UniProt/Swiss-Prot ID

miRNA Registry ID  UniProt/Swiss-Prot Accession

MIM Gene Accession  Unified UniProt ID

MIM Morbid accession  Unified UniProt Accession

PDB ID  Uniprot varsplic ID

### 3. Results

Export all results to:  File  HTML  CSV  XLS  ADF  Unique results only (Go)

Email notification to:

View:  rows as HTML  Unique results only

Gene Symbol	Ensembl Gene ID	Differential Expression Analysis	Direction of Regulation	IPI ID
CENPM_HUMAN	ENS000001193192	Pancreatic adenocarcinoma (PDAC) / Normal pancreas (normal ductal cells) (both microdissected)	Up-regulated	IPI00031566
CEP55_HUMAN	ENS00000138180	Pancreatic adenocarcinoma (PDAC) / Normal pancreas (normal ductal cells) (both microdissected)	Up-regulated	IPI00101532
CEP55_HUMAN	ENS00000138180	Pancreatic adenocarcinoma (PDAC) / Normal pancreas (normal ductal cells) (both microdissected)	Up-regulated	IPI00478215
QSOX2-2	ENS00000138180	Pancreatic adenocarcinoma (PDAC) / Normal pancreas (normal ductal cells) (both microdissected)	Up-regulated	IPI00478215
QSOX2-2	ENS00000138180	Pancreatic adenocarcinoma (PDAC) / Normal pancreas (normal ductal cells) (both microdissected)	Up-regulated	IPI00478543
NR_002794.1	ENS00000178717	Pancreatic adenocarcinoma (PDAC) / Normal pancreas (normal ductal cells) (both microdissected)	Up-regulated	IPI00025908
ATOX1_HUMAN	ENS00000177556	Pancreatic adenocarcinoma (PDAC) / Normal pancreas (normal ductal cells) (both microdissected)	Up-regulated	IPI00010863
TRIB2_HUMAN	ENS00000071675	Pancreatic adenocarcinoma (PDAC) / Normal pancreas (normal ductal cells) (both microdissected)	Up-regulated	IPI00012516
OBFBV1_HUMAN	ENS00000136231	Pancreatic adenocarcinoma (PDAC) / Normal pancreas (normal ductal cells) (both microdissected)	Up-regulated	IPI00739282
OBFBV1_HUMAN	ENS00000136231	Pancreatic adenocarcinoma (PDAC) / Normal pancreas (normal ductal cells) (both microdissected)	Up-regulated	IPI00746216

**Figure 8**

**Meta-analysis of pancreatic cancer data.** In this example, our goal was to retrieve the sets of overlapping genes between user's data obtained by a proteomics platform with IPI identifiers and those reported in the studies stored in the Pancreatic Expression database. 1. Filter: one can restrict the query by clicking on the "Filters" node on the left and expanding the filter section "GENE RELATED ACCESSIONS:" using the "+" box. Next one can choose the "ID" list limit simply by clicking on the left checkbox and selecting the identifiers for the experimental results from the right drop-down menu. In this example, the "IPI ID(s)" is used but one can choose any other identifiers such as Uniprot, Affymetrix or Illumina etc. One can copy and paste the list of identifiers in the box or browse to read directly from a file. 2. Attributes: one can click on the "Attributes" node on the left and start choosing attributes from the "General Gene Features" page. One can expand the "PANCREATIC EXPRESSION:" section or other sections and check other attributes to add to the results. 3. Results: for saving the whole result set, one can choose the appropriate action(s) at the top of the page, select the format, for example Excel, and click the "Go" button to export the results.

Please restrict your query using criteria below

**1. Filter**

- PANCREATIC SPECIMEN/CELL TYPE:
- PANCREATIC DIFFERENTIAL EXPRESSION INFORMATION:
- GENES DIFFERENTIALLY EXPRESSED IN:
  - Pancreatic adenocarcinoma (PDAC) vs normal pancreas NP (bulk tissue) Only
  - Pancreatic adenocarcinoma (PDAC) vs normal pancreas ND (microdissected normal ductal cells) Only
  - Chronic pancreatitis (CP) vs normal pancreas ND (microdissected normal ductal cells) Only
  - Chronic pancreatitis (CP) vs normal pancreas NP (bulk tissue) Only
  - Chronic pancreatitis (CP) vs pancreatic adenocarcinoma (PDAC) Only
  - Pancreatic adenocarcinoma (PDAC) vs chronic pancreatitis (CP) (bulk tissue) Only
  - Intraductal papillary mucinous neoplasms (IPMN) vs cell line Only

