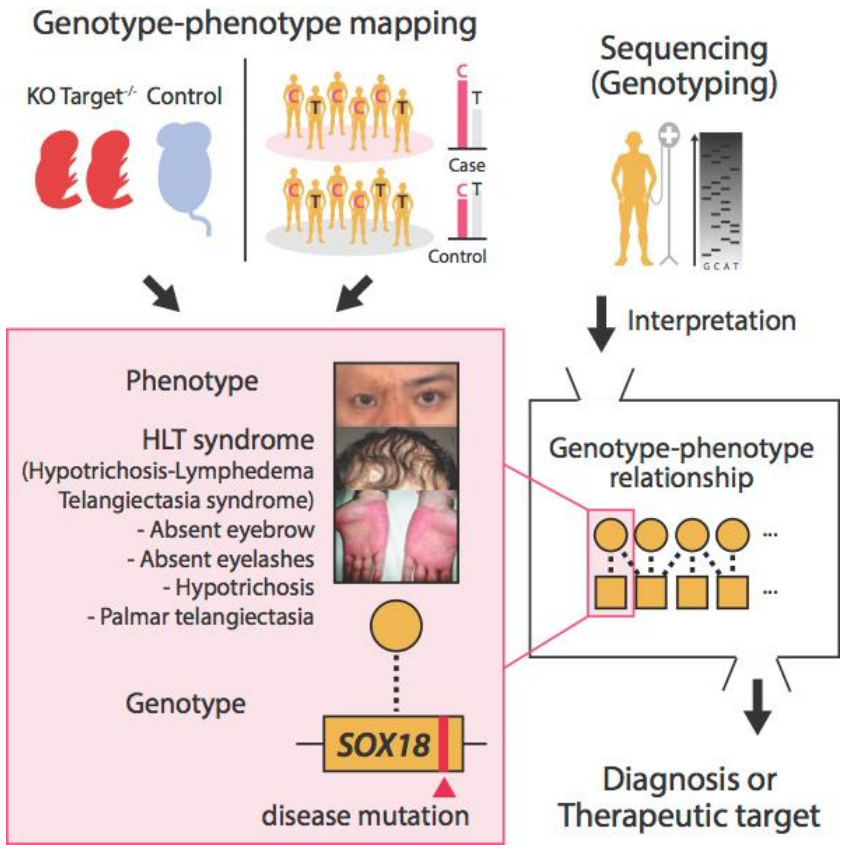
# A systematic gene-phenotype comparison between human and mouse to find accurate phenologs

Sanguk Kim, Ph.D.

Structural Bioinformatics Laboratory

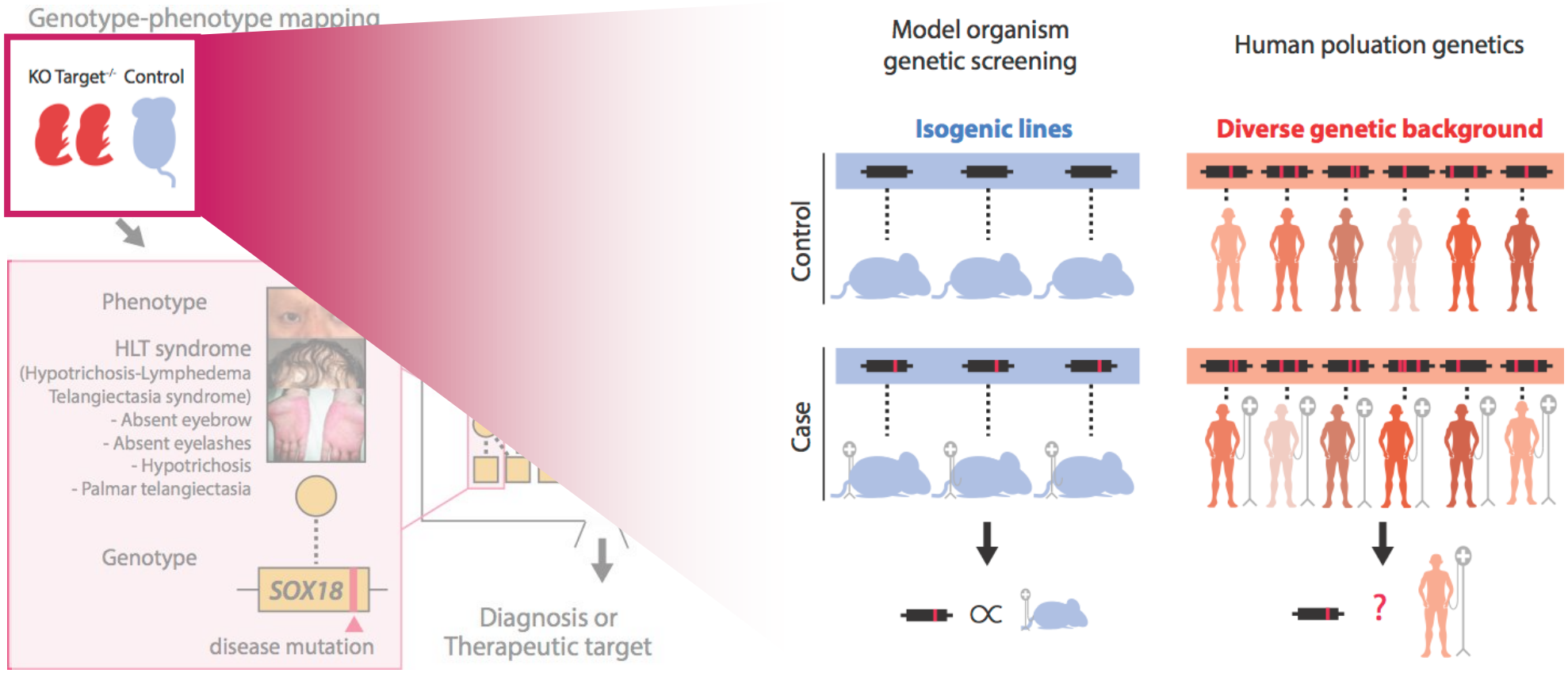
Department of Life Sciences, POSTECH

# Human genotype-phenotype relationship for disease studies.




- From personal-genome sequencing of patients, screening disease-associated mutations is helpful for prevention of diseases.
- Targeted therapy could be served from the development of drugs that correct the malfunction of genes carrying the mutations.

# Model organism: a crucial tool for understanding human genotype-phenotype relationship.

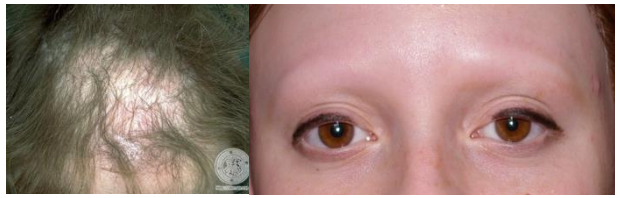



- With the advent of genome sequencing technique,
  genome-wide association studies on patient cohorts have become crucial for human disease research.

- Model organisms play a complementary role to the human population studies,
  because genetic background in model organisms can be controlled through the breeding of isogenic lines.

