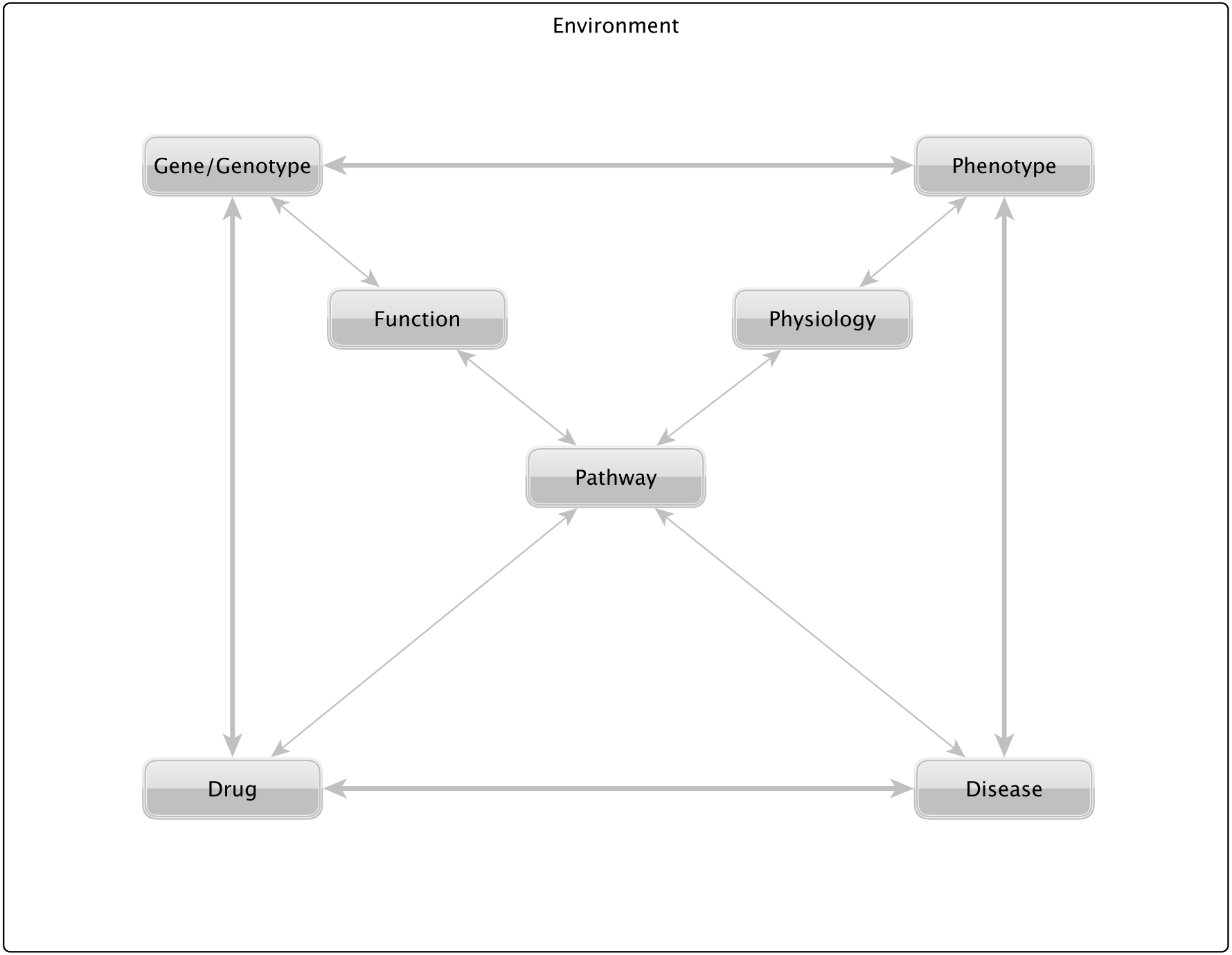
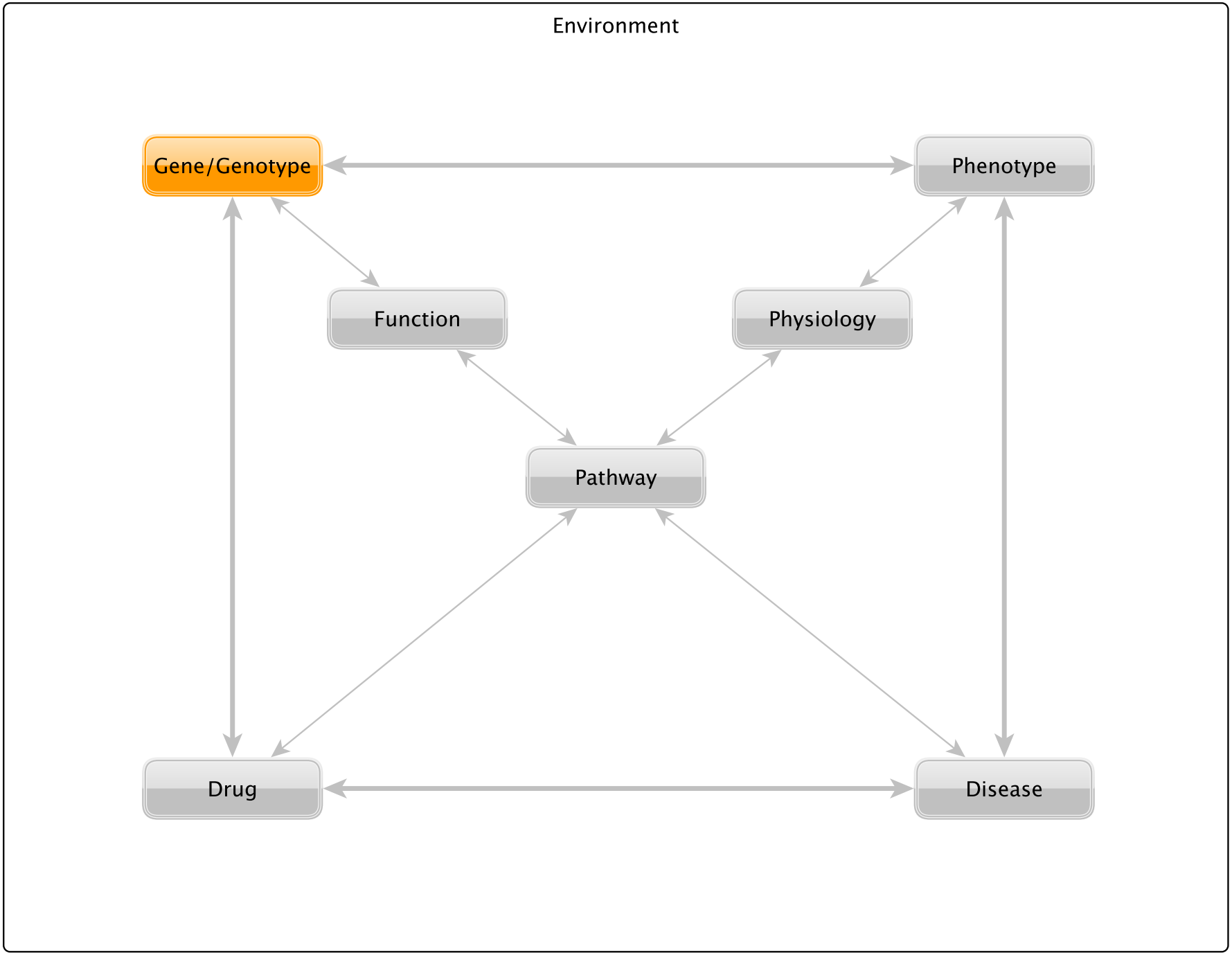
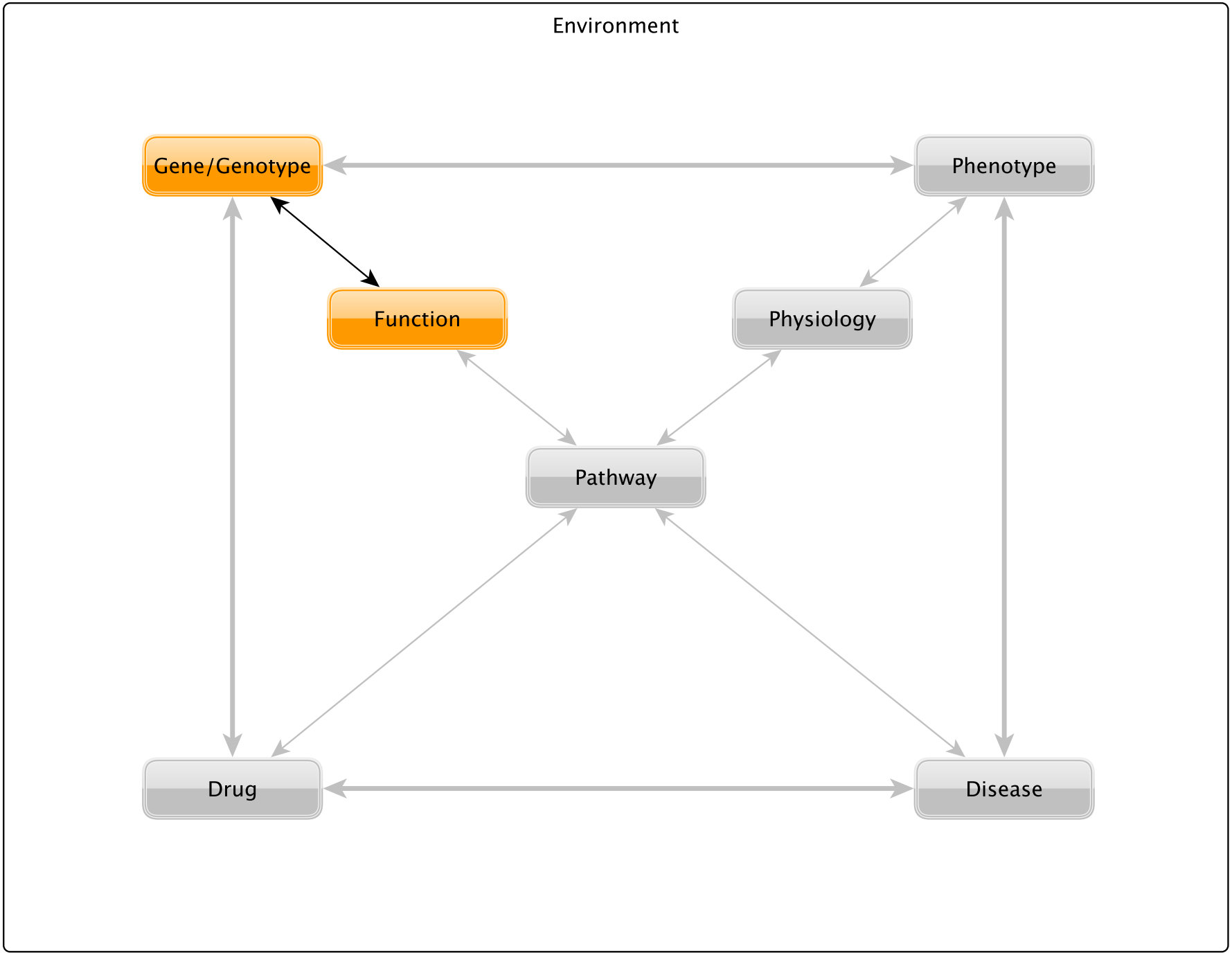
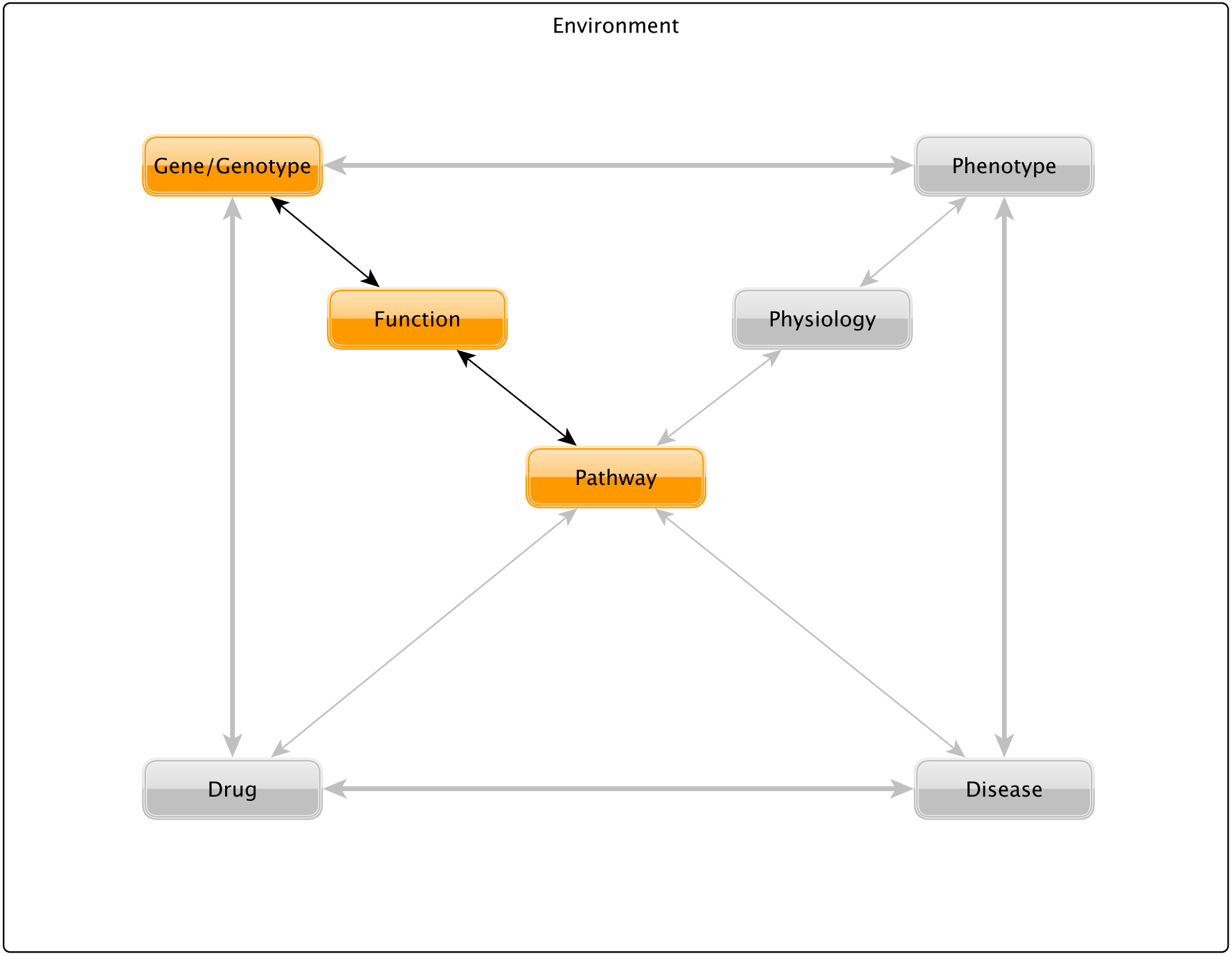
From Systems Genetics to Translational Medicine

