Published online 18 October 2010 *Nucleic Acids Research, 2011, Vol. 39, Database issue* ***D1023–D1028***
doi:10.1093/nar/gkq937

The Pancreatic Expression database: 2011 update

Rosalind J. Cutts[1], Emanuela Gadaleta[1], Stephan A. Hahn[2], Tatjana Crnogorac-Jurcevic[1], Nicholas R. Lemoine[1] and Claude Chelala[1,*]

[1]Centre for Molecular Oncology and Imaging, Institute of Cancer and CR-UK Centre, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, Charterhouse Square, London EC1M 6BQ, UK and [2]Molecular GI-Onkologie (MGO), University of Bochum, 44780 Bochum, Germany

Received August 9, 2010; Revised September 27, 2010; Accepted September 30, 2010

ABSTRACT

The Pancreatic Expression database (PED, http://www.pancreasexpression.org) has established itself as the main repository for pancreatic-derived -omics data. For the past 3 years, its data content and access have increased substantially. Here we describe several of its new and improved features, such as data content, which now includes over 60 000 measurements derived from transcriptomics, proteomics, genomics and miRNA profiles from various pancreas-centred reports on a broad range of specimen and experimental types. We also illustrate the capabilities of its interface, which allows integrative queries that can combine PED data with a growing number of biological resources such as NCBI, Ensembl, UniProt and Reactome. Thus, PED is capable of retrieving and integrating different types of -omics, annotations and clinical data. We also focus on the importance of data sharing and interoperability in the cancer field, and the integration of PED into the International Cancer Genome Consortium (ICGC) data portal.

INTRODUCTION

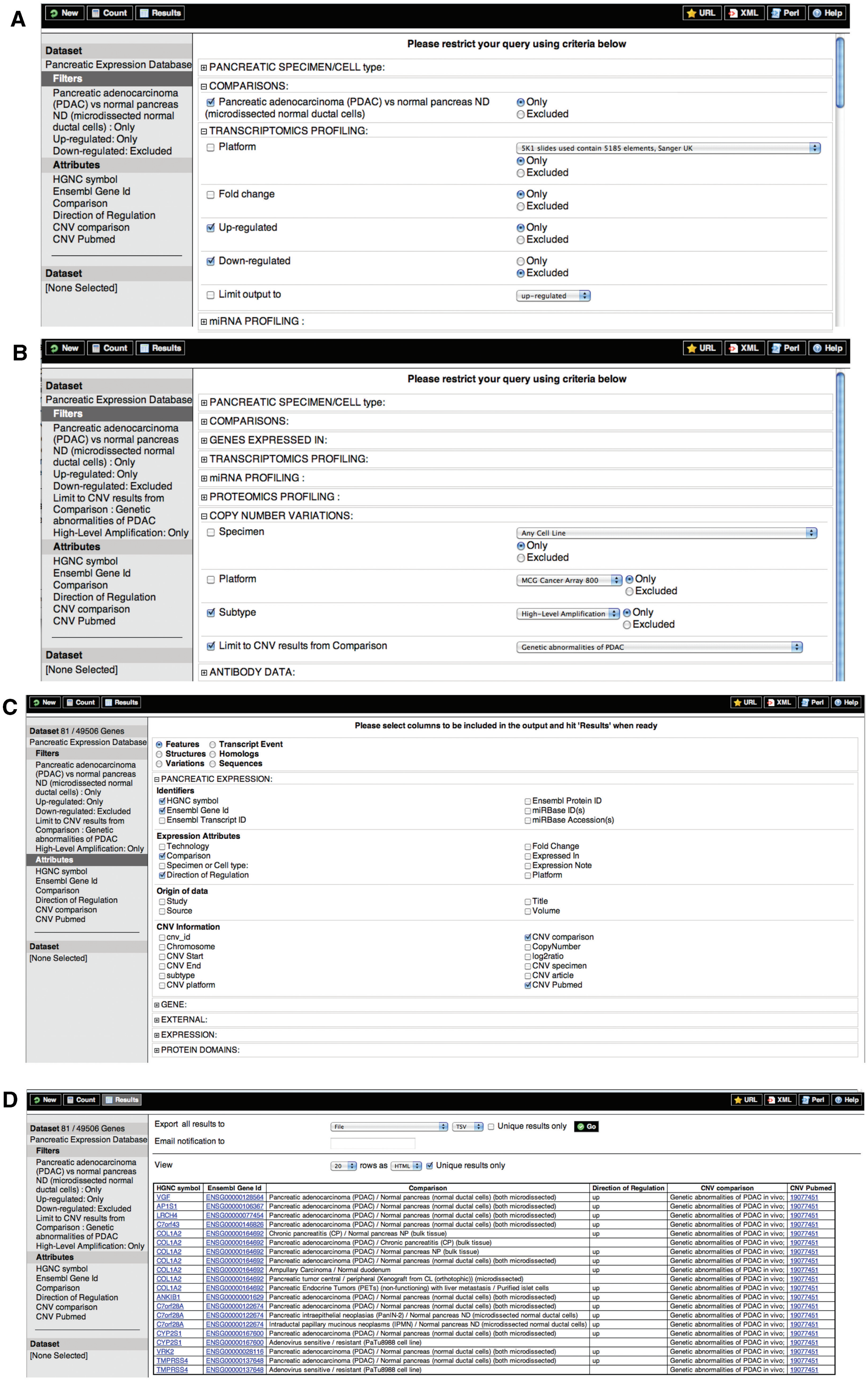
Pancreatic cancer is the fourth leading cause of cancer-related death world-wide (1) with surgical intervention and radiotherapy having a minimal impact on 5-year survival rates. As a result, patient survival rates have remained relatively unchanged over the past 30 years. Because of the poor prognosis associated with pancreatic cancer, a multitude of studies have been dedicated to elucidating the pathogenesis of this malignancy (2). Augmented by advances in high-throughput technologies, this has resulted in a plethora of -omics data. Despite the magnitude of information available, the heterogeneity and isolation of public datasets prevents researchers from effectively mining, extracting and integrating relevant data into their current research.

