

Integrative Machine Learning Algorithm for Comparative Cross-species Analysis of Protostomes



State of the art

Technological advancements in next-generation sequencing have opened up new possibilities to study molecular mechanisms across numerous organisms. Machine learning algorithms have bolstered these analyses, but they're often limited by the need for large data quantities, scarce in less-studied organisms. A recently developed Species-Agnostic-Transfer-Learning (SATL), aims to address these limitations. It facilitates the transfer of knowledge across various mammalian species to another, bypassing the incomplete gene orthology constraints. Additionally, work on explainable AI has identified vital features of machine learning models that drive prediction, lending insights into the underpinnings of target phenotypes.

Figure 1 SATL species transfer figure
See figure SATL

Primary question

Can we transfer knowledge on molecular mechanisms across different species using advanced machine learning and comprehensive OMICs data from public databases and generated GönomiX?

Objectives

Develop and tailor cross-insecta transfer learning (CITL) for prediction of Wnt signaling related cell types and stages
Extend and validate the CITL approach for insects with more divergence time like fly and beetle
Adapting and applying CITL towards distant invertebrates using the GönomiX data

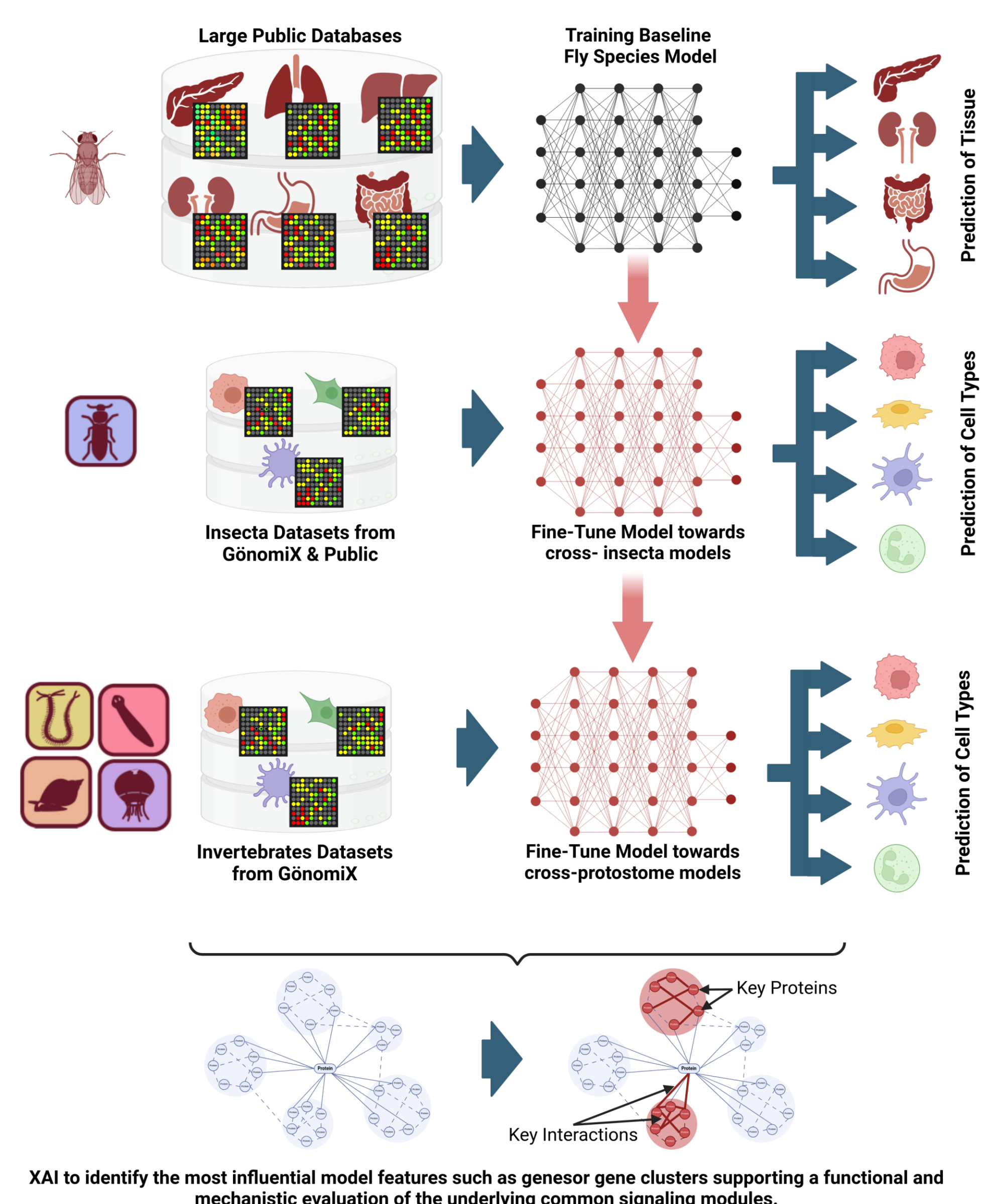
Workplan

A) Developing and validating cross-insecta transfer learning for prediction of Wnt signaling related cell types and stages:

first build a fly model on available public multi-omics data (such as GEO, FlyAtlas) and transfer the model to data available Within the Gönomix consortium. Subsequently, to extend to insects with more divergence time (e.g. BeetleAtlas and iBeetle-Base). Build and validate cross species models will be trained to predict matching cell types and stages of these target species.

B) Adapting and applying the cross-species algorithm towards distant invertebrates

Employ the Cross-Insecta transfer learning algorithm to the data generated within GönomiX to refine and adapt the Insect-SATL models developed in A. We focus on Multi-Omics from various species studied within the Gönomix consortium with increasing phylogenetic distance. Develop and employ explainable artificial intelligence methods to identify the most influential model features such as genes or gene clusters supporting a functional and mechanistic evaluation of the underlying common signaling modules.



Synergy and collaborations

Collaborative Project 3: Novel bioinformatics and genetic tools

Development of machine learning methods based on the generated data to be used by the other groups.

Collaborative Project 2: Reconstructing evolving GRNs

Analyzing diverse omics data generated across all projects and clades (RNA-seq, ATAC-seq, chromatin structure, single-cell RNA-seq).

Transfer of knowledge across Drosophila species and between fly and beetle. Ultimately, these models will be trained on GönomiX data to predict matching cell types and stages of these target species across clades within the groups in the consortium).

Interaction will be essential with the biologist postdoctoral researcher in order to test and interpret the results and adapt the algorithms.

Technical innovation

Development of Advanced Cross-Species Transfer Learning and adaptation towards protostoma
Integrating Cross-Species Model for gene regulatory element prediction
Advance study of WNT signaling pathways

Specific qualification

Bioinformatics for High-Throughput Omics Data Analysis and Systems Biology
Transfer Learning and Explainable AI
Designing and Deploying Tools and Workflows



Anne-Christin
Hauschild
Medical Center
University

References

Fu* J, Posnien* N, Bolognesi R, Fischer TD, Rayl P, Oberhofer G, Kitzmann P, Brown* SJ, Bucher* G. Asymmetrically expressed axin required for anterior development in Tribolium. Proc Natl Acad Sci U S A 2012
Ansari S, Troelenberg N, Dao VA, Richter T, Bucher* G, Klingler* M. Double abdomen in a short-germ insect: Zygotic control of axis formation revealed in the beetle Tribolium castaneum. Proc Natl Acad Sci 2018

