Relationship between spiking activity and simultaneously recorded fluorescence signals in transgenic mice expressing GCaMP6

Lawrence Huang[1*], Ulf Knoblich[1*], Peter Ledochowitsch[1], Jérôme Lecoq[1], R. Clay Reid[1], Saskia E. J. de Vries[1], Michael A. Buice[1], Gabe J. Murphy[1], Jack Waters[1], Christof Koch[1], Hongkui Zeng[1#], Lu Li[1,2,3,4#]

1 Allen Institute for Brain Science, Seattle, WA 98109, USA
2 Guangdong Provincial Key Laboratory of Malignant Tumor Epigenetics and Gene Regulation, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou, Guangdong Prov., China 510120
3 Medical Research Center, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou, Guangdong Prov., China 510120
4 Centre for Brain Science and Brain-Inspired Intelligence, Guangdong-Hongkong-Macao Greater Bay Area, China

*These authors contributed equally
#Correspondence should be addressed to hongkuiz@alleninstitute.org (H.Z.) or lilu67@mail.sysu.edu.cn (L.L.)

Keywords
calcium imaging, genetically encoded calcium indicator, action potential, excitatory neurons, calibration

Abstract

Two-photon calcium imaging is often used with genetically encoded calcium indicators (GECIs) to investigate neural dynamics, but the relationship between fluorescence and action potentials (spikes) remains unclear. Pioneering work linked electrophysiology and calcium imaging *in vivo* with viral GECI expression, albeit in a small number of cells. Here we characterized the spike-fluorescence transfer function *in vivo* of 91 layer 2/3 pyramidal neurons in primary visual cortex in four transgenic mouse lines expressing GCaMP6s or GCaMP6f. We found that GCaMP6s cells have spike-triggered fluorescence responses of larger amplitude, lower variability and greater single-spike detectability than GCaMP6f cells. Single spike detection rates differed substantially across neurons in each line. They declined from ~40-90% at 5% false positive rate under high-resolution imaging to ~10-15% when imaging hundreds of neurons across a larger field of view. Our dataset thus provides quantitative insights to support more refined inference of neuronal activity from calcium imaging data.

Introduction

Genetically encoded calcium indicators (GECIs) are widely used with two-photon (2-p) laser scanning microscopy to report neuronal activity within local populations *in vivo* (Luo et al., 2018). By reporting changes in calcium-dependent fluorescence, this optical approach is minimally invasive and enables simultaneous measurement of activity from hundreds or even thousands of neurons at single-cell resolution. As such, calcium imaging has become an attractive alternative to extracellular electrophysiology for surveying large neuronal populations of genetically identified cells, in particular over multiple sessions. Indeed, using a newly developed GECI such as GCaMP6s, fluorescence changes associated with isolated single spikes (action potentials) *in vivo* can be reliably detected when imaged at sufficiently high spatiotemporal resolution (Chen et al., 2013).

However, despite recent advances in imaging approaches and GECI development, calcium imaging remains an indirect measure of a neuron's spiking activity. Inferring the underlying spike train or firing rate from calcium imaging remains challenging (Theis et al., 2016; Berens et al., 2018), in part because the relationship between spiking activity and calcium dynamics is complex and has not been fully characterized. The spike-to-calcium fluorescence transfer function depends on several factors, including (1) the relationship between changes in GECI fluorescence and changes in intracellular calcium concentration, and buffering effects (Rose et al., 2014); (2) the neuronal cell types expressing the GECI, due to differences in intracellular calcium dynamics; (3) GECI expression level (*e.g.* differential viral uptake and expression when using viral GECI delivery); and (4) the behavioral state of the animal, due to changes in calcium dynamics, *e.g.* by neuromodulators (Nadim and Bucher, 2014). Together, these variables underlie the difficulty of linking calcium imaging studies to the large body of existing knowledge concerning spiking activity.

Compared to viral expression, transgenic mouse lines offer convenience (e.g. bypassing virus injection and associated procedures) and achieve more uniform GECI expression in genetically defined neuronal populations (Madisen et al., 2015; Daigle et al., 2018). This should permit a more straightforward comparison of activity across cells and animals. Using our intersectional transgenic mouse lines that enable Cre recombinase-dependent expression of GCaMP6s or GCaMP6f, we simultaneously characterized the spiking activity and fluorescence of individual GECI-expressing pyramidal neurons in layer (L) 2/3 of mouse primary visual cortex (V1). Consisting of 91 neurons from 4 mainstream transgenic lines, this ground truth dataset provides

quantitative insight into the relationship between *in vivo* spiking activity and observed fluorescence signals, and will aid the interpretation of existing and future calcium imaging datasets.

Results

Dataset overview

To characterize the single-cell transfer function between observed fluorescence signals and underlying spikes *in vivo*, we performed simultaneous calcium imaging and cell-attached recordings in V1 L2/3 excitatory pyramidal neurons in anesthetized mice (**Figure 1A**, see Methods). The mice were from 4 transgenic lines, including 2 GCaMP6s-expressing lines, *Emx1-IRES-Cre;Camk2a-tTA;Ai94* (referred to as Emx1-s in this study) and *Camk2a-tTA;tetO-GCaMP6s* (tetO-s), and 2 GCaMP6f-expressing lines, *Emx1-IRES-Cre;Camk2a-tTA;Ai93* (Emx1-f) and *Cux2-CreERT2;Camk2a-tTA;Ai93* (Cux2-f) (**Table 1**). A total of 237 neurons, all of which had apparent baseline fluorescence which was clearly excluded from the nuclei, were randomly selected to record and image for spontaneous activity or visually-evoked responses. To directly compare our results to virally-expressed GCaMP6f and GCaMP6s (Chen et al., 2013), calcium imaging was performed at high optical zoom focused on individual cells (field of view of ~20 x 20 μm at ~158 frames per second; fps). We also patched and imaged a subset of these neurons at a lower zoom factor, *i.e.* one at which the responses of many neurons can be characterized in parallel (field of view of ~400 x 400 μm at ~30 fps), allowing direct comparison of calcium fluorescence responses acquired at higher spatiotemporal resolution with that more commonly employed by other studies including our Brain Observatory dataset (de Vries et al., 2019).

A large number (>550) of individual recording/imaging sessions, 2-4 min each, were obtained from these 237 cells (multiple sessions were collected for some cells in cases where the patch remained stable). Upon careful inspection of these sessions, we selected 91 cells with high-quality recording and imaging conditions from the 4 mouse lines (**Table 1**) for analysis, each with only one recording/imaging session for unbiased sampling.

Mouse line	Acronym	GECI	# Mice	# Cells			
				Spon	Visstim	Total	Low-zoom
Emx1-IRES-Cre;Camk2a-tTA;Ai94	Emx1-s	GCaMP6s	7	21	11	32	10
Camk2a-tTA;tetO-GCaMP6s	tetO-s	GCaMP6s	2	6	0	6	0

Emx1-IRES-Cre;Camk2a-tTA;Ai93	Emx1-f	GCaMP6f	6	9	17	26	10*
Cux2-CreERT2;Camk2a-tTA;Ai93	Cux2-f	GCaMP6f	7	4	23	27	17
Total			22	40	51	91	37

Table 1. Dataset. Sample size for each mouse line. Spon and Visstim: spontaneous and visually-evoked recordings, respectively. Total: total number of cells in high-zoom dataset. Low-zoom: cells imaged at both high- and low- optical zooms. *One Emx1-f cell was imaged at the same low optical zoom but with smaller field of view and at ~60 fps; see **Results**. The duration of the exclusion and summation windows used for each mouse line is described in **Methods** and **Figure 1B**.

Constructing single-cell spike-to-calcium fluorescence response curves

Throughout the manuscript, “spike” and “spiking activity” always refer to fast, presumably Na+ mediated, all-or-none electrical events. To delineate the relationship between spiking activity and GCaMP6 fluorescence, we focused on isolated spiking events whose calcium responses were well-separated from, and thus minimally contaminated by, those of adjacent spikes. A spiking event is defined as a group of spikes within a spike summation window (150 ms and 50 ms for GCaMP6s and GCaMP6f, respectively) with no spikes in the pre-event and post-event exclusion windows (150 ms and 50 ms pre and post respectively for GCaMP6s, and 50 ms both pre and post (after the last spike) for GCaMP6f) (**Figure 1B**). We summed the number of spikes within each spiking event, aligned fluorescence responses to the first spike within the event, and computed the peak fluorescence change (dF/F peak) during the calcium response window (200 ms for GCaMP6s, 50 ms or 75 ms for GCaMP6f in single-spike or multi-spike events) (**Figure 1B-D**; see **Methods** for spike exclusion, spike summation, and calcium response windows). Spiking events were binned based on the number of spikes within each event. Following this approach, we constructed single-cell spike-to-calcium fluorescence response curves (dF/F peak as a function of the number of spikes, with a minimum requirement of 3 events per bin; **Figure 2**). Events with >5 spikes were excluded from analysis due to the low frequency of such events (*e.g.*, ≥3 6-spike events were observed in 2 out of 32 Emx1-s cells, 0 out of 6 tetO-s cells, 5 out of 26 Emx1-f cells, and 1 out of 27 Cux2-f cells; also see **Figure 6C, i** below).

For calcium imaging data, we first examined the effect of fluorescence background subtraction on the peak dF/F signal. Following methods for virally expressed GCaMP6 (Chen et al., 2013), the background neuropil signal was estimated as the mean fluorescence of all pixels within 20-

133 μm from the cell center, excluding the selected cell, and neuropil subtraction was performed as:
134 $F_{corrected}(t) = F_{measured}(t) - r \times F_{neuropil}(t)$, where r was the neuropil contamination ratio (**Figure 2–**
135 **figure supplement 1**). Neuropil subtraction using a constant r-value for all cells (ranging from
136 0.1 to 0.7, where 0.7 was used for virally expressed GCaMP6) did not consistently decrease
137 within-cell variability of dF/F peak (*i.e.* variability across spiking events) (**Figure 2–figure**
138 **supplement 1D**). Given the potential for artificially boosting dF/F due to lower baseline
139 fluorescence (see example cell in **Figure 2–figure supplement 1B-C**), we chose not to perform
140 neuropil correction in this study (however, see **Figure 3–figure supplement 3** and **Figure 5–**
141 **figure supplement 3** below).
142
143 **Population spike-to-calcium fluorescence response curves**
144 To compare the relationship between spikes and fluorescence for the 4 mouse lines, population
145 response curves were constructed by averaging together responses of individual neurons
146 (**Figure 3A**). Spontaneous and visually-evoked responses were similar (**Figure 3–figure**
147 **supplement 1**) even though they were not recorded from the same cell and/or animal, and were
148 therefore pooled in the population responses. Data from mice anesthetized with isoflurane (n =
149 77 cells from 18 mice) and urethane (n = 14 cells from 4 mice; see **Methods**) were also pooled
150 as their response curves and firing rates were consistent with each other (**Figure 3–figure**
151 **supplement 2**).
152
153 Of the 4 mouse lines, tetO-s exhibited the largest single-spike fluorescence response (dF/F
154 peak: 13.6±3.7%, mean ± sd), followed by Emx1-s (7.8±2.6%), Cux2-f (6.0±1.6%), and Emx1-f
155 (4.5±1.6%) (**Figure 3A-D**). The response curve slopes (representing the average dF/F increase
156 per spike, computed from linear regression of dF/F peak against the number of spikes) of the 4
157 lines were comparable to each other, with tetO-s being somewhat higher (**Figure 3E**).
158
159 For a more direct comparison with that of viral GCaMP6 (Chen et al., 2013), we additionally
160 analyzed our data using a neuropil contamination ratio of 0.7 (**Figure 3–figure supplement 3**).
161 Aggregating spiking events across cells, mean dF/F peak in response to 5-spike events within
162 150 ms and 50 ms for GCaMP6s and GCaMp6f, respectively, was 135% for Emx1-s, 117% for
163 tetO-s (130% in response to 4-spike events), 130% for Emx1-f, and 93% for Cux2-f. In
164 comparison, Chen et al. reported ~200% and ~100% mean dF/F peak in response to 5-spike
165 events within 250 ms for viral GCaMP6s and GCaMP6f, respectively. Without neuropil

166 subtraction, mean dF/F peak in response to 5-spike events was 62% for Emx1-s, 73% for tetO-
167 s, 62% for Emx1-f, and 52% for Cux2-f (**Figure 3B**, bottom).
168
169 **Fluorescence response variability**
170 Within-cell variability of dF/F peak, as measured by the mean coefficient of variation across all
171 spiking events, was significantly larger in GCaMP6f-expressing cells compared to that in
172 GCaMP6s (**Figure 4A**). Likewise, within-cell variability for single-spike events was significantly
173 larger in GCaMP6f-expressing cells compared to that in GCaMP6s. In GCaMP6f-expressing
174 cells (but not GCaMP6s-expressing cells), single-spike events exhibited significantly larger
175 within-cell variability compared to multi-spike events (**Figure 4B**). Between-cell (cell-to-cell)
176 variability of dF/F peak was generally similar across the 4 lines (**Figure 4C**).
177
178 To determine the signal-to-noise ratio (SNR) and whether single-spike events could be
179 distinguished from zero-spike events (imaging noise), we first identified no-spike intervals of ≥1
180 s, separated by ≥4 s and ≥1 s from previous and subsequent spikes, respectively, for analysis
181 (**Figure 4–figure supplement 1A**). We then randomly sampled the corresponding calcium
182 fluorescence traces to obtain snippets of fluorescence with the same length as single-spike
183 events, and computed dF/F peak and the standard deviation of dF/F (an estimate of the noise
184 floor) using the same calcium response windows as single-spike events (50 ms and 200 ms for
185 GCaMP6f and GCaMp6s, respectively; **Figure 4–figure supplement 1B**). The mean single-
186 spike response was different from the mean imaging noise in 90% or 27 out of 30 Emx1-s cells
187 with at least 1 no-spike interval (out of 32 total cells), 100% or 4 out of 4 tetO-s cells with at least
188 1 no-spike interval (out of 6 total cells), 96% or 22 out of 23 Emx1-f cells with at least 1 no-spike
189 interval (out of 26 total cells), and 83% or 19 out of 23 Cux-f cells with least 1 no-spike interval
190 (out of 27 total cells) (**Figure 4–figure supplement 1C-E**).
191
192 Emx1-f and tetO-s cells exhibited the lowest and highest noise floor, respectively. Normalizing
193 the single-spike dF/F peak by the noise floor, the single-spike SNR for GCaMP6s was
194 significantly larger than that of GCaMP6f (**Figure 4–figure supplement 2A**), although a
195 difference in single-spike SNR was not observed between Emx1-s and tetO-s as in the single-
196 spike dF/F peak (**Figure 3D**). The population response curve expressed in SNR (**Figure 4–**
197 **figure supplement 2B**) largely resembled that expressed in dF/F peak (**Figure 3B**), with the
198 exception of tetO-s cells, whose SNR was lower due to a higher noise floor. However, tetO-s,

