An Optrode Array for Spatiotemporally Precise Large-Scale Optogenetic Stimulation of Deep Cortical Layers in Non-human Primates

Andrew M. Clark[1], Alexander Ingold[1], Christopher F. Reiche[2], Donald Cundy III[1], Justin L. Balsor[1], Frederick Federer[1], Niall McAlinden[3], Yunzhou Cheng[3], John D. Rolston[4], Loren Rieth[5,6], Martin D. Dawson[3], Keith Mathieson[3], Steve Blair[2*], and Alessandra Angelucci[1*]

[1] Department of Ophthalmology and Visual Science, Moran Eye Institute, University of Utah, Salt Lake City, UT

[2] Department of Electrical and Computer Engineering, University of Utah, Salt Lake City, UT

[3] SUPA, Institute of Photonics, Department of Physics, University of Strathclyde, Glasgow, UK

[4] Department of Neurosurgery, University of Utah, Salt Lake City, UT

[5] Mechanical and Aerospace Engineering, West Virginia University, Morgantown, WV

[6] Feinstein Institute for Medical Research, Manhasset, NY

*Corresponding authors.

Corresponding authors' address:

A. Angelucci:

65 Mario Capecchi Drive
Salt Lake City, UT 84132, USA
Tel: (801) 5857489
Email: alessandra.angelucci@hsc.utah.edu

S. Blair:

50 S. Central Campus Dr., Rm. 2110
Salt Lake City, UT 84112, USA
Tel: (801) 5856157
Email: blair@ece.utah.edu

ABSTRACT

Optogenetics has transformed studies of neural circuit function, but remains challenging to apply in non-human primates (NHPs). A major challenge is delivering intense and spatially precise patterned photostimulation across large volumes in deep tissue. Here, we have developed and tested the Utah Optrode Array (UOA) to meet this critical need. The UOA is a 10×10 glass waveguide array bonded to an electrically-addressable µLED array. *In vivo* electrophysiology and immediate early gene (c-fos) immunohistochemistry demonstrate that the UOA allows for large-scale spatiotemporally precise neuromodulation of deep tissue in macaque primary visual cortex. Specifically, the UOA permits either focal (confined to single layers or columns), or large-scale (across multiple layers or columns) photostimulation of deep cortical layers, simply by varying the number of simultaneously activated µLEDs and/or the light irradiance. These results establish the UOA as a powerful tool for studying targeted neural populations within single or across multiple deep layers in complex NHP circuits.

Keywords

Optogenetics, macaque, neocortex, feedback, AAV, cFos, current source density, multi-channel electrophysiology

A central goal in neuroscience is to understand how neural circuits generate the computations underlying perception and behavior. Optogenetics has transformed the study of neural circuit function by allowing for the selective modulation of neural activity on a physiologically relevant timescale[1]. While very successful in mice, its application to non-genetically tractable model species, such as the non-human primate (NHP), has lagged behind[2]. Extending optogenetics to NHP studies is crucial, as, due to their similarity to humans, NHPs represent the most important model for understanding neural circuit function and dysfunction[3-6], and provide an essential technology testbed towards human use[7] and the potential application of optogenetics as therapeutic interventions in humans[8,9]. The continuing refinement of viral methods for selectively delivering opsins to particular neural circuit elements, either defined projections[10,11] or particular cell types[12-14], is rapidly opening up new opportunities to study neural circuits in NHPs[2,15]. Despite these advances, one of the most significant remaining obstacles in the application of optogenetics to NHPs is the lack of devices for reliably delivering light to deep neural tissue across relatively large brain volumes with both: (i) sufficient intensity to reach optogenetic activation threshold, and (ii) sufficient spatial resolution to selectively modulate relevant circuit elements. For example, it is presently challenging to optogenetically perturb the activity of neuronal populations in the deep layers of the cerebral cortex across large cortical volumes.

There are several features of cortical networks that both make development of such a device critical and provide design requirements for the device First, a subset of cortico-cortical feedback connections, which are critical for the contextual modulation of sensory processing[10,16], as well as cognitive phenomena such as selective attention[17] and working memory[18], arises from deep cortical layers[19]. Selective optogenetic modulation of these projections would greatly inform work examining the role of cortico-cortical feedback in neural computation. Second, cortico-thalamic projections, which are part of thalamo-cortico-thalamic loops (ubiquitous features of mammalian neural networks), arise exclusively from deep layers[20]. Dissecting these circuits will require selective perturbation of deep layer cortical neurons with high spatiotemporal precision. Finally, in NHP sensory cortex, specific stimulus features are mapped in an orderly, columnar, fashion. That is, nearby neurons prefer similar stimulus features, and this pattern of selectivity extends throughout the cortical depth[21]. While methods for selective illumination of the cortical surface at the spatial scale of these columns have recently been

93 developed in smaller animals[22,23], they unfortunately only allow for stimulation of the superficial
94 layers in the NHP.

95 State-of-the-art optogenetic experiments in NHPs have mainly followed two light
96 delivery approaches: through-surface illumination and penetrating probes. Surface
97 photostimulation utilizes either a laser- or LED-coupled optical fiber positioned above the
98 cortical surface[10], or chronically-implantable surface LED arrays[24]. These approaches enable
99 photoactivation of a large area, but only to a depth of <1mm, due to light attenuation and
100 scattering in tissue, as well as to unintended superficial layer neuron activation and even damage
101 caused by the heat generated by the light intensity required to reach deeper layers[10,25]. In
102 contrast, penetrating optical fibers, integrated with single[26,27] or multiple[28] recording probes,
103 allow photoactivation at depths >1mm, but only of a small tissue volume (a few hundred microns
104 in diameter), and due to their size and shape, can cause significant superficial layer damage.

105 To overcome the limitations of these light delivery methods, we have developed the Utah
106 Optrode Array (UOA), a 10x10 array of glass needle shanks tiling a 4x4 mm^2 area and bonded to
107 an electrically-addressable μLED array independently delivering light through each shank[29,30].
108 Device bench testing and *in vivo* testing in macaque primary visual cortex (V1) demonstrate that
109 the UOA allows for spatiotemporally patterned photostimulation of deep cortical layers with sub-
110 millimeter resolution (at the scale of single layers and columns) over a large expanse of cortical
111 tissue. This spatial selectivity can be scaled up to multiple layers and columns simply by varying
112 the number of simultaneously activated μLEDs and/or the light irradiance. These results
113 establish the UOA as a powerful tool for studying local and large-scale populations of deep layer
114 neurons in NHP cortex.

RESULTS

The UOA: Geometry and Optical Properties

Figure 1 about here

The spatiotemporally multiplexed UOA was developed based on the geometry of the widely utilized Utah Electrode Array (UEA)[31]. The UOA is a 10x10 array of penetrating optical light guides made in glass, with customizable length (up to 2.5mm) and shank width (80-120µm diameter) on a 400µm pitch tiling a 4x4mm area. For deep layer light delivery, a custom blue (~450nm) µLED array grown on sapphire is directly integrated with the device, with each electrically addressable 80µm µLED delivering light through a single needle shank (**Fig. 1A-E**). A second 9x9 array of "interstitial" µLEDs can be interleaved on the same device and matrix-addressed for independent surface stimulation (as shown in **Fig. 1B**, but the interstitial µLED array was not used in this study). To limit the spatial spread of light coupled into the optrode, the first generation UOA used a metal pinhole array[29]. Bench testing of this first generation device demonstrated the capacity for delivering irradiances in excess of activation threshold across a range of commonly employed depolarizing[32] and hyperpolarizing[33] opsins. Estimates from bench testing suggested a 50% decrease in irradiance within tissue within approximately 200µm of a needle tip and a beam cross section of approximately 150-200µm (full-width at half-maximum - FWHM)[29]. These initial results suggested that direct optogenetic activation through the UOA is on a spatial scale commensurate with the functional architecture of primate cortex.

Building upon these initial results, we have developed the second generation UOA reported here, which incorporates an optically opaque (silicon) interposer layer with optical "vias" on the array backplane to eliminate uncontrolled surface illumination and inter-needle crosstalk, and facilitate alignment during bonding to the µLED-array (**Fig. 1A, C**; see Online Methods for manufacturing details). This second-generation device **(Fig. 1A-E**) was bench tested (**Fig. 1F**) and its *in vivo* optical performance was estimated via ray tracing (**Fig. 1G**). From optical measurements, the output power (in mW) emitted through the tip of each needle in the 10x10 array at different drive voltages is shown in **Fig. 1F** and in **Extended Data Fig. 1**, where we also show the estimated output irradiance (mW/mm^2) for each needle tip (assuming a pyramidal tip and uniform light output). At 3V, estimated irradiance levels are below the 1 mW/mm^2 threshold for the excitatory opsin Channelrhodopsin-2 (ChR2); average output optical

power across all needles ± SD = 0.0057±0.004 mW, corresponding to an average estimated irradiance of 0.21±0.14 mW/mm^2 (**Extended Data Table 1**). There is variation across the array, due primarily to variations in the resistance (and therefore slope efficiency) of each µLED. At 3.5V, about 30% of the stimulation sites reach or exceed this threshold (average optical power ± SD = 0.022 ± 0.013 mW; average irradiance: 0.82 ± 0.49 mW/mm^2), while at 5V, more than 90% of the sites emit above threshold (average optical power ± SD = 0.1 ± 0.056 mW; average irradiance: 3.79 ± 2.08 mW/mm^2). In principle, software modifications in the matrix driver interface could be made to better equalize stimulation levels across the array.

We used optical ray tracing to estimate the performance of the device *in vivo* based upon calibration curves obtained from bench testing (see Online Methods for details). In particular, we were interested in estimating the direct neural stimulation volume (based upon the local irradiance in tissue) as a function of drive voltage and pattern of lit needles in order to interpret the *in vivo* results. These results are shown in **Figure 1G**, where the left column panels show the stimulation volume along the first UOA column as produced by the needle (column 1, row 8) nearest one of the electrode penetrations (penetration 2, or P2) in the *in vivo* experiments, and the right column panels show the activation volume when the entire UOA column 1 is lit. Each row of figures is for a different drive voltage. At low drive voltage (~3V, equal to 38% of the maximum input voltage used), highly localized stimulation in tissue near the needle tips is produced (note also that the irradiance across the tip surface is non-uniform – concentrated near the apex – explaining why above-threshold irradiance levels can be achieved at 3V as demonstrated in the *in vivo* experiments). At higher voltages (~5V/64% max intensity and above), the stimulation volume overlaps that of adjacent needles, while also extending deeper into tissue. When driving an entire column, at 3V, stimulation localized near each tip is mostly retained, whereas a nearly continuous stimulation volume is obtained at 3.2V due to overlapping fields. At 5V (64% of max intensity) and 7.8V (100% max intensity), the depth of this continuous volume increases, both above and below the tips.

In Vivo Testing: Electrophysiology

Figure 2 about here

We used *in vivo* linear electrode array (LEA) recordings to assess the utility of UOAs for precise modulation of neural activity in deep cortical layers of monkey cortex expressing the excitatory opsin *Channelrhodopsin-2* (ChR2). ChR2 was expressed in area V1 of a macaque monkey by injecting a mixture of Cre-expressing and Cre-dependent adeno-associated virus (AAV9) carrying the genes for ChR2 and the red reporter protein tdTomato (tdT). We have previously shown that this viral mixture results in almost exclusive anterograde neuronal infection at the injected site[10]. Following a survival period to allow for viral expression, we recorded neural spiking activity using a 24-contact LEA inserted nearby a UOA (fully integrated with a µLED array, as described above and in **Fig. 1A-E**) that was implanted into a region of dense ChR2 expression in V1 (**Fig. 2A-C; Extended Data Figs. 2-3A**). We performed 3 LEA penetrations (P1-P3), but for only 2 of them (P2, P3) photostimulation via the UOA modulated neural activity, likely because P1 was located farthest from the region of tdT/ChR2 expression, as revealed by postmortem histology (**Extended Data Fig. 3A**). Below we report data from P2 and P3.

Comparison of Surface and UOA Photostimulation

Figure 2D shows neural responses recorded in P2 to simultaneous stimulation of µLEDs at all UOA sites at an irradiance level (average ± SD across the whole UOA = 7.4 ± 3.19 mW/mm^2, induced by an input intensity of 7.8V; see **Extended Data Table 1**) almost an order of magnitude greater than ChR2 activation threshold[32]. To examine the spatiotemporal distribution of responses to UOA stimulation across V1 layers, we first performed a current source density (CSD) analysis of the local field potential (LFP) recorded across the LEA around the time of a UOA pulse. The CSD, computed as the second spatial derivative of the LFP, reveals the location of current sinks (negative voltage deflections reflecting neuronal depolarization) and sources (positive voltage deflections reflecting return currents) throughout the cortical depth. Current sinks in response to UOA stimulation were confined to L4C and below (**Fig. 2D,** Left), suggesting that direct optogenetic activation was confined to a region below the UOA needle tips. Analysis of the multiunit (MU) spiking activity across contacts demonstrated strong phasic responses to UOA stimulation that extended from L4C to the white matter boundary (**Fig. 2D**, Right). **Figure 2E**, shows the CSD and MU signals under the same photostimulation and recording conditions as in **Figure 2D**, but at lower light irradiance (0.82 ± 0.49 mW/mm^2 -

3.5V). At this lower intensity, CSD signals and MU activity were mostly localized to L4C and the lower part of the deep layers, with L4C activation preceding in time that of deeper layers. These laminar patterns of neural activity suggest that the UOA needle tips closest to P2 terminated in L4C and that at low photostimulation intensities light spread nearby the UOA tips. Additional analysis demonstrating that response onset latency and light activation threshold were lowest for the P2 contacts located in L4C, together with postmortem histological assessment, further confirmed that the UOA needle tips closest to P2 were indeed located in L4C (**Extended Data Fig. 3A, B Right**). Comparison of the above laminar patterns of response with that elicited by direct surface stimulation in a different animal at a lower suprathreshold irradiance ($2.2mW/mm^2$) revealed a sharp dissociation. Specifically, surface stimulation of ChR2 evoked responses starting in superficial layers and terminating in L4C **(Fig. 2F**).

UOA Stimulation Parameters Can Be Tuned to Achieve Laminar Specificity

To systematically assess the impact of UOA stimulation on neural spiking responses we varied: (i) the spatial pattern of UOA stimulation, from single µLED sites, to entire columns, to the entire device, and (ii) stimulation intensity across different spatial patterns. In all conditions, we used phasic stimulation (5Hz, 100 msec pulses for 1 sec with 1-10 sec inter-trial intervals, with the longer intervals used at the higher stimulation intensities) with a slow on/off ramping to eliminate the potential of any electrical artifacts induced by capacitive coupling at the array/tissue interface[34]. As an example, **Figure 2G-J** shows responses from P2. An analysis of firing rate increase across layers induced by systematically stimulating a single µLED along column 1, indicated that the UOA needles in column 1 closest to P2 were those in rows 8 and 9 (C1-R8, C1-R9), and that the tips of these needles terminated into L4C (**Extended Data Fig. 3B Left**). The laminar distributions of MU activity in P2 varied considerably in strength across conditions, but were reliably confined to deeper layers. By varying the spatial pattern of stimulation and/or the stimulation intensity, MU activation could be confined to single layers or spread across multiple layers. For example, activation of the whole µLED array (**Fig. 2G**) at low light intensity (>2.8 and up to 5V), evoked a peak of MU activation localized to L4C, the layer of needle tip terminations nearest P2, and this peak increased in magnitude with increasing stimulation intensity. At and near the peak of this intensity range, a second smaller peak of MU activation was also present in L6 but not in L5. However, when the whole UOA was activated at

the highest intensity used (7.8V), the MU spiking peak in L4C decreased in magnitude, while activation of L5 and L6 increased. In macaque V1, L4C projects to both L5 and L6[35], but its net effect is to suppress the former[36] and activate the latter[37], consistent with the interpretation that at lower light intensities lack of L5 responses and increases in L6 responses may have resulted from synaptic spread from L4C neurons directly activated by UOA stimulation. In contrast, at the higher photostimulation intensity (black curve in **Fig. 2G**) light scattering through a larger volume may have directly contributed to firing rate increase in deeper layers 5 and 6, while the reduced MU peak in L4C could have resulted from activation of higher threshold inhibitory networks. In a later section of the Results, we provide evidence supporting this interpretation. Intensities ≥ 5V evoked similar laminar patterns and magnitude of neural activity irrespective of whether a single µLED or an entire column nearest the LEA or the whole UOA were illuminated (**Fig. 2G, H, J**). However, at lower photostimulation intensities, for a given intensity, firing rate increased with the number of µLEDs activated (e.g. compare blue curves in **Fig. 2G, H, J**), and higher intensities (>3.2V) were required to modulate neural activity via a single active µLED (**Fig. 2J**). Moving the µLED activation column from 1 to 5 resulted in a 10 times reduction in the amplitude of MU activity (**Fig. 2I**). No increase in firing rate could be evoked by stimulation of an entire column beyond column 5 or a single µLED in column 1 beyond row 4 (distances from the LEA of about 2.6-2.7 mm – estimated on the postmortem histology) even at the highest drive voltage used, i.e. 7.8V (corresponding to an average irradiance of 6-11 mW/mm^2, see **Extended Data Table 1**).

