



HAL
open science

**THEMA: Identification of molecular mechanisms by
which Tumor HEterogeneity influences disease outcome:
high-dimension Mediation Analysis**

Florence Pittion, Olivier François, Magali Richard

► **To cite this version:**

Florence Pittion, Olivier François, Magali Richard. THEMA: Identification of molecular mechanisms by which Tumor HEterogeneity influences disease outcome: high-dimension Mediation Analysis. 31st Annual Intelligent Systems For Molecular Biology and 22nd Annual European Conference on Computational Biology (ISCM/ECCB 2023), Jul 2023, Lyon, France. ISCM/ECCB 2023 Proceedings. hal-04667942

HAL Id: hal-04667942

<https://hal.science/hal-04667942v1>

Submitted on 5 Aug 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.

THEMA: Identification of molecular mechanisms by which Tumor Heterogeneity influences disease outcome: high-dimension Mediation Analysis



Florence Pittion, Olivier François, Magali Richard

MAGe team, TIMC, CNRS UMR 5525, Grenoble

corresponding author: florence.pittion@univ-grenoble-alpes.fr



Introduction

As tumors are dynamic heterogeneous ecosystems, it is essential to study the underlying heterogeneity for a better understanding of cancer biology. Currently, there is no framework to infer the causal mechanisms explaining effects of **tumor heterogeneity** on **disease outcome**. We want to investigate the share of effect attributed to molecular mechanisms such as **gene expression** and **DNA methylation (DNAm)** an epigenetic mechanism.

A tumor is composed of cancer cells and cells from the host tissue (stromal and immune components). During oncogenesis and cancer progression, all those cells evolve heterogeneously. We want study intra-tumor heterogeneity, as it has a major impact on cancer cells growth, division, resistance and metastasis.

THEMA project will concentrate on one cancer: development of **pancreatic ductal adenocarcinoma (PDAC)**. Incidence of PDAC increases regularly in Western countries and is expected to become the second leading cause of cancer-related mortality by 2025. According to our collaborators [3] PDACs are **highly heterogeneous** cancers, with an abundant differentiated stroma associated with prognostic relevance.

