

Anisha Jain¹, Chandan Shivamallu², Amogha Shivakumar³, Chandan Dharmashekhar², Bhargav Shreevatsa², Rafael Rosell⁴, Ashwini Prasad², M N Nagendra Prasad¹

¹JSS Science and Technology University, Mysuru/IN, ²JSS Academy of Higher Education & Research, Mysuru/IN, ³KVC Super Speciality Hospital, Mysuru/IN, ⁴Instituto Oncológico Dr Rosell (IOR), Dexeus University Institute, Barcelona/ES

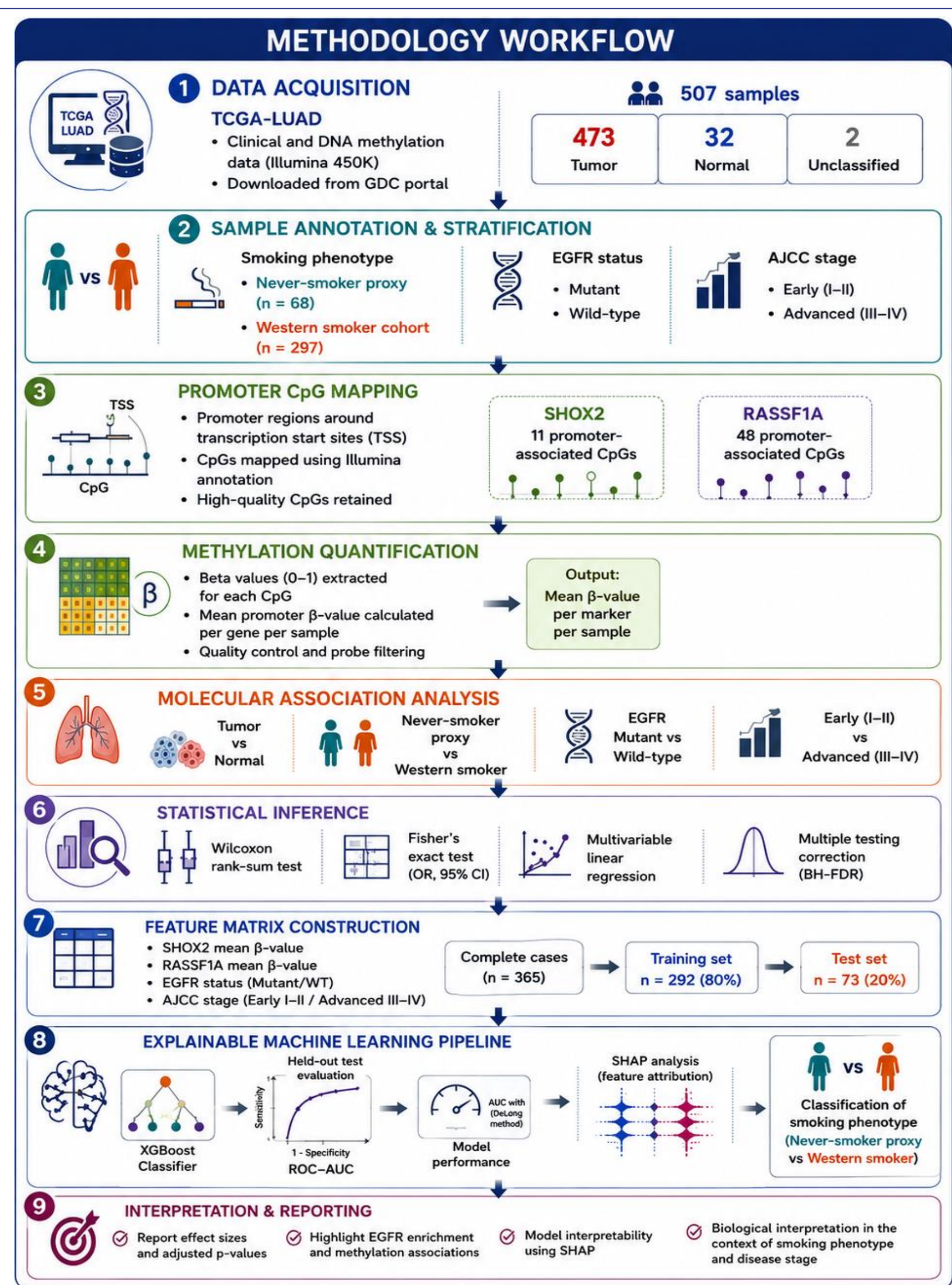
Correspondence: mnnagendraprasad@jssstuniv.in

INTRODUCTION

Lung adenocarcinoma in never-smokers determines a distinct molecular landscape, including an enrichment of EGFR alterations, highlighting the need for molecularly informed biomarker strategies. Aberrant promoter methylation of SHOX2 and RASSF1A represents a promising epigenetic approach for detecting lung cancer. The characterization of clinically relevant LUAD subgroups may be improved by integrating methylation patterns with smoking phenotype, EGFR status and pathological stage using explainable machine learning.

