Gene panel selection for targeted spatial transcriptomics

Yida Zhang[1,2], Viktor Petukhov[1], Evan Biederstedt[1], Peter V. Kharchenko[1,3#]

[1]Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA.
[2]Department of Neurobiology, Duke University, Durham, NC, USA.
[3]San Diego Institute of Science, Altos Labs, San Diego, CA, USA.
[#]Correspondence: peter_kharchenko@hms.harvard.edu

Abstract

Methods for targeted spatial transcriptomic analysis hold particular promise in analysis of complex tissues. Most such methods, however, measure only a limited panel of transcripts, which need to be selected in advance to inform on the cell types or processes being studied. Here we describe gpsFISH, a computational method to perform such selection in a way that optimizes detection of previously annotated cell types or cell type hierarchies. We show systematic difference of transcript detection rates between different technologies, which distorts the resulting transcriptional profile estimates. Such distortions reduce the resulting ability to distinguish cell types in the targeted spatial transcriptomics data relative to what would be expected from scRNA-seq data. gpsFISH was designed to model and adjust for the platform effects, yielding more informative gene panels and better cell type classification. In addition, gpsFISH can incorporate the hierarchical structure of cell types to better account for complex cell type relationships. Finally, we observed a high redundancy of gene panels, and extended gpsFISH to incorporate custom gene preferences. Overall, gpsFISH outperforms other gene selection methods, improving support for hierarchical cell classification as well as flexible options to accommodate diverse design requirements.

Key words: gene panel selection, targeted spatial transcriptomics, single cell RNA sequencing, platform effect, cell type hierarchy

Background

The building block of complex tissues is the diverse range of cell types [1–4]. Knowing the identity and spatial location of cells from different cell types is the key for understanding how they communicate with each other to carry out specific functions and how diseases emerge when this complex network of interactions goes awry [5–11]. Single-cell RNA sequencing (scRNA-seq) provides a powerful tool to study the identity of cell types and cell states [12–17]. However, the spatial information is lost due to cell disassociation during library preparation. Recent advances in spatial transcriptomics technologies have overcome this limitation by providing ways to quantify gene expression while keeping the spatial information of cells, leading to more comprehensive and detailed understanding of diseases and normal functions [18–23].

Based on the number of transcripts that can be probed, spatial transcriptomics technologies can be broadly categorized as (1) targeted, measuring a limited panel of transcripts; and (2) untargeted, capturing all transcripts from the transcriptome. Targeted spatial transcriptomics include in situ hybridization (ISH)-based [24–28] and most in situ sequencing (ISS)-based methods [29–32]. Untargeted spatial transcriptomics include next-generation sequencing (NGS)-based methods [33–39].

Compared to untargeted spatial transcriptomics, targeted spatial transcriptomics can achieve high sensitivity and subcellular resolution. However, their targeted nature requires a panel of genes (from a few hundred to thousand) to be selected in advance to recognize cell types or processes relevant to the tissue being studied.

Gene selection methods are used to design gene panels. They can be classified into two major categories based on their gene selection objectives. One category with an imputation-based objective aims to select genes based on their ability to capture as much of transcriptional variation in the scRNA-seq data as possible. Examples range from simply selecting highly-variable genes to more advanced methods like L1000 [40], geneBasis [41], and SCMER [42]. Specifically, L1000 identifies the optimal set of 'landmark' transcripts that construct a reduced representations of the transcriptome. geneBasis finds genes that can yield a *k*-nearest neighbor graph that is similar to the "true" graph constructed using the entire transcriptome. SCMER aims to select genes that preserve the manifold of scRNA-seq data. Another category of gene selection method with a classification-based objective selects genes given their ability to reconstruct cell classifications or relationships. Examples range from selecting differentially expressed genes to more advanced methods like scGeneFit [43], RankCorr [44], and NS-Forest [45]. scGeneFit selects marker genes that jointly optimize cell type recovery using a label-aware compressive classification method. RankCorr is a rank-based one-vs-all feature selection method that selects marker genes for each cell type based on a sparsity parameter that controls the number of marker genes selected per cell type. NS-Forest is a machine learning-based marker

gene selection algorithm that uses the nonlinear attributes of random forest feature selection and a binary expression scoring approach to select the minimal combination of marker genes that captures the cell type identity in scRNA-seq data. All these methods can be used to design gene panels for targeted spatial transcriptomics technologies.

A key limitation of current gene selection methods is that they select genes purely based on scRNA-seq data without considering potential differences between scRNA-seq and the target spatial transcriptomics technologies. Such platform effects include systematic differences in capture efficiency of genes between platforms caused by technology-dependent factors such as detection technique and library preparation chemistry. Platform effects have been previously noted when comparing gene expression measurements from single-cell and single-nucleus RNA-seq on the same biological sample [46]. It has been recently demonstrated that platform effects also exist between scRNA-seq and spatial transcriptomics technologies [47–49], posing a challenge when transferring cell type information from scRNA-seq to spatial transcriptomics technologies. Such systematic difference of transcript detection rate across platforms can cause distortions of transcriptional profile estimates and lead to reduced performance of gene panels selected using only scRNA-seq data in terms of distinguishing cell types in the resulting targeted spatial transcriptomics data.

94 Besides platform effects, there are other complications involved in gene panel
95 selection. First, current classification-based gene selection methods [43–45] treat
96 cell types as equally dissimilar from each other. However, cell types are organized in
97 a hierarchical manner with cell subpopulations belonging to the same broad cell
98 type more similar to each other than subpopulations belonging to different broad
99 cell types [50–56]. Depending on the biological questions and capabilities of the
100 assays, a gene selection method could optimize for fidelity at lower cell type
101 resolution, or place increased emphasis on certain subgroups of cell types. More
102 generally, this is not only useful for selecting genes that inform on cell types but can
103 also be extended to selecting genes for other biological entities with a hierarchical
104 structure, e.g., gene ontology and pathways [57,58]. Second, both imputation-based
105 and classification-based gene selection methods select genes solely based on a pre-
106 defined objective function. However, in practice of gene panel design for targeted
107 spatial transcriptomics, there can be other criteria contributing to the gene
108 selection. Examples range from technical factors, such as ability to design probes for
109 targeting certain transcripts, to biological factors such as preferences for certain
110 pathways or marker genes commonly used in the literature. A framework that takes
111 such orthogonal preferences into consideration can be helpful in practice.
112
113 To address these challenges, we developed gpsFISH, a classification-based gene
114 selection method that models and adjusts for the platform effects between scRNA-
115 seq and targeted spatial transcriptomics technologies, yielding more informative
116 gene panels and better cell type classification compared to previously published

classification-based gene selection methods. In addition, gpsFISH provides options to account for cell type hierarchy and gene-specific custom preferences during gene panel design, offering flexible and finer control of cell type granularity and gene selection for different biological questions.

Results

Platform effects between scRNA-seq and targeted spatial transcriptomics

Even molecule counting assays carry inherent detection biases, posing challenges for joint analysis of multiple assays, such as scRNA-seq and spatially-resolved counts [47–49]. Indeed, we observed a systematic difference of transcript detection rate across platforms (**Fig. 1A-D**), which distorts the resulting transcriptional profile estimates. Consequently, a panel of genes selected based on scRNA-seq that works well on distinguishing cell types may not achieve similar level of performance when measured by targeted spatial transcriptomics.

To address this challenge, we estimate the level of gene expression distortion in targeted spatial transcriptomics data relative to scRNA-seq and from the same tissue using a Bayesian model (**Supplementary Fig. 1, Methods**). Bayesian inference estimates the posterior distribution of distortion magnitudes, which will be used to predict the potential distortion levels for genes that have not yet been observed in a given assay. Specifically, we assume platform effects are on a per gene basis. γ_i and c_i represent gene specific multiplicative and additive platform effect for each gene i, respectively. These distortion parameters are assumed to follow two

140 normal distributions with μ_γ, μ_c as mean and σ_γ, σ_c as standard deviation,
141 respectively. The posterior distribution of σ_γ and σ_c can be considered as a
142 generalized description of the magnitude of multiplicative and additive platform
143 effects. We can use them to sample the magnitudes of gene specific multiplicative
144 and additive platform distortions for unobserved genes. The model is fitted for a
145 given pair of scRNA-seq and targeted spatial transcriptomics platforms to account
146 for the differences between them.
147
148 To check the extent to which the model is able to capture platform biases, we used
149 three paired scRNA-seq and targeted spatial transcriptomics datasets [24,26]
150 (**Methods**). Fitting a model for each pair of datasets, we then performed posterior
151 predictive check, i.e., we simulated spatial transcriptomics measurements from
152 scRNA-seq data using the fitted Bayesian model (**Methods**). By comparing the
153 distribution of simulated and observed spatial transcriptomics measurements, we
154 observed that the Bayesian model can accurately recapitulate the platform effects
155 from different pairs of technologies (**Fig. 1E, Supplementary Fig. 2A-C**). The
156 posterior distributions of σ_γ and σ_c (**Fig. 1F-G**) on the three datasets showed
157 distinct levels of additive and multiplicative platform effects, indicating the need to
158 account for platform-specific properties during gene panel selection.
159
160 **Gene panel selection using genetic algorithm**
161 To take platform distortions into account during selection of the gene panels, we use
162 Bayesian model of the platform and scRNA-seq data to simulate spatial

163 transcriptomics measurements with distortions (**Methods**). We then evaluate a
164 gene panel's ability to recover the scRNA-seq cell type labels from such simulated
165 spatial measurements, rather than the original scRNA-seq measurements. Such
166 approach is intended to provide a more accurate estimation of panel performance in
167 a real spatially-resolved measurement. Instead of selecting top-performing genes,
168 gpsFISH optimizes the entire gene panel in its ability to recover cell type labels. To
169 optimize within this combinatorial gene space, gpsFISH uses genetic algorithm
170 [59,60] (**Fig. 2, Methods**).
171
172 Within each iteration of optimization, multiple cross validations of classification are
173 performed on each proposed gene panel. To avoid biasing towards a specific
174 realization of spatial transcriptomics distortions, gpsFISH performs the simulation
175 separately in each cross validation. As a result, gene panels that are more robust to
176 unexpected platform distortions will be favored. This gene panel selection
177 framework ensures the evaluation is reflective of the gene panel's real classification
178 performance when measured by specific targeted spatial transcriptomics
179 technologies.
180
181 We first tested gpsFISH on the Moffitt dataset, by optimizing a 200 gene panel to
182 distinguish "level 1" cell type annotation, which includes 12 broadly defined cell
183 types (**Fig 3A**). Most of the cells are correctly classified, yielding an overall accuracy
184 of 0.983 and high area under the receiver-operator curve (AUC) across all cell types
185 (**Fig 3B, C**). The optimized gene panel was also more successful in separating the 12

186 cell types on the resulting UMAP embedding compared to the gene panel selected
187 without considering platform effect (**Fig. 3D, E**).
188
189 To quantify performance of different methods we simulated spatial transcriptomics
190 data from scRNA-seq, separating training and test sets (**Methods**). Simulations
191 were performed both with and without distortions in order to evaluate how taking
192 platform effects into account impacts gene panel performance. We also compared
193 our method with two previously published classification-based gene selection
194 methods: RankCorr and scGenefit. Both methods rely on the scRNA-seq expression
195 profiles without considering platform effects. RankCorr is a rank-based one-vs-all
196 feature selection method that selects marker genes for each cell type given a
197 sparsity parameter, which controls the number of marker genes selected per cell
198 type. We tuned this parameter to make sure the panels generated using RankCorr
199 have the same size (200 genes). scGenefit selects gene markers that jointly optimize
200 cell label recovery using label-aware compressive classification methods. As control,
201 we provided a naïve way to simulate spatial transcriptomics measurements without
202 platform effects (**Methods**). In addition, we also generated a panel of randomly
203 selected genes as baseline.
204
205 The objective of gpsFISH optimization is to achieve high quality cell type
206 classification on the spatial transcriptomics data. This entails two tasks: (1)
207 selecting a good gene panel, and (2) using the gene panel for accurate cell type
208 classification. In practice, while the design of an initial gene panel may rely on the

209 scRNA-seq data, optimization of subsequent panels can take advantage of the probe-
210 specific distortions that have already been observed in earlier measurements.
211 Similarly, as more and more spatial transcriptomics data are generated, when
212 classifying cell types in a newly generated spatially-resolved measurement, it is
213 likely that some partial annotations may already be available for that platform
214 either on the current or previously acquired datasets. Regardless of the cell type
215 granularity of partial annotation, it contains gene-specific platform effect
216 information of genes in the spatial transcriptomics data, which can be estimated
217 using our Bayesian model to improve cell type classification. Following this logic, we
218 used two benchmark strategies, which evaluate the impact of platform effects on the
219 two tasks (**Methods**). Both strategies share the same general framework in which a
220 classifier is trained on the training data with gene expression profiles for all cell
221 types, and then applied onto the testing data for cell type classification. The
222 difference is how the two strategies incorporate partial annotation into the training
223 data when available. Specifically, for the first strategy, evaluation with platform
224 effect re-estimation (**Methods**), platform effects are estimated from the partial
225 annotation data and incorporated into the training data for all gene selection
226 methods. Since under this strategy the gene panels from all methods are evaluated
227 in the same manner, it is useful in evaluating the impact of platform effect on the
228 first task, i.e., selecting a good gene panel. In contrast, under the second evaluation
229 strategy, evaluation without platform effect re-estimation (**Methods**), only gpsFISH
230 panels are evaluated with platform effect estimation as described above (**Methods**),

231 illustrating the impact of platform effects on both tasks.

232

233 Evaluation with platform effect re-estimation on the Moffit dataset using naïve
234 Bayes as the classifier shows gpsFISH outperforms the control with naïve simulation
235 and other gene selection methods (**Fig. 4A**), indicating that taking platform effects
236 into consideration leads to more informative gene panels. Similar results were
237 observed for the Zhang and Codeluppi datasets (**Supplementary Fig. 3A and 3C**)
238 and using random forest as classifier (**Supplementary Fig. 4A-C**). From the
239 normalized confusion matrix of the gene panel selected by gpsFISH with
240 hierarchical tree on the left showing the relationship between cell types (**Fig. 4C**)
241 we can see that most of the misclassifications are within the complex
242 subpopulations of inhibitory and excitatory neurons.

243

244 A larger performance difference between gpsFISH and other gene selection methods
245 is observed using evaluation without platform effect re-estimation, especially when
246 the level of partial annotation is low (**Fig. 4B, Supplementary Fig. 3B and 3D,**
247 **Supplementary Fig. 4D-F**), indicating that considering platform effects can lead to
248 more accurate cell type classification.

249

250 Overall, the comparison results show that gpsFISH outperforms other gene selection
251 methods and considering platform effects can result in more informative gene
252 panels and better cell type classification.

253

254 **Redundancy in gene space across independent gene panel optimizations**

255 **enables incorporation of customized preferences**

256 Independent panel optimizations performed multiple times (10) for each of the

257 three datasets showed high level of redundancy in the gene space (**Fig. 5A**).

