PieParty: Visualizing cells from scRNA-seq data as pie charts

Stefan Kurtenbach*, James J. Dollar, Michael A. Durante, and J. William Harbour

Bascom Palmer Eye Institute, Sylvester Comprehensive Cancer Center, Interdisciplinary Stem Cell Institute, University of Miami Miller School of Medicine, Miami, FL, USA.

*Corresponding author

Abstract

Single cell RNA sequencing (scRNA-seq) has been a transformative technology in many research fields. Dimensional reduction techniques such as UMAP and tSNE are used to visualize scRNA-seq data in two or three dimensions in order for cells to be clustered in biologically meaningful ways. Subsequently, gene expression is frequently mapped onto these plots to show the distribution of expression across cell types. However, plotting each cell with only one color leads to repetitive and unintuitive representations. Here, we present Pie Party, which allows scRNA-seq data to be plotted such that every cell is represented as a pie chart, where every slice in the pie charts correspond to the gene expression of individual genes, which allows for the simultaneous visualization of the expression of multiple genes and gene networks. The resulting figures are information dense, space efficient and highly intuitive. PieParty is publicly available on GitHub at https://github.com/harbourlab/PieParty.

Introduction

Gene expression data at single-cell resolution has brought a lot of opportunity to gain detailed understanding of heterogenous cell populations, but also many challenges regarding its processing and visualization. Principle component analysis (PCA) was initially used to reduce dimensions on single-cell RNA sequencing (scRNA-seq) datasets and to plot them in two or three dimensions. t-SNE and UMAP were developed subsequently, offering superior and global resolution, and are currently the most commonly used dimensional reduction techniques for scRNA-seq [1, 2]. For most applications, gene expression is mapped onto t-SNE or UMAP plots, for instance to highlight genes that differentiate cell types. The common and only approach nowadays is to color the respective cell dots in UMAP or tSNE plots, where the intensity of the color correlates with how high the gene is expressed. However, this only allows to plot one gene per cell, which is inefficient, and results in many repetitive illustrations when expression of multiple different genes needs to be shown. This not only consumes valuable figure space, but is also not intuitive in many cases. Here we present PieParty, a python script that allows users to display

each cell in t-SNE, UMAP, or any other single-cell plots with coordinates, as a pie charts. The pie chart can be used to visualize expression of multiple genes at once, and can be customized using different colors or color palettes. PieParty also offers additional settings for normalization and customized plotting.

Results

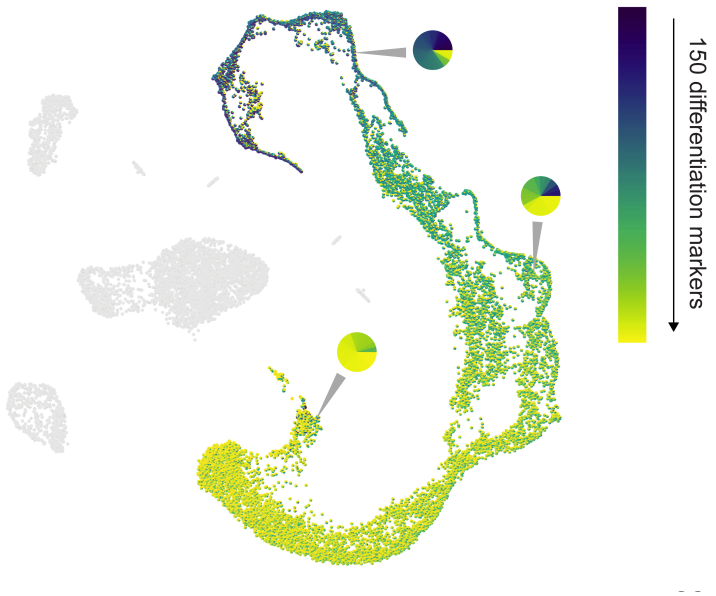

Figure 1 PieParty plot depicting 15,479 cells. Around 150 differentiation markers were sorted from early (blue) to late (yellow), and auto-assigned a color.

The basic principle of the PieParty visualization is to generate pie charts for every cell in a single cell sequencing plot, like t-SNE and UMAP plots. The user provides a gene list of all genes that he wants to visualize, and PieParty will generate pie charts where each slice in a pie chart represents the normalized gene expression of one individual gene in the list. Each gene (slice) can be assigned a unique color, or a color palette can be chosen to auto-assign unique colors for every gene. Choosing a color palette is useful if larger gene lists need to be plotted. As an example, we analyzed scRNA-seq data from human testis, and clustered the cells with UMAP [3]. The analysis reveals a developmental trajectory, ranging from stem cells to differentiated sperm, in very defined cell stages. Figure 1 and Supplementary Figure 1 shows PieParty plots, where every cell (n = 15,479) is represented by a pie chart, with 145 different differentiation markers plotted per pie chart. Each differentiation marker is automatically assigned a unique color on a color map, sorted from early markers in dark violet to markers of differentiation in yellow. The resulting plot shows that even plotting a large number of genes as pie charts can still yield a very informative and intuitive UMAP plot, showing that the expression of these differentiation markers forms a continuum along the differentiation axis.

