

## Different histological types of breast cancer share common epigenetic and transcriptomic signature that predicts clinical outcome

Jan Bińkowski<sup>1</sup>, Tomasz Kazimierz Wojdacz<sup>1\*</sup>

<sup>1</sup> Independent Clinical Epigenetics Laboratory, Pomeranian Medical University in Szczecin, Unii Lubelskiej 1, Szczecin 70-204, Poland

\*Correspondence: Tomasz K. Wojdacz, [tomasz.wojdacz@pum.edu.pl](mailto:tomasz.wojdacz@pum.edu.pl)

**Background:** Breast cancer is the leading cancer diagnosis in women accounting for over 10% of new cancer cases annually. It is also the most frequently diagnosed malignant tumor and primary cause of cancer deaths in women globally. Despite substantial improvement in the understanding of breast cancer biology, we are still deciphering the contribution of methylome changes to breast cancer pathology and association of those changes with expression of genes involved in cancer development. In our study we analyzed methylomics and transcriptomics profiles of three different breast cancer types: infiltrating duct carcinoma (n=50), breast lobular carcinoma (n=50) and mucinous adenocarcinoma (n=14)) as well as healthy breast tissue (n=46) to assess whether there is methylation signature common for those different cancer types associated with the gene expression changes observed during breast cancer pathology.

**Methods:** Statistical analyses were conducted using eDAVE platform that we recently developed [1] and Kaplan-Meier Plotter [2] tool. Gene set enrichment analyses as well as protein-protein interactions networks were performed using FUMA [3] and STRING [4] tools respectively.

**Results:** We found a subset of 615 genes and 13714 CpG sites with statistically significant aberrant expression ( $FDR \leq 0.05$ ,  $\log_2(FC) > 2$ ) and methylation levels ( $FDR \leq 0.05$ ,  $|\Delta| > 0.1$ ) respectively, between cancer and healthy breast tissue, that were common for all analyzed cancer types. The subset of 171 of the identified genes harbored aberrantly methylated CpG sites. The Gene Set Enrichment Analyses (GSEA) based on this subset of genes confirmed that expression of those genes is specific for adipose and breast mammary tissues ( $FDR \leq 0.05$ ) and those genes are significantly enriched in molecular pathways involved in: epithelial-mesenchymal transition, estrogen response as well as Wnt signaling. Survival analysis based on over 2000 breast cancer samples deposited in the TCGA, GEO and EGA databases showed that expression of 10 of the identified genes is significantly ( $p\text{-value} \leq 0.05$ ) associated with patients relapse free survival as well as overall survival.

**Conclusion:** Our study showed that there is common methylation signature associated with transcriptome of breast cancer and uniformly shared between histologically different types of this cancer. The identified methylation signature is furthermore strongly associated with the expression changes of genes involved in breast cancer pathology and strongly predicts clinical outcomes of patients such as overall survival and relapse free survival.

**Funding:** This study was funded by Polish Returns grant program from Polish National Agency for Academic Exchange, grant ID: PPN/PPO/2018/1/00088/U and OPUS22 grant from National Science Centre, grant ID: 2021/43/B/NZ2/02979.

### References:

1. Bińkowski J, Taryma-Leśniak O, Sokolowska KE, Przybylowicz PK, Staszewski M, Wojdacz TK: **eDAVE – Extension of GDC data analysis, visualization, and exploration tools.** *Computational and Structural Biotechnology Journal* 2023, **21**:5446-5450.
2. Gyórfy B: **Survival analysis across the entire transcriptome identifies biomarkers with the highest prognostic power in breast cancer.** *Comput Struct Biotechnol J* 2021, **19**:4101-4109.
3. Watanabe K, Taskesen E, van Bochoven A, Posthuma D: **Functional mapping and annotation of genetic associations with FUMA.** *Nature Communications* 2017, **8**(1):1826.
4. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P *et al*: **STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets.** *Nucleic Acids Res* 2019, **47**(D1):D607-d613.