Droplets: A Marker Design for visually enhancing Local Cluster Association

Stefan Lengauer*, Peter Waldert, and Tobias Schreck
Institute of Computer Graphics and Knowledge Visualization, Graz University of Technology

ABSTRACT

The objective of the Redesign Challenge of the Bio+MedVis Challenge @ IEEE VIS 2024 is to redesign an existing visualization of multi-cell gene expressions of tissue samples. In this, multiple cells are accumulated into pixels. For each pixel the visualization should convey the prevalence and extent of cell types it is composed of, i.e., a proportional relation. The provided baseline technique of superimposed Pie charts – a common technique for this kind of relation – is not an ideal choice as the cell-type quantities of neighboring pixels are hard to compare due to a spatial disarray inherent to pie charts. This limits the perception of regions with coherent cell-type compositions, which constitutes one of the essential visual analytics tasks. We propose a novel marker design: *Droplets* – a space-saving design for visually enhancing the presence of clusters and regional borders. We evaluate this concept for the given tissue sample and compare it to the given baseline and other alternatives.

Index Terms: Association, Gestalt laws, Small multiples, Glyphs.

1 INTRODUCTION

The *STdeconvolve* [4] accurately recovers cell-type proportions in multi-cellular spatial transcriptomic (ST) tissue profiling. Miller et al. [4] present a visualization where the components of the deconvolved cell-types pixels are illustrated as pie charts (Fig. 2a[1]) – a standard visualization for proportional relations. They can be superposed to the respective hematoxylin and eosin (H&E) stain, as the locations of pixels is given by the center of each circle. I.e., the visualization has to convey two vital types of information: locations and proportions. One of the fundamental tasks of visual analytics is to recognize coherent areas of homogeneous cell-type compositions, a quest that is inherently difficult by comparing pie charts, as quantities are not aligned across pie charts. With the words of Tufte [9]: *"[...] the only worse design than a pie chart is several of them, for then the viewer is asked to compare quantities located in spatial disarray both within and between pies [...]."*

Other examples for visualizing part-to-whole relationships are bar charts, stacked bar/column charts or treemaps. While suitable for the comparison of quantities between locations, bar charts come with the obvious downside that they exhibit a suboptimal shape for conveying locality, since it is difficult for humans to perceive a chart's center (Fig. 2b). Also their layout poorly uses the available space and does not scale well for a larger number of components. Stacked bar/column charts can be visualized with a quadratic design (Fig. 2c), which fixes this issue. Yet, also with this design we observe a level of spatial disarray of quantities, making it hard for viewers to spot homogeneous regions. Also, in regions with multiple equally strong quantities it is comparably difficult to decipher their respective contributions. Yet another common visualization for displaying proportional relations are *treemaps* (Fig. 2d). This layout makes optimal use of the available space, which is iteratively split into regions reflecting the quantities' respective contributions. While it is a very helpful design for a single ST pixel, using treemaps in a small multiples visualization leads to a very turbulent looking result. The spatial disarray is even aggravated through the sorting of quantities. In summary, none of the baseline visualizations for proportional relations leads to a substantial improvement for the problem at hand. More advanced approaches for displaying small multiples in their spatial context exist [7, 5], but they are not applicable for showing proportional relations.

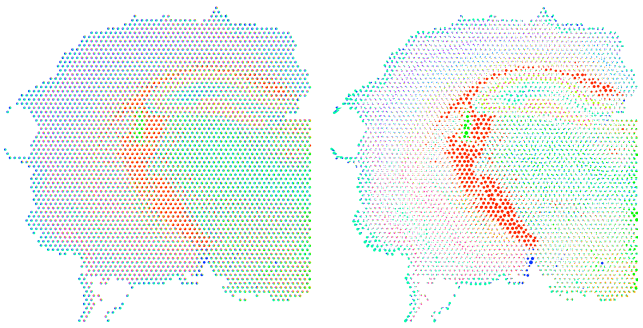


Figure 1: Original small multiples cell type visualization with pie charts (left) next to our proposed *Droplets* design (right).