The publicly available Pancreatic Expression database (PED) was developed to overcome these obstacles by enabling complex pancreatic datasets to be manipulated, mined and integrated with ease (3). Since its inception in 2007, PED has undergone extensive improvements. The range of clinical and -omics data types available for queries has broadened substantially, facilitating the systematic study of pancreatic cancer.

Unlike many cancer databases specialised in providing single-type information, PED stores four different kinds of -omics data: transcriptomics, proteomics, miRNA and genomics. These profiles are derived from a broad range of specimens from tissues and body fluids of healthy people or patients, cell lines and mouse models as well as different treatments and drugs. This is key to providing a comprehensive overview of the molecular changes in cancer. Another important feature of PED is that it allows for its data to be interrogated from major gene repositories such as NCBI EntrezGene (4) and Ensembl GeneView (5), third-party software such as R statistical environment (6) and Cytoscape (7) or jointly with data from major biological resources such as the Reactome Pathway project (8), PRIDE (9) and UniProt (10). Increased functionality also allows for greater interoperability with international cancer efforts, such as the International Cancer Genome Consortium (ICGC, http://www.icgc.org), a major international collaboration designed to identify the key genetic mutations involved in up to 50 types of cancer, which will enable the development of new and better ways of diagnosing, treating and preventing cancer (11).

To the best of our knowledge, there are neither tools nor databases that provide the same information for cancer research as our platform. By allowing for integration and mining of published pancreatic cancer data in the context of a wide range of annotations, PED offers more options than raw data repositories such as

*To whom correspondence should be addressed. Tel: +44 20 7014 0474; Fax: +44 20 7014 0431; Email: c.chelala@qmul.ac.uk



A

Dataset
Pancreatic Expression Database
Filters
Pancreatic adenocarcinoma (PDAC) vs normal pancreas ND (microdissected normal ductal cells) : Only
Up-regulated: Only
Down-regulated: Excluded
Attributes
HGNC symbol
Ensembl Gene Id
Comparison
Direction of Regulation
CNV comparison
CNV Pubmed
Dataset
[None Selected]

Please restrict your query using criteria below

⊞ PANCREATIC SPECIMEN/CELL type:

⊟ COMPARISONS:
☑ Pancreatic adenocarcinoma (PDAC) vs normal pancreas ND (microdissected normal ductal cells) — ● Only ○ Excluded

⊟ TRANSCRIPTOMICS PROFILING:
☐ Platform — 5K1 slides used contain 5185 elements, Sanger UK — ● Only ○ Excluded
☐ Fold change — ● Only ○ Excluded
☑ Up-regulated — ● Only ○ Excluded
☑ Down-regulated — ○ Only ● Excluded
☐ Limit output to — up-regulated

⊞ miRNA PROFILING :

B

Dataset
Pancreatic Expression Database
Filters
Pancreatic adenocarcinoma (PDAC) vs normal pancreas ND (microdissected normal ductal cells) : Only
Up-regulated: Only
Down-regulated: Excluded
Limit to CNV results from Comparison : Genetic abnormalities of PDAC
High-Level Amplification: Only
Attributes
HGNC symbol
Ensembl Gene Id
Comparison
Direction of Regulation
CNV comparison
CNV Pubmed
Dataset
[None Selected]

Please restrict your query using criteria below

⊞ PANCREATIC SPECIMEN/CELL type:
⊞ COMPARISONS:
⊞ GENES EXPRESSED IN:
⊞ TRANSCRIPTOMICS PROFILING:
⊞ miRNA PROFILING :
⊞ PROTEOMICS PROFILING :
⊟ COPY NUMBER VARIATIONS:
☐ Specimen — Any Cell Line — ● Only ○ Excluded
☐ Platform — MCG Cancer Array 800 — ● Only ○ Excluded
☑ Subtype — High-Level Amplification — ● Only ○ Excluded
☑ Limit to CNV results from Comparison — Genetic abnormalities of PDAC
⊞ ANTIBODY DATA:

C

Dataset 81 / 49506 Genes
Pancreatic Expression Database
Filters
Pancreatic adenocarcinoma (PDAC) vs normal pancreas ND (microdissected normal ductal cells) : Only
Up-regulated: Only
Down-regulated: Excluded
Limit to CNV results from Comparison : Genetic abnormalities of PDAC
High-Level Amplification: Only
Attributes
HGNC symbol
Ensembl Gene Id
Comparison
Direction of Regulation
CNV comparison
CNV Pubmed
Dataset
[None Selected]

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

● Features ○ Transcript Event
○ Structures ○ Homologs
○ Variations ○ Sequences

⊟ PANCREATIC EXPRESSION:

Identifiers
☑ HGNC symbol
☑ Ensembl Gene Id
☐ Ensembl Transcript ID
☐ Ensembl Protein ID
☐ miRBase ID(s)
☐ miRBase Accession(s)

Expression Attributes
☐ Technology
☑ Comparison
☐ Specimen or Cell type:
☑ Direction of Regulation
☐ Fold Change
☐ Expressed In
☐ Expression Note
☐ Platform