Objective: To characterize SHOX2 and RASSF1A promoter methylation in TCGA-LUAD, evaluate associations with clinical and molecular features, and integrate these using an explainable XGBoost ML model.

METHODS



DISCUSSION AND CONCLUSION

Actionable Take Away Message

- EGFR mutations were strongly enriched in never-smoker LUAD (36.8% vs 9.0%; OR = 5.84), highlighting a distinct molecular phenotype.
- SHOX2/RASSF1A methylation provided complementary information, assisting multi-marker rather than single-biomarker approaches.
- Combining SHOX2/RASSF1A methylation, EGFR status and pathological stage using XGBoost achieved moderate smoking-phenotype discrimination (AUC = 0.752; 95% CI: 0.619 - 0.885).

Future Direction for Research

- To validate in larger Asian never-smoker cohorts.
- Translate SHOX2/RASSF1A methylation to plasma/ctDNA and integrate methylation, EGFR + CEA/CYFRA21-1/NSE/pro-GRP for early detection.
- To evaluate early cancer by the development of CaDet-SHIELD.

Limitations

Limited Asian representation in TCGA; never-smoker proxy used. The current analysis is exploratory and tissue-based. Methylation associations were not significant after BH-FDR correction.

REFERENCES & ACKNOWLEDGEMENTS

References

[1] Chen Y, et al. Detection of SHOX2 and RASSF1A methylation for early-stage lung adenocarcinoma. *Mol Clin Oncol.* 2025;23(6):112. [2] Machado DI, et al. Circulating cell-free DNA methylation as biomarker for lung cancer detection: a systematic review and meta-analysis of diagnostic studies. *Syst Rev.* 2025;14(1):120. [3] Zhou Y, et al. Decoding the SHOX2 methylation-expression paradox in lung adenocarcinoma: Dual-regulatory methylation patterns drive oncogenic activation and prognostic relevance. *Cancer Lett.* 2026;648:218456. [4] Trombetta D, et al. Methylation Analyses in Liquid Biopsy of Lung Cancer Patients: A Novel and Intriguing Approach Against Resistance to Target Therapies and Immunotherapies. *Cancers (Basel).* 2025;17(18):3021. [5] Karlow JA, et al. Non-small Cell Lung Cancer Epigenomes Exhibit Altered DNA Methylation in Smokers and Never-smokers. *Genomics Proteomics Bioinformatics.* 2023;21(5):991-1013.

Acknowledgment

We thank JSS Science and Technology University and JSS Academy of Higher Education & Research, Mysuru for institutional and research support. A.J. acknowledges DBT, Government of India, for the Research Associateship (DBT RA/2024-25/Call-I/RA/49) and ANRF, Government of India, for financial assistance to attend and present at WCLC 2026 (File Number: ITS/2026/004007).

(AI) Use Disclaimer

OpenAI ChatGPT was used to troubleshoot data-analysis code and generating the initial infographic for the methodology overview which was subsequently modified and finalized by the authors using Microsoft 365 PowerPoint. All scientific content, analyses, and interpretations were independently reviewed and verified by the authors.