We propose a design (Fig. 1) which optimally uses the available space, being as translucent as possible (so that the underlying H&E image can be perceived), while still conveying the proportional information. Additionally the component markers should visually reflect their association to local clusters.

2 PROPOSED DESIGN

The essential idea of our proposed design is to visualize the different cell-type contributions as separated, yet tightly-packed, smooth shapes which reflect both a component's contribution and its belonging to local accumulations. To this end, a circular base shape is equipped with a protruding spike (tail) pointing towards the center of such a local cluster (Fig. 3, right), with the length of this spike reflecting the size of the respective cluster. This is in accordance with the Gestalt principles of *Continuity* and *Common Fate*. Continuity states that viewers group elements that seem to follow a continuous path – in our case towards a local accumulation. Common Fate refers to viewers' tendency to group elements which seem to move in a common direction. This marker design emphasizes latent borders of regions with similar cell-type contributions (Fig. 3, left) and facilitates the recognition of – otherwise invisible – coherent regions. The idea is inspired by the winglets analogy by Lu et al. [3] for scatter plots, where each point is fitted with dual-sided trails, reflecting the affiliation confidence to a cluster and protruding perpendicular to the direction of the respective cluster. However, to allow for a tight packing we opt for said protruding spike as opposed to such dual-sided trails.

Regarding the positioning of the individual components within

*e-mail: s.lengauer@cgv.tugraz.at

[1] Fig. 2–5 given in the supplemental material.

a ST pixel, we intend to arrange them such that they appear to be "drawn towards" their belonging local cluster centers (if any are present). This ensures both that components' markers are closer to their respective clusters and that their tails are unlikely to obfuscate other markers, while still maintaining a tight packing. This requires precise knowledge of the prevalence and the direction of the nearest local cluster. Let $\{p_d\} \subset \mathbb{R}^2$ be the set of ST pixels, with the index set $d \in D$, and $k \in K$ be the index set of cell-types. The center (and thus the direction to) a cluster for a component k can be obtained with the multivariate L_1 median, the point $q = \arg\min_x\{\sum_{d\in D}\|p_d - x\|\}$, which minimizes the Euclidean distance to all points $\{p_d\}$. As we do not want to compute the global L_1 but a localized version, we employ a variant of the Locally Optimal Projection (LOP) by Lipmann et al. [1], giving $\nabla_d^k = \frac{\sum_{d'\in D} p_{d'}\omega_{d'}^k \theta(\|p_{d'}-p_d\|)}{\sum_{d'\in D}\omega_{d'}^k\theta(\|p_{d'}-p_d\|)} - p_d$, for the directivity of component k in ST pixel d, with ω_d^k as component k's contribution to pixel d. Locality is introduced through the fast decaying weight function θ. ∇_d^k also contains the information how pronounced the directivity to a cluster is, which is used for the eccentricity $a \propto |\nabla_d^k|$ of a marker's tail. The circular base shape with radius r_d^k is extended with two circle arcs meeting at the point with distance a in direction of ∇_d^k (Fig. 3, right). There are different options for defining a marker's radius, as there are three distinct retinal variables that encode a component in a pie chart: angle, area of the circle wedge, and arc length. While it has been shown that the angle is the least important visual cue, the arc length and area exhibit similar accuracy in terms of understanding [6]. As a consequence, the latter two have been evaluated ($r_d^k \propto \omega_d^k$ and $r_d^k \propto (\omega_d^k)^{1/2}$, respectively), with the area preserving option (used in all subsequent experiments) leading to substantially more expressive layouts.

