Crossing the Species Divide

Chris Mungall
Lawrence Berkeley Laboratory

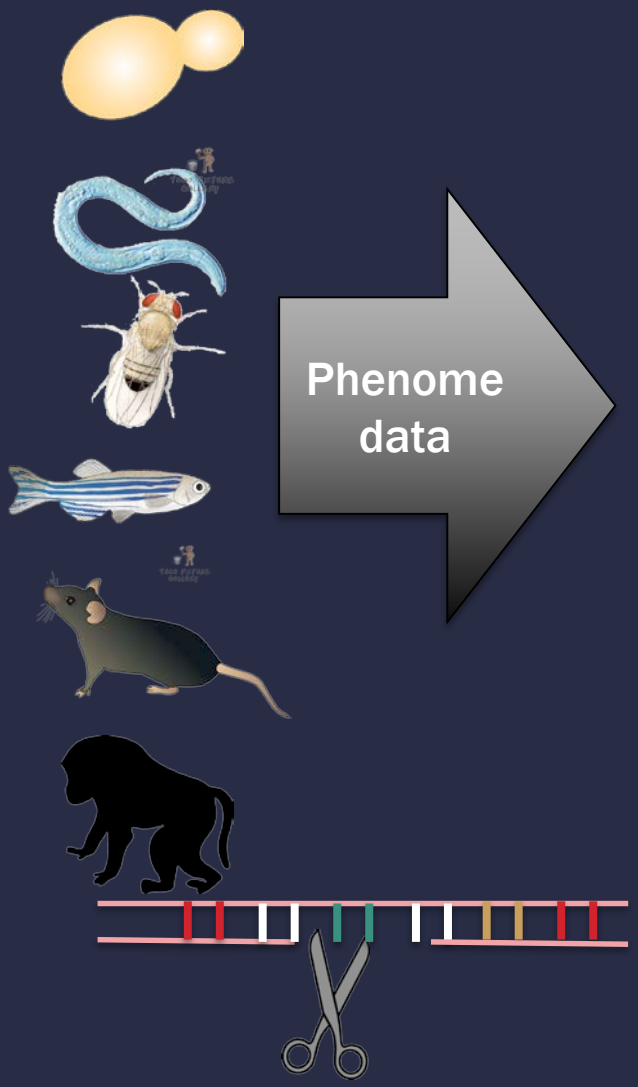
NIH Symposium: Linking Disease Model Phenotypes to Human Conditions

