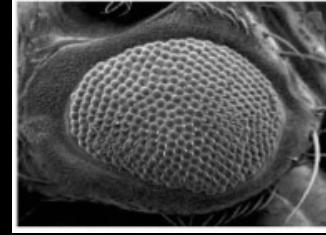
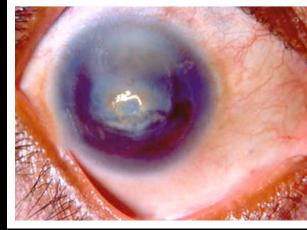


# Linking Animal Models and Human Diseases

EU-US Animal Biotechnology Working Group Workshop

Monte Westerfield - November 8, 2011



Supported by NIH P41 HG002659, U54 HG004028, & R01 HG004838

Berkeley, Cambridge & Oregon Universities

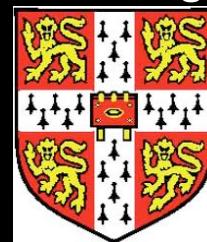


**Yvonne Bradford**  
**Melissa Haendel**  
**Barbara Ruef**  
**Kevin Schaper**  
**Erik Segerdell**



**Michael Ashburner**  
**Rachel Drysdale**  
**David Osumi-Sutherland**

**Cambridge**



**George Gkoutos**  
**Paul Schofield**  
**Damian Smedley**



**Sandra Dölken**  
**Sebastian Köhler**  
**Peter Robinson**



**Suzi Lewis**  
**Chris Mungall**  
**Nicole Washington**

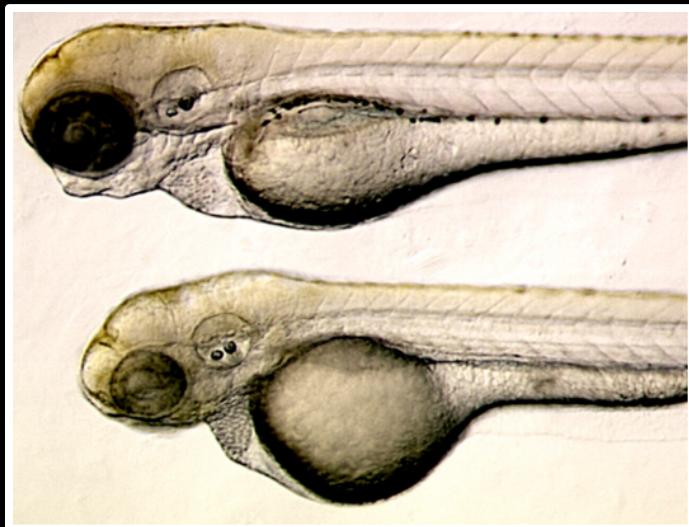


# Linking Genes to Human Diseases

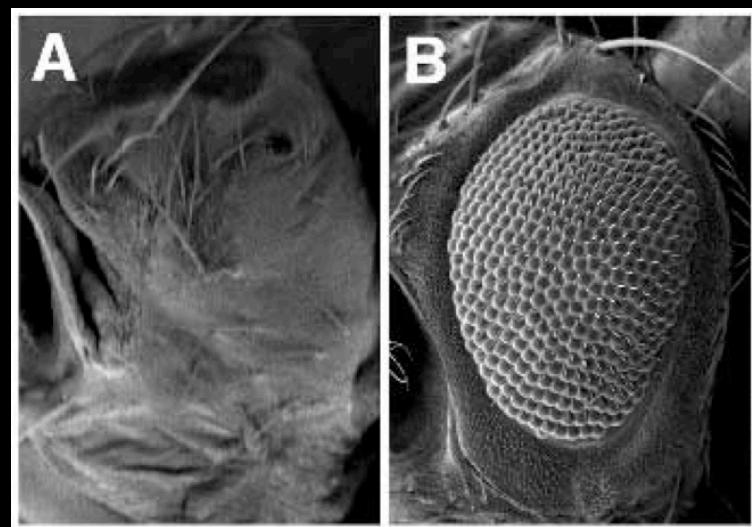
**Develop methods to describe  
& compare annotations**

