Database

Open Access

Pancreatic Expression database: a generic model for the organization, integration and mining of complex cancer datasets Claude Chelala^{*1}, Stephan A Hahn², Hannah J Whiteman¹, Sayka Barry¹, Deepak Hariharan¹, Tomasz P Radon¹, Nicholas R Lemoine^{†1} and Tatjana Crnogorac-Jurcevic^{†1}

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

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

* Corresponding author †Equal contributors

Published: 28 November 2007

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

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 (<u>http://creativecommons.org/licenses/by/2.0</u>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received: 8 August 2007 Accepted: 28 November 2007

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 webbased 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 userfriendly; 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 deregulated 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 I

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 an URL in a notification email that allows the query results to be downloaded. I. 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 Struc-tures", "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 Expres-sion 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



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)



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.

000		Ensembl release	45: Homo sapien	s Gene repor	t for ENSG0000038427	7				
 C C<	http://www.ensembl.org/Homo	_sapiens/geneview?gene=ENSG0	0000038427			RSS *	Q. Google			
e! Ensembl Hur	nan GeneView				Search: of Hur	man 💿 EBI 🔘 Sanger	e.g. ENSG000001	Go 39618, ENSG00000138411		
Ensembl release 45 - Jun 2007							HOME BLAST	BIOMART SITEMAP HELP		
Your EnsEMBL	Ensembl Gene Re	port for ENSG0000003	8427							
Login or Register About User Accounts	Gene	VCAN (HGNC Symbol) This gene is a member of the H	uman CCDS set: <u>CCI</u>	To view all E DS4060	nsembl genes linked to the nam	ne <u>click here</u> .				
ENSG0000038427	Ensembl Gene ID	ENSG0000038427								
Gene Information Gene regulation info. Genomic sequence alignment Gene spiles site image Gene tree info. Cone variation info. LD info LD info	Genomic Location	This gene can be found on Chro The start of this gene is located	gene can be found on Chromosome 5 at location <u>82,803,339-82,912,737</u> . start of this gene is located in <u>Contig AC016568.5.1.220932</u> .							
	Description	Versican core protein precurso Source: Uniprot/SWISSPROT P13611	r (Large fibroblast p	oroteoglycan)	(Chondroitin sulfate protec	oglycan core protein 2) (PG-M) (Glial hyaluronate	-binding protein) (GHAP).		
	Prediction Method	Genes were annotated by the E by an ORF prediction. GeneWise Genome Res. 2004 14:942-50.	nsembl automatic a e/Exonerate models)	nalysis pipelin are further co	e using either a GeneWise. In the state of t	/Exonerate model from ned cDNAs to annotat	n a database protein or a s e UTRs (For more informat	et of aligned cDNAs followed ion see V.Curwen et al. ,		
 ID history Compare SNPs in transcript Transcript information Exon information Protein information Export gene data 	Transcripts	ENST0000265077 ENST00000342785 ENST00000343200 ENST00000348134	ENSP00000265 ENSP00000342 ENSP00000345 ENSP00000345 Features ▼	5077 2768 0062 5946	<u>CSPG2_HUMAN</u> <u>P13611-3</u> <u>P13611-2</u> <u>P13611-4</u>	[Transcript info [Transcript info [Transcript info [Transcript info] [Exon info]] [Exon info]] [Exon info]] [Exon info]	[Peptide info] [Peptide info] [Peptide info] [Peptide info]		
Chromosome 5 82,803,339 - 82,912,737			Chr. 5 Length	82.80 Mb Forward stra P13611- Ensemb	82.82 Mb 82.84 Mb d 12 4 > Known Protein Coding	82.88 Mb 82.88 M 29.40 Kb	b 82.00 Mb 82.0			
Graphical view Graphical overview Export information about region Export sequence as FASTA Export EMBL file Export GMB file Export SNP info in region Export SNP info in region			Ensembl trans. DNA(contigs)	P13611- Ersemb P13611- Ersemb CSPG2 Ersemb	3 > Known Protein Coding 2 > Known Protein Coding HUMAN > Known Protein Coding Known Protein Coding					
Ensembl Archive			Length	82.80 Mb	12 82.82 Mb 82.84 Mb	82.86 Mb 82.88 M	Reverse strand b 82.90 Mb 82.1			
e! View previous release of	+ Alignments	To show this information click t	he + to the left							
All Stable Archive! link for this page	Orthologue Prediction	Fo show this information click the + to the left								
	Paralogue Prediction	To show this information click the + to the left								
≶ sanger 👔 👔	Gene DAS Report									
Sea lamprey Petromyzon marinus	Pancreatic Expression	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 Fine needle aspirates (FNA) / 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 Fine needle aspirates (FNA) / Normal pancreas (normal ductal cells) (both microdissected) PANCREATIC EXPRESSION (Build 36) Up-regulated Manual carcinoma / Normal ducdenum								
New in Ensembl Pres	DAS Sources	AlfSplice (Alternative splice AlfSrlice (Alternative splice AlfSrlice (Alternative splice) AlfSrlice (Alternative splice) AlfSrlice (Alternative Trans And Splice) Add Splice AlfSrlice (Alternative Splice) AlfSrlice (Alternat	ARCREATC Extraction (10 a) 50, uprogramme and relate additional (UNC) / Rotmal parcreas (normal ductal cells) (60th Microdissected) ARCREATC EXPRESSION (alloid 36) Up-regulated Ampullary Carcinoma / Normal ducdenum AltSplice (Alternative splice database) AltTans (Alternative splice database) GAD (Genetic Association Database) GAD (Genetic Association Database) GAD (Genetic Association Database) HAGO text (PubMed text-mining via HGNC symbol) DMA (Orthology Prediction from ETH Zurich) # Pancreatic Expression (Pancreatic Expression Database) Phenotypes (Associated directly or via orthologues or protein families) Protonet; (Global classification of proteins into hierarchical clusters) R22D servif, cDMA (R2PD sequence verified non-redundant cDNA clone sets) R22D servif, cDMA (R2PD clones ready for protein expression) Reactome (Knowledgebase) UniProt (Protein knowledgebase) Quefate Manage Sources							

