Visual Clustering of Transcriptomic Data from Primary and Metastatic Tumors – Dependencies and Novel Pitfalls

André Marquardt[1,2,3], Philip Kollmannsberger[4], Markus Krebs[1,5], Markus Knott[6,7], Antonio Giovanni Solimando[8,9]* & Alexander Kerscher[1]*

[1] Comprehensive Cancer Center Mainfranken, University of Würzburg, 97080 Würzburg, Germany.
[2] Institute of Pathology, University of Würzburg, 97080 Würzburg, Germany.
[3] Bavarian Cancer Research Center (BZKF), 97080 Würzburg, Germany.
[4] Center for Computational and Theoretical Biology, University of Würzburg, 97074 Würzburg, Germany.
[5] Department of Urology and Pediatric Urology, University Hospital Würzburg, 97080 Würzburg
[6] Klinikum Stuttgart, Department of Hematology, Oncology, Stem Cell Transplantation and Palliative Care, 70174 Stuttgart, Germany
[7] Klinikum Stuttgart, Stuttgart Cancer Center – Tumor Unit Eva Mayr-Stihl, 70174 Stuttgart, Germany
[8] Guido Baccelli Unit of Internal Medicine, Department of Biomedical Sciences and Human Oncology, School of Medicine, Aldo Moro University of Bari, 70124 Bari, Italy.
[9] IRCCS Istituto Tumori “Giovanni Paolo II” of Bari, 70124 Bari, Italy.

* joint senior authors.

1. Abstract

Personalized Oncology is a rapidly evolving area and offers cancer patients therapy options more specific than ever. Yet, there is still a lack of understanding regarding transcriptomic similarities or differences of metastases and corresponding primary sites. Approaching this question, we used two different unsupervised dimension reduction methods – t-SNE and UMAP – on three different metastases datasets – prostate cancer, neuroendocrine prostate cancer, and skin cutaneous melanoma – including 682 different samples, with three different underlying data transformations – unprocessed FPKM values, log10 transformed FPKM values, and log10+1 transformed FPKM values – to visualize potential underlying clusters. The approaches resulted in formation of different clusters that were independent of respective resection sites. Additionally, data transformation critically affected cluster formation in most cases. Of note, our study revealed no tight link between the metastasis resection site and specific transcriptomic features. Instead, our analysis demonstrates the dependency of cluster formation on the underlying data transformation and the dimension reduction method applied. These observations propose data transformation as another key element in the interpretation

of visual clustering approaches apart from well-known determinants such as initialization and parameters. Furthermore, the results show the need for further evaluation of underlying data alterations based on the biological question and subsequently used methods and applications.

2. Introduction

From a clinical perspective different metastatic patterns obviously occur for various entities (***Wu et al., 2017***). Depending on the location, the site of metastasis has a considerable effect on patients' prognosis. For example, liver metastases from a pancreatic adenocarcinoma are prognostically worse than lymph node or lung metastases (***Liu et al., 2020; Thomas et al., 2012***). Still, it is unclear whether there is a biological or genetic determination for tumors to develop regional metastases or even distant metastasis with preferred target regions (***Nishizaki et al., 1997; Siraj et al., 2020; Yachida et al., 2010***).

From a biological point of view, tumor cells develop through clonal evolution, favoring tumor heterogeneity reflected by different driver mutations or genomic alterations. Due to the accumulation of different alterations metastases can occur, but the circumstances of local or distant metastases still need to be fully answered, as it was already shown that distant metastasis can develop without previous local metastasis (***Esmaeli, 2001***). Additionally, transcriptomic differences of the linear (***Cairns, 1975***) and the parallel progression (***Klein***, C. A.***, 2009***) model are still not fully clarified.

Due to the advances in personalized oncology, comprehensive elucidation of primary tumors and metastases is an ongoing process. Still, the determination of possible differences or similarities between metastasis – especially from several different resection sites – and primary tumors is an unmet need, especially regarding the transcriptome. Previous approaches frequently analyzed the differences between two specific groups (e.g. bone and brain metastasis (***Klein***, A. ***et al., 2009***), or the mutational evolution – yet showing a high concordance of primary and metastatic tumors (***Brannon et al., 2014; Goswami et al., 2015; Vignot et al., 2015***). As a result, a comprehensive study of the transcriptomic characteristics of metastases and corresponding primary tumors is necessary.

Visual clustering – based on data dimension reduction methods – is one potential approach to determine transcriptomic differences and proximities of different metastasis sites. The visualization methods mainly used for this purpose - t-SNE (***Laurens van der Maaten and Geoffrey Hinton, 2008***) and UMAP (***McInnes et al., 09.02.2018***) - have already been widely applied in the field of single cell sequencing (***Cillo et al., 2020; Puram et al., 2017; Zhang et al., 2021***) but also bulk RNA-sequencing (***Marquardt et al., 2021a; Marquardt et al., 2021b; Zhao et al., 2020; Zheng et al., 2021***) to visually separate transcriptionally similar cell

populations from other diverging populations. Furthermore, there a recent studies also showing the critical impact of the initialization (***Kobak and Linderman, 2021***) and used parameters (***Kobak et al., 2020***) on dimension reduction methods.

To now search for transcriptomic dependencies caused by the site of metastasis, we used both t-SNE and UMAP to analyze three different metastasis datasets – prostate cancer (PCa), neuroendocrine PCa, and skin cutaneous melanoma – consisting of 682 samples in total. For a more comprehensive analysis, we used unprocessed data – the FPKM values obtained after normalization of the mapped sequencing reads – but also to two different data transformations – log10 and log10+1 – as log transformations are commonly used when analyzing gene expressions.

3. Material and Methods

3.1. Data acquisition

For our analyses, we used RNA sequencing data from three different metastasis datasets. The first dataset contains n=266 samples from metastatic prostate carcinoma (PRAD-SU2C - Dream Team (***Abida et al., 2019***)), the second consists of n=49 samples from metastatic neuroendocrine prostate carcinomas (NEP-WCM (***Beltran et al., 2016***), and the third dataset consists of n=367 metastatic melanomas (TCGA-SKCM-Metastatic (***Rehan Akbani et al., 2015***)). All datasets indicate the site of resection, which serves as the basis for further analyses.

For evaluation purposes, we used an additional dataset known to form distinct clusters based on histopathologic subgroups. The TCGA-KIPAN dataset, consisting of three renal cell carcinoma (RCC) subgroups–TCGA-KIRC (clear cell RCC, n=538), TCGA-KIRP (papillary RCC, n=288), and TCGA-KICH (chromophobe RCC, n=65).

For further testing similarities and differences between primary and metastatic tumors, we used the complete TCGA-SKCM dataset – adding 103 primary tumor samples for a total of n=470 samples – and a metastatic breast cancer dataset (MBCproject cBioPortal (***Cerami et al., 2012; Gao et al., 2013***) data version February 2020 (***Wagle et al., 2016***)) consisting of n=120 primary tumor and n =26 metastatic samples.

3.2. Bioinformatic Analysis

We used both t-SNE plots and UMAPs for the analysis, combined with three data transformation approaches, to get a more comprehensive view of the data. First, we used the unprocessed FPKM values – obtained by normalizing the mapped sequencing reads – log10 transformed values, and log10+1 transformed values. The so-called log10 transformed values, are the log10 values of the unprocessed FPKM values, where values equal to 0 are set to 0.

The so-called log10+1 transformed values are the log10 values which are obtained after the unprocessed FPKM value plus 1 has been calculated.

Subsequently, all data obtained were examined and compared with the t-SNE plot as well as with UMAP - both of which represent the high-dimensional data in 2D space. t-SNE plots were created based on a PCA with 50 components, a learning rate of 300, a perplexity of 27. Further details on the procedure are given elsewhere (***Marquardt et al., 2021b***). UMAP plots were generated based on an adapted UMAP approach as previously described (***Marquardt et al., 2021a***), using raw values as input. In brief, the squared pairwise Euclidean distance was used to calculate the distance between samples with a subsequent binary search for the optimal rho based on a fixed number of 15 nearest neighbors. The symmetry calculation was simplified, by just dividing the sum of probabilities by 2. Furthermore, mind_dist = 0.25 was used, as well as cross-entropy as cost function with normalized Q parameter. Last, gradient descent learning was used with 2 dimensions and 50 neighbors, were applicable (NEP-WCM dataset used 25 neighbors). After generating the unbiased low-dimensional representations of the high-dimensional input (RNA-sequencing) data we performed manual cluster interpretation.

4. Results

4.1. Analysis of the PRAD-SU2 (Dream Team) Dataset

The first dataset in our analysis represented prostate carcinoma. In the t-SNE plots, up to three clusters were observable according to data transformation applied. Using t-SNE, the raw FPKM values resulted in one additional visible cluster (**Figure 1** A), whereas log10 (**Figure 1** B) and log10+1 (**Figure 1** C) approaches showed two additional smaller clusters. These clusters mainly contained bone or liver samples and were named accordingly.

