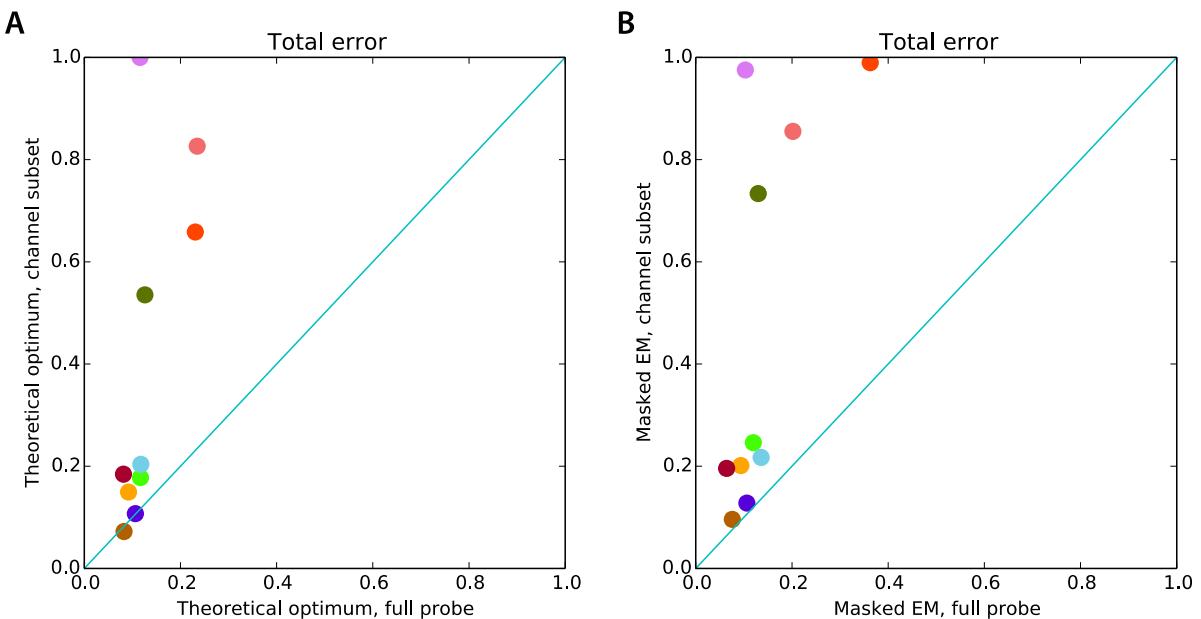
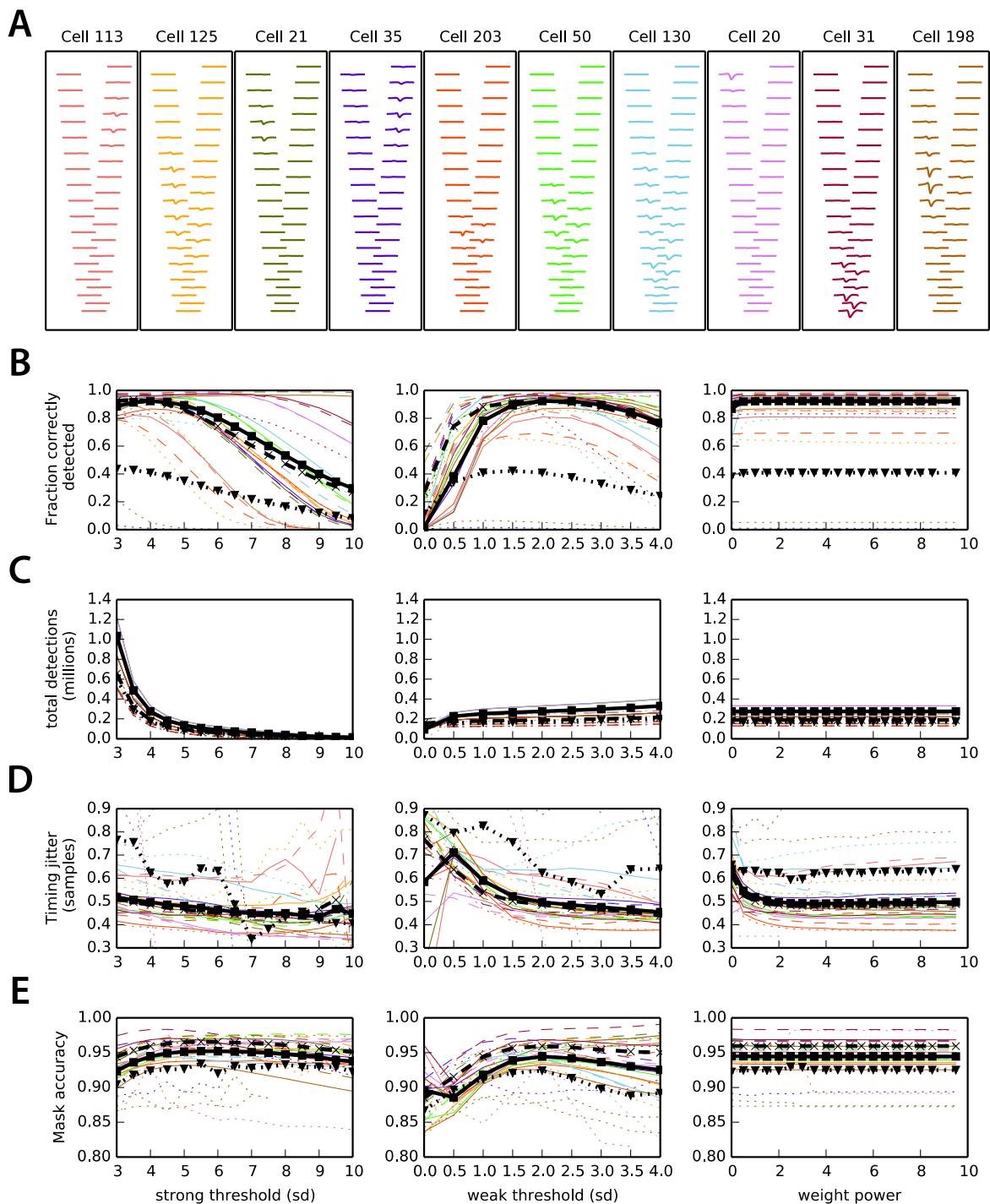


