Evidence of off-target probe binding affecting 10x Genomics Xenium gene panels compromise accuracy of spatial transcriptomic profiling

Caleb Hallinan[1,2], Hyun Joo Ji[1,3], Edmund Tsou[1,2,3], Steven L Salzberg[1,2,3,4], Jean Fan[1,2,3]

[1]Center for Computational Biology, Whiting School of Engineering, Johns Hopkins University, Baltimore, MD 21211, USA

[2]Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD 21218, USA

[3]Department of Computer Science, Johns Hopkins University, Baltimore, MD 21218, USA

[4]Department of Biostatistics, Johns Hopkins University, Baltimore, MD 21205, USA

Correspondence should be addressed to:

Jean Fan (jeanfan@jhu.edu)

Abstract

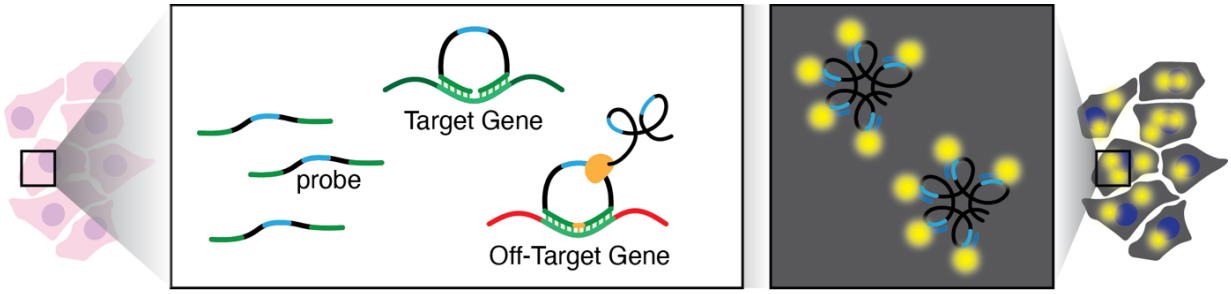
The accuracy of spatial gene expression profiles generated by probe-based *in situ* spatially-resolved transcriptomic technologies depends on the specificity with which probes bind to their intended target gene. Off-target binding, defined as a probe binding to something other than the target gene, can distort a gene's true expression profile, making probe specificity essential for reliable transcriptomics. Here, we investigated off-target binding affecting the 10x Genomics Xenium technology. We developed a software tool, Off-target Probe Tracker (OPT), to identify putative off-target binding via alignment of probe sequences and assessing whether mapped loci corresponded to the intended target gene across multiple reference annotations. Applying OPT to a Xenium human breast gene panel, we identified at least 14 out of the 313 genes in the panel potentially impacted by off-target binding to protein-coding genes. To substantiate our predictions, we leveraged a Xenium breast cancer dataset generated using this gene panel and compared results to orthogonal spatial and single-cell transcriptomic profiles from Visium CytAssist and 3' single-cell RNA-seq derived from the same tumor block. Our findings indicate that for some genes, the expression patterns detected by Xenium demonstrably reflect the aggregate expression of the target and predicted off-target genes based on Visium and single-cell RNA-seq rather than the target gene alone. We further applied OPT to identify potential off-target binding in custom gene panels and integrate tissue-specific RNA-seq data to assess effects. Overall, this work enhances the biological interpretability of spatial transcriptomics data and improves reproducibility in spatial transcriptomics research.

Introduction

Recent advances in high-throughput spatially-resolved transcriptomic profiling technologies have enabled the investigation of gene expression with high spatial resolution within tissues. One such commercially available spatial transcriptomics platform is Xenium from 10x Genomics, a publicly traded company with a market capitalization exceeding $2 billion as of November 2025 (Yahoo Finance 2025). Xenium achieves spatial gene expression profiling at single-cell resolution for targeted genes using a probe-based *in situ* detection approach. 10x Genomics currently offers targeted gene panels with pre-designed probe sets. As of December 2024, over 16,000 Xenium consumable reactions have been sold, with each tissue slide profiled costing approximately $5,000, underscoring the platform's widespread use and high commercial value (10x Genomics 2025b; 2025a).

Briefly, Xenium uses padlock probes that include sequences complementary to the RNA of target genes. Once a padlock probe binds to its target, it is ligated and subsequently amplified via rolling circle amplification (RCA). Fluorescently labeled decoder probes then hybridize to the amplified RCA product, enabling the simultaneous detection and decoding of the optical signature, or codeword, specific to each gene in the panel through successive rounds of fluorescence imaging. When combined with cell-segmentation, this approach allows for spatially-resolved single-cell quantification of gene expression.

56 The accuracy of these gene expression measurements thus relies on the specificity of the probes
57 to bind to their intended target gene. We define off-target binding as when a probe binds to
58 something other than the RNA sequence intended to correspond to the target gene (Figure 1).
59 We note once ligation and RCA occur, the resulting fluorescent signal cannot be distinguished
60 between on-target and off-target binding. As such, off-target binding can distort the
61 quantification of the intended target gene's expression, as the observed expression would
62 represent a combination of the target as well as off-target expression.
63



64

65 **Figure 1. Schematic of potential off-target binding in 10x Genomics Xenium.** In this illustration,
66 the arms of the padlock probes were designed to bind an RNA sequence intended to correspond
67 to a target gene (green). However, these probes exhibit off-target binding and bind to an RNA
68 sequence in a different off-target gene (red). The probe is circularized and subsequently
69 amplified via rolling circle amplification (RCA). Hybridization of fluorescent probes to the RCA
70 product enables the generation a fluorescent signal that is used to quantify RNA expression
71 within cells.
72

73 To predict for such potential off-target binding, we developed Off-target Probe Tracker (OPT), a
74 software tool that aligns probe sequences to an annotated transcriptome with the option to

allow for mismatches that may still permit probe binding. Using OPT, we identify putative off-target probe binding to protein-coding genes affecting at least 14 out of 313 genes in a 10x Genomics Xenium human breast gene panel, compromising the accuracy of their spatial transcriptomic profiles. We substantiate our predictions using data from orthogonal spatial and single-cell gene expression profiling technologies. We further apply OPT to identify potential off-target binding in custom gene panels and integrate tissue-specific RNA-seq data from The Human BioMolecular Atlas Program (HuBMAP) to assess whether such off-target binding could meaningfully affect assayed expression patterns in specific tissues. By facilitating a more rigorous evaluation of probe sequence specificity, tools like OPT can aid in future probe design decisions to help ensure that probes are optimized to minimize off-target binding based on current transcriptome annotations.

Results

OPT predicts potential off-target probe binding

To identify potential off-target binding impacting the 10x Genomics Xenium technology, we require the probe sequences for a specific gene panel of interest, generally represented in a FASTA file. To this end, we first focus on a human breast gene panel used in the Janesick et al. publication, courtesy of 10x Genomics (Methods; Supplementary Table 1). This file includes 2,582 probe sequences that are 40bp in length and designed to target 313 genes, including 33 genes

97 targeted by custom probes, with an average of 8 probe sequences per gene (ranging from 2 to
98 21 probe sequences per gene). We note this panel represents an earlier iteration of and is highly
99 similar to the commercially available pre-designed Xenium v1 Human Breast Gene Expression
100 Panel (Supplementary Note).
101
102 To enable the prediction of potential off-target binding, we developed a software tool called OPT
103 (Methods) that uses nucmer (Marçais et al. 2018) to align probe sequences to various reference
104 transcriptomes, which comprise curated collections of transcript isoforms for all genes in a
105 species. OPT features adjustable parameters for binding strictness (e.g., number of mismatches)
106 and generates a summary file that details all targeted genes along with their potential off-targets
107 based on the sequence alignments. 10x Genomics designed its probe sequences using the
108 GENCODE "basic" annotation (Mudge et al. 2025), so initially we also used the latest GENCODE
109 "basic" annotation (v47) to predict off-target binding for these probes.
110
111 We first sought to predict if a probe has off-target binding based on perfect sequence homology
112 (i.e., if it aligns with 100% identity) with any annotated transcripts other than those that belong
113 to the intended target gene. Of the 2,582 probe sequences in this gene panel, using GENCODE
114 v47, OPT identified 121 probe sequences across 37 genes as having off-target binding based on
115 perfect sequence homology (Table 1). Among the 37 genes with predicted off-target binding, the
116 number of affected probe sequences per gene ranged from 1 to 8. Overall, these off-target
117 probes matched 71 other genes, including 20 protein-coding genes, 31 pseudogenes, 10 long
118 non-coding RNA, 9 transcripts labeled as nonsense-mediated decay, and one microRNA gene.

119

Target gene	Number of probes	Predicted binding genes	Number of probes aligned	Gene types - GENCODE (v47)
ADH1B	8	ADH1B,ADH1A,ADH1C	8,2,1	PC,PC,PC
AKR1C1	9	AKR1C1,AKR1C2,AKR1C3,AKR1C4,AKR1C5P	9,1,1,1,1	PC,PC,PC,PC,PG
APOBEC3A	8	APOBEC3A,APOBEC3B	8,2	PC,PC
APOBEC3B	8	APOBEC3B,APOBEC3D,APOBEC3F,ENSG00000284554	8,2,2,2	PC,PC,PC,PC
AQP1	10	AQP1,ENSG00000250424	10,4	PC,PC
C15orf48	6	C15orf48,MIR147B	6,1	PC,miRNA
C1QA	4	C1QA,ENSG00000289692	4,2	PC,PC
CD68	7	CD68,ENSG00000264772	7,6	PC,lncRNA
CD79B	5	CD79B,ENSG00000285947	5,3	PC,PC
CD8B	16	CD8B,CD8B2	16, 2	PC,PC
CEACAM6	8	CEACAM6,ENSG00000267881	8,1	PC,PC
CLECL1; CLECL1P	3	CLECL1P,ENSG00000293488	3,3	PG,lncRNA
DPT	8	DPT,LINC00970	8,8	PC,lncRNA
EPCAM	8	EPCAM,ENSG00000225356	8,1	PC,PG
HMGA1	7	HMGA1,HMGA1P1,HMGA1P2,HMGA1P3	7,1,1,1	PC,PG,PG,PG
IL2RG	9	IL2RG,ENSG00000285171	9,8	PC,NMD
KRT14	6	KRT14,KRT16P6,ENSG00000290977	6,1,1	PC,PG,lncRNA
KRT8	16	KRT8,KRT8P3,KRT8P2,KRT8P33,KRT8P45,CDK5R2-AS1,ENSG00000304440,KRT8P11,KRT8P17,KRT8P22, KRT8P30,KRT8P32,KRT8P36,KRT8P37,KRT8P42	16,3,2,2,2, 1,1,1,1,1,1, 1,1,1,1	PC,PG,PG,PG,PG,lncRNA,lncRNA ,PG,PG,PG,PG,PG,PG,PG,PG
LDHB	8	LDHB,ENSG000002854	8,5	PC,NMD
LILRA4	8	LILRA4,ENSG00000275210	8,1	PC,lncRNA
MYLK	11	MYLK,MYLKP1	11,1	PC,PG
MYO5B	8	MYO5B,MYO5BP1,MYO5BP2, ENSG00000266997	8,1,1, 4	PC,PG,PG,NMD
PCLAF	8	PCLAF,ENSG00000259316	8,1	PC,NMD
POLR2J3	10	POLR2J3,POLR2J4,POLR2J,ENSG00000270249,POLR2J2,POLR2J2-UPK3BL1,ENSG00000291154	10,4,3,2,2, 2,1	PC,lncRNA,PG,PC,PC,PC,NMD,lncRNA
PTGDS	5	PTGDS,ENSG00000284341	5,3	PC,NMD
SCD	8	SCD,SCDP1	8,2	PC,PG
SERHL2	8	SERHL2,SERHL	8,7	PC,PG
SERPINA3	8	SERPINA3,ENSG00000273259	8,8	PC,NMD
SLAMF1	10	SLAMF1, ENSG00000228863	10, 1	PC, lncRNA
SMS	8	SMS,ENSG00000213080,ENSG00000232389,ENSG00000249779	8,3,1,1	PC,PG,PG,PG
THAP2	13	THAP2,ENSG00000258064	13,2	PC,NMD
TPD52	8	TPD52,ENSG00000276418	8,5	PC,NMD
TPSAB1	2	TPSAB1,TPSB2,TPSD1	2,2,1	PC,PC,PC
TRAF4	9	TRAF4,ENSG00000225869	9,1	PC,PG
TUBB2B	8	TUBB2B,TUBB2BP1	8,1	PC,PG
VOPP1	11	VOPP1,ENSG00000223612	11,1	PC,PG
VWF	8	VWF,VWP1	8,1	PC,PG

120

121 **Table 1. OPT output of genes with predicted off-target binding based on perfect sequence**
122 **homology using GENCODE v47.** This table shows the 37 genes whose probes in the 10x Genomics
123 Xenium v1 Human Breast Gene Expression Panel exhibit predicted off-target probe binding,
124 where each off-target alignment involves a perfect 40bp match to the probe sequence. Although
125 OPT predicted off-target binding of *CCPG1* probe sequences to the *DNAAF1-CCPG1* gene, we

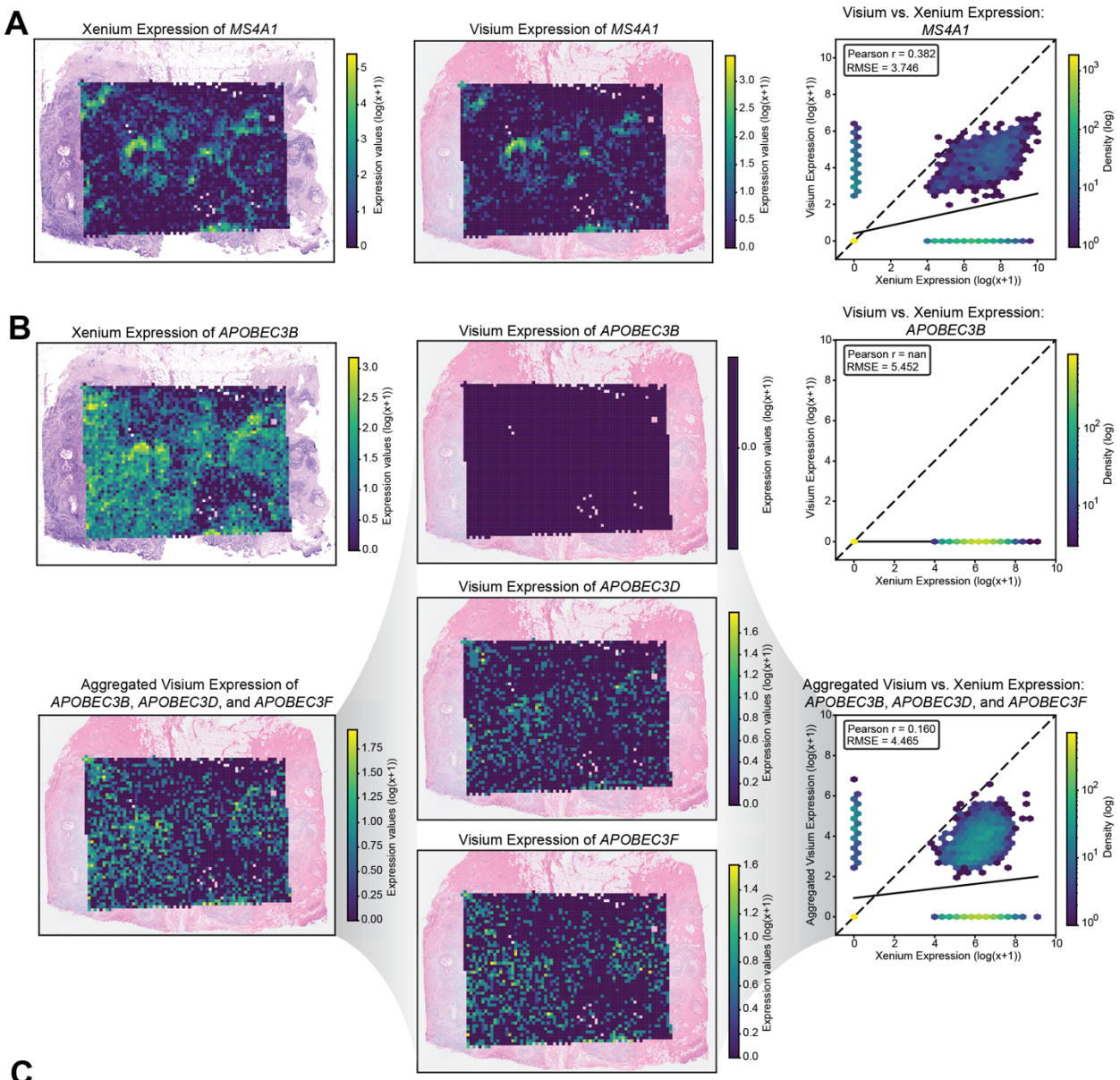
126 manually excluded it from our list because *DNAAF1-CCPG1* is a read-through gene containing
127 portions of both *DNAAF1* and *CCPG1*. The final column shows the gene types, in order, of each
128 of the off-target genes shown in column 3. Abbreviations: PC = protein-coding; PG = pseudogene;
129 NMD = nonsense-mediated decay; lncRNA = long non-coding RNA.
130
131
132 Off-target binding predictions vary across different annotations
133
134 OPT relies on alignments to an annotated transcriptome, which ideally reflects all genes and gene
135 variants stably transcribed in a given species. However, genome annotation is still an active area
136 of research (Varabyou et al. 2023), with discrepancies across annotations in gene counts,
137 isoforms, and many other features. We therefore further used OPT to predict for off-target
138 binding and affected genes using two additional human genome annotation sets, RefSeq (v110),
139 (O'Leary et al. 2016) and CHESS (v3.1.3) (Varabyou et al. 2023), and compared to our previous
140 results from GENCODE (v47) (Mudge et al. 2025) (Methods).
141
142 When considering only perfect sequence homology, while we previously found 37 affected genes
143 using GENCODE, we found 14 when using RefSeq and 23 when using CHESS (Supplementary Table
144 2, 3). Given that RefSeq and CHESS have more transcripts than GENCODE, these discrepancies in
145 off-target binding predictions was not simply an artifact of the difference in transcript set sizes.
146