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

2. Attributes

	Please restrict your query using criteria below
PANCREATIC SPECIMEN/CELL TYPE:	
PANCREATIC DIFFERENTIAL EXPRESSION	N INFORMATION:
GENES DIFFERENTIALLY EXPRESSED IN:	
Pancreatic adenocarcinoma (PDAC) vs norm tissue)	nal pancreas NP (bulk Only :
Pancreatic adenocarcinoma (PDAC) vs norm (microdissected normal ductal cells)	nal pancreas ND Only
Chronic pancreatitis (CP) vs normal pancrea normal ductal cells)	s ND (microdissected Only :
Chronic pancreatitis (CP) vs normal pancreatities	s NP (bulk tissue) Only
Chronic pancreatitis (CP) vs pancreatic ader	nocarcinoma (PDAC) Only
Pancreatic adenocarcinoma (PDAC) vs chro (bulk tissue)	nic pancreatitis (CP) Only
GENES EXPRESSED IN:	
Normal pancreas ND (microdissected nor	mal ductal cells) Only
✓ Plasma	Only
Urine	Only

Please select colur	nns to be included in the output and hit 'Results' when ready
General Gene Features O Homologs Genomic Structures O Sequences GENE SNP Associations	
PANCREATIC EXPRESSION:	
Expression Attributes Gene Symbol Source Gene Symbol Source Sembli Gene ID Differential Expression Analysis Specimen or Cell type:	Direction of Regulation Fold Change Expressed In Expression Note Platform
Origin of data □ Study □ Source	□ Title □ Volume
GENE:	
Ensembl Attributes Transcript ID Peptide ID CDS length cDNA length Peptide length	 □% GC content ☑ Description □ Biotype □ Source □ Status

3. Results

New Count Results		XML Perl Help					
Dataset 37 / 31206 Entries PANCREATIC EXPRESSION (Build 36) Filters	Export all results to File Email notification to	E TSV E Unique results only					
Pancreatic adenocarcinoma (PDAC) vs normal pancreas NP (bulk tissue) : Only	View 10 rows a	IS HTML : VIIque results only					
Pancreatic adenocarcinoma (PDAC) vs	Ensembl Gene ID	Description					
chronic pancreatitis (CP) (bulk tissue) :	ENSG00000157601 Interferon-induced GTP-binding protein M [Source:Uniprot/SWISSPROT;Acc:P2059	57601 Interferon-induced GTP-binding protein Mx1 (Interferon-regulated resistance GTP-binding protein MxA) (Interferon-induced protein p78) (IFI-78K). [Source:Uniprot/SWISSPROT;Acc:P20591]					
Placma : Only	ENSG00000122641 Inhibin beta A chain precursor (Activin beta A chain) (Erythroid differentiation protein) (EDF). [Source:Uniprot/SWISSPROT;Acc:P08476]						
Flasma . Only	ENSG00000146674 Insulin-like growth factor-binding protein 3	precursor (IGFBP-3) (IBP-3) (IGF-binding protein 3). [Source:Uniprot/SWISSPROT;Acc:P17936]					
Attributes	ENSG00000106366 Plasminogen activator inhibitor 1 precurso	xr (PAI-1) (Endothelial plasminogen activator inhibitor) (PAI). [Source:Uniprot/SWISSPROT:Acc:P05121]					
Ensembl Gene ID	ENSG00000173636 Trypsin-1 precursor (EC 3.4.21.4) (Trypsin	 I) (Cationic trypsinogen). [Source:Uniprot/SWISSPROT:Acc:P07477] 					
Description	ENSG00000115414 Fibronectin precursor (FN) (Cold-insoluble globulin) (CIG). [Source:Uniprot/SWISSPROT;Acc:P02751]						
	ENSG00000115461 Insulin-like growth factor-binding protein 5	0115461 Insulin-like growth factor-binding protein 5 precursor (IGFBP-5) (IBP- 5) (IGF-binding protein 5). [Source:Uniprot/SWISSPROT;Acc:P24593]					
	ENSG00000109072 Vitronectin precursor (Serum-spreading fa [Source:Uniprot/SWISSPROT;Acc:P0400	2 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 (Source:Uniprot/SWISSPROT;Acc:P0723	5.3.4.1) (PDI) (Prolyl 4- hydroxylase subunit beta) (Cellular thyroid hormone-binding protein) (p55). 7)					
	ENSG00000117298 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 **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. I. 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

2. Attributes

New Count Results		New Count Results					
Dataset 10/31206 Entries	Please restrict your query using criteria below	Dataset 10 / 31206 Entries	Please select columns to be included in the output and hit 'Results' when rear				
PANCREATIC EXPRESSION (Build 36) Filters	EXPRESSION (Build 38) PANCREATIC SPECIMENCELL TYPE: PANCREATIC DIFFERENTIAL EXPRESSION INFORMATION:	PANCREATIC EXPRESSION (Build 36) Filters Pancreatic intraepithelial neoplasias	General Gene Features Homologs Genomic Structures Sequences GENE SNP Associations				
Panceatic intracepitelial neoplasias (Panik-2) vanima jancenas ND (microdissected normal ductal cells) : Panceatic as (Panik-3) va normal jancenas ND (microdissected normal ductal cells) : Panceatic as (Panch-3) va normal jancenas ND (microdissected normal ductal cells) : Panceatic as Panceatic as (Panceatic as Panceatic as (Panceatic a	B GENES DIFFERENTIALLY EXPRESSED IN: P Ancreatic adenocarcinoma (PDAC) vs normal pancreas NP (bulk only the second second normal ducta cells) Pancreatic adenocarcinoma (PDAC) vs normal pancreas ND only the conduct cells) Chancia expectation (PDAC) vs normal pancreas ND (incrediscented on mail ducta cells)	(PanIN-2) vs normal pancreas ND (microdissected normal ductal cells) : Only Pancreatic intraepithelial neoplasias (PanIN-9) vs normal pancreas ND (microdissected normal ductal cells) : Only Pancreatic adenocarcinome (PDAC) vs normal neorceas ND Pink Issue) : Only	PANCREATIC EXPRESSION: Expression Attributes Gene Symbol Gene Symbol Gene B General Decession Analysis Specimero Cell type:	Direction of Regulation Fold Change Expressed In Expression Note Platform			
	normal ductal cells) Chronic pancreatils (CP) vs normal pancreas NP (bulk tissue) Chronic pancreatils (CP) vs pancreatic adenocarcinoma (PDAC) Only	Attributes Ensembl Gene ID Description	Origin of data Study Source © GENE: Ensemb/Attributes	□ Title □ Volume			
	Pancreatic Intraepithelial neoplasias (PanIN-2) vs normal pancreas Only Pancreatic Intraepithelial neoplasias (PanIN-3) vs normal pancreas Only Pancreatic Intraepithelial neoplasias (PanIN-3) vs normal pancreas Only O		Transcript ID Peptide ID CDS length Peptide length Peptide length Transcript count	 % GC content ✓ Description Biotype Source Status 			

