

**DATA**

**SOURCES**

(November 2016)

The Euretos Knowledge Platform is an innovative bioinformatics platform used by some of the world’s largest pharma & biotech companies, academic hospitals and research institutes. The platform is unparalleled in terms of scope and breadth, containing phenotypical down to molecular data. In addition, the platform supports powerful embedded (predictive) analytics that are too resource intensive to run in traditional environments.

The Euretos Knowledge platform powers advanced applications specifically designed to cater for the most essential research questions. At every stage the researcher can easily access supporting references and review all the underlying evidence.

Our unique configurable data loading platform enables us to update and extend the content of the platform efficiently. As a result, the Euretos Knowledge Platform is the most up to date and fastest growing interconnected Knowledge Base. In the November 2016 release, the platform integrates and interconnects data from the following sources.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Source** | **Description** | **Content** |
| 1 | APB | Australian Phenome Bank | Relation mouse model / strain - genotype from relational database |
| 2 | Aracyc | pathways in Arabidopsis Thaliana | Relations between Genes - Enzymatic reactions; Enzymatic Reactions – Pathways and Gene - Pathways |
| 3 | ARC | Animal Resources Centre | Relation mouse model / strain - genotype from relational database |
| 4 | Baseline Expression Atlas | RNA expresssion in healthy tissue | RNAseq expression levels in healthy tissue, defines relations: gene – expression level – tissue. All read level data has been processed in a single pipeline. Used to calculate expression specificity and display expression profiles per gene. |
| 5 | BIND | molecular interactions | biomolecular interaction, complex and pathway information, defines relations: “gene”-interacts with “gene” ; “gene”-forms protein complex with – “gene” ; gene- affects – “pathway” |
| 6 | BindingDB | biochemical interaction data for small molecules and protein targets | Binding affinities of small molecules and target proteins with mechanism of action; “molecule” – inhibits/stimulates – “protein“ including IC50/EC50 measurement values |
| 7 | Biogrid | protein-protein interactions and protein- small molecule interactions | genetic interactions, chemical associations and post translational modifications. Defines relations:  “gene” – interacts with – “gene” ; “gene” – controls phosphorylation of – “gene” ; “gene” – controls state change of – “gene” ; “ gene” – produces - “molecule” ; “molecule” – consumption controlled by – “gene” ; “molecule” – affects – “gene” |
| 8 | Biomodels | computational models of biological processes | genetic and molecular interactions ; defines relations “gene“– affects – “pathway” “gene” – interacts with – “gene” ; “gene” – interacts with – “molecule” |
| 9 | BioPlex | protein interaction networks | protein interaction networks ; define relations: “gene” - interacts with – “gene” |
| 10 | CARD | Centre for Animal Resources and Development | Relation mouse model / strain - genotype from relational database |
| 11 | CellLine | curated cell line overview | Lists cell lines with relations to species, condition, etc. |
| 12 | Chebi | vocabulary for chemical entities of biological interest | Classifies compounds and defines relations between small molecules and between small molecules and (functional) classifications: molecule – is a – molecule class ; molecule – classified as - chemical function ; molecule – converts to – molecule ; “molecule” – has part – “molecule” ; “molecule” – functionally related to - “molecule” |
| 13 | ChEMBL | compounds, targets and biochemical assays | Binding affinities of small molecules and target proteins with mechanism of action; molecule – binds to – “protein” ; “molecule” – inhibits/stimulates – “protein“ including IC50/EC50 measurement values |
| 14 | Clinical Trials | clinical trials | relations pertaining to drugs tested for indications, including measures for levels of success, trial phase, status. Defines relations: ‘drug” – tested for – indication ; “drug – tested with – “drug” |
| 15 | Clinvar | clinical genomic variation | clinical genomic variation, define relations: “gene” – variant causes – “disease” with information on genotype |
| 16 | CMMR | Canadian Mouse Mutant Repository | Relation mouse model / strain - genotype from relational database |
| 17 | Comparative Toxicogenomics Database | gene - phenotype - chemical interactions | gene - phenotype - chemical association annotations from article publications ; defines relations: “gene” – associated with - “ disease” ; “molecule” - inhibits - “gene” ; “molecule” - stimulates – “gene” ; “molecule” – produces – “gene” ; “molecule” - affects – “gene” ; “molecule” – indicates – “disease” ; “molecule” – treats – “disease” ; pathway – functionally related to – “disease” |
| 18 | ConceptWiki | vocabulary | Lists mappings of concepts and synonyms |
| 19 | CORUM | protein complexes | Defines protein interactions and protein complexes ; defines relations”: “gene” – forms protein complex with – “gene” ; “gene” – interacts with – “gene” |
| 20 | CROP | genetic annotations and diversity of traits and environmental treatments for plants | Genetic annotations and diversity of traits and environmental treatments for plants, based on gene ontology, plant ontology, environment ontology and trait ontology |
