Bioimage informatics

ATHENA: analysis of tumor heterogeneity from spatial omics measurements

Adriano Luca Martinelli and Maria Anna Rapsomaniki *

IBM Research Europe, Zurich, 8803 Rüschlikon, Switzerland

*To whom correspondence should be addressed.

Received on January 8, 2022; revised on April 9, 2022; editorial decision on April 23, 2022; accepted on April 26, 2022

Abstract

Summary: Tumor heterogeneity has emerged as a fundamental property of most human cancers, with broad implications for diagnosis and treatment. Recently, spatial omics have enabled spatial tumor profiling, however computational resources that exploit the measurements to quantify tumor heterogeneity in a spatially aware manner are largely missing. We present ATHENA (Analysis of Tumor Heterogeneity from Spatial omics measurements), a computational framework that facilitates the visualization, processing and analysis of tumor heterogeneity from spatial omics measurements. ATHENA uses graph representations of tumors and bundles together a large collection of established and novel heterogeneity scores that quantify different aspects of the complexity of tumor ecosystems.

Availability and implementation: ATHENA is available as a Python package under an open-source license at: https://github.com/AI4SCR/ATHENA. Detailed documentation and step-by-step tutorials with example datasets are also available at: https://ai4scr.github.io/ATHENA/. The data presented in this article are publicly available on Figshare at https://figshare.com/articles/dataset/zurich_pkl/19617642/2.

Contact: aap@zurich.ibm.com

Supplementary information: Supplementary data are available at Bioinformatics online.

1 Introduction

Tumor ecosystems exhibit a considerable degree of inter- and intratumor phenotypic heterogeneity, arising from both intrinsic and extrinsic sources (Marusyk et al., 2012). Within these ecosystems, a number of phenotypically diverse cancer cell subpopulations coexist alongside other cell types of the tumor microenvironment, notably immune and stromal cells. However, their synergistic or antagonistic interactions are not fully understood. Growing evidence suggests that tumor heterogeneity is clinically relevant and not only determines disease progression, but also affects therapeutic response (Tabassum and Polyak, 2015). Accurate and biologically meaningful quantification of tumor heterogeneity at diagnosis or during treatment has the potential to translate biological complexity into actionable insight, enabling precise and personalized therapy approaches.

To achieve this goal, several heterogeneity quantification scores have been devised, largely based on spatial statistics and information theory, that, when applied to different omics or tissue imaging data showed prognostic capabilities in various cancer types (Kashyap et al., 2021). At the same time, the advancement of spatial omics (e.g. imaging mass cytometry (Giesen et al., 2014), multiplexed ion beam imaging (Angelo et al., 2014) and spatial transcriptomics (Ståhl et al., 2016]) is revolutionizing cancer biology, enabling the simultaneous quantification of dozens of proteins or thousands of transcripts in millions of cells while preserving the architecture of the tumor ecosystem. Although a number of useful data analysis platforms have emerged [e.g. Squidpy (Palla et al., 2022), Giotto (Dries et al., 2021) and MISTy (Tanevski et al., 2022)], there is a need for a dedicated resource that facilitates tumor heterogeneity quantification. Here, we introduce ATHENA (Analysis of Tumor Heterogeneity from Spatial omics measurements), an open-source computational framework that brings together established and novel scores able to capture the heterogeneity of the tumor ecosystem, with a strong focus on visualization at each step of the analysis.

2 Overview of ATHENA

Starting with raw multiplexed images, cell segmentation masks and cell phenotypes, ATHENA first constructs a cell–cell graph encoding the tumor topology and then applies a number of heterogeneity scores that quantify the complexity of the tumor architecture (Fig. 1A). ATHENA supports any spatial omics data modality [e.g. imaging mass cytometry (IMC), multiplexed ion beam imaging (MBI), seqFISH, Visium], as well as standard tissue imaging data [e.g. multiplexed immunohistochemistry (mIHC)].
ATHENA is implemented in a highly modular, extendable and scalable fashion, based on two main components: SpatialOmics, a new data structure inspired by AnnData (Wolf et al., 2018) that accommodates storing spatial omics data in a technology-agnostic way, and SpatialHeterogeneity, a module that implements all processing, analysis and visualization steps integral to ATHENA’s functionalities (Supplementary Methods S1.1–S1.2). At the core of ATHENA resides a graph representation of the tissue, based on three different biologically inspired graph builders that model different levels of cell–cell interactions: (i) a radius graph, where cells within a given radius \( r \) are connected, modeling paracrine signaling, (ii) a contact graph, where only cells that are in physical contact are connected, modeling juxtacrine signaling and (iii) a \( k \)NN graph, where \( k \) nearest neighboring cells are connected, a popular graph choice in digital pathology (Pati et al., 2022) (Supplementary Methods S1.3). Next, ATHENA computes a number of established and novel heterogeneity scores, broadly classified into the following categories depending on their underlying mathematical foundations: (i) spatial statistics scores that quantify the degree of clustering or dispersion of each phenotype individually, (ii) graph-theoretic scores that examine structural properties of the tumor graph, (iii) information-theoretic scores that quantify how diverse the tumor is with respect to different phenotypes present and their relative proportions and (iv) cell interaction scores that assess the pairwise relationships between different phenotypes in the tumor ecosystem (Supplementary Methods S1.4). This large collection of scores includes established approaches previously used in multiple tumor heterogeneity studies [see Kashyap et al. (2021) for a thorough review], as well as novel scores introduced in ATHENA (Supplementary Table S1). Specifically, ATHENA implements a new concept of local heterogeneity scores, computed at a single-cell level using the graph topology that capture spatial heterogeneity and highlight local variations of tumor diversity or cell interactions (Fig. 1D and G). Finally, ATHENA simplifies interoperability with other spatial omics analysis platforms such as Squipdy (Palla et al., 2022), and allows exporting the resulting scores for downstream analysis using any statistics or machine learning platform.