RESULTS

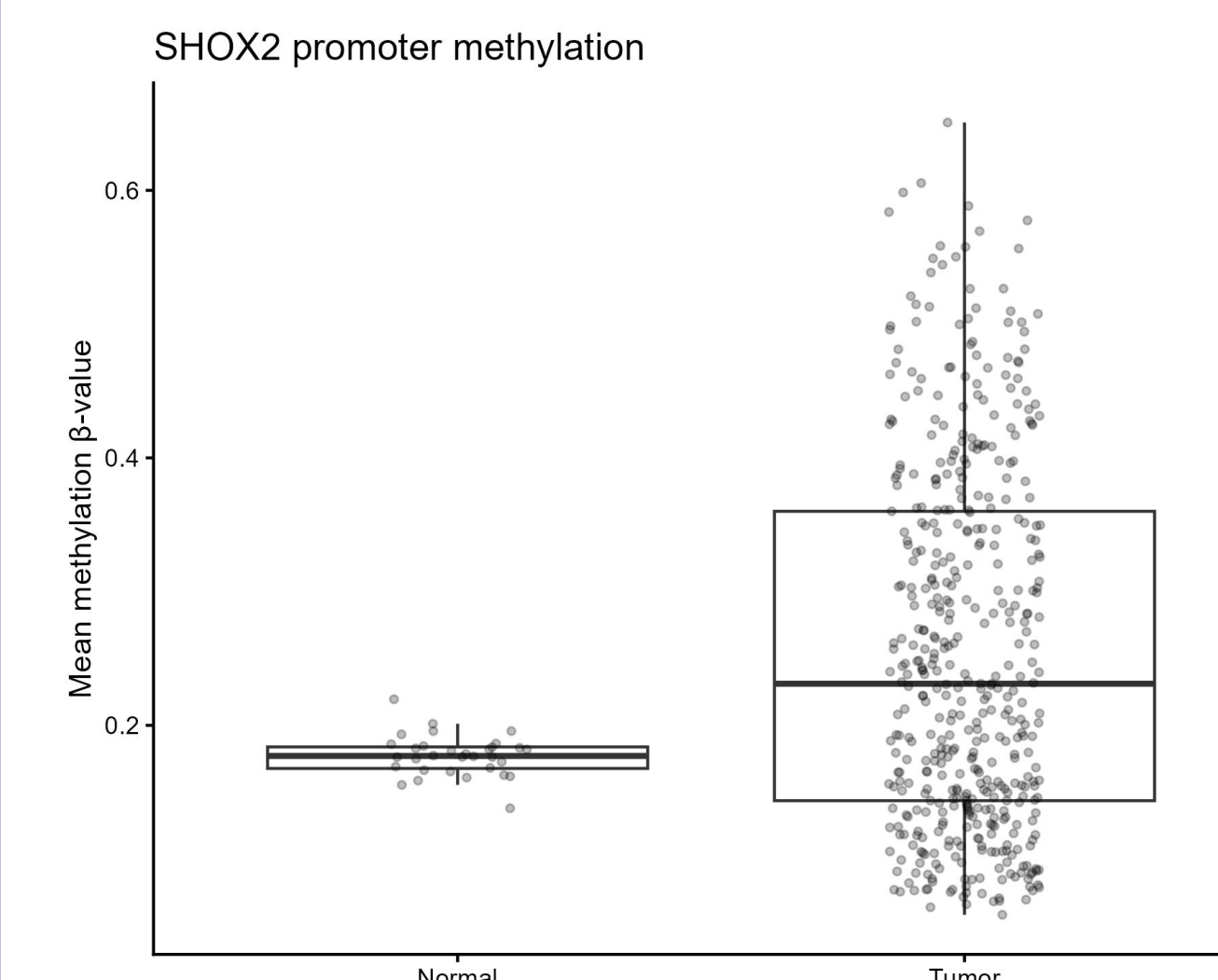


Figure 1. SHOX2 methylation differs between LUAD tumor and normal tissues. Mean β -values were higher in tumours (Wilcoxon $p = 0.01398$).

