Co-funded by the European Union

MINISTRY OF EDUCATION, YOUTH AND SPORTS

MUNI | RECETOX

ENVIRONMENTAL GENOMICS

MULTI-OMIC INTEGRATION OF FOUR DATASETS: INSIGHTS INTO BARRETT'S ESOPHAGUS AND ESOPHAGEAL CARCINOMA

Jan Böhm, Petra Bořilová Linhartová *et al.*

RECETOX, Faculty of Science, Masaryk University, Kotlarska 2, Brno, Czech Republic

BACKGROUND

Due to long-term **gastric acid** reflux, **Barrett's esophagus** manifests as a metaplastic change in the esophageal lining, transitioning towards intestinal-like epithelium.

Barrett's esophagus is often progressing to **esophageal adenocarcinoma**, a malignant cancer of the esophagus.

In this study, samples were obtained from 12 patients diagnosed with Barrett's esophagus, comprising both **pathological tissue** (denoted **BE**) and **adjacent tissue** without pathology (**BEadj**).

We have obtained samples from 12 patients with esophageal carcinoma, both pathological tissue (**EAC**) and adjacent tissue without pathology (**EACadj**).

AIMS

Comprehensive Analysis Aim: To integrate multiple omic datasets, including Whole Exome Sequencing (WES), transcriptomics, metatranscriptomics and metagenomics (16S rRNA sequencing) to achieve a more comprehensive understanding of the differences between BE/EAC pathological tissues and adjacent esophageal tissues.

Discriminative Power Aim: To compare the discriminative power of features from different omic datasets.

Biomarker Identification Aim: To identify specific features, either alone or in combination, that can serve as effective biomarkers for differentiating between BE and EAC.

DATA PROCESSING

WHOLE EXOME SEQUENCING (WES):
⚙Retained only somatic mutations that are rare in the European population, have pathogenic potential, and possess a minimum of 100 supporting reads.
⚙Data aggregated by genes and expressed as a ratio of reads in pathological tissue (somatic) compared to leucocytes (germline).

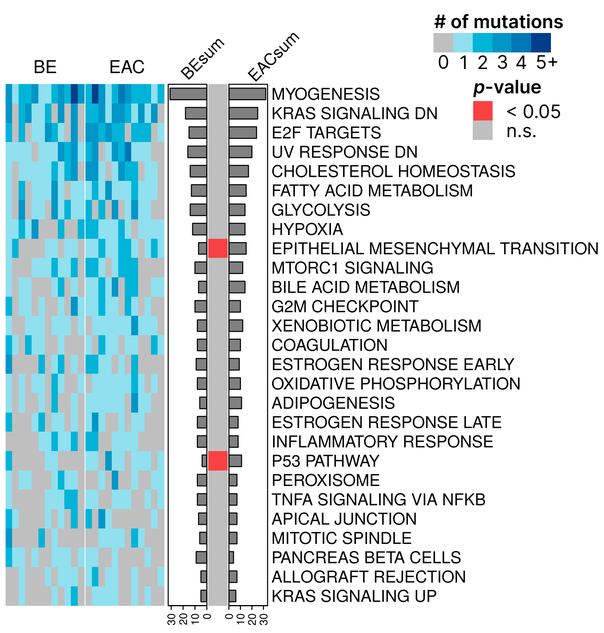
TRANSCRIPTOMICS & METATRANSCRIPTOMICS:
⚙Transformed data to counts per million (CPM); inclusion criterion set at a minimum of 10 CPM for genes.
⚙Log_2 transformation followed by quantile normalization.
⚙Metatranscriptomic data aggregated at the pathway level.

METAGENOMICS (16S rRNA SEQUENCING):
⚙Aggregated at the genus taxonomic level.
⚙Removed MOCK reads; all samples met the 5000 reads minimum inclusion criterion.
⚙Applied Total Sum Scaling (TSS) and Central Log-Ratio (CLR) transformations.

