



HAL
open science

Exploring the metabolic effects of different genotoxic compounds from transcriptomics data: using gene selection to unravel significant effects

Patricia KEMBIA KALOMBO, Nathalie Poupin, Cécile Canlet, Roselyne Gautier, Yannick Lippi, Gladys Mirey, Marie Tremblay-Franco

► To cite this version:

Patricia KEMBIA KALOMBO, Nathalie Poupin, Cécile Canlet, Roselyne Gautier, Yannick Lippi, et al.. Exploring the metabolic effects of different genotoxic compounds from transcriptomics data: using gene selection to unravel significant effects. Journées Ouvertes Biologie Informatique Mathématiques (JOBIM), Jun 2024, Toulouse, France. 10.1126/science.1136800 . hal-04635933

HAL Id: hal-04635933

<https://hal.science/hal-04635933v1>

Submitted on 25 Jul 2024

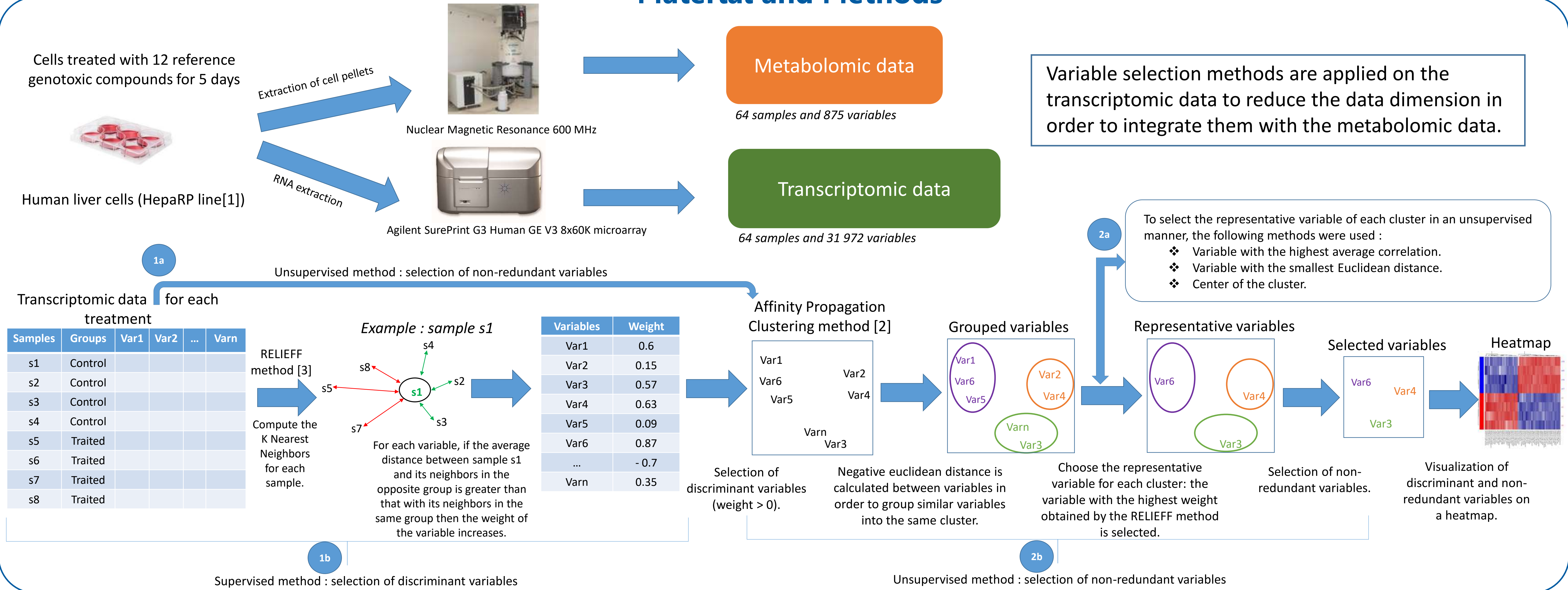
HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

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.