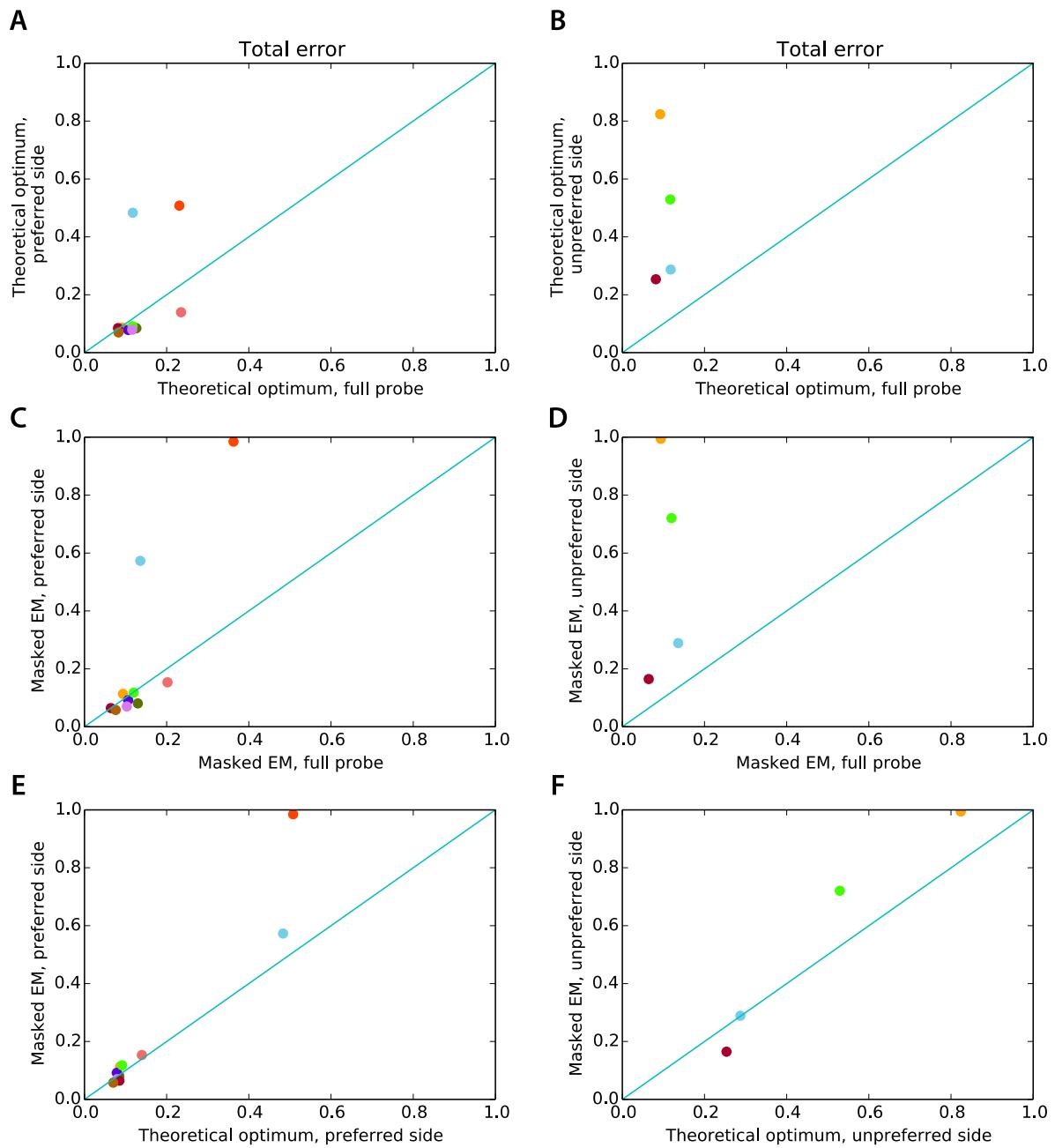
**Supplementary Figure 1: Spike detection performance for simulations of lowered site density.** To evaluate the effect of probe site spacing on unit isolation, we considered a simulation in which half the probe sites had been removed. **A**, Mean waveforms for hybrid cells analyzed (c.f. Figure 3A). Solid lines indicate channels that were included in the simulation analyses; dashed lines indicate channels that were excluded. **B-E**, Statistics assessing spike detection accuracy for full data (solid lines) and increased site spacing (dashed lines); colors correspond to the color scheme in A, and black represents the averaged over all cells (c.f. Figures 3B-E). Note that lowered density has a substantial effect on the fraction correctly detected, reducing it on average to below 80%.



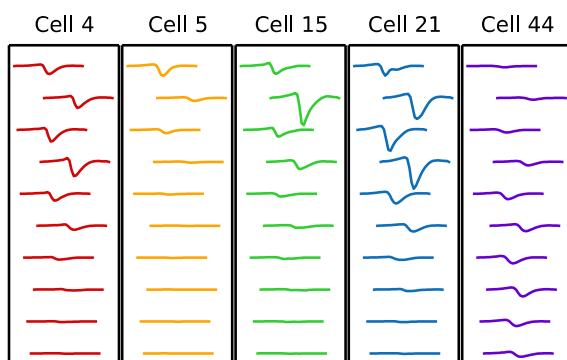
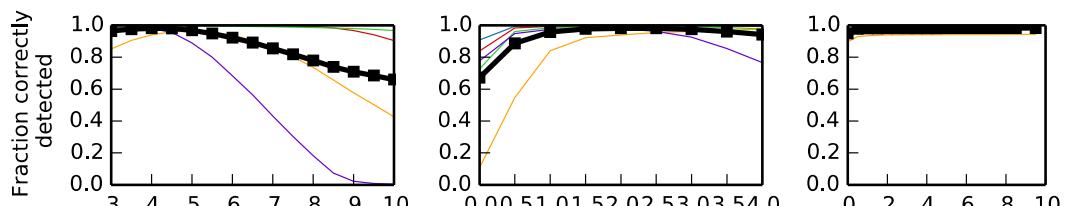
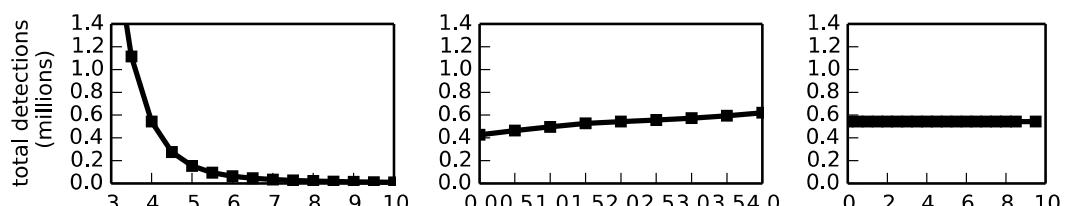
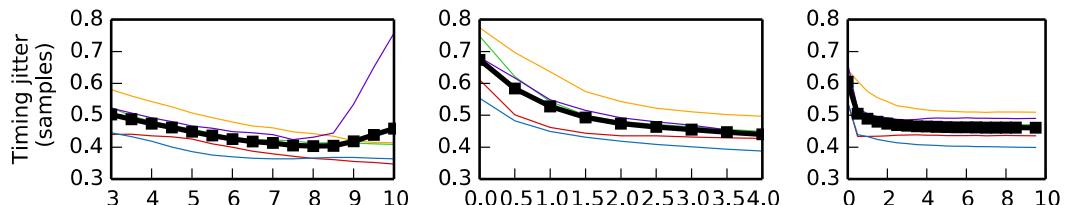
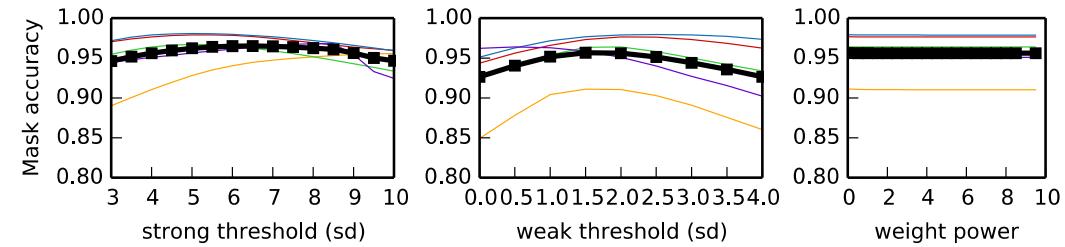
**Supplementary figure 2: Classification performance for simulations of lowered probe density. A,** Comparison of theoretical optimum error rates obtained by supervised learning (support vector machine with quadratic kernel) for the full probe (x-axis) and the subset simulating lower site density (y-axis). Colors match waveforms in Supplementary Figure 1A. Note that performance decreases for all cells after lowering site spacing, in some cases substantially. **B,** Analogous plot comparing total errors obtained for masked EM on both the full probe, vs the less dense probe. Results are similar to those for the theoretical optimum.



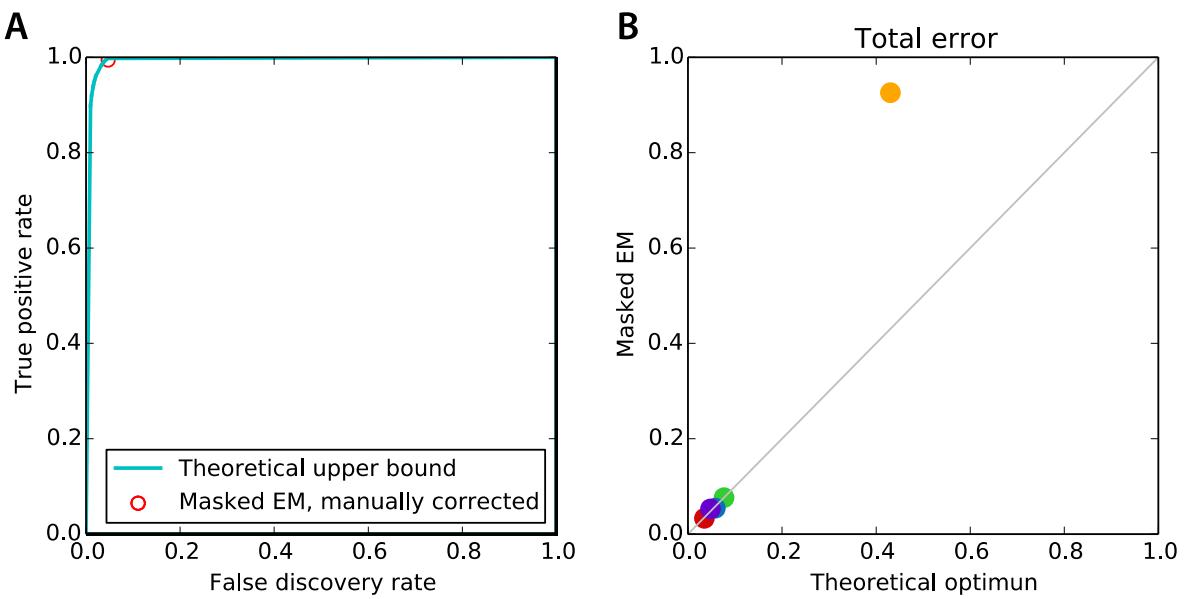
**Supplementary figure 3: Detection performance for simulations of single-sided probes.** **A**, Hybrid spike waveforms. **B-E**, Statistics assessing spike detection accuracy for full data (solid lines), preferred side (dashed lines), and unpreferred side (dotted lines). Colors are as in (A); black indicates an average over all cells. Note that while performance for the unpreferred side is extremely poor (detection rates of <40%), performance for the favored side is comparable, and can be slightly better than the full probe by some metrics.



