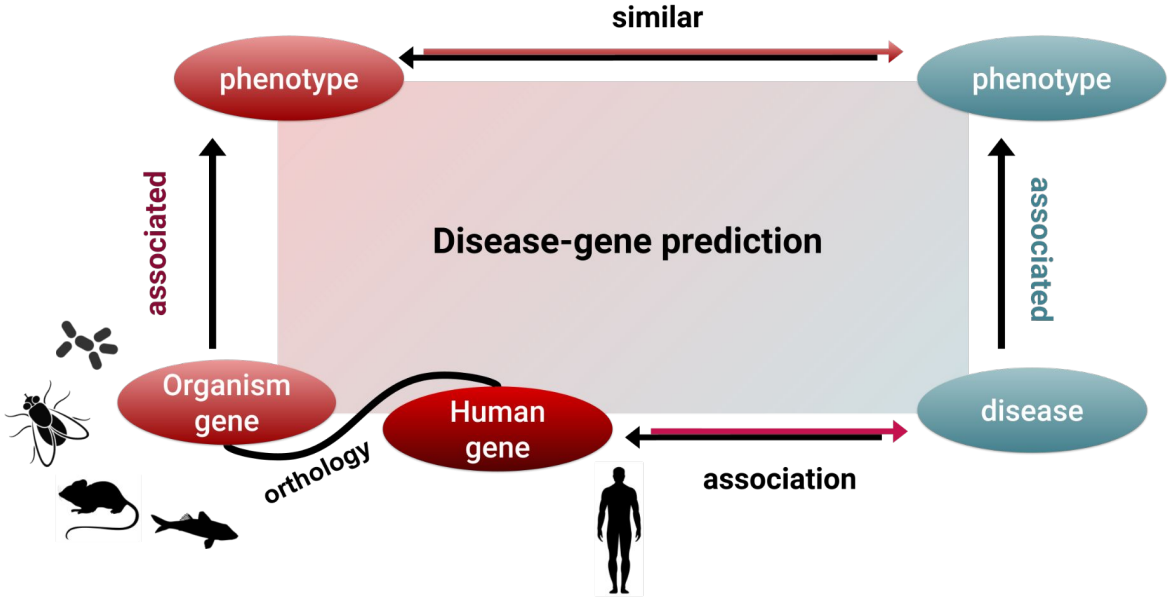
Applications of Ontology Embeddings

- Predicting gene-disease associations based on phenotypic similarity
- Diagnosis of disease based on phenotypic similarity
- Predict protein–protein interactions based on their functional similarity

….

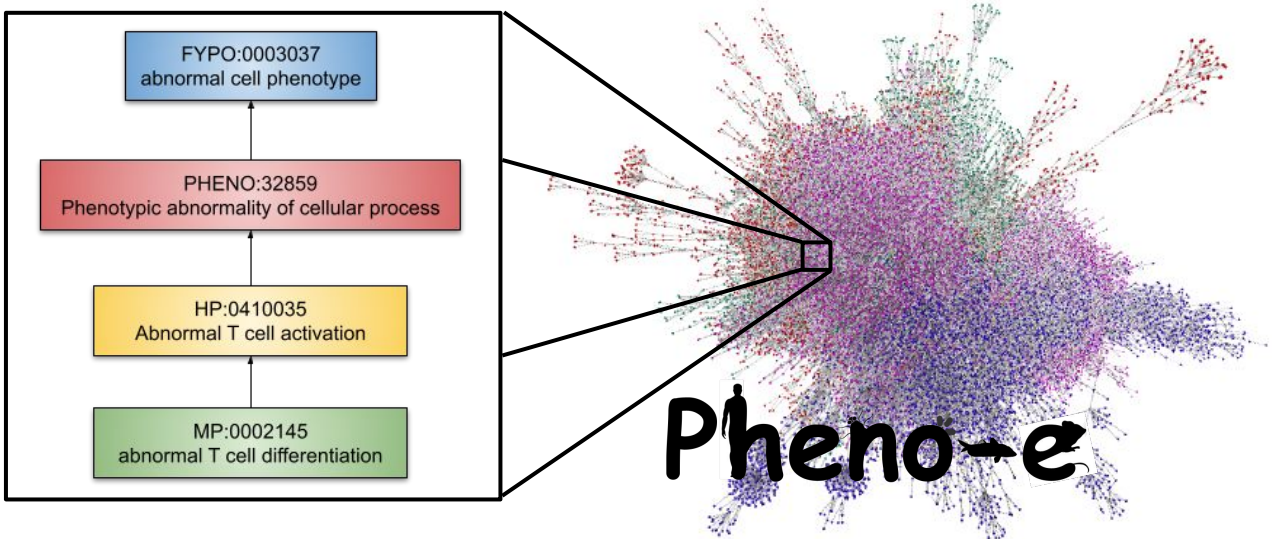
Predicting gene-disease associations

- Based on Phenotypic similarity
- Using the phenotypes of model organism genes and the diseases' phenotypes



Predicting gene-disease associations

- PhenomeNet-Extension **(Pheno-e)** and **uPheno** are cross-species phenotype ontologies that can be utilized here.
- Both with the objective of allowing similar phenotypes from the same or different organisms ontologies to be logically defined in similar form.



Example:

Isolated anhidrosis with normal morphology and number sweat glands (ANHD)

Itpr2
inositol 1,4,5-triphosphate receptor 2

Human disease phenotypes:

Generalized anhidrosis

Heat intolerance

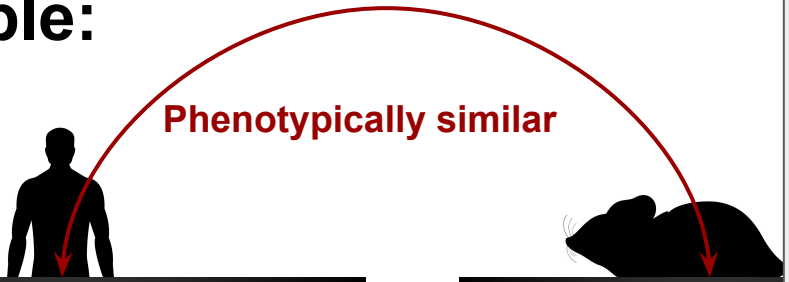
Anhidrosis

Mouse gene phenotypes:

abnormal sweat gland physiology

Hypohidrosis

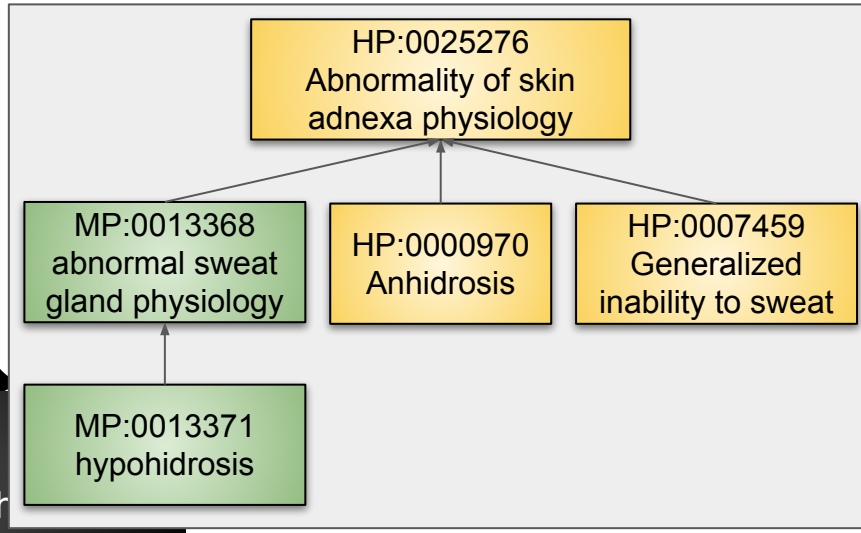
Example:



Phenotypically similar

Isolated anhidrosis with normal morphology and number sweat glands (ANHD)

Itpr2
inositol 1,4,5-triph
receptor 2



Human disease phenotypes:

Generalized anhidrosis

Heat intolerance

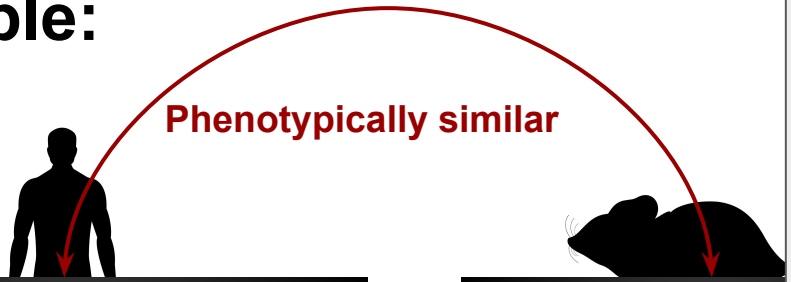
Anhidrosis

Mouse gene phenotypes:

abnormal sweat gland physiology

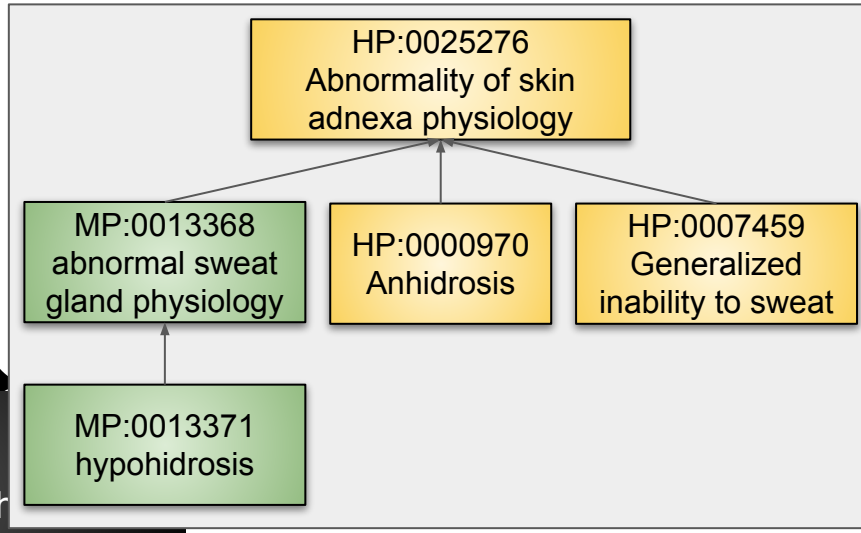
Hypohidrosis

Example:



Isolated anhidrosis with normal morphology and number sweat gla
(ANHD)

Itpr2
inositol 1,4,5-triph
receptor 2



How do we calculate the phenotypic similarity between genes and diseases using ontology embbeddings?

Human disease
phenotypes:

Generalized anhidrosis

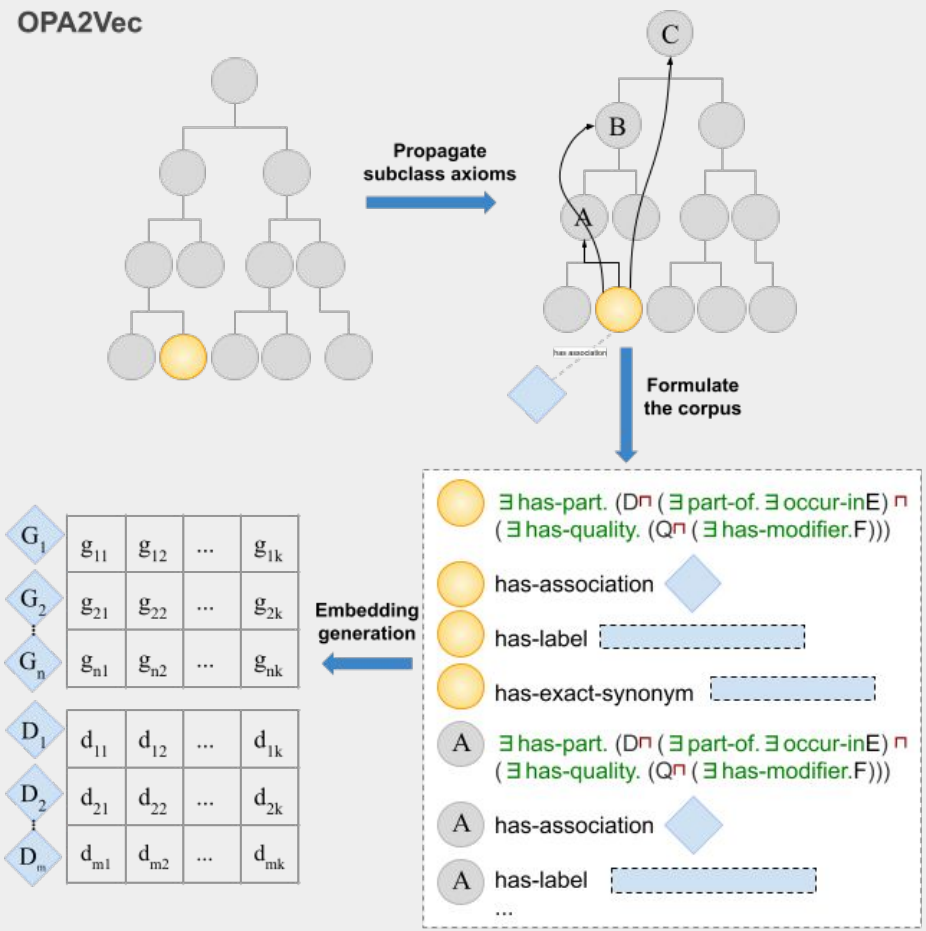
Heat intolerance

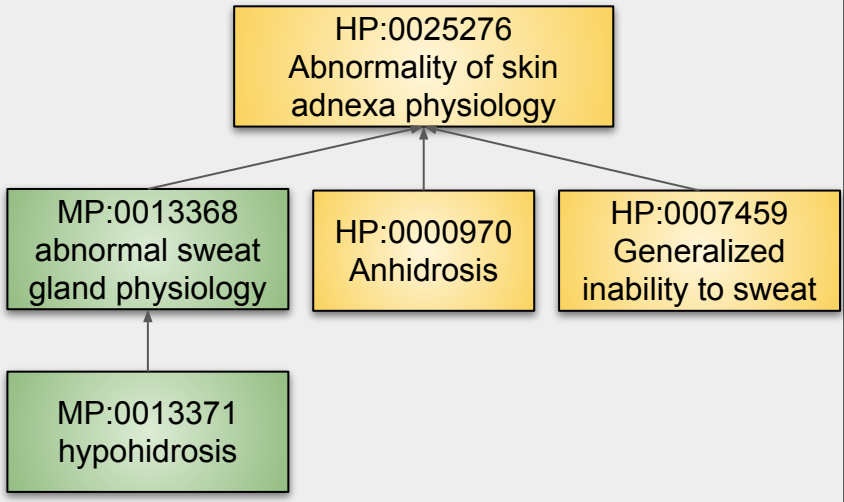
Anhidrosis

phenotypes:

abnormal sweat gland physiology

Hypohidrosis


OPA2Vec
Propagate subclass axioms
C
B
A
Formulate the corpus
∃ has-part. (D⊓ (∃ part-of. ∃ occur-inE) ⊓ (∃ has-quality. (Q⊓ (∃ has-modifier.F)))
has-association
has-label
has-exact-synonym
A ∃ has-part. (D⊓ (∃ part-of. ∃ occur-inE) ⊓ (∃ has-quality. (Q⊓ (∃ has-modifier.F)))
A has-association
A has-label
...
Embedding generation
G1 g11 g12 ... g1k
G2 g21 g22 ... g2k
Gn gn1 gn2 ... gnk
D1 d11 d12 ... d1k
D2 d21 d22 ... d2k
Dm dm1 dm2 ... dmk

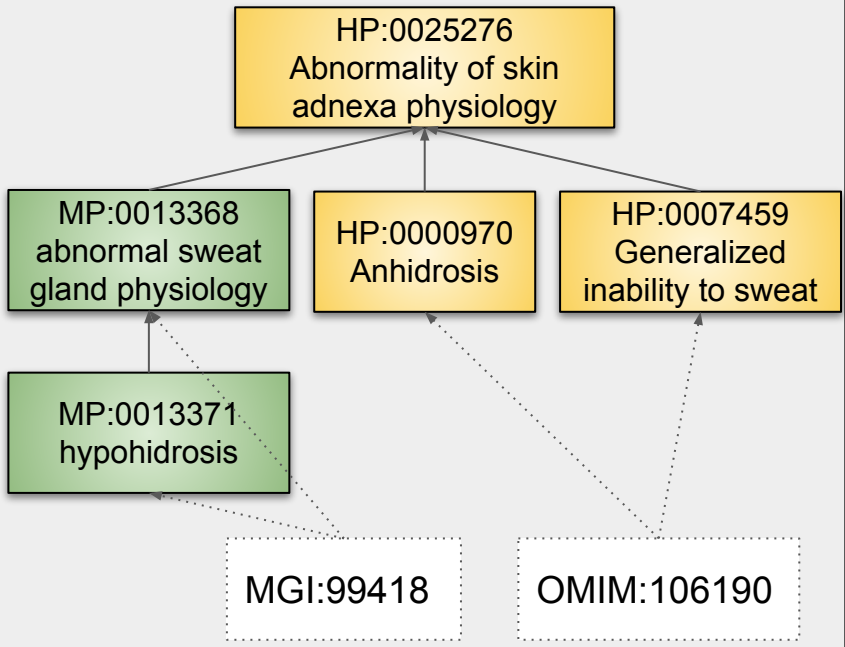

HP:0025276
Abnormality of skin adnexa physiology
MP:0013368
abnormal sweat gland physiology
HP:0000970
Anhidrosis
HP:0007459
Generalized inability to sweat
MP:0013371
hypohidrosis