## Human phenotyping relying on model organism has been possibly due to ‘orthology-function conjecture’.

HLT syndrome

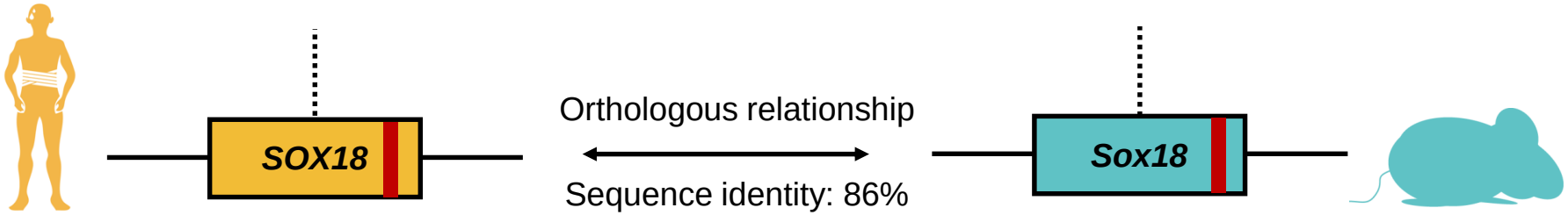


- **Absent eyebrow / eyelashes**
- **Hypotrichosis**
- **Palmar telangiectasia**
- Abnormality of the nail

≈

Mouse phenotypes

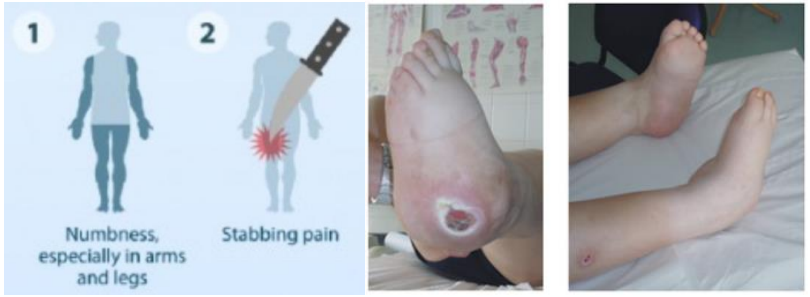
- **Alopecia**
- **Hairless / Sparse hair**
- **Abnormal vibrissae morphology**
- Cyanosis




**Orthology-function conjecture:**
Perform similar function between orthologous genes

# Phenotypic differences which occurred in the orthologous genes: Great challenges of human disease study using mouse genetic approaches

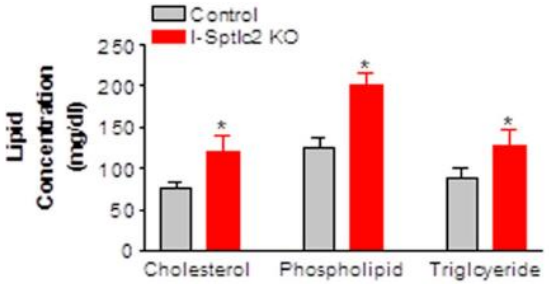
Hereditary sensory and autonomic neuropathy



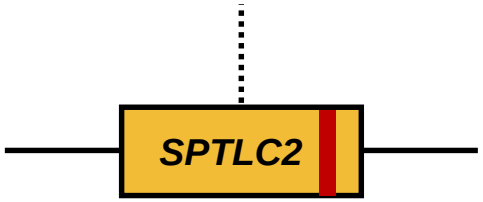

- Sensorimotor neuropathy
- Distal sensory impairment
- Abnormality of foot

≠

Mouse phenotypes

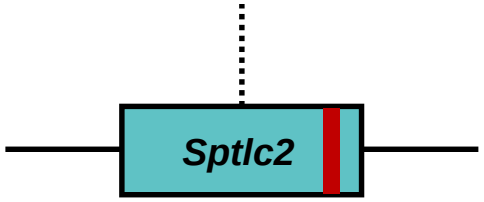



- Abnormal lipid level
- Abnormal liver physiology
- Abnormal cholesterol/sphingomyelin level




Orthologous relationship

Sequence identity: 96%

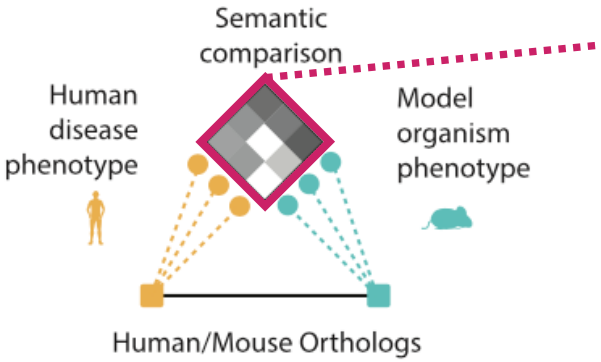



## Section I.

# Analysis of molecular evolutionary events accounting for phenotypic differences of human and mouse orthologous genes

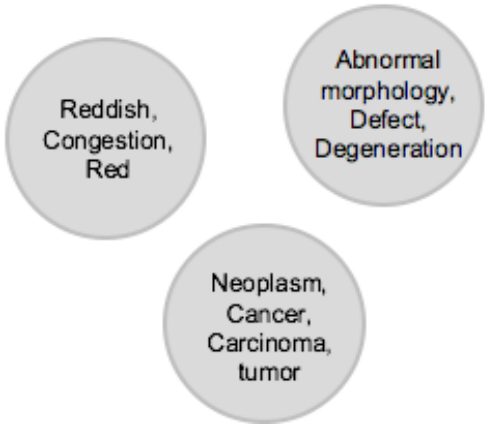
Han SK, Kim D, Lee H, Kim I, Kim S.
Divergence of non-coding regulatory elements explains gene-phenotype differences between human and mouse orthologous genes
*(Molecular Biology and Evolution, 2018)*

# PS score: a statistical framework for a systematic quantification of the phenotypic differences occurred in human and mouse orthologous genes

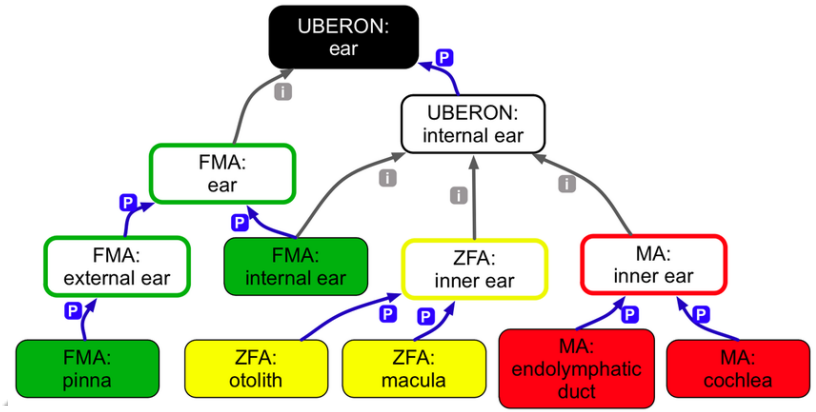


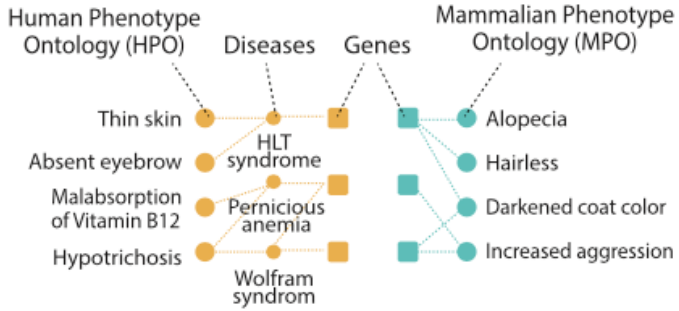

**Calculation of semantic similarity**
Cross-species comparison between HPO and MPOs

Phenotype quality (synonym)




Phenotype entity (anatomical homology)

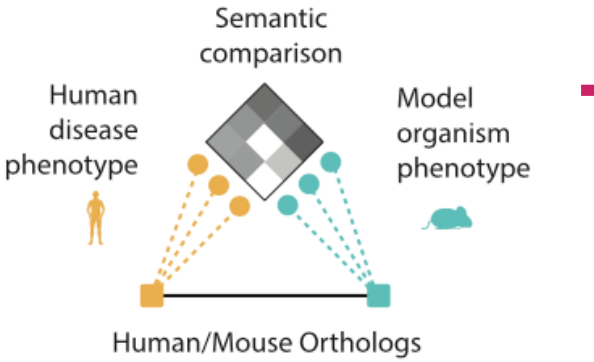


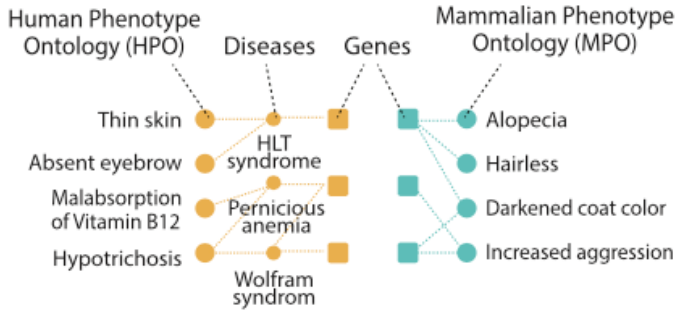

*Washington et al., PLoS Biol* **7(11)**, e1000247 (2009)