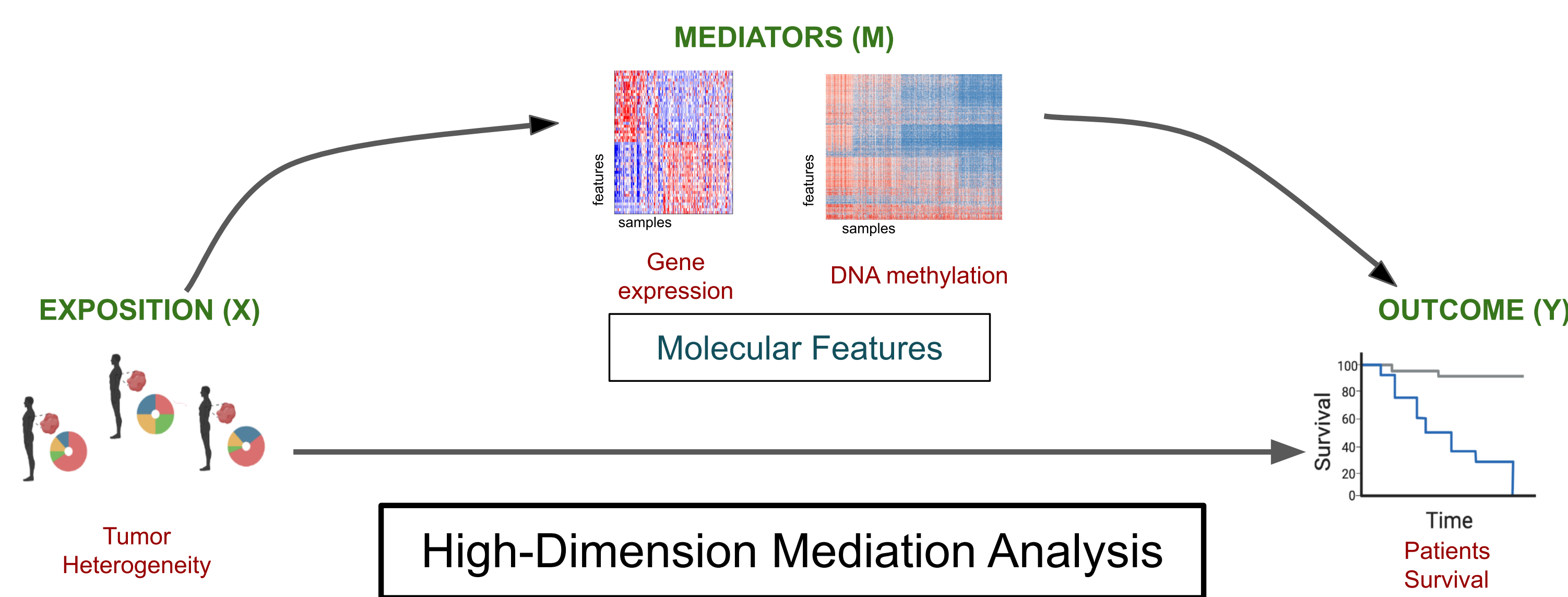
Model

Mediation analysis is a statistical approach for understanding the causal structure between an external exposure and a patient outcome, through the inclusion of mediator variables.

Classification based on heterogeneity will be used as **Exposure (X)** and patients survival associated analyses as **Outcome (Y)** in our mediation approach.

Potential **mediators** will be DNAm and gene expression. DNAm contributes to cell type differentiation through the control of gene expression