Origin of data
☐ Study
☐ Source
☐ Title
☐ Volume

CNV Information
☐ cnv_id
☐ Chromosome
☐ CNV Start
☐ CNV End
☐ subtype
☐ CNV platform
☑ CNV comparison
☐ CopyNumber
☐ log2ratio
☐ CNV specimen
☐ CNV article
☑ CNV Pubmed

⊞ GENE:
⊞ EXTERNAL:
⊞ EXPRESSION:
⊞ PROTEIN DOMAINS:

D

Dataset 81 / 49506 Genes
Pancreatic Expression Database
Filters
Pancreatic adenocarcinoma (PDAC) vs normal pancreas ND (microdissected normal ductal cells) : Only
Up-regulated: Only
Down-regulated: Excluded
Limit to CNV results from Comparison : Genetic abnormalities of PDAC
High-Level Amplification: Only
Attributes
HGNC symbol
Ensembl Gene Id
Comparison
Direction of Regulation
CNV comparison
CNV Pubmed
Dataset
[None Selected]

Export all results to — File — TSV — ☐ Unique results only — Go
Email notification to

View — 20 rows as HTML — ☑ Unique results only

HGNC symbol	Ensembl Gene Id	Comparison	Direction of Regulation	CNV comparison	CNV Pubmed
VGF	ENSG00000128564	Pancreatic adenocarcinoma (PDAC) / Normal pancreas (normal ductal cells) (both microdissected)	up	Genetic abnormalities of PDAC in vivo;	19077451
AP1S1	ENSG00000106367	Pancreatic adenocarcinoma (PDAC) / Normal pancreas (normal ductal cells) (both microdissected)	up	Genetic abnormalities of PDAC in vivo;	19077451
LRCH4	ENSG00000077454	Pancreatic adenocarcinoma (PDAC) / Normal pancreas (normal ductal cells) (both microdissected)	up	Genetic abnormalities of PDAC in vivo;	19077451
C7orf43	ENSG00000146826	Pancreatic adenocarcinoma (PDAC) / Normal pancreas (normal ductal cells) (both microdissected)	up	Genetic abnormalities of PDAC in vivo;	19077451
COL1A2	ENSG00000164692	Chronic pancreatitis (CP) / Normal pancreas NP (bulk tissue)	up	Genetic abnormalities of PDAC in vivo;	19077451
COL1A2	ENSG00000164692	Pancreatic adenocarcinoma (PDAC) / Chronic pancreatitis (CP) (bulk tissue)		Genetic abnormalities of PDAC in vivo;	19077451
COL1A2	ENSG00000164692	Pancreatic adenocarcinoma (PDAC) / Normal pancreas NP (bulk tissue)	up	Genetic abnormalities of PDAC in vivo;	19077451
COL1A2	ENSG00000164692	Pancreatic adenocarcinoma (PDAC) / Normal pancreas (normal ductal cells) (both microdissected)	up	Genetic abnormalities of PDAC in vivo;	19077451
COL1A2	ENSG00000164692	Ampullary Carcinoma / Normal duodenum	up	Genetic abnormalities of PDAC in vivo;	19077451
COL1A2	ENSG00000164692	Pancreatic tumor central / peripheral (Xenograft from CL (orthotophic)) (microdissected)		Genetic abnormalities of PDAC in vivo;	19077451
COL1A2	ENSG00000164692	Pancreatic Endocrine Tumors (PETs) (non-functioning) with liver metastasis / Purified islet cells	up	Genetic abnormalities of PDAC in vivo;	19077451
ANKIB1	ENSG00000001629	Pancreatic adenocarcinoma (PDAC) / Normal pancreas (normal ductal cells) (both microdissected)	up	Genetic abnormalities of PDAC in vivo;	19077451
C7orf28A	ENSG00000122674	Pancreatic adenocarcinoma (PDAC) / Normal pancreas (normal ductal cells) (both microdissected)	up	Genetic abnormalities of PDAC in vivo;	19077451
C7orf28A	ENSG00000122674	Pancreatic intraepithelial neoplasias (PanIN-2) / Normal pancreas ND (microdissected normal ductal cells)	up	Genetic abnormalities of PDAC in vivo;	19077451
C7orf28A	ENSG00000122674	Intraductal papillary mucinous neoplasms (IPMN) / Normal pancreas ND (microdissected normal ductal cells)	up	Genetic abnormalities of PDAC in vivo;	19077451
CYP2S1	ENSG00000167600	Pancreatic adenocarcinoma (PDAC) / Normal pancreas (normal ductal cells) (both microdissected)	up	Genetic abnormalities of PDAC in vivo;	19077451
CYP2S1	ENSG00000167600	Adenovirus sensitive / resistant (PaTu8988 cell line)		Genetic abnormalities of PDAC in vivo;	19077451
VRK2	ENSG00000028116	Pancreatic adenocarcinoma (PDAC) / Normal pancreas (normal ductal cells) (both microdissected)	up	Genetic abnormalities of PDAC in vivo;	19077451
TMPRSS4	ENSG00000137648	Pancreatic adenocarcinoma (PDAC) / Normal pancreas (normal ductal cells) (both microdissected)	up	Genetic abnormalities of PDAC in vivo;	19077451
TMPRSS4	ENSG00000137648	Adenovirus sensitive / resistant (PaTu8988 cell line)		Genetic abnormalities of PDAC in vivo;	19077451

Figure 1. Integrated view of genomic and transcriptomic changes in pancreatic cancer. This figure integrates the results from both genome-wide DNA copy number and genome-wide gene expression profiling. In a few seconds, this allows for visual inspection of copy number-driven expression changes. Here the query is to find genes differentially up-regulated in PDAC versus normal using microdissected ductal cells (**A**) and then combine this information with data on genes that are also associated with genomic variations in PDAC samples by combining with results for copy number changes on high level amplifications (**B**). Pick attributes for display (**C**). A summary of the first 20 results is shown in (**D**).

