

TOTAL NUMBER OF INTEGRATED DATA SOURCES: 200+

Euretos provides an AI platform used by pre-clinical researchers for in-silico discovery & validation of targets and biomarkers. World leading pharma, biotech and academic institutions use it to accelerate their multi omics research. Based on the latest big data analytics technologies, researchers discover and evaluate how molecular mechanisms influence cell and tissue functions, and in turn mediate phenotypes and disease pathology.



The platform integrates and interconnects three types of data sources:

- **Life Sciences databases** – These are sources that life scientists directly consult some of which are widely known such as Pubmed, Uniprot, ChEMBL and many other biological sources.
- **Textual & publication sources** – These are text based sources such as Pubmed, patent data and clinical records
- **Semantic and Ontological Sources** – These are sources like ontologies, structured vocabularies and thesauri that are often not consulted directly by life scientists but play a crucial role in classifying life sciences data and recognizing individual concepts.

Below is a selection from these sources. A complete list of all semantic, ontological and life sciences sources is available to our customers on request where we provide full disclosure.

Source	Source
APB	Australian Phenome Bank
Aracyc	Pathways in Arabidopsis Thaliana
ARC	Animal Resources Centre
Baseline Expression Atlas	RNA expression in healthy tissue
BIND	Molecular interactions
BindingDB	Biochemical interaction data for small molecules and protein targets
Biogrid	Protein-protein interactions and protein- small molecule interactions
Biomodels	Computational models of biological processes
BioPlex	Protein interaction networks
CARD	Centre for Animal Resources and Development
CellLine	Curated cell line overview
Chebi	Vocabulary for chemical entities of biological interest

ChEMBL	Compounds, targets and biochemical assays
Clinical Measurement Ontology	
Clinical Trials	Clinical trials
Clinvar	Clinical genomic variation
CMMR	Canadian Mouse Mutant Repository
Comparative Toxicogenomics Database	Gene - phenotype - chemical interactions
ConceptWiki	Vocabulary
CORUM	Protein complexes
CROP	Genetic annotations and diversity of traits and environmental treatments for plants
DECIPHER	Clinical phenotypic - genotypic data
Differential Expression Atlas	Gene expression patterns in different biological conditions
DIP	Protein interactions
DisGeNet	Gene - disease associations
Drugbank	Chemical, pharmacological and pharmaceutical information for approved and experimental drugs and protein therapies
EMMA	European Mouse Mutant Archive
Ensembl Plant Variation (Gramene)	Genetic variation and phenotype association for plants
Enzyme	Enzyme characterisation
Experimental Factor ontology	Ontology for variables in biological experiments
Fantom 5	CAGE expression data in healthy tissue
Foodb	Database for food and ingredients
Gene Ontology	Molecular functions, biological processes and cellular components
GTEx	RNA-seq Expression data in healthy tissue
GWAS	Curated resource of SNP-trait associations

HGNC	Nomenclature for human genes
HPRD	Protein interactions and pathways
Human Metabolome database	Small molecule metabolites
Human Phenotype Ontology	Phenotypic abnormalities encountered in human disease
Human Phenotype Ontology - Pubmed	Datamining results for Phenotype-Disease associations
Human Protein Atlas	RNAseq and protein abundance measurement in healthy tissues and cancer tissues; Gene classifications
Human Protein Map	Protein abundance measurements in healthy tissues
Humancyc	Molecular pathways
ICD10	Disease classification
INOH	Manually curated pathway database
Intact	Molecular interactions
IntAct Complex	Protein complexes
International Mouse Strain Resource (IMSR)	Mouse model resources
JAX	Strain database
KOMP	Knockout Mouse Project
MeSH	Vocabulary
miRTarBase	MicroRNA target interactions
MMRRC	Mutant Mouse Regional Resource Center
Monarch	Gene - phenotype associations from model organisms
Monarch Cross Species Phenotype Ontology	Ontology
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.
Mouse Phenotype Ontology	Ontology

MRC Harwell	Harwell Science and Innovation Campus
MUGEN	Mugen Mouse Database
National Cancer Institute	Vocabulary for cancers, agents and related substances
NCBI gene annotation	Gene information and annotation
NCBI gene orthologues	Gene orthologues
NCBI gene sequence	Reference sequences
NCI PID	Molecular pathways
NCIMR	National Cancer Institute at Frederick
Netpath	Manually curated pathway database
NIG	National Institute of Genetics (Japan)
OMIM	Genetic phenotypes
Orphanet	Rare diseases and orphan drugs
Panther	Protein classifications
Pathway ontology	Classic metabolic, regulatory, signaling, drug and disease pathways
Phosphosite Plus	Protein post-translational modifications
Phytozome	Gene orthologs in plants
Plant environment ontology	Vocabulary of plant environmental conditions and experimental treatments
Plant ontology	Plant anatomy, morphology and stages of development for all plants
Pubmed	Scientific literature; abstracts
RBRC	Riken BioResource Center
Reactome	Molecular pathways
Recon X	Human metabolism
RxNorm	Drugs vocabulary

Semantic Medline	Text mining results of Pubmed abstracts
SGD	Genetic annotations for <i>Saccharomyces cerevisiae</i>
SMPDB	Small molecule pathway database
SNOMED CT	Clinical health vocabulary
String	Protein interactions
TAC	Taconic Biosciences
TAIR	Genetic and molecular biology for model organism <i>Arabidopsis Thaliana</i>
The Cancer Genome Atlas (TCGA)	RNA-seq Expression data in cancer patients
TIGM	Texas A&M Institute of Genome Medicine
Trait ontology	Vocabulary of plant traits
Transfac Public	Gene regulation
Uberon	Cross species anatomy ontology
UMLS	Biomedical vocabulary
Uniprot	Proteomics
Vertebrate trait ontology	
Wikipathways	Manually curated pathway database