The placement of the component markers $\{m_d^k\}$ within a pixel d is achieved through a physical simulation of multiple simultaneously acting forces ($\boldsymbol{f}_{col} \gg \boldsymbol{f}_g > \boldsymbol{f}_\nabla$): (i) a gravitational force $\boldsymbol{f}_g \sim \boldsymbol{g}$ pulling them towards the pixel's center, with $\boldsymbol{g}$ as the gravitational acceleration; (ii) a fast decaying collision force $\boldsymbol{f}_{col} \sim \sum_{k'\in K\setminus\{k\}}(m_d^k - m_d^{k'})\psi_{d_k,d_{k'}}$, with $\psi_{d_k,d_{k'}} \propto (r_d^k + r_d^{k'})^{-1}$, and (iii) a clustering force $\boldsymbol{f}_\nabla \sim \left|\nabla_d^k\right| \nabla_d^k m_d^k$, pulling them towards the local cluster center (Fig. 3, middle). Initially the circular markers are placed at an offset from the pixel's center relative to their ∇_d^k, before their positions are iteratively refined by applying said forces. In most cases this leads to a blossom-like arrangement with differently large and elongated petals, facing away from the pixel center.

3 First Results and Discussion

We evaluate our proposed design both with circular- and droplet-shaped markers. Fig. 4 shows the respective results in comparison with the original pie chart design. Column (a) features the spot cell-types and (b) the spot cell-types superimposed onto the H&E image. Already from the macro perspective in column (a) and (b), it is apparent that our proposed design reveals several fine-grained intensity structures which are latent in the pie charts layout, such as the turquoise X5 cell-types, present all around the border, or the violet and pink marblings of the X8/X9 cell-types to the top and left borders of tissue sample. A closer inspection of the regions where homogeneous cell-type compositions border each other (column (c)), clearly demonstrates the advantage of our design. This part of the tissue exhibits a coherent region dominated by the red X1 cell type, which encases a cluster of the green X4 cell type. To the top and right we observe an area dominated by the X5 cell type which is intersected with a streak, composed in equal parts of X3 and X9 cell type. This also correlates with darker marbling in the H&E image. Even though this information can also be derived from the original pie chart visualization, our proposed design conveys this information more efficiently with less visual disturbance. In cases, where a dominant cluster has offshoots into neighboring regions – such as the X1 cell type cluster in (d) – the droplets shape lead to a more pronounced cluster outline, in accordance with the *Common Faith* Gestalt principle.

4 Next Steps and Further Improvements

For a proof-of-concept, visualizations were generated with a Python script[2] without optimization in mind. The implementation runs fast on commodity hardware but will need further improvements regarding efficiency if the concept is to be applied in real-time on tissue data. To this end, many approaches such as parallelization or approximation are possible. The iteratively generated force layout can be optimized with *early stopping* or similar schemes, be replaced by an approximate algorithmic approach altogether.

Another target for improvement is the employed color scheme. The provided color scheme contains easy-to-confuse color pairs, such as for components X4/X5 or X8/X9. The latter are also hard to discern from the underlying H&E diagram, which features a similar colorization. Hence, future tasks will include the improvement of the color scheme with strategies for enhancing the distinguishability of adjacent color complexes [2] (Fig. 5).

Besides improvements of the marker design, we have yet to explore dynamic/interactive concepts such as filters, which limit the visibility of certain components or balance the opacity of both droplets and H&E image, or interactive lenses [8]. Combined with a perception-centered usability testing, investigating such concepts constitute our next research efforts.

Acknowledgment

This work has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement no 101137074 - HEREDITARY.

References

[1] Y. Lipman, D. Cohen-Or, D. Levin, and H. Tal-Ezer. Parameterization-free projection for geometry reconstruction. *ACM Trans. Graph.*, 26(3):22–es, July 2007. doi: 10.1145/1276377.1276405 2

[2] K. Lu, M. Feng, X. Chen, M. Sedlmair, O. Deussen, D. Lischinski, Z. Cheng, and Y. Wang. Palettailor: Discriminable colorization for categorical data. *IEEE TVCG*, 27(2):475–484, 2020. doi: 10.1109/TVCG.2020.3030406 2

[3] M. Lu, S. Wang, J. Lanir, N. Fish, Y. Yue, D. Cohen-Or, and H. Huang. Winglets: Visualizing association with uncertainty in multi-class scatterplots. *IEEE TVCG*, 26(1):770–779, 2020. doi: 10.1109/TVCG.2019.2934811 1

[4] B. F. Miller, F. Huang, L. Atta, A. Sahoo, and J. Fan. Reference-free cell type deconvolution of multi-cellular pixel-resolution spatially resolved transcriptomics data. *Nature communications*, 13(1):2339, 2022. doi: 10.1038/s41467-022-30033-z 1

