Analysis of epitranscriptome for grading of glioma

Eric Rivals

Analysis of epitranscriptome for grading of glioma

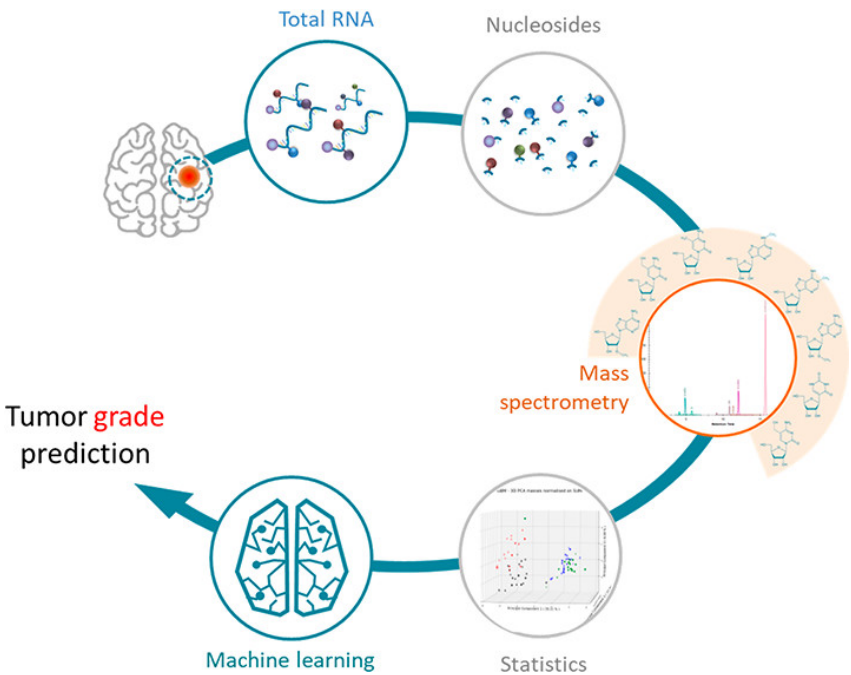
Eric Rivals

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



[7]

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

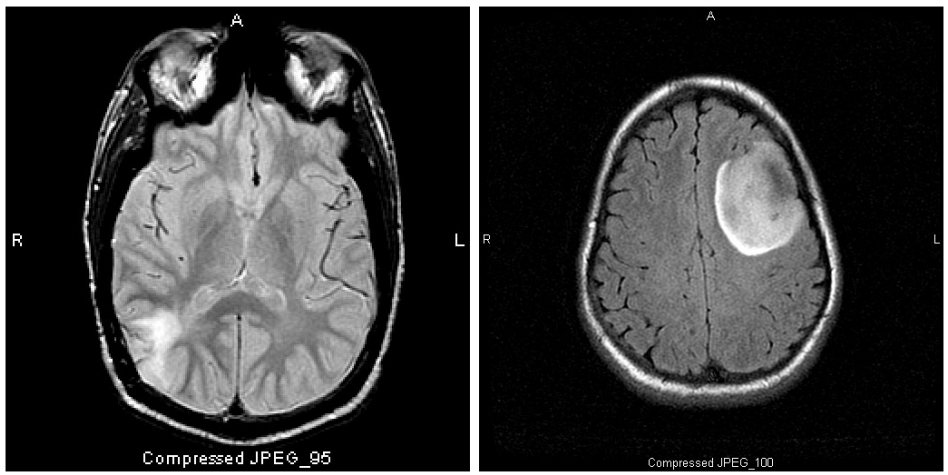


Figure: **Left**: MRI of a grade II oligodendroglioma. **Right**: MRI of a grade III anaplasic astrocytoma. Source: Manuel MSD, Steven A. Goldman

Cohort

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.)