Context

- **ANR GENOSHIFT** : toxic agents are a major risk for human/animal health and biosystems, to understand their modes of action, a multi-omic study is conducted to identify biomarkers of exposure following chronic exposure of human liver cells to 12 reference genotoxic compounds at low doses.
- The differential gene analysis revealed Differentially Expressed Genes (DEGs) for some of these 12 molecules and no DEGs for others. But, variable selection highlighted a transcriptomic effect for these latter compounds.
- Statistical integration of metabolomic and transcriptomic data: identify genes and metabolites correlated and modulated by chronic exposure of human liver cells to 12 genotoxic compounds.
 - Selection of the most "informative" variables for statistical integration by removing non-informative, redundant and noisy variables in transcriptomic data using the following methods:
 - ❑ **Affinity Propagation Clustering**[2] (apcluster R package), unsupervised method of selection by grouping similar variables: allows to remove redundancy.
 - ❑ **RELIEFF**[3] (mt R package), supervised method of selecting discriminant variables.

Material and Methods



Results and Discussions

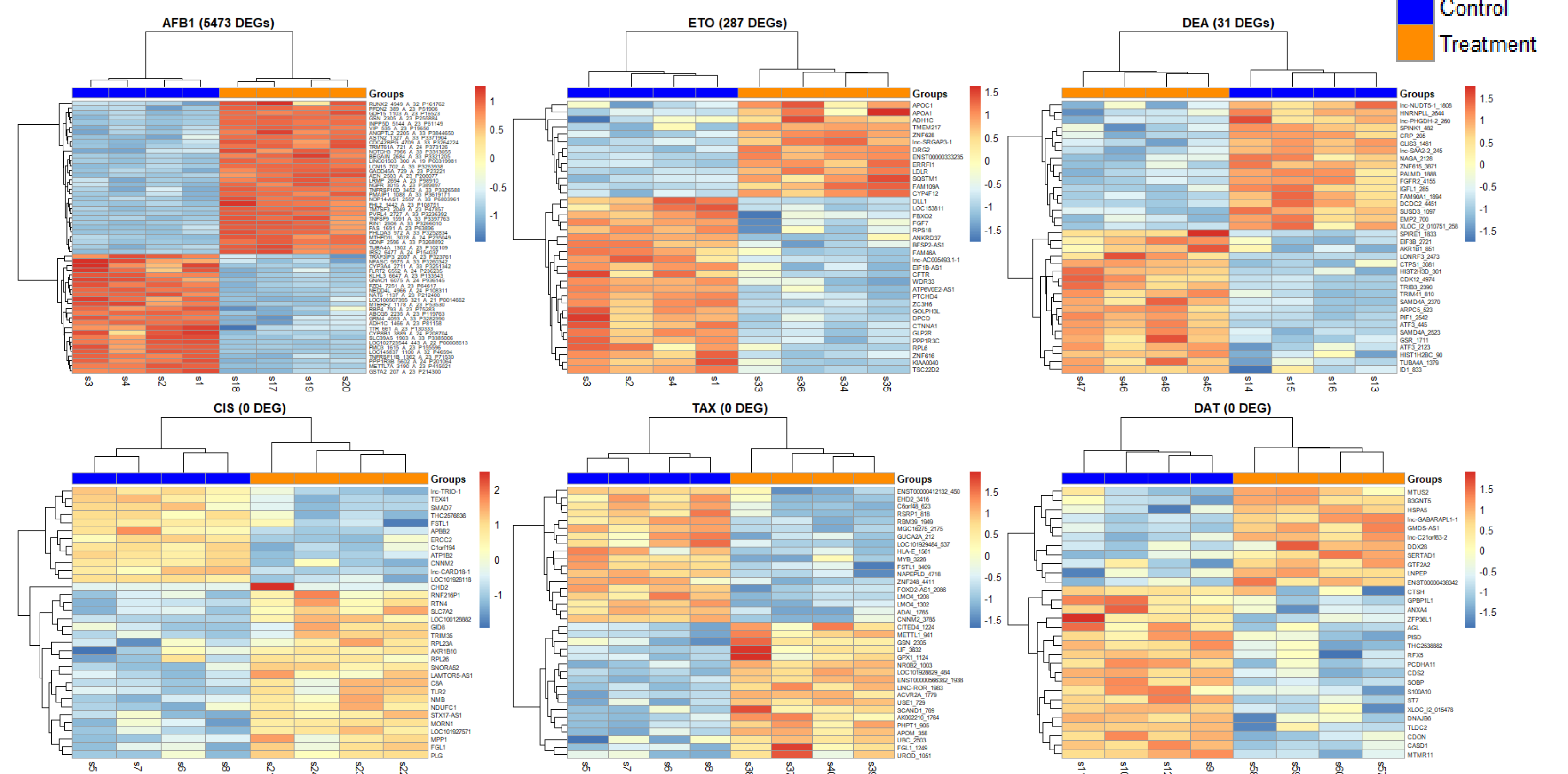
Affinity Propagation Clustering

| Compounds vs Control | Number of selected variables |
|----------------------|------------------------------|
| AFB1 vs Control | 1456 |
| ETO vs Control | 1405 |
| CIS vs Control | 1232 |
| TAX vs Control | 1278 |
| DAT vs Control | 1185 |
| DEA vs Control | 1255 |

Number of non-redundant variables selected by Affinity Propagation clustering: each line indicates the number of variables selected for a compound compared to its control.

- ❖ Affinity Propagation clustering[2] allowed to remove redundancy between variables and thus reduce the dimension of transcriptomic data.
- ❖ The table indicates the number of variables selected for each compound: approximately one thousand variables were selected to have a number of variables of the order of those of metabolomic data for data integration.

