MerQuaCo: a computational tool for quality control in image-based spatial transcriptomics

Naomi Martin[1], Paul Olsen[1], Jacob Quon, Jazmin Campos, Nasmil Valera Cuevas, Josh Nagra, Marshall VanNess, Zoe Maltzer, Emily C Gelfand, Alana Oyama, Amanda Gary, Yimin Wang, Angela Alaya, Augustin Ruiz, Cade Reynoldson, Cameron Bielstein, Christina Alice Pom, Cindy Huang, Cliff Slaughterbeck, Elizabeth Liang, Jason Alexander, Jeanelle Ariza, Jocelin Malone, Jose Melchor, Kaity Colbert, Krissy Brouner, Lyudmila Shulga, Melissa Reding, Patrick Latimer, Raymond Sanchez, Stuard Barta, Tom Egdorf, Zachary Madigan, Chelsea M Pagan, Jennie L Close, Brian Long, Michael Kunst, Ed S Lein, Hongkui Zeng, Delissa McMillen, Jack Waters[2].

Allen Institute for Brain Science, 615 Westlake Ave N, Seattle WA.

[1]Equal contributions.

[2]Author for correspondence. jackw@alleninstitute.org.

Acknowledgements

This research was funded by grants from National Institute of Mental Health, U19MH114830 to H.Z. and U19AG060909 to E.L. This work was also supported by the Allen Institute for Brain Science. The authors thank the Allen Institute founder, Paul G. Allen, for his vision, encouragement, and support. We thank Vizgen for their assistance maintaining MERSCOPE hardware throughout the study and for their feedback on this manuscript.

ABSTRACT

Image-based spatial transcriptomics platforms are powerful tools often used to identify cell populations and describe gene expression in intact tissue. Spatial experiments return large, high-dimensional datasets and several open-source software packages are available to facilitate analysis and visualization. The outputs of spatial transcriptomics platforms are typically imperfect. For example, local variations in transcript detection probability are common. Software tools to characterize imperfections and their impact on downstream analyses are lacking so the data quality is assessed manually, a laborious and often a subjective process. Here we describe imperfections in a dataset of 641 fresh-frozen adult mouse brain sections collected using the Vizgen MERSCOPE. Common imperfections included the local loss of tissue from the section, tissue outside the imaging volume due to detachment from the coverslip, transcripts missing due to dropped images, varying detection probability through space, and differences in transcript detection probability between experiments. We describe the incidence of each imperfection and the likely impact on the accuracy of cell type labels. We develop MerQuaCo, open-source code that detects and quantifies imperfections without user input, facilitating the selection of sections for further analysis with existing packages. Together, our results and MerQuaCo facilitate rigorous, objective assessment of the quality of spatial transcriptomics results.

INTRODUCTION

The recent advent of spatially resolved molecular imaging methods has enabled the investigation of gene expression patterns within cells in their native tissue context, revealing the organization of transcriptomically-defined cell types (Close *et al.*, 2021). Researchers have leveraged emerging spatial technologies to create comprehensive cell-type atlases of a variety of tissue types, including human heart (Asp *et al.,* 2019), breast cancer (Wu *et al.,* 2021), and lung (Madissoon *et al.,* 2023). In brain, in particular, molecular imaging methods have been deployed to unravel the complex spatial relationships of thousands of cell types, resulting in atlases of cell types in adult mouse brain (Zhang *et al.,* 2021; Langlieb *et al.*, 2023; Shi *et al.*, 2023; Yao *et al.,*2023; Zhang *et al.,* 2023); cell types in adult human brain (Jorstad *et al*, 2023a) and in developing human brain (Braun *et al.*, 2023; Velmeshev *et al.*, 2023; Kim *et al.*, 2023); non-neuronal cells in the mouse nervous system (Zeisel *et al.*, 2018); interneurons in mouse, human and non-human primates (Bugeon *et al.*, 2022; Chartrand *et al.*, 2023; Jorstad *et al.*, 2023b; Lee *et al.*, 2023); DNA methylation and epigenomics in mouse brain (Liu *et al.*, 2023; Zhou *et al.*, 2023); and brain cell populations in Alzheimer's Disease (Gabitto *et al.,* 2023).

Already spatial technologies have enabled many discoveries in biology, but the field of spatial transcriptomics remains immature. Errors may arise during tissue preparation, chemistry, and imaging, resulting in erroneous detection and identification of transcripts. In principle, the sources of these many errors are known. In practice, often it's unclear how often errors occur, how to best detect and describe the resulting imperfections in the results, and how these imperfections impact downstream analyses such as cell type identification.

For more mature technologies, often the main sources of error are known, there's consensus on correction strategies, and these corrections are implemented in widely used analysis software suites. In high-throughput RNA sequencing, for example, RNA is fragmented, reverse transcribed to cDNA, and mapped to a known genome, and the number of raw counts per transcript varies with transcript length, GC content, and sequencing depth. No single procedure corrects for all possible errors but various normalization strategies are widely used to minimize within-sample and between-sample effects (Leek *et al.*, 2010; Oshlack *et al.*, 2010; Conesa *et al.*, 2016; Evans *et al.*, 2018) and there's awareness that misinterpretation of results may occur where biological and technical effects are correlated and normalization is inadequate (Conrads *et al.*, 2004; Baggerly *et al.*, 2004; Liotta *et al.*, 2004; Spielman *et al.*, 2007; Akey *et al.*, 2007, Spielman & Cheung, 2007). Ideally, there would be a consensus around common errors and corrections for spatial transcriptomics, where there is not yet the same emphasis on quality control of results before downstream analyses.

Here we characterize imperfections on MERSCOPE, a commercial platform using Multiplexed Error-Robust single molecule Fluorescence In Situ Hybridization (MERFISH) chemistry (Chen *et al.*, 2015; Moffitt & Zhuang, 2016; Moffitt *et al.*, 2016). We collected results from 641 adult mouse sections over ~2 years, developed code to detect and characterize the most common imperfections, and describe the frequency with which each imperfection occurred and its likely impact on cell type identification. Our results indicate that imperfections are common and reduce the accuracy of cell type labels but are rarely severe enough to prevent investigation of the spatial organization of cell populations in adult mouse brain.

Our code, called MerQuaCo, complements existing packages that facilitate the analysis of spatial transcriptomics datasets, including packages focused on data storage and access, e.g. Pysodb, SpatialData (Lin *et al.,* 2024; Marconato *et al.,* 2024); cell segmentation, e.g. cellpose, Baysor (Stringer *et al.,* 2021; Petukhov *et al.,* 2022); and analysis of high-dimensionality spatial data, e.g. Seurat, scanpy, Giotto, squidpy (Sajita *et al.*, 2015; Wolf, Angerer & Theis, 2018; Dries *et al.,* 2021; Palla *et al.,* 2022; Hao *et al.,* 2024). Adding MerQuaCo, or comparable procedures, to existing workflows offers an alternative to time-consuming and subjective manual assessment of data quality, streamlining the analysis of large spatial datasets and supporting rigorous, objective assessment of results generated with spatial molecular imaging technologies.

METHODS

We collected MERFISH results using the Vizgen MERSCOPE platform (https://vizgen.com/products/). Procedures for sample preparation were as described by the Vizgen User Guide (https://vizgen.com/resources/fresh-and-fixed-frozen-tissue-sample-preparation) with modifications in Yao *et al.* (2023). Experiments were conducted on fresh frozen P14-56 mouse brain tissue sectioned at 10 µm onto MERSCOPE coverslips, fixed and permeabilized, hybridized with encoding probes, gel embedded and cleared, and stained with DAPI and polyT to facilitate the

94 identification of somata. Samples were then loaded into the MERSCOPE, which manages sequential fluid exchange
95 and imaging. We excluded 11 experiments with transcripts per μm^2 per gene <0.0002, yielding 641 sections.
96
97 **Pixel Classification**
98 Sample preparation-related errors can arise during MERFISH experiments. Common problems include damage to
99 the section, resulting in the loss of a region of tissue, and detachment from the coverslip, resulting in a localized
100 region of tissue being too far from the coverslip surface to be within the imaging volume.
101 We built a pixel classifier to quantify the area of each section affected by common prep-related problems. The
102 classifier generates a series of binary masks then combines these masks in a final step, resulting in the
103 classification of each location in the section into one of 5 categories: tissue (tissue within the imaging volume),
104 detachment (tissue present but outside the imaging volume), ventricle (no tissue in the imaging volume but no loss
105 of tissue), damage (no tissue in the imaging volume due to loss of tissue), off-tissue (outside the section).
106 The intermediate transcript, DAPI, detachment, ventricle, and damage masks were created via mostly binary
107 image operations based on two outputs of the MERSCOPE: the transcript table (which provides the location and
108 gene identity for each transcript) and the DAPI image. To generate these masks we used Random Forest models
109 implemented in ilastik, an interactive image classification, segmentation, and analysis tool
110 (https://www.ilastik.org/).
111 From the transcript table, we plot an image of transcript counts with 10 x 10 µm pixels, including all transcripts
112 except blanks. The transcript image was converted into a binary mask, the transcript mask, by application of a
113 random forest model trained on binary images from 7 sections, manually annotated to distinguish tissue from all
114 other pixel classes.
115 From the high resolution DAPI image, we selected plane 0, downsampled in x and y by a factor of 100, and
116 thresholded to remove off-tissue pixels. The resulting modified DAPI image was converted into a binary mask, the
117 DAPI mask, by application of a random forest model trained on modified DAPI images from 6 sections, manually
118 annotated to distinguish DAPI-positive and -negative regions.
119 The detachment mask was created by subtracting the transcript mask from the DAPI mask. Detached tissue
120 manifests as regions with completely missing transcripts and blurry DAPI signal so the transcript mask excludes
121 detached regions and the DAPI mask includes them.
122 Our probe panels generally include a few genes expressed preferentially around the boundary of ventricles. We
123 leveraged these genes to distinguish ventricles from regions of tissue damage. We made a list of 11 ventricle-
124 associated genes: Crb2, Glis3, Inhbb, Naaa, Cd24a, Dsg2, Hdc, Shroom3, Vit, Rgs12, Trp73. For each section, we
125 plot transcript density images for all ventricle-associated genes in the probe panel, thresholded each image, and
126 combined all images via an AND operation to create a ventricle outline image (figure 1D). The ventricle outline was
127 summed with the DAPI mask (creating an image with pixel values of 0, 1 and 2) to which we applied a random
128 forest model trained on images from 7 mouse brain sections with annotated ventricles, resulting in the ventricle