# PS score: a statistical framework for a systematic quantification of the phenotypic differences occurred in human and mouse orthologous genes

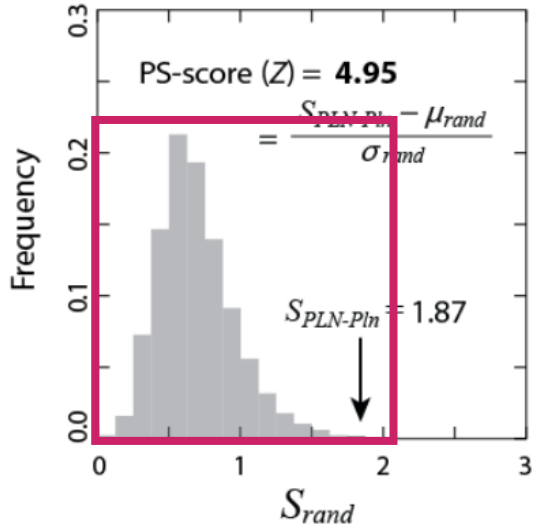
PS-score assesses statistical significance of the phenotypic similarity based on the comparisons of the semantic-similarity scores of orthologous gene-pairs to those of random gene-pairs.



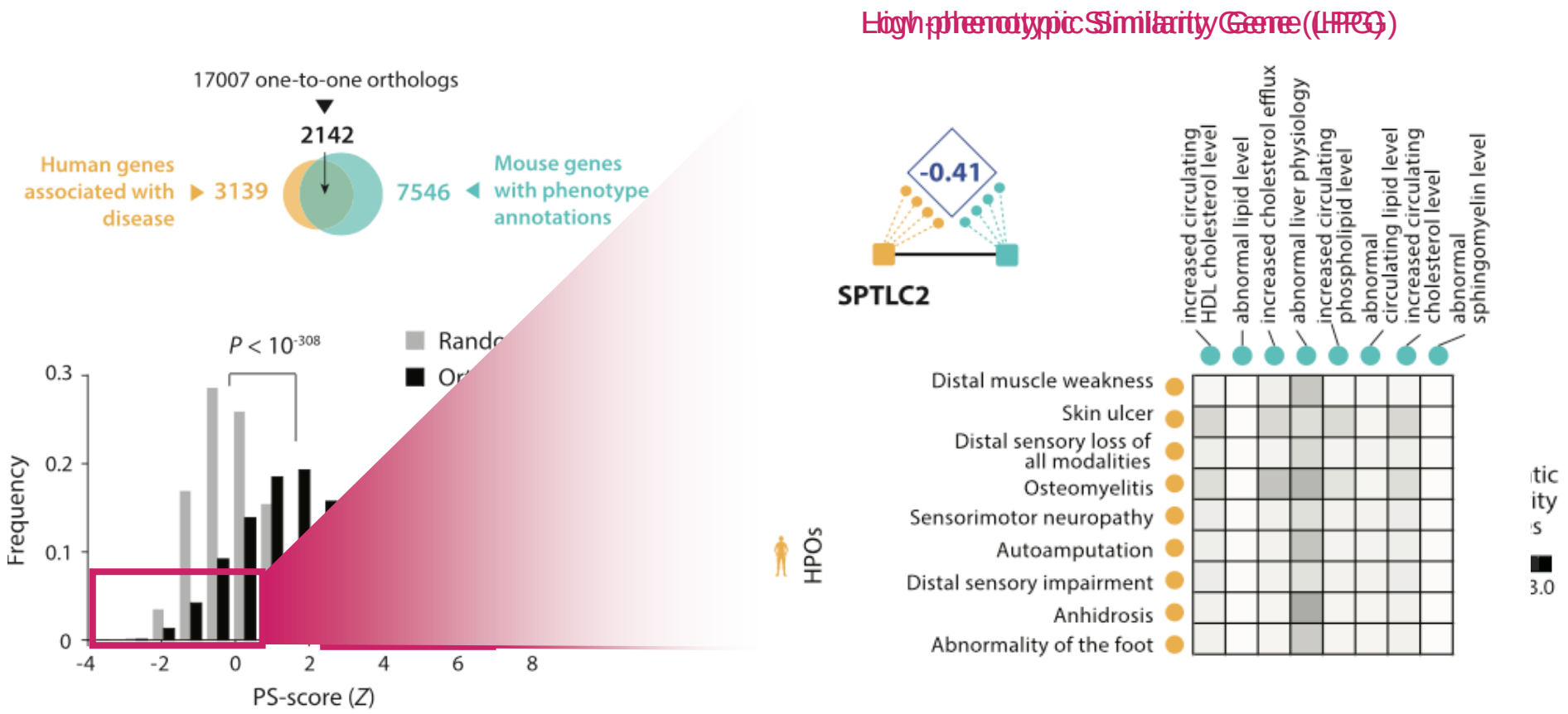

Assessment of statistical significance on the semantic similarity



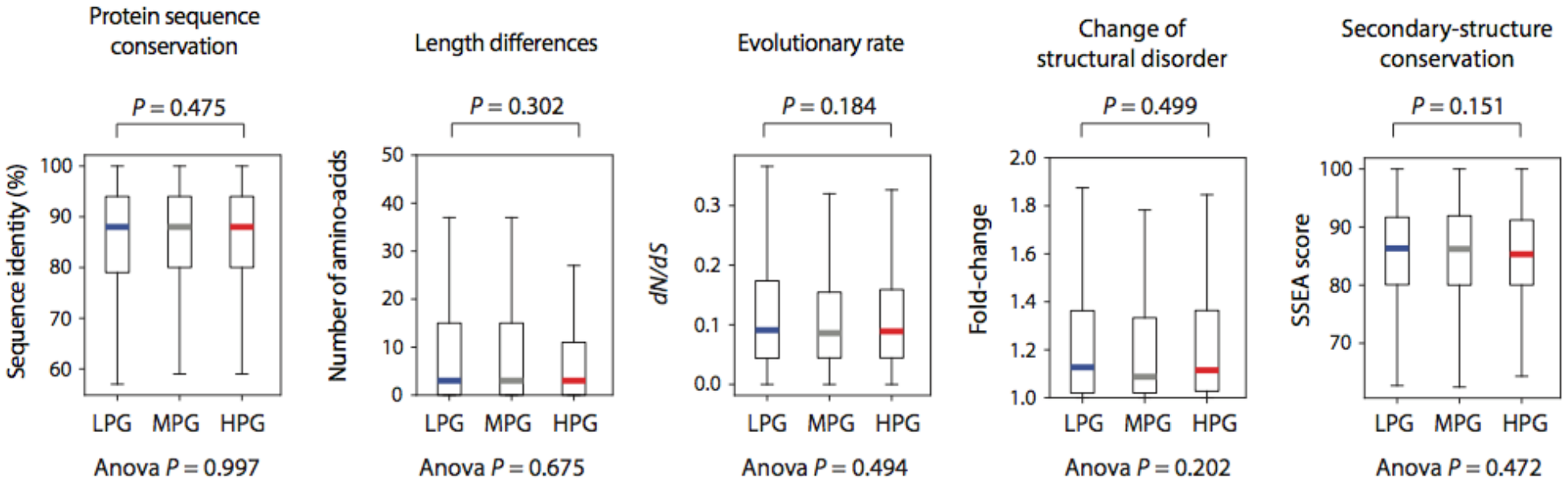


# Orthologous genes cover a wide range of phenotype similarity



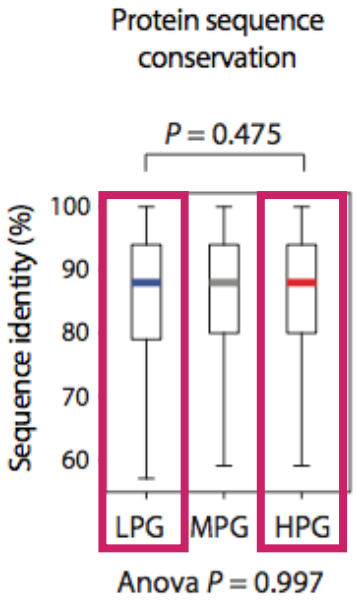

Orthologous genes turn out to have significantly higher PS-scores than random gene-pairs which might be easily explained by orthology-function conjecture.

