Unsupervised Clustering and Heatmap Analysis of Human Schwannoma Metabolites

Mark Dougherty
Project Homework #1 - BIOL 4386: Intro to Scientific Computing
Due February 13, 2023
Revision written & submitted February 21, 2023

Reference Masalha, W., Daka, K., Woerner, J. *et al.* Metabolic alterations in meningioma reflect the clinical course. *BMC Cancer* 21, 211 (2021). https://rdcu.be/c5yzG

Introduction Schwannomas are benign (WHO grade 1) tumors that grow on peripheral nerves, originating from the Schwann cells that myelinate the nerve axons. Vestibular schwannomas are those that occur on the vestibular nerve intracranially, and account for about 8% of all primary brain tumors. Surgery and radiation are first-line treatments for these tumors, but if they fail there is no second-line therapy. Thus, novel medical treatments are needed. What's more, although all schwannomas are considered Grade 1 tumors, some are more aggressive than others; we do not currently have a biological explanation for this inter-tumor heterogeneity or a good way to predict this behavior in order to modify our clinical care. Recent literature has suggested a possible role of the tumor-immune microenvironment, as there are differences in macrophage infiltration between different tumors. Another recent development in the literature is that DNA methylation profiling can identify tumors with more aggressive phenotypes better than genomic or transcriptomic analysis, suggesting that schwannomas may be driven largely by *epigenetic* changes rather than the classical genetic mutation paradigm in cancer.

My research focuses on trying to improve our understanding of the underlying biology of schwannomas in hope of finding vulnerabilities that could serve as drug targets, as well as further explain differences in tumor recurrence/aggressiveness. Specifically, I have been using metabolomics, which is a method of analyzing levels of many (~100-150) metabolites in a tissue at a given time. Metabolomic analysis has not previously been used to study schwannomas, but has shown promise in finding novel treatment targets in other tumors/cancers. Thus, our aims are twofold: identify novel drug targets in metabolic pathways, and identify biological differences that might explain differences in tumor behavior. In the future, I also hope to integrate this analysis with other '-omics' data and clinical outcomes, but that is likely beyond the scope of the current project.

In this project, my aim is to use R to process data from metabolomic analysis of primary human schwannoma samples, and then use unsupervised clustering analysis and a heatmap to evaluate whether there are meaningful clusters of tumors that seem to be metabolically similar. I will use the same unsupervised PAM cluster analysis method as the authors of this paper use, although at this time I do not know the specifics of this cluster method or why they chose it over

other unsupervised clustering methods. This is an exploratory analysis, so it is possible that I will not find clean clustering as is the case with the reference figure. On the other hand, if I do identify strong data clusters, further steps would then be needed to determine what the groups/clusters mean. In order to evaluate the clustering visually, I will combine the clustering analysis to arrange the metabolites, and then visualize with a heatmap as in Figure 2A of Masalha et al. Notably, I do not intend to perform the analyses in parts B & C of the same figure, nor do I anticipate including a 'Silhouette width' graph as they do at the top part of their figure.

Figure to reproduce: Figure 2A

Materials and Methods

Specimen collection

- Schwannoma specimens are collected directly from surgical patients at UIHC. A *primary tumor* specimen is flash-frozen in liquid nitrogen in the operating room. When available, additional tissue is implanted in 8-9 nude mice per human tumor (*patient-derived xenografts*); after the mice recover (~2-4 weeks), these xenografts are treated with radiation (0, 10, 20 Gy) and harvested 72 hours post-treatment. The *primary tumors* and *xenografts* are then metabolically profiled with GC-MS and/or LC-MS (AKA metabolomics).

Data Preprocessing & Cleaning

- The UI Metabolomics Core performs the mass spectrometry analysis and provides the data to our lab as relative concentrations of each metabolite in a **labeled Excel spreadsheet**. Each sample has ~100-150 metabolite levels measured. *Critically, these are relative levels rather than absolute concentrations. This means that we can compare one metabolite between different samples (e.g. glutamine 2x higher in Sample X than in Sample Y), but we cannot directly compare levels of different metabolites (e.g. cannot state "glutamine is 2x higher than glutamate").*
- Non-metabolic information must then be manually associated with samples. For example, in the above figure 2A this would include Edema, Proliferation, Gender, and WHO Grade at the bottom of the heatmap. Our samples are labeled with: NF2 status (categorical), prior radiation (categorical), prior surgery (categorical), and proliferation (continuous; from EdU assay, [xenografts only]). Some samples also have freeze time data (continuous)
- Data will then be imported from Excel into R
- Double check that undesired samples are excluded from further analyses (eg 2022.3.11 known ischemic sample)
- **Outlier detection**: Grubbs' test, alpha = 0.01. This will be done in R; I do not currently know how to do this but assume it is doable.