258 Specifically, despite similar levels of overall performance, the overlap between

259 independently optimized 200-gene panels was around 85, 65, and 35 genes, and

260 more than 20%, 30%, and 45% of the genes showed up in only one of the 10

261 optimized gene panels for the Zhang, Moffit, and Codeluppi datasets, respectively

262 (**Supplementary Fig. 5 A-C**). We observed similar level of redundancy even when

263 the optimization was performed for a more granular "level 2" cell type annotations

264 (46, 87, and 47 cell types for Zhang, Moffit, and Codeluppi dataset, **Supplementary**

265 **Fig. 5D**). The ability to achieve similar level of performance with different gene sets

266 suggests that the panels can be further optimized to accommodate secondary

267 criteria, such as inclusion of pre-selected genes, specific pathways, etc.

268

269 To illustrate how redundancy can be used to incorporate secondary preferences

270 with little impact on the classification performance, we added customized

271 preferences to gene panel selection. During the generation of ISH-based spatial

272 transcriptomics data, probes are designed to detect target genes of interest. The

273 number of probes that can be designed to target each gene is determined by gene-

274 specific factors like gene length. Since probes may fail during data generation, genes

275 with more probes are preferred due to the potential better sensitivity. Therefore,

276 the predicted number of probes for each gene in DARTFISH data (**Methods**) was
277 used to determine gene weight during gene panel selection. Of note, we capped the
278 probe count at 15 to avoid bias towards a small portion of genes with extremely
279 high number of probes (**Supplementary Fig. 6A**). This also agrees with the fact that
280 sensitivity will saturate when we have enough probes for a gene.
281
282 Following this approach, we performed 10 optimizations with and without probe
283 count gene weights on the Zhang dataset using "level 1" cell type annotations. As
284 expected, the optimizations with gene weight had slightly lower accuracy (**Fig. 5B**)
285 but achieved a significantly higher number of total probes (**Fig. 5C**). This
286 demonstrates that the redundancy of gene spaces allows one to incorporate
287 additional customized constraints/preferences based on orthogonal information to
288 design gene panels with preferred features without sacrificing the overall cell type
289 classification performance.
290
291 **Hierarchical gene selection based on cell type hierarchy**
292 Cell types are organized in a hierarchical manner with broad cell types divided into
293 more detailed subpopulations. This hierarchical relationship can be considered
294 when evaluating cell classification errors. For example, failure to distinguish two
295 closely related subtypes, such as Th1 and Th17, of cells is likely to be considered
296 less severe than mis-annotation of a Th cell into a different major cell type such as B
297 cells.
298

299 In addition to the default "flat" cell type evaluation, gpsFISH therefore, implements a
300 hierarchical classification option (**Fig. 6A, Methods**), in which correct classifications
301 or misclassification between different cell types will receive different credit/penalty
302 specified by a weighted penalty matrix according to cell type hierarchy. Using this
303 hierarchical classification framework, gpsFISH provides flexibility to customize
304 optimization based on desired level of cell type granularity.
305
306 To evaluate the effect of the hierarchical classification for gene selection, we
307 performed hierarchical gene selection at level 2 cell annotation of all three datasets.
308 Under a hierarchical penalty scheme, misclassifications of cells between different
309 level 1 categories incur a fixed penalty, whereas misclassifications within the same
310 level 1 category were given partial credit, proportional to the expression similarity
311 between the called and true subtypes (**Methods, Fig. 6B**). To quantify to what
312 extent this hierarchical classification framework reduces misclassifications across
313 broad cell types at level 1, we calculated the percentage of across broad cell type
314 mistakes over all mistakes (**Methods**). We observed that the optimized gene panels
315 using hierarchical classification tend to make significantly fewer misclassifications
316 across broad cell types at level 1 compared to flat classification (**Fig. 6C-E**),
317 indicating that cell type granularity can be controlled through the hierarchical
318 classification framework.
319

320 ## Discussion and conclusions

321 Accurate cell type classification is crucial for understanding the spatial relationship
322 of cells in complex tissues. We implemented gpsFISH, a method for gene panel
323 design of targeted spatial transcriptomics. By accounting for platform effects
324 between scRNA-seq and targeted spatial transcriptomics technologies, gpsFISH is
325 able to find more robust and informative gene panels and achieve better cell type
326 classification.
327
328 Different technology has different patterns of platform effects. Specifically, we
329 decomposed platform effects into two components: multiplicative and additive
330 platform effects. While the multiplicative effect has been considered in
331 deconvolution contexts (e.g., RCTD [47]), neither type of platform-specific
332 distortions have been considered by other gene selection methods. Among other
333 things, the additive platform effect enables gpsFISH to describe situations where
334 specific genes show no expression in scRNA-seq data, but is detected in spatial
335 transcriptomics data (dots forming the vertical line in **Fig. 1A** and **1B**). This
336 observation is common for osmFISH and MERSFISH, and cannot be modelled using
337 only multiplicative platform effect.
338
339 Comparing the three targeted spatial transcriptomics platforms, we found highest
340 levels of additive platform effects in DARTFISH (Zhang) data, followed by osmFISH
341 (Codeluppi) and MERFISH (Moffit) (**Fig. 1G**). More specifically, DARTFISH had the
342 lowest μ_c, indicating the highest level of signal reduction compared to MERFISH and
343 osmFISH (**Supplementary Fig. 2E**). Signal reduction increases the possibility of

344 good marker genes from scRNA-seq lose cell type specificity in spatial
345 transcriptomics data, which is a main scenario where platform effects affect gene
346 panel selection. Higher level of signal reduction for DARTFISH agrees with our result
347 that the performance difference between gpsFISH and other gene selection methods
348 is the largest in the Zhang dataset compared to the other two datasets. In addition,
349 platform effects have higher impact on gene panel selection when there is a higher
350 level of signal reduction also explains our evaluation with platform effect re-
351 estimation result on the Zhang dataset that gene panel selection using naive
352 simulation without considering platform effects performs similarly to gpsFISH that
353 considers platform effects (**Supplementary Fig. 3A and 4A**). Specifically, the
354 average number of detected molecules per cell per gene in the DARTFISH data is
355 lower than the MERFISH and osmFISH data (0.055 vs. 2.176 and 5.839). In addition,
356 the probe failure rate for DARTFISH (Zhang) was estimated to be around 3.5%,
357 compared to 0% for the other two datasets. Therefore, although there is no explicit
358 platform effect considered during naïve simulation, the significantly lower empirical
359 cell depth and non-zero probe failure rate for DARTFISH creates a mechanism for
360 signal reduction, which has similar effect to a negative additive platform effect (i.e.,
361 low μ_c) and make the gene panel selected from naïve simulation robust to platform
362 effects. To test this hypothesis, we performed gene panel selection using a modified
363 naïve simulation model where we increased the empirical cell depth of observed
364 DARTFISH data by 100-fold to match the cell depth magnitude in the osmFISH
365 dataset, and set the probe failure rate to 0 to remove signal reduction and see if the
366 resulted gene panel is still robust to platform effects. Evaluation with platform effect

367 re-estimation showed that after removing signal reduction in the modified naïve
368 simulation model, the gene panel from gpsFISH outperforms the one from the
369 modified naïve simulation model, followed by other published gene selection
370 methods (**Supplementary Fig. 7**). This verifies our hypothesis that the similar
371 performance between gpsFISH and naïve simulation we observed from the Zhang
372 dataset (**Supplementary Fig. 3A and 4A**) is because the naïve simulation
373 procedure implicitly includes "additive platform effects" to account for signal
374 reduction through low cell depth and non-zero probe failure rate of the DARTFISH
375 data. This result serves as another evidence that considering platform effects yield
376 more informative and robust gene panels, especially for targeted spatial
377 transcriptomics technologies with higher level of signal reduction.
378
379 In addition to additive platform effect, multiplicative platform effect also contributes
380 to the systematic difference of transcripts detection rate across technologies, posing
381 a challenge when transferring cell type information from scRNA-seq to spatial
382 transcriptomics technologies. Comparison of three targeted spatial transcriptomics
383 platforms shows osmFISH has the highest level of multiplicative platform effect,
384 followed by MERFISH and then DARTFISH (**Fig. 1F**). The high level of multiplicative
385 platform effect leads to poor cell type classification when there is no or low level of
386 partial annotation (**Fig. 4A and 4B**, **Supplementary Fig. 3C and 3D**), especially for
387 evaluation without platform effect re-estimation due to distorted expression
388 profiles between scRNA-seq and targeted spatial transcriptomics technologies.
389 However, for the Zhang dataset, the cell type classification accuracy has minor

390 variation across different levels of partial annotation (**Supplementary Fig. 3A** and
391 **3B**). We suspect this is due to the relatively lower level of multiplicative platform
392 effects of DARTFISH, and the classifier we used (naïve Bayes) is not sensitive to the
393 expression difference caused by this low level of platform effect between the
394 training and testing data. Therefore, we performed the same evaluation using
395 random forest as classifier and this time, we observed poor classification
396 performance when there is no or low level of partial annotation (**Supplementary**
397 **Fig. 4**). Specifically for evaluation with platform effect re-estimation, low level of
398 partial annotation provided limited statistical power to accurately estimate gene
399 specific platform effects, thus not able to increase the classification performance.
400 This reduced performance is gone when we have more than one cell type included
401 in the partial annotation, indicating partial annotation of a few cell types is enough
402 to enhance cell type classification. Similar results are observed for the Moffit and
403 Codeluppi dataset (**Supplementary Fig. 4**).
404
405 Redundancy across independent optimizations allows incorporation of customized
406 preferences into gene selection. However, gene weight needs to be carefully
407 specified to ensure no sacrifice on overall gene panel performance. For the result in
408 **Fig. 5B** and **5C**, we capped the number of probes for each gene at 15. For cutoffs
409 lower than 15, gene weight difference between genes are small, leading to gene
410 panels with similar performance but also similar total number of probes. However,
411 for cutoffs higher than 15, the optimization will bias towards a small group of genes
412 with high probe count, resulting in local minimum during optimization

413 (**Supplementary Fig. 6B and C**). This does achieve panels with significantly higher
414 total number of probes, but the classification accuracy is dropped. This emphasizes
415 the need to test different ways for gene weight specification in order to get the
416 expected result without sacrificing performance.
417
418 Similarly, in our test of hierarchical gene selection, we specified the weighted
419 penalty matrix directly from cell type hierarchy. Although we reduced
420 misclassifications across broad cell types, the overall accuracy is slightly lower than
421 flat classification (**Supplementary Fig. 8**). This shows that partial credit of
422 misclassifications needs to be given carefully, especially when there are many
423 similar subpopulations within the same broad cell type like in the Moffit dataset. In
424 real usage, it is suggested to prone the weighted penalty matrix constructed from
425 the cell type hierarchy to remove unnecessary partial credit. Gene panel selection
426 using flat classification can be run first to help adjust the weighted penalty matrix
427 constructed using cell type hierarchy. In addition, the hierarchical classification
428 provides a generic framework to fine tune emphasize of classification on certain cell
429 types. Here we showed its usage to incorporate cell type hierarchy, but it is not
430 restricted to cell type hierarchy. Customized weighted penalty matrix can be
431 constructed using other information that provides preferences towards different
432 classifications.
433
434 A major goal of spatial transcriptomics is to understand the spatial distribution of
435 cell types and their corresponding cellular environment. gpsFISH facilitates this by

436 selecting more informative and robust gene panels and providing ways for better
437 cell type annotation. We also provide options to account for various custom
438 preferences. As more targeted spatial transcriptomics data are generated, we expect
439 that gpsFISH can facilitate the study of cellular organization of complex tissues
440 under different biological contexts.
441

442 Method

443 Datasets

444 We used three datasets in our study. Two of them (Moffit and Codeluppi) are
445 publicly available. We specifically focused on datasets that have both scRNA-seq and
446 targeted spatial transcriptomics data. In addition, the datasets should have cell type
447 annotation at multiple levels of granularity. Information regarding the three
448 datasets is summarized in **Table 1**. Further processing details are discussed below.
449

450 ***Moffit dataset***

451 scRNA-seq data was downloaded from Gene Expression Omnibus (GEO,
452 www.ncbi.nlm.nih.gov/geo) (GSE113576). MERFISH data was downloaded from
453 Dryad (https://datadryad.org/stash/dataset/doi:10.5061/dryad.8t8s248). Of note,
454 the MERFISH data from Dryad is normalized and batch corrected. We undid the
455 volume normalization and batch correction to get the original data.
456

457 In the scRNA-seq data, we first filtered out cells annotated as "Ambiguous" and
458 "Unstable". We then used information in the supplementary Table 1 of the original

459 study to assign cell types. “Cell class (determined from clustering of all cells)”
460 column was used as level 1 cell type annotation. “Neuronal cluster (determined
461 from clustering of inhibitory or excitatory neurons)” and “Non-neuronal cluster
462 (determined from clustering of all cells)” were used as level 2 cell annotation.
463 Normalization was performed as described in the original study.
464
465 Only MERFISH data from naïve mice was used (to match scRNA-seq data). In
466 addition, we also filtered out cells annotated as “Ambiguous” and “Unstable”. Fos
467 gene and five blank genes were filtered out. 135 genes imaged in the combinatorial
468 smFISH imaging were kept. Following the naming of cell types in Fig. 3D of the
469 original study, we modified the cell type annotation of MERFISH data to make it
470 consistent with the scRNA-seq data. Specifically, at level 1 cell type annotation, cells
471 annotated as "Endothelial 1", "Endothelial 2", "Endothelial 3" were merged into
472 “Endothelial”. “Astrocyte” was changed to “Astrocytes”. "OD Immature 1" and "OD
473 Immature 2" were changed to “Immature_oligodendrocyte”. "OD Mature 1", "OD
474 Mature 2", "OD Mature 3", and "OD Mature 4" were changed to
475 “Mature_oligodendrocyte”. “Pericytes” was changed to “Mural”. At cell type level 2,
476 “Endothelial 1”, “Endothelial 2”, and “Endothelial 3” were changed to
477 “Endothelial_1”, “Endothelial_2”, and “Endothelial_3”, respectively. “Ependymal” was
478 changed to “Ependymal_1”. “OD Immature 1” and “OD Immature 2” were changed to
479 “Immature_oligodendrocyte_1” and “Immature_oligodendrocyte_2”, respectively.
480 “OD Mature 1”, “OD Mature 2”, “OD Mature 3”, and “OD Mature 4” were changed to

481 “Mature_oligodendrocyte_1”, “Mature_oligodendrocyte_2”,

482 “Mature_oligodendrocyte_3”, and “Mature_oligodendrocyte_4”, respectively.

483

484 After the processing above, additional filters were applied on the raw and

485 normalized scRNA-seq data before gene panel selection. Genes with maximum cell

486 type average expression lower than 1 were filtered out. In addition, long non-coding

487 RNAs were also removed. As a result, 2886 and 5100 genes were used for gene

488 panel selection at level 1 and 2, respectively. For platform effects estimation, we

489 filtered out cells from cell types that only show up in either scRNA-seq data or

490 MERFISH data.

491

492 ***Codeluppi dataset***

493 scRNA-seq data was downloaded from GEO under accession code GSE60361.

