Journal of Translational Medicine

LETTER TO THE EDITOR Open Access

Harnessing sample preparation for RNA-sequencing toward a reliable bioinformatics analysis

Marianna Talia[1†], Eugenio Cesario[2†], Francesca Cirillo[1†], Domenica Scordamaglia[1], Marika Di Dio[1], Azzurra Zicarelli[1], Adelina Assunta Mondino[1], Maria Antonietta Occhiuzzi[1], Ernestina Marianna De Francesco[3], Antonino Belfiore[4], Anna Maria Miglietta[5], Michele Di Dio[6], Carlo Capalbo[1,7], Marcello Maggiolini[1*†] and Rosamaria Lappano[1*†]

Dear Editor,

We aim to clarify the issues raised by Tan et al. [1] regarding our recent publication "Cancer-associated fibroblasts (CAFs) gene signatures predict outcomes in breast and prostate tumor patients" [2], which provided changes in the transcriptomic milieu of these malignancies largely occurring in women and men, respectively [3]. We appreciated that the letter of Tan and colleagues dealt with our experimental model and data, giving us the opportunity to further detail the evidence shown.

The first issue was related to the patient's characteristics and sample processing before RNA-sequencing (RNA-seq) analysis. Regarding the patient data, we selected 20 female patients with luminal invasive breast cancer characterized by estrogen receptor (ER)-positivity, human epidermal growth factor receptor 2 (HER2)-negativity and Ki67 ≥ 30% as well as 20 male patients with prostate cancer characterized by Gleason score at least of 8 or PSA > 20 mg/l. Concerning the sample processing method, we specified the experimental procedure in the "cell cultures" paragraph of the Methods section of our manuscript [2]. In particular, we stated that breast and prostate CAFs were isolated respectively from 20 mammary ductal carcinomas and 20 prostate adenocarcinomas. Then, a unique population of breast and prostate CAFs was obtained by pooling the 20 isolated cell cultures of each tumor type. From the single population of breast and prostate CAFs, we obtained three biological replicates, as previously recommended for the RNA-seq experiments [4]. Next, the RNA extraction from each biological replicate was performed by employing the same sample processing method.

As it concerns the grouping settings of our analysis, we aimed to identify the differences in the gene expression profiles occurring in CAFs of the two malignancies considered, similarly to previous studies that evaluated different types of tumors by RNA-seq analyses [5]. Therefore, comparing the differentially expressed genes in CAFs from breast and prostate cancer, we uncovered the unique molecular signatures and pathways that may

†Marianna Talia, Francesca Cirillo, Domenica Scordamaglia contributed equally to this work. Marcello Maggiolini and Rosamaria Lappano are co-last authors.

*Correspondence:
Marcello Maggiolini
marcello.maggiolini@unical.it
Rosamaria Lappano
rosamaria.lappano@unical.it
[1]Department of Pharmacy, Health and Nutritional Sciences, University of Calabria, 87036 Rende, Italy
[2]Department of Cultures, Education and Society, University of Calabria, 87036 Rende, Italy
[3]Department of Medicine and Surgery, University of Enna "Kore", 94100 Enna, Italy
[4]Endocrinology, Department of Clinical and Experimental Medicine, University of Catania, Garibaldi-Nesima Hospital, 95122 Catania, Italy
[5]Breast and General Surgery Unit, Annunziata Hospital Cosenza, 87100 Cosenza, Italy
[6]Division of Urology, Department of Surgery, Annunziata Hospital, 87100 Cosenza, Italy
[7]Complex Operative Oncology Unit, Annunziata Hospital Cosenza, 87100 Cosenza, Italy

trigger the action of CAFs in these malignancies. Our data may also help in pinpointing potential specific biomarkers and therapeutic targets of the breast and prostate tumor microenvironment, thereby supporting the improvement of personalized medicine and targeted strategies.

We esteemed the suggestion of Tan and colleagues [1] regarding the use of specific tools that may allow the identification of molecular and biological functions of the differentially expressed genes. However, we preferred the use of the enrichment analyses rather than to explore gene correlation patterns. Additionally, we clarify that our study did not compare tumor versus normal TCGA samples, as stated in the letter of Tan and colleagues. Instead, we specifically compared breast versus prostate samples of the TCGA dataset patients, thereafter we intersected the identified genes with those obtained in CAFs.