129 mask. Of our 641 sections, 20 were imaged with panels lacking any of the 11 ventricle-associated genes. The
130 results in figure 2 were therefore generated from 621 sections.
131 We investigated two strategies for generating the damage mask. The first began with the modified DAPI image,
132 which was binarized, dilated and eroded, and subtracted from the DAPI mask. This intermediate image often
133 created a mask in which regions of damage were detected but incomplete and with a thin, erroneous strip of
134 damage around the section boundary. The second strategy involved inversion of the DAPI mask followed by
135 elimination of pixels outside the tissue boundary (via a flood fill operation seeded at the origin). This intermediate
136 image often captured ventricles and damage within the section but excluded regions of damage along the section
137 boundary. We summed the two intermediate images, capturing damage within the section and along its boundary.
138 The resulting binary image commonly included ventricles, removed by subtraction of the ventricle mask.
139 The transcript, detachment, ventricle, and damage masks can contain conflicting pixel labels. To obtain the final
140 pixel classification the 4 masks were combined, with unassigned pixels being off-tissue.
141
142 **Perfusion Rate**
143 MERSCOPE outputs a log file that includes the perfusion flow rate (in unspecified units, likely milliliters per minute)
144 at one second intervals during solution exchange. Post-hoc examination of the log file can reveal possible
145 inconsistencies of the flow rate during the experiment, due to blockage of the tubing for example. During a typical
146 MERSCOPE run, median flow was consistently > 1.5 ml/min. Occasionally, solution failed to flow throughout an
147 experiment. We identified experiments with ≥1 solution exchange with a median flow rate <0.5 ml/min. In the 303
148 experiments for which perfusion log files were available, 30 (10%) had ≥1 compromised reagent solution exchange.
149
150 **Data Loss**
151 Like other image-based spatial technologies, MERSCOPE acquires and stitches together many images to map
152 transcript density across a section of tens of millimeters. The absence of an image is readily visualized as a square
153 hole in a plot of transcript locations. To capture this type of data loss, we developed an iterative algorithm that
154 calculates the ratio of transcript counts between every on-tissue field of view (FOV) and its cardinal neighbor FOVs
155 for every gene. In the first iteration, we preliminarily assigned a target FOV as experiencing data loss if the ratio of
156 transcript counts is below 0.15 for any 3 of its 4 cardinal neighbors. We did this to allow the possibility of two
157 neighboring FOVs experiencing data loss; requiring a difference below 0.15 for all 4 cardinal neighbors would result
158 in a high false negative rate in the case of adjacent FOVs with data loss. In a second iteration, we began by
159 removing target FOVs from consideration if the mean transcript counts of their 4 cardinal neighbors is below 100,
160 effectively filtering regions of overall low transcript detection. We then assigned a target FOV as experiencing data
161 loss if the ratio of transcript counts is below 0.15 for all 4 cardinal neighbors, unless one of its neighbors was
162 preliminarily determined to be experiencing data loss in the first iteration, at which point we considered that
163 target FOV to be experiencing data loss in comparison to 3 neighbors.

164 The third and final iteration aimed to eliminate false positives: FOVs marked as a potential site of data loss after
165 the second iteration where data was not lost. In the third iteration, codebook information was used to determine
166 whether genes in each FOV identified in iteration 2 were in the same round. For each FOV, data loss was excluded
167 where the lost genes were not compatible with the codebook.
168
169 **Detection efficiency across the section: periodicity metric**
170 In MERSCOPE images, the transcript count often varies along the x and y axes with a periodicity of ~200 µm, the
171 size of a field-of-view (FOV). We developed a periodicity metric to describe the uniformity of detection efficiency in
172 the two cardinal (x- and y-) axes across the section. We began by computing a histogram of transcript density in
173 one dimension for one imaging plane (transcripts per µm, either the x- or y-dimension). We divided the histogram
174 into 202 µm segments, approximating the dimensions of a FOV, normalized each segment to its mean transcript
175 density, and calculated the mean of all segments, finally calculating the minimum/maximum density ratio. We
176 repeated this procedure for each of the 7 z-planes and for x- and y-axes, resulting in 14 minimum/maximum ratios.
177 The periodicity metric was the least of these 14 ratios.
178
179 **Detection efficiency through the section: p6/p0 ratio**
180 Sections were 10 µm thick. The MERSCOPE images z-planes at 1.5 µm intervals, starting 1.5 µm from the coverslip
181 surface (plane 0). For a 10 µm section, MERSCOPE acquires a stack of seven image planes extending to 10.5 µm
182 (plane 6). Transcript counts often differed across z-planes, generally declining with distance from the coverslip
183 surface. We quantified the gradient by taking the ratio of transcript counts in planes 6 and 0, the p6/p0 ratio. A
184 p6/p0 ratio of 1 corresponds to uniform transcript detection along the z-axis, while a p6/p0 ratio of 0 indicates a
185 failure to detect transcripts in the plane furthest from the coverslip.
186
187 **Transcript Density**
188 Transcript density should vary across a tissue section due to differences in gene expression but the mean density
189 per gene should vary little between sections, for sections with comparable RNA quality. The mean transcript
190 density therefore provides an overview of the quality, particularly when benchmarked against a dataset of similar
191 experiments. We calculated transcript density as the mean counts for all on-tissue transcript species (excluding
192 blanks) divided by the area of the on-tissue regions of the sections. Units of transcript density are counts per
193 transcript species per μm^2.
194
195 **Public datasets**
196 We downloaded and analyzed publicly available datasets from four commercial platforms: Vizgen MERSCOPE, 10x
197 Genomics Xenium, NanoString CosMx, and Resolve Molecular Cartography. All datasets include transcript tables,
198 which form the basis for our analyses. All datasets were accessed in July 2024.

199 The MERSCOPE datasets were animal 1 replicate 2 from the Vizgen MERFISH Mouse Liver Map, a 10 µm thick
200 section imaged with a 347 gene panel (https://info.vizgen.com/mouse-liver-access) and three 10 µm thick coronal
201 sections from three mouse brains in the Vizgen Data Release V1.0 May 2021, imaged with a 483-gene panel
202 (https://info.vizgen.com/mouse-brain-data).
203 The Xenium datasets were a 5 µm thick formalin-fixed paraffin-embedded (FFPE) coronal mouse brain
204 hemisection from the TgCRND8 mouse model of amyloid precursor protein overexpression, 17.9 months of age,
205 imaged with a 347 gene panel (https://www.10xgenomics.com/datasets/xenium-in-situ-analysis-of-alzheimers-
206 disease-mouse-model-brain-coronal-sections-from-one-hemisphere-over-a-time-course-1-standard) and three 10
207 µm thick fresh frozen coronal mouse brain sections, imaged with a 248 gene panel
208 (https://www.10xgenomics.com/datasets/fresh-frozen-mouse-brain-replicates-1-standard).
209 The CosMx dataset is a 10 µm thick FFPE human frontal cortex section imaged with a 6078 gene panel
210 (https://nanostring.com/products/cosmx-spatial-molecular-imager/ffpe-dataset/human-frontal-cortex-ffpe-
211 dataset/).
212 The Molecular Cartography dataset is a coronal mouse brain hemisection imaged with a 100 gene panel
213 (thickness not stated, https://resolvebiosciences.com/open-dataset/?dataset=mouse-brain-2021)
214
215 **Variability of transcript density across sections**
216 In figure 8D, we estimated the variability in transcript counts or density across experiments for MERSCOPE, Xenium
217 and CosMx. Molecular Cartography was excluded since results were available for only one tissue section. Datasets
218 varied in size (2 experiments each in Xenium and CosMx from figure 2A of Cook *et al.* (2023); 3 experiments in
219 Xenium fresh-frozen-mouse-brain-replicates-1-standard dataset; 59 experiments in Yao *et al.* (2023) MERSCOPE
220 dataset) and metric measured (median transcript count per cell in Cook *et al.* (2023); transcript density per gene in
221 Xenium fresh-frozen-mouse-brain-replicates-1-standard dataset; transcript density per gene in Yao *et al.* (2023)
222 MERSCOPE dataset). To enable comparison across metrics and datasets, we calculated the mean coefficient of
223 variation (CV) of all pairwise combinations of experiments. Importantly, the mean CV is independent of sample size
224 and is identical when calculated from transcript density or counts per cell. Equal transcript counts between
225 experiments would result in a CV of 0. CV increases linearly with differences in transcript counts.
226
227 **Availability of MerQuaCo**
228 MerQuaCo (for MERSCOPE Quality Control) is a Python package available on Github:
229 https://github.com/AllenInstitute/merquaco.
230 Documentation: https://merquaco.readthedocs.io/en/latest.

231 **RESULTS**

232 Our aim was to develop code and characterize data quality for each tissue section processed on our MERSCOPE
233 platforms. We developed code to quantify commonplace imperfections and assess quality by comparing each
234 section to the distribution across 641 mouse brain sections. Our dataset, collected over 2 years on 8 MERSCOPE
235 systems, includes the 59 adult mouse brain coronal sections published in Yao *et al.* (2023) and freely available
236 through the Allen Brain Cell Atlas (https://portal.brain-map.org/atlases-and-data/bkp/abc-atlas). Below, for each
237 imperfection we provide an example, describe our code, and consider the likely effects of each imperfection on
238 cell type identification.

239

240 **Tissue preparation**

241 A MERSCOPE experiment starts with sectioning, each tissue section being placed onto a coverslip. After several
242 benchtop chemistry steps, the coverslip is assembled into a flow chamber and then loaded into a MERSCOPE for
243 automated imaging. Common failures during tissue preparation include localized damage resulting in the loss of
244 part of the tissue section, and localized detachment of part of the section from the coverslip. Both result in data
245 loss, the former because tissue is lost and the latter because tissue is present but outside the volume imaged by
246 the MERSCOPE, which extends 10.5 µm from the coverslip surface.

247 Even in the absence of damage and detachment, some regions of the coverslip lack tissue. These include regions
248 outside the section boundary and tissue-free regions within the section. In brain sections, the latter includes
249 ventricles. Some of the quality metrics we measure in our MERSCOPE experiments, such as transcript density,
250 require that we distinguish regions of the coverslip with and without tissue and we therefore begin our analysis
251 with code that locates tissue.

252 Ideally, the transcript table would include transcripts only where there's tissue; there would be no transcripts in
253 regions of the coverslip without tissue. In practice, every experiment includes transcripts where there's no tissue.
254 Often, the transcript density in some off-tissue regions exceeds that in some on-tissue regions, preventing the use
255 of a simple threshold to identify regions of the coverslip containing tissue. We therefore developed a pixel
256 classifier, which converts the transcript table into a transcript density image and applies a random forest classifier,
257 trained using 10 manually annotated images. The result is the transcript mask, a binary mask which initially
258 classifies each pixel as on- or off-tissue (figure 1A).

259 We developed our classifier to further categorize off-tissue pixels, resulting in a classification of each pixel into
260 one of five categories: tissue, detached, ventricle, damage, and off-tissue. As inputs, our pixel classifier takes two
261 outputs of the MERSCOPE experiment: the DAPI image and the transcript table. Our strategy was to generate four
262 image masks, each a binary map of one class of pixel (transcript mask, damage mask, detachment mask, ventricle
263 mask) and combine the masks into a single image with our five pixel classes.