199 with a small sample size (n = 4 cells with at least 1 no-spike interval), also had few multi-spike
200 events (e.g. only 1 cell with 4-spike events).
201
202 **Spike detection**
203 We next examined spike detection at the single-trial level. Mean (± sd) detection rates were
204 63±25% and 94±11% for Emx1-s, 87±15% and 100% for tetO-s, 48±21% and 82±26% for
205 Emx1-f, and 40±27% and 74±22% for Cux2-f for single-spike and 2-spike events, respectively,
206 at 5% false positive rate (**Figure 5**). Spike detection rates differed substantially between cells,
207 from 0 to 100% for isolated spikes detected with GCaMP6s at 1% false positive rate (**Figure 5**).
208 Single-spike detection rate was ≥90% in 13% (4 of 30) Emx1-s cells, 75% (3 of 4) tetO-s cells,
209 4% (1 of 23) Emx1-f cells and 0% (0 of 22) Cux2-f cells, at 5% false positive rate with no
210 neuropil subtraction.
211
212 Previous studies have documented substantial contamination of somatic traces with
213 fluorescence from the surrounding neuropil, due to the extended nature of the microscope point
214 spread function. Neuropil contamination is often removed by subtracting a scaled version of the
215 neuropil fluorescence from the somatic fluorescence. The scale factor is often referred to as the
216 r-value and many studies employ the same r-value, often 0.7, for all neurons (Akerboom et al.,
217 2012). For each neuron, we found the optimum r-value, defined as the r-value which resulted in
218 the maximum number of detected spikes at the minimum false positive rate and calculated as
219 the r-value that maximized the area under the ROC curve. Optimal r-values differed greatly
220 between cells. Mean±sd optimal r-values were 0.39±0.30 for Emx1-s (range 0 to 0.8, with r<0.5
221 in 14 of 30 cells or 47%), 0.28±0.31 for tetO-s (range 0 to 0.7, with r<0.5 in 3 of 4 cells or 75%),
222 0.43±0.33 for Emx1-f (range 0 to 1, with r<0.5 in 13 of 22 cells or 59%), and 0.33±0.36 for Cux2-
223 f (range 0 to 1, with r<0.5 in 15 of 22 cells or 68%) (**Figure 5–figure supplement 1**). In the
224 Allen Brain Observatory, r-values were calculated individually for each neuron and follow similar
225 distributions to these optimal r-values. At optimal r-values, the single-spike detection rate at 5%
226 false positive rate was 76±21% (mean±sd) for Emx1-s, 92±11% for tetO-s, 52±23% for Emx1-f,
227 and 40±27% for Cux2-f.
228
229 **Spiking characteristics**
230 Spontaneous or visually evoked firing rates of individual cells, computed from the total number
231 of spikes detected during a recording, were similar across mouse lines (Emx1-s: 1.17±0.71 Hz,
232 mean±sd; tetO-s: 0.96±0.40; Emx1-f: 1.52±1.34; Cux2-f: 1.47±1.43; referred to as spikes

233 detected in **Figure 6A**). After preprocessing (selecting isolated spiking events with ≤5 spikes
234 within spike summation windows of 150 ms and 50 ms for GCaMP6s and GCaMP6f,
235 respectively, and excluding atypical spiking events - see **Methods**), 70-78% of detected spikes
236 were included in the response curve analysis (**Figure 6B**; referred to as spikes analyzed in
237 **Figure 6A**). For all mouse lines, >50% of analyzed spikes occurred in multi-spike events and
238 not as isolated single spikes (**Figure 6C**; see **Figure 6–figure supplement 1** for a comparison
239 of spiking characteristics between spontaneous and visually-evoked cells). The fraction of
240 single-spike events within spike summation windows was largest in Cux2-f (48%), followed by
241 tetO-s (37%), Emx1-s (32%), and Emx1-f (23%).
242
243 In multi-spike events, we assessed whether variability in dF/F peak could be attributed to the
244 mean inter-spike interval (ISI) within the spiking event (here, the maximum ISI corresponded to
245 the spike summation window: 150 ms and 50 ms for GCaMP6s and GCaMP6f, respectively).
246 Pooling multi-spike events from all cells, the mean ISI was negatively correlated with dF/F but
247 explained little of its variance (R^2 ranging from 0.03 to 0.21; **Figure 6–figure supplement 2**).
248 Breaking this down into individual cells, we found a small fraction of cells having spiking events
249 with significant correlation coefficients of dF/F peak and mean ISI (**Figure 6–figure**
250 **supplement 3A-B**). Within these cells, a negative correlation between mean ISI and dF/F peak
251 was observed, and for 4- and 5-spike events, R^2 ranged from ~0.4 to 0.9 (**Figure 6–figure**
252 **supplement 3C**), likely contributing to the overall negative correlation at the population level.
253
254 **Comparing spike-to-calcium fluorescence response curves imaged at high and low**
255 **spatiotemporal resolutions**
256 For a subset of cells, we performed calcium imaging at both high zoom (focusing on individual
257 cells; ~20 × 20 μm field of view at 0.2 μm per pixel and ~158 fps) and low zoom (population
258 imaging; ~400 x 400 μm field of view at 0.8 μm per pixel and ~30 fps, with 1 Emx1-f cell at 1.6
259 μm per pixel and ~60 fps) (**Table 1, Figure 7A**), enabling a direct comparison of their spike-to-
260 calcium fluorescence response curves (**Figure 7B**). For Emx1-f and Cux2-f, the slope of the
261 spike-to-calcium fluorescence response curve was significantly smaller when imaged at lower
262 spatiotemporal resolution, despite no detected difference in mean firing rates between groups
263 (**Figure 7C-D; Figure 7–figure supplement 1A**). A similar trend was observed in Emx1-s cells
264 ($p = 0.05$, paired t-test) between response curve slopes at high- and low-zoom. Across all 3
265 lines, imaging at low-zoom consistently resulted in higher variability of dF/F peak, both within
266 individual cells (**Figure 7B, Figure 7–figure supplement 1B-C**) and between cells (**Figure 7–**

267 **figure supplement 1D**). Furthermore, the mean single-spike response was different from the
268 mean imaging noise in fewer cells at low-zoom compared to that at high-zoom for all mouse
269 lines (**Figure 7–figure supplement 1E**), with higher noise floors and lower SNRs (**Figure 7–**
270 **figure supplement 1F**).
271
272 At the single-trial level, detection of single-spikes was 15±12% for Emx1-s (compared to
273 52±28% for the same cells imaged at high-zoom), 15±8% for Emx1-f (47±18% at high-zoom),
274 and 9±7% for Cux2-f (42±26% at high-zoom) at 5% false positive rate (**Figure 7–figure**
275 **supplement 2**). For 2-spike events, detection rate at low-zoom was 36±22% for Emx1-s
276 (compared to 92±16% for the same cells imaged at high-zoom), 16±11% for Emx1-f (86±23% at
277 high-zoom), and 19±12% for Cux-f (73±24% at high-zoom) at 5% false positive rate.
278
279
280 **Discussion**
281 Calcium imaging is widely used to report neuronal spiking activity *in vivo*. However, accurate
282 spike inference from calcium imaging remains a challenge, and there are relatively few ground
283 truth datasets with simultaneous calcium imaging and electrophysiology to aid the development
284 of more accurate spike inference algorithms. In a recent challenge (Spike Finder;
285 http://spikefinder.codeneuro.org/) (Berens et al., 2018), ~40 algorithms were trained and tested
286 on datasets consisting of 37 GCaMP6-expressing cells, underscoring the need for additional
287 GCaMP6 calibration data. In addition to supporting efforts toward spike inference, an improved
288 understanding of the relationship between spiking and observed fluorescence signals is
289 necessary to further broaden the utility and impact of calcium imaging. To these ends, we
290 contribute a ground truth dataset consisting of 91 V1 L2/3 excitatory neurons recorded at single-
291 cell resolution (available at https://brain-map.org/explore/circuits/oephys), and characterized
292 their spike-to-calcium fluorescence transfer function. Complementing existing datasets with viral
293 GECI expression (Chen et al., 2013; Theis et al., 2016; Dana et al., 2016), our work facilitates
294 interpretation of existing and future calcium imaging studies using mainstream transgenic
295 mouse lines, such as the Allen Institute's Brain Observatory Visual Coding dataset
296 (http://observatory.brain-map.org/visualcoding) (de Vries et al., 2019).
297
298 We used the same ROI extraction and neuropil correction methods as Chen et al. to compare
299 our results in transgenic mice to that of virally expressed GCaMP6 (Chen et al., 2013). To avoid
300 contamination of fluorescence responses by subsequent spiking events, we used spike

301 summation windows of 150 ms and 50 ms for GCaMP6s and GCaMP6f, respectively, instead of
302 250 ms as reported in Chen et al., a change that was necessitated by differences in firing rates
303 between our dataset and theirs. Choosing a constant neuropil contamination ratio of 0.7, dF/F
304 magnitudes of Emx1-f and Cux2-f were comparable to that of viral GCaMP6f (~100% for 5-spike
305 events in viral GCaMP6f), whereas Emx1-s and tetO-s were lower than that of viral GCaMP6s
306 (~200% for 5-spike events in viral GCaMP6s). To avoid potential confounding factors
307 associated with choosing a constant neuropil contamination ratio for all cells (e.g. increased
308 within-cell variability and artificially boosted dF/F magnitude), we chose not to perform neuropil
309 subtraction in subsequent analyses, except to quantify the relationship between neuropil
310 contamination ratio and the efficiency of detecting single-trial spiking events (discussed below).
311
312 Consistent with viral GCaMP6 (Chen et al., 2013), GCaMP6s cells in transgenic mice produced
313 a larger single-spike dF/F peak compared to that of GCaMP6f. Additionally, GCaMP6s cells
314 exhibited lower within-cell variability compared to GCaMP6f, both for single-spikes and multi-
315 spike events. Interestingly, tetO-s cells exhibited a larger single-spike dF/F peak and a larger
316 response curve slope compared to that of Emx1-s (albeit with a small sample size of n=6 cells).
317 However, tetO-s cells exhibited a higher noise floor (standard deviation of dF/F during no-spike
318 intervals) compared to that of Exm1-s, such that no difference in single-spike SNR was detected
319 between the two GCaMP6s lines. For multi-spike events, SNR for tetO-s was lower than that of
320 Emx1-s and Emx1-f, which could be attributed to a smaller sample size and fewer mutli-spike
321 events in the tetO-s cells with no-spike intervals (n=4 cells with no-spike intervals out of 6 total
322 cells). Nevertheless, with the large dF/F response shown here and previous work showing
323 comparable expression levels to Emx1-s and no observed epileptiform events (Steinmetz et al.,
324 2017), the tetO-s line represents an alternative to Emx1-s for expressing GCaMP6s.
325
326 Spike detection at the single-trial level was the most accurate in tetO-s, with mean single-spike
327 detection rate of 87±15% (mean±sd) at 5% false positive rate, or 85±16% at 1% false positive
328 rate. Thus, single spikes could be reliably detected in transgenic mouse lines expressing
329 GCaMP6s, and in tetO-s the detection rate was not far short of that reported for viral GCaMP6s
330 (99% at 1% false positive rate) (Chen et al., 2013). The difference in reported values could be
331 partially attributed to the different spike exclusion windows used: 150/200 ms (pre and post) in
332 this study and 1 s in Chen et al. Although spike detection rates were variable across cells, we
333 observed single-spike detection rates of ≥90% in a number of neurons (including 1 Emx1-f cell),
334 thus validating our experimental setup. Additionally, we found that spike detection rate, similar

335 to dF/F magnitude, was dependent on the neuropil contamination ratio r. We found optimal r-
336 values for maximizing single-spike detection rates in individual cells, by maximizing the area
337 under the ROC curve for classifying spiking events. The effect of optimizing r-values on spike
338 detection rate was most pronounced in Emx1-s, with an increase of 36±62% per cell compared
339 to r = 0 (at 5% false positive rate; the corresponding mean single-spike detection rate increased
340 from 63±25% to 76±20%). These data highlight the potential benefit of tuning r-values for
341 individual neurons, as was done for our Brain Observatory dataset (de Vries et al., 2019), over
342 the traditional method of selecting a constant r-value for the population. However, it is worth
343 noting that increased r-values also typically resulted in increased within-cell variability of dF/F
344 peak and increased dF/F magnitude, as discussed above.
345
346 2-p calcium imaging of large neuronal populations in parallel is typically performed on the order
347 of 30 fps, providing a small fraction (~20% in this study) of the number of temporal samples of
348 that imaged at high-zoom. Additionally, assuming an 8 × 8 μm square cell, the number of pixels
349 at low-zoom was 6.25% ($10^2/40^2$) of that at high-zoom. Coupled with differences in the number
350 of samples per pixel (13 or 18 at high-zoom and 3 at low-zoom), the number of spatial samples
351 at low-zoom was ~1.5% of that at high-zoom. Not surprisingly, we observed reduced response
352 curve slope, higher variability, and lower SNR when imaging at the lower spatiotemporal
353 resolution. Furthermore, single-spike detection rates (at 5% false positive rate) dropped from
354 ~40-50% when imaged at high-zoom to ~10-15% at low-zoom in the same cells. At low-zoom,
355 detection of 2-spike events remained unreliable: <40% for Emx1-s and <20% for Emx1-f and
356 Cux2-f. Due to the nature of our experiments (e.g. acute surgical preparation, with brain motion
357 controlled through agarose, and the presence of a recording micropipette), our low-zoom data
358 may be underestimating the data quality for a typical population imaging session (a chronic,
359 more stable preparation with a coverslip cranial window, no micropipette and associated tissue
360 damage). Nevertheless, these data from the same cells offer a point of comparison between
361 data acquired under high-zoom and population imaging conditions.
362
363 In summary, in this study we present a ground truth dataset with simultaneous electrophysiology
364 and calcium imaging. This unique dataset allowed us to systematically compare multiple
365 aspects of spike-response properties between different GECIs (GCaMP6s and GCaMP6f) and
366 among several transgenic lines. For example, we found that GCaMP6s and GCaMP6f cells
367 have similar spike-fluorescence response curves, but GCaMP6f cells have greater variability
368 and lower single-spike detection probability than GCaMP6s cells. While both GCaMP6f and

369 GCaMP6s cells have reasonably good SNR and single-spike detectability at high-zoom
370 conditions, both types of properties deteriorate substantially under the commonly used low-
371 zoom population imaging conditions, with single-spike detection probabilities dropping to 4-20%
372 (at 1-10% false positive rates). This finding has important implications for the interpretation of
373 many large-scale in vivo calcium imaging datasets. By making our data freely available, we
374 hope that it will serve the community as a further resource to better understand quantitatively
375 the link between calcium-evoked fluorescent imaging signals and spiking activity.
376
377
378 **Materials and Methods**
379 Experimental procedures were in accordance with NIH guidelines and approved by the
380 Institutional Animal Care and Use Committee (IACUC) of the Allen Institute for Brain Science.
381 All experiments were conducted at the Allen Institute for Brain Science.
382
383 **Mice.** 2-p-targeted electrophysiology and 2-p calcium imaging was conducted in adult
384 transgenic mice (2-6 months old, both sexes, n = 22), including *Emx1-IRES-Cre;Camk2a-*
385 *tTA;Ai94* (simplified as Emx1-s, n = 7), *Camk2a-tTA;tetO-GCaMP6s* (Wekselblatt et al., 2016)
386 (simplified as tetO-s, n = 2), *Emx1-IRES-Cre;Camk2a-tTA;Ai93* (simplified as Emx1-f, n = 6),
387 and *CreERT2;Camk2a-tTA;Ai93* (simplified as Cux2-f, n = 7). Ai93 and Ai94 containing mice
388 (Madisen et al., 2015) included in this dataset did not show behavioral signs for epileptic brain
389 activity (Steinmetz et al., 2017).
390
391 **Surgery.** Mice were anesthetized with either isoflurane (0.75-1.5% in O_2) or urethane (1.5 g/kg,
392 30% aqueous solution, intraperitoneal injection), then implanted with a metal head-post. A
393 circular craniotomy was performed with skull thinning over the left V1 centering on 1.3 mm
394 anterior and 2.6 mm lateral to the Lambda. During surgery, the craniotomy was filled with
395 Artificial Cerebrospinal Fluid (ACSF) containing (in mM): NaCl 126, KCl 2.5, NaH_2PO_4 1.25,
396 $MgCl_2$ 1, $NaHCO_3$ 26, glucose 10, $CaCl_2$ 2, in ddH_2O; 290 mOsm; pH was adjusted to 7.3 with
397 NaOH to keep the exposed V1 region from overheating or drying. Durotomy was performed to
398 expose V1 regions of interest that were free of major blood vessels to facilitate the penetration
399 of recording micropipettes. A thin layer of low melting-point agarose (1-1.3% in ACSF, Sigma-
400 Aldrich) was then applied to the craniotomy to control brain motion. The mouse body
401 temperature was maintained at 37°C with a feedback-controlled animal heating pad (Harvard
402 Apparatus).

