The Omics Discovery Index - OmicsDI

Henning Hermjakob
Team Leader Molecular Networks
European Bioinformatics Institute
hhe@ebi.ac.uk
Finding a publication

- Straightforward through PubMed or (Europe) PubMed Central
Finding a Dataset

- Many disconnected search entry points
- Google does not work well, as it does not separate out datasets
OmicsDI Vision

A PubMed for (omics) datasets
Home Page – Graphical Browsing
Statistics

- 11 DBs
- 3 continents
- 78,000 datasets
- PRIDE, PeptideAtlas, MassIVE, GPMdb, ArrayExpress, Metabolights, EGA, MetabolomExpress, ...

www.omicsDI.org
Access Metrics

Latest Datasets
- Jun 29 Identification of protein-protein interaction...
- Jun 29 Bacillus cereus proteomics (Exoproteome)
- Jun 29 SNW1 is required for sister chromatid...
Search results overview

Omics Discovery Index

6 Q. Results for search term: human glioma IDH1 IDH2

Repository

Find your repositories

Refine by:
Organism

Find your species
- Mus musculus (3)
- Homo sapiens (8)

Tissue

www.omicsDI.org
Search results overview

Omics Discovery Index

Results for search term: human glioma IDH1 IDH2 AND TAXONOMY:"Homo sapiens"

Show results for
- Proteomics (3)
- Metabolomics (5)
- Transcriptomics (3)
- Genomics (2)
- Multi-Omics (3)

Repository

Find your repositories
- ArrayExpress (3)

Refine by:
Organism

Find your species
- Homo sapiens (3)

Tissue

Find your Tissue

1. Transcription profiling by array of human oligodendroglioma tumor samples with IDH1 and IDH2 mutations

To study the effects of IDH mutations, we collected and performed gene expression microarray analysis with tumor specimens from patients with grade II-III oligodendrogliomas. Sequencing for mutations of IDH1 and IDH2 was done. Gene expression was compared between IDH wildtype vs. mutant samples.

ORGANISM(S): Homo sapiens

2014-06-03 | E-MEXP-3239 | ArrayExpress

2. DNA hypermethylation in intrahepatic cholangiocarcinomas and glioblastomas

This SuperSeries is composed of the following subset Series: GSE32079: Mutations in IDH1 and IDH2 are associated with DNA hypermethylation in intrahepatic cholangiocarcinomas GSE32283: Mutations in IDH1 are associated with DNA hypermethylation in glioblastomas Refer to individual Series

ORGANISM(S): Homo sapiens

2012-06-21 | E-GEOD-32288 | ArrayExpress

3. IDH1 mutation defines methylation class and survival in human glioma

Astrocytoma, oligodendroglioma, oligoastrocytoma, and ependymoma are the main histologic subtypes of glioma. The molecular character of these subtypes has profound implications for understanding their causes and treatment. We describe the epigenetic landscape of these tumor types using novel DNA methylation profiling tools. There is a consistent association of methylation profile with tumor histology and IDH1 mutation status. Furthermore, tumors with ...

ORGANISM(S): Homo sapiens

2012-03-27 | E-GEOD-20395 | ArrayExpress

www.omicsDl.org
Dataset Information

**Transcription profiling by array of human oligodendroglioma tumor samples with IDH1 and IDH2 mutations**

**ABSTRACT:** To study the effects of IDH mutations, we collected and performed gene expression microarray analysis with tumor specimens from patients with grade II/III oligodendroglioma. Sequencing for mutations of IDH1 and IDH2 was done. Gene expression was compared between IDH wildtype vs. mutant samples.