APPROACHES & RESULTS (single-omic)

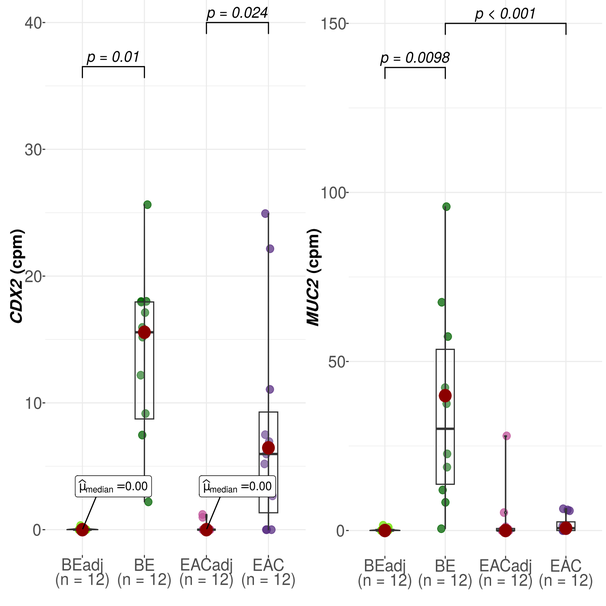
WES

Whole Exome Sequencing (WES) focuses on sequencing the protein-coding regions of the genome, offering insights into genetic variants that could influence traits or disease susceptibility.
We identified two hallmark pathways, the **P53 pathway** and **epithelial-mesenchymal transition**, that exhibit a higher mutational burden in EAC samples compared to BE.



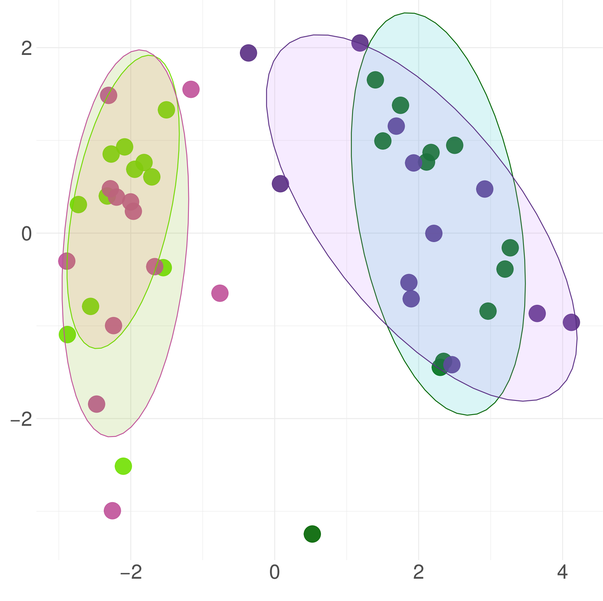
TRANSCRIPTOMICS

Transcriptomic data encompass the complete set of RNA transcripts produced by the genome, providing a comprehensive snapshot of gene expression levels in a cell or tissue at a specific time.
Distinct gene expression patterns differentiate between groups; for instance, *CDX2* **effectively distinguishes between pathological and adjacent tissue**, while *MUC2* **discriminates between BE and EAC** (Wilcoxon test).



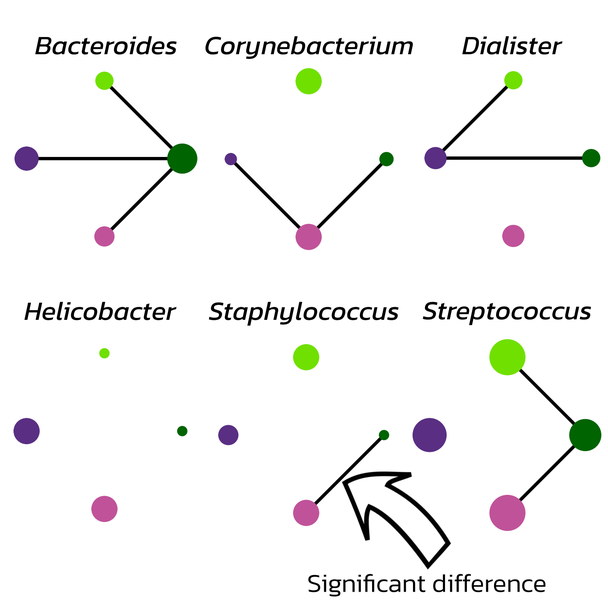
METATRANSCRIPTOMICS

Metatranscriptomic data provide insights into the RNA transcripts from esophageal microbiota, highlighting their active metabolic functions and interactions within the host environment.
Multidimensional scaling revealed that **metatranscriptomic data can distinguish between pathological and adjacent esophageal tissue**, yet they do not differentiate between BE and EAC.