and is often disrupted in cancer. Therefore it presents a potential therapeutic target.

These two molecular characteristics are on a large scale (20k for gene expression, 450k for DNAm), so we will carry out **high dimensional** mediation analysis.



Objectives

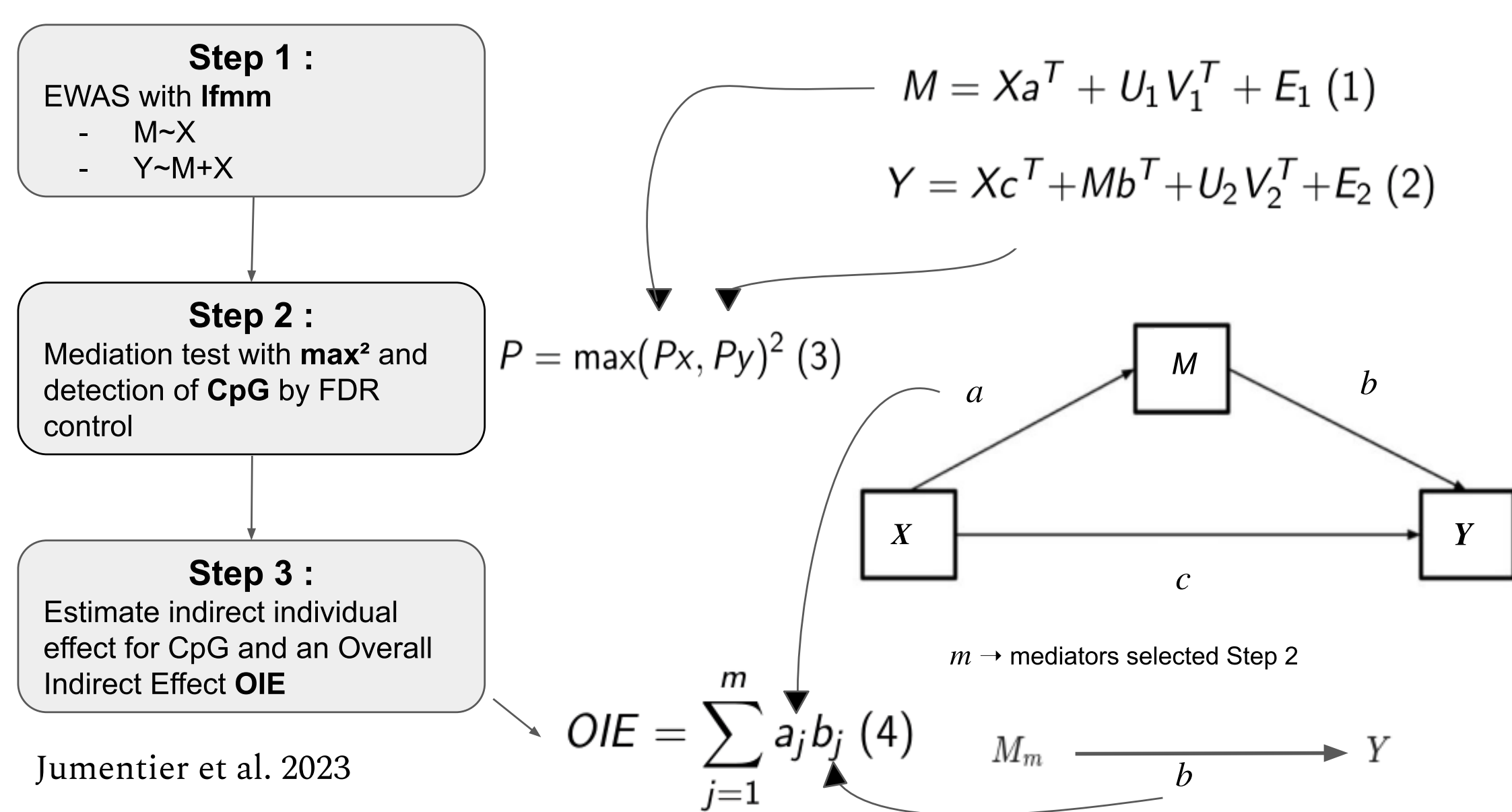
- Identify molecular mediators of tumor heterogeneity, both at the DNAm and gene expression level.
- Improve understanding of carcinogenesis
- Offer perspectives developing new biomarkers and personalized therapeutic treatments.