2015-09-10

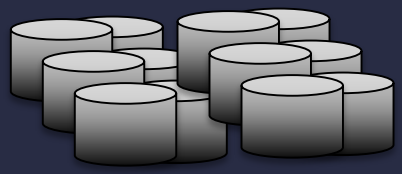
monarch INITIATIVE

BERKELEY LAB

The growth of phenomic data



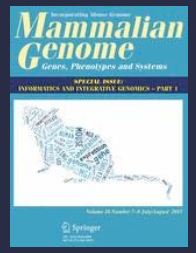
Phenome data

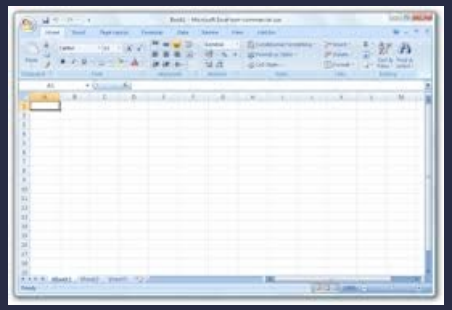


PLOS

PDF

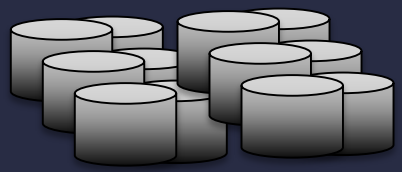
Adobe





The growth of phenomic data

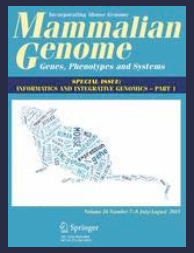
Phenome data



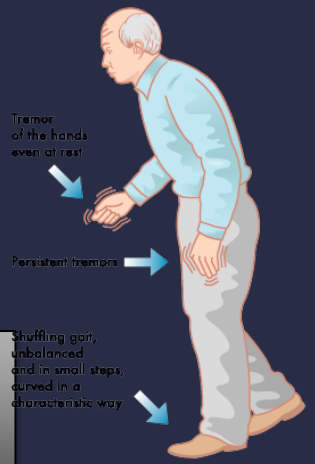
PLOS

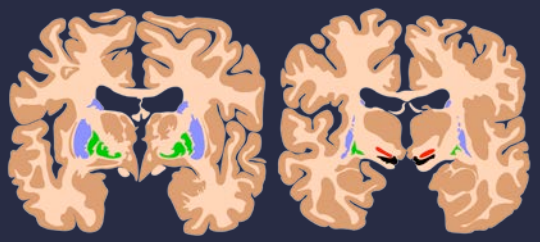
PDF

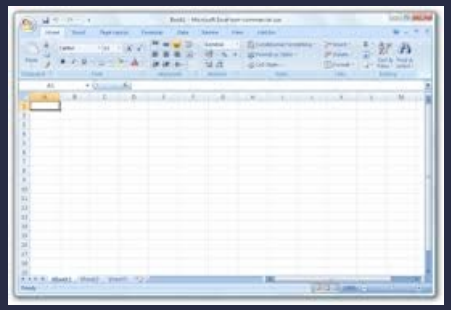
Adobe



Human Phenotype data

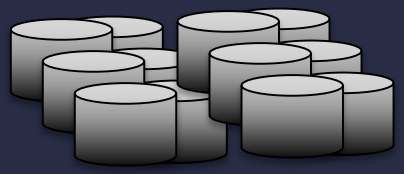






monarchinitiative.org

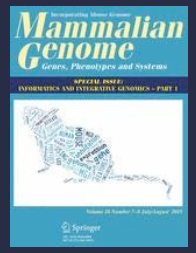
Building a phenomic knowledge base



PLOS

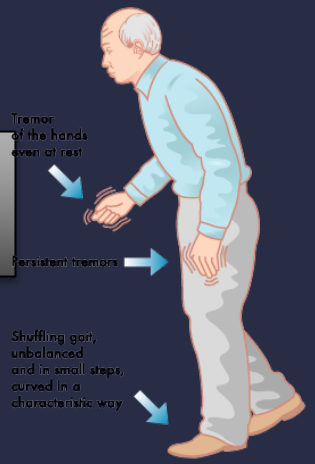
PDF

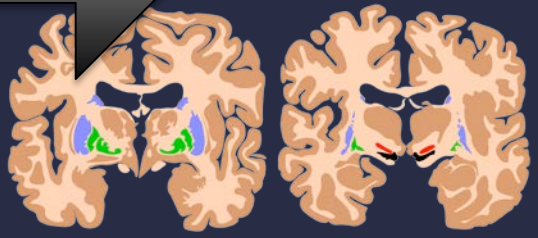
Adobe

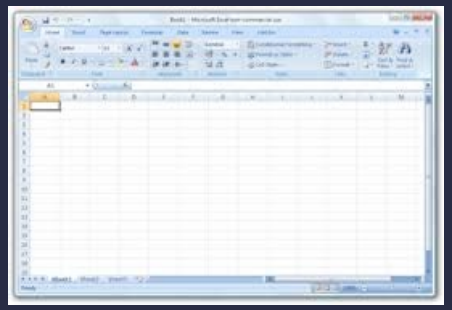


Phenome search

Inform

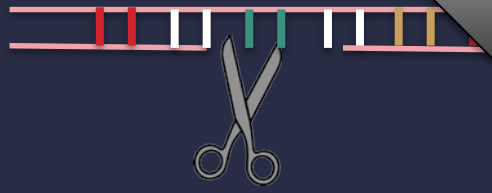




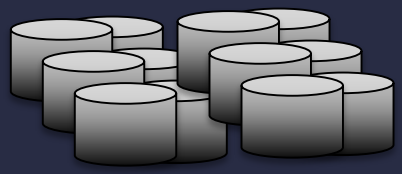


monarchinitiative.org

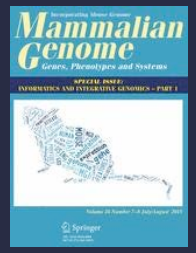
Building a phenomic knowledge base



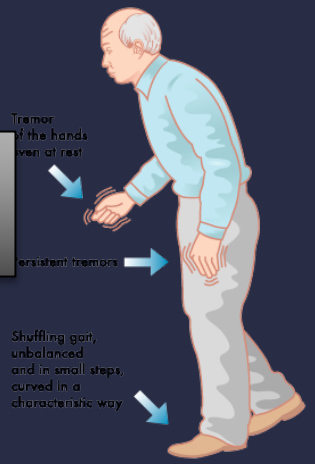
Phenome data



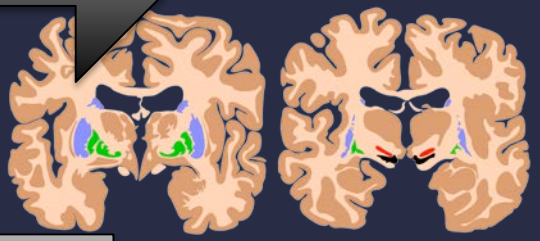
PLOS