The UMAP approach showed similar results compared to the t-SNE approach, where the unprocessed FPKM values (**Figure 1** D) did not provide any clustering information. The log10 (**Figure 1** E) and log10+1 (**Figure 1** F) transformations showed three visible and distinct clusters. Again, one of these clusters consisted of only bone samples, one consisted mainly of liver samples and the last and largest cluster consisted of all remaining samples. These results indicate that the resection site was not the main cause for clustering, instead visualization techniques (t-SNE vs. UMAP) and data transformation (unprocessed vs. log10 vs. log10+1 transformed data) heavily affected clustering results.

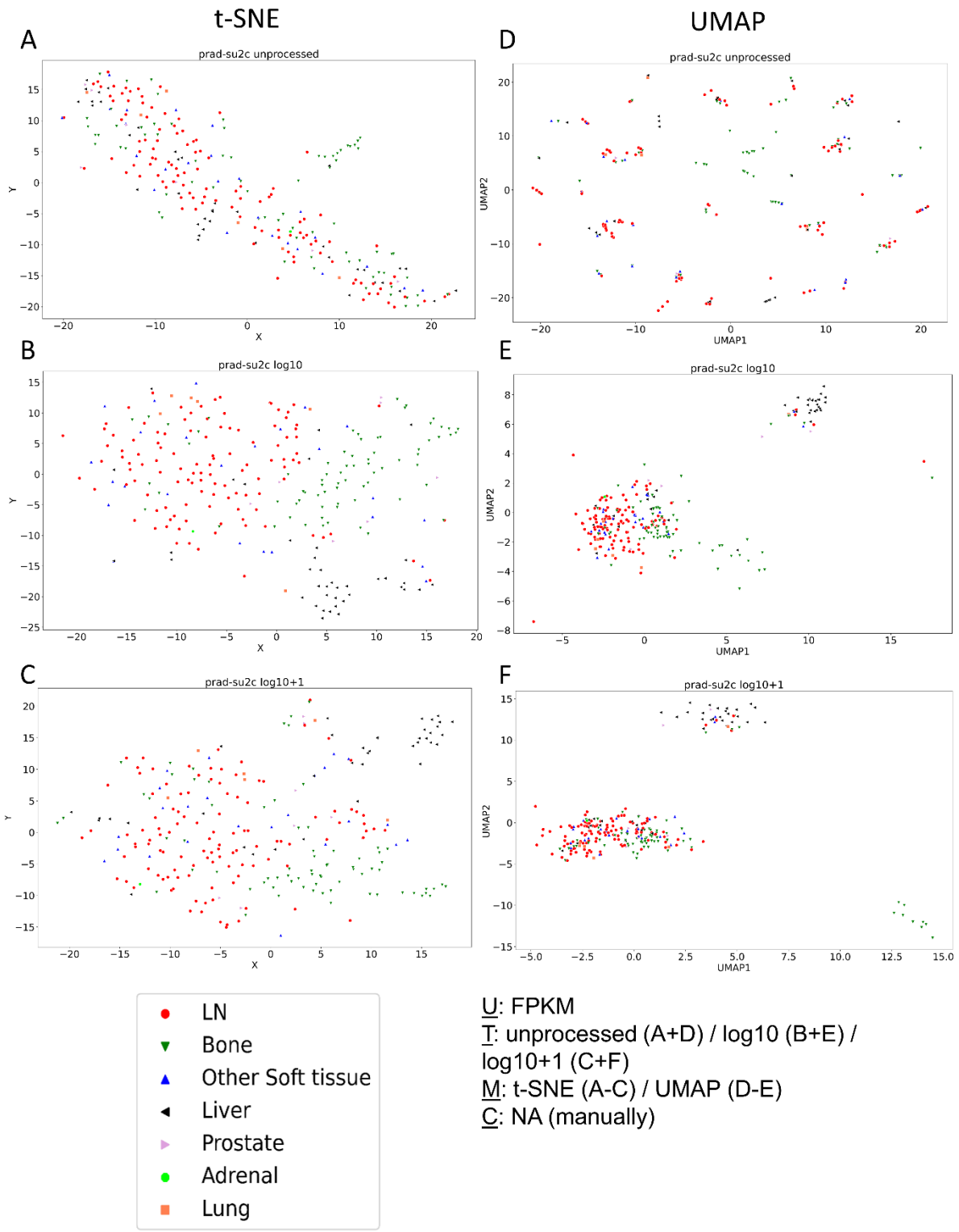


Figure 1 Visual clustering of the DreamTeam dataset*: Dataset consisting of metastatic prostate cancers with indicated resection site, applying different data dimension reduction methods. t-SNE plot approach for (A) unprocessed, (B) log10 transformed, and (C) log10+1 transformed FPKM values and UMAP approach using (D) unprocessed, (E) log10 transformed, and (F) log10+1 transformed FPKM values. FPKM: Fragments Per Kilobase Million; U: Unit, T: Transformation, M: data dimension reduction Method, C: Clustering method, NA: not applicable.*

4.2. Analysis of the NEPC WCM (Neuroendocrine Prostate Cancer) Dataset

The second dataset represented neuroendocrine prostate cancer. No clusters were detectable using the t-SNE approach (Figure 2 A to C), whereas the UMAP approach consistently resulted in three distinct clusters (Figure 2 D-F). Of note, resulting clusters were very similar throughout all applied data transformations – thereby not displaying any resection site specificities.

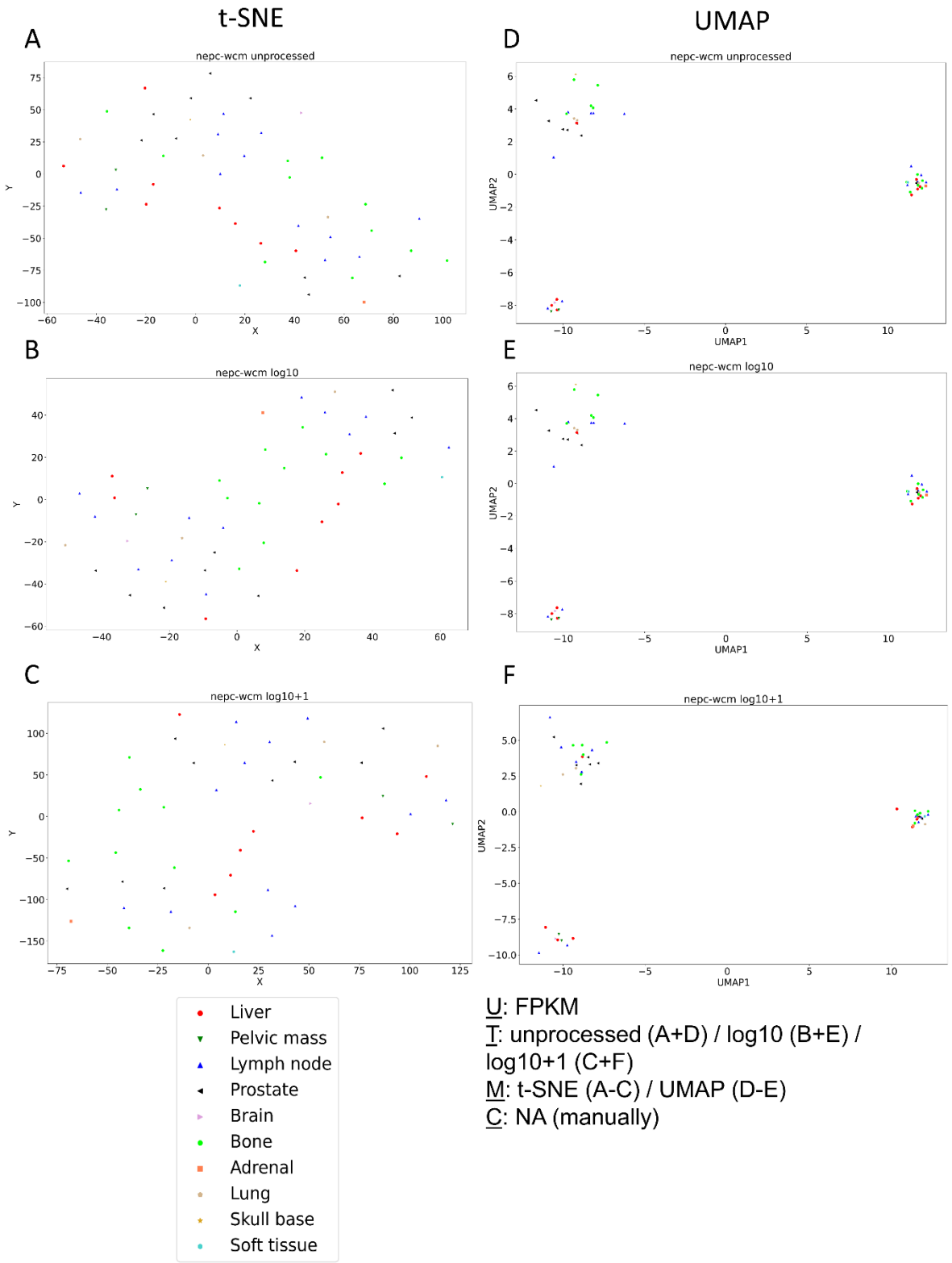


Figure 2 ***Visual clustering of the NEPC WCM dataset****: Dataset consisting of neuroendocrine metastatic prostate cancers with indicated resection site, applying different data dimension reduction methods. t-SNE plot approach for (A) unprocessed, (B) log10 transformed, and (C) log10+1 transformed FPKM values and UMAP approach using (D) unprocessed, (E) log10 transformed, and (F) log10+1 transformed FPKM values. FPKM: Fragments Per Kilobase Million; U: Unit, T: Transformation, M: data dimension reduction Method, C: Clustering method, NA: not applicable.*

4.3. Analysis of the Metastatic Samples of TCGA-SKCM Dataset

The SKCM-TCGA dataset served as the third dataset. Again, no clusters were detectable using t-SNE plots with the different data transformations (Figure 3 A to C). Considering the UMAP approaches, unprocessed FPKM values did not provide any useful clustering information (Figure 3 D). The log10 transformed values formed one large cluster containing nearly all samples with only few outliers (Figure 3 E). Only the log10+1 transformed values formed distinct clusters without site-specific agglomeration (Figure 3 F), again showing the critical impact of data transformation on cluster formation.

