Bayesian ST Inference: Uncertainty Quantification in Spatial Transcriptomics

Thanosan Prathifkumar, Nehan Mohamed
Emerging Diagnostic and Investigative Technologies, Department of Pathology, Dartmouth Hitchcock Medical Center

Dartmouth Health

ABSTRACT

- Built a Bayesian neural network with Monte Carlo sampling + heteroscedastic loss to estimate epistemic and aleatoric uncertainty
- Core/metabolic pathways show low uncertainty; stress-response and immune pathways show high uncertainty
- Cancer tissues show high uncertainty in cell-death and immune regulation-related pathways
- Implementing a Bayesian framework on VRI predictions provides context towards predictions

INTRODUCTION

Virtual RNA Inference: Virtual RNA inference (VRI) models are able to predict spatial transcriptomic data from hematoxylin and eosin stained tissue images.

- Spatial transcriptomic data provide spatial context to gene expression data, but they can be very expensive
- Virtual RNA inference models are able to predict ST-level data from hematoxylin and eosin stained slides through neural networks, but it is hard to quantify their uncertainty

Bayesian Neural Networks: Bayesian neural networks can quantify uncertainty within predictions by creating distributions of neural network parameters.

- Variational inference is performed. This technique estimates probability distributions via gradient descent.
- Heteroscedastic loss is implemented by having the BNN predict mean and log variance for each gene, allowing genes to have different uncertainty levels

Uncertainty Quantification: To quantify uncertainty during testing, techniques such as monte carlo sampling must be implemented.

- Monte carlo sampling is implemented to estimate a posterior distribution
- Epistemic uncertainty is uncertainty due to lack of learning; aleatoric uncertainty is uncertainty due to noise and measurement uncertainty.
- Epistemic uncertainty can be calculated as the variance of mean predictions; aleatoric uncertainty can be calculated as the mean variance across samples.

METHODS

Goal: Create a Bayesian framework that is able to quantify uncertainty within VRI predictions

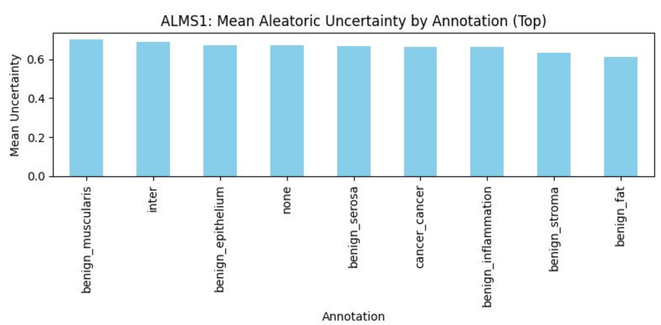
Data: Hematoxylin and eosin stained tissue slides paired with Visium spatial transcriptomic data and annotations.

Experimental Design:

- Using *bayesian-torch* to implement the VRI models into a Bayesian model
- *bayesian-torch* implements variational inference, allowing different weights to be sampled each time
- Heteroscedastic loss is used to give different uncertainty predictions for different genes
- Trained VRI models on 512x512 patches with whole slide images as input and ST-level data as output
- Using Bayesian model to make predictions on testing data, and extracted epistemic and aleatoric uncertainty on a gene-by-gene level using monte-carlo sampling

RESULTS

- Among genes with high uncertainty, aleatoric uncertainty is consistent across all architectures; for genes with low uncertainty, these uncertainties tend to vary
- Core/metabolic pathways (such as oxidative phosphorylation or mitochondrial ATP synthesis) exhibit low uncertainty, while stress-response pathways exhibit high uncertainty (such as positive regulation of apoptotic process)
- Aleatoric uncertainty is tied to metabolic/homeostasis pathways; epistemic uncertainty is tied to immune signalling and RHO GTPase pathways.
- In cancerous tissue, there is high uncertainty in cell-death and immune regulation-related pathways



DISCUSSION

- Core/metabolic pathways are predicted with high confidence as these genes are prevalent throughout various tissue contexts; stress-response and immune pathways can vary across different tissue contexts and can give varying results.
- In benign tissue, immune-related pathways are harder to predict, potentially signifying that these are hard to predict in non-pathological contexts; in cancerous tissue, there is high uncertainty in pathways related to cell death, likely signaling the high amount of heterogeneity in tumor microenvironments
- Aleatoric pathways tend to show metabolic or homeostasis pathways, signifying that these pathways are related to biological randomness
- Epistemic pathways relate to immune signaling and RHO GTPase pathways, which makes sense as these can be heterogeneous

FUTURE WORK

- Work towards spot-level or single-cell level resolution instead of just patch-level
- See the effect multi-modality has on epistemic uncertainty in predictions

CONCLUSION

- This study's finding shows the benefit that uncertainty quantification can provide in virtual RNA inference
- Uncertainty quantification can tell us about what genes are context-dependent and what genes are inherently noisy in their measurements
- Uncertainty quantification can also show pathways that are prone to biological randomness and pathways that are heterogenous and vary from tissue-to-tissue

edit.