Gene Expression Omnibus (GEO) (12) or ArrayExpress (13). One of the main strengths of our system is the possibility of setting many filters to ask very specific pancreatic cancer-related questions and obtain a focused annotated data output. Here we outline details of the improved data content, query interfaces and interoperability.

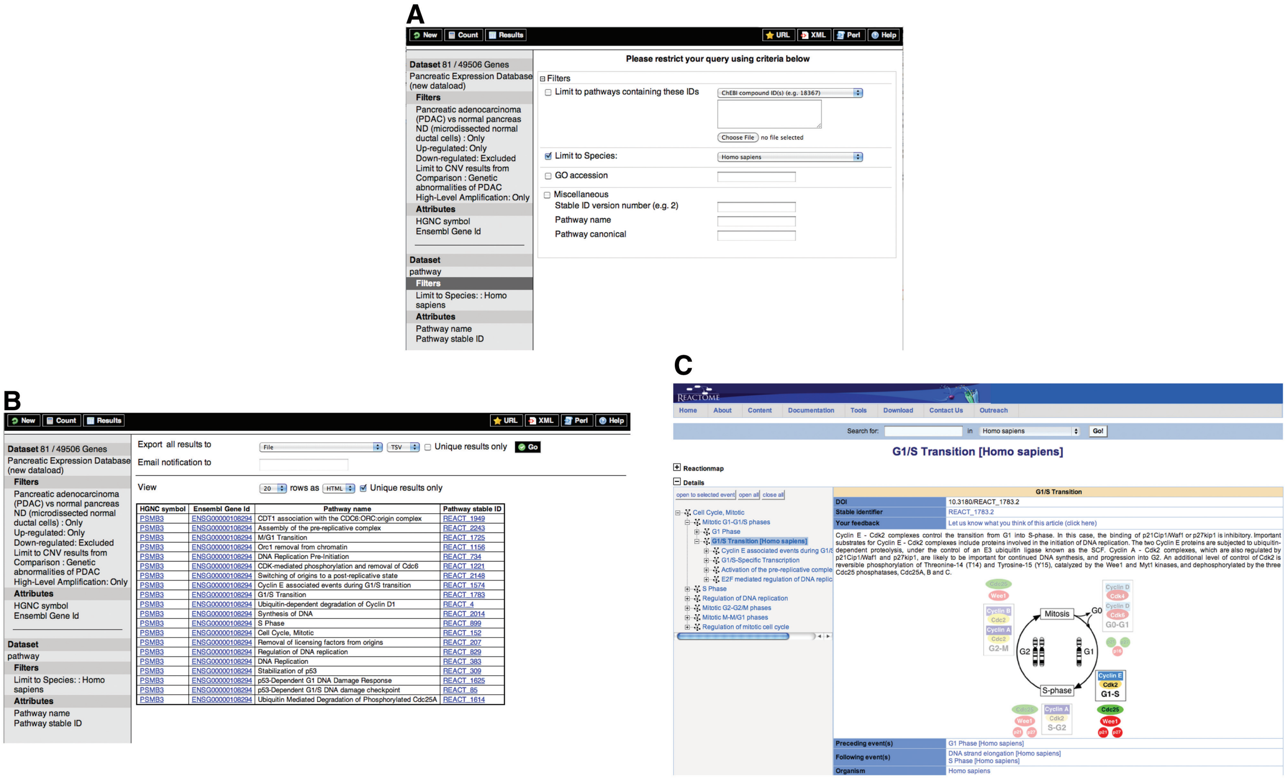
IMPROVED DATA CONTENT

The database contains 56 015 differential or expression measurements and 6363 DNA copy number alterations. These data values are extracted from over 59 published studies and include profiles from a large number of specimen types, such as pancreatic tissues and various body fluids obtained from healthy subjects and patients with benign, pre-malignant and malignant diseases. In addition, studies using pancreatic cancer cell lines and murine models have also been incorporated (Supplementary Table S1). Where applicable, information on the different treatment conditions applied is provided to users.

The collected samples were profiled on a wide range of transcriptomics, proteomics, miRNAs and genomics platforms (Supplementary Table S2). To date, PED describes pancreatic-related regulation events in 8229 genes/proteins, 27 327 transcripts and 279 miRNA as well as 2771 gains, 1073 losses, 347 homozygous deletions, 1297 high-level amplifications and 875 loss of heterozygosity events occurring in distinct genomic areas.

The data collection pipeline from the original database was enhanced to cover not only transcriptomics but also genomics, proteomics and miRNA information. Data from the primary literature were manually curated, reviewed for accuracy and consistency, and loaded into a relational database. The Ensembl annotations originally used for mapping probe identifiers and gene names to standard values (version 48) have been updated to a more current version (version 56). This ensures that our database remains up-to-date with ongoing improvements to annotation and microarray probe set mappings and helps avoid data integrity errors. The data collection process was expanded to encompass new data types such as DNA copy number changes. Here, chromosomal coordinates for each copy number variation identified from papers were mapped to genome release GRCh37 with conversions between genome versions carried out using the liftover tool from UCSC (14).

