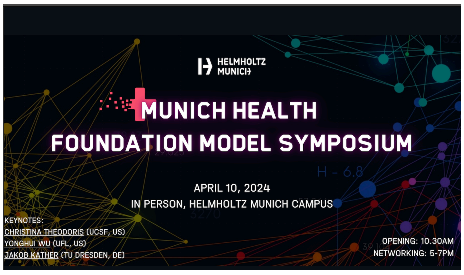
HELMHOLTZ MUNICH
MUNICH HEALTH FOUNDATION MODEL SYMPOSIUM
APRIL 10, 2024
IN PERSON, HELMHOLTZ MUNICH CAMPUS
KEYNOTES:
CHRISTINA THEODORIS (UCSF, US)
YONGHUI WU (UFL, US)
JAKOB KATHER (TU DRESDEN, DE)
OPENING: 10.30AM
NETWORKING: 5-7PM

Contribution ID: **23** Type: **Poster**

Understanding the epithelial-to-mesenchymal transition in cancer with a single-cell large language model

The epithelial-to-mesenchymal transition (EMT) is pivotal in tumour progression and resistance to treatment, yet its heterogeneity complicates the precise assessment of EMT status of individual tumour cells in different cancer types. Furthermore, while key epithelial and mesenchymal genes driving the transformation are well characterised, other regulators, especially at intermediate stages of the process, are less well understood.

In this study, we employ RNA-seq data from single cells profiled at 0 hours, 8 hours, 1 day, 3 days and 7 days during EMT transformation from xx et al, and leverage a pre-trained single-cell language model (scLLM) to develop a generalisable classifier of EMT status in single cell cancer data. Our method, scMultiNet, demonstrates an average prediction accuracy of EMT state of 90% AUROC across various cancers. scMultiNet incorporates a simple yet efficient multiplication mechanism and widely considered Parameter-Efficient Fine-Tuning (PEFT) strategy, offering an effective way to adapt the self-supervised pre-trained language model for specific EMT processes. Our approach enables the model to achieve good performance even with limited training data. We further propose a Attention-Driven Expression Significance Index (ADESI), which considers both attention scores from our model and the original gene expression values, to uncover genes that are critical in regulating the entire timeline of EMT transformation. The top regulators uncovered include genes involved in mitochondrial function (e.g., NDUFB10, MRPL51) and oxidative stress response (e.g., PRDX1) suggesting a metabolic reprogramming during EMT. Other genes such as TUBA1B and TUBB, which form microtubules crucial for cell shape and transport during migration, have not been specifically linked with EMT previously. Finally, we employ the derived gene signatures to explore the association of distinct EMT states with survival outcomes and disease recurrence in the METABRIC dataset and find that that patients exhibiting the 8h and 3d EMT signatures, as identified by genes with high attention scores in these categories, showed a notable decrease in survival rates. .

In conclusion, scMultiNet exemplifies the effective application of language models in cancer biology research, offering a novel approach to EMT status prediction and identifying clinically relevant gene signatures reflecting the plasticity of the EMT programme.

Primary authors: Dr SECRIER, Maria (university college london); PAN, Shi (University College London)

Session Classification: Break + Posters session

Track Classification: Poster