



# Gene–environment interaction analysis under the Cox model

Kuangnan Fang<sup>1</sup> · Jingmao Li<sup>1</sup> · Yaqing Xu<sup>2</sup> · Shuangge Ma<sup>3</sup> ·  
Qingzhao Zhang<sup>1,4</sup>

Received: 29 August 2022 / Revised: 21 February 2023 / Accepted: 24 February 2023 /

Published online: 10 April 2023

© The Institute of Statistical Mathematics, Tokyo 2023

## Abstract

For the survival of cancer and many other complex diseases, gene–environment (G-E) interactions have been established as having essential importance. G-E interaction analysis can be roughly classified as marginal and joint, depending on the number of G variables analyzed at a time. In this study, we focus on joint analysis, which can better reflect disease biology and is statistically more challenging. Many approaches have been developed for joint G-E interaction analysis for survival outcomes and led to important findings. However, without rigorous statistical development, quite a few methods have a weak theoretical ground. To fill this knowledge gap, in this article, we consider joint G-E interaction analysis under the Cox model. Sparse group penalization is adopted for regularizing estimation and selecting important main effects and interactions. The “main effects, interactions” variable selection hierarchy, which has been strongly advocated in recent literature, is satisfied. Significantly advancing from some published studies, we rigorously establish the consistency properties under high dimensionality. An effective computational algorithm is developed, simulation demonstrates competitive performance of the proposed approach, and analysis of The Cancer Genome Atlas (TCGA) data on stomach adenocarcinoma (STAD) further demonstrates its practical utility.

**Keywords** Gene–environment interaction analysis · Cox model · Penalized estimation · Asymptotic consistency

---

✉ Qingzhao Zhang  
qzzhang@xmu.edu.cn

Extended author information available on the last page of the article

## Authors and Affiliations

**Kuangnan Fang<sup>1</sup> · Jingmao Li<sup>1</sup> · Yaqing Xu<sup>2</sup> · Shuangge Ma<sup>3</sup> · Qingzhao Zhang<sup>1,4</sup>**

<sup>1</sup> Department of Statistics and Data Science, School of Economics, Xiamen University, No.422, Siming South Road, Xiamen 361005, Fujian, China

<sup>2</sup> School of Public Health, Shanghai Jiao Tong University School of Medicine, 227 South Chongqing Road, Shanghai 200240, China

<sup>3</sup> Department of Biostatistics, Yale School of Public Health, 60 College Street, New Haven, CT 06520, USA

<sup>4</sup> The Wang Yanan Institute for Studies in Economics, Xiamen University, No.422, Siming South Road, Xiamen 361005, Fujian, China