

Mapping Cancer Gene Expression with *catmap*

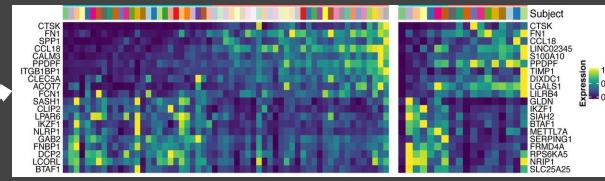
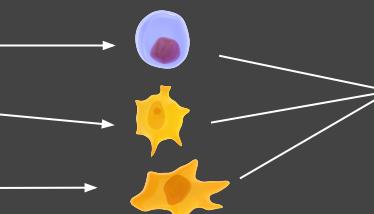
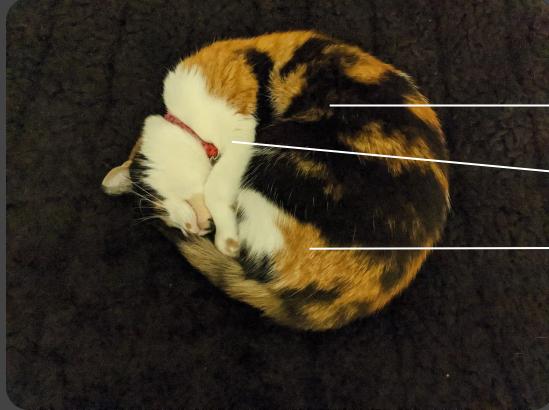
catmap: **c**a(ncer) **t**(ranscriptomics) **m**ap.

Marissa Fujimoto
Alyssa Mell
Erik Pedersen
Vaishnavi Verma



By *Ellisn95* - Own work, CC BY-SA 4.0,
<https://commons.wikimedia.org/w/index.php?curid=124199574>

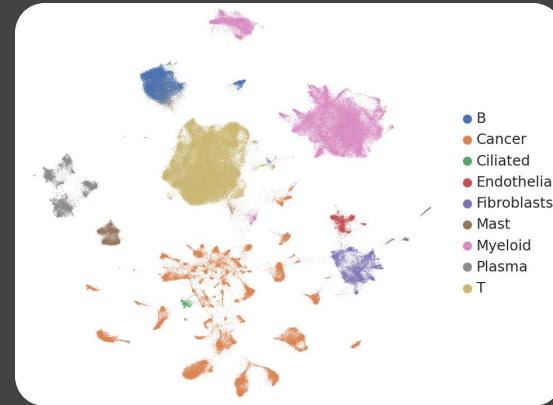
What is Gene Expression Data?



Cells have the same *DNA*,
but express different *genes!*

EvaWest, CC BY-SA 4.0
<<https://creativecommons.org/licenses/by-sa/4.0/>>, via Wikimedia Commons

Goal: Visualize Cancer Gene Expression



Mikael Häggström, M.D. Author info - Reusing images- Conflicts of interest: None
Mikael Häggström, M.D. Consent note: Consent from the patient or patient's relatives is regarded as redundant, because of absence of identifiable features (List of HIPAA identifiers) in the media and case information (See also HIPAA case reports guidance)., CC0, via Wikimedia Commons

Let our eyes *see cancer gene expression* !

Cancer Gene Expression Data

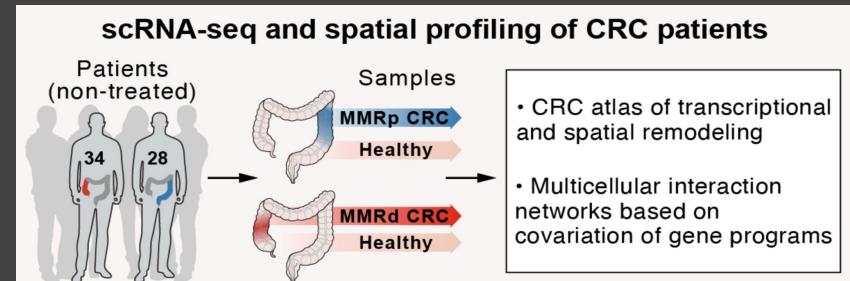
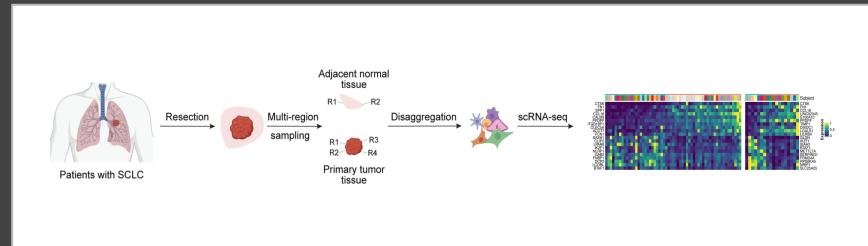
Single cell transcriptomics data from lung cancer patients.

