

# Geneformer for cell exposure prediction

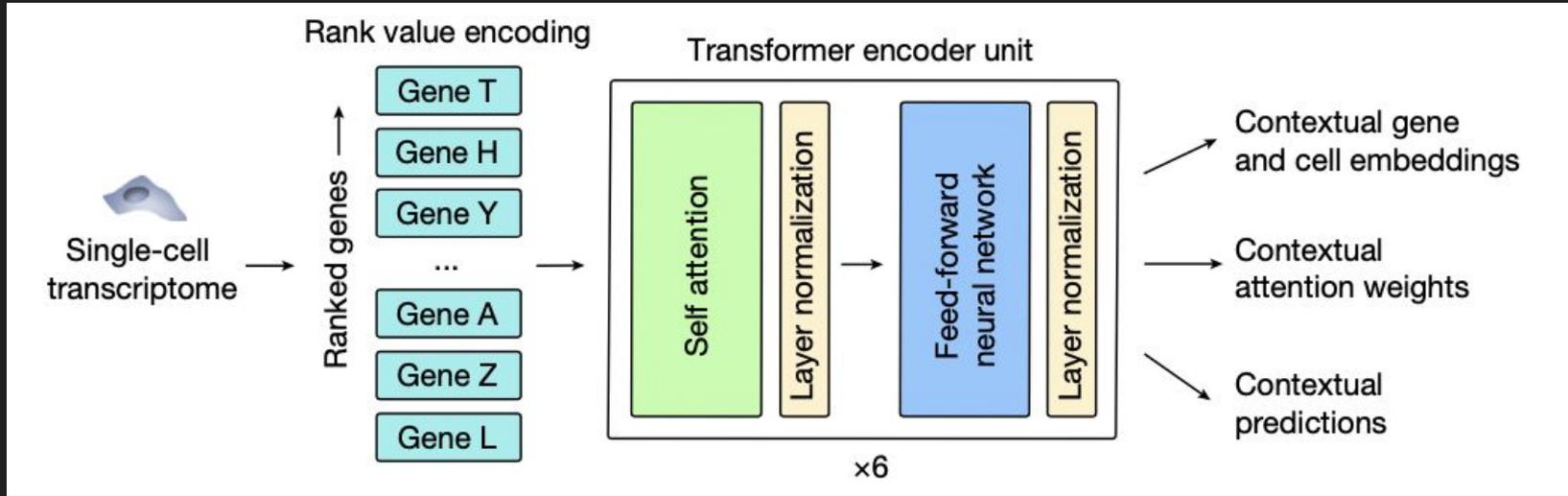
## Overarching goal

Fine-tune a pre-trained transformer model architecture to predict the exposure of a cell to a particular hormonal pathway or an EDC, leveraging on scRNA-seq data

# Structure

- The Geneformer architecture
- Motivation and purpose
- Results
- Limitations and inspiration for further research