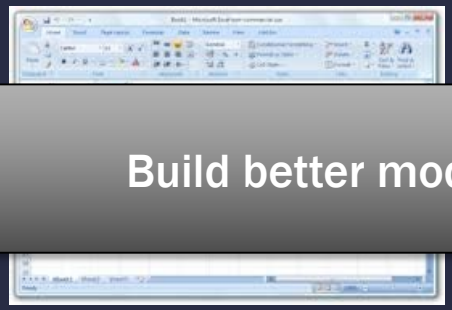


Phenome search



Inform

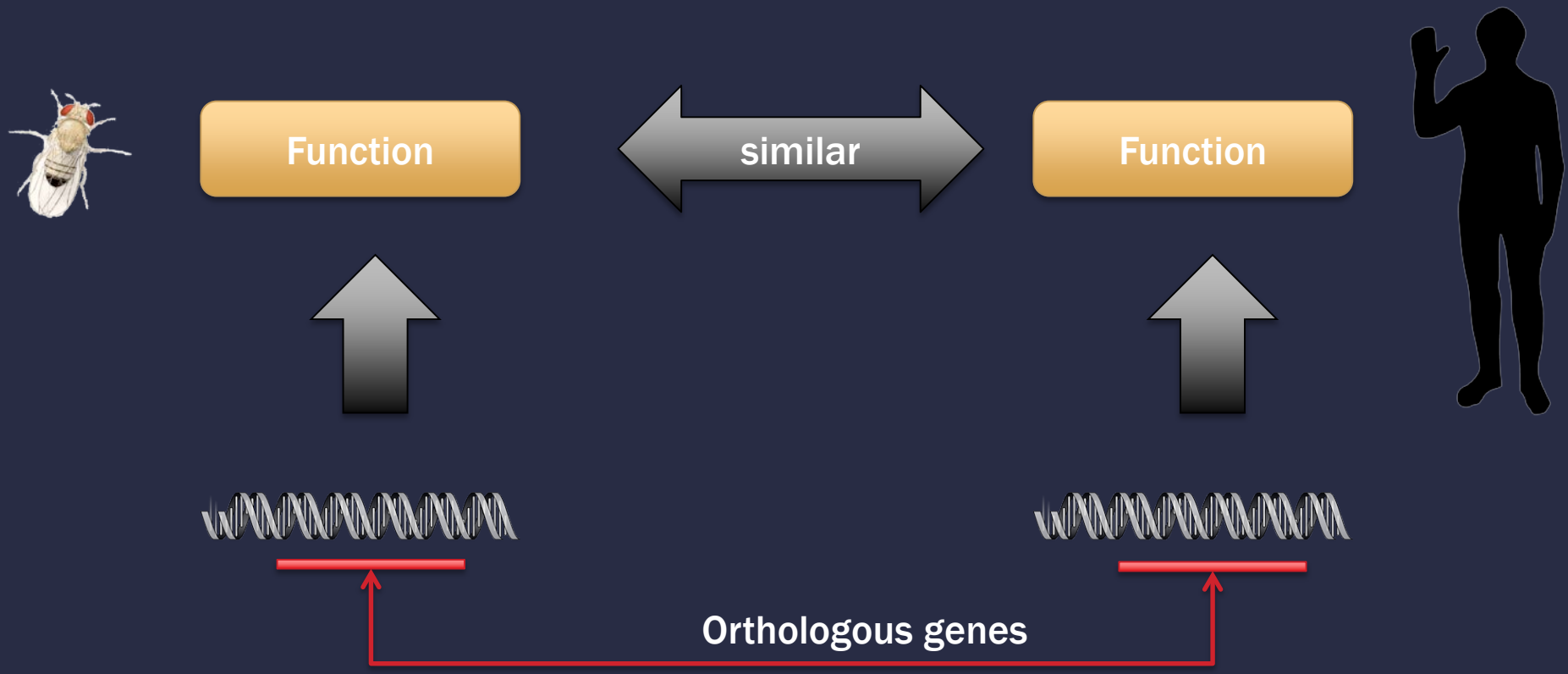




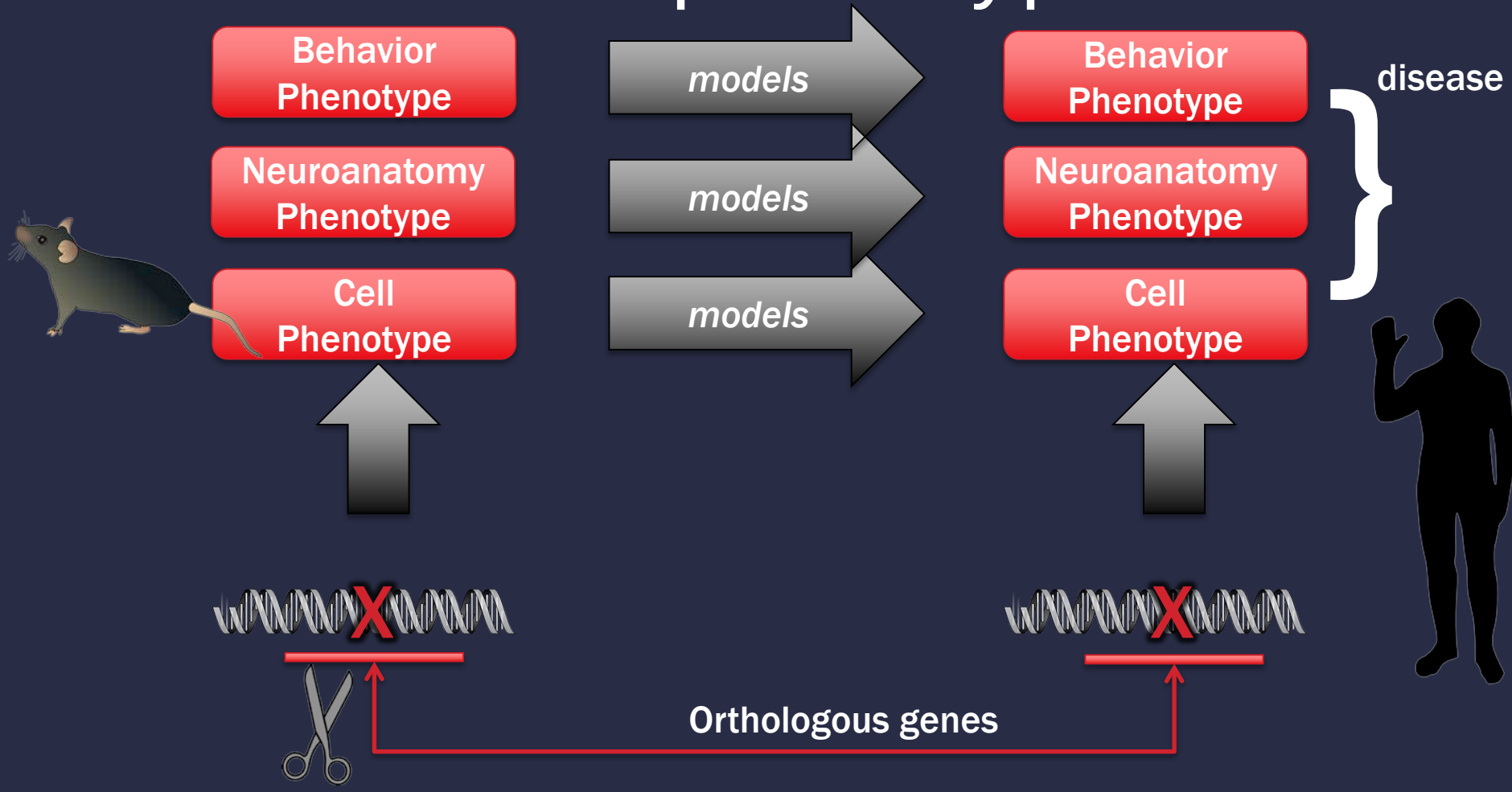
Build better models

monarchinitiative.org

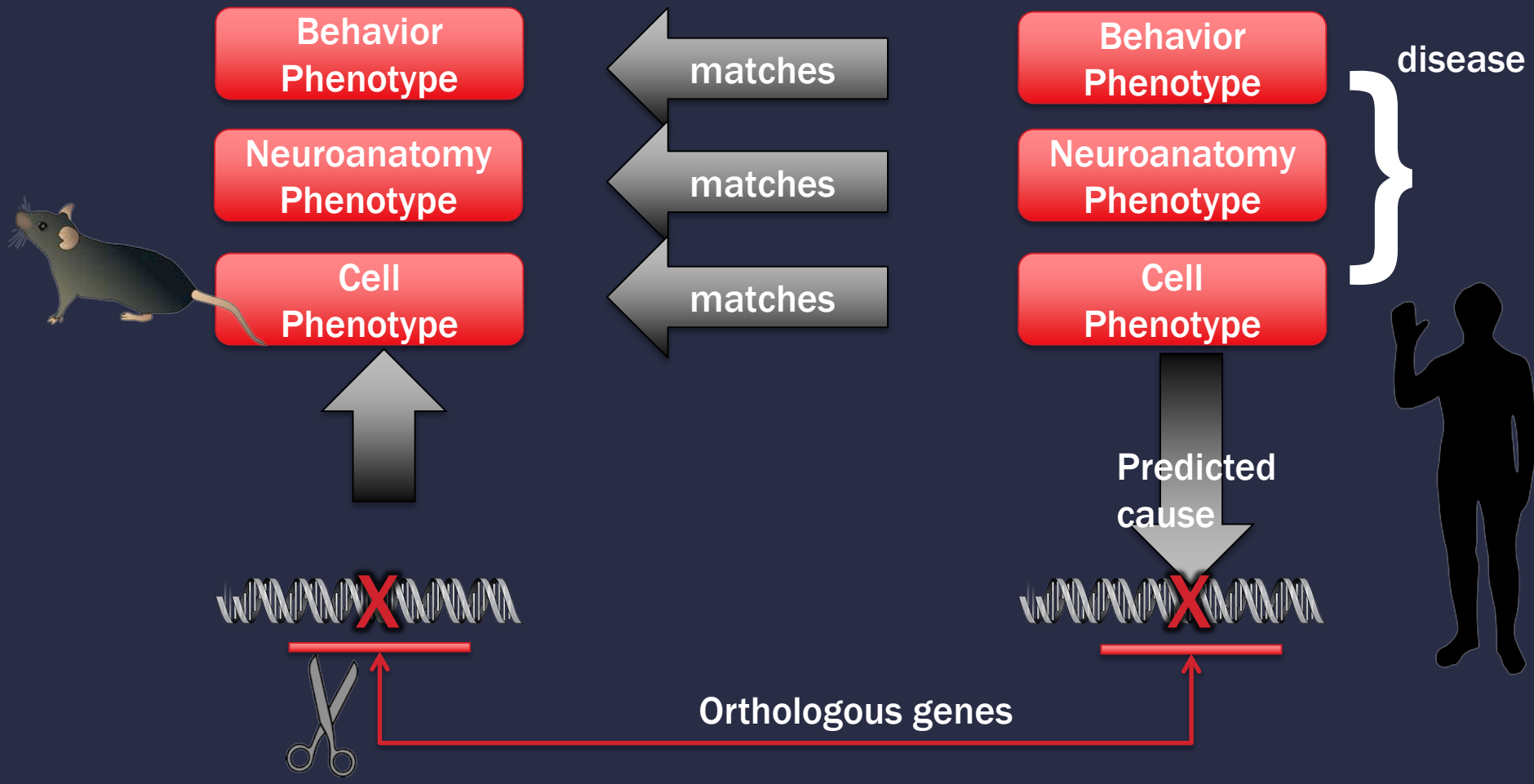
Shared genes, shared biology



Model mutants recapitulate disease phenotypes



Phenotype matching to find gene candidates



Different models for different aspects

Cell Phenotype

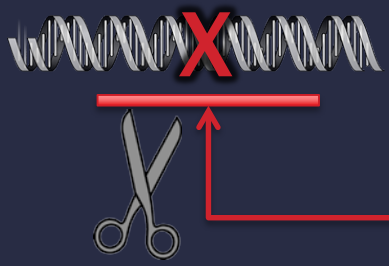
models

Behavior Phenotype

Neuroanatomy Phenotype

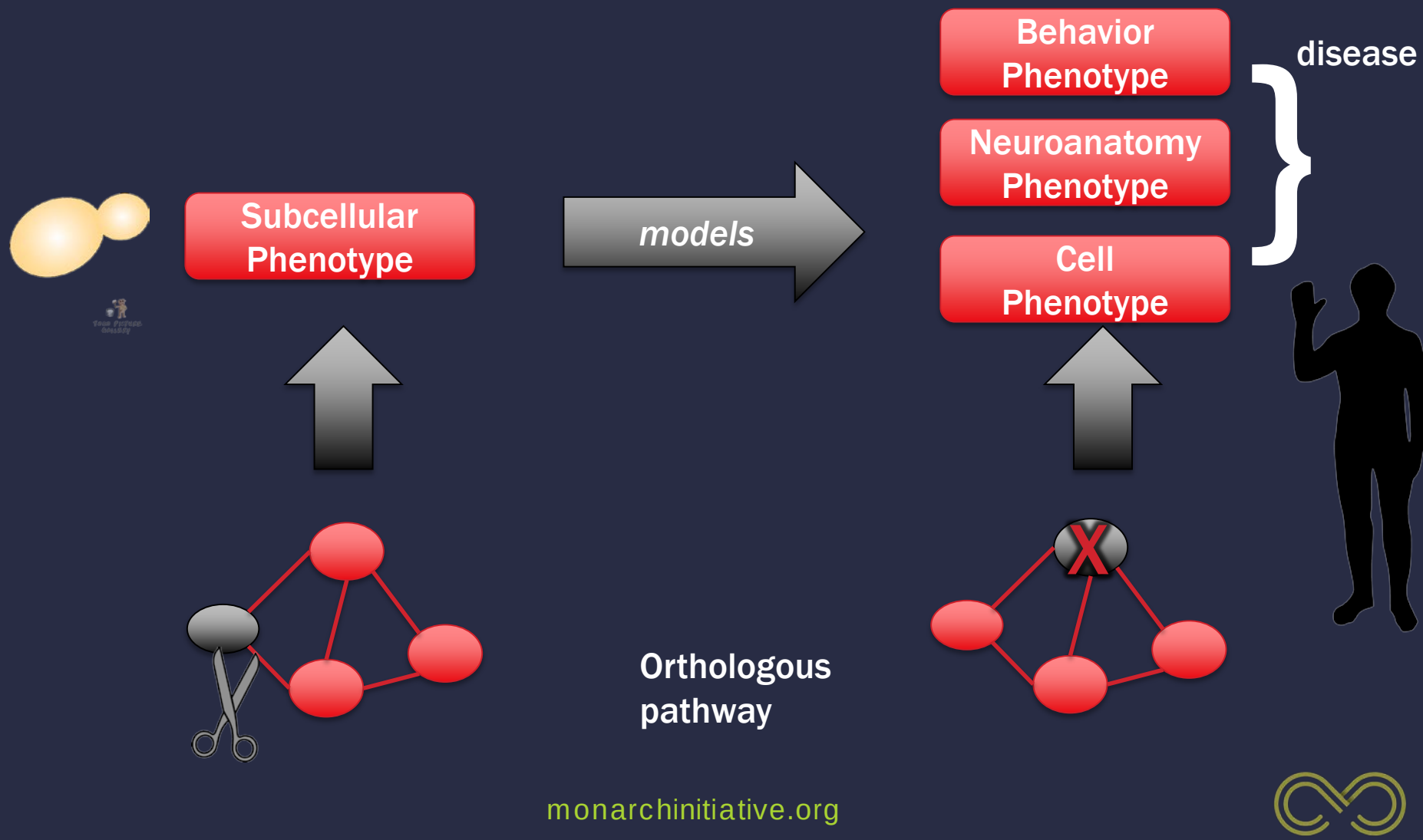
Cell Phenotype

disease



Orthologous genes

Deep homology of phenotypes



Can we quantify?

Abnormal Phenotype

What features?

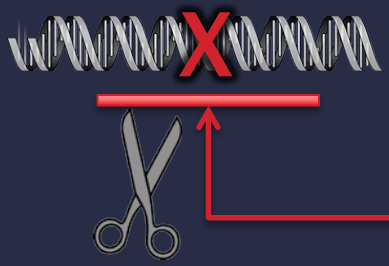
Abnormal Phenotype

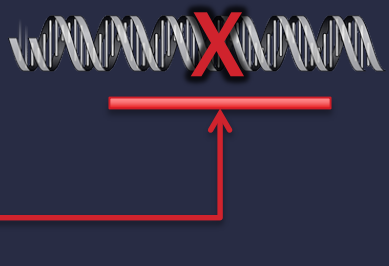
Abnormal Phenotype

Abnormal Phenotype

} disease

How good a model?





How do we get to precision comparative phenomics?

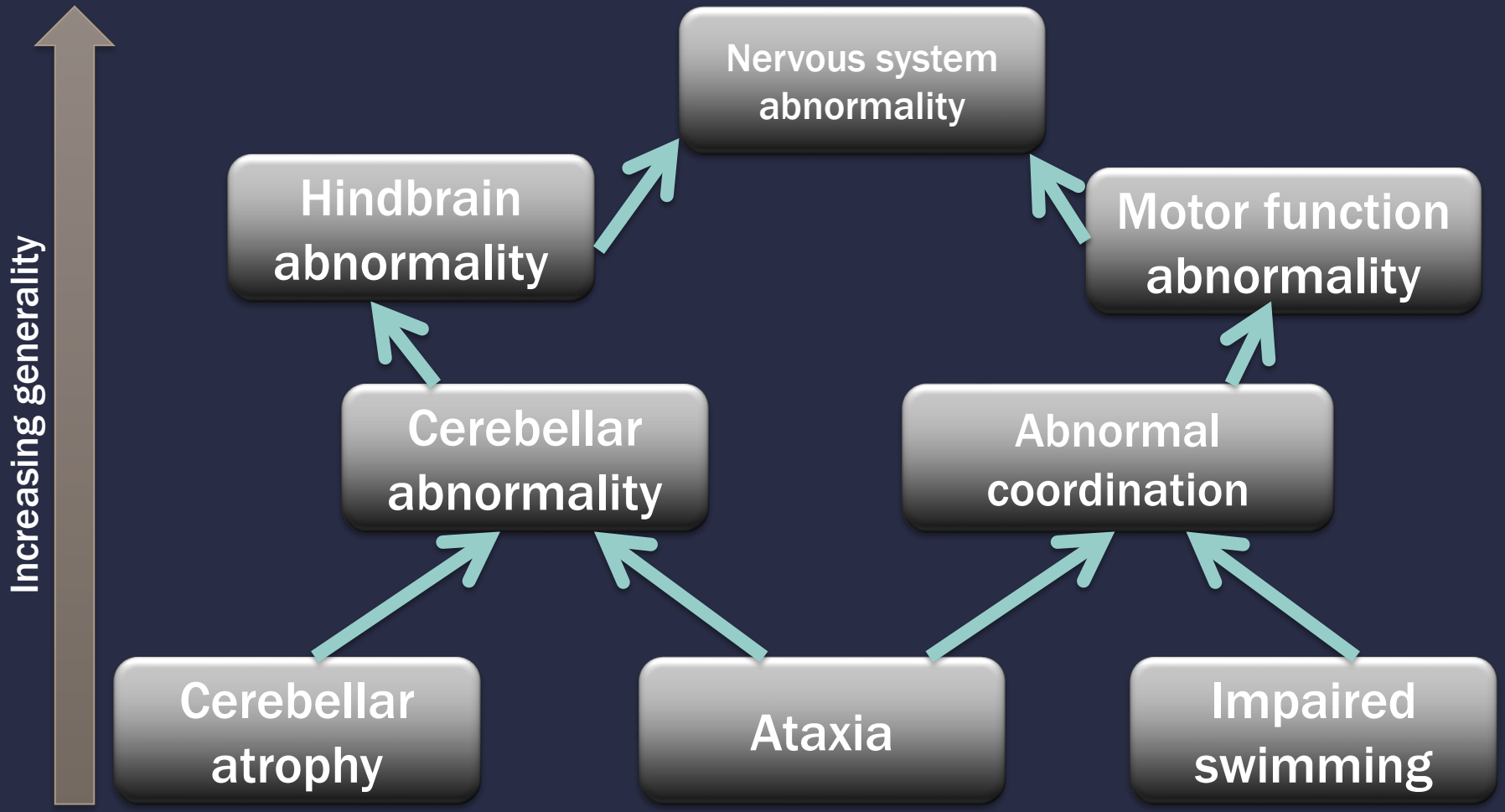
Sequence Similarity

- Increase in data
  - Sanger->NGS
- Enabled bioinformatics revolution
  - Search genome databases
- Edit distance is quantifiable
- Unifying model of molecular evolution