**Supplementary figure 4: Classification performance for simulated single-sided probes.** Each plot shows a comparison of total error rates for two conditions. **A**, theoretical optimum performance of full probe vs. preferred side only. **B**, theoretical optimum performance for full probe vs. unpreferred side only. **C**, Cluster analysis for full probe vs. preferred side only. **D**, Cluster analysis for full probe vs. preferred side only. **E**, Theoretical optimum vs. cluster analysis for preferred side only. **F**, Theoretical optimum vs. cluster analysis for unpreferred side only. Note that performance on the preferred side is sometimes worse and sometimes better than performance on the full probe for both theoretical optimum and cluster analysis, whereas performance on the unpreferred side is always worse, often dramatically so. Also note that performance of cluster analysis is close to theoretical optimum in all cases except those where the theoretical optimum is itself poor ( $\sim 50\%$  error, cells 125 and 203).

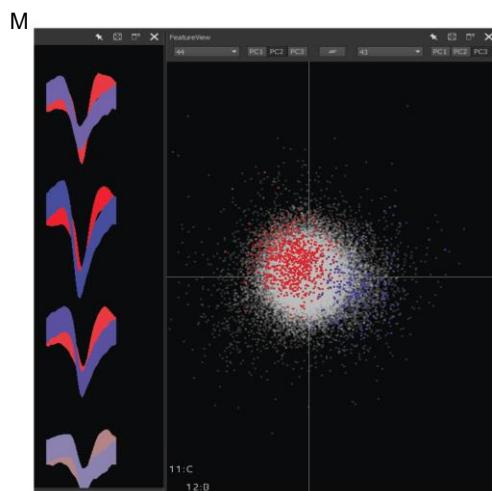
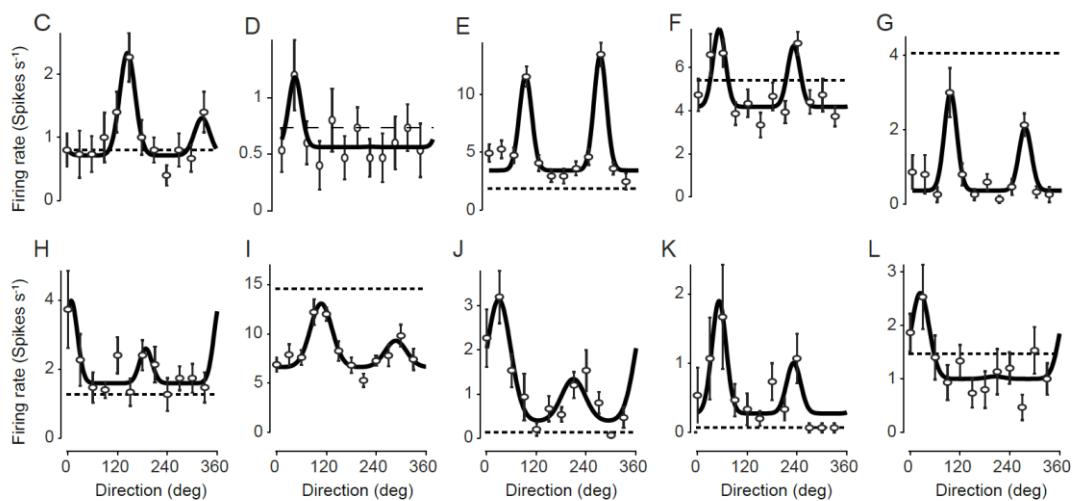
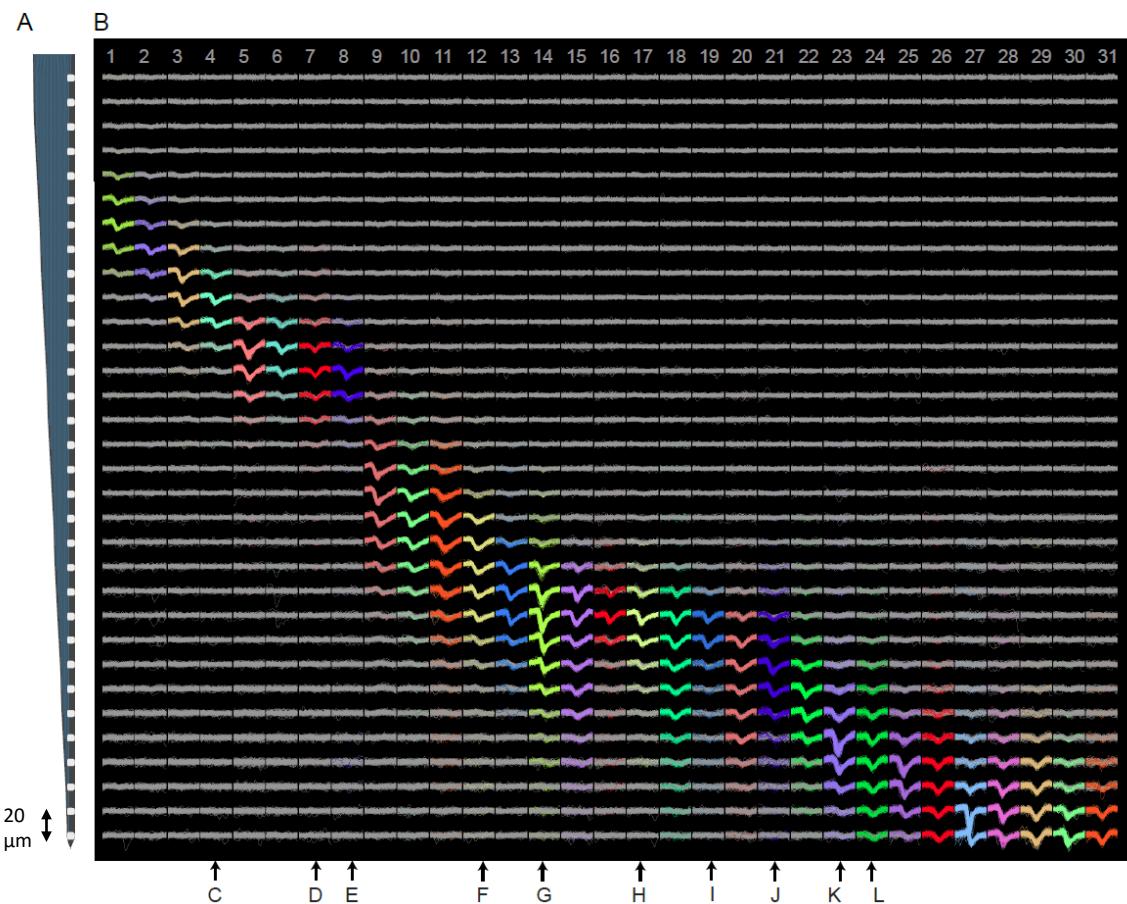
**A****B****C****D****E**

**Supplementary figure 5: Detection performance for data from rat hippocampal area CA1 . A, Hybrid spike waveforms. B-E, Detection statistics. The same parameters that gave optimal performance in neocortex (strong threshold of 4 S.D., weak threshold of 2 S.D.) were also optimal in CA1.**

