Annotation of UV-Damaged Skin Histologies for Virtual RNA Inference Research in Photoaging

Edwardia Fosah

Dartmouth Health

ABSTRACT

- Skin cancer is the **most common** form of cancer in the United States
- Photoaging, skin damage resulting from acute UV radiation, is a **major risk factor** for skin cancer but **lacks reliable forms of measurement**
- Virtual RNA Inference (VRI) - a computational method that uses hematoxylin and eosin (H&E) stained tissue images to **predict spatial gene expression**
- VRI could be used to identify photoaging with **comparable performance** to direct measurement (spatial transcriptomics [ST])
- VRI could also be applied to **downstream tasks** (like cell type abundances prediction and age prediction) that can be verified through annotations of skin whole slides images (WSIs).

INTRODUCTION

- Photoaging **lacks reliable measurement** due to: unstandardized histological assessments and inconsistencies in self-assessments of UV exposure
- Identification of photoaging can increase likelihood of early cancer detection, thereby increasing 5-year survival rate to over 99%
- Inferred ST data from VRI can assist with cell-type specific features of photoaging **without the challenges of ST** technologies: high cost, low throughput, and reproducibility issues
- A major limitation in the current Levy Lab photoaging VRI research is the lack of histology annotations of skin tissue whole slide images which this study sets out to alleviate through the generation of high quality, digital annotations

METHODOLOGY

Data:

- Data comprises 261 H&E WSIs of skin tissues coming from various parts of the body, such as the cheek, nasal sidewall, forehead, and temple, with c**heek tissue being the most common**
- Significant majority of the WSIs are of **UV-damaged skin tissue**
- Tissues were collected at Dartmouth Health Dermatology in the Mohs Clinic
- Mohs surgeons cut out **Burow's triangles** at different sites of the body, but mostly the face
- Triangles were bisected with one half going to methylation and one half going to Visium for ST data measurement
- **2 cohorts of data**: the Visum cohort with 16 tissues from 16 patients who have matching WSIs and a remaining cohort of 261 WSIs
- First cohort will be used to further **develop the VRI** model, and the second cohort will be used to **validate the VRI algorithm**
- Data had originally been annotated for 4 out of the 16 Visium WSIs and 25 out of 261 of the WSIs of the remaining cohort
- 14,000 Visium spots per WSI, 5,000 spots per patient, each spot is 55 microns wide with 100 microns between each center