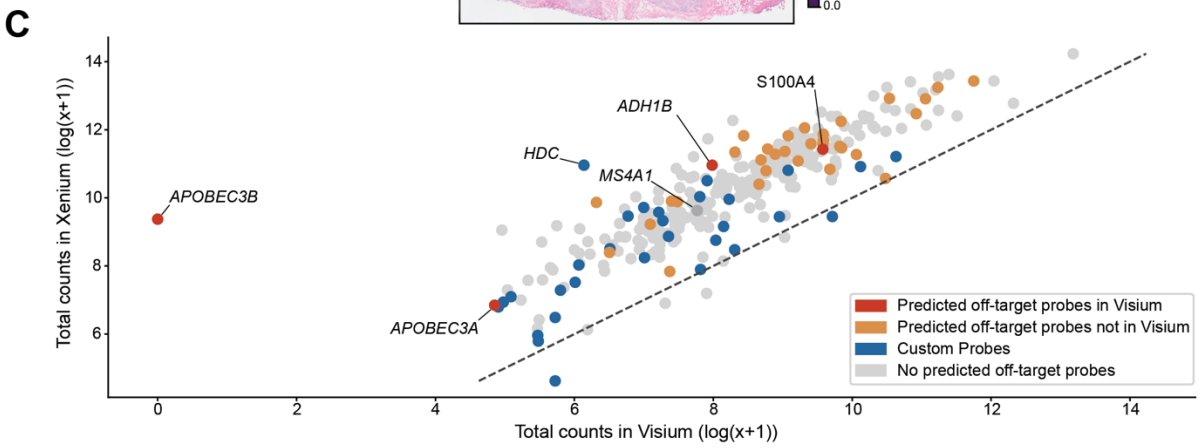
147 While the human annotation databases mostly agree on the number of protein-coding genes in
148 the genome, they remain widely divergent on pseudogenes and lncRNA genes. Therefore, we
149 focused on how the results change when we restrict our analysis to only protein-coding genes.
150 By excluding pseudogenes (which are presumably not expressed), lncRNAs, and other non-
151 protein-coding RNAs when using OPT, the number of affected genes fell to 11 for GENCODE, 10
152 for RefSeq and 9 for CHESS (Supplementary Table 4). Again, these discrepancies reflect
153 annotation differences. For example, the probe sequence
154 (ENSG00000196154|S100A4|ab4e3dc), which was designed to target *S100A4* based on
155 GENCODE annotations, also aligns to *S100A5* in RefSeq (Supplementary Figure 1). We reason that
156 if a probe sequence aligns off-target to a protein-coding gene based on any of these annotations,
157 it could result in off-target binding. We therefore focused further analysis on the union of genes
158 with predicted off-target binding to protein-coding genes across the 3 annotations, resulting in
159 14 genes: *ADH1B, AKR1C1, APOBEC3A, APOBEC3B, AQP1, C1QA, CD79B, CD8B, CEACAM6,*
160 *POLR2J3, S100A4, TOMM7, TPD52,* and *TPSAB1* (Supplementary Figure 2; Supplementary Table
161 5).
162
163
164 <u>Comparison with Visium CytAssist reveals spatial gene expression patterns consistent with off-</u>
165 <u>target binding</u>
166
167 To investigate the potential effects of our predicted off-target binding for this Xenium human
168 breast gene panel in experimental settings, we compared spatial gene expression patterns

169 detected in two previously published spatial transcriptomics datasets from serial sections of the
170 same breast cancer tissue: one section assayed with Xenium using this gene panel, and another
171 assayed using Visium CytAssist, an orthogonal spatial transcriptomics platform (Janesick et al.
172 2023). Briefly, Visium CytAssist is a sequencing-based spatial transcriptomics technology in which
173 RNA is hybridized to spatially barcoded capture spots on a slide, enabling spatial transcriptomic
174 mapping after sequencing. However, while Xenium offers single-cell resolution gene expression
175 quantification, Visium quantifies gene expression within 55μm spots. To enable direct
176 comparison, we first structurally aligned the Xenium and Visium tissue sections using STalign
177 (Clifton et al. 2023), restricting our analysis to overlapping regions since different parts of the
178 tissue were profiled (Supplementary Figure 3A). To improve visual comparability, we aggregated
179 the Xenium gene expression data at the aligned locations to match the Visium spatial resolution
180 and visualized using resolution-matched tiles (Supplementary Figure 3B; Methods).
181
182 Among the 14 genes exhibiting off-target binding to protein-coding genes, 4 were present in the
183 Visium dataset that had at least one corresponding off-target gene also detected in the dataset.
184 For genes with no predicted off-target binding based on perfect sequence homology such as
185 *MS4A1*, we observed a visually similar spatial pattern between the two technologies (Figure 2A),
186 suggesting that spatially aligned groups of cells across the two technologies express this gene at
187 comparable relative magnitudes. This can be quantitively assessed by comparing the pseudo-log
188 gene expression values from both Visium and Xenium at matched spatial locations and
189 computing the RMSE and Pearson correlation in a manner similar to STcompare (Clifton et al.
190 2025). For *MS4A1*, the RMSE is relatively low at 3.746, and the Pearson correlation of 0.382

191 indicates a moderate degree of concordance between the two technologies. However, for genes
192 with predicted off-target binding based on perfect sequence homology such as *APOBEC3B*, we
193 observed a visually dissimilar spatial pattern between the two technologies (Figure 2B).
194 Consistent with this, the RMSE is relatively high at 5.452, and the Pearson correlation is *nan*
195 because *APOBEC3B* is not expressed in the Visium dataset. Importantly, its predicted off-target
196 genes, *APOBEC3D* and *APOBEC3F*, in Visium shows a visually more similar spatial pattern to the
197 Xenium *APOBEC3B*. To better visualize the effect of off-target binding within the Xenium data,
198 we aggregated the expression of each gene along with its predicted off-targets found in the
199 Visium dataset and visually compared across spatial locations. Notably, the spatial pattern of the
200 aggregated expression of *APOBEC3B, APOBEC3D* and *APOBEC3F* in Visium is visually more similar
201 to the spatial pattern of *APOBEC3B* in Xenium. Quantitatively, comparing this aggregated
202 expression results in a decrease in RMSE to 4.465 and a non-*nan* Pearson correlation of 0.160,
203 consistent with the prediction that Xenium *APOBEC3B* probes exhibit off-target binding to
204 *APOBEC3D* and *APOBEC3F*. We further visually confirmed using the Integrative Genomics Viewer
205 (IGV) where two probe sequences intending to bind to *APOBEC3B* were found to perfectly align
206 to sequences in both *APOBEC3D* and *APOBEC3F*, consistent across all annotations evaluated
207 (Supplementary Figure 4).
208
209 Overall, when comparing the total gene expression between the two technologies, we observed
210 a generally strong positive correlation, consistent with the previously published work (Janesick
211 et al. 2023). We do not observe an obvious trend between gene expression magnitude and the
212 presence of predicted off-target probes (Figure 2C), suggesting that off-target binding prediction

alone does not explain the observed higher expression magnitude in Xenium compared to Visium, which may still be attributed to variation in detection efficiency, sequencing depth, and other factors.

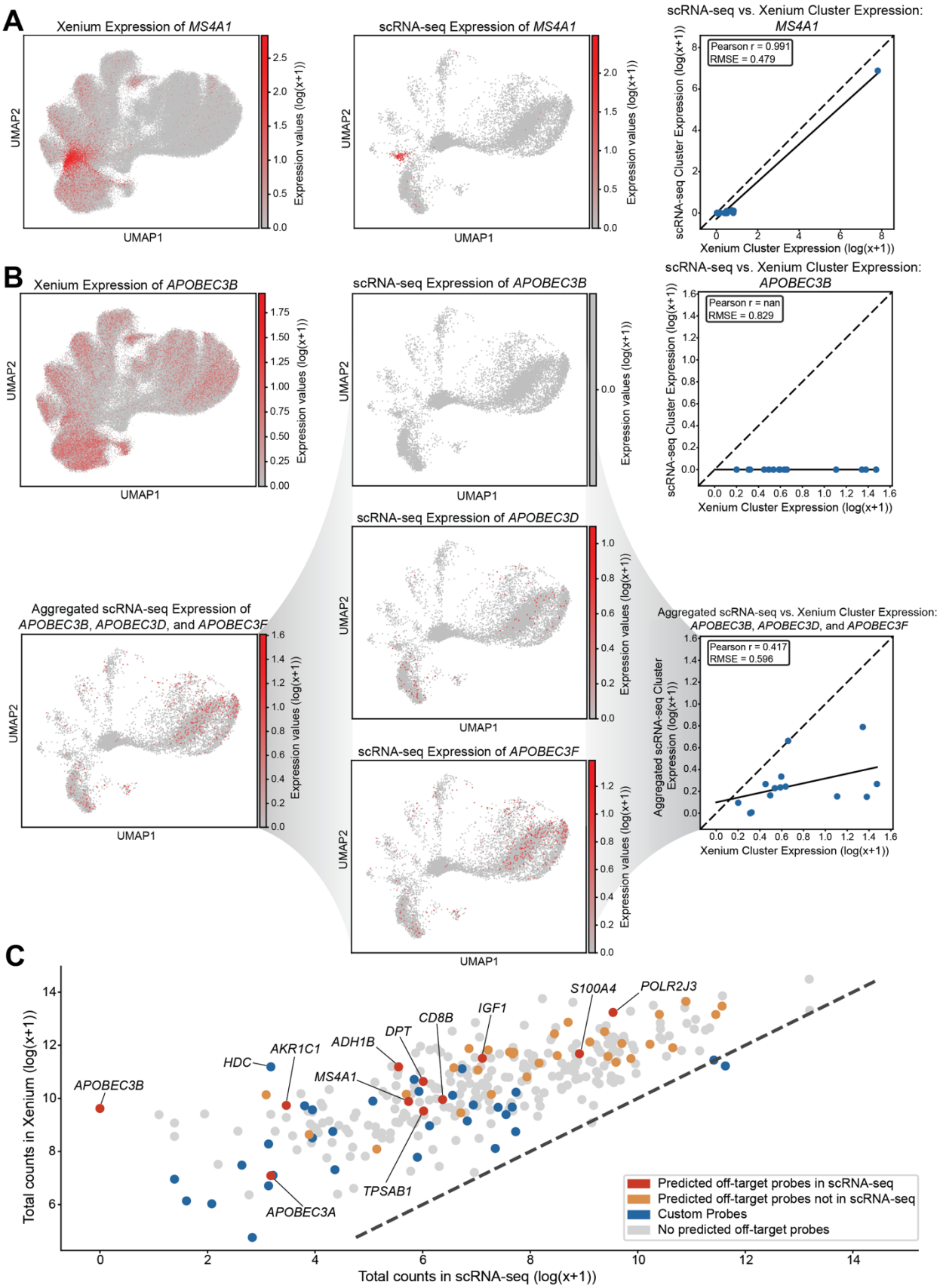

A
Xenium Expression of MS4A1
Expression values (log(x+1))
Visium Expression of MS4A1
Expression values (log(x+1))
Visium vs. Xenium Expression: MS4A1
Pearson r = 0.382
RMSE = 3.746
Visium Expression (log(x+1))
Xenium Expression (log(x+1))
Density (log)
B
Xenium Expression of APOBEC3B
Expression values (log(x+1))
Visium Expression of APOBEC3B
Expression values (log(x+1))
Visium vs. Xenium Expression: APOBEC3B
Pearson r = nan
RMSE = 5.452
Visium Expression (log(x+1))
Xenium Expression (log(x+1))
Density (log)
Visium Expression of APOBEC3D
Expression values (log(x+1))
Aggregated Visium Expression of APOBEC3B, APOBEC3D, and APOBEC3F
Expression values (log(x+1))
Aggregated Visium vs. Xenium Expression: APOBEC3B, APOBEC3D, and APOBEC3F
Pearson r = 0.160
RMSE = 4.465
Aggregated Visium Expression (log(x+1))
Xenium Expression (log(x+1))
Density (log)
Visium Expression of APOBEC3F
Expression values (log(x+1))


C
APOBEC3B
HDC
ADH1B
S100A4
MS4A1
APOBEC3A
Predicted off-target probes in Visium
Predicted off-target probes not in Visium
Custom Probes
No predicted off-target probes
Total counts in Xenium (log(x+1))
Total counts in Visium (log(x+1))

217 **Figure 2. Comparison of spatial gene expression patterns between Xenium and Visium. (A)**
218 Spatial gene expression of *MS4A1* overlaid on the corresponding histological images for Xenium
219 and Visium, accompanied by a density plot comparing Xenium vs. Visium *MS4A1* expression. The
220 dotted line indicates the identity line (X = Y), and the solid line represents the line of best fit. **(B)**
221 Gene expression patterns for *APOBEC3B*: Xenium expression, Visium expression, the aggregated
222 Visium expression combining *APOBEC3B* and its predicted off-target gene's expression
223 *APOBEC3D* and *APOBEC3F*, and Visium expression of *APOBEC3B*'s predicted off-targets
224 *APOBEC3D* and *APOBEC3F*. Two density plots are shown: one comparing Xenium vs. Visium for
225 *APOBEC3B* alone, and one comparing Xenium vs. the aggregated Visium expression of *APOBEC3B*
226 with all off-targets. The dotted line indicates the identity line (X = Y), and the solid line represents
227 the line of best fit. (**C)** Scatterplot of log-transformed total expression counts (with a
228 pseudocount) for 307 genes comparing Visium and Xenium data. The dotted line indicates the
229 identity line (X = Y), and points (genes) are colored by probe information.
230
231
232 <u>Comparison with scRNA-seq reveals single-cell gene expression patterns consistent with off-</u>
233 <u>target binding</u>
234
235 To further investigate the potential effects of our predicted off-target binding for this Xenium
236 human breast gene panel, we compared the detected single-cell gene expression patterns in the
237 same previously published work using Chromium Next GEM Single Cell 3' (Janesick et al. 2023).
238 Briefly, single-cell RNA sequencing (scRNA-seq) with 3' end capture is a technique used to profile

239 gene expression at the single-cell level by profiling the 3' ends of mRNA transcripts with
240 sequencing followed by alignment to a genome or transcriptome for quantification. While this
241 approach provides single-cell resolution gene expression quantification, it lacks spatial
242 information. To enable a single-cell comparison with Xenium, we use Harmony (Korsunsky et al.
243 2019) to remove batch effects and project cells into a shared UMAP embedding (Supplementary
244 Figure 5A-B). We also performed Leiden clustering on the harmonized principal components to
245 quantitively compare cluster expression (Supplementary Figure 5C; Methods).
246
247 Among the 14 genes exhibiting off-target binding to protein-coding genes, 10 were present in
248 the scRNA-seq dataset that had at least one corresponding off-target gene also detected in the
249 dataset. Again, for genes with no predicted off-target binding based on perfect sequence
250 homology such as *MS4A1*, we observed a visually similar gene expression pattern in the
251 harmonized UMAP across both technologies (Figure 3A), suggesting that transcriptionally similar
252 clusters of cells or cell-types across the two technologies express this gene at comparable relative
253 magnitudes. We further quantitatively assessed the data by comparing the pseudo-log gene
254 expression values obtained from the clusters within the clustered harmonized UMAP and
255 computed the RMSE and Pearson correlation. For *MS4A1*, the RMSE is relatively low at 0.479,
256 and the Pearson correlation of 0.991 indicates a strong degree of concordance between the two
257 technologies. Likewise, again, for genes with predicted off-target binding based on perfect
258 sequence homology such as *APOBEC3B*, we observed a visually dissimilar gene expression
259 pattern on the harmonized UMAP (Figure 3B). Consistent with this, the RMSE is relatively high at
260 0.829, and the Pearson correlation is *nan* because *APOBEC3B* is not expressed in the scRNA-seq

261 dataset. Again, its predicted off-target genes, *APOBEC3D* and *APOBEC3F,* in scRNA-seq showed a
262 visually more similar expression pattern in the harmonized UMAP embedding to the Xenium
263 *APOBEC3B*. To better illustrate the impact of the off-target probes, we again aggregated the
264 expression of a gene and its predicted off-target genes present in the scRNA-seq data and visually
265 compared across the harmonized UMAP embedding. The aggregated expression of *APOBEC3B*,
266 *APOBEC3D* and *APOBEC3F* in the scRNA-seq data shows a visually more similar gene expression
267 pattern in the harmonized UMAP embedding to *APOBEC3B* in the Xenium data (Figure 3B).
268 Quantitatively, comparing this aggregated expression results in a decrease in RMSE to 0.596 and
269 a non-nan in Pearson correlation to 0.417, consistent with the prediction that Xenium *APOBEC3B*
270 probes exhibit off-target binding with these paralogs.
271
272 Overall, when comparing total gene expression between the two technologies, we again
273 observed a generally strong positive correlation (Figure 3C), similar to the Visium comparison
274 results and consistent with the previously published work (Janesick et al. 2023).


A
Xenium Expression of MS4A1
scRNA-seq Expression of MS4A1
scRNA-seq vs. Xenium Cluster Expression: MS4A1
Pearson r = 0.991
RMSE = 0.479
B
Xenium Expression of APOBEC3B
scRNA-seq Expression of APOBEC3B
scRNA-seq vs. Xenium Cluster Expression: APOBEC3B
Pearson r = nan
RMSE = 0.829
scRNA-seq Expression of APOBEC3D
Aggregated scRNA-seq Expression of APOBEC3B, APOBEC3D, and APOBEC3F
scRNA-seq Expression of APOBEC3F
Aggregated scRNA-seq vs. Xenium Cluster Expression: APOBEC3B, APOBEC3D, and APOBEC3F
Pearson r = 0.417
RMSE = 0.596
UMAP1
UMAP2
Expression values (log(x+1))
scRNA-seq Cluster Expression (log(x+1))
Xenium Cluster Expression (log(x+1))
Aggregated scRNA-seq Cluster Expression (log(x+1))
C
APOBEC3B
HDC
AKR1C1
ADH1B
MS4A1
DPT
CD8B
IGF1
S100A4
POLR2J3
TPSAB1
APOBEC3A
Predicted off-target probes in scRNA-seq
Predicted off-target probes not in scRNA-seq
Custom Probes
No predicted off-target probes
Total counts in Xenium (log(x+1))
Total counts in scRNA-seq (log(x+1))

275

276 **Figure 3. Comparison of single-cell gene expression patterns between Xenium and scRNA-seq.**
277 **(A)** Harmonized UMAP visualization of *MS4A1* expression for Xenium and scRNA-seq data,
278 accompanied by a scatterplot comparing Xenium vs. scRNA-seq *MS4A1* cluster expression. The
279 dotted line indicates the identity line (X = Y), and the solid line represents the line of best fit. **(B)**
280 Comparison of *APOBEC3B* expression patterns on harmonized UMAP: Xenium expression, scRNA-
281 seq expression, an aggregated scRNA-seq profile combining *APOBEC3B* and its predicted off-
282 target gene's expression *APOBEC3D* and *APOBEC3F*, and scRNA-seq expression of *APOBEC3B*'s
283 predicted off-targets *APOBEC3D* and *APOBEC3F*. Two scatterplots are shown: one comparing
284 Xenium vs. scRNA-seq for *APOBEC3B* cluster expression alone, and one comparing Xenium vs. the
285 aggregated scRNA-seq cluster expression of *APOBEC3B* and it's predicted off-targets. The dotted
286 line indicates the identity line (X = Y), and the solid line represents the line of best fit. **(C)**
287 Scatterplot of log-transformed total expression counts (with a pseudocount) for 313 genes
288 between Visium and scRNA-seq data. The dotted line indicates the identity line (X = Y), and points
289 (genes) are colored by probe information.
290
291
292 <u>OPT results when allowing mismatches at the terminal ends of the probe sequences identifies</u>
293 <u>additional off-target candidates</u>
294
295 Thus far, we have focused on predicting off-target binding based on perfect sequence homology.
296 However, we reason that imperfect sequence matching could still result in off-target binding.
297 Specifically, for Xenium v1, which underlies the Xenium human breast gene panel, padlock probes

298 with two 20 base pair (bp) arms bind complementary mRNA regions, forming a 40bp probe
299 sequence. A ligase then circularizes the padlock probe, favoring specific 2-bp junctions.
300 Importantly, if there is a sequence mismatch, particularly outside the ligation site towards the
301 terminal ends of the probe sequence, hybridization may still occur and result in off-target
302 binding, albeit with reduced hybridization efficiency (Supplementary Figure 6A). We therefore
303 added an option in OPT to allow imperfect alignments at the ends of the probe sequences,
304 specifying the sequence length at either ends, where mismatches, insertions, deletions, or
305 clipping can occur (Methods).
306
307 Allowing for 10bp mismatches on either end of the 40bp probe sequence (i.e., requiring a 20bp
308 match covering the middle of the probe sequence including the ligation site) revealed 18
309 additional genes, including to protein-coding genes, with potential off-target binding when using
310 GENCODE v47 (Supplementary Table 6), among which *ACTG2* was included. Additionally, 10 of
311 the 37 genes previously predicted to be affected by off-target binding based on perfect sequence
312 homology were now predicted to have additional off-target genes, including to protein-coding
313 genes (Supplementary Table 7), among which *TUBB2B* was included. Both genes showed visually
314 dissimilar spatial patterns between Xenium and Visium, accompanied by comparably high RMSE
315 and low Pearson correlation (Supplementary Figure 7A-B). In contrast, the spatial pattern of the
316 aggregate of *TUBB2B* and *ACTG2* with their predicted off-target protein-coding genes (*TUBB2A*
317 and *ACTB*/*ACTA1*/*POTEM* respectively) in Visium more closely resembled the spatial pattern of
318 *TUBB2B* and *ACTG2* observed in Xenium, accompanied by a corresponding decrease in RMSE for
319 both genes and increase in Pearson correlation for *TUBB2B*. Likewise, a similar trend is observed

320 in the scRNA-seq comparison (Supplementary Figure 8A-B). Ultimately, these findings suggest
321 that off-target binding, even with imperfect sequence matching, can contribute to the expression
322 patterns observed in Xenium.
323
324
325 RNA-seq reference atlases suggest off-target binding can variably impact results in Xenium
326 custom probe panels
327
328 Our analyses to this point focused on the Xenium human breast gene panel from Janesick et al.,
329 a predecessor to and thus similar to the commercially available pre-designed Xenium v1 Human
330 Breast Gene Expression Panel (Supplementary Note). We next sought to determine whether such
331 off-target binding could affect custom Xenium gene panels, in which gene selection is specified
332 by the user rather than provided as pre-designed panels by 10x Genomics. To investigate this, we
333 leveraged two custom Xenium gene panels used by the Human BioMolecular Atlas Program
334 (HuBMAP) (Jain et al. 2023): one designed for the placenta and another designed for the kidney,
335 lung, and heart (i.e. multi-organ). Applying OPT to the associated FASTA files using the GENCODE
336 v47 annotation and allowing for 10bp mismatches on either end of the 40bp probes, we found
337 that 49 genes out of the 300 targeted genes in the placenta panel (Supplementary Table 8) and
338 24 genes out of the 300 targeted genes in the multi-organ (kidney, lung, heart) panel had
339 predicted off-target binding (Supplementary Table 9). Of these, 30 of the 49 placenta panel genes
340 and 11 of the 24 multi-organ panel genes had predicted off-targets that were protein-coding
341 genes.