Please restrict your query using criteria below

- PANCREATIC SPECIMEN/CELL TYPE:
- PANCREATIC DIFFERENTIAL EXPRESSION INFORMATION:
- GENES DIFFERENTIALLY EXPRESSED IN:
- GENES EXPRESSED IN:
- PLATFORM TECHNOLOGY:
- ANTIBODY DATA:
- ARTICLES:
- GENOMIC REGION:
- GENE RELATED ACCESSIONS:
- GENE ONTOLOGY:
  - Evidence code (Molecular function) ISS
  - Molecular function browse
  - Evidence code (Biological process) ISS
  - Biological process CO:0008283 browse

Please select columns to be included in the output and hit 'Results' when ready

**2. Attributes**

- General Gene Features
- Homologs
- Genomic Structures
- Sequences
- GENE SNP Associations

**PANCREATIC EXPRESSION:**

- Expression Attributes
- Direction of Regulation
- Gene Symbol
- Fold Change
- Gene Symbol Source
- Expressed In
- Ensembl Gene ID
- Expression Note
- Differential Expression Analysis
- Platform
- Specimen or Cell type:

**GENE:**

- GO Attributes
- GO evidence code
- GO ID
- GO description

**3. Results**

Dataset 92 / 31206 Entries  
PANCREATIC EXPRESSION (Build 36)

Export all results to: File | CSV | Unique results only

Filters: Intraductal papillary mucinous neoplasms (IPMN) vs cell line - Only

View: 150 rows as HTML | Unique results only

Gene Symbol	Ensembl Gene ID	GO description
BACE2_HUMAN	ENS0000018294	pepsin A activity
BACE2_HUMAN	ENS0000018294	peptidase activity
BACE2_HUMAN	ENS0000018294	aspartic-type signal peptidase activity
BACE2_HUMAN	ENS0000018294	protein modification
BACE2_HUMAN	ENS0000018294	membrane protein ectodomain proteolysis
BACE2_HUMAN	ENS0000018294	protein secretion
BACE2_HUMAN	ENS0000018294	peptide hormone processing
BACE2_HUMAN	ENS0000018294	negative regulation of amyloid precursor protein biosynthesis
BACE2_HUMAN	ENS0000018294	membrane fraction
BACE2_HUMAN	ENS0000018294	membrane
BACE2_HUMAN	ENS0000018294	integral to membrane
BACE2_HUMAN	ENS0000018294	proteolysis
BACE2_HUMAN	ENS0000018294	extracellular space
BACE2_HUMAN	ENS0000018294	hydrolase activity
NP_60476.1	ENS0000018294	pepsin A activity
NP_60476.1	ENS0000018294	peptidase activity

Dataset 92 / 31206 Entries  
PANCREATIC EXPRESSION (Build 36)