264 Where tissue becomes detached from the coverslip, and is outside the imaging planes of the MERSCOPE, the
265 transcript count is low. Although slightly out of focus, DAPI fluorescence is usually present in regions of

266 detachment (figure 1B, DAPI image). By subtracting the transcript mask from a DAPI mask (generated from the
267 DAPI image with the use of a random forest classifier) we created a detachment mask (figure 1C). For the ventricle
268 mask, we mapped transcripts associated with endothelial cells, which line the ventricle, again using a random
269 forest classifier to convert transcript density to a mask (figure 1D). The damage mask was generated from the DAPI
270 image via a series of binary operations (figure 1E). The final classification was created by summing damage,
271 transcript, detachment, and ventricle masks in sequence (figure 1F).
272 To assess the accuracy of the pixel classifier, a test dataset of a 1 x 1 mm subregion from each of 621 sections
273 was manually annotated for damage, tissue, detachment, ventricles, and off-tissue, to which the pixel classification
274 results were compared. (The remaining 20 sections were imaged without probes for endothelial cell marker
275 genes.) Pixel classification was >90% accurate for 567 (91%) of 621 subregions (Figure 2A). Typically, the tissue
276 classifier reported <10% detachment, <5% ventricles, and <10% damage (figure 2B, 12 sections with detachment
277 >10%, 4 sections with ventricles >5%, 11 sections with damage >10%). The classifier was prone to detect minor
278 detachment, ventricles and damage in their absence. To quantify the false positive rate for detachment, ventricles
279 and damage, we ran the classifier on 20 sections with no detachment, 20 without ventricles, and 20 undamaged
280 sections. False positive rates were <2% detachment, <1% ventricles, and <4% damage (figure 2B, insets). Only 78
281 (12.6%) of 621 sections had >2% detachment and likely were partially detached from the coverslip during imaging.
282 117 (18.8%) of 621 sections had >4% damage and likely suffered some tissue loss due to damage during
283 preparation. There was no significant change in detachment, tissue area or damage over ~2 years of MERSCOPE
284 experiments so a few percent detachment and damage is routine in our MERSCOPE experiments (figure 2C;
285 Pearson correlation coefficients and p-values: detachment -0.043, 0.28; tissue 0.031, 0.45; damage 0.072, 0.073).
286 In summary, the classifier estimated tissue area, detachment and damage with reasonable accuracy. We used
287 tissue area in the calculation of subsequent metrics, such as transcript density, and the incidence of common
288 problems in tissue preparation to monitor our tissue preparation and handling procedures.
289
290 **Differences in transcript density between MERSCOPE experiments**
291 Once imaging is complete the MERSCOPE runs automated image analysis procedures, returning a transcript table,
292 with the locations and gene identity for each transcript, and a cell-by-gene table, with cell locations and a list of
293 transcripts within each soma. Our analyses of data quality focus on the transcript table. Ideally, the probability of
294 detection of an RNA molecule would be invariant: detection probability would be identical in every experiment,
295 through space within each section, and for different genes. Gene-specific differences in detection probability are
296 almost inevitable with probe-based methods in which the number of target sequences and probe binding differs
297 between genes, but we find that detection probability also varies from experiment to experiment, and often
298 through space within each experiment.
299 Transcript counts often varied substantially between sections, even for two neighboring sections from the same
300 mouse (figure 3A). For closely spaced sections of similar area, probed with the same gene panel, we expect modest

301 inter-section differences in transcript density due to differing expression patterns through the brain. In practice,
302 inter-section differences in transcript density were approximately 2-fold for a series of sections from a single
303 mouse (collected in a single sectioning session and processed over several weeks; figure 3B). Across 641 sections,
304 transcript density was distributed approximately normally with a mean of 0.0056 and standard deviation of 0.0023
305 transcripts per gene per square micrometer of tissue (figure 3C). Transcript density differed across probe panels
306 but the variability in transcript density changed little through time (figure 3D).
307 Why does transcript density vary across sections? Variability was substantial across sections from a single mouse,
308 where tissue quality would have been comparable for all sections, so differences in preparation are unlikely to be
309 responsible. Transcript count in plane 0, just 1.5 µm from the coverslip surface, correlated tightly with total
310 transcript count (figure 3E) and mean transcripts per soma correlated with transcript density (figure 3F) so varying
311 thickness across sections and variable loss of transcripts from the tissue surface during MERSCOPE chemistry are
312 unlikely mechanisms. We conclude that transcript detection efficiency varies across sections, resulting in a broad
313 range of 200-1000 transcripts per soma in a single adult mouse (figure 3G). The mechanism underlying these batch
314 effects remains unclear, but variability in MERSCOPE chemistry, imaging and image analysis all remain candidates.
315
316 **Data loss**
317 Transcript density should vary across each section due to differences in gene expression through the tissue.
318 Additionally, artifactual gradients and abrupt changes in transcript density can be introduced by the MERSCOPE.
319 The most abrupt changes in transcript density result from simple data loss. The field of view (FOV) of the
320 MERSCOPE is ~200 µm x ~200 µm so to image a tissue section the MERSCOPE tiles many images. If an image is lost,
321 the likely result is the loss of transcripts, readily visualized as a square hole in a plot of transcript locations (figure
322 4A). The MERSCOPE images three spectrally distinct readout bits in each imaging round. Loss of one of the three
323 images would result in the loss of information on one readout probe. Each readout probe binds to transcripts from
324 tens to hundreds of genes (for panels used here, 60-104 genes). The loss of one bit from the barcode may
325 complicate decoding and decrease the accuracy of detection for many transcript species, but the effect will likely
326 be greatest for transcript species to which the missing readout probe binds. Hence data loss tends to occur for
327 groups of genes, linked by a shared readout probe. Whether the data loss is visible for each of the genes depends
328 on the density of transcripts for each gene in the surrounding regions. In short, data loss typically occurs for
329 multiple but rarely all transcript species and the number of species may not be readily apparent from the
330 transcript table.
331 For each transcript species, we quantified data loss by comparing transcript counts in each field of view to its
332 four cardinal neighbors. A transcript species was considered lost from a field of view if the counts were less than
333 15% of its cardinal neighbors, with a subsequent false positive correction step. Where a gene was lost its transcript
334 count was 5.1 ± 3.9% of the mean of its cardinal neighbors. As expected, where data was lost from a field of view,
335 often tens of transcript species were missing (figure 4B). Data loss occurred in 201 (31%) of 641 sections and was

336 mostly limited to a few isolated locations with <3 missing fields of view in 133 (66%) of 201 sections (figure 4C).
337 Occasionally more substantial data loss was observed, including loss from up to 51 fields of view for a single
338 section, and 120 transcript species in a single field of view. For sections in the Yao *et al.* (2023) Allen Brain Cell
339 Atlas dataset, 16 of 59 (27%) sections suffered data loss but for no section was there loss from >8 fields of view,
340 with ≤32 transcript species lost from each field of view.
341 The loss of transcript species reduces the accuracy of cell type labels. For the Yao *et al.* (2023) Allen Brain Cell
342 Atlas dataset with a 500-probe panel (VA142) we calculated the effect of omitting 40-200 transcript species (figure
343 4D). The loss of 40 transcript species changed the cluster labels of ~10% of cells so we expect label transfer to be
344 less accurate in fields of view where even one readout bit is lost from the barcode through the loss of an image.
345 The prevalence of data loss changed substantially over ~2 years on our MERSCOPEs, as acquisition firmware was
346 updated (figure 4E). Data loss was relatively common with versions 232b and 233 of the acquisition firmware,
347 available in mid-to-late 2023. Data loss has been less common with more recent firmware, such as version 233b,
348 but occurs with all versions of the acquisition firmware.
349
350 **Spatially heterogeneity of transcript detection**
351 Ideally, detection efficiency would be uniform throughout the tissue. In practice, detection efficiency is not
352 spatially uniform and there are inter-experiment differences in the non-uniformity. We characterized transcript
353 density in all three cardinal optical axes of the MERSCOPE.
354 Across the tissue section (in the x and y axes) we expect transcript counts to vary due to differences in gene
355 expression. We observed an additional source of variation: transcript counts varied along x and y axes with a
356 periodicity of ~200 µm, indicating that detection efficiency varied systematically across each field of view (figure
357 5A). We characterized the uniformity of detection efficiency with a periodicity metric. We calculated the variation
358 in transcript density across the mean field of view, in x- and y-axes independently, and calculated the
359 minimum/maximum density ratio for x- and y-dimensions for each of the seven z-planes (figure 5B), using the
360 minimum of these 14 values to describe non-uniformity of detection efficiency for each section. A periodicity
361 metric of 1 indicates that detection efficiency was uniform; a value of 0 indicates that no transcripts are detected
362 in part of the field of view.
363 Detection efficiency varied across the field of view for all sections, with the variation differing substantially
364 across sections. The median periodicity metric for all 641 sections was 0.80, with a long tail extending towards zero
365 (figure 5C). 40 exhibited a periodicity metric of <0.6. For the 59 sections in the Yao *et al.* (2023) Allen Brain Cell
366 Atlas dataset, the median periodicity metric was 0.84 and the range 0.66-0.93. Based on simulations with the Yao
367 *et al.* (2023) Allen Brain Cell Atlas dataset, we expect reduced detection efficiency, effectively the loss of
368 transcripts, to reduce the accuracy of label transfer. The loss of 20% of transcripts changes the cluster labels of 10-
369 15% of cells, and the loss of 40% of transcripts changes the cluster labels of 15-20% of cells (figure 5D). The
370 differences in detection efficiency across the field of view have changed little over ~2 years (figure 5E).

371 Along the optical axis (z axis, perpendicular to the plane of the tissue section), the MERSCOPE acquires images in
372 7 locations separated by 1.5 µm. Ideally, transcript detection efficiency would be equal in all 7 images, but we
373 routinely observed more transcripts in imaging planes closer to the coverslip than in planes near the tissue-
374 solution interface. Figure 6A shows transcripts from two neighboring sections from the same mouse brain
375 (collected from A-P locations 200 µm apart). In the first section, the transcript count is uniform along the optical
376 axis (figure 6A, B). In the second section, the transcript count is comparable to that in the first section near the
377 coverslip, consistent with similar gene expression in two closely spaced sections, but transcript count declines with
378 distance from the coverslip, to ~10% 10.5 µm from the coverslip (Figure 6A, B). We quantify homogeneity of
379 detection efficiency along the optical axis with the ratio of transcript counts in planes 6 and 0 (10.5 and 1.5 µm
380 from the coverslip, p6/p0 ratio). Uniform detection efficiency along the optical axis corresponds to a p6/p0 ratio of
381 1. A p6/p0 ratio of 0 indicates a steep decline in transcript detection with distance from the coverslip, such that no
382 transcripts are detected 10.5 µm from the coverslip.
383 The p6/p0 ratio was skewed towards 0 with a median of 0.34 (mean of 0.39, figure 6C). The p6/p0 ratio
384 distribution was shifted towards 1 for sections in the Yao *et al.* (2023) Allen Brain Cell Atlas dataset, with a median
385 of 0.74. p6/p0 ratio changed little over time (figure 6D). To improve homogeneity in detection efficiency in the
386 imaged volume, we cut thicker tissue sections while maintaining the number and separation of imaging planes (7
387 planes at 1.5 µm intervals, figure 6E). The p6/p0 ratio was ~1 for 20 µm sections, but at a cost of fewer transcripts
388 close to the coverslip and fewer total transcripts (mean ± SEM transcript count per μm^2, summed along the z-axis:
389 2.31 ± 0.20 for six 10 µm sections, 2.53 ± 0.28 for three 14 µm sections, 2.20 ± 0.33 for three 20 µm sections). 14
390 µm thick sections proved a good compromise, with transcript numbers near the coverslip comparable to 10 µm
391 sections (0.38 ± 0.04, 3 sections vs 0.39 ± 0.04, 6 sections) and a p6/p0 ratio of 0.79 ± 0.04 (three sections, vs 0.42
392 ± 0.11 for six 10 µm sections).
393 The decline in transcripts with distance from the coverslip differed between sections from a brain so tissue
394 quality is unlikely to be a major factor. Our results provide little further insight into possible mechanisms, but the
395 access of solutions to the deep (near the coverslip) and superficial faces of the section differ during benchtop
396 chemistry and on the MERSCOPE, with the deep face being less accessible than the superficial face. Loss of RNA
397 from the section, preferentially from the superficial face, might cause the gradient in transcript detection.
398 Similarly, unbinding and loss of readout probes into wash solution during imaging would have a similar effect.
399 Although the mechanism is unclear, detection of transcripts is rarely uniform through the depth of a MERSCOPE
400 section.
401 In summary, the detection of transcripts in MERSCOPE experiments is rarely homogenous, varying in all 3 spatial
402 dimensions and between sections. Our simulations provide some sense of the magnitude of the resulting effects
403 on cell labels, but the variation in detection efficiency is complex enough that it's likely not possible to map the
404 accuracy of cell type labels throughout a section. More homogenous detection efficiency would facilitate the
405 interpretation of spatial results.