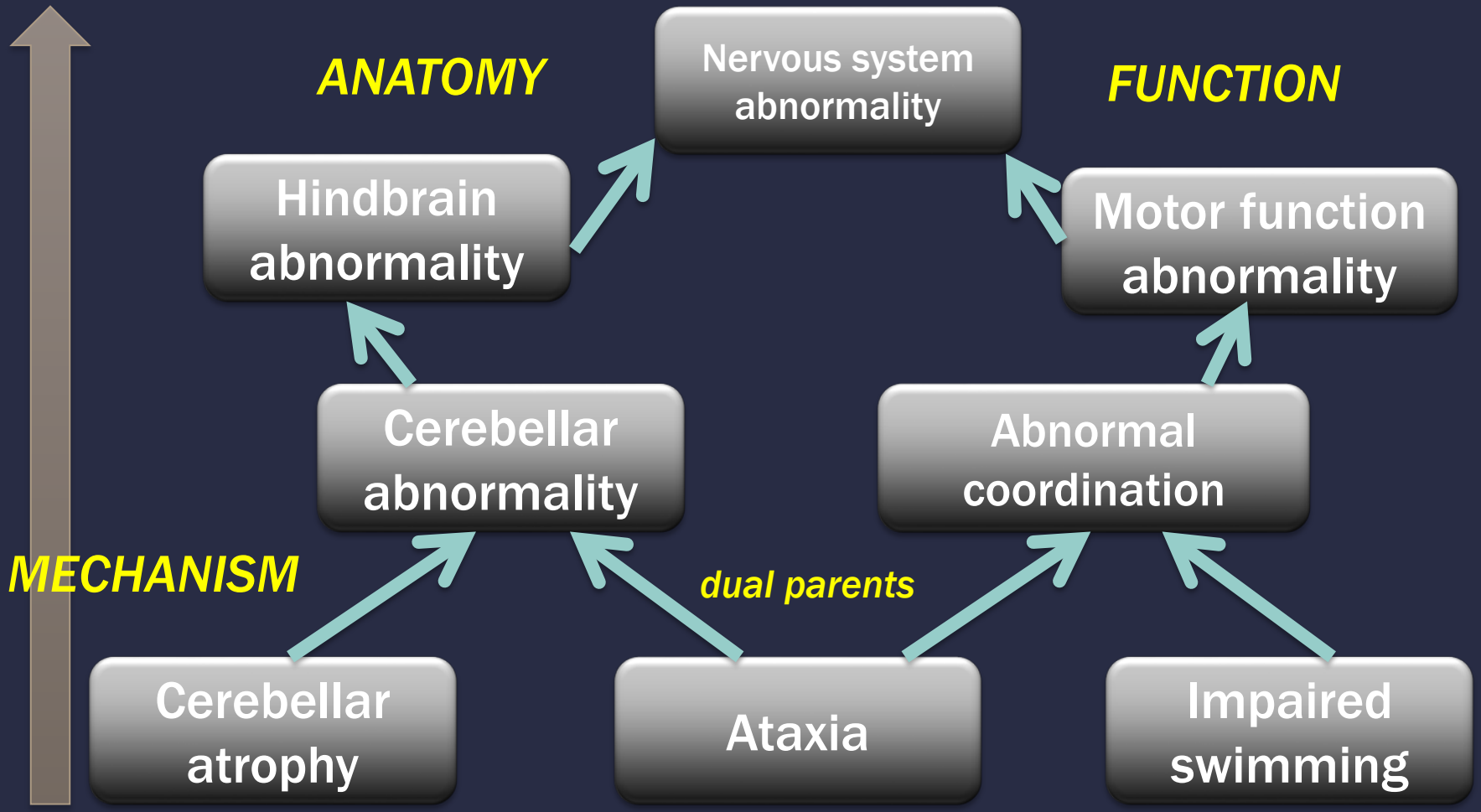
Phenotype Similarity

- Increase in data
  - Knockouts->CRISPR
- **Potential** to revolutionize
  - Search phenome databases
- Quantifying distance is **complex**
- Single unifying model is elusive