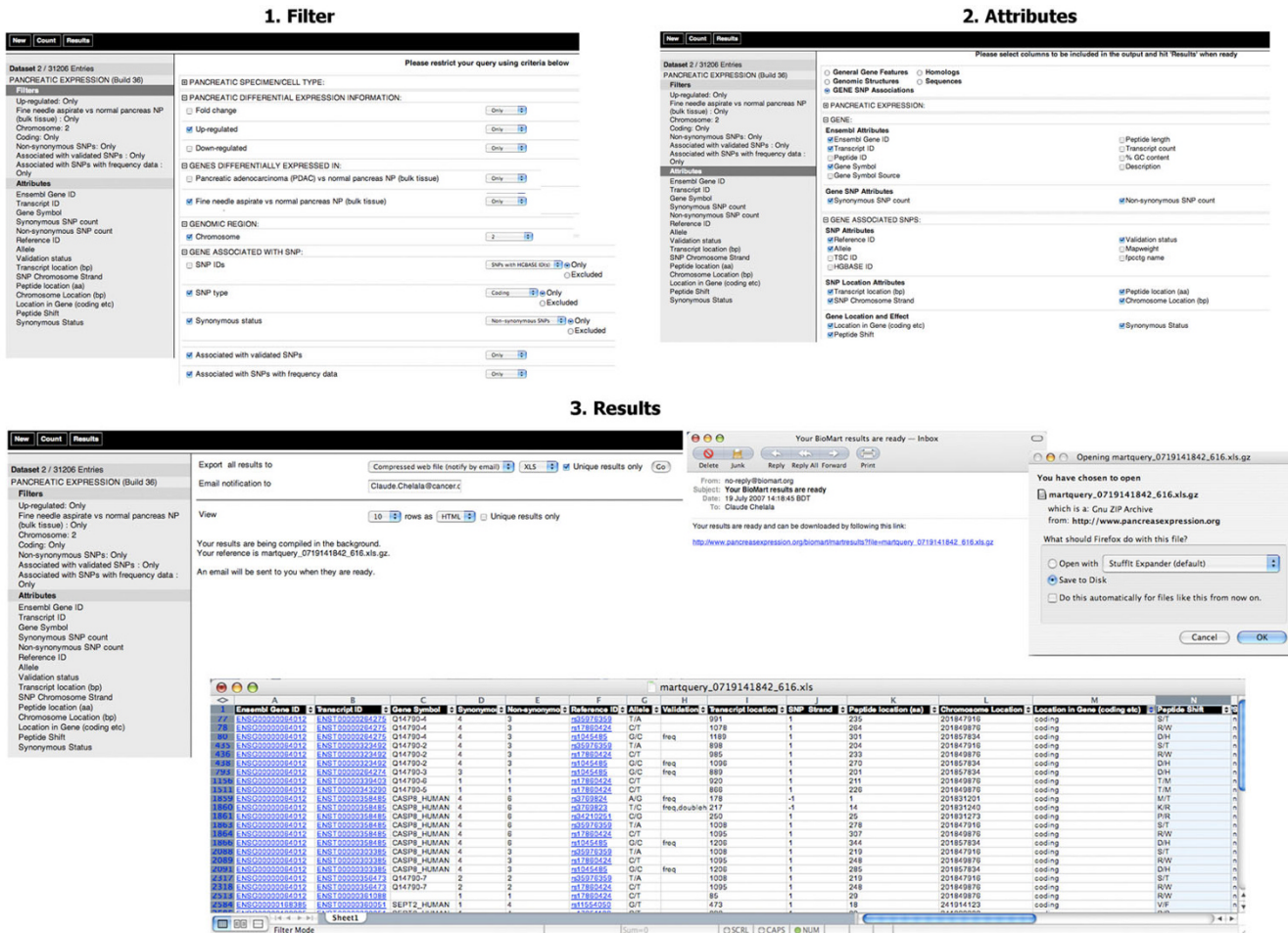
Export all results to: File | CSV | Unique results only

Filters: Intraductal papillary mucinous neoplasms (IPMN) vs cell line - Only  
Biological process: GO:0008283

View: 150 rows as HTML | Unique results only

Gene Symbol	Ensembl Gene ID	GO description
ANG_HUMAN	ENS0000018194	DNA binding
ANG_HUMAN	ENS0000018194	actin binding
ANG_HUMAN	ENS0000018194	peptidomimetic activity
ANG_HUMAN	ENS0000018194	pancreatic nonoblastase activity
ANG_HUMAN	ENS0000018194	ribonuclease activity
ANG_HUMAN	ENS0000018194	receptor binding
ANG_HUMAN	ENS0000018194	zinc ion binding
ANG_HUMAN	ENS0000018194	hydrolase activity
ANG_HUMAN	ENS0000018194	ATP binding
ANG_HUMAN	ENS0000018194	angiogenesis
ANG_HUMAN	ENS0000018194	normal female development
ANG_HUMAN	ENS0000018194	response to hypoxia
ANG_HUMAN	ENS0000018194	pancreatic adenocarcinoma
ANG_HUMAN	ENS0000018194	positive regulation of endothelial cell proliferation
ANG_HUMAN	ENS0000018194	ribosyltransferase activity
ANG_HUMAN	ENS0000018194	oblastase biosynthesis
ANG_HUMAN	ENS0000018194	cell communication
ANG_HUMAN	ENS0000018194	phenylethanolamine O-methylation
ANG_HUMAN	ENS0000018194	ATP binding
ANG_HUMAN	ENS0000018194	negative regulation of protein biosynthesis