- Mendelian diseases
- Multigenic diseases

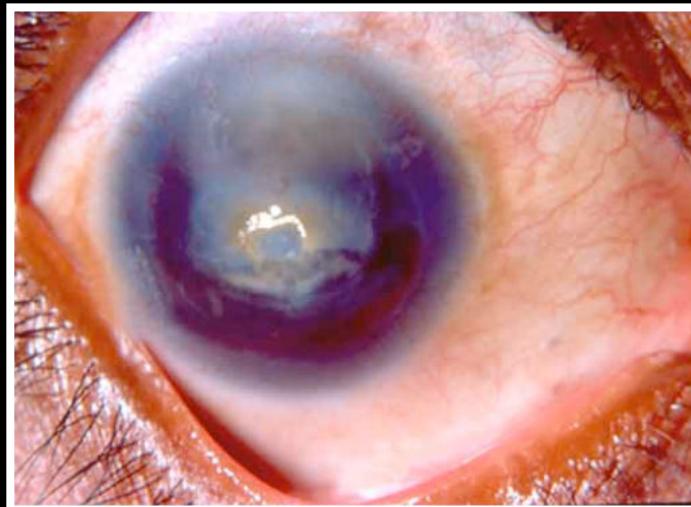
# *EYA* gene mutants



zebrafish

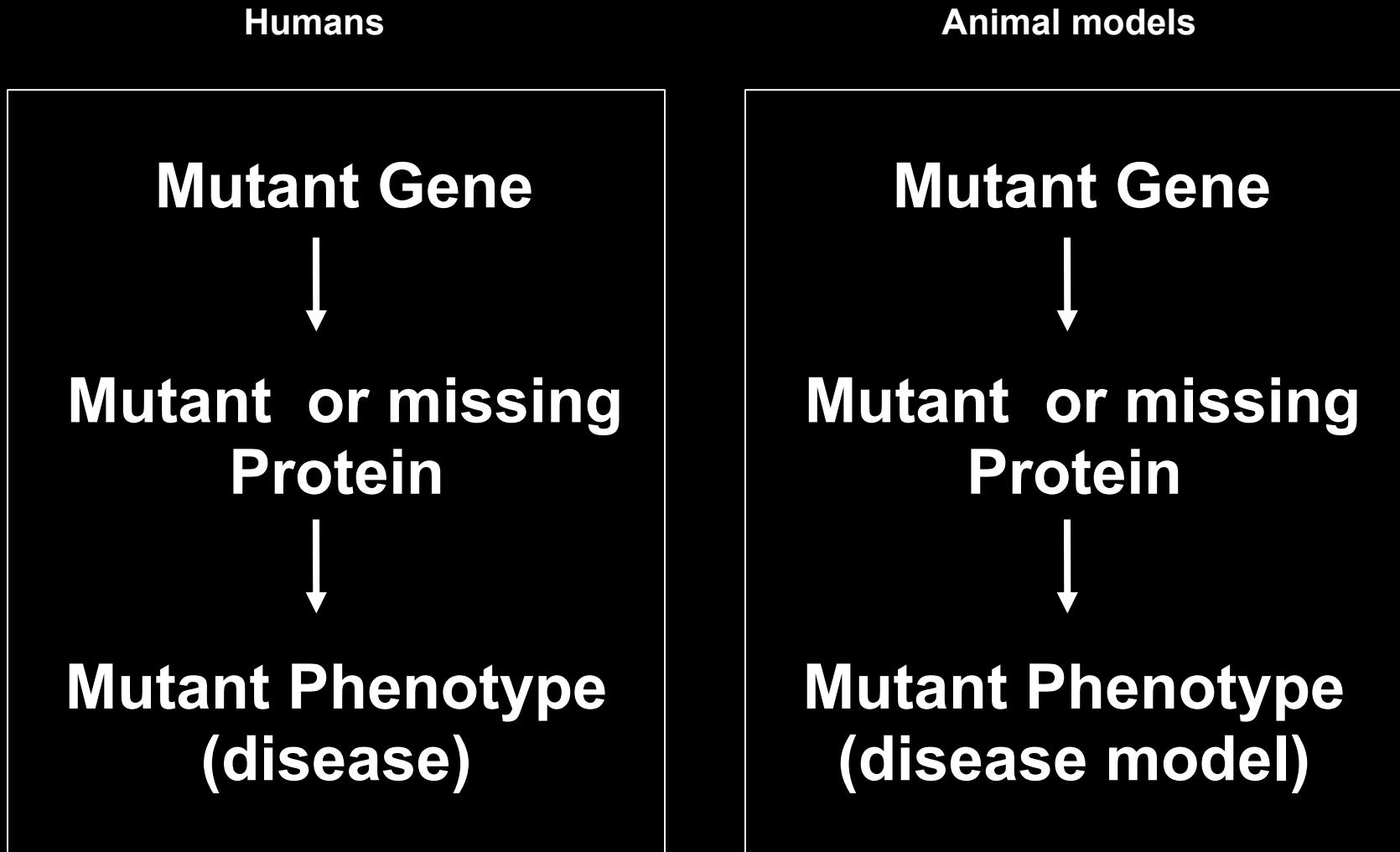


fly

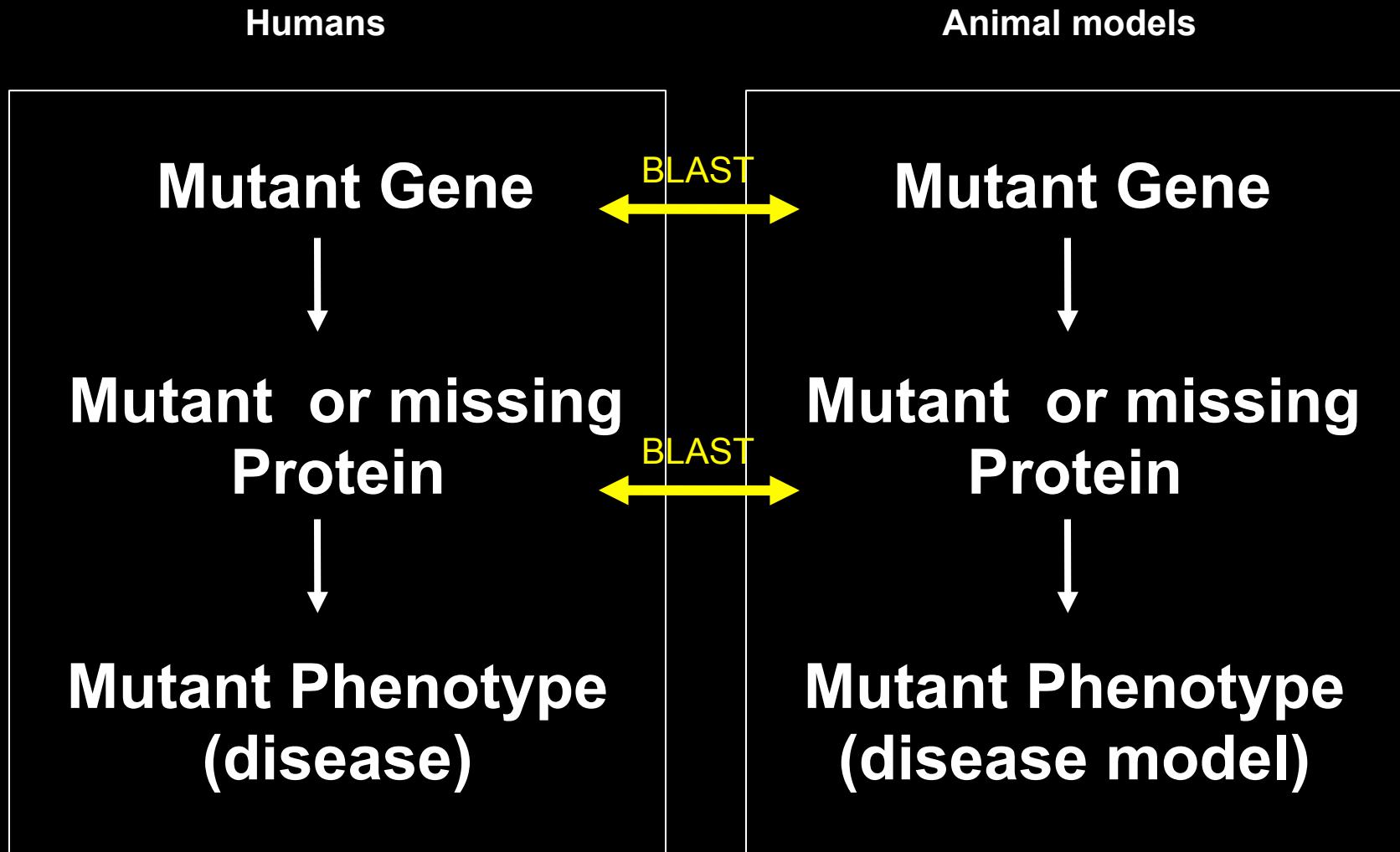


human

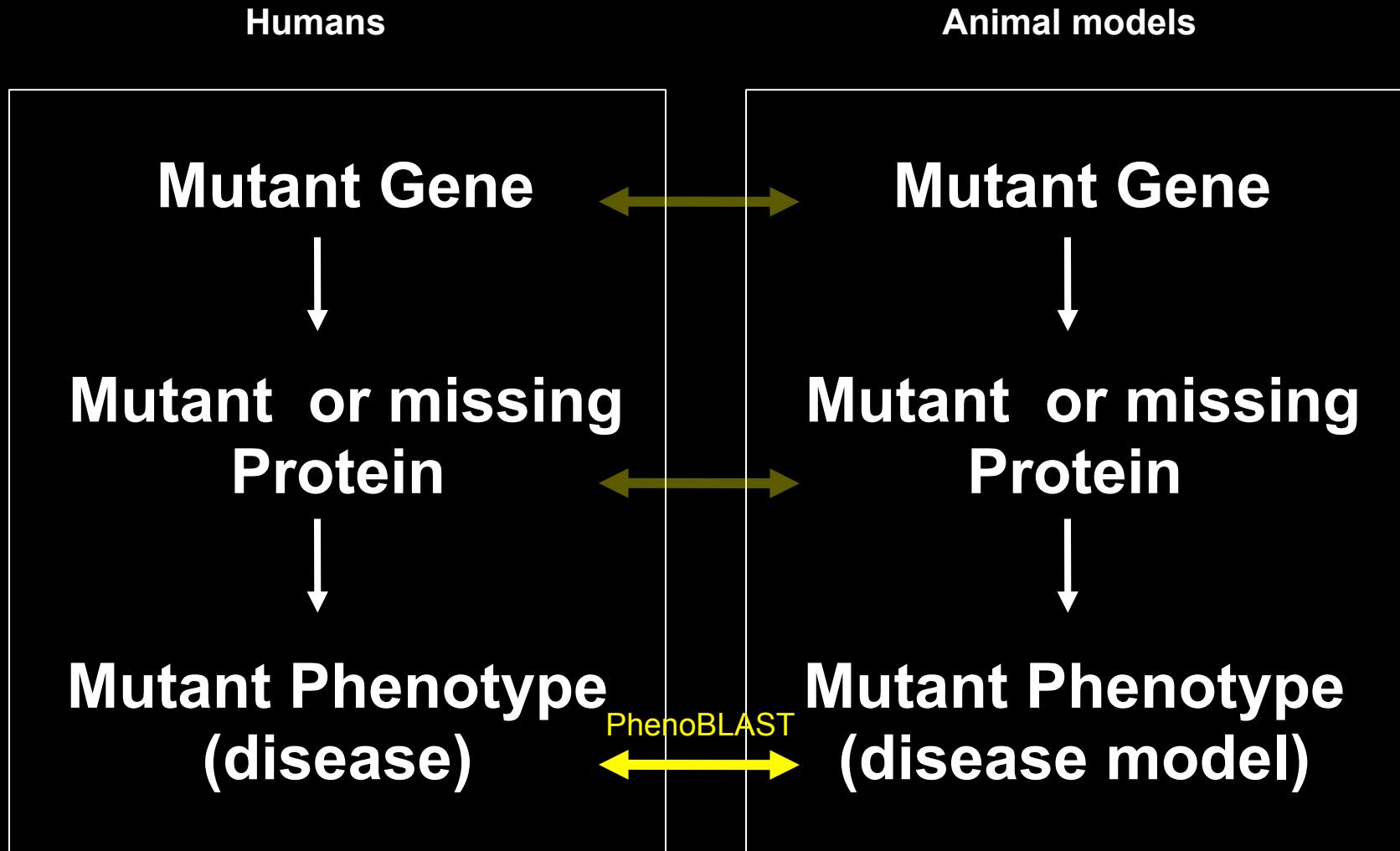
# Animal disease models



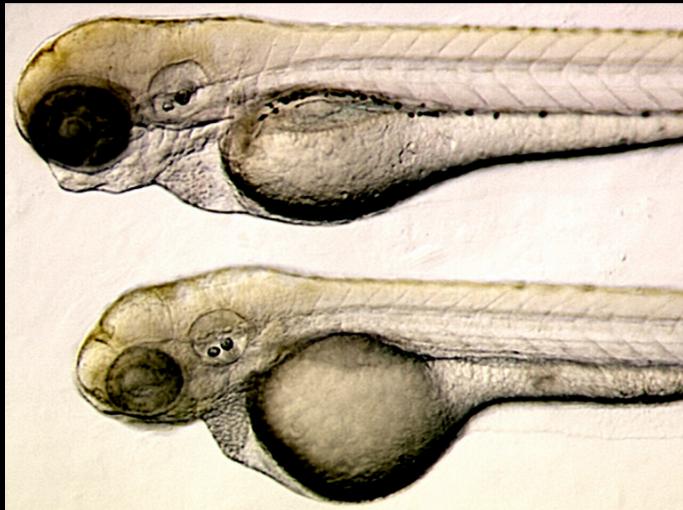
# Animal disease models



# Animal disease models



# Annotation of *eya* mutant phenotype using ontologies and EQ syntax



Phenotype = Entity + Quality