[5] E. Mörth, I. S. Haldorsen, S. Bruckner, and N. N. Smit. Paraglyder: Probe-driven interactive visual analysis for multiparametric medical imaging data. In *Advances in Computer Graphics*, pp. 351–363. Springer, 2020. doi: 10.1007/978-3-030-61864-3_29 1

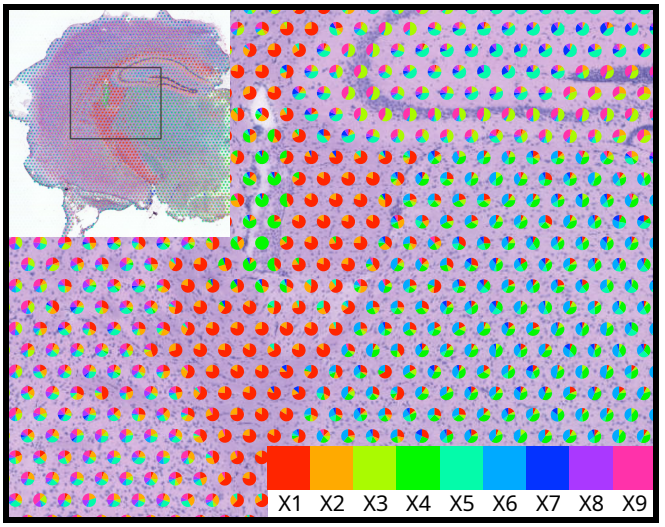
[6] D. Skau and R. Kosara. Arcs, angles, or areas: Individual data encodings in pie and donut charts. *Computer Graphics Forum*, 35(3):121–130, 2016. doi: 10.1111/cgf.12888 2

[7] S. Stoppel, E. Hodneland, H. Hauser, and S. Bruckner. Graxels: information rich primitives for the visualization of time-dependent spatial data. In *VCBM/MedViz*, pp. 183—192. Eurographics Association, 2016. doi: 10.2312/vcbm.20161286 1

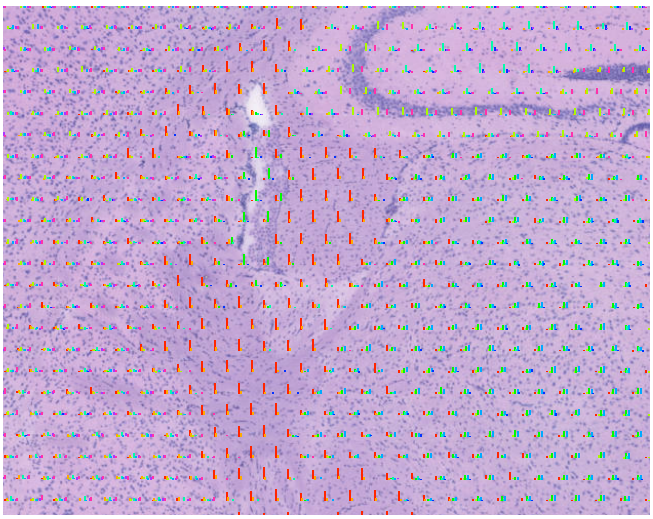
[8] C. Tominski, S. Gladisch, U. Kister, R. Dachselt, and H. Schumann. A survey on interactive lenses in visualization. In *EuroVis (STARs)*, pp. 1–20, 2014. doi: 10.2312/eurovisstar.20141172 2

[9] E. R. Tufte and P. R. Graves-Morris. *The visual display of quantitative information*, vol. 2. Graphics press Cheshire, CT, 1983. 1