403

404 **Calcium imaging.** Individual GCaMP6+ neurons within ~100-300 μm underneath the pial
405 surface were visualized under adequate depth of anesthesia (Stage III-3) using a Bruker
406 (Prairie) 2-p microscope with 8 kHz resonant-galvo scanners, coupled with a Chameleon Ultra II
407 Ti:sapphire laser system (Coherent). Fluorescence excited at 920 nm wavelength, with <70 mW
408 laser power measured after the objective, was collected in two spectral channels using green
409 (510/42 nm) and red (641/75 nm) emission filters (Semrock) to visualize GCaMP6 and the Alexa
410 Fluor 594-containing micropipette, respectively. Fluorescence images were acquired at various
411 framerates (117.6-158.3 fps) through a 16x water-immersion objective lens (Nikon, NA 0.8), with
412 or without visual stimulations.
413

414 **Electrophysiology.** 2-p targeted cell-attached recording was performed following established
415 protocols (Margrie et al., 2003; Kitamura et al., 2008; Knoblich et al., 2019). Long-shank
416 borosilicate (KG-33, King Precision Glass) micropipettes (5-10 MΩ) were pulled with a P-97
417 puller (Sutter) and filled with ACSF and Alexa Fluor 594 to perform cell-attached recordings on
418 GCaMP6+ neurons. Micropipettes were installed on a MultiClamp 700B headstage (Molecular
419 Devices), which was mounted onto a Patchstar micromanipulator (Scientifica) with an
420 approaching angle of 31 degrees from horizontal plane. Minimal seal resistance was 20 MΩ.
421 Data were acquired under "I = 0" mode (zero current injection) with a Multiclamp 700B, recorded
422 at 40 kHz using Multifunction I/O Devices (National Instruments) and custom software written in
423 LabVIEW (National Instruments) and MATLAB (MathWorks). Isoflurane level was intentionally
424 adjusted during recording sessions to keep the anesthesia depth as light as possible, resulting
425 in fluctuation of the firing rates of recorded neurons.
426

427 **Visual stimulation.** Whole-screen sinusoidal static and drifting gratings were presented on a
428 calibrated LCD monitor spanning 60° in elevation and 130° in azimuth to the contralateral eye.
429 The mouse's eye was positioned ~22 cm away from the center of the monitor. For static
430 gratings, the stimulus consisted of 4 orientations (45° increment), 4 spatial frequencies (0.02,
431 0.04, 0.08, and 0.16 cycles per degree), and 4 phases (0, 0.25, 0.5, 0.75) at 80% contrast in a
432 random sequence with 10 repetitions. Each static grating was presented for 0.25 seconds, with
433 no inter-stimulus interval. A gray screen at mean illuminance was presented randomly a total of
434 60 times. For drifting gratings, the stimulus consisted of 8 orientations (45° increment), 4 spatial
435 frequency (0.02, 0.04, 0.08, and 0.16 cycle per degree) and 1 temporal frequency (2 Hz), at
436 80% contrast in a random sequence with up to 5 repetitions. Each drifting grating lasted for 2

437 seconds with an inter-stimulus interval of 2 seconds. A gray screen at mean illuminance was
438 presented randomly for up to 15 times.
439
440 **Data analysis.** Electrophysiology and calcium imaging data were analyzed using custom
441 MATLAB and Python scripts. For electrophysiology, Vm was filtered between 250 Hz and 5 kHz,
442 and automated spike detection was performed using a threshold criterion (5×std of Vm).
443 Unusually prolonged transient increases in calcium fluorescence were excluded from analysis
444 with an adaptive Vm threshold: $0.25\times V_{baseline}+(V_{peak}-V_{baseline})$, where $V_{baseline}$ was the mean Vm
445 over 2 ms before the first spike of the spiking event and V_{peak} was the amplitude of the first spike
446 of the spiking event. The cumulative time above this Vm threshold was compared against a time
447 threshold (6 ms × the number of spikes within the spiking event), and the fluorescence
448 associated with spiking events with cumulative time above the time threshold were not
449 analyzed.
450
451 For calcium imaging, in-plane motion artifacts were corrected (Dombeck et al., 2007), and
452 cell/region of interest (ROI) selection was performed using a semi-automatic algorithm (Chen et
453 al., 2013) (kindly provided by Karel Svoboda, Janelia Research Campus). Ring-shaped ROIs
454 were used to select GCaMP6+ excitatory neurons, with GCaMP6 expression typically excluded
455 from the nucleus and restricted to the cytoplasm. For 2 of 91 cells, a satisfactory ring-shaped
456 ROI could not be found automatically, and a circular ROI covering the entire soma was used
457 instead. Calcium imaging data was smoothed using a local regression method using weighted
458 linear least squares and a 1st degree polynomial model (built-in “smooth” function in MATLAB
459 with “rlowess” method), and an averaging window of 5 frames.
460
461 To construct spike-calcium fluorescence response curves, we first identified all isolated spiking
462 events regardless of the number of spikes they contained. For GCaMP6s, spiking events were
463 separated from previous and subsequent spiking events by >150 ms and >50 ms, respectively,
464 due to the higher firing rates in our study. The 50 ms post-event exclusion window ensured the
465 absence of additional spikes in the calcium response window of 200 ms. For GCaMP6f, spiking
466 events were separated from previous and subsequent spiking events by >50 ms. Within each
467 spiking event, spike summation windows were 150 ms and 50 ms for GCaMP6s and GCaMP6f,
468 respectively, chosen based on the rise time of the GECIs (Chen et al., 2013). To determine
469 spike-triggered calcium fluorescence responses, fluorescence traces were aligned to the first
470 spike in each spiking event. The change in fluorescence, dF/F, for each spiking event was

calculated as $(F-F_0)/F_0$, where F_0 was computed locally as the mean fluorescence over 50 ms and 20 ms before the first spike for GCaMP6s and GCaMP6f, respectively. For GCaMP6s, peak dF/F was found within 200 ms after the first spike. For GCaMP6f, a variable calcium response window was used: peak dF/F was found within 50 ms and 75 ms after the first spike for single-spike and multi-spike events, respectively. Bursts of >5 spikes were excluded from analysis due to the low frequency of such events. Alignment jitter intrinsic to the imaging frame rate was ≤6.3 ms in 86 of 91 cells (imaged at 158.7 fps), with mean expected error of 3.2 ms, and between ≤5.6 ms to ≤8.5 ms in others (imaged at 117.6-178.6 fps), with mean expected error of 2.8 to 4.3 ms.

To compare mean single-spike dF/F peak responses to imaging noise, imaging noise was estimated as follows: we found no-spike intervals of ≥1 s, separated by ≥4 s and ≥1 s from previous and subsequent spikes, respectively, and randomly sampled the m corresponding calcium fluorescence traces n times, where $m \times n$ approximately matched the number of single-spike events. Noise (zero-spike) dF/F peak was computed using the same calcium response windows as single-spike events (50 ms and 200 ms for GCaMP6f and GCaMp6s, respectively).

To quantify the efficiency of detecting spiking events at a single-trial level, we compared dF/F traces of the response (single-spike events or 2-spike events) to that of imaging noise. To estimate imaging noise, calcium fluorescence traces during no-spike intervals were randomly sampled as described above, for a minimum of 25 times. For each cell, the mean response trace was used as the template vector. The template vector was normalized after subtracting the mean to create the unit vector, and the scalar results of projecting the response and noise traces on the unit vector were computed: r_i and n_i for response and noise scalars, respectively. The detection threshold was defined as the xth percentile of n_i values, where 1-x represented the false positive rate (e.g., x = 95 for 95th percentile or 5% false positive rate), and the detection rate (true positive rate) was the fraction of r_i values above the detection threshold.

Acknowledgements

We are grateful for the Animal Care, Transgenic Colony Management, and Lab Animal Services teams for mouse husbandry, and Carol Thompson and John Phillips for providing project management support. We thank Karel Svoboda, Hod Dana and Tsai-Wen Chen for sharing analysis software. This work was also supported by grants from National Natural Science Foundation of China (NSFC31871055) and Guangdong Science and Technology Department

(2017B030314026) to L.L. We thank the Allen Institute founders, Paul G. Allen and Jody Allen, for their vision, encouragement, and support.

Author contributions

CK, HZ, RCR conceived the project. LL, UK, LH performed experiments. LL, LH, UK, PL, JW, JL, SEJdV, MAB, GJM analyzed data. LL, HZ, CK, RCR, MAB, GJM supervised the project. LH wrote the paper with input from all co-authors.

Competing interests

The authors declare no competing financial interests.

Data availability

The dataset is available at https://brain-map.org/explore/circuits/oephys.

References

Akerboom, J., Chen, T.-W., Wardill, T.J., Tian, L., Marvin, J.S., Mutlu, S., Calderón, N.C., Esposti, F., Borghuis, B.G., Sun, X.R., et al. (2012). Optimization of a GCaMP calcium indicator for neural activity imaging. J. Neurosci. Off. J. Soc. Neurosci. *32*, 13819–13840.

Berens, P., Freeman, J., Deneux, T., Chenkov, N., McColgan, T., Speiser, A., Macke, J.H., Turaga, S.C., Mineault, P., Rupprecht, P., et al. (2018). Community-based benchmarking improves spike rate inference from two-photon calcium imaging data. PLoS Comput. Biol. *14*, e1006157.

Chen, T.-W., Wardill, T.J., Sun, Y., Pulver, S.R., Renninger, S.L., Baohan, A., Schreiter, E.R., Kerr, R.A., Orger, M.B., Jayaraman, V., et al. (2013). Ultrasensitive fluorescent proteins for imaging neuronal activity. Nature *499*, 295–300.

Daigle, T.L., Madisen, L., Hage, T.A., Valley, M.T., Knoblich, U., Larsen, R.S., Takeno, M.M., Huang, L., Gu, H., Larsen, R., et al. (2018). A Suite of Transgenic Driver and Reporter Mouse Lines with Enhanced Brain-Cell-Type Targeting and Functionality. Cell *174*, 465-480.e22.

Dana, H., Mohar, B., Sun, Y., Narayan, S., Gordus, A., Hasseman, J.P., Tsegaye, G., Holt, G.T., Hu, A., Walpita, D., et al. (2016). Sensitive red protein calcium indicators for imaging neural activity. ELife *5*.

Dombeck, D.A., Khabbaz, A.N., Collman, F., Adelman, T.L., and Tank, D.W. (2007). Imaging large-scale neural activity with cellular resolution in awake, mobile mice. Neuron *56*, 43–57.

Kitamura, K., Judkewitz, B., Kano, M., Denk, W., and Häusser, M. (2008). Targeted patch-clamp recordings and single-cell electroporation of unlabeled neurons in vivo. Nat. Methods *5*, 61–67.

Knoblich, U., Huang, L., Zeng, H., and Li, L. (2019). Neuronal cell-subtype specificity of neural synchronization in mouse primary visual cortex. Nat. Commun. *10*, 2533.

542 Luo, L., Callaway, E.M., and Svoboda, K. (2018). Genetic Dissection of Neural Circuits: A
543 Decade of Progress. Neuron *98*, 256–281.

544 Madisen, L., Garner, A.R., Shimaoka, D., Chuong, A.S., Klapoetke, N.C., Li, L., van der Bourg,
545 A., Niino, Y., Egolf, L., Monetti, C., et al. (2015). Transgenic mice for intersectional targeting of
546 neural sensors and effectors with high specificity and performance. Neuron *85*, 942–958.

547 Margrie, T.W., Meyer, A.H., Caputi, A., Monyer, H., Hasan, M.T., Schaefer, A.T., Denk, W., and
548 Brecht, M. (2003). Targeted whole-cell recordings in the mammalian brain in vivo. Neuron *39*,
549 911–918.

550 Nadim, F., and Bucher, D. (2014). Neuromodulation of neurons and synapses. Curr. Opin.
551 Neurobiol. *29*, 48–56.

552 Rose, T., Goltstein, P.M., Portugues, R., and Griesbeck, O. (2014). Putting a finishing touch on
553 GECIs. Front. Mol. Neurosci. *7*, 88.

554 Steinmetz, N.A., Buetfering, C., Lecoq, J., Lee, C.R., Peters, A.J., Jacobs, E.A.K., Coen, P.,
555 Ollerenshaw, D.R., Valley, M.T., de Vries, S.E.J., et al. (2017). Aberrant Cortical Activity in
556 Multiple GCaMP6-Expressing Transgenic Mouse Lines. ENeuro *4*.

557 Theis, L., Berens, P., Froudarakis, E., Reimer, J., Román Rosón, M., Baden, T., Euler, T.,
558 Tolias, A.S., and Bethge, M. (2016). Benchmarking Spike Rate Inference in Population Calcium
559 Imaging. Neuron *90*, 471–482.

560 de Vries, S.E.J., Lecoq, J., Buice, M.A., Groblewski, P.A., Ocker, G.K., Oliver, M., Feng, D.,
561 Cain, N., Ledochowitsch, P., Millman, D., et al. (2019). A large-scale, standardized physiological
562 survey reveals higher order coding throughout the mouse visual cortex. Nat. Neurosci. *In press*.

563 Wekselblatt, J.B., Flister, E.D., Piscopo, D.M., and Niell, C.M. (2016). Large-scale imaging of
564 cortical dynamics during sensory perception and behavior. J. Neurophysiol. *115*, 2852–2866.

565

566

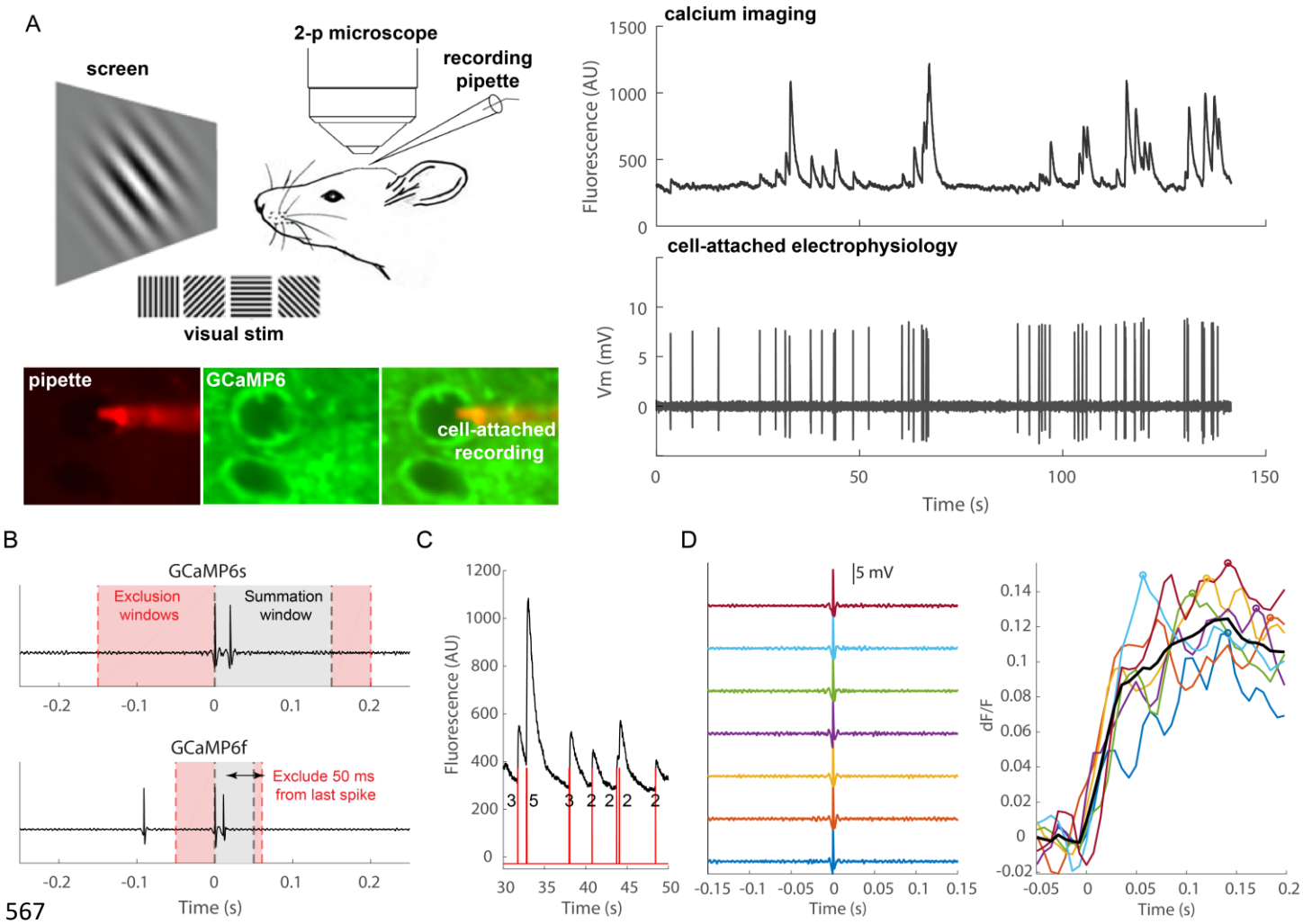


Figure 1. Simultaneous calcium imaging and electrophysiology *in vivo*. (**A**) Experimental design and example fluorescence and Vm traces from an example Emx1-s neuron. (**B**) Identification of spiking events. Spike-to-calcium fluorescence analysis was performed for spiking events that were separated from previous and next spikes by durations of the exclusion windows (red shading) or longer. The summation window (gray shading) determined the number of spikes within each spiking event. (**C**) Spike-triggered fluorescence responses. Red lines indicate spikes and numbers indicate the number of spikes in each spiking event. (**D**) Fluorescence changes (dF/F) in response to representative single-spike events from the recorded cell shown in (A). For GCaMP6s, the spike summation window (left panel) and calcium response window (right panel) were 150 ms and 200 ms, respectively. Colors indicate correspondence between spikes and dF/F. Black line indicates mean dF/F. Circles indicate peak fluorescence changes (dF/F peak) within the calcium response window.