Tangential Extent of Responses Induced by Photostimulation Via the UOA

Figure 3 about here

Having determined that C1-R8 and C1-R9 were the UOA needles in column 1 closest to P2, and that these needles' tips terminated in L4C, we then performed a similar analysis to determine the location of P3 relative to the UOA. The results of this analysis, reported in **Extended Data Fig. 3C**, suggested that C1-R7 was the UOA needle in column 1 closest to P3 and that this needle tip terminated in the superficial layers.

We then asked whether MU responses across LEA contacts were tuned for the spatial site of UOA stimulation, how such tuning varies with intensity, and whether UOA stimulation can be

tuned to selectively activate volumes on the scale of functional columns in NHP cortex. Moreover, different spatial tuning across LEA contacts would provide us with information about the angle of LEA insertion into cortex relative to the UOA. We performed this analysis separately for each penetration.

To estimate the MU response to stimulation at UOA sites between columns 1-5 and rows 3-10, we fit a multiple linear regression model to the MU spiking recorded at each LEA contact, with row, column, and intensity (volts) as independent variables (see Online Methods). We included in this analysis only contacts on which there was a significant difference in MU firing rates during the stimulation and control periods for at least one of the row or column conditions (ANOVA, $p < 0.01$). On average, including a quadratic term explained more of the variance in the MU response (mean $R^2 \pm$ SD: 0.58 ± 0.14 vs. 0.31 ± 0.11 for a linear model; Kolmogorov-Smirnov, $p < 10^{-7}$). **Figure 3A, E** shows plots of MU responses, evoked by a mid-level photostimulation intensity (3.5V), estimated from the fitted data for the contact in P2 and P3 that showed the greatest relative response modulation. We normalized each contact's fitted responses to the peak and averaged across contacts to get a sense of whether MU responses preferred stimulation at different UOA sites on different LEA penetrations (**Fig. 3B**, **F**). Consistent with postmortem histological assessment (**Extended Data Fig. 3A**) and the analysis in **Extended Data Figure 3B-C**, the peaks for P2 contacts tended to cluster mostly near columns 1-2 and rows 8-9, while those for P3 contacts clustered mostly near columns 1-3 and rows 4-7, and the spatial pattern of peak activity suggested that for both penetrations, particularly P3, the LEA was inserted at a slightly oblique angle. Statistical analysis revealed that peak locations differed significantly across the two penetrations (ANOVA, $p < 0.01$).

The data in **Figures 2H-I, 3A, B, E, F** suggested that the magnitude of evoked MU responses decreased with increasing distance between photostimulation and recording sites. To quantify this observation and better characterize the extent of photostimulation-evoked responses across the tangential domain of the cortical tissue, as well as how this extent is affected by photostimulation intensity, we examined response amplitude as a function of distance on the UOA (in a straight line extending along either the row or column axis) from the LED site that evoked the peak response. Plots of relative response (percent of the peak fitted response) as a function of distance in either the column or the row direction, sorted by input intensity, are shown in **Figure 3C-H** separately for each penetration. As is evident from the steeper decrease

in responses along the column versus the row axis, as well as the difference in relative response across stimulus intensities, there was a significant main effect of UOA axis and input intensity on this relative response measure (ANOVA, both $p < 10^{-21}$) as well as a significant difference across penetration (ANOVA, $p < 10^{-14}$). Finally, there was a significant interaction between intensity and UOA axis as well as UOA axis and penetration (ANOVA, both $p < 0.01$). These results indicate that the response decrease from peak is greater in the column versus the row direction, that intensity has a different effect on this drop-off when looking at the row and column directions, and that this differed across penetrations. In the column direction, at the lowest stimulation intensity (2.8V), evoked responses dropped to 16% of the peak at a distance of about 1.6mm from the location evoking peak response, but at the higher intensities (5-7.8V) responses only dropped to about 50% of the peak at the same distance (**Fig. 3C,G**). In contrast, in the row direction, at the lowest stimulation intensity (2.8V), evoked responses only dropped to about 80% of the peak at a distance of about 2.8mm from the peak location, and to about 90% at ≥5V intensity (5-7.8V) (**Fig. 3D,H**). The difference in response drop-off with distance in the column vs. row directions is likely explained by the greater differences in irradiance, for a given input intensity, along the column as compared to the row axis (see **Extended Data Fig. 1**).

In summary, the spatial spread of MU activation along the tangential domain of cortex varied according to UOA stimulation site and intensity. Importantly, the extent of this spread was more limited at lower intensities, suggesting that increasing intensity increased the volume over which cells were optogenetically activated, consistent with the model simulations in **Fig. 1G**.

UOA Activation Parameters Can Be Tuned to Activate Distinct Cortical Networks

Given the spatial separation between the LEA and the UOA (estimated to be about 1-1.1mm for P2 and 700-800µm for P3 on the histological sections; **Extended Data Fig. 3**), the reported sharp falloff in light intensity over short distances in tissue[38,39], and our bench estimates of light spread from the UOA tips (≤ 600µm laterally and radially from the needle tip at intensities ≤ 5V, and 0.9-1.0mm at 7.8V intensity; **Fig. 1G** and[29]), we reasoned that the UOA-driven MU activity we observed, at least at ≤ 5V photostimulation intensities across all contacts and at any intensity in the deeper layer contacts, was not caused by direct activation of ChR2-expressing cells nearby the LEA recording site, but rather by indirect activation of ChR2-expressing cells in the vicinity of the UOA needle tips. Given the greater spread of light at the highest intensity used, direct

activation of neurons recorded at LEA contacts closest to the UOA tips was, however, possible. Thus, to estimate the tangential spread of direct as well as indirect activation to the LEA recording sites, we measured the onset latency of UOA-driven MU responses across layers.

Figure 4 about here

Example latency data from P2 are shown in **Figure 4A**. Here, the UOA stimulus was a single µLED (C1-R8) nearest the recording location with an input voltage of 5V. Pulse-aligned MU rasters on each contact are shown next to a schematic of the recording LEA. The fastest response occurred in mid layers with an onset latency of about 15 msec, suggesting the recorded neurons in these layers were activated indirectly via ChR2-expressing cells in mid-layers nearby the C1-R8 needle tip. Moreover, deep layer response onset (mean ± s.e.m: 30 ± 7 msec) lagged that in mid-layers, as would be expected if optogenetic activation first propagated through L4C before being synaptically relayed to deeper layers, via L4C-to-L5/6 connections. Averaged PSTHs for the peri-pulse period on one example L4C and one L6 contacts are shown in **Figure 4B**. There was a significant pulse-by-pulse difference in onset latency across contacts (ANOVA, $p < 10^{-30)}$, as well as a significant pairwise difference across these two LEA recording sites (Tukey HSD test, $p < 10^{-6}$; **Fig. 4B**, Right).

To better visualize onset latency of evoked responses throughout V1, **Figure 4C** shows average peri-pulse PSTHs across all LEA contacts as a function of normalized cortical depth for an example whole µLED array (top panels), single column (middle panels), and single µLED (bottom panels) experiment at different photostimulation intensities and different spatial separations between UOA stimulation and LEA sites. The primary effects of increasing total stimulus area at the lower intensities were to increase the number of responsive contacts and the amplitude of driven responses, as well as to shorten onset latencies (e.g. compare panels in the left column of **Fig. 4C**). At higher intensities (e.g. 5V, middle column), there was little change in these parameters across these large differences in total stimulated area. The main effect of decreasing the stimulus intensity for a fixed area (**Fig. 4C** compare middle to left columns), or increasing the separation between the stimulated UOA site/s and the LEA for a fixed stimulus intensity (**Fig. 4C** compare middle to right panels in the center and bottom rows) was an increased delay in onset latencies across all contacts. For example, mean onset latency at 5V and 3.2V across all contacts was 17 ± 1.7 msec and 25.4 ± 2 msec for the whole array condition, 19.8

± 1.4 msec and 37.5 ± 1.9 msec for the C1 condition, and 21.4 ± 2.3 msec and 74.1 ± 1.6 msec for the single µLED (C1-R8) condition. Mean onset latency at 5V for the C3 and C1-R6 conditions was 47.6 ± 4.3 msec and 59.4 ± 4.1 msec, respectively. Calculating onset latency on a pulse-by-pulse basis and looking at the effects on latency of cortical depth, stimulation pattern, and stimulation intensity, we observed significant main effects of pattern and intensity, significant two-way interactions between depth and pattern, depth and intensity, pattern and intensity, and a significant three-way interaction between all three factors (ANOVA, all $p < 10^{-4}$). Limiting our analysis to each pattern, we observed a significant main effect of intensity and distance from the LEA on onset latency for the single column conditions in **Fig. 4C** (ANOVA, all $p < 10^{-4}$), and a significant main effect of distance for the single µLED conditions (ANOVA, $p = 0.03$). Furthermore, in many conditions, pairwise comparisons across contacts revealed a delayed response onset in deep layers (and in superficial layers in some conditions) relative to mid-layers; this time lag varied with intensity and distance of the stimulation site from the LEA, increasing at lower intensities and greater distances. These differences in onset latency between the fastest mid-layer response and those in L6 were statistically significant for most conditions shown in **Fig. 4C** at 5V, and for some at 3.2V (Tukey Kramer, all $p < 0.01$; see **Extended Data Fig. 4** for specific comparisons). There was also a significant difference in onset latency between mid- and superficial layers in some conditions, namely C1 at 5V and the whole array at 5V and 3.2V (Tukey Kramer, all $p < 0.01$; **Extended Data Fig. 4**). Notably, however, when the whole µLED array was stimulated at the highest intensity used (7.8V), there was no difference in onset latencies between the deep and middle layers, suggesting the former were directly activated by light spreading through deeper tissue in this condition (**Fig. 4C** top right, and **Extended Data Fig. 4**). The fastest onset latency in this condition was 11 msec, possibly a long enough delay to suggest that light did not directly activate the L4C neurons recorded at the LEA site; however, a shorter latency, indicative of direct light activation of L4C contacts in this condition, is also possible given the slow on ramping of the photostimulating current used in our experiments, and the possibility that a lower threshold criterion would yield shorter latencies.

Figure 5 about here

To quantify these effects across the population (n= 33 significantly responsive contacts, across 2 LEA penetrations), we first calculated the distance between each LEA contact and the contact with the shortest onset latency, and plotted this distance versus onset latency, separately for each unique combination of UOA stimulation site(s) and intensity. Similar to the P2 data shown in **Fig. 4C**, the population data showed 2 main effects. (1) Onset latency decreased significantly across all contacts with increasing stimulation intensity (ANOVA, main effect of intensity, all $p < 0.01$; **Fig. 5A, 5B Left, 5C Left**) and proximity to the recording LEA site (ANOVA, main effect of row or column on UOA, all $p < 10^{-4}$; **Fig. 5B Right, 5C Right**). (2) Onset latency increased significantly with contact distance on the LEA from the fastest contact (**Fig. 5A-C**, main effect of distance on the LEA, ANOVA all $p < 0.01$), except for the whole array condition for which, post-hoc comparisons revealed that at the highest intensity there was no significant difference in latency across contacts (as also shown in **Extended Data Fig. 4**). This suggests that at lower irradiance the increase in firing rate caused by direct light started near the site of UOA tip termination (in L4C in P2, but in L3 in P3) and then spread to more distant sites indirectly via interlaminar networks; in contrast, at higher irradiances, lack of difference in onset latency across contacts suggests that light spreads through a larger tissue volume.

Across the three categories of UOA stimulation (whole array, column, and single µLED), only for the whole array and single µLED conditions did we observe a significant interaction between the effects of distance along the LEA and UOA photostimulation intensity on onset latency (**Fig. 5A and C Left**; both $p < 0.05$, ANOVA). Specifically, in these conditions, lowering photostimulation intensity decreased the slope of the curves, indicating that the difference in onset latency with distance on the LEA increased at lower intensity. Additionally, for the single µLED condition, we also observed a significant decrease in the slope of the curves when photostimulating at increasing UOA-LEA separation, but only when we moved the single µLED stimulus to sites that were far enough from the LEA to necessitate stimulation at the very highest powers to elicit any response (dashed lines in **Fig. 5C Right**, µLED in rows 4-7; ANOVA, LEA distance x UOA row x intensity interaction, $< 10^{-3}$). For the single column condition, there was no significant interaction between contact distance and either photostimulation intensity or UOA-LEA separation (**Fig. 5B**; ANOVA, all $p > 0.09$).

In summary, by varying photostimulation intensity and/or number of stimulated sites, the UOA allows activation of single or multiple layers, while by varying the spatial separation

between the site of UOA stimulation and that of the recording, the UOA allows investigations of local vs long-range intra and interlaminar circuits.

In Vivo Testing: c-Fos Expression

Figure 6 about here

As an additional approach to validate the performance of the UOA for large scale photostimulation of deep cortical tissue, we measured changes in c-fos expression, an immediate early gene whose expression rapidly increases when neurons are stressed or activated[40,41]. The product of this gene, c-fos protein, can be identified using immunohistochemistry (IHC) and used as an indirect measure of the spatial pattern of neural activation in post-mortem brains. We analyzed patterns of c-fos expression using IHC (see Online Methods) in two control and two experimental hemispheres from 3 animals. In one of the experimental cases (MK414-RH), a "passive" UOA (a device similar to the one used for the electrophysiological experiments but without an integrated µLED array) was implanted in area V1 of the right hemisphere in a region of high ChR2 expression (**Fig. 6A-B**), 10 weeks after injecting a mixture of Cre-expressing and Cre-dependent AAV9 vectors carrying the genes for ChR2 and tdT. Photostimulation of deep cortical layers via the passive UOA was performed by illuminating a subset of UOA needles using a fiber-coupled blue laser and a collimating lens, while shielding from light the non-stimulated cortex and UOA needles (see Online Methods). About 75 minutes after completion of the photostimulation protocol, the animal was euthanized, and the brain processed for histology and c-fos IHC. Results from this case are shown in **Figure 6A-D**. Histological analysis revealed that the UOA was inserted at an angle (due to the brain curvature), so that its needle tips ended at the bottom of the superficial layers, dorsally, and in progressively deeper layers ventrally, with most of the needle tips ending in L4C and only the most ventral ones reaching into L6 (**Fig. 6A-B**). There was extensive c-fos expression encompassing all cortical layers, and extending far beyond the site of UOA insertion and photostimulation. Specifically, c-fos positive (c-fos+) cells were found throughout V1 (**Fig. 6A, C**), as well as in extrastriate cortical areas known to receive inputs from V1, including V2 (**Fig. 6A, C**), the third-tier cortex and area MT (not shown). Qualitatively, c-fos expression appeared densest at the site of UOA photostimulation in the

superficial and mid-layers, and decreased with distance from this site (**Fig. 6C, D**). This extensive pattern of c-fos+ cells suggested c-fos expression was induced by both direct neural activation by light as well as indirect activation via synaptic activity. To test this hypothesis, as well as to determine the extent of neural activation directly induced by UOA-photostimulation, we repeated the same experiment in a different animal (MK422-RH) in which the AMPA receptor antagonist NBQX was applied to ChR2-expressing cortex prior to passive-UOA insertion and photostimulation, in order to block most glutamatergic synaptic transmission. The UOA was only partially inserted in this case, so that most of its needle tips only reached the bottom of the superficial layers (**Fig. 6E-F**). As expected, C-fos expression in this case was much less extensive than in case MK414-RH, being largely confined to the superficial layers at the site of UOA photostimulation (**Fig. 6G-H**).