494 Annotation data was downloaded from http://linnarssonlab.org/cortex. osmFISH

495 and corresponding annotation data was downloaded from

496 http://linnarssonlab.org/osmFISH/.

497

498 For scRNA-seq data, cell labels in row 9 of the annotation were used as level 1 cell

499 type annotation, and row 11 were used as level 2 cell type annotation. However, the

500 level 1 cell type annotation is too broad (only 5 major cell types). Therefore, we

501 regenerated level 1 cell type annotation by merging similar cell types at level 2

502 following descriptions from the original study. Specifically, in generating data for

503 gene panel selection at level 1, "S1PyrDL", "S1PyrL23", "S1PyrL4", "S1PyrL5",

504 "S1PyrL5a", "S1PyrL6", S1PyrL6b", "ClauPyr" were merged into "S1_Excitatory".
505 “CA1Pyr1", "CA1Pyr2", "CA1PyrInt", "CA2Pyr2", "SubPyr” were merged into
506 "Hippocampus_Excitatory". 16 subclasses of interneurons (“Int1” to “Int16”) were
507 merged into "Interneuron". "Astro1" and "Astro2" were merged into "Astrocyte".
508 "Mgl1" and "Mgl2" were merged into "Microglia". "Pvm1" and "Pvm2" were merged
509 into "Pvm". Six subpopulations of oligodendrocytes ("Oligo1" to "Oligo6") were
510 merged into "Oligodendorcyte". "Vend1" and "Vend2" were merged into
511 "Endothelial". To make cell type labels consistent between scRNA-seq and osmFISH,
512 "Peric" was changed to "Pericyte". "Choroid" was changed to "Ventricle". "Epend"
513 was changed to "Ependymal".
514
515 To generate the data for platform effect estimation, cell type labels were modified
516 slightly differently to reflect the correspondence between cell types in scRNA-seq
517 and osmFISH as shown in Fig. 2C and Fig. 2D of the original study). Specifically,
518 three CA1 subclasses (“CA1Pyr1”, “CA1Pyr2”, “CA1PyrInt”) were merged into
519 “Hippocampus_Excitatory”. 16 subclasses of interneurons (“Int1” to “Int16”) were
520 merged into “Interneuron”. Two subclasses of microglia (“Mgl1” and “Mgl2”) were
521 merged into “Microglia”. Two subclasses of perivascular macrophages (“Pvm1” and
522 “Pvm2”) were merged into “Pvm”. Subclasses of S1 pyramidal cells were also
523 merged: “S1PyrL4” and “S1PyrL5a” were merged into “S1_Excitatory_L45a”,
524 “S1PyrL5” and “S1PyrL6b” were merged into “S1_Excitatory_L56b”. In addition, to
525 make the cell type labels consistent between scRNA-seq and osmFISH , we changed
526 “Astro1” and “Astro2” to “Astrocyte1” and “Astrocyte2”, respectively. We changed

527 “Oligo6” to “Oligo_Mature”, “Oligo5” to “Oligo_MF”, “Oligo1”, to “Oligo_COP”, “Vend1”
528 to “Endothelial1”, “Vend2” to “Endothelial2”, “Peric” to “Pericyte”, “Choroid” to
529 “Ventricle”, “Epend” to “Ependymal”, “S1PyrL23” to “S1_Excitatory_L23”, and
530 “S1PyrL6” to “S1_Excitatory_L6”. Cell types with fewer than 50 cells were removed.
531
532 For osmFISH data, we first filtered out invalid cells based on the “Valid” column of
533 the annotation data. Then, similar to scRNA-seq data, we modified cell type labels
534 according to Fig. 2C and Fig.2D in the original study, which shows correspondence
535 between cell types in scRNA-seq and osmFISH. Specifically, "Astrocyte Gfap" was
536 changed to “Astrocyte1”. “Astrocyte Mfge8” was changed to “Astrocyte2”.
537 “Hippocampus” was changed to “Hippocampus_Excitatory”. “pyramidal L4” was
538 changed to “S1_Excitatory_L45a”. “Pyramidal L5” was changed to
539 “S1_Excitatory_L56b”. “Pyramidal L6” was changed to “S1_Excitatory_L6”.
540 “Perivascular Macrophages” was changed to “Pvm”. "Oligodendrocyte COP" was
541 changed to “Oligo_COP”. “"Oligodendrocyte Mature"” was changed to
542 “Oligo_Mature”. “Oligodendrocyte MF” was changed to “Oligo_MF”. “Endothelial 1”
543 was changed to “Endothelial1”, and “Endothelial” was changed to “Endothelial2”.
544 “Pericytes” was changed to “Pericyte”. “Vascular Smooth Muscle” was changed to
545 “Vsmc”, “C. Plexus” was changed to “Ventricle”. "Pyramidal L2-3" and "Pyramidal L2-
546 3 L5" were merged into “S1_Excitatory_L23”. "Inhibitory Cnr1", "Inhibitory CP",
547 "Inhibitory Crhbp", "Inhibitory IC", "Inhibitory Kcnip2", "Inhibitory Pthlh", and
548 "Inhibitory Vip" were merged into “Interneuron”.
549

550 scRNA-seq data was normalized using the count_normalize function in the scran
551 package. Similar to the Moffit dataset, the raw and normalized scRNA-seq were
552 further filtered before gene panel selection using the same filters.
553 6123 and 9052 genes were used for gene panel selection at level 1 and 2,
554 respectively. For platform effect estimation, the subset of the raw scRNA-seq and
555 osmFISH data with cells from overlapping cell types were used.
556
557 ***Zhang dataset***
558 Raw and normalized scRNA-seq data from kidney were obtained from [61]. They
559 were further filtered before gene panel selection using the same filters. 2920 and
560 3796 genes were used for gene panel selection at level 1 and 2, respectively.
561 The DARTFISH data is unpublished. We annotated the cells in the DARTFISH data
562 manually using curated marker genes (**Table 2**) at subclass level (third column).
563 For platform effect estimation, the subset of the raw scRNA-seq and DARTFISH data
564 with cells from overlapping cell types were used.
565
566 **Platform effects estimation using a Bayesian model**
567 To estimate the platform effects between scRNA-seq and targeted spatial
568 transcriptomics, we assume the observed number of molecules y_{ij} in the spatial
569 transcriptomics data for gene i in cell j follows a zero-inflated negative bimonial
570 (ZINB) distribution with:
571
572 $$y_{ij} \sim ZINB(\mu_{ij}, \theta_{ij}, \pi) \qquad (1)$$

where π is the zero inflation parameter which is assumed to be constant across genes and cells. μ_{ij} is the mean parameter determined by a global intercept α, true expression level of gene i in cell j denoted as λ_{ij}, and the cell depth (total number of molecules) of cell j from spatial transcriptomics data as CD_j^{SP}:

$$\ln(\mu_{ij}) = \alpha + \ln(\lambda_{ij}) + \ln(CD_j^{SP}) \qquad (2)$$

To account for platform effects, we assume the true expression level λ_{ij} is a random variable defined by:

$$\text{logit}(\lambda_{ij}) = \gamma_i \times \sqrt{x_{ij}} + c_i \qquad (3)$$

where γ_i is a gene specific coefficient representing multiplicative platform effects, and c_i is a gene specific intercept representing additive platform effects. x_{ij} represents the relative expression of gene i in cell j calculated from scRNA-seq data:

$$x_{ij} = \frac{c_{ij}}{\sum_{i=1}^{N} c_{ij}}$$

where c_{ij} is the number of count for gene i in cell j from the scRNA-seq data, and N is the totol number of genes. When fitting the Bayesian model, in order to match measurement between scRNA-seq data and targeted spatial transcriptomics data,

we used cell type average relative expression to replace individual cell level relative expression:

$$x_{ij,j\subset k} = x_{ik} = \frac{\sum_{j=1}^{M_k} c_{ij}}{\sum_{i=1}^{N}\sum_{j=1}^{M_k} c_{ij}}$$

where M_k is the number of cells in cell type k.

For the dispersion parameter θ_{ij} of the ZINB distribution, we assume it is also dependent on λ_{ij}:

$$\ln(\theta_{ij}) = \beta + \lambda_{ij} \qquad (4)$$

where β is the intercept.

A Beta prior distribution is assumed for π. For α, β, and c_i, we assume they follow normal distribution. γ_i is assumed to follow a log-normal distribution:

$$\pi \sim \text{Beta}(1, 1)$$

$$\alpha \sim \text{Normal}(0, \sigma_\alpha)$$

$$\beta \sim \text{Normal}(0, \sigma_\beta)$$

$$c_i \sim \text{Normal}(\mu_c, \sigma_c)$$

$$\gamma_i \sim \text{LogNormal}(\mu_\gamma, \sigma_\gamma)$$

610

611 where the hyperparameters are assumed to follow Cauchy and half Cauchy

612 distribution:

613

$$\mu_c, \mu_\gamma \sim \text{Cauchy}(0, 5)$$

614 $$\sigma_\alpha, \sigma_\beta, \sigma_c, \sigma_\gamma \sim \text{HalfCauchy}(0, 5)$$

615

616 scRNA-seq and targeted spatial transcriptomics data from overlapping genes and

617 overlapping cell types were used as input. Additional filters were applied on the

618 MERFISH data to reduce the totol number of cells for more efficient estimation.

619 Specifically, cells with cell depth lower than 100 were filtered out. Cell types with

620 fewer than 1000 cells were filtered out. Then we subsampled each cell type to keep

621 at most 1000 cells for each cell type.

622

623 Variational inference in Stan was used for model fitting. The ultimate goal is to

624 estimate the posterior distribution of σ_c and σ_γ, which represents the level of

625 additive and multiplicative platform effects.

626

627 **Simulation of spatial transcriptomics measurements from scRNA-seq data**

628 **with platform effects**

629 After fitting the Bayesian model, we used it to simulate spatial transcriptomics

630 measurements from scRNA-seq data. Specifically, $\alpha, \beta, \pi, \mu_c, \sigma_c, \mu_\gamma, \sigma_\gamma$ were

631 randomly sampled from their estimated posterior distribution. c_i, and γ_i were

632 randomly sampled from their corresponding normal and log normal distribution for
633 each new gene that is not observed in the data used to fit the Bayesian model. If a
634 gene is already seen during fitting the Bayesian model, we can either use the
635 empirical c_i, and γ_i estimated during model fitting or randomly generate it from the
636 corresponding normal and log norml distribution. CD_j^{SP} was randomly sampled
637 from empirical cell depth distribution from observed targeted spatial
638 transcriptomics data. x_{ij} was calculated from scRNA-seq data. It can be cell type
639 average as we used in model fitting but it can also be calculated within each
640 individual cell. In our study, the latter was used when simulating spatial
641 transcriptomics measurements to maintain the cell level heterogenity in scRNA-seq
642 data. Finally, the generated values were plugged into equations (1), (2), (3), and (4)
643 to generate spatial transcriptomics measurements.
644
645 **Simulation of spatial transcriptomics measurements from scRNA-seq data**
646 **without platform effects (naïve simulation)**
647 In additional to the simulation method above, we provided a naïve way to simulate
648 spatial transcriptomics measurements without platform effects as control. During
649 the simulation without platform effects, cell depth of simulated spatial
650 transcriptomics cell were randomly sampled from the empirical cell depth
651 distribution of observed targeted spatial transcriptomics data. Of note, the empirical
652 cell depth distribution was adjusted proportionally based on the ratio between
653 relative expression of new genes for simulation and relative expression of
654 overlapping genes used in fitting the Bayesian model. If new genes for simulation

655 have higher overall relative expression, the simulated cell depth will also be higher
656 accordingly. After having the simulated cell depth for each cell, the number of
657 molecules for each gene within each cell was sampled from a multinomial
658 distribution with size equal to the simulated cell depth and probability equal to each
659 gene's relative expression in that cell. At the end, genes were randomly selected
660 given the probe failure rate. Then, their simulated molecule count were set to 0 to
661 reflect probe failure.
662
663 **Genetic algorithm for gene panel selection**
664 We used genetic algorithm as the framework for gene panel selection. In genetic
665 algorithm, the fittest individual is selected from a population of candidates through
666 iteratively updating the population based on individuals' fitness. In this study, each
667 individual is one candidate gene panel. We set the gene panel size to 200 genes.
668 Each population contains 200 individuals.
669
670 The first step of genetic algorithm is to initialize a population of candidate gene
671 panels. The genes can be either randomly selected from all candidate genes or
672 selected based on their differential expression between cell types. In this study, we
673 took a hybrid approach. 95% of the 200 gene panels were initiated randomly from
674 all candidate genes to maintain population diversity. The rest 5% were initialized
675 using DEGs for each cell type. DE analysis was performed using Pagoda2. Genes
676 with AUC greater than 0.7 were considered significant.
677

678 After having an initial population, the second step is to evaluate the fitness of each
679 candidate gene panel in the population. Here we defined fitness as the average
680 classification accuracy over 5 cross validations. Classification was performed on
681 simulated spatial transcriptomics measurements from scRNA-seq data. Cell type
682 annotation from scRNA-seq data was used as ground truth. To improve the
683 efficiency, scRNA-seq data was subsampled to reduce the number of cells for large
684 cell types and resampled to increase the number of cells for small cell types.
685 Specifically, for level 1 cell type annotation, cell type size was capped at 1500 cells.
686 The lower bound was set as 1000 cells for Moffit dataset and 500 for Zhang and
687 Codeluppi dataset. For level 2 cell type annotation, 250 and 500 were used as the
688 cell type size range for Moffit dataset. The range for Zhang and Codeluppi dataset
689 was 300 and 900. We provided two classifiers, random forest and naïve Bayes. In
690 this study we used naïve Bayes due to its fast speed and relatively similar level of
691 accuracy compared to random forest. The accuracy was calculated based on the
692 classification confusion matrix. For flat classification, we calculated accuracy using
693 the original confusion matrix. For hierarchical classification, we calculated a
694 weighted confusion matrix by element-wise multiplying the original confusion
695 matrix with the weighted penalty matrix. Then the accuracy was calculated using
696 the weighted confusion matrix.
697
698 The third step is selection and mutation. The selection strategy we used is
699 tournaments. Specifically, randomly selected candidate gene panels face each other
700 1 vs. 1. The one with a higher fitness value was used as parent. In addition,

701 candidate gene panels with higher fitness values were more likely to be selected in
702 the tournaments. After having the parent gene panels, uniform crossover was
703 performed to generate the offspring gene panels. Duplicated genes after uniform
704 crossover were replaced by randomly sampled genes in the parent candidate gene
705 panels but not in the offspring gene panel. Mutation was then performed to maintain
706 gene diversity and prevent premature convergence. We set the mutation rate to 1%.
707
708 Finally, the same process was repeated for the offspring population. The candidate
709 gene panel with the highest fitness value for one iteration was considered as the
710 optimal gene panel. If the iteration after it has a candidate gene panel with higher
711 fitness value, the optimal panel will be replaced by this new candidate gene panel.
712 Otherwise, the optimal panel will stay the same. This way, we guarantee that the
713 final optimal gene panel is the best one through all the iterations. The iterative
714 process will end either when it reaches a given number of iterations, or the accuracy
715 doesn't improve more than a threshold for a given number of iterations. Both can be
716 specified by the user. In our study, we ran all the optimizations for at least 500
717 iterations to ensure convergence although in all cases the saturation point came a
718 lot earlier.
719
720 Two additional features are provided to further customize the optimization. First, if
721 there is a list of genes that must be included, usually canonical marker genes based
722 on previous knowledge, it can be provided, and the genes will be kept in each
723 population as well as the final optimal gene panel. In addition, for the second