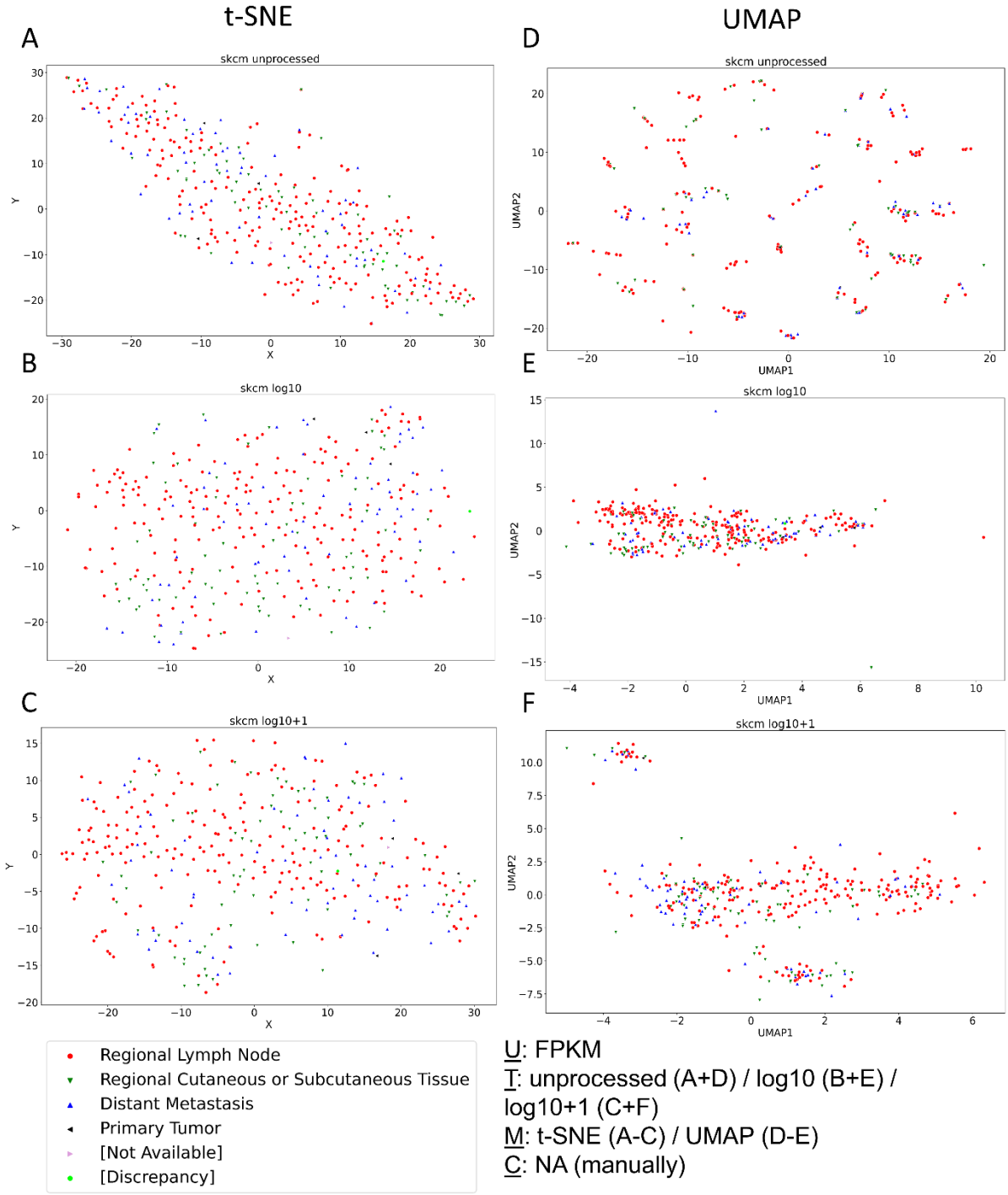


Figure 3 ***Visual clustering of themetastatic TCGA-SKCM dataset****: Dataset consisting of the metastatic sampels of the TCGA-SKCM dataset (skin cutanous melanoma) with indicated resection site, applying different data dimension reduction methods. t-SNE plot approach for (A) unprocessed, (B) log10 transformed, and (C) log10+1 transformed FPKM values and UMAP approach using (D) unprocessed, (E) log10 transformed, and (F) log10+1 transformed FPKM values. FPKM: Fragments Per Kilobase Million; U: Unit, T: Transformation, M: data dimension reduction Method, C: Clustering method, NA: not applicable.*

In summary, in none of the three datasets a continuous dependence of the resection site could be seen. Instead, we observed a strong dependence of the visual cluster formation on the applied method and data transformation. Only one small bone cluster in the Dream Team dataset was detectable independent of data transformation and dimension reduction approach. To validate our results and to test previous observations of subgroup-dependent clustering, we additionally considered the KIPAN dataset, consisting of three known biologically distinct subgroups of renal cell carcinoma (RCC).

4.4. Further Evaluation of Cluster Formation based on Data Dimension Reduction Methods and Data Transformations

To investigate the influence of data transformation and data dimension reduction methods on forming visually distinct clusters, the TCGA data sets of the three largest RCC subgroups – clear cell (KIRC), papillary (KIRP), and chromophobe (KICH) – were combined to one data set (KIPAN). Due to the nature of the histopathologic origin of the dataset, a specific clustering was expected.

t-SNE (Figure 4 B+C) and UMAP (Figure 4 E+F) approaches based on log10 or the log10+1 transformed data yielded a separation of samples matching the histopathologic expectation. Further, the importance and clinical relevance of subgroups identified by t-SNE using unprocessed data for the TCGA-KIPAN dataset has previously been published (Figure 4 A).Yet, the unprocessed FPKM values yielded no useful information regarding the resection site specific agglomeration of samples in the UMAP approach (Figure 4 D).

Although both applied data transformations showed a separation based on the histopathological subgroups for both data dimension reduction methods, clusters were not exclusively subgroup-specific and displayed certain outliers.