To control for potential c-fos activation induced by UOA insertion independent of photostimulation and/or ChR2 activation of neural tissue, in case MK414, we inserted a passive UOA in the supplementary motor area (SMA) of the hemisphere contralateral to the experimental one; the control hemisphere (LH) was not injected with virus, thus did not express opsins. The animal was euthanized 4 hrs following UOA insertion without receiving any photostimulation. Histological analysis revealed that the UOA was fully inserted in this case, its tips reaching into the upper part of the deep layers (L5). C-fos expression encompassed all layers, but was largely confined to the site of insertion (**Fig. 6I-J).** To control for potential c-fos activation induced by light, independent of ChR2-induced neural activation and/or UOA insertion, in a different animal (MK421-RH) we performed surface photostimulation of SMA cortex not expressing ChR2, using a fiber-coupled laser and a collimating lens; no UOA was inserted in this case. We found a few c-fos+ cells immediately underneath the illuminated area largely within L1 (**Fig. 6K-L**). To quantify these results, for each experimental and control case we counted c-fos+ cells in 3 regions of interest (ROIs) encompassing all cortical layers, one centered in the region of UOA insertion and/or light stimulation, the other two located 4 and 8 mm, respectively, from the first one (*white boxes* numbered 1-3 in **Fig. 6A-L**; see Online Methods for details). **Figure 6M** plots the average number of c-fos+ cells across samples, as a function of distance from the UOA insertion site, while **Figure 6N** shows the laminar distributions of c-fos+ neurons at each distance. Confirming our qualitative observations, we found significant local and long-range c-fos expression only in case MK414-RH, which received

photostimulation of ChR2-expressing cortex via the UOA. Application of the glutamate blocker prior to photostimulation prevented long-range c-fos expression, and reduced its expression by 5 fold in the area of UOA stimulation, where it was largely confined to the directly photostimulated layers near the UOA tips. UOA insertion-only led to as much local c-fos expression as the glutamate block case, but to greater interlaminar, as well as intra- and inter-areal long-range spread, suggesting that neurons activated by the insertion trauma, in turn indirectly activated downstream networks. Finally, surface photostimulation of cortex not-expressing ChR2, without UAO insertion, caused virtually no c-fos expression, except for a few cells in L1 and upper L2. Statistical analysis (one way ANOVA with pairwise comparisons and Bonferroni correction for multiple comparisons) revealed a significant difference in the number of c-fos+ cells at each distance between the experimental case (MK414-RH) and the glutamate block case (MK422-RH; $p<0.001$, at all distances) as well as between MK414-RH and each of the control cases (MK414-LH: $p<0.001$, at all distances; MK 421-RH: $p<0.001$, at all distances). There was no significant difference between the glutamate-block and UOA-insertion-only (MK414-LH) cases at any distance ($p=1$ at 0mm, $p=0.23$ at 4 mm, and $p=0.44$ at 8mm distance), but both cases differed significantly from the light-only case (MK421-RH) at 0mm distance ($p<0.05$ for all comparisons). Finally, the number of c-fos+ cells decreased significantly with distance for cases MK414-RH ($p<0.001$), MK422-RH ($p=0.001$), and MK414-LH ($p=0.003$), but not for case MK421-RH ($p=0.079$).

DISCUSSION

We have developed and tested *in vivo* in the NHP cortex a novel device, the UOA, a 10x10 array of penetrating light waveguides with integrated μLEDs, which has the potential to open novel avenues for furthering optogenetic research in larger brain species, particularly NHPs. Current optogenetic approaches in NHPs allow for light delivery either over a large area but at limited depth[10,24], or to deeper tissue but over a small area[26-28,39]. Multi-site optical stimulation probes for larger volume stimulation have also been developed and combined with singe[42] or multisite[43,44] electrical recordings, but these multi-fiber-based approaches are typically cumbersome to assemble and don't easily scale to precisely target multiple small tissue volumes. The UOA combines the advantages of all these approaches. Namely, it allows for both focal and

larger-scale neuronal activation of single or multiple deep layers simply by varying the number of simultaneously activated µLEDs and/or the light irradiance. Moreover, although here we only used the needle-aligned µLED array for deeper layer activation, the integrated interleaved interstitial µLED array allows for selective photostimulation of superficial layers either independently or in conjunction with deep layers.

This novel device has the potential to significantly impact neuroscience allowing dissection of neural circuit function in animal species capable of complex behavior. By design, the UOA is intended to achieve spatial resolution in cortical application in NHPs, and eventually humans, and is, thus, ideal for addressing neuroscience questions that require large-scale manipulations of deep and/or superficial cortical layers. Even with its current limitation of lacking recording capability, we have demonstrated that the UOA used as a stimulation-only device in conjunction with LEA recordings can be used to study inter-laminar interactions. By varying light intensity, the UOA allowed us to localize photostimulation to single or multiple cortical layers, and by varying the depth of insertion (or the customizable shank length) it is possible to target distinct layers. Differences in onset latency of light-evoked responses across the cortical depth could be used to distinguish distinct network activity patterns following different patterns of UOA stimulation. For example, at low light irradiance, direct neuronal activation was initially localized to the layers nearest optrode tip termination before spreading trans-synaptically to other layers. Increasing light irradiance directly contributed to firing rate increases in deep layers below the UOA tips while firing rates in L4C increased less compared to firing rates evoked at lower intensities (suggesting local activation of higher threshold inhibitory networks).

We also demonstrated that by varying the distance between the stimulation site/s on the UOA and the recording electrode, the UOA can be used to study local versus long-distance intra-areal interactions, while when coupled to recordings in a different cortical area (using LEAs or UEAs) the UOA allows investigations of the functions of inter-areal feedforward and feedback circuits. Moreover, used in conjunction with c-fos IHC, we were able to identify multisynaptic interactions within and beyond the photostimulated area. Photostimulation via the UOA increased c-fos expression over distances much > 8mm (well beyond the stimulated cortical area), but spiking activity could not be evoked beyond ~3 mm from the stimulated site, indicating that c-fos expression can be used to reveal subthreshold activity induced by network

interactions. This is consistent with previous demonstrations of c-fos expression several synapses away from the electrically stimulated site, indicating c-fos can be used for functional mapping of neuronal circuits[40].

Importantly, despite its limited shank length (up to 2.5 mm), the UOA can still be employed to study cortico-subcortical interactions, e.g. through modulation of axon terminals of deep nuclei within cortex, and recordings of postsynaptic cortical neurons in the same cortical area and/or layer.

Future developments of this device will involve: (i) addition of red μLEDs for dual color optogenetic neuromodulation through each needle shank and interstitial site, (ii) addition of electrical recording capabilities and (iii) optimization for chronic use in NHPs and humans.

In conclusion, the UOA will enable studies addressing long-standing fundamental questions in neuroscience, e.g., regarding the role of cortico-cortical feedback and cortical layers in the model system closest to humans. As many human neurological and psychiatric disorders have been linked to abnormalities in cortical circuit[4,5], this technology can improve our understanding of the circuit-level basis of human brain disorders, and will pave the way for a new generation of precise neurological and psychiatric therapeutic interventions via cell type-specific optical neural control prosthetics.

ONLINE METHODS

Device Fabrication, Characterization, and Benchmarking

Fabrication and testing of the first generation UOA devices was previously reported[29,45]. The second-generation devices used in this study included an optical interposer layer that limits emission from the μLED array to the shank sites for illumination of deep cortical tissue.

Fabrication. A 2 mm-thick, 100mm diameter Schott Borofloat 33 glass wafer used to construct the optrode needles was anodically bonded to a freshly cleaned 0.1mm thick, 100 mm diameter intrinsic Si wafer serving as an optical interposer. The Si and Borofloat wafers were coarsely aligned and bonding performed using an EVG 520 anodic bonder. The optical vias were patterned in the Si interposer by deep reactive ion etching (DRIE) using a Bosch process. A 10-μm-thick AZ9260 soft mask was photolithographically patterned to define the array of 80×80

μm^2 optical vias for shank and interstitial illumination for the DRIE process. The bonded wafer was then sub-diced into *modules* of 9 to 16 UOAs using a DISCO 3220 dicing saw.

UOA modules were mounted to a carrier wafer using WaferGrip™ (Dynatex International, Santa Rosa, CA). The glass shanks were cut with the DISCO 3220 using the previously reported process[29,45]. Briefly, beveled blades were first used to generate pyramidal tips on the surface, followed by standard profile blades to form the shanks. The shanks on a module were then etched to a nominal 110 µm thickness using a mixture of hydrofluoric (49%) and hydrochloric (37%) acid in a 9:1 ratio. The die was then demounted and cleaned, and the shanks were smoothened to decrease light scattering using a 725 °C heat treatment for 2 hours in a vacuum furnace. UOA modules were then singulated into individual 4×4 mm^2 UOAs using the DISCO 3220.

Arrays of µLEDs on thinned (150µm) sapphire substrates, from the Institute of Photonics at University of Strathclyde, were integrated with the UOA using closed-loop optical alignment to the optical vias on individual UOAs at Fraunhofer IZM (Berlin, Germany)[29], and bonded using index-matched epoxy. At the University of Utah, passive matrix µLED pads were wire bonded to an ICS-96 connector (Blackrock Microsystems, Salt Lake City, UT) using insulated gold alloy wire. The wire bundle and back-side of the UOA were then potted in NuSil MED-4211 silicone, respectively, followed by overcoating with a 6µm-layer of Parylene C.

Bench Testing. To characterize the electrical and optical performance of the finalized devices, the latter were attached to a custom switch board for matrix addressing the individual optrode shanks. The switch board consisted of a matrix arrangement of parallel connected mechanical switches and electrical relays, 10 sets for the anodes and 10 sets for the cathodes. This enabled both manual and automated activation of individual optrode shanks or optrode patterns. For the automated activation and testing, the relays were connected to Arduino boards which received commands from the lab computer. To prevent voltage spikes originating from the switching of the channels from damaging the µLEDs, the anode paths also contained a small filter circuit consisting of capacitors and Zehner diodes (break-down voltage: 8.2V). For the automated testing, the UOAs were inserted into the opening of an integrating sphere that was, in turn, connected to a photodetector and power meter (Newport 2832-C Dual-Channel Power Meter). The calibration factor of the integrating sphere was determined using a fiber coupled LED prior to the experiment. Then the UOAs were connected to the switch board, and the latter

was connected to a source measure unit (Keithley 236 Source Measure Unit) for the measurement. The automated characterization was conducted as follows: the switch board's Arduino boards received the command to switch to an individual optrode shank using the relays. Then the source measure unit applied a voltage pulse measurement pattern (pulse length 100 msec, pause between pulses 1900 msec to prevent heat buildup) sweeping the voltage from 0 to 7.2V (or until the compliance current of 100mA was reached) with each pulse increasing by 100mV. For each pulse, the resulting current and the output optical power were recorded; the optical power was then corrected using the integrating sphere calibration factor. This was repeated for each individual optrode shank of the device for a full characterization.

To ensure the stability of the device for an acute *in vivo* experiment, additional voltage transient measurements were made before and after a 48-hour soak test in phosphate-buffered saline (PBS) at 37 °C. Further, an electrode was immersed in solution to verify encapsulation integrity, as evidenced by lack of shorting to solution.

For the *in vivo* experiments, the switch board was upgraded two-fold: first, transistors were added to the cathode channels to allow for turning the device on and off based on an external TTL trigger. However, we found that turning on the optrodes using the trigger signal directly induced too strong a capacitively-coupled voltage signal in the recording. Therefore, as a second upgrade, an additional Arduino board with digital-analog-converter was added that received the external trigger and introduced rise and fall times to the square wave. This reduced the capacitively-coupled interference to a level below measurable when both the LEA and the UOA were in close proximity in 1xPBS solution prior to the *in vivo* experiment. During the experiment, the voltage for the UOA was supplied by a lab power supply via the switch board, and the switches were operated manually to define the required patterns.

Modeling. To understand light spread in tissue, the optical output of the device was modeled using ray-tracing software (Optics Studio 12, in non-sequential mode). This model has been described previously[29]. Brain tissue was modeled using a Henyey-Greenstein scattering model, with a scattering coefficient of 10 mm^{-1}, absorption coefficient of 0.07 mm^{-1}, and anisotropy of 0.88[46]. Each needle was modelled individually using its measured optical output at the given voltage level. To generate the cross-section images from a simultaneously illuminated column (**Fig. 1G**), the light output from the 10 needles in that column were summed.

638 **Animals**

639 A total of 3 adult female Cynomolgus monkeys (*Macaca fascicularis*) were used in this study.
640 The left hemisphere of one animal (case MK421-LH) was used for the *in vivo*
641 electrophysiological testing of the active UOA (integrated with the µLED array). The right
642 hemisphere from the same animal (MK42-RH), and 3 hemispheres from 2 additional animals
643 (MK414RH and LH, and MK422-RH) were used for c-fos testing of the passive UOA (without
644 an integrated µLED array). All procedures conformed to the National Institutes of Health Guide
645 for the Care and Use of Laboratory Animals and were approved by the University of Utah
646 Institutional Animal Care and Use Committee.

647

648 **Survival Surgical Procedures and Viral Injections**

649 Animals were pre-anesthetized with ketamine (10 mg/kg, i.m.), intubated, placed in a
650 stereotaxic apparatus, and artificially ventilated. Anesthesia was maintained with isofluorane (1–
651 2.5% in 100% oxygen). Heart rate, end tidal CO_2, oxygen saturation, electrocardiogram, and
652 body temperature were monitored continuously. I.V. fluids were delivered at a rate of 5/cc/kg/hr.
653 The scalp was incised and a craniotomy and durotomy were performed over area V1 (n=2
654 animals, MK421-LH and MK414-RH), or rostral to the precentral gyrus, roughly above the
655 supplementary motor area (SMA; n=1, MK422-RH). We injected a 1:1 viral mixture of
656 AAV9.CamKII.4.Cre.SV40 and AAV9.CAG.Flex.ChR2.tdTomato (Addgene Catalog #s:
657 105558, and 18917, respectively). We have previously found that this method nearly eliminates
658 retrograde expression of transgenes[10]. The viral mixture was slowly (~15nl/min) pressure-
659 injected (250-350nl repeated at 2 or 3 cortical depths between 0.5 and 1.5 mm from the cortical
660 surface) using a picospritzer (World Precision Instruments, FL, USA) and glass micropipettes
661 (35-45µm tip diameter). After each injection, the pipette was left in place for 5-10 min before
662 retracting, to avoid backflow of solution. A total of 5-6 such injections, each 500-750nl in total
663 volume, and spaced 1.5-2mm apart, were made in two animals (MK421-LH, MK414-RH) while
664 the third animal (MK422-RH) received 2 x 1,050nl injections. These injections resulted in a
665 region of high viral expression roughly 4-6 mm in diameter (as an example see **Extended Data**
666 **Fig. 3A Right**). Following viral injections, a sterile silicone artificial dura was placed on the
667 cortex, the native dura was sutured and glued onto the artificial dura, covered with Gelfoam to
668 fill the craniotomy, and the latter was sealed with sterile parafilm and dental acrylic. Anesthesia

669 was discontinued and the animal returned to its home cage. After a survival period of 5-10
670 weeks, to allow for robust ChR2 expression, the animals were prepared for a terminal UOA
671 photostimulation procedure.
672

673 **Terminal Surgical Procedures and UOA Insertion**
674 Monkeys were pre-anesthetized and prepared for experiments as described above. Anesthesia
675 and paralysis were maintained by continuous infusion of sufentanil citrate (5–10 μg/kg/h) and
676 vecuronium bromide (0.3 mg/kg/h), respectively. Vital signs were continuously monitored for
677 the duration of the experiment, as described above. Following suture removal and scalp incision,
678 the craniotomy and durotomy were enlarged to allow space for device implantation, and ChR2
679 expression was verified *in* vivo using a custom fluorescent surgical microscope (Carl Zeiss,
680 GmbH; **Fig. 2B**). UOAs were positioned over cortical regions of high tdT/ChR2 expression (e.g.
681 **Figs. 2B,6B,F**), and then inserted using a high speed pneumatic hammer typically used for
682 insertion of Utah Electrode Arrays[34] (Blackrock MicroSystems, Salt Lake City, UT). Parameters
683 used for insertion were 20 psi for 30 msec, using a 1 mm-long inserter, in order to achieve partial
684 insertion of the UOA, so as to minimize insertion trauma on the cortex. In two animals used for
685 c-fos experiments after partial insertion with the pneumatic inserter, the UOA was gently pushed
686 down to achieve deeper insertion.
687

688 **Photostimulation**
689 We implanted two types of UOA devices: (i) a 10x10 UOA with fully integrated μLED arrays
690 (also referred to as "active" device; n=1 device in 1 animal, MK421-LH; see **Fig. 2A-C**), and
691 (ii): 10x10 UOAs with an optical interposer integrated into the sapphire backplane, but with no
692 μLED array for light delivery (referred to as "passive" devices; n=3 devices in 3 hemispheres
693 from 2 animals, MK414-RH, MK414-LH, MK422-RH). The active device was used for
694 electrophysiological testing experiments, while the passive devices were used for the c-fos
695 experiments.
696 ***Active Device (Electrophysiology).*** Photostimulation with the active UOA occurred via
697 the integrated μLED array. Photostimulation parameters were 5Hz, 100 msec-pulse duration for
698 1 sec, followed by 1-10sec inter-trial interval (longer intervals were used at the higher

699 photostimulation intensities). We varied the spatial pattern (single μLED along column 1, whole
700 single columns, and all μLEDs across the entire UOA) and intensity (from 2.8 to 7.8V input
701 intensity) of photostimulation as described in the Results section.