**SAMPLE PROTOCOL:** Hybridization - Affymetrix Generic Hybridization. Labeling - Label protocol using the Affymetrix GeneChip Labeling Assay according to the manufacturer’s instructions. (Parameters: Amount of nucleic acid labeled = 1, Amplification = none, Mass unit = Micro gram) Nucleic Acid Extraction - tumor sample RNA was extracted with Trizol and purified with Qiagen RNeasy. (Parameters: Extracted product = total RNA Amplification = none)

**DATA PROTOCOL:** Feature Extraction - Affymetrix CEL analysis

**INSTRUMENT(S):** 418 (Affymetrix);

**ORGANISM(S):** Homo sapiens;

**DISEASE(S):** oligodendroglioma

2014-06-03 E-MEXP-3230 | ArrayExpress

---

**Similar Datasets**

1. **Isocitrate Dehydrogenase (IDH) Mutations Promote a Reversible ZEB1/miR-200-Dependent Epithelial Mesenchymal Transition (EMT)**  
   2012-11-08 | E-GEOD-41802 | ArrayExpress

2. **Somatic mosaic IDH1 and IDH2 mutations are associated with enchondroma and spindle cell haemangioma in Ollier disease and Maffucci syndrome**  
   2014-05-04 | E-GEOD-30844 | ArrayExpress

3. **DNA hypermethylation in intrahepatic cholangiocarcinomas and glioblastomas**  
   2012-06-21 | E-GEOD-32286 | ArrayExpress

4. **Mutant IDH inhibits HNF4a to disrupt hepatocyte differentiation and promote cholangiocarcinoma**  
   2014-06-03 | E-GEOD-57002 | ArrayExpress

5. **IDH2 mutation induced histone and DNA hypermethylation is progressively reversed by small molecule inhibition**  
   2014-11-28 | E-GEOD-51352 | ArrayExpress

---

**Publications**

**IDH mutation impairs histone demethylation and results in a block to cell differentiation.**


---

www.omicsDI.org
Ontology-aware Indexing

Omics Discovery Index

63 Q Results for search term: cardiovascular AND omics_type:"Proteomics"

Show results for
- Proteomics [63]
- Metabolomics [2]
- Transcriptomics [1]
- Genomics [2]
- Multi-Omics [2]

Repository

Find your repositories
- GPMDB [3]
- MassIVE [1]
- PeptideAtlas [12]

Page 1 2 3 4 5
Sort by: Accession Relevance Publication date
Page size 15 20 50
Showing 1 - 15 of 63

Phosphodiesterase 9A Controls Nitric-oxide Independent cGMP and Hypertrophic Heart Disease
Rat neonatal cardiac myocytes were treated with PDE5 and PDE9 inhibitor in the presence and absence of L-NAME. ORGANISM(S): Rattus norvegicus
2016-06-24 | PXD001585 | PRIDE
cyclic GMP cGMP PKG PDE9 Phosphodiesterase

A Circular RNA Controls Ribosomal RNA Maturation and Atherosclerosis in Humans
Circular RNAs (circRNAs) are broadly expressed in eukaryotic cells, but their role in human health and disease remains obscure. Here, we show that circular antisense non-coding RNA in the INK4 locus (circANRIL), which is transcribed at a locus of atherosclerotic cardiovascular disease on chromosome 9p21, confers athero-protection by controlling ribosomal RNA (rRNA) maturation and modulating pathways of atherogenesis. At the molecular level, circA ...
Ontology Highlighting

NCBO tools
Related by Publication

Use of Apoe −/− Mice in an 8 Month Systems Toxicology Inhalation/Cessation Study to Investigate Cardiovascular and Respiratory Exposure Effects of a Candidate Modified Risk Tobacco Product, THS 2.2, Compared with Conventional Cigarettes

OTHER RELATED OMICS DATASETS IN: E-MTAB-3681

INSTRUMENT(S): Q Exactive;

ORGANISM(S): Mus musculus;

TISSUE(S): lung;

2015-11-30 | PXD002530 | PRIDE

Publications

Effects of Cigarette Smoke, Cessation, and Switching to Two Heat-Not-Burn Tobacco Products on Lung Lipid Metabolism in Apoe−/− Mice
Proteomics and phosphoproteomics analysis of melanoma BRAF

ABSTRACT: Proteomics and phosphoproteomics analysis of BRAF mutant cells

DATA PROTOCOL: The raw MS data were processed and analyzed using MaxQuant software version 1.3.0.5. Protein and peptide false discovery rates were set at 1% and a minimum peptide length of 6 amino acids was required. In addition, Andromeda minimum score was set to 60 (~ Mascot score 20 PMID: 21254760). Enzyme specificity was set to trypsin/P allowing up to two missed cleavages. The search included cysteine carbamidomethylation as a fixed modification; oxidation of methionine and phospho-Tyrosine as variable modifications.