*224,611 cells, 72,131 genes per cell.
Collected from 104 patients*

Metadata and t-SNE dimensionality reduction data from colon cancer patients.

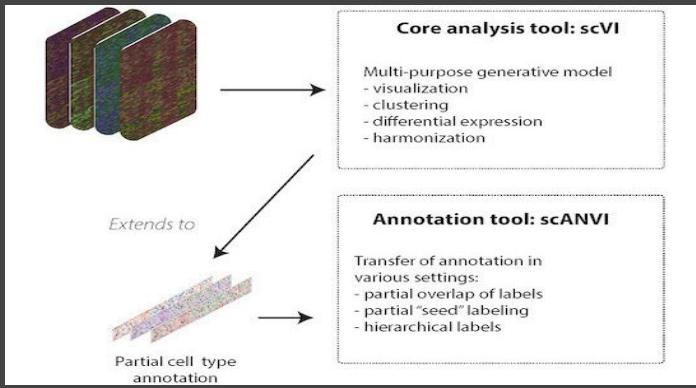
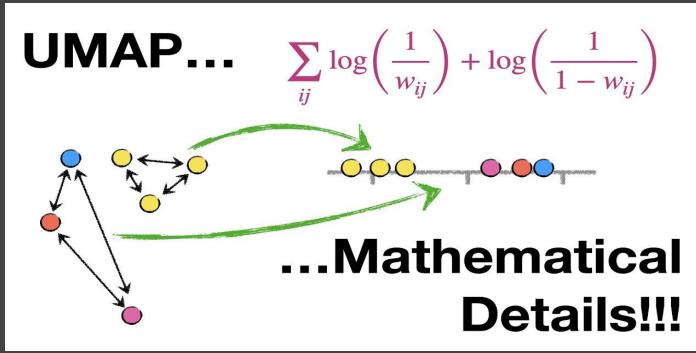
*371,223 cells, 43,282 genes per cell.
Collected from 62 patients*

Data: Single cell gene expression from cancer tumors
Modified from
<https://www.nature.com/articles/s41392-022-01150-4/figures/1>
<https://www.biorxiv.org/content/10.1101/2022.02.07.479293v1.full.pdf>

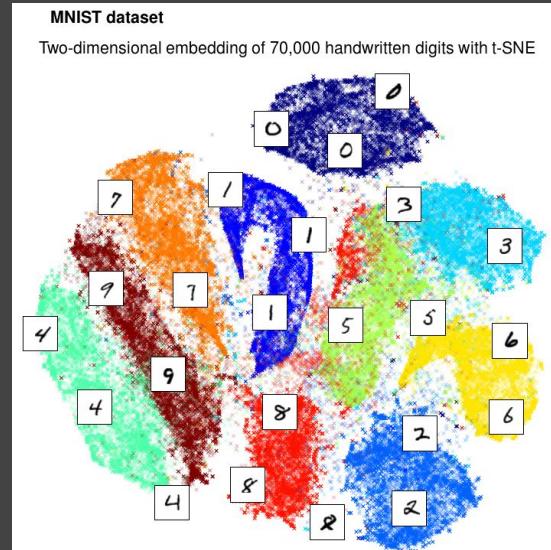


Machine Learning Methods

Lung
cancer
dataset



Colon cancer dataset



User Experience

Focus: Intuitive & Clean

Doctors

*Study Visualizations
Embed Data*

Patient

*Learn more about cancer
Fit their data into catmap*

Public Health Officer

Learn about cancer at a population level



Software Design



Streamlit

Technology

Streamlit, Plotly

- Faster to learn, good performance for lots of data

Implementation

Components

- Machine learning models, pages, individual components within pages
- Easier to work in parallel, separate UI from machine learning