406 Across our fleet of 8 MERSCOPEs, we observed significant differences in the magnitudes of all imperfections
407 (ANOVA, $p<0.05$), but differences were slight. Overall, performance was similar across MERSCOPEs.
408
409 **Visual inspection**
410 MerQuaCo characterizes the most common imperfections in each section, based on the transcript table. There are
411 imperfections that are not detected by MerQuaCo, most often imperfections that are not evident in the transcript
412 table or imperfections that become apparent when comparing nearby sections. For every experiment, we view
413 results in the MERSCOPE Vizualizer, manually searching for imperfections. Figure 7 provides two examples of
414 imperfections that were rare, not detected by MerQuaCo, but were observed multiple times by manual inspection.
415 Figure 7A-C illustrate data loss in the DAPI image, visible as horizontal stripes through the left half of the section
416 (figure 7A, B) and resulting in the local loss of somata within the image, and transcripts that cannot be assigned to
417 a soma. Although DAPI information is lost, transcripts are observed throughout the section (figure 7C), preventing
418 this imperfection from being detected by inspection of the transcript table or a transcript image.
419 Occasionally we observed imperfections that are evident only when comparing sections. For example, in figure
420 7D-F a region of thalamus is missing transcripts in one section (figure 7D). An abrupt change in transcript density
421 running along an anatomical boundary might result from localized expression, but in this instance the neighboring
422 section displays no comparable change in transcript density (figure 7E). Moreover, transcripts are lost from
423 thalamus in 4 of 6 neighboring sections (figure 7F). Clearly these differences are not biological: detection is
424 reduced >50% in thalamus in 2 of 6 sections, likely resulting in a marked decline in accuracy with which cell
425 populations in thalamus can be identified in these sections. MerQuaCo operates on individual sections so will not
426 detect imperfections that are evident only when comparing sections. We search for intra-section changes in
427 detection probability manually, by comparing results from nearby sections.
428
429 **Variations in transcript density on commercial spatial transcriptomics platforms**
430 Many of the imperfections described above occur in datasets collected with multiple spatial transcriptomics
431 platforms. We examined publicly accessible datasets from four commercial spatial transcriptomics platforms:
432 Vizgen MERSCOPE, 10x Genomics Xenium, NanoString CosMx, and Resolve Molecular Cartography. For some
433 sections, uneven detection across the field of view was visible by eye and was captured by our periodicity metric
434 (figure 8A, B). For all sections, transcript density varied along the z axis (figure 8C). For platforms where results
435 from multiple sections were available, we estimated differences in transcript count between sections, a proxy for
436 inter-experiment differences in detection efficiency (figure 8D). With only small numbers of sections available, the
437 results of this comparison should be considered preliminary but our results indicate that the imperfections we
438 have described, and that we detect and quantify with MerQuaCo, occur on spatial platforms other than
439 MERSCOPE. In some instances, imperfections are pronounced, underlining the potential value of applying
440 MerQuaCo to other platforms.

442 **DISCUSSION**

443 Here we have documented the incidence and magnitude of imperfections in image-based spatial transcriptomics
444 datasets, focusing on the most common imperfections in a dataset of hundreds of sections collected over ~2 years
445 on the MERSCOPE platform. In time, these imperfections may be eliminated by equipment manufacturers, but not
446 all the technical challenges have been solved in this new and rapidly evolving field and there is a need to
447 characterize imperfections that persist in spatial datasets. Unfortunately, residual imperfections are often not
448 obvious upon inspection of transcript or cell images. Our code, MerQuaCo, allows the user to detect and visualize
449 imperfections, assisting in the process of quality control.

450 Like many other groups, we use gene expression profiles from spatial datasets as the basis for cell type labels
451 (e.g. Yao *et al.*, 2023). Most imperfections do not prevent the identification of cell types but impact the accuracy of
452 labels, reducing confidence in labels and perhaps limiting the granularity with which cell populations can be
453 characterized. Which imperfections have the greatest impact on the accuracy of cell type labels?

454 Tissue damage and detachment from the coverslip, both of which result in local data loss, prevent all
455 downstream analyses for the affected regions of the section. These two imperfections affect all transcripts and are
456 therefore obvious on visual inspection of the dataset and are unlikely to lead to hidden errors in interpretation.

457 Like tissue damage and detachment, dropped images result in local data loss. In contrast with tissue damage and
458 detachment, typically dropped images result in loss of only a subset of transcript species. This is a critical
459 difference since many probe panels designed to identify cell types include some redundancy. In our experiments,
460 dropped images eliminated tens of genes from a panel of 500. Often, the effect on the accuracy of cell type labels
461 is modest, particularly for class and subclass labels. Dropped images may be more problematic where the aim of
462 the experiment is other than cell typing. Where the aim is to measure the expression of one or a small number of
463 genes, for example, dropped images may simply eliminate information on the genes of interest in affected regions.

464 When using spatial transcriptomics to locate genetically defined cell populations, the most impactful
465 imperfections are differences in transcript density between sections and through space within a section. Our
466 results indicate that transcript densities differ ~2-fold between sections, ~30% along the x and y axes, and ~5-fold
467 in z, and these effects are presumably multiplicative. The consequences can be substantial. Perhaps only ~50-60%
468 of cluster labels are accurate near the surface of a typical section. Furthermore, the effects of local changes in
469 transcript density are difficult to assess. One solution might be to discard results from sections or from regions of a
470 section where transcript density drops below a critical threshold. This threshold will depend on the aims of the
471 experiment, but MerQuaCo could facilitate such a solution by quantifying transcript density.

472 Our analyses of publicly accessible datasets indicate that some of the most common imperfections in our
473 MERSCOPE dataset also occur on other platforms. The accessible datasets are relatively small, often a few sections,
474 sometimes just part of a section, preventing a thorough comparison of imperfections across platforms. As a result,
475 our analyses only hint at the some of the relative strengths and weaknesses of different platforms. We expect that,
476 as with MERSCOPE, imperfections will differ across experiments on each platform, necessitating quality control to

477 identify experiments that meet the needs of the study. MerQuaCo could form the basis of such a quality control
478 process, with only minor changes to the code needed to enable the analysis of results from other platforms.
479 Previous authors have compared results across spatial transcriptomics platforms, focusing on high-dimensional
480 analysis of transcripts and cell expression profiles (Cook *et al.*, 2023; Wang *et al.*, 2023; Hartman & Satija, 2024).
481 Cook *et al.* (2023) compared Xenium and CosMx results from prostate adenocarcinoma samples; Wang *et al.*
482 (2024) compared MERSCOPE, Xenium, and CosMx results with FFPE tissue from multiple organs; and Hartman &
483 Satija (2024) compared results from fresh-frozen mouse brain slices across 6 spatial transcriptomics platforms.
484 These authors focused primarily on platform-specific differences in transcript specificity and sensitivity, cell
485 boundary identification, and the resulting differences in cell RNA content and classification. A consistent
486 conclusion was that results were generally reproducible, across samples processed on each platform, and across
487 platforms. These authors discussed the criteria by which they might select datasets for further analysis, and
488 discard others, implying that there is enough variability between experiments that not all datasets are equally
489 informative. For example, Wang *et al.* (2024) compared transcript counts per gene and pairwise correlation
490 coefficients and suggested that these measures might form the basis for decisions on which datasets to include or
491 discard. However, it remained unclear which parameters might best differentiate higher and lower quality
492 datasets, and how to use these parameters to select the highest quality datasets.
493 Our results indicate that variability across experiments is substantial and the consequences for downstream
494 interpretation can be significant, making it difficult to compare platforms. Firstly, one needs a large enough dataset
495 from each platform to differentiate between results of differing quality since collecting a small number of samples
496 from each platform leaves open the possibility that experiments were unusually successful on one platform and
497 unusually unsuccessful on another and that the resulting comparison misleads. Secondly, one needs to develop
498 and apply inclusion criteria for each platform that support a fair comparison. To better understand how spatial
499 transcriptomics platforms compare, we need additional studies based on large datasets with intentional filtering of
500 datasets before comparison. In short, our analyses complement previous studies, quantifying imperfections in
501 spatial results rather than comparing platforms. Combining our approach and those described by previous authors
502 would likely bring further insight into the relative performance of different platforms.
503 Despite the imperfections we've documented, numerous groups have published reliable results with spatial
504 methods. An example is our recent atlas of cell types in the adult mouse brain (Yao *et al.*, 2023). Many
505 imperfections exist in our published dataset, documented here. In our experience, the presence of imperfections
506 rarely prevents the collection of valuable results with spatial platforms. Rather, we regard the characterization of
507 imperfections as a quality control step that allows the identification and perhaps elimination of the weakest
508 datasets and the accurate interpretation of results, aware of the remaining imperfections and their possible
509 consequences. Ideally our study of imperfections combined with future studies will build a consensus and more
510 software tools for quality control, standardizing and streamlining data processing and further enhancing the
511 reliability of results, thereby facilitating more discoveries with spatially resolved molecular imaging methods.

512 **REFERENCES**

513 Akey JM, Biswas S, Leek JT, Storey JD (2007) On the design and analysis of gene expression studies in human
514 populations. Nature Genetics 39, 807–808.

515 Asp M *et al.* (2019) A spatiotemporal organ-wide gene expression and cell atlas of the developing human heart.
516 Cell 179, 1647–1660.

517 Baggerly KA, Edmonson SR, Morris JS, Coombes KR (2004) High-resolution serum proteomic patterns for ovarian
518 cancer detection. Endocrine-Related Cancer 11, 583–584.