We imported the available Ensembl human gene annotations (5) for genes, proteins, SNP information, sequences, gene structure and multi-species data, enabling the integration and annotation of heterogeneous pancreatic data (Supplementary Table S3).



A

Dataset 81 / 49506 Genes
Pancreatic Expression Database (new dataload)
Filters
Pancreatic adenocarcinoma (PDAC) vs normal pancreas ND (microdissected normal ductal cells) : Only
Up-regulated: Only
Down-regulated: Excluded
Limit to CNV results from Comparison : Genetic abnormalities of PDAC
High-Level Amplification: Only
Attributes
HGNC symbol
Ensembl Gene Id

Dataset
pathway
Filters
Limit to Species: : Homo sapiens
Attributes
Pathway name
Pathway stable ID

Please restrict your query using criteria below

⊟ Filters
☐ Limit to pathways containing these IDs — ChEBI compound ID(s) (e.g. 18367) — Choose File no file selected
☑ Limit to Species: — Homo sapiens
☐ GO accession
☐ Miscellaneous
Stable ID version number (e.g. 2)
Pathway name
Pathway canonical

B

Dataset 81 / 49506 Genes
Pancreatic Expression Database (new dataload)
Filters
Pancreatic adenocarcinoma (PDAC) vs normal pancreas ND (microdissected normal ductal cells) : Only
Up-regulated: Only
Down-regulated: Excluded
Limit to CNV results from Comparison : Genetic abnormalities of PDAC
High-Level Amplification: Only
Attributes
HGNC symbol
Ensembl Gene Id

Dataset
pathway
Filters
Limit to Species: : Homo sapiens
Attributes
Pathway name
Pathway stable ID

Export all results to — File — TSV — ☐ Unique results only — Go
Email notification to
View — 20 — rows as — HTML — ☑ Unique results only

HGNC symbol	Ensembl Gene Id	Pathway name	Pathway stable ID
PSMB3	ENSG00000108294	CDT1 association with the CDC6:ORC:origin complex	REACT_1949
PSMB3	ENSG00000108294	Assembly of the pre-replicative complex	REACT_2243
PSMB3	ENSG00000108294	M/G1 Transition	REACT_1725
PSMB3	ENSG00000108294	Orc1 removal from chromatin	REACT_1156
PSMB3	ENSG00000108294	DNA Replication Pre-Initiation	REACT_734
PSMB3	ENSG00000108294	CDK-mediated phosphorylation and removal of Cdc6	REACT_1221
PSMB3	ENSG00000108294	Switching of origins to a post-replicative state	REACT_2148
PSMB3	ENSG00000108294	Cyclin E associated events during G1/S transition	REACT_1574
PSMB3	ENSG00000108294	G1/S Transition	REACT_1783
PSMB3	ENSG00000108294	Ubiquitin-dependent degradation of Cyclin D1	REACT_4
PSMB3	ENSG00000108294	Synthesis of DNA	REACT_2014
PSMB3	ENSG00000108294	S Phase	REACT_899
PSMB3	ENSG00000108294	Cell Cycle, Mitotic	REACT_152
PSMB3	ENSG00000108294	Removal of licensing factors from origins	REACT_207
PSMB3	ENSG00000108294	Regulation of DNA replication	REACT_829
PSMB3	ENSG00000108294	DNA Replication	REACT_383
PSMB3	ENSG00000108294	Stabilization of p53	REACT_309
PSMB3	ENSG00000108294	p53-Dependent G1 DNA Damage Response	REACT_1625
PSMB3	ENSG00000108294	p53-Dependent G1/S DNA damage checkpoint	REACT_85
PSMB3	ENSG00000108294	Ubiquitin Mediated Degradation of Phosphorylated Cdc25A	REACT_1614

C

Search for: in Homo sapiens

G1/S Transition [Homo sapiens]

Cell Cycle, Mitotic
Mitotic G1-G1/S phases
G1 Phase
G1/S Transition [Homo sapiens]
Cyclin E associated events during G1/S
G1/S-Specific Transcription
Activation of the pre-replicative comple
E2F mediated regulation of DNA replic
S Phase
Regulation of DNA replication
Mitotic G2-G2/M phases
Mitotic M-M/G1 phases
Regulation of mitotic cell cycle

G1/S Transition	
DOI	10.3180/REACT_1783.2
Stable identifier	REACT_1783.2
Your feedback	Let us know what you think of this article (click here)

Cyclin E - Cdk2 complexes control the transition from G1 into S-phase. In this case, the binding of p21Cip1/Waf1 or p27kip1 is inhibitory. Important substrates for Cyclin E - Cdk2 complexes include proteins involved in the initiation of DNA replication. The two Cyclin E proteins are subjected to ubiquitin-dependent proteolysis, under the control of an E3 ubiquitin ligase known as the SCF. Cyclin A - Cdk2 complexes, which are also regulated by p21Cip1/Waf1 and p27kip1, are likely to be important for continued DNA synthesis, and progression into G2. An additional level of control of Cdk2 is reversible phosphorylation of Threonine-14 (T14) and Tyrosine-15 (Y15), catalyzed by the Wee1 and Myt1 kinases, and dephosphorylated by the three Cdc25 phosphatases, Cdc25A, B and C.

Preceding event(s)	G1 Phase [Homo sapiens]
Following event(s)	DNA strand elongation [Homo sapiens] S Phase [Homo sapiens]
Organism	Homo sapiens

Figure 2. Cross-linking pathways information from reactome. Results from the previous query (shown in Figure 1) can be further mined and merged with data from Reactome. Here we select Reactome pathways as a second dataset to combine in queries and then restrict obtained information to '*Homo sapiens*' data (**A**). Pick attributes and display a summary of the first 20 results (**B**). Query results return pathway name and stable ID with a hyper-link to the Reactome website allowing to instantly extract data (**C**).