$EQ_1$  = eye + small

$EQ_2$  = kidney + hypoplastic

# Ontologies for Phenotype Annotation

**Phenotype**  
**(clinical sign)** = **Entity** + **Quality**

Anatomical ontology

Cell & tissue ontology

Developmental ontology

Gene ontology

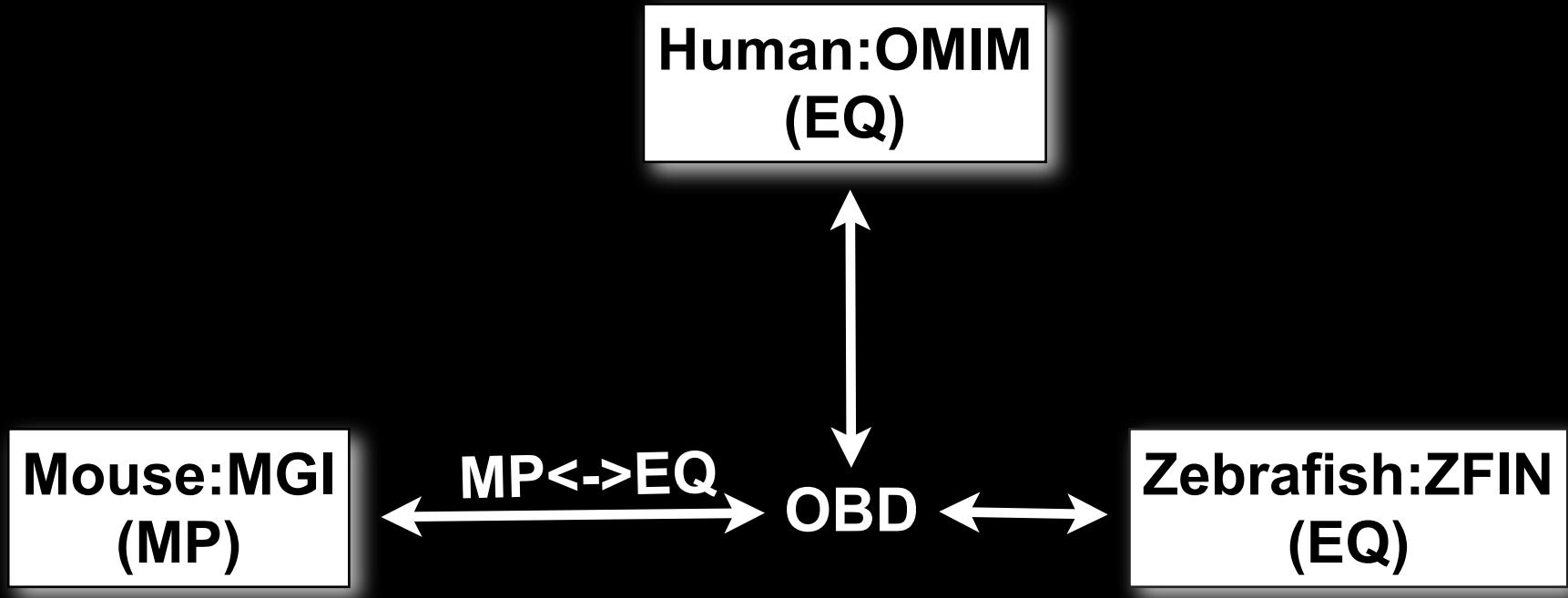
+ PATO

biological process

molecular function

cellular component

# Ontology mappings support cross-species comparisons



(Funded by NIH R01 HG004838)