| 21 | DECIPHER | clinical phenotypic - genotypic data | Clinical genotypic to phenotypic annotation ; gene – product variant causes – phenotype (disease) |
| 22 | Differential Expression Atlas | gene expression patterns in different biological conditions | Gene expression patterns in different biological conditions ;  Experiment – evaluates – condition  Gene – over/underregulated - Experiment |
| 23 | DIP | protein interactions | Protein interactions |
| 24 | DisGeNet | gene - disease associations | Gene - disease associations, based on text mining, clinical data and annotation databases. Covers biomarker, variation and general literature annotations. DisGeNet scores (0-1) are translated to scientific evidence score (1, 4, or 5) |
| 25 | Drugbank | chemical,  pharmacological and pharmaceutical information for approved and experimental drugs and protein therapies | Drug – Target associations. |
| 26 | EMMA | European Mouse Mutant Archive | Relation mouse model / strain - genotype from relational database |
| 27 | Ensembl Plant Variation | genetic variation and phenotype association for plants | genetic variation and phenotype association for plants based on trait ontology and environment ontology |
| 28 | Enzyme | Enzyme characterisation | Enzyme characterisation |
| 29 | Experimental Factor ontology | ontology for variables in biological experiments | ontology for variables in biological experiments |
| 30 | Gene Ontology | molecular functions, biological processes and cellular components | molecular functions, biological processes and cellular components |
| 31 | GWAS | curated resource of SNP-trait associations | curated resource of SNP-trait associations; gene – variant causes – disease ; gene – variant results in abnormal - phenotype |
| 32 | HGNC | nomenclature for human genes | nomenclature for human genes |
| 33 | HPRD | protein interactions and pathways | protein interactions and pathways |
| 34 | Human Metabolome database | small molecule metabolites | small molecule metabolites |
| 35 | Human Phenotype Ontology | phenotypic abnormalities encountered in human disease | Ontology used for gene annotations |
| 36 | Human Phenotype Ontology - Pubmed | Datamining results for Phenotype-Disease associations | Datamining results for Phenotype-Disease associations; phenotype – is manifestation of - disease |
| 37 | Human Protein Atlas | RNAseq and protein ubandance measurement in healthy tissues and cancer tissues ; Gene classifications | gene – expressed in – Tissue  gene – specifically expressed in –Tissue protein – gene product expressed in – Tissue  protein – gene product expressed in - Cancer  gene – gene product not expressed in - Cancer |
| 38 | Human Protein Map | Protein abundance measurements in healthy tissues | Protein abundance measurements in healthy tissues. |
| 39 | Humancyc | molecular pathways | molecular pathways |
| 40 | ICD10 | disease classification | Disease naming and classification ; Disease – parent of - Disease |
| 40 | Intact | molecular interactions | molecular interactions |
| 41 | IntAct Complex | protein complexes | protein complexes;  gene – form protein complex with - gene |
| 42 | International Mouse Strain Resource (IMSR) | Mouse model resources | Relation mouse model / strain - genotype from relational database |
| 43 | JAX | Strain database | Relation mouse model / strain - genotype from relational database |
| 44 | KOMP | Knockout Mouse Project | Relation mouse model / strain - genotype from relational database |
| 45 | MeSH | vocabulary | vocabulary for Medical Subject headers ; also contains ontological relations between concepts. |
| 46 | miRTarBase | microRNA target interactions | microRNA target interactions ; miRNA – interacts with - gene |
| 47 | MMRRC | Mutant Mouse Regional Resource Center | Relation mouse model / strain - genotype from relational database |
| 48 | Monarch | gene - phenotype associations from model organisms | associations identified from model organisms. Gene – ortholog associated with – Phenotype / Disease |
| 49 | Monarch Cross Species Phenotype Ontology | Ontology | Mapping of phenotype concepts and gene – phenotypes |
| 50 | Mouse Genome Informatics | MGI is the international database resource for the laboratory mouse, providing integrated genetic, genomic, and biological data to facilitate the study of human health and disease. | Gene - Phenotype relations ; Gene – disease relations ; Mouse gene – Ortholog – Human gene relations. |
| 51 | Mouse Phenotype Ontology | Ontology | Ontology containing concepts for mouse phenotypes and mappings to human phenotypes. |
| 52 | MRC Harwell | Harwell Science and Innovation Campus | Relation mouse model / strain - genotype from relational database |
| 53 | MUGEN | Mugen Mouse Database | Relation mouse model / strain - genotype from relational database |
| 54 | National Cancer Institute | vocabulary for cancers, agents and related substances | Contains relations between gene- pathways and gene – diseases. Annotated database. |
| 55 | NCBI gene annotation | gene information and annotation | . Gene – part of chromosome ; gene – part of – chromosome band ; gene – neighbor of – gene ; Gene – disease annotations from curated database. |
| 56 | NCBI gene orthologues | gene orthologues | gene orthologues |
