Exploring the Proteomics Services

Sandra Orchard
Why study proteins

• Proteins are the work horses of the cell – and it is proteins which are largely the targets of therapeutic agents

• Knowing what proteins are being expressed in your healthy cells and how this alters in disease is critical to understanding the disease process

• Proteins cannot be studied in isolation – they need to be seen in the context of the network of interactions they make in the cell
Protein identification

Protein Function

Network analysis

Pathway

Protein Structure
UniProt mission

• Provide the scientific community with a comprehensive, high-quality and freely accessible resource of protein sequence and functional information.
UniProt Knowledgebase (UniProtKB)

UniProtKB/Swiss-Prot

- Non-redundant
- High-quality
- Manually curated

*Reviewed*

UniProtKB/TrEMBL

- Redundant
- Computationally generated
- Automatically annotated

*Unreviewed*
Manual annotation of UniProtKB/Swiss-Prot

Splice variants

Sequence

Ontologies

Annotations

Protein names

Protein quaking

Also known as:

Nomenclature

Gene names

Name: QKI

Synonyms: Qk, Qk1, Qka

References


With


Molecular interactions [interaction networks] [proteins] [mutations] [expression] [imaging]

Citeseer: NUCLEIC ACID SEQUENCE (GENOMIC DNA) BINDING [ISOFORMS 1, 3, 4 AND 7], ALTERNATIVE SPlicing [ISOFORM 2], [STUDIES 1, 2, 3, 4, 5, 6, 7, 8, 9, 10].
Automatic annotation

- Allows annotation of UniProtKB/TrEMBL in an efficient and scalable manner with a high degree of accuracy
- Based on annotation rules which are created, tested and validated against published experimental data in UniProtKB/Swiss-Prot
- Rules are linked to InterPro member database signatures and define annotations to be added and conditions which must be fulfilled
- Signatures identify family members in UniProtKB
- Common annotation in Swiss-Prot is transferred to related family members in TrEMBL if they fulfil rule conditions
Extensively cross-referenced
Data import

Data from structures

Gene names

Additional bibliography

Interactions

GO terms

**Sequence annotation (Features)**

<table>
<thead>
<tr>
<th>Feature key</th>
<th>Position(s)</th>
<th>Length</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metal binding</td>
<td>58</td>
<td>1</td>
<td>Zinc 1: via kolo nitrogen</td>
</tr>
<tr>
<td>Metal binding</td>
<td>149</td>
<td>1</td>
<td>Zinc 1: via kolo nitrogen</td>
</tr>
<tr>
<td>Metal binding</td>
<td>182</td>
<td>1</td>
<td>Zinc 2: via kolo nitrogen</td>
</tr>
<tr>
<td>Metal binding</td>
<td>238</td>
<td>1</td>
<td>Zinc 2: via kolo nitrogen</td>
</tr>
<tr>
<td>Metal binding</td>
<td>314</td>
<td>1</td>
<td>Zinc 1: via kolo nitrogen</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Amino acid modifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modified residue</td>
</tr>
</tbody>
</table>

**Gene names**

![Gene names table]