IMPROVED QUERY CAPABILITIES

While the original PED system included tools for querying across different tumour stages of pancreatic cancer and different pancreatic disease types in a simple integrated way (3), the interface has been greatly improved to provide the user access to the expanded data content. By including new platforms, comparisons and data types, it is possible to combine information and, therefore, to perform more complex queries than by viewing expression data alone. For example, it is possible to highlight the impact of copy number aberrations on gene expression patterns to filter out genes whose expression levels are not consistent with their DNA copy number status and point to the subset of candidate oncogenes or tumour suppressor genes showing copy number-driven expression changes (Figure 1). There are now options available for proteomics, transcriptomics and miRNA profiling, allowing these data types to be queried in isolation or combined to look for genes consistently identified across different data technologies or to extract specific data such as microRNAs deregulated in the different stages of pancreatic cancer.

The new interface incorporates the full functionality and data from Ensembl Mart 56 including richer data content and more advanced querying capabilities and also makes better use of the BioMart interface capabilities (15) for cross-linking different datasets that have data types in common; for example by combining PED data with Reactome pathway data of human reactions and pathways (8) (Figure 2). This integration allows researchers to quickly identify both pancreatic cancer genes and the pathways they control.

IMPROVED DATA ACCESS

PED is freely accessible through a BioMart web-based query interface at http://www.pancreasexpression.org. PED is also a DAS server (16) providing DAS annotations for the wider community, so it can be used in other resources or browsers such as Ensembl GeneView using the GeneDAS protocol. Access is also possible *via* web services through third party software tools that have been made compatible with BioMart resources such as Bioconductor (http://www.bioconductor.org) (6,17), Galaxy (18) and Cytoscape (7). Interoperability with the ICGC is also possible through this web services layer and PED is now available through the ICGC data portal (http://dcc.icgc.org) (Figure 3) allowing researchers to conduct combined queries on ICGC experimental data alongside PED literature data. This further increases the ways in which the database can be accessed and its exposure to a variety of disciplines and interests in the scientific community. The database is also accessible as a Linkout resource from NCBI EntrezGene (4). This allows EntrezGene users to be alerted to pancreatic expression data by the presence of a data link for relevant genes that are in the database.

DISCUSSION

We have described how PED has evolved from its original role as a repository for cancer transcriptomics data into a comprehensive resource capable of providing a quick overview of molecular changes at the transcriptome, proteome, genome and/or miRNA level. Consequently, there has been a huge growth in its -omics, specimen/clinical and annotation data content.

We believe that interoperability is a key factor in the utility and productive use of any current and future cancer databases. This is essential to ensure the sustainability of any cancer database and facilitate its integration with major international efforts in cancer research such as the ICGC. This will also allow the design and implementation of more sophisticated analysis portals. The cancer research community needs open source, fully interoperable resources allowing information connectivity and data sharing. Only these types of resource can ensure that cancer data generated across different organisations are shared, thereby maximising the impact of cancer research. By using the BioMart technology for its data management system, PED is fully interoperable with the ICGC. This ensures that PED is integrated with The Cancer Genome Atlas (TCGA, http://cgap.nci.nih.gov) data available through the ICGC data portal. The BioMart web service layer also allows PED to be integrated with several other data sources that also use the BioMart technology such as Reactome, PRIDE, UniProt and Ensembl. Moreover, PED is a Linkout resource integrated with NCBI.

Our database fills the urgent requirement of the pancreatic cancer community for resources capable of integrating the overflowing influx of data generated by novel high-throughput technologies.

The architectural flexibility of PED is easily extendable to other disease types, with this model being used to create a similar resource for malignant (breast cancer) and non-malignant (neurodegenerative) diseases. Reuse of a similar database design will facilitate complex query capabilities across multiple diseases and data types.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