Besides merely plotting a single list of genes, PieParty also allows for more complex visualizations. As an example, we used scRNA-seq data from Uveal Melanoma (UM) tumors to demonstrate a more complex application [4]. There are two main classes of UM tumors, class 1,

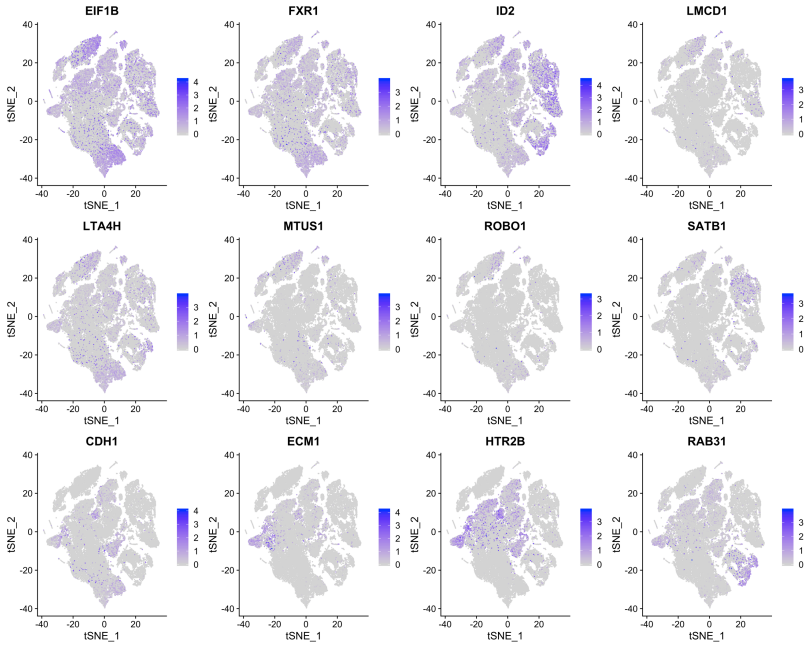

Figure 2: 12 tSNE plots showing the classical way of depicting gene expression of 12 different genes, which is not very intuitive.

which rarely metastasize, and class 2, which frequently metastasize and cause a high mortality rate in patients. There is a PCR-based clinical test available, using gene expression of 12 genes to discriminate both tumor classes [5]. 4 genes are generally up-regulated in class 2 tumors, whereas expression of 8 marker genes is used to classify class 1 tumors. The standard way of visualizing this is by showing 12 UMAP plots, where expression of these 12 genes are visualized separately. Supplementary Figure 2 shows all 12 genes plotted separately, which is very space consuming, and while still possible for 12 genes, not immediately comprehensible. Not every class 1 tumor cell expresses all 8 markers of class 1 tumors, same is true for class 2 tumors. Another problem with this visualization is that the class 1 and class 2 tumor cells are intermixed in tSNE plots to some extent. This is even more the case when the UMAP technique is applied that better preserves the global structure of the data (data not shown). This makes it hard to asses gene expression relationships, even when an overview is provided of where class 1 and class 2 tumor cells are located. For these and other cases, it is useful to plot class 1 and class 2 tumor cells separately, which is depicted in Supplemental Figures 3 and 4. However, with 24 total plots this consumes precious space within a figure. PieParty offers multiple options to visualize this complex relationship of class 1 and class 2 genes by applying different colors and plotting cells as pie charts. One consideration to make is that plotting gene lists with different number of genes, in this case four genes for class 2 versus eight genes for class 1, comes with a