Next, we remark that the cumulative impact of the genes belonging to the identified signatures was robustly held by k-means analysis. Importantly, we demonstrated that a common prognosis characterizes patients clustered according to comparable gene expression patterns. The reliability of our analysis was further validated through a classification task, which provided attribute usage for each gene along with metrics for accuracy, recall and precision. The attribute usage percentage served as a quantitative measure of each gene's influence in patient classification, thereby assigning varying degrees of importance to the different genes. This metric reflects the frequency by which each gene is employed as a decision criterion in our predictive models, thus showing its significance toward patient outcomes. We also appreciated the suggestion of Tan and colleagues regarding the time-dependent ROC analysis. Considering that our clustering approach inherently stratifies patients on the basis of gene expression patterns leading to distinct groups with built-in prognostic differences, we believe that a time-dependent ROC analysis is not suitable for our current workflow. Of note, we carried out our tests by (i) exploiting the k-cross validation, (ii) computing multiple evaluation metrics (accuracy, precision and recall) and (iii) showing the confusion matrices for the different cases. This methodology guarantees a statistically robust estimation of classification model performances as well as a complete explainability in terms of true positive, true negative, false positive and false negative rates. Overall, we recognize the value of the analyses proposed by Tan et al., therefore it would be useful to take into consideration their suggestions for subsequent studies.

Acknowledgements
Not applicable.

Author contributions
All authors contributed equally, read and approved the final manuscript.

Funding
Fondazione AIRC supported E.M.D.F. (Start-Up Grant 21651), A.B. (IG n. 23369), R.L. (IG n. 27386). Ministero della Salute (Italy) supported A.B., M.M. and R.L. (RF-2019-12368937). Ministero dell'Università e Ricerca supported E.M.D.F. (Prin 2022 PNRR P2022MALRP), A.B. (Prin 2022 2022Y79PT4), M.M. (Prin 2022 2022Y79PT4) and R.L. (Prin 2022 202282CMEA; Prin 2022 PNRR P2022MALRP). This work was also funded by: (1) The Next Generation EU - project Tech4You - Technologies for climate change adaptation and quality of life improvement, n. ECS0000009; (2) The National Plan for NRRP Complementary Investments - project n. PNC0000003 - AdvaNced Technologies for Human-centrEd Medicine (ANTHEM); (3) The Next Generation EU - Project Age-It: "Ageing Well in an Ageing Society" [DM 1557 11.10.2022]; (4) POS RADIOAMICA project funded by the Italian Minister of Health (CUP: H53C22000650006); (5) POS CAL.HUB. RIA project funded by the Italian Minister of Health (CUP H53C22000800006); (6) Proof of Concept (PoC) - Patent Enhancement Program Unical Pathways (UP) (CUP C28H23000330002).

Data availability
Not applicable.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
All the authors agree to publish this paper.

Competing interests
Not applicable.

Received: 31 July 2024 / Accepted: 4 August 2024
Published online: 16 September 2024

References

1. Tan G, Shao Y, Zhu Z, Letter. Sample preparation before sequencing and bioinformatics analysis method selection are more important than the results. J Transl Med. 2024;22(1):699.
2. Talia M, Cesario E, Cirillo F, Scordamaglia D, Di Dio M, Zicarelli A, Mondino AA, Occhiuzzi MA, De Francesco EM, Belfiore A, Miglietta AM, Di Dio M, Capalbo C, Maggiolini M, Lappano R. Cancer-associated fibroblasts (CAFs) gene signatures predict outcomes in breast and prostate tumor patients. J Transl Med. 2024;22(1):597.
3. Risbridger GP, Davis ID, Birrell SN, Tilley WD. Breast and prostate cancer: more similar than different. Nat Rev Cancer. 2010;10(3):205–12.
4. Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A, Szcześniak MW, Gaffney DJ, Elo LL, Zhang X, Mortazavi A. A survey of best practices for RNA-seq data analysis. Genome Biol. 2016;17:13.
5. Na W, Lee IJ, Koh I, Kwon M, Song YS, Lee SH. Cancer-specific functional profiling in microsatellite-unstable (MSI) colon and endometrial cancers using combined differentially expressed genes and biclustering analysis. Med (Baltim). 2023;102(19):e33647.

Publisher's Note