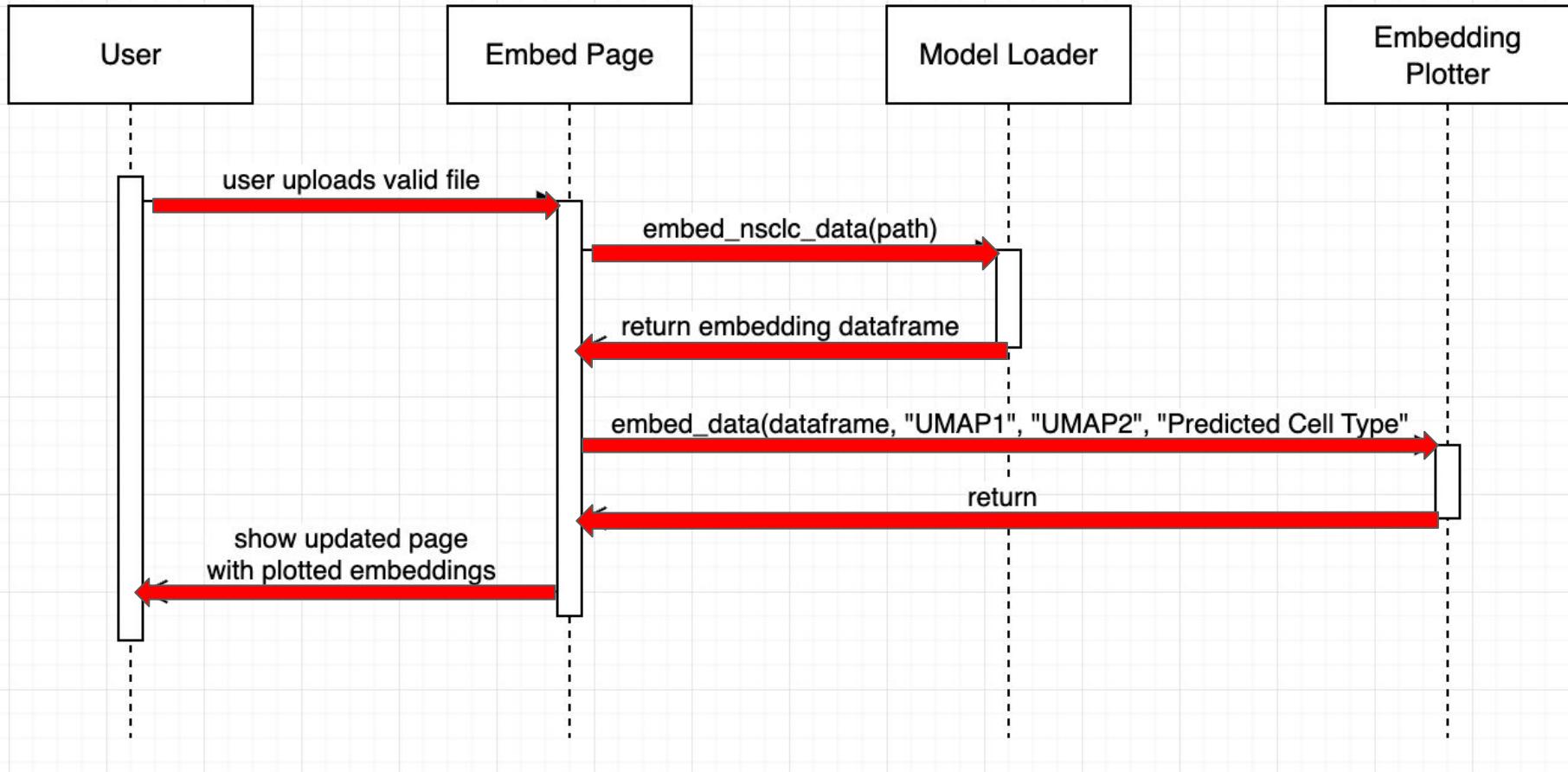
Preprocessing

- Stored dataframe from machine learning model
- Lazily loading for new embedding