RELIEFF method combined with Affinity Propagation clustering



Heatmap of discriminant and non-redundant variables selected by the RELIEFF method combined with Affinity Propagation clustering. In line, there are the selected variables and in column, the control samples (in blue) and treated samples (in orange).

- ❖ The RELIEFF [3] method combined with Affinity Propagation clustering [2] selected discriminant and non-redundant variables, particularly for compounds where there was no differentially expressed gene (e.g. CIS, TAX, DAT).
- ❖ We can see on the heatmaps that the selected variables have a differential expression between the control vs treatment groups.

Conclusions and Perspectives

- ✓ These results show that the RELIEFF method combined with Affinity Propagation clustering is able to select discriminant and non-redundant genes, particularly for compounds for which there is a no clear transcriptomic effect (e.g., CIS).
- ✓ The unsupervised method (Affinity Propagation clustering) reduced the size of the data by selecting around 3% of variables.
- ✓ The selected variables will be used for statistical integration with the metabolomic variables.

Acknowledgements

[1] ANR grant Genoshift (ANR-20-CE34-0016)

References

[1] Brun C., Allain C., Ferron P.J., Younoussa H., Colicchio B., Jeandidier E., M'Kacher R., Guguen-Guillouzo C., Bertile F. Extended lifespan and improved genome stability in HepaRG-derived cell lines through reprogramming by high-density stress. Proc Natl Acad Sci. 2023 Sep 5;120(36):e2219298120.

[2] Frey B. J. and Dueck D. Clustering by passing messages between data points. Science 315, 972-976. DOI: doi:10.1126/science.1136800. 2007 Feb 16;315(5814):972-6.

[3] Kononenko I., Simes E. and Robnik-Sikonja M. Overcoming the Myopia of Induction Learning Algorithms with RELIEFF. Applied Intelligence, Vol.7, 1, 39-55. 1997.