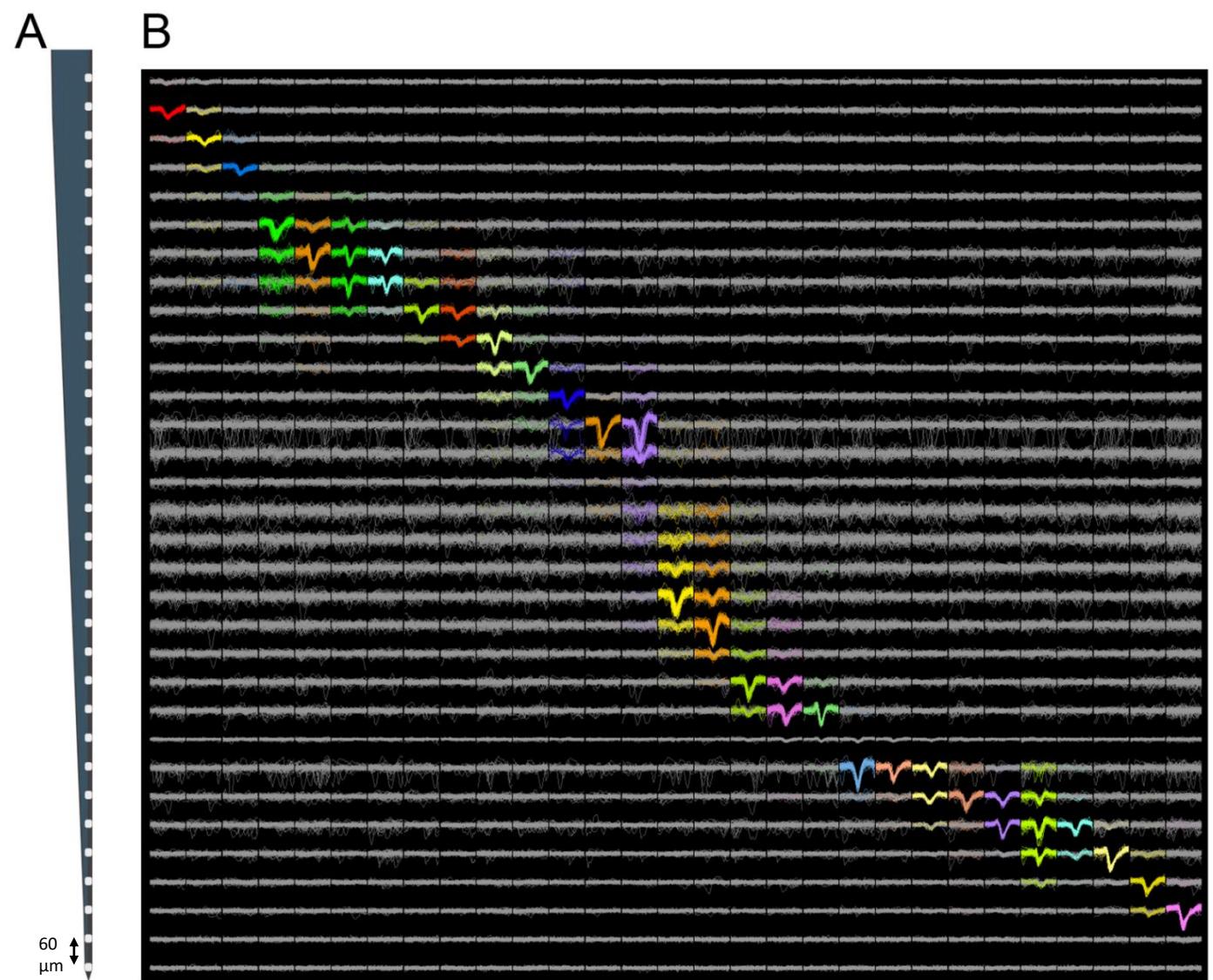


**Supplementary figure 6: Clustering performance for data from rat hippocampal area CA1 . A,**

Theoretically optimal performance (cyan line) and performance of the clustering software (red). **B**, Comparison of actual performance and theoretical optimum across all hybrids. Performance is similar except in cases where the theoretical optimum is itself very poor (~50%; cell 5).

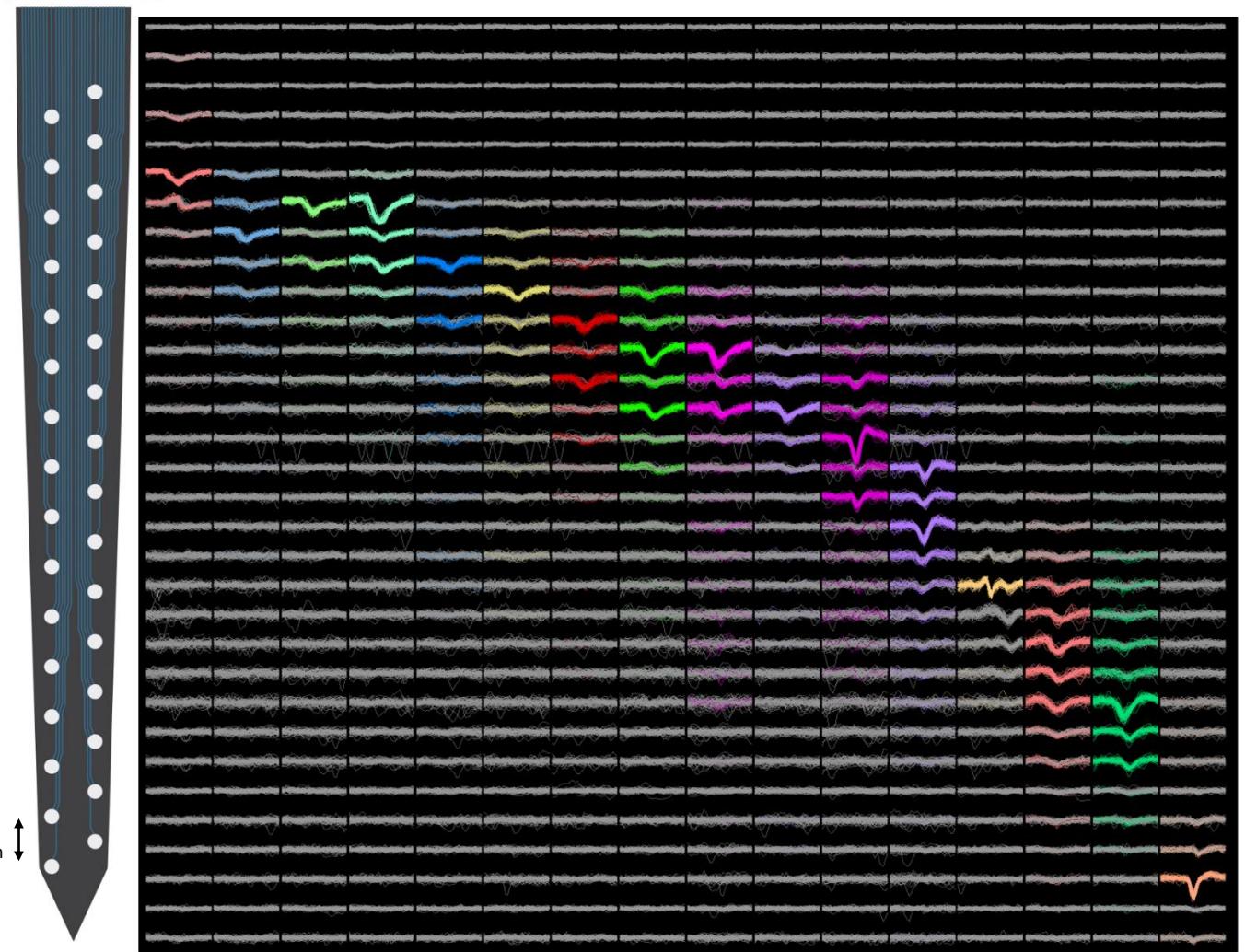


**Supplementary Figure 7: Example analysis of mouse V1 data.** **A.** Schematic of electrode array used to record in primary visual cortex (NeuroNexus “edge probe”, 20  $\mu\text{m}$  spacing; tip depth of 800  $\mu\text{m}$ ). **B.** Screenshot showing a subset of detected waveforms, arranged by neuronal depth. Each column corresponds to a clustered unit, while each row corresponds to a recording site. Colored traces indicate unmasked spikes. **C-L.** Direction tuning curves for a subset of these units (drifting sinusoidal grating 0.05 cycles/deg; 3 cycles/sec; 1 sec; 40 deg). Error bars: s.e.m.; dotted line: baseline firing rate. **M.** KlustaView screenshot showing the fine waveform and feature space differences distinguishing two clusters with different sensory tuning (clusters 7 and 8; receptive fields shown in D, E).