# Ontologies support comparisons

Annotation 1

E: Cornea  
Q: Opaque

inheres\_in

Cornea

is\_a

Fibrous layer  
of eyeball

part\_of

Wall of  
Eyeball

part\_of

Eyeball

Annotation 2

E: Middle layer of  
corneal epithelium  
Q: Opacity

is\_a

Annotation 3

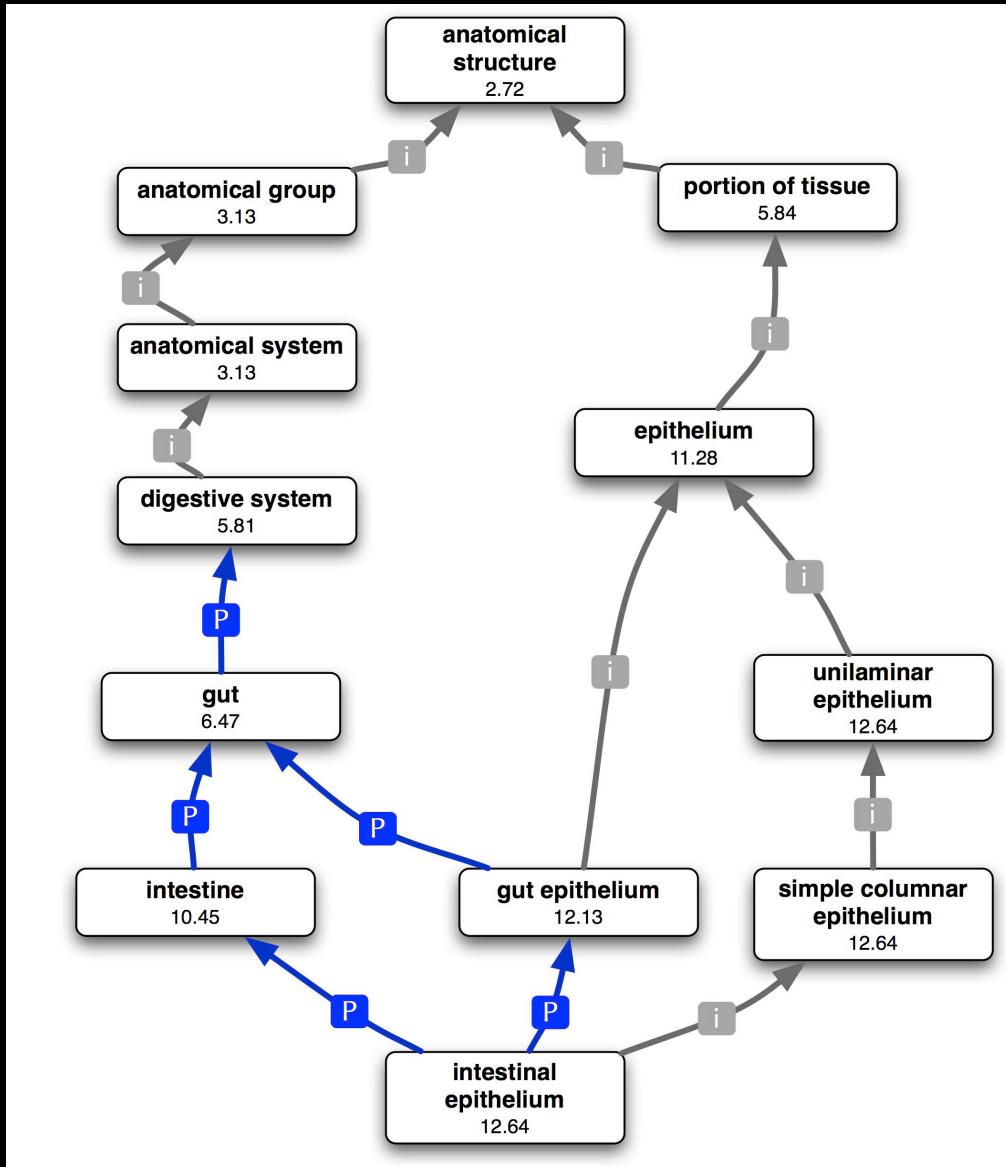
E: Lens quarter  
Q: Opaque

inheres\_in

Lens

opaque

# Subsumption reasoning for similarity scoring



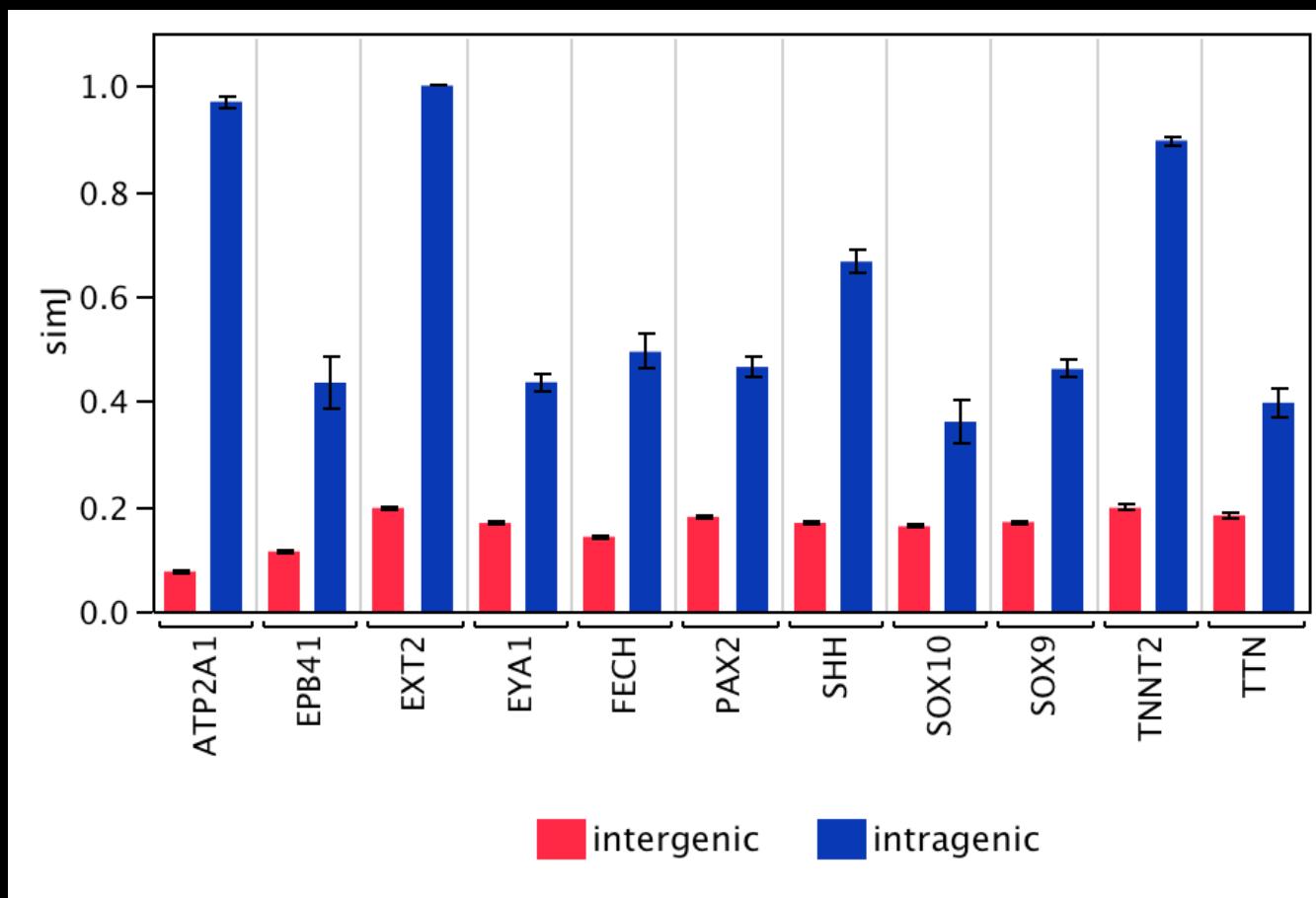
Information content (IC) is calculated based on depth within the ontology and annotation frequency

$$\text{UI}(p, q) = \frac{|g^p \cap g^q|}{|g^p \cup g^q|}$$

Similarity is calculated based on ratio of IC values

$$\text{sim}_{\text{GIC}}(p, q) = \frac{\sum_{t \in g^p \cap g^q} \text{IC}(t)}{\sum_{t \in g^p \cup g^q} \text{IC}(t)}$$