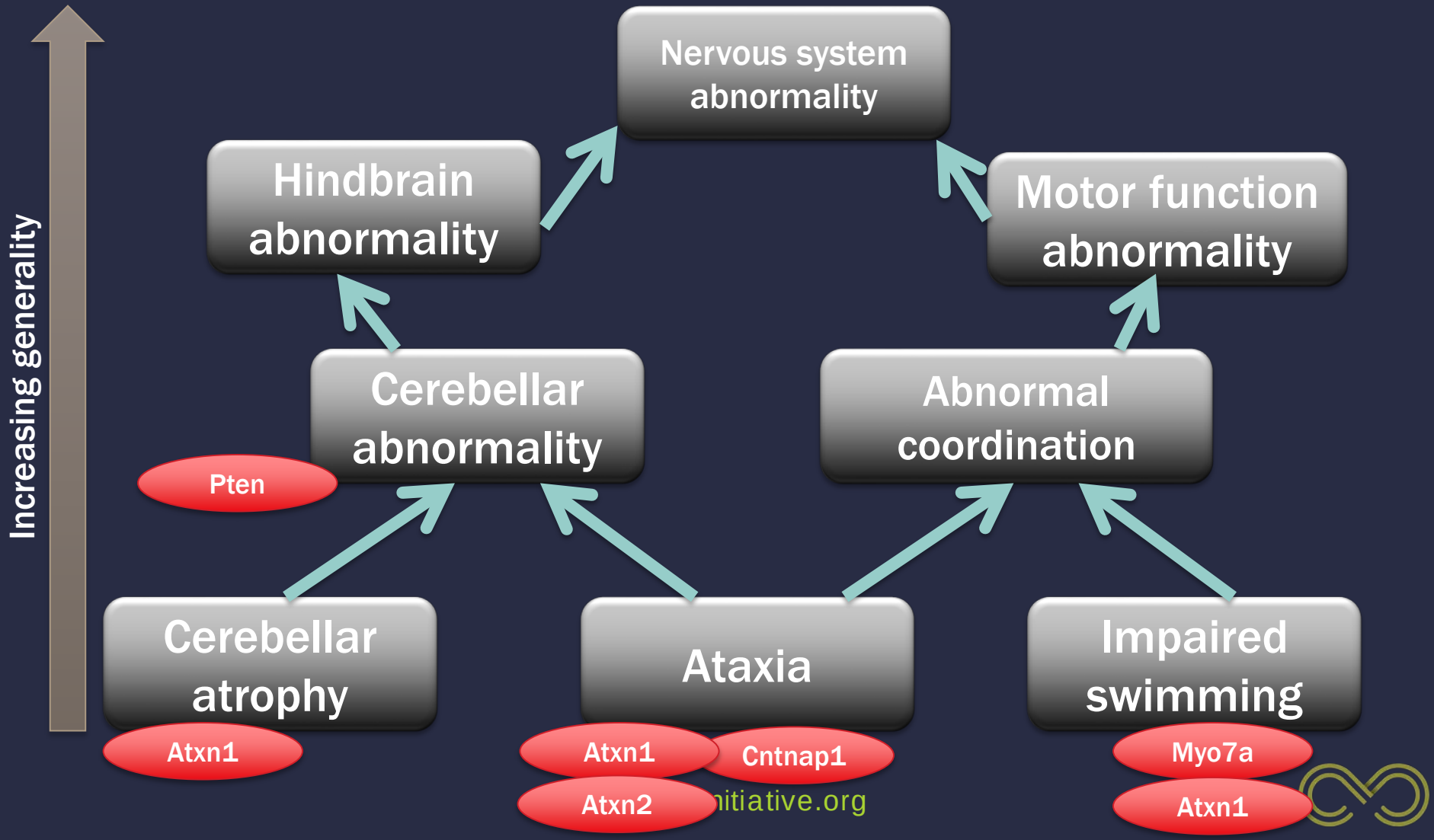
Classifying phenotypes in a graph



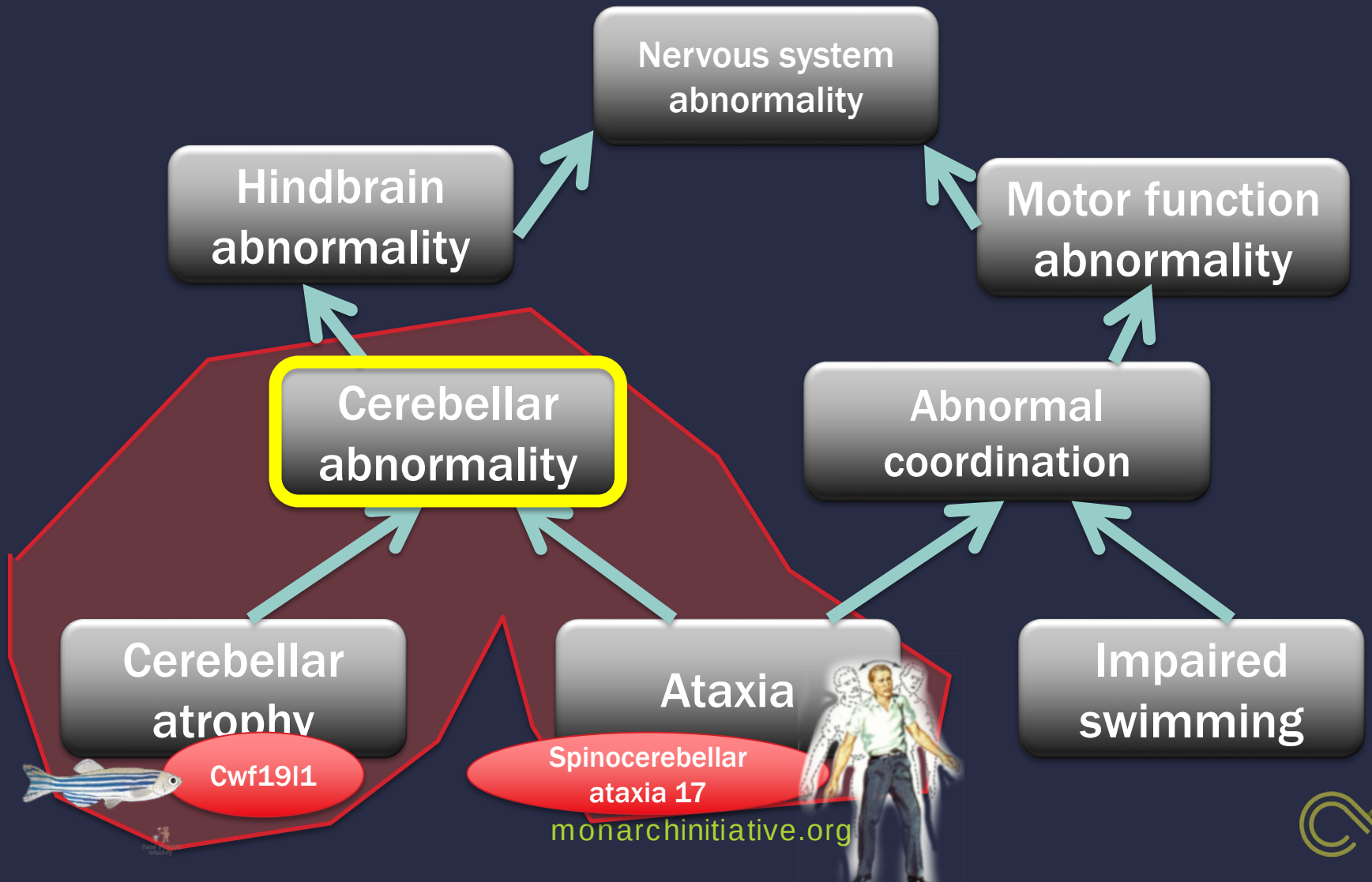
Classifying phenotypes in a graph



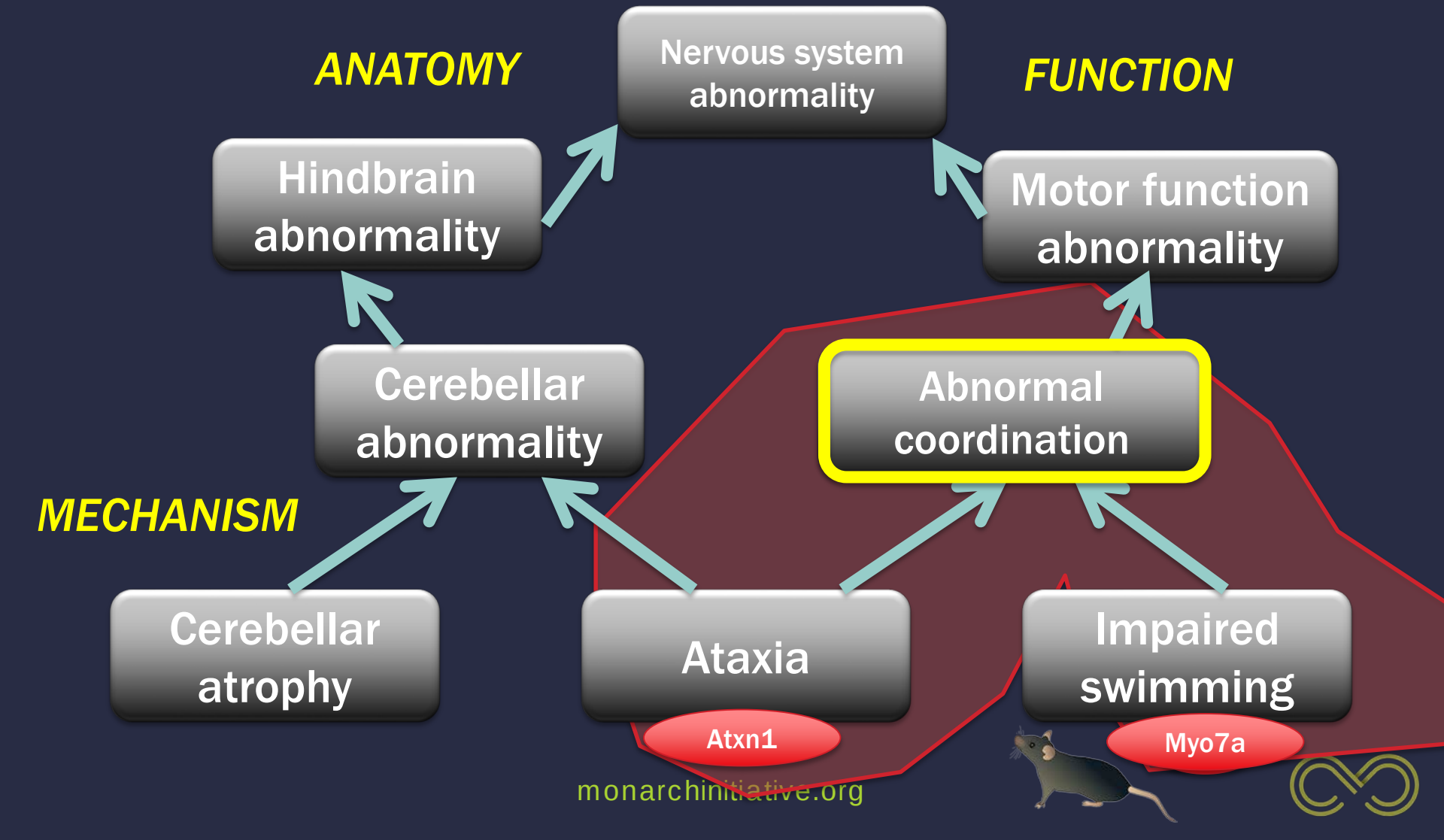
Classifying genomic entities



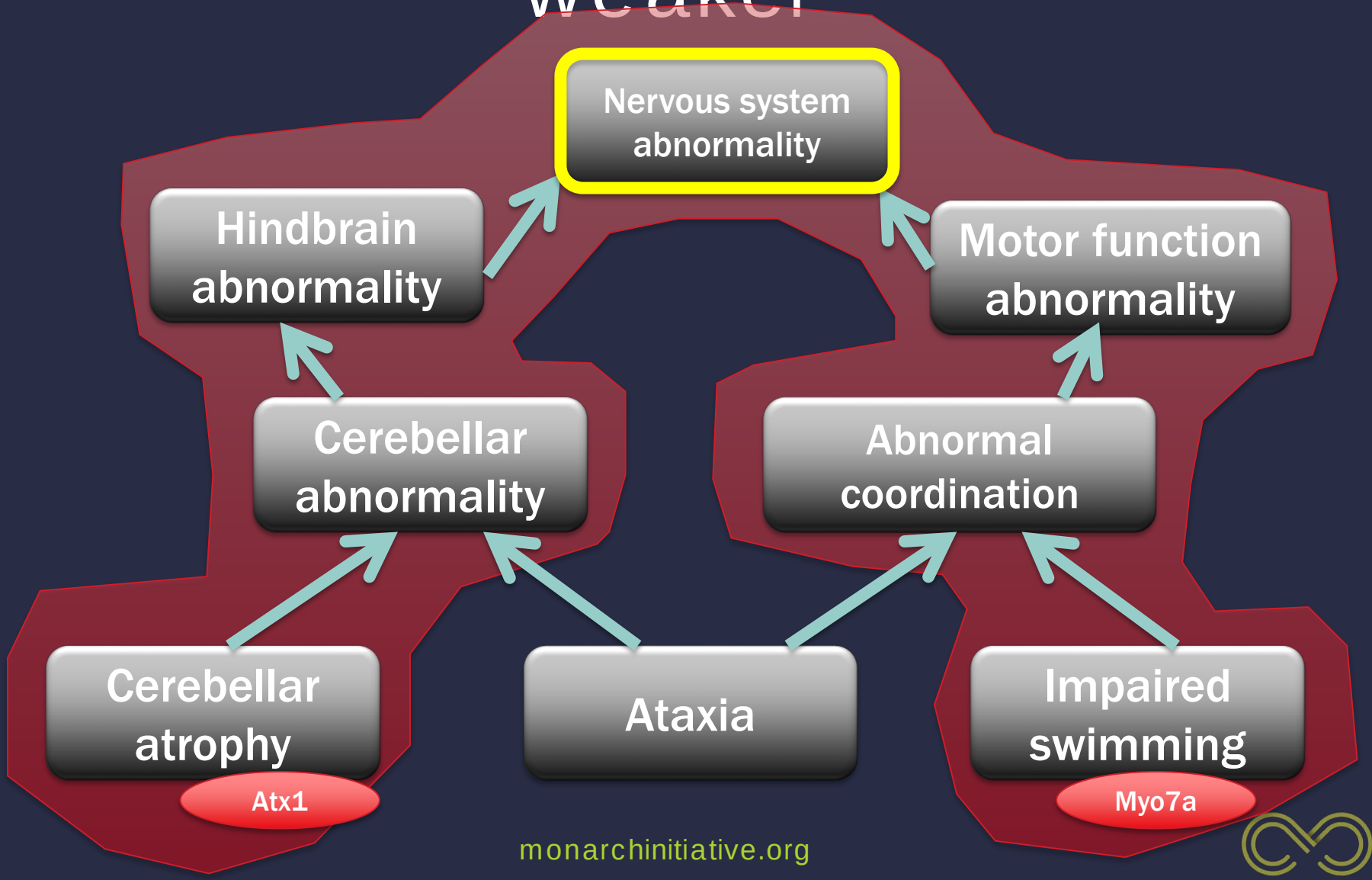
Graph-based similarity



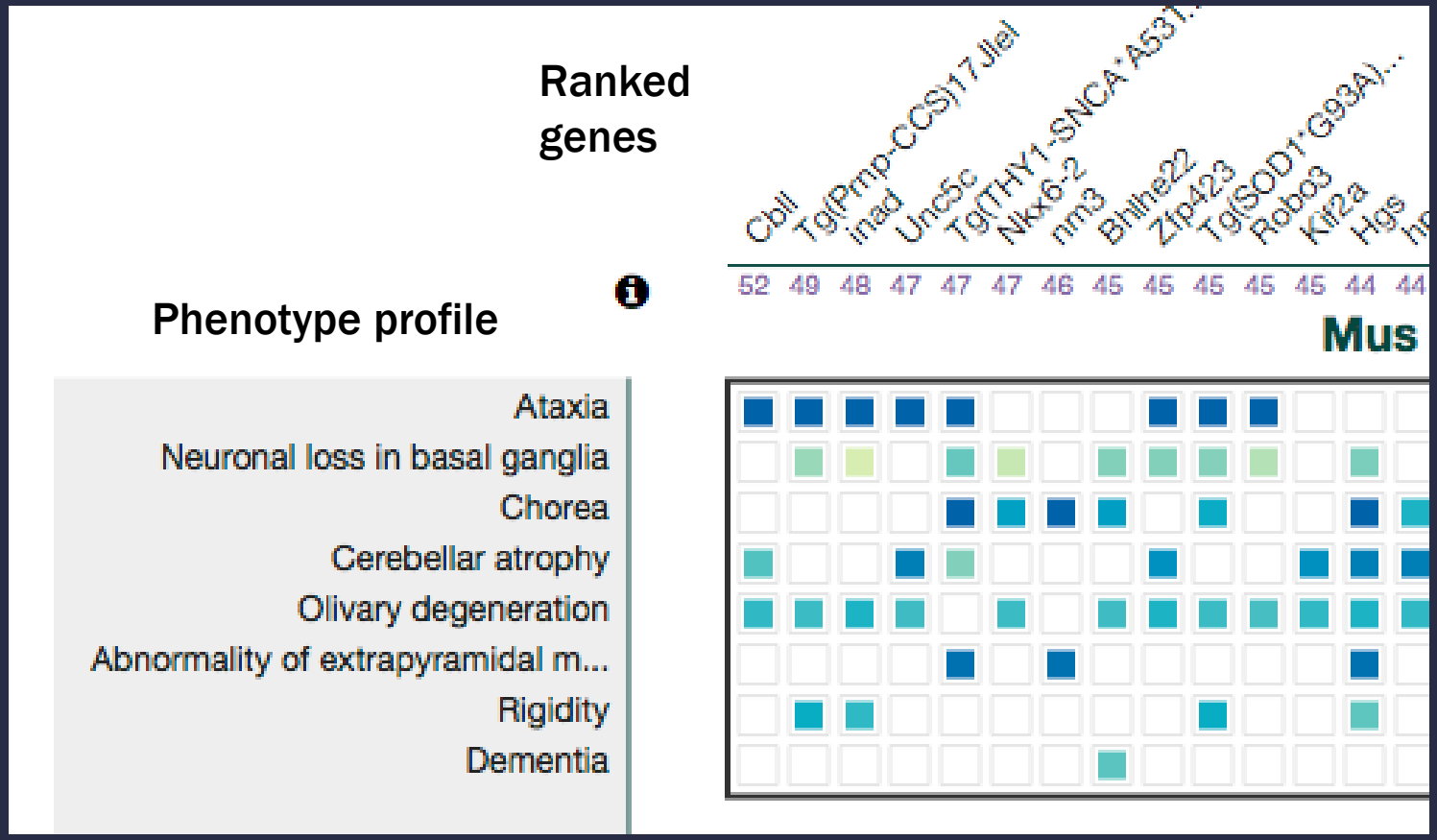
Similarity based on different criteria



Less specific matches are weaker

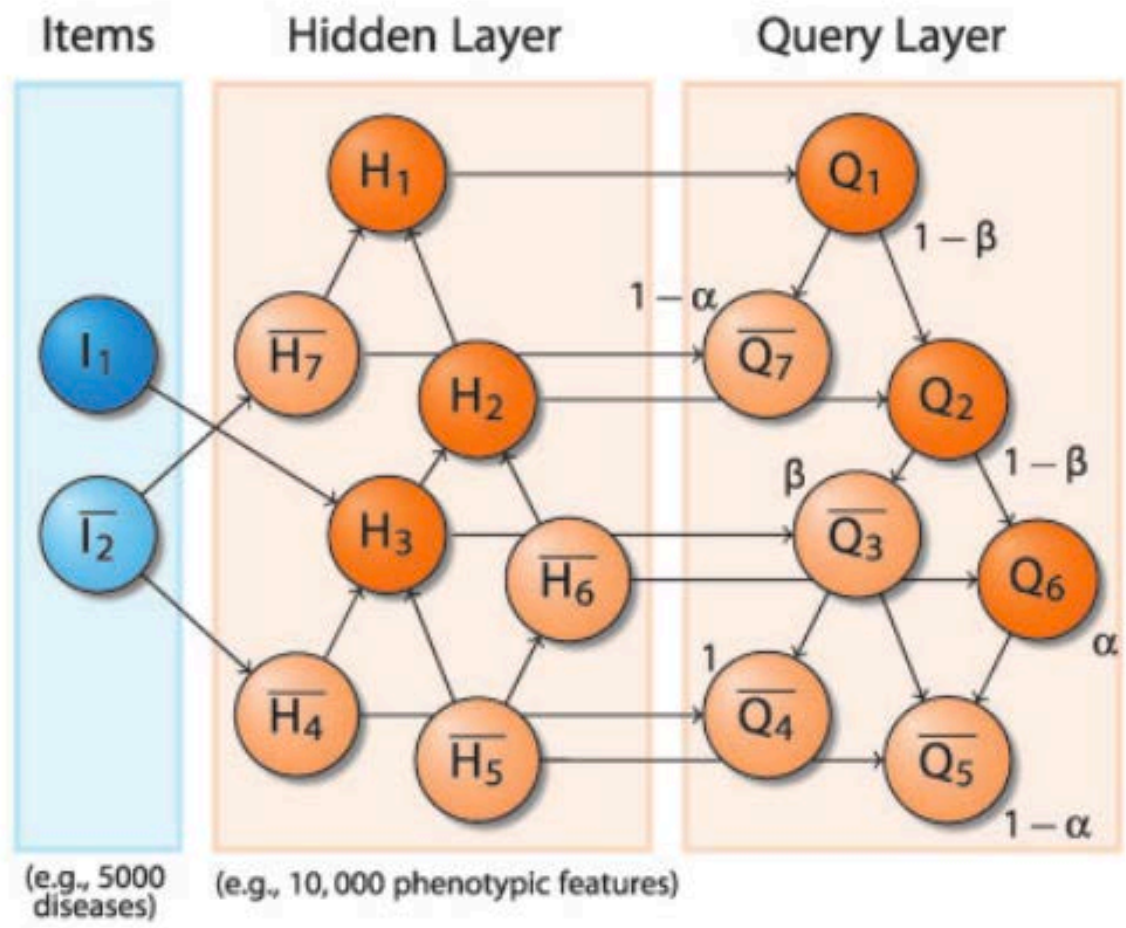


Multiple phenotypes can be compared at once



http://monarchinitiative.org/analyze/phenotypes

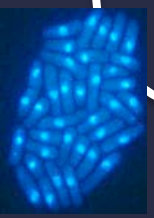
Probabilistic approaches



Explicit modeling of unobserved data

Bauer, S., Köhler, S., Schulz, M. H., & Robinson, P. N. (2012). Bayesian ontology querying for accurate and noise-tolerant semantic searches. Bioinformatics (Oxford, England), 28(19), 2502–8. doi:10.1093/bioinformatics/bts471

Challenge: multiple phenotype ontologies and terminologies



DPO

VT

MP

OMIA

WPO

ZFA

FYPO

APO

n/a

HP

OMIM

SNO MED

...

monarchinitiative.org

Helping machines understand phenotype terms

Human phenotype

"Degeneration of substantia nigra pars compacta"

I have absolutely no idea what that means