342

343 To assess the potential effects of our predicted off-target binding for these custom Xenium gene
344 panels in experimental settings, we examined the expression of the corresponding predicted off-
345 target genes in matched scRNA-seq or bulk RNA-seq data from the HuBMAP consortium for each
346 of the relevant tissue types. For the placenta custom panel, 34 of the 49 genes with predicted
347 off-target genes were detected with non-zero expression magnitudes in the placenta bulk RNA-
348 seq dataset (Supplementary Figure 9). For the multi-organ panel, of the 21 genes with predicted
349 off-target genes, 13 were detected with non-zero expression magnitudes in the heart, 12 in the
350 kidney, and 13 in the lung (Supplementary Figure 10). Together, these results suggest that off-
351 target binding can impact custom Xenium gene panels by distorting observed Xenium gene
352 expression measurements in a tissue-dependent manner, particularly where the off-target gene
353 is expressed at a higher magnitude compared to the target gene within the tissue of interest.

354

355

356 **Discussion**

357

358 Our study presents evidence of off-target probe binding that may distort gene expression profiles
359 affecting the 10x Genomics Xenium spatial transcriptomics technology. We identified at least 14
360 out of the 313 genes in a Xenium human breast gene panel, which is highly similar to the
361 commercially-available pre-designed Xenium v1 Human Breast Gene Expression Panel
362 (Supplementary Note), that may be affected by off-target probe binding based on sequence
363 similarity, supported by spatial and single-cell comparative analyses using Xenium with serial

364 section datasets from Visium CytAssist and 3' single-cell RNA-seq respectively. We further
365 identified potential off-target probe binding affecting custom Xenium gene panels. To assist in
366 the interpretation of existing probe-based gene expression data as well as future probe design,
367 we provide OPT as a software tool for predicting potential off-target probe binding. For future
368 reference, we have run OPT on all publicly available 10x Genomics pre-designed Xenium gene
369 panels and supply them as a ZIP file (Supplementary Table 10).
370
371 Although we have predicted off-target binding based on sequence alignment, its effect on gene
372 expression quantification may still vary. One reason is that the off-target protein-coding or non-
373 protein-coding gene may not be expressed (Supplementary Figure 11). For example, in the
374 Xenium human breast gene panel, although *ADH1B* probes have predicted off-target binding to
375 *ADH1A* and *ADH1C* based on perfect sequence homology, the sparse expression of *ADH1A* and
376 *ADH1C* in both the Visium and scRNA-seq breast cancer data led to only a minor difference in the
377 aggregated expression and quantitative results (Supplementary Figure 12). Our analysis of
378 HuBMAP custom probe panels demonstrated how to evaluate for the potential impact of
379 predicted off-targets by integrating tissue-specific single-cell or bulk RNA-seq data from
380 reference atlases (Supplementary Figures 9, 10). Overall, we anticipate evaluating whether
381 predicted off-target genes are expressed in a tissue-specific manner will be useful for gauging
382 whether predicted off-target binding is likely to meaningfully affect observed gene expression
383 and interpretation when applied to a tissue of interest.
384

385 Other sources of non-specific signal may also arise, including probe self-hybridization or probe-
386 probe interactions (Supplementary Fig 6B). In general, probe binding specificity is influenced by
387 numerous factors, with many methods previously developed to aid in the design of probe
388 sequences while taking these factors into consideration (Wang and Seed 2003; Rouillard et al.
389 2003; Wernersson and Nielsen 2005; Chou 2010; Li et al. 2011; Hu et al. 2020; Hershberg et al.
390 2021; Fornace et al. 2022; Stenberg et al. 2005; Kuemmerle et al. 2024). For example, in the
391 Xenium human breast gene panel, *HDC*, a custom gene not included in the pre-designed Xenium
392 v1 Human Breast Gene Expression Panel, did not have any off-targets predicted by OPT. Yet in
393 Xenium, *HDC* exhibited a distinct spatial pattern and high global expression level, whereas in both
394 Visium and scRNA-seq, *HDC* showed a minimal spatial pattern and sparse expression level
395 respectively (Supplementary Figure 13). This illustrates how discrepancies across platforms can
396 signal potential off-target activity not captured by alignment-based predictions alone and
397 highlights the general importance of experimental validation with orthogonal technologies since
398 sequence alignment-based tools such as OPT may not flag all potential discrepancies.
399
400 We note that most off-target binding impacts paralogs, homologous genes that have diverged
401 following gene duplication events. Paralogs often belong to large gene families whose members
402 can share high sequence similarity, increasing the risk of off-target probe activity. Unlike
403 orthologs with conserved functions across species, paralogs are additional copies that can
404 acquire function-altering mutations (Platt et al. 2000; Pevny et al. 1991; Tsai et al. 1994). Pooling
405 expression signals across paralogs can therefore prevent researchers from capturing their distinct
406 functional roles.

We also found that 10 of the 2,582 Xenium human breast gene panel probe sequences did not align to any reference transcripts in the GENCODE v47 annotation. Upon manually aligning these probes to the GRCh38 genome, we determined that each unmapped sequence corresponded either to regions immediately upstream or downstream of annotated transcripts, or to intronic sequences that would not typically be present in mature RNA. Interestingly, when we aligned several of these probes, such as (ENSG00000125878|TCF15|5d3cbc2) and (ENSG00000169083|AR|a0c6719), to an earlier annotation (GENCODE v28), they instead mapped to exonic regions (Supplementary Figure 14). This suggests that these intronic or intergenic probe sequences were likely designed using older GENCODE versions in which those regions were annotated as exons. This finding illustrates the importance of disclosing the specific annotation version to promote reproducibility, as well as the ongoing variability of human gene annotation. Likewise, as evidenced by our analysis across GENCODE, RefSeq, and CHESS, we emphasize the variation across these reference annotations and therefore recommend using multiple annotations when designing probe sequences and evaluating them for off-target effects to ensure a more comprehensive assessment.

Given these challenges, we advise probes with predicted off-target binding to protein-coding genes based on high sequence homology be avoided in future experiments. Likewise, we encourage the use of tools like OPT to aid in future probe design decisions and help ensure that probes are optimized to minimize off-target binding based on the most current transcriptome annotations. When probes with predicted off-targets cannot be avoided, we encourage the

429 integration of tissue-specific RNA-seq data from HuBMAP and other reference atlases to evaluate
430 for its potential impact. Further, such integration of tissue-specific RNA-seq data from reference
431 atlases into the probe design process itself may offer a data-driven opportunity to minimize the
432 impact of potential off-target binding by enforcing stricter probe-design constraints only where
433 potential off-target genes are highly expressed in the tissue of interest. For datasets that have
434 already been generated using probes with predicted off-target binding, we generally recommend
435 taking into consideration these predictions to avoid drawing misleading conclusions. For
436 example, we recommend expression measurements for genes with predicted off-target binding
437 be omitted from training foundation models to avoid error propagation. Alternatively, when
438 performing integrative analyses that compare or align gene expression with measurements
439 across technologies, it may be necessary to incorporate off-target binding predictions. For
440 instance, integration could be performed between the observed Xenium gene expression and the
441 aggregated expression of the target and predicted off-target genes for the orthogonal
442 technology. Finally, existing literature that base conclusions on genes with predicted off-target
443 binding should be interpreted with caution.
444
445 We emphasize that these findings were missed in the previous Janesick et al. publication from
446 10x Genomics (Janesick et al. 2023). Consistent with previously published observations, we
447 observed a highly correlated total gene expression magnitude between Xenium and Visium as
448 well as scRNA-seq. However, a notable exception is *APOBEC3B*, which is not expressed according
449 to both Visium and scRNA-seq but highly expressed according to Xenium (Figure 2B; Figure 3B) –
450 a discrepancy that Janesick et al. omitted. We emphasize that positive significant average gene

451 expression correlation is a necessary but not sufficient metric for consistency across technologies
452 and that individual data points should be scrutinized. Likewise, validation with orthogonal
453 technologies could have helped identify discrepancies suggestive of off-target effects. We note
454 Janesick et al. used immunofluorescence to validate two genes, *ERBB2* and *MS4A1*, which by our
455 analysis were predicted to exhibit no off-target binding. Although Xenium incorporates blank and
456 negative control probes that are intended to help quantify the rate of non-specific and potential
457 off-target binding, our findings suggest that relying solely on such probes for error detection may
458 be insufficient. Implementing probe redundancy, where the same gene is targeted using different
459 codewords, could provide an additional internal control to enable the detection of off-target
460 binding.
461
462 Although we focus here on the 10x Genomics Xenium technology, we do not exclude the
463 possibility that off-target binding may similarly affect other probe-based gene detection
464 approaches from other commercial vendors. Any technology that relies on hybridization-based
465 detection is inherently susceptible to off-target probe binding when sequence similarity exists.
466 Further, hybridization-based detection often inherently involves a trade-off between sensitivity
467 and specificity. Given these inherent technological limitations, we therefore emphasize the
468 importance of transparency through sharing probe sequences. However, many companies do not
469 release the probe sequences used in their assays, limiting the consumer's ability to fully interpret
470 their results as well as the community's ability to effectively characterize and benchmark
471 performance variation across platforms. Therefore, we strongly recommend that companies
472 publish probe sequences for pre-designed panels and likewise that researchers using these

473 technologies should obtain and publish probe sequences used in their studies to support
474 transparent and reproducible science.
475
476 This is not the first instance in which a commercially available platform has encountered
477 challenges in probe design (McCartney et al. 2016; Harbig 2005; Mecham et al. 2004; Liu et al.
478 2014). These findings underscore the critical role of academic researchers towards ensuring the
479 robustness of industry-led product development by providing oversight, free of financial conflicts
480 of interest through independent federal funding. This complementarity between industry and
481 academia fosters a more rigorous, transparent, and reliable scientific process, ultimately to the
482 benefit of consumers and the public. By shedding light on putative off-target probe binding as
483 well as by providing a tool to enable such off-target binding predictions, this work will help
484 enhance the quality of spatial transcriptomics data and improve the overall reproducibility in
485 spatial transcriptomics research.
486
487
488 **Methods**
489
490 Off-target Probe Tracker Tool (OPT)
491
492 OPT (Off-target Probe Tracker) is a Python program that runs nucmer (Marçais and Kingsford
493 2011) for alignment and then processes the results to predict probe binding based on sequence
494 homology. OPT is available as an open-source Python toolkit at https://github.com/JEFworks-

495 Lab/off-target-probe-tracker. When a user provides a query probe sequence file, a target
496 transcript sequence file, and the annotation used to extract these transcripts, OPT outputs which
497 gene each probe is likely to bind to. Nucmer is a fast nucleotide sequence aligner that uses
498 maximal exact matches as anchors which it then joins together to find longer alignments. By
499 default, OPT saves nucmer results in SAM format and finds perfect sequence matches between
500 a query probe and a target transcript, requiring that alignments consist of only matches and cover
501 the entirety of the query. OPT consists of four modules: (1) `flip` for reverse complementing
502 probe sequences aligned to the opposite strand of their target genes; (2) `track` for aligning
503 probe sequences and processing alignment results; (3) `stat` for compiling summary statistics on
504 the number of off-target binding probes and affected genes; and (4) `all` for running the `flip`,
505 `track`, and `stat` modules at once.
506
507 In the case that a probe's target gene has synonyms, we consider alignments to genes annotated
508 with one of its synonym to still be on-target. For example, if a probe that targets *NARS* shows
509 alignments to a gene called *NARS1*, we don't consider it to be off-target binding. We gathered
510 relevant gene synonym relationships using the GeneCards and HGNC online database.
511
512 OPT also provides a “pad” mode in which imperfect alignments are allowed at either end of the
513 query (i.e., probe sequence). The -pl parameter sets the pad length at either end of the query
514 and OPT allows for any number of mismatches in these padded regions. For example, if the pad
515 length is 10 and the probe sequence length is 40bp, then the middle 20bp are the only part of
516 the probe sequence required to match. As long as the critical region is intact, OPT reports an off-

517 target binding site based on this alignment. By default, -pl is set to 0, and the pad mode is
518 activated by providing a non-zero integer to -pl.
519
520 Obtaining probe sequences for the Xenium v1 Human Breast Gene Expression Panel
521
522 To identify potential off-target binding impacting the 10x Genomics Xenium v1 Human Breast
523 Gene Expression Panel, we obtained the FASTA file of probe sequences from the Janesick et al.
524 publication courtesy of 10x Genomics and available as Supplementary Table 1 for preservation.
525 Notably, this panel slightly deviates from the commercially available Xenium v1 Human Breast
526 Gene Expression Panel (Supplementary Note).
527
528 The target gene names and IDs were extracted from the probe IDs of the following format:
529
530 `> gene_id|gene_name|accession`
531
532 We expected the provided FASTA file to contain probe sequences to be the reverse
533 complemented sequence of their intended target genes and hence align to the reverse strand of
534 their target isoforms. However, when we aligned the breast panel probe sequences to the
535 GENCODE basic (v47) reference transcripts using nucmer, we found that 2,563 / 2,582 of probe
536 sequences aligned on the reverse strand of their target transcripts (i.e., isoforms of their target
537 genes). For consistency, we enforced that all probe sequences be oriented in the same direction
538 and align to the forward strand of their target genes and transcripts. As such, we reverse-

539 complemented these 2,563 probe sequences. We also added this functionality as an OPT module
540 called "flip" in which probe sequences aligned to the reverse strand of their targets are reverse
541 complemented. We expect probe sequences to align to the forward strand of transcripts (i.e.,
542 both oriented in the same direction) during the downstream probe sequence binding prediction
543 step.
544
545 Visium Comparison
546
547 The Visium CytAssist dataset, collected from a breast cancer tissue block utilized in Janesick et
548 al., was downloaded from the 10x Genomics website
549 (https://www.10xgenomics.com/products/xenium-in-situ/preview-dataset-human-breast). This
550 dataset originally contained 4,992 spots with x-y coordinates and included 18,085 genes per spot.
551 Of the 313 unique genes in the Xenium dataset, 307 were shared with the Visium dataset; the
552 other six genes (*AKR1C1*, *ANGPT2*, *BTNL9*, *CD8B*, *POLR2J3*, and *TPSAB1*) were excluded from the
553 analysis because they were absent from the Visium dataset.
554
555 To compare spatial gene expression patterns from Visium and Xenium technologies, we first
556 mapped all the data to the same coordinate space. We used STalign (v1.0.1), a computational
557 tool that utilizes affine transformations along with diffeomorphic metric mapping to align target
558 and source datasets (Clifton et al. 2023). The initial alignment involved only affine
559 transformations and eight manually determined landmarks to align the Visium histology image
560 (source) to the Xenium histology image (target). This transformation brought the Visium image

561 into the coordinate space of the higher-resolution Xenium image. We then applied this learned
562 transformation to the Visium spots, ensuring that they were correctly positioned relative to both
563 histology images. Next, we used STalign to map the Xenium transcripts (source) onto their
564 corresponding Xenium histology image (target) using both affine and diffeomorphic metric
565 mapping. The transcripts were rasterized at 30μm resolution, with an initial affine transformation
566 guided by four manually defined landmarks. Diffeomorphic metric mapping was then performed
567 with the following parameters: a = 2500, epV = 1, niter = 2000, sigmaA = 0.11, sigmaB = 0.10,
568 sigmaM = 0.15, sigmaP = 50, muA = [1, 1, 1], muB = [0, 0, 0], with all other settings left at their
569 defaults. We extracted the overlapping regions between the two datasets (Supplementary Figure
570 3A), which reduced the total spots in the Visium dataset to 3,958. Finally, we aggregated the
571 Xenium gene expression data to ~55μm x 55μm patches that correspond to the spatial locations
572 of the Visium spots, resulting in matched-resolution spatial gene expression for both
573 technologies (Supplementary Figure 3B). The Visium spatial gene expression data is displayed as
574 patches rather than spots to enhance visual saliency and ensure consistency with the Xenium
575 spatial gene expression plots.
576
577 After obtaining matched-resolution spatial gene expression matrices, we quantified agreement
578 between the Visium and Xenium data. For each gene shared between the two platforms, we
579 constructed expression vectors across the aligned spatial spots, where each element
580 corresponded to the log-normalized gene expression at its matched location in the tissue section.
581 We then compared the Visium and Xenium vectors for each gene using two metrics: root-mean-
582 square error (RMSE), computed relative to the line y = x, and Pearson correlation (*r*). To assess

583 whether discrepancies in Xenium measurements could be attributed to predicted off-target
584 genes, we compared the Xenium data of a target gene to the aggregated Visium expression of
585 the target genes with its predicted off-targets. We summed the raw Visium counts of all off-target
586 genes associated with each Xenium gene with predicted off-targets, re-normalized the
587 aggregated counts using CPM followed by log(x + 1), and again calculated the RMSE and Pearson
588 correlation. This approach enabled us to evaluate whether Xenium expression patterns aligned
589 more closely with the intended target gene alone or with the combined expression of its
590 predicted off-target genes.
591
592 Single-cell RNA-seq Comparison
593
594 The Chromium Next GEM 3' scRNA-seq dataset, collected from a breast cancer tissue block
595 utilized in Janesick et al., was downloaded from the 10x Genomics website
596 (https://www.10xgenomics.com/products/xenium-in-situ/preview-dataset-human-breast). This
597 dataset contained 12,388 cells with 36,601 genes per cell. All 313 unique genes present in the
598 Xenium dataset are also in the scRNA-seq dataset, hence both datasets were subsetted to these
599 genes for the analysis.
600
601 Both scRNA-seq and Xenium provide single-cell resolution data. To integrate these datasets, we
602 first removed cells lacking detectable gene expression. We then normalized the combined gene
603 expression data using counts per million (CPM) and applied a log transformation with a
604 pseudocount of 1. Principal component analysis (PCA) is then applied to the normalized data, and

