Identifying and imputing gene programs across-age, tissues and inflammatory diseases using machine learning

Supervisor: Prof Muzlifah Haniffa (mh32@sanger.ac.uk)

Department/Institute: Wellcome Sanger Institute

Co-supervisor: Dr Mohammad Lotfollahi (Wellcome Sanger Institute)

Research area: Machine learning and human health

Project outline:

The Human Developmental Cell Atlas (HDCA) initiative aims to construct a comprehensive reference map of cells during development essential to understand normal organogenesis, the effect of mutations, environmental factors, and infectious agents on human development, congenital and childhood disorders, and the cellular basis of aging, cancer, and regenerative medicine. Yet, mapping cell states from the developmental stage to adult atlases remains difficult due to state-plasticity and unique gene programs across different tissues and ages. Here, we aim to construct a meta single-cell atlas cross-age, developmental, and tissues atlas using interpretable machine-learning models (Lotfollahi et al. 2022) . The prospective student will leverage deep neural networks to identify conserved or novel gene programs across different conditions to discover and link cell states and gene programs from our comprehensive developmental datasets (Suo, Dann et al., Science 2022 and unpublished whole human embryo). By extending the data to map age-related changes (paediatric skin cell atlas, paediatric thymus data)   and disease states (e.g. unpublished large cohort eczema data before and after treatment with duplimuab, JAK1 inhibitor and methotrexate), we aim to reveal gene programs across different ages extracted from healthy data and those altered in disease conditions to identify and reveal disease-associated programs. Next, we will extend the model to enable counterfactual predictions to in silico impute the molecular information for the unstudied population of cells in disease conditions (Lotfollahi et al. 2021) to further guide data generation and experimental design.

The project's interdisciplinary nature, encompassing machine learning and biology, will provide the student with a unique training environment.

Reference

Lotfollahi, Mohammad, Sergei Rybakov, Karin Hrovatin, Soroor Hediyeh-zadeh, Carlos Talavera-López, Alexander V. Misharin, and Fabian J. Theis. 2022. "Biologically Informed Deep Learning to Infer Gene Program Activity in Single Cells." bioRxiv. https://doi.org/10.1101/2022.02.05.479217.

Lotfollahi, Mohammad, Anna Klimovskaia Susmelj, Carlo De Donno, Yuge Ji, Ignacio L. Ibarra, F. Alexander Wolf, Nafissa Yakubova, Fabian J. Theis, and David Lopez-Paz. 2021. "Compositional Perturbation Autoencoder for Single-Cell Response Modeling." bioRxiv. https://doi.org/10.1101/2021.04.14.439903.

Suo C*, Dann E*, Goh I, Jardine L, Kleshchevnikov V, Park JE, Botting RA, Stephenson E, Engelbert J, Tuong ZK, Polanski K, Yayon N, Xu C, Suchanek O, Elmentaite R, Domínguez Conde C, He P, Pritchard S, Miah M, Moldovan C, Steemers AS, Mazin P, Prete M, Horsfall D, Marioni JC, Clatworthy MR†, Haniffa M†,Teichmann SA†. Mapping the developing human immune system across organs. Science. 2022 Jun 3;376(6597):eabo0510. doi: 10.1126/science.abo0510.

BBSRC DTP main strategic theme: Bioscience for an integrated understanding of health

BBSRC DTP secondary strategic theme: Understanding the rules of life, Transformative technologies