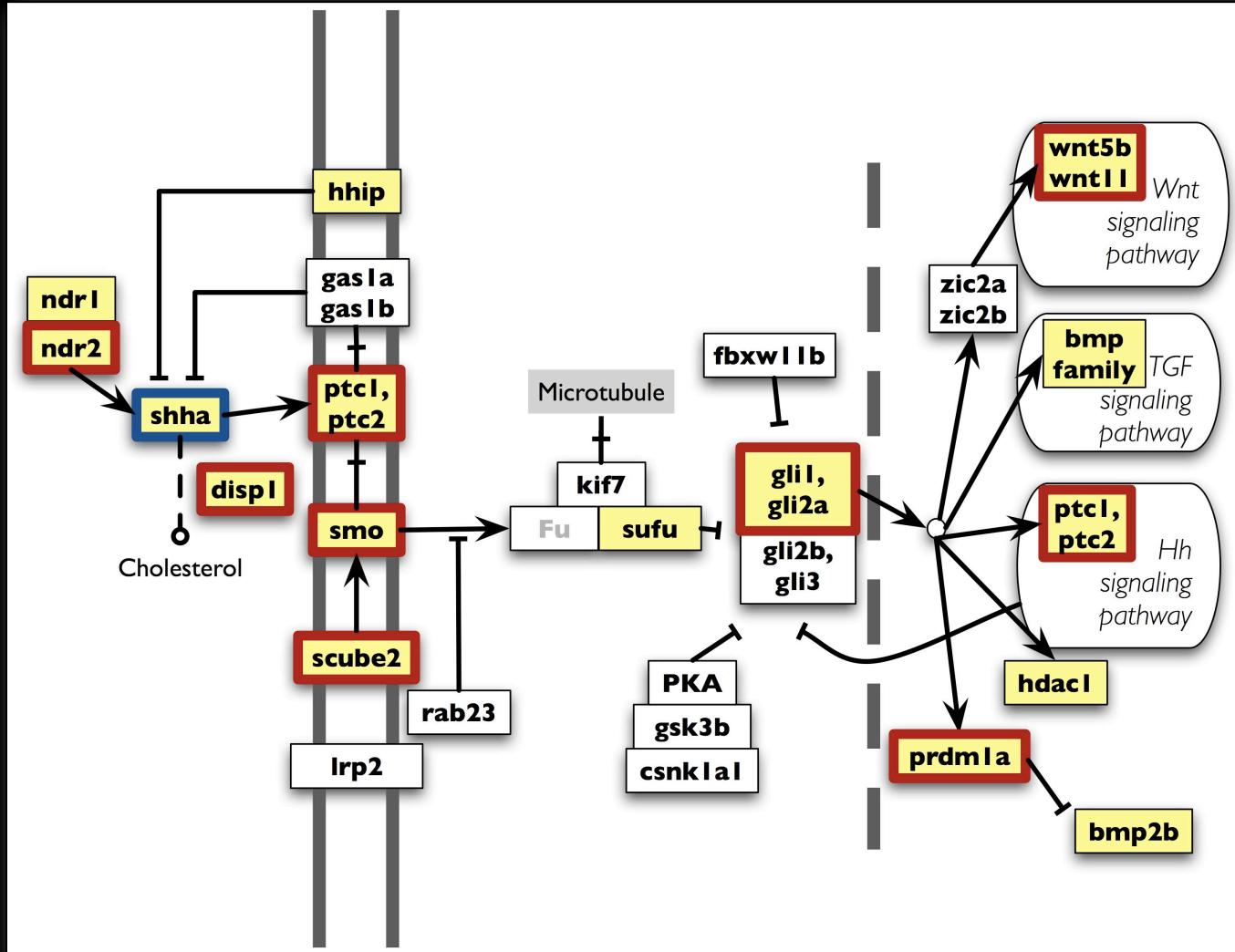
# Phenotypes identify other alleles of the same gene



All alleles are significantly more similar to alleles of the same gene than to alleles of other genes  $p<0.0001$

# Annotations can identify other pathway members

Similarity search for zebrafish *shha<sup>t4/t4</sup>* identifies pathway members



# Human phenotypes identify mutations in orthologous model organism genes

A search for phenotypes similar to:

**Human *EYA1* variant OMIM:601653**

MP:deafness = E = Sensory perception of sound Q = absent

# Human phenotypes identify mutations in orthologous model organism genes

A search for phenotypes similar to:

Human *EYA1* variant OMIM:601653

MP:deafness = E = Sensory perception of sound Q = absent

returns:

Mouse *Eya1* *bor/bor* and *Eya1<sup>tm1Rilm/tm1Rilm</sup>*

E = Sensory perception of sound Q = decreased

# Linking Genes to Human Diseases

**Develop methods to describe  
& compare annotations**

- Mendelian diseases
- Multigenic diseases

# Diseases of multiple genes - the DECIPHER project

## Phenotype

### disease features

(e.g. Williams Beuren syndrome annotations)

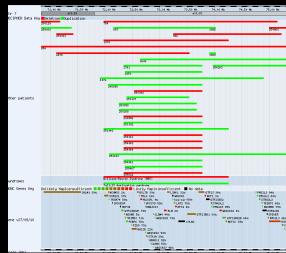


Nasal hypoplasia	HP:0003196
Glucose intolerance	HP:0001952
Hypercalcemia	HP:0003072 ...

## Genotype

### heterozygous deletions or duplications (CNVs)

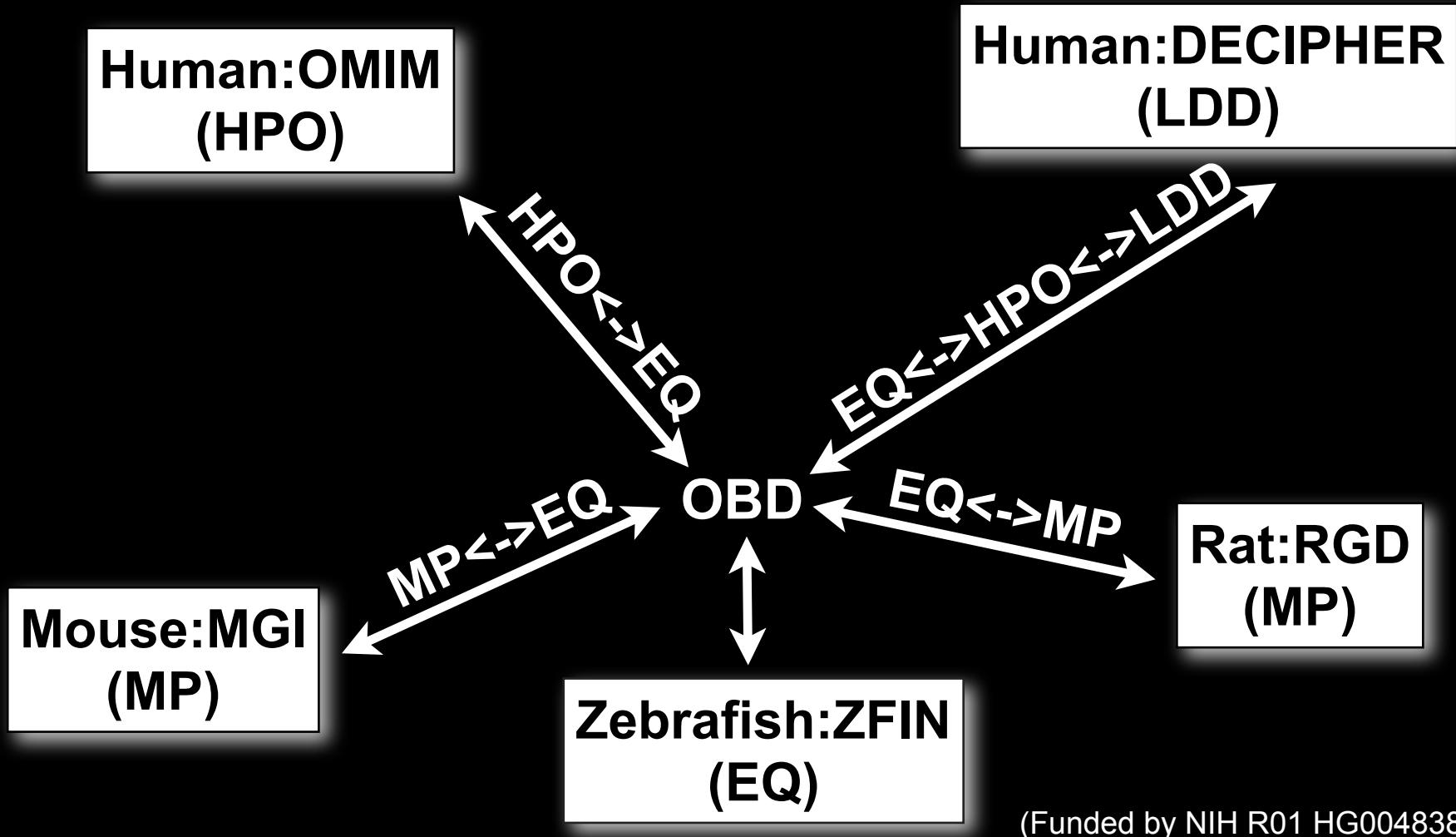
(e.g. heterozygous deletion of ~ 34 genes ~ 1.55Mb on 7q11.23)



GTF2I	EntrezGene-ID: 2969
BAZ1B	EntrezGene-ID: 9031
GTF2IRD2	EntrezGene-ID: 84163 ...