**Figure 9**  
**Ontology for pancreatic cancer genes.** The goal of this example was to obtain the gene ontology classification of the genes differentially expressed in intraductal papillary mucinous neoplasms (IPMN) and/or mine the retrieved genes for cell proliferation process using the ontology term GO:0008283. 1. Filter: one can restrict the query by clicking on the "Filters" node on the left and selecting filters on the right. A. restriction to a cancer subtype comparison, for example by expanding the filter section "GENES DIFFERENTIALLY EXPRESSED IN:" and selecting for genes differentially expressed in "Intraductal papillary mucinous neoplasms (IPMN) vs cell line" B. restriction to a cancer subtype comparison and cell proliferation biological process "GO:0008283" by expanding the filter section "GENES DIFFERENTIALLY EXPRESSED IN:" and selecting for genes differentially expressed in "Intraductal papillary mucinous neoplasms (IPMN) vs cell line", expanding the filter section "GENE ONTOLOGY:", selecting "Biological process" and typing in "GO:0008283". 2. Attributes: one can select attributes by clicking on the "Attributes" node on the left, choosing the "General Gene Features" attribute page, expanding the attribute section "PANCREATIC EXPRESSION" to remove "Differential Expression Analysis" and "Direction of Regulation" default options, expanding the attribute section "GENE" and selecting "GO ID" under the "GO Attributes". 3. Results: one can preview the results by clicking the "Results" button on the toolbar. By default this shows the first 10 results for A and B in hyper-linked HTML. The number and format can be altered using the drop downs above this preview.



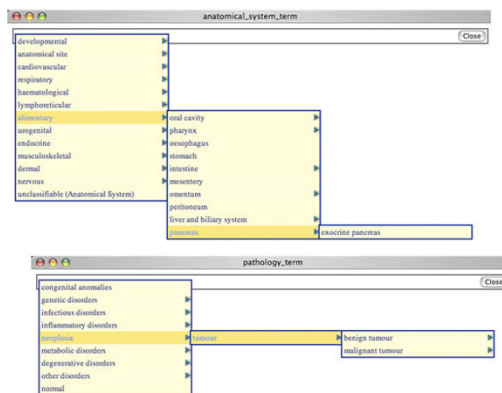
**Figure 10**  
**Design of pancreatic cancer association studies.** The goal of this example was to target SNPs for pancreatic cancer association studies by selecting for coding non-synonymous, validated SNPs with frequency data associated with genes involved in pancreatic cancer. 1. Filter: one can restrict the query by clicking on the "Filters" node on the left and selecting "Up-regulated" from the filter section "PANCREATIC DIFFERENTIAL EXPRESSION INFORMATION:", "Fine needle aspirate vs normal pancreas NP (bulk tissue)" from the filter section "GENES DIFFERENTIALLY EXPRESSED IN:" and "Chromosome 2" from the filter section "GENOMIC REGION". For SNP filters, one can expand the filter section "GENE ASSOCIATED WITH SNP:", select "SNP type" as "coding" from its drop down menu, "Synonymous status" as "Non-synonymous SNPs" from its drop down menu as well as "Associated with validated SNPs" and "Associated with SNPs with frequency data". 2. Attributes: one can click on the "Attributes" node on the left and start choosing attributes on the right from the "Gene SNP associations" attribute page. In this example, one can expand the attribute section "PANCREATIC EXPRESSION" to remove "Differential Expression Analysis" and "Direction of Regulation" default options, expand the attribute section "GENE" to select "Transcript ID" under the "Ensembl Attributes", "Synonymous SNP count" and "Non-synonymous SNP count" under "Gene SNP Attributes", expand the "GENE ASSOCIATED SNPS:" section to select the "Reference ID", "Allele" and "Validation status" under the "SNP Attributes", "Transcript location (bp)", "SNP Chromosome Strand", "Peptide location (aa)" and "Chromosome Location (bp)" under the "SNP Location Attributes:", and finally, "Location in Gene (coding etc)", "Peptide Shift" as well as "Synonymous Status" under the "Gene Location and Effect". 3. Results: one can preview the results by clicking the "Results" button on the toolbar. For particularly intensive queries, which is true in this example, one can use the web file options by entering an email address. An email will then be sent to the given address when the results are ready and one can click on the link in the email to download the results. In this example, results are viewed and filtered with Excel.



### 1. Filter

Please restrict your query using criteria below

- PANCREATIC SPECIMEN/CELL TYPE:
- PANCREATIC DIFFERENTIAL EXPRESSION INFORMATION:
- GENES DIFFERENTIALLY EXPRESSED IN:
  - Pancreatic adenocarcinoma (PDAC) vs normal pancreas NP (bulk tissue) Only
  - Neuroendocrine tumours (PEN) vs islet cell Only
- EXPRESSION PATTERNS:
  - eGenetics/SANBI EST anatomical system data pancreas browse
  - eGenetics/SANBI EST development stage data browse
  - eGenetics/SANBI EST cell type data browse
  - eGenetics/SANBI EST pathology data tumour browse
  - GNF anatomical system data browse
  - GNF development stage data browse
  - GNF cell type data browse