OMIM:106190 has_annotaion **HP:0007459**
HP:0007459 has_label Generalized anhidrosis
HP:0007459 is_a **HP:0025276**
HP:0007459 has_synonyms Generalized anhydrosis
HP:0007459 has_synonyms Generalized inability to sweat

OMIM:106190 has_annotaion **HP:0000970**
HP:0000970 has_label Anhidrosis
HP:0000970 is_a **HP:0025276**
HP:0000970 has_database_cross_reference SNOMEDCT_US:39659002
HP:0000970 has_database_cross_reference UMLS:C0003028
HP:0000970 has_database_cross_reference MSH:D007007
HP:0000970 has_database_cross_reference SNOMEDCT_US:14662005
HP:0000970 has_database_cross_reference MEDDRA:10002512
HP:0000970 has_synonyms Anhydrosis
HP:0000970 has_synonyms Sweating dysfunction
HP:0000970 has_synonyms Sudomotor dysfunction,
HP:0000970 has_synonyms Lack of sweating

MGI:99418 has_annotaion **MP:0013368**
MP:0013368 Equivilant_to has_part some (functionality and (characteristic of some sweat gland) and (has modifier some abnormal))
MP:0013368 has_label abnormal sweat gland physiology
MP:0013368 has_definition Inability to sweat
MP:0013368 is_a **HP:0025276**
MP:0013368 has_synonyms sudomotor dysfunction
MP:0013368 has_synonyms abnormal sweat response
MP:0013368 has_synonyms sweating dysfunction

MGI:99418 has_annotaion **MP:0013371**
……



Word2Vec

Embedding

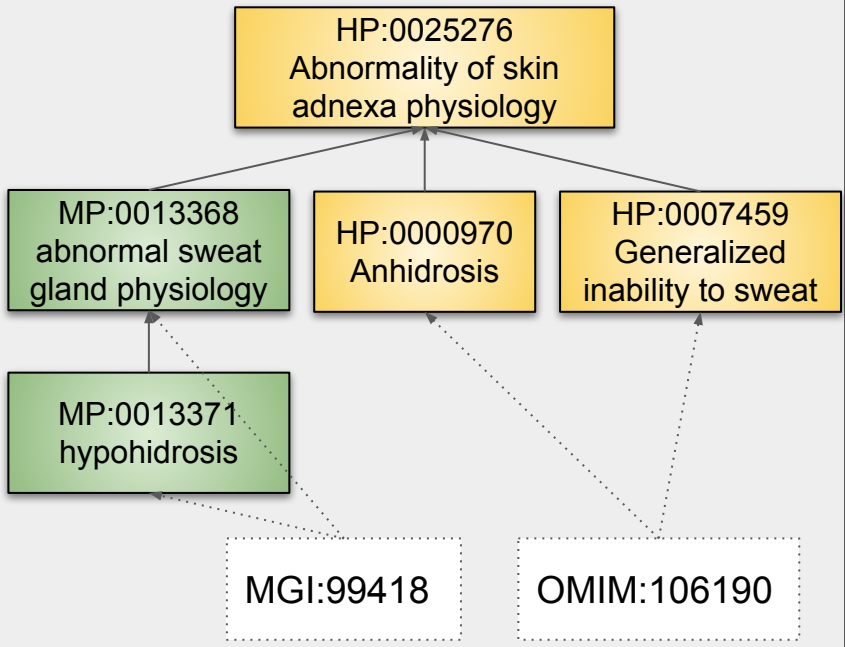
OMIM:106190 [-0.3482, -0.2413, 0.6085, 0.0490, …]
MGI:99418 [-0.5776, 0.0502, 0.0963, -0.2741, …]

OMIM:106190 has_annotaion **HP:0007459**
HP:0007459 has_label Generalized anhidrosis
HP:0007459 is_a **HP:0025276**
HP:0007459 has_synonyms Generalized anhydrosis
HP:0007459 has_synonyms Generalized inability to sweat

OMIM:106190 has_annotaion **HP:0000970**
HP:0000970 has_label Anhidrosis
HP:0000970 is_a **HP:0025276**
HP:0000970 has_database_cross_reference SNOMEDCT_US:39659002
HP:0000970 has_database_cross_reference UMLS:C0003028
HP:0000970 has_database_cross_reference MSH:D007007
HP:0000970 has_database_cross_reference SNOMEDCT_US:14662005
HP:0000970 has_database_cross_reference MEDDRA:10002512
HP:0000970 has_synonyms Anhydrosis
HP:0000970 has_synonyms Sweating dysfunction
HP:0000970 has_synonyms Sudomotor dysfunction,
HP:0000970 has_synonyms Lack of sweating

MGI:99418 has_annotaion **MP:0013368**
MP:0013368 Equivilant_to has_part some (functionality and (characteristic of some sweat gland) and (has modifier some abnormal))
MP:0013368 has_label abnormal sweat gland physiology
MP:0013368 has_definition Inability to sweat
MP:0013368 is_a **HP:0025276**
MP:0013368 has_synonyms sudomotor dysfunction
MP:0013368 has_synonyms abnormal sweat response
MP:0013368 has_synonyms sweating dysfunction

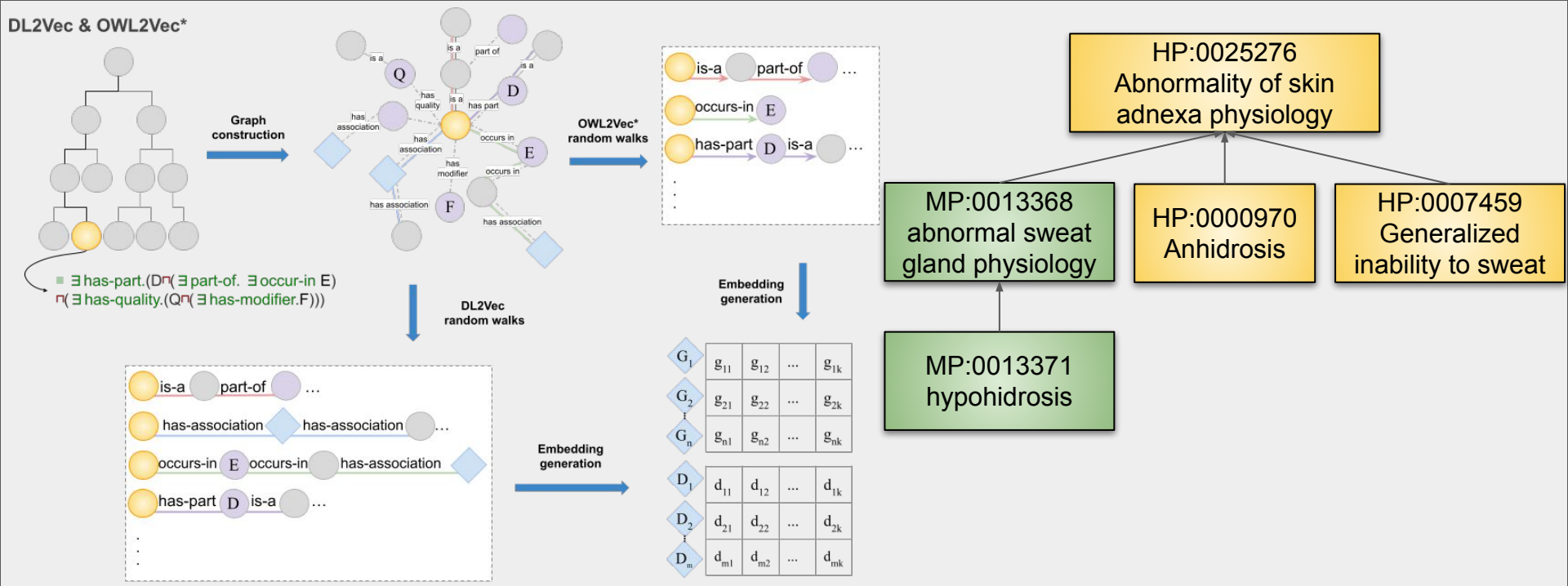
MGI:99418 has_annotaion **MP:0013371**
……