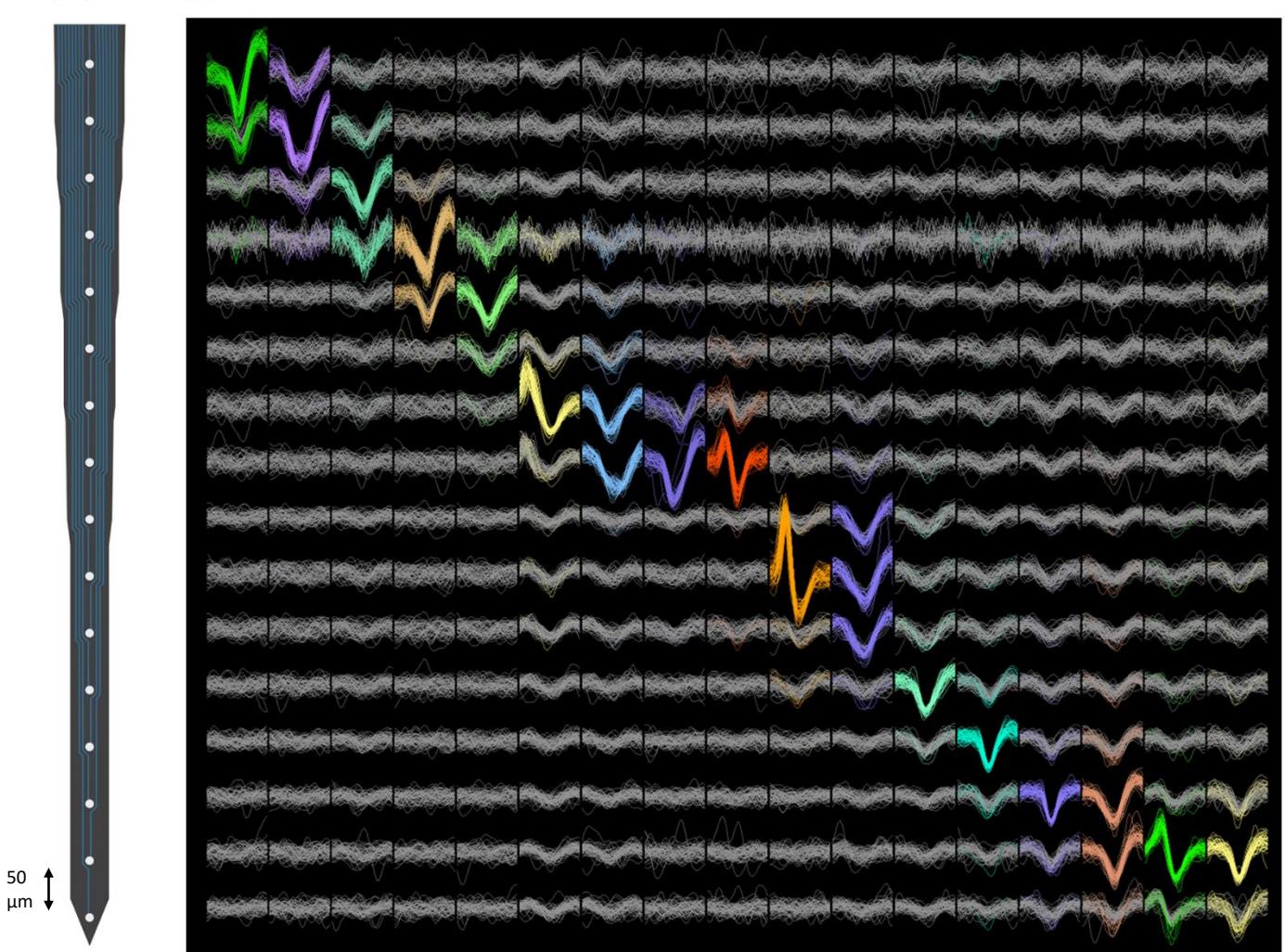


**Supplementary figure 8:** Spike sorting of 32-site data in V1 of behaving macaque. **A**, diagram of 32-site Neuronexus “edge” probe, 60  $\mu\text{m}$  spacing. **B**, Screenshot showing a subset of detected waveforms, arranged by neuronal depth. Each column corresponds to a clustered unit, while each row corresponds to a recording site. Colored traces indicate unmasked spikes.

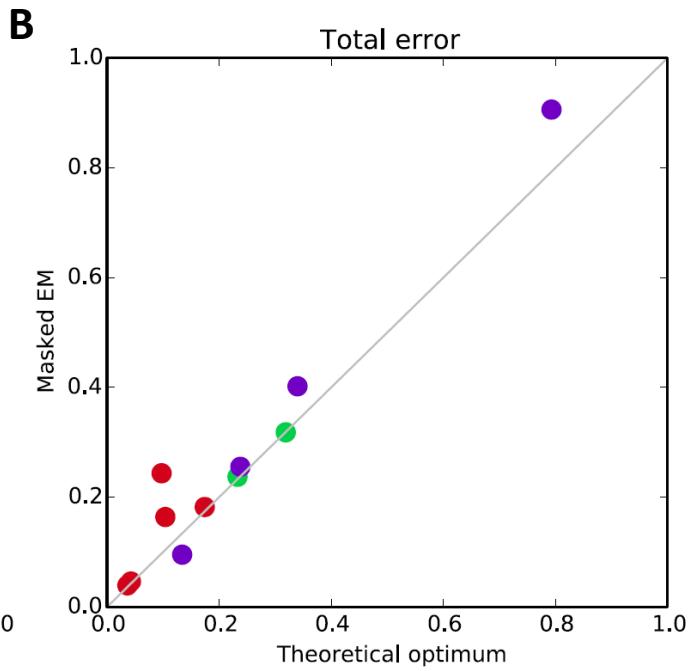
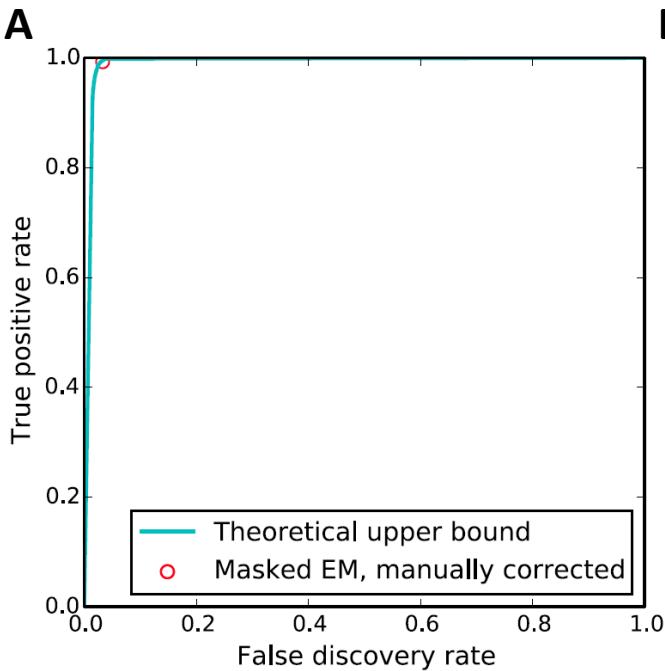
A B



**Supplementary figure 9:** Spike sorting of 32-site data in V1 of behaving macaque. **A**, diagram of 32-site “poly2” probe, 50  $\mu\text{m}$  same-side spacing. **B**, Superimposed waveforms of spikes belonging to individual clusters. Each column corresponds to a clustered unit, while each row corresponds to a recording site. Colored traces indicate unmasked spikes.