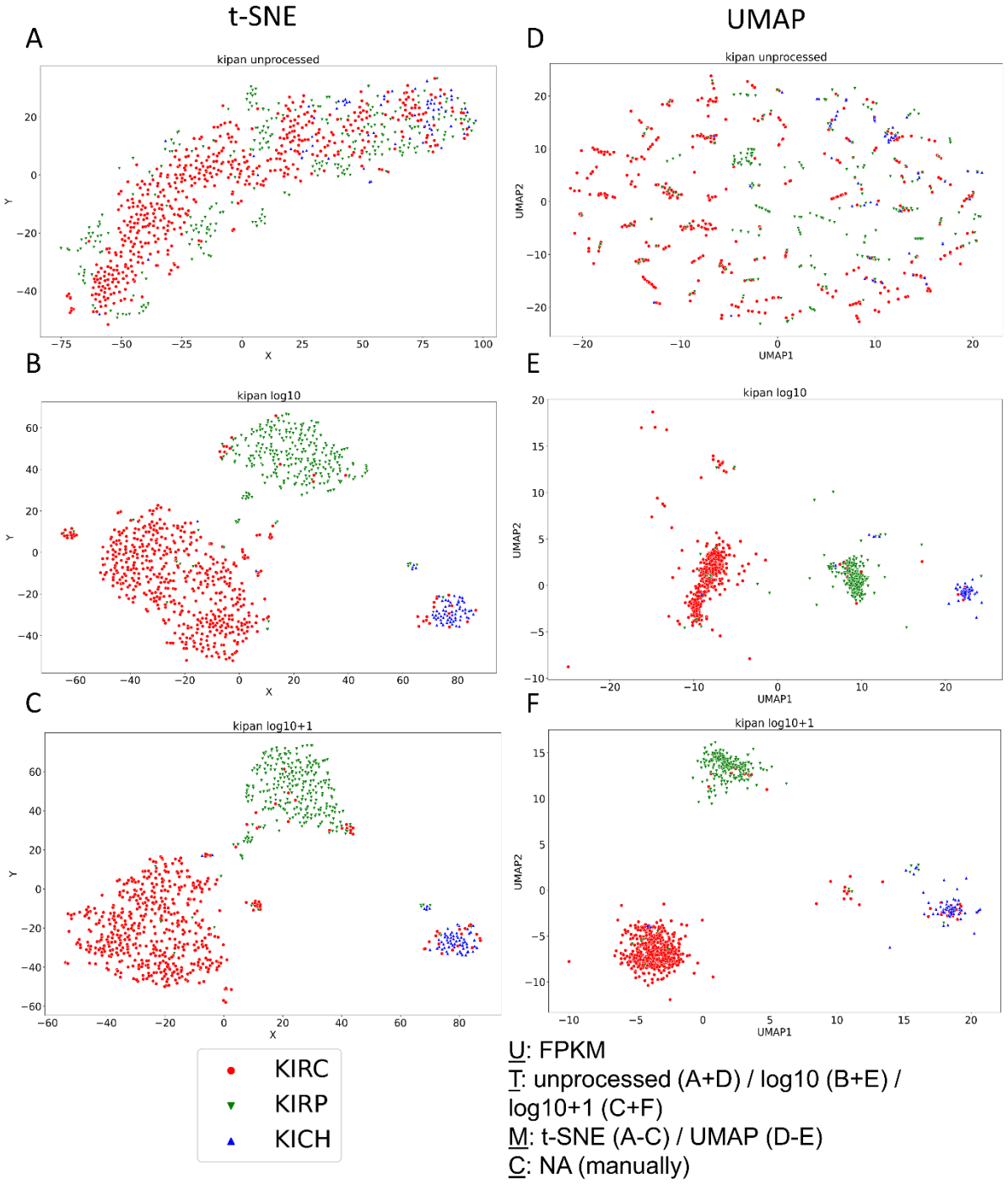


Figure 4 ***Visual clustering of the TCGA-KIPAN dataset****: Dataset consisting of the three major histopatholgic subgroups of renal cell carcinoma (RCC), namely clear cell RCC (KIRC), papillary RCC (KIRP), and chromophobe RCC (KICH) , applying different data dimension reduction methods. t-SNE plot approach for (A) unprocessed, (B) log10 transformed, and (C) log10+1 transformed FPKM values and UMAP approach using (D) unprocessed, (E) log10 transformed, and (F) log10+1 transformed FPKM values. FPKM: Fragments Per Kilobase Million; U: Unit, T: Transformation, M: data dimension reduction Method, C: Clustering method, NA: not applicable.*

4.5. Combined Analysis of Primary and Metastatic Samples of the same Entity

Motivated by the results of the KIPAN cohort, we performed all analyses for the complete TCGA-SKCM dataset as well – in order to analyze the transcriptomic relation of primary and metastatic melanoma. As shown in Figure 5, there was no distinct separation between metastatic and primary melanoma samples. Moreover, only the UMAP log10+1 transformed approach displayed two distinct clusters (each containing primary and metastatic samples, Figure 5 F). For both clusters, no complete subgroup-specific (primary tumor vs. metastasis) clustering resulted. This indicates that there are differences between primary and metastatic tumors with respect to their transcriptomic features, but without previous knowledge of the subgroup, no assumptions could be made in separating both groups. Moreover, some metastatic tumors seem to still harbour primary tumor transcriptomic features, whereas there are also primary tumors already harbouring metastatic tumor transcriptome features.

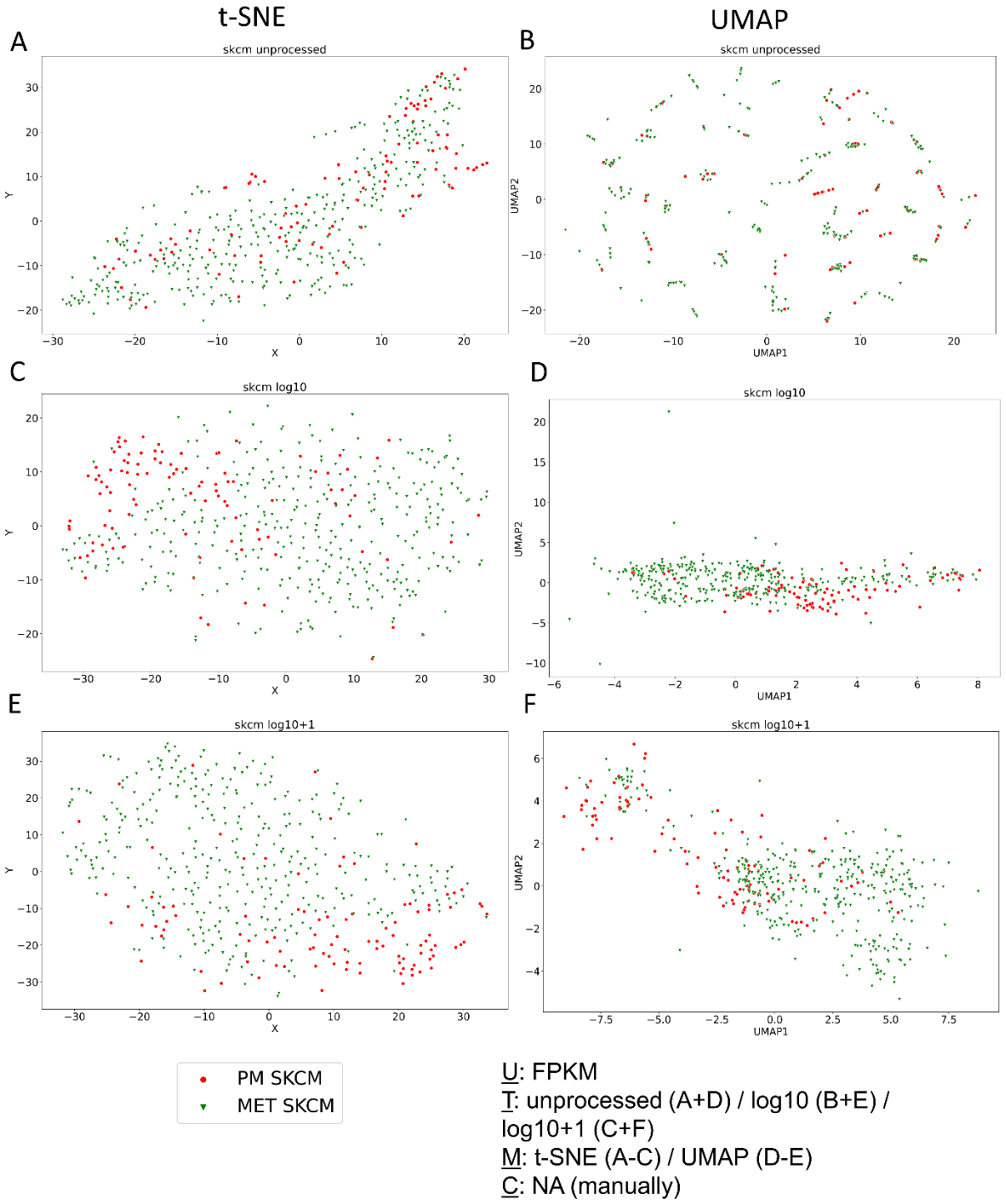
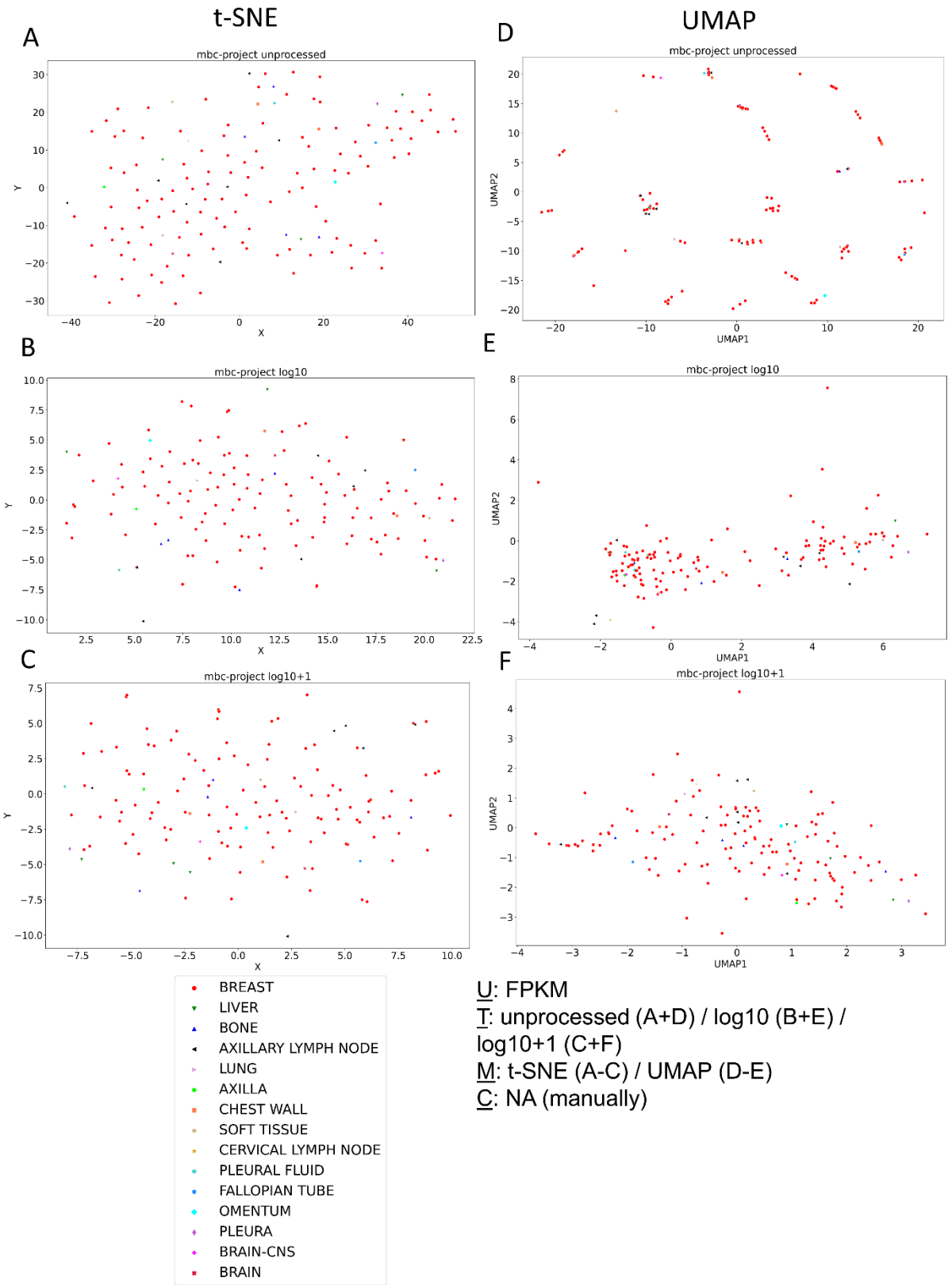