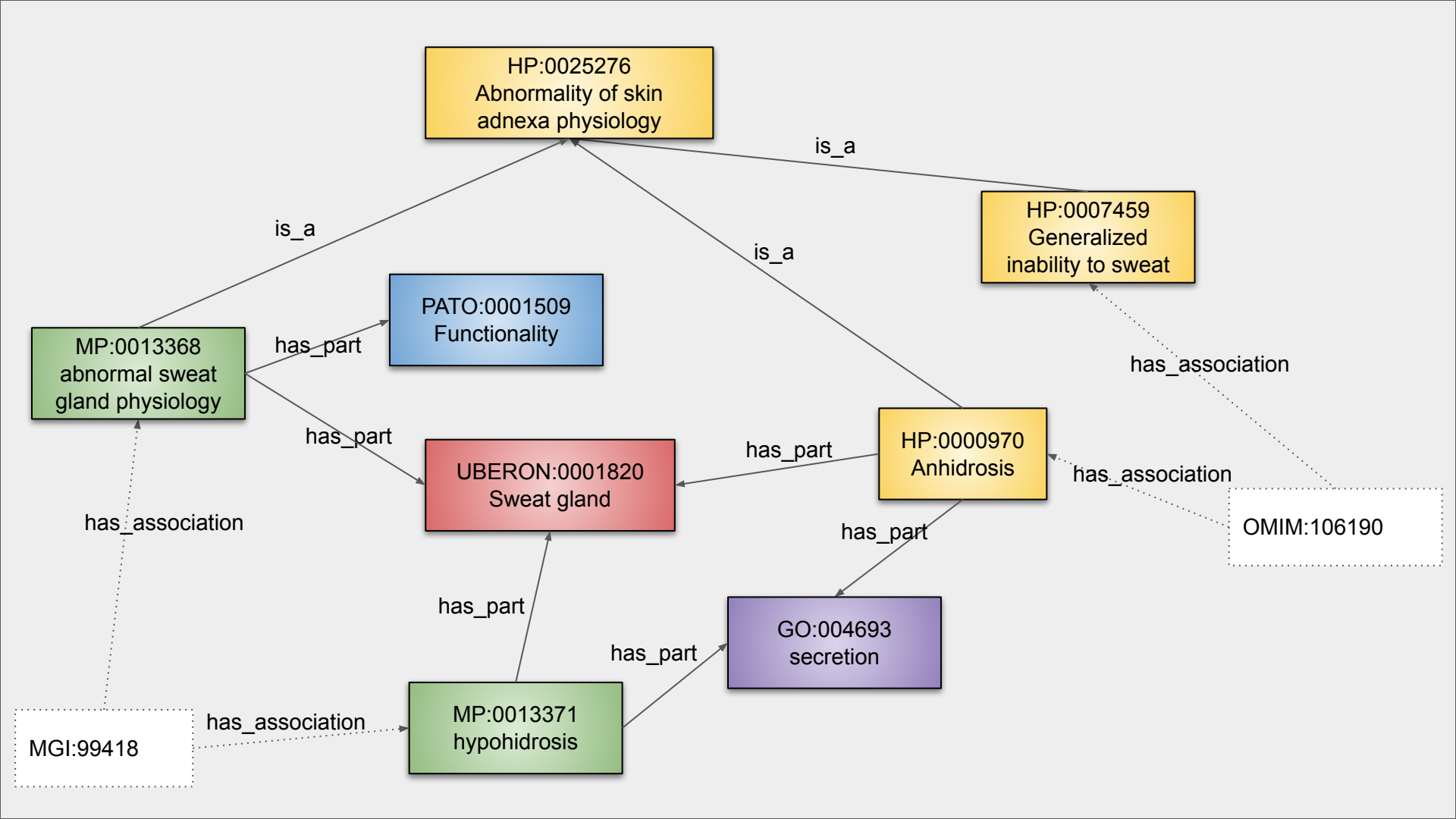


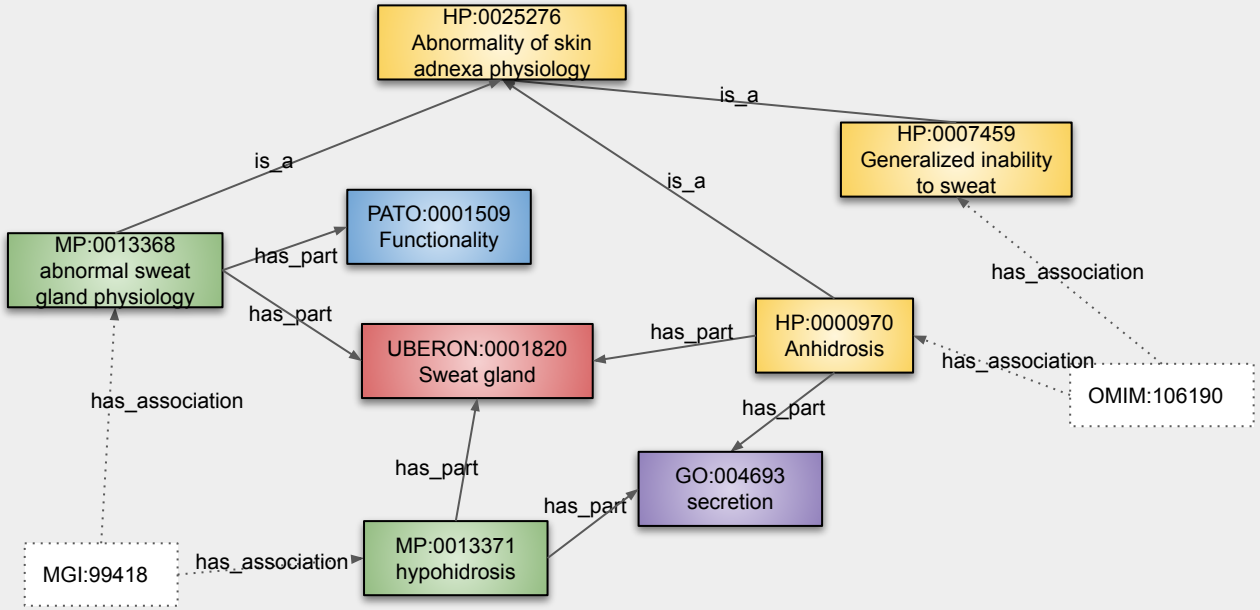
Word2Vec

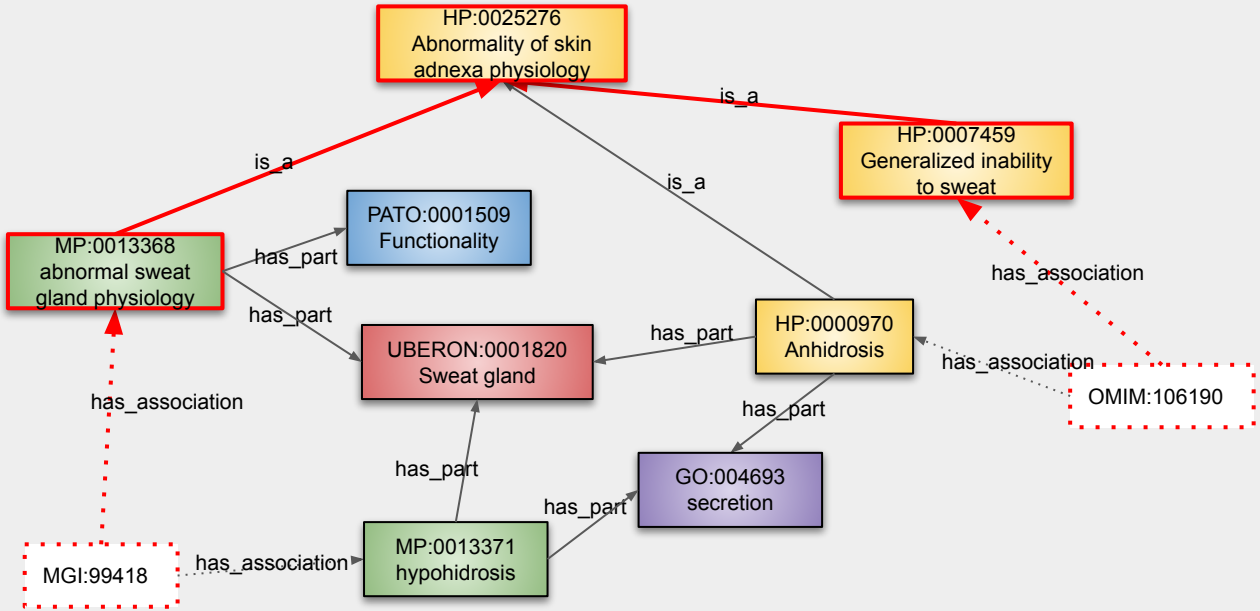
Embedding

OMIM:106190 [-0.3482, -0.2413, 0.6085, 0.0490, …]
MGI:99418 [-0.5776, 0.0502, 0.0963, -0.2741, …]