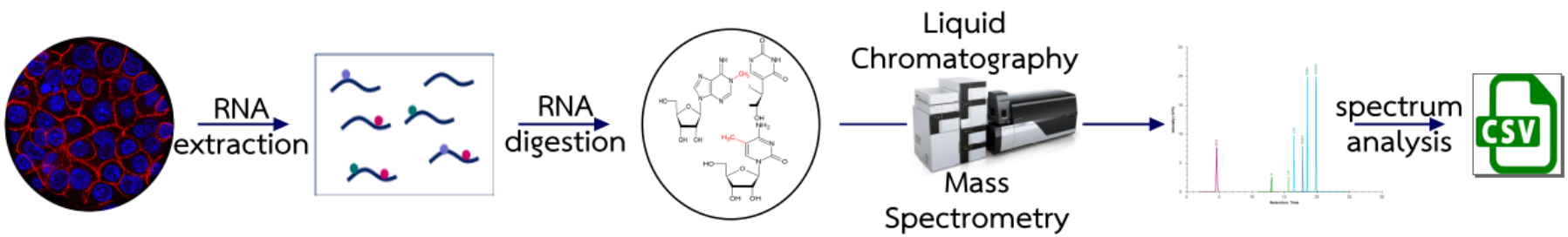
Distribution

type	grade-II	grade-III	grade-IV	control
name	glioma	glioma	glioblastoma	non-tumor
number	20	20	18	19

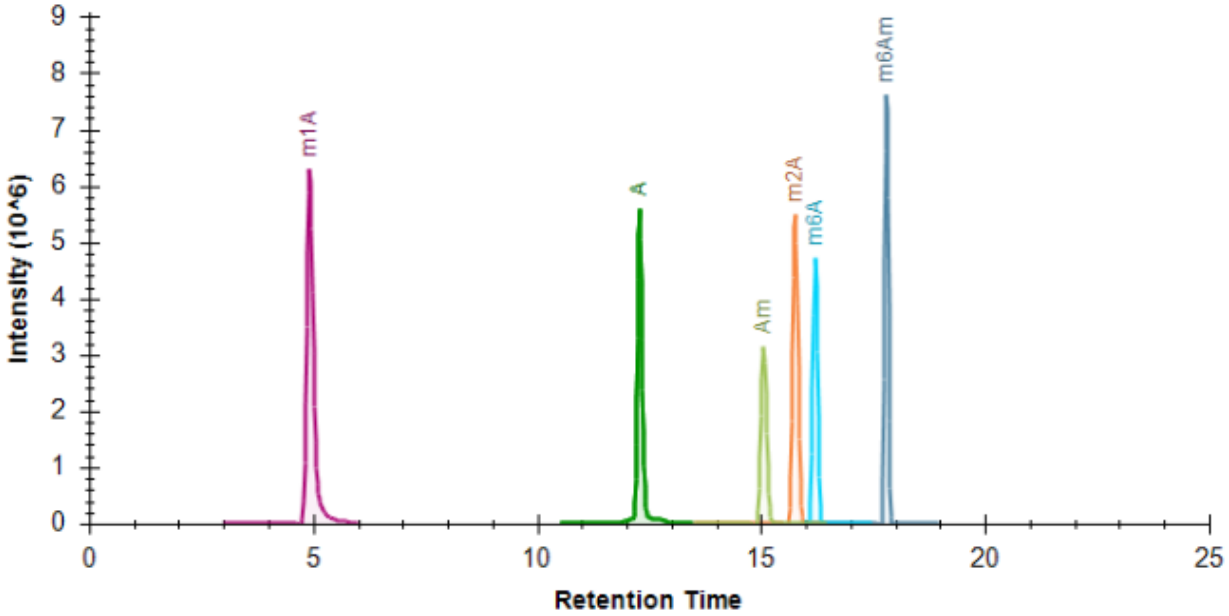
Methods - sample analysis

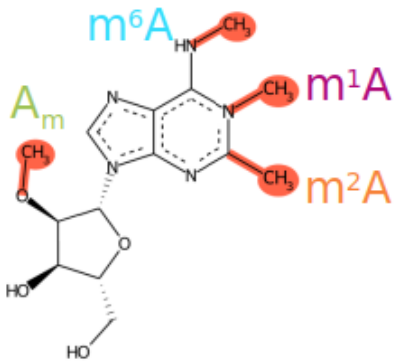
From tissue to data file:

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





List of quantified modified nucleosides (25)

Standard	modified nucleosides
A	Am m1A m66A m66Am m6A m6Am I
C	Cm ac4C m3C m5C hm5C
G	Gm m1G m227G m27G m7G oxo8G
U	Um m3Um mcm5U mcm5s2U ncm5U Psi
	Queuosine

Data analysis and machine learning

For each sample, 29 MS measures for modified and standard nucleosides.
Math: a vector of 29 dimensions $\Rightarrow$ epitranscriptomic profile

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?