Figure 5 ***Visual clustering of the complete TCGA-SKCM dataset****: Dataset consisting of primary tumors (PM SKCM – red) and metastatic tumors (MET SKCM - green), applying different data dimension reduction methods. t-SNE plot approach for (A) unprocessed, (B) log10 transformed, and (C) log10+1 transformed FPKM values and UMAP approach using (D) unprocessed, (E) log10 transformed, and (F) log10+1 transformed FPKM values. FPKM: Fragments Per Kilobase Million; U: Unit, T: Transformation, M: data dimension reduction Method, C: Clustering method, NA: not applicable.*

To further validate these results, we also analyzed the metastatic breast cancer project (MBC Project) dataset, consisting of both primary and metastatic tumors of different resection sites. Using the t-SNE approach on this dataset did not lead to any cluster formation for any of the used data transformations (Figure 6 A to C). The unprocessed FPKM values in combination with the UMAP again did not provide any useful information about the dataset (Figure 6 D). Additionally, the logarithmic transformations in the UMAP approaches do not form any distinct clusters, either (Figure 6 E+F) – thereby confirming the findings from the TCGA-SKCM dataset.



U: FPKM
T: unprocessed (A+D) / log10 (B+E) / log10+1 (C+F)
M: t-SNE (A-C) / UMAP (D-E)
C: NA (manually)

Figure 6 ***Visual clustering of the MBC Project dataset****: Dataset consisting of the primary and metastatic breast cancer samples with indicated resection site,, applying different data dimension reduction methods. t-SNE plot approach for (A) unprocessed, (B) log10 transformed, and (C) log10+1 transformed FPKM values and UMAP approach using (D) unprocessed, (E) log10 transformed, and (F) log10+1 transformed FPKM values. FPKM: Fragments Per Kilobase Million; U: Unit, T: Transformation, M: data dimension reduction Method, C: Clustering method, NA: not applicable*

5. Discussion

5.1. The Impact of Data Transformations on Cluster Formation in Data Dimension Reduction Methods

In this approach, we were looking for transcriptomic similarities and differences of metastasis representing different resection sites. It has already been shown in several studies that there is no clustering of samples depending on the underlying resection site of the metastasis (***Brasó-Maristany et al., 2021***). Nevertheless, within these studies, clustering was frequently observed (***Cejalvo et al., 2017***). These clusters were often attributed to biologically distinct subgroups in one entity – also stating preferred metastasis sites for different subgroups (***Wu et al., 2017***). Additionally, there are studies showing transcriptional differences between two different resection sites (***Klein***, A. ***et al., 2009***). Due to this, we compared the clustering results of three different datasets. Since previous analyses did not specifically investigate the transcriptomic dependency of the resection site, our approach considered not only different unbiased data dimension reduction methods – subsequently used for visual clustering – but also different data transformations. It could be observed that especially log10+1 transformed data frequently resulted in a clearer and more distinct cluster formation when analysed with UMAP. In line with this observation, UMAP analysis of the TCGA-KIPAN dataset showed a cluster dependency mainly based on underlying RCC histopathology. However, histopathological clustering was evident in the UMAP log10+1 and in the UMAP log10 data transformations and the results of the corresponding t-SNE approaches.

We already showed in a previous publication, that clusters obtained by using unprocessed FPKM values in a t-SNE approach yielded prognostically relevant and biologically distinct characteristics for RCC (***Marquardt et al., 2021b***). These findings were also in line with previous literature (***Zhou et al., 2014***). Additionally, using UMAP data dimension reduction with logarithmically transformed data of the TCGA-ACC (adrenocortical carcinoma) dataset revealed two clusters that closely matched the already known ACC subgroups (***Marquardt et al., 2021a***). This suggests that histopathological and cancer subgroup-specific differences can be represented with a UMAP log10+1 approach, even though clusters seen within TCGA-KIPAN analysis were not completely subgroup-specific, which can also be observed in t-SNE plot using unprocessed data. Since t-SNE and UMAP show biologically meaningful clustering

results – known histopathological or cancer entity subgroups – based on different data transformations, both data dimension reduction methods are useable and valid, depending on the underlying biological question.

Another remarkable element is the bone cluster we named within the Dream Team dataset. This cluster appears with minor changes and depending on the area considered in each of our analyses, regardless of the data transformation and the data dimension reduction method. Taking into account previous results, we conclude that all present methods have their justification and can be used depending on the research question. For example, the UMAP log10+1 approach is suitable for bulk RNA sequencing to identify subgroups within specific entities. However, clusters based on different histopathological tissues, for example, and thus generally showing a quite different transcriptome, can also be seen in the unprocessed data, where the t-SNE plot seems to be more suitable for bulk RNA sequencing than UMAP, which does not seem to be suitable for unprocessed data of bulk RNA-sequencing in general.

When looking at the number of clusters previously identified for our datasets, it becomes apparent that log10+1 approaches were mostly in line with previously shown results. For the TCGA-SKCM dataset, three clusters were identified in in the first publication of this dataset (***Rehan Akbani et al., 2015***). The initial description of the NEP-WCM dataset showed three (based on the main branches of the dendrogram) different main clusters based on unsupervised clustering, overlapping with our UMAP results. Additionally, they described a smaller neuroendocrine subgroup inside the Dream Team dataset, which might be one of the shown clusters in our approach. This further proves the clusters found by unsupervised clustering in the Dream Team original publication, stating the independence of the metastasis site (***Abida et al., 2019***), and confirming the different molecular phenotype of neuroendocrine prostate cancers (***Beltran et al., 2011***). Regarding breast cancer metastases, our results confirm previous findings showing the cluster dependency on biological subgroups rather than on resection site (***Brasó-Maristany et al., 2021***).

Looking more closely at the differences in the resulting clusters between the different data transformations of the individual data dimension reduction methods, changes are similar to the ones caused by parameters such as the number of neighbors. Considering very recent research, we believe that the data transformation used is just an equally important factor to consider in the initialization of the data (***Kobak and Linderman, 2021***), as respective kernel transformations (***Kobak et al., 2020***).

5.2. Primary Tumors and Metastases of the same Entity share Common Transcriptomic Features

Our findings that primary and metastatic tumors share common transcriptomic features and are inseparable when analysed with data dimension reduction methods, overlaps with findings of previous research. In metastatic pancreas carcinoma samples, a distinction between primary tumor and metastatic tumor cells was not possible using single cell RNA-sequencing (***Lin et al., 2020; Pan et al., 2021***). This could also be seen in breast cancer single cell RNA-sequencing comparing lymph node metastasis with primary tumors (***Xu et al., 2021***), which is totally in line with our findings regarding the MBC project dataset, not forming visual clusters in any considered approach. The presented results support the linear progression model to some extent, at least for the transcriptomic differences between metastasis and primary tumor, indicating the need for further research to combine genomic alterations with transcriptomic features to clarify the (clonal) evolution of metastasis. In conclusion, our results suggest that there is no general transcriptomic resection site dependency for metastasis of the same primary tumor and that obtained clusters can be mostly attributed to existing subgroups. The genetic diversity, using bulk sequencing and analytical deconvolution is a major hallmark of cancer in general. Pre-metastatic and pre-treatment diversity can help to predict the clinical and evolutionary outcome of the disease. Nonetheless, the regulatory wiring that underpins the metastatic process is likely to dynamically change across the transcriptional landscape.