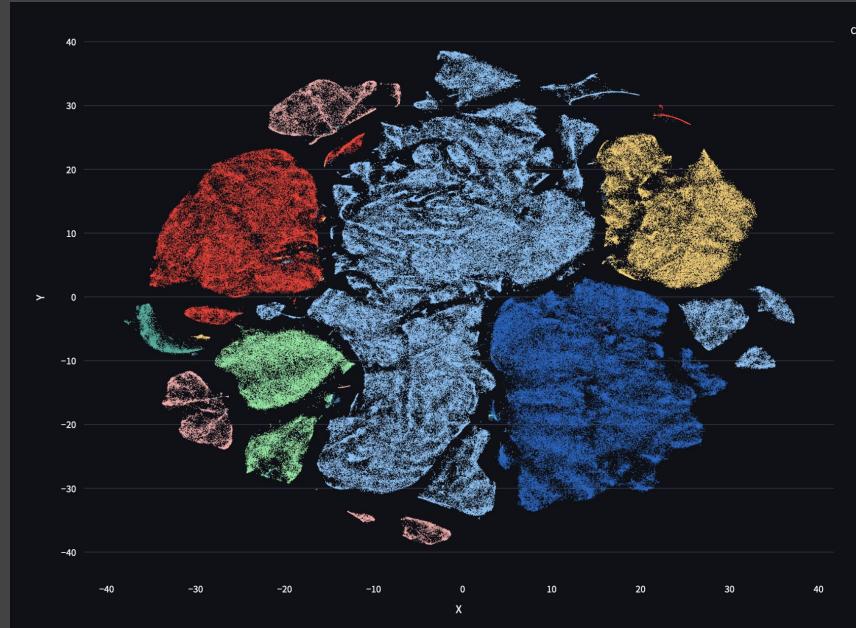
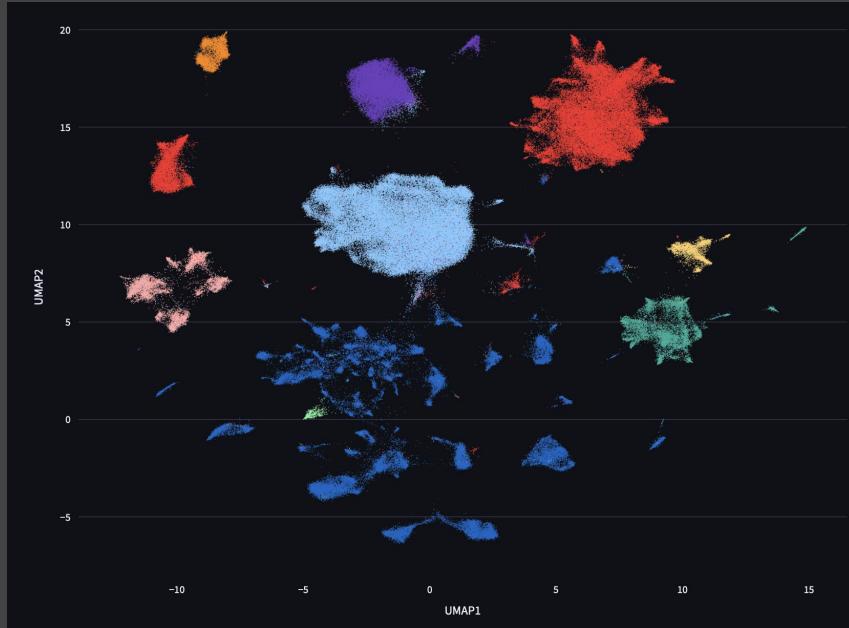
Example

- Use case: doctor uploads new data





Demo Part 1: Visualization



Demo Part 2: Embedding New Data / Transferring Labels



Next Steps?

Better embedding!

Include preprocessing workflow

More tools to inspect results

More powerful visualizations!

Multiple filters

Side by side comparisons

More datasets!

Explore more cancer types!

Train better ML models!



Sofocle77, CC BY-SA 4.0
<<https://creativecommons.org/licenses/by-sa/4.0/>>, via Wikimedia Commons

Conclusion

catmap: **ca**(ncer) **t**(ranscriptomics) **map**.

A data visualization application and tool

Embed and visualize single cell gene expression data from cancer cells



Robert Calderon, CC BY 3.0
<<https://creativecommons.org/licenses/by/3.0/>>, via Wikimedia Commons