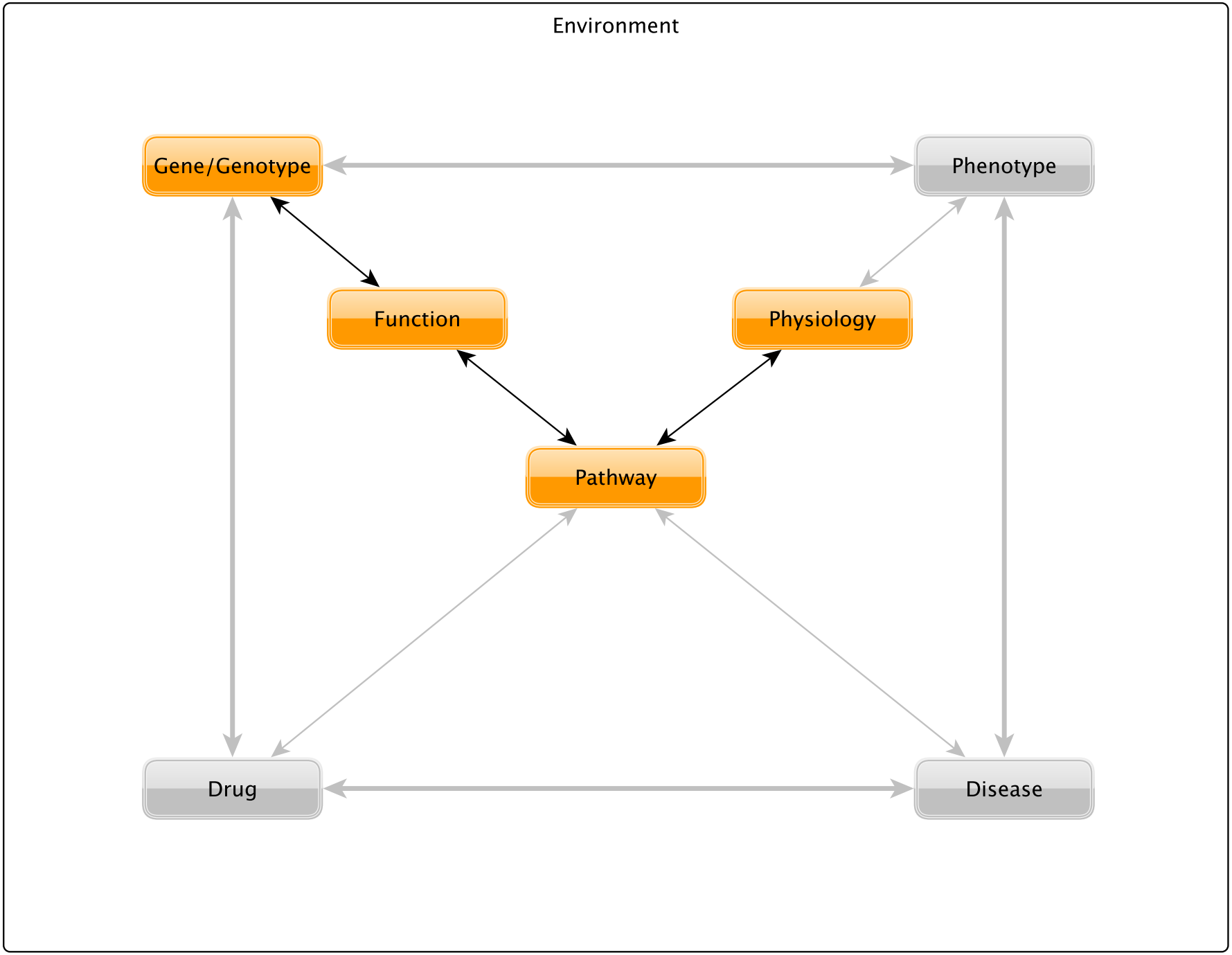
George Gkoutos

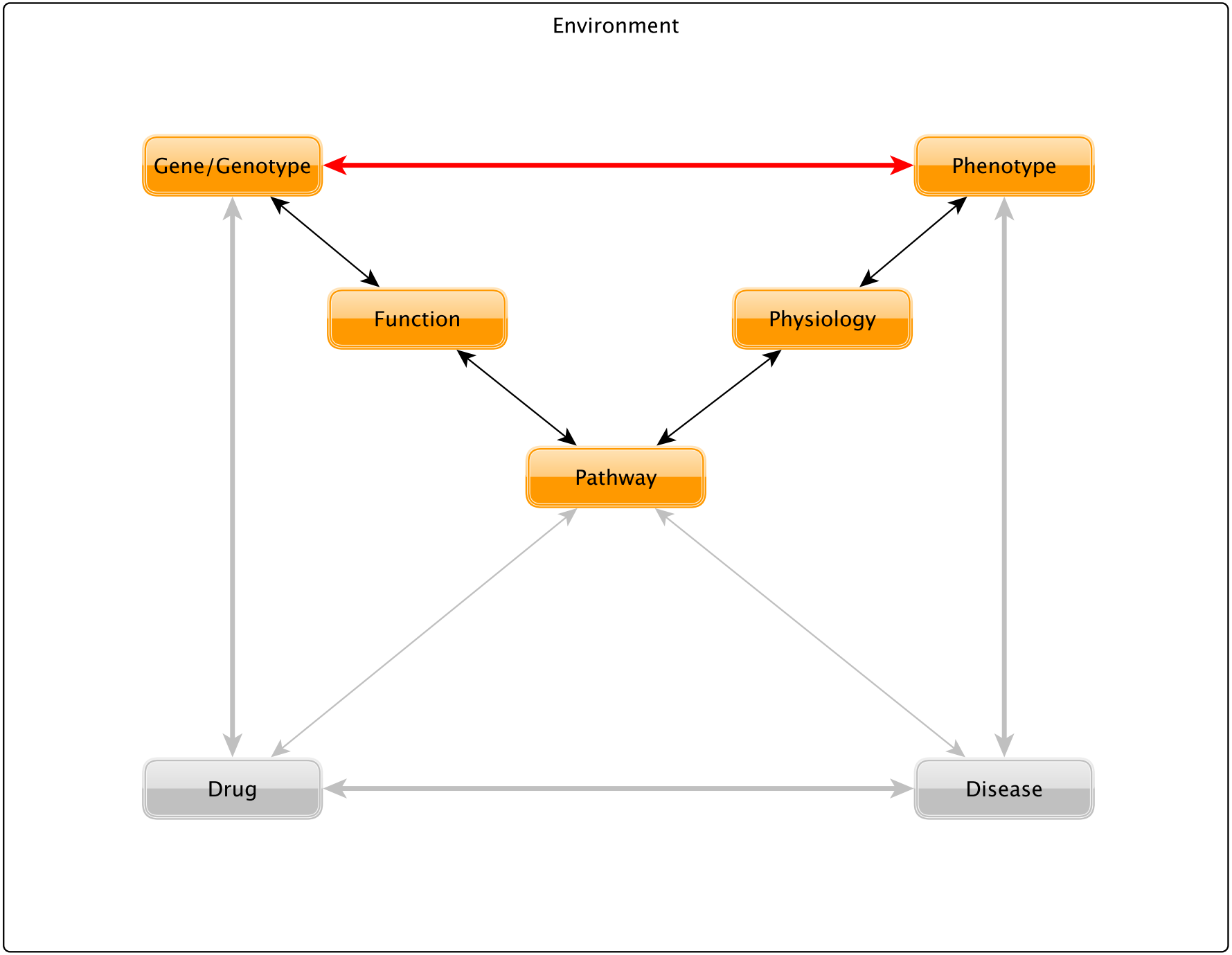

Environment
Gene/Genotype
Phenotype
Function
Physiology
Pathway
Drug
Disease

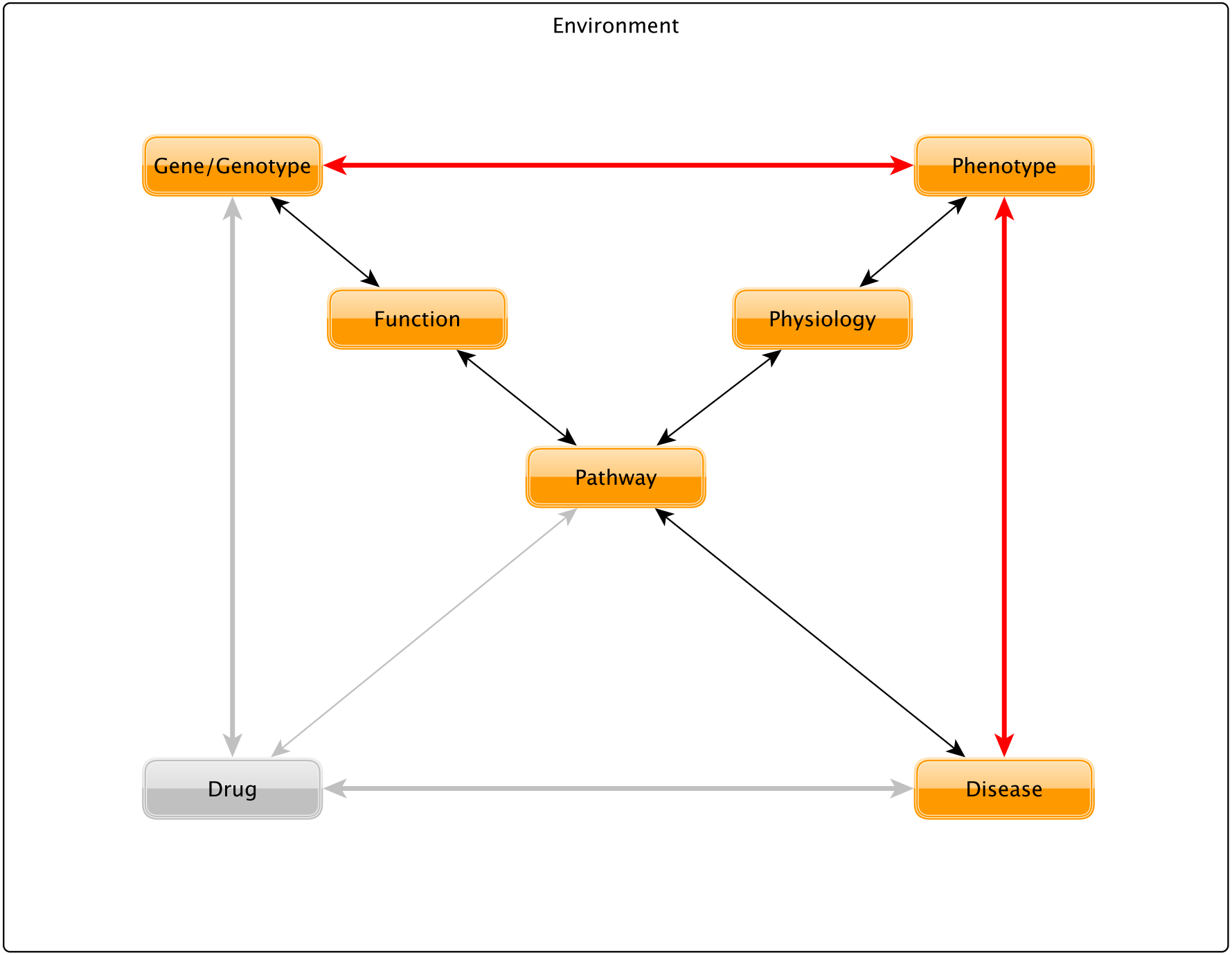

Environment
Gene/Genotype
Phenotype
Function
Physiology
Pathway
Drug
Disease

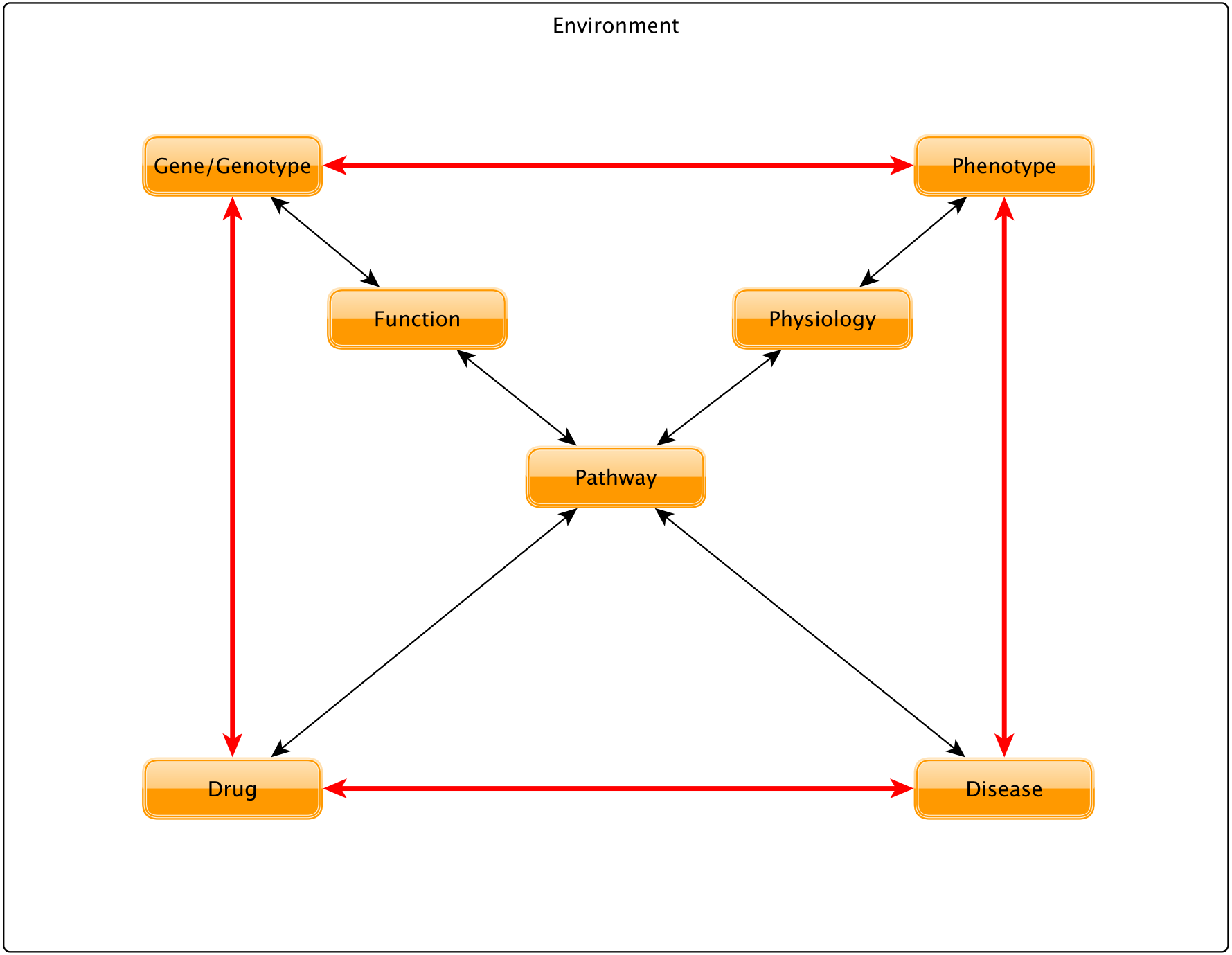

Environment
Gene/Genotype
Phenotype
Function
Physiology
Pathway
Drug
Disease


Environment
Gene/Genotype
Phenotype
Function
Physiology
Pathway
Drug
Disease

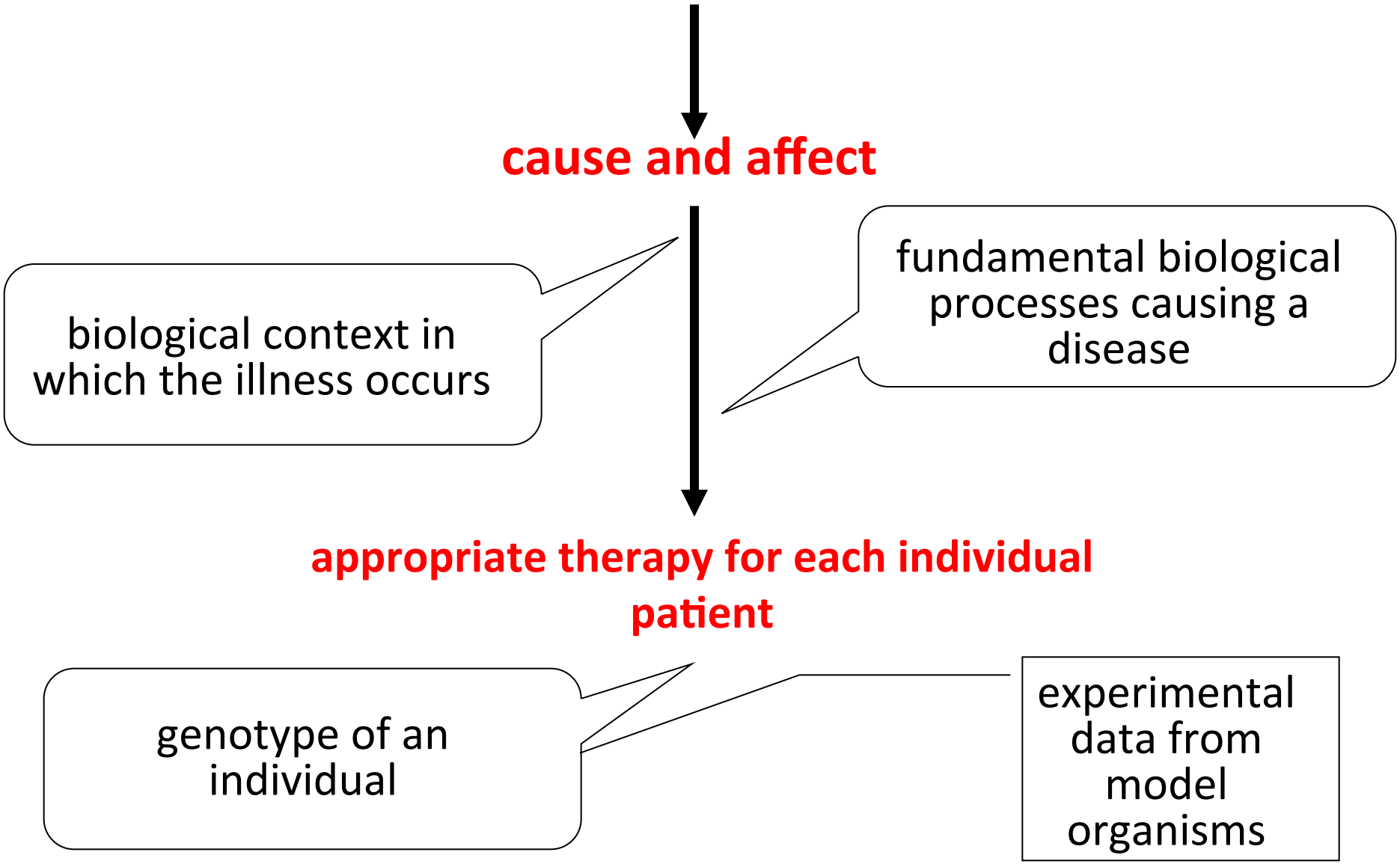

Environment
Gene/Genotype
Phenotype
Function
Physiology
Pathway
Drug
Disease


Environment
Gene/Genotype
Phenotype
Function
Physiology
Pathway
Drug
Disease

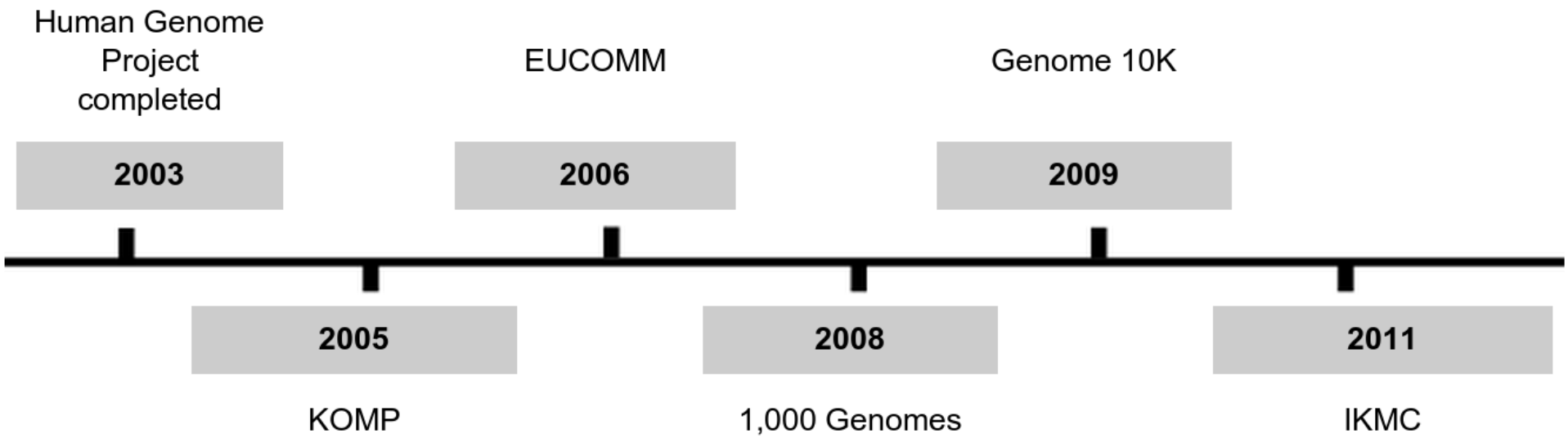

Environment
Gene/Genotype
Phenotype
Function
Physiology
Pathway
Drug
Disease


Environment
Gene/Genotype
Phenotype
Function
Physiology
Pathway
Drug
Disease

Gain an understanding of the molecular basis of human disease



Exploring the Phenome



Key EU/NIH missions:

- integration and analysis of disease data within and across species → diagnostic and therapeutic advances at the clinical level
- identification of causative genes for Mendelian orphan diseases

Power of the Phenotype

The meaningful cross species and across domain translation of phenotype is essential → phenotype-driven gene function discovery and comparative pathobiology