# Sequence divergence may not explain the phenotypic differences of orthologous genes.



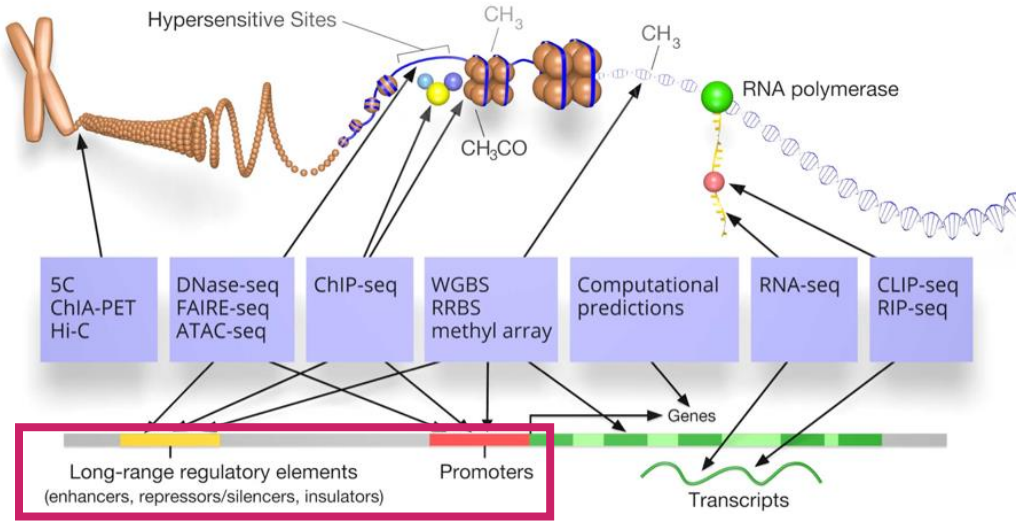

- One might expect that sequence divergence would affect the phenotypic differences of the LPGs.
- Orthologous genes usually have high sequence similarity but genetic drift between species triggered the divergence of coding region, changing the phenotypes between the species.

# Sequence divergence may not explain the phenotypic differences of orthologous genes.



# Non-coding regulatory evolution: a source of phenotypic evolution

ENCODE Encyclopedia of DNA Elements

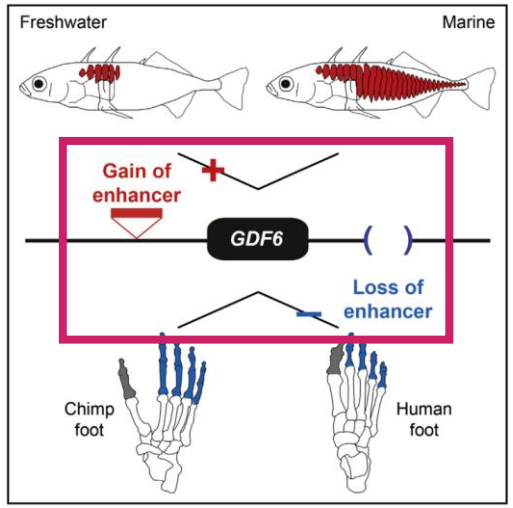



*Yue et al., Nature* **515(7527)**, 355-64 (2014)

Article

**Cell**

**Evolving New Skeletal Traits by *cis*-Regulatory Changes in Bone Morphogenetic Proteins**



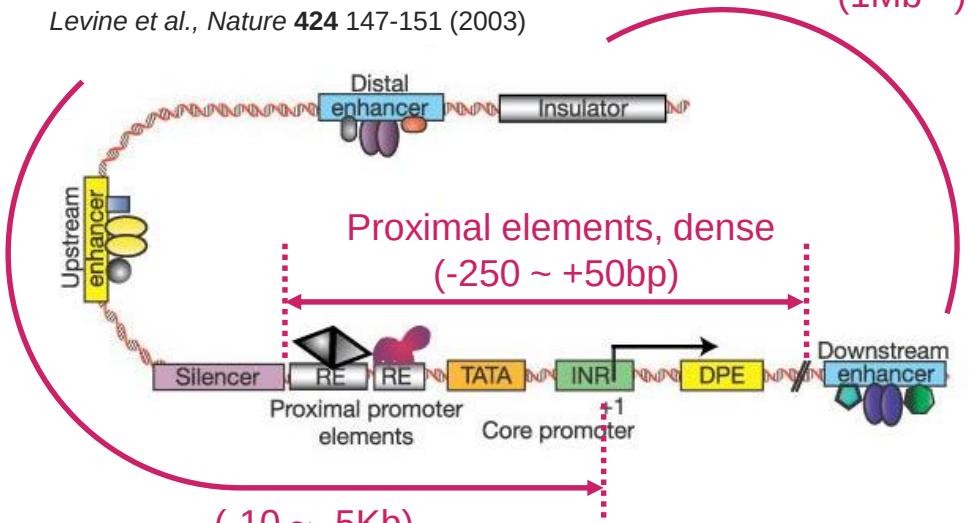

*Indjeian et al., Cell* **164(1-2)**, 45-56 (2016)

Phenotypic differences may be the result of changes in non-coding regulatory sequences, which are frequently observed within mammalian species.

# Divergence of non-coding regulatory sequences is correlated with phenotypic differences.

Intergenic,
Very distal elements
(1Mb ~)

*Levine et al., Nature* **424** 147-151 (2003)




(-10 ~ -5Kb)