Methods

Python scripts using libraries Scikit-learn [5], Lifelines [1], Pandas [6], Matplotlib [2], Seaborn [8]

Nucleosides with decreasing quantities wrt cancer grade

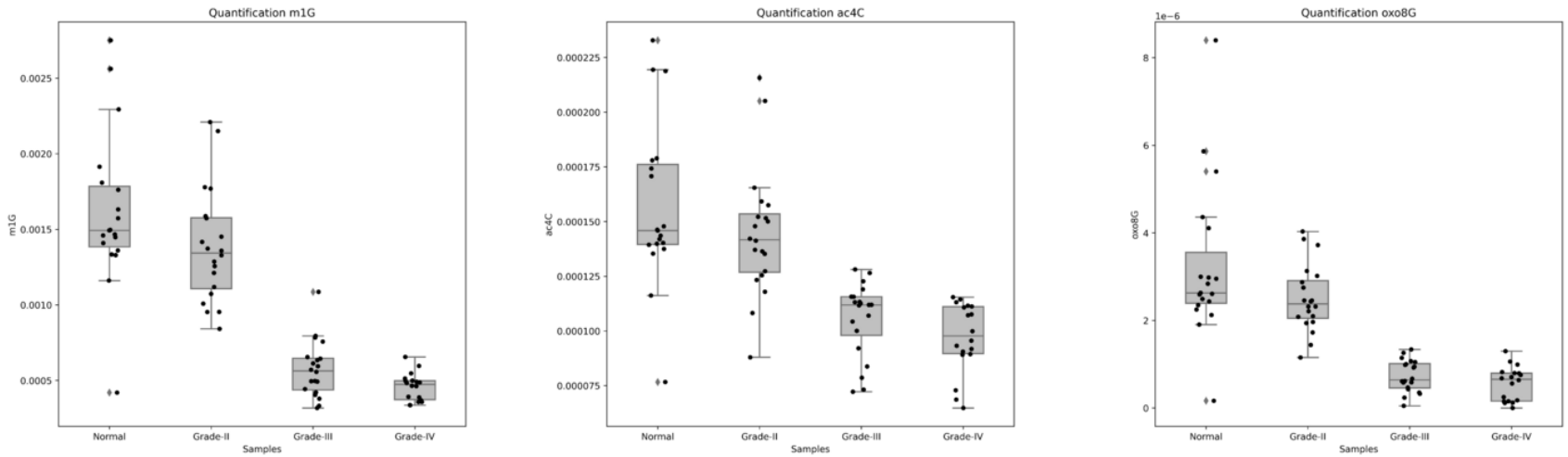


Figure: Boxplots wrt grade of m1G, ac4C, and oxo8G.