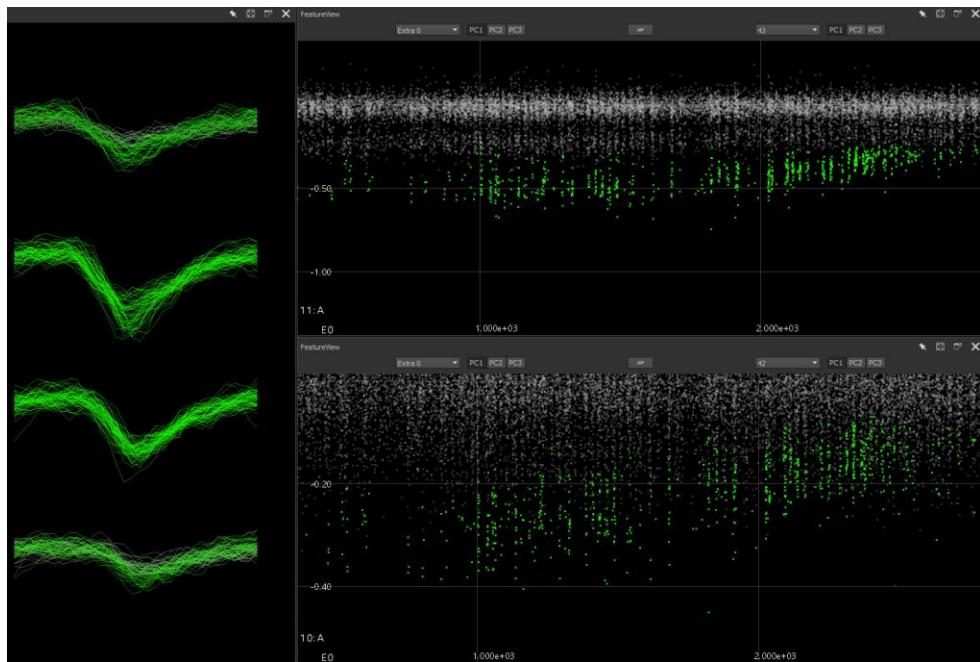
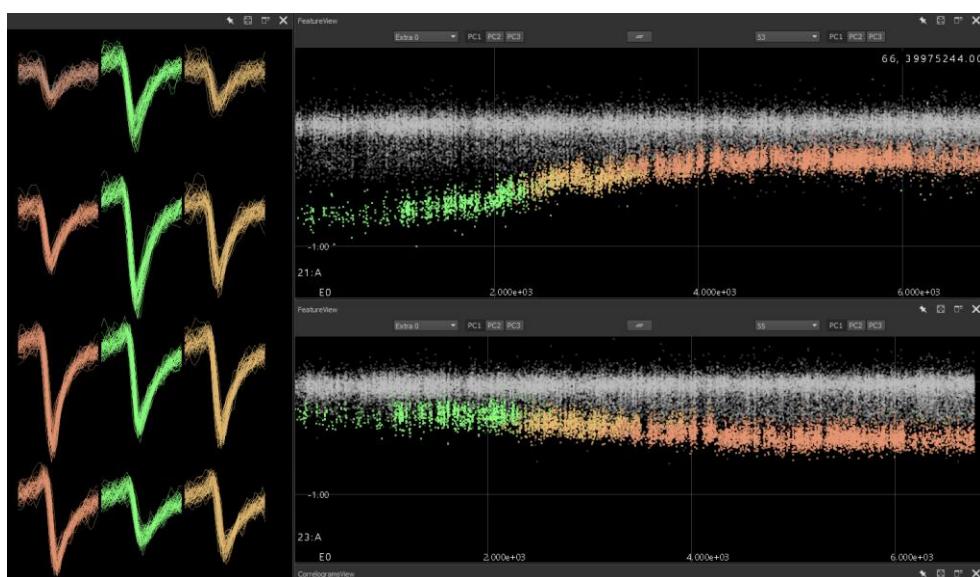
**A****B**

**Supplementary figure 10:** Spike sorting of 16-site data in LGN of anesthetized marmoset. **A**, diagram of one 16-site linear probe shank. **B**, Superimposed waveforms of spikes belonging to individual clusters. Each column corresponds to a clustered unit, while each row corresponds to a recording site. Colored traces indicate unmasked spikes.

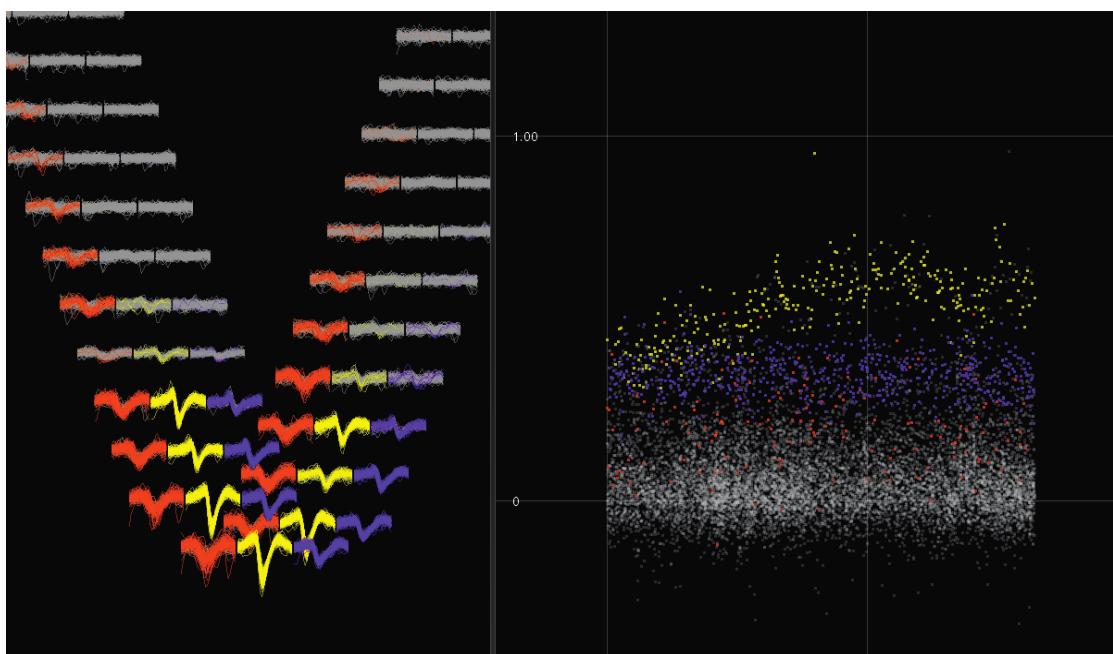
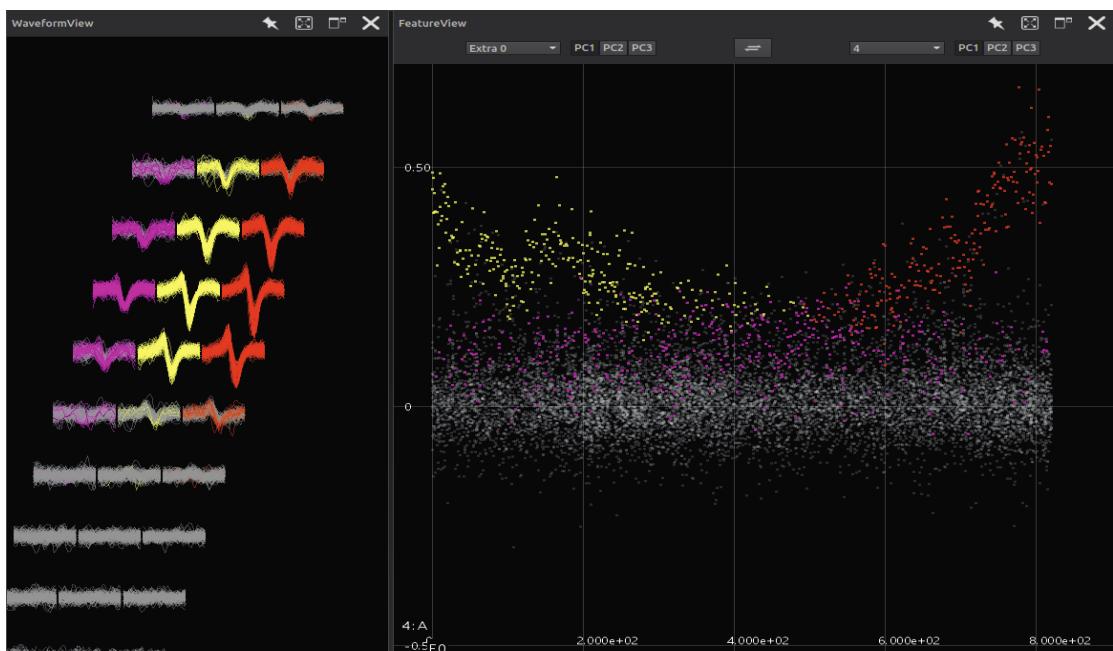