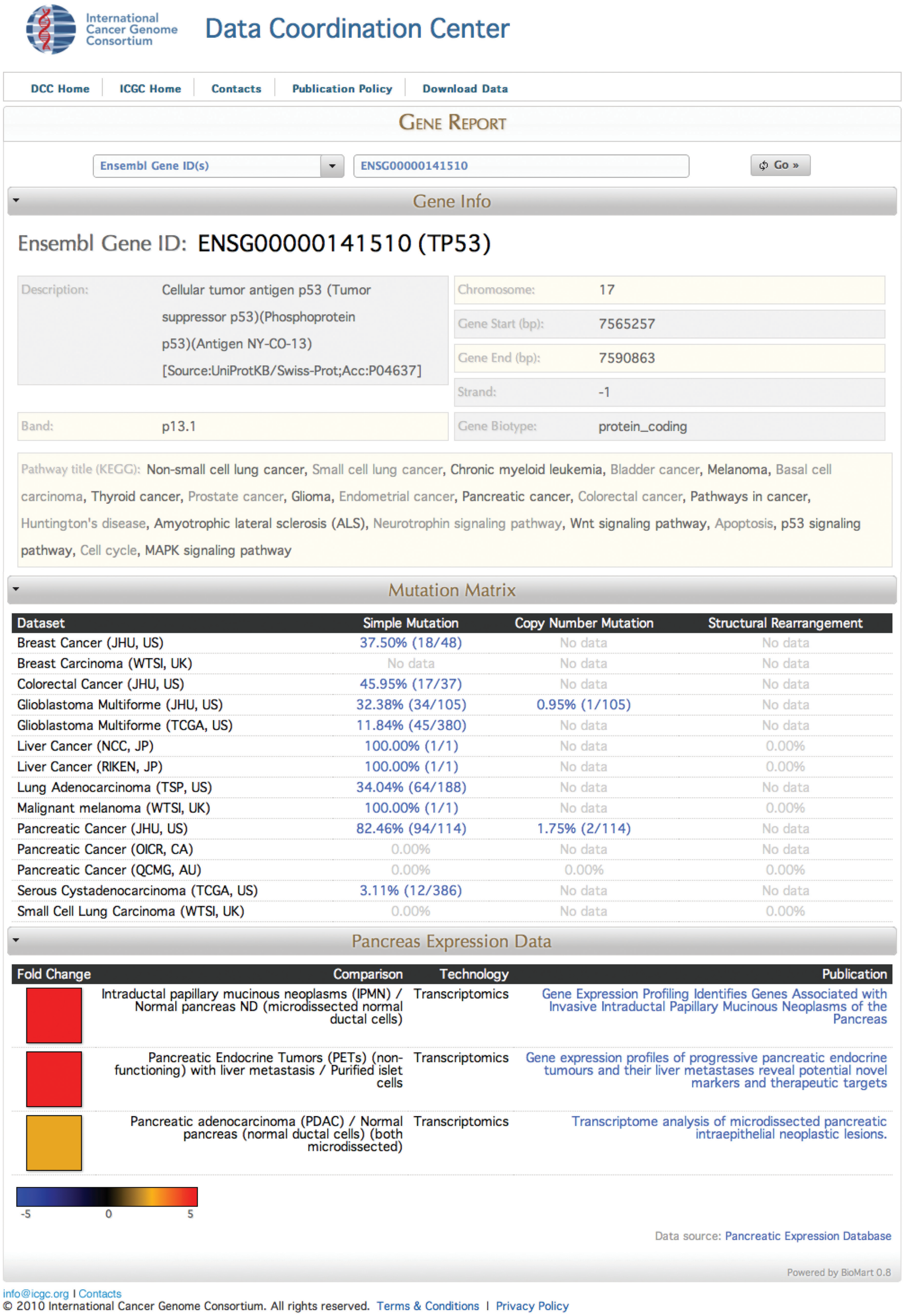
ACKNOWLEDGEMENTS

The authors thank our colleagues who have suggested or contributed data for current or previous versions of the database. They also would like to thank the swift and helpful advice from the ICGC DCC and BioMart team.

FUNDING

Cancer Research UK (programme grant C355/A6253) and FW6 EU project MolDiag-Paca. R.C. is funded by Breast Cancer Campaign. Funding for open access charge: Cancer Research UK.

Conflict of interest statement. None declared.



International Cancer Genome Consortium

Data Coordination Center

GENE REPORT

Gene Info

Ensembl Gene ID: ENSG00000141510 (TP53)

Description:	Cellular tumor antigen p53 (Tumor suppressor p53)(Phosphoprotein p53)(Antigen NY-CO-13) [Source:UniProtKB/Swiss-Prot;Acc:P04637]	Chromosome:	17
		Gene Start (bp):	7565257
		Gene End (bp):	7590863
		Strand:	-1
Band:	p13.1	Gene Biotype:	protein_coding

Pathway title (KEGG): Non-small cell lung cancer, Small cell lung cancer, Chronic myeloid leukemia, Bladder cancer, Melanoma, Basal cell carcinoma, Thyroid cancer, Prostate cancer, Glioma, Endometrial cancer, Pancreatic cancer, Colorectal cancer, Pathways in cancer, Huntington's disease, Amyotrophic lateral sclerosis (ALS), Neurotrophin signaling pathway, Wnt signaling pathway, Apoptosis, p53 signaling pathway, Cell cycle, MAPK signaling pathway

Mutation Matrix

Dataset	Simple Mutation	Copy Number Mutation	Structural Rearrangement
Breast Cancer (JHU, US)	37.50% (18/48)	No data	No data
Breast Carcinoma (WTSI, UK)	No data	No data	No data
Colorectal Cancer (JHU, US)	45.95% (17/37)	No data	No data
Glioblastoma Multiforme (JHU, US)	32.38% (34/105)	0.95% (1/105)	No data
Glioblastoma Multiforme (TCGA, US)	11.84% (45/380)	No data	No data
Liver Cancer (NCC, JP)	100.00% (1/1)	No data	0.00%
Liver Cancer (RIKEN, JP)	100.00% (1/1)	No data	0.00%
Lung Adenocarcinoma (TSP, US)	34.04% (64/188)	No data	No data
Malignant melanoma (WTSI, UK)	100.00% (1/1)	No data	0.00%
Pancreatic Cancer (JHU, US)	82.46% (94/114)	1.75% (2/114)	No data
Pancreatic Cancer (OICR, CA)	0.00%	No data	No data
Pancreatic Cancer (QCMG, AU)	0.00%	0.00%	0.00%
Serous Cystadenocarcinoma (TCGA, US)	3.11% (12/386)	No data	No data
Small Cell Lung Carcinoma (WTSI, UK)	0.00%	No data	0.00%

Pancreas Expression Data

Fold Change	Comparison	Technology	Publication
🟥	Intraductal papillary mucinous neoplasms (IPMN) / Normal pancreas ND (microdissected normal ductal cells)	Transcriptomics	Gene Expression Profiling Identifies Genes Associated with Invasive Intraductal Papillary Mucinous Neoplasms of the Pancreas
🟥	Pancreatic Endocrine Tumors (PETs) (non-functioning) with liver metastasis / Purified islet cells	Transcriptomics	Gene expression profiles of progressive pancreatic endocrine tumours and their liver metastases reveal potential novel markers and therapeutic targets
🟧	Pancreatic adenocarcinoma (PDAC) / Normal pancreas (normal ductal cells) (both microdissected)	Transcriptomics	Transcriptome analysis of microdissected pancreatic intraepithelial neoplastic lesions.