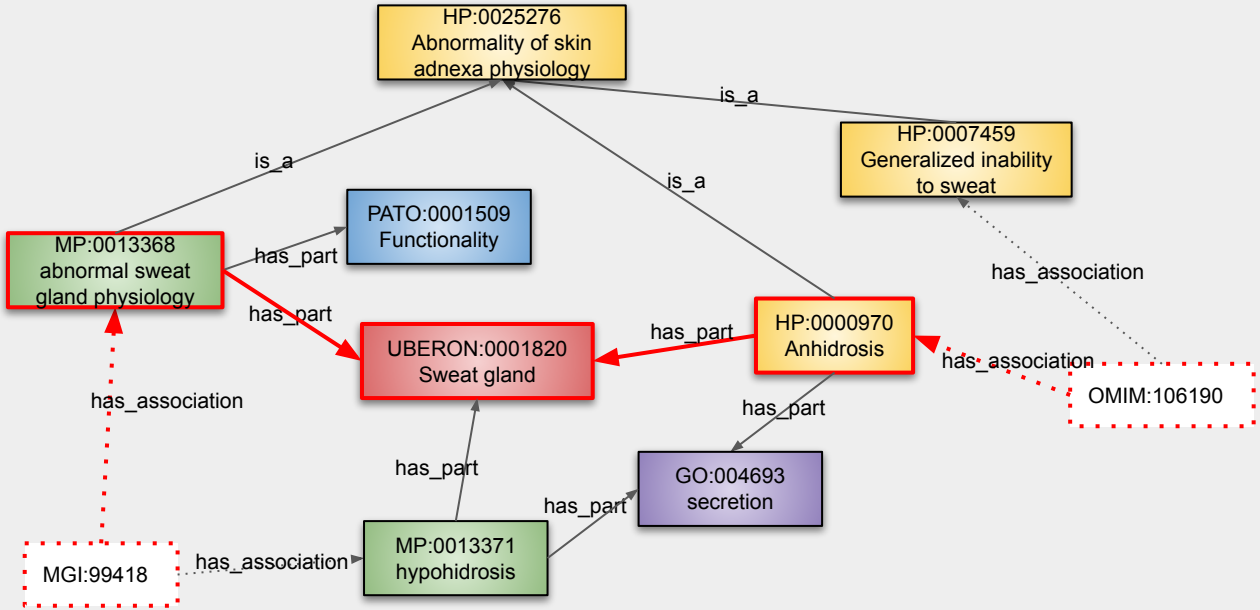

DL2Vec & OWL2Vec*
Graph construction
OWL2Vec* random walks
∃ has-part.(D⊓(∃ part-of. ∃ occur-in E)
⊓(∃ has-quality.(Q⊓(∃ has-modifier.F)))
is-a
part-of
occurs-in
has-part
has-association
has-quality
has-modifier
DL2Vec random walks
Embedding generation
Embedding generation
HP:0025276
Abnormality of skin adnexa physiology
MP:0013368
abnormal sweat gland physiology
HP:0000970
Anhidrosis
HP:0007459
Generalized inability to sweat
MP:0013371
hypohidrosis


HP:0025276
Abnormality of skin adnexa physiology
is_a
HP:0007459
Generalized inability to sweat
is_a
is_a
MP:0013368
abnormal sweat gland physiology
has_part
PATO:0001509
Functionality
has_association
has_part
UBERON:0001820
Sweat gland
has_part
HP:0000970
Anhidrosis
has_association
OMIM:106190
has_association
has_part
has_part
GO:004693
secretion
has_part
MGI:99418
has_association
MP:0013371
hypohidrosis


HP:0025276
Abnormality of skin adnexa physiology
is_a
HP:0007459
Generalized inability to sweat
is_a
is_a
MP:0013368
abnormal sweat gland physiology
has_part
PATO:0001509
Functionality
has_part
UBERON:0001820
Sweat gland
has_part
HP:0000970
Anhidrosis
has_association
has_association
OMIM:106190
has_association
has_part
has_part
GO:004693
secretion
has_part
MP:0013371
hypohidrosis
has_association
MGI:99418

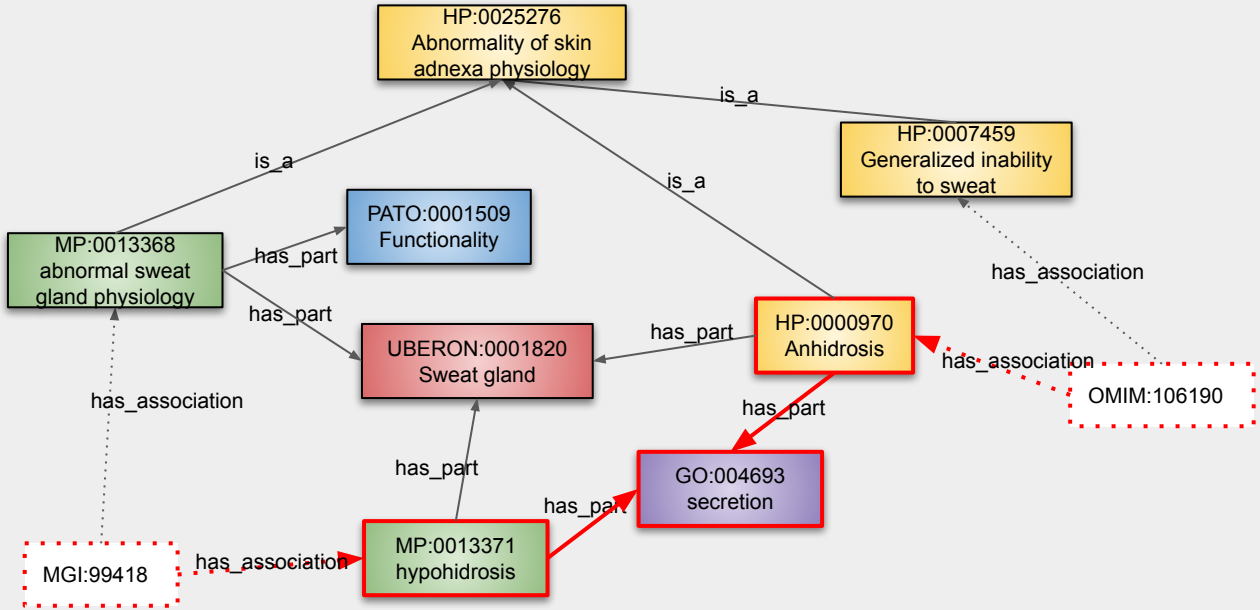


OMIM:106190 has_association HP:0007459 is_a HP:0025276 is_a MP:0013368 has_association **MGI:99418** …



OMIM:106190 has_association HP:0007459 is_a HP:0025276 is_a MP:0013368 has_association **MGI:99418** …