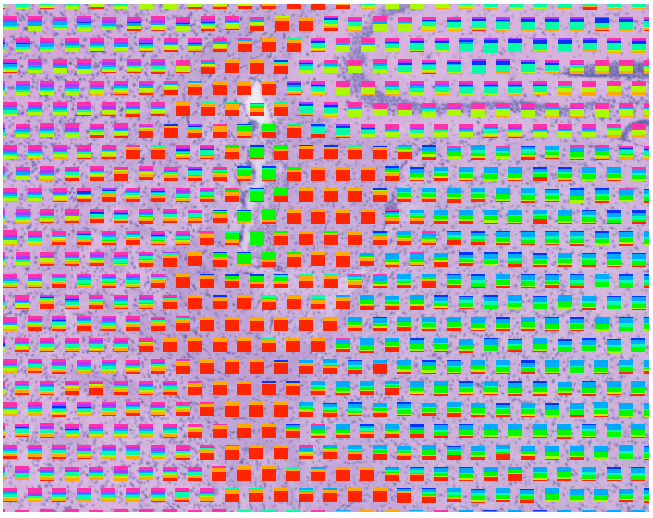
[2]Code @ https://pluto.cgv.tugraz.at/slengaue/droplets.



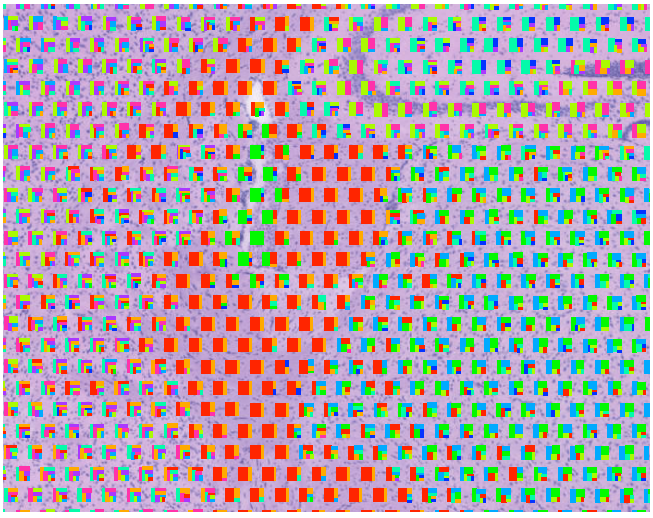
(a) Reference: Pie Charts



(b) Bar Charts



(c) Stacked Bar Charts



(d) Treemaps

Figure 2: Common alternatives for visualizing proportional relationships – besides (a) Pie Charts – include (b) Bar Charts, (c) Stacked Bar Charts or (d) Treemaps. All of them are evaluated on the given tissue sample with cell-type components X1 to X9.