5.3. Addressing Pitfalls in Visual Clustering

To address these challenges, we propose an additional standard-legend for visual clustering approaches based on data dimension reduction methods and machine learning, as represented by the UTMC legend in this work. The information required by this new standard includes the unit (**U**) (such as FPKM, TPM, RPKM or read counts), data transformations (**T**), represented visualization or data dimension reduction method (**M**) and if applicable the applied cluster identification algorithm (**C**). This enables reproducibility of illustrations and analyses and makes visual clustering approaches much more transparent.

Taken together, our work further extends the knowledge of tumor heterogeneity in different biological context (***Russano et al., 2020***), by providing sufficient evidence for the linear progression model of metastasis, since no dependency of clusters based on resection site was observable in any of the three considered data sets. The applied transformation tended to have the biggest impact on clustering results, and thus needs more in-depth analysis. Nevertheless, our results cannot identify a favourite approach, as all answer different questions. Using transformed data, independent of the data dimension reduction method, tend to visualize

subgroups very specifically, whereas using unprocessed data in t-SNE seems to be closer to the biological nature of samples, demonstrating the need for further research in this area.

6. Conclusion

In summary, using two different data dimension reduction methods, we showed that there was no visual association between resection site and the transcriptome for three considered metastatic datasets. Instead, there was a significant dependence of clustering according to data transformation and the data dimension reduction method applied. Additionally, the analysis of primary and metastatic samples of specific entities did not show distinct clusters or visible differences. Combining recent works and the results of our study, visual clustering seems highly vulnerable towards data and parameter alterations. To avoid pitfalls in analyzing visual clustering and to enhance reproducibility. We recommend using standardized nomenclature, such as the UTMC-legend introduced in this manuscript.

Acknowledgments

The results shown here are in part based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga.

The results included here include the use of data from The Metastatic Breast Cancer Project (https://www.mbcproject.org/), a project of Count Me In (https://joincountmein.org/).

Data Availability

All datasets in this study are publicly available. Datasets were either accessed via GDC-portal (https://portal.gdc.cancer.gov/projects) or cBioPortal (https://www.cbioportal.org/) (***Cerami et al., 2012; Gao et al., 2013***).

Funding

This project was supported in part by the Apulian Regional Project Medicina di Precisione to AGS. Moreover, MK was funded by a personal grant from Else-Kröner-Foundation (Else Kröner Integrative Clinician Scientist College for Translational Immunology, University Hospital Würzburg, Germany).

References

Abida W, Cyrta J, Heller G, Prandi D, Armenia J, Coleman I, Cieslik M, Benelli M, Robinson D, van Allen EM, Sboner A, Fedrizzi T, Mosquera JM, Robinson BD, Sarkar N de, Kunju LP, Tomlins S, Wu YM, Nava Rodrigues D, Loda M, Gopalan A, Reuter VE, Pritchard CC, Mateo J, Bianchini D, Miranda S, Carreira S, Rescigno P, Filipenko J, Vinson J, Montgomery RB, Beltran H, Heath EI, Scher HI, Kantoff PW, Taplin M-E, Schultz N, deBono JS, Demichelis F, Nelson PS, Rubin MA, Chinnaiyan AM, Sawyers CL. 2019. Genomic correlates of clinical outcome in advanced prostate cancer. *Proceedings of the National Academy of Sciences* **116**:11428–11436. doi: 10.1073/pnas.1902651116.

Beltran H, Prandi D, Mosquera JM, Benelli M, Puca L, Cyrta J, Marotz C, Giannopoulou E, Chakravarthi BVSK, Varambally S, Tomlins SA, Nanus DM, Tagawa ST, van Allen EM, Elemento O, Sboner A, Garraway LA, Rubin MA, Demichelis F. 2016. Divergent clonal evolution of castration-resistant neuroendocrine prostate cancer. *Nature Medicine* **22**:298–305. doi: 10.1038/nm.4045.

Beltran H, Rickman DS, Park K, Chae SS, Sboner A, MacDonald TY, Wang Y, Sheikh KL, Terry S, Tagawa ST, Dhir R, Nelson JB, La Taille A de, Allory Y, Gerstein MB, Perner S, Pienta KJ, Chinnaiyan AM, Wang Y, Collins CC, Gleave ME, Demichelis F, Nanus DM, Rubin MA. 2011. Molecular characterization of neuroendocrine prostate cancer and identification of new drug targets. *Cancer Discovery* **1**:487–495. doi: 10.1158/2159-8290.CD-11-0130.

Brannon AR, Vakiani E, Sylvester BE, Scott SN, McDermott G, Shah RH, Kania K, Viale A, Oschwald DM, Vacic V, Emde A-K, Cercek A, Yaeger R, Kemeny NE, Saltz LB, Shia J, D'Angelica MI, Weiser MR, Solit DB, Berger MF. 2014. Comparative sequencing analysis reveals high genomic concordance between matched primary and metastatic colorectal cancer lesions. *Genome Biology* **15**:454. doi: 10.1186/s13059-014-0454-7.

Brasó-Maristany F, Paré L, Chic N, Martínez-Sáez O, Pascual T, Mallafré-Larrosa M, Schettini F, González-Farré B, Sanfeliu E, Martínez D, Galván P, Barnadas E, Salinas B, Tolosa P, Ciruelos E, Carcelero E, Guillén C, Adamo B, Moreno R, Vidal M, Muñoz M, Prat A. 2021. Gene expression profiles of breast cancer metastasis according to organ site. *Molecular oncology*. doi: 10.1002/1878-0261.13021.

Cairns J. 1975. Mutation selection and the natural history of cancer. *Nature* **255**:197–200. doi: 10.1038/255197a0.

Cejalvo JM, Martínez de Dueñas E, Galván P, García-Recio S, Burgués Gasión O, Paré L, Antolín S, Martinello R, Blancas I, Adamo B, Guerrero-Zotano Á, Muñoz M, Nucíforo P, Vidal M, Pérez RM, Chacón López-Muniz JI, Caballero R, Peg V, Carrasco E, Rojo F, Perou CM, Cortés J, Adamo V, Albanell J, Gomis RR, Lluch A, Prat A. 2017. Intrinsic Subtypes and Gene Expression Profiles in Primary and Metastatic Breast Cancer. *Cancer Research* **77**:2213–2221. doi: 10.1158/0008-5472.CAN-16-2717.

Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, Antipin Y, Reva B, Goldberg AP, Sander C, Schultz N. 2012. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discovery* **2**:401–404.

Cillo AR, Kürten CHL, Tabib T, Qi Z, Onkar S, Wang T, Liu A, Duvvuri U, Kim S, Soose RJ, Oesterreich S, Chen W, Lafyatis R, Bruno TC, Ferris RL, Vignali DAA. 2020. Immune Landscape of Viral- and Carcinogen-Driven Head and Neck Cancer. *Immunity* **52**:183-199.e9. doi: 10.1016/j.immuni.2019.11.014.

Esmaeli B. 2001. Patterns of regional and distant metastasis in patients with conjunctival melanoma Experience at a cancer center over four decades. *Ophthalmology* **108**:2101–2105. doi: 10.1016/S0161-6420(01)00782-5.

Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E, Cerami E, Sander C, Schultz N. 2013. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Science signaling* **6**:pl1.

Goswami RS, Patel KP, Singh RR, Meric-Bernstam F, Kopetz ES, Subbiah V, Alvarez RH, Davies MA, Jabbar KJ, Roy-Chowdhuri S, Lazar AJ, Medeiros LJ, Broaddus RR, Luthra R, Routbort MJ. 2015. Hotspot mutation panel testing reveals clonal evolution in a study of 265 paired primary and metastatic tumors. *Clinical Cancer Research* **21**:2644–2651. doi: 10.1158/1078-0432.CCR-14-2391.

Klein, A., **Olendrowitz**, C., **Schmutzler**, R., **Hampl**, J., **Schlag**, P. M., **Maass**, N., **Arnold**, N., **Wessel**, R., **Ramser**, J., **Meindl**, A., **Scherneck**, S., **Seitz**, S. 2009. Identification of brain- and bone-specific breast cancer metastasis genes. *Cancer Letters* **276**:212–220. doi: 10.1016/j.canlet.2008.11.017.

Klein, C. A. 2009. Parallel progression of primary tumours and metastases. *Nature Reviews Cancer* **9**:302–312. doi: 10.1038/nrc2627.

Kobak D, Linderman G, Steinerberger S, Kluger Y, Berens P. 2020. Heavy-tailed kernels reveal a finer cluster structure in t-SNE visualisations. *Machine learning and knowledge discovery in databases : European Conference, ECML PKDD … : proceedings. ECML PKDD (Conference)* **11906**:124–139. doi: 10.1007/978-3-030-46150-8_8.

Kobak D, Linderman GC. 2021. Initialization is critical for preserving global data structure in both t-SNE and UMAP. *Nature Biotechnology* **39**:156–157. doi: 10.1038/s41587-020-00809-z.

Laurens van der Maaten, **Geoffrey Hinton**. 2008. Visualizing Data using t-SNE. *Journal of Machine Learning Research* **9**:2579–2605.