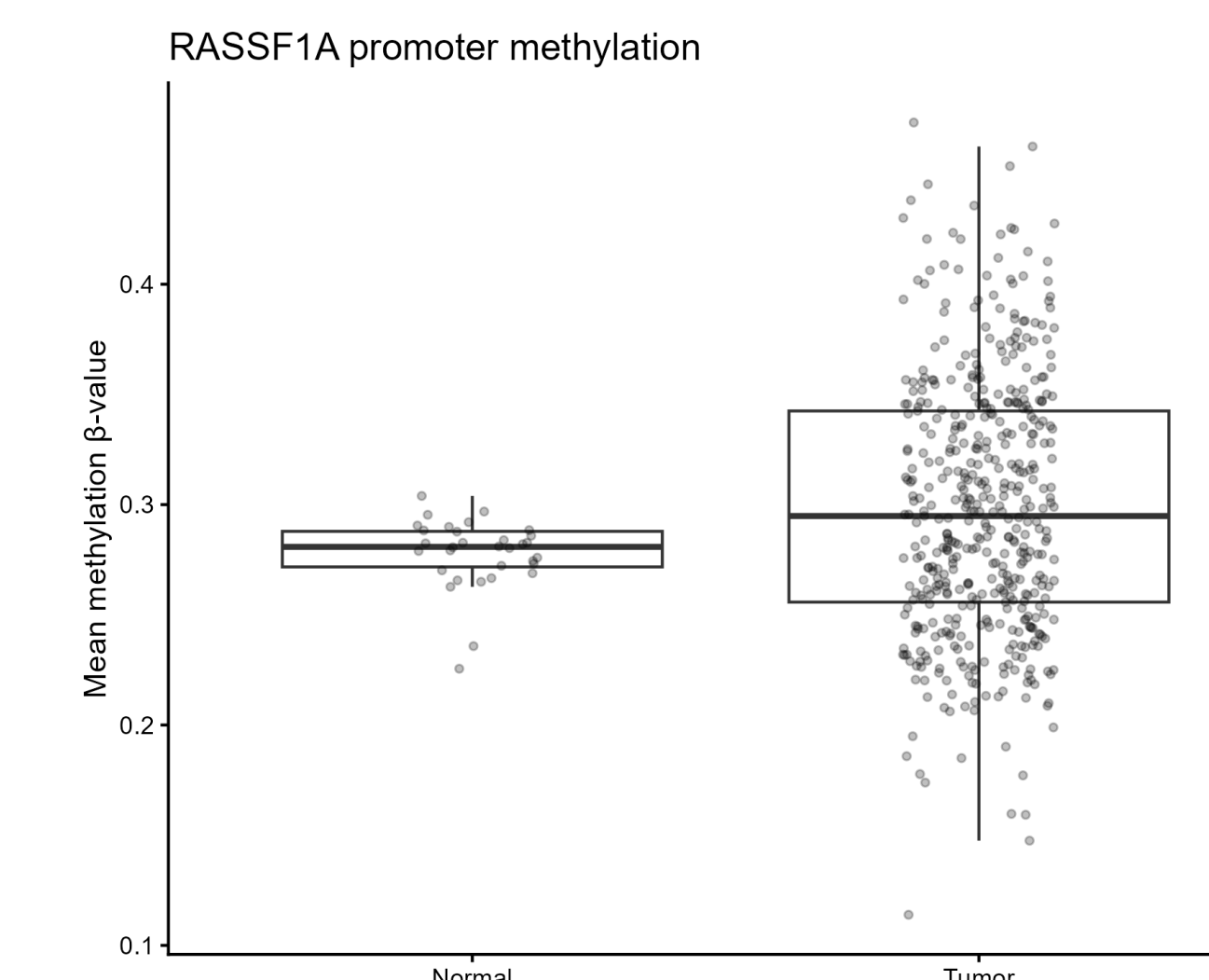


Figure 2. RASSF1A promoter-associated methylation in LUAD tumor and normal tissues. The difference was not statistically significant (Wilcoxon $p = 0.07117$).



Figure 3. SHOX2 methylation showed no statistically significant difference between the never-smoker proxy and Western smoker cohorts ($p = 0.1609$).