724 feature, a gene weight can be provided and used to guide optimization. A gene with
725 higher weight will be (1) more likely to be selected during crossover, (2) less likely
726 to be mutated if it is already in the population, (3) and more likely to be introduced
727 into the population through mutation if it is not in the current population.
728
729 **Hierarchical classification using cell type hierarchy**
730 During genetic algorithm optimization, a weighted penalty matrix can be provided
731 to assign partial credit or extra penalty to classification between certain cell types.
732 The weighted penalty matrix is a square matrix with each row and each column
733 representing one cell type. This is the same with the confusion matrix. For each
734 value p_{ij} $(i \neq j)$ in the weighted penalty matrix, if $p_{ij} > 1$, an extra penalty is given
735 to misclassifying cells from cell type j to cell type i. If $p_{ij} < 1$, a partial credit is given
736 to misclassifying cells from cell type j to cell type i. $p_{ij} = 1$ means no penalty or
737 partial credit. The weighted penalty matrix was incorporated to the confusion
738 matrix by element-wise multiplication to provide a weighted confusion matrix. The
739 accuracy of the weighted confusion matrix was used to evaluate the fitness of
740 candidate gene panels.
741
742 Essentially, the weighted penalty matrix can be constructed arbitrarily by user's
743 preference. In this study, we constructed the weighted penalty matrix from cell type
744 hierarchy. First, pairwise distance between cell types was calculated. Specifically,
745 average expression profile of each cell type was calculated using normalized count
746 by taking average expression of all cells in each cell type. Top 1000 genes with

747 highest standard deviation were used to calculate pairwise Pearson correlation
748 coefficient. One minus the pairwise Pearson correlation coefficient was used as
749 pairwise distance between cell types. Second, the pairwise distance matrix was
750 normalized by the largest distance so the values range from 0 to 1. Third, the
751 pairwise distance matrix was then adjusted based on cell type hierarchy.
752 Specifically, a level of cell type annotation was selected as reference. For cell types
753 below the reference level that are from the same cell type at the reference level, the
754 pairwise distance (between 0 and 1) between them was kept unchanged to reflect
755 partial credit to wrong classifications among them. For cell types below the
756 reference level that are from different cell types at the reference level, the pairwise
757 distance between them was changed to a user defined value where 1 means no extra
758 penalty and greater than 1 means extra penalty. In this study, we used 1 for no extra
759 penalty and level 1 cell type annotation was used as reference. Finally, the diagonal
760 value was changed to 1 to reflect no extra credit to correct classifications. This
761 weighted penalty matrix was used for hierarchical classification in our study.
762
763 **Calculating the percentage of across broad cell type mistakes over all mistakes**
764 We performed 5 optimizations with flat classification and hierarchical classification
765 for all three datasets, respectively. Average confusion matrix over 5 optimizations
766 for each data was calculated. After that, for each cell type, we counted the total
767 number of misclassifications and among all the misclassifications, what proportion
768 of them misclassifies cells to cell types at level 2 that don't belong to the same cell
769 type at level 1.

770

771 **Gene panel selection using RankCorr and scGeneFit**

772 The same scRNA-seq data from the three datasets after filtering were used as input
773 for marker gene selection. For RankCorr, raw scRNA-seq data before normalization
774 was used as suggested. The lamb parameter was tuned to make sure the output
775 marker gene list has 200 genes. For scGeneFit, normalized scRNA-seq data was used
776 by following the examples on the GitHub page. Panel size was set to 200.

777

778 **Evaluation of optimized gene panel**

779 To evaluate optimized gene panels, we first simulated spatial transcriptomics
780 measurements with platform effects based on the gene panel's expression profile in
781 scRNA-seq data. Then this simulated spatial transcriptomics data was split into
782 training and testing data. The training data contains cells from a subset of cell types
783 whose cell type labels are known. This was used as the partial annotation of the
784 simulated spatial transcriptomics data. The testing data contains cells from all cell
785 types, which is considered as part of the simulated spatial transcriptomics data that
786 hasn't been annotated yet. We varied the number of cell types in the training data
787 from zero to all the cell types to reflect different levels of partial annotation. When
788 there was no partial annotation, scRNA-seq data was used as the final training data
789 for classifier training. When there was partial annotation, information in the partial
790 annotation was included in the final training data. The final training data was then
791 subsampled using the same cell type size used in the genetic algorithm. After that, a
792 classifier (naïve Bayes or random forest) was trained using the final training data.

793 Finally, the classifier was applied on the testing data for cell type classification
794 evaluation. Since the testing data was simulated from scRNA-seq data, the cell type
795 labels in scRNA-seq data were used as ground truth to calculate the classification
796 accuracy, which was used as the metric to evaluate a gene panel. At each level of
797 partial annotation, we repeated the same calculation 10 times. To separately
798 evaluate the impact of platform effects on gene panel selection and cell type
799 classification, within the same framework described here, we designed two different
800 strategies to evaluate a gene panel by varying whether platform effect distortions
801 that can be learned from partial annotation examples are used to produce more
802 realistic training data for cell type classification.
803
804 ***Evaluation with platform effect re-estimation***
805 This evaluation strategy was designed to focus on the performance of the optimized
806 gene panels, and not on the differences in the cell type classification (evaluation)
807 stage. In this strategy, partial annotation was first used to estimate gene specific
808 platform effects using the Bayesian model. We then used these estimated gene
809 specific platform effects to simulate an updated spatial transcriptomics training data,
810 which will be used for training cell type classifiers for all the methods being
811 evaluated. Only cell types not already available in the partially annotated spatial
812 data were simulated. When partial annotation was available for 5 or fewer cell types,
813 the final training data combined the simulated spatial transcriptomics data with
814 scRNA-seq data. When more than 5 cell types were available, training was
815 performed on the simulated spatial data only. The final training data and testing

816 data were normalized by the total number of transcripts within each cell and scaled
817 by 10000. It was then log transformed after adding 1 pseudocount. This normalized
818 training and testing data were used for classifier training and testing.
819
820 ***Evaluation without platform effect re-estimation***
821 In this evaluation strategy, only gpsFISH is able to make use of the platform effects
822 information in the partial annotation (as described above). All the other methods
823 used the partial annotation according to their own method design. Specifically, for
824 the control which used naïve simulation during gene panel selection, the empirical
825 cell depth distribution of the complete testing data was used to simulate a spatial
826 transcriptomics data. We use the complete instead of partially annotated testing
827 data because naïve simulation does not rely on cell type annotation. This simulated
828 spatial transcriptomics data was used in the same way as described above to get the
829 final training data. For RankCorr, scGeneFit, and the random panel, since the gene
830 selection was solely based on scRNA-seq data, cells in the partial annotation were
831 directly combined with the scRNA-seq data of cell types not already available in the
832 partially annotated spatial data. The combined data were used as the final training
833 data. Same normalization was performed on the final training data and testing data
834 before classifier training and testing.
835
836 **Calculate the number of probes for each gene for the DARTFISH data**

837 During the generation of the DARTFISH data, ppDesigner (http://genome-
838 tech.ucsd.edu/public/Gen2_BSPP/ppDesigner/ppDesigner.php) was used to
839 calculate the number of probes that can be designed to target each gene.
840
841 **Availability of data and materials**
842 gpsFISH's open-source code is maintained and documented on Github at
843 https://github.com/kharchenkolab/gpsFISH and is publicly available under the MIT
844 license.
845 Pre-fitted Bayesian models based on the Zhang, Moffit, and Codeluppi dataset
846 respectively are deposited on Zenodo (DOI:
847 https://doi.org/10.5281/zenodo.6946054).
848
849

850 ## Reference

851 1. Arendt D. The evolution of cell types in animals: emerging principles from
852 molecular studies. Nat Rev Genet. 2008;9:868–82.

853 2. Elmentaite R, Domínguez Conde C, Yang L, Teichmann SA. Single-cell atlases:
854 shared and tissue-specific cell types across human organs. Nat Rev Genet.
855 2022;23:395–410.

856 3. Lindeboom RGH, Regev A, Teichmann SA. Towards a Human Cell Atlas: Taking
857 Notes from the Past. Trends in Genetics. 2021;37:625–30.

858 4. Zeng H. What is a cell type and how to define it? Cell. 2022;185:2739–55.

859 5. Kölsch Y, Hahn J, Sappington A, Stemmer M, Fernandes AM, Helmbrecht TO, et al.
860 Molecular classification of zebrafish retinal ganglion cells links genes to cell types to
861 behavior. Neuron. 2021;109:645-662.e9.

862 6. Osterhout JA, Kapoor V, Eichhorn SW, Vaughn E, Moore JD, Liu D, et al. A preoptic
863 neuronal population controls fever and appetite during sickness. Nature.
864 2022;606:937–44.

865 7. Xu S, Yang H, Menon V, Lemire AL, Wang L, Henry FE, et al. Behavioral state
866 coding by molecularly defined paraventricular hypothalamic cell type ensembles.
867 Science. 2020;370:eabb2494.

868 8. Elmentaite R, Kumasaka N, Roberts K, Fleming A, Dann E, King HW, et al. Cells of
869 the human intestinal tract mapped across space and time. Nature. 2021;597:250–5.

870 9. Armingol E, Officer A, Harismendy O, Lewis NE. Deciphering cell–cell interactions
871 and communication from gene expression. Nat Rev Genet. 2021;22:71–88.

872 10. Chen W-T, Lu A, Craessaerts K, Pavie B, Sala Frigerio C, Corthout N, et al. Spatial
873 Transcriptomics and In Situ Sequencing to Study Alzheimer's Disease. Cell.
874 2020;182:976-991.e19.

875 11. Hwang WL, Jagadeesh KA, Guo JA, Hoffman HI, Yadollahpour P, Reeves JW, et al.
876 Single-nucleus and spatial transcriptome profiling of pancreatic cancer identifies
877 multicellular dynamics associated with neoadjuvant treatment. Nature Genetics
878 [Internet]. 2022; Available from: https://doi.org/10.1038/s41588-022-01134-8

879 12. Darmanis S, Sloan SA, Zhang Y, Enge M, Caneda C, Shuer LM, et al. A survey of
880 human brain transcriptome diversity at the single cell level. Proc Natl Acad Sci USA.
881 2015;112:7285–90.

882 13. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, et al. Highly
883 Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter
884 Droplets. Cell. 2015;161:1202–14.

885 14. Tasic B, Menon V, Nguyen TN, Kim TK, Jarsky T, Yao Z, et al. Adult mouse cortical
886 cell taxonomy revealed by single cell transcriptomics. Nat Neurosci. 2016;19:335–
887 46.

888 15. Ner-Gaon H, Melchior A, Golan N, Ben-Haim Y, Shay T. JingleBells: A Repository
889 of Immune-Related Single-Cell RNA–Sequencing Datasets. JI. 2017;198:3375–9.

890 16. Trapnell C. Defining cell types and states with single-cell genomics. Genome Res.
891 2015;25:1491–8.

892 17. Regev A, Teichmann SA, Lander ES, Amit I, Benoist C, Birney E, et al. The Human
893 Cell Atlas. eLife. 2017;6:e27041.

894 18. Marx V. Method of the Year: spatially resolved transcriptomics. Nat Methods.
895 2021;18:9–14.

896 19. Chen R, Blosser TR, Djekidel MN, Hao J, Bhattacherjee A, Chen W, et al. Decoding
897 molecular and cellular heterogeneity of mouse nucleus accumbens. Nat Neurosci.
898 2021;24:1757–71.

899 20. Fernandez J. Molecular atlas of the adult mouse brain. SCIENCE ADVANCES.
900 2020;14.

901 21. Rao A, Barkley D, França GS, Yanai I. Exploring tissue architecture using spatial
902 transcriptomics. Nature. 2021;596:211–20.

903 22. Zhang M, Eichhorn SW, Zingg B, Yao Z, Cotter K, Zeng H, et al. Spatially resolved
904 cell atlas of the mouse primary motor cortex by MERFISH. Nature. 2021;598:137–
905 43.

906 23. Wang Y, Eddison M, Fleishman G, Weigert M, Xu S, Wang T, et al. EASI-FISH for
907 thick tissue defines lateral hypothalamus spatio-molecular organization. Cell.
908 2021;184:6361-6377.e24.

909 24. Moffitt JR, Bambah-Mukku D, Eichhorn SW, Vaughn E, Shekhar K, Perez JD, et al.
910 Molecular, spatial, and functional single-cell profiling of the hypothalamic preoptic
911 region. Science. 2018;362:eaau5324.

912 25. Chen KH, Boettiger AN, Moffitt JR, Wang S, Zhuang X. Spatially resolved, highly
913 multiplexed RNA profiling in single cells. Science. 2015;348:aaa6090–aaa6090.

914 26. Codeluppi S, Borm LE, Zeisel A, La Manno G, van Lunteren JA, Svensson CI, et al.
915 Spatial organization of the somatosensory cortex revealed by osmFISH. Nat
916 Methods. 2018;15:932–5.

917 27. Shah S, Lubeck E, Zhou W, Cai L. In Situ Transcription Profiling of Single Cells
918 Reveals Spatial Organization of Cells in the Mouse Hippocampus. Neuron.
919 2016;92:342–57.

920 28. Cai M, Zhang K. Spatial mapping of single cells in human cerebral cortex using
921 DARTFISH: A highly multiplexed method for in situ quantification of targeted RNA
922 transcripts. eScholarship, University of California; 2019.

923 29. Wang X, Allen WE, Wright MA, Sylwestrak EL, Samusik N, Vesuna S, et al. Three-
924 dimensional intact-tissue sequencing of single-cell transcriptional states. Science.
925 2018;361:eaat5691.

926 30. Chen X, Sun Y-C, Church GM, Lee JH, Zador AM. Efficient in situ barcode
927 sequencing using padlock probe-based BaristaSeq. Nucleic Acids Research.
928 2018;46:e22–e22.

929 31. Chen X, Sun Y-C, Zhan H, Kebschull JM, Fischer S, Matho K, et al. High-
930 Throughput Mapping of Long-Range Neuronal Projection Using In Situ Sequencing.
931 Cell. 2019;179:772-786.e19.

932 32. Gyllborg D, Langseth CM, Qian X, Choi E, Salas SM, Hilscher MM, et al.
933 Hybridization-based *in situ* sequencing (HybISS) for spatially resolved
934 transcriptomics in human and mouse brain tissue. Nucleic Acids Research.
935 2020;48:e112–e112.

936 33. Chen A, Liao S, Cheng M, Ma K, Wu L, Lai Y, et al. Spatiotemporal transcriptomic
937 atlas of mouse organogenesis using DNA nanoball-patterned arrays. Cell.
938 2022;185:1777-1792.e21.

