Multiomics integration of transcriptomics, proteomics, and lipidomics for toxicological assessment

Alain Sewer, Björn Titz, Florian Martin, Manuel C. Peitsch, and Julia Hoeng

PMI R&D, Philip Morris Products S.A., Quai Jeanrenaud 5, CH-2000 Neuchâtel, Switzerland (Part of Philip Morris International group of companies)

Introduction and Objectives

The ability to efficiently disentangle the relevant information contained in multi-omics datasets has become an essential aspect of Systems Toxicology. While the complexity of the data generated during the toxicological assessment studies has increased, the analysis outcome has to remain quantitative and biologically interpretable. We illustrate such a multi-omics analysis using transcriptomics, proteomics, and lipidomics data from a published study **[1]**. Its aim was to investigate general exposure effects over eight months in ApoE$^{-/-}$ mice exposed to conventional cigarette smoke (CS) and aerosol from a heat-not-burn tobacco product, and to assess the consequences of smoking cessation or switching to a heat-not-burn tobacco product after two months of exposure to cigarette smoke.

We applied sparse partial least squares correlation analysis that performs a L1-penalized multivariate complexity reduction scheme to extract relevant directions that correlate between data modalities. The first identified component captured the cigarette smoke exposure effect, while not distinguishing the other treatments. The loadings, which define this direction, were subsequently annotated using functional association clustering to enable the biological interpretation. The identified clusters included lipid metabolism, oxidative stress, and inflammation processes – all positively associated with cigarette smoke exposure. Notably, this analysis showed a concordant induction of lipid and protein components of the lung surfactants upon cigarette smoke exposure.

By identifying biological mechanisms that are relevant across data modalities, our approach supports a holistic interpretation of multiomics experiments and provides the basis for the quantitative assessment of toxicologically relevant mechanisms.

Material and Methods

Toxicology study for product assessment

ApoE-/- mice were exposed to mainstream aerosols for 5 days a week, 3 hours per day at a target nicotine concentration of 30 µg/L. The five exposure groups and durations are detailed in Fig. 1: Control (sham), Cigarette Smoke (3R4F), heat-not-burn tobacco product (THS 2.2), cessation, switching to heat-not-burn tobacco product (THS 2.2). Groups were composed of 8 animals **[2]**.

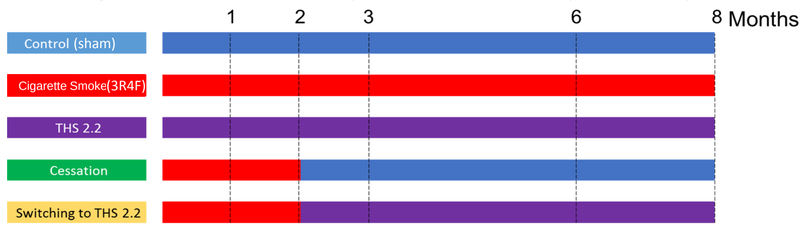

Figure 1: Experimental design. The biological material was collected at the indicated timepoints and the relevant effects are obtained by performing the time-matched "treatment vs. control" comparisons.

Generation of transcriptomics, proteomics, and lipidomics data

The relevant molecular species were extracted from the collected biological material and processed through their respective workflows to quantify their abundance (Fig. 2) **[3,4,5]**.