702 ***Passive Devices (c-Fos).*** Selective photostimulation via passive devices was obtained by
703 illuminating a subset of UOA needles with an appropriately positioned fiber-coupled 473nm
704 laser (400 μm multimode optic fiber, ThorLabs Newton, NJ; laser: Laserwave, Beijing, China)
705 held in place with a stereotaxic tower. We used a collimating lens (ThorLabs, Newton, NJ) to
706 restrict spot size to ~1.5mm in diameter. To shield stray light, we covered any exposed tissue
707 around the illuminated area, as well as the non-illuminated portions of the UOA, with an opaque
708 (black) artificial dura. For each UOA we stimulated 2 or 3 separate sites. At each site we used
709 phasic photostimulation (50Hz for 2.5 min, 2.5 min pause, and 20Hz for an additional 2.5 min;
710 pulse duration was 10 msec) at 3.8mW power output (corresponding to an estimated irradiance
711 of 15-19mW/mm^2).
712

713 **Electrophysiological Recordings**

714 Extracellular recordings were made in V1 with 24-channel linear electrode arrays (LEAs; V-
715 Probe, Plexon, Dallas, TX; 100μm contact spacing, 300μm from tip to first contact, 20μm
716 contact diameter). The LEAs were inserted into the cortex next to the UOA to a depth of 2.4-
717 2.6mm, slightly angled laterally (towards the UOA) and posteriorly. We made a total of 3
718 penetrations (P1-P3; **Extended Data Fig. 3A**), of which only P2 and P3 provided useful data.
719 After UOA and LEA were inserted into the cortex, we applied a layer of Dura-Gel
720 (CambridgeNeuroTech, Cambridge, UK) over the cortex and UOA, to prevent the cortex from
721 drying and stabilize the recordings. A 128-channel recording system (Cerebus, Blackrock
722 Microsystems, Salt Lake City, UT) was used for standard signal amplification and filtering.
723 Multi-unit spiking activity was defined as any signal deflection that exceeded a voltage threshold
724 (set at 4 x the SD of the signal on each channel). Threshold crossings were timestamped with
725 sub-millisecond accuracy. We did not record responses to visual stimuli but only to UOA
726 photostimulation performed as described above; thus, the monkey's eyes were closed during the
727 duration of the experiment.
728

729 **Analysis of Electrophysiological Data**

We analyzed MU spiking responses from a total of 45 contacts deemed to lie within the parafoveal representation of V1 in two penetrations (out of 3 total, see above) for which neural activity was modulated by photostimulation via the active UOA. For the results presented in **Figures 3-5**, quantitative analysis was limited to contacts on which MU activity was stimulus modulated (one-way ANOVA comparing spike rates during full one-second photostimulation trials with spike rates during control periods of equivalent duration, $p < 0.01$).

To quantify the change in MU firing rates, relative to background, during photostimulation we calculated firing rates for all pulse epochs within all trials and then compared them to the average background rate. To estimate the preference at each recording site for stimulation across the full range of tested UOA locations (**Fig. 3**), we regressed average evoked-responses on UOA stimulation site and intensity. Preliminary analyses had revealed a non-monotonic relationship between stimulation intensity and response on many contacts (cf. **Fig. 2G**), thus we included a quadratic term in the regression model.

CSD analysis. For the CSD analysis shown in **Fig. 2D-F,** current source density (CSD) was calculated from the band-pass filtered (1-100Hz) and pulse-aligned and averaged LFP, using the kernel CSD toolbox (kCSD_Matlab)[47]. CSD was calculated as the second spatial derivative of the LFP signal, reflecting the net local transmembrane currents generating the LFP. The depth profile of the CSD was estimated by interpolating every 10μm. To facilitate comparisons across conditions, CSDs from different conditions were normalized to the standard deviation (SD) of the baseline (50 msec prior to pulse onset) after subtraction of the baseline mean.

Onset Latency. To quantify the onset latency of MU responses, we either: (i) calculated the average peri-stimulus time histogram (PSTH) from all pulse-aligned responses (e.g. **Fig. 4**) or (ii) estimated a PSTH separately for the response to each pulse (**e.g. Extended Data Fig. 4**). Peristimulus time histograms (PSTHs) were estimated via an adaptive algorithm in which the MU raster was first convolved with a Gaussian kernel of fixed width (3 msec bandwidth), kernel width was then adapted so that the number of spikes falling under the kernel was the same on average across the response (http://chronux.org[48]). We then subtracted the mean baseline response from the stimulus-evoked response. For each response measure, i.e. either the average or pulse-by-pulse PSTHs, we took the time at which the response reached 25% of the peak as the onset latency (results were qualitatively similar using 15% and 35% criteria; data not shown). We report the former measure as the mean onset latency in **Figures 4-5**. We used the latter

761 measure to test for differences in onset latency across contacts within and across UOA
762 stimulation parameters (**Figs. 4-5 and Extended Data Fig. 4**).

763 ***Statistical Analysis***. Stimulus-evoked firing rates were calculated from pulse-aligned or
764 trial-aligned responses and baseline corrected (mean baseline activity subtracted). We
765 determined responsiveness to stimulation via a one-way ANOVA comparing firing rates during
766 the full 1-second trial period with inter-leaved control periods of equivalent duration; MU
767 activity at an LEA recording site was deemed responsive if there was a significant difference
768 between stimulation and control trials at the p=0.01 level. To estimate the selectivity of MU
769 activity for stimulation at different UOA sites we performed a multiple linear regression, with
770 UOA column, row, and intensity as independent variables and pulse-aligned, baseline corrected,
771 firing rates as the dependent measure. To test for differences in the goodness-of-fit of models
772 with- and without a quadratic term, we used a two-sample Kolmogorov-Smirnov test. We
773 assessed the effects of varying UOA stimulation site and intensity on response amplitude or
774 onset latency using ANOVA models followed by the Tukey-Kramer test for post-hoc
775 comparisons.
776
777 **c-Fos Experiments**
778 We used 4 hemispheres from 3 animals for these experiments (MK414-RH and LH, MK422-RH,
779 and MK421-RH). Two of these animals (MK422 and MK414) were prepared for a terminal
780 experiment (as described above) 5 or 10 weeks, respectively, after the viral injections, and a
781 passive UOA was inserted in regions of high tdT/ChR2 expression in the injected hemisphere. In
782 one of these animals (MK422-RH), UOA insertion was preceded by glutamate block (see
783 below). After UOA insertion, photostimulation was performed via an optical fiber-coupled laser
784 through the UOA, as described above. Two additional hemispheres in 2 animals (MK414-LH
785 and MK421-RH) were used as controls. Specifically, case MK414-LH received insertion of a
786 passive UOA in non-opsin expressing SMA cortex, and was euthanized 4 hrs following UOA
787 insertion without receiving any photostimulation. As a separate control, in case MK421-RH we
788 performed surface photostimulation of SMA cortex not expressing opsins, using a fiber-coupled
789 laser and a collimating lens and the same photostimulation protocol described above for other c-
790 fos experiments; no UOA was inserted in this case. In all animals, UOA insertion and/or
791 photostimulation were performed after a 10-14-hour period of eye closure and at least 5 hours

after completion of surgery, and the animals were euthanized 75 minutes after completion of the photostimulation protocol.

Pharmacological Blockade of Local Glutamate Signaling. To compare changes in c-fos expression due to direct local optogenetic activation with indirect local and long-range changes due to synaptic increases in excitatory glutamatergic neurotransmission downstream of the directly-activated neurons, in one case (MK422-RH) we applied the selective glutamate AMPA receptor antagonist 2,3-dihydroxy-6-nitro-7-sulfamoyl-benzoquinoxaline-2,3-dione (NBQX, 5mM) (Tocris BioSciences, Minneapolis, MN). NBQX was applied topically prior to UOA insertion, by soaking a piece of Gelfoam placed over ChR2-expressing SMA cortex with 1ml of the drug solution. The drug was allowed to passively diffuse through the cortical layers for 90 minutes, during which 100-200µl of the solution were applied every 15 minutes to ensure saturation of the Gelfoam, after which the Gelfoam was removed and the passive UOA inserted over the region of glutamate block. Photostimulation was performed as described above for the passive device.

Histology

On completion of the experiments, the animals were euthanized by an overdose of Beuthanasia (0.22 ml/kg, i.v.) and perfused transcardially with saline for 2–3 min, followed by 4% paraformaldehyde (PFA) in 0.1M phosphate buffer for 20 min to fix the brain. The brains were post-fixed overnight in the same fixative, sunk in cryoprotectant 30% sucrose solution, and sectioned at 40µm on a freezing microtome. The hemisphere used for electrophysiological testing of the active UOA (MK421-LH) was sectioned tangentially. One in 3 sections were wet-mounted and imaged for fluorescent tdT-label at 10x magnification. The same sections were then reacted for cytochrome oxidase (CO) to reveal cortical layers and the location of UOA and LEA insertions visible as discolorations in CO staining (**Extended Data Fig. 3A Left**).

All other hemispheres used for c-fos experiments were sectioned sagittally. One full series of sections (1:3) were immunoreacted for c-fos by overnight incubation in primary antibody (1:500 rabbit anti-c-fos, Ab 19089, Abcam, MA) at room temperature, followed by 2 hrs incubation in near-infrared secondary antibody (1:200 donkey anti-rabbit IgG-AF647, Jackson ImmunoResearch, PA) at room temperature. Sections were then wet-mounted, counterstained with blue fluorescent Nissl (1:100 N21479, Thermo Fisher Scientific, MA), by

dripping the solution onto the slide-mounted sections every 5 min for 20 min, rinsed, and coverslipped and sealed with CoverGrip™ Coverslip Sealant (Biodium, CA).

Tissue Imaging

Imaging of tissue sections was performed on a Zeiss Axio Imager.Z2 fluorescent microscope (Zeiss GmbH, Germany) with a Zeiss X-cite 120 LED Boost light source, using a 10x objective and an Axiocam 506 mono camera (Zeiss GmbH, Germany). Image files were created and analyzed using Zen 2.6 Blue Software (Zeiss GmbH, Germany). The light intensity was set to 100%, and the exposure time for each channel was kept the same between images. The tangentially-sectioned hemisphere (MK421-LH) was imaged as described above. In all other cases, each sagittal section was imaged in 3 channels simultaneously, one channel for tdT/ChR2 (red- but note the color was artificially changed to green in **Fig. 6B, F**), one channel for Alexa-647-c-Fos (far-red), and the third channel for 435-455 Nissl (blue).

Analysis of c-Fos Expression

To quantify c-fos expression, c-fos+ cells were plotted and counted in sampled areas, using Neurolucida software 2006 (Microbrightfield Bioscience, VT). For each case, we selected for counts 5 sections spaced 1 mm apart encompassing the area of UOA insertion and/or photostimulation (for the light-only case). In each section, we plotted and counted cells within three 200µm-wide windows spanning all cortical layers, one positioned at or near the center of the UOA insertion region (or of photostimulation-only), and the other two located at distances of 4mm and 8mm, respectively, from the center of the UO insertion (**Fig. 6**). Thus, a total of 15 regions of interest (ROIs) were counted for each case. The laminar distribution of c-fos+ cells was analyzed by tracing the layers on the Nissl stain and counting the number of c-fos+ cells within each layer in Neurolucida. Statistical differences in c-fos+ cell counts among experimental and control cases, and across distances were estimated using a one-way ANOVA with post-hoc comparisons and Bonferroni correction).

REFERENCES

1. Deisseroth, K. Optogenetics: 10 years of microbial opsins in neuroscience. *Nat. Neurosci.* 18, 1213-1225 (2015).

2. El-Shamayleh, Y. & Horwitz, G.D. Primate optogenetics: Progress and prognosis. *Proc. Natl. Acad. Sci. U.S.A.* 116, 26195-26203 (2019).

3. Goldberg, M.E. The neurology clinic needs monkey research. *Proc. Natl. Acad. Sci. U.S.A.* 116, 26255-26258 (2019).

4. Lynall, M.-E. et al. Functional Connectivity and Brain Networks in Schizophrenia. *J. Neurosci.* 30, 9477-9487 (2010).

5. Belmonte, M.K. et al. Autism and Abnormal Development of Brain Connectivity. *J. Neurosci.* 24, 9228-9231 (2004).

6. Vitek, J.L. & Johnson, L.A. Understanding Parkinson's disease and deep brain stimulation: Role of monkey models. *Proc. Natl. Acad. Sci. U.S.A.* 116, 26259-26265 (2019).

7. Andersen, R.A., Aflalo, T. & Kellis, S. From thought to action: The brain-machine interface in posterior parietal cortex. *Proc. Natl. Acad. Sci. U.S.A.* 116, 26274-26279 (2019).

8. Picaud, S. et al. The primate model for understanding and restoring vision. *Proc. Natl. Acad. Sci. U.S.A.* 116, 26280-26287 (2019).

9. Jarvis, S. & Schultz, S.R. Prospects for Optogenetic Augmentation of Brain Function. *Front. Syst. Neurosci.* 9, 157 (2015).

10. Nurminen, L., Merlin, S., Bijanzadeh, M., Federer, F. & Angelucci, A. Top-down feedback controls spatial summation and response amplitude in primate visual cortex. *Nature Commun.* 9, 2281 (2018).

11. Siu, C., Balsor, J., Merlin, S., Federer, F. & Angelucci, A. A direct interareal feedback-to-feedforward circuit in primate visual cortex. *Nat Commun* 12, 4911 (2021).

12. Mehta, P. et al. Functional Access to Neuron Subclasses in Rodent and Primate Forebrain. *Cell Rep.* 26, 2818-2832 e2818 (2019).

13. Vormstein-Schneider, D.C. et al. Viral manipulation of functionally distinct interneurons in mice, non-human primates and humans. *Nat. Neurosci.* 23, 1629-1636 (2020).

14. Dimidschstein, J. et al. A viral strategy for targeting and manipulating interneurons across vertebrate species. *Nat. Neurosci.* 19, 1743-1749 (2016).

15. Tremblay, S. et al. An Open Resource for Non-human Primate Optogenetics. *Neuron*

884 108, 1075-1090 e1076 (2020).

885 16. Angelucci, A. et al. Circuits and mechanisms for surround modulation in visual cortex.
886 *Ann. Rev. Neurosci.* 40, 425-451 (2017).

887 17. Moore, T. & Zirnsak, M. Neural Mechanisms of Selective Visual Attention. *Annu. Rev.*
888 *Psychol.* 68, 47-72 (2017).

889 18. Barbas, H., J., W., Joyce, M.K.P. & Garcia-Cabezas, M.A. *J. Neurophysiol.* 120, 2659-
890 2678 (2018).

891 19. Rockland, K.S. & Pandya, D.N. Laminar origins and terminations of cortical connections
892 of the occipital lobe in the Rhesus monkey. *Brain Res.* 179, 3-20 (1979).