**Supplementary figure 11: Clustering performance for hybrid data (macaque and marmoset).** **A,**

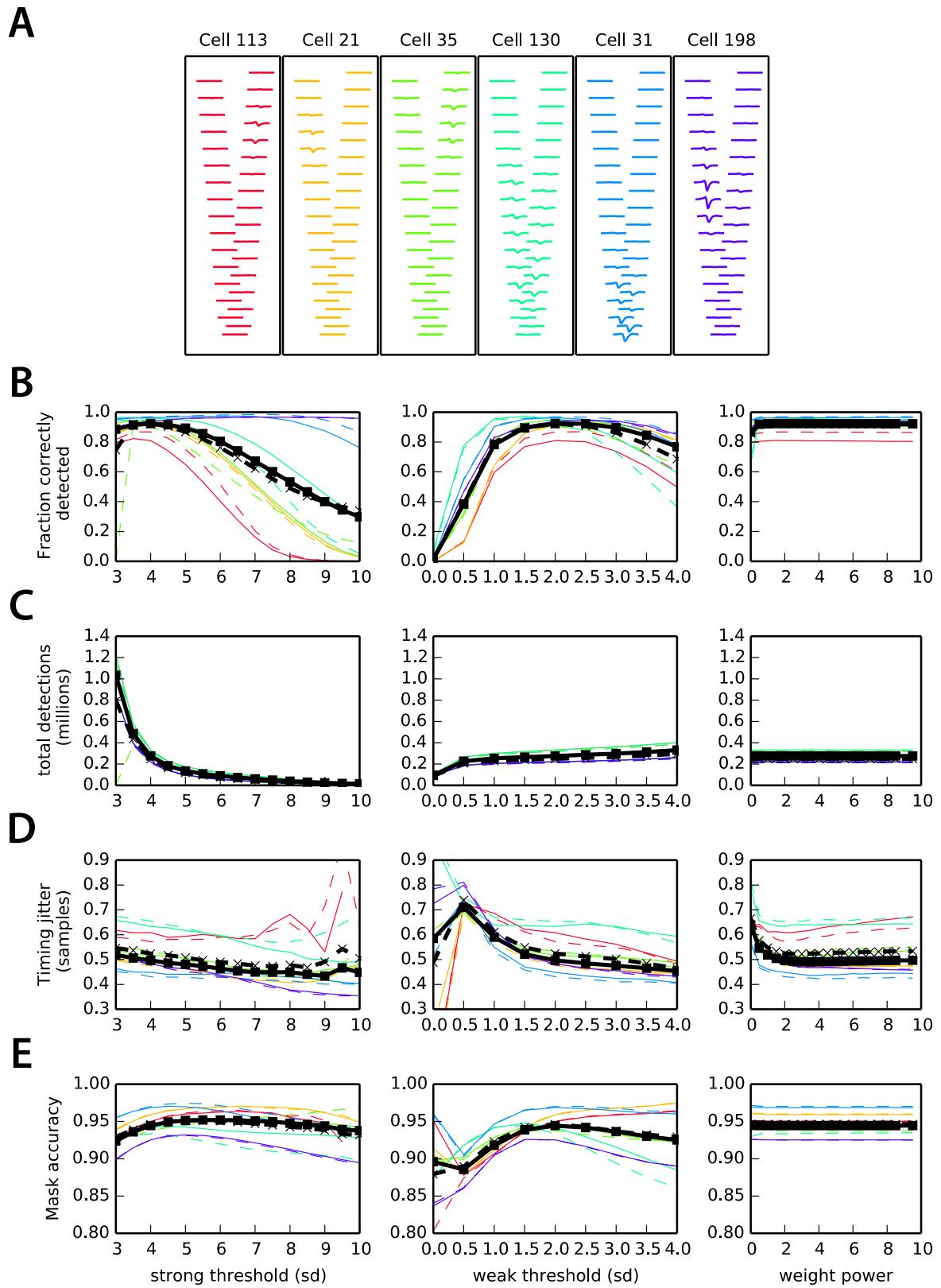
Theoretically optimal performance (cyan line) and performance of the clustering software (red open circle) for example hybrid unit taken from “poly2” probe in macaque V1. **B,** Comparison of actual performance and theoretical optimum across all hybrids. Red, macaque V1 Poly2 probe; green, macaque V1 Edge probe; Purple; marmoset LGN, linear probe. Error rates were statistically indistinguishable from the theoretical optimum ( $p=0.72$ , t-test).

**A****B**

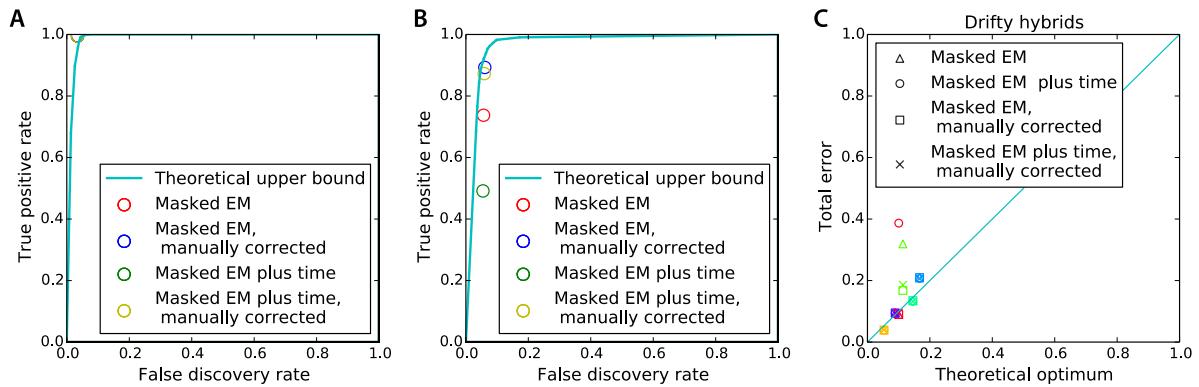
**Supplementary figure 12: Examples of cluster drift in real-world data. A,** Klustaviewa screenshot showing the output of KlustaKwik (without manual curation) for a cell recorded acutely in mouse V1, whose amplitude drifted, and the drift was tracked without requiring manual correction. **B,** Klustaviewa screenshot showing the output of KlustaKwik (without manual curation), in a case where the single cell to be split into multiple clusters requiring manual merging.

**A****B**

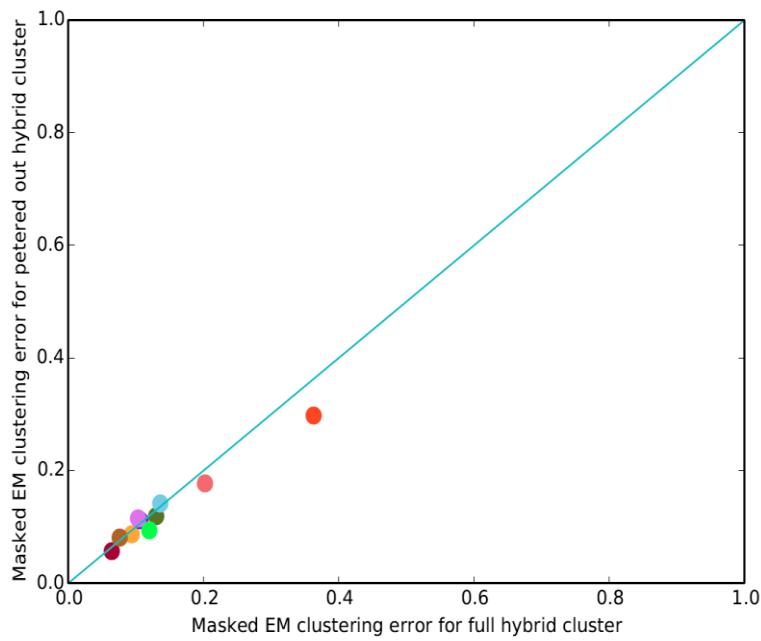
**Supplementary figure 13: Simulation of electrode drift.** Drift was simulated by modulating the amplitude of the hybrid spikes by a geometric random walk, such that the mean amplitude remained constant. **A.** Screen capture from KlustaViewa showing the output of KlustaKwik prior to manual correction. The panel shows an example hybrid cell that did not require manual correction (yellow) together with two possible confounding units (orange and purple), plus all other spikes (white). **B.** An example cell that did require manual merging, which was split by KlustaKwik into two clusters (yellow and orange). Magenta shows a possible confounding unit, and white shows all other spikes.