3. Results

New Count Results	XML Perl Holp					
Dataset 10 / 31206 Entries	Export all results to File I Unique results only Co					
Filters						
Pancreatic intraepithelial neoplasias	View 50 🕫 rows as HTML 🛊 🗹 Unique results only					
(microdissected normal ductal cells) :	Ensembl Gene ID Description					
Only	ENSG00000169910 Argininosuccinate lyase (EC 4.3.2.1) (Arginosuccinase) (ASAL). [Source:Uniprot/SWISSPROT;Acc:P04424]					
Pancreatic intraenithelial neonlasias	ENSG00000142789 Elastase-3B precursor (EC 3.4.21.70) (Elastase IIIB) (Protease E). [Source:Uniprot/SWISSPROT;Acc:P08861]					
(PanIN-3) vs normal pancreas ND	ENSG0000184292 Tumor-associated calcium signal transducer 2 precursor (Pancreatic carcinoma marker protein GA733-1) (Cell surface glycoprotein Trop-2). [Source:Uniprot/SWISSPROT.Acc:P09758]					
(microuissected normal ductal cens).	ENSG00000168925 chymotrypsinogen B2 [Source:RefSeq_peptide;Acc:NP_001020371]					
Panarastia adapagarajnama (PDAC) va	ENSG00000165949 Interferon-alpha-induced 11.5 kDa protein (p27) (ISG12(a) protein). [Source:Uniprot/SWISSPROT;Acc:P40305]					
normal paperoas NP (bulk tissue) : Only	ENSG00000175535 Pancreatic triacylglycerol lipase precursor (EC 3.1.1.3) (Pancreatic lipase) (PL). [Source:Uniprot/SWISSPROT;Acc:P16233]					
normal parcreas IVF (bulk ussue) . Only	ENSG00000165862 Pancreatic lipase-related protein 2 precursor (EC 3.1.1.3). [Source:Uniprot/SWISSPROT;Acc:P54317]					
Attributes	ENSG00000196565 Hemoglobin subunit epsilon (Hemoglobin epsilon chain) (Epsilon- globin). [Source:Uniprot/SWISSPROT;Acc:P02100]					
Ensembl Gene ID Description	ENSG00000170835 Bile salt-activated lipase precursor (EC 3.1.1.3) (EC 3.1.1.3) (BAL) (Bile salt-stimulated lipase) (BSSL) (Carboxyl ester lipase) (Sterol esterase) (Cholesterol esterase) (Pancreatic lysophospholipase). [Source:Uniprot/SWISSPROT;Acc:P19835]					
	ENSG00000163993 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 DIFEREN-TIALLY EXPRESSED IN:". 2. Attributes: one can choose the "General Gene Features" attribute page, expand the attribute section "PANCRE-ATIC 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		2. Attributes			
Please restrict your q	uery using criteria below		Please select columns to be included in the output and hit 'Results' when read		
PANCREATIC SPECIMEN/CELL TYPE:			General Gene Features O Homologs		
PANCREATIC DIFFERENTIAL EXPRESSION INFORMATION:			Genomic Structures Sequences GENE SNB Accession		
GENES DIFFERENTIALLY EXPRESSED IN:			GENE SNP Associations		
 Pancreatic adenocarcinoma (PDAC) vs normal pancreas NP (bulk tissue) 	Only		PANCREATIC EXPRESSION: Expression Attributes		
 Pancreatic adenocarcinoma (PDAC) vs normal pancreas ND (microdissected normal ductal cells) 	Only		Gene Symbol Gene Symbol Source	 Direction of Regulation Fold Change Eventeed la 	
PLATFORM TECHNOLOGY:			Differential Expression Analysis	Expression Note	
Affymetrix array	Affymetrix GeneChip Human Full Length Array HuGeneFL	Only	Specimen or Cell type:	Platform	
		OExcluded	Origin of data		
SAGE	Only		⊟Study	⊟ Title	
aRNA_longSAGE	Only		Source	Volume	
E 2D PAGE	Only		GENE:		
			Ensembl Attributes		
CDNA array	Only		Transcript ID	GC content	
⊻ Oligo array	Only		CDS length CDNA length Peptide length Transcript count	Biotype Source Status	
New Count Results		3. Resu	lts	XML Peri Help	
Dataset 141 /21206 Entrice	Export all results to	File	TSV : Unique results only	0	
Dataset 1417 ST200 Entries	Email potification to				
PANCREATIC EXPRESSION (Build 36)	Email nouncation to				
Filters			The second states and the se		
cDNA array: Only Oligo array: Only	View	50 😯 rows as	HTML 💽 🗹 Unique results only		
Pancreatic adenocarcinoma (PDAC) vs	Ensembl Gene ID		Description	1	
normal pancreas NP (bulk tissue) : Only	ENSG00000160180 Trefoil factor 3 precu	rsor (Intestinal trefoil fac	ctor) (hP1.B). [Source:Uniprot/SWISSPROT;Acc:Q07654]	10 (00)	
Attributes	ENSG0000160181 Trefoil factor 2 precu	rsor (Spasmolytic polyp	eptide) (SP) (Spasmolysin), [Source:Uniprof/SWISSPHOT;Acc:QC A) (Preset concernent inducible amtein) (PNP 2), [Source:Uniprof/SWISSPHOT;Acc:QC	IS4US]	
Ensembl Gene ID	Integrin beta-2 precu	rsor (psz protein) (HP1.	A) (breast cancer estrogen-inducible protein) (PNH-2), [Source:On ion alycoproteins LEA, 1/CB3/o150.95 subunit beta) (Complement)	recentor C3 subunit beta) (CD18 antigen)	
Description	ENSG00000160255 [Source:Uniprot/SWI	SSPROT;Acc:P05107]	ion grycoproteina er re- nonerproo, ao subunit beta) (complement	receptor oo aubumit betaj (ob to antigen).	

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: I. 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 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: hiting 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.