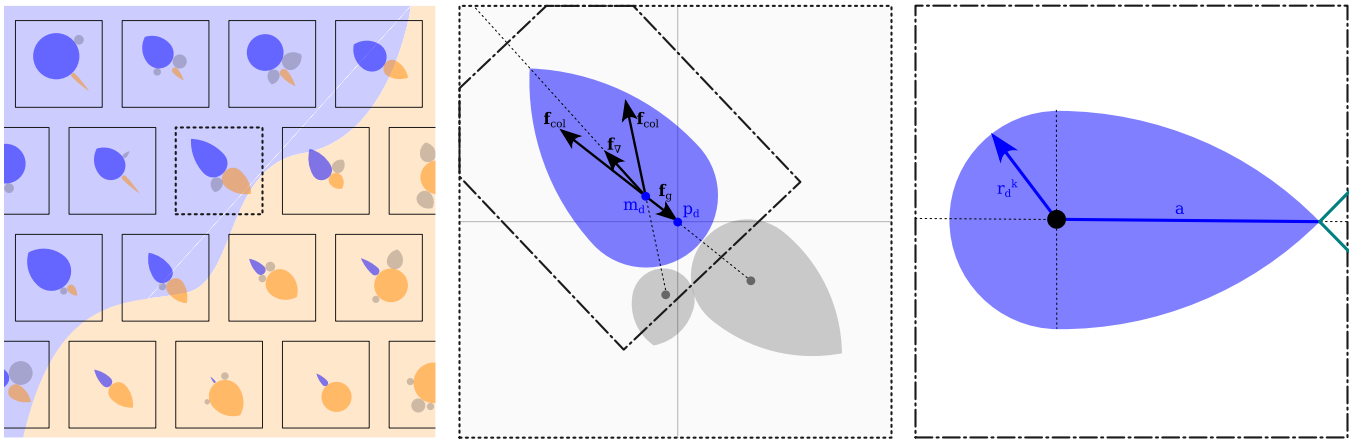


Figure 3: Left: pixel components of two symbolic clusters (blue and orange) strive towards their respective cluster centers as, indicated by their respective tails. Middle: different forces simultaneously act upon each droplets marker, typically resulting in a blossom pattern for a ST pixel. Right: the eccentricity a, responsible for the droplets shape is governed by the prevalence and manifestation of component clusters.