893 20. Usrey, W.M. & Sherman, S.M. Corticofugal circuits: Communication lines from the
894 cortex to the rest of the brain. *J. Comp. Neurol.* 527, 640-650 (2019).

895 21. Hubel, D.H. & Wiesel, T.N. Receptive fields and functional architecture of monkey
896 striate cortex. *J. Physiol. (Lond.)* 195, 215-243. (1968).

897 22. Huang, X., Elyada, Y.M., Bosking, W.H., Walker, T. & Fitzpatrick, D. Optogenetic
898 assessment of horizontal interactions in primary visual cortex. *J. Neurosci.* 34, 4976-4990
899 (2014).

900 23. Roy, A. et al. Optogenetic spatial and temporal control of cortical circuits on a columnar
901 scale. *J. Neurophysiol.* 115, 1043-1062 (2016).

902 24. Rajalingham, R. et al. Chronically implantable LED arrays for behavioral optogenetics in
903 primates. *Nat. Methods.* 18, 1112-1116 (2021).

904 25. Owen, S.F., Liu, M.H. & Kreitzer, A.C. Thermal constraints on in vivo optogenetic
905 manipulations. *Nat. Neurosci.* 22, 1061-1065 (2019).

906 26. Wang, J. et al. Integrated device for combined optical neuromodulation and electrical
907 recording for chronic in vivo applications. *J. Neural Eng.* 9, 016001 (2012).

908 27. Ozden, I. et al. A coaxial optrode as multifunction write-read probe for optogenetic
909 studies in non-human primates. *J. Neurosci. Methods* 219, 142-154 (2013).

910 28. Anikeeva, P. et al. Optetrode: a multichannel readout for optogenetic control in freely
911 moving mice. *Nat. Neurosci.* 15, 163-170 (2011).

912 29. McAlinden, N. et al. Multisite microLED optrode array for neural interfacing.
913 *Neurophotonics* 6, 035010 (2019).

914 30. Scharf, R. et al. Depth-specific optogenetic control in vivo with a scalable, high-density
915 muLED neural probe. *Sci. Rep.* 6, 28381 (2016).

916 31. Maynard, E.M., Nordhausen, C.T. & Normann, R.A. The Utah intracortical Electrode

Array: a recording structure for potential brain-computer interfaces. *Electroencephalogr. Clin. Neurophysiol.* 102, 228-239 (1997).

32. Lin, J.Y. Optogenetic excitation of neurons with channelrhodopsins: light instrumentation, expression systems, and channelrhodopsin variants. *Prog. Brain Res.* 196, 29-47 (2012).

33. Wiegert, J.S., Mahn, M., Prigge, M., Printz, Y. & Yizhar, O. Silencing Neurons: Tools, Applications, and Experimental Constraints. *Neuron* 95, 504-529 (2017).

34. Kim, K. et al. Artifact-free and high-temporal-resolution in vivo opto-electrophysiology with microLED optoelectrodes. *Nat. Commun.* 11, 2063 (2020).

35. Callaway, E.M. & Wiser, A.K. Contributions of individual layer 2-5 spiny neurons to local circuits in macaque primary visual cortex. *Vis. Neurosci.* 13, 907-922 (1996).

36. Pluta, S. et al. A direct translaminar inhibitory circuit tunes cortical output. *Nat. Neurosci.* 18, 1631-1640 (2015).

37. Qi, G. & Feldmeyer, D. Dendritic Target Region-Specific Formation of Synapses Between Excitatory Layer 4 Neurons and Layer 6 Pyramidal Cells. *Cereb. Cortex* 26, 1569-1579 (2016).

38. Acker, L., Pino, E.N., Boyden, E.S. & Desimone, R. FEF inactivation with improved optogenetic methods. *Proc. Natl. Acad. Sci. U.S.A.* 113, E7297-E7306 (2016).

39. Andrei, A.R., Pojoga, S., Janz, R. & Dragoi, V. Integration of cortical population signals for visual perception. *Nat. Commun.* 10, 3832 (2019).

40. Sagar, S.M., Sharp, F.R. & Curran, T. Expression of c-fos protein in brain: metabolic mapping at the cellular level. *Science* 240, 1328-1331 (1988).

41. Kovacs, K.J. c-Fos as a transcription factor: a stressful (re)view from a functional map. *Neurochem. Int.* 33, 287-297 (1998).

42. Tamura, K. et al. A glass-coated tungsten microelectrode enclosing optical fibers for optogenetic exploration in primate deep brain structures. *Journal of neuroscience methods* 211, 49-57 (2012).

43. Stark, E., Koos, T. & Buzsaki, G. Diode probes for spatiotemporal optical control of multiple neurons in freely moving animals. *J. Neurophysiol.* 108, 349-363 (2012).

44. Royer, S. et al. Multi-array silicon probes with integrated optical fibers: light-assisted perturbation and recording of local neural circuits in the behaving animal. *Eur. J. Neurosci.* 31, 2279-2291 (2010).

45. Scharf, R. et al. A compact integrated device for spatially-selective optogenetic neural stimulation based on the Utah Optrode Array. *Proc. SPIE* 10482, 104820M (2018).

46. Bernstein, J.G. et al. Prosthetic systems for therapeutic optical activation and silencing of genetically-targeted neurons. *Proc. SPIE Int. Soc. Opt. Eng.* 6854, 68540H (2008).

47. Potworowski, J., Jakuczun, W., Leski, S. & Wojcik, D. Kernel current source density method. *Neural. Comput.* 24, 541-575 (2012).

48. Mitra, P. & Bokil, H. Observed Brain Dynamics. (Oxford University Press, New York; 2008).

ACKNOWLEDGMENTS

We thank Kesi Sainsbury for histological assistance, Seminare Ta'afua for help with experiments, and Julian Haberland and Christine Kallmayer at the Fraunhofer-Institut für Zuverlässigkeit und Mikrointegration (IZM) for µLED-to-interposer bonding. This work was supported primarily by a BRAIN grant (U01 NS099702) from the National Institute of Health (NIH) to S.B., A.A. L.R and S.M. Additional support was provided by grant from the NIH (R01 EY026812, R01 EY019743, R01 EY031959) and the National Science Foundation (IOS 175543) to A.A, and an unrestricted grant from Research to Prevent Blindness, Inc. and a core grant from the NIH (EY014800) to the Department of Ophthalmology, University of Utah.

AUTHOR CONTRIBUTIONS

Conceptualization: A.C., A.I., C.F.R., N.M., L. R., K.M., S.B. A.A. Device Fabrication and Bench Testing: C.F.R., N.M., L.R.,K.M., M.D.D., S.B. Modeling: N.M., K.M. *In Vivo* Electrophysiology Testing: A.C., D.C., C.F.R., F.F., A.A. *In Vivo* cFos Testing: A.I., J.B., F.F., J.D.R., A.A., A.C. Analysis of Electrophysiology Data: A.C. Analysis of cFos Data: A.I., J.B., F.F., A.A. Writing-Original Draft: A.C., L.R., S.B., A.A. Writing–Review/Editing: all authors. Supervision & Funding Acquisition: A.A., S.B., K.M., L.R.

COMPETING INTERESTS STATEMENT

The authors declare no competing interests.

FIGURE LEGENDS

Figure 1. UOA Design and Optical Properties

(A) Schematics of UOA design superimposed to a Nissl-stained coronal section of macaque V1 showing the layers. The UOA consists of 3 main components: a µLED array (B), an optical interposer (C) and a glass needle array (D). **(B)** Two interleaved µLED arrays on a sapphire substrate are shown in this image; the first 10x10 array is needle-aligned for deep layer stimulation, the second 9x9 interstitial array lies in-between the first for surface stimulation. The interstitial array, although built into the UOA, was not used in this study. Scale bar: 1mm. **(C)** A region of the silicon optical interposer corresponding to approximately the size of the *white box* in (B); the optical "vias" are etched through the silicon and matched to the size of a µLED ($80x80\mu m^2$). Scale bar: 200µm. **(D)** High magnification image of the glass needle shanks bonded to the interposer. Scale bar: 200µm. **(E)** LEFT: The µLED on sapphire and needle array components are integrated into the final device, wire-bonded, and encapsulated. The image shown is a representative device. The integrated UOA used in this study consisted of 10x10 glass needle shanks, 1.6 mm long (to target deeper layers) and 100-110µm wide, with tip apex angles about 64º. An image of the actual device used in the *in vivo* testing studies, after completion of the experiment and explantation is shown in **Extended Data Fig. 2**. Scale bar: 1mm. RIGHT: Example spatial patterns of device operation. **(F)** Average output optical power (in mW) across each needle tip at different drive voltages (currents), when the entire UOA was turned on (*top left inset*). *Blue and gray bars*: needle shanks with estimated tip irradiances above and below, respectively, the $1mW/mm^2$ threshold for ChR2 activation. **(G)** LEFT: Ray trace model of light spread in cortical tissue when a single µLED (in column 1 and row 8, i.e. the closest to the linear electrode array —LEA– in penetration 2 –P2– used for the electrophysiological testing experiment, and indicated as a *black dot*) is activated at various input voltages (% of maximum intensity used), with power output calibrated to the bench tests. RIGHT: Model of light spread in tissue when column 1 (the nearest to the LEA in P2 and P3) is activated at various input voltages. *Green contour* encloses tissue volume within which the light irradiance is above $1mW/mm^2$, the threshold for ChR2 activation. Scale bars: 400µm.

Figure 2. Laminar Distribution of Responses Induced by UOA Photostimulation.

(A) The UOA inserted in macaque V1. **(B)** Same field of view as in (A) shown under fluorescent illumination to reveal expression of the red fluorescent protein tdTomato (*arrow*). **(C)** Preparation for recording electrophysiological responses to photostimulation. A 24 channel linear electrode array (LEA) was inserted next to the UOA (guide tube protecting array marked "LEA") slightly angled laterally (towards the UOA) and posteriorly. Here the UOA is partially covered with a piece of Gelfoam. **(D)** Current Source Density analysis (CSD; Left) and multiunit (MU) spiking activity (Right) signals recorded through the depth of V1 in P2 in response to phasic UOA photostimulation (pulse parameters: 100 msec pulse duration, 5Hz, 7.4mW/mm^2; pulse periods denoted as *blue bars* above MU plot). Here, all 100 needle-aligned µLED sites ("whole µLED array" condition) were activated simultaneously. CSD responses to each 100 msec pulse were zero-aligned, while MU activity is shown for the full 5Hz pulse train. *The dashed lines in the CSD panel* demarcate the borders of layer 4C (L4C)*; the gray shaded region in the MU activity panel* delimits the extent of L4C. **(E)** Same as in (D), but for photostimulation irradiance of 0.82 mW/mm^2. **(F)** Same as in (D-E), but following surface photostimulation of V1 via a laser-coupled optical fiber with pulse parameters of 10 msec, 5Hz, 2.2mW/mm^2. **(G-J)** Left: Relative cortical depth of each contact on P2 (*black dot in the insets*) is plotted versus the relative response (% firing rate increase over baseline) to UOA stimulation for different 450nm µLED illumination patterns (*insets*). Different colored traces are data for different photostimulation intensities (expressed as voltage or percent of max intensity used). *Gray area:* extent of L4C; *dashed lines*: approximate location of the L4A/4B (upper) and L5/L6 (lower) borders. Right: PSTHs with and without µLED activation are shown for the same contact on the LEA in L4C (marked by the *black circle)* across conditions. *Dashed line in the PSTH*: pulse periods.

Figure 3. Tangential Extent of Responses Induced by UOA Photostimulation.

A) Examples of model fits to single µLED and single column photostimulation for an example contact from P2, the one that showed the largest relative response increase across these stimulation conditions. This contact preferred stimulation in the proximal UOA columns 1-2, at sites closer to the top of the device (rows 9-7). The schematics on the left of the UOA and of the LEA-P2 indicates as *blue shading* the UOA sites represented in the heat map, and as a *red dot* the contact on the LEA whose response is mapped on the right. The *horizontal lines* and *gray*

shading on the LEA schematics mark the pial and white matter, and L4C boundaries, respectively. Color scale applies to panels (A-B, E-F). **(B)** Average normalized fitted responses across all responsive contacts in P2 (*red dots* in schematics of LEA to the left). **(C)** Change in response in the column direction for P2. Average relative response amplitude (% of peak response) is plotted as a function of stimulation intensity and distance along a straight line extending from the preferred UOA site in the column direction. Data averaged across all contacts. **(D)** Change in response in the row direction for P2. Average relative response amplitude (% of peak response) is plotted as a function of stimulation intensity and distance along a straight line extending from the preferred UOA site in the row direction. Data averaged across all contacts. **(E-H**) Same as in (A-D) but for P3. (E-F) P2 preferred stimulation in the proximal UOA columns 1-3, at sites closer to the middle of the UOA (rows 4-7).

Figure 4. Onset Latencies Reveal Local Networks Activated by Focal Optogenetic Modulation. (A) Left: Schematics of UOA stimulation through a single μLED site (C1-R8) and of LEA in P2. Right: Pulse-aligned raster plots for all 21 channels on the LEA through the depth of V1. *Black lines* separate data from different channels. *Gray shaded region:* channels in L4C. *Blue line above plot*: 100ms pulse period at the input voltage (irradiance) indicated. *Red and black arrows denote example contacts in L4C and 6, respectively*. A graded shift in MU onset latency is apparent. **(B)** Left: Pulse-aligned PSTHs for the two channels indicated by arrows in the raster plot in (A). Responses are plotted as baseline-subtracted firing rate versus time. Response onset latency at the L6 contact (35 msec) clearly lagged that on the L4C contact nearest the UOA needle tips (17 msec). Right: Histograms of pulse-by-pulse onset latencies for the two example contacts. **(C)** Heatmaps of MU response (firing rate) through the depth of V1 during the peri-pulse period, for the UOA stimulation condition indicated by the insets at the top left of each plot. Stimulation intensity (average irradiance) is reported above each plot. The firing rate color scale applies to all panels. *White dots* mark the onset latency (estimated from the mean PSTH- see Online Methods) for each contact that was significantly responsive to UOA stimulation.

Figure 5. Population Onset Latencies as a Function of UOA Stimulation Intensity and Spatial Pattern.

(A) Distance on the LEA of each contact from the contact with the fastest onset latency is plotted against onset latency; lines are linear fits. Each line is from simultaneous stimulation throughout the whole µLED array at each indicated intensity. **(B)** Left: Effect of varying photostimulation intensity for a fixed column (C1). Right: Effect of varying stimulated column (C1 to C4) for a fixed photostimulation intensity (5V). Either lowering intensity for a given column or increasing the distance between an activated column and the LEA had similar effects on the latency of network activation. **(C)** As in (B), but for a single µLED stimulation condition. On the left panel, photostimulation intensity was varied for a fixed µLED (C1-R8), while on the right panel, the stimulated µLED was varied along column 1 (from row 3 to 9) at a fixed intensity (5V for µLEDS in rows 8-10, but 7.8V for those in rows 4-7, as lower intensities did not evoke a response from many of these latter µLEDs). **(D)** The shortest onset latency across all intensities (here expressed as percent of max- see legend in **Fig. 2G** for corresponding input voltage) is plotted for the whole array condition (Left), and selected columns (Middle) or µLEDs (Right).

Figure 6. Local Optogenetic Activation Through the UOA Spreads Through Cortico-Cortical Networks.