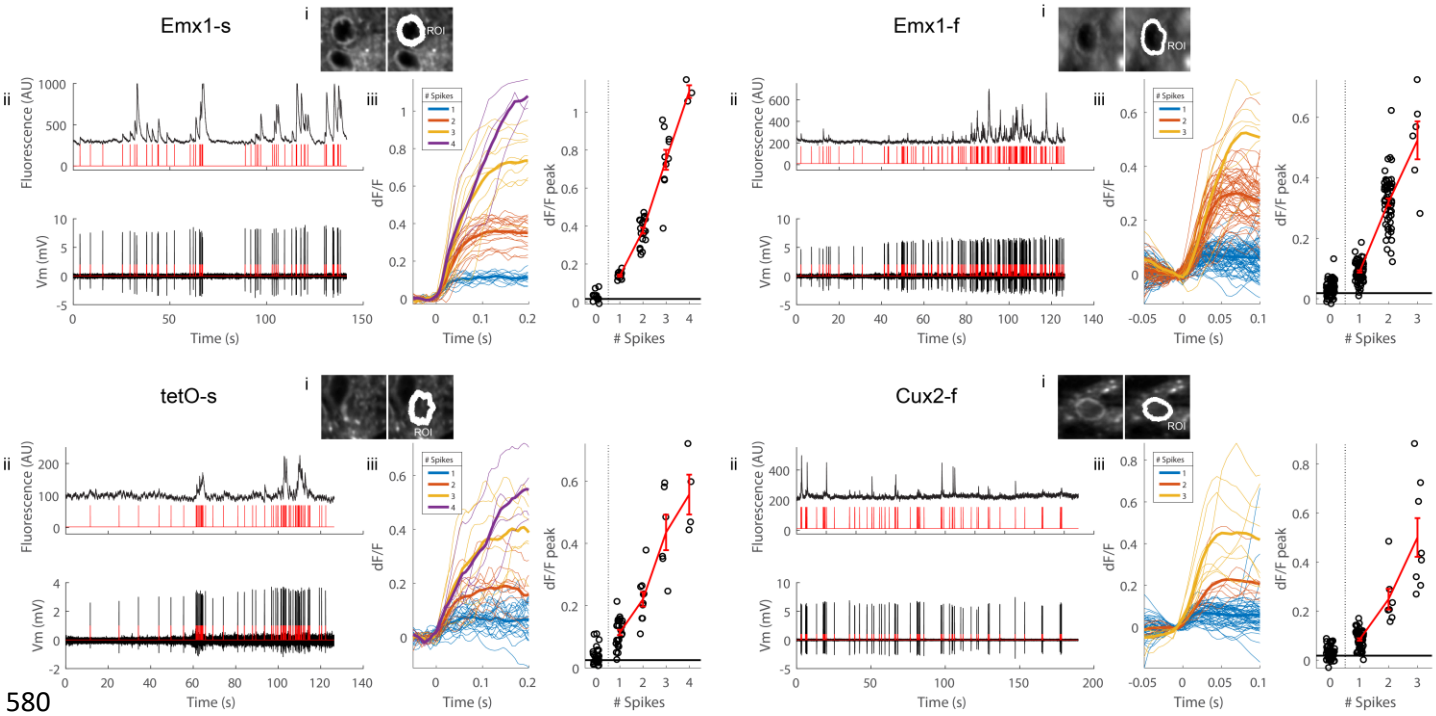


Figure 2. Example single-cell spike-to-calcium fluorescence response curves. For each type of mice: (i) Cell ROI and segmentation mask. (ii) Corresponding fluorescence (top) and Vm traces (bottom). Red lines indicate detected spikes. (iii) dF/F traces (left) and spike-to-calcium fluorescence response curve (dF/F peak as a function of the number of spikes, with a minimum requirement of 3 events per bin; right). Horizontal black lines indicate the standard deviation of dF/F during no-spike intervals (i.e. estimated noise floor; see **Figure 4–figure supplement 2**). Estimated zero-spike dF/F peak were plotted for comparison (see **Figure 4-figure supplement 1**). Spike summation windows were 150 ms and 50 ms for GCaMP6s and GCaMP6f, respectively.

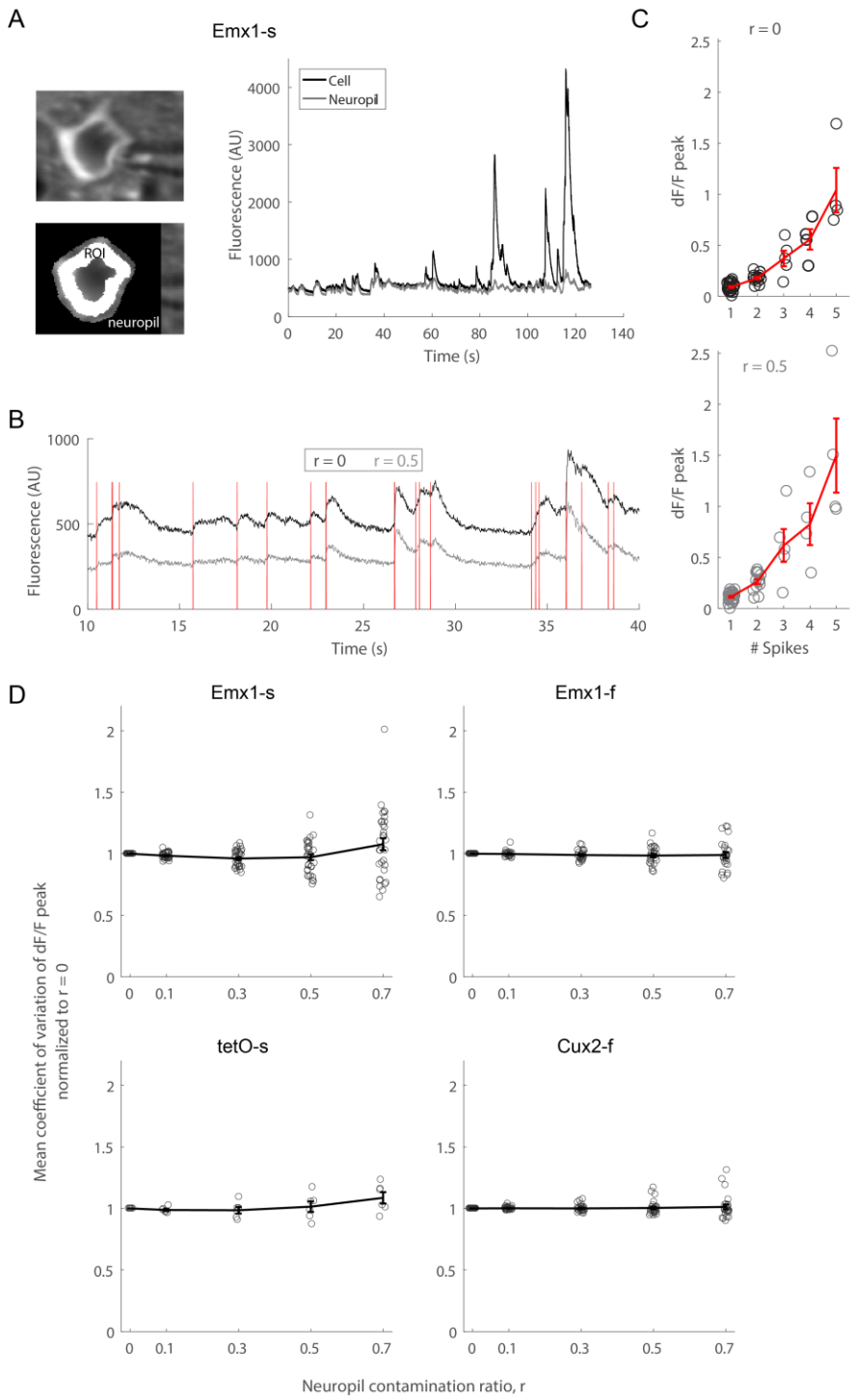


Figure 2–figure supplement 1. Effect of neuropil subtraction on within-cell variability of dF/F peak. (**A**) An example Emx1-s neuron, with labeled regions where cell and neuropil fluorescence were measured, and the corresponding mean fluorescence traces. (**B**) Comparison of cell fluorescence without and with neuropil correction, where the contamination ratio r = 0.5 was chosen. Fluorescence changes related to spikes (red lines) were also subtracted. (**C**) Comparison of single-cell response curve without and with neuropil correction (r = 0.5). Red lines indicate mean ± sem. (**D**) Mean coefficient of variation of dF/F peaks as a function of r, normalized to no neuropil subtraction (r = 0), for all mouse lines. Circles indicate individual cells. Black crosses represent the Emx1-s cell shown in (A).

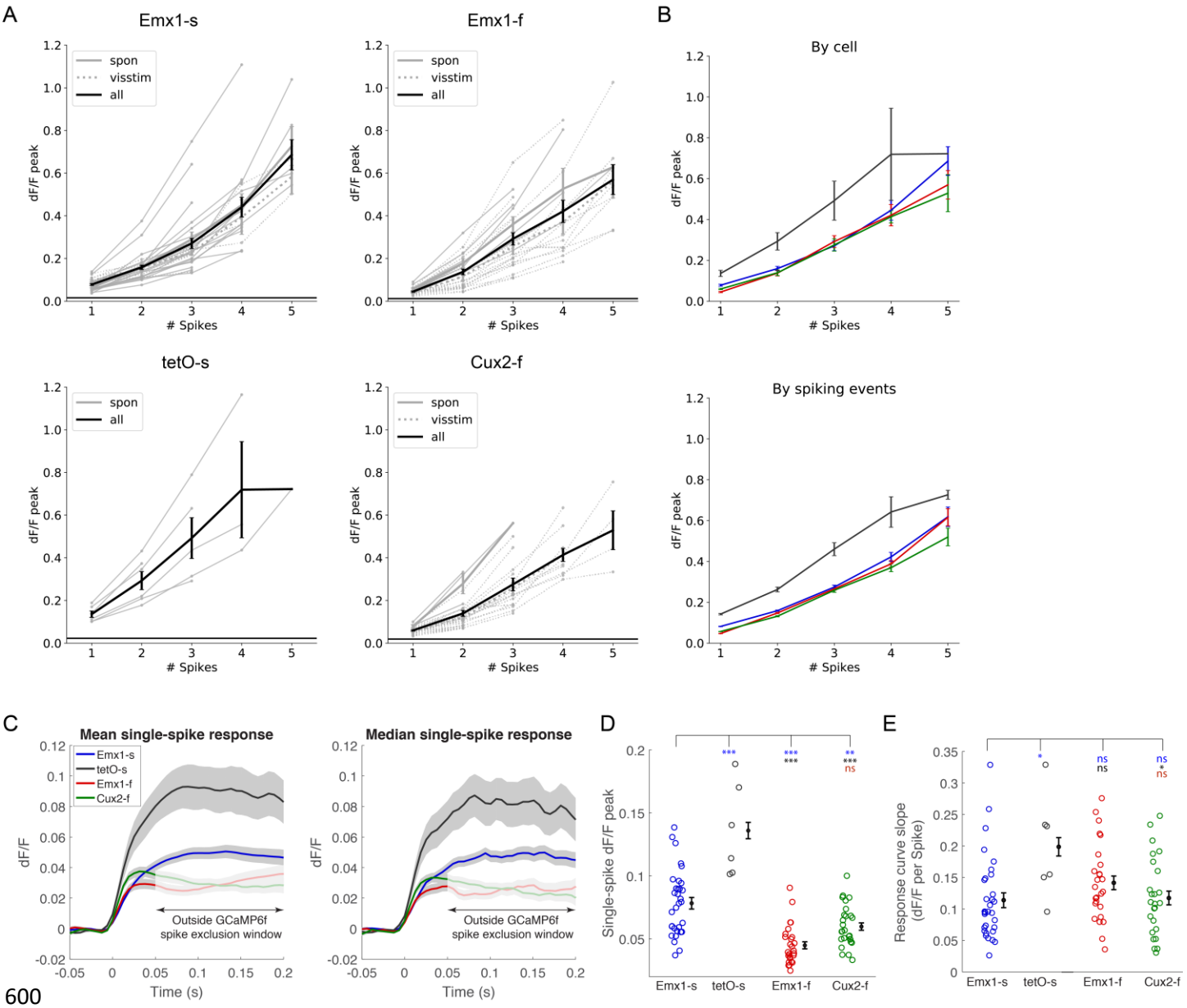


Figure 3. Population response. (**A**) Population spike-to-calcium fluorescence response curves. Each line represents one cell. Bold lines indicate mean responses for spontaneous, visually-evoked, and all recordings. Horizontal black lines indicate the mean standard deviation of dF/F during no-spike intervals (i.e. mean noise floor; see **Figure 4–figure supplement 2**). (**B**) Overlay of population response curves. By cell (top): population means from (A), computed from the mean response of individual cells. Emx1-s: n = 32 cells; tetO-s: n = 6; Emx1-f: n = 26; Cux2-f: n = 27. By spiking events (bottom): population response computed from spiking events pooled from all cells as in Chen et al., 2013. Emx1-s: n = 1160, 599, 255, 103, and 52 spiking events for 1-5 spikes; tetO-s: n = 250, 103, 48, 17, and 6; Emx1-f: n = 1204, 921, 416, 157, and 60; Cux2-f: n = 2667, 809, 209, 118, and 46. (**C**) Mean and median dF/F in response to single-spike events. For GCaMP6f, the single-spike calcium response window was 50 ms. Shading corresponds to 2×sem. (**D**) Mean single-spike dF/F peak of each cell. (**E**) Response curve slope of each cell, computed from linear regression of dF/F peaks against the numbers of spikes.