605 batch effects are corrected using Harmony (v1.2.3) on the top 30 principal components (PCs)
606 using default parameters except for theta, which was set to 8, to promote further mixing with
607 clusters across technologies. Finally, Uniform Manifold Approximation and Projection (UMAP) is
608 performed on the harmonized PCs, generating a shared 2D embedding across the two
609 technologies, and the data is further facetted by technology for visualization (Supplementary
610 Figure 5).
611
612 After generating a shared embedding, we quantified differences in gene expression patterns
613 between scRNA-seq and Xenium. We first computed Leiden clusters on the harmonized principal
614 components (resolution = 1.0) to identify transcriptionally similar groups of cells shared across
615 both technologies. For each Leiden cluster, we calculated the mean expression of every gene
616 present in both datasets. Clusters containing fewer than ten cells from either modality were
617 excluded to ensure robust gene-level estimates. To compare expression patterns between
618 scRNA-seq and Xenium for a given gene, we constructed vectors of cluster-level mean expression
619 from both technologies and evaluated their similarity using two metrics: root-mean-square error
620 (RMSE), computed relative to the line y = x, and Pearson correlation (*r*). To investigate whether
621 predicted off-target genes may contribute to the observed target-gene expression in Xenium, we
622 compared the Xenium target gene to the aggregated scRNA-seq expression of itself and its
623 predicted off-targets. We summed the raw counts of a gene and all of its predicted off-targets in
624 the scRNA-seq dataset, re-normalized the aggregated counts using CPM followed by log(x + 1),
625 and again calculated the RMSE and Pearson correlation. This allowed us to test whether the

626 Xenium gene expression more closely reflected the intended gene alone or the combined
627 expression of the intended gene and its predicted off-targets.
628
629 Obtaining custom probe panels from HuBMAP
630
631 To evaluate potential off-target binding in the HuBMAP placenta and multi-tissue (heart, kidney,
632 and lung) custom probe panels, we first downloaded the corresponding BED files from the
633 HuBMAP portal
634 (https://portal.hubmapconsortium.org/browse/dataset/28fe8e4ac8a4193f82fdd9f4d4eb0bb2;
635 https://portal.hubmapconsortium.org/browse/dataset/6f597ca43db80f2499443f5c5bfac97c).
636 Using pyfaidx and pandas in Python, we extracted each probe's target gene name, gene identifier,
637 and genomic coordinates from the BED files, and then generated FASTA files by retrieving the
638 corresponding sequences from the reference genome (GRCh38). These FASTA files were then
639 used as input to OPT to predict potential off-target binding.
640
641 HuBMAP custom probe-panel evaluation
642
643 To assess whether predicted off-target genes were likely to confound Xenium results in the
644 HuBMAP custom probe panels, we evaluated the expression of the predicted off-target genes in
645 matched HuBMAP RNA-seq datasets corresponding to the tissues for which each panel was
646 designed. Four RNA-seq datasets were downloaded from the HuBMAP portal: a bulk RNA-seq
647 dataset for the placenta

648 (https://portal.hubmapconsortium.org/browse/dataset/cad68e5e910ca90bda79cc0d6aa586d1
649) and scRNA-seq datasets for the heart
650 (https://portal.hubmapconsortium.org/browse/dataset/e62b7f592efe97894de86cfabbfddde8),
651 kidney
652 (https://portal.hubmapconsortium.org/browse/dataset/43f83f08bedac2bc4d58767362dfebe2)
653 , and lung
654 (https://portal.hubmapconsortium.org/browse/dataset/6c2b7f1b8bc7e4fb6a1cadf29b09b83a).
655
656 For each tissue, raw count matrices were normalized to counts per million (CPM) and averaged
657 across all cells to obtain a bulk-like mean expression profile. Log1p-transformed mean CPM
658 values were used for all downstream comparisons. OPT was run on the two HuBMAP custom
659 probe panels with all RNA species included and a pad length of 10. Since OPT reports transcript-
660 level identifiers, we removed transcript-specific suffixes from Ensembl IDs to align them with
661 gene symbols present in the RNA-seq datasets. For every target gene in the custom panels, we
662 then compiled its predicted off-target genes based on OPT results and evaluated whether these
663 off-targets were expressed in the matched tissue's RNA-seq profile. To visualize these results, we
664 generated heatmaps in which rows correspond to intended target genes and columns represent
665 their predicted off-targets ordered by decreasing expression. This enabled direct comparison of
666 the magnitude and tissue specificity of potential off-target expression across the HuBMAP
667 datasets.
668
669 Cross-annotation analysis

To compare OPT's results with different reference annotations, we used the most recent releases of GENCODE basic (v47), GENCODE comprehensive (v47), RefSeq (v110), and CHESS (v3.1.3) annotation of the GRCh38 genome. Note that GENCODE "basic" is the more reliable version of the annotation and is much closer to RefSeq and CHESS. GENCODE "comprehensive" includes hundreds of thousands of low-quality annotations, which we included in some of our analyses for completeness. Note also that GRCh38 has many non-reference sequences called "alternative scaffolds"; we removed these for our analysis. We then used gffread to extract transcripts as defined in these annotations by running:

```
$ gffread -w transcripts.fa -g grch38.p12/14.fa annotation.gff
```

The GRCh38.p14 assembly was used during transcript sequence extraction for all reference annotations, except for CHESS which specifies that the annotation maps genes and transcripts onto the GRCh38.p12 assembly. For RefSeq, we renamed the VD(J) segment features as transcript features to ensure consistency, and we also removed transcript sequences with the gene_biotype "pseudogene." RefSeq has a separate biotype called "transcribed_pseudogene," but doesn't annotate transcripts for these features. We considered transcripts annotated for a small subset of just pseudogenes an error in the annotation.

Conflicting interests

692 The authors declare no conflicting interests.

693

694

695 **Acknowledgements**

696 We thank Reza Kalhor for his input on the project and feedback on the manuscript. We thank Ian
697 Fiddes for sharing the Xenium gene panel from the Janesick et al. publication and explaining its
698 relation to the commercially available pre-designed Xenium v1 Human Breast Gene Expression
699 Panel. We thank Sergii Domanskyi, Scott Lindsay-Hewett, and Chenchen Zhu for sharing the
700 HuBMAP custom gene panels. Research reported in this publication was supported by the
701 National Institute of General Medical Sciences of the National Institutes of Health under Awards
702 R35-GM142889 and R35-GM130151, the HuBMAP Integration, Visualization, and Engagement
703 (HIVE) Initiative under Award Number OT2-OD033760, and the National Science Foundation
704 under Grant No. 2047611.

705

706

707

708

References

10x Genomics. 2025a. “10x Genomics Product Catalog.” March 19. www.10xgenomics.com/store/product-catalog?query=xenium.

10x Genomics. 2025b. “TXG_Q4-and-Full-Year-2024-Earnings-Press-Release.” 10x Genomics Reports Fourth Quarter and Full Year 2024 Financial Results and Provides Outlook for 2025, March 19. https://s28.q4cdn.com/592666581/files/doc_financials/2024/q4/TXG_Q4-and-Full-Year-2024-Earnings-Press-Release.pdf.

Chou, Hui-Hsien. 2010. “Shared Probe Design and Existing Microarray Reanalysis Using PICKY.” *BMC Bioinformatics* 11 (1): 196. https://doi.org/10.1186/1471-2105-11-196.

Clifton, Kalen, Manjari Anant, Gohta Aihara, et al. 2023. “STalign: Alignment of Spatial Transcriptomics Data Using Diffeomorphic Metric Mapping.” *Nature Communications* 14 (1): 8123. https://doi.org/10.1038/s41467-023-43915-7.

Clifton, Kalen, Vivien Jiang, Rafael dos Santos Peixoto, et al. 2025. “STcompare: Comparative Spatial Transcriptomics Data Analysis of Structurally Matched Tissues to Characterize Differentially Spatially Patterned Genes.” Preprint, bioRxiv, November 25. https://doi.org/10.1101/2025.11.21.689847.

Fornace, Mark E., Jining Huang, Cody T. Newman, Nicholas J. Porubsky, Marshall B. Pierce, and Niles A. Pierce. 2022. “NUPACK: Analysis and Design of Nucleic Acid Structures, Devices, and Systems.” Preprint, Chemistry, November 11. https://doi.org/10.26434/chemrxiv-2022-xv98l.

Harbig, J. 2005. “A Sequence-Based Identification of the Genes Detected by Probesets on the Affymetrix U133 plus 2.0 Array.” *Nucleic Acids Research* 33 (3): e31–e31. https://doi.org/10.1093/nar/gni027.

Hershberg, Elliot A., Conor K. Camplisson, Jennie L. Close, et al. 2021. “Author Correction: PaintSHOP Enables the Interactive Design of Transcriptome- and Genome-Scale Oligonucleotide FISH Experiments.” *Nature Methods* 18 (10): 1265–1265. https://doi.org/10.1038/s41592-021-01273-6.

Hu, Mengwei, Bing Yang, Yubao Cheng, et al. 2020. “ProbeDealer Is a Convenient Tool for Designing Probes for Highly Multiplexed Fluorescence in Situ Hybridization.” *Scientific Reports* 10 (1): 22031. https://doi.org/10.1038/s41598-020-76439-x.

Jain, Sanjay, Liming Pei, Jeffrey M. Spraggins, et al. 2023. “Advances and Prospects for the Human BioMolecular Atlas Program (HuBMAP).” *Nature Cell Biology* 25 (8): 1089–100. https://doi.org/10.1038/s41556-023-01194-w.

Janesick, Amanda, Robert Shelansky, Andrew D. Gottscho, et al. 2023. “High Resolution Mapping of the Tumor Microenvironment Using Integrated Single-Cell, Spatial and in Situ Analysis.” *Nature Communications* 14 (1): 8353. https://doi.org/10.1038/s41467-023-43458-x.

Korsunsky, Ilya, Nghia Millard, Jean Fan, et al. 2019. “Fast, Sensitive and Accurate Integration of Single-Cell Data with Harmony.” *Nature Methods* 16 (12): 1289–96. https://doi.org/10.1038/s41592-019-0619-0.

Kuemmerle, Louis B., Malte D. Luecken, Alexandra B. Firsova, et al. 2024. “Probe Set Selection for Targeted Spatial Transcriptomics.” *Nature Methods* 21 (12): 2260–70. https://doi.org/10.1038/s41592-024-02496-z.

Li, Qiyuan, Nicolai J Birkbak, Balazs Gyorffy, Zoltan Szallasi, and Aron C Eklund. 2011. “Jetset: Selecting the Optimal Microarray Probe Set to Represent a Gene.” *BMC Bioinformatics* 12 (1): 474. https://doi.org/10.1186/1471-2105-12-474.

Liu, Hongfang, Ionut Bebu, and Xin Li. 2014. “Microarray Probes and Probe Sets.” *Front Biosci* 2 (1): 325–38. https://doi.org/10.2741/e93.

Marçais, Guillaume, Arthur L. Delcher, Adam M. Phillippy, Rachel Coston, Steven L. Salzberg, and Aleksey Zimin. 2018. “MUMmer4: A Fast and Versatile Genome Alignment System.” *PLOS Computational Biology* 14 (1): e1005944. https://doi.org/10.1371/journal.pcbi.1005944.

Marçais, Guillaume, and Carl Kingsford. 2011. “A Fast, Lock-Free Approach for Efficient Parallel Counting of Occurrences of *k* -Mers.” *Bioinformatics* 27 (6): 764–70. https://doi.org/10.1093/bioinformatics/btr011.

McCartney, Daniel L., Rosie M. Walker, Stewart W. Morris, Andrew M. McIntosh, David J. Porteous, and Kathryn L. Evans. 2016. “Identification of Polymorphic and Off-Target Probe Binding Sites on the Illumina Infinium MethylationEPIC BeadChip.” *Genomics Data* 9 (September): 22–24. https://doi.org/10.1016/j.gdata.2016.05.012.

Mecham, Brigham H., Daniel Z. Wetmore, Zoltan Szallasi, Yoel Sadovsky, Isaac Kohane, and Thomas J. Mariani. 2004. “Increased Measurement Accuracy for Sequence-Verified Microarray Probes.” *Physiological Genomics* 18 (3): 308–15. https://doi.org/10.1152/physiolgenomics.00066.2004.

Mudge, Jonathan M, Sílvia Carbonell-Sala, Mark Diekhans, et al. 2025. “GENCODE 2025: Reference Gene Annotation for Human and Mouse.” *Nucleic Acids Research* 53 (D1): D966–75. https://doi.org/10.1093/nar/gkae1078.

O’Leary, Nuala A., Mathew W. Wright, J. Rodney Brister, et al. 2016. “Reference Sequence (RefSeq) Database at NCBI: Current Status, Taxonomic Expansion, and Functional Annotation.” *Nucleic Acids Research* 44 (D1): D733–45. https://doi.org/10.1093/nar/gkv1189.

Pevny, Larysa, M. Celeste Simon, Elizabeth Robertson, et al. 1991. “Erythroid Differentiation in Chimaeric Mice Blocked by a Targeted Mutation in the Gene for Transcription Factor GATA-1.” *Nature* 349 (6306): 257–60. https://doi.org/10.1038/349257a0.

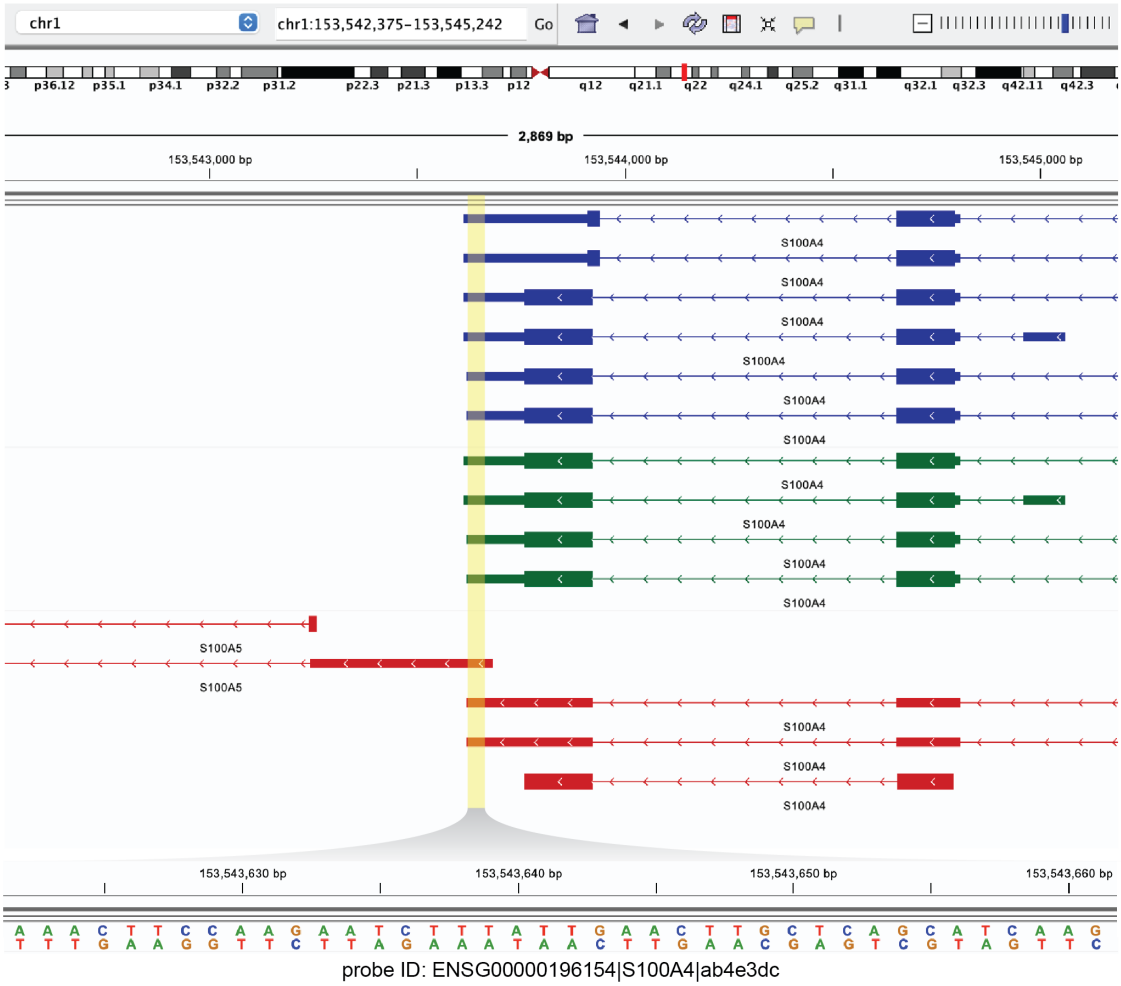
Platt, Adam, Helen C. Ross, Steven Hankin, and Richard J. Reece. 2000. “The Insertion of Two Amino Acids into a Transcriptional Inducer Converts It into a Galactokinase.” *Proceedings of the National Academy of Sciences* 97 (7): 3154–59. https://doi.org/10.1073/pnas.97.7.3154.

Rouillard, J.-M., Micahel Zuker, and Erdogan Gulari. 2003. “OligoArray 2.0: Design of Oligonucleotide Probes for DNA Microarrays Using a Thermodynamic Approach.” *Nucleic Acids Research* 31 (12): 3057–62. https://doi.org/10.1093/nar/gkg426.

Stenberg, Johan, Mats Nilsson, and Ulf Landegren. 2005. “ProbeMaker: An Extensible Framework for Design of Sets of Oligonucleotide Probes.” *BMC Bioinformatics* 6 (1): 229. https://doi.org/10.1186/1471-2105-6-229.