### 2. Attributes

Please select columns to be included in the output and hit 'Results' when ready

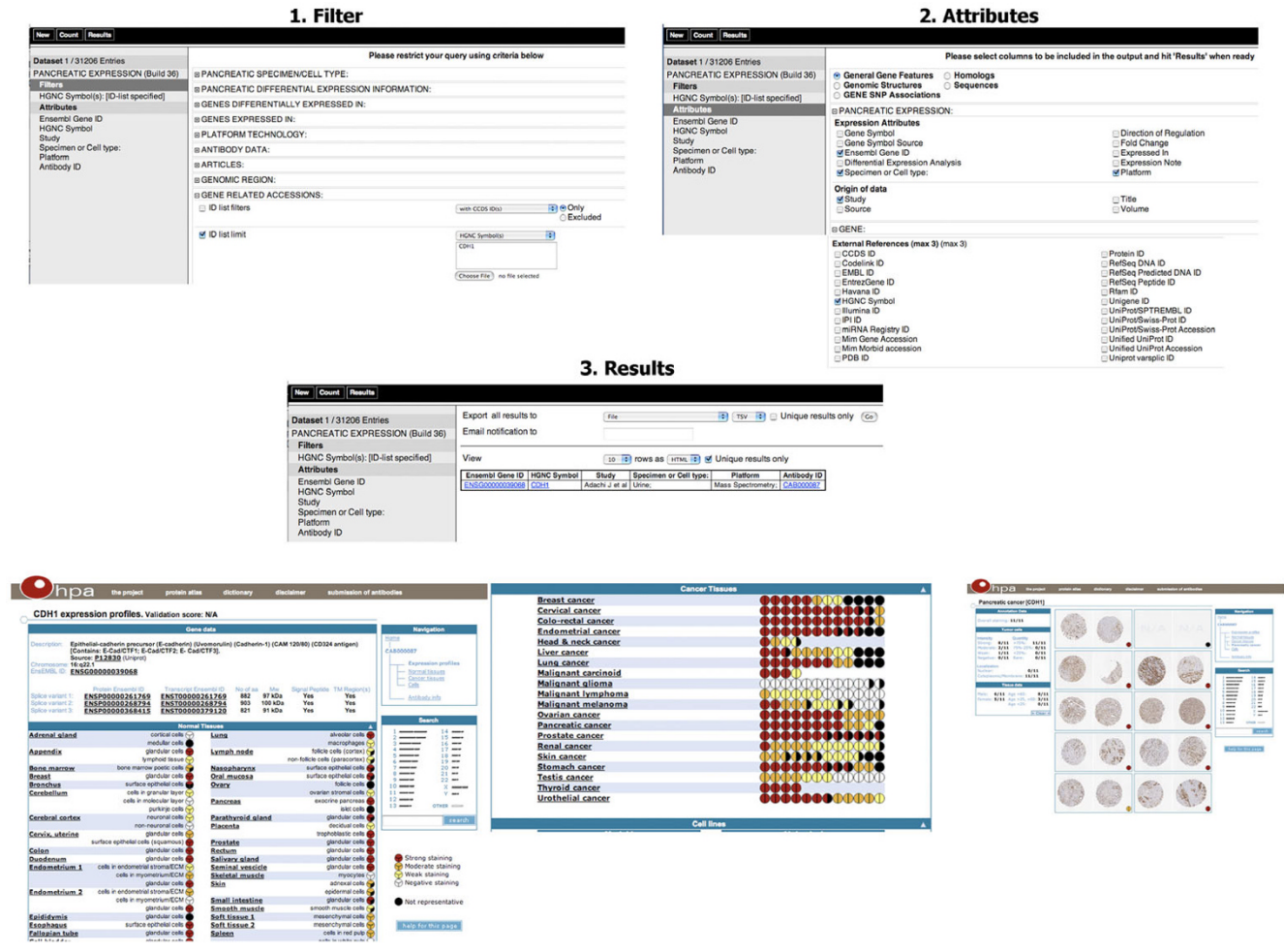
- General Gene Features
- Homologs
- Genomic Structures
- Sequences
- GENE SNP Associations
- PANCREATIC EXPRESSION:
  - Expression Attributes
    - Gene Symbol
    - Gene Symbol Source
    - Ensembl Gene ID
    - Differential Expression Analysis
    - Specimen or Cell type:
  - Direction of Regulation
  - Fold Change
  - Expression In
  - Expression Note
  - Platform
- Origin of data
  - Study
  - Source
  - Title
  - Volume
- GENE:
- REGION:
- EXPRESSION:
  - Source of Expression Data
    - eGenetics/SANBI
    - GNF (AFFY HG U95)
    - EST Profile Unigene