**Binary Interactions**

<table>
<thead>
<tr>
<th>With</th>
<th>Entry</th>
<th>#Exp.</th>
<th>IntAct</th>
</tr>
</thead>
<tbody>
<tr>
<td>itself</td>
<td>Q9YJD8</td>
<td>3</td>
<td>EBI:945751, EBI:945751</td>
</tr>
<tr>
<td>AK5</td>
<td>P54253</td>
<td>3</td>
<td>EBI:945751, EBI:930964</td>
</tr>
<tr>
<td>ATN2</td>
<td>P54253-1</td>
<td>6</td>
<td>EBI:945751, EBI:930975</td>
</tr>
<tr>
<td>CEP70</td>
<td>Q0NHQ1</td>
<td>3</td>
<td>EBI:945751, EBI:739624</td>
</tr>
<tr>
<td>LNK1</td>
<td>Q8TEB1</td>
<td>2</td>
<td>EBI:945751, EBI:739832</td>
</tr>
<tr>
<td>PICK1</td>
<td>Q9NRD5</td>
<td>3</td>
<td>EBI:945751, EBI:79165</td>
</tr>
<tr>
<td>PSME3</td>
<td>P61289</td>
<td>3</td>
<td>EBI:945751, EBI:355546</td>
</tr>
<tr>
<td>VRK1</td>
<td>Q99986</td>
<td>9</td>
<td>EBI:945751, EBI:1769146</td>
</tr>
<tr>
<td>XRCC5</td>
<td>P13010</td>
<td>6</td>
<td>EBI:945751, EBI:357997</td>
</tr>
<tr>
<td>XRCC6</td>
<td>P12956</td>
<td>3</td>
<td>EBI:945751, EBI:353208</td>
</tr>
</tbody>
</table>

**The 4.1/ezrin/radixin/moesin domain of the PAL-1/Protein 4.1B tumour suppressor interacts with 14-3-3 proteins.**

Yu T., Robb V.A., Singh V., Gutmann D.H., Newham I.F.


Results show that three 14-3-3 isoforms beta gamma and eta are DAL-1/Protein 4.1B-binding proteins.

**Regulation of TSC2 by 14-3-3 binding.**

Li Y., Inoki K., Yeung R., Guan K.L.


**GO terms**

- **Biological process**
  - Ras protein signal transduction
  - Inferred from Experiment. Source: Reactome
  - activation of pro-apoptotic gene products
  - Inferred from Experiment. Source: Reactome
  - cytoplasmic sequestration of protein
  - Inferred from direct assay. Source: BHF-UCL
  - negative regulation of protein amino acid dephosphorylation
  - Inferred from direct assay. Source: BHF-UCL
  - positive regulation of catalytic activity
  - Inferred from direct assay. Source: BHF-UCL

- **Cellular component**
  - cytosol
  - Inferred from Experiment. Source: Reactome
  - melanosome
  - Inferred from electronic annotation. Source: UniProtKB:SubCell
  - peripheral region of cytoplasm
  - Inferred from direct assay. Source: UniProtKB

- **Molecular function**
  - histone deacetylase binding
  - Inferred from physical interaction. Source: BHF-UCL
  - phosphate binding
  - Inferred from physical interaction. Source: BHF-UCL
  - protein domain specific binding
  - Inferred from physical interaction. Source: UniProtKB
Sequence curation

Improve sequence curation and representation, focus on capturing complexity of protein products – splicing, protein cleavage, transcriptional read-through, uORFS
PRIDE – MS identifications

- PRIDE-Archive - stores submitted MS identifications plus meta-data describing sample
- PRIDE-Proteomes will hold a filtered subset - high quality identifications
- Will then build an expression atlas to give organism/cellular/subcellular map at the protein level
- Sample annotation enables comparison of normal/cell line/diseased cell protein profile
- Protein identification provides proof of protein existence and will also be a source of PTM data → UniProtKB.
782 Search results

Filter your results

Sort by: Accession Title Relevance Publication date (Descending)

PXD000226
Human Uterine Smooth Muscle S-nitrosoproteome
Species: Homo sapiens (Human)
Project description: S-nitrosoproteome of human uterine smooth muscle in different s (More)
Made public: 2014-03-03

PXD000581
Combined Proteomic and Transcriptomic Interrogation of the Venom Gland of Conus geographus
Species: Conus geographus (Geography cone) (Nubecula geographus)
Project description: Combined Proteomic and Transcriptomic Interrogation of the Venom (More)
Made public: 2014-02-14

PXD000460
Pithovirus particle proteome
Species: unidentified
### PRIDE Inspector – suite of analysis tools