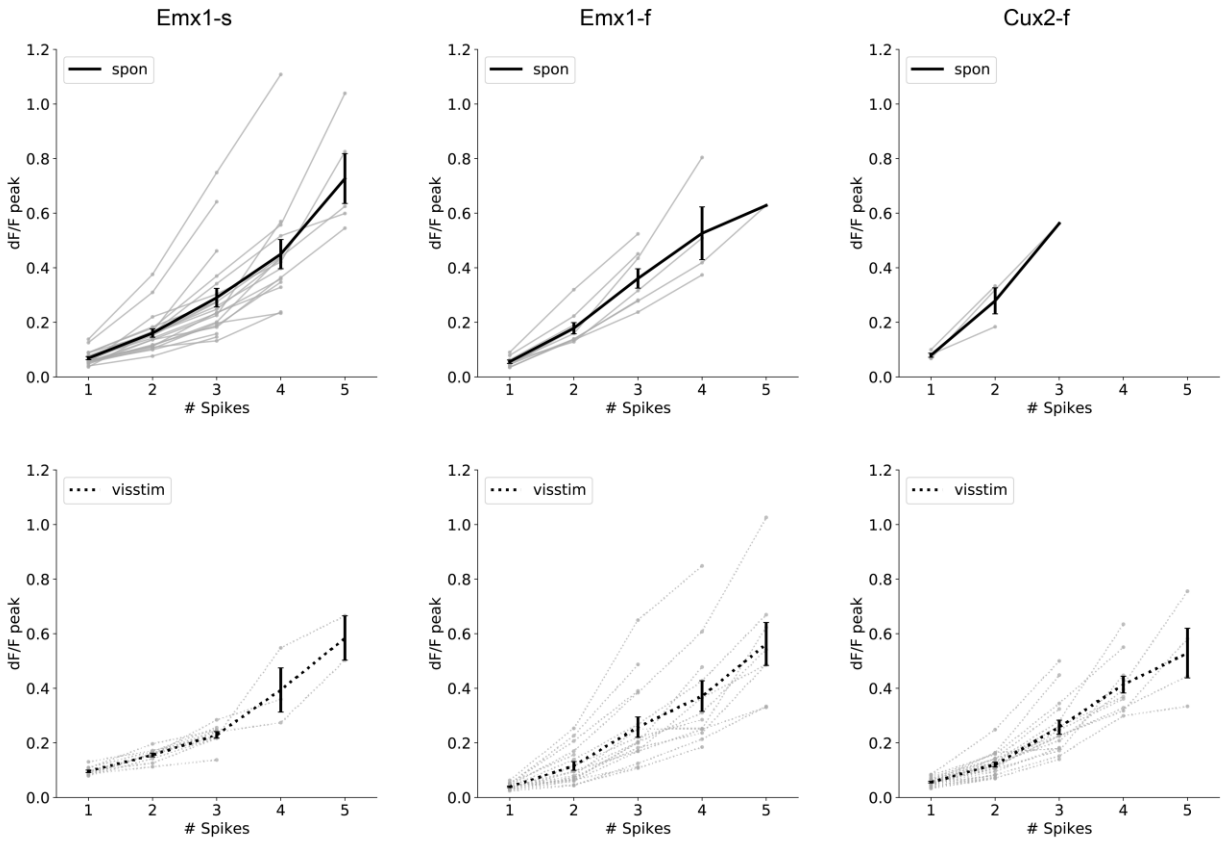
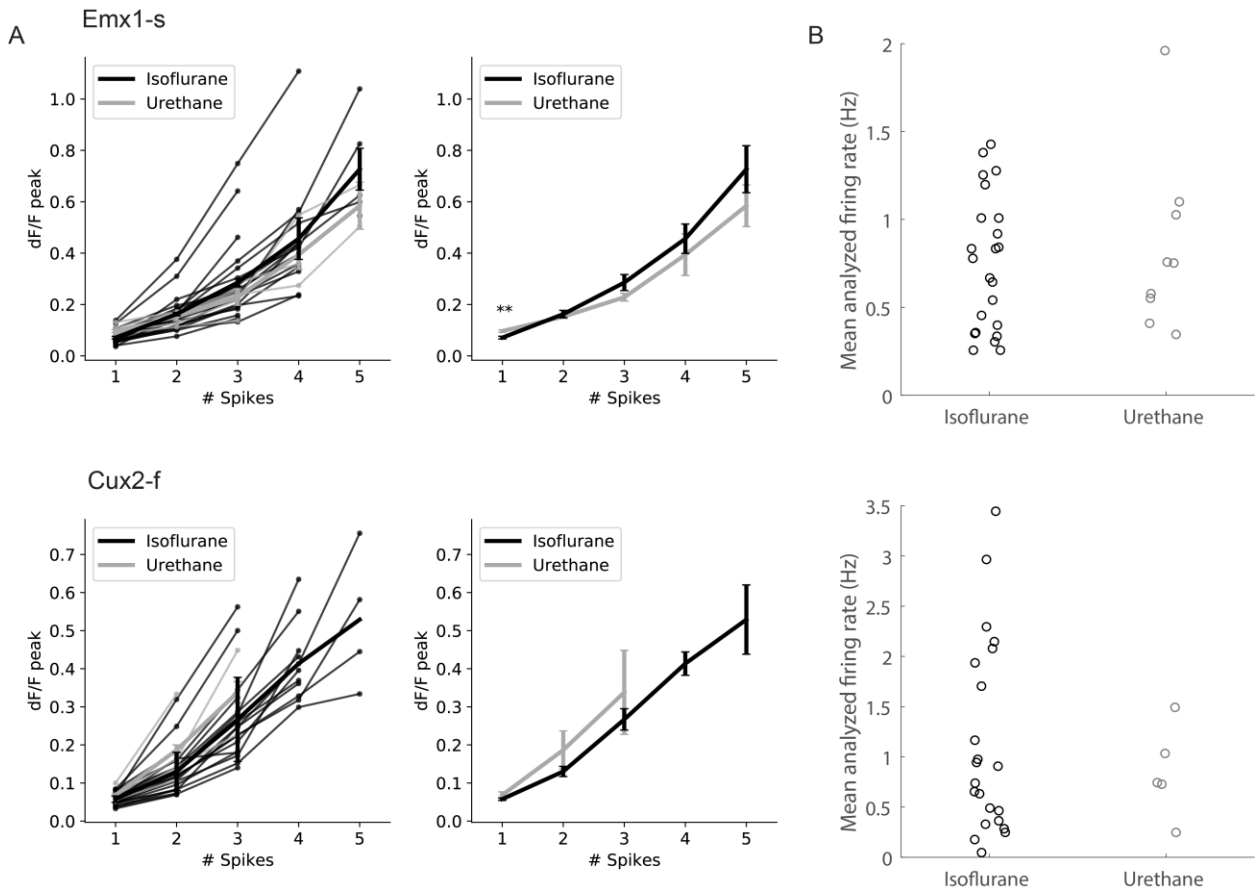


Figure 3–figure supplement 1. Population response for all genotype and visual stimulation groups. Population spike-to-calcium fluorescence response curves for spontaneous or visually-evoked responses. Bold lines indicate mean responses. No visually-evoked responses were recorded from tetO-s mice. Spike summation windows were 150 ms and 50 ms for GCaMP6s and GCaMP6f, respectively.



621
622
623 **Figure 3–figure supplement 2. Spike-to-calcium fluorescence response curve and mean firing rate**
624 **under isoflurane or urethane anesthesia**. (**A**) Single-cell spike-to-calcium fluorescence response
625 curves acquired under isoflurane or urethane anesthesia (left; bold lines indicate mean population
626 response) and overlay of population response (right). (**B**) Mean firing rate from analyzed spikes acquired
627 under isoflurane or urethane anesthesia (See **Results** for description of spiking characteristics). Emx1-s:
628 n = 9 cells from 2 mice using urethane (0 spontaneous and 9 visually-evoked recordings); n = 23 cells
629 from 5 mice using isoflurane (21 spontaneous and 2 visually-evoked recordings). Cux2-f: n = 5 cells from
630 2 mice using urethane (3 spontaneous and 2 visually-evoked recordings); n = 22 cells from 5 mice using
631 isoflurane (1 spontaneous and 21 visually-evoked recordings). Spike summation windows were 150 ms
632 and 50 ms for GCaMP6s and GCaMP6f, respectively.

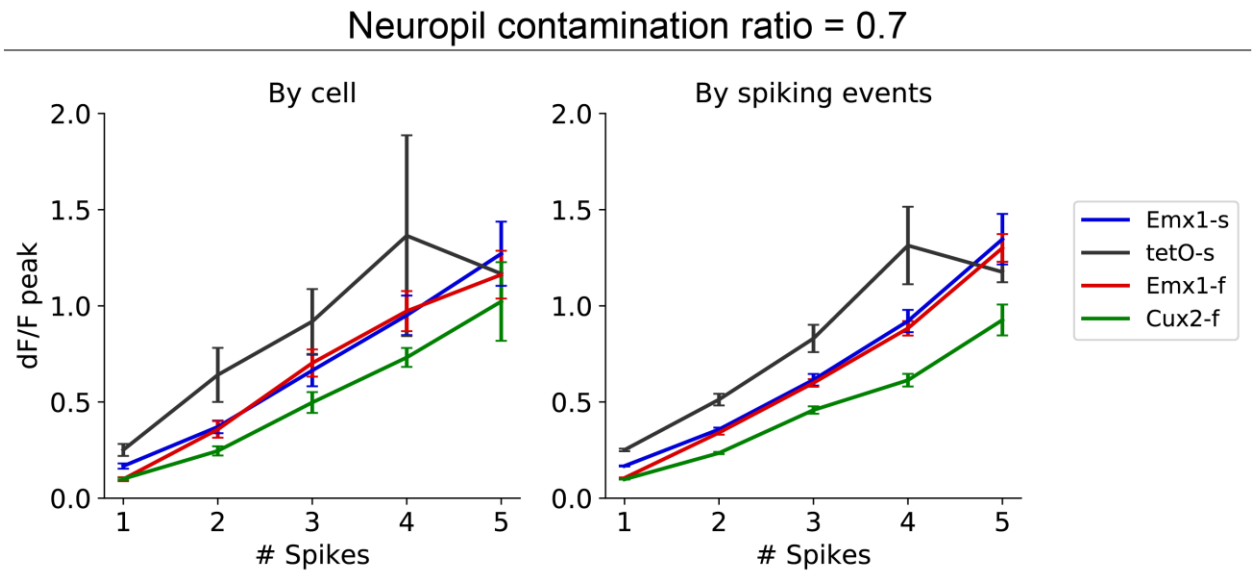


Figure 3–figure supplement 3. Population spike-to-calcium response curves with neuropil correction. Population spike-to-calcium fluorescence response curves computed from all spiking events (no exclusion of atypical spiking events) using a neuropil contamination ratio of 0.7. By cell: population response was computed from the mean response of individual cells (left). Emx1-s: n = 32 cells; tetO-s: n = 6; Emx1-f: n = 25; Cux2-f: n = 26. By spiking events: population response was computed from spiking events pooled from all cells as in Chen et al., 2013 (right). Emx1-s: n = 1264, 668, 298, 118, and 57 spiking events for 1-5 spikes; tetO-s: n = 254, 103, 48, 17, and 6; Emx1-f: n = 1300, 1010, 476, 183, and 70; Cux2-f: n = 2673, 788, 207, 118, and 46. For both analyses, only cells with ring-shaped ROI were included (see **Methods**). Spike summation windows were 150 ms and 50 ms for GCaMP6s and GCaMP6f, respectively.

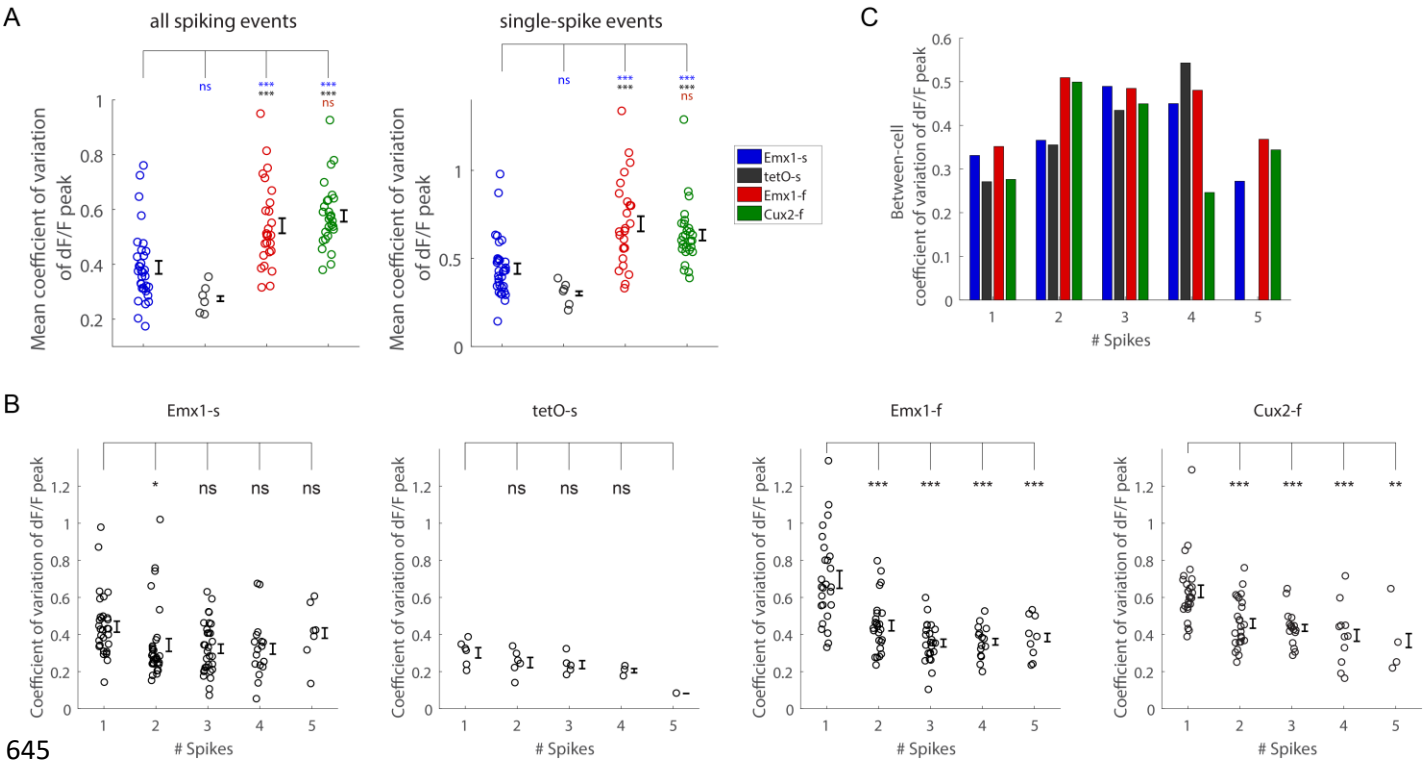


Figure 4. Within-cell and between-cell variability of dF/F peak. (**A**) Within-cell variability. Mean coefficient of variation of dF/F peaks of each cell computed from all spiking events (left panel; p = 2e-8, ANOVA) or single-spike events (right panel; p = 3e-7, ANOVA). (**B**) Within-cell variability (coefficient of variation of dF/F peak) for each spike bin (Emx1-s: p = 0.04; tetO-s: p = 0.03; Emx1-f: p = 3e-11; Cux2-f: p = 6e-6, ANOVA). ns, p > 0.05; *, p < 0.05; **, p < 0.01; ***, p < 0.001, multiple comparison test using Tukey's honest significant difference criterion. (**C**) Cell-to-cell variability. Coefficient of variation of dF/F peak from the population response in Fig. 3B (top). Spike summation windows were 150 ms and 50 ms for GCaMP6s and GCaMP6f, respectively.