### 3. Results

Gene Symbol	Ensembl Gene ID	eGenetics/SANBI
CN21_HUMAN	ENSG00000298983	ANAT SYSTEM Anatomical System -> urogenital -> genital -> female genital -> placenta DEV STAGE Development Stage -> urogenital Development Stage PATH Pathology -> infectious (Pathology) CELL TYPE Cell Type -> urogenital Cell Type
CN21_HUMAN	ENSG00000298983	Development Stage PATH Pathology -> neoplasia -> tumour -> malignant tumour -> neuroblastoma CELL TYPE Cell Type -> neuroblastoma CELL Type -> not profiled
CN21_HUMAN	ENSG00000298983	ANAT SYSTEM Anatomical System -> endocrinal (Anatomical System) DEV STAGE Development Stage -> endocrinalizable
CN21_HUMAN	ENSG00000298983	ANAT SYSTEM Anatomical System -> urogenital -> genital -> female genital -> uterus -> cervix DEV STAGE Development Stage -> urogenital -> urogenital Pathology -> neoplasia -> tumour -> malignant tumour -> carcinoma -> adenocarcinoma CELL TYPE Cell Type -> neuroblastoma CELL Type -> not profiled
CN21_HUMAN	ENSG00000298983	ANAT SYSTEM Anatomical System -> urogenital -> genital -> female genital -> uterus -> cervix DEV STAGE Development Stage -> urogenital -> urogenital Pathology -> neoplasia -> tumour -> malignant tumour -> carcinoma -> adenocarcinoma CELL TYPE Cell Type -> neuroblastoma CELL Type -> not profiled

Figure 1 |

**Pancreatic cancer gene expression.** In this example, our goal was to retrieve additional expression data from a specific anatomical site for genes differentially regulated in neuroendocrine tumours (PEN). 1. Filter: one can restrict the query by clicking on the "Filters" node on the left, expanding the filter section "GENES DIFFERENTIALLY EXPRESSED IN:" to restricted the query for genes differentially expressed in "Neuroendocrine tumours (PEN) vs islet cell". Next, one can expand the filter section "EXPRESSION PATTERNS:" and restrict the query further to genes having expression profiles in pancreas and tumour by selecting, "eGenetics/SANBI EST anatomical system data" for "pancreas" (eGenetics/SANBI EST anatomical system data -> alimentary -> pancreas) and the "eGenetics/SANBI EST pathology data" for "tumour" (eGenetics/SANBI EST pathology data -> neoplasia -> tumour). 2. Attributes: one can click on the "Attributes" node on the left and start choosing attributes on the right. In this example, one can choose the "General Gene Features" attribute page, expand the attribute section "PANCREATIC EXPRESSION" to remove "Differential Expression Analysis" and "Direction of Regulation" options selected by default, expand the attribute section "EXPRESSION:" and select the "eGenetics/SANBI" under the "Source of Expression Data". 3. Results: one can preview the results by clicking the Results button on the toolbar.



**Figure 12 Immunohistochemistry data.** In this example, the goal was to show how to obtain immunohistochemistry data for the E-cadherin gene. 1. Filter: one can restrict the query by clicking on the "Filters" node on the left, expanding the filter section "GENE RELATED ACCESSIONS:" using the "+" box, choosing the "ID" list limit simply by clicking on the left checkbox and selecting the "HGNC Symbol(s)" and typing in the E-cadherin gene symbol "CDH1". 2. Attributes: one can click on the "Attributes" node on the left and start choosing attributes from the "General Gene Features" page. In this example, one can expand the "PANCREATIC EXPRESSION:" section, check "Ensembl Gene ID", "Specimen or Cell type", "Platform" under "Expression Attributes" and "Study" under "Origin of data". Once can also expand the "GENE:" section to check "HGNC Symbol" under the "External References (max 3)" and "ANTIBODIES:" to check "Antibody ID". 3. Results: one can click the "Results" button to view the results. The antibody ID for "CDH1" is displayed with a hyper-link to the Human Protein Atlas where immunohistochemistry data for normal and cancer tissues, as well as for cell lines, can be displayed.

**Acknowledgements**

The funding for this project was obtained through FW6 EU project Mol-Diag-Paca and Cancer Research UK programme grant C355/A6253.