519 Braun E *et al.* (2023) Comprehensive cell atlas of the first-trimester developing human brain. Science 382,
520 eadf1226.

521 Bugeon *et al.* (2022) A transcriptomic axis predicts state modulation of cortical interneurons. Nature 607, 330-338.

522 Chartrand T *et al.* (2023) Morphoelectric and transcriptomic divergence of the layer 1 interneuron repertoire in
523 human versus mouse neocortex. Science 382, eadf0805.

524 Chen KH, Boettiger AN, Moffitt JR, Wang S, Zhuang X (2015) Spatially resolved, highly multiplexed RNA profiling in
525 single cells. Science 348(6233), 412.

526 Conrads TP *et al.* (2004) High-resolution serum proteomic features for ovarian cancer detection. Endocrine-Related
527 Cancer 11, 163-178.

528 Cook DP *et al.* (2023) A comparative analysis of imaging-based spatial transcriptomics platforms. bioRxiv
529 2023.12.13.571385.

530 Close JL, Long BR, Zeng H (2021) Spatially resolved transcriptomics in neuroscience. Nature Methods 18(1), 23-25.

531 Dries R *et al.* (2021) Giotto: a toolbox for integrative analysis and visualization of spatial expression data. Genome
532 Biol 22, 78.

533 Gabitto M *et al.* (2023) Integrated multimodal cell atlas of Alzheimer's disease. bioRxiv 2023.05.08.539485.

534 Hao Y *et al.* (2024) Dictionary learning for integrative, multimodal and scalable single-cell analysis. Nat Biotechnol
535 42, 293–304.

536 Hartman A, Satija R (2024) Comparative analysis of multiplexed in situ gene expression profiling technologies.
537 bioRxiv 2024.01.11.575135.

538 Jorstad NL *et al.* (2023a) Transcriptomic cytoarchitecture reveals principles of human neocortex organization.
539 *Science* 382, eadf6812.

540 Jorstad NL *et al.* (2023b) Comparative transcriptomics reveals human-specific cortical features. Science 382,
541 eade9516.

542 Kim CN *et al.* (2023) Spatiotemporal molecular dynamics of the developing human thalamus. Science 382,
543 eadf9941.

544 Langlieb J *et al.* (2023) The molecular cytoarchitecture of the adult mouse brain. Nature 624, 333–342.

545 Lee et al. (2023) Signature morphoelectric properties of diverse GABAergic interneurons in the human neocortex.
546 Science 382, eadf6484.

547 Liotta LA, Petricoin EF, Veenstra TD, Conrads TP (2004) High-resolution serum proteomic patterns for ovarian
548 cancer detection. Endocrine-Related Cancer 11, 585–587.
549 Liu H *et al.* (2023) Single-cell DNA methylome and 3D multi-omic atlas of the adult mouse brain. Nature 624, 366–
550 377.
551 Lin S *et al.* (2024) Streamlining spatial omics data analysis with Pysodb. Nat Protoc 19, 831–895.
552 Madissoon E *et al.* (2023) A spatially resolved atlas of the human lung characterizes a gland-associated immune
553 niche. Nat Genet 55, 66–77.
554 Marconato L *et al.* (2024) SpatialData: an open and universal data framework for spatial omics. Nat Methods.
555 Moffitt JR, Hao J, Wang G, Chen KH, Babcock HP, Zhuang X (2016) High-throughput single-cell gene-expression
556 profiling with multiplexed error-robust fluorescence in situ hybridization. Proc. Natl. Acad. Sci. 113 (39) 11046-
557 11051.
558 Moffitt JR, Zhuang X (2016) RNA imaging with Multiplexed Error-Robust Fluorescence In Situ Hybridization
559 (MERFISH). Meth Enzymol 572, 1-49.
560 Palla G *et al.* (2022) Squidpy: a scalable framework for spatial omics analysis. Nat Methods 19, 171–178.
561 Petukhov V *et al.* (2022) Cell segmentation in imaging-based spatial transcriptomics. Nat Biotechnol 40, 345-354.
562 Sajita et al. (2015) Spatial reconstruction of single-cell gene expression data. Nature Biotechnol 33, 495-502.
563 Shi H *et al.* (2023) Spatial atlas of the mouse central nervous system at molecular resolution. Nature 622, 552–561.
564 Spielman RS, Bastone LA, Burdick JT, Morley M, Ewens WJ, Cheung VG (2007) Common genetic variants account
565 for differences in gene expression among ethnic groups. Nature Genetics 39, 226–231 (2007).
566 Spielman RS, Cheung VG (2007) Reply to “On the design and analysis of gene expression studies in human
567 populations”. Nature Genetics 39, 808-809 (2007).
568 Stringer C *et al.* (2021) Cellpose: a generalist algorithm for cellular segmentation. Nat Methods 18, 100–106.
569 Velmeshev D *et al.* (2023) Single-cell analysis of prenatal and postnatal human cortical development. Science 382,
570 eadf0834.
571 Wang H *et al.* (2023) Systematic benchmarking of imaging spatial transcriptomics platforms in FFPE tissues. bioRxiv
572 2023.12.07.570603.
573 Wu SZ *et al.* (2021) A single-cell and spatially resolved atlas of human breast cancers. Nat Genet 53, 1334–1347.
574 Yao Z *et al.* (2023) A high-resolution transcriptomic and spatial atlas of cell types in the whole mouse brain. Nature
575 624, 317–332.
576 Zeisel A *et al.* (2018) Molecular architecture of the mouse nervous system. Cell 174, 999-1014.
577 Zhang M *et al.* (2021) Spatially resolved cell atlas of the mouse primary motor cortex by MERFISH. Nature 598, 137
578 143.
579 Zhang M *et al.* (2023) Molecularly defined and spatially resolved cell atlas of the whole mouse brain. Nature 624,
580 343–354.

581 Zhou J *et al.* (2023) Brain-wide correspondence of neuronal epigenomics and distant projections. Nature 624, 355–
582 365.

583 **FIGURE LEGENDS**

584 Figure 1. A classifier to assess section integrity.

585 (A) Generation of the transcript mask. The transcript density image was converted to a binary mask using a
586 random forest classifier. (B) Generation of the DAPI mask. The modified DAPI image was converted to a binary
587 mask using a random forest classifier. (C) Generation of the detachment mask. The detachment mask was the
588 difference between DAPI and transcript masks. (D) Generation of the ventricle mask. A binary image summarizing
589 the locations of 11 ventricle boundary genes was summed with the DAPI mask and converted to a binary mask
590 with a random forest classifier. (E) Generation of the damage mask. Two intermediate masks were created via a
591 series of binary operations on the modified DAPI image and DAPI mask, then summed. The ventricle mask was
592 subtracted to remove ventricles. (F) Sequential combination of ventricle, detachment, transcript, and damage
593 masks resulted in the final 5-category image.

594

595 Figure 2. Tissue area, detachment and damage.

596 (A) Accuracy of the pixel classifier, evaluated on a test dataset consisting of a 1 mm^2 subregion from each of 621
597 tissue sections (grey). Black: 59 sections in the Yao et al. (2023) Allen Brain Cell Atlas dataset. (B) Probability
598 distributions describing the percentage detachment, ventricle, and damage for each of 621 sections. Insets: false
599 positive distributions calculated for 20 sections without detachment, ventricles, or damage. (C) Percentage of the
600 section identified as on-tissue, detachment, and damage, plot over time (grey), and for the 59 sections in the Yao
601 et al. (2023) Allen Brain Cell Atlas dataset (black).

602

603 Figure 3. Transcript density.

604 (A) Transcript locations for two neighboring sections from the same mouse brain, separated along the A-P axis by
605 200 µm. (B) Transcript density across A-P locations for a single mouse. 59 sections in the Yao et al. (2023) Allen
606 Brain Cell Atlas dataset. (C) Histograms of transcript density per transcript species per square micrometer for 641
607 sections. Summed results from 4 gene panels (VA142, VA373, BP0770, VZG147). Black: 59 sections in the Yao et al.
608 (2023) Allen Brain Cell Atlas dataset (using VA142). Arrowheads: transcript densities for the two sections in A
609 (panel VA142). (D) Transcript density over time. Symbols indicate gene panels. (E) Comparison of transcript
610 densities in plane 0 and summed across all planes. Each data point represents one section from the Yao et al.
611 (2023) dataset. Pearson correlation coefficient 0.86, $p = 3.1 \times 10^{-18}$. (F) Mean transcripts per soma vs transcript
612 density. One data point per section, Yao et al. (2023) dataset. Pearson correlation coefficient 0.69, $p = 1.07 \times 10^{-9}$.
613 (G) Distribution of transcripts per soma. Yao et al. (2023) dataset.