Mungall, C. J., Gkoutos, G., Smith, C., Haendel, M., Lewis, S., & Ashburner, M. (2010). Integrating phenotype ontologies across multiple species. *Genome Biology*, **11(1), R2. doi:10.1186/gb-2010-11-1-r2**

Decomposition of complex concepts

Human phenotype

"Degeneration of substantia nigra pars compacta"

=

PATO

degeneration

+

Uberon

Substantia nigra pas compacta

Species neutral ontologies, homologous concepts

Mungall, C. J., Gkoutos, G., Smith, C., Haendel, M., Lewis, S., & Ashburner, M. (2010). Integrating phenotype ontologies across multiple species. *Genome Biology*, **11(1), R2. doi:10.1186/gb-2010-11-1-r2**

Matching using subconcepts

Human phenotype

"Degeneration of substantia nigra pars compacta"

=

PATO

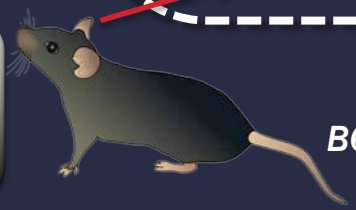
degeneration

+

Uberon

Substantia nigra pas compacta

"abnormal SN pars compacta morphology"



*B6.Cg-Tg(THY1-SNCA*A53T)M53Sud/J*

Source: MGI

Mungall, C. J., Gkoutos, G., Smith, C., Haendel, M., Lewis, S., & Ashburner, M. (2010). Integrating phenotype ontologies across multiple species. *Genome Biology*, **11(1), R2. doi:10.1186/gb-2010-11-1-r2**

Anatomical homology

Human phenotype

"Degeneration of substantia nigra pars compacta"

=

PATO

degeneration

+

Uberon

Substantia nigra pas compacta

Homologous to

Caudal tuberculum (teleost)

dla<hi781Tg>/dla<hi781Tg>; dld<tr233>/dld<tr233> [n.s.]