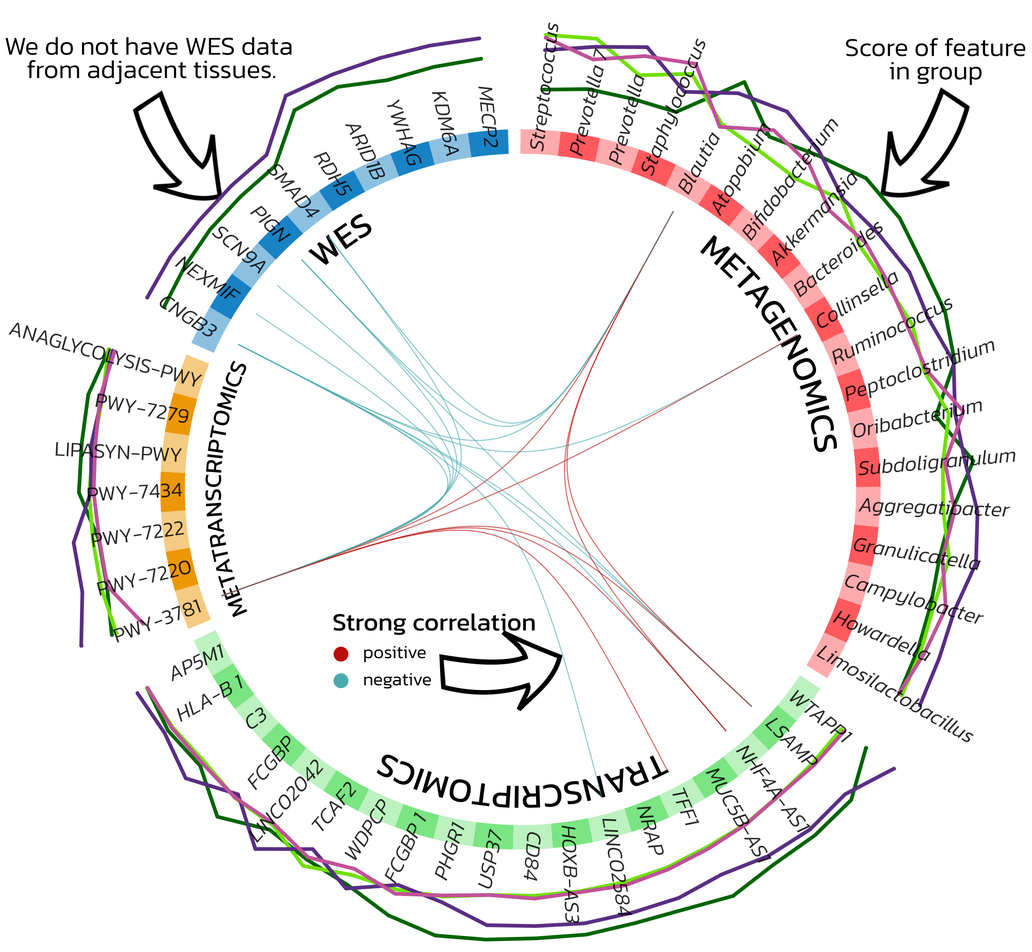
METAGENOMICS

Metagenomics data reveal the relative abundances of bacterial genera within samples.
Figure below: Discs represent the mean relative abundance of specific bacterial genera in each group. Groups connected by a line show a significant difference ($p < 0.05$, *t*-test) in their relative abundances.
For example, ***Bacteroides*** **are significantly more abundant in BE samples** compared to other groups.



RESULTS (multi-omics DIABLO* framework)

TOP FEATURES (CIRCOS PLOT)



TOP FEATURES

✲ Multi-omics approaches reveal distinct features not identifiable through single-omics analyses.

✲ Transcriptomical data exhibit the highest discriminative power, effectively distinguishing between pathological and adjacent tissues, as well as between BE and EAC (second component of block SPLSDA).

