Analysis of epitranscriptome for grading of glioma
Eric Rivals

To cite this version:
Eric Rivals. Analysis of epitranscriptome for grading of glioma: Multivariate Analysis of RNA Chemistry Marks Uncovers Epitranscriptomics-Based Biomarker Signature for Adult Diffuse Glioma Diagnostics. Symposium on Cancer genomics and epitranscriptomics: from the bench to the clinic, Comprehensive Cancer Center (CIC), University of Salamanca, Nov 2022, Salamanca, Spain. lirmm-04287393

HAL Id: lirmm-04287393
https://hal-lirmm.ccsd.cnrs.fr/lirmm-04287393
Submitted on 15 Nov 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution - NonCommercial - ShareAlike| 4.0 International License
Analysis of epitranscriptome for grading of glioma

Eric Rivals

LIRMM, CNRS Univ Montpellier
Objective and summary

Glioma: the most frequent brain cancer, is difficult to cure and to follow up.

Aim:

1. determine whether a panel of epitranscriptomic marks can serve as biomarkers for glioma grading
2. pipeline for grade prediction using epitranscriptomic profile
Diffuse glioma: the most frequent of brain tumors

Glioma: intracranial tumor of glial cells

- Diverse forms and locations in the brain.
- 2nd most frequent tumor among child cancers
- Diagnostic: using Magnetic Resonance Imaging (MRI), scanner, and biopsy.
- Three possible grades denoted II, III, IV

Clinical question: distinguish the grades

Figure: Left: MRI of a grade II oligodendroglioma. Right: MRI of a grade III anaplastic astrocytoma. Source: Manuel MSD, Steven A. Goldman
Composition:
58 tumor samples and 19 control samples
- Tumors: surgery from adult patients diagnosed with diffuse glioma.
- Histopathological types according to revised World Health Organization classification
- Control samples (n = 19): from non tumoral brain surgeries (epilepsy, benign lesion, etc.)

Distribution

<table>
<thead>
<tr>
<th>type name number</th>
<th>grade-II glioma</th>
<th>grade-III glioma</th>
<th>grade-IV glioblastoma</th>
<th>control non-tumor</th>
</tr>
</thead>
<tbody>
<tr>
<td>number</td>
<td>20</td>
<td>20</td>
<td>18</td>
<td>19</td>
</tr>
</tbody>
</table>
Methods - sample analysis

From tissue to data file:

1. RNA extraction (10 µg)
2. Digestion into nucleosides
3. Liquid Chromatography Mass Spectrometry
4. Quantification (spectrum analysis)
5. Statistical & Machine Learning analysis
Mass spectrometry - spectrum of A and modified nuc
<table>
<thead>
<tr>
<th>Standard</th>
<th>modified nucleosides</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Am m1A m66A m66Am m6A m6Am I</td>
</tr>
<tr>
<td>C</td>
<td>Cm ac4C m3C m5C hm5C</td>
</tr>
<tr>
<td>G</td>
<td>Gm m1G m227G m27G m7G oxo8G</td>
</tr>
<tr>
<td>U</td>
<td>Um m3Um mcm5U mcm5s2U ncm5U Psi Queuosine</td>
</tr>
</tbody>
</table>
For each sample, 29 MS measures for modified and standard nucleosides. 
\textbf{Math}: a vector of 29 dimensions $\Rightarrow$ \textit{epitranscriptomic profile}

\textbf{Questions}
1. Do epitranscriptome modification measures vary with cancer grade? 
2. Can one predict glioma grade from a sample using machine learning? 
3. Which combination of nucleosides are most important for prediction?