Goal

Developing a new multimodal high dimension mediation analysis framework to test to which extent the effect of tumor heterogeneity on disease outcome is explained by molecular features of the tumor.

Method

HDMAX 2: A framework for High Dimensional Mediation Analysis



Only methylation markers will be proposed as potential mediators in first approach.

HDMAX2 [1] pipeline will be used for mediation analysis:

- Two successive EWAS (Epigenome Wide association study) are performed using **lfmm** (Latent Factor Mixed Model) algorithms (equation (1) and (2)).
- Detection of causal probes with max^2 test (3).
- Estimation of individual indirect and overall indirect effects (4)

References

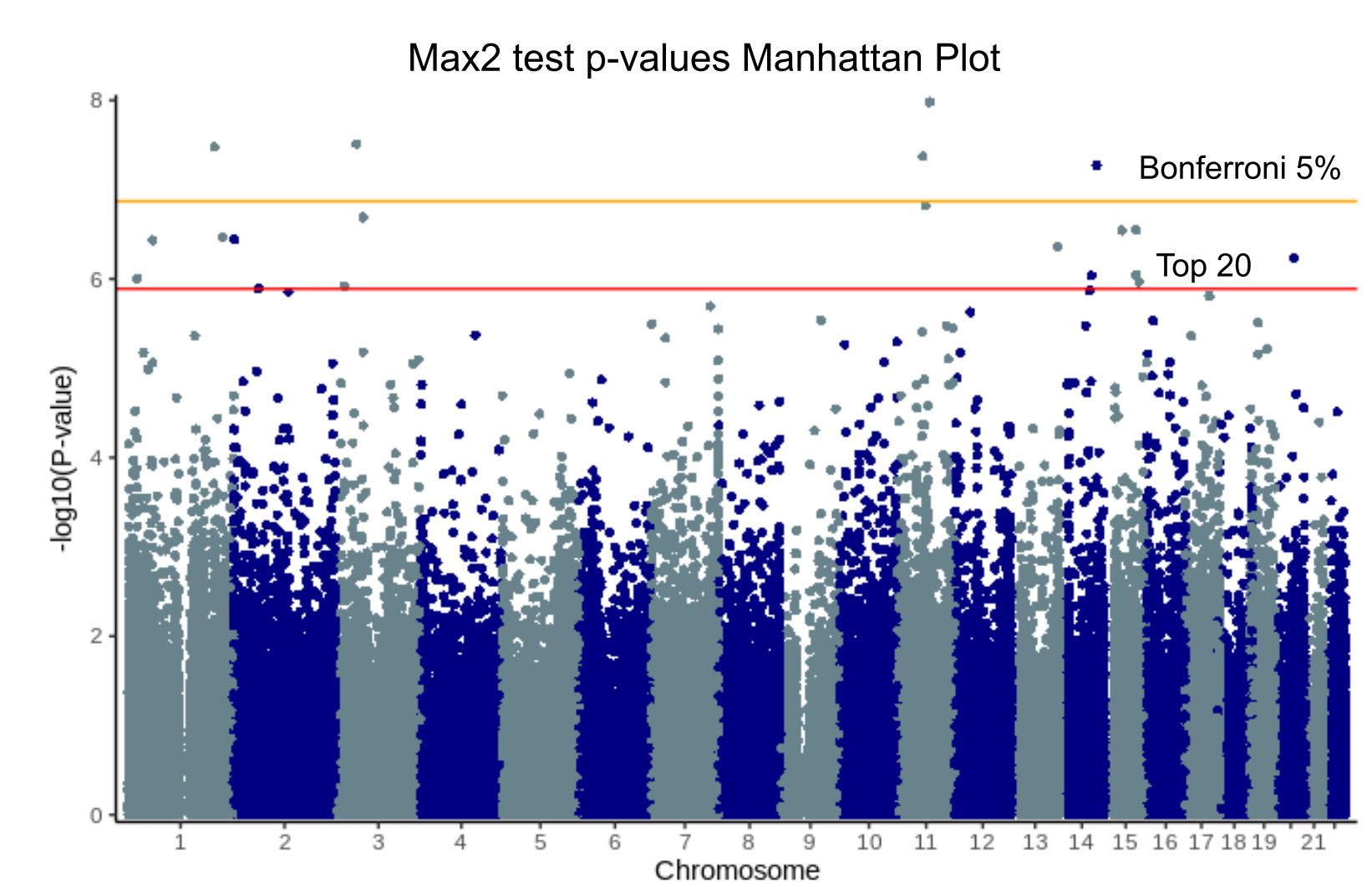
- [1] Basile Jumentier et al. "High-Dimensional Mediation Analysis: A New Method Applied to Maternal Smoking, Placental DNA Methylation, and Birth Outcomes". In: *Environmental Health Perspectives* 131.4 (). Publisher: Environmental Health Perspectives, p. 047011. DOI: 10.1289/EHP11559. URL: <https://ehp.niehs.nih.gov/doi/10.1289/EHP11559>.
- [2] Rémy Nicolle et al. "Establishment of a pancreatic adenocarcinoma molecular gradient (PAMG) that predicts the clinical outcome of pancreatic cancer". en. In: *EBioMedicine* 57 (July 2020), p. 102858.
- [3] Francesco Puleo et al. "Stratification of Pancreatic Ductal Adenocarcinomas Based on Tumor and Microenvironment Features". en. In: *Gastroenterology* 155.6 (Dec. 2018), 1999–2013.e3. ISSN: 00165085. DOI: 10.1053/j.gastro.2018.08.033.