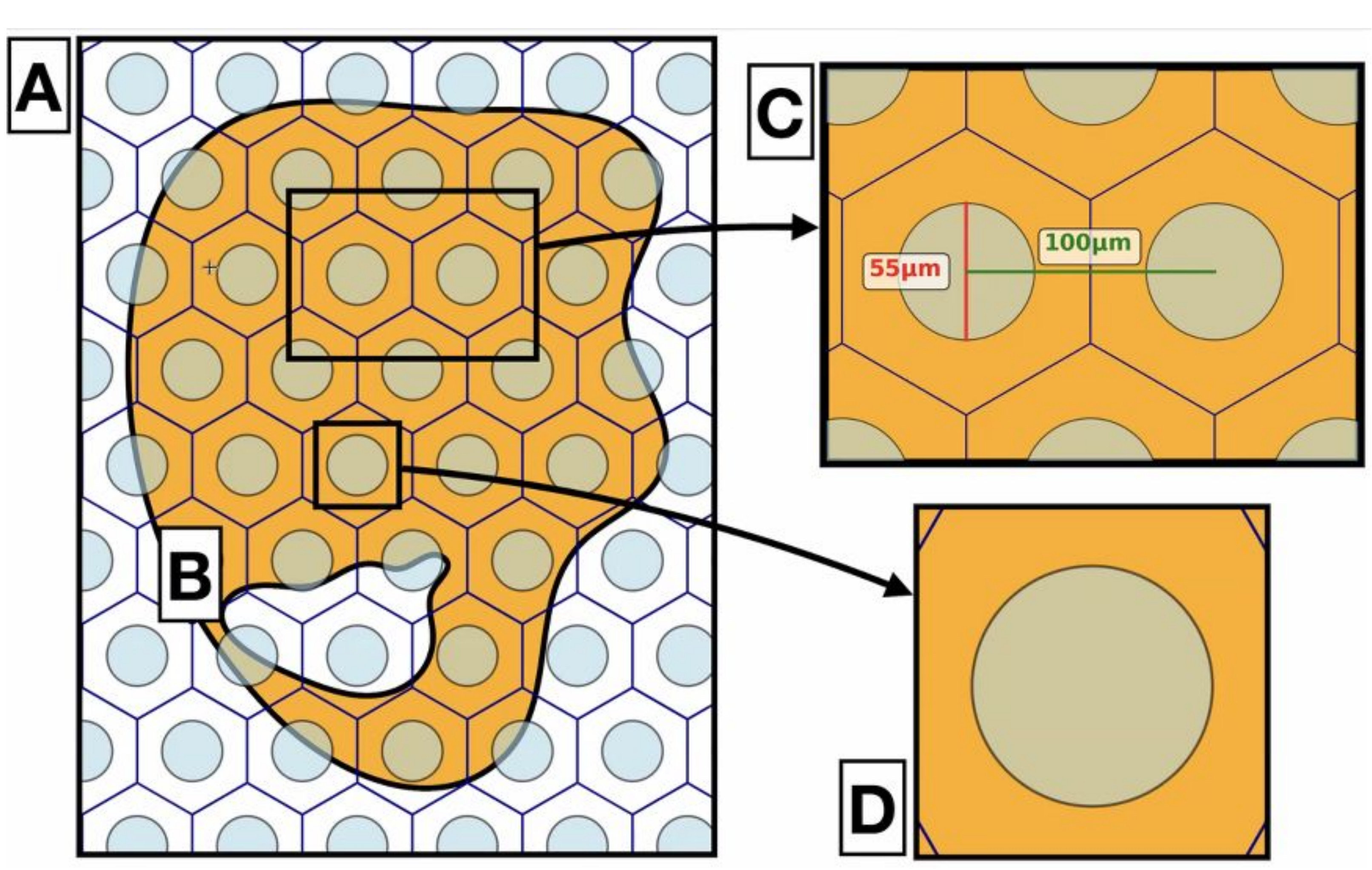

Figure 1: Structure and spacing of Visium spots

Annotation:

- Software used for annotation was **QuPath**, an open software for bioimage analysis
- Annotations were created in a by pixel, **segmentation** format
- **Classes**: sebaceous gland, hair follicle, vessel, eccrine gland, nerve, fat, epidermis, and smooth muscle
- Annotations were primarily created using the brush tool in QuPath while magnified
- Larger annotation areas, such as extensive collections of fat, occasionally utilized the magic wand tool with manual touch ups
- Annotations were saved as **GeoJSONs**, exported as a **FeatureCollection**, and included **measurements**