Goal - "A platform for facilitating mutual understanding and interoperability of phenotype information *across*

- **species**,
- **domains of knowledge**,

and amongst people and machines" …..

Phenotype And Trait Ontology (PATO)

- phenotypes may be described in many different dimensions, e.g.
  - the biochemical ('alcohol dehydrogenase null')
  - the cellular ('cell division arrested at metaphase’)
  - the anatomical ('eye absent')
  - the behavioral (‘hyperactive’)
  - etc.

- in whatever dimension and granularity, however, there is a commonality so that phenotypic descriptions can be decomposed into two parts

  - An **entity** that is affected. This entity may be an enzyme, an anatomical structure or a complex biological process.

  - The **qualities** of that entity.

PATO today

PATO is now being used as a community standard for phenotype description

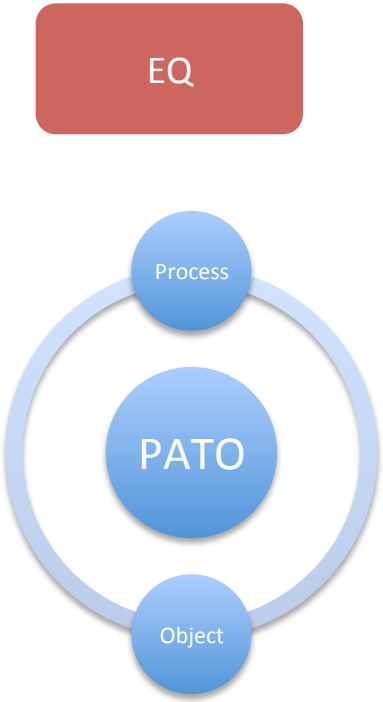
- many consortia (e.g. Phenoscape, The Virtual Human Physiology project (VPH), IMPC, BIRN, NIF)
- most of the major model organism databases, (e.g. example Flybase, Dictybase, Wormbase, Zfin, Mouse genome database (MGD))
- international projects

PATO's Semantic Framework

- Conceptual Layer
- Semantic Components Layer
- Unification Layer
- Integration Layer

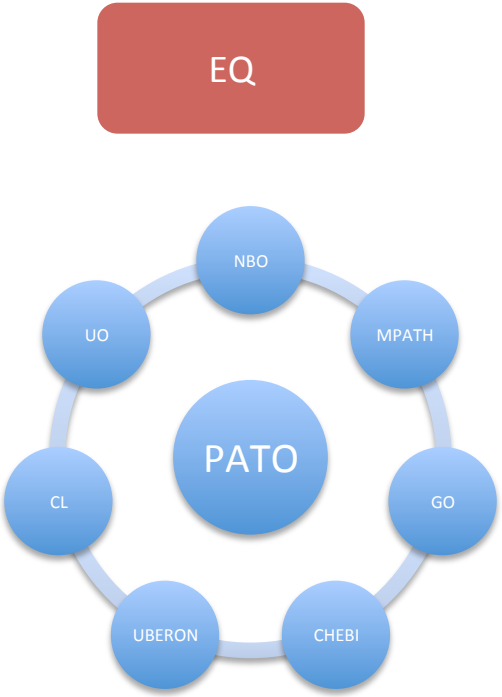
PATO Conceptual Layer

EQ Model
link Entities (E) from GO, CheBI, FMA etc. to Qualities (Q) from PATO →
EQ statements