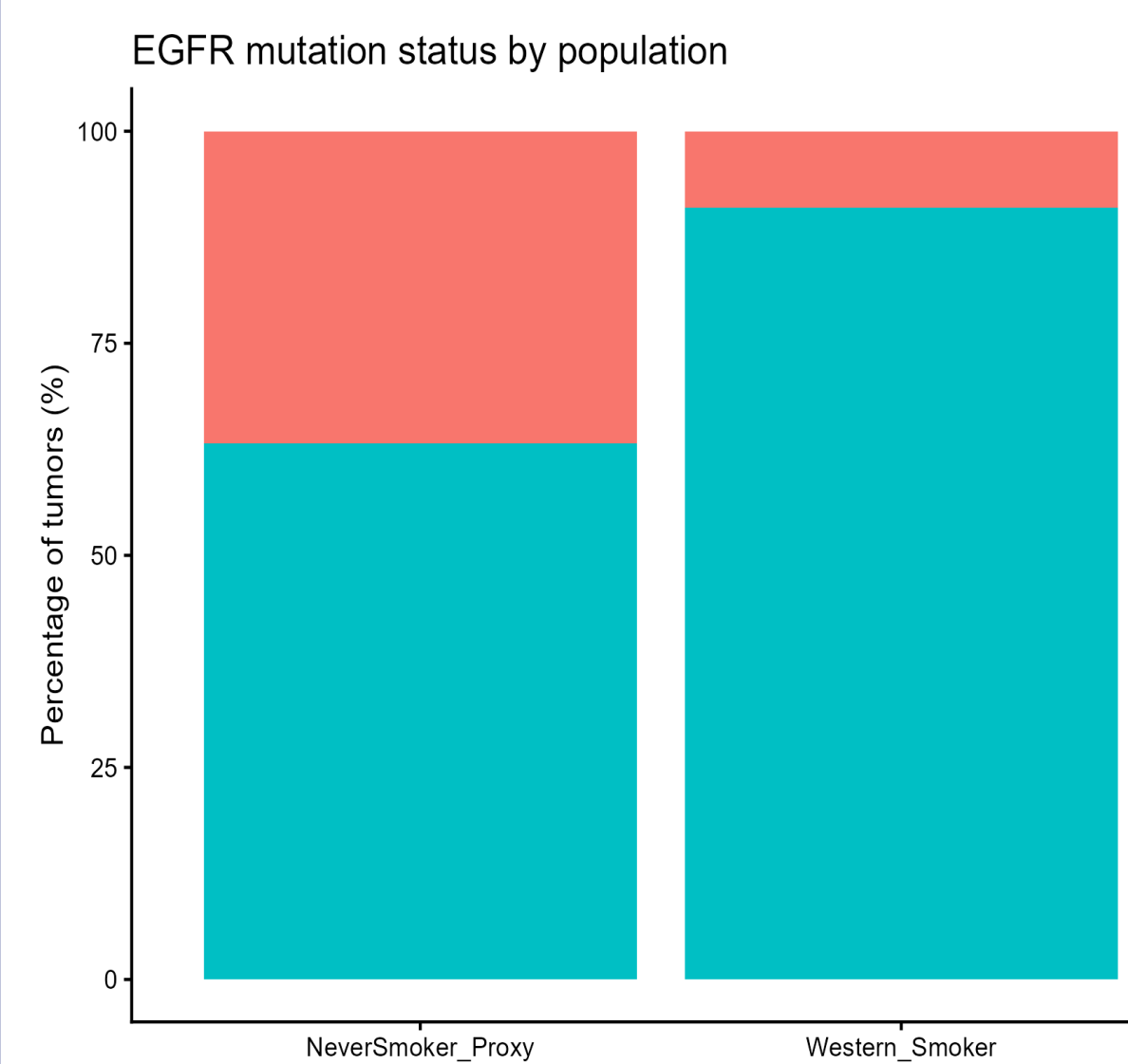


Figure 4. EGFR mutation enrichment according to smoking phenotype. EGFR mutations were more frequent in the never-smoker proxy cohort (36.8% vs 9.0%; OR=5.84, 95% CI: 2.96-11.56; Fisher's exact $p = 8.83 \times 10^{-8}$).