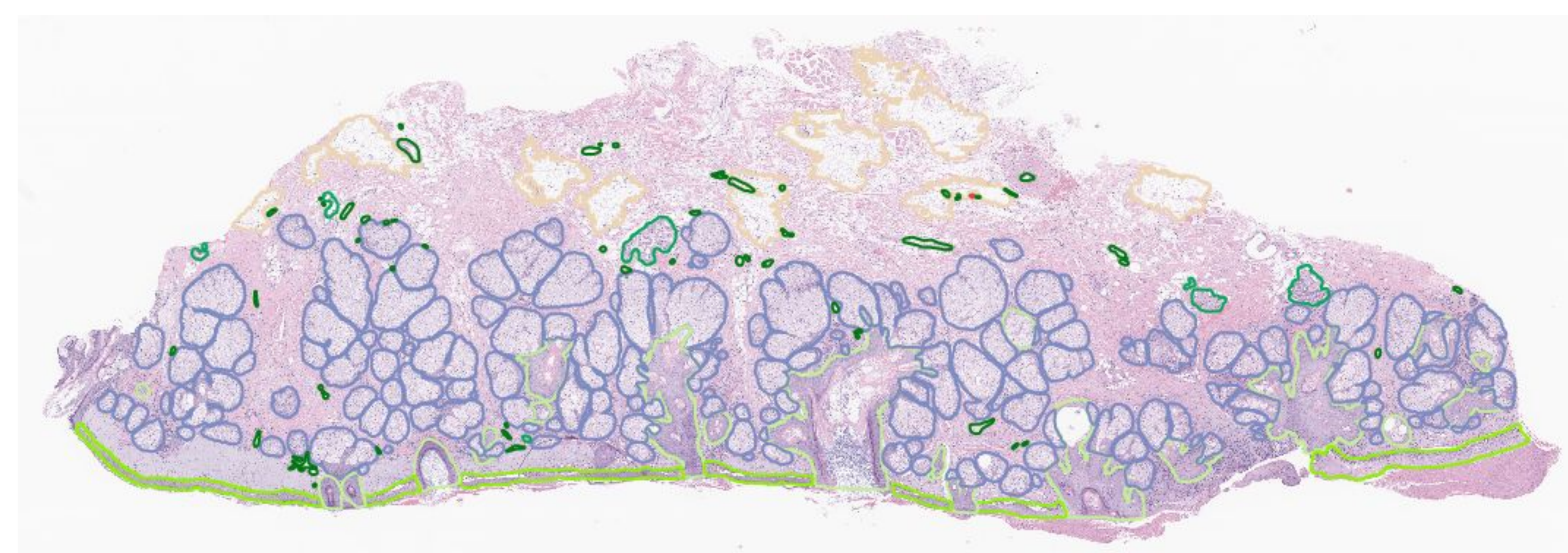

Figure 2: Example of QuPath annotated WSI

RESULTS

- **Nine** different skin tissues with matching Visium data were annotated
- 1228 individual segmentations of skin tissue structures
  - **428** sebaceous gland annotations, **232** hair follicle annotations, **275** vessel annotations, **46** eccrine gland annotations, **39** nerve annotations, **107** fat annotations, **82** epidermis annotations, and **10** smooth muscle annotations

DISCUSSION

- These annotations can act as the data for the training of **segmentation models** that can automatically annotate WSIs
- **Example** - UNet model with a pretrained ResNet backbone
- **Workflow** - hold out entire slides for our validation set, use OpenSlide to work with SVS files, use early stopping
- Annotations can also train **patch-level classification models**
- **Examples** - ResNet with a pathology specific model OR CNN/ViT setup
- **Workflow** - patch and feature extraction conducted with TRIDENT
- **Limitations** - shallower histology knowledge of a non-pathologist; genes with a clear histological basis are more likely to be accurately predicted compared to genes without; our models are trained on only skin tissue WSIs thus they can only be expected to perform well on skin tissue data

CONCLUSION

- Goal of the greater Levy Lab photoaging project is to **validate the downstream task prediction** of inferred ST data from VRI, such as cell type abundances and age
- **Different baseline models** that can be used for age prediction with inferred ST data: XGBoost regression on the inferred gene matrix; Denoising Autoencoder (DAE) embedding model with a Graph Convolutional Network (GCN) on the embedded gene matrices; GCN on the inferred gene matrices
- **Future directions** include creating different VRI models using different embeddings, eventually taking advantage of **multimodal modeling** with embedded H&E patches and inferred gene expression to predict age
- In this, we expect inferred ST data to be able to predict age and cell type abundance at a **comparable performance** to measured ST data
- More expansive research that includes **more areas of anatomy** would be required to generate accurate spatial gene expression data **across domains**

edit.