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. I. Filter: one can restrict the query by clicking on the "*Filters*" node on the left and expanding the filter section "*GENE RELATED ACCES-SIONS:*" 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 o	query using	g criteria below				Please restrict you	r query using crite	ria below	
	PANCREATIC SPECIMEN/C	ELL TYPE	:			PANCREATIC SPECIMEN/CELL TYPE:						
	PANCREATIC DIFFERENTIA		XPRESSION INFORMATION:			B PANCREATIC DIFFERENTIAL EXPRESSION INFORMATION:						
	B GENES DIFFEBENTIALLY E	B GENES DIFFERENTIALLY EXPRESSED IN:					GENES DIFFERENTIALLY	EXPRESSED IN				
	Pancreatic adenocarcinoma	Pancreatic adenocarcinoma (PDAC) vs normal pancreas NP (bulk			•		GENES EXPRESSED IN:					
	tissue)					⊕ F	PLATFORM TECHNOLOGY	Y:				
1. Filter	Pancreatic adenocarcinoma (PDAC) vs normal pancreas ND (microdissected normal ductal cells)			Only	•	œ A	B ANTIBODY DATA:					
1. Filter	Chronic pancreatitis (CP) vs normal pancreas ND (microdissected normal ductal cells)			Only		₽A	ARTICLES:					
				Only	•	± 0	GENOMIC REGION:					
	Chronic pancreatitis (CP) vs	s normal pa	ancreas NP (bulk tissue)	Only	•	± 0	GENE RELATED ACCESS	IONS:				
				<u> </u>		GENE ONTOLOGY:			(T T T T T T T T T T T T T T T T T T T			
	Chronic pancreatitis (CP) vs	s pancreati	c adenocarcinoma (PDAC)	Only	:		Evidence code (Molecular	r function)		ISS		
	Pancreatic adenocarcinoma (bulk tissue)	Pancreatic adenocarcinoma (PDAC) vs chronic pancreatitis (CP) (hulk tiesue)			•	=	Molecular function				browse	
	Intraductal papillary muc	 Intraductal papillary mucinous neoplasms (IPMN) vs cell line 		е	Only 🛟		Evidence code (Biological	ISS 🔹				
							Biological process			GO:0008283	browse	
2. Attributes	X		Genomic Structures St Genomic Structures St GENE SNP Associations PANCREATIC EXPRESSION: Expression Attributes Gene Symbol Source Finambi Gene ID Differential Expression Analysi Specimen or Cell type: GO EXEL: GO Attributes GO ID GO ID GO ID GO ID GO ID GO ID	Iomologs Sequences			Direction of Regulation Fold Change Expressed In Expression Note Platform 30 evidence code		8			
3. Results	Totol Detailed Pancel End Dataset 692 / 31 206 Entries Emministration of the second o	port all results to meno Symbol Eno (22 HUMAN S) C22 HUMAN S) C24 HUMA	Enc. Enc.	Tsy Tsy Juliane receiption Svity oteolysis coursor protein biosy	Unique results only	٢	Cont Control Cont	Epsot al results to Email residuation to Mail results and the second to the second to	Tax (1)	Inv B Unique results only	va results only (ar	

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. I. 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". 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 "PANCRE-ATIC 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.