<table>
<thead>
<tr>
<th>Protein</th>
<th>Threshold</th>
<th># Peptides</th>
<th># Distinct Peptides</th>
<th># PTMs</th>
</tr>
</thead>
<tbody>
<tr>
<td>SELH10</td>
<td>29.0</td>
<td>151</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>P07195</td>
<td>29.0</td>
<td>54</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>P05787</td>
<td>29.0</td>
<td>114</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>P29401</td>
<td>29.0</td>
<td>75</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>P21333</td>
<td>29.0</td>
<td>72</td>
<td>12</td>
<td>30</td>
</tr>
<tr>
<td>P30101</td>
<td>29.0</td>
<td>52</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>P077387</td>
<td>20.0</td>
<td>47</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>P02497</td>
<td>29.0</td>
<td>73</td>
<td>4</td>
<td>19</td>
</tr>
<tr>
<td>P07355</td>
<td>29.0</td>
<td>78</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

**Protein [P14518] Modified AAs: [M - 15.9949] [C - 57.0215]**

<table>
<thead>
<tr>
<th>Peptide</th>
<th>Ranking</th>
<th>Delta m/z</th>
<th>Charge</th>
<th>Precursor m/z</th>
<th>Modifications</th>
<th>Mascot Score</th>
<th>charge state</th>
<th>Length</th>
<th>Start</th>
<th>Stop</th>
<th>Spectrum ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>DLDIEPAEK</td>
<td></td>
<td></td>
<td></td>
<td>571.31</td>
<td></td>
<td>45.21</td>
<td>2.0</td>
<td>11</td>
<td>295</td>
<td>365</td>
<td>5863</td>
</tr>
<tr>
<td>LDIDSPFRITAR</td>
<td></td>
<td>-0.04</td>
<td>2</td>
<td>590.29</td>
<td>33.63</td>
<td>2.0</td>
<td>11</td>
<td>33</td>
<td>43</td>
<td>7220</td>
<td></td>
</tr>
<tr>
<td>LDIDSPFRITAR</td>
<td></td>
<td>-0.02</td>
<td>2</td>
<td>599.31</td>
<td>43.23</td>
<td>2.0</td>
<td>11</td>
<td>33</td>
<td>43</td>
<td>7236</td>
<td></td>
</tr>
<tr>
<td>LDIDSPFRITAR</td>
<td></td>
<td>0.01</td>
<td>2</td>
<td>599.32</td>
<td>43.23</td>
<td>2.0</td>
<td>11</td>
<td>33</td>
<td>43</td>
<td>7240</td>
<td></td>
</tr>
</tbody>
</table>

**Spectrum**

- Fragment ion: y Ion
- Show All

**Amino Acid**

- Show All
‘Omics analysis

• The end result of such experiments is usually a list of identifiers

• Biological context needs to be derived from this list

• Tools available for this, using EBI data
  • Network analysis
  • Pathway analysis
  • GO representation analysis
Network analysis

• Protein interaction networks aim to represent the complex web of molecular interactions within a cell

• Overlay of expression data on those networks can help identify clusters of interacting (i.e. probably functionally related) proteins which are up(down)-regulated in your sample

• Graph analysis tools (e.g. as supplied in R) can enable a deeper understanding of the data
IntAct

- Publicly available repository of molecular interactions (mainly PPIs) - ~445K binary interactions taken from >12,500 publications (February 2014)

- Data is standards-compliant and available via our website, for download at our ftp site or via PSICQUIC

http://www.ebi.ac.uk/intact
ftp://ftp.ebi.ac.uk/pub/databases/intact
www.ebi.ac.uk/Tools/webservices/psicquic/view/main.xhtml

- Provide open-access versions of the software to allow installation of local IntAct nodes.
IntAct – Home Page

IntAct provides a freely available, open source database system and analysis tools for molecular interaction data. All interactions are derived from literature curation or direct user submissions and are freely available. To perform a search in the IntAct database use the search box in the top left corner.