Distal upstream
elements, sparse

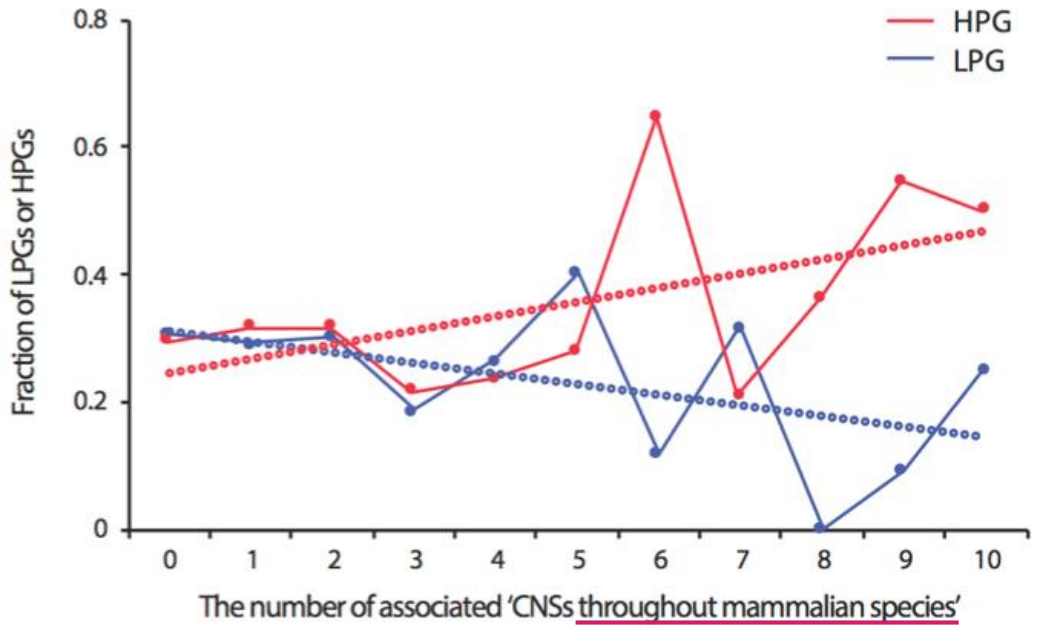
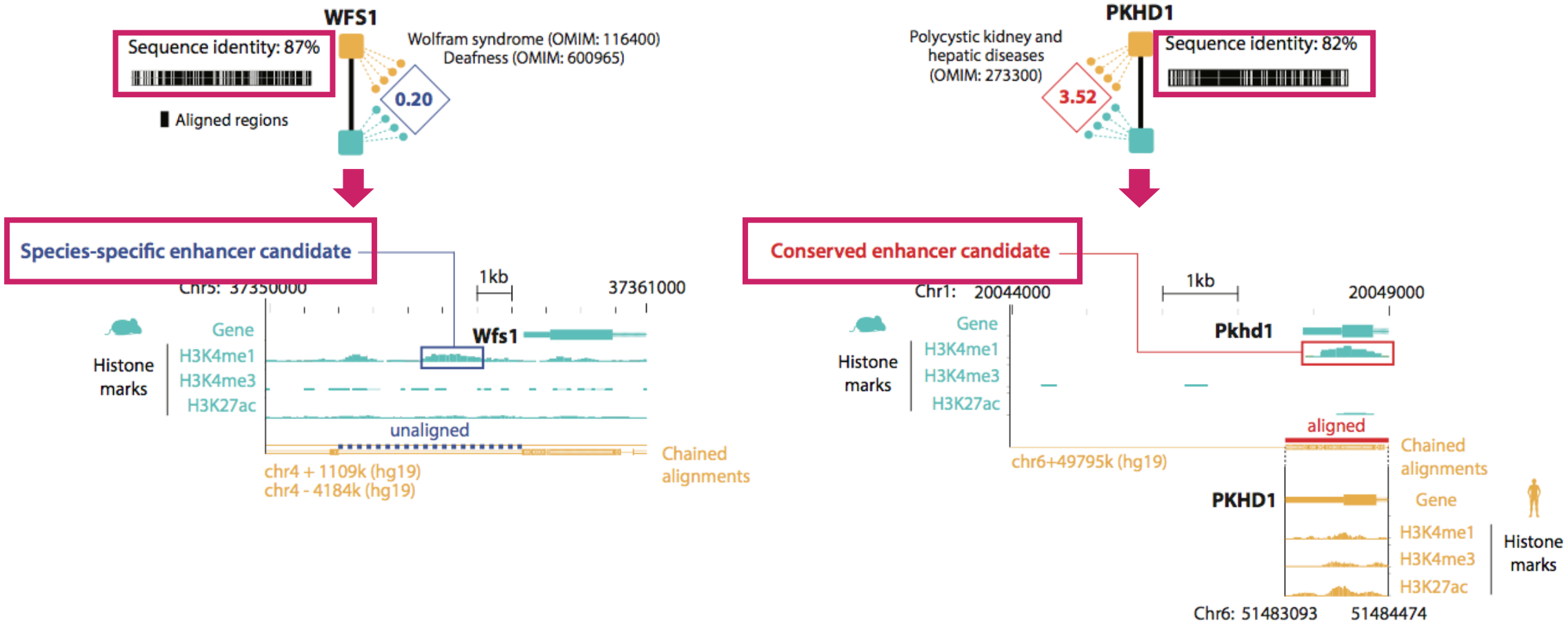



Table 1. Conservation of *cis*-Regulatory Elements in LPGs

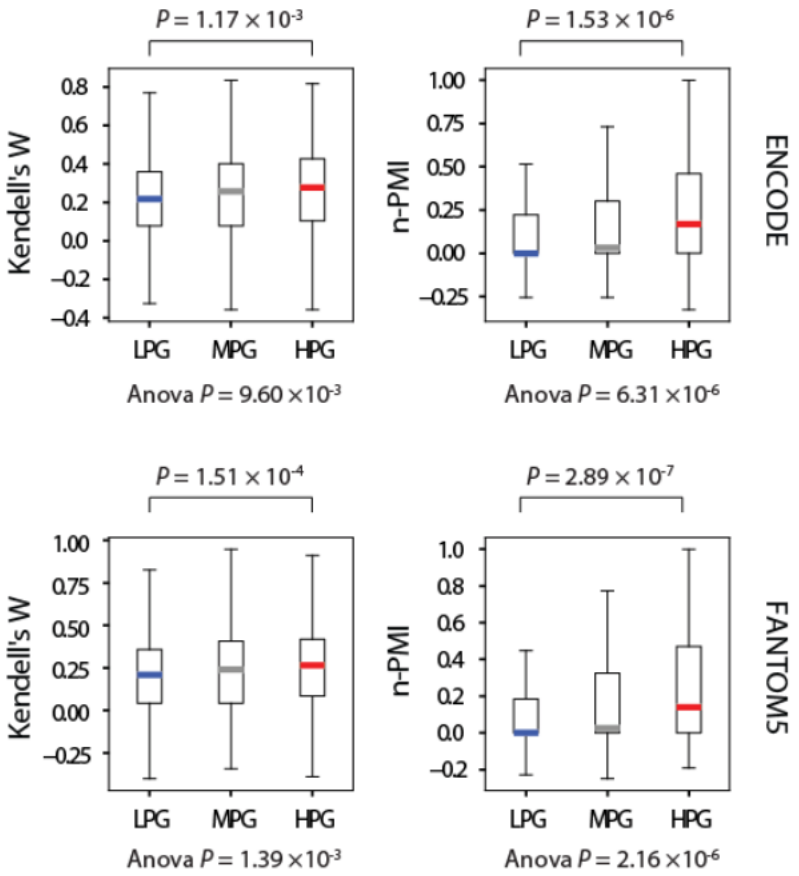
| Functional elements | | LPG | | HPG | |
|---|---|---|---|---|---|
| Enhancers | Species-specific | 28.4% | | | |
| | Conserved | 40.6% | (370/912) | 49.6% | (413/833) |
| TF-binding sites | Species-specific | 17.5% | (200/1, 143) | 10.3% | (173/1, 828) |
| | Conserved | 66.2% | (757/1, 143) | 76.2% | (1, 436/1, 828) |
| Promoters | Species-specific | 4.4% | (27/621) | 2.4% | (13/535) |
| | Conserved | 81.2% | (504/621) | 86.9% | (465/535) |

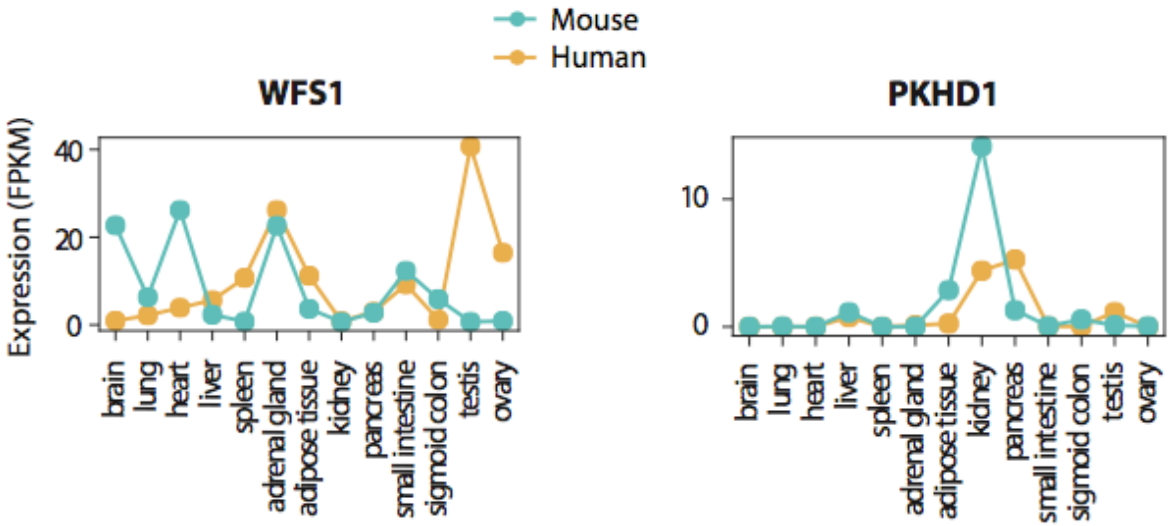
# Divergence of non-coding regulatory sequences is correlated with phenotypic differences.