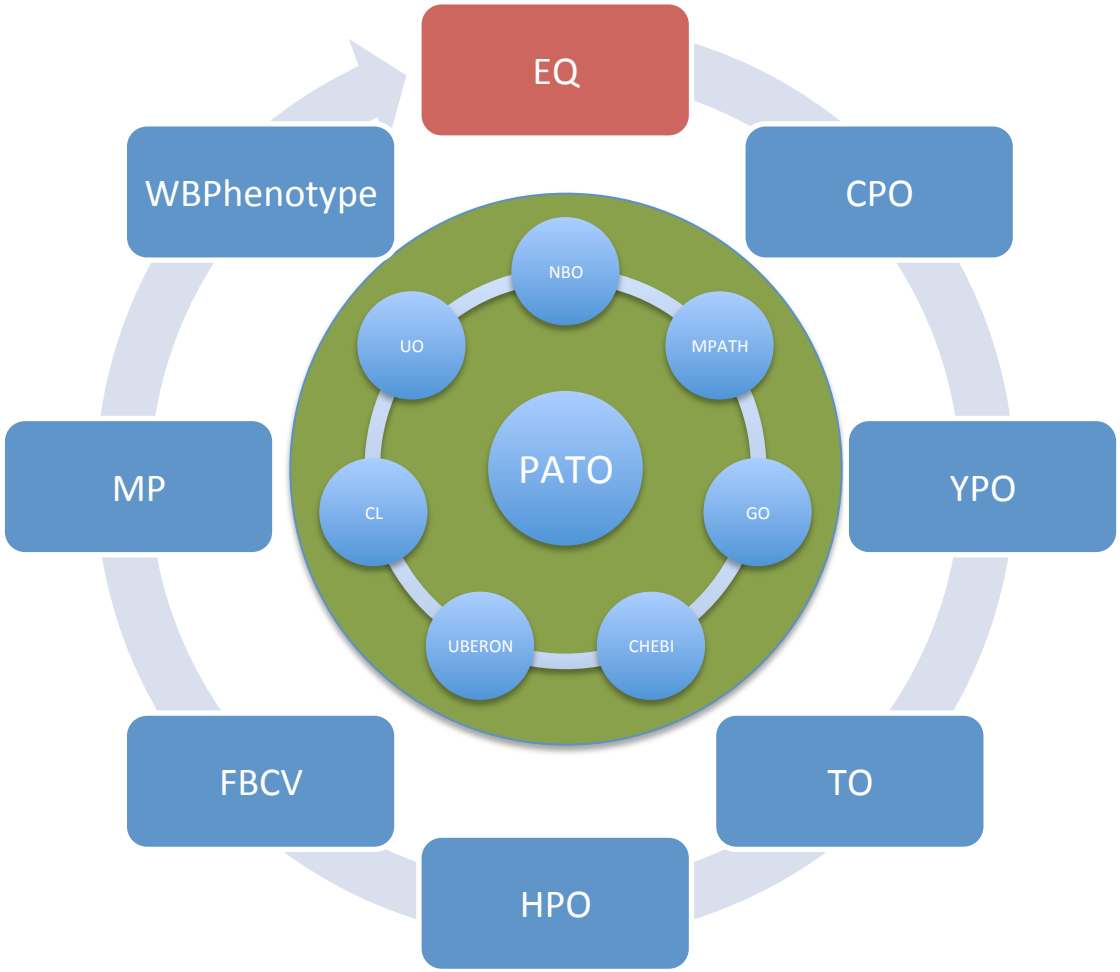
Semantic Components Layer

- Behavior
- Cell Phenotype
- Pathology
- UBERON
- Measurements (Units Ontology)

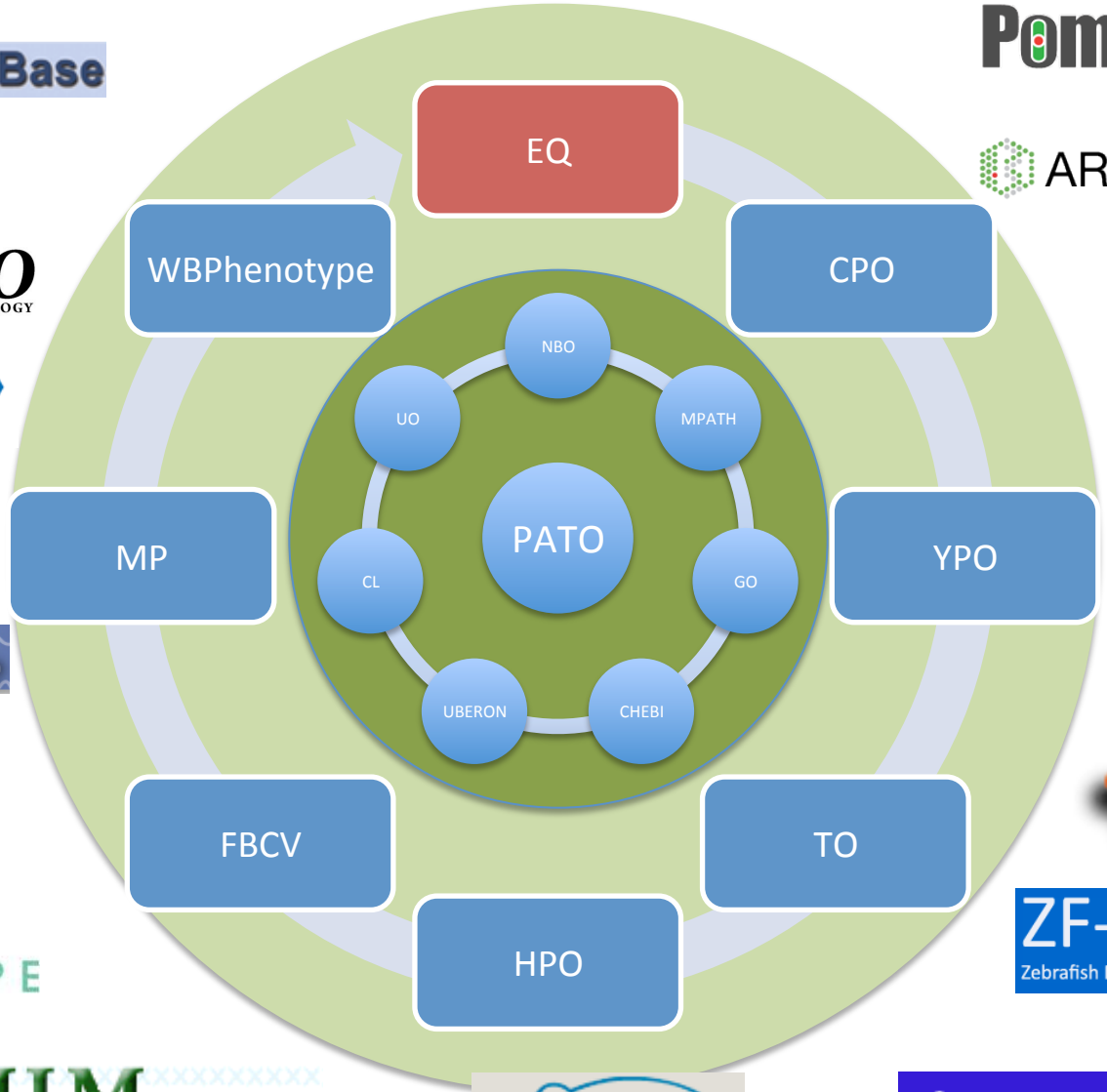


Unification Layer

Provision of PATO based equivalence definitions



Integration Layer



Pathbase
European mutant mouse pathology database

dictyBase

RICORDO
INTEROPERABLE ANATOMY AND PHYSIOLOGY

EMPRESS

QB

WormBase

MGI

PHENOSCAPE

OMIM
Online Mendelian Inheritance in Man

europhenome

RIKEN

PomBase

ARRAYEXPRESS

EUMODIC

IMPC
International Mouse Phenotyping Consortium

eumorphia

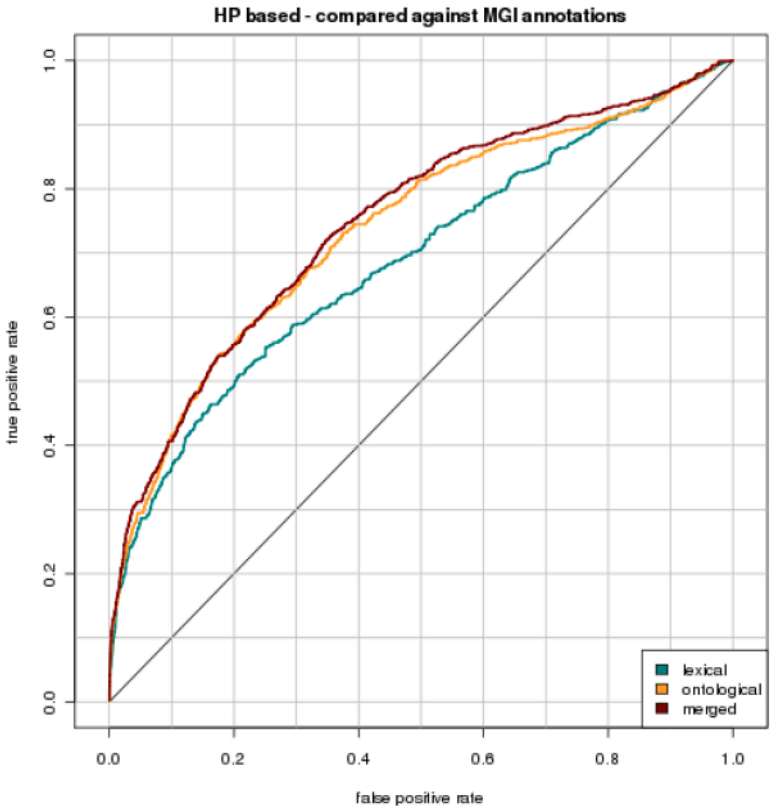
SGD

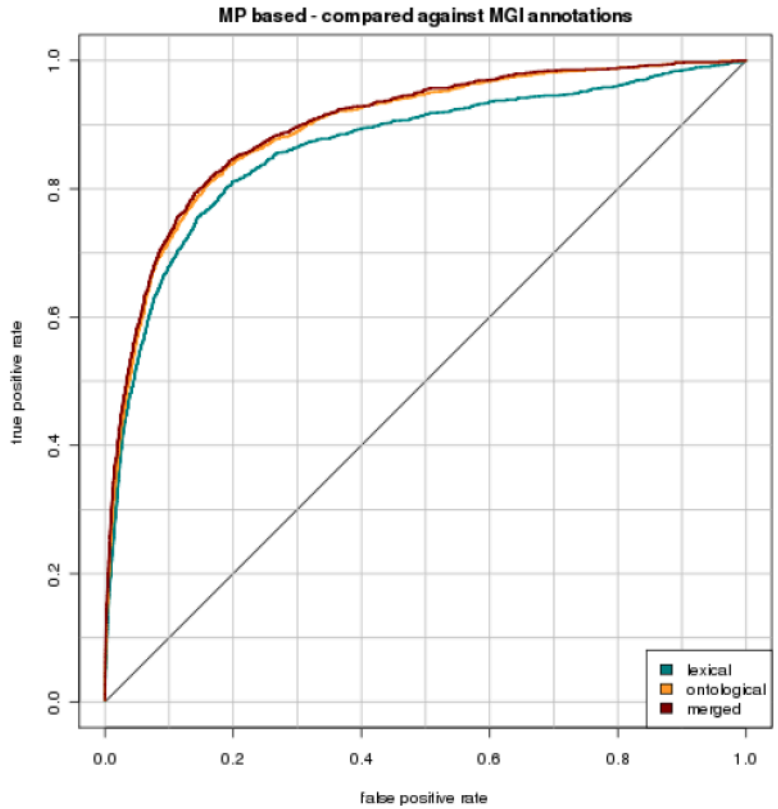
ZF-HEALTH
Zebrafish Regulomics for Human Health

Cross Species Data Integration

Cross species integration framework

- A PATO-based cross species phenotype network based on experimental data from 5 model organisms yeast, fly, worm, fish and mouse and human disease phenotypes (OMIM)
- integration of anatomy and phenotype ontologies
  - more than 500,000 classes and 1,500,000 axioms
- PhenomeNET forms a network with more than 111.000 complex phenotype nodes representing complex phenotypes
- Semantic phenotype similarity - pairwise comparison of disease and animal phenotypes