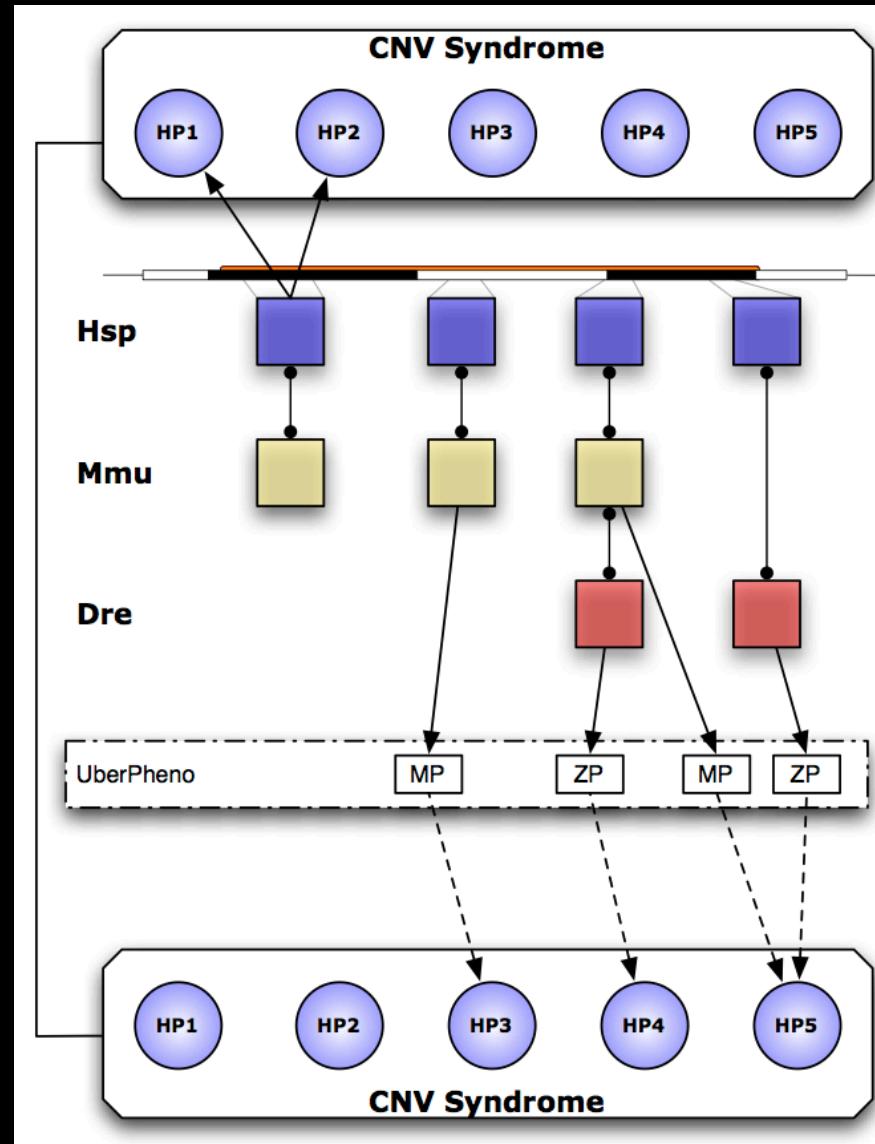
Questions: Which genes explain the phenotypic features of the patients?  
What are the phenotype/genotype correlations?

# Ontology mappings support cross-species comparisons

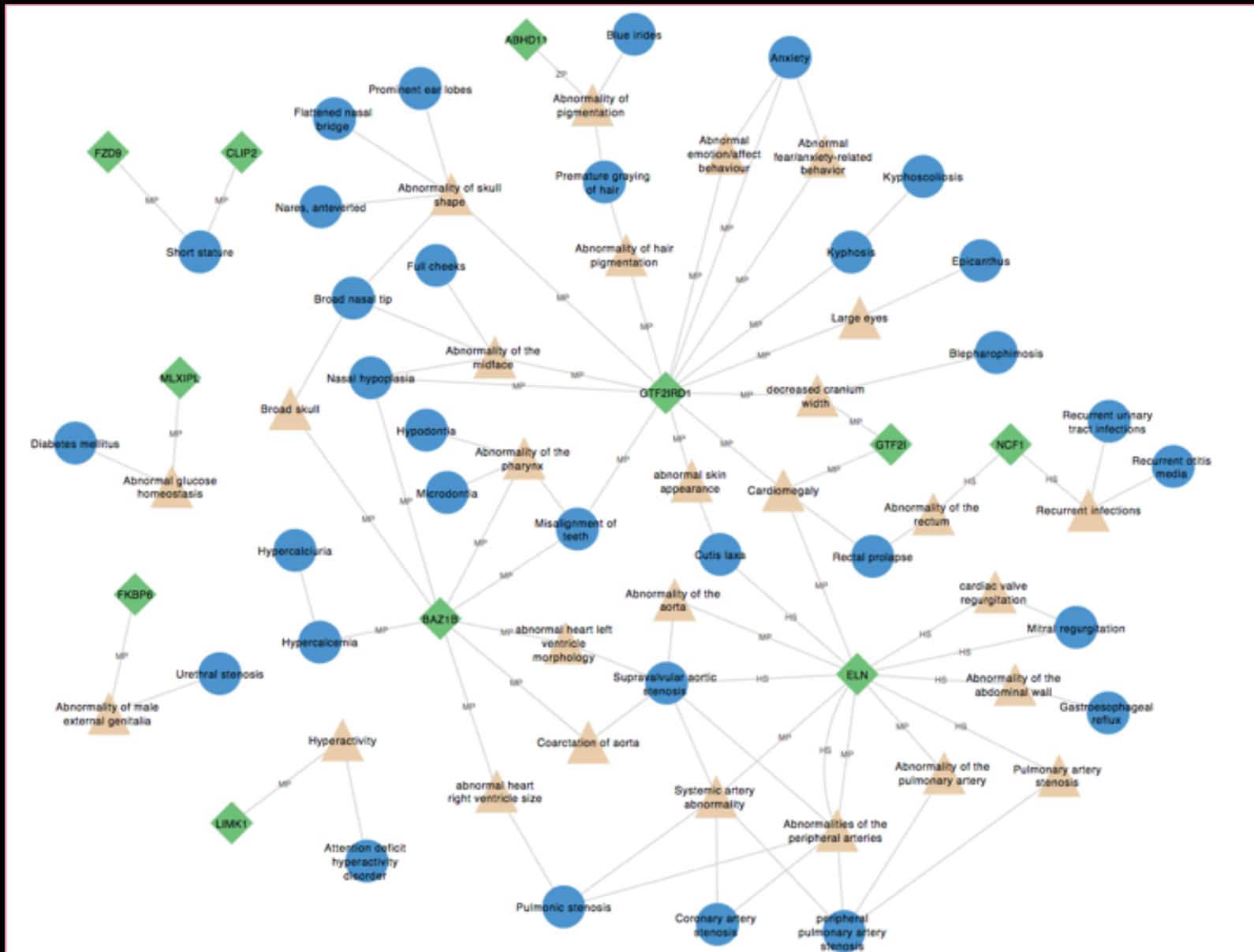


(Funded by NIH R01 HG004838)