\textbf{Methods}
Python scripts using libraries Scikit-learn [5], Lifelines [1], Pandas [6], Matplotlib [2], Seaborn [8]
Nucleosides with decreasing quantities wrt cancer grade

**Figure:** Boxplots wrt grade of m1G, ac4C, and oxo8G.
Nucleosides with increasing quantities wrt cancer grade

Figure: Boxplots wrt grade of Gm, Um, and m6Am.
Correlation between nucleoside quantities

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>C</th>
<th>U</th>
<th>m1G</th>
<th>oxoG</th>
<th>QuoBase34C</th>
<th>G</th>
<th>m6A</th>
<th>Um</th>
<th>Gm</th>
<th>Am</th>
<th>m3C</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1.0</td>
<td>0.1</td>
<td>0.7</td>
<td>0.2</td>
<td>0.3</td>
<td>0.4</td>
<td>0.2</td>
<td>0.6</td>
<td>0.4</td>
<td>0.7</td>
<td>0.3</td>
<td>0.2</td>
</tr>
<tr>
<td>C</td>
<td>0.1</td>
<td>1.0</td>
<td>0.3</td>
<td>0.2</td>
<td>0.4</td>
<td>0.6</td>
<td>0.3</td>
<td>0.4</td>
<td>0.3</td>
<td>0.5</td>
<td>0.4</td>
<td>0.3</td>
</tr>
<tr>
<td>U</td>
<td>0.7</td>
<td>0.3</td>
<td>1.0</td>
<td>0.4</td>
<td>0.6</td>
<td>0.7</td>
<td>0.4</td>
<td>0.4</td>
<td>0.5</td>
<td>0.7</td>
<td>0.4</td>
<td>0.3</td>
</tr>
<tr>
<td>m1G</td>
<td>0.2</td>
<td>0.2</td>
<td>0.4</td>
<td>1.0</td>
<td>0.5</td>
<td>0.6</td>
<td>0.3</td>
<td>0.3</td>
<td>0.4</td>
<td>0.7</td>
<td>0.4</td>
<td>0.3</td>
</tr>
<tr>
<td>oxoG</td>
<td>0.3</td>
<td>0.4</td>
<td>0.6</td>
<td>0.5</td>
<td>1.0</td>
<td>0.7</td>
<td>0.4</td>
<td>0.4</td>
<td>0.5</td>
<td>0.7</td>
<td>0.4</td>
<td>0.3</td>
</tr>
<tr>
<td>QuoBase34C</td>
<td>0.4</td>
<td>0.6</td>
<td>0.7</td>
<td>0.6</td>
<td>0.7</td>
<td>1.0</td>
<td>0.5</td>
<td>0.5</td>
<td>0.7</td>
<td>0.7</td>
<td>0.4</td>
<td>0.3</td>
</tr>
<tr>
<td>G</td>
<td>0.2</td>
<td>0.3</td>
<td>0.4</td>
<td>0.4</td>
<td>0.5</td>
<td>0.6</td>
<td>1.0</td>
<td>0.6</td>
<td>0.5</td>
<td>0.7</td>
<td>0.5</td>
<td>0.4</td>
</tr>
<tr>
<td>m6A</td>
<td>0.6</td>
<td>0.4</td>
<td>0.7</td>
<td>0.6</td>
<td>0.7</td>
<td>0.7</td>
<td>0.5</td>
<td>1.0</td>
<td>0.7</td>
<td>0.7</td>
<td>0.6</td>
<td>0.5</td>
</tr>
<tr>
<td>Um</td>
<td>0.3</td>
<td>0.6</td>
<td>0.4</td>
<td>0.5</td>
<td>0.6</td>
<td>0.7</td>
<td>0.4</td>
<td>0.7</td>
<td>1.0</td>
<td>0.7</td>
<td>0.5</td>
<td>0.4</td>
</tr>
<tr>
<td>Gm</td>
<td>0.2</td>
<td>0.3</td>
<td>0.4</td>
<td>0.4</td>
<td>0.5</td>
<td>0.6</td>
<td>0.5</td>
<td>0.6</td>
<td>0.7</td>
<td>1.0</td>
<td>0.7</td>
<td>0.7</td>
</tr>
<tr>
<td>Am</td>
<td>0.4</td>
<td>0.4</td>
<td>0.7</td>
<td>0.5</td>
<td>0.5</td>
<td>0.7</td>
<td>0.6</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
<td>1.0</td>
<td>0.7</td>
</tr>
<tr>
<td>m3C</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.4</td>
<td>0.6</td>
<td>0.7</td>
<td>0.4</td>
<td>0.5</td>
<td>0.6</td>
<td>0.7</td>
<td>0.7</td>
<td>1.0</td>
</tr>
</tbody>
</table>

High correlations:

- Triplet: (m1A, m5C, m3C) 0.97; (Am, Cm, Gm) 0.95
- Pairs: (ac4C, m1G) 0.94; (Gm, Um) 0.93; (Am, m6A) 0.92
PCA: 3-dimensional (3D) view of PCA

PCA using full epitranscriptomic profile: colors = grades.