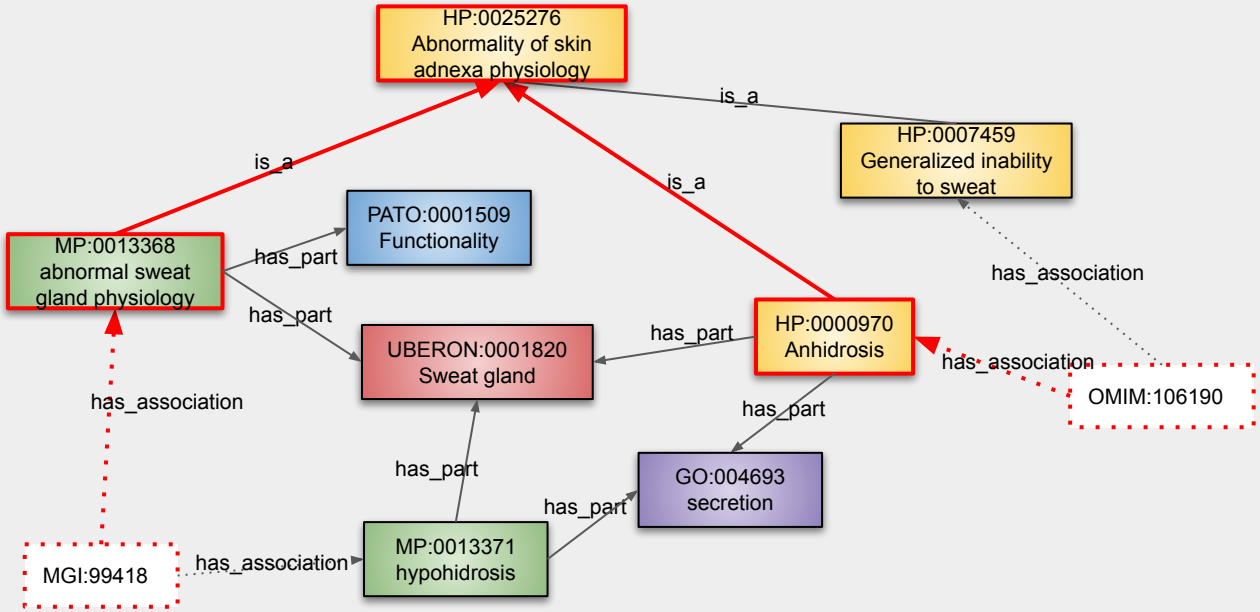
OMIM:106190 has_association HP:0000970 has_part UBERON:0001820 has_part MP:0013368 has_association **MGI:99418** …



OMIM:106190 has_association HP:0007459 is_a HP:0025276 is_a MP:0013368 has_association **MGI:99418** …

OMIM:106190 has_association HP:0000970 has_part UBERON:0001820 has_part MP:0013368 has_association **MGI:99418** …

OMIM:106190 has_association HP:0000970 has_part GO:004693 has_part MP:0013371 has_association **MGI:99418 …**

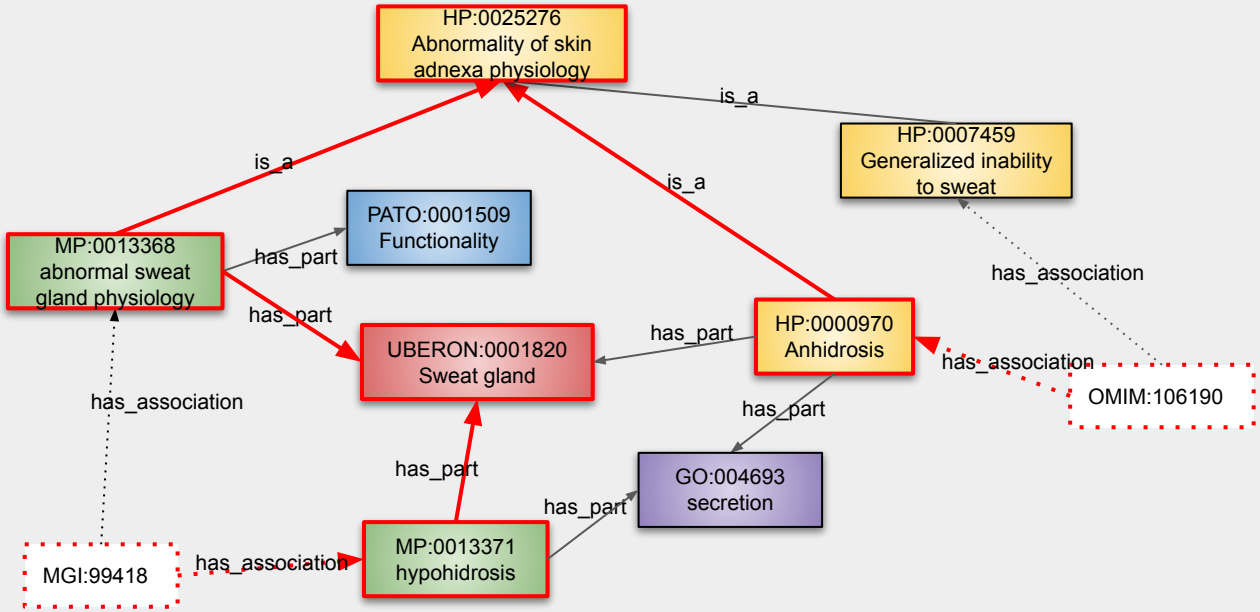


OMIM:106190 has_association HP:0007459 is_a HP:0025276 is_a MP:0013368 has_association **MGI:99418** …

OMIM:106190 has_association HP:0000970 has_part UBERON:0001820 has_part MP:0013368 has_association **MGI:99418** …

OMIM:106190 has_association HP:0000970 has_part GO:004693 has_part MP:0013371 has_association **MGI:99418 …**

OMIM:106190 has_association HP:0000970 is_a HP:0025276 is_a MP:0013368 has_association **MGI:99418** …



OMIM:106190 has_association HP:0007459 is_a HP:0025276 is_a MP:0013368 has_association **MGI:99418** …

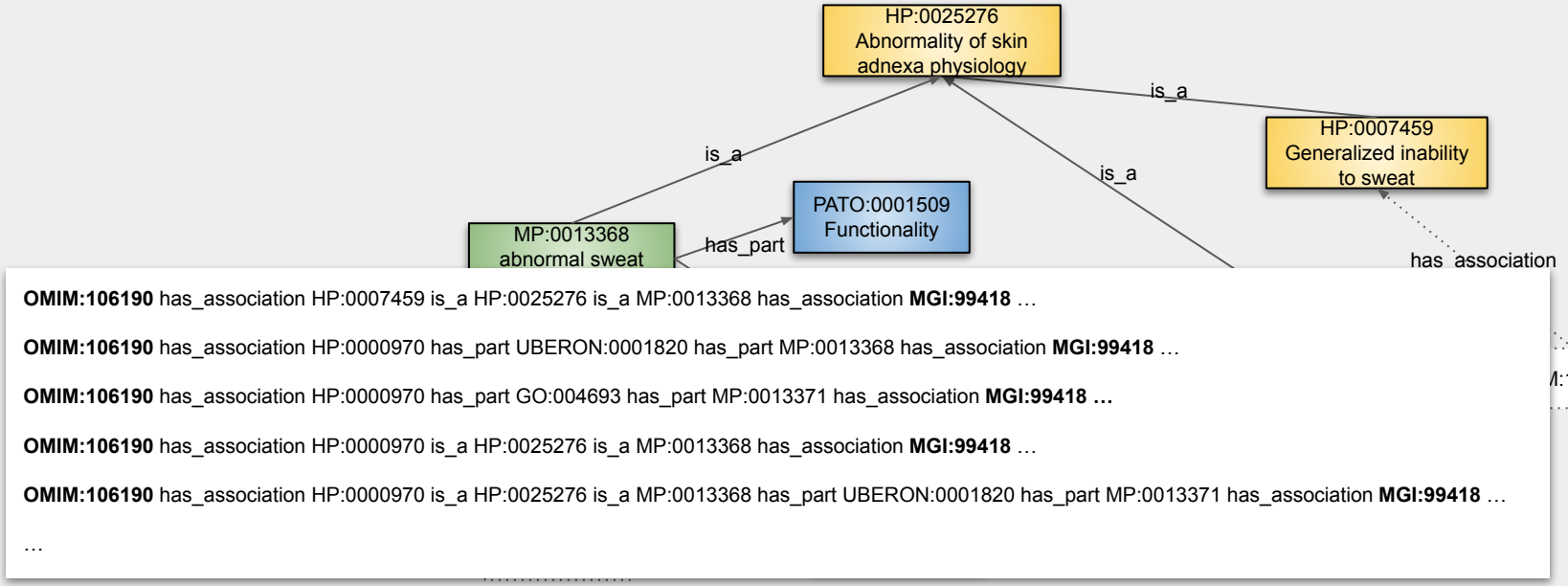
OMIM:106190 has_association HP:0000970 has_part UBERON:0001820 has_part MP:0013368 has_association **MGI:99418** …

OMIM:106190 has_association HP:0000970 has_part GO:004693 has_part MP:0013371 has_association **MGI:99418 …**

