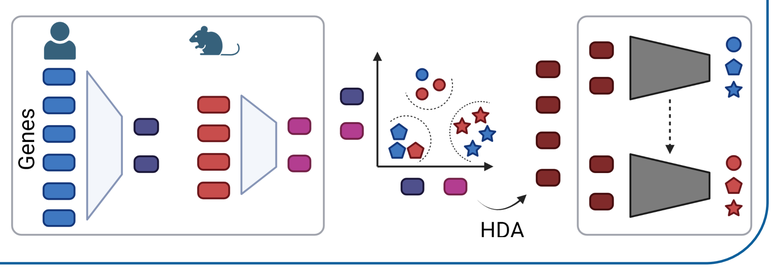
P9

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

GönomiX

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 [1]. It facilitates the transfer of knowledge across various mamalian 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 under-pinnings of target phenotypes [2].



Primary question

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

Objectives

- Develop and taylor 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

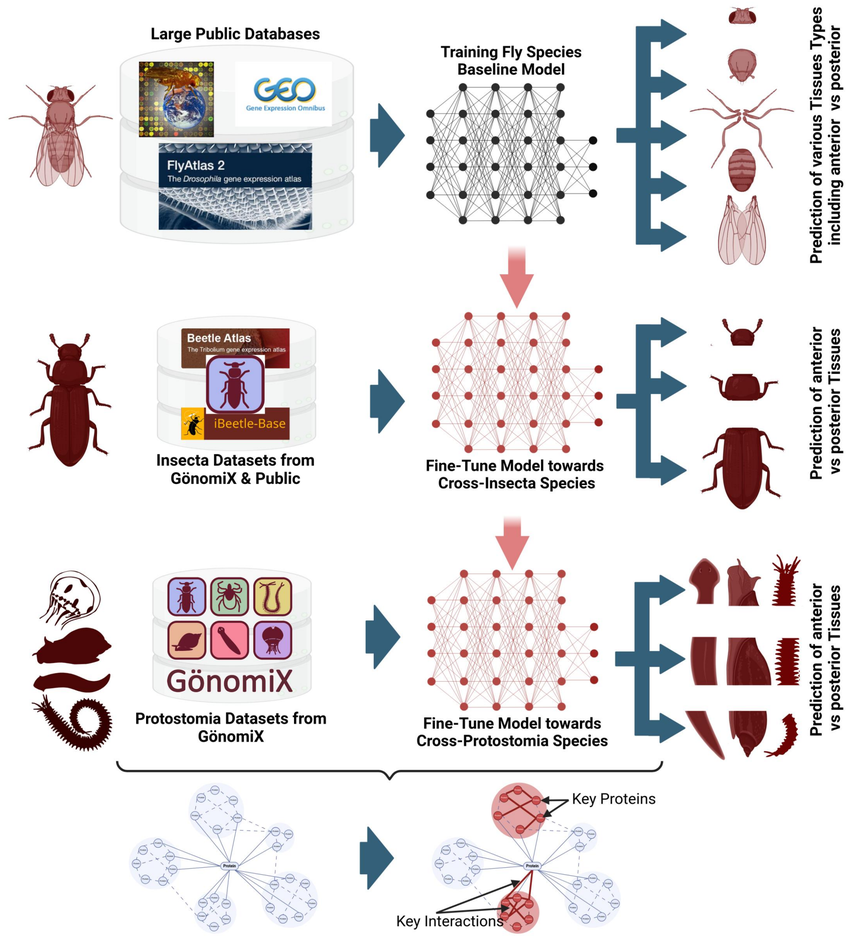
Workplan

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

First build a baseline model on available public multi-omics data in fly (such as GEO, FlyAtlas) and transfer the model to data available within the Gönomix consortium. Subsequently, we will extend towards insects with more divergence time (e.g. BeetleAtlas and iBeetle-Base). Therefore, cross species models will be trained and validated 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 signalling modules.



XAI to identify the most influential model features such as genesor 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 **all** groups in the consortium.
- Essential interaction with the students and postdoctoral researchers in order to test and interpret the results and adapt the algorithms.

Technical innovation

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

Specific qualification

- Bioinformatics and Machine Learning and Transfer Learning for High-Throughput Omics Data Analysis
- Interpretable and Explainable AI for Systems Biology Investigation
- Designing Interactive Visualization Tools and Workflows

Anne-Christin Hauschild
University Medical Center

References

[1] Park Y, Muttray NP, Hauschild AC Species-Agnostic Transfer Learning for Cross-species Transcriptomics Data Integration without Gene Orthology Briefings in Bioinformatics, 25(2), 2024. bbae004

[2] Beinecke J, Saranti A, Angerschmid A, Pfeifer B, Klemt V, Holzinger A, Hauschild AC CLARUS: An Interactive Explainable AI Platform for Manual Counterfactuals in Graph Neural Networks Journal of Biomedical Informatics. 2024 Jan 30:104600.

GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN IN PUBLICA COMMODA SEIT 1737

MAX PLANCK INSTITUTE FOR MULTIDISCIPLINARY SCIENCES

UNIVERSITÄTSMEDIZIN GÖTTINGEN UMG Göttingen Campus