| 57 | NCBI gene sequence | reference sequences | Genome build and gene locations for protein coding genes, various types of RNA’s and pseudo genes |
| 58 | NCI PID | molecular pathways | molecular signaling and metabolic pathways |
| 59 | NCIMR | National Cancer Institute at Frederick | Relation mouse model / strain - genotype from relational database |
| 60 | NIG | National Institute of Genetics (Japan) | Relation mouse model / strain - genotype from relational database |
| 61 | OMIM | genetic phenotypes | genetic phenotypes and diseases; gene – variant causes - disease |
| 62 | Orphanet | rare diseases and orphan drugs | Annotated relations Gene – disease assocations ; gene – phenotype assocations. |
| 63 | Panther | protein classifications | Gene /Pprotein classifications and families |
| 64 | Pathway ontology | classic metabolic, regulatory, signaling, drug and disease pathways | classic metabolic, regulatory, signaling, drug and disease pathways |
| 65 | Phosphosite Plus | protein post-translational modifications | Curated gene interactions; gene – phosphorylates - gene |
| 66 | Phytozome | gene orthologs in plants | gene – ortholog of - gene |
| 67 | Plant environment ontology | vocabulary of plant environmental conditions and experimental treatments | Curated relations from annotations of experimental conditions; gene – associated – condition. |
| 68 | Plant ontology | plant anatomy, morphology and stages of development for all plants | Curated relations from annotations of experimental conditions and QTL analyses : gene – associated - trait |
| 69 | Pubmed | scientific literature; abstracts | Abstracts of scientific literature with source information; serves as reference for annotated databases. |
| 70 | RBRC | Riken BioResource Center | Relation mouse model / strain - genotype from relational database |
| 71 | Reactome | molecular pathways | Molecular signaling and metabolic pathways ; gene – gene and gene – metabolite interactions. |
| 72 | Recon X | human metabolism | molecular signaling and metabolic pathways; gene – gene and gene – metabolite interactions. |
| 73 | RxNorm | drugs vocabulary | Normalized naming system for generic and branded drugs; relations between drugs and their ingredients. |
| 74 | Semantic Medline | text mining results of Pubmed abstracts | Textmining of Pubmed Abstracts with ontologies in the platform to uncover biologicial/ biochemical relations. |
| 75 | SGD | genetic annotations for Saccharomyces cerevisiae | Functional annotations for Saccharomyces cerevisiae genes related to genetic function and organism function. |
| 76 | SNOMED CT | clinical health vocabulary | Normalised naming system for capturing and exchanging clinical health records, relating to diagnostics and treatments. |
| 77 | String | Protein interactions | specific gene/protein relations (inhbits, stimulates, controls expression) . Scoring model (0 – 1) is mapped to scientific values parameter (1 and 4) |
| 78 | TAC | Taconic Biosciences | Relation mouse model / strain - genotype from relational database |
| 79 | TAIR | genetic and molecular biology for model organism Arabidopsis Thaliana | Specific manual annotations for genes, genotypes and phenotypes based on gene ontology and plant ontology |
| 80 | TIGM | Texas A&M Institute of Genome Medicine | Relation mouse model / strain - genotype from relational database |
| 81 | Trait ontology | vocabulary of plant traits | Vocabulary of plant traits and relationships between the traits. |
| 82 | Transfac Public | gene regulation | Gene regulation ; gene – controls expression of - gene |
| 83 | Uberon | Cross species anatomy ontology | Annotations between genes and anatomicalentities and relations between anatomical entities, mapped for different species. |
| 84 | UMLS | biomedical vocabulary | Broad biomedical vocabulary containing annotated relations between genes, disease, phenotypes, clinical diagnosis and treatments. |
| 85 | Uniprot | proteomics | Curated annotations of proteins to sequence and structures, protein characteristics, location and involvement in disease. |
| 86 | Vertebrate trait ontology |  | Curated annotations of treats across model organisms and mappings to species specific phenotypes |
| 87 | Clinical Measurement Ontology |  | Vocabulary for clinical measurements of phenotypes and the underlying relations between these measurements. Enables relations between genes, phenotypes and their measurements. |
| 88 | GTEx | RNA-seq Expression data in healthy tissue | Used in expression analysis algorithms |
| 89 | Fantom 5 | CAGE expression data in healthy tissue | Used in expression analysis algorithms |
| 90 | The Cancer Genome Atlas (TGCA) | RNA-seq Expression data in cancer patients | Identification of genes with a high expression (mean Z-score >2.5) in the majority of cancer patients (> 60%).  gene – expressed in – cancer |
| 91 | SMPDB | Small molecule pathway database | Curated relations from Metabolic pathways between genes, substrates and products. |
| 92 | Wikipathways | Manually curated pathway database | Curated relations from metabolic and signaling pathways between genes, substrates and products. |
| 93 | Netpath | Manually curated pathway database | Curated relations from signal transduction pathways. |
| 94 | INOH | Manually curated pathway database | Curated relations from signal transduction pathways. |
| 95 | Foodb | Database for food and ingredients | Curated relations for the chemical composition of food and metabolites. |