7th Central German Meeting on Bioinfomatics - Mittelerde Meeting (26-27.3.26)

Aligning host dependency factors of Influenza A virus with gene regulatory network dynamics in infected scRNA-seq data using a context-aware deep learning model

Alicia Hiemisch[1], Fionn Keogh[1], Andre Dietz[1], Rainer König[1]

[1]Systemsbiology Research Group, Institute for Infectious Diseases and Infection Control, Jena University Hospital, Jena, Germany

Influenza A virus (IAV) has long been a major threat to human health and demonstrated its pandemic potential through several outbreaks in recent decades. Besides vaccines and directly acting antivirals, host-directed antivirals are promising therapeutic options. Traditionally, targets of host-directed antivirals are genes being essential for the viral life cycle, known as host dependency factors (HDFs). While CRISPR/Cas9 or RNA interference screens have been employed in numerous studies to identify relevant HDFs, achieving robust and consistent identification poses a challenge. Moreover, viral pathogenesis is not solely driven by viral replication but also by the propagation of infection-associated signals through host gene regulatory network. The identification of core regulatory elements within these networks may therefore offer disease-modifying intervention points. However, inference of gene regulatory networks typically requires large-scale datasets, which are rarely available in the context of viral infection studies.

In the present work, we used Geneformer, a context-aware, attention-based deep learning model pretrained on approximately 30 million human single-cell transcriptomes, to model host regulatory responses during IAV infection. We fine-tuned Geneformer to classify infected versus non-infected human primary cells using a publicly available single-cell RNA-sequencing dataset of 30,000 normal human bronchial epithelial cells infected with IAV (A/California/04/2009, MOI 0.5) or mock treated. Following fine-tuning, we performed in silico gene perturbation to identify genes whose deletion in infected cells shifts cellular embeddings toward a non-infected state.

The fine-tuned model achieved an accuracy of 90%, successfully distinguishing between infected and non-infected cells, based solely on host gene expression. In silico perturbation analysis identified 16 candidate regulatory genes whose deletion induced a significant shift in the embedding space towards non-infected cells compared to random perturbations. Notably, these candidates showed significant enrichment for previously reported IAV host dependency factors compiled from 15 independent screening studies.

Together, these preliminary results suggest that transfer learning–based modelling of gene regulatory context may provide a complementary strategy for prioritizing host-directed antiviral targets, even in settings with limited infection-specific data.