**Disclaimer:** Information in the Pancreatic Expression database is curated from highly relevant published pancreatic cancer papers. However, the quality and accuracy of the published data are solely the responsibility of the authors. The Pancreatic Expression database is a mining tool to the literature rather than a substitute for the experiments. We highly recommend researchers to trace the origin of the data to check if the data may comply

with their quality standards. We also recommend researchers to apply independent technologies to confirm data retrieved through our mining tool prior to integrating them into the individual research efforts.

**References**

- Schneider G, Siveke JT, Eckel F, Schmid RM: **Pancreatic cancer: basic and clinical aspects.** *Gastroenterology* 2005, **128(6):**1606-1625.
- Pancreatic Expression database** [<http://www.pancreasexpression.org>]
- MySQL** [<http://www.mysql.com>]
- BioMart** [<http://www.biomart.org>]
- Ensembl** [<http://www.ensembl.org>]

6. **Ensembl microarray probeset mapping** [[http://www.ensembl.org/info/about/docs/microarray\\_probe\\_set\\_mapping.html](http://www.ensembl.org/info/about/docs/microarray_probe_set_mapping.html)]
7. **Human Protein Atlas** [<http://www.proteinatlas.org>]
8. Van Heek NT, Maitra A, Koopmann J, Fedarko N, Jain A, Rahman A, Iacobuzio-Donahue CA, Adsay V, Ashfaq R, Yeo CJ, Cameron JL, Offerhaus JA, Hruban RH, Berg KD, Goggins M: **Gene expression profiling identifies markers of ampullary adenocarcinoma.** *Cancer Biol Ther* 2004, **3(7)**:651-656.
9. Adachi J, Kumar C, Zhang Y, Olsen JV, Mann M: **The human urinary proteome contains more than 1500 proteins, including a large proportion of membrane proteins.** *Genome Biol* 2006, **7(9)**:R80.
10. Anderson NL, Polanski M, Pieper R, Gatlin T, Tirumalai RS, Conrads TP, Veenstra TD, Adkins JN, Pounds JG, Fagan R, Lohley A: **The human plasma proteome: a nonredundant list developed by combination of four separate sources.** *Mol Cell Proteomics* 2004, **3(4)**:311-326.
11. Buchholz M, Braun M, Heidenblut A, Kestler HA, Kloppel G, Schmiegel W, Hahn SA, Luttgies J, Gress TM: **Transcriptome analysis of microdissected pancreatic intraepithelial neoplastic lesions.** *Oncogene* 2005, **24(44)**:6626-6636.
12. Crnogorac-Jurcevic T, Efthimiou E, Capelli P, Blaveri E, Baron A, Terris B, Jones M, Tyson K, Bassi C, Scarpa A, Lemoine NR: **Gene expression profiles of pancreatic cancer and stromal desmoplasia.** *Oncogene* 2001, **20(50)**:7437-7446.
13. Crnogorac-Jurcevic T, Efthimiou E, Nielsen T, Loader J, Terris B, Stamp G, Baron A, Scarpa A, Lemoine NR: **Expression profiling of microdissected pancreatic adenocarcinomas.** *Oncogene* 2002, **21(29)**:4587-4594.
14. Crnogorac-Jurcevic T, Gangeswaran R, Bhakta V, Capurso G, Lattimore S, Akada M, Sunamura M, Prime W, Campbell F, Brentnall TA, Costello E, Neoptolemos J, Lemoine NR: **Proteomic analysis of chronic pancreatitis and pancreatic adenocarcinoma.** *Gastroenterology* 2005, **129(5)**:1454-1463.
15. Crnogorac-Jurcevic T, Missiaglia E, Blaveri E, Gangeswaran R, Jones M, Terris B, Costello E, Neoptolemos JP, Lemoine NR: **Molecular alterations in pancreatic carcinoma: expression profiling shows that dysregulated expression of S100 genes is highly prevalent.** *J Pathol* 2003, **201(1)**:63-74.
16. Friess H, Ding J, Kleeff J, Fekell L, Rosinski JA, Guweidhi A, Reidhaar-Olson JF, Korc M, Hammer J, Buchler MW: **Microarray-based identification of differentially expressed growth- and metastasis-associated genes in pancreatic cancer.** *Cell Mol Life Sci* 2003, **60(6)**:1180-1199.