939 34. Ståhl PL, Salmén F, Vickovic S, Lundmark A, Navarro JF, Magnusson J, et al.
940 Visualization and analysis of gene expression in tissue sections by spatial
941 transcriptomics. Science. 2016;353:78–82.

942 35. Rodriques SG, Stickels RR, Goeva A, Martin CA, Murray E, Vanderburg CR, et al.
943 Slide-seq: A scalable technology for measuring genome-wide expression at high
944 spatial resolution. 2019;6.

945 36. Vickovic S, Eraslan G, Salmén F, Klughammer J, Stenbeck L, Schapiro D, et al.
946 High-definition spatial transcriptomics for in situ tissue profiling. Nat Methods.
947 2019;16:987–90.

948 37. Liu Y, Yang M, Deng Y, Su G, Enninful A, Guo CC, et al. High-Spatial-Resolution
949 Multi-Omics Sequencing via Deterministic Barcoding in Tissue. Cell. 2020;183:1665-
950 1681.e18.

951 38. Cho C-S, Xi J, Si Y, Park S-R, Hsu J-E, Kim M, et al. Microscopic examination of
952 spatial transcriptome using Seq-Scope. Cell. 2021;184:3559-3572.e22.

953 39. Fu X, Sun L, Chen JY, Dong R, Lin Y, Palmiter RD, et al. Continuous Polony Gels for
954 Tissue Mapping with High Resolution and RNA Capture Efficiency. bioRxiv.
955 2021;2021.03.17.435795.

956 40. Subramanian A, Narayan R, Corsello SM, Peck DD, Natoli TE, Lu X, et al. A Next
957 Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. Cell.
958 2017;171:1437-1452.e17.

959 41. Missarova A, Jain J, Butler A, Ghazanfar S, Stuart T, Brusko M, et al. geneBasis: an
960 iterative approach for unsupervised selection of targeted gene panels from scRNA-
961 seq. Genome Biology. 2021;22:333.

962 42. Liang S, Mohanty V, Dou J, Miao Q, Huang Y, Müftüoğlu M, et al. Single-cell
963 manifold-preserving feature selection for detecting rare cell populations. Nat
964 Comput Sci. 2021;1:374–84.

965 43. Dumitrascu B, Villar S, Mixon DG, Engelhardt BE. Optimal marker gene selection
966 for cell type discrimination in single cell analyses. Nat Commun. 2021;12:1186.

967 44. Vargo AHS, Gilbert AC. A rank-based marker selection method for high
968 throughput scRNA-seq data. BMC Bioinformatics. 2020;21:477.

969 45. Aevermann BD, Zhang Y, Novotny M, Keshk M, Bakken TE, Miller JA, et al. A
970 machine learning method for the discovery of minimum marker gene combinations
971 for cell-type identification from single-cell RNA sequencing. Genome Res.
972 2021;gr.275569.121.

973 46. Bakken TE, Hodge RD, Miller JA, Yao Z, Nguyen TN, Aevermann B, et al. Single-
974 nucleus and single-cell transcriptomes compared in matched cortical cell types.
975 Soriano E, editor. PLoS ONE. 2018;13:e0209648.

976 47. Cable DM, Murray E, Zou LS, Goeva A, Macosko EZ, Chen F, et al. Robust
977 decomposition of cell type mixtures in spatial transcriptomics. Nat Biotechnol.
978 2022;40:517–26.

979 48. Okochi Y, Sakaguchi S, Nakae K, Kondo T, Naoki H. Model-based prediction of
980 spatial gene expression via generative linear mapping. Nat Commun. 2021;12:3731.

981 49. Andersson A, Bergenstråhle J, Asp M, Bergenstråhle L, Jurek A, Fernández
982 Navarro J, et al. Single-cell and spatial transcriptomics enables probabilistic
983 inference of cell type topography. Commun Biol. 2020;3:565.

984 50. the FANTOM Consortium, Liang C, Forrest ARR, Wagner GP. The statistical
985 geometry of transcriptome divergence in cell-type evolution and cancer. Nat
986 Commun. 2015;6:6066.

987 51. Pliner HA, Shendure J, Trapnell C. Supervised classification enables rapid
988 annotation of cell atlases. Nat Methods. 2019;16:983–6.

989 52. Tasic B. Single cell transcriptomics in neuroscience: cell classification and
990 beyond. Current Opinion in Neurobiology. 2018;50:242–9.

991 53. Zeng H, Sanes JR. Neuronal cell-type classification: challenges, opportunities and
992 the path forward. Nat Rev Neurosci. 2017;18:530–46.

993 54. Yuste R, Hawrylycz M, Aalling N, Aguilar-Valles A, Arendt D, Armañanzas R, et al.
994 A community-based transcriptomics classification and nomenclature of neocortical
995 cell types. Nat Neurosci. 2020;23:1456–68.

996 55. Bard J, Rhee SY, Ashburner M. An ontology for cell types. Genome Biology.
997 2005;5.

998 56. Bakken T, Cowell L, Aevermann BD, Novotny M, Hodge R, Miller JA, et al. Cell
999 type discovery and representation in the era of high-content single cell phenotyping.
1000 BMC Bioinformatics. 2017;18:559.

1001 57. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene
1002 Ontology: tool for the unification of biology. Nat Genet. 2000;25:25–9.

1003 58. The Gene Ontology Consortium, Carbon S, Douglass E, Good BM, Unni DR, Harris
1004 NL, et al. The Gene Ontology resource: enriching a GOld mine. Nucleic Acids
1005 Research. 2021;49:D325–34.

1006 59. Goldberg DE. Genetic Algorithms in Search, Optimization and Machine Learning.
1007 1st ed. USA: Addison-Wesley Longman Publishing Co., Inc.; 1989.

1008 60. Holland JH. Adaptation in Natural and Artificial Systems: An Introductory
1009 Analysis with Applications to Biology, Control, and Artificial Intelligence [Internet].
1010 The MIT Press; 1992 [cited 2022 Jul 30]. Available from:
1011 https://doi.org/10.7551/mitpress/1090.001.0001

1012 61. Lake BB, Menon R, Winfree S, Hu Q, Ferreira RM, Kalhor K, et al. An atlas of
1013 healthy and injured cell states and niches in the human kidney. bioRxiv.
1014 2021;2021.07.28.454201.

1015

1016

1017

1018

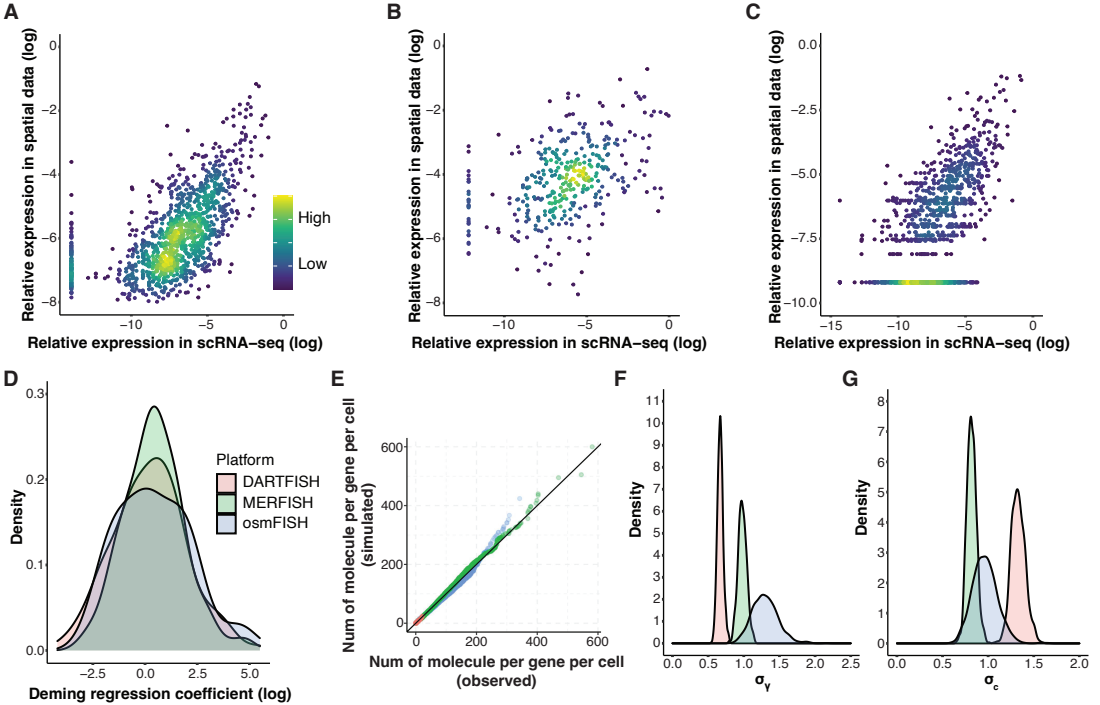


Figure 1 Platform effect between scRNA-seq and targeted spatial transcriptomics technologies

A-C:

Scatter plot showing the log transformed relative expression of genes measured by both scRNA-seq and targeted spatial transcriptomics across three datasets, Moffit (A), Codeluppi (B), and Zhang (C), respectively. A small value is added to avoid negative infinity after log transformation. Each dot represents the relative expression of one gene in one cell type. Denominator for relative expression calculation is from all genes measured by both technologies. Color indicates density of dots. Dots should fall on the diagonal when there is no platform effect.

D:

Density plot of Deming regression coefficient for each dataset. Deming regression is fitted for each gene using relative expression measured by scRNA-seq and spatial transcriptomics data with intercept fixed to 0.

E:

Posterior predictive check of the Bayesian models fitted using each of the three datasets. QQ plot showing the distribution of simulated vs. observed spatial transcriptomics measurements.

F-G:

Density plot showing the estimated posterior distribution of σ_γ (F) and σ_c (G).

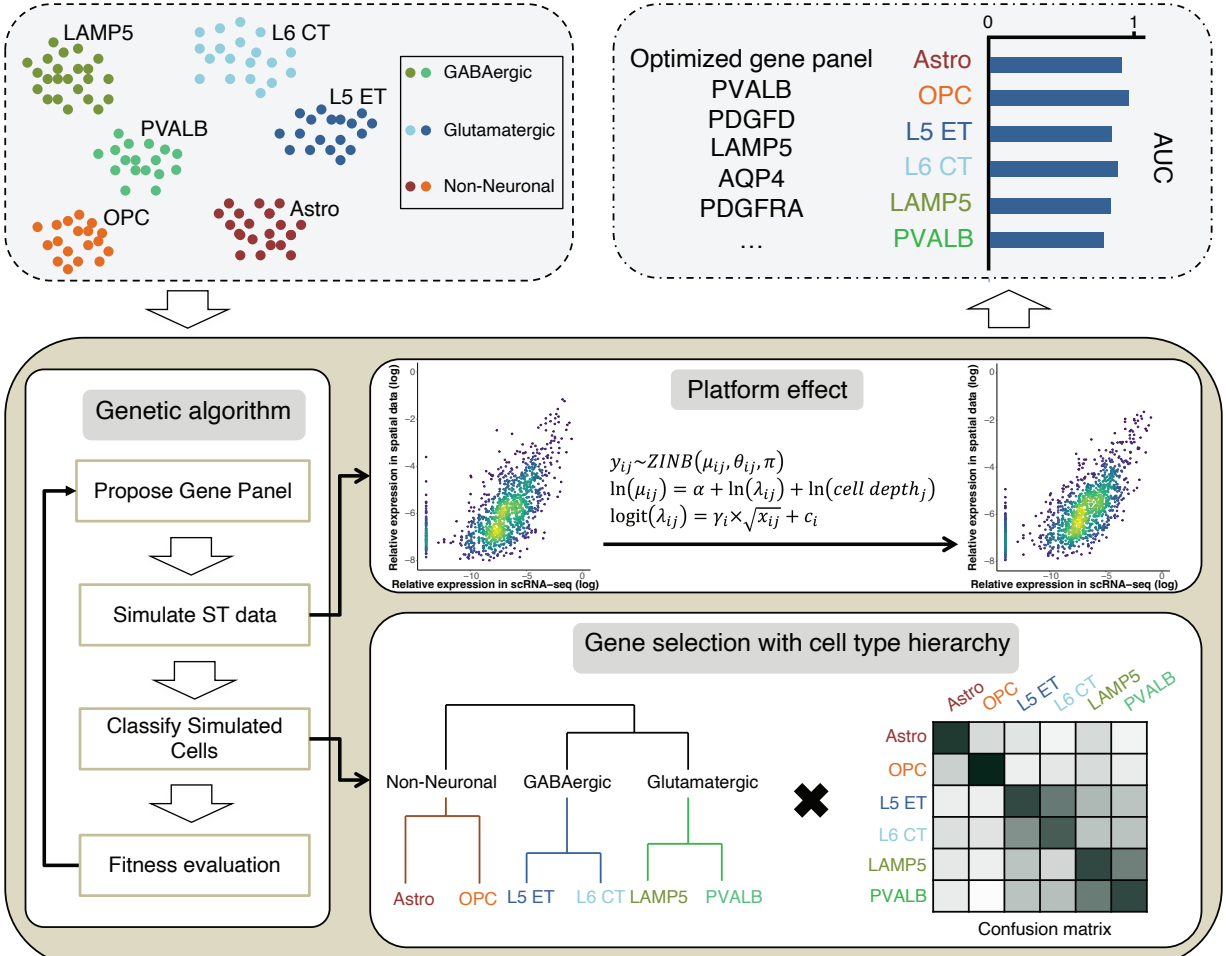


Figure 2: Schematic overview of gpsFISH
Upper left, an scRNA-seq dataset with cell type annotation is used as input. Bottom, a genetic algorithm framework is used for gene panel selection. Platform effects are accounted for using a Bayesian model. Cell type hierarchy can also be incorporated. Upper right, output includes optimized gene panel with classification statistics.