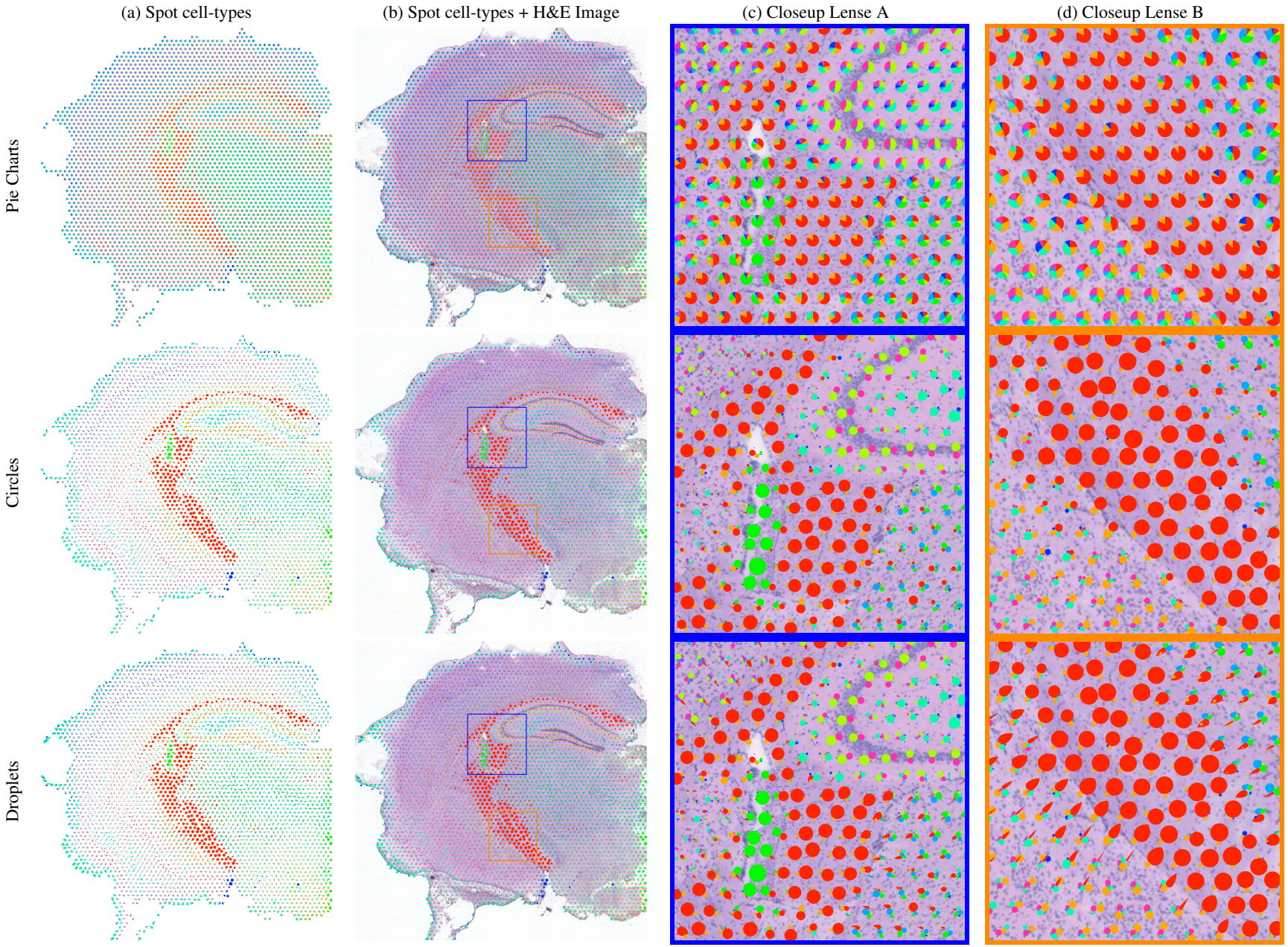


Figure 4: Original cell-type visualization with pie charts (top row) and our proposed redesign with circular markers (2nd row) and the droplets shape (3rd row). The columns show the (a) spot cell-types of the tissue as a whole, the (b) spot cell types of the tissue as a whole superimposed to the H&E image and two close-ups on significant regions (columns (c) and (d)). The same colormap as in Fig. 2a is used.

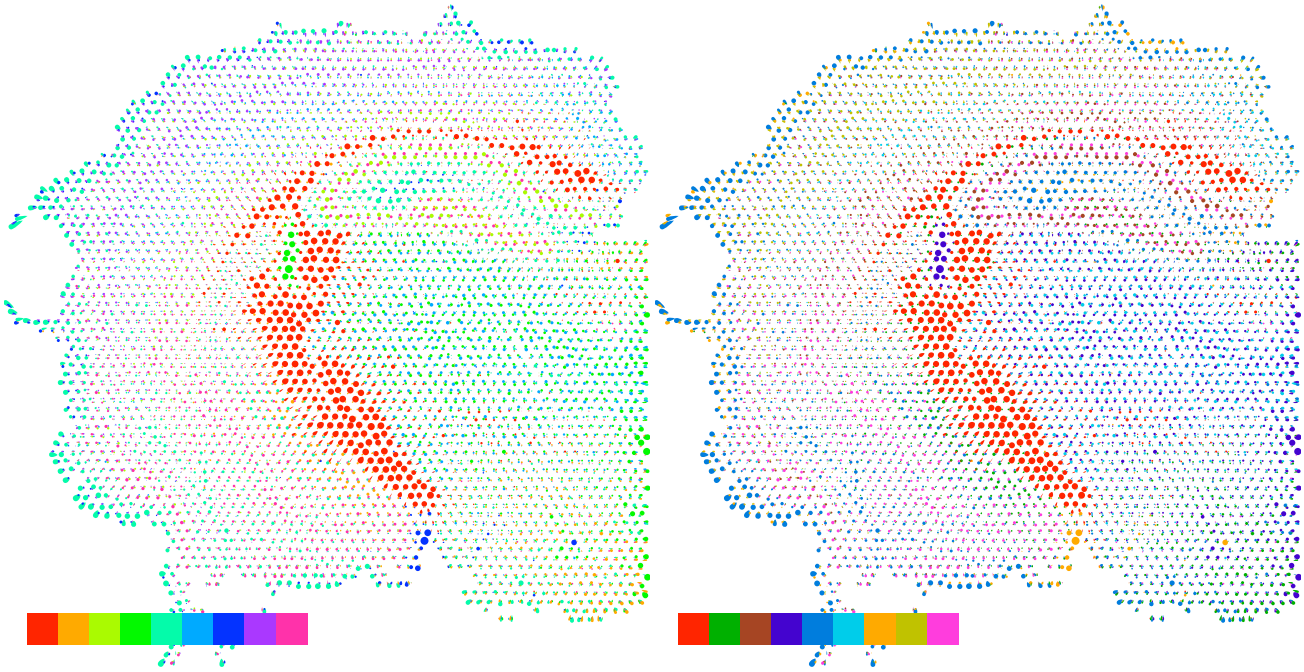


Figure 5: Left: pairs of easy-to-confuse colors (X4/X5 or X8/X9) are challenging to discern type contributions if they appear together (see, e.g., closeup). Right: using a alternate color scheme can prohibit similar colors bordering each others and reveal otherwise-hard-to-see structures, such as with the green and turquoise cell-type components in the lower right part of the tissue.