Data source: Pancreatic Expression Database

Powered by BioMart 0.8

info@icgc.org I Contacts

© 2010 International Cancer Genome Consortium. All rights reserved. Terms & Conditions I Privacy Policy

Figure 3. Interoperability with the ICGC experimental data. Access available from http://dcc.icgc.org. The figure shows the ICGC data portal report for the TP53 gene including gene information; ICGC experimental sequencing results obtained from the participating centres; and PED data. In the PED Report section, a heatmap represents, visually, the level of de-regulation as extracted from the original publication and stored in PED. If you hold your mouse over the coloured box, it will display the fold change value (when available).

REFERENCES

1. Hariharan,D., Saied,A. and Kocher,H.M. (2008) Analysis of mortality rates for pancreatic cancer across the world. *HPB*, **10**, 58–62.
2. Goonetilleke,K.S. and Siriwardena,A.K. (2008) Current status of gene expression profiling of pancreatic cancer. *Int. J. Surg.*, **6**, 81–83.
3. Chelala,C., Hahn,S.A., Whiteman,H.J., Barry,S., Hariharan,D., Radon,T.P., Lemoine,N.R. and Crnogorac-Jurcevic,T. (2007) Pancreatic Expression database: a generic model for the organization, integration and mining of complex cancer datasets. *BMC Genomics*, **8**, 439.
4. Tatusova,T. Genomic databases and resources at the National Center for Biotechnology Information. *Methods Mol. Biol.*, **609**, 17–44.
5. Hubbard,T.J., Aken,B.L., Ayling,S., Ballester,B., Beal,K., Bragin,E., Brent,S., Chen,Y., Clapham,P., Clarke,L. *et al.* (2009) Ensembl. *Nucleic Acids Res.*, **37**, D690–D697.
6. R Development Core Team. (2009) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org.
7. Cline,M.S., Smoot,M., Cerami,E., Kuchinsky,A., Landys,N., Workman,C., Christmas,R., Avila-Campilo,I., Creech,M., Gross,B. *et al.* (2007) Integration of biological networks and gene expression data using Cytoscape. *Nat. Protoc.*, **2**, 2366–2382.
8. Matthews,L., Gopinath,G., Gillespie,M., Caudy,M., Croft,D., de Bono,B., Garapati,P., Hemish,J., Hermjakob,H., Jassal,B. *et al.* (2009) Reactome knowledgebase of human biological pathways and processes. *Nucleic Acids Res.*, **37**, D619–D622.
9. Jones,P., Cote,R.G., Cho,S.Y., Klie,S., Martens,L., Quinn,A.F., Thorneycroft,D. and Hermjakob,H. (2008) PRIDE: new developments and new datasets. *Nucleic Acids Res.*, **36**, D878–D883.
10. UniProt Consortium. (2010) The Universal Protein Resource (UniProt) in 2010. *Nucleic Acids Res.*, **38**, D142–D148.
11. International network of cancer genome projects. (2010) *Nature*, **464**, 993–998.
12. Barrett,T., Troup,D.B., Wilhite,S.E., Ledoux,P., Rudnev,D., Evangelista,C., Kim,I.F., Soboleva,A., Tomashevsky,M., Marshall,K.A. *et al.* (2009) NCBI GEO: archive for high-throughput functional genomic data. *Nucleic Acids Res.*, **37**, D885–D890.
13. Parkinson,H., Kapushesky,M., Kolesnikov,N., Rustici,G., Shojatalab,M., Abeygunawardena,N., Berube,H., Dylag,M., Emam,I., Farne,A. *et al.* (2009) ArrayExpress update–from an archive of functional genomics experiments to the atlas of gene expression. *Nucleic Acids Res.*, **37**, D868–D872.
14. Rhead,B., Karolchik,D., Kuhn,R.M., Hinrichs,A.S., Zweig,A.S., Fujita,P.A., Diekhans,M., Smith,K.E., Rosenbloom,K.R., Raney,B.J. *et al.* The UCSC Genome Browser database: update 2010. *Nucleic Acids Res.*, **38**, D613–D619.
15. Haider,S., Ballester,B., Smedley,D., Zhang,J., Rice,P. and Kasprzyk,A. (2009) BioMart Central Portal – unified access to biological data. *Nucleic Acids Res.*, **37**, W23–W27.
16. Dowell,R.D., Jokerst,R.M., Day,A., Eddy,S.R. and Stein,L. (2001) The distributed annotation system. *BMC Bioinformatics*, **2**, 7.
17. Durinck,S., Moreau,Y., Kasprzyk,A., Davis,S., De Moor,B., Brazma,A. and Huber,W. (2005) BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics*, **21**, 3439–3440.
18. Blankenberg,D., Von Kuster,G., Coraor,N., Ananda,G., Lazarus,R., Mangan,M., Nekrutenko,A. and Taylor,J. Galaxy: a web-based genome analysis tool for experimentalists. *Curr. Protoc. Mol. Biol.*, Chapter 19, Unit 19.10.1–19.10.21.