2nd INTERNATIONAL CONFERENCE ON
Artificial Intelligence
31 May - 1 June 2026

IICAI
Second International Interdisciplinary Conference on Artificial Intelligence

Enhancing Single-Cell Transcriptomic Quality in Gastric Cancer: An Intersective Approach Combining Adaptive and Manual Quality Control

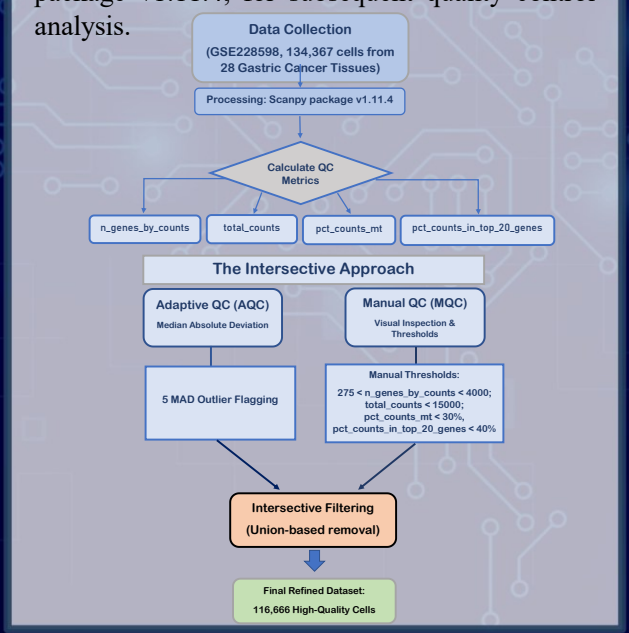
Fatemeh Khalaji-Pirbalouti*, Mojtaba Emadi-Baygi
*- MSc student of Genetics, Department of Genetics, Faculty of Basic Sciences, Shahrekord University, Shahrekord, Iran.
2- Corresponding author: Associate Professor of Genetics, Department of Genetics, Faculty of Basic Sciences, Shahrekord University, Shahrekord, Iran.
fkhalaji078@gmail.com; emadi-m@sku.ac.ir

Abstract

Single-cell RNA sequencing (scRNA-seq) has emerged as an indispensable tool for dissecting cellular heterogeneity in complex malignancies such as gastric cancer (GC). We present a robust QC framework, implemented in the Python environment using Scanpy package, and applied to publicly available GC scRNA-seq dataset from the Gene Expression Omnibus (GEO) database. Our approach intersects adaptive QC, based on the median absolute deviation (MAD) for objective, outlier-based filtering, with manual QC, where thresholds are visually determined via histogram and violin plot inspection to remove low-quality cells. Specifically, we targeted three critical metrics: the number of detected genes, total unique molecular identifier counts (UMI), and the fraction of counts from mitochondrial genes per cell. Here, we profiled the transcriptomes of 134,367 cells from 28 patients with gastric cancer. By intersecting the set of cells removed by both methods, we derive a refined, high-quality cell population. In conclusion, our intersective QC framework delivers a reproducible, high-quality gastric cancer scRNA-seq dataset by comprehensively mitigating technical artifacts.

Research method

ScRNA-seq data from 28 gastric cancer tissue specimens were obtained from the Gene Expression Omnibus under accession number GSE228598. The raw count matrix was loaded into the Python environment using the Scanpy package v1.11.4, for subsequent quality control analysis.



conclusion

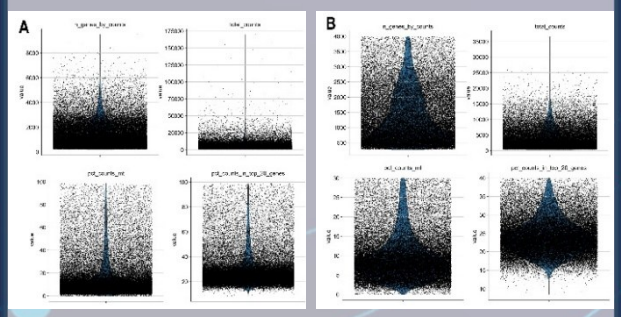
We developed and implemented a robust, intersective quality control strategy using adaptive quality control (AQC) and manual quality control (MQC) filtering methods on gastric cancer scRNA-seq dataset. This dual-approach provides a more comprehensive mitigation of technical artifacts than either method used alone. The resulting high-quality dataset is optimized for revealing true cellular heterogeneity and molecular mechanisms underlying gastric cancer pathogenesis. This reproducible workflow is critical for high-quality scRNA-seq studies, particularly in contexts of high cellular heterogeneity like tumor biology.

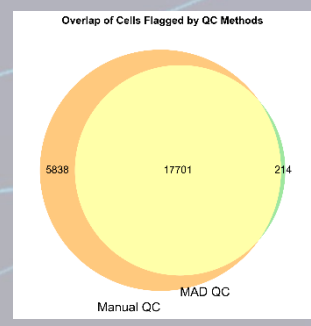
The purpose of the research

Due to the heterogeneity and plasticity of cancers including gastric cancer (GC) and the amount of data produced by scRNA-seq technologies, we focused our study on GC data analysis. Scanpy package facilitates quality control (QC) and visualization to dissect tumor heterogeneity and identify meaningful biological patterns from scRNA-seq datasets. Rigorous QC is fundamental to the fidelity of scRNA-seq data; technical noise and artifacts propagate directly into spurious biological interpretations, particularly in heterogeneous TMEs where rare but clinically relevant cell states may be obscured by low-quality cells. Our intersective framework harmonizes adaptive filtering with manual curation, leveraging the strengths of both paradigms while mitigating their individual limitations. MAD-based AQC provides objective, scalable thresholds that adapt to dataset-specific noise profiles without requiring a priori assumptions. This statistical approach is computationally efficient and reproducible.

Practical or simulation results

After applying the robust QC approach, a total of 17,701 low-quality barcodes were removed. This resulted in a final high-quality dataset of 116,666 cells retained.





References

[1]Li Y, Hu X, Lin R, Zhou G, Zhao L, Zhao D, Zhang Y, Li W, Zhang Y, Ma P, Ren H, Liao X, Niu P, Wang T, Zhang X, Wang W, Gao R, Li Q, Church G, He J, Chen Y. Single-cell landscape reveals active cell subtypes and their interaction in the tumor microenvironment of gastric cancer. Theranostics. 2022 May 9;12(8):3818-3833.
[2]Jovic D, Liang X, Zeng H, Lin L, Xu F, Luo Y. Single-cell RNA sequencing technologies and applications: A brief overview. Clin Transl Med. 2022 Mar;12(3):e694.
[3]Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. 2018 Feb 6;19(1):15.
[4]Germain PL, Sonrel A, Robinson MD. pipeComp, a general framework for the evaluation of computational pipelines, reveals performant single cell RNA-seq preprocessing tools. Genome Biol. 2020 Sep 1;21(1):227.