

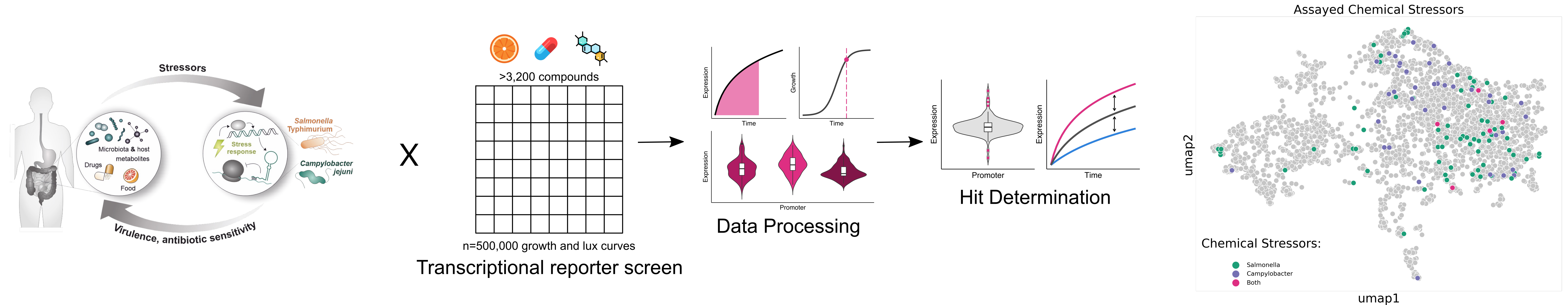
StressRegNet: Molecular representations decipher stress response in gut pathogens

Roberto Olayo-Alarcon^{1,2}, Martin K. Amstalden³, Christoph Binsfeld³, Susanne Brenzinger³, Mona Alzheimer M⁴, Annamaria Zannoni⁴, Fabian Koenig⁴, Cynthia Sharma⁴, Ana Rita Brochado³, Mina Rezaei¹, Christian L. Müller^{1,2,5}

¹Department of Statistics, Ludwig-Maximilians-Universität München, Munich, ²Institute of Computational Biology, Helmholtz Munich, Munich, ³Department of Molecular Infection Biology II, Institute of Molecular Infection Biology (IMB), Julius-Maximilians-Universität Würzburg, Würzburg, ⁴Department of Microbiology, Julius-Maximilians-Universität Würzburg, Biocenter, Würzburg, ⁵Center for Computational Mathematics, Flatiron Institute, New York

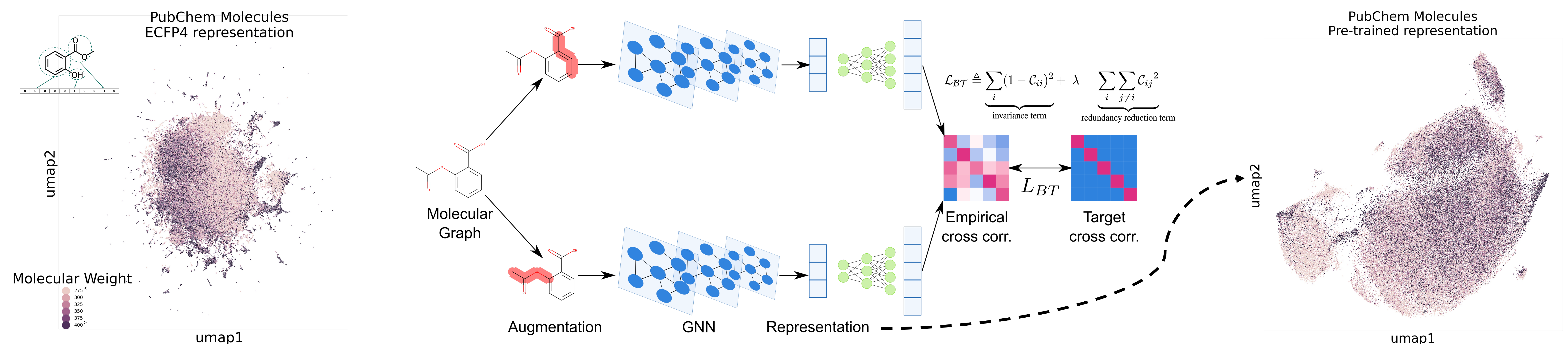
A diverse set of molecules can induce a transcriptional response

We screen a large chemical library against transcriptional reporters in gut pathogens. We find that structurally diverse chemicals induce a transcriptional response in gut pathogens. It becomes an interesting task to relate chemical structure to the observed stress response.



We leverage large databases to create a novel molecular representation

By adapting the Barlow-Twins pre-training strategy, we are able to use the structural information of various molecules available in PubChem to create a pre-trained chemical representation that is more informative than commonly used ECFP4 fingerprints.



Our pre-trained representations can aid various computational tasks

Our pre-trained representation can improve the performance of predictive algorithms in various tasks of biological relevance. Furthermore, the representation can be fine-tuned to certain outcomes. With this, we can begin to clarify how structurally diverse compounds can induce similar responses in our StressRegNet data compendium.