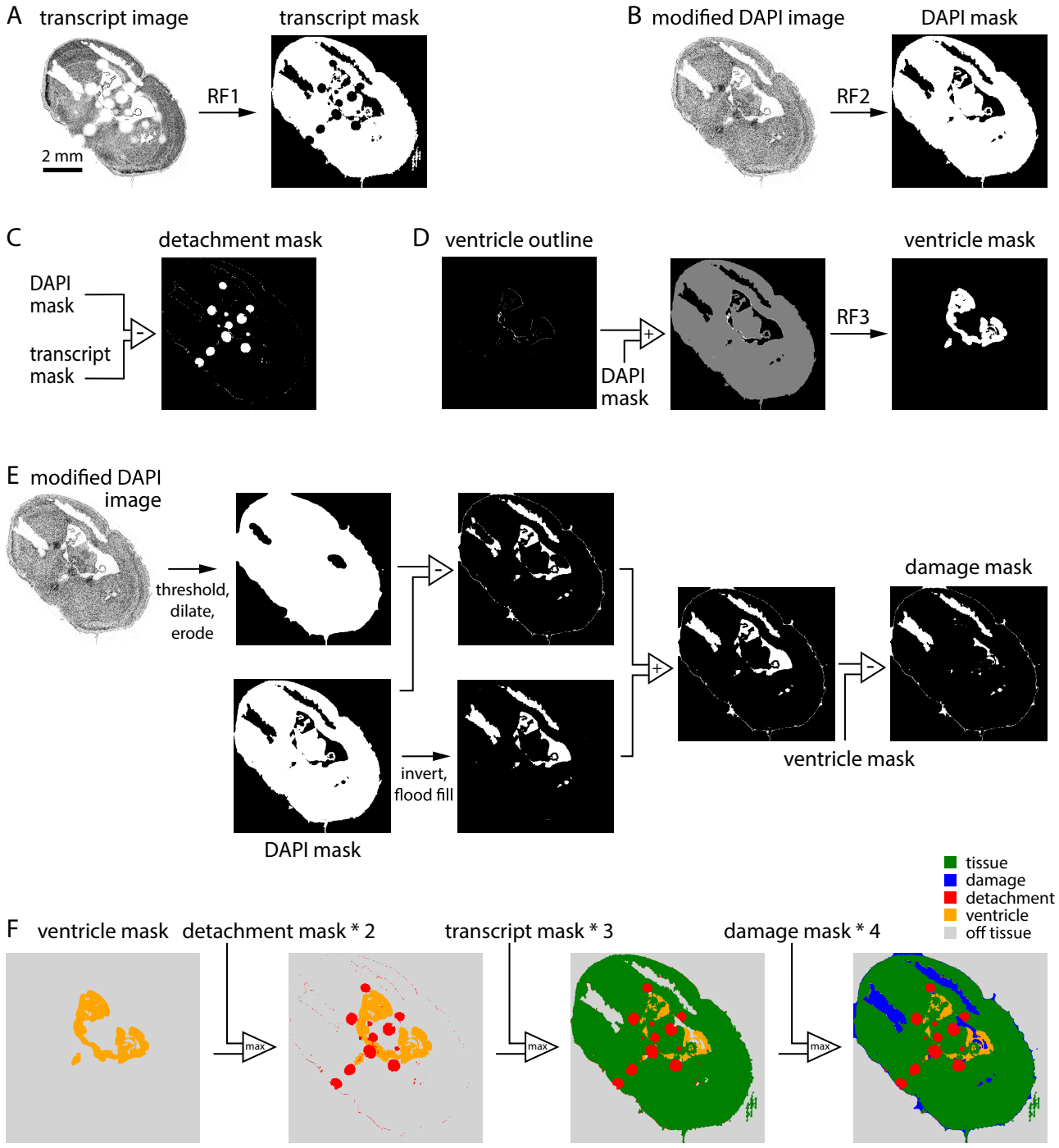
614

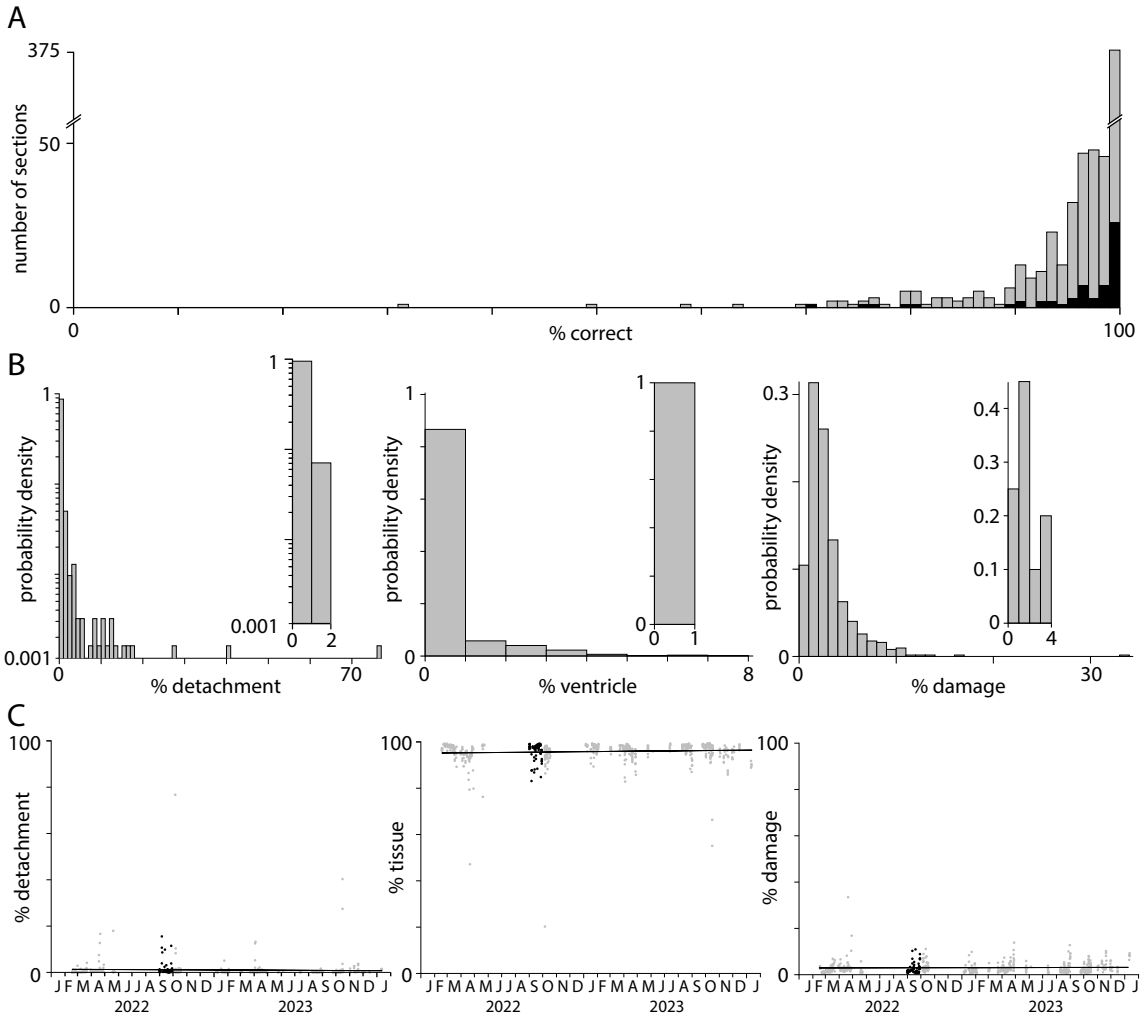
615 Figure 4. Dropped images cause local data loss.

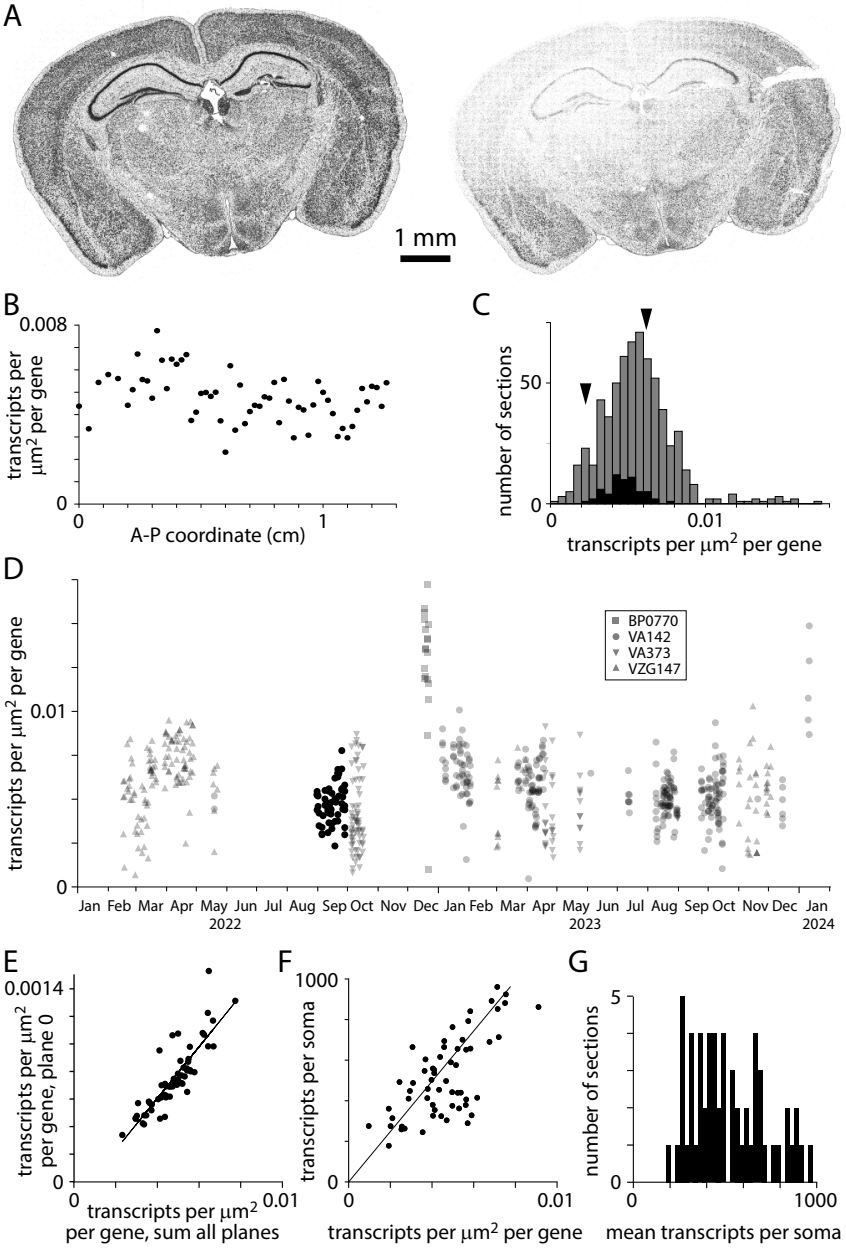
616 (A) Transcript locations for one transcript species: Gja1. Each point represents one Gja1 transcript. Inset:
617 transcripts around one region of data loss. (B) Output of the data loss detection routine, showing the number of