17. Grutzmann R, Pilarsky C, Ammerpohl O, Luttgies J, Bohme A, Sipos B, Foerder M, Alldinger I, Jahnke B, Schackert HK, Kalthoff H, Kremer B, Kloppel G, Saeger HD: **Gene expression profiling of microdissected pancreatic ductal carcinomas using high-density DNA microarrays.** *Neoplasia* 2004, **6(5)**:611-622.
18. Hu L, Evers S, Lu ZH, Shen Y, Chen J: **Two-dimensional protein database of human pancreas.** *Electrophoresis* 2004, **25(3)**:512-518.
19. Iacobuzio-Donahue CA, Maitra A, Shen-Ong GL, van Heek T, Ashfaq R, Meyer R, Walter K, Berg K, Hollingsworth MA, Cameron JL, Yeo CJ, Kern SE, Goggins M, Hruban RH: **Discovery of novel tumor markers of pancreatic cancer using global gene expression technology.** *Am J Pathol* 2002, **160(4)**:1239-1249.
20. Logsdon CD, Simeone DM, Binkley C, Arumugam T, Greenon JK, Giordano TJ, Misek DE, Kuick R, Hanash S: **Molecular profiling of pancreatic adenocarcinoma and chronic pancreatitis identifies multiple genes differentially regulated in pancreatic cancer.** *Cancer Res* 2003, **63(10)**:2649-2657.
21. Lu Z, Hu L, Evers S, Chen J, Shen Y: **Differential expression profiling of human pancreatic adenocarcinoma and healthy pancreatic tissue.** *Proteomics* 2004, **4(12)**:3975-3988.
22. Maitra A, Hansel DE, Argani P, Ashfaq R, Rahman A, Naji A, Deng S, Geradts J, Hawthorne L, House MG, Yeo CJ: **Global expression analysis of well-differentiated pancreatic endocrine neoplasms using oligonucleotide microarrays.** *Clin Cancer Res* 2003, **9(16 Pt 1)**:5988-5995.
23. Nakamura T, Furukawa Y, Nakagawa H, Tsunoda T, Ohigashi H, Murata K, Ishikawa O, Ohgaki K, Kashimura N, Miyamoto M, Hirano S, Kondo S, Katoh H, Nakamura Y, Katagiri T: **Genome-wide cDNA microarray analysis of gene expression profiles in pancreatic cancers using populations of tumor cells and normal ductal epithelial cells selected for purity by laser microdissection.** *Oncogene* 2004, **23(13)**:2385-2400.
24. Segara D, Biankin AV, Kench JG, Langusch CC, Dawson AC, Skalicky DA, Gotley DC, Coleman MJ, Sutherland RL, Henshall SM: **Expression of HOXB2, a retinoic acid signaling target in pancreatic cancer and pancreatic intraepithelial neoplasia.** *Clin Cancer Res* 2005, **11(9)**:3587-3596.
25. Shen J, Person MD, Zhu J, Abbruzzese JL, Li D: **Protein expression profiles in pancreatic adenocarcinoma compared with normal pancreatic tissue and tissue affected by pancreatitis as detected by two-dimensional gel electrophoresis and mass spectrometry.** *Cancer Res* 2004, **64(24)**:9018-9026.
26. Tan ZJ, Hu XG, Cao GS, Tang Y: **Analysis of gene expression profile of pancreatic carcinoma using cDNA microarray.** *World J Gastroenterol* 2003, **9(4)**:818-823.
27. Terris B, Blaveri E, Crnogorac-Jurcevic T, Jones M, Missiaglia E, Ruszniewski P, Sauvanet A, Lemoine NR: **Characterization of gene expression profiles in intraductal papillary-mucinous tumors of the pancreas.** *Am J Pathol* 2002, **160(5)**:1745-1754.
28. **Pancreatic Expression database web-based query interface** [<http://www.pancreasexpression.org/biomart/martview>]
29. **Bioconductor** [<http://www.bioconductor.org>]
30. Durinck S, Moreau Y, Kasprzyk A, Davis S, De Moor B, Brazma A, Huber W: **BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis.** *Bioinformatics* 2005, **21(16)**:3439-3440.
31. **R project** [<http://www.r-project.org>]
32. Hull D, Wolstencroft K, Stevens R, Goble C, Pocock MR, Li P, Oinn T: **Taverna: a tool for building and running workflows of services.** *Nucleic Acids Res* 2006, **34(Web Server issue)**:W729-32.
33. **Galaxy** [<http://main.g2.bx.psu.edu>]
34. **Pancreatic Expression database access through web services** [<http://www.pancreasexpression.org/biomart/martservice>]

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