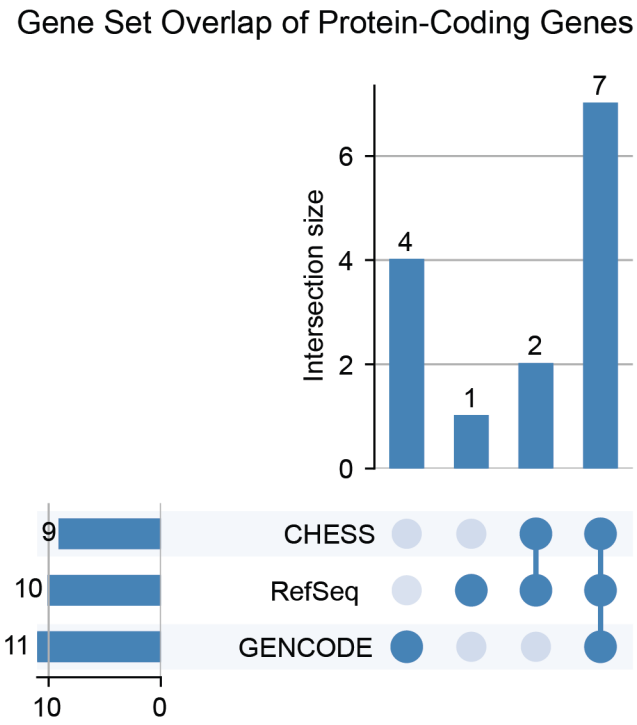
794 Tsai, Fong-Ying, Gordon Keller, Frank C. Kuo, et al. 1994. “An Early Haematopoietic Defect in
795 Mice Lacking the Transcription Factor GATA-2.” *Nature* 371 (6494): 221–26.
796 https://doi.org/10.1038/371221a0.
797 Varabyou, Ales, Markus J. Sommer, Beril Erdogdu, et al. 2023. “CHESS 3: An Improved,
798 Comprehensive Catalog of Human Genes and Transcripts Based on Large-Scale
799 Expression Data, Phylogenetic Analysis, and Protein Structure.” *Genome Biology* 24 (1):
800 249. https://doi.org/10.1186/s13059-023-03088-4.
801 Wang, Xiaowei, and Brian Seed. 2003. “Selection of Oligonucleotide Probes for Protein Coding
802 Sequences.” *Bioinformatics* 19 (7): 796–802.
803 https://doi.org/10.1093/bioinformatics/btg086.
804 Wernersson, R., and H. B. Nielsen. 2005. “OligoWiz 2.0--Integrating Sequence Feature
805 Annotation into the Design of Microarray Probes.” *Nucleic Acids Research* 33 (Web
806 Server): W611–15. https://doi.org/10.1093/nar/gki399.
807 Yahoo Finance. 2025. “10x Genomics, Inc. (TXG).” 10x Genomics, Inc. (TXG), November 20.
808 https://finance.yahoo.com/quote/TXG/.
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823

Supplementary Figures

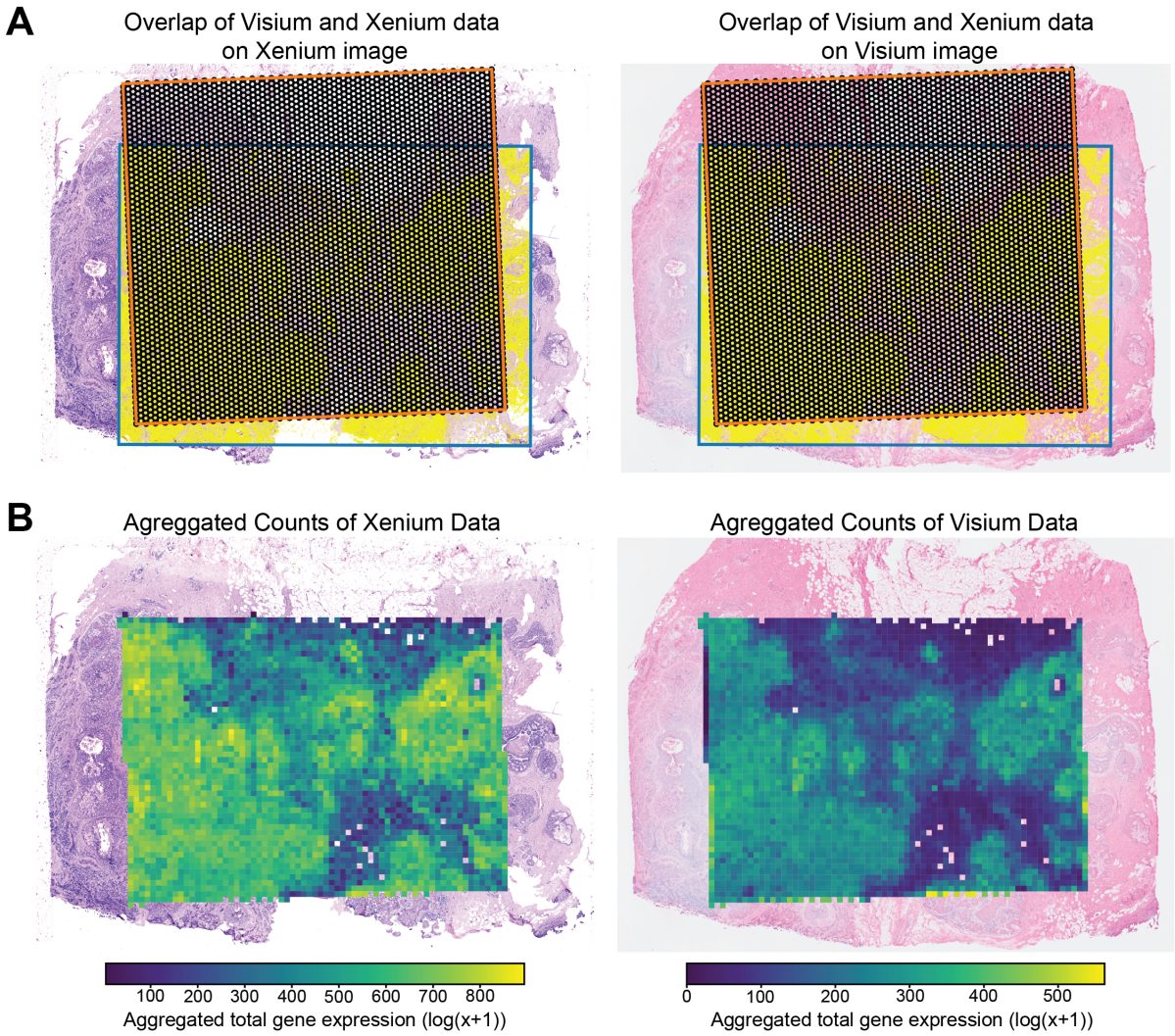


Red: RefSeq Blue: CHESS Green: GENCODE

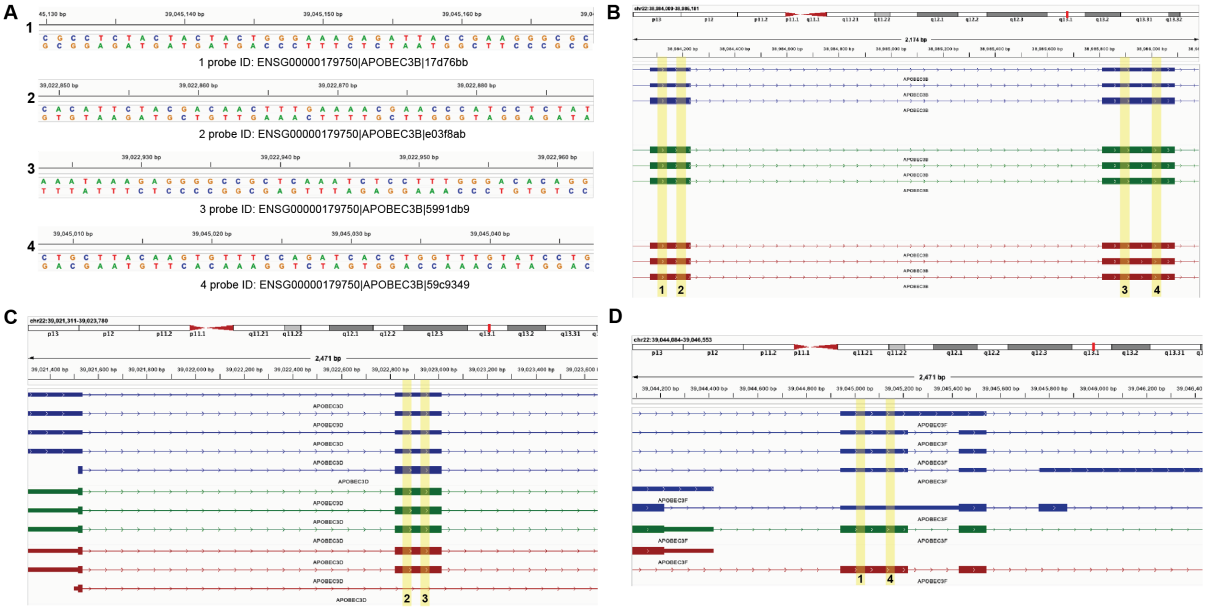
Supplementary Figure 1. Screenshot from the Integrated Genome Viewer (IGV) showing the 40bp probe sequence (ID: ENSG00000196154|S100A4|ab4e3dc) that matches both *S100A5* and *S100A4*. Shown are 6 isoforms from the CHESS v3.1 annotation, 4 from GENCODE basic v47, and 3 from RefSeq v110 for *S100A4*, as well as 2 RefSeq isoforms for the neighboring *S100A5*. The probe sequence aligns to the overlapping region between *S100A5* and *S100A4* gene loci. Matching probe shown in a zoomed-in view below. The forward- and reverse-strand sequences of the probe are shown, and the highlighted area indicates approximately where the probe falls within the gene.



Supplementary Figure 2. UpSet plot illustrating the overlap of protein-coding genes across three genome annotations: GENCODE basic, RefSeq, and CHESS.

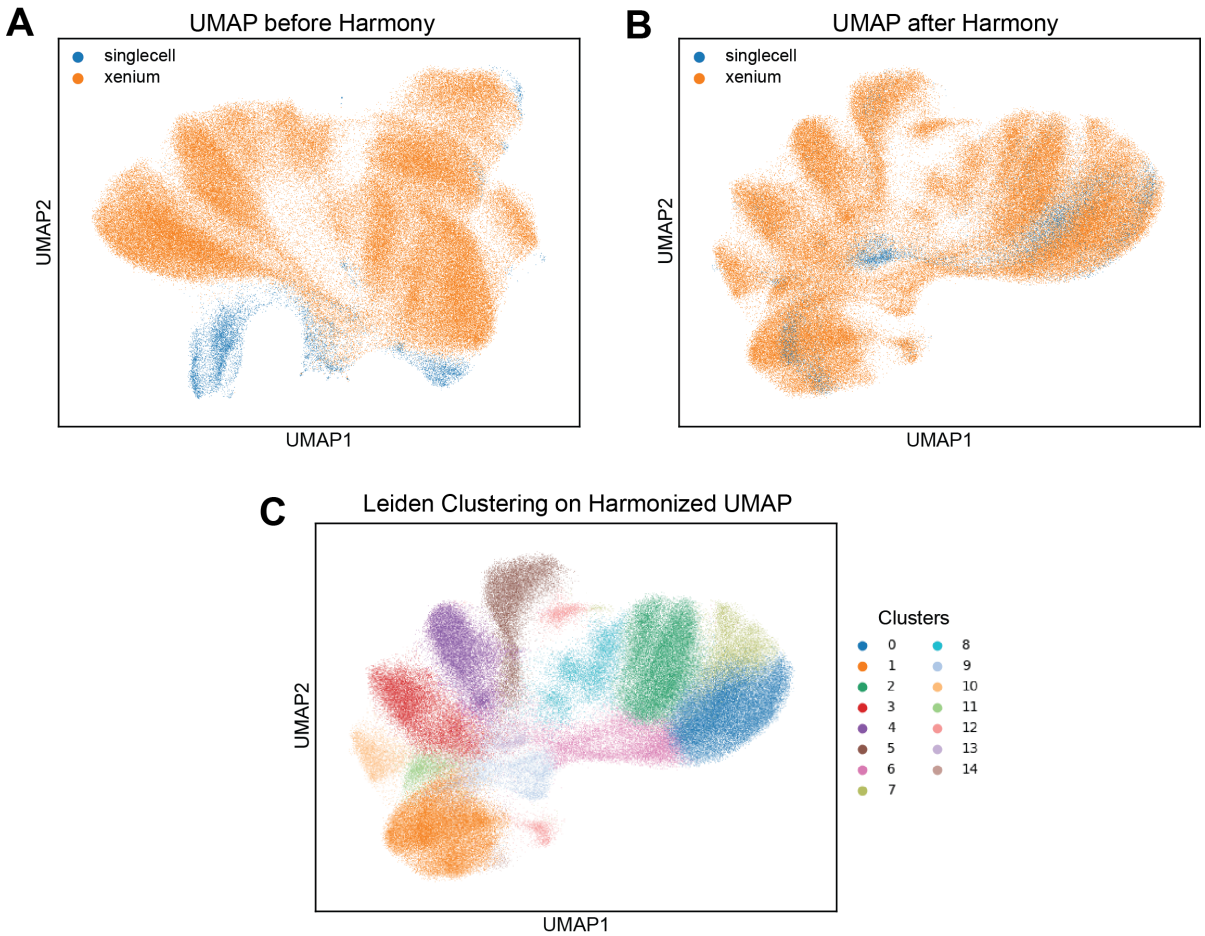


Supplementary Figure 3. (A) Overlap regions between Visium (orange outline) and Xenium (blue outline) data, shown on the Xenium histological image and the Visium histological image, respectively. **(B)** Log transformed aggregated total gene counts for spots (~55μm x 55μm) in both Xenium and Visium datasets, overlaid on their corresponding histological image.

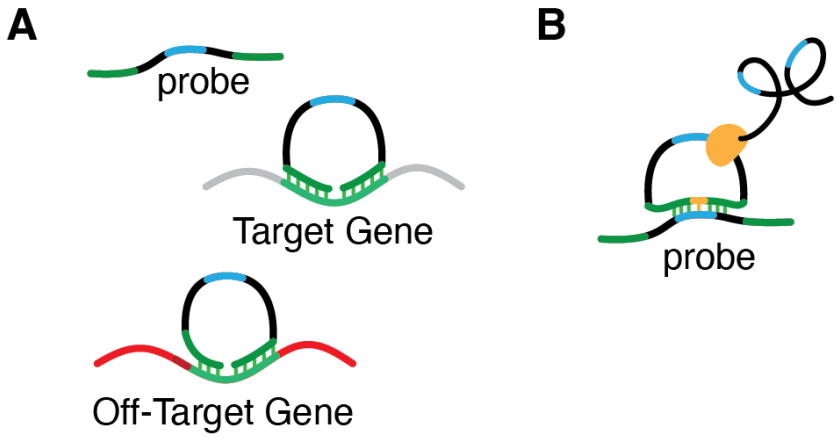


Red: RefSeq Blue: CHESS Green: GENCODE

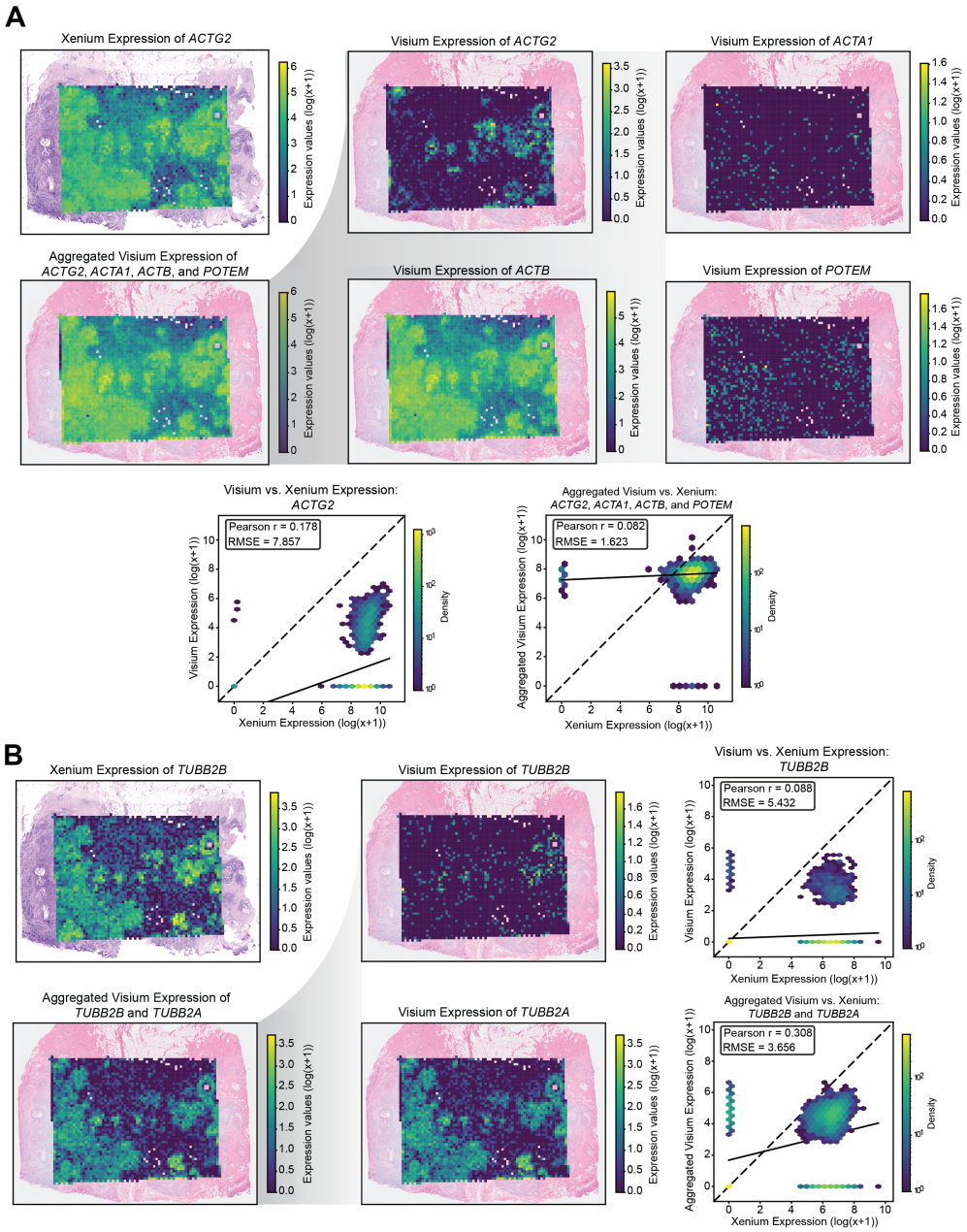
Supplementary Figure 4. Screenshots from the Integrated Genome Viewer (IGV) showing alignments of **(A)** four 40bp probes targeting *APOBEC3B* (ENSG00000179750|APOBEC3B|17d76bb, ENSG00000179750|APOBEC3B|e03f8ab, ENSG00000179750|APOBEC3B|5991db9, and ENSG00000179750|APOBEC3B|59c9349). **(B)** All four probes align to their intended target gene *APOBEC3B* while two of the four probes align to each off-target gene: **(C)** *APOBEC3D* and **(D)** *APOBEC3F*. The forward- and reverse-strand sequences of the probe are shown, and the highlighted areas indicate approximately where the probe falls within the gene. Panels **(B-D)** share a common legend.



Supplementary Figure 5. UMAP visualization of integrated scRNA-seq and Xenium datasets: **(A)** before harmony batch correction and **(B)** after harmony batch correction. **(C)** Leiden clustering results on the harmonized UMAP.