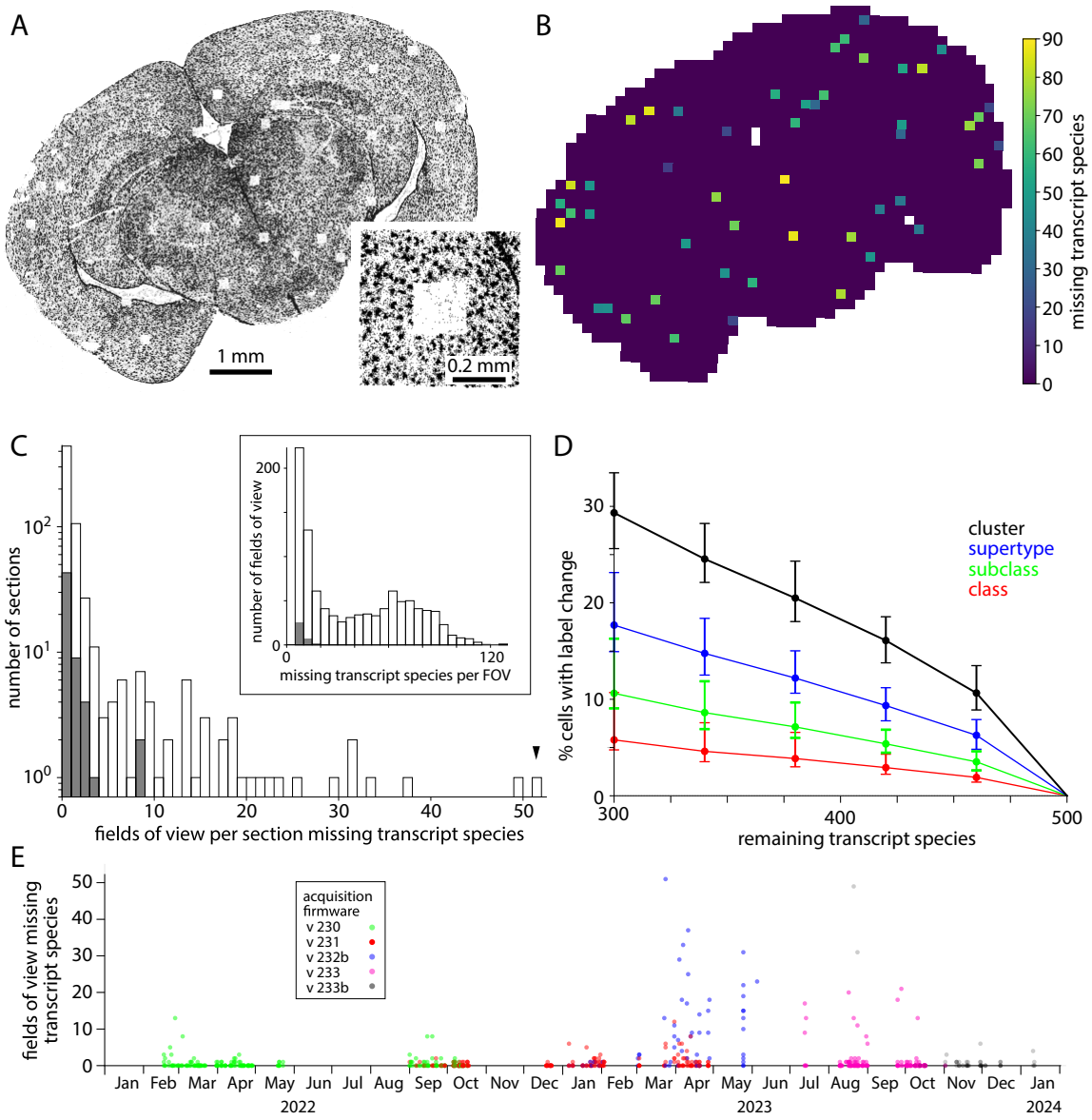
618 transcript species missing from each FOV. Not all FOVs with dropped genes are missing Gja1. White: off-tissue, as
619 determined by the classifier. (C) Histogram of missing fields of view across 641 mouse sections (grey) and for the
620 59 sections in the Yao et al. (2023) Allen Brain Cell Atlas dataset (blue). Arrowhead: results for the section in panels
621 A and B. Inset: Number of transcript species missing per affected field of view. (D) Effect of missing transcript
622 species on label transfer. Change in class, subclass, supertype and cluster labels calculated for 10,000 cells from the
623 Yao et al. (2023) Allen Brain Cell Atlas dataset. Median, minimum and maximum % change from 100 trials. Genes
624 to be removed were selected at random. (E) Number of missing fields of view over time and with different
625 acquisition software versions. Each point represents a tissue section.
626
627 Figure 5. Uneven detection efficiency across each field of view.
628 (A) Transcript locations for two coronal sections from the same brain, separated by 100 µm. To the left and below,
629 transcript densities summed along x and y axes. For the section on the left, changes in transcript density occur at
630 anatomical boundaries with little indication of variations in detection efficiency along x or y axes. For the section
631 on the right, superimposed on differences in genes expression are variations in detection efficiency with a
632 periodicity of 200 µm. (B) Periodicity metric, calculated for each z-plane along x and y axes, for the two sections in
633 A. Filled symbols, left example in A. Open symbols, right example in A. Black and grey, metric along x and y axes,
634 respectively. (C) Histogram of minimum periodicity metrics for 641 sections (grey) and for the 59 sections in the
635 Yao et al. (2023) Allen Brain Cell Atlas dataset (black). Arrowheads, the two sections in A. (D) Effect of reduced
636 detection efficiency on label transfer. Change in class, subclass, supertype and cluster labels for 10,000 cells from
637 the Yao et al. (2023) Allen Brain Cell Atlas dataset (VA142 500 probe panel). Median, minimum and maximum %
638 change from 100 trials. (E) Periodicity metric over time. Black: periodicity metric for the 59 sections in the Yao et
639 al. (2023) Allen Brain Cell Atlas dataset.
640
641 Figure 6. Uneven detection efficiency along the optical axis.
642 (A) Transcript locations in three z planes for each of two neighboring sections from the same mouse brain,
643 separated along the A-P axis by 200 µm. Distances are from the coverslip surface. (B) Transcript counts along the z-
644 axis for the section in panel A. (C) Distribution of p6/p0 ratio for 641 sections (grey) and the 59 sections in the Yao
645 et al. (2023) Allen Brain Cell Atlas dataset (black). Arrowheads, sections in A. (D) p6/p0 ratio over time. (E) Mean ±
646 SEM transcript counts along the z-axis for 10, 14 and 20 µm thick sections. 3 sections each at 14 and 20 um, 6 at 10
647 um.
648
649 Figure 7. Imperfections identified by manual inspection.
650 (A) DAPI image. Data loss results in horizontal stripes in the left half of the image. (B) DAPI in the sub-region in the
651 box in panel A. (C) Transcripts in the corresponding region. (D) Transcripts in a section 6.2 mm posterior to bregma.

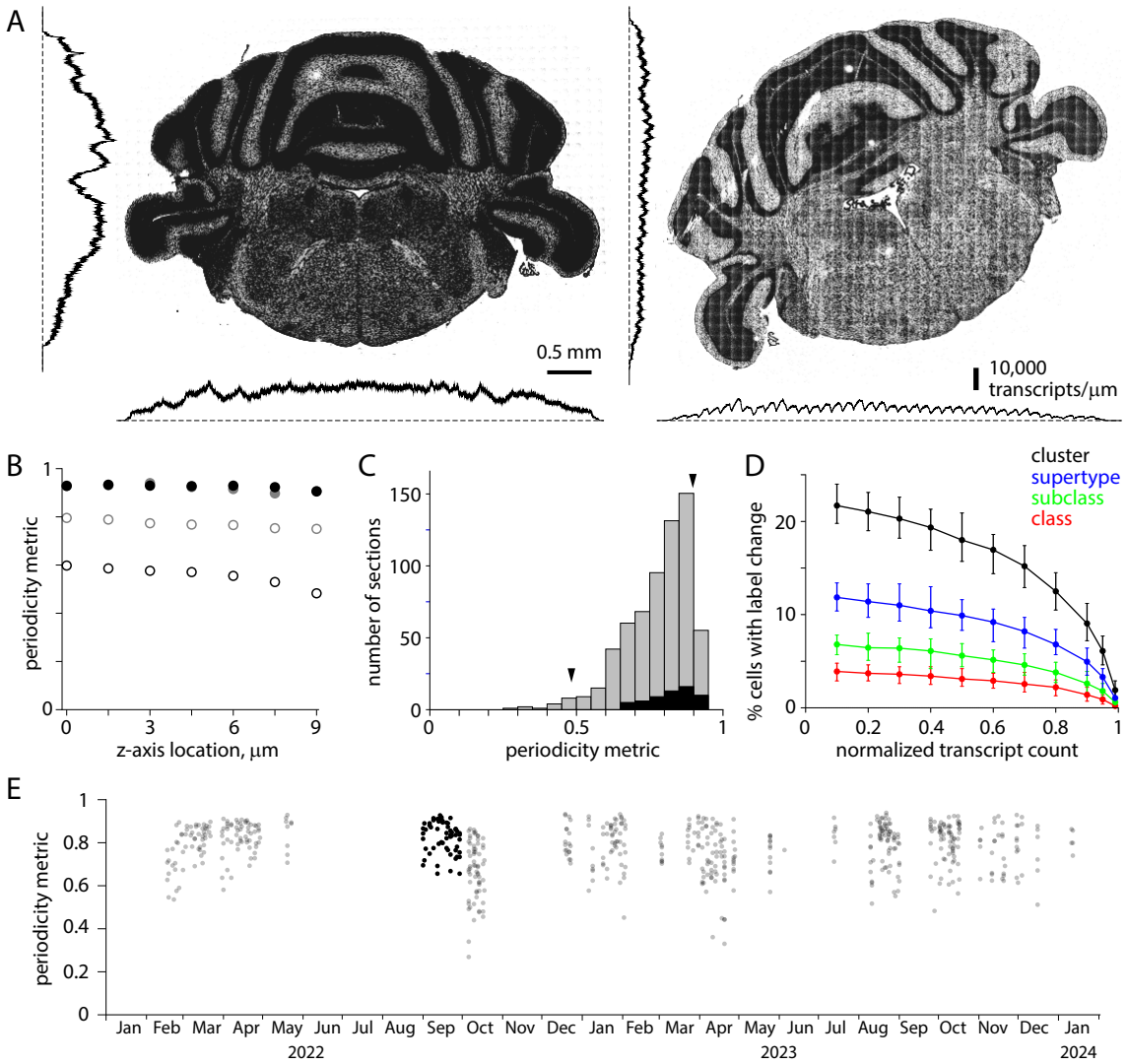
652 (E) Transcripts from a neighboring section. (F) Transcript density in cortex and thalamus (boxes in panels D and E)
653 for 6 neighboring sections.
654
655 Figure 8. Transcript density across spatial transcriptomics platforms.
656 (A) Example sections from four commercial platforms. MERSCOPE, mouse liver section. CosMx, human brain.
657 Xenium and Molecular Cartography, mouse brain. Scale bars, 1 mm. Histograms indicate transcript density in
658 (cardinal axes, normalized to peak). (B) Periodicity metric for public datasets (4 sections in A and 3 Vizgen mouse
659 brain datasets). Grey: mean ± stdev periodicity metric for the Yao et al. (2023) Allen Brain Cell Atlas dataset. (C)
660 Transcript count along the z axis for sections in A. (D) Pairwise CV of transcripts across experiments. MERSCOPE, 3
661 Vizgen mouse brain datasets and Yao et al. (2023) Allen Brain Cell Atlas dataset (open symbol). Xenium, fresh-
662 frozen-mouse-brain-replicates-1-standard dataset from 10x. Grey datapoints (Xenium and CosMx) from Cook et al.
663 (2023).