Nucleosides with increasing quantities wrt cancer grade

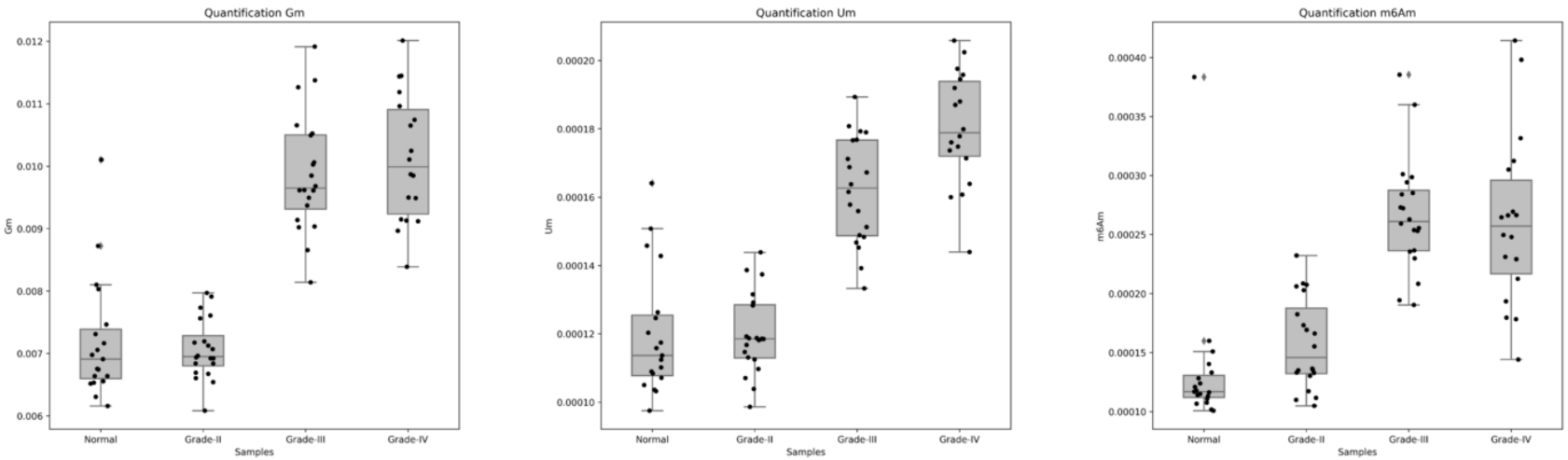
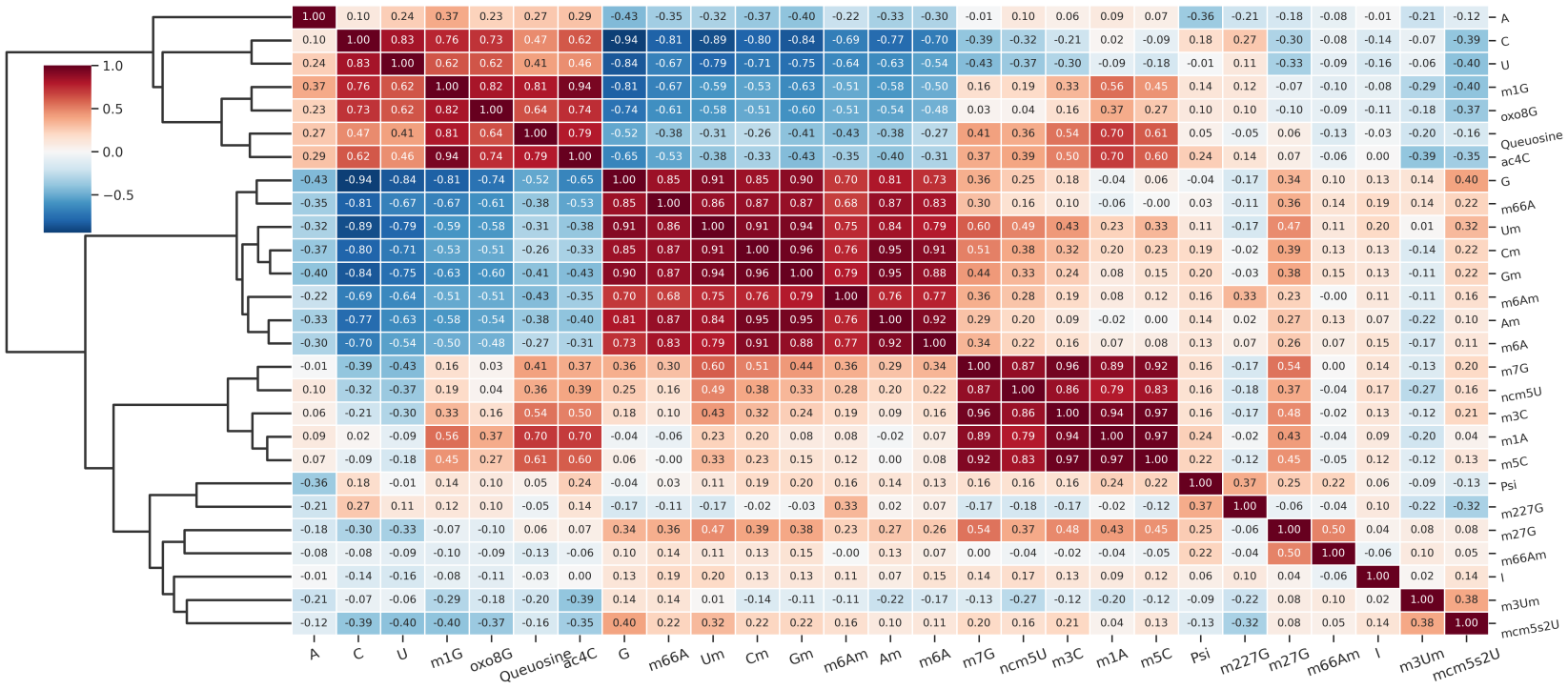


Figure: Boxplots wrt grade of Gm, Um, and m6Am.