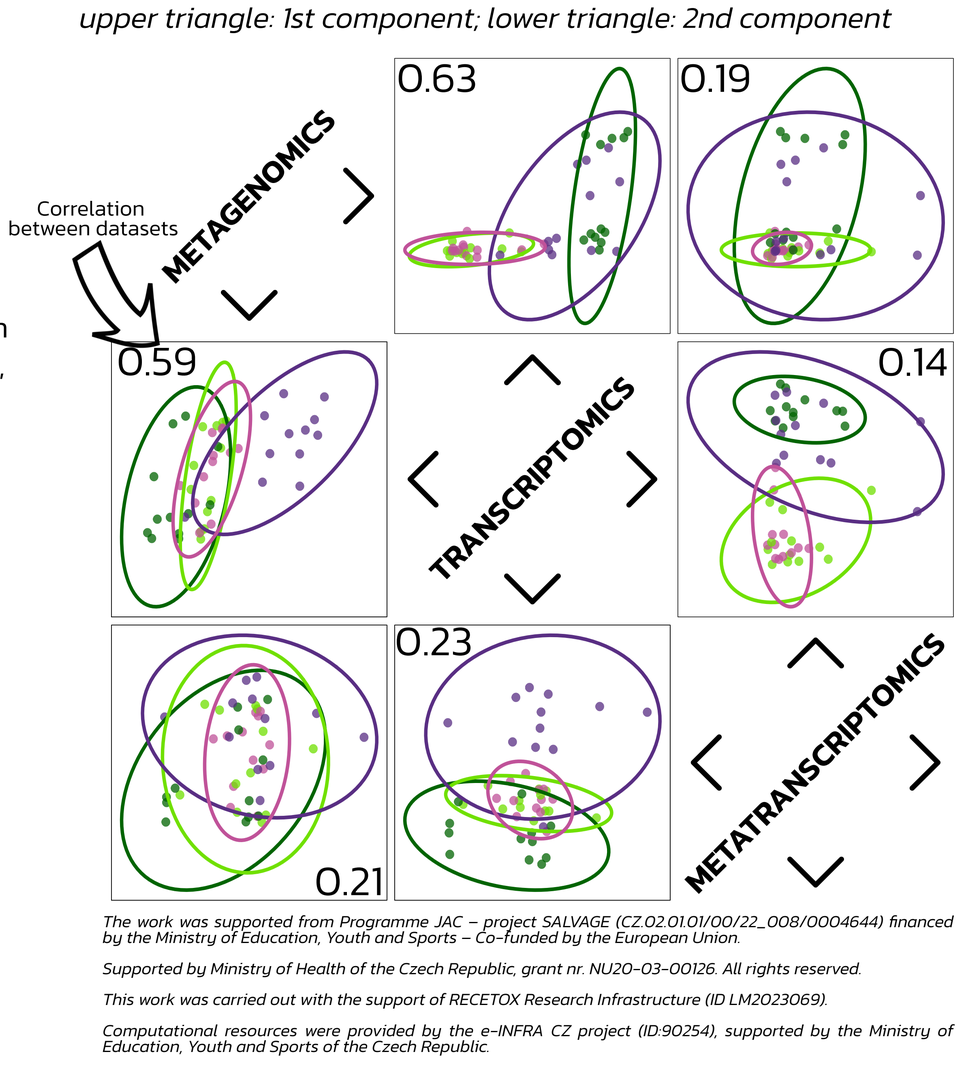
✲ Metatranscriptomics show limited discriminative ability.

✲ DIABLO analysis identifies bacterial genera that are able to differentiate BE samples from other types.

✲ BE and EAC samples exhibit greater variability compared to adjacent tissue samples.

BLOCK SPLSDA

upper triangle: 1st component; lower triangle: 2nd component



The work was supported from Programme JAC – project SALVAGE (CZ.02.01.01/00/22_008/0004644) financed by the Ministry of Education, Youth and Sports – Co-funded by the European Union.

Supported by Ministry of Health of the Czech Republic, grant nr. NU20-03-00126. All rights reserved.

This work was carried out with the support of RECETOX Research Infrastructure (ID LM2023069).

Computational resources were provided by the e-INFRA CZ project (ID:90254), supported by the Ministry of Education, Youth and Sports of the Czech Republic.

*Not the game nor black magic, but **D**ata **I**ntegration analysis for **B**iomarker discovery using **L**atent variable approaches for **O**mics studies developed by `mixOmics`: Rohart F, Gautier B, Singh A, and Le Cao K-A (2017) mixOmics: An R package for 'omics feature selection and multiple data integration. PLoS computational biology 13(11):e1005752