OMIM:106190 has_association HP:0000970 is_a HP:0025276 is_a MP:0013368 has_association **MGI:99418** …

OMIM:106190 has_association HP:0000970 is_a HP:0025276 is_a MP:0013368 has_part UBERON:0001820 has_part MP:0013371 has_association **MGI:99418** …

…



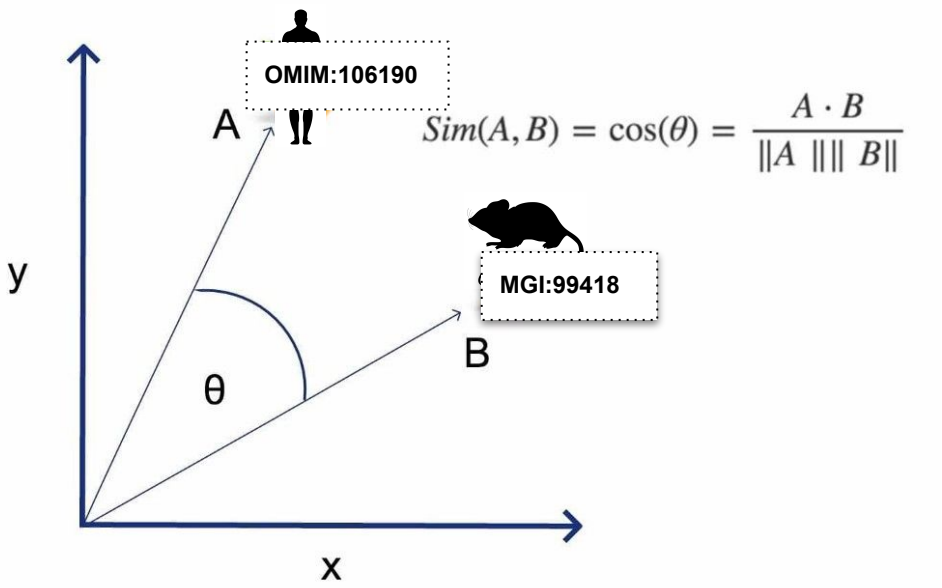
Embedding

OMIM:106190
MGI:99418

[0.6909, 0.6992, 0.2646, -0.0663, …]
[0.4071, 0.8932, -0.1988, -0.7038, …]

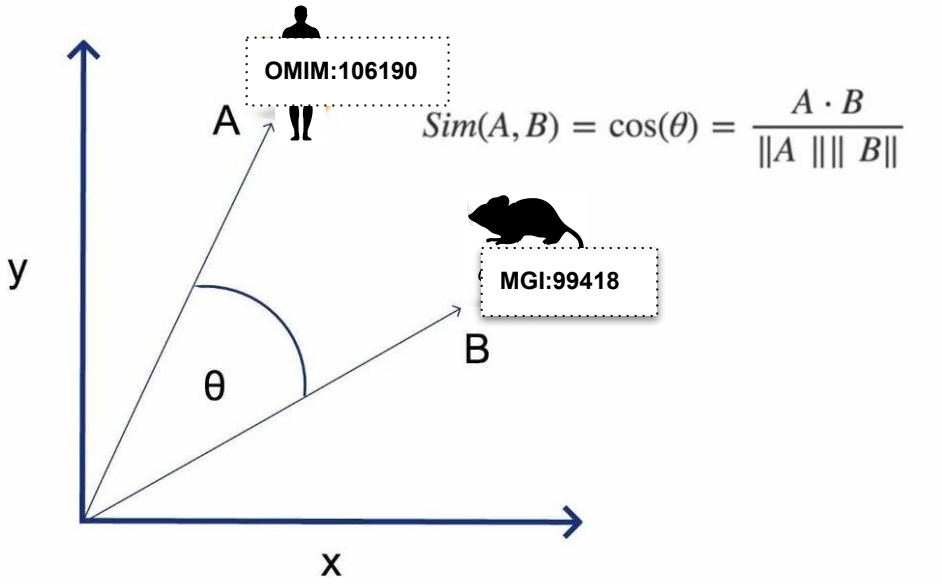
Calculating Phenotypic Similarity Approaches

- Unsupervised Approach
  - Cosine similarity



Calculating Phenotypic Similarity Approaches

- Unsupervised Approach
  - Cosine similarity



diseases

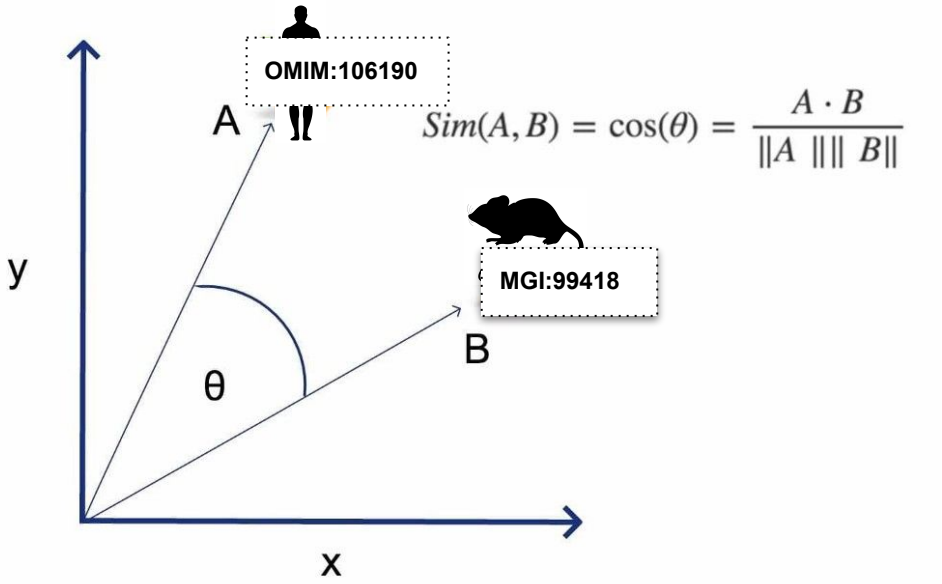
D_1	d_{11}	d_{12}	...	d_{1k}
D_2	d_{21}	d_{22}	...	d_{2k}
D_m	d_{m1}	d_{m2}	...	d_{mk}

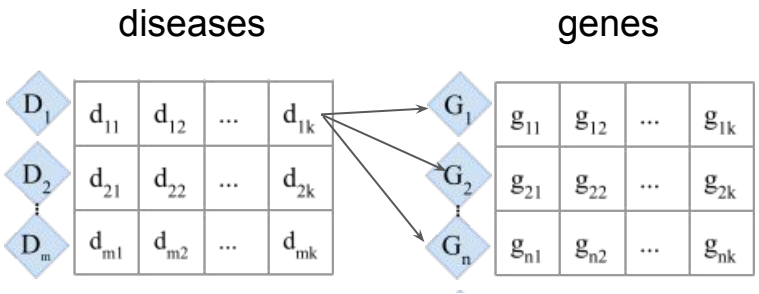
genes

G_1	g_{11}	g_{12}	...	g_{1k}
G_2	g_{21}	g_{22}	...	g_{2k}
G_n	g_{n1}	g_{n2}	...	g_{nk}

Calculating Phenotypic Similarity Approaches

- Unsupervised Approach
  - Cosine similarity

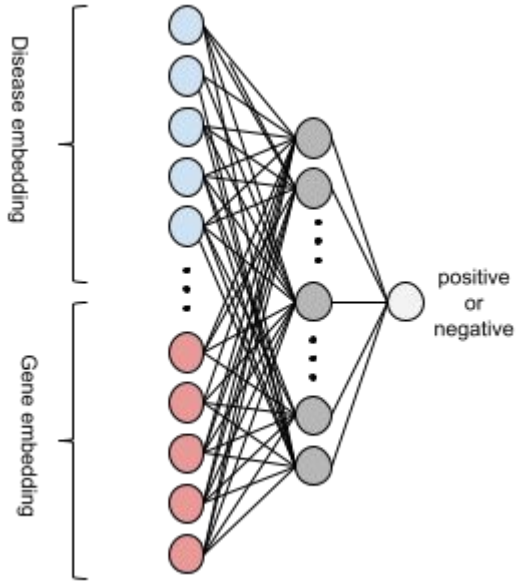




Predict associated genes to disease D_1:
$\max_{G \in genes}(Sim(D_1,G))$

Calculating Phenotypic Similarity Approaches

- Unsupervised Approach
  - Cosine similarity
- Supervised Approach
  - MLP



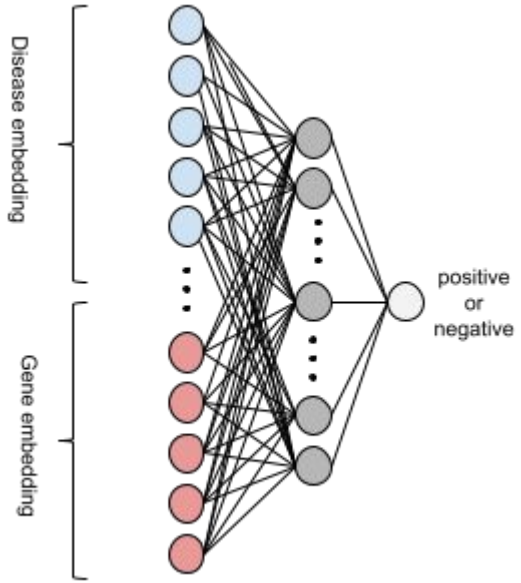
Calculating Phenotypic Similarity Approaches