Correlation between nucleoside quantities

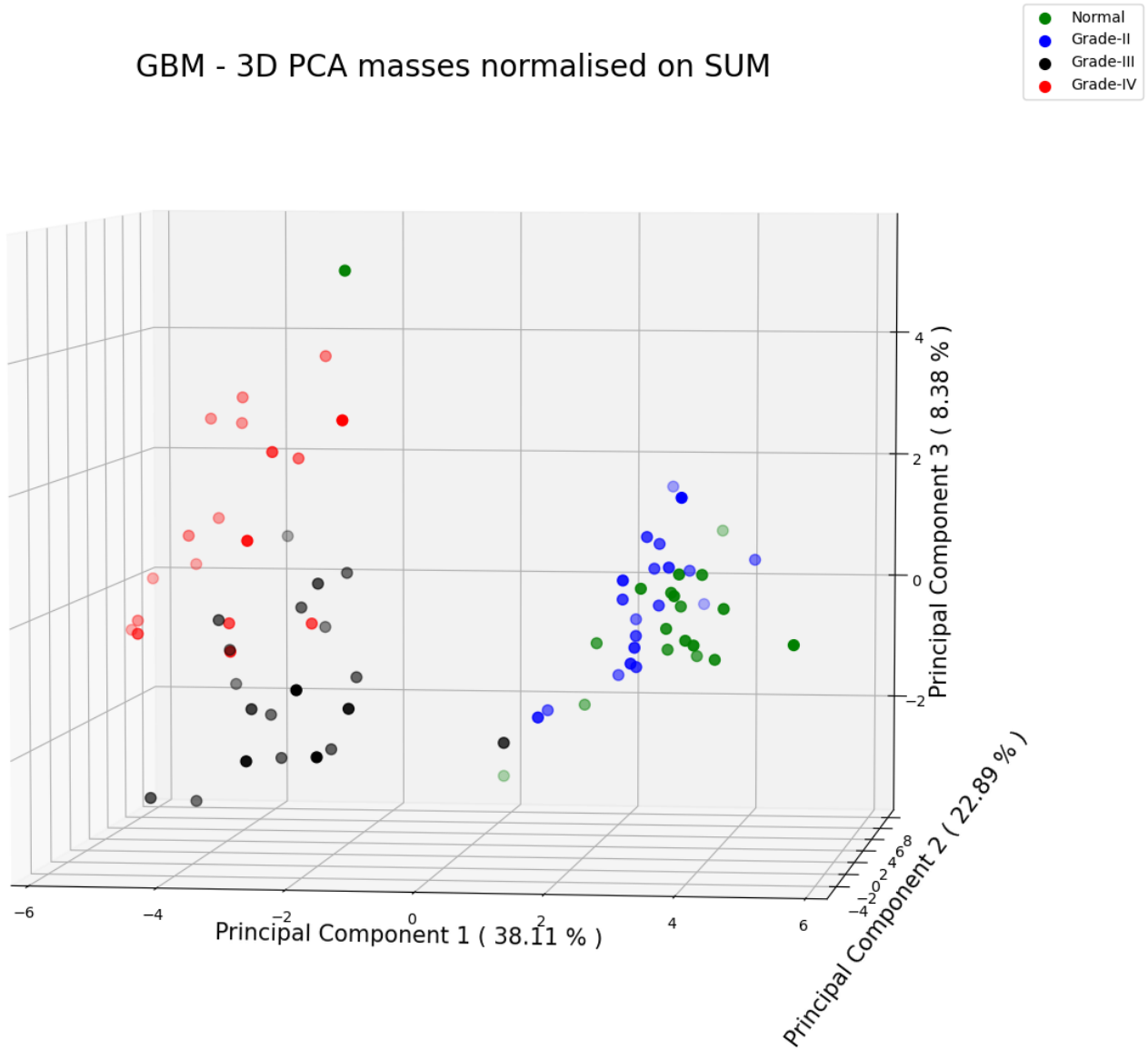


High correlations:

- Triplet: (m1A, m5C, m3C) 0.97; (Am, Cm, Gm) 0.95
- Pairs: (ac4C, m1G) 0.94; (Gm, Um) 0.93; (Am, m6A) 0.92