Species-specific regulatory elements could explain the existence of orthologous genes with very low PS scores, which could not be accounted for by the divergence of coding regions.

# Transcriptomic divergence also explains phenotypic differences between orthologous genes.

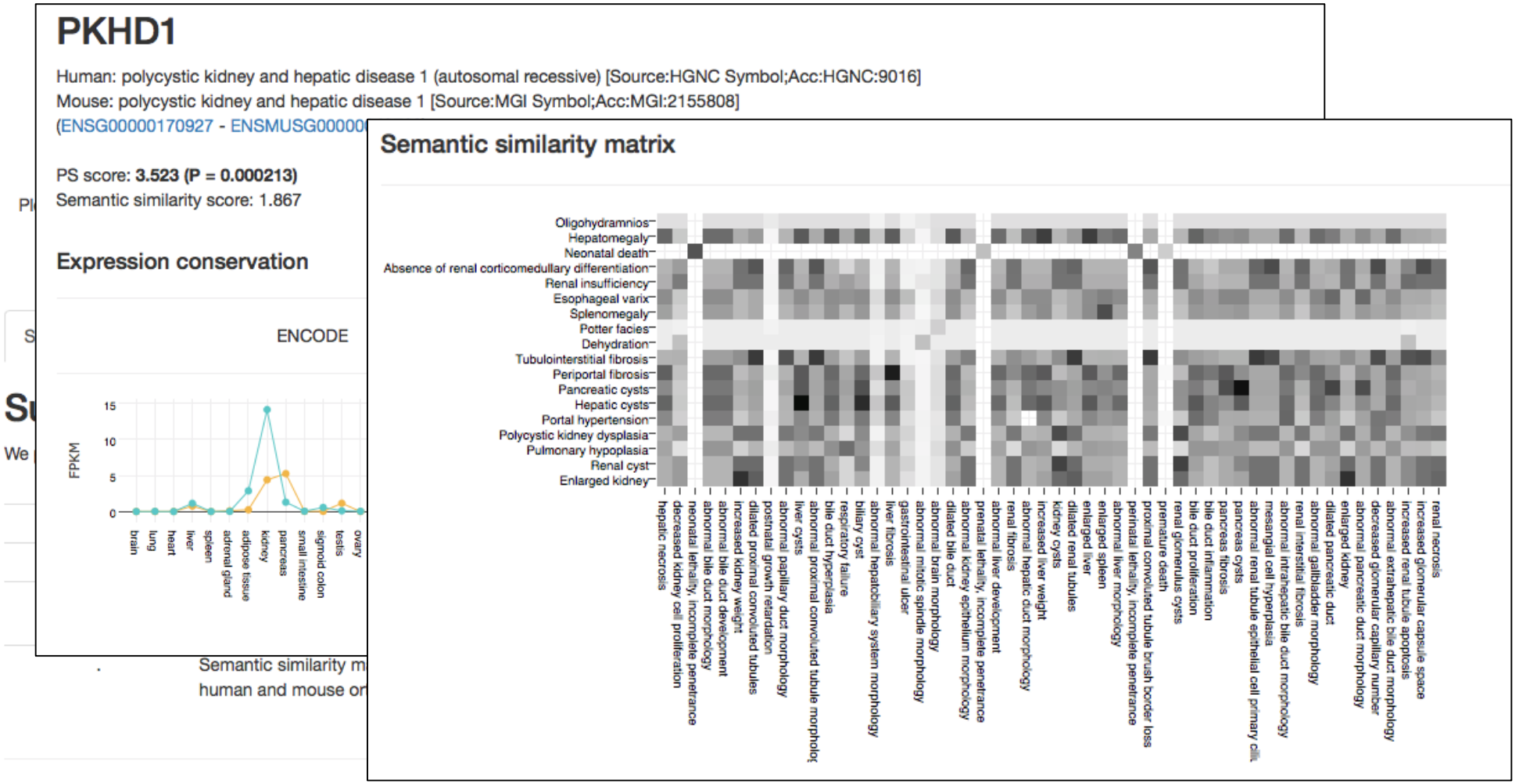



Changes in gene regulatory sequences could trigger divergence in transcription across species that might repurpose functionally conserved proteins in different cell types or tissues, and consequently change the phenotypes.

# Our companion website for investigating human disease phenotypes through mouse genetic approaches. (https://sbi.postech.ac.kr/w/PS)

- We provide tissue-specific expression conservation in all human and mouse orthologous genes using the FANTOM and ENCODE databases.
- Orthologous genes with high expression conservation are likely to be useful for identifying putative phenologs; orthologous genes with phenotypes that are identical across species.



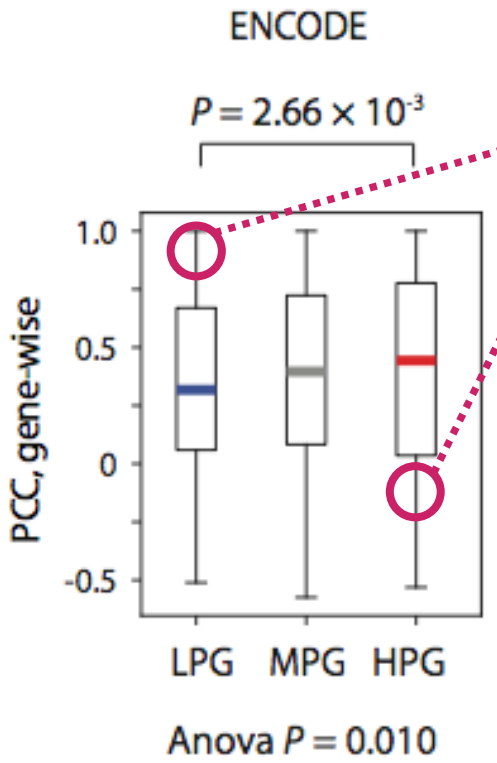

Structural Bioinformatics Lab.

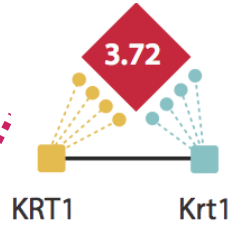
## Section II.