(A-C) Case MK414-RH. The same sagittal section encompassing parts of V1 and V2 is shown under 3 different fluorescent illuminations, to reveal Nissl stain (A), tdT/Chr2 expression (B; the red tdT fluorescence was converted to green for purpose of illustration), and c-fos IHC (C). *White solid contour*: V1/V2 border; *dashed contours:* layer boundaries (layers are indicated); *white boxes:* ROIs (numbered 1-3 in panel C) where c-fos+ cells were counted. *White Arrows in (B)* point to the visible damage caused by each UOA needle, while the *gray arrow* points to the likely location of one of the UOA needles which did not cause visible damage in this section. *Asterisks in (B)* mark the core of the viral injections, and sites of highest tdT/ChR2 expression. *P:* posterior; *V:* ventral. C-fos expression in this case is observed throughout all layers (local) and across cortical areas (long-range). Scale bar in (A): 1mm (valid for A-C). **(D)** Higher magnification of c-fos IHC in and around each ROI. Scale bar: 0.2mm. **(E-H)** Case MK422-RH. Same as in (A-D) but for a different case in which an AMPA receptor antagonist was injected into the SMA prior to UOA insertion and photostimulation. The sagittal section is from the SMA. *D:* dorsal; *A:* anterior. Scale bars: 1mm (E and valid for E-G); 0.2 mmm (H). Blocking AMPA receptors demonstrates that initial optogenetic activation is limited to the stimulated

layers in the region of UOA insertion. **(I-J)** Case MK414-LH. C-fos IHC in a sagittal section of SMA cortex (I) and at higher magnification in and around each ROI used for cell counts (J), in a case which only received UOA insertion. Scale bars: 1mm (I), 0.2mm (J). **(K-L)** Case MK421-RH. Same as in (I-J), but for a control case in which SMA cortex only received surface photostimulation via an optical fiber-coupled laser. Here only one ROI is shown at higher magnification to reveal the few labeled cells in L1. Scale bars: 0.5mm (K), 0.2mm (L). Increases in cFos expression cannot be explained by device insertion or surface illumination. **(M)** Average number of c-fos+ cells across sections used for quantification, as a function of distance from the center of UOA insertion for the 4 different cases. Error bars: s.e.m **(N)** Distribution of c-fos+ cells across layers at each distance.