REANALYZED by: GPM32310005623

INSTRUMENT(S): LTQ Orbitrap Elite; Q Exactive;

ORGANISM(S): Homo sapiens;

TISSUE(S): Unspecified in relation to dataset;

2014-12-19 | PXD000497 | PRIDE

Publications

ROCK1 is a potential combinatorial drug target for BRAF mutant melanoma.

...
Data re-use

ABSTRACT: The availability of human genome sequence has transformed biomedical research over the past decade. However, an equivalent map for the human proteome with direct measurements of proteins and peptides does not exist yet. Here, we report a draft map of the human proteome based on high resolution Fourier transform mass spectrometry-based proteomics technology. In depth proteomic profiling of 30 histologically normal samples including 17 adult tissues, 7 fetal tissues.

SAMPLE PROTOCOL: 17 adult tissues, 7 fetal tissues, and 6 hematopoietic cell types were lysed in lysis buffer containing 4% SDS, 100 mM DTT and 100 mM Tris pH 7.5, homogenized, sonicated, heated for 10-15 min at 750°C cooled and centrifuged at 2,000 rpm for 10 minutes. The protein concentration of the cleared lysate was estimated using BCA assay and equal amounts from three donors were pooled for further fractionation. Proteins from SDS lysates were separated on SDS-PAGE and in gel digestion was carried out using trypsin.

DATA PROTOCOL: Mass spectrometry data obtained from all LC-MS analysis were searched against Human RefSeq50 database (containing 33,833 entries along with common contaminants) using Sequel and Mascot version 2.2) search algorithms through Proteome Discoverer 1.3 Thermo Scientific, Bremen, Germany). Enzyme specificity was set as trypsin with maximum one missed cleavage allowed. The minimum peptide length was specified to be 6 amino acids. Carbamidomethylation of cysteine was specified as fixed.

REANALYZED by: P Aae004737, P Aae004901, P Aae004728, P Aae004828, P Aae004738, P Aae004826

OTHER RELATED OMICS DATASETS IN: E-PROT-1, E-PROT-1

www.omicsDL.org
ROCK1 is a potential combinatorial drug target for BRAF mutant melanoma.


Abstract: Treatment of BRAF mutation carriers may eventually relapse due to drug resistance mechanisms. Here we show that RAF kinase inhibitors can be reactivated by inducing endogenous inhibitors of RAF kinases (so-called combinatorial drug targets). We identified a small molecule able to induce combinatorial drug target (CDT) formation in multiple cell lines and primary melanoma cell cultures. This molecule, which was previously identified as a BRAF inhibitor, appears to be induced upon PLX4720 treatment. We propose that other CDTs can be discovered by targeting the endogenous inhibitors of RAF kinases.

Comment

Yasset Perez-Riverol · 2 months ago
This dataset has been reanalyzed in two different publications showing that the original results and claims were incorrect. Bromenshenk et al. [21], approach resulted in their identifying peptides of Iridovirus and Nosema origin in North American honey bees. However, subsequent studies by Foster [http://www.ncbi.nlm.nih.gov/pu...] and Knudsen and Chalkley [http://www.ncbi.nlm.nih.gov/pu...] have proved findings of Bromenshenk et al. wrong, with the misidentification of viral proteins being caused by inappropriate usage of viral-only sequence database for mass-spectrometry data interpretation.

www.omicsDI.org
Mandatory Fields:

- Repository Id
- Dataset Title
- Publication date
- Submitter information (Name, Affiliation)
- Original URL

Desired Fields:

- Description/Abstract
- Sample and Data Protocols
- PubMed Id
- Organism, Tissue, Disease