# Expression divergence of neighbors impacts on phenotypic differences of human and mouse orthologous genes

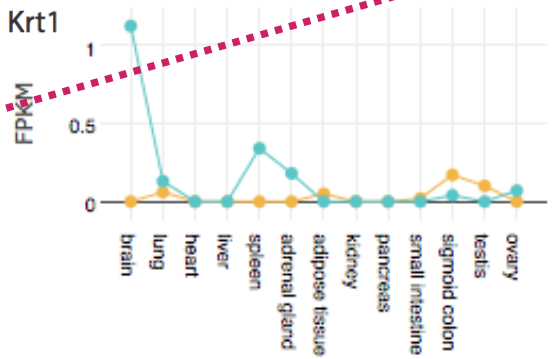
*(Ongoing Project 1)*

# Limitation on the explanation of phenotypic differences with ’single-gene’ expression divergence.



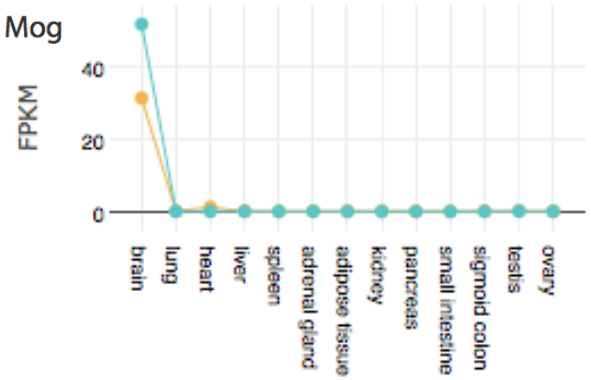




Expression Conservation
(PCC) = -0.22

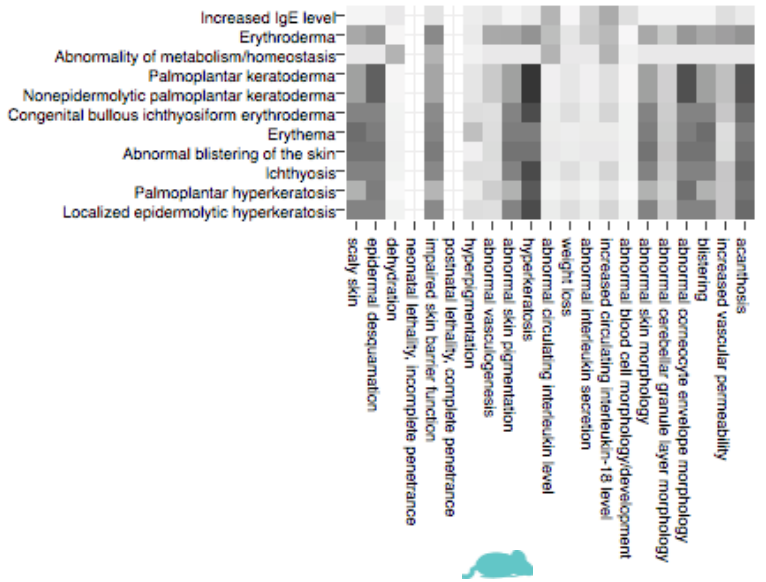




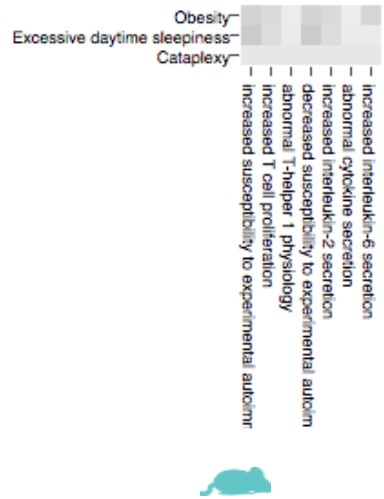
Expression Conservation
(PCC) = 1.00



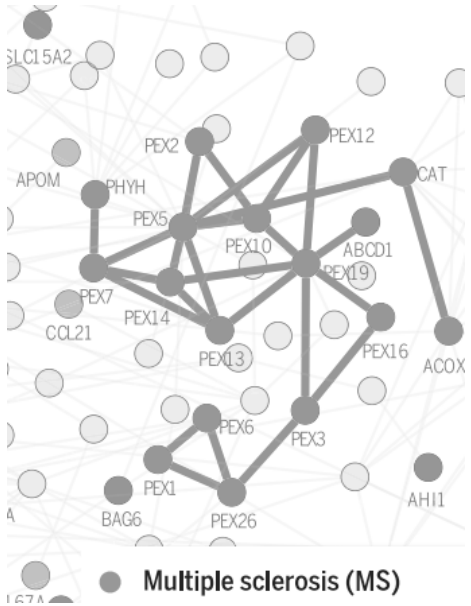

Human and mouse phenotypes



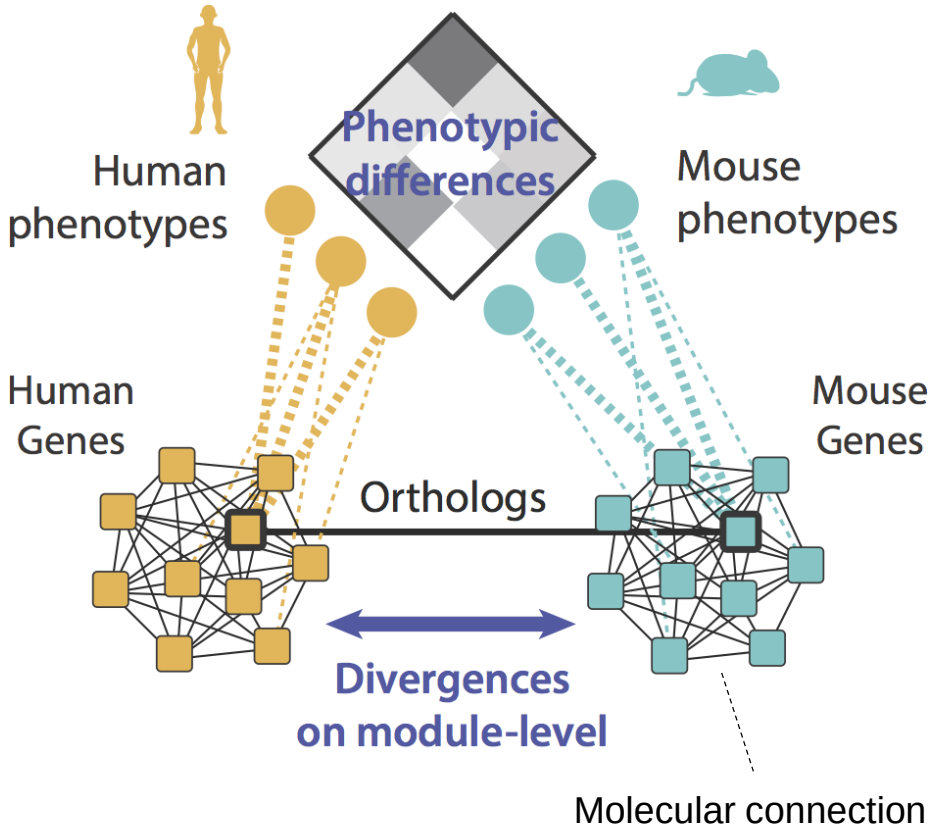

Human and mouse phenotypes



# Molecular evolution of genetic modules impacts on phenotypic evolution

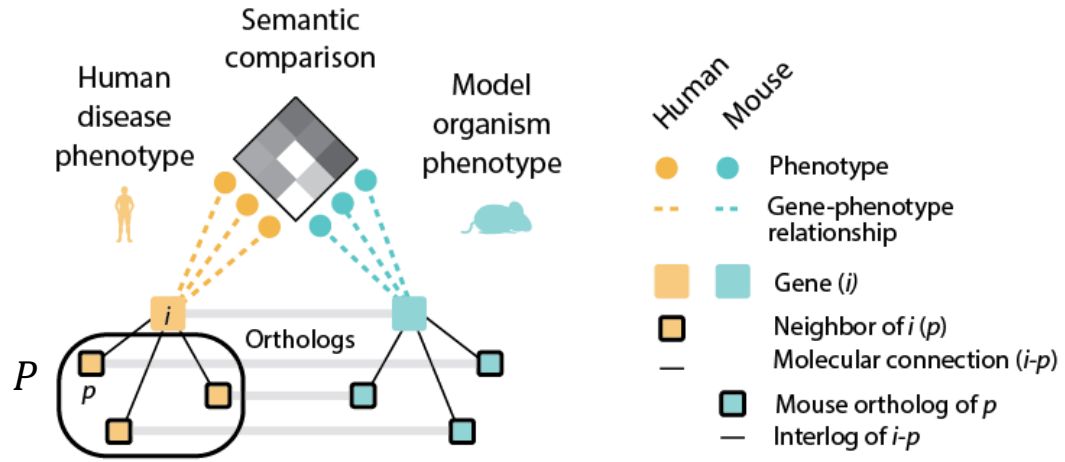



*Menche et al., Science* **347(6224)**, 1257601 (2015)




- Phenotypes are not caused by single gene but affected by group of genes in a module.
- The evolutionary divergence of the gene networks underlying may eventually affect to the phenotypic differences.