Mungall, C. J., Gkoutos, G., Smith, C., Haendel, M., Lewis, S., & Ashburner, M. (2010). Integrating phenotype ontologies across multiple species. *Genome Biology*, **11(1), R2. doi:10.1186/gb-2010-11-1-r2**

ZFIN

Cellular ontologies

Human phenotype

"Degeneration of substantia nigra pars compacta"

=

PATO

degeneration

has

GO

Cell death

Uberon

+

Substantia nigra pas compacta

has

Cell type

Dopamine neuron

Caudal tuberculum

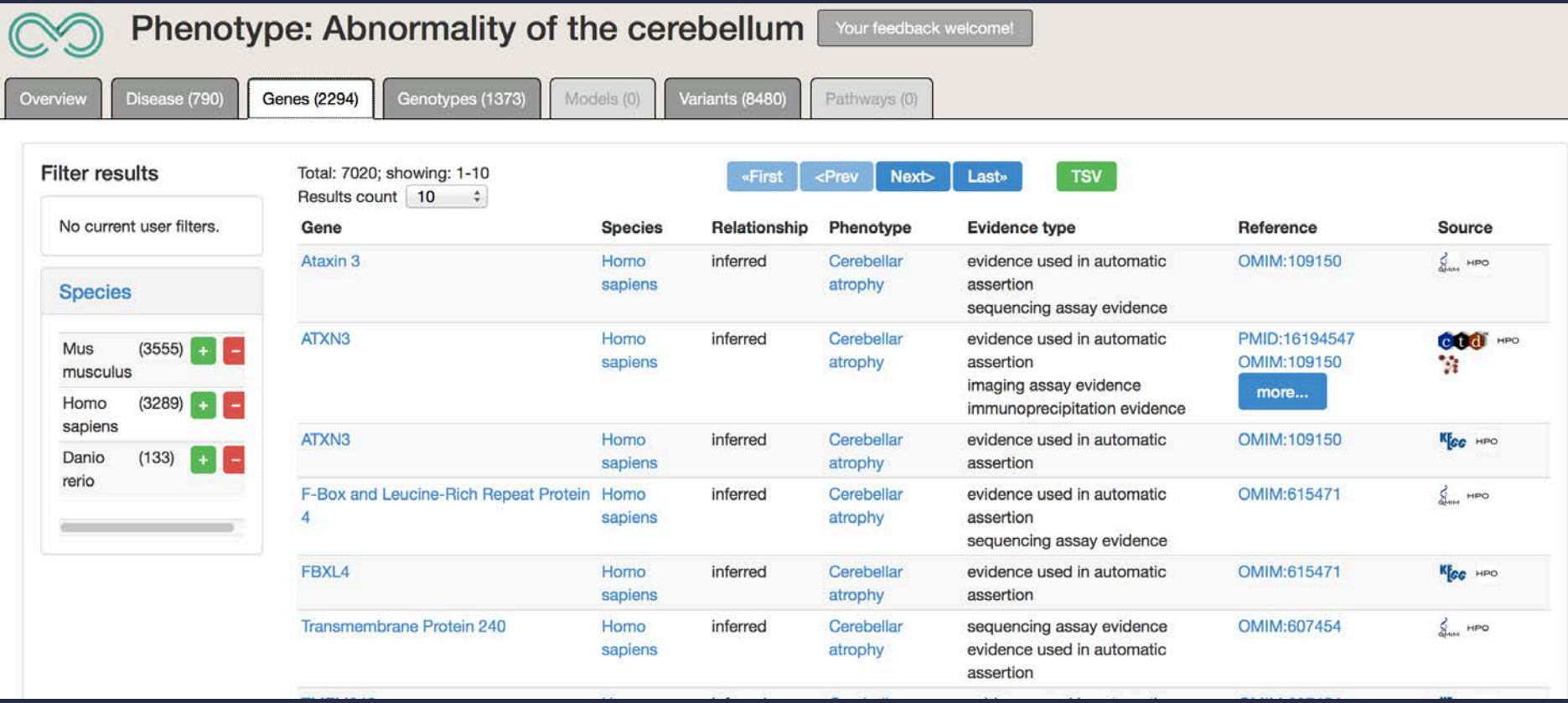
Mungall, C. J., Gkoutos, G., Smith, C., Haendel, M., Lewis, S., & Ashburner, M. (2010). Integrating phenotype ontologies across multiple species. *Genome Biology*, **11(1), R2. doi:10.1186/gb-2010-11-1-r2**

Monarch Ontology

- Integrates
  - Phenotypes
  - Diseases
    - DO, Orphanet, OMIM, ...
  - Species-neutral ontologies
    - Anatomy
    - GO
    - ...

https://github.com/monarch-initiative/monarch-ontology/

Köhler, S., Doelken, S. C., Ruef, B. J., Bauer, S., Washington, N., Westerfield, M., ... Mungall, C. J. (2013). Construction and accessibility of a cross-species phenotype ontology along with gene annotations for biomedical research. F1000Research, 1–12. doi:10.3410/f1000research.2-30.v1