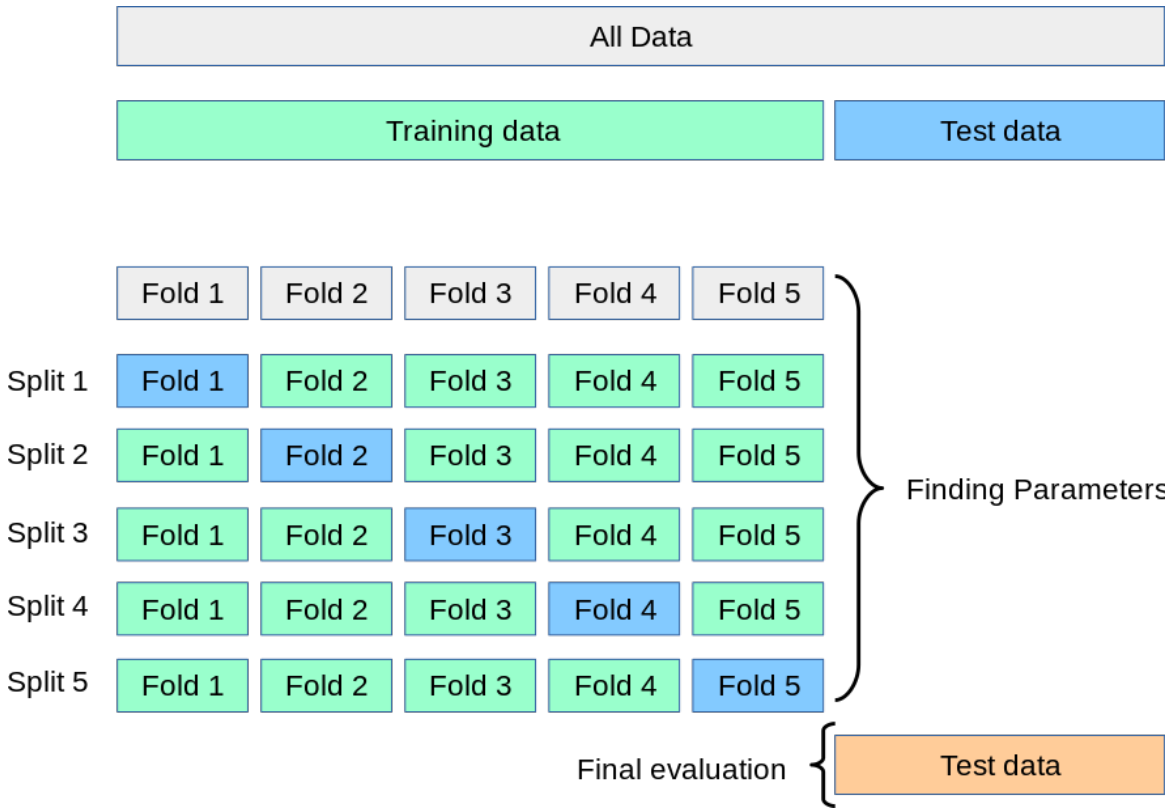
PCA: 3-dimensional (3D) view of PCA

PCA using full epitranscriptomic profile: colors = grades.



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)