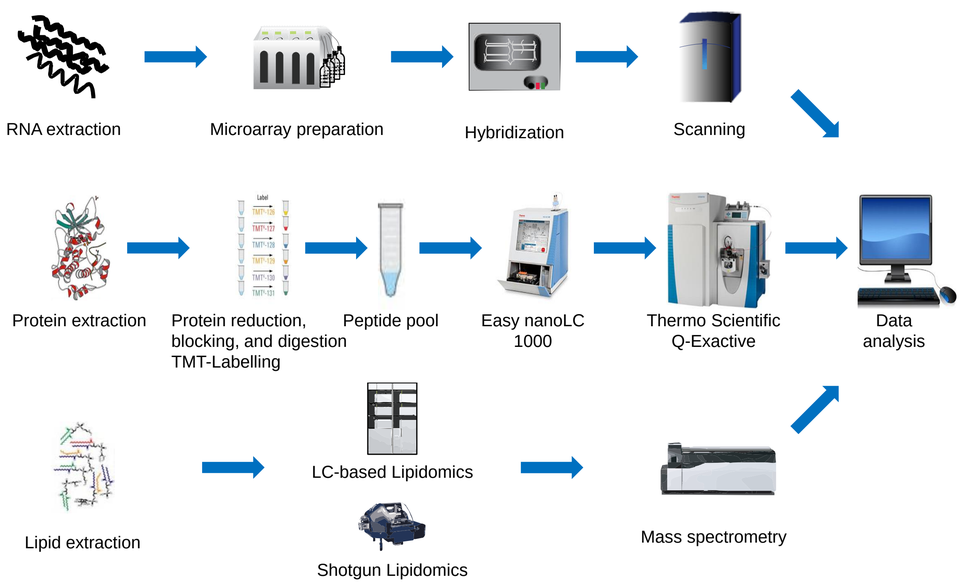

Figure 2: Workflows for the generation of transcriptomics, proteomics, and lipidomics data. All the data used in this study have been deposited in public repositories [1].

Results ("Singleomics")

The quantified molecular responses to the various exposure conditions showed consistent patterns across the three data modalities in ApoE-/- mouse lungs (Fig. 3). A large number of molecules were significantly affected by CS exposure (thousands of probed mRNAs, hundreds of identified proteins and lipids), whereas the effects observed for the other exposure groups (exposed to aerosol from the THS2.2 heat-not-burn tobacco product, cessation and THS2.2 switching) were much more limited.

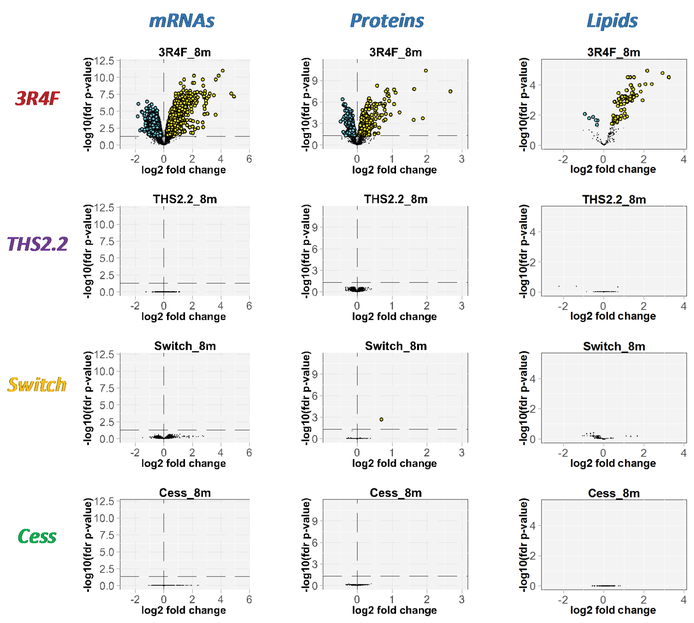

*Figure 3: Volcano plots provide a global view of the molecular-level response. The x-axis represents the amplitude of the response ("differential expression") while the y-axis quantifies its statistical significance ("-log10 false discovery rate"). For space reasons, only the 8-month time point is shown (see **[1]** and **[2]** for data from all time points).*

Results ("Multiomics")

The ApoE-/- mouse lung proteome and lipidome both showed a broad response to CS exposure. Preliminary functional analysis of the single-omics results indicated that CS exposure induced alterations in lipid metabolism pathways and multiple lipid classes. In order to analyze these mechanisms in an integrated manner, a multivariate complexity reduction approach named "sPLS-can" was applied to identify the directions ("components") that best correlate between to the matched proteomics and lipidomics datasets **[6]**. Expectedly, the CS-exposed lung samples were all captured in the first component (Fig. 4A). The proteins contributing to the proteomics first component were subsequently subjected to a functional clustering analysis that revealed their associations with several lipid metabolism, oxidative stress, and inflammation processes (Fig. 4B) **[7]**.