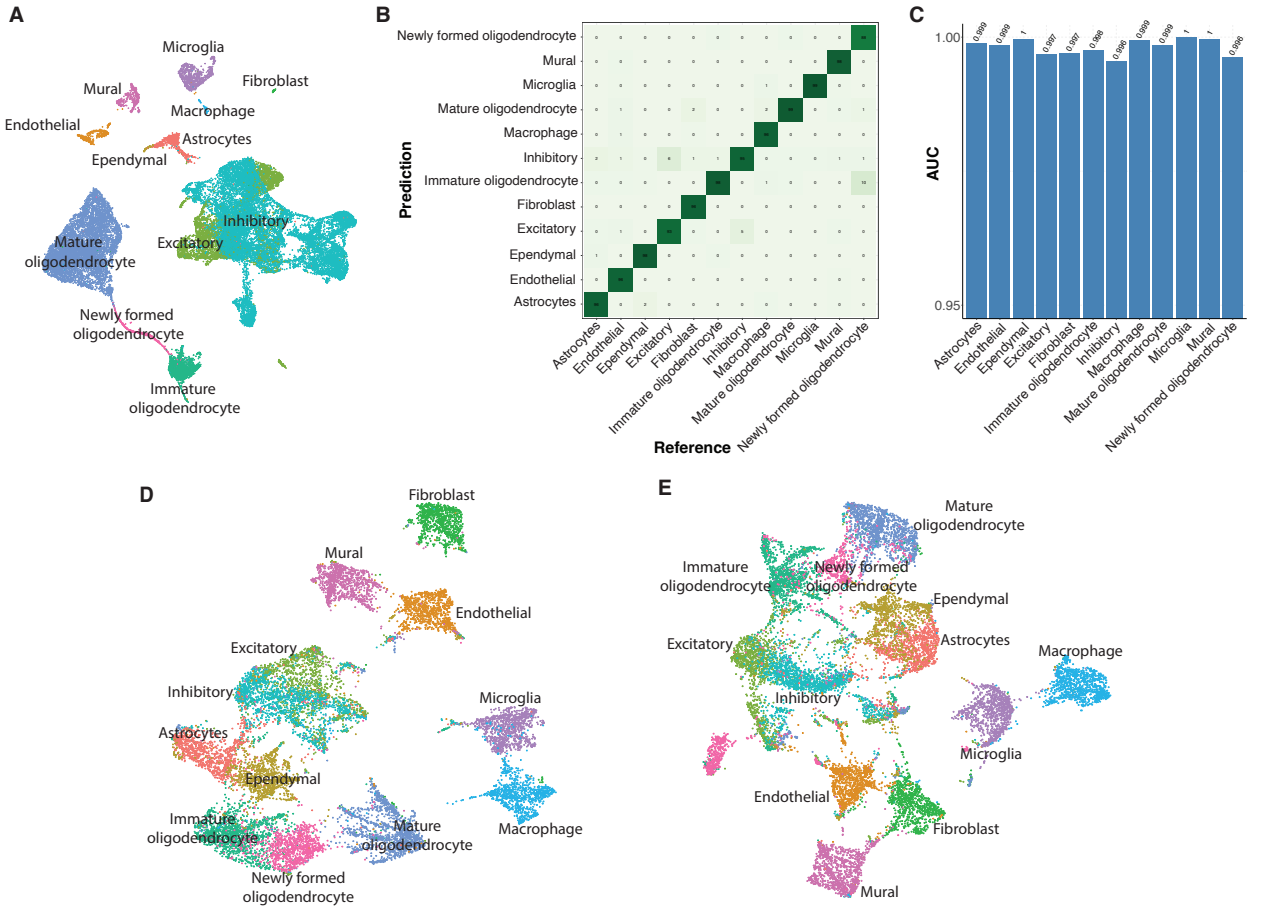


Figure 3: Gene panel selection using gpsFISH

A:

UMAP of cells based on the scRNA-seq data from Moffit dataset at level 1 cell type annotation.

B:

Normalized confusion matrix of the optimized gene panel for Moffit dataset at level 1 cell type annotation.

C:

AUC for each cell type of the same gene panel.

D-E:

UMAP of cells based on simulated spatial transcriptomics measurements of the optimized gene panel selected with (D) and without (E) considering platform effect at level 1 cell type annotation.

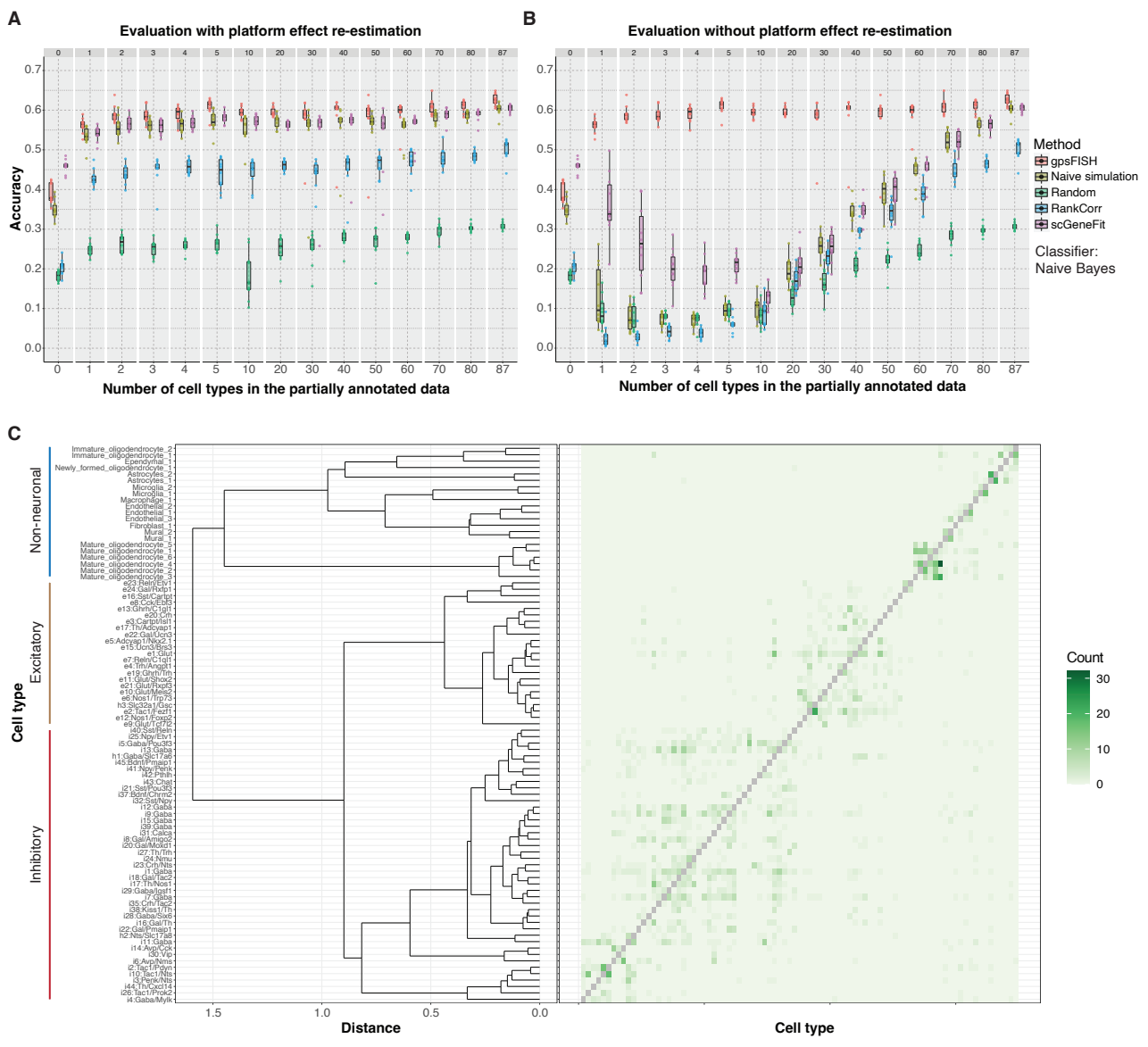


Figure 4: Comparison between gpsFISH and other gene selection methods

A-B:

Box plot showing classification accuracy distribution of gene panels selected by 5 gene panel selection methods at different levels of partial annotation. The result is based on the Moffit dataset using evaluation with (A) and without (B) platform effect re-estimation. Naïve Bayes is used as classifier.

C:

Normalized confusion matrix of the optimized gene panel for Moffit dataset at level 2 cell type annotation with dendrogram showing the cell type hierarchy. Diagonal values of the confusion matrix are removed for better visualization of misclassifications.

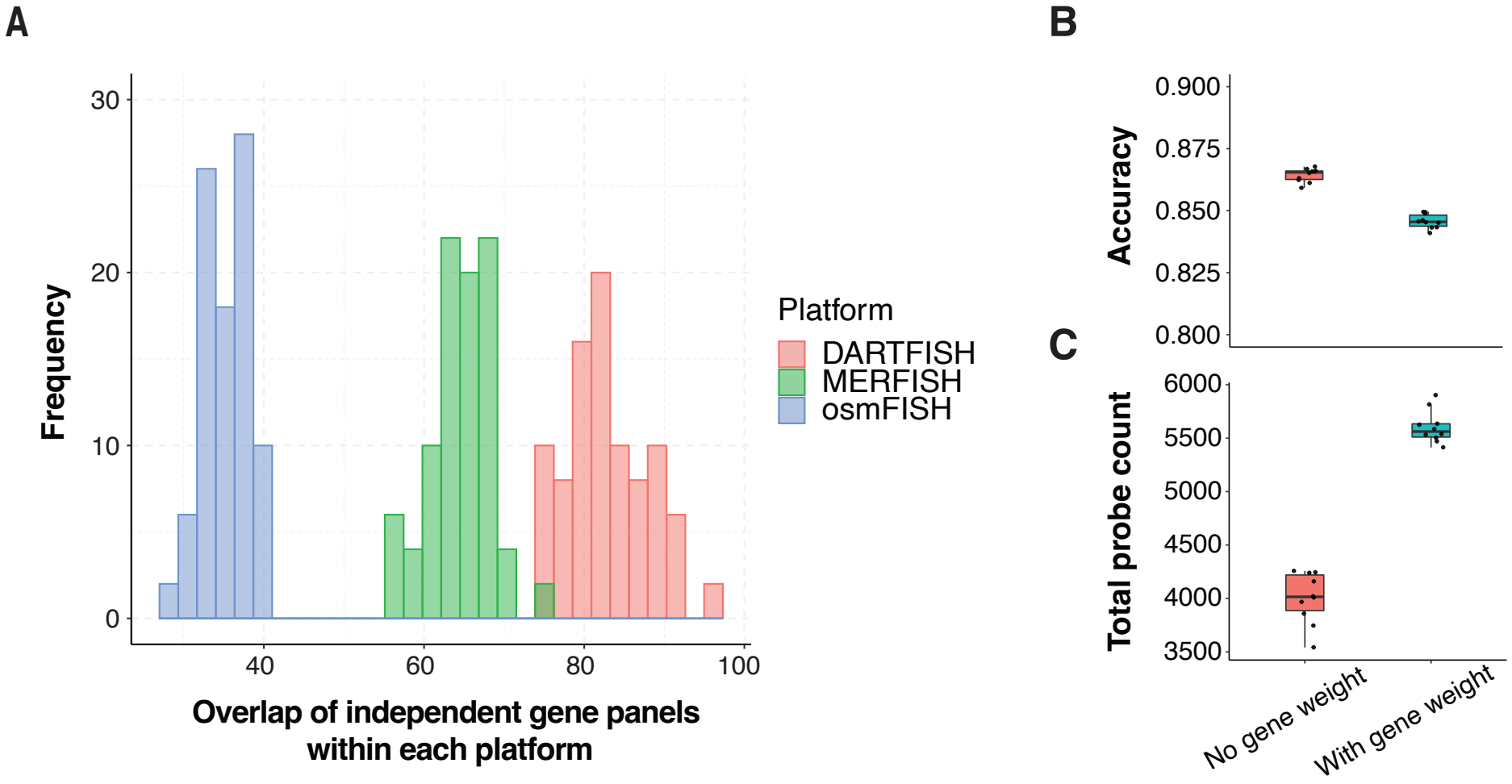


Figure 5: Redundancy in gene space across independent gene panel optimizations enables incorporation of customized preferences

A:

Distribution of overlap of independent gene panels across 10 optimizations within each platform at level 1 cell type annotation.

B:

Accuracy of optimized gene panels without vs. with gene weight across 10 optimizations.

C:

Total number of probes of optimized gene panels without vs. with gene weight across 10 optimizations.

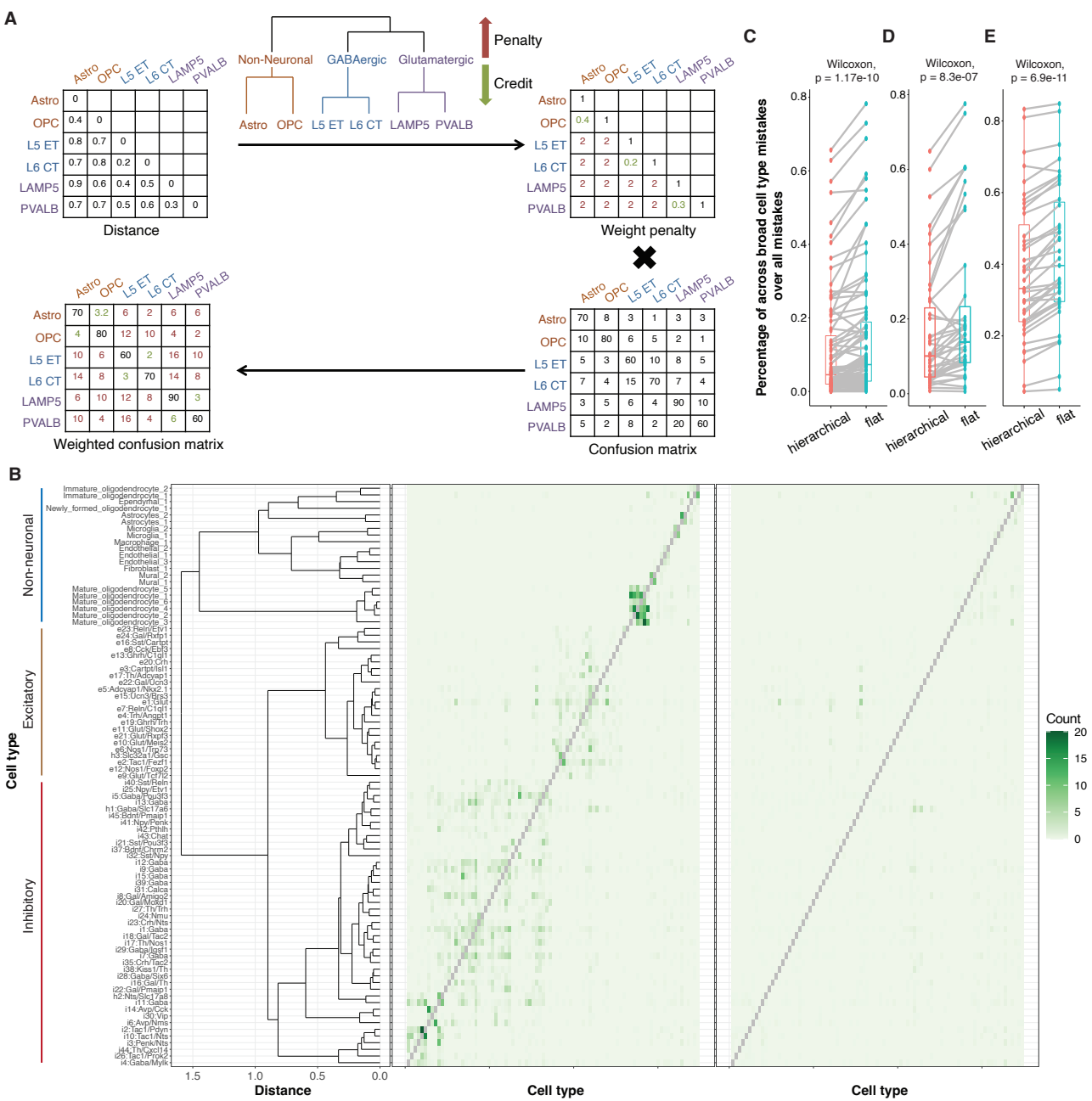


Figure 6: Gene panel selection with cell type hierarchy

A:

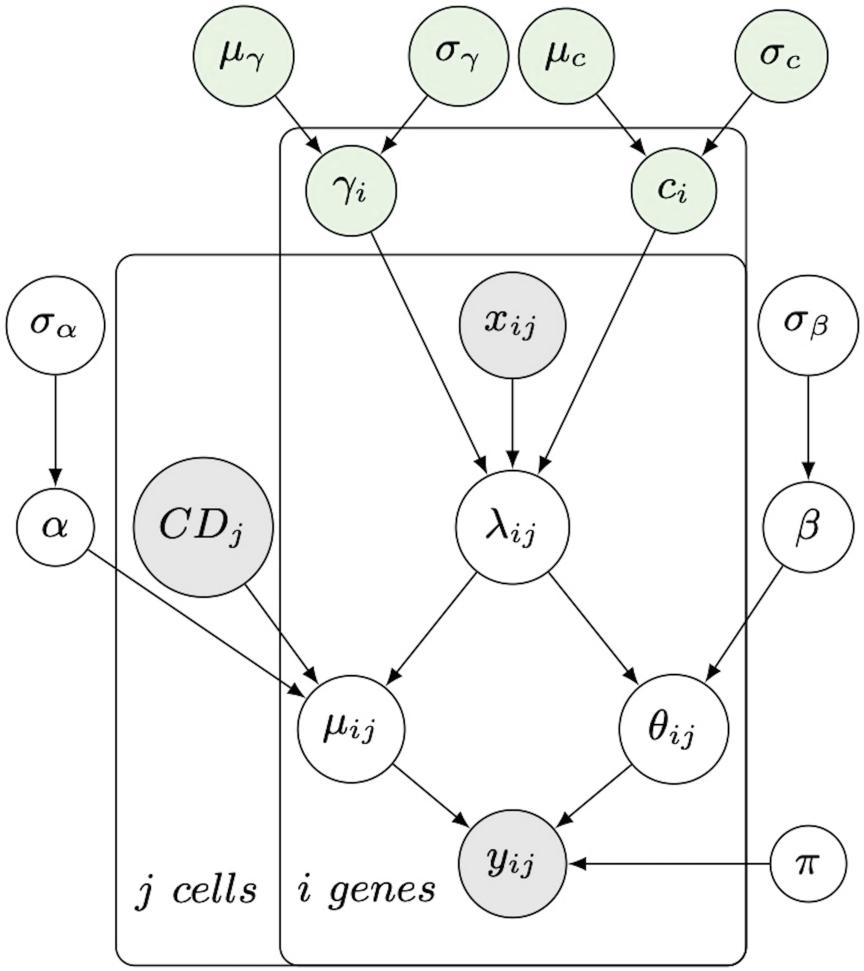
Schematic of hierarchical gene selection using cell type hierarchy. A weighted penalty matrix is constructed using cell type hierarchy information quantified by pairwise distance between cell types. Additional penalty can be specified according to the cell type hierarchy. The weighted penalty matrix is then multiplied element-wise with the original confusion matrix to get the weighted confusion matrix for fitness evaluation.

B:

Original vs. weighted confusion matrix of the same optimized gene panel from Moffit dataset at level 2 cell type annotation with dendrogram showing the cell type hierarchy. Diagonal values of the confusion matrix are removed for better visualization of misclassifications.

C-E:

Percentage of across broad cell type (level 1) misclassifications over all misclassifications for flat vs. hierarchical classification on the Moffit (C), Codeluppi (D), and Zhang (E) dataset. Each dot represents one cell type with dots representing the same cell type connected. Wilcoxon paired test is performed between the percentages from flat vs. hierarchical classification and the p value is shown.

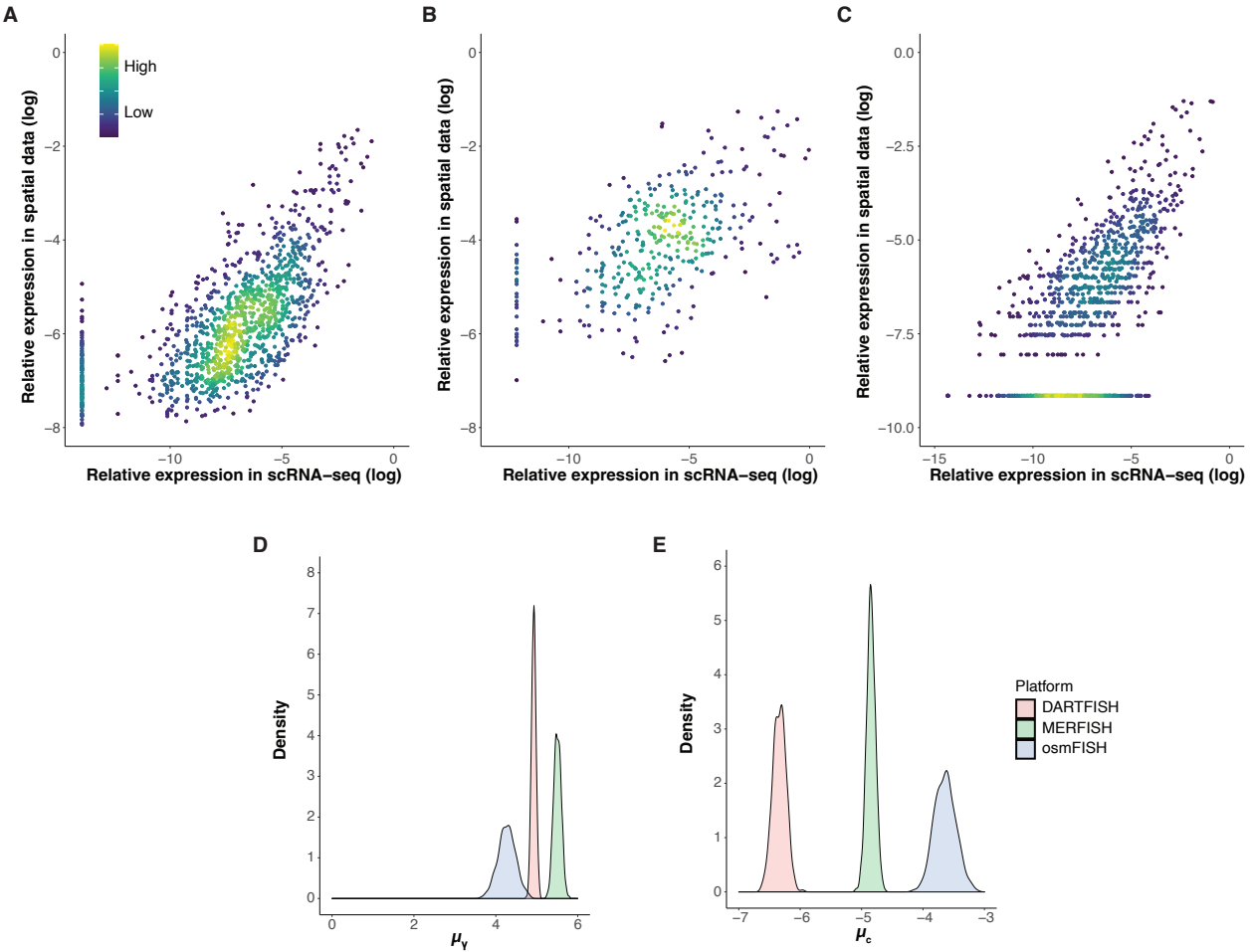


$$y_{ij} \sim \mathrm{ZINB}(\mu_{ij}, \theta_{ij}, \pi)$$
$$\ln(\mu_{ij}) = \alpha + \ln(\lambda_{ij}) + \ln(CD_j)$$
$$\ln(\theta_{ij}) = \beta + \lambda_{ij}$$
$$\mathrm{logit}(\lambda_{ij}) = \gamma_i \times \sqrt{x_{ij}} + c_i$$

$$\alpha \sim \mathcal{N}(0, \sigma_\alpha)$$
$$\beta \sim \mathcal{N}(0, \sigma_\beta)$$
$$\gamma_i \sim \mathcal{LN}(\mu_\gamma, \sigma_\gamma)$$
$$c_i \sim \mathcal{N}(\mu_c, \sigma_c)$$
$$\sigma_\alpha, \sigma_\beta, \sigma_\gamma, \sigma_c \sim \mathrm{HalfCauchy}(0, 5)$$
$$\mu_\gamma, \mu_c \sim \mathrm{Cauchy}(0, 5)$$
$$\pi \sim \mathcal{B}(1, 1)$$

y_{ij}: number of molecules in gene i and cell j from spatial data
x_{ij}: relative expression of gene i in cell j from scRNA-seq data
CD_j: total number of molecules in cell j from spatial data
λ_{ij}: corrected relative expression of gene i in cell j
γ_i and c_i: platform effect magnitude
$\mu_\gamma, \sigma_\gamma, \mu_c, \sigma_c$: platform effect hyperparameters
μ_{ij}: mean expression of gene i in cell j
θ_{ij}: dispersion of gene i in cell j
π: zero inflation parameter

Supplementary figure 1:
Schematic of the Bayesian model for platform effect estimation. Circles in gray represent observed variables. Circles in green correspond to platform effect related variables to be estimated.



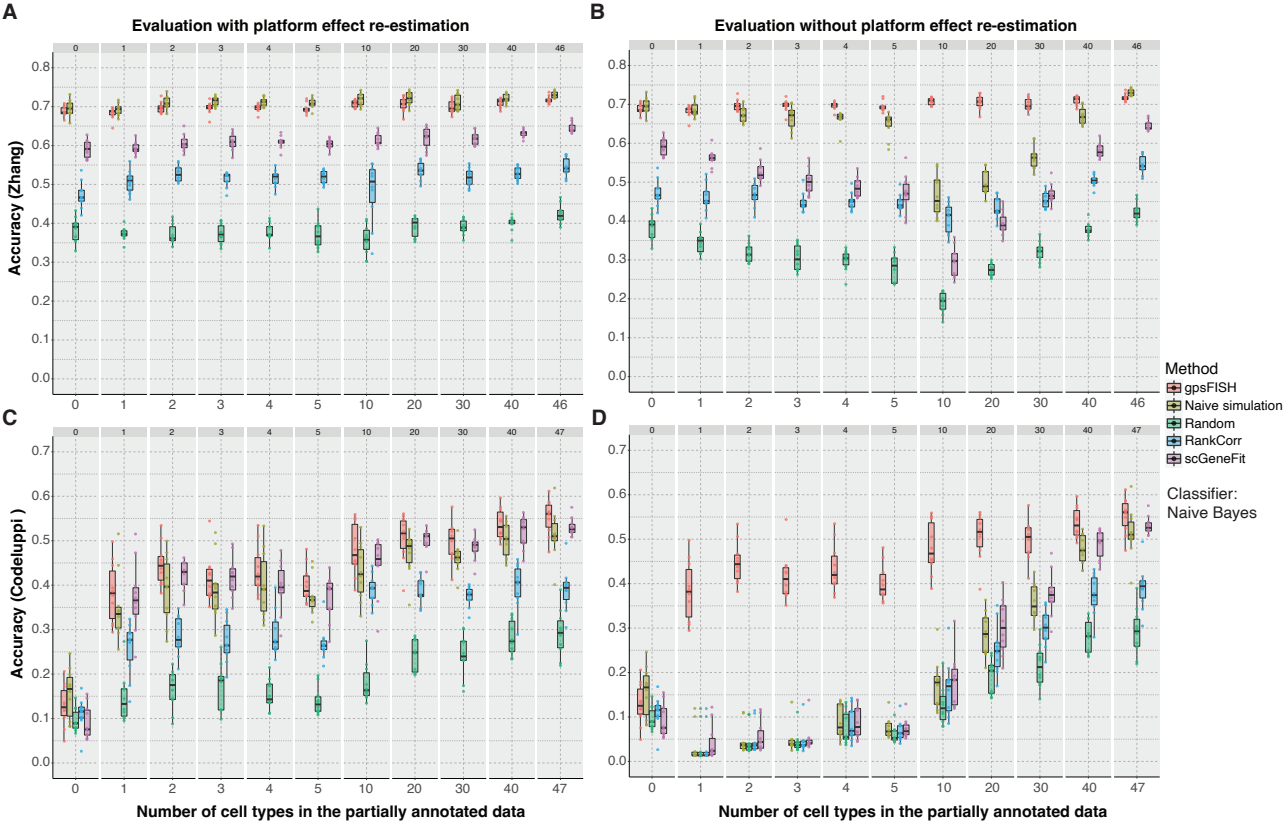
Supplementary figure 2: Bayesian model captures platform effect between scRNA-seq and targeted spatial transcriptomics technologies

A-C:

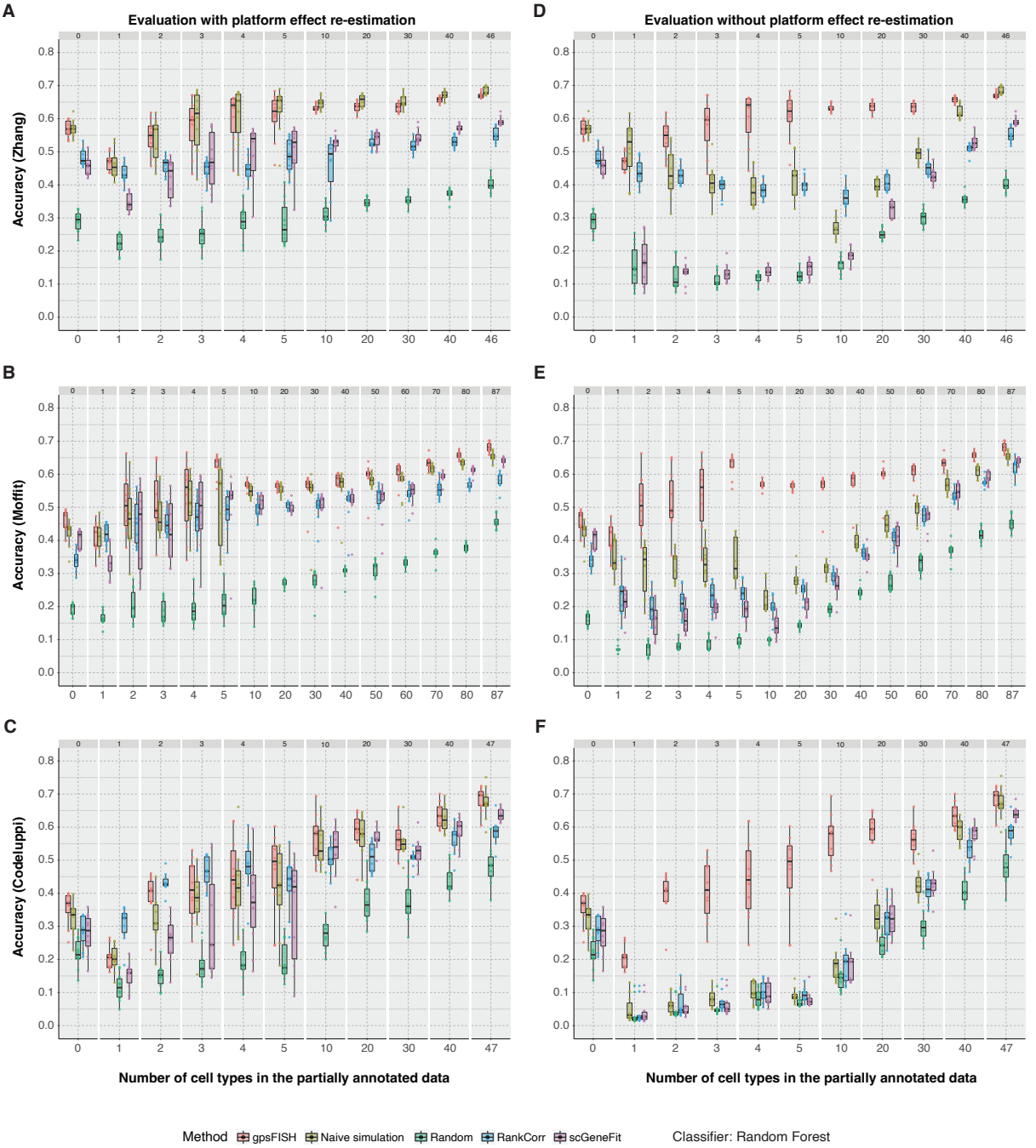
Scatter plot showing the log transformed relative expression of genes measured by scRNA-seq vs. simulated spatial transcriptomics data for Moffit (A), Codeluppi (B), and Zhang (C), respectively.