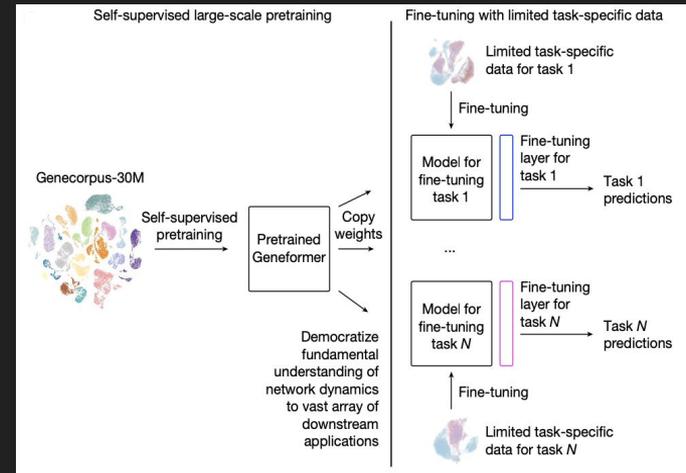
# The Geneformer architecture



# Motivation and purpose

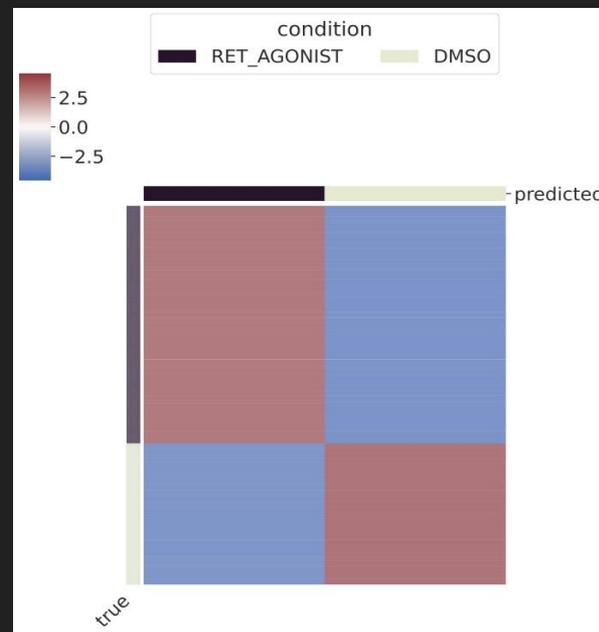
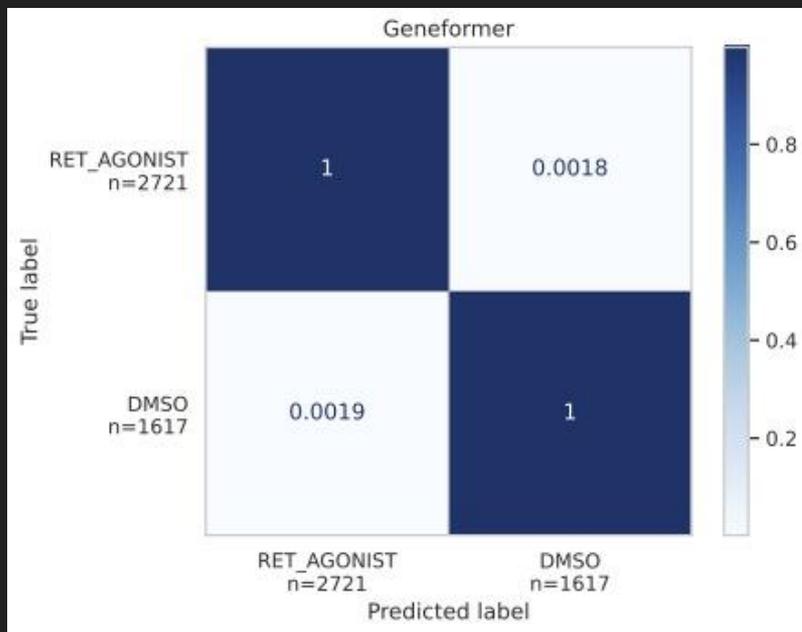
Geneformer is an encoder only transformer architecture: a context-aware, attention-based deep learning model

- A large - scale pretraining corpus, Genecorpus-30M, comprising 29.9 million human single-cell transcriptomes from a broad range of tissues from publicly available data was assembled
- Pre-training was then carried out on this corpus through a self-supervised masked learning objective
- This way, Geneformer gained a fundamental knowledge which can be transferred towards a broad range of fine-tuning tasks and downstream applications to accelerate discovery of key network regulators and candidate therapeutic targets



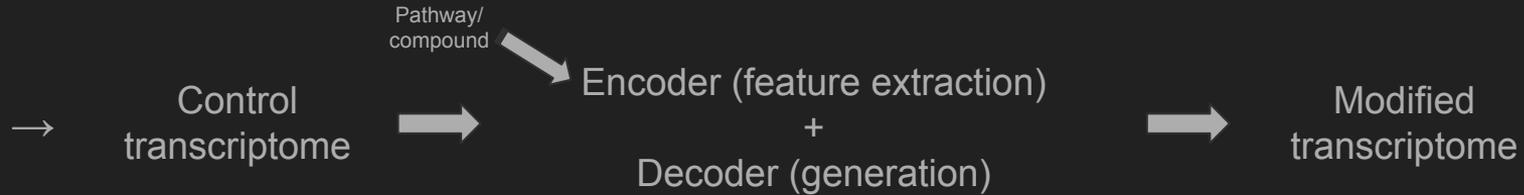
# Results

Fine-tuning to distinguish between *RET\_AGONIST* and *DMSO* exposure



# Limitations and inspiration for further research

- Unfortunately, the fine-tuning process was computationally very expensive  
→ enable Geneformer to distinguish among a wider variety of categories
- Our model is able to, starting from a modified transcriptome, deduce which pathway/compound the cell was exposed to



... experiment with different transformer models e.g. scGPT, or theories e.g. Optimal Transport, Diffusion Models, Diffusion Schrödinger Bridges (DSB)