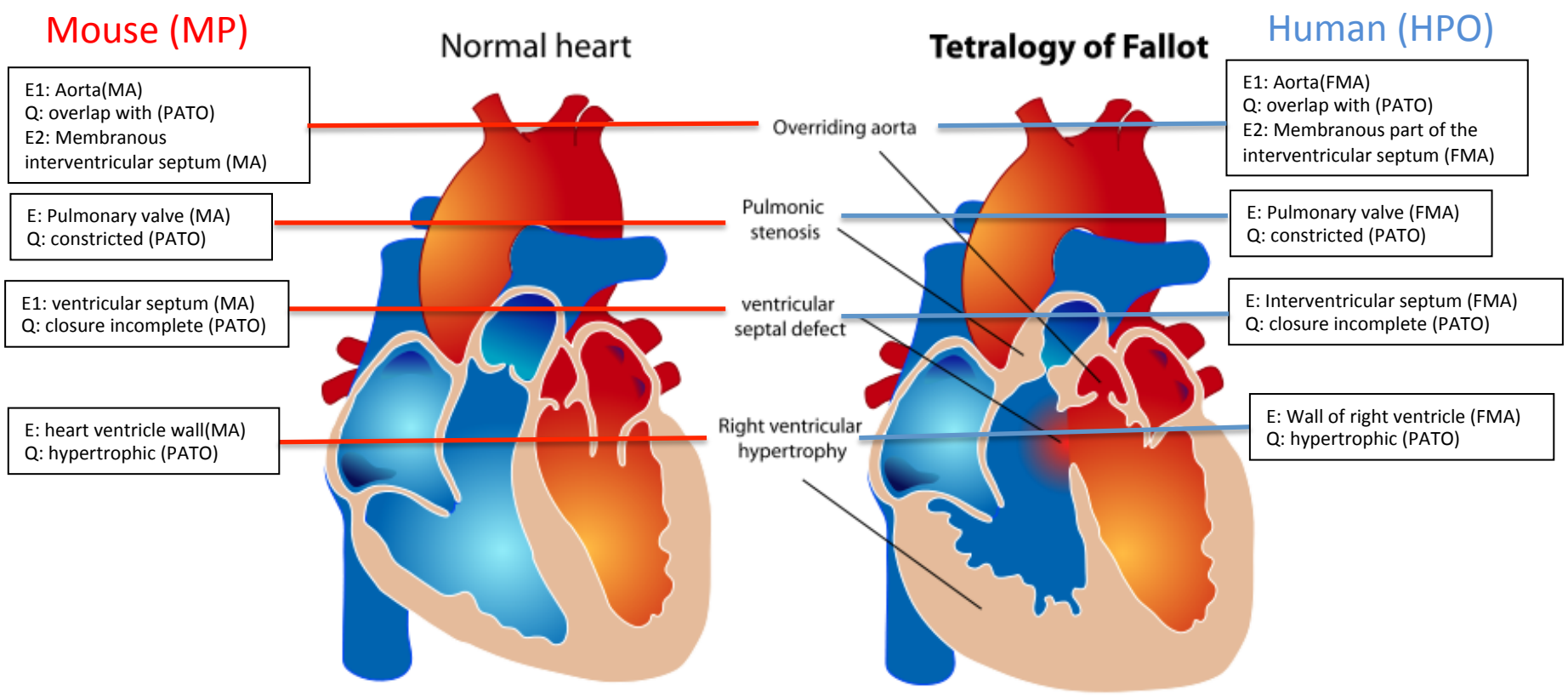




AUC = 0.9

- Enhance the network e.g.
  - Semantics e.g Behavior and pathology related phenotypes etc.
  - Methods e.g. text mining, machine learning etc.
  - Resources e.g. OrphaNet
- PhenomeNET now significantly outperforms previous phenotype-based approaches of predicting gene–disease associations
- Performance matches gene prioritization methods based on prior information about molecular causes of a disease

Candidate disease gene prioritization



- Predict all known human and mouse disease genes
- Adam19 and Fgf15 mouse genes
- using zebrafish phenotypes - mammalian homologues of Cx36.7 and Nkx2.5 are involved in TOF

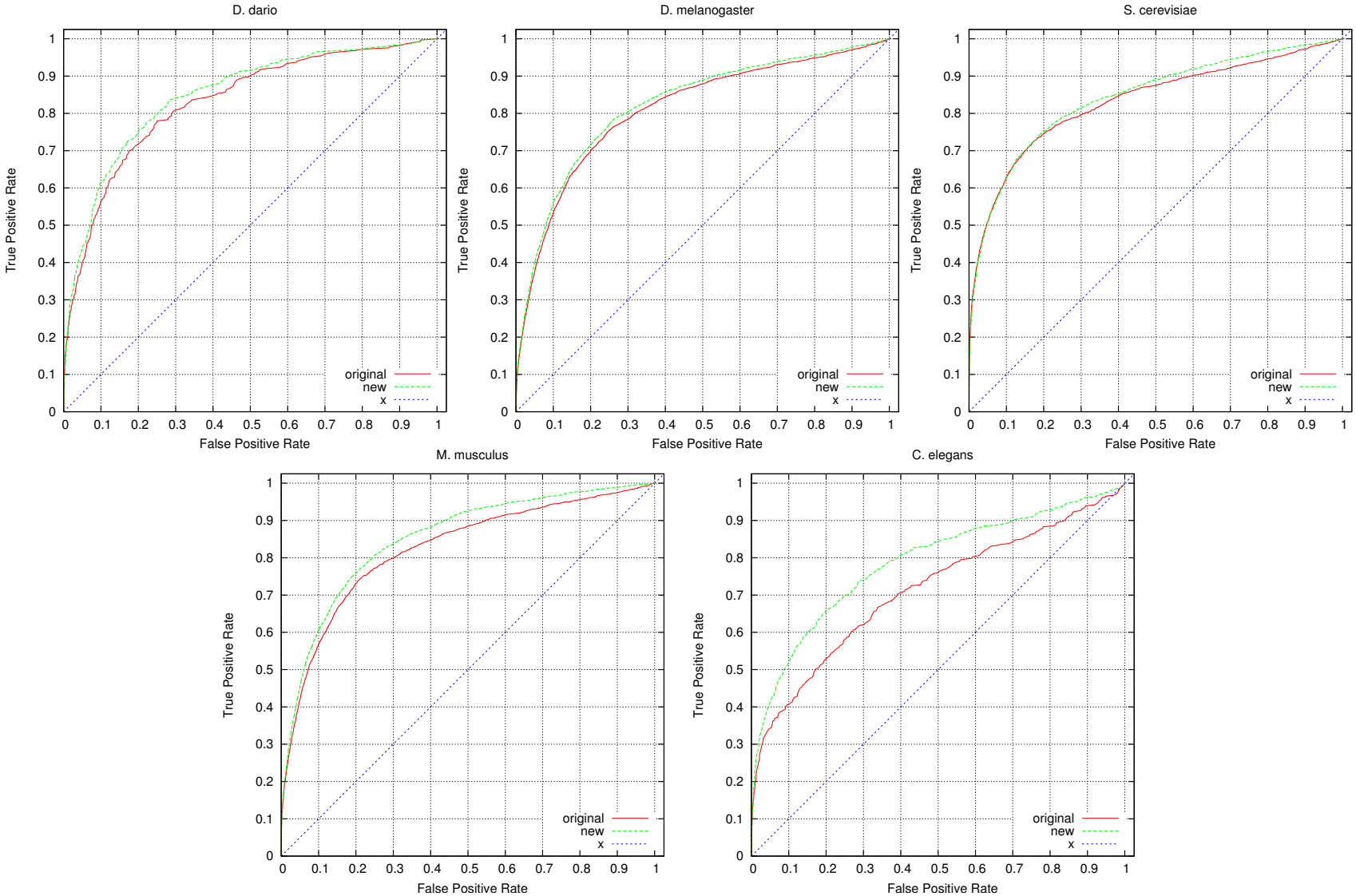
Data Integration Across Domains of Knowledge

Gene function determination

- phenotype data are commonly annotated, automatically, semi-automatically and manually, with terms from species-specific phenotype ontologies,

- utilize experimental data → infer the functions (impaired given a phenotype observation)

- phenotype data from 5 different species → more than 30000 novel gene functions
- evaluation:
  - manually for biological correctness
  - predict genetic interactions (semantic similarity between genes)



Species	Inferred	p-value
Fish	1717	0.24
Yeast	14047	0.162
Fly	1633	0.041
Worm	9221	0.03
Mouse	15693	0.00007

Data Integration Across Species and Across Domains of Knowledge

Rare and orphan diseases

Number of Entries:

Prefix	Autosomal	X Linked	Y Linked	Mitochondrial	Totals
* Gene description	12,750	627	48	35	13,460
+ Gene and phenotype, combined	250	14	0	2	266
# Phenotype description, molecular basis known	2,836	240	4	28	3,108
% Phenotype description or locus, molecular basis unknown	1,628	135	5	0	1,768
Other, mainly phenotypes with suspected mendelian basis	1,819	130	2	0	1,951
Totals	19,283	1,146	59	65	20,553

- at least 4000 diseases without known molecular basis
- disease-gene identification methods have a limited focal range that may include up to 300 genes

- necessary to suggest possible causative genes

Identifying causative genes for rare and orphan disease

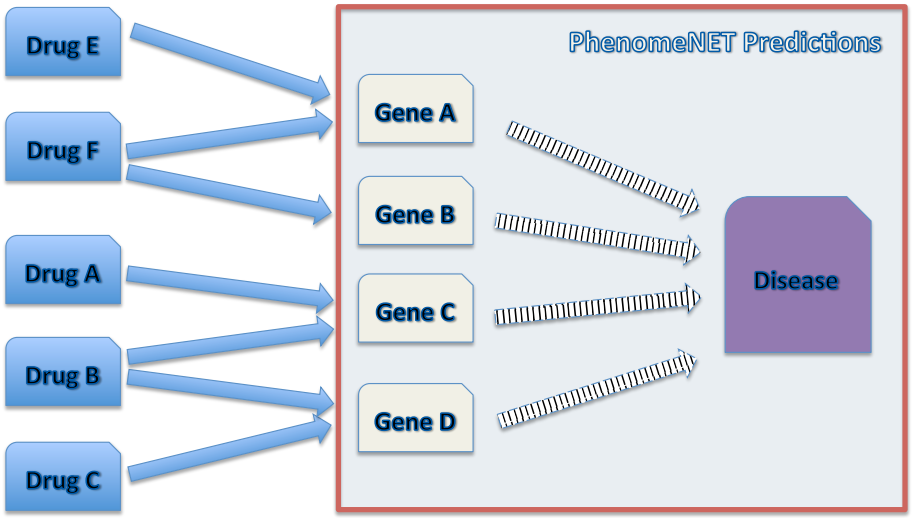
- Bassoe syndrome (OMIM:25400) (visual, musculoskeletal, reproductive systems, behavior)
  - HIP1 gene → Huntingtin-Interacting Protein 1 (HIP1)
  - interacting partner of huntingtin → Huntington Disease
  - expressed in many tissues and different brain regions
  - involved in clathrin-mediated endocytosis of cell surface receptors
  - role in development and tumorigenesis

Affected System	Bassoe Syndrome	HIP1 -/- mice
Visual System	cataract	cataract
Musculoskeletal System	kyphosis	spinal defects kyphosis kypholordosis
	muscle hypotrophy hypotonia	Developmental abnormalities failure to thrive muscle hypotrophy/wasting tremors
	cubitus valgus	
Reproductive System	hypogonadism	hypogonadism
	Hypogenitalism abnormal ovaries hypoplastic testis	decreased testicular weight due testicular degeneration increased apoptosis of postmeiotic spermatids and oligospermia decreased fertility ovarian abnormalities
Behavior	abnormal gait	abnormal gait ataxia
	neurobehavioral abnormalities intellectual disabilities, hyperactivity aggression epilepsy amimia	defects in presynaptic function delayed recovery from chemically induced long-term depression epilepsy hyperactivity aggression

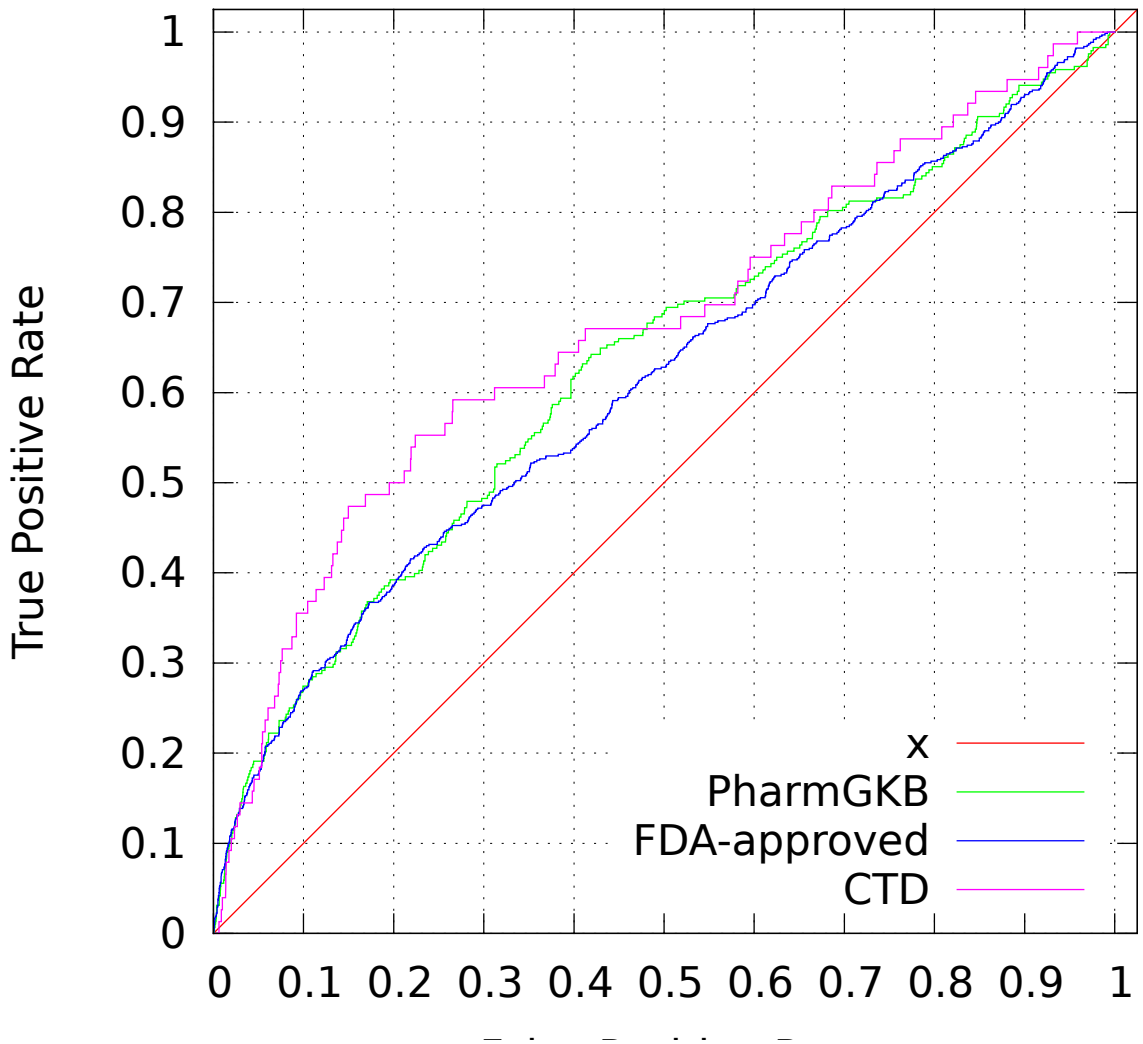
Novel drug discovery and repurposing

Variety of methods successfully being applied for drug repositioning and the suggestions of potentially novel drugs

Can a phenotype of gene which the drug interacts be used to predict diseases in which the drug is active?