Lin W, Noel P, Borazanci EH, Lee J, Amini A, Han IW, Heo JS, Jameson GS, Fraser C, Steinbach M, Woo Y, Fong Y, Cridebring D, Hoff DD von, Park JO, Han H. 2020. Single-cell transcriptome analysis of tumor and stromal compartments of pancreatic ductal adenocarcinoma primary tumors and metastatic lesions. *Genome Medicine* **12**:80. doi: 10.1186/s13073-020-00776-9.

Liu Q, Zhang R, Michalski CW, Liu B, Liao Q, Kleeff J. 2020. Surgery for synchronous and metachronous single-organ metastasis of pancreatic cancer: a SEER database analysis and systematic literature review. *Scientific Reports* **10**:4444. doi: 10.1038/s41598-020-61487-0.

Marquardt A, Landwehr L-S, Ronchi CL, Di Dalmazi G, Riester A, Kollmannsberger P, Altieri B, Fassnacht M, Sbiera S. 2021a. Identifying New Potential Biomarkers in Adrenocortical Tumors Based on mRNA Expression Data Using Machine Learning. *Cancers* **13**:4671. doi: 10.3390/cancers13184671.

Marquardt A, Solimando AG, Kerscher A, Bittrich M, Kalogirou C, Kübler H, Rosenwald A, Bargou R, Kollmannsberger P, Schilling B, Meierjohann S, Krebs M. 2021b. Subgroup-Independent Mapping of Renal Cell Carcinoma-Machine Learning Reveals Prognostic Mitochondrial Gene Signature Beyond Histopathologic Boundaries. *Frontiers in Oncology* **11**:621278. doi: 10.3389/fonc.2021.621278.

McInnes L, Healy J, Melville J. 09.02.2018. *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction*.

Nishizaki T, DeVries S, Chew K, Goodson WH, Ljung B-M, Thor A, Waldman FM. 1997. Genetic alterations in primary breast cancers and their metastases: Direct comparison using modified comparative genomic hybridization. *Genes, Chromosomes and Cancer* **19**:267–272. doi: 10.1002/(SICI)1098-2264(199708)19:4<267::AID-GCC9>3.0.CO;2-V.

Pan H, Diao H, Zhong W, Wang T, Wen P, Wu C. 2021. A Cancer Cell Cluster Marked by LincRNA MEG3 Leads Pancreatic Ductal Adenocarcinoma Metastasis. *Frontiers in Oncology* **11**:656564. doi: 10.3389/fonc.2021.656564.

Puram SV, Tirosh I, Parikh AS, Patel AP, Yizhak K, Gillespie S, Rodman C, Luo CL, Mroz EA, Emerick KS, Deschler DG, Varvares MA, Mylvaganam R, Rozenblatt-Rosen O, Rocco JW, Faquin WC, Lin DT, Regev A, Bernstein BE. 2017. Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer. *Cell* **171**:1611-1624.e24. doi: 10.1016/j.cell.2017.10.044.

Rehan Akbani, **Kadir C. Akdemir**, **B. Arman Aksoy**, **Monique Albert**, **Adrian Ally**, **Samirkumar B. Amin**, **Harindra Arachchi**, **Arshi Arora**, **J. Todd Auman**, **Brenda Ayala**, **Julien Baboud**, **Miruna Balasundaram**, **Saianand Balu**, **Nandita Barnabas**, **John Bartlett**, **Pam Bartlett**, **Boris C. Bastian**, **Stephen B. Baylin**, **Madhusmita Behera**, **Dmitry Belyaev**, **Christopher Benz**, **Brady Bernard**, **Rameen Beroukhim**, **Natalie Bir**, **Aaron D. Black**, **Tom Bodenheimer**, **Lori Boice**, **Genevieve M. Boland**, **Riccardo Bono**, **Moiz S. Bootwalla**, **Marcus Bosenberg**, **Jay Bowen**, **Reanne Bowlby**, **Christopher A. Bristow**, **Laura Brockway-Lunardi**, **Denise Brooks**, **Jakub Brzezinski**, **Wiam Bshara**, **Elizabeth Buda**, **William R. Burns**, **Yaron S.N. Butterfield**, **Michael Button**, **Tiffany Calderone**, **Giancarlo Antonini Cappellini**, **Candace Carter**, **Scott L. Carter**, **Lynn Cherney**, **Andrew D. Cherniack**, **Aaron Chevalier**, **Lynda Chin**, **Juok Cho**, **Raymond J. Cho**, **Yoon-La Choi**, **Andy Chu**, **Sudha Chudamani**, **Kristian Cibulskis**, **Giovanni Ciriello**, **Amanda Clarke**, **Stephen Coons**, **Leslie Cope**, **Daniel Crain**, **Erin Curley**, **Ludmila Danilova**, **Stefania D'Atri**, **Tanja Davidsen**, **Michael A. Davies**, **Keith A. Delman**, **John A. Demchok**, **Qixia A. Deng**, **Yonathan Lissanu Deribe**, **Noreen Dhalla**, **Rajiv Dhir**, **Daniel DiCara**, **Michael Dinikin**, **Michael Dubina**, **J. Stephen Ebrom**, **Sophie Egea**, **Greg Eley**, **Jay Engel**, **Jennifer M. Eschbacher**, **Konstantin V. Fedosenko**, **Ina Felau**, **Timothy Fennell**, **Martin L. Ferguson**, **Sheila Fisher**, **Keith T. Flaherty**, **Scott Frazer**, **Jessica Frick**, **Victoria Fulidou**, **Stacey B. Gabriel**, **Jianjiong Gao**, **Johanna Gardner**, **Levi A. Garraway**, **Julie M. Gastier-Foster**, **Carmelo Gaudioso**, **Nils Gehlenborg**, **Giannicola Genovese**, **Mark Gerken**, **Jeffrey E. Gershenwald**, **Gad Getz**, **Carmen Gomez-Fernandez**, **Thomas Gribbin**, **Jonna Grimsby**, **Benjamin Gross**, **Ranabir Guin**, **Tony Gutschner**, **Angela Hadjipanayis**, **Ruth Halaban**, **Benjamin Hanf**, **David Haussler**, **Lauren E. Haydu**, **D. Neil Hayes**, **Nicholas K. Hayward**, **David I. Heiman**, **Lynn Herbert**, **James G. Herman**, **Peter Hersey**, **Katherine A. Hoadley**, **Eran Hodis**, **Robert A. Holt**, **Dave SB. Hoon**, **Susan Hoppough**, **Alan P. Hoyle**, **Franklin W. Huang**, **Mei Huang**, **Sharon Huang**, **Carolyn M. Hutter**, **Matthew Ibbs**, **Lisa Iype**, **Anders Jacobsen**, **Valerie Jakrot**, **Alyssa Janning**, **William R. Jeck**, **Stuart R. Jefferys**, **Mark A. Jensen**, **Corbin D. Jones**, **Steven J.M. Jones**, **Zhenlin Ju**, **Hojabr Kakavand**, **Hyojin Kang**, **Richard F. Kefford**, **Fadlo R. Khuri**, **Jaegil Kim**, **John M. Kirkwood**, **Joachim Klode**, **Anil Korkut**, **Konstanty Korski**, **Michael Krauthammer**, **Raju Kucherlapati**, **Lawrence N. Kwong**, **Witold Kycler**, **Marc Ladanyi**, **Phillip H. Lai**, **Peter W. Laird**, **Eric Lander**, **Michael S. Lawrence**, **Alexander J. Lazar**, **Radoslaw Łaźniak**, **Darlene Lee**, **Jeffrey E. Lee**, **Junehawk Lee**, **Kenneth Lee**, **Semin Lee**, **William Lee**, **Ewa Leporowska**, **Kristen M. Leraas**, **Haiyan I. Li**, **Tara M. Lichtenberg**, **Lee Lichtenstein**, **Pei Lin**, **Shiyun Ling**, **Jia Liu**, **Ouida Liu**, **Wenbin Liu**, **Georgina V. Long**, **Yiling Lu**, **Singer Ma**, **Yussanne Ma**, **Andrzej Mackiewicz**, **Harshad S. Mahadeshwar**, **Jared Malke**, **David Mallery**, **Georgy M. Manikhas**, **Graham J. Mann**, **Marco A. Marra**, **Brenna Matejka**, **Michael Mayo**, **Sousan Mehrabi**, **Shaowu Meng**, **Matthew Meyerson**, **Piotr A. Mieczkowski**, **John P. Miller**, **Martin L. Miller**, **Gordon B. Mills**, **Fedor Moiseenko**, **Richard A. Moore**, **Scott Morris**, **Carl Morrison**, **Donald Morton**, **Stergios Moschos**, **Lisle E. Mose**, **Florian L. Muller**, **Andrew J. Mungall**, **Dawid Murawa**, **Pawel Murawa**, **Bradley A. Murray**, **Luigi Nezi**, **Sam Ng**, **Dana Nicholson**, **Michael S. Noble**, **Adeboye Osunkoya**, **Taofeek K. Owonikoko**, **Bradley A. Ozenberger**, **Elena Pagani**, **Oxana V. Paklina**, **Angeliki Pantazi**, **Michael Parfenov**, **Jeremy Parfitt**, **Peter J. Park**, **Woong-Yang Park**, **Joel S. Parker**, **Francesca Passarelli**, **Robert Penny**, **Charles M. Perou**, **Todd D. Pihl**, **Olga Potapova**, **Victor G. Prieto**, **Alexei Protopopov**, **Michael J. Quinn**, **Amie Radenbaugh**, **Kunal Rai**, **Suresh S. Ramalingam**, **Ayush T. Raman**, **Nilsa C. Ramirez**, **Ricardo Ramirez**, **Uma Rao**, **W. Kimryn Rathmell**, **Xiaojia Ren**, **Sheila M. Reynolds**, **Jeffrey Roach**, **A. Gordon Robertson**, **Merrick I. Ross**, **Jason Roszik**, **Giandomenico Russo**, **Gordon Saksena**, **Charles Saller**, **Yardena Samuels**, **Chris Sander**, **Cindy Sander**, **George Sandusky**,