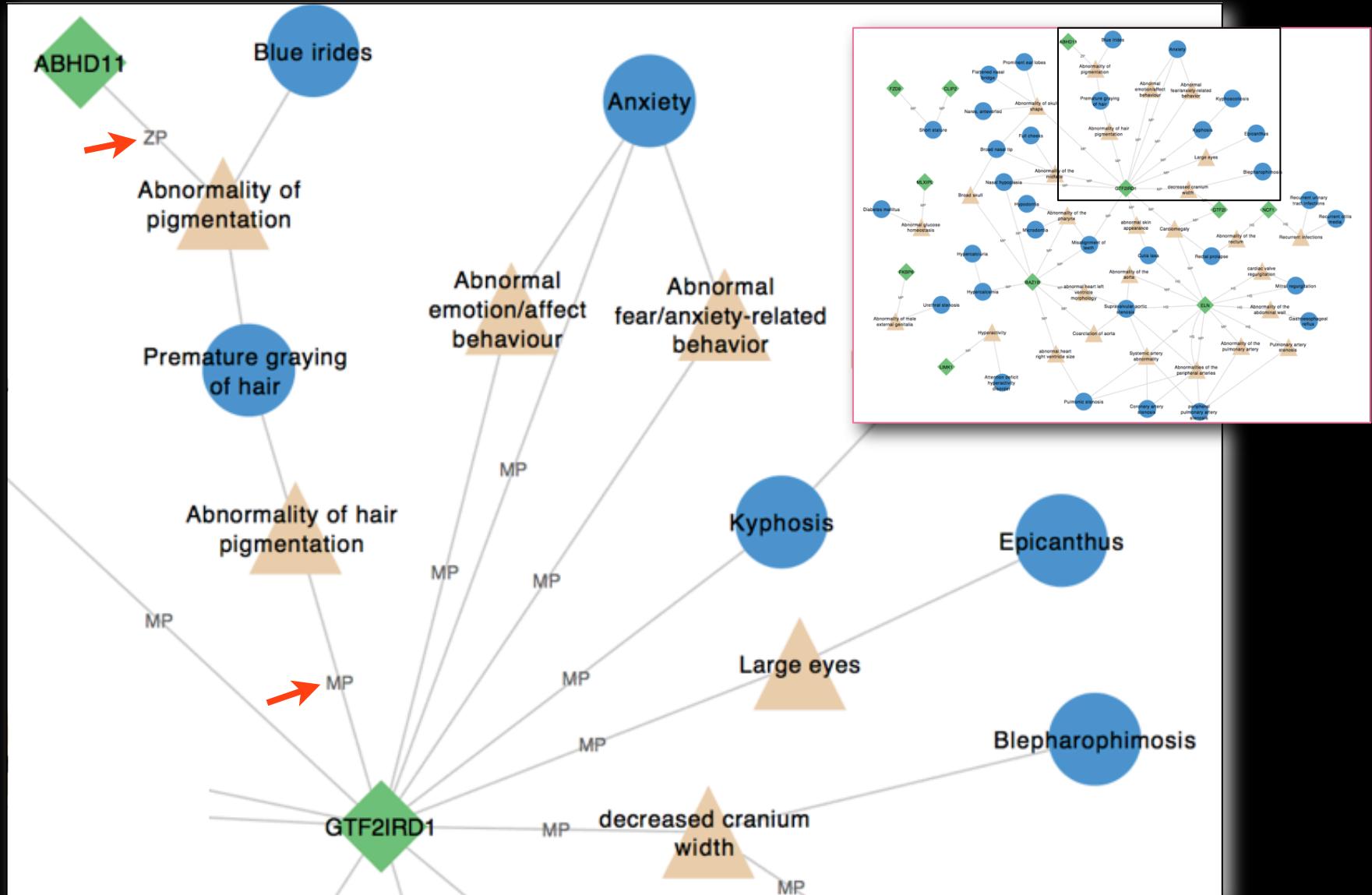
# Strategy for identifying CNV disease genes