3 Results

We tested ATHENA on a publicly available IMC dataset that contains a number of breast cancer tissue microarray images (Jackson et al., 2020), annotated by the authors in a number of phenotypes (Fig. 1B). Depending on the preferred choice of graph topology, the tumor can be represented as a radius or contact graph (Fig. 1C), connecting cells within a certain distance, or cells in contact, respectively. Quantification of local Rao’s quadratic entropy highlights tumor regions with high diversity of interactions between phenotypically distant cells (Fig. 1D). Spatial cluster analysis scores quantify the degree of spatial clustering or dispersion as a function of distance, revealing a strong clustering of the three most abundant cancer cell phenotypes at a medium distance (peaks in the three curves of Fig. 1E—top), and random distribution of the three most abundant immune and stromal cell types (curves falling on the zero axis in Fig. 1E—bottom). Neighborhood analysis scores quantify cell type interactions, suggesting avoidance of cells high for Hormone Receptors (HR\( ^{\text{hi}} \)) and positive for Cytokeratins (CK\( ^{\text{+}} \)) by most immune and stromal cell types and attraction by most other cancer cell phenotypes (last column indicated by arrow in Fig. 1F), and self-attraction of Basal CK\( ^{\text{+}} \) cells (heatmap entry indicated by arrow in Fig. 1F). Last,
infiltration maps (Fig. 1G) point to immune-tumor ‘hotspots’, i.e. tumor regions locally penetrated by few immune cells.

4 Conclusion
We developed ATHENA to offer to the research community a modular computational framework that allows the quantification of custom heterogeneity scores from spatial omic datasets. ATHENA is highly interoperable with popular single-cell analysis platforms [e.g. Scanpy (Wolf et al., 2018), Squidpy (Palla et al., 2022)] and can be easily extended to accommodate tailored heterogeneity scores. As spatial omics datasets become increasingly available, we hope that ATHENA will be a valuable resource for accurate tumor heterogeneity analyses.

Acknowledgements
The authors thank all members of the computational pathology team at IBM Research Zurich and Prof. Mark Robinson, University of Zurich, for fruitful discussions and feedback.

Funding
This work was supported by the Swiss National Science Foundation (SNSF) Sinergia Grant [CRSII5_202297].

Conflict of Interest: none declared.

References
Angelo,M. et al. (2014) Multiplexed ion beam imaging of human breast tumors. Nat. Med., 20, 436–442.
Dries,R. et al. (2021) Giotto: a toolbox for integrative analysis and visualization of spatial expression data. Genome Biol., 22, 78.
Giesen,C. et al. (2014) Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. Nat. Methods, 11, 417–422.
Jackson,H. et al. (2020) The single-cell pathology landscape of breast cancer. Nature, 578, 615–620.
Kashyap,A. et al. (2021) Quantification of tumor heterogeneity: from data acquisition to metric generation. Trends Biotechnol., https://doi.org/10.1016/j.tibtech.2021.11.006.
Marusyk,A. et al. (2012) Intra-tumour heterogeneity: a looking glass for cancer? Nat. Rev. Cancer, 12, 323–334.
Palla,G. et al. (2022) Squidpy: a scalable framework for spatial single cell analysis. Nat. Methods, 19, 171–178.
Pati,P. et al. (2022) Hierarchical graph representations in digital pathology. Med. Image Anal., 75, 102264.
Stähl,P. et al. (2016) Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. Science, 353, 78–82.
Tabassum,D.P. and Polyak,K. (2015) Tumorigenesis: it takes a village. Nat. Rev. Cancer, 15, 473–483.
Tanevski,J. et al. (2022) Explainable multi-view framework for dissecting inter-cellular signaling from highly multiplexed spatial data, Genome Biol., 23, 97.
Wolf,F.A. et al. (2018) SCANPY: large-scale single-cell gene expression data analysis. Genome Biol., 19, 15.