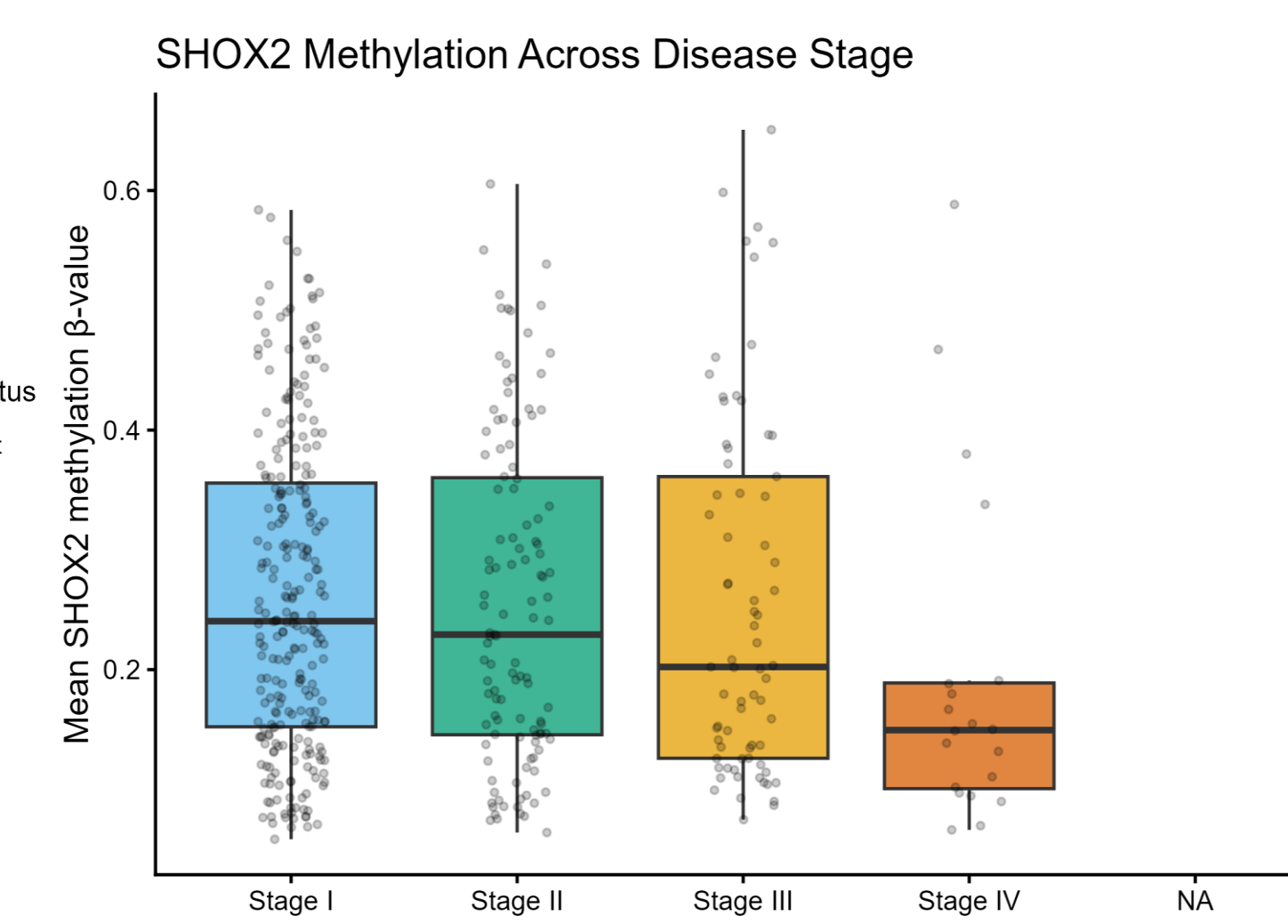


Figure 5. SHOX2 methylation showed an exploratory association with disease stage, with higher methylation in early-stage tumours than advanced-stage tumours (unadjusted $p = 0.07694$). Multivariable regression: $\beta = +0.0412$; $p = 0.0232$