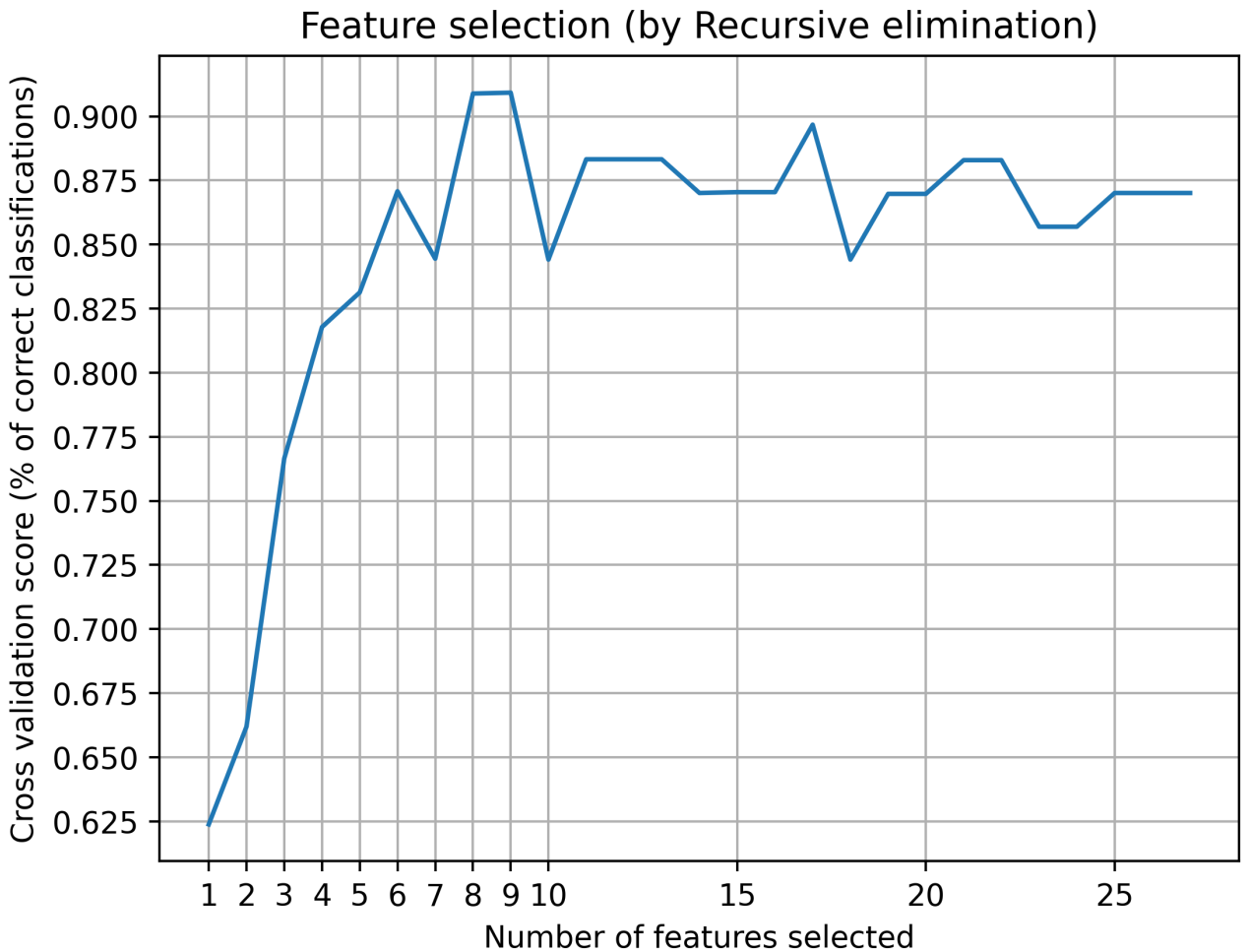
Table: Results for grade prediction with full profiles using LDA

Grade	precision	recall	f1-score	support
Grade-II	0.67	1.00	0.80	2
Grade-III	1.00	0.83	0.91	6
Grade-IV	0.88	1.00	0.93	7
Ctrl	1.00	0.80	0.89	5
weighted avg	0.92	0.90	0.90	20
accuracy			**0.90**	

Selection of most informative nucleosides for grading

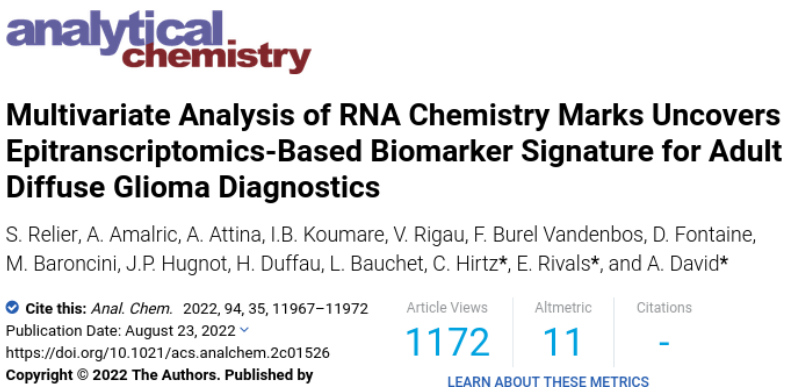
Use Recursive Feature Elimination on a SVM (linear kernel):

- a profile with 9 nucleosides suffice for grade prediction
- one automatic selection yields
  {Cm , Psi, Q, Um, m1G, m227G, m5C, m66A, m6Am}



Conclusion – Summary

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.


analytical chemistry

Multivariate Analysis of RNA Chemistry Marks Uncovers Epitranscriptomics-Based Biomarker Signature for Adult Diffuse Glioma Diagnostics

S. Relier, A. Amalric, A. Attina, I.B. Koumare, V. Rigau, F. Burel Vandenbos, D. Fontaine, M. Baroncini, J.P. Hugnot, H. Duffau, L. Bauchet, C. Hirtz*, E. Rivals*, and A. David*

Cite this: *Anal. Chem.* 2022, 94, 35, 11967–11972
Publication Date: August 23, 2022
https://doi.org/10.1021/acs.analchem.2c01526
Copyright © 2022 The Authors. Published by

Article Views 1172 | Altmetric 11 | Citations -

LEARN ABOUT THESE METRICS

doi:10.1021/acs.analchem.2c01526

Advantages & Future work

Advantages:

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

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.