D-E:

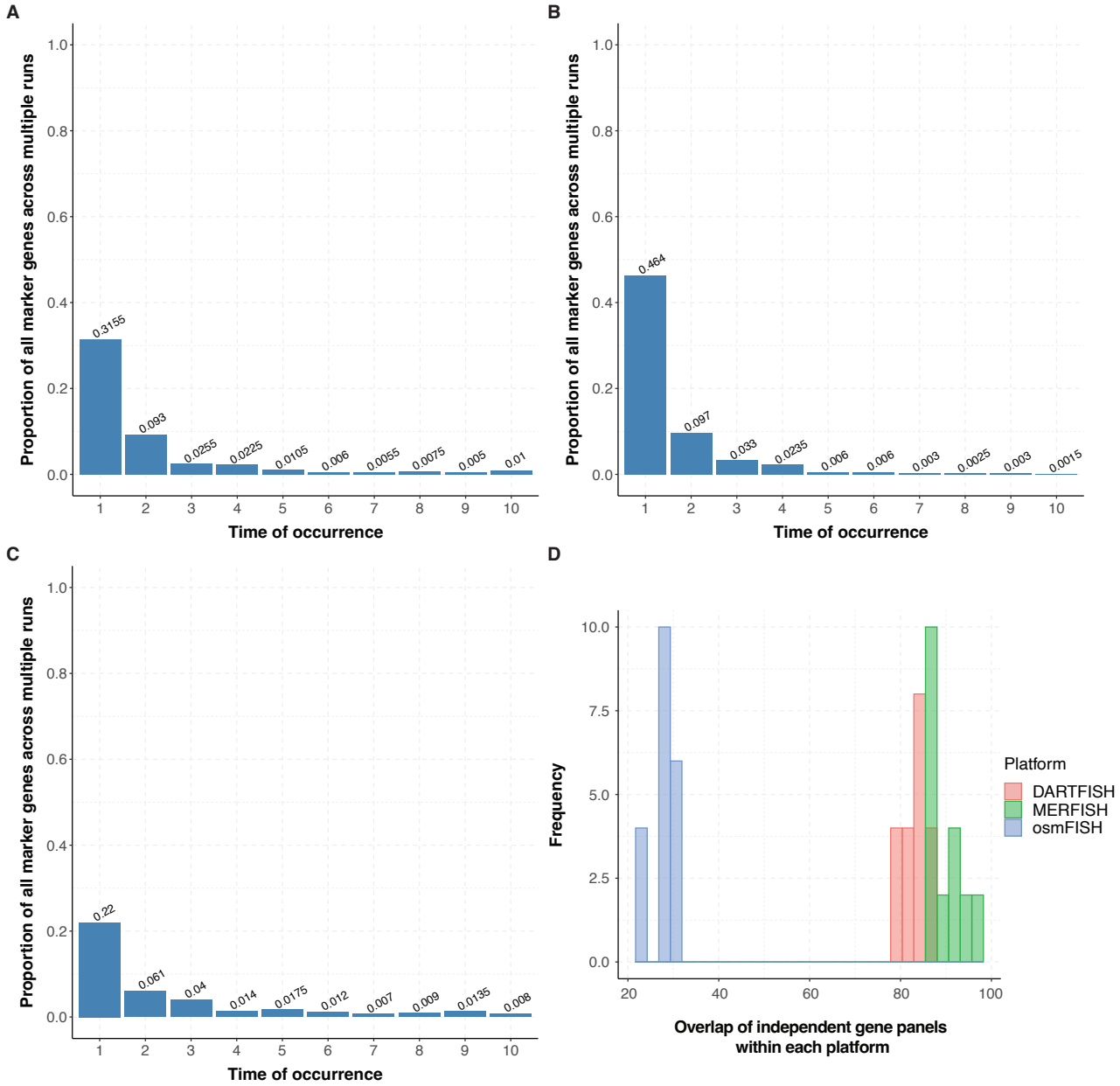
Density plot showing the estimated posterior distribution of μ_γ (D) and μ_c. (E).



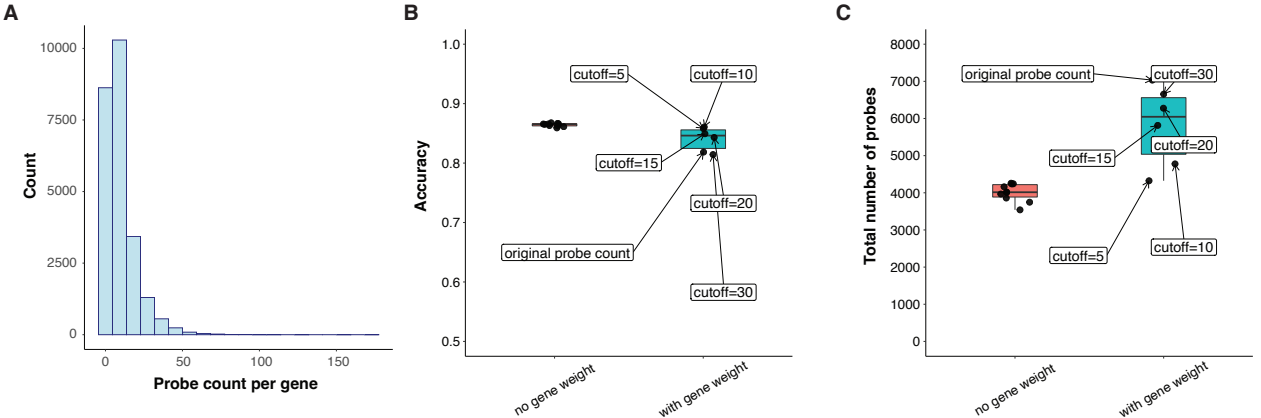
Supplementary figure 3: Comparison between gpsFISH and other gene selection methods on the Zhang and Codeluppi dataset
Box plot showing classification accuracy distribution of gene panels selected by 5 gene panel selection methods at different levels of partial annotation.
(A) Zhang dataset using evaluation with platform effect re-estimation. (B) Zhang dataset using evaluation without platform effect re-estimation.
(C) Codeluppi dataset using evaluation with platform effect re-estimation. (D) Codeluppi dataset using evaluation without platform effect re-estimation.
Naïve Bayes is used as classifier.



Supplementary figure 4: Comparison between gpsFISH and other gene selection methods using random forest as classifier
Box plot showing classification accuracy distribution of gene panels selected by 5 gene panel selection methods at different levels of partial annotation for the three datasets using evaluation with (A-C) and without (D-E) platform effect re-estimation. Random forest is used as classifier.



Supplementary figure 5: High redundancy across optimizations using gpsFISH
A-C:
Bar plot showing among all the genes selected in 10 optimizations, the percentage of them that are included in 1 to 10 optimized panels for Moffit (A), Codeluppi (B), and Zhang (C) dataset at level 1 cell type annotation.
D:
Distribution of overlap of independent gene panels across 10 optimizations within each platform at level 2 cell type annotation.



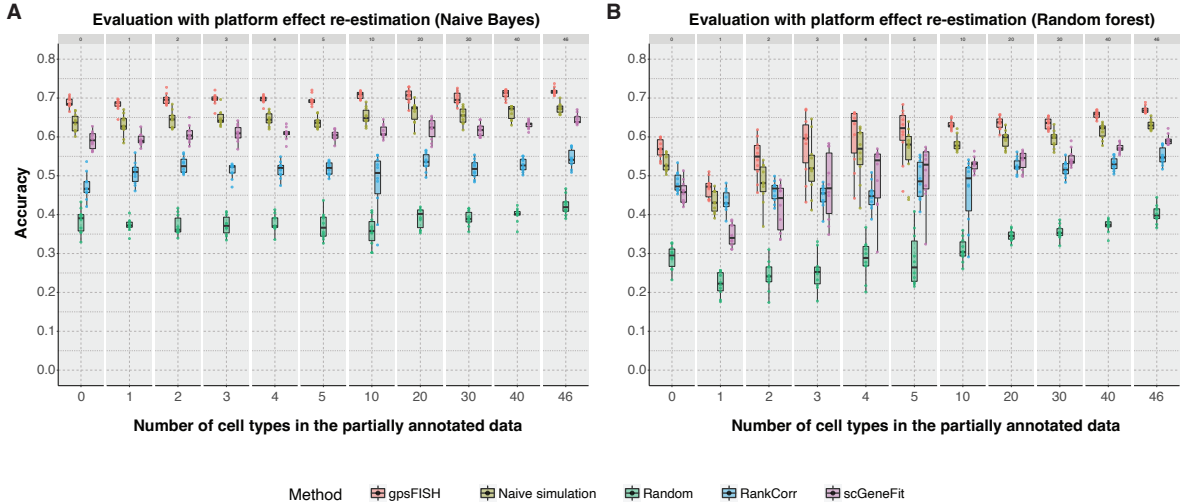
Supplementary figure 6: Weighted gene panel selection based on probe count per gene

A:

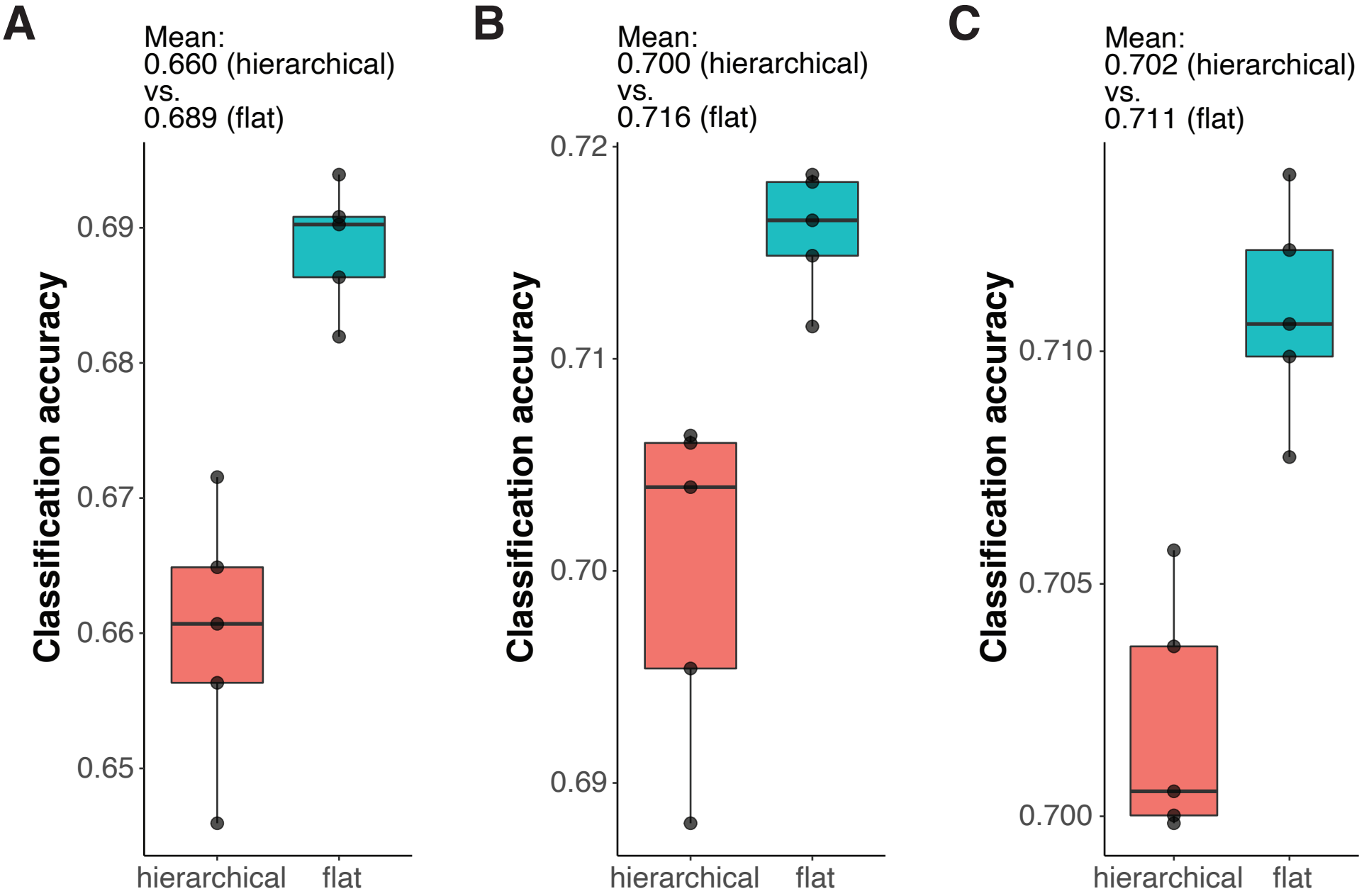
Distribution of probe count per gene for the Zhang dataset.

B-C:

Distribution of accuracy (B) and total number of probes (C) of optimized gene panels from optimization with and without gene weight. Optimization without gene weight is performed 10 times. Optimization with gene weight is performed 6 times, each time with a different probe count cutoff (no cutoff, 5, 10, 15, 20, 30).



Supplementary figure 7: Comparison between gpsFISH and other gene selection methods based on modified naïve simulation model
Box plot showing classification accuracy distribution of gene panels selected by 5 gene panel selection methods at different levels of partial annotation for the Zhang datasets using evaluation with platform effect re-estimation. The gene panel selected by naïve simulation is based on the modified naïve simulation model. Naïve Bayes (A) and random forest (B) is used as classifier, respectively.



Supplementary figure 8: Accuracy of optimized gene panels using flat vs. hierarchical gene selection

A-C:

Distribution of accuracy of optimized gene panels using flat vs. hierarchical gene selection for Moffit (A), Codeluppi (B), and Zhang (C), respectively. Both flat and hierarchical gene selection are performed 5 times.