New Count Results			2. Attributes							
			New Count Results							
Dataset 2 / 31206 Entries PANCREATIC EXPRESSION (Build 36)	PIe	ase restrict your query using criteria below	Dataset 2 / 31206 Entries PANCREATIC EXPRESSION (Build 36) Filters	Please sel General Gene Features Genomic Structures Sequences	ect columns to be included in the output and hit 'Results' when ready					
Filters	E PANCREATIC DIFFERENTIAL EXPRESSION INFORMATION:		Up-regulated: Only	GENE SNP Associations						
Fine needle aspirate vs normal pancreas NP	Fold change	Only	Fine needle aspirate vs normal pancreas NP (bulk tissue) : Only	@ PANCREATIC EXPRESSION:						
(bulk tissue) : Only Chromosome: 2	Up-regulated	Only	Chromosome: 2 Coding: Only	G GENE: Ensembl Attributes						
Coding: Only Non-synonymous SNPs: Only			Non-synonymous SNPs: Only Associated with validated SNPs : Only	Ensembl Gene ID	Peptide length					
Associated with validated SNPs : Only	Commegnated	Cont I	Associated with SNPs with frequency data :	Peptide ID	Stranscript count					
Associated with SNPs with frequency data : Only Attributes Ensembl Gene ID Transcript ID Gene Symbol	B GENES DIFFERENTIALLY EXPRESSED IN: Pancreatic adenocarcinoma (PDAC) vs normal pancreas NP (bulk)	tissue) Only	Attributes	Gene Symbol Gene Symbol Source	Description					
	E Eine naartie aspirate us normal paperase ND (bulk tissue)		Transcript ID Gene Symbol	Gene SNP Attributes	Vioneyronymus SNP court					
			Synonymous SNP count Non-synonymous SNP count							
Synonymous SNP count Non-synonymous SNP count	E GENOMIC REGION:		Reference ID Allele	SNP Attributes						
Reference ID	Chromosome	2 0	Validation status	Reference ID	Validation status					
Validation status	E GENE ASSOCIATED WITH SNP:		SNP Chromosome Strand	TSC ID	_ fpcctg name					
Transcript location (bp) SNP Chromosome Strand	SNP IDs	SNPs with HOBASE IDIG CONTY	Peptide location (aa) Chromosome Location (bp)	HGBASE ID						
Peptide location (aa) Chromosome Location (bp) Location in Gene (coding etc)	SNP type	Coding Only	Location in Gene (coding etc) Peptide Shift Synonymous Status	SNP Location Attributes @Transcript location (bp) @SNP Chromosome Strand	# Peptide location (aa) # Chromosome Location (bp)					
Peptide Shift Synonymous Status	Synonymous status	Non-synonymous Shihs 10 @ Only O Excluded		Gene Location and Effect Subscription in Gene (coding etc) Peptide Shift	Synonymous Status					
	S Associated with validated SNPs	Only 10								
	Associated with SNPs with frequency data	Only								
Dataset 2 / 31206 Entries PANCREATIC EXPRESSION (Build 36) Filters Up-regulated: Only	Export all results to Compressed Email notification to Claude.Cheli	web file (notify by email) 👔 🔀 XLS 📑 🗹 Unique results	s only Go Delete Junk Reply Reply A From: no-reply@biomatr.org Subject: Your BioMatr results are re Date: 19 July 2007 14:18:45 BDT To: Claude Chalable	ady	Vos have chosen to gen martquery.0719141842,616.xls.gz which is a: Gna 2P Archive from: http://www.pancreasexpression.org What should Firefox do with this file? Open with Stuffit Expander (default) @ Save to Disk Do this automatically for files like this from now or					
Free needle aspirate vs normal pancreas NP (buik tissue): Only Oriomosome: 2 Non-synonymous SNPs: Only Associated with validated SNPs : Only Associated with SNPs with frequency data : Only Associated with SNPs with frequency data : Attributes Ensembl Gene ID Transcript ID	View <u>10 8</u> row Your releases are being compiled in the background. Your reference is markquery, 0719116142, 516 x.H.g. An email will be sent to you when they are ready.	s as <u>iππκ i9</u> Ο Unique results only	You result a une south of white you have be done You result a une result a une result and the done http://www.pancreasexpression.org/bio	nibaded by following the line: machinades subs?file-marquery, 0713141842, 616xis gz	which is a: Cur VArchive from: http://Www.pancreasexpression.org What should Firefox do with this file? Open with Stuffit Expander (default) @ Save to Disk Do this automatically for files like this from now on.					
From needed aspirate vs normal panciesas NP Directionaisana Coding: Only Nonsyrroommoa: SIMPs: Osly Associated with SIMPs with frequency data : Only Attributes ame (ID Grans SIMPs) Granscript ID Grans SIMP Count Reference ID Reference ID Atelia	Ver Lo 1 of the second	s as <u>Litter, is</u> _ Unique results only	No read with an endoy and can be doe http://www.patoresensoresion.org/bio	nibaded by following tha line: marthadresula?Newnanguery, 0713141842, 616xis gz	which is a: Con Yarchive from: http://www.parcreasespression.org What should Firefox do with this file? Open with Suffit Expander (default) @ Save to Disk Do this automatically for files like this from now on Cancel					
Frien reedel aspirate vis normal panciesas NP (por fissue) 20 Coding: Only Non-groopmous SNPs: Only Association with validated Stars Only Association with validated Stars Only Astrobutes Ensemble Gene ID Gave Symbol Synonymous SNP count Non-synonymous SNP count Non-synonymous SNP count Atribio Validation status Transcript Location (bp)	Vere Vere Image: The set of the set	e as <u>Little (3</u> Unique results only	The called list in the day and calls be day the call and make and the day and calls be day the days and the day an	nbaded by following this link: machadrosuba/filewinarquery, 0719141642, 016.xis.gz	which is a: Contract Archive from: http://www.pancreasexpression.org What should Firefox do with this file? Open with Stuffit Expander (default) @ Save to Disk Do this automatically for files like this from now on Cancel					
Frie neede aspirate vis normal panciess NP (buck fissue) Coty Coding: Only Non-synonymous SNPs: Only Associated with validated SNPs: Only Associated with validated SNPs: Only Associated with validated SNPs: Only Attributes Erreneth Gene ID Transcript ID Gene Symbol: SNP count Synonymous SNP count Parlemence ID Allele Validation status SNP Count Bank Allele Validation status SNP Chromosome Strand Peptide location (tag)	Vere Vere Vour results are being complete in the backgourd. Vour research are being complete in the backgourd. An email will be sent to you when they are ready.		martquery_0719141842_616.xks	nikaded by billowing the link: marchadresuld.thermarca.edy. 0713141642_016486.gz	which is a: Carl VP Archive from: http://www.gancreasespression.org What should Firefox do with this file? Open with Suffit Expander (default) Save to Disk Do this automatically for files like this from now on Cancel					
From needle aspicete vs momal panciesas NP (profilingia) 20 Coding: Cohy Non-synonymous SNPs: of Sty Associated with SNPs with frequency data : Cry Attributes Ensembl Gito ID Gans Symbol Synonymous SNP count Non-synonymous SNP count Non-synonymous SNP count Non-synonymous SNP count Attributes Ensemble Count Attributes File Count Non-Synonymous SNP count Counts NP counts Non-Synonymous SNP counts Association (hg) NP Counts SNP counts NP Counts SNP coun	Vere Vere Vere Vere Vor results are being complete in the background. Vor release is mailquery, CPT9442, Biska ga. An email will be sent to you when they are ready.	C Department : C Depa	martquery_0719141842_616.xls	nikaded by billowing the line: marthandreuds https://doi.org/10.1141.642_016.xis.gz	which is a: Con UP Archive from: http://www.pancreasespression.org What should Firefox do with this file? Open with Stuffit Expander (default) Save to Disk Do this automatically for files like this from now on. Cancel Cancel					
Frien neede aspirate vis normal panciess NP (buck fissue) Coly Coding: Only Non-synonymous SNPs: Only Associated with validated SNPs: Only Associated with validated SNPs: Only Associated with validated SNPs: Only Ensemblic Gene (D) Transcript Loop and SNP count Non-synonymous SNP count Reference (D) Alles Alles Transcript Loopling (SNP Count Participation (SNP Count) Participation (SNP Cou	Vere Vere Vere Vere Vour results are being complete in the backgound. Vour releance is mantgoury, 07194142, 86 kt stig at. An email will be sent to you when they are ready.	c Disper results only C D E F C D E F 2 C D E F 2 C D E F 2 C S 0 2 2 C S 0 2 2 C S 0 2 2 C S 0 2 2	martquery, 0719141842,616.xls	nkaded by billowing the link: markhadreuds/file-markadery, 0719341642, 016x8.pz	which is a: Ciru JP Archive from: http://www.pancreasespression.org What should Firefox do with this file? Open with Suffic Expander (default) Cancel Cancel Suffic Expander (default) Up of the Suffic Expander (default) Open with Suffic Expander (default) Cancel Suffic Expander (default) Up of the Suffic Expand					
Frien reede aspirate vs normal panciess NP Coding: Only Non-synonymean SIMPs: of SMPs: Only Non-synonymean SIMPs: of SMPs: Only Associated with SMPs with frequency data : Only Astrobuses Foremotic Of D Gene Synonymea SMP Count Particulates SMP Controlscores SMPs of Alaba Validation (bit) SMP Controlscores SMPs of Peptide location (bit) SMP Controlscores SMPs of Peptide location (bit) Peptide location (bit) Peptide SMIt Synonymea Status	Ver Ver Vor results are being complete in the backgound. Vor releance as matiguery, 6719442, 654, as g.t. An email will be sent to you when they are ready. Termed Cites to 9 sectod 10 complete in the sectod 10 complete in th	C D E	martquery_0719141842_616.xts G H Land S and S an	Inbaded by billowing the line: matchedreuids The-matchedy, 0713141842, 816.bb p2 Unreal 2 (Organized Society on yan) 2 2014 (2014) 2014 (2014) 2015 (2014)	which is a: Contract reaspression arg Ymm: http://www.gancreaspression.org What should Firefox do with this file? Open with Stuffit Expander (default) Save to Disk Do this automatically for files like this from now on Cancel Cancel Conting and Contracting style Matter Same Save Save Save Save Save Save Save Sav					