Netty Santoso, **Melissa Saul**, **Robyn PM. Saw**, **Dirk Schadendorf**, **Jacqueline E. Schein**, **Nikolaus Schultz**, **Steven E. Schumacher**, **Charles Schwallier**, **Richard A. Scolyer**, **Jonathan Seidman**, **Pedamallu Chandra Sekhar**, **Harmanjatinder S. Sekhon**, **Yasin Senbabaoglu**, **Sahil Seth**, **Kerwin F. Shannon**, **Samantha Sharpe**, **Norman E. Sharpless**, **Kenna R. Mills Shaw**, **Candace Shelton**, **Troy Shelton**, **Ronglai Shen**, **Margi Sheth**, **Yan Shi**, **Carolyn J. Shiau**, **Ilya Shmulevich**, **Gabriel L. Sica**, **Janae V. Simons**, **Rileen Sinha**, **Payal Sipahimalani**, **Heidi J. Sofia**, **Matthew G. Soloway**, **Xingzhi Song**, **Carrie Sougnez**, **Andrew J. Spillane**, **Arkadiusz Spychała**, **Jonathan R. Stretch**, **Joshua Stuart**, **Wiktoria M. Suchorska**, **Antje Sucker**, **S. Onur Sumer**, **Yichao Sun**, **Maria Synott**, **Barbara Tabak**, **Teresa R. Tabler**, **Angela Tam**, **Donghui Tan**, **Jiabin Tang**, **Roy Tarnuzzer**, **Katherine Tarvin**, **Honorata Tatka**, **Barry S. Taylor**, **Marek Teresiak**, **Nina Thiessen**, **John F. Thompson**, **Leigh Thorne**, **Vesteinn Thorsson**, **Jeffrey M. Trent**, **Timothy J. Triche**, **Kenneth Y. Tsai**, **Peiling Tsou**, **David J. Van Den Berg**, **Eliezer M. Van Allen**, **Umadevi Veluvolu**, **Roeland G. Verhaak**, **Douglas Voet**, **Olga Voronina**, **Vonn Walter**, **Jessica S. Walton**, **Yunhu Wan**, **Yuling Wang**, **Zhining Wang**, **Scot Waring**, **Ian R. Watson**, **Nils Weinhold**, **John N. Weinstein**, **Daniel J. Weisenberger**, **Peter White**, **Matthew D. Wilkerson**, **James S. Wilmott**, **Lisa Wise**, **Maciej Wiznerowicz**, **Scott E. Woodman**, **Chang-Jiun Wu**, **Chia-Chin Wu**, **Junyuan Wu**, **Ye Wu**, **Ruibin Xi**, **Andrew Wei Xu**, **Da Yang**, **Liming Yang**, **Lixing Yang**, **Travis I. Zack**, **Jean C. Zenklusen**, **Hailei Zhang**, **Jianhua Zhang**, **Wei Zhang**, **Xiaobei Zhao**, **Jingchun Zhu**, **Kelsey Zhu**, **Lisa Zimmer**, **Erik Zmuda**, **Lihua Zou**. 2015. Genomic Classification of Cutaneous Melanoma. *Cell* **161**:1681–1696. doi: 10.1016/j.cell.2015.05.044.

Russano M, Napolitano A, Ribelli G, Iuliani M, Simonetti S, Citarella F, Pantano F, Dell'Aquila E, Anesi C, Silvestris N, Argentiero A, Solimando AG, Vincenzi B, Tonini G, Santini D. 2020. Liquid biopsy and tumor heterogeneity in metastatic solid tumors: the potentiality of blood samples. *Journal of experimental & clinical cancer research : CR* **39**:95. doi: 10.1186/s13046-020-01601-2.

Siraj S, Masoodi T, Siraj AK, Azam S, Qadri Z, Ahmed SO, AlBalawy WN, Al-Obaisi KA, Parvathareddy SK, AlManea HM, AlHussaini HF, Abduljabbar A, Alhomoud S, Al-Dayel FH, Alkuraya FS, Al-Kuraya KS. 2020. Clonal Evolution and Timing of Metastatic Colorectal Cancer. *Cancers* **12**:2938. doi: 10.3390/cancers12102938.

Thomas RM, Truty MJ, Nogueras-Gonzalez GM, Fleming JB, Vauthey J-N, Pisters PWT, Lee JE, Rice DC, Hofstetter WL, Wolff RA, Varadhachary GR, Wang H, Katz MHG. 2012. Selective reoperation for locally recurrent or metastatic pancreatic ductal adenocarcinoma following primary pancreatic resection. *Journal of Gastrointestinal Surgery* **16**:1696–1704. doi: 10.1007/s11605-012-1912-8.

Vignot S, Lefebvre C, Frampton GM, Meurice G, Yelensky R, Palmer G, Capron F, Lazar V, Hannoun L, Miller VA, André F, Stephens PJ, Soria J-C, Spano J-P. 2015. Comparative analysis of primary tumour and matched metastases in colorectal cancer patients: evaluation of concordance between genomic and transcriptional profiles. *European journal of cancer (Oxford, England : 1990)* **51**:791–799. doi: 10.1016/j.ejca.2015.02.012.

Wagle N, Painter C, Krevalin M, Oh C, Anderka K, Larkin K, Lennon N, Dillon D, Frank E, Winer EP, Lander E, Golub T. 2016. The Metastatic Breast Cancer Project: A national direct-to-patient initiative to accelerate genomics research. *Journal of Clinical Oncology* **34**:LBA1519-LBA1519. doi: 10.1200/JCO.2016.34.18_suppl.LBA1519.

Wu Q, Li J, Zhu S, Wu J, Chen C, Liu Q, Wei W, Zhang Y, Sun S. 2017. Breast cancer subtypes predict the preferential site of distant metastases: a SEER based study. *Oncotarget* **8**:27990–27996. doi: 10.18632/oncotarget.15856.

Xu K, Wang R, Xie H, Hu L, Wang C, Xu J, Zhu C, Liu Y, Gao F, Li X, Wang C, Huang J, Zhou W, Zhou G, Shu Y, Guan X. 2021. Single-cell RNA sequencing reveals cell

heterogeneity and transcriptome profile of breast cancer lymph node metastasis. *Oncogenesis* **10**:66. doi: 10.1038/s41389-021-00355-6.

Yachida S, Jones S, Bozic I, Antal T, Leary R, Fu B, Kamiyama M, Hruban RH, Eshleman JR, Nowak MA, Velculescu VE, Kinzler KW, Vogelstein B, Iacobuzio-Donahue CA. 2010. Distant metastasis occurs late during the genetic evolution of pancreatic cancer. *Nature* **467**:1114–1117. doi: 10.1038/nature09515.

Zhang Y, Narayanan SP, Mannan R, Raskind G, Wang X, Vats P, Su F, Hosseini N, Cao X, Kumar-Sinha C, Ellison SJ, Giordano TJ, Morgan TM, Pitchiaya S, Alva A, Mehra R, Cieslik M, Dhanasekaran SM, Chinnaiyan AM. 2021. Single-cell analyses of renal cell cancers reveal insights into tumor microenvironment, cell of origin, and therapy response. *Proceedings of the National Academy of Sciences* **118**:e2103240118. doi: 10.1073/pnas.2103240118.

Zhao Y, Pan Z, Namburi S, Pattison A, Posner A, Balachander S, Paisie CA, Reddi HV, Rueter J, Gill AJ, Fox S, Raghav KPS, Flynn WF, Tothill RW, Li S, Karuturi RKM, George J. 2020. CUP-AI-Dx: A tool for inferring cancer tissue of origin and molecular subtype using RNA gene-expression data and artificial intelligence. *EBioMedicine* **61**:103030. doi: 10.1016/j.ebiom.2020.103030.

Zheng H, Liu H, Ge Y, Wang X. 2021. Integrated single-cell and bulk RNA sequencing analysis identifies a cancer associated fibroblast-related signature for predicting prognosis and therapeutic responses in colorectal cancer. *Cancer Cell International* **21**:552. doi: 10.1186/s12935-021-02252-9.

Zhou H, Zheng S, Truong LD, Ro JY, Ayala AG, Shen SS. 2014. Clear cell papillary renal cell carcinoma is the fourth most common histologic type of renal cell carcinoma in 290 consecutive nephrectomies for renal cell carcinoma. *Human Pathology* **45**:59–64. doi: 10.1016/j.humpath.2013.08.004.