Thanks for your attention

IGF
INSTITUT DE GÉNOMIQUE FONCTIONNELLE
Sébastien Relier
Amandine Bastide
Amandine Amalric
Hélène Guillorit
Jean-P. Hugnot

LIRMM
MART
CHU MONTPELLIER
CHU LILLE
Luc Bauchet
I.B. Koumare, V. Rigau, H. Duffau
F. Burel Vandenbos, D. Fontaine, M. Baroncini
ATGC

PPC
Christophe Hirtz
Jérôme Vialaret
Aurore Attina

IBMM
Jean-Jacques Vasseur
Françoise Debart

amabiotics
advanced microbiome medicine
SIRIC MONTPELLIER CANCER
Site de Recherche Intégrée sur le Cancer
INSTITUT NATIONAL DU CANCER
LA LIGUE CONTRE LE CANCER
anrt association nationale recherche technologie
L'EUROPE S'ENGAGE L'OCCITANIE AGIT
UNION EUROPÉENNE
La Région Occitanie Pyrénées · Méditerranée
Projet cofinancé par le Fonds Européen de Développement Régional

Bibliographic references I


Cameron Davidson-Pilon.
lifelines: survival analysis in python.
Journal of Open Source Software, 4(40):1317, 2019.


J. D. Hunter.
Matplotlib: A 2d graphics environment.
Computing in Science & Engineering, 9(3):90–95, 2007.


Phillip J. McCown, Agnieszka Ruszkowska, Charlotte N. Kunkler, Kurtis Breger, Jacob P. Hulewicz, Matthew C. Wang, Noah A. Springer, and Jessica A. Brown.
Naturally occurring modified ribonucleosides.
WIREs RNA, 11(5), Apr 2020.


Quinn T Ostrom, Mackenzie Price, Corey Neff, Gino Cioffi, Kristin A Waite, Carol Kruchko, and JillS Barnholtz-Sloan.
CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2015-2019.
Neuro-Oncology, 24(Supp. 5):v1–v95, Oct 2022.

Bibliographic references II

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay.
Scikit-learn: Machine learning in Python.
Journal of Machine Learning Research, 12:2825–2830, 2011.

Jeff Reback, Wes McKinney, jbrockmendel, Joris Van den Bossche, Tom Augspurger, Phillip Cloud, gfyoung, Sinhrks, Adam Klein, Matthew Roeschke, Simon Hawkins, Jeff Tratner, Chang She, William Ayd, Terji Petersen, Marc Garcia, Jeremy Schendel, Andy Hayden, MomIsBestFriend, Vytautas Jancauskas, Pietro Battiston, Skipper Seabold, chris b1, h vetinari, Stephan Hoyer, Wouter Overmeire, alimcmaster1, Kaiqi Dong, Christopher Whelan, and Mortada Mehyar.
pandas-dev/pandas: Pandas 1.0.3, March 2020.

Bibliographic references III

S. Relier, A. Amalric, A. Attina, I.B. Koumare, V. Rigau, F. Burel Vandenbos, D. Fontaine, M. Baroncini, J.P. Hugnot, H. Duffau, L. Bauchet, C. Hirtz, E. Rivals, and A. David.
Multivariate analysis of rna chemistry marks uncovers epitranscriptomics-based biomarker signature for adult diffuse glioma diagnostics.
Analytical Chemistry, Aug 2022.

Michael L. Waskom.
seaborn: statistical data visualization.
Journal of Open Source Software, 6(60):3021, 2021.

Evaluation metrics

Recall and precision

$$\text{recall} = TP/(TP+FN) \qquad \text{precision} = TP/(TP+FP)$$

F1-score

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

$$F1score = 2*(precision*recall)/(precision+recall)$$

Accuracy

is the fraction of correct predictions over all tested samples

Principal Component Analysis: % of explained variance

First 3 components capture 69 % of variance.