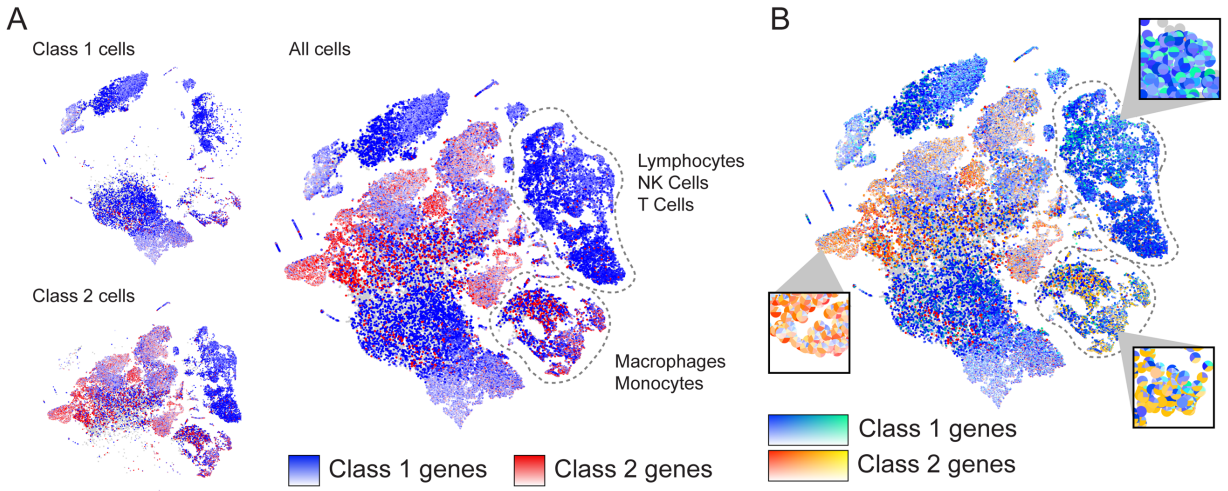

Figure 3: (A) left -Class 1 and Class 2 tumor cells plotted separately, with class 1 associated genes in blue, and class 2 associated genes in red. In contrast to Figure 1, color intensity of each slice in all pies was set to reflect gene expression, in addition to slice size. Right – merged image. (B) Same plot as in Figure 3A right, but auto assigned color palettes, giving each gene a unique color.

bias. Hence, PieParty allows to normalize for that fact, and weighs gene sets equally if desired, by applying a normalization factor to account for the difference in number of genes. This normalization technique was applied for all following plots, and the color intensity of the individual pie slices was set to correlate with gene expression, in addition to the size of the pies itself.

Figure 3A depicts the tSNE plots of 11 UM tumor samples, (3 class 1, 8 class 2, n = 59,915 cells) plotted with PieParty and a simple blue/red color scheme. Class 1 and Class 2 tumor cells were plotted individually, as well as combined in one big plot. The eight class 1 genes are plotted in blue in these examples, and the four class 2 genes are plotted in red. It is immediately evident, that the tumor cell markers are not only found in tumor cells, but also significantly in immune cells. Class 2 tumor markers are predominantly enriched in macrophages and monocytes, whereas class 1 tumor markers are present in other immune cells including Lymphocytes, NK cells, T-Cells. Especially interesting is the fact that in this scRNA-seq dataset one class 1 tumor marker is expressed in some class 2 immune cells. While all this information is technically visible when all genes are plotted individually (see Supplemental Figures 3 and 4), the PieParty plots are intuitively readable and visualize these relationships in a space efficient way. Besides assigning single colors to multiple gene lists and shading colors by gene expression, the information density can be further increased by assigning individual colors to each gene as in Figure 1. Figure 3B shows the same dataset, with color palettes applied to the two gene lists. This visualization shows

that for class 2 tumors the genes expressed by immune cells are different from the tumor cells, but for class 1 tumors there seems to be overlap.

Discussion

Here we present PieParty, which allows single cells from scRNA-seq data to be represented as pie charts instead of single-colored dots. PieParty plots are intended to complement, rather than replace, the conventional method of depicting the expression of one gene per plot, which is the clearest way to show which cells a gene is expressed in. By contrast, PieParty plots are far more information rich, allowing multi-dimensional single cell sequencing data to be represented in an intuitive and space-efficient way, allowing for the identification of previously unrecognized transcriptional heterogeneity. In our example, one PieParty plot allowed for visualization of the GEP class for each tumor cell based on the expression of 12 genes (Figure 3). This approach can be further expanded to include larger gene lists to determine the transcriptional profiles defining each tumor cell or subset of tumor cells [6]. Together, the PieParty plot tool can provide deeper insights into biology derived from single cell sequencing by allowing simultaneous multidimensional visualization of high-density datasets.

References

1. van der Maaten, L. and G. Hinton, *Viualizing data using t-SNE.* Journal of Machine Learning Research, 2008. **9**: p. 2579-2605.
2. McInnes, L., J. Healy, and J. Melville, *Umap: Uniform manifold approximation and projection for dimension reduction.* arXiv preprint arXiv:1802.03426, 2018.
3. Guo, J., et al., *The adult human testis transcriptional cell atlas.* Cell Res, 2018. **28**(12): p. 1141-1157.
4. Durante, M.A., et al., *Single-cell analysis reveals new evolutionary complexity in uveal melanoma.* Nat Commun, 2020. **11**(1): p. 496.
5. Onken, M.D., et al., *An accurate, clinically feasible multi-gene expression assay for predicting metastasis in uveal melanoma.* J Mol Diagn, 2010. **12**(4): p. 461-8.
6. Subramanian, A., et al., *Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.* Proc Natl Acad Sci U S A, 2005. **102**(43): p. 15545-50.