Examples:
- Gene name: e.g. BRCA2
- UniProtKB Ac: e.g. Q05609
- UniProtKB Id: e.g. dmc1
- Pubmed Id: e.g. 10831611

Please supply your feedback to helpdesk. We thank you for your help in further developing IntAct.

Citing IntAct
- The IntAct molecular interaction database is http://www.nature.com/nmeth/journal/v9/n4/full/nmeth.1931.html?FPID=221212201

Dataset of the month: May
- Systematic analysis of dimeric E3-RING interactions reveals increased combinatorial...
Use IntAct to:

- Search for the interactors of one, or a set of, proteins
- Perform an ontology-driven search (e.g. find all the interaction of human peroxisomal proteins – GO:0005777 peroxisome)
- Download a large dataset (e.g. the human interactome) in standard formats (XML, tab-delimited, XGMML, RDF)
Network analysis – adding quantitative expression data to an interaction network
Gene ontology enrichment analysis

Choose Layouts → Cytoscape Layouts → Hierarchical layout
Pathway analysis of large datasets

• Network analysis highlights interacting clusters of molecules (often indicates protein complexes stable/transient) – may have similar function and subcellular location

• To group molecules by Process, can also subject identifier link to pathway analysis
A Database of human biological pathways
Reactome is...

Free, online, open-source curated database of pathways and reactions in human biology

Authored by expert biologists, maintained by Reactome editorial staff (curators)

Mapped to cellular compartment – agnostic of cell/tissue type/state
Data Expansion - Link-outs From Reactome

- GO
  - Molecular Function
  - Compartment
  - Biological process

- KEGG, ChEBI – small molecules
- UniProt – proteins
- Sequence dbs – Ensembl, OMIM, Entrez Gene, RefSeq, HapMap, UCSC, KEGG Gene
- PubMed references – literature evidence for events
Species Selection

Switch Species: Homo sapiens

Search results:
- Homo sapiens
- Saccharomyces cerevisiae
- Schizosaccharomyces pombe
- Canis familiaris
- Rattus norvegicus
- Bos taurus
- Drosophila melanogaster
- Plasmodium falciparum
- Staphylococcus aureus N315
- Taeniopygia guttata
- Sus scrofa
- Dictyostelium discoideum
- Arabidopsis thaliana
- Mus musculus
- Mycobacterium tuberculosis
- Gallus gallus
- Xenopus tropicalis
- Oriza sativa
- Caenorhabditis elegans
- Escherichia coli
Data Expansion – Projecting to Other Species

Human

\[ A + ATP \rightarrow A-P + ADP \]

Mouse

\[ A + ATP \rightarrow A-P + ADP \]

Drosophila

\[ \text{No reaction} \]

No orthologue - Protein not inferred

Reaction not inferred
Reactome Tools

• Interactive Pathway Browser

• Pathway Mapping and Over-representation

• Expression overlay onto pathways

• Molecular Interaction overlay

• Biomart
# Expression analysis

## Analysis results, per pathway

This table provides an overview of your expression data in a pathway context. For each Reactome pathway, the total number of proteins is shown, plus the number of genes/proteins in your dataset that match. By clicking on a pathway name, you will be taken to an interactive graphical representation of the pathway, where your expression levels are represented as coloration of proteins.

Select format to download this table:  
- [microsoft excel (tsv)](download_url)