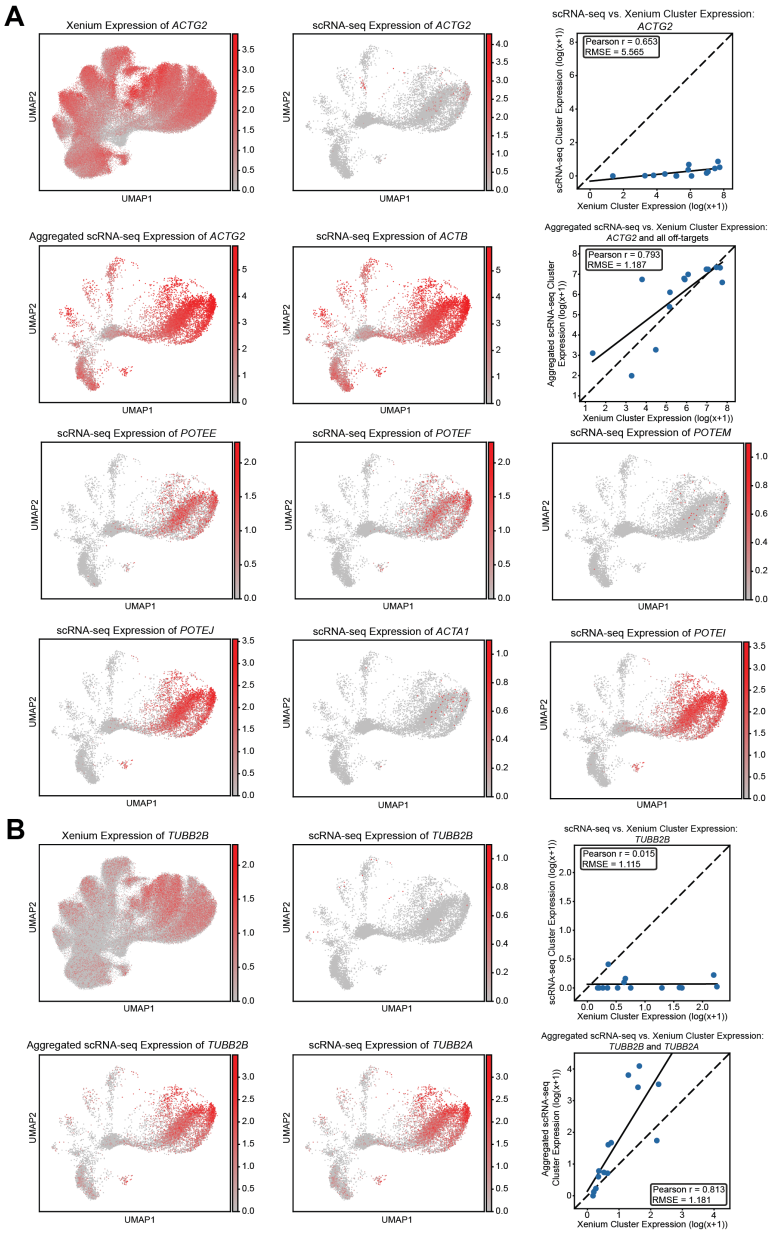


Supplementary Figure 6. (A) Schematic illustrating that hybridization may still occur even when there is a sequence mismatch at the non-ligated ends of the probe sequence. **(B)** Schematic depicting how probes could bind to each other instead of to their intended target.



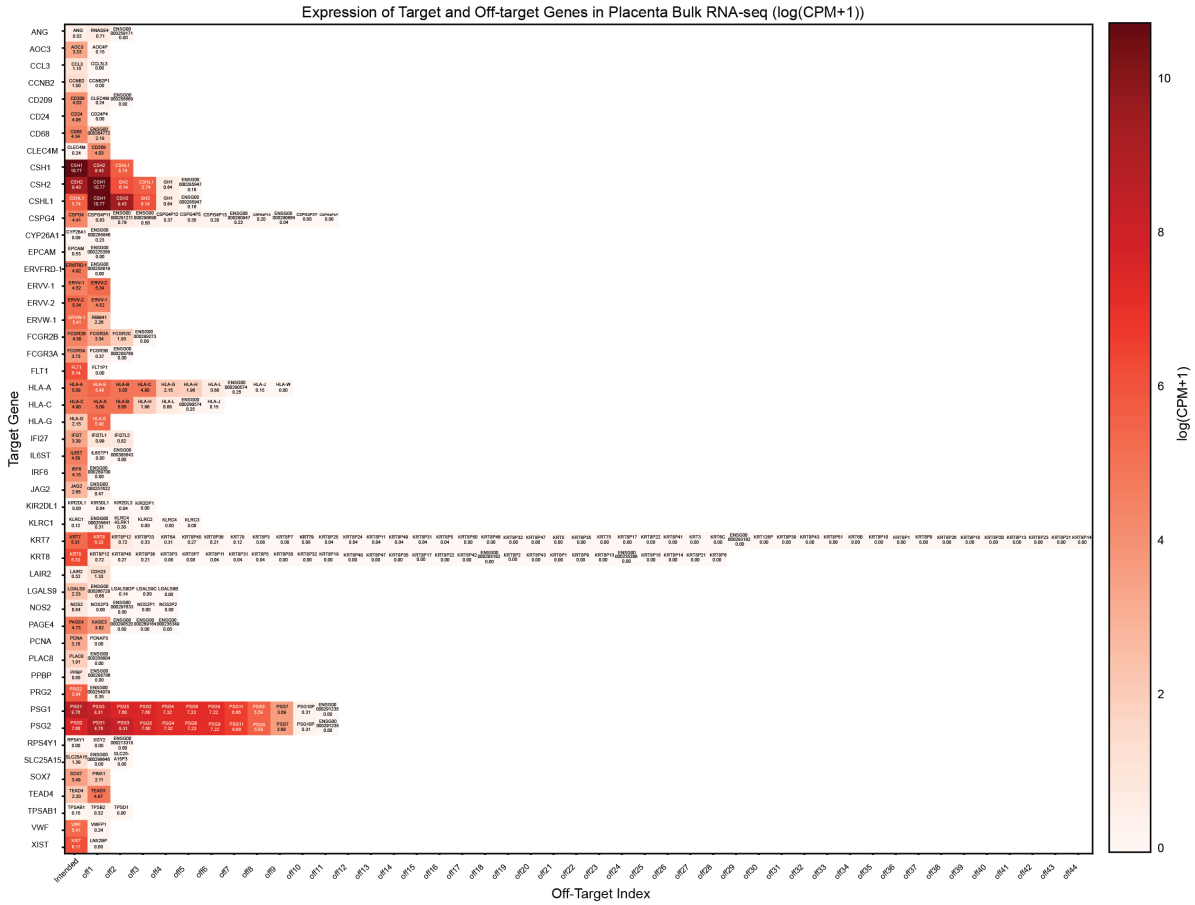
Supplementary Figure 7. (A) Gene expression patterns for *ACTG2*: Xenium expression, Visium expression, the aggregated Visium expression combining *ACTG2* and its predicted off-target gene's expression ACTA1, ACTB, and *POTEM*, and Visium expression of *ACTG2*'s predicted off-targets ACTA1, ACTB, and *POTEM*. Two density plots are shown: one comparing Xenium vs. Visium for *ACTG2* alone, and one comparing Xenium vs. the aggregated Visium expression. The dotted line indicates the identity line (X = Y), and the solid line represents the line of best fit. **(B)** Gene expression patterns for *TUBB2B*: Xenium expression, Visium expression, the aggregated Visium expression combining *TUBB2B* and its predicted off-target gene's expression *TUBB2B* and *TUBB2A*, and Visium expression of *TUBB2B*'s predicted off-target *TUBB2A*. Two density plots are shown: one comparing Xenium vs. Visium for *TUBB2B* alone, and one comparing Xenium vs. the aggregated Visium expression. The dotted line indicates the identity line (X = Y), and the solid line represents the line of best fit.

862

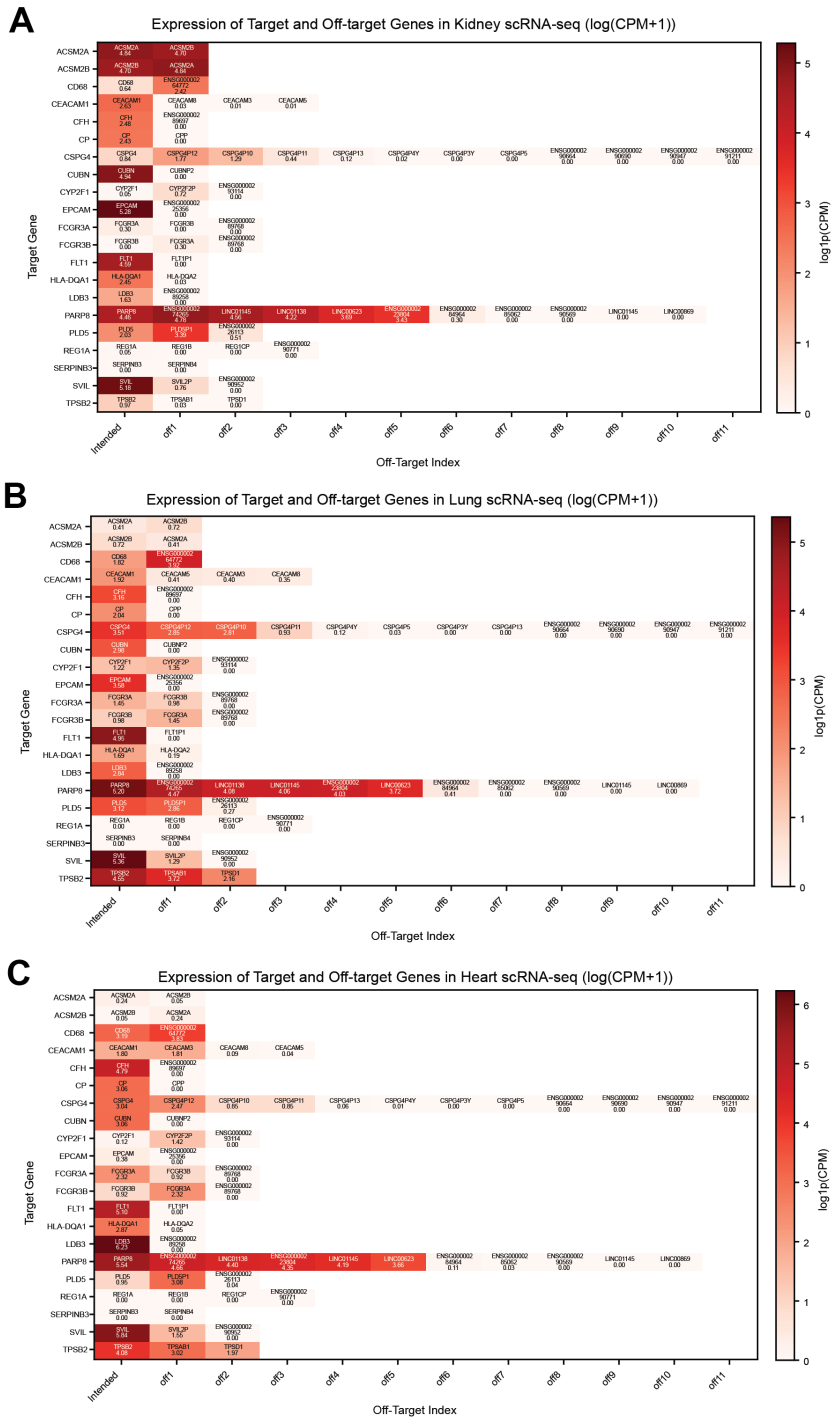


Supplementary Figure 8. (A) Comparison of *ACTG2* expression patterns on harmonized UMAP: Xenium expression, scRNA-seq expression, an aggregated scRNA-seq profile combining *ACTG2* and its predicted off-target gene's expression *ACTB*, *POTEM*, *POTEE*, *POTEF*, *POTEI*, *POTEJ*, and *ACTA1*, and scRNA-seq expression of *ACTG2*'s potential off-targets. Two scatterplots are shown: one comparing Xenium vs. scRNA-seq for *ACTG2* cluster expression alone, and one comparing Xenium vs. the aggregated scRNA-seq cluster expression. The dotted line indicates the identity line (X = Y), and the solid line represents the line of best fit. **(B)** Comparison of *TUBB2B* expression patterns on harmonized UMAP: Xenium expression, scRNA-seq expression, an aggregated scRNA-seq profile combining *TUBB2B* and its predicted off-target gene's expression *TUBB2A*, and scRNA-seq expression of *TUBB2B*'s potential off-target *TUBB2A*. Two scatterplots are shown: one comparing Xenium vs. scRNA-seq for *TUBB2B* cluster expression alone, and one comparing Xenium vs. the aggregated scRNA-seq cluster expression. The dotted line indicates the identity line (X = Y), and the solid line represents the line of best fit.

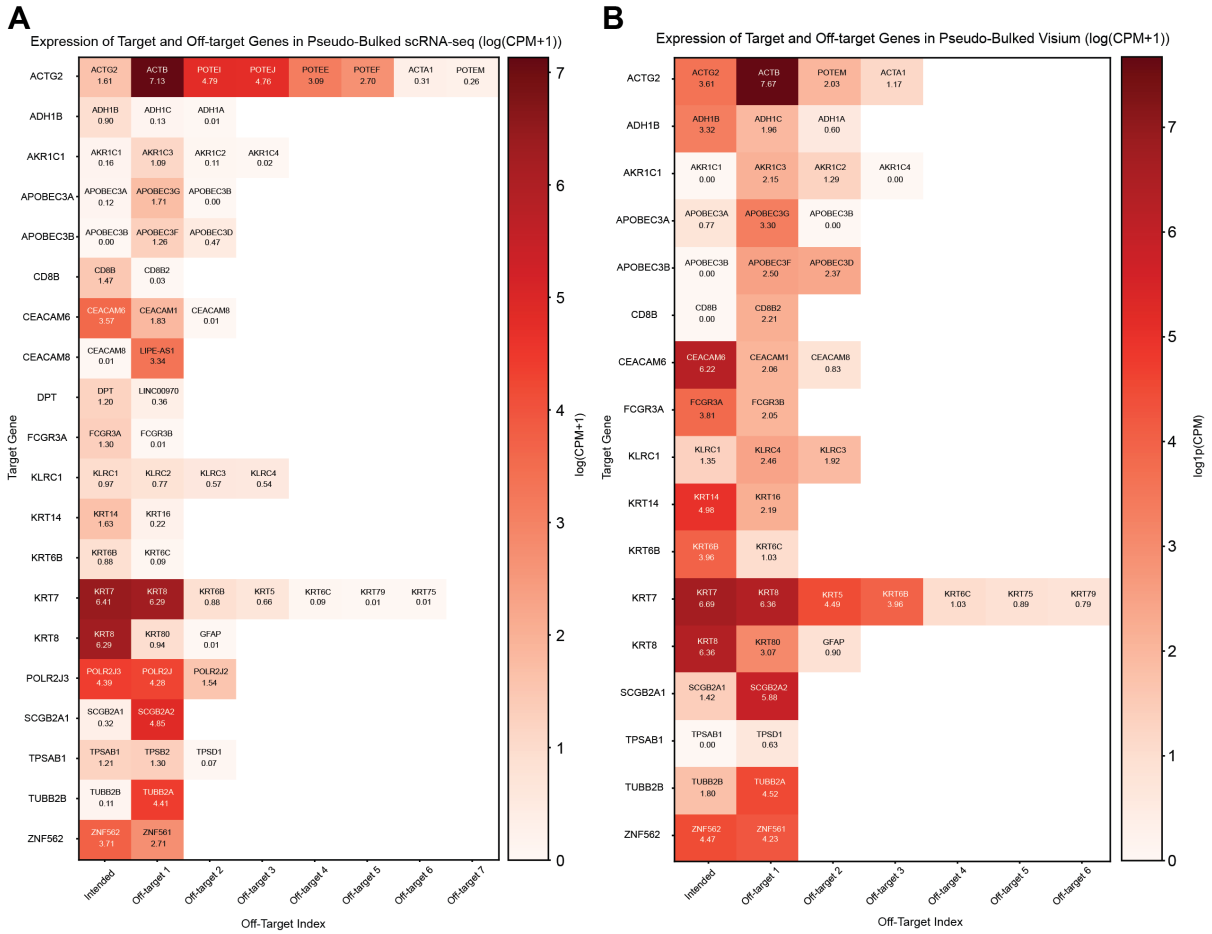
863



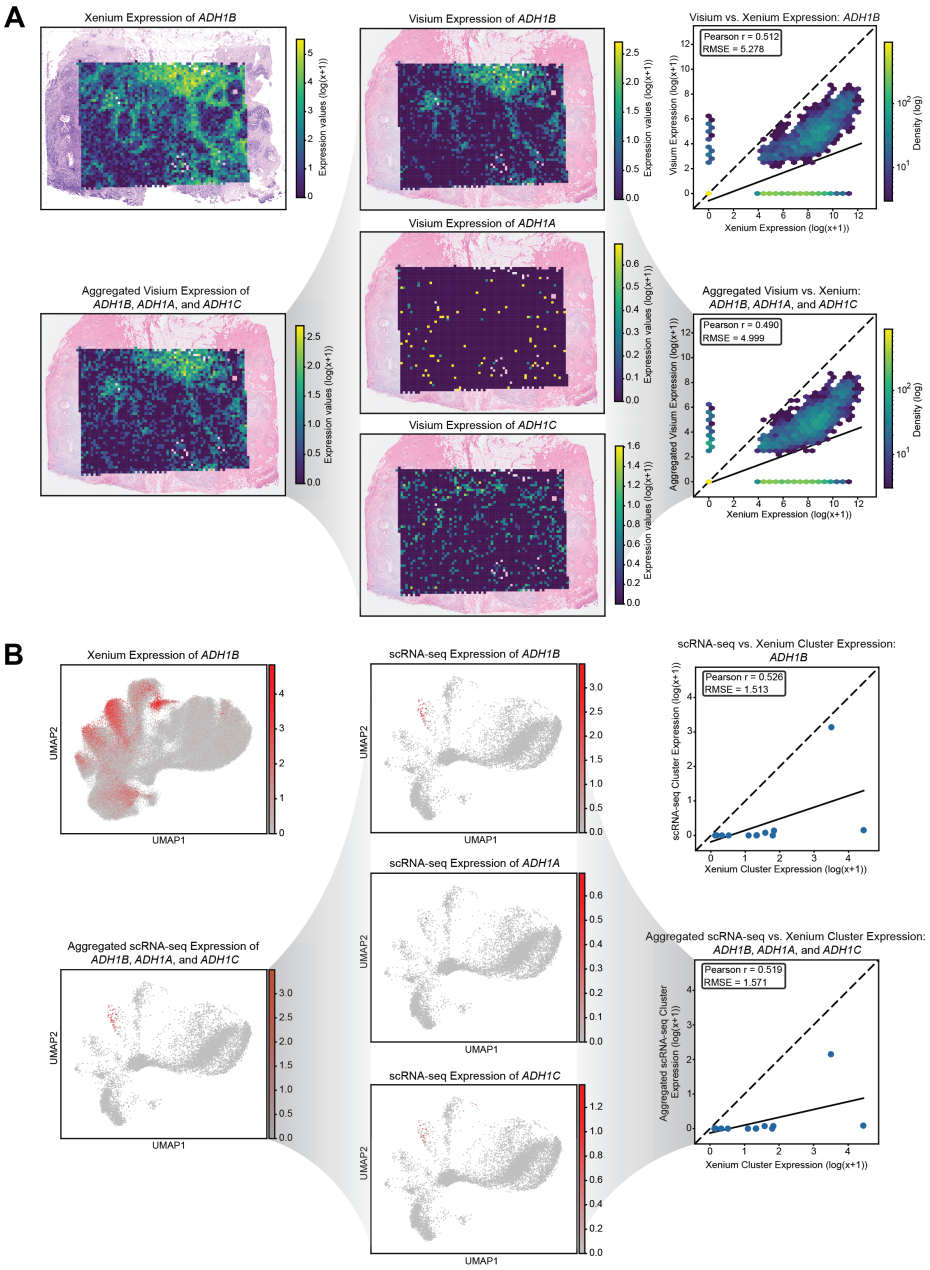
Supplementary Figure 9. Heatmap visualization of target genes and their predicted off-target genes of the HuBMAP placenta custom probe panel using a corresponding placenta bulk RNA-seq dataset. Gene expression values are counts-per-million (CPM) normalized, with a pseudocount added prior to log transformation. The first column shows the expression of each target gene in the placenta bulk RNA-seq dataset, while the remaining columns display the expression of the corresponding predicted off-target genes.