http://.monarchinitiative.org/phenotype/HP:0001317

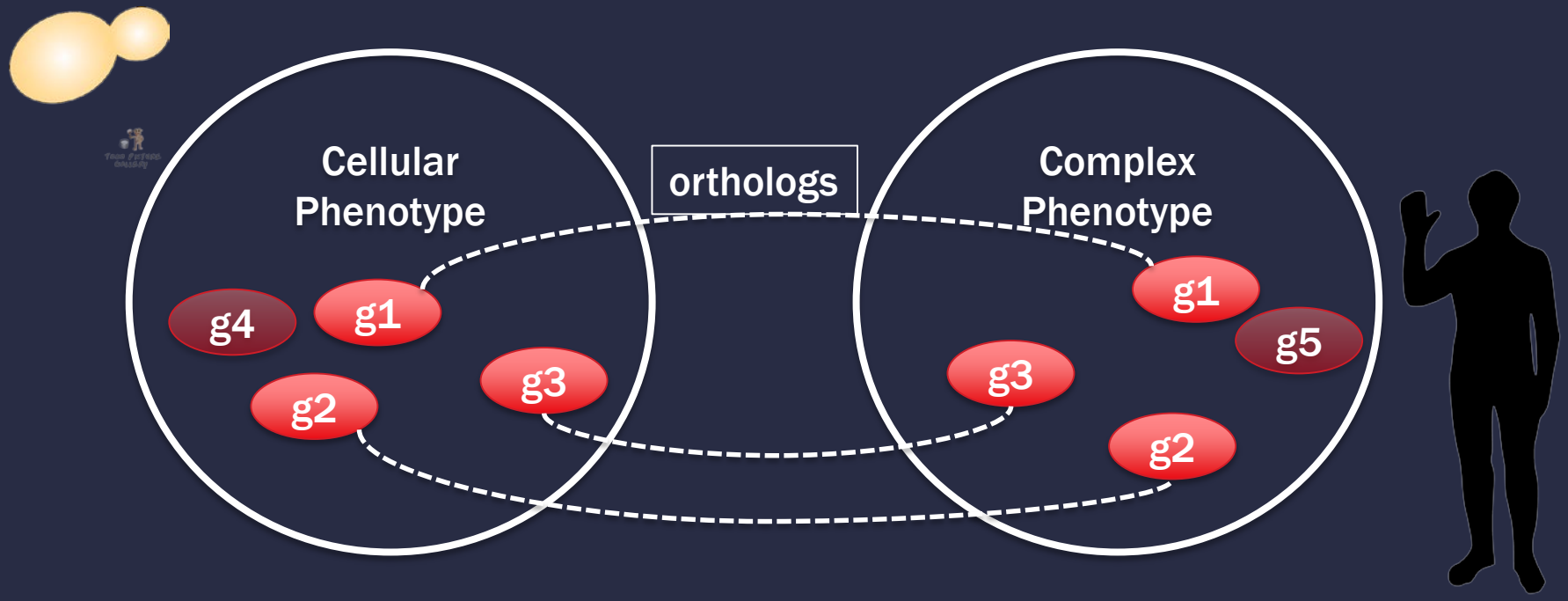
monarchinitiative.org

Network approaches for deep phenotype matching

Cellular Phenotype

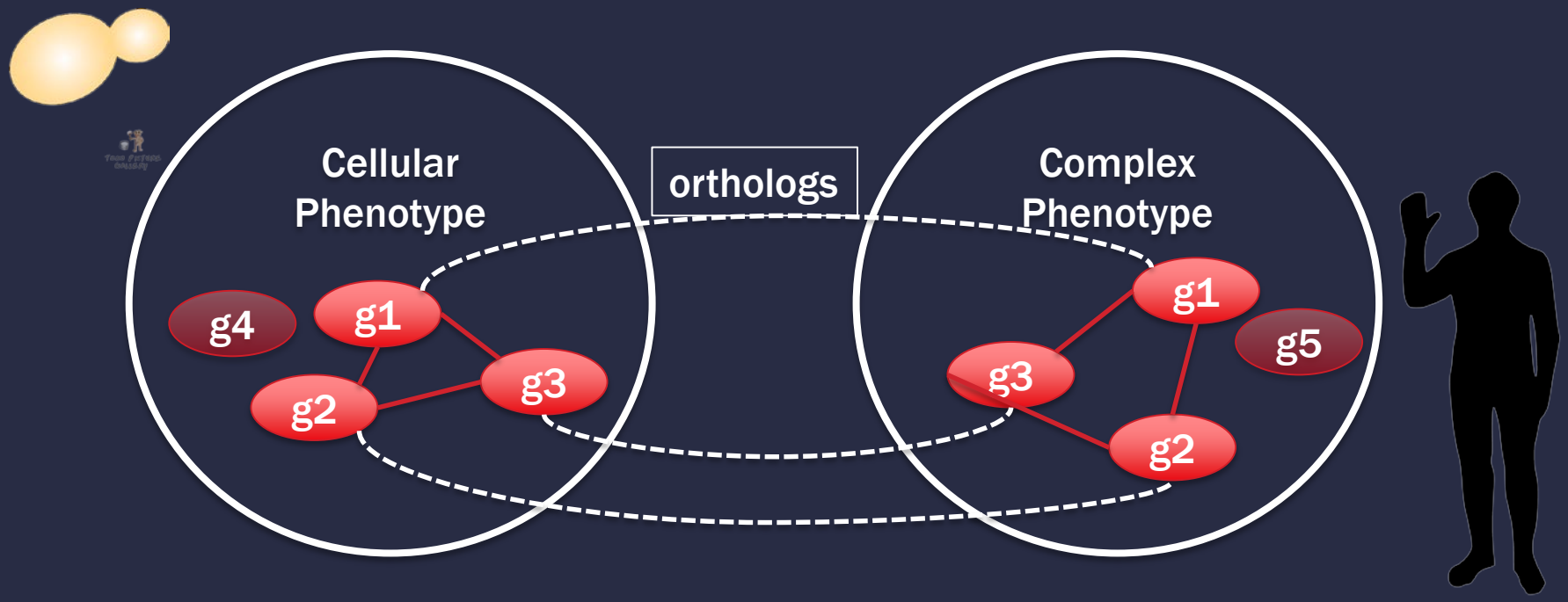
Complex Phenotype

Network approaches for deep phenotype matching



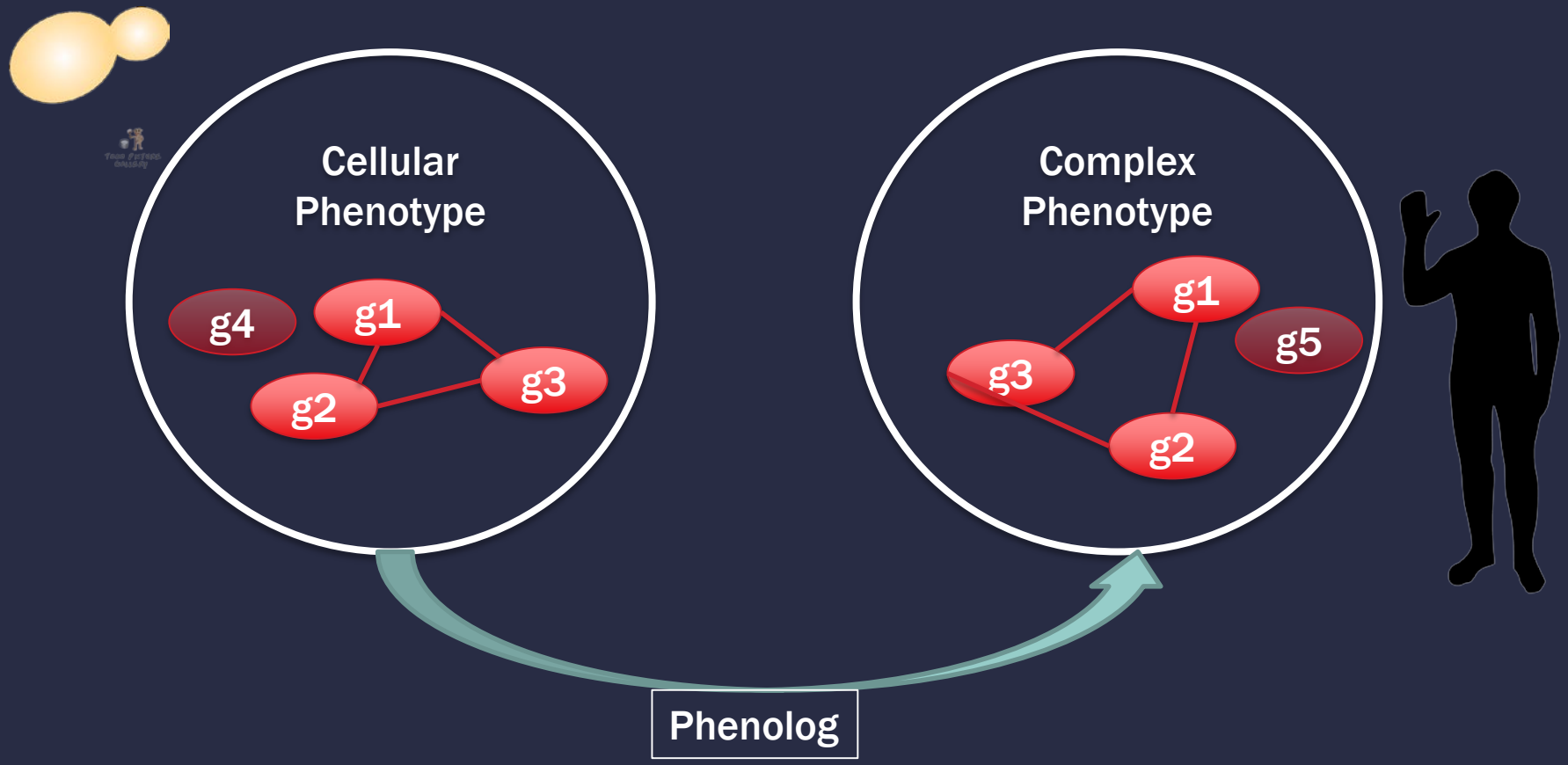
Distant phenotypes that involve the same genes

Network approaches for deep phenotype matching



Distant phenotypes that involve the same genes,
Similar interaction patterns

Network approaches for deep phenotype matching

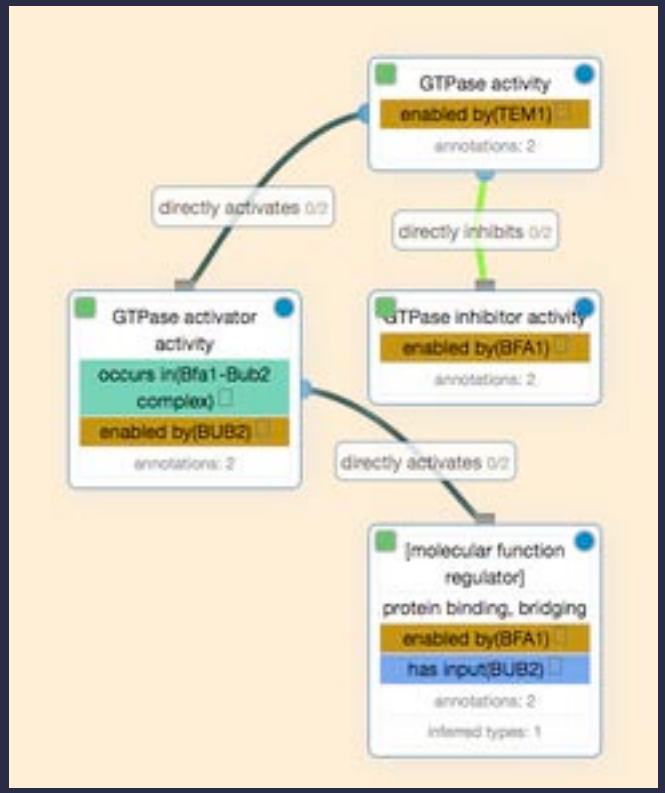


McWhite, C. D., et al Applications of comparative evolution to human disease genetics. *Current Opinion in Genetics & Development*, 35, 16–24. doi:10.1016/j.gde.2015.08.004

Next Steps

- Joint genome-phenome comparisons
  - Use in the clinic
  - Exomes to genomes
- Probabilistic modeling
  - evolutionary models of phenotypic change
  - Incorporation of negation uncertainty
- Ontologies to causal models
  - From genotype-phenotype associations
  - To temporal and causal models of phenotype progression

Next generation phenomic curation



http://create.monarchinitiative.org

http://noctua.berkeleybop.org

What we need

- Pan-species approach
  - Comparative phenomics perspective
  - Enable collaboration
  - Easy to integrate new organisms
- Ramp up knowledge curation
  - Provide support for expert curators
  - Develop tools to enhance efficiency
  - Use open standards
- Integrated phenotype curation
  - Ontology
  - Genotype-phenotype curation
  - Causal networks

Acknowledgments

- LBNL
  - Nicole Washington
  - Jeremy Nguyen-Xuan
  - Suzanna Lewis
- OHSU
  - Melissa Haendel
  - Nicole Vasilesky
  - Matt Brush
  - Kent Shefchek
- Charite
  - Peter Robinson
  - Sebastian Koehler
- Sanger
  - Damian Smedley
  - Anika Ohlrich
- Garvan
  - Tudor Graza
- University of Pittsburg
  - Harry Hochheiser
- USC/GO
  - Paul Thomas

- Model Organism Database, Phenotyping Resources and Human Phenomics Resource curators
  - MGI
  - RGD
  - ZFIN
  - Xenbase
  - FlyBase
  - WormBase
  - SGD
  - PomBase
  - OMIA
  - OMIM
  - Orphanet
  - ...

NIH R24OD011883.

END