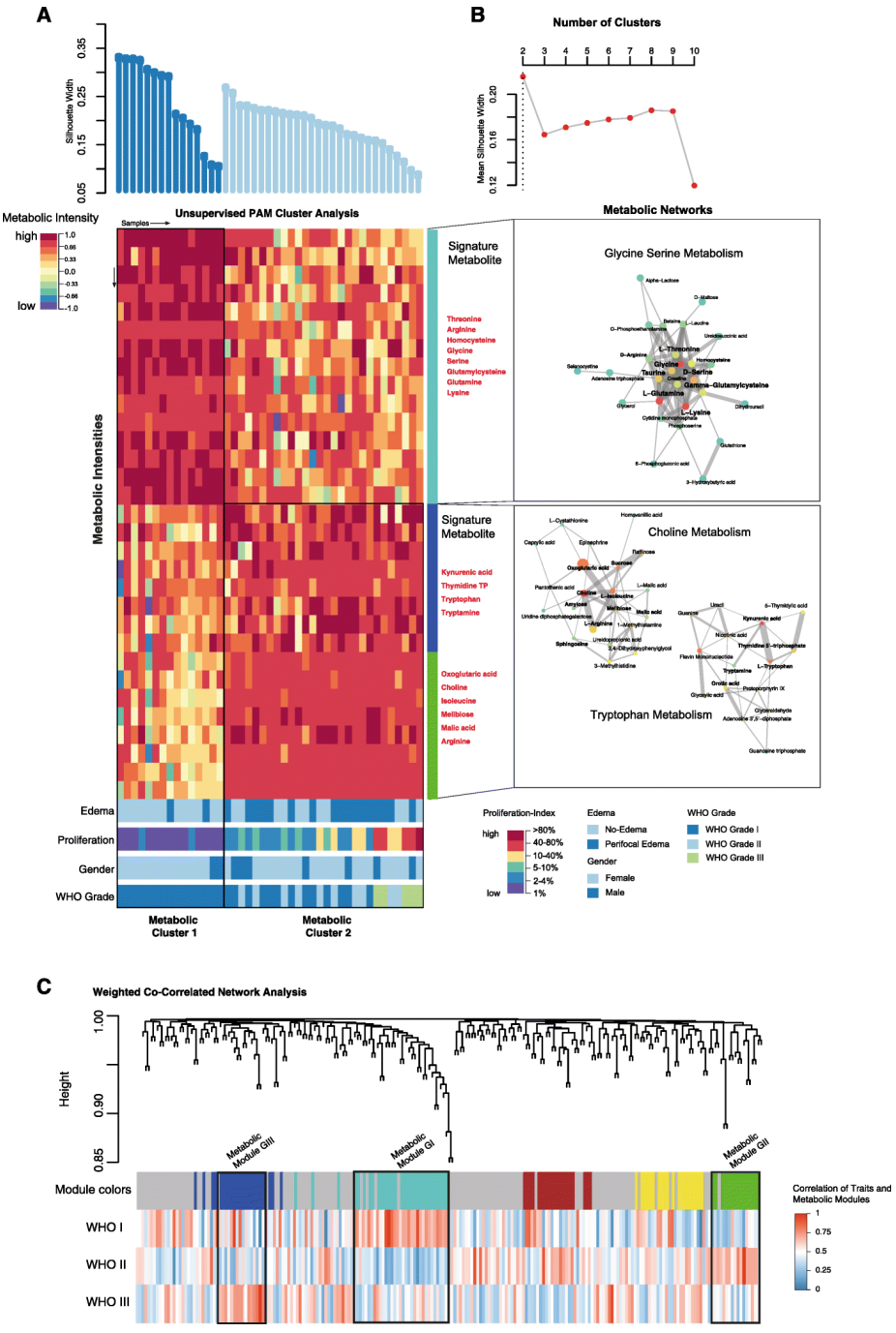

Figure 1: Figure 2A taken from Masalha et al

Data analysis: Primary Tumor Samples (primary goal)

- **Clustering & Heatmap**
  - As described in the Methods of the journal article cited above, I hope to perform cluster analysis using their R package **AutoPipe**, which they make available on GitHub. I hope to follow their method of
  - If this does not work for some reason, I will need to explore other clustering algorithms/R packages such as K-means. I am not certain from a statistical standpoint which clustering algorithm is best for my data.

Data analysis: Xenografts (radiation treatment) (secondary goal)

- **Normalization**: Fold change calculations (xenografts only)
  - *Xenograft fold change calculations are complicated because want to normalize to mean of control group for each specific tumor*
  - Example: Tumor from patient 123 was implanted into 9 mice, and those 9 mice were randomized to 0, 10, or 20 Gy radiation treatment (3 per group). Tumor from patient 456 was also implanted into 9 mice. However, to evaluate the fold change of a given metabolite after radiation, we want to compare the 10 & 20 Gy treatment groups from Patient 123 to the control tumors of Patient 123, and Radiated Tumors from Patient 456 to Control Tumors from Patient 456.
- **Test for normality** - Shapiro-Wilk test
- **Transformation** (when needed) - for non-normally distributed metabolites, LogTransform the values
  - Easily done in GraphPad. Don't know how to do in R but should be possible
- **Statistical analysis (primarily for xenografts)**
  - Correlation with radiation dose (per metabolite)
  - Two-way ANOVA with Holm-Sidak test for xenografts to compare radiation treatment doses
- **Graph**