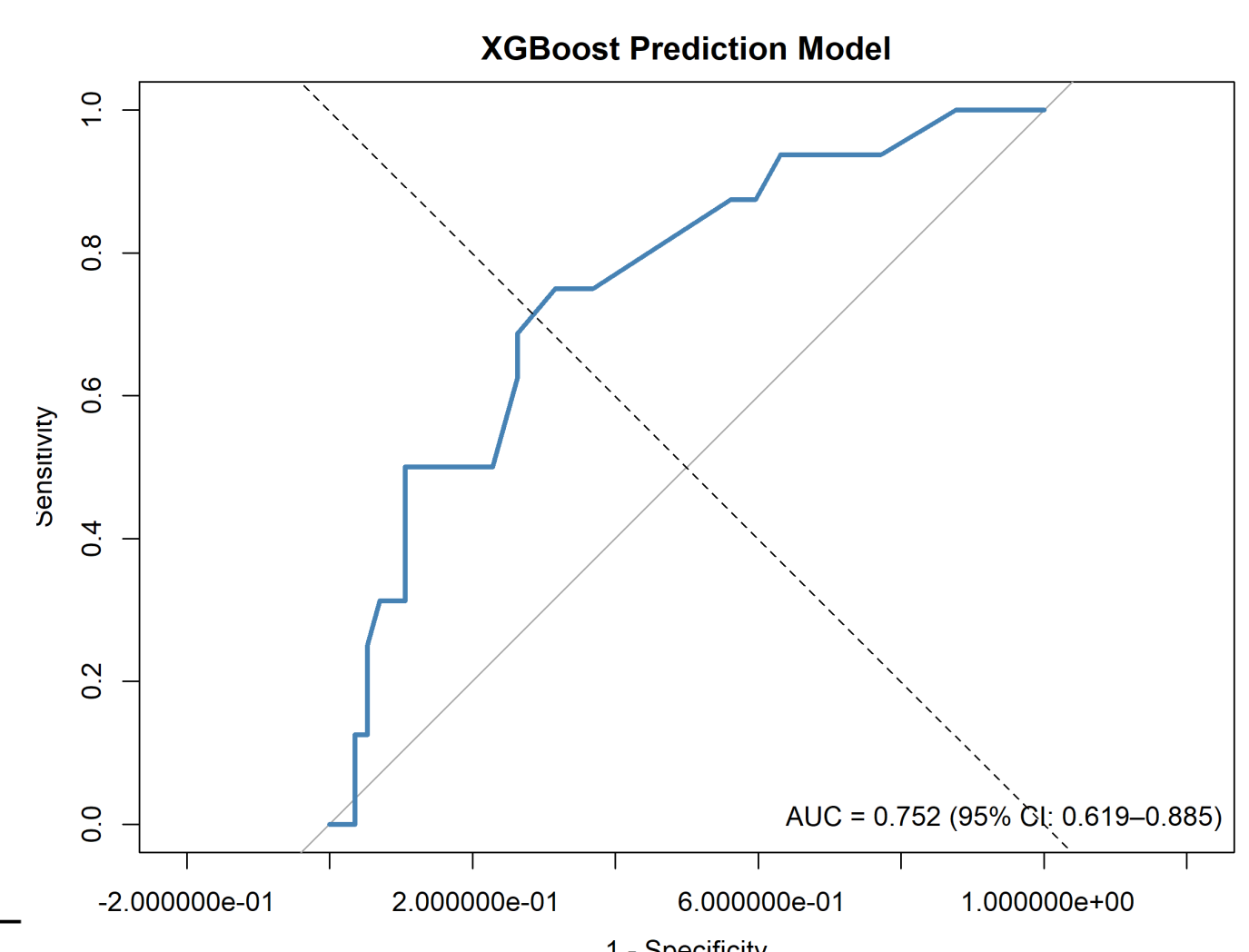


Figure 6. XGBoost ROC- Integrated SHOX2/RASSF1A methylation, EGFR and stage model for smoking-phenotype classification (AUC=0.752; 95% CI: 0.619-0.885)

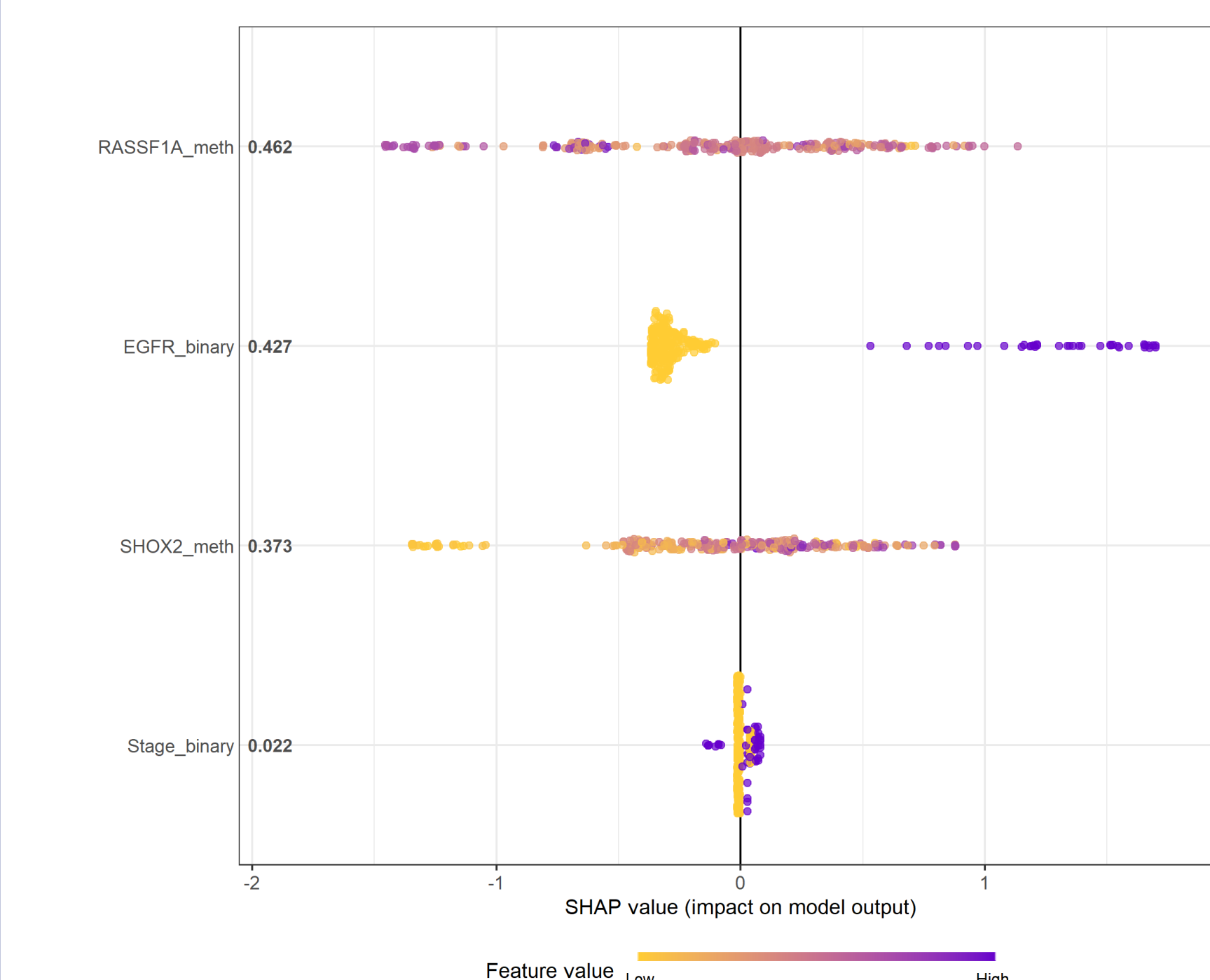


Figure 7. SHAP feature attribution- RASSF1A methylation, EGFR status and SHOX2 methylation were the principal contributors to XGBoost predictions; stage showed minimal contribution.