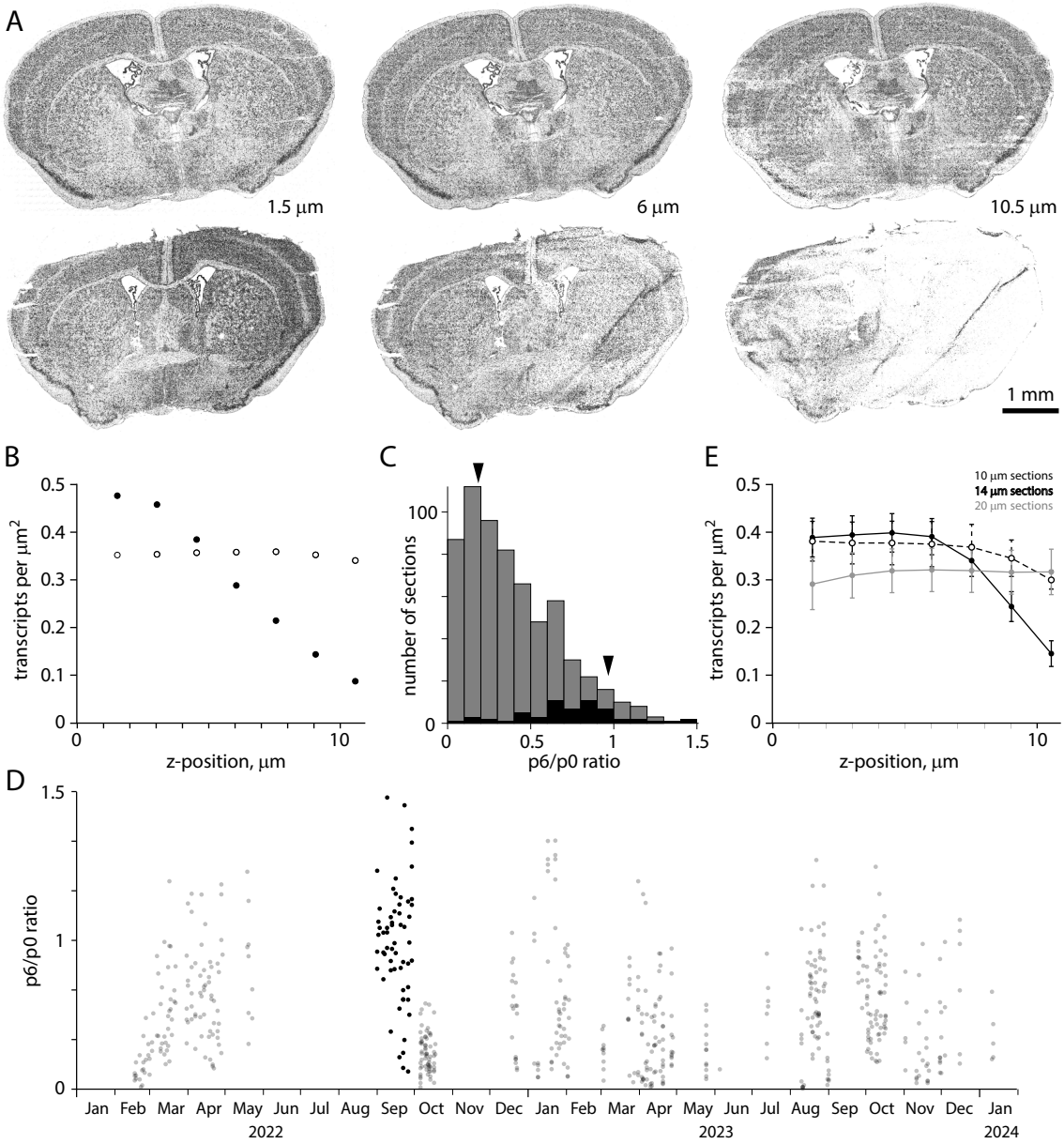

A
transcript image
transcript mask
RF1
2 mm
B
modified DAPI image
DAPI mask
RF2
C
detachment mask
DAPI mask
transcript mask
D
ventricle outline
ventricle mask
DAPI mask
RF3
E
modified DAPI image
threshold, dilate, erode
invert, flood fill
DAPI mask
damage mask
ventricle mask
tissue
damage
detachment
ventricle
off tissue
F
ventricle mask
detachment mask * 2
transcript mask * 3
damage mask * 4
max
max
max

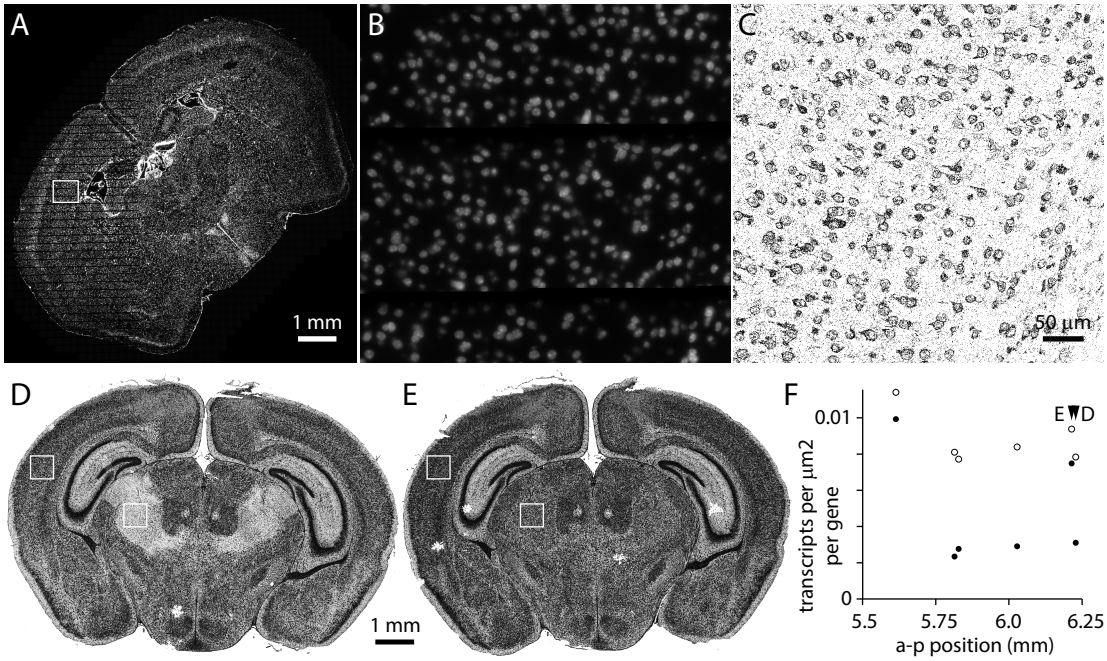

A
375
50
0
number of sections
0
% correct
100
B
1
0.001
probability density
0
% detachment
70
1
0.001
0
2
1
0
probability density
0
% ventricle
8
1
0
0
1
0.3
0
probability density
0
% damage
30
0.4
0.3
0.2
0.1
0
0
4
C
100
0
% detachment
J F M A M J J A S O N D J F M A M J J A S O N D J
2022
2023
100
0
% tissue
J F M A M J J A S O N D J F M A M J J A S O N D J
2022
2023
100
0
% damage
J F M A M J J A S O N D J F M A M J J A S O N D J
2022
2023

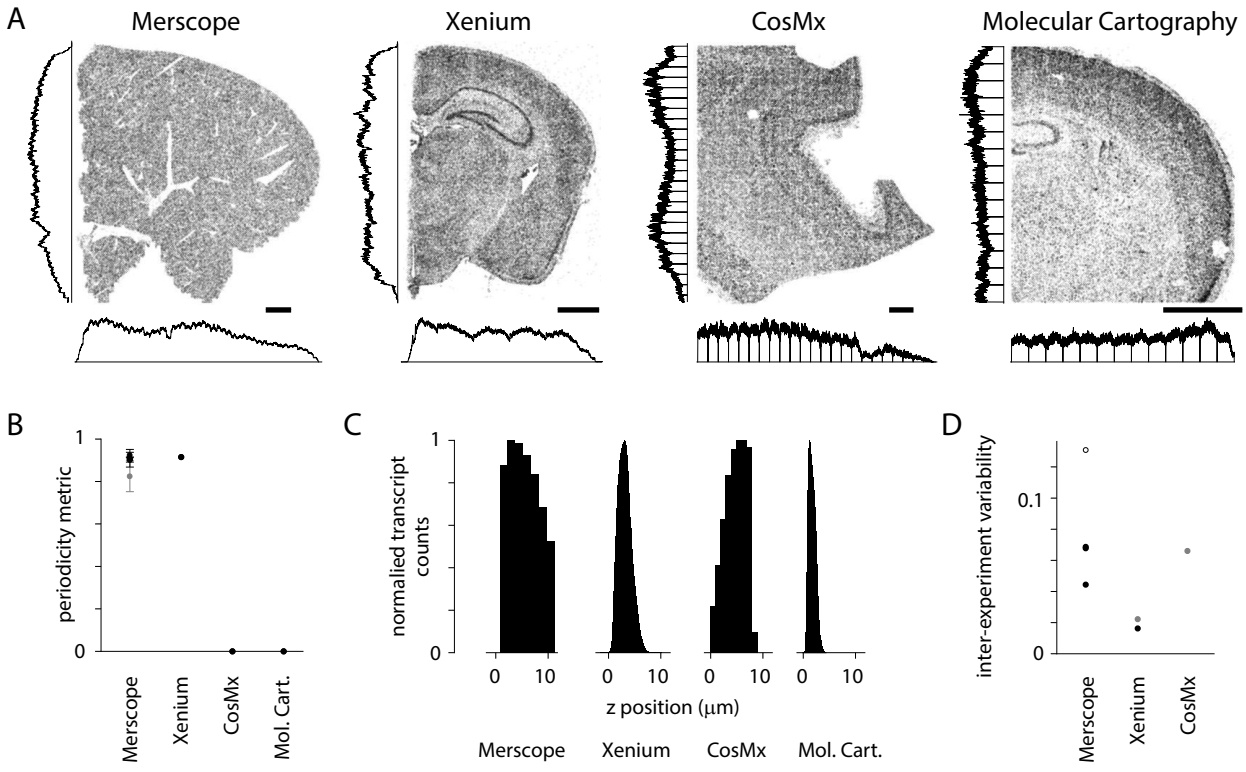

A
1 mm
B
0.008
transcripts per μm² per gene
0
0
1
A-P coordinate (cm)
C
50
number of sections
0
0
0.01
transcripts per μm² per gene
D
BP0770
VA142
VA373
VZG147
0.01
transcripts per μm² per gene
0
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec Jan
2022
2023
2024
E
0.0014
transcripts per μm² per gene, plane 0
0
0
0.01
transcripts per μm² per gene, sum all planes
F
1000
transcripts per soma
0
0
0.01
transcripts per μm² per gene
G
5
number of sections
0
0
1000
mean transcripts per soma


A
1 mm
0.2 mm
B
missing transcript species
90
80
70
60
50
40
30
20
10
0
C
number of sections
fields of view per section missing transcript species
number of fields of view
missing transcript species per FOV
D
% cells with label change
remaining transcript species
cluster
supertype
subclass
class
E
fields of view missing transcript species
acquisition firmware
v 230
v 231
v 232b
v 233
v 233b
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec Jan
2022
2023
2024


A
0.5 mm
10,000
transcripts/µm
B
periodicity metric
z-axis location, µm
C
number of sections
periodicity metric
D
cluster
supertype
subclass
class
% cells with label change
normalized transcript count
E
periodicity metric
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec Jan
2022
2023
2024


A
1.5 µm
6 µm
10.5 µm
1 mm
B
transcripts per µm²
z-position, µm
C
number of sections
p6/p0 ratio
E
10 µm sections
14 µm sections
20 µm sections
transcripts per µm²
z-position, µm
D
p6/p0 ratio
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec Jan
2022
2023
2024


A
1 mm
B
C
50 µm
D
E
1 mm
F
0.01
0
E▼D
transcripts per µm2
per gene
5.5
5.75
6.0
6.25
a-p position (mm)


A
Merscope
Xenium
CosMx
Molecular Cartography
B
periodicity metric
1
0
Merscope
Xenium
CosMx
Mol. Cart.
C
normalied transcript counts
1
0
0
10
z position (μm)
Merscope
Xenium
CosMx
Mol. Cart.
D
inter-experiment variability
0.1
0
Merscope
Xenium
CosMx