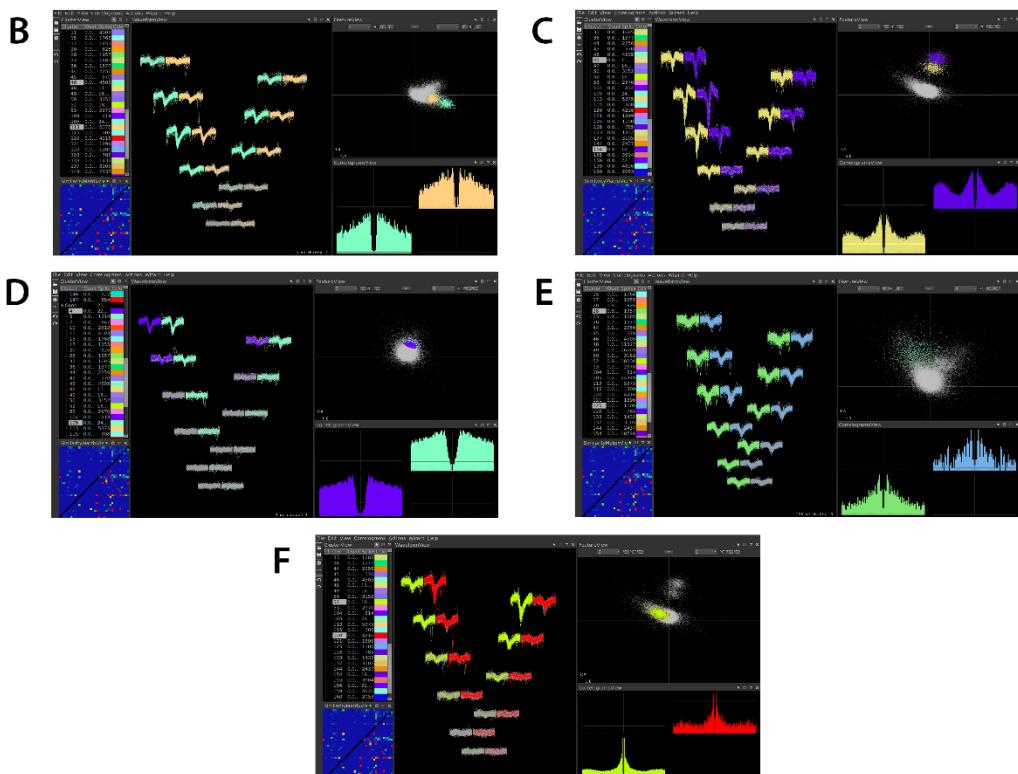
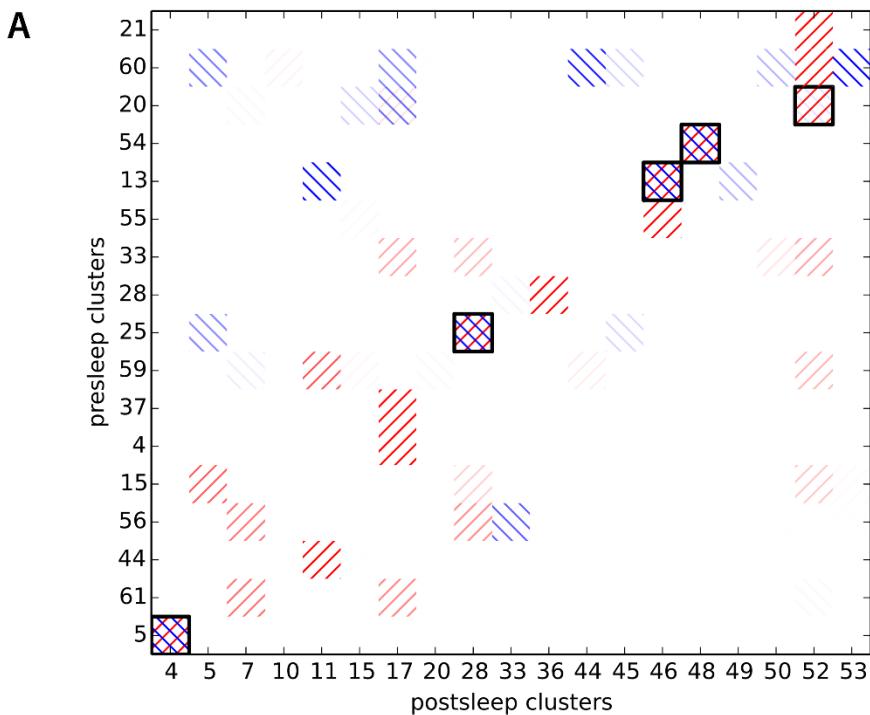
**Supplementary figure 14: Detection performance for simulations of electrode drift.** **A**, Hybrid spike waveforms. **B-E**, Statistics assessing spike detection accuracy for non-drift case (solid lines), and drift simulations (dashed lines). Colors are as in (A); black indicates an average over all cells. Drift has no major impact on spike detection performance.



**Supplementary figure 15: Including time as a feature did not improve clustering performance for drifty cells.** **A**, For cell 31 (which did not require manual correction), including time as a feature made no difference to performance (only one circle is visible as performance was equal for all conditions). **B**, For cell 35, manual correction was required (see supplementary figure 13B). Including time as a feature worsened performance, both before and after manual correction. **C**, Summary statistics, showing that while performance of the masked EM algorithm is comparable to theoretical optimum (after manual correction if required), adding time as a feature serves no benefit. Symbol colors correspond to the code in supplementary figure 14A; symbol shapes as shown in legend.



**Supplementary figure 16: Hybrid cell analysis for a different drift model.** To model a different type of drift, we performed a new analysis in which hybrid spikes are inserted for only half of a recording, and “peter out” in the second half. Error rates were statistically indistinguishable whether or not the spikes petered out ( $p=0.75$ ; t-test).



**Supplementary figure 17: Using the similarity metric to suggest correspondences between multiple recordings.** Data was used from a CA1 recording of sleep on two consecutive days (“presleep” and “postsleep”), which were clustered separately. **A**, Clusters from the two recordings were combined and the similarity metric was computed. For each cluster, the most similar cluster in the other recording was found; because the similarity metric is asymmetrical, this need not be a reciprocal relationship. Blue shading indicates the most likely candidate cluster in the second experiment for each of the clusters in the first experiment; red shading diagonal represents the reverse comparison; color intensity represents the similarity score (range 0-1). A correspondence was suggested when both highest-likelihood pairs are the same, indicated by a black border drawn around the pair. The conservative nature of this metric is indicated by the small number of correspondences suggested. **B-F**: Screenshots of the suggested correspondences viewed in KlustaViewa, showing more or less evidence of electrode drift. Note that while the waveforms may be quite different, the similarity of the autocorrelogram shapes suggests that these are indeed the same unit. **B**: 5 (presleep) vs 4 (postsleep), **C**: 54 (presleep) vs 48 (postsleep), **D**: 25 (presleep) vs 28 (postsleep), **E**: 13 (presleep) vs 46 (postsleep), **F**: 20 (presleep) vs 52 (postsleep).

Cell	113	125	21	35	203	50	130	20	31	198
Preferred side	R	L	L	R	R	L	R	L	L	L
Right side included	✓	✓		✓	✓	✓	✓		✓	
Left side included		✓	✓			✓	✓	✓	✓	✓

**Supplementary Table 1: Analysis of simulated single-sided probes.** To simulate the effect of a single-sided electrode, we removed the channels from one side of the electrode and recomputed all analyses. Because some of the hybrid cells only had substantial energy on one side of the probe, they gave essentially zero performance when the wrong side was taken; these were excluded from further analysis. The table shows, for each hybrid cell, the side of the probe that gave best performance, and whether the unpreferred side was also included in statistical analysis.