GBM - 3D PCA masses normalised on SUM
Supervised Machine Learning approach

- Two steps: Learning + Evaluation; then prediction
- Different algorithms: Support Vector Machine, Random Forest, LDA, etc.
Prediction with full epitranscriptomic profile (I)

With Linear Discriminant Analysis (LDA)

<table>
<thead>
<tr>
<th>Grade</th>
<th>precision</th>
<th>recall</th>
<th>f1-score</th>
<th>support</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade-II</td>
<td>0.67</td>
<td>1.00</td>
<td>0.80</td>
<td>2</td>
</tr>
<tr>
<td>Grade-III</td>
<td>1.00</td>
<td>0.83</td>
<td>0.91</td>
<td>6</td>
</tr>
<tr>
<td>Grade-IV</td>
<td>0.88</td>
<td>1.00</td>
<td>0.93</td>
<td>7</td>
</tr>
<tr>
<td>Ctrl</td>
<td>1.00</td>
<td>0.80</td>
<td>0.89</td>
<td>5</td>
</tr>
<tr>
<td>weighted avg accuracy</td>
<td>0.92</td>
<td>0.90</td>
<td>0.90</td>
<td>20</td>
</tr>
</tbody>
</table>

Table: Results for grade prediction with full profiles using LDA
Selection of most informative nucleosides for grading

Use Recursive Feature Elimination on a SVM (linear kernel):

▶ a profile with 9 nucleosides suffice for grade prediction
▶ one automatic selection yields
{Cm, Psi, Q, Um, m1G, m227G, m5C, m66A, m6Am}
1. State-of-the-art MS/MS quantifies modified nucleosides from biopsy tissue
2. Levels of some modified nuc. vary with diffuse glioma grades
3. Full epitranscriptomic profile allows grade prediction with 90% accuracy using machine learning approaches
4. but single nuc quantities do not!
5. Feature selection exhibits a selection of 9 nucleosides, which is sufficient for accurate prediction.

doi:10.1021/acs.analchem.2c01526
Advantages & Future work

**Advantages:**

- requires little material: 10 µg
- lasts less than 6 hours
- cheap

**Future:**

- Link to known influent mutations: IDH1/2, MGMT, BRAF, ATRX, EGFR, TERT
- Patient stratification
- Analysis from blood samples, or other fluids.
- Other cancers: breast, colorectal, pancreas.
Advantages:

- requires little material: 10 µg
- lasts less than 6 hours
- cheap

Future:

- Link to known influent mutations: IDH1/2, MGMT, BRAF, ATRX, EGFR, TERT
- Patient stratification
- Analysis from blood samples, or other fluids.
- Other cancers: breast, colorectal, pancreas.
Thanks for your attention
Cameron Davidson-Pilon.
lifelines: survival analysis in python.

J. D. Hunter.
Matplotlib: A 2d graphics environment.

Naturally occurring modified ribonucleosides.

Quinn T Ostrom, Mackenzie Price, Corey Neff, Gino Cioffi, Kristin A Waite, Carol Kruchko, and JillS Barnholtz-Sloan.

Scikit-learn: Machine learning in Python.


Evaluation metrics

Recall and precision

\[
\text{recall} = \frac{TP}{TP + FN} \quad \text{precision} = \frac{TP}{TP + FP}
\]

F1-score

a harmonic mean of the precision and recall (best value is 1, worse is 0).

\[
F1\text{score} = \frac{2 \times (\text{precision} \times \text{recall})}{\text{precision} + \text{recall}}
\]

Accuracy

is the fraction of correct predictions over all tested samples
Principal Component Analysis: % of explained variance

First 3 components capture 69 % of variance.