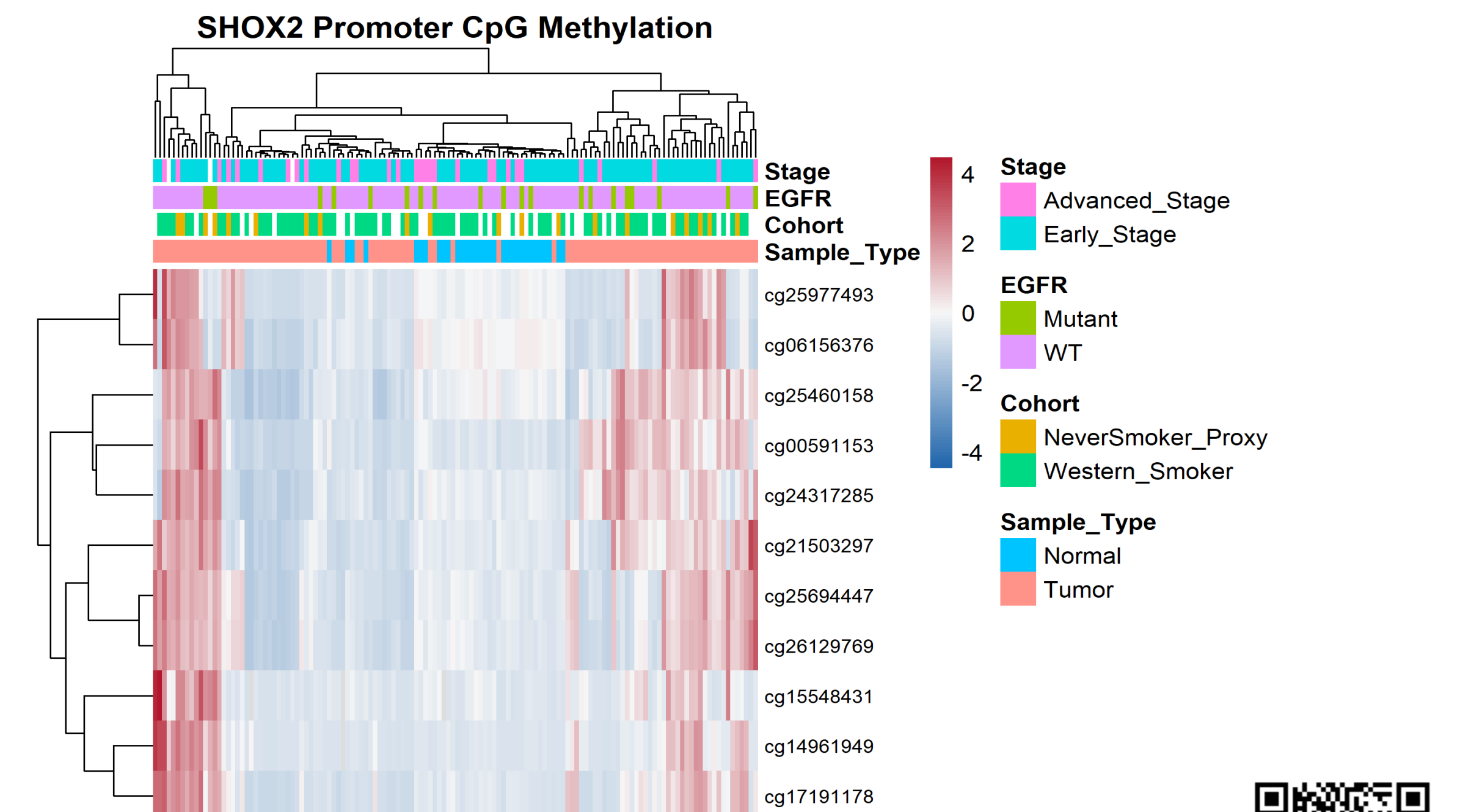


Figure 8. SHOX2 CpG heatmap- Row-scaled methylation patterns across 11 promoter-associated SHOX2 CpGs, annotated by stage, EGFR status, smoking cohort and sample type.