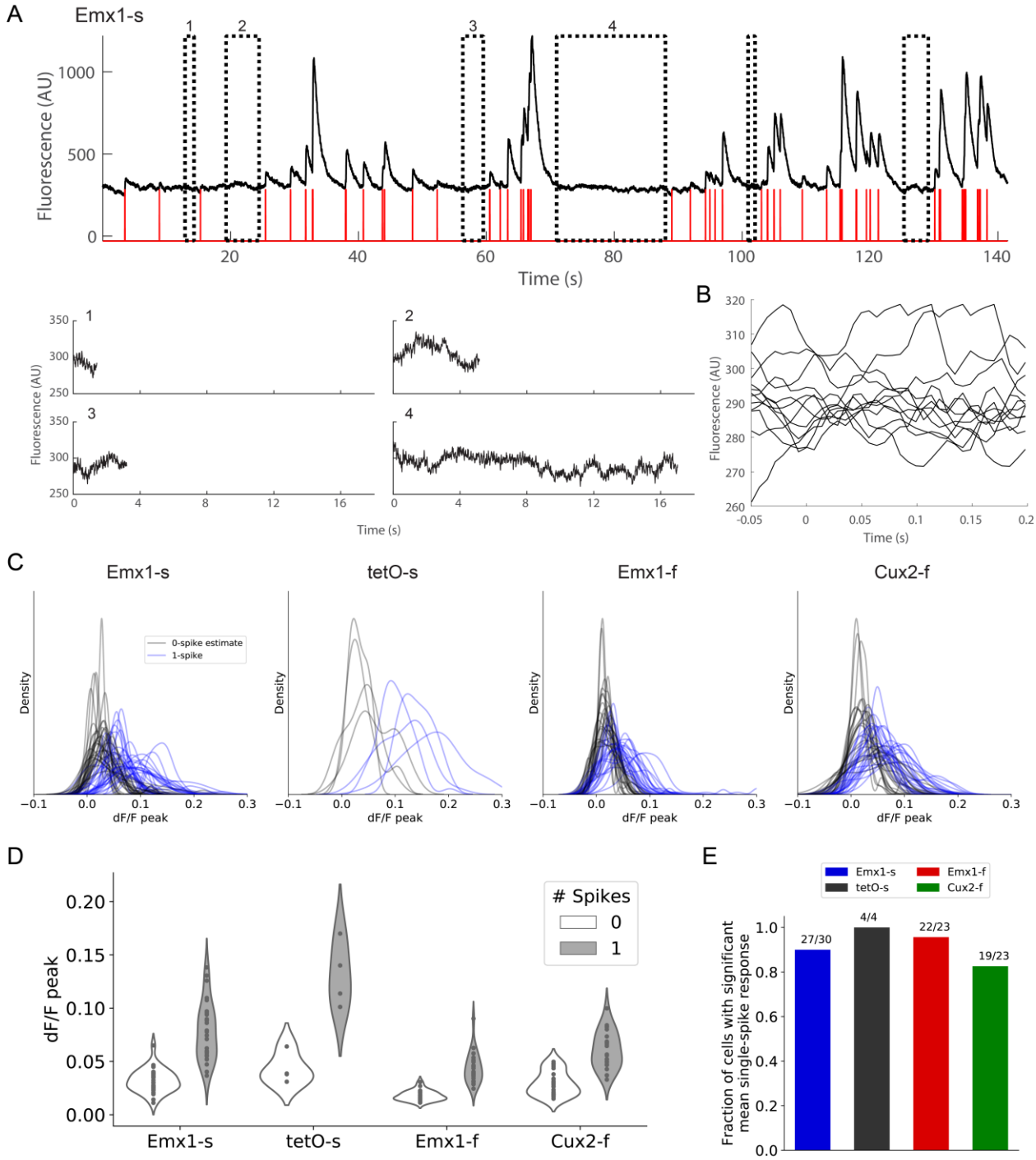


Figure 4–figure supplement 1. Comparing mean single-spike response and imaging noise. (**A**) No-spike intervals of ≥1 s, separated by ≥4 s and ≥1 s from previous and subsequent spikes, respectively, and zoomed-in view of their corresponding calcium fluorescence traces in an example Emx1-s cell. Red lines indicate detected spikes. Boxed regions indicate no-spike intervals. Zoomed-in views are shown for numbered boxed regions. (**B**) Fluorescence traces randomly sampled from all no-spike intervals in (A). (**C**) Kernel density estimates of dF/F peak sampled from no-spike intervals (zero-spike estimate) and in response to single-spikes. (**D**) Mean zero-spike and single-spike dF/F peak. Circles indicate individual cells. (**E**) Fraction of cells in which mean single-spike response was different from mean imaging noise ($p < 0.05$, t-test).

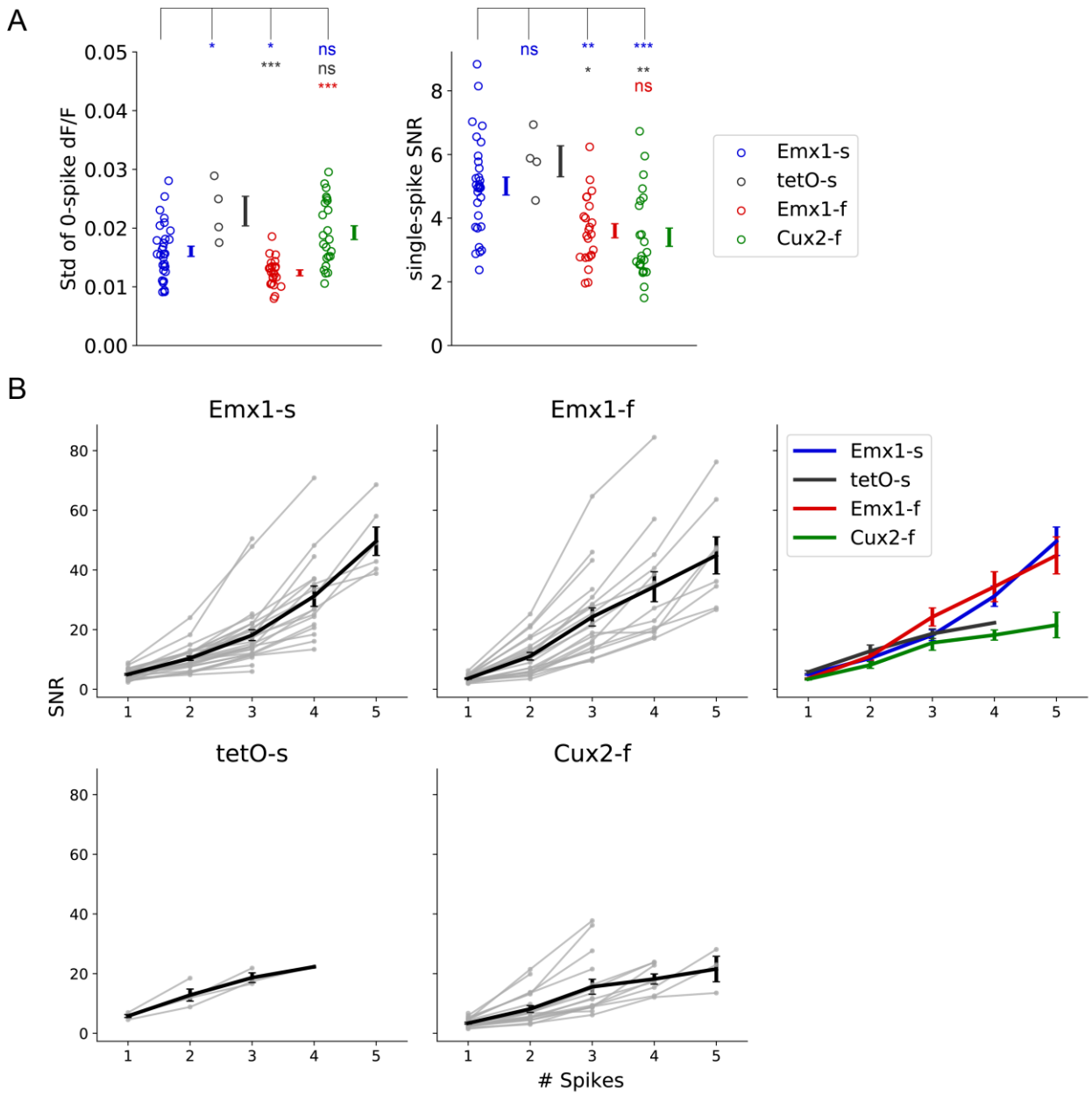


Figure 4–figure supplement 2. Noise floor and signal-to-noise ratio. (**A**) Noise floor (standard deviation of dF/F during no-spike intervals; left; p = 2e-6, ANOVA) and single-spike signal-to-noise ratio (SNR; single-spike dF/F peak normalized by standard deviation of dF/F during no-spike intervals; right; p = 2e-5, ANOVA). *, p < 0.05; **, p < 0.01, ***, p < 0.001, multiple comparison test using Tukey's honest significant difference criterion. (**B**) SNR as a function of the number of spikes (left panels), and overlay of mean SNR for the 4 mouse lines (right panel). Spike summation windows were 150 ms and 50 ms for GCaMP6s and GCaMP6f, respectively.

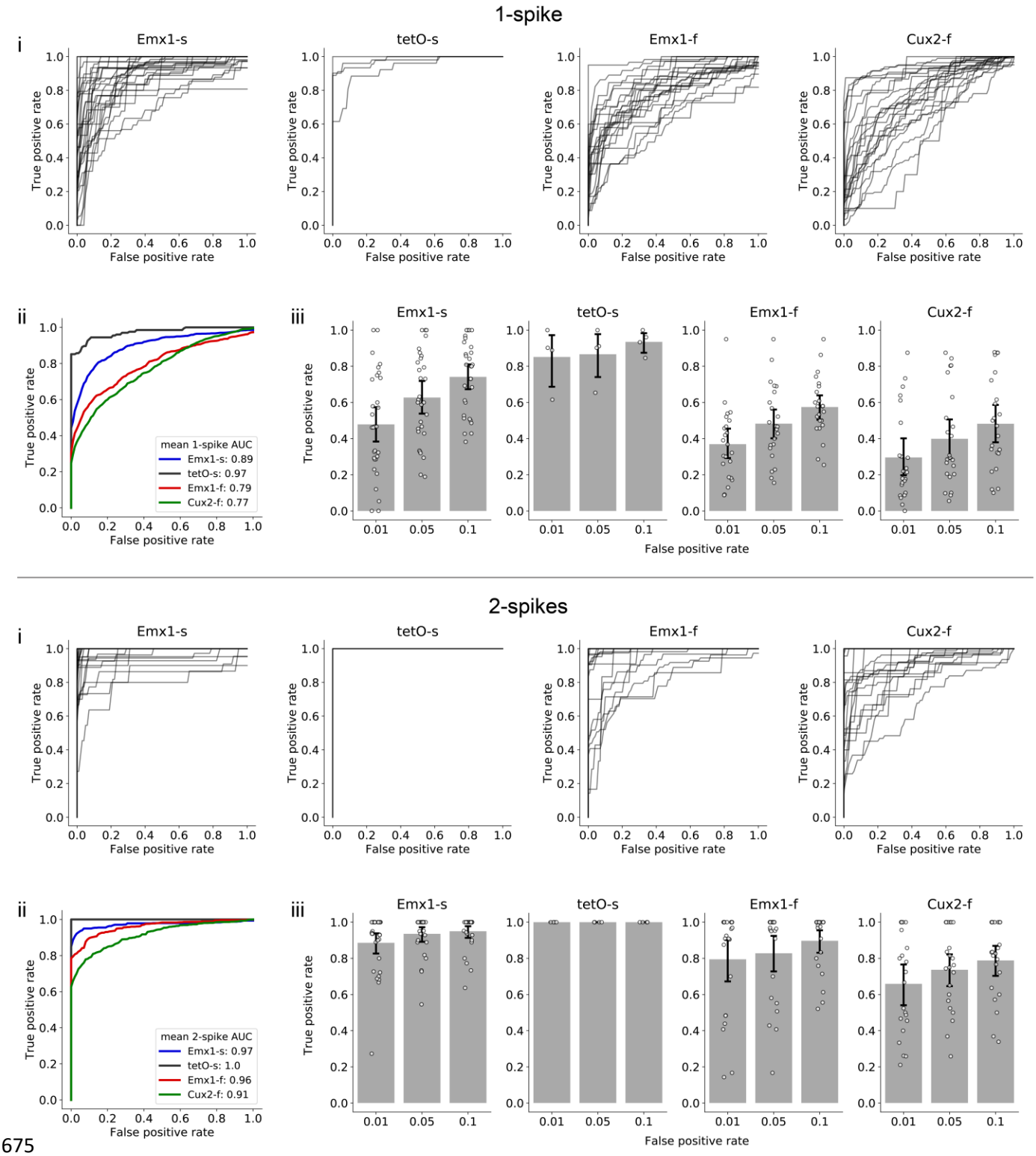


Figure 5. Spike detection at the single-trial level. (i) Receiver operating characteristic (ROC) curves for classifying 1-spike (top) and 2-spike (bottom) events in individual cells. (ii) Mean population response. (iii) Detection rate (true positive rate) at 1%, 5%, and 10% false positive rate. Emx1-s: n = 30 cells with at least 1 no-spike interval out of 32 total cells; tetO-s: n = 4 out of 6; Emx1-f: n = 23 out of 26; Cux2-f: n = 23 out of 27.

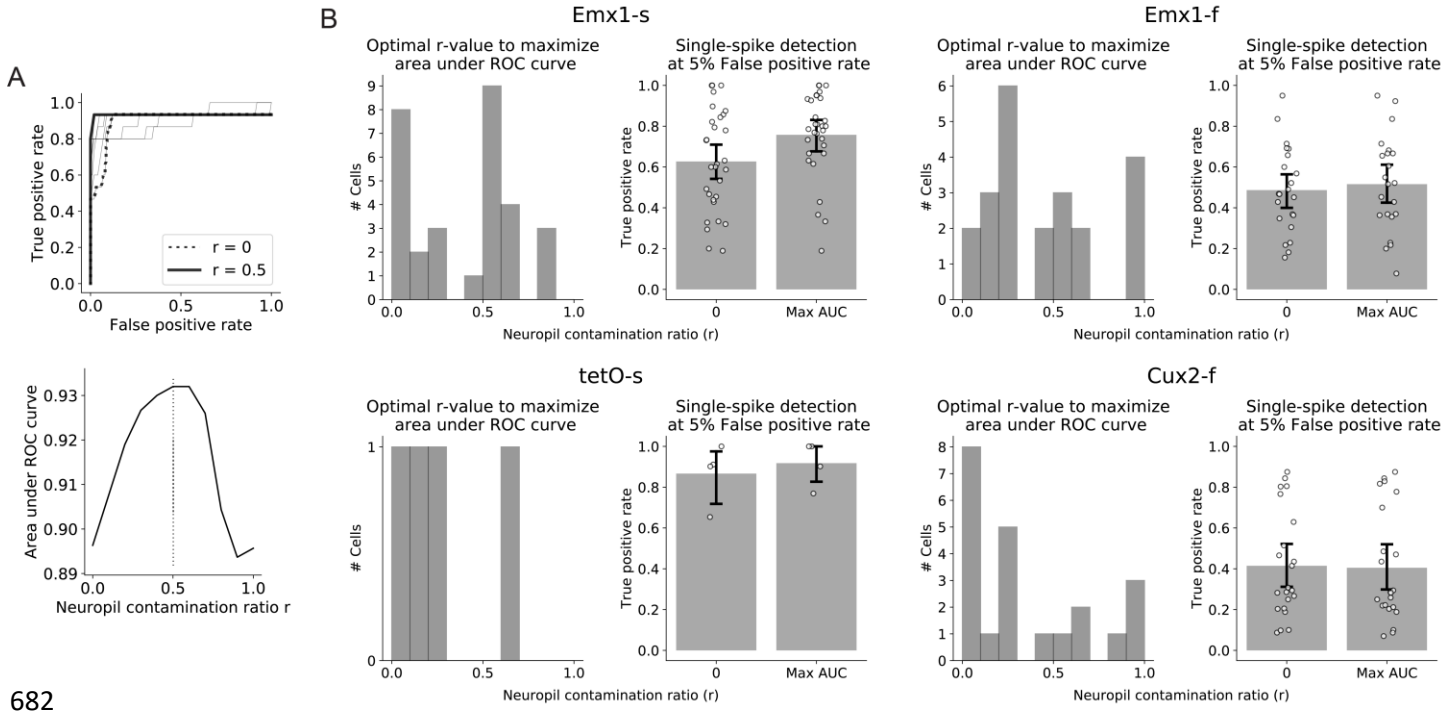


Figure 5–figure supplement 1. Effect of neuropil subtraction on single-spike detection rate. (A) Identifying the optimal neuropil contamination ratio r for maximizing spike detection efficiency. (Top) ROC curves for classifying single-spike events in an example Emx1-s cell, where the neuropil contamination ratio r was varied from 0 to 1. (Bottom) Area under ROC curves as a function of r. (B) (Left) Distribution of optimal r-values in individual cells for maximizing area under ROC curve. (Right) Single-spike detection at 5% false positive rate before (r = 0) and after optimization. (Emx1-s: n = 30 cells with at least 1 no-spike interval; tetO-s: n = 4; Emx1-f: n = 22; Cux2-f: n = 22. One Emx1-f and 1 Cux2-f cell were not analyzed because a satisfactory ring-shaped ROI could not be found automatically as described in **Methods**.)