Supplementary Figure 10. Heatmap visualizations of target genes and their predicted off-target genes of the HuBMAP multi custom probe panel using a corresponding **(A)** kidney, **(B)** lung, and **(C)** heart scRNA-seq datasets. Gene expression values are counts-per-million (CPM) normalized, with a pseudocount added prior to log transformation. The first column shows the expression of each target gene in their respective scRNA-seq dataset, while the remaining columns display the expression of the corresponding predicted off-target genes.



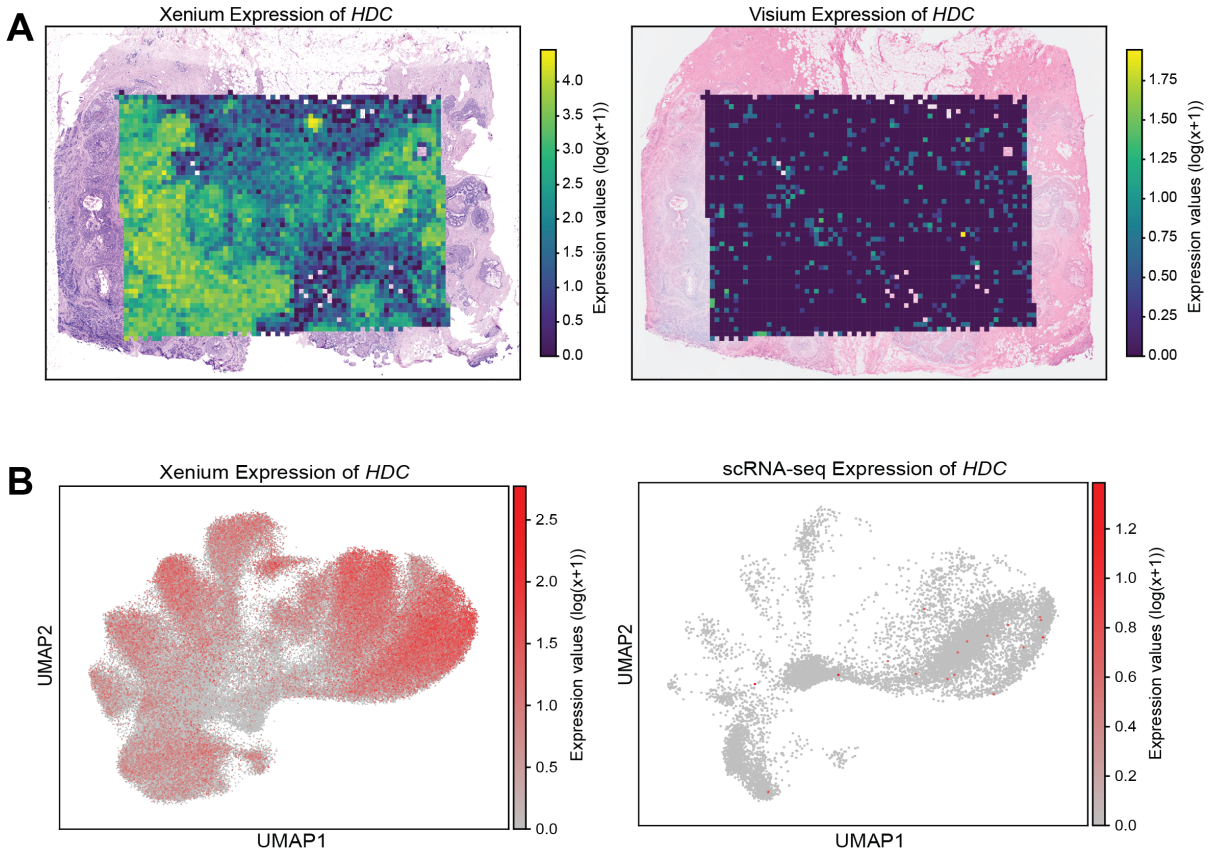
Supplementary Figure 11. Heatmap visualizations of target genes and their predicted off-target genes for the Janesick et al. probes using **(A)** pseudo-bulked scRNA-seq and **(B)** Visium data. Gene expression values are counts-per-million (CPM) normalized, with a pseudocount added prior to log transformation. The first column shows the expression of each target gene in their respective scRNA-seq dataset, while the remaining columns display the expression of the corresponding predicted off-target genes.



Supplementary Figure 12. (A) Gene expression patterns for *ADH1B*: Xenium expression, Visium expression, the aggregated Visium expression combining *ADH1B* and its predicted off-target gene's expression *ADH1A* and *ADH1C*, and Visium expression of *ADH1B*'s predicted off-targets *ADH1A* and *ADH1C*. Two density plots are shown: one comparing Xenium vs. Visium for *ADH1B* alone, and one comparing Xenium vs. the aggregated Visium expression. The dotted line indicates the identity line (X = Y), and the solid line represents the line of best fit. **(B)** Comparison of *ADH1B* expression patterns on harmonized UMAP: Xenium expression, scRNA-seq expression, an aggregated scRNA-seq profile combining *ADH1B* and its predicted off-target gene's expression *ADH1A* and *ADH1C*, and scRNA-seq expression of *ADH1B*'s potential off-targets *ADH1A* and *ADH1C*. Two scatterplots are shown: one comparing Xenium vs. scRNA-seq for *ADH1B* cluster expression alone, and one comparing Xenium vs. the aggregated scRNA-seq cluster expression. The dotted line indicates the identity line (X = Y), and the solid line represents the line of best fit.

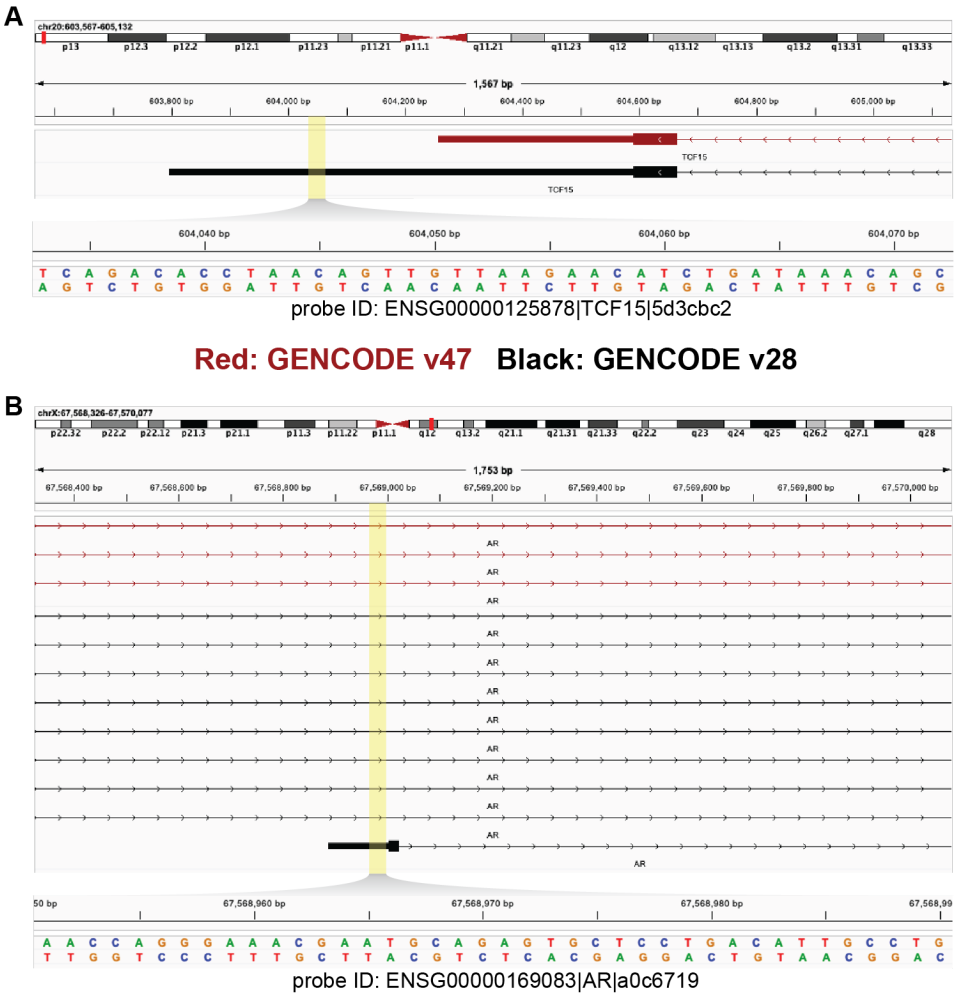
868

869



Supplementary Figure 13. (A) Spatial gene expression of *HDC* overlaid on the corresponding histological images for Xenium and Visium. **(B)** Harmonized UMAP visualization of *HDC* expression for Xenium and scRNA-seq data.

870



Supplementary Figure 14. Screenshots from the Integrated Genome Viewer (IGV) illustrating annotation-dependent differences in probe alignment. **(A)** A 40bp probe (ID: ENSG00000125878|TCF15|5d3cbc2) aligns to an exonic region in GENCODE v28 but to an upstream region in GENCODE v47. **(B)** A 40bp probe (ID: ENSG00000169083|AR|a0c6719) aligns to an exonic region in GENCODE v28 but to an intronic region in GENCODE v47. Matching probe shown in a zoomed-in view below. The forward- and reverse-strand sequences of the probe are shown, and the highlighted areas indicate approximately where the probe falls within the gene. Panels **(A-B)** share a common legend.

Supplementary Tables

FASTA FILE

Supplementary Table 1. Fasta file of probe sequences from 10x Genomics corresponding to the 10x Genomics Xenium v1 Human Breast Gene Expression Panel used in Janesick et al.

Target gene	Number of probes	Predicted binding genes	Number of probes aligned	Gene types
ADH1B	8	*ADH1B,ADH1A,ADH1C*	8,2,1	PC,PC,PC
AKR1C1	9	*AKR1C1,AKR1C2,AKR1C3,AKR1C4*	8,1,1,1	PC,PC,PC,PC
APOBEC3A	10	*APOBEC3A,APOBEC3B*	7,2	PC,PC
APOBEC3B	8	*APOBEC3B,APOBEC3D,APOBEC3F*	8,2,2	PC,PC,PC
C15orf48	6	*C15orf48,MIR147B*	6,1	PC, precursor_RNA
CD8B	16	*CD8B,CD8B2*	15,2	PC,PC
IGF1	13	*IGF1,LINC02456*	11,1	PC,ncRNA
POLR2J3	12	*SPDYE2,SPDYE18,SPDYE5,SPDYE21, SPDYE7P,SPDYE2B,SPDYE16,POLR2J, POLR2J3-UPK3BL2,POLR2J2, POLR2J2-UPK3BL1, POLR2J3,POLR2J4*	2,1,1,1,1,1, 1,3,3,2,2,3,2	PC,PC,PC,PC , misc_RNA, PC,PC,PC,ncRNA,PC, ncRNA,PC,ncRNA
S100A4	5	*S100A4,S100A5*	3,1	PC,PC
SERHL2	8	*SERHL2,SERHL*	8,7	PC,misc_RNA
SLAMF1	11	*SLAMF1,LOC107985220*	10,1	PC,ncRNA
TOMM7	10	*TOMM7,LOC107986346*	3,2	PC,PC
TPD52	8	*TPD52,TPD52-MRPS28*	8,5	PC,PC
TPSAB1	2	*TPSAB1,TPSB2,TPSD1*	2,2,1	PC,PC,PC

Supplementary Table 2. OPT output of genes with predicted off-target binding based on perfect sequence homology in RefSeq. This table shows the 14 genes whose probes in the 10x Genomics Xenium v1 Human Breast Gene Expression Panel that exhibit predicted off-target probe binding, where each off-target alignment involves a perfect 40bp match to the probe sequence. The final column shows the gene types, in order, of each of the off-target genes shown in column 3. Off-target alignments between *CCPG1* probes and *DNAAF1-CCPG1* were excluded. Abbreviations: PC = protein-coding; PG = pseudogene; precursor_RNA = precursor RNA; misc_RNA = miscellaneous RNA; ncRNA = non-coding RNA.

Target gene	Number of probes	Predicted binding genes	Number of probes aligned	Gene types
ADH1B	8	*ADH1B,ADH1A,ADH1C*	8,2,1	PC,PC,PC
AKR1C1	9	*AKR1C1,AKR1C2,AKR1C3, AKR1C4*	8,1,1,1	PC,PC,PC,PC
APOBEC3A	10	*APOBEC3A,APOBEC3B*	8,2	PC,PC
APOBEC3B	8	*APOBEC3B,APOBEC3D,APOBEC3F*	8,2,2	PC,PC,PC
C15orf48	6	*C15orf48,MIR147B*	6,1	PC,miRNA
CD8B	16	*CD8B,CD8B2*	16,2	PC,PC
EPCAM	8	*EPCAM,LOC132386*	8,1	PC,PG
HMGA1	7	*HMGA1,HMGA1P1,HMGA1P2, HMGA1P3,HMGA1P6*	7,1,3,2,2	PC,PG,PG,PG, PG
KRT14	6	*KRT14,KRT16P6*	6,1	PC,PG
KRT8	16	*KRT8,KRT8P33,KRT8P32,KRT8P22, KRT8P3,KRT8P42,KRT8P45,KRT8P17, KRT8P11,KRT8P2,KRT8P36,KRT8P37, KRT8P30*	8,2,1,1,3, 1,2,1,1,2, 1,1,1	PC,PG,PG,PG, PG,PG,PG,PG, PG,PG,PG,PG, PG
MYLK	11	*MYLK,MYLKP1*	11,1	PC,PG
MYO5B	8	*MYO5B,MYO5BP1,MYO5BP2, LOC102724787*	8,1,1,1	PC,PG,PG,PG
POLR2J3	12	*SPDYE2,SPDYE7P,SPDYE21,SPDYE16, SPDYE5,SPDYE18,SPDYE2B,POLR2J, POLR2J3,POLR2J2-UPK3BL1,POLR2J2, POLR2J3-UPK3BL2,POLR2J4*	1,1,1,1, 1,1,1,3, 5,2,3,4,3, 2	PC,PG,PC,PC, PC,PC,PC,PC, PC,PC,lncRNA, lncRNA,PG
SCD	8	*SCD,SCDP1*	8,2	PC,PG
SERHL2	8	*SERHL2,SERHL*	8,7	PC,PG
SMS	8	*SMS,LOC646347,LOC100419037, LOC642590*	8,3,1,1	PC,PG,PG,PG
TOMM7	10	*TOMM7,LOC107986346*	3,2	PC,PC
TPD52	8	*TPD52,TPD52-MRPS28*	8,5	PC,PC
TPSAB1	2	*TPSAB1,TPSB2,TPSD1*	2,2,1	PC,PC,PC
TRAF4	9	*TRAF4,LOC100420513*	9,1	PC,PG
TUBB2B	8	*TUBB2B,TUBB2BP1*	8,1	PC,PG
VOPP1	11	*VOPP1,LOC729086*	10,8	PC,PG
VWF	11	*VWF,VWFP1*	8,1	PC,PG

Supplementary Table 3. OPT output of genes with predicted off-target binding based on perfect sequence homology in CHESS. This table shows the 23 genes whose probes in the 10x Genomics Xenium v1 Human Breast Gene Expression Panel that exhibit predicted off-target probe binding, where each off-target alignment involves a perfect 40bp match to the probe sequence. The final column shows the gene types, in order, of each of the off-target genes shown in column 3. Off-target alignments between *CCPG1* probes and *DNAAF1-CCPG1* were excluded. Abbreviations: PC = protein-coding; PG = pseudogene; miRNA = microRNA.

Perfect match		**GENCODE Basic**	**GENCODE Comprehensive**	**RefSeq**	**CHESS**
Number of transcripts		158,339	387,944	181,627	158,363
All genes	Off-target probes	128	126	44	73
	Affected genes	37	38	14	23
Pseudogenes removed	Off-target probes	101	107	44	35
	Affected genes	29	29	15	11
Protein-coding genes only	Off-target probes	26	28	26	26
	Affected genes	11	11	10	9

Supplementary Table 4. The number of off-target probes and affected genes (from the set of 313 genes in the Xenium panel) found when looking for perfect matches between probe sequences and transcripts in four different reference annotations: GENCODE basic, GENCODE comprehensive, RefSeq, and CHESS. Off-target alignments between *CCPG1* probes and *DNAAF1-CCPG1* were excluded.

Gene	GENCODE	RefSeq	CHESS
ADH1B	✓	✓	✓
AKR1C1	✓	✓	✓
APOBEC3A	✓	✓	✓
APOBEC3B	✓	✓	✓
CD8B	✓	✓	✓
POLR2J3	✓	✓	✓
TPSAB1	✓	✓	✓
AQP1	✓	x	x
C1QA	✓	x	x
CD79B	✓	x	x
CEACAM6	✓	x	x
S100A4	x	✓	x
TOMM7	x	✓	✓
TPD52	x	✓	✓

Supplementary Table 5. Union set of protein-coding genes that OPT predicts to be affected by off-target binding, across three different reference annotations: GENCODE basic, RefSeq, and CHESS.