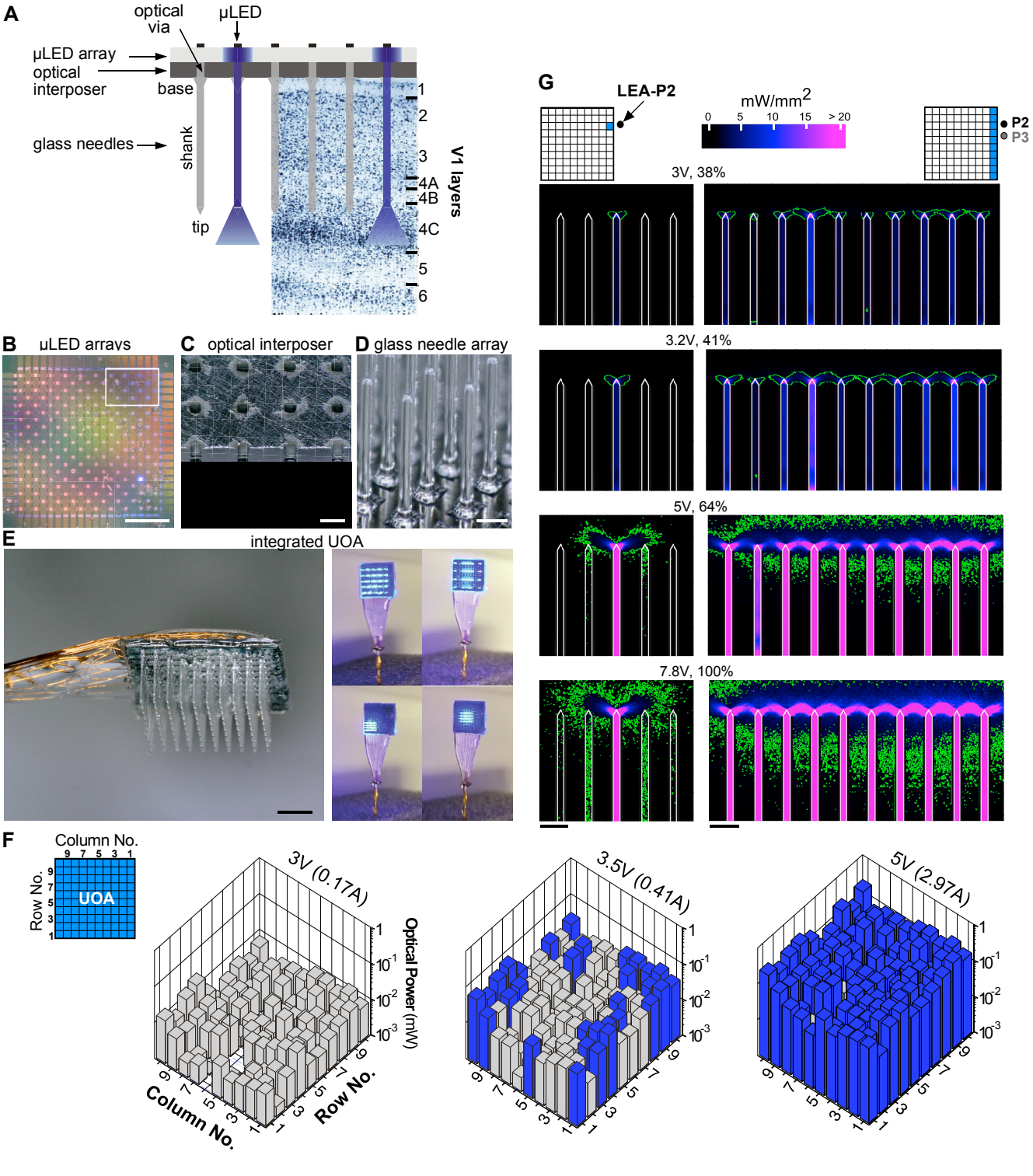


Figure 1

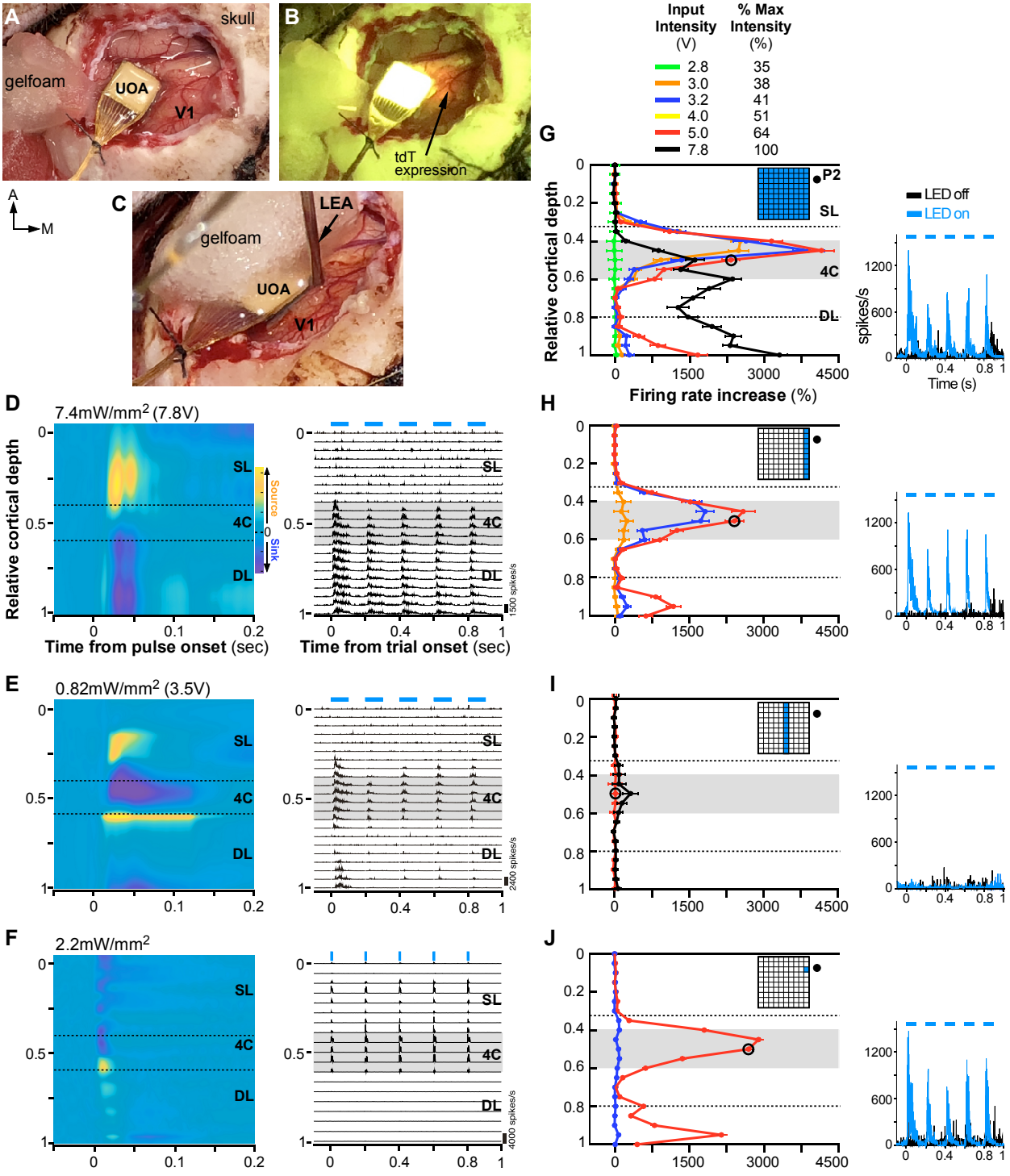


Figure 2

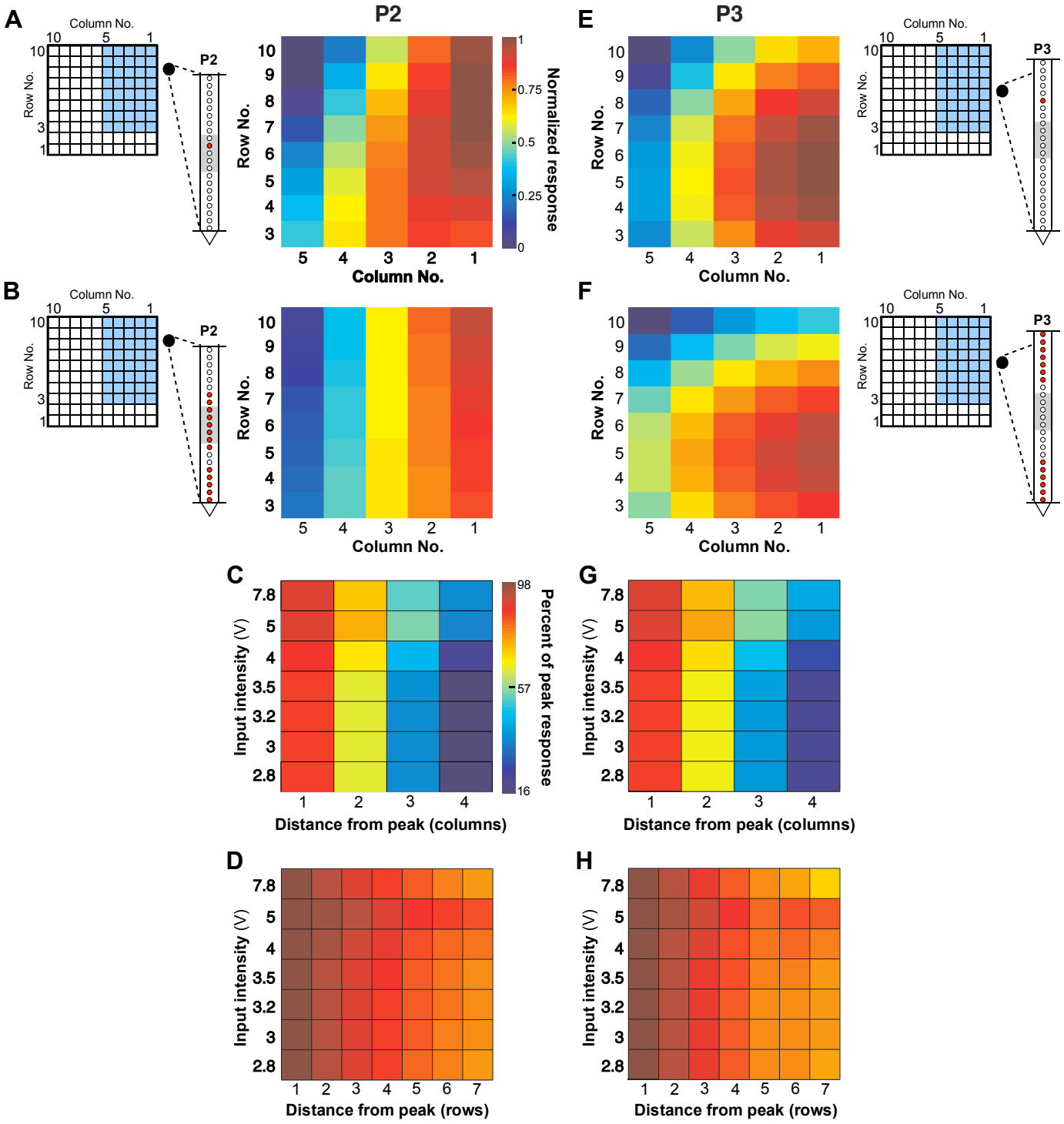


Figure 3

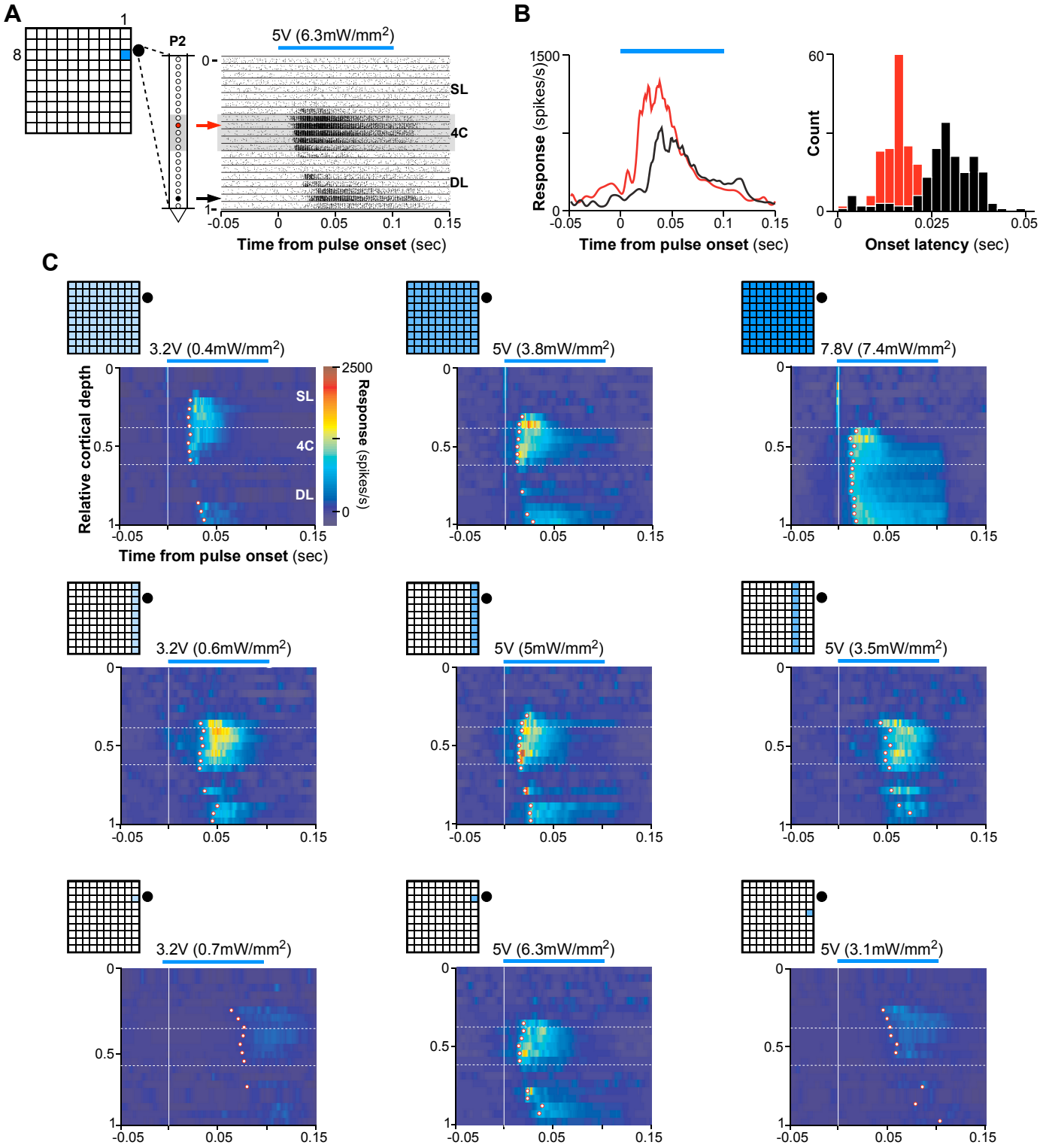


Figure 4

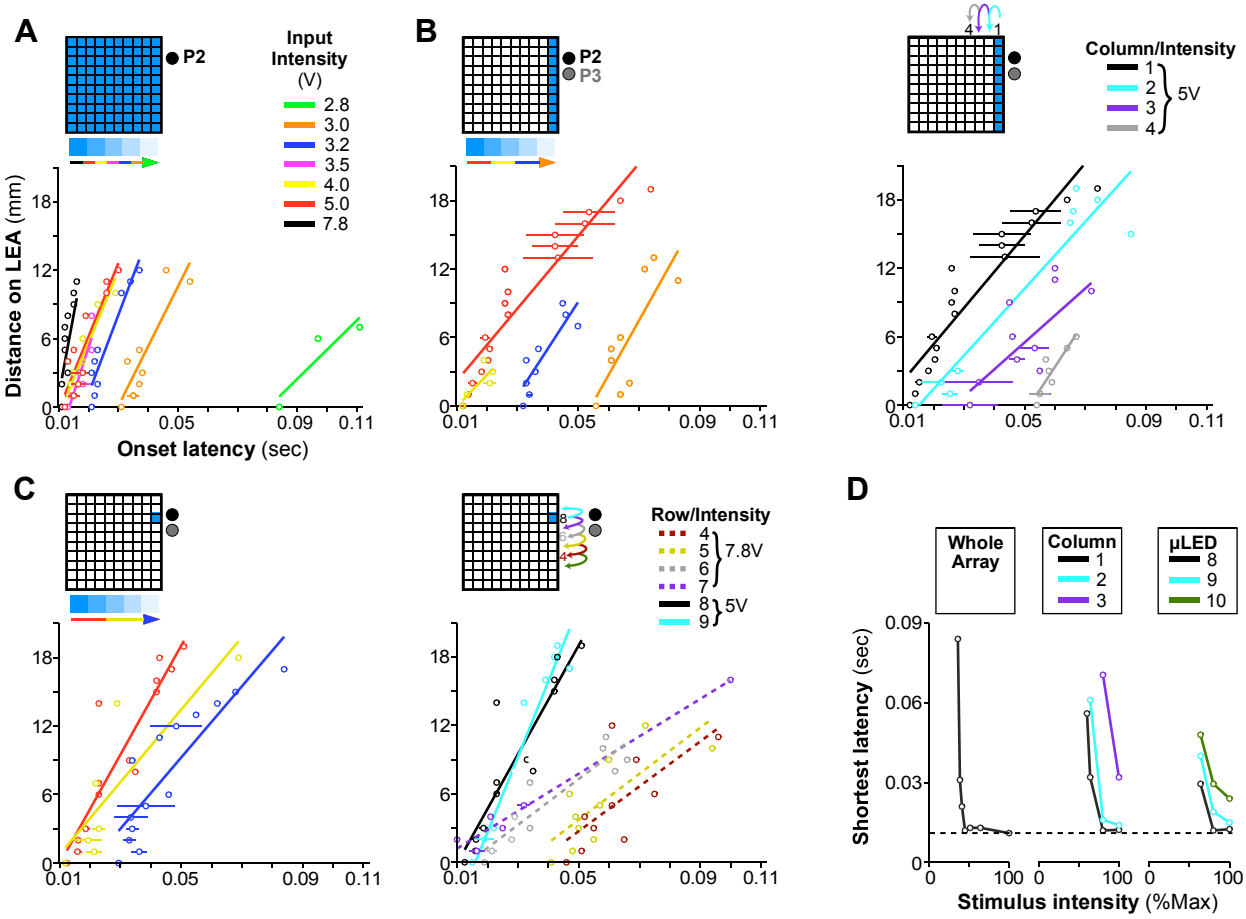


Figure 5

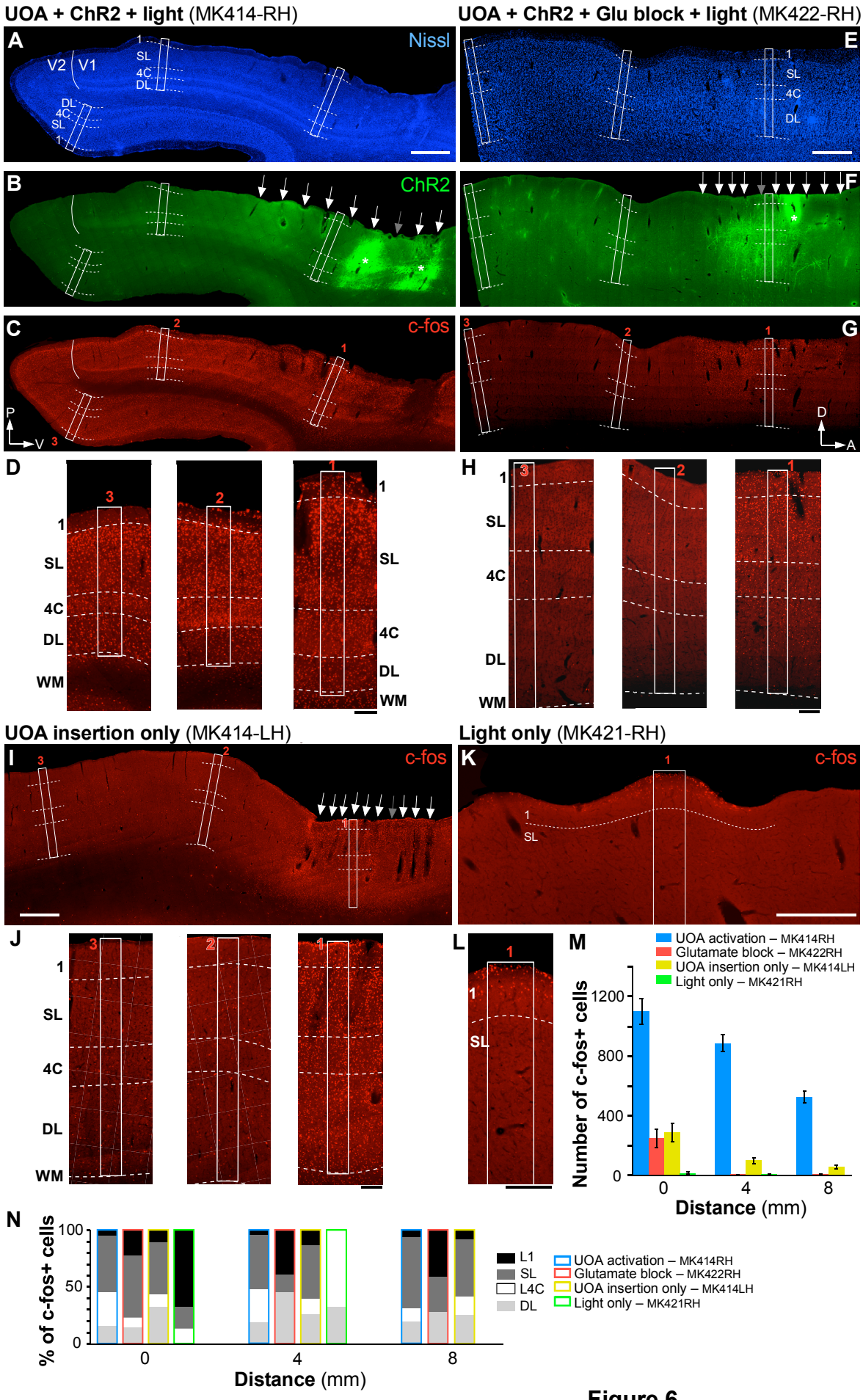


Figure 6

Measured Mean Output Photostimulation Intensities for Different Input Voltages

WHOLE ARRAY

Input Voltage (V)	**Output Optical Power** (mW)					
	Mean	***SD***	***Median***	***Min***	***Max***	***IQR***
2.8	0.0022	0.0016	0.0019	0	0.010	0.0014
3	0.0057	0.0040	0.0050	0.0005	0.024	0.0051
3.2	0.011	0.0072	0.010	0.0010	0.042	0.0091
3.5	0.022	0.013	0.020	0.0024	0.075	0.0170
4	0.044	0.026	0.041	0.0075	0.13	0.0313
5	0.10	0.056	0.088	0.018	0.27	0.0629
7.8	0.19	0.09	0.18	0.039	0.42	0.12

Input Voltage (V)	**Output Irradiance** (mW/mm^2)					
	Mean	***SD***	***Median***	***Min***	***Max***	***IQR***
2.8	0.08	0.06	0.07	0.01	0.38	0.05
3	0.21	0.14	0.20	0.02	0.91	0.19
3.2	0.41	0.27	0.38	0.04	1.56	0.33
3.5	0.82	0.49	0.75	0.09	2.79	0.62
4	1.67	0.95	1.53	0.28	4.98	1.16
5	3.79	2.08	3.33	0.67	9.88	2.48
7.8	7.4	3.19	6.9	1.45	15.6	4.46

COLUMN 1

Input Voltage (V)	**Output Optical Power** (mW)					
	Mean	***SD***	***Median***	***Min***	***Max***	***IQR***
2.8	0.0030	0.0016	0.0024	0.0010	0.0068	0.0013
3	0.0079	0.0046	0.0071	0.0019	0.0181	0.0043
3.2	0.0156	0.0082	0.0146	0.0040	0.0321	0.0091
3.5	0.0305	0.0146	0.0301	0.0089	0.0570	0.0170
4	0.0611	0.0276	0.0627	0.0181	0.1016	0.0354
5	0.1350	0.0597	0.1324	0.0400	0.2424	0.0873
7.8	0.2548	0.0981	0.2719	0.1016	0.4168	0.1411

Input Voltage (V)	**Output Irradiance** (mW/mm^2)					
	Mean	***SD***	***Median***	***Min***	***Max***	***IQR***
2.8	0.11	0.06	0.09	0.04	0.25	0.05
3	0.29	0.17	0.27	0.07	0.67	0.16

3.2	0.58	0.31	0.54	0.15	1.19	0.34
3.5	1.13	0.54	1.12	0.33	2.11	0.63
4	2.26	1.02	2.32	0.67	3.77	1.31
5	5.00	2.21	4.90	1.48	8.98	3.23
7.8	9.43	3.63	10.07	3.77	15.44	5.23

COLUMN 3

Input Voltage (V)	**Output Optical Power** (mW)					
	Mean	***SD***	***Median***	***Min***	***Max***	***IQR***
2.8	0.0020	0.0010	0.0022	0.0005	0.0032	0.0021
3	0.0055	0.0023	0.0058	0.0019	0.0089	0.0043
3.2	0.0106	0.0044	0.0110	0.0032	0.0159	0.0078
3.5	0.0211	0.0080	0.0216	0.0081	0.0321	0.0119
4	0.0421	0.0145	0.0432	0.0181	0.0657	0.0143
5	0.0933	0.0303	0.0932	0.0443	0.1460	0.0232
7.8	0.1836	0.0550	0.1820	0.0981	0.2871	0.0202

Input Voltage (V)	**Output Irradiance** (mW/mm^2)					
	Mean	***SD***	***Median***	***Min***	***Max***	***IQR***
2.8	0.07	0.04	0.08	0.02	0.12	0.08
3	0.20	0.09	0.22	0.07	0.33	0.16
3.2	0.39	0.16	0.41	0.12	0.59	0.29
3.5	0.78	0.30	0.80	0.30	1.19	0.44
4	156	0.54	1.60	0.67	2.43	0.53
5	3.46	1.12	3.46	1.64	5.40	0.86
7.8	6.80	2.03	6.74	3.63	10.63	0.75

COLUMN 5

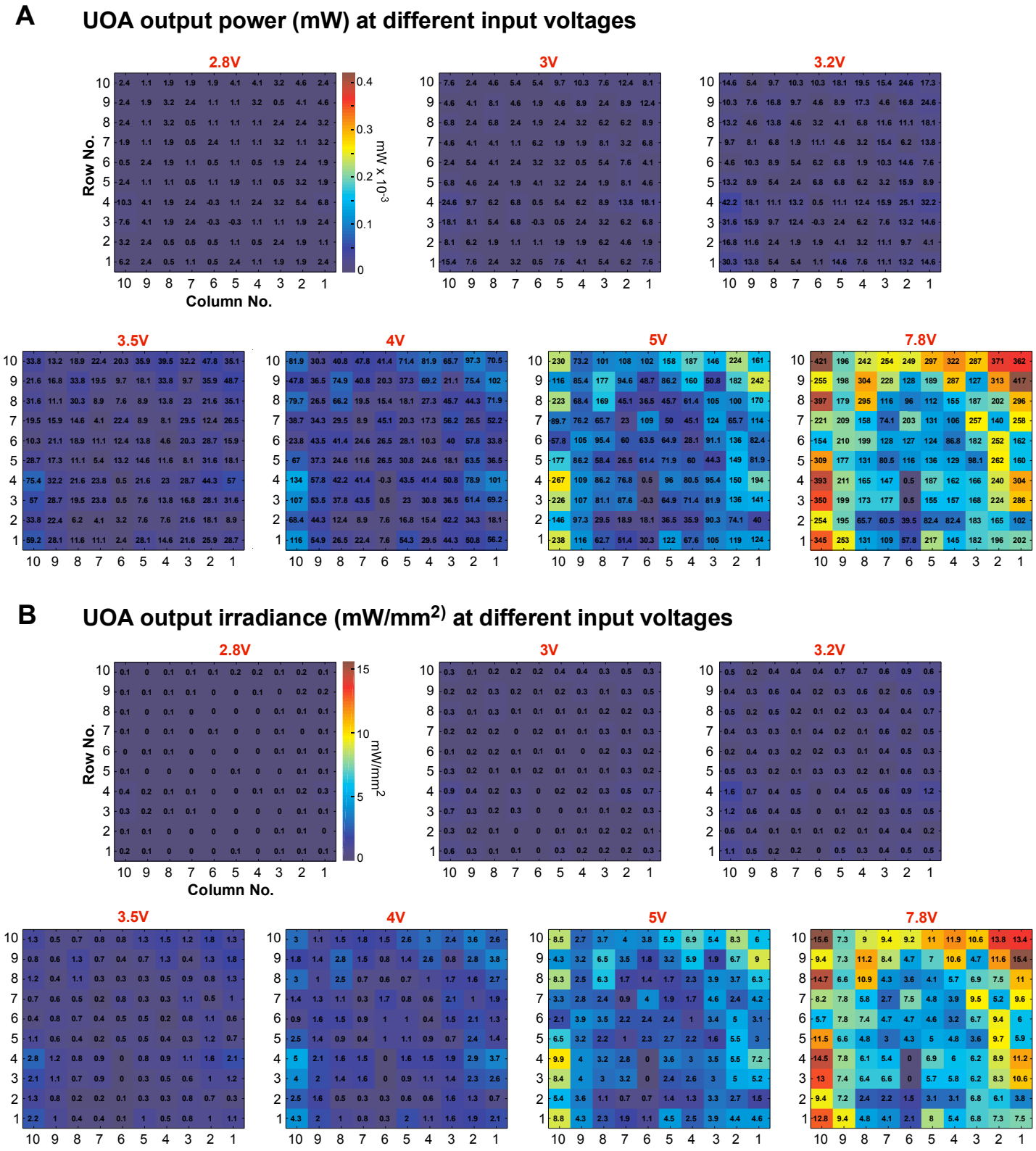
Input Voltage (V)	**Output Optical Power** (mW)					
	Mean	***SD***	***Median***	***Min***	***Max***	***IQR***
2.8	0.0014	0.0011	0.0010	0.0003	0.0041	0.0008
3	0.0041	0.0028	0.0032	0.0005	0.0097	0.0035
3.2	0.0081	0.0051	0.0068	0.0024	0.0181	0.0070
3.5	0.0165	0.0110	0.0142	0.0076	0.0360	0.0127
4	0.0344	0.0177	0.0249	0.0168	0.0713	0.0232
5	0.0797	0.0376	0.0683	0.0365	0.1583	0.0460
7.8	0.1623	0.0620	0.1456	0.0824	0.2971	0.0651