# Expression divergence of PPI network partners correlates with phenotypic differences of orthologous genes.

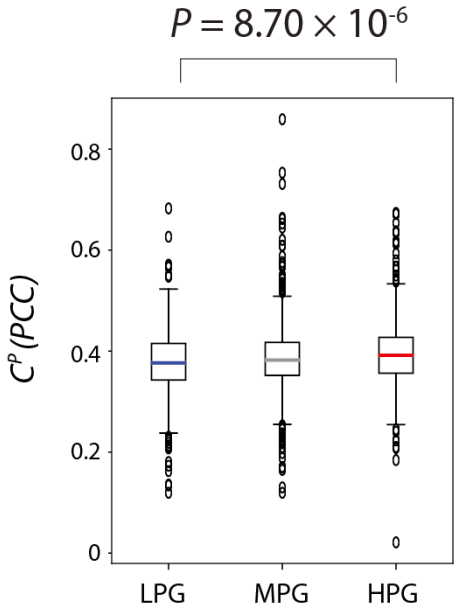


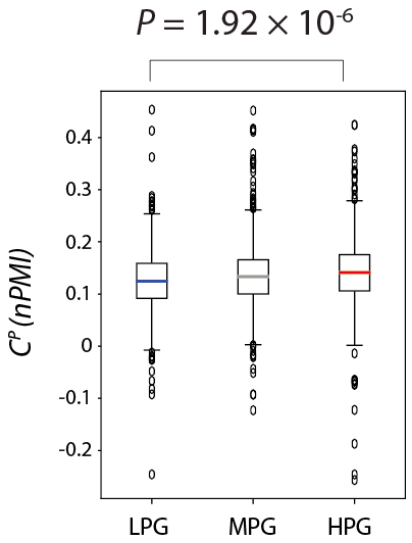

Expression conservation of Single-gene ($c$)

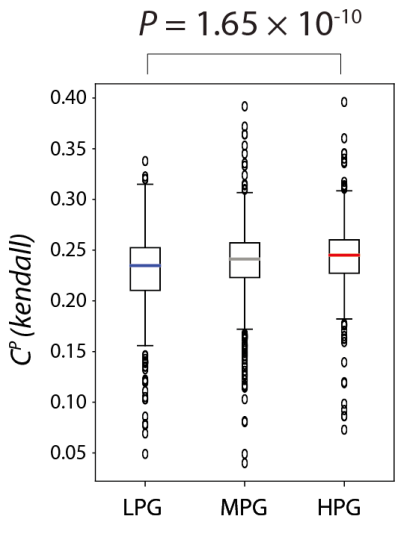
$$c_i = \begin{cases} \text{Pearson's R} \\ \text{Kendall's W} \\ \text{normalized PMI} \end{cases}$$

Expression conservation of Partners ($c^p$)

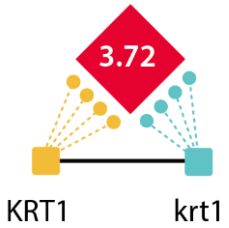
$$c_i^P = \frac{\sum_{p \in P} c_p}{|P|} \quad (i \notin P)$$



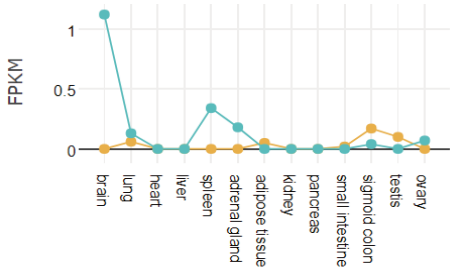



# Molecular evolution of genetic modules impacts on phenotypic changes of orthologous genes.




Expression conservation of single gene
*c* (PCC) = -0.22



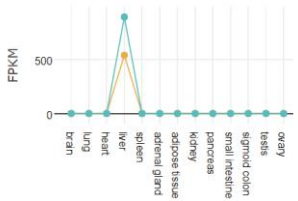

Functional module :

- Negative regulation of coagulation
- Regulation of water loss via skin
- Fibrinolysis

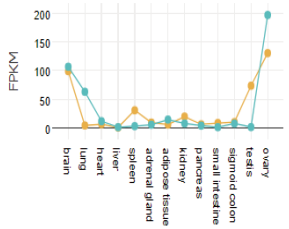
Neighbors in module:
HS3ST5, TSPAN8, F11, PLAUR, …, F2, PROC, PLAT

**Expression conservation of module partners**
***c*$^p$ (PCC) = 0.50**

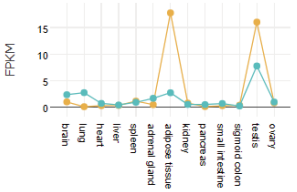
F2, c(PCC) = 1.00



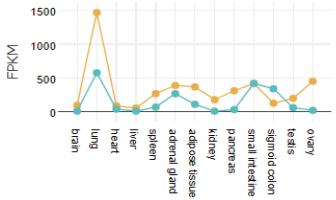

SERPINE2, c(PCC=0.81)




CYP26B1, c(PCC) = 0.75

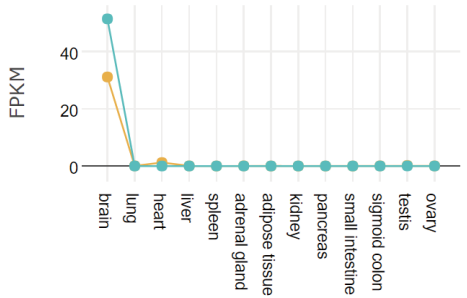



ANXA2, c(PCC) = 0.73)





Expression conservation of single gene
*c* (PCC) = 1.00



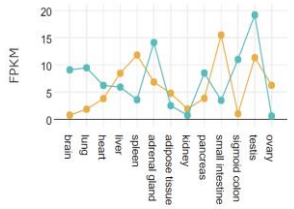

Functional module :

- Response to folic acid

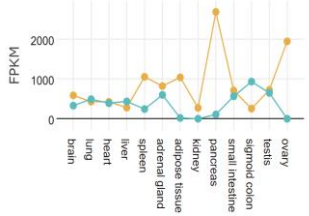
Neighbors in module:
TIMP3, MTHFR, ASCL1, TYMS, BCHE, EEF2, MGMT, OGG1, NRF1, FOLR1, GSN

**Expression conservation of module partners**
***c*$^p$ (PCC) = 0.14**

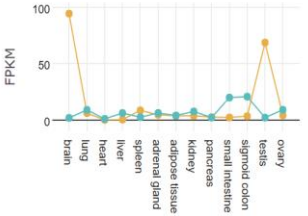
MTHFR, c(PCC) = 0.01



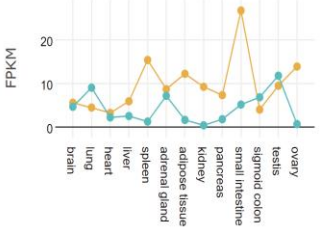

EEF2, c(PCC=-0.51)




TYMS, c(PCC) = -0.35




OGG1, c(PCC) = -0.15)



# Acknowledgements


- Seong Kyu Han
- Donghyo Kim
- Doyeon Ha
- Inhae Kim
- Heetak Lee
- Jungho Kong
- Kwanghwan Lee

**Structural Bioinformatics Lab @POSTECH**

**https://sbi.postech.ac.kr**