Results



AUC =

0.67 PharmGKB
0.65 FDA
0.69 CTD

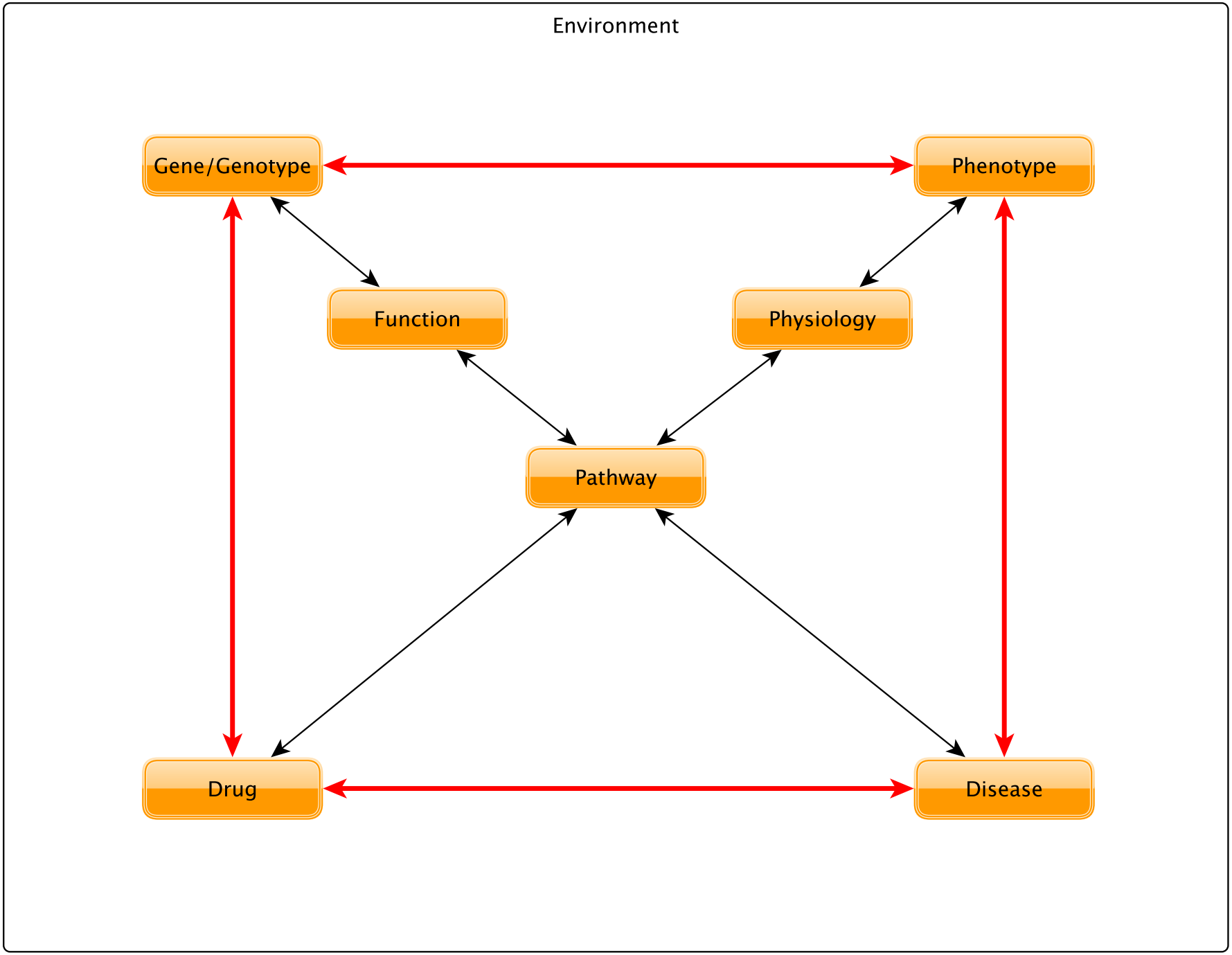
Disease and drug pathways

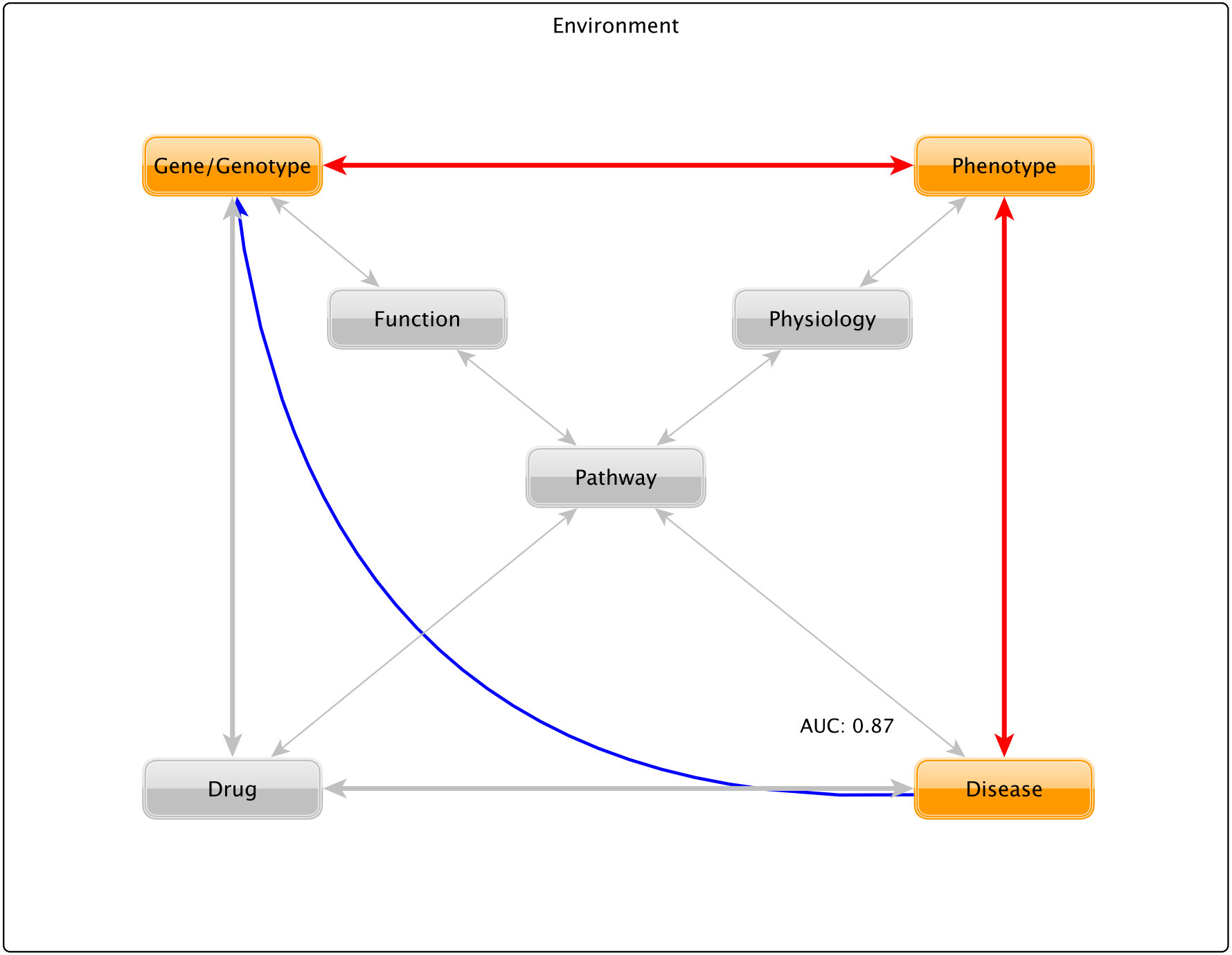
- integration of chemical ontologies and disease ontologies
  - Chemicals: ATC, ChEBI, MeSH (chemicals)
  - MeSH (diseases, phenotypes), HumanDO
- integration of databases containing drug, gene and disease information
  - DrugBank, PharmGKB, CTD, Pathway Interaction Database
  - associations: drug-gene, gene-disease, drug-disease, biological pathways