Target gene	Number of probes	Predicted binding genes	Number of probes aligned	Gene types – GENCODE (v47)
ACTG2	14	*ACTG2,POTEE,ACTB,POTEJ,ACTBP16, ACTA1,POTEI,POTEM,ACTG1P14, ACTG1P12,POTEF,POTEKP,ACTBP15, ENSG00000241838,ACTG1P15, ENSG00000230080,ACTBP1,ACTG1P19*	14,1,1,1, 1,1,1, 1,1, 1,1,1, 1,1, 1,1,1,1	PC,PC,PC,PC,PG,PC, PC,PC,PG,PG, PC,PG,PG,PG,PG, PG,PG,PG
APOC1	2	*APOC1,APOC1P1*	2,1	PC, PG
CD83	8	*CD83,CD83P1*	8,2	PC,PG
CEACAM8	13	*CEACAM8,CEACAMP5,LIPE-AS1, CEACAMP8*	8,1,1,1	PC,PG,lncRNA,PG
ENAH	8	*ENAH,ENAHP1*	8,3	PC,PG
FBLIM1	9	*FBLIM1,FBLIM1P1*	9,1	PC,PG
FCGR3A	8	*FCGR3A,FCGR3B,ENSG00000289768*	8.4.1	PC,PC,PC
FLNB	14	*FLNB, ENSG00000263343*	14,1	PC,PG
KLRC1	8	*KLRC1,KLRC2,KLRC3,KLRC4, ENSG00000255641,KLRC4-KLRK1*	8,1,1,1,1,1	PC,PC,PC,PC,PC, PC
KRT16	5	*KRT16,ENSG00000290978 ,KRT16P3, KRT16P2,ENSG00000290657*	5,2,2,1,1	PC,lncRNA, lncRNA,PG, lncRNA
KRT6B	8	*KRT6B,KRT6C*	8,1	PC,PC
KRT7	5	*KRT7,KRT75,KRT8,KRT8P26,KRT8P20,KRT8P13,KRT8P7,KRT8P12,KRT8P3,KRT5, KRT8P21,KRT8P51,KRT79,KRT8P14 ,KRT8P33,KRT8P35,KRT8P17, KRT6C,KRT8P18,KRT8P10,KRT8P11, KRT8P50,KRT8P25,KRT8P32,KRT6B, KRT8P36,KRT126P,ENSG00000293192, KRT8P46,KRT8P39,KRT8P49*	5,1,1,1,1,1, 1, 1,1,1,1,1,1, 1, 1,1,1,1,1,1, 1, 1,1,1,1,1,1, 1,1,1,1	PC,PC,PC,PG,PG, PG,PG,PG,PG,PC, PG,PG,PC,PG,PG, PG,PG,PC,PG,PG, PG,PG,PG,PG,PC, PG,PG,lncRNA, PG,PG,PG
OCIAD2	8	*OCIAD2,ENSG00000253760, ENSG00000294977*	8,1,1	PC,PG,lncRNA
REXO4	8	*REXO4,ENSG00000259696*	8,1	PC,PG
SCGB2A1	6	*SCGB2A1,SCGB2A2*	6,1	PC,PC
SVIL	8	*SVIL,SVIL2P,ENSG00000290952*	8,1,1	PC,PG,lncRNA
ZEB2	21	*ZEB2,ENSG00000254198*	21,1	PC,PG

ZNF562	10	*ZNF562,ENSG00000291231,ZNF561, ZNF812P,ENSG00000226810, ENSG00000267490*	10,1, 1,1, 1,1	PC,lncRNA,PC,PG, PG,PG

965
966 **Supplementary Table 6.** 18 additional genes that were identified to exhibit potential off-target probe
967 binding when allowing for a 10bp error margin on either side of the binding site when using GENCODE
968 v47. Abbreviations: PC = protein-coding; PG = pseudogene; lncRNA = long non-coding RNA.
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008

Target gene	Number of probes	Predicted binding genes	Number of additional probes aligned	Gene types – GENCODE (v47)
AKR1C1	9	*AKR1C6P*	1	PG
APOBEC3A	10	*APOBEC3G*	1	PC
CEACAM6	8	*CEACAM1,CEACAM8*	1,1	PC,PC
HMGA1	7	*HMGA1P6*	2	PG
KRT14	6	*ENSG00000290978,KRT16,KRT16P2*	1,1,1	lncRNA,PC,PG
KRT8	16	*KRT8P8,KRT8P9,KRT8P44,KRT8P31, KRT8P46,KRT8P48,KRT8P6, ENSG00000235286,KRT8P43,KRT8P47, KRT8P13,KRT8P14,KRT8P50,KRT8P5, KRT8P35,KRT8P7,KRT8P15,KRT8P1, KRT8P19,GFAP,KRT80,KRT8P41, KRT8P34,KRT90P,KRT8P4,KRT8P49, KRT8P40,ENSG00000293192,KRT8P12, KRT8P21*	1,1,1,1,1, 1,2,1,1,1, 1,1,1,1,1,2, 1,2,1,2,2,2, 1,1,2,2,2,2, 1,1,1	PG,PG,PG,PG, PG,PG,PG,PG, PG,PG,PG,PG, PG,PG,PG,PG, PG,PG,PG,PC, PC,PG,PG,PG, PG,PG,PG, lncRNA,PG,PG
LDHB	8	*LDHBP1,LDHBP2*	2,4	PG,PG
MYO5B	8	*MYO5BP3*	2	PG
SMS	8	*ENSG00000251342*	2	PG
TUBB2B	8	*TUBB2A*	1	PC

Supplementary Table 7. 10 genes that were previously predicted to be affected by off-target binding based on perfect matching now show additional predicted off-target interactions when a 10bp error margin is allowed on either side of the binding site when using GENCODE v47. This effect is observed either through the accumulation of new probes with predicted off-target binding or via the identification of additional predicted off-target genes per probe. Abbreviations: PC = protein-coding; PG = pseudogene; lncRNA = long non-coding RNA.

1033

Target gene	Number of probes	Predicted binding genes	Number of additional probes aligned	Gene types – GENCODE (v47)
ANG	8	*ANG,RNASE4,ENSG00000259171*	8,1,1	PC,PC,PC
AOC3	8	*AOC3,AOC4P*	8,3	PC,PG
CCL3	6	*CCL3, CCL3L3*	6,2	PC,PG
CCNB2	8	*CCNB2,CCNB2P1*	7,1	PC,PG
CD209	11	*CD209,ENSG00000288669,CLEC4M*	10,2,2	PC,PC,PC
CD24	12	*CD24,CD24P4*	12,2	PC,PG
CD68	9	*CD68,ENSG00000264772*	9,8	PC,lncRNA
CLEC4M	9	*CLEC4M,CD209*	9,2	PC,PC
CSH1	5	*CSH1,CSHL1,CSH2*	5,2,5	PC,PC,PC
CSH2	8	*CSH2,CSH1,GH2,GH1,ENSG00000285947,CSHL1*	8,8,2,2,2,3	PC,PC,PC,PC,PC,PC
CSHL1	9	*CSHL1,CSH1,CSH2,GH2,ENSG00000285947,GH1*	9,3,3,2,1,2	PC,PC,PC,PC,PC,PC
CSPG4	8	*CSPG4,ENSG00000290664,CSPG4P13,CSPG4P5,CSPG4P10,CSPG4P11,ENSG00000290947,ENSG00000290690,CSPG4P12,ENSG00000291211,CSPG4P4Y,CSPG4P3Y*	8,1,5,2,2,2,1,2,2,1,1,1	PC,lncRNA,PG,PG,PG,PG,lncRNA,lncRNA,PG,lncRNA,PG,PG
CYP26A1	8	*CYP26A1,ENSG00000285846*	5,5	PC,lncRNA
EPCAM	8	*EPCAM,ENSG00000225356*	5,1	PC,PG
ERVFRD-1	8	*ERVFRD-1,ENSG00000258619*	8,1	PC,PG
ERVV-1	8	*ERVV-1,ERVV-2*	8,5	PC,PC
ERVV-2	8	*ERVV-2,ERVV-1*	8,1	PC,PC
ERVW-1	8	*ERVW-1,RBM41*	8,4	PC,PC
FCGR2B	8	*FCGR2B,FCGR2A,ENSG00000289273,FCGR2C*	7,2,1,3	PC,PC,PG,PC
FCGR3A	8	*FCGR3A,FCGR3B,ENSG00000289768*	8,4,1	PC,PC,PG
FLT1	11	*FLT1,FLT1P1*	10,2	PC,PG
HLA-A	8	*HLA-H,HLA-A,HLA-L,HLA-B,HLA-J,HLA-C,ENSG00000290574,HLA-G,HLA-W,HLA-E*	7,3,1,1,1,1,1,1,1,1	PC,PG,PG,PC,PG,PC,lncRNA,PC,PG,PC
HLA-C	8	*HLA-C,HLA-B,HLA-L,HLA-J,HLA-H,HLA-A,ENSG00000290574*	8,2,1,1,1,1,1	PC,PC,PG,PG,PG,PC,lncRNA
HLA-G	8	*HLA-G,HLA-E*	7,1	PC,PC
IFI27	5	*IFI27,IFI27L1,IFI27L2*	5,1,1	PC,PC,PC
IL6ST	10	*IL6ST,IL6STP1,ENSG00000305843*	10,2,1	PC,PG,lncRNA
IRF6	8	*IRF6,ENSG00000289700*	8,8	PC,PC

JAG2	8	*JAG2,ENSG00000257622*	8,3	PC,lncRNA
KIR2DL1	8	*KIR2DL1,KIR2DP1,KIR3DL1,KIR2DL3*	7,1,1,2	PC,PG,PC,PC
KLRC1	8	*KLRC1,KLRC2,KLRC4,KLRC3,ENSG00000255641,KLRC4-KLRK1*	8,1,1,1,1,1	PC,PC,PC,PC,PC,PC
KRT7	7	*KRT7,KRT8P25,KRT8P14,KRT8P7,KRT6C,KRT8P50,KRT8,KRT8P39,KRT8P3,KRT75,KRT8P20,KRT8P18,KRT8P21,KRT126P,KRT8P12,KRT8P10,KRT8P32,KRT5,KRT8P33,ENSG00000293192,KRT8P13,KRT8P49,KRT8P26,KRT8P35,KRT6B,KRT8P51,KRT8P36,KRT79,KRT8P17,KRT8P11,KRT8P46,KRT8P1,KRT8P9,KRT8P31,KRT8P5,KRT8P41,KRT8P45,KRT8P23,KRT8P47,KRT8P43,KRT78,KRT6A,KRT3,KRT8P24,KRT8P22,KRT7-AS*	3,1,2,1,1,1,2,1,2,2,2,2,1,1,2,1,2,1,2,2,1,2,1,2,1,1,2,2,2,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1	
KRT8	10	*KRT8,ENSG00000293192,KRT8P21,KRT8P12,KRT8P32,KRT8P48,KRT8P36,KRT8P3,ENSG00000235286,KRT8P19,KRT8P47,KRT8P1,KRT8P11,KRT8P31,KRT8P7,KRT8P2,KRT8P35,KRT8P13,KRT8P22,KRT8P15,KRT8P17,KRT8P43,KRT8P14,KRT8P42,KRT8P50,KRT8P6,KRT8P5,KRT8P9,KRT8P45*	10,1,1,1,3,1,2,3,2,1,2,1,2,1,1,1,1,2,2,1,1,1,1,1,1,1,1,1,1	PC,lncRNA,PG,PG,PG,PG,PG,PG,PG,PG,PG,PG,PG,PG,PG,PG,PG,PG,PG,PG,PG,PG,PG,PG,PG,PG,PG,PG,PG,PG,PG,PG
LAIR2	6	*LAIR2,CDH23*	5,1	PC,PC
LGALS9	8	*LGALS9,ENSG00000266728,LGALS9DP,LGALS9B,LGALS9C*	7,3,3,2,2	PC,PC,PG,PC,PC
NOS2	8	*NOS2,NOS2P2,ENSG00000297633,NOS2P3,NOS2P1*	8,1,1,1,1	PC,PG,lncRNA,PG,PG
PAGE4	6	*PAGE4,ENSG00000235349,XAGE3,ENSG00000290520,ENSG00000169164*	6,1,1,1,1	PC,PG,PC,lncRNA,PG
PCNA	8	*PCNA,PCNAP3*	8,2	PC,PG
PLAC8	11	*PLAC8,ENSG00000256804*	11,1	PC,PG
PPBP	8	*PPBP,ENSG00000288796*	8,4	PC,PC
PRG2	4	*PRG2,ENSG00000254979*	4,1	PC,PC
PSG1	8	*PSG1,PSG10P,PSG5,PSG2,ENSG00000291235,PSG6,PSG7,PSG11,PSG4,PSG8,PSG3,PSG9*	8,4,3,3,4,2,3,5,2,3,4,1	PC,PG,PC,PC,lncRNA,PC,PC,PC,PC,PC,PC,PC
PSG2	8	*PSG2,PSG9,PSG5,PSG8,PSG4,PSG6,PSG3,PSG1,PSG7,PSG10P,PSG11,ENSG00000291235*	8,3,4,5,2,2,4,4,3,2,3,1	PC,PC,PC,PC,PC,PC,PC,PC,PC,PG,PC,lncRNA
RPS4Y1	8	*RPS4Y1,XGY2,ENSG00000213318*	8,3,1	PC,PG,PG
SLC25A15	8	*SLC25A15,SLC25A15P3,ENSG00000299646*	7,1,1	PC,PG,lncRNA
SOX7	8	*SOX7,PINX1*	8,4	PC,PC
TEAD4	8	*TEAD4,TEAD3*	5,1	PC,PC

TPSAB1	8	*TPSAB1,TPSB2,TPSD1*	7,6,3	PC,PC,PC
VWF	8	*VWF,VWFP1*	8,1	PC,PG
XIST	15	*XIST,LNX2BP*	15,3	lncRNA,PG

Supplementary Table 8. OPT output for 49 genes with predicted off-target binding in the HuBMAP placenta custom probe panel, generated using GENCODE v47 and allowing a 10bp mismatch on either end of each probe. The final column lists the gene types, in order, corresponding to the off-target genes shown in column 3. Abbreviations: PC = protein-coding; PG = pseudogene; precursor_RNA = precursor RNA; misc_RNA = miscellaneous RNA; ncRNA = non-coding RNA.

Target gene	Number of probes	Predicted binding genes	Number of additional probes aligned	Gene types – GENCODE (v47)
ACSM2A	8	*ACSM2A,ACSM2B*	6,4	PC,PC
ACSM2B	8	*ACSM2B,ACSM2A*	8,4	PC,PC
ATRNL1	8	*ATRNL1,ENSG00000285582*	7,1	PC,lncRNA
CD68	7	*CD68,ENSG00000264772*	7,6	PC,lncRNA
CEACAM1	8	*CEACAM1,CEACAM3,CEACAM8,CEACAM5*	8,1,1,1	PC,PC,PC,PC
CFH	8	*CFH,ENSG00000289697*	7,7	PC,PC
CP	8	*CP,CPP*	8,3	PC,PG
CSPG4	8	*CSPG4,CSPG4P11,ENSG00000290690,ENSG00000291211,CSPG4P12,CSPG4P13,CSPG4P10,CSPG4P5,ENSG00000290947,ENSG00000290664,CSPG4P4Y,CSPG4P3Y*	8,2,2,1,2,5,2,2,1,1,1,1	PC,PG,lncRNA,lncRNA,PG,PG,PG,PG,lncRNA,lncRNA,PG,PG
CUBN	8	*CUBN,CUBNP2*	8,1	PC,PG
CYP2F1	8	*CYP2F1,ENSG00000293114,CYP2F2P*	6,2,3	PC,lncRNA,PG
EPCAM	8	*EPCAM,ENSG00000225356*	6,1	PC,PG
FCGR3A	8	*FCGR3A,FCGR3B,ENSG00000289768*	8,4,1	PC,PC,PC
FCGR3B	12	*FCGR3B,FCGR3A,ENSG00000289768*	10,7,4	PC,PC,PC
FLT1	17	*FLT1,FLT1P1*	16,4	PC,PG
HLA-DQA1	8	*HLA-DQA1,HLA-DQA2*	7,2	PC,PC
LDB3	13	*LDB3,ENSG00000289258*	11,7	PC,PC
PARP8	15	*PARP8,LINC00869,LINC01138,LINC00623,ENSG00000290569,ENSG00000223804,LINC01145,ENSG00000274265,ENSG00000284964,ENSG00000285062*	14,2,2,2,2,2,2,2,1,1	PC,lncRNA,lncRNA,lncRNA,lncRNA,PG,lncRNA,lncRNA,PG,PG
PLD5	8	*PLD5,ENSG00000226113,PLD5P1*	8,2,2	PC,PG,PC
REG1A	8	*REG1A,ENSG00000290771,REG1CP,REG1B*	7,2,2,1	PC,lncRNA,PG,PC
SERPINB3	8	*SERPINB3,SERPINB4*	8,3	PC,PC
SLAMF1	11	*SLAMF1,ENSG00000228863*	10,1	PC,lncRNA
SVIL	8	*SVIL,ENSG00000290952,SVIL2P*	8,1,1	PC,lncRNA,PG
TPSB2	1	*TPSB2,TPSD1,TPSAB1*	1,1,1	PC,PC,PC
VIM	8	*VIM,ENSG00000234961*	7,1	PC,lncRNA

Supplementary Table 9. OPT output for 24 genes with predicted off-target binding in the HuBMAP multi custom probe panel, generated using GENCODE v47 and allowing a 10bp mismatch on either end of each probe. The final column lists the gene types, in order, corresponding to the off-target genes shown in column 3. Abbreviations: PC = protein-coding; PG = pseudogene; precursor_RNA = precursor RNA; misc_RNA = miscellaneous RNA; ncRNA = non-coding RNA.

1064 **ZIP FILE**

1065 **Supplementary Table 10.** OPT results for all publicly available 10x Genomics probe sets. Results include
1066 all possible RNA species and using a pad length (-pl) of 10.

1067

1068

1069 **Supplementary Note: Differences between 3 Xenium breast gene panels and the impact on**
1070 **interpretation of results**

1071

1072 In the previous iteration of this manuscript, we used the publicly available Xenium v1 Human
1073 Breast Gene Expression Panel probe sequences from the 10x Genomics website (prior to April
1074 10, 2025) and assumed it was used to generate the Xenium breast cancer data from Janesick et
1075 al. based on the paper's methodological descriptions. 10x Genomics has since noted on their
1076 public website as well as through private communications that this publicly available file
1077 erroneously included extra probe sequences that are not used. Further, the Xenium breast cancer
1078 data presented in Janesick et al. uses a human breast gene panel with probe sequences that
1079 represent an earlier iteration of the commercially available Xenium v1 Human Breast Gene
1080 Expression Panel. In particular, compared to the commercially available panel, the Janesick et al.
1081 panel has 24 genes with one or more probes modified in the final panel, along with the probe
1082 sequences targeting the 33 custom genes added by Janesick et al. This impacts our interpretation
1083 of observed results to provide stronger evidence of putative imperfect sequence homology based
1084 off-target probe binding.

1085

1086 For clarity, we refer to the three probesets as:

1087 - Probeset A: the previously analyzed publicly available Xenium v1 Human Breast Gene
1088 Expression Panel probe sequences (prior to April 10, 2025)
1089 - Probeset B: the Xenium probe sequences used in the Janesick et al. panel
1090 - Probeset C: the currently commercial pre-designed Xenium v1 Human Breast Gene
1091 Expression Panel (after April 10, 2025)

1092

1093 To provide a specific example, in the previous iteration of this manuscript, OPT identified a probe
1094 sequence targeting *TUBB2B* that exhibited perfect sequence homology with *TUBB2A*:

1095 `>ENSG00000137285|TUBB2B|1dec8c0`

1096 `GTTCATGATGCGGTCTGGGTACTCTTCCCGGATCTTGCTG`

1097 Analysis with orthogonal spatial and single-cell transcriptome profiling technologies suggested
1098 the *TUBB2B* gene expression pattern observed in the Xenium breast cancer data from Janesick
1099 et al. represented an aggregation of both *TUBB2B* and *TUBB2A* consistent with off-target binding.
1100 Given our assumption at the time that the Xenium breast cancer data had been generated using
1101 Probeset A, we therefore believed this perfect homology probe was responsible for the observed
1102 off-target signal.

1103

1104 However, we now understand that the Xenium breast cancer data from Janesick et al. were
1105 generated using Probeset B. Probeset B no longer contains this specific probe for *TUBB2B* with
1106 perfect sequence homology with *TUBB2A*. However, the observed *TUBB2B* gene expression
1107 pattern still represents an aggregation of both *TUBB2B* and *TUBB2A* based on analysis with

1108 orthogonal spatial and single-cell transcriptome profiling technologies. Although Probeset B no

1109 longer contains this specific probe for *TUBB2B* with perfect sequence homology with *TUBB2A*, it

1110 contains one probe with imperfect sequence homology with *TUBB2A* (=30X1=6X3):

1111 `>ENSG00000137285|TUBB2B|ed52e1c`

1112 `TTGTCAATGCAGTAGGTTTCATCTGTGTTTTCCACCAGCT`

1113 The observed comparative gene expression patterns therefore provide newfound strong support

1114 for putative imperfect sequence homology off-target binding.

1115

1116 While we do not have Xenium breast cancer data generated using Probeset C, this *TUBB2B* probe

1117 is also present in Probeset C:

1118 `>TUBB2B (ENSG00000137285) | 3 | 5`

1119 `TTGTCAATGCAGTAGGTTTCATCTGTGTTTTCCACCAGCT`

1120 This suggests that off-target effects may still impact commercially available pre-designed panels

1121 such as Probeset C (Supplementary Table 10).

1122