Additional Fields:

- Protein Id (Ensembl or Uniprot)
- Metabolite Id (ChEMBL)
- More…
Indexing System

Lucence-based Indexer System:

Strength:
- Already implemented
- Open source if we need to migrate the infrastructure.
- Indexed with all the EBI information facilitates cross-references.
- Indexes all of EBI (1.1 B entries), known to scale well

Limitations:
- Only an indexing system, not a database -> no persistence
- Relies on EBI infrastructure
DDI application

• Database (Mongo)
  • Access statistics

• Web Service
  • Search
  • Statistics

• Web Application
  • Statistics
  • Browsing
  • Knowledge Discovery
Indexing system

Proteomic dataset

Genomic dataset

Metabolomic dataset

publication

metadata

publication

metadata

publication

metadata

PUBMED

EBI Search Indexer
- More than 30 databases
- 1,376,278,505 records
Indexing system

Proteomic dataset

- Publication
- Metadata
- Cross-references
- UNIPROT
- ENSEMBL

Genomic dataset

- Publication
- Metadata
- Cross-references
- G
- ENSEMBL

Metabolomic dataset

- Publication
- Metadata
- Cross-references
- M
- CHEMBL

EBI Search Indexer
- More than 30 databases
- 1,376,278,505 records
Related by Shared Molecules

Only show the datasets with similarity scores above: 0.56

Threshold: 0.56

Comprehensive Assessment of...
Comprehensive Assessment of...
Proteome profiling of breast cancer...
Chr16-HPP, Shotgun Analysis...
Proteomic characterization of...
Chr16-HPP.Pilot experiment.JPR HP...
Perspective: Beyond Omics Datasets

Proteomic dataset

Genomic dataset

Metabolomic dataset

Cross-references

publication

metadata

UNIPROT
ENSEMBL

ENSEMBL

CHEMBL

PUBMED

EBI Search Indexer
- More than 30 databases
- 1,376,278,505 records
Perspective: Beyond Omics Datasets

Proteomic dataset

- Publication
- Metadata
- Cross-references

Genomic dataset

- Publication
- Metadata
- Cross-references

Metabolomic dataset

- Publication
- Metadata
- Cross-references

UNIPROT
ENSEMBL

P
P
P
P

ENSEMBL

G
G
G

CHEMBL

M
M
M

EBI Search Indexer
- More than 30 databases
- 1,376,278,505 records

Systems Biology

REACTOME Pathways

P
G
M
OmicsDI and bioCADDIE

- Originally independently funded
- Administrative supplement 2016
  - PIs
    - Peipei Ping, UCLA
    - Lucila Ohno-Machado, UCSD
    - Susanna Assunta-Sansone, U Oxford
    - Eric Deutsch, ISB
    - Henning Hermjakob, EBI
  - WPs
    - Map OmicsDI, bioCADDIE data model
    - Re-usable visualisation widgets
    - Access metrics

- Collaboration
  - Mapping from OmicsDI data model to DATS model
  - OmicsDI provides metadata from “its” repositories to bioCADDIE
  - OmicsDI goes more into the “depth” for omics
  - bioCADDIE focuses on breadth
Acknowledgements

Metabolights Team
Kenneth Haug
Pablo Conesa Mingo
Christoph Steinbeck

EGA Team
Dylan Spalding
Helen Parkinson

NIH BD2K Center of Excellence @ UCLA
Grant number 1U54GM114833-01

PRIDE Team
Attila Csordas
Tobias Ternent
Noemi del Toro
Juan Antonio Vizcaíno

Reactome Team
Antonio Fabregat
Steve Jupe
Peter D’Eustachio
Lincoln Stein
NIH P41 HG003751

EBI Search Team
Silvano Squizzato
Young Mi Park
Rodrigo Lopez

Yasset Perez-Riverol

Mingze Bai
ProteomeXchange: A Global, distributed proteomics database

Mandatory raw data deposition since July 2015
ProteomeXchange: 3,802 datasets up until April 1st, 2016