# Phenogram visualization of Williams syndrome

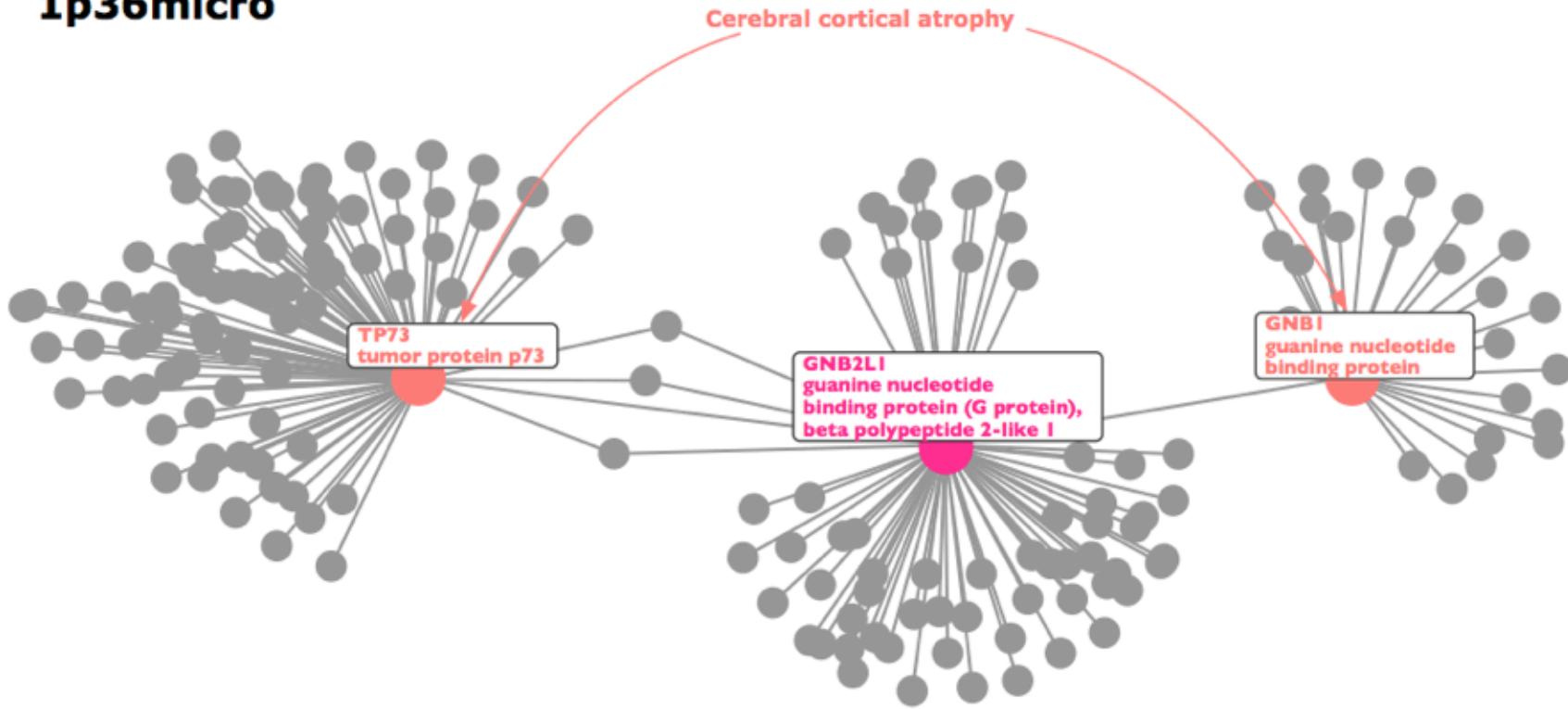


# Phenotypes due to multiple genes in CNV (phenoclusters)

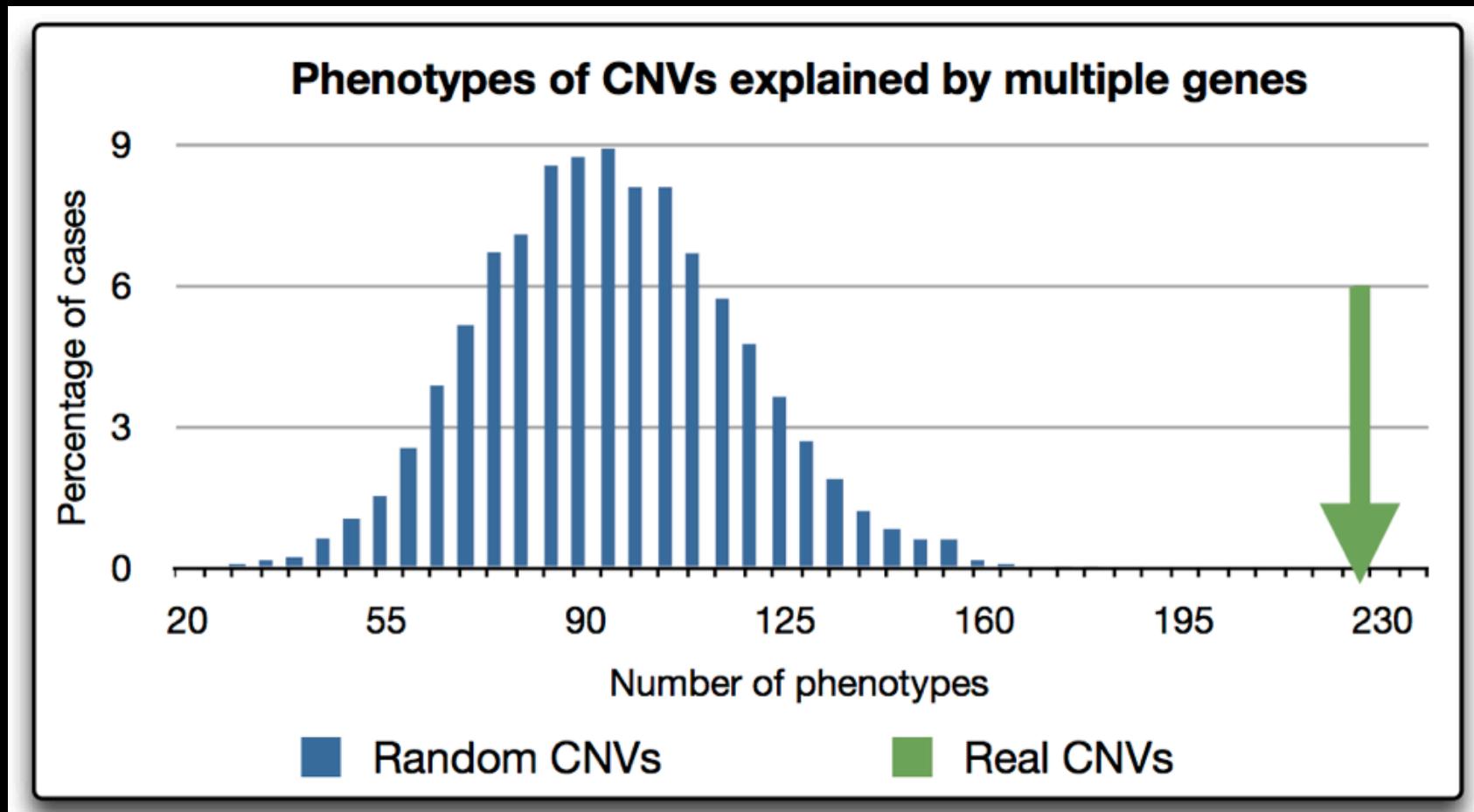


# Analysis shows clustering of phenotypes

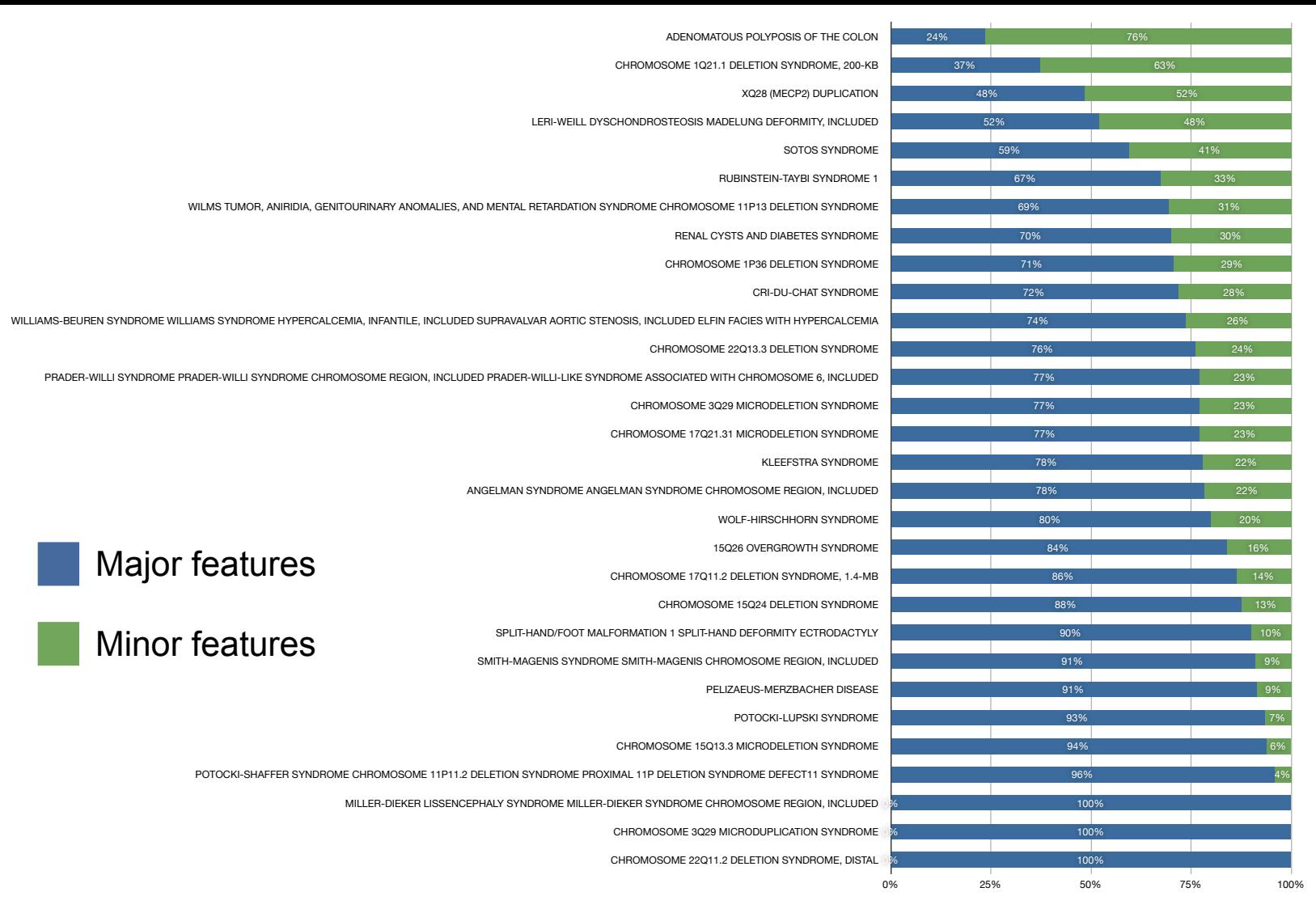
1p36micro



# Correlations are statistically significant



# Correlations explain both major & minor disease features



Major features

Minor features

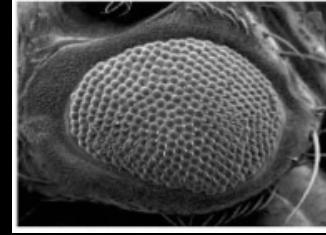
# Linking Genes to Human Diseases

**Develop methods to describe  
& compare annotations**

- Mendelian diseases
- Multigenic diseases

# Linking Animal Models and Human Diseases

**EU-US Animal Biotechnology Working Group Workshop**  
**Monte Westerfield - November 8, 2011**



**Supported by NIH P41 HG002659, U54 HG004028, & R01 HG004838**  
**Berkeley, Cambridge & Oregon Universities**