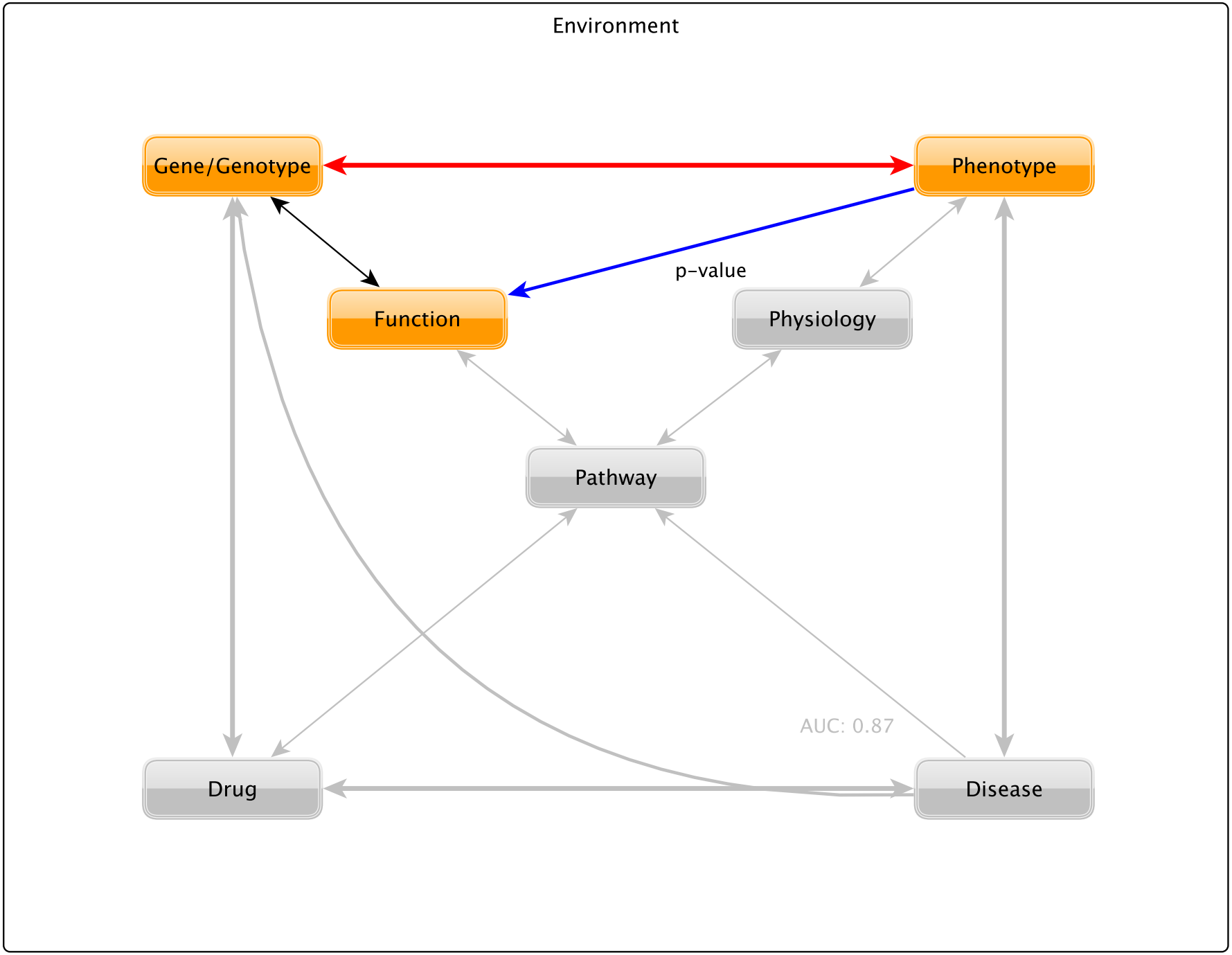
OntoFUNC – ontology enrichment analysis

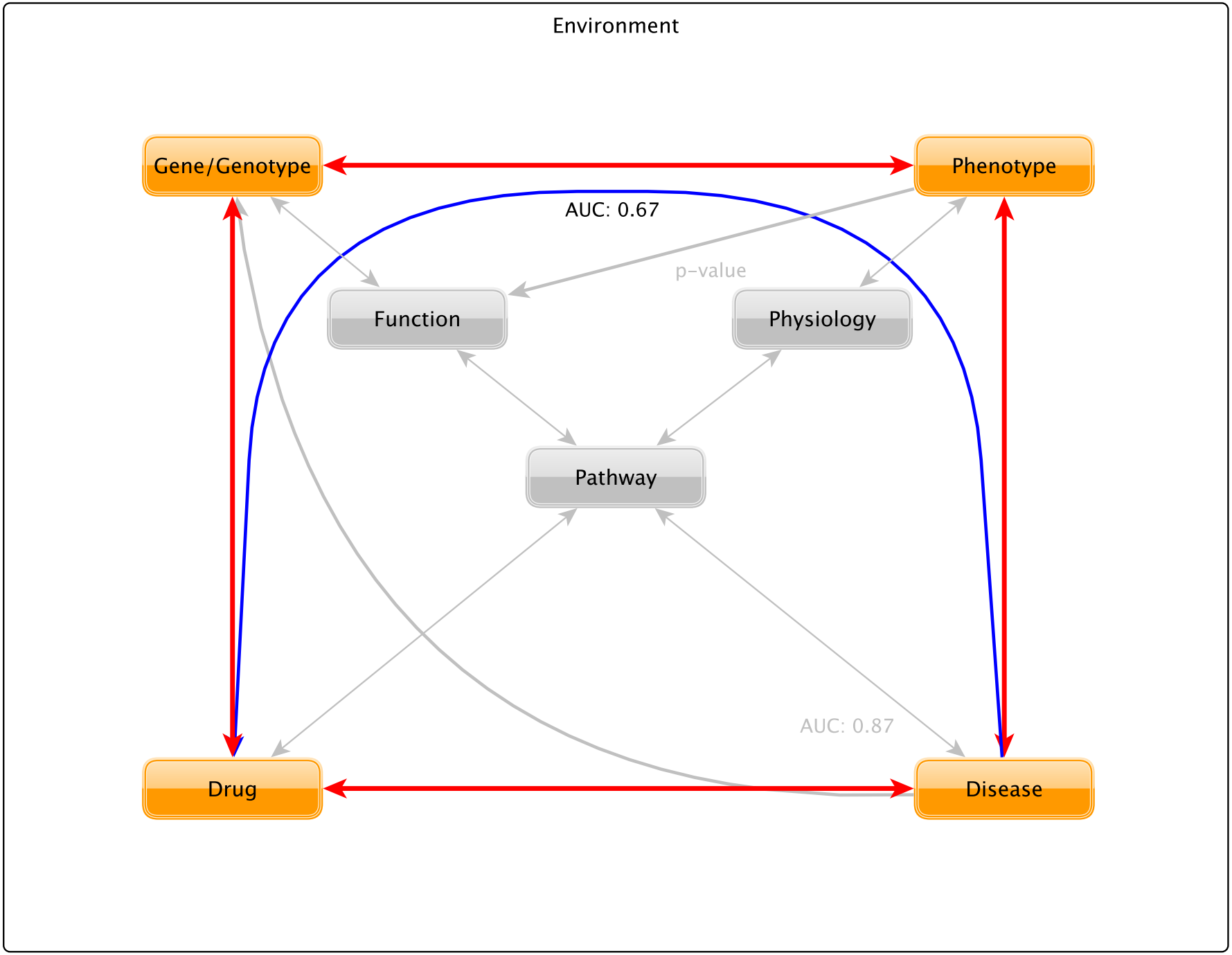
- use graph structure of ontology to identify statistically over- and under-represented ontology classes
- identify over-represented disease classes (from HumanDO) for genes in a pathway (disease pathways)
  - carcinosarcoma (DOID:4236) and Zidovudine Pathway(PharmGKB:PA165859361) ($p < 10^{-10}$)
- identify over-represented chemical classes (from ChEBI) for genes in a pathway (drug pathways)
  - mood disorder (DOID:3324) and Zidovudine Pathway **(**PharmGKB:PA165859361) ($p < 10^{-3}$)

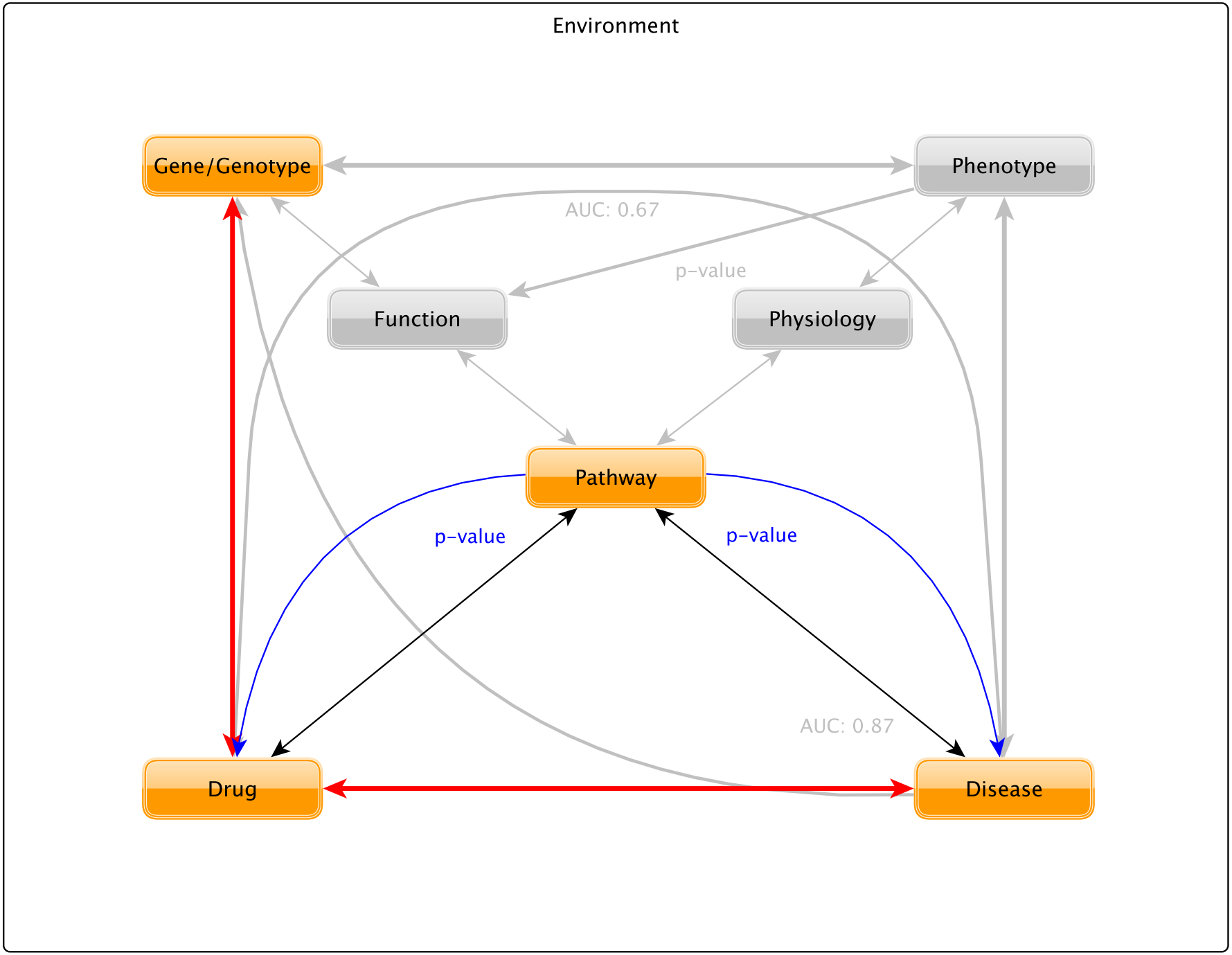
From Genotype to Phenotype to Disease and Intervention

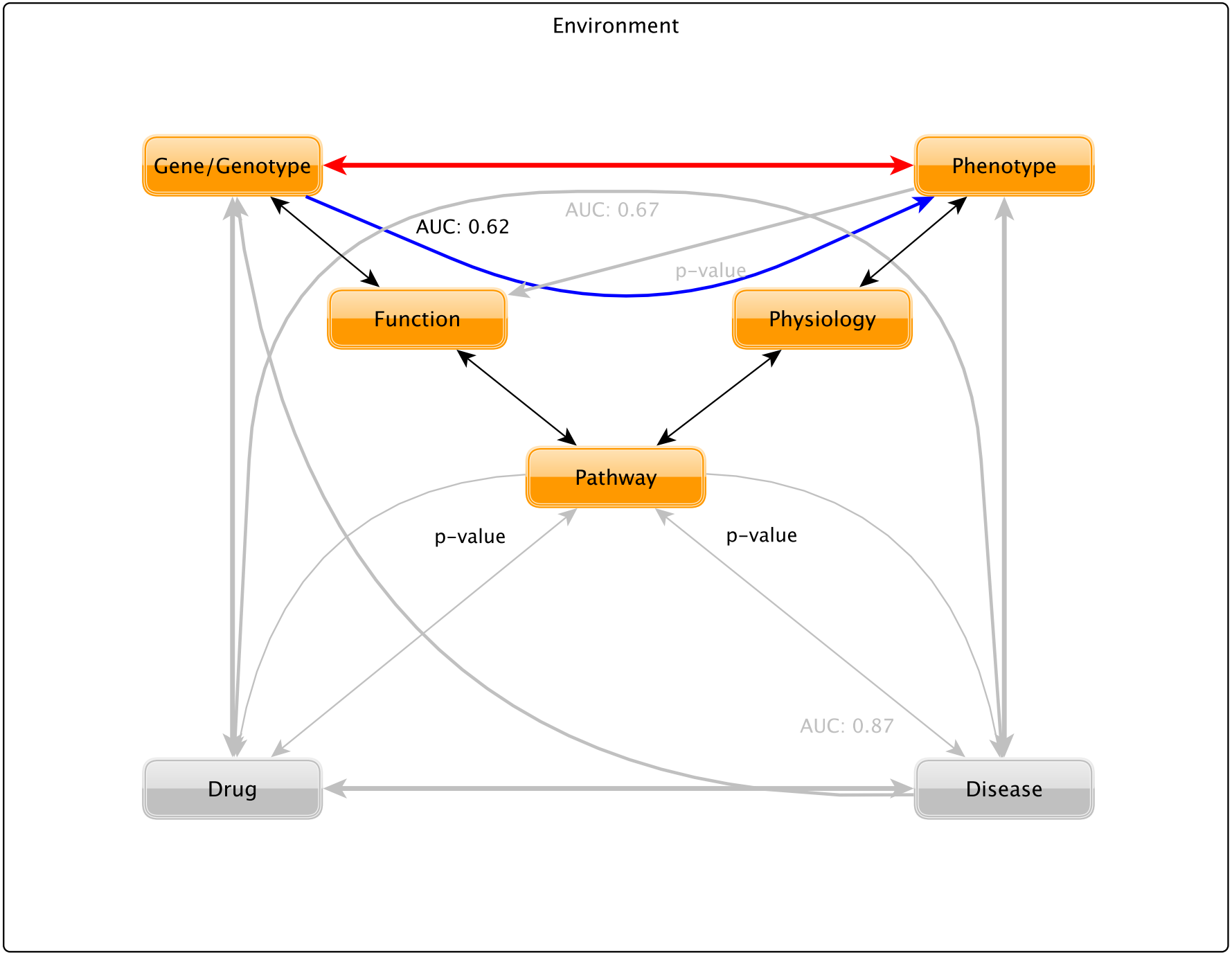

Environment
Gene/Genotype
Phenotype
Function
Physiology
Pathway
Drug
Disease

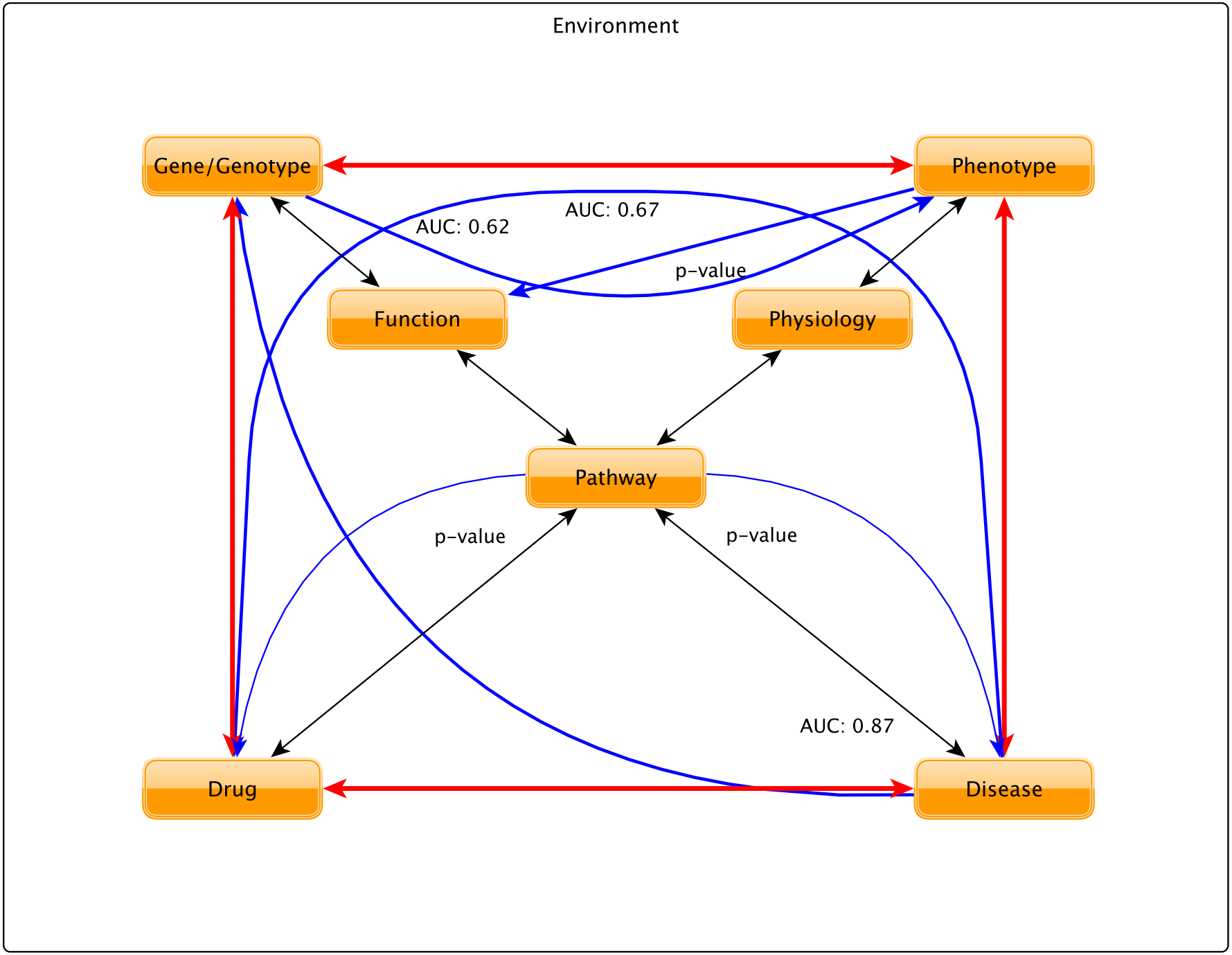

Environment
Gene/Genotype
Phenotype
Function
Physiology
Pathway
AUC: 0.87
Drug
Disease