Origin:
885 USA
465 Germany
342 United Kingdom
264 China
194 France
158 Netherland
136 Canada
126 Switzerland
107 Denmark
104 Spain
99 Australia
95 Japan
72 Belgium
68 Austria
63 Sweden
61 India
51 Norway
43 Taiwan
30 Italy
29 Brazil
28 Singapore
28 Finland
27 Ireland
27 Russia
26 Israel …

Datasets/year:
2012: 102
2013: 527
2014: 963
2015: 1758
2016: 452

Data volume:
Total: ~150 TB
Number of all files: ~400,000
PXD001860: ~ 12 TB
PXD000320-324: ~ 4 TB
PXD002319-26 ~2.4 TB
PXD001471 ~1.6 TB

Top Species studied by at least 20 datasets:
1526 Homo sapiens
485 Mus musculus
150 Saccharomyces cerevisiae
121 Arabidopsis thaliana
102 Rattus norvegicus
86 Escherichia coli
44 Bos taurus
35 Drosophila melanogaster
32 Glycine max

~ 700 species in total

Downloads

<table>
<thead>
<tr>
<th>Downloads</th>
<th>Hits/ No files = dataset</th>
<th>Dataset Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>PXD001641</td>
<td>31808/91 = 350</td>
<td>Single muscle fiber proteomics reveals unexpected mitochondrial specialization</td>
</tr>
<tr>
<td>PXD001126</td>
<td>4897/26 = 188</td>
<td>Building high-quality assay libraries for targeted analysis of SWATH MS data</td>
</tr>
<tr>
<td>PXD001574</td>
<td>4436/32 = 139</td>
<td>Phospho-iTRAQ</td>
</tr>
<tr>
<td>PXD000475</td>
<td>6638/50 = 133</td>
<td>Yersinia enterocolitica SOR17</td>
</tr>
<tr>
<td>PXD000700</td>
<td>266/2 = 133</td>
<td>Proteomic analysis of accessory gland in sexually mature Eriocheir sinensis</td>
</tr>
<tr>
<td>PXD000561</td>
<td>46578/2383=20</td>
<td>A draft map of the human proteome</td>
</tr>
</tbody>
</table>
Search results overview
ABSTRACT: proteomics data demonstrate that skin mucus of the air exposed large yellow croaker had a complex composition, with an unexpectedly high number of proteins (3,209), suggesting its multiple protective mechanisms possibly involved in antioxidant functions, oxygen transport, immune defence, and osmotic and ionic regulation. These results expand our knowledge of skin mucus secretion function, oxygen transport, immune defence, and osmotic and ionic regulation. These results expand our knowledge of skin mucus secretion and function in fish, highlighting its importance in response to stress.

SAMPLE PROTOCOL: skin mucus was collected from six healthy L. crocea individuals under air exposure as previously described [1]. Briefly, the fish were anaesthetised with a sub-lethal dose of tricaine -S (100 mg/l), and transferred gently to a sterile plastic bag for 3 min to slough off the mucus under air exposure. Proteins were extracted from a pool of skin mucus of six fish by the trichloroacetic acid acetone precipitation method and then digested by the trypsin gold promega, USA with a ratio of protein: trypsin...

DATA PROTOCOL: peptide and protein identification ---All spectra were mapped by mascot sever version 2.3.02 against the database of L. crocea genome with the parameters as follows: peptide mass tolerance 0.05 Da; fragment mass tolerance 0.1 Da; fixed modifications “Carbamidomethyl (C)”; variable modifications “Gln- pyro glu (N term Q), Oxidation (M), Deamidated (NQ)”. The 25,026 peptides were identified, which belong to 4,489 proteins encoded by L. crocea genome. For further analyses of the function of the mucus...

OTHER RELATED OMICS DATASETS IN: E-GEOD-57608

INSTRUMENT(S): TripleTOF 5600;

ORGANISM(S): Larimichthys crocea;

TISSUE(S): skin;

2015-04-17| PXD001218 | PRIDE