Input Voltage (V)	**Output Irradiance** (mW/mm^2)					
	Mean	***SD***	***Median***	***Min***	***Max***	***IQR***
2.8	0.05	0.04	0.04	0.01	0.15	0.03
3	0.15	0.11	0.12	0.02	0.36	0.13
3.2	0.30	0.18	0.25	0.09	0.67	0.26
3.5	0.61	0.36	0.53	0.28	1.33	0.47
4	1.27	0.66	1.09	0.62	2.64	0.86
5	2.95	1.39	2.53	1.35	5.87	1.7
7.8	6.03	2.30	5.39	3.05	11.00	2.41

SINGLE µLEDs IN COLUMN 1

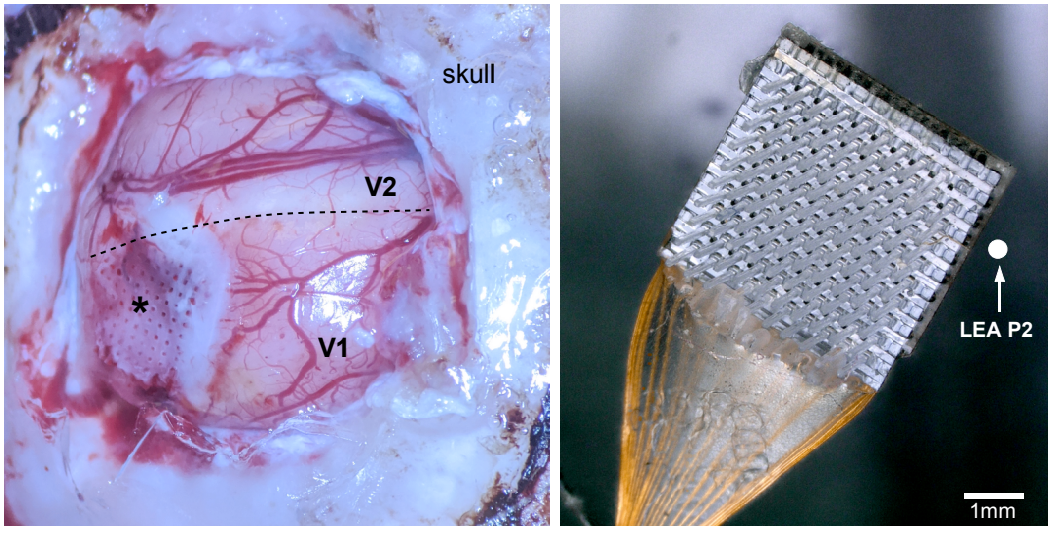
Input Voltage (V)	**Mean Output Optical Power** (mW)								
	Row 1	***Row 2***	***Row 3***	***Row 4***	***Row 5***	***Row 6***	***Row 7***	***Row 8***	***Row 9***
2.8	0.0024	0.0011	0.0024	0.0068	0.0019	0.0019	0.0032	0.0032	0.0046
3	0.0076	0.0019	0.0068	0.0181	0.0046	0.0041	0.0068	0.0089	0.0124
3.2	0.0146	0.0040	0.0146	0.0322	0.0089	0.0076	0.0138	0.0181	0.0246
3.5	0.0287	0.0089	0.0316	0.0570	0.0181	0.0159	0.0265	0.0351	0.0487
4	0.0562	0.0181	0.0691	0.1010	0.0365	0.0338	0.0521	0.0719	0.1016
5	0.1243	0.0400	0.1406	0.1938	0.0819	0.0824	0.1138	0.1697	0.2425
7.8	0.2016	0.1016	0.2857	0.3035	0.1597	0.1624	0.2581	0.2957	0.4168

Input Voltage (V)	**Mean Output Irradiance** (mW/mm^2)								
	Row 1	***Row 2***	***Row 3***	***Row 4***	***Row 5***	***Row 6***	***Row 7***	***Row 8***	***Row 9***
2.8	0.1	0	0.1	0.3	0.1	0.1	0.1	0.1	0.2
3	0.3	0.1	0.3	0.7	0.2	0.2	0.3	0.3	0.5
3.2	0.5	0.2	0.5	1.2	0.3	0.3	0.5	0.7	0.9
3.5	1.1	0.3	1.2	2.1	0.7	0.6	1	1.3	1.8
4	2.1	0.7	2.6	3.7	1.4	1.3	1.9	2.7	3.8
5	4.6	1.5	5.2	7.2	3	3.1	4.2	6.3	9
7.8	7.5	3.8	10.6	11.2	5.9	6	9.6	11	15.4



Extended Data Figure 1

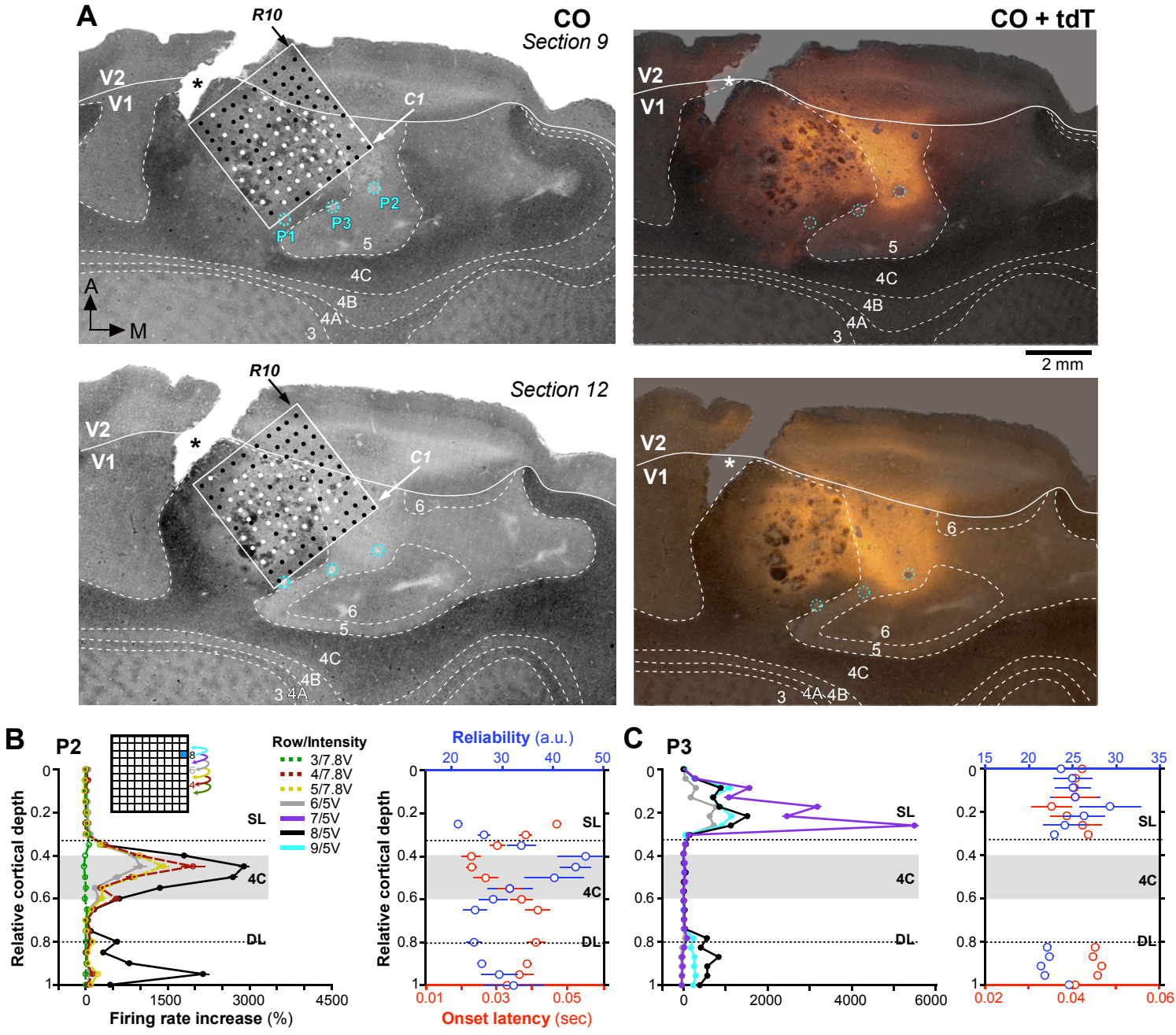
Output optical power and irradiance of the UOA
Each heat map represents the output optical power **(A)** or irradiance **(B)** measured at each needle tip across the entire UOA for different input voltages (indicated at the top of each map). Needles C6-R3 and C6-R4 did not emit light at their tips, thus, they were not included in the descriptive statistics in Extended Data Table 1.



Extended Data Figure 2

UOA after explantation

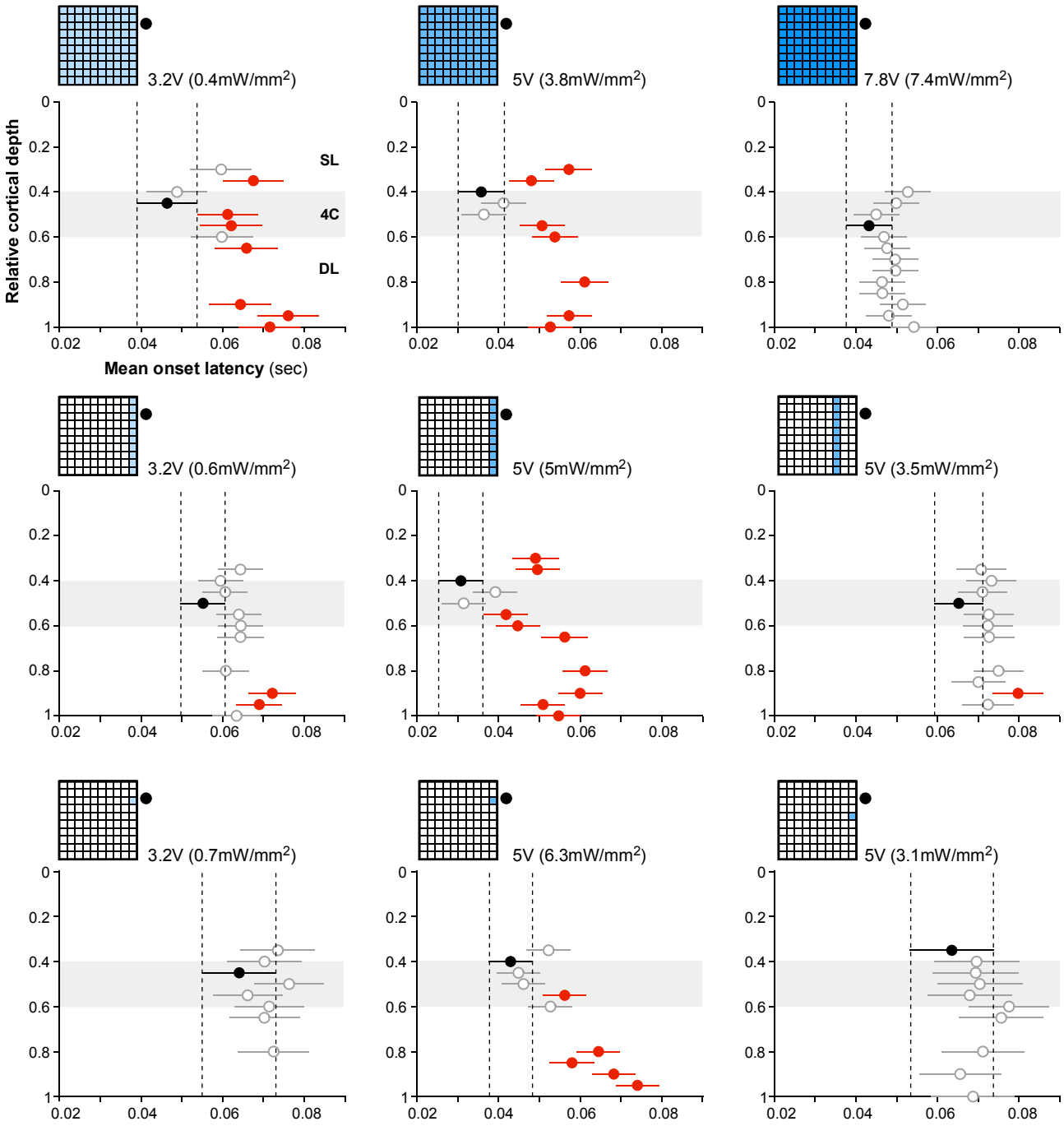
(A) Image of the brain after explantation of the UOA at the end of the *in vivo* testing experiment. The *asterisk* marks the center of the UOA implantation site. The white area at the site of the explantation is the *Duragel* that was placed on the brain after insertion of the UOA to protect the cortical surface and prevent deydration. The *dashed line* marks the border between V2 and V1. **(B)** The UOA after explantation. The *white dot* indicates the approximate location of the LEA used for penetration 2 (P2) relative to the UOA. The shank at the top right corner (column 1 row 10) was broken prior to the UOA insertion. The remaining shanks are intact.



Extended Data Figure 3

Laminar and tangential location of UOA and LEA penetrations

(A) **Left:** The locations of the UOA (*white box*) and of 3 LEA penetrations *(P1-P3, cyan dashed circles*) are shown on two cytochrome-oxidase (CO)-stained tangential sections through V1 and V2 (top section is more superficial). *Solid white contour*: V1/V2 border; *dashed white contours* delineate V1 layers (indicated). The location of column 1 (*C1*) and row 10 (*R10*) are indicated by *arrows*. *White dots* mark the locations of UOA needles visible in these sections as cortical damage. *Black dots* mark the location of UOA needles visible in more superficial sections, but not in these sections. Note that the majority of white dots are located in layer 4C in both sections. There are no white dots in column 1 in section #12 and only 2 in section #9, as the needle tips in this column terminated in the sections just above, thus in the superficial part of L4C. The postero-lateral half of the UOA terminated in slightly more superficial layers compared to its anterio-medial half. *A:*anterior; *M:* medial. **Right:** Same CO-stained section as shown on the left, with superimposed image of the same section viewed under tdT fluorescence. The fluorescent image was rendered transparent in Adobe Photoshop. P2 and P3 were located inside or near, respectively, the region of tdT/ChR2 expression, whereas P1 was more distant from it; accordingly, only the neurons recorded in P2 and P3, but not in P1, could be modulated by the laser. P2 was located about 1-1.1 mm medial to the nearest UOA needle (C1-R8, C1-R9, in these sections), while P3 was located about 800µm from the UOA (C1-R5 and C1-R4 are the nearest needles to P3 in these sections, but, as this penetration was not vertical, the more superficial LEA contacts were closer to needles C1-R6 and C1-R7 and more distant from the UOA, as also indicated by the physiological recordings in (C), and in **Fig. 3F**). The *asterisk* in all panels marks a crack in the tissue caused by histological processing, not by the UOA insertion. **(B)** **Left:** Relative cortical depth of each contact on the LEA in P2 is plotted versus the increase in firing rate caused by stimulation of single µLEDs along column 1 (*inset*). Different color traces are data for different µLEDs (rows 3-9) at 5 or 7.8V stimulation intensity (the most distant µLEDs only evoked responses at the higher intensity). µLED C1-R8 evoked the max response, indicating this needle tip was the closest to P2. **Right:** Relative cortical depth on the LEA-P2 is plotted versus the mean onset latency (*red*) or the mean onset latency reliability (*blue*; inverse of the SD of the distribution of pulse by pulse onset latencies) of responses at each contact evoked by stimulation of the whole µLED array; means are averages across all photostimulation intensities ≤5V. The shortest and most reliable response latencies are for contacts in L4C, indicating the UOA tips nearest P2 ended in this layer. **(C)** Same as in (B) but for P3. The data indicate that the µLED closest to P3 was C1-R7 whose tip terminated in the superficial layers.



Extended Data Figure 4

Statistical analysis of onset latencies for penetration 2 (data shown in Figure 4C)
Mean onset latency (± s.e.m) for each contact in P2 which showed significant response to UOA stimulation, for the UOA stimulation condition indicated by the *insets* at the top left of each plot. The mean latency was estimated from distributions of single-trial latency estimates. The *black dot* indicates the contact with the shortest latency in each condition. The *red dots* indicate the contacts that showed a statistically significant (Tukey HSD test) pairwise difference with the shortest latency contact (*black dot*), and the *gray dots* the contacts that did not differ significantly from the black dot. The *vertical dashed lines* indicate the points beyond which comparisons are significant.