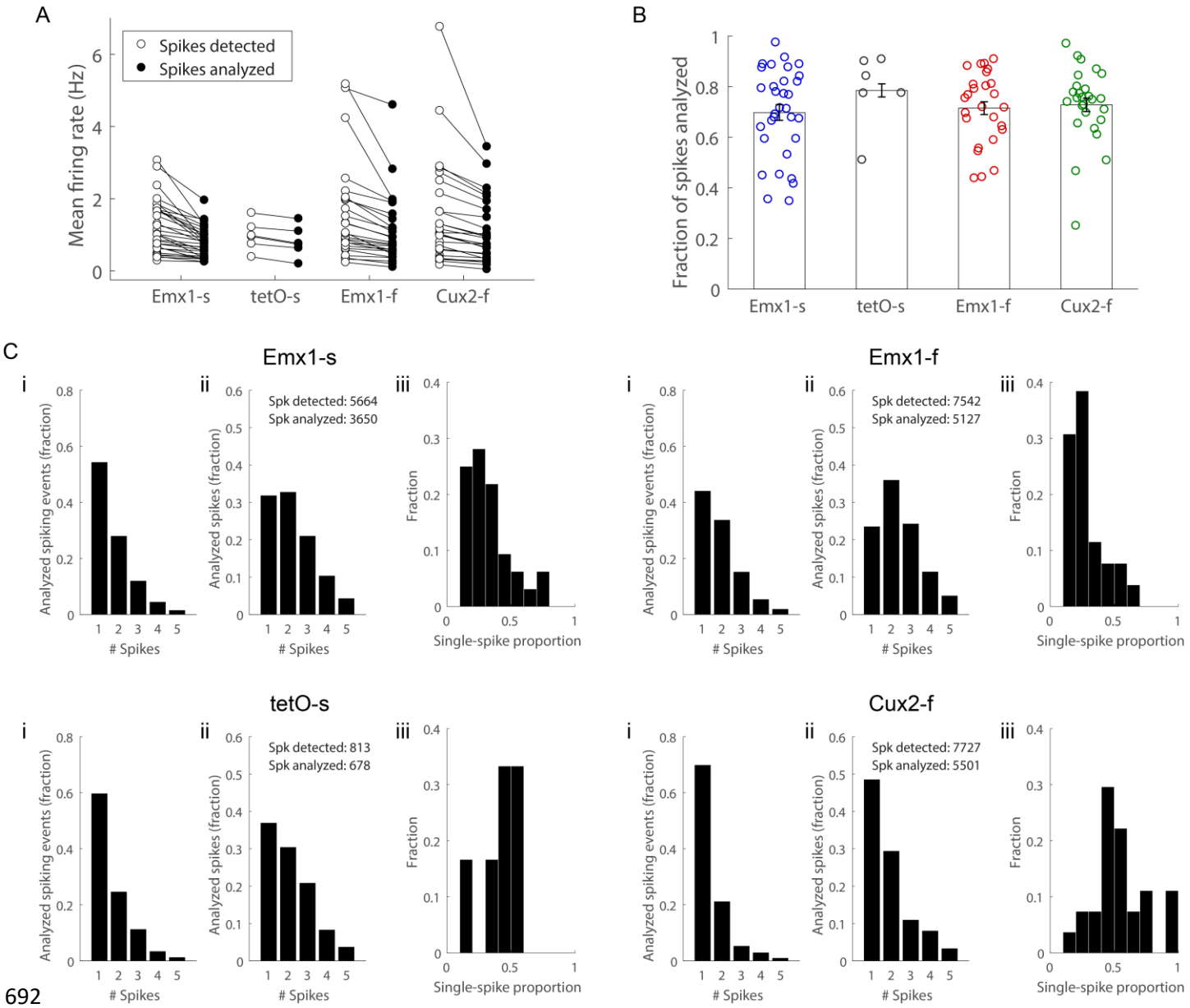


Figure 6. Spiking characteristics. (**A**) Mean firing rate computed from every detected spike or analyzed spikes within the same time window (i.e. isolated spiking events with atypical spiking events excluded). Each circle (pair) indicates one cell. (**B**) Fraction of spikes analyzed. Each circle indicates one cell. (**C**) (i) Fraction of spiking events in each spike bin; (ii) Fraction of number of spikes in each spike bin; and (iii) Single-spike proportion by cell. Spike summation windows were 150 ms and 50 ms for GCaMP6s and GCaMP6f, respectively.

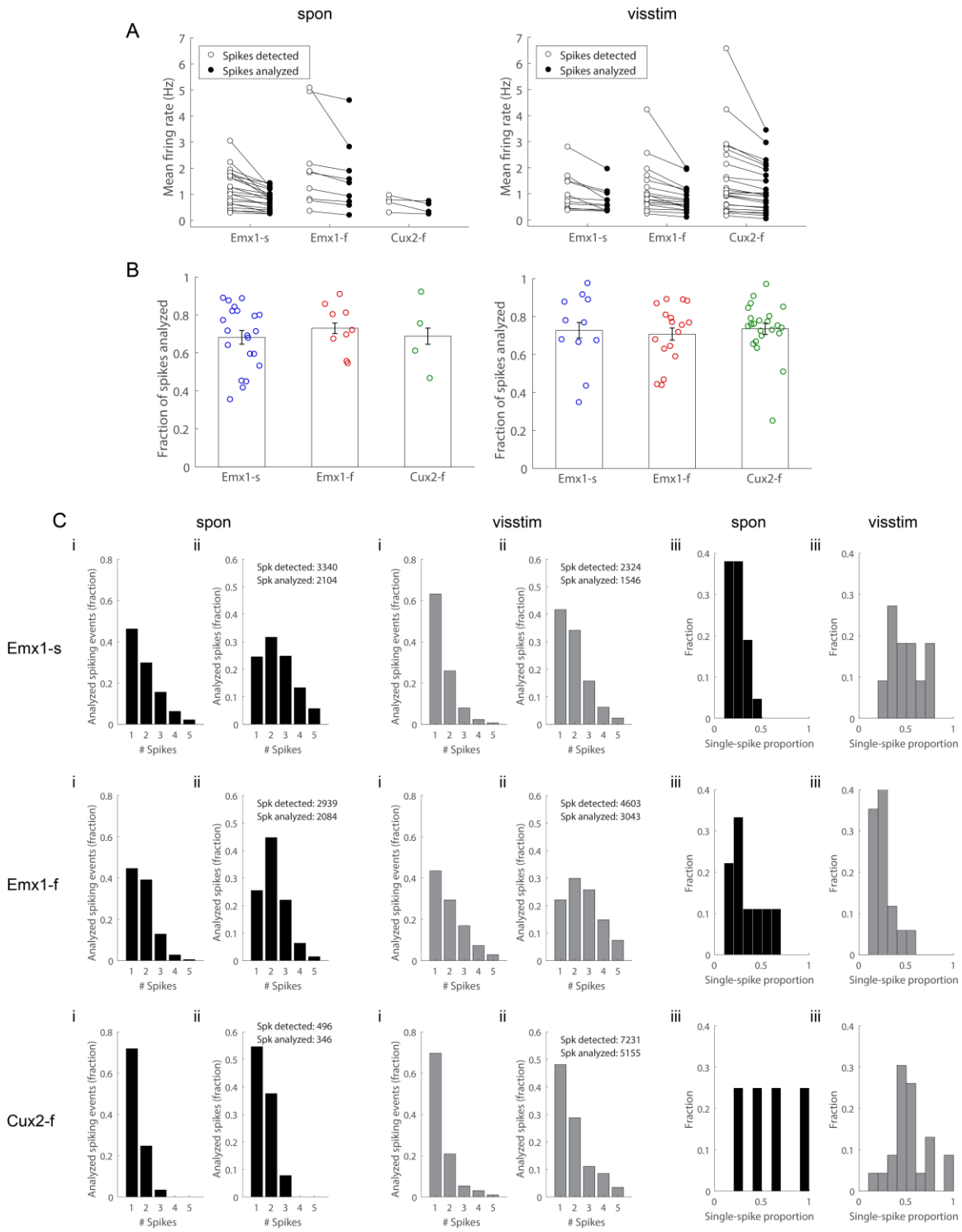


Figure 6–figure supplement 1. Spiking characteristics for spontaneous and visually-evoked cells. (**A**) Mean firing rate from every detected spike or analyzed spikes (i.e. isolated spiking events with atypical spiking events excluded). Each circle (pair) indicates 1 cell. (**B**) Fraction of spikes analyzed. Each circle indicates 1 cell. (**C**) (i) Fraction of spiking events in each spike bin; (ii) Fraction of number of spikes in each spike bin; and (iii) Single-spike proportion by cell. Spon and visstim indicate spontaneous and visually-evoked responses, respectively. Spike summation windows were 150 ms and 50 ms for GCaMP6s and GCaMP6f, respectively.

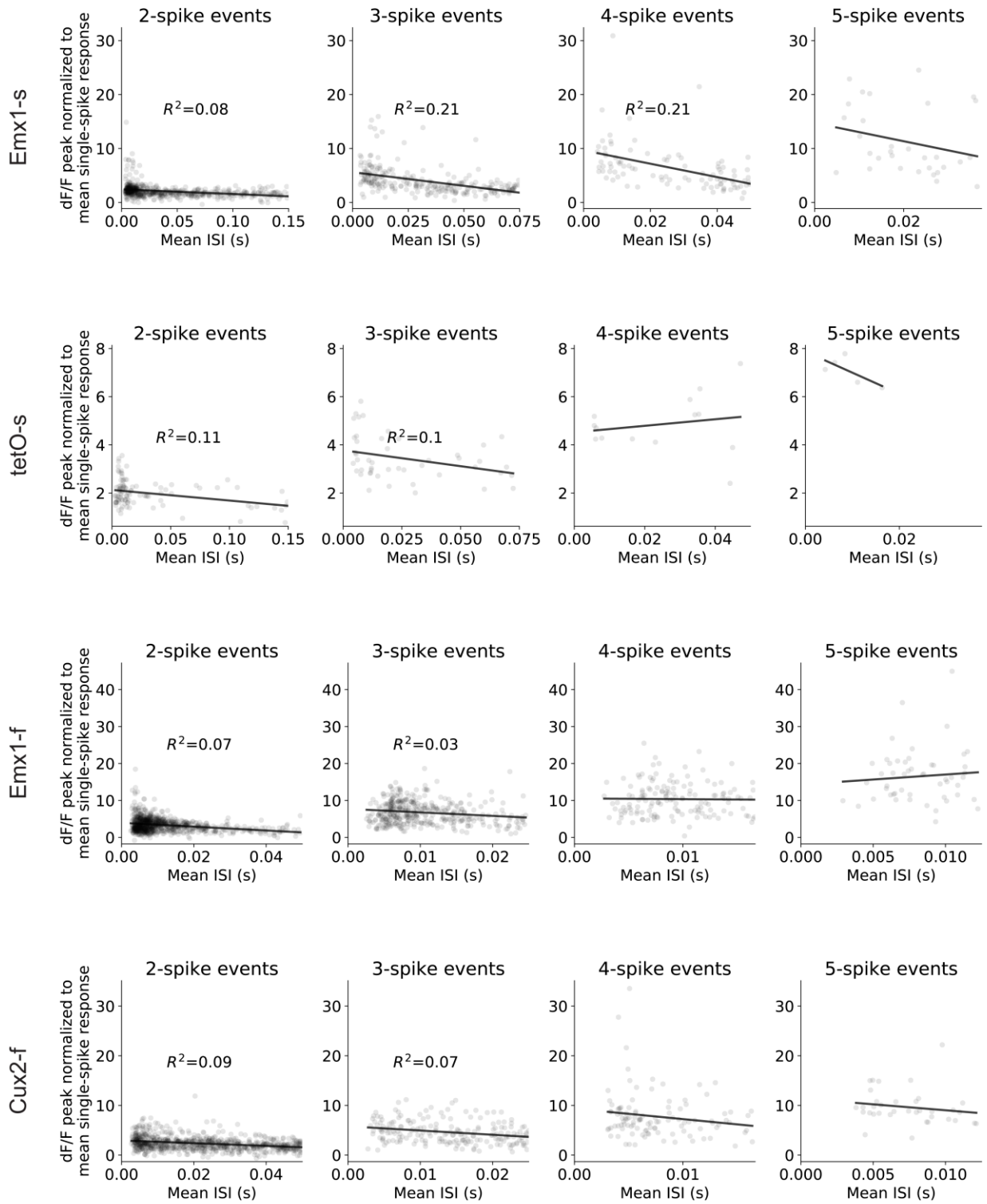


Figure 6–figure supplement 2. Relationship between mean inter-spike interval and dF/F peak at the population level. Relationship between mean inter-spike interval (ISI) and dF/F peak (normalized to the mean single-spike response) for multi-spike events pooled from all cells. Circles and lines indicate spiking events and linear regression fit. Spike summation windows were 150 ms and 50 ms for GCaMP6s and GCaMP6f, respectively.