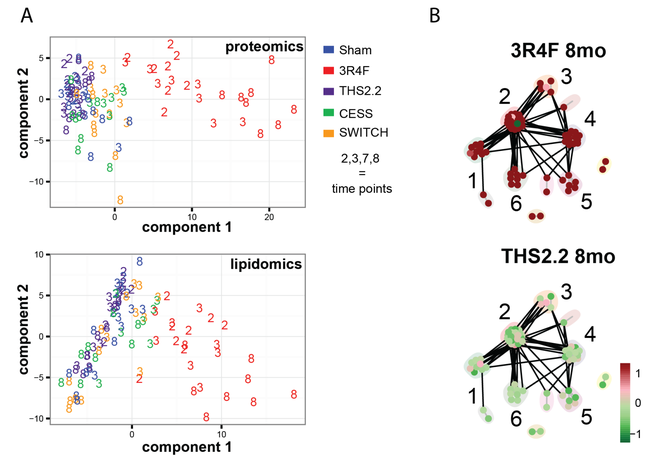

Figure 4. Panel A: sPLS-can main components in terms of samples (see Figure 1). "sPLS-can" stands for sparse partial least square canonical analysis. sPLS-can is related to principal component analysis. However, rather than identifying the components that best explain the variance of individual datasets, sPLS-can identifies the components in the data that best correlate between the two datasets. Panel B: Clustered functional association network for proteins positively contributing to sPLS-can component 1. The identified clusters are the following: lipid-related functions (1), components of the pentose-phosphate pathway (2), immune-related proteins (4), surfactants (5), and xenobiotic response proteins (6). The colors correspond to the (normalized) protein differential expression values obtained for the 8-month time-point.

The disentanglement of the complex global responses to CS exposure in the proteomics and lipidomics datasets revealed several mechanisms that were then examined in a more targeted manner **[1]**. This analysis included the transcriptomics data. Only the lipid metabolism-related results are discussed hereafter (clusters 1 and 5 in Figure 4B).
Quantitative proteomics results showed that CS exposure had statistically significant effects on several surfactant and surfactant metabolism proteins (Fig. 5A). These alterations in surfactant metabolism likely contribute to the CS-induced changes in lipidome profiles of surfactant lipids (Fig. 5B). The quantification of the activities of transcriptional regulators implicated in lipid metabolic processes (C/EBP, SREBP, and PPAR families) were found to be significantly upregulated in the lungs of CS-exposed Apoe−/− mice (Fig. 5C) **[3]**.

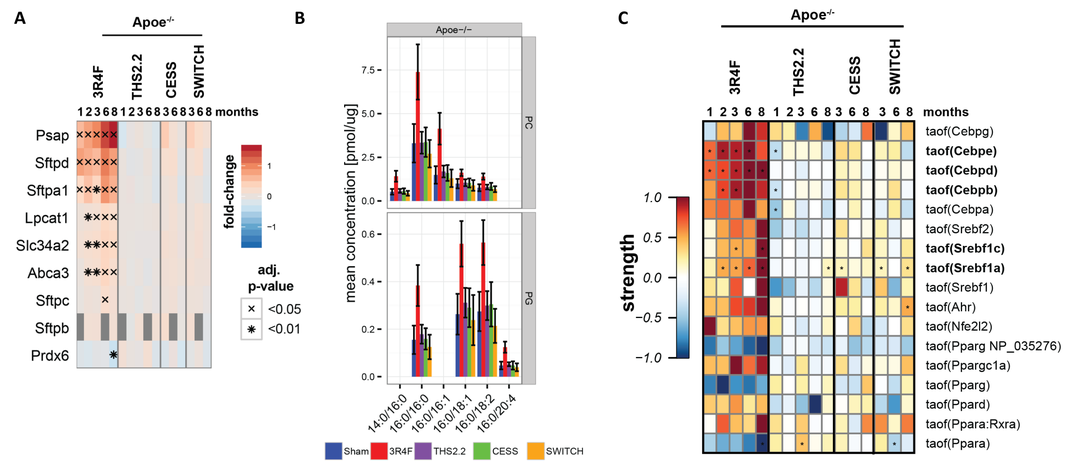

Figure 5. Panel A: Differential expression profiles for surfactant-related proteins. The fold-change is color-coded and statistical significance is marked. Panel B: Comparison of lipid concentration profiles for the surfactant-associated PC and PG lipid classes. Panel C: Transcription regulator activity quantification. The "strength" perturbation metric value is color-coded and statistical significance is indicated.

