**Brain Metastasis Database**
Shijie Fan, Joyce Fang, James Jiang, Yutong Qiu
Advisor: Dr. Debashis Sahoo

**MOTIVATION**

What is brain metastasis?
It is one of the most devastating effects of cancer which occurs when cancer cells originating in another region or organ of the body travels to the brain.

Why study brain metastasis?
The median survival time is a dismal 7.16 months, and our current medical tools have difficulty detecting brain metastasis in its early stages.

**BACKGROUND**

One method to help diagnose brain metastasis early in patients is to analyze genes relevant to the disease. Therefore, microarray data, which directly measures the amount of RNA in genes, is an ideal source of data for our research goal.

**TOOLS**

- GEO (Gene Expression Omnibus), a public genomics data repository, for collecting data
- Google Excel Sheet API to export annotated samples and characteristics
- Hegemon to visualize organized data
- Python to parse data and automate changes

**REFERENCES**


---

**OUR GOALS**

1. Build a brain metastasis database
2. Provide a website for other researchers to access the organized data
3. Analyze data to find interesting relationships between genes that may indicate a biomarker for brain metastasis

**CONTRIBUTIONS**

- Search for datasets in GEO
- Manually annotate each sample
- Produce six files from samples for data visualization with our bash script
- Hegemon will display each gene expression

We organized and annotated 2347 samples in Google Excel Sheets with specified characteristics and exported the data to produce organized files for Hegemon to parse and display.

<table>
<thead>
<tr>
<th>Array ID</th>
<th>c Sample Site</th>
<th>c Tissue Type</th>
<th>c Cancer Type</th>
<th>c Sample Type</th>
<th>c Cell Line/Patient ID</th>
<th>c Primary Cancer Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSM354034</td>
<td>Brain</td>
<td>Lung</td>
<td>Metastasis</td>
<td>Cancer</td>
<td>YX_U133_2_100874T</td>
<td>Distant Organ Metastasis</td>
</tr>
</tbody>
</table>

**DATA SUMMARY**

- **Cancer Type**
  - Metastasis: 13.4%
  - Primary: 86.6%

- **Metastasis Sites**
  - Osseous: 12.3%
  - Liver: 6.9%
  - Lung: 5.4%
  - Brain: 16.5%
  - Other: 16.5%
  - Omentum: 15.3%
  - Peritoneum: 4.2%
  - Lymph Nodes: 2.2%
  - Pleura: 2%
  - Pleural: 2%
  - Other: 15.3%

This is a summary of the 2347 samples we collected and annotated in GPL570, demonstrating the diversity of the metastasis sites.

**CONCLUSION**

We built a brain metastasis database from gene expression profiles of large patient-derived datasets and have designed the website for displaying the data we have annotated.

**FUTURE WORK**

- Continuously collect and annotate data
- Analyze more genes to find a biomarker
- Implement interactive features and functionality for the website

**ACKNOWLEDGEMENTS**

We would like to thank Debashis Sahoo, Christine Alvarado, and Aditi Mavalankar for all their guidance and support. This material is based upon work supported by the National Science Foundation under Grant No. CNS-1339335.