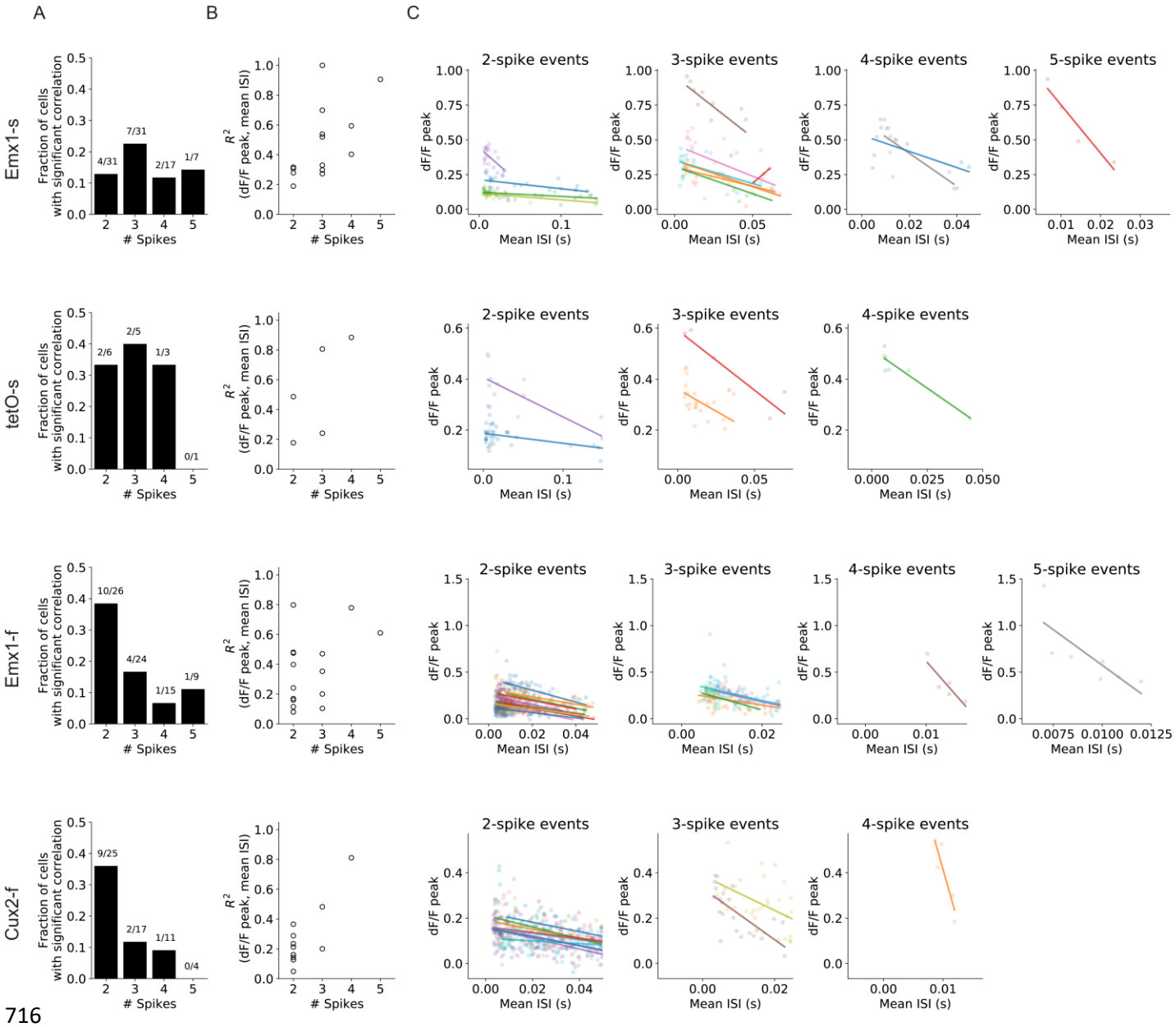
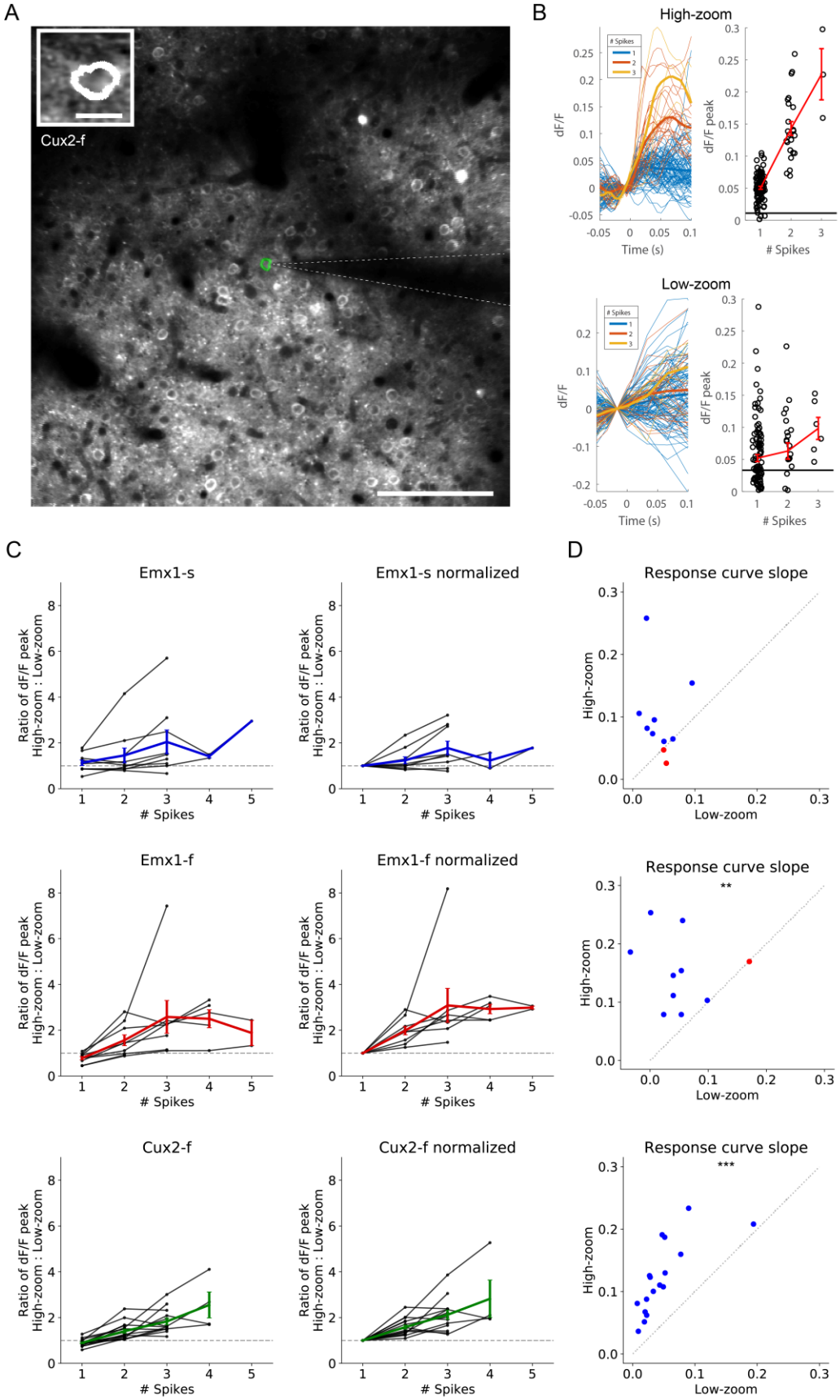
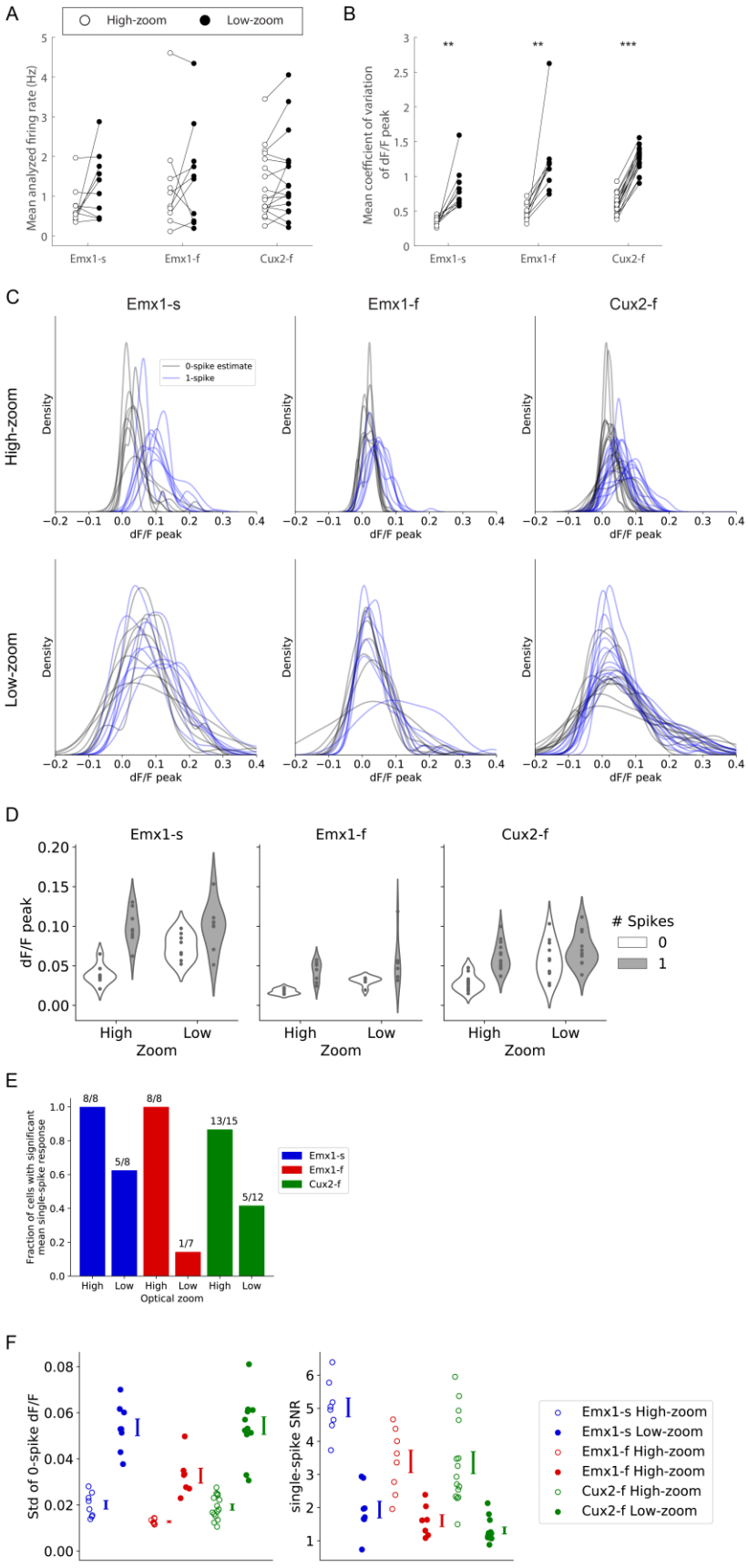


Figure 6–figure supplement 3. Relationship between mean ISI and dF/F peak in individual cells. (**A, B**) Between-cell variability. (**A**) Fraction of cells with significant correlation coefficients of dF/F peak and mean ISI ($p < 0.05$). Denominators indicate the number of cells with the corresponding multi-spike event. Overall, significant correlations were observed in 14 out of 86 (16%) multi-spike events for Emx1-s, 5 out of 14 (36%) multi-spike events for tetO-s, 16 out of 74 (22%) multi-spike events for Emx1-f, and 12 out of 57 (21%) multi-spike events for Cux2-f. (**B**) R^2 of dF/F peak and mean ISI for cells with significant correlation coefficients. (**C**) Within-cell relationship between mean ISI and dF/F peak for cells with significant correlation coefficients. Circles and lines indicate spiking events and linear regression fit; colors indicate individual cells and do not correspond across figure panels. Spike summation windows were 150 ms and 50 ms for GCaMP6s and GCaMP6f, respectively.


A
Cux2-f
B
High-zoom
Spikes
dF/F
Time (s)
dF/F peak
Spikes
Low-zoom
C
Emx1-s
Emx1-s normalized
Emx1-f
Emx1-f normalized
Cux2-f
Cux2-f normalized
Ratio of dF/F peak
High-zoom : Low-zoom
Spikes
D
Response curve slope
High-zoom
Low-zoom
**
***

728

Figure 7. Spike-to-calcium fluorescence response curves imaged at high and low spatiotemporal resolutions. (**A**) Fields of view and cell ROIs for an example Cux2-f neuron imaged at both high (white ROI, inset) and low (green ROI) optical zooms. Scale bars, high-zoom: 10 μm; low-zoom: 100 μm. (**B**) Corresponding dF/F traces and spike-to-calcium fluorescence response curves for the neuron shown in (A). Horizontal black lines indicate the standard deviation of dF/F during no-spike intervals (i.e. estimated noise floor; see **Figure 4–figure supplement 2**). (**C**) Population response of cells imaged at both high- and low-zooms, expressed as the ratio of high-zoom to low-zoom dF/F peak, without (left) and with normalization (right) by the mean single-spike response. Colored bold lines indicate the mean response. One Emx1-f cell with near zero (<0.01) dF/F peak for 2-spikes at low-zoom was excluded from this plot. (**D**) Response curve slope imaged at high-zoom and low-zoom, computed from linear regression of dF/F peak against the number of spikes. Blue or red indicates response curve slope was larger when imaged at high-zoom or low-zoom, respectively. Diagonal line indicates equal response curve slope when imaged at high-zoom and low-zoom. (***, $p < 0.001$; **, $p < 0.01$, paired t-test.) Spike summation windows were 150 ms and 50 ms for GCaMP6s and GCaMP6f, respectively.


A
High-zoom
Low-zoom
Mean analyzed firing rate (Hz)
Emx1-s
Emx1-f
Cux2-f
B
Mean coefficient of variation of dF/F peak
**
**
***
C
Emx1-s
Emx1-f
Cux2-f
High-zoom
Low-zoom
Density
0-spike estimate
1-spike
dF/F peak
D
dF/F peak
Spikes
0
1
High
Low
Zoom
E
Fraction of cells with significant mean single-spike response
8/8
5/8
8/8
1/7
13/15
5/12
Emx1-s
Emx1-f
Cux2-f
Optical zoom
F
Std of 0-spike dF/F
single-spike SNR
Emx1-s High-zoom
Emx1-s Low-zoom
Emx1-f High-zoom
Emx1-f High-zoom
Cux2-f High-zoom
Cux2-f Low-zoom

744

745 **Figure 7–figure supplement 1. Fluorescence response variability imaged at high and low**
746 **spatiotemporal resolutions.** (**A**) Mean firing rate from analyzed spikes. (**B**) Mean coefficient of variation
747 of dF/F peak. Each circle (pair) indicates 1 cell. **, $p < 0.01$; ***, $p < 0.001$, paired t-test. (**C**) Kernel
748 density estimates of dF/F peak sampled from no-spike intervals (zero-spike estimate) and in response to
749 single-spikes. (**D**) Mean zero-spike and single-spike dF/F peak. Circles indicate individual cells. (**E**)
750 Fraction of cells in which mean single-spike response was different from mean imaging noise ($p < 0.05$, t-
751 test). (**F**) Noise floor (standard deviation of dF/F during no-spike intervals; left) and single-spike SNR
752 (single-spike dF/F peak normalized by standard deviation of dF/F during no-spike intervals; right).

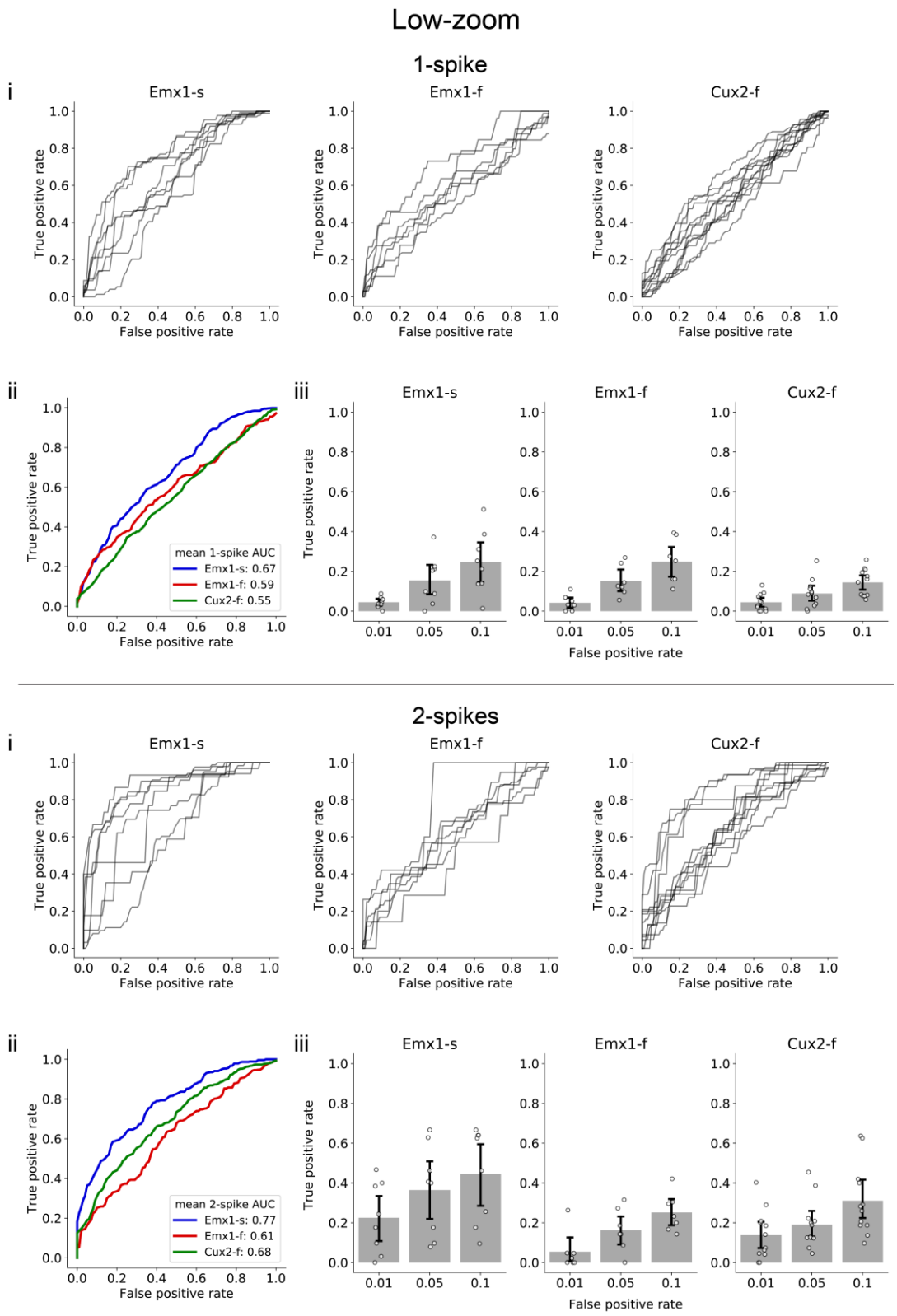


Figure 7–figure supplement 2. Spike detection at low spatiotemporal resolution. (i) ROC curves for classifying 1-spike (top) and 2-spike (bottom) events in individual cells. (ii) Mean population response. (iii) Detection rate (true positive rate) at 1%, 5%, and 10% false positive rate.