Conclusions

- The integration of lung transcriptomics, proteomics, and lipidomics data enabled to assess the complex effects of CS exposure and revealed a complex response in terms of lipid-related processes (Fig. 6).
- Overall, CS exposure resulted in extensive lipid metabolism changes, potentially associated with adverse effects on the lung, including the observed emphysematous changes and increased inflammation **[2]**.
- By identifying biological mechanisms that are consistent across data modalities, this integrative approach supports a holistic interpretation of multiomics experiments and provides the basis for the quantitative assessment of toxicologically relevant mechanisms.

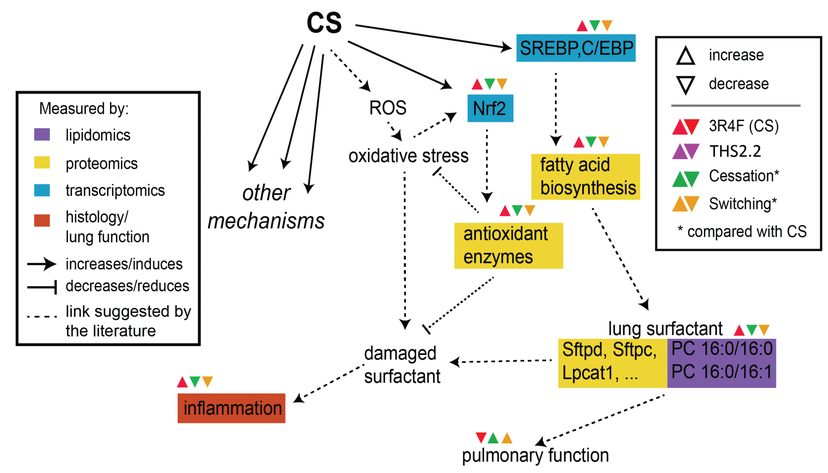

Figure 6: Schematic network for the interplay among proteins and lipids in CS-induced lung response. The node colors indicate the source data used for quantifying the lung responses to the various exposure treatments.

References

[1] Titz *et al.*, Effects of Cigarette Smoke, Cessation, and Switching to Two Heat-Not-Burn Tobacco Products on Lung Lipid Metabolism in C57BL/6 and Apoe-/- Mice, Toxicol Sci. 2016.
[2] Philipps *et al.*, An 8-Month Systems Toxicology Inhalation/Cessation Study in Apoe-/- Mice to Investigate Cardiovascular and Respiratory Exposure Effects of a Candidate Modified Risk Tobacco Product, THS 2.2, Compared With Conventional Cigarettes, Toxicol Sci. 2016.
[3] Martin *et al.*, Assessment of network perturbation amplitudes by applying high-throughput data to causal biological networks, BMC Syst Biol. 2012.
[4] Titz *et al.*, Proteomics for systems toxicology, Comput Struct Biotechnol J. 2014.
[5] Boué *et al.*, Modulation of atherogenic lipidome by cigarette smoke in apolipoprotein E-deficient mice, Atherosclerosis 2012.
[6] Lê Cao *et al.*, Sparse PLS discriminant analysis: biologically relevant feature selection and graphical displays for multiclass problems, BMC Bioinformatics 2011.
[7] Chen *et al.*, ToppGene Suite for gene list enrichment analysis and candidate gene prioritization, Nucleic Acids Res. 2009.

PMI SCIENCE
PHILIP MORRIS INTERNATIONAL

EMBL Heidelberg, Germany
12 - 14 Nov 2017

Competing Financial Interest

All authors are employees of Philip Morris International (PMI). PMI is the sole source of funding and sponsor of this project.