Preliminary Results

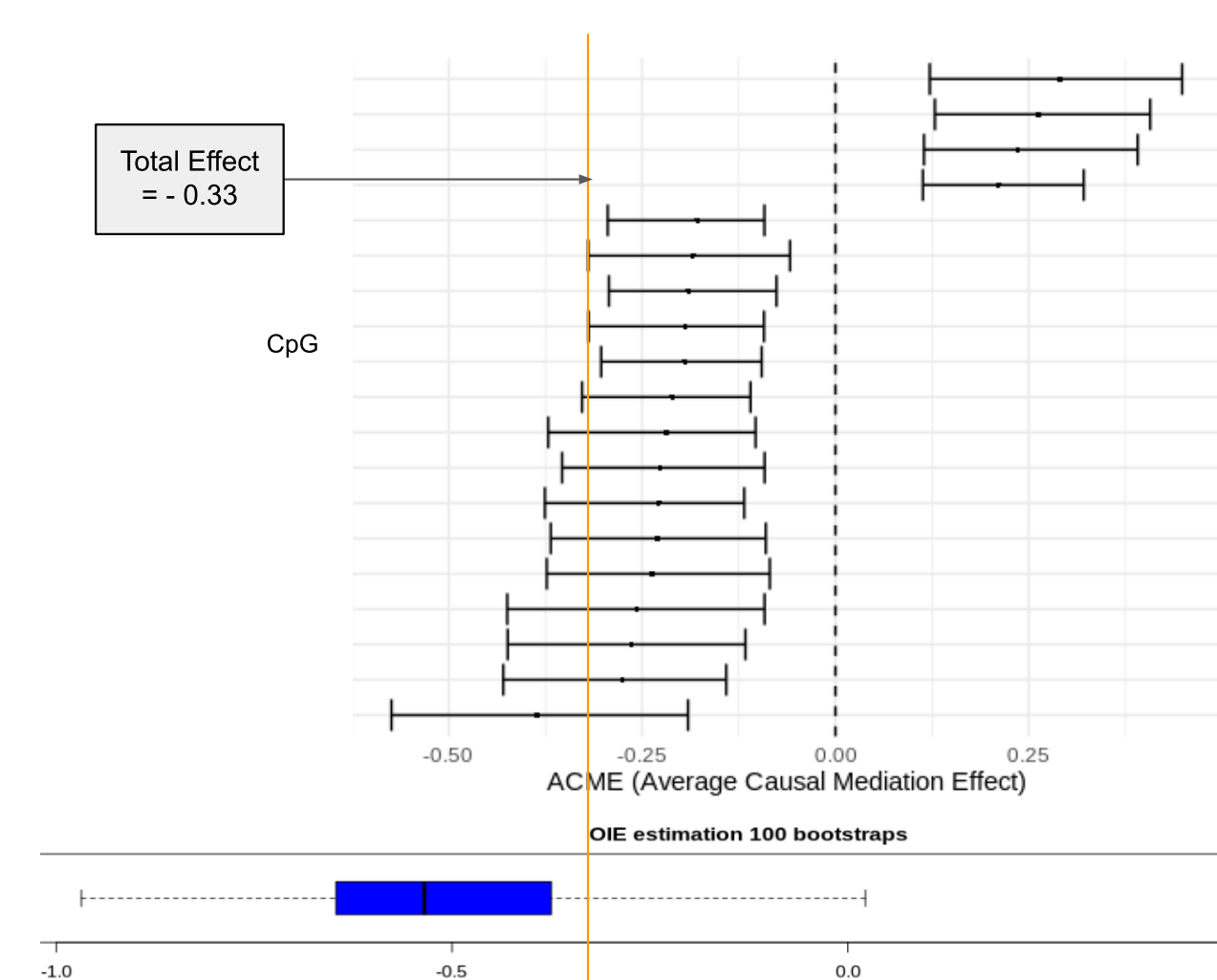
Our first mediation approach use:

- continuous exposition, tumor heterogeneity **molecular gradient** based on transcriptome [2]
- simplified outcome, **survival state at one year** follow up.

We apply HDMAX2 method on PDAC TCGA public dataset. Results after step 2 (max^2) illustrated in Mahanttan plot (on the right), with Bonferroni multiple testing controlling 5 probes are selected. This procedure is very conservative so we decided to selected **top 20** probes to be applied in the next step.



Mediators Individual Indirect Effects and Overall Indirect Effect



For each of the 20 selected probes **Individual Indirect Effect** is provided as well as Overall Indirect Effect (**OIE**). In forest plot (above) these results are compared with each other and set in perspective in relation to the total effect.

Future Tasks

- 1 Improve this first approach with categorical exposition and Cox model (adapted for survival data)
- 2 Expand this model to study the indirect effect of gene expression
- 3 Develop a statistical framework to perform multimodal (gene expression and DNAm) high dimension mediation analysis
- 4 Test how the exposure to therapeutic treatments affect the identified indirect effects