- Unsupervised Approach
  - Cosine similarity

- Supervised Approach
  - MLP
    - Train/Test split

      10-fold cross validation

      Stratified by disease

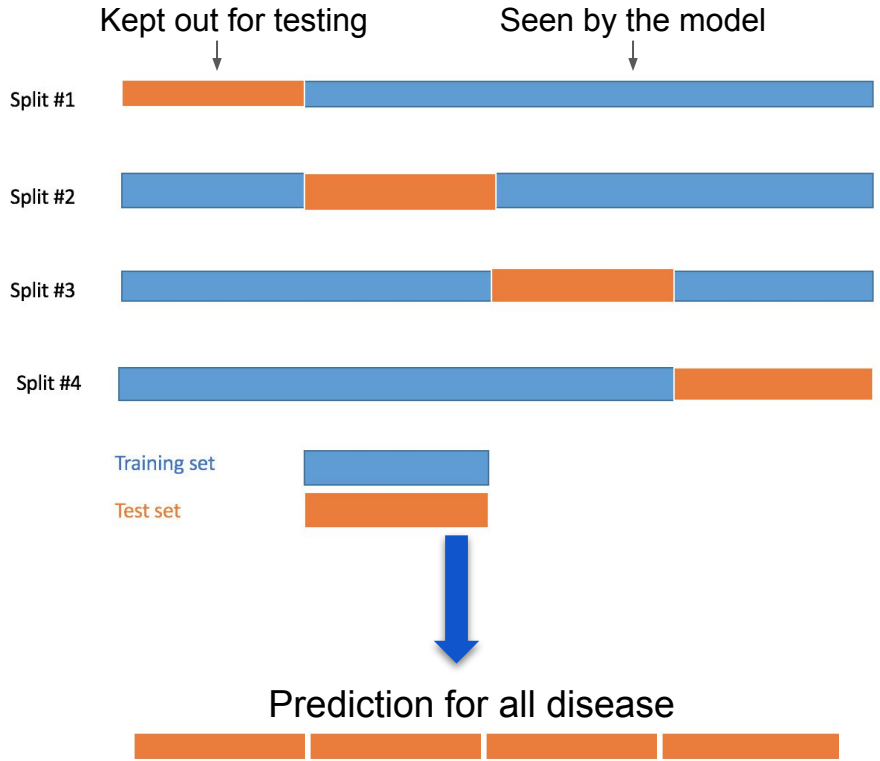


Calculating Phenotypic Similarity Approaches

- Unsupervised Approach
  - Cosine similarity
- Supervised Approach
  - MLP
    - Train/Test split

      10-fold cross validation

      Stratified by disease



Calculating Phenotypic Similarity Approaches

- Unsupervised Approach
  - Cosine similarity

- Supervised Approach
  - MLP
    - Train/Test split

      10-fold cross validation

      Stratified by disease

    - Positives/Negatives

Positives/Negatives imbalance …

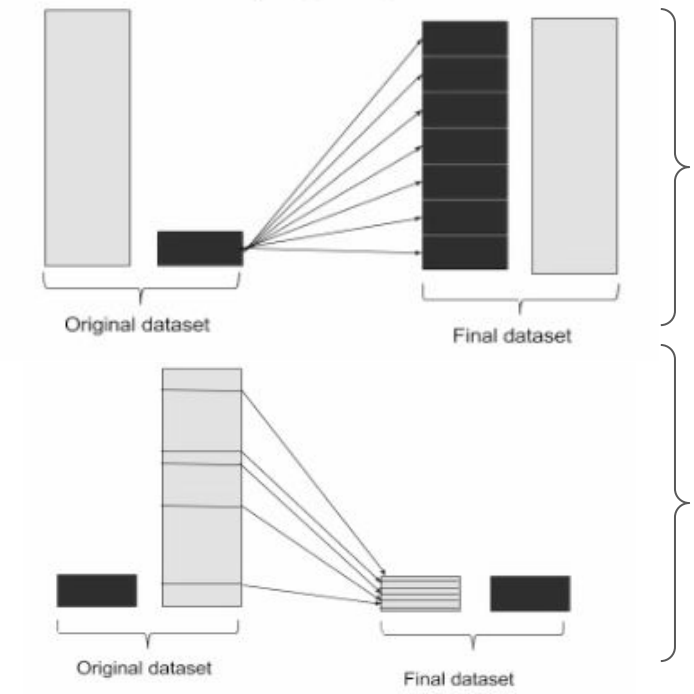
Calculating Phenotypic Similarity Approaches

- Unsupervised Approach
  - Cosine similarity
- Supervised Approach
  - MLP
    - Train/Test split

      10-fold cross validation

      Stratified by disease
    - Positives/Negatives

Positives/Negatives imbalance …



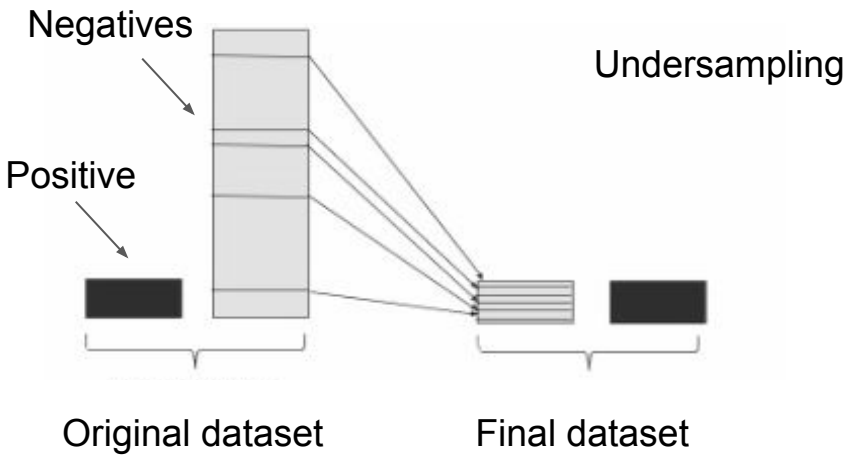
Calculating Phenotypic Similarity Approaches

- Unsupervised Approach
  - Cosine similarity
- Supervised Approach
  - MLP
    - Train/Test split

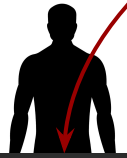
      10-fold cross validation

      Stratified by disease
    - Positives/Negatives

Positives/Negatives imbalance …



Example:



Phenotypically similar

Isolated anhidrosis with normal morphology and number sweat glands (ANHD)

Itpr2
inositol 1,4,5-triphosphate receptor 2

ITPR2
inositol 1,4,5-trisphosphate receptor type 2

Orthologous

Human disease phenotypes:

Generalized anhidrosis

Heat intolerance

Anhidrosis

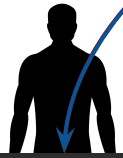
Mouse gene phenotypes:

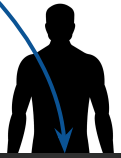
abnormal sweat gland physiology

Hypohidrosis

Human gene

Example:





Associated

Phenotypically similar

Orthologous

Isolated anhidrosis with normal morphology and number sweat glands (ANHD)

Itpr2
inositol 1,4,5-triphosphate receptor 2

ITPR2
inositol 1,4,5-trisphosphate receptor type 2

Human disease phenotypes:

Generalized anhidrosis

Heat intolerance

Anhidrosis

Mouse gene phenotypes:

abnormal sweat gland physiology

Hypohidrosis

Human gene

Hands On tutorial ..