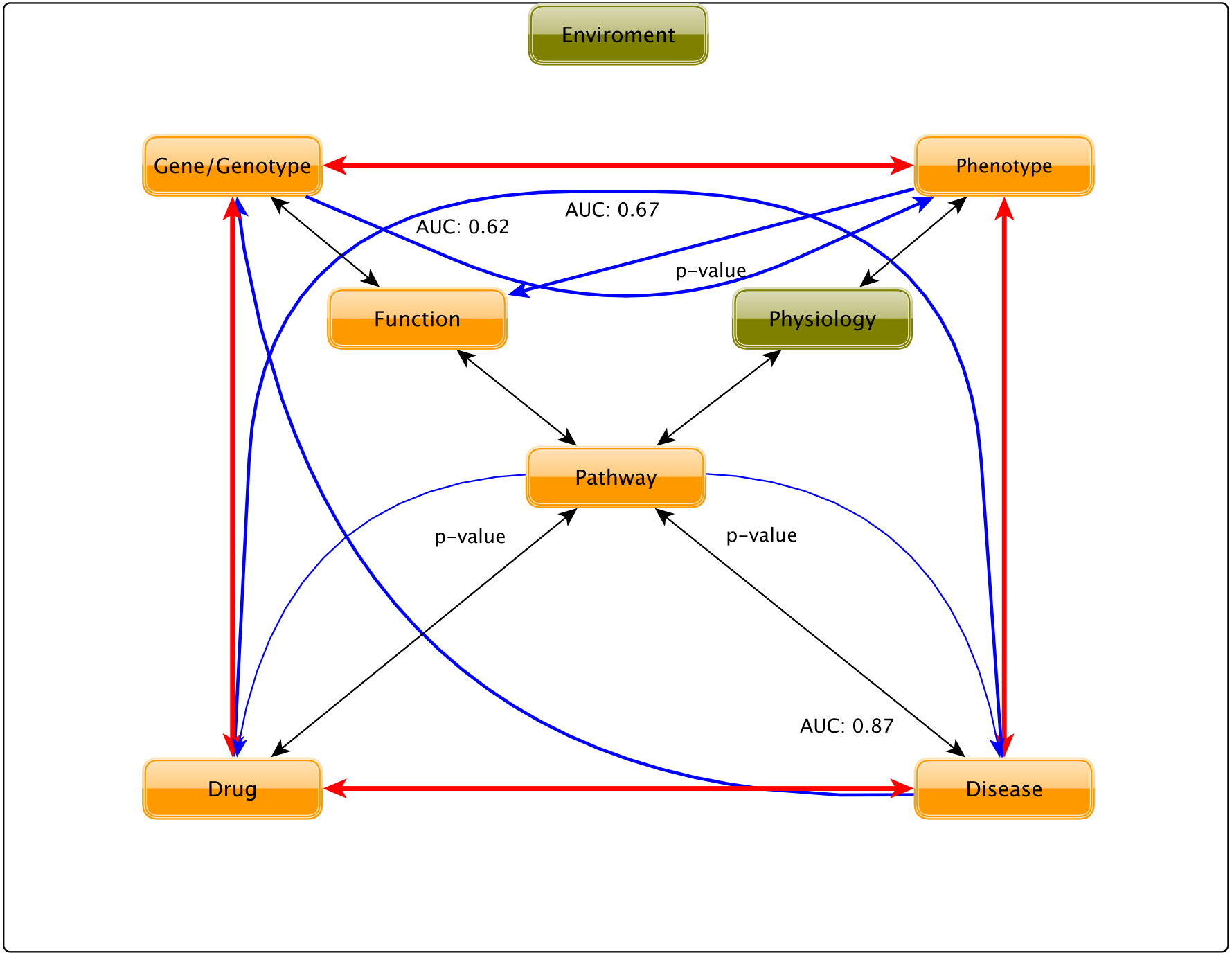

Environment
Gene/Genotype
Phenotype
p-value
Function
Physiology
Pathway
AUC: 0.87
Drug
Disease

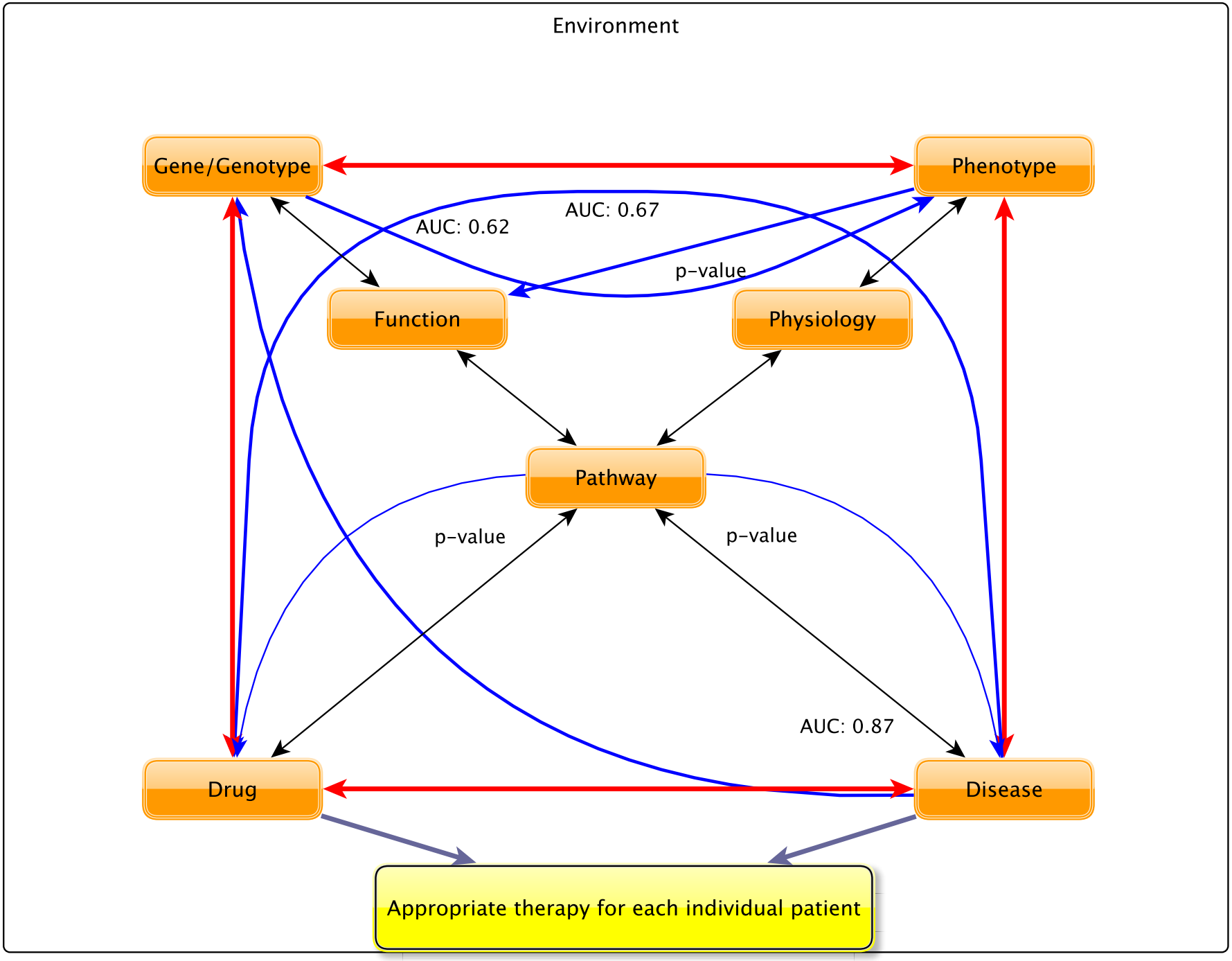

Environment
Gene/Genotype
Phenotype
AUC: 0.67
p-value
Function
Physiology
Pathway
AUC: 0.87
Drug
Disease

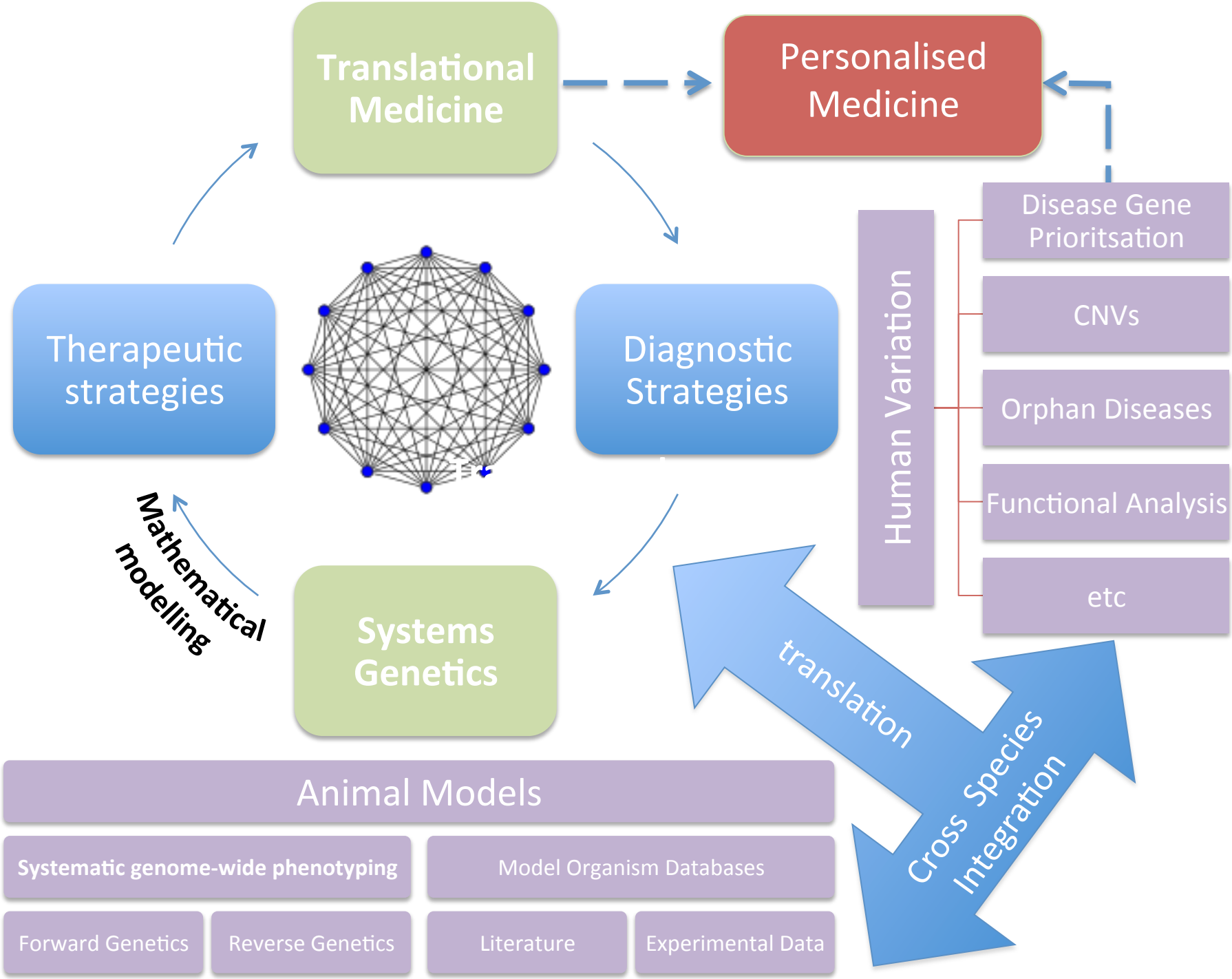

Environment
Gene/Genotype
Phenotype
AUC: 0.67
p-value
Function
Physiology
Pathway
p-value
p-value
AUC: 0.87
Drug
Disease


Environment
Gene/Genotype
Phenotype
AUC: 0.67
AUC: 0.62
p-value
Function
Physiology
Pathway
p-value
p-value
AUC: 0.87
Drug
Disease


Environment
Gene/Genotype
Phenotype
AUC: 0.67
AUC: 0.62
p-value
Function
Physiology
Pathway
p-value
p-value
AUC: 0.87
Drug
Disease


Enviroment
Gene/Genotype
Phenotype
AUC: 0.67
AUC: 0.62
p-value
Function
Physiology
Pathway
p-value
p-value
AUC: 0.87
Drug
Disease


Environment
Gene/Genotype
Phenotype
AUC: 0.67
AUC: 0.62
p-value
Function
Physiology
Pathway
p-value
p-value
AUC: 0.87
Drug
Disease
Appropriate therapy for each individual patient


Translational Medicine
Personalised Medicine
Therapeutic strategies
Diagnostic Strategies
Systems Genetics
Mathematical modelling
Human Variation
Disease Gene Prioritsation
CNVs
Orphan Diseases
Functional Analysis
etc
translation
Cross Species Integration
Animal Models
Systematic genome-wide phenotyping
Model Organism Databases
Forward Genetics
Reverse Genetics
Literature
Experimental Data