Prior needia aspirate vs normal panciess NP Guki Issuai (50% Coding: Ohy Non-synonymous INPs: Only Associated with validated SMPs: Only Associated with validated SMPs: Only Associated with validated SMPs: Ohy Associated With Validated With Validated Associated With Validated With Validated With Validated Associated With Validated With Validated Associated With Validated With Validated Associated With Validated With Validated With Validated Associated With Validat	View Query results are being complete in the backgound. Your release is managenergy CHIPMER 26 Markage. An email will be sent to you when they are ready.	cs D E F Cold 0 <td>martquery_0719141842_616.xls</td> <td>nkaded by following the link: markinadreads Triller-Inactourcy, 07/1914/1642, 016.kts.pz</td> <td>which is a: Car UZ Marchive from: http://www.pancreasespression.org What should Firefox do with this file? Open with Suffic Exander (default) Cancel Cancel Open open with Suffic Exander (default) Open open with Suffic Exander (default)</td>	martquery_0719141842_616.xls	nkaded by following the link: markinadreads Triller-Inactourcy, 07/1914/1642, 016.kts.pz	which is a: Car UZ Marchive from: http://www.pancreasespression.org What should Firefox do with this file? Open with Suffic Exander (default) Cancel Cancel Open open with Suffic Exander (default)					
From needs aspirate vs normal panciess NP Decremensors 2 Coding: Only Non-syrroomnood SNP1: SNP1 Non-syrroomnood SNP1: SNP1 Associated with SNP1 with frequency data : Criy Atributed Gene Synrobil Synroomnood SNP count Generation (ID) Gene Synrobil SNP1 Count SNP1 SNP1 SNP1 SNP1 SNP1 SNP1 SNP1 SNP1	Ver Ver Vor results are being complete in the backgound. Your results are being complete in the backgound. Your results are being complete in the backgound. An email will be sent to you when they are ready. An email will be sent to you when they are ready. Termend Gree to 1 sector 1 to 2 to	C D E	martquery_0719141842_616.xtis	nbaded by billowing the line: matchedrawing file-matchedy, 07/1914/1942, 016/ship p2 2016/2016/2016/2016/2016/2016/2016/2016/	which is a: Carl JP Archive from: http://www.pancreasespression.org What should Firefox do with this file? Open with Stuffit Expander (default) Save to Disk Do this automatically for files like this from now on Cancel Concel Concel C					
Free needs aspirate vs normal panciess NP Book Issuei, Sofy Coding: Only Non-synonymous SNPs: Only Associated with validated SNPs: Only Associated with validated SNPs: Only Associated with validated SNPs: Only Transcript Coden ID Transcript Coden ID SNP Count SNP Count Non-synonymous SNP Count Non-synonymous SNP Count Non-synonymous SNP Count Non-synonymous SNP Count Non-synonymous SNP Count Non-synonymous SNP Count Transcript Coden (b) SNP Chromosome Stand SNP Chromosome Stand	Vere	C Unique results only	martquery_0719141842_616.xts	Naddel by following the line: markinarteauto file-markaety, 27/3341642, 615.45 pt 2016 2017 2017 2017 2017 2017 2017 2017 2017	which is a: Cour JP Archive from: http://www.pancreasespression.org What should Firefox do with this file? Open with Suffic Exampler (default) Cancel Cancel Cancel Concel					
From needed ascirict vis normal panciess NP (hordnorman), 2 Coding: Only Non-syrroymous SNP1: Coly Associated with SNPs with frequency data Crity Atributes are 10 Canes Synthol Synchrymous SNP court Reference 10 Canes SNP SNP court Reference 10 Alise Validation (hord) SNP Concoments Stratu Peptide Boht Synchrymous Status Synchrymous Status	Ver Ver Lo et al. Ver Vor results are being complete in the backgound. Vor results are being complete in the backgound. An email will be sent to you when they are ready. An email will be sent to you when they are ready.	C D E	martquery_0719141842_616.x/s martquery_0719141842_616.x/s 100,0000 000000000000000000000000000000	Noaded by billowing the line: matchedresult Trite-matchedy, 07131411642, 016416 p2 102451 2 021451 021411642, 016416 p2 2016	which is a: Car UP Archive from: http://www.pancreasespression.org What should Firefox do with this file? Open with Stuffit Expander (default) Save to Disk Do this automatically for files like this from now on Cancel					
Frie neede aspice vs normal panceas NP (both Staue) Coly Coding: Only Non-synonymous SNPs: Only Associated with validated SNPs: Only Associated with validated SNPs: Only Associated with validated SNPs: Only Attributes Ensembl Green (D Transcript Location (B) SNP Grownymous SNP Court Non-synonymous SNP Court Non-synonymous SNP Court Antancipt Location (b) SNP Grownscens Stell SNP Grownscens Stell Development SNP Grownscens Stell Development SNP Grownscens Stell Development SNP Grownscens Stell SNP Grow	Vere Vere Vere Vere Vere Vere Vere Ver		Control and the disk martquery_0719141842_616.xts G H H H H H H H H H	Noaded by following the line: machinetreeds 7/8+machaety, 27/3741842, 615486 pt	which is a: Car UP Archive from: http://www.pancreasespression.org What should Firefox do with this file? Open with Suffic Exampler (default) Cancel Can					
From needed aspirate vs normal pancinas NP (profinational) - 2 Coding: Only Non-synonymous SNPs: cb NPs Non-synonymous SNPs: cb NPs with frequency data : Only Associated with SNPs of the SNPs SNP Connotes SNP Count SNP Connotes SNP Count Peptide SNH Synonymous Status SNP Connotes SNPs Connotes SNP Count SNP Connotes SNP Count SNP Count	Ver Ver Vorresults are being complete in the backgound. Vorresults are being complete in the backgound. Vorresults are being complete in the backgound. An email will be sent to you when they are ready. An email will be sent to you when they are ready.	C D E	The Class density of the	Noaded by billowing the line: matchedresult Thermatiquery, 07/1914/1942, 0164/bp2 01/07/1914/1942, 0164/bp2 02/1914/1942, 0164/bp2 02/1914/1942, 0164/bp2 02/1914/1942, 0164/bp2 02/1914/1942, 0164/bp2 02/1914/1942, 0164/bp2 02/1914/1942, 0164/bp2 02/1914/1942, 0164/bp2 02/1914/1942, 0164/bp2 02/1914/1942, 0164/bp2 02/1914/bp2 02/	which is a: Car UP Archive from: http://www.gancreasespression.org What should Firefox do with this file? Open with Suffit Expander (default) Same to Disk Do this automatically for files like this from now on Cancel					
From needle aspicet vis normal panciess NP (bit fistud) : 2 Coding: Only Non-synonymous SNPs: Only Associated with validated SNPs: Only Associated with validated SNPs: Only Associated with validated SNPs: Only Chry Attributes Errandit Grow (D Transcript Coaling) SNP Chromosome SNP count Non-synonymous SNP count Reference SNP Chromosome SNP Validation status Transcript Coalino (bp) SNP Chromosome Status SNP Chromosome Status SNP Chromosome Status SNP Chromosome Status SNP Chromosome Status	Very Univer Very Constant of the statistication of the statistica	C D E D E	Control and the device of the devic	Noaded by following the line: machinerseuld Titler-Instagery, 27/19/1642, 615/16.pc United 2 Difference of the line of the l	which is a: Ciru JP Archive from: http://www.pancrasespression.org What should Firefox do with this file? Open with Suffit Exander (default) Cancel Cande					
From needle aspicet vs momal panciess NP Governmedon; 2 Coding: Only Non-synonymous SNPs; dSNPs; Ochy Associated with SNPs with frequency data : Only Astrobuses Frametal Control to Care Synbol SNPs on the SNPs with frequency data : Only Astrobuses SNPs Control SNPs Control SNPs	Ver	C Disput results only 2 0	The Control Hart Hart Hart Hart Hart Hart Hart Hart	Noaded by billowing the line: machine dressed in Thermatique y, 07/1914/1842, 016.nb.pz	which is a: Car UP Archive frem: http://www.pancrease.pression.org What should Firefox do with this file? Open with Staff Save to Disk Do this automatically for files like this from now on Cancel Cancel <td< td=""></td<>					
Frie neede aspice vs normal panceas NP (bit fissue) Coly Coding: Only Non-gronymous SNPs: Only Associated with validated SNPs: Only Associated with validated SNPs: Only Associated with validated SNPs: Only Attributes Ensemblic One ID Transcript Coding SNP: One ID SNP Count Non-gronymous SNP Count Non-gronymous SNP Count Revenue Validation status Transcript Coding (b) SNP Chromosome SNP Location in Gene (coding etc) Peptide SNP Snpromymous Status	View	C D E E E Control 0 </td <td>Control and the device of the devic</td> <td>Noaded by following the line: machine results (2019) 11642 - 0154</td> <td>which is a: Cour JP Archive from: http://www.pancreasespression.org What should Firefox do with this file? Open with Suffit Exampler (default) Cancel Ca</td>	Control and the device of the devic	Noaded by following the line: machine results (2019) 11642 - 0154	which is a: Cour JP Archive from: http://www.pancreasespression.org What should Firefox do with this file? Open with Suffit Exampler (default) Cancel Ca					
From needed aspirate vs mormal pancinas NP (por fissue) 2 Coding: Only Non-synonymous SNPs: dSNPs: Only Associated with SNPs with frequency data : Only Attributes Extended Col: 10 Care Synonymous SNP count Non-synonymous SNP count Non-synonymous SNP count Non-synonymous SNP count Non-synonymous SNP count Non-synonymous SNP count Non-synonymous SNP count NP Conconsciones Strata Peptide SNH: Synonymous Status	Ver V	Company Encode Encode Encode 2 0 6 0	Control are in do in or do to the dot mantquery_0719141842_615.xls c mantquery_0719141842_615.xls c r	Noaded by billowing the line: machinetresult Thermatiquery, 07/1914/1842, 016.th.pz	which is a: Ciru IP Archive from: http://www.pancreasespression.org What should Firefox do with this file? Open with Stiffit Expander (default) State to Disk Do this automatically for files like this from now on Cancel Cancel Cancel State to Disk What should Firefox do with this file?					
Free needs applied vs normal pancess NP Double staval, SON Coding: Only Non-synonymous SNPs: Only Associated with vskidated SNPs: Only Associated with vskidated SNPs: Only Associated with vskidated SNPs: Only Associated with vskidated SNPs: Only Attributes Ensembl Gene ID Transcript Coden ID SNP Count SNP Count Non-synonymous SNP Count Participation SNP Count SNP Count SNP Count SNP Construction Count SNP Count SNP Co	Very Univer Very Univer Very results are being complete in the backgound. Your results are being complete in the backgound. An email will be sent to you when they are ready. An email will be sent to you when they are ready.	C Unique results only	Control and the disk martquery_0719141842_616.xts G H H H G H G H G H	Noaded by following the line: machinetreeds 7/8+machaety, 27/3741842, 615.45 pt	which is a: Chu 2P Archive from: they / www.pancer.case.orgression.org What should Firefox do with this file? Open with Suffit Expander (default) Open with Suffit Expander (default) Do this automatically for files like this from now or Cancel Canc					

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. I. Filter: one can restrict the query by clicking on the "Filters" node on the left and selecting "Up-regulated" from the filter section "PAN-CREATIC 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 Attributes:", and finally, "Location in Gene (coding etc)", "Peptide Shift" as well a "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.

Please restrict your q	uery using crit	eria below	000	anatomical_system
PANCREATIC SPECIMEN/CELL TYPE:			developmental	•
PANCREATIC DIFFERENTIAL EXPRESSION INFORMATION:			anatomical site cardiovascular	
GENES DIFFERENTIALLY EXPRESSED IN:			respiratory haematological	
 Pancreatic adenocarcinoma (PDAC) vs normal pancreas NP (bulk tissue) 	Only		lymphoreticular alimentary	oral cavity
Veuroendocrine tumours (PEN) vs islet cell	Only		endocrine musculoskeletal	 pnarynx ocsophagus stomach
B EXPRESSION PATTERNS:			dermal	intestine
eGenetics/SANBI EST anatomical system data	pancreas	browse	nervous unclassifiable (Anatomical System)	omentum
eGenetics/SANBI EST development stage data		browse		peritoneum liver and biliary system punctuas
eGenetics/SANBI EST cell type data		browse		patholog
eGenetics/SANBI EST pathology data	tumour	browse	congenital anomalies	
GNF anatomical system data		browse	genetic disorders infectious disorders	A A 2
GNF development stage data		browse	neoplasia metabolic disorders	tumour
GNF cell type data		browse	degenerative disorders other disorders	A A

2. Attributes



3. Results

	Please select columns to be include	ed in the output and hit 'Results' when ready	New Count Results					XML Perl Help
 General Gene Features Genomic Structures GENE SNP Associations 	General Gene Features O Homologs Genomic Structures O Sequences GENE SNP Associations		Dataset PANCREATIC EXPRESSION (Build 36) Filters Neuroendocrine tumours (PEN) vs islet cell : Only	Export all results to Tex B Tex B Tex B Unique results only (a) Unique results only User (b) B Texes as (mix, B) (b) Unique results only				
PANCREATIC EXPRESSIO	N:		eGenetics/SANBI EST pathology data : tumour	Gene Symbol CN021 HUMAN	Ensembl Gene ID	ANAT SYSTEM Anatomic unclassifiable (Developme	eGenetics/SANBI cal System -> uropenital -> penital -> placenta DEV STAGE et Stagel PATH Pathology -> unclassifiable plathology/ CELL TYPE Cal Typ	E Development Stage -> ne -> unclassifiable (Cell Type)
Expression Attributes Gene Symbol Gene Symbol Source Ensembl Gene ID Differential Expression Ana Specimen or Cell type:	lysis	Direction of Regulation Fold Change Expressed In Expression Note Platform	eGeneticaSAN91EST anatomical system data: pancreas Attributes Gene Symbol Ensemb Gene ID eGenetica/SAN91	CN021_HUMAN CN021_HUMAN CN021_HUMAN	ENSCOULD18943	POCL-Pooling >= not pool ANAT 5YSTEM Anatomic Development Blagg) PAT POCL-Pooling >= not pool ANAT 5YSTEM Anatomic abut >= 31 years PATHP Anatomic Anatomic ANAT 5YSTEM Anatomic abut >= 31 years PATHP kenatinocyte POOL-Poolin	ed - of goods taxa. If pymm - unclassifiable (hostimical gystem) CKV 317AGE Development (1) shring - unclassifiable (hostimical gystem) CKV 317AGE Development (1) shring - unclassifiable - binary or hasplanet homo - is exclusion or (1) shring - ungredent - general groups - binary general covers - cencer (0) shring - ungredent - general or homo general covers - cencer (0) shring - ungredent - unclassifiable through - cenceronal - address (1) shring - ungredent - binary previous - cenceronal - address (1) shring - expension - binary - binary - binary - binary shring - expension - binary - miligrant timory - cenceronal - address (1) shring - expension - binary - binary - binary shrings - expension - binary - miligrant timory - cenceronal - address (1) - nd podel - nd podel binary	Blage → unclassFlable ELL TPE Cell Type → neutribiat STAGE Development Stage → accorna CELL TYPE Cell Type → STAGE Development Stage → accorna CELL TYPE Cell Type →
Origin of data Study Source		□ Title □ Volume						
GENE:								
REGION:								
EXPRESSION:								
Source of Expression Data eGenetics/SANBI GNF (AFFY HG U95)		EST Profile Unigene						

Figure II

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). I. 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 -> alimentary -> pancreas) and the "eGenetics/SANBI EST pathology data" for "pancreas" (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 "PAN-CREATIC 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.

1. Filter



Figure I2

Immunohistochemistry data. In this example, the goal was to show how to obtain immunohistochemistry data for the E-cadherin gene. I. Filter: one can restrict the query by clicking on the "Filters" node on the left, expanding the filter section "GENE RELATED ACCES-SIONS:" 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.
- 2. Pancreatic Expression database [http://www.pancreasexpres sion.org]
- 3. MySQL [http://www.mysql.com]
- 4. BioMart [http://www.biomart.org]
- 5. Ensembl [http://www.ensembl.org]

6. Ensembl microarray probeset mapping [http:// www.ensembl.org/info/about/docs/

microarray_probe_set_mapping.html]
7 Human Protein Atlas [http://www.protei

- 7. Human Protein Atlas [http://www.proteinatlas.org]
- Van Heek NT, Maitra A, Koopmann J, Fedarko N, Jain A, Rahman A, lacobuzio-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.
- 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.
- Anderson NL, Polanski M, Pieper R, Gatlin T, Tirumalai RS, Conrads TP, Veenstra TD, Adkins JN, Pounds JG, Fagan R, Lobley A: The human plasma proteome: a nonredundant list developed by combination of four separate sources. *Mol Cell Proteomics* 2004, 3(4):311-326.
- Buchholz M, Braun M, Heidenblut A, Kestler HA, Kloppel G, Schmiegel W, Hahn SA, Luttges J, Gress TM: Transcriptome analysis of microdissected pancreatic intraepithelial neoplastic lesions. Oncogene 2005, 24(44):6626-6636.
- Crnogorac-Jurcevic T, Etchimiou 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.
- 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.
- 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. *Gastro*enterology 2005, 129(5):1454-1463.
- 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.
- Friess H, Ding J, Kleeff J, Fenkell 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.
- Grutzmann R, Pilarsky C, Ammerpohl O, Luttges 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.
- Hu L, Evers S, Lu ZH, Shen Y, Chen J: Two-dimensional protein database of human pancreas. *Electrophoresis* 2004, 25(3):512-518.
- 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.
- Logsdon CD, Simeone DM, Binkley C, Arumugam T, Greenson 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.
- 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.

- 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.
- 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.
- 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.
- 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]
- 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]
- 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]