<table>
<thead>
<tr>
<th>Pathway</th>
<th>Species</th>
<th>IDs in pathway (%)</th>
<th>10h_control</th>
<th>10h</th>
<th>14h</th>
<th>18h</th>
<th>24h</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not assigned</td>
<td>Not known</td>
<td>21 (0%)</td>
<td>7.3</td>
<td>6.8</td>
<td>7.0</td>
<td>6.0</td>
<td>6.5</td>
</tr>
<tr>
<td>Apoptosis</td>
<td>Homo sapiens</td>
<td>74 (43%)</td>
<td>7.9</td>
<td>7.6</td>
<td>7.7</td>
<td>7.7</td>
<td>7.4</td>
</tr>
<tr>
<td>Binding and Uptake of Ligands by Scaveng</td>
<td>Homo sapiens</td>
<td>6 (3%)</td>
<td>4.8</td>
<td>5.2</td>
<td>5.2</td>
<td>5.0</td>
<td>5.4</td>
</tr>
<tr>
<td>Cell Cycle</td>
<td>Homo sapiens</td>
<td>115 (25%)</td>
<td>8.4</td>
<td>7.8</td>
<td>8.0</td>
<td>8.0</td>
<td>7.5</td>
</tr>
<tr>
<td>Cell-Cell communication</td>
<td>Homo sapiens</td>
<td>4 (3%)</td>
<td>7.8</td>
<td>7.8</td>
<td>7.9</td>
<td>7.7</td>
<td>7.8</td>
</tr>
<tr>
<td>Cellular responses to stress</td>
<td>Homo sapiens</td>
<td>24 (12%)</td>
<td>7.6</td>
<td>7.1</td>
<td>7.2</td>
<td>7.1</td>
<td>6.9</td>
</tr>
<tr>
<td>Chromatin organization</td>
<td>Homo sapiens</td>
<td>4 (4%)</td>
<td>8.1</td>
<td>7.6</td>
<td>7.5</td>
<td>7.8</td>
<td>7.2</td>
</tr>
<tr>
<td>Circadian Clock</td>
<td>Homo sapiens</td>
<td>4 (2%)</td>
<td>10.3</td>
<td>10.1</td>
<td>10.1</td>
<td>10.2</td>
<td>9.3</td>
</tr>
<tr>
<td>Developmental Biology</td>
<td>Homo sapiens</td>
<td>19 (5%)</td>
<td>5.5</td>
<td>5.5</td>
<td>5.6</td>
<td>5.5</td>
<td>5.6</td>
</tr>
<tr>
<td>Disease</td>
<td>Homo sapiens</td>
<td>180 (17%)</td>
<td>3.9</td>
<td>8.7</td>
<td>8.8</td>
<td>8.8</td>
<td>8.5</td>
</tr>
<tr>
<td>DNA Repair</td>
<td>Homo sapiens</td>
<td>74 (69%)</td>
<td>7.6</td>
<td>7.1</td>
<td>7.2</td>
<td>7.3</td>
<td>6.9</td>
</tr>
<tr>
<td>DNA Replication</td>
<td>Homo sapiens</td>
<td>77 (77%)</td>
<td>3.8</td>
<td>8.1</td>
<td>8.3</td>
<td>8.3</td>
<td>7.7</td>
</tr>
<tr>
<td>Extracellular matrix organization</td>
<td>Homo sapiens</td>
<td>22 (8%)</td>
<td>5.7</td>
<td>5.9</td>
<td>5.8</td>
<td>5.8</td>
<td>6.0</td>
</tr>
<tr>
<td>Gene Expression</td>
<td>Homo sapiens</td>
<td>272 (38%)</td>
<td>9.0</td>
<td>8.6</td>
<td>8.8</td>
<td>8.8</td>
<td>8.4</td>
</tr>
<tr>
<td>Hemostasis</td>
<td>Homo sapiens</td>
<td>76 (17%)</td>
<td>6.0</td>
<td>6.0</td>
<td>6.0</td>
<td>6.0</td>
<td>6.1</td>
</tr>
<tr>
<td>Immune System</td>
<td>Homo sapiens</td>
<td>95 (8%)</td>
<td>3.0</td>
<td>7.7</td>
<td>7.8</td>
<td>7.8</td>
<td>7.5</td>
</tr>
</tbody>
</table>
Exploring proteomics

- UniProt can act as a central hub to explore other resources
- All resources designed for large-scale queries – web services, APIs etc
